Point process models and local asymptotics in statistics

II – Local asymptotic normality – comparison of estimators

Reinhard Höpfner, University of Mainz **NOMP II, March 22–24, 2021**

We explain local asymptotic normality (LAN) in the sense of Le Cam, and give the main theorems on comparison of estimators under LAN. I will not give the proofs. All results below admit an extension to the –important and more general– setting of local asymptotic mixed normality (LAMN), but I will give only some very few hints in this direction. Key references are

- Le Cam, L.: Théorie asymptotique de la décision statistique. Montreal 1969.
- Hájek, J.: A characterization of limiting distributions for regular estimators. Zeitschrift f. Wahrscheinlichkeitstheorie u. verw. Geb. **14**, 232–330 (1970)
- Le Cam, L.: Limits of experiments. Proc. 6th Berkeley symposium on mathematical statistics and probbability. Vol. I, 245–261. Univ. Calif. Press 1972.
- Davies, R.: Asymptotic inference when the amount of information is random. In: LeCam, L., Olshen, R., Eds: Proc. of the Berkeley Symposium in in honor of J. Neyman and J. Kiefer, Vol. II: Wadsworth, 1985.
- Le Cam, L., Yang, G.: Asymptotics in statistics: some basic concepts. Springer 1990.

See also the books

- Hájek, J., Sidak: Theory of rank tests. Academic Press 1967.
- Strasser, H.: Mathematical theory of statistics. deGruyter 1985.
- Pfanzagl, J.: Parametric statistical theory. deGruyter 1994
- Liese, F., Miescke, K.: Statistical decision theory. Springer 2008.
- Höpfner, R.: Asymptotic statistics with a view to stochastic processes. deGruyter 2014.

When I give precise links (to proofs, background results, or LAMN) I will refer to the last title.

Quoting the work of Le Cam and of Hájek up to $\approx$ 1975, the russian school took a different road to a local asymptotic minimax theorem:

- Ibragimov, I., Khasminskii, R.: Statistical estimation. Springer 1981.
- Kutoyants, Y.: Statistical inference in ergodic diffusion processes. Springer 2004.

6 L^2-differentiable statistical models and iid experiments

This section deals with iid models. Consider dominated experiments

$$\mathcal{E} \;:=\; (\,\Omega\,,\,\mathcal{A}\,,\,\mathcal{P}:=\{P_\vartheta:\vartheta\in\Theta\}\,)\quad,\quad \mathcal{P}\ll\nu\,,\;\; f_\vartheta:=\frac{dP_\vartheta}{d\nu}\,,\;\; \Theta\subset\mathbb{R}^d \text{ open}$$

and n-fold product experiments

$$\mathcal{E}_n \;:=\; \left(\,\Omega_n:=\bigtimes_{j=1}^n\Omega\,,\;\mathcal{A}_n:=\bigotimes_{j=1}^n\mathcal{A}\,,\;\mathcal{P}_n:=\left\{P_\vartheta^n:=\bigotimes_{j=1}^n P_\vartheta:\vartheta\in\Theta\right\}\right)$$

and let $X=(X_1,\ldots,X_n)$ denote the canonical statistics on $(\Omega_n,\mathcal{A}_n)$.

Definition : $\mathcal{E}$ is called $\underline{L^2\text{-differentiable at } \vartheta \text{ with derivative } V_\vartheta}$ if there is some

$$\mathbb{R}^d\text{-valued random variable } V_\vartheta \text{ with components } V_{\vartheta,1},\ldots,V_{\vartheta,d} \text{ in } L^2(\Omega,\mathcal{A},P_\vartheta)$$

such that

$$\frac{1}{|\xi-\vartheta|^2}\int_\Omega\left|f_\xi^{1/2}-f_\vartheta^{1/2}-\frac{1}{2}f_\vartheta^{1/2}\,(\xi-\vartheta)^\top V_\vartheta\right|^2 d\nu \quad\longrightarrow\quad 0 \quad \text{as } \xi\to\vartheta\;.$$

We call

$$J(\vartheta)\;:=\;E_\vartheta\left(V_\vartheta^\top V_\vartheta\right)$$

$\underline{\text{Fisher-Information in } \vartheta}$ (when L^2-differentiability holds, V_ϑ does not depend on the choice of the dominating measure, and is centred: $E_\vartheta(V_\vartheta)=0$ in all components).

