# A BINOMIAL DISTRIBUTION WITH DEPENDENT TRIALS AND ITS USE IN STOCHASTIC MODEL EVALUATION

*Evdokia Xekalaki and John Panaretos*
*Department of Statistics,*
*Athens University of Economics and Business,*
*Athens, Greece*
*June 21,2003*

## 1. INTRODUCTION

An interesting problem in statistics is that of model selection. This amounts to the evaluation of a set of competing functional representations of the mechanism that produces the data and the selection of one of them on the basis of a preference ordering of the models and a "measurable difference" among them.

The existing evaluation and selection criteria have been developed on the basis of at least one of the following four types of inductive inference: Likelihood, entropy, Bayesian and

prediction. Predictive approaches are based on the predictive ability of the models and are typically cross- validatory. A good reference for such criteria is Lahiri (1992).

One possible strategy for constructing model evaluation criteria is to adopt an approach of sequential nature based on the predictions produced by the model. Such a strategy was suggested by Xekalaki and Katti (1984) who put forward various alternative schemes of this nature. Xekalaki et al. (2003), along the above lines of thinking, introduced a model evaluation scheme that utilizes the standardized prediction errors as scoring rules.

In this paper we concentrate our attention to an alternative evaluation scheme, that utilizes again the model based prediction. In this scheme (proposed originally by Xekalaki and Katti (1984)), the forecasting potential of a model is measured based on a predictive approach of a non cross-validatory nature. This evaluation scheme consists of the sequential construction of an interval centered at the model's prediction with length that is increased or decreased depending on the degree of concordance between observed and predicted values of' the dependent variable. The evaluation of the model and the selection of a model among several candidate models is effected through the use of a scoring rule.

In order to be able to apply the suggested methodology in model evaluation, we need to study its theoretical foundation. To do this, we develop some distribution theory that leads to a new binomial distribution with dependent trials. Properties of this distribution are studied and used to formulate the theoretical basis for making inference on the forecasting behaviour of a model by exploiting the sequential nature of the model-based predictions.

In particular, section 2 provides the necessary background on Xekalaki and Katti's (1984) method. The statistical behaviour of one of the scoring rules suggested by them is studied leading to a new binomial distribution with dependent trials whose properties are discussed in section 3. Sections 4 and 5 concentrate on a special case of the model and suggest its use for constructing confidence intervals or testing hypotheses concerning an appropriately chosen parameter that would reflect the forecasting potential of the model in question. Large sample inference is also made.

## 2. DESIGN OF THE EVALUATION SCHEME

Most often, a regression model describing the relationship between a set of predictor variables $X_1, X_2, ..., X_m$ and one response Y is or the linear form

$$E(Y \mid X_1 = x_1, ..., X_m = x_m) = b_0 + b_1 x_1 + ... + b_m x_m \tag{2.1}$$

where the b's are coefficients to be estimated from the data. Let $X_t$ be the $l_t \times m$ matrix of predictor variables associated with t observations at time t ($l_t \geq m$, $|X_t'X_t| \neq 0$), $Y_t$ the corresponding $l_t \times 1$ vector of observations on the dependent random variable (r.v.) Y and let b be the $m \times 1$ vector of regression coefficents. Then the model in (2.1) can alternatively be represented by

$$Y_t = X_t b + \varepsilon_t \tag{2.2}$$

where $\varepsilon_t$ is an $l_t \times 1$ vector of normal error random variables with mean $E(\varepsilon_t)=0$ and dispession matrix $V(\varepsilon_t)=\sigma^2 I_t$. Here $I_t$ is an $l_t \times l_t$ indentity matrix.

Of course, the methodology to be studied is of general nature, it does not depend on the functional form of $E(Y \mid X_1 = x_1, ..., X_m = x_m)$ and hence it can be applied to any type of model.

