Sankhyā : The Indian Journal of Statistics
Special Issue on Biostatistics
2000, Volume 62, Series B, Pt. 1, pp. 104–118

A BIRTH AND DEATH URN FOR RANDOMIZED CLINICAL TRIALS: ASYMPTOTIC METHODS

By ANASTASIA IVANOVA
University of North Carolina, Chapel Hill
WILLIAM F. ROSENBERGER[1]
University of Maryland, Baltimore
STEPHEN D. DURHAM[2]
University of South Carolina, Columbia
and
NANCY FLOURNOY[3]
American University, Washington

SUMMARY. Consider the situation in which subjects arrive sequentially in a clinical trial. There are K possibly unrelated treatments. Suppose balls in an urn are labeled with treatments. When a subject arrives, a ball is drawn randomly from the urn, the subject receives the treatment indicated on the ball and the ball is returned to the urn. If the treatment is successful, one ball of the same type is added to the urn. Otherwise, one ball of the same type is taken out from the urn. Under certain assumptions, the urn process can be embedded in a continuous time birth and death process, and associated distributional results can thereby be obtained. Since certain treatments can "die out," we introduce a Poisson immigration process to replenish the urn. We derive the maximum likelihood estimators of the success probabilities, prove their consistency and obtain their limiting distributions using martingale theory. We then derive inference procedures for the comparison of the K treatments.

1. Introduction

In comparing K treatments in a clinical trial, subjects arrive sequentially to be randomized to a treatment. Suppose that the response of the nth subject to treatment is known before the $(n+1)$th subject enters the trial. The major goal of the trial is to gather data from subjects to derive information about the effectiveness

AMS (1991) *subject classification.* Primary 62L05, 60J85.
Key words and phrases. Adaptive designs, birth and death process, immigration, sequential design, urn models.
[1]Research supported by grant 1 R29-DK51017-03 from the National Institute of Diabetes and Digestive and Kidney Diseases.
[2]Research supported by grant DMS 95-03104 from the National Science Foundation.
[3]Research supported by grant CR822636-02 from the Environmental Protection Agency.

of the K treatments. There is, however, an ethical constraint: to provide the best possible medical care for the individual subject. The ethical imperative may be compromised if an equal allocation scheme is used, as accruing evidence begins to favor some therapies over others. Adaptive designs have been proposed to remedy this situation. A response-adaptive design seeks to skew assignment probabilities to favor the treatments performing better thus far in the study, according to the magnitude of the treatment effect. In an adaptive urn design, balls are given labels that correspond to different treatments. When a ball is drawn, the next subject is given the corresponding treatment. Urn designs differ according to the rule that governs putting balls back into the urn. In this paper, we discuss a new adaptive urn design, called *a birth and death urn*. This design extends a previous model described by Durham, Flournoy and Li (1998) by allowing balls to be taken out of the urn.

Zelen (1969) originated the idea of using an urn model for clinical trials. A birth and death urn differs from other urn models suggested for randomized clinical trials in the way that it treats failures. Wei and Durham (1978) proposed the *randomized play-the-winner rule* which Wei (1979) generalized to accommodate more than two treatments. This design adds one ball of each opposing treatment type to the urn if there is a treatment failure. Anderson, Faries, and Tamura (1994) proposed an urn model which adds fractional balls to opposing treatments in the same proportions as the current urn composition. Durham and Yu (1990) proposed a two arm model which ignores failures entirely; the urn composition only changes when there is a success. Li (1995), and Durham, Flournoy and Li (1998) developed a *randomized Pòlya urn* model as an extension of Durham and Yu (1990) for more then two treatments.

In this paper we derive the asymptotic properties of the new birth and death urn. In Section 2, we define a birth and death urn design and then an extension of the above design, a birth and death urn *with immigration*. In Section 3, we derive the generating functions for the number of births and splits and obtain limiting results for the birth and death urn with immigration. Likelihood-based estimation is derived in Section 4. In Section 5, we derive the asymptotic distribution for a likelihood-ratio test and use this to explore power. Finally, we make some concluding remarks in Section 6.

2. A Birth and Death Urn

2.1 Definition of a birth and death urn. We define a *birth and death urn* as follows:

An urn contains balls of K types, representing K possibly unrelated treatments. Subjects arrive sequentially to be assigned to a treatment in a clinical trial. A ball is drawn at random and replaced. If it is a type i ball, the subject receives treatment i. Assume that response is dichotomous and immediate. If a success is observed, a type i ball is added to the urn; if a failure is observed, a type i ball is removed. The process continues sequentially with the arrival of the next subject.

If we start the trial having Z_{0i} balls of type i, then $\mathbf{Z_0} = \{Z_{01}, ..., Z_{0K}\}$ is the initial *urn composition*. We assume that the vector $\mathbf{Z_0}$ is fixed. Let $p_i = Pr\{\text{success} \mid \text{treatment } i\}$. Let $\mathbf{Z_n} = \{Z_{n1}, ..., Z_{nK}\}$ denote the composition of the urn after n consecutive draws, where Z_{ni} is the number of type i balls. Then $\{\mathbf{Z_n}, n = 0, 1, 2, ...\}$ is a stochastic process defined on K dimensional integer lattice.

As will be shown later, if the probability of success for some treatments does not exceed 0.5, there will be a point when the type of balls corresponding to this treatment will become extinct and no subjects will be assigned to this ineffective treatment.

2.2 Embedding the urn scheme into a continuous time birth and death process. To assist in the characterization of $\mathbf{Z_n}$ we employ the technique of embedding $\mathbf{Z_n}$ into a continuous-time K-type Markov process, which in the present case is a collection of K independent linear birth and death processes (given by Theorem 2.1). Conventionally, continuous-time birth and death processes describe the behavior of particles which split (produce or die) at certain time points. In our context, "particles" correspond to balls in the urn, and "splitting" of a type i particle corresponds to drawing a type i ball from the urn followed by random replacement (i.e., a birth or a death). We impose upon these independent processes the condition that, for each $i, i = 1, ..., K$, the growth rate is equal to the probability of success on the ith treatment, p_i, and a death rate equal to $q_i = 1 - p_i$.

