In the example in the previous paragraph, there was no UMVUE for $\tau(\theta) = \theta$, the parametric function of interest. But, what if a UMVUE exists for $\tau(\theta)$ and there is a minimally sufficient but not complete statistic T? Could we use T to Rao–Blackwellize an unbiased estimator h of $\tau(\theta)$ to obtain the UMVUE (i.e., could it serve the same purpose as the complete sufficient statistic in the Rao–Blackwell and Lehmann–Scheffé theorems)? The answer is negative because even though the UMVUE of $\tau(\theta)$, which can be made into a function of T alone by using the Rao–Blackwell Theorem, is unique (cf., Rohatgi 1976, p. 353), there can still exist many unbiased estimators of $\tau(\theta)$ that are functions of T. An example of this phenomenon is provided by taking X from the probability function

$$f_X(k;\theta) = \mathbf{P}_\theta\{X = k\} = \theta \quad \text{if } k = -1$$
$$= (1-\theta)^2\theta^k \quad \text{if } k = 0, 1, 2, \ldots,$$

where $\theta \in \Theta = (0, 1)$. Clearly, X is minimally sufficient but not complete. Letting $\tau(\theta) = (1-\theta)^2$, it is known (cf., Lehmann 1983, pp. 76–77) that the UMVUE of $\tau(\theta)$ is

$$h_1^*(X) = 1 \quad \text{if } X = 0$$
$$= 0 \quad \text{if } X \neq 0.$$

But another unbiased estimator of $\tau(\theta)$ that also depends on X is $h_2(X) = h_1^*(X) - aX$ for any $a \in (-\infty, \infty)$. This phenomenon shows the importance of the concept of completeness in guaranteeing that the process of Rao–Blackwellizing an unbiased estimator h of $\tau(\theta)$ will produce the UMVUE.

REFERENCES

Bickel, P., and Doksum, K. (1977), *Mathematical Statistics*, San Francisco: Holden-Day.

DeGroot, M. (1989), *Probability and Statistics* (2nd ed.), Reading, MA: Addison-Wesley.

Hogg, R., and Craig, A. (1978), *Introduction to Mathematical Statistics* (4th ed.), New York: Macmillan.

Lehmann, E. (1983), *Theory of Point Estimation*, New York: John Wiley.

——— (1986), *Testing of Hypotheses* (2nd ed.), New York: John Wiley.

Mood, A., Graybill, F., and Boes, D. (1974), *Introduction to the Theory of Statistics* (3rd ed.), New York: McGraw-Hill.

Rohatgi, V. (1976), *An Introduction to Probability Theory and Mathematical Statistics*, New York: John Wiley.

Binomial and Negative Binomial Analogues Under Correlated Bernoulli Trials

Román VIVEROS, K. BALASUBRAMANIAN, and N. BALAKRISHNAN

Several statistical applications demand the adoption of models in which the response is binary but the outcomes of different trials exhibit some degree of correlation. Although the independent case is well known and treated even in elementary textbooks, results on correlated Bernoulli trials are hardly found in the literature. Analogues of the binomial and negative binomial distributions are presented in this article when the correlation is of the Markovian type. Probability-generating function, probability mass function, mean, and variance are derived. The analysis allows illustration of a variety of techniques useful in the study of discrete distributions appropriate for second-level probability courses. An example on customer brand switching discussed by Olkin, Glesser, and Derman is presented as illustration.

KEY WORDS: Central limit theorem; Difference equation; Markov chain; Probability generating function; Probability mass function; Recursion.

Román Viveros is Associate Professor, Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada L8S 4K1. K. Balasubramanian is Associate Professor, Indian Statistical Institute, New Delhi, India 110 016. N. Balakrishnan is Associate Professor, Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada L8S 4K1.

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

1. INTRODUCTION

A daily customer has to choose between two brands, A and B, of a given product. A model for representing the customer switching patterns discussed by Olkin, Glesser, and Derman (1980, pp. 459–460) assumes that only the customer's choice of brand on the immediately preceding day, day i, affects his or her choice on day $i + 1$. Suppose that when the customer chooses Brand A on one day, then he or she will choose Brand A the next day with probability .8; when the customer chooses Brand B on one day, then he or she will choose Brand A the next day with probability .3. The customer chooses a brand randomly the first day he or she buys the product. The following are questions of concern to the product makers.

