Two Applications of Urn Processes: The Fringe Analysis of Search Trees and the Simulation of Quasi-Stationary Distributions of Markov Chains

David Aldous[1]
Department of Statistics
University of California
Berkeley, CA 94720

Barry Flannery
Department of Mathematics and Computer Science
Valdosta State College
Valdosta, GA 31698

José Luis Palacios[2]
Universidad Simón Bolívar
Departamento de Matemáticas y C. C.
Apartado 89.000
Caracas, Venezuela

Technical Report No. 122
November 1987

[1]Research supported through National Science Foundation Grant MCS84-03239.
[2]Research partially supported by the NSF-AMS Program for Exchange and Scientific Collaboration with Latin American Countries

Department of Statistics
University of California
Berkeley, California

Two Applications of Urn Processes: The Fringe Analysis of Search Trees and the Simulation of Quasi-Stationary Distributions of Markov Chains

David Aldous[1]
Department of Statistics
University of California
Berkeley, CA 94720

Barry Flannery
Department of Mathematics and Computer Science
Valdosta State College
Valdosta, GA 31698

José Luis Palacios[2]
Universidad Simón Bolívar
Departamento de Matemáticas y C. C.
Apartado 89.000
Caracas, Venezuela

[1]Research supported through National Science Foundation Grant MCS84-03239.
[2]Research partially supported by the NSF-AMS Program for Exchange and Scientific Collaboration with Latin American Countries

Abstract

From an urn containing colored balls draw one ball and replace a random number of differently-colored balls, with the distribution of the added balls depending only on the color of the ball drawn. Under mild regularity conditions, the proportions of different colors will converge to deterministic limits. Two applications of this standard result are described. The average efficiency of binary trees, 2-3 trees and other structures for information storage has been studied with diverse techniques. One such technique for estimation is called fringe analysis, and this turns out to involve urn processes. The other topic is the simulation of quasi-stationary distributions of Markov chains. These arise when the chain is conditioned to avoid a taboo set. Here urn processes are used to prove an analog of the classical result on convergence of empirical distributions to the stationary distribution.

Key Words: urn process, branching process, binary search tree, 2-3 trees, empirical convergence, quasi-stationary distribution.

AMS (MOS) Subject Classification: 60J20, 68P05

Abbreviated Title: Applications of Urn Processes

1. Urn Processes

In this section we define a class of urn processes and state a limit theorem. Except for minor differences in hypotheses, these processes and the limit theorem are those described in Athreya and Ney (1972) section V.9.

Let $(c_1, c_2, \ldots, c_m)$ be a vector of constants. Consider an urn and a supply of balls of m colors. The composition of the urn after n draws is described by a random vector $X(n) = (X_1(n), \ldots, X_m(n))$, where $X_i(n)$ is the number of balls of color i. A draw is performed as follows: pick one ball from the urn, with each ball of color i having chance $\dfrac{c_i}{\sum_i c_i X_i(n)}$ of being drawn. We see that the vector c gives the relative probability of an individual i-colored ball being drawn. Now, given that a ball of color i has been drawn, remove it and replace it by $Z_j^{(i)}$ balls of color j $(1 \le j \le m)$. The random choices on each draw are independent. Finally, let the initial distribution $X(0)$ be arbitrary.

This describes an urn process parametrized by m constants $(c_1, \ldots, c_m)$ and m distributions $(Z_j^{(i)}, 1 \le j \le m)$, $1 \le j \le m$. We make the following hypotheses concerning these parameters:

(1.1) $c_i > 0$ for all i

(1.2) $Z_j^{(i)}$ is non-negative integer-valued, $\sum_j Z_j^{(i)} \ge 1$ for each i and $P(\sum_j Z_j^{(i)} > 1) > 0$ for some i.

(1.3) $E(Z_j^{(i)} \log Z_j^{(i)}) < \infty$.

Let $Q = (q_{i,j})$ be the matrix with entries $q_{i,j} = EZ_j^{(i)}$. We make the last assumption

(1.4) Q is irreducible, that is to say for each pair i,j there exists an n such that $q_{i,j}^n > 0$.

Next, let $\hat{X}(n)$ be the vector of proportions after n draws:

$$\hat{X}_i(n) = \frac{X_i(n)}{\sum_j X_j(n)}.$$

Let $\mathbf{R}$ be the matrix with entries $r_{ij} = c_i(q_{i,j} - \delta_{i,j})$, where $\delta_{i,j} = 1_{(i=j)}$.

Theorem 1.5. *Consider an urn process satisfying (1.1) - (1.4). Then*

(a) the equations $\sum_i x_i r_{i,j} = \lambda x_j$, $1 \le j \le m$ *have a unique solution* (x, λ) *for which* $x_i > 0$, $\Sigma x_i = 1$ *and* $\lambda > 0$.

(b) $\hat{\mathbf{X}}(n) \to \mathbf{x}$ *a.s. as* $n \to \infty$, *where* $\mathbf{x}$ *is the eigenvector in (a).*

This theorem is essentially the result of Athreya and Ney (1972), except for some minor changes in hypotheses; they use the hypotheses

(1.1)′ $c_i = 1$ for each i

(1.2)′ the drawn ball is replaced, so that $Z_i^{(i)} \geq 1$ for all i and

(1.3)′ the matrix $\mathbf{Q}$ is regular, that is to say there exists n such that $q_{i,j}^n > 0$ for all pairs i, j.

The proof uses the natural embedding of the urn process into a continuous-time multi-type branching process, where individuals of type i have exponential (mean 1) life-times and on death are replaced by $Z_j^{(i)}$ $(1 \leq j \leq m)$ offspring of type j. Minor modifications of the proof allow our relaxations to be made. To modify (1.1)′ make type i individuals have exponential (with rate c_i) lifetimes. Condition (1.2)′ is used to ensure that the branching process is supercritical with extinction probability 0, and our hypothesis (1.2) has the same effect. Finally, we see (1.3)′ is to guarantee that the process is "positive regular", and our weaker hypothesis (1.4) does the job.

For later use, we remark that in the context of Theorem 1.5,

$$\text{if} \quad \sum_j q_{i,j} c_j = c_i + 1 \quad \text{for all i, then } \lambda = 1. \tag{1.6}$$

To show this, we observe

$$\begin{aligned} \lambda \sum_j c_j x_j &= \sum_j \sum_i c_j x_i r_{i,j} \qquad \text{using (a)} \\ &= \sum_i \sum_j c_j x_i c_i (q_{i,j} - \delta_{i,j}) \quad \text{by definition of } r_{i,j} \\ &= \sum_i c_i x_i \quad \text{by (1.6).} \end{aligned}$$

The purpose of this paper is to exhibit two completely different applications of this theorem. The "novel content" of the paper is simply the (trivial) observation that the urn result does apply to the problems discussed (without this observation, the problems look more difficult). We have chosen to describe the problems in Section 2 in some detail, rather than refer to other papers. They would perhaps make interesting non-traditional examples in a course on branching processes or urn models.

