APPROXIMATE NORMALITY OF LARGE PRODUCTS

by

DAVID BLACKWELL

6

Technical Report No. 54

December 1985

Department of Statistics
University of California
Berkeley, California

Approximate Normality of Large Products

D. Blackwell

It is well-known that the beta density, proportional to

$$P(x)=x^A(1-x)^B, \quad 0<x<1,$$

and the gamma density, proportional to

$$P(x)=x^A \exp(-x), \ x>0,$$

are for large A and B nearly $N(A/(A+B), AB/(A+B)^3)$, $N(A,A)$ respectively. These are instances of the following general observation:

(*) If the product P of a large number of smooth positive functions has a unique maximum at x_0, then P is often nearly $N(x_0,\sigma^2)$, where $\sigma^2=-P(x_0)/P''(x_0)$, i.e. the standardized P_s defined by

$$P_s(z)=P(x_0+\sigma z)/P(x_0)$$

is often nearly $\exp(-z^2/2)$ for moderate z, say $|z|\leqslant 5$.

The smooth positive functions are of course x and 1-x for the beta, and x, exp(-x) for the gamma.

We shall give below a crude, slightly messy general result that makes (*) precise and covers the beta and gamma cases. First, let us recast (*). With L=log P, $L_s=\log P_s$, it is easily checked that

$$L_s(0)=0, \ L_s'(0)=0,$$
$$L_s''(0)=\sigma^2L''(x_0)= \sigma^2P''(x_0)/P(x_0)=-1,$$
$$L_s'''(z)=\sigma^3L'''(x_0+\sigma z)=L'''(x_0+\sigma z)/|L''(x_0)|^{3/2},$$

so that

$$L_s(z)=-z^2/2+(z^3/6)\ L'''(x_0+\theta\sigma z)/|L''(x_0)|^{3/2}, \quad 0<\theta<1,$$ and (*) asserts that, often, $E=\max_{|z|\leqslant 5}|L'''(x_0+\sigma z)|/|L''(x_0)|^{3/2}$ is small.

It is easier to state the hypotheses of our general result, not in terms of the functions p whose product is P, but in terms of their logarithms $f=\log p$, whose sum is L. Here is the result.

THEOREM. Let (a,b) be any interval, let ε, M be positive numbers, and denote by $\mathcal{F}$ the class of all smooth functions f on (a,b) that satisfy $f'' \leq -\varepsilon$, $|f'''| \leq M$ on (a,b). If the sum L of n functions in $\mathcal{F}$ has a maximum at a point x_0 with $a+\delta \leq x_0 \leq b - \delta$, then $E \leq M/(\varepsilon^{3/2} n^{1/2})$ for $n \geq (5/\delta)^2/\varepsilon$.

PROOF. On (a,b) we have $L'' \leq -n\,\varepsilon$, $|L'''| \leq n\,M$, so that, if $x_0 \pm 5\sigma$ is within (a,b), we shall have $E \leq n\,M/(n\varepsilon)^{3/2} = M/(\varepsilon^{3/2} n^{1/2})$ as claimed. Since $\sigma^2 = -1/L''(x_0) \leq 1/n\varepsilon$, we shall have $x_0 \pm 5\sigma$ within (a,b) if $5/(n\varepsilon)^{1/2} \leq \delta$, i.e. $n \geq (5/\delta)^2/\varepsilon$.

To apply our Theorem to the beta case, note that $L=A\log x + B\log(1-x)$ is the sum of $n=A+B$ functions, each of which is either $f_1(x)=\log x$ or $f_2(x)=\log(1-x)$. For any a, b with $0 < a < b < 1$, both f_1 and f_2 will be in $\mathcal{F}$ for suitable ε, M.

Our Theorem has the following

COROLLARY. If p is a smooth positive function on (a_0,b_0) that has a unique maximum at an interior point x_0 of (a_0,b_0) with $p''(x_0) < 0$, then $P=p^n$ is nearly normal for large n.

PROOF. With $f=\log p$, we shall have $f'' \leq -\varepsilon = f''(x_0)/2$ in some interval (a,b) around x_0. With $M=\max|f'''|$ on (a,b), we have $f \in \mathcal{F}$ and the Theorem applies.

Note that in our Theorem $\sigma^2 \to 0$ as $n \to \infty$, so that it does not apply directly to the gamma case, where $\sigma^2=A$ is large in the case of near normality. But for any P, if we change the x-scale by choosing $c > 0$ and defining $Q(x)=P(cx)$, it is easily checked that $Q_s \equiv P_s$, so that, if Q satisfies the hypotheses of our Theorem, then P satisfies the conclusion. For the family defined by

$$P(x)=x^A \exp(-x^k),\ k > 0,\ x > 0,$$

where gamma is the special case k=1, we have

$$Q(x)=c^A x^A \exp(-c^k x^k),$$

so that, with c defined by $c^k=A$, we have

$$Q(x)=c^A [x \exp(-x^k)]^A,$$

and approximate normality for large A follows from our Corollary, with $P(x)= x \exp(-x^k),\ x > 0$.

