NOTES

A MARKOV CHAIN DERIVATION OF DISCRETE DISTRIBUTIONS

By F. G. Foster

Magdalen College, Oxford

Let an irreducible, aperiodic Markov chain[1] have the matrix of transition probabilities, $\mathbf{A} = [p_{ij}]$ $(i, j = 0, 1, 2, \cdots)$. Then as usual we shall have

$$p_{ij} \geqq 0 \qquad \text{for all } i \text{ and } j,$$

$$\sum_{j=0}^{\infty} p_{ij} = 1 \qquad \text{for all } i.$$

It is known ([1], p. 325) that the nth power of $\mathbf{A}$, $\mathbf{A}^n$, tends to a limiting matrix as $n \rightarrow \infty$

$$\lim_{n \to \infty} \mathbf{A}^n = \mathbf{B},$$

and $\mathbf{B}$ will either be null or have the identical rows,

$$\mathbf{x} = (x_0, x_1, \cdots),$$

such that $x_i > 0$ for all i and $\sum_{i=0}^{\infty} x_i = 1$. Moreover we shall have

$$\mathbf{xA} = \mathbf{x}.$$

In this way we may make correspond to any matrix $\mathbf{A}$, of the type under consideration, either the null vector or a probability distribution represented by $\mathbf{x}$. Conversely, to any distribution $\mathbf{x}$ there will correspond a matrix $\mathbf{A}$ (not necessarily unique). A method of constructing such a matrix is given below and illustrated with some examples.

Let $\{a_i\}$ $(i = 0, 1, 2, \cdots)$ be a sequence of positive numbers and define $A_n = \sum_{i=0}^{n} a_i$ $(n = 0, 1, 2, \cdots)$. Now let

$$\mathbf{A} = \begin{bmatrix} \frac{a_0}{A_1} & \frac{a_1}{A_1} & 0 & 0 & 0 & \cdots \\ \frac{a_0}{A_2} & \frac{a_1}{A_2} & \frac{a_2}{A_2} & 0 & 0 & \cdots \\ \frac{a_0}{A_3} & \frac{a_1}{A_3} & \frac{a_2}{A_3} & \frac{a_3}{A_3} & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}.$$

Then $\mathbf{A}$ satisfies the usual conditions for being a transition probability matrix;

[1] For definitions of all terms used see [1].

moreover it is clearly irreducible, and, since its diagonal elements are all positive, it is also aperiodic. Now suppose $\mathbf{x}$ is a vector such that

$$\mathbf{xA} = \mathbf{x}.$$

If we regard this as an equation in $\mathbf{x}$, we find that in our special case the solution is easily obtained. We have

$$x_1 = \frac{a_1}{a_0} x_0,$$

$$\frac{x_{n+1}}{a_{n+1}} = \frac{x_n}{a_n} - \frac{x_{n-1}}{A_n}, \qquad n \geqq 1.$$

It follows by induction that

$$x_n = x_0 \frac{a_1 a_2 \cdots a_n}{A_0 A_1 \cdots A_{n-1}}, \qquad n \geqq 1,$$

and so $\mathbf{x}$ is uniquely determined, apart from a common factor. For $\mathbf{x}$ to be a distribution, we must have in addition

$$1 = \sum_{n=0}^{\infty} x_n = x_0 + x_0 \sum_{n=1}^{\infty} \frac{a_1 \cdots a_n}{A_0 \cdots A_{n-1}}.$$

Thus

$$x_0 = 1 \Big/ \left(1 + \sum_{n=1}^{\infty} \frac{a_1 \cdots a_n}{A_0 \cdots A_{n-1}}\right),$$

and it follows that the matrix $\mathbf{B}$ ($= \lim \mathbf{A}^n$) is non-null if and only if

$$\sum_{n=1}^{\infty} \frac{a_1 \cdots a_n}{A_0 \cdots A_{n-1}} < \infty,$$

and each row of $\mathbf{B}$ will consist of the distribution $\mathbf{x}$.

Conversely, if we have given a distribution $\mathbf{x}$, we may calculate the sequence $\{a_i\}$ which possesses the required property. We have only to put

$$\frac{a_n}{A_{n-1}} = \frac{x_n}{x_{n-1}}, \qquad n \geqq 1.$$

Then we find as required that

$$x_0 \frac{a_1 \cdots a_n}{A_0 \cdots A_{n-1}} = x_n.$$

The sequence $\{a_i\}$ is now easily calculated. We have

$$A_n - A_{n-1} = a_n = A_{n-1} \frac{x_n}{x_{n-1}}.$$

Therefore

$$A_n = A_{n-1}\left(1 \quad \frac{x_n}{x_{n-1}}\right),$$

and, by iteration,

$$A_n = a_0 \prod_{i=1}^{n} \left(1 + \frac{x_i}{x_{i-1}}\right).$$

Hence, putting (as we may, since a common factor is unimportant) $a_0 = 1$, we have

$$a_n = \frac{x_n}{x_{n-1}} \prod_{i=1}^{n-1} \left(1 + \frac{x_i}{x_{i-1}}\right), \qquad n \geqq 1.$$

