A Time-Dependent Version of Pólya's Urn

Robin Pemantle
Department of Statistics
U.C. Berkeley
Berkeley, CA 94720

Research supported by an NSF Graduate & an NSF Postdoctoral fellowship

Technical Report No. 192
February 1989

Department of Statistics
University of California
Berkeley, California

A TIME-DEPENDENT VERSION OF PÓLYA'S URN

Robin Pemantle
Dept. of Statistics
U.C. Berkeley
Berkeley, CA 94720

ABSTRACT:

A process is defined that consists of drawing balls from an urn and replacing them along with extra balls. The process generalizes the well-known *Pólya urn process.* It is shown that the proportion of red balls in the urn converges to a random limit that may have a nonzero probability of being 0 or 1, but is nonatomic elsewhere.

The Pólya urn process was introduced in Eggenberger and Pólya (1923). To run this process, let an urn contain R red balls and B black balls at time $n = 1$. A ball is drawn at random from the urn and replaced in the urn along with another ball of the same color, so that at time $n = 2$ there are three balls, with red ounumbering black by two to one half the time and black outnumbering red by two to one the rest of the time. The draw and replacement are repeated ad infinitum, with the probability of drawing a red ball always equal to the proportion of balls in the urn that are red at that time. It is a well-known fact that the proportions of red balls converge almost surely to a limit that is random and has beta distribution with parameters R and B; see Feller (1957) for a discussion of this.

This paper considers a Pólya urn with the single change that the number of extra balls added of the color drawn is a function of time. For some other generalizations of the Pólya urn process see Friedman (1949) and Hill, Lane and Sudderth (1980). Let $F : \mathbf{Z}^{\geq 0} \to (0, \infty)$ be any function. Let $\{v_1, v_2, \ldots\}$ be the successive proportions of red balls in an urn that begins with one red ball and one black ball and evolves as follows: at discrete times $n = 1, 2, \ldots$, a ball is drawn and replaced in the urn along with $F(n)$ balls of the same color ($F(n)$ need not be an integer). The usual Pólya urn scheme is the case where $F(n) = 1$ for all n. We show that v_n must converge for any F and that the limit has no atoms except possibly at 0 and 1. Necessary and sufficient conditions for the limit to concentrate entirely on the two point set $\{0, 1\}$ are given. The proofs are based on variance calculations for a discrete-time martingale and are completely elementary.

The following formal definition of the process is completely routine and may be skipped. Let $\mathbf{\Omega}$ be $[0, 1]^{\mathbf{Z}^{\geq 0}}$ with the product uniform measure. All probabilities will be with respect to this space and all functions will be functions of ω where ω is a generic point in $\mathbf{\Omega}$, but the notation will supress the role of ω when no ambiguity arises. Let z_n be the n^{th} coordinate of ω so that $\{z_n : n = 0, 1, 2, \ldots\}$ is a set of independent uniformly distributed variables on $[0, 1]$, and let $\mathcal{F}_n$ be the σ-algebra generated by $\{z_i : i \leq n\}$. Let

$S_1(0) = S_2(0) = 1$ and recursively define

$$\begin{aligned} S_1(n+1) &= S_1(n) + F(n)\mathbf{1}(z_n \leq S_1(n)/(S_1(n)+S_2(n))) \\ S_2(n+1) &= S_2(n) + F(n)\mathbf{1}(z_n > S_1(n)/(S_1(n)+S_2(n))) \end{aligned} \tag{1}$$

where $\mathbf{1}$ denotes the indicator function of a set. So $S_1(n)$ and $S_2(n)$ represent the numbers of red and black balls in the urn after n draws. For convenience we let

$$\delta_n = F(n)/(2 + \sum_{i=0}^{n-1} F(i))$$

denote the fractional additions. Let $v_n = S_1(n)/(S_1(n) + S_2(n))$ denote the proportion of red balls at time n. The following results will be proved.

