Statistical inference and Malliavin calculus*

José M. Corcuera
Universitat de Barcelona
Gran Via de les Corts Catalanes. 585
E-08007, Barcelona
Spain
Email: jmcorcuera@ub.edu
Arturo Kohatsu-Higa
Ritsumeikan University and Japan Science and Technology Agency
Department of Mathematical Sciences
1-1-1 Nojihigashi,
Kusatsu, Shiga, 525-8577, Japan
E-mail: arturokohatsu@gmail.com

May 7, 2013

Keywords: Diffusion processes, Malliavin calculus, parametric estimation, Cramer-Rao lower bound, LAN property, LAMN property, jump-diffusion processes, score function.

Abstract

The derivative of the log-likelihood function, known as score function, plays a central role in parametric statistical inference. It can be used to study the asymptotic behavior of likelihood and pseudo-likelihood estimators. For instance, one can deduce the local asymptotic normality property which leads to various asymptotic properties of these estimators. In this article we apply Malliavin Calculus to obtain the score function as a conditional expectation. We then show, through different examples, how this idea can be useful for asymptotic inference of stochastic processes. In particular, we consider situations where there are jumps driving the data process.

This is an abridged version of the article: Statistical Inference and Malliavin Calculus . Seminar on Stochastic Analysis, Random Fields and Applications VI, 2011

1 Introduction

In classical statistical theory, the Cramer-Rao lower bound is obtained by using two steps: an integration by parts and the Cauchy-Schwarz inequality. Therefore it seems natural that the integration by parts formulas of Malliavin calculus will play a role in this context. In recent times, the theory of Malliavin Calculus has attracted attention from Computational Finance to derive expressions for the calculation of the *greeks* which measure the sensitivity of option prices (conditional expectations) with respect to certain parameters, see e.g. [**?**],[**?**]. We show in this article, that similar techniques can be used to derive expressions for the score function in a parametric statistical model and consequently we obtain the Fisher information and the

*This research has been supported by grants of the Japanese government.

Cramer-Rao's bounds. The advantage of these expressions is that they do not require the explicit expression of the likelihood function directly and that its form is appropriate to study asymptotic properties of the model and some estimators. Gobet [?] was the first contribution in this direction. He uses the duality property in a *Gaussian space* to study the local asymptotic mixed normality of parametric diffusion models when the observations are discretely observed. Recently, duality properties have been used to obtain a Stein's type estimator in the context of a *Gaussian space* (see [?] and [?]), and in the context of a *Poisson process* (see [?]). An interesting discussion of different Cramer-Rao's bounds is given in [?].

In this paper we use Malliavin calculus with the aim of giving alternative expressions for the score function as conditional expectations of certain expressions involving Skorohod's integrals and we show how to use them to study the asymptotic properties of the statistical model, to derive Cramer-Rao lower bounds and expressions for the maximum likelihood estimator.

2 Background

A parametric statistical model is defined as the triplet $(\mathcal{X}, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$ where $\mathcal{X}$ is the sample space that corresponds to the possible values of certain n-dimensional random vector $X = (X_1, X_2, ..., X_n)$, $\mathcal{F}$ is the σ-field of observable events and $\{P_\theta, \theta \in \Theta\}$ is the family of possible probability laws of X. However, when the vector X corresponds to observations of a random process where certain parameter θ is involved, it is better to assume that X itself depends explicitly θ.

Then we define *a parametric statistical model* as a triplet consisting of a probability space $(\Omega, \mathcal{F}, P)$, a parameter space Θ, an open set in $\mathbb{R}^d$, and a measurable map

$$\begin{aligned} X : \Omega \times \Theta &\rightarrow \mathcal{X} \subseteq \mathbb{R}^n \\ (\omega, \theta) &\mapsto X(\omega, \theta), \end{aligned}$$

where in Θ we consider its Borelian σ-field.

