STATISTICAL RESEARCH REPORT
Institute of Mathematics
University of Oslo

June 1978

A MODEL FOR MULTINOMIAL TRIALS WITH DEPENDENCE

by

Bo Lindqvist

Abstract: Klotz (1972,1973) develops a model for Bernoulli trials with dependence of Markov chain type. The present paper generalizes this model to the multinomial case. The model studied involves the usual frequency parameters $p_1,\ldots,p_r$ and a dependence parameter c. We assume that $p_1,\ldots,p_m,c$ are unknown, while $p_{m+1},\ldots,p_r$ are apriori known $(0\le m\le r)$. The paper gives a system of equations for determining the maximum likelihood estimator $\hat{\theta}$ of $\theta = (p_1,\ldots,p_{m-1},c)$. Asymptotic normality is verified and the asymptotic covariance matrix is derived. An easily computed estimator $\tilde{\theta} = (\tilde{p}_1,\ldots,\tilde{p}_{m-1},\tilde{c})$ is proposed, where $\tilde{p}_j$ are relative frequencies. This estimator is compared with $\hat{\theta}$, and it is shown that $\tilde{\theta}$ is efficient if and only if $r = m = 2$. However, we conclude that the relative frequency estimator $(\tilde{p}_1,\ldots,\tilde{p}_{m-1})$ of $(p_1,\ldots,p_{m-1})$ is rather robust against a reasonable dependence among the trials.

1. Introduction and model

Klotz (1972,1973) develops a model for Bernoulli trials with dependence, the dependence being of Markov chain type. Some comments to Klotz's papers are given by Devore (1976) and Lindqvist (1978). The present paper generalizes the ideas and some of the results of Klotz (1972,1973) to the multinomial case.

Consider a sequence of n trials, each of them resulting in one of the r outcomes $A_1,\dots,A_r$ with probabilities $p_1,\dots,p_r$, respectively. We assume that $p_i \geq 0$ $(i=1,\dots,r)$ and $\sum_{i=1}^{r} p_i = 1$. Let $Y_k = i$ if the k-th trial results in A_i. We shall assume that $Y_1, Y_2, \dots, Y_n$ is a stationary Markov chain on $\{1,2,\dots,r\}$ with stationary distribution $p = (p_1,\dots,p_r)$ and with transition matrix

$$T = (1-c)\,P + c\,I \tag{1.1}$$

where P is the $r \times r$ matrix with each row equal to p and I is the $r \times r$ identity matrix. In order that all entries of T are nonnegative, we must require

$$\max_{1 \leq i \leq r} [1 - (1-p_i)^{-1}] \leq c \leq 1 \tag{1.2}$$

We shall in this paper assume that $p_{m+1},\dots,p_r$ are apriori known and that $p_1,\dots,p_m,c$ are unknown parameters, where $0 \leq m \leq r$. (Note that m=1 is impossible.) We shall let $q_m = 1 - \sum_{i=m+1}^{r} p_i$ (an empty sum will be defined as 0). Hence q_m is known, and since $p_m = q_m - \sum_{i=1}^{m-1} p_i$ it follows that $p_1,\dots,p_{m-1}$ are free to vary. The natural parameter set of our model is thus the set of m-tuples $\theta = (p_1,\dots,p_{m-1},c)$ such that $p_i \geq 0\,(i=1,\dots,m-1)$, $\sum_{i=1}^{m-1} p_i \leq q_m$ and such that (1.2) is satisfied.

It is seen that $c > 0$ and $c < 0$ correspond to, respectively, clustering and lack of clustering among the outcomes. If $c=0$, then we have independent trials and hence $c=0$ corresponds to the usual multinomial case. It is well known that in this case we will still have a multinomial situation if some of the outcomes $A_1,\dots,A_r$ are lumped together. It follows from Theorem 3 of Burke & Rosenblatt (1958) that in the model defined by (1.1) the Markov chain property of $Y_1,Y_2,\dots$ will also be preserved if outcomes are lumped together. In fact it follows that T given in (1.1) is the only $r \times r$ transition matrix having stationary distribution p and having the property that any lumping of outcomes results in a Markov chain.

If we put $r = m = 2$, then the model (1.1) coincides with the one studied by Lindqvist (1978) and which is essentially the model given by Klotz (1973).