Theorem 1 *For any function F as above, the random variables v_n converge almost surely to some random variable v.*

Theorem 2 *The limit v satisfies prob$(v = 0)$ = prob$(v = 1) = 1/2$ if and only if $\sum_{i=1}^{\infty} {\delta_n}^2 = \infty$.*

This theorem applies, for example, when $F(n) = 2^n$. Roughly, the hypothesis means that F grows faster than polynomially, but one needs to look more closely if the growth is irregular since the function

$$F(n) = \begin{cases} n \text{ if } n \text{ is a power of } 2; \\ 2^{-n} \text{ otherwise ;} \end{cases}$$

satisfies the hypothesis.

Theorem 3 *The distribution of v has no atoms on $(0, 1)$.*

Remark: It is possible for the distribution of v to have atoms at 0 and 1 of weight less than 1/2 each; then the remainder of the time v is in $(0,1)$ and this part of the distribution is nonatomic. An example where this occurs is if $F(n) = n$. In this case the probability that all draws are of the same color is $\frac{2}{3} \times \frac{6}{7} \times \frac{15}{16} \times \cdots > 0$, but according to Theorem 2 the distribution is not entirely concentrated on $\{0,1\}$. I do not know of a counterpart to Theorem 2 giving a necessary condition for the probability that $v \in \{0,1\}$ to be nonzero.

Proof of Theorem 1: $\{v_n : n = 1,2,\ldots\}$ is a martingale. To see this, calculate

$$\begin{aligned}
\mathbf{E}(v_{n+1} \mid \mathcal{F}_n) &= v_n(S_1(n) + F(n))/(S_1(n) + F(n) + S_2(n)) \\
&+ (1 - v_n)S_1(n)/(S_1(n) + S_2(n) + F(n)) \\
&= [S_1(n) + v_n F(n)]/(S_1(n) + S_2(n) + F(n)) \\
&= v_n.
\end{aligned}$$

Now since $\{v_n\}$ is bounded, it converges almost surely to some v. □

Proof of Theorem 2: We calculate the expected value of v^2. By symmetry this is at most 1/2 with equality if and only if $prob(v = 0) = prob(v = 1) = 1/2$. Necessary and sufficient conditions for this will follow from the simple recurrence relation (2) below for the values of $1/2 - \mathbf{E}({v_n}^2)$, which are denoted W_n.

Since v_n converges almost surely to v and the variables are bounded by 1, we know that $\mathbf{E}(v_n^2)$ converges to $\mathbf{E}(v^2)$. Let V_n denote $\mathbf{E}({v_n}^2)$. For a fixed F, v_n takes on only finitely many values and V_n can be recursively calculated as follows. If $v_{n-1}(\omega) = x = S_1(n-1)/(S_1(n-1) + S_2(n-1))$ then

$$\begin{aligned}
v_n(\omega) &= S_1(n-1)/(S_1(n-1) + S_2(n-1) + F(n)) \\
&= x/(1 + \delta_n)
\end{aligned}$$

with probability $1 - x$, and

$$\begin{aligned}
v_n(\omega) &= (S_1(n-1) + F(n))/(S_1(n-1) + S_2(n-1) + F(n)) \\
&= (x + \delta_n)/(1 + \delta_n)
\end{aligned}$$

with probability x. So

$$\begin{aligned}
V_n &= \mathbf{E}{v_n}^2 \\
&= \mathbf{E}[(1-v_{n-1}){v_{n-1}}^2/(1+\delta_n)^2 + v_{n-1}(v_{n-1}+\delta_n)^2/(1+\delta_n)^2] \\
&= 1/(1+\delta_n)^2\mathbf{E}[(1-v_{n-1}){v_{n-1}}^2 + v_{n-1}(v_{n-1}+\delta_n)^2] \\
&= 1/(1+\delta_n)^2\mathbf{E}[{v_{n-1}}^2 + 2\delta_n {v_{n-1}}^2 + v_{n-1}\delta_n^2] \\
&= (\delta_n^2/2 + V_{n-1}(1+2\delta_n))/(1+\delta_n)^2.
\end{aligned}$$

