Journal of Multivariate Analysis 154 (2017) 135–146

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Functional Cramér–Rao bounds and Stein estimators in Sobolev spaces, for Brownian motion and Cox processes

Eni Musta [a], Maurizio Pratelli [b], Dario Trevisan [b,*]

[a] Delft University of Technology, Netherlands

[b] Università degli Studi di Pisa, Italy

ARTICLE INFO

Article history:
Received 20 July 2015
Available online 29 October 2016

AMS 2010 subject classifications:
primary 62J07
secondary 60J65

Keywords:
Cramer–Rao bound
Stein phenomenon
Malliavin calculus
Cox model

ABSTRACT

We investigate the problems of drift estimation for a shifted Brownian motion and intensity estimation for a Cox process on a finite interval $[0, T]$, when the risk is given by the energy functional associated to some fractional Sobolev space $H_0^1 \subset W^{\alpha,2} \subset L^2$. In both situations, Cramér–Rao lower bounds are obtained, entailing in particular that no unbiased estimators (not necessarily adapted) with finite risk in H_0^1 exist. By Malliavin calculus techniques, we also study super-efficient Stein type estimators (in the Gaussian case).

1. Introduction

In this paper, we focus on two problems of nonparametric (or, more rigorously, infinite-dimensional parametric) statistical estimation: drift estimation for a shifted Brownian motion and intensity estimation for a Cox process, on a finite time interval $[0, T]$. Our investigation stems from the articles [10,11], where N. Privault and A. Réveillac developed an original approach to these problems, by employing techniques from Malliavin calculus to study Cramér–Rao bounds and super-efficient "shrinkage" estimators, originally developed by C. Stein in [5] and then expanded in [13], to fit in infinite-dimensional frameworks. Such a combination of these two powerful techniques can be cast into a more general picture, where Malliavin calculus tools provide insights in statistics and more generally, on probabilistic approximations: let us mention here the monograph [8], which collects many results of the fruitful meeting of another great contribution of C. Stein (the so-called Stein method) with Malliavin calculus, and other recent articles such as [2,4,7,12].

As in [10,11], here we assume that the unknown function to be estimated belongs to the Hilbert space $H_0^1(0, T)$ (which is a reasonable choice, at least in the case of shifted Brownian motion, because of the Cameron–Martin and Girsanov theorems) but we move further by addressing the following question, which is rather natural but has apparently not yet been considered: what about estimators that also take values in H_0^1? Indeed, in [10,11], estimators are seen as functions with values in $L^2([0, T], \mu)$ (where μ is any finite measure) or, equivalently, the associated risk is computed with respect to the L^2 norm and not the (stronger) H_0^1 norm.

To investigate this problem, we first provide Cramér–Rao bounds with respect to different risks, by considering the estimation in the interpolating fractional Sobolev space $H_0^1 \subset W^{\alpha,2} \subset L^2$, for $\alpha \in [0, 1]$. It turns out that no unbiased

* Corresponding author.
E-mail addresses: e.musta@tudelft.nl (E. Musta), pratelli@dm.unipi.it (M. Pratelli), dario.trevisan@unipi.it (D. Trevisan).

http://dx.doi.org/10.1016/j.jmva.2016.10.011
0047-259X/

estimator exists in H_0^1 (Theorem 2.5) and even in $W^{\alpha,2}$, for $\alpha \geq 1/2$ (Theorem 2.9). Although a bit surprising, these results reconcile with the following intuition: since the estimator is a function of the realization of the process, whose paths also do not belong to H_0^1 (nor $W^{\alpha,2}$, for $\alpha \geq 1/2$), it is "too risky" to estimate (without bias) the parameter on that scale of regularity. Therefore, besides answering a rather natural question, our results highlight the delicate role played by the choice of different norms in such estimation problems, and one might expect that similar phenomena might appear in other situations, technically more demanding, e.g., stochastic differential equations.

