Les deux complices

Piet Groeneboom, Delft University

Juin 28, 2018

Monsieur Teste (Paul Valéry)

Je me suis rarement perdu de vue

Je me suis détesté
je me suis adoré

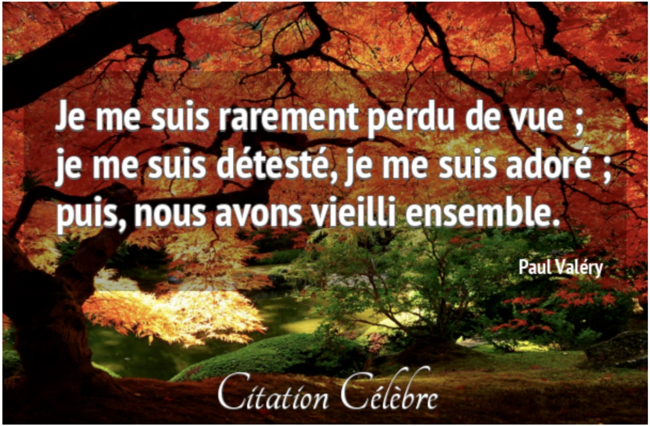
Puis, nous avons vieilli ensemble

(La soirée avec Monsieur Teste)

Internet

Images pour vos réseaux sociaux

Belle citation avec photo de paysage



Monsieur Teste (Paul Valéry)

La recherche du succès entrain nécessairement une perte de temps.

En échange du pourboire public, il donne le temps qu'il faut pour se rendre perceptible.

Lucien, Montceau-les-Mines

Lucien, école normale supérieure

La mémoire

La mémoire est l'avenir du passé (Paul Valéry: Cahiers).

My lecture at the école normale supérieure

I recently tested my memories of the day on Cécile and Lucien.

Din(n)er the same evening with Lucien at the house of Jean Bretagnolle.

Lucien sportif

La mémoire

European meeting of statisticians, Brighton, 1980.

Bien étonnés de se trouver ensemble.
Gábor Tusnády organizer invited session.

Common ground? Kullback-Leibler "distance"? Efficiencies?
Fano's lemma versus Assouad's lemma? Non-local versus local?
Minimax?

Bahadur efficiency of weighted Kolmogorov-Smirnov test (dividing by $\sqrt{\mathbb{F}_n(1-\mathbb{F}_n)}$) is zero!
Bahadur, Abramson, Galen Shorack (1980)

Lucien talked on matters related to his dissertation too (which I read afterwards). I later talked about it in Amsterdam, with Aad van der Vaart, Sara van de Geer, and others in the audience. They liked it!

Lucien on dimension (metric entropy)

The “old days”:

- Parameters belong to a finite-dimensional space

Present situation:

- Parameters belong to an infinite-dimensional space

- L. Birgé (1979). Dimension métrique et vitesse d’estimation. C.R.A.S. Sér. A 289, 291-293.
- L. Birgé (1980). Thése, 3^e partie.
- L. Birgé (1983). Approximation dans les espaces métriques et théorie de l’estimation. Z. Warsch. Verw. Gebiete 65, 181-237

Lucien on relation between estimation and testing

- L. Birgé (1979). Un estimateur construit à partir de tests. C.R.A.S. Sér. A 289, 361-363.
- L. Birgé (1983). Vitesses maximales de décroissance des erreurs et tests optimaux associés. Z. Wahrsch. verw. Gebiete 65, 181-237.
- L. Birgé (1984). Sur un théoréme de minimax et son application aux tests. Probab. Math. Statist. 3, 259-282
- L. Birgé (1986). On estimating a density using Hellinger distance and some other strange facts. Probab. Th. Rel. Fields 71, 271-291.
- L. Birgé (2005). Model selection via testing: an alternative to (penalized) maximum likelihood estimators. Annales de l' IHP.

Second encounter

Berkeley 1983, first year of MSRI (1982-1983).

Le Cam's Thursday seminar. Let us thank ...

