Le Cam's Asymptotic Theory in a Nutshell

Marc Hallin

ECARES, Université Libre de Bruxelles

London, December 11, 2015

Outline

An introduction to Le Cam's asymptotic theory
of statistical experiments

- 1. Local Asymptotic Normality (LAN)
- 2. Contiguity. Le Cam's first Lemma
- 3. Le Cam's third Lemma
- 4. Convergence of statistical experiments
- 5. Locally asymptotically optimal tests
- 6. Locally asymptotically optimal estimators

Outline

An introduction to Le Cam's asymptotic theory
of statistical experiments

- 1. Local Asymptotic Normality (LAN). Examples
- 2. Contiguity. Le Cam's first Lemma
- 3. Le Cam's third Lemma
- 4. Convergence of statistical experiments
- 5. Locally asymptotically optimal tests
- 6. Locally asymptotically optimal estimators

local parameters

Denote by $\mathcal{E}^{(n)} := (\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)} := \{P^{(n)}_{\boldsymbol{\theta}} \,|\, \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^k\})$ a sequence of parametric statistical models (statistical *experiments*) indexed by $\boldsymbol{\theta}$. For convenience, let $\boldsymbol{\Theta} = \mathbb{R}^k$

Consider sequences of parameter values of the form

$$\boldsymbol{\theta} + \boldsymbol{\nu}_n \boldsymbol{\tau}, \quad \boldsymbol{\tau} \in \mathbb{R}^k,$$

where $\boldsymbol{\nu}_n$ is a sequence of $k \times k$ full-rank matrices such that $\boldsymbol{\nu}_n \to 0$ at some adequate rate: $\boldsymbol{\tau}$ will be called a local parameter (localized at $\boldsymbol{\theta}$). More generally, any sequence of the form

$$\boldsymbol{\theta}^{(n)} := \boldsymbol{\theta} + \boldsymbol{\nu}_n \boldsymbol{\tau}^{(n)}$$

where $\boldsymbol{\tau}^{(n)}$ is a bounded sequence of $\mathbb{R}^k$ will be called a local alternative (to $\boldsymbol{\theta}$)

examples

The matrix $\boldsymbol{\nu}$ reflects the local rate(s) of asymptotics in the experiment. In most "classical" cases, this rate is $n^{1/2}$, and $\boldsymbol{\nu}^{(n)}$ is simply $n^{-1/2}\mathbf{I}$

Example 1. In the classical one-sample location model, local alternatives are of the form $\theta^{(n)} = \theta + n^{-1/2}\tau^{(n)}$

Example 2. In the general linear model with independent observations $\mathbf{Y}_i = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, local parameters are of the form $\boldsymbol{\beta}^{(n)} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1/2}\boldsymbol{\tau}^{(n)}$

LAN

In the sequence of experiments
$\mathcal{E}^{(n)} = (\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)} := \{P^{(n)}_{\boldsymbol{\theta}} \,|\, \boldsymbol{\theta} \in \mathbb{R}^k\})$, denote by

$$\Lambda^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}_n/\boldsymbol{\theta}} := \log \frac{dP^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}_n}}{dP^{(n)}_{\boldsymbol{\theta}}}$$

the logarithm of the local log-likelihood ratio $\frac{dP^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}_n}}{dP^{(n)}_{\boldsymbol{\theta}}}$
($P^{(n)}_{\boldsymbol{\theta}}$-uniquely defined)

Definition: the sequence $\mathcal{E}^{(n)}$ is LAN (*locally asymptotically normal*) if, for all $\boldsymbol{\theta}$, there exist a sequence $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}}$ of $\mathcal{A}^{(n)} \otimes \mathcal{B}_k$- -measurable k-dimensional random vectors, and a $(k \times k)$ matrix $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}$ such that, for all bounded $\boldsymbol{\tau}_n$, under $P^{(n)}_{\boldsymbol{\theta}}$, as $n \to \infty$,

- (i) $\Lambda^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}_n/\boldsymbol{\theta}} = \boldsymbol{\tau}'_n \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}} - \frac{1}{2}\boldsymbol{\tau}'_n \boldsymbol{\Gamma}_{\boldsymbol{\theta}} \boldsymbol{\tau}_n + o_{\mathrm{P}}(1)$, and
- (ii) $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \boldsymbol{\Gamma}_{\boldsymbol{\theta}})$.

Remarks

- $\mathbf{\Delta}_{\boldsymbol{\theta}}^{(n)}$ is called a central sequence (localized at $\boldsymbol{\theta}$); $\mathbf{\Gamma}_{\boldsymbol{\theta}}$ is the information matrix .
- $\mathbf{\Delta}_{\boldsymbol{\theta}}^{(n)}$ is defined up to $o_{\mathrm{P}_{\boldsymbol{\theta}}^{(n)}}(1)$ quantities
- LAN does not depend on the version of the log-likelihood $\Lambda^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}_n/\boldsymbol{\theta}}$ adopted in (i)
- more generally, under $P_{\boldsymbol{\theta}}^{(n)}$, as $n \to \infty$

$$(\boldsymbol{\tau}_n'\mathbf{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\tau}_n)^{-1/2}(\Lambda^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}_n/\boldsymbol{\theta}} + \frac{1}{2}\boldsymbol{\tau}_n'\mathbf{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\tau}_n) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

ULAN

Actually, for most applications, we require Uniform Local Asymptotic Normality (ULAN), that is, replacing (i) with the slightly stronger condition

- (i′) for any local sequence $\boldsymbol{\theta}^{(n)}$ such that $\boldsymbol{\nu}_n^{-1}(\boldsymbol{\theta}^{(n)} - \boldsymbol{\theta}) = O(1)$ and all bounded sequence $\boldsymbol{\tau}_n$, $\Lambda^{(n)}_{\boldsymbol{\theta}^{(n)}+\boldsymbol{\nu}_n\boldsymbol{\tau}_n/\boldsymbol{\theta}^{(n)}} = \boldsymbol{\tau}_n' \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}^{(n)}} - \frac{1}{2}\boldsymbol{\tau}_n'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\tau}_n + o_{\mathrm{P}}(1)$, with $\boldsymbol{\theta} \mapsto \boldsymbol{\Gamma}(\boldsymbol{\theta})$ continuous
- moreover, for convenience, we also assume that $\boldsymbol{\Gamma}(\boldsymbol{\theta})$ has full rank k
- then, it is easy to show that ULAN is equivalent to LAN and asymptotic linearity of the central sequence $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}}$, namely,

$$\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}^{(n)}} - \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}} = -\boldsymbol{\Gamma}(\boldsymbol{\theta})\boldsymbol{\tau}^{(n)} + o_{\mathrm{P}}(1) \quad \text{under } \mathrm{P}^{(n)}_{\boldsymbol{\theta}}, \text{ as } n \to \infty$$

Why "LAN"?