If we think of x as an unknown parameter, and P as the likelihood function for x after many experiments, our observation says that, often, P is nearly normal with mean x_0, the maximum likelihood estimate, and variance $\sigma^2=-P(x_0)/P''(x_0)$. This σ^2 is clearly analogous to the reciprocal of Fisher information in classical statistics. Since (Savage's principle of stable estimation; see Savage (1962)) the likelihood function is the first approximation to the posterior density of the parameter, our observation says that, often, the posterior distribution of a parameter after many observations is nearly normal, so that x_0 and σ^2 are nearly sufficient. For the case of iid observations, these ideas have been deeply explored by Le Cam, DeGroot and others: see Le Cam (1966), Le Cam's comments in Berkson (1980) and the references cited there, and DeGroot (1970), pp. 212-218.

Are Cauchy likelihoods nearly normal, i.e. is $P(x)=\prod_{i=1}^{n} c(x-x_i)$, where $c(x)=1/(1+x^2)$ and $x_1 \ldots x_n$ are real numbers, nearly normal for large η? It depends on the x_i. If the x_i are uniformly distributed over a fixed interval say (-H,H), then, with $x_0=0$, we have $L''(x_0)=\Sigma f''(-x_i) \sim \int_{-H}^{H} f''(-t)dt \cdot n/2H$ for large n, where $f(x)= \log c(x)$. But $\int_{-H}^{H} f''(-t)dt$ is negative for all H, so that

$L''(x_0) \leqslant -\varepsilon n$ for some $\varepsilon > 0$ and all large n. Since f''' is uniformly bounded, we have $|L'''| \leqslant Mn$ for some M, and the proof of our Theorem shows that $E \to 0$ as $n \to \infty$ (all the proof uses is $L'' \leqslant -\varepsilon n$, $|L'''| \leqslant Mn$). But if the x_i are the integers $-k,\ldots,k$, so that $n=2k+1$, we shall have

$$L''(0)=\sum_{-k}^{k} f''(-x_i) \to \sum_{-\infty}^{\infty} f''(-x_i)=-.148\ldots$$

as $n \to \infty$, so that $\sigma \to 1/(.148)^{1/2} = 2.599\ldots$. Also

$$L'''(x)=\sum_{-k}^{k} f'''(x-x_i) \to \sum_{-\infty}^{\infty} f''(x-x_i).$$

In particular $L'''(.8) \to -.885$ as $n \to \infty$ (there is nothing special about .8), so that, since .8 is well within $0\pm 5\sigma$, E will be at least nearly $.885/(.148)^{3/2}=$ 15.5..for large n.

REFERENCES

1. Berkson, J. (1980). Minimum chi-square, not maximum likelihood! *Annals of Statistics* 8, 457-487.

2. De Groot, M.H. (1970). *Optimal Statistical Decisions*. McGraw-Hill Book Company.

3. Le Cam, L. (1966). Likelihood functions for large numbers of independent observations. *Festschrift for J. Neyman*, F.N. David, editor. John Wiley and Sons.

4. Savage, L.J. (1962). Bayesian statistics. *The writings of Leonard Jimmie Savage*. American Statistical Association and Institute of Mathematical Statistics, 1971.

University of California, Berkeley

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.

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. To appear in 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, 1980 Wald Lectures, (W. Chernoff, ed.) 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, Academic Press, New York, 1983.

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. To appear in 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. To appear in J. of Forecasting.

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. To appear in the Proc. of the Neyman-Kiefer Conference, (L. Le Cam, ed.) Wadsworth, 1984.

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). Proportional hazards, transformation models, partial likelihood, the order bootstrap, and adaptive inference, I.

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

29 BICKEL, P. J. and FREEDMAN, D. A. (Jan. 1984). Asymptotic Normality and the bootstrap in stratified sampling. To appear in Ann. Statist.

30 FREEDMAN, D. A. (Jan. 1984). The mean vs. the median: a case study in 4-R Act litigation. To appear in JBES.

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. To appear in Proc. of the Neyman-Kiefer Conference, (L. Le Cam, ed.) Wadsworth, 1985.

35 FREEDMAN, D. A. and NAVIDI, W. C. (Sept. 1984, revised Jan. 1985). Regression models for adjusting the 1980 Census.

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.

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.

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

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). 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.

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

51 O'SULLIVAN, F. and COX, D. D. (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). 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). As others see us: A case study in path analysis. Prepared for the Journal of Educational Statistics.

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
Technical Reports
University of California
Berkeley, California 94720

Cost: $1 per copy.