# MARKOV CHAIN CLUSTERING OF BIRTHS BY SEX

JEROME KLOTZ
UNIVERSITY OF WISCONSIN, MADISON

## 1. Introduction and summary

This paper is concerned with a simple generalization of the Bernoulli trials model to a Markov chain which has an additional parameter that measures dependence between trials. Small and large sample distribution theories are worked out for the model with a new and simple closed form expression obtained for the exact distribution of the sufficient statistics.

The model is applied to a sample of birth order data from an appropriate human population and a slight dependence of sex on that of the previous child is found to be significant.

## 2. Notation and model

In the Bernoulli model, denote two valued random variables by $X_i = 1$ with probability $p$ and 0 with probability $q = 1 - p$, for $i = 1, 2, \cdots, n$. The joint distribution for a sequence of independent trials is given by

$$P[X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n] = p^s q^{n-s}, \tag{2.1}$$

where $s = x_1 + x_2 + \cdots + x_n$ and $x_i = 1$ or 0. To generalize this model to permit dependence between successive trials, consider a Markov chain with symmetric conditional probabilities given by

$$P[X_i = 1|X_{i-1} = 1] = P[X_i = 1|X_{i+1} = 1] = \theta p, \tag{2.2}$$

with the remaining conditional probabilities completely determined by symmetry:

$$P[X_i = 0|X_{i\pm1} = 1] = 1 - \theta p, \tag{2.3}$$

$$P[X_i = 1|X_{i\pm1} = 0] = \frac{P[X_i = 1, X_{i\pm1} = 0]}{P[X_{i\pm1} = 0]} = \frac{(1 - \theta p)p}{q}, \tag{2.4}$$

$$P[X_i = 0|X_{i\pm1} = 0] = 1 - \frac{(1 - \theta p)p}{q} = \frac{1 - 2p + \theta p^2}{q}, \tag{2.5}$$

and unconditionally

$$P[X_i = 1] = 1 - P[X_i = 0] = p. \tag{2.6}$$

With the partial support of the National Institutes of Health PHS. Research Grant GM-10525-08 and the National Science Foundation under Grant GP-12093.

The parameter $\theta$ is a measure of dependence. The value $\theta = 1$ gives the Bernoulli model (2.1) and independence, while values of $\theta > 1$ ($\theta < 1$) imply a tendency for pairwise clustering of like (unlike) values in the sequence of random variables. In order to avoid negative conditional probabilities (2.2) through (2.5), we restrict $\theta$ to the range $\max(0, (2p - 1)/p^2) \leqq \theta \leqq 1/p$ which contains the value 1.

The joint distribution of a sequence can be written

$$
\begin{aligned}
(2.7)\quad & P[X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n] \\
&= P[X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \cdots, X_1 = x_1] \\
&\qquad\qquad P[X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \cdots, X_1 = x_1] \\
&= P[X_n = x_n | X_{n-1} = x_{n-1}] P[X_{n-1} = x_{n-1} | X_{n-2} = x_{n-2}] \\
&\qquad\qquad \cdots P[X_2 = x_2 | X_1 = x_1] P[X_1 = x_1],
\end{aligned}
$$

repeatedly using the Markov dependence assumption. Using (2.2) through (2.6) and $x_i = 1$ or $0$, we can write

$$
\begin{aligned}
(2.8)\quad & P[X_i = x_i | X_{i-1} = x_{i-1}] \\
&= (\theta p)^{x_i x_{i-1}} (1 - \theta p)^{(1 - x_i) x_{i-1}} \left[\frac{(1 - \theta p)\, p}{q}\right]^{x_i(1 - x_{i-1})} \left[\frac{1 - 2p + \theta p^2}{q}\right]^{(1 - x_i)(1 - x_{i-1})},
\end{aligned}
$$

$$
(2.9)\qquad P[X_1 = x_1] = p^{x_1} q^{1 - x_1}.
$$

Substituting (2.8) and (2.9) into (2.7), the joint distribution becomes the product

$$
\begin{aligned}
(2.10)\quad & \left\{\prod_{i=2}^{n} (\theta p)^{x_i x_{i-1}} (1 - \theta p)^{(1 - x_i) x_{i-1}} \left[\frac{(1 - \theta p) p}{q}\right]^{x_i (1 - x_{i-1})} \right. \\
&\qquad\qquad \left. \left[\frac{1 - 2p + \theta p^2}{q}\right]^{(1 - x_i)(1 - x_{i-1})}\right\} p^{x_1} q^{1 - x_1} \\
&= \theta^{n_{11}} (1 - \theta p)^{(n_{01} + n_{10})} (1 - 2p + \theta p^2)^{n_{00}} p^{\sum_{i=1}^{n} x_i} q^{-\sum_{i=2}^{n-1} (1 - x_i)},
\end{aligned}
$$

where

$$
(2.11)\qquad
\begin{aligned}
n_{11} &= \sum_{i=2}^{n} x_i x_{i-1} = \sum_{i=1}^{n-1} x_i x_{i+1}, \\
n_{01} &= \sum_{i=1}^{n-1} (1 - x_i) x_{i+1}, \\
n_{10} &= \sum_{i=1}^{n-1} x_i (1 - x_{i+1}), \\
n_{00} &= \sum_{i=1}^{n-1} (1 - x_i)(1 - x_{i+1}),
\end{aligned}
$$

so that $n_{11} + n_{01} + n_{10} + n_{00} = n - 1$.