2. Fringe Analysis of Search Trees

Binary search trees are a well-known structure for storage, retrieval and addition of information labelled by numerical keys. Perhaps the simplest example is that described in Knuth (1973). Keys are stored at the internal nodes. To search for a key x, one compares x with the key y (say) occupying the top node. If $x \neq y$, then move

to the left or right according to whether $x < y$ or $x > y$, respectively. At the next mode, a comparison is made with the key located there. Ultimately, either x is located or the search terminates at an external node. In the latter case, x may be appended at that node thus creating two new external nodes. The figure below shows the effect of adding keys 13 and 1 to the tree. A standard convention is to depict internal nodes with ovals and external nodes with squares.

Figures 1 and 2 here.

Another scheme for information storage, called "2-3 tree" (see Yao (1978)) allows the internal nodes to contain either 1 or 2 keys and has all the external nodes at the same bottom level. To put a new key into a node that contains only one key, we simply insert it as a second key. If the node already contains 2 keys, the node is split into two nodes containing respectively the minimum and the maximum of the three keys and the middle key is inserted into the parent node by repeating the process. When thee is no node above, a new root is created to hold the middle key. For example, to this 2-3 tree

Figure 3 here.

we add the keys 20 and 65 to obtain:

Figure 4 here.

Write t_n to denote a tree with n keys. With such a tree one may associate various statistics, $f(t_n)$, which measure aspects of the "efficiency" of using the search tree. For instance, one may be interested in the average number of comparisons needed in a search or the height of the tree. Consider any tree with $j - 1$ keys in it. These $j - 1$ keys divide all possible key values into j intervals (notice that there are as many such intervals as external nodes). The insertion of a new key into the tree is said to be a *random insertion* if the new key has equal probabilities of being in any of the j intervals defined before. A random insertion is accomplished if we take as the successive keys $K_1, K_2, \ldots$ a sequence of i.i.d. continuous random variables. Any given algorithm for tree construction will then generate a random sequence of trees T_n, and we will speak of "random binary trees", "random 2-3 trees", etc.

One may compare different algorithms by choosing a measure, f, of efficiency and comparing the random variables $f(T_n)$. It turns out that the algorithm for 2-3 trees is efficient whereas that of the binary tree is inefficient. The latter algorithm can be

improved by means of a simple heuristic proposed and analyzed by Poblete and Munro (1985). Trees of the form shown in the following diagram are produced. To this tree

Figure 5 here.

we add the record 8. The elementary addition algorithm would yield

Figure 6 here.

But this new algorithm would yield the more balanced tree

Figure 7 here.

this algorithm may be described as follows: as comparisons are being made, if the basic algorithm would dictate a new level to be created emanating from the last node while the prior level (from this same node) is not full, then rotate the three keys in order that the median of the three keys is at the top, and the least key is at the left and the largest is at the right.

Let us define the fringe of a tree as the set of subtrees at the bottom of the tree which contain the external nodes. Then the algorithm just described produces trees of two types at the fringe:

Figure 8 here.

Let $(x_1(n), x_2(n))$ be the number of type 1 and type 2 subtrees in the fringe of the tree t_n produced by this algorithm. Of course, this pair will not uniquely determine the tree, but it turns out that most statistics $f(t_n)$ can be either computed or bounded by functions $g(x_1(n), x_2(n))$. The same thing can be said of the 2-3 trees, whose fringe exhibits subtrees of either of these two types:

Figure 9 here.

This technique of estimating efficiency by analyzing the vector $x(n)$ was introduced by Yao (1978) and is now called *fringe analysis*. It was further studied in Eisenbarth et al (1982), who proposed the following general framework. Given an algorithm, suppose we can define m types of subtrees such that

(2.1) the fringe of the sequences of trees, t_n, produced by the algorithm consists only of subtrees of these m types (in our examples m is 2);

(2.2) when an element is inserted into a subtree $\hat{t}$, say, of the fringe, the other subtrees of the fringe are unaffected, and $\hat{t}$ is changed into one or more subtrees in a way depending only on the type of $\hat{t}$.

Consider now the random trees T_n grown from i.i.d. keys $K_1, K_2, \ldots$ and let $X(n) = (X_1(n), \ldots, X_m(n))$ count the number of subtrees of types 1 to m in the fringe of T_n. Then $X(n)$ is an urn process as described in Section 1 for which:

$Z_j^{(i)}$ is the (usually deterministic) number of type j subtrees created by the insertion into a type i subtree;
c_i is the number of external nodes in a type i subtree (= the number of subintervals that the keys of a type i subtree determine).

A draw of a ball is the arrival of a new key, and to select a ball of color i means now to have the new coming key landing in any one of the subtrees of type i in the fringe (c_i "good" places to land out of a total of $\Sigma c_i X_i(n)$ places, yielding the appropriate value of the probability).

We will assume

$$\text{the matrix } Q \text{ with } q_{ij} = EZ_j^{(i)} \text{ is irreducible.} \tag{2.3}$$

This hypothesis is met by every reasonable algorithm that creates random trees, the reason being the cyclic nature of the creating process: usually insertion into a type i subtree ($i < m$) originates a type $i + 1$ subtree and insertion into a type m subtree originates trees of several types including type 1. Of course, more complicated situations preserving irreducibility can arise. As the hypotheses (1.1) - (1.3) are automatically satisfied we may apply Theorem (1.5). The conclusion may be simplified by using some special structure. Because the binary tree grows one node at a time, the urn processes arising here have a special property:

$$\sum_j c_j Z_j^{(i)} = c_i + 1; \quad 1 \le i \le m \tag{2.4}$$

and hence

$$\sum_j c_j q_{i,j} = c_i + 1. \tag{2.5}$$

This allows us to apply (1.6) and conclude that in Theorem 1.5 we have $\lambda = 1$. Also, since the number of external nodes in a search tree equals the number of intervals determined by the first n keys ($n + 1$ such intervals) we can conclude that

$$\sum_i c_i X_i(n) = n + 1. \tag{2.6}$$

Theorem 1.5 says

$$\frac{\sum_i c_i X_i(n)}{\sum_i X_i(n)} \to b = \sum_i c_i x_i,$$

and then (2.6) gives

$$\frac{1}{n}\sum_i X_i(n) \to 1/b.$$

Thus in Theorem 1.5 we can substitute normalization by n for normalization by $\sum_i X_i(n)$, to obtain the following result. Recall $r_{i,j} = c_i(q_{i,j} - \delta_{i,j})$.

Theorem 2.7. *In a fringe analysis satisfying (2.3),* $X_i(n)/n \to x_i/b$ *a.s. as* $n \to \infty$, *where* x *is the unique solution of* $\sum_i x_i r_{i,j} = x_j$, $1 \le j \le m$; $x_i > 0$; $\sum x_i = 1$, *and where* $b = \sum_i c_i x_i$.

Remarks. Previous fringe analysis — Yao (1978), Eisenbarth et al (1982), Poblete and Munro (1985), Cunto and Gascón (1987) and references therein — used recurrence relations for $EX_i(n)$ in order to conclude convergence of $EX_i(n)/n$ to the same limit. The urn process representation gives the stronger conclusion of a.s. convergence. Bagchi and Pal (1985) noted that the fringe of 2-3 trees had an urn process representation, and gave a proof of a central limit theorem: however, they gave an ad hoc proof, rather than exploiting known theory of urn processes.