As a second task, we study super-efficient "shrinkage" estimators in the spaces $W^{\alpha,2}$. It is often suggested on heuristic grounds that the ideal situation for the problem of estimation would be to have an unbiased estimator with low variance, but that allowing for a little bias may allow one to find estimators with lower risks, in many situations: we strongly rely on the recent extensions and combinations of the original approach by Stein with Malliavin calculus to these frameworks developed in [10,11]. Using a similar approach, we give sufficient conditions for the existence of super-efficient estimators in $W^{\alpha,2}$, for $\alpha < 1/2$, and we give explicit examples of such estimators, in the case of Brownian motion (Example 3.3). In the case of Cox processes, although it is possible to define a suitable version of Malliavin calculus and provide sufficient conditions for Stein estimators, we are currently unable to provide explicit examples.

The paper is organized as follows. In Section 2 we deal with drift estimation for a shifted Brownian motion, addressing Cramér–Rao lower bounds with respect to risks computed in H_0^1 and fractional Sobolev spaces. In Section 3, we discuss super-efficient estimators. Finally, analogous results on intensity estimators for Cox processes are given in Section 4.

2. Drift estimation for a shifted Brownian motion

In this section, we fix $T \geq 0$ and let $X = (X_t)_{t\in[0,T]}$ be a Brownian motion (on the finite interval $[0, T]$), defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$. Instead of choosing a fixed (infinite-dimensional) space of parameters Θ, we simply notice that our arguments apply to any set Θ of absolutely continuous, adapted processes $u_t = \int_0^t \dot{u}_s ds$ (for $t \in [0, T]$) such that

(1) $(\dot{u}_t)_{t\in[0,T]}$ satisfies the conditions of Girsanov's theorem;
(2) Θ contains the Cameron–Martin space H_0^1;
(3) for any $u \in \Theta$, $v \in H_0^1$, one has $u + v \in \Theta$.

Let us recall that H_0^1 $(= H_0^1(0, T))$ is defined as the space of (continuous) functions of the form $h(t) = \int_0^t \dot{h}(s)ds$, for $t \in [0, T]$, with $\dot{h} \in L^2(0, T)$. In particular, we may let $\Theta = H_0^1$.

For $u \in \Theta$, we define the probability measure $P^u = L^u P$, with

$$L^u = \exp\left(\int_0^T \dot{u}_s dX_s - \frac{1}{2}\int_0^T \dot{u}_s^2 ds\right).$$

Girsanov's theorem entails that, with respect to the probability measure P^u, the process $X_t^u = X_t - u_t$ is a Brownian motion on $[0, T]$.

We address the problem of estimating the drift with respect to P^u on the basis of a single observation of X (of course, repeated and independent observations can improve the estimates, but this amounts to a simple generalization). Such a problem is of interest in different fields of application: for example, we can interpret X as the observed output signal of some unknown input signal u, perturbed by a Brownian noise. Such a problem is investigated, e.g., in [10], where the following definition is given.

Definition 2.1. Any measurable stochastic process $\xi : \Omega \times [0, T] \to \mathbb{R}$ is called an estimator of the drift u. An estimator of the drift u is said to be unbiased if, for every $u \in \Theta$, $t \in [0, T]$, ξ_t is P^u-integrable and one has $\mathrm{E}^u(\xi_t) = \mathrm{E}^u(u_t)$.

In this section, we forego the specification of "the drift u" and simply refer to estimators. Moreover, we refer to the quantity $\mathrm{E}^u(\xi_t - u_t)$ as the bias of the estimator ξ (whenever it is well-defined).

By introducing as a risk associated to any estimator ξ the quantity

$$\mathrm{E}^u(\|\xi - u\|^2_{L^2(\mu)}) = \mathrm{E}^u\left\{\int_0^T |\xi_t - u_t|^2 \mu(dt)\right\}, \tag{1}$$

where μ is any finite Borel measure on $[0, T]$, Privault and Réveillac provide the Cramér–Rao lower bound stated next for adapted and unbiased estimators [10, Proposition 2.1]. In what follows, Θ being the space of all absolutely continuous, adapted processes, whose derivatives satisfy the conditions of Girsanov's theorem.