Denote by $\mathbf{\Delta}$ the (unique) observation in the k-dimensional Gaussian location (or Gaussian shift) model

$$\mathcal{E}_{\mathbf{\Gamma}}^{\mathcal{N}} = \Big(\mathbb{R}^k, \mathcal{B}^k, \mathcal{P}_{\mathbf{\Gamma}} = \big\{P_{\boldsymbol{\tau}} = \mathcal{N}_k(\mathbf{\Gamma}\boldsymbol{\tau}, \mathbf{\Gamma}) \,|\, \boldsymbol{\tau} \in \mathbb{R}^k\big\}\Big),$$

where $\mathbf{\Gamma}$ is a specified covariance matrix. Straightforward calculation yields

$$\Lambda_{\boldsymbol{\tau}/\mathbf{0}}(\mathbf{\Delta}) := \log \frac{dP_{\boldsymbol{\tau}}}{dP_{\mathbf{0}}}(\mathbf{\Delta}) = \boldsymbol{\tau}'\mathbf{\Delta} - \frac{1}{2}\boldsymbol{\tau}'\mathbf{\Gamma}\boldsymbol{\tau}$$

This resemblance with Gaussian location, as we shall see, is not just a coincidence ...

Examples

(a) the one-sample location model, under which the observation $\mathbf{X}^{(n)} = (X_1^{(n)}, \ldots, X_n^{(n)})$ satisfies

$$X_i^{(n)} = \theta + \varepsilon_i^{(n)}, \quad \theta \in \mathbb{R}, \quad , \varepsilon_i^{(n)} \text{ i.i.d., with density } f,$$

where f is such that

- $f(x) > 0 \; \forall x \in \mathbb{R}$,
- $\int_{\mathbb{R}} x f(x)\, dx = 0$,
- f absolutely continuous over finite intervals, that is, there exists a function $\dot{f} : \mathbb{R} \to \mathbb{R}$ such that, $\forall a < b$, $f(b) - f(a) = \int_a^b \dot{f}(x)\, dx$, and,
- letting $\varphi := -\dot{f}/f$, $\mathcal{I}_f := \int_{\mathbb{R}} \varphi_f^2(x)\, f(x)\, dx < \infty$

(actually, absolute continuity of f can be weakened into quadratic mean differentiability of $f^{1/2}$)

local alternatives are of the form $\theta + n^{-1/2}\tau_n$, with log-likelihoods

$$\Lambda^{(n)}_{\theta+n^{-1/2}\tau_n/\theta;f} := \log \frac{d\mathrm{P}^{(n)}_{\theta+n^{-1/2}\tau_n;f}}{d\mathrm{P}^{(n)}_{\theta;f}} = \log \left[\frac{\prod_{i=1}^n f\big(X_i^{(n)} - \theta - n^{-1/2}\tau_n\big)}{\prod_{i=1}^n f\big(X_i^{(n)} - \theta\big)} \right]$$

$$= \sum_{i=1}^n \Big[\log f\big(Z_i^{(n)}(\theta) - n^{-1/2}\tau_n\big) - \log f\big(Z_i^{(n)}(\theta)\big)\Big],$$

with $Z_i^{(n)}(\theta) := X_i^{(n)} - \theta$

This model is ULAN, with central sequence

$$\Delta^{(n)}_{\theta;f} := n^{-1/2}\sum_{i=1}^n \varphi_f\big(Z_i^{(n)}(\theta)\big) \quad \text{and information} \quad \Gamma_\theta := \Gamma_{\theta;f} := \mathcal{I}_f$$

(b) the general linear model , under which the observation $\mathbf{X}^{(n)} = (X_1^{(n)}, \ldots, X_n^{(n)})$ satisfies

$$X_i^{(n)} = \left(\mathbf{c}_i^{(n)}\right)'\boldsymbol{\beta} + \varepsilon_i^{(n)} = \sum_{k=1}^{K} c_{ik}^{(n)} \beta_k + \varepsilon_i^{(n)}, \quad \varepsilon_i^{(n)} \text{ i.i.d., with density } f,$$

with $\mathbf{c}_i^{(n)} := (c_{i1}^{(n)}, \ldots, c_{iK}^{(n)})'$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)'$; f is assumed to satisfy the same assumptions as in the location model, and the regression constants are such that

$$\lim_{n\to\infty} \frac{\max_{1\leq i\leq n}(c_{ik}^{(n)} - \bar{c}_k^{(n)})^2}{\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}_k^{(n)})^2} = 0, \text{ with } \bar{c}_k^{(n)} := \frac{1}{n}\sum_{i=1}^{n} c_{ik}^{(n)},$$

$\mathbf{C}^{(n)} := \frac{1}{n}\sum_{i=1}^{n} \mathbf{c}_i^{(n)}\left(\mathbf{c}_i^{(n)}\right)'$ is positive definite for all n, and $(\mathbf{R}^{(n)})^{-2} := \mathbf{C}^{(n)} \to \mathbf{C} =: \mathbf{R}^{-2}$, where $\mathbf{C}$ is of full rank.

The parameter is $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)'$; local alternatives are of the form $\boldsymbol{\beta} + \boldsymbol{\nu}_n \boldsymbol{\tau}_n$, with

$$\boldsymbol{\nu}_n = n^{-1/2}\mathbf{R}^{(n)} \ \text{ and } \ \boldsymbol{\tau}_n \in \mathbb{R}^K.$$

This model is ULAN, with central sequence

$$\boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(n)} := \boldsymbol{\Delta}_{\boldsymbol{\beta};f}^{(n)} := n^{-1/2}\big(\mathbf{R}^{(n)}\big)' \sum_{i=1}^{n} \varphi_f\big(Z_i^{(n)}(\boldsymbol{\beta})\big)\mathbf{c}_i^{(n)}$$

where $Z_i^{(n)}(\boldsymbol{\beta}) := X_i^{(n)} - \big(\mathbf{c}_i^{(n)}\big)'\boldsymbol{\beta}$, and with information matrix $\boldsymbol{\Gamma}_{\boldsymbol{\beta};f} := \mathcal{I}_f \mathbf{I}_K$,

(c) AR(K) models; here the observation $\mathbf{X}^{(n)} = (X_1^{(n)}, \ldots, X_n^{(n)})$ satisfies

$$X_t^{(n)} = \sum_{k=1}^{K} \theta_k X_{t-k}^{(n)} + \varepsilon_t^{(n)}, \quad \varepsilon_t^{(n)} \text{ i.i.d., with density } f.$$

with $1 - \sum_{k=1}^{K} \theta_k z^k \neq 0$ for $|z| \leq 1$, $z \in \mathbb{C}$ (causality or fundalmentalness); f is assumed to satisfy the same assumptions as in the location model, with, in addition, $\sigma_f^2 := \int x^2 f(x) dx < \infty$.