To see better how the value of V_n relates to the value of V_{n-1}, we let W_k denote $1/2 - V_k$. Then

$$\begin{aligned}
W_n &= 1/2 - [(\delta_n^2/2) + V_{n-1}(1+2\delta_n)]/(1+\delta_n)^2 \\
&= (1/2 + \delta_n - V_{n-1} - 2\delta_n V_{n-1})/(1+\delta_n)^2 \\
&= W_{n-1}(1+2\delta_n)/(1+\delta_n)^2 \\
&= W_{n-1}(1-\delta_n^2/(1+\delta_n)^2). \qquad (2)
\end{aligned}$$

Thus the value V_n converges to $1/2$ if and only if W_n converges to 0, which happens whenever the product of the values $(1-\delta_n^2/(1+\delta_n)^2)$ converges to 0. This happens whenever $\sum_{n=1}^\infty \delta_n^2/(1+\delta_n)^2$ diverges, which in turn happens exactly when $\sum_{n=1}^\infty \delta_n^2$ diverges, and Theorem 2 is proved. □

Proof of Theorem 3: Fix $p \in (0,1)$. If $prob(v_n \to p) > 0$ then there is some n and some event $\mathcal{A} \in \mathcal{F}_n$ such that $prob(v_n \to p \,|\, \mathcal{A})$ is arbitrarily close to 1. In fact, n can be taken to be as large as desired. Define

$$\alpha_n = \sum_{i=n}^\infty {\delta_i}^2.$$

The quantity α_n can be thought of as the "remaining variance", since the expected square increments of the martingale $\{v_i\}$ are bounded between constant multiples of ${\delta_i}^2$ when v_i is near p. According to Theorem 2 there is no loss of generality in assuming α_n to be finite. Also assume without loss of generality that $p \le 1/2$ since the case $p > 1/2$ is identical but with red balls and black balls interchanged.

Since $\alpha_n \to 0$ there is an N for which $n \geq N$ implies $\alpha_n < p/10$. Choose $c > 0$ small enough so that

$$9c^2 \leq 81p^2/800. \tag{3}$$

The essence of the proof is in the following two claims, holding for any $n > N$.

$$\text{Claim 1: } prob(\sup_{k\geq n} |v_k - p| > c\sqrt{\alpha_n} \,|\, \mathcal{F}_n) \geq 9p/10 \tag{4}$$

$$\text{Claim 2: } prob(\inf_{k\geq n} |v_k - p| \geq c\sqrt{\alpha_n}/2 \,|\, \mathcal{F}_n, \mathcal{B}) \geq c^2/16 \tag{5}$$

where $\mathcal{B}$ is the event $|v_n - p| \geq c\sqrt{\alpha_n}$.

Putting these two claims together, we see that for any value of $n > N$, the probability geven $\mathcal{F}_n$ is at least $9p/10 \cdot c^2/16$ that some v_{n+k} will be at least $c\sqrt{\alpha_n}$ away from p and that no subsequent v_{n+k+l} will ever return to the interval $[p - c\sqrt{\alpha_n}/2\,,\, p + c\sqrt{\alpha_n}/2]$. This contradicts the existence of the event $\mathcal{A}$ above, and the theorem follows.

Proof of Claim 1: Let $\tau = \inf\{k \geq n : |v_k - p| > c\sqrt{\alpha_n}\}$. We need to show that $prob(\tau < \infty \,|\, \mathcal{F}_n) \geq 9p/10$. We will calculate the variance of $v_{i\wedge\tau}$. On the one hand, this is limited by the fact that $v_{i\wedge\tau}$ is never very far from p. (If the increment on which the stopping time is reached may be very large, then a different argument is used; see Case 1 below.) On the other hand, the variance always grows by at least a constant multiple of $\delta_i{}^2$ until τ is reached, and c is chosen to be much smaller than this constant. These two facts together will imply that the stopping time is reached often enough for Claim 1 to be true.