Theorem 2.2 (*Cramér–Rao Inequality in $L^2(\mu)$*). *For any adapted and unbiased estimator ξ, one has*

$$\mathrm{E}^u(\|\xi - u\|^2_{L^2(\mu)}) \geq \int_0^T t\mu(dt), \quad \text{for every } u \in \Theta. \tag{2}$$

Equality is attained by the (efficient) estimator $\hat{u} = X$.

Before giving our results, we briefly report the original proof in [10] but observe that the requirement made therein to the effect that ξ is adapted is actually unnecessary.

Proof. The inequality follows from an application of the Cauchy–Schwarz inequality to the crucial identity

$$v(t) = \mathrm{E}^u \left\{ (\xi_t - u_t) \int_0^T \dot{v}(s) dX_s^u \right\}, \quad \text{for } t \in [0, T], \tag{3}$$

valid for every *deterministic* process $v \in \Theta$ (thus, $v(t) = \int_0^t \dot{v}(s)ds$). Indeed, if we choose, for any $t \in [0, T]$, $\dot{v}(s) = \mathbf{1}_{[0,t]}(s)$, then $v(t) = t$ and $\int_0^T \dot{v}(s)dX_s^u = X_t^u$. We obtain, from (3),

$$t = \mathrm{E}^u \left\{ (\xi_t - u_t) X_t^u \right\} \leq \mathrm{E}^u \left\{ (\xi_t - u_t)^2 \right\}^{1/2} \mathrm{E}^u \{ (X_t^u)^2 \}^{1/2} = \mathrm{E}^u \{ (\xi_t - u_t)^2 \}^{1/2} \sqrt{t},$$

since X^u is a Brownian motion under P^u. After dividing by $\sqrt{t}$ and squaring on both sides, we integrate with respect to μ for $t \in [0, T]$, to obtain (2).

In turn, to prove (3) we use the fact that, for every $\varepsilon \in \mathbb{R}$, one has $u + \varepsilon v \in \Theta$, and hence

$$\mathrm{E}^{u+\varepsilon v}(\xi_t) = \mathrm{E}^{u+\varepsilon v}\{u_t + \varepsilon v(t)\} = \mathrm{E}^{u+\varepsilon v}(u_t) + \varepsilon v(t), \quad \text{for } t \in [0, T].$$

We then differentiate with respect to ε, at $\varepsilon = 0$. Exchanging between differentiation and expectation is justified by the finiteness of the left-hand side in (2), for μ-a.e. $t \in [0, T]$; otherwise there is nothing to prove. We obtain

$$\begin{aligned} \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \mathrm{E}^{u+\varepsilon v}(\xi_t - u_t) &= \mathrm{E} \left\{ (\xi_t - u_t) \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} L_T^{u+\varepsilon v} \right\} \\ &= \mathrm{E}^u \left\{ (\xi_t - u_t) \int_0^T \dot{v}(s) dX_s^u \right\}. \quad \square \end{aligned}$$

Remark 2.3. Once again, let us stress the fact that in the above proof, ξ need not be adapted. Concerning the issue of comparing adapted with non-adapted estimators, it would be desirable to argue that general (not necessarily adapted) estimators cannot perform better than adapted ones, and the following argument might seem to go in that direction. However, it does not allow us to conclude. Let ξ be any unbiased estimator and for $u \in \Theta$, consider the optional projection η of ξ, with respect to the probability P^u, so that $\eta_t = \mathrm{E}^u(\xi_t|\mathcal{F}_t)$, for $t \in [0, T]$. Then $\mathrm{E}^u(\eta_t) = u_t$ and one has

$$\mathrm{E}^u(|\eta_t - u_t|^2) = \mathrm{E}^u \left\{ \mathrm{E}^u(\xi_t - u_t|\mathcal{F}_t)^2 \right\} \leq \mathrm{E}^u(|\xi_t - u_t|^2).$$

However, this does not entail that η performs better that ξ, since $\eta = \eta^u$ depends also on u; thus it is not an estimator. Note, however, that if we keep $\bar{u} \in \Theta$ fixed, then $\eta^{\bar{u}}$ could be biased, i.e., $\mathrm{E}^u(\eta_t^{\bar{u}}) \neq \mathrm{E}^u(u_t)$ for some $u \in \Theta$, $t \in [0, T]$. A similar issue appears in [10].