Lucien: on estimating a density using Hellinger distance and some other strange facts.

Me: paper with Ronald Pyke on the L_2 distance of the Grenander estimator to 1 in a sample from the uniform distribution.

$$\frac{1}{\sqrt{3\log n}}\left\{n\left\|\hat{f}_n - f_0\right\|_2^2 - \log n\right\} \xrightarrow{\mathcal{D}} N(0,1). \tag{1}$$

Second encounter

Or:

$$\frac{2}{\sqrt{3}}\left\{n^{1/2}\left\|\hat{f}_n - f_0\right\|_2 - \sqrt{\log n}\right\} \xrightarrow{\mathcal{D}} N(0,1).$$

For a *strictly monotone* density f_0 the following result was stated in Groeneboom (1985) for the L_1-distance, with a sketch of proof:

$$n^{1/6}\left\{n^{1/3}\left\|\hat{f}_n - f_0\right\|_1 - \mu\right\} = n^{1/2}\left\|\hat{f}_n - f_0\right\|_1 - n^{1/6}\mu \xrightarrow{\mathcal{D}} N\left(0,\sigma^2\right),$$

for constants μ and σ, depending on the inverse process

$$V(c) = \sup\{t : W(t) - (t-c)^2 \text{ is maximal}\}.$$

In Cécile Durot's dissertation under supervision of Pascal Massart V is called the "Groeneboom process".

Groeneboom process

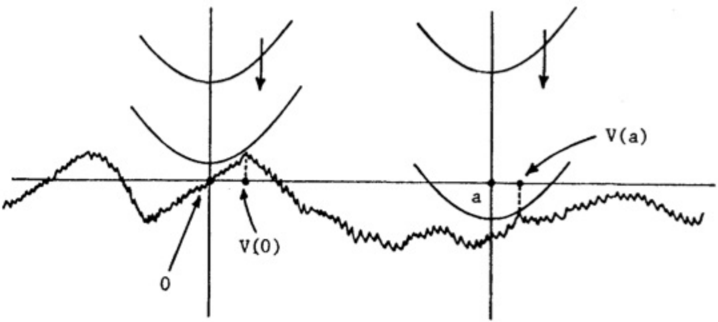


Fig. 4.1

La mémoire

Piet to Pascal: Cécile and you call this the Groeneboom process?

Pascal to Piet: do you object to us calling this the Groeneboom process?

Piet: no, not particularly.

Monsieur Teste: it is a block in the stream, of course. Better to stay anonymous than to be the man of the "Groeneboom process".

Second encounter, Grenander and minimax

$$n^{1/6}\left\{n^{1/3}\left\|\hat{f}_n - f_0\right\|_1 - \mu\right\} = n^{1/2}\left\|\hat{f}_n - f_0\right\|_1 - n^{1/6}\mu \xrightarrow{\mathcal{D}} N\left(0, \sigma^2\right).$$

What are μ and σ?

$$\mu = 2E|V(0)| \int_0^1 \left|\frac{1}{2} f'(t) f(t)\right|^{1/3} dt \tag{2}$$

$$\sigma^2 = 8 \int_0^\infty \operatorname{covar}\left(|V(0)|, |V(c) - c|\right) dc. \tag{3}$$

Rigorous proof is given in Groeneboom, Hooghiemstra, and Lopuhaä (1999). The result is also discussed in Groeneboom and Jongbloed (2014).

Lucien: (2) is the real expression for the L_1 minimax risk, in contrast with the expression for it in Bretagnolle and Huber.

Monotone density estimation and isotonic regression

The log likelihood of a specific density f is given by

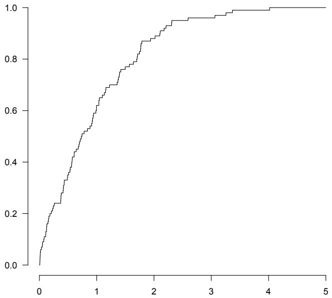
$$\ell(f) = \frac{1}{n}\sum_{i=1}^{n} \log f(X_i) = \int \log f(x)\, d\mathbb{F}_n(x).$$

The Grenander maximum likelihood estimator maximizes this function over all decreasing densities on $[0, \infty)$.