Hence, define $\{\mathbf{Z(t)} = (Z_1(t), ..., Z_K(t)) \mid t \geq 0\}$, where $Z_i(t)$, $i = 1, ..., K$, is a collection of independent linear birth and death processes described above. Consider the split times, τ_ω, $\omega = 0, 1, 2, ...$, of $\mathbf{Z(t)}$, and the discrete process $\{\mathbf{Z}(\tau_\omega), \omega = 0, 1, 2, ...\}$. We have the following theorem.

THEOREM 2.1. *Given that* $\mathbf{Z}(\mathbf{0}) = \mathbf{Z_0}$, *the stochastic processes* $\mathbf{Z_n} = \{Z_{n1}, ..., Z_{nK}\}$ *and* $\{\mathbf{Z}(\tau_\omega), \omega = 0, 1, 2, ...\}$ *are equivalent.*

PROOF. The proof is similar to that of Athreya and Ney (1972), page 221. □

The distribution and other characteristics of the number of particles in a single linear birth and death process at time t are known (see, for example, Anderson, 1991). In terms of our urn process, time is a construct that has no meaning. However, asymptotic results are the same for both the continuous time and the urn processes. Using the embedding theorem (Theorem 2.1), we obtain asymptotic results for the urn. But first, we extend our design to include immigration.

2.3 A birth and death urn with immigration. If a treatment is particularly harmful, the type of balls corresponding to the treatment will eventually become extinct. From the embedding theorem and theory of the continuous time birth and death processes, it follows that, without immigration, the probability of eventual extinction of type i balls (Anderson, 1991, page 109) is equal to

$$\begin{cases} 1, & \text{if } p_i \leq 0.5 \\ (\dfrac{q_i}{p_i})^{Z_{0i}}, & \text{if } p_i > 0.5. \end{cases}$$

However, sometimes the natural extinction of balls of certain types is not desirable. For instance, there is a positive probability that a type can become extinct even if

its probability of success is high. Thus we also consider the following supplemented birth and death urn, which we call a *birth and death urn with immigration*:

An urn contains balls of K types representing K treatments, and aK immigration balls. When a subject arrives a ball is drawn at random and replaced. If it is an immigration ball, one ball is added to the urn with each ball type having probability $1/K$ of being the one added, and the next ball is drawn. The procedure is repeated until the ball representing an actual treatment is drawn (i.e., not an immigration ball). If it is a type i ball, the subject receives treatment i. If a success is observed, a type i ball is added to the urn; if a failure is observed, a type i ball is removed. The process continues sequentially with the arrival of the next subject.

The parameter a is called the rate of immigration. If $a = 0$, we have the birth and death design considered earlier. As in Section 2.2, the birth and death urn with immigration can be embedded into a continuous-time K-type Markov process, which in this case is a collection of K independent linear birth and death processes with immigration. The birth rate is p_i, the death rate is $q_i = 1 - p_i$, and the immigration rate is a.

3. Limiting Sample Sizes for the Urn with Immigration

In this section we provide limiting results for the birth and death urn. In Section 3.1, we derive the generating functions for the number of births and splits in the birth and death process with immigration. In Section 3.2, we link them to the urn using Theorem 2.1 and then derive corresponding asymptotic results.

3.1 Generating functions for the number of births and splits. Consider the process $Z_i(t)$, the number of particles at time t for the ith continuous-time process. Consider also the processes $N_i(t)$ and $X_i(t)$, the number of splits (i.e., trials on treatment i) and the number of births (i.e., successes on treatment i) by or at time t, respectively.

The probability generating function for the process $Z_i(t)$ can be found, for example, in Anderson (1991). To obtain the probability generating function for $X_i(t)$ we use an idea of Cox and Miller (1965) and consider a two-dimensional process $(X_i(t), Z_i(t))$, with state space as the set of integer pairs (x, z). The probability generating function for the number of births, $G_i(w, t \mid Z_{0i})$, is the solution of the following partial differential equation [see Ivanova (1998) for details]:

$$\frac{\partial G_i(w,s,t \mid Z_{0i})}{\partial t} = (p_i w s^2 - s + q_i)\frac{\partial G_i(w,s,t \mid Z_{0i})}{\partial s} - a(s-1)G_i(w,s,t \mid Z_{0i}).$$

We assume $Z_{0i} = 0$ throughout the remainder of the paper. If $Z_{0i} \neq 0$ the process is the sum of two processes: the first process initiated by the balls which were in the urn from the beginning and the second process initiated by the balls constantly immigrating into the urn. The first process is a birth and death process without immigration. The limiting results for this process separately can be found in Ivanova (1998). Limiting results for the case $Z_{0i} \neq 0$ may be of interest for future research.

The probability generating function for the number of births for $p_i \neq 0$ is

$$G_i(w,t \mid 0) = \left[\frac{2\rho}{e^{-\rho t}(2p_iw-1+\rho)-(2p_iw-1-\rho)}\right]^{\frac{a}{p_iw}} e^{-\frac{ta}{2p_iw}(2p_iw-1+\rho)}, \quad (1)$$

where $\rho = \sqrt{1-4p_iq_iw}$, $|w| \leq 1$. The probability generating function, $F_i(w,t \mid 0)$, for the total number of splits, $N_i(t)$, can be obtained similarly and is

$$F_i(w,t \mid 0) = \left[\frac{2\sigma}{e^{-\sigma t}(2p_iw-1+\sigma)-(2p_iw-1-\sigma)}\right]^{\frac{a}{p_iw}} e^{-\frac{ta}{2p_iw}(2p_iw-1+\sigma)}, \quad (2)$$

where $\sigma = \sqrt{1-4p_iq_iw^2}$, $|w| \leq 1$.