Using (2.2), one may predict the value of Y at time t+1 to be $\hat{Y}^0_{t+1} = X^0_{t+1}\hat{b}_t$ where $\hat{b}_t$ is the least squares estimator of b at time t+1 given by

$$\hat{b}_t = (X_t'X_t)^{-1} X_t'Y_t \tag{2.3}$$

and $X^0_{t+1}$ is a $(1 \times m)$ vector of values of the regressors at time t+1. The value of $V(\hat{Y}^0_{t+1})$ is then obtained by

$$V(\hat{Y}^0_{t+1}) = \sigma^2 \{X^0_{t+1}(X_t'X_t)^{-1} X^{0'}_{t+1} + 1\}$$

and is estimated by replacing $\sigma^2$ by

$$S_t^2 = (Y_t - X_t\hat{b}_t)'(Y_t - X_t\hat{b}_t)/(l_t - m). \tag{2.4}$$

After the actual value $Y^0_{t+1}$ of Y at time t+1 has been observed, the model for predicting the value $Y^0_{t+2}$ of Y at time t+2 can be represented by

$$Y_{t+1} = X_{t+1} b + \varepsilon_{t+1}$$

where the matrices $X_{t+1}$ and $Y_{t+1}$ are defined by

$$X_{t+1} = \begin{bmatrix} X_t \\ X^0_{t+1} \end{bmatrix} \text{ and } Y_{t+1} = \begin{bmatrix} Y_t \\ Y^0_{t+1} \end{bmatrix}$$

with dimensions $(l_t + l) \times m$ and $(l_t + 1) \times 1$, respectively.

Then, obviously, at time t+2 the vector b will be estimated by $\hat{b}_{t+2}$ obtained by (2.3) substituting t+1 for t.

Of course, incorporating the observed value $Y^0_{t+1}$ in the data set and reestimating the regression coefficients of the model so that it be ready for the next prediction on the value of

Y, pressuposes that the statistical behaviour of the model in the future will be similar to its statistical behaviour in the past. On this assumption, the evaluation scheme proposed by Xekalaki and Katti (1984) is an n-stage sequential technique that is described by the following steps:

1. At time t+1 obtain $\hat{Y}^0_{t+1}$ using (2.2).
2. Construct the interval

$$C_{t+1} = [\hat{Y}^0_{t+1} - k_t S_t, \ \hat{Y}^0_{t+1} + k_t S_t]$$

where $S_t$ is as given by (2.3) and $k_t$ a positive constant whose initial value is set by the experimenter.

3. Observe the actual value $Y^0_{t+1}$ of Y at time t+l.
4. Choose a scoring rule to assign a score to each of the two complementary outcomes

$$I_t = \{Y^0_{t+1} \in C_{t+1}\} \text{ and } O_t = \{Y^0_{t+1} \notin C_{t+1}\}$$

5. Incorporate $X^0_{t+1}$ and $Y^0_{t+1}$ to the data set and re-estimate the regression coefficents.
6. Construct $C_{t+2}$ as in step 2 using the rule

$$k_{t+1} = \begin{cases} (1-\alpha_{t+1})\, k_t & \text{if } Y^0_{t+1} \in C_{t+1} \\ (1+\gamma_{t+1})\, k_t & \text{if } Y^0_{t+1} \notin C_{t+1} \end{cases}$$

where $\alpha_t$, $\gamma_t$ are non-negative quantities less than 1 defined by the experimenter. These may well be functions of the frequencies of the events $I_t$ and $O_t$, respectively.

The process is repeated for as many times as the number of times the model was applied, say n. As a final rating reflecting the forecasting potential of the model Xekalaki and Katti (1984) suggested the average of the scores from step 4.

Such scores can be obtained through a scoring rule the choice of which is a matter of the experimenter's personal judgement. Among the rules suggested by Xekalaki and Katti (1984), the simplest possible, amounts to assigning a score

$$Z_{t+1} = \begin{cases} 1 & \text{if } I_t \text{ is observed} \\ 0 & \text{if } O_t \text{ is observed} \end{cases} \tag{2.5}$$

leading to $S_n/n$ as the final rating of the model where

$$S_n = \sum_{i=1}^{n} Z_i \ .$$

In the sequel, attention is given to this rule and an attempt is made to develop some theory that will enable, us to obtain insight as to the statistical significance of an observed

value of $S_n/n$ as well as of information provided by $S_n$ -based confidence intervals or tests of hypothesis for appropriate parameters as to the merit of the model in question.

