Lucien Le Cam's Mathematical Framework:

Abstract Nonsense or An Inspired Simplification?

Talk by David Pollard at the Atlanta JSM 2001
6 August 2001

How many would-be readers of Le Cam's master work—his 1986 book on "Asymptotic Methods in Statistical Decision Theory"—have gotten stuck on the first chapter? Instead of families of probability measures and (bounded) random variables on a common sample space, Le Cam offered bands in abstract L-spaces and their dual M-spaces.

The translation into more traditional notation is not hard, but it does raise a key question. To paraphrase Le Cam, Why cling to a sample space that causes trouble, if there are other choices that ensure nice properties for the objects we care about? The answer might persuade the would-be readers to explore beyond the first chapter.

References

Le Cam, L. (1986), *Asymptotic Methods in Statistical Decision Theory*, Springer-Verlag, New York.

Le Cam, L. & Yang, G. L. (2000), *Asymptotics in Statistics: Some Basic Concepts*, 2nd edn, Springer-Verlag.

Torgersen, E. (1991), *Comparison of Statistical Experiments*, Cambridge University Press.

For more information:

http://www.stat.yale.edu/~pollard/ (then follow link to Paris lectures)

Dictionary **(imperfect translation)**

CLASSICAL	LE CAM GENERALIZATION
statistical model $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ on $(\mathcal{X}, \mathcal{A})$	L-space $L(\mathcal{P})$ generated by $\mathcal{P}$
bounded measurable (real) functions on $\mathcal{X}$	M-space $M(\mathcal{P})$ of continuous linear functionals on $L(\mathcal{P})$
test function	nonnegative element of unit ball of $M(\mathcal{P})$
randomization, Markov kernel	transition = generalized randomization

- Break with tradition?
- Why generalize?
- Interpretation?
- Advantages? Disadvantages?

Randomization

$\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ on $(\mathcal{X}, \mathcal{A})$

$\mathcal{Q} = \{\mathbb{Q}_\theta : \theta \in \Theta\}$ on $(\mathcal{Y}, \mathcal{B})$

Randomization from $\mathcal{X}$ to $\mathcal{Y}$:

? Markov kernel $\{K_x : x \in \mathcal{X}\}$, with each K_x a probability measure on $\mathcal{B}$

For each finite (signed) measure μ on $\mathcal{A}$ define $\nu := K\mu$ on $\mathcal{B}$ by

$$\nu(B) = \int K_x(B)\,\mu(dx)$$

Le Cam distance: want randomization to make

$$\sup_{\theta\in\Theta} \|K\mathbb{P}_\theta - \mathbb{Q}_\theta\|$$

small (where $\|\cdot\|$ denotes total variation norm).

Transitions

Le Cam (1986, page 4):

> Let L' and L'' be two L-spaces.
> A transition from L' to L'' is a positive linear map of L' into L'' such that
> $\|T\mu^+\| = \|\mu^+\|$ for every $\mu \in L'$.

For example: L' might be the set of all finite (signed) measures on $\mathcal{A}$ that are dominated by a fixed measure and L'' might be the set of all finite (signed) measures on $\mathcal{B}$. The transition might be defined by a Markov kernel.

Conditioning

Let T be a measurable map from $(\mathcal{X}, \mathcal{A})$ to $(\mathcal{Y}, \mathcal{B})$.
Let $\mathbb{P}$ be a probability measure on $\mathcal{A}$, and $\mathbb{Q}$ be the distribution of T under $\mathbb{P}$.
That is, $\mathbb{Q}$ is the image of $\mathbb{P}$ under T.