The parameter is $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)'$; local alternatives are of the form $\boldsymbol{\theta} + \boldsymbol{\nu}_n \boldsymbol{\tau}_n$, with $\boldsymbol{\nu}_n = n^{-1/2}\mathbf{I}_K$ and $\boldsymbol{\tau}_n \in \mathbb{R}^K$

This model is ULAN, with central sequence

$$\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta};f} := n^{-1/2}\sum_{t=K+1}^{n} \varphi_f\big(Z_t^{(n)}(\boldsymbol{\theta})\big)\begin{pmatrix} X_{t-1} \\ \vdots \\ X_{t-K}\end{pmatrix} = \frac{n^{-1/2}}{\sigma_f}\sum_{t=1}^{n}\varphi_{f_1}\Big(\frac{Z_t(\boldsymbol{\theta})}{\sigma_f}\Big)\begin{pmatrix} \sum_{u=0}^{t-1} g_u(\boldsymbol{\theta}) Z_{t-1-u} \\ \vdots \\ \sum_{u=0}^{t-1} g_u(\boldsymbol{\theta}) Z_{t-K-u}\end{pmatrix}$$

where $Z_i^{(n)}(\boldsymbol{\theta}) := X_t^{(n)} - \sum_{k=1}^K \theta_k X_{t-k}^{(n)}$ and impulse-response coefficients (the Green's functions) g_u characterized by $\left(1 - \sum_{k=1}^K \theta_k L^k\right)^{-1} =: \sum_{u=0}^{\infty} g_u L^u$, and with information matrix $\boldsymbol{\Gamma}_{\boldsymbol{\theta};f_1} := \mathcal{I}_{f_1}\boldsymbol{\Gamma}(\boldsymbol{\theta})$ where $\boldsymbol{\Gamma}(\boldsymbol{\theta})$ is the $K \times K$ covariance matrix of the stationary solution of the autoregressive equation

Outline

An introduction to Le Cam's
asymptotic theory of statistical experiments

- 1. Local Asymptotic Normality (LAN)
- 2. Contiguity. Le Cam's first Lemma
- 3. Le Cam's third Lemma
- 4. Convergence of statistical experiments
- 5. Locally asymptotically optimal tests
- 6. Locally asymptotically optimal estimators

Contiguity

Denote by $\mathrm{P}^{(n)}$ and $Q^{(n)}$ two sequences of probability measures over $(\mathcal{X}^{(n)}, \mathcal{A}^{(n)})$

Definition: $Q^{(n)}$ is contiguous to $\mathrm{P}^{(n)}$ (notation: $Q^{(n)} \lhd P^{(n)}$) if, for all $A^{(n)} \in \mathcal{A}^{(n)}$, $\mathrm{P}^{(n)}[A^{(n)}] \to 0$ implies $Q^{(n)}[A^{(n)}] \to 0$

Contiguity thus is a form of asymptotic absolute continuity of the sequence $Q^{(n)}$ wrt the sequence $\mathrm{P}^{(n)}$

Remarks:

- whenever $Q^{(n)} \lhd P^{(n)}$ and $\mathrm{P}^{(n)} \lhd Q^{(n)}$, we say that $\mathrm{P}^{(n)}$ and $Q^{(n)}$ are mutually contiguous (notation: $\mathrm{P}^{(n)} \bowtie Q^{(n)}$).
- If $Q^{(n)} \lhd \mathrm{P}^{(n)}$, $S_n = o_{\mathrm{P}^{(n)}}(h_n)$ implies that $S_n = o_{Q^{(n)}}(h_n)$

Contiguity is also a form of asymptotic non-distinguishability:

consider a sequence of nonrandomized tests with critical region $A^{(n)}$ for the testing problem

$$\left\{ \begin{array}{ll} \mathcal{H}_0^{(n)} & : \{\mathrm{P}^{(n)}\} \\ \mathcal{H}_1^{(n)} & : \{Q^{(n)}\} \end{array} \right.$$

If $\mathrm{P}^{(n)} \bowtie Q^{(n)}$, then

- (a) $\mathrm{P}^{(n)}[A^{(n)}] \to 0 \Rightarrow Q^{(n)}[A^{(n)}] \to 0$: if the probability level tends to 0, so does the power
- (b) $Q^{(n)}[A^{(n)}] \to 1 \Rightarrow \mathrm{P}^{(n)}[A^{(n)}] \to 1$: if the power tends to 1, so does type I risk

Le Cam's first Lemma

A famous sufficient condition for $Q^{(n)} \lhd \mathrm{P}^{(n)}$ is given by the so-called Le Cam first Lemma:

Lemma (Hájek and Šidák): *Let* $F_n(x) := \mathrm{P}^{(n)}[\frac{dQ^{(n)}}{d\mathrm{P}^{(n)}} \leq x]$. *If* F_n *converges weakly to* F, *where* $\int x\, dF(x) = 1$, *then* $Q^{(n)} \lhd \mathrm{P}^{(n)}$
If moreover $F(0) = 0$, *then* $\mathrm{P}^{(n)} \bowtie Q^{(n)}$

Application to LAN families ($\mathrm{P}^{(n)} = \mathrm{P}^{(n)}_{\boldsymbol{\theta}}$ and $Q^{(n)} = P^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}}$):

under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$, $\frac{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}}}{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}}} = e^{\Lambda^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}/\boldsymbol{\theta}}} \xrightarrow{\mathcal{L}} e^Z$, where $Z \sim \mathcal{N}\big(-\frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\tau},\ \boldsymbol{\tau}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\tau}\big)$

hence, $F_n(x) := \mathrm{P}^{(n)}_{\boldsymbol{\theta}}[\frac{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}}}{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}}} \leq x]$ converges to $F(x) = \mathrm{P}[e^Z \leq x]\ \forall x$

But for $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mathrm{E}[e^X] = e^{\mu + \frac{1}{2}\sigma^2}$, so that

$$\int x\, dF(x) = \mathrm{E}[e^Z] = \exp\Big[-\frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\tau} + \frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\tau} \Big] = 1.$$

Le Cam's first Lemma thus implies that $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}} \lhd \mathrm{P}^{(n)}_{\boldsymbol{\theta}}$.

Since moreover $F(0) = \mathrm{P}[e^Z \leq 0] = 0$, we have that $\mathrm{P}^{(n)}_{\boldsymbol{\theta}} \bowtie \mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}}$ for all $\boldsymbol{\tau}$: loca

---

More generally, $\mathrm{P}^{(n)}_{\boldsymbol{\theta}} \bowtie \mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}_n}$

Outline

2. An introduction to Le Cam's
asymptotic theory of statistical experiments

- 1. Local Asymptotic Normality (LAN)
- 2. Contiguity. Le Cam's first Lemma
- 3. Le Cam's third Lemma
- 4. Convergence of statistical experiments
- 5. Locally asymptotically optimal tests
- 6. Locally asymptotically optimal estimators

Le Cam's Third Lemma

Contiguity is essentially about convergence in probability; under similar conditions, can we also say something about convergence in distribution?

