

Applied Statistics for Civil and Environmental Engineers

## SECOND EDItION

Nathabandu T. Kottegoda and Renzo Rosso

## Problem Solution Manual

BlackwellPublithing

# Applied Statistics for Civil and Environmental Engineers <br> Problem Solution Manual <br> by N.T. Kottegoda and R. Rosso <br> Chapter 1 - `Preliminary Data Analysis 

1.1. Earthquake records. Measurements of engineering interest have been recorded during earthquakes in Japan and in other parts of the world during the period 1802 to 1968. One of the critical recordings is of apparent relative density, RDEN. After the commencement of a strong earthquake, a saturated fine, loose sand undergoes vibratory motion and consequently the sand may liquefy without retaining any shear strength, thus behaving like a dense liquid. This will lead to failures in structures supported by the liquefied sand. These are often catastrophic. The standard penetration test is used to measure RDEN. Another measurement taken to estimate the prospect of liquefaction is that of the intensity at which the ground shakes. This is the peak surface acceleration of the soil during the earthquake, ACCEL. The data are from J.T. Christian and W.F.Swiger (1975) in J. Geotech. Eng. Div., Proc. ASCE, 101, GT111, 1135-1150 and are reproduced by permission of the publisher (ASCE):

| RDEN <br> (\%) | ACCEL <br> (units of $\mathbf{~}$ ) | RDEN <br> (\%) | ACCEL <br> (units of $\mathbf{~}$ ) | RDEN <br> (\%) | ACCEL <br> (units of g) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 53 | 0.219 | 30 | 0.138 | 50 | 0.313 |
| 64 | 0.219 | 72 | 0.422 | 44 | 0.224 |
| 53 | 0.146 | 90 | 0.556 | 100 | 0.231 |
| 64 | 0.146 | 40 | 0.447 | 65 | 0.334 |
| 65 | 0.684 | 50 | 0.547 | 68 | 0.419 |
| 55 | 0.611 | 55 | 0.204 | 78 | 0.352 |
| 75 | 0.591 | 50 | 0.170 | 58 | 0.363 |
| 72 | 0.522 | 55 | 0.170 | 80 | 0.291 |
| 40 | 0.258 | 75 | 0.192 | 55 | 0.314 |
| 58 | 0.250 | 53 | 0.292 | 100 | 0.377 |
| 43 | 0.283 | 70 | 0.299 | 100 | 0.434 |
| 32 | 0.419 | 64 | 0.292 | 52 | 0.350 |
| 40 | 0.123 | 53 | 0.225 | 58 | 0.334 |

g denotes acceleration due to gravity ( 9.81 meters per second per second)
Compute the sample mean $\bar{X}$, standard deviation $\hat{S}$ and the coefficient of skewness $g_{1}$ for RDEN and ACCEL. Construct stem-and-leaf plots for each set. Comment on the distributions. Plot the scatter diagram and calculate the correlation coefficient $r$. What conclusions can be reached?

## Solution.

Stem-and-leaf-plot for RDEN: For stem 10\%; for leaves 1\%

```
    2 3|0 2
    7 4|0 00 0 3 4
(15) 5
17 6
11 700 2 2 5 5 8
5 80
4 90
310|0 0
```

Stem-and-leaf plot for ACCEL: For stem 0.1 units of $g$; for leaves 0.01 units of $g$

```
    7 1|\lllllll
19
(9) 3|0}10
114}
6}
2 6|1 8
```

RDEN
ACCEL

| Mean | Std dev | Skewness coef | Cor coef |
| ---: | :---: | :---: | :---: |
| 61.0 | 17.39 | 0.63 | 0.28 |
| 0.327 | 0.142 | 0.70 |  |

Scatter diagram plot is shown in Fig. P1.1.

P1.1 Scatter diagram of Rden vs. Accel


Fig. P1.1
Poor correlation; low coefficients of variation; positive skewness; distribution of the gamma or similar type.
1.2. Flood discharge. Annual maximum flood flows in the Po river at Pontelagoscuro, Italy over a 61year period from 1918 to 1978 are given in the second column of Table E.7.2. Compute the sample mean $\bar{X}$ and standard deviation $\hat{S}$. Sketch a histogram and the cumulative relative frequency diagram. Compute the quartiles and draw a box and whiskers plot. Comment on the distribution. Flood embankments along the banks of the river can withstand a flow of $5000 \mathrm{~m}^{3} / \mathrm{s}$. What is the probability that this will be exceeded during a 12 -month period?

Solution. Po river at Pontelagoscoro in Italy. Ranked annual maximum flow data in $\mathrm{m}^{3} / \mathrm{s}$ :

| 2,240 | 2,400 | 2,470 | 2,590 | 2,980 | 3,000 | 3,170 | 3,260 | 3,270 | 3,460 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 3,660 | 3,700 | 3,760 | 3,900 | 3,920 | 4,030 | 4,150 | 4,200 | 4,240 | 4,380 |
| 4,450 | 4,540 | 4,600 | 4,690 | 4,880 | 5,090 | 5,130 | 5,270 | 5,360 | 5,390 |
| 5,400 | 5,420 | 5,460 | 5,540 | 5,590 | 5,630 | 5,680 | 5,940 | 6,080 | 6,110 |
| 6,430 | 6,510 | 6,620 | 6,620 | 6,630 | 6,810 | 6,830 | 6,870 | 6,990 | 7,220 |
| 7,220 | 7,240 | 7,400 | 7,700 | 7,730 | 7,800 | 7,830 | 8,030 | 8,600 | 8,850 | 8,940

$\mathrm{n}=61$; Mean $=5408 \mathrm{~m}^{3} / \mathrm{s}$;
Standard deviation $=1735 \mathrm{~m}^{3} / \mathrm{s}$; coefficient of variation $=32 \%$, median $Q_{2}(31$ st value $)=5,400 \mathrm{~m}^{3} / \mathrm{s} ;$ range $r=8,940-2,220=6,720$; medians of the bottom and top halves of the ranked data:

$$
Q_{1}=(3,920+4,030) / 2=3,975 \mathrm{~m}^{3} / \mathrm{s} ; \quad Q_{2}=(6,810+6,830) / 2=6,820 \mathrm{~m}^{3} / \mathrm{s}:
$$

interquartile range, iqr $=6,820-3,975=2,845 \mathrm{~m}^{3} / \mathrm{s}$;
$n_{c}=1+3.3 \log _{10} 61=6.89$ or
$n_{c}=r n^{1 / 3} /(2$ iqr $)=6720 \times 61^{1 / 3} /(2 \times 2845)=4.65$.
Take $n_{c}=6$ : class width $=r / 6=1,120 \mathrm{~m}^{3} / \mathrm{s}$.
$\begin{array}{lrlllllllll}\text { Class limits: } & 2,220 & 3,360 & 4,460 & 5,580 & 6,700 & 7,820 & 8,940 \mathrm{~m}^{3} / \mathrm{s} \\ \text { Values within limits: } & 9 & 12 & 13 & 11 & 11 & 5\end{array}$
Sketch of histogram:



Cumulative plot as in Fig.1.1.6.
$\begin{array}{llllllll}2,000 & 3,000 & 4,000 & 5,000 & 6,000 & 7,000 & 8,000 & 9,000 \\ 10,000 & \mathrm{~m}^{3} / \mathrm{s} .\end{array}$ Distribution has a small positive skewness.

Sketch of box-and-whiskers plot:


Frequency based probability of exceedance of $5,000 \mathrm{~m}^{3} / \mathrm{s}=36 / 61=0.59$.
1.3 Flood discharge. The following are the annual maximum flows in $\mathrm{m}^{3} / \mathrm{s}$ in the Colorado River at Black Canyon for the 52-year period from 1878 to 1929:

| 1980 | 1130 | 3120 | 2120 | 1700 | 2550 | 8500 | 3260 | 3960 | 2270 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 1700 | 1570 | 2830 | 2120 | 2410 | 2550 | 1980 | 2120 | 2410 | 2410 |
| 1420 | 1980 | 2690 | 3260 | 1840 | 2410 | 1840 | 3120 | 3290 | 3170 |
| 1980 | 4960 | 2120 | 2550 | 4250 | 1980 | 4670 | 1700 | 2410 | 4550 |
| 2690 | 2270 | 5660 | 5950 | 3400 | 3120 | 2070 | 1470 | 2410 | 3310 |
| 3230 | 3090 |  |  |  |  |  |  |  |  |

[Adapted from Gumbel E.J. (1954): "Statistical Theory of Extreme Values and Some Practical Applications," National Bureau of Standards, Applied Mathematics Series 33, U.S. Govt. Printing Office, Washington D.C.].
Compute the mean $\bar{X}$ and standard deviation $\hat{S}$. Sketch a histogram and the relative frequency diagram. Compute the quartiles and draw a box-and-whiskers plot. How does this distribution differ from that of Problem 1.2?

Solution. Colorado river at Black Canyon, USA. Ranked annual maximum flow data in $\mathrm{m}^{3} / \mathrm{s}$ :

| 1,130 | 1,420 | 1,470 | 1,560 | 1,700 | 1,700 | 1,700 | 1,840 | 1,840 | 1,980 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 1,980 | 1,980 | 1,980 | 1,980 | 2,070 | 2,120 | 2,120 | 2,120 | 2,120 | 2,270 |
| 2,270 | 2,410 | 2,410 | 2,410 | 2,410 | 2,410 | 2,410 | 2,550 | 2,550 | 2,550 |
| 2,690 | 2,690 | 2,830 | 3,090 | 3,120 | 3,120 | 3,120 | 3,170 | 3,230 | 3,260 |
| 3,260 | 3,290 | 3,310 | 3,460 | 3,960 | 4,250 | 4,530 | 4,670 | 4,960 | 5,660 |
| 5,950 | 8,500 |  |  |  |  |  |  |  |  |

$n=52$. Median $Q_{2}=(2,410+2,410)=2,410 \mathrm{~m}^{3} / \mathrm{s}$. Medians of the bottom and top halves of the ranked data:
$Q_{1}=(1,980+1,980) / 2=1,980 \mathrm{~m}^{3} / \mathrm{s} ; \quad Q_{3}=(3,230+3,260) / 2=3,245 \mathrm{~m}^{3} / \mathrm{s}:$
Range $r=8,500-1,130=7,370 \mathrm{~m}^{3} / \mathrm{s} ; \mathrm{iqr}=Q_{3}-Q_{1}=3,245-1,980=1,265 \mathrm{~m}^{3} / \mathrm{s}$.
Mean $=2,837 \mathrm{~m}^{3} / \mathrm{s} ;$ Standard deviation $=1,301 \mathrm{~m}^{3} / \mathrm{s}$;
Coefficient of variation $=45 \%$
$n_{c}=1+3.3 \log _{10} 52=6.66$ or
$n_{c}=r n^{1 / 3} /(2$ iqr $)=7370 \times 52^{1 / 3} /(2 \times 1265)=10.87$
Take $n_{c}=8$; class width $=r / 8 \approx 1,000 \mathrm{~m}^{3} / \mathrm{s}$. To demarcate high outliers add 1.5 iqr
$=1.5 \times 1265 \approx 1,900 \mathrm{~m}^{3} / \mathrm{s}$ to $Q_{3}$, hence demarcation point $=5,145 \mathrm{~m}^{3} / \mathrm{s}$
Sketch of histogram:



Sketch of box and whiskers plot:
(Data signposts:lowest value, 3 quartiles, outlier limit, outliers and highest value)


## Comparison:

Colorado river: higher coefficient of variation, high positive skewness, 3 outliers.
Po river: lower coefficient of variation, almost symmetrical distribution, no outliers.
1.4. Welding joints for steel. At the University of Birmingham, England, laboratory measurements were taken of the horizontal legs $x$ and vertical legs $y$ of numerous welding joints for steel buildings. The main objective was to make the legs equal to 6 mm . A part of the results is listed below in millimeters. The data were provided by Dr.A.G.Kamtekar.

```
x= 5.5,5.0,5.0,6.0,7.0,5.2, 5.5,5.5,6.0,6.0, 4.5,6.0, 5.5,7.7,7.5, 6.0,5.6,5.0,5.5,5.5,
    6.0,6.5,5.5,5.0, 5.5, 5.5, 6.5, 6.5,7.0, 5.5, 6.5,5.5,6.0, 6.5, 8.5, 5.0, 6.0, 6.5, 5.0, 7.0,
    5.0,5.0,6.5,6.5,6.0, 4.7, 8.0,7.0, 5.5,7.0, 6.6,6.5,7.0,6.0,6.5,5.0,7.0,7.5,7.0,7.0
y = \quad 6 . 5 , 6 . 5 , 5 . 5 , 7 . 5 , 6 . 0 , 7 . 0 , 5 . 0 , 8 . 0 , 6 . 7 , 7 . 8 , 5 . 7 , 6 . 5 , 5 . 5 , 8 . 0 , 8 . 0 , 6 . 3 , 6 . 0 , 6 . 0 , 6 . 0 , 5 . 5 ,
    6.5,6.0,6.0,6.0,6.0, 6.5, 6.5,6.0, 6.0, 6.5,7.5,7.5, 6.0, 4.5, 7.0, 7.0, 6.0, 4.0, 4.0, 7.0,
    7.0,6.5,7.0,5.0, 5.0, 5.7,5.0, 5.0,6.0,7.0,6.0,7.0,6.0,5.5,6.0, 4.0, 5.5, 8.0, 7.5, 6.5
```

Draw a scatter diagram for these data. Draw a line through the ideal point ( $x=y=6 \mathrm{~mm}$ ) and the origin. Draw two lines through the origin that are symmetrical about the first line and envelope all of the points. Comment on the results.

Draw the cumulative sum (cusum) plots,

$$
C x_{n}=\sum_{i}^{n}\left(x_{i}-\mu_{x}\right) \quad \text { and } \quad C y_{n}=\sum_{i}^{n}\left(y_{i}-\mu_{y}\right)
$$

for $n=1,2, \ldots, 60$ and $\mu_{X}=\mu_{y}=6$. Let

$$
d x_{n}=C x_{n}-\min _{i=1}^{n-1}\left[C x_{i}\right]
$$

and the critical limit be $\max \left(d x_{n}\right)=12 \mathrm{~mm}$. Is the critical limit reached? Repeat for the vertical legs $y$. (Further details of cusum plots are given by Woodall, W.H., and B.M. Adams (1993), "The statistical design of cusum charts," Quality Eng., Vol. 5(4), pp.550-570; the associated control chart is the subject of Problem 5.11)

Solution. See 3 graphs annexed:
Poor correlation with larger deviations for the $x$ legs; $x$ legs are relatively higher than $y$ legs;
Cusum plots: for $x$ legs the critical limit is not reached for $y$ legs the critical limit is reached



Problem 1.4 Cusum charts

1.5. Frost frequency. Excessive frost can be harmful to roads. Frequencies of the number of days of frost during April in Greenwich, England, over a 65 year period are given by C.E. Brooks and N.Carruthers (1953), Handbook of Statistical Methods in Hydrology, H.M.Stationary Office, London and are listed below:

| Days of frost: | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Frequency: | 15 | 11 | 5 | 11 | 7 | 6 | 2 | 3 | 2 | 1 | 2 |

Draw a line diagram of the data. Comment on the results. Compute the mean number of days of frost in 1April. What is the probability of a frostfree April in a given year? What change would you expect in the frequency distribution for a month in midwinter?

## Solution.

Line diagram: frequency vs. number of days of frost


The occurrences of frosts in April have a geometric type of distribution with mean $\bar{x}=(0 \times 15+1 \times 11+2 \times 5+3 \times 11+4 \times 7+5 \times 6+6 \times 2+7 \times 3+8 \times 2+9 \times 1+10 \times 2) / 65=2.92$ The probability of a frost free April is $15 / 65=0.23$. In midwinter the shape of the frequency distribution is expected to be reversed with lower frequencies for zero or low numbers of days of frosts
1.6. Concrete cube test. From 28 -day concrete cube tests made in England in 1990, the following results of maximum load at failure in kilonewtons and compressive strength in newtons per square millimeter were obtained:

Maximum load: 950, 972, 981, 895, 908, 995, 646, 987, 940, 937, 846, 947, 827, 961, 935, 956

Compressive strength: $42.25,43.25,43.50,39.25,40.25,44.25,28.75,44.25,41.75,41.75$, 38.00, 42.50, 36.75, 42.75, 42.00, 33.50

The data were supplied by Dr.J.E.Ash, University of Birmingham, England.
Calculate the means $\bar{X}$, standard deviations $\hat{S}$, mean absolute deviations $d$ and the coefficients of skewness $g_{1}$. Draw two stem-and-leaf plots of the data. Draw a scatter diagram and calculate the coefficient of correlation. Are there any unexpected results?

## Solution.

Sketch of stem-and-leaf-plot for maximum load

```
16*|5
17
1 7*
283
3 8* 5
(5) \(9 \times 10 \begin{array}{llll}9 & 4 & 4 & 4\end{array}\)
    \begin{tabular}{ll|lllllll}
8 & \(9 *\) & 5 & 5 & 6 & 6 & 7 & 8 & 9
\end{tabular}
    \(110 \mid 0\)
```

Stem-and-leaf-plot for compressive strength
02
12*|9
23 4
5 3*|789
(11) $4 \left\lvert\, \begin{array}{lllllllllll}0 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & 4 & 4 & 4\end{array}\right.$

Alternative stem-and-leaf-plot for compressive strength


|  | Mean | Std dev | Mean abs dev | Skew coef | Cor coef |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Maximum load | 918 | 86 | 58 | -2.1 | 0.846 |
| Compressive strength | 40 | 4.2 | 3.2 | -1.5 |  |

Negative skewness, positive correlation and $10 \%$ coefficient of variation.


See scatter diagram for Problem 1.6. One suspected outlier.
1.7. Timber strength. For the timber strength data of Table E.1.1 determine the following measures of dispersion:
(a) Interquantile range, iqr
(b) Mean absolute deviation, $d$
(c) Gini's mean difference, $g$

Compare results with the standard deviation $\hat{s}$ of Table 1.2.2. Repeat these determinations after deleting the zero value. Rank the measures of dispersion in increasing order of susceptibility to the exclusion of the zero value on the basis of percentage change.

## Solution.

Interquartile range
Mean absolute deviation

|  | All data | Exc. zero | Perc. reduction exc. zero |
| ---: | ---: | ---: | :---: |
| iqr | 11.66 | 11.47 | 1.63 |
| $d$ | 7.55 | 7.36 | 2.48 |
| $g$ | 10.89 | 10.54 | 3.20 |
| $\hat{s}$ | 9.92 | 9.46 | 4.63 |

As expected the interquartile range is the most robust statistic.
1.8. Population growth. From past records, the population of an urban area has doubled every 10 years. Currently, it has a population of 200,000 . An engineer needs to make an estimate of the requirements for water supply during the next 23 years. What maximum population does one assume for the period?

## Solution.

$200000 \times 2^{23 / 10}=985000$
1.9. Traffic speed. The following is the frequency distribution of travel times of motor cars on the M1 motorway from Coventry, England to M10, St. Albans according to a survey conducted in England (see Ph.D. thesis of Andrew W. Evans, University of Birmingham, England, 1967).

Mean times (min): $53,58,63,68,73,78,83,88,93,98,103,108,113,118,123,128,133$, 138, 143, 148, 153, 158, 163, 168
Corresponding frequencies: $10,24,109,127,122,119,97,102,104,92,68,72,66,61,36$, $33,17,15,10,8,9,6,7,3$

Draw the histogram. Describe the salient features. What is the likely reason for the twin peaks? What inference can be made from the mean time interval between the two peaks?

## Solution.

Sketch of histogram:

$\begin{array}{lllllllllllll}53 & 63 & 73 & 83 & 93 & 103 & 113 & 123 & 133 & 143 & 153 & 163 & m i n s .\end{array}$
(Mean times of travel)
Shape of histogram with twin peaks indicate a mixed distribution.
Those driving non-stop and those stopping in rest areas.
From the mean distances between successive peaks at the top, the mean stopping time is about 25 minutes.
1.10. Average speed. On a certain country road that runs from a coastal town to a village in the mountains, the average speed of motor cars is $80 \mathrm{~km} / \mathrm{hr}$ uphill and $100 \mathrm{~km} / \mathrm{hr}$ downhill. What is the average speed for a journey from the town to the village and back?

## Solution.

Let distance $=x$ kilometers
Average speed $=2 x /(x / 80+x / 100)=88.9$ kilometers per hour, using the harmonic mean.
1.11. Annual rainfall. Catchment-averaged annual rainfall in the Po River basin of Italy for the 61year period from 1918 to 1978 are given in the penultimate column of Table E.7.2. Draw a stem-andleaf plot and a box plot of the data. Comment on the type of distribution.

## Solution.

Ranked mean annual rainfall in mm in the Po river basin in Italy from 1918 to 1978:

| 807 | 846 | 876 | 885 | 886 | 896 | 909 | 913 | 922 | 940 |
| ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
| 940 | 950 | 959 | 969 | 978 | 986 | 987 | 993 | 995 | 997 |
| 999 | 1011 | 1015 | 1017 | 1026 | 1028 | 1029 | 1046 | 1046 | 1051 |
| 1090 | 1096 | 1100 | 1110 | 1112 | 1123 | 1133 | 1133 | 1142 | 1159 |
| 1171 | 1196 | 1197 | 1215 | 1210 | 1228 | 1259 | 1264 | 1290 | 1318 |
| 1323 | 1345 | 1349 | 1356 | 1362 | 1422 | 1496 | 1501 | 1529 | 1564 |
| 1654 |  |  |  |  |  |  |  |  |  |

$n=61$. Median $Q_{2}=1,090 \mathrm{~mm}$. Medians of the bottom and top halves of the ranked data:
$Q_{1}=(978+986) / 2=982 \mathrm{~mm} ; \quad Q_{3}=(1,228+1,259) / 2=1,243.5 \mathrm{~mm}:$
Range $r=1,654-807=847 \mathrm{~mm}$; iqr $=Q_{3}-Q_{1}=1,243.5-982=261.5 \mathrm{~mm}$
.To demarcate high outliers add 1.5 iqr
$=1.5 \times 261.5=392.25 \mathrm{~mm}$ to $Q_{3}$; hence demarcation point $\approx 1,636 \mathrm{~mm}$

Stem-and-leaf plot

| 5 | 8 | 1 | 5 | 8 | 9 | 9 |  |  |  |  |  |  |  |  |
| ---: | ---: | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 18 | 9 | 0 | 1 | 1 | 2 | 4 | 5 | 5 | 6 | 7 | 8 | 9 | 9 | 9 |
| $(13)$ | 10 | 0 | 0 | 0 | 1 | 2 | 2 | 3 | 3 | 3 | 5 | 5 | 5 | 9 |
| 30 | 11 | 0 | 0 | 0 | 1 | 3 | 3 | 3 | 4 | 6 | 7 |  |  |  |
| 20 | 12 | 0 | 0 | 2 | 2 | 3 | 6 | 6 | 9 |  |  |  |  |  |
| 12 | 13 | 2 | 2 | 5 | 5 | 6 | 6 |  |  |  |  |  |  |  |
| 6 | 14 | 2 |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | 15 | 0 | 0 | 3 | 6 |  |  |  |  |  |  |  |  |  |
| 1 | 16 | 5 |  |  |  |  |  |  |  |  |  |  |  |  |

Sketch of box and whiskers plot:
(Data signposts:lowest value, 3 quartiles, high outlier limit, 1 outlier)

|  |  |  |
| :---: | :---: | :---: | :---: |
| 982 | 1,09 | $1,243.5$ |
| $Q_{1}$ | $Q_{2}$ | $Q_{3}$ |


$8009001,0001,1001,2001,3001,4001,5001,6001,700 \mathrm{~mm}$


Comments: The distribution is not symmetrical. It has a long right tail, The highest value is a suspected outlier.
1.12. Rock test. A contractor engaged in building part of a sewer tunnel claimed that the rock was harder than described in his contract with a District Council in the United Kingdom and thus more work was required to construct the tunnel than anticipated. An independent company made tests to verify the contractor's claim. Among these were uniaxial compressive strengths, of which 123 specimens are listed here, in meganewtons per square meter.
2.40, 22.08, 16.80, 4.80, 21.36, 9.12, 9.36, 3.60, 15.36, 15.60, 6.24, 9.84, 16.08, 30.00, 20.40, $12.96,19.20,10.32,15.84,62.40,40.80,4.80,7.20,8.88,14.40,14.88,5.76,18.72,12.48$, $11.04,8.64,19.20,8.16,18.96,8.64,12.00,14.88,17.52,12.48,13.44,9.36,11.28,8.88$, $15.12,9.36,17.28,26.40,4.32,11.28,7.92,13.92,11.76,9.60,8.40,9.84,27.60,6.00,14.40$, $8.88,17.04,12.48,9.84,10.80,12.24,12.00,13.20,11.28,11.7611 .76,8.00,9.36,15.12$, $11.52,16.08,10.8014 .64,8.40,13.44,10.56,9.12,13.44,12.72,13.68,11.28,5.52,11.04$, $12.00,7.20,8.64,11.76,8.64,7.68,7.68,13.92,6.48,7.20,7.92,9.60,8.64,9.12,12.96$, $9.36,14.64,9.12,8.88,20.40,17.28,8.64,11.76,7.92,7.68,11.04,12.48,14.40,9.84,9.12$, 8.40, 12.00, 4.80, 12.72, 9.60, 8.64, 9.84

Draw histograms using Eqs. (1.1.1) and (1.1.2) for the class widths. What do you notice about the histograms in general? Draw a box and whiskers plot. What evidence is there to support the contractor's claim?

## Solution.

Ranked uniaxial compressive strengths in meganewtons per square meter:

| 2.40 | 3.60 | 4.32 | 4.80 | 4.80 | 4.80 | 5.52 | 5.76 | 6.00 | 6.24 |
| ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
| 6.48 | 7.20 | 7.20 | 7.20 | 7.68 | 7.68 | 7.68 | 7.92 | 7.92 | 7.92 |
| 8.00 | 8.16 | 8.40 | 8.40 | 8.40 | 8.64 | 8.64 | 8.64 | 8.64 | 8.64 |
| 8.64 | 8.65 | 8.88 | 8.88 | 8.88 | 8.88 | 9.12 | 9.12 | 9.12 | 9.12 |
| 9.12 | 9.36 | 9.36 | 9.36 | 9.36 | 9.36 | 9.60 | 9.60 | 9.60 | 9.84 |
| 9.84 | 9.84 | 9.84 | 9.84 | 10.32 | 10.56 | 10.80 | 10.80 | 11.04 | 11.04 |
| 11.04 | 11.28 | 11.28 | 11.28 | 11.28 | 11.52 | 11.76 | 11.76 | 11.76 | 11.76 |
| 11.76 | 12.00 | 12.00 | 12.00 | 12.00 | 12.24 | 12.48 | 12.48 | 12.48 | 12.48 |
| 12.72 | 12.72 | 12.96 | 12.96 | 13.20 | 13.44 | 13.44 | 13.44 | 13.68 | 13.92 |
| 13.92 | 14.40 | 14.40 | 14.40 | 14.64 | 14.64 | 14.88 | 14.88 | 15.12 | 15.12 |
| 15.36 | 15.60 | 15.84 | 16.08 | 16.08 | 16.80 | 17.04 | 17.28 | 17.28 | 17.52 |
| 18.72 | 18.96 | 19.20 | 19.20 | 20.40 | 20.40 | 21.36 | 22.08 | 26.40 | 27.60 |
| 30.00 | 40.80 | 62.40 |  |  |  |  |  |  |  |

$n=123$. Median $Q_{2}=11.28$. Medians of the bottom and top halves of the ranked data: $Q_{1}=8.64 ; \quad Q_{3}=14.40$.
Range $r=62.40-2.40=60.00 ; \mathrm{iqr}=Q_{3}-Q_{1}=14.40-8.64=5.76$
$n_{c}=1+3.3 \log _{10} 123=7.9$ or
$n_{c}=r n^{1 / 3} /(2 i q r)=60 \times 123^{1 / 3} /(2 \times 5.76)=25.9$
To demarcate high outliers add 1.5 iqr
$=1.5 \times 5.76 \approx 8.64$ to $Q_{3}$, hence demarcation point $=23.04$

Sketch histograms. (1)Take $n_{c}=8$; class width $=r / 8 \approx 8$

(2) Take $n_{c}=25$; class width $=r / 25 \approx 3$


Assymetrical histograms.

Sketch of box-and-whiskers plot:
(Data signposts:lowest value, 3 quartiles, outlier limit, 5 outliers incl. highest value)



Large skewness and 5 outliers. The shape of the histograms and the high number of outliers support the contractor's claim.
1.13. Soil erosion. Measurements taken on farm lands of the amounts of soil washed away by erosion suggest a relationship with flow rates. The following results are taken from G.R. Foster, W.R.Ostercamp and L.J.Lane, "Effect of Discharge Rate on Rill Erosion", Winter 1982 Meeting of the American Society of Agricultural Engineers:

| Flow (L/s) | 0.31 | 0.85 | 1.26 | 2.47 | 3.75 |
| :--- | :--- | :--- | :--- | :--- | :--- |
| Soil eroded (kg) | 0.82 | 1.95 | 2.18 | 3.01 | 6.07 |

Draw a plot of the data. Comment on the results.

## Solution.

Sketch of soil eroded in kilograms vs. flow in liters per second



Soil erosion tends to increase linearly with flow rate. Hence a relationship can be made for predictive purposes. However one should be cautioned about the validity of extrapolations beyond the range of flows observed.
1.14. Concrete cube test. The following 28-day compressive strengths, in newtons per square millimeter, were obtained from test results on concrete cubes in England:
50.5, 45.8, 49.6, 47.7, 54.0, 49.4, 54.1, 53.1, 56.5, 55.2, 52.7, 52.0, 54.2, 55.2, 53.4, 51.0, 53.1, 48.5, 51.0, 58.6, 52.5, 49.5, 51.1, 48.1, 50.2, 49.3, 47.3, 52.9, 52.8, 49.5, 48.8, 53.8, 47.3, 47.7, 52.2 45.7, 53.4, 48.5, 49.1, 43.3

The data were supplied by Dr.J.E.Ash, University of Birmingham, England
Compare these results with the compressive strengths in Table E.1.2 by drawing back-to-back stem-and-leaf plots. For this purpose, plot the foregoing results on the left of the stem with reference to Fig. 1.3.1 and omit the cumulative frequencies. Comment on the differences in the distributions.

## Solution.

Back-to-back stem-and-leaf plot of compressive strengths: Table E1.2 data on right:

| 3 | \| 43 |  |
| :---: | :---: | :---: |
|  | 44 |  |
| 87 | 45 |  |
|  | 46 |  |
| 7733 | 47 |  |
| 8551 | 48 |  |
| 655431 | 49 | 9 |
| 52 | 50 | 7 |
| 100 | 51 |  |
| 987520 | 52 | 5 |
| 84411 | 53 | 24 |
| 210 | 54 | 46 |
| 22 | 55 | 8 |
| 5 | 56 | 379 |
|  | 57 | 89 |
| 6 | 58 | 89 |
|  | 59 | 0688 |
|  | 60 | $\begin{array}{lllllll}0 & 2 & 5 & 5 & 9\end{array}$ |
|  | 61 | 159 |
|  | 62 |  |
|  | 63 | 34 |
|  | 64 | 99 |
|  | 65 | 7 |
|  | 66 |  |
|  | 67 | 23 |
|  | 68 | 139 |
|  | 69 | 5 |

Alternative back-to-back stem-and-leaf plot of strengths: Table E1.2 data on right


The two distributions of compressive strengths of concrete (given in $\mathrm{N} / \mathrm{mm}^{2}$ ). Both distributions have low coefficients of skewness. The mean and variance are higher for data from Table E1.2
1.15. Water quality. Water quality measurements are taken daily on the River Ouse at Clapham, England. The concentrations of chlorides and phosphates in solution, given below in milligrams per liter, are determined over a 30-day period.

Chloride: 64.0, 66.0, 64.0, 62.0, 65.0, 64.0, 64.0, 65.0, 65.0, 67.0, 67.0, 74.0 69.0, 68.0, 68.0, 69.0, 63.0, 68.0, 66.0, 66.0, 65.0, 64.0, 63.0, 66.0, 55.0, 69.0, 65.0, 61.0, 62.0, 62.0 Phosphate: 1.31, 1.39, 1.59, 1.68, 1.89, 1.98, 1.97, 1.99, 1.98, 2.15, 2.12, 1.90 1.92, 2.00, $1.90,1.74,1.81,1.86,1.86,1.65,1.58,1.74,1.89,1.94,2.07,1.58,1.93,1.72,1.73,1.82$

Compare the coefficients of variation $v$. Draw a scatter diagram and compute the correlation coefficient $r$. Comment on the results. Do you see any role in this association for predictive purposes?

## Solution.

Coefficients of variation of $4.9 \%$ and $10.7 \%$. Scatter diagram follows. $r=0.027$.


There is no role in the association for predictive purposes.
1.16. Timber strength. From the timber strength data of Table 1.1.3, compute the 3 percent trimmed mean by omitting 3 percent of the observations from the highest and the lowest extremities of the ranked data. Compute the standard deviation $\hat{S}$ and the coefficients of skewness $g_{1}$ and kurtosis $g_{2}$. Compare with the results for the full sample (as given in Table 1.2.2).

Solution.

|  | Mean | Std dev | Coef. of variation | Coef of skew | Coef of kurt |
| :--- | ---: | :--- | :---: | :---: | :---: |
| Sample 3\% trimmed | 8.94 | 7.94 | $20 \%$ | 0.14 | 2.48 |
| Full sample | 39.09 | 9.92 | $25 \%$ | 0.15 | 4.46 |

It is seen that trimming reduces all the statistics.
1.17. Concrete beam. Joist-hanger tests carried out at the University of Birmingham, England, on concrete beams gave observations of deflections in millimeters and failure load in kilograms. The following results pertain to $75 \mathrm{~mm} \times 150 \mathrm{~mm}$ hangers on which timber joists rest:

Failure load: 1903, 1665, 1903, 1991, 2229, 1910, 2025, 1991, 1882, 2032, 1896, 1346
Deflection: $0.69,0.67,0.80,0.50,0.74,0.78,0.57,0.91,0.54,0.50,0.97,0.62$
Determine by drawing a scatter diagram and computing the correlation coefficients whether there is any association between the two variables. Discuss your results.

## Solution.

Scatter diagram follows.

$r=0.069$. There is no role in the association for predictive purposes.
1.18. Hurricane frequency. Hurricane damage is of great concern to civil engineers. The frequencies of hurricanes affecting the east coast of the United States each year during a period of 69 years are given as follows by H.C.S.Thom (1966), Some Methods of Climatological Analysis, World Meteorological Organisation, Geneva:

| Number of hurricanes: | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Frequency: | 1 | 6 | 10 | 16 | 19 | 5 | 7 | 3 | 1 | 1 |

Draw a line diagram and comment on its form. Discuss differences or similarities between this diagram and Fig. 1.1.1.

## Solution.

Sketch of line diagram: frequency vs. number of hurricanes


More symmetrical than Fig. 1.1.1 in text. Can be approximated to a normal distribution.
1.19. Air pollution. On April 13, 1994, the following concentration of pollutants were recorded at eight stations of the monitoring system for pollution control located in the downtown area of Milan, Italy.

| Station: | Aquileia | Cenisio | Juvara | Liguria | Marche | Senato | Verziere | Zavattari |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\mathrm{NO}_{2}$, | 130 | 130 | 115 | 120 | 135 | 142 | 90 | 116 |
| ( $\mu \mathrm{g} / \mathrm{m}^{3}$ ) |  |  |  |  |  |  |  |  |
| $\begin{aligned} & \text { CO, } \\ & \left(\mathrm{mg} / \mathrm{m}^{3}\right) \end{aligned}$ | 2.9 | 4.4 | 3.6 | 4.1 | 3.3 | 5.7 | 4.8 | 7.3 |

Compare the coefficients of variation $v$ of the pollutants and determine their correlation $r$.

## Solution.

Coef. of variation Coef. of correlsation
$\mathrm{NO}_{2} \quad 1.2 \% \quad-0.15$
CO
3.0\%
$\mathrm{NO}_{2}$ and CO have low variation and are uncorrelated
1.20. Storm rainfall. The analysis of storm data is essential for predicting flood hazards in urban areas. Maximum rainfall depths recorded at Genoa University in Italy, for durations varying from 5 minutes to 3 hours, are presented here for the years 1974-1987.

| Maximum rainfall depth recorded at Genoa University, Italy (mm) |  |  |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Duration (min) | 5 | 10 | 20 | 30 | 40 | 50 | 60 | 120 | 180 |
| Year: |  |  |  |  |  |  |  |  |  |
| 1974 | 10.1 | 19.5 | 28.8 | 30.5 | 32.4 | 35.5 | 38.7 | 48.0 | 51.6 |
| 1975 | 17.9 | 20.0 | 26.7 | 31.2 | 34.7 | 38.2 | 40.2 | 55.0 | 56.0 |
| 1976 | 20.0 | 32.6 | 52.6 | 37.2 | 41.1 | 51.0 | 55.7 | 67.1 | 80.6 |
| 1977 | 5.1 | 13.6 | 16.0 | 21.3 | 90.1 | 108.8 | 118.9 | 146.5 | 157.3 |
| 1978 | 20.5 | 26.1 | 36.3 | 46.1 | 49.3 | 24.6 | 25.0 | 40.7 | 49.9 |
| 1979 | 10.0 | 15.7 | 20.9 | 25.0 | 30.5 | 38.0 | 45.6 | 65.2 | 90.1 |
| 1980 | 12.0 | 27.9 | 47.9 | 56.0 | 70.0 | 80.0 | 89.4 | 106.9 | 114.2 |
| 1981 | 10.0 | 14.4 | 20.0 | 23.3 | 25.1 | 26.4 | 27.2 | 34.3 | 41.2 |
| 1982 | 10.0 | 12.1 | 17.3 | 19.2 | 22.1 | 27.3 | 32.7 | 54.4 | 66.5 |
| 1983 | 20.1 | 32.8 | 60.0 | 65.7 | 76.1 | 92.8 | 105.7 | 122.3 | 122.3 |
| 1984 | 7.6 | 8.1 | 13.0 | 16.5 | 21.6 | 25.3 | 25.3 | 27.0 | 32.3 |
| 1985 | 8.7 | 11.7 | 20.0 | 22.9 | 26.1 | 26.3 | 27.6 | 41.1 | 56.7 |
| 1986 | 24.6 | 36.7 | 56.7 | 73.9 | 93.9 | 110.1 | 128.5 | 180.8 | 188.7 |
| 1987 |  |  |  |  |  |  |  |  |  |

Compute the mean $\bar{X}$ and standard deviation $\hat{S}$ and coefficient of skewness $g_{1}$ for each duration. Are there some regularities in the growth of these statistics with increasing duration? Comment on the results and the physical relevance to storm characteristics.

Solution.

| Duration, min | 5 | 10 | 20 | 30 | 40 | 50 | 60 | 120 | 180 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Mean, mm | 13.5 | 20.4 | 32.0 | 38.7 | 45.5 | 52.5 | 50.8 | 60.5 | 69.4 |
| Std dev, mm | 5.9 | 9.2 | 16.1 | 20.4 | 26.0 | 31.7 | 36.5 | 47.3 | 48.9 |
| Coef of var, \% | 44 | 45 | 50 | 53 | 57 | 60 | 72 | 78 | 70 |
| Coef of skew | 0.50 | 0.49 | 0.60 | 0.68 | 0.85 | 0.88 | 1.12 | 1.58 | 1.39 |

There is higher uncertainty for longer storms. The means and standard deviations increase with increasing durations, also distributions become more skewed. The lower variability of short bursts of rainfall suggest that rainfalls of short duration, 30 minutes or less, have similar physical characteristics.
1.21. Carbon dioxide. The records of atmospheric trace gases are used in the study of global climatic changes. Monthly carbon dioxide concentrations recorded at Mount Cimone, Italy, from 1980 to 1988 are given here.

Carbon dioxide concentration recorded at Mount Cimone, Italy, in parts per million in volume


Compute the mean $\bar{X}$ and standard deviation $\hat{S}$ for each year (by rows) and for each month (by columns). Because the temporal evolution of the annual mean indicates that carbon dioxide increases (probably resulting in global warming) compute the annual rate of increase. Comment on the results.

## Solution

Measurements in parts per million in volume

| Year | 1980 | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Mean | 337.0 | 338.2 | 340.2 | 341.3 | 345.1 | 345.2 | 346.5 | 348.7 | 350.2 |
| Std dev | 4.8 | 4.8 | 4.6 | 4.0 | 4.8 | 5.1 | 5.3 | 4.9 | 4.6 |

Annual increase in mean carbon dioxide concentration $=(350.2-337.0) / 8=1.65$

| Month | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dev |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Mean | 346 | 348 | 349 | 348 | 345 | 341 | 337 | 336 | 338 | 342 | 346 | 347 |
| Std dev | 4.9 | 5.9 | 5.0 | 4.5 | 4.9 | 4.7 | 5.4 | 5.2 | 4.1 | 4.6 | 5.0 | 4.2 |

This shows an increase in carbon dioxide in winter months
1.22. Historical records of earthquake intensity. Catalogo dei terremoti italiani dall'anno 1000 al 1980 ("Catalog of Italian earthquakes from year 1000 to 1980") was edited by D. Postpischl in 1985, and is available through the National Research Council of Italy. This directory contains all of the available historical information on earthquakes that occurred in Italy during the past (nearly) 1000 years. It also includes values of earthquake intensity in terms of the Mercalli-Canconi-Sieber (MCS) index. The following table gives the values of MCS intensity for the city of Rome.

| MCS intensity 2 |  | 3 | 4 |  | 5 |  | 6 |  | 7 |  | Total |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Century |  |  |  |  |  |  |  |  |  |  |  |
| XI |  |  |  |  | 2 |  |  |  |  |  | 2 |
| XII |  |  |  |  | 1 |  |  |  |  |  | 1 |
| XIII |  |  |  |  | 1 |  |  |  |  |  | 1 |
| XIV |  |  |  |  |  |  |  |  |  |  | 0 |
| XV |  |  |  |  | 1 |  | 1 |  | 1 |  | 3 |
| XVI |  |  |  |  |  |  |  |  |  |  | 0 |
| XVII |  |  |  |  | 1 |  |  |  |  |  | 1 |
| XVIII |  | 7 |  | 4 |  | 2 |  | 2 |  |  | 15 |
| XIX | 110 | 125 |  | 50 |  | 14 |  | 1 |  | 1 | 301 |
| XX | 3 |  |  | 2 |  |  |  |  |  |  | 5 |
| Total | 113 | 132 |  | 56 |  | 22 |  | 4 |  | 2 | 329 |

Draw the line diagram for the whole data and for those recorded in each century. Compare the data recorded in the eighteenth century with those recorded in the other centuries.

## Solution.

Line diagram for the whole data: frequency vs. MCS intensty


Line diagram for nineteenth century: frequency vs. MCS intensty


The nineteenth century was extraordinary. When the line diagram for this century is compared with that for the whole data from the eleventh century onwards, the differences in the frequencies are very small.
1.23. Sea waves. Because of scarcity of records, the characteristics of sea waves are often derived from other climatological data. For the purpose, the SMB method (named after Sverdrup, Munk, and Bretschneider) is widely used in engineering practice [see: U.S. Army Corps of Engineers(1977), Shore Protection Manual, Vol. 1, Coastal Engineering Research Center, Washington, DC]. Liberatore and Rosso (1983) used this model to simulate sea waves in the upper Adriatic Sea. They investigated two different strategies for model calibration, called "no.1" and "no.2" in the table presented here. The table also includes the observed and the simulated values of the height of the highest sea wave and of its period for measurements taken from August 1977 to September 1978.

| Measured values |  | Simulated values |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Calibration strategy no. 1 |  |  |  |
| Height $(\mathbf{m})$ | Period $(\mathbf{s})$ | Height $(\mathbf{m})$ | Period $(\mathbf{s})$ | Height $(\mathbf{m})$ | Period $\mathbf{( s )}$ |
| 2.26 | 6.1 | 1.81 | 5.4 | 1.54 | 5.8 |
| 3.10 | 4.3 | 2.93 | 6.8 | 2.54 | 6.4 |
| 3.22 | 5.7 | 3.24 | 7.2 | 2.80 | 6.7 |
| 3.84 | 7.7 | 3.18 | 7.1 | 2.69 | 6.6 |
| 2.56 | 5.3 | 2.74 | 6.6 | 2.32 | 6.1 |
| 2.74 | 5.7 | 3.49 | 7.4 | 3.00 | 6.9 |
| 2.28 | 4.9 | 2.12 | 5.8 | 1.80 | 5.4 |
| 3.88 | 6.7 | 5.10 | 9.0 | 4.43 | 8.4 |
| 2.49 | 5.0 | 2.14 | 5.8 | 1.81 | 5.4 |
| 4.22 | 6.9 | 4.45 | 8.8 | 3.77 | 7.7 |
| 2.01 | 5.0 | 2.57 | 6.4 | 2.19 | 5.9 |
| 2.77 | 5.9 | 2.68 | 6.5 | 2.27 | 6.0 |
| 3.61 | 6.5 | 3.86 | 7.8 | 3.36 | 7.3 |
| 3.51 | 7.4 | 4.02 | 8.0 | 3.51 | 7.5 |
| 2.52 | 5.0 | 3.39 | 7.3 | 2.95 | 6.9 |
| 2.12 | 5.1 | 2.61 | 6.5 | 2.21 | 6.0 |
| 2.73 | 6.5 | 2.22 | 6.0 | 1.88 | 5.5 |
| 3.30 | 5.4 | 4.05 | 8.0 | 3.49 | 7.5 |

Draw a scatter diagram to compare the observed and simulated values of wave heights and periods. Compute the correlation coefficients $r$. Compute the deviations of the simulated data from the observed data, and find the mean $\bar{X}_{1}$, standard deviation $\hat{S}_{1}$ and coefficient of variation $v$ of these deviations. Do these results indicate which of the two investigated strategies provides the better representation of sea waves from climatological data?

Solution.

|  | H meas | Pmeas | $\mathrm{H}_{1}$ | $\mathrm{H}-\mathrm{H}_{1}$ | $\mathrm{H}_{2}$ | $\mathrm{H}-\mathrm{H}_{2}$ | $\mathrm{P}_{1}$ | $\mathrm{P}-\mathrm{P}_{1}$ | $\mathrm{P}_{2}$ | $\mathrm{P}-\mathrm{P}_{2}$ |
| :--- | :--- | :--- | :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 2.26 | 6.1 | 1.81 | 0.45 | 1.54 | 0.72 | 5.4 | 0.7 | 5.8 | 0.3 |
|  | 3.10 | 4.3 | 2.93 | 0.17 | 2.54 | 0.56 | 6.8 | -2.5 | 6.4 | -2.1 |
|  | 3.22 | 5.7 | 3.24 | -0.02 | 2.80 | 0.42 | 7.2 | -1.5 | 6.7 | -1.0 |
|  | 3.84 | 7.7 | 3.18 | 0.66 | 2.69 | 1.15 | 7.1 | 0.6 | 6.6 | 1.1 |
|  | 2.56 | 5.3 | 2.74 | -0.18 | 2.32 | 0.24 | 6.6 | -1.3 | 6.1 | -0.8 |
|  | 2.74 | 5.7 | 3.49 | -0.75 | 3.00 | -0.26 | 7.4 | -1.7 | 6.9 | -1.2 |
|  | 2.28 | 4.9 | 2.12 | 0.16 | 1.80 | 0.48 | 5.8 | -0.9 | 5.4 | -0.5 |
|  | 3.88 | 6.7 | 5.10 | -1.22 | 4.43 | -0.55 | 9.0 | -2.3 | 8.4 | -1.7 |
|  | 2.49 | 5.5 | 2.14 | 0.35 | 1.81 | 0.68 | 5.8 | -0.3 | 5.4 | 0.1 |
|  | 4.22 | 6.9 | 4.45 | -0.23 | 3.77 | 0.45 | 8.8 | -1.9 | 7.7 | -0.8 |
|  | 2.01 | 5.0 | 2.57 | -0.56 | 2.19 | -0.18 | 6.4 | -1.4 | 5.9 | -0.9 |
|  | 2.77 | 5.9 | 2.68 | 0.09 | 2.27 | 0.50 | 6.5 | -0.6 | 6.0 | -0.1 |
|  | 3.61 | 6.5 | 3.86 | -0.25 | 3.36 | 0.25 | 7.8 | -1.3 | 7.3 | -0.8 |
|  | 3.51 | 7.4 | 4.02 | -0.51 | 3.51 | 0.00 | 8.0 | -0.6 | 7.5 | -0.1 |
|  | 2.52 | 5.0 | 3.39 | -0.87 | 2.95 | -0.43 | 7.3 | -2.3 | 6.9 | -1.9 |
|  | 2.12 | 5.1 | 2.61 | -0.49 | 2.21 | -0.09 | 6.5 | -1.4 | 6.0 | -0.9 |
|  | 2.73 | 6.5 | 2.22 | 0.51 | 1.88 | 0.85 | 6.0 | 0.5 | 5.5 | 1.0 |
|  | 3.30 | 5.4 | 4.05 | -0.75 | 3.49 | -0.19 | 8.0 | -2.6 | 7.5 | -2.1 |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| mean | 2.953 | 5.867 | 3.144 | -0.191 | 2.698 | 0.256 | 7.022 | -1.156 | 6.556 | -0.689 |
| stdev | 0.660 | 0.925 | 0.893 | 0.529 | 0.786 | 0.471 | 1.020 | 1.038 | 0.872 | 0.950 |
| cv\% | 22.3 | 15.8 | 28.4 | 276 | 29.1 | 184 | 14.5 | 89.8 | 0.133 | 138 |
|  |  |  |  |  |  |  |  |  |  |  |
| Cor | 0.809 |  |  |  |  |  | 0.443 |  |  |  |

## Prob 1.23 Measured sea waves vs. simulat ed



Summary of statistics of deviations from observed heights and periods F rom above table

|  | Strategy 1 |  | Strategy 2 |  |
| :--- | :---: | :---: | :---: | :---: |
|  | Height, $m$ Period, $s$ | Height, $m$ | Period, $s$ |  |
| Coef of cor with | 0.81 | 0.43 | 0.80 | 0.44 |
| $\quad$ measured data | 0.191 | 1.156 | 0.256 | 0.689 |
| Mean | 0.529 | 1.038 | 0.471 | 0.950 |
| Std dev <br> Coef of var, \% | 276 | 90 | 184 | 137 |

Very high variabilities of the deviations from observed heights, also in the case of periods of waves. The coefficients of correlation are artificially high. There is no clear indication of the better strategy.
1.24. Surveying. A triangulated network is used to determine the position of three points in space, denoted by $\mathbf{u}_{1} \equiv\left(x_{1}, y_{1}\right), \mathbf{u}_{2} \equiv\left(x_{2}, y_{2}\right)$, and $\mathbf{u}_{3} \equiv\left(x_{3}, y_{3}\right)$, by measuring their mutual distances, and their distances from two reference points, $\mathbf{u}_{A} \equiv\left(x_{A}, y_{A}\right)$,and $\mathbf{u}_{B} \equiv\left(x_{B}, y_{B}\right)$, as shown in Fig. 1.P1.


The Cartesian coordinates of the reference points are $x_{A}=y_{A}=0, x_{B}=92$, and $y_{B}=40 \mathrm{~m}$. The table of the measured distances is given next.

|  | $\mathbf{u}_{A}$ | $\mathbf{u}_{B}$ | $\mathbf{u}_{\mathbf{1}}$ | $\mathbf{u}_{2}$ | $\mathbf{u}_{3}$ |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{u}_{A}$ | $\mathbf{0}$ | $\mathbf{1 0 0}$ | 50 | 71 | 92 |
| $\mathbf{u}_{B}$ | $\mathbf{1 0 0}$ | $\mathbf{0}$ | 86 | 70 | 40 |
| $\mathbf{u}_{1}$ | 50 | 86 | $\mathbf{0}$ | 26 | 99 |
| $\mathbf{u}_{2}$ | 71 | 70 | 26 | $\mathbf{0}$ | 93 |
| $\mathbf{u}_{3}$ | 92 | 40 | 99 | 93 | $\mathbf{0}$ |

Using appropriate trigonometric methods find the average location and coefficients of variation of the coordinates of point $\mathbf{u}_{1} \equiv\left(x_{1}, y_{1}\right)$.

## Solution.

Use $A=2 \operatorname{atan}(\operatorname{sqrt}((s-b)(s-c) / s /(s-a))) \quad$ where $s=(a+b+c)$.

| $1 \mathrm{a} 3=1.444044$ | $\mathrm{a} 13=1.172796$ | $\mathrm{a} 31=0.524752$ | total $=3.141593$ |
| :--- | :--- | :--- | :--- |
| $\mathrm{ba} 3=0.411500$ | $\mathrm{ab} 3=1.167992$ | $\mathrm{a} 3 \mathrm{~b}=1.562101$ | total $=3.141593$ |
| $\mathrm{ba} 1=1.035146$ | $\mathrm{ab} 1=0.523557$ | $\mathrm{a} 1 \mathrm{~b}=1.582890$ | total $=3.141593$ |
| $\mathrm{oa} 2=0.258003$ | $\mathrm{a} 12=2.370727$ | $\mathrm{a} 21=0.512863$ | total $=3.141593$ |
| $\mathrm{ba} 2=0.775380$ | $\mathrm{a} 2 \mathrm{~b}=1.576732$ | $\mathrm{ab} 2=0.789480$ | total $=3.141593$ |
| $\mathrm{ta} 3=1.189948$ | $\mathrm{a} 23=1.163935$ | $\mathrm{a} 32=0.787710$ | total $=3.141593$ |
| $\mathrm{ob} 3=1.688071$ | $\mathrm{~b} 31=1.040624$ | $\mathrm{~b} 13=0.412898$ | total $=3.141593$ |
| $\mathrm{ob} 2=0.264909$ | $\mathrm{~b} 12=0.782288$ | $\mathrm{~b} 21=2.094395$ | total $=3.141593$ |
| $\mathrm{tb} 3=1.964650$ | $\mathrm{~b} 23=0.408439$ | $\mathrm{~b} 32=0.768504$ | total $=3.141593$ |
| $\mathrm{o} 23=1.669384$ | $\mathrm{o} 32=0.264422$ | $\mathrm{t} 13=1.207787$ | total $=3.141593$ |


| Coordin | es of poin |  |  | x | $y$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 11 a 3 | \& 1A |  |  | 6.320654 | 49.598880 |
| 2 bal | - ba3 \& 1A |  |  | 6.191542 | 49.615170 |
| 3 ba3 | - ba2 + 1 a2 | 1A |  | 6.279009 | 49.604180 |
| 4 a31 | \& 1A |  |  | 6.320652 | 49.598880 |
| 5 ba2 | ba3\& 2A \& | $2+\mathrm{ba} 3-\mathrm{a} 2$ | \& 12 | 6.279009 | 49.604180 |
| 61 a 2 | 2a3 \& 1A |  |  | 6.126843 | 49.623200 |
| 7 2a3 | \& 2A \& 2a3 | $1 \& 12$ |  | 6.126842 | 49.623200 |
| 8 a32 | \& 23 \& ta3- | \& 12 |  | 5.822723 | 49.623200 |
| $9 \mathrm{~B}+$ | ba-ba3 \& |  |  | 6.539369 | 49.616680 |
| 10 B \& | ob3-ab3-b | \& 1B |  | 6.506444 | 49.319430 |
| 11 lb 2 | 1b3 \& 2B \& |  |  | 6.561309 | 49.250940 |
| 12 lb 2 | 1ba -ba3 \& | \& 12 |  | 6.650522 | 49.477430 |
| $13 \mathrm{2b3}$ | \& 2B \& 12 |  |  | 6.863803 | 50.009780 |
| 14 a 31 | o32\&23 \& | -a21 \& 12 |  | 5.726768 | 49.719150 |
| 15 23a | \& 23 \& 12 |  |  | 6.126844 | 49.623200 |
| 16231 | 13a \& 23 \& |  |  | 6.223382 | 49.719150 |
| Mean $x$ | 6.291607 | $\operatorname{std} \operatorname{dev} x$ | 0.291388 | 8 coef var $x$ | x 4.631372 |
| Mean $y$ | 49.601670 | std $\operatorname{dev} y$ | 0.166066 | 6 coef var $y$ | y 0.334798 |
| $y$ coord | nate is more | able |  |  |  |

## Applied Statistics for Civil and Environmental Engineers <br> Problem Solution Manual <br> by N.T. Kottegoda and R. Rosso <br> Chapter2- Basic Probability Concepts

2.1. Football stadium balcony. A civil engineer is asked to assess the reliability of a balcony overlooking a football stadium. The maximum number of people who can be accommodated in the balcony is 20 . The weight of an individual can be approximately $50 \mathrm{~kg}, 75 \mathrm{~kg}$, or 100 kg .
(a) Sketch the sample space.
(b) Show the following events involving numbers of people and their weights at any time:
$A \equiv\{$ there are more than 16 people in the balcony $\}$,
$B \equiv\{$ the total weight of people in the balcony is 1500 kg$\}$,
$C \equiv\{$ there are more than 15 people of the maximum weight $\}$.

## Solution.

Let $N_{50}, N_{75}$ and $N_{100}$ denote the numbers of people on the balcony weighing 50, 75 and 100 kg respectively, $N=N_{50}+N_{75}+N_{100}$ is constrained by $0 \leq N \leq 20$. These inequalities define a tetrahedron in the space $\left(N_{50}+N_{75}+N_{100}, 0\right)$. This represents the (DISCRETE/LATTICE) sample space $\Omega$. The events $A, B$ and $C$ are proper subsets of $\Omega$. The sample space is as follows.

$A=\{16<N \leq 20\}$ is represented by the space bounded by the two tetrahedrons $(0,20,20,20)$ and $(0,17,17,17)$; integer values only:

$B$ is represented by the intersection of the plane $\left\{50 \times N_{50}+75 \times N_{75}+100 \times N_{100}=1500\right\}$ with the tetrahedron ( $0,20,20,20$ ). This means the integer values that are common to both.

$C \equiv\left\{N_{100}>15\right\}$ represents integer values bounded by the horizontal plane passing through $\left\{N_{100}=16\right\}$ and the tetrahedron $(0,20,20,20)$ :


## additional sketch


2.2. Reservoir inflows. A reservoir impounds water from a stream $X$ and receives water $Y$ deviated via a tunnel from an adjoining catchment. The annual inflow from source $X$ can be approximated to 1 or 2 or 3 units of $10^{6} \mathrm{~m}^{3}$, and that from source $Y$ is 2 or 3 or 4 units of $10^{6} \mathrm{~m}^{3}$. On appropriate Venn diagrams show the following events.
(a) $A \equiv\{$ source $X$ is less than 3 units $\}$.
(b) $B \equiv$ \{source $Y$ is more than 2 units $\}$.
(c) $A+B$.
(d) $A B$.

## Solution.


2.3. Sequential construction. The sequence of construction of a structure involves two phases. Initially, the foundation is built, then work commences on the superstructure. The completion of the foundation can take 4 or 5 months, which are equally likely to be needed. The superstructure requires 5 , 6 or 7 months to be completed, with equal likelihood for each period. The time of completion of the superstructure is independent of that taken to complete the foundation. List the possible combinations of times for the completion of the project, and determine the associated probabilities.

## Solution.



| MONTHS FOMPLETION OF PROJECT | 9 | 10 | 11 | 12 |
| :--- | :--- | :--- | :--- | :--- |
| PROBABI.LITIES | $1 / 6$ | $1 / 3$ | $1 / 3$ | $1 / 6$ |

2.4. Dam spillway. An engineer is designing a spillway for a dam. The evaluation of maximum flow data is based on a short period of recordkeeping. The critical flow rates and their probabilities are estimated from $A$, discharge measurements, $B$, rainfall observations, and $C$, combination of flow discharge and rainfall data, as follows:

$$
\begin{array}{lrl}
\text { Event } A \text { from flow data: } & 8,000 \text { to } 12,000 \mathrm{~m}^{3} / \mathrm{s}, & \operatorname{Pr}[A]=.5 . \\
\text { Event } B \text { from rainfall data: } & 10,000 \text { to } 15,000 \mathrm{~m}^{3} / \mathrm{s}, & \operatorname{Pr}[B]=.6 . \\
\text { Event } C=A+B: & 8,000 \text { to } 15,000 \mathrm{~m}^{3} / \mathrm{s}, & \operatorname{Pr}[C]=.9 .
\end{array}
$$

(a) Sketch the foregoing events.
(b) Show on the sketch $A B, A C$, and $A^{\mathrm{c}_{+}} B^{\mathrm{C}}$.
(c) Determine the probabilities $\operatorname{Pr}[A B]$ and $\operatorname{Pr}\left[A^{\mathrm{c}}+B^{\mathrm{C}}\right]$.
(d) Determine the conditional probabilities $\operatorname{Pr}[A \mid B]$ and $\operatorname{Pr}[B \mid A]$.

## Solution.


$\operatorname{Pr}[A B]=\operatorname{Pr}[A]+\operatorname{Pr}[B]-\operatorname{Pr}[A+B]=0.5+0.6-0.9=0.2$
$\operatorname{Pr}\left[A^{\mathrm{c}}+B^{\mathrm{C}}\right]=\operatorname{Pr}\left[(A-B)^{\mathrm{c}}\right]=1-\operatorname{Pr}(A B)=1-0.2=0.8$
$\operatorname{Pr}[A / B]=\operatorname{Pr}(A-B) / \operatorname{Pr}(B)=0.2 / 0.6=1 / 3$
$\operatorname{Pr}[B / A]=\operatorname{Pr}[A B] / \operatorname{Pr}[A]=0.2 / 0.5=2 / 5$
additional sketch

2.5. Wind direction and intensity. Strong winds in a particular area come uniformly from any direction from north, $\theta=0^{\circ}$, to east, $\theta=90^{\circ}$. Wind speed $V$ is also variable, and it can exceed $50 \mathrm{~km} / \mathrm{h}$ with a probability of .04 , and $100 \mathrm{~km} / \mathrm{h}$ with a probability of .01 .
(a) Sketch the sample space for wind speed and direction.
(b) Sketch the following events: $A \equiv\{V>50 \mathrm{~km} / \mathrm{h}\}, B \equiv\{50<V<100 \mathrm{~km} / \mathrm{h}\}, A B, A+B, C \equiv\{30<\theta$ $\left.<60^{\circ}\right\}, A C$, and $B C$.
(c) Find $\operatorname{Pr}[B]$ and $\operatorname{Pr}[B C]$ assuming that wind speed and direction are stochastically independent.

## Solution.






$\operatorname{Pr}[A]=\operatorname{Pr}[V>50]=0.04$
$\operatorname{Pr}[D]=\operatorname{Pr}[V>100]=0.01$
$\operatorname{Pr}[B]=\operatorname{Pr}[50<V<100]=\operatorname{Pr}[A]-\operatorname{Pr}[D]=-0.4-0.1=.03$
$\operatorname{Pr}[C]=1 / 3$
$\operatorname{Pr}[B C]=\operatorname{Pr}[B] \times \operatorname{Pr}[C]=0.03 \times 1 / 3=0.01$
additional sketch

2.6. Irrigation water supply. A dam is designed to supply water to three separate irrigation schemes, $I_{1}, I_{2}$, and $I_{3}$. The demand for the first scheme $I_{1}$ is 0 or 1 or $2 \mathrm{~m}^{3} / \mathrm{s}$, whereas that for $I_{2}$ and $I_{3}$ is 0 or 2 or $4 \mathrm{~m}^{3} / \mathrm{s}$ in each case.
(a) Sketch the sample space for $I_{1}, I_{2}$, and $I_{3}$ separately, and for $I_{1}, I_{2}$, and $I_{3}$ jointly.
(b) Show the following events:

- $A \equiv\left\{I_{1}>1 \mathrm{~m}^{3} / \mathrm{s}\right\}$;
- $B \equiv\left\{I_{2} \geq 2 \mathrm{~m}^{3} / \mathrm{s}\right\}$;
- $C \equiv\left\{I_{3}<4 \mathrm{~m}^{3} / \mathrm{s}\right\}$;
- $A^{c} ; A B ; A+B ;(\mathrm{A}+\mathrm{B})^{c} ; A B^{c} ; A C ; A^{c} C ; B^{c} C ; B^{c} C^{c} ;$ (where feasible).
(c) Assuming that the demands from the three schemes are independent of each other, and that all possible demands are equally likely to occur, find the probability that the total water demand exceeds $5 \mathrm{~m}^{3} / \mathrm{s}$.


## Solution.

A simple sketch of the sample space for $I_{1}, I_{2}$, and $I_{3}$ jointly and the events $A \equiv\{$ $\left.I_{1}>1 \mathrm{~m}^{3} / \mathrm{s}\right\}$ and $A^{c}\left\{I_{1} \leq 1 \mathrm{~m}^{3} / \mathrm{s}\right\}$.
$4 \mathrm{~m}^{3} / \mathrm{s} ; 2 \mathrm{~m}^{3} / \mathrm{s} \uparrow \mathrm{I}_{1} ; 4 \mathrm{~m}^{3} / \mathrm{s}$



A simple sketch of the events $A B$ and $A B^{c}$ and some of the events $A^{c} C$
$4 \mathrm{~m}^{3} / \mathrm{s} ; 2 \mathrm{~m}^{3} / \mathrm{s} \uparrow \mathrm{I}_{1} ; 4 \mathrm{~m}^{3} / \mathrm{s}$


$$
\begin{aligned}
\operatorname{Pr}\left[I_{1}+I_{2}\right. & \left.+I_{3}>5\right]= \\
& =\operatorname{Pr}\left[I_{1}=0, I_{2}=2, I_{3}=4\right]+ \\
& +\operatorname{Pr}\left[I_{1}=0, I_{2},=4, I_{3}=2\right]+ \\
& +\operatorname{Pr}\left[I_{1}=0, I_{2}=4, I_{3}=4\right]+ \\
& +\operatorname{Pr}\left[I_{1}=1, I_{2},=2, I_{3}=4\right]+ \\
& +\operatorname{Pr}\left[I_{1}=1, I_{2}=4, I_{3}=2\right]+ \\
& +\operatorname{Pr}\left[I_{1}=1, I_{2},=4, I_{3}=4\right]+ \\
& +\operatorname{Pr}\left[I_{1}=2, I_{2}=0, I_{3}=4\right]+ \\
& +\operatorname{Pr}\left[I_{1}=2, I_{2},=2, I_{3}=2\right]+ \\
& +\operatorname{Pr}\left[I_{1}=2, I_{2}=2, I_{3}=4\right]+ \\
& +\operatorname{Pr}\left[I_{1}=2, I_{2},=4, I_{3}=0\right]+ \\
& +\operatorname{Pr}\left[I_{1}=2, I_{2}=4, I_{3}=2\right]+ \\
& +\operatorname{Pr}\left[I_{1}=2, I_{2},=4, I_{3}=4\right]= \\
& =12 \times 1 / 3 \times 1 / 3 \times 1 / 3=4 / 9
\end{aligned}
$$

additional sketch

2.7. Port occupancy. An experiment consists of counting the number of ships in a small harbor on a particular day and estimating the total tonnage. The maximum number of vessels permitted at a given time in the port is six, while each vessel can have a tonnage from 5,000 to 25,000 . Only the total number of ships and the total tonnage is recorded.
(a) Sketch a sample space for this experiment.
(b) Indicate on the diagram the regions corresponding to the following events:

- $A \equiv\{$ the number of ships is less than 5$\} ;$
- $B \equiv\{$ the total tonnage is less less than 35,000$\}$;
- $C \equiv\{$ three ships each of maximum tonnage are present $\} ;$
- $A+B ; A B ; A^{c}+B^{c} ; A^{c} B^{c} ; A C ; A^{c} C$ (where feasible).


## Solution.

Sketches of the sample space and events $A, B, C, A C, A^{C}$ and $A^{C}+B^{C}$


Vertical scale is uneven

2.8. Simply supported beam. A load of 200 kg is placed on a simply supported beam of length 6 m . If $R_{1}$, and $R_{2}$ denote the reactions at the left and right supports, respectively, $R_{1}+R_{2}=200 \mathrm{~kg}$ for any location of the load.
(a) Define, and sketch the sample space for this experiment.
(b) Sketch on the diagram the following events.

- $A \equiv\{$ the load is located at 1 m from support 1$\}$;
- $B \equiv\{$ the load is located between 2 and 4 m from support 1$\}$;
- $C \equiv\{$ the load is located between 3 and 5 m from upport 1$\}$;
- $A+B ; A B ; B+C ; B C ; A^{C}+B C ;$ and $A^{c} B^{C} C^{C}$ (where feasible).
(c) If the load can vary from 100 to 400 kg , define and sketch the new sample space. Sketch on this diagram the following events.
- $D \equiv\{$ a load heavier than 100 kg is located at 2 to 4 m from support 1$\}$;
- $E \equiv\{$ a load heavier than 200 kg is located at 3 to 5 m from support 1$\}$;
- $D+E ; D E$; and $D E^{c}$ (where feasible).


## Solution.

$B \cdot \quad ■ C$
Sketch $1(a, b)$ shows sample space of the experiment, the event $\bullet \mathbf{A}$ and the events $\backslash$ and $\backslash$ $B \cdot \quad ■ C$
with distances from support 1 at point $X$, where $X Y=6$ meters, and support 2 is at point $Y$. The reactions $R_{1}$ and $R_{2}$ are given on the two axes.
Sketch $2(a, b)$ shows the events $B+C$, and $B C$ but $A B$ is a null event. The event $A^{C} \equiv A^{C}+B C$ covers the full distance $X Y=6$ meters excluding the event $\bullet \mathbf{A}$. The event
$A^{C} B^{C} C^{C}$ covers a distance of 2 meters from $X($ excluding the event $\bullet \mathbf{A})$ and 1 meter from $Y$.
Sketch 1(c) gives the sample space if the load can vary from 100 to 400 kg .

| $R_{1}$ | $R_{1}$ | $R_{1}$ |
| :---: | :---: | :---: |
| $\uparrow$ | $\uparrow$ | $\uparrow$ |
| $\mathbf{2 0 0 \\|} \backslash X^{\prime}$ | 200 | 400 |
| \| | \| | \|/\} |
| - $A 1$ | $\bullet$ A | \|/// |
| 150\| \} | 150\| | 300\|//// |
| $\boldsymbol{\\|} \cdot \backslash \quad 2$ | $\boldsymbol{B} \cdot \backslash$ | \|///// \} |
| $\backslash \backslash$ | \| 11 | \|/I/II/^ |
| 100\| \ \■C3 | 100 \ \ | 200\|//////// |
| \ \ \ | 111 | \|/IIIIII/込 |
| $B \cdot \backslash \backslash 4$ | $+\backslash \cdot B$ | \|////////// \} |
| 50\| |  |  |
|  | 50\| \ \ \ BC | 100 $\backslash / / / / / / / / / / \wedge$ |
| $\backslash \mathrm{C} 5$ | \| $C$ ■ \■C | \| \////////// |
| 1 | 1 | \| VIIIIIII分 |
| $0\|\quad\| Y 6$ | $0 \mid$ । | 0\| V/IIIIII/ |
|  |  | $L_{\sim} \quad L_{\sim} \rightarrow R_{2}$ |
| 0100000 | $\overline{0 \quad 100 \quad 200}$ | $\begin{array}{llll}0 & 100 & 400\end{array}$ |
| $1(a, b)$ | 2(a,b) | 1(c) |

Events $D$ and $E$ are shaded in the following sketches $2(c)$ and $2(d)$, respectively. Distances from support 1 at $X$ are shown as before. Events $D+E, D+E$ and $D E^{C}$ follow easily.

additional sketch




2.9. Storm rainfall. Analysis of the data of Problem 1.20 indicates that the estimated probability of a storm resulting in more than 40 mm of rainfall in one hour is about 0.5 . Using relative frequencies, compute the probability that in any year the same rainfall intensity is exceeded over a duration of (a) 20 minutes, and of $(b) 3$ hours.