In this paper attention is restricted to inference on $\theta$ and it is assumed that $p$ is known. The case where $p$ is an unknown nuisance parameter is of interest but somewhat more complicated and will not be considered here.

## 3. Large sample theory

The model (2.10) is a particular case of the general Markov processes discussed by Billingsley [1], and the large sample theory developed there can be applied directly. Expression (2.8) corresponds to the notation $f(X_{i-1}, X_i; \theta)$ in [1]. If terms not depending on $\theta$ are neglected, the log likelihood (the log of (2.10)) can be written

$$(3.1)\quad L_n(\theta) = n_{11} \log \theta + (n_{01} + n_{10}) \log (1 - \theta p) + n_{00} \log (1 - 2p + \theta p^2).$$

The term $f(X_1, \theta)$ of [1] corresponds to (2.9) for the model and does not depend on $\theta$. Thus, the exact likelihood (3.1) and the large sample approximate likelihood used by Billingsley are equivalent. Theorems (2.1) and (2.2) of [1] give the following.

THEOREM 1. *If $\theta$ is restricted to an open interval, there exists a sequence of estimators $\hat{\theta}(X_1, X_2, \cdots, X_n)$ which converge in probability to $\theta$. The sequence $\hat{\theta}(X_1, X_2, \cdots, X_n)$ is a solution of*

$$(3.2)\qquad \frac{dL_n(\theta)}{d\theta} = \frac{n_{11}}{\theta} - \frac{(n_{01} + n_{10})p}{(1 - \theta p)} + \frac{n_{00}p^2}{(1 - 2p + \theta p^2)} = 0,$$

*with probability going to one as $n \to \infty$. Further, the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta)$ is normal $N(0, \tau^2(\theta))$, where*

$$(3.3)\qquad \tau^2(\theta) = \frac{\theta(1 - \theta p)(1 - 2p + \theta p^2)}{p^2(1 - 2p + \theta p)}.$$

*For testing the hypothesis $H:\theta = \theta_0$ against the alternative $A:\theta \neq \theta_0$, the* log *likelihood ratio test statistic has an asymptotic chi square distribution* (1 *degree of freedom*):

$$(3.4)\qquad 2[L_n(\hat{\theta}) - L_n(\theta)] \xrightarrow{\mathcal{L}} \chi_1^2$$

*as $n \to \infty$ under $H$.*

A large sample confidence interval for $\theta$ can be obtained by using (3.4). If $\chi_{1,\alpha}^2$ is the upper $100\alpha$ per cent critical value of the chi square distribution (1 degree of freedom), then the set of $\theta$ values which satisfy

$$(3.5)\qquad 2[L_n(\hat{\theta}) - L_n(\theta)] \leqq \chi_{1,\alpha}^2$$

is a confidence interval for $\theta$ with confidence coefficient approaching $1 - \alpha$ as $n \to \infty$. It is an interval since $L_n(\theta)$ is concave ($L_n''(\theta) < 0$).

## 4. The exact distribution of the sufficient statistics

The joint distribution (2.10) can be rewritten in the form

$$(4.1)\quad \frac{(1 - 2p + \theta p^2)^{n-1}}{q^{n-2}} \left[\frac{\theta(1 - 2p + \theta p^2)}{(1 - \theta p)^2}\right]^{\sum_{i=1}^{n-1} x_ix_{i+1}} \left[\frac{pq(1 - \theta p)^2}{(1 - 2p + \theta p^2)^2}\right]^{\sum_{i=1}^{n} x_i} \left[\frac{1 - 2p + \theta p^2}{(1 - \theta p)q}\right]^{x_1+x_n}.$$

Denote $R = \sum_{i=1}^{n-1} X_iX_{i+1}$, $S = \sum_{i=1}^{n} X_i$, and $T = X_1 + X_n$. By the factorization theorem they are sufficient although not minimal since $N_{11} = R$ and

$$(4.2)\quad (N_{01} + N_{10}) = \sum_{i=1}^{n-1} X_{i+1}(1 - X_i) + X_i(1 - X_{i+1}) = 2(S - R) - T$$

are also sufficient, which can be seen from the likelihood expression (2.10) using $N_{00} = (n - 1) - N_{11} - (N_{01} + N_{10})$. However, the joint distribution of $(R, S, T)$ seems easier to derive than that of $(N_{11}, N_{01} + N_{10})$.

Since the joint distribution (4.1) of $X_1, X_2, \cdots, X_n$ is constant for fixed values of $R$, $S$, and $T$, using (4.1), it follows that

$$(4.3)\qquad P[R = r, S = s, T = t] = M_n(r, s, t)C_{1n}\eta_1^r\eta_2^s\eta_3^t,$$

where

$$(4.4)\qquad \begin{aligned} C_{1n} &= \frac{(1 - 2p + \theta p^2)^{n-1}}{q^{n-2}}, & \eta_1 &= \frac{\theta(1 - 2p + \theta p^2)}{(1 - \theta p)^2},\\ \eta_2 &= \frac{pq(1 - \theta p)^2}{(1 - 2p + \theta p^2)^2}, & \eta_3 &= \frac{(1 - 2p + \theta p^2)}{(1 - \theta p)q},\end{aligned}$$

and $M_n(r, s, t)$ is the number of sequences $(x_1, x_2, \cdots, x_n)$ of zeros and ones which have $\sum_{i=1}^{n-1} x_ix_{i+1} = r$, $\sum_{i=1}^{n} x_i = s$, and $x_1 + x_n = t$.

