Statistica Neerlandica (1999) Vol. 53, nr. 3, pp. 376–382

# Problem Section

## Solutions

*Problem 345* (M. N. DESHPANDE, INDIA)
Let $m$, $n$ and $k$ be three integers such that $m + n = 3k$. Let us consider an urn containing $m$ white and $n$ black balls. We go on drawing 3 balls at a time without replacement at random from the urn and obtain $k$ groups of 3 balls. Let $X$ denote the number of groups which contain at least one black and at least one white ball. Clearly $X$ is a random variable and $0 \leq X \leq \min(k, m, n)$. Calculate the expectation and the variance of $X$.

Solutions were submitted by A. A. Jagers and the proposer, who solves the problem by solving nonhomogeneous linear difference equations with variable coefficients. Jagers applies symmetry arguments. We give his slightly modified solution. Moreover, he solves the problem for $m + n \geq 3k$.

*Solution by* A. A. JAGERS
We write $X = X_1 + X_2 + \cdots + X_k$, where $X_i$ is a $\{0, 1\}$-valued random variable with value 1 iff the $i$-th group of 3 balls shows two colours. Note that the joint probability distribution of $(X_1, \ldots, X_k)$ is invariant under permutation of the indices. So, the $X_i$ are identically distributed and something similar holds for the pairs $(X_i, X_j)$ with $i \neq j$. This implies that

$$\begin{aligned} \mathrm{E}X &= k\mathrm{E}X_1 = kP\{X_1 = 1\} \\ &= \frac{3kmn}{(m+n)(m+n-1)}. \end{aligned}$$

The last equality can easily be established. Similarly we have

$$\begin{aligned} \mathrm{E}X^2 &= k\mathrm{E}X_1^2 + k(k-1)\mathrm{E}X_1X_2 \\ &= kP\{X_1 = 1\} + k(k-1)P\{X_1 = X_2 = 1\}. \end{aligned}$$

It is not difficult to derive that

$$\mathrm{E}X^2 = \frac{3kmn}{(m+n)(m+n-1)}\left[1 + \frac{3(k-1)(m-1)(n-1)}{(m+n-2)(m+n-3)}\right].$$

Note that

$$\mathrm{E}X^2 = \mathrm{E}X[1 + \mathrm{E}X^-],$$

where $X^-$ denotes a random variable just like $X$ but with the parameters $k$, $m$, $n$ replaced with $k - 1$, $m - 1$, $n - 1$.

In case $m + n = 3k$, the following results are obtained:

$$\mathrm{E}X = \frac{mn}{m+n-1}$$
$$\mathrm{Var}\ X = \frac{mn(m-1)(n-1)}{(m+n-1)^2(m+n-2)}.$$

*Problem 346* (F. W. Steutel)
A game is played as follows. Two players $A$ and $B$ take turns in drawing random numbers, independent and uniformly distributed on (0, 1). The person who first draws a number smaller than the last number drawn by his opponent loses the game. Person $A$ draws the first number. Question: What is the probability $P$ that $A$ wins the game?

Solutions were submitted by D. Alma, A. A. Jagers, J. J. Lok and J. G. Schouten, M. H. van Raalte, and the proposer. Alma and Jagers have remarked that the solution will hold for any probability distribution with a continuous distribution function. Jagers and the proposer solved the problem by obtaining the solution of an integral equation. We present two solutions.