Example: i) Consider a location model $\mathcal{E}$ on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ with Lebesgue densities

$$f_\vartheta(x)\;=\;\frac{1}{2}e^{-|x-\vartheta|}\quad,\quad \vartheta\in\Theta:=\mathbb{R}\quad,\quad x\in\mathbb{R}\,.$$

$\mathcal{E}$ is L^2-differentiable at every ϑ with derivative $V_\vartheta(x):=\operatorname{sign}(x-\vartheta)\,.$
The Fisher information is $J(\vartheta)=E_\vartheta(V_\vartheta^2)=1$ for all $\vartheta\in\Theta$.

ii) In smoothly parametrized models we can prove L^2-differentiability under mild assumptions, with derivative V_ϑ of type $\frac{\partial}{\partial\vartheta}\log f_\vartheta(\cdot)\,.$

'2nd Le Cam lemma' : Consider a point ϑ where $\mathcal{E}$ is L^2-differentiable with derivative V_ϑ. Then for bounded sequences (h_n) in $\mathbb{R}^d$ (we always assume that $\vartheta+n^{-1/2}h_n$ belongs to Θ), log-likelihood ratios in $\mathcal{E}_n$ admit at ϑ quadratic expansions

$$\log L_n^{(\vartheta+n^{-1/2}h_n)\,/\,\vartheta} \quad = \quad h_n^\top S_n(\vartheta) \;-\; \frac{1}{2} h_n^\top J(\vartheta)\, h_n \;+\; o_{P_\vartheta^n}(1) \quad , \quad n \to \infty$$

with

$$S_n(\vartheta) \;:=\; n^{-1/2} \sum_{j=1}^{n} V_\vartheta(X_j)$$

(called score at ϑ in $\mathcal{E}_n$) converging in law:

$$\mathcal{L}\,(\,S_n(\vartheta) \mid P_\vartheta^n\,) \quad \stackrel{w}{\longrightarrow} \quad \mathcal{N}\,(\,0\,,\,J(\vartheta)\,) \qquad \text{(weak convergence in } \mathbb{R}^d \text{ as } n \to \infty) \;\; .$$

Remark: Compare this to normal distributions $\mathcal{P} := \left\{\, P_h := \mathcal{N}(Jh, J) \,:\, h \in \mathbb{R}^d \,\right\}$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $J \in \mathbb{R}^{d\times d}$ symmetric and strictly positive definite, where we have

$$\frac{dP_h}{dP_0}(x) \;=\; \exp\left\{ h^\top S(x) - \frac{1}{2} h^\top J\, h \right\} \quad , \quad S(x) := x$$

with $\mathcal{L}(S \mid P_0) = \mathcal{N}(0, J)$.

Proofs in many books, e.g. sections 4.1 and 4.2 in H. 2014.

Results of type '2nd Le Cam Lemma' can be proved in broad classes of stochastic process models under ergodicity assumptions: then martingales and their angle brackets appear in place of score and Fisher information, and one applies martingale convergence theorems (e.g. Jacod-Shiryaev 1987, VIII.3.22).

- Löcherbach, E.: LAN and LAMN for systems of interacting diffusions with branching and immigration. Ann. H. Poincaré Proba. Stat. **38**, 59–90 (2002).
- Holbach, S.: Local asymptotic normality for shape and periodicity of a signal in the drift of a degenerate diffusion with internal variables. Electronic J. Probability **13**, 4884–4915 (2019).

In a variety of other stochastic process models, results of type '2nd Le Cam Lemma' exist, with information processes (to which limit theorems apply) in place of deterministic information.

- Höpfner, R., Kutoyants, Y.: On a problem of statistical inference in null recurrent diffusions. Statist. Inference Stoch. Processes **6**, 25–42 (2003).

