Statistical inference and Malliavin calculus

José M. Corcuera and Arturo Kohatsu-Higa

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.

Mathematics Subject Classification (2000). Primary 62Mxx, 60Hxx; Secondary 60Fxx, 62Fxx.

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

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. [3],[4]. We show in this article, that similar techniques can be used to derive expressions for the score function in a parametric statistical model and

This work is supported by the MEC Grant No. MTM2006-03211.

consequently we obtain the Fisher information and the 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 [5] 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 [9] and [10]), and in the context of a *Poisson process* (see [11]). An interesting discussion of different Cramer-Rao's bounds is given in [2].

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.

The paper is organized as follows. In the first section we formulate the statistical model in a way that appears clearly that the Cramer-Rao lower bound can be obtained after integrating by parts and using the Cauchy-Schwarz inequality. In the second section, we consider the parametric models associated with diffusion processes where the driving process is a Wiener process. In the last section we consider the case of diffusions with jumps. The basic reference on Malliavin calculus is [7]. In particular, we refer the reader to this textbook for definitions and notation.

As our goal is to concentrate in general principles rather than technical details, we briefly sketch the mathematical framework required and explain in the examples the procedure.

2. The Cramer-Rao lower bound

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 $\partial_\theta = \frac{\partial}{\partial \theta}$.

Definition 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, $\text{supp}(X)$, 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). $\frac{\partial_{x_j}(E(\partial_\theta X_j | X) p(X;\theta))}{p(X;\theta)} \in L^2(\Omega)$, $j = 1, \dots, n$; 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 $\text{supp}(X)$ is regular.

Remark 2.2. Note that if $T \in C^1$ has compact support in the interior of $\mathrm{supp}(X)$ 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 $\forall x_0 \in \partial \mathrm{supp}(X)$, where $\partial \mathrm{supp}(X)$ 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 $\partial supp(X)$ 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.

Proof. By the boundary condition (2.2) we have that

$$E\left(\sum_{j=1}^n \frac{\partial_{x_j} \left(E(\partial_\theta X_j | X) p(X;\theta) \right)}{p(X;\theta)} \right) = 0.$$

Then, by condition iii) and since $\partial_\theta X_j \in L^2$ and T is regular we also have that $E[|\partial_\theta X_j T(X)|] < \infty$ and therefore

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

$\forall \theta \in \Theta$ and $\forall x_0 \in \partial \text{supp}(X)$, then

$$\partial_\theta E(T(X)) = -E\left(T(X) \sum_{j=1}^{n} \frac{\partial_{x_j} \left(E(\partial_\theta X_j | X) p(X;\theta) \right)}{p(X;\theta)} \right) \tag{2.4}$$

and we can write

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

The result follows from the Cauchy-Schwarz inequality. □

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.

Example 1. Assume that $X = (X_1, X_2, ..., X_n)$ where $X_i = \frac{U_i}{\theta}$ are i.i.d. r.v. with $U_i \sim \exp(1)$. This situation corresponds to a usual parametric model of n independent observations exponentially distributed with parameter θ. We have that

$$\begin{aligned}
\partial_\theta X_j &= -\frac{X_j}{\theta}; \\
E(\partial_\theta X_j|X)p(X;\theta) &= -X_j\theta^{n-1}\exp\{-\theta\sum_{j=1}^n X_j\}; \\
\frac{\partial_{x_j}\left(E(\partial_\theta X_j|X)p(X;\theta)\right)}{p(X;\theta)} &= -\frac{1}{\theta}+X_j; \\
\sum_{j=1}^n \frac{\partial_{x_j}\left(E(\partial_\theta X_j|X)p(X;\theta)\right)}{p(X;\theta)} &= \sum_{j=1}^n X_j - \frac{n}{\theta}; \\
E\left(\sum_{j=1}^n \frac{\partial_{x_j}\left(E(\partial_\theta X_j|X)p(X;\theta)\right)}{p(X;\theta)}\right)^2 &= Var(\sum_{j=1}^n X_j) \\
&= Var(\partial_\theta \log(p(X;\theta)) = I(\theta),
\end{aligned}$$

where $I(\theta)$ is the Fisher information. So we see that, in this case, (2.3) is the classical Cramer-Rao lower bound.

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 (2.3) is just the usual Cramer-Rao inequality.

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

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

Proof. We have seen in (2.1) that

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

On the other hand,

$$\begin{aligned}\partial_\theta E(T(X)) &= \partial_\theta \int_{\mathbb{R}^n} T(x)p(x;\theta)dx \\ &= \int_{\mathbb{R}^n} \partial_\theta \log p(x;\theta)T(x)p(x;\theta)dx \\ &= E(T(X)\partial_\theta \log p(X;\theta)).\end{aligned}$$

Using a density argument w.r.t. T, we have that

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

□

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.

Example 2. Assume that $X = (X_1, X_2, ..., X_n)$ where $X_i = \theta U_i$, with $U_i \sim$ Uniform$(0,1)$ independent r.v.'s for $i = 1, ..., n$ and $\theta \in \mathbb{R}$. Consider a function $T = T(X) \in C^1$ and function $\pi : [0,1]^n \to \mathbb{R}$ also belonging to C^1 in $(0,1)^n$ and continuous in $[0,1]^n$ such that $\pi(U) = 0$

if there exists $i=1,...,n$ with $U_i\in\{0,1\}$. Then

$$
\begin{aligned}
&\partial_\theta E(T(X)\pi(U)) \\
&= E(\partial_\theta T(X)\pi(U)) = E(\sum_{j=1}^n \partial_{x_j} T(X)\partial_\theta X_j \pi(U)) \\
&= \sum_{j=1}^n E(\partial_{x_j} T(X)\partial_\theta X_j \pi(U)) \\
&= \sum_{j=1}^n \frac{1}{\theta^n}\int_{[0,\theta]}\cdots\int_{[0,\theta]} \partial_{x_j} T(x) E(\partial_\theta X_j | X=x)\pi(\frac{x}{\theta})dx \\
&= -\sum_{j=1}^n \frac{1}{\theta^n}\int_{[0,\theta]}\cdots\int_{[0,\theta]} T(x)\partial_{x_j}(E(\partial_\theta X_j | X=x)\pi(\frac{x}{\theta}))dx \\
&= -E(T(X)\sum_{j=1}^n \partial_{x_j}(\partial_\theta X_j \pi(\frac{X}{\theta}))) \\
&= -E((T(X)-E(T(X)))\sum_{j=1}^n \partial_{x_j}(\partial_\theta X_j \pi(\frac{X}{\theta})))
\end{aligned}
$$

then we have that

$$
Var(T) \geq \frac{(\partial_\theta E(T(X)\pi(U)))^2}{E(\sum_{j=1}^n \partial_{x_j}(\partial_\theta X_j \pi(\frac{X}{\theta})))^2}
$$

where

$$
\partial_{x_j}(\partial_\theta X_j \pi(\frac{X}{\theta})) = \frac{1}{\theta}\partial_{u_j}(U_j\pi(U)).
$$

Note that if $\pi(U)=1$ (and this could be considered as a limit case) then the Cramer-Rao bound is of order n^{-2} which corresponds to the variance of $\max(X_1, X_2, ..., X_n)$ (the maximum likelihood estimator of θ) .