The expected number of splits by time t, $E(N_i(t)) = E(N_i(t) \mid 0)$, can be obtained from the generating function and is

$$E(N_i(t)) = \begin{cases} a(e^{(p_i-q_i)t}-(p_i-q_i)t-1)/(p_i-q_i)^2 & \text{when } p_i \neq 0.5 \\ at^2/2 & \text{when } p_i = 0.5. \end{cases} \quad (3)$$

3.2 Limiting distribution of the number of subjects assigned to the ith treatment. Suppose we have split times $\tau_1, ..., \tau_n$. Now we define the sample sizes on each treatment after n subjects to be $N_i(n)$, where, in terms of the continuous-time birth and death process, $N_i(n) = N_i(\tau_n)$. Note that, by construction, $N(\tau_n) = n$.

Consider the proportion of subjects assigned to the ith treatment for each i, $i = 1, ..., K$, after n subjects, $N_i(n)/n$. We are interested in the limit of the proportion in probability when $n \to \infty$. It follows from Theorem 2.1 that

$$\lim_{n\to\infty} \frac{N_i(n)}{n} = \lim_{t\to\infty} \frac{N_i(t)}{\sum_{j=1}^{K} N_j(t)}, \quad (4)$$

where $N_i(t)$ is the number of type i particles at time t. Consider $E^*(N_i(t))$, the dominant term of $E(N_i(t))$ in the sense $E(N_i(t)) = E^*(N_i(t)) + o(E^*(N_i(t)))$:

$$E^*(N_i(t)) = \begin{cases} ae^{(p_i-q_i)t}/(p_i-q_i)^2 & \text{when } p_i > 0.5 \\ at^2/2 & \text{when } p_i = 0.5 \\ at/(q_i-p_i) & \text{when } p_i < 0.5. \end{cases}$$

The following theorem gives the rates of divergence for the various subsample sizes as a function of t.

THEOREM 3.1.

$$N_i(t)/E^*(N_i(t)) \xrightarrow{P} W_i(p_i), \quad (5)$$

where

$$W_i(p_i) = \begin{cases} Gamma(a/p_i, p_i/a) & \textit{when } p_i > 0.5 \\ \textit{has characteristic function } [\cosh(\tau\sqrt{-1}/a)^{1/2}]^{-2a} & \textit{when } p_i = 0.5 \\ 1 & \textit{when } p_i < 0.5 \end{cases} \quad (6)$$

PROOF. Characteristic functions of the random variable $\lim_{t\to\infty} N_i(t)/E^*(N_i(t))$ can be obtained by substituting $w = exp(\sqrt{-1}\tau/E^*(N_i(t)))$ in $G_i(w, s, t \mid 0)$ and calculating the limit when t goes to infinity. When $p_i > 0.5$, the limit is $[1 - p_i/a\sqrt{-1}\tau]^{a/p_i}$, which is the characteristic function of the Gamma distribution. When $p_i = 0.5$, the limit is $[\cosh(\tau\sqrt{-1}/a)^{1/2}]^{-2a}$. When $p_i < 0.5$, the limit is $exp(\sqrt{-1}\tau)$ and the limiting distribution is degenerate. Limits exist if $p_i \geq 0.5$ when the limiting random variables are not degenerate since we have the following convergence uniformly in $\delta > 0$:

$$\lim_{t\to\infty} E[N_i(t+\delta)/E^*(N_i(t+\delta)) - N_i(t)/E^*(N_i(t))]^2 = 0. \qquad \square$$

We assume without loss of generality that

$$p_1 \geq p_2 \geq ... \geq p_K.$$

THEOREM 3.2.
Case 1. $p_1 \geq 0.5$ *and* $p = p_1 = p_2 = ... = p_h > p_{h+1} \geq ... \geq p_K$.

$$\frac{N_i(n)}{n} \underset{n\to\infty}{\overset{P}{\to}} D_i = \begin{cases} \dfrac{W_i(p)}{W_1(p) + ... + W_h(p)} & \text{if } i \leq h \\ 0 & \text{if } i > h, \end{cases}$$

where $W_i(p), i = 1, ..., h$, *are random variables as in (6).*
Case 2. $p_1 < 0.5$

$$\frac{N_i(n)}{n} \underset{n\to\infty}{\overset{P}{\to}} D_i = \frac{\frac{1}{q_i - p_i}}{\frac{1}{q_1 - p_1} + ... + \frac{1}{q_K - p_K}}.$$

Note. In Case 1 the joint distribution of $(D_1, ..., D_h)$ is a $(h-1)$-variate Dirichlet distribution with all parameters equal to a/p.

PROOF. For the first case, dividing the numerator and denominator of (4) by $E^*(N_1(t))$, we have

$$\lim_{t\to\infty} \frac{N_i(t)}{\sum_{j=1}^{K} N_j(t)} = \frac{c_i W_i(p)}{\sum_{j=1}^{K} W_j c_j},$$

where $W_i(p) = \lim_{t\to\infty} N_i(t)/E^*(N_i(t))$ is the random variable which depends on p according to (6), and

$$c_i = \lim_{t\to\infty} E^*(N_i(t))/E^*(N_1(t)) = \begin{cases} 1 & \text{if } i \leq h \\ 0 & \text{if } i > h. \end{cases}$$

Then

$$\frac{N_i(t)}{\sum_{j=1}^{K} N_j(t)} \underset{t\to\infty}{\overset{P}{\to}} \begin{cases} \dfrac{W_i(p)}{W_1(p) + ... + W_h(p)} & \text{if } i \leq h \\ 0 & \text{if } i > h. \end{cases}$$

The result of the theorem for $p_1 \geq 0.5$ now follows from (4). When $p_1 < 0.5$ dividing the numerator and the denominator by $at/(q_i - p_i)$ and using Theorem 3.1 and (4), we get the result. □

Note that if $h = 1$, and the maximum success probability is equal to or greater than 0.5, the limiting proportion of subjects receiving the (unique) best treatment is one and the limiting proportion of subjects on all other treatments is zero. More precise rates of divergence for the treatments subsamples are given in the following corollary.

