Lucien Le Cam: An Appreciation

Rudolf Beran
University of California, Berkeley, Department of Statistics
Berkeley, CA 94720–3860, USA
beran@stat.berkeley.edu

1. Biographical Sketch

Lucien Le Cam, one of the great mathematical statisticians of the twentieth century, died on April 25, 2000 after many years as Professor of Statistics and Mathematics at the University of California, Berkeley. His scholarly career was a life-long quest to create mathematical structures capable of expressing statistical theory.

Born on November 18, 1924 in Croze, Creuse, France, Le Cam was a son of farmers. His father died young. Higher education after graduation from a Catholic boarding school in 1942 would require scholarships. His first choice, attending a Catholic seminary, ended after one day. His next step, which provided partial financial support, was to take mathematics courses at a lycée in Clermont-Ferrand. His interests turned from chemistry to mathematics. In 1944, he registered as a student at the University of Paris and received the degree Licence es Science in 1945, after passing examinations in calculus, rational mathematics, and statistics.

For the next five years, Le Cam worked at Electricité de France on efficient operation of dams and on estimating probabilities of drought or flood. His participation in a seminar on statistics mentored by Darmois at the University of Paris resulted in an encounter with Jerzy Neyman around Easter 1950 and an invitation to visit Berkeley for one year as an instructor. Under Neyman's supervision, he stayed on as a graduate student, receiving his Ph.D. in 1952. He was promoted to Assistant Professor of Mathematics at Berkeley in 1953, joined the new Department of Statistics at its founding in 1955, became full Professor in 1960, and served as Department Chairman from 1961 to 1965.

With the exception of two years (1972–73) as Director of the Centre de Recherches Mathématiques at the University of Montreal, Le Cam spent his entire career at Berkeley. The 38 Ph.D. students he supervised at Berkeley include prominent names in our field. His office door at Berkeley was usually open. He was generous with ideas. Students and colleagues were welcome to engage him on any matter, whether scholarly or practical. His mathematical scholarship was only one projection of a richly talented personality.

Further biographical details may be found in Le Cam's published conversation with Grace Yang (1999), in an obituary by Beran and Yang (2000), and through the web-site `<www.stat.berkeley.edu/lecam/>`.

2. Le Cam's Early Asymptotics

I will discuss only Le Cam's earliest work on parametric models and how it helped rebuild the asymptotic theory of statistics while setting the stage for his later investigations into convergence of experiments. From his Ph.D. dissertation onwards, Le Cam sought to understand the asymptotic properties of maximum likelihood estimators and other procedures through statistically interpretable mathematical structures. This was a reaction to the obscure technical conditions that arose in pioneering mathematical studies of statistical procedures by H. Cramér and A. Wald. In conversation, Le Cam often cited Wald's (1943) paper as a key influence.

Consider a sample of n independent, identically distributed observations whose joint distribution is $P_{\theta,n}$. Here θ is an unknown element of parameter space Θ, which is an open subset of R^k. The problem is to estimate $\tau(\theta) \in R^m$, where τ is a differentiable function with $m \times k$ derivative matrix $\nabla\tau(\theta)$ of full rank. Let $T_{ML,n}$ denote the maximum likelihood estimator and let $\Sigma_\tau(\theta) = \nabla\tau(\theta)I^{-1}(\theta)\nabla'\tau(\theta)$, where $I(\theta)$ is the information matrix. In the first half of the twentieth century, it was widely believed that, in regular parametric models:

C1. The limiting distribution of $n^{1/2}(T_{n,ML} - \tau(\theta))$ under $P_{\theta,n}$ is $N(0, \Sigma_\tau(\theta))$ and

$$\lim_{n\to\infty} n\mathrm{E}_\theta|T_{ML,n} - \tau(\theta)|^2 = \mathrm{tr}\Sigma_\tau(\theta) \qquad \forall\theta \in \Theta;$$

C2. For any estimator T_n of $\tau(\theta)$,

$$\liminf_{n\to\infty} n\mathrm{E}_\theta|T_n - \tau(\theta)|^2 \geq \mathrm{tr}\Sigma_\tau(\theta) \qquad \forall\theta \in \Theta.$$

If true, these conjectures would support the conclusion that maximum likelihood estimators are asymptotically efficient in the sense that their risk attains the asymptotic lower bound in C2 at every θ. The long, complex history of these ideas includes contributions by Bernoulli, Laplace, Gauss and Edgeworth. Particularly influential was R. A. Fisher's (1925) paper.