The following result answers that question

Lemma (Hájek and Šidák): *Denote by* Λ_n *a version of* $\log \frac{dQ^{(n)}}{d\mathrm{P}^{(n)}}$, *and let* $S^{(n)}$ *be* $\mathcal{A}^{(n)}$*-mesurable. Assume that, under* $\mathrm{P}^{(n)}$*, as* $n \to \infty$*,*

$$\begin{pmatrix} S^{(n)} \\ \Lambda^{(n)} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} \mu \\ -\frac{1}{2}d^2 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma_{12} \\ \sigma_{12} & d^2 \end{pmatrix}\right).$$

Then,

- *(i)* $\mathrm{P}^{(n)} \bowtie Q^{(n)}$*;*
- *(ii) under* $Q^{(n)}$*, as* $n \to \infty$*,* $S^{(n)} \xrightarrow{\mathcal{L}} \mathcal{N}(\mu + \sigma_{12}, \sigma^2)$

Application to LAN families

Applying the Lemma to the central sequence of a LAN model, we readily obtain that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}}$, as $n \to \infty$,

$$\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta};f} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\boldsymbol{\Gamma}_{\boldsymbol{\theta};f}\boldsymbol{\tau},\ \boldsymbol{\Gamma}_{\boldsymbol{\theta};f}\right)$$

Thus, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}}$

$$\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta};f} \xrightarrow{\mathcal{L}} \boldsymbol{\Delta}$$

where $\boldsymbol{\Delta}$ is the (unique) observation in the Gaussian shift model

$$\mathcal{E}_{\boldsymbol{\theta}} := \left(\mathbb{R}^K, \mathcal{B}^K, \mathcal{N}\left(\boldsymbol{\Gamma}_{\boldsymbol{\theta};f}\boldsymbol{\tau},\ \boldsymbol{\Gamma}_{\boldsymbol{\theta};f}\right) \mid \boldsymbol{\tau} \in \mathbb{R}^K\right)$$

The relation between the local experiments

$$\mathcal{E}^{(n)}_{\boldsymbol{\theta}} := \left(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \{\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}_n\boldsymbol{\tau}} \mid \boldsymbol{\tau} \in \mathbb{R}^K\}\right)$$

(parameter: $\boldsymbol{\tau}$) and the Gaussian shift $\mathcal{E}_{\boldsymbol{\theta}}$ (parameter: $\boldsymbol{\tau}$) actually is much stronger and essential: as we shall see now, $\mathcal{E}^{(n)}_{\boldsymbol{\theta}}$ converges (in a sense to be defined) to $\mathcal{E}_{\boldsymbol{\theta}}$

Outline

2. An introduction to Le Cam's
asymptotic theory of statistical experiments

- 1. Local Asymptotic Normality (LAN)
- 2. Contiguity. Le Cam's first Lemma
- 3. Le Cam's third Lemma
- 4. Convergence of statistical experiments
- 5. Locally asymptotically optimal tests
- 6. Locally asymptotically optimal estimators

Convergence of statistical experiments

Consider a statistical experiment $\mathcal{E} := (\mathcal{X}, \mathcal{A}, \mathcal{P} = \{\mathrm{P}_{\boldsymbol{\theta}}^{(\backslash)}\})$ indexed by $\boldsymbol{\theta}$. A statistical procedure ρ (with value in a decision space $\mathcal{D}$) and a loss function W define a risk function $\boldsymbol{\theta} \mapsto R_W^\rho(\boldsymbol{\theta})$

Denote by $\mathcal{R}(\mathcal{P}, \mathcal{D}, W)$ the set of all risk functions that can be implemented from the family of distributions $\mathcal{P}$ with $\mathcal{D}$-valued statistical procedures and the loss function W. More precisely, let

$$\mathcal{R}(\mathcal{P}, \mathcal{D}, W) := \{r : \mathbb{R}^K \to \mathbb{R} \mid \exists \rho \text{ such that } R_W^\rho(\boldsymbol{\theta}) \leq r(\boldsymbol{\theta}) \ \forall \boldsymbol{\theta}\}$$

and denote by $\bar{\mathcal{R}}(\mathcal{P}, \mathcal{D}, W)$ the pointwise closure of $\mathcal{R}(\mathcal{P}, \mathcal{D}, W)$:

$$\bar{\mathcal{R}}(\mathcal{P}, \mathcal{D}, W) := \{r : \mathbb{R}^K \to \mathbb{R} \mid \forall \boldsymbol{\theta},\ r(\boldsymbol{\theta}) = \lim_{i \to \infty} r_i(\boldsymbol{\theta}) \text{ for some sequence } r_i \in \mathcal{R}(\mathcal{P}, \mathcal{D}, W)\}$$

The Le Cam distance

Consider two statistical experiment, $\mathcal{E}_1 := (\mathcal{X}_1, \mathcal{A}_1, \mathcal{P}_1 = \{\mathrm{P}^{(n)}_{1:\boldsymbol{\theta}}\})$ and $\mathcal{E}_2 := (\mathcal{X}_2, \mathcal{A}_2, \mathcal{P}_2 = \{\mathrm{P}^{(n)}_{2:\boldsymbol{\theta}}\})$, indexed by the same parameter $\boldsymbol{\theta}$

Definition 1. The *deficiency* $\delta(\mathcal{E}_1, \mathcal{E}_2)$ of $\mathcal{E}_1$ with respect to $\mathcal{E}_2$ is

$$\delta(\mathcal{E}_1, \mathcal{E}_2) := \inf \Big\{\epsilon \in [0,1] \,\Big|\, \forall \mathcal{D}, \forall W : \mathbb{R}^K \to [0,1], \forall r_2 \in \mathcal{R}(\mathcal{P}_2, \mathcal{D}, W)$$

$$\exists\, r_1 \in \bar{\mathcal{R}}(\mathcal{P}_1, \mathcal{D}, W) : \; r_1(\boldsymbol{\theta}) \leq r_2(\boldsymbol{\theta}) + \epsilon \;\forall \boldsymbol{\theta} \in \mathbb{R}^K \Big\}$$

Definition 2. The *Le Cam distance* $\Delta(\mathcal{E}_1, \mathcal{E}_2)$ between $\mathcal{E}_1$and $\mathcal{E}_2$ is

$$\Delta(\mathcal{E}_1, \mathcal{E}_2) := \max\left(\delta(\mathcal{E}_1, \mathcal{E}_2), \; \delta(\mathcal{E}_2, \mathcal{E}_1)\right)$$

Interpretation: all risks functions $\boldsymbol{\theta} \mapsto r(\boldsymbol{\theta})$ that can be achieved via $\mathcal{E}_1$ also can be achieved, within $\pm\Delta$, via $\mathcal{E}_2$

Weak convergence of experiments

The Le Cam distance naturally induces a concept of convergence for sequences of experiments. Uniform (in $\boldsymbol{\theta}$) convergence however would be too demanding; the following convergence is uniform over all risk functions, but pointwise in $\boldsymbol{\theta}$:

Definition 3. A sequence of experiments $\mathcal{E}^{(n)} := (\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)} = \{\mathrm{P}_{\boldsymbol{\theta}}^{(n)}\})$ (all indexed by the same parameter $\boldsymbol{\theta}$) converges weakly to the (limit-experiment) $\mathcal{E} := (\mathcal{X}, \mathcal{A}, \mathcal{P} = \{\mathrm{P}_{\boldsymbol{\theta}}\})$ if, for all $m \in \mathbb{N}$ and all $\boldsymbol{\Theta}_0 := \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m\}$, the subexperiments $\mathcal{E}_{\boldsymbol{\Theta}_0}^{(n)} := (\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}_{\boldsymbol{\Theta}_0}^{(n)} := \{\mathrm{P}_{\boldsymbol{\theta}_i}^{(n)}, i = 1, \ldots, m\})$ and $\mathcal{E}_{\boldsymbol{\Theta}_0} := (\mathcal{X}, \mathcal{A}, \mathcal{P}_{\boldsymbol{\Theta}_0} := \{\mathrm{P}_{\boldsymbol{\theta}_i}, i = 1, \ldots, m\})$ are such that

$$\lim_{n\to\infty} \Delta(\mathcal{E}_{\boldsymbol{\Theta}_0}^{(n)}, \mathcal{E}_{\boldsymbol{\Theta}_0}) = 0$$

Interpretation: the set of all risks functions $\boldsymbol{\theta} \mapsto r(\boldsymbol{\theta})$ that can be achieved in the limit experiment $\mathcal{E}$ is the the “uniform pointwise” limit of those that can be achieved in $\mathcal{E}^{(n)}$

Le Cam's weak convergence theorem

This definition of weak convergence looks quite formidable and, at first sight, completely inapplicable

The following result shows that this is not the case. Associated with the experiment $\mathcal{E} := (\mathcal{X}, \mathcal{A}, \mathcal{P} := \{\mathrm{P}_{\boldsymbol{\theta}} \mid \boldsymbol{\theta} \in \mathbb{R}^K\})$, define the log-likelihood process

$$\left\{\Lambda(\mathbf{s}, \mathbf{t}) := \log(d\mathrm{P}_{\mathbf{s}}/d\mathrm{P}_{\mathbf{t}}) \mid \mathbf{s}, \mathbf{t} \in \mathbb{R}^K\right\}.$$

Similarly, associated with the sequence $\mathcal{E}^{(n)} := (\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)} := \{\mathrm{P}^{(n)}_{\boldsymbol{\theta}} \mid \boldsymbol{\theta} \in \mathbb{R}^K\})$, is

$$\left\{\Lambda^{(n)}(\mathbf{s}, \mathbf{t}) := \log(d\mathrm{P}^{(n)}_{\mathbf{s}}/d\mathrm{P}^{(n)}_{\mathbf{t}}) \mid \mathbf{s}, \mathbf{t} \in \mathbb{R}^K\right\}$$

Le Cam (1969) then proved the following very simple necessary and sufficient condition for weak convergence of $\mathcal{E}^{(n)}$ to $\mathcal{E}$

Le Cam's weak convergence theorem

Theorem. (Le Cam 1969) *The sequence of experiments $\mathcal{E}^{(n)}$ converges weakly to $\mathcal{E}$ iff for all $\mathbf{t} \in \mathbb{R}^k$, the finite-dimensional vectors of $\mathcal{E}^{(n)}$'s log-likelihood process converge in distribution to the corresponding ones for $\mathcal{E}$, namely, iff*

$$\forall J \in \mathbb{N},\ \forall \mathbf{s}_1, \ldots, \mathbf{s}_J,\ \forall \mathbf{t} \qquad \begin{pmatrix} \Lambda^{(n)}(\mathbf{s}_1, \mathbf{t}) \\ \vdots \\ \Lambda^{(n)}(\mathbf{s}_J, \mathbf{t}) \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} \Lambda(\mathbf{s}_1, \mathbf{t}) \\ \vdots \\ \Lambda(\mathbf{s}_J, \mathbf{t}) \end{pmatrix}$$

under $\mathrm{P}^{(n)}_{\mathbf{t}}$, as $n \to \infty$

LAN: convergence of local experiments

This theorem is taylor-made for local experiments under LAN. Indeed, denoting by $\Lambda^{(n)}(\mathbf{s}, \mathbf{t})$ the log-likelihood processes associated with the sequence of local experiments at $\boldsymbol{\theta}$ ($\mathbf{s}$, $\mathbf{t}$ here stand for values of the local parameter $\boldsymbol{\tau}$), under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\mathbf{t}}$, as $n \to \infty$,

$$\begin{pmatrix} \Lambda^{(n)}(\mathbf{s}_1, \mathbf{t}) \\ \vdots \\ \Lambda^{(n)}(\mathbf{s}_J, \mathbf{t}) \end{pmatrix} = \begin{pmatrix} \Lambda^{(n)}(\mathbf{s}_1, \mathbf{0}) - \Lambda^{(n)}(\mathbf{t}, \mathbf{0}) \\ \vdots \\ \Lambda^{(n)}(\mathbf{s}_J, \mathbf{0}) - \Lambda^{(n)}(\mathbf{t}, \mathbf{0}) \end{pmatrix}$$

$$= \begin{pmatrix} (\mathbf{s}_1 - \mathbf{t})' \\ \vdots \\ (\mathbf{s}_J - \mathbf{t})' \end{pmatrix} \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}} - \frac{1}{2} \begin{pmatrix} (\mathbf{s}_1 - \mathbf{t})' \\ \vdots \\ (\mathbf{s}_J - \mathbf{t})' \end{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\theta}} \begin{pmatrix} (\mathbf{s}_1 + \mathbf{t})' \\ \vdots \\ (\mathbf{s}_J + \mathbf{t})' \end{pmatrix}' + o_{\mathrm{P}}(1)$$

Similarly, denoting by $\Lambda(\mathbf{s}, \mathbf{t})$ the log-likelihood processes associated with the Gaussian shift experiment $\boldsymbol{\Delta} \sim \mathcal{N}\left(\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\tau},\ \boldsymbol{\Gamma}_{\boldsymbol{\theta}}\right)$, $\boldsymbol{\tau} \in \mathbb{R}^K$, we obtain (exactly so)

$$\begin{pmatrix} \Lambda(\mathbf{s}_1, \mathbf{t}) \\ \vdots \\ \Lambda(\mathbf{s}_J, \mathbf{t}) \end{pmatrix} = \begin{pmatrix} \Lambda(\mathbf{s}_1, \mathbf{0}) - \Lambda(\mathbf{t}, \mathbf{0}) \\ \vdots \\ \Lambda(\mathbf{s}_J, \mathbf{0}) - \Lambda(\mathbf{t}, \mathbf{0}) \end{pmatrix}$$
$$= \begin{pmatrix} (\mathbf{s}_1 - \mathbf{t})' \\ \vdots \\ (\mathbf{s}_J - \mathbf{t})' \end{pmatrix} \boldsymbol{\Delta} - \frac{1}{2} \begin{pmatrix} (\mathbf{s}_1 - \mathbf{t})' \\ \vdots \\ (\mathbf{s}_J - \mathbf{t})' \end{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\theta}} \begin{pmatrix} (\mathbf{s}_1 + \mathbf{t})' \\ \vdots \\ (\mathbf{s}_J + \mathbf{t})' \end{pmatrix}'$$