If the annual 30 -minutes and 1 -hour rainfalls refer to the same storm events, what is the conditional probability that the intensity does not decrease from $60 \mathrm{~mm} /$ hour or more during the first 30 minutes by more than 25 percent during a one hour period?

## Solution.

Storm rainfall. 60 minute rainfall: The durations of 7 out of 14 annual maximum rainfall events exceed 40 min . Hence probability of exceedence is 0.5 . Intensity $=40$ $\mathrm{min} /$ hour. With same intensity in 20 minutes rainfall should exceed 13.3 min . Probability of exceedence is $13 / 14$. In 3 hours with the same intensity of 40 $\mathrm{min} /$ hours the rainfall $=120 \mathrm{~min}$. Probability of exceedence $=3 / 14$.

|  | Year | 1974 | 1975 | 1976 | 1977 | 1979 | 1981 | 1984 | 1987 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 30 min rainfall exceeding 30 min | (1) | 30.5 | 31.2 | 37.2 | 72.4 | 46.1 | 56.0 | 65.7 | 73.9 |
| 60 hours rainfall | (2) | 45.75 | 46.8 | 55.8 | 108.6 | 69.15 | 84.0 | 98.55 | 111 |
| $\quad 30$ min x1.50 |  |  |  |  |  |  |  |  |  |
| 60 min recorded rainfall | (3) | 38.7 | 40.2 | 55.7 | 118.9 | 55.6 | 89.4 | 105.7 | 128.5 |
| It is seen that $(3)>(2)$ |  |  |  |  |  |  |  |  |  |

Probability that the intensity does not decrease from $60 \mathrm{~min} /$ hour or more during the first 30 minutes by more than 25 percent during a one hour period $=0.5$
2.10. Hydropower. Run-of-river hydroelectrical plants convert the natural potential energy of surface water in a stream into electrical energy. The plant capacities depend on natural river flow, which generally varies during the year according to season and precipitation regime. Assume that the design flow of a given power station, say, $Q_{D}$, is the natural flow, which is exceeded during 274 days in a year on average. At other times, when the river flow is lower than the design flow, the plant is nevertheless capable of producing some power if the flow is not lower than $Q_{0}$. Moreover, during floods it is not possible to convey water to the plant due to sedimentation, which occurs when the natural river flow $Q$ exceeds $Q_{1}$. (a) If $\operatorname{Pr}\left[Q<Q_{0}\right]=.1$ and $\operatorname{Pr}\left[Q>Q_{1}\right]=.05$, for how many days in a year will the plant be incapable of supplying electric energy? $(b)$ What is the probability that the plant works at full capacity? (c) What is the probability that the plant fulfills its minimum target? Note that $Q_{0}<Q_{D}<Q_{1}$.

## Solution.

Sketch of $Q$ vs. days per year $Q$ is exceeded

$\begin{array}{lll}\mathrm{P}\left(Q<Q_{0}\right)=0.1 & 365 \times 0.1 \approx 37 & \text { insufficient } \mathrm{Q} \\ \mathrm{P}\left(Q>Q_{1}\right)=0.5 & 365 \times 0.05 \approx & \text { sedimentation }\end{array}$

$$
N=0.15 \times 365 \approx 55 \text { days }
$$

Probability (full capacity) $=(274-18) / 365 \approx \underline{0.70}$
Probability $($ minimum target $)=(328-18) / 365 \approx \underline{0.85}$

2.11. Reservoir operational policy. Consider the water storage $S$ in a reservoir as described in Example 2.1 and Fig. 2.1.1. The manager must release in a year an amount of water $R$ that depends on the amount of the annual inflow $I$, the storage $S$ at the beginning of that year, and the demand $d$ in that year. The manager follows the following "normal operational policy" for water releases:

$$
\begin{array}{lr}
R=d, & \text { if } d \leq I+S<d+c, \\
R=I+S, & \text { if } I+S<d, \\
R=I+S-c &
\end{array}
$$

$$
\text { if } I+S \geq d+c
$$

with $c$ denoting the effective storage capacity of the reservoir. If $\operatorname{Pr}[d \leq I+S \leq d+c]=.6, \operatorname{Pr}[I+S<$ $d]=.1$, and $\operatorname{Pr}[I+S>d+c]=.3$, find the probability that the demand is satisfied.

## Solution.

Sketch of reservoir


Sketch of operational policy


Probability that the demand is satisfied $=0.3+0.6=0.9$

## additional sketch


2.12. Industrial park utilities. Consider the design requirements of water supply and wastewater removal systems in a new industrial park, which consists of five independent buildings. Assume that the water demand $S$ of each of the five industrial buldings can be 10 or 15 units, whereas the required wastewater removal capacity $R$ can be 8,10 or 15 units. After some interviews with potential clients, the designer has estimated that the combined requirements of the two systems are likely to occur with the following probabilities at the $i$ th site:

|  | $R=\mathbf{1 5}$ | $R=\mathbf{1 0}$ | $R=\mathbf{8}$ |
| :---: | :---: | :---: | :---: |
| $S=10$ | .00 | .25 | .15 |
| $S=15$ | .20 | .35 | .05 |

Stochastic independence can also be assumed among the requirements of different buildings.
(a) What is the probability that the total water demand exceeds 60 units?
(b) What is the probability that the total wastewater removal capacity exceeds 50 units?

## Solution.

Summary of probabilities

| $S / R$ | $\mathbf{1 5}$ | $\mathbf{1 0}$ | $\mathbf{8}$ | Sum of Probs. |
| :---: | :---: | :---: | :---: | :---: |
| $\mathbf{1 0}$ | 0 | 0.25 | 0.15 | 0.40 |
| $\mathbf{1 5}$ | 0.20 | 0.35 | 0.05 | 0.60 |
| Sum of Probs. | 0.20 | 0.60 | 0.20 | 1.00 |

$\operatorname{Pr}[S>60]=\operatorname{Pr}[5 \times 15]+\operatorname{Pr}[4 \times 15,1 \times 10]+\operatorname{Pr}[3+15,2 \times 10]$
$\left.=\binom{5}{5} 0.6^{5}+\binom{5}{4} 0.6^{4} \times 0.4+\binom{5}{3}\right) 0.6^{3} \times 0.4^{2}$
$=0.07776+0.2592+0.3456=0 . \underline{68256}$

$$
\begin{aligned}
& \leftarrow 40 \rightarrow 42 \quad \rightarrow \quad \leftarrow 47 \rightarrow \\
& \operatorname{Pr}[R \leq 50]=\operatorname{Pr}[0 \times 15,0 \times 10,5 \times 8]+\operatorname{Pr}[0 \times 15,1 \times 10,4 \times 8]+\operatorname{Pr}[1 \times 15,1 \times 10,3 \times 8] \\
& +\operatorname{Pr}[1 \times 15,1 \times 10,3 \times 8]+\underset{\operatorname{Pr}[0 \times 15,2 \times 10,3 \times 8]}{\leftarrow} \stackrel{44}{\leftarrow}+\underset{\operatorname{Pr}[0 \times 15,3 \times 10,}{\leftarrow} \stackrel{46}{\leftarrow} \overrightarrow{8}] \\
& +\stackrel{48}{\stackrel{4}{*}} \stackrel{\stackrel{4}{0 \times 15}, 4 \times 10,1 \times 8]}{\overrightarrow{4}} \stackrel{\operatorname{Pr}[0 \times 15,}{\leftarrow} \stackrel{50}{5 \times 10,0 \times 8]} \overrightarrow{0} \\
& =0.2^{5}+5 \times 0.6 \times 0.2^{4}+5 \times 0.2 \times 0.2^{4}+20 \times 0.2 \times 0.6 \times 0.2^{3}+10 \times 0.6^{2} \times 0.2^{3}+10 \times 0.6^{3} \times 0.2^{2} \\
& +5 \times 0.6^{4} \times 0.2+0.6^{5} \\
& =0.00032+0.00480+0.00160+0.01920+0.0 .0288+0.0864+0.1296+0.07776=0.34848 \\
& \text { Therefore, } \operatorname{Pr}[R>50]=1-0.34848=0.65152
\end{aligned}
$$

2.13. Construction scheduling. Consider the sequential construction scheme of Problem 2.3 , and assume that both the foundation and the superstructure can be completed at three different rates, say, $a$, $b$, or $c$. These rates modify the probability of completion of each phase of construction as shown in the table given here. Also, monthly costs vary for the different rates.

| Phase | Rate | Cost per month at rate, \$ | Probability of time of completion |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 4 months | 5 months | 6 months | 7 months |
| Foundation | a | 30,000 | . 3 | . 7 | 0 | 0 |
| Foundation | $b$ | 36,000 | . 5 | . 5 | 0 | 0 |
| Foundation | c | 42,000 | . 3 | . 7 | 0 | 0 |
| Superstructure | a | 25,000 | 0 | . 1 | . 4 | . 5 |
| Superstructure | $b$ | 40,000 | 0 | . 3 | . 3 | . 3 |
| Superstructure | c | 50,000 | 0 | . 5 | . 3 | . 2 |

In addition, if the construction is not completed in 11 months, the contractor must pay a penalty of $\$ 300,000$ per month.
(a) Compute the expected cost of foundation performed at rate $a$ as the summation for all times of completion of the product between the total cost (the product of the number of required months and the cost per month) and probability.
(b) Compute all expected costs.
(c) Compute the total expected penalty for each possible strategy of completion of the whole structure.
(d) Determine the best strategy by minimizing the sum of total expected cost and penalty.

## Solution.

(a) Expected cost is $30,000 \times 4 \times 0.3+30,000 \times 5 \times 0.7=36,000+105,000=$ 141,000
(b) $E\left[F_{a}\right]=30,000 \times(4 \times 0.3+5 \times 0.7)=\$ 141,000$
$E\left[F_{b}\right]=36,000 \times(4 \times 0.5+5 \times 0.5)=\$ 162,000$
$E\left[F_{c}\right]=42,000 \times(4 \times 0.3+5 \times 0.7)=\$ 197,000$
$E\left[S_{a}\right]=25,000 \times(5 \times 0.1+6 \times 0.4+7 \times 0.5)=\$ 160,000$
$E\left[S_{b}\right]=40,000 \times(5 \times 0.3+6 \times 0.3+7 \times 0.4)=\$ 244,000$
$E\left[S_{c}\right]=50,000 \times(5 \times 0.5+6 \times 0.3+7 \times 0.2)=\$ 285,000$
(c) $E\left[t F_{a}\right]=4 \times 0.3+5 \times 0.7=4.7$ months;
$E\left[t F_{b}\right]=4 \times 0.5+5 \times 0.5=4.5$ months;
$E\left[t F_{c}\right]=4 \times 0.3+5 \times 0.7=4.7$ months;
$E\left[t S_{a}\right]=5 \times 0.1+6 \times 0.4+7 \times 0.5=6.4$ months
$E\left[t S_{b}\right]=5 \times 0.3+6 \times 0.3+7 \times 0.4=6.1$ months
$E\left[t S_{c}\right]=5 \times 0.5+6 \times 0.3+7 \times 0.2=5.7$ months
$E\left[t_{a}\right]=4.7+6.4=11.1$ months; penalty $\$ 300,000 \times 0.1=\$ 30,000$
$E\left[t_{b}\right]=4.5+6.1=10.6$ months; penalty $=0$
$E\left[t_{c}\right]=4.7+5.7=10.4$ months; penalty $=0$.
(d) Total expected cost with rate $a$, say, $T E C_{a}=141,000+160,000+30,000$ $=\$ 331,000$

Also $T E C_{b}=162,000+244,000=\$ 406,000$ and
$T E C_{c}=197,000+285,000=\$ 482,000$.
Hence best strategy is rate $a$.
2.14. Research project ranking. A committee consisting of three independent referees ( $R_{1}, R_{2}$, and $\left.R_{3}\right)$ is to rank four different research project applications $(A, B, C$ and $D)$. Each referee ranks the four projects as 3 (for the best), 2, 1, and 0 , and then the assigned ranks for each project are summed. Assume that the referees are unable to discriminate between projects so that the rankings are randomly assigned. What is the probability that project $A$ will receive a total score of 4 ?

## Solution.

$R_{1} R_{2} R_{3}$
(1) Total score for project $A=0$ :

Ranks for $A \quad R_{1} 0$

$$
R_{2} 0
$$

$$
R_{3} 0 \quad \text { number of ways } \quad \rightarrow 1
$$

(2) Total score for project $A=1$ :

Ranks for $A \quad R_{1} 100$

$$
R_{2} 010
$$

$$
R_{3} 001 \quad \text { number of ways } \quad \rightarrow 3
$$

(3) Total score for project $A=2$ :

Ranks for $A \quad R_{1} 20211$
0

$$
\begin{array}{lllllll}
R_{2} & 0 & 2 & 1 & 1 & 1 \\
R_{3} & 0 & 0 & 20 & 1 & \\
\end{array} \quad \quad \text { number of ways } \quad \rightarrow 6
$$

(4) Total score for project $A=3$ :

Ranks for $A \quad R_{1} 1212100300$
$R_{2} 1120021030$
$R_{3} 1001212303 \quad$ number of ways $\quad \rightarrow 10$
(5) Total score for project $A=4$ :

Ranks for $A$

$$
\begin{aligned}
& R_{1} \\
& R_{2} 31010202211 \\
& R_{2} \\
& R_{3}
\end{aligned} 0101301220121 \quad l+\quad \rightarrow 12
$$

6) Total score for project $A=5$ :

Ranks for $A \quad R_{1} 323200221113$
$R_{2} 230032212131$
$R_{3} 002323122311$ number of ways $\quad \rightarrow 12$
For $A \geq 5$, by symmetry:- for $A=6$, number of ways $=10$; for $A=7$, number of ways $=6$; for $A=8$, number of ways $=3$; for $A=9$, number of ways $=1$. The total number of ways $=64$. Hence, the probability that project $A$ will receive a total score of $4=12 / 64=$ 0.1875 .
2.15. Probabilities of reservoir storage. Consider the water storage $S$ in a reservoir described by a sequence of four states $\omega_{1}, \omega_{2}, \omega_{3}, \omega_{4}$, where each state describes water volumes ranging from 0 to $c / 4$, from $c / 4$ to $2 c / 4$, and so on (see Example 2.10 and Fig. 2.1.2). The reservoir manager is interested in the simple events given by $A_{i, k} \equiv\{(i-1) c / 4 \leq S<i c / 4\}$ for $i=1,2,3,4$ and annual time periods $k=$ $1,2,3, \ldots$

The manager has estimated the following conditional probabilities: $\operatorname{Pr}\left[A_{j, k+1} \mid A_{i, k}\right]=1 / 2$ for $j$ $=i$, and $\operatorname{Pr}\left[A_{j, k+1} \mid A_{i, k}\right]=1 / 6$ for $j \neq i$. What is the transition probability matrix $p_{i j}$ from the $i$ th to the $j$ th state after one step?

What is the probability that state 1 occurs in the third operational period, given that the reservoir was in state 4 in the first period?

## Solution.

$$
\text { Transition probability matrix } Q=\left[p_{i j}\right]=\begin{array}{llll} 
& \begin{array}{l}
i=1 \\
j=1 \\
j=1 \\
j=2
\end{array} & i=2 & i=3
\end{array} \quad i=4
$$

$$
\operatorname{Pr}[j, 1]=\left[\begin{array}{l}
0 \\
0 \\
0 \\
1
\end{array}\right], \text { for } j=1,4 . \operatorname{Pr}[j, 2]=Q \times \operatorname{Pr}[j, 1]=\left[\begin{array}{l}
1 / 6 \\
1 / 6 \\
1 / 6 \\
1 / 2
\end{array}\right] \text {, for } j=1,4
$$

$$
\operatorname{Pr}[j, 3]=Q \times \operatorname{Pr}[j, 2]=\left[\begin{array}{c}
2 / 9 \\
x \\
y \\
z
\end{array}\right], \text { for } j=1,4
$$

[The top element in the column is the sum $(1 / 12+1 / 36+1 / 36+1 / 12)$; the other elements, $x, y$ and $z$ are not of interest here.]
Hence answer is $2 / 9$.
2.16. Pumping station. Two pumps operate in parallel to provide water supply of a village located in a recreational area. Water demand is subject to considerable weekly and seasonal fluctuations. Each unit has a capacity so that it can supply the demand 80 percent of the time in case the other unit fails. The probability of failure of each unit is 10 percent, whereas the probability that both units fail is 3 percent. What is the probability that the village demand will be satisfied?

## Solution.

$\operatorname{Pr}\left[F_{1}\right]=0.1-0.03=0.07$
$\operatorname{Pr}\left[F_{2}\right]=0.1-0.03=0.07$
$\operatorname{Pr}\left[F_{12}\right]=0.03$
where $F_{i} \equiv$ failure of only pump $i$
and $\quad F_{i j} \equiv$ both pumps $i$ and $j$ fail.
The 2 pumps are the independent because
$\mathrm{P}\left(F_{12}\right)=0.03 \neq \mathrm{P}\left(F_{1}\right) \times \mathrm{P}\left(F_{2}\right)$.
$\operatorname{Pr}[$ Village unsatisfied $]=\operatorname{Pr}[U]=$
$\operatorname{Pr}\left[U \mid F_{12}\right] \operatorname{Pr}\left[F_{12}\right]+\operatorname{Pr}\left[U \mid F_{1}\right] \operatorname{Pr}\left[F_{1}\right]+\operatorname{Pr}\left[U \mid F_{2}\right] \operatorname{Pr}\left[F_{2}\right]$
$=1 \times \operatorname{Pr}\left[F_{12}\right]+0.2 \times \operatorname{Pr}\left[F_{1}\right]+0.2 \times \operatorname{Pr}\left[F_{2}\right]$
$=0.03+0.2 \times 0.07+0.2 \times 0.07=0.058$
$\operatorname{Pr}[$ Village satisfied $]=\operatorname{Pr}[S]=1-0.058=\underline{0.942}$

## additional sketch


2.17. Analysis of reservoir lifetime. A reservoir is designed for an area with high erosional rates. The engineer is interested in determining the lifetime of the reservoir, which can come to an end either because the impounding dam can be destroyed by a flood exceeding the spillway capacity or because excessive sedimentation results in a severe loss in reservoir capacity. It is necessary to determine the probability that the structure will come to an end of its useful life in each of the years after construction. One can assume a constant probability $q$ that in any year a flow exceeding the spillway capacity can occur, and an exponentially increasing probability $p_{i}$ that reservoir sedimentation can occur in the $i$ th year after construction, given that no significant sedimentation has occurred prior to the $i$ th year, that is, $p_{i}=1-\exp (-\beta i)$, with $\beta>0$.

Denote by $A_{n}$ the event associated with a destructive flood occurring in the $n$th year after construction, and by $B_{n}$ that associated with excessive sedimentation.
(a) What is the probability that the system will survive for $n$ years, that is

$$
\operatorname{Pr}\left[\left(A_{1}^{c} B_{1}^{c}\right)\left(A_{2}^{c} B_{2}^{c}\right) \ldots\left(A_{n}^{c} B_{n}^{c}\right)\right] ?
$$

(b) What is the probability that the system will come to an end in the $n$th year, where $S_{n}$ denotes survival up to the $n$th year,

$$
\operatorname{Pr}\left[\left(A_{n}+B_{n}\right) \mid S_{n-1}\right] \operatorname{Pr}\left[S_{n-1}\right] ?
$$

(c) Compute the foregoing probabilities for $q=.01, \beta=.002$, and $n=25$.

## Solution.

Analysis of reservoir lifetime: $A_{n}$ destructive flood in year $n, B_{n}$ same with respect to sediment
(a) $\operatorname{Pr}\left[\left(A^{\mathrm{c}}{ }_{1} B^{\mathrm{c}}{ }_{1}\right)\left(A^{\mathrm{c}}{ }_{2} B^{\mathrm{c}}{ }_{2}\right) . .\left(A^{\mathrm{c}}{ }_{n} B^{\mathrm{c}}{ }_{n}\right)\right]=(1-q)^{n} \prod_{i=1}^{n} e^{-\beta i}=(1-q)^{n} e^{-\beta \frac{n+1}{2} n}$
$=0.95^{25} \mathrm{e}^{-0.002 \times 13 \times 25}=0 . \underline{4061}$
[for $q=0.01, \beta=.0 .02, n=25]$
(b) $\operatorname{Pr}\left[\left(A_{n}+B_{n}\right) \mid S_{n-1}\right] \operatorname{Pr}\left[S_{n-1}\right]$, where $S$ denotes survival up to $n$th year, $=\left[1-e^{-\beta n}+q e^{-\beta n}\right]\left[(1-q)^{n-1} e^{-\beta n(n-1) / 2}\right]$
$\left\{\right.$ This is obtained from $\operatorname{Pr}\left[A_{n}+B_{n}\right]=\operatorname{Pr}\left[A_{n}\right]+\operatorname{Pr}\left[B_{n}\right]-\operatorname{Pr}\left[A_{n} B_{n}\right]$
$\left.=q+\left(1-e^{-\beta n}\right)-q\left(1-e^{-\beta n}\right)\right\}$
For $q=0.1, \beta=.002$ and $n=25$,
Probability $=\left[1-0.99 e^{-0.05}\right]\left[0.99^{24} \times \mathrm{e}^{-.002 \times 25 \times 12}\right]$
$=. \underline{02513}$
2.18. Highway system. To reach Grenoble, France, from Turin, Italy, one can follow either of two routes. The first directly connects Turin and Grenoble, whereas the second passes through Chambery, France. During extreme weather conditions in winter, travel between Turin and Grenoble is not always possible because some parts of the highway may not be open to traffic. Denote with A, B, and C the events that highways from Turin to Grenoble, Turin to Chambery and Chambery to Grenoble are open, respectively. In anticipation of driving from Turin to Grenoble, a traveler listens to the next day's weather forecast. If snow is forecast for the next day over the southern Alps, one can assume (on the basis of past records) that $\operatorname{Pr}[\mathrm{A}]=.6, \operatorname{Pr}[\mathrm{~B}]=.7, \operatorname{Pr}[\mathrm{C}]=.4, \operatorname{Pr}[\mathrm{C} \mid \mathrm{B}]=.5$, and $\operatorname{Pr}[\mathrm{A} \mid \mathrm{BC}]=.4$.
(a) What is the probability that the traveler will be able to reach Grenoble from Turin?
(b) What is the probability that the traveler will be able to drive from Turin to Grenoble by way of Chambery?
(c) Which route should be taken in order to maximize his chance to reach Grenoble?

## Solution.

$A \operatorname{Turin}(T)-\operatorname{Grenoble}(G), \operatorname{Pr}[A]=0.6$;
$B$ Turin - Chambery $(C), \operatorname{Pr}[B]=0.7$;.
$C$ Chambery - Grenoble, $\operatorname{Pr}[C]=0.4$;
$\operatorname{Pr}(C \mid B)=0.5 ; \operatorname{Pr}(A \mid B C)=0.7$
[1] $\operatorname{Pr}(T G)=\mathrm{P}[A]+\operatorname{Pr}[B C]-\operatorname{Pr}[A B C]=\operatorname{Pr}[A]+\operatorname{Pr}[C \mid B] \operatorname{Pr}[\mathrm{B}]-\operatorname{Pr}[A \mid B C] \operatorname{Pr}[B C]$ $=\operatorname{Pr}[A]+\operatorname{Pr}[C \mid B] \operatorname{Pr}[B]-\operatorname{Pr}[A \mid B C) \operatorname{Pr}[C \mid B] \operatorname{Pr}[B]=0.6+0.5 \times 0.7-0.4 \times 0.5 \times 0.7=$.
$\underline{81}$
[2] $\operatorname{Pr}(T G /$ Chambery $)=\operatorname{Pr}[B]+\operatorname{Pr}[C]-\operatorname{Pr}[B C]=\operatorname{Pr}[B]+\operatorname{Pr}[C]-\operatorname{Pr}[C \mid B] \operatorname{Pr}[B]$

$$
=0.7+0.4-0.5 \times 0.7=0.75
$$

[3] $\operatorname{Pr}[T G \mid$ Directly $]=\mathrm{P}[A]=0.6$. Therefore, take route [2] via Chambery
2.19. Wastewater treatment. The wastewater from an industrial plant requires treatment before disposal in the sea. This process consists of three sequential stages. For simplicity, define these stages as primary, secondary, and tertiary treatments, respectively. The result for each stage can be rated as unsatisfactory, incomplete, and satisfactory. Denote with $A_{k}$ the event that the $k$ th stage of the treatment process is unsatisfactory, with $B_{k}$ the event that it is incomplete, and with $C_{k}$ the event that it is satisfactory. The associated probabilities are given in the following table.

|  | $\operatorname{Pr}\left[\boldsymbol{A}_{\boldsymbol{k}}\right]$ | $\operatorname{Pr}\left[\boldsymbol{B}_{\boldsymbol{k}}\right]$ | $\operatorname{Pr}\left[\boldsymbol{C}_{\boldsymbol{k}}\right]$ |
| :---: | :---: | :---: | :---: |
| $k=1$ | .1 | .3 | .6 |
| $k=2$ | .2 | .3 | .5 |
| $k=3$ | .1 | .5 | .4 |

Further, assume that the three stages of the process are stochastically independent. If the satisfactory overall treatment requires that none of the three stages is unsatisfactory and at least two of these stages are satisfactory, what is the probability of this event?

## Solution.

The 5 combinations satisfying these conditions and the associated probabilities:
$\operatorname{Pr}\left[B_{1} C_{2} C_{3}\right]=0.3 \times 0.5 \times 0.4=0.060$
$\operatorname{Pr}\left[C_{1} B_{2} C_{3}\right]=0.6 \times 0.3 \times 0.4=0.072$
$\operatorname{Pr}\left[C_{1} C_{2} B_{3}\right]=0.6 \times 0.5 \times 0.5=0.150$
$\operatorname{Pr}\left[C_{1} C_{2} C_{3}\right]=0.6 \times 0.5 \times 0.4=0.120$
Sum of probabilities $\quad=\underline{0.402}$
2.20. Earthquake occurrence and intensity. Because of the uncertainties associated with the occurrence and intensity of earthquakes, one must consider earthquakes occurring in a given location as random phenomena. MCS intensity is a measure based on earthquake impact on the landscape, buildings, and population. In Problem 1.22 records of earthquake intensity in terms of MCS index are given for a period of about 1000 years in Rome, Italy. They are ranked from 2 to 7 for increasing intensities. In 10 centuries 329 earthquakes were reported in the study area, and in only two centuries there were no occurrences. Calculate a frequency-based estimate of the probability that at least one earthquake is likely to occur in a century. What is the probability that a recorded earthquake is of intensity 7 ?

## Solution.

(1) The frequency-base estimate of the probability that at least one earthquake is likely to occur in a century. Data from 10 centuries are given. Only in 2 centuries earthquakes not observed The probability of such an event $=\underline{0.8}$
(2) The probability that a recorded earthquake is of intensity $7=2 / 329$.
2.21. Air pollution control. The air pollution in Milan, Italy, is mainly caused by industrial, automobile, and heating emissions. A newly elected local government wishes to control these three sources of pollution within a period of four years. The chances of successfully controlling these sources are 80 percent, 70 percent and 50 percent, respectively. The government assumes that if only one of these three sources is successfully controlled, the probability of bringing air pollution below the acceptable level would be 50 percent only, but this probability increases to 80 percent if two of them are successfully controlled. The government also assumes stochastic independence among controlling industrial, heating, and automobile exhausts. What is the probability that two of the sources of air pollution will be successfully controlled in Milan during the four-year period?

## Solution.

Let $I, A$ and $H$ represent the successful control of pollution caused by industrial, automobile, and heating emissions, respectively.
The probability of controlling all 3 sources of pollution $\operatorname{Pr}[I A H]=\operatorname{Pr}[I] \times \operatorname{Pr}[A] \times$ $\operatorname{Pr}[H]=0.8 \times 0.7 \times 0.5=0.28$
There are 3 combinations of controlling only two sources of pollution successfully:
$\operatorname{Pr}\left[I A H^{c}\right]=0.8 \times 0.7 \times(1-0.5)=0.28$
$\operatorname{Pr}\left[I A^{c} H\right]=0.8 \times(1-0.7) \times 0.5=0.12$
$\operatorname{Pr}\left[I^{c} A H\right]=(1-0.8) \times 0.7 \times 0.5=0.07$
Sum of probabilities $\quad=0.47$
Therefore the probability of successfully controlling air pollution with respect to at least 2 of the sources $=0.28+0.47=\underline{0.75}$.
2.22. Imperfect concrete testing. An existing reinforced concrete building must be tested for possible obsolescence. Based on professional judgement, the engineer classifies concrete quality as either 35 to $39.9,40$ to $44.9,45$ to 49.9 , or 50 to $60 \mathrm{~N} / \mathrm{mm}^{2}$ based on a 28 -day test of compressive strength of concrete cubes. The relative likelihoods assigned to these four states are $.2, .3, .4$, and .1 , respectively. Concrete cores are to be cut and tested to help ascertain the true state, although the engineer knows that results from test cores are not conclusive. Therefore, conditional probabilities are estimated to account for the uncertainties involved in examining the cores. These probabilities describe the likelihood that the value of core strength indicated predicts a given unknown state. For example, if the true state is 35 to $39.9 \mathrm{~N} / \mathrm{mm}^{2}$, there is a 70 percent chance that the tested core strength also lies between 35 and $39.9 \mathrm{~N} / \mathrm{mm}^{2}$, but there is a 20 percent chance that it will lie between 40 and $44.9 \mathrm{~N} / \mathrm{mm}^{2}$, and a 10 percent chance that it lies in the range 45 to $49.9 \mathrm{~N} / \mathrm{mm}^{2}$. The conditional probabilities are tabulated next.

|  | State |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | $x_{1}$ | $x_{2}$ | $x_{3}$ | $x_{4}$ |  |
| $35-39.9 \mathrm{~N} / \mathrm{mm}^{2}$ | $40-44.9 \mathrm{~N} / \mathrm{mm}^{2}$ | $45-49.9 \mathrm{~N} / \mathrm{mm}^{2}$ | $50-60 \mathrm{~N} / \mathrm{mm}^{2}$ |  |  |


| Core Strength |  |  | 0.1 | 0.0 |
| :--- | :--- | :--- | :--- | :--- |
| $y_{1}: 35-39.9 \mathrm{~N} / \mathrm{mm}^{2}$ | 0.7 | 0.2 | 0.2 | 0.1 |
| $y_{\mathbf{2}}: 40-44.9 \mathrm{~N} / \mathrm{mm}^{2}$ | 0.2 | 0.6 | 0.6 | 0.2 |
| $y_{\mathbf{3}}: 45-49.9 \mathrm{~N} / \mathrm{mm}^{2}$ | 0.1 | 0.1 | 0.1 | 0.7 |
| $y_{\mathbf{4}}: 50-60 \mathrm{~N} / \mathrm{mm}^{2}$ | 0.0 | 0.1 | 0.1 |  |

If the engineer takes three subsequent cores, and the laboratory tests yield $z_{(1)}=41, z_{(2)}=49$ and $z_{(3)}$ $=44 \mathrm{~N} / \mathrm{mm}^{2}$, respectively, what are the posterior probabilities of the four states at the end of the experiment? The required posterior probability is given by $\operatorname{Pr}\left[\right.$ state $x_{i} \mid$ sample $\left.z_{(3)}=y_{2}\right]$.

## Solution.

Imperfect concrete testing. The prior probabilities are
$\left.\mathrm{C}_{1}\right) 35$ to $39.9 \mathrm{~N} / \mathrm{mm}^{2}$
$\underline{0.2}$
C2) 40 to 44.9
0.3
C3) 45 to 49.9
0.4
C4) 50 to 80
0.1

## Conditional probabilities

| Measured state $\downarrow$ | True state $\rightarrow$ | $\mathrm{C}_{1}$ | $\mathrm{C}_{2}$ | $\mathrm{C}_{3}$ | $\mathrm{C}_{4}$ |
| :--- | :--- | :--- | :--- | :--- | :--- |
| Cample 1) $\rightarrow \mathrm{C}_{2}$ |  | 0.7 | 0.2 | 0.1 | 0.0 |
| Sample 3) $\rightarrow$ | 0.2 | 0.6 | 0.2 | 0.1 |  |
| Sample 2) $\rightarrow \mathrm{C}_{3}$ |  |  |  |  |  |
|  | $\mathrm{C}_{4}$ | 0.1 | 0.1 | 0.6 | 0.2 |
|  |  | 0.0 | 0.1 | 0.1 | 0.7 |

SAMPLE (1). The denominator for Bayes theorem
$=0.2 \times 0.2+0.6 \times 0.3+0.2 \times 0.4+0.1 \times 0.1=0.04+0.18+0.08+0.01=0.31$
Posterior probabilities $\left.\operatorname{Pr}\left[C_{1} \mid S_{1}\right] \quad \operatorname{Pr}\left[C_{2} \mid S_{1}\right)\right] \quad \operatorname{Pr}\left[C_{3} \mid S_{1}\right] \quad \operatorname{Pr}\left[C_{4} \mid S_{1}\right]$
$0.04 / 0.31=0.129 \quad 0.18 / 0.31=0.581 \quad 0.08 / 0.31=0.258 \quad 0.01 / 0.31=0.0320$
SAMPLE (2). The denominator for Bayes theorem $=$
$0.1 \times 0.129+0.1 \times 0.581+0.6 \times 0.258+0.2 \times 0.032=0129+0.0581+0.1548+0.0064=0.2322$
Posterior probs $\operatorname{Pr}\left[\mathrm{C}_{1} / \mathrm{S}_{2}\right] \quad \operatorname{Pr}\left[\mathrm{C}_{2} / \mathrm{S}_{2}\right] \quad \operatorname{Pr}\left[\mathrm{C}_{3} / \mathrm{S}_{2}\right]$
$\operatorname{Pr}\left[\mathrm{C}_{4} / \mathrm{S}_{2}\right]$
$0.0129 / .2322=0.055 \quad .0 .0581 / .2322=0.25 \quad 0.1548 / .2322=0.667 \quad 0.0064 / .2322=$
0.028

SAMPLE (3) The denominator for Bayes theorem $=0.2 \times 0.055+0.6 \times 0.25+0.2 \times 0.667$ $+0.1 \times 0.028=0.011+0.15+0.1334+.0028=0.2972$

Posterior probabilities $\operatorname{Pr}\left[C_{1} / S_{3}\right]=.011 / .2972=\underline{0.04} \operatorname{Pr}\left[C_{2} / S_{3}\right]=0.15 / 0.2972=\underline{0.50}$
$\operatorname{Pr}\left[C_{3} / S_{3}\right]=0.1334 / .2972=\underline{0.45} \quad \operatorname{Pr}\left[C_{4} / S_{3}\right]=0.0028 / 0.2972=\underline{0.01}$
2.23. Highway pavement. Before any $250-\mathrm{m}$ length of a pavement is accepted by the State Highway Department, the thickness of a $30-\mathrm{cm}$ is monitored by an ultrasonics instrument to verify compliance to specification. Each section is rejected if the measured thickness is less than 10 cm ; otherwise, the entire section is accepted. From past experience, the State Highway engineer knows that 85 percent of all sections constructed by the contractor comply with specifications. However, the reliability of ultrasonic thickness testing is only 75 percent, so that there is a 25 percent chance of erroneous conclusions based on the determination of thickness with ultrasonics.
(1) What is the probability that a poorly constructed section is accepted on the basis of the ultrasonics test?
(2) What is the probability that if a section is well constructed, it will be rejected on the basis of the ultrasonics test?

## Solution.

Let $B_{1}$ represent a well constructed section and $B_{2}$ represent a poorly constructed section
Prior probabilities: $\operatorname{Pr}\left[B_{1}\right]=0.85: \operatorname{Pr}\left[B_{2}\right]=0.15$.
Likelihoods:

$$
\underline{\text { True State }} B_{j} \rightarrow
$$

Measured state $B_{i} \downarrow \quad j=1$ : good construction $j=2$ : poor construction
$i=1$ : good construction
$i=2$ : poor construction
0.75
0.25
0.25
0.75

$$
\operatorname{Pr}\left[B_{k} \mid \text { Test Result }\right]=\frac{\operatorname{Pr}\left[\text { TestResult } \mid B_{\mathrm{k}}\right] \operatorname{Pr}\left[B_{k}\right]}{\operatorname{Pr}\left[\operatorname{TestResult} \mid B_{1}\right] \operatorname{Pr}\left[B_{1}\right]+\operatorname{Pr}\left[\operatorname{TestResult} \mid B_{2}\right] \operatorname{Pr}\left[B_{2}\right]} \text { for } k=1,2
$$

$$
\begin{align*}
& \operatorname{Pr}\left[B_{2}: \text { Poorcons } \mid \text { GoodTest Result }\right]  \tag{1}\\
& =\frac{\operatorname{Pr}[\text { GoodTestResult } \mid \text { Poorcons }] \operatorname{Pr}[\text { Poorcons }]}{\operatorname{Pr}[\text { GoodTestResult } \mid \text { Poorcons }] \operatorname{Pr}[\text { Poorcons }]+\operatorname{Pr}[\text { GoodTestResult } \mid \text { Goodcons }] \operatorname{Pr}[\text { Goodcons }]} \\
& =\frac{0.25 \times 0.15}{0.25 \times 0.15+0.75 \times 0.85}=\frac{0.0375}{0.675}=0.055 \\
& \operatorname{Pr}\left[B_{1}: \text { Goodcons } \mid \text { PoorTest Result }\right]  \tag{2}\\
& =\frac{\operatorname{Pr}[\text { PoorTestResult } \mid \text { Goodcons }] \operatorname{Pr}[\text { Goodcons }]}{\operatorname{Pr}[\text { PoorTestResult } \mid \text { Goodcons }] \operatorname{Pr}[\text { Goodcons }]+\operatorname{Pr}[\text { PoorTestResult } \mid \text { Poorcons }] \operatorname{Pr}[\text { Poorcons }]} \\
& =\frac{0.25 \times 0.85}{0.25 \times 0.85+0.75 \times 0.15}=\frac{0.2125}{0.325}=0.65
\end{align*}
$$

2.24. Remote sensing of inundated areas. Two independent satellite-borne sensors are used to determine the extension of inundated areas after a flood. Sensor $A$ has a reliability of 70 percent, that is, the probability of detecting a pixel (picture element) whose characteristics reflect inundation is .7 , whereas sensor $B$ has a reliability of 90 percent. Also, the probability of both sensors detecting a pixel is .65 .
(a) Find the probability that a pixel reflecting inundation is detected, that is, it is detected by at least one of the two sensors.
(b) Determine the probability that a pixel reflecting flooding is detected by only one sensor.

## Solution.

$|\leftarrow---------A------------\rightarrow|$

| 0.05 | $0.65=\boldsymbol{A B}$ | 0.25 |
| :---: | :---: | :---: |
|  |  |  |

(a) $\operatorname{Pr}[A+B]=\operatorname{Pr}[A]+\operatorname{Pr}[B]-\operatorname{Pr}[A B]=0.7+0.9-0.65=\underline{0.95}$.
(b) $\operatorname{Pr}[$ only one sensor detection $]=$

$$
=\{\operatorname{Pr}[A]-\operatorname{Pr}[A B]\}+\{\operatorname{Pr}[B]-\operatorname{Pr}[A B]\}=0.05+0.25=0.30 .
$$

2.25. Runoff production. Characterization of the soils of a small catchment includes 40 percent of well-drained sand and gravel (type A hydrologic soil group), 35 percent of fine-textured soils (type C hydrologic soil group), and 25 percent of clay soils (type D hydrologic soil group). Type A and Type D terrains have been contoured and are covered with small grains in poor condition, 60 percent of type C terrains are covered by pasture in fair condition, and the remaining type C terrain is sparsely forested land without forest litter. The engineer evaluates runoff production using the Soil Conservation Service procedure (see: Soil Conservation Service, "Section 4: Hydrology", 1985). This procedure gives surface runoff $R$ as

$$
R=(P-0.2 S)^{2} /(P+0.8 S)
$$

where $P$ is the rainfall depth of the design storm, and $S$ is the maximum soil potential retention, which is given by

$$
S=25 \cdot 4(1000 / \mathrm{CN}-10)
$$

where CN is a dimensionless parameter known as the "Curve Number". The values of CN range from 0 to 100 depending on the joint categories of "hydrologic soil group", and "land use" according to the table below. $R, P$, and $S$ are measured in millimeters per unit area.

Values of CN obtained by matching hydrologic soil group with land use.

|  | Hydrologic Soil Group |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  | A | B | C | D |
| Land use |  |  |  |  |
| Straight row crops in poor condition | 72 | 81 | 88 | 91 |
| Contoured row crops in poor condition | 70 | 79 | 84 | 88 |
| Contoured row crops in good condition | 65 | 75 | 82 | 86 |
| oured small grain in poor condition | 63 | 74 | 82 | 85 |
| Pe.. | $\ldots$ | ... |  | ... |
| Pasture in fair condition | 49 | 69 | 79 | 84 |
|  | $\ldots$ | $\ldots$ |  | ... |
| Wood and forest land with thin stand, poor cover, no mulch | 45 | 66 | 77 | 83 |
| Woods protected from grazing with adequate brush coverage | 30 | 55 | 70 | 77 |
| Commercial and business areas ( 85 percent inpervious) | 89 | 92 | 94 | 95 |
| ... | ... | ... | ... | ... |

(a) Determine the expected surface runoff caused by a heavy storm resulting in 120 mm of rainfall per unit area.
A new commercial and business area is planned ( 85 percent impervious). The site includes 40 percent of type A terrains but 60 percent of this is pasture land. The engineer has two alternatives: (1) designing a large culvert to carry runoff excess due to urbanization, or (2) improving the hydrologic conditions of the surrounding forest land (for example, by protecting woods from grazing and providing adequate brush coverage) so that the expected runoff from the catchment does not change. The design storm is 120 mm .
(b) Determine the expected excess runoff due to urbanization.
(c) Evaluate the feasibility of the second alternative under (a).

## Solution.

Design storm $=120 \mathrm{~mm}$
(a) Terrain
A (40\%)
C (35\%)
D (25\%)
$\mathrm{CN}=63 \quad \mathrm{CN}=79 \times 0.6+77 \times 0.4=78.2$
CN=85

Mean CN= $63 \times 0.4+78.2 \times 0.35+85 \times 0.25=73.82$
$S=25.4[(1000 / 73.82)-10]=90.08 \mathrm{~mm}: \quad R=(120-18.02)^{2} / 192.06=54.15 \mathrm{~mm}$
(b) New $\mathrm{CN}=(49 \times 0.6+89 \times 0.4) \times 0.4+94 \times 0.35+95 \times 0.25=82.65$
$S=25.4[(1000 / 82.65)-10]=53.32 \mathrm{~mm}: \quad R=(120-10.66)^{2} / 162.66=73.50 \mathrm{~mm}$
Therefore, expected runoff due to urbanization $=73.50-54.15=19.35$.
(c) Do not increase expected runoff under (b) if possible.

Keep the pasture land same under Type $A$ group. But the remaining $40 \%$ of Type $A$ group change to woods protected with $\mathrm{CN}=30$. New CN is
$(49 \times 0.6+30 \times 0.4) \times 0.4+94 \times 0.35+95 \times 0.25=73.21$.
Hence $S=92.95 \mathrm{~mm}$ and $R=52.9 \mathrm{~mm}$.
Yes.
2.26. Universal soil loss equation. In the United States the prediction of upland erosion amounts is frequently made by the universal soil loss equation (USLE) developed by the U.S.D.A. Agricultural Research Service in cooperation with U.S.D.A. Soil Conservation Service and certain experimental stations (see: Soil Conservation Service, "National Engineering Handbook, Section 3, Sedimentation", 1983). The USLE gives the annual soil loss due to erosion in kilograms per square meter per year, say, $A$, as

$$
A=c R \times K \times L \times S \times C \times P
$$

where $c$ is a constant, $R$ denotes the rainfall factor, $K$ the soil erodibility factor, $L$ the slope length factor, $S$ the slope gradient factor, $C$ the crop-management factor, and $P$ the erosion control practice factor. The engineer must analyze the effects of crop management on the annual soil loss in a small forested catchment. From previous computations $c=1, R=185, K=0.38, L S=1.4$, and $P=1$. The values of $C$ vary from 0.0005 to 0.009 depending on the joint variation of the percentage of area covered by the canopy of trees and undergrowth $C_{1}$ and of the percentage of area covered by litter, $C_{2}$, as shown in the table below.

|  |  | $\mathrm{C}_{1}=$ | 100 to 90.1 \% | 90 to 70.1 \% | 70 to 40 \% |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $\mathrm{C}_{2}=$ | 100 to $70 \cdot 1$ \% |  | 0.0005 | 0.0008 | 0.0010 |
|  | $\begin{aligned} & 70 \text { to } 401 \% \\ & 40 \text { to } 20 \% \end{aligned}$ |  | $\begin{aligned} & 0.0020 \\ & 0.0030 \end{aligned}$ | $\begin{aligned} & 0.0030 \\ & 0.0060 \end{aligned}$ | $\begin{aligned} & 0.0040 \\ & 0.0090 \end{aligned}$ |

(a) Assuming that all the foregoing categories of crop management are equally likely, compute the probability that $A$ exceeds $0.3 \mathrm{~kg} / \mathrm{m}^{2}$ per year.
(b) Assuming that the catchment is partitioned as in the foregoing table into nine subcatchments equal in area, each having a different crop management, compute the expected annual soil loss from the catchment.
(d) What is the minimum number of subcatchments where crop management must be improved in order to reduce the expected annual soil loss from the catchment to a value lower than 0.2 ?

## Solution.

(a) $R=185, K=0.38, L S=1.40, P=1.0, C=1$
$C$ as given : 9 values $\rightarrow 0.0005$ to 0.009
$A=c R \times K \times L \times S \times C \times P=98.42 C$.
For $A>0.30 .3 \mathrm{~kg} / \mathrm{m}^{2}$ per year
$C>0.3 / 98.42=0.003048$
Probability $=3 / 9$
(b) Mean $\bar{C}=0.0293 / 9=0.00326$

Mean $\bar{A}=98.42 \times 0.00326=0.32 \mathrm{~kg} / \mathrm{m}^{2}$
(c) Expected annual soil loss $<0.2 \mathrm{~kg} / \mathrm{m}^{2}$ $\bar{C}=0.002032$.
Before, the sum of $C$ from 9 sub catchments $=0.0293$.
Now it should be $0.002032 \times 9=0.0183$.
Reduction $=0.0293-0.0183=0.0110$.
Maximum number of sub catchments where crop management must be improved $=2$ (the last two on the bottom row of the table of 9 subcatchments).

Blank page

## Applied Statistics for Civil and Environmental Engineers

Problem Solution Manual
by N.T. Kottegoda and R. Rosso

## Chapter 3 - Random Variables and their Properties

3.1. Sea waves. The pmf of the number of days per month of high-amplitude waves, $X$, acting on a sea pier is given below.

| $X=$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | $\geq 7$ |
| ---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $p_{X}(x)$ | $=$ | 0.38 | 0.22 | 0.18 | 0.13 | 0.09 | 0.06 | 0.03 |

Determine the expected value and variance of $X$.

## Solution.

```
    \(X=\begin{array}{lllllllll}0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & >7\end{array}\)
\(\mathrm{p}_{X}(x)=0.28 \quad 0.22 \quad 0.18 \quad 0.13 \quad 0.09 \quad 0.06\)
\(E[X]=0 \times 0.28+1 \times 0.22+2 \times 0.18+3 \times 0.13+4 \times 0.09+5 \times 0.06+6 \times 0.03\)
\(+7 \times 0.01=1.88\).
\(E\left[X^{2}\right]=0 \times 0.28+1 \times 0.22+2 \times 2 \times 0.18+3 \times 3 \times 0.13+4 \times 4 \times 0.09+5 \times 5 \times\)
\(0.06+6 \times 6 \times 0.03+7 \times 7 \times 0.01=6.66\).
\(\operatorname{Var}[X]=\mathrm{E}\left[X^{2}\right]-[\mathrm{E}[X]]^{2}=6.66-1.88 \times 1.88=3.13\)
```

3.2. Tensile strength. The tensile strength of a structural material is found to be highly variable, although tests showed that there is an increasing number of specimens of high strengths with a possible limit of $20 \mathrm{~N} / \mathrm{mm}^{2}$ in strength. Based on observations and as a first approximation, the pdf of tensile strength $X$ is represented by the function $f_{X}(x)=a x^{2}, 0 \leq x \leq 20 \mathrm{~N} / \mathrm{mm}^{2}$.
(a) Determine the constant $a$ in the function.
(b) What is the probability of $X>10 \mathrm{~N} / \mathrm{mm}^{2}$ ?

Solution. Find constant $a$
$\int_{0}^{20} f_{X}(x) d x=\int_{0}^{20} a x^{2} d x=a\left[x^{3} / 3\right]_{0}^{20}=1$.
Hence $a=3 / 8000$
$\operatorname{Pr}[X>10]=\int_{10}^{20}(3 / 8000) x^{2} d x=(3 / 8000) \times(1 / 3)(8000-1000)=7 / 8=0.875$.
3.3. Wind load. A tower is subject to a horizontal force caused by high winds. An important factor which should be taken into account when strengthening the tower is the duration of the winds. The duration $T$ of winds in the area is a random variable with a maximum of 18 hours. From observations of wind data, the pdf of $T$ can be approximated to the form $f_{T}(t)=c t^{1.5}$, with a maximum ordinate of $k$.
(a) Evaluate $c$ and $k$.
(b) Find the mean and coefficient of variation of $T$.
(c) What is the probability of a wind lasting more than nine hours?

## Solution.

(a) $\int_{0}^{18 . .} f_{T}(t) d t=\int_{0}^{18 .} c t^{1.5} d t=c\left[t^{5 / 2} /(5 / 2)\right]_{0}^{18 .}=1$

Hence $c=(5 / 2) / 18^{5 / 2}=0.00182$ and
$f_{T}(t)=(5 / 2)\left(1 / 18^{5 / 2}\right) t^{1.5}$
$k=\operatorname{Max} f_{T}(t)=(5 / 2)\left(1 / 18^{5 / 2}\right) 18^{1.5}=(5 / 2) \times(1 / 18)=5 / 36=0.138$
(b) $E[T]=\mu_{T}=\int_{0}^{18} t c t^{1.5} d t=c \int_{0}^{18} t^{2.5} d t=5 /\left(2 \times 18^{5 / 2}\right) 18^{3.5} / 3.5=18 \times 5 / 7=12.857$
$E\left[T^{2}\right]=\int_{0}^{18} t^{2} c t^{1.5} d t=c \int_{0}^{18} t^{3.5} d t=5 /\left(2 \times 18^{5 / 2}\right) 18^{4.5} / 4.5=18^{2} \times 5 / 9$
$\operatorname{Var}[T]=18 \times 18(5 / 9-25 / 49)=18 \times 18 \times 25 \times 4 /(45 \times 49)$
$\sigma_{T}=12 \times 5^{0.5} / 7=3.83$
$V_{T}=\sigma_{T} / \mu_{T}=\left(12 \times 5^{0.5} / 7\right) \times 7 /(18 \times 5)=2 / 3 / 5^{0.5}=.298$
$\operatorname{Pr}[T>9]=c \int_{9}^{18} t^{1.5} d t=\left(18^{5 / 2}-9^{5 / 2}\right) 5 /\left(2 \times 18^{5 / 2} \times 5 / 2\right)=1-1 / 2^{2.5}=0.823$
3.4. Flood exceedance. A flow of magnitude $40 \mathrm{~m}^{3} / \mathrm{sec}$ is exceeded at a particular site on a river once in three months on average. What is the probability of having at least one such high flow in a year? State assumptions made.

Solution. Assume that the time $T$ between exceedances of $40 \mathrm{~m}^{3} / \mathrm{sec}$ are exponentially distributed.
The mean time interval $\mathrm{E}[T]=1 / \lambda=1 / 4$ year. Hence $\lambda=4$.
$\operatorname{Pr}[T>\mathrm{t}]=1-\exp (-\lambda t) . \quad$ For $t=1$ (year),
$\operatorname{Pr}[T>1]=1-\exp (-4 \times 1)=0.982$
3.5. Compressive strength of concrete. The expected value of the compressive strength of a particular concrete is $60 \mathrm{~N} / \mathrm{mm}^{2}$ and the coefficient of variation is 10 percent. Assuming that the theoretical probability distribution is symmetrical but is unknown, calculate the probability that the compressive strength will be greater than $50 \mathrm{~N} / \mathrm{mm}^{2}$.

Solution. The distribution is symmetrical but is unknown. Use the Chebyshev inequality. Given that the mean $\mu=60 \mathrm{~N} / \mathrm{m}^{3}$, the standard deviation $\sigma=6 \mathrm{~N} / \mathrm{m}^{3}$ and let
$\sigma k=10$; then $k=10 / 6$.
$\operatorname{Pr}[\mu-\sigma k<X<\mu+\sigma k] \geq 1-1 / k^{2}=1-36 / 100=0.64$. Then
$\operatorname{Pr}[50<X<70] \geq 0.64$. Also because the distribution is symmetrical,
$\operatorname{Pr}[X \geq 70]=\operatorname{Pr}[X \leq 50]<0.18$
$\operatorname{Pr}[X>50] \geq 0.82$.
3.6. Highway accidents. Highway accidents along a busy highway leading away from a city have the following pmf (see Example 3.18 for this Poisson pmf):

$$
p_{X}(x)=v^{x} \frac{e^{-v}}{x!}, \quad \text { for } x=0,1,2, \ldots
$$

Originally $v$ has been estimated as 0.9 . Subsequently the exit road was widened and the parameter was estimated as 0.5 . Plot the pmf in each case and determine the probabilities of $\operatorname{Pr}[X>0]$.

## Solution.

(1)
$p_{X}(x)=e^{-0.9} 0.9^{x} / x!$
(2) $p_{X}(x)=e^{-0.5} 0.5^{x} / x$ !

| ${ }^{X}$ |  |  |  |  |  |  |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | :--- |
| (1) | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|  | 0.407 | 0.366 | 0.165 | 0.049 | 0.011 | 0.002 | 0.0003 |
| (2) |  | 0.607 | 0.303 | 0.076 | 0.013 | 0.002 | 0.0016 |

3.7. Earthquake occurrence. During a period of 125 years, 16 major earthquakes have occurred in the San Francisco area. Assuming these are Poisson events (see problem 3.6 and Example 3.18), determine
(a) the probability of more than one such earthquake during a five-year period, and
(b) the mean time between such earthquakes.

## Solution.

$$
p_{X}(x)=e^{-\lambda} \lambda^{x} / x!
$$

(a) Mean number of occurrences per year $\lambda=16 / 25$. Hence $p_{X}(0)=e^{-16 / 25}$ and $p_{X}(1)=e^{-16 / 25} \times 16 / 25$.
Therefore $\operatorname{Pr}[X>1]=1-p_{X}(1)-p_{X}(0)$
$=1-e^{-16 / 25} \times(1+16 / 25)=0.135$. Also
(b)Mean time between such earthquakes $=125 / 16=7.81$ years
3.8. Computer system failure. The times to failure in months of several identical computer systems are observed as follows: $21,53,43,56,18,17,40,14,13$. Assuming these are distributed as $F_{T}(t)=1$ -$e^{-\lambda t}$, estimate the parameter $\lambda$ by the method of maximum likelihood. Repeat the procedure using the method of moments.

## Solution.

$f_{X}(x)=\lambda e^{-\lambda t}$
Mean number of occurrences $\bar{x}=\sum_{i=1}^{n} x_{i} / n=275 / 9$. Hence,
$\hat{\lambda}=9 / 275$, by the method of moments.
Maximum likelihood $\quad L=\prod_{i=1}^{9} \lambda e^{-\lambda t_{i}}=\lambda^{9} e^{-275 \lambda}$
Log likelihood, $L L=9 \ln \lambda-275 \lambda$
$\frac{\partial L L}{\partial \lambda}=9 / \lambda-257$
Hence by equating the derivative to zero, we also obtain
$\hat{\lambda}=9 / 275$, by the method of maximum likelihood.
3.9. Maximum flows. In some applications the exponential distribution of Problem 3.8 is written with a lower bound $\varepsilon$ and this takes $F_{T}(t)=1-\exp [-\lambda(t-\varepsilon)]$. Show how the parameters may be estimated using the probability-weighted moments procedure.

## Solution.

$$
\begin{aligned}
& F(x)=1-\exp [-\lambda(x-t)] \\
& x(F)=\varepsilon-(1 / \lambda) \ln (1-F),
\end{aligned}
$$

where $F \equiv F(x)$.
$M_{10 k}=\mathrm{E}\left[X(1-F)^{\mathrm{k}}\right]=\varepsilon \int_{0}^{1}(1-F)^{k} d F-(1 / \lambda) \int_{0}^{1} \ln (1-F)(1-F)^{k} d F$.
Let $z=1-F$. Hence, $\mathrm{d} z=-\mathrm{d} F$; for $F=0, z=1$; for $F=1, z=0$.
$M_{10 k}=\varepsilon \int_{0}^{1} z^{k} d z-(1 / \lambda) \int_{0}^{1}(\ln z) z^{k} d z$
$=\varepsilon\left[z^{k+1} /(k+1)\right]_{0}^{1}-(1 / \lambda)(-1) /(k+1)^{2}$
$=\varepsilon /(k+1)+(1 / \lambda)\left[1 /(k+1)^{2}\right]$
$M_{100}=\varepsilon+(1 / \lambda) ; M_{101}=(1 / 2)[\varepsilon+1 /(2 \lambda)]$. Hence
$\lambda=(1 / 2)\left[1 /\left(M_{100}-2 M_{101}\right] \quad\right.$ and $\quad \varepsilon=4 M_{101}-M_{100}$
Proceed by equating moments to plotting positions, as in Example 3.21, but inserting $\left(1-p_{i}\right)^{k}$
3.10. Occurrence of volcanic eruptions. There are frequent volcanic eruptions at a particular site. The times of the occurrences are unpredictable. From past observations, the pmf of occurrences $X$ over periods of ten years is as follows:

| $X=$ | 0 | 1 | 2 | 3 |
| ---: | :---: | :---: | :---: | :---: |
| $p_{X}(x)=$ | 0.1 | 0.3 | 0.4 | 0.2 |

What entropy does this distribution represent? What is the maximum possible entropy for possible four values of probability?

## Solution.

| X | 0 | 1 | 2 | 3 |
| :--- | :--- | :--- | :--- | :--- |
| $p_{X}(x)$ | 0.1 | 0.3 | 0.4 | 0.2 |

From Eq. (3.2.27) entropy is given by
$-\sum(0.1 \ln 0.1+0.3 \ln 0.3+0.4 \ln 0.4+0.2 \ln 0.2)$
$=+0.2303+0.3612+0.3665+0.3219=1.2799$.
Maximum $=-\ln 0.25=1.3863 \quad$ (for uniform distribution)
3.11. Pipe settlement. Three subcontractors laid water pipes running through a flat part of a city. Excavations made at 3 -meter intervals along the pipelines after a period of five years showed that settlements had taken place from the original levels. The following table gives the settlements in millimeters at each excavation:

| Sub-Contractor 1 | 181 | 190 | 71 | 55 | 105 |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Sub-Contractor 2 | 99 | 78 | 25 | 50 | 198 |
| Sub-Contractor 3 | 23 | 23 | 197 | 75 | 189 |

If in a particular case, the settlements have been the same at each point of observation along the pipeline, no problem will arise with regard to the system. On the basis of entropy, determine the relative settlement of the pipes laid by each subcontractor. Which system has the least relative settlement? What is the entropy of a particular system with no relative settlement?

## Solution.

$$
1^{\text {st }} \text { subcontractor } \quad \text { 2nd subcontractor } \quad \text { 3rd }
$$

subcontractor
Sum
of settlements: 502450
507
Entropy calculations;
$\begin{array}{llll}(-181 / 502) \ln (181 / 502)= & 3678 & (-99 / 450) \ln (99 / 450)= & .3331 \\ (-190 / 502) \ln (190 / 502)=.3677 & (-78 / 450) \ln (78 / 450)= & .3038 & (-23 / 507) \ln (23 / 507) \ln (23 / 507)=.1403 \\ (-71 / 502) \ln (71 / 502)= & .2766(-25 / 450) \ln (25 / 450)= & .1606(-197 / 507) \ln (197 / 507)=.3673 \\ (-55 / 502) \ln (55 / 502)= & .2443(-50 / 450) \ln (50 / 450)= & .2441 \quad(-75 / 507) \ln (75 / 507)=.2827 \\ (-105 / 502) \ln (105 / 502)=.3273(-198 / 450) \ln (198 / 450)=.3612(-189 / 507) \ln (189 / 507)=.3678 \\ \text { Sums } & 1.5817 & 1.4028 & 1.2984\end{array}$
On the basis of entropy ,pipes laid by $1^{\text {st }}$ contractor has the least relative settlement. Maximum entropy $=\sum-0.2 \times \ln (0.2)=\ln (5)=1.61$
3.12. Project scheduling. In a building project, the construction of the foundations takes time $T_{1}$ and the construction of the superstructure takes time $T_{2}$. On account of inclement weather, labor problems and other factors, $T_{1}$ and $T_{2}$ behave like random variables with empirical pmf's as follows:

| Time, in weeks | $\mathbf{1}$ | $\mathbf{2}$ | $\mathbf{3}$ | $\mathbf{4}$ | $\mathbf{5}$ | $\mathbf{6}$ | $\mathbf{7}$ |
| ---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $p_{T_{1}}\left(t_{1}\right)=$ | 0.1 | 0.3 | 0.4 | 0.2 | 0.0 | 0.0 | 0.0 |
| $p_{T_{2}}\left(t_{2}\right)=$ | 0.0 | 0.0 | 0.0 | 0.1 | 0.5 | 0.4 | 0.0 |

(a) Calculate the mean times taken for the foundations and the superstructure.
(b) Evaluate the pmf of the total time spent on the foundations and superstructure.
(c) What is the probability that the total work will be completed in less than seven weeks?

Solution. Consider all possibilities for time in weeks: $\boldsymbol{F}$, foundations: $\boldsymbol{S}$, Superstructure.

| $\boldsymbol{F}$ | $\boldsymbol{S}$ | $\boldsymbol{F}$ | $\boldsymbol{S}$ | $\boldsymbol{F}$ | $\boldsymbol{S}$ | $\boldsymbol{F}$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 1 | 4 | 2 | 4 | 3 | 4 | 4 |
|  | 5 |  | 5 | 5 |  |  |
|  | $\mathbf{6}$ | $\mathbf{6}$ |  | $\mathbf{6}$ |  |  |
|  |  | $\mathbf{6}$ |  |  |  |  |

(a) Mean $T_{1}$ foundations $=1 \times 0.1+2 \times 0.3+3 \times 0.4+4 \times 0.2=.1+.6+1.2+.8=2.7$ weeks Mean $T_{1}$ Superstructure $=4 \times 0.1+5 \times 0.5+6 \times 0.4=0.4 \quad+2.5 \quad+2.4=5.3$ weeks
(b) Pmf of total time spent on foundations and superstructure $T=T_{1}+T_{2}$

|  | $\boldsymbol{F}$ | $\boldsymbol{S}$ | $\boldsymbol{F}$ | $\boldsymbol{S}$ |
| :--- | :---: | :---: | :---: | :---: |
| 5 weeks | $1 \times 0.1+4 \times 0.1$ | $\boldsymbol{F}$ | $\boldsymbol{S}$ | sum |
| 6weeks | $1 \times 0.1+5 \times 0.5$ | $+2 \times 0.3+4 \times 0.1$ |  | $=0.5$ |
| 7weeks | $1 \times 0.1+6 \times 0.4$ | $+2 \times 0.3+5 \times 0.5$ | $+3 \times 0.4+4 \times 0.1$ | $=3.6$ |
| 8weeks | $2 \times 0.3+6 \times 0.4$ | $+3 \times 0.4+5 \times 0.5$ | $+4 \times 0.2+4 \times 0.1$ | $=7.9$ |
| 9weeks | $3 \times 0.4+6 \times 0.4$ | $+4 \times 0.2+5 \times 0.1$ |  | $=6.9$ |
| 10 weeks | $4 \times 0.2+6 \times 0.4$ |  |  | $=3.2$ |


| Total time $T$ in weeks | 5 | 6 | 7 | 8 | 9 | 10 |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: |
| Sum from above right | 0.5 | 3.6 | 7.2 | 7.9 | 6.9 | $3.2=29.3$ |
| $\mathrm{P}(T)$ | 0.017 | 0.123 | 0.246 | 0.270 | 0.235 | $0.109=1.00$ |

(c) $\operatorname{Pr}[T<7]=0.017+0.123=0.140$
3.13. Sea pier construction. With reference to the data given in Problem 3.1, a contractor is assigned to work on an extension to the sea pier. The contractor finds that the profits $Y$ of the job are directly decreased by the number of days per month $X$ of high-amplitude waves acting on the sea front. It is estimated that $Y=10000(10-X)$. Determine the pmf of $Y$ and the mean and variance of $Y$.

## Solution.

| $Y$ | $=$ | 100,000 | 90,000 | 80,000 | 70,000 | 60,000 | 50,000 | 40,000 | 30,000 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $X$ | $=$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| $p_{X}(x)$ | $=$ | 0.28 | 0.22 | 0.18 | 0.13 | 0.09 | 0.06 | 0.03 | 0.01 |

$Y=10,000(10-X)$
$E[Y]=100,000-10,000 E[X]=100,000-10,000 \times 1.88=81,200$
$\operatorname{Var}[Y]=10000^{2} \operatorname{Var}[X]=10000^{2} \times 3.14=314,000,000$
3.14. Head loss in a pipe. The head loss $H$ in a pipe is related to the mean velocity of flow $V$ as $H=k V^{2}$, where $k$ is a constant depending on pipe length, diameter and roughness. In a particular case, $V$ varies randomly between limits $\nu_{1}$ and $\nu_{2}$. Assuming a symmetrical triangular pdf for $V$, derive the pdf of $H$.

## Solution.

Case [a]. Head loss in a pipe, symmetrical pdf for flow $V$
Sketch of pdf with $v_{1}=-v^{*}$ and $v_{2}=+v^{*}$


$$
\begin{aligned}
& \begin{aligned}
f_{V}(v) & =\frac{1}{v^{*}}\left(1-v / v^{*}\right) \quad \text { for } 0 \leq v \leq v^{*} \\
& =\frac{1}{v^{*}}\left(1+v / v^{*}\right) \quad \text { for }-v^{*} \leq v \leq 0
\end{aligned} \\
& H=k V^{2} \Rightarrow f_{H}(h)=\frac{1}{\sqrt{k h}}\left[f_{v}\left(\sqrt{\frac{h}{k}}\right)\right]=\frac{1}{\sqrt{k h}} \frac{1}{v^{*}}\left(1-\frac{\sqrt{h / k}}{v^{*}}\right) .
\end{aligned}
$$

Case [b]. This is simply a translation of the previous case, because physically $v_{1} \geq 0$.
Let

$$
Y=V+\left(v_{1}+v_{2}\right) / 2
$$


$0 \quad v_{1} \frac{v_{1}+v_{2}}{2} \quad v_{2}$
$f_{Y}(y)=\frac{4}{\left(v_{2}-v_{1}\right)^{2}}\left(v_{2}-y\right) \quad$ for $\frac{v_{1}+v_{2}}{2} \leq y \leq v_{2}$
$f_{Y}(y)=\frac{4}{\left(v_{2}-v_{1}\right)^{2}}\left(y-v_{1}\right) \quad$ for $v_{1} \leq y \leq \frac{v_{1}+v_{2}}{2}$
$H=k Y^{2} \Rightarrow f_{H}(h)=\frac{1}{2 \sqrt{k h}}\left[f_{y}\left(\sqrt{\frac{h}{k}}\right)\right] \Rightarrow$
$f_{H}(h)=\frac{1}{2 \sqrt{k h}}\left[\frac{4}{\left(v_{2}-v_{1}\right)^{2}}\left(v_{2}-\sqrt{\frac{h}{k}}\right)\right]$ for $\frac{v_{1}+v_{2}}{2} \leq \sqrt{\frac{h}{k}} \leq v_{2}$
$f_{H}(h)=\frac{1}{2 \sqrt{k h}}\left[\frac{4}{\left(v_{2}-v_{1}\right)^{2}}\left(\sqrt{\frac{h}{k}}-v_{1}\right)\right]$ for $v_{1} \leq \sqrt{\frac{h}{k}} \leq \frac{v_{1}+v_{2}}{2}$
3.15. Joint wind measurements. For the joint pdf of the number of days of occurrences of high winds recorded by two instruments and given in Table 3.3.1, evaluate the probability that the differences between the readings of the two instruments are not greater than 1 .

Solution.
$\begin{array}{llllll}\text { Diff. } 1 & 0.0400 & 0.0600 & 0.0100 & 0.0300 & 0.1750\end{array}$

Diff. 2
Diff. 3

| $\mathrm{y}=0$ | $\mathrm{y}=1$ | $\mathrm{y}=2$ | $\mathrm{y}=3$ | Sum |
| :---: | :---: | :---: | :---: | :---: |
| 0.2910 | 0.3580 | 0.1135 | 0.0505 | 0.8130 |
| 0.0400 | 0.0600 | 0.0100 | 0.0300 | 0.1750 |
|  | 0.0250 | 0.0100 | 0.0000 |  |
| 0.0100 | 0.0015 | 0.0000 | 0.0000 | 0.0115 |
| 0.0005 | 0.0000 | 0.0000 | 0.0000 | 0.0005 |
|  |  |  |  | 1.0000 |

Sum

Hence $\operatorname{Pr}[$ Diff. $<2]=0.988$.
3.16. Contract analysis. A contractor's financial outlay $X$ and labor force $Y$ are random variables with bivariate pdf given by:

$$
\begin{aligned}
& f_{X, Y}(x, y)=k x y, \quad \text { for } 10,000<x<100,000 \text { and } 10<y<20, \\
& \text { and }=0, \quad \text { elsewhere. }
\end{aligned}
$$

(a) Evaluate constant $k$.
(b) Determine the marginal pdf's of $X$ and $Y$.

## Solution.

$\iint f_{X, Y}(x, y)=\iint k x y=k \int_{10,000}^{100,000} x \int_{10}^{20} y d y d x$
$k\left[\frac{20^{2}-10^{2}}{2}\right]\left[\frac{100000^{2}-10000^{2}}{2}\right]=1$.
Hence $k=4 /(300 \times 90,000 \times 110,000)$ for $10,000<x<100,000$.
$f_{X}(x)=\int_{0}^{20} k x y d y=k x\left(\frac{20^{2}-10^{2}}{2}\right)=2 x /\left(99 \times 10^{8}\right)$
Similarly $\quad f_{Y}(y)=\frac{1}{150} y \quad$ for $10<y<20$
3.17. Welding legs. The joint pdf of the lengths of horizontal and vertical legs, $X$ and $Y$, of welding joints (similar to the ones referred to in Problem1.4) is given by

$$
f_{X, Y}(x, y)=\frac{1}{16} x y, \quad \text { for } 4.0<x, y<8.0
$$

and $=0$, elsewhere.
Determine the probability $\operatorname{Pr}[5.5<X<6.5 ; 5.5<Y<6.5]$.

## Solution.

Verify $\iint f_{X, Y}(x, y) d y d x=\int_{4}^{8} \int_{4}^{8} \frac{1}{16 \times 36} x y d y d x$
$=\frac{\left(8^{2}-4^{2}\right)\left(8^{2}-4^{2}\right)}{16 \times 36 \times 4}=1$.
$\operatorname{Pr}[5.5<x<6.5 ; 5.5<y<6.5]=\int_{5.5}^{6.5} x \int_{5.5}^{6.5} \frac{y d y d x}{16 \times 36}$
$=\frac{12 \times 1 \times 12 \times 1}{16 \times 36 \times 4}=\frac{1}{16}$.
3.18. Density and compressive strength of concrete. Estimate the correlation in the case of the simplified joint distribution of concrete density and compressive strength given in Example 3.37 and shown in Fig. 3.3.5 from the data of Table E.1.2.