Lindqvist (1978) notes that when $r=2$ we have

$$\rho(Y_j,Y_k) = c^{|j-k|}, \tag{1.3}$$

giving a simple interpretation of the parameter c. A direct computation, using the fact that $T^n = (1-c^n)P + c^n I$ shows that this result is valid also in the general case.

The eigenvalues of T may be shown to be 1 and c. Hence it follows from Lindqvist (1977) that the information (in the sense defined in that paper) that Y_k gives about Y_j $(j<k)$ is proportional to $|c|^{k-j}$. This provides another interpretation of c, relating to the memory of the sequence $Y_1,Y_2,\dots$.

We will finally, as a curiosity, mention a simple process $Z_1,Z_2,\dots$ for which the probabilities $z_{ij} = \Pr(Z_{k+1} = j|Z_k = i)$ are also given by (1.1). Let $X_1,X_2,\dots$ be i.i.d. with

$\Pr(X_1=i) = p_i$ $(i=1,\ldots,r)$. Let further $U_1, U_2, \ldots$ be i.i.d. with $\Pr(U_1=1) = 1 - \Pr(U_1=0) = \delta$ where $0<\delta<1$ and assume that the X's and the U's are independent. We define for $k=1,2,\ldots$ $Z_k = U_k X_k + (1-U_k)X_{k+1}$. The z_{ij}'s are now given by (1.1) with $c = \delta(1-\delta)$. The process $Z_1, Z_2, \ldots$ is, however, not a Markov chain, since clearly Z_j and Z_k are independent if $|j-k| \geq 2$.

2. Maximum likelihood estimation

We shall derive the maximum likelihood estimator (MLE) of $(p_1,\ldots,p_{m-1},c)$ using the theory in Section 2 of Billingsley (1961a).

Let $Y_1,\ldots,Y_n$ be given as in Section 1. Define the transition frequences N_{ij} $(1\leq i,j\leq r)$ as the number of integers k $(1\leq k\leq n-1)$ such that $Y_k = i$ and $Y_{k+1} = j$. Let further t_{ij} denote the entries of T. Then the likelihood is given by

$$\prod_{i=1}^{r} p_i^{\delta(Y_1,i)} \prod_{i,j} t_{ij}^{N_{ij}}$$

where δ is the Kronecker function. Taking the logarithm and ignoring the terms corresponding to the first product sign (see note on p.4 in Billingsley, 1961a) we get

$$L_n(p_1,\ldots,p_{m-1},c) = \sum_{i=1}^{r}\Big[\sum_{j\neq i} N_{ij}\ln(p_j(1-c)) + N_{ii}\ln(p_i(1-c)+c)\Big] \tag{2.1}$$

Differentiating with respect to $p_1,\ldots,p_{m-1},c$ and putting the derivatives equal to 0, we get the following m equations to determine the MLE $(\hat{p}_1,\ldots,\hat{p}_{m-1},\hat{c})$:

$$\sum_{i=1}^{m} \frac{N_{ii}}{\hat{p}_i(1-\hat{c})+\hat{c}} + \sum_{i=m+1}^{r} \frac{N_{ii}}{p_i(1-\hat{c})+\hat{c}} = n-1$$
$$\frac{M_i}{\hat{p}_i(1-\hat{c})} + \frac{N_{ii}}{\hat{p}_i(1-\hat{c})+\hat{c}} = \frac{M_m}{\hat{p}_m(1-\hat{c})} + \frac{N_{mm}}{\hat{p}_m(1-\hat{c})+\hat{c}}, \quad (i=1,\ldots,m-1) \tag{2.2}$$

where $\hat{p}_m = q_m - \sum_{i=1}^{m-1} \hat{p}_i$.

The equations (2.2) are easily solved when $r = 2$ (see Lindqvist, 1978). If $r \geq 3$, however, a closed form expression seems difficult to obtain. We will then have to solve (2.2) by some numerical method, e.g. Newton-Raphson's process. A useful set of initial values for numerical iteration is given by the estimator $\tilde{\theta}$ considered in Section 4. It is believed that the system (2.2) has a unique solution at least when all $N_{ii} > 0$.

We shall study in some detail the case $m = 0$. This occurs in practice if the distribution of each Y_k is well known and the dependence parameter c is of interest. It also occurs when testing the hypothesis that $(p_1,\ldots,p_{m-1})$ equals some vector $(p_1^o,\ldots,p_{m-1}^o)$ (see Section 5).