Convergence of the local experiments to the Gaussian shift thus readily follows from the convergence in law, under any $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\mathbf{t}}$, of $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}}$ to $\boldsymbol{\Delta}$

Locally and asymptotically, thus,

- the risk functions (in the vicinity of $\boldsymbol{\theta}$) of the original experiment are those of the Gaussian shifts $\boldsymbol{\Delta} \sim \mathcal{N}\left(\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\tau},\ \boldsymbol{\Gamma}_{\boldsymbol{\theta}}\right)$; locally, the "asymptotically optimal" risk functions in the original experiment thus are the (exactly) optimal risk functions of the corresponding Gaussian shifts—which are well known
- moreover, since $\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)}$ converges in law to $\boldsymbol{\Delta}$, those risk functions are achieved by treating the central sequence $\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)}$ the way one would treat $\boldsymbol{\Delta}$ in the Gaussian shift (bounded loss functions: Helly-Bray applies)
- for instance, if a test $\phi(\boldsymbol{\Delta})$ is exactly most powerful at level α in the Gaussian shift, the sequence of tests $\phi(\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)})$ will be locally (at θ) and asymptotically most powerful (at asymptotic level α) in the original experiment

Outline

2. An introduction to Le Cam's
asymptotic theory of statistical experiments

- 1. Local Asymptotic Normality (LAN)
- 2. Contiguity. Le Cam's first Lemma
- 3. Le Cam's third Lemma
- 4. Convergence of statistical experiments
- 5. Locally asymptotically optimal tests
- 6. Locally asymptotically optimal estimators

Locally asymptotically maximin tests

Consider the problem of testing $\mathcal{H}^{(n)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $\mathcal{K}^{(n)} : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$

- local version (in the vicinity of $\boldsymbol{\theta}_0$, in terms of the local parameter $\boldsymbol{\tau}$): $\mathcal{H}^{(n)} : \boldsymbol{\tau} = \mathbf{0}$ against $\mathcal{K}^{(n)} : \boldsymbol{\tau} \neq \mathbf{0}$
- in the Gaussian shift: $\mathcal{H}^{(n)} : \boldsymbol{\Delta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\theta}})$ against $\mathcal{K}^{(n)} : \boldsymbol{\Delta} \sim \mathcal{N}(\boldsymbol{\mu} \neq \mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\theta}})$
- optimal (maximin at level α) test in the Gaussian shift: reject $\mathcal{H}^{(n)}$ whenever $\boldsymbol{\Delta}'\boldsymbol{\Gamma}'_{\boldsymbol{\theta}_0}\boldsymbol{\Delta}$ exceeds the $(1-\alpha)$-quantile $\chi^2_{1-\alpha;K}$ of the chi-square distribution with K d.f.
- locally asymptotically optimal (locally asymptotically maximin, at asymptotic level α) test (in the original experiment): reject $\mathcal{H}^{(n)}$ whenever $\boldsymbol{\Delta}^{(n)\prime}_{\boldsymbol{\theta}_0}\boldsymbol{\Gamma}_{\boldsymbol{\theta}_0}\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}_0}$ exceeds the $(1-\alpha)$-quantile $\chi^2_{1-\alpha;K}$ of the chi-squaredistribution with K d.f.

maximinity

the adequate concept of optimality here is *maximinity* (Wald 1943)

consider testing $\mathcal{H}$ against $\mathcal{K}$; let $\mathcal{C}$ be some class of tests

for all $\phi \in \mathcal{C}$, the "worst performance" against $\mathcal{K}$ is $\inf_{\mathrm{P}\in\mathcal{K}} \mathrm{E}_{\mathrm{P}}[\phi]$

call ϕ^* maximin for $\mathcal{H}$ against $\mathcal{K}$ within $\mathcal{C}$ if

- $\phi^* \in \mathcal{C}$
- $\inf_{\mathrm{P}\in\mathcal{K}} \mathrm{E}_{\mathrm{P}}[\phi^*] \geq \inf_{\mathrm{P}\in\mathcal{K}} \mathrm{E}_{\mathrm{P}}[\phi]$ for all $\phi \in \mathcal{C}$

In the Gaussian shift experiment $\mathbf{\Delta} \sim \mathcal{N}(\boldsymbol{\mu},\, \mathbf{\Gamma})$, the maximin test for $\mathcal{H} : \ \boldsymbol{\mu} = \mathbf{0}$ against $\mathcal{K} : \ \boldsymbol{\mu} \neq \mathbf{0}$ within the class $\mathcal{C}$ of α-level tests is trivially $\phi^* \equiv \alpha$

therefore, the classical optimality concept for this testing problem is the maximin test for $\mathcal{H} : \ \boldsymbol{\mu} = \mathbf{0}$ against $\mathcal{K}_c := \ \{\boldsymbol{\mu} : |\ \boldsymbol{\mu}'\mathbf{\Gamma}^{-1}\boldsymbol{\mu} \geq c\}$ within the class $\mathcal{C}$ of α-level tests; irrespective of $c > 0$, the solution is $\phi^* = I[\mathbf{\Delta}'\mathbf{\Gamma}^{-1}\mathbf{\Delta} > \chi^2_{K;1-\alpha}]$

Local powers

- power at $\boldsymbol{\tau} \neq \mathbf{0}$ of the α-level maximin test in the Gaussian shift: $1 - F^{\chi^2_K}(\chi^2_{1-\alpha;K}; \boldsymbol{\tau}'\boldsymbol{\Gamma}_{\boldsymbol{\theta}_0}\boldsymbol{\tau})$, where $F^{\chi^2_K}(\,.\,; \lambda^2)$ denotes the distribution function of a noncentral chi-square distribution with K d.F., and noncentrality parameter λ^2
- local asymptotic power (in the original experiment) at $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_0 + \boldsymbol{\nu}(n)\boldsymbol{\tau}}$ of the locally asymptotically maximin test: SAME

testing "subhypotheses"