COROLLARY 3.1 *(a) If $p_1 > p_j > 0.5$, $N_j(n)n^{-(p_j-q_j)/(p_1-q_1)}$ converges to a random variable in probability.*
(b) If $p_1 > p_j = 0.5$, $N_j(n)log^{-2}(n)$ converges to a random variable in probability.
(c) If $p_1 > 0.5 > p_j$, $N_j(n)log^{-1}(n)$ converges to a random variable in probability.
(d) If $p_1 = 0.5 > p_j$, $N_j(n)n^{-1/2}$ converges to a random variable in probability.

4. **Estimation**

4.1 Consistency of the maximum likelihood estimator. Consider a birth and death urn design with immigration. Suppose there are n subjects, each of which will be randomly assigned to one of K treatments during the course of the study.

Let $\mathbf{D^n}$ denote all the treatment assignments and their associated outcomes as well as the history of additions to the urn, including immigration, up to and including the nth subject. The likelihood function can be derived as in Rosenberger, Flournoy and Durham (1997). The likelihood $\mathcal{L}_n$ reduces to

$$\mathcal{L}_n \propto \prod_{i=1}^{K} p_i^{X_i(n)} q_i^{N_i(n)-X_i(n)},$$

where $N_i(n)$, the number of type i balls drawn from the urn after n draws, and $X_i(n)$, the number of successes on treatment i after n draws.

Consider

$$\frac{\partial \log \mathcal{L}_n(\mathbf{p})}{\partial p_i} = \sum_{j=1}^{n} \frac{\partial}{\partial p_i}\{\log \mathcal{L}_j(\mathbf{p}) - \log \mathcal{L}_{j-1}(\mathbf{p})\} = \sum_{j=1}^{n} \frac{\partial L_j(\mathbf{p})}{\partial p_i},$$

where $\mathcal{L}_0(\mathbf{p}) = 1$, and $L_j(\mathbf{p}) = \log \mathcal{L}_j(\mathbf{p}) - \log \mathcal{L}_{j-1}(\mathbf{p})$. The first derivative of the loglikelihood is given by

$$\frac{\partial \log \mathcal{L}_n(\mathbf{p})}{\partial p_i} = \sum_{j=1}^{n} \frac{\partial L_j(\mathbf{p})}{\partial p_i} = \frac{X_i(n)}{p_i} - \frac{N_i(n) - X_i(n)}{(1 - p_i)}. \tag{7}$$

Hence, the maximum likelihood estimator of p_i is

$$\hat{p}_i = \frac{X_i(n)}{N_i(n)}, \tag{8}$$

the proportion of observed successes at treatment i.

For each i, $i = 1, ..., K$, consider the following quantity:

$$I_n^i(\mathbf{p}) = \sum_{j=1}^{n} E_{j-1}\left(\frac{\partial L_j(\mathbf{p})}{\partial p_i}\right)^2 = \frac{N_i(n)}{p_i(1-p_i)}, \tag{9}$$

where E_{j-1} denote an expectation conditional on data accrued through the $(j-1)$st trial. The quantity $I_n^i(\mathbf{p})$ reduces to the standard Fisher information when random variables are independent. Consider also

$$J_n^i(\mathbf{p}) = \sum_{j=1}^{n} \frac{\partial^2 L_j(\mathbf{p})}{\partial p_i^2} = -\left(\frac{X_i(n)}{p_i^2} + \frac{N_i(n) - X_i(n)}{(1-p_i)^2}\right). \tag{10}$$

It can be shown that the following two conditions are satisfied by the loglikelihood:

(A1) $\int \mathcal{L}_n(\mathbf{p}) dD^n$ can be partially differentiated twice (with respect to each p_i, $i = 1, ..., K$) under the integral sign and the first partials have finite moments of order two.

(A2) $I_n^i(\mathbf{p}) \overset{a.s.}{\to} \infty$ as $n \to \infty$, $i = 1, ..., K$.

THEOREM 4.1. *There exists a sequence of maximum likelihood estimators* (8), $\{\hat{\mathbf{p}}_\mathbf{n} = (\hat{p}_{1n}, ..., \hat{p}_{Kn})\}$ *such that the* $K-$ *dimensional vector* $(\hat{\mathbf{p}}_\mathbf{n} - \mathbf{p})$ *converges almost surely to* $\mathbf{0}$ *as* $n \to \infty$.

PROOF. Since (A1) and (A2) hold, the proof is similar to that of Theorem 6.4 in Hall and Heyde (1980, page 175). □

4.2 Limiting distribution related to the maximum likelihood estimator. Without loss of generality, we assume that the p_i occur in nonincreasing order. The following theorem is a first step toward characterizing the asymptotic distribution of $\hat{\mathbf{p}}_\mathbf{n}$.

THEOREM 4.2.

Case 1. $p \geq 0.5$ *Assume* $p_1 = p_2 = ... = p_h = p > p_{h+1}$. *Consider the set* A, $A = \{1, 2, ..., h\}$ *of treatments. The vector with components* $\sqrt{N_i(n)}(\hat{p}_{in} - p)$, $i \in A$, *is asymptotically multivariate normal with mean zero, variances with components* $p(1-p)$, *and all covariances zero.*

Case 2. $p < 0.5$ *The vector with components* $\sqrt{N_i(n)}(\hat{p}_{in} - p_i)$, $i = 1, ..., K$, *is asymptotically multivariate normal with mean zero, variances with components* $p_i(1-p_i)$, *and all covariances zero.*

PROOF. *Case* 1. $p \geq 0.5$

First we show that the following conditions are satisfied:

(A3) For all $\epsilon > 0$,

$$\sum_{j=1}^{n} E_{j-1}\left[\frac{1}{n}\left(\frac{\partial L_j(\mathbf{p})}{\partial p_i}\right)^2 I\left(\left|\frac{1}{\sqrt{n}}\frac{\partial L_j(\mathbf{p})}{\partial p_i}\right| > \epsilon\right)\right] \overset{P}{\to} 0$$

$n \to \infty$, $i = 1, ..., K$;

(A4) $n^{-1} I_n^i(\mathbf{p}) \overset{P}{\to} \eta^2 > 0$, $i = 1, ..., K$, where η^2 is a.s. a positive random variable.