*Solution by* A. A. Jagers
For $0 < t < 1$, let $f(t)$ denote the probability that a player wins the game, if his opponent has just drawn a record value $t$ at his last turn so far. Then by conditioning on the outcome of the next drawing we have

$$f(t) = t \cdot 0 + \int_t^1 (1 - f(u))du \tag{1}$$

or, equivalently,

$$f'(t) = f(t) - 1 \text{ with } f(1-) = 0. \tag{2}$$

Either way, the unique solution is given by

$$f(t) = 1 - e^{t-1}$$

and

$$P = f(0+) = 1 - e^{-1}.$$

*Combined Solution by* D. Alma and M. H. van Raalte
Let $X_n$ denote the number drawn at the $n$th turn. The probability that the game ends at this turn is

$$P\{X_1 \leq X_2 \leq \cdots \leq X_{n-1} \text{ and } X_{n-1} \geq X_n\} = \frac{n-1}{n!}$$

The probability $P$ that $A$ wins is equal to

$$\begin{aligned} P &= \sum_{n \text{ is even}} \frac{n-1}{n!} = \sum_{n=1}^{\infty} \frac{2n-1}{(2n)!} \\ &= \sum_{n=1}^{\infty} \frac{1}{(2n-1)!} - \sum_{n=1}^{\infty} \frac{1}{(2n)!} \\ &= \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!} = 1 - e^{-1} = 0.63212. \end{aligned}$$

*Problem 347* (V. C. Hombas and A. S. Xenakis, Greece)
Consider a random sample of size $n$ from the Bernoulli distribution with parameter $p \in (0, 1)$.

(i) Find the Cramér-Rao lower bound for the variance of unbiased estimators of $p(1-p)$.
(ii) Find the uniformly minimum variance unbiased estimator of $p(1-p)$.

The problem has been solved by C. van Eeden, J. J. Lok, and the proposers. As Mrs. van Eeden points out, for $n = 1$ there does not exist such an unbiased estimator. If $\delta(X_1)$ would be an unbiased estimator, then we should have

$$p\delta(1) + (1-p)\delta(0) = p(1-p) \text{ for all } p \in (0, 1)$$

This is impossible. For $n \geq 2$ there exist unbiased estimators, e.g. $X_1(1 - X_2)$.

*Solution by* J. J. Lok

(i) The Cramér-Rao lower bound for estimating $g(p) = p(1-p)$ equals $\{g'(p)\}^2/(nI(p))$, where $I(p)$ denotes the Fisher information for the Bernoulli distribution. If $\ell(p; X_1)$ denotes the log-likelihood, then

$$\begin{aligned} I(p) &= -\mathrm{E}\frac{d^2\ell(p; X_1)}{dp^2} \\ &= \mathrm{E}\left[\frac{X_1}{p^2} + \frac{1-X_1}{(1-p)^2}\right] = \frac{1}{p(1-p)}. \end{aligned}$$

The Cramér-Rao lower bound for estimating $p(1-p)$ equals

$$\frac{(1-2p)^2}{np(1-p)}.$$

(ii) Let $X_1, \ldots, X_n$ denote a random sample from the Bernoulli distribution. From the theory on exponential families we know that $T = X_1 + \cdots + X_n$ is a complete sufficient statistic. Because of the Lehmann-Scheffé theorem any UMVU estimator of $g(p) = p(1-p)$ is a function of $T$. Since $\hat{p} = T/n$ is an obvious estimator for $p$, we try $\hat{p}(1-\hat{p})$:

$$\begin{aligned} \mathrm{E}\hat{p}(1-\hat{p}) &= p - (\mathrm{Var}\,\hat{p} + (\mathrm{E}\hat{p})^2) \\ &= p - \frac{p(1-p)}{n} - p^2 \\ &= \frac{n-1}{n} p(1-p). \end{aligned}$$

So, $n(n-1)^{-1}\hat{p}(1-\hat{p})$ is the required UMVU estimator.