The system (2.2) is now reduced to

$$\sum_{i=1}^{r} \frac{N_{ii}}{p_i(1-\hat{c})+\hat{c}} = n-1. \tag{2.3}$$

We may without loss of generality assume that $p_1=\ldots=p_u < p_{u+1} \leq \ldots \leq p_r$ where $1 \leq u \leq r$. Put $s_i = 1-(1-p_i)^{-1}$ $(i=1,\ldots,r)$. The restriction (1.2) may now be written $s_1 \leq c \leq 1$.

Let the function $f(c)$ be given by the left hand side of (2.3) with $\hat{c}$ replaced by c. f is well defined for $c > s_1$ and is strictly decreasing if and only if $N_{ii} > 0$ for some i. If all

$N_{ii} = 0$, then $f(c) \equiv 0$.

Assume first that $N_{11} = \ldots = N_{uu} = 0$. Then $f(s_1)$ exists and two cases may occur:

(i) $f(s_1) \geq n-1$. Since $f(1) = \sum_{i=1}^{r} N_{ii} \leq n-1$ and f is strictly decreasing, it follows that (2.3) has a unique solution $\hat{c} \in [s_1, 1]$.

(ii) $f(s_1) < n-1$. Since f is decreasing, (2.3) has no solution in $[s_1, 1]$. But clearly $\partial L_n / \partial c < 0$ for all $c \geq s_1$, which implies that L_n is maximized by $c = s_1$ and hence that $\hat{c} = s_1$.

Assume finally that $N_{ii} > 0$ for some $i \leq u$. Then $f(c) \uparrow \infty$ as $c \downarrow s_1$ and since $f(1) \leq n-1$ it follows that (2.3) has a unique solution $\hat{c} \in \langle s_1, 1]$.

We have thus proved that equation (2.3) has at most one solution $\hat{c} \in \langle s_1, 1]$. The MLE $\hat{c}$ is given by this solution if it exists and is otherwise equal to s_1.

Multiplying each term of (2.3) by $\prod_{i=1}^{r} [p_i(1-\hat{c})+\hat{c}]$, the equation (2.3) is transformed to a polynomial equation $h(\hat{c}) = 0$, where h is a polynomial of degree r. When $r=2$ it follows from Theorem 3 in Klotz (1972) that $\hat{c}$ in any case is given by the largest solution of the resulting quadratic equation. It is not difficult to see from what we already have proved that $\hat{c}$ also in the general case is always given by the largest solution of $h(\hat{c}) = 0$.

3. Asymptotic distribution of the MLE

We shall assume that the parameter $\theta = (p_1, \ldots, p_{m-1}, c)$ is contained in the interior of the natural parameter set defined in Section 1. The Markov chain with transition matrix T given in (1.1) clearly satisfies Condition 5.1 of Billingsley (1961a).

For notational convenience, let $\theta_i = p_i$; $i=1,\dots,m-1$ and $\theta_m = c$. Then $\theta = (\theta_1,\dots,\theta_m)$ is the parameter of our model.

By Theorem 5.1 in Billingsley (1961a) there is a consistent solution $\hat{\theta}$ of the equations (2.2). Moreover, if $A(n) = n^{\frac{1}{2}}(\hat{\theta}-\theta)$, then by Theorem 2.2 in Billingsley (1961a) $A_n \xrightarrow{D} N(0,\sigma^{-1}(\theta))$, where $\sigma(\theta) = (\sigma_{ij}(\theta))$ is a $m \times m$ matrix given by

$$\sigma_{ij}(\theta) = E_\theta\left(\frac{\partial \ln t_{Y_1Y_2}}{\partial\theta_i} \cdot \frac{\partial \ln t_{Y_1Y_2}}{\partial\theta_j}\right).$$

A computation shows that

$$\sigma(\theta) = \begin{bmatrix} b_1+b_m & b_m & \cdots & b_m & a_1-a_m \\ b_m & b_2+b_m & \cdots & b_m & a_2-a_m \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ b_m & b_m & \cdots & b_{m-1}+b_m & a_{m-1}-a_m \\ a_1-a_m & a_2-a_m & \cdots & a_{m-1}-a_m & h \end{bmatrix}$$

where

$$a_i = p_i t_{ii}^{-1}$$

$$b_i = (p_i^{-1} - ct_{ii}^{-1})(1-c) \qquad (i=1,\dots,m)$$

$$h = (1-c)^{-1} \sum_{i=1}^{r} p_i(1-p_i)t_{ii}^{-1}$$

That $\sigma(\theta)$ is non-singular, follows from p.24 i Billingsley (1961a). The inverse matrix $\sigma^{-1}(\theta)$, which will be denoted $S = (s_{uv})$, may be computed e.g. by the cofactor method.

