Stochastic Processes and their Applications 31 (1989) 283-305
North-Holland

SAMPLING A BRANCHING TREE

Ellen MAKI

Department of Biostatistics, Ontario Cancer Institute, 500 Sherbourne St., Toronto, Ontario, Canada, M4X 1K9

Philip McDUNNOUGH

Department of Statistics, University of Toronto, Toronto, Ontario, Canada

Received 23 April 1987
Revised 25 October 1988

A Galton-Watson branching tree is sampled, yielding a derived vector process of family sizes. Exact and asymptotic distributions for this process are derived, rates of convergence given, and the probability of selecting different families is shown to converge rapidly to one. Consistent, asymptotically normal nonparametric estimates of the underlying offspring distribution are obtained and for power series distributions approximate MLE's are shown to be asymptotically normal and efficient.

branching process * sampling * inference in stochastic processes * likelihood inference * asymptotic theory

1. Introduction

Consider N distinguishable "individuals" of the same type. These N individuals will be taken as the initial ancestors of a (Bienaymé) Galton-Watson branching tree. That is, each of the N individuals dies but leaves a number of offspring (of the same type) in an i.i.d. way. These offspring then constitute the first generation of the tree and in an identical fashion generate a second generation and so on.

A variety of observational patterns may arise in the study of a Galton-Watson tree. By far the most studied has been the Galton-Watson process $\{Z_n, n=0, 1, \ldots\}$, where Z_n denotes the total population size at generation n. If the offspring distribution is denoted by $p(\cdot|1)$, then it is well-known that $\{Z_n\}$ is a Markov process with transition probabilities $P(Z_n=i|Z_{n-1}=j)=p(i|j)$, where $p(\cdot|j)$ is the j-fold convolution of $p(\cdot|1)$.

The problem of estimating the mean, μ, and variance, σ^2, of $p(\cdot|1)$, in a nonparametric context, has been considered by many authors (cf. Dion and Keiding (1978)). Consistent estimates have been provided and with the exception of the unresolved question of conditional versus unconditional inference (cf. Feigin and Reiser (1979), Sweeting (1986)), most of the problems seem to have been resolved, at least from the asymptotic classical view. An interesting and important result in support of this

0304-4149/89/$3.50

statement, due to Lockhart (1982), is that under the assumption of the existence of the third moment of $p(\cdot|1)$, only the first two moments can be consistently estimated.

The above discussion refers to nonparametric inference about μ and σ^2 based on observing the process $\{Z_n\}$. In this context it is of interest to note that Harris (1948) obtained an MLE for μ on the basis of (a) observing $\{Z_n\}$ and (b) maximizing a likelihood based on the (unobserved) underlying tree. Dion (1974) and Feigin (1977) independently showed that the Harris estimate was identical to an MLE obtained from a likelihood based solely on (a). It is clear that complete observation of the underlying branching tree leads to consistent estimation of the whole offspring distribution. Intermediate situations between observation of the underlying tree and recording the total population size Z_n seem not to have been considered from a statistical point of view. It is the purpose of this paper to provide such an analysis in cases based on recording family (i.e. all those individuals having the same parent) sizes of individuals selected at random (i.e. uniformly) from the population. In particular, we consider incomplete observational schemes which occur when a random sample of individuals in a generation or successive ones is taken and their family sizes are recorded. Only processes with $p(0|1)=0$ and $\mu>1$ are considered (which restricts us to supercritical ones). We derive the exact (when two individuals are sampled) and limiting (for 'not too rapidly' increasing sample sizes) joint distributions of the sampled individuals' family sizes. In addition, some rates of convergence are given and conditions are found under which the probability that all sampled individuals belong to different families converges to one. These results are then used in order to obtain the asymptotic behaviour of the non-Markovian process $\{\boldsymbol{X}_n\}$, where $\boldsymbol{X}_n$ denotes the family sizes of r_n randomly chosen individuals from generation n. From this we are then able to present various inference techniques, with respect to $p(\cdot|1)$ (in a nonparametric context) based on observing the process $\{\boldsymbol{X}_n\}$. Parametric inference for the case of power series offspring distributions is considered. In this case, the 'MLE' of n (based on an approximate likelihood) is again shown to be consistent and asymptotically normal. In addition, the 'MLE' is shown to be efficient and a comparison is made with a method of moments type estimator. This theoretical comparison is then supplemented via a Monte Carlo experiment involving a modified geometric distribution. Also, the asymptotic behaviour of certain approximate weighted likelihood procedures is estimated and extensions to more general parametric families of offspring distributions are indicated.

Finally, it should be noted that our study of randomly sampled branching trees is not solely motivated by statistical questions. Indeed, the random selection of individuals in an evolving population occurs quite naturally in demography and certain biological processes. For example, Joffe and Waugh (1982), in considering a kin number problem, have studied the case where a single individual is randomly sampled from some generation of a supercritical Galton–Watson tree. They have obtained both exact and limiting distributions of the generations sizes in the whole family tree of this one particular individual.

2. Sampling two individuals in a generation

Consider a Galton–Watson branching tree starting from N individuals and having offspring distribution $p(\cdot\,|\,1)$. Let $\{Y_n^{(k)}\}$ be i.i.d. random variables with $Y_n^{(k)}$ denoting the number of offspring of the kth individual in generation n. Now pick two individuals at random from the nth generation and denote their respective family sizes by $X_n^{(1)}$ and $X_n^{(2)}$. We now proceed to calculate the joint distribution of $X_n^{(1)}$ and $X_n^{(2)}$.

We first note that

$$\begin{aligned} P(X_n^{(1)}=i, X_n^{(2)}=j) &= E[P(X_n^{(1)}=i, X_n^{(2)}=j\,|\,Z_{n-1})] \\ &= \sum_N P(X_n^{(1)}=i, X_n^{(2)}=j\,|\,Z_{n-1}=N)P(Z_{n-1}=N) \\ &= \sum_N P(X_1^{(1)}=i, X_1^{(2)}=j)P(Z_{n-1}=N). \end{aligned} \tag{1}$$

In the above Z_n denotes the size of the nth generation.

Now, to evaluate $P(X_1^{(1)}=i,\ X_1^{(2)}=j)$ we consider the three cases $i=j=1$, $i\neq j$, and $i=j>1$. First, note that there are N families in the first generation of sizes $Y_0^{(1)}, Y_0^{(2)}, \ldots, Y_0^{(N)}$ respectively. So

$$\begin{aligned} P(X_1^{(1)}=1, X_1^{(2)}=1) &= P(\text{two different families of size 1 are selected}) \\ &= \sum_{s<t}^{N} P(\text{families } s \text{ and } t \text{ are selected and are of size 1}) \\ &= \sum_{s<t}^{N} P(\text{families } s \text{ and } t \text{ are selected}\,|\,\text{they are of size 1}) \\ &\qquad \times P(\text{families } s \text{ and } t \text{ are of size 1}). \end{aligned}$$

Clearly $P(\text{families } s \text{ and } t \text{ are of size 1}) = p(1\,|\,1)^2$ while by symmetry

$$\begin{aligned} &P(\text{families } s \text{ and } t \text{ are selected}\,|\,\text{they are of size 1}) \\ &\quad = P(\text{families 1 and 2 are selected}\,|\,\text{they are of size 1}) \end{aligned}$$

and hence

$$\begin{aligned} &P(X_1^{(1)}=1, 2X_1^{(2)}=1) \\ &\quad = \binom{N}{2} p(1\,|\,1)^2 P(\text{families 1 and 2 are selected}\,|\,\text{they are of size 1}) \\ &\quad = \binom{N}{2} p(1\,|\,1)^2 \\ &\quad \times E[P(\text{families 1 and 2 are selected}\,|\,\text{they are of size 1})\,|\,Y_0^{(1)}, \ldots, Y_0^{(N)}] \\ &\quad = \binom{N}{2} p(1\,|\,1)^2 E\left[\left(\frac{2+Y_0^{(3)}+\cdots+Y_0^{(N)}}{2}\right)^{-1}\right]. \end{aligned} \tag{2}$$

Notice, incidentally, that

$$E\left[\binom{2+Y_0^{(3)}+\cdots+Y_0^{(N)}}{2}^{-1}\right]$$

$$=\sum_{k_0^{(3)},\ldots,k_0^{(N)}}\left[\binom{2+k_0^{(3)}+\cdots+k_0^{(N)}}{2}^{-1} P(Y_0^{(3)}=k_0^{(3)})\cdots P(Y_0^{(N)}=k_0^{(N)})\right]$$

$$=\sum_{k_0^{(3)},\ldots,k_0^{(N)}}\left[\binom{2k_0^{(3)}+\cdots+k_0^{(N)}}{2}^{-1} p(k_0^{(3)}\,|\,1)\cdots p(k_0^{(N)}\,|\,1)\right].$$

For $i\neq j$ a similar argument to that which led to (2) yields

$$P(X_1^{(1)}=i, X_1^{(2)}=j)=\left[\binom{N}{2}p(i\,|\,1)p(j\,|\,1)\right]E\left[ij\binom{i+j+Y_0^{(3)}+\cdots+Y_0^{(N)}}{2}^{-1}\right]. \quad (3)$$

