In Memory of
Lucien Le Cam (1924 - 2000):

Seven theorems every statistician should know

Jon A. Wellner

Talk given at University of Washington, Seattle,
May 14, 2001.

Email: jaw@stat.washington.edu
http: //www.stat.washington.edu/jaw/jaw.research.html

OUTLINE:

0. Introduction: Lucien Le Cam

1. Poisson approximations

2. Contiguity, Le Cam's three lemmas

3. Hellinger distance and Independence

4. Quadratic mean differentiability (QMD) and asymptotic normality of the MLE

5. Local asymptotic minimax and convolution theorems

6. Preservation of LAN under information loss

7. Dimensionality and Estimation: Rates of convergence

8. Convergence of Experiments

References

Appeal

Lucien Le Cam: a few facts

- 77 published papers, 4-6 books.
- 38 PhD students (Stigler, Ferguson, Yang , ...)
- Chair of Statistics Dept., Berkeley, 1961-1965.
- President of the IMS, 1973
- Editor, Proceedings of the 5th and 6th Berkeley Symposia

Further Reading:

- Lucien Le Cam. In *More Mathematical People*, 161-180, D. J. Albers, G. L. Alexanderson, and C. Reid, editors. Harcourt Brace Jovanovich, Boston.
- Lehmann, E. L. (1997). Le Cam at Berkeley. In *Festschrift for Lucien Le Cam*, 297 - 304. D. Pollard, E. Torgersen, G. Yang, editors. Springer-Verlag, New York.
- Le Cam, L. (1998). Recollections on my contacts with Jaroslav Hájek. In *Collected Works of Jaroslav Hájek - With Commentary*, 21 - 28. Compiled by M. Huskova, R. Beran, and V. Dupac. Wiley, New York.

1. Poisson approximation

$X_i \sim \text{Bernoulli}(p_i)$, $i = 1, 2, \ldots$.
$S_n = X_1 + \cdots + X_n$, $n = 1, 2, \ldots$.
$\lambda = \sum_1^n p_i$.
Let $Y \sim \text{Poisson}(\lambda)$: thus

$$P(Y = y) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \ldots$$

Theorem 1. (Le Cam, 1960).

$$\sum_{k=0}^{\infty} |P(S_n = k) - P(Y = k)| \le \sum_{i=1}^{n} p_i^2 .$$

Theorem 2. (Le Cam, 1960). If $\max_{1 \le i \le n} p_i \le 1/4$, then

$$\sum_{k=0}^{\infty} |P(S_n = k) - P(Y = k)| \le \frac{8}{\lambda} \sum_{i=1}^{n} p_i^2 .$$

References:

- Le Cam (1960a),
- Le Cam (1960b),
- Serfling (1975),
- Barbour, Holst and Janson (1992).

2. Contiguity and Le Cam's three Lemmas

$(X_1, \ldots, X_n) \sim P_n \quad (= P^n_{\theta_0})$
or
$(X_1, \ldots, X_n) \sim Q_n \quad (= P^n_{\theta_n})$

$Q_n \lhd P_n$ (in words, $\{Q_n\}$ is *contiguous* w.r.t. $\{P_n\}$ if, $P_n(A_n) \to 0 \quad \text{implies} \quad Q_n(A_n) \to 0$.

Le Cam's first lemma. (Le Cam, 1960).

If $L_n \equiv dQ_n/dP_n$ satisfies

$$L_n \to_d L \;\; \text{under} \;\; P_n$$

with $E(L) = 1$, then $\{Q_n\} \lhd \{P_n\}$.

Suppose:
$P_n = P_{n1} \times \cdots \times P_{nn}$ with densities $p_{n1}, \ldots, p_{nn}$;
$Q_n = Q_{n1} \times \cdots \times Q_{nn}$ with densities $q_{n1}, \ldots, q_{nn}$.