To count the number of such sequences for given $(r, s, t)$, first note that there are $\binom{2}{t}$ different ways of getting a sum of $t$ from $x_1 + x_n$. Next, since $\sum_{i=1}^{n} x_i = s$, there are a total of $s$ ones in the sequence. If we count the number $z$ of zero runs between the first and the last of these ones in the sequence, we have the relationship

$$(4.5)\qquad r = s - 1 - z.$$

The reason for this is that every time a run of zeros is inserted between consecutive ones, the value of $R$ is decreased by one. The number of ways of putting $n - s$ zeros into $z + (2 - t)$ cells of zero runs where $z = s - 1 - r$ (from 4.5) and no cell is empty is given by

$$(4.6)\qquad \binom{n - s - 1}{z + (2 - t) - 1} = \binom{n - s - 1}{s - r - t}$$

(see, for example, Feller [5], p. 37). We define throughout $\binom{-1}{-1} = 1$, as is required for the special case $s = n$. Finally, the number of ways of inserting $s$ ones into $z + 1$ cells of ones (where the cells of ones are separated by the $z$ zero runs between the first and last ones) with no cell empty is given by

$$(4.7)\qquad \binom{s - 1}{z + 1 - 1} = \binom{s - 1}{s - 1 - r} = \binom{s - 1}{r},$$

again using Feller [5], p. 37, and (4.5). Thus, the total number of ways that these three conditions can hold (and which must hold so that $R = r$, $S = s$, and

$T = t$) is the product of

$$M_n(r, s, t) = \binom{2}{t}\binom{n-s-1}{s-r-t}\binom{s-1}{r}. \tag{4.8}$$

Summarizing, we have proved the following.

THEOREM 2. *If $X_1, X_2, \cdots, X_n$ is a Markov chain satisfying* (2.2) *through* (2.6), *then the sufficient statistics $R$, $S$, and $T$ have joint distribution given by*

$$P[R = r, S = s, T = t] = \binom{2}{t}\binom{n-s-1}{s-r-t}\binom{s-1}{r} C_{1n}\eta_1^r\eta_2^s\eta_3^t, \tag{4.9}$$

*where $C_{1n}$, $\eta_1$, $\eta_2$, and $\eta_3$ are given by* (4.4).

Using (4.9), the joint distribution of $N_{11} = R$ and $N_{01} + N_{10} = 2(S - R) - T$ can be derived since

(4.10)
$$\begin{aligned} P[N_{11} = r, (N_{01} + N_{10}) = w] &= \sum_t P[R = r, 2(S-R) - T = w, T = t] \\ &= \sum_t P[R = r, S = r + \tfrac{1}{2}(w+t), T = t]. \end{aligned}$$

For even values $w = 2u$,

(4.11) $P[N_{11} = r, (N_{01} + N_{10}) = 2u]$
$$\begin{aligned} &= P[R = r, S = r + u, T = 0] + P[R = r, S = r + u + 1, T = 2] \\ &= \binom{r+u-1}{r}\binom{n-r-u-1}{u} C_{1n}\eta_1^r\eta_2^{r+u} \\ &\qquad + \binom{r+u}{r}\binom{n-r-u-2}{u-1} C_{1n}\eta_1^r\eta_2^{r+u+1}\eta_3^2 \\ &= C_{1n}(\eta_1\eta_2)^r\eta_2^u\left[\binom{r+u-1}{r}\binom{n-r-u-1}{u}\right. \\ &\qquad \left. + \binom{r+u}{u}\binom{n-r-u-2}{u-1}\frac{p}{q}\right]. \end{aligned}$$

For odd values $w = 2u + 1$, similarly,

(4.12) $P[N_{11} = r, N_{01} + N_{10} = 2u + 1]$
$$\begin{aligned} &= P[R = r, S = r + u + 1, T = 1] \\ &= 2\binom{r+u}{r}\binom{n-r-u-2}{u} C_{1n}(\eta_1\eta_2)^r\eta_2^u\eta_2\eta_3. \end{aligned}$$

## 5. The exact maximum of the likelihood

For maximizing the likelihood, consider the derivative equation $L_n'(\theta) = 0$ given by (3.2). This equation leads to a quadratic equation and we prove that the solution with the positive square root term maximizes the likelihood in all cases.

THEOREM 3. *The maximum likelihood estimator of $\theta$ is given by*

$$(5.1)\quad \hat{\theta}_+(n_{11}, (n_{01} + n_{10})) = \frac{1}{2}\left[\left\{\frac{2n_{11} + (n_{01} + n_{10}) + m}{mp} - \frac{(n_{11} + n_{10} + n_{01})}{mp^2}\right\}\right.$$
$$\left. + \left(\left\{\frac{2n_{11} + (n_{01} + n_{10}) + m}{mp} - \frac{(n_{11} + n_{01} + n_{10})}{mp^2}\right\}^2 + \frac{4n_{11}(1 - 2p)}{mp^3}\right)^{1/2}\right],$$

*where* $m = n - 1$.

PROOF. See the Appendix.

For the special case of $p = 1/2$, (5.1) reduces to $\hat{\theta}_+ = 2[1 - ((n_{01} + n_{10})/m)]$.

