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:

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

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

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

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.