For completeness, let us give some examples, even though the numerical calculations are the same as those in the original papers. We can apply simultaneously Theorem 2.7 to the trees created by the Poblete-Munro algorithm (P.M. trees from now on) and to the 2-3 trees. In both cases we have $c_1 = 2$, $c_2 = 3$ and Q is the irreducible matrix $\begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}$. The equations have solution

$$(x_1, x_2) = (2/3, 1/3); \quad b = 7/3$$

so that almost surely we have the convergence:

$$\left[\frac{X_1(n)}{X_1(n)+X_2(n)}, \frac{X_2(n)}{X_1(n)+X_2(n)}\right] \to \left[\frac{2}{3}, \frac{1}{3}\right] \tag{2.8}$$

$$\left[\frac{X_1(n)}{n}, \frac{X_2(n)}{n}\right] \to \left[\frac{2}{7}, \frac{1}{7}\right]. \tag{2.9}$$

Now in order to be able to compute statistics of the efficiency of the trees using the above results, we need equations relating the fringe to the totality of the tree. Let $\nu(n)$ and $\mu(n)$ denote respectively the number of nodes above the fringe and the

number of internal nodes of a search tree. For P.M. trees we have the equality:

$$X_1(n) + X_2(n) = \nu(n) - 1 \tag{2.10}$$

which is a simple consequence of the fact that for binary trees the number of external nodes equals the number of internal nodes minus one. For the 2-3 trees we have a set of inequalities (Yao (1978) Lemma 2.2):

$$\frac{3}{2}(X_1(n) + X_2(n)) - \frac{1}{2} \le \mu(n) \le 2(X_1(n) + X_2(n)) - 1. \tag{2.11}$$

Using (2.9) in (2.10) we can conclude that for P.M. trees

$$\frac{\nu(n)}{n} \to \frac{3}{7} \approx 0.43$$

which is a measure of the balance of these trees, because for a perfectly balanced complete binary tree $\frac{\nu(n)}{n} = \frac{2^K}{2^{K+1} - 1} \approx 0.5$ whereas for a totally imbalanced tree (a linear array) $\frac{\nu(n)}{n} = 0$. For the 2-3 trees, using (2.9) in (2.11) we get:

$$0.64 \approx \frac{9}{14} \le \lim_n \frac{\mu(n)}{n} \le \frac{6}{7} \approx 0.86$$

which is basically the corollary to Theorem 2.7 in Yao, except that there, $\mu(n)$ is replaced by the expected value $E\mu(n)$. Yao defines the *storage utilization* for a 2-3 tree of n keys as $\frac{n}{2E\mu(n)}$ (if all internal nodes had two keys, there would be $\frac{n}{2}$ of them, so the storage utilization is the ratio $\frac{\text{smallest possible \# of internal nodes}}{\text{expected \# of internal nodes}}$; Eisenbarth et.al. (1982) discuss the *inverse* of this measure) and gives the asymptotic value $\log 2 \approx 0.69$ for the storage utilization of a B tree (a generalization of 2-3 trees). Here, using (2.9) and (2.11) we can obtain for the point storage utilization the inequalities:

$$0.58 \approx \frac{7}{12} \le \lim_n \frac{n}{2\mu(n)} \le \frac{14}{18} \approx 0.78. \tag{2.12}$$

Our result (2.7) also applies in more complicated settings such as the generalized $k - t$ binary search trees of Cunto and Gascón (1987), a generalization of P.M.-trees where there are up to t keys per node and rotations are made each time there is a linear array of $2k - 1$ nodes in the fringe. As a final example, Yao's second order analysis of 2-3 trees (i.e., the consideration of the *two* bottom tiers of nodes) from our viewpoint is simply the fringe problem for which $(c_1, c_2, c_3, c_4, c_5, c_6, c_7) = (4, 5, 6, 6, 7, 8, 9)$ and Q is the (irreducible) matrix

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2/5 & 3/5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 6/7 & 0 & 0 & 0 & 0 & 4/7 & 0 \\ 3/4 & 3/4 & 0 & 0 & 0 & 0 & 1/4 \\ 2/3 & 2/3 & 2/3 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

An application of (2.7) yields

$$x_1 = \frac{138}{485} \approx 0.28, \quad x_2 = \frac{132}{485} \approx 0.27, \quad x_3 = \frac{304}{3395} \approx 0.09,$$

$$x_4 = \frac{396}{3395} \approx 0.12 \quad x_5 = \frac{15}{97} \approx 0.15 \quad x_6 = \frac{20}{291} \approx 0.07 \quad X_7 = \frac{4}{291} \approx 0.01$$

$$b = \frac{7991}{1455} \approx 5.49.$$

Another combinatorial equation (Yao's lemma 2.8) relates $\mu(n)$ to the two bottom tiers of nodes:

$$\frac{7}{2}\sum_{i=1}^{3} X_i(n) + \frac{9}{2}\sum_{i=4}^{7} X_i(n) - \frac{1}{2} \le \mu(n) \le 4\sum_{i=1}^{3} X_i(n) + 5\sum_{i=4}^{7} X_i(n) - 1.$$

Taking limits after dividing by n and using the values for b and the x_i's just found we get

$$0.70 \approx \frac{78501}{111874} \le \lim_n \frac{\mu(n)}{n} \le \frac{44343}{55937} \cong 0.79$$

the a.s. result analogous to Theorem 2.12 of Yao. We see also that our inequalities in (2.12) for the storage utilization are improved now to

$$0.63 \approx \frac{55937}{88686} \le \lim_n \frac{n}{2\mu(n)} \le \frac{55937}{78561} \cong 0.71.$$

3. Empirical Convergence to Quasi-Stationary Distributions

Let Y_n be a Markov chain in a finite state space I with irreducible and aperiodic transition matrix P. Classical limit theory gives existence and uniqueness of a stationary distribution π satisfying

$$\sum_i \pi_i P_{i,j} = \pi_j \quad \text{for all } j;\ \pi_j > 0;\ \Sigma\pi_j = 1 \tag{3.1}$$

$$P(Y_n = j) \to \pi_j \text{ as } n \to \infty \tag{3.2}$$

$$N^{-1}\sum_{n=1}^{N} 1_{(Y_n = j)} \to \pi_j \text{ a.s. as } N \to \infty. \tag{3.3}$$

So (3.2) asserts convergence of the distribution of Y_n, while (3.3) gives convergence of the empirical distribution. Next we fix $A \subset I$, and let $\hat{P}$ be the matrix P with the rows and columns corresponding to states in A^c deleted. We suppose $\hat{P}$ is aperiodic and irreducible as well. Furthermore, let T be the first hitting time of Y_n on A^c. Formally, $T = \min\{n: Y_n \in A^c\}$. Then there exists a unique *quasi-stationary distribution* α on A satisfying

$$\sum_{i \in A} \alpha_i \hat{P}_{i,j} = \lambda \alpha_j;\ \alpha_j > 0;\ \Sigma \alpha_j = 1 \tag{3.4}$$

where λ is the largest eigenvalue of $\hat{P}$, $0 < \lambda < 1$, and λ is real. See e.g. Darroch and Seneta (1965). This α has the interpretation of being the limiting distribution of the process Y_n conditioned on $\{T > n\}$:

$$P(Y_n = j \mid T > n) \to \alpha_j \quad \text{as} \quad n \to \infty. \tag{3.5}$$

Now (3.4) and (3.5) are plainly analogous to (3.1) and (3.2). There is no immediate analog of (3.3), but we shall motivate an indirect analog.