Q_1: How likely is that the customer will choose Brand A on 20 days out of the first month (30 days) he or she buys the product?

Q_2: How likely is it that the twelfth time the customer chooses a Brand A product occurs in 15 days?

If the probability of choosing each brand was constant from day to day, then any student with an elementary course on probability could provide an answer to the aforementioned questions by invoking the binomial and negative binomial distributions because the chain of brand choices across the various days will simply be a sequence of independent Bernoulli trials. In the present application, however, the trials are not independent, and the answers to Q_1 and Q_2 demand careful examination.

The problem just described is an example of Markov-correlated Bernoulli trials. These correlated trials find application in many applied disciplines. For instance, Estes (1950) and others since then have used binary Markov models to represent learning (see Olkin et al. 1980, pp. 449–450). An application to binary signal transmission in communications was discussed by Pfeiffer and Schum (1973, p. 358). Kemeny and Snell (1960, p. 31) discussed a model for representing gambler's behavior in choosing between two slot machines. An application to model the starting-up reliability of power-generation equipment operated on gas (e.g., lawnmowers) is presented in Viveros and Balakrishnan (1993). Dry–rainy weather patterns (Ross 1993, p. 138) have also been modeled with Markov-correlated Bernoulli trials. An interesting application to modeling vowel–consonant patterns in biblical text, and more generally in languages, was made by Newman (1951). Among the eight languages investigated, Newman (1951) found that Samoan and Lifu (the largest of the Loyalty Islands, a group located in the south Pacific Ocean east of New Caledonia) are the only languages whose vowel–consonant patterns appear to follow first-order Markov chains. King James English has patterns that appear to follow a third-order chain.

To examine this class of problems more generally, let p_0, p_1, and p_2 be $p_0 = \Pr(O_1 = S)$, $p_1 = \Pr(O_i = S \mid O_{i-1} = S)$, and $p_2 = \Pr(O_i = S \mid O_{i-1} = F)$, where O_i is the outcome of trial i $(i \geq 1)$, and S and F denote *success* and *failure*, respectively. Let $q_j = 1 - p_j$ $(j = 0, 1, 2)$.

This simple correlation structure is tantamount to a stationary two-state Markov chain with state space $\{S, F\}$ and transition matrix with rows (p_1, q_1) and (p_2, q_2). Therefore, any situation giving rise to such Markov chains will provide an application for the results discussed in this article (e.g., see Kemeny and Snell 1960).

This article centers merely on probabilistic considerations; we do not discuss statistical issues such as the estimation of p_0, p_1, and p_2 and the testing of the Markovian assumption. We refer the interested reader to the work of Bishop, Fienberg, and Holland (1975, chap. 7) and Cox and Snell (1989, pp. 98–105) for the statistical analysis of binary Markov chains and time series.

The focus of this article is on the distribution of the random variables: X_n denotes the number of successes in the first n trials; N_x denotes the number of trials needed to get x successes. Note that n is fixed in X_n and x is fixed in N_x. Clearly, X_n and N_x will have the familiar binomial and negative binomial distributions when $p_0 = p_1 = p_2$, that is, when the Bernoulli trials are independent. To the best of our knowledge, the elementary treatment of the extensions to correlated trials discussed in this article do not appear to have been hitherto presented. As will be seen, the richness of the analysis provides an opportunity for the classroom teacher to illustrate a variety of techniques useful in the study of discrete distributions.

In contrast to the corresponding results for independent trials, we found that the distribution of N_x is simpler to describe than that of X_n. This explains the order of presentation we followed.