Remark 2.4. Beyond the mean squared error, one can consider the risk defined by L^p norms, for $p \in (1, \infty)$, viz.

$$\int_0^T \mathrm{E}^u(|\xi_t - u_t|^p)\mu(dt).$$

Again, by direct inspection of the proof in [10], applying Hölder's inequality (with conjugate exponents (p, q)) instead of the Cauchy–Schwarz inequality in (3), we obtain an inequality of the form

$$\mathrm{E}^u(|\xi_t - u_t|^p) \geq \frac{|v(t)|^p}{c_q^{p/q} \left\{ \int_0^t \dot{v}^2(s)ds \right\}^{p/2}} \geq \frac{1}{c_q^{p/q}} t^{p/2}, \quad \text{for } t \in [0, T],$$

where $c_q = \mathrm{E}(|Y|^q)$ is the qth moment of a $\mathcal{N}(0, 1)$ random variable Y. Integration with respect to μ then provides a Cramér–Rao type lower bound. However, letting $\xi = X$, one has

$$\mathrm{E}^u(|X_t - u_t|^p) = \mathrm{E}^u(|X_t^u|^p) = c_p t^{p/2}, \quad \text{for } t \in [0, T].$$

Thus X is not an efficient estimator in $L^p(\Omega \times [0, T])$ for $p \neq 2$.

Let us recall that the Cameron–Martin space H_0^1 is a Hilbert space, endowed with the norm induced by the natural Sobolev "energy" functional, namely $\|h\|_{H_0^1} = \|\dot{h}\|_{L^2(0,T)}$. For simplicity of notation, we extend such a functional identically to $+\infty$ for any Borel curve $h : [0, T] \to \mathbb{R}$ that does not belong to H_0^1.

We observe that H_0^1 is continuously included in $\mathcal{C}^{1/2}(0, T)$, the space of 1/2-Hölder continuous functions: since the paths of the Brownian motion are not 1/2-Hölder continuous, we deduce that the process X is not H_0^1-valued (negligibility of the Cameron–Martin space holds true also for abstract, infinite-dimensional, Wiener spaces). However, since the drift u takes values in H_0^1, it is natural to look for an estimator ξ sharing this property. Our first result shows that, if we require ξ to be unbiased, this is not possible, i.e., such an estimator ξ has necessarily infinite H_0^1 risk.

Theorem 2.5 (*Estimators in H_0^1*)**.** *Let ξ be an estimator such that, for some $u \in \Theta$, one has $\mathrm{E}^u(\|\xi - u\|^2_{H_0^1}) < \infty$. Then ξ is not unbiased (in particular, the bias in not zero at u).*

Before we address the proof for general, possibly non-adapted estimators, we give the following argument that exploits Itô's formula. Actually it is longer, but we feel that it has more of a stochastic flavor.

Proof (*Case of Adapted Estimators*)**.** Arguing by contradiction, we assume that ξ is unbiased and the risk at u is finite, i.e., $\xi - u \in L^2(\Omega, P^u; H_0^1)$. For every (deterministic) $v \in H_0^1$, arguing exactly as above for the deduction of (3), we obtain

$$v(t) = \mathrm{E}^u \left\{ \int_0^t (\dot{\xi}_s - \dot{u}_s) ds \int_0^t \dot{v}(s) dX_s^u \right\}, \quad \text{for } t \in [0, T],$$

where stochastic integration reduces to the interval $[0, t]$ because of the adaptedness assumption. Integrating by parts (i.e., using Itô's formula) we rewrite the random variable above as

$$\int_0^t \left\{ \int_0^s \dot{v}(r) dX_r^u \right\} (\dot{\xi}_s - \dot{u}_s) ds + \int_0^t \left\{ \int_0^s (\dot{\xi}_r - \dot{u}_r) dr \right\} \dot{v}(s) dX_s^u.$$