## 6. Large sample distributions and small sample comparisons

Using (4.9) and writing $X = (R - (n - 1)\theta p^2)/\sqrt{n}$, $Y = (S - np)/\sqrt{n}$, it can be shown that the limiting distribution of $(X, Y)$ and $T$ is that of a bivariate normal $N((0, 0), \Sigma)$ and an independent binomial $B(2, p)$, where the asymptotic variance covariance matrix of $X$ and $Y$ is

$$(6.1)\quad \Sigma = \begin{pmatrix} 4\theta p^3 q + \dfrac{\theta p^2(1 - 2p + \theta p)(1 - 2p + \theta p^2)}{1 - \theta p} & \dfrac{2\theta p^2 q^2}{1 - \theta p} \\ \dfrac{2\theta p^2 q^2}{1 - \theta p} & \dfrac{pq(1 - 2p + \theta p)}{1 - \theta p} \end{pmatrix}.$$

This result, which should take no longer than a day to verify, was obtained by writing out the factorials in the binomial coefficients in (4.9), using Stirling's approximation and a log expansion, and taking the limit as $n \to \infty$.

Because of its greater simplicity, one is tempted to use the estimator of $\theta$ given by $R/[(n - 1)p^2]$, which is unbiased. However, from the asymptotic variance of $X$, we note for $\max\{0, (2p - 1)/p^2\} < \theta < 1/p$, $0 < p < 1$,

$$(6.2)\quad \lim_{n\to\infty} n \operatorname{Var}\left[\frac{R}{(n - 1)p^2}\right]$$
$$= \frac{1}{p^4}\left[4\theta p^2 q + \frac{\theta p^2(1 - 2p + \theta p)(1 - 2p + \theta p^2)}{1 - \theta p}\right]$$
$$> \frac{\theta(1 - \theta p)(1 - 2p + \theta p^2)}{p^2(1 - 2p + \theta p)} = \tau^2(\theta) = \lim_{n\to\infty} nE(\hat{\theta}_+ - \theta)^2.$$

For small sample comparisons, because of the complexity of $\hat{\theta}_+$ and the distribution (4.9), it seems unlikely that there exists a computationally convenient closed form expression for its mean and variance except for $p = 1/2$, $\theta = 1$. In this exceptional case, $\hat{\theta}_+ = 2[1 - ((N_{01} + N_{10})/(n - 1))]$ and it can be shown, summing (4.11) and (4.12), that

$$(6.3)\qquad P[(N_{01} + N_{10}) = w] = \binom{n - 1}{w}\frac{1}{2^{n-1}},$$

so that $E\theta_+ = 1$, $\operatorname{Var}\theta_+ = 1/(n-1)$, and $n\operatorname{Var}\theta_+/\tau^2 = n/(n-1)$. Also for $\theta = 1$ but $0 < p < 1$, we can calculate

$$\frac{1}{n}\operatorname{Var} R < p^2q(1+3p)\left[1 - \frac{(1+5p)}{n(1+3p)}\right], \tag{6.4}$$

which has the extra factor $[1 - (1+5p)/(n(1+3p))]$ compared with the asymptotic value $p^2q(1+3p)$ obtained from $\Sigma$. Although the maximum likelihood estimator is still preferred, the finite sample ratio of its variance or mean squared error to that of the unbiased is smaller than the asymptotic ratio for these cases.

For the general case, expectations and variances were numerically computed, in a metallurgical application of the model [8], by Dr. Charles A. Johnson at Argonne Laboratories and are reproduced in Table I. The computations were performed using (4.9) and so forth. For example,

$$E\hat{\theta}_+ = \sum_r \sum_w \hat{\theta}_+(r, w)p[N_{11} = r, (N_{01} + N_{10}) = w], \tag{6.5}$$

with similar expressions for the variances. Table I gives $E\hat{\theta}_+$ and $\operatorname{Var}\hat{\theta}_+$ for selected combinations of $n$, $\theta$, $p$, and compares them with the asymptotic values $\theta$ and $\tau^2(\theta)$. It is interesting to note the oscillatory behavior of both $E\hat{\theta}_+$ and $\operatorname{Var}\hat{\theta}_+$ and the size of the samples for good asymptotic approximation.

TABLE I

FINITE SAMPLE MEANS AND VARIANCES OF $\hat{\theta}_+$ AND COMPARISON WITH THE ASYMPTOTIC APPROXIMATIONS $\theta$, $\tau^2(\theta)$

| $p$ | $\theta$ | $\tau^2(\theta)$ | $n$ | $E\hat{\theta}_+$ | $100(E\hat{\theta}_+ - \theta)/\theta$ | $n\operatorname{Var}\hat{\theta}_+$ | $100[n\operatorname{Var}\hat{\theta}_+ - \tau^2]/n\operatorname{Var}\hat{\theta}_+$ |
|---|---|---|---|---|---|---|---|
| 0.5 | 1.0 | 1.0 | 10 | 1.000 | 0.0% | 1.111 | 10.0% |
| | | | 40 | 1.000 | 0.0 | 1.026 | 2.5 |
| | | | 100 | 1.000 | 0.0 | 1.010 | 1.0 |
| 0.1 | 5.0 | 163.5 | 10 | 1.221 | −75.6 | 197.1 | 17.1 |
| | | | 40 | 4.173 | −16.5 | 252.6 | 35.3 |
| | | | 100 | 4.995 | −0.1 | 205.5 | 20.5 |
| | | | 200 | 5.007 | 0.1 | 181.5 | 9.9 |
| | | | 500 | 4.999 | 0.0 | 169.8 | 3.8 |
| | | | 1000 | 4.996 | −0.1 | 166.5 | 1.8 |
| 0.06 | 2.0 | 433.7 | 10 | .360 | −87.0 | 1230.0 | 64.7 |
| | | | 40 | 1.861 | −7.0 | 1082.0 | 59.9 |
| | | | 100 | 2.192 | 9.6 | 736.3 | 41.1 |
| | | | 200 | 2.019 | 1.0 | 591.6 | 26.7 |
| | | | 500 | 1.997 | −0.1 | 493.4 | 12.1 |
| | | | 1000 | 1.999 | 0.0 | 458.7 | 5.5 |
| 0.06 | 10.0 | 687.7 | 10 | 1.363 | −86.4 | 4513. | 84.8 |
| | | | 40 | 5.002 | −50.0 | 1140. | 39.6 |
| | | | 100 | 8.961 | −10.4 | 1008. | 31.8 |
| | | | 200 | 9.964 | −0.4 | 843. | 18.4 |
| | | | 500 | 10.008 | 0.0 | 741. | 7.2 |
| | | | 1000 | 10.000 | 0.0 | 713. | 3.5 |