Consider the setting (3.1) - (3.2). In applications we are usually interested not in the distribution π itself, but in distributions or expectations of functions defined on the state space. Since the distribution of a function g can be obtained from expectations of functions $f(i) = 1_{(g(i) \le x)}$, we need only consider expectations. We know the long-run averages converge:

$$N^{-1} \sum_{n=1}^{N} f(Y_n) \to \Sigma f(i) \pi_i, \text{ a constant.} \tag{3.6}$$

To calculate this constant we have two options:

a) Solve (3.1) for π; or

b) run a computer simulation of the chain for N steps, and use the observed value of (3.6) to estimate the limit.

Method (a) is preferable, but when the number of states is large it becomes impractical, and 9b) becomes more attractive.

The same issue arises in consideration of the quasi-stationary distribution, α. Here (3.5) implies $E(f(Y_n) \mid T > n) \to \Sigma f(i) \alpha_i$, and in the case of a large state space we would like a method of estimating $\Sigma f(i) \alpha_i$ without having to solve equations (3.4).

In this setting, simulations to estimate α by running the chain and discarding runs that hit A^c are unsatisfactory because $P(T > n)$ may be small. Instead, we define a process whose empirical distribution converges to α. First, pretend we know α, and consider a process V_n on A defined as follows: V_n evolves as Y_n until the first hit on A^c, at which point the hit on A^c is not counted as a transition and instead the chain is put back into A according to the distribution α. The process V_n defined in this manner is

a Markov chain with stationary distribution α. And (3.3) shows that the empirical distribution θ_N converges to α, where $\theta_N(j) = N^{-1} \sum_{n=1}^{N} 1_{(V_n = j)}$. Of course, since we do not know α, we cannot simulate this process. However, we can simulate the natural "adaptive" process for which a transition at step $N+1$ which would go into A^c is instead sent into A according to the empirical distribution θ_N. Theorem 3.8 below says for this adaptive process, the empirical distribution does, in fact, converge to α.

We define this Markov process formally.

Definition 3.7. Let M be the set of counting measures on A, and let $\delta_j(i) = 1_{(i=j)}$. Let (V_n, μ_n) be the $A \times M$-valued Markov chain with $V_1 = i_1$, $\mu_1 = \delta_{i_1}$, and with transitions

$$P(V_{n+1} = j, \mu_{n+1} = \mu + \delta_j \mid V_n = i, \mu_n = \mu) = P_{i,j} + \mu(j) \sum_{k \in A^c} P_{i,k}.$$

(Here we are using the following notational convention: μ is an arbitrary vector, μ_n is the empirical counting measure at time n, and $\mu(j)$ is the jth entry of the arbitrary vector.)

Theorem 3.8. *For the process defined in (3.7),* $n^{-1}\mu_n \to \alpha$ *a.s. as* $n \to \infty$, *where* α *is the quasi-stationary distribution (3.4).*

This result was proved in Flannery (1986) in a direct, but complicated manner. It can be deduced easily from Theorem 1.5. To do so, include in the definition (3.7) the count C_n of the number of aborted visits to A^c prior to step n. Precisely, let $C_1 = 0$ and rewrite (3.7) as

$$P(X_{n+1} = j, \mu_{n+1} = \mu + \delta_j, C_{n+1} = c \mid X_n = i, \mu_n = \mu, C_n = c) = P_{i,j}$$

$$P(X_{n+1} = j, \mu_{n+1} = \mu + \delta_j, C_{n+1} = c + 1 \mid X_n = i, \mu_n = \mu, C_n = c) = \mu(j) \sum_{k \in A^c} P_{i,k}.$$

Next, let $S_n = \min\{m : C_m = n\}$ be the time of the n^{th} aborted visit to A^c, and let $\nu_n = \mu_{S_n}$ be the empirical counting measure at that time. It is easy to see that ν_n is the urn process where, for each i, $c_i = 1$ and the distribution of $(Z_j^{(i)}; j \in A)$ is the distribution of $(\sum_{n=0}^{T-1} 1_{(V_n = j)}; j \in A)$ given $V_0 = i$. We want to apply Theorem 1.5. Hypotheses (1.1) - (1.3) are clearly satisfied. For $Q = (q_{i,j})$ defined above (1.4),

$$\begin{aligned} q_{i,j} &= E Z_j^{(i)} \\ &= E\left(\sum_{n=0}^{T-1} 1_{(V_n = j)} \mid V_0 = i\right) \\ &= E\left(\sum_{n=0}^{\infty} 1_{(V_n = j, T > n)} \mid V_0 = i\right) \end{aligned}$$

$$= \sum_{n=0}^{\infty} \hat{P}_{i,n}^{(n)}, \text{ for } \hat{P} \text{ as in (3.4).}$$

In other words,

$$Q = \sum_{n=0}^{\infty} \hat{P}^n. \tag{3.9}$$

In particular, the hypothesis of irreducibility of $\hat{P}$ implies irreducibility of Q, condition (1.4). Setting $r_{i,j} = q_{i,j} - \delta_{i,j}$ and taking (λ, α) as in (3.4), we see that (3.4) implies

$$\sum_{i \in A} \alpha_i r_{i,j} = \frac{\lambda}{1-\lambda} \alpha_j.$$

Applying Theorem 1.5 we see that α is the required eigenvector and $\frac{\nu_n}{\sum_j \nu_n(j)} \to \alpha$ a.s.

That is

$$\frac{\mu_{S_n}}{S_n} \to \alpha \text{ a.s.} \tag{3.10}$$

Since the state space is finite, $E(S_{n+1} - S_n)^2$ is bounded, and this easily implies

$$\frac{S_{n+1}}{S_n} \to 1 \text{ a.s.} \tag{3.11}$$

Noting that μ_n is monotone in n, the desired conclusion $\frac{\mu_n}{n} \to \alpha$ a.s. follows from (3.10) and (3.11) by simple analysis.

REFERENCES

Athreya, K. and Ney, P. *Branching Processes*, Springer-Verlag, New York, 1972.

Bagchi, A. and Pal, A.K. *Asymptotic Normality in the Generalized Polya-Eggenberger Urn model, with an application to Computer Data Structures.* SIAM. J. Alg. Disc. Meths. 6, (1985), pp.394-405.

Cunto, W. and Gascón, J.L. *Improving time and space efficiency in generalized binary search trees.* Acta Informatica, **24**, (1987), pp.583-594.

Darroch, J., Seneta, E. *On Quasi-Stationary Distributions in Absorbing Discrete Time Finite Markov Chains*, J. Appl. Probability, **2**, (1965), pp.88-100.