3. Statistical Inference in Gaussian spaces

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*

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

Proof. Without loss of generality we can assume that $||h|| = 1$ and that $T = T(W(h), W(h_2), ..., W(h_n))$ is in $\mathcal{S}$ with h_i , $i = 2, ..., n$

orthogonal to h. Then

$$\begin{aligned} & E(\langle DT,h\rangle_H) \\ = \; & E(\partial_1 T) = E(\int_{\mathbb{R}} \partial_1 T(x_1, W(h_2),...,W(h_n))\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_1^2}dx_1 \\ = \; & E(\int_{\mathbb{R}} x_1 T(x_1, W(h_2),...,W(h_n))\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_1^2}dx_1 \\ = \; & E(TW(h)). \end{aligned}$$

□

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

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

Proof.

$$\begin{aligned} & E(T\delta(u)) \\ = \; & \sum_{j=1}^{n} E(TF_j W(h_j)) - \sum_{j=1}^{n} E(T\langle DF_j, h_j\rangle_H) \\ = \; & \sum_{j=1}^{n} E(\langle D(TF_j), h_j\rangle) - \sum_{j=1}^{n} E(T\langle DF_j, h_j\rangle_H \\ = \; & \sum_{j=1}^{n} E(\langle D(TF_j) - TDF_j, h_j\rangle_H) \\ = \; & \sum_{j=1}^{n} E(F_j\langle DT, h_j\rangle_H) \\ = \; & E(\langle DT, \sum_{j=1}^{n} F_j h_j\rangle_H) \\ = \; & E(\langle DT, u\rangle_H). \end{aligned}$$

□

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

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

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

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

Proof.

$$\partial_\theta E_\theta(T(X)) = \sum_{k=1}^{n} E(\partial_{x_k} T(X) \partial_\theta X_k).$$

If we have Z with

$$\langle Z, DX_j \rangle_H = \partial_\theta X_j$$

then

$$\partial_\theta E(T(X)) = E(\partial_{x_k} T(X) \langle Z, DX_k \rangle_H).$$

By the chain rule for the derivative operator D,

$$DT(X) = \partial_{x_k} T(X) DX_k,$$

then

$$\begin{array}{rcl} \partial_\theta E(T(X)) & = & E(\langle Z, DT(X) \rangle_H) \\ & = & E(T(X)\delta(Z)). \end{array}$$

Finally, by the Cauchy-Schwarz inequality

$$(\partial_\theta E(T(X)))^2 \leq Var(T(X))Var(E(\delta(Z)|X)).$$

Next, we have

$$\begin{aligned} E(T(X)\delta(Z)) &= \partial_\theta E(T(X)) \\ &= \partial_\theta \int_{\mathbb{R}^n} T(x)p(x;\theta)dx \\ &= \int_{\mathbb{R}^n} T(x)\partial_\theta p(x;\theta)dx \\ &= \int_{\mathbb{R}^n} T(x)\partial_\theta \log p(x;\theta)p(x;\theta)dx \\ &= E(T(X)\partial_\theta \log p(X;\theta)). \end{aligned}$$

Therefore the result follows from a density argument. □

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 (3.1) 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 [7]) that the random vector X has a density function $p(x;\theta)$.

In order to understand the role of each of the elements in Theorem 3.3, we treat some classical examples from time series analysis in the present framework.

Example 3. Let $X_j = \varepsilon_j + \theta, 1 \leq j \leq n$, where ε_j are independent standard normally distributed random variables for $j = 1, ..., n$. Let $\mathcal{H}$ be the linear space generated by the sequence $\varepsilon_1, \varepsilon_2, ...$, then

$$DX_j = e_j$$

where $W(e_j) = \varepsilon_j$, and $\langle e_j, e_k \rangle_H = E(\varepsilon_j \varepsilon_k) = \delta_{jk}$, so $A_{kj} = \delta_{kj}$, and

$$Z = \sum_{j,k=1}^{n} \partial_\theta X_k A_{kj} DX_j = \sum_{j=1}^{n} e_j,$$

since $\partial_\theta X_k = 1$. Then

$$\delta(Z) = \sum_{j=1}^{n} W(e_j) = \sum_{j=1}^{n} \varepsilon_j = \sum_{j=1}^{n} X_j - n\theta.$$

Therefore we obtain the classical Cramer-Rao lower bound

$$Var_\theta(T) \geq \frac{(\partial_\theta E_\theta(T))^2}{n}.$$

Example 4. Let $X_j = \theta \varepsilon_j$ where ε_j are independent standard normally distributed r.v. Let $\mathcal{H}$ be again the linear space generated by the sequence $\varepsilon_1, \varepsilon_2, ...$, then

$$DX_j = \theta e_j$$

where $W(e_j) = \varepsilon_j, A_{jk} = \frac{1}{\theta^2}\ \delta_{jk}$, and $\partial_\theta X_k = \varepsilon_j$, so we can take

$$Z = \sum_{j,k=1}^{n} \partial_\theta X_k A_{kj} DX_j = \sum_{j=1}^{n} \frac{1}{\theta} \varepsilon_j e_j.$$

then using Proposition 3.2,

$$\delta(Z) = \sum_{j=1}^{n} \frac{1}{\theta} \varepsilon_j^2 - \sum_{j=1}^{n} \frac{1}{\theta} \langle e_j, e_j \rangle_H = \sum_{j=1}^{n} \frac{X_j^2}{\theta^3} - \frac{n}{\theta}.$$

Therefore we also obtain the classical Cramer-Rao lower bound

$$Var(T) \geq \frac{\theta^2 (\partial_\theta E_\theta(T))^2}{2n}$$

As we mention in remark 3.4, the problem is how to find Z. The following elementary proposition can be useful in this sense.

Proposition 3.5. *Let $X = (X_1, X_2, ..., X_n)^T$ and $Y = (Y_1, Y_2, ..., Y_n)^T$ be two random vectors with components in $\mathbb{D}^{1,2}$, where $Y = h(\theta, X)$ with $h \in C^{1,1}$ and $h(\theta, \cdot)$ one to one for all θ. Assume that*

$$\langle Z, DY_j \rangle_H = \partial_\theta Y_j.$$

Then,

$$\langle Z, DX_j \rangle_H = \partial_\theta X_j.$$

Also we have that

$$\sum_{r,l=1}^{n} DY_r B_{rl}(d_\theta Y_l - (\partial_\theta h_l)(\theta, X)) = \sum_{r,l=1}^{n} DX_r A_{rl} \partial_\theta X_l$$

where $B_{kj} = (\langle DY_j, DY_k \rangle_H)^{-1}$, $A_{kj} = (\langle DX_j, DX_k \rangle_H)^{-1}$ and d_θ means total derivative with respect to θ. That is, $d_\theta Y_l = \partial_\theta(h(\theta, X))$.

Proof. It is straightforward □

Next, we give an application of the above proposition.

