The Le Cam methodology for directional data

Yvik Swan (Université de Liège)

with Christophe Ley (ULB), Baba Thiam (Lille 3) and Thomas Verdebout (Lille 3)

ADISTA, Brussels 2014

Consider two statistical experiments $\mathcal{E}^P$ and $\mathcal{E}^Q$.

Suppose that $\mathcal{E}^P$ and $\mathcal{E}^Q$ are “close”.

Le Cam’s method : with the “correct notion of closeness” there is an automatic transfer of solutions to certain types of decision theoretic problems from $\mathcal{E}^Q$ to $\mathcal{E}^P$.

In other words, optimal procedures for $\mathcal{E}^Q$ become (nearly) optimal procedures for $\mathcal{E}^P$ and vice-versa.

In this talk the “correct notion of closeness” is Local Asymptotic Normality (LAN).

Consider $\mathbf{X}^{(n)} = (X_1^{(n)}, \ldots, X_n^{(n)})$ observations described by

$$\mathcal{P}^{(n)} = \left\{ \mathrm{P}_{\boldsymbol{\theta}}^{(n)} : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^K \right\}.$$

The sequence of models is LAN if, for all $\boldsymbol{\theta} \in \Theta$, there exist

- a central sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta})(= \boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, \mathbf{X}^{(n)}))$
- a Fisher information matrix $\boldsymbol{\Gamma}(\boldsymbol{\theta})$

such that, for $\boldsymbol{\tau}_n \in \mathbb{R}^K$,

$$\log \left[\frac{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{n}^{-1/2}\boldsymbol{\tau}_n}}{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}}} \right] = \boldsymbol{\tau}_n' \boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}) - \frac{1}{2} \boldsymbol{\tau}_n' \boldsymbol{\Gamma}(\boldsymbol{\theta}) \boldsymbol{\tau}_n + o_P(1)$$

and $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}(\boldsymbol{\theta}))$, both under $P_{\boldsymbol{\theta}}^{(n)}$ as $n \to \infty$.

The log-likelihood reminds us of

$$\log\left(\frac{f_{\boldsymbol{\tau}}^{\boldsymbol{\Delta}}(\boldsymbol{\Delta})}{f_{\mathbf{0}}^{\boldsymbol{\Delta}}(\boldsymbol{\Delta})}\right) = \boldsymbol{\tau}'\boldsymbol{\Delta} - \frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\Gamma}\boldsymbol{\tau}$$

the (exact) log-likelihood for a Gaussian shift model $\mathcal{N}(\boldsymbol{\Gamma}\boldsymbol{\tau}, \boldsymbol{\Gamma})$.

LAN means that the likelihoods of $\mathcal{P}^{(n)}$ resemble the likelihoods of a Gaussian shift model.

More importantly LAN means that all risk functions can be approximated by the risk functions for the Gaussian model !

Consequence : We know how, given $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta})$ and $\boldsymbol{\Gamma}(\boldsymbol{\theta})$, to construct locally and asymptotically optimal (parametric) procedures for $\boldsymbol{\theta}$.

Consider observations $\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathcal{S}^{k-1}$ from model

$$\mathcal{P}^{(n)} = \left\{ \mathrm{P}^{(n)}_{\boldsymbol{\theta};f_1} \,|\, \boldsymbol{\theta} \in \mathcal{S}^{k-1} \right\}$$

where $\mathrm{P}_{\boldsymbol{\theta};f_1}$ has rotationally symmetric density

$$\mathbf{x} \mapsto f_{\boldsymbol{\theta}}(\mathbf{x}) = c_{k,f_1}\, f_1(\mathbf{x}'\boldsymbol{\theta}), \quad \mathbf{x} \in \mathcal{S}^{k-1}, \tag{1}$$

with respect to the usual surface area measure on spheres and where $f_1 : [-1, 1] \to \mathbb{R}_0^+$ is absolutely continuous and (strictly) monotone increasing.

Question : is it possible to obtain a LAN result for this family of distributions ?

In other words, do small perturbations of the parameter $\boldsymbol{\theta}$ result in the "quasi-Gaussian" representation for the log-likelihood ?

The answer to the question is yes (but it's not easy to prove).

1. Pass into spherical coordinates

$$\boldsymbol{\eta} \in \mathbb{R}^{k-1} \to \boldsymbol{\theta} = \hbar(\boldsymbol{\eta}) \in \mathcal{S}^{k-1}.$$

2. Prove that, under reasonable assumptions on f_1, the likelihood will have correct behaviour under perturbations in $\boldsymbol{\eta}$, i.e. the family

$$\left\{ \mathrm{P}^{(n);\hbar}_{\boldsymbol{\eta};f_1} \mid \boldsymbol{\eta} \in \mathbb{R}^{k-1} \right\}$$

is LAN.

