Statistica Neerlandica (2000) Vol. 54, nr. 3, pp. 378–382

# Problem Section

## Solutions

V. C. Hombas and A. S. Xenakis (Greece) have submitted a solution to problem 351. The solution by the proposer, F. W. Steutel, has already been published in Statistica Neerlandica, vol. 54 (2000), page 111.

*Problem 354*** (M. N. DESHPANDE, INDIA)
Let $n$ and $k$ be positive integers greater than or equal to 3. Divide integers (1, 2, 3, ..., $nk-1$, $nk$) into $n$ groups, each group containing $k$ elements, 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

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

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

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

The problem has been solved by M. S. Keane. The proposer has given a solution to the cases $n = 7$, $k = 4$ and $n = 4$, $k = 7$. These solutions differ from those according to the general solution by Keane.

*Solution by* M. S. KEANE
It is notationally slightly easier to work with solutions satisfying

$$P\{X_i < X_{i+1}\} > \tfrac{1}{2} \text{ for } i = 1, 2, \ldots, n-1 \quad (3)$$

$$P\{X_n < X_1\} > \tfrac{1}{2}; \quad (4)$$

this just means that the order in which the groups are numbered is reversed. Here is a somewhat inelegant solution to this problem. First let $k \geqslant 3$ be fixed. We represent a solution for $n = 3$ and this fixed $k$ by a sequence

$$x_1, x_2, \ldots, x_{3k}, \quad (5)$$

where each $x_j = 1$, 2 or 3 and where we interpret $x_j = i$ to mean that the integer $j$ belongs to group $i$. Given such a solution we can obtain a solution for any $n \geqslant 4$ by simply replacing each occurrence $x_j = 3$ in the sequence (5) by the finite string

$$3, 4, \ldots, n.$$

It is then obvious that the new sequence thus obtained satisfies (3) and (4), if the original sequence did. So, if we have a solution for $n = 3$ and some fixed $k$, then we have a solution for arbitrary $n$ and this $k$.

Now we consider the problem for $n = 3$. If $k = 3$ then the sequence

$$123 \quad 312 \quad 231 \tag{6}$$

satisfies (3) and (4); the three groups are

$$\{1, 5, 9\}, \quad \{2, 6, 7\}, \quad \{3, 4, 8\}.$$

To obtain a solution for $k = 3$ and $n = 5$, use the sequence

$$12345 \quad 34512 \quad 23451$$

yielding the five groups

$$\{1, 9, 15\}, \quad \{2, 10, 11\}, \quad \{3, 6, 12\}, \quad \{4, 7, 13\}, \quad \{5, 8, 14\}.$$

Now comes the inelegance. For $k = 3$ sequence (6) suffices, and for any $k \geqslant 5$ periodic repetition of the sequence suffices. E.g. if $k = 7$ then

$$123 \quad 312 \quad 231 \quad 123 \quad 312 \quad 231 \quad 123$$

satisfies (3) and (4). This works for all values of $k$ except $k = 4$. For $k = 4$, the sequence

$$123 \quad 122 \quad 333 \quad 112$$

clearly satisfies (3) and (4).

*Solution by* THE PROPOSER, $n = 7$, $k = 4$ and $n = 5$, $k = 7$.
The solution uses the original formulation of the problem, i.e. the requirements (1) and (2). The case $n = 7$, $k = 4$:

group

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| 7 | 6 | 5 | 4 | 3 | 2 | 1 |
| 13 | 12 | 11 | 10 | 9 | 8 | 23 |
| 18 | 17 | 16 | 15 | 14 | 25 | 24 |
| 22 | 21 | 20 | 19 | 28 | 27 | 26 |

The case $n = 4$, $k = 7$:

group

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| 4 | 3 | 2 | 1 |
| 7 | 6 | 5 | 11 |
| 9 | 8 | 13 | 12 |
| 10 | 16 | 15 | 14 |
| 29 | 19 | 18 | 17 |
| 24 | 23 | 22 | 21 |
| 28 | 27 | 26 | 25 |

*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 one of the four classes are $\frac{1}{4}(2+\theta)$, $\frac{1}{4}(1-\theta)$, $\frac{1}{4}(1-\theta)$, $\frac{1}{4}\theta$, where $0 \leqslant \theta \leqslant 1$. Calculate the maximum likelihood estimator for $\theta$ and obtain an estimate for its variance.

*Solution by* THE PROPOSER
Let $X = (X_1, \ldots, X_4)^t$ follow a multinomial distribution $M(n;\ p)$, where $n = 3839$ and $p = p(\theta) = (\frac{1}{4}(2+\theta), \frac{1}{4}(1-\theta), \frac{1}{4}(1-\theta), \frac{1}{4}\theta)^t$. The log-likelihood is of the form

$$\ln L(\theta;\ X) = C + X_1 \ln(2+\theta) + (X_2 + X_3)\ln(1-\theta) + X_4 \ln\theta.$$

The first order derivative with respect to $\theta$ is

$$\frac{\partial \ln L(\theta;\ X)}{\partial\theta} = \frac{X_1}{2+\theta} - \frac{X_2+X_3}{1-\theta} + \frac{X_4}{\theta}.$$