(A5) $J_n^i(\mathbf{p})/I_n^i(\mathbf{p}) \xrightarrow{P} -1$ as $n \to \infty$, $i = 1, ..., K$.

Condition (A3) can be proven by using the Cauchy-Schwartz inequality. According to Theorem 3.2 condition (A4) is satisfied on the set A. Condition (A5) follows from (8), (9), (10), and the consistency of the estimator (8).

Consider the Taylor series expansion of the function $\sum_{j=1}^{n} \partial L_j(\mathbf{p})/\partial p_i$, $i \in A$, around $\mathbf{p}$, where $\mathbf{p}$ is the vector of true parameters:

$$\sum_{j=1}^{n} \frac{\partial L_j(\mathbf{p}')}{\partial p_i} = \sum_{j=1}^{n} \frac{\partial L_j(\mathbf{p})}{\partial p_i} + \sum_{j=1}\sum_{k\in A}^{n} (p_k' - p_k)\frac{\partial^2 L_j(\mathbf{p}^*)}{\partial p_i \partial p_k}, \tag{11}$$

where $\mathbf{p}^*$ is a point on a line segment connecting $\mathbf{p}'$ and $\mathbf{p}$. From (7), it is clear that

$$\frac{\partial^2 L_j(\mathbf{p})}{\partial p_i \partial p_k} = 0 \text{ if } i \neq k,$$

and thus, (11) can be written as

$$\sum_{j=1}^{n} \frac{\partial L_j(\mathbf{p}')}{\partial p_i} = \sum_{j=1}^{n} \frac{\partial L_j(\mathbf{p})}{\partial p_i} + \sum_{j=1}^{n} (p_i' - p_i)\frac{\partial^2 L_j(\mathbf{p}^*)}{\partial p_i^2}. \tag{12}$$

When we replace $\mathbf{p}'$ with $\hat{\mathbf{p}}$, the left hand side of equation (12) vanishes, and we obtain

$$\sum_{j=1}^{n} \frac{\partial L_j(\mathbf{p})}{\partial p_i} = -\sum_{k\in A}(\hat{p}_k - p_k) \sum_{j=1}^{n} \frac{\partial^2 L_j(\mathbf{p}^*)}{\partial p_i^2}.$$

For each $i \in A$, consider the equation

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\partial L_j(\mathbf{p})}{\partial p_i} = \sqrt{n}(\hat{p}_i - p_i)\frac{1}{n} I_n(\mathbf{p}) - \frac{1}{\sqrt{n}}(\hat{p}_i - p_i)(J_n(\mathbf{p}^*) + I_n(\mathbf{p})).$$

Then according to Corollary 3.1 (Hall and Heyde, 1980, page 58)

$$\sum_{j=1}^{n} \frac{\partial L_j(\mathbf{p})}{\partial p_i} \xrightarrow{D} \eta N(0,1),\ i \in A,$$

where η is a random variable from (A4). Furthermore, from (A4) and (A5), it follows that

$$\sqrt{I_n^i(\mathbf{p})}(\hat{p}_i - p) \xrightarrow{D} N(0,1),\ i \in A,$$

or equivalently,

$$\sqrt{N_i(n)}(\hat{p}_i - p) \xrightarrow{D} N(0, p(1-p)),\ i \in A.$$

Consider

$$\frac{1}{\sqrt{N_i(n)}} \sum_{j=1}^{n}\sum_{i\in A} c_i \frac{\partial L_j(\mathbf{p})}{\partial p_i},$$

where $\{c_i\}$, $i \in A$, are arbitrary nonzero constants. Then

$$\frac{1}{\sqrt{N_i(n)}} \sum_{j=1}^{n} \sum_{i=1}^{K} c_i \frac{\partial L_j(\mathbf{p})}{\partial p_i} \xrightarrow{D} N(0, \sum_{i=1}^{K} c_i^2 p_i(1-p_i))$$

for any nonzero vector $\mathbf{c}$. Employing the Cramér-Wold device gives us the desired result.

Case 2. $p < 0.5$

The proof is similar to that of Case 1. □

Now we can combine Theorem 3.2 and Theorem 4.2 to get the main result first for the case $p \geq 0.5$. Introduce the following notation: $\mathbf{d} = (d_1, ..., d_h)$ is a realization of $(D_1, ..., D_h)$, and $\mathbf{v} = (z_1, ..., z_h)$ is a realization of $\mathbf{V_n} = \sqrt{n}(\hat{p}_{1n} - p, ..., \hat{p}_{hn} - p)$.

THEOREM 4.3. *If $p \geq 0.5$, the joint density for $\mathbf{V_n}$ and $\mathbf{D} = (D_1, ..., D_h)$ is the product of two densities:*

$$f_n(\mathbf{v}, \mathbf{d}) = f_n(\mathbf{v}|\mathbf{d}) f(\mathbf{d}).$$

The conditional density for $\mathbf{V_n}$ given $\mathbf{D} = \mathbf{d}$, $f_n(\mathbf{v}|\mathbf{d})$, converges to a multivariate normal density with mean 0, variances $p(1-p)/d_i$, $i = 1, ..., h$, and all cross-covariances equal to zero. The marginal density, $f(\mathbf{d})$, for $\mathbf{D}$ is a $(h-1)$-variate Dirichlet density with all parameters equal to a/p if $p > 0.5$, and $f(\mathbf{d})$ is the density from Case 1 of Theorem 3.2 if $p = 0.5$.

PROOF. The proof follows directly from Theorem 4.2 and Theorem 3.2. □

Theorem 4.4 states that for $p_1 < 0.5$ the D_i are nonrandom, and therefore the joint density is essentially the same as the conditional, but now over all treatments, $i = 1, .., K$.