Set

$$W_n = 2 \sum_{i=1}^{n} \left\{ \frac{q_{ni}^{1/2}}{p_{ni}^{1/2}}(X_{ni}) - 1 \right\} .$$

Le Cam's second lemma. (Le Cam, 1960).
If $W_n \rightarrow_d N(-\sigma^2/4, \sigma^2)$ and

$$P_n \left(\max_{1 \le i \le n} |\frac{q_{ni}}{p_{ni}}(X_{ni}) - 1| > \epsilon \right) \rightarrow 0$$

for every $\epsilon > 0$, then

$$\log \frac{dQ_n}{dP_n}(\underline{X}) \rightarrow_d N(-\sigma^2/2, \sigma^2) \quad \text{under} \quad P_n .$$

Suppose:
T_n is a (real-valued) statistic for each n.

Le Cam's third lemma. (Le Cam, 1960).
Suppose that

$$\begin{pmatrix} T_n \\ \log \frac{dQ_n}{dP_n} \end{pmatrix} \rightarrow_d N_2 \left(\begin{pmatrix} 0 \\ -\sigma^2/2 \end{pmatrix}, \begin{pmatrix} \tau^2 & c \\ c & \sigma^2 \end{pmatrix} \right)$$

under $\{P_n\}$. Then

$$\begin{pmatrix} T_n \\ \log \frac{dQ_n}{dP_n} \end{pmatrix} \rightarrow_d N_2 \left(\begin{pmatrix} c \\ +\sigma^2/2 \end{pmatrix}, \begin{pmatrix} \tau^2 & c \\ c & \sigma^2 \end{pmatrix} \right)$$

under $\{Q_n\}$.

Corollary: asymptotic power of test based on T_n under the alternative(s) Q_n.

References:

- Le Cam (1960c),
- Hajek and Sidak (1967),
- Hall and Loynes (1977),
- Pollard (1997).

3. Hellinger distance of product laws - Independence

$$
\begin{aligned}
H^2(P,Q) &= \frac{1}{2}\int (\sqrt{p}-\sqrt{q})^2 d\mu \\
&= 1 - \int \sqrt{pq}\, d\mu \\
&= \text{squared Hellinger distance from } P \text{ to } Q \\
&= 1 - \rho(P,Q)
\end{aligned}
$$

where $p = dP/d\mu$, $q = dQ/d\mu$, $\mu = P+Q$ always works. $\rho(P,Q) =$ the affinity between P and Q.

$$(X_{n,1},\ldots,X_{n,n}) \sim P_n, \quad X_{n,i} \sim P_{ni} \ \text{ indep.}$$

or

$$(X_{n,1},\ldots,X_{n,n}) \sim Q_n \quad X_{n,i} \sim Q_{ni} \ \text{ indep.}$$

Theorem. (Le Cam, 1970)

$$\begin{aligned} H^2(P_n, Q_n) &= 1 - \rho(P_n, Q_n) = 1 - \prod_{i=1}^{n} \rho(P_{ni}, Q_{ni}) \\ &= 1 - \prod_{i=1}^{n} \left\{ 1 - H^2(P_{ni}, Q_{ni}) \right\} \end{aligned}$$

Corollary. If $P_n = P_{\theta_0}^n$, $Q_n = P_{\theta_0 + c/\sqrt{n}}^n$, and

$$nH^2(P_{\theta_0 + c/\sqrt{n}}, P_{\theta_0}) \to c^T I(\theta_0) c \,,$$

then

$$\begin{aligned} H^2(P_n, Q_n) &= 1 - (1 - H^2(P_{\theta_0}, P_{\theta_0 + c/\sqrt{n}}))^n \\ &\to 1 - \exp(-c^T I(\theta_0) c) \,. \end{aligned}$$

References:

- Kakutani (1948),
- Le Cam (1969),
- Le Cam (1970),
- Le Cam and Yang (1990).