Case 1: $\delta_i > 2c\sqrt{\alpha_n}/(1 - p - c\sqrt{\alpha_n})$ for some $i \geq n$. Basically what happens in this case is that there is a good enough chance of stopping on the $i + 1^{st}$ draw:

$$\tau > i \text{ and draw } i \text{ is red}$$

$$\Rightarrow \quad v_i \geq p - c\sqrt{\alpha_n} \text{ and draw } i \text{ is red}$$

$$\Rightarrow \quad v_{i+1} = v_i + (1 - v_i)\delta_i/(1 + \delta_i) \geq p + c\sqrt{\alpha_n}$$

$$\Rightarrow \quad \tau = i + 1.$$

The probability of a red draw is always at least $p - c\sqrt{\alpha_i}$ until τ is reached,which is at least $> 9p/10$ by choice of c and N. This easily implies that $prob(\tau < \infty \,|\, \mathcal{F}_n) \geq 9p/10$.

Case 2: No δ_i is that big. Then the increment on which τ is reached cannot be bigger than $2c\sqrt{\alpha_n}$, and so

$$|v_{i\wedge\tau} - p| \leq 3c\sqrt{\alpha_n} \text{ for all } i \geq n. \tag{6}$$

Pick any $i \geq n$ and use the fact that $v_{i\wedge\tau} - p$ is a martingale to get

$$\mathbf{E}((v_{(i+1)\wedge\tau} - p)^2|\mathcal{F}_n) = \mathbf{E}((v_{i\wedge\tau} - p)^2|\mathcal{F}_n) + \mathbf{E}(\mathbf{1}_{\tau>i}(v_{i+1} - v_i)^2|\mathcal{F}_n). \tag{7}$$

But

$$(v_{i+1} - v_i)^2 = \begin{cases} v_i^2(\delta_i/(1+\delta_i))^2 \text{ with probability } 1 - v_i, \\ (1 - v_i)^2(\delta_i/(1+\delta_i))^2 \text{ with probability } v_i. \end{cases} \tag{8}$$

Now since $1+\delta_i < 2$ by the assumption that $\alpha_n < p/10$ and since $\tau > i \Rightarrow \min\{v_i, 1-v_i\} \geq p - c\sqrt{\alpha_n} \geq 9p/10$, it follows that

$$(v_{i+1} - v_i)^2 \geq 81p^2\delta_i^2\mathbf{1}_{\tau>i}/400.$$

So the right hand side of equation (7) is at least

$$\mathbf{E}((v_{i\wedge\tau} - p)^2|\mathcal{F}_n) + 81\delta_i^2 prob(\tau = \infty|\mathcal{F}_n)/400.$$

Now summing over i and dropping the positive term $(v_n - p)^2$ gives

$$\mathbf{E}((v_{(n+M)\wedge\tau} - p)^2|\mathcal{F}_n) \geq \left(81p^2 \sum_{i=n}^{n+M-1} \delta_i^2/400\right) prob(\tau = \infty|\mathcal{F}_n).$$

But equation (6) implies that

$$\mathbf{E}((v_{(n+M)\wedge\tau} - p)^2|\mathcal{F}_n) \leq 9c^2\alpha_n = 9c^2\sum_{i=n}^{\infty}\delta_i^2,$$

so letting $M \to \infty$ gives

$$prob(\tau = \infty \,|\, \mathcal{F}_n) \leq 9c^2/(81p^2/400) \leq 1/2$$

by the choice of c in (3) above. So Claim 1 is proved. □.

Proof of Claim 2: The idea this time is that the remaining variance is not enough to give a high probability of getting back to within $c\sqrt{\alpha_n}/2$ of p. The inequality we use is a one-sided Tschebysheff inequality saying that if v_n has a probability greater than $1-\epsilon$ of reentering the interval, then since it is a martingale, the other ϵ of the time its average is on the order of ϵ^{-1} in the other direction, and this gives a contribution to the variance that gets impossibly large as ϵ goes to 0.