Example 5. Let $X_j = \theta X_{j-1} + \varepsilon_j$, $j = 1, ..., n$ where X_0 is a constant and ε_j are independent standard normal distributed. Let $\mathcal{H}$ be the linear space generated by the sequence $\varepsilon_1, \varepsilon_2, ...$, then define $Y = h(\theta, X)$, where $h_j(\theta, x) = x_j - \theta x_{j-1}$ with $x_0 = X_0$. Therefore $Y_j = \varepsilon_j$ and

$$d_\theta Y_j = d_\theta \varepsilon_j = 0, (\partial_\theta h_j)(\theta, X) = -X_{j-1},$$

so we have that

$$\sum_{r,l=1}^{n} DY_r A_{rl}(d_\theta Y_l - (\partial_\theta h_l)(\theta, X)) = \sum_{r,l=1}^{n} e_r \delta_{rl} X_{l-1}.$$

As $W(e_j) = \varepsilon_j$, then

$$\begin{aligned}
&\sum_{r,l=1}^{n} \delta(DY_r A_{rl}(d_\theta Y_l - (\partial_\theta h_l)(\theta, X))) \\
= &\sum_{r,l=1}^{n} \delta(e_r \delta_{rl} X_{l-1}) \\
= &\sum_{l=1}^{n} W(e_l) X_{l-1} - \sum_{l=1}^{n} \langle DX_{l-1}, e_l \rangle_H. \\
= &\sum_{l=1}^{n} \varepsilon_l X_{l-1} = \sum_{l=1}^{n} (X_l - \theta X_{l-1}) X_{l-1}
\end{aligned}$$

and we obtain the score function. Another way to obtain Z in Theorem 3.3 would be to use the relation

$$\partial_\theta X_j = X_{j-1} + \theta \partial_\theta X_{j-1}$$
$$DX_j = e_j + \theta DX_{j-1}$$

so that

$$\langle \sum_{l=1}^{n} e_l X_{l-1}, DX_j \rangle = X_{j-1} + \theta \langle \sum_{l=1}^{n} e_l X_{l-1}, DX_{j-1} \rangle$$

and therefore by uniqueness of solutions for difference equations, we have that

$$\partial_\theta X_j = \langle \sum_{l=1}^{n} e_l X_{l-1}, DX_j \rangle.$$

So, if we define

$$Z = \sum_{l=1}^{n} e_l X_{l-1},$$

then

$$\delta(Z) = \sum_{l=1}^{n} \varepsilon_l X_{l-1} = \sum_{l=1}^{n} (X_l - \theta X_{l-1}) X_{l-1}.$$

And consequently

$$\partial_\theta \log p(X;\theta) = \sum_{l=1}^{n} (X_l - \theta X_{l-1}) X_{l-1}.$$

The continuous version of the previous example is given by the following one.

Example 6. Let $X^{(n)} = (X_{t_1}, X_{t_2}, ..., X_{t_n})$, $0 < t_1 < t_2... < t_n$, be observations of the Ornstein-Uhlenbeck process

$$dX_t = -\theta X_t dt + dB_t, t \geq 0, X_0 = 0.$$

Or, by integrating,

$$X_t = \int_0^t e^{-\theta(t-s)} dB_s.$$

Here $\{B_t, t \geq 0\}$ is a standard one dimensional Brownian motion. Let $\mathcal{H}$ be the closed linear space generated by the random variables $\{B_t, 0 \leq t \leq T\}$ and $H = L^2([0,T], dx)$. Then the map

$$\begin{aligned} W : H &\rightarrow \mathcal{H} \\ \mathbf{1}_{[0,t]}(\cdot) &\mapsto W(\mathbf{1}_{[0,t]}) \equiv \int_0^T \mathbf{1}_{[0,t]}(s) dB_s = B_t \end{aligned}$$

defines a linear isometry, and $W(h)$ is the stochastic integral of the function h. Consequently $DX_t = e^{-\theta(t-\cdot)} \mathbf{1}_{[0,t]}(\cdot)$. We have

$$d\partial_\theta X_t = -X_t dt - \theta \partial_\theta X_t dt, t \geq 0, \partial_\theta X_0 = 0,$$

so

$$\begin{aligned} \partial_\theta X_t &= -\int_0^t e^{-\theta(t-s)} X_s ds \\ &= -\int_0^T X_s D_s X_t ds = -\langle X, DX_t \rangle_H, \end{aligned}$$

where we write $DX_t(s) = D_s X_t$. Then

$$\begin{aligned} \partial_\theta \log p(X^{(n)}; \theta) &= -E(\delta(X)|X^{(n)}) = -E(\int_0^T X_s dB_s | X^{(n)}) \\ &= -E(\int_0^T X_s dX_s + \theta \int_0^T X_s^2 ds | X^{(n)}). \end{aligned}$$

In particular the maximum likelihood estimator of θ is given by

$$\hat{\theta} = -\frac{E(\int_0^T X_s dX_s | X^{(n)})}{E(\int_0^T X_s^2 ds | X^{(n)})}.$$

Take now $\Delta t_i = \Delta_n$ and n observations in such a way that $\Delta_n \to 0$ and $n\Delta_n \to \infty$ when n goes to infinity. Then

$$\frac{1}{\sqrt{n\Delta_n}} \partial_\theta \log p(X^{(n)};\theta) = -E\left(\frac{1}{\sqrt{n\Delta_n}} \int_0^{n\Delta_n} X_s dB_s \middle| X^{(n)} \right).$$

By ergodicity

$$\frac{1}{n\Delta_n} \int_0^{n\Delta_n} X_s^2 ds \xrightarrow{P} \frac{1}{2\theta},$$

then we have that $\frac{1}{\sqrt{n\Delta_n}}\delta(X) \xrightarrow{\mathcal{L}} N(0, \frac{1}{2\theta})$ (see Theorem 1.19 in [6]). On the other hand

$$\begin{aligned} \delta(X) &= \int_0^{n\Delta_n} X_s dB_s = \int_0^{n\Delta_n} X_s dX_s + \theta \int_0^{n\Delta_n} X_s^2 ds \\ &= \sum_{i=1}^n X_{t_{i-1}} \Delta X_{t_i} + \theta \sum_{i=1}^n X_{t_{i-1}}^2 \Delta_n + \sum_{i=1}^n R_i, \end{aligned}$$

where

$$R_i = \int_{t_{i-1}}^{t_i} (X_s - X_{t_{i-1}}) dB_s + \theta \int_{t_{i-1}}^{t_i} X_{t_{i-1}} (X_s - X_{t_{i-1}}) ds.$$

By straightforward calculations one can see that $\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n R_i \xrightarrow{L^2} 0$, so

$$\frac{1}{\sqrt{n\Delta_n}}\delta(X) - \frac{1}{\sqrt{n\Delta_n}} \left(\sum_{i=1}^n X_{t_{i-1}} \Delta X_{t_i} + \theta \sum_{i=1}^n X_{t_{i-1}}^2 \Delta_n \right) \xrightarrow{L^2} 0,$$

and, since the second term is $X^{(n)}$ measurable, we have that

$$\frac{1}{\sqrt{n\Delta_n}}\delta(X) - E\left(\frac{1}{\sqrt{n\Delta_n}}\delta(X) \middle| X^{(n)} \right) \xrightarrow{L^2} 0,$$

so finally

$$E\left(\frac{1}{\sqrt{n\Delta_n}}\delta(X) \middle| X^{(n)} \right) \xrightarrow{\mathcal{L}} N(0, \frac{1}{2\theta}).$$