Devroye, L. *Branching Processes in the Analysis of the Heights of Trees*, Acta Informatica, **24**, (1987) pp.277-298.

Eisenbarth, B., Ziviani, N., Gonnet, G., Mehlhorn, K., and Wood, D. *The Theory of Fringe Analysis and Its Application to 2-3 Trees and B-Trees,* Information and Control, **55** (1982), pp.125-174.

Flannery, B. *On Quasi-Stationary Distributions of Markov Chains* Ph. D. Dissertation, Department of Statistics, University of California, Berkeley, 1986.

Knuth, D. *The Art of Computer Programming, volume 3,* Addison-Wesley, Reading, 1973.

Poblete, P. and Munro, J. *The Analysis of a Fringe Heuristic for Binary Search Trees,* Journal of Algorithms, **6** (1985), pp.336-350.

Yao, A. *On Random 2-3 Trees*, Acta Informatica, **9** (1978), pp.159-170.

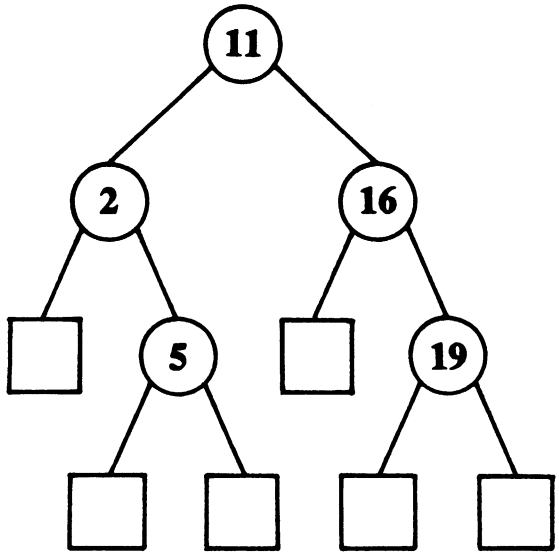


Figure 1

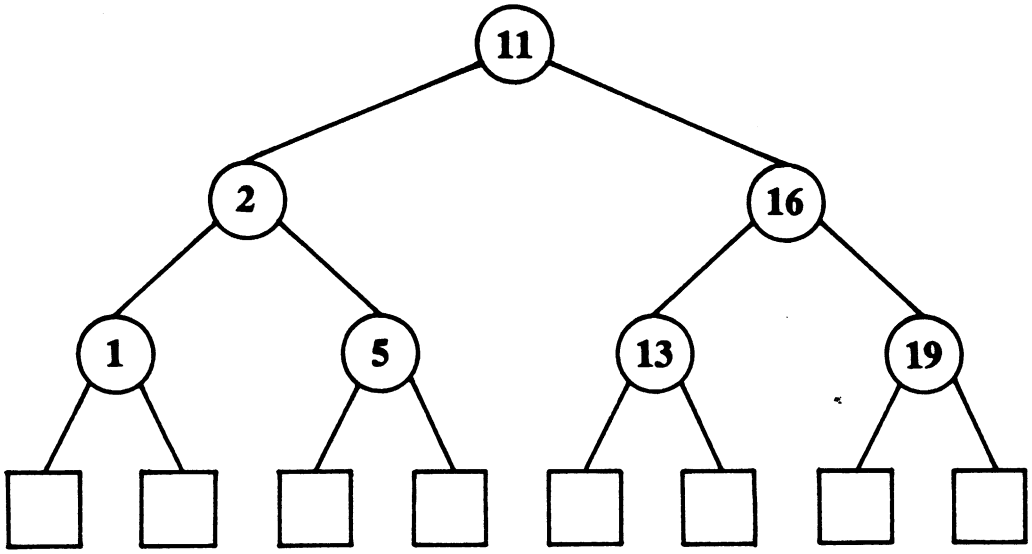


Figure 2

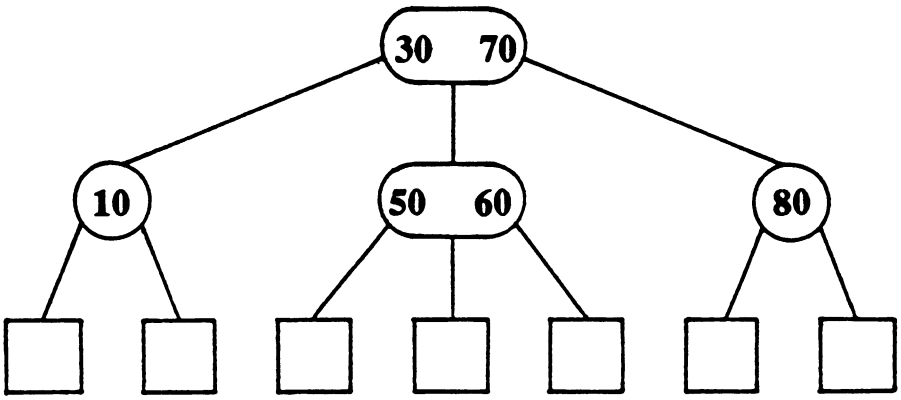


Figure 3

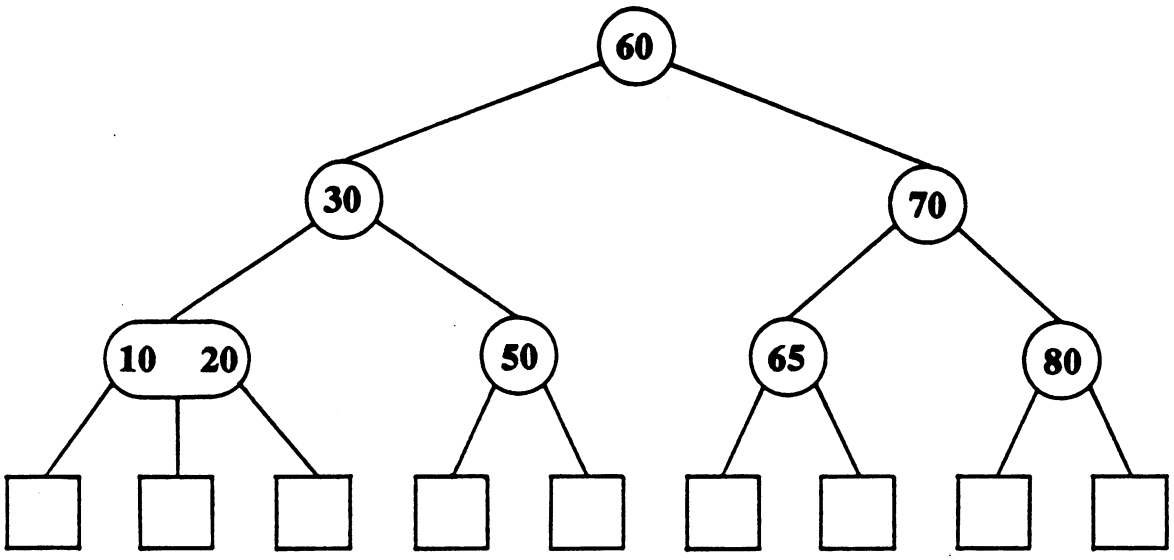


Figure 4

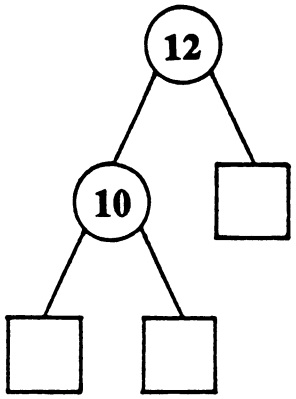


Figure 5

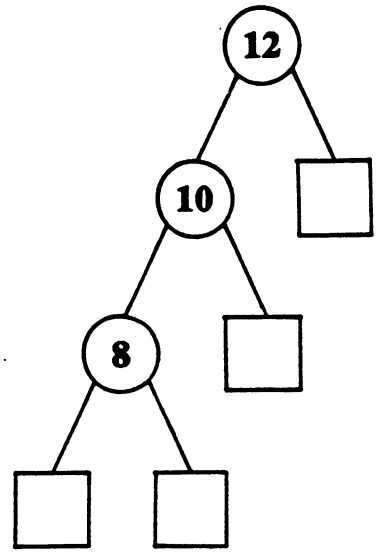


Figure 6

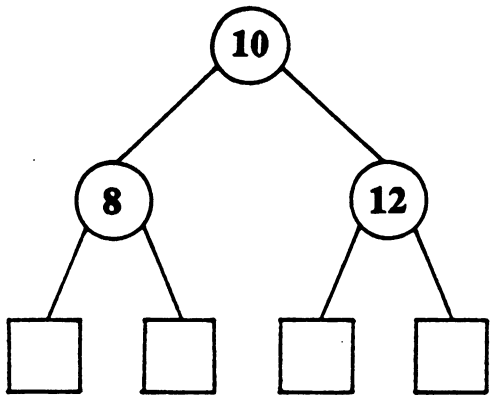
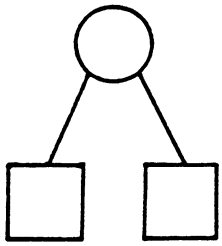
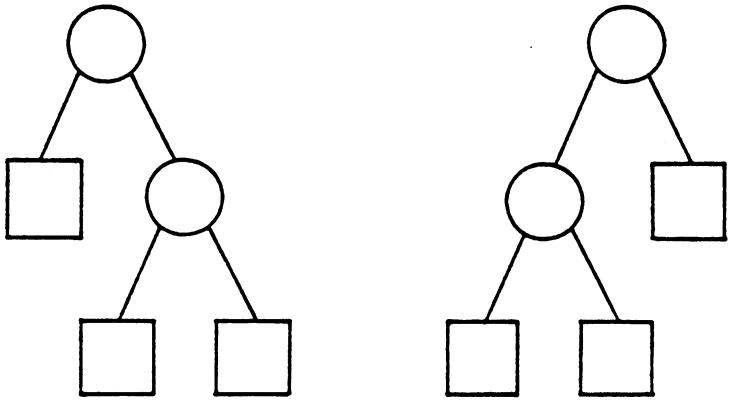


Figure 7



type 1



type 2

Figure 8

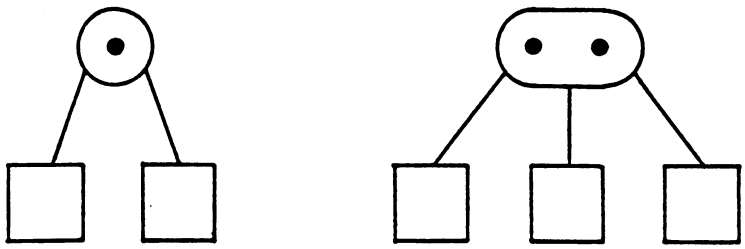


Figure 9

TECHNICAL REPORTS

Statistics Department

University of California, Berkeley

1. BREIMAN, L. and FREEDMAN, D. (Nov. 1981, revised Feb. 1982). How many variables should be entered in a regression equation? Jour. Amer. Statist. Assoc., March 1983, 78, No. 381, 131-136.

2. BRILLINGER, D. R. (Jan. 1982). Some contrasting examples of the time and frequency domain approaches to time series analysis. Time Series Methods in Hydrosciences. (A. H. El-Shaarawi and S. R. Esterby, eds.) Elsevier Scientific Publishing Co., Amsterdam, 1982, pp. 1-15.

3. DOKSUM, K. A. (Jan. 1982). On the performance of estimates in proportional hazard and log-linear models. Survival Analysis. (John Crowley and Richard A. Johnson, eds.) IMS Lecture Notes - Monograph Series, (Shanti S. Gupta, series ed.) 1982, 74-84.