The above procedure may be given the following interpretation. Consider a particle performing a random walk on the integers 0, 1, 2, $\cdots$ in such a manner that when in position n it has probabilities in the ratios

$$a_0 : a_1 : a_2 : \cdots : a_{n+1}$$

of jumping at the next move into one of the positions 0, 1, 2, $\cdots$, $n + 1$. That is, the particle can move either one step along or back to any previous position. The distribution $\{x_i\}$ may then be interpreted as giving the asymptotic probabilities for its position after a large number of moves, and we have shown how the sequence $\{a_i\}$ may be calculated to give any required asymptotic distribution $\{x_i\}$ (with $x_i > 0$ for all i).

In some cases where $\mathbf{x}$ is a recognised distribution the sequence $\{a_i\}$ has a particularly simple form.

Example (a). *The Poisson distribution.* Let us take

$$x_n = e^{-\lambda}\lambda^n/n!, \qquad \lambda > 0, \qquad n = 0, 1, 2 \cdots.$$

We find that

$$a_n = \frac{1}{n!}\lambda(\lambda + 1) \cdots (\lambda + n - 1), \qquad n = 1, 2, \cdots,$$

with $a_0 = 1$. Thus the nth row of $\mathbf{A}$ is a truncated negative binomial distribution having $n + 1$ terms. In particular, when $\lambda = 1$, $\mathbf{A}$ takes the very simple form wherein $a_n \equiv 1$ for all n, and the rule governing the motion of the particle is that when it is in the nth position it has *equal* probabilities of jumping into any of the positions 0, 1, 2, $\cdots$, $n + 1$. We have then the result that its asymptotic position is a random variable with the Poisson distribution $\{1/(en!)\}$ $(n = 0, 1, 2, \cdots)$.

Example (b). *The negative binomial distribution.* Let us take

$$x_n = (1 - \beta)^{\lambda} \frac{1}{n!} \lambda(\lambda + 1) \cdots (\lambda + n - 1)\beta^n, \quad n = 1, 2, \cdots,$$

$$x_0 = (1 - \beta)^{\lambda},$$

where $0 < \beta < 1, \lambda > 0$. We find that

$$a_n = \frac{\lambda + n - 1}{n!} \beta(1 + \lambda\beta) \cdots (n - 1 + [\lambda + n - 2]\beta), \qquad n = 1, 2, \cdots,$$

with $a_0 = 1$. In particular, when $\lambda = 1$, we have

$$a_n = (1 + \beta)^{n-1}\beta, \qquad n = 1, 2, \cdots,$$

with $a_0 = 1$, and each row of **A** is a truncated modified geometric distribution.

REFERENCE

[1] WILLIAM FELLER. *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, 1950.

ON MINIMUM VARIANCE ESTIMATORS[1]

BY J. KIEFER

Cornell University

Chapman and Robbins [1] have given a simple improvement on the Cramér-Rao inequality without postulating the regularity assumptions under which the latter is usually proved. The purpose of this note is to show by examples how a similarly derived stronger inequality (see equation (2)) may be used to verify that certain estimators are uniform minimum variance unbiased estimators. This stronger inequality is that which (under additional restrictions) was shown in [2] to be the best possible, but is in a more useful form for applications than the form given in [2]. For simplicity we consider only an inequality on the variance of unbiased estimators, but inequalities on other moments than the second (see [2]), or for biased estimators, may be found similarly. The two examples considered here are ones where the regularity conditions of [2] are not satisfied, where the method of [1] does not give the best bound, and where the method of this note is used to find the best bound and thus to verify that certain estimators are uniform minimum variance unbiased. (For the examples considered this also follows from completeness of the sufficient statistic; the method used here applies, of course, more generally.)

Let X be a chance variable with density $f(x; \theta)$ with respect to some fixed σ-finite measure μ. ($\theta \varepsilon \Omega$, $x \varepsilon \mathfrak{X}$). We suppose suitable Borel fields to be given and $f(x; \theta)$ to be measurable in its arguments. Ω is a subset of the real line. For each θ, let $\Omega_\theta = \{h \mid (\theta + h) \varepsilon \Omega\}$. For fixed θ, let λ_1 and λ_2 be any two probability measures on Ω_θ such that $E_i h = \int_{\Omega_\theta} h d\lambda_i(h)$ exists for $i = 1, 2$. Then, for any

[1] Research sponsored by the Office of Naval Research.