As usual, a statistic is a measurable map

$$\begin{aligned} T \;:\; & \mathcal{X} \rightarrow \mathbb{R}^m \\ & x \mapsto T(x) = y. \end{aligned}$$

For simplicity, take $m = d = 1$ and Θ an open interval in $\mathbb{R}$. Let us denote

$$g(\theta) := E_\theta(T) = E\left(T(X)\right),$$

then, under smoothness conditions on $g(\theta)$, we can evaluate

$$\partial_\theta E(T(X)),$$

where we write $\partial_\theta = \frac{\partial}{\partial \theta}$.

Hypothesis 2.1. *A square integrable statistic $T \in C^1$ is said to be regular if*

$$\partial_\theta E(T(X)) = E\left(\partial_\theta T(X)\right)$$

and $Var(T(X)) < \infty$.

Suppose that the family of random variables $\{X(\cdot, \theta), \theta \in \Theta\}$ is also regular in the following sense:

i) X has a density $p(\cdot;\theta) \in C^1$ for all $\theta \in \Theta$ with support, supp(X), independent of θ.

ii) $X(\omega,\cdot) \in C^1$ as a function of θ, $\forall \omega \in \Omega$. Furthermore, $\partial_\theta X_j \in L^2(\Omega)$ and $E(\partial_\theta X_j | X = x) \in C^1$ as a function of $x, \forall \theta \in \Theta$ and $j = 1, ..., n$.

iii) $\frac{\partial_{x_j}(E(\partial_\theta X_j|X)p(X;\theta))}{p(X;\theta)} \in L^2(\Omega)$, $j = 1, \ldots, n$; where, for any smooth function h, we denote $\partial_{x_j} h(X) := \partial_{x_j} h(x)_{|x=X}$

iv) Any statistic $T \in C^1$ with compact support in the interior of supp(X) is regular.

Remark 2.2. *Note that if $T \in C^1$ has compact support in the interior of supp(X) and the family $\{X(\cdot,\theta), \theta \in \Theta\}$ is regular then*

$$\begin{aligned}
\partial_\theta E(T(X)) &= E(\partial_\theta T(X)) = E(\sum_{j=1}^n \partial_{x_j} T(X) \partial_\theta X_j) \\
&= \int_{\mathbb{R}^n} \sum_{j=1}^n \partial_{x_j} T(x) E(\partial_\theta X_j | X = x) p(x;\theta) dx \\
&= -\int_{\mathbb{R}^n} T(x) \sum_{j=1}^n \partial_{x_j} \left(E(\partial_\theta X_j | X = x) p(x;\theta) \right) dx,
\end{aligned}$$

since

$$\lim_{x \to x_0} E(\partial_\theta X_j | X = x) p(x;\theta) T(x) = 0, j = 1, ..., n,$$

$\forall \theta \in \Theta$ and $\forall x_0 \in \partial$supp(X), where ∂supp(X) is the boundary of the support of X. So,

$$\partial_\theta E(T(X)) = -E\left(T(X) \sum_{j=1}^n \frac{\partial_{x_j}\left(E(\partial_\theta X_j|X)p(X;\theta)\right)}{p(X;\theta)} \right). \tag{2.1}$$

Here the quotient is defined as 0 if $p(X;\theta) = 0$.

The following result is a statement of Cramer-Rao's inequality.

Proposition 2.3. *Let T be a regular statistic satisfying i)-iii) and*

$$\lim_{x \to x_0} E(\partial_\theta X_j | X = x) p(x;\theta) = 0, j = 1, ..., n, \tag{2.2}$$

$\forall \theta \in \Theta$ and $\forall x_0 \in \partial$supp(X), including $x_0 = \infty$ in ∂supp(X) if the supp(X) is not compact, then

$$Var(T(X)) \geq \frac{(\partial_\theta E(T(X)))^2}{E\left(\sum_{j=1}^n \frac{\partial_{x_j}(E(\partial_\theta X_j|X)p(X;\theta))}{p(X;\theta)}\right)^2}, \tag{2.3}$$

provided that the denominator is not zero.