THEOREM 4.4. *If $p_1 < 0.5$, the vector with components $\sqrt{n}(\hat{p}_{in} - p_i)$, $i = 1, ..., K$, is asymptotically a multivariate normal with mean zero, variances with components $p_i(1-p_i)/D_i$, and all cross-covariances equal to zero. The constants, D_i, are given in Case 2, Theorem 3.2, and are equal to*

$$D_i = \frac{\frac{1}{q_i - p_i}}{\frac{1}{q_1 - p_1} + ... + \frac{1}{q_K - p_K}}.$$

5. **Inference**

The problem of testing the hypothesis of equal treatment effects is a central problem in a clinical trial. Consider testing the following hypothesis

$$\begin{aligned} H_0 &: p_1 = p_2 = ... = p_K = p \\ H_1 &: \text{not all } p_i \text{ are equal,} \end{aligned}$$

where p is used for the common probability of success under H_0. We first construct the likelihood ratio test statistic:

$$\lambda_n = \frac{\mathcal{L}_n(\mathbf{p}_\mathbf{n}^*)}{\mathcal{L}_n(\hat{\mathbf{p}}_\mathbf{n})},$$

where $\mathbf{p}_\mathbf{n}^*$ is the maximum likelihood estimator over the null hypothesis and $\hat{\mathbf{p}}_\mathbf{n}$ is the maximum likelihood estimator over the whole parameter space. Let $S = \sum_{i=1}^{K} X_i$. Then the logarithm of the likelihood ratio can be written as

$$\begin{aligned}\log \lambda_n &= \log \mathcal{L}_n(\mathbf{p}_\mathbf{n}^*) - \log \mathcal{L}_n(\hat{\mathbf{p}}_\mathbf{n}) \\ &= n\left((S/n)\log(S/n) + (1 - S/n)\log(1 - S/n)\right) \\ &\quad - \sum_{i=1}^{K} N_i \left[(X_i/N_i)\log(X_i/N_i) + (1 - X_i/N_i)\log(1 - X_i/N_i)\right].\end{aligned}$$

The test rejects H_0 when the likelihood ratio test statistic, or its logarithm, is too small.

Theorem 5.1 states that the limiting results for $-2\log\lambda_n$ are similar to the case of independent random variables. Surprisingly, the null distribution is the same, regardless of the value of p. The theorem is an extension of Wilks' Theorem to dependent random variables. Ferguson (1996, pp. 145-147) proves Wilks' Theorem for independent random variables. The proof is very similar to Ferguson's for $p < 0.5$, except that we must employ martingale techniques (using Theorems 4.2 and 4.4). For $p \geq 0.5$, an additional conditioning argument can be used to achieve the same result.

THEOREM 5.1. *Under H_0,*

$$-2log\lambda_n \xrightarrow{D} \chi^2_{K-1}.$$

PROOF. Let us restate the null hypothesis as follows:

$$H_0 : \theta_1 = \theta_2 = ... = \theta_{K-1} = 0, 0 \leq \theta_K \leq 1,$$

where $\theta_K = p$, and $\theta_1 = p_1 - p, ..., \theta_{K-1} = p_{K-1} - p$. We proceed as in Ferguson (1996). We have $-2log\lambda_n = 2[\log \mathcal{L}_i(\hat{\boldsymbol{\theta}}_\boldsymbol{n}) - \log \mathcal{L}_i(\boldsymbol{\theta}_\boldsymbol{n}^*)]$, where $\hat{\boldsymbol{\theta}}_\boldsymbol{n}$ is the unrestricted maximum likelihood estimator, and $\boldsymbol{\theta}_\boldsymbol{n}^*$ is maximum likelihood estimator restricted to H_0. Expand $\log \mathcal{L}_n(\boldsymbol{\theta}_n^*)$ about $\hat{\boldsymbol{\theta}}_\boldsymbol{n}$:

$$\begin{aligned}&\log \mathcal{L}_n(\boldsymbol{\theta}_n^*) \\ = \;& \log \mathcal{L}_n(\hat{\boldsymbol{\theta}}_\boldsymbol{n}) + (\log \mathcal{L}_n(\hat{\boldsymbol{\theta}}_\boldsymbol{n}))'(\boldsymbol{\theta}_\boldsymbol{n}^* - \hat{\boldsymbol{\theta}}_\boldsymbol{n}) - \frac{1}{2} n(\boldsymbol{\theta}_\boldsymbol{n}^* - \hat{\boldsymbol{\theta}}_\boldsymbol{n})^T \frac{1}{n} \boldsymbol{J}_\boldsymbol{n}(\boldsymbol{\theta}_\boldsymbol{n}^+)(\boldsymbol{\theta}_\boldsymbol{n}^* - \hat{\boldsymbol{\theta}}_\boldsymbol{n}),\end{aligned}$$

where $\boldsymbol{\theta}_\boldsymbol{n}^+$ is a point on a segment connecting $\boldsymbol{\theta}_\boldsymbol{n}^*$ and $\hat{\boldsymbol{\theta}}_\boldsymbol{n}$, and $\boldsymbol{J}_\boldsymbol{n}(\boldsymbol{\theta}_\boldsymbol{n}^+)$ is a matrix of the second derivatives of the loglikelihood. Because $\boldsymbol{\theta}^\mathbf{0}$ is the true value under H_0, $\boldsymbol{\theta}_\boldsymbol{n}^+$ converges to $\boldsymbol{\theta}^\mathbf{0}$ by the Theorem 4.1 and the definition of $\boldsymbol{\theta}_\boldsymbol{n}^+$.