## 3. THE PROBABILITY DISTRIBUTION OF THE SCORES

For a series of n points in time, consider the sequence of pairs $\left\{ \left[\hat{Y}_i^0, Y_i^0\right]: i = 1, 2, ..., n \right\}$. Each pair $\left[\hat{Y}_i^0, Y_i^0\right]$ can be regarded as a trial whose outcome can be designated as

$$Z_i = \begin{cases} 1 \ \text{(success)} & \text{if} \quad Y_i^0 \in C_i \\ 0 \ \text{(failure)} & \text{if} \quad Y_i^0 \notin C_i \end{cases} \tag{3.1}$$

i = l, 2, ..., n. Hence $\{Z_1, Z_2, \ldots, Z_n\}$ is a sequence of Bernoulli variables with

$$p_i = P(Z_i = 1) = 1 - P(Z_i = 0), \qquad i = 1, 2, ..., n. \tag{3.2}$$

It is obvious from the evaluation scheme that the probability of the outcome of a trial depends on the outcome of the previous trial. Let

$$\begin{aligned} p_{11}^{(i)} &= P\left[Z_i = 1 \mid Z_{i-1} = 1\right] = 1 - p_{10}^{(i)} \\ p_{01}^{(i)} &= P\left[Z_i = 1 \mid Z_{i-1} = 0\right] = 1 - p_{00}^{(i)} \end{aligned} \tag{3.3}$$

$$\pi_{i-1,i} = \begin{bmatrix} p_{00}^{(i)} & p_{01}^{(i)} \\ p_{10}^{(i)} & p_{11}^{(i)} \end{bmatrix} \tag{3.4}$$

Therefore, $\{Z_t, t=1, 2, ..., n\}$ defines a non-homogeneous Markov process with first order transition probability matrix given by (3.4).

Before proceeding to the further study of the statistical inference related to this evaluation scheme it is necessary to study the distributional properties of the joint distribution of $Z_1, Z_2, \ldots, Z_n$ as well as of the distribution of $S_n = Z_1 + Z_2 + \ldots + Z_n$. In this respect we first derive two results, which specify the probability $p_i$ of success at the i-th step and the mean and variance of the total number of successes.

*Theorem 3.1.* Consider the sequence of r.v. 's $Z_1, Z_2, \ldots, Z_n$ as defined by (3.1), (3.2) and (3.3). Then, for i = 1, 2, 3, …

$$p_i = p_{01}^{(i)} + \sum_{j=2}^{1} p_{01}^{(j-1)} \prod_{I=j}^{i} \left[p_{11}^{(I)} - p_{01}^{(I)}\right] + p_1 \prod_{I=2}^{i} \left[p_{11}^{(I)} - p_{01}^{(I)}\right] \tag{3.5}$$

where $p_{11}^{(0)} = 0$.

*Proof.* It is obvious that

$$p_i = p_{i-1}\, p_{11}^{(i)} + \left[1 - p_{i-1}\right] p_{01}^{(i)}, \quad i = 2, 3, ..., n$$

or, equivalently, that

$$p_i = p_{01}^{(i)} + \left[p_{11}^{(i)} - p_{01}^{(i)}\right] p_{i-1}, \quad i = 2, 3, ..., n \tag{3.6}$$