Let $\mathcal{B}$ be the event $|v_n - p| \geq c\sqrt{\alpha_n}$ as in (5) above. Define a new stopping time by $\tau = \inf\{k \geq n \,;\, |v_k - p| \leq c\sqrt{\alpha_n}/2\}$. From (7) again, calculate

$$\mathrm{Var}(v_{(n+M)\wedge\tau}|\mathcal{F}_n) = \sum_{i=n}^{n+M-1}\mathbf{E}(\mathbf{1}_{\tau>i}(v_{i+1} - v_i)^2|\mathcal{F}_n) \leq \sum_{i=n}^{\infty}\delta_i^2 \tag{9}$$

according to the values for $(v_{i+1} - v_i)^2$ given in (8). So $\{v_{(n+i)\wedge\tau}\}$ is an L^2-bounded martingale with variance $\mathbf{E}((v_\tau - v_n)^2 \,|\, \mathcal{F}_n)$ at most $\sum_{i=n}^{\infty}\delta_i^2 = \alpha_n$. On the other hand,

$$\begin{aligned}
&\mathbf{E}((v_\tau - v_n)^2|\mathcal{F}_n, \mathcal{B}) \\
\geq\ & prob(\tau < \infty \,|\, \mathcal{F}_n, \mathcal{B})\,(c\sqrt{\alpha_n}/2)^2 + prob(\tau = \infty \,|\, \mathcal{F}_n, \mathcal{B})\mathbf{E}((v_\infty - v_n)^2 \,|\, \mathcal{F}_n, \mathcal{B}, \tau = \infty) \\
\geq\ & prob(\tau = \infty \,|\, \mathcal{F}_n, \mathcal{B})\mathbf{E}(v_\infty - v_n \,|\, \mathcal{F}_n, \mathcal{B}, \tau = \infty)^2 \\
\geq\ & prob(\tau = \infty \,|\, \mathcal{F}_n, \mathcal{B})\left(\frac{c\sqrt{\alpha_n}}{2}\frac{prob(\tau < \infty \,|\, \mathcal{F}_n, \mathcal{B})}{prob(\tau = \infty \,|\, \mathcal{F}_n, \mathcal{B})}\right)^2 \\
=\ & \frac{c^2\alpha_n}{4}\frac{prob(\tau < \infty \,|\, \mathcal{F}_n, \mathcal{B})^2}{prob(\tau = \infty \,|\, \mathcal{F}_n, \mathcal{B})}
\end{aligned} \tag{10}$$

where the penultimate term is calculated from the fact that $|\mathbf{E}(v_\infty - v_n \,|\, \mathcal{F}_n, \mathcal{B}, \tau < \infty)| > c\sqrt{\alpha_n}/2$ while $\mathbf{E}(v_\infty - v_n)$ must be zero. Combining the two inequalities (9) and (10) gives

$$\alpha_n \geq (c^2/4)\alpha_n \frac{prob(\tau < \infty \,|\, \mathcal{F}_n, \mathcal{B})^2}{prob(\tau = \infty \,|\, \mathcal{F}_n, \mathcal{B})}.$$

It follows easily from this that $prob(\tau = \infty \,|\, \mathcal{F}_n, \mathcal{B}) \geq \min\{1/2, c^2/16\} = c^2/16$ and Claim 2 is proved, along with thoerem 3. □

Knowing that the distribution of v is nonatomic on $(0, 1)$, it is logical to ask when the distribution is absolutely continuous with respect to Lebesgue measure. Nothing is known about this except when F is constant and the distribution of v is known to be a beta, or when $F(n)$ goes to zero faster than 2^{-n} and v is supported on a Cantor set.

Bibliography

[1] Eggenberger, F. and Pólya, G. (1923). Über die Statistik Verketter Vorgänge. *Zeitschrift für Angewandte Mathematik und Mechanik* **3** 279-289.

[2] Feller, W. (1966). Introduction to probability theory and its applications, volume 2. John Wiley & Sons: New York.

[3] Friedman, B. (1949). A simple urn model. *Comm. Pure Appl. Math.* **2** 59 - 70.

[4] Hill, B., Lane, D. and Sudderth, W. (1980). A strong law for some generalized urn processes. *Ann. Prob.* **8** 214 - 226.

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. Statistical Science, 1988, Vol.3 No. 2 239-271.

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, 579-611.

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 estimate 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. Scandinavian J. Statist., 1988, 15, 1-23.

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. (June 1987). As others see us: A case study in path analysis. Journal of Educational Statistics. 12, 101-128.

56. LE CAM, L. and YANG, G. L. (January 1986). Replaced by No. 68.

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. Biometrika, 1987, 74, 799-808.

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 polynomial, 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. To appear in Statistical Science.