## 7. Distribution theory for independent samples

When several independent, identically distributed samples are combined, the distribution theory is modified slightly. Denote each sample size by $N_k$, the corresponding individual sufficient statistics by $(R_k, S_k, T_k)$ or $N_{11k}$ and so forth, for $k = 1, 2, \cdots, K$, where $K$ is the total number of samples to be combined. Using the factorization theorem on the combined joint distribution of $(R_k, S_k, T_k)$, $k = 1, 2, \cdots, K$, or on $N_{11k}$, $(N_{01k} + N_{10k})$, $k = 1, 2, \cdots, K$, we have $R = \sum_{k=1}^{K} R_k$, $S = \sum_{k=1}^{K} S_k$, $T = \sum_{k=1}^{K} T_k$, are sufficient or, more minimally $N_{11} = \sum_k N_{11k} = R$, $(N_{01} + N_{10}) = \sum_k (N_{01k} + N_{10k}) = 2(S - R) - T$ are sufficient. We have the following.

THEOREM 4. *If $(R_k, S_k, T_k)$ are independent for $k = 1, 2, \cdots, K$, then the joint distribution of $R$, $S$, $T$ is given by*

$$(7.1)\qquad P[R = r, S = s, T = t] = \binom{2K}{t}\binom{n - s - K}{s - r - t}\binom{s - K}{r} C_{Kn}\eta_1^r\eta_2^s\eta_s^t,$$

*where $C_{Kn} = (1 - 2p + \theta p^2)^{n-K}/q^{n-2K}$, $n = \sum_k n_k$, and $\eta_1, \eta_2, \eta_3$ are given by* (4.4).

PROOF. We can prove (7.1) by induction on $K$. For $K = 1$, (7.1) reduces to (4.9). For $K + 1$, we compute the joint distribution by convoluting (7.1) and (4.9):

$$(7.2)\qquad \sum_{r_{K+1}=0}^{r}\ \sum_{s_{K+1}=0}^{s}\ \sum_{t_{K+1}=0}^{t}$$

$$\binom{2K}{t - t_{K+1}}\binom{n - n_{K+1} - (s - s_{K+1}) - K}{s - s_{K+1} - (r - r_{K+1}) - (t - t_{K+1})}\binom{s - s_{K+1} - K}{r - r_{K+1}}$$

$$\times\ \eta_1^{r - r_{K+1}}\eta_2^{s - s_{K+1}}\eta_3^{t - t_{K+1}}C_{Kn - n_{K+1}}$$

$$\times\binom{2}{t_{K+1}}\binom{n_{K+1} - s_{K+1} - 1}{s_{K+1} - r_{K+1} - t_{K+1}}\binom{s_{K+1} - 1}{r_{K+1}}\eta_1^{r_{K+1}}\eta_2^{r_{K+1}}\eta_3^{t_{K+1}}C_{1n_{K+1}}$$

$$= \binom{2(K + 1)}{t}\binom{n - s - (K + 1)}{s - r - t}\binom{s - (K + 1)}{r} C_{K+1n}\eta_1^r\eta_2^s\eta_3^t,$$

provided the combinatorial identity can be shown for the sums of the products of the binomial coefficients. To do this use the following three identities:

$$(7.3)\qquad \sum_{t_{K+1}=0}^{t}\binom{2K}{t - t_{K+1}}\binom{2}{t_K} = \binom{2(K + 1)}{t},$$

$$(7.4)\qquad \sum_{r_{K+1}=0}^{r}\binom{s - s_{K+1} - K}{r - r_{K+1}}\binom{s_{K+1} - 1}{r_{K+1}} = \binom{s - (K + 1)}{r},$$

$$(7.5)\qquad \sum_{s_{K+1}=0}^{s}\binom{n - n_{K+1} - (s - s_{K+1}) - K}{s - s_{K+1} - (r - r_{K+1}) - (t - t_{K+1})}\binom{n_{K+1} - s_{K+1} - 1}{s_{K+1} - r_{K+1} - t_{K+1}}$$

$$= \binom{n - s - (K + 1)}{s - r - t}.$$

Both (7.3) and (7.4) hold since the hypergeometric distribution sums to one and (7.5) can be proved by induction on $s$. Note that (7.4) holds for any $s_{K+1}$ and

(7.5) holds for any $r_{K+1}$, $t_{K+1}$. Taking the products of (7.3), (7.4), and (7.5) and using the above remark for appropriate orders of summation gives the required result.

The distribution of the more minimal sufficient statistics $N_{11}$, $(N_{01} + N_{10})$ can be obtained using (4.10) and (7.1).