Therefore, (3.5) is valid for i=2. Assume that it is true for i=2, 3, ..., n. It will be shown that it is true for i=n+1. Indeed, from (3.5) and the induction hypothesis we have that

$$p_{n+1} = p_{01}^{(n+1)} + \left[p_{11}^{(n+1)} - p_{01}^{(n+1)}\right] p_n$$

$$= p_{01}^{(n+1)} + \left[p_{11}^{(n+1)} - p_{01}^{(n+1)}\right] \times$$

$$\left\{ p_{01}^{(n)} + \sum_{j=2}^{n} p_{01}^{(j-1)} \prod_{i=j}^{n} \left[p_{11}^{(i)} - p_{01}^{(i)}\right] + p_1 \prod_{i=2}^{n} \left[p_{11}^{(i)} - p_{01}^{(i)}\right] \right\}$$

$$= p_{01}^{(n+1)} + p_1 \prod_{i=2}^{n+2} \left[p_{11}^{(i)} - p_{01}^{(i)}\right] + \left\{ \sum_{j=2}^{n} p_{01}^{(j-1)} \prod_{i=j}^{n} \left[p_{11}^{(i)} - p_{01}^{(i)}\right] + p_{01}^{(n)} \left[p_{11}^{(n+1)} - p_{01}^{(n+1)}\right] \right\}$$

$$= p_{01}^{(n+1)} + p_1 \prod_{i=2}^{n+1} \left[p_{11}^{(i)} - p_{01}^{(i)}\right] + \sum_{j=2}^{n+1} p_{01}^{(j-1)} \prod_{i=j}^{n+1} \left[p_{11}^{(i)} - p_{01}^{(i)}\right]$$

Hence the result.

*Theorem* 3.2. Let $Z_1, Z_2, \ldots, Z_n$ be defined as in theorem 3.1 and let $S_n = \sum_{i=1}^{n} Z_i$. Then

(i) $$E(S_n) = \sum_{i=1}^{n} p_i \tag{3.7}$$

(ii) $$V(S_n) = \sum_{i=1}^{n} p_i (1 - p_i) + 2 \sum_{i<j} (W' P_{ij} W - p_i p_j) \tag{3.8}$$

with $p_i$, i=1, 2, ..., n as defined by (3.5), $W' = (0, 1)$ and, for i<j.

$$P_{ij} = \pi_{j-1,j}\ \pi_{j-2,j-1} \cdots\ \pi_{i+1,i+2}\ \pi_{i,i+1}\ \rho_i$$

where $\pi_{j-1,j}$ is given by (3.4) and

$$\rho_i = \begin{bmatrix} 1-p_i & 0 \\ 0 & p_i \end{bmatrix}, \quad i = 1, 2, \ldots, n \tag{3.9}$$

*Proof:* The proof of (i) is straightforward. To prove (ii) observe that

$$V(S_n) = V\left[\sum_{i=1}^{n} Z_i\right] = \sum_{i=1}^{n} V(Z_i) + 2\sum_{i<j} Cov(Z_i, Z_j)$$

But,

$$V(Z_i) = p_i\,(1-p_i) \tag{3.10}$$

and,

$$Cov(Z_i, Z_j) = E(Z_i\, Z_j) - p_i\, p_j \tag{3.11}$$

$$i,j = 1, 2, \ldots, n; \quad i \neq j.$$

Moreover, since $\{Z_t,\ t=1, 2, \ldots,\}$ is a Markov process with first order transition probability matrix given by (3.4) it follows that, for a non-negative integer m

$$\pi_{i,\ i+m} = \pi_{i+m-1,\ i+m}\ \pi_{i+m-2,\ i+m-1} \cdots\ \pi_{i+1,\ i+2}\ \ \pi_{i,\ i+1}\ .$$

Hence, setting j–i =m the probability $P(Z_i = x, Z_j = y)$, x, y =0, 1 is the (x, y) element of the matrix

$$p_{ij} = \pi_{ij}\ \rho_i$$

which implies that

$$E(Z_i Z_j) = (0,1)\ P_{ij} \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Combining (3.10) and (3.11) leads to (3.8) and hence the theorem is established.

In what follows we derive the probability generating function (p.g.f.) of $S_n = Z_1 + Z_2 + \ldots + Z_n$. As an intermediate step we specify the p.g.f. of $(Z_1, Z_2, \ldots, Z_n)$.