For $F \subseteq \{1,2,\dots,m\}$ let $b(F)$ denote the product $\prod_i b_i$ where the index i runs through the set $\{1,2,\dots,m\} \setminus F$. Let $b(\emptyset) = 1$. Then we have

$$\det \sigma(\theta) \equiv d = h \sum_{i=1}^{m} b(i) - \sum_{i<j} b(i,j)(a_i - a_j)^2$$

and for $1 \leq u,v \leq m$ we get

$$s_{uv} = d_{uv}/d \tag{3.1}$$

where

$$\left.\begin{aligned}
d_{uu} &= h \sum_{i \neq u} b(u,i) - \sum_{\substack{i<j \\ i,j \neq u}} b(u,i,j)(a_i - a_j)^2 && \text{for } u=1,2,\ldots,m-1 \\
d_{mm} &= \sum_{i} b\,(i) \\
d_{uv} &= \sum_{i \neq u,v} b(u,v,i)(a_u - a_i)(a_v - a_i) - hb(u,v) \\
& && \text{for } u \neq v,\ 1 \leq u,v \leq m-1 \\
d_{um} &= d_{mu} = \sum_{i \neq u} b(u,i)(a_i - a_u) && \text{for } u=1,2,\ldots,m-1
\end{aligned}\right\} \tag{3.2}$$

As an example, if $p_1 = \ldots = p_r = r^{-1}$, then we find that

$$\left.\begin{aligned}
s_{uu} &= (m-1)m^{-1}r^{-1}(1-c)^{-1}[1+(r-1)c][1+(r-2)c]^{-1} \\
& && \text{for } u=1,\ldots,m-1 \\
s_{mm} &= (r-1)^{-1}(1-c)[1+(r-1)c] \\
s_{uv} &= -\ m^{-1}r^{-1}(1-c)^{-1}[1+(r-1)c][1+(r-2)c]^{-1} \\
s_{um} &= s_{mu} = 0 && \text{for } u=1,\ldots,m-1
\end{aligned}\right\} \tag{3.3}$$

If $m=0$, i.e. $p_1,\ldots,p_r$ are all known, then it is seen that

$$n^{\frac{1}{2}}(\hat{c}-c) \xrightarrow{D} N(0,h^{-1}) \tag{3.4}$$

4. Easily computed estimator

For $i=1,\ldots,r$ define N_i as the number of Y_k such that $Y_k = i$. If $c=0$, then $(N_1,\ldots,N_r)$ has the usual multinomial distribution. It is easily verified that in this case the MLE of $(p_1,\ldots,p_{m-1})$ is given by $(\tilde{p}_1,\ldots,\tilde{p}_{m-1})$ where

$$\tilde{p}_i = q_m N_i \left(\sum_{i=1}^{m} N_i\right)^{-1} \qquad (i=1,\ldots,m-1) \qquad (4.1)$$

If $r = m = 2$, Klotz (1973) (see also Lindqvist, 1978) proves that this $\tilde{p}_i$ is asymptotically equivalent to $\hat{p}_i$ also if $c \neq 0$. As will be seen from the present section, this result will not remain valid for $r \geq 3$.

We shall consider the estimator $\tilde{\theta} = (\tilde{p}_1,\ldots,\tilde{p}_{m-1},\tilde{c})$ where $\tilde{p}_i$ is given by (4.1) and $\tilde{c}$ is defined by

$$\tilde{c} = (r-1)^{-1}\left(\sum_{i=1}^{r} N_{ii} N_i^{-1} - 1\right) \qquad (4.2)$$

If $r = m = 2$, then this is the MLE for c derived in Lindqvist (1978).