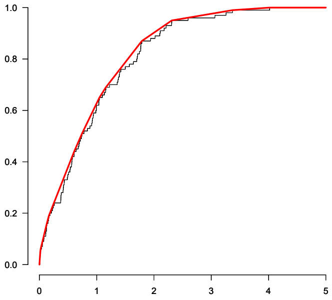
Theorem (Grenander (1956))

The maximum likelihood estimator is the left derivative of the least concave majorant $\hat{F}_n$ of the empirical distribution function $\mathbb{F}_n$.

The Grenander estimator



The Grenander estimator



The Grenander estimator

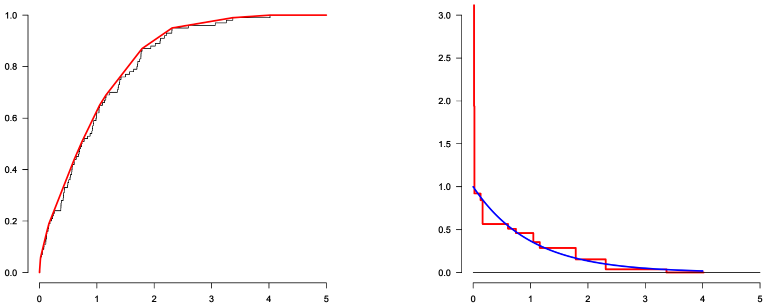


Figure: The least concave majorant $\hat{F}_n$ and its derivative $\hat{f}_n$ (the Grenander estimator) for a sample of size $n = 100$ from a standard exponential.

Ulf Grenander

The Grenander estimator is also least squares estimator

The Grenander estimator $\hat{f}_n$ maximizes

$$\ell(f) = \frac{1}{n}\sum_{i=1}^{n} \log f(X_i) = \int_0^\infty \log f(x)\, d\mathbb{F}_n(x).$$

but minimizes

$$\int_0^\infty f(t)^2\, dt - 2\int_0^\infty f(t)\, d\mathbb{F}_n(t)$$

over all decreasing densities f on $[0, \infty)$.

Cator (2011): $\hat{f}_n$ is locally asymptotically minimax

Lucien's interest

Lucien was interested in the result

$$n^{1/2}\|\hat{f}_n - f_0\|_1 - n^{1/6}\mu \xrightarrow{\mathcal{D}} N(0,\sigma^2).$$

$$\mu = 2E|V(0)|\int_0^1 |\frac{1}{2}f'(t)f(t)|^{1/3}\,dt \tag{4}$$

$$\sigma^2 = 8\int_0^\infty \mathrm{covar}\left(|V(0)|, |V(c)-c|\right)dc.$$

in particular in the integral (4).

Would the Grenander estimator also be globally minimax, using the L_1-distance? Answer is still unknown.

Lucien's non-asymptotic point of view

- L. Birgé (1985). Non-asymptotic minimax risk for Hellinger balls. Probab. Math. Statistics. 5, 21-29.
- L. Birgé (1987). Estimating a density under order restrictions: non-asymptotic minimax risk. Ann. Statist. 15, 995-1012.
- L. Birgé (1989). The Grenander estimator: a non-asymptotic approach. Ann. Statist. 17, 1532-1549.
- L. Birgé (1999). Interval censoring: a non-asymptotic point of view. Math. Methods of Statist. 8. 285-298.
- L. Birgé and Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. Bernoulli 4, 329-375.

La mémoire

Letter David Donoho late eighties to my home address

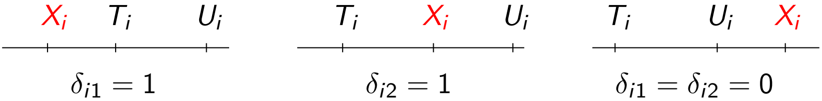
Third encounter

Berkeley, MSRI, 1991. Office assignment.