When considering the case $i=j>1$ we must take into account the possibility of both individuals belonging to the same family. The number of ways that this can happen is of course N and the probability that the common parent produces i offspring is $p(i\,|\,1)$. Hence, splitting the expectation into a part corresponding to different families and another corresponding to identical ones yields

$$P(X_1^{(1)}=i, X_1^{(2)}=i)$$

$$=\left[\binom{N}{2}p(i\,|\,1)^2\right]E\left[i^2\binom{2i+Y_0^{(3)}+\cdots+Y_0^{(N)}}{2}^{-1}\right]$$

$$+Np(i\,|\,1)E\left[\binom{i}{2}\binom{i+Y_0^{(2)}+\cdots+Y_0^{(N)}}{2}^{-1}\right]. \quad (4)$$

In order to obtain $P(X_n^{(1)}=i,\ X_n^{(2)}=j)$ we use (1), which is just

$$P(X_n^{(1)}=i,\ X_n^{(2)}=j)=E[P(X_1^{(1)}=i,\ X_1^{(2)}=j\,|\,Z_0')]$$

where $Z_0' \overset{d}{=} Z_{n-1}$ is independent of $Y_n^{(k)}$, $n\geqslant 1$, along with (2), (3), (4) replacing N by Z_0'. This can be equivalently summarized as

$$P(X_n^{(1)}=i, X_n^{(2)}=j)$$

$$=p(1\,|\,1)^2E\left[\binom{Z_{n-1}}{2}\binom{2+Z_{n-1}-Y_{n-1}^{(1)}-Y_{n-1}(2)}{2}^{-1}\right], \quad i=j=1$$

$$=ijp(i\,|\,1)p(j\,|\,1)E\left[\binom{Z_{n-1}}{2}\binom{i+j+Z_n-Y_{n-1}^{(1)}-Y_{n-1}^{(2)}}{2}^{-1}\right], \quad i\neq j$$

$$=i^2p(i\,|\,1)^2E\left[\binom{Z_{n-1}}{2}\binom{2i+Z_n-Y_{n-1}^{(1)}-Y_{n-1}^{(2)}}{2}^{-1}\right]$$

$$+\binom{i}{2}E\left[Z_{n-1}\binom{i+Z_n-Y_{n-1}^{(1)}}{2}^{-1}\right], \quad i=j>1. \quad (5)$$

Of course, as before, each expectation in (5) may be replaced by an infinite sum involving only the offspring distribution.

Now consider the following heuristic approach to the behaviour of (5) for large n (or equivalently for large N). It is well-known that Z_n grows exponentially so that the terms within the first three expectations behave as

$$Z_{n-1}^2 Z_n^{-2} \to \mu^{-2} \quad \text{a.s.,} \quad \text{as } n \to \infty,$$

while the last expectation tends to 0. Formally we then have

$$P(X_n^{(1)} = i,\ X_n^{(2)} = j) \to [ip(i\,|\,1)][jp(j\,|\,1)]/\mu^2, \tag{6}$$

so that $X_n^{(1)}$, $X_n^{(2)}$ are approximately i.i.d. with marginals as found in Joffe and Waugh (1982). This informal argument will be made rigorous in the next section. We do note, however, that the methods used in determining (5) can be extended to the sampling of more than two individuals (at a considerable cost in notational complexity). Moreover, the resulting family sizes can still be shown to be asymptotically i.i.d. with marginals as in (6); for a rigorous statement and proof see Section 4. As we shall see, this result is of considerable importance from a statistical viewpoint.

3. The probability of selecting different families

When sampling $r_n > 2$ individuals from generation n, we noted that the joint distribution of family sizes becomes increasingly complex. This is simply due to the fact that we must take into account the possibility of different individuals coming from the same family. We are led, therefore, to consider the event

$$D_n = \{\text{all } r_n \text{ selected individuals belong to different families}\}.$$

In this section we will show that $P(D_n) \to 1$ as $n \to \infty$, the convergence being rapid. We also demonstrate similar results if r_n is not necessarily constant but may increase 'not too rapidly' with n.

Consider first $P(D_1)$ with $r_1 = 2$. Then

$$\begin{aligned} P(D_1) &= \sum_{s<t} P(\text{families } s \text{ and } t \text{ are selected}) \\ &= \binom{N}{2} P(\text{families 1 and 2 are selected}) \\ &= \binom{N}{2} E\left[Y_0^{(1)} Y_0^{(2)} \binom{Z_1}{2}^{-1} \right]. \end{aligned}$$

Now, as in the argument leading to (5), condition on Z_{n-1} to obtain

$$P(D_n) = E\left[Y_{n-1}^{(1)} Y_{n-1}^{(2)} \binom{Z_{n-1}}{r_n} \binom{Z_n}{r_n}^{-1} \right].$$

An identical argument also yields the extension

$$P(D_n) = E\left[Y_{n-1}^{(1)} \cdots Y_{n-1}^{(r_n)} \binom{Z_{n-1}}{r_n} \binom{Z_n}{r_n}^{-1} \right]. \tag{7}$$

The behaviour of $P(D_n)$ for large n is extremely important. An answer is provided by the following result.

Theorem 1. *Let $p(\cdot|1)$ have finite mean and variance, μ and σ^2, respectively. Assume*

$$r_n^2 \mu^{-n} \to 0. \tag{8}$$

Then, for $0 \leq \alpha < -\ln[E(1/Y)]/\ln[\mu]$, where $Y \sim p(\cdot|1)$, we have
(a) $P(D_n) \to 1$ *as* $n \to \infty$,
(b) $\mu^{\alpha n} r_n^{-2}(1 - P(D_n)) \to 0$ *as* $n \to \infty$.

We will require the following Lemma in the proof of Theorem 1.

Lemma 1. *For $0 \leq \alpha < -\ln[E(1/Y)]/\ln \mu$, $\mu^{\alpha n} E(1/Z_n) \to 0$, $n \to \infty$.*

Proof. By using the inequality $a_1 + \cdots + a_k \geq k(a_1 \cdots a_k)^{1/k}$ for $a_i \geq 0$, $i = 1, \ldots, k$, we have

$$E[Z_n^{-1} | Z_{n-1}] \leq Z_{n-1}^{-1} \{E[Y^{-1/Z_{n-1}} | Z_{n-1}]\}^{Z_{n-1}} \leq Z_{n-1}^{-1} E(1/Y),$$

for $Y \sim p(\cdot|1)$ independent of Z_n, Z_{n-1}. Hence

$$E(\mu^{\alpha n} Z_n^{-1}) \leq E(\mu^{\alpha(n-1)} Z_{n-1}^{-1}) \mu^{\alpha} E(1/Y),$$

and so by iteration,

$$E(\mu^{\alpha n} Z_n^{-1}) \leq E(\mu^{\alpha} Z_1^{-1})[\mu^{\alpha} E(1/Y)]^{n-1} \to 0.$$

This last statement is a consequence of

$$0 \leq \alpha < -[\ln E(1/Y)]/\ln \mu \quad \text{implies } \mu^{\alpha} E(1/Y) < 1.$$

Proof of Theorem 1. Consider first the case where $p(\cdot|1)$ is of bounded support, say $\{k_1, k_1+1, \ldots, k_2\}$. Now, for degenerate $p(\cdot|1)$ it is easily shown that the probability of selecting r_n different families of size k when $p(k|1) = 1$ is greater than the probability of selecting r_n different families of size $k+1$ when $p(k+1|1) = 1$. Therefore

$$k_2^{r_n} \binom{Z_{n-1}}{r_n} \binom{k_2 Z_{n-1}}{r_n}^{-1} \leq P(D_n | Z_{n-1}) \leq k_1^{r_n} \binom{Z_{n-1}}{r_n} \binom{k_1 Z_{n-1}}{r_n}^{-1} \leq 1. \tag{9}$$

(Notice that the LHS of (9) is simply the probability of selecting r_n different families from the first generation when the initial population size is Z_{n-1} and the offspring distribution degenerates at k_2. A similar interpretation holds for the third term of (9).) Now

$$Z_nE(Z_n)^{-1}=Z_nN^{-1}\mu^{-n}\xrightarrow{\text{a.s.}} W\overset{\text{a.s.}}{>}0$$

so that $r_n^2/\mu^n\to 0$ implies $r_n^2/Z_n\xrightarrow{\text{a.s.}}0$ and since

$$\begin{aligned}k_2^{r_n}\binom{Z_{n-1}}{r_n}\binom{k_2Z_{n-1}}{r_n}^{-1}&\geq k_2^{r_n}(Z_{n-1}-r_n+1)^{r_n}(k_2Z_{n-1})^{-r_n}\\&=\{1-(r_n-1)/Z_{n-1}\}^{r_n}\\&\geq 1-r_n(r_n-1)/Z_{n-1}\end{aligned}$$

we have

$$k_2^{r_n}\binom{Z_{n-1}}{r_n}\binom{k_2Z_{n-1}}{r_n}^{-1}\to 1\quad\text{a.s.}$$

Now use (9) to get $P(D_n\,|\,Z_{n-1})\to 1$ a.s. and the Dominated Convergence Theorem to conclude that $P(D_n)=E[P(D_n\,|\,Z_{n-1})]\to 1$.

For the case of offspring distributions of (possibly) unbounded support consider the event

$$G_n=\{Y_n^{(1)}\leq Z_{n-1},\ldots,Y_n^{(Z_{n-1})}\leq Z_{n-1}\}.$$

If we denote the second moment of $p(\cdot\,|\,1)$ by μ_2 then

$$P(G_n\,|\,Z_{n-1})\geq(1-\mu_2Z_{n-1}^{-2})^{Z_{n-1}}\to 1\quad\text{a.s.}$$

so that $P(G_n)\to 1$.