The Itô integral has zero expectation, since $\xi - u \in L^2(\Omega, P^u; H_0^1) \subseteq L^2(\Omega, P^u; \mathcal{C}^{1/2}(0, T))$ and $\dot{v} \in L^2(0, T)$, hence the integrand is an adapted and square-integrable process. Therefore, taking expectation, we obtain the analogue of (3) for the study of H_0^1 energy:

$$v(t) = \mathrm{E}^u \left[\int_0^t \left\{ \int_0^s \dot{v}(r) dX_r^u \right\} (\dot{\xi}_s - \dot{u}_s) ds \right], \quad \text{for } t \in [0, T].$$

Indeed, the Cauchy–Schwarz inequality and Itô's isometry give

$$\begin{aligned} v(t)^2 &\le \mathrm{E}^u \left[\int_0^t \left\{ \int_0^s \dot{v}(r) dX_r^u \right\}^2 ds \right] \mathrm{E}^u \left\{ \int_0^t (\dot{\xi}_s - \dot{u}_s)^2 ds \right\} \\ &= \int_0^t \left\{ \int_0^s \dot{v}^2(r) dr \right\} ds \int_0^t \mathrm{E}^u \{ (\dot{\xi}_s - \dot{u}_s)^2 \} ds \\ &= \int_0^t (t-s) \dot{v}^2(s) ds \int_0^t \mathrm{E}^u \{ (\dot{\xi}_s - \dot{u}_s)^2 \} ds. \end{aligned}$$

In particular, choosing $t = T$, we deduce

$$\mathrm{E}^u(\|\xi - u\|^2_{H_0^1}) \ge \frac{v(T)^2}{\int_0^T (T-t) \dot{v}^2(t) dt}.$$

To reach a contradiction, it is enough to prove that for every constant $c > 0$, there exists $\dot{v} \in L^2(0, T)$ such that the left-hand side above is greater than c, i.e.,

$$\left\{ \int_0^T \dot{v}(t) dt \right\}^2 \ge c \int_0^T (T-t) \dot{v}(t)^2 dt. \tag{4}$$

Indeed, if we let $\dot{v}(t) = (T-t)^{-\alpha}$ for some $0 < \alpha < 1$, we get

$$\left\{ \int_0^T \dot{v}(t) dt \right\}^2 = \left(\frac{T^{1-\alpha}}{1-\alpha} \right)^2 \quad \text{and} \quad \int_0^T (T-t) \dot{v}^2(t) dt = \frac{T^{2(1-\alpha)}}{2(1-\alpha)}.$$

It is then sufficient to let $\alpha \uparrow 1$ to conclude. □

Remark 2.6. Instead of the explicit construction of $v \in H_0^1$ above, to obtain a contradiction we can also use the following duality result. On a measure space $(E, \mathcal{E}, \mu)$, suppose that $g \ge 0$ is a measurable function such that, for some constant $c > 0$, the following condition holds:

$$\int_E fg d\mu \le c \left(\int_E f^2 d\mu \right)^{1/2}, \quad \text{for every } f \in L^\infty(\mu), f \ge 0.$$

Then $g \in L^2(\mu)$ with $\|g\|_{L^2(\mu)} \le c$. The easy proof follows from considering the continuous, linear functional ϕ initially defined on $L^\infty \cap L^2(\mu)$ by $f \mapsto \int_E fg d\mu$ and then applying Riesz's theorem on its extension to $L^2(\mu)$. In the proof above, a contradiction immediately follows from (4), letting $\mu(dt) = (T-t)dt$ and $g(t) = (T-t)^{-1}$.

We now provide a complete proof of Theorem 2.5.