## Solution.

$f_{X, Y}(x, y)=\frac{1}{2000} \frac{y-40}{20} \quad$ for $40 \leq y \leq 60 ; 2400 \leq x \leq 2,500$

$$
=\frac{1}{2000}\left[1-\frac{y-60}{20}\right] \text { for } 60 \leq y \leq 80 ; 2400 \leq x \leq 2,500
$$

Verify $\quad \int_{x} \int_{y} f_{x, y}(x, y) d y d x=\frac{1}{20}\left\{\left[\frac{y^{2}}{40}-2 y\right]_{40}^{60}+\left[4 y-\frac{y^{2}}{40}\right]_{60}^{80}\right\}$
$=\frac{1}{20}\left[\frac{60^{2}}{40}-2 \times 60-\frac{40^{2}}{40}+2 \times 40+4 \times 20-\frac{80^{2}}{40}+\frac{60^{2}}{40}\right]=1$
$f_{X}(x)=\frac{1}{100}$ for $2400 \leq x \leq 2,500$
$E[X]=\int_{2400}^{2500} x f_{X}(x) d x=\frac{1}{100}\left[\frac{x^{2}}{2}\right]_{2400}^{2500}=2450$,
which is simply the mid-range.

$$
\begin{aligned}
f_{Y}(y) & =0.5 \frac{y-40}{20} & & \text { for } 40 \leq y \leq 60 \\
& =0.5\left(1-\frac{y-60}{20}\right) & & \text { for } 60 \leq y \leq 80
\end{aligned}
$$

$E[Y]=\int y f_{Y}(y) d y=0.05\left\{\left[\frac{y^{3}}{60}-y^{2}\right]_{40}^{60}+\left[2 y^{2}-\frac{y^{3}}{60}\right]_{60}^{80}\right\}$
$=\frac{1}{20}\left(\frac{60^{3}}{60}-60^{2}-\frac{40^{3}}{60}+40^{2}+2 \times 80^{2}-\frac{80^{3}}{60}-2 \times 60^{2}+\frac{60^{3}}{60}\right)$
$=60$.
$E[X Y]=\iint x y f_{X, Y}(x, y) d x d y=\int_{2400}^{2500} \frac{x}{2000}\left[\int_{40}^{60} y \frac{y-40}{20} d y+\int_{60}^{80} y\left[1-\frac{y-60}{20}\right] d y\right] d x$
$=\frac{4900 \times 100}{2000 \times 2}\left(\left[\frac{y^{3}}{60}-y^{2}\right]_{40}^{60}+\left[2 y^{2}-\frac{y^{3}}{60}\right]_{60}^{80}\right)$
$=147,000$.
$\operatorname{Cov}[X Y]=\mathrm{E}[X Y]-\mathrm{E}[X] \mathrm{E}[Y]=147,0002450 \times 60=0$
3.19. Contractor's profits, financial outlay, and labor force. For the pdf given in Problem 3.16, the contractor's profits $P$ may be assumed to be related to his financial outlay $X$ and labor force $Y$ as follows:
(a) $P=1.3 X+15,000$
(b) $\quad P=1.2 X+1000 Y+10,000$

Determine the pdf of $P$ in each case.

## Solution.

(a) $\frac{d P}{d x}=1.3$

$$
f_{P}(p)=f_{X}(x) \frac{d x}{d p}=\frac{p-15000}{1.3} \frac{2}{99 \times 10^{8}}=\frac{p-15000}{8.3655 \times 10^{9}}
$$

Check: for $x=10,000, p=28,000$; also $x=100,000, p=145,000$

$$
\int_{28000}^{145000} f_{P}(p) d p=\frac{1}{8.3655 \times 10^{9}}\left[\frac{p^{2}}{2}-15000 p\right]_{28000}^{145000}=1
$$

$$
\begin{equation*}
f_{X, Y}(x, y)=k x y \quad \text { for } 10<y<20 . \tag{b}
\end{equation*}
$$

Also $\quad X=\frac{P-1000 Y-10000}{1.2}$
$f_{P}(p)=\frac{k}{1.2} \int_{10}^{20} y(p-1000 y-10000) d y=\frac{k}{1.2}\left(\left[\frac{y^{2}}{2}\right]_{10}^{20}(p-1000)-1000\left[\frac{y^{3}}{3}\right]_{10}^{20}\right)$
$=k\left[125 p-\frac{11500000}{3 \times 1.2}\right]$
From Problem 3.16, $k=\frac{4}{300 \times 90000 \times 110000}$. Hence

$$
f_{P}(p)=\frac{p}{5.94 \times 10^{9}}-\frac{1}{232434.78}
$$

3.20. Rivet production. Two machines produce rivets for a factory job. The numbers of sub-standard rivets per hour by the two machines are random variables denoted by $X_{1}$ and $X_{2}$. The bivariate pmf of $X_{1}$ and $X_{2}$ is given by the following table:

|  | $\boldsymbol{X}_{\mathbf{2}}=\mathbf{0}$ | $\boldsymbol{X}_{\mathbf{2}}=\mathbf{1}$ | $\boldsymbol{X}_{\mathbf{2}}=\mathbf{2}$ | $\boldsymbol{X}_{\mathbf{2}}=\mathbf{3}$ | $\boldsymbol{p}_{\boldsymbol{X}_{\mathbf{1}}}\left(\boldsymbol{x}_{\mathbf{1}}\right)$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $X_{1}=0$ | .07 | .05 | .02 | .01 | .15 |
| $X_{1}=1$ | .05 | .16 | .12 | .02 | .35 |
| $X_{1}=2$ | .02 | .12 | .17 | .05 | .36 |
| $X_{1}=3$ | .01 | .01 | .05 | .07 | .14 |
| $p_{X_{2}}\left(x_{2}\right)$ | .15 | .34 | .36 | .15 | $\sum=1.00$ |

(a) Determine the probability that the number of substandard rivets produced do not differ by more than 1 between one machine and the other.
(b) Determine the conditional distribution of $P_{X_{2} \mid X_{1}}\left(x_{2}, x_{1}\right)$.
(c) The factory manager estimated that an older machine, which was replaced, produced $X_{1}+$ $X_{2}$ substandard rivets per hour. Estimate its marginal pmf.

## Solution.

(a) $\operatorname{Pr}\left[-1 \leq X_{1}-X_{2} \leq+1\right]$

$$
\begin{aligned}
= & (0.07+0.05)+(0.05+0.16+0.12)+(0.12+0.17+0.05)+(0.05+0.07)= \\
& 0.91
\end{aligned}
$$

(b) Conditional distributions

|  | $X_{2}=0$ | $X_{2}=1$ | $X_{2}=2$ | $X_{2}=3$ | Sum |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $X_{1}=0$ | $\frac{0.07}{0.15}=0.47$ | $\frac{0.05}{0.15}=033$ | $\frac{0.02}{0.15}=0.13$ | $\frac{0.01}{0.15}=0.07$ | 1.00 |
| $X_{1}=0$ | $\frac{0.05}{0.35}=0.14$ | $\frac{0.16}{0.35}=0.46$ | $\frac{0.12}{0.35}=0.34$ | $\frac{0.02}{0.35}=0.06$ | 1.00 |
| $X_{1}=0$ | $\frac{0.02}{0.36}=0.06$ | $\frac{0.12}{0.36}=0.33$ | $\frac{0.17}{0.36}=0.47$ | $\frac{0.05}{0.36}=0.14$ | 1.00 |
| $X_{1}=0$ | $\frac{0.01}{0.14}=0.07$ | $\frac{0.01}{0.14}=0.07$ | $\frac{0.05}{0.14}=0.36$ | $\frac{0.07}{0.14}=0.50$ | 1.00 |

(c)

| $Y=0$ | 0 | 0 |
| :--- | :--- | :--- |
| $Y=1$ | 0 | 1 |
| $Y=2$ | 1 | 0 |
|  | 0 | 2 |
| $Y=3$ | 1 | 1 |
|  | 2 | 0 |
|  | 0 | 3 |
| $Y=4$ | 1 | 2 |
|  | 2 | 1 |
|  | 3 | 0 |
| $Y=5$ | 1 | 3 |
|  | 2 | 2 |
| $Y=6$ | 3 | 1 |
|  | 2 | 3 |
|  | 3 | 2 |
|  | 3 | 3 |

$$
p_{Y}(y)
$$

$Y=0$

3.21. Earthquake hazard. Two adjoining regions are subject to earthquakes at irregular intervals. The first region experiences $X_{1}$ earthquakes over a period of time, and $X_{2}$ earthquakes occur in the second region over the same period, where $X_{1}$ and $X_{2}$ are random variables. It is estimated that the joint distribution of earthquakes over the two regions is as follows:
and

$$
p_{X_{1}, X_{2}}\left(x_{1}, x_{2}\right)=\frac{x_{1}+x_{2}}{21}, \quad \text { for } x_{1}=0,1,2 \quad \text { and } \quad x_{2}=2,3
$$

$$
=0, \quad \text { elsewhere } .
$$

Determine the probabilities $p_{X_{1} \mid X_{2}}\left(x_{1} \mid x_{2}\right)$ and the expected values $E\left[X_{1} \mid X_{2}\right]$.

## Solution.

|  | $X_{2}=2$ | $X_{2}=3$ |
| :--- | :---: | :---: |
| $X_{1}=0$ | $\frac{2}{21}$ | $\frac{3}{21}$ |
| $X_{1}=1$ | $\frac{3}{21}$ | $\frac{4}{21}$ |
| $X_{2}=2$ | $\frac{4}{21}$ | $\frac{5}{21}$ |
| Sum of marginals | $\frac{9}{21}$ | $\frac{12}{21}$ |

Conditional probabilities $p_{X_{1} \mid X_{2}}\left(x_{1} \mid x_{2}\right)$

|  | $X_{2}=2$ | $X_{2}=3$ |
| :--- | :---: | :---: |
| $X_{1}=0$ | $\frac{2}{9}$ | $\frac{3}{12}$ |
| $X_{1}=1$ | $\frac{3}{9}$ | $\frac{4}{12}$ |
| $X_{2}=2$ | $\frac{4}{9}$ | $\frac{5}{12}$ |

## Conditional expectations

$E\left[X_{1} \mid X_{2}=2\right]=\frac{2}{9} \times 0+\frac{3}{9} \times 1+\frac{4}{9} \times 2=\frac{11}{9}$
$E\left[X_{1} \mid X_{2}=3\right]=\frac{3}{12} \times 0+\frac{4}{12} \times 1+\frac{5}{12} \times 2=\frac{14}{12}$
3.22. Water treatment plant. A water treatment plant has two units which are designed to perform with identical characteristics. The consequenses of both units failing simultaneously are severe on the community. The times to failure in days are denoted by $X_{1}$ and $X_{2}$ and their bivariate pdf is given by

$$
f_{X_{1}, X_{2}}\left(x_{1}, x_{2}\right)=a e^{-b\left(x_{1}+x_{2}\right)}, \quad \text { or } x_{1}, x_{2} \geq 0
$$

(1) What is the relationship between the constants $a$ and $b$ ?
(2) How may they be estimated in practice?
(3) What is the chance that both units will fail within a year?

## Solution.

In general $f_{X, Y}(x, y)=a b e^{-a x-b y}$
But the two plants have identical characteristics and it is given by
$f_{X_{1}, X_{2}}\left(x_{1}, x_{2}\right)=a e^{-b\left(x_{1}+x_{2}\right)}$
Therefore, $a=b^{2}$
$f_{X_{1}, X_{2}}\left(x_{1}, x_{2}\right)=b^{2} e^{-b x_{1}} e^{-b x_{2}}$
$F_{X_{1}, X_{2}}\left(x_{1}, x_{2}\right)=1-e^{-b x_{1}}-e^{-b x_{2}}+e^{-b\left(x_{1}+x_{2}\right)}$
$F_{X_{1}, X_{2}}(365,365)=1-e^{-b \times 365}-e^{-b \times 365_{2}}+e^{-2 b \times 365}$
3.23. Water treatment plant. In Problem 3.22 a change in design is made so that only one of the units needs to operate at a time. The second will be brought into operation only on failure of the first, whenever that happens. What is the probability that the plant will be inoperative within a year?

## Solution.

Let $X=$ time to failure of first unit in days
Let $Y=$ time to failure of first unit in days
$S=X+Y$ is the time to failure of the plant
The required probability is $\operatorname{Pr}[S=X+Y \leq 365$ days $]$.
The pdf of S is simply the convolution of $X, \mathrm{Y}$, where
$f_{X, Y}(x, y)=a b e^{-a x-b y} \quad$, for $x, y>0$.
$f_{S}(s)=\int_{0}^{s} f_{X, Y}(x, s-x) d x \quad$, for $s>0$.
$=a \int_{0}^{s} e^{-b(x+s-x)} d x=a s e^{-b s}$.
Then $\operatorname{Pr}[S=X+Y \leq 365$ days $]=\int_{0}^{365} a s e^{-b s} d s=\frac{a}{b^{2}}\left[1-365 b e^{-365 b}-e^{-365 b}\right]$
(integration by parts).
3.24. Sewerage pollution discharge. Two sewage plants serving different communities discharge a pollutant into a stream. The concentrations of the respective discharges are measured as $X$ and $Y$ parts per million. Suppose the bivariate distribution is given by

$$
f_{X, Y}(x, y)=2-x-y
$$

for $0 \leq x, y \leq 1$, and 0 elsewhere.
(a) Determine the joint probability $\operatorname{Pr}[X<0.5, Y<0.6]$,
(b) If $x \leq 0.5$, determine the distribution of $Y$.
(c) Determine the coefficient of linear correlation between $X$ and $Y$.

## Solution.

(a) $\operatorname{check} \int_{0}^{1} \int_{0}^{1}(2-x-y) d y d x=\int_{0}^{1}(3 / 2-x) d x=1$.
$\int_{0}^{0.50 .6} \int_{0}^{0.6}(2-x-y) d y d x=\int_{0}^{0.5}\left[(2-x) y-y^{2} / 2\right]_{0}^{0.6} d x$
$\int_{0}^{0.5} 0.6(2-x-0.3) d x=0.6\left[1.7 x-x^{2} / 2\right]_{0}^{0.5}=0.435$.
(b) $f_{Y \mid X \leq 0.5}(y, x)=\left[\int_{0}^{0.5}(2-x-y) d x\right] / \int_{0}^{0.5} \int_{0}^{1}(2-x-y) d y d x$

$$
=\frac{0.875-0.5 y}{0.625}=1.4-0.8 y
$$

$\operatorname{check} \int_{0}^{1}(1.4-0.8 y) d y=1.0$.
(c) $f_{X, Y}(x, y)=2-x-y ; f_{X}(x)=\frac{3-2 x}{2} ; f_{Y}(y)=\frac{3-2 y}{2}$.
$E[X]=\int_{0}^{1} x \frac{3-2 x}{2} d x=5 / 12$.
Hence $E[Y]=5 / 12$
$E\left[X^{2}\right]=\int_{0}^{1} x^{2} \frac{3-2 x}{2} d x=1 / 4$.
$\operatorname{Var}[X]=E\left[X^{2}\right]-(E[x])^{2}=11 / 144$.
$\sigma_{X}=\sigma_{Y}=\sqrt{11} / 12$.
$E[X Y]=\int_{0}^{1} \int_{0}^{1} x y(2-x-y) d x d y=\int_{0}^{1} \int_{0}^{1}\left[x\left(2 y-y^{2}\right)-x^{2} y\right] d x d y$
$=\int_{0}^{1}\left[\left(2 y-y^{2}\right) / 2-y / 3\right] d y=\left[\left(y^{2}-y^{3} / 3\right) / 2-y^{2} / 6\right]_{0}^{1}=1 / 6$.
$\operatorname{Cor}[X, Y]=\frac{E[X Y]-E[X] E[Y]}{\sigma_{X} \sigma_{Y}}=\frac{1 / 6-(5 / 12)^{2}}{11 / 144}=-1 / 11$.
3.25. Dam construction. The times spent in months by a contractor, engaged in the construction of small earthen dams, on the substructure and conduit on the one hand and the dam itself on the other are random variables (on account of frequent interruptions by weather and other unpredictable factors) denoted by $X_{1}$ and $X_{2}$, respectively. These times have common expectations, and past experience suggests that the bivariate pdf can be approximated by

$$
f_{X_{1}, X_{2}}\left(x_{1}, x_{2}\right)=a x_{1} x_{2} e^{-b\left(x_{1}+x_{2}\right)}, \quad \text { for } x_{1}, x_{2}>0
$$

Determine the probability that the time spent on the earthwork is greater than or equal to 1.5 times that on the earthwork.

## Solution.

$f_{X_{1}, X_{2}}\left(x_{1}, x_{2}\right)=a x_{1} x_{2} e^{-b\left(\left(x_{1}+x_{2}\right)\right.}$ for $x_{1}, x_{2}>0$
$\operatorname{Pr}\left[X_{1}>(3 / 2) X_{2}\right]=\operatorname{Pr}\left[X_{2}<(2 / 3) X_{1}\right]$
$=\int_{0}^{\infty} \int_{0}^{2 x / 3_{1}} f_{X_{1}, X_{2}}\left(x_{1}, x_{2}\right) d x_{2} d x_{1}=a \int_{0}^{\infty} x_{1} e^{-b x_{1}} \int_{0}^{2 x_{1} / 3} x_{2} e^{-b x_{2}} d x_{2} d x_{1}=\frac{44 a}{125 b^{4}}$.
3.26. Maximum annual flood. Flood flows at a given river site are assumed to be independent identically distributed variables. The peak flow $X$ for each flood exceeding a level of $a$ is assumed to have a distribution with cdf

$$
F_{X}(x)=1-(a / x)^{\theta}
$$

with $x>a$ (Pareto with parameters $a$ and $\theta$ ). Since flood events occur randomly, the number $N$ of flood flows exceeding $a$ in a year is assumed to be distributed as

$$
p_{N}(n)=\frac{v^{n} e^{-v}}{n!}
$$

for $n=0,1,2, \ldots,($ Poisson with parameter $v)$. Show that the probability distribution of the annual maximum peak flow, $Y$ is

$$
F_{Y}(y)=\exp \left[-\left(\frac{x_{0}}{x}\right)^{\beta}\right]
$$

[Extreme value Type II (Fréchet) with parameters $x_{0}$ and $\beta$ ]. Find the relationships linking parameters $x_{0}$ and $\beta$ with $v, \theta$ and $a$.

## Solution.

$F_{X}(x)=\operatorname{Pr}[X \leq x]=1-(a / x)^{\theta}, \quad p_{N}(n)=e^{-v} v^{n} / n!$
For fixed $n, F_{X_{(n)}}(x)=\left[F_{X}(x)\right]^{n}=\left[1-\left(\frac{a}{x}\right)^{\theta}\right]^{n}=\operatorname{Pr}\left[X_{(n)} \leq x\right]$
$=\sum_{n=0}^{\infty} \operatorname{Pr}\left[X_{(n)} \leq x \mid N=n\right] \operatorname{Pr}[N=n]$
$=\sum_{n=0}^{\infty}\left[1-\left(\frac{a}{x}\right)^{\theta}\right]^{n} e^{-v} v^{n} / n=e^{-v} \sum_{n=0}^{\infty}\left[v\left[1-\left(\frac{a}{x}\right)^{\theta}\right]\right]^{n} / n!=e^{-v} e^{v\left[1-\left(\frac{a}{x}\right)^{\theta}\right]}=e^{-\left(\frac{a}{x}\right)^{v}}$.
Since this should equal $e^{-\left(\frac{x_{0}}{x}\right)^{\beta}}$ we get $\theta=\beta$.
Also, $v a^{\theta}=x_{0}{ }^{\theta}$.
Hence $x_{0}=a v^{1 / \theta}$

# Applied Statistics for Civil and Environmental Engineers <br> Problem Solution Manual <br> by N.T. Kottegoda and R. Rosso <br> Chapter 4 - Probability distributions 

4.1. Protective sea embankment. To counteract the effects of erosion and damage caused by sea waves, an embankment wall is built alongside a railway line. From recorded data, the annual maximum wave height exceeds that of the embankment, on average, once in eight years. What is the probability that the embankment will be over- topped at least once during the next 10 years? Assume that the events are independent and identically distributed.

Solution. $\operatorname{Pr}[H>h]=1 / 8=p$ (say).
$p_{X}(x)=\operatorname{Pr}[X=x ; n, p]=\binom{n}{x} p^{x}(1-p)^{n-x}$.
$p_{x}(0)=\left(\frac{1}{8}\right)(1 / 8)^{0}(1 / 8)^{10}=0.263$, for $n=10$
$p_{X}(x>0)=1-0.263=0.737$.
4.2. Dam design. Determine the return period that should be used in a design for a small dam so that the design flood is exceeded with a probability of not more than .05 during a 50 -year economic time horizon. Assume that the events are independent and identically distributed.

Solution. $(1-p)^{50}=0.95$.
Hence $1-p=0.99897 ; p=0.00103$
$T=1 / p=975 \approx 1000$.
4.3. Bridge design. A bridge is to be constructed over a river. The design criterion is that a flood should rise above the high-level marks on the piers in not more than once in 25 years with a probability not exceeding .1. What return period should be used in the flood design? Assume that the events are independent and identically distributed.

## Solution.

$\binom{25}{0} p^{0}(1-p)^{25}+\binom{25}{1} p^{1}(1-p)=0.90$
$=(1-p)^{25}+25 p(1-p)^{24}=(1-p)^{24}(1-p+25 p)$
Hence $(1-p)=[0.90 /(1+24 p)]^{1 / 24}$.
By numerical solution $p=0.0216$. Hence $T=1 / 0.0216=46.3 \approx 50$.
4.4. Revision of dam design. In a situation similar to that of Problem 4.3, supposing the engineer adopts a 100-year design period, determine (a) the probability that the design flood level is not exceeded during a 100 -year period (b) the probability that the design flood is exceeded just after the tenth year but not during the first10 years.

Solution. $\quad(a) p=0.01 ; p_{X}(x)=\binom{100}{0} p^{0}(1-p)^{100}=0.366$
(b) The required probability is $0.99^{10} \times 0.01=0.00904$.
4.5. Frequent flooding. Calculate the probability of having two 10 -year flows in a 5 - year period assuming that the events are independent and identically distributed.

Solution. $p_{x}(x=2 ; 5,0.1)=\binom{5}{2} p^{2}(1-p)^{3}=10 \times 0.1^{2} \times 0.9^{3}=0.0729$.
4.6. Storm sewer design. For a storm sewer design, an engineer uses the annual maximum one-hour rainfall with a five-year return period as a design criterion. As shown on the city plan, sewer A drains one area of the city and sewer B drains the remaining area. However, there is no correlation in the intensive rainfalls which occur in the two parts of the city, although the storm characteristics are the same. What is the probability that there will not be more than two design events in the city during a five- year period?

Solution. $n=5+5=10 ; p=0.2$.
$p_{X}(0)=\binom{10}{0} p^{0}(1-p)^{10}=(1-p)^{10}=0.10737$
$p_{X}(1)=\binom{10}{1} p^{1}(1-p)^{9}=10 p(1-p)^{9}=0.26844$
$p_{X}(2)=\binom{10}{2} p^{2}(1-p)^{8}=45 p^{2}(1-p)^{8}=0.30199$
$p_{x}(0)+p_{x}(1)+p_{x}(2) \quad=0.678$
4.7. Vehicle count. The following count is made on the number of vehicles that pass an observation point every 10 minutes for one hour. What counts are expected theoretically if the distribution is Poisson?

| Count $i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Frequency $f$ | 220 | 94 | 23 | 11 | 4 | 2 | 1 |

## Solution.

| $i=$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | Sum |
| :--- | ---: | :---: | :---: | :---: | :---: | :---: | :---: | :--- |
| $f=$ | 220 | 94 | 23 | 11 | 4 | 2 | 1 | $355=\Sigma f=n$ |
| $i f=$ | 0 | 94 | 46 | 33 | 16 | 10 | 6 | $205=\Sigma i f$ |

$\hat{\lambda}=\Sigma$ if $/ \Sigma f=205 / 355=0.5775$
Poisson counts $=$
$n p_{X}=355 \frac{.5775^{x} e^{-.5775}}{x!}$

$$
\begin{array}{lllllll}
199 & 115 & 33 & 6 & 1 & 0 & 0
\end{array}
$$

4.8. Machine failure. The probability that a certain make of piling machine breaks down is 0.00002 per 100 m of piles made. What is the probability of having one breakdown after 1000 m and before 1010 m of piles?

## Solution.

$$
\begin{aligned}
& p_{X}(X=x ; p)=p(1-p) \quad \text { for } x=1,2,3, \ldots \\
& =0 \quad \text { otherwise } \\
& p=0.00002 ; x=11 \\
& p_{X}(X=11 ; 0.00002)=0.00002(0.99998)^{10}=0.000019996
\end{aligned}
$$

4.9. First-time failure. Taking the probability given in problem 4.8 as the probability of failure during a week's work and an average weekly production of 1000 m of piles, determine the probability of failure for the first time after three months. How does the first-time probability of failure vary with time?

## Solution.

3 months +1 week; $x=13+1=14$ weeks
Just after 3 months, $p_{X}(X=14 ; 0.00002)=0.00002(0.99998)^{13}=0.00002$
Just after lyear $\quad p_{X}(X=53 ; 0.00002)=0.00002(0.99998)^{53}=0.0000200$
Just after 5years $p_{X}(X=261 ; 0.00002)=0.00002(0.99998)^{261}=0.0000199$
Just after 10years $p_{X}(X=521 ; 0.00002)=0.00002(0.99998)^{521}=0.0000198$
For very low values of the probability of failure $p$, the probability of first time failure shows insignificant decreases as time increases.
4.10. Transportation. An operator runs a small bus which conveys people from a town centre to a large shopping complex. The bus leaves as soon as 12 people have arrived. If we assume that the passenger arrivals are independent and are at a mean rate of nine per hour, what is the probability that the time between two consecutive departures is more than 60 minutes? Assume that there are no delays caused by the nonarrival of the bus because standby buses are available.

## Solution.

Nine people arrive every 60 minutes on average.
Therefore, 12 people arrive every 80 minutes on average
Mean time between departures of the bus $=80$ minutes; i.e., $\lambda=1 / 80 \mathrm{~min}^{-1}$.
Assume exponential distribution. Hence $F_{X}(X>60)=e^{-60 / 80}=0.472$
4.11. Traffic: number of cars waiting to turn. For the control of vehicles at a traffic light one needs to determine the length of the left-turn lane (right-turn lane in countries where vehicles are driven on the left). The occurrences of left (right)-turns are assumed to have a Poisson distribution in time. The mean uninterrupted rate of left (right) turns is 160 per hour and the red light is on for 50 seconds. What is the expected number of vehicles awaiting a left (right) turn?

## Solution.

Mean uninterrupted rate of left turns $=160 / 60$ per minute
For the number of vehicles awaiting a turn, the estimated parameter $v=$ $(160 / 60) \times(50 / 60)=20 / 9$ per cycle of 50 seconds . The mean number is close to 2 .
4.12. Traffic: length of lane. In Problem 4.11 the design criterion for the length of the left (right) lane is that it should be sufficient for 95 percent of the time. What should be the minimum length of the lane as a multiple of the average length of a vehicle?

## Solution.

$\operatorname{Pr}[X \leq i]=\sum_{x=0}^{i} \frac{1}{x!}\left(\frac{20}{9}\right)^{x} e^{-20 / 9}$
For $i=0 \quad \operatorname{Pr}[X=0]=\frac{1}{0!}\left(\frac{20}{9}\right)^{0} e^{-20 / 9}=0.1084$
Similarly,
For $i=1, \quad \operatorname{Pr}[X \leq 1]=\operatorname{Pr}[X=0] \quad+\quad \operatorname{Pr}[X=1] \quad=0.1084$
$+\frac{1}{1!}\left(\frac{20}{9}\right)^{1} e^{-20 / 9}=0.1084+0.2408=0.3492$.
For $i=2, \operatorname{Pr}[X \leq 2]=\operatorname{Pr}[X \leq 1]+\operatorname{Pr}[X=2]=0.3492+0.2676=0.6168$
For $i=3, \operatorname{Pr}[X \leq 3]=\operatorname{Pr}[X \leq 2]+\operatorname{Pr}[X=3]=0.6168+0.1982=0.8150$
For $i=4, \operatorname{Pr}[X \leq 4]=\operatorname{Pr}[X \leq 3]+\operatorname{Pr}[X=4]=0.8150+0.1101=0.9251$
For $i=5, \operatorname{Pr}[X \leq 5]=\operatorname{Pr}[X \leq 4]+\operatorname{Pr}[X=5]=0.9251+0.0489=0.9740$
The required minimum length of lane $=5$ vehicle lengths.
4.13. Wet spells The following distribution of wet spells was observed at the Dharamjaigarh rainfall station in central India during the monsoon season:

| $i$, length of wet spell in days | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $O_{i}$, observed number of spells | 161 | 52 | 32 | 17 | 8 | 6 | 4 | 1 |

What is the maximum length of wet spell which is exceeded with probability less than .05 assuming a geometric distribution?

## Solution.

| $i$ | $=$ | 1 | 2 |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $i$ | $=$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Sum |
| $O_{i}$ | $=161$ | 52 | 5 | 6 | 7 | 8 | Sum |  |  |  |
| $O_{i}$ | $=161$ | 52 | 32 | 17 | 8 | 6 | 4 | 1 | 271 |  |
| $i O_{i}$ | $=161$ | 104 | 96 | 68 | 40 | 36 | 28 | 8 | 541 |  |
| $\hat{p}=271 / 541=0.5009 \approx 0.5$. |  |  |  |  |  |  |  |  |  |  |

Find minimum $n$ so that
$p_{X}(X=n ; 0.5)=0.5 \times 0.5^{n-1} \leq 0.05$

Minimum $n=4.32$, Nearest minimum integer $=5$. In detail:

$$
\begin{array}{ll}
p_{X}(X=1 ; 0.5)=0.5 \times 0.5^{0}=0.5 & 0.5000 \\
p_{X}(X=2 ; 0.5)=0.5 \times 0.5^{1}=0.25 & 0.7500 \\
p_{X}(X=3 ; 0.5)=0.5 \times 0.5^{2}=0.125 & 0.8750 \\
p_{X}(X=4 ; 0.5)=0.5 \times 0.5^{3}=0.0625 & 0.9375 \\
p_{X}(X=5 ; 0.5)=0.5 \times 0.5^{4}=0.03125 & 0.9688
\end{array}
$$

Required length in days $=5$

$$
p_{X}(X=n ; 0.5)=0.5 \times 0.5^{n-1} \leq 0.05
$$

4.14. Stream pollution. Traces of toxic wastes from an unknown source are found in a stream. From tests made on the water the mean concentration is found to be $1 \mathrm{mg} / \mathrm{L}$. What is the probability that the concentration of the pollutant will be in the range 0.5 to $2 \mathrm{mg} / \mathrm{L}$ assuming the distribution is (a) exponential (b) normal?

## Solution.

(a) Exponential. $F_{X}(2)=1-e^{-2 / 1}=1-0.13533 ; \quad \quad F_{X}(0.5)=1-e^{-0.5 / 1}=1-$ 0.60653 ;'

Hence $F_{X}(2)-F_{X}(0.5)=0.4712$
(b) Normal is not an appropriate distribution here. Assume the standard deviation is also $1 \mathrm{mg} / \mathrm{L}$
Standard normal variates $z_{1}=(2-1) / 1=1 \quad z_{2}=(0.5-1) / 1=-0.5$
$\Phi(1)=0.84134$ and $\Phi(-0.5)=0.30854$. Hence $\Phi(1)-\Phi(-0.5)=0.53280$.
4.15. Failure of pumps installed in parallel. A pumped storage power supply scheme has five pumps of identical specification installed in parallel. The mean life span of a pump is estimated as 10 years from previous experience. What is the minimum number of pumps required in parallel so that the probability of not having a failure of the system during a three-year period is more than .95 ?

Solution. Assuming an exponential distribution, the probability of failure of a single unit in a parallel system is $1-e^{-3 / 10}$ Hence for $n$ units, we find minimum $n$ such that
$0.05 \leq\left[1-e^{-3 / 10}\right]^{n}$
$n \geq \ln (0.5) / \ln \left[1-e^{-3 / 10}\right]=2.218$
Hence $n=3$
4.16. Failure of pumps in a compound system. Suppose that in the scheme described in Problem 4.15, two pumps are placed in parallel, one of which must work. This subsystem is combined in series with another identical pump. Determine the probability of not having a failure of the system in any year.

Solution. The reliability of a single unit is $e^{-3 / 10}$. The reliability of a parallel system of 2 units is $1-\left[1-e^{-3 / 10}\right]^{2}$. Hence the required probability is
$\left(1-\left[1-e^{-3 / 10}\right]^{2}\right) e^{-3 / 10}=0.691$.
4.17. Traffic accidents. From experience it is found that there are about three accidents per year at an intersection. If the occurrences are Poisson-distributed, what is the pdf of the time till the fourth accident?

Solution. The Poisson parameter $\lambda$ is estimated as $1 / 3$. We apply the Erlang distribution (Eq. 4.2.7), with $X$ denoting the time to the $r$ th arrival of the Poisson process, and $r=4$.

$$
f_{X}(x)=\frac{\lambda(\lambda x)^{r-1} e^{-\lambda x}}{(r-1)!}=\frac{1}{2 \times 3} \times \frac{1}{3}\left(\frac{x}{3}\right)^{3} e^{-x / 3}=x^{3} e^{-x / 3} / 486 .
$$

4.18. Defective valves. A manufacturer supplies nine valves for a pumping scheme. Two faulty valves were included in the consignment. However, the scheme had been completed using three of the nine valves. What is the probability that no faulty valves were used?

Solution. Use multinomial distribution.
$\operatorname{Pr}[X=0]=\frac{\binom{2}{0}\binom{7}{3}}{\binom{9}{3}}=\left(\frac{7!}{3!4!}\right) /\left(\frac{9!}{3!6!}\right)=\left(\frac{7 \times 6 \times 5}{3 \times 2}\right) /\left(\frac{9 \times 8 \times 7}{3 \times 2}\right)=5 / 12$.
4.19. Gamma-distributed annual runoff. The annual runoff in the Cave Creek, near Fort Spring, Kentucky, U.S.A. are given as follows in millimeters over an 18-year period:

$$
\begin{array}{llllllllllll}
337 & 84 & 385 & 394 & 361 & 538 & 196 & 448 & 582 & 480 & 326 & 294 \\
385 & 264 \\
458 & 413 & 299 & 455 .
\end{array}
$$

Assuming independence and a gamma distribution for the annual runoff, determine the probability that the runoff will be greater than 100 mm in a given year. Data from Haan (1977); used with permission, copyright 1977, The Iowa State University Press.

Solution. Cave Creek flows. Mean $\bar{x}=372.17$; variance $\hat{S}^{2}=14,490.74$.
$\hat{\lambda}=\bar{x} / \hat{s}^{2}=\frac{372.167}{14490.74}=0.02567$.
$\hat{r}=\hat{\lambda} \bar{x}=0.02567 \times 372.17=9.558$.
For $\mathrm{x}=100 \mathrm{~mm}$ in a given year, the chi-squared variate is
$2 x \hat{\lambda}=2 \times 100 \times 0.02567=5.134$.
The degrees of freedom $v=2 r=2 \times 9.558 \approx 19$
$\operatorname{Pr}[X>100] \approx 1-0.001=0.999$.
4.20. Low river flows in a tropical region. The following annual minimum mean daily flows, given in $\mathrm{m}^{3} / \mathrm{s}$, were recorded at the proposed Bango diversion site in the Hasdo subcatchment of the Mahanadi basin in India over a 22 -year period:

$$
\begin{array}{lllllllllll}
2.78 & 2.47 & 1.64 & 3.91 & 1.95 & 1.61 & 2.72 & 3.48 & 0.85 & 2.29 & 1.72 \\
2.41 & 1.84 & 2.52 & 4.45 & 1.93 & 5.32 & 2.55 & 1.36 & 1.47 & 1.02 & 1.73
\end{array}
$$

Assuming a two-parameter Weibull distribution, determine the probability that the annual minimum low flow will be less than $2 \mathrm{~m}^{3} / \mathrm{s}$ over a two-year period.

## Solution.

We follow the least squares procedure of Eqs. (4.2.18). The ranked flows $x$ preceded by the $y$ and $z$ values are as follows:

| $y$ | -3.51 | -2.55 | -2.05 | -1.71 | -1.44 | -1.21 | -1.02 | -0.85 | -0.69 | -0.55 | -0.41 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $z$ | -0.16 | 0.02 | 0.31 | 0.39 | 0.48 | 0.49 | 0.54 | 0.55 | 0.61 | 0.66 | 0.67 |
| $x$ | 0.85 | 1.02 | 1.36 | 1.47 | 1.61 | 1.64 | 1.72 | 1.73 | 1.84 | 1.93 | 1.95 |
|  |  |  |  |  |  |  |  |  |  |  |  |
| $y$ | -0.28 | -0.16 | -0.03 | 0.09 | 0.22 | 0.35 | 0.48 | 0.63 | 0.80 | 1.03 | 1.42 |
| $z$ | 0.83 | 0.88 | 0.90 | 0.92 | 0.94 | 1.00 | 1.02 | 1.25 | 1.36 | 1.49 | 1.67 |
| $x$ | 2.29 | 2.41 | 2.47 | 2.52 | 2.55 | 2.72 | 2.78 | 3.48 | 3.91 | 4.45 | 5.32 |

Hence $\bar{z}=0.7644 ; \quad \bar{y}=2.591 ; \quad \hat{\beta}=2.7728 ; \quad \hat{\lambda}=2.591$.
$F_{X}(2)=1-\exp \left[-\left(\frac{2}{2.591}\right)^{2.7728}\right]=1-e^{-0.488}=0.386$.
Assuming that the low flows are independent, the probability that the annual minimum low flow will be less than $2 \mathrm{~m}^{3} /$ s over a two-year period is $\left[F_{X}(x)\right]^{2} \approx 0.15$.
4.21. Low river flows in a temperate region. Ten years of annual minimum daily mean low-flow data from the River Pang at Pangbourne in hydrometric area 39 in England are ranked and given here in cubic meters per second:

$$
\begin{array}{llllllllll}
11.5 & 23.6 & 29.1 & 32.7 & 34.5 & 37.0 & 39.8 & 49.0 & 54.6 & 53.5 .
\end{array}
$$

Fit a Weibull distribution to the data, estimating the parameters using Eqs. (4.2.17) and Tables C. 5 and C. 6 in Appendix C, noting that $\Gamma(r+1)=\Gamma(r)$. If it is not permissible to pump water from the river when the daily mean low flow is less than $20 \mathrm{~m}^{3} / \mathrm{s}$, estimate the return period of such an event. Data are used by permission from Institute of Hydrology(1980), "Low flow studies report", Institute of Hydrology, Wallingford.

## Solution.

For the given data, mean $\bar{x}=36.53$; variance $\hat{s}^{2}=2183.178$.
Square of the coefficient of variation is $\hat{v}^{2}=183.178 / 36.53^{2}$.
From Eq. $(4.2 .17 c), \frac{1}{\hat{v}^{2}+1}=\frac{\left[\Gamma(1+1 / \hat{\beta}]^{2}\right.}{\Gamma(1+2 / \hat{\beta})}=\frac{1}{183.178 / 36.53^{2}+1}=0.8793$.
From Table C. $6 \quad 1 / \hat{\beta}=0.359$ and hence $\hat{\beta}=2.786$.
From Eq.(4.2.17a), and from Table C. 5 with some interpolation,
$\hat{\lambda}=\frac{\bar{x}}{\Gamma(1+0.359)}=\frac{36.53}{0.359 \Gamma(0.359)}=\frac{36.53}{0.359 \times 2.483}=40.98$. Hence,
$F_{X}(20)=1-\exp \left[-\left(\frac{20}{40.98}\right)^{2.786}\right]=1-0.8733=0.1267$.
$T=1 / F_{X}(20) \approx 7.9$ years.
4.22. Ferry transport schedule. A ferry boat is designed to carry 35 passengers across a lagoon from station A during the busy hours of the day. If the passengers arrive at an average rate of two per five minutes and ferries leave every 70 minutes, what is the probability that there will be more than the stipulated number of passengers waiting to take the boat? How often should a ferry be scheduled to leave station A if the chance of an excess is to be less than 5 percent? Assume that the arrivals of the passengers constitute a Poisson process.

## Solution.

On average, two passengers arrive per 5 minutes.
Ferries leave every 70 minutes.
That is, 28 passengers arrive in 70 minute, on average. Hence,

$$
\operatorname{Pr}[X>35]=e^{-35 / 28}=0.2865 .
$$

For the next question, we condition $e^{-35 \lambda}=0.05$.
$\lambda=0.08559$. That is, $1 / \lambda=11.68$ passengers.
With two passengers arriving at the ferry terminal every 5 minutes, on average, the minimum time required to get the full capacity of the ferry boat of 35 passengers $=$ $11.68 \times 5 / 2 \approx 29$ minutes.
4.23. Distribution of concrete strengths. The compressive strengths of concrete in Table 1.2 .2 have an estimated mean of $60.14 \mathrm{~N} / \mathrm{mm}^{2}$ and a standard deviation of $5.02 \mathrm{~N} / \mathrm{mm}^{2}$ and are assumed to be normally distributed. What is the probability that in ten random tests the compressive strength will be in the range 45 to $75 \mathrm{~N} / \mathrm{mm}^{2}$ ?

## Solution.

$\Phi\left(\frac{72-60.14}{5.02}\right)=\Phi(2.363)=0.99094$
$\Phi\left(\frac{48-60.14}{5.02}\right)=\Phi(-2.418)=1-0.99220$.
The required probability is $[0.99094-(1-0.99220)]^{10} \approx 0.844$.
4.24. Ferry transport: weight restriction. Suppose there is a weight restriction of $2,900 \mathrm{~kg}$ for a ferryboat. Random tests carried out on a large number of incoming passengers establish a mean weight of 75 kg per person and a standard deviation of 25 kg . What is the probability that the total weight of an incoming batch of 35 passengers will exceed the limit?

## Solution.

$E[T]=35 \times 75=2625 ; \operatorname{Var}[T]=35^{2} \times 25^{2}$.
$\operatorname{Pr}[T>2900]=1-\Phi\left(\frac{2900-2625}{35 \times 25}\right)=1-\Phi(0.31429)$
$=1-0.62235 \approx 0.378$.
4.25. Monthly rainfalls. Monthly rainfalls in a locality are independent and identically distributed normal variates with mean 20 cm . and variance $12 \mathrm{~cm}^{2}$. Determine the probability that 220 cm . of rainfall occurs over a period of 6 months. What is the probability of having less than 18 cm rainfall each month for a period of 6 months?

Solution. Let $R$ represent monthly rainfall and $T$ represent total rainfall over 6 months.

$$
\begin{aligned}
& E[T]=6 \times 20 ; \operatorname{Var}[T]=6 \times 12 \\
& \operatorname{Pr}[T>220]=1-\Phi\left(\frac{220-120}{6 \sqrt{12}}\right)=1-\Phi(4.8113)=0.0000 .
\end{aligned}
$$

For the second question we assume that monthly rainfalls are independent.
$[\operatorname{Pr}[R<18]]^{6}=\left[\Phi\left(\frac{18-20}{12}\right)\right]^{6}=[\Phi(-0.16667)]^{6}=(1-0.56618)^{6}$
$=0.0067$.
4.26. Relationship between rainfall and runoff. Annual rainfall is usually normally distributed over many river basins around the world. In a particular catchment, annual rainfall $X$ has a mean of 1000 mm and a standard deviation of 200 mm . The annual runoff $Y$ is related to the rainfall as follows:

$$
Y=100+0.4 X
$$

Specify the complete distribution of $Y$. What is the probability that $Y$ will be less than 350 mm in a year?

## Solution.

$E[X]=1000 \mathrm{~mm} ; \operatorname{Var}[X]=200^{2} \mathrm{~mm}^{2}$. Also
$Y=100+0.4 X$
$E[Y]=100+0.4 E[X]=500 \mathrm{~mm}$
$\operatorname{Var}[Y]=0.4^{2} \operatorname{Var}[X]=0.4^{2} \times 200^{2} \mathrm{~mm}^{2}$
$Y \sim N\left[500,80^{2}\right]$.
$\operatorname{Pr}[Y<350]=\Phi\left(\frac{350-500}{0.4 \times 200}\right)=\Phi(-1.875)$
$=1-0.96961 \approx 0.03$.
4.27. River diversion. A river with annual flows $X \sim N(300,50)$ is joined by a major tributary with annual flows $Y \sim N(150,75)$ at point $P$. At point $Q$ on the river below $P$ there is a diversion with annual flows $Z \sim N(100,25)$. The units are in $1000 \mathrm{~m}^{3}$. Below $Q$, suppose the annual flows are denoted by $R$. If $X, Y$ and $Z$ are independent, determine the following:
(a)the distribution of $R$.
(b) $\operatorname{Pr}(R>300)$.

Recalculate $(a)$ and $(b)$ if there are miscellaneous withdrawals and net losses affecting $X$ and $Y$ which total 15 percent in each case.

## Solution.

Sketch of flow system
$\downarrow X$
$\mathrm{P} \leftarrow Y$
$\downarrow X_{1}$
$\mathrm{Q} \rightarrow Z$
$\downarrow R$
(a) $X_{1}=X+Y \sim N\left(\mu_{X}+\mu_{Y}, \sigma_{X}{ }^{2}+\sigma_{Y}{ }^{2}\right)$
$R=X_{1}-Z=X+Y-Z$
$\sim N\left(\mu_{X}+\mu_{Y}-\mu_{Z}, \sigma_{X}{ }^{2}+{\sigma_{Y}}^{2}+\sigma_{Y}{ }^{2}\right)$
$\sim N(300+150-100,50+75+25) \sim N(350,150)$,
(see mgf of normal distribution).
(b) $\operatorname{Pr}\left[\frac{R-\mu_{R}}{\sigma_{R}}>\frac{300-\mu_{R}}{\sigma_{R}}\right]=\operatorname{Pr}\left[Z \sim N(0,1)>\frac{300-350}{\sqrt{150}}\right]$
$=1-\operatorname{Pr}[Z<-50 / \sqrt{150}]=1-\operatorname{Pr}[Z<-4.0825]=1.000$.
(c) $\tilde{\mu}_{X}=0.85 \mu_{X}$ and $\tilde{\mu}_{Y}=0.85 \mu_{Y}$

$$
\widetilde{R} \sim N(282.5,150) ;=1-0.9235 \approx 0.08
$$

4.28. Lognormal distribution of annual river flows. The annual flows, in cubic meters per second, at the Weldon River at Mill Grove, Missouri for the period 1930 to 1960 are averaged as follows:

| 3.06 | 1.52 | 16.60 | 2.78 | 1.15 | 13.39 | 2.74 | 6.16 | 1.21 | 5.90 |  |
| ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
| 4.06 | 2.66 | 11.29 | 8.46 | 7.04 | 12.51 | 10.91 | 16.09 | 3.46 | 4.28 |  |
| 6.92 | 11.35 | 6.95 | 3.23 | 18.70 | 3.75 | 1.25 | 2.06 | 3.83 | 18.02 | 14.41. |

Fit the lognormal distribution to this data. What is the probability that the annual river flow is in the range 2 to $15 \mathrm{~m}^{3} / \mathrm{s}$ ? These data are from Markovic (1965) and are used with permission of Colorado State University.

Solution. Mean $\bar{x}=7.282$; $\operatorname{Var} \hat{\sigma}^{2}=30.079$. We substitute these values in the theoretical equations for the lognormal distribution.
$\sigma_{\ln (X)}=\left[\ln \left(V_{X}^{2}+1\right)\right]^{1 / 2}=\left[\ln \left(\frac{30.079}{7.282^{2}}+1\right)\right]^{1 / 2}=0.6703$
$\mu_{\ln (X)}=\ln \left[\frac{\mu_{X}}{\left(V_{X}{ }^{2}+1\right)^{1 / 2}}\right]=\ln \left[\frac{7.282}{\left(\frac{30.079}{7.282^{2}}+1\right)^{1 / 2}}\right]=1.7607$
$\operatorname{Ln}(15)=2.70821 ; \ln (2)=0.6931$
$\operatorname{Pr}[2<X<15]=\Phi\left(\frac{2.7081-1.7607}{0.6703}\right)-\Phi\left(\frac{0.6931-1.7607}{0.6703}\right)$
$=\Phi(21.4134)-\Phi(-1.5927)=0.92123-(1-0.94438)=0.866$.
4.29. Lognormal distribution of low flows in the Po River, Italy. Low flows in the Po River basin in northern Italy are affected by irrigation releases and return flows. The following are the annual minimum low flows in cubic meters per second occurring at Pontelagoscuro during the period 1 October to 14 April, a period that is outside the irrigation season. There are 18 occurrences during the period 1920 to 1991:

| 735 | 429 | 742 | 828 | 554 | 855 | 787 | 668 | 655 |
| ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | :--- |
| 830 | 732 | 577 | 1030 | 650 | 620 | 561 | 588 | 635 |

The low flows in the lower reaches of the Po river have a two-component lognormal distribution on account of the intervention caused by irrigation (from Kottegoda and Natale, 1994). For the data given, which represents one component, determine the probability that an annual minimum of $400 \mathrm{~m}^{3} / \mathrm{sec}$ can be maintained in the Po at Pontelagoscuro over a three-year period?

Solution Mean $\bar{x}=693.111$; Var $\hat{\sigma}^{2}=19697.8$. We substitute these values in the theoretical equations for the lognormal distribution.
$\mu_{\ln (X)}=\ln \left[\frac{\mu_{X}}{\left(V_{X}{ }^{2}+1\right)^{1 / 2}}\right]=\ln \left[\frac{693.11}{\left(\frac{19697.8}{693.11^{2}}+1\right)^{1 / 2}}\right]=6.521$
$\sigma_{\ln (X)}=\left[\ln \left(V_{X}^{2}+1\right)\right]^{1 / 2}=\left[\ln \left(\frac{19697.8}{693.11^{2}}+1\right)\right]^{1 / 2}=0.20046$
$\operatorname{Ln}(400)=5.9915$.
$\operatorname{Pr}[X>400]=\operatorname{Pr}[Y>5.9915]=1-\Phi\left(\frac{5.9915-6.5211}{0.20046}\right)$
$=1-[1-\Phi(2.6419)=0.99587$.
The required probability is $0.99587^{3}=0.988$
4.30. Ratios of densities of concrete. Densities of concrete (such as those given in Table 1.2.1) can be approximated by a uniform distribution. Taking data from two similar mixes of concrete, determine the distribution of the ratios of the densities, after transformation to $U(0,1)$.

## Solution.

$$
X \sim \mathrm{U}(a x, b y)
$$

$$
Y \sim \mathrm{U}(a y, b y)
$$

$$
\begin{aligned}
& \left\{\frac { Z = X / Y } { W = Y } \Rightarrow \left\{\frac{x=z y}{y=w} \Rightarrow|J|=\left|\frac{w}{-0} \frac{z}{1}\right|=|w|\right.\right. \\
& f_{Z, W}(z, w)=|w| f_{X, Y}(z w, w) \Rightarrow f_{Z}(z)=\int_{I R} f_{Z, W}(z, w) d w=\int_{I R} f_{X, Y}(z w, w)|w| d w
\end{aligned}
$$

Assume $X$ and $Y$ are independent.

$$
\begin{aligned}
f_{X, Y}(x, y)= & f_{X}(x) f_{Y}(y) \Rightarrow f_{Z}(z)=\int_{I R} f_{X}(z w) f_{Y}(w)|w| d w \\
& \int_{I R} \mathrm{I}_{(a x, b x)}(z w) \mathrm{I}_{(a y, b y)}(w)|w| d w \Rightarrow\left\{\begin{array}{l}
\frac{a x<z w<b x}{a y<w<b y}
\end{array}\right.
\end{aligned}
$$

Note: if we could normalize $X, Y$ into $U(0,1) . \Rightarrow f_{Z}(z) \sim \frac{1}{2} \mathrm{I}_{(0,1)}(z)+\frac{1}{2} Z^{-2} \mathrm{I}_{(1, \infty)}(z)$

4.31. Relationship between strengths of construction materials. The strength of a construction material $X$, in newtons per square millimeter, is found to be normally distributed. It is claimed that a new material $Y$ can be produced that is proportional in strength to the square of the strength of $X$. Derive the distribution of $Y$ assuming that $X$ is standardized to zero mean and unit variance

## Solution.

$X \sim N(0,1)$
$Y=X^{2}$
$f_{Y}(y)=\frac{1}{2 \sqrt{y}}\left\{f_{X}(-\sqrt{y})+f_{X}(+\sqrt{y})\right\}, Y \geq 0$.
$X \sim N(0,1) \Rightarrow f_{X}$ even function (symmetrical w.r.t. 0)
i.e., $f_{X}(t)=f_{X}(-t)$
$f_{Y}(y)=\frac{1}{2 \sqrt{y}}\left[2 f_{X}(\sqrt{y})\right]=\frac{1}{\sqrt{y}} f_{X}(\sqrt{y})$
$=\frac{1}{\sqrt{y}} \frac{1}{\sqrt{2 \pi}} e^{-y / 2} \sim \Gamma(1 / 2,1 / 2) \sim \chi^{2}{ }_{1}$ (chi-squared variable with 1 d.f.)
$Z=a X^{2}=a Y \Rightarrow f_{Z}(z)=\frac{1}{|a|} f_{Y}(z / a) \quad\{a>0 \Rightarrow Z>0$.
$f_{Z}(z)=\frac{1}{a} \frac{1}{\sqrt{z / a}} \frac{1}{\sqrt{2 \pi}} e^{-z /(2 a)}$
$=\frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{2 a}} \frac{1}{\sqrt{2}} e^{-z /(2 a)} \Rightarrow \Gamma(\alpha, \lambda) ; \alpha=1 / 2, \lambda=1 / 2$
$=\lambda^{\alpha} z^{\alpha-1} e^{-\lambda z} / \Gamma(\alpha)$

Blank page

# Applied Statistics for Civil and Environmental Engineers 

Problem Solution Manual
by N.T. Kottegoda and R. Rosso

## Chapter 5 - Model Estimation and Testing

5.1. Piling failures. A contractor involved in driving piles for foundations in a region has a good record of success. Nevertheless, some piles have been unsuccessful. The following failures have been recorded from 50 driven piles in each set.

| Set number, $i$ | Number unsuccessful |
| :---: | :---: |
| 1 | 2 |
| 2 | 3 |
| 3 | 1 |
| 4 | 2 |
| 5 | 4 |
| 6 | 0 |
| 7 | 1 |
| 8 | 3 |
| 9 | 0 |
| 10 | 2 |

Assume the probability of failure is a constant and the trials are independent.
(a) What type of statistical process generates the numbers given in the second column?
(b) What is the distribution of the average failure rate, for various $i$, when based on large sizes of sets?
(c) What is the estimated fraction of failures $p$ from all the sets?
(d) Provide 95 percent confidence limits on the true value of $p$, stating the assumptions made.
(e) Draw a line diagram of the observed and theoretical distributions based on the above table and state whether the data are compatible with it.

Solution. (a) Bernoulli $(0,1)$ process
(b) $2 / 50,3 / 50,1 / 50,2 / 50,4 / 50,0 / 50,1 / 50,3 / 50,0 / 50,2 / 50$,
(c) $18 / 500=0.036$,
(d) $\hat{\sigma}_{p}=\sqrt{\frac{p(1-p)}{n}}=\sqrt{\frac{18}{500}} \times \sqrt{\frac{482}{500}} \times \sqrt{\frac{1}{500}}=0.008331$.

We substitute the sample value of $p$ in this equation because the true value is unknown.

Assuming asymptotic normality in the sampling distribution, the $95 \%$ confidence limits of the true value of $p$ are
$0.036 \pm 1.96 \times 0.008331=0.020,0.052$
(e)The theoretical distribution of the proportions of failures and expected failures in 10 sets:
$p_{X}(0)=\operatorname{Pr}[X=0 ; 50, p]=(1-p)^{50}=0.160,1.60$

$$
\begin{aligned}
& p_{X}(1)=\operatorname{Pr}[X=1 ; 50, p]=50 p(1-p)^{49}=0.299,2.99 \\
& p_{X}(2)=\operatorname{Pr}[X=2 ; 50, p]=\binom{50}{2} p^{2}(1-p)^{48}=0.273,2.73 \\
& p_{X}(3)=\operatorname{Pr}[X=3 ; 50, p]=\binom{50}{3} p^{3}(1-p)^{47}=0.163,1.63 \\
& p_{X}(4)=\operatorname{Pr}[X=4 ; 50, p]=\binom{50}{4} p^{4}(1-p)^{46}=0.072,0.72
\end{aligned}
$$

Sketch of line diagram: frequency (observed and non-integer theoretical) of failures vs. number of failures in 10 sets, each with 50 occurrences.

$\begin{array}{lllll}0 & 1 & 2 & 3 & 4\end{array}$
5.2. Confidence limits for concrete densities. Suppose that only the top 20 of the concrete densities listed in Table E.1.2 are available.
(a) Assuming a normal population, provide 95 percent confidence limits for the mean density of concrete.
(b) Revise the confidence limits for the mean density if the population standard deviation is 16 $\mathrm{kg} / \mathrm{m}^{3}$.

Solution. The data are:

| 2437 | 2437 | 2425 | 2427 | 2428 | 2448 | 2456 | 2436 | 2435 | 2446 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 2441 | 2456 | 2444 | 2447 | 2433 | 2429 | 2435 | 2471 | 2472 | 2445 |

$n=20$; mean $\bar{x}=2442.4 ;$ standard deviation $\hat{s}=13.31$;
significance level $\alpha=1-0.95=0.05$
$t_{n-1, \alpha / 2}=t_{19, \alpha / 2}=2.093$.
(a) Confidence limits for $\mu: \bar{x}-t_{n-1, \alpha / 2} \hat{s} / \sqrt{n} \leq \mu \leq \bar{x}+t_{n-1, \alpha / 2} \hat{s} / \sqrt{n}$,
i.e., $2442.4-6.2 \leq \mu \leq 2442.4+6.2$, i.e., (2436.2, 2448.6).
(b) known variance $=16 \mathrm{mg} / \mathrm{m}^{3}$. Use normal distribution with $z_{\alpha / 2}=1.96$.

Confidence limits for $\mu$ : $\bar{x}-z_{\alpha / 2} s / \sqrt{n} \leq \mu \leq \bar{x}+z_{\alpha / 2} s / \sqrt{n}$;
i.e., $2442.4-7.0 \leq \mu \leq 2442.4+7.0$, i.e., (2435.4, 2449.4).

Although the standard deviation is known, it is higher than in case $(a)$, hence the confidence limits are wider.
5.3. Minimum sample size for estimating mean dissolved oxygen (DO) concentration. Monitoring of pollution levels of similar streams in a region indicates that the standard deviation of DO is 1.95 $\mathrm{mg} / \mathrm{L}$ over a long period of time.
(a) What is the minimum number of observations required to estimate the mean DO within $\pm 0.5$ $\mathrm{mg} / \mathrm{L}$ with 95 percent confidence?
(b) If only 30 observations are taken, what should be the percentage level in the confidence limits for the same difference in means?

## Solution.

(a) $1.96 \times \frac{1.95}{\sqrt{n}}=0.5$. Hence $n=58.43$. Minimum sample size $=59$
(b) $z_{\alpha / 2} \times \frac{1.95}{\sqrt{n}}=0.5 . \quad$ For $n=30, z_{\alpha / 2}=\frac{0.5 \times \sqrt{30}}{1.95}=1.404$.
$\alpha / 2=0.08 ; 1-\alpha=0.84$.
Hence $84 \%$ confidence limits are applicable in this case.
5.4. Yield strength of steel rods. Tests done on a new make of steel rods indicated that, on average, loads up to 1990 kg can be withstood before exceeding the yield strength. This value is based on estimates from 50 specimens chosen at random. The standard deviation of the load is 183 kg . If a more stringent design is based on a 99.9 lower confidence limit, determine the mean yield strength to meet this specification.

## Solution.

$\alpha=0.001 ; z_{\alpha}=3.09$.

$$
z_{\alpha} s / \sqrt{n}=3.09 \times 183 / \sqrt{50} \approx 80 \mathrm{~kg}
$$

The required mean yield strength $=1990-80=1910 \mathrm{~kg}$.
5.5. Confidence intervals on the variance of concrete densities. For the data of Problem 5.2a, provide 95 percent confidence limits on the population variance.

## Solution.

For $\alpha=0.05$ and $n=20$, from Eq. 5.3.14b, $\operatorname{Pr}\left[\frac{(n-1) \hat{S}^{2}}{\chi_{n-1, \alpha / 2}^{2}} \leq \sigma^{2} \leq \frac{(n-1) \hat{S}^{2}}{\chi_{n-1,1-\alpha / 2}^{2}}\right]=1-\alpha$
$\operatorname{Pr}\left[\frac{19 \times 13.31^{2}}{32.9} \leq \sigma^{2} \leq \frac{19 \times 13.31^{2}}{8.91}\right]=0.95$
$\operatorname{Pr}\left[102.3 \leq \sigma^{2} \leq 377.8\right]=0.95$,
i.e., confidence limits for the population variance are 102.3 and 377.8
5.6. Confidence limits on proportions of wet days. A building contractor who works in a relatively dry area is planning to acquire additional work in a newly developing area but is somewhat doubtful of progress because of the adverse effects of rainfall in many months of the year. However, the contractor knows that March is a month of low rainfall with independently distributed daily rainfalls and no apparent relationship between the weather on successive days. Therefore, the thought is that this may be a suitable month to work on the foundations. The proportion of wet days in March is 0.10 from data of the past three years. Suppose it is possible to put off the decision for some time in order to make further observations of daily rainfalls in March. Determine the total number of years of data necessary before one can be 95 percent confident of estimating the true proportion of wet days to within 0.05 .

## Solution.

$\hat{\sigma}_{p}=\sqrt{p(1-p) / n}=\sqrt{0.1 \times 0.9 / n}$.
Confidence limits: $0.10 \pm 1.96 \sqrt{0.1 \times 0.9 / n}$.
$1.96 \sqrt{0.1 \times 0.9 / n}=0.05$. Hence $n=138.3$ days.
March has 31 days, hence number of years required $=138.3 / 31=4.46$, i.e. 5 years.
5.7. Significance of change in temperature. A water supply engineer is concerned that possible climatic change with respect to temperature may have an effect on forecasts for future demands for water to a city. The long-period mean and standard deviation of the annual average temperature measured at mid-day are $33^{\circ} \mathrm{C}$ and $0.75^{\circ} \mathrm{C}$. The alarm is caused by the mean temperature of $34.3^{\circ} \mathrm{C}$ observed for the previous year. Does this suggest that there is an increase in the mean annual temperature at a 5 percent level of significance.

Solution. NH: $\mu_{1}=\mu_{0}=33$; AH: $\mu_{1}>\mu_{0}=33$.
$\mathrm{z}=(34.3-33) / 0.75=1.733$.
$z_{\alpha}=1.645$ for $\alpha=0.05$.
There is a significant increase in temperature. Reject NH.
5.8. Time intervals between passing vehicles. In Example 4.21 the parameter of the fitted exponential distribution was estimated as $1.81 \mathrm{~min}^{-1}$ for the time gaps between vehicles in traffic from 204 observations. By probability plotting methods, this is estimated in Example 5.41 as $1.75 \mathrm{~min}^{-1}$. If these were field estimates over different time periods, do they constitute a significant difference in the mean time intervals, using $\alpha=$. 05 ?

## Solution.

$1 / \hat{\lambda}_{1}=1 / 1.81 \mathrm{~min}^{-1} \quad 1 / \hat{\lambda}_{2}=1 / 1.75 \mathrm{~min}^{-1}$
Mean estimate $=(1 / 1.81+1 / 1.75) / 2=0.562 \mathrm{~min}^{-1}$
Mean $\lambda=1 / 0.562=1.78 \mathrm{~min}^{-1}$
$\mathrm{NH}: \lambda_{1}=\lambda_{2} ; \mathrm{AH}: \lambda_{1} \neq \lambda_{2}$.
For the exponential distribution, $F_{X}(x)=1-e^{-\lambda x}$.
Mean $=1 / \lambda$ and variance $=1 / \lambda^{2}$.
For large sample sizes, the mean rate is distributed approximately as $N\left[1 / \lambda, 1 /\left(n \lambda^{2}\right)\right]$.
Under the NH, $z=\frac{\left(\frac{1}{\hat{\lambda}_{1}}-\frac{1}{\hat{\lambda}_{2}}\right)-0}{1 /(\hat{\lambda} \sqrt{n})}=\frac{\left(\frac{1}{1.81}-\frac{1}{1.75}\right)}{1 /(1.78 \sqrt{204})}=-0.48$
Because $z_{\alpha / 2}=1.96$, for $\alpha=.05$, the difference is not significant; do not reject NH
5.9. Comparing outputs of waste water plants. Two treatment plants are built in an area to treat waste-water from a city. Their relative performances are compared from the results of BOD tests made on the outputs. Eight preliminary results are listed below as differences in BOD between plant 1 and 2.

| Test | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Difference +1.2 +0.2 | -1.6 | +0.7 | +1.3 | -0.9 | -0.1 | -1.9 |  |  |
| in BOD, |  |  |  |  |  |  |  |  |
| $\mathrm{mg} / \mathrm{L}$ |  |  |  |  |  |  |  |  |

Test the difference in the outputs at the 5 percent level of significance.

## Solution.

| Test | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $x=$ Diff <br> in BOD | +1.2 | +0.2 | -1.6 | +0.7 | +1.3 | -0.9 | -0.1 | -1.9 |
| $x^{2}$ | 1.44 | 0.04 | 2.56 | 0.49 | 1.69 | 0.81 | 0.01 | 3.61 |

From the above table the mean of $x=-1.1 / 8$; mean of $x^{2}=10.65 / 8$.
$\hat{S}^{2}=\sqrt{\frac{n}{n-1}}\left(\frac{\Sigma x^{2}}{n}-\bar{x}^{2}\right)^{1 / 2}=\left\{\frac{8}{7}\left[\frac{10.65}{8}-\left(\frac{-1.1}{8}\right)^{2}\right]\right\}^{1 / 2}=1.225$.
NH: $\mu_{1}=\mu_{2}$; AH: $\mu_{1} \neq \mu_{2}$
Test statistic $\frac{\left(\bar{x}_{1}-\bar{x}_{2}\right)-\left(\mu_{1}-\mu_{2}\right)}{\hat{S} / \sqrt{n}}=\frac{-1.1 / 8}{1.225 / \sqrt{8}}=-0.317$.
Degrees of freedom $v=n-1=7$
$t_{7,0.025}=2.365$.
The difference is not significant; do not reject NH.
5.10. Change in the mean and variance of flood flows. Annual maximum flows of the Tevere (Tiber) River recorded at Ripetta in Rome are given in Table E.5.8 for the period 1921 to 1974. The observation of numerous low maximum flows during the last 20 years led to a suspicion that the flow regime or climatic conditions had changed. Divide the record into two halves. Determine if the mean annual maximum flow in the second half is lower than those in the first half at a level of significance $\alpha$ $=.01$ under the following conditions:
(a) If the standard deviation is $450 \mathrm{~m}^{3} / \mathrm{sec}$.
(b) If the standard deviation is estimated from the data but is assumed to be constant.
(c) If the standard deviations are estimated separately for the two halves and are assumed to be different.
Using the estimated variances in part $c$, above, determine whether the change in the variance is significant for $\alpha=.01$ ?

## Solution.

Flows in the first half are
$\begin{array}{llllllllllllll}1092 & 1099 & 1440 & 1083 & 1621 & 1132 & 935 & 1540 & 1966 & 775 & 1166 & 843 & 1508 & 1876\end{array}$
$\begin{array}{lllllllllllllllll}1696 & 1690 & 2730 & 1440 & 985 & 1346 & 1553 & 1370 & 743 & 1340 & 896 & 1600 & 2189\end{array}$
Flows in the second half are

| 1600 | 714 | 794 | 1460 | 1240 | 1230 | 1270 | 861 | 1355 | 612 | 822 | 1370 | 1380 | 510 |
| ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
| 810 | 735 | 259 | 1290 | 1325 | 528 | 622 | 355 | 468 | 472 | 664 | 717 | 950 |  |

$\bar{x}_{1}=1395 ; \quad \bar{x}_{2}=904 ; \quad \hat{s}_{1}=457.9 ; \quad \hat{s}_{2}=385.5$.
(a) For this one-tailed test, $z=(1395-904) /\left(\frac{450^{2}}{27}+\frac{450^{2}}{27}\right)^{1 / 2}=4.01$ $z_{\alpha}=2.325$ for $\alpha=0.01$. Highly significant. Reject NH
(b) From Eq. 5.4.9, $t=\frac{1395-904}{\left(26 \times 457.9^{2}-26 \times 385.5^{2}\right.} \sqrt{\frac{(54-2) \times 27 \times 27}{54}}=10.32$

Degrees of freedom $v=27+27-2=52 ; t_{52,0.01} t_{52,0.01}=2.42 .4$
The test result is highly significant, reject NH.
(c) Under the NH and using a one-tailed test,
$t=\frac{\bar{x}_{1}-\bar{x}_{2}}{\sqrt{\frac{\hat{s}_{1}^{2}}{n_{1}}+\frac{\hat{s}_{2}^{2}}{n_{2}}}}=\frac{4.91}{\sqrt{\frac{457.9^{2}}{27_{1}}+\frac{385.5^{2}}{27_{2}}}}=4.26$
Degrees of freedom are: $v$
$\left.\frac{\left[\hat{s}_{1}{ }^{2} / n_{1}+\hat{s}_{2}{ }^{2} / n_{2}\right]^{2}}{\left.\left[\hat{s}_{1}{ }^{2} / n_{1}\right]^{2} /\left(n_{1}-1\right)+\hat{s}_{2}{ }^{2} / n_{2}\right]^{2} /\left(n_{2}-1\right)}=\frac{13269.7^{2}}{\left[\left(457.9^{2} / 27\right)^{2} / 26+\left(385.5^{2} / 27\right)^{2} / 26\right.}\right]=50.5$
$t_{50.5,0.01} \approx 2.4$ The test result is highly significant, reject NH .
(d) $F=\left(\frac{457.9}{385.5}\right)^{2}=1.41 ; \quad F_{26,26,0.05} \approx 1.96$

The test result is not significant, do not reject NH.
5.11. Control chart for quality control of concrete. Control charts were introduced in 1924 by Walter A. Shewhart [see Shewhart, W. A. (1931), Economic Control of Quality of Manufactured Products, D.Van Nostrand, New York] in order to detect and control any unwanted deviations in a process so that quality can be maintained.

Suppose tests based on compressive strengths have been made on concrete cubes to determine the ultimate loads that can be carried by concrete being used at a construction site. From past data the mean and standard deviation are estimated as 61.1 and $4.9 \mathrm{~N} / \mathrm{mm}^{2}$, respectively, and each day 5 test cubes are tested at random and the mean is computed. The following results are obtained from the tests of 12 working days:

| Batch <br> number, $i$ | Mean compressive <br> strength, $\mathrm{N} / \mathrm{mm}^{2}$ |
| :--- | :--- |
| 1 | 58.1 |
| 2 | 60.9 |
| 3 | 62.5 |
| 4 | 59.9 |
| 5 | 56.1 |
| 6 | 58.7 |
| 7 | 61.5 |
| 8 | 61.9 |
| 9 | 63.5 |
| 10 | 58.1 |
| 11 | 67.1 |
| 12 | 60.1 |

Draw control charts using bands that are two standard errors from the mean. (Three standard errors are commonly used.)
(a) Do any of the above results suggest that corrective action is necessary?
(b) What is the probability that a Type I error is made, that is, action as in $(a)$ is taken without any need for it?
(c) What is the probability of making one or more of such errors during a six-day working week?
(d) What is the probability of making a Type II error, if the use of aggregates of lower quality have reduced the mean strength to $57.5 \mathrm{~N} / \mathrm{mm}^{2}$ ?
(e) How does one reduce the foregoing errors?