Now,

$$\begin{aligned}P(D_n)&\geq P(D_n\,|\,G_n)P(G_n)\\&\geq E\left[Z_{n-1}^{r_n}\binom{Z_{n-1}}{r_n}\binom{Z_{n-1}^2}{r_n}^{-1}\right]E\left[\left(1-\frac{\mu_2}{Z_{n-1}^2}\right)^{Z_{n-1}}\right]\\&\geq E\left[Z_{n-1}^{r_n}\binom{Z_{n-1}}{r_n}\binom{Z_{n-1}^2}{r_n}^{-1}\right]E\left[1-\frac{\mu_2}{Z_{n-1}}\right].\end{aligned}$$

Using $F_n=\binom{Z_{n-1}}{r_n}\binom{Z_{n-1}}{r_n}^{-1}$, we have

$$\begin{aligned}\mu^{\alpha n}r_n^{-2}[1-P(D_n)]&\leq\mu^{\alpha n}r_n^{-2}\{1-E[Z_{n-1}^{r_n}F_n]\}+\mu^{\alpha n}r_n^{-2}\{E[Z_{n-1}^{r_n}F_n]E(\mu_2Z_{n-1}^{-1})\}\\&\leq\mu^{\alpha n}r_n^{-2}\{E[1-Z_{n-1}^{r_n}F_n]+E[\mu_2Z_{n-1}^{-1}]\}\\&\leq\mu^{\alpha n}r_n^{-2}\{E[(r_n-1)^2Z_{n-1}^{-1}(1-(r_n-1)^2Z_{n-1}^{-2})]+E[\mu_2Z_{n-1}^{-1}]\}\\&\leq\mu^{\alpha n}r_n^{-2}\{E[r_n^2Z_{n-1}^{-1}]+E[\mu_2Z_{n-1}^{-1}]\}\\&\to 0\end{aligned}\tag{10}$$

by Lemma 1 thus proving (b). To see (a) we note from (10) that

$$1-P(D_n)\leq E[r_n^2Z_{n-1}^{-1}]+E[\mu_2Z_{n-1}^{-1}]$$
$$=(r_n^2\mu^{-n})\mu^nE(Z_{n-1}^{-1})+E(\mu_2Z_{n-1}^{-1})$$
$$\to 0,$$

again by Lemma 1 (along with (8)). This completes the proof of Theorem 1.

4. Asymptotic behaviour of family sizes

In the previous section we saw that the probability of picking r_n different families rapidly approaches one, exponentially fast for constant r_n. This suggests approximating the distribution of $\boldsymbol{X}_n\equiv(X_n^{(1)},\ldots,X_n^{(r_n)})'$ by the joint distribution of $\boldsymbol{X}_n$ and I_{D_n}, where I_{D_n} denotes the indicator function of the event of selecting r_n different families.

Thus, arguing along the lines of the derivations of Section 2 we see that

$$P(\boldsymbol{X}_n=\boldsymbol{i}_n,\ I_{D_n}=1)$$
$$=\left[\prod_{j=1}^{r_n} i_n^{(j)}p(i_n^{(j)}\,|\,1)\right]E\left[\binom{Z_{n-1}}{r_n}\binom{\boldsymbol{i}_n'\mathbf{1}+Y_{n-1}^{(r_n+1)}+\cdots+Y_{n-1}^{(Z_{n-1})}}{r_n}^{-1}\right],\tag{11}$$

where $\boldsymbol{i}_n'=(i_n^{(1)},\ldots,i_n^{(r_n)})$ and $\mathbf{1}'=(1,\ldots,1)$.

Set $\boldsymbol{t}_{r_n}'=(t_1,\ldots,t_{r_n})$ and denote the characteristic function of $\boldsymbol{X}_n$ by $c_n(\boldsymbol{t}_{r_n})$. In addition, let $c(\boldsymbol{t}_{r_n})$ denote the characteristic function of r_n i.i.d. random variables each having probability function $ip(i\,|\,1)/\mu$. The main result of this section deals with the convergence of c_n to c and is given in the following Theorem.

Theorem 2. *Under the assumptions of Theorem* 1,

$$\lim_{n\to\infty}|c_n(\boldsymbol{t}_{r_n})-c(\boldsymbol{t}_{r_n})|=0.$$

Remarks. From Theorem 2 we see that any fixed subset of r, say, coordinates of $\boldsymbol{X}_n$ are asymptotically i.i.d. with marginal probability functions $ip(i\,|\,1)/\mu$. Moreover, since all r.v.'s are positive integer valued the corresponding probability functions converge. A rate of convergence of these probability functions is given in Theorem 3.

Proof of Theorem 2. We have $c_n(\boldsymbol{t}_{r_n})=E[\exp\{\mathrm{i}\boldsymbol{t}_{r_n}'\boldsymbol{X}_n\}]$ while

$$c(\boldsymbol{t}_{r_n})=E[\mu^{-r_n}Y_{n-1}^{(1)}\cdots Y_{n-1}^{(r_n)}\exp\{\mathrm{i}\boldsymbol{t}_{r_n}'\boldsymbol{Y}_{n-1}\}],$$

where $\boldsymbol{Y}_{n-1}'=(Y_{n-1}^{(1)},\ldots,Y_{n-1}^{(r_n)})$. We now approximate $c_n(\boldsymbol{t}_{r_n})$ by $c_n^a(\boldsymbol{t}_{r_n})$ defined by

$$c_n^a(\boldsymbol{t}_{r_n})=E[I_{D_n}\exp\{\mathrm{i}\boldsymbol{t}_{r_n}'\boldsymbol{X}_n\}].$$

(Recall that $P(D_n) \approx 1$ so that I_{D_n} will 'usually' equal 1.) Using (11), a straightforward calculation yields

$$c_n^a(t_{r_n}) = E\left[Y_{n-1}^{(1)} \cdots Y_{n-1}^{(r_n)} \binom{Z_{n-1}}{r_n} \binom{Z_n}{r_n}^{-1} \exp\{it'_{r_n} Y_{n-1}\} \right].$$

Now consider

$$\lim_{n\to\infty} |c_n(t_{r_n}) - c(t_{r_n})| \le \lim_{n\to\infty} |c_n(t_{r_n}) - c_n^a(t_{r_n})| + \lim_{n\to\infty} |c_n^a(t_{r_n}) - c(t_{r_n})|.$$

The first term in the RHS is simply

$$\lim_{n\to\infty} |E[(1-I_{D_n}) \exp\{it'_{r_n} X_n\}]| \le \lim_{n\to\infty} E(1-I_{D_n}) = \lim_{n\to\infty}[1-P(D_n)] = 0,$$

while the second term is

$$\lim_{n\to\infty} \left| E\left[Y_{n-1}^{(1)} \cdots Y_{n-1}^{(r_n)} \left\{ \mu^{-r_n} - \binom{Z_{n-1}}{r_n}\binom{Z_n}{r_n}^{-1} \right\} \exp\{it'_{r_n} Y_{n-1}\} \right] \right|$$

$$\le \lim_{n\to\infty} E\left(Y_{n-1}^{(1)} \cdots Y_{n-1}^{(r_n)} \left| \binom{Z_{n-1}}{r_n}\binom{Z_n}{r_n}^{-1} - \mu^{-r_n} \right| \right)$$

$$= \lim_{n\to\infty} E\left((C_n \mu^{r_n})^{-1} \sum_{j=1}^{C_n} P_j \left| \mu^{r_n} \binom{Z_{n-1}}{r_n}\binom{Z_n}{r_n}^{-1} - 1 \right| \right)$$

where we have used

$$C_n = \binom{Z_{n-1}}{r_n},$$

and the P_j are the C_n r.v.'s of the form $Y_{n-1}^{(j_1)} Y_{n-1}^{(j_2)} \cdots Y_{n-1}^{(j_{r_n})}$. We will show that $(C_n\mu^{r_n})^{-1} \sum_{j=1}^{C_n} P_j \xrightarrow{P} 1$ and

$$\left| \mu^{r_n} \binom{Z_{n-1}}{r_n}\binom{Z_n}{r_n}^{-1} - 1 \right| \xrightarrow{P} 0,$$

and therefore that their product converges to zero in probability. Denote the product by V_n, and define A_n to be the event $A_n = \{|V_n| > \varepsilon\}$. Now,

$$0 \le V_n \le \sum_{j=1}^{C_n} P_j \left(\binom{Z_n}{r_n} + C_n \mu^{r_n} \right)^{-1} \le 1 + \sum_{j=1}^{C_n} P_j (C_n \mu^{r_n})^{-1}.$$

Also,

$$E(V_n) = P(A_n)E(V_n | A_n) + P(A_n^c)E(V_n | A_n^c)$$

$$\le P(A_n)E\left[1 + (C_n\mu^{r_n})^{-1} \sum_{j=1}^{C_n} P_j\right] \varepsilon P(A_n^c)$$

$$\to \varepsilon.$$

Since ε is arbitrary, we have $E(V_n) \to 0$ as $n \to \infty$.