Split $\boldsymbol{\theta}$ into $(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$, $\boldsymbol{\theta}_1 \in \mathbb{R}^{K_1}$, $\boldsymbol{\theta}_2 \in \mathbb{R}^{K_2}$ ($K_1 + K_2 = K$), and consider the problem of testing $\mathcal{H}^{(n)} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{1;0}$ against $\mathcal{K}^{(n)} : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_{1;0}$, with $\boldsymbol{\theta}_2$ playing the role of a nuisance. Partition $\boldsymbol{\tau}$ and $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}$ similarly into $\boldsymbol{\tau} =: (\boldsymbol{\tau}_1', \boldsymbol{\tau}_2')'$ and $\boldsymbol{\Gamma}_{\boldsymbol{\theta}} = \begin{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\theta};11} & \boldsymbol{\Gamma}_{\boldsymbol{\theta};12} \\ \boldsymbol{\Gamma}_{\boldsymbol{\theta};12}' & \boldsymbol{\Gamma}_{\boldsymbol{\theta};22} \end{pmatrix}$

- local version (in the vicinity of $(\boldsymbol{\theta}_{1;0}', \boldsymbol{\theta}_2')'$, in terms of the local parameter $\boldsymbol{\tau}$): $\mathcal{H}^{(n)} : \boldsymbol{\tau} = (\mathbf{0}', \boldsymbol{\tau}_2')'$
  against $\mathcal{K}^{(n)} : \boldsymbol{\tau} = (\boldsymbol{\tau}_1' \neq \mathbf{0}', \boldsymbol{\tau}_2')'$
- in the Gaussian shift:

  $$\mathcal{H}^{(n)} : \boldsymbol{\Delta} \sim \mathcal{N}(\boldsymbol{\Gamma}_{(\boldsymbol{\theta}_{1;0}', \boldsymbol{\theta}_2')'}(\mathbf{0}', \boldsymbol{\tau}_2')',\ \boldsymbol{\Gamma}_{(\boldsymbol{\theta}_{1;0}', \boldsymbol{\theta}_2')'}), \quad \boldsymbol{\tau}_2 \text{ unspecified}$$

  against

  $$\mathcal{K}^{(n)} : \boldsymbol{\Delta} \sim \mathcal{N}(\boldsymbol{\Gamma}_{(\boldsymbol{\theta}_{1;0}', \boldsymbol{\theta}_2')'}(\boldsymbol{\tau}_1', \boldsymbol{\tau}_2')',\ \boldsymbol{\Gamma}_{(\boldsymbol{\theta}_{1;0}', \boldsymbol{\theta}_2')'}),\ \boldsymbol{\tau}_1 \neq \mathbf{0},\ \boldsymbol{\tau}_2 \text{ unspecified}$$

stringency

the adequate concept of optimality is *stringency* (Wald 1943)
consider testing $\mathcal{H}$ against $\mathcal{K}$; let $\mathcal{C}$ be some class of tests

for all $\phi \in \mathcal{C}$, define the "regret of ϕ" within $\mathcal{C}$:

$$r(\phi) := \sup_{\mathrm{P}\in\mathcal{K}} \left\{ \sup_{\phi'\in\mathcal{C}} \mathrm{E}_{\mathrm{P}}[\phi'] - \mathrm{E}_{\mathrm{P}}[\phi] \right\}$$

call ϕ^* most stringent for $\mathcal{H}$ against $\mathcal{K}$ within $\mathcal{C}$ if

- $\phi^* \in \mathcal{C}$
- $r(\phi^*) \le r(\phi)$ for all $\phi \in \mathcal{C}$

stringency in Gaussian shifts

In the Gaussian shift model under which

$$\boldsymbol{\Delta} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\Gamma}_{11} & \boldsymbol{\Gamma}_{12} \\ \boldsymbol{\Gamma}'_{12} & \boldsymbol{\Gamma}_{22} \end{pmatrix}\begin{pmatrix} \boldsymbol{\tau}_1 \\ \boldsymbol{\tau}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Gamma}_{11} & \boldsymbol{\Gamma}_{12} \\ \boldsymbol{\Gamma}'_{12} & \boldsymbol{\Gamma}_{22} \end{pmatrix}\right),$$

- the most stringent α-level test of $\mathcal{H}: \boldsymbol{\tau}_1 = \mathbf{0}$ against $\mathcal{K}: \boldsymbol{\tau}_1 \neq \mathbf{0}$ is obtained by considering the residual $\boldsymbol{\Delta}_1^{\perp}$ of the regression of $\boldsymbol{\Delta}_1$ with respect to $\boldsymbol{\Delta}_2$ in the covariance matrix $\begin{pmatrix} \boldsymbol{\Gamma}_{11} & \boldsymbol{\Gamma}_{12} \\ \boldsymbol{\Gamma}'_{12} & \boldsymbol{\Gamma}_{22} \end{pmatrix}$: $\boldsymbol{\Delta}_1^{\perp} := \boldsymbol{\Delta}_1 - \boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_{22}^{-1}\boldsymbol{\Delta}_2$;
- elementary calculation shows that, irrespective of $\boldsymbol{\tau}_2$,

$$\boldsymbol{\Delta}_1^{\perp} \sim \mathcal{N}\left(\boldsymbol{\Gamma}_{11}^{\perp}\boldsymbol{\tau}_1,\ \boldsymbol{\Gamma}_{11}^{\perp}\right) \quad \text{with} \quad \boldsymbol{\Gamma}_{11}^{\perp} := \boldsymbol{\Gamma}_{11} - \boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_{22}^{-1}\boldsymbol{\Gamma}'_{12}$$

  (another Gaussian shift, with information $\boldsymbol{\Gamma}_{11}^{\perp} \leq \boldsymbol{\Gamma}_{11}$—equality iff $\boldsymbol{\Gamma}_{12} = \mathbf{0}$)