## Solution.

Sketch of control bands


The two control bands are set at distances $2 \times 4.9 / \sqrt{5}=4.38 \mathrm{~N} / \mathrm{mm}^{2}$ above and below the mean of $61.1 \mathrm{~N} / \mathrm{mm}^{2}$
(a) 5 and 11
(b) $\alpha=2[1-\Phi(2)]=2(1-0.97725)=0.0455$
(c) $p_{X}(x)=\operatorname{Pr}[X=x ; n, p]=\binom{n}{x} p^{x}(1-p)^{n-x}$
$\operatorname{Pr}[X=0 ; 30,0.0455]=(1-0.0455)^{30}=0.247$
$\operatorname{Pr}[X>0 ; 30,0.0455]=(1-0.0455)^{30}=1-0.247=0.752$
(d) $\beta=\Phi\left(\frac{65.5-57.5}{2.19}\right)-\Phi\left(\frac{56.7-57.5}{2.19}\right)=1-(1-0.642)=0.642$
(e) To reduce the Type I error, increase distance from mean to control band from 2 to 3 standard errors. To reduce the Type II error, increase $n$
5.12. Power curve for concrete strengths. In Example 5.9, 95 percent confidence limits of 58.54 and $61.76 \mathrm{~N} / \mathrm{mm}^{2}$ were provided for the 40 concrete strengths with mean and standard deviation 60.14 and $5.02 \mathrm{~N} / \mathrm{mm}^{2}$ listed in Table 1.2.2. Determine the Type II errors made if the population values are equal to each of the following values, all in newtons per square millimeter:
60.5
61.5
62.0
62.5
63.5.

Draw the power curve for the corresponding points.

## Solution.

Population mean $\quad \beta \quad$ Power $=1-\beta$
(a) 60.5

$$
\begin{align*}
& \beta=\Phi\left(\frac{61.76-60.5}{0.7937}\right)-\Phi\left(\frac{58.54-60.5}{0.7937}\right) \\
= & \Phi(1.5875)-\Phi(-2.47)=0.944-0.007=0.937
\end{align*}
$$

(b) 61.5

$$
\beta=\Phi\left(\frac{61.76-61.5}{0.7937}\right)-\Phi\left(\frac{58.54-61.5}{0.7937}\right)
$$

$$
=\Phi(0.3275)-\Phi(-3.729)=0.628
$$

(c) 62.0

$$
\beta=\Phi\left(\frac{61.76-62.0}{0.7937}\right)-\Phi\left(\frac{58.54-62.0}{0.7937}\right)
$$

$$
=\Phi(-0.3024)-\Phi(-4.359)=0.382
$$

(d) 62.5

$$
\beta=\Phi\left(\frac{61.76-62.5}{0.7937}\right)-\Phi\left(\frac{58.54-62.5}{0.7937}\right)
$$

$$
=\Phi(0-0.9323)-\Phi(-4.9892)=0.175
$$

$$
0.825
$$

(e) 63.5

$$
\begin{aligned}
& \beta=\Phi\left(\frac{61.76-63.5}{0.7937}\right)-\Phi\left(\frac{58.54-63.5}{0.7937}\right) \\
= & \Phi(-2.1923)-\Phi(-6,2492)=0.0 .014
\end{aligned}
$$

$$
0.986
$$

Sketch of power curve: Power vs. Mean
1.0 $\qquad$ .986 $\qquad$
0.9
0.8 .825
0.7
-
.825
0.6 .618
0.5
-
0.4 .372
0.3 -
0.2
0.1

0.0 $\qquad$ .063 | 61 | 62 | 63 | $64 \mathrm{~N} / \mathrm{mm}^{2}$ |
| :--- | :--- | :--- | :--- |

5.13. Irrigation and rain. Irrigation usually commences on 15th April in the Po River basin, Italy. An engineer is interested in the probability of rain during the seven days from April 15 to 21. From rainfall data of the past 100 years in a particular area, the following distribution of rainy days is obtained for the period.

| Rainy days | 0 | 1 | 2 | 3 | 4 | $5,6,7$ | Total |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Frequency | 57 | 30 | 9 | 3 | 1 | 0 | 100 |

The binomial model $B(M=m \mid 7,0.1)$ is postulated. Can this be justified at the 5 percent level of significance on the basis of a chi-squared test?

| Solution. | Rainy days, | $i=$ | 0 | 1 | 2 | 3 | 4 | 5,6,7 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Observed num |  | 57 | 30 | 9 | 3 | 1 | 0 |
|  | Expected nu | $E_{i}=$ | 47.8 | 37.2 | 12.4 | 2.3 | 0.3 | 0 |

$$
\begin{aligned}
& E_{i}=100\binom{n}{i} p^{i}(1-p)^{n-i} \text { for } n=7, p=0.1 . \\
& X^{2}=\sum_{i=0}^{7}\left(O_{i}-E_{i}\right)^{2} / E_{i}=9.2^{2} / 47.8+7.2^{2} / 37.2+3.4^{2} / 12.4+0.7^{2} / 2.3+0.7^{2} / 0.3 \\
& =1.77+1.39+0.93+0.21+1.63=5.93 . \\
& \chi^{2}{ }_{3,0.05}=7.81 . \text { Do not reject the NH of binomial model } B(M=m \mid 7,0.1) .
\end{aligned}
$$

5.14. One-sample sign test on flows. The following is a sample from the recorded annual flows in the St Lawrence River which runs out from the Great Lakes of North America. The data are in standardized units obtained by dividing the original observations by the annual mean. Test the null hypothesis that the median is 1.006 against the alternative hypothesis that it is greater or less than this value, at the 5 percent level of significance.

$$
\begin{array}{llllllllll}
0.942 & 0.947 & 1.005 & 0.988 & 1.001 & 1.013 & 1.013 & 1.088 & 1.000 & 0.959 .
\end{array}
$$

## Solution.

For the stated median of 1.006 the signs are as follows:
$n=10, k=3$, i.e., $k<n / 2, \quad z=\frac{(k+1 / 2)-n / 2}{\sqrt{n} / 2}=\frac{3.5-5}{\sqrt{10} / 2}=-0.95$
Because $z_{\alpha / 2}=1.96$, for $\alpha=.05$, the difference is not significant; do not reject NH
5.15. Sign test applied to paired observations. We reexamine the concrete densities listed in Table E.1.2. Divide the record into two samples of equal length and apply the sign test to correspondingly paired observations from the two halves. Test the hypothesis that the mean density is unchanged at the 5 percent level of significance.

Solution. The number of non-zero differences $n=20-1=19$.
The number of positive differences $k=9$.
$z=\frac{(k+1 / 2)-n / 2}{\sqrt{n} / 2}=\frac{0}{\sqrt{19} / 2}=0$
Because $z_{\alpha / 2}=1.96$, for $\alpha=.05$, the difference is not significant; do not reject NH
5.16. Wilcoxon signed-rank test on flows. Use the Wilcoxon signed-rank test to ascertain whether the mean of the annual maximum flows of the Tevere River has changed from the first half to the second half of the period given in Table E.5.8. Test the null hypothesis that the means are the same against the alternative hypothesis that the mean flow is less in the second half at the 1 percent level of significance.

Solution. $\mathrm{NH} \quad H_{0}: \mu_{A}=\mu_{B}$

$$
\mathrm{AH} \quad H_{1}: \mu_{A}>\mu_{B}
$$

For $n=27$, differences $\mu_{A}-\mu_{B}$ and ranks commencing with lowest absolute value:

| -508 | 385 | 646 | -377 | 381 | -98 | -335 | 679 | 611 | 163 | 344 | -527 | 128 | 1366 |
| ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
| 13 | 12 | 16 | 10 | 11 | 1 | 7 | 17 | 15 | 4 | 9 | 14 | 2 | 26 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 886 | 955 | 2471 | 150 | -340 | 818 | 931 | 1015 | 275 | 868 | 232 | 883 | 1240 |  |
| 21 | 23 | 27 | 3 | 8 | 18 | 22 | 24 | 6 | 19 | 5 | 20 | 25 |  |

The total sum of ranks $=n(n+1) / 2=27 \times 28 / 2=378$.
Sum of ranks of negative differences $=T^{-}=53$.
Sum of ranks of positive differences $=T^{+} \quad=325$.
$\mu_{T}=n(n+1) / 4=189 \quad \sigma_{T}{ }^{2}=n(n+1)(2 n+1) / 24=27 \times 28 \times 55 / 24=1732.5$
$z=\left(T^{+}-\mu_{T}\right) / \sigma_{T}=(325-189) / \sqrt{1732.5}=3.27$
Because $z_{\alpha}=2.325$, for $\alpha=.01$, the test result is highly significant; reject NH.
5.17. Runs test on the Wolfer sunspot numbers. Wolfer sunspot numbers are an index of activity on the solar surface. They have been investigated for their impact on terrestrial climate and for the resulting environmental effects. Twenty annual observations are listed below for the period 1770-1789.

| 101 | 82 | 66 | 35 | 31 | 7 | 20 | 92 | 154 | 125 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 85 | 68 | 38 | 23 | 10 | 24 | 83 | 132 | 131 | 118 |

Apply a runs test for randomness. Do these represent a random series at the 5 percent level of significance?

Solution. We underline the runs above the median, 70, as follows:
$\begin{array}{llllllllllllllll}101 & 82 & 66 & 35 & 31 & 7 & 20 & \underline{92} & 154 & 125 & 85 & 68 & 38 & 23 & 10 & 24 \\ 83 & 132 & 131 & 118\end{array}$ The total number of runs $=5$.
$n=10 ; m=10 ; \quad \hat{\mu}_{R}=1+\frac{2 n m}{n+m}=1+\frac{2 \times 10 \times 10}{10+10}=11$
$\operatorname{Var}[R]=\frac{2 n m(2 n m-n-m)}{(n+m)^{2}(n+m-1)}=\frac{2 \times 10 \times 10(200-20)}{20^{2} \times 19}=\frac{180}{38}=4.737$.
$z=\frac{\left[(r-1 / 2)-\hat{\mu}_{R}\right]}{\sqrt{\operatorname{Var} R}}=\frac{(4.5-11)}{\sqrt{4.737}}=2.99$
Because $z_{\alpha / 2}=1.96$, for $\alpha=.05$, the test result is highly significant; reject NH that the sunspots represent a random series.
5.18. Spearmen rank correlation test on the DO-BOD relationship. With reference to the data in Table E.1.3 determine the rank correlation coefficient for the relationship between DO and BOD. Compare with the result in Example 1.30.

Solution. The following table gives a summary of the calculations.

| Item <br> number | DO, <br> $x$ | BOD, <br> $y$ | DO <br> ranked | BOD <br> ranked | Rank <br> of $x$ | Rank <br> of $y$ | Diff in <br> ranks, <br> $d$ | $d^{2}$ | Cumulative <br> Sum of $d^{2}$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 5 | 3.4 | 6.2 | 6.2 | 2.7 | 35 | 12 | 23 | 529 | 4272.3 |
| 10 | 3.9 | 6.7 | 6.7 | 2.9 | 30 | 5 | 25 | 625 | 7409.6 |
| 15 | 3.2 | 7.2 | 7.2 | 3.0 | 19 | 21 | -2 | 4 | 8827.6 |
| 20 | 3.2 | 7.6 | 7.6 | 3.2 | 22.5 | 17 | -1 | 1 | 9330.9 |
| 25 | 3.0 | 8.0 | 8.0 | 3.3 | 15 | 24 | -9 | 81 | 9507.5 |
| 30 | 2.9 | 8.4 | 8.4 | 3.7 | 9 | 29 | -20 | 400 | 10777.5 |
| 35 | 2.7 | 8.8 | 8.8 | 4.0 | 4 | 34 | -30 | 900 | 14197.5 |
| 38 | 2.5 | 9.4 | 9.4 | 4.4 | 1 | 37 | -36 | 1296 | $\mathbf{1 7 6 7 3 . 5}$ |

The sample rank correlation coefficient is
$r=1-\frac{6 \sum_{i=1}^{n} d_{i}{ }^{2}}{n\left(n^{2}-1\right)}=1-\frac{6 \times 17673.5}{38\left(38^{2}-1\right)}=-0.934$.
In Example 1.30, the sample product-moment correlation coefficient $=-0.9$.
5.19. Poisson distribution of numbers of days of high waves. High waves in a coastal area where further development is planned cause property damage and erosional problems but measurements of wave heights are scanty. A researcher has obtained the following information of the number of days of high waves in a year from local chronicles and residents.

| Number of days of high waves | 0 | 1 | 2 | 3 | 4 | 5 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Frequency | 26 | 13 | 6 | 3 | 2 | 0 |

Sketch a histogram of the number of days of high waves recorded in the area during a 50 -year period.
Test the hypothesis that the occurrence of high waves is Poisson distributed at the 5 percent level of significance using the chi-squared test. What is the probability that the mean rate will be more than one day per year?

Solution. Number of days
$\begin{array}{lllllllll}\text { of high wind, } i, & = & 0 & 1 & 2 & 3 & 4 & 5\end{array}$

| Observed number, $O_{i}=$ | 26 | 13 | 6 | 3 | 2 | 0 |
| ---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $i O_{i}=$ | 0 | 13 | 12 | 9 | 8 | 0 |
| Expected number, $E_{i}=$ | 21.5 | 18.1 | 7.6 | 2.1 | 0.5 | 0.1 |

The sum of the $i O_{i}=42$. Therefore, the parameter $\hat{v}=42 / 50=0.84$. The Poisson distributed expected numbers, $E_{i}$, are calculated as $n \times \operatorname{Pr}[I=i ; v]=n \frac{\nu^{i} e^{-v}}{i!}$
$X^{2}=\sum_{i=0}^{7}\left(O_{i}-E_{i}\right)^{2} / E_{i}=4.5^{2} / 21.5+5.1^{2} / 18.1+1.6^{2} / 7.6+0.9^{2} / 2.1+1.5^{2} / 0.5+$
$0.2^{2} / 0.1$

$$
=0.94+1.44+0.34+0.39+4.50+0.2=7.81 .
$$

$\chi^{2}{ }_{5,0.05}=11.1$. Do not reject the NH of the Poisson model.
5.20. All-red phase of traffic lights. At 12 four-way junctions in London, England, brief "all red" phases were introduced. The numbers of accidents causing injuries were recorded for 2 years before and after the installation as given here:

| Site | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Before | 27 | 4 | 18 | 20 | 17 | 12 | 18 | 24 | 18 | 19 | 3 | 8 |
| After | 20 | 9 | 14 | 14 | 16 | 3 | 13 | 4 | 9 | 11 | 3 | 6 |

(With the kind courtesy of the Transport and Road Research Laboratory, England.)
Test the reduction in the number of accidents at the 1 percent level of significance. It is thought that sites with high rates of accidents are highly weighted. At a given site the variance is expected to be proportional to the mean over consecutive time periods. Taking the square roots of the numbers will reduce the differences in variances. Repeat the test at the 5 percent and 1 percent levels of significance for the variance-adjusted data.

Solution. Denote After data as A and the Before data as B. Then

| Site | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | Sum |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| A-B | -7 | 5 | -4 | -6 | -1 | -9 | -5 | -20 | -9 | -8 | 0 | -2 | -66 |
| $(\mathrm{~A}-\mathrm{B})^{2}=49$ | 25 | 16 | 36 | 1 | 81 | 25 | 400 | 81 | 64 | 0 | 4 | 782 |  |

Mean $\bar{x}=-66 / 12=-5.5$; standard deviation $\hat{s}=\sqrt{\left(782 / 12-5.5^{2}\right) \times 12 / 11}=6.17$.
The null hypothesis is $\quad H_{0}: \mathrm{A}-\mathrm{B}=0$, i.e., the "all red" phase has no effect.
The alternate hypothesis is $H_{1}: \mathrm{A}-\mathrm{B}<0$.
$T=\left(\bar{X}_{A-B}-\mu_{A-B}\right) / S_{A-B} \sqrt{n}$.
Under NH $t=\bar{x} /(\hat{s} / \sqrt{n})=-5.5 \times \sqrt{12} / 6.17=-3.09$.
For the one-tailed test with d.f. $=11$ and $\alpha=0.01, t_{11,0.01}=2.718$.
Because the test result is highly significant reject the null hypothesis that the "all red" phase has no effect.
For the variance reduction techniques, we take the square roots of A and B .

| Site | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | Sum |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | :--- |
| A | $=4.47$ | 3 | 3.74 | 3.74 | 4 | 1.73 | 3.61 | 2 | 3 | 3.32 | 1.73 | 2.45 |  |
| B | $=5.20$ | 2 | 4.24 | 4.47 | 4.12 | 3.46 | 4.24 | 4.90 | 4.24 | 4.36 | 1.73 | 2.83 |  |
| A-B | $=-0.73$ | -0.50 | -0.73 | -0.12 | -1.73 | -0.63 | -2.90 | -1.24 | -1.04 | 0 | -0.39 | -9.01 |  |
| $(\mathrm{~A}-\mathrm{B})^{2}=0.53$ | 1 | 0.25 | 0.53 | 0.01 | 2.99 | 0.40 | 8.41 | 1.54 | 1.08 | 0 | 0.15 | 16.89 |  |
| Mean | $\bar{x}$ | $=-9.01 / 12$ |  | $=$ | $-0.75 ;$ |  | standard | deviation |  |  |  |  |  |
| $\hat{s}=\sqrt{\left(16.89 / 12-0.75^{2}\right) \times 12 / 11}=0.96=6.17$. |  |  |  |  |  |  |  |  |  |  |  |  |  |

The null hypothesis is $\quad H_{0}: \mathrm{A}-\mathrm{B}=0$, i.e., the "all red" phase has no effect.
The alternate hypothesis is $H_{1}: \mathrm{A}-\mathrm{B}<0$.
$T=\left(\bar{X}_{A-B}-\mu_{A-B}\right) / S_{A-B} \sqrt{n}$.
Under NH $t=\bar{x} /(\hat{s} / \sqrt{n})=-.75 \times \sqrt{12} / 0.96=-2.706$.

For the one-tailed test with d.f. $=11$ and for $\alpha=0.05, t_{11,0.05}=1.796$; also for $\alpha=$ $0.01, t_{11,0.01}=2.718$. The test result is significant at the $\alpha=0.05$ level, but it is not significant at the $\alpha=0.01$ level, just marginally. Reject NH .
5.21. Speed limit: USA. Speeds of cars were estimated on rural interstate roads in the United States during 1973 and 1975. The numbers of cars within certain categories of speeds are listed here.

|  | Less than $\mathbf{4 5} \mathbf{~ m p h}$ | $\mathbf{4 5}$ to $\mathbf{5 5} \mathbf{~ m p h}$ | $\mathbf{5 5}$ to $\mathbf{7 0} \mathbf{~ m p h}$ | $\mathbf{7 0}$ to85 $\mathbf{~ m p h}$ | Total |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Upper limits kph | 72.5 | 88.5 | 112.6 | 137 |  |
| 1973 | 0 | 7 | 63 | 30 | 100 |
| 1975 | 1 | 28 | 69 | 2 | 100 |

Determine whether there is a significant decrease, at the 1 percent level, of the proportion of cars exceeding the speed limit of $55 \mathrm{mph}(88.5 \mathrm{kph})$ between the two years. Data from Transportation Research, Vol. 17, D.B. Kamerud, "The 55 mph speed limit: Costs benefits, and implied trade-offs", pp. 51-64, Copyright (1983) with the kind permission from Elsevier Science Ltd., The Boulevard, Langford Lane, Kidlington, OX5 1GB, England.

## Solution.

Proportion of cars exceeding the speed limit in 1973, $\hat{p}_{73}=93 / 100$ and in 1975, $\hat{p}_{75}=71 / 100$. The average proportion $\hat{p}=(93+71) / 200=0.82$

The null hypothesis is $\quad H_{0}: p_{75}=p_{73}=p$, i.e., there is no decrease in the proportions
The alternate hypothesis is $H_{1}: p_{75}<p_{73}$.
Difference in proportions $\hat{p}_{75}-\hat{p}_{73}=-0.22$
Variance of difference between observed proportions,
$\operatorname{Var}\left[\hat{\mathrm{p}}_{75}-\hat{\mathrm{p}}_{73}\right]=\operatorname{Var}\left[\hat{\mathrm{p}}_{75}\right]+\operatorname{Var}\left[\hat{\mathrm{p}}_{73}\right]=\frac{p_{75}\left(1-p_{75}\right)}{n_{75}}+\frac{p_{73}\left(1-p_{73}\right)}{n_{73}}=\frac{2 \times 0.22 \times 0.78}{100}=0.003432$
because $n_{75}=n_{75}=100$ and also, under the NH, $p_{75}=p_{73}=p=0.22$.
Under the NH, $\left[\hat{p}_{75}-\hat{p}_{73}\right] \sim N(0,0.003432)$ approximately.
Normal score $z=\frac{-0.22-0}{\sqrt{0.003432}}=-3.76$.
For the one-tailed test with $\alpha=0.01, z_{0.01}=2.326$.
Because the test result is highly significant reject the null hypothesis that there is no decrease in the proportions
5.22 Speed limit: England. To test the effect of a 65 kilometers per hour ( 40 miles per hour) speed limit on the A 4123 road in England, speeds of vehicles were calculated from observations taken at sites during one day before and one day after the introduction of the limit. The following results were obtained:

| Day and Site | Mean speed in kph. of private cars. |  |  |
| :---: | :--- | :---: | :---: |
|  |  | Before | After |
| Monday | Northbound | $68.3(42.4)$ | $63.4(39.4)$ |
| 1 | Southbound | $61.4(38.1)$ | $58.9(36.6)$ |
| Tuesday | Northbound | $72.8(45.2)$ | $64.1(39.8)$ |
| 2 | Southbound | $69.9(43.4)$ | $64.6(40.1)$ |
| Wednesday | Northbound | $61.4(38.1)$ | $56.8(35.3)$ |
| 3 | Southbound | $59.1(36.7)$ | $55.1(34.2)$ |

(With the kind courtesy of the Transport and Road Research Laboratory, England.)
Note: Values in parentheses are in mph . as originally calculated.
In considering that there may be other factors, such as weather, that could have caused the differences, observations were also made on the same days over a similar part of the road where no speed limit was imposed. The following changes in mean speeds, [Before-After] were recorded in the same units:

| Monday | -2.42 | $(-1.5)$ |
| :--- | :--- | :--- |
| Tuesday | -0.16 | $(-0.1)$ |
| Wednesday | -1.29 | $(-0.8)$ |

Test the change at the 5 percent level of significance.

Solution. As in Problem 5.20 we calculate the differences [After - Before] and then add the changes caused by other factors with the sign reversed

## Site 1

Site 2
$-8.7+0.2=-8.5$ Site 3

N -Bound $-4.9+2.4=-2.5$
3.3

S-Bound $-2.5+2.4=-0.1$
$-5.3+0.2=-5.1$
$-4.0+1.3=-$
2.7

Mean $\bar{x}=-22.2 / 6=-3.77$; standard deviation $\hat{s}=\sqrt{\left(122.6 / 6-3.77^{2}\right) \times 6 / 5}=2.75$.
Under NH $t=-(3.77 / 2.75) \sqrt{6}=-3.36$
For the one-tailed test with d.f. $=5$ and for $\alpha=0.05, t_{5,0.05}=-2.015$; Reject NH. That there is no change in speeds after the introduction of the speed limits.
5.23 Machine failures. The following are intervals in hours between failures of the air conditioning system of a Boeing 720 jet airplane:

$$
\begin{aligned}
& 23,261,87,7,120,14,62,47,225,71,246,21,42,20,5,12,120,11,3,14,71,11, \\
& 14,11,16,90,1,16,52,95
\end{aligned}
$$

Test whether the data are exponentially distributed at the 5 percent level of significance. Draw a probability plot.

Data with the kind courtesy of the publishers from F. Proshan (1963), "Theoretical explanation of observed decreasing failure rate", Technometrics, Vol. 5, pp. 375-383.

## Solution.

Ranked failures of the air conditioning system of a Boeing 720 jet airplane:

| 1 | 3 | 5 | 7 | 11 | 11 | 11 | 12 | 14 | 14 | 14 | 16 | 16 | 20 | 21 | 23 |  | 42 | 47 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 52 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 62 | 71 | 71 |  |  | 87 | 90 | 95 |  |  | 120 | 120 |  | 225 | 246 | 261 |  |  |  |

$n=30 ; \sum x=1788 ; \bar{x}=1788 / 30=59.6$.
Exponential model: $F(x)=1-e^{-30 x / 1788}$.

| $\quad$$x \leq 20$ <br> $x>120$ <br> $x>$ | $20<x \leq 40$ | $40<x \leq 60$ | $60<x \leq 80$ | $80<x \leq 100$ | $100<x \leq 120$ |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: |
| $O_{i}$ | 14 | 2 | 3 | 3 | 3 | 2 |
| 3 |  |  |  |  |  |  |

limiting

| $F(x)$ | 0.285 | 0.489 | 0.635 | 0.739 | 0.813 | 0.866 |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: |
| $1-F(x) .715$ | .511 | .365 | .261 | .187 | .134 |  |
| $\operatorname{Ln}[1-F(x)]$ |  |  |  |  |  |  |
| -0.335 | -.671 | -1.008 | -1.343 | -1.677 | -2.010 |  |
| $n F(x) 8.552$ | 14.666 | 19.037 | 22.163 | 24.397 | 25.994 |  |
| $E_{i}=8.552$ | 6.114 | 4.371 | 3.126 | 2.234 | 1.597 | 4.006 |
| $\frac{\left(O_{i}-E_{i}\right)^{2}}{E_{i}}$ |  |  |  |  |  |  |
| $=3.471$ | 2.768 | 0.430 | 0.005 | 0.262 | 0.102 | 0.252 |

$\sum \frac{\left(O_{i}-E_{i}\right)^{2}}{E_{i}}=7.29$. The degrees of freedom $v=7-1-1=5 . \quad \chi^{2}{ }_{5,0.05}=11.1$. Do not reject the NH of the exponential model. See sketch of probability plot below.

5.24 Bacterials counts. The following are counts of the number of fields of bacteria reported by Bliss, C. and R.A. Fisher (1953): "Fitting the negative Binomial Distribution to Biological Data", Biometrics, Vol. 9, pp. 176-196.

| Bacteria for field | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 and more |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Number of fields | 11 | 17 | 31 | 24 | 29 | 18 | 19 | 16 | 13 | 17 | 6 | 8 | 31 |

As a preliminary step fit the geometric distribution to these data. Apply a chi-squared goodness-of-fit test at a level of significance $\alpha=.05$, combining the counts for fields 0 and 1 . (Data used with the kind courtesy of the International Biometric Society, 808 17th Street NW, Suite 200, Washington D.C., 20006-3910 USA.).

## Solution.

| $i=$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | $\geq 12$ | Sum |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
| $O_{i}=11$ | 17 | 31 | 24 | 29 | 18 | 19 | 16 | 13 | 17 | 6 | 8 | 31 | 240 |
| $i O_{i}=0$ | 17 | 62 | 72 | 116 | 90 | 114 | 112 | 104 | 153 | 60 | 88 | 434 | 1422 |
| $E_{i}=$ | 40.5 | 33.7 | 28.0 | 23.3 | 19.3 | 16.1 | 13.4 | 11.1 | 9.2 | 7.7 | 6.4 | 31.3 |  |

(For the column $\geq 12$ we take an average count of 14.) $n=240 \bar{x}=1422 / 240=$ 5.925.
$p=1 / \bar{x}=0.1668$.
$E_{i}=n p(1-p)^{i-1}$
$X^{2}=\sum_{i=0}^{7}\left(O_{i}-E_{i}\right)^{2} / E_{i}=$
$=(28-40.5)^{2} / 40.5+(31-33.7)^{2} / 33.7+(24-28.0)^{2} / 28.0+(29-23.3)^{2} / 23.3$
${ }_{i}+(18-19.3)^{2} / 19.3+(19-16.1)^{2} / 16.1+(16-13.4)^{2} / 13.4+(13-11.1)^{2} / 11.1$
$+(17-9.2)^{2} / 9.2+(6-7.7)^{2} / 7.7+(8-6.4)^{2} / 6.4+(31-31.3)^{2} / 31.3$
$=3.86+0.22+0.57+1.39+0.09+0.52+0.50+0.33+6.61+0.38+0.40+0.00$
$=14.87$.
The degrees of freedom $v=12-1-1=10 . \chi^{2}{ }_{10,0.05}=18.3$. Do not reject the NH of the geometric model
5.25. Pump failures. Two manufacturers, A and B, supply pumps to the same specification of 500 hours on average to failure. Twenty pumps of each manufacturer have been installed and the times to failure for each pump are as follows:

| A | 510 | 450 | 478 | 512 | 506 | 485 | 501 | 481 | 452 | 494 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
|  | 514 | 507 | 487 | 467 | 502 | 508 | 503 | 492 | 502 | 499 |
| B | 510 | 513 | 497 | 506 | 493 | 501 | 547 | 514 | 487 | 490 |
|  | 495 | 497 | 508 | 493 | 522 | 502 | 527 | 486 | 531 | 497 |

(a) Test whether the mean time of failure for A is less than that for B .
(b) Test whether the proportion of pumps not reaching specification is less for B than for A .

Use $\alpha=.01$ and an appropriate test in each case.

Solution. (a)
$\mathrm{NH} \mu_{A}=\mu_{B}$
AH $\mu_{A}<\mu_{B}$
Sample sizes are $n_{A}=n_{B}=20$. The sample means are $\bar{x}_{A}=492.5$ and $\bar{x}_{B}=505.8$
The sample standard deviations are
$\hat{s}_{A}=\left[\left(242892-492.5^{2}\right) \times \frac{20}{19}\right]^{1 / 2}=18.80 ; \quad \hat{s}_{B}=\left[\left(256078-505.8^{2}\right) \times \frac{20}{19}\right]^{1 / 2}=16.05$.
The sample $t$ value $i s$
$t=\frac{\left(\bar{x}_{A}-\bar{x}_{B}\right)-\left(\mu_{A}-\mu_{B}\right)}{\left[\left(n_{A}-1\right) \hat{s}_{A}^{2}+\left(n_{B}-1\right) \hat{s}_{B}^{2}\right]^{1 / 2}}\left[\frac{\left(n_{A}+n_{B}-2\right) n_{A} n_{B}}{\left(n_{A}+n_{B}\right)}\right]^{1 / 2}$
$=\frac{(492.5-505.8)}{\left[19 \times 18.8^{2}+19 \times 16.05^{2}\right]^{1 / 2}}[380]^{1 / 2}=2.406$, under the NH.
For the one-tailed test with d.f. $=2 n-2=38$ and $\alpha=0.01, t_{38,0.01}=2.43$.
The difference is not significant. Do not reject NH.
(b) NH $p_{A}=p_{B}$
$\mathrm{AH} p_{B}<p_{A}$
$\hat{p}_{A}=10 / 20$ and $\hat{p}_{B}=9 / 20$. The average proportion $\hat{p}=(10+9) /(20+20)=19 / 40$
Difference in proportions $\hat{p}_{B}-\hat{p}_{A}=-1 / 20=-0.05$
Under the NH, the variance of difference between observed proportions is
$\operatorname{Var}\left[\hat{p}_{B}-\hat{p}_{A}\right]=\operatorname{Var}\left[\hat{p}_{B}\right]+\operatorname{Var}\left[\hat{p}_{A}\right]=2 \frac{\hat{p}(1-\hat{p})}{n_{B}}=2 \times(19 / 40) \times(21 / 40)=0.0249$
because $n_{B}=n_{A}=20$ and also, under the NH, $p_{B}=p_{A}=p$.
Under the NH, $\left[\hat{p}_{B}-\hat{p}_{A}\right] \sim N\left(0,(0.0249)^{1 / 2}\right)$ approximately.
Normal score $z=\frac{-0.05-0}{\sqrt{0.0249}}=-0.3169$.
For the one-tailed test with $\alpha=0.01, z_{0.01}=2.326$.
Because the test result is not significant do not reject the null hypothesis
5.26. Groundwater quality. The following are measurements of concentrations of chloride in milligrams per liter in a shallow unconfined aquifer taken at intervals of three months [from J. Harris, J.C. Loftis, and R.H. Montgomery (1987). "Statistical models for characterizing ground-water quality", Groundwater, Vol. 25, pp. 185-193]:

| 38 | 40 | 35 | 37 | 32 | 37 | 37 | 32 | 45 | 38 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 33.8 | 14 | 39 | 46 | 48 | 41 | 35 | 49 | 64 | 73 |
|  | 67 | 67 | 59 | 73 | 92.5 | 45.5 | 40.4 | 33.9 | 28.1 |

Compute the coefficients of skewness and kurtosis and make an approximate test for normality, using $\alpha=.05$. (Data used with the kind courtesy of the publishers).

## Solution.

For groundwater quality, sample size $n=29$; sample coefficient of skewness $g_{1}=$ 0.9660 and sample coefficient of kurtosis $=3.737$.
$\operatorname{Var}\left[\mathrm{g}_{1}\right]=\frac{6 n(n-1)}{(n-2)(n+1)(n+3)}=\frac{6 \times 29 \times 28}{27 \times 30 \times 32}=0.187$

For the population coefficient of skewness $\gamma_{1}$, the $95 \%$ confidence limits are $\pm 1.96 \times \sqrt{0.187}= \pm 0.85$

For the population coefficient of skewness $\gamma_{1}$, the approximate $95 \%$ confidence limits are (from the text, for the given sample size) are 4.60, 1.85.
Reject NH of normality in distribution.
5.27. Kolmogorov-Smirnov two sample test on flows. Annual rainfall from 1918 to 1978 in the Po River basin of northern Italy are given in the penultimate column of Table E.7.2. Divide the record into two parts of 30 and 31 years. Determine whether the rainfall regime has changed by testing whether the two parts belong to the same population at the 5 percent level of significance using the KolmogorovSmirnov two-sample test.

Solution. Po annual rainfall from Table E7.2

| First half of data $m=30$ :- 1133 | 999 | 1501 | 807 | 1051 | 969 | 997 | 1090 | 1356 | 1133 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1171 | 876 | 1159 | 993 | 1112 | 1128 | 1345 | 1290 | 1259 | 1529 |
| 940 | 1196 | 1046 | 1218 | 948 | 896 | 950 | 846 | 1011 | 1096 |
| Second half of data $n=31:-1100$ | 922 | 978 | 1496 | 913 | 1046 | 1100 | 886 | 1028 | 1215 |
| 1142 | 1422 | 1654 | 987 | 909 | 1362 | 1026 | 1015 | 1228 | 885 |
| 1264995 | 986 | 017 | 349 | 1029 | 95 | 323 | 1318 | 156 |  |

The (critical) $d_{m, n}$ occurs after the $24^{\text {th }}$ item of the ranked data of the first part with m $=30$, and the $19^{\text {th }}$ item of the ranked data of the second part with $n=31$.
$d_{m, n}=0.1871$ and $\quad\left[\frac{m n}{m+n}\right]^{1 / 2} \times d_{m, n}=\left[\frac{30 \times 31}{61}\right]^{1 / 2} \times 1871=0.7305$.
This is low when compared with the critical value in Table C.7. Therefore do not reject the NH that the two parts belong to the same population.
5.28. Lilliefor's test. The following are the ranked annual inflows in $10^{6}$, for the period 1950 to 1974 , to the Warragamba reservoir, which supplies water to the city of Sydney, Australia:

| 724 | 1,505 | 3,310 | 6,551 | 6,915 | 7,114 | 7,811 | 8,962 | 9,219 | 9,664 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 9,840 | 10,134 | 10,299 | 10,824 | 11,953 | 12,566 | 13,969 | 14,941 | 15,449 | 16,800 |
| 17,601 | 18,250 | 18,483 | 19,081 | 20,242 |  |  |  |  |  |

(By kind courtesy of the University of New South Wales, Sydney)
Test whether the distribution is normal using Table C. 8 of Appendix C, which is Lilliefors' test for normality corrected by Dallal and Wilkinson (1986) for the purpose.

## Solution.

| Ranked <br> data | 724 |  |  |  |  |  |  |  |  |  |  | 10134 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| $F_{n}(x)$ | .04 | .08 | .12 | .16 | .20 | .24 | .28 | .32 | .36 | .40 | .44 | .48 |
| $F_{0}(x)$ | .026 | .036 | .071 | .192 | .211 | .221 | .261 | .334 | .352 | .383 | .395 | .416 |


| Ranked <br> data |  | 10824 | 11953 |  |  |  |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $k$ | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
| $F_{n}(x)$ | .52 | .56 | .60 | .64 | .68 | .72 | .76 | .80 | .84 | .88 | .92 | .96 | 1.00 |
| $F_{0}(x)$ | .428 | .466 | .549 | .593 | .689 | .75 | .78 | .84 | .88 | .90 | .91 | .92 | .95 |

The critical value $d_{n}=0.094$, which occurs at $k=14$. From Table C.8, the limiting value $D_{n, 0.05}=0.173$. Therefore do not reject the NH that the distribution is normal
5.29. Chi-squared test. We transformed the Warragamba annual flows (introduced in Problem 5.28 but extended over a 103-year period) to natural logarithms. The data were ranked and sorted into 10 classes with equal class intervals as follows:

| Class boundary | 11.61 | 12.07 | 12.52 | 12.98 | 13.44 | 13.89 | 14.35 | 14.81 | 15.26 | $\infty$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Number of class | 2 | 6 | 8 | 14 | 19 | 16 | 17 | 11 | 8 | 2 |

The total is $n=103$, the estimated mean is 13.50 , and the variance is 0.874 .
Using the chi-squared goodness-of-fit procedure, test the hypothesis that the log-transformed flows are normally distributed with $\alpha=.05$.

## Solution.

As given, the mean $\bar{x}=13.50$ and the standard deviation $\hat{s}=\sqrt{0.874}=0.935$

| $x$ | $z=(x-\bar{x}) / \hat{s}$ | $\Phi(z)$ | $E_{i}$ | $O_{i}$ | $\left(E_{i}-O_{i}\right)^{2} / E_{i}$ |
| :--- | :--- | :--- | :--- | :--- | :--- |
| 11.61 | -2.022 | 0.0216 | 2.26 | 2 | 0.020 |
| 12.67 | -1.530 | 0.0630 | 4.30 | 6 | 0.672 |
| 12.52 | -1.048 | 0.1473 | 8.71 | 8 | 0.058 |
| 12.98 | -0.556 | 0.2891 | 14.65 | 14 | 0.029 |
| 13.44 | -0.064 | 0.4737 | 19.12 | 19 | 0.001 |
| 13.89 | 0.417 | 0.6617 | 19.26 | 16 | 0.552 |
| 14.35 | 0.909 | 0.8183 | 10.10 | 17 | 0.050 |
| 14.81 | 1.401 | 0.9194 | 10.36 | 11 | 0.040 |
| 15.26 | 1.883 | 0.9702 | 5.19 | 8 | 1.521 |
|  |  |  | 3.02 | 2 | 0.361 |
| SUM |  |  | 103.0 | 103 | 3.314 |

Thus from the last entry of the above table, $X^{2}=3.314$. The degrees of freedom $v$ $=10-1-2=7 . \chi^{2}{ }_{7,0.05}=14.1$. Do not reject the NH of the lognormal model.
5.30. Anderson-Darling test. Reconsider the data of Problem 5.28. Test the hypothesis of normality using the Anderson-Darling test at the 5 percent level of significance.

Solution.

| Rank | data | First F term | Sec (1-F) term | Sum term |
| :--- | :--- | :--- | :--- | :--- |
| 1 | 724 | .02608 | .04989 | -6.64 |
| 2 | 1505 | .03606 | .07600 | -24.34 |
| 3 | 3310 | .07124 | .09299 | -49.43 |
| 4 | 6551 | .19193 | .10032 | -77.08 |
| 5 | 6915 | .21073 | .12293 | -109.96 |
| 6 | 7114 | .22145 | .15548 | -147.01 |
| 7 | 7811 | .26136 | .22219 | -184.01 |
| 8 | 8962 | .33448 | .25098 | -221.18 |
| 9 | 9219 | .35184 | .31110 | -258.78 |


| Rank | data | First F term | Sec (1-F) term | Sum term |
| :--- | :--- | :--- | :--- | :--- |
| 10 | 9664 | .38263 | .40715 | -294.11 |
| 11 | 9840 | .39503 | .45136 | -330.32 |
| 12 | 10134 | .41598 | .53402 | -364.92 |
| 13 | 10299 | .42784 | .57216 | -400.11 |
| 14 | 10824 | .46598 | .58402 | -435.24 |
| 15 | 11953 | .54864 | .60497 | -467.23 |
| 16 | 12566 | .59285 | .61737 | -498.39 |
| 17 | 13969 | .68890 | .64816 | -524.99 |

-524.99

| Rank | data | First F term | Sec (1-F) term | Sum term |
| :--- | :--- | :--- | :--- | :--- |
| 18 | 14941 | 0.74902 | 0.66552 | -549.360 |
| 19 | 15449 | 0.77781 | 0.73864 | -569.87 |
| 20 | 16800 | 0.84452 | 0.77855 | -586.22 |
| 21 | 17601 | 0.87707 | 0.78927 | -601.30 |
| 22 | 18250 | 0.89968 | 0.80807 | -615.01 |
| 23 | 18483 | 0.90701 | 0.92876 | -622.73 |
| 24 | 19081 | 0.92400 | 0.96394 | -628.17 |
| 25 | 20242 | 0.95011 | 0.97392 | -631.97 |

The estimated $A^{2}$ value is $-25-(-631.9708) / 25=0.27883$. This is less than the limiting value, as given in the text, at the 5 percent level of significance. Do not reject the NH of the normal model.
5.31. Harmonic coefficients. Monthly inflows into Warragamba reservoir in New South Wales, Australia were computed over the period 1881-1983 in units of $1000 \mathrm{~m}^{3}$.
It is proposed to fit the following harmonic model to the periodicity in the means:

$$
\mu_{\tau}=\mu+\sum_{i=1}^{6} \alpha_{i} \sin (2 \pi i \tau / 6)+\sum_{i=1}^{6} \beta_{i} \cos (2 \pi i \tau / 6)
$$

where $\mu_{\tau}$ is the harmonic mean in month $\tau, \tau=1,2, \ldots, 12 ; 1$ denotes January and so on; $\mu$ is the annual mean; $\alpha_{i}, \beta_{i}, i=1,2, \ldots, 6$ are harmonic coefficients. The following harmonic coefficients have been computed in $\mathrm{m}^{3}$.

| Harmonic $i$ | 1 | 2 | 3 | 4 | 5 | 6 |
| ---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\alpha_{i}$ | 6,066 | 8,568 | $-12,629$ | $-2,135$ | 8,954 | 0 |
| $\beta_{i}$ | $-39,393$ | 7,062 | $-16,877$ | 6,586 | $-5,204$ | 875 |
| SS | 9,817 | 762 | 2,746 | 296 | 663 | 5 |

Note: The last row gives the sum of squares associated with each harmonic in units $10^{8}\left(\mathrm{~m}^{3}\right)^{2}$.
Determine by an analysis of variance how many of the harmonics are significant using $\alpha=.05$. There are 1,236 items of data and the total sum of squares is $457,175 \times 10^{8}\left(\mathrm{~m}^{3}\right)^{2}$.

Calculate the fitted means using the significant harmonics.

## Solution.

| Variables | Sum of squares | Degrees of freedom | Mean square | F |
| :--- | :---: | :---: | :---: | :---: |
| $\alpha_{i}, \beta_{i} \mathrm{n}$ | 4,472 | 9 | $4,472 / 9=4,969$ | 1.37 |
| $i=2,3,4,56$ |  |  |  |  |
| $\alpha_{1}, \beta_{1}$ | 9,817 | 2 | $9,817 / 2=4,908.5$ | 13.6 |
| Residuals | 442,886 | 1,224 | $44,886 / 1,224=361.8$ |  |
| Total | 457,175 | 1,235 |  |  |

From Table C.4, $F_{2, \infty, 0.05}=4.63$.
Only $\alpha_{1}, \beta_{1}$ (which represent the first harmonic) are significant. The fitted means are obtained as follows:
$\mu_{\tau}=\mu+6066 \sin (2 \pi \tau / 6)-639393 \cos (2 \pi \tau / 6)$
5.32. Analysis of variance of road data. Using the data from Table 5.7 .5 for the road rutting experiment of Example 5.34 test whether the base thickness has a significant effect on the depth of the rutting. For this test combine the results from the two types of base material.

## Solution.

From Eq. 5.7.5a $\quad S S_{T}=\sum_{i=1}^{k} \sum_{j=1}^{n} x_{i j}{ }^{2}-\frac{1}{k n} T^{2}=194.37-1446.33 / 10=49.73$
$S S_{T r}=\frac{1}{n} \sum_{i=1}^{k} T_{i}^{2}-\frac{1}{k n} T^{2}=148.08-144.63=3.45$

| Source | Degrees of freedom | SS | mean SS | F |
| :--- | :---: | :---: | :---: | :---: |
| Base thickness | 2 | 3.45 | $3.45 / 2$ | 1.725 |
| Error | 27 | 46.28 | $46.28 / 27$ | 1.714 |
| Total | 29 | 49.73 |  |  |

$F_{2,27,0.05} \approx 3.35$
Do not reject the NH that all treatments are the same.
5.33 Analysis of variance of dynamic effect of vehicles. We return again to the road rutting experiment of Example 5.34. From Fig. 5.7.2 and other data sets in Tables E.5.1 to E.5.6, it is suggested that the movement of heavy vehicles can have a dynamic effect (that is, time-dependent) on the road surface. This can change with road material and thickness. Analyze the variance of this effect by using the estimated error sum of squares $\left(x_{i}-\hat{\beta}_{0}-\hat{\beta}_{1} u_{i}\right)^{2}$ to represent it. Hence determine whether it is significant with $\alpha=.05$. Assume (a) that the intercepts and gradients for the 30 series studied in Example 5.34 are variable; (b) Assume a constant intercept and gradient.

## Solution.

(a) Analysis of variance for a two way classification: Dynamic effects. Intercepts and gradients are variable in the series studied
Source Degrees of freedom SS mean SS F

| Treatment | 1 | 76.3 | 76.3 | 1.67 |
| :--- | ---: | ---: | ---: | :--- |
| Blocks | 2 | 167.6 | 83.8 | 1.84 |
| Interaction | 2 | 131.8 | 65.9 | 1.4 |
| Error | 24 | 1094.9 | 45.6 |  |
| Total | 29 |  |  |  |

$F_{2,25,0.05} \approx 3.39$
Therefore do not reject the NH that the dynamic effect of vehicles has no influence on the variances considered, assuming variable intercepts and gradients in the series studied.
(b) Analysis of variance for a two way classification: Dynamic effects. Intercepts and gradients are constant in the series studied

| Source | Degrees of freedom | SS | mean SS | F |
| :--- | :---: | :--- | :--- | :--- |
| Treatment | 1 | 0.00264 | 0.00264 | 0.25 |
| Blocks | 2 | 0.0296 | 0.0148 | 1.408 |
| Interaction | 2 | 0.048 | 0.024 | 2.33 |
| Error | 24 | 0.25 | 0.0104 |  |
| Total | 29 |  |  |  |

$F_{2,25,0.05} \approx 3.39$
Therefore do not reject the NH that the dynamic effect of vehicles has no influence on the variances considered, assuming constant intercepts and gradients in the series studied.
5.34. Normal plots. In Table E.1.2 and Problems 1.6 and 1.14 of Chapter 1 there are lists of compressive strengths of concrete in newtons per square millimeter obtained by testing three lots of test cubes. Make comparative normal probability plots of these three sets of test results. Comment on the results.

Solution. For the Douglas data (Table E.1.2) and Ash2data (Problem 1.14) ranked below, we use the following plotting position: $p_{i}=(i-0.375) /(n+0.25)$ where $n=$ 40

| $i=$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Douglas | 49.9 | 50.7 | 52.5 | 53.2 | 53.4 | 54.4 | 54.6 | 55.8 | 56.3 | 56.7 |
| Ash2data | 43.3 | 45.7 | 45.8 | 47.3 | 47.3 | 47.7 | 47.2 | 48.1 | 48.5 | 48.5 |
| $p_{i}=$ | .016 | .04 | .065 | .090 | .0925 | .14 | .165 | .189 | .214 | .239 |
| $Z=$ | -2.16 | -1.75 | -1.52 | -1.34 | -1.35 | -1.08 | -0.98 | -0.88 | -0.79 | -0.71 |


| $i=$ | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Douglas | 56.9 | 57.8 | 57.9 | 58.8 | 58.9 | 59.0 | 59.6 | 59.8 | 59.8 | 60.0 |
| Ash2data | 48.8 | 49.1 | 49.3 | 49.4 | 49.5 | 49.5 | 49.6 | 50.2 | 50.5 | 51.0 |
| $p_{i}=$ | .264 | .289 | .314 | .339 | .363 | .388 | .413 | .453 | .463 | .488 |
| $Z=$ | -0.63 | -0.56 | -0.49 | -0.42 | -0.35 | -0.30 | -0.22 | -0.12 | -0.10 | -0.02 |


| $i=$ | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Douglas | 60.2 | 60.5 | 60.5 | 60.5 | 60.9 | 60.9 | 61.1 | 61.5 | 61.9 | 63.3 |
| Ash2data | 51.0 | 51.1 | 52.0 | 52.2 | 52.5 | 52.7 | 52.8 | 52.9 | 53.1 | 53.1 |
| $p_{i}=$ | .512 | .537 | .562 | .587 | .612 | .637 | .662 | .686 | .711 | .736 |
| $z=$ | 0.03 | 0.095 | 0.16 | 0.22 | 0.28 | 0.35 | 0.42 | 0.485 | 0.56 | 0.635 |


| $i=$ | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Douglas | 63.4 | 63.9 | 63.9 | 65.7 | 67.2 | 67.3 | 68.1 | 68.3 | 68.9 | 69.5 |
| Ash2data | 53.4 | 53.4 | 53.8 | 54.0 | 54.1 | 54.2 | 55.2 | 55.2 | 56.5 | 58.6 |
| $p_{i}=$ | .761 | .786 | .811 | .835 | .866 | .885 | .910 | .935 | .960 | .985 |
| $z=$ | 0.71 | 0.79 | 0.88 | 0.99 | 1.11 | 1.20 | 1.34 | 1.52 | 1.75 | 2.17 |

For the smaller sample of Ash1data (Problem 1.16) ranked below we use the following plotting position: $p_{i}=(i-0.35) / n$, where $n=16$.

| $i=$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Ash1data | 28.75 | 33.5 | 36.75 | 38.0 | 39.25 | 40.25 | 41.75 | 41.75 |
| $p_{i}$ | .038 | .100 | .162 | .223 | .285 | .346 | .408 | .469 |
| $z$ | -1.78 | -1.78 | -0.99 | -0.76 | -0.57 | -0.40 | -0.24 | -0.75 |


| $i=$ | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Ash1data | 42.0 | 42.25 | 42.50 | 42.75 | 43.25 | 43.50 | 44.25 | 44.25 |
| $p_{i}$ | .531 | .592 | .654 | .715 | .777 | .839 | .902 | .962 |
| $z$ | .095 | .235 | .395 | .570 | .765 | .990 | 1.29 | 1.77 |



Sketch of normal probability plot for ASH 2data (Problem 1.14)


Sketch of normal probability plot for ASH1data (Problem 1.6)

44 |
$42 \mid$
$40 \mid$
$38 \mid$
$36 \mid$
$34 \mid$
$32 \mid$
$30 \mid$
$28 \mid$

| $z=$ | -2.5 | -2.0 | -1.5 | -1.0 | -0.5 | 0.0 | 0.5 | 1.0 | 1.5 | 2.0 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 2.5 |  |  |  |  |  |  |  |  |  |  |

Comments: For the concrete compressive strengths of the Douglas data (Table E.1.2) and ASH2 data (Problem 1.16) indicate that the normal distribution is a good fit. However, the normal fit is less satisfactory for ASH1 data (Problem 1.4). The small sample size is a likely cause.
5.35. Lognormal probability plotting of sunspot data. Make normal and lognormal probability plots the Wolfer sunspot data of Problem 5.17. Comment on the results.

## Solution.

The lognormal distribution provides a better fit compared to the normal as shown in the 3 following graphs.
INSERT 3 GRAPHS FOR PROBLEM 5.35
5.36. Number of vehicles passing using a Poisson probability plot. Draw curves to represent for each Poisson occurrence, say from 0 to 6 , the relationship between the probability of occurrence and the value of the parameter from say 0.1 to 1.0 . Plot the probabilities of the following observer counts of the number of vehicles passing a point of observation:

| Count | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Frequency | 221 | 95 | 24 | 12 | 5 | 2 | 1 |

Is the Poisson a reasonable model? What is the estimated parameter from the plot?

Solution. Calculate probabilities obtained from the data

| Count | 0 | 1 | 2 | 3 | 4 | 5 | 6 | Sum |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Frequency | 221 | 95 | 24 | 12 | 5 | 2 | 1 | 360 |
| $p_{X}(x)$ | 0.613 | 0.263 | 0.067 | 0.013 |  |  |  |  |

Calculate Poisson probabilities for counts $0,1,2$,

| Count/Poisson parameter | 0 | 1 | 2 |
| :---: | :---: | :---: | :---: |
| 0.2 | 0.818 | 0.164 | 0.016 |
| 0.3 | 0.741 | 0.222 | 0.033 |
| 0.4 | 0.670 | 0.268 | 0.054 |
| 0.5 | 0.607 | 0.303 | 0.076 |
| 0.6 | 0.549 | 0.329 | 0.099 |
| 0.7 | 0.497 | 0.348 | 0.122 |

Sketch of Poisson probability vs. parameter for counts

- zero
- one
two

From given data $\circ$


Estimated parameter $\approx 0.5$ (see vertical line).
The vertical straight line suggests that the Poisson provides a good approximation.
5.37. Tensile strengths on Weibull probability paper. The original work of Weibull (1951) on the strengths of materials suggests that the breaking tensile stress of concrete has a Weibull distribution. The following results of tensile strengths, in newtons per square millimeter, were obtained from 12 tests conducted in a laboratory.

$$
\begin{array}{llllllllllll}
14.8 & 15.7 & 15.1 & 13.8 & 14.3 & 16.6 & 14.1 & 16.4 & 16.1 & 13.7 & 13.9 & 14.6
\end{array}
$$

Plot these results on Weibull probability paper. Fit a straight line by eye and comment on the results.

## Solution.

See plot on Weibull paper. There is much scatter. A larger sample may have provided a better fit.

INSERT GRAPH FOR PROBLEM 5.37
5.38. Hanging histogram of the Tevere flood flows. For the annual maximum flows of the Tevere River of Problem 5.16, draw a hanging histogram using the lognormal distribution. Comment on the results.

Solution. Sketch of hanging histogram: Frequency vs. Flow


Sketch of hanging histogram for lognormal


There is indication of some misfit at the mode.
5.39. Accommodation of outliers. Consider the annual maximum flows in the North Fork Sun River listed in Table E.5.7, series 1. It is noted that there is one suspected outlier in the series of 25 annual maximum flows.
(a) Plot the data on normal probability paper
(b) Fit a straight line by eye without considering the outlier
(c) Excluding the outlier, calculate the mean $\bar{x}$ and unbiased variance $\hat{S}^{2}$.

For incorporating outliers, the procedure adopted by W.R.C.(1981) is to empirically increase the data base using historical evidence, if available, and then revise the mean and variance. Suppose $l$ outliers are identified in a record of $n_{R}$ years; and from past information, such as marks on bridge piers, it is found that the highest recorded flood level has not been exceeded in $n_{T}$ years. Then the revised mean $\tilde{x}$ and revised variance $\widetilde{S}^{2}$ of the extended data base, are calculated as follows:

$$
\begin{gathered}
\tilde{x}=\left[\left(n_{T}-l\right) \bar{x}+\sum_{i=n_{R}-l+1}^{i=n_{R}} x_{(i)}\right] / n_{T} \\
\left(n_{T}-1\right) \widetilde{s}^{2}=\left[\hat{s}^{2}\left(n_{T}-l\right)+\left(n_{T}-l\right)(\bar{x}-\tilde{x})^{2}+\sum_{i=n_{R}-l+1}^{i=n_{R}}\left(x_{(i)}-\tilde{x}\right)^{2}\right] .
\end{gathered}
$$

[The first equation is a direct adjustment of the mean and the second equation follows from the ANOVA methods of Section 5.7.]

If the magnitude of the outlier has not been exceeded for 500 years, calculate the annual maximum flow with a return period of 500 years assuming a normal distribution.

## Solution.

Mean $\bar{x}=3127.9 \mathrm{f}^{3} / \mathrm{s}$; standard deviation $\hat{s}==785.9 \mathrm{f}^{3} / \mathrm{s}$.

$$
x_{(25)}=51100 \mathrm{f}^{3} / \mathrm{s} . \quad \text { Also, } z_{(25)}=\frac{51100-3127.9}{785.9}=61.04
$$

Clearly, the value of 51100 is an outlier

$$
\begin{aligned}
& \tilde{x}=[499 \times 3127.9+51100] / 500=3223.9 \mathrm{f}^{3} / \mathrm{s} . \\
& 499 \widetilde{s}^{2}=\left[785.9^{2} \times 499+499(3127.9-3223.9)^{2}+(51100-3223.9)^{2}\right] \\
& \widetilde{s}=1042.2 \mathrm{f}^{3} / \mathrm{s} \\
& \Phi\left(z_{(500)}\right)=0.998 \text { and } z_{(500)}=2.88 \\
& x_{(500)}=2.88 \times 1042.2+3223.9=6225 \mathrm{f}^{3} / \mathrm{s} .
\end{aligned}
$$

Compared to the outlier of $51,100 \mathrm{f}^{3} / \mathrm{s}$ this is an under estimate. It may be because 500 years of past observations are insufficient for this type of application, considering the magnitude of the given outlier. The alternative reason is that, outliers, such as the value of 51100, have a different probability distribution from that of the other values (see Eq. 5.9.1. in text)
INSERT GRAPH FOR PROBLEM 5.39

Blank page

## Applied Statistics for Civil and Environmental Engineers <br> Problem Solution Manual <br> by N.T. Kottegoda and R. Rosso <br> Chapter 6 - Methods of Regression and Multivariate Analysis

6.1. Dissolved oxygen The following observations of dissolved oxygen (DO) were made with respect to time of travel downstream from a point of regulation in a river.

| Time of travel, days | 0 | 0.6 | 1.1 | 1.7 | 1.9 | 2.4 | 2.8 | 3.3 | 3.7 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| DO, ppm | 0.39 | 0.37 | 0.31 | 0.28 | 0.27 | 0.25 | 0.20 | 0.17 | 0.16 |

Fit a linear regression and estimate the parameters. Calculate the coefficient of determination. Does a straight line provide a reasonable fit?

## Solution.

$\bar{x}=1.944 \quad \bar{y}=0.267 \quad S_{x x}=12.22 \quad S_{y y}=0.05 \quad S_{x y}=0.05$
$\hat{\beta}_{0}=0.394 \quad \hat{\beta}_{1}=-0.0656 \quad r=-0.992 \quad r^{2}=0.985$
For $t=4$ days $\quad \mathrm{DO}=0.1491 \mathrm{ppm}$
$95 \%$ confidence limits: $0.1391,0.1146$.
Sample $t$ value for slope $=21.2$ and for intercept $=56.2$. Obviously very high .
Straight line: good fit to the data.
6.2. Population growth. A small city has doubled in population in 9 years. The following approximate counts have been made during the period.

| Year | 1 | 2 | 3 | 4 | 5 |
| :--- | ---: | ---: | ---: | ---: | ---: |
| Population in 1000's | 100 | 107 | 115 | 124 | 135 |
| Year | 6 | 7 | 8 | 9 | 10 |
| Population in 1000's | 146 | 158 | 171 | 185 | 200 |

Plot the data. Determine the sample correlation coefficient. Decide whether a linear model provides a good fit or whether there should be a transformation of the response variable (population). Give 95 percent confidence limits for the mean or expected population in year 15 if growth patterns do not change.

## Solution.

Sketch of population growth in time: population in 1000's years vs. year number


| Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\hat{\beta}_{0}=82.80$ | $\hat{\beta}_{1}=11.145$ | $r=-0.994$ | $r^{2}=0.988$ |  |  |  | 10 |  |  |
| $\bar{x}=5.5$ | $\hat{\sigma}^{2}=15.5819$ | $S_{x x}=82.5$ |  |  |  |  |  |  |  |

The straight line
$P=11.145 Y+82.8 \quad$ provides a good approximation
$t_{8,0.025}=2.306$.
Confidence limits for $y=15$ :-
$11.145 \times 15+82.8 \pm 2.306 \sqrt{15.5819}\left[1+\frac{1}{10}+\frac{(15-5.5)^{2}}{82.5}\right]^{1 / 2} \Rightarrow 240,260$
6.3. Water quality For the water quality measurements on the River Ouse at Clapham, England, data given in Problem 1.15 (Chapter 1) determine the linear regression equation by least squares using phosphate as the explanatory variable. Comment on the model and the results.

## Solution.

The linear regression model (see Eqs. 6.1 to 6.7) takes the form
CHL $=0.454$ phos $+64.4+\varepsilon$
With $r=0.03$, this serves no purpose for prediction purposes.
6.4. Correlation of low flows. The lowest annual flows measured, in cubic meters per second, at stations $X$ and $Y$ on the Jackson and Cowpasture rivers, respectively, in the United States are to be correlated in order to extend the shorter record at station $Y$. A simple linear regression model is to be used. The following summary statistics have been computed over a 12 -year period:
Sum $X=28.77$; Sum $Y=28.23$; Sum $X X=73.14$; Sum $Y Y=71.20$; Sum $X Y=71.53$.
(1) Find least squares estimates of the parameters.
(2) What is the standard error of the residuals?
(3) Estimate the coefficient of correlation.
(4) Find approximate 95 percent confidence limits for the population correlation coefficient.

Solution.
(a) $\hat{\beta}_{1}=\frac{S_{x y}}{S_{x x}}=\frac{\Sigma x y-(\Sigma x \Sigma y) / n}{\Sigma x^{2}-(\Sigma x)^{2} / n}=\frac{71.53-28.77 \times 28.23 / 12}{73.14-28.77^{2} / 12}=0.924$.
$\hat{\beta}_{0}=\bar{y}-\hat{\beta}_{1} \bar{x}=\frac{28.23}{12}-0.924 \times \frac{28.77}{12}=0.137$
The regression model takes the form
$Y=0.924 \mathrm{x}+0.137+\varepsilon$
(b) $\hat{\sigma}^{2}=\frac{1}{n-2}\left(S_{y y}-\frac{S_{x y}{ }^{2}}{S_{x x}}\right)=\frac{1}{10}\left[\left(\Sigma y^{2}-\frac{(\Sigma y)^{2}}{12}\right)-\frac{\left(\Sigma x y-\frac{\Sigma x \Sigma y}{12}\right)^{2}}{\Sigma x^{2}-\frac{(\Sigma x)^{2}}{12}}\right]=0.123$
$\hat{\sigma}=0.352$
(c) $r=\frac{S_{x y}}{\left[S_{x x} S_{y y}\right]^{1 / 2}}=0.862$
(d) $0.572,0.961$
6.5. Extension of steel wires Ten steel wires of diameter 0.5 mm and length 2.5 m were extended in a laboratory by applying vertical forces of varying magnitudes. Results are as follows

| Force, kg | 15 | 19 | 25 | 35 | 42 | 48 | 53 | 56 | 62 | 65 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Increase in length, mm | 1.7 | 2.1 | 2.5 | 3.4 | 3.9 | 4.9 | 5.4 | 5.7 | 6.6 | 7.2 |

(1) Estimate the parameters of a simple linear regression model with force as the explanatory variable
(2) Find 95 percent confidence limits for the two parameters
(3) Test the hypothesis that the intercept is zero
(4) What are the conclusions?

Solution. Sketch of increase in length, mm vs. force, kg

(a) The parameters are $\hat{\beta}_{0}=-0.1212$ and $\hat{\beta}_{1}=0.1062$ The regression model takes the form
$I N C=0.1062$ force $-0.1212+\varepsilon$
(b) The $95 \%$ confidence limits for $\beta_{1}$ are 0.095 and 0.118 and for
$\beta_{0}$ are -0.633 and 0.391
(c) Do not reject the NH that the intercept is zero
(d) The fit is good.
6.6. Rainfall-runoff relationship. Table E7.2 gives 61 years of rainfall and runoff (see columns 6 and 7) at Pontelagoscuro, on the Po river, in northeast Italy. Fit a simple regression model. Test the hypothesis that the slope is zero. Comment on the results. Suggest methods of forming a multiple regression model and the inclusion of other measurements that can enhance the relationship.

## Solution.

The parameters are $\hat{\beta}_{0}=475$ and $\hat{\beta}_{1}=0.943$ The regression model takes the form
RUNOFF $=0.943$ rainfall $+475+\varepsilon$
The $r$ value is 0.911
The sample $t$ value for the slope is 17.015
Clearly, the NH of zero slope is rejected
The $95 \%$ confidence limits for $\beta_{1}$ are 0.7077 and 0.9752
To improve the regression model, include antecedent rainfall, allow for evaporation losses, consider the inclusion of ground ware contributions
6.7. Asbestos concentrations. Keifer, M.J., R.M. Buchan, T.J. Keefe and K.D. Blehm (1987) ("A predictive model for determining asbestos concentrations for fibres less than five millimeters in length," Environmental Research, Vol. 43, pp. 31-38) give the following data for PCM (phase contrast microscopy) and SEM (scanning electron microscopy) concentrations:

| Filter | PCM | SEM | Filter | PCM | SEM |
| ---: | ---: | ---: | ---: | ---: | ---: |
| 1 | 3.14 | 7.79 | 16 | 0.41 | 1.86 |
| 2 | 2.61 | 6.85 | 17 | 0.77 | 2.90 |
| 3 | 3.03 | 7.60 | 18 | 1.63 | 4.92 |
| 4 | 4.03 | 9.29 | 19 | 3.99 | 9.22 |
| 5 | 7.82 | 14.8 | 20 | 2.94 | 7.44 |
| 6 | 5.61 | 11.72 | 21 | 1.02 | 3.54 |
| 7 | 4.23 | 9.61 | 22 | 1.67 | 5.00 |
| 8 | 0.62 | 2.49 | 23 | 6.33 | 12.76 |
| 9 | 1.09 | 3.71 | 24 | 2.38 | 6.42 |
| 10 | 0.73 | 2.79 | 25 | 1.93 | 5.34 |
| 11 | 0.70 | 2.71 | 26 | 6.29 | 12.70 |
| 12 | 7.92 | 14.94 | 27 | 3.77 | 8.86 |
| 13 | 2.50 | 6.64 | 28 | 4.50 | 10.04 |
| 14 | 1.91 | 5.50 | 29 | 4.54 | 12.10 |
| 15 | 4.98 | 10.78 | 30 | 0.48 | 2.08 |

Plot the data with PCM as the explanatory variable. Estimate the parameters of a simple linear regression model and show the straight line and the 95 percent confidence limits. Comment on the model. For a future value of $\mathrm{PCM}=8.5$ give the predicted value of SEM and the 95 percent prediction interval. (Data used with permission from the Academic Press Inc., Orlando, Florida 32887-6777 and the authors.)

Solution. PCM is the explanatory variable, $X$; SEM is the response variable $Y$
$\bar{x}=3.052 \quad \bar{y}=7.413 \quad S_{x x}=141.5 \quad S_{y y}=438.4 \quad S_{x y}=242.25$
$\hat{\beta}_{0}=2.188 \quad \hat{\beta}_{1}=1.712 \quad r=0.9721 \quad$ The regression model takes the form
$S E M=1.712 \mathrm{pcm}+2.188+\varepsilon$
For $p c m=8.5$, SEM $=16.74$
$95 \%$ confidence limits for SEM: 14.64 and 18.84
6.8. Alternative Least Squares. Rewrite the multiple regression model with two explanatory variables subtracting the sample means from each of the variables. Hence write equations for the matrix and parameters.

Solution. The model takes the form
$Y_{i}-\bar{Y}=\beta_{1}\left(x_{i 1}-\bar{x}_{1}\right)+\beta_{2}\left(x_{i 2}-\bar{x}_{2}\right)+\varepsilon_{i}$, for $i=1,2, \ldots, n$
Let $\mathrm{X}^{\mathrm{T}} \mathrm{X}=\left[\begin{array}{ll}S_{11} & S_{12} \\ S_{21} & S_{22}\end{array}\right]$ where
$S_{j k}=\sum_{i=1}^{n}\left(x_{i j}-\bar{x}_{j}\right)\left(x_{i k}-\bar{x}_{k}\right) \quad$ for $j, k=1,2$. Thus
$\mathrm{X}^{\mathrm{T}} \mathrm{X}=\left[\begin{array}{ccc}x_{11}-\bar{x}_{1} & x_{21}-\bar{x}_{1} \ldots & x_{n 1}-\bar{x}_{1} \\ x_{12}-\bar{x}_{2} & x_{22}-\bar{x}_{2} \ldots & x_{n 2}-\bar{x}_{2}\end{array}\right]\left[\begin{array}{c}x_{11}-\bar{x}_{1} x_{12}-\bar{x}_{2} \\ x_{21}-\bar{x}_{1} x_{22}-\bar{x}_{2} \\ \cdot \\ x_{n 1}-\bar{x}_{1} x_{n 2}-\bar{x}_{2}\end{array}\right]$ and
$\mathrm{X}^{\mathrm{T}} \mathrm{y}=\left[\begin{array}{ccc}x_{11}-\bar{x}_{1} & x_{21}-\bar{x}_{1} \ldots . & x_{n 1}-\bar{x}_{1} \\ x_{12}-\bar{x}_{2} & x_{22}-\bar{x}_{2} \ldots & x_{n 2}-\bar{x}_{2}\end{array}\right]\left[\begin{array}{c}y_{1}-\bar{y} \\ y_{2}-\bar{y} \\ . \\ y_{n}-\bar{y}\end{array}\right]$.
The determinant $d=S_{11} S_{22}-S_{12}{ }^{2}$. Then
$\left[\mathrm{X}^{\mathrm{T}} \mathrm{X}\right]^{-1}=\left[\begin{array}{cc}S_{22} / d & -S_{12} / d \\ -S_{12} / d & S_{11} / d\end{array}\right]$ and parameters are estimated as
$\hat{\beta}=\left[X^{T} X\right]^{-1} X^{T} y$.
6.9. Road rutting. The rate of cutting of road ruts was measured with properties of asphalt and road materials in 31 experiments by Gorman, J.W. and R.J. Toman [(1966), "Selection of variables for fitting variables to data," Technometrics, Vol. 8, pp. 27-51]. The following is a modified and reduced form of the equation with specified variables and residual sums of squares (RSS)

$$
y=\beta_{0}+\beta_{1} x_{1}+\beta_{2} x_{2}+\beta_{3} x_{3}+\beta_{4} x_{4}+\varepsilon
$$

where $y=$ passes)
$\log$ (change of rut depth per million wheel
$x_{1}=\log$ (viscosity of asphalt)
$x_{2}=$ percent asphalt in surface course
$x_{3}=$ percent asphalt in base course
$x_{4}=$ percent fines in surface course

RSS
11.058
0.607
0.499
0.498
0.475

Determine the "best" form of equation to use. (Many more variables are used in the original work.)
Solution. For $n=31$
For one explanatory variable:

## Degrees of freedom

Sum of squares due to regression
Error sum of squares
Total sum of squares

| 2 | 10.451 |
| :---: | ---: |
| $n-3=28$ | 0.607 |
| $n-1=30$ | 11.058 |

For two explanatory variables:
Degrees of freedom
Sum of squares due to regression
Error sum of squares
Total sum of squares
2
$n-4=27$
10.559
$n-1=30$
From Eq. 6.2.19, the sample $F$ value corresponding to the increase from one to two explanatory values:
$F=\frac{10.559-10.451}{0.499 /(n-3)}=\frac{28 \times 0.108}{0.499}=6.06$
$F_{1,28,0.05} \approx 4.2$. This is significant.
Further increases of the number of explanatory variables are not of significant.
For example, increase from three to four variables, gives the corresponding $F$ value:-
$F=\frac{0.498-0.475}{0.475 /(n-5)}=\frac{26 \times 0.023}{0.475}=1.26$
Therefore, use a model with the first two explanatory variables.
6.10. Weighted least squares. The simple linear regression model

$$
Y=\beta_{0}+\beta_{1} x+\varepsilon
$$

is modified so that the variance of $Y$ depends in the magnitude of the $x$ as

$$
\sigma^{2}\left(Y_{i} \mid x_{i}\right)=\sigma_{i}^{2}, i=1,2, \ldots, n
$$

Rewrite the least squares equations.