Case 1. $p < 0.5$

From conditions (A4), and (A5), it follows that, when $p < 0.5$,

$$-\frac{1}{n}\boldsymbol{J_n}(\boldsymbol{\theta_n^+}) \xrightarrow{P} \mathbf{G}.$$

If $\boldsymbol{\theta^{00}} = (\theta_1^0 + \theta_K^0, ..., \theta_{K-1}^0 + \theta_K^0, \theta_K^0)$, and $(D_1, ..., D_K)$ are constants from Case 2 of Theorem 3.2, with $\boldsymbol{\theta^{00}}$ as a vector of success probabilities, then $\mathbf{G}$ is given by

$$\mathbf{G} = \begin{pmatrix} \frac{D_1}{p_1(1-p_1)} & 0 & \dots & 0 & \frac{D_1}{p_1(1-p_1)} \\ 0 & \frac{D_2}{p_2(1-p_2)} & \dots & 0 & \frac{D_2}{p_2(1-p_2)} \\ & & \ddots & & \\ \frac{D_1}{p_1(1-p_1)} & \frac{D_2}{p_2(1-p_2)} & \dots & \frac{D_{K-1}}{p_{K-1}(1-p_{K-1})} & \sum_{i=1}^{K} \frac{D_i}{p_i(1-p_i)} \end{pmatrix}. \tag{13}$$

Also define the diagonal $K \times K$ matrix $\mathbf{H} = diag(0, 0, ..., 0, [\sum_{i=1}^{K} D_i/(p_i(1-p_i))]^{-1})$.

For sufficiently large n, $(\log \mathcal{L}_n(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}))' = 0$, so that

$$-2log\lambda_n \sim n(\boldsymbol{\theta_n^*} - \hat{\boldsymbol{\theta}}_{\boldsymbol{n}})^T \mathbf{G}(\boldsymbol{\theta_n^*} - \hat{\boldsymbol{\theta}}_{\boldsymbol{n}}).$$

As in Ferguson, we expand $(log\mathcal{L}_n(\boldsymbol{\theta_n^*}))'$ about $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ and $(\log \mathcal{L}_n(\boldsymbol{\theta_n^*}))'$ about $\boldsymbol{\theta^0}$ in Taylor series. Then from the proof of Theorem 4.2,

$$\frac{1}{\sqrt{n}}(\log \mathcal{L}_n(\boldsymbol{\theta^0}))' = \sqrt{n}\frac{1}{n}(\log \mathcal{L}_n(\boldsymbol{\theta^0}))' \xrightarrow{D} N(\mathbf{0}, \mathbf{G}). \tag{14}$$

Hence,

$$\frac{1}{\sqrt{n}}(\log \mathcal{L}_n(\boldsymbol{\theta_n^*}))' \xrightarrow{D} [\mathbf{I} - \mathbf{HG}]\mathbf{Y},$$

where $\mathbf{Y}$ is distributed as $N(\mathbf{0}, \mathbf{G})$, so that

$$-2log\lambda_n \xrightarrow{D} \mathbf{Y^T}\left[\mathbf{G^{-1}} - \mathbf{H}\right]\mathbf{Y} \sim \chi^2_{K-1}.$$

Case 2. $p \geq 0.5$

The proof in this case is similar to Case 1. If $p \geq 0.5$, according to Theorem 3.2 the matrix $\mathbf{G} = \mathbf{G}(\mathbf{D})$ in (13) has random entries depending on $\mathbf{D}$. In any case $\mathbf{G}(\mathbf{D})$ is invertible with probability one. Let us condition on the values $D_i = d_i$, where d_i are realizations of D_i, and assume without loss of generality that the nonrandom matrix $\mathbf{G}(\mathbf{d})$ is invertible. Replacing $\mathbf{G}$ in the previous argument by $\mathbf{G}(\mathbf{d})$ we will obtain similar conclusions. The asymptotic normality of $1/(\sqrt{n})(\log \mathcal{L}_n(\boldsymbol{\theta^0}))'$ is implied by Theorem 4.3 and

$$-2log\lambda_n \xrightarrow{D} \mathbf{Y(d)^T}\left[\mathbf{G(d)^{-1}} - \mathbf{H}\right]\mathbf{Y(d)},$$

where $\mathbf{Y}(\mathbf{d})$ is distributed $N(\mathbf{0}, \mathbf{G}(\mathbf{d}))$. The limit is independent of $\mathbf{d}$, and has a χ^2_{K-1} distribution. □

We are also interested in the distribution of the test statistic when the true value of parameter $\boldsymbol{\theta}$ does not satisfy H_0. Define $\boldsymbol{\delta} = \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta^0})$, where $\boldsymbol{\theta^0}$ is the closest value in H_0 to the true value $\boldsymbol{\theta}$. In terms of the components of the vector $\mathbf{p}$ from the initial problem $\boldsymbol{\delta} = \sqrt{n}(p_1 - p_K, ..., p_{K-1} - p_K, 0)$.

COROLLARY 5.1. *For $p < 0.5$,*

$$-2log\lambda_n \xrightarrow{D} \chi^2_{K-1}(\varphi),$$

where the noncentrality parameter φ is

$$\varphi = \sum_{i=1}^{K-1} \frac{D_i \delta_i^2}{p_i(1-p_i)} - \left(\sum_{i=1}^{K} \frac{D_i}{p_i(1-p_i)} \right)^{-1} \left(\sum_{i=1}^{K-1} \frac{D_i \delta_i}{p_i(1-p_i)} \right)^2,$$

where $\boldsymbol{\delta} = \sqrt{n}(p_1 - p_K, ..., p_{K-1} - p_K, 0)$.

PROOF. The proof is as in Theorem 5.1 but the limiting distribution of $(\frac{1}{\sqrt{n}}(\log \mathcal{L}_n(\mathbf{p^0}))'$ is different and can be found by the expansion,

$$\begin{aligned} \frac{1}{\sqrt{n}}(\log \mathcal{L}_n(\mathbf{p^0}))' &= \frac{1}{\sqrt{n}}(\log \mathcal{L}_n(\mathbf{p}))' + \frac{1}{n}\mathbf{J_n}(\mathbf{p_n^+})\sqrt{n}(\mathbf{p^0} - \mathbf{p}) \\ &\sim N(\mathbf{0}, \mathbf{G}) + \mathbf{G}\boldsymbol{\delta} = N(\mathbf{G}\boldsymbol{\delta}, \mathbf{G}) \end{aligned}$$

The rest of the proof is as in Ferguson (1996, page 148). □

6. Discussion

Our purpose in introducing the birth and death urn has been to deal with the ethical issue of assigning more patients to better treatments. We have focused on asymptotic properties of the birth and death urn design with immigration. We show how this is done in Section 2 and, in particular, there is a threshold phenomenon which occurs according to whether the best treatment has success probability less than 0.5, equal to 0.5 or greater than 0.5. If the highest success probability, p_1, is less than 0.5, the allocation proportions approach fixed positive values favoring the better treatments proportionately to $1/(q_i - p_i)$. This is similar to the behavior of play-the-winner rules. If $p_1 > 0.5$, then the proportions of allocation to any of the inferior treatments are driven to zero at a rate which is either proportional to a power of n if the inferior treatment also has $p_i > 0.5$ or to the $\log(n)$ if $p_i < 0.5$. The first case exhibits behavior similar to that of a pure birth urn [see Durham, Flournoy and Li (1998)].