2. NEGATIVE BINOMIAL

Let $\varphi_x(t)$ denote the probability generating function (PGF) of N_x, namely, $\varphi_x(t) = E(t^{N_x})$. It is obvious that

$$N_x = Y_1 + Y_2 + \cdots + Y_x, \tag{1}$$

where Y_i is the number of Bernoulli trials elapsed since the $(i-1)$th success up to and including the ith success, $1 \leq i \leq x$. Because of the Markovian property, $Y_1, Y_2, \ldots, Y_x$ are independent. Furthermore, $Y_2, Y_3, \ldots, Y_x$ are identically distributed with probability mass function (PMF) $f_B(1) = p_1$, $f_B(y) = q_1 q_2^{y-2} p_2$ for $y \geq 2$, whereas the PMF of Y_1 is of the same form but with (p_1, q_1) replaced with (p_0, q_0). As a result, each of $Y_2, Y_3, \ldots, Y_x$ has PGF $\varphi_B(t) = p_1 t + q_1 p_2 t^2/(1 - q_2 t)$. Similarly, the PGF of Y_1 is $\varphi_A(t) = p_0 t + q_0 p_2 t^2/(1 - q_2 t)$. These results, when used in conjunction with the multiplicative property of the PGF over sums of independent variates, yield

$$\varphi_x(t) = \left[p_0 t + \frac{q_0 p_2 t^2}{1 - q_2 t}\right]\left[p_1 t + \frac{q_1 p_2 t^2}{1 - q_2 t}\right]^{x-1}. \tag{2}$$

An alternative derivation of (2) is possible by working directly with the contributions to the PGF of all possible sequences of outcomes ending in a success and containing exactly x successes.

When expanded in powers of t, the coefficients of (2) give the PMF of N_x. A convenient way of obtaining such coefficients is by first writing

$$\begin{aligned}\varphi_x(t) = q_1^{-1} q_2^{-x} t^x \{ & q_2(p_0 - p_1)[q_1 p_2 (1 - q_2 t)^{-1} \\ & + (p_1 - p_2)]^{x-1} + q_0[q_1 p_2 (1 - q_2 t)^{-1} \\ & + (p_1 - p_2)]^x\},\end{aligned}$$

and then applying the binomial theorem in conjunction with the negative binomial series (e.g., see Kalbfleisch 1985, p. 24) $(1 - q_2 t)^{-i} = \sum_{j=0}^{\infty} \binom{i+j-1}{j} (q_2 t)^j$. The final result is

$$\begin{aligned} f_{N_x}(n) = {} & q_1^{-1}(p_0 - p_1) q_2^{n-2x+1} \\ & \times \sum_{i=0}^{x-1} \binom{x-1}{i} \binom{i+n-x-1}{n-x} \\ & \times (q_1 p_2)^i (p_1 - p_2)^{x-i-1} \\ & + q_1^{-1} q_0 q_2^{n-2x} \sum_{i=0}^{x} \binom{x}{i} \binom{i+n-x-1}{n-x} \\ & \times (q_1 p_2)^i (p_1 - p_2)^{x-i}, \end{aligned} \tag{3}$$

for $n \geq x$. It is easily verified that Equation (3) reduces to the familiar negative binomial PMF,

$$f_{N_x}(n) = \binom{n-1}{n-x} p^x q^{n-x}, \qquad n \geq x, \tag{4}$$

when $p_0 = p_1 = p_2 = p$.

A referee has correctly pointed out that explicit calculation of $f_{N_x}(n)$, n small, can become a cumbersome task when direct use of (3) is attempted. An alternative simpler version of (3) is obtained by noticing that

$$\begin{aligned} & \sum_{i=0}^{x-1} \binom{x-1}{i} \binom{i+n-x-1}{n-x} (q_1 p_2)^i (p_1 - p_2)^{x-i-1} \\ & = \frac{1}{(n-x)!} \left\{ \frac{d^{n-x}[u^{n-x-1}(p_1 - p_2 + q_1 p_2 u)^{x-1}]}{du^{n-x}} \right\}_{u=1}. \end{aligned}$$