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. (Nov 1986, revised Aug 1988). 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. Ann. Inst. Henri Poincaré, 1987, 23, 397-423.

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 January 1988). Estimating and testing in a two sample generalized odds rate model. J. Amer. Statist. Assoc., 1988, 83, 744-749.

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

91. DIACONIS, P and FREEDMAN, D.A. (April 1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. To appear in the Journal of Applied Probability.

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, revised Feb 1988). Nonparametric quantile regression with censored data.

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

95. CANCELLED

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 carcinogenesis. To appear in Environmental Health Perspectives.

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. Annals of Statistics, June, 1988.

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. IEEE Computer Graphics and applications, June, 1988.

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. (Aug 1987, revised Oct 1988). 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. Annals of Statistics, June, 1988.

111. BICKEL, P.J. and GHOSH, J.K. (August 1987, revised June 1988). 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. (Sept. 1987, revised Aug 1988). Large sample theory of estimation in biased sampling regression models I.

116. RITOV, Y. and BICKEL, P.J. (Sept.1987, revised Aug. 1988). Achieving 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. and LIU, R. (Oct. 1987, revised Mar. 1988, Oct. 1988). 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., MACGIBBON, B. and LIU, R.C. (Nov.1987, revised July 1988). 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, revised June 1988). Optimal stopping of two-parameter processes on nonstandard probability spaces.

127. Same as No. 133.

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

129. SMITH, D.L. (December 1987). Exponential bounds in Vapnik-Červonenkis classes of index 1.

130. STONE, C.J. (Nov.1987, revised Sept. 1988). Uniform error bounds involving logspline models.

131. Same as No. 140

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.

135. FREEDMAN, D.A. and NAVIDI, W. (December 1987). On the risk of lung cancer for ex-smokers.

136. LE CAM, L. (January 1988). On some stochastic models of the effects of radiation on cell survival.

137. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the uniform consistency of Bayes estimates for multinomial probabilities.

137a. DONOHO, D.L. and LIU, R.C. (1987). Geometrizing rates of convergence, I.

138. DONOHO, D.L. and LIU, R.C. (January 1988). Geometrizing rates of convergence, III.

139. BERAN, R. (January 1988). Refining simultaneous confidence sets.

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

141. BRILLINGER, D.R. (January 1988). Two reports on trend analysis: a) An Elementary Trend Analysis of Rio Negro Levels at Manaus, 1903-1985 b) Consistent Detection of a Monotonic Trend Superposed on a Stationary Time Series

142. DONOHO, D.L. (Jan. 1985, revised Jan. 1988). One-sided inference about functionals of a density.

143. DALANG, R.C. (Feb. 1988, revised Nov. 1988). Randomization in the two-armed bandit problem.

144. DABROWSKA, D.M., DOKSUM, K.A. and SONG, J.K. (February 1988). Graphical comparisons of cumulative hazards for two populations.

145. ALDOUS, D.J. (February 1988). Lower bounds for covering times for reversible Markov Chains and random walks on graphs.

146. BICKEL, P.J. and RITOV, Y. (Feb.1988, revised August 1988). Estimating integrated squared density derivatives.

147. STARK, P.B. (March 1988). Strict bounds and applications.

148. DONOHO, D.L. and STARK, P.B. (March 1988). Rearrangements and smoothing.

149. NOLAN, D. (March 1988). Asymptotics for a multivariate location estimator.

150. SEILLIER, F. (March 1988). Sequential probability forecasts and the probability integral transform.

151. NOLAN, D. (March 1988). Limit theorems for a random convex set.

152. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On a theorem of Kuchler and Lauritzen.

153. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the problem of types.