The question of the feasibility of inference under such a plan is then undertaken. Asymptotic distributions for the maximum likelihood estimators are given and the likelihood ratio test for the equality of the p_i's is constructed. The expected normality and chi-square asymptotics are obtained, except that, in the supercritical case

$p_i > 0.5$, they must be conditional, since the asymptotic allocation proportions are still random variables. Interestingly, the basic Wilks' chi-square is the same in all cases; but the power under the alternative is not. Standard methods such as used by Basawa, Billard and Srinivasan (1984) for time series models are applicable to the subcritical case, directly and, with a conditioning argument, to the null distribution in the supercritical and critical cases. But the fundamental rates of convergence conditions are violated in the alternative. The existence of useful asymptotics for power in the supercritical case is still an open question. More refined alternatives corresponding to selecting the best treatment or discarding the worst treatment are also open research areas.

It is widely thought that equal allocation is always preferred in clinical trials. We find that this is not so. Suppose we have three treatments and the vector of success probabilities is $\mathbf{p} = (0.4, 0.1, 0.1)$. If the sample size is $n = 100$, then $\delta = \sqrt{100}(0.3, 0, 0)$. When equal allocation is used the noncentrality parameter is $\varphi' = 0.1053$, and the approximate power is 0.83. Under the birth and death urn design with immigration the noncentrality parameter is $\varphi = 0.1429$, and the approximate power is 0.93. Being a subcritical case, there will be a still more powerful fixed allocation. But the supercritical case presents a more intriguing possibility. Preliminary simulations suggest that when $p_1 > p_2 = p_3$, for example, the urn design appears to be more powerful than any fixed design.

Exact properties can be derived by incorporating stopping rules in the urn design, using techniques found in Durham, Flournoy, and Li (1998). The interested reader is referred to Ivanova (1998), who describes these methods and applies them with the birth and death urn.

Acknowledgment. The authors thank Professor Bimal Sinha for his encouragement and helpful comments. Sections 1-4 represent a portion of Dr. Ivanova's doctoral thesis under the direction of Professor Rosenberger.

References

Andersen, J., Faries, D. and Tamura, R. (1994). A randomized play-the-winner design for multi-arm clinical trials. *Communications in Statistics - Theory and Methods*, **23**, 309-323.

Anderson, W. J. (1991). *Continuous-Time Markov Chains.* Springer Verlag, New York.

Athreya, K. B. and Ney, P. E. (1972). *Branching Processes.* Springer Verlag, Berlin.

Basawa, I.V., Billard, L. and Srinivasan, R. (1984). Large-sample tests of homogeneity for time series models. *Biometrika*, **71**, 203-206.

Cox, D. R. and Miller, H. D. (1965). *The Theory of Stochastic Processes.* John Wiley, New York.

Durham, S. D., Flournoy, N. and Li, W. (1998). Sequential designs for maximizing the probability of a favorable response. *Canadian Journal of Statistics*, **3**, 479-495.

Durham, S. D. and Yu, C. F. (1990). Randomized play-the-leader rules for sequential sampling from two populations. *Probability in Engineering and Information Science*, **4**, 355-367.

Ferguson, T. S. (1996). *A Course in Large Sample Theory.* Chapman and Hall, New York.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Application.* Academic Press, New York.

Ivanova, A. (1998). *A Birth and Death Urn for Randomized Clinical Trials.* University of Maryland Graduate School, Baltimore (doctoral dissertation).

KARLIN, S. AND TAYLOR, H. M. (1974). *A First Course in Stochastic Processes.* Academic Press, New York.

LI, W. (1995). *Sequential Designs for Opposing Failure Functions.* American University, Washington (doctoral dissertation).

ROSENBERGER, W. F., FLOURNOY, N. AND DURHAM, S. D. (1997). Asymptotic normality of maximum likelihood estimators from multiparameter response- driven designs. *Jour. Statistical Planning and Inference*, **60**, 69-76.

SEVASTIANOV, B. A. (1971). *Branching Processes* (in Russian). Nauka, Moscow.

WEI, L. J. (1979). The generalized Pólya's urn design for sequential medical trials. *Ann. Statist.*, **7**, 291-296.

WEI, L. J. AND DURHAM, S. (1978). The randomized play-the-winner rule in medical trials. *Jour. Amer. Statist. Assoc.*, **73**, 840-843.

ANASTASIA IVANOVA
DEPARTMENT OF BIOSTATISTICS
3103-C MCGAVRAN-GREENBERG HALL, CB# 7400
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
CHAPEL HILL, NC 27599-7400 U.S.A.
E-mail: aivanova@bios.unc.edu

WILLIAM F. ROSENBERGER
DEPT. OF MATH. AND STAT.
UNIVERSITY OF MARYLAND,
BALTIMORE COUNTY
1000 HILLTOP CIRCLE
BALTIMORE, MD 21250 U.S.A.
E-mail: billr@math.umbc.edu

STEPHEN D. DURHAM
DEPARTMENT OF STATISTICS
UNIVERSITY OF SOUTH CAROLINA
COLUMBIA, SC 29208 U.S.A.

NANCY FLOURNOY
DEPT. OF MATH. AND STAT.
AMERICAN UNIVERSITY
WASHINGTON, DC 20016-8050 U.S.A.
E-mail: flournoy@american.edu