If $m = 0$ and $p_i = r^{-1}$ for all i, then the solution of (2.2) is

$$\hat{c} = (r-1)^{-1}\left(\sum_{i=1}^{r} N_{ii} r(n-1)^{-1} - 1\right)$$

Comparing this with (4.2), it is seen that $\hat{c}$ is obtained from $\tilde{c}$ by replacing N_i by $(n-1)r^{-1} = (1-n^{-1})EN_i$. Hence it may seem reasonable that one should replace the N_i's in (4.2) corresponding to known p_i's, by $EN_i = np_i$, which gives the estimator

$$c^* = (r-1)^{-1}\left(\sum_{i=1}^{m} N_{ii}N_i^{-1} + \sum_{i=m+1}^{r} N_{ii}(np_i)^{-1} - 1\right)$$

A computation shows, however, that in the case $c=0$ we have

$$\text{as var } n^{\frac{1}{2}}\hat{c} = (r-1)^{-1}$$

$$\text{as var } n^{\frac{1}{2}}c^* = (r-1)^{-1}[1+(r-1)^{-1}q_m(1-q_m)].$$

Thus $\hat{c}$ and c^* are equivalent only if $q_m = 0$ or 1, i.e. if $m=0$ or r, and $\hat{c}$ is the better estimator otherwise.

That $\hat{c}$ is to be preferred to c^* may also seem reasonable from the following intuitive considerations. We have for $i = m+1,\dots,r$ a choice between using $N_{ii}N_i^{-1}$ and $N_{ii}(np_i)^{-1}$. Assume now that some N_i happens to be too large (small). It seems reasonable that the same will happen to N_{ii}. Hence the value of $N_{ii}N_i^{-1}$ will not be influenced as much as $N_{ii}(np_i)^{-1}$. If $m=0$, then intuitively such deviations will add to 0, which may not happen if $0< m<r$.

The ergodic theorem (see e.g. Billingsley, 1965,p.13) implies that $N_i n^{-1} \xrightarrow{a.s.} p_i$, $N_{ii}N_i^{-1} \xrightarrow{a.s.} t_{ii} = p_i(1-c) + c$ and that $N_{ii}n^{-1} \xrightarrow{a.s.} t_{ii}p_i$ as $n \to \infty$. This implies by (4.1) and (4.2) that $\tilde{\theta}$ is a consistent estimator of θ.

Let now $B(n) = n^{\frac{1}{2}}(\tilde{\theta}-\theta)$. We shall prove that $B(n)$ converges in distribution to a certain multinormal distribution.

Define for $1 \leq i,j \leq r$, $Z_{ij}(n) = n^{-\frac{1}{2}}(N_{ij}-t_{ij}N_i)$. From Theorem 3.1 in Billingsley (1961b) follows that the r^2-dimensional vector $(Z_{ij}(n))$ converges to the multinormal distribution $N(0,\Lambda)$ where $\Lambda = (\lambda_{ij\cdot kl})$ is given by

$$\lambda_{ij\cdot kl} = p_i t_{ij} \delta_{ik} (\delta_{jl} - t_{il}) \tag{4.3}$$

(δ is the Kronecker function).

Define for $j=1,\ldots,r$, $W_j(n) = n^{-\frac{1}{2}}(N_j - np_j)$. Using the facts that $p_j = \sum_{k=1}^{r} p_k t_{kj}$ and that $N_j = \sum_{k=1}^{r} N_{kj} + I(Y_1=j)$ we get

$$W_j(n) = \sum_{k=1}^{r} n^{-\frac{1}{2}}(N_{kj} - t_{kj}N_k) + \sum_{k=1}^{r} t_{kj} n^{-\frac{1}{2}}(N_k - np_k) + n^{-\frac{1}{2}} I(Y_1=j)$$

$$= \sum_{k=1}^{r} Z_{kj}(n) + \sum_{k=1}^{r} t_{kj} W_k(n) + n^{-\frac{1}{2}} I(Y_1=j)$$

$$= \sum_{k=1}^{r} Z_{kj}(n) + p_j(1-c) \sum_{k=1}^{r} W_k(n) + cW_j(n) + n^{-\frac{1}{2}} I(Y_1=j).$$

Clearly $\sum_{k=1}^{r} W_k(n) = 0$. Hence it follows that

$(1-c)W_j(n) = \sum_{k=1}^{r} Z_{kj}(n) + n^{-\frac{1}{2}} I(Y_1=j)$ which implies that

$$W_j(n) \overset{a}{=} (1-c)^{-1} \sum_{k=1}^{r} Z_{kj}(n) \qquad (1 \le j \le r) \tag{4.4}$$

(where $U_n \overset{a}{=} V_n$ shall mean $U_n - V_n \overset{P}{\to} 0$).