*Theorem 3.3.* Let $S_n$ be defined as in theorem 3.2. Then the p.g.f. of $S_n$ is given by

$$G_{S_n}(s) = (1-p_1, p_1 s) \begin{bmatrix} p_{00}^{(2)} & p_{01}^{(2)} s \\ p_{10}^{(2)} & p_{11}^{(2)} s \end{bmatrix} \cdots \begin{bmatrix} p_{00}^{(n)} & p_{01}^{(n)} s \\ p_{10}^{(n)} & p_{11}^{(n)} s \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{3.12}$$

*Proof*: Let $G_W(s)$ denote the p.g.f. of a r.v. W. Then

$$G_{S_n}(s) = G_{Z_1} + \ldots + Z_n(s) = G_{Z_1, Z_2, \ldots, Z_n}(s, s, \ldots, s) \tag{3.13}$$

But from the definition of a p.g.f. we have

$$G_{Z_1,\ldots,Z_n}(s_1,\ldots,s_n) = \sum_{Z_1=0}^{1}\cdots\sum_{Z_n=0}^{1} P(Z_1 = z_1,\ldots, Z_n = z_n)\prod_{i=1}^{n} s_i^{z_i}$$

$$= \sum_{Z_1=0}^{1}\cdots\sum_{Z_n=0}^{1} P(Z_1 = z_1)\, P(Z_2 = z_2 \mid Z_1 = z_1)\ldots P(Z_n = z_n \mid Z_{n-1} = z_{n-1})\prod_{i=1}^{n} s_i^{z_i}$$

The latter equality follows from Markov's property of the process defined by $\{Z_t, t=1, 2, \ldots, n\}$ and implies that

$$G_{Z_1,\ldots,Z_n}(s_1,\ldots,s_n) = (1-p_1, p_1 s_1)\begin{bmatrix} p_{00}^{(2)} & p_{01}^{(2)} s_2 \\ p_{10}^{(2)} & p_{11}^{(2)} s_2 \end{bmatrix}\cdots\begin{bmatrix} p_{00}^{(n)} & p_{01}^{(n)} s_n \\ p_{10}^{(n)} & p_{11}^{(n)} s_n \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad (3.14)$$

Relationship (3.12) follows immediately from the successive application of (3.13) and (3.14). This completes the proof of the theorem.

Theorems 3.1, 3.2 and 3.3 can be used to derive the joint distribution of $(Z_1, Z_2, \ldots, Z_n)$ and the distribution of $S_n$.

*Theorem* 3.4. Let $Z_1, Z_2, \ldots, Z_n$ be defined as in theorem 3.1. Then

$P(Z_1=z_1, Z_2=z_2, \ldots, Z_n=z_n) =$

$$= p_1^{z_1}(1-p_1)^{1-z_1}\prod_{i=2}^{n}\left\{\left[p_{11}^{(i)}\right]^{z_{i-1} z_i}\left[p_{10}^{(i)}\right]^{z_{i-1}(1-z_i)}\left[p_{01}^{(i)}\right]^{(1-z_{i-1}) z_i}\left[p_{00}^{(i)}\right]^{(1-z_{i-1})(1-z_i)}\right\}$$

$$z_i = 0, 1\ ;\ i = 1, 2, \ldots, n \qquad (3.15)$$

*Theorem* 3.5. If $S_n$ is defined as in theorem 3.2 then $P(S_n = s) =$

$$\sum_{\sum Z_i = s} p_1^{z_1}(1-p_1)^{1-z_1}\prod_{i=2}^{n}\left\{\left[p_{11}^{(i)}\right]^{z_{i-1} z_i}\left[p_{10}^{(i)}\right]^{z_{i-1}(1-z_i)}\left[p_{01}^{(i)}\right]^{(1-z_{i-1}) z_i}\times\left[p_{00}^{(i)}\right]^{(1-z_{i-1})(1-z_i)}\right\}$$

$$s = 0, 1, 2, \ldots, n \qquad (3.16)$$

We call the distribution defined by (3.15) Markov-dependent multivariate Bernoulli distribution with variable transition probabilities and the distribution defined by (3.16) Markov-dependent binomial distribution with variable transition probabilities.