Counterexamples in the 1950's revealed that conjectures C1 and C2 are too simple. Maximum likelihood estimators can violate C1 even in natural models such as the three-parameter lognormal or normal mixtures. The Hodges and James-Stein estimators both have smaller asymptotic risk at a superefficiency point than is permitted by C2. The former estimator has high risk near its superefficiency point (cf. Le Cam (1953)) while the latter dominates the MLE uniformly, especially near its superefficiency point. Subsequent investigations by Le Cam and a few others built a mathematical structure that unravels these paradoxes.

For any $\theta_0 \in \Theta$ and every $h \in R^k$, let $L_n(h, \theta_0)$ denote the log-likelihood ratio of the absolutely continuous part of $P_{\theta_0+n^{-1/2}h,n}$ with respect to $P_{\theta_0,n}$. The model is *locally asymptotically normal* (LAN) at θ_0 if there exists a random column vector $Y_n(\theta_0)$ and a nonsingular symmetric matrix $I(\theta_0)$ such that:

(a) Under $P_{\theta_0,n}$, $L_n(h_n, \theta_0) - h'Y_n(\theta_0) - 2^{-1}h'I(\theta_0)h = o_p(1)$ for every $h \in R^k$ and every sequence $h_n \to h$;

(b) $\mathcal{L}[Y_n(\theta_0)|P_{\theta_0,n}] \Rightarrow N(0, I(\theta_0))$

(cf. Le Cam (1960)). For an LAN model, the log-likelihood ratio behaves asymptotically like the log-likelihood ratio of $N(h, I^{-1}(\theta_0))$ with respect to $N(0, I^{-1}(\theta_0))$. The LAN property is possessed by regular parametric models such as smoothly parametrized exponential families.

For any sequence of estimators $\{T_n\}$ of $\tau(\theta)$, let $H_n(\theta)$ denote $\mathcal{L}[n^{1/2}(T_n - \tau(\theta))|P_{\theta,n}]$. The estimators $\{T_n\}$ are *locally asymptotically equivariant* (LAE) at θ_0 if, for every $h \in R^k$ and every sequence $h_n \to h$, $H_n(\theta_0 + n^{-1/2}h_n) \Rightarrow H(\theta_0)$.

Estimators $\{T_{n,E}\}$ that are classically efficient for $\tau(\theta_0)$ in a model that is LAN at θ_0 satisfy

$$T_{n,E} = \tau(\theta_0) + n^{-1/2}\nabla\tau(\theta_0)I^{-1}(\theta_0)Y_n(\theta_0) + o_p(1)$$

under $P_{\theta_0,n}$ and are LAE at θ_0 (cf. Le Cam (1969)). The precise role of these estimators, which include (possibly emended) MLE's, is revealed by Hájek's Convolution Theorem. This result draws on the LAN and LAE concepts and on earlier work by Wolfowitz and others.

Convolution Theorem. Let $K_n(\theta) = \mathcal{L}[(n^{1/2}(T_n - T_{n,E}), Y_n(\theta))|P_{\theta,n}]$. Suppose that the model is LAN and that $H_n(\theta_0) \Rightarrow H(\theta_0)$ as n increases. The following two statements are equivalent:

- $\{T_n\}$ is LAE at θ_0 with limit distribution $H(\theta_0)$.
- For every $h \in R^k$ and every sequence $h_n \to h$, $K_n(\theta_0 + n^{-1/2}h_n) \Rightarrow D(\theta_0) \times N(0, I(\theta_0))$ for some distribution $D(\theta_0)$ such that $H(\theta_0) = D(\theta_0) * N(0, \Sigma_\tau(\theta_0))$.