Now for $i=1,\ldots,m$,

$$n^{\frac{1}{2}}(\tilde{p}_i - p_i) = n^{-\frac{1}{2}} \frac{q_m N_i - p_i \sum_{j=1}^{m} N_j}{\sum_{j=1}^{m} N_j n^{-1}} = \frac{q_m W_i(n) - p_i \sum_{j=1}^{m}(n)}{\sum_{j=1}^{m} N_j n^{-1}}$$

Since $\sum_{j=1}^{m} N_j n^{-1} \overset{P}{\to} \sum_{j=1}^{m} p_j = q_m$ it follows (e.g. from Theorem 4.4 in Billingsley, 1968) that

$$n^{\frac{1}{2}}(\tilde{p}_i - p_i) \overset{a}{=} W_i(n) - p_i q_m^{-1} \sum_{j=1}^{m} W_j(n) \tag{4.5}$$

Next, by the definition of $\tilde{c}$ we get

$$n^{\frac{1}{2}}(\tilde{c}-c) = (r-1)^{-1} \sum_{i=1}^{r} nN_i^{-1} Z_{ii}(n)$$

and hence

$$n^{\frac{1}{2}}(\tilde{c}-c) \overset{a}{=} (r-1)^{-1} \sum_{i=1}^{r} p_i^{-1} Z_{ii}(n) \tag{4.6}$$

since $N_i n^{-1} \overset{P}{\to} p_i$.

We have thus proved that $B(n) \overset{a}{=} C(n) \equiv (C_1(n),\ldots,C_m(n))$ where $C_i(n)$ for $i=1,\ldots,m-1$ are given by the right hand side of (4.5) and $C_m(n)$ is given by the right hand side of (4.6).

By (4.4) and the property of the $Z_{ij}(n)$, it follows from Theorem 4.4 in Billingsley (1968) that $C(n)$ converges in distribution to a multinormal distribution $N(0,\Sigma)$ and that $B(n)$ converges to the same limit. Σ is easily computed from (4.3-4.6) and we get

$$\left.\begin{aligned}
\Sigma_{uu} &= p_u(q_m-p_u)q_m^{-1}(1+c)(1-c)^{-1} && \text{for } u=1,2,\ldots,m-1 \\
\Sigma_{mm} &= (r-1)^{-2}(1-c)\sum_{i=1}^{r} t_{ii}(1-p_i)p_i^{-1} && \\
\Sigma_{uv} &= -\,q_m^{-1}p_u p_v(1+c)(1-c)^{-1} && \text{for } u\neq v,\ 1\leq u,v\leq m-1 \\
\Sigma_{um} &= \Sigma_{mu} = c(r-1)^{-1}[1-rp_u+p_u(rq_m-m)q_m^{-1}] && \\
& && \text{for } u=1,2,\ldots,m-1
\end{aligned}\right\} \tag{4.7}$$

Putting $p_i = r^{-1}$ for all i, we get

$$\left.\begin{array}{ll} \Sigma_{uu} = r^{-1}m^{-1}(m-1)(1+c)(1-c)^{-1} & \text{for } u=1,2,\ldots,m-1 \\ \Sigma_{mm} = [1+(r-1)c](r-1)^{-1}(1-c) & \\ \Sigma_{uv} = -\,r^{-1}m^{-1}(1+c)(1-c)^{-1} & \text{for } u\neq v,\ 1\leq u,v\leq m-1 \\ \Sigma_{um} = \Sigma_{mu} = 0 & \text{for } u=1,2,\ldots,m-1 \end{array}\right\} \quad (4.8)$$

Assume now that $m \geq 2$ and $p_i = r^{-1}$. The relative asymptotic efficiency of $\tilde{p}_u$ w.r.t. $\hat{p}_u$ is given by (see (3.3) and (4.8))

$$e_r(c) = [1+(r-1)c][1+(r-2)c]^{-1}(1+c)^{-1}.$$

Hence $e_2(c) \equiv 1$, while $e_r(0) = 1$ and $e_r(c) < 1$ for $c \neq 0$ if $r \geq 3$. This proves that $\tilde{\theta}$ is not an efficient estimator in our model if $r \geq 3$ and $m \geq 2$. That **it** is when $r = m = 2$ is shown in Lindqvist (1978.)