The interesting thing about this Markov-dependent binomial distribution is that it is not a mathematical creation but it arises as a consequence of a real problem.

The immediate objective would be to develop tables for the probability distributions defined by (3.15) and (3.16) which can be used for either deriving confidence intervals for p or for testing hypotheses pertaining to p. Such a tabulation would be immense and will be the subject of future research. Another problem is the estimation of the parameters involved.

The special case of this distribution where the transition probabilities are constant, i.e., $p_{11}^{(i)} = p(z_i = 1 | z_{i-1} = 1) = p,$ was considered by Klotz (1973) as a model for rainfall data. Klotz (1973), treated the problem of estimating the parameters of this special case by the method of maximum likelihood while Devore (1976), suggesting a modification on the log likelihood, came up with explicit expressions for the solutions of the resulting likelihood equation namely

$$\hat{\lambda} = n_{11}/(n_{10} + n_{11}) \tag{3.17}$$

$$\hat{p} = n_{01}(n_{10} + n_{11})/\{n_{01}(n_{10} + n_{11}) + n_{10}(n_{00} + n_{01})\} \tag{3.18}$$

where $n_{ij}$ is the observed frequency of (i, j) ; i, j = 0, 1.

Also Klotz (1973) showed that the asymptotic distribution of $(S_n - np)/\sqrt{n}$ is normal with mean 0 and variarice $p(1-p)(1-2p+\lambda)/(1-\lambda)$ and remarked that this variance coincides with the variance of $\hat{p}$ in (3.18).

For this special case, of the Markov-dependent binomial distribution with constant $p_i$ considered by Klotz we can show that the following properties are valid, as immediate consequences of the above theorems.

*Theorem 3.6.* Let $Z_1, Z_2, ..., Z_n$ be a sequence of r.v's with $P(Z_i=l)=l-P(Z_i=0)=p$, i=1, 2, ... and $P(Z_i=1| Z_{i-1}=1)=\lambda \le p$, i= 2, 3, ..., n and let $S_n = \sum_{i=l}^{n} Z_i$ . Then

(i) $E(S_n) = np$

(ii) $$V(S_n) = np(1-p) + \frac{2p(1-p)(\lambda - p)}{1-\lambda}\left\{n - 1 - \frac{\lambda - p}{1-\lambda}\left[1 - \left[\frac{\lambda - p}{1-p}\right]^n\right]\right\} \tag{3.19}$$

*Theorem 3.7.* Let $S_n$ be defined as in theorem 3.6 then

$$G_{S_n}(s) = (1-p, ps)\begin{bmatrix} 1 - \frac{1-\lambda}{1-p} & \frac{1-\lambda}{1-p}ps \\ 1-\lambda & \lambda s \end{bmatrix}^{n-1}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{3.20}$$

*Theorem 3.8.* Let $Z_1, Z_2, ..., Z_n$ and $S_n$ be defined as in theorem 3.6. Then

(i) $P(Z_1 = z_1, ..., Z_n = z_n) =$

(3.21)

$$\frac{(1-\lambda_0)^{n-1}}{1-p}\left\{\frac{\lambda(1-\lambda_0)(1-p)}{(1-\lambda)^2 p}\right\}^{\sum_{i=2}^{n} Z_{i-1}Z_i}\left\{\frac{(1-\lambda)^2 p}{(1-\lambda_0)^2(1-p)}\right\}^{\sum_{i=1}^{n} Z_i}\left\{\frac{1-\lambda_0}{1-\lambda}\right\}^{Z_1+Z_n}$$

(ii) $$P(S_n = s) = \frac{(1-\lambda_0)^{n-1}}{1-p}\frac{(1-\lambda)^{2s}p^s}{(1-\lambda_0)^{2s}(1-p)^s}\times$$

$$\sum_{\sum Z_i = s}\left\{\frac{\lambda(1-\lambda_0)(1-p)}{(1-\lambda)^2 p}\right\}^{\sum_{i=2}^{n} Z_{i-1}Z_i}\left(\frac{1-\lambda_0}{1-\lambda}\right)^{Z_1+Z_n} \tag{3.22}$$

where $\lambda_0$ is given by (4.3).