Remark. Mrs. van Eeden also showed how this estimator can be obtained by direct computation. If $\delta(T)$ denotes the required estimator, then we have to solve

$$\sum_{i=0}^{n} \delta(i)\binom{n}{i} p^i (1-p)^{n-i} = p(1-p) \text{ for all } p \in (0, 1).$$

Letting $\lambda = p/(1-p)$, this is equivalent to

$$\sum_{i=0}^{n} \delta(i)\binom{n}{i} \lambda^{i-1} = (1+\lambda)^{n-2} \text{ for all } \lambda \geq 0.$$

On comparing the powers of $\lambda$ the following unique solution is obtained

$$\begin{aligned} \delta(0) &= \delta(n) = 0 \\ \delta(i)\binom{n}{i} &= \binom{n-2}{i-1}, \quad i = 1, \ldots, n-1. \end{aligned}$$

Hence it follows that

$$\delta(T) = \frac{T(n-T)}{n(n-1)} = \frac{n}{n-1}\hat{p}(1-\hat{p}).$$

*Problem 348* (A. G. M. Steerneman)

Let $X$, $Y$ be standard normally distributed random variables with correlation coefficient $\rho$. Define

$$F(x, y, \rho) = P\{X \leq x, Y \leq y\}.$$

Moreover, let $f(x, y, \rho)$ denote the joint probability density function of $X$ and $Y$. Show that

$$\frac{\partial F(x, y, \rho)}{\partial \rho} = f(x, y, \rho).$$

Is there a short proof?

The problem has been solved by A. Müller (Germany) and the proposer; the much shorter solution is due to A. Müller.

*Solution by* A. Müller

A short and elegant proof of this result can be derived, if one writes the density $f(x, y, \rho)$ as the transform of its characteristic function. Let

$$\begin{aligned}\varphi(t_1, t_2) &= \mathrm{E}\exp(it_1X + it_2Y)\\ &= \exp(-\tfrac{1}{2}(t_1^2 + 2\rho t_1t_2 + t_2^2))\end{aligned}$$

be the characteristic function of the random vector $(X, Y)$. It is well known that

$$f(x, y, \rho) = \frac{1}{(2\pi)^2}\int\int \exp\left(-it_1x - it_2y - \frac{1}{2}(t_1^2 + 2\rho t_1t_2 + t_2^2)\right)dt_1dt_2.$$

Hence

$$\begin{aligned}\frac{\partial}{\partial\rho}f(x, y, \rho) &= \frac{1}{(2\pi)^2}\int\int -t_1t_2\exp(-it_1x - it_2y)\varphi(t_1, t_2)dt_1dt_2\\ &= \frac{\partial^2}{\partial x\partial y}f(x, y, \rho).\end{aligned}$$

Integrating both sides of this equality with respect to $x$ and $y$ yields.

$$\frac{\partial}{\partial\rho}F(x, y, \rho) = f(x, y, \rho).$$

The idea of this proof can be found e.g. on page 8 ff. in the book of Y. L. Tong (1980), Probability inequalities in multivariate distributions; Academic Press, New York, who uses this identity to prove Slepian's inequality.

*Problem 349* (V. C. Hombas, Greece)

In proving Paul Lévy's theorem, the following result is used:

$$\lim_{n\to\infty}\int_0^{\pi}\sin\left(\left(n + \frac{1}{2}\right)x\right)\left[\frac{1}{\frac{1}{2}x} - \frac{1}{\sin\frac{1}{2}x}\right]dx = 0.$$

Prove this result.

A. A. Jagers and the proposer have submitted solutions to this problem.

*Solution by* A. A. Jagers

The statement of the problem is a particular instance of the renowned Riemann–Lebesgue lemma:

"If $\hat{f}$ denotes the Fourier transform of the $L^1$-function $f$, then

$$\lim_{y\to\pm\infty}\hat{f}(y) = 0."$$

Cf. e.g. E. Hewitt & K. Stromberg, Real and Abstract Analysis, Springer (1965), New York.

Note that

$$\frac{1}{t} - \frac{1}{\sin t} = \frac{\sin - t}{t \sin t} \sim \frac{t}{6}(t \to 0),$$

so that $x = 0$ is a removable singularity in the case at hand.

*Problem 350* (A. G. M. Steerneman)
Let $X$ and $Y$ be independently distributed according to a Cauchy distribution, i.e. their probability density function is

$$f(x) = \frac{\gamma}{\pi(\gamma^2 + x^2)}, x \in \mathbb{R}.$$

Show that $XY$ and $X/Y$ have the same distribution and derive its probability density function.