## Solution.

The estimated error variance is found as
$\hat{\sigma}^{2}=\frac{1}{n-2} \sum_{i=1}^{n} \hat{\varepsilon}_{i}{ }^{2}=\frac{1}{n-2} \sum_{i=1}^{n}\left(y_{i}-\hat{\beta}_{0}-\hat{\beta}_{1} x_{i}\right)^{2}$.
In the case of weighted least squares, the variances of the $Y_{i}$ are $V\left[Y_{i}\right]=\sigma_{i}^{2}=\sigma^{2} / k_{i}$ for $i=1,2, \ldots, n$, where the $k_{i}$ are the weights.
But, as stated, $\sigma_{i}{ }^{2}=V\left[Y_{i}\right]=c x_{i}$
Hence, $k_{i}=\sigma^{2} / c x_{i}$
The sum of squared errors,

$$
S=\sum_{i=1}^{n} k_{i} \hat{\varepsilon}_{i}^{2}=\sum_{i=1}^{n} k_{i}\left(y_{i}-\hat{\beta}_{0}-\hat{\beta}_{1} x_{i}\right)^{2} .
$$

Taking partial derivatives
$\frac{\partial S}{\partial \beta_{0}}=\sum_{i=1}^{n} k_{i} y_{i}-\hat{\beta}_{0} \sum_{i=1}^{n} k_{i}-\hat{\beta}_{1} \sum_{i=1}^{n} k_{i} x_{i}=0$
$\frac{\partial S}{\partial \beta_{1}}=\sum_{i=1}^{n} k_{i} x_{i} y_{i}-\hat{\beta}_{1} \sum_{i=1}^{n} k_{i} x_{i}-\hat{\beta}_{1} \sum_{i=1}^{n} k_{i} x_{i}{ }^{2}=0$
Hence, the parameters are estimated.
6.11. Trend in precipitation. Annual precipitation in millimeters at Saracay in the Puyango Basin, Ecuador are given in the last column of Tables E.10.1. By using an appropriate regression equation, test the hypothesis that there is a trend in the precipitation.

Solution. $n=23$
$y$ series:- $1097.4 \quad 71.0 \quad 1110.1 \quad 576.2 \quad 705.0 \quad 192.6$ $\begin{array}{llllllllll}1500.0 & 2272.9 & 636.1 & 1645.5 & 2327.2 & 846.0 & 1818.0 & 201.3 & 191.5\end{array}$ $\begin{array}{lllll}194.7 & 252.2 & 272.4 & 221.3 & 818.8\end{array}$
$x$ series $1.0 \quad 2.0 \quad 3.0 \ldots \ldots \ldots \ldots . . .23 .0$
$\bar{x}=n(n+1) /(2 n)=12 \quad \bar{y}=837.5 \quad S_{x x}=1012 \quad S_{y y}=10018140$
$\hat{\beta}_{1}=-15.94 \hat{\beta}_{0}=1023.5 \quad r=-0.16 \quad r^{2}=0.0256 \hat{\sigma}^{2}=\operatorname{Var}[\varepsilon]=464813$.
The $95 \%$ confidence limits for $\beta_{1}$ are $(-60.5,28.6)$.
The sample $t$ statistic for the slope parameter $=-0.743$.
Therefore, the slope is not significant and there is no trend.
Reject NH of a trend in the precipitation
6.12. Extending flow records. The River Oba in western Nigeria has been gauged near Imo and a five year record is available. Also, sixty-year records are available for monthly rainfalls measured in the cities of Illorin and Ibadan with estimates of evaporation losses during the same period. It is proposed to extend the Imo flow record $Q$ by correlating with the Illorin (current $R$, antecedent $R A$ ) and the Ibadan (current $S$ ) residual rainfalls over the five-year period of flow observations. A multiple regression model is to be used. The following statistics are provided for the 4 monthly variables (in Imperial units):

| Variable | Mean | Standard Deviation |
| :---: | :---: | :---: |
| $R$ | 1.58 | 2.73 |
| $R A$ | 1.58 | 2.73 |
| $S$ | 1.45 | 2.40 |
| $Q$ | 55.75 | 111.68 |

The following are the sums of squares and cross-products of deviations from the mean:

|  | $\boldsymbol{R}$ | $\boldsymbol{R A}$ | $\boldsymbol{S}$ | $\boldsymbol{Q}$ |
| :--- | :---: | :---: | :---: | :---: |
| $\boldsymbol{R}$ | 441.49 | 160.04 | 197.80 | 9426.61 |
| $\boldsymbol{R} \boldsymbol{A}$ | 160.041 | 441.49 | 176.21 | 13005.22 |
| $\boldsymbol{S}$ | 197.80 | 176.21 | 340.16 | 10642.07 |
| $\boldsymbol{Q}$ | 9426.61 | 13005.22 | 10642.07 | 735890.35 |

1. Write the four normal equations from which the parameters are estimated.
2. If the variable $R A$ is not taken into account,
(a) estimate the parameters,
(b) estimate the standard error of estimate of Q from R and S , and
(c) estimate the coefficient of determination.

Solution. The regression equation for three explanatory variables is written as:
$Y_{i}=\beta_{0}+\beta_{1} x_{1 i}+\beta_{2} x_{2 i}+\beta_{3} x_{3 i}+\varepsilon_{i}$
In application we may use the following equation which includes the sample means of the three explanatory variables:
$y_{i}=a+b_{1}\left(x_{1 i}-\bar{x}_{1}\right)+b_{2}\left(x_{2 i}-\bar{x}_{2}\right)+b_{3}\left(x_{3 i}-\bar{x}_{3}\right)+\varepsilon_{i}$,
that is $\bar{x}_{1}=\left(\sum_{i=1}^{n} x_{1 i}\right) / n ; \bar{x}_{2}=\left(\sum_{i=1}^{n} x_{2 i}\right) / n$ and $\bar{x}_{3}=\left(\sum_{i=1}^{n} x_{3 i}\right) / n$.
We can relate the constants as follows:
$b_{i}=\beta_{i}$ for $i=1,2,3 ; \beta_{0}=a-\beta_{1} \bar{x}_{1}-\beta_{2} \bar{x}_{2}-\beta_{3} \bar{x}_{3}$.
(1a)
The sum of squared errors is given by
$S^{2}=\sum_{i=1}^{n} \varepsilon_{i}^{2}=\sum_{i=1}^{n}\left[y_{i}-a-\beta_{1}\left(x_{1 i}-\bar{x}_{1}\right)-\beta_{2}\left(x_{2 i}-\bar{x}_{2}\right)-\beta_{1}\left(x_{3 i}-\bar{x}_{3}\right)\right]^{2}$
Solving $\frac{\partial S^{2}}{\partial a}=0, \frac{\partial S^{2}}{\partial \beta_{1}}=0, \frac{\partial S^{2}}{\partial \beta_{2}}=0$ and $\frac{\partial S^{2}}{\partial \beta_{3}}=0$ and if $\sum$ represents $\sum_{i=1}^{n}$, we have
$\Sigma y_{i}-\hat{a} n-\hat{\beta}_{1} \Sigma\left(x_{1 i}-\bar{x}_{i}\right)-\hat{\beta}_{2} \Sigma\left(x_{2 i}-\bar{x}_{i}\right)-\hat{\beta}_{3} \Sigma\left(x_{3 i}-\bar{x}_{i}\right)=0$. That is,
$\hat{a}=\frac{1}{n} \Sigma y_{i}=\bar{y}$
$\Sigma\left(y_{i}-\bar{y}\right)\left(x_{1 i}-\bar{x}_{i}\right)=\hat{\beta}_{1} \Sigma\left(x_{1 i}-\bar{x}_{1}\right)^{2}+\hat{\beta}_{2} \Sigma\left(x_{2 i}-\bar{x}_{2}\right)\left(x_{1 i}-\bar{x}_{1}\right)+\hat{\beta}_{3} \Sigma\left(x_{3 i}-\bar{x}_{3}\right)\left(x_{1 i}-\bar{x}_{1}\right)$
$\Sigma\left(y_{i}-\bar{y}\right)\left(x_{2 i}-\bar{x}_{2}\right)=\hat{\beta}_{1} \Sigma\left(x_{1 i}-\bar{x}_{1}\right)\left(x_{2 i}-\bar{x}_{2}\right)+\hat{\beta}_{2} \Sigma\left(x_{2 i}-\bar{x}_{2}\right)^{2}+\hat{\beta}_{3} \Sigma\left(x_{3 i}-\bar{x}_{3}\right)\left(x_{2 i}-\bar{x}_{2}\right)$
$\Sigma\left(y_{i}-\bar{y}\right)\left(x_{3 i}-\bar{x}_{3}\right)=\hat{\beta}_{1} \Sigma\left(x_{1 i}-\bar{x}_{1}\right)\left(x_{3 i}-\bar{x}_{3}\right)+\hat{\beta}_{3} \Sigma\left(x_{3 i}-\bar{x}_{3}\right)^{2}+\hat{\beta}_{2} \Sigma\left(x_{3 i}-\bar{x}_{3}\right)\left(x_{2 i}-\bar{x}_{2}\right)$
In the tables $R \equiv x_{1} ; R A \equiv x_{2} ; S \equiv x_{3} ; Q \equiv \mathrm{y}$
We delete Eq.(3) and the terms involving $x_{2}$, and also $\hat{\beta}_{2}$, from Eqs.(2) and (4).
(a). Hence from Eqs.(2) and (4), $\hat{\beta}_{1}=9.92$ and $\hat{\beta}_{3}=25.51$. Also from Eq.(1b)
$\hat{a}=\bar{y}=55.75$ and, from Eq.(1a), $\hat{\beta}_{0}=3.09$.
(b). For the total sum of squares $S S_{y y}=\Sigma\left(y_{i}-\bar{y}\right)^{2}=735,890.35$. Also, $S S_{x_{1} y}=\Sigma\left(x_{1 i}-\bar{x}_{i}\right)\left(y_{i}-\bar{y}\right)$ and $S S_{x_{3} y}=\Sigma\left(x_{3 i}-\bar{x}_{i}\right)\left(y_{i}-\bar{y}\right)$.
Also the residual sum of squares $S S_{\varepsilon}=S S_{y y}-S S_{R}=S S_{y y}-\left[\hat{\beta}_{1} S S_{x_{1} y}+\hat{\beta}_{3} S S_{x_{3} y}\right]$ $S S_{\varepsilon}=735890.35-[9.92 \times 9426.61+25.51 \times 10642.07]=735890.35-364991.18=370899.17$

Hence $\hat{\sigma}=[370899.17 / 57]=80.67$
(c) $R^{2}=\frac{S S_{R}}{S S_{y y}}=\frac{364991.18}{735890.35}=0.50$
6.13. Salinity data The following are part of the salinity data reported by Rupport and Carroll (1980) for water during the spring season in Pamlico Sound, North Carolina, U.S.A.:

| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Salinity | 7.6 | 7.7 | 4.3 | 5.9 | 5.0 | 6.5 | 8.3 | 8.2 | 13.2 |  |
| Lagged Salinity | 8.2 | 7.6 | 4.6 | 4.3 | 5.9 | 5.0 | 6.5 | 8.3 | 10.1 |  |
| Discharge | 23.01 | 23.87 | 26.42 | 24.89 | 29.90 | 24.20 | 23.22 | 21.86 | 22.27 |  |
| Trend | 4 | 5 | 0 | 1 | 2 | 3 | 4 | 5 | 0 |  |
| Index | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |  |
| Salinity | 12.6 | 10.4 | 10.8 | 13.1 | 12.3 | 10.4 | 10.5 | 7.7 | 9.5 |  |
| Lagged Salinity | 13.2 | 12.6 | 10.4 | 10.8 | 13.1 | 13.3 | 10.4 | 10.5 | 7.7 |  |
| Discharge | 23.83 | 25.14 | 22.43 | 21.79 | 22.38 | 23.93 | 33.49 | 24.86 | 22.69 |  |
| Trend | 1 | 2 | 3 | 4 | 5 | 0 | 1 | 2 | 3 |  |
| Index | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 |
| Salinity | 12. | 12.6 | 13.6 | 14.1 | 13.5 | 11.5 | 12.0 | 13.0 | 14.1 | 15.1 |
| Lagged Salinity | 10.0 | 12.0 | 12.1 | 13.6 | 15.0 | 13.5 | 11.5 | 12.0 | 13.0 | 14.4 |
| Discharge | 21.79 | 22.04 | 21.03 | 21.01 | 25.87 | 26.29 | 22.93 | 21.31. | 20.77 | 21.39 |
| Trend | 0 | 1 | 4 | 4 | 0 | 1 | 2 | 3 | 4 | 5 |

The biweekly average salinity is given in milligrams per liter with salinity lagged two weeks, discharge in cubic millimeters per second, and trend as a dummy variable for the time period.
(1) Estimate the parameters for a linear model with salinity as the response variable.
(2) Determine the 'best' form of model.
(3) Determine Cook's distances and leverage measures
(4) Test for any outliers.
(5) Comment on the foregoing results (3 and 4).

Data used with permission from the Journal of the American Statistical Association, Copyright (1980) by the American Statistical Association. All rights reserved.

## Solution.

(a)The estimated salinity by the least squares method using the full data:
$\hat{s}=9.595+0.777 s_{l}-0.0256 d-0.295 t$,
where $\hat{s}$ denotes estimated salinity, $s_{l}$ denotes lagged salinity, $d$ denotes discharge and $t$ denotes trend.
(b), (c) and (d): Item 16 is an influential observation and an outlier as shown by

|  | Residual | Stud residual | Leverage $h_{i}$. | $h_{i} /\left(1-h_{i}\right)$ | Cook's $d_{i}$ |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Item 16 | 2.72 | 3.03 | 0.546 | 1.204 | 3.70 |

We revise the model by deleting item 16, This is given by $\hat{s}=18.49+0.697 s_{l}-0.157 d-0.6303 t$
For which $r^{2}=0.8934$.

|  | SS | Degrees of freedom | Mean SS | F |
| :--- | ---: | :---: | :---: | :---: |
|  |  |  |  |  |
| Regression | 218.56 | 3 | 72.85 | 64.23 |
| Residual | 26.09 | 23 | 1.13 |  |
| Total | 244.65 | 26 |  |  |

The $F$ value is clearly highly significant. Also, $\hat{\beta}^{\mathrm{T}} \mathrm{X}^{\mathrm{T}} \mathrm{y}=3226.89$
In addition, if we delete the trend component, we have the new model:
$\hat{s}=16.24+0.713 s_{l}-0.5587 d$
For which $r^{2}=0.8865$.

|  | SS | Degrees of freedom | Mean SS | F |
| :--- | ---: | :---: | :---: | :---: |
| Regression | 216.87 | 3 | 108.44 | 93.7 |
| Residual | 27.77 | 24 | 1.16 |  |
| Total | 244.65 | 26 |  |  |

The $F$ value is again highly significant. Also, $\hat{\beta}^{\mathrm{T}} \mathrm{X}^{\mathrm{T}} \mathrm{y}=3225.21$
Note that the difference in the above $\hat{\beta}^{\mathrm{T}} \mathrm{X}^{\mathrm{T}} \mathrm{y}$ values is not significant.
Suppose we adopt simple regression models, firstly, using only lagged salinity and then trend as the single explanatory variable

|  | $\hat{\beta}_{0}$ | $\hat{\beta}_{1}$ | $\hat{\beta}^{\mathrm{T}} \mathrm{X}^{\mathrm{T}} \mathrm{y}$ |
| :--- | ---: | ---: | ---: |
| Simple regression: $\hat{s}, s_{l}$ | 1.809 | 0.842 | 3194.1 |
| Simple regression: $\hat{s}, d$ | 32.63 | -0.944 | 3110.6 |

(b), (c), (d) and $(e)$, The difference between the first two $\hat{\beta}^{\mathrm{T}} \mathrm{X}^{\mathrm{T}} \mathrm{y}$ values, which are almost equal and those for the simple regression models are not insignificant. Therefore adopt a multiple regression model using the two explanatory variables: lagged salinity and trend Also Item 16 is deleted from the data. The 'best model' is given by
$\hat{s}=16.24+0.713 s_{l}-0.5587 d$
Final check on other observations

|  | Residual | Stud residual | Leverage $h_{i}$. | $h_{i} /\left(1-h_{i}\right)$ | Cook's $d_{i}$ |
| :--- | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |
| Item 5 | 1.246 | 1.511 | 0.412 | 0.7006 | 0.5330 |
| Item 9 | 2.192 | 2.091 | 0.0507 | 0.0534 | 0.0779 |

There are no problem items.
6.14. Hald cement data. The following is a part of the data reported by Hald (1952, p. 647) for the heat generated $H$ in calories per gram, during hardening, for a type of cement as a function of four additives. The table gives $H$ and four additives $A 1, A 2, A 3$, and $A 4$.

| Item | $\boldsymbol{H}$ | $\boldsymbol{A 1}$ | A2 | $\boldsymbol{A}$ 3 | $\boldsymbol{A 4}$ |
| :---: | ---: | ---: | ---: | ---: | ---: |
| 1 | 78.5 | 7 | 26 | 6 | 60 |
| 2 | 74.3 | 1 | 29 | 15 | 52 |
| 3 | 104.3 | 11 | 56 | 8 | 20 |
| 4 | 87.6 | 11 | 31 | 8 | 47 |
| 5 | 95.9 | 7 | 52 | 6 | 33 |
| 6 | 109.2 | 11 | 55 | 9 | 22 |
| 7 | 102.7 | 3 | 71 | 17 | 6 |
| 8 | 72.5 | 1 | 31 | 22 | 44 |
| 9 | 93.1 | 2 | 54 | 18 | 22 |
| 10 | 115.9 | 21 | 47 | 4 | 26 |
| 11 | 83.8 | 1 | 40 | 23 | 34 |
| 12 | 113.3 | 11 | 66 | 9 | 12 |
| 13 | 109.4 | 10 | 68 | 8 | 12 |

Complete a ridge analysis and write a predictive equation for $H$.
(Data used with the kind permission of the author.)

## Solution.

The data are standardized as in Example 6.17. See following figure INSERT FIGURE FOR PROBLEM 6.14
6.15. Principle components. The following covariance matrix $\mathbf{C}$ was computed in a study of catchment characteristics of a river basin.

$$
\left[\begin{array}{ccc}
3.67 & -4.93 & 2.08 \\
-4.93 & 98.1 & -3.01 \\
2.08 & -3.01 & 2.01
\end{array}\right] .
$$

Determine the eigenvalues and the eigenvectors, and comment on the results.

## Solution.

$\mathrm{C}=\left[\begin{array}{ccc}3.67 & -4.93 & 2.08 \\ -4.93 & 98.1 & -3.01 \\ 2.08 & -3.01 & 2.01\end{array}\right] . \quad \mathrm{C}-\lambda=\left[\begin{array}{ccc}3.67-\lambda & -4.93 & 2.08 \\ -4.93 & 98.1-\lambda & -3.01 \\ 2.08 & -3.01 & 2.01-\lambda\end{array}\right]$
$|\mathrm{C}-\lambda \mathrm{I}|=(3.67-\lambda)(98.1-\lambda)(2.01-\lambda)-3.01^{2}(3.67-\lambda)-4.93^{2}(2.01-\lambda)+4.93 \times 3.01 \times 2.08+$ $2.08 \times 4.93 \times 3.01-2.08^{2}(98.1-\lambda)=0$
The eigen values are $\lambda_{1}=98.4572752, \lambda_{2}=4.723437$ and $\lambda_{3}=0.5996811$.

$$
\left[\begin{array}{ccc}
3.67-98.46 & -4.93 & 2.08 \\
-4.93 & 98.1-98.46 & -3.01 \\
2.08 & =3.01 & 2.01-98.46
\end{array}\right]\left[\begin{array}{l}
a_{11} \\
a_{21} \\
a_{31}
\end{array}\right]=0
$$

Also $a_{11}^{2}+a_{21}^{2}+a_{31}^{2}=1$. Hence $a_{11}=0.05262043, a_{21}=-0.9980926$, $a_{31}=0.03228406$.
Proceeding further
$A=\left[\begin{array}{ccc}0.05262043 & 0.82395523 & 0.564206404 \\ -0.99809260 & 0.06165978 & 0.003039957 \\ 0.03228406 & 0.56329020 & -0.825628180\end{array}\right]$
6.16. Nitrates in river. For the data given in Table E.6.2, where water quality is given at $1-\mathrm{km}$ intervals, draw a semivariogram of the nitrate values, $h=1,2, \ldots, 30$. What type of model is suggested?

## Solution.

From Table E.6.2 of Appendix E a partial list of semivariogram values for various distances in km are given by

| Distance | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Semivar, | 3.728 | 3.270 | 3.049 | 3.278 | 3.227 | 3.019 | 3.644 | 3.780 | 4.055 | 4.321 |


| Distance | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| :--- | :---: | :---: | ---: | ---: | ---: | ---: | ---: | ---: | :---: | :---: |
| Semivar, | 4.743 | 3.468 | 4.98 | 4.12 | 4.80 | 4.48 | 4.29 | 5.37 | 5.273 | 5.373 |

Solution. Sketch of semivariogram vs. distance in km


An exponential model is suggested for the semivariogram.
6.17. Salinity of groundwater. The following salinity observations were recorded in 25 wells in a coastal aquifer, in milligrams per liter The wells are spaced at distances of approximately 1 km in the NS and EW. Directions.

| 10.5 | 9.3 | 10.4 | 9.1 | 10.0 |
| ---: | :---: | :---: | :---: | :---: |
| 9.2 | 10.1 | 11.1 | 10.2 | 10.3 |
| 11.2 | 10.8 | 10.2 | 11.5 | 11.5 |
| 10.9 | 9.5 | 11.5 | 11.0 | 12.0 |
| 11.1 | 10.5 | 11.0 | 10.7 | 12.5 |

1. Determine the semivariogram under conditions of isotropy.
2. Determine the semivariograms in the NE-SW and NW-SE directions
3. Comment on the results

## Solution.

The differences observed in salinity for various distances apart of the wells and directions are given below.
$\mathbf{1} \mathbf{~ k m}$ [total $n=40]$ :-
$\leftrightarrow 1.21 .11 .30 .9 ; 0.91 .00 .90 .1 ; 0.40 .61 .30 ; 1.42 .00 .51 .0 ; 0.60 .50 .31 .8 ;$
$\uparrow 1.32 .00 .30 .2 ; 0.80 .71 .31 .0 ; 0.70 .91 .30 .5 ; 1.11 .30 .50 .3 ; 0.31 .20 .50 .5$;
$\mathbf{2} \mathbf{~ k m}$ [total $n=30]$ :-
$\leftrightarrow 0.10 .20 .3 ; 1.90 .10 .8 ; 1.00 .71 .3 ; 0.61 .50 .5 ; 0.10 .21 .5$;
§ $0.71 .70 .1 ; 1.50 .60 .3 ; 0.20 .40 .8 ; 2.40 .8$ 0.8; 1.5 1.7 1.0;
$3 \mathbf{~ k m}$ [total $n=20]$ :-
$\leftrightarrow 1.40 .7 ; 1.00 .2 ; 0.30 .7 ; 0.12 .5 ; 0.42 .0$
ई 0.4 1.9; 0.2 0.4; 1.1 0.1; $1.90 .5 ; 2.02 .2$
4 km [total $n=10]:-\leftrightarrow 0.5,1.10 .31 .11 .4 ; \downarrow 0.61 .20 .61 .62 .5$;
NE $\sqrt{2} \mathbf{k m}[t \operatorname{talal} n=16] .11 .1 .3 .1 .32 .01 .6 .70 .00 .21 .00 .01 .20 .01 .20 .0 .51 .3$

SW $\sqrt{18} \mathbf{k m}[$ total $n=4] \begin{array}{lllll}1.8 & 0.9 & 0.5 & 0.2\end{array}$ $\sqrt{32} \mathbf{~ k m}[$ total $n=1] 1.0$

NW $\quad \sqrt{2} \mathbf{k m}[$ total $n=16] 1.2 .21 .31 .8 .4 .5 .4 .1$. 8 1.3 1.6.7.7.8 1.71 .5 . 4

$\mathbf{S E} \sqrt{18} \mathbf{k m}\left[\begin{array}{lllll}\text { total } n=4] & 2.7 & 0.5 & 2.2 & 1.5\end{array}\right.$ $\sqrt{32} \mathbf{k m}[$ total $n=1] 2.0$
(a)Semovariogram for vertical and horizontal directions


Distance in km $\rightarrow$
(b)Semovariogram for NE-SW directions

(c)Semovariogram for NW-SE directions


Distance in $\mathrm{km} \rightarrow$
The approximate linear relationships suggest that isotropy is a reasonable assumption.
6.18. Linear semivariogram model. The following data are used for the semivariogram of the grouped data of Fig. 6.4.3. Fit a linear model by regression.

| Item | $\mathbf{k m}$ | Semivariogram |
| :--- | :--- | :--- |
| 1 | 15.8 | 43423.3 |
| 2 | 27.2 | 43886.0 |
| 3 | 36.0 | 55604.7 |
| 4 | 44.0 | 51576.5 |
| 5 | 52.2 | 55397.1 |
| 6 | 61.6 | 65899.6 |
| 7 | 72.4 | 65089.6 |
| 8 | 85.6 | 73292.7 |
| 9 | 103.2 | 67044.6 |
| 10 | 131.9 | 68610.7 |
| 11 | 166.1 | 88218.9 |

## Solution.

Please see diagram of semivariogram
INSERT FIGURE FOR PROBLEM 1.18
6.19. Exponential semivariogram model. The following data are used for the semivariogram of the annual rainfall data of Fig. 6.4.1. Fit an exponential model stating any assumptions made. If one takes account of the results from Example 6.21, what can be said about the properties of the annual rainfall in the region?

| Item | km | Semivariogram | Item | km | Semivariogram |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | 8.3 | 37851.9 | 47 | 58.6 | 64391.7 |
| 2 | 11.7 | 35326.8 | 48 | 59.8 | 68915.5 |
| 3 | 13.6 | 68326.4 | 49 | 60.4 | 102468.4 |
| 4 | 14.9 | 48446.3 | 50 | 61.0 | 49045.3 |
| 5 | 16.1 | 31541.6 | 51 | 62.1 | 41186.5 |
| 6 | 17.5 | 41053.2 | 52 | 63.3 | 49151.1 |
| 7 | 18.7 | 49047.8 | 53 | 65.2 | 79739.4 |
| 8 | 20.2 | 33951.9 | 54 | 66.5 | 66704.0 |
| 9 | 21.4 | 45263.4 | 55 | 67.6 | 71922.7 |
| 10 | 22.5 | 53088.1 | 56 | 69.0 | 52255.0 |
| 11 | 23.8 | 34547.4 | 57 | 70.0 | 94147.4 |
| 12 | 25.1 | 51042.2 | 58 | 71.1 | 64181.3 |
| 13 | 26.1 | 39733.1 | 59 | 72.6 | 62752.7 |
| 14 | 27.0 | 49519.7 | 60 | 74.0 | 56775.6 |
| 15 | 28.2 | 32489.5 | 61 | 74.9 | 49858.3 |
| 16 | 29.5 | 55755.0 | 62 | 75.6 | 75493.2 |
| 17 | 30.8 | 47523.9 | 63 | 77.0 | 58419.7 |
| 18 | 31.6 | 31275.0 | 64 | 78.6 | 48015.3 |
| 19 | 32.2 | 56937.5 | 65 | 80.8 | 81717.9 |
| 20 | 32.9 | 51891.9 | 66 | 82.4 | 84134.3 |
| 21 | 33.9 | 51271.0 | 67 | 84.2 | 59251.8 |
| 22 | 35.0 | 56113.3 | 68 | 85.8 | 68939.1 |
| 23 | 36.2 | 53660.4 | 69 | 87.3 | 89176.8 |
| 24 | 37.1 | 77947.4 | 70 | 88.6 | 63794.1 |
| 25 | 37.9 | 38722.0 | 71 | 90.6 | 66877.4 |
| 26 | 38.8 | 53830.4 | 72 | 92.3 | 97726.8 |
| 27 | 39.9 | 60068.5 | 73 | 94.1 | 56009.0 |
| 28 | 41.0 | 62769.7 | 74 | 96.9 | 51560.0 |
| 29 | 41.9 | 61731.6 | 75 | 98.4 | 72857.7 |
| 30 | 42.7 | 36311.1 | 76 | 100.6 | 58023.1 |
| 31 | 43.6 | 47268.9 | 77 | 102.4 | 81231.9 |
| 32 | 44.1 | 27706.2 | 78 | 104.9 | 51767.3 |
| 33 | 44.6 | 75630.6 | 79 | 107.5 | 88162.3 |
| 34 | 45.3 | 40477.0 | 80 | 110.6 | 60517.2 |
| 35 | 46.1 | 64279.0 | 81 | 113.5 | 83272.4 |
| 36 | 47.0 | 48014.5 | 82 | 116.1 | 62780.6 |
| 37 | 47.9 | 48837.3 | 83 | 119.1 | 72207.9 |
| 38 | 48.7 | 55357.7 | 84 | 122.0 | 88796.7 |
| 39 | 50.3 | 56927.1. | 85 | 126.3 | 71400.8 |
| 40 | 51.8 | 70026.0 | 86 | 130.2 | 58951.1 |
| 41 | 52.5 | 49162.2 | 87 | 134.3 | 81787.8 |
| 42 | 53.4 | 43975.7 | 88 | 138.9 | 75341.3 |
| 43 | 54.4 | 38393.8 | 89 | 144.9 | 57169.7 |
| 44 | 55.0 | 72487.2 | 90 | 155.1 | 49060.9 |
| 45 | 55.7 | 63407.0 | 91 | 166.1 | 88218.9 |
| 46 | 57.1 | 71495.0 |  |  |  |

## Solution.

Please see diagram of semivariogram
INSERT FIGURE FOR PROBLEM 1.18
6.20. Exponential model. Repeat the Kriging analysis of the groundwater quality example of Example 6.22 replacing the linear model by an exponential model stating any assumption made. Estimate the value at K . Comment on the models and the results.

## Solution.

The exponential model and substitution from given data
$F(h)=A_{0} \delta(h)+\varpi[1-\exp (-h / a)]$
$0.5+\varpi[1-\exp (-2 / 2)]=4.5$
Hence $\varpi=6.3279$
$\hat{\gamma}(h)=0.5+\varpi[1-\exp (-h / 2)]$
$\hat{\gamma}(0.5)=0.5+\varpi[1-\exp (-0.5 / 2)]=1.90$
$\hat{\gamma}(1.0)=0.5+\varpi[1-\exp (-1.0 / 2)]=2.99$
$\hat{\gamma}(1.12)=0.5+\varpi[1-\exp (-1.12 / 2)]=3.21$
$\lambda+0 \times \lambda_{1}+1.9 \times \lambda_{2}+2.99 \times \lambda_{3}=3.21$
$\lambda+1.9 \times \lambda_{1}+0 \times \lambda_{2}+3.21 \times \lambda_{3}=2.99$
$\lambda+2.99 \times \lambda_{1}+3.21 \times \lambda_{2}+0 \times \lambda_{3}=1.90$
$\lambda_{1}+\lambda_{2}+\lambda_{3}=1.00$, Hence $\lambda_{1}=0.084 ; \lambda_{2}=0.274 ; \lambda_{3}=0.642 ; \lambda=0.769$
Hence $\hat{z}_{k}=0.084 \times 12+0.274 \times 15+0.642 \times 10=11.54 \mathrm{mg} / \mathrm{L}$

Blank page

# Applied Statistics for Civil and Environmental Engineers <br> Problem Solution Manual <br> by N.T. Kottegoda and R. Rosso 

## Chapter 7 - Frequency Analysis of Extreme Events

7.1. Observed frequency of maximum annual storm rainfall data. Consider the 58 -year data of maximum annual hourly storm depth reported in Table E.7.1.
(a) Find the expected frequency of nonexceedance of the largest recorded value.
(b) Find the theoretical probability of nonexceedance of this value resulting from the fitted GEV distribution.
(c) Compute the plotting positions of the observations using the equation $p_{i}=(i-0.35) / n$ where $i$ is the rank in increasing order and $n$ is the number of items of data. Compare the observed frequency estimates with these expected frequencies and with the theoretical ones.

## Solution

(a) $\mathrm{P}[\mathrm{X} \leq 128.50]=\frac{n}{n+1}=\frac{58}{59}=0.983$
(b) If $X \sim \operatorname{GEV}(\kappa, \alpha, \varepsilon)$

To estimate the shape parameter of the GEV distribution by the method of moments the sampling skewness is used in Eq. (7.2.62), which is then solved for k by numerical iteration. Thus,
$\mu_{x}=48.16 \mathrm{~mm}$

$$
\sigma_{x}=23.76 \mathrm{~mm}
$$

$\mathrm{k}=-0.05$.
Then, from Eq. (7.2.63),
$\alpha=\sqrt{\frac{k^{2} \sigma_{X}^{2}}{\Gamma(1+2 k)-\Gamma^{2}(1+k)}}=\sqrt{\frac{0.0029 \times 564.33}{\Gamma(1-0.107)-\Gamma^{2}(1-0.05)}}=17.17 \mathrm{~mm}$,
from Eq. (7.2.64),
$\varepsilon=\mu_{x}-\frac{\alpha}{k}[1-\Gamma(1+k)]=48.16+\frac{17.17}{0.05}[1-\Gamma(1-0.05)]=37.29 \mathrm{~mm}$
and the theoretical probability of non-exceedence of the largest value is

$$
F_{x_{\max }}=\exp \left\{-\left[1-\frac{k(x-\varepsilon}{\alpha}\right]^{1 / k}\right\}=\exp \left\{-\left[1-\frac{-0.05 \times(128.50-37.29}{17.17}\right]^{1 /-0.05}\right\}=0.991
$$

(c)


Figure 7.S1
7.2. Hurst effect in hydrologic data. Using rescaled range analysis shown in Example 7.8 compute the Hurst exponents for annual rainfall and runoff in the Po river at Pontelagoscuro, Italy (see data in Table E.7.2). Use equispaced values of $\ln n$.

## Solution

0.63 (rainfall); 0.72 (runoff)
7.3. Minimum flight delay. An airport is designed to receive $n$ daily flight arrivals. Find the mean and variance of the expected minimum delay if the interarrival time $X$ is a shifted exponentially distributed variate with scale parameter $\lambda$ and location parameter $x_{0}$.

Solution. If the interarrival time between two successive flights, $X$, is shifted exponentially distributed

$$
F_{X}(x)=1-e^{-\lambda\left(x-x_{0}\right)}
$$

the expected minimum delay is given by Eq. (7.1.13) that is

$$
\begin{aligned}
& E\left[X_{\mathrm{man}}\right]=n \int_{x_{0}}^{\infty} x\left[1-1+e^{-\lambda\left(x-x_{0}\right)}\right]^{(n-1)} \lambda e^{-\lambda\left(x-x_{0}\right)} d x=\int_{x_{0}}^{\infty} n \lambda x e^{-n \dot{\lambda}\left(x-x_{0}\right)} d x \\
& E\left[X_{\mathrm{man}}\right]=-\left[x e^{-n \lambda\left(x-x_{0}\right)}\right]_{x_{0}}^{-x}-\frac{1}{n \lambda} \int_{x_{0}}^{\infty} d e^{-n \lambda\left(x-x_{0}\right)}=x_{0}+\frac{1}{n \lambda}
\end{aligned}
$$

while the variance of the minimum delay can be determined from Eq. (7.1.14), that is

$$
\begin{aligned}
& \operatorname{Var}\left[X_{\mathrm{mn}}\right]=\int_{x_{0}}^{-\infty}\left(x-E\left[X_{\mathrm{min}}\right]\right)^{2} f_{X}(x) d x=\int_{x_{0}}^{-x}\left(x-x_{0}+\frac{1}{n \lambda}\right)^{2} \lambda e^{-\lambda\left(x-x_{0}\right)} d x \\
& \operatorname{Var}\left[X_{\mathrm{min}}\right]=\left(x_{0}^{2}+2 \frac{x_{0}}{\lambda}+\frac{2}{\lambda^{2}}\right)-2 E\left[X_{\mathrm{min}}\left[x_{0}+\frac{1}{\lambda}\right]+E^{2}\left[X_{\mathrm{min}}\right] .\right.
\end{aligned}
$$

7.4. Flood Discharge. Consider the data of maximum annual flood flows in the Tevere river at Ripetta, Italy, reported in Table E. 5.8. Compute the 100-year flood discharge using (a) the Gumbel distribution, (b) the GEV distribution, $(c)$ the lognormal distribution, and $(d)$ the gamma distribution. Perform a goodness-of-fit test using the chi-squared, Kolgomorov-Smirnov and Anderson-Darling tests. Consider $\alpha$ $=0.10$

Solution. From data available in the Tevere river at Ripetta station, it is possible to estimate the mean, the standard deviation and the skewness of maximum annual flood flows, $X$
$\mu_{x}=1149.41 \mathrm{~m}^{3} / \mathrm{s}$
$\sigma_{x}=486.89 \mathrm{~m}^{3} / \mathrm{s}$
$\gamma_{1 x}=0.614$
The method of moments is used to estimate the parameters of four extreme value distributions so
(1) if $X_{\text {max }} \sim \operatorname{Gumbel}(\alpha, \mathrm{b})$

The Gumbel parameters can be calculated using respectively Eq. (7.2.21) and Eq.(7.2.22).
$\alpha=\frac{\sqrt{6}}{\pi} \sigma_{x}=379.63 \mathrm{~m}^{3} / \mathrm{s}, \quad \mathrm{b}=\mu_{x}-\mathrm{n}_{\mathrm{e}} \alpha=930.29 \mathrm{~m}^{3} / \mathrm{s}$.
For a return period $\mathrm{T}=100$ years, the required design value is given from Eq. (7.2.26)

$$
\begin{aligned}
& \xi_{\max }(100)=\mu-\sigma_{x}\left\{\frac{\sqrt{6}}{\pi}\left[n_{e}+\ln \left(\ln \frac{T}{T-1}\right)\right]\right\}=1149.41-486.89\left\{\frac{\sqrt{6}}{\pi}\left[0.5772+\ln \left(\ln \frac{1}{0.99}\right)\right]\right\} \\
& \xi_{\max }(100)=2676.62 \mathrm{~m}^{3} / \mathrm{s}
\end{aligned}
$$

(2) If $X_{\text {max }} \sim \operatorname{GEV}(\kappa, \alpha, \varepsilon)$

To estimate the shape parameter of the GEV distribution by the method of moments the sampling skewness is used in Eq. (7.2.62), which is then solved for $k$ by numerical iteration. Thus,
$\mathrm{k}=0.11$.
Then, from Eq. (7.2.63),
$\alpha=\sqrt{\frac{k^{2} \sigma_{V}^{2}}{\Gamma(1+2 k)-\Gamma^{2}(1+k)}}=\sqrt{\frac{0.0112 \times 237060}{\Gamma(1+0.21)-\Gamma^{2}(1+0.11)}}=427.74 \mathrm{~m}^{3} / \mathrm{s}$,
and from Eq. (7.2.64),
$\varepsilon=\mu_{x}-\frac{\alpha}{k}[1-\Gamma(1+k)]=1149.41-\frac{427}{0.11}[1-\Gamma(1+0.11)]=943.32 \mathrm{~m}^{3} / \mathrm{s}$
The Eq.(7.2.66) gives the 100 -year flood discharge for GEV distribution
$\xi_{\max }(100)=\varepsilon+\frac{\alpha}{k}\left[1-\left(\ln \left(\frac{T}{T-1}\right)\right)^{k}\right]=943.32+\frac{427.74}{0.11}\left[1-\left(\ln \left(\frac{1}{0.99}\right)\right)^{0.11}\right]=2501.70 \mathrm{~m}^{3} / \mathrm{s}$
(3) If $X_{\text {max }} \sim \operatorname{LN}\left(\mu_{\operatorname{tr}(.) 7}, \sigma_{\ln (.)}^{2}\right)$
$\sigma_{\operatorname{In}(.1)}^{2}=\ln \left(1+V_{N}^{2}\right)=\ln \left(1+0.424^{2}\right)=0.165$
and
$\mu_{\ln (.)}=\ln \left(\mu_{x}\right)-0.5 \sigma_{\ln (.)}^{2}=\ln (1149)-0.5 \times 0.165=6.964$.
The design value corresponding to the return period of 100 years is calculated using Eq.(7.2.81)
$\xi_{\max }(100)=\exp \left(\mu_{\ln (X)}+\Phi_{0.99}^{-1} \sigma_{\ln (\lambda)}\right)=\exp (6.964+2.326 \times 0.406)=2723.17 \mathrm{~m}^{3} / \mathrm{s}$
where $\Phi_{0.99}^{-1}$ is the 0.99 -th quantile of the standard normal variate.
(4) If $X_{\text {max }} \sim \operatorname{Gamma}(\lambda, \mathrm{r})$
$\mathrm{r}=4 \gamma_{1 . x}^{-2}=5.57$
$\lambda=\frac{\sigma_{x} \gamma_{1 x}}{2}=206.25 \mathrm{~m}^{3} / \mathrm{s}$
The 100 -th quantile is obtained from Eq. (7.2.87) using the corresponding quantile of the standard gamma variate $G_{0.99}^{-1}$

$$
\xi_{\max }(100)=\lambda G_{0.99}^{-1}(r)=206.25 \times 12.472=2572.31 \mathrm{~m}^{3} / \mathrm{s}
$$

For a formal assessment of these fits we can use the $\chi^{2}$, Kolmogorov-Smirnov and Anderson-Darling goodness-of -fit of these distributions to observed data.

In these tests :
the null hypothesis $H_{0}$ : the random variable, $X_{\max }$, has a (*) distribution,
where (*) is one of the four distributions considered.
The alternate hypothesis $H_{l}$ : the random variable does not have the specified distribution.
The level of significance: $\alpha=0.10$.
$\chi^{2}$ test
The critical region: $X^{2} \geq \chi_{v, a=0.10}^{2}$
Fixed $1=6$ classes,

Table 7.S4.1- $\chi^{2}$ test

| Classes | Observed | Gumbel | GEV | LN2 | GA2 | Gumbel | GEV | LN2 | GA2 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Classes | N(i) | npi | npi | npi | npi | $\left(\mathrm{Ni}\right.$-npi) ${ }^{2}($ (npi) | $\left(\mathrm{Ni}\right.$-npi) ${ }^{2} /(\mathrm{npi})$ | $(\mathrm{Ni} \text {-npi) })^{2}(\mathrm{npi})$ | $\left(\mathrm{Ni}\right.$-npi) ${ }^{2} /($ npi) |
| 0-664 | 9 | 7.2 | 8.2 | 6.8 | 8.1 | 0.457 | 0.078 | 0.726 | 0.110 |
| 664-843 | 9 | 8.2 | 7.1 | 8.8 | 7.6 | 0.088 | 0.527 | 0.007 | 0.240 |
| 843-1132 | 9 | 14.7 | 13.3 | 15.0 | 13.6 | 2.185 | 1.372 | 2.408 | 1.540 |
| 1132-1355 | 9 | 9.0 | 9.0 | 8.8 | 8.8 | 0.000 | 0.000 | 0.005 | 0.005 |
| 1355-1553 | 9 | 5.5 | 6.0 | 5.3 | 5.8 | 2.177 | 1.449 | 2.505 | 1.792 |
| $>1553$ | 9 | 9.5 | 10.4 | 9.3 | 10.1 | 0.028 | 0.182 | 0.011 | 0.128 |
|  |  |  | Degrees 1 |  |  | 3 | 2 | 3 | 3 |
|  |  |  |  | $\mathrm{X}^{2}$ |  | 4.935 | 3.608 | 5.663 | 3.814 |
|  |  |  | $\chi^{2} 1_{1_{a}=0.10}$ |  |  | 6.251 | 4.605 | 6.251 | 6.251 |
|  |  |  |  |  |  | unrejected | unrejected | unrejected | unrejected |

From table 7.S4.1,
(1) if $X_{\text {max }} \sim \operatorname{Gumbel}(\alpha, \mathrm{b}), \mathrm{k}=2, v=6-2-1=3$ and $\chi_{3, a=0.10}^{2}=6.251$.
$\mathrm{X}^{2}=4.935 \leq \chi_{3, \alpha=0.10}^{2}=6.251$, the null hypothesis is not rejected.
(2) if $X_{\text {max }} \sim \operatorname{GEV}(\kappa, \alpha, \varepsilon), \mathrm{k}=3, v=6-3-1=2$ and $\chi_{2, \alpha=0.10}^{2}=4.605$
$\mathrm{X}^{2}=3.608 \leq \chi_{2, \alpha=0.10}^{2}=4.605$, the nuil hypothesis is not rejected.
(3) If $X_{\max } \sim \mathrm{LN}\left(\mu_{\ln (.) 7}, \sigma_{\ln (X)}^{2}\right), \mathrm{k}=2, v=6-2-1=3$ and $\chi_{3 . a=0.10}^{2}=6.251$.
$\mathrm{X}^{2}=5.663 \leq \chi_{3 ., 40.0 .10}^{2}=6.251$, the null hypothesis is not rejected.
(4) If $X_{\text {max }} \sim \operatorname{Gamma}(\lambda, \mathrm{r}), \mathrm{k}=2, v=6-2-1=3$ and $\chi_{3, \alpha=0.10}^{2}=6.251$.
$\mathrm{X}^{2}=3.814 \leq \chi_{3, a=0.10}^{2}=6.251$, the null hypothesis is not rejected.

Kolmogorov-Smirnov test.
Critical region: $D_{n} \geq D_{54, a=0.10}=0.166$
From Table 7.S4.2 it is shown that the null hypothesis is not rejected for all the four distributions.

Table 7.S4.2 - Kolmogorov-Smirnov test.

|  | i/n | X | Gumbel | GEV | LN2 | GA2 | Gumbel | GEV | LN2 | GA2 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Rank | $\mathrm{F}_{0}(\mathrm{x})$ | asc. order | $\mathrm{F}_{\mathrm{n}}(\mathrm{x})$ | $\mathrm{F}_{\mathrm{n}}(\mathrm{x})$ | $\mathrm{F}_{\mathrm{n}}(\mathrm{x})$ | $\mathrm{F}_{\mathrm{n}}(\mathrm{x})$ | $\mid \mathrm{F}_{\mathrm{a}}(\mathrm{x})$ - $\mathrm{F}_{\mathrm{o}}(\mathrm{x}) \mid$ | $\left\|\mathrm{F}_{0}(\mathrm{x})-\mathrm{F}_{0}(\mathrm{x})\right\|$ | $\mathrm{F}_{\mathrm{a}}(\mathrm{x})$ - $\mathrm{F}_{0}(\mathrm{x}) \mid$ | $\mathrm{F}_{3}(\mathrm{x})-\mathrm{F}_{0}(\mathrm{x}) \mid$ |
| i |  |  |  |  |  |  |  |  | 0.017 | 0.015 |
| 1 | 0.019 | 259 | 0.003 | 0.012 | 0.002 | 0.004 | 0.016 | 0.006 | 0.017 | 0.015 |
| 2 | 0.037 | 355 | 0.011 | 0.027 | 0.011 | 0.015 | 0.026 | 0.010 | 0.026 | 0.022 |
| 3 | 0.056 | 468 | 0.034 | 0.057 | 0.044 | 0.045 | 0.021 | 0.002 | 0.011 | 0.011 |
| 4 | 0.074 | 472 | 0.035 | 0.059 | 0.046 | 0.046 | 0.039 | 0.015 | 0.028 | 0.028 |
| 5 | 0.093 | 510 | 0.049 | 0.073 | 0.064 | 0.062 | 0.044 | 0.020 | 0.028 | 0.031 |
| 6 | 0.111 | 528 | 0.056 | 0.080 | 0.074 | 0.070 | 0.055 | 0.031 | 0.037 | 0.041 |
| 7 | 0.130 | 612 | 0.099 | 0.122 | 0.129 | 0.115 | 0.031 | 0.008 | 0.001 | 0.014 |
| 8 | 0.148 | 622 | 0.105 | 0.127 | 0.136 | 0.122 | 0.043 | 0.021 | 0.012 | 0.027 |
| 9 | 0.167 | 664 | 0.133 | 0.153 | 0.169 | 0.149 | 0.034 | 0.014 | 0.002 | 0.017 |
| 10 | 0.185 | 714 | 0.171 | 0.186 | 0.211 | 0.185 | 0.014 | 0.000 | 0.025 | 0.000 |
| 11 | 0.204 | 717 | 0.173 | 0.188 | 0.213 | 0.188 | 0.031 | 0.016 | 0.009 | 0.016 |
| 12 | 0.222 | 735 | 0.188 | 0.200 | 0.229 | 0.202 | 0.034 | 0.022 | 0.007 | 0.021 |
| 13 | 0.241 | 743 | 0.194 | 0.206 | 0.236 | 0.208 | 0.046 | 0.035 | 0.005 | 0.033 |
| 14 | 0.259 | 775 | 0.222 | 0.230 | 0.264 | 0.234 | 0.037 | 0.029 | 0.005 | 0.026 |
| 15 | 0.278 | 794 | 0.239 | 0.244 | 0.281 | 0.249 | 0.039 | 0.033 | 0.004 | 0.029 |
| 16 | 0.296 | 810 | 0.253 | 0.257 | 0.296 | 0.263 | 0.043 | 0.039 | 0.000 | 0.034 |
| 17 | 0.315 | 822 | 0.264 | 0.266 | 0.307 | 0.273 | 0.050 | 0.048 | 0.008 | 0.042 |
| 18 | 0.333 | 843 | 0.284 | 0.283 | 0.326 | 0.291 | 0.049 | 0.050 | 0.007 | 0.042 |
| 19 | 0.352 | 861 | 0.301 | 0.298 | 0.342 | 0.306 | 0.051 | 0.054 | 0.010 | 0.045 |
| 20 | 0.370 | 896 | 0.335 | 0.327 | 0.374 | 0.337 | 0.036 | 0.043 | 0.004 | 0.033 |
| 21 | 0.389 | 935 | 0.372 | 0.361 | 0.409 | 0.372 | 0.016 | 0.028 | 0.020 | 0.017 |
| 22 | 0.407 | 950 | 0.387 | 0.374 | 0.422 | 0.385 | 0.020 | 0.034 | 0.015 | 0.022 |
| 23 | 0.426 | 985 | 0.421 | 0.404 | 0.452 | 0.416 | 0.005 | 0.022 | 0.026 | 0.010 |
| 24 | 0.444 | 1083 | 0.512 | 0.488 | 0.533 | 0.501 | 0.068 | 0.044 | 0.088 | 0.057 |
| 25 | 0.463 | 1092 | 0.520 | 0.496 | 0.540 | 0.509 | 0.057 | 0.033 | 0.077 | 0.046 |
| 26 | 0.481 | 1099 | 0.527 | 0.502 | 0.545 | 0.515 | 0.045 | 0.020 | 0.064 | 0.033 |
| 27 | 0.500 | 1132 | 0.556 | 0.529 | 0.570 | 0.542 | 0.056 | 0.029 | 0.070 | 0.042 |
| 28 | 0.519 | 1166 | 0.584 | 0.557 | 0.594 | 0.570 | 0.066 | 0.038 | 0.076 | 0.051 |
| 29 | 0.537 | 1230 | 0.635 | 0.607 | 0.638 | 0.619 | 0.098 | 0.070 | 0.101 | 0.082 |
| 30 | 0.556 | 1240 | 0.643 | 0.615 | 0.644 | 0.626 | 0.087 | 0.059 | 0.089 | 0.071 |
| 31 | 0.574 | 1270 | 0.665 | 0.637 | 0.663 | 0.648 | 0.090 | 0.063 | 0.089 | 0.074 |
| 32 | 0.593 | 1290 | 0.679 | 0.651 | 0.675 | 0.662 | 0.086 | 0.059 | 0.083 | 0.069 |
| 33 | 0.611 | 1325 | 0.702 | 0.676 | 0.695 | 0.686 | 0.091 | 0.065 | 0.084 | 0.075 |
| 34 | 0.630 | 1340 | 0.712 | 0.686 | 0.704 | 0.696 | 0.082 | 0.056 | 0.074 | 0.066 |
| 35 | 0.648 | 1346 | 0.716 | 0.690 | 0.707 | 0.699 | 0.068 | 0.042 | 0.059 | 0.051 |
| 36 | 0.667 | 1355 | 0.721 | 0.696 | 0.712 | 0.705 | 0.055 | 0.029 | 0.045 | 0.038 |
| 37 | 0.685 | 1370 | 0.731 | 0.706 | 0.720 | 0.715 | 0.045 | 0.021 | 0.035 | 0.029 |
| 38 | 0.704 | 1370 | 0.731 | 0.706 | 0.720 | 0.715 | 0.027 | 0.002 | 0.016 | 0.011 |
| 39 | 0.722 | 1380 | 0.736 | 0.712 | 0.725 | 0.721 | 0.014 | 0.010 | 0.003 | 0.002 |
| 40 | 0.741 | 1440 | 0.770 | 0.748 | 0.754 | 0.756 | 0.029 | 0.008 | 0.014 | 0.015 |
| 41 | 0.759 | 1440 | 0.770 | 0.748 | 0.754 | 0.756 | 0.011 | 0.011 | 0.005 | 0.004 |
| 42 | 0.778 | 1460 | 0.781 | 0.760 | 0.763 | 0.766 | 0.003 | 0.018 | 0.014 | 0.011 |
| 43 | 0.796 | 1508 | 0.804 | 0.786 | 0.784 | 0.791 | 0.008 | 0.011 | 0.012 | 0.005 |
| 44 | 0.815 | 1540 | 0.818 | 0.802 | 0.797 | 0.806 | 0.003 | 0.013 | 0.018 | 0.009 |
| 45 | 0.833 | 1553 | 0.824 | 0.808 | 0.802 | 0.812 | 0.010 | 0.025 | 0.031 | 0.021 |
| 46 | 0.852 | 1600 | 0.843 | 0.829 | 0.819 | 0.832 | 0.009 | 0.023 | 0.033 | 0.019 |
| 47 | 0.870 | 1600 | 0.843 | 0.829 | 0.819 | 0.832 | 0.028 | 0.041 | 0.051 | 0.038 |
| 48 | 0.889 | 1621 | 0.850 | 0.838 | 0.826 | 0.841 | 0.039 | 0.051 | 0.062 | 0.048 |
| 49 | 0.907 | 1690 | 0.874 | 0.865 | 0.848 | 0.866 | 0.034 | 0.043 | 0.059 | 0.041 |
| 50 | 0.926 | 1696 | 0.875 | 0.867 | 0.850 | 0.868 | 0.051 | 0.059 | 0.076 | 0.058 |
| 51 | 0.944 | 1876 | 0.921 | 0.920 | 0.895 | 0.918 | 0.024 | 0.025 | 0.050 | 0.026 |
| 52 | 0.963 | 1966 | 0.937 | 0.938 | 0.912 | 0.937 | 0.026 | 0.025 | 0.051 | 0.026 |
| 53 | 0.981 | 2190 | 0.964 | 0.970 | 0.943 | 0.967 | 0.017 | 0.012 | 0.038 | 0.015 |
| 54 | 1.000 | 2730 | 0.991 | 0.996 | 0.980 | 0.994 | 0.009 | 0.004 | 0.020 | 0.006 |
|  |  |  |  |  |  |  | 2 0.098 | 0.070 | 0.101 | 0.082 |

## Anderson-Darling test.

Critical region: $\mathrm{A}^{2} \geq \mathrm{A}^{2}{ }_{\alpha=0.05}=2.492$
From Table 7.S4.2, $\mathrm{A}^{2} \leq \mathrm{A}^{2}{ }_{\alpha=0.05}$ is checked for all the distribution, so also in this case the null hypothesis is not rejected.
Table 7.S4.3 - Anderson-Darling test.

7.5. Depth-duration-frequency curves of storm rainfall. Consider the statistical summaries of Table 7.3.1 for the annual maximum storm depth for various durations observed at Lanzada, Italy. The estimated $L$-moments for the normalized annual maximum storm depth (extreme value data divided for various durations divided by the corresponding mean for the specified duration) are $L_{1}=1, L_{2}=0.1330$, and $L_{3}=0.0182$, respectively.
(a) Compute the parameters of the Gumbel distribution by the method of L-moments and compare this distribution with that estimated in Example 7.34 by the method of moments.
(b) Compute the parameters of the GEV distribution by the methods of moments and $L$-moments, and compare these distributions with the Gumbel model on a Gumbel probability plot.
(c) Find the depth-duration-frequency curve for a return period of 100 years using the GEV model estimated by the method of $L$-moments.

## Solution

(a) The Gumbel parameters estimated with $L$-moments are
$\alpha=\frac{L_{2}}{\ln (2)}=\frac{0.133}{\ln (2)}=0.192$
$b=L_{1}-n_{e} \alpha=1-0.5772 \times 0.192=0.889$.
It is seen that the $L$-moments estimates are very close to those estimated by the method of moments $\alpha=0.191, b=0.89$ reported in Illustration E7.33.
(b) Method of moments.

The shape parameters of the GEV distribution, k , is estimated by the sampling skewness in Eq. (7.2.62), that is solved by numerical approximations.
$\mathrm{k}=0.033$
Then, from Eq. (7.2.63),
$\alpha=\sqrt{\frac{k^{2} \sigma_{1}^{2}}{\Gamma(1+2 k)-\Gamma^{2}(1+k)}}=\sqrt{\frac{0.033 \times 0.060}{\Gamma(1+0.066)-\Gamma^{2}(1+0.033)}}=0.199$,
and from Eq. (7.2.64),
$\varepsilon=\mu_{w_{0}}-\frac{\alpha}{k}[1-\Gamma(1+k)]=1-\frac{0.199}{0.033}[1-\Gamma(1+0.033)]=0.891$.
where $\mu_{\mathrm{x}}$ and $\sigma^{2} \mathrm{x}$ - are respectly the mean and variance of the normalised annual maximum storm depth.
Method of $L$-moments.

$$
\begin{aligned}
k & =7.8590\left(\frac{2 L_{2}}{L_{3}+3 L_{2}}-\frac{\ln (2)}{\ln (3)}\right)+2.9554\left(\frac{2 L_{2}}{L_{3}+3 L_{2}}-\frac{\ln (2)}{\ln (3)}\right)^{2} \\
& =7.8590\left(\frac{2 \times 0.133}{0.018+3 \times 0.133}-\frac{\ln (2)}{\ln (3)}\right)+2.9554\left(\frac{2 \times 0.133}{0.018+3 \times 0.133}-\frac{\ln (2)}{\ln (3)}\right)^{2}=0.052
\end{aligned}
$$

then, the evaluation of $\alpha$ is obtained from Eq. (7.2.67b) as
$\alpha=\frac{k L_{2}}{\left(1-2^{-k}\right) \Gamma(1+k)}=\frac{0.052 \times 0.133}{\left(1-2^{-0.052}\right) \Gamma(1+0.052)}=0.201$
and the location parameter, $\varepsilon$, is found using Eq. (7.2.67b) that is

$$
\varepsilon=L_{1}-\frac{\alpha}{k}[1-\Gamma(1+k)]=1-\frac{0.201}{0.052}[1-\Gamma(1+0.052)]=0.894 \text {. }
$$

The comparation between the Gumbel and GEV distributions is reported in Figure 7.S5.1 .


Figure 7.S5.1
(c) Using the scaling model, the depth-duration-frequency curve for a return period of 100 years, can be determined, if the variate $\mathrm{X}_{\max }(100, \mathrm{t})$ is a GEV distributed by Eq. (7.3.5)

$$
x_{\max }(100, t)=\left[\varepsilon+\frac{\alpha}{\mathrm{K}}\left(1-e^{-\gamma_{0}, p}\right)\right] \mu_{1} t^{n}=\left[0.894+\frac{0.201}{0.052}\left(1-e^{-0.052 \times 4.600}\right)\right] 14.0 \times t^{0.457}=23.01 t^{0.457} \mathrm{~mm}
$$

where the scaling parameters, $\mu_{1}$ and $n$ are determined in illustration (E7.33)
7.6. Dry spells. A period of days on which no rainfall is experienced continuously is called a dry run if preceded and succeeded by one or more wet days. As shown in Example 4.16, the run length $X$ of a dry spell can be modeled as a log-series distributed variate. Suppose that $X$ has a mean of 5 days, so the estimated $p$ is 0.07 , and the number of dry spells in a year is a Poisson variate with a mean of 40 . Find the cdf of the annual maximum run length of a dry spell. Compute the return period of a dry spell 60 days long.

Solution. Since the run length X of a dry spell can be modeled as a logarithmic series distributed variate, its $p m f$ and $c d f$ are respectively

$$
f_{x}(x)=-\frac{(1-p)^{x}}{x \ln p} \quad F_{x}(x)=\sum_{u=1}^{x}-\frac{(1-p)^{u}}{u \ln p}
$$

then the $c d f$ of maximum run length is

$$
\left.\left.F_{X_{-\infty}}(x)=e^{-v\left[1-F_{x}(x)\right]}=e^{\left\{-v\left[1+\frac{\sum}{w-1} \frac{\left.(1-p)^{*}\right]}{u \ln p}\right]\right\}}=e^{\left\{-40\left[1-0.376 \sum_{i=1}^{0.93^{*}} u\right.\right.}\right]\right\}
$$

So the return period of a dry spell of 60 days is calculated in the following manner

$$
\begin{aligned}
& \left.F_{X_{-}}(60)=e^{\left\{-40\left[1-0.376 \sum_{i=1}^{00.93^{*}}\right]\right.}\right\} \\
& T=\frac{1}{1-F_{X_{\text {ma }}}(60)}=\frac{1}{1-0.949} \approx 20 \text { years } .
\end{aligned}
$$

7.7. Highest sea wave in a storm. The highest sea wave $X$ in a storm is modeled as a Rayleighdistributed variate with $\operatorname{pdf} f_{X}(x)=\left(x / \lambda^{2}\right) \exp \left[-(x / \lambda)^{2} / 2\right]$. Suppose that parameter $\lambda$ varies randomly from one storm to another, and assume that the number of storms in a year with $\lambda \geq \lambda_{0}$ is a Poisson-distributed variate with mean $v$. Find the cdf of the annual maximum sea wave height, $X_{\max }$, if $\lambda-\lambda_{0}$ is an exponentially distributed variate with parameter $\alpha$.

Solution. If the number of events in a year with $\lambda \geq \lambda_{0}, N \sim \operatorname{Poisson}(v), p_{N}(n)=\frac{v^{n} e^{-v}}{n!}$
if the parameter $\lambda-\lambda_{0}$ is exponentially distributed $f_{\Lambda}(\lambda)=\alpha \mathrm{e}^{-\alpha(\lambda-\lambda .0)}$,
and if the highest sea wave X is modeled with a Rayleigh distribution, $\mathrm{f}_{\mathrm{x}}(\mathrm{x})=\left(\mathrm{x} / \lambda^{2}\right) \exp \left[-(\mathrm{x} / \lambda)^{2} / 2\right]$, the $c d f$ of the annual maximum sea wave height, $\mathrm{X}_{\text {max }}$, is
$F_{X_{-\infty}}(x)=\int_{\lambda_{0}}^{+\infty} f_{\Lambda}(\lambda)\left\{\sum_{n=0}^{+\infty}\left[F_{X}(x)\right]^{n} p_{N}(n)\right\} d \lambda=\int_{\lambda_{0}}^{\infty} f_{\Lambda}(\lambda)\left\{e^{-v\left(1-E_{x}(x)\right]}\right\} d \lambda$
$F_{x_{\pi m}}(x)=\int_{\lambda_{0}}^{-x} \alpha e^{-\alpha\left(\lambda-\lambda_{0}\right)} e^{--e^{-\frac{1}{2}\left(\frac{1}{\lambda}\right)^{2}}} d \lambda$
7.8. Overflooding. The annual maximum flood discharge $X_{\max }$ at a given river site is a Gumbeldistributed variate with parameters $\alpha=625 \mathrm{~m}^{3} / \mathrm{s}$ and $b=1152 \mathrm{~m}^{3} / \mathrm{s}$. The overflooding volume $Y$ during a flood with peak discharge $X$ exceeding a value of $\quad \varepsilon=2000 \mathrm{~m}^{3} / \mathrm{s}$ is modeled as an exponentially distributed variate with mean $c(X / \varepsilon)^{\beta}$, where $c=5 \times 10^{6} \mathrm{~m}^{3}$ and $\beta=0.5$. Assume that $N \sim \operatorname{Poisson}(v)$ is the number flood events in a year with peak discharge exceeding $\varepsilon$ with $v=2.1$. Find the cdf of the annual maximum overflooding volume, $Y_{\max }$, and compute the 100 -year overflooding volume.

## Solution

The peak discharge X exceeding a threshold $\varepsilon$ of $2000 \mathrm{~m}^{3} / \mathrm{s}$ is shifted exponentially distributed with the parameter $\lambda$ that can be determined by the parameters of the annual maximum flood discharge variate.
Since $\mathrm{X}_{\max } \sim \operatorname{Gumbel}(\alpha, \mathrm{b})$, and $F_{X_{\max }}(x)=\left[F_{Y_{1}}(x)\right]^{n}=\left[1-e^{-\lambda\left(x-x_{0}\right)}\right]^{n}=\exp \left[-e^{-\frac{x-x_{0}-b}{\alpha}}\right]$,
$\lambda=1 / \alpha=1 / 625=0.0016\left(\mathrm{~m}^{3} / \mathrm{s}\right)^{-1}$.
Then the mean of X variate is
$E[X]=\varepsilon+1 / \lambda=2000+1 / 0.0016=2625 \mathrm{~m}^{3} / \mathrm{s}$,
and the mean of Y is
$E[Y]=c\left(\frac{E[X]}{\varepsilon}\right)^{\beta}=5 \times 10^{6}\left(\frac{2625}{2000}\right)^{0.5}=5728220 \mathrm{~m}^{3}$.
The overflooding volume Y is exponentially distributed with parameter $\lambda^{\circ}=1 / E[Y]=1.75 \times 10^{-7} \mathrm{~m}^{-3}$ and $c d f$

$$
F_{\gamma}(y)=1-e^{-\lambda^{\cdot} y}=1-e^{-1.75 \times 10^{-7} y},
$$

so the maximum overflooding volume $\mathrm{Y}_{\max }$ presents the following distribution

$$
F_{Y_{\max }}(y)=e^{-v\left[1-F_{Y}(y)\right]}=\exp \left[-v e^{-\lambda y}\right]=\exp \left[-e^{-\lambda^{*}\left(y-\lambda^{*-1} \ln v\right)}\right]=\exp \left[-e^{-1.75 \times 10^{-7}(y-5728220 \ln 2.1)}\right],
$$

that is a Gumbel distribution with shape parameter $\alpha^{*}=1 / \lambda^{*}$ and location $b^{*}=\lambda^{-1} \ln v$.
$\alpha^{*}=1 / \lambda^{*}=5728220 \mathrm{~m}^{3}$
$b^{*}=\lambda^{-1} \ln v=4249980 \mathrm{~m}^{3}$.
The 100 -year overflooding volume is
$y_{\max }(100)=b^{*}-\alpha^{*} \ln \left(\frac{T}{T-1}\right)=23.8 \times 10^{6} \mathrm{~m}^{3}$.
7.9. Sea waves. Consider the data set of simulated highest sea waves above a threshold in the upper Adriatic sea of Problem 1.23. Find the 10 -year design wave height resulting from the data set obtained by using calibration strategy no.1, and that for calibration strategy no. 2 (assume Poisson events and the shifted exponential distribution to fit the data). Compare these values with that obtained from the analysis of observed data reported in Example 7.41.

Solution. Considering the data sets of the highest sea waves in upper Adriatic sea X , obtained from the simulations no. 1 and no.2 given in Problem 1.23, and assuming these ones are shifted exponential distributed, $f_{x}(x)=\lambda e^{-\lambda\left(x-x_{0}\right)}$, the corresponding parameters, $\lambda$ and $\mathrm{x}_{0}$, can be derived from the mean and the variance in the following manner
$\mu_{\mathrm{X}}=\mathrm{x}_{0}+\frac{1}{\lambda}$
$\sigma^{2} x=\frac{1}{\lambda^{2}}$

$$
\begin{array}{ll}
\text { strategy no.1 } & \text { strategy no.2 } \\
\mu_{\mathrm{x}}=3.144 \mathrm{~m} & \mu_{\mathrm{x}}=2.698 \mathrm{~m} \\
\sigma_{\mathrm{x}}^{2}=0.797 \mathrm{~m}^{2} & \sigma^{2}=0.617 \mathrm{~m}^{2} \\
\lambda=1 / \sigma_{\mathrm{x}}=1 / \sqrt{0.797}=1.120 \mathrm{~m}^{-1} & \lambda=1 / \sqrt{0.617}=1.273 \mathrm{~m}^{-1} \\
x_{0}=\mu_{\mathrm{x}}-1 / \lambda=3.144-1 / 1.120=2.252 \mathrm{~m} & x_{0}=2.698-1 / 1.273=1.912 \mathrm{~m}
\end{array}
$$

Assuming a Poissonian process of occurrences, $\mathrm{N} \sim$ Poisson( $v$ ), the mean number of storm events above the threshold of 2 m in a year is obtained as
strategy no. 1

$$
\text { strategy no. } 2
$$

$v=(14 / 13) \times 12=12.9$

$$
v=(12 / 13) \times 12=11.1
$$

The $c d f$ of maximum annual sea wave is a Gumbel distribution with parameters $\alpha=1 / \lambda$ and $\mathrm{b}=\mathrm{x}_{0}+\frac{1}{\lambda} \ln v$,
strategy no. 1
$\alpha=0.893 \mathrm{~m}$

$$
\begin{aligned}
& \text { strategy no. } 2 \\
& \alpha=0.786 \mathrm{~m}
\end{aligned}
$$

$\mathrm{b}=4.536 \mathrm{~m}$

So for $\mathrm{T}=10$ years the design value is estimated fro the two strategies as
strategy no. 1
$x_{\max }(10)=4.536-0.893 \ln \left(\frac{1}{0.9}\right)=6.545 \mathrm{~m} \quad x_{\max }(10)=3.801-0.786 \ln \left(\frac{1}{0.9}\right)=5.569 \mathrm{~m}$.

$$
\text { strategy no. } 2
$$

$$
x_{\max }(10)=3.801-0.786 \ln \left(\frac{1}{0.9}\right)=5.569 \mathrm{~m} .
$$

These values are superior to one estimated in Illustration (E7.40).
7.10. Maximum annual wind speed predictions. Find the probability distribution of maximum annual wind speed and the 50 -year wind velocity for $(a)$ Cagliari and (b) Pantelleria, Italy, by fitting the Gumbel distribution to the extreme value data shown in Table E.7.3 using the method of moments.