4. BICKEL, P. J. and BREIMAN, L. (Feb. 1982). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. Ann. Prob., Feb. 1982, 11. No. 1, 185-214.

5. BRILLINGER, D. R. and TUKEY, J. W. (March 1982). Spectrum estimation and system identification relying on a Fourier transform. The Collected Works of J. W. Tukey, vol. 2, Wadsworth, 1985, 1001-1141.

6. BERAN, R. (May 1982). Jackknife approximation to bootstrap estimates. Ann. Statist., March 1984, 12 No. 1, 101-118.

7. BICKEL, P. J. and FREEDMAN, D. A. (June 1982). Bootstrapping regression models with many parameters. Lehmann Festschrift. (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) Wadsworth Press, Belmont, 1983, 28-48.

8. BICKEL, P. J. and COLLINS, J. (March 1982). Minimizing Fisher information over mixtures of distributions. Sankhyā, 1983, 45, Series A, Pt. 1, 1-19.

9. BREIMAN, L. and FRIEDMAN, J. (July 1982). Estimating optimal transformations for multiple regression and correlation.

10. FREEDMAN, D. A. and PETERS, S. (July 1982, revised Aug. 1983). Bootstrapping a regression equation: some empirical results. JASA, 1984, 79, 97-106.

11. EATON, M. L. and FREEDMAN, D. A. (Sept. 1982). A remark on adjusting for covariates in multiple regression.

12. BICKEL, P. J. (April 1982). Minimax estimation of the mean of a mean of a normal distribution subject to doing well at a point. Recent Advances in Statistics, Academic Press, 1983.

14. FREEDMAN, D. A., ROTHENBERG, T. and SUTCH, R. (Oct. 1982). A review of a residential energy end use model.

15. BRILLINGER, D. and PREISLER, H. (Nov. 1982). Maximum likelihood estimation in a latent variable problem. Studies in Econometrics, Time Series, and Multivariate Statistics, (eds. S. Karlin, T. Amemiya, L. A. Goodman). Academic Press, New York, 1983, pp. 31-65.

16. BICKEL, P. J. (Nov. 1982). Robust regression based on infinitesimal neighborhoods. Ann. Statist., Dec. 1984, 12, 1349-1368.