7 The Gaussian shift model

Definition: A model $(\Omega, \mathcal{A}, \{P_h : h \in \mathbb{R}^d\})$ is called Gaussian shift experiment $\mathcal{E}(J)$ if there exists a statistic S, $\mathbb{R}^d$-valued, and a deterministic matrix J, symmetric and strictly positive definite, such that for every $h \in \mathbb{R}^d$,

$$\omega \longrightarrow \exp\left(h^\top S(\omega) - \frac{1}{2} h^\top J\, h\right) =: L^{h/0}(\omega)$$

is a version of the likelihood ratio of P_h with respect to P_0. Call $Z := J^{-1}S$ central statistic in $\mathcal{E}(J)$.

Remark: In a Gaussian shift experiment $\mathcal{E}(J)$, all probability measures are equivalent, and we have

$$L^{(h_0+h)/h_0} = L^{(h_0+h)/0} / L^{h_0/0} = \exp\left(h^\top (S - Jh_0) - \frac{1}{2} h^\top J\, h\right) \qquad \text{for all } h,\ h_0 \text{ in } \mathbb{R}^d\ .$$

Taking expectation $E_{h_0}(\ldots)$ on both sides we identify (Laplace transform!)

$$\mathcal{L}\,(S - Jh_0 \mid P_{h_0}) = \mathcal{N}(0, J)$$

and thus obtain for the central statistic $Z := J^{-1}S$ and arbitrary $h_0 \in \mathbb{R}^d$

$$\mathcal{L}\,(Z - h_0 \mid P_{h_0}) = \mathcal{N}(0, J^{-1}) \text{ does not depend on } h_0 \in \mathbb{R}^d.$$

We all estimators κ for the unknown parameter equivariant if laws of estimation errors $\mathcal{L}\,(\kappa - h \mid P_h)$ do not depend on the value of the parameter h. In particular, in a Gaussian shift model $\mathcal{E}(J)$, the central statistic Z is an equivariant estimator for the unknown parameter.

Convolution theorem (Boll 1955): In a Gaussian shift experiment $\mathcal{E}(J)$, for every equivariant estimator κ for the unknown parameter $h \in \mathbb{R}^d$, there is a probability law Q on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that

$$\mathcal{L}\,(\kappa - h \mid P_h) = \mathcal{N}(0, J^{-1}) \star Q \quad \text{for all} \quad h \in \mathbb{R}^d\ .$$

Convolutions $\mathcal{N}(0, J^{-1}) \star Q$ are always 'more spread out' than the law $\mathcal{N}(0, J^{-1})$ itself. Hence in $\mathcal{E}(J)$, the central statistic Z is a best concentrated equivariant estimator for the unknown parameter.

However, we would like to be able to compare quite arbitrary estimators for the unknown parameter. For this, the main technical (and somewhat difficult) step:

Lemma: In a Gaussian shift experiments $\mathcal{E}(J)$, for arbitrary estimators η for the unknown parameter $h \in \mathbb{R}^d$, there is a sequence $(Q_n)_n$ of probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that total variation distance between the laws

$$\int_{\{|h| \leq n\}} \mathcal{L}\,(\,\eta - h \mid P_h\,)\; dh \qquad \text{and} \qquad \mathcal{N}(0, J^{-1}) \star Q_n$$

vanishes as $n \to \infty$.

Definition: We call a loss function $\ell : \mathbb{R}^d \to [0, \infty)$ subconvex or bowl-shaped if level sets

$$\Big\{ x \in \mathbb{R}^d : \ell(x) \leq c \Big\} \quad , \quad c \geq 0$$

are convex and symmetric with respect to the origin.

With respect to a loss function ℓ, the risk of an estimator η for the unknown parameter h is the function

$$h \quad \longrightarrow \quad R_\ell(\eta, h) \;:=\; E_h\,(\,\ell(\eta - h)\,)$$

which may take the value ∞.