To show that $(C_n\mu^{r_n})^{-1}\sum_{j=1}^{C_n} P_j \xrightarrow{P} 1$ we calculate its expected value and variance to be 1 and

$$E(C_n^{-1})\left[\left(1+\frac{\sigma^2}{\mu^2}\right)^{r_n}-1\right]+2\sum_{j=0}^{r_n-1} E\left(C_n\binom{r_n}{j}\binom{Z_{n-1}-r_n}{r_n-j}\right)$$
$$\times[(\sigma^2+\mu^2)^j\mu^{2r_n-j}-\mu^{2r_n}]$$

respectively. It is easily verified that both terms in the variance go to zero.

Finally, to show that

$$\left|\mu^{r_n}\binom{Z_{n-1}}{r_n}\binom{Z_n}{r_n}^{-1}-1\right|\xrightarrow{P}0$$

notice that

$$\left|\mu^{r_n}\binom{Z_{n-1}}{r_n}\binom{Z_n}{r_n}^{-1}-1\right|$$
$$\leqslant\left|\mu^{r_n}\binom{Z_{n-1}}{r_n}\binom{Z_n}{r_n}^{-1}-(\mu Z_{n-1})^{r_n}Z_n^{-r_n}\right|+|(\mu Z_{n-1})^{r_n}Z_{n-1}^{-r_n}-1|.$$

The first term on the right-hand side can be expressed as

$$\left(\frac{\mu Z_{n-1}}{Z_n}\right)^{r_n}\left|\frac{\prod_{j=1}^{r_{n-1}}\left[1-\dfrac{j}{Z_{n-1}}\right]}{\prod_{j=1}^{r_{n-1}}\left[1-\dfrac{j}{Z_n}\right]}-1\right|.$$

Clearly the quantity in absolute values $\xrightarrow{\text{a.s.}} 0$ if $r_n^2/Z_{n-1}\xrightarrow{\text{a.s.}} 0$. Thus it remains to show that

$$\left|\left(\frac{\mu Z_{n-1}}{Z_n}\right)^{r_n}-1\right|\xrightarrow{P}0.$$

Now,

$$P\left(\left|\left(\frac{\mu Z_{n-1}}{Z_n}\right)^{r_n}-1\right|<\varepsilon\right)\geqslant P\left(\left|\frac{Z_n}{\mu Z_{n-1}}-1\right|<\{1-(1+\varepsilon)^{-1/r_n}\}\right)$$
$$\geqslant 1-\operatorname{Var}\left(\frac{Z_n}{\mu Z_{n-1}}\right)\{1-(1+\varepsilon)^{-1/r_n}\}^{-2}$$
$$=1-\frac{\sigma^2}{\mu^2}E(Z_{n-1}^{-1}\{1-(1+\varepsilon)^{-1/r_n}\}^{-2}).$$

Using a Taylor series expansion, we find that

$$Z_{n-1}\{1-(1+\varepsilon)^{-1/r_n}\}^2=Z_{n-1}\varepsilon^2 r_n^{-2}[1-\varepsilon/2(1+r_n^{-1})(1+z)^{-(1/r_n+2)}]^2$$

where $|z|<|\varepsilon|$. Clearly this is bounded below by $\varepsilon^2[1-\varepsilon/2(1+z)^{-3}]^2$ and converges almost surely to 0 if $r_n^2/Z_{n-1} \xrightarrow{\text{a.s.}} 0$. Thus

$$\left|\mu^{r_n}\binom{Z_{n-1}}{r_n}\binom{Z_n}{r_n}^{-1}-1\right| \xrightarrow{\text{a.s.}} 0 \quad \text{as } n\to\infty,$$

and this completes the proof.

Theorem 3. *Assume the conditions of Theorem* 1 *hold. Then, for any fixed subset of r coordinates of* $\mathbf{X}_n$, $X_n^{(1)},\ldots,X_n^{(r)}$, *say, and any* $\alpha\in[0,-\ln[E(1/Y)]/\ln\mu)$, *we have*

$$\lim_{n\to\infty}\frac{\mu^{n\alpha}}{r_n^2}\left\{P(X_n^{(1)}=i_1,\ldots,X_n^{(r)}=i_r)-\prod_{j=1}^{r}\frac{i_jp(i_j\,|\,1)}{\mu}\right\}=0.$$

Proof. Consider the case where $r=1$ and notice that

$$\begin{aligned}\frac{\mu^{n\alpha}}{r_n^2}&\left\{P(X_n^{(1)}=i)-\frac{ip(i\,|\,1)}{\mu}\right\}\\ &=\frac{\mu^{n\alpha}}{r_n^2}\left\{P(X_n^{(1)}=i)-P(X_n^{(1)}=i,\ I_{D_n}=1)\right\}\\ &\quad+\frac{\mu^{n\alpha}}{r_n^2}\left\{P(X_n^{(1)}=i,\ I_{D_n}=1)-\frac{ip(i\,|\,1)}{\mu}\right\}.\end{aligned}\tag{12}$$

Since the first term in (12) is greater than or equal to zero and $\leqslant\mu^{n\alpha}r_n^{-2}\{1-P(D_n)\}$, it converges to zero by Theorem 1.

Consider the second term in (12). Summing over all i we have

$$\frac{\mu^{n\alpha}}{r_n^2}\sum_{i=1}^{\infty}\left\{P(X_n^{(1)}=i,I_{D_n}=1)-\frac{ip(i\,|\,1)}{\mu}\right\}=\mu^{n\alpha}r_n^{-2}(P(D_n)-1)$$
$$\to 0.$$

Thus, if we can show that $(\mu^{n\alpha}/r_n^2)\{P(X_n^{(1)}=i,I_{D_n}=1)-ip(i\,|\,1)/\mu\}$ has the same limit for all values of i, then this limit must be zero.

Let

$$U_{n,i}=i+\sum_{j=2}^{Z_{n-1}}Y_{n-1}^{(j)}\quad\text{and}\quad b_{n,i}=E\left(Y_{n-1}^{(2)}\cdots Y_{n-1}^{(r_n)}\binom{Z_{n-1}}{r_n}\binom{U_{n,i}}{r_n}^{-1}\right).$$

We need to show that $(\mu^{n\alpha}/r_n^2)\{b_{n,i}-b_{n,i+1}\}\to 0$. Now,

$$\begin{aligned}\mu^{n\alpha}r_n^{-2}\{b_{n,i}-b_{n,i+1}\}&=\frac{\mu^{n\alpha}}{r_n^2}E\left(Y_{n-1}^{(2)}\cdots Y_{n-1}^{(r_n)}\binom{Z_{n-1}}{r_n}\left[\binom{U_{n,i}}{r_n}^{-1}-\binom{U_{n,i+1}}{r_n}^{-1}\right]\right)\\ &=\frac{\mu^{n\alpha}}{r_n^2}E\left(\frac{1}{C_n'}\sum_{k=1}^{C_n'}P_k'\binom{Z_{n-1}}{r_n}\left[\binom{U_{n,i}}{r_n}^{-1}-\binom{U_{n,i+1}}{r_n}^{-1}\right]\right)\end{aligned}$$

where $C'_n = \binom{Z_{n-1}-1}{r_n-1}$ and the P'_k's are products of the form $Y^{(2)}_{n-1} \cdots Y^{(r_n)}_{n-1}$. (Note: none of them includes $Y^{(1)}_{n-1}$.) We can write

$$
\begin{aligned}
0 &\leq \frac{\mu^{n\alpha}}{r_n^2} \frac{1}{C'_n} \sum_{k=1}^{C'_n} P'_k \binom{Z_{n-1}}{r_n} \left[\binom{U_{n,i}}{r_n}^{-1} - \binom{U_{n,i+1}}{r_n}^{-1} \right] \\
&= \frac{\mu^{n\alpha}}{r_n^2} \frac{Z_{n-1}}{U_{n,0}+i+1} \sum_{k=1}^{C'_n} P'_k \binom{U_{n,i}}{r_n}^{-1} \\
&\leq \frac{\mu^{n\alpha}}{r_n^2} \sum_{1}^{C'_n} P'_k \binom{U_{n,i}}{r_n}^{-1} \\
&\leq \frac{\mu^{n\alpha}}{r_n^2} \sum_{1}^{C'_n} P'_k \binom{U_{n,0}}{r_n}^{-1} \\
&= \frac{\mu^{n\alpha}}{r_n^2} \frac{r_n}{(U_{n,0}-r_n+1)} \sum_{1}^{C'_n} P'_k \binom{U_{n,0}}{r_n-1}^{-1} \\
&\leq \mu^{n\alpha} r_n^{-1} (U_{n,0}-r_n+1)^{-1}.
\end{aligned}
$$

By Lemma 1, the expected value of this last r.v. $\to 0$. Thus, we have $\mu^{n\alpha} r_n^{-2}\{b_{n,i} - b_{n,i+1}\} \to 0$ and consequently $\mu^{n\alpha} r_n^{-2}\{b_{n,i} - \mu^{-1}\}$ has the same limit for all i.

This completes the proof for $r = 1$. The proof for $r \geq 2$ is similar to that given for $r = 1$ and is omitted.