Proof (*General Case*)**.** Arguing by contradiction, we assume that ξ is unbiased and the risk at u is finite, i.e., $\xi - u \in L^2(\Omega, P^u; H_0^1)$. For every (deterministic) $v \in H_0^1$, arguing as above for the deduction of (3), we obtain instead

$$v(t) = \mathrm{E}^u \left\{ \int_0^t (\dot{\xi}_s - \dot{u}_s) ds \int_0^T \dot{v}(s) dX_s^u \right\}, \quad \text{for } t \in [0, T].$$

Then we differentiate with respect to $t \in [0, T]$ (exchanging derivatives and expectation is ensured by the finite risk assumption), and we obtain, for a.e. $t \in [0, T]$,

$$\dot{v}(t) = \mathrm{E}^u \left\{ (\dot{\xi}_t - \dot{u}_t) \int_0^T \dot{v}(s) dX_s^u \right\}.$$

At this stage, the Cauchy–Schwarz inequality and Itô's isometry together yield

$$|\dot{v}(t)|^2 \leq \mathrm{E}^u \left(|\dot{\xi}_t - \dot{u}_t|^2 \right) \int_0^T |\dot{v}(s)|^2 ds, \quad \text{for a.e. } t \in [0, T]. \tag{5}$$

From this inequality, we easily obtain a contradiction, arguing as follows. Let $A \subseteq [0, T]$ be a non-negligible Borel subset such that $\int_A \mathrm{E}^u(|\dot{\xi}_t - \dot{u}_t|^2) dt < 1$, which exists because of the finite risk assumption and uniform integrability (notice that A does not depend upon v). Then, integrating the above inequality for $t \in A$, we obtain

$$\int_A |\dot{v}(t)|^2 dt \leq \int_A \mathrm{E}^u(|\dot{\xi}_t - \dot{u}_t|^2) dt \int_0^T |\dot{v}(t)|^2 dt,$$

for every $\dot{v} \in L^2(0, T)$, in particular for every $\dot{v} \in L^2(A)$. Simply taking $\dot{v} = \mathbf{1}_A$, we obtain the required contradiction. □

Actually, the result on the absence of unbiased estimators in H_0^1 can be slightly strengthened, allowing for estimators whose bias is sufficiently regular. We state it as a corollary (of the proof), remarking that similar deductions could be performed also in the cases that we consider below.

Corollary 2.7. *Let ξ be an estimator such that, for every $u \in \Theta$, $t \in [0, T]$, ξ_t is P^u-integrable, and one has, for some $C = (C_t)_{t \in [0,T]} \in L^2(0, T)$ (possibly depending upon $u \in \Theta$),*

$$\left| \frac{d}{dt} \frac{d}{d\varepsilon} \bigg|_{\varepsilon=0} \mathrm{E}^{u+\varepsilon v} \{ \xi_t - u_t - \varepsilon v(t) \} \right| \leq C_t \|v\|_{L^2(0,T)}, \quad \textit{a.e. } t \in [0, T], \textit{ for every } v \in H_0^1.$$

Then the H_0^1 risk of the estimator ξ is infinite, i.e.,

$$\mathrm{E}^u(\|\xi - u\|_{H_0^1}^2) ds = \infty, \quad \textit{for every } u \in \Theta.$$

Proof. We argue exactly as in the proof above, but we write

$$\mathrm{E}^{u+\varepsilon v}(\xi_t) = \mathrm{E}^{u+\varepsilon v}(u_t) + \varepsilon v(t) + b_t^{u+\varepsilon v}$$

where $b_t^u = \mathrm{E}^u(\xi_t - u_t)$ is the bias. After differentiation with respect to ε and t, we obtain (5) with $\mathrm{E}^u(|\dot{\xi}_t - \dot{u}_t|^2) + C_t^2$ in place of $\mathrm{E}^u(|\dot{\xi}_t - \dot{u}_t|^2)$ and we conclude arguing as in the proof above. □

We address now analogous results for the intermediate spaces $H_0^1 \subset W^{\alpha,2} \subset L^2$, for $\alpha \in (0, 1)$, defined as follows.