The local asymptotics just described have global implications. Suppose that $H_n(\theta) \Rightarrow H(\theta)$ for every $\theta \in \Theta$. Then there exists a Lebesgue nullset E in Θ such that the estimators $\{T_n\}$ are LAE at every $\theta \in \Theta - E$ (cf. Le Cam and Yang (1990), p. 76). It follows that the limit distribution $H(\theta)$ must have convolution structure for almost every $\theta \in \Theta$. Moreover, if $H(\theta)$ and $\Sigma_\tau(\theta)$ are both continuous in θ, the former in the topology of weak convergence, then $H(\theta)$ has convolution structure for *every* θ. This is often the case for classical parametric estimators of the maximum likelihood or minimum distance types.

The foregoing discussion yields corrected forms of conjectures C2 and C1. Let w be any symmetric, subconvex, continuous, non-negative function. If $\{T_n\}$ is LAE at θ_0, then the second assertion in the Convolution Theorem implies the risk bound

$$\liminf_{n\to\infty} \mathrm{E}_{\theta_0} w[n^{1/2}(T_n - \tau(\theta_0))] \geq \mathrm{E}w[\Sigma_\tau^{1/2}(\theta_0)Z],$$

where Z has the standard normal distribution on R^k. Whenever w is also bounded, this risk bound is attained asymptotically by the estimators $\{T_{n,E}\}$. No estimator sequence can achieve smaller asymptotic risk than $\{T_{n,E}\}$ *except* at its non-LAE points in Θ, which form a Lebesgue null set.

3. Concluding Remarks

That superefficiency points form a Lebesgue null set in the parameter space does not make them unimportant. Sophisticated modern estimators, among them certain scatterplot smoothers and model selection procedures, necessarily possess superefficiency points because they are constructed to dominate asymptotically MLE's that overfit the data.

The foregoing arguments incidentally yield necessary and sufficient conditions for correct convergence of parametric bootstrap distributions (cf. Beran (1996)). This asymptotic analysis explains why naive bootstrapping fails at superefficiency points, provides diagnostic plots for bootstrap failure based on the Convolution Theorem, and suggests possible remedies.

Le Cam (1972) generalized the Convolution Theorem and the Local Asymptotic Minimax Theorem to models whose limit experiments are not necessarily normal. These results, now known as the Hájek-Le Cam theorems, stand on the border between the mathematical structures of Le Cam's early work and those of his later studies on convergence of experiments.

Le Cam's life-long quest for pertinent mathematical structures is a guiding example for the reconstruction of statistical theory in the computer age.

REFERENCES

Beran, R. (1996). Diagnosing bootstrap success. *Ann. Inst. Statist. Math.*, **49**, 1–24.

Beran, R. and Yang, G. (2000). Lucien Le Cam. *IMS Bull.*, **29**, 464–466.

Fisher, R.A. (1925). Theory of statistical estimation. *Proc. Cambridge Math. Soc.*, **22**, 700–725.

Le Cam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. Calif. Publ. Statist.*, **1**, 277–330.

Le Cam, L. (1960). Locally asymptotically normal families of distributions. *Univ. Calif. Publ. Statist.*, **3**, 27–98.

Le Cam, L. (1969). *Théorie Asymptotique de la Décision Statistique.* University of Montreal Press.

Le Cam, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, **I**, L. Le Cam, J. Neyman, and E. L. Scott, eds., University of California Press, 245–261.

Le Cam, L. and Yang, G.L. (1990). *Asymptotics in Statistics.* Springer Verlag.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.*, **54**, 426–482.

Yang, G.L. (1999). A Conversation with Lucien Le Cam. *Statist. Science*, **14**, 223–241.

RESUMÉ

Lucien Le Cam a creé des structures mathématiques pour exprimer la théorie statistique developée sous des modèles probabilistes.