5. Sampling from consecutive generations

A more complete picture of the underlying branching tree is obtained when family sizes of individuals, selected at random in each of $T+1$ consecutive generations, are recorded. In an obvious notation, this yields data $\boldsymbol{X}_n, \ldots, \boldsymbol{X}_{n+T}$ of sizes $r_n, \ldots, r_{n+T}$ respectively. In this section we will state conditions under which for large n $\boldsymbol{X}_n, \ldots, \boldsymbol{X}_{n+T}$ are 'approximately independent' with 'asymptotically i.i.d.' components. As before these results will be seen to hold when r_n increases 'not too rapidly' and also, under conditions, for certain increasing T_n. As the techniques used in providing the following theorems are similar to those used in previous sections, we shall simply provide the proof of the first and state the others. For convenience, set

$$
\boldsymbol{X}'_{n,T} = (\boldsymbol{X}'_n, \ldots, \boldsymbol{X}'_{n+T}), \qquad \boldsymbol{t}'_{n,T} = (\boldsymbol{t}'_{r_n}, \ldots, \boldsymbol{t}'_{r_{n+T}}),
$$

where $\boldsymbol{t}'_{r_n} = (t_1, \ldots, t_{r_n})$, $\boldsymbol{t}'_{r_{n+1}} = (t_{r_n+1}, \ldots, t_{r_n+r_{n+1}})$, etc. ..., and denote the event of selecting $r_n + r_{n+1} + \cdots + r_{n+T}$ different families by $D_{n,T}$.

Theorem 4. *Under the assumptions of Theorem* 1,

(a) $P(D_{n,T}) \to 1$ *as* $n \to \infty$.

(b) $\mu^{\alpha(n+T)} r_{n+T}^{-2}[1 - P(D_{n,T})] \to 0$, *for any* $0 \leq \alpha < -\ln[E(1/Y)]/\ln \mu$.

Proof of Theorem 4. Since $D_{n,T}=\bigcap_{i=n}^{n+T} D_i$,

$$1-P(D_{n,T})\le \sum_{j=0}^{T}[1-P(D_{n+j})]\to 0 \quad \text{as } n\to\infty,$$

by Theorem 1. Also

$$\begin{aligned}0&\le \mu^{\alpha(n+T)}r_{n+T}^{-2}[1-P(D_{n,T})]\\ &\le \sum_{j=0}^{T}\mu^{\alpha(T-j)}\mu^{\alpha(n+j)}r_{n+j}^{-2}(1-P(D_{n+j}))\to 0,\end{aligned}$$

again as a consequence of Theorem 1. This completes the proof.

Theorem 5. *Under the assumptions of Theorem* 1 *we have*:

(a) $|c_{n,T}(\boldsymbol{t}_{n,T})-c(\boldsymbol{t}_{n,T})|\to 0$ *as* $n\to\infty$.

(b) *For any fixed r and* $0\le\alpha<-\ln[E(1/Y)]/\ln\mu$

$$\mu^{\alpha(n+T)}r_{n+T}^{-2}\left|P(X_n^{(1)}=i_1,\ldots,X_n^{(r)}=i_r)-\prod_{j=1}^{r}i_jp(i_j\,|\,1)\mu^{-1}\right|\to 0$$

where $c_{n,T}$ *denotes the characteristic function of* $\boldsymbol{X}_{n,T}$ *while c is the characteristic function of* $r_n+r_{n+1}+\cdots+r_{n+T}$ *i.i.d. random variables with probability function* $ip(i\,|\,1)/\mu$.

Theorem 6. *Under the assumptions of Theorem* 1 *we have*:

(a) $P(D_{n,T_n})\to 1$ *if, for some* $\alpha<-\ln[E(1/Y)]/\ln\mu$,

$$\sum_{j=n}^{n+T_n} r_j^2\mu^{-\alpha j}\to 0.$$

(b) $\mu^{\alpha n}T_n^{-1}r_n^{-2}[1-P(D_{n,T_n})]\to 0$, *for all* $0\le\alpha<-[\ln E(1/Y)]/\ln\mu$.

Results concerning the limiting behaviour of $\boldsymbol{X}_{n,T_n}$ are derived as before. For instance, Theorem 6 may be used to show that

$$|c_{n,T_n}(\boldsymbol{t}_{n,T_n})-c(\boldsymbol{t}_{n,T_n})|\to 0 \tag{13}$$

and that for any fixed subset of r coordinates of $\boldsymbol{X}_{n,T_n}$ Theorem 5(b) holds with rate of convergence $n^{\alpha n}(T_nr_n^2)^{-1}$. In the case where sample sizes remain constant from the nth generation to the $(n+T_n)$th, this can, of course, be strengthened to $n^{\alpha n}r_n^{-1}$.

Now the following two conditions play a key role in the development of our statistical results.

Condition (A): $r_n^2\mu^{-n}\to 0$, as $n\to\infty$.

Condition (B): For some $\alpha<-\ln[E(1/Y)]/\ln(\mu)$, $Y\stackrel{d}{=}Y_n^{(k)}$, $\sum_{j=n}^{n+T_n} r_j^2/\mu^{\alpha j}\to 0$, as $n\to\infty$.

Recall, since $Z_n/N\mu^n \xrightarrow{\text{a.s.}} W \stackrel{\text{a.s.}}{>} 0$, that (A) is equivalent to $r_n^2/Z_n\xrightarrow{\text{a.s.}}0$. Moreover, since $\alpha>1$

$$r_n^2\mu^{-n}\le\sum_{j=n}^{n+T_n} r_j^2/\mu^{\alpha j}$$

so that (B) implies (A). The following three lemmas will be used repeatedly.

Lemma 2. *Condition* (B) *implies* $P(D_{n,T_n}) \to 1$ *while Condition* (A) *implies* $P(D_n) \to 1$.

Lemma 3.

$$E[I(D_{n,T_n})g(\boldsymbol{X}_{n,T_n})] = E\left[\prod_{j=n}^{n+T_n}\prod_{k=1}^{r_j} Y_{j-1}^{(k)}\binom{Z_{j-1}}{r_j}\binom{Z_j}{r_j}^{-1} g(\boldsymbol{Y}_{n-1,T_n})\right]$$

where $\boldsymbol{Y}_{n-1,T_n} = (\boldsymbol{Y}'_{n-1}, \ldots, \boldsymbol{Y}'_{n-1+T_n})$ *and* $\boldsymbol{Y}_n = (Y_n^{(1)}, \ldots, Y_n^{(r_n)})'$. *Moreover, if we take* $\boldsymbol{X}_{\infty,T_n}$ *to be an* $R_n \times 1$ *random vector consisting of i.i.d.* $p(i)$ *r.v.'s, then*

$$\mu^{-R_n}E\left[\prod_{j=n}^{n+T_n}\prod_{k=1}^{r_j} Y_{j-1}^{(k)} g(\boldsymbol{Y}_{n-1,T_n})\right] = E[g(\boldsymbol{X}_{\infty,T_n})].$$

Lemma 4. *Condition* (B) *implies*

$$E\left\{\prod_{j=n}^{n+T_n}\prod_{k=1}^{r_j} Y_{j-1}^{(k)}\left|\prod_{j=n}^{n+T_n}\binom{Z_{j-1}}{r_j}\binom{Z_j}{r_j}^{-1} - \mu^{-R_n}\right|\right\} \to 0, \tag{14}$$

while Condition (A) *implies* (14) *with* $T_n = 0$.

A proof of Lemma 2 is identical to that of Theorem 4(a) (see also Theorem 6(a)). The first part of Lemma 3 follows on considering the joint distribution of $I(D_{n,T_n})$ and $\boldsymbol{X}_{n,T_n}$ while the second assertion is a consequence of the relationships $p(i) = ip(i\,|\,1)/\mu$. Of course this lemma holds only for those g's for which the relevant expectations exist. Finally, we note that the proof of (14) in Lemma 4 is contained in the proof of Theorem 2. The extension to $T_n > 0$ follows along similar lines.

6. Asymptotic method of moments estimates

Suppose that our partially observed Galton–Watson tree $\{\boldsymbol{X}_n\}$ yields $\boldsymbol{X}_{n,T_n}$. Then under assumption (B) we may approximate the distribution of $\boldsymbol{X}_{n,T_n}$ by that of $R_n = r_n + \cdots + r_{n+T_n}$ i.i.d. random variables having probability function $p(i) = ip(i\,|\,1)/\mu$. If $X \sim p(i)$ then

$$E(1/X) = 1/\mu, \qquad E[I(\{X = i\})] = p(i)$$

which suggests the 'approximate' moment estimators

$$\left(\widehat{\frac{1}{\mu}}\right)_{\text{AM}} = \frac{1}{R_n}\sum_{j=1}^{R_n}(X_{n,T_n}^{(j)})^{-1}, \qquad \hat{p}_{\text{AM}}(i) = \frac{1}{R_n}\sum_{j=1}^{R_n} I(\{X_{n,T_n}^{(j)} = i\})$$

where $X_{n,T_n}^{(j)}$ is the jth component of $\boldsymbol{X}_{n,T_n}$. Now set $\hat{\mu}_{\text{AM}} = (\hat{1}/\mu)_{\text{AM}}^{-1}$ and take

$$\hat{p}_{\text{AM}}(i\,|\,1) = \hat{\mu}_{\text{AM}}\hat{p}_{\text{AM}}(i)/i.$$

The following theorem then demonstrates the asymptotic normality ($\overset{a}{\sim}$) of these estimators.