A similar expression holds for the second sum in (3). Use of the Lagrange formula for high-order derivatives of products yields

$$f_{N_x}(n) = (p_0 - p_1)p_2 q_2^{n-2x+1} \sum_{j=1}^{(x-1)\wedge(n-x)} \binom{n-x-1}{j-1} \times \binom{x-1}{j} (p_1 q_2)^{x-j-1} (q_1 p_2)^{j-1} + q_0 p_2 q_2^{n-2x} \sum_{j=1}^{x\wedge(n-x)} \binom{n-x-1}{j-1}\binom{x}{j} \times (p_1 q_2)^{x-j} (q_1 p_2)^{j-1}, \quad n \geq x+1, \tag{5}$$

where $a \wedge b = \min\{a, b\}$. When n is small, (5) requires only a few terms for the explicit calculation of $f_{N_x}(n)$; these expressions may also be readily verified by direct probability arguments.

It is easily seen from (2) that $\varphi_x(t) = \varphi_{x-1}(t)[p_1 t + (p_2 - p_1)t^2]/(1 - q_2 t)$. This relationship can be rewritten as

$$(1 - q_2 t)\varphi_x(t) = [p_1 t + (p_2 - p_1)t^2]\varphi_{x-1}(t). \tag{6}$$

Equating coefficients of t^n on both sides of (6) yields the recursion

$$f_{N_x}(n) = q_2 f_{N_x}(n-1) + p_1 f_{N_{x-1}}(n-1) + (p_2 - p_1) f_{N_{x-1}}(n-2), \tag{7}$$

for $n > x$. Thus (7) provides an alternative method for computing probabilities associated with N_x in a simple recursive manner. Note that $f_{N_x}(x) = p_0 p_1^{x-1}$.

The elementary method for the derivation of geometric moments suggested by Samaniego (1992) can be adopted here to obtain the mean and variance of each Y_i in (1). In fact, each Y_i $(2 \leq i \leq x)$ can be shown to have mean and variance $\mu_B = 1 + q_1/p_2$ and $\sigma_B^2 = (q_1/p_2) + q_1(2q_2 - q_1)/p_2^2$, respectively. Similarly, the mean and variance of Y_1, μ_A, and σ_A^2 can be obtained from μ_B and σ_B^2 by replacing q_1 with q_0. Thus the mean and variance of N_x are derived from (1) to be $\mu_{N_x} = \mu_A + (x-1)\mu_B$ and $\sigma_{N_x}^2 = \sigma_A^2 + (x-1)\sigma_B^2$ and are given by

$$\mu_{N_x} = x + [q_0 + (x-1)q_1]/p_2,$$
$$\sigma_{N_x}^2 = \{p_2[q_0 + (x-1)q_1] + q_0(2q_2 - q_0) + (x-1)q_1 \times (2q_2 - q_1)\}/p_2^2. \tag{8}$$

Alternatively, μ_{N_x} and $\sigma_{N_x}^2$ can be obtained from (2) by differentiation.

A normal approximation to the PMF of N_x can be derived as follows. Define $S_x = Y_2 + Y_3 + \cdots + Y_x$, and note that

$$(N_x - \mu_{N_x})/\sigma_{N_x} = Z_1 \left[1 + \frac{(x-1)\sigma_B^2}{\sigma_A^2}\right]^{-1/2} + Z_x \left[1 + \frac{\sigma_A^2}{(x-1)\sigma_B^2}\right]^{-1/2}, \tag{9}$$

where $Z_1 = (Y_1 - \mu_A)/\sigma_A$ and $Z_x = [S_x - (x-1)\mu_B]/(x-1)^{1/2}\sigma_B$. By central limit theorem, $Z_x \approx N(0, 1)$ for x large. Note that the coefficient of Z_1 in (9) converges to 0, whereas that of Z_x converges to 1, as $x \to \infty$. Because Z_1 has finite variance, then

$$(N_x - \mu_{N_x})/\sigma_{N_x} \approx N(0, 1), \tag{10}$$

for x large. As usual, a continuity correction may enhance the quality of the normal approximation.