Note that the asymptotic Fisher information is given by $\frac{1}{2\theta}$. Also if we consider

$$Z_n(\theta, \theta + \frac{u}{\sqrt{n\Delta_n}}) := \log p(X^{(n)}; \theta + \frac{u}{\sqrt{n\Delta_n}}) - \log p(X^{(n)}; \theta),$$

we have

$$
\begin{aligned}
Z_n(\theta, \theta + \frac{u}{\sqrt{n\Delta_n}}) &= -\int_{\theta}^{\theta + \frac{u}{\sqrt{n\Delta_n}}} E(\delta(X) | X_{\theta'}^{(n)} = X_{\theta}^{(n)}) d\theta' \\
&= -\int_{\theta}^{\theta + \frac{u}{\sqrt{n\Delta_n}}} \left(\sum_{i=1}^{n} X_{t_{i-1}} \Delta X_{t_i} + \theta' \sum_{i=1}^{n} X_{t_{i-1}}^2 \Delta_n \right) d\theta' \\
&\quad -M_n
\end{aligned}
$$

where

$$
M_n = \int_{\theta}^{\theta + \frac{u}{\sqrt{n\Delta_n}}} E\left(\sum_{i=1}^{n} R_i | X_{\theta'}^{(n)} = X_{\theta}^{(n)} \right) d\theta'.
$$

We can show that $M_n \xrightarrow{L^2} 0$. In fact

$$
\begin{aligned}
E(M_n^2) &= \sum_{i=1}^{n} E\left(\int_{\theta}^{\theta + \frac{u}{\sqrt{n\Delta_n}}} E(R_i | X_{\theta'}^{(n)} = X_{\theta}^{(n)}) d\theta' \right)^2 \qquad (3.3) \\
&+2 \sum_{i<j}^{n} E\left(\int_{\theta}^{\theta + \frac{u}{\sqrt{n\Delta_n}}} E(R_i | X_{\theta'}^{(n)} = X_{\theta}^{(n)}) d\theta' \int_{\theta}^{\theta + \frac{u}{\sqrt{n\Delta_n}}} E(R_j | X_{\theta''}^{(n)} = X_{\theta}^{(n)}) d\theta'' \right),
\end{aligned}
$$

and

$$
\begin{aligned}
& E\left(\int_{\theta}^{\theta + \frac{u}{\sqrt{n\Delta_n}}} E(R_i | X_{\theta'}^{(n)} = X_{\theta}^{(n)}) d\theta' \right)^2 \\
\leq & \frac{u}{\sqrt{n\Delta_n}} \int_{\theta}^{\theta + \frac{u}{\sqrt{n\Delta_n}}} E(E(R_i^2 | X_{\theta'}^{(n)} = X_{\theta}^{(n)})) d\theta' \\
= & \frac{u}{\sqrt{n\Delta_n}} \int_{\theta}^{\theta + \frac{u}{\sqrt{n\Delta_n}}} E\left(E(R_i^2 | X_{\theta'}^{(n)}) \frac{p_i(\theta)}{p_i(\theta')} \right) d\theta',
\end{aligned}
$$

where $p_i(\theta)$ is the joint density of $\left((X_\theta)_{t_{i-1}}, (X_\theta)_{t_i} \right)$ evaluated at $X_{\theta'}^{(n)}$. Then, since

$$
\begin{aligned}
E\left(E(R_i^2 | X_{\theta'}^{(n)}) \frac{p_i(\theta)}{p_i(\theta')} \right) &\leq \left(E(E(R_i^2 | X_{\theta'}^{(n)})^2) \right)^{1/2} \left(E\left(\frac{p_i(\theta)}{p_i(\theta')} \right)^2 \right)^{1/2} \\
&\leq C \left(E(E(R_i^2 | X_{\theta'}^{(n)})^2) \right)^{1/2} \leq C \left(E(R_i^4) \right)^{1/2} \\
&\leq C\Delta_n^2
\end{aligned}
$$

we have that

$$E\left(\int_{\theta}^{\theta+\frac{u}{\sqrt{n\Delta_n}}} E(R_i|X_{\theta'}^{(n)}=X_{\theta}^{(n)})d\theta'\right)^2 \le C\frac{u^2}{n}\Delta_n.$$

Here we have used Burkholder's inequality and, in order to prove that $E\left(\frac{p_i(\theta)}{p_i(\theta')}\right)^2$ is uniformly bounded, we used the explicit form of the Gaussian density of $\left((X_\theta)_{t_{i-1}},(X_\theta)_{t_i}\right)$. Finally the term (3.3) goes to zero since the covariance function of X goes to zero exponentially. Then

$$\begin{aligned} Z_n(\theta,\theta+\frac{u}{\sqrt{n\Delta_n}}) &= -\frac{u}{\sqrt{n\Delta_n}}\left(\sum_{i=1}^n X_{t_{i-1}}\Delta X_{t_i}+\theta\sum_{i=1}^n X_{t_{i-1}}^2\Delta_n\right) \\ &\quad -\frac{1}{2}\frac{u^2}{n\Delta_n}\sum_{i=1}^n X_{t_{in}}^2\Delta_n - M_n. \end{aligned}$$

and, since X is ergodic, we have that

$$Z_n(\theta,\theta+\frac{u}{\sqrt{n\Delta_n}}) \xrightarrow{\mathcal{L}} uN(0,\frac{1}{2\theta})-\frac{u^2}{4\theta}.$$

That is, the model satisfies the LAN (Local Asymptotic Normality) property.

3.1. Elliptic diffusions

Let $X^{(n)}=(X_{t_1},X_{t_2},...,X_{t_n})$ be the vector of observations. Here, $t_i=i\Delta_n$, with $n=\Delta_n^{-1}$ where we omit the dependence of the partition on n. X is defined as the solution to

$$X_t = x+\int_0^t b_s(\theta,X_s)ds+\int_0^t \sigma_s(\theta,X_s)dB_s.$$

and B is a standard Brownian motion. Assume that b_s, σ_s and their derivatives with respect to θ and x are $C_b^{1.3}$ (as functions of t and x) and that σ_s is uniformly bounded and uniformly elliptic. Sometimes, when the arguments are clear, we will write σ_s and b_s instead of $\sigma_s(\theta,X_s)$ and $b_s(\theta,X_s)$ respectively. Note that the setting here is different from the previous example where the time frame for the data satisfied that $n\Delta_n\to\infty$ while in this subsection we have that $n\Delta_n=1$.