4. Quadratic mean differentiability (QDM) and asymptotic normality of the MLE

The model $\{P_\theta : \ \theta \in \Theta \subset R^k\}$ is **differentiable in quadratic mean** at θ_0 if there exist a vector-valued function $\dot{l}_{\theta_0}$ such that the densities $p_\theta = dP_\theta/d\mu$ satisfy

$$\int \left\{ \sqrt{p_\theta} - \sqrt{p_{\theta_0}} - \frac{1}{2}(\theta - \theta_0)^T \dot{l}_{\theta_0} \sqrt{p_{\theta_0}} \right\}^2 d\mu = o(|\theta - \theta_0|^2) \, .$$

Note that

$$\frac{\partial}{\partial\theta}\sqrt{p_\theta} = \frac{1}{2\sqrt{p_\theta}} \frac{\partial}{\partial\theta} p_\theta = \frac{1}{2}\left(\frac{\partial}{\partial\theta} \log p_\theta \right) \sqrt{p_\theta} \, ,$$

so the function $\dot{l}_{\theta_0}$ is really the score function of the model, and

$$I(\theta_0) = P_{\theta_0}(\dot{l}_{\theta_0} \dot{l}_{\theta_0}^T)$$

is the Fisher information matrix. However QDM does *not* require existence of $(\partial/\partial\theta) \log p_\theta(x)$ for every x.

Theorem.

Le Cam (1970) + van der Vaart (1998).

Suppose that $\{P_\theta : \theta \in \Theta \subset R^k\}$ is differentiable in quadratic mean at an interior point θ_0 of Θ. Furthermore, suppose that there exists a measurable function $\dot{l}$ with $P_{\theta_0}(\dot{l}^2) < \infty$ such that, for every θ_1 and θ_2 in a neighborhood of θ_0,

$$|\log p_{\theta_1}(x) - \log p_{\theta_2}(x)| \leq \dot{l}(x)|\theta_1 - \theta_2|\,.$$

If $I(\theta_0)$ is non-singular and the MLE $\widehat{\theta}_n$ is consistent, then

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = I(\theta_0)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\dot{l}_{\theta_0}(X_i) + o_{P_{\theta_0}}(1)$$

and hence

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d N_k\left(0, I^{-1}(\theta_0)\right)\,.$$

Theorem. (Le Cam, 1970).
Suppose that $\{P_\theta : \theta \in \Theta \subset R^1\}$ is differentiable in quadratic mean at an interior point θ_0 of Θ. Furthermore, suppose that for θ in a neighborhood of θ_0

$$\limsup_{\delta \to 0} \frac{1}{|\delta|} H(P_\theta, P_{\theta+\delta}) \equiv \sigma(\theta) < \infty \, ,$$

$\sigma(\theta_0) > 0$, and

$$\lim_{\delta \searrow 0} \delta^{-1} \int_{\theta_0-\delta}^{\theta_0+\delta} |\sigma(\theta) - \sigma(\theta_0)| d\theta = 0 \, .$$

Then with probability tending to 1 the MLE $\widehat{\theta}_n$ of θ exists and satisfies

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d N_k(0, V(\theta_0))$$

where

$$V(\theta_0) = (I(\theta_0)/4)/(\sigma^2(\theta_0) + I(\theta_0)/4)^2 \leq I^{-1}(\theta_0) \, .$$

References:

- Le Cam (1970);
- van der Vaart and Wellner (1996), page 306;
- van der Vaart (1998), page 65.

5. Local asymptotic minimax and convolution theorems

Local asymptotic normality or LAN:
$\underline{X} \sim P_{n,\theta}$, $\theta \in \Theta \subset R^k$.
The family $\{P_{n,\theta}\}$ is LAN at θ_0 if there exist matrices r_n, $I(\theta_0)$, and random vectors $S_n(\theta_0)$ with $S_n(\theta_0) \rightarrow_d N(0, I(\theta_0))$ such that for every $h_n \rightarrow h$

$$\log \frac{dP_{n,\theta_0+h_n/r_n}}{dP_{n,\theta_0}} = h^T S_n(\theta_0) - \frac{1}{2} h^T I(\theta_0) h + o_p(1)$$