*Theorem 3.9.* If $Z_1, Z_2, ..., Z_n$ are defined as in theorem 3.6, the pairs $(Z_{i-1}, Z_i)$, i=2, 3, ..., n are negatively correlated with correlation coefficient given by the formula

$$\rho = (\lambda - p)(l - p). \tag{3.23}$$

## 4. LARGE SAMPLE INFERENCE PERTAINING TO THE ADEQUACY OF MODEL

The method suggested for the evaluation of a forecasting model and the results obtained in the previous section provide a framework whereby a sequence of Bernoulli trials is realised with varying probabilities of success $p_i$ with Markov dependence reflected by the parameter $p_{11}^{(i)} = P(Z_i = 1 \mid Z_{i-1} = 1)$ at each step.

The question that naturally arises is how well the model is going to perform at a certain point in time given its recent performance history. In this section, this question is considered through the construction of confidence limits or through testing hypotheses concerning an appropriate parameter that will "reflect" the forecasting potential of the model. It seems that $p_n = \sum_{i=1}^{n} p_i/n$ is a suitable choice and it is obvious from theorem 3.2 that $S_n/n$ is an unbiased estimator of $p_n$.

To illustrate how our methods works in the model-evaluation problem we confine ourselves to the special case of the Markov-dependent binomial distribution with constant transition probabilities, i.e., with $p_i = p$, a constant, for which statistical inference is available.

From the scheme described in section 2 it becomes apparent that the random variables $Z_{i-1}$, $Z_i$ are negatively correlated since $P(Z_i = 1 \mid Z_{i-1} = 1) \le P(Z_i = 1) \le P(Z_i = 1 \mid Z_{i-1} = 0)$.

This suggests a variant of the evaluation scheme whereby step 6 can be modified as follows:

Construct $C_{t+2}$ defining $k_{t+1}$ so that

$$p_i = p, \quad i = 1, 2, ..., n; \quad p \in (0,1) \tag{4.1}$$

$$p^{(i)}_{z_{i-1}} = \lambda^{Z_i} \lambda_0^{1-z_i}, \quad \lambda \le p \le \lambda_0, \quad z_i = 0,1; \quad i = 1, 2, ..., n.$$

The validity of (4.1) can be ensured by a value of $k_{t+1}$ which is an appropriate multiple of

$$\begin{cases} k_t \dfrac{t_{l_t - m + t + 1}, (1+\lambda)/2}{t_{l_t - m + t,}(1+\lambda)/2} & \text{if } Y^0_{t+1} \in C_{t+1} \\ k_t \dfrac{t_{l_t - m + t + 1}, (1+\lambda_0)/2}{t_{l_t - m + t,}(1+\lambda_0)/2} & \text{if } Y^0_{t+1} \notin C_{t+1} \end{cases} \tag{4.2}$$

where $t_{v,\alpha}$ denotes the 100a percentile of the t distibution with $v$ degrees of freedom.

Since $p_1 = p_2 = \ldots = p_n = p$ it follows from (3.6) that $p^{(i)}_{11} = 1 - (1-p)p^{(i)}_{01} / p$, i.e.

$$\lambda_0 = (1-\lambda)\, p / (1-p) \tag{4.3}$$

Hence the probability distribution of $(Z_1, ..., Z_n)$ can be specified completely by the parameters p and $\lambda$.

Using the large sample results of Klotz mentioned in the previous section one may proceed to construct confidence intervals or to test hypotheses for p. Thus, the approximate 100(1-$\alpha$)% confidence limits for p can be obtained for known $\lambda$ by

$$\hat{p} \pm z_{1-\alpha/2}\, \hat{p}\,(1-\hat{p})\,(1-2\hat{p}+\lambda)\,/(1-\lambda) \tag{4.4}$$

where $Z_\alpha$ represents the 100$\alpha$ percentile of the standard normal distribution and $\hat{p}$ stands for either $S_n/n$ or the right hand side of (3.18).