Theorem 7. *Under conditions* (A) *and* (B), *the following hold*:

(a) $(1/\hat{\mu}_{\mathrm{AM}}, \hat{p}_{\mathrm{AM}}(i))' \overset{a}{\sim} N((1/\mu, p(i))', R_n^{-1}\Sigma)$ *where*

$$\Sigma = \begin{bmatrix} \sigma_{1/x}^2 & \left(\dfrac{1}{i}-\dfrac{1}{\mu}\right)p(i) \\ \left(\dfrac{1}{i}-\dfrac{1}{\mu}\right)p(i) & p(i)(1-p(i)) \end{bmatrix}$$

($\sigma_{1/x}^2 \equiv \mathrm{Var}(1/X)$, *where* $X \sim p(i)$),

(b) $\hat{\mu} \overset{a}{\sim} N(\mu, R_n^{-1}\mu^4\sigma_{1/x}^2)$,

(c) $\hat{p}(i\,|\,1)_{\mathrm{AM}} \overset{a}{\sim} N(p(i\,|\,1), \sigma_{p,i}^2/R_n)$, *where*

$$\sigma_{p,i}^2 \equiv \mu^2 p(i\,|\,1)^2\sigma_{1/x}^2 - 2\mu\left(\frac{1}{i}-\frac{1}{\mu}\right)p(i\,|\,1)^2 + \frac{1}{i}\,\mu p(i\,|\,1)(1-p(i\,|\,1)).$$

Proof of Theorem 7. Results (b) and (c) follow directly, after some calculation, from (a) using the (well-known) method of "statistical differentials". To show (a) we employ the Cramér–Wold technique and consider the linear combination $\hat{\beta}$ given by

$$\hat{\beta} = c_1/\hat{\mu}_{\mathrm{AM}} + c_2\hat{p}_{\mathrm{AM}}(i)$$

where $c_1, c_2 \in R$. If the $X_n^{(j)}$ were i.i.d. $X \sim p(i)$ then $\hat{\beta}$ would have variance σ_β^2/R_n where

$$\sigma_\beta^2 = c_1^2\sigma_{1/x}^2 + c_2^2 p(i)(1-p(i)) + 2c_1c_2\left(\frac{1}{i}-\frac{1}{\mu}\right)p(i).$$

Set $\beta = c_1/\mu + c_2(i)$ and $V_n = (\hat{\beta}-\beta)/(\sigma_\beta/\sqrt{R_n})$. We now show $V_n \xrightarrow{d} N(0,1)$, $\forall (c_1, c_2) \neq (0,0)$ which is sufficient to demonstrate (a). Thus let g be a bounded (by 1, say) continuous function and $Z \sim N(0,1)$. It suffices to show $E(g(V_n)) \to E(g(Z))$. Consider then

$$\begin{aligned} |E[g(V_n)-g(Z)]| &\leq |E[(1-I(D_{n,T_n}))(g(V_n)-g(Z))]| \\ &\quad + |E[I(D_{n,T_n})(g(V_n)-g(Z))]| \\ &\leq 2[1-P(D_{n,T_n})] + |E[I(D_{n,T_n})(g(V_n)-g(Z))]|. \end{aligned}$$

Now $P(D_{n,T_n}) \to 1$ while

$$\begin{aligned} E[I(D_{n,T_n})g(Z)] &= E[I(D_{n,T_n})]E[g(Z)] \\ &= P(D_{n,T_n})E[g(Z)] \to E[g(Z)]. \end{aligned}$$

Write $V_n = V_n(\boldsymbol{X}_{n,T_n})$. Then by Lemma 3

$$E(I(D_{n,T_n})g[V_n(\boldsymbol{X}_{n,T_n})])$$

$$= E\left\{\prod_{j=n}^{n+T_n}\prod_{k=1}^{r_j}\left[Y_{j-1}^{(k)}\binom{Z_{j-1}}{r_j}\binom{Z_j}{r_j}^{-1}\right]g(V_n(\boldsymbol{Y}_{n-1,T_n}))\right\}$$

$$= E\left\{\left[\prod_{j=n}^{n+T_n}\binom{Z_{j-1}}{r_j}\binom{Z_j}{r_j}^{-1} - \mu^{-R_n}\right]\left[\prod_{j=n}^{n+T_n}\prod_{k=1}^{r_j} Y_{j-1}^{(k)}\right][g(V_n(\boldsymbol{Y}_{n-1,T_n}))]\right\}$$

$$+ E\left\{\mu^{-R_n}\left[\prod_{j=n}^{n+T_n}\prod_{k=1}^{r_j} Y_{j-1}^{(k)}\right][g(V_n(\boldsymbol{Y}_{n-1,T_n}))]\right\}.$$

Using Lemma 3 and the Central Limit Theorem we obtain

$$E\left\{\mu^{-R_n}\left[\prod_{j=n}^{n+T_n}\prod_{k=1}^{r_j} Y_{j-1}^{(k)}\right][g(V_n(\boldsymbol{Y}_{n-1,T_n}))]\right\} = E[g(V_n(\boldsymbol{X}_{\infty,T_n}))] \to E[g(z)].$$

On the other hand Lemma 4 yields

$$\left|E\left\{\left[\prod_{j=n}^{n+T_n}\binom{Z_{j-1}}{r_j}\binom{Z_{j-1}}{r_j}^{-1} - \mu^{-R_n}\right]\left[\prod_{j=n}^{n+T_n}\prod_{k=1}^{r_j} Y_{j-1}^{(k)}\right][g(V_n(\boldsymbol{Y}_{n-1,T_n}))]\right\}\right|$$

$$\leq E\left\{\prod_{j=n}^{n+T_n}\prod_{k=1}^{r_j} Y_{j-1}^{(k)}\left|\prod_{j=n}^{n+T_n}\binom{Z_{j-1}}{r_j}\binom{Z_j}{r_j}^{-1} - \mu^{-R_n}\right|\right\} \to 0.$$

Piecing everything back together then yields the desired result $|E[g(\nu_n) - g(Z)| \to 0$.

Note, incidentally, that one consequence of Theorem 7 is the weak consistency ($\xrightarrow{P}$) of $\hat{\mu}_{AM}$, $\hat{p}_{AM}(i/1)$, and $\hat{p}_{AM}(i)$. In fact, it can be shown that the latter two estimates are uniformly (in i) weakly consistent.

7. Approximate likelihood estimation

In the classical i.i.d. case, it is well-known that method of moments estimates need not be asymptotically efficient (in the sense of attaining the Cramér–Rao bound in their limiting normal distributions), while maximum likelihood estimates usually are (under suitable regularity conditions). Since the process of family sizes $\{\boldsymbol{X}_n\}$ behaves asymptotically in an i.i.d. fashion, one may expect similar results to hold here. Indeed, despite having to base our methods on approximate likelihood, we shall, in this section, see that analogous results continue to hold for incompletely observed branching trees having power series offspring distributions. The case of more general parametric families of offspring distributions is discussed briefly in our concluding section.

Consider offspring distributions with probability functions

$$P(k\,|\,1)=\frac{\theta^k a_k}{A(\theta)},\quad n=1,2,\ldots$$

where $\theta>0$. The constants $a_k\geqslant 0$ are assumed known while

$$A(\theta)=\sum_{k=1}^{\infty}\theta^k a_k,$$

so that $A(\theta)$, $A'(\theta)$, $A''(\theta)>0$. Notice that the existence of the rth moment of $p(\cdot\,|\,1)$ implies the existence of $A^{(r)}(\theta)$ as well as ensuring $A^{(r)}(\theta)>0$. The offspring mean μ may be expressed as

$$\mu=\theta\frac{\mathrm{d}}{\mathrm{d}\theta}[A(\theta)]$$

which implies a one-to-one relationship between μ and θ. Thus, as is well-known, the power series family (with finite first moment) may also be parametrized by its mean. Now, since $p(i)=ip(i\,|\,1)/\mu$ we find that the asymptotic family size distribution is also in the power series class and is given by

$$P(i)=\frac{\theta^i i a_i}{\theta A'(\theta)},\quad i=1,2,\ldots.$$

Consider now the approximate likelihood $L_A(\theta)$ obtained by assuming $\boldsymbol{X}_{n,T_n}$ to consist of random variables which are i.i.d. $p(i)$. This yields

$$L_A(\theta)\propto A'(\theta)^{-R_n}\theta^{S_n-R_n},$$

where the arbitrary positive constant of proportionality does not involve θ (but may depend on $\boldsymbol{X}_{n,T_n}$), and

$$S_n=\sum_{j=n}^{n+T_n}\sum_{k=1}^{r_j}X_j^{(k)}.$$

The approximate score function is therefore

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L_A(\theta)=-R_n\frac{A''(\theta)}{A'(\theta)}+\frac{1}{\theta}(S_n-R_n).$$

The score equation

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L_A(\theta)=0$$

is equivalent to

$$\frac{S_n}{R_n}=\theta\frac{A''(\theta)}{A'(\theta)}+1. \tag{15}$$

Now, it is easily shown that the derivative of the RHS of (15) is >0, for every $\theta > 0$, while

$$\theta \frac{A''(\theta)}{A'(\theta)} = \left(\sum_{k=0}^{\infty} k(k+1)a_{k+1}\theta^k \right) \Big/ \left(\sum_{k=0}^{\infty} (k+1)a_{k+1}\theta^k \right)$$

$$\to \infty \quad \text{as } \theta \to \infty,$$

$$\to 0 \quad \text{as } \theta \downarrow 0,$$

so that (15) has a unique solution, say $h(S_n/R_n)$. Moreover, since $L_A(\theta) \to 0$ as either $\theta \to \infty$ or $\theta \downarrow 0$ we can conclude that $L_A(\theta)$ is unimodal with a maximum occurring at

$$\hat{\theta}_{\text{AML}} \equiv h(S_n/R_n).$$

Note that h is simply the inverse function of $(\theta A''(\theta)/A''(\theta)) + 1$, while S_n/R_n is the average of all R_n observations. For convenience set $\bar{X} = S_n/R_n$. We have the following result.