Theorem. (Minimax theorem;
Hájek (1970), (1972);
Le Cam (1972), (1979)).
Suppose that $q : \Theta \mapsto R$ is differentiable with derivative $\dot{q}(\theta_0)$ at θ_0. For any estimators $\{T_n\}$ of $q(\theta)$ and any bowl-shaped loss function l

$$\liminf_{n\to\infty} \sup\{E_\theta \left(l(r_n(T_n - q(\theta))\right) : r_n|\theta - \theta_0| \leq M\} \geq El(Z_0)\,,$$

where

$$Z_0 \sim N\left(0, \dot{q}^T(\theta_0) I(\theta_0)^{-1} \dot{q}(\theta_0)\right) \,.$$

Locally regular estimators:
$\{T_n\}$ is locally regular at θ_0 if: for every h, with $\theta_n = \theta_0 + h/r_n$,

$$r_n\left(T_n - q(\theta_n)\right) \to_d E_0$$

under P_{n,θ_n} where the distribution of E_0 does not depend on h.

Theorem. (Convolution theorem;
Hájek (1970), (1972);
Le Cam (1972), (1979)).
If $\{T_n\}$ is a sequence of regular estimators of $q(\theta)$ with limiting distribution given by $\mathcal{L}(E_0)$, and $\{P_{n,\theta}\}$ satisfies LAN at θ_0, then

$$E_0 =_d Z_0 + W_0$$

where W_0 is independent of

$$Z_0 \sim N\left(0, \dot{q}^T(\theta_0) I(\theta_0)^{-1} \dot{q}(\theta_0)\right) .$$

References:

- Le Cam (1972),
- Le Cam (1979),
- Hajék (1970), Inagaki (1970)
- Hajék (1972) .

6. Preservation of LAN

Suppose that $\mathcal{Q} = \{Q_\theta : \theta \in \Theta\}$ is differentiable in quadratic mean at θ_0 with derivative $\dot{l}_0(\cdot, \mathcal{Q})$ and Fisher information matrix

$$I(\theta_0, \mathcal{Q}) = Q_{\theta_0}(\dot{l}_0(\cdot, \mathcal{Q})\dot{l}_0(\cdot, \mathcal{Q})^T) \, .$$

Suppose that $X \sim Q_\theta \in \mathcal{Q}$ and $Y \equiv T(X) \sim P_\theta$ for a (measurable) function T.
Let $\mathcal{P} \equiv \{P_\theta : \theta \in \Theta\}$.

Theorem. (Le Cam and Yang, 1988).
If $\mathcal{Q}$ is differentiable in QM at θ_0, then $\mathcal{P}$ is differentiable in quadratic mean at θ_0 with score function

$$\dot{l}(Y, \mathcal{P}) = E\{\dot{l}_{\theta_0}(X; \mathcal{Q})|Y\}$$

and

$$I(\theta_0; \mathcal{P}) \leq I(\theta_0; \mathcal{Q}) \, .$$

References:

- Le Cam and Yang (1988),
- Ibragimov and Has'minskii (1979), page 70,
- Bickel, Klaassen, Ritov and Wellner (1993), page 461.

7. Dimensionality and Estimation: Rates of convergence

$\theta : \mathcal{P} \to \Theta$, (Θ, d) a metric space
$\widehat{\theta}$ an estimator of $\theta(P)$.