If $X \in L_1(\mathbb{P})$ then $Z(t) := \mathbb{E}(X \mid T = t)$ is the element of $L_1(\mathbb{Q})$ defined (up to almost sure equivalence) by

$$(*) \qquad \int X(x)h(Tx)\,\mathbb{P}(dx) = \int Z(t)h(t)\,\mathbb{Q}(dt) \qquad \text{for all bounded, } \mathcal{B}\text{-measurable } h.$$

Define μ_X as the measure with density $d\mu_X/d\mathbb{P} = X$ and ν_Z as the measure with density $d\nu_Z/d\mathbb{Q} = Z$. Then (*) becomes

$$\int h(Tx)\,\mu_X(dx) = \int h(t)\,\nu_Z(dt) \qquad \text{for all bounded, } \mathcal{B}\text{-measurable } h.$$

That is, ν_Z is the image of μ_X under T.

$$\begin{array}{ccc}
L_1(\mathbb{P}) & \xrightarrow{\mathbb{E}(\cdot \mid T = t)} & L_1(\mathbb{Q}) \\
\updownarrow & & \updownarrow \\
\begin{array}{c}\text{L-space of}\\ \text{finite measures}\\ \text{dominated by } \mathbb{P}\end{array} & \begin{array}{c}\longrightarrow\\ \text{image measure}\\ \text{under } T\end{array} & \begin{array}{c}\text{L-space of}\\ \text{finite measures}\\ \text{dominated by } \mathbb{Q}\end{array}
\end{array}$$

- The Kolmogorov conditional expectation represents a transition between two L-spaces.
- Work with $\mathbb{E}(\cdot \mid T = t)$ even if we really want the full conditional distribution?
- Martingales (don't need the full conditional distribution)
- Strong Markov property
- Bayes theory (posterior distributions?)
- Sufficiency (Rao-Blackwell, ...)

Two Le Cam-equivalent models

$\mathcal{P}$ model: probabilities on $[0,2)$,

$$\mathbb{P}_\theta = \begin{cases} \text{point mass at } \theta & \text{for } 0 \le \theta < 1 \\ \text{Lebesgue measure on } [1,2) & \text{for } \theta = 1 \end{cases}$$

$L(\mathcal{P})$ consists of all finite (signed) measures of the form $\mu = \mu_d + \mu_a$ where μ_d is a discrete measure on $[0,1)$ and μ_a is a measure on $[1,2)$ that is absolutely continuous with respect to Lebesgue measure.

$\mathcal{Q}$ model: probabilities on $[0,1)$,

$$\mathbb{Q}_\theta = \begin{cases} \text{point mass at } \theta & \text{for } 0 \le \theta < 1 \\ \text{Lebesgue measure on } [0,1) & \text{for } \theta = 1 \end{cases}$$

$L(\mathcal{Q})$ consists of all finite (signed) measures of the form $\mu = \mu_d + \mu_a$ where μ_d is a discrete measure on $[0,1)$ and μ_a is a measure on $[0,1)$ that is absolutely continuous with respect to Lebesgue measure.

Want to test hypotheses

$$\theta = 1 \qquad \text{versus} \qquad 0 \le \theta < 1$$

For $\mathcal{P}$ use test function:

$$\psi(x) = \begin{cases} 1 & \text{if } x < 1 \\ 0 & \text{if } x \ge 1 \end{cases}$$

Get $\mathbb{P}_1\psi = 0$ and $\mathbb{P}_\theta\psi = 1$ for $0 \le \theta < 1$: a perfect test.

Note that $\mu \mapsto \int \psi \, d\mu$ defines a continuous linear functional on $L(\mathcal{P})$.

For $\mathcal{Q}$, all traditional tests useless. Generalized test defined by the continuous linear functional $\mu \mapsto \|\mu_d\|$ on $L(\mathcal{Q})$.

Some ways to use Le Cam framework

- Add regularity assumptions so that all generalized objects reduce to their classial analogs.
- Use Le Cam framework as a convenient way of finding traditional solutions:

  (i) Find generalized solution.

  (ii) Show that solution from (i) can actually be identified with a traditional solution.

- Rethink what we mean by a statistical model. For example, what does it mean to say that data are observations on a fixed distribution? Why are sample spaces needed? . . .