The second order derivative is

$$\frac{\partial^2 \ln L(\theta;\ X)}{\partial\theta^2} = -\frac{X_1}{(2+\theta)^2} - \frac{X_2+X_3}{(1-\theta)^2} - \frac{X_4}{\theta^2}.$$

The maximum likelihood estimator $\hat{\theta}$ for $\theta$ is obtained putting the first order derivative equal to zero. So, by solving

$$\frac{1997}{2+\theta} - \frac{1810}{1-\theta} + \frac{32}{\theta} = 0,$$

which is elementary, we obtain $\hat{\theta} = 0.035712$. The asymptotic variance of $\hat{\theta}$ is obtained by considering the Fisher information:

$$I(\theta) = -E_\theta \frac{\partial^2 \ln L(\theta;\ X)}{\partial\theta^2}$$

$$= \frac{n}{4}\left[\frac{1}{2+\theta} + \frac{2}{1-\theta} + \frac{1}{\theta}\right].$$

By plugging in $\hat{\theta}$ we obtain as an estimate for its variance $I^{-1}(\hat{\theta}) = 0.00003408$. So, the estimated standard deviation is 0.005838.

**New Problems**

*Problem 357* (V. C. HOMBAS, GREECE)
A discrete random vector $(X, Y)$ has a joint distribution that is uniform on the four points $(-l, -k-1)$, $(-l, k-1)$, $(l, k-1)$, $(l, k+1)$, where $l, k \neq 0$.

(a) Show that the correlation coefficient $\rho$ for $(X, Y)$ does not depend on the magnitude $l$ and find the linear regression of $Y$ on $X$.
(b) Show that perfect linearity will not be achieved no matter how close $|\rho|$ is to one and if $k = l \to \infty$, then $|\rho| \to 1$.

*Problem 358* (F. W. STEUTEL)
Let $k$ balls be distributed randomly over $n$ boxes numbered $1, 2, \ldots, n$ with $k < n$. Calculated the probability that the first box contains less than one ball (i.e. zero balls), the first two boxes together contain less than two balls, ..., the first $n-1$ boxes together contain less than $k-1$ balls.

*Problem 359*** (N. SCHMITZ, GERMANY) ('Prophet inequality for the general $S_n/n$-problem')
Let $X_1, \ldots, X_n$, $n \geqslant 2$, be [0,1]-valued random variables (with arbitrary dependencies), $S_i := \sum_{j=1}^{i} X_j$ and

$$V(X_1, \ldots, X_n) := \sup\{E(S_\tau/\tau)\colon \tau \text{ stopping rule } \leqslant n\},$$
$$M(X_1, \ldots, X_n) := E(\max_{1\leqslant i\leqslant n} S_i/i),$$
$$\gamma_n := \sup_{(X_1,\ldots,X_n)} M(X_1, \ldots, X_n)/V(X_1, \ldots, X_n).$$

The problem is to determine $\gamma_n$, for $n \geqslant 3$, and to decide whether $\gamma_n$ is attained, i.e. whether there exist $X_1^*, \ldots, X_n^*$ such that $M(X_1^*, \ldots, X_n^*)/V(X_1^*, \ldots, X_n^*) = \gamma_n$.

*Remark*: Obviously $\gamma_2 = 3/2$; this bound is not attained.

*Conjectures*:

a) Special case $n = 3$:
(i) For $x \in (0, 1)$

$$X_1 \equiv x,$$

$$X_2 = \begin{cases} 1 & \\ & \text{with prob.} \\ 0 & \end{cases} \quad \begin{matrix} x \\ \\ 1-x \end{matrix}$$

$$X_3 = \begin{cases} 1 & \\ & \text{with prob.} \\ 0 & \end{cases} \quad \begin{matrix} S_2/2 \\ \\ 1-S_2/2 \end{matrix}$$

yields

$$\gamma_3(x) := \sup_{(x,X_2,X_3)} M(x, X_2, X_3)/V(x, X_2, X_3)$$

$$= \begin{cases} (7-4x+x^2)/4 & \text{for } x \leqslant \frac{1}{2}, \\ (19-6x-x^2)/12 & \text{for } x \geqslant \frac{1}{2}. \end{cases}$$

(ii) $\gamma_3 = \lim_{x\downarrow 0} \gamma_3(x) = \frac{7}{4}$.

(iii) $\gamma_3$ fails to be attained.

b) General $n$:

(i) For $x \in (0,1)$

$$X_1 \equiv x,\ X_{i+1} = \begin{cases} 1 & \\ & \text{with prob.} \\ 0 & \end{cases} \begin{matrix} S_i/i \\ \\ 1 - S_i/i \end{matrix},\ 1 \leqslant i \leqslant n-1$$

yields $\gamma_n(x) := \sup_{(x,X_2,\ldots,X_n)} M(x, X_2, \ldots, X_n)/V(x, X_2, \ldots, X_n)$.

(ii) $\gamma_n = \lim_{s\downarrow 0} \gamma_n(x)$.

(iii) $\gamma_n$ fails to be attained.

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, 2001. 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 the processing of the Problem Section.