Minimax theorem: In a Gaussian shift experiment $\mathcal{E}(J)$, the central statistic Z minimizes the maximal risk with respect to any subconvex loss function $\ell(\cdot)$, i.e.: arbitrary estimators η for the unknown parameter $h \in \mathbb{R}^d$ can be compared to Z via

$$\sup_{h \in \mathbb{R}^d} R_\ell(\eta, h) \quad \geq \quad \int_{\mathbb{R}^d} \ell(z)\, \mathcal{N}(0, J^{-1})(dz) \;=\; R_\ell(Z, 0) \;=\; \sup_{h \in \mathbb{R}^d} R_\ell(Z, h)\;.$$

References, proofs, further reading: H. 2014 Section 5.1; mixed normal experiments: Section 6.1.

8 Local asymptotic normality (LAN)

Consider a sequence of statistical experiments parametrized by the same parameter set $\Theta \subset \mathbb{R}^d$ open:

$$\mathcal{E}_n \;:=\; (\,\Omega_n\,,\,\mathcal{A}_n\,,\,\{P_{n,\vartheta} : \vartheta \in \Theta\}\,) \quad , \quad n \in \mathbb{N}\,.$$

Definition: For $\vartheta \in \Theta$ fixed, the sequence of experiments $(\mathcal{E}_n)_n$ is called locally asymptotically normal (LAN) at ϑ if there is a sequence of positive real numbers (or matrices in $\mathbb{R}^{d\times d}$)

$$\delta_n \;=\; \delta_n(\vartheta) \quad \text{decreasing to 0 as } n \to \infty$$

called local scale at ϑ, and a matrix in $\mathbb{R}^{d\times d}$

$$J(\vartheta)\ \text{, deterministic, symmetric and strictly positive definite}$$

called limit information at ϑ, such that the following holds: for all bounded sequences $(h_n)_n$ in $\mathbb{R}^d$ (we always assume that $\vartheta + \delta_n(\vartheta)h_n$ belongs to Θ), log-likelihood ratios in local models at ϑ

$$\mathcal{E}_{n,\vartheta} := \left\{\, P_{n\,,\,\vartheta+\delta_n(\vartheta)h} \;:\; h \text{ in } \mathbb{R}^d \text{ such that } \vartheta + \delta_n(\vartheta)h \,\in\, \Theta \,\right\} \quad , \quad n \to \infty$$

admit quadratic expansions

$$\log L_n^{(\vartheta+\delta(\vartheta)h_n)/\vartheta} \;=\; h_n^\top S_n(\vartheta) \;-\; \frac{1}{2} h_n^\top J_n(\vartheta)\, h_n \;+\; o_{P_{n,\vartheta}}(1)$$

as $n \to \infty$ where

$$\mathcal{L}\,(\,S_n(\vartheta)\,,\,J_n(\vartheta)\mid P_{n,\vartheta}\,) \quad \xrightarrow{\;w\;} \quad \mathcal{N}\,(\,0\,,\,J(\vartheta)\,) \otimes \epsilon_{J(\vartheta)} \quad \text{(weakly in } \mathbb{R}^d\times\mathbb{R}^{d\times d}\,)\,.$$

Remarks: 1) Equivalently, we can write the last assertion as

$$\mathcal{L}\,(\,S_n(\vartheta)\mid P_{n,\vartheta}\,) \;\xrightarrow{\;w\;}\; \mathcal{N}\,(\,0\,,\,J(\vartheta)\,) \quad \text{together with} \quad J_n(\vartheta) \;=\; J(\vartheta) \;+\; o_{P_{n,\vartheta}}(1)$$

as $n \to \infty$ (weak convergence to a deterministic object is convergence in probability).

2) Local models at ϑ are parametrized by $h \in ...\mathbb{R}^d...$: thus, when LAN holds at ϑ, the Gaussian shift experiment $\mathcal{E}(J(\vartheta))$ appears as limit model for $\mathcal{E}_{n,\vartheta}$ as $n \to \infty$.

Definition: Assuming LAN at ϑ, a sequence of estimators for the unknown parameter $\vartheta \in \Theta$

$$T_n \text{ on } (\Omega_n, \mathcal{A}_n)\,,\ n \geq 1$$

is called regular at ϑ if there is some probability law $F = F(\vartheta)$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that

$$\mathcal{L}\left(\,\delta_n^{-1}(\vartheta)\,(T_n - (\vartheta + \delta_n(\vartheta)h)) \mid P_{n\,,\,\vartheta+\delta_n(\vartheta)h}\,\right) \qquad \xrightarrow{w} \qquad F$$

(weak convergence in $\mathbb{R}^d$, as $n \to \infty$) for every $h \in \mathbb{R}^d$.