3. BINOMIAL

A convenient way of connecting X_n and N_x is by means of the fundamental identity

$$\Pr(X_n < x) = \Pr(N_x > n). \tag{11}$$

In the independent case, (11) is usually used to compute probabilities about N_x from those about X_n. Here the roles are reversed. The PMF of X_n can be calculated from (11) as

$$f_{X_n}(x) = \Pr(N_x \leq n) - \Pr(N_{x+1} \leq n), \tag{12}$$

for $0 \leq x \leq n$. Summing (7) over the obvious range yields a corresponding recursion for the cumulative distribution function of N_x. Substituting this recursion in (12) gives

$$f_{X_n}(x) = q_2 f_{X_{n-1}}(x) + p_1 f_{X_{n-1}}(x-1) + (p_2 - p_1) f_{X_{n-2}}(x-1) \tag{13}$$

for $0 < x < n$.

Perhaps the most convenient way of computing $f_{X_n}(x)$ is by using the identity

$$f_{X_n}(x) = f_{N_x}(n) + f_{M_{n-x}}(n), \tag{14}$$

where M_y denotes the number of trials needed to get y failures. The PMF of M_y can be obtained from (3) or (5) simply by taking $x = y$ and reversing the roles of p_0 and q_0, p_1 and q_2, and p_2 and q_1. Equation (14) finds justification in the fact that the occurrence of x successes in the first n trials can take place by observing either of the disjoint events that the xth success comes on trial n or the $(n-x)$th failure comes on trial n.

Because the Markov-dependent model has three probability parameters, the range of shapes exhibited by $f_{X_n}(x)$ is considerably larger compared with that for the standard binomial. In particular, bimodal and trimodal shapes are possible. This fact widens the scope for data fitting.

Let $\varphi_n(t)$ denote the PGF of X_n, $\varphi_n(t) = E(t^{X_n})$. The derivation of $\varphi_n(t)$ illustrates another important tool in the study of discrete distributions, namely the solution of difference equations. It is convenient to decompose $\varphi_n(t)$ as

$$\varphi_n(t) = \varphi_n^{(S)}(t) + \varphi_n^{(F)}(t), \tag{15}$$

where $\varphi_n^{(S)}(t)$ and $\varphi_n^{(F)}(t)$ are the net contributions to the PGF from n sequences of trials ending in a success and in a failure, respectively. By conditioning on the outcome of the nth trial, one immediately obtains

$$\varphi_{n+1}^{(S)}(t) = p_1 t \varphi_n^{(S)}(t) + p_2 t \varphi_n^{(F)}(t),$$
$$\varphi_{n+1}^{(F)}(t) = q_1 \varphi_n^{(S)}(t) + q_2 \varphi_n^{(F)}(t). \tag{16}$$

A straightforward manipulation of (16) yields the difference equation

$$\varphi_{n+2}^{(S)}(t) - (p_1 t + q_2)\varphi_{n+1}^{(S)}(t) - (p_2 - p_1)t\varphi_n^{(S)}(t) = 0. \tag{17}$$

The general solution of (17) takes the form

$$\varphi_n^{(S)}(t) = A(t)[\alpha(t)]^n + B(t)[\beta(t)]^n, \tag{18}$$

where $\alpha(t)$ and $\beta(t)$ are the roots of the quadratic polynomial $Q_t(z) = z^2 - (p_1 t + q_2)z - (p_2 - p_1)t$, and $A(t)$ and $B(t)$ are arbitrary coefficients (e.g., see Johnson and

Kotz 1977, pp. 13–14). The quadratic equation $Q_t(z) = 0$ has solutions

$$\alpha(t) = \theta_1 + \theta_2, \quad \beta(t) = \theta_1 - \theta_2;$$
$$\theta_1 = (p_1 t + q_2)/2, \; \theta_2 = [(p_1 t - q_2)^2 + 4q_1 p_2 t]^{1/2}/2. \tag{19}$$

Noting that $\varphi_1^{(S)}(t) = p_0 t$, $\varphi_2^{(S)}(t) = p_0 p_1 t^2 + q_0 p_2 t$, the particular coefficients $A(t)$ and $B(t)$ of interest are calculated as the solution of the 2×2 system of linear equations obtained from (18) for $n = 1, 2$ after substituting for $\alpha(t)$ and $\beta(t)$ from (19).