As to the testing of the statistical hypothesis, $H_0$: $p \ge p_0$ versus $H_1$: $p < p_0$ a critical region of size $\alpha$ can be determined by

$$\{s \in [0,1] : \sqrt{n}(1-\lambda)(s/n - p_0) \big/ (p_0(1-p_0)(1-2p_0+\lambda)) < z_{\alpha/2}\} \tag{4.5}$$

In the case where $\lambda$ is not known approximate confidence limits or critical region can be obtained by substituting $\lambda$ by its maximum likelihood estimator $\hat{\lambda}$ as given by (3.17) in (4.4) or (4.5), respectively.

## 5. SMALL SAMPLES APPROACH

Trying to assess the merit of the large sample inference approach pertaining to the methodology suggested by Xekalaki and Katti (1984) or its variant of section 3, one may well argue that it would be of small value: The suggested scheme aims at evaluating models of a time series by exploiting the sequential nature of model-based prediction on the presupposition that the model does not change substantially over the entire study time period. Hence allowing the length of the period to increase by letting n increase, i.e. allowing "eternal" use of model may not be meaningful. It becomes therefore evident that the development of an exact inferential approach is necessary. This will have to depend upon the exact distribution of $S_n$ as given by (3.12). Developing a computational algorithm for generating the cumulative distribution of $S_n$ will be the subject of future research. Going back to the variant of Xekalaki and Katti's scheme of section 3 one can be led to exact confidence statements or tests concerning p based on Ladd's (1975) tabulation of a reparameterized version of the cummulative distribution of (3.22).

Hence, to make a confidence statement of the type:

"$p_l \leq p \leq p_u$ at a confidence level $l - \alpha_l - \alpha_u$"

the limits $p_l$ and $p_u$ can be determined so that

$$\alpha_l = P(S_n \geq s \mid n,\ p = p_l) = 1 - S(n, s-1; p_l)$$

$$\alpha_u = P(S_n \leq s \mid n,\ p = p_u) = 1 - S(n, s; p_u)$$

where S(., .; p) can be determined by entering Ladd's table for $P_{hh}=\lambda$, $P_{hm}=(1-\lambda)p/(1-p)$ and $p=(\lambda-p)\ (1-p)$.

The same table can be used for determining a critical region of size $\alpha$ to test the hypothesis $H_0$: $p \geq p_0$ against $H_1$: $p<p_1$. This will be defined by

$$\{s \in [0, 1]: s \leq t \text{ where t is such that } \alpha = P(S_n \leq t \mid n,\ \ p = p_0\}.$$

## REFERENCES


1. Devore, J.L., (1976). A note on the estimation of parameters in a Bernoulli model with dependence. Annals of Statistics 4, 990-992.
2. Klotz, J., (1973). Statistical inference in Bernoulli trials with dependence. Annals of Statistics, 1, 373- 379.
3. Ladd, D.W., (1975). An Algorithm for the Binomial Distribution with dependent trials. J. Amer. Statist. Assoc., 70, 333-340.
4. Lahiri, P., (1992). Model Selection. IMS Monograph Series, vol. 38.
5. Xekalaki, E. and Katti, S., (1984). A Technique for Evaluating Forecasting Models. Biom. J., 26, 173-184.
6. Xekalaki, E., Panaretos, J. and Psarakis, S., (2003). A Predictive Model Evaluation And Selection Approach - The Correlated Gamma Ratio Distribution. Stochastic Musings: Perspectives from the Pioneers of the Late 20th Century. J.Panaretos (ed.). Laurence Erlbaum, USA, 188-202.



Mailing Address:

John Panaretos

Department of Statistics

Athens University of Economics and Business

76 Patision St., 10434 Athens, GREECE.

jpan@aueb.gr