In most classical models, the above calculation can be straightforward, as the explicit form of the density function p is available. This is carried out in the following examples. Our goal later is to show that in some cases where the explicit density is not known, the above bound can also be written without using directly the form of p. This will be the case of elliptic diffusions.

If we also assume that

v) $p(x;\theta)$ is smooth as function of θ, for every fixed x.

vi) For any smooth statistic T with compact support in the interior of supp(X), $\partial_\theta \int_{\mathbb{R}^n} T(x)p(x;\theta)dx = \int_{\mathbb{R}^n} T(x)\partial_\theta p(x;\theta)dx, \forall\theta \in \Theta$,

we have the following proposition that shows that (**??**) is just the usual Cramer-Rao inequality.

Proposition 2.4. *Assume conditions i)-vi). Then, a.e.*

$$-\sum_{j=1}^{n} \frac{\partial_{x_j}\left(E(\partial_\theta X_j|X=x)p(x;\theta)\right)}{p(x;\theta)} = \partial_\theta \log p(x;\theta), \forall\theta \in \Theta.$$

We have assumed in the above proof that the support of X does not depend on θ. In the following example we suggest a localization method to treat the case where the support depends on θ.

3 Main results and conclusions

In this section, we give an alternative derivation of the Cramer-Rao bound using the integration by parts formula of Malliavin Calculus. We start explaining some of the basic concepts of Malliavin Calculus. Consider a probability space $(\Omega, \mathcal{F}, P)$ and a Gaussian subspace $\mathcal{H}$ of $L^2(\Omega, \mathcal{F}, P)$ whose elements are zero-mean Gaussian random variables. Let H be a separable Hilbert space with scalar product denoted by $\langle \cdot, \cdot \rangle_H$ and norm $||\cdot||_H$, we will assume that there exists an isometry

$$\begin{aligned} W: H &\rightarrow \mathcal{H} \\ h &\mapsto W(h) \end{aligned}$$

in the sense that

$$E(W(h_1)W(h_2)) = \langle h_1, h_2 \rangle_H.$$

Let $\mathcal{S}$ be the class of smooth random variables $T(W(h_1), W(h_2), ..., W(h_n))$ such that T and all its derivatives have polynomial growth. Given $T \in \mathcal{S}$ we can define its differential as

$$DT = \sum_{i=1}^{n} \partial_i T(W(h_1), W(h_2), ..., W(h_n))h_i.$$

DT can be seen as a random variable with values in H. Then we can define the stochastic derivative operator as

$$\begin{aligned} D: \mathbb{D}^{1,2} \subseteq L^2(\Omega, \mathbb{R}) &\longrightarrow L^2(\Omega, H) \\ T &\mapsto DT. \end{aligned}$$

where $\mathbb{D}^{1,2}$ is the closure of the class of smooth random variables with respect to the norm

$$||T||_{1,2} = \left(E(|T|^2) + E(||DT||_H^2)\right)^{1/2}$$

Let u be an element of $L^2(\Omega, H)$ and assume there is an element $\delta(u)$ belonging to $L^2(\Omega)$ and such that

$$E(\langle DT, u \rangle_H) = E(T\delta(u))$$

for any $T \in \mathbb{D}^{1,2}$, then we say that u belongs to the domain of δ (denoted by $dom(\delta)$) and that δ is the adjoint operator of D.