Solution. From the data set reported in Table E7.3 of maximum annual wind speed for Cagliari and Pantelleria, we compute the mean and standard deviation
(a) Cagliari
(b) Pantelleria
$\mu_{\mathrm{x}}=17.82 \mathrm{~m} / \mathrm{s}$

$$
\mu_{\mathrm{x}}=26.82 \mathrm{~m} / \mathrm{s}
$$

$\sigma_{\mathrm{x}}=1.88 \mathrm{~m} / \mathrm{s}$

$$
\sigma_{\mathrm{X}}=4.08 \mathrm{~m} / \mathrm{s} .
$$

Then the parameters of their Gumbel distributions are,

$$
\begin{array}{ll}
\alpha=\sigma_{x} \frac{\sqrt{6}}{\pi}=1.88 \frac{\sqrt{6}}{\pi}=1.46 \mathrm{~m} / \mathrm{s} & \alpha==4.08 \frac{\sqrt{6}}{\pi}=3.18 \mathrm{~m} / \mathrm{s} \\
\mathrm{~b}=\mu_{x}-n_{e} \alpha=17.82-0.5772 \times 1.46=16.97 \mathrm{~m} / \mathrm{s} & \mathrm{~b}=26.82-0.5772 \times 3.18=24.98 \mathrm{~m} / \mathrm{s}
\end{array}
$$

and the 50 -year quantiles are,
(a) Cagliari , $x_{\max }(50)=b-\alpha \ln \left(\frac{T}{T-1}\right)=16.97-1.46 \ln \left(\frac{50}{49}\right)=22.68 \mathrm{~m} / \mathrm{s}$
(b) Pantelleria, $x_{\max }(50)=24.98-3.18 \ln \left(\frac{50}{49}\right)=37.38 \mathrm{~m} / \mathrm{s}$.
7.11. Rescaling of wind speed estimates. The maximum annual wind speed $X(t, z)$ averaged over a period of length $t$ that is recorded at ground elevation $z$ in a particular site with roughness length $z_{0}$ scales as

$$
X(\lambda t, \eta z)=X(t, z)\left[1+0.98 c(\lambda) / \ln \left(z / z_{0}\right)\right] \ln \left(\eta z / z_{0}\right) / \ln \left(z / z_{0}\right)
$$

where $t<1,0<\lambda<1$ and $\eta$ are two scaling factors, with $c(\lambda)$ denoting a scaling function, $c(1)=0$. Suppose that the available records are averaged for 5 minutes, say, $\lambda=1 / 12$, for $t$ in hours, with $c(1 / 12)=$ 0.54 , the gauging station (site $A$ ) is located at elevation of $z=8 \mathrm{~m}$ in open terrain with $z_{0}=0.01 \mathrm{~m}$, and the sample mean and standard deviation of the annual maximum data are $15 \mathrm{~m} / \mathrm{s}$ and $3 \mathrm{~m} / \mathrm{s}$, respectively. Reference wind speed $X_{A}$ is taken as the 10 -minute average wind speed at a standard elevation of 10 m , that is, $X_{A}=X(1 / 6$ hours, 10 m$)$, with $\eta=1.25, \lambda=1 / 6$, and $c(1 / 6)=0.36$.
(a) Find the cdf of annual maximum reference wind speed for site $A$ using the Gumbel distribution and the method of moments. To design a building located in the downtown area, one must determine the 10 -minute average wind speed at a ground elevation of 50 m knowing that the roughness length for this location (site $B$ ) is 0.3 m . The vertical profile of wind velocity is given by $2.5 u_{*} \ln \left(z / z_{0}\right)$, where $u_{*}$ denotes the friction velocity, which scales as $u_{*_{\mathrm{B}}} / u_{*_{\mathrm{A}}}=\left(z_{0 \mathrm{~B}} / z_{0 \mathrm{~A}}\right)^{\gamma}$ for two sites $A$ and $B$ with different roughness length (see Figure 7.P1). Therefore, one must rescale the reference wind speed as

$$
X_{B}=X\left(t, z_{B}\right)=\frac{X\left(t, z_{A}\right)}{\ln \left(z_{A} / z_{0 A}\right)}\left(\frac{z_{0 B}}{z_{0 A}}\right)^{\gamma} \ln \left(z_{B} / z_{0 B}\right)
$$

with $t=1 / 6$ hours and $\gamma=0.07$.

(b) Find the cdf of annual maximum wind speed at $z_{B}=50 \mathrm{~m}$ for site $B$ using the Gumbel distribution and the method of moments.
(c) Compute the 50 -year design wind speed.

## Solution

(a) $\mathrm{E}[X(1 / 12$ hours, 8 m$)]=13.90 \mathrm{~m} / \mathrm{s}$
$\operatorname{Stdev}[X(1 / 12$ hours, 8 m$)]=2.78 \mathrm{~m} / \mathrm{s}$
$\mathrm{E}\left[X_{A}\right]=\mathrm{E}[X(1 / 6$ hours, 10 m$)]=$

$$
\mathrm{E}[X(1 / 12 \text { hours } 8 \mathrm{~m})][1+0.98 c(1 / 12) / \ln (8 / 0.01)] \ln (10 / 0.01) / 1 \ln (8 / 0.01)=15.12 \mathrm{~m} / \mathrm{s},
$$

$\operatorname{Stdev}\left[X_{A}\right]=\operatorname{Stdev}[X(1 / 6$ hours, 10 m$)]=$
$=\operatorname{Stdev}[X(1 / 12$ hours. 8 m$)][1+0.98 c(1 / 12) / \ln (8 / 0.01)] \ln (10 / 0.01) / 1 \ln (8 / 0.01)=3.02 \mathrm{~m} / \mathrm{s}$,
Then the Gumbel parameters, calculated with method of moments, are
$\alpha=\frac{\sqrt{6}}{\pi} \sigma_{x_{d}}=2.36 \mathrm{~m} / \mathrm{s}$,

$$
\mathrm{b}=\mu_{x_{d}}-\mathrm{n}_{\mathrm{e}} \alpha=13.76 \mathrm{~m} / \mathrm{s}
$$

(b) To determine the $c d f$ of annual maximum wind speed at $z_{\mathrm{B}}=50 \mathrm{~m}$ for site B we can use these scale relations to estimate mean and standard deviation:
$E\left[X_{B}\right]=E[X(1 / 6$ hours, $50 \mathrm{~m})]=\frac{E[X(1 / 6 \text { hour }, 10 \mathrm{~m})]}{\ln (10 \mathrm{~m} / 0.01)}\left(\frac{0.3}{0.01}\right)^{0.07} \ln (50 / 0.3)=14.21 \mathrm{~m} / \mathrm{s}$
$\operatorname{Stdev}\left[X_{B}\right]=\operatorname{Stdev}[X(1 / 6$ hours, $50 \mathrm{~m})]=\frac{\operatorname{Stdev}[X(1 / 6 \text { hour }, 10 \mathrm{~m})]}{\ln (10 \mathrm{~m} / 0.01)}\left(\frac{0.3}{0.01}\right)^{0.07} \ln (50 / 0.3)=2.84 \mathrm{~m} / \mathrm{s}$
So the Gumbel parameters are
$\alpha=\frac{\sqrt{6}}{\pi} \sigma_{x_{B}}=2.22 \mathrm{~m} / \mathrm{s}, \quad \mathrm{b}=\mu_{x_{A}}-\mathrm{n}_{\mathrm{e}} \alpha=12.93 \mathrm{~m} / \mathrm{s}$
(c) The 50 -year design wind speed in site $B$ is

$$
x_{\max . B}(50)=14.21-2.84\left\{\frac{\sqrt{6}}{\pi}\left[0.5772+\ln \left(\ln \frac{50}{49}\right)\right]\right\}=19.47 \mathrm{~m} / \mathrm{s}
$$

7.12. Wind speed prediction for Milan Park Tower. The 110-m Park Tower in Milan, Italy, is a steel tower built in 1933 on the occasion of the Fifth Triennial Decorative Arts Exhibition. From 1951 to 1973 the following twelve thunderstorms were recorded by an anemometer located at the elevation of 108 m with 10-minute average wind speed $X$ exceeding the critical mean velocity of $20.18 \mathrm{~m} / \mathrm{s}$.

| Date | Wind Direction, ${ }^{\circ} \mathbf{N}$ | Average Velocity, m/s |
| :---: | :---: | :---: |
| $29 / 04 / 53$ | 15 | 20.86 |
| $08 / 01 / 58$ | 135 | 20.98 |
| $05 / 01 / 59$ | 315 | 25.06 |
| $08 / 01 / 59$ | 315 | 25.06 |
| $09 / 01 / 59$ | 315 | 22.19 |
| $20 / 04 / 59$ | 45 | 22.17 |
| $28 / 07 / 59$ | 345 | 20.26 |
| $10 / 02 / 61$ | 315 | 21.48 |
| $12 / 02 / 61$ | 315 | 27.21 |
| $03 / 04 / 71$ | 315 | 21.48 |
| $20 / 11 / 71$ | 315 | 21.48 |
| $15 / 12 / 73$ | 345 | 26.34 |

Find the cdf of the annual maximum 10-minute average wind speed if the probability distribution of $X$ is (a) exponential, and (b) Pareto, and compare the corresponding extreme values for a return period of 50 years. Use a Gumbel probability plot to compare these results with $X_{\max } \sim \operatorname{Gumbel}(2.89 \mathrm{~m} / \mathrm{s}, 15.79 \mathrm{~m} / \mathrm{s})$ which is obtained by rescaling the extreme value data recorded at the Forlanini Airport station [Ballio, G., F. Mamberini, and G. Solari (1992) A 60-Year-Old 100-m-High Steel Tower: Limit States Under Wind Action, J. Wind Engin. and Industrial Aerodynamics, ASCE, Vol. 41-44, pp. 2089-2100].

## Solution

(a) X~Exponential $\left(\lambda, \mathrm{x}_{0}\right), \lambda=0.370(\mathrm{~m} / \mathrm{s})^{-1}, \mathrm{x}_{0}=20.18 \mathrm{~m} / \mathrm{s}$

The mean of occurrences above the threshold in a year is $v=0.521$.
$\mathrm{X}_{\max } \sim$ Gumbel $(\alpha, b), \alpha=1 / \lambda=2.70 \mathrm{~m} / \mathrm{s}, \mathrm{b}=\mathrm{x}_{0}+(1 / \lambda) \ln \nu=18.42 \mathrm{~m} / \mathrm{s}$
$x_{\text {max }}(50)=b-\alpha \ln \left(\frac{T}{T-1}\right)=18.42-2.79 \ln \left(\frac{50}{49}\right)=28.96 \mathrm{~m} / \mathrm{s}$
(b) $\mathrm{X} \sim$ Pareto $\left(\mathrm{x}_{0}, \theta\right), \mathrm{x}_{0}=20.18 \mathrm{~m} / \mathrm{s}, \theta=8.47$
$X_{\max } \sim$ Frèchet $\left(x_{0}{ }^{*}, \theta^{*}\right), x_{0}{ }^{\circ}=x_{0} v^{\frac{1}{\theta}}=18.69 \mathrm{~m} / \mathrm{s}, \theta^{*}=\theta=8.47$.
$x_{\max }=x_{0} \exp \left(\frac{y}{\theta}\right)=18.69 \times \exp \left(\frac{3.902}{8.47}\right)=29.62 \mathrm{~m} / \mathrm{s}$
These distributions are compared with the Gumbel $(2.89 \mathrm{~m} / \mathrm{s}, 15.79 \mathrm{~m} / \mathrm{s})$ which is obtained by rescaling the extreme value recorded at the Forlanini Airport station in Figure 7.S12.


Figurre 7.S12
7.13. Annual maximum wind speed in Pisa. Consider the following data set of 41 annual maximum 10minute average wind speeds at Pisa Airport, Italy.

| year | 1951 | 1952 | 1953 | 1954 | 1955 | 1956 | 1957 | 1958 | 1959 | 1960 | 1961 | 1962 | 1963 | 1964 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\mathrm{~m} / \mathrm{s}$ | 15.43 | 15.43 | 15.37 | 15.43 | 22.03 | 18.52 | 16.46 | 18.00 | 19.55 | 19.03 | 18.00 | 19.41 | 18.52 | 13.89 |
| year | 1965 | 1966 | 1967 | 1968 | 1969 | 1970 | 1971 | 1972 | 1973 | 1974 | 1975 | 1976 | 1977 | 1978 |
| $\mathrm{~m} / \mathrm{s}$ | 18 | 14.40 | 16.46 | 14.40 | 16.46 | 13.43 | 21.50 | 13.37 | 15.43 | 20.58 | 11.32 | 15.43 | 11.32 | 14.40 |
| year | 1979 | 1980 | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 |  |
| m/s | 16.46 | 16.98 | 13.59 | 22.63 | 15.95 | 13.89 | 13.89 | 19.05 | 13.89 | 13.89 | 16.98 | 19.95 | 12.33 |  |

Use the Kolmogorov-Smirnov and Anderson-Darling goodness-of-fit tests to compare the observed and theoretical cumulative frequencies as predicted by the (a) Gumbel, (b) Fréchet, (c) lognormal, (d) gamma, (e) GEV, (f) shifted-lognormal, and (g) shifted-gamma distributions. Discuss the decision of rejecting the null hypothesis when applicable. Consider $\alpha=0.01$ and 0.05 .

Solution. Using the method of moments we calculate the parameters of the considered distributions
(1) $X_{\text {max }} \sim \operatorname{Gumbel}(\alpha, \mathrm{b})=\operatorname{Gumbel}(2.2,15.1)$
(2) $X_{\text {max }} \sim \operatorname{Frèchet}\left(\mathrm{x}_{0}, \theta\right)=\operatorname{Frechet}(14.0,4.89)$
(3) $X_{\max } \sim \operatorname{Lognormal}\left(\mu_{\ln (.) 7}, \sigma_{\ln (.) 7}^{2}\right)=\operatorname{Lognormal}(2.78,0.17)$
(4) $X_{\text {max }} \sim \operatorname{Gamma}(\lambda, \mathrm{r})=\operatorname{Gamma}(0.49,33.45)$
(5) $X_{\text {max }} \sim G E V(\kappa, \alpha, \varepsilon)=G E V(0.31,2.87,15.40)$
(6) $X_{\max } \sim \operatorname{Shifted}-\operatorname{lognormal}\left(\mu_{\operatorname{tn}(. X-a)}, \sigma_{\ln (. X-a)}^{2}, a\right)=\operatorname{Shifted-lognormal}(3.14,0.121,-7.01)$
(7) $X_{\text {max }} \sim \operatorname{Shifted-gamma}(\lambda, \mathrm{r}, \mathrm{a})=\operatorname{Shifted}$-gamma $(0.52,30.03,0.86)$

For a formal assessment of these fits we can use the Kolmogorov-Smirnov and Anderson-Darling goodness-of -fit of these distributions to observed data.
In these tests :
the null hypothesis $H_{0}$ : the random variable, $X_{\max }$, has a (*) distribution,
where $\left(^{*}\right)$ is one of the seven distributions considered.
The alternate hypothesis $H_{t}$ : the random variable does not have the specified distribution.
Considering the following levels of significance: $\alpha=0.01, \alpha=0.05$
Kolmogorov-Smimov test.
Critical region: $\mathrm{D}_{\mathrm{n}} \geq \mathrm{D}_{41, \alpha=0.01}=0.254$
Critical region: $\mathrm{D}_{\mathrm{n}} \geq \mathrm{D}_{41, \alpha=0.05}=0.212$
From Table 7.S13.1 it is shown that the null hypothesis is not rejected for all the four distributions.

Anderson-Darling test.
Critical region: $\mathrm{A}^{2} \geq \mathrm{A}^{2}{ }_{\alpha=0.01}=3.857$
Critical region: $\mathrm{A}^{2} \geq \mathrm{A}^{2}{ }_{\alpha=0.05}=2.492$
From Table 7.S4.2, $\mathrm{A}^{2} \leq \mathrm{A}^{2}{ }_{\alpha=0.01}$ or $\mathrm{A}^{2} \leq \mathrm{A}^{2}{ }_{\alpha=0.05}$ is checked for all the distributions, so also in this case the null hypothesis is not rejected.


Table 7.S13.1-Kolmogorov-Smirnov test.

| Gumbel | Frechet | LN2 | GA2 | GEV | LN3 | GA3 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Summ. term. | Summ. term. | Summ. term. | Summ. term. | Summ. term | mm. term. | Summ. term. |
| 0.219 | 0.127 | 0.186 | 0.184 | 0.197 | 0.184 | 0.184 |
| 0.636 | 0.373 | 0.531 | 0.522 | 0.533 | 0.522 | 0.523 |
| 0.784 | 0.490 | 0.714 | 0.713 | 0.729 | 0.714 | 0.713 |
| 0.804 | 0.548 | 0.771 | 0.774 | 0.779 | 0.775 | 0.774 |
| 0.962 | 0.668 | 0.912 | 0.910 | 0.898 | 0.913 | 0.910 |
| 1.090 | 0.774 | 1.034 | 1.030 | 1.010 | 1.033 | 1.031 |
| 1.189 | 0.867 | 1.141 | 1.140 | 1.120 | 1.143 | 1.140 |
| 1.317 | 0.971 | 1.256 | 1.251 | 1.221 | 1.255 | 1.251 |
| 1.489 | 1.098 | 1.419 | 1.414 | 1.380 | 1.418 | 1.414 |
| 1.567 | 1.173 | 1.482 | 1.472 | 1.428 | 1.477 | 1.473 |
| 1.732 | 1.297 | 1.638 | 1.627 | 1.578 | 1.632 | 1.628 |
| 1.580 | 1.258 | 1.516 | 1.509 | 1.470 | 1.513 | 1.509 |
| 1.718 | 1.367 | 1.648 | 1.640 | 1.598 | 1.644 | 1.641 |
| 1.855 | 1.477 | 1.780 | 1.771 | 1.726 | 1.776 | 1.772 |
| 1.374 | 1.249 | 1.345 | 1.342 | 1.326 | 1.344 | 1.342 |
| 1.451 | 1.326 | 1.421 | 1.418 | 1.401 | 1.420 | 1.418 |
| 1.396 | 1.313 | 1.377 | 1.376 | 1.369 | 1.378 | 1.376 |
| 1.481 | 1.393 | 1.461 | 1.460 | 1.451 | 1.461 | 1.460 |
| 1.566 | 1.472 | 1.544 | 1.543 | 1.534 | 1.544 | 1.543 |
| 1.650 | 1.552 | 1.628 | 1.626 | 1.617 | 1.628 | 1.626 |
| 1.387 | 1.419 | 1.389 | 1.392 | 1.401 | 1.391 | 1.391 |
| 1.143 | 1.284 | 1.166 | 1.172 | 1.196 | 1.171 | 1.172 |
| 1.196 | 1.343 | 1.221 | 1.227 | 1.252 | 1.225 | 1.226 |
| 1.249 | 1.403 | 1.275 | 1.281 | 1.307 | 1.279 | 1.281 |
| 1.302 | 1.463 | 1.329 | 1.336 | 1.363 | 1.334 | 1.335 |
| 1.215 | 1.443 | 1.233 | 1.236 | 1.259 | 1.234 | 1.235 |
| 1.240 | 1.482 | 1.263 | 1.266 | 1.293 | 1.264 | 1.266 |
| 0.753 | 1.120 | 0.804 | 0.811 | 0.848 | 0.807 | 0.810 |
| 0.780 | 1.161 | 0.833 | 0.840 | 0.878 | 0.837 | 0.839 |
| 0.808 | 1.202 | 0.862 | 0.869 | 0.909 | 0.866 | 0.869 |
| 0.607 | 1.029 | 0.673 | 0.681 | 0.722 | 0.678 | 0.681 |
| 0.627 | 1.063 | 0.695 | 0.704 | 0.745 | 0.700 | 0.703 |
| 0.577 | 1.046 | 0.633 | 0.638 | 0.666 | 0.634 | 0.637 |
| 0.593 | 1.077 | 0.650 | 0.654 | 0.683 | 0.651 | 0.654 |
| 0.568 | 1.076 | 0.618 | 0.619 | 0.639 | 0.616 | 0.619 |
| 0.488 | 0.994 | 0.549 | 0.553 | 0.576 | 0.550 | 0.552 |
| 0.424 | 0.935 | 0.482 | 0.486 | 0.503 | 0.483 | 0.485 |
| 0.371 | 0.895 | 0.423 | 0.424 | 0.430 | 0.422 | 0.424 |
| 0.161 | 0.549 | 0.208 | 0.212 | 0.218 | 0.211 | 0.212 |
| 0.091 | 0.328 | 0.108 | 0.111 | 0.111 | 0.111 | 0.110 |
| 0.073 | 0.310 | 0.090 | 0.093 | 0.094 | 0.093 | 0.092 |
|  |  |  |  |  | 0.3287 | 0.3242 |
| $\mathrm{A}^{2} \quad 0.5125$ | 2.4129 | 0.3083 | 0.3280 | 0.4590 | 0.3287 | 0.3242 |

Table 7.S13.2 - Anderson-Darling test.
7.14. Design return period tornado wind speed . For the case study of Example 7.40, compute (a) the return period of a design tornado wind speed of $160 \mathrm{~m} / \mathrm{s}$ and $(b)$ the associated probability of exceedance using the contagious model of Eq. (7.3.17) with $X \sim \operatorname{lognormal}(43 \mathrm{~m} / \mathrm{s}, 16.34 \mathrm{~m} / \mathrm{s})$ and $v=1.5 \times 10^{-3}$.

## Solution

(a) $X \sim \operatorname{lognormal}(43 \mathrm{~m} / \mathrm{s}, 16.34 \mathrm{~m} / \mathrm{s})$ with
$\sigma_{\ln x_{\max }}=\sqrt{\ln \left[1+\left(\frac{\sigma}{\mu}\right)^{2}\right]}=0.367, \quad \quad \mu_{\ln x_{\max }}=\ln (\mu)-\frac{1}{2} \ln \left[1+\left(\frac{\sigma}{\mu}\right)^{2}\right]=3.694$
$\mathrm{F}_{\mathrm{X}}(\mathrm{x}=160 \mathrm{~m} / \mathrm{s})=0.999915, \mathrm{~F}_{\mathrm{xmax}}(\mathrm{x})=\exp \left\{-\mathrm{v}\left[1-\mathrm{F}_{\mathrm{X}}(\mathrm{x})\right]\right\}=0.99999987$ and $T=7.89 \times 10^{6}$ years.
(b) The associated probability of exceedance is $p=1-\mathrm{F}_{\mathrm{X} \max }(\mathrm{x}=160)=1.27 \times 10^{-7}$.
7.15. Confidence limits of design values. Consider the Gumbel distribution given in inverse form by Eq. (7.2.26) where the sampling mean and standard deviation $\hat{\mu}_{X}$ and $\hat{\sigma}_{X}$ are used to estimate the population mean and standard deviation, $\mu$ and $\sigma$, respectively. Assuming that $\hat{\mu}_{X}$ and $\hat{\sigma}_{X}$ are asymptotically normally distributed show that

$$
\operatorname{Var}\left[\hat{\xi}_{q}\right]=\operatorname{Var}\left[\frac{\pi^{2}}{6}+1.1396\left(y-n_{e}\right)+1.1\left(y-n_{e}\right)^{2}\right] \frac{\sigma^{2}}{n}
$$

where $y$ denotes the reduced variate. This expression can be used to determine the confidence interval by approximating the sampling distribution of $\xi_{q}$ as $N \sim\left(\hat{\xi}_{q}, \operatorname{Var}\left[\hat{\xi}_{q}\right]\right)$ and substituting the sampling variance for $\sigma^{2}$. Using this procedure, compute the 95 percent confidence interval for the annual maximum hourly storm rainfall predicted in Example 7.16. It can be shown that the variance of any estimator of a parameter is larger than, or at least equal to, a theoretically specified variance known as the Cramer-Rao lower bound, which makes use of the Cramer-Rao inequality of Subsection 5.2.3. This method may be used to derive the variance of quantile estimates from a given extreme value distribution.

Solution. From Illustration (E7.16) $\mu=48.16 \mathrm{~mm}, \sigma=23.76 \mathrm{~mm}, \alpha=18.53 \mathrm{~mm}, b=37.47 \mathrm{~mm}$.
To estimate the confidence interval one can use the following expression


For a level of significance $\alpha=0.05$ and $n=58$, the cumulative distribution of Student's $t$ is $t_{n-1 . \alpha / 2}=0.51$. Considering $\mathrm{q}=0.9$ and 0.99 the quantiles for annual maximum hourly storm rainfall are reported in Illustration (E7.16) $\xi_{0.9}=79.1 \mathrm{~mm} \xi_{0.99}=122.7 \mathrm{~mm}$ and the variance are $\operatorname{Var}\left[\xi_{0.9}\right]=64.54 \mathrm{~mm}^{2} \operatorname{Var}\left[\xi_{0.99}\right]=$ 233.91 mm . So the corresponding intervals of confidence are
$74 \mathrm{~mm} \leq \xi_{0.9}=79.1 \mathrm{~mm} \leq 84.2 \mathrm{~mm}$
$104.2 \mathrm{~mm} \leq \xi_{0.9}=122.7 \mathrm{~mm} \leq 141.2 \mathrm{~mm}$
7.16. Southern California earthquakes. Consider the ordered sample of Magnitudes of southern California annual maximum earthquakes from 1932 to 1962 reported by Lomnitz (1974).

```
4.9}55.3 5.3 5.5 5.5 5.5 5.5 5.6 5.6 5.6 5.8 5.8 5.8 5.9 6.0 6.0 
6.0
```

Originally, the Gumbel distribution was fitted to these data, but other potential candidates are the Fréchet and lognormal distributions. Use the Anderson-Darling goodness-of-fit test to compare the observed and theoretical cumulative frequencies as predicted by the (1) Gumbel, (2) Frèchet, (3) lognormal distributions. Consider $\alpha=0.05$. Compare the theoretical and observed cdfs on a Gumbel probability plot.

Solution. From data set
(1) $X_{\text {max }} \sim \operatorname{Gumbel}(\alpha, b)=\operatorname{Gumbel}(0.4,5.7)$
(2) $X_{\text {max }} \sim \operatorname{Frèchet}\left(\mathrm{x}_{0}, \theta\right)=\operatorname{Frèchet}(5.12,4.89)$
(3) $X_{\text {max }} \sim \operatorname{Lognormal}\left(\mu_{\ln (X)}, \sigma_{\operatorname{tr}(X)}^{2}\right)=\operatorname{Lognormal}(1.8,0.09)$

The null hypothesis $H_{0}$ : the random variable, $X_{\max }$, has a (*) distribution,
where $\left({ }^{*}\right)$ is one of the three considered distributions.
The alternate hypothesis $H_{7}$ : the random variable does not have the specified distribution.
Considering the following levels of significance: $\alpha=0.05$
Critical region: $\mathrm{A}^{2} \geq \mathrm{A}^{2}{ }_{\alpha=0.05}=2.492$
From Table 7.S16, $\mathrm{A}^{2} \leq \mathrm{A}^{2}{ }_{\alpha=0.05}$ is checked for Gumbel distribution and LN distribution where the null hypothesis is not rejected. For Frèchet distribution $\mathrm{A}^{2} \geq \mathrm{A}^{2}{ }_{\alpha=0.05}$ so the null hypothesis is rejected.
Table 7.S16

| Rank | M | Gumbel | Frèchet | LN | Gumbel | Frèchet | LN |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | $\mathrm{F}(\mathrm{x})$ | $\mathrm{F}(\mathrm{x})$ | $\mathrm{F}(\mathrm{x})$ | Sum term | Sum term | Sum term |
| 1 | 4.9 | 0.001 | 0.289 | 0.017 | 0.372 | 0.107 | 0.321 |
| 2 | 5.3 | 0.063 | 0.429 | 0.100 | 0.574 | 0.246 | 0.563 |
| 3 | 5.3 | 0.063 | 0.429 | 0.100 | 0.744 | 0.349 | 0.651 |
| 4 | 5.5 | 0.175 | 0.493 | 0.190 | 0.810 | 0.457 | 0.767 |
| 5 | 5.5 | 0.175 | 0.493 | 0.190 | 1.041 | 0.587 | 0.986 |
| 6 | 5.5 | 0.175 | 0.493 | 0.190 | 1.198 | 0.695 | 1.122 |
| 7 | 5.5 | 0.175 | 0.493 | 0.190 | 1.416 | 0.821 | 1.326 |
| 8 | 5.6 | 0.250 | 0.524 | 0.247 | 1.361 | 0.888 | 1.297 |
| 9 | 5.6 | 0.250 | 0.524 | 0.247 | 1.542 | 1.006 | 1.470 |
| 10 | 5.6 | 0.250 | 0.524 | 0.247 | 1.603 | 1.085 | 1.522 |
| 11 | 5.8 | 0.418 | 0.580 | 0.381 | 1.424 | 1.130 | 1.389 |
| 12 | 5.8 | 0.418 | 0.580 | 0.381 | 1.286 | 1.142 | 1.269 |
| 13 | 5.8 | 0.418 | 0.580 | 0.381 | 1.398 | 1.242 | 1.380 |
| 14 | 5.9 | 0.501 | 0.606 | 0.453 | 1.353 | 1.303 | 1.339 |
| 15 | 6 | 0.578 | 0.630 | 0.525 | 1.320 | 1.363 | 1.299 |
| 16 | 6 | 0.578 | 0.630 | 0.525 | 1.411 | 1.457 | 1.389 |
| 17 | 6 | 0.578 | 0.630 | 0.525 | 1.502 | 1.551 | 1.478 |
| 18 | 6 | 0.578 | 0.630 | 0.525 | 1.404 | 1.572 | 1.408 |
| 19 | 6 | 0.578 | 0.630 | 0.525 | 1.302 | 1.586 | 1.340 |
| 20 | 6 | 0.578 | 0.630 | 0.525 | 1.372 | 1.672 | 1.413 |
| 21 | 6.2 | 0.708 | 0.675 | 0.662 | 1.173 | 1.667 | 1.179 |
| 22 | 6.2 | 0.708 | 0.675 | 0.662 | 0.879 | 1.574 | 0.966 |
| 23 | 6.3 | 0.760 | 0.695 | 0.723 | 0.817 | 1.605 | 0.883 |
| 24 | 6.3 | 0.760 | 0.695 | 0.723 | 0.853 | 1.676 | 0.922 |
| 25 | 6.4 | 0.804 | 0.714 | 0.777 | 0.648 | 1.607 | 0.732 |
| 26 | 6.4 | 0.804 | 0.714 | 0.777 | 0.674 | 1.673 | 0.762 |
| 27 | 6.5 | 0.841 | 0.732 | 0.824 | 0.624 | 1.696 | 0.691 |
| 28 | 6.5 | 0.841 | 0.732 | 0.824 | 0.647 | 1.760 | 0.717 |
| 29 | 6.5 | 0.841 | 0.732 | 0.824 | 0.436 | 1.604 | 0.551 |
| 30 | 7.1 | 0.958 | 0.816 | 0.970 | 0.205 | 1.452 | 0.259 |
| 31 | 7.7 | 0.989 | 0.873 | 0.997 | 0.023 | 0.939 | 0.039 |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  | $\mathrm{~A}^{2}$ | 0.410 | 6.511 |
|  | 0.430 |  |  |  |  |  |  |

7.17. Historical records in extreme value analysis. One wishes to supplement the information available from a $s$-year sample of observed extreme value data with a historical record of $h$ years given that a perception threshold or detection limit was exceeded $l$ times. This occurs, for example, for paleoflood data, and also for water quality data exceeding a prescribed level, for sea wave heights estimated by sailors, and for earthquake intensity estimated from earthquake effects on landscapes. If $e$ denotes the number of observations that exceeded the threshold in the $s$-year sample, a total of $r=l+e$ observations exceeded this threshold for the $n=s+h$ years of record, which is referred to as a censored sample. The natural estimator of the probability of exceedance of the detection threshold is $r / n$. If these $r$ values are indexed by $j=1, \ldots, r$, the reasonable plotting positions accommodating the probabilities of exceedance within the interval $(0, r / n)$ are

$$
p_{j}=\frac{r}{n}\left(\frac{j-\eta}{r+1-2 \eta}\right)
$$

where $p_{j}$ is the probability of exceedance of the $j$ th observation arranged in descending order, and $\eta$ is a value depending on the underlying distribution, say, $\eta=0.4$. Note that $e$ observations that exceeded the threshold are counted among the $r$ exceedances of that threshold. Plotting positions within $(r / n, 1)$ for the remaining $(s-e)$ data in the $s$-year sample are

$$
p_{j}=\frac{r}{n}+\left(1-\frac{r}{n}\right)\left(\frac{j-\eta}{s-e+1-2 \eta}\right)
$$

for $j=1, \ldots, s-e$. For instance, consider the $s=58$ years of data of maximum annual hourly storm depth shown in Table E.7.1. Suppose that during a supplementary historical period of $h=98$ years, the maximum annual hourly storm depth in Genoa exceeded 100 mm in $l=5$ years. The total length of the record is $s+h=156$ years, and $r=l+e=5+3=8$. The observed frequencies are thus modified as shown in Figure 7.P2, assuming that all historical storm depths exceeded the largest observed value.


Fit the GEV distribution to the censored sample of maximum annual hourly storm depth at Genoa University using $L$-moments.

Consider the following data of annual maximum flood flows in $\mathrm{m}^{3} / \mathrm{s}$ for the Arno River in Florence, Italy, with $s=40$ years.
year 19291930193119321933193419351936193719381939194019411942194319441945194619471948194919501951 x 1642 - 12641130122017801520110014906331350125013451079 - 2068 na 9781594120614259221780 year 19521953195419551956195719581959196019611962196319641965196619671968196919701971197219731974

It is reported that the discharge in the Arno exceeded $l=3$ times a threshold of about $2400 \mathrm{~m}^{3} / \mathrm{s}$ in a historical period of $h=145$ years. None of these floods exceeded the 1966 flood, which had a peak
discharge of 3540 cubic meters per second. Fit the GEV distribution to the censored sample by the method of $L$-moments. Find the return period of a flood with peak discharge exceeding $3000 \mathrm{~m}^{3} / \mathrm{s}$.

## Solution

Maximum annual hourly storm depth at Genoa University.
Using this formula, $p_{j}=\frac{r}{n}+\left(1-\frac{r}{n}\right)\left(\frac{j-\eta}{s-e+1-2 \eta}\right)$, to estimate the probability of exceedance of observations that non-exceeded the threshold and this one , $p_{j}=\frac{r}{n}\left(\frac{j-\eta}{r+1-2 \eta}\right)$, to calculate the probability of exceedance of the observations that exceeded the threshold we determine the plotting position of censored data. The table reported here indicates the probability of exceedance of censored records.

| Rank | Data set | $F$ | Rank | Data set | $F$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |
| 1 | 22.80 | 0.010 | 30 | 40.10 | 0.508 |
| 2 | 23.30 | 0.027 | 31 | 40.20 | 0.526 |
| 3 | 24.50 | 0.045 | 32 | 40.20 | 0.543 |
| 4 | 24.70 | 0.062 | 33 | 42.70 | 0.560 |
| 5 | 25.00 | 0.079 | 34 | 43.20 | 0.577 |
| 6 | 25.30 | 0.096 | 35 | 46.70 | 0.594 |
| 7 | 26.90 | 0.113 | 36 | 48.50 | 0.612 |
| 8 | 27.20 | 0.131 | 37 | 50.40 | 0.629 |
| 9 | 27.40 | 0.148 | 38 | 52.40 | 0.646 |
| 10 | 27.60 | 0.165 | 39 | 53.80 | 0.663 |
| 11 | 27.80 | 0.182 | 40 | 53.90 | 0.680 |
| 12 | 29.20 | 0.199 | 41 | 54.50 | 0.697 |
| 13 | 29.30 | 0.216 | 42 | 55.60 | 0.715 |
| 14 | 30.00 | 0.234 | 43 | 55.70 | 0.732 |
| 15 | 30.00 | 0.251 | 44 | 58.10 | 0.749 |
| 16 | 30.20 | 0.268 | 45 | 58.60 | 0.766 |
| 17 | 32.70 | 0.285 | 46 | 64.10 | 0.783 |
| 18 | 32.90 | 0.302 | 47 | 66.50 | 0.801 |
| 19 | 33.20 | 0.320 | 48 | 66.50 | 0.818 |
| 20 | 33.70 | 0.337 | 49 | 69.40 | 0.835 |
| 21 | 33.80 | 0.354 | 50 | 72.60 | 0.852 |
| 22 | 34.70 | 0.371 | 51 | 76.20 | 0.869 |
| 23 | 34.80 | 0.388 | 52 | 79.20 | 0.886 |
| 24 | 38.60 | 0.405 | 53 | 80.00 | 0.904 |
| 25 | 38.70 | 0.423 | 54 | 80.00 | 0.921 |
| 26 | 38.80 | 0.440 | 55 | 89.40 | 0.938 |
| 27 | 39.30 | 0.457 | 56 | 105.70 | 0.952 |
| 28 | 39.60 | 0.474 | 57 | 118.90 | 0.959 |
| 29 | 39.80 | 0.491 | 58 | 128.50 | 0.965 |

Then the parameters of GEV distribution, estimated with method of L-moments are $\mathrm{k}=-0.121, \alpha=15.28 \mathrm{~mm}$, and $\varepsilon=37.27 \mathrm{~mm}$.
Annual maximum flood flow in the Arno river in Florence.
In the same way the table reported here indicates the probability of exceedance of censored records.

| Rank | $Q(\mathrm{mc} / \mathrm{s})$ | $F$ | Rank | $Q(\mathrm{mc} / \mathrm{s})$ | $F$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |
| 1 | 385 | 0.015 | 21 | 1220 | 0.512 |
| 2 | 428 | 0.040 | 22 | 1250 | 0.537 |
| 3 | 540 | 0.065 | 23 | 1264 | 0.562 |
| 4 | 633 | 0.090 | 24 | 1345 | 0.587 |
| 5 | 738 | 0.114 | 25 | 1350 | 0.612 |
| 6 | 776 | 0.139 | 26 | 1390 | 0.637 |
| 7 | 820 | 0.164 | 27 | 1425 | 0.662 |
| 8 | 899 | 0.189 | 28 | 1430 | 0.686 |
| 9 | 901 | 0.214 | 29 | 1490 | 0.711 |
| 10 | 922 | 0.239 | 30 | 1520 | 0.736 |
| 11 | 937 | 0.264 | 31 | 1594 | 0.761 |
| 12 | 978 | 0.289 | 32 | 1600 | 0.786 |
| 13 | 1000 | 0.313 | 33 | 1642 | 0.811 |
| 14 | 1060 | 0.338 | 34 | 1670 | 0.836 |
| 15 | 1079 | 0.363 | 35 | 1760 | 0.861 |
| 16 | 1100 | 0.388 | 36 | 1780 | 0.885 |
| 17 | 1120 | 0.413 | 37 | 1780 | 0.910 |
| 18 | 1120 | 0.438 | 38 | 2068 | 0.935 |
| 19 | 1130 | 0.463 | 39 | 2070 | 0.960 |
| 20 | 1206 | 0.488 | 40 | 3540 | 0.997 |

Then the parameters of GEV distribution, estimated with method of L-moments are
$\mathrm{k}=0.05, \alpha=421.75 \mathrm{~m}^{3} / \mathrm{s}$, and $\varepsilon=1050.20 \mathrm{~m}_{3} / \mathrm{s}$. The return period of a flood with peak discharge exceeding $3000 \mathrm{~m}^{3} / \mathrm{s}$ is $\geq 190$ years.
7.18. Maximum local earthquake intensity and ground motion. Using the epicentral intensity data in the Charleston area, South Carolina, from 1893 to 1984 the following recurrence realtionship is found

$$
\log (V)=1.02-Y,
$$

where $v$ is the number of earthquakes with Intensity larger than $Y$ in a year [see Amick, D. and P. Talwani (1986) Earthquake Recurrence Rates and Probability Estimates for the Occurrence of Significant Seismic Activity in the Charleston Area: the Next 100 years, in Proceedings of the Third Annual Conference on Earthquake Engineering, Charleston, South Carolina, Vol.1, pp.55-64]. Find the cdf of annual maximum earthquake intensity. Suppose that peak ground motion is related to local site intensity as

$$
\log (Z)=0.3 Y+0.014
$$

where $Z$ denotes the average horizontal peak acceleration in $\mathrm{m} / \mathrm{s}^{2}$ peak [see Trifunac, M.D., and A.G. Brady (1975). "On the correlation of seismic intensity scales with peaks of recorded strong ground motion," Bull. Seismological Soc. Amer., Vol. 65, pp. 139-162]. Find the cdf of the annual maximum average horizontal component of epicentral peak acceleration.

Solution. Assuming the annual number of earthquakes to be a Poisson variate with mean $v=10 \exp (1.02$ y), the cdf of annual maximum earthquake Intensity is

$$
F_{Y_{\max }}(y)=e^{-v}=\exp \left[-10 e^{(1-02-y)}\right] .
$$

Because $Y=3.33 \log (Z)-0.0467$, the $c d f$ of maximum annual average horizontal component of epicentral peak acceleration cab be determined from one of the annual maximum earthquake Intensity as
$F_{Z_{\max }}(z)=\exp \left[-10 e^{(1-02-3.33 \log (z)+0.047)}\right]=\exp \left[-29.06 z^{-1.446}\right]$
7.19. Ground motion acceleration in earthquakes. The horizontal peak ground motion acceleration $Z$ is a basic quantity in seismic hazard analysis at a particular site (as discussed in Problem 1.1). It depends on different factors, because it increases with the epicentral intensity $Y$ of an earthquake and with its magnitude $X$, and it decreases with the epicentral distance $r$ of the site. The relationship between these variables also depends on the geographic region, and it is determined using multiple regression. Suppose that $\log Z=0.14 Y+0.24 X-0.68 \log (r)+0.60$, where $Y$ is the epicentral intensity as measured in the modified Mercalli scale, $r$ is the epicentral distance in km , and Z is measured in meters per square second. This relationship provides a good fit for data from the western United States [see Murphy, J.R., and L.J. O'Brian (1977). "The correlation of the peak ground acceleration amplitude with seismic intensity and other physical parameters," Bull. Seismological Soc. Amer., Vol. 67, pp. 877-915]. Suppose that $X$ is a Gumbel-distributed variate with parameters $\alpha=0.49$ and $b=1.4 \times 10^{5}$, and $X=2 Y / 3+1$. Evaluate the $\operatorname{cdf}$ of $Z$ at an epicentral distance of 100 km .

$$
\begin{aligned}
& \text { Solution. For } \mathrm{r}=100 \mathrm{~km} \\
& \log (Z)=0.45 X-0.97, \\
& \text { and } \\
& X=2.22 \log (Z)+2.15 .
\end{aligned}
$$

Assuming that $X$ is a Gumbel distributed variate with parameters $\alpha=0.49$ and $b=1.4 \times 10^{5}$,

the $c d f$ of Z is obtained as

$$
F_{Z}(z)=\exp \left[-e^{-\frac{2.22 \log (z)+2.15-1.4 \times 10^{5}}{0.49}}\right]=\exp \left[-e^{2.857 \times 10^{5}} z^{-1.967}\right]
$$

7.20. Ground motion acceleration in earthquakes. The horizontal peak ground motion acceleration $z$ is a basic quantity in seismic hazard analysis at a particular site. For a magnitude- $x$ earthquake that occurred at a distance $z$ from a given site, an estimate of $z$ can be obtained as $z=A e^{B x} /(u+C)^{2}$.
Typical values are $A=1230, B=0.8, C=25 \mathrm{~km}$ for $u$ in kilometers and $z$ in centimeters per square second [see, for example, Newmark, N.M., and E. Rosenblueth (1977) Fundamentals of Earthquake Engineering, Prentice-Hall, Englewood Cliffs, N.J.]. Suppose that, in a homogeneous area, the annual number of earthquakes exceeding a given threshold $x_{0}$ is a Poisson variate with mean $v$, and $X-x_{0}$ has an exponential distribution with scale parameter $\lambda$. If there are no recognized point sources or active faults, one can assume the epicentral distance as $U \sim$ uniform $(0, l)$, with $l$ denoting the maximum distance between two points in the given area. Show that the annual maximum of $Z$ follows the Fréchet distribution. This distribution can be used to predict design values of horizontal peak ground motion acceleration in this area.

## Solution

Solution. $F_{Z \max }=e p\left\{-\left[\frac{1}{z} \frac{1230}{(l+25)^{2}}\left[\exp \left(\lambda x_{0}+\ln v\right)\right]^{0.8 / \lambda}\right]^{\lambda / 0.8}\right\}$
7.21. Southern California earthquakes. Consider the magnitude data listed in Table E.7.3.
(a) Check the Poisson assumption for the occurrence of earthquakes exceeding magnitude 6 by fitting the exponential distribution to the interarrival time $W$. Compare the observed and fitted cdf of $W$ on an exponential probability plot.
(b) Compute the parameters of the Gutenberg-Richter law for type A zones, and find the return period of a magnitude- 7 earthquake assuming that magnitude is bounded by $X_{\min }=6$ and $X_{\max }=8.22$.

## Solution

(a) $W$ ~exponential( 0.32 years).
(b) $a=2.55 \times 10^{4}, b=0.91$, and $T=92$ years.
7.22. Design return period of snow load. Snow load is evaluated as the product $Z=X W$, with $X$ denoting the depth of snow cover, and $W$ its specific weight. Based on a long record of observations of snow cover in the Italian Apennines, one models the depth $X$ of snow delivered by a snow storm as $X \sim \operatorname{lognormal}(0.32 \mathrm{~m}, 0.29 \mathrm{~m})$. The specific weight of snow $W$ depends on weather and season, and one should model snow pack dynamics to achieve accurate estimates of $Z$. However, measurements of density and temperature of snow yield $W \sim \operatorname{lognormal}\left(3500 \mathrm{~N} / \mathrm{m}^{3}, 800 \mathrm{~N} / \mathrm{m}^{3}\right)$. Also, $X$ and $W$ are positively correlated with $\rho_{X, W}=0.60$. If 4.7 snow storms are expected to occur in a year on average, show that the cdf of maximum annual snow load (see Figure 7.P3) can be written as

$$
F_{Z_{\max }}(z)=\exp \left\{-v\left[1-\Phi\left(\frac{z-\mu_{\ln (Z)}}{\sigma_{\ln (Z)}}\right)\right]\right\},
$$

with $v=4.7, \mu_{\ln (Z)}=6.696$, and $\sigma_{\ln (Z)}=0.837$. Note that
$\rho_{\ln (X), \ln (Y)} \sigma_{\ln (X)} \sigma_{\ln (Y)}=\ln \left(1+V_{X} V_{Y} \rho_{X, Y}\right)$
if $\ln (X)$ and $\ln (Y)$ have a bivariate normal distribution.


Find the return period for design values of (a) $z_{\max }=8000 \mathrm{~N} / \mathrm{m}^{2}$, and (b) $x_{\max }=2.15 \mathrm{~m}$.

## Solution

(a) $T=69$ years. (b) $T=97$ years.

# Applied Statistics for Civil and Environmental Engineers <br> Problem Solution Manual <br> by N.T. Kottegoda and R. Rosso <br> Chapter 8 - Simulation Techniques for Design 

8.1. Flood regionalization. Develop an algorithm to generate random numbers to simulate the twocomponent extreme value distribution of flood flows with cdf

$$
F_{X}(x)=\exp \left(-e^{-\frac{x-b_{1}}{\alpha_{1}}}-e^{-\frac{x-b_{2}}{\alpha_{2}}}\right)
$$

Generate 100 samples each with 1000 items for given values of parameters and find the sampling probability distribution of the coefficients of variation and skewness. Let $\quad \alpha_{1}=1.15 \mathrm{~m}^{3} / \mathrm{s}, \mathrm{b}_{1}=$ $10 \mathrm{~m}^{3} / \mathrm{s}, \alpha_{2}=2.20 \mathrm{~m}^{3} / \mathrm{s}, b_{2}=15 \mathrm{~m}^{3} / \mathrm{s}$. This method may be used to compare the theoretical and sampling variability of these coefficients as estimated from maximum annual flow data observed at different gauging stations in a region.

## Solution

1) Assigning numerical values to the parameters $x_{1}, x_{2}, \alpha_{2}$.
2) The $p d f$ corresponding to the given $c d f$ reads

$$
\left.f_{x}(x)=\left[\frac{1}{\alpha} e^{-\frac{x-x_{1}}{\alpha_{1}}}+\frac{1}{\alpha_{2}} e^{\frac{x-x_{2}}{\alpha_{2}}}\right] e^{-e^{-\frac{x(x-x}{\alpha_{1}}}-e^{-\frac{x-x}{\alpha_{2}}}}\right]
$$

Applying the rejection method, simulate the samples as done in Illustration E8.13 (the same technique is applicable here since the two component extreme values distribution has a bell-shaped $p d f$ ). Find the optimum values of the parameters $c, a$ and $b$ of the proposed $g(x)$.
3) Calculate the coefficient of variation and skewness for each sample.
4) Find the best $p d f$ for the obtained samples of coefficients applying the KolmogorovSmirnov test.
5) The distributions are approximately $N(0.32,0.02)$ and $N(7.91,1.63)$
8.2. Percolation cluster. A fluid spreading randomly through a medium is represented by particles moving on a square grid, that is, a quadratic lattice, where each node is occupied by a pore with a probability of $p$ and neighboring pores are connected by small capillary channels (see Figure 8.P1).

A fluid injected into any given pore may only invade another adjacent pore that is directly connected to that pore through a capillary channel. The pores connected to the injection point form a cluster.
(a)
$p=0.5$

(b)

$$
p=0.7
$$



## Figure 8.P1

(a) Find the minimum probability, $p_{c}$, that a fluid injected into a site on the left edge of the lattice reaches the right edge for the structure shown with $16 \times 16$ nodes. This cluster is called the spanning cluster, or the percolation cluster. Simulations on very large clusters showed that the probability of having a percolation cluster tends to zero as $n \rightarrow \infty$ and $p<0.593$ [from Ziff, R.M. (1986). "Test of scaling exponents for percolation-clusters perimeters," Phys. Rev. Lett., Vol.56, pp. 545-548].
(b) The percolation probability $p_{\infty}(p)$ is defined as the probability that a fluid injected at a site, chosen at random, will wet infinite number of pores. Then, $p_{\infty}(p)=0$ for $\quad p \leq p_{c}$. Design a Monte Carlo experiment to show that the percolation probability vanishes as a power law near $p_{c}$; that is $p_{\infty}(p) \propto\left(p-p_{c}\right)^{\alpha}$ for $p>p_{c}$, and $p \rightarrow p_{c}$. The exponent $\alpha$ is $5 / 36$ for two-dimensional percolation and about $2 / 5$ for three-dimensional percolation.
(c) Design a Monte Carlo experiment to show that for large $n$ the number of sites of the largest cluster increases as $\ln (n)$ for $p<p_{c}$; as $n^{2}$ for $p>p_{c}$; and as $n^{\alpha}$ for $p=p_{c}$; with a value of $\alpha$ of about 1.89 [from Feder, J. (1988). Fractals, Plenum Press, New York, Sec.7.2, "The infinite cluster at $p_{c}$ "].

## Solution

1) Assign uniform (0.1) random numbers to each situ of the lattice.
2) Simulate uniform numbers so as to occupy a fraction $p$ of the sites in the lattice.
3) Verify with repeated simulations which is the smallest fraction $p$ that allows get a
1. spanning cluster.
4) The critical probability $p_{c}$ amounts to 0.59275 .
8.3. Invasion percolation. In a porous medium, oil is displaced by water, which is injected very slowly. Invasion percolation occurs when one neglects any pressure drops both in the invading fluid (water) and in the defending fluid (oil) because the capillary forces completely dominate the viscous forces, and the dynamics of the process is determined at the pore level. Simulation of the process on a lattice consists of following the motion of the water particle injected at a given site on the lattice as it advances through the smallest available pore, thus filling the pores with the invading fluid. As the invader advances, it traps regions of the defending fluid by completely surrounding regions of this fluid, that is, by disconnecting finite clusters of the defending fluid from the exit sites of the sample (see Figure 8.P2).


Figure 8.P2

For a $n \times n$ lattice, the following simulation algorithm describes invasion percolation [from Wilkinson, D.J. and J.F. Willemsen (1983). "Invasion percolation: A new form of percolation theory," J. Phys. A, Vol.16, pp. 3365-3376].

1. One assigns uniform $(0,1)$ random numbers to each site of the lattice.
2. The injection for the invading fluid is assumed to occur at the upper-left corner, and extraction for the defending fluid at the lower-right corner.
3. Growth sites are defined as the sites belonging to the defending fluid and neighbors to the invading fluid.
4. The invading fluid advances to the growth site that has the lowest random number.
5. Trapping is obtained by eliminating the growth sites in regions completely surrounded by the invading fluid from the list of growth sites.
6. The invasion process ends when the invading fluid reaches the exit site.

This algorithm is based on the fact that oil is incompressible (thus, water cannot invade trapped regions of oil). Using this simulation algorithm, show that the number of sites in the central $m \times m$ portion of the $n \times n$ lattice (with $m \ll n$ ) that are occupied by water is proportional to $m^{\alpha}$ with a value of $\alpha$ of about 1.89 [from Dias, M.M. and D.J. Wilkinson (1986). "Percolation with trapping," J. Phys. A, Vol.19, pp. 3131-3146). This is, for instance, one origin of the phenomenon of residual oil.

## Solution.

The procedure is similar to that of Problem 8.2 after assigning uniform $(0,1)$ random numbers to each site of the lattice.
8.4. Water storage. Water storage $X$ in a large reservoir is modeled as a truncated normal variate with pdf

$$
f_{X}(x)=\frac{1}{[\Phi(1)-\Phi(-2)]} \frac{1}{3 \sqrt{2 \pi}} \exp \left[-\left(\frac{x-6}{3}\right)^{2}\right]=\frac{1.5389}{3 \sqrt{2 \pi}} \exp \left[-\left(\frac{x-6}{3}\right)^{2}\right]
$$

for $0 \leq x \leq 9$ units, and zero elsewhere. Find $F_{X}(7)$ by Monte Carlo integration using 1000 simulation cycles, and compare this result with that obtained using tables of the normal distribution. What is the number of simulation cycles required to achieve a standard error of estimation not larger than 10 percent of the true value? Assume the mode of $X$ as the maximum ordinate for the rectangular envelope of $f_{X}(x)$, with
$0 \leq x \leq 7$ units.

## Solution

1) Find the maximum $\alpha$ of the given $p d f$ in the interval $(0,9)$.
2) Generate 2 samples of uniform variates $U_{1}(0,9)$ and $U_{2}(0, \alpha)$.
3) Compute the number $c$ of points with coordinates $\left(U_{1}, U_{2}\right)$ lying in the area under the given $p d f$ in the interval $(0,7)$.
4) Compute the ratio $d$ between the $c$ and the total number of points generated.
5) Multiple $d$ for the area given by $9 \alpha$.
6) By computing analytically $F_{x}(7)$, find the requested number of cycles.
7) The resulting $F_{x}(7)$ is approximately 0.7388 . The requested number of cycles is approximately 361 .
8.5. Storm rainfall. The total amount of water $Z$ delivered by a storm in a given location is evaluated as $Z=X Y$ from independent duration $X$ and average rainfall rate $Y$ of a storm, with $X \sim \operatorname{lognormal}(1.2 \mathrm{hr}$, $6 \mathrm{hr}^{2}$ ), and $Y \sim$ lognormal $\left(10 \mathrm{~mm} / \mathrm{hr}, 100 \mathrm{~mm}^{2} / \mathrm{hr}^{2}\right.$ ). Assume that the number of storms in a year is a Poisson variate with a mean of 25 . Using Monte Carlo simulation find the cdf of the annual maximum hourly storm depth, that is, the maximum amount of rainfall in a year which is delivered in the specified duration of one hour.

## Solution

1) Simulate a sample from the Poisson variate with mean of 25 , as done in Illustration E8.15. This given the position of each storm within the year.
2) Simulate samples from the two given lognormal distributions. Generation of lognormal random numbers can be done by generating normal random numbers with assigned mean and variance and performing an exponential transformation. The mean and variance of the corresponding normal distribution (see Illustration 4.32) are
$\mu_{\ln (X)}=\ln \left[\frac{\mu_{X}}{\left(V_{X}^{2}+1\right)^{1 / 2}}\right], \quad V_{\ln (X)}=\sqrt{\ln \left(V_{x}^{2}+1\right)}$
This allows to compute the duration and intensity of each storm.
3) Compute the maximum rainfall height for the given durations.
4) Find to best probability distribution is approximately a Gumbel (19.06, 0.10).
8.6. Storm rainfall. Solve Problem 8.5 under the assumption that the duration $X \sim \operatorname{lognormal}(1.2 \mathrm{hr}$, $6 \mathrm{hr}^{2}$ ) of a storm and its average intensity $Y \sim \operatorname{lognormal}\left(10 \mathrm{~mm} / \mathrm{hr}, 100 \mathrm{~mm}^{2} / \mathrm{hr}^{2}\right)$ are negatively correlated variates with $\rho_{X, Y}=-0.3$. Note that if two jointly distributed variates $U$ and $W$ follow the bivariate normal distribution, then the covariance between $X=\exp (U)$ and $Y=\exp (W)$ is given by $\operatorname{Cov}(X$, $Y)=\mu_{X} \mu_{Y}\{\exp [\operatorname{Cov}(U, W)]-1\}$. One can thus generate correlated values of $X$ and $Y$ from bivariate normal random numbers distributed as $U=\ln (X) \sim N\left(\mu_{\ln (X)}, \sigma_{\ln (X)}{ }^{2}\right)$ and $W=\ln (Y) \sim N\left(\mu_{\ln (Y)}, \sigma_{\ln (Y)}{ }^{2}\right)$ having a correlation coefficient of
$\rho_{U, W}=\frac{\ln \left(1+V_{X} V_{Y} \rho_{X, Y}\right)}{\sqrt{\ln \left(1+V_{X}^{2}\right) \ln \left(1+V_{Y}^{2}\right)}}$.

## Solution

1) Use the same procedure given for the Exercise 8.5; generate the samples the two lognormal distributions by assigning the specified correlation. This can be done generating jointly distributed normal random numbers with mean and variance

$$
\begin{aligned}
& \mu_{\ln (X)}=\ln \left[\frac{\mu_{X}}{\left(V_{X}^{2}+1\right)^{1 / 2}}\right] \\
& \left.\ln _{n(X)}=\sqrt{\ln \left(V_{x}^{2}\right.}+1\right)
\end{aligned}
$$

and correlation given by the indicated relationship. Generation of multivariate normal random numbers is done in Illustration E8.16.
$2)$ The resulting distribution is approximately a Gumbel $(7.00,0,74)$.
8.7. Generation of beta variates. Let $X \sim \operatorname{beta}(a, b)$ with $0 \leq x \leq 1$. Develop an algorithm to generate beta random numbers based on the rejection method. Compare the cdf resulting from simulation of 100 samples of $X \sim$ beta $(1,3)$ with its analytical form by using the Kolmogorov-Smirnov test.

## Solution

1) Applying the rejection method, simulate the samples as done in Illustration E18.13 (the same technique is applicable here since the beta distribution has a bell-shaped $p d f$ ). Find the optimum values of the parameters $c, \alpha$ and $b$ of the proposed $g(x)$.
8.8. Wastewater treatment plant. An activated-sludge plant includes five serial processes: (1) coarse screening, (2) grit removal, (3) plain sedimentation, (4) contact treatment, and (5) final settling. Let $X_{i}$ denote the efficiency of the $i$ th treatment, that is, the fraction of remaining pollutant after removal by the $i$ th serial treatment. For example, $X_{1}$ is the fraction of pollutant removed by treatment process $1, X_{2}$ is the fraction of the remaining pollutant after removal by treatment process 2 , and so on. The amount $Q_{\text {out }}$ of pollutant in the effluent is given by

$$
Q_{\text {out }}=\left(1-X_{1}\right)\left(1-X_{2}\right)\left(1-X_{3}\right)\left(1-X_{4}\right)\left(1-X_{5}\right) Q_{\text {in }},
$$

where $Q_{\text {in }}$ denote the amount of pollutant in the untreated inflow. A quality indicator of the performance of the plant is then defined as

$$
Y=\left(1-X_{1}\right)\left(1-X_{2}\right)\left(1-X_{3}\right)\left(1-X_{4}\right)\left(1-X_{5}\right) .
$$

Consider a plant with the following single-process mean efficiencies in the removal of the 5 -day $20^{\circ} \mathrm{C}$ biological oxygen demand (BOD):

$$
\mu_{1}=0.05, \mu_{2}=0.05, \mu_{3}=0.20, \mu_{4}=0.70, \mu_{5}=0.10
$$

where $\mu_{i}=E\left[X_{i}\right]$. Suppose that $X_{1}, X_{2}, X_{3}$, and $X_{5}$ are normal variates with common coefficient of variation of 0.2 , and $X_{4} \sim$ uniform $(0.6,0.8)$. Find the pdf and cdf of $Y$ by simulation assuming that the five processes are independent of each other. Compare the mean of $Y$ with the nominal value.

## Solution

1) Simulate samples of the quantities involved, given their $p d f$. Generation of normal random numbers can be done as explained in Illustration E8.11. Simulation of uniform random numbers can be done as explained in Section 8.2.1.
2) The nominal value of the mean of $Y$ can be computed as

$$
\mathrm{E}(\mathrm{Y})=\left(1-\mathrm{E}\left(\mathrm{X}_{1}\right)\right)\left(1-\mathrm{E}\left(\mathrm{X}_{2}\right)\right)\left(1-\mathrm{E}\left(\mathrm{X}_{3}\right)\right)\left(1-\mathrm{E}\left(\mathrm{X}_{4}\right)\right)\left(1-\mathrm{E}\left(\mathrm{X}_{5}\right)\right) .
$$

8.9. Underground pipeline subject to corrosion. An underground pressured pipeline is subject to stresses caused by external soil pressure and by internal (fluid) pressure. Assuming the radius of pipe $r$ is much larger than the thickness of the pipe wall $t$, the circumferential stress $s_{f}$ due to internal pressure is estimated as
$s_{f}=p r / t$,
where $p$ is the internal pressure. The bending stress $s_{s}$ in the circumferential direction produced in the pipe wall by the external soil loading can be estimated from
$s_{s}=\frac{6 k_{m} C_{d} \gamma B_{d}^{2} E t r}{E t^{3}+24 k_{d} p r^{3}}$.
Here $C_{d}$ is a dimensionless calculation coefficient for soil load, $\gamma$ is the unit weight of soil backfill, $B_{d}$ is the width of the ditch at the top of the pipe, $E$ is the modulus of elasticity of the pipe metal, $k_{m}$ is a bending moment coefficient dependent on the distribution of vertical load and reaction and $k_{d}$ is a deflection coefficient dependent on the distribution of vertical load and reaction. The circumferential bending stress $s_{t}$ produced in the pipe wall due to traffic loads (such as that resulting from roadway, railway, or airplane traffic) may be estimated from
$s_{t}=\frac{6 k_{m} I_{c} C_{t} F E t r}{A\left(E t^{3}+24 k_{d} p r^{3}\right)}$,
where $I_{c}$ is a dimensionless impact factor, $C_{t}$ is the dimensionless surface load coefficient, $F$ is the wheel load on surface, and $A$ is the effective length of pipe on which load is computed. If the pipe remains in the elastic range under load, the maximum circumferential stress is given at the critical sections by $s_{f}+s_{s}+s_{t}$. By using simulation, compute the expected maximum circumferential stress and its coefficient of variation. Suppose that the quantities involved have the following distributions [from Ahammed, M. and R.E. Melchers (1994). "Reliability of underground pipelines subject to corrosion," J. of Transp. Engin. Div., ASCE, Vol. 120, pp. 989-1002, reproduced by the permission of the publisher, ASCE].

| Variate | Distribution | Mean | Coefficient of variation |
| :---: | :---: | :---: | :---: |
| $p$ | normal | 6.205 MPa | 0.20 |
| $r$ | normal | 228.6 mm | 0.05 |
| $t$ | normal | 8.73 mm | 0.05 |
| $k_{m}$ | lognormal | 0.235 | 0.20 |
| $C_{d}$ | lognormal | 1.32 | 0.20 |
| $\gamma$ | normal | $18.85 \times 10^{-6} \mathrm{~N} / \mathrm{mm}^{3}$ | 0.10 |
| $B_{d}$ | normal | 762 mm | 0.15 |
| $E$ | normal | 206800 MPa | 0.05 |
| $k_{d}$ | lognormal | 0.108 | 0.20 |
| $I_{c}$ | normal | 1.5 | 0.25 |
| $C_{t}$ | lognormal | 0.12 | 0.20 |
| $F$ | normal | 267000 N | 0.25 |
| $A$ | normal | 914 mm | 0.20 |

The main effect of corrosion is weight loss. Because we are mainly interested is general corrosion, it is assumed that the loss of wall thickness can be modeled empirically by a power law, $d=k \tau^{n}$, where $\tau$ is the time of exposure in years, $k$ is a multiplying coefficient, and $n$ is a constant. Accordingly, one will substitute $(t-d)$ or $\left(t-k t^{n}\right)$ for $t$ in the above equations to account for corrosion. Suppose that both $k$ and $n$ are normal variates with means of 0.3 and 0.6 , respectively, and coefficients of variation of 0.3 and 0.2 , respectively. Evaluate the expected maximum circumferential stress and its coefficient of variation after an exposure of 30 years.

## Solution

1) Simulate samples of the quantities involved, given their $p d f$. Generation of normal random numbers can be done as explained in Illustration E8.11. Generation of lognormal random numbers can be done by generating normal random numbers with assigned mean and variance and performing an exponential transformation. The mean and variance of the corresponding normal distribution (see Illustration 4.32) are

$$
\begin{aligned}
& \mu_{\ln (X)}=\ln \left[\frac{\mu_{X}}{\left(V_{X}^{2}+1\right)^{1 / 2}}\right] \\
& \left.1_{n(X)}=\sqrt{\ln \left(V_{x}^{2}\right.}+1\right)
\end{aligned}
$$

2) Compute the circumferential stresses given the items of the quantities involved. This allows to obtain a sample of circumferential stressed.
3) Compute the expected value and coefficient of variation of the obtained sample. They are approximatively 288 and 0.22 .
8.10. Debris flow. Debris flows, also referred as mudflows, are a significant hazard in many parts of the world, causing extensive damage to engineering structures such as buildings, bridges and culverts, as well as causing loss of life. From data analysis in the Los Angeles area, California, the following empirical formula was proposed to estimate the debris volume $X$ in cubic meters:

$$
X=56.56 Y^{0.75} a^{1.25}\left(1+80 e^{-0.239 a-0.537 W}\right)^{0.5}
$$

where $a$ denotes the watershed area in square kilometers, $Y$ the 72-hour maximum annual rainfall depth in millimeters, and $W$ the time interval between watershed burning in years [from McCuen, R.H., Ayyub, B.M. and T.V. Hromadka (1990). "Risk of Debris-Basin failure, ASCE J. Water Resources Plan. and Man. Div., ASCE, Vol. 116, pp.473-483, reproduced by the permission of the publisher, ASCE]. Assume that $Y$ and $W$ are independent variates, $Y$ is a Gumbel-distributed variate with a mean of 100 mm and a coefficient of variation of 0.444 , and $W$ is a lognormal-distributed variate with a mean of 8 years and a coefficient of variation of 1.375 . Consider a drainage area $a$ of $2.5 \mathrm{~km}^{2}$, and find the probability distribution of $X$ by simulation.

## Solution

1) Simulate samples of the quantities involved, given their $p d f$. Generation of normal random numbers ca be don as explained in Illustration E8.11. Generation of lognormal random numbers can be done by generating normal random numbers with assigned mean and variance and performing an exponential transformation. The mean and variance of the corresponding normal distribution ( see Illustration 4.32) are
$\mu_{1 n(X)}=\ln \left[\frac{\mu_{X}}{\left(V_{X}^{2}+1\right)^{1 / 2}}\right]$
$\left.1_{n(X)}=\sqrt{\ln \left(V_{x}^{2}\right.}+1\right)$
2) Compute the variable $X$ given the items of the quantities involved. This allows to obtains a sample of debris volumes.
3) Find the best $p d f$ for the obtained sample using the Kolmogorov-Smirnov test.
4) The $p d f$ is approximately lognormal (14.200, 9915 ${ }^{2}$ ).
8.11. Reservoir capacity. In determining the optimal capacity of a reservoir let us assume that the manager will follow the so-called normal operating rule shown in Fig. 8.P3.


Figure 8.P3

In this case, the draft or release $R_{t}$ is obtained as

1. $R_{t}=S_{t}+X_{t}$,
if $S_{t}+X_{t} \leq d_{\tau}$,
2. $R_{t}=d_{\tau}$,
if $d_{\tau}<S_{t}+X_{t}<d_{\tau}+c$,
3. $R_{t}=S_{t}+X_{t}-c$,
if $S_{t}+X_{t} \geq d_{\tau}+c$,
where $c$ denotes the capacity of the reservoir. The rate of demand of water supply, $d_{\tau}$, is equal to the mean annual runoff in March, April, November and December. It is reduced to 85 percent in May, August, September and October, and to 70 percent in January, February, June and July. Using this rule and the other data of Example 8.22, find the optimal capacity of the reservoir for an average annual deficit of 1 percent of the annual demand. Assume full reservoir as the initial condition. Compare this result with that of Example 8.22.

## Solution

1) Find the total annual demand of water supply by using the data of Illustration E8.23. Compute the maximum deficit allowed ( $1 \%$ of the total annual demand).
2) Using the recursive stochastic equation of Illustration E8.19, simulate 100 years of monthly inflows.
3) Using these data for a given value $c$ of the capacity, and applying the relationships (a), (b) and (c) find the release $R_{t}$ and the annual deficit.
4) Change the value of $c$ to find the optimum capacity.
5) Verify with other simulations the correctness of the results.
6) The optimum capacity is approximately 0.340 .
8.12. Model selection for extreme value data. Let $X$ denote a GEV-distributed random variable with parameters $\varepsilon=0, \alpha=1$ and $k=-0.2$. Perform the following experiment.
(a) Generate a sample of 100 outcomes of this variate.
(b) Fit the (1) Gumbel, (2) Fréchet, (3) lognormal, (4) gamma, (5) GEV, (6) shifted-lognormal, (7) shifted-gamma and (8) log-Pearson Type III distributions to the generated sample.
(c) Perform a goodness-of-fit testing procedure using the chi-squared, Kolgomorov-Smirnov and Anderson-Darling tests.

Determine the probability models for which the null hypothesis is not rejected.
Repeat the experiment for a sample of 10,000 outcomes.

## Solution

Follow procedures for simulation in Example 8.9 and methods of fitting and testing in Chapter 7, including that adopted in Example 7.32.
8.13. River network. A river network can be described as a random binary tree, as shown in Fig. 8.P4a.

(b)

$$
X_{i+1}
$$

Figure 8.P4

A mathematical tree originates from a root (ancestor) and it grows by subsequent branching, through a bifurcation process. A link is defined as the line segment between two vertices of the tree; external links are those connecting an internal vertex (junction) with an external vertex (source), and internal links are those joining two junctions. The total number of external links is called the "magnitude" of the tree. A tree of magnitude $m$ has $n=2 \mathrm{~m}-1$ links (total progeny). A hierarchical order can be assigned to each element of the tree by indexing a link by its "level" of branching, that is by progressively numbering the links from 1 , which is assigned to the root, to $k$, which is the level of the source having the highest distance from the root. Let $X_{i}$ denote the number of links at branching level $i$. In a standard model of river networks, the tree randomly branches with a constant branching probability $p$ for all the links independently of the bifurcation level. Therefore, the number of links at level $i+1, X_{i+1}$, depends only on $X_{i}$, the number of links at the previous level. The process of branching through upstream growth is called Markovian, because each stage of development depends only on the immediately previous one. If $p=1 / 2$, the probability that $X_{i-1}$ links at level $i-1$ will originate $X_{i}$ links at level $i$ is

$$
\operatorname{Pr}\left[X_{i}=x_{i} \mid X_{i-1}=x_{i-1}\right]=2^{-x_{i-1}}\binom{x_{i-1}}{x_{i} / 2}
$$

where

$$
\binom{x_{i-1}}{x_{i} / 2}
$$

denotes the combinations of $x_{i-1}$ links taking $x_{i} / 2$ at a time, and $X_{i-1} \leq 2 X_{i}$. For example, if $X_{i-1}=8$ at level $i$ - 1 , the corresponding transition probabilities $p_{X}$ for $X_{i}$ are those listed in the following table with the associated conditional $\operatorname{cdf} F_{X}$.

| $X_{i}=$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $p_{X}=$ | 0.0039 | 0.0313 | 0.1094 | 0.2188 | 0.2734 | 0.2188 | 0.1094 | 0.0313 | 0.0039 |
| $F_{X}=$ | 0.0039 | 0.0352 | 0.1445 | 0.3633 | 0.6367 | 0.8555 | 0.9648 | 0.9961 | 1.0000 |

One can simulate a river network by using the probability integral transform method as shown in Fig. 8.3.P4b. The process terminates when $X_{k}=0$ (adsorbing state), and level $k$ is called the "diameter" of the river network. Using this model find the probability distribution of the level $j$ for which the number $X_{j}$ of links is a maximum in trees with diameter of $k=8$.