3. Prove that one can read this approximation entirely "on the sphere" without needing the spherical coordinates.

Specifically, define

1

$$\mathbf{\Delta}^{(n)}_{\boldsymbol{\theta};f_1} := n^{-1/2} \sum_{i=1}^{n} \varphi_{f_1}(\mathbf{X}_i'\boldsymbol{\theta})(1 - (\mathbf{X}_i'\boldsymbol{\theta})^2)^{1/2} \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_i)$$

where

$$\varphi_{f_1} := \dot{f}_1 / f_1 \quad \text{and} \quad \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_i) := \frac{\mathbf{X} - (\mathbf{X}'\boldsymbol{\theta})\boldsymbol{\theta}}{\|\mathbf{X} - (\mathbf{X}'\boldsymbol{\theta})\boldsymbol{\theta}\|}.$$

2

$$\mathbf{\Gamma}_{\boldsymbol{\theta};f_1} := \frac{\mathcal{J}_k(f_1)}{k-1}(\mathbf{I}_k - \boldsymbol{\theta}\boldsymbol{\theta}')$$

with, for $\tilde{f}_1$ the density of $\mathbf{X}'\boldsymbol{\theta}$,

$$\mathcal{J}_k(f_1) := \int_{-1}^{1} \varphi^2_{f_1}(t)(1 - t^2)\tilde{f}_1(t)dt.$$

Theorem (ULAN)

Assume three technical assumptions.

Then,

1. *for any* $\boldsymbol{\theta}^{(n)} \in \mathcal{S}^{k-1}$ *such that* $\boldsymbol{\theta}^{(n)} - \boldsymbol{\theta} = O(n^{-1/2})$
2. *for any sequence* $\mathbf{t}_n$ *such that* $\boldsymbol{\theta}^{(n)} + n^{-1/2}\mathbf{t}_n \in \mathcal{S}^{k-1}$,

we have

$$\log\left(\frac{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}^{(n)}+n^{-1/2}\mathbf{t}_n;f_1}}{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}^{(n)};f_1}}\right) = \mathbf{t}_n'\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}^{(n)};f_1} - \frac{1}{2}\mathbf{t}_n'\boldsymbol{\Gamma}_{\boldsymbol{\theta};f_1}\mathbf{t}_n + o_{\mathrm{P}}(1)$$

and

$$\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}^{(n)};f_1} \xrightarrow{\mathcal{L}} \mathcal{N}_{k-1}(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\theta};f_1}),$$

both under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}^{(n)};f_1}$, *as* $n \to \infty$.

The ULAN result is very nice but still insufficient as procedures will be based on

$$\mathbf{\Delta}^{(n)}_{\boldsymbol{\theta}^{(n)};f_1} = n^{-1/2}\sum_{i=1}^{n}\varphi_{f_1}(\mathbf{X}_i'\boldsymbol{\theta}^{(n)})(1-(\mathbf{X}_i'\boldsymbol{\theta}^{(n)})^2)^{1/2}\mathbf{S}_{\boldsymbol{\theta}^{(n)}}(\mathbf{X}_i)$$

and thus only be valid for known angular function f_1.

To ensure validity over all f_1, the next step in the Le Cam method is to

replace $\mathbf{\Delta}^{(n)}_{\boldsymbol{\theta};f_1}$ with $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{\boldsymbol{\theta};K}$

a rank-based version of the central sequence, such that

1. the observations are replaced by their ranks
2. the score-function $\varphi_{f_1}(\cdot)$ is replaced by an arbitrary score function K.

To this end we need an appropriate notion of “spherical ranks”.

A natural requirement for $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\boldsymbol{\theta};K}$ is its distribution-freeness under

$$\bigcup_{g_1} \mathrm{P}^{(n)}_{\boldsymbol{\theta};g_1}.$$

Invariance principle : If the family of distributions is invariant under the action of some group of transformations, then it is best to express $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\boldsymbol{\theta};K}$ in terms of the corresponding maximal invariant.

In order to proceed we thus need to answer the question :

which group of transformations generates $\bigcup_{g_1} \mathrm{P}^{(n)}_{\boldsymbol{\theta};g_1}$?

Fix $\boldsymbol{\theta} \in \mathcal{S}^{k-1}$ and note that

$$\mathbf{X} := (\mathbf{X}'\boldsymbol{\theta})\boldsymbol{\theta} + \sqrt{1 - (\mathbf{X}'\boldsymbol{\theta})^2}\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X})$$

for all $\mathbf{X} \in \mathcal{S}^{k-1}$. Consider $\mathcal{G}_h^{(n)}$ transformations

$$g_h^{(n)} : (\mathbf{X}_1, \ldots, \mathbf{X}_n) \mapsto (g_h(\mathbf{X}_1), \ldots, g_h(\mathbf{X}_n))$$

with

$$g_h(\mathbf{X}_i) := h(\mathbf{X}_i'\boldsymbol{\theta})\boldsymbol{\theta} + \sqrt{1 - h(\mathbf{X}_i'\boldsymbol{\theta})^2}\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_i)$$

for nondecreasing $h : [-1, 1] \to [-1, 1]$ such that $h(-1) = h(1) = 0$.