17. DRAPER, D. C. (Feb. 1983). Rank-based robust analysis of linear models. I. Exposition and review.

18. DRAPER, D. C. (Feb 1983). Rank-based robust inference in regression models with several observations per cell.

19. FREEDMAN, D. A. and FIENBERG, S. (Feb. 1983, revised April 1983). Statistics and the scientific method, Comments on and reactions to Freedman, A rejoinder to Fienberg's comments. Springer New York 1985 Cohort Analysis in Social Research. (W. M. Mason and S. E. Fienberg, eds.).

20. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Jan. 1984). Using the bootstrap to evaluate forecasting equations. J. of Forecasting, 1985, Vol. 4, 251-262.

21. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Aug. 1983). Bootstrapping an econometric model: some empirical results. JBES, 1985, 2, 150-158.

22. FREEDMAN, D. A. (March 1983). Structural-equation models: a case study.

23. DAGGETT, R. S. and FREEDMAN, D. (April 1983, revised Sept. 1983). Econometrics and the law: a case study in the proof of antitrust damages. Proc. of the Berkeley Conference, in honor of Jerzy Neyman and Jack Kiefer. Vol I pp. 123-172. (L. Le Cam, R. Olshen eds.) Wadsworth, 1985.

24. DOKSUM, K. and YANDELL, B. (April 1983). Tests for exponentiality. Handbook of Statistics, (P. R. Krishnaiah and P. K. Sen, eds.) 4, 1984.

25. FREEDMAN, D. A. (May 1983). Comments on a paper by Markus.

26. FREEDMAN, D. (Oct. 1983, revised March 1984). On bootstrapping two-stage least-squares estimates in stationary linear models. Ann. Statist., 1984, 12, 827-842.

27. DOKSUM, K. A. (Dec. 1983). An extension of partial likelihood methods for proportional hazard models to general transformation models. Ann. Statist., 1987, 15, 325-345.

28. BICKEL, P. J., GOETZE, F. and VAN ZWET, W. R. (Jan. 1984). A simple analysis of third order efficiency of estimates. Proc. of the Neyman-Kiefer Conference, (L. Le Cam, ed.) Wadsworth, 1985.

29. BICKEL, P. J. and FREEDMAN, D. A. Asymptotic normality and the bootstrap in stratified sampling. Ann. Statist. 12 470-482.

30. FREEDMAN, D. A. (Jan. 1984). The mean vs. the median: a case study in 4-R Act litigation. JBES. 1985 Vol 3 pp. 1-13.

31. STONE, C. J. (Feb. 1984). An asymptotically optimal window selection rule for kernel density estimates. Ann. Statist., Dec. 1984, 12, 1285-1297.

32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.

33. STONE, C. J. (Oct. 1984). Additive regression and other nonparametric models. Ann. Statist., 1985, 13, 689-705.

34. STONE, C. J. (June 1984). An asymptotically optimal histogram selection rule. Proc. of the Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer (L. Le Cam and R. A. Olshen, eds.), II, 513-520.

35. FREEDMAN, D. A. and NAVIDI, W. C. (Sept. 1984, revised Jan. 1985). Regression models for adjusting the 1980 Census. Statistical Science, Feb 1986, Vol. 1, No. 1, 3-39.

36. FREEDMAN, D. A. (Sept. 1984, revised Nov. 1984). De Finetti's theorem in continuous time.

37. DIACONIS, P. and FREEDMAN, D. (Oct. 1984). An elementary proof of Stirling's formula. Amer. Math Monthly, Feb 1986, Vol. 93, No. 2, 123-125.

38. LE CAM, L. (Nov. 1984). Sur l'approximation de familles de mesures par des familles Gaussiennes. Ann. Inst. Henri Poincaré, 1985, 21, 225-287.

39. DIACONIS, P. and FREEDMAN, D. A. (Nov. 1984). A note on weak star uniformities.

40. BREIMAN, L. and IHAKA, R. (Dec. 1984). Nonlinear discriminant analysis via SCALING and ACE.

41. STONE, C. J. (Jan. 1985). The dimensionality reduction principle for generalized additive models.

42. LE CAM, L. (Jan. 1985). On the normal approximation for sums of independent variables.

43. BICKEL, P. J. and YAHAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?

44. BRILLINGER, D. R. (1985). The natural variability of vital rates and associated statistics. Biometrics, to appear.

45. BRILLINGER, D. R. (1985). Fourier inference: some methods for the analysis of array and nonGaussian series data. Water Resources Bulletin, 1985, 21, 743-756.

46. BREIMAN, L. and STONE, C. J. (1985). Broad spectrum estimates and confidence intervals for tail quantiles.

47. DABROWSKA, D. M. and DOKSUM, K. A. (1985, revised March 1987). Partial likelihood in transformation models with censored data.

48. HAYCOCK, K. A. and BRILLINGER, D. R. (November 1985). LIBDRB: A subroutine library for elementary time series analysis.

49. BRILLINGER, D. R. (October 1985). Fitting cosines: some procedures and some physical examples. Joshi Festschrift. 1986. D. Reidel.

50. BRILLINGER, D. R. (November 1985). What do seismology and neurophysiology have in common? - Statistics! Comptes Rendus Math. Rep. Acad. Sci. Canada. January, 1986.

51. COX, D. D. and O'SULLIVAN, F. (October 1985). Analysis of penalized likelihood-type estimators with application to generalized smoothing in Sobolev Spaces.

52. O'SULLIVAN, F. (November 1985). A practical perspective on ill-posed inverse problems: A review with some new developments. To appear in Journal of Statistical Science.

53. LE CAM, L. and YANG, G. L. (November 1985, revised March 1987). On the preservation of local asymptotic normality under information loss.

54. BLACKWELL, D. (November 1985). Approximate normality of large products.

55. FREEDMAN, D. A. (December 1985, revised Dec. 1986). As others see us: A case study in path analysis. Prepared for the Journal of Educational Statistics.

56. LE CAM, L. and YANG, G. L. (January 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies.

57. LE CAM, L. (February 1986). On the Bernstein - von Mises theorem.

58. O'SULLIVAN, F. (January 1986). Estimation of Densities and Hazards by the Method of Penalized likelihood.

59. ALDOUS, D. and DIACONIS, P. (February 1986). Strong Uniform Times and Finite Random Walks.

60. ALDOUS, D. (March 1986). On the Markov Chain simulation Method for Uniform Combinatorial Distributions and Simulated Annealing.

61. CHENG, C-S. (April 1986). An Optimization Problem with Applications to Optimal Design Theory.

62. CHENG, C-S., MAJUMDAR, D., STUFKEN, J. & TURE, T. E. (May 1986, revised Jan 1987). Optimal step type design for comparing test treatments with a control.

63. CHENG, C-S. (May 1986, revised Jan. 1987). An Application of the Kiefer-Wolfowitz Equivalence Theorem.

64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.

65. ALDOUS, D. (JUNE 1986). Finite-Time Implications of Relaxation Times for Stochastically Monotone Processes.

66. PITMAN, J. (JULY 1986, revised November 1986). Stationary Excursions.

67. DABROWSKA, D. and DOKSUM, K. (July 1986, revised November 1986). Estimates and confidence intervals for median and mean life in the proportional hazard model with censored data.

68. LE CAM, L. and YANG, G.L. (July 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies (Fourth edition).

69. STONE, C.J. (July 1986). Asymptotic properties of logspline density estimation.

71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.