It can be shown that (see [7])

$$DX_t = \partial_x X_t \left(\partial_x X_\cdot\right)^{-1} \sigma_\cdot(\theta, X_\cdot)\mathbf{1}_{[0,t]}(\cdot).$$

Define

$$\beta_t = \left(\partial_x X_t\right)^{-1} \partial_\theta X_t$$

and

$$\tilde{\beta}_t = \partial_x X_t \left(\sigma_t(\theta, X_t)\right)^{-1} \sum_{i=1}^{n} a(t)(\beta_{t_i} - \beta_{t_{i-1}})\mathbf{1}_{\{t_{i-1}\leq t<t_i\}},$$

where

$$a \in L^2([0,T]), \int_{t_{i-1}}^{t_i} a(t)dt = 1, i = 1, ..., n.$$

Then, it is easy to see that

$$\partial_\theta X_{t_i} = \langle \tilde{\beta}, DX_{t_i} \rangle_H.$$

In fact

$$\begin{aligned}
\langle \tilde{\beta}, DX_{t_i} \rangle_H &= \partial_x X_{t_i} \int_0^{t_i} \sum_{j=1}^{n} a(t)(\beta_{t_j} - \beta_{t_{j-1}})\mathbf{1}_{\{t_{j-1}\leq t<t_j\}} dt \\
&= \partial_x X_{t_i} \sum_{j=1}^{i} (\beta_{t_j} - \beta_{t_{j-1}}) \int_{t_{j-1}}^{t_j} a(t)dt \\
&= \partial_x X_{t_i} \sum_{j=1}^{i} (\beta_{t_j} - \beta_{t_{j-1}}) = \partial_x X_{t_i}(\beta_{t_i} - \beta_{t_0}) \\
&= \partial_x X_{t_i} \beta_{t_i} = \partial_\theta X_{t_i}.
\end{aligned}$$

By the assumption of uniform ellipticity and smoothness of the coefficients, we have that $\tilde{\beta}$ is in the domain of δ and that the score function is given by

$$E(\delta(\tilde{\beta})|X^{(n)}),$$

where we have for $a(t) = n$,

$$\begin{aligned}
\frac{1}{n}\delta(\tilde{\beta}) &= \sum_{i=1}^{n} (\beta_{t_i} - \beta_{t_{i-1}}) \int_{t_{i-1}}^{t_i} \partial_x X_t \left(\sigma_t(\theta, X_t)\right)^{-1} dB_t \\
&\quad - \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} D_t \beta_{t_i} \partial_x X_t \left(\sigma_t(\theta, X_t)\right)^{-1} dt.
\end{aligned}$$

Conditions i) and ii) of Theorem 3.3 are fulfilled since supp(X) is $\mathbb{R}^n$ and $\partial_\theta p(x;\theta)$ is uniformly bounded with respect to θ by an integrable function. In addition,

$$\partial_x X_t = 1 + \int_0^t \partial_x b_s \partial_x X_s ds + \int_0^t \partial_x \sigma_s \partial_x X_s dB_s,$$

and

$$\partial_\theta X_t = \int_0^t \left(\partial_\theta b_s + \partial_x b_s \partial_\theta X_s\right) ds + \int_0^t \left(\partial_\theta \sigma_s + \partial_x \sigma_s \partial_\theta X_s\right) dB_s,$$

so, by applying the Itô formula we have that

$$\beta_t = \frac{\partial_\theta X_t}{\partial_x X_t} = \int_0^t \mu_s ds + \int_0^t \frac{\partial_\theta \sigma_s}{\partial_x X_s} dB_s, \tag{3.4}$$

for a certain adapted process μ that can be explicitly calculated. Then, by Theorem 2.1 in [1] and the polarization identity, we have that

$$\sqrt{n}\left(\sum_{i=1}^n (\beta_{t_i} - \beta_{t_{i-1}}) \int_{t_{i-1}}^{t_i} \partial_x X_t \sigma_t^{-1} dB_t \right.$$
$$\left. - \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \frac{\partial_\theta \sigma_t}{\sigma_t} dt \right) \xrightarrow{\mathcal{L}} \sqrt{2} \int_0^1 \left(\frac{\partial_\theta \sigma_s}{\sigma_s}\right) dW_s.$$

Here W is a Brownian motion independent of B. Moreover, for $t \leq t_i$,

$$D_t \beta_{t_i} = \frac{\partial_\theta \sigma_t}{\partial_x X_t} + \int_t^{t_i} D_t \mu_s D_t X_s ds + \int_t^{t_i} \partial_x \left(\frac{\partial_\theta \sigma_s}{\partial_x X_s}\right) D_t X_s dB_s.$$

and therefore we have

$$\sqrt{n}\left(\sum_{i=1}^n \int_{t_{i-1}}^{t_i} D_t \beta_{t_i} \partial_x X_t {\sigma_t}^{-1} dt - \int_{t_{i-1}}^{t_i} \frac{\partial_\theta \sigma_t}{\sigma_t} dt \right)$$
$$\xrightarrow{L^2} 0.$$

Consequently

$$\frac{1}{\sqrt{n}} \delta(\tilde{\beta}) \xrightarrow{\mathcal{L}} \sqrt{2} \int_0^1 \left(\frac{\partial_\theta \sigma_s}{\sigma_s}\right) dW_s. \tag{3.5}$$

Moreover, by using the polarization identity and Burkholder's inequality we can see that

$$\frac{1}{\sqrt{n}}\delta(\tilde{\beta}) - \sqrt{n}\sum_{i=1}^{n}\left\{\frac{\partial_\theta \sigma_{t_{i-1}}}{\sigma^3_{t_{i-1}}}(\Delta X_{t_i})^2 - \frac{1}{n}\frac{\partial_\theta \sigma_{t_{i-1}}}{\sigma_{t_{i-1}}}\right\} \xrightarrow{L^\alpha} 0,$$

for any $\alpha > 0$. So, finally

$$\frac{1}{\sqrt{n}}E(\delta(\tilde{\beta})|X^{(n)}) \xrightarrow{\mathcal{L}} \sqrt{2}\int_0^1 \left(\frac{\partial_\theta \sigma_s}{\sigma_s}\right) dW_s.$$

In particular this implies that the asymptotic Fisher information for θ is given by

$$2E\left[\int_0^1 \left(\frac{\partial_\theta \sigma_s}{\sigma_s}\right)^2 ds\right].$$

If we define

$$Z_n(\theta, \theta + \frac{u}{\sqrt{n}}) = \log p(X^{(n)}; \theta + \frac{u}{\sqrt{n}}) - \log p(X^{(n)}; \theta)$$

$$Z_n(\theta, \theta + \frac{u}{\sqrt{n}}) = \int_\theta^{\theta + u/\sqrt{n}} E(\delta(\tilde{\beta})|X^{(n)}_{\theta'} = X^n_\theta)d\theta'.$$

It can be seen that

$$E(\delta(\tilde{\beta})|X^{(n)}_\theta) = n\sum_{i=1}^{n}\left\{\frac{\partial_\theta \sigma_{t_{i-1}}}{\sigma^3_{t_{i-1}}}(\Delta X_{t_i})^2 - \frac{1}{n}\frac{\partial_\theta \sigma_{t_{i-1}}}{\sigma_{t_{i-1}}}\right\} + \sum_{i=1}^{n} R_i(\theta, (X_\theta)_{t_{i-1}}, (X_\theta)_{t_i}) \tag{3.6}$$

where $E(|R_i(\theta, (X_\theta)_{t_{i-1}}, (X_\theta)_{t_i})|^\alpha)^{1/\alpha} = O(1/n)$, uniformly in i, for any $\alpha > 0$. Then,

$$\begin{aligned} E_\theta\left(|R_i(\theta', (X_\theta)_{t_{i-1}}, (X_\theta)_{t_i})|\right) &= E_{\theta'}\left(|R_i(\theta', (X_{\theta'})_{t_{i-1}}, (X_{\theta'})_{t_i})|\frac{p_i(\theta)}{p_i(\theta')}\right) \\ &= (E_{\theta'}(|R_i|^\alpha))^{1/\alpha}\left(E_{\theta'}\left[\left(\frac{p_i(\theta)}{p_i(\theta')}\right)^\beta\right]\right)^{1/\beta}, \end{aligned}$$

where $1/\alpha + 1/\beta = 1, \alpha > 1, \beta > 1$, $R_i = R_i(\theta', (X_\theta)_{t_{i-1}}, (X_\theta)_{t_i})$ and $p_i(\theta)$ is the joint density of $\left((X_\theta)_{t_{i-1}}, (X_\theta)_{t_i}\right)$ evaluated at $X_{\theta'}$. The expectation with respect to this process is denoted by $E_{\theta'}$.