Consider finally the case $m = 0$. By (3.4) we have

$$\text{as var } n^{\frac{1}{2}}\hat{c} = h^{-1} = (1-c)\left[\sum_{i=1}^{r} p_i(1-p_i)t_{ii}^{-1}\right]^{-1} \quad \text{and by (4.7)}$$

$$\text{as var } n^{\frac{1}{2}}\tilde{c} = (r-1)^{-2}(1-c)\sum_{i=1}^{r} t_{ii}(1-p_i)p_i^{-1}.$$

Application of Schwarz' inequality gives as var $n^{\frac{1}{2}}\hat{c} \leq$ as var $n^{\frac{1}{2}}\tilde{c}$, where equality holds if and only if $p_i = r^{-1}$ for all i or $c = 0$. Hence $\hat{c}$ is (asymptotically) better than $\tilde{c}$ when $m = 0$.

We have thus proved that $\tilde{\theta}$ is asymptotically efficient if and only if $r = m = 2$. However, a computation shows that in any case, $\Sigma = S + O(c^2)$ as $c \to 0$. Hence we may conclude that the "natural" estimators $\tilde{p}_1,\ldots,\tilde{p}_{m-1}$ are rather robust against the dependence of trials considered in this paper.

5. Testing hypotheses

Assume that $m \geq 2$ and let $1 \leq k < m$. We consider first the hypothesis $H_o : p_i = p_i^o$ $(i=k+1,\ldots,m)$. Let $p_1^*,\ldots,p_{k-1}^*,c^*$ denote the MLE of $p_1,\ldots,p_{k-1},c$ when $p_i = p_i^o$ for $i=k+1,\ldots,m$. From the theory on pp. 17-18 in Billingsley (1961a) it follows that under H_o, $R_n = 2[L_n(\hat{p}_1,\ldots,\hat{p}_{m-1},\hat{c}) - L_n(p_1^*,\ldots,p_k^*,p_{k+1}^o,\ldots,p_{m-1}^o,c^*)]$ converges in distribution to χ^2_{m-k}. This enables us to derive a test with approximate level ε, rejecting when R_n is large.

Finally, consider the hypothesis of independence, $H_1 : c = 0$. The MLE of $p_1,\ldots,p_{m-1}$ under H_1 is clearly $\tilde{p}_1,\ldots,\tilde{p}_{m-1}$. Hence from Billingsley (1961a, pp.17-18) it follows that under H_1,

$$U_n = 2[L_n(\hat{p}_1,\ldots,\hat{p}_{m-1},\hat{c}) - L_n(\tilde{p}_1,\ldots,\tilde{p}_{m-1},0)]$$

converges in distribution to χ^2_1. We reject H_1 if U_n is large.

REFERENCES

Billingsley, P. (1961a). *Statistical inference for Markov processes.* University of Chicago Press, Chicago.

Billingsley, P. (1961b). Statistical methods in Markov chains. *Ann. Math. Statist.* 32, 12-40, 1343.

Billingsley, P. (1965). *Ergodic theory and information.* Wiley, New York.

Billingsley, P. (1968). *Convergence of probability measures.* Wiley, New York.

Burke, C.K. & Rosenblatt, M. (1958). A Markovian function of a Markov chain. *Ann. Math. Statist.* 29, 1112-1122.

Devore, J.L. (1976). A note on the estimation of parameters in a Bernoulli model with dependence. *Ann. Statist.* 4, 990-992.

Klotz, J. (1972). Markov chain clustering of births by sex. *Proc. Sixth Berkeley Symp. Math. Statist. Prob., Vol. 4,* pp. 173-185. University of California Press.

Klotz, J. (1973). Statistical inference in Bernoulli trials with dependence. *Ann. Statist.,* 1, 373-379.

Lindqvist, B. (1977). How fast does a Markov chain forget the initial state? A decision theoretical approach. *Scand. J. Statist.,* 4, 145-152.

Lindqvist, B. (1978). A note on Bernoulli trials with dependence. To appear in *Scand. J. Statist.*