Proposition 3.1. *Let h be an element of H, then*

$$\delta(h) = W(h).$$

Proposition 3.2. *If*

$$u = \sum_{i=1}^{n} F_j h_j$$

where $F_j \in \mathcal{S}$ and h_j are elements of H then

$$\delta(u) = \sum_{j=1}^{n} F_j W(h_j) - \sum_{j=1}^{n} \langle DF_j, h_j \rangle_H.$$

In the following results we assume that our observations are expressed as the measurable map

$$\begin{aligned} X : \Omega \times \Theta &\rightarrow \mathbb{R}^n \\ (\omega, \theta) &\mapsto x = X(\omega, \theta), \end{aligned}$$

Θ an open subset of $\mathbb{R}$ with the Borelian σ-field and the σ-field in Ω is the σ-field generated by $\mathcal{H}$.

theorem 3.3. *Let $X_j \in \mathbb{D}^{1,2}, j = 1, ..., n$ and Z be a random variable with values in H, in the domain of δ, such that*

$$\langle Z, DX_j \rangle_H = \partial_\theta X_j. \tag{3.1}$$

If T is a regular unbiased estimator of $g(\theta)$ then

$$Var(T(X))Var(E(\delta(Z)|X)) \geq g'(\theta)^2. \tag{3.2}$$

Furthermore, assume that

i) X has density $p(x;\theta) \in C^1$ as function of θ with support, supp(X), independent of θ,

ii) Any smooth statistic with compact support in the interior of supp(X) is regular and $\partial_\theta \int_{\mathbb{R}^n} T(x)p(x;\theta)dx = \int_{\mathbb{R}^n} T(x)\partial_\theta p(x;\theta)dx$, for all $\theta \in \Theta$,

then

$$E(\delta(Z)|X) = \partial_\theta \log p(X;\theta), \text{ a.s. and for all } \theta \in \Theta.$$

Remark 3.4. *The next goal is to provide a somewhat standard way of finding Z. For example, if there exists U, an n-dimensional random vector with values on H such that*

$$\langle U_k, DX_j \rangle_H = \delta_{kj}$$

where δ_{kj} is Kronecker's delta then

$$Z = \sum_{k=1}^{n} U_k \partial_\theta X_k$$

*verifies condition (**??**) if $Z \in dom(\delta)$. In particular, if*

$$(A_{kj}) = (\langle DX_k, DX_j \rangle_H)^{-1}$$

is well defined then we can take

$$U_k = \sum_{j=1}^{n} A_{kj} DX_j.$$

*The matrix A is called the Malliavin covariance matrix and the property that its inverse is well defined implies (see Theorem 2.1.2 in [**?**]) that the random vector X has a density function $p(x;\theta)$.*

References

[1] O.E. Barndorff-Nielsen, S.E. Graversen, J. Jacod, M. Podolskij, N. Shephard, A central limit theorem for realised power and bipower variations of continuous semimartingales, in: Yu. Kabanov, R. Liptser and J. Stoyanov (Eds.), *From Stochastic Calculus to Mathematical Finance. Festschrift in Honour of A.N. Shiryaev*, Heidelberg: Springer, 2006, pp. 33–68.

[2] B.Z. Bobrovsky, E. Mayer-Wolf and M. Zakai, *Some Classes of Global Cramer-Rao Bounds*, Annals of Statistics, 15 (1987), 1421-1428.

[3] E. Gobet *Local asymptotic mixed normality property for elliptic disffusion: a Malliavin calculus approach.* Bernoulli **7**(6)(2001), 899–912.

[4] Y. A. Kutoyants, *Statistical inference for ergodic diffusion processes*, Springer, 2004.

[5] D. Nualart, *The Malliavin calculus an related topics.* 2nd edition, Springer-Verlag, 2006.

[6] N. Privault and A. Réveillac Superefficient estimation on the Wiener space. *C.R. Acad. Sci. Paris Sér. I Math.*,**343** (2006), 607–612.

[7] N. Privault and A. Réveillac, *Stein estimation for the drift of Gaussian processes using the Malliavin calculus.* Preprint 2007. To appear in the Annals of Statistics.

[8] N. Privault and A. Réveillac, *Stein estimation of Poisson process intensities.* Preprint 2007.