Now, since $\sigma \in C^{1,3}$ is uniformly elliptic, transition densities can be bounded from above and below uniformly by the Gaussian kernel, and then by continuity there exists $\beta > 1$ such that

$$E_{\theta'}\left[\left(\frac{p_i(\theta)}{p_i(\theta')}\right)^{\beta}\right] < C,$$

see Proposition 5.1 in ([5]). So

$$\begin{aligned}\int_{\theta}^{\theta+u/\sqrt{n}} E|R_i(\theta', (X_\theta)_{t_{i-1}}, (X_\theta)_{t_i})|d\theta' &\leq C\int_{\theta}^{\theta+u/\sqrt{n}} \left(E|R_i|^{\alpha}\right)^{1/\alpha} d\theta' \\ &\leq C\frac{u}{n^{3/2}}.\end{aligned}$$

Now, we only need to calculate the derivatives, with respect to θ fixed $X^{(n)}$, of the expression (3.6). We obtain, after some calculations,

$$\begin{aligned}\frac{1}{n}\partial_\theta E(\delta(\tilde{\beta})|X^{(n)}) &= \sum_{i=1}^{n}\left\{\frac{\partial_\theta^2\sigma_{t_{i-1}}}{\sigma_{t_{i-1}}^3}\left(\Delta X_{t_i}\right)^2 - \frac{1}{n}\frac{\partial_\theta^2\sigma_{t_{i-1}}}{\sigma_{t_{i-1}}}\right\} \\ &+\sum_{i=1}^{n}\left\{-3\frac{\left(\partial_\theta\sigma_{t_{i-1}}\right)^2}{\sigma_{t_{i-1}}^4}\left(\Delta X_{t_i}\right)^2 + \frac{1}{n}\frac{\left(\partial_\theta\sigma_{t_{i-1}}\right)^2}{\sigma_{t_{i-1}}^2}\right\} \\ &+o_p(1).\end{aligned}$$

Finally, by taking into account the continuity of the derivatives with respect to θ, we have,

$$Z_n(\theta, \theta + \frac{u}{\sqrt{n}}) \xrightarrow{\mathcal{L}} u\int_0^1 \sqrt{2}\frac{\partial_\theta\sigma_s}{\sigma_s}dW_s - \frac{u^2}{2}\int_0^1 2\left(\frac{\partial_\theta\sigma_s}{\sigma_s}\right)^2 ds.$$

That is, the model satisfies the LAMN (Local Asymptotic Mixed Normality) property.

4. Statistical inference for models with jumps

In this section, we show how to treat cases where the observed process have jumps. Essentially the question reduces to the use of an appropriate extension of Malliavin's Calculus to processes with jumps. That is, the integration by parts formulas. First, we consider the case when assume that the vector of observations $Z^{(n)}$ has a Gaussian component (c.f. [4]). That is, Z is a sequence of $\mathcal{G} = \mathcal{F} \times \mathcal{H}$-measurable random variables where $\mathcal{F}$ is the σ-field generated by the isonormal Gaussian process and $\mathcal{H}$ is a σ-field independent of $\mathcal{F}$. For instance

$$Z_k = F(\theta, X_k, Y_k), k = 1, ..., n$$

where $X^{(n)} = (X_1, ..., X_n)$ is a $\mathcal{G}$-measurable random vector and $Y^{(n)} = (Y_1, ..., Y_n)$ is independent of $\mathcal{G}$. We also set $Z^{(n)} = (Z_1, ..., Z_n)$.

Let T be a regular statistic $T = T(Z^{(n)})$ and $g(\theta) = E_\theta(T)$. Then we have

$$\begin{aligned} \partial_\theta E_\theta(T) &= \partial_\theta E(T(Z^{(n)})) = E(\partial_\theta T(Z^{(n)})) \\ &= E(\partial_{z_i} T(Z^{(n)}) \partial_\theta Z_i). \end{aligned}$$

Now we extend the operator D to the product filtration (this is also called partial Malliavin Calculus) so that

$$DZ_k = \partial_{x_k} F DX_k$$

Then using this duality property we have that

$$DT = \partial_{z_i} T(Z^{(n)}) DZ_i = \partial_{z_i} T(Z^{(n)}) \partial_{x_i} F DX_i.$$

As before, if there exists a random variable V with values in H such that

$$\langle DZ_i, V \rangle_H = \partial_\theta Z_i,$$

we will have that

$$\partial_\theta \log p(Z^{(n)}; \theta) = E(\delta(V) | Z^{(n)}).$$

4.1. Jump diffusions

Let $X^{(n)} = (X_{t_1}, X_{t_2}, ..., X_{t_n})$, $t_i = i\Delta_n$, where $n = \Delta_n^{-1}$ and

$$X_t = x + \int_0^t b_s(\theta, X_s) ds + \int_0^t \sigma_s(\theta, X_s) dB_s + \int_0^t c_{s,z}(\theta, X_{s-}) M(dz, ds).$$

Where $M(dz, ds)$ is a compensated Poisson random measure with intensity $\nu(dz, ds) = \nu_s(dz) ds$ and B is an independent standard Brownian motion. Assume that b_s, σ_s, $c_{s,z}$ and their derivatives with respect to θ and x are $C_b^{1.3}$ (as functions of t and x) and that σ_s is uniformly bounded below by a positive constant. By using the same arguments that in the continuous case and assuming that $\partial_x c_{s,z}(\theta, x) > -1$, we have that

$$DX_t = \partial_x X_t \left(\partial_x X_\cdot\right)^{-1} \sigma_\cdot(\theta, X_\cdot) \mathbf{1}_{[0,t]}(\cdot).$$