154. DOKSUM, K.A. (May 1988). On the correspondence between models in binary regression analysis and survival analysis.

155. LEHMANN, E.L. (May 1988). Jerzy Neyman, 1894-1981.

156. ALDOUS, D.J. (May 1988). Stein's method in a two-dimensional coverage problem.

157. FAN, J. (June 1988). On the optimal rates of convergence for nonparametric deconvolution problem.

158. DABROWSKA, D. (June 1988). Signed-rank tests for censored matched pairs.

159. BERAN, R.J. and MILLAR, P.W. (June 1988). Multivariate symmetry models.

160. BERAN, R.J. and MILLAR, P.W. (June 1988). Tests of fit for logistic models.

161. BREIMAN, L. and PETERS, S. (June 1988). Comparing automatic bivariate smoothers (A public service enterprise).

162. FAN, J. (June 1988). Optimal global rates of convergence for nonparametric deconvolution problem.

163. DIACONIS, P. and FREEDMAN, D.A. (June 1988). A singular measure which is locally uniform. (Revised by Tech Report No. 180).

164. BICKEL, P.J. and KRIEGER, A.M. (July 1988). Confidence bands for a distribution function using the bootstrap.

165. HESSE, C.H. (July 1988). New methods in the analysis of economic time series I.

166. FAN, JIANQING (July 1988). Nonparametric estimation of quadratic functionals in Gaussian white noise.

167. BREIMAN, L., STONE, C.J. and KOOPERBERG, C. (August 1988). Confidence bounds for extreme quantiles.

168. LE CAM, L. (August 1988). Maximum likelihood an introduction.

169. BREIMAN, L. (August 1988). Submodel selection and evaluation in regression-The conditional case and little bootstrap.

170. LE CAM, L. (September 1988). On the Prokhorov distance between the empirical process and the associated Gaussian bridge.

171. STONE, C.J. (September 1988). Large-sample inference for logspline models.

172. ADLER, R.J. and EPSTEIN, R. (September 1988). Intersection local times for infinite systems of planar brownian motions and for the brownian density process.

173. MILLAR, P.W. (October 1988). Optimal estimation in the non-parametric multiplicative intensity model.

174. YOR, M. (October 1988). Interwinings of Bessel processes.

175. ROJO, J. (October 1988). On the concept of tail-heaviness.

176. ABRAHAMS, D.M. and RIZZARDI, F. (September 1988). BLSS - The Berkeley interactive statistical system: An overview.

177. MILLAR, P.W. (October 1988). Gamma-funnels in the domain of a probability, with statistical implications.

178. DONOHO, D.L. and LIU, R.C. (October 1988). Hardest one-dimensional subfamilies.

179. DONOHO, D.L. and STARK, P.B. (October 1988). Recovery of sparse signals from data missing low frequencies.

180. FREEDMAN, D.A. and PITMAN, J.A. (Nov. 1988). A measure which is singular and uniformly locally uniform. (Revision of Tech Report No. 163).

181. DOKSUM, K.A. and HOYLAND, ARNLJOT (Nov. 1988, revised Jan. 1989). A model for step-stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution.

182. DALANG, R.C., MORTON, A. and WILLINGER, W. (November 1988). Equivalent martingale measures and no-arbitrage in stochastic securities market models.

183. BERAN, R. (November 1988). Calibrating prediction regions.

184. BARLOW, M.T., PITMAN, J. and YOR, M. (Dec. 1988). On Walsh's Brownian Motions.

185. DALANG, R.C. and WALSH, J.B. (Dec. 1988). Almost-equivalence of the germ-field Markov property and the sharp Markov property of the Brownian sheet.

186. HESSE, C.H. (Dec. 1988). Level-Crossing of integrated Ornstein-Uhlenbeck processes

187. PITMAN, J.W. (Dec. 1988). Neveu's branching process in Brownian excursions.

188. PITMAN, J.W. (Dec. 1988). The equilibrium law of Brownian extrema.

189. BARLOW, M.T., PITMAN, J.W. and Yor, M. (Jan. 1989). Some extensions of the arcsine law.

190. STARK, P.B. (Dec. 1988). Duality and discretization in linear inverse problems.

191. LEHMANN, E.L. and SCHOLZ, F.W. (Jan. 1989). Ancillarity.

192. PEMANTLE, R. (Feb. 1989). A time-dependent version of Pólya's urn.

193. PEMANTLE, R. (Feb. 1989). Nonconvergence to unstable points in urn models and stochastic approximations.

194. PEMANTLE, R. (Feb. 1989). When are touchpoints limits for generalized Pólya urns.

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.