The maximum likelihood estimate $\hat{\theta}_+$ is similarly given by (5.1) except that $m = n - K$ in the formula, since now $N_{00} + N_{01} + N_{10} + N_{11} = n - K$.

For independent samples, Theorem 1 is modified slightly to read:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N(0, \gamma\tau^2(\theta)) \tag{7.6}$$

as $n \rightarrow \infty$, where $\gamma = \lim [n/(n - K)]$. The proof follows Billingsley [1] (p. 14) with $r = 1$ and $g$ replaced by a $g^{(k)}$, which is the sum of the log of (2.8) over the $k$th sample sequence. A law of large numbers for independent nonidentical random variables and a similar central limit theorem is combined with the martingale argument there.

## 8. An application to birth order data

To apply the model and investigate possible Markov dependence of sex between successive births, appropriate family data was sought. Data was required from families not practicing family limitation by sex so that the fixed sample analysis of the model would be appropriate instead of complicated mixtures of inverse sampling schemes. Family limitation is a common difficulty in sex ratio studies [10] (p. 175), [3], [7].

A geneology of Amish families [6] appears to provide appropriate data. Birth control was considered sinful and the prescript "be fruitful and multiply" was followed [4]. The only type of limitation that might have been practiced was that of limiting the total number once the family size was considered enough. The data also support these considerations since, according to Edwards [3] (p. 343), families practicing limitation by sex composition would have an increased number of girl-boy or boy-girl outcomes in the last two children and the count of such sequences is 86 out of the 195 families in the sample.

Table II gives the coded sex composition in order of birth for selected Amish families. Families were chosen from Old Order Amish or Amish Mennonite parents born before or up to 1900–1910 and who were mostly farmers or carpenters. Families with multiple births were eliminated. To conserve space in the table, the girl and boy family sequence is considered as a binary number with girls corresponding to ones and boys to zeros, a leading one is placed at the front of the sequence and then the resulting sequence converted to the corresponding octal number code. Thus, for example, ggb, gbg, bbg, ggg, bbg is coded 165171. The leading one is used to denote the start of the sequence so that zero sequences of different length are not confused.

In the application of the model, we assume that data from different families are independent and identically distributed and that $p$ and $\theta$ do not differ be-

TABLE II

OCTAL CODED BIRTH ORDER DATA FOR 195 AMISH FAMILIES
Convert the octal number to binary, drop the leading one, and associate girls with ones and boys with zeros in the resulting sequence.

| | | | | | |
|---:|---:|---:|---:|---:|---:|
| 17 | 336 | 7443 | 106 | 240 | 42 |
| 167 | 235 | 264 | 63 | 6 | 4 |
| 203 | 17 | 16 | 75 | 24 | 1137 |
| 24 | 115 | 54 | 13726 | 2153 | 3004 |
| 21 | 11117 | 1670 | 760 | 35 | 237 |
| 342 | 17760 | 306 | 434 | 2636 | 130 |
| 1275 | 367 | 24 | 576 | 254 | 3200 |
| 63 | 1350 | 471 | 115 | 460 | 136 |
| 1703 | 503 | 211 | 13403 | 43677 | 24220 |
| 304 | 156 | 6 | 1402 | 25175 | 367 |
| 267 | 37 | 1575 | 5621 | 33 | 7702 |
| 4676 | 23241 | 56 | 234 | 56 | 4 |
| 206 | 12 | 120 | 1413 | 237 | 41 |
| 32 | 1556 | 37 | 70 | 13 | 1675 |
| 144654 | 2410 | 6211 | 353436 | 1100 | 7 |
| 4456 | 165171 | 16 | 640 | 5377 | 73 |
| 1205 | 4 | 4233 | 302 | 32 | 223 |
| 15020 | 110 | 171 | 100564 | 4602 | 4320 |
| 143 | 4 | 20 | 55 | 27431 | 33 |
| 713 | 103 | 46 | 7 | 161 | 147 |
| 1550 | 43 | 23 | 70 | 3054 | 267 |
| 10372 | 606 | 217 | 237 | 45207 | 70 |
| 7552 | 4350 | 2462 | 27 | 115 | 130 |
| 64 | 763 | 334 | 225 | 317 | 773 |
| 744 | 1175 | 11222 | 46240 | 61 | 561 |
| 131 | 3015 | 1135 | 1363 | 7450 | 2177 |
| 1254 | 47 | 14261 | 71 | 13 | 53 |
| 312134 | 13357 | 4 | 445 | 14 | 37 |
| 510 | 37 | 627 | 1112 | 127 | 652 |
| 70 | 11001 | 62 | 731 | 125 | 650 |
| 14 | 15533 | 41 | 57 | 52 | 1066 |
| 130 | 765 | 14 | 16240 | 144062 | 6370 |
| 47 | 3056 | 134137 | | | |

tween trials or families. Although there is some evidence that $p$ can vary between families [11] (p. 645) and between trials [9] (p. 447), other evidence [2] (p. 249) suggests that this is a not unrealistic assumption for the overall model since the variation is slight. Applying model (7.1) with $K = 195$ families, we compute from Table II, $n = \Sigma n_k = 1466$, $R = \Sigma R_k = 337$, $S = \Sigma S_k = 723$, and $T = \Sigma T_k = 184$. Using either $p = 0.48$ or $p = 0.49$ for the probability of a female birth, we estimate $\hat{\theta}_+ \doteq 1.08$ applying (5.1) with $m = n - K = 1271$. Using (3.4) and $p = 0.49$, we reject the hypothesis that $\theta = 1$ in favor of $\theta \neq 1$ at the 0.01 level of significance, since $2[L_n(\hat{\theta}) - L_n(1)] \doteq 7.10 > 6.6 \doteq \chi^2_{1,.01}$. The 95 per cent confidence interval for $\theta$, using (3.5), is $1.02 \leqq \theta \leqq 1.13$ to two decimals. Although the model is different, the finding of an increase in the conditional probability of a girl given a previous girl is in agreement with Edwards [3] (p. 343) and Renkonen [9].