Solutions have been submitted by C. van Eeden and A. A. Jagers. Both they correctly remark that $\gamma = 1$ should be taken in the formulation of the problem.

*Solution by* C. van Eeden
Note that $X$ and $Y$ are distributed as the ratio of two i.i.d. standard normal random variables. So, $Y$ and $1/Y$ have the same distribution. Hence $XY$ and $X/Y$ have the same distribution. Define $T = X/Y$. For $t > 0$ we obtain

$$P\{T \le t\} = \frac{1}{\pi^2}\left(\int_{-\infty}^{0} \frac{1}{1+y^2} \int_{ty}^{\infty} \frac{1}{1+x^2}\, dxdy + \int_{0}^{\infty} \frac{1}{1+y^2} \int_{-\infty}^{ty} \frac{1}{1+x^2}\, dxdy\right).$$

Taking the derivative with respect to $t$ gives the following expression for the density $g(t)$ of $T$ for $t > 0$:

$$\pi^2 g(t) = -\int_{-\infty}^{0} \frac{y}{(1+y^2)(1+t^2y^2)}\, dy + \int_{0}^{\infty} \frac{y}{(1+y^2)(1+t^2y^2)}\, dy$$
$$= 2\int_{0}^{\infty} \frac{y}{(1+y^2)(1+t^2y^2)}\, dy.$$

Using the fact that

$$\frac{y}{(1+y^2)(1+t^2y^2)} = \frac{y}{t^2-1}\left(\frac{t^2}{1+t^2y^2} - \frac{1}{1+y^2}\right)$$

gives, for $t > 0$,

$$\pi^2 g(t) = \frac{\log t^2}{t^2 - 1}.$$

Further, of course, $g(t) = g(-t)$ for all $t \neq 0$.

## New Problems

*Problem 354*** (M. N. Deshpande, India)
Let $n$ and $k$ be positive integers greater than or equal to 3. Divide integers $(1, 2, 3, \ldots, nk-1, nk)$ in $n$ groups, every group containing $k$ elements each, such that the following conditions are satisfied. Let $X_i$ denote the number randomly selected from group $i$; all integers in every group have equal chance of selection. Then it is required that

$$P\{X_i > X_{i+1}\} > \tfrac{1}{2} \text{ for } i = 1, 2, \ldots, n-1 \qquad (1)$$

$$P\{X_n > X_1\} > \tfrac{1}{2}. \qquad (2)$$

The problem is to obtain these groups, e.g. for $n = 7$, $k = 4$ and $n = 4$, $k = 7$.

*Problem 355* (V. C. Hombas, Greece)
The following are the Carver's data (cited by Norton (1956), One likelihood adjustment may be inadequate, Biometrics, vol. 12, 79–81) for two factors in maize, starchy versus sugary and green versus white.

| Starchy | | Sugary | | Total |
|---|---|---|---|---|
| Green | White | Green | White | |
| 1997 | 906 | 904 | 32 | 3839 |

These are the frequencies that have been observed. Assume that the multinomial model is adequate and that the probabilities of belonging to the four classes are $\frac{1}{4}(2+\theta)$, $\frac{1}{4}(1-\theta)$, $\frac{1}{4}(1-\theta)$, $\frac{1}{4}\theta$, where $0 \leq \theta \leq 1$. Calculate the maximum likelihood estimator of $\theta$ and obtain an estimate for its variance.

Problems marked with * are nonelementary, of problems marked with ** no solution is known to the editor; unmarked problems are not necessarily simple. Solutions of the problems in this issue should arrive before January 31, 2000. Problems (preferably with solutions) and solutions (type-written on separate sheets bearing the name of the solver) are welcomed by the column editor. Contributors are requested to use electronic mail and to submit material in LATEX format, whenever this is possible, because this speeds up processing the problem section.