These results can be combined to yield an expression for $\varphi_n(t)$. First, note that by using (15) and (16), one can write $\varphi_n(t)$ in terms of $\varphi_{n+1}^{(S)}(t)$ and $\varphi_n^{(S)}(t)$. Then (18) gives

$$p_2\varphi_n(t) = [(p_2 - p_1)t + \alpha(t)]A^*(t)[\alpha(t)]^{n-1} + [(p_2 - p_1)t + \beta(t)]B^*(t)[\beta(t)]^{n-1}, \tag{20}$$

where $A^*(t) = [p_0 p_1 t + q_0 p_2 - p_0\beta(t)]/[\alpha(t) - \beta(t)]$ and $B^*(t) = [p_0 p_1 t + q_0 p_2 - p_0\alpha(t)]/[\beta(t) - \alpha(t)]$.

A careful examination of (20) reveals that as expected, $\varphi_n(t)$ is a polynomial of degree n in t. Finally, evaluating $\varphi_n'(t)$ at $t = 1$ gives the mean of X_n as

$$\mu_{X_n} = [np_2 + p_0 - p_2/a + (a-1) \times (p_0 q_1 - q_0 p_2)(p_1 - p_2)^{n-1}/a]/a, \tag{21}$$

where $a = q_1 + p_2$. Note that (21) reduces to the familiar result $\mu_{X_n} = np$ when $p_0 = p_1 = p_2 = p$.

4. EXAMPLE: BRAND SWITCHING

Consider the customer brand-switching model described in the introduction. From the information provided, $p_0 = .5$, $p_1 = .8$, and $p_2 = .3$, where success and failure correspond to *choosing Brand A* and *choosing Brand B*, respectively.

The answer to Q_2 requires the calculation of probabilities about N_{12}. Some of these probabilities, which were computed by using (3), are reported in Table 1.

The answer to Q_2 is $f_{N_{12}}(15)$ or 6.4%. In addition, the chances that the customer opts for Brand A at least a dozen times during the first three weeks he or she buys the product is $\Pr(N_{12} \leq 21)$ or 60%, whereas the chances that the customer opts for Brand A at least a dozen times during the first month is $\Pr(N_{12} \leq 30)$ or 91%. The mean and standard deviation of N_{12}, as calculated from (8), are 21 and 6.6 days, respectively.

In contrast, if the customer made his or her selections randomly, from the ordinary negative binomial distribution with $p = .5$ one would obtain $f_{N_{12}}(15) = 0.011$, $\Pr(N_{12} \leq 21) = .33$, $\Pr(N_{12} \leq 30) = .9$, with mean and standard deviation for N_{12} at 24 and 4.9 days, respectively. Similarly, if the customer opted for Brand A, 80% of the time regardless of his or her previous selections, again from the ordinary negative binomial distribution with $p = .8$, one would obtain $f_{N_{12}}(15) = 0.2$, $\Pr(N_{12} \leq 21) = .996$, $\Pr(N_{12} \leq 30) = 1$, with respective mean and standard deviation for N_{12} of 15 and 1.9 days.

Table 1. PMF of N_{12} for the Brand Switching Model

n	$f_{N_{12}}(n)$	n	$f_{N_{12}}(n)$	n	$f_{N_{12}}(n)$	n	$f_{N_{12}}(n)$
12	.043	17	.069	22	.053	27	.029
13	.048	18	.068	23	.048	28	.025
14	.058	19	.066	24	.043	29	.021
15	.064	20	.062	25	.038	30	.018
16	.068	21	.058	26	.033	≥ 31	.064

Table 2. PMF of X_{30} for the Brand Switching Model

x	$f_{X_{30}}(x)$	x	$f_{X_{30}}(x)$	x	$f_{X_{30}}(x)$	x	$f_{X_{30}}(x)$
≤ 6	.008	11	.029	16	.076	21	.072
7	.006	12	.039	17	.082	22	.062
8	.010	13	.049	18	.085	23	.051
9	.015	14	.059	19	.084	24	.039
10	.021	15	.068	20	.079	≥ 25	.068