## Solution

1) Simulate the river network using the probability integral transform method. Keep only the simulation whose diameter $k$ is equal to 8 . In order to reduce the number of simulations required, assign value 1 to the branching probability $p$ of the root (first link).
2) Find the best $p d f$ of the obtained sample of levels $j$ for which the number of links is a maximum in the obtained trees, using the Kolmogorov-Smirnov test.
3) The resulting $p d f$ is approximately lognormal ( $2.82,0.96$ ).
8.14. Seismic hazard. In a period of 600 years about 330 earthquakes occurred in Central Italy having epicentral MCS intensity $X$ exceeding 6. Also, $X$ is modeled as an exponential variate with scale and location parameters of 0.91 and 6 , respectively. Seismic hazard in a specific site is represented by MCS intensity $Y$ as evaluated from the following attenuation law

$$
Y=X-\frac{1}{\ln \psi} \ln \left[1+\frac{\psi-1}{\psi_{0}}\left(\frac{Z \varphi^{x_{0}-X}}{z_{0}}-1\right)\right]
$$

where $Z$ denotes the distance from the picenter, $z_{0}=9.5 \mathrm{~km}$ is the distance of the isoseismical line for epicentral intensity $x_{0}=10$, and $\psi_{0}=1, \psi=1.5$ and $\varphi=1.3$ are the estimated values of parameters $\psi_{0}, \psi$ and $\varphi$ for Central Italy [see Grandori, G., Drei, A., Perotti, F. and A. Tagliani (1991). "Macroseismic intensity versus epicentral distance": The case of Cental Italy, in: Stucchi, M., Postpischl, D,. and D. Slejko, eds., "Investigations of Historical Earthquakes in Europe", Tectonophysics, Vol. 193, pp.175181]. Suppose $Z \sim \operatorname{uniform}(3 \mathrm{~km}, 25 \mathrm{~km})$. and find the probability distribution of $Y$ by simulation. Compute t

The 100-year MCS intensity for this region assuming that $Y$ is a Gumbel variate.

## Solution

1) Applying the rejection method, simulate a sample of the given exponential distribution as done in Illustration E8.13 (the same technique is applicable here since the exponential distribution has a bell-shaped $p d f$ ). Find the optimum values of the parameters $c, \alpha$ and $b$ of the proposed $g(x)$.
2) Simulate a sample of equal length of the Uniform (3.25) distribution.
3) Using the two previous samples, calculate the sample of the $Y$ 's.
4) Find the best $p d f$ for the sample of $Y$ 's.
5) The distributions is approximately a Gumbel $(0.70,4.65)$.

# Applied Statistics for Civil and Environmental Engineers 

# Problem Solution Manual 

by N.T. Kottegoda and R. Rosso

## Chapter 9 - Risk and Reliability Analysis

9.3. Structural safety factor. Consider a structure designed with a central safety factor of 2 , and loaded with a non-random load $y$. Determine its risk of failure for (a) normally, (b) lognormally, and (c) gamma distributed load-carrying capacity $X$ with mean $\mu_{X}=y$ and coefficient of variation $V_{X}=0.5$.

Solution. The probability of failure is given by
$p_{f}=\operatorname{Pr}[X>\zeta y]=1-F_{X}(\zeta y)=1-F_{X}(2 y)$.
(a) If $X \sim N\left(y,\left(y V_{X}\right)^{2}\right)$,

$$
p_{f}=\int_{2 y}^{\infty} \frac{1}{y V_{X} \sqrt{2 \pi}} \exp \left[-\frac{1}{2}\left(\frac{x-y}{y V_{X}}\right)^{2}\right] d x=\Phi\left(-\frac{2 y-y}{y V_{X}}\right)=\Phi\left(-\frac{1}{V_{X}}\right)=\Phi(-2)=0.023,
$$

where $\Phi($.$) denotes the c d f$ of the standard normal variate.
(b) If $X \sim L N\left(\mu_{\ln (X)}, \sigma_{\ln (X)}{ }^{2}\right)$,
$p_{f}=\int_{2 y}^{\infty} \frac{1}{x \sqrt{2 \pi \sigma_{\ln (X)}^{2}}} \exp \left[-\frac{\left(\ln x-\mu_{\ln (X)}\right)^{2}}{2 \sigma_{\ln (X)}^{2}}\right] d x$.
Since
$\sigma_{\ln (X)}{ }^{2}=\ln \left(1+V_{X}{ }^{2}\right)=\ln \left(1+0.5^{2}\right)=0.223$
and

$$
\begin{aligned}
& \mu_{\ln (X)}=\ln \left(\mu_{X}\right)-0.5 \sigma_{\ln (X)}^{2}=\ln y-0.5 \times 0.223=\ln y-0.112 \\
& p_{f}=1-\Phi\left(\frac{\ln (2 y)-\mu_{\ln (X)}}{\sigma_{\ln (X)}}\right)=1-\Phi\left(\frac{\ln (2 y)-\ln y+0.112}{\sqrt{0.223}}\right)= \\
& =1-\Phi\left(\frac{\ln 2+0.112}{0.472}\right)=1-\Phi(1.704)=0.044
\end{aligned}
$$

(c) If $X \sim \operatorname{gamma}(\lambda, \gamma)$,
$p_{f}=\int_{2 y}^{\infty} \frac{1}{\lambda \Gamma(\gamma)}\left(\frac{x}{\lambda}\right)^{\gamma-1} \exp \left(-\frac{x}{\lambda}\right) d x$.
Since
$\gamma=1 / V_{X}^{2}=4$
and

$$
\begin{aligned}
& \lambda=\mu_{X} / \gamma=y / 4, \\
& \begin{aligned}
& p_{f}=\int_{2 y}^{\infty} \frac{4}{y \Gamma(4)}\left(\frac{4 x}{y}\right)^{3} \exp \left(-\frac{4 x}{y}\right) d x=\frac{1}{6} \int_{2 y}^{\infty} \frac{4}{y}\left(\frac{4 x}{y}\right)^{3} \exp \left(-\frac{4 x}{y}\right) d x=. \\
&=\frac{1}{6} \int_{8}^{\infty} \frac{4}{y} \frac{y}{4} u^{3} e^{-u} d u=\frac{1}{6} \int_{8}^{\infty} u^{3} e^{-u} d u=\frac{1}{6}\left[-\left(u^{3}+3 u^{2}+6 u+6\right) e^{-u}\right]_{u=8}^{u=\infty}=0.042 .
\end{aligned}
\end{aligned}
$$

9.4. Pile. The conventional safety factor of a pile is $z^{*}=1.2$. Both load-carrying capacity and strength are independent normal variates with coefficients of variation of $30 \%$ and $50 \%$, respectively. Find the sigma bound $h_{X}=h_{Y}=h$, if the central safety factor is $\zeta=1.6$.

Solution. By combining Eq. (9.1.4),

$$
\zeta=\mu_{X} / \mu_{\mathrm{Y}},
$$

with Eq. (9.1.6),

$$
z^{*}=\frac{\mu_{X}-h_{X} \sigma_{X}}{\mu_{Y}+h_{Y} \sigma_{Y}}=\frac{\mu_{X}}{\mu_{Y}} \frac{1-h V_{X}}{1+h V_{Y}}
$$

one gets

$$
h=\frac{1-z^{*} / \zeta}{V_{X}+V_{Y} z^{*} / \zeta}=\frac{1-1.2 / 1.6}{0.3+0.5 \times 1.2 / 1.6}=0.370
$$

9.5. Flow meters. The reliability of a standard flow meter used for rating municipal water supply to private buildings is $95 \%$, and it is estimated that a defective meter underestimates flow of $20 \%$. If the tolerable loss is $2 \%$ for each supplied building, what is the reliability of the municipal system if 100 buildings are supplied? Calculate also the reliability if 1000 buildings are supplied. Note that for large $n$ the binomial distribution can be approximated by the normal (with the same mean and variance) as shown in Sec. 4.2.
Solution. The probability that a flow meter is defective is $p=1-0.95=0.05$. If $n=100$ buildings are supplied, one assumes that the number $N$ of defective meters is a binomial variate with parameters $p$ and $n$. Thus, the expected number of defective meters is
$\mu_{N}=p n=0.05 \times 100=5$,
and the variance is
$\sigma_{N}{ }^{2}=p n(1-p)=0.05 \times 100(1-0.05)=4.75$.
Hence,
$\sigma_{N}=\sqrt{ } 4.75=2.179$.
If $l=0.2$ is the loss due to a defective meter, the total loss $X$ has a mean of
$\mu_{X}=\mu_{N} l=5 \times 0.2=1$,
and a standard deviation of
$\sigma_{X}=\sigma_{N} l=2.179 \times 0.2=0.436$.
If the tolerable loss is $t=0.02$, the required system reliability is the probability that the total loss $X$ is less than $t n=0.02 \times 100=2$, that is $r=\operatorname{Pr}[X \leq 2]$. Assuming that $X \sim N\left(1,0.436^{2}\right)$, system reliability is
$r=\Phi((2-1) / 0.436)=\Phi(2.294)=0.989$,
with $\Phi(\cdot)$ denoting the $c d f$ of the standard normal variate.
For $n=1000$,
$\mu_{N}=p n=0.05 \times 1000=50$,
$\sigma_{N}^{2}=p n(1-p)=0.05 \times 1000(1-0.05)=47.5$,
$\sigma_{N}=\sqrt{ } 47.5=6.892$,
$\mu_{X}=\mu_{N} l=50 \times 0.2=10$,
$\sigma_{X}=\sigma_{N} l=6.892 \times 0.2=1.378$.
Thus,
$r=\Phi((20-10) / 1.378)=\Phi(7.255) \approx 1$.
9.6. Uniform capacity and demand. The joint capacity-demand distribution of a supply system is uniform, say, $f_{X, Y}(X, y)=(a b)^{-1}$ units ${ }^{-2}$, for $0 \leq X \leq a$ units and $0 \leq Y \leq b$ units, with $a \geq b$. What is the reliability of the system?
Solution. The reliability of the system is given by

$$
\operatorname{Pr}[X \geq Y]=\int_{0}^{b} d x \int_{0}^{x} \frac{1}{a b} d y+\int_{b}^{a} d x \int_{0}^{b} \frac{1}{a b} d y=\int_{0}^{b} \frac{x}{a b} d x+\int_{b}^{a} \frac{1}{a} d x=\frac{b^{2}}{2 a b}+\frac{a}{a}-\frac{b}{a}=1-\frac{b}{2 a} .
$$

It is also seen that the shaded area in Figure 9.56 is $b^{2} / 2+b(b-a)$, which is normalized by $a b$ to obtain $1-b /(2 a)$.

Figure 9.S6

9.7. Pipe flow. The pressure $p$ and water flow $q$ in a circular pipe are measured as $p=7 \mathrm{kpascal}\left(\mathrm{kN} / \mathrm{m}^{2}\right)$ and $q=0.08 \mathrm{~m}^{3} / \mathrm{s}$, respectively. The pipe is located 2 m above the reference level and its diameter is $d=20 \mathrm{~cm}$. The total head $h$ (energy) in the pipe at the point of interest is given by Bernoulli's equation,
$h=\frac{u^{2}}{2 g}+\frac{p}{\gamma}+z$,
where $X_{1}=u^{2} / 2 g$ is called the kinetic head, $X_{2}=p / \gamma$ the pressure head, and $X_{3}=z$ the elevation head. Assuming $X_{1}, X_{2}$, and $X_{3}$ to be normal variates with a coefficient of variation of 0.05 , an engineer needs to determine the reliability of system operation for $h>h_{0}$, with $h_{0}=3 \mathrm{~m}$ (the flow velocity is defined as the ratio between flow and cross sectional area of the pipe, say, $u=q /\left(\pi d^{-2} / 4\right), g$ is the acceleration due to gravity, say $g=9.806 \mathrm{~m} / \mathrm{s}^{2}$, and $\gamma$ is the specific weight of water, say, $\gamma=9.806 \mathrm{kN} / \mathrm{m}^{3}$ ). Assume that all variates are independent of each other.
Solution. The mean values of kinetic, pressure and elevation heads are given by
$\mu_{1}=\left[0.08 /\left(3.142 \times 0.2^{2} / 4\right)\right]^{2} /(2 \times 9.806)=0.331 \mathrm{~m}$,
$\mu_{2}=7 / 9.806=0.714 \mathrm{~m}$
and
$\mu_{3}=2 \mathrm{~m}$,
respectively. If $V_{1}=V_{1}=V_{3}=0.05$, the values of the corresponding standard deviation are $\sigma_{1}=0.05 \times 0.331=0.017 \mathrm{~m}, \sigma_{2}=0.05 \times 0.714=0.036 \mathrm{~m}$ and $\sigma_{3}=0.05 \times 2=0.107 \mathrm{~m}$.
Under tha assumption of independent normal variates, the total head $Y$ is a normal variate with mean
$\mu_{\mathrm{Y}}=\mu_{1}+\mu_{2}+\mu_{3}=0.331+0.714+2=3.044 \mathrm{~m}$,
and standard deviation
$\sigma_{\mathrm{Y}}=\left(\sigma_{1}^{2}+\sigma_{2}^{2}+\sigma_{3}^{2}\right)^{1 / 2}=\left(0.017^{2}+0.036^{2}+107^{2}\right)^{1 / 2}=0.107 \mathrm{~m}$.
The required reliability is given by $r=\operatorname{Pr}\left[Y>h_{0}\right]$ with $h_{0}=3 \mathrm{~m}$. It can be computed as $r=\operatorname{Pr}[Y>3]=1-\Phi((3-3.044) / 0.107)=\Phi(-0.414)=0.661$,
with $\Phi(\cdot)$ denoting the $c d f$ of the standard normal variate.
9.8. Column load. A column of a building is designed with a central safety factor of 1.6. The coefficient of variation of its strength is $25 \%$. The total column load is the sum of several factors: live load, dead load, wind load, and snow load. If these factors are independent normal variates as:

| Factor | Expected value, in kN | Coefficient of variation |
| :---: | :---: | :---: |
| Live load | 70 | 0.15 |
| Dead load | 90 | 0.05 |
| Wind load | 30 | 0.30 |
| Snow load | 20 | 0.20 |

find
(a) the expected value and coefficient of variation of the total column load, $Y$;
(b) the reliability index and the corresponding risk of failure of the column, if the strength is assumed to be a normal variate independent of load; and
(c) the reliability index and risk of failure, if the strength and load are correlated normal variates with $\rho=0.6$.

Solution. Denote with $X_{1}, X_{2}, X_{3}$, and $X_{4}$ the live, dead, wind and snow loads, respectively. The values of the corresponding standard deviations are $\sigma_{1}=0.15 \times 70=10.5 \mathrm{kN}, \quad \sigma_{2}=0.05 \times 90=4.5 \mathrm{kN}, \quad \sigma_{3}=0.30 \times 30=9 \mathrm{kN} \quad$ and $\sigma_{4}=0.20 \times 20=4 \mathrm{kN}$.
(a) The mean and variance of the total column load, $Y$, are the sums of the load component means and variances, that is
$\mu_{\mathrm{Y}}=\mu_{1}+\mu_{2}+\mu_{3}+\mu_{4}=70+90+30+20=210 \mathrm{kN}$
and
$\sigma_{Y}=\left(\sigma_{1}^{2}+\sigma_{2}^{2}+\sigma_{3}^{2}+\sigma_{4}^{2}\right)^{1 / 2}=\left(10.5^{2}+4.5^{2}+9^{2}+4^{2}\right)^{1 / 2}=15.1 \mathrm{kN}$.
Also, the coefficient of variation of $Y$ is $V_{Y}=15.1 / 210=0.072$.
(b) From $\mu_{X}=336 \mathrm{kN}$ and $V_{X}=0.25, \sigma_{X}=336 \times 0.25=84 \mathrm{kN}$. For normal column strength $X$, the safety margin $S=X-Y$ is a normal variate with mean
$\mu_{S}=\mu_{X}-\mu_{Y}=336-210=126 \mathrm{kN}$
and standard deviation
$\sigma_{S}=\left(\sigma_{X}^{2}+\sigma_{Y}^{2}\right)^{1 / 2}=\left(84^{2}+15.1^{2}\right)^{1 / 2}=85.3 \mathrm{kN}$.
From Eq. (9.1.13), the reliability index is
$\beta=126 / 85.3=1.476$.
From Eq. (9.1.15) the probability of failure is
$p_{f}=1-\Phi(1.476)=0.070$,
with $\Phi(\cdot)$ denoting the $c d f$ of the standard normal variate.
(c) From Eq. (9.1.14)

$$
\beta=\frac{\mu_{X}-\mu_{Y}}{\sqrt{\sigma_{X}^{2}-2 \rho_{X Y} \sigma_{X} \sigma_{Y}+\sigma_{Y}^{2}}}=\frac{336-210}{\sqrt{84^{2}-2 \times 0.6 \times 84 \times 15.1+15.1^{2}}}=1.660 .
$$

The corresponding probability of failure is
$p_{f}=1-\Phi(1.660)=0.048$.
Note that positive correlation between load and strength increases the reliability of the system.
9.9. Earth embankment. For the stability of an earth embankment the overturning moment $e W$, must not exceed the resisting moment $r\left(L_{A} R_{A}+L_{B} R_{B}\right)$, as shown schematically in Figure 9.P1. For the given configuration $L_{A}=21 \mathrm{~m}, L_{B}=4 \mathrm{~m}, r=12 \mathrm{~m}, e=3 \mathrm{~m}$, and $W=2000 \mathrm{kN} / \mathrm{m}^{2}$. Find the reliability of the system if $R_{A}$ and $R_{B}$ are joint normally distributed variates with means 35 and $20 \mathrm{kN} / \mathrm{m}^{2}$, respectively, coefficient of variation of $20 \%$, and coefficient of correlation of 0.7.

Solution. Since the standard deviations of $R_{A}$ and $R_{B}$ are $\sigma_{A}=0.20 \times 35=7.0 \mathrm{kN} / \mathrm{m}^{2}$ and $\sigma_{B}=0.20 \times 20=4.0 \mathrm{kN} / \mathrm{m}^{2}$, respectively, the mean and standard deviation of the resisting moment $X$ are
$\mu_{X}=12 \times(21 \times 35+4 \times 20)=9780.0 \mathrm{kN} / \mathrm{m}^{2}$
and
$\sigma_{X}=\left\{[12 \times(21 \times 7)]^{2}+2 \times 0.7 \times 12 \times(21 \times 7) \times 12 \times(4 \times 4)+[12 \times(4 \times 4)]^{2}\right\}^{1 / 2}=1903.3 \mathrm{kN} / \mathrm{m}^{2}$.
The overturning moment $Y=e W$ has a mean of $6000 \mathrm{kN} / \mathrm{m}^{2}$, and null variance. Thus, the safety margin $S=X-Y$ is a normal variate with mean
$\mu_{S}=\mu_{X}-\mu_{Y}=9780-6000=3780 \mathrm{kN} / \mathrm{m}^{2}$,
and standard deviation
$\sigma_{S}=\left(\sigma_{X}^{2}+\sigma_{Y}^{2}\right)^{1 / 2}=\left(1903.3^{2}+0^{2}\right)^{1 / 2}=1903.3 \mathrm{kN} / \mathrm{m}^{2}$.
From Eq. (9.1.13) the reliability index is
$\beta=3780 / 1903.3=1.986$.
The probability of failure is obtained from Eq. (9.1.15), that is
$p_{f}=1-\Phi(1.986)=0.024$,
with $\Phi(\cdot)$ denoting the $c d f$ of the standard normal variate. Thus,
$r=1-p_{f}=0.976$
is the reliability of the system.
9.10. Slope stability. The wedge method for analysing the stability of an earth slope assumes a linear critical surface, such as AB in Figure 9.P2. The factor of safety is then obtained as
$z=\frac{2 c \sin \theta \cos \varphi}{h \gamma \sin ^{2}[(\theta-\varphi) / 2]}$,
where $c$ is the cohesion parameter, $\varphi$ is the internal angle of friction or friction angle, $\theta$ is the slope angle, $\gamma$ is specific weight, and $h$ is the slope height. If these factors are independent normal variates as:

| Factor | Expected value | Coefficient of variation |
| :---: | :---: | :---: |
| Friction angle | $21^{\circ}$ | 0.12 |
| Cohesion parameter | $15 \mathrm{kN} / \mathrm{m}^{2}$ | 0.40 |
| Specific weight | $20 \mathrm{kN} / \mathrm{m}^{3}$ | 0.10 |

find the risk of failure for a slope with $h=10 \mathrm{~m}$, and $\theta=55^{\circ}$.
Solution. The following limiting state equation is considered
$g(\varphi, c, \gamma)=2 X_{2} \sin \theta \cos X_{1}-h X_{3} \sin ^{2}\left[\left(\theta-X_{1}\right) / 2\right]=0$,
where $X_{1}=\varphi, X_{2}=c$ and $X_{3}=\gamma$ are three random variables with
$\mu_{1}=21 \mathrm{deg}, V_{1}=0.12$, and $\sigma_{1}=2.52 \mathrm{deg} ;$
$\mu_{2}=15 \mathrm{kN} / \mathrm{m}^{2}, V_{2}=0.40$, and $\sigma_{2}=6.0 \mathrm{kN} / \mathrm{m}^{2}$;
$\mu_{3}=20 \mathrm{kN} / \mathrm{m}^{3}, V_{3}=0.10$, and $\sigma_{3}=2 \mathrm{kN} / \mathrm{m}^{3}$.

The partial derivates of the performance function with respect to each of the variables evaluated at the failure point are determined as
$\partial g / \partial X_{i}^{\prime}=\left(\partial g / \partial X_{i}\right)\left(\partial X_{i} / \partial X_{i}^{\prime}\right)=\left(\partial g / \partial X_{i}\right) \sigma_{i}$,
which follows directly from Eq. (9.1.23a). Thus,
$\left(\partial g / \partial X_{1}\right)_{f}=\left\{-2 x_{2} \sin (55) \sin x_{1}+10 x_{3} \sin \left[\left(55-x_{1}\right) / 2\right] \cos \left[\left(55-x_{1}\right) / 2\right]\right\} \times 2.52$,
$\left(\partial g / \partial X_{2}\right)_{f}=\left\{2 \sin (55) \cos x_{1}\right\} \times 6$,
and
$\left(\partial g / \partial X_{3}\right)_{f}=\left\{-10 \sin ^{2}\left[\left(55-x_{1}\right) / 2\right]\right\} \times 2$,
The means are taken as initial values, that is $x_{1}=21 \mathrm{deg}, x_{2}=15 \mathrm{kN} / \mathrm{m}^{2}$ and $x_{3}=20 \mathrm{kN} / \mathrm{m}^{3}$. Hence, $\left(\partial g / \partial X_{1}\right)_{f}=118.72,\left(\partial g / \partial X_{2}\right)_{f}=9.177$ and $\left(\partial g / \partial X_{3}\right)_{f}=-1.710$.
Also, $\Sigma\left(\partial g / \partial X_{i}\right)_{f}^{2}=14182.4$. From Eq. (9.1.31),
$\alpha_{1}=118.72 / \sqrt{ } 14182.4=0.997$,
$\alpha_{2}=9.177 / \sqrt{ } 14182.4=0.077$,
and
$\alpha_{3}=-1.710 / \sqrt{ } 14182.4=-0.014$.
Thus, the new failure point is given by
$x_{1(\text { new })}=\mu_{1}-\alpha_{1} \sigma_{1} \beta=21-(0.997 \times 2.52) \beta=21-2.512 \beta$,
$x_{2(\text { new })}=\mu_{2}-\alpha_{2} \sigma_{2} \beta=15-(0.077 \times 6.0) \beta=15-0.462 \beta$
and
$x_{3 \text { (new) }}=\mu_{3}-\alpha_{3} \sigma_{3} \beta=20-(-0.014 \times 2.0) \beta=20+0.028 \beta$.
The limit state equation
$2(15-0.462 \beta) \sin (55) \cos (21-2.512 \beta)-$

$$
-10 \times(20+0.028 \beta) \sin ^{2}[(55-21+2.512 \beta) / 2]=0
$$

is solved numerically for $\beta$ to get $\beta=1.94$.
To perform the second iteration, one makes use of the new failure point, that is
$x_{1(\text { new })}=21-2.512 \beta=21-2.512 \times 1.94=16.1$,
$x_{2(\text { new })}=15-0.462 \beta=15-0.462 \times 1.94=14.1$
and
$x_{3 \text { (new) }}=20+0.028 \beta=20+0.028 \times 1.94=20.1$.
Then the values of the partial derivates are computed as $\left(\partial g / \partial X_{1}\right)_{f}=142.38$, $\left(\partial g / \partial X_{2}\right)_{f}=9.443$ and $\left(\partial g / \partial X_{3}\right)_{f}=-2.214$. Also, $\Sigma\left(\partial g / \partial X_{i}\right)_{f}^{2}=20365.8$. Hence,
$\alpha_{1}=142.38 / \sqrt{ } 20365.8=0.998$,
$\alpha_{2}=9.443 / \sqrt{ } 20365.8=0.066$,
and
$\alpha_{3}=-2.214 / \sqrt{ } 20365.8=-0.016$.
This gives a new failure point
$x_{1 \text { (new) }}=21-(0.998 \times 2.52) \beta=21-2.515 \beta$,
$x_{2(\text { new })}=15-(0.066 \times 6.0) \beta=15-0.396 \beta$
and
$x_{3(\text { new })}=20-(-0.016 \times 2.0) \beta=20+0.032 \beta$.
Accordingly, the new limit state equation is
$2(15-0.396 \beta) \sin (55) \cos (21-2.515 \beta)-$

$$
-10 \times(20+0.032 \beta) \sin ^{2}[(55-21+2.515 \beta) / 2]=0
$$

Table 9.S10 - Slope stability.

| Design data | Unit | Mean | Coefficient of Variation | Standard Deviation |
| :---: | :---: | :---: | :---: | :---: |
| Friction angle, $\varphi$ | deg | 21 | 0.12 | 2.52 |
| Cohesion, c | kN/m ${ }^{2}$ | 15 | 0.40 | 6.0 |
| Specific weight, $\gamma$ | $\mathrm{kN} / \mathrm{m}^{3}$ | 20 | 0.10 | 2.0 |
| Slope height, $h$ | m | 10 | - | - |
| Slope angle, $\theta$ | deg | 55 | - | - |
| Nominal capacity, $x$ | kN/m ${ }^{2}$ | 22.9 | - | - |
| Nominal demand, $y$ | $\mathrm{kN} / \mathrm{m}^{2}$ | 17.1 | - | - |
| Nominal safety factor, z | - | 1.34 | - | - |
| Limit state of interest is Iteration process | $g(\varphi, C, \theta)=2 c \sin \theta \cos \varphi-h \gamma \sin ^{2}[(\theta-\varphi) / 2]=0$ |  |  |  |
| Initial $\varphi=\chi_{1 f}$ | 21 | 16.1 | 16.0 | 16.0 |
| Initial $c=x_{2 f}$ | 15 | 14.1 | 14.2 | 14.2 |
| Initial $\gamma=x_{3 f}$ | 20 | 20.1 | 20.1 | 20.1 |
| $(\partial g / \partial \varphi)_{\text {f }}$ | 118.72 | 142.38 | 143.00 | 143.01 |
| $(\partial g / \partial)_{f}$ | 9.177 | 9.443 | 9.450 | 9.450 |
| $(\partial g / \partial \gamma)_{f}$ | -1.710 | -2.214 | -2.231 | -2.231 |
| $\Sigma\left(\partial g / \partial X_{i}\right){ }^{2}$ | 14182.4 | 20365.8 | 20543.4 | 20547.5 |
| $\alpha_{1 f}$ | 0.997 | 0.998 | 0.998 | 0.998 |
| $\alpha_{2 f}$ | 0.077 | 0.066 | 0.066 | 0.066 |
| $\alpha_{3 f}$ | -0.014 | -0.016 | -0.016 | -0.016 |
| New $\varphi=\chi_{1 f}$ | 16.1 | 16.0 | 16.0 | 16.0 |
| New $c=x_{2 f}$ | 14.1 | 14.2 | 14.2 | 14.2 |
| New $\gamma=\chi_{3 f}$ | 20.1 | 20.1 | 20.1 | 20.1 |
| $g($. | 7E-07 | 1E-06 | 6E-06 | 6E-06 |
| yields $\beta=$ | 1.94 | 2.00 | 2.00 | 2.00 |
| Reliability, $\Phi(\beta)$ | 0.974 | 0.977 | 0.977 | 0.977 |
| Risk, 1- $\Phi(\beta)$ | 0.026 | 0.023 | 0.023 | $\underline{0.023}$ |

Hence $\beta=2.00$. Further iterations indicate that $\beta=2.00$ is the required reliability index. This yields a probability of failure of
$p_{f}=1-\Phi(2.00)=0.023$,
with $\Phi(\cdot)$ denoting the $c d f$ of the standard normal variate. The corresponding reliability is
$r=1-p_{f}=0.977$.
Detailed computations are shown in Table 9.S10.
9.11. Elastic collapse of a steel beam. Consider a simply supported steel beam with normally distributed strength $X$, with mean of $25 \mathrm{KN} / \mathrm{cm}^{2}$, and coefficient of variation of $15 \%$. The bending moment $Y$ is also a normal variate with mean 900 kN cm , and coefficient of variation of $20 \%$. Find the reliability of the beam if its section modulus $W$ is normally distributed with mean $20 \mathrm{~cm}^{3}$, and coefficient of variation of $5 \%$. Note that the limit state of interest is given by $Y / W-X=0$, and assume mutually independent $X, Y$ and W.

Solution. The following limiting state equation is considered
$g(Y, W, X)=x_{1} / x_{2}-x_{3}=0$,
where $X_{1}=Y, X_{2}=W$ and $X_{3}=X$ are three random variables with
$\mu_{1}=900 \mathrm{kN} \times \mathrm{cm}, V_{1}=0.20$, and $\sigma_{1}=180 \mathrm{kN} \times \mathrm{cm}$;
$\mu_{2}=20 \mathrm{~cm}^{3}, V_{2}=0.05$, and $\sigma_{2}=1 \mathrm{~cm}^{3}$;
$\mu_{3}=25 \mathrm{kN} / \mathrm{cm}^{2}, V_{3}=0.15$, and $\sigma_{3}=3.75 \mathrm{kN} / \mathrm{cm}^{2}$.
The partial derivates of the performance function with respect to each of the variables evaluated at the failure point are determined as
$\partial g / \partial X_{i}^{\prime}=\left(\partial g / \partial X_{i}\right)\left(\partial X_{i} / \partial X_{i}^{\prime}\right)=\left(\partial g / \partial X_{i}\right) \sigma_{i}$,
which follows directly from Eq. (9.1.23a). Thus,
$\left(\partial g / \partial X_{1}\right)_{f}=\left(1 / x_{2}\right) \times 180$,
$\left(\partial g / \partial X_{2}\right)_{f}=\left(-x_{1} / x_{2}^{2}\right) \times 1$
and
$\left(\partial g / \partial X_{3}\right)_{f}=(-1) \times 3.75$,
The means are taken as initial values, that is $x_{1}=900 \mathrm{kN} \times \mathrm{cm}, x_{2}=20 \mathrm{~cm}^{3}$ and $x_{3}=25 \mathrm{kN} / \mathrm{cm}^{3}$. Hence, $\left(\partial g / \partial X_{1}\right)_{f}=9,\left(\partial g / \partial X_{2}\right)_{f}=-2.25$ and $\left(\partial g / \partial X_{3}\right)_{f}=-3.75$. Also, $\Sigma\left(\partial g / \partial X_{i}\right)_{f}^{2}=100.1$. From Eq. (9.1.31)
$\alpha_{1}=9 / \sqrt{ } 100.1=0.899$,
$\alpha_{2}=-2.25 / \sqrt{ } 100.1=-0.225$,
and
$\alpha_{3}=-3.75 / \sqrt{ } 100.1=-0.375$.
Thus, the new failure point is given by
$x_{1 \text { (new) }}=\mu_{1}-\alpha_{1} \sigma_{1} \beta=900-(0.899 \times 180) \beta=900-161.8 \beta$,
$x_{2(\text { new })}=\mu_{2}-\alpha_{2} \sigma_{2} \beta=20-(-0.225 \times 1) \beta=20+0.225 \beta$
and
$x_{3(\text { new })}=\mu_{3}-\alpha_{3} \sigma_{3} \beta=25-(-0.375 \times 3.75) \beta=25+1.406 \beta$.
The limit state equation
$(900-161.8 \beta) /(20+0.225 \beta)-(25+1.406 \beta)=0$
is solved numerically for $\beta$ to get $\beta=2.04$.
To perform the second iteration, one makes use of the new failure point, that is
$x_{1(\text { new })}=900-161.8 \beta=900-161.8 \times 2.04=570.1$,
$x_{2(\text { new })}=20+0.225 \beta=20-0.225 \times 2.04=20.5$
and
$x_{3(\text { new })}=25+1.406 \beta=25+1.406 \times 2.04=27.9$.
Then the values of the partial derivates are computed as
$\left(\partial g / \partial X_{1}\right)_{f}=8.798$,
$\left(\partial g / \partial X_{2}\right)_{f}=-1.362$
and
$\left(\partial g / \partial X_{3}\right)_{f}=-3.750$.
Also, $\Sigma\left(\partial g / \partial X_{i}\right)_{f}^{2}=93.3$. Hence,
$\alpha_{1}=8.798 / \sqrt{93.3}=0.911$,
$\alpha_{2}=-1.362 / \sqrt{ } 93.3=-0.141$,
and
$\alpha_{3}=-3.750 / \sqrt{ } 93.3=-0.388$.
This gives a new failure point
$x_{1 \text { (new) }}=900-(0.911 \times 180) \beta=900-164.0 \beta$,
$x_{2(\text { new })}=20-(-0.141 \times 1) \beta=20+0.141 \beta$
and
$x_{3 \text { (new) }}=25-(-0.388 \times 3.75) \beta=25+1.455 \beta$.
Accordingly, the new limit state equation is
$(900-164.0 \beta) /(20+0.141 \beta)-(25+1.455 \beta)=0$.
Hence, $\beta=2.03$. Further iterations indicate that $\beta=2.03$ is the required reliability index. This yields a probability of failure of
$p_{f}=1-\Phi(2.03)=0.021$,
with $\Phi(\cdot)$ denoting the $c d f$ of the standard normal variate, and a reliability of $r=1-p_{f}=0.979$.
Detailed computations are shown in Table 9.S11.

Table 9.S11 - Elastic collapse of a steel beam.

| Design data | Unit | Mean | Coefficient of Variation | Standard Deviation |
| :---: | :---: | :---: | :---: | :---: |
| Bending moment, $Y$ | kN cm | 900 | 0.20 | 180 |
| Section modulus, $W$ | $\mathrm{cm}^{3}$ | 20 | 0.05 | 1 |
| Strength, $X$ | $\mathrm{kN} / \mathrm{cm}^{2}$ | 25 | 0.15 | 3.75 |
| Limit state of interest is Iteration process | $g(Y, W, X)=y / W-x=0$ |  |  |  |
| Initial $y=x_{1 f}$ | 900 | 570.1 | 567.1 | 566.7 |
| Initial $w=x_{2 f}$ | 20 | 20.5 | 20.3 | 20.3 |
| Initial $x=x_{3 f}$ | 25 | 27.9 | 28 | 28 |
| $(\partial g / \partial M)_{\mathrm{f}}$ | 9.000 | 8.798 | 8.873 | 8.872 |
| $(\partial g / \partial W)_{f}$ | -2.250 | -1.362 | -1.378 | -1.377 |
| $(\partial g / \partial X)_{f}$ | -3.750 | -3.750 | -3.750 | -3.750 |
| $\Sigma\left(\partial g / \partial X_{i}\right){ }^{2}$ | 100.1 | 93.3 | 94.7 | 94.7 |
| $\alpha_{1 f}$ | 0.899 | 0.911 | 0.912 | 0.912 |
| $\alpha_{2 f}$ | -0.225 | -0.141 | -0.142 | -0.142 |
| $\alpha_{3 f}$ | -0.375 | -0.388 | -0.385 | -0.385 |
| New $m=\chi_{1 f}$ | 570.1 | 567.1 | 566.7 | 566.7 |
| New $w=x_{2 f}$ | 20.5 | 20.3 | 20.3 | 20.3 |
| New $x=x_{3 f}$ | 27.9 | 28.0 | 27.9 | 27.9 |
| $g()=.x_{1 f} f \chi_{2 f}-X_{3 f}$ | 2E-07 | 9E-08 | 9E-08 | -7E-07 |
| yields $\beta=$ | 2.04 | 2.03 | 2.03 | $\underline{2.03}$ |
| Reliability, $\Phi(\beta)$ | 0.979 | 0.979 | 0.979 | $\underline{0.979}$ |
| Risk, 1-Ф( $\beta$ ) | 0.021 | 0.021 | 0.021 | $\underline{0.021}$ |

9.12. Flexure formula. Consider a timber beam subject to flexure. The stress at the extreme fiber at a distance $X_{2}$ from the neutral axis acted upon by a bending moment $X_{3}$ is given by $X_{2} X_{3} / X_{4}$, where $X_{4}$ denotes the moment of inertia of the section. We assume that the factors are normal variates as

| Factor | Expected value | Coefficient of variation |
| :---: | :---: | :---: |
| Bending moment, $X_{3}$ | 6 kN cm | 0.25 |
| Moment of inertia, $X_{4}$ | $90 \mathrm{~cm}^{4}$ | 0.10 |
| Distance from neutral axis, $X_{2}$ | 20 cm | 0.05 |

and assume that $X_{2}, X_{3}$, and $X_{4}$ are independent each other. Find the reliability of the system if the capacity $X_{1}$ of the beam is a normal variate with a mean of $4 \mathrm{kN} / \mathrm{cm}^{2}$, and a coefficient of variation of 30\%.

Solution. The following limiting state equation is considered
$g\left(X_{1}, X_{2}, X_{3}, X_{4}\right)=x_{1}-x_{2} x_{3} / x_{4}=0$,
where $X_{1}, X_{2}, X_{3}$ and $X_{4}$ are four random variables, with
$\mu_{1}=4 \mathrm{kN} / \mathrm{cm}^{2}, V_{1}=0.30$, and $\sigma_{1}=1.2 \mathrm{kN} / \mathrm{cm}^{2}$;
$\mu_{2}=20 \mathrm{~cm}, V_{2}=0.05$, and $\sigma_{2}=1 \mathrm{~cm}$;
$\mu_{3}=6 \mathrm{kN} \times \mathrm{cm}, V_{3}=0.25$, and $\sigma_{3}=1.5 \mathrm{kN} \times \mathrm{cm}$.
$\mu_{4}=90 \mathrm{~cm}^{4}, V_{3}=0.10$, and $\sigma_{4}=9 \mathrm{~cm}^{4}$.
The partial derivates of the performance function with respect to each of the variables evaluated at the failure point are determined as
$\partial g / \partial X_{i}^{\prime}=\left(\partial g / \partial X_{i}\right)\left(\partial X_{i} / \partial X_{i}^{\prime}\right)=\left(\partial g / \partial X_{i}\right) \sigma_{i}$,
which follows directly from Eq. (9.1.23a). Thus,
$\left(\partial g / \partial X_{1}\right)_{f}=(1) \times 1.2$,
$\left(\partial g / \partial X_{2}\right)_{f}=\left(-x_{3} / x_{4}\right) \times 1$,
$\left(\partial g / \partial X_{3}\right)_{f}=\left(-x_{2} / x_{4}\right) \times 1.5$
and
$\left(\partial g / \partial X_{4}\right)_{f}=\left(x_{2} x_{3} / x_{4}\right) \times 9$,
The means are taken as initial values, that is $x_{1}=4 \mathrm{kN} / \mathrm{cm}^{2}, x_{2}=20 \mathrm{~cm}, x_{3}=6 \mathrm{kN} \times \mathrm{cm}$ and $x_{4}=90 \mathrm{~cm}^{4}$. Hence,
$\left(\partial g / \partial X_{1}\right)_{f}=1.2$,
$\left(\partial g / \partial X_{2}\right)_{f}=-0.067$,
$\left(\partial g / \partial X_{3}\right)_{f}=-0.333$
and
$\left(\partial g / \partial X_{4}\right)_{f}=0.133$.
Also, $\Sigma\left(\partial g / \partial X_{i}\right)_{f}^{2}=1.573$. From Eq. (9.1.31),
$\alpha_{1}=1.2 / \sqrt{ } 1.573=0.957$,
$\alpha_{2}=-0.067 / \sqrt{ } 1.573=-0.053$,
$\alpha_{3}=-0.333 / \sqrt{ } 1.573=-0.266$,
and
$\alpha_{4}=0.133 / \sqrt{1.573}=0.106$.
Thus, the new failure point is given by

$$
\begin{aligned}
& x_{1(\text { new })}=\mu_{1}-\alpha_{1} \sigma_{1} \beta=4-(0.957 \times 1.2) \beta=4-1.148 \beta, \\
& x_{2(\text { new })}=\mu_{2}-\alpha_{2} \sigma_{2} \beta=20-(-0.053 \times 1) \beta=20+0.053 \beta, \\
& x_{3(\text { new })}=\mu_{3}-\alpha_{3} \sigma_{3} \beta=6-(-0.266 \times 1.5) \beta=6+0.399 \beta
\end{aligned}
$$

and
$x_{4 \text { (new) }}=\mu_{4}-\alpha_{4} \sigma_{4} \beta=90-(0.106 \times 9) \beta=90-0.954 \beta$.
The limit state equation
$(4-1.148 \beta)-(20+0.053 \beta)(6+0.399 \beta) /(90+0.954 \beta)=0$,
is solved numerically for $\beta$ to get $\beta=2.121$.
To perform the second iteration, one makes use of the new failure point, that is
$x_{1(\mathrm{new})}=4-1.148 \beta=4-1.148 \times 2.121=1.57$,
$x_{2 \text { (new) }}=20+0.053 \beta=20-0.053 \times 2.121=20.11$,
$x_{3(\text { new })}=6+0.399 \beta=6+0.399 \times 2.121=6.87$
and
$x_{4(\text { new })}=90-0.954 \beta=90-0.954 \times 2.121=88.0$.
Then the values of the partial derivates are computed as
$\left(\partial g / \partial X_{1}\right)_{f}=1.2$,
$\left(\partial g / \partial X_{2}\right)_{f}=-0.078$,
$\left(\partial g / \partial X_{3}\right)_{f}=-0.343$
and
$\left(\partial g / \partial X_{4}\right)_{f}=0.160$.
Also, $\Sigma\left(\partial g / \partial X_{i}\right)_{f}^{2}=1.589$. Hence, from Eq. (9.1.31),
$\alpha_{1}=1.2 / \sqrt{ } 1.589=0.952$,
$\alpha_{2}=-0.078 / \sqrt{ } 1.589=-0.062$,
$\alpha_{3}=-0.343 / \sqrt{ } 1.589=-0.272$,
and
$\alpha_{4}=0.160 / \sqrt{ } 1.589=0.127$.
Thus, the new failure point is given by
$x_{1 \text { (new) }}=\mu_{1}-\alpha_{1} \sigma_{1} \beta=4-(0.952 \times 1.2) \beta=4-1.142 \beta$,
$x_{2(\text { new })}=\mu_{2}-\alpha_{2} \sigma_{2} \beta=20-(-0.062 \times 1) \beta=20+0.062 \beta$,
$x_{3(\text { new })}=\mu_{3}-\alpha_{3} \sigma_{3} \beta=6-(-0.272 \times 1.5) \beta=6+0.408 \beta$
and
$x_{4(\text { new })}=\mu_{4}-\alpha_{4} \sigma_{4} \beta=90-(0.127 \times 9) \beta=90-1.143 \beta$.
Accordingly, the new limit state equation is
$(4-1.142 \beta)-(20+0.062 \beta)(6+0.408 \beta) /(90+1.143 \beta)=0$,
Hence $\beta=2.120$. Further iterations indicate that $\beta=2.120$ is the required reliability index. This yields a probability of failure of
$p_{f}=1-\Phi(2.120)=0.017$,
with $\Phi(\cdot)$ denoting the $c d f$ of the standard normal variate, and a reliability of $r=1-p_{f}=0.983$.
Detailed computations are shown in Table 9.S12.

Table 9.S12 - Flexure formula.

| Design data | Unit | Mean | Coefficient of Variation | Standard Deviation |
| :---: | :---: | :---: | :---: | :---: |
| Section modulus, $X_{4}$ | $\mathrm{cm}^{4}$ | 90 | 0.10 | 9 |
| Bending moment, $X_{3}$ | kN cm | 6 | 0.25 | 1.5 |
| Distance, $X_{2}$ | cm | 20 | 0.05 | 1 |
| Capacity, $X_{1}$ | $\mathrm{kN} / \mathrm{cm}^{2}$ | 4 | 0.30 | 1.2 |
| Limit state of interest is Iteration process | $g\left(x_{1}, x_{2}, x_{3}, x_{4}\right)=x_{1}-x_{2} x_{3} / x_{4}=0$ |  |  |  |
| Initial $X_{4 f}$ | 90 | 88.0 | 87.6 | 87.5 |
| Initial $x_{3 f}$ | 6 | 6.85 | 6.87 | 6.87 |
| Initial $x_{2 f}$ | 20 | 20.11 | 20.13 | 20.13 |
| Initial $x_{1 f}$ | 4 | 1.57 | 1.58 | 1.58 |
| $\left(\partial g / \partial X_{4}\right)_{\mathrm{f}}$ | 0.133 | 0.160 | 0.162 | 0.162 |
| $\left(\partial g / \partial X_{3}\right)_{\mathrm{f}}$ | -0.333 | -0.343 | -0.345 | -0.345 |
| $\left(\partial g / \partial X_{2}\right)_{\mathrm{f}}$ | -0.067 | -0.078 | -0.078 | -0.078 |
| $\left(\partial g / \partial X_{1}\right)_{\mathrm{f}}$ | 1.200 | 1.200 | 1.200 | 1.200 |
| $\Sigma\left(\partial g / \partial X_{i}\right){ }^{2}$ | 1.573 | 1.589 | 1.591 | 1.592 |
| $\alpha_{4 f}$ | 0.106 | 0.127 | 0.129 | 0.129 |
| $\alpha_{3 f}$ | -0.266 | -0.272 | -0.273 | -0.273 |
| $\alpha_{2 f}$ | -0.053 | -0.062 | -0.062 | -0.062 |
| $\alpha_{1 f}$ | 0.957 | 0.952 | 0.951 | 0.951 |
| New $\chi_{4 f}$ | 88.0 | 87.6 | 87.5 | 87.5 |
| New $x_{3 f}$ | 6.85 | 6.87 | 6.87 | 6.87 |
| New $\chi_{2 f}$ | 20.11 | 20.13 | 20.13 | 20.13 |
| New $\chi_{1 f}$ | 1.57 | 1.58 | 1.58 | 1.58 |
| $g()=.\mathrm{x}_{2} \mathrm{f}_{3} / \mathrm{X}_{4 \mathrm{f}}-\chi_{1 f}$ | -6E-07 | 8E-09 | 9E-09 | 9E-09 |
| yields $\beta=$ | 2.121 | 2.120 | 2.120 | 2.120 |
| Reliability, $\Phi(\beta)$ | 0.983 | 0.983 | 0.983 | 0.983 |
| Risk, 1-Ф( $\beta$ ) | 0.017 | 0.017 | 0.017 | $\underline{0.017}$ |

9.13. Surveying using Geosatellite Positioning System. The values of latitude $Y_{\text {GPS }}$ and longitude $X_{\text {GPS }}$ obtained by GPS readings at a point are affected by a certain random multiplicative error, $Z$. Thus, $X_{\mathrm{GPS}}=Z x$ and $Y_{\mathrm{GPS}}=Z y$, respectively, with $x$ and $y$ denoting longitude and latitude of the point. Find the reliability of measuring the planar distance $w$ between two points with a tolerance of $3 \%$ if $Z$ is a lognormally distributed variate with unit mean and coefficient of variation of 0.02 , assuming that all (four) readings used in measuring the distance are independent of each other (this reliability can be evaluated as $\left.\operatorname{Pr}\left[0.97<W_{G P S} / w \leq 1.03\right]\right)$.

Solution. The planar distance between two points with coordinates $\left(x_{1}, y_{1}\right)$ and $\left(x_{2}, y_{2}\right)$ is given by
$w=\left[\left(x_{2}-x_{1}\right)^{2}+\left(y_{2}-y_{1}\right)^{2}\right]^{1 / 2} ;$
while that obtained by GPS readings is
$W_{\mathrm{GPS}}=\left[\left(Z x_{2}-Z x_{1}\right)^{2}+\left(Z y_{2}-Z y_{1}\right)^{2}\right]^{1 / 2}=Z w$.

Thus, $W_{\text {GPs }} / w=Z$. The required reliability is computed as
$r=\operatorname{Pr}\left[z_{\text {sup }}<W_{G P S} / w \leq z_{\text {inf }}\right]=\Phi\left[\left(\ln z_{\text {sup }}-\mu_{\ln (Z))}\right) / \sigma_{\ln (z)}\right]-\Phi\left[\left(\ln z_{\text {inf }}-\mu_{\ln (Z))} / \sigma_{\ln (Z)}\right]\right.$.
with $\Phi[\cdot]$ denoting the $c d f$ of the standard normal variate. Since, $Z_{\text {sup }}=0.97$ and $z_{\text {inf }}=1.03, \ln z_{\text {sup }}=0.0296$ and $\ln z_{\text {inf }}=-0.0305$. Also,
$\sigma_{\ln (Z)}=\left[\ln \left(1+V_{Z}^{2}\right)\right]^{1 / 2}=\left[\ln \left(1+0.02^{2}\right)\right]^{1 / 2}=0.0200$
and
$\mu_{\ln (Z)}=\ln \left(\mu_{Z}\right)-0.5 \sigma_{\ln (Z)^{2}}{ }^{2}=0.0002$. Hence,
$r=\Phi[(0.0296+0.0002) / 0.02]-\Phi[(-0.0305+0.0002) / 0.02]=$

$$
=\Phi(1.488)-\Phi(-1.513)=0.932-0.065=0.867
$$

9.14. Column load. The strength of column of a building is normally distributed with mean of 336 kN , and coefficient of variation of $25 \%$. The total column load is the sum of several components, say, the live load, dead load, wind load, and snow load. If these factors are independent variates as:

| Factor | Expected value, in kN | Coefficient of variation |
| :---: | :---: | :---: |
| Normal live load | 70 | 0.15 |
| Normal dead load | 90 | 0.05 |
| Weibull wind load | 30 | 0.30 |
| Gumbel snow load | 20 | 0.20 |

find the reliability index and the risk of failure of the column.
Solution. Since $\mu_{X}=336 \mathrm{kN}$ and $V_{X}=0.25, \sigma_{X}=0.25 \times 336=84 \mathrm{kN}$ is the standard deviation of column strength, $X$. Denote $Y_{1}, Y_{2}, Y_{3}$, and $Y_{4}$ the live, dead, wind and snow loads, respectively. The values of the corresponding standard deviations are given by
$\sigma_{1}=0.15 \times 70=10.5 \mathrm{kN}$,
$\sigma_{2}=0.05 \times 90=4.5 \mathrm{kN}$,
$\sigma_{3}=0.30 \times 30=9 \mathrm{kN}$
and
$\sigma_{4}=0.20 \times 20=4 \mathrm{kN}$.
From Eqs. (4.2.15) and (4.2.16)
$F_{Y_{3}}\left(y_{3}\right)=1-\exp \left[-\left(\frac{y_{3}}{\lambda}\right)^{\gamma}\right]$ and $f_{Y_{3}}\left(y_{3}\right)=\frac{\gamma}{\lambda}\left(\frac{y_{3}}{\lambda}\right)^{\gamma-1} \exp \left[-\left(\frac{y_{3}}{\lambda}\right)^{\gamma}\right]$,
where $\gamma$ is found by solving numerically Eq. (4.2.17c) for $\gamma$. Thus,
$\Gamma(1+2 / \gamma) / \Gamma^{2}(1+1 / \gamma)-\left(1+V_{3}{ }^{2}\right)=0$,
that is
$\Gamma(1+2 / \gamma) / \Gamma^{2}(1+1 / \gamma)-\left(1+0.3^{2}\right)=0$,
which yields $\gamma=3.71$. From Eq. (4.217a)
$\lambda=\mu_{1} / \Gamma(1+1 / \gamma)=30 / \Gamma(1+1 / \gamma)=30 / \Gamma(1.27)=33.24 \mathrm{kN}$.

From Eqs. (7.2.17) and (7.2.18)

$$
F_{Y_{4}}\left(y_{4}\right)=\exp \left[-\exp \left(-\frac{y_{4}-b}{\alpha}\right)\right] \text { and } f_{Y_{4}}\left(y_{4}\right)=\frac{1}{\alpha} \exp \left[-\frac{y_{4}-b}{\alpha}-\exp \left(-\frac{y_{4}-b}{\alpha}\right)\right] .
$$

The values of $\alpha$ and $b$ are evauated from Eqs. (7.2.21) and (7.2.22). Thus, $\alpha=(\sqrt{ } 6 / \pi) \sigma_{4}=0.780 \times 4=3.12 \mathrm{kN}$
and
$b=\mu_{4}-0.5772 \alpha=20-0.5772 \times 3.12=18.20 \mathrm{kN}$.
Because $Y_{3}$ has the Weibull distribution, the equivalent normal variate is determined using $y_{3}{ }^{*}=20 \mathrm{kN}$ as a starting point. Thus,

$$
\begin{aligned}
\sigma_{3}{ }^{*}= & \phi\left\{\Phi^{-1}\left[F_{Y_{3}}\left(y_{3}{ }^{*}\right)\right]\right\} / f_{Y_{3}}\left(y_{3}{ }^{*}\right)=\phi\left\{\Phi^{-1}\left[F_{Y_{3}}(20)\right]\right\} / f_{Y_{3}}(20)= \\
& =\phi\left[\Phi^{-1}(0.1407)\right] / 0.0242=\phi(-1.077) / 0.0242=0.223 / 0.0242=9.230
\end{aligned}
$$

and
$\mu_{3}{ }^{*}=y_{3}{ }^{*}-\sigma_{3}{ }^{*} \Phi^{-1}\left[F_{Y_{3}}\left(X_{3}{ }^{*}\right)\right]=20-9.230 \times(-1.077)=29.942$.
Similarly for $Y_{4} \sim G u m b e l(18.2 \mathrm{kN}, 3.12 \mathrm{kN})$ the equivalent normal variate is determined using $y_{4}{ }^{*}=15 \mathrm{kN}$ as a starting point. Thus,

$$
\begin{aligned}
\sigma_{4}^{*}= & \phi\left\{\Phi^{-1}\left[F_{Y_{4}}\left(y_{4}{ }^{*}\right)\right]\right\} / f_{Y_{4}}\left(y_{4}{ }^{*}\right)=\phi\left\{\Phi^{-1}\left[F_{Y_{4}}(15)\right]\right\} / f_{Y_{4}}(15)= \\
& =\phi\left[\Phi^{-1}(0.0614)\right] / 0.0550=\phi(-1.543) / 0.0550=0.121 / 0.0550=2.208
\end{aligned}
$$

and
$\mu_{4}{ }^{*}=y_{4}{ }^{*}-\sigma_{4}{ }^{*} \Phi^{-1}\left[F_{Y_{4}}\left(y_{4}{ }^{*}\right)\right]=15-2.208 \times(-1.543)=12.620$.
From Eq. (9.1.13)

$$
\begin{aligned}
& \beta=\left[\mu_{X}-\left(\mu_{1}+\mu_{2}+\mu_{3}^{*}+\mu_{4}^{*}\right)\right] /\left(\sigma_{X}^{2}-\sigma_{1}^{2}+\sigma_{2}^{2}+\sigma_{3}^{* 2}+\sigma_{4}^{*^{2}}\right)^{1 / 2}= \\
&=(336-70-90-29.94-12.62) /\left(84^{2}-10.5^{2}+4.5^{2}+9.23^{2}+2.21^{2}\right)^{1 / 2}= \\
&=1.546 .
\end{aligned}
$$

This procedure is iterated as shown in Table 9.S14a until convergence to the value of $\beta=1.451$ is achieved.

Table 9.S14a - Column load. Analytical solution.

| Design data | Unit | Mean | Coefficient of Variation | Standard Deviation | Scale Paramete r | Shape Paramete $r$ | Location Parameter |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Normal Stength, $X$ | kN | 336 | 0.25 | 84.0 |  |  |  |
| Normal Live Load, $Y_{1}$ | kN | 70 | 0.15 | 10.5 |  |  |  |
| Normal Dead Load, $Y_{2}$ | kN | 90 | 0.05 | 4.5 |  |  |  |
| Weibull Wind Load, $Y_{3}$ | kN | 30 | 0.30 | 9.0 | 33.24 | 3.71 |  |
| Gumbel Snow Load, $Y_{4}$ Iteration process | kN | 20 | 0.20 | 4.0 | 3.12 |  | 18.20 |
| $y_{3}$ | = | 20.0 | 25.0 | 30.0 | 32.0 | 33.0 | 33.5 |
| $y_{4}$ | $=$ | 15.0 | 18.0 | 20.0 | 21.0 | 22.0 | 22.5 |
| $F\left(y_{3}{ }^{*}\right)$ | $=$ | 0.1407 | 0.2934 | 0.4952 | 0.5805 | 0.6224 | 0.6429 |
| $f\left(y_{3}{ }^{*}\right)$ | = | 0.0242 | 0.0365 | 0.0427 | 0.0423 | 0.0414 | 0.0408 |
| $\Phi^{-1}\left[F\left(y_{3}{ }^{*}\right)\right]$ | $=$ | -1.077 | -0.543 | -0.012 | 0.203 | 0.312 | 0.366 |
| $\phi\left\{\Phi^{-1}\left[F\left(y_{3}{ }^{*}\right)\right]\right\}$ | = | 0.223 | 0.344 | 0.399 | 0.391 | 0.380 | 0.373 |
| $\mu_{3}{ }^{*}$ | $=$ | 29.942 | 30.130 | 30.113 | 30.122 | 30.137 | 30.148 |
| $\sigma_{3}{ }^{*}$ | $=$ | 9.230 | 9.441 | 9.338 | 9.240 | 9.182 | 9.151 |
| $F\left(y_{4}{ }^{*}\right)$ | = | 0.0614 | 0.3443 | 0.5704 | 0.6653 | 0.7440 | 0.7773 |
| $f\left(y_{4}{ }^{*}\right)$ | = | 0.0550 | 0.1177 | 0.1027 | 0.0869 | 0.0705 | 0.0628 |
| $\Phi^{-1}\left[F\left(y_{4}{ }^{*}\right)\right]$ | = | -1.543 | -0.401 | 0.177 | 0.427 | 0.656 | 0.763 |
| $\phi\left\{\Phi^{-1}\left[F\left(y_{4}^{*}\right)\right]\right\}$ | = | 0.121 | 0.368 | 0.393 | 0.364 | 0.322 | 0.298 |
| $\mu_{4}{ }^{*}$ | $=$ | 12.620 | 17.839 | 19.969 | 20.818 | 21.570 | 21.917 |
| $\sigma_{4}{ }^{*}$ | $=$ | 2.208 | 3.128 | 3.825 | 4.190 | 4.561 | 4.749 |
| $\mu_{X}-\mu_{1}-\mu_{2}-\mu_{3}{ }^{*}-\mu_{4}{ }^{*}$ | $=$ | 133.44 | 128.03 | 125.92 | 125.06 | 124.29 | 123.93 |
| $\sigma_{X}{ }^{2}+\sigma_{1}{ }^{2}+\sigma_{2}{ }^{2}+\sigma_{3}{ }^{*}+\sigma_{4}{ }^{2}$ | = | 7276.57 | 7285.42 | 7288.33 | 7289.43 | 7291.62 | 7292.80 |
| yields $\beta$ | = | 1.564 | 1.500 | 1.475 | 1.465 | 1.456 | 1.451 |
| Reliability, $\Phi(\beta)$ | = | 0.941 | 0.933 | 0.930 | 0.929 | 0.927 | 0.927 |
| Risk, 1-Ф( $\beta$ ) | = | 0.059 | 0.067 | 0.070 | 0.071 | 0.073 | $\underline{0.073}$ |

From Eq. (9.1.15) the probability of failure is
$p_{f}=1-\Phi(1.451)=1-0.927=0.073$.
Simulations are performed by generating values of $X, Y_{1}, Y_{2}, Y_{3}$ and $Y_{4}$. Since $X \sim N\left(336 \mathrm{kN}, 84^{2}(\mathrm{kN})^{2}\right)$, an outcome of $x$ can be obtained using the Box-Muller method of Illustration E8.10, that is
$x=\mu_{X}+\left(-2 \ln u_{1}\right)^{1 / 2} \sin \left(2 \pi u_{2}\right) \sigma_{X}$,
where $u_{1}$ and $u_{2}$ denote two uniform ( 0,1 ) random numbers. Let, for example, $u_{1}=0.4$ and $u_{2}=0.7$. For $\mu_{X}=15^{\circ}$ and $\sigma_{X}=1.5^{\circ}$ one gets
$x=336+84[-2 \ln (0.4)]^{1 / 2} \sin (2 \times 0.7 \pi)=227.85$.
The same procedure is used to generate values $y_{1}$ and $y_{2}$ of $Y_{1} \sim N\left(70 \mathrm{kN}, 10.5^{2}(\mathrm{kN})^{2}\right)$ and $Y_{2} \sim N\left(90 \mathrm{kN}, 4.5^{2}(\mathrm{kN})^{2}\right)$. Since $Y_{3} \sim$ Weibull( $33.24 \mathrm{kN}, 3.71$ ), the probability integral transform is used to get a value $y_{3}$ of $Y_{3}$, that is
$y_{3}=33.24\left(-\ln u_{3}\right)^{1 / \gamma}$,
with $u_{3}$ denoting a uniform (0,1) random number. Similarly, a value $y_{4}$ of $Y_{4} \sim G u m b e l(18.2 \mathrm{kN}, 3.12 \mathrm{kN})$ is obtained from
$y_{4}=18.2-3.12 \ln \left(-\ln u_{4}\right)$,
with $u_{4}$ denoting a uniform $(0,1)$ random number. This procedure is repeated $n$ times to perform $n$ simulation cycles. If one denotes with $x_{i}, y_{1 i}, y_{2 i}, y_{3 i}, y_{4 i}$ the generated values for the $i$-th cycles, a failure occurs if $x_{i}<y_{1 i}+y_{2 i}+y_{3 i}+y_{4 i}$. In this case, let $\eta_{i}=1$, where $\eta_{i}$ is the failure counter. If $x_{i} \geq y_{1 i}+y_{2 i}+y_{3 i}+y_{4 i}$, then $\eta_{i}=0$. The simulated probability of failure is estimated as $p_{f}=\Sigma \eta_{i} / n$. The results of 10 runs with $n=1000$ cycles for each run are shown in Table 9.S14b.

Table 9.S14b - Column load. Simulation results ( $n=1000$ ).

| Run \#: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Average |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $r=$ | 0.914 | 0.923 | 0.919 | 0.921 | 0.928 | 0.938 | 0.917 | 0.933 | 0.928 | 0.927 | $\underline{0.925}$ |
| $p_{f}=$ | 0.086 | 0.077 | 0.081 | 0.079 | 0.072 | 0.062 | 0.083 | 0.067 | 0.072 | 0.073 | $\underline{0.075}$ |

It is seen that the average probability of failure is 0.075 .
9.15. Stormwater removal. In the Italian method for designing storm sewer systems the system capacity is estimated by
$W=q X\left[1-\exp \left(-\frac{q X}{s}\right)\right]^{-1}$,
where $W$ is the stormwater removal capacity (i.e., the volume of stormwater which can be appropriately drained by the system) for a storm with duration $X, q$ is the outlet discharge capacity (i.e., the maximum discharge which can be conveyed by the outlet channel under uniform flow conditions), and $s$ is system storage capacity (i.e., the volume of water which can be stored in the whole system, including the outlet channel, upstream channel network, and surface detention). System reliability can be evaluated using the concepts of capacity and load by considering the volume of stormwater delivered in a storm event as the storm depth multiplied by drainage area $a$, that is, $a X Y$, where $Y$ denotes the average intensity of a storm (see: Rosso, R. and E. Caroni (1977) "Storm sewer capacity design under risk", Proc. XVII Congr. Int. Assoc. Hydraul. Res., Baden-Baden, August 15-19, Vol.4, pp.537-543). Consider the stormwater removal system for a drainage area of $205 \times 10^{3} \mathrm{~m}^{2}$ in a location where the duration $X$ and intensity $Y$ of a severe storm are independent exponentially distributed variates with means of 1.4 hours and $18 \mathrm{~mm} /$ hour, respectively. The outlet discharge capacity of the system is $4 \mathrm{~m}^{3} \mathrm{~s}^{-1}$, and its storage capacity is $1500 \mathrm{~m}^{3}$. Using coherent units,
(a) compute the system reliability for a storm by simulation.
(b) Because of gradually varied flow in the system, the system storage during a storm may not achieve storage capacity, and a random variate $Z$ is substitued for $s$ to evaluate the stormwater removal capacity $W$. Simulate system reliability for a storm if $f_{Z}(z)=3 z^{2} / s^{3}$ for $0 \leq z \leq s$, and 0 elsewhere.
Solution. (a) Since $\mu_{X}=1.4$ hours $=5040 \mathrm{~s}$ and $\mu_{Y}=18 \mathrm{~mm} / \mathrm{hour}=5 \times 10^{-6} \mathrm{~m} / \mathrm{s}$, the exponential assumption for $X$ and $Y$ gives
$F_{X}(x)=1-\exp (-x / 5040)$ and $F_{Y}(y)=1-\exp \left(-2 \times 10^{5} y\right)$.

Each simulation cycle is then performed by generating values $x$ and $y$ of $X \sim \operatorname{exponential}\left(1 / 5040 \mathrm{~s}^{-1}\right)$ and $Y \sim$ exponential $\left(2 \times 10^{5} \mathrm{~s} / \mathrm{m}\right)$. From the probability integral transform,
$x=-5040 \ln u$
and
$y=-5 \times 10^{-6} \ln v$,
where $u$ and $v$ denote two uniform ( 0,1 ) random numbers. Let, for example, $u=0.9$ and $v=0.2$. Thus,
$x=-5040 \ln (0.9)=531 \mathrm{~s}$ and $y=-5 \times 10^{-6} \ln (0.3)=8.05 \times 10^{-6} \mathrm{~m} / \mathrm{s}$,
which are used to evaluate system capacity as
$w=q x\left[1-\exp \left(-\frac{q x}{s}\right)\right]^{-1}=\frac{4 \times 531}{1-\exp \left(-\frac{4 \times 531}{1500}\right)}=2805 \mathrm{~m}^{3}$.
This is compared with the effective volume of stormwater, that is
$a x y=205,000 \times 531 \times 8.05 \times 10^{-6}=876 \mathrm{~m}^{3}$.
Since $w \geq a x y$, the system performs properly for this cycle. In this case, let $\eta_{i}=0$, where $\eta_{i}$ is the failure counter for the $i$-th simulation cycle. On the contrary, $\eta_{i}=0$ if $w_{i} \geq a x_{i} y_{i}$. If $n$ cycles are performed, the simulated probability of failure is estimated as $p_{f}=\Sigma \eta_{i} / n$, and the associated system reliability is $p_{f}=1-\Sigma \eta_{i} / n$. The results of 10 runs with $n=1000$ cycles for each run are shown in Table 9.S15.

Table 9.S15 - Stormwater removal. Simulation results ( $n=1000$ ).

| Run \# | $\Sigma \eta_{i(a)}$ | $p_{f(a)}$ | $r_{(a)}$ | $\Sigma \eta_{i(b)}$ | $p_{f(b)}$ | $r_{(b)}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | 18 | 0.018 | 0.982 | 18 | 0.018 | 0.982 |
| 2 | 15 | 0.015 | 0.985 | 16 | 0.016 | 0.984 |
| 3 | 12 | 0.012 | 0.988 | 13 | 0.013 | 0.987 |
| 4 | 25 | 0.025 | 0.975 | 26 | 0.026 | 0.974 |
| 5 | 17 | 0.017 | 0.983 | 18 | 0.018 | 0.982 |
| 6 | 13 | 0.013 | 0.987 | 13 | 0.013 | 0.987 |
| 7 | 14 | 0.014 | 0.986 | 14 | 0.014 | 0.986 |
| 8 | 15 | 0.015 | 0.985 | 17 | 0.017 | 0.983 |
| 9 | 20 | 0.020 | 0.980 | 22 | 0.022 | 0.978 |
| 10 | 19 | 0.019 | 0.981 | 21 | 0.021 | 0.979 |
| Average | 16.8 | 0.017 | $\mathbf{0 . 9 8 3}$ | 17.8 | 0.018 | $\mathbf{0 . 9 8 2}$ |

It is seen that the average probability of failure is 0.017 and system reliability is 0.983 .
(b) The randomness of storage capacity is considered here. This is a random variable $Z$ with $c d f$
$F_{Z}(z)=z^{3} / s^{3}=z^{3} / 1500^{3}$,
for $z \leq s$, and 1 for $z>s$. A random outcome $z$ of $Z$ is obtained from
$z=s t^{1 / 3}=1500 t^{1 / 3}$,
with $t$ denoting a uniform $(0,1)$ random number. Let, for example, $t=0.1$. Then, $z=15000.1^{1 / 3}=696 \mathrm{~m}^{3}$.
Using the above generated values of $x$ and $y$, system capacity is geiven by
$w=q x\left[1-\exp \left(-\frac{q x}{z}\right)\right]^{-1}=\frac{4 \times 531}{1-\exp \left(-\frac{4 \times 531}{696}\right)}=2230 \mathrm{~m}^{3}$,
which is larger than the effective volume of stormwater of $876 \mathrm{~m}^{3}$. The above counting procedure is repeated $n$ times, and the results of 10 runs with $n=1000$ cycles for each run are also shown in Table 9.S15. It is seen that the average probability of failure is 0.018 and system reliability is 0.982 .
9.16. Dilution requirements. The amount of water into which wastewater can be discharged without creating objectionable conditions is represented by a dilution parameter which is commonly expressed for combined systems as the minimum streamflow $Y$ required. If wastewater with a first-stage BOD of $W$ in N per capita is discharged daily into a stream with a permissible loading of $Z$ in $\mathrm{N} / \mathrm{m}^{3}$, the required streamflow becomes $Y=0.012 \mathrm{~W} / \mathrm{Z}$ in $\mathrm{m}^{3} / \mathrm{s}$ per 1,000 population (see: Fair, G.M., Geyer, J.C. and D.A. Okun (1971): "Elements of water supply and wastewater disposal", 2nd edition, John Wiley and Sons, New York, p.659). Assume that population is 10,000 , permissible loading $Z$ can vary uniformly from 0.23 to $0.12 \mathrm{~N} / \mathrm{m}^{3}$, and load $W$ is a normal variate with a mean of 1.2 N per capita and coefficient of variation of $20 \%$. Find the reliability of the system by simulation if streamflow $X$ is a gamma variate with mean and standard deviation of 2 and $0.5 \mathrm{~m}^{3} / \mathrm{s}$, respectively.