◇ ◇ ◇ ◇ ◇

## APPENDIX

Proof of Theorem 3. To show that (5.1) maximizes the likelihood, we consider two cases: $(n_{01} + n_{10}) > 0$, and $= 0$.

*Case* 1: $n_{01} + n_{10} > 0$. If we examine the derivative equation (3.2), we note $L_n'(1/p) = -\infty$ and $L_n''(\theta) < 0$. Assume first that $L_n'(\max \{0, (2p - 1)/p^2\}) \leqq 0$. Then $L_n(\theta)$ is decreasing and the maximum occurs on the boundary

$$\theta = \max \{0, (2p - 1)/p^2\}; \tag{A.1}$$

we must show $\hat{\theta}_+$ gives this value. If $p \leqq 1/2$, the assumption reduces to $L_n'(0) \leqq 0$ so that from (3.2), $n_{11} = 0$ and $-(n_{01} + n_{10})p + [n_{00}p^2/(1 - 2p)] \leqq 0$. Using $n_{11} = 0$, $n_{11} + n_{01} + n_{10} + n_{00} = m$, this condition becomes

$$(mp - (n_{01} + n_{10})q) \leqq 0. \tag{A.2}$$

Consequently, evaluating (5.1),

(A.3)

$$\begin{aligned}\hat{\theta}_+ &= \frac{1}{2}\left[\left\{\frac{n_{01} + n_{10} + m}{mp} - \frac{n_{01} + n_{10}}{mp^2}\right\} + \left(\left\{\frac{n_{01} + n_{10} + m}{mp} - \frac{(n_{01} + n_{10})}{mp^2}\right\}^2\right)^{1/2}\right] \\ &= \frac{1}{2}\left[\frac{mp - (n_{01} + n_{10})q}{mp^2} - \frac{mp - (n_{01} + n_{10})q}{mp^2}\right] = 0.\end{aligned}$$

If $p > 1/2$, the assumption becomes $L_n'((2p - 1)/p^2) \leqq 0$ so that $n_{00} = 0$ and $n_{11}p - (2p - 1)m \leqq 0$. Evaluating (5.1),

(A.4)

$$\begin{aligned}\hat{\theta}_+ &= \frac{1}{2}\left[\left\{\frac{n_{11} + 2m}{mp} - \frac{m}{mp^2}\right\} + \left(\left\{\frac{n_{11}p}{mp^2} + \frac{2p - 1}{p^2}\right\}^2 - \frac{4n_{11}pm(2p - 1)}{m^2p^4}\right)^{1/2}\right] \\ &= \frac{1}{2}\left[\frac{2p - 1}{p^2} + \frac{n_{11}p}{mp^2} + \left|\frac{n_{11}p}{mp^2} - \frac{2p - 1}{p^2}\right|\right] = \frac{2p - 1}{p^2}.\end{aligned}$$

Thus, under the assumption $\hat{\theta}_+ = \max \{0, (2p - 1)/p^2\}$. Next assume $L_n'(\max \{0, (2p - 1)/p^2\}) > 0$. Under this assumption there will be a root which maximizes $L_n(\theta)$ in the open interval $(\max \{0, (qp - 1)/p^2\}, 1/p)$ since $L_n''(\theta) < 0$ and $L_n'(1/p) = -\infty$. This root will be one of the two roots of the quadratic equation

$$\theta(1 - \theta p)(1 - 2p + \theta p^2)\left(\frac{n_{11}}{\theta} - \frac{(n_{01} + n_{10})p}{1 - \theta p} + \frac{n_{00}p^2}{1 - 2p + \theta p^2}\right) = 0 \tag{A.5}$$

or

$$\theta^2 - \theta\left\{\frac{2n_{11} + n_{01} + n_{10} + m}{mp} - \frac{(n_{11} + n_{01} + n_{10})}{mp^2}\right\} - \frac{n_{11}(1 - 2p)}{mp^3} = 0. \tag{A.6}$$

Since $\hat{\theta}_+$ is one of the roots of this equation, it remains to rule out the other root

$$\text{(A.7)}\quad \hat{\theta}_- = \frac{1}{2}\left[\left\{\frac{2n_{11} + n_{01} + n_{10}}{mp} - \frac{(n_{11} + n_{01} + n_{10})}{mp^2}\right\} - \left(\left\{\frac{2n_{11} + n_{01} + n_{10} + m}{mp} - \frac{(n_{11} + n_{01} + n_{10})}{mp^2}\right\}^2 + \frac{4n_{11}(1 - 2p)}{mp^3}\right)^{1/2}\right],$$

by showing $\hat{\theta}_- \leqq \max\,\{0, (2p - 1)/p^2\}$. Substituting $n_{01} + n_{10} = m - n_{11} - n_{00}$, we can rewrite

$$\text{(A.8)}\quad \hat{\theta}_- = \frac{1}{2}\left[\left\{\frac{2p - 1}{p^2} + \frac{n_{11}p + n_{00}q}{mp^2}\right\} - \left(\left\{\frac{2p - 1}{p^2} + \frac{n_{11}p + n_{00}q}{mp^2}\right\}^2 + \frac{4n_{11}(1 - 2p)}{mp^3}\right)^{1/2}\right].$$