Lemma 5. *Under either condition* (A) *or* (B) *and assuming a power series offspring distribution with finite second moment*

$$\bar{X} \overset{a}{\sim} N(\mu_x, \sigma_x^2/R_n)$$

where $\mu_x = E(X)$, $\sigma_x^2 = \text{Var}(X)$, $X \sim p(i)$.

Using the mean value theorem, the above lemma, the equation $\hat{\theta}_{\text{AML}} = h(\bar{X})$ then yields the following theorem.

Theorem 8. *Under the conditions of Lemma* 2,

$$\hat{\theta}_{\text{AML}} \overset{a}{\sim} N(\theta, \sigma_x^2\{h'(\mu_x)\}^2/R_n).$$

Proof of Lemma 5. Proceeding as in the proof of Theorem 7, we select a bounded (by 1) continuous function g and show that $E[g(Z_n(\boldsymbol{X}_{n,T_n})) - g(Z)] \to 0$, where

$$Z_n(\boldsymbol{X}_{n,T_n}) = \frac{\bar{X} - \mu_x}{\sigma_x/\sqrt{R_n}}$$

and $Z \sim N(0, 1)$. As before

$$|E[(1 - I(D_{n,T_n}))(g(Z_n(\boldsymbol{X}_{n,T_n})) - g(Z))]| \leq 2P(D^c_{n,T_n}) \to 0,$$

while $E[I(D_{n,T_n})g(Z)] \to E[g(Z)]$. Now we use the Central Limit Theorem and

Lemmas 3 and 4 to obtain (after some simplification),

$$
\begin{aligned}
&|E[I(D_{n,T_n})(g(Z_n(\boldsymbol{X}_{n,T_n}))-g(Z))]| \\
&\quad \leqslant |E[g(Z_n(\boldsymbol{X}_{n,T_n})-g(Z))]|+|E[I(D_{n,T_n})g(Z)-g(Z)]| \\
&\quad\quad +E\left\{\prod_{j=n}^{n+T_n}\prod_{k=1}^{r_j} Y_{j-1}^{(k)}\left|\prod_{j=n}^{n+T_n}\binom{Z_{j-1}}{r_j}\binom{Z_j}{r_j}^{-1}-\mu^{-R_n}\right|\right\} \\
&\quad \to 0,
\end{aligned}
$$

from which we conclude $E[g(Z_n(\boldsymbol{X}_n, T_n))]\to E(g(Z))$.

At this point we are in a position to compare the asymptotic distributions of estimates of μ based on the approximate likelihood and moment procedures. Indeed, the offspring mean μ is related to θ via

$$\mu(\theta)=\theta\frac{\mathrm{d}}{\mathrm{d}\theta}\log[A(\theta)]$$

which is easily seen to be one-to-one and continuous. The approximate MLE of μ is therefore

$$\hat{\mu}_{\mathrm{AML}}=\mu(\hat{\theta}_{\mathrm{AML}})\overset{a}{\sim} N(\mu,[\mu'(\theta)h'(\mu_x)]^2/R_n),$$

this last result again following from the method of statistical differentials. Now notice that the Fisher information about θ contained in a sample of R_n i.i.d. $p(i)=ia_i\theta^{i-1}/A'(\theta)$ random variables is precisely

$$R_n/\sigma_x^2[h'(\mu_x)]^2,$$

and this is the reciprocal of the asymptotic variance of $\hat{\theta}_{\mathrm{AML}}$. In terms of μ the Fisher information is

$$R_n/\sigma_x^2[\mu'(\theta)h'(\mu_x)]^2.$$

Applying the method of moments to this i.i.d. sample yields

$$\hat{\theta}_{\mathrm{MM}}\sim N(\theta, R_n^{-1}\mu'(\theta)^{-2}\mu(\theta)^4\sigma_{1/x}^2)$$

so that

$$[\mu'(\theta)]^{-2}[\mu(\theta)]^4\sigma_{1/x}^2\geqslant\sigma_x^2h'(\mu_x)^2$$

from which we conclude that $\hat{\theta}_{\mathrm{AML}}$ is asymptotically efficient relative to $\hat{\theta}_{\mathrm{AMM}}$. Hence $\hat{\mu}_{\mathrm{AML}}$ is asymptotically efficient relative to $\hat{\mu}_{\mathrm{AMM}}$. A measure of the relative efficiency of $\hat{\mu}_{\mathrm{AMM}}$ to $\hat{\mu}_{\mathrm{AML}}$ is the ratio

$$e(\theta)\equiv e(\hat{\mu}_{\mathrm{AMM}},\hat{\mu}_{\mathrm{AML}})\equiv\frac{[\mu'(\theta)h'(\mu_x)]^2\sigma_x^2}{\mu(\theta)^4\sigma_{1/x}^2},$$

where we have used the result $\hat{\mu}_{\mathrm{AMM}}\overset{a}{\sim} N(\mu,\mu^4\sigma_{1/x}^2/R_n)$ given in Theorem 7.

We know, from the preceding arguments, that $0 \leq e(\theta) \leq 1$. This inequality is not of course strict and there are cases where both $\hat{\mu}_{\mathrm{AML}}$ and $\hat{\mu}_{\mathrm{AMM}}$ are equally efficient for all $\theta > 0$. For example, suppose the offspring distribution is the shifted Bernoulli

$$p(i\,|\,1) = \theta^i[\theta(1+\theta)]^{-1}, \quad i = 1, 2.$$

Then both $\hat{\mu}_{\mathrm{AML}}$ and $\hat{\mu}_{\mathrm{AMM}}$ have asymptotic variance

$$\frac{\theta(1+2\theta)^2}{2(1+\theta)^4 R_n}$$

so that $e(\theta) \equiv 1$. On the other hand, for the shifted geometric

$$p(i\,|\,1) = \theta^i\left(\frac{1-\theta}{\theta}\right), \quad i = 1, 2, 3, \ldots,$$

$$e(\theta) = \frac{[(1-\theta)^{-2}(1-\theta)^2/2]^2[2\theta(1-\theta)^{-2}]}{[(1-\theta)^{-1}]^4[-(1-\theta)^2(\theta^{-1}\log(1-\theta)+1)]}$$
$$= [2(\tfrac{1}{2} + \theta/3 + \theta^2/4 + \cdots)]^{-1},$$

which varies continuously from 0 ($\theta = 1$) to 1 ($\theta = 0$). In this case a Monte Carlo experiment was conducted in order to compare $\hat{\mu}_{\mathrm{AML}}$ and $\hat{\mu}_{\mathrm{AMM}}$ from a different perspective. A Galton–Watson process was started from 10 ancestors. The shifted geometric offspring distribution had $\theta = 0.25$ and hence $\mu = 4/3$. A sample of $r_j = r = 5$ individuals was taken from each of the first 10 generations and $\hat{\mu}_{\mathrm{AML}}$, $\hat{\mu}_{\mathrm{AMM}}$ were calculated from the data at each generation (thus $T_j = 0$). This procedure was replicated 402 times. Two measures of the error made in approximating the exact distributions of $\hat{\mu}_{\mathrm{AML}}$ and $\hat{\mu}_{\mathrm{AMM}}$ may be obtained via the Central Limit Theorem and the known asymptotic distribution of the Kolmogrov–Smirnov statistic (cf. Serfling (1980, p. 62)). These yield

$$P(|\hat{F}_{\hat{\mu}}(x) - F_{\hat{\mu}}(x)| \leq 0.04) \approx 0.90 \tag{16}$$

and

$$P(\sup_x |\hat{F}_{\hat{\mu}}(x) - F_{\hat{\mu}}(x)| \leq 0.06) \approx 0.90, \tag{17}$$

respectively. The bounds in (16) and (17) apply to both $\hat{\mu}_{\mathrm{AML}}$ and $\hat{\mu}_{\mathrm{AMM}}$ and refer to errors obtained through use of a finite (402) number of replications. The "≈" results from use of asymptotic approximations to the distributions involving the empirical distribution function. To obtain tighter bounds in (16) and (17) a larger number of replications would be necessary. For example to reduce 0.04 to 0.01 in (17) requires about 1700 replications, while a further reduction to 0.001 necessitates 169 000. For a bound of 0.01 in (4) we would require 15 000 replications and a 0.001 bound would result through the use of about 1.5 million! The results of the experiment are summarized in Table 1. A corresponding summary for a similar Monte Carlo study with $r_1 = 7$, $r_5 = 11$, $r_{10} = 16$ is provided in Table 2. These results point to a better performance of $\hat{\mu}_{\mathrm{AML}}$ over $\hat{\mu}_{\mathrm{AMM}}$—from the point of view of lower observed

Table 1
Estimated properties of the distributions of $\hat{\mu}_{AML}$ and $\hat{\mu}_{AMM}$ based on 402 replications. True mean $= \frac{4}{3}$, $\sigma_A(\hat{\mu}_{AML}) = 0.2108$, $\sigma_A(\hat{\mu}_{AMM}) = 0.2315$