Definition 2.8. For $\alpha \in (0, 1)$, $p \in (1, \infty)$, the fractional Sobolev space $W^{\alpha,p} (= W^{\alpha,p}(0, T))$ is defined as the space of functions $u \in L^p(0, T)$ such that their "energy" functional is finite, i.e.,

$$\|u\|_{W_0^{\alpha,p}}^p = \int_0^T \int_0^T \frac{|u_t - u_s|^p}{|t - s|^{p\alpha+1}} dt ds < \infty.$$

The notation $W_0^{\alpha,p}$, with subscript 0, is introduced here to distinguish the energy functional from the usual norm in the theory of fractional Sobolev spaces, for which we refer throughout to the survey [3]. For our purposes, we need nothing more than the definition above, but let us stress some further (well-known) facts. The space $W^{\alpha,p}$ (endowed with a suitable norm) interpolates between the Sobolev space $W^{1,p}$ and L^p; for example, one has $W^{\alpha',p} \subseteq W^{\alpha,p}$ for $0 < \alpha \leq \alpha' < 1$, and $W^{\alpha,2} \subseteq H^1$, with

$$\|u\|_{W_0^{\alpha,2}}^2 \leq 2 \int_0^T |\dot{u}_r|^2 \int_r^T \int_0^r \frac{1}{|t - s|^{2\alpha}} ds dt dr \leq C_{\alpha,T} \|u\|_{H_0^1}^2. \tag{6}$$

From this inequality, the above theorem for estimators in H_0^1 could also be obtained from the next results. Moreover, if $\alpha p > 1$, then one can prove a continuous embedding of $W^{\alpha,p}(0,T)$ into $\mathcal{C}^{\beta}(0,T)$, with $\beta = \alpha - 1/p$.

Let us first consider the Cramér–Rao bound in the quadratic case.

Theorem 2.9 (*Cramér–Rao Inequality in* $W^{\alpha,2}$). *Let ξ be an unbiased estimator. For every $\alpha \in (0,1)$, one has*

$$\mathrm{E}^u\left(\|\xi - u\|^2_{W_0^{\alpha,2}}\right) \geq \int_0^T \int_0^T \frac{1}{|t-s|^{2\alpha}}\, dtds, \quad \text{for every } u \in \Theta.$$

Equality is attained by the (efficient) estimator $\xi = X$.

In particular, if an estimator ξ has finite $W^{\alpha,2}$ risk for some $\alpha \in [1/2, 1)$ and some $u \in \Theta$, then it is not unbiased. This is consistent with the qualitative and informal fact that the paths of Brownian motion do not possess "half of a derivative" in time, even measured in a L^2 sense.

Proof. We introduce the notation $\Delta_t = \xi_t - u_t$, for $t \in [0,T]$, so that, by Fubini's theorem, we write

$$\mathrm{E}^u\left(\|\xi - u\|^2_{W_0^{\alpha,2}}\right) = \int_0^T \int_0^T \frac{\mathrm{E}^u(|\Delta_t - \Delta_s|^2)}{|t-s|^{2\alpha+1}}\, dtds.$$

If ξ is an unbiased estimator and $v \in H_0^1$, we argue (once again) to obtain (3), and subtract the corresponding identities for $s, t \in [0,T]$, thus

$$v(t) - v(s) = \mathrm{E}^u\left\{(\Delta_t - \Delta_s)\int_0^T \dot{v}(r)dX_r^u\right\}.$$

Hence, the Cauchy–Schwarz inequality and Itô's isometry give the lower bound

$$\mathrm{E}^u(|\Delta_t - \Delta_s|^2) \geq \frac{|v(t) - v(s)|^2}{\int_0^T \dot{v}^2(s)ds}, \quad \text{for } s, t \in [0,T].$$