Consider Q_1. The PMF of X_{30} has been computed through (14) and is reported in Table 2. The answer to Q_1 is $f_{X_{30}}(20)$ or 7.9%. In addition, the chances that the customer opts for Product A at least half of the time during the first month he or she buys the product is $\Pr(X_{30} \geq 15)$ or 76.4%. From (21), the expected number of days in which the customer chooses Product A out of the first month is $\mu_{X_{30}} = 17.8$ days.

By comparison, if the customer made his or her selections randomly, from the familiar binomial distribution with $p = .5$ one would obtain $f_{X_{30}}(20) = 0.028$, $\Pr(X_{30} \geq 15) = .572$, and $\mu_{X_{30}} = 15$ days. Similarly, when the customer chooses Product A with probability $p = .8$ regardless of previous selections, the ordinary binomial model gives $f_{X_{30}}(20) = 0.035$, $\Pr(X_{30} \geq 15) = 1$, and $\mu_{X_{30}} = 24$ days.

5. DISCUSSION

The theory of Markov-dependent Bernoulli trials is fascinating and rich in many ways. From the teaching–learning process point of view, the benefit is twofold: the students have the opportunity to understand the concepts and appreciate the consequences of a simple dependence structure, and the instructor is able to introduce in one model several techniques useful in the study of discrete distributions. From the applications point of view, as indicated in the introduction, the scope is fairly wide.

Although the focus of this article has been the binomial and negative binomial analogues under the proposed Markov-correlated Bernoulli trials, many other interesting problems arise under such dependence model. For instance, Viveros and Balakrishnan (1993) and Balakrishnan, Balasubramanian, and Viveros (1994) have extended the start-up demonstration test model of Hahn and Gage (1983) to the practical situation in which the trials are Markov dependent. Here the emphasis is on runs of successes and failures. More interesting and somewhat involved extensions to Markov dependence models of various sampling procedures (e.g., frequency and run quota sampling) have been examined by Balasubramanian, Viveros, and Balakrishnan (1993).

REFERENCES

Balakrishnan, N., Balasubramanian, K., and Viveros, R. (1994), "Probability Models and Inference for Start-Up Demonstration Tests Under Correlation and Corrective Action," in *Naval Research Logistics*.

Balasubramanian, K., Viveros, R., and Balakrishnan, N. (1993), "Sooner and Later Waiting Time Problems for Markovian Bernoulli Trials," *Statistics & Probability Letters*, 18, 153–161.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975), *Discrete Multivariate Analysis*, Cambridge: MIT Press.

Cox, D. R. and Snell, E. J. (1989), *Analysis of Binary Data*, London: Chapman & Hall.

Estes, W. K. (1950), "Toward a Statistical Theory of Learning," *Psychological Review*, 57, 94–107.

Hahn, G. J., and Gage, J. B. (1983), "Evaluation of a Start-Up Demonstration Test," *Journal of Quality Technology*, 15, 103–105.

Johnson, N. L., and Kotz, S. (1977), *Urn Models and Their Application*, New York: John Wiley.

Kalbfleisch, J. G. (1985), *Probability and Statistical Inference*, Vol. 1: Probability, New York: Springer-Verlag.

Kemeny, J. G., and Snell, J. L. (1960), *Finite Markov Chains*, Princeton: D. Van Nostrand Co.

Newman, E. B. (1951), "The Pattern of Vowels and Consonants in Various Languages," *American Journal of Psychology*, 64, 369–379.

Olkin, I., Glesser, L. J., and Derman, C. (1980), *Probability Models and Applications*, New York: Macmillan.

Pfeiffer, P. E., and Schum, D. A. (1973), *Introduction to Applied Probability*, New York: Academic Press.

Ross, S. M. (1993), *Introduction to Probability Models* (5th ed.), New York: Academic Press.