Solution. Each simulation cycle is performed by generating values $w$ and $z$ of $W \sim N\left(1.2 \mathrm{~N}, 0.24^{2} \mathrm{~N}\right)$ and $Z \sim \operatorname{uniform}\left(0.12 \mathrm{~N} / \mathrm{m}^{3}, 0.23 \mathrm{~N} / \mathrm{m}^{3}\right)$. From the probability integral transform,
$z=0.12+(0.23-0.12) u_{1}$,
with $u_{1}$ denoting a uniform $(0,1)$ random number. For example, if $u_{1}=0.3$,
$z=0.12+(0.23-0.12) \times 0.3=0.153 \mathrm{~N} / \mathrm{m}^{3}$.
An outcome of $w$ can be obtained using the Box-Muller method of Illustration E8.10, that is
$w=\mu_{W}+\left(-2 \ln u_{2}\right)^{1 / 2} \sin \left(2 \pi u_{3}\right) \sigma_{W}$,
where $u_{2}$ and $u_{3}$ denote two uniform $(0,1)$ random numbers. Let, for example, $u_{2}=0.4$ and $u_{3}=0.7$. For $\mu_{X}=1.2^{\circ}$ and $\sigma_{X}=0.24^{\circ}$ one gets
$w=1.2+0.24[-2 \ln (0.4)]^{1 / 2} \sin (2 \times 0.7 \pi)=0.891 \mathrm{~N}$ per capita.
Then,
$y=10 \times 0.012 \times 0.981 / 0.153=0.699 \mathrm{~m}^{3} / \mathrm{s}$
is the simulated discharge requirement or system load, which is compared with system capacity, that is streamflow $X$. An outcome $x$ of $X \sim$ gamma $\left(16,0.125 \mathrm{~m}^{3} / \mathrm{s}\right.$ ) is obtained
by generating a standard gamma random number $v$ with shape parameter $r=16$ by using the rejection method of Illustration E8.12. Let, for example, $v=4.2$, then $x=4.2 \times 0.125=0.525 \mathrm{~m}^{3} / \mathrm{s}$.
Since $x<y$, the system fails for this cycle. In this case, let $\eta_{i}=1$, where $\eta_{i}$ is the failure counter for the $i$-th simulation cycle. On the contrary, $\eta_{i}=0$ if $x \geq y$. If $n$ cycles are performed, the simulated probability of failure is estimated as $p_{f}=\Sigma \eta_{i} / n$, and the associated system reliability is $p_{f}=1-\Sigma \eta_{i} / n$. The results of 10 runs with $n=1000$ cycles for each run are shown in Table 9.S16.

Table 9.S16-Diluition requirements.
Simulation results ( $n=1000$ ).

| Run \# | $\Sigma \eta_{i}$ | $r$ | $p_{f}$ |
| :---: | :---: | :---: | :---: |
| 1 | 9 | 0.991 | 0.009 |
| 2 | 13 | 0.987 | 0.013 |
| 3 | 18 | 0.982 | 0.018 |
| 4 | 9 | 0.991 | 0.009 |
| 5 | 17 | 0.983 | 0.017 |
| 6 | 11 | 0.989 | 0.011 |
| 7 | 13 | 0.987 | 0.013 |
| 8 | 9 | 0.991 | 0.009 |
| 9 | 9 | 0.991 | 0.009 |
| 10 | 16 | 0.984 | 0.016 |
| Average | 12.4 | $\mathbf{0 . 9 8 8}$ | 0.012 |

It is seen that the simulation average of system reliability is 0.988 .
9.17. Law of diminishing returns. Consider a system with $n$ serial components, each of them constituted by $m$ redundant independent subcomponents with equal probability of failure $p$. Find (a) the overall reliability of this system. (b) Show that, by determining the rate of increase of system reliability with increasing number $m$, the advantage of introducing additional redundant subcomponents in each serial component rapidly vanishes.
Solution. (a) From Eq. (9.2.7) the probability of failure of the $i$-th serial component is $p_{f i}=p^{m}$. From Eq. (9.2.1)
$r=\prod_{i=1}^{n}\left(1-p_{f i}\right)=\prod_{i=1}^{n}\left(1-p^{m}\right)=\left(1-p^{m}\right)^{n}$
is the required system reliability.
(b) The rate of increase of $r$ with increasing $m$ is given by

$$
\frac{d r}{d m}=\frac{d}{d m}\left(1-p^{m}\right)^{n}=n p^{m}\left(1-p^{m}\right)^{n-1} \ln \frac{1}{p},
$$

which is shown to rapidly decrease in Figure 9.S17.

Figure 9.S17

9.18. Pipe network. Consider the portion of a pipeline network for urban water supply of Illustration E9.18 (see, also, Figure 9.2.2). Assuming independent failure modes find the probability that node $c$ remains isolated if there is a common probability of rupture of $1 \%$ for all pipes.

Solution. From the network it is seen that for this condition to hold at least one of the following routes must work: (2), $(1,5),(1,3,4)$. Thus, the required probability of failure is $p_{f}=\operatorname{Pr}[A B C]$, where the events $A, B$ and $C$ are the failures of route $(2),(1,5)$ and $(1,3,4)$, respectively. For each route the probability of failure is obtained as shown at the extreme right-hand side of Eq. (9.2.8). That is
$\operatorname{Pr}[A]=1-(1-p)=p$ (that is route $A$ does not work);
$\operatorname{Pr}[B]=1-(1-p)(1-p)=2 p-p^{2}$ (that is route $B$ does not work);
$\operatorname{Pr}[C]=1-(1-p)(1-p)(1-p)=3 p-3 p^{2}+p^{3}$ (that is route $C$ does not work).
It is seen that these routes form three parallel-serial (that is redundant) systems. However, the events described here are not independent. For example, routes $B=(1,5)$ and $C=(1,3,4)$ have pipe 1 in common. Therefore, their joint effects are considered. Route $A$ is independent from routes $B$ and $C$. From Eq. (9.2.8)
$p_{f}=\operatorname{Pr}[A B C]=\operatorname{Pr}[A] \operatorname{Pr}[B C]$.
On the other hand, routes $B$ and $C$ are not independent. From the addition rule of probability,

$$
\begin{aligned}
& \operatorname{Pr}[B C]=\operatorname{Pr}[B]+\operatorname{Pr}[C]-\operatorname{Pr}[B+C]=\left(2 p-p^{2}\right)+\left(3 p-3 p^{2}+p^{3}\right)-\left[1-(1-p)^{4}\right]= \\
&=p+2 p^{2}-3 p^{3}+p^{4} .
\end{aligned}
$$

Thus,
$p_{f}=\operatorname{Pr}[A B C]=\operatorname{Pr}[A] \operatorname{Pr}[B C]=p\left(p+2 p^{2}-3 p^{3}+p^{4}\right)=p^{2}+2 p^{3}-3 p^{4}+p^{5}$.
For $p=0.01$,
$p_{f}=0.01^{2}+2 \times 0.01^{3}-3 \times 0.01^{4}+0.01^{5}=0.000102$.
9.19. Retaining wall. The retaining wall for road embankment sketched in Figure 9.P3a can fail due to several factors. The failure modes are schematically indicated in the block diagram for system reliability analysis shown in Figure 9.P3b as given by Harr (1987). Under the assumption of independent modes compute the reliability of the overall system if the individual probabilities of failure are as indicated in parentheses.
Solution. The system can be viewed as a serial system with components given by redundant subsystems. Denote with $p_{i j}$ the probability of failure of the $j$-th redundant component of the $i$-th serial component. System reliability is found by combining Eqs. (9.2.1) and (9.2.6), that is
$r_{s}=\prod_{i=1}^{n}\left[1-\prod_{j=1}^{m_{i}}\left(1-p_{i j}\right)\right]$,
where $m_{i}$ denotes the number of redundant components of the $i$-th serial component, and $n$ is the number of serial components. The individual probabilities are as follows.
Serial component 1 has $m_{1}=1$ subcomponents, with $p_{11}=0.001$ (earthquake).
Serial component 2 has $m_{2}=2$ subcomponents, with $p_{21}=0.0001$ (heavy load) and with $p_{22}=0.001$ (weakened condition of pavement).
Serial component 3 has $m_{3}=3$ subcomponents, each of which is a serial system. The corresponding probabilities of failure are found as
$p_{31}=1-\left(1-p_{311}\right)\left(1-p_{312}\right)=1-(1-0.001)(1-0.005)=0.005995$, which accounts for bearing capacity and retaining wall,
$p_{32}=1-\left(1-p_{321}\right)\left(1-p_{322}\right)=1-(1-0.001)(1-0.001)=0.001999$, which accounts for bearing capacity and soil stability, and
$p_{33}=1-\left(1-p_{331}\right)\left(1-p_{332}\right)=1-(1-0.005)(1-0.001)=0.005995$, which accounts for retaining wall and soil stability.
Serial component 4 has $m_{4}=3$ subcomponents, with $p_{41}=0.02$ (groundwater table), $p_{42}=0.02$ (drainage) and $p_{43}=0.001$ (freezing).
Thus,

$$
\begin{aligned}
& \prod_{j=1}^{1}\left(1-p_{1 j}\right)=(1-0.001)=0.999 \\
& \prod_{j=1}^{2}\left(1-p_{2 j}\right)=(1-0.0001)(1-0.001)=0.9989 \\
& \prod_{j=1}^{3}\left(1-p_{3 j}\right)=(1-0.005995)(1-0.001999)(1-0.005995)=0.98607
\end{aligned}
$$

$$
\prod_{j=1}^{2}\left(1-p_{4 j}\right)=(1-0.02)(1-0.02)(1-0.001)=0.95944
$$

and
$r_{s}=(1-0.999)(1-0.9989)(1-0.98607)(1-0.95944)=0.9441$.
9.20. Raingage network. The raingage network for stormwater management in the metropolitan area of Milan, Italy, is constituted by 16 stations. It is seen that real time operation of the urban drainage control system requires telemetered data from at least 12 stations in order to have sufficient information of spatial precipitation. Find the reliability of the network if the failures occur independently with a probability of 10\%.

Solution. From Eq. (9.2.11)
$r=1-p_{f}=\sum_{x=k}^{m}\binom{m}{x}(1-p)^{x} p^{m-x}$,
with $m=16, k=12$ and $p=0.1$. Thus,

$$
\begin{aligned}
& r=\sum_{x=12}^{16}\binom{16}{x}(1-p)^{x} p^{m-x}= \\
& =1820 \times 0.9^{12} \times 0.1^{4}+560 \times 0.1^{13} \times 0.1^{3}+120 \times 0.9^{14} \times 0.1^{2}+16 \times 0.9^{15} \times 0.1+1 \times 0.9^{16} \times 1= \\
& \quad=0.0514+0.1423+0.2745+0.3294+0.1853=0.9830 .
\end{aligned}
$$

9.21. Improved reliability bounds for a $\boldsymbol{k}$-out-of-m system. Assuming that failures occur as rare events, one can substitute the Poisson distribution for the binomial distribution to compute the reliability of a $k$-out-of-m system. Accordingly, Serfling (1974) found
$F_{X}(m-k)-l \leq r \leq F_{X}(m-k)+l$,
where $F_{X}($.$) is the c d f$ of a Poisson variate $X$ with mean $\Sigma p_{i}$, and $2 l=\Sigma p_{i}^{2}$, where the summation is made over all $m$ components, each of them having an individual probability of failure of $p_{i}$. Compute these bounds for the pumping system shown in Figure 9.2.3 with $p=0.03$.
Solution. This is a 3-out-of-4 system with equal probability of failure for each component, that is $m=4, k=3$ and $p_{i}=p=0.03$. The mean of $X$ is given by $m p=4 \times 0.03=0.12$. Thus,
$F_{X}(m-k)=F_{X}(1)=\sum_{x=0}^{1} \frac{(m p)^{x} e^{-m p}}{x!}=\frac{0.12^{0} e^{-0.12}}{0!}+\frac{0.12 e^{-0.12}}{1!}=0.9934$.
Since $l=4 \times 0.03^{2} / 2=0.0018$, the lower and upper bound are found as
$r_{-}=F_{X}(1)-l=0.9934-0.0018=0.9916$
and
$r_{+}=F_{X}(1)+l=0.9934+0.0018=0.9952$,
respectively.
9.22. Levee collapse. The design elevation of a levee built for flood protection at a site in the Po river plain was determined using the estimated 200-year flood, so that overtopping occurs with a risk of $0.5 \%$ in a year. However, this structure can fail also due to excessive seepage through the foundation material at high stages of the river, and it is estimated that this can occur with a probability of $10 \%$ if the 10 -year flood stage is exceeded. Assuming that the two failure modes are normal and correlated with $\rho=0.7$, find the minimum reliability of the levee and compare this probability with that corresponding to the independent case.

Solution. Denote with $A, B$ and $C$ the occurrence of the 200-year flood, of the 10 -year flood and of excessive seepage, respectively. Since $\operatorname{Pr}[C \mid B]=0.10$, the corresponding probabilities are
$\operatorname{Pr}[A]=1 / 200=0.005$,
$\operatorname{Pr}[B]=1 / 10=0.1$
and
$\operatorname{Pr}[C]=\operatorname{Pr}[C \mid B] \operatorname{Pr}[B]=0.1 \times 0.1=0.01$.
From Eq. (9.2.17) two weak bounds for system reliability are found as
$\min (1-\operatorname{Pr}[A], 1-\operatorname{Pr}[C]) \leq r \leq 1-\operatorname{Pr}[A] \operatorname{Pr}[C]$,
that is
$\min (0.995,0.99) \leq r \leq 1-0.005 \times 0.01$,
$0.99 \leq r \leq 0.99995$.
Under the assumption of normal failure modes, the reliability bounds are found by using Eq. (9.2.25). The individual reliability indeces of the two serial failure modes $A$ and $C$ are
$\beta_{A}=\Phi^{-1}(1-\operatorname{Pr}[A])=\Phi(1-0.005)=\Phi(0.995)=2.576$,
$\beta_{C}=\Phi^{-1}(1-\operatorname{Pr}[C])=\Phi(1-0.01)=\Phi(0.99)=2.326$.
Also,

$$
p_{a}=\Phi\left(-\beta_{A}\right) \Phi\left(-\frac{\beta_{C}-\rho \beta_{A}}{\sqrt{1-\rho^{2}}}\right)=\Phi(-2.576) \Phi\left(-\frac{2.326-0.7 \times 2.576}{\sqrt{1-0.7^{2}}}\right)=0.0029
$$

and

$$
p_{c}=\Phi\left(-\beta_{C}\right) \Phi\left(-\frac{\beta_{A}-\rho \beta_{C}}{\sqrt{1-\rho^{2}}}\right)=\Phi(-2.326) \Phi\left(-\frac{2.576-0.7 \times 2.326}{\sqrt{1-0.7^{2}}}\right)=0.0038
$$

From Eq. (9.2.22)
$1-\operatorname{Pr}[A]-\operatorname{Pr}[C]+\max \left(p_{a}, p_{c}\right) \leq r \leq 1-\operatorname{Pr}[A]-\operatorname{Pr}[C]+p_{a}+p_{c}$,
that is
$1-0.005-0.01+\max (0.0029,0.0038) \leq r \leq 1-0.005-0.01+0.0029+0.0038$, $0.9888 \leq r \leq 0.9917$.

The minimum reliability of the levee is 0.9888 . From Eq. (9.2.1)
$r=(1-\operatorname{Pr}[A])(1-\operatorname{Pr}[C])=(1-0.005)(1-0.01)=0.9851$
for independent failure modes.
9.23. Redundant and serial equally reliable components. A system with a given overall reliability $r$ is composed of $l$ positively correlated components; each of them has an identical probability of failure, say, p. Find the bounds for the reliability of each component if (a) they are serial; (b) they are redundant.

Solution. By substituting the reliability $r_{i s}=1-p$ of the $i$-th serial positively correlated component for $p$ in Eq. (9.2.16)
$\left[1-\left(1-r_{i s}\right)\right]^{l} \leq r \leq 1-\left(1-r_{i s}\right)$,
one gets
$r \leq r_{i s} \leq r^{1 / l}$.
From Eq. (9.2.18)
$1-\left(1-r_{i s}\right) \leq r \leq 1-\left(1-r_{i r}\right)^{l}$,
that is
$1-(1-r)^{1 / l} \leq r_{i s} \leq r$.
where $r_{i r}=1-p$ is the reliability of the $i$-th redundant positively correlated component.
9.24. Repeated design. The reliability of a particular design procedure to prevent the collapse of buildings caused by earthquakes was found to be $99 \%$ over a long period of time. Since it is planned to construct 10 structures using this design, evaluate the probability that none of the 10 similar structures fails over the same time span.

Solution. The risk of failure of a new structure is $p=1-r=1-0.99=0.01$. The probability that none of 10 structures will collapse is given by the binomial $p m f$ for $X=0$ with individual failure probability $p$ and $n=10$ number of trials, that is

$$
f_{X}(0)=\binom{10}{0} p^{0}(1-p)^{10-0}=(1-0.01)^{10}=0.9044
$$

9.25. Reservoir sedimentation. When reservoir sedimentation exceeds the dead level in a reservoir, it can affect its efficiency in meeting the target demand. Expensive maintainance work is than necessary to remove sedimentation excess. Assuming that the annual sediment yield trapped by a reservoir with dead capacity $c$ is an exponentially distributed variate with mean $\mu$, find expressions of (a) $R(t)$ and (b) $h(t)$. (c) For $c=5 \times 10^{6} \mathrm{~m}^{3}$ and $\mu=4 \times 10^{5} \mathrm{~m}^{3}$, determine the design life for a reliability level of $90 \%$. Note that the sum of $t$ exponentially distributed variates having a common scale parameter is a gamma distributed variate with the same scale parameter and shape parameter equal to $t$.
Solution. (a) The cumulative sedimentation $X$ after $t$ years is a gamma distributed variate with $p d f$

$$
f_{X}(x)=\frac{1}{\mu \Gamma(t)}\left(\frac{x}{\mu}\right)^{t-1} \exp \left(-\frac{x}{\mu}\right)
$$

The reliability function for the reservoir is given by the probability that $X$ is less than $c$, that is

$$
R(t)=\int_{0}^{c} \frac{1}{\mu \Gamma(t)}\left(\frac{x}{\mu}\right)^{t-1} \exp \left(-\frac{x}{\mu}\right) d x .
$$

This is the $c d f$ of $X$ for $X=c$.
(b) The probability that $X$ reaches $c$ after $t$ years equals the probability that the survival time $W$ is $t$ years, that is

$$
f_{W}(t)=\frac{1}{\mu \Gamma(t)}\left(\frac{c}{\mu}\right)^{t-1} \exp \left(-\frac{c}{\mu}\right) .
$$

From Eq. (9.4.17)
$h(t)=\frac{f_{W}(t)}{R(t)}=\frac{\frac{1}{\mu \Gamma(t)}\left(\frac{c}{\mu}\right)^{t-1} \exp \left(-\frac{c}{\mu}\right)}{\int_{0}^{c} \frac{1}{\mu \Gamma(t)}\left(\frac{x}{\mu}\right)^{t-1} \exp \left(-\frac{x}{\mu}\right) d x}$.
The reliability and hazard functions are shown in Figure 9.525 for $c=5 \times 10^{6} \mathrm{~m}^{3}$ and $\mu=4 \times 10^{5} \mathrm{~m}^{3}$.

Figure 9.S25

(c) The reliable life for $r=0.9$ is obtained by iterative computations of $R(t)$. Since
$R(8)=\int_{0}^{c} \frac{1}{\mu \Gamma(8)}\left(\frac{x}{\mu}\right)^{8-1} \exp \left(-\frac{x}{\mu}\right) d x=0.930$
and
$R(9)=\int_{0}^{c} \frac{1}{\mu \Gamma(9)}\left(\frac{x}{\mu}\right)^{8-1} \exp \left(-\frac{x}{\mu}\right) d x=0.875$,
the reliable life of the reservoir is from 8 to 9 years.
9.26. Road pavement. Suppose that a road is made of 1000 pavements sections. The number of surviving pavements $n_{s}$ after the $j$-th year in service is as follows:

| $j=$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $n_{s}=$ | 865 | 782 | 701 | 362 | 201 | 157 | 86 | 47 | 40 | 36 |
| $j=$ | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| $n_{s}=$ | 31 | 27 | 26 | 16 | 10 | 6 | 4 | 3 | 2 | 1 |

Estimate the $p d f$ and $c d f$ of the survival time distribution, $f_{W}(t)$ and $F_{W}(t)$, the reliability function $R(t)$, and the hazard function $h(t)$ for a pavement section. Find the reliable life of a pavement section for a specified reliability level of $70 \%$. Note that for the discrete case the hazard function is the ratio between the number of failures in a time interval and the average number of survivors for the period.
Solution. The $c d f$ of survival time $W$ is computed as
$F_{W}(t)=\left[n_{s}(0)-n_{s}(t)\right] / n_{s}(0)$,
where $t=j$ and $n_{s}(t)$ denotes the number of surviving pavements. The pmf of $W$ is given by
$p_{W}(t)=F_{W}(t)-F_{W}(t-1)=\left[n_{s}(t-1)-n_{s}(t)\right] / n_{s}(0)$.
From Eq. (9.4.1)
$R(t)=1-F_{W}(t)=1-\left[n_{s}(0)-n_{s}(t)\right] / n_{s}(0)=n_{s}(t) / n_{s}(0)$.
From Eq. (9.4.17)
$h(t)=p_{W}(t) / R(t)=\left[n_{s}(t-1)-n_{s}(t)\right] / n_{s}(t)$.
The computed values of $F_{W}(t), p_{W}(t), R(t)$ and $h(t)$ are reported in Table 9.526 for $t=1, \ldots 20$. Functions $F_{W}(t), R(t)$ and $h(t)$ are also shown in Figure 9S.26.

Table 9.S26-Road pavement.

| $j=t$ | $n_{s, t}$ | $F_{w}(t)$ | $f_{w}(t)$ | $R(t)$ | $h(t)$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $\mathbf{0}$ | $\mathbf{1 0 0 0}$ | 0.000 | 0.000 | 1.000 | 0.000 |
| $\mathbf{1}$ | $\mathbf{8 6 5}$ | 0.135 | 0.135 | 0.865 | 0.156 |
| $\mathbf{2}$ | $\mathbf{7 8 2}$ | 0.218 | 0.083 | 0.782 | 0.106 |
| $\mathbf{3}$ | $\mathbf{7 0 1}$ | 0.299 | 0.081 | 0.701 | 0.116 |
| $\mathbf{4}$ | $\mathbf{3 6 2}$ | 0.638 | 0.339 | 0.362 | 0.936 |
| $\mathbf{5}$ | $\mathbf{2 0 1}$ | 0.799 | 0.161 | 0.201 | 0.801 |
| $\mathbf{6}$ | $\mathbf{1 5 7}$ | 0.843 | 0.044 | 0.157 | 0.280 |
| $\mathbf{7}$ | $\mathbf{8 6}$ | 0.914 | 0.071 | 0.086 | 0.826 |
| $\mathbf{8}$ | $\mathbf{4 7}$ | 0.953 | 0.039 | 0.047 | 0.830 |
| $\mathbf{9}$ | $\mathbf{4 0}$ | 0.960 | 0.007 | 0.040 | 0.175 |
| $\mathbf{1 0}$ | $\mathbf{3 6}$ | 0.964 | 0.004 | 0.036 | 0.111 |
| $\mathbf{1 1}$ | $\mathbf{3 1}$ | 0.969 | 0.005 | 0.031 | 0.161 |
| $\mathbf{1 2}$ | $\mathbf{2 7}$ | 0.973 | 0.004 | 0.027 | 0.148 |
| $\mathbf{1 3}$ | $\mathbf{2 6}$ | 0.974 | 0.001 | 0.026 | 0.038 |
| $\mathbf{1 4}$ | $\mathbf{1 6}$ | 0.984 | 0.010 | 0.016 | 0.625 |
| $\mathbf{1 5}$ | $\mathbf{1 0}$ | 0.990 | 0.006 | 0.010 | 0.600 |
| $\mathbf{1 6}$ | $\mathbf{6}$ | 0.994 | 0.004 | 0.006 | 0.667 |
| $\mathbf{1 7}$ | $\mathbf{4}$ | 0.996 | 0.002 | 0.004 | 0.500 |
| $\mathbf{1 8}$ | $\mathbf{2}$ | 0.997 | 0.001 | 0.003 | 0.333 |
| $\mathbf{1 9}$ | $\mathbf{1}$ | 0.998 | 0.001 | 0.002 | 0.500 |
| $\mathbf{2 0}$ | $\mathbf{0}$ | 0.999 | 0.001 | 0.001 | 1.000 |

Figure 9.S26


From Table 9.S26 $R(3)=0.701$ and $R(4)=0.362$. Thus, $t_{0.7} \approx 3$ years.
9.27. Weibull reliability function. The reliability function of a salt-water conversion unit is taken as $R(t)=\exp \left[-(t / \tau)^{\gamma}\right]$, where $\tau$ is its characteristic lifetime, $t$ is the test time, and $\gamma$ is a parameter estimated from observations of several units of this type. For $\gamma=1$, one has the exponential distribution with constant rate of failure $1 / \tau$, and $\gamma=2$ gives the Rayleigh model associated with a linearly increasing
hazard function. (a) Find the expression of $h(t)$ associated with $R(t)$. (b) If $\tau=10$ years, $\gamma=1.5$, and the unit is observed to be performing properly for 5 years, find its conditional reliability at the end of this period for $t=10$ years.
Solution. (a) From Eq. (9.4.1) the $c d f$ of survival time $W$ is given by
$F_{W}(t)=1-R(t)=1-\exp \left[-(t / \tau)^{2}\right]$.
Thus,
$f_{W}(t)=d F_{W}(t) / d t=(\gamma / \tau)(t / \tau)^{\gamma-1} \exp \left[-(t / \tau)^{\gamma}\right]$.
From Eq. (9.4.17)
$h(t)=f_{w}(t) / R(t)=(\gamma / \tau)(t / \tau)^{\gamma-1} \exp \left[-(t / \tau)^{\gamma}\right] / \exp \left[-(t / \tau)^{\gamma}\right]=(\gamma / \tau)(t / \tau)^{\gamma-1}$,
which is shown in Figure 9.S27.
Figure 9.S27

(b) The required conditional reliability function is given by
$R(t \mid 5)=1-\operatorname{Pr}[5<W \leq t \mid W>5]=1-\operatorname{Pr}[5<W \leq t] / \operatorname{Pr}[W>5]=1-\left[F_{W}(t)-F_{W}(5)\right] / R(5)$.
Hence
$R(t \mid 5)=1-\left\{\exp \left[-(5 / 10)^{1.5}\right]-\exp \left[-(t / 10)^{1.5}\right\} / \exp \left[-(5 / 10)^{1.5}\right]\right.$.
For $t=10$,
$R(10 \mid 5)=1-\left\{\exp \left[-(5 / 10)^{1.5}\right]-\exp \left[-(10 / 10)^{1.5}\right\} / \exp \left[-(5 / 10)^{1.5}\right]=0.5239\right.$.
9.28. Combined hazards in bridge construction. An engineer evaluates the reliability of a bridge to be constructed in an earthquake prone area, where the estimated return period of catastrophic earthquakes is 250. The 200-year flood must be taken as a design guidance for that area. The bridge is exposed to the hazard due to obsolescence, which increases in time as a logistic function, say,
$h_{\text {obs }}(t)=0.05 /\{1+\exp [-0.25(t-25)]\}$
where $t$ is in years. Assuming that earthquakes and floods are independent sequences of Poisson events find (a) the reliability function of the bridge, and (b) its design life for a reliability level of $90 \%$.

Solution. (a) Under the Poisson assumption, the hazard functions for earthquakes and floods are given by
$h_{e}(t)=(1 / 250) \exp (-t / 250) / \exp (-t / 250)=1 / 250=0.004$ year $^{-1}$
and
$h_{f}(t)=(1 / 200) \exp (-t / 200) / \exp (-t / 200)=1 / 250=0.005$ year $^{-1}$,
respectively. Since the three processes are mutually independent, the hazard rates can be added. Thus,
$h(t)=h_{\text {obs }}(t)+h_{e}(t)+h_{f}(t)=0.05 /\{1+\exp [-0.25(t-25)]\}+0.009$.
From Eq. (9.4.18)

$$
R(t)=\exp \left[-\int_{0}^{t} h(u) d u\right]=\exp \left\{-0.009 t-\int_{0}^{t} \frac{0.05}{1+\exp [-0.25(u-25)]} d u\right\},
$$

which is shown in Figure 9.S28.

Figure 9.S28

(b) The design life for a reliability level of $90 \%$ is found by solving numerically
$0.9=\exp \left\{-0.009 t_{0.9}-\int_{0}^{t_{0.9}} \frac{0.05}{1+\exp [-0.25(u-25)]} d u\right\}$,
which yields $t_{0.9} \approx 11$ years.
9.29. Bearing capacity of soil. Bearing capacity of soil depends on the following three factors:
$Y=\tan ^{4}(45+X / 2)$,
$W=e^{\tan (X)} \tan ^{2}(45+X / 2)$,
$Z=\operatorname{cotan}(X)\left[e^{\pi \tan (X)} \tan ^{2}(45+X / 2)-1\right]$,
which are functions of the friction angle $X$ (see: Vannucchi, 1986). Using the point estimate method find the mean and coefficient of variation of $Y, W$ and $Z$ for mean friction angles of $15^{\circ}, 25^{\circ}$ and $35^{\circ}$, and coefficients of variations of $X$ of $0.1,0.2$ and 0.3 . Compare these estimates with those obtained by simulation.

Solution. For $\mu_{X}=15^{\circ}$ and $V_{X}=0.1, \sigma_{X}=1.5^{\circ}$. Application of Eq. (9.1.35) gives
$y_{1}=\tan ^{4}\left(45+\mu_{X} / 2+\sigma_{X} / 2\right)=\tan ^{4}(45+15 / 2+1.5 / 2)=3.216$
and
$y_{2}=\tan ^{4}\left(45+\mu_{X} / 2-\sigma_{X} / 2\right)=\tan ^{4}(45+15 / 2-1.5 / 2)=2.589$.
These point estimates are used in Eqs. (9.1.40) and (9.1.41) to obtain
$\mu_{Y}=0.5\left[\tan ^{4}\left(45+\mu_{X} / 2+\sigma_{X} / 2\right)+\tan ^{4}\left(45+\mu_{X} / 2-\sigma_{X} / 2\right)\right]=0.5(3.216+2.589)=2.903$,
and
$\sigma_{X}{ }^{2}=0.25\left[\tan ^{8}\left(45+\mu_{X} / 2+\sigma_{X} / 2\right)+\tan ^{8}\left(45+\mu_{X} / 2-\sigma_{X} / 2\right)-\right.$

$$
\begin{aligned}
& \left.-2 \tan ^{4}\left(45+\mu_{X} / 2+\sigma_{X} / 2\right) \tan ^{4}\left(45+\mu_{X} / 2-\sigma_{X} / 2\right)\right]= \\
& =0.25\left(3.216^{2}+2.589^{2}-2 \times 3.216 \times 2.589\right)=0.098
\end{aligned}
$$

hence,
$V_{Y}=\sigma_{Y} / \mu_{Y}=0.098^{1 / 2} / 2.903=0.108$.
Similarly,

$$
\begin{aligned}
w_{1}=\exp \left[\pi \operatorname { t a n } \left(\mu_{X}\right.\right. & \left.\left.+\sigma_{X}\right)\right] \tan ^{2}\left(45+\mu_{X} / 2+\sigma_{X} / 2\right)= \\
& =\exp [\pi \tan (15+1.5)] \tan ^{2}(45+15 / 2+1.5 / 2)=2.536 \times 1.793=4.548
\end{aligned}
$$

and

$$
\begin{aligned}
& w_{2}=\exp \left[\pi \tan \left(\mu_{X}-\sigma_{X}\right)\right] \tan ^{2}\left(45+\mu_{X} / 2-\sigma_{X} / 2\right)= \\
& \quad=\exp [\pi \tan (15+1.5)] \tan ^{2}(45+15 / 2-1.5 / 2)=2.126 \times 1.609=3.421
\end{aligned}
$$

Then,
$\mu_{W}=0.5(4.548+3.421)=3.984$,
and
$\sigma_{W}{ }^{2}=0.25\left(4.548^{2}+3.421^{2}-2 \times 4.548 \times 3.421\right)=0.318 ;$
hence
$V_{Y}=\sigma_{Y} / \mu_{Y}=0.318^{1 / 2} / 3.984=0.141$.
Also,
$z_{1}=\operatorname{cotan}\left(\mu_{X}+\sigma_{X}\right)\left\{\exp \left[\pi \tan \left(\mu_{X}+\sigma_{X}\right)\right] \tan ^{2}\left(45+\mu_{X} / 2+\sigma_{X} / 2\right)-1\right\}=$

$$
\begin{aligned}
=\operatorname{cotan}(15+1.5)\left\{\exp [\pi \tan (15+1.5)] \tan ^{2}(45\right. & +15 / 2+1.5 / 2) 2)-1\}= \\
& =3.376 \times 3.548=11.978
\end{aligned}
$$

and
$z_{2}=\operatorname{cotan}\left(\mu_{X}-\sigma_{X}\right)\left\{\exp \left[\pi \tan \left(\mu_{X}-\sigma_{X}\right)\right] \tan ^{2}\left(45+\mu_{X} / 2-\sigma_{X} / 2\right)-1\right\}=$

$$
\begin{array}{r}
\left.=\operatorname{cotan}(15-1.5)\left\{\exp [\pi \tan (15-1.5)] \tan ^{2}(45+15 / 2-1.5 / 2) 2\right)-1\right\}= \\
= \\
4.165 \times 2.421=10.084 .
\end{array}
$$

Then,
$\mu_{Z}=0.5(11.978+10.084)=11.031$,
and
$\sigma_{Z}^{2}=0.25\left(11.978^{2}+10.084^{2}-2 \times 11.978 \times 10.084\right)=0.897 ;$
hence
$V_{Y}=\sigma_{Y} / \mu_{Y}=0.897^{1 / 2} / 11.031=0.086$.
The same procedure is carried out for $V_{X}$ equal to 0.2 and 0.3 . The complete results are shown in Table 9.S29 for $\mu_{X}=15^{\circ}, 25^{\circ}$ and $35^{\circ}$.
Simulations are performed by generating values of $Y, W$ and $Z$ from that of friction angle $X \sim N\left(\mu_{X}, \sigma_{X}^{2}\right)$. An outcome of $x$ can be obtained using the Box-Muller method of Illustration E8.10, that is
$x=\mu_{X}+\left(-2 \ln u_{1}\right)^{1 / 2} \sin \left(2 \pi u_{2}\right) \sigma_{X}$,
where $u_{1}$ and $u_{2}$ denote two uniform ( 0,1 ) random numbers. Let, for example, $u_{1}=0.4$ and $u_{2}=0.7$. For $\mu_{X}=15^{\circ}$ and $\sigma_{X}=1.5^{\circ}$ one gets
$x=15+1.5[-2 \ln (0.4)]^{1 / 2} \sin (2 \times 0.7 \pi)=13.069 ;$
hence
$y=\tan ^{4}(45+13.069 / 2)=\tan ^{4}(51.534)=2.510$,
$w=\exp [\pi \tan (13.069)] \tan ^{2}(45+13.069 / 2)=3.285$
and
$z=\operatorname{cotan}(13.069)\left\{\exp [\pi \tan (13.069)] \tan ^{2}(45+13.069 / 2)-1\right\}=9.845$.
This procedure is repeated and the means and coefficients of variation of $Y, W$ and $Z$ are estimated from the generated samples. The results are shown in Table 9.S29.

Table 9.S29-Bearing capacity of soil.

|  |  |  | (a) Point Estimate Method |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\mu_{X}$ | $V_{X}$ | $\mu_{Y}$ | $V_{Y}$ | $\mu_{W}$ | $V_{W}$ | $\mu_{Z}$ | $V_{Z}$ |  |  |  |
| $\mathbf{1 5}$ | $\mathbf{0 . 1}$ | 2.903 | 0.108 | 3.984 | 0.141 | 11.031 | 0.086 |  |  |  |
| $\mathbf{1 5}$ | $\mathbf{0 . 2}$ | 2.957 | 0.214 | 4.116 | 0.277 | 11.194 | 0.171 |  |  |  |
| $\mathbf{1 5}$ | $\mathbf{0 . 3}$ | 3.049 | 0.315 | 4.339 | 0.404 | 11.471 | 0.253 |  |  |  |
| $\mathbf{2 5}$ | $\mathbf{0 . 1}$ | 6.196 | 0.190 | 11.083 | 0.257 | 21.151 | 0.175 |  |  |  |
| $\mathbf{2 5}$ | $\mathbf{0 . 2}$ | 6.580 | 0.368 | 12.400 | 0.484 | 22.487 | 0.340 |  |  |  |
| $\mathbf{2 5}$ | $\mathbf{0 . 3}$ | 7.249 | 0.523 | 14.797 | 0.662 | 24.867 | 0.489 |  |  |  |
| $\mathbf{3 5}$ | $\mathbf{0 . 1}$ | 14.321 | 0.290 | 37.084 | 0.411 | 49.272 | 0.309 |  |  |  |
| $\mathbf{3 5}$ | $\mathbf{0 . 2}$ | 16.560 | 0.537 | 50.047 | 0.706 | 59.755 | 0.568 |  |  |  |
| $\mathbf{3 5}$ | $\mathbf{0 . 3}$ | 20.770 | 0.719 | 78.120 | 0.870 | 81.309 | 0.754 |  |  |  |
|  |  |  | (b) Simulation |  |  |  |  |  |  |  |
| $\mu_{X}$ | $V_{X}$ | $\mu_{Y}$ | $V_{Y}$ | $\mu_{W}$ | $V_{W}$ | $\mu_{Z}$ | $V_{Z}$ |  |  |  |
| $\mathbf{1 5}$ | $\mathbf{0 . 1}$ | 2.902 | 0.108 | 3.983 | 0.143 | 11.029 | 0.086 |  |  |  |
| $\mathbf{1 5}$ | $\mathbf{0 . 2}$ | 2.924 | 0.163 | 4.037 | 0.419 | 11.096 | 0.138 |  |  |  |
| $\mathbf{1 5}$ | $\mathbf{0 . 3}$ | 2.956 | 0.213 | 4.115 | 0.590 | 11.192 | 0.175 |  |  |  |
| $\mathbf{2 5}$ | $\mathbf{0 . 1}$ | 6.191 | 0.194 | 11.081 | 0.268 | 21.147 | 0.180 |  |  |  |
| $\mathbf{2 5}$ | $\mathbf{0 . 2}$ | 6.354 | 0.295 | 11.645 | 0.419 | 21.712 | 0.278 |  |  |  |
| $\mathbf{2 5}$ | $\mathbf{0 . 3}$ | 6.590 | 0.406 | 12.515 | 0.590 | 22.563 | 0.388 |  |  |  |
| $\mathbf{3 5}$ | $\mathbf{0 . 1}$ | 14.329 | 0.307 | 37.242 | 0.474 | 49.357 | 0.342 |  |  |  |
| $\mathbf{3 5}$ | $\mathbf{0 . 2}$ | 15.300 | 0.481 | 43.439 | 0.820 | 54.197 | 0.575 |  |  |  |
| $\mathbf{3 5}$ | $\mathbf{0 . 3}$ | 16.873 | 0.709 | 55.763 | 1.380 | 63.123 | 0.940 |  |  |  |

9.30. Partial Load Factors for Beams. Consider two simply supported beams made of different materials which are designed to carry the same live load, with a nominal value of $q_{l}=2 \mathrm{kN} / \mathrm{m}$. The nominal dead loads are $q_{d}=1 \mathrm{kN} / \mathrm{m}$ for beam 1 , and $q_{d}=3 \mathrm{kN} / \mathrm{m}$ for beam 2, respectively. The mean value of the nominal resistance is $\left(\gamma_{d} q_{d}+\gamma_{1} q_{l}\right) / \phi$ with $\phi=0.9, \gamma_{d}=1.33$ and $\gamma_{1}=1.5$, and its coefficient of variation is 0.1 . Use the FOSM method to compute the values of the reliability index $\beta$ for the two beams assuming that both dead and live loads are normally distributed variates with means equal to the nominal values and coefficients of variation of 0.1 for the dead loads, and 0.25 for the live load, respectively. Modify the value of the partial factor $\gamma_{d}$ for beam 1 in order to obtain the same reliability of beam 2.
Solution. The mean, standard deviation and coefficient of variation of load $Y$ are
$\mu_{Y(\text { Beam } 1)}=q_{d(\text { Beam 1) }}+q_{l}=1+2=3 \mathrm{kN} / \mathrm{m}$,
$\sigma_{Y(\text { Beam 1) }}=\left(\sigma_{d(\text { Beam 1) }}^{2}+\sigma_{l}^{2}\right)^{1 / 2}=\left[(0.1 \times 1)^{2}+(0.25 \times 2)^{2}\right]^{1 / 2}=0.510 \mathrm{kN} / \mathrm{m}$
and
$V_{Y(\text { Beam 1) }}=\sigma_{Y(\text { Beam 1) }} / \mu_{Y(\text { Beam 1) }}=0.510 / 3=0.170$,
for beam 1; and
$\mu_{Y(\text { Beam 2) }}=q_{d(\text { Beam 2) }}+q_{l}=3+2=5 \mathrm{kN} / \mathrm{m}$,
$\sigma_{Y(\text { Beam 2) }}=\left(\sigma_{d(\text { Beam 2) }}{ }^{2}+\sigma_{l}^{2}\right)^{1 / 2}=\left[(0.1 \times 3)^{2}+(0.25 \times 2)^{2}\right]^{1 / 2}=0.583 \mathrm{kN} / \mathrm{m}$
and
$V_{Y(\text { Beam 2) }}=\sigma_{Y(\text { Beam 2) }} / \mu_{Y(\text { Beam 2) }}=0.583 / 3=0.117$,
for beam 2. The mean of resistance $X$ is
$\mu_{X(\text { Beam 1) }}=\left(\gamma_{d} q_{d(\text { Beam 1) }}+\gamma q_{I I}\right) / \phi=(1.33 \times 1+1.5 \times 2) / 0.9=4.811 \mathrm{kN} / \mathrm{m}$, for beam 1; and
$\mu_{X(\text { Beam 2) }}=\left(\gamma_{d} q_{d(\text { Beam 2) }}+\gamma_{1} q_{I I}\right) / \phi=(1.33 \times 3+1.5 \times 2) / 0.9=7.767 \mathrm{kN} / \mathrm{m}$, for beam 2, with $V_{X(\text { Beam 1) }}=V_{X(\text { Beam 2) }}=V_{X}=0.1$.

From Eq. (9.1.34)

$$
\beta_{\text {Beam1 }} \approx \frac{\ln \left(\mu_{X(\text { Beam 1) }} / \mu_{Y(\text { Beam 1) }}\right)}{\sqrt{V_{X}^{2}+V_{Y(\text { Beam1 })}^{2}}}=\frac{\ln (4.811 / 3)}{\sqrt{0.1^{2}+0.170^{2}}}=2.395
$$

and

$$
\beta_{\text {Beam } 2} \approx \frac{\ln \left(\mu_{X(\text { Beam 2) }} / \mu_{Y(\text { Beam } 2)}\right)}{\sqrt{V_{X}^{2}+V_{Y(\text { Beam2) }}^{2}}}=\frac{\ln (7.767 / 5)}{\sqrt{0.1^{2}+0.117^{2}}}=2.867 .
$$

Eq.(9.1.34) is inverted to obtain the modified partial load factor $\gamma_{d m}$ for beam 1 for a specified value of the reliability index, that is

$$
\mu_{X(\text { Beam 1) }}=\mu_{Y(\text { Beam 1 })} \exp \left(\beta \sqrt{V_{X}^{2}+V_{Y(\text { Beam1 })}^{2}}\right) .
$$

Hence,

$$
\begin{aligned}
\gamma_{d(\text { Beam 1, new })}=\left[\phi \mu_{Y(\text { Beam 1) }}\right. & \exp \left(\beta \sqrt{\left.V_{X}^{2}+V_{Y(\text { Beam1) }}^{2}\right)}-\gamma_{l} q_{l}\right] q_{d}^{-1}= \\
& =\left[0.9 \times 3 \times \exp \left(2.867 \sqrt{0.1^{2}+0.170^{2}}\right)-1.5 \times 2\right] \times 1^{-1}=1.75 .
\end{aligned}
$$

## Blank page

# Applied Statistics for Civil and Environmental Engineers <br> Problem Solution Manual <br> by N.T. Kottegoda and R. Rosso <br> <br> Chapter 10 - Bayesian Decision Methods and Parameter <br> <br> Chapter 10 - Bayesian Decision Methods and Parameter Uncertainty 

 Uncertainty}
10.1. Treatment plant design. Design I of a wastewater plant for a new community is based on the expectation that some heavy industries will be established in the area. This eventuality has an estimated probability of .9. Some alternative designs are considered. The cost of implementing Design $I(a)$ is $\$ 300,000$ and Design $I(b)$ is $\$ 400,000$. These designs are based on effluents of two types (depending on types of possible heavy industries) that have equal probabilities, given a positive decision on citing the industries here. If these industries are not established in the area, a loss $\$ 150,000$ will be incurred because of modifications to the plant in Design I(a). Furthermore, if Design I(a) is implemented, whereas subsequently Design I(b) is found to be necessary, an additional cost of $\$ 150,000$ will be incurred. [Design $\mathrm{I}(\mathrm{b})$ is versatile in all these aspects.]

Design II, on the other hand, costs $\$ 170,000$ to implement and does not take account of the extra industrial effluent; however, if the industries are cited in the area, it is deemed that extensions costing an estimated $\$ 330,000$ will be necessary to meet the increased demand.

Sketch a tree diagram and show the expected risks. What decision should be taken?

## Solution

Solution. Sketch of tree diagram showing estimated probabilities of the true states of nature, costs in dollars, and, in bold, expected risks in dollars for different designs; o represents a node.

$$
\text { o - } \$ 300,000
$$

$$
/ 0.5
$$

- \$375,000 o
/ $\backslash 0.5$
0.9 / o-\$450,000
- \$382,500 0

Design / $\backslash 0.1 \quad \mathbf{0}-\$ 400,000$
$\mathrm{I}(\mathrm{a}) / \boldsymbol{0}$ / $\$ 450,000 \quad / 0.9$
o -------------------Design I(b)------------- -\$400,000 o
Design $\backslash$ o- $\$ 500,000 \quad$ o- $\$ 400,000$
II \ / 0.9

- $\$ 467,000$ o
$\backslash 0.1$
o -\$170,000
Choose Design $\mathrm{I}(a)$. The design risk is $-\$ 382,500$.
10.2. Decision tree and utility curve. A structural engineer has to choose an action from three alternatives $a_{1}, a_{2}$, and $a_{3}$. The state of nature to cope with is the modulus of elasticity used in the design. Suppose the unknown states of nature are approximated discretely by $\theta_{1}, \theta_{2}$ and $\theta_{3}$. The monetary values of action $a_{1}$ consequent to the three states of nature are estimated as $\$ 500,000, \$ 150,000$, and $-\$ 300,000$, respectively. The corresponding values are $\$ 250,000, \$ 200,000$, and $-\$ 100,000$ for action $a_{2}$ and $\$ 200,000 ; \$ 50,000$, and $-\$ 50,000$ for action $a_{3}$. Sketch the decision tree. Assigning a utility value of +1 unit to the highest of these monetary values and -1 unit to the lowest value, draw three utility curves to typically represent $(a)$ a risk seeker, $(b)$ a risk avoider and $(c)$ a large organization with a balanced view on risk taking. In case (c) what utility should be assigned for the outcome of the second action and the third state of nature.


## Solution

Sketch of decision tree showing actions $a$ and, in bold, monetary values of actions in dollars, consequent to the true states of nature $\theta$; $\mathbf{o}$ represents a node.


Sketch of three typical utility curves;- for $(a)$ a risk seeker $■(b)$ a risk avoider $\bullet$ and (c) a large organization with a balanced view on risk taking *.


Utility for second action and third state of nature using curve (c) is -0.5 .
10.3. Rural water supply. A contractor has the job of providing water to communities in a rural area by drilling boreholes. There is uncertainty regarding the depth of the ground water level. Experience elsewhere suggests an assumption of either 10 m or 20 m . It is necessary to acquire well-casing, pumps, and other equipment in advance because of time factors. If an incorrect choice of depth is made, a loss will be incurred in monetary units as follows:

| Water depth |  |  |
| :---: | :---: | :---: |
| Purchase equipment for depth | $\mathbf{1 0 m}$ | $\mathbf{2 0 m}$ |
| 10 m | no loss | 150 units |
| 20 m | 50 units | no loss |

From data of other wells in the region the following prior probabilities are assumed.
$\operatorname{Pr}($ depth 10 m$)=.7$
$\operatorname{Pr}($ depth 20 m$)=.3$.
Draw the decision tree and show the expected risks. Determine the Bayes rule. Compare with the minimax solution.

A hydrogeologist is consulted on the optimum depth. The following likelihoods are assigned to the predictions:

## Actual depth

| Indicated depth | 10 m | 20 m |
| :---: | :---: | :---: |
| 10 m | 0.7 | 0.1 |
| 20 m | 0.3 | 0.9 |

Obtain the posterior probabilities of the two states of nature conditional to the predictions. Determine the expected risks for each prediction.

## Solution

Sketch of decision tree showing water depth $d$ for which equipment is purchased, estimated prior probabilities of the true states of nature $\theta$, that is, depth of water of 10 m or 20 m , loss in monetary units if an incorrect decision is made and, in bold and boxed, expected risks in monetary units; $\mathbf{o}$ represents a node.
o 0
$\theta_{1} / 0.7$
[45] 0--- $\theta_{2}---\quad 150$
/ 0.3
$\begin{array}{cr}d_{1} / 10 m & \text { o } 50 \\ / & \theta_{1} / 0.7\end{array}$
$\begin{array}{cc}\mathbf{0}-------d_{2}-----------\left[\mathbf{3 5 ] - 0 - - - \theta _ { 2 } - \cdots \mathbf { o } 0} 0\right. \\ 20 m & 0.3\end{array}$
The following are given in monetary units.
Expected risks: $R\left(d_{1}\right)=45$ and $R\left(d_{2}\right)=35$.
Hence Bayes Rule is decision $d_{2}$
Minimax: For decision $d_{1}$, risk $=150$
For decision $d_{2}$, risk $=50$
Hence choose decision $d_{2}$
Posterior:

$$
\begin{aligned}
& P_{1}\left(\theta_{1} \mid d_{1}\right)=\frac{P_{0}\left(\theta_{1}\right) P\left(d_{1} \mid \theta_{1}\right)}{\sum_{i=1}^{2} P_{0}\left(\theta_{i}\right) P\left(d_{1} \mid \theta_{i}\right)}=\frac{0.49}{0.49+0.03}=\frac{0.49}{0.52}=0.9423 \\
& P_{1}\left(\theta_{2} \mid d_{1}\right)=1-0.9423=0.0577 . \\
& P_{1}\left(\theta_{2} \mid d_{2}\right)=\frac{P_{0}\left(\theta_{2}\right) P\left(d_{2} \mid \theta_{2}\right)}{\sum_{i=1}^{2} P_{0}\left(\theta_{i}\right) P\left(d_{2} \mid \theta_{i}\right)}=\frac{0.27}{0.27+0.21}=\frac{0.27}{0.48}=0.5625
\end{aligned}
$$

$$
P_{1}\left(\theta_{1} \mid d_{2}\right)=1-0.5625=0.4375 .
$$

Expected risks for decision $d_{1}, R\left(d_{1}\right)=0 \times 0.9423+150 \times 0.0577=8.7$
for decision $d_{2}, R\left(d_{2}\right)=50 \times 0.4375+0 \times 0.5625=21.9$
Hence choose decision $d_{2}$
10.4. Pipes and cofferdam. Find solutions to the pipes-for-water-supply (Example10.3) and cofferdam (Example10.4) problems using the following criteria:
(a) Bayesian theory with uniform prior
(b) Maximax (maximize maximum profit or minimize minimum loss)

## Solution

From Example 10.3 but assuming a vague prior distribution

| (a) True states of nature: | $\theta_{1}$ | $\theta_{2}$ | Expected costs |  |
| :--- | ---: | ---: | :---: | :--- |
| Decisions: | $d_{1}$ | -18.0 | -3.0 | -10.50 |
|  | $d_{2}$ | -18.3 | 35.4 | 8.55 |
|  | $d_{3}$ | -32.7 | -1.4 | $-17.05 \rightarrow$ optimum |
|  | $d_{4}$ | -33.0 | 37.0 | +2.00 |

(b) Maximax

Benefits
Decisions:

| $d_{1}$ | 18.0 |
| :--- | :--- |
| $d_{2}$ | 18.3 |
| $d_{3}$ | 32.7 |
| $d_{4}$ | 33.0 |

From Example 10.4 but assuming a vague prior distribution

| (a) True states of nature: | $\theta_{1}$ | $\theta_{2}$ | Expected costs |  |
| :--- | :--- | :---: | :---: | :---: |
| Decisions: | $d_{1}$ | -0 | $-60,000$ | $-30,000$ |
|  | $d_{2}$ | -750 | $-18,000$ | $-9,375$ |
|  | $d_{3}$ | $-3,000$ | $-6,000$ | $-4,500$ |
|  | $d_{4}$ | $-7,500$ | 0 | $-3,750 \rightarrow$ optimum |
|  | $d_{5}$ | $-15,000$ | 0 | $-7,500$ |

(b) Maximax ( or minimize costs)

Decisions

| $d_{1}$ | -0 |
| :---: | :---: |
| $d_{2}$ | -750 |
| $d_{3}$ | $-3,000$ |
| $d_{4}$ | $-7,500$ |
| $d_{5}$ | $-15,000$ |


| $-60,000$ |  |
| :---: | :--- |
| $-18,000$ |  |
| $-6,000$ |  |
| 0 | $\rightarrow$ optimum |
| 0 | $\rightarrow$ optimum |

10.5. Water projects. Water supply schemes are planned for three new towns, $X, Y$ and $Z$. Designs are based on projected populations 10 years hence. Future populations with approximated probabilities are as follows:

| Town | .25 | .50 | .25 |
| :---: | :---: | :---: | :---: |
| $X$ | 90,000 | 100,000 | 120,000 |
| Y | 125,000 | 150,000 | 175,000 |
| Z | 160,000 | 190,000 | 250,000 |

Assume that the water demand is 100 liters per day per head of population. The cost C in dollars per million liters per day varies with size $S$ of water supply scheme in liters per day as follows:

$$
\mathrm{C}=-\mathrm{S} / 10+100,000
$$

Assume additional supply is sold to local industries at $\$ 70,000$ per million liters per day and any shortfall is met from alternate sources at $\$ 130,000$ per million liters per day. Determine the optimum sizes of plants at $\mathrm{X}, \mathrm{Y}$ and Z on the basis of expected minimum costs and sizes given by the above forecasts of population.

## Solution

For town X with water supply of size S, $10^{6} \mathrm{~L} \quad 9 \quad 10 \quad 12$
$\begin{array}{llll}\text { Probabilities of } 3 \text { future populations as tabulated } & 0.25 & 0.5 & 0.25\end{array}$
Initial cost $C$ of supply $=\$(S / 10+100,000), \$ 10^{6} \quad 1.0 \quad 1.1 \quad 1.3$
SizeS $=10 \times 10^{6} \mathrm{~L}, \mathrm{E}[\mathrm{C}]=\$ 1,100,000-70,000 \times 0.25+0 \times 0.50+260,000 \times 0.25=$ \$1,147,500.
SizeS $=11 \times 10^{6} \mathrm{~L}, \mathrm{E}[\mathrm{C}]=\$ 1,200,000-140,000 \times 0.25-70,000 \times 0.50+130,000 \times 0.25=$ \$ 1,162,500.
SizeS $=12 \times 10^{6} \mathrm{~L}, \mathrm{E}[\mathrm{C}]=\$ 1,300,000-210,000 \times 0.25-140,000 \times 0.50+0 \times 0.25=$ \$ 1,177,500.
Optimum size of supply $=10 \times 10^{6} \mathrm{~L}$ and expected cost $=\$ 1,147,500$
$\begin{array}{lllll}\text { For town } Y & \text { with water supply of size } \mathrm{S}, & 10^{6} \mathrm{~L} & 12.5 & 15\end{array}$
$\begin{array}{lllll}\text { Probabilities of } 3 \text { future populations as tabulated } & 0.25 & 0.5 & 0.25\end{array}$
Initial cost C of supply $=\$(\mathrm{~S} / 10+100,000), \$ 10^{6} \quad 1.35 \quad 1.6 \quad 1.85$
Size $S=14 \times 10^{6} \mathrm{~L}, \mathrm{E}[\mathrm{C}]=\$ 1,500,000-105,000 \times 0.25+130,000 \times 0.50+3.5 \times 130,000 \times$ $0.25=\$ 1,652,500$.
SizeS $=15 \times 10^{6} \mathrm{~L}, \mathrm{E}[\mathrm{C}]=\$ 1,600,000-175,000 \times 0.25-0 \times 0.50+2.5 \times 130,000 \times 0.25=$ \$ 1,637,500.
SizeS $=16 \times 10^{6} \mathrm{~L}, \mathrm{E}[\mathrm{C}]=\$ 1,700,000-3.5 \times 70,000 \times 0.25-70,000 \times 0.50+195,000 \times 0.25=$ $\$ 1,652,500$.
Optimum size of supply $=15 \times 10^{6} \mathrm{~L}$ and expected cost $=1,637,500$
For town Z $\begin{array}{lllll}\text { with water supply of size } S, 10^{6} \mathrm{~L} & 16 & 19 & 25\end{array}$
$\begin{array}{llll}\text { Probabilities of } 3 \text { future populations as tabulated } & 0.25 & 0.5 & 0.25\end{array}$
Initial cost C of supply $=\$(\mathrm{~S} / 10+100,000), \$ 10^{6} \quad 1.7 \quad 2.0 \quad 2.6$
SizeS $=18 \times 10^{6} \mathrm{~L}, \mathrm{E}[\mathrm{C}]=\$ 1,900,000-140,000 \times 0.25+(1 \times 0.50+7 \times 0.25) \times 130,000$ $=\$ 2,157,500$.
SizeS $=19 \times 10^{6} \mathrm{~L}, \mathrm{E}[\mathrm{C}]=\$ 2,000,000-210,000 \times 0.25+0 \times 0.5+6 \times 0.25 \times 130,000=$ $\mathbf{\$ 2 1 4 2 , 5 0 0}$.
SizeS $=20 \times 10^{6} \mathrm{~L}, \mathrm{E}[\mathrm{C}]=\$ 2,100,000-(4 \times 0.25+0.50) \times 70,000+650,000 \times 0.25=$ \$ 2,157,500.
Optimum size of supply $=19 \times 10^{6} \mathrm{~L}$ and expected cost $=2,142,500$
10.6. Contractor's utility function. Suppose the utility function of the contractor in Example10.1 is defined by the following pairs of utilities (in the range 0 to 100 units) and gains in units of $\$ 100,000$ :

| Utility | Gain |
| ---: | ---: |
| 100 | 2.30 |
| 80 | 1.25 |
| 60 | 0.90 |
| 40 | 0.60 |
| 20 | 0.50 |
| 0 | 0.25 |

What decision should be taken if the expected utility were to be maximized and the probabilities of winning either contract are equal.

## Solution

Sketch of utility curve: Utility vs. Gain in \$


For action $d_{1} \quad R\left(\theta_{1}, d_{1}\right)=-241,000 \quad R\left(\theta_{2}, d_{1}\right)=-38,000$
Expected utility $=E[U]=0.5 \times(99+6)=52.5$.
For action $d_{2} \quad R\left(\theta_{1}, d_{2}\right)=-55,000 \quad R\left(\theta_{2}, d_{2}\right)=42,000$
Expected utility $=E[U]=0.5 \times(28+7)=17.5$
Take decision $d_{1}$ to maximize the expected utility.
10.7. Earth dam. Two designs are submitted for an earth dam. Design I is based on locally available materials, and its implementation is estimated to cost $\$ 1,000,000$. For Design II $5,000 \mathrm{~m}^{3}$ of a particular type of clay is required. The engineer's estimated pdf of the availability $X$ of the clayey material in the vicinity of the dam is uniform $\left(0,7000 \mathrm{~m}^{3}\right)$. The estimated cost of implementing Design II is $\$ 650,000$; however the average cost of hauling any extra material from outside the area at $\$ 100$ per $\mathrm{m}^{3}$ should be added. Which design should be accepted on the basis of expected least cost? What decision should be taken if the engineer's pdf for $X$ is

$$
f_{X}(x)=\frac{1}{3500} \exp (-x / 3500) ?
$$

## Solution

Design I Cost $=\$ 1,000,000$ based on local materials
Design II Cost $=\$ 650,000$ based on $5,000 \mathrm{~m}^{3}$ of a clayey material
(a) Uniform $\left(0,7000 \mathrm{~m}^{3}\right)$ distribution assumed for the pdf of the availability $X$ of the clayey material in the vicinity of the dam
$\mathrm{C}=$ cost of extra material $=\int_{0}^{5000} \frac{100(5000-x)}{7000} d x=\frac{100}{7000}\left[5000 x-x^{2} / 2\right]_{0}^{5000}=\$ 178,571$
Total cost of Design II $=\$ 650,000+\$ 178,591=\$ 828,571$.
(b) Simple exponential distribution with parameter $1 / 3500$ assumed for the pdf of the availability $X$ of the clayey material in the vicinity of the dam
$\mathrm{C}=$ cost of extra material $=\int_{0}^{5000} \frac{100(5000-x)}{3500} e^{-x / 3500} d x$
$=\frac{500}{3.5} \int_{0}^{5000} e^{-x / 3500} d x-\frac{1}{35} \int_{0}^{5000} x e^{-x / 3500} d x$
$=500 \times 1000\left[1-\frac{1}{e^{5 / 35}} \cdot\right]-\frac{1}{35} \mathrm{I}$
Where $\mathrm{I}=\left[3500 x e^{-x / 3500}\right]+\int_{0}^{5000} 3500 e^{-x / 3500} d x$
$=\frac{-3500 \times 5000}{e^{5 / 3.5}}+3500^{2}\left[1-\frac{1}{e^{5 / 3.5}}\right]$
$\mathrm{C}=150 \times 1000+350 \times 1000 / e^{10 / 7}=233,878$.
Total cost of Design II $=\$ 650,000+\$ 233,878=\$ 888,878$.
The answer is Design II in each case.
10.8. Soil strengths with uniform prior. In Example 10.11, the engineer decided that the prior distribution of the soil strength is $N\left(85,000 ; 11,000^{2}\right)$. Determine the posterior distribution and the optimum design strength assuming that the prior distribution is uniform $(75,000 ; 95,000)$.

## Solution

From Eqs. (10.2.17) and (10.2.18) and the ensuing discussion, the posterior mean $\mu_{1}=\bar{x}=70,000 \mathrm{~N} / \mathrm{m}^{2}$ and the posterior variance $\sigma_{1}^{2}=\sigma^{2} / n=15,000^{2} / 3$.
That is, the posterior distribution $N\left(70,000,15,000^{2} / 3\right)$.
For the given constants, $b=0.8, c=90,000$ and $k=0.00001$, the optimum design strength

$$
\begin{aligned}
a & =\mu_{1}+\frac{1}{k} \ln \left(\frac{b}{c k}\right)-\frac{1}{2} \times k \times \sigma_{1}^{2} \\
& =70000+\frac{1}{0.00001} \ln \left(\frac{0.8}{90000 \times 0.00001}\right)-\frac{1}{2} \times 0.00001 \times \frac{15,000^{2}}{3} \\
& =70,000-11,778-375=57,847 \mathrm{~N} / \mathrm{m}^{2}
\end{aligned}
$$

10.9. Loss functions. It was shown that in the case of squared loss function, the Bayes estimator is the mean value of the state of nature $\theta$. What estimator is obtained if the loss function is (a) constant and (b) linear with respect to $\theta$ ?

## Solution

(a) Constant loss function. Let us assume that
$l(\widetilde{\theta}, \theta)=\left\{\begin{array}{c}c ;-\infty<\theta<\widetilde{\theta}-\varepsilon / 2 \\ 0 ; \widetilde{\theta}-\varepsilon / 2<\theta<\widetilde{\theta}+\varepsilon / 2 ; \varepsilon \ll 1 \\ c ; \widetilde{\theta}+\varepsilon / 2<\theta<\infty\end{array}\right.$
Then $R(\widetilde{\theta}, d)=E[l(\widetilde{\theta}, \theta)]$
$=\int_{-\infty}^{\tilde{\theta}-\varepsilon / 2} c f(\theta \mid x) d \theta+\int_{\tilde{\theta}+\varepsilon / 2}^{\infty} c f(\theta \mid x)=\int_{-\infty}^{\infty} c f(\theta \mid x) d \theta-\int_{\tilde{\theta}-\varepsilon / 2}^{\tilde{\theta}+\varepsilon / 2} c f(\theta \mid x) d \theta$.
$\approx c-c \varepsilon f(\widetilde{\theta} \mid x)$
Thus $\frac{d R(\widetilde{\theta}, d)}{d \widetilde{\theta}}=-c \varepsilon \frac{d f(\widetilde{\theta} \mid x)}{d \widetilde{\theta}}$.
Then imposing $\frac{d R(\widetilde{\theta}, d)}{d \widetilde{\theta}}=0$, we get $\frac{d f(\widetilde{\theta} \mid x)}{d \widetilde{\theta}}=0$.
Therefore, the Bayes estimator for a constant loss function is the posterior mode of $\theta$.
(b) Linear loss function. Let us assume that the loss function is $l(\widetilde{\theta}, \theta)=c|\widetilde{\theta}-\theta|$. Then
$R(\widetilde{\theta}, d)=E[l(\widetilde{\theta}, \theta)]=\int l(\widetilde{\theta}, \theta) f(\theta \mid x) d \theta=\int c|\widetilde{\theta}-\theta| f(\theta \mid x) d \theta$ is the associated risk.
Simple probabilistic considerations would immediately indicate that the above
expression for $R(\widetilde{\theta}, d)$ is minimized by $\widetilde{\theta}=\underline{\text { median } \mathbf{~ o f ~} \boldsymbol{\theta}}$, which corresponds to the Bayes estimator of $\theta$. We can show this as follows:
$\frac{d R(\widetilde{\theta}, d)}{d \widetilde{\theta}}=\frac{d}{d \widetilde{\theta}}\left[\int_{-\infty}^{\tilde{\theta}} c(\widetilde{\theta}-\theta) f(\theta \mid x) d \theta+\int_{\widetilde{\theta}}^{\infty} c(\theta-\widetilde{\theta}) f(\theta \mid x) d \theta\right]$
$=\int_{-\infty}^{\tilde{\theta}} c f(\theta \mid x) d \theta-\int_{\theta}^{\infty} c f(\theta \mid x) d \theta$.
Then since $\int_{-\infty}^{\infty} f(\theta \mid x) d \theta=\int_{-\infty}^{\tilde{\theta}} f(\theta \mid x) d \theta+\int_{\theta}^{\infty} f(\theta \mid x) d \theta=1$,
and imposing $\frac{d R(\widetilde{\theta}, d)}{d \widetilde{\theta}}=0 \Rightarrow \int_{-\infty}^{\tilde{\theta}} f(\theta \mid x) d \theta=\int_{\tilde{\theta}}^{\infty} f(\theta \mid x) d \theta$, we get
$\int_{-\infty}^{\tilde{\theta}} f(\theta \mid x) d \theta=1 / 2=\int_{\theta}^{\infty} f(\theta \mid x) d \theta$.
That is, $\widetilde{\theta}=\underline{\text { median }} \underline{\text { of } \theta}$.
10.10. Traffic rates. In Example 10.10 ten exponentially distributed waiting times between successive vehicles are given.. Using this data, formulate and apply a likelihood ratio test in which the null hypothesis is that the parameter is 1 minute and the alternative hypothesis is that it is 0.9 minute, if (a) the prior probabilities are 0.4 and 0.6 , respectively, $(b)$ the prior probabilities are unknown. Show how the Type I and II errors of the test, $\alpha$ and $\beta$ respectively, can be calculated.

## Solution

NH: exponential parameter $\lambda=\theta_{0}=1.0$ minute
AH: exponential parameter $\lambda=\theta_{1}=0.9$ minute
$\frac{L_{0}}{L_{1}}=\frac{\prod_{i=1}^{n} f\left(x_{i} \mid \theta_{0}\right)}{\prod_{i=1}^{n} f\left(x_{i} \mid \theta_{1}\right)}=\left(\frac{\theta_{0}}{\theta_{1}}\right)^{n} e^{\left(\theta_{1}-\theta_{0}\right) \sum_{i=1}^{n} x_{i}}=\left(\frac{1.0}{0.9}\right)^{10} e^{(0.9-1.0) \times 10.2}=1.034176$
(a) The prior probability for value $\theta_{0}=1.0 \mathrm{~min}$ of the exponential parameter is $p_{0}=$
0.4 ; The prior probability for value $\theta_{1}=0.9 \mathrm{~min}$ of the exponential parameter is $p_{1}=$ 0.6

The critical region is limited by $\frac{L_{0}}{L_{1}} \leq \frac{p_{1}}{p_{0}}$
From the foregoing calculations $\frac{L_{0}}{L_{1}}<\frac{p_{1}}{p_{0}}=\frac{0.6}{0.4}=1.5$
Therefore we reject the NH.
(b) For the uniform prior distribution, $p_{0}=p_{1}=0.5$, Then $>\frac{p_{1}}{p_{0}}=1$
$\frac{L_{0}}{L_{1}}>\frac{p_{1}}{p_{0}}$. Therefore we not reject the NH.
(c) Under the NH, $\sum_{i=1}^{n} x_{i} \sim \operatorname{gamma}\left(n, \theta_{0}\right) \sim \chi_{2 n}^{2} /\left(2 \theta_{0}\right)$

This gives a procedure to calculate the probabilities $\alpha$ and $\beta$ of the Type I and II errors.
10.11. Traffic rates. In Example 10.10 estimate the Poisson parameter $\lambda$ using the posterior mean. Compare with the moments or ML estimator. What is the significance of the difference?

## Solution

For gamma, $\bar{x} \lambda=\frac{r+n}{\alpha+S_{x}}=\frac{0.5+10}{0.25+10.2}=\frac{10.5}{10.45}=1.005$
From sample data of intervals between vehicles $S_{x}=10.2$ and $n=10$ Hence

$$
\bar{x}=10.2 / 10=1.02
$$

For exponential, $\hat{\lambda}=\tilde{\lambda}=1 / \bar{x}=1 / 1.02=0.980$.
The difference is on account of the prior information.
10.12. Ecuador rainfalls. From the data used in Example 10.13 choose some series which meet the model requirements more closely and repeat the exercise of comparing the past averages with the JamesStein estimators for predicting the future averages (see Table E10.2). The lag-l serial correlation that are $l$ units apart in time) may be estimated, say, for $l=1,2$ and 3, as follows:

$$
r_{l}=\frac{\sum_{t=1}^{n-l}\left(x_{t}-\bar{x}\right)\left(x_{t+l}-\bar{x}\right)}{\left.\sum_{t=1}^{n}\left(x_{t}-\bar{x}\right)^{2}\right)}
$$

In an independent time series the $r_{l}, l \neq 0$ have an approximate $N(0,1 / n)$ distribution. Are the conclusions from the reduced data set substantially different?

## Solution

We follow the same procedure as in Example 10.14. The conclusions are the same.