We let $\dot{v}(r) = \mathbf{1}_{[s\wedge t, s\vee t]}(r)$, so that

$$\mathrm{E}^u(|\Delta_t - \Delta_s|^2) \geq |t-s| \quad \text{for } s, t \in [0,T].$$

The Cramér–Rao bound then follows, viz.

$$\int_0^T \int_0^T \frac{\mathrm{E}^u(|\Delta_t - \Delta_s|^2)}{|t-s|^{2\alpha+1}}\, dtds \geq \int_0^T \int_0^T \frac{1}{|t-s|^{2\alpha}}\, dtds.$$

Finally, if $\xi = X$, then $X - u = X^u$, thus one has

$$E^u(|X_t^u - X_s^u|^2) = |t-s|, \quad \text{for } s, t \in [0,T].$$

Hence the Cramér–Rao lower bound is attained, i.e.,

$$\int_0^T \int_0^T \frac{\mathrm{E}^u(|X_t^u - X_s^u|^2)}{|t-s|^{2\alpha+1}}\, dtds = \int_0^T \int_0^T \frac{1}{|t-s|^{2\alpha}}\, dtds. \quad \square$$

In the case of a general exponent $p \in (1,\infty)$ (with $q = p/(p-1)$), arguing similarly, we obtain the following bound, in $W^{\alpha,p}$. As above, we let $c_q = \mathrm{E}(|Y|^q)$ be the qth moment of a standard Gaussian (Normal) random variable.

Theorem 2.10 (*Cramér–Rao Inequality in* $W^{\alpha,p}$). *Let ξ be an unbiased estimator. For every $\alpha \in (0,1)$, $p \in (1,\infty)$, one has*

$$\mathrm{E}^u\left(\|\xi - u\|^p_{W_0^{\alpha,p}}\right) \geq \frac{1}{c_q^{p/q}} \frac{2T^{1-p\alpha+p/2}}{p\{1 + p(1/2 - \alpha)\}\max\{0, (1/2-\alpha)\}}.$$

Since

$$\mathrm{E}^u(|X_t^u - X_s^u|^p) = c_p|t-s|^{p/2},$$

the risk of the estimator $\xi = X$ is given by

$$\int_0^T \int_0^T \frac{\mathrm{E}^u(|X_t^u - X_s^u|^p)}{|t-s|^{p\alpha+1}}\, dtds = c_p \int_0^T \int_0^T \frac{1}{|t-s|^{p\alpha+1-p/2}}\, dtds.$$

As in Remark 2.4, we conclude that X is not an efficient estimator with respect to the risk in $W^{\alpha,p}$, for $p \neq 2$.

Remark 2.11. Before we conclude this section, we remark that all the bounds above can be generalized (at least) to the case of a continuous Gaussian martingale, with quadratic variation process $\int_0^t \sigma_s^2 ds$, $t \in [0, T]$ and also by introducing different energies, such as

$$\int_0^T \int_0^T \frac{|u(t) - u(s)|^p}{|t - s|^{\alpha p+1}} \mu(dt, ds),$$

where μ is a measure on $[0, T]$ (a natural choice would be to take μ somehow related to σ^2). However, we choose to restrict the discussion to the case of the Brownian motion, to limit technicalities and emphasize the role played by the norm chosen to estimate the risk.

3. Super-efficient estimators

In this section, we address the problem of Stein type, super-efficient estimators for the drift of a shifted Brownian motion, with respect to risks computed in the Sobolev spaces introduced above.

For $L^2(\mu)$-type risks, super-efficient estimators in the form $X + \xi$ were first studied in [10], using tools from Malliavin calculus. Before we discuss their approach and our extension to Sobolev spaces, let us review some facts about Malliavin calculus on the classical Wiener space (we refer to the monograph [9] for details), limiting ourselves to what is essential for our purpose.

3.1. Malliavin Calculus on the Wiener space

In the framework of Section 2, i.e., if $X = (X_t)_{t\in[0,T]}$ is a Brownian motion (on the finite interval $[0, T]$), defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$, we introduce the space $\mathcal