Lucien et Pascal: il se sont rarement perdus de vue.

Interval censoring, case 2

$X_1, X_2, \ldots, X_n \sim F_0$.
Instead of observing the X_i's, one only observes for each i whether $X_i \le T_i$ or $X_i \in (T_i, U_i]$ or $X_i > U_i$, for some random pair (T_i, U_i), where $T_i < U_i$, where (T_i, U_i) is independent of X_i.



So, instead of observing X_i's, one observes

$$(T_i, U_i, \delta_{i1}, \delta_{i2}) = (T_i, U_i, 1_{\{X_i \le T_i\}}, 1_{\{X_i \in (T_i, U_i]\}}).$$

Interval censoring, case 2

Asymptotic local distribution?

Conjecture (Groeneboom (1991)):
Let H be the distribution function of (T_i, U_i) and let F_0 and H be continuous differentiable at t_0 and (t_0, t_0), respectively, with strictly positive derivatives $f(t_0)$ and $h(t_0, t_0)$. Let $\hat{F}_n$ be the MLE of F_0. Then

$$(n \log n)^{1/3} \left\{ \hat{F}_n(t_0) - F_0(t_0) \right\} \left\{ 6 f_0(t_0)^2 / h(t_0, t_0) \right\}^{1/3} \xrightarrow{\mathcal{D}} Z,$$

where $Z = \text{argmax}_t \{ W(t) - t^2 \}$.

Still not proved!

Local rate

1. Piet (Groeneboom (1991), Groeneboom and Wellner (1992)): the conjecture is true for a "toy" estimator, obtained by doing one step of the iterative convex minorant algorithm, starting the iterations at the underlying distribution function F_0. Lucien (Birgé (1999)) has constructed a histogram-type estimator, achieving the local rate $(n \log n)^{1/3}$ in this model. Faster rate than in current status model!
2. If the times T_i and U_i are separated, that is:

$$\mathbb{P}\{U_i - T_i < \epsilon\} = 0,$$

   for some $\epsilon > 0$, the rate drops to $n^{1/3}$.
3. The asymptotic distribution of the MLE can in this case be proved to be the same as the distribution of the toy estimator (Groeneboom (1996), St-Flour). Limit distribution is again $Z = \text{argmax}\{W(t) - t^2\}$. Variance: Wellner (1995).

Efficient estimates, based on the MLE

Theorem (Groeneboom (1991),Groeneboom and Wellner (1992))

Let F_0 be differentiable on $[0, B]$ and let a number of regularity condtions be fulfilled. Then, both in the separated and in the non-separated case,

$$\sqrt{n}\left\{\int x\,d\hat{F}_n(x) - \int x\,dF_0(x)\right\} \xrightarrow{\mathcal{D}} N(0,\sigma^2),$$

where σ^2 can be found (numerically) by solving an integral equation.

$\int x\,d\hat{F}_n(x)$ is efficient and asymptotically minimax: Geskus and Groeneboom (1996, 1997, 1999).

Visits of Lucien and Pascal to my office at MSRI in 1991 (after the switch between Lucien and Nemirovski).

Famous open problem

Let, in the single index regression problem, $\hat{\boldsymbol{\alpha}}_n$ be the estimate of $\boldsymbol{\alpha}$, which minimizes

$$\frac{1}{n}\sum_{i=1}^{n}\left\{Y_i - \hat{\psi}_{n\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{X}_i)\right\}^2, \tag{5}$$

assuming that $\hat{\psi}_{n\boldsymbol{\alpha}}$ is the least squares estimate of ψ for fixed $\boldsymbol{\alpha}$ under the restriction that ψ is monotone increasing.
Is $\hat{\boldsymbol{\alpha}}_n$ a rate $\sqrt{n}$ estimate of $\boldsymbol{\alpha}$?

Balabdaoui, Durot, and Jankowski (2016) prove it has rate $n^{1/3}$.