Then, writing

$$\beta_t = (\partial_x X_t)^{-1} \partial_\theta X_t$$

and

$$\tilde{\beta}_t = \partial_x X_t \left(\sigma_t(\theta, X_t)\right)^{-1} \sum_{i=1}^{n} a(t)(\beta_{t_i} - \beta_{t_{i-1}}) \mathbf{1}_{\{t_{i-1} \leq t < t_i\}},$$

where

$$a \in L^2([0,T]), \int_{t_{i-1}}^{t_i} a(t)dt = 1, i = 1, ..., n,$$

we have

$$\partial_\theta X_{t_i} = \langle \tilde{\beta}, DX_{t_i} \rangle_H.$$

Consequently,

$$E(\delta(\tilde{\beta})|X^{(n)}),$$

where

$$\begin{aligned} \delta(\tilde{\beta}) \quad = \quad & \sum_{i=1}^{n} (\beta_{t_i} - \beta_{t_{i-1}}) \int_{t_{i-1}}^{t_i} a(t) \partial_x X_t \left(\sigma_t(\theta, X_t)\right)^{-1} dB_t \\ & - \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} a(t) D_t \beta_{t_i} \partial_x X_t \left(\sigma_t(\theta, X_t)\right)^{-1} dt. \end{aligned}$$

Also we will have that

$$Z_n(\theta, \theta + \frac{u}{\sqrt{n}}) \xrightarrow{\mathcal{L}} u \int_0^1 \sqrt{2} \frac{\partial_\theta \sigma_s}{\sigma_s} dW_s - \frac{u^2}{2} \int_0^1 2 \left(\frac{\partial_\theta \sigma_s}{\sigma_s} \right)^2 ds.$$

Notice that we would obtain another asymptotic regime if σ_s would not depend on θ as we have seen in example 6. This is also the case in the next subsections.

4.2. The pure jump case

In the previous calculation, we have used the Brownian motion in order to use the integration by parts property that will lead to an expression of the score. In this section we use the jump structure and, for the sake of simplicity, we assume that there is no Brownian component. Clearly, this also leads to different expressions for the same models if we consider models with Brownian and jump components.

4.2.1. Processes with different intensity. Let Y_t be an observation of a compound Poisson process of parameter λ and with jumps given by a random variable X, then

$$\begin{aligned} &\partial_\lambda E(\mathbf{1}_{\{Y_t\in A\}}) = \partial_\lambda P(Y_t\in A) \\ = &\ \partial_\lambda \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} P(X_1+...+X_n\in A) \\ = &\ -t\sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} P(X_1+...+X_n\in A) \\ &+t\sum_{n=0}^{\infty} \frac{ne^{-\lambda t}(\lambda t)^{n-1}}{n!} P(X_1+...+X_n\in A) \\ = &\ -tP(Y_t\in A)+\frac{1}{\lambda}\sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} nP(X_1+...+X_n\in A) \\ = &\ -tE(\mathbf{1}_{\{Y_t\in A\}})+\frac{1}{\lambda}E(\mathbf{1}_{\{Y_t\in A\}}N_t) = E(\mathbf{1}_{\{Y_t\in A\}}(\frac{N_t}{\lambda}-t)), \end{aligned}$$

where we should understand that $X_1+...+X_n:=0$ if $n=0$. Suppose now that we observe $Y_{t_1}, Y_{t_2}, .., Y_{t_n}$, $0\le t_1\le .. \le t_n\le T$. Since the increments are independent, the score function is given by

$$E(\frac{N_T}{\lambda}-T|Y_{t_1}, Y_{t_2}, ..., Y_{t_n}) = \lambda^{-1}\sum_{i=0}^{n-1} E(N_{t_{i+1}}-N_{t_i}|Y_{t_{i+1}}-Y_{t_i})-T.$$

Here

$$E(N_t|Y_t) = \sum_{k=1}^{\infty} k\frac{f^{(k)}(Y_t)(\lambda t)^k}{k!\sum_{i=0}^{\infty} f^{(i)}(Y_t)\frac{(\lambda t)^i}{i!}},$$

where $f^{(i)}$ stands for the i-th convolution of the density of the random variable X. We remark that this is related to Proposition 3.6 in [3], where the authors use Girsanov's theorem.

The trajectories of a compound Poisson process can be described by indicating the time and amplitude of the jumps. To see this, let $\Omega = \cup_{n=1}^{\infty}(\mathbb{R}_+\times\mathbb{R})^n$ be the sample space and denote an element $\omega\in\cup_{n=1}^{\infty}(\mathbb{R}_+\times\mathbb{R})^n$ as $\omega=((s_1,x_1),...,(s_k,x_k))$ for certain natural k. In this space we assume that the canonical random variables $s_j-s_{j-1}\sim\exp(\lambda)$ and $x_j\sim f(x)$ are independent sequences and also independent of each other. Next, consider the maps

$$\begin{aligned} Y_t : \Omega &\rightarrow \mathbb{R} \\ ((s_1, x_1), ..., (s_k, x_k)) &\rightarrow Y_t(\omega) = \sum_{i=1}^{k} x_i \mathbf{1}_{[s_i, \infty]}(t) \end{aligned}$$

the family $(Y_t)_{t\geq 0}$ is a compound Poisson process. Then, note that

$$\begin{aligned} &\partial_\lambda E(\mathbf{1}_{\{Y_t\in A\}}) \\ = \; & t\sum_{n=0}^{\infty} \frac{e^{-\lambda t}\,(\lambda t)^n}{n!}(P(X_1+...+X_{n+1}\in A)-P(X_1+...+X_n\in A)) \\ = \; & tE(\sum_{n=0}^{\infty} \frac{e^{-\lambda t}\,(\lambda t)^n}{n!}(P(X_1+...+X_n+x\in A)-P(X_1+...+X_n\in A))_{x=X_{n+1}}) \\ = \; & \int_0^T E(\Psi_{s,x}(\mathbf{1}_{\{Y_t\in A\}})_{x=X_{n+1}})ds = E\left(\int_{\mathbb{R}}\int_0^T \Psi_{s,x}(\mathbf{1}_{\{Y_t\in A\}})f(x)dxds\right), \end{aligned}$$

where for any random variable $\mathbf{F}$, we let $\Psi_{s,x}(\mathbf{F})(\omega) = \mathbf{F}(\omega_{s,x}) - \mathbf{F}(\omega)$, where $\omega_{s,x}$ is a trajectory like ω but where we add a jump of amplitude x at time s. We also can write

$$\partial_\lambda E(\mathbf{1}_{\{Y_t\in A\}}) = \frac{1}{\lambda}E\left(\int_{\mathbb{R}}\int_0^T \Psi_{s,x}(\mathbf{1}_{\{Y_t\in A\}})\nu(dx)ds\right),$$

where $\nu(dx) = \lambda f(x)dx$ is the Lévy measure of the compound Poisson process Y. If we denote by δ the adjoint operator of $\Psi_{\cdot,\cdot}$ we deduce that

$$\delta(1) = N_T - \lambda T.$$

It can be seen that if $u_{s,x}$ is a predictable random field on $[0,T]\times\mathbb{R}$ belonging to the domain of the operator δ then

$$\delta(u) = \int_{\mathbb{R}}\int_0^T u_{s,x}\tilde{M}(dx, ds),$$

where $\tilde{M}$ is the compensated random measure associated with the compound Poisson process (we assume here that $\int_{|x|>1} x\nu(dx) < \infty$), see [8] for the Poisson process and [12] for the general case, note that the authors use the operator $\frac{\Psi_{s,x}}{x}$ instead of $\Psi_{s,x}$.