Theorem. (Le Cam, 1973). Suppose there are subsets $\Theta_1, \Theta_2 \subset \Theta$ that are $2\delta-$ separated. Suppose that $\mathcal{P}_1$, $\mathcal{P}_2$ are subsets of $\mathcal{P}$ with $\theta(P) \in \Theta_i$ for $P \in \mathcal{P}_i$, $i = 1, 2$. Then

$$\sup_{P \in \mathcal{P}} E_P d(\widehat{\theta}, \theta(P)) \geq \delta \cdot \sup_{P_i \in co(\mathcal{P}_i)} \|P_1 \wedge P_2\|$$

where

$$\|P_1 \wedge P_2\| = \int p_1 \wedge p_2 \, d\mu$$

for any common dominating measure μ of P_1, P_2, and $co(\mathcal{P}_i)$ = the convex hull of $\mathcal{P}_i$. Here

$$\begin{aligned} d_{TV}(P, Q) &= \frac{1}{2} \int |p - q| d\mu \\ &= 1 - \|P \wedge Q\|, \end{aligned}$$

so $\|P \wedge Q\|$ is the “affinity” for total variation distance.

Let $D(\epsilon, \Theta)$ be the $\epsilon-$ packing number of Θ for the metric d; i.e. the maximal number of $\epsilon-$ separated points in Θ.

Corollary. Suppose that $X_1, \ldots, X_n$ are i.i.d. P_θ, $\theta \in \Theta$. Suppose that ϵ_n satisfies

$$D(\epsilon_n) = 4n\epsilon_n^2 + 2\log 2\,.$$

Then

$$\min_{\widehat{\theta}} \max_{\theta \in \Theta} E_\theta d^2(\widehat{\theta}, \theta) \geq (A^2/8)\epsilon_n^2\,.$$

References:

- Le Cam (1973),
- Yu (1997),
- Birgé (1983), (1986)
- Yang and Barron (1999).

8. Approximation of Experiments

.

Another talk!

.

References:

- Le Cam (1972),
- Le Cam (1986),
- van der Vaart (1998, Chapter 9, pp. 125 - 137,
- van der Vaart (200?).

References:

- Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press.

- Birgé, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrschein. verw. Gebiete* **65**, 181 - 237.

- Hodges, J. L. and Le Cam, L. (1960). The Poisson approximation to the Poisson binomial distribution. *Ann. Math. Statist.* **31**, 737- 740.

- Inagaki, N. On the limiting distribution of a sequence of estimators with uniformity properties. *Ann. Inst. Statist. Math.* **22**, 1 - 13.

- Le Cam, L. (1960). Poisson approximation for the Poisson binomial distribution. *Pacific Journal of Mathematics* **10**, 1181-1197.

- Le Cam, L. (1960). Locally asymptotically normal families of distributions. *Univ. of California Publ. in Statistics* **3**, 37 - 98.

- Le Cam, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Statist.* **41**, 802-828.

- Le Cam, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, Vol. 1, University of California Press, Berkeley, 245 - 261.
- Le Cam, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1**, 38 - 53.
- Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York.
- Le Cam, L. and Yang, G. L. (1988). On the preservation of local asymptotic normality under information loss. *Ann. Statist.* **16**, 483 - 520.
- Le Cam, L. (1990). Maximum likelihood: an introduction. *International Statistical Review* **58**, 153 - 172.
- Pollard, D. (1997). Another look at differentiability in quadratic mean, In *Festschrift for Lucien Le Cam*, D. Pollard, E. Torgersen, and G. L. Yang, editors, pp. 205 - 214. Springer-Verlag, Berlin.
- Serfling, R. J. (1975). A general Poisson approximation theorem. *Ann. Probability* **3**, 726 - 731.
- van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press.
- Yang, Y. and Barron, A. (1999). Information -theoretic determination of minimax rates of convergence. *Ann. Statist.* **27**, 1564 - 1599.

- Yu, Bin (1997). Assouad, Fano, and Le Cam, In *Festschrift for Lucien Le Cam*, D. Pollard, E. Torgersen, and G. L. Yang, editors, pp. 423 - 435. Springer-Verlag, Berlin.

Fund Raising for a Permanent IMS Le Cam Lecture. Visit

www.stat.berkeley.edu/lecam

for the IMS Council resolution and more information. Please send donations, payable to

IMS-Le Cam Endowment Fund

to:

IMS-Le Cam Endowment Fund
PO Box 22718
Beachwood OH 44122