It is proved in Balabdaoui, Groeneboom, and Hendrickx (2018) that the estimate $\hat{\boldsymbol{\alpha}}_n$, solving

$$\left(\boldsymbol{I} - \boldsymbol{\alpha}\boldsymbol{\alpha}^T\right)\sum_{i=1}^{n}\boldsymbol{X}_i\left\{Y_i - \hat{\psi}_{n\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{X}_i)\right\} = 0$$

has rate $\sqrt{n}$. No tuning parameters!

Bibliography I

Fadoua Balabdaoui, Cécile Durot, and Hanna Jankowski. Least squares estimation in the monotone single index model. *arXiv preprint arXiv:1610.06026*, 2016.

Fadoua Balabdaoui, Piet Groeneboom, and Kim Hendrickx. Score estimation in the monotone single index model. *arXiv preprint arXiv:1712.05593v3*, 2018.

L. Birgé. Interval censoring: a nonasymptotic point of view. *Math. Methods Statist.*, 8:285–298, 1999. ISSN 1066-5307.

Eric Cator. Adaptivity and optimality of the monotone least-squares estimator. *Bernoulli*, 17(2):714–735, 2011. ISSN 1350-7265. URL `http://dx.doi.org/10.3150/10-BEJ289`.

R. B. Geskus and P. Groeneboom. Asymptotically optimal estimation of smooth functionals for interval censoring. I. *Statist. Neerlandica*, 50:69–88, 1996. ISSN 0039-0402. URL `http://dx.doi.org/10.1111/j.1467-9574.1996.tb01481.x`.

Bibliography II

R. B. Geskus and P. Groeneboom. Asymptotically optimal estimation of smooth functionals for interval censoring. II. *Statist. Neerlandica*, 51:201–219, 1997. ISSN 0039-0402. URL `http://dx.doi.org/10.1111/1467-9574.00050`.

R.B. Geskus and P. Groeneboom. Asymptotically optimal estimation of smooth functionals for interval censoring, case 2. *Ann. Statist.*, 27:627–674, 1999. ISSN 0090-5364. URL `http://dx.doi.org/10.1214/aos/1018031211`.

U. Grenander. On the theory of mortality measurement. II. *Skand. Aktuarietidskr.*, 39:125–153 (1957), 1956.

P. Groeneboom. Estimating a monotone density. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*, Wadsworth Statist./Probab. Ser., pages 539–555, Belmont, CA, 1985. Wadsworth.

Bibliography III

P. Groeneboom. Nonparametric maximum likelihood estimators for interval censoring and deconvolution. Technical Report 378, Department of Statistics, Stanford University., 1991. URL `https://statistics.stanford.edu/resources/technical-reports`.

P. Groeneboom. Lectures on inverse problems. In *Lectures on probability theory and statistics (Saint-Flour, 1994)*, volume 1648 of *Lecture Notes in Math.*, pages 67–164. Springer, Berlin, 1996. URL `http://dx.doi.org/10.1007/BFb0095675`.

P. Groeneboom, G. Hooghiemstra, and H.P. Lopuhaä. Asymptotic normality of the L_1 error of the Grenander estimator. *Ann. Statist.*, 27:1316–1347, 1999. ISSN 0090-5364. URL `http://dx.doi.org/10.1214/aos/1017938928`.

P. Groeneboom and J.A. Wellner. *Information bounds and nonparametric maximum likelihood estimation*, volume 19 of *DMV Seminar*. Birkhäuser Verlag, Basel, 1992. ISBN 3-7643-2794-4.

Bibliography IV

Piet Groeneboom and Geurt Jongbloed. *Nonparametric Estimation under Shape Constraints*. Cambridge Univ. Press, Cambridge, 2014.

J.A. Wellner. Interval censoring, case 2: alternative hypotheses. In *Analysis of censored data (Pune, 1994/1995)*, volume 27 of *IMS Lecture Notes Monogr. Ser.*, pages 271–291. Inst. Math. Statist., Hayward, CA, 1995. URL `http://dx.doi.org/10.1214/lnms/1215452225`.