Samaniego, F. J. (1992), "Elementary Derivations of Geometric Moments," *The American Statistician*, 46, 108–109.

Viveros, R., and Balakrishnan, N. (1993), "Statistical Inference From Start-Up Demonstration Test Data," *Journal of Quality Technology*, 25, 119–130.

Accent on Teaching Materials

Harry O. Posten, Section Editor

In this section *The American Statistician* publishes announcements and selected reviews of teaching materials of general use to the statistical field. These may include (but will not necessarily be restricted to) curriculum material, collections of teaching examples or case studies, modular instructional material, transparency sets, films, filmstrips, videotapes, probability devices, audiotapes, slides, and data deck sets (with complete documentation).

Authors, producers, or distributors wishing to have such materials announced or reviewed should submit a single, complete copy of the product (three copies of printed material double-spaced) to Section Editor Harry O. Posten, Statistics Department, University of Connecticut, Storrs, CT 06268. A statement of intention that the material will be available to all requesters for a minimum of a two-year period should be provided, along with information on the cost (including postage) and special features of the material. Information on classroom experience may also be included. All materials submitted must be of general use for teaching purposes in the area of probability and statistics.

Review of *Teaching Statistical Concepts*

This book by Anne Hawkins, Flavia Jolliffe, and Leslie Glickman provides an overview of research in teaching introductory probability and statistics. It is published in London by Longman Group UK Limited (Longman House, Burntmill, Harlow Essex, CA20 2JE) as part of the Effective Teacher Series. Cost: £11.99.

Probability and statistics are useful topics for all students to learn but are difficult to teach. If it were obvious how to explain things like conditional probability and standard errors, then we would all be good statistics teachers, and ours would be among the favorite subjects of most students. This is not the case, as is well known. There is a lot we do not know about how to teach probability and statistics, but many people are doing research in the area, and progress is being made. This book provides an overview of the research, along with general comments and suggestions for improving statistics teaching, particularly at the introductory level.

If you think of yourself as an innovative statistics teacher who keeps a close eye on the statistics education movement, then this book was written for you. By reading it you will find out what research has been done and be reminded of the key issues that call for future research. If you do not put much stock in the scholarly work done in the area of statistics education, then you may safely forego reading this book—but then you probably would not be reading this review in the first place.

Much of the discussion in the book deals with the United Kingdom (reading Chap. 7, "Statistics Practicals and Projects," I was reminded of the quote by Shaw that England and America are two countries separated by the same language), but Hawkins, Jolliffe, and Glickman cite research done around the world and display knowledge of a surprisingly broad literature on research in statistics education. I found it refreshing to hear about statistics education at the school level in England and to contrast it, in my mind, to the situation in the United States, where we have tended to postpone the introduction of probability and statistics until college.

The book is more of an overview of general issues than it is a compilation of anecdotes and teaching ideas. Several solid teaching ideas are mentioned, but they are not the main focus of the book. Rather, Hawkins et al. provide a detailed resource to turn to when you hear an anecdote about teaching statistics and ask yourself "But has anyone done serious, scholarly research on this?" If your goal is to pick up a collection of ideas to use in the classroom, you will be disappointed by the amount of reading you have to do to get to the next novel classroom tip. Hawkins et al. are serious about their task and expect the reader to be diligent as well. This is not a book for someone who has only a passing interest in statistics education; this book is for people who take statistics education seriously and are systematic in searching for ways to improve their teaching. Hawkins et al. do not take a chatty, informal approach that says "Here are some neat ideas; give them a try." They approach their subject as scholars and experimenters who address difficult questions, eschew simplistic answers, and seek solid evidence to support hypotheses regarding the effectiveness of teaching methods.

The book begins with a brief history of statistics teaching, noting that mathematics has overshadowed the applied nature of the development of the field and that computers are now changing what is taught, as well as how it is taught. Computers have made real data analysis possible in the schools, but few school-level teachers have training in statistics. Thus, there is much to be done if we are to improve the general level of statistical thinking among students at all levels.

The chapters on teaching descriptive statistics, probability, and inference contain a set of recommendations that we have heard be-