Thus regular means 'asymptotically equivariant with respect to the local parameter' at ϑ.
Regularity can be checked by proving joint weak convergence of pairs 'rescaled estimation errors, log-likelihood ratios' under $P_{n,\vartheta}$ as $n \to \infty$ to limit laws of suitable structure.
A key tool is 'Le Cam's 3rd lemma', e.g. H. 2014 section 3.1.

Hájek's convolution theorem: Assume LAN at ϑ, and consider any sequence of estimators $(T_n)_n$ for the unknown parameter $\vartheta \in \Theta$ which is regular at ϑ.
a) Any limit distribution F arising in the definition above can be written as

$$F \;=\; \mathcal{N}\left(0\,,\, J^{-1}(\vartheta)\right) \star Q$$

for some probability law Q on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.
b) A sequence of estimators $(T_n)_n$ is regular and efficient at ϑ if and only if

$$\delta_n^{-1}(\vartheta)\,(T_n - \vartheta) \;\;=\;\; Z_n(\vartheta) \;+\; o_{P_{n,\vartheta}}(1) \quad \text{as } n \to \infty$$

i.e. rescaled estimation errors under ϑ are coupled to the central sequence $Z_n(\vartheta) = J_n^{-1}(\vartheta) S_n(\vartheta)$ at ϑ.

Local asymptotic minimax theorem: Assume LAN at ϑ, and consider any sequence of estimators $(T_n)_n$ for the unknown parameter $\vartheta \in \Theta$ whose rescaled estimation errors at ϑ are tight:

$$\mathcal{L}\left(\delta_n^{-1}(\vartheta)\,(T_n - \vartheta) \mid P_{n,\vartheta}\right)\ ,\ n \geq 1\ , \quad \text{is tight in } \mathbb{R}^d \text{ as } n \to \infty\ .$$

a) For every loss function $\ell(\cdot)$ which is continuous, subconvex and bounded:

$$\lim_{c\uparrow\infty} \limsup_{n\to\infty} \sup_{|h|\leq c} E_{\vartheta+\delta_n(\vartheta)h}\left(\ell\left(\delta_n^{-1}(\vartheta)\left(\widetilde{\vartheta}_n - (\vartheta + \delta_n(\vartheta)h)\right)\right)\right) \quad \geq \quad \int_{\mathbb{R}^d} \ell(z)\,\mathcal{N}(0, J^{-1}(\vartheta))(dz)$$

where $J(\vartheta)$ is the limiting information.

b) Sequences $(T_n)_n$ with the property

$$\delta_n^{-1}(\vartheta)\,(T_n - \vartheta) \ = \ Z_n(\vartheta) \ + \ o_{P_{n,\vartheta}}(1) \quad \text{as } n \to \infty$$

achieve the local asymptotic minimax bound in a): for every $0 < c < \infty$,

$$\lim_{n\to\infty} \sup_{|h|\leq c} E_{\vartheta+\delta_n(\vartheta)h}\left(\ell\left(\delta_n^{-1}(\vartheta)\left(\widetilde{\vartheta}_n - (\vartheta + \delta_n(\vartheta)h)\right)\right)\right) \quad = \quad \int_{\mathbb{R}^d} \ell(z)\,\mathcal{N}(0, J^{-1}(\vartheta))(dz)\ .$$

Remark: Starting from any sequence $(\widetilde{T}_n)_n$ of preliminary estimators for the unknown parameter $\vartheta \in \Theta$ such that rescaled estimation errors at ϑ are tight, it is possible (under additional assumptions which often do hold) to construct explicitely a new sequence $(T_n)_n$ which is efficient in the sense of (the convolution theorem and) the local asymptotic minimax theorem: this is Le Cam's 'one-step correction'.

References, proofs, further reading, extension to LAMN: H. 2014 Sections 7.1, 7.2, 7.3.