4.2.2. Processes with different amplitude parameter. Let

$$L_i := Y_{t_i} - Y_{t_{i-1}} = \int_{t_{i-1}}^{t_i} \int_{\mathbb{R}} g(\theta, x) M(dx, ds),$$

$t_i = i\Delta_n$, $i = 1, ...n$, with

$$M(dx, ds) = \sum_{k=1}^{\infty} \delta_{\{\tau_k\}}(ds) \delta_{\{X_k\}}(dx),$$

where $0 < \tau_1 < \tau_2 < ... < \tau_n$ is a sequence of jump times of a Poisson process with intensity one and $X_{i,} i = 1, ..., n, ...$ is an iid sequence independent of $(\tau_i)_{i \geq 1}$ and with density f, $g(\theta, x)$ is a deterministic function. Note that the intensity of the Lévy process is assumed to be known and the inference is on the amplitude of the jumps. The compound Poisson process associated with $M(dx, ds)$ is

$$Z_t = \int_0^t \int_{\mathbb{R}} x M(dx, ds).$$

Note that when Z jumps x units, $Y_t = \int_0^t \int_{\mathbb{R}} g(\theta, x) M(dx, ds)$ jumps $g(\theta, x)$ but with same frequency as Z. In fact, Y is a compound Poisson process with random Poisson measure

$$M^Y(dx, ds) = \sum_{k=1}^{\infty} \delta_{\{\tau_k\}}(ds) \delta_{\{g(\theta, X_k)\}}(dx)$$

In such a situation we have for a regular statistic $T \equiv T(L_1, ..., L_n)$

$$\partial_\theta E(T) = \sum_{j=1}^{n} E(\partial_{l_j} T \partial_\theta L_j),$$

and

$$\begin{aligned} \partial_\theta L_j &= \int_{t_{j-1}}^{t_j} \int_{\mathbb{R}} \partial_\theta g(\theta, x) M(dx, ds) \\ &= \sum_{k=1}^{\infty} \partial_\theta g(\theta, X_k) \mathbf{1}_{(\tau_k \in (t_{j-1}, t_j)\}}, \end{aligned}$$

so

$$\begin{aligned}\partial_\theta E(T) &= \sum_{j=1}^{n}\sum_{k=1}^{\infty} E(\partial_{l_j} T \partial_\theta g(\theta, X_k)\mathbf{1}_{(\tau_k\in(t_{j-1},t_j)\}})\\ &= \sum_{j=1}^{n}\sum_{k=1}^{\infty} E(\partial_{l_j} T \partial_{x_k} L_j \partial_\theta g(\theta, X_k)\frac{1}{\partial_x g(\theta, X_k)}\mathbf{1}_{(\tau_k\in(t_{j-1},t_j)\}}).\end{aligned}$$

Since

$$\partial_{x_k} L_j = \partial_x g(\theta, X_k)\mathbf{1}_{(\tau_k\in(t_{j-1},t_j)\}}$$

we can write

$$\begin{aligned}\partial_\theta E(T) &= \sum_{j=1}^{n}\sum_{k=1}^{\infty} E(\partial_{l_j} T \partial_{x_k} L_j \partial_\theta g(\theta, X_k)\frac{1}{\partial_x g(\theta, X_k)}\mathbf{1}_{(\tau_k\in(t_{j-1},t_j)\}})\\ &= \sum_{j=1}^{n}\sum_{k=1}^{\infty} E(\partial_{x_k} T \frac{\partial_\theta g(\theta, X_k)}{\partial_x g(\theta, X_k)}\mathbf{1}_{(\tau_k\in(t_{j-1},t_j)\}})\\ &= -\sum_{j=1}^{n}\sum_{k=1}^{\infty} E\left(T\frac{1}{f(X_k)}\partial_x\left(\frac{\partial_\theta g(\theta, X_k)f(X_k)}{\partial_x g(\theta, X_k)}\right)\mathbf{1}_{(\tau_k\in(t_{j-1},t_j)\}}\right),\end{aligned}$$

where in the last inequality we use the independence between $(X_k)_k$ and the jump times and where we assume there are not "border" effects. That is,

$$T\frac{\partial_\theta g(\theta, x)f(x)}{\partial_x g(\theta, x)}\bigg|_{x\in\partial\mathrm{supp}(f)} = 0.$$

Finally

$$\begin{aligned}\partial_\theta E(T) &= -\sum_{j=1}^{n}\sum_{k=1}^{\infty} E\left(T\frac{1}{f(X_k)}\partial_x\left(\frac{\partial_\theta g(\theta, x)f(x)}{\partial_x g(\theta, x)}\right)_{x=X_k}\mathbf{1}_{(\tau_k\in(t_{j-1},t_j)\}}\right)\\ &= -E\left(T\sum_{j=1}^{n}\int_{\mathbb{R}}\int_{t_{j-1}}^{t_j}\frac{1}{f(x)}\partial_x\left(\frac{\partial_\theta g(\theta, x)f(x)}{\partial_x g(\theta, x)}\right)M(dx, ds)\right)\\ &= -E\left(T\int_{\mathbb{R}}\int_{t_0}^{t_n}\frac{1}{f(x)}\partial_x\left(\frac{\partial_\theta g(\theta, x)f(x)}{\partial_x g(\theta, x)}\right)M(dx, ds)\right).\end{aligned}$$

So, the score function is given by

$$-E\left(\int_{\mathbb{R}}\int_{t_0}^{t_n}\frac{1}{f(x)}\partial_x\left(\frac{\partial_\theta g(\theta,x)f(x)}{\partial_x g(\theta,x)}\right)M(dx,ds)\middle| Y_{t_1},Y_{t_2},...,Y_{t_n}\right).$$

If we define the random variables $V:=g(\theta,X)$ the previous expression can be written as

$$-E\left(\int_{\mathbb{R}}\int_{t_0}^{t_n}\frac{\partial_v\partial_\theta v f_V(v)}{f_V(v)}M(dv,ds)\middle| Y_{t_1},Y_{t_2},...,Y_{t_n}\right),$$

and we can compare it with expression (2.4).

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] M.H.A. Davis and P. Johansson, *Malliavin Monte Carlo Greeks for jump diffusions.* Stoch. Proc. Appl. 116 (2006), 101-129.

[4] V. Debelley and N. Privault, *Sensitivity Analysis of European options in jump-diffusiom models via Malliavin calculus on the Wiener space.* Preprint 2006.

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

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

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

[8] D. Nualart and J. Vives, *A duality formula on the Poisson space and some applications.* In R. Dalang, M. Dozzi, and F. Russo, editors, Seminar on Stochastic Analysis, Random Fields and Applications(Ascona, 1993), volume 36 of Progress in Probability, pages 205–213. Birkhäuser, 1995.

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

[10] 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.

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

[12] J.L. Solé, F. Utzet and J. Vives *Canonical Lévy Process and Malliavin Calculus.* Stochastic Processes and their Applications, **117**(2),(2007),165–187.

José M. Corcuera
Universitat de Barcelona
Gran Via de les Corts Catalanes. 585
E-08007, Barcelona
Spain
e-mail: jmcorcuera@ub.edu

Arturo Kohatsu-Higa
Osaka University
Graduate School of Engineering Sciences
Machikaneyama cho. 1-3
Osaka 560-8531
Japan.
e-mail: kohatsu@sigmath.es.osaka-u.ac.jp