72. LEHMANN, E.L. (July 1986). Statistics - an overview.

73. STONE, C.J. (August 1986). A nonparametric framework for statistical modelling.

74. BIANE, PH. and YOR, M. (August 1986). A relation between Lévy's stochastic area formula, Legendre polynomials, and some continued fractions of Gauss.

75. LEHMANN, E.L. (August 1986, revised July 1987). Comparing Location Experiments.

76. O'SULLIVAN, F. (September 1986). Relative risk estimation.

77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.

78. PITMAN, J. & YOR, M. (September 1987). Further asymptotic laws of planar Brownian motion.

79. FREEDMAN, D.A. & ZEISEL, H. (November 1986). From mouse to man: The quantitative assessment of cancer risks.

80. BRILLINGER, D.R. (October 1986). Maximum likelihood analysis of spike trains of interacting nerve cells.

81. DABROWSKA, D.M. (November 1986). Nonparametric regression with censored survival time data.

82. DOKSUM, K.J. and LO, A.Y. (November 1986). Consistent and robust Bayes Procedures for Location based on Partial Information.

83. DABROWSKA, D.M., DOKSUM, K.A. and MIURA, R. (November 1986). Rank estimates in a class of semiparametric two-sample models.

84. BRILLINGER, D. (December 1986). Some statistical methods for random process data from seismology and neurophysiology.

85. DIACONIS, P. and FREEDMAN, D. (December 1986). A dozen de Finetti-style results in search of a theory.

86. DABROWSKA, D.M. (January 1987). Uniform consistency of nearest neighbour and kernel conditional Kaplan - Meier estimates.

87. FREEDMAN, D.A., NAVIDI, W. and PETERS, S.C. (February 1987). On the impact of variable selection in fitting regression equations.

88. ALDOUS, D. (February 1987, revised April 1987). Hashing with linear probing, under non-uniform probabilities.

89. DABROWSKA, D.M. and DOKSUM, K.A. (March 1987, revised November 1987). Estimating and testing in a two sample generalized odds rate model.

90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.

91. DIACONIS, P and FREEDMAN, D.A. (March 1987). A finite version of de Finetti's theorem for exponential families, with uniform asymptotic estimates.

92. DABROWSKA, D.M. (April 1987, revised September 1987). Kaplan-Meier estimate on the plane.

92a. ALDOUS, D. (April 1987). The Harmonic mean formula for probabilities of Unions: Applications to sparse random graphs.

93. DABROWSKA, D.M. (June 1987). Nonparametric quantile regression with censored data.

94. DONOHO, D.L. & STARK, P.B. (June 1987). Uncertainty principles and signal recovery.

95. RIZZARDI, F. (Aug 1987). Two-Sample t-tests where one population SD is known.

96. BRILLINGER, D.R. (June 1987). Some examples of the statistical analysis of seismological data. To appear in *Proceedings, Centennial Anniversary Symposium, Seismographic Stations, University of California, Berkeley.*

97. FREEDMAN, D.A. and NAVIDI, W. (June 1987). On the multi-stage model for cancer.

98. O'SULLIVAN, F. and WONG, T. (June 1987). Determining a function diffusion coefficient in the heat equation.

99. O'SULLIVAN, F. (June 1987). Constrained non-linear regularization with application to some system identification problems.

100. LE CAM, L. (July 1987, revised Nov 1987). On the standard asymptotic confidence ellipsoids of Wald.

101. DONOHO, D.L. and LIU, R.C. (July 1987). Pathologies of some minimum distance estimators.

102. BRILLINGER, D.R., DOWNING, K.H. and GLAESER, R.M. (July 1987). Some statistical aspects of low-dose electron imaging of crystals.

103. LE CAM, L. (August 1987). Harald Cramér and sums of independent random variables.

104. DONOHO, A.W., DONOHO, D.L. and GASKO, M. (August 1987). Macspin: Dynamic graphics on a desktop computer.

105. DONOHO, D.L. and LIU, R.C. (August 1987). On minimax estimation of linear functionals.

106. DABROWSKA, D.M. (August 1987). Kaplan-Meier estimate on the plane: weak convergence, LIL and the bootstrap.

107. CHENG, C-S. (August 1987). Some orthogonal main-effect plans for asymmetrical factorials.

108. CHENG, C-S. and JACROUX, M. (August 1987). On the construction of trend-free run orders of two-level factorial designs.

109. KLASS, M.J. (August 1987). Maximizing $E \max_{1 \le k \le n} S_k^+ / ES_n^+$: A prophet inequality for sums of I.I.D. mean zero variates.

110. DONOHO, D.L. and LIU, R.C. (August 1987). The "automatic" robustness of minimum distance functionals.

111. BICKEL, P.J. and GHOSH, J.K. (August 1987). A decomposition for the likelihood ratio statistic and the Bartlett correction — a Bayesian argument.

112. BURDZY, K., PITMAN, J.W. and YOR, M. (September 1987). Some asymptotic laws for crossings and excursions.

113. ADHIKARI, A. and PITMAN, J. (September 1987). The shortest planar arc of width 1.

114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.

115. BICKEL, P.J. and RITOV, Y. (September 1987). Large sample theory of estimation in biased sampling regression models I.

116. RITOV, Y. and BICKEL, P.J. (September 1987). Unachievable information bounds in non and semiparametric models.

117. RITOV, Y. (October 1987). On the convergence of a maximal correlation algorithm with alternating projections.

118. ALDOUS, D.J. (October 1987). Meeting times for independent Markov chains.

119. HESSE, C.H. (October 1987). An asymptotic expansion for the mean of the passage-time distribution of integrated Brownian Motion.

120. DONOHO, D. (October 1987). Geometrizing rates of convergence, II.

121. BRILLINGER, D.R. (October 1987). Estimating the chances of large earthquakes by radiocarbon dating and statistical modelling. To appear in *Statistics a Guide to the Unknown.*

122. ALDOUS, D., FLANNERY, B. and PALACIOS, J.L. (November 1987). Two applications of urn processes: The fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains.

123. DONOHO, D.L. and MACGIBBON, B. (November 1987). Minimax risk for hyperrectangles.

124. ALDOUS, D. (November 1987). Stopping times and tightness II.

125. HESSE, C.H. (November 1987). The present state of a stochastic model for sedimentation.

126. DALANG, R.C. (December 1987). Optimal stopping of two-parameter processes on hyperfinite probability spaces.

128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.

129. SMITH, D.L. (December 1987). Vapnik-Červonenkis classes.

130. STONE, C.J. (November 1987). Uniform error bounds involving logspline models.

131. HESSE, C.H. (December 1987). Numerical and statistical aspects of neural networks.

132. HESSE, C.H. (December 1987). A Bahadur - Type representation for empirical quantiles of a large class of stationary, possibly infinite - variance, linear processes

133. DONOHO, D.L. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean, I.

134. DUBINS, L.E. and SCHWARZ, G. (December 1987). A sharp inequality for martingales and stopping-times.

Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:

Department of Statistics
University of California
Berkeley, California 94720

Cost: $1 per copy.