Gen.	Mean		Standard deviation		Skewness		Kurtosis	
	MME	MLE	MME	MLE	MME	MLE	MME	MLE
1	1.3941	1.3184	0.3499	0.2571	1.8335	1.2658	5.7304	1.9683
2	1.3512	1.2925	0.3036	0.2235	1.4717	0.8499	3.1380	0.7203
3	1.3357	1.3107	0.3022	0.2537	1.5365	1.4585	3.8079	3.1734
4	1.4029	1.3396	0.3211	0.2387	1.3983	1.2279	2.4807	2.4345
5	1.3911	1.3326	0.3356	0.2464	1.8983	1.2715	5.0919	2.0410
6	1.3996	1.3368	0.3338	0.2461	1.7614	1.0086	5.5897	1.3621
7	1.3705	1.3264	0.2734	0.2233	1.5067	1.2796	4.1876	2.8340
8	1.3711	1.3204	0.2907	0.2181	1.6374	0.9236	3.9523	1.0090
9	1.4179	1.3517	0.3061	0.2192	1.5023	0.6845	4.0772	0.2784
10	1.3699	1.3281	0.2502	0.2040	1.0512	0.6672	1.8259	0.4260

Table 2

Gen.	MME				
	Mean	Standard deviation	Asymptotic standard deviation	Skewness	Kurtosis
1	1.3794	0.3186	0.1957	1.5439	3.4435
5	1.3534	0.1992	0.1561	1.2670	3.0651
10	1.3453	0.1585	0.1294	0.5477	0.0204
Gen.	MLE				
	Mean	Standard deviation	Asymptotic standard deviation	Skewness	Kurtosis
1	1.3102	0.2419	0.1782	1.1862	1.6166
5	1.3236	0.1740	0.1421	1.1126	2.3907
10	1.3261	0.1414	0.1179	0.5630	0.3259

mean square error and in the adequacy of the normal approximation to their respective distributions.

8. Conclusion

This article has provided some techniques for estimating offspring parameters on the basis of a sampled branching tree $\{\boldsymbol{X}_n\}$. While the emphasis has been on estimation other modes of inference can be developed. For example, from a likelihood point of view, and for power series offspring distributions, conditions (A)/(B) yield

$$\frac{L_A(\hat{\theta}_{\mathrm{AML}}+t\hat{\sigma}_n)}{L_A(\hat{\theta}_{\mathrm{AML}})} \xrightarrow{\mathrm{P}} \mathrm{e}^{-t^2/2},$$

where $\hat{\sigma}_n^{-2} = -(\mathrm{d}^2/\mathrm{d}\theta^2) \ln L_A|_{\theta=\hat{\theta}_{\mathrm{AML}}}$. If a weighted likelihood procedure is to be employed then for continuous strictly positive weight functions $w(\theta)$ satisfying $\int w(\theta)L_A(\theta)\,\mathrm{d}\theta < \infty$ we have

$$\frac{w(\hat{\theta}_{\mathrm{AML}}+\hat{\sigma}_n t)L_A(\hat{\theta}_{\mathrm{AML}}+\hat{\sigma}_n t)}{\hat{\sigma}_n \int w(\theta)L_A(\theta)\,\mathrm{d}\theta} \xrightarrow{\mathrm{P}} \frac{\mathrm{e}^{-t^2/2}}{\sqrt{2\pi}}.$$

The results can be demonstrated as in Fraser and McDunnough (1984).

While the discussion of the asymptotic distribution of $\hat{\theta}_{\mathrm{AML}}$ has taken place in the context of power offspring distributions, the results extend readily to more general parametric families sharing the property that $L_A(\theta)$ be unimodal. Indeed, since

$$E\left(\frac{\mathrm{d}}{\mathrm{d}\theta} \ln p_\theta(X)\right) = 0$$

when $\theta = \theta_0$ (where $X \sim p_{\theta_0}(i)$) and it can be shown, as in the proof of Theorem 7 that under (A)/(B)

$$\frac{1}{R_n}\frac{\mathrm{d}}{\mathrm{d}\theta} \ln L_A(\theta) \xrightarrow{\mathrm{P}} E\left(\frac{\mathrm{d}}{\mathrm{d}\theta} \ln p_\theta(X)\right)$$

we obtain

$$\hat{\theta}_{\mathrm{AML}} \xrightarrow{\mathrm{P}} \theta_0.$$

Thus follows from the unimodality of $L_A(\theta)$ and the fact that $E((\mathrm{d}/\mathrm{d}\theta) \ln p_\theta(X))$ is strictly decreasing in a neighborhood of $\theta=\theta_0$. Hence the usual differential method yields

$$(\hat{\theta}_{\mathrm{AML}}-\theta) \overset{\mathrm{d}}{\approx} -\frac{\mathrm{d}}{\mathrm{d}\theta} \ln(L_A(\theta)) \bigg/ \frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \ln(L_A(\theta)).$$

Now apply the method of proof found in Theorem 7 again to get (under (A)/(B))

$$\frac{-1}{R_n}\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \ln(L_A(\theta)) \to -E\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \ln p_\theta(X)\right) = \left\{\mathrm{Var}\left(\frac{\mathrm{d}}{\mathrm{d}\theta} \ln p_\theta(X)\right)\right\}^{-1}$$

and

$$\sqrt{R_n}\frac{\mathrm{d}}{\mathrm{d}\theta} \ln(L_A(\theta)) \xrightarrow{\mathrm{d}} N\left(0, \mathrm{Var}\left(\frac{\mathrm{d}}{\mathrm{d}\theta} \ln p_\theta(X)\right)\right)$$

so that

$$\sqrt{R_n}(\hat{\theta}_{\mathrm{AML}}-\theta) \xrightarrow{\mathrm{d}} N\left(0, \mathrm{Var}\left(\frac{\mathrm{d}}{\mathrm{d}\theta} \ln p_\theta(X)\right)\right).$$

That is,

$$\hat{\theta}_{\text{AML}} \overset{a}{\sim} N\left(\theta, \frac{1}{R_n}\operatorname{Var}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\ln p_\theta(X)\right)\right).$$

Notice that $R_n[\operatorname{Var}((\mathrm{d}/\mathrm{d}\theta)\ln p_\theta(X))]^{-1}$ is the Fisher information about θ contained in a sample of R_n i.i.d. X's.

Finally, we mention an unresolved question. This article has dealt with some exact and approximate distributions of the non-Markovian process $\{\boldsymbol{X}_n\}$, where $\boldsymbol{X}_n$ represents the family sizes of r_n individuals selected at random in the nth generation of a Galton–Watson branching tree. Our results indicate that the process behaves, asymptotically as a sequence of independent random vectors with i.i.d. components (provided r_n increases moderately). This suggests that the offspring distribution may be completely estimated from $\{\boldsymbol{X}_n\}$. In addition complete estimation of the offspring distribution could possibly be achieved by only recording $\{\boldsymbol{X}'_n\boldsymbol{1}\}$, where $\boldsymbol{X}'_n\boldsymbol{1}$ is simply the total size of all families corresponding to the r_n selected individuals from the nth generation. Since we found $\boldsymbol{X}'_n\boldsymbol{1}$ to be approximately independent, with $\boldsymbol{X}'_n\boldsymbol{1}$ approximately distributed as an r_n*-fold* convolution of the distribution $ip(i\,|\,1)/\mu$, it would appear likely that the offspring distribution could still be completely estimated from even this more limited information provided r_n increasing exponentially on the order of Z_n brings only the first two moments of the offspring distribution can be estimated consistently. The authors (Maki and McDunnough (1989)) have investigated this matter, as well as related questions arising from an incomplete observation of a random walk (cf. Guttorp and Siegel (1985)).

References

[1] J.P. Dion, Estimation of the mean and the initial probabilities of a branching process. J. Appl. Prob. 11 (1974) 687–694.

[2] J.P. Dion and N. Keiding, Statistical Inference in Branching Processes. Branching Processes (A. Joffe and P. Ney, eds.) (Marcel Dekker, New York, 1978) 105–140.

[3] P.D. Feigin, A note on maximum likelihood estimation for simple branching processes, Austral. J. Statist. 19 (1977) 152–154.

[4] P.D. Feigin and B. Reiser, On asymptotic ancillarity and inference for Yule and regular nonergodic processes, Biometrika 66 (1979) 279–283.

[5] D.A.S. Fraser and P. McDunnough, Further remarks on asymptotic normality of likelihood and conditional analyses, Can. J. Statist. 12 (1984) 163–172.

[6] P. Guttorp and A.F. Siegel, Consistent estimation in partially observed random walks, Ann. Statist. 13 (1985) 958–969.

[7] T.E. Harris, Branching processes, Ann. Math. Statist. 19 (1948) 474–494.

[8] A. Joffe and W.A.O'N. Waugh, Exact distributions of kin numbers in a Galton–Watson process, J. Appl. Prob. 19 (1982) 767–775.

[9] R.A. Lockhart, On the nonexistence of consistent estimates in Galton–Watson processes, J. Appl. Prob. 19 (1982) 842–846.

[10] E. Maki and P. McDunnough, What can or can't be estimated in branching and related processes?, Stochastic Process. Appl. 31 (1989) 307–314 (this issue).

[11] T.J. Sweeting, Asymptotic conditional inference for the offspring mean of a supercritical Galton–Watson process, Ann. Statist. 14 (1986) 925–933.