- the most stringent test then rejects $\mathcal{H}: \boldsymbol{\tau}_1 = \mathbf{0}$ whenever $\boldsymbol{\Delta}_1^{\perp\prime}\left(\boldsymbol{\Gamma}_{11}^{\perp}\right)^{-1}\boldsymbol{\Delta}_1^{\perp}$ exceeds $\chi^2_{K_1;1-\alpha}$
- accordingly, the locally (at $(\boldsymbol{\theta}'_{1;0}, \boldsymbol{\theta}'_2)'$) asymptotically most stringent test for $\mathcal{H}^{(n)}: \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{1;0}$ against $\mathcal{K}^{(n)}: \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_{1;0}$ consists in rejecting $\mathcal{H}^{(n)}$ whenever $\boldsymbol{\Delta}^{(n)\perp\prime}_{(\boldsymbol{\theta}'_{1;0},\boldsymbol{\theta}'_2)';1}\left(\boldsymbol{\Gamma}^{\perp}_{(\boldsymbol{\theta}'_{1;0},\boldsymbol{\theta}'_2)';11}\right)^{-1}\boldsymbol{\Delta}^{(n)\perp}_{(\boldsymbol{\theta}'_{1;0},\boldsymbol{\theta}'_2)';1}$ exceeds $\chi^2_{K_1;1-\alpha}$
- finally, replacing $\boldsymbol{\theta}_2$ with a $\boldsymbol{\nu}(n)$-consistent estimate $\hat{\boldsymbol{\theta}}_2^{(n)}$ (i.e., such that $\boldsymbol{\nu}^{-1}(n)(\hat{\boldsymbol{\theta}}_2^{(n)} - \boldsymbol{\theta}_2) = O_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}_{(\boldsymbol{\theta}'_{1;0},\boldsymbol{\theta}'_2)'}$ as $n \to \infty$) does not affect the asymptotic distribution of $\boldsymbol{\Delta}_1^{\perp}$. It follows that the locally (at any $(\boldsymbol{\theta}'_{1;0}, \boldsymbol{\theta}'_2)'$, $\boldsymbol{\theta}_2 \in \mathbb{R}^{K_2}$) asymptotically most stringent test for $\mathcal{H}^{(n)}: \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{1;0}$ against $\mathcal{K}^{(n)}: \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_{1;0}$ consists in rejecting $\mathcal{H}^{(n)}$ whenever $\boldsymbol{\Delta}^{(n)\perp\prime}_{(\boldsymbol{\theta}'_{1;0},\hat{\boldsymbol{\theta}}_2^{(n)\prime})';1}\left(\boldsymbol{\Gamma}^{\perp}_{(\boldsymbol{\theta}'_{1;0},\hat{\boldsymbol{\theta}}_2^{(n)\prime})';11}\right)^{-1}\boldsymbol{\Delta}^{(n)\perp}_{(\boldsymbol{\theta}'_{1;0},\hat{\boldsymbol{\theta}}_2^{(n)\prime})';1}$ exceeds $\chi^2_{K_1;1-\alpha}$

Outline

2. An introduction to Le Cam's
   asymptotic theory of statistical experiments

- 1. Local Asymptotic Normality (LAN)
- 2. Contiguity. Le Cam's first Lemma
- 3. Le Cam's third Lemma
- 4. Convergence of statistical experiments
- 5. Locally asymptotically optimal tests
- 6. Locally asymptotically optimal estimators

estimation in the Gaussian shift model

- optimal estimation of $\boldsymbol{\tau}$ in the Gaussian shift model

$$\boldsymbol{\Delta} \sim \mathcal{N}(\boldsymbol{\Gamma}\tau, \boldsymbol{\Gamma}) \quad \boldsymbol{\tau} \in \mathbb{R}^K$$

  is achieved through

$$\hat{\boldsymbol{\tau}} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Gamma}^{-1})$$

- in principle, thus, a locally asymptotically optimal estimator $\hat{\boldsymbol{\theta}}^{(n)}$ of $\boldsymbol{\theta}$ should achieve the same performance, asymptotically, under $\mathrm{P}_{\theta}^{(n)}$, as $n \to \infty$

one-step estimators

Let $\boldsymbol{\theta}_*^{(n)}$ be a sequence of $\boldsymbol{\nu}(n)$-consistent (and $\boldsymbol{\nu}(n)$discrete) estimators of $\boldsymbol{\theta}$, i.e., $\boldsymbol{\nu}^{-1}(n)(\boldsymbol{\theta}_*^{(n)} - \boldsymbol{\theta}) = O_{\mathrm{P}_{\boldsymbol{\theta}}^{(n)}}(1)$

define (the one-step estimator)

$$\hat{\boldsymbol{\theta}}^{(n)} := \boldsymbol{\theta}_*^{(n)} + \boldsymbol{\nu}(n)\boldsymbol{\Gamma}_{\boldsymbol{\theta}_*^{(n)}}^{-1}\boldsymbol{\Delta}_{\boldsymbol{\theta}_*^{(n)}}^{(n)}$$

- then,

$$\begin{aligned}\boldsymbol{\nu}^{-1}(n)(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}) &= \boldsymbol{\nu}^{-1}(n)(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}_*^{(n)}) + \boldsymbol{\nu}^{-1}(n)(\boldsymbol{\theta}_*^{(n)} - \boldsymbol{\theta}) \\ &= \boldsymbol{\Gamma}_{\boldsymbol{\theta}_*^{(n)}}^{-1}\boldsymbol{\Delta}_{\boldsymbol{\theta}_*^{(n)}}^{(n)} + \boldsymbol{\nu}^{-1}(n)(\boldsymbol{\theta}_*^{(n)} - \boldsymbol{\theta})\end{aligned}$$

But (ULAN) $\boldsymbol{\Delta}_{\boldsymbol{\theta}_*^{(n)}}^{(n)} = \boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\nu}^{-1}(n)(\boldsymbol{\theta}_*^{(n)} - \boldsymbol{\theta}) + o_{\mathrm{P}_{\boldsymbol{\theta}}^{(n)}}(1)$; hence,

$$\begin{aligned}\boldsymbol{\nu}^{-1}(n)(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}) &= \boldsymbol{\Gamma}_{\boldsymbol{\theta}_*^{(n)}}^{-1}\left[\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\nu}^{-1}(n)(\boldsymbol{\theta}_*^{(n)} - \boldsymbol{\theta})\right] + \boldsymbol{\nu}^{-1}(n)(\boldsymbol{\theta}_*^{(n)} - \boldsymbol{\theta}) + o_{\mathrm{P}}(1) \\ &= \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)} + o_{\mathrm{P}}(1) \approx \mathcal{N}(\boldsymbol{\tau},\, \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1}) \quad \text{under } \mathrm{P}_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\boldsymbol{\tau}}^{(n)}\end{aligned}$$

Hájek convolution Theorem

the one-step estimator thus asymptotically achieves the optimality result expected from the convergence to the Gaussian shift

This optimality property, which could be explicitated in terms of convergence of risk functions, is substantiated by the elegant Hájek convolution Theorem

Theorem. (Hájek 1970) *Let* $\mathbf{T}^{(n)}$ *be a* $\mathbb{R}^K$*-valued statistic (in a "local" sense, i.e., it may depend on* $\boldsymbol{\theta}$*, but not on* $\boldsymbol{\tau}$*) such that, under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\boldsymbol{\tau}}$*, the distribution of* $\mathbf{T}^{(n)} - \boldsymbol{\tau}$ *converges to some distribution* H *that does not depend on* $\boldsymbol{\tau}$*. Then,* H *is the distribution of* $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1/2}\mathbf{Z} + \mathbf{U}$*, where* $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$*, and* $\mathbf{Z}$ *and* $\mathbf{U}$ *are mutually independent.*

optimality of one-step

The one-step estimator $\hat{\boldsymbol{\theta}}^{(n)}$ just constructed

- satisfies the condition; more precisely, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\boldsymbol{\tau}}$, $\mathbf{T}^{(n)} := \boldsymbol{\nu}^{-1}(n)(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}) - \boldsymbol{\tau}$ is asymptotically $\mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1})$
- is (asymptotically) optimal in the sense that its asymptotic distribution corresponds to the "most favorable convolution" of a $\mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^{-1})$ distribution with a Dirac at $\mathbf{0}$ (viz., $\mathbf{U} = \mathbf{0}$ a.s.)