Then

$$g_h(\mathbf{X}_i) := h(\mathbf{X}_i'\boldsymbol{\theta})\boldsymbol{\theta} + \sqrt{1 - h(\mathbf{X}_i'\boldsymbol{\theta})^2}\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_i)$$

is such that

1. $\|g_h(\mathbf{X}_i)\| = 1$ so that $g_h^{(n)}$ sends $\mathcal{S}^{k-1}$ to $\mathcal{S}^{k-1}$
2. $g_h(g_l(X_i)) = g_{h\circ l}(X_i)$, i.e. group action
3. $\mathcal{G}_h^{(n)}$ is generating group for $\bigcup_{g_1\in\mathcal{F}} \mathrm{P}^{(n)}_{\boldsymbol{\theta};g_1}$

Moreover

$$\mathbf{S}_{\boldsymbol{\theta}}(g_h(\mathbf{X}_i)) = \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_i),$$

i.e. signs are invariant under the action of $\mathcal{G}_h^{(n)}$.

Finally, letting

$$R_i = \text{Rank}(X_i) = \text{ the rank of } \mathbf{X}_i'\boldsymbol{\theta} \text{ among } \mathbf{X}_1'\boldsymbol{\theta}, \ldots, \mathbf{X}_n'\boldsymbol{\theta}$$

we see that

$$\text{Rank}(g_h(X_i)) = \text{Rank}(X_i)$$

i.e. the ranks R_i are invariant under the action of $\mathcal{G}_h^{(n)}$.

It follows that any statistic measurable with respect to the signs $\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_i)$ and ranks R_i is distribution-free under $\bigcup_{g_1} \mathrm{P}^{(n)}_{\boldsymbol{\theta};g_1}$.

Our spherical version of the Le Cam methodology : base inference procedures on sign- and rank-based version of the parametric central sequence

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\boldsymbol{\theta};K} := n^{-1/2} \sum_{i=1}^{n} K\left(\frac{R_i}{n+1}\right) \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_i),$$

where K is a score function satisfying some regularity assumption.

Some more maths (including asymptotic linearity results) then allow to construct

- Rank based estimators for $\boldsymbol{\theta}$ (Paper 1, Statistica Sinica)
- Pseudo-FVML and Rank based tests for ANOVA problems

$$\mathcal{H}_0 : \boldsymbol{\theta}_1 = \cdots = \boldsymbol{\theta}_k$$

  (Paper 2, submitted)

in both cases based on properties of $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\boldsymbol{\theta};K}$.

These procedures are

1. valid under the whole family of rotationally symmetric distributions ;
2. optimal for given f_1 ;
3. behave quite well with respect to the more common estimators / tests from the literature even in situations where the latter are known to be optimal.

In the estimation problem, comparing to the spherical mean

$$\hat{\boldsymbol{\theta}} = \sum_{i=1}^{n} \mathbf{X}_i / || \sum_{i=1}^{n} \mathbf{X}_i ||$$

(MLE of $\boldsymbol{\theta}$ under the FVML distribution).

	AREs with respect to the spherical mean (ARE($\hat{\boldsymbol{\theta}}_{f_0} / \hat{\boldsymbol{\theta}}_{\mathrm{Mean}}$))					
Underlying density	$\hat{\boldsymbol{\theta}}_{\mathrm{FVML}(2)}$	$\hat{\boldsymbol{\theta}}_{\mathrm{FVML}(6)}$	$\hat{\boldsymbol{\theta}}_{\mathrm{Lin}(2)}$	$\hat{\boldsymbol{\theta}}_{\mathrm{Lin}(4)}$	$\hat{\boldsymbol{\theta}}_{\mathrm{Log}(2.5)}$	$\hat{\boldsymbol{\theta}}_{\mathrm{Logis}(1,1)}$
FVML(1)	0.9744	0.8787	0.9813	0.9979	0.9027	0.9321
FVML(2)	1	0.9556	0.9978	0.9586	0.9749	0.9823
FVML(6)	0.9555	1	0.9381	0.8517	0.9768	0.9911
Lin(2)	1.0539	0.9909	1.0562	1.0215	1.0212	1.0247
Lin(4)	0.9709	0.8627	0.9795	1.0128	0.8856	0.9231
Log(2.5)	1.1610	1.1633	1.1514	1.0413	1.1908	1.1625
Log(4)	1.0182	0.9216	1.0261	1.0347	0.9503	0.9741
Logis(1,1)	1.0768	1.0865	1.0635	0.9991	1.0701	1.0962
Logis(2,1)	1.3182	1.4426	1.2946	1.0893	1.4294	1.3865

TABLE : Asymptotic relative efficiencies (AREs) of R-estimators with respect to the spherical mean under various 3-dimensional rotationally symmetric densities.