For $p \leqq \frac{1}{2}$, if we replace the second term in the radical by zero, we obtain

$$\text{(A.9)}\quad \hat{\theta}_- \leqq \frac{1}{2}\left[\left\{\frac{2p - 1}{p^2} + \frac{n_{11}p + n_{00}q}{mp^2}\right\} - \left|\frac{2p - 1}{p^2} + \frac{n_{11}p + n_{00}q}{mp^2}\right|\right] \leqq 0.$$

For $p > \frac{1}{2}$, rewrite $\hat{\theta}_-$ in the form

$$\text{(A.10)}\quad \hat{\theta}_- = \frac{1}{2}\left[\left\{\frac{2p - 1}{p^2} + \frac{n_{11}p + n_{00}q}{mp^2}\right\} - \left(\frac{2p - 1}{p^2} + \frac{2(p - 1)(-n_{11}p + n_{00}q)}{mp^2} + \left\{\frac{n_{11}p + n_{00}q}{mp^2}\right\}^2\right)^{1/2}\right].$$

Replacing the $+n_{00}q$ term in the center of the radical by $-n_{00}q$, we obtain

$$\text{(A.11)}\quad \hat{\theta}_- \leqq \frac{1}{2}\left[\left\{\frac{2p - 1}{p^2} + \frac{n_{11}p + n_{00}q}{mp^2}\right\} - \left(\left\{\frac{2p - 1}{p^2} + \frac{(n_{11}p - n_{00}q)}{mp^2}\right\}^2\right)^{1/2}\right] \leqq \frac{2p - 1}{p^2},$$

so that $\hat{\theta}_- \leqq \max\,\{0, (2p - 1)/p^2\}$.

*Case* 2: $n_{01} + n_{10} = 0$. If $n_{01} + n_{10} = 0$, then $L_n'(\theta) > 0$ so that $L_n(\theta)$ is increasing and maximized at the end point $\theta = 1/p$. Substituting in $\hat{\theta}_+$ with $n_{01} + n_{10} = 0$ gives

$$\text{(A.12)}\quad \begin{aligned} \hat{\theta}_+ &= \frac{1}{2}\left[\left\{\frac{mp - n_{11}(1 - 2p)}{mp^2}\right\} + \left(\left\{\frac{mp - n_{11}(1 - 2p)}{mp^2}\right\}^2 + \frac{4n_{11}(1 - 2p)}{mp^3}\right)^{1/2}\right] \\ &= \frac{1}{2}\left[\left\{\frac{mp - n_{11}(1 - 2p)}{mp^2}\right\} + \left|\frac{mp + n_{11}(1 - 2p)}{mp^2}\right|\right] = \frac{1}{p}. \end{aligned}$$

*Q.E.D.*

Note, $mp + n_{11}(1 - 2p) = (n_{11} + n_{01} + n_{10} + n_{00})p + n_{11}(1 - 2p) \geqq 0$. Methods similar to those used to rule out $\hat{\theta}_-$ can be used to verify $\max\,\{0, (2p - 1)/p^2\} \leqq \hat{\theta}_+ \leqq 1/p$.

◇ ◇ ◇ ◇ ◇

The model was first considered in a metallurgical application with the late Charles Johnson. Conversations with John McDonald, Nathan Keyfitz, and Tom Espenshade (who provided the data) were most helpful.

## REFERENCES

[1] P. Billingsley, *Statistical Inference for Markov Processes*, Chicago, University of Chicago Press, 1961.

[2] A. W. F. Edwards and M. Fraccaro, "Distribution and sequences of sexes in a selected sample of Swedish families," *Ann. Hum. Genet.*, Vol. 24 (1960), pp. 245–252.

[3] A. W. F. Edwards, "Sex-ratio data analysed independently of family limitation," *Ann. Hum. Genet.*, Vol. 29 (1966), pp. 337–346.

[4] T. J. Espenshade, "A new method for estimating the level of natural fertility in populations practicing birth control," *Demography*, Vol. 8 (1971), pp. 525–536.

[5] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I, New York, Wiley, 1957 (2nd ed.).

[6] J. M. Fisher, *Descendants and History of the Christian Fisher Family*, published by Amos L. Fisher, Rt. 1, Ranks, Pa., 1957.

[7] L. A. Goodman, "Some possible effects of birth control on the human sex ratio," *Ann. Hum. Genet.*, Vol. 25 (1961), pp. 75–81.

[8] C. A. Johnson and J. H. Klotz, "The atom probe and Markov chain statistics of clustering," University of Wisconsin Statistics Department Technical Report No. 267, 1971.

[9] K. O. Renkonen, "Is the sex ratio between boys and girls correlated to the sex of precedent children?" *Ann. Med. Exp. Biol. Fenn.*, Vol. 34 (1956), pp. 447–451.

[10] K. O. Renkonen, O. Mäkelä, and R. Lehtovaara, "Factors affecting the human sex ratio," *Ann. Med. Exp. Biol. Fenn.*, Vol. 39 (1961), pp. 173–184.

[11] L. B. Shettles, "Factors influencing sex ratios," *Int. J. Gynaecol. Obstet.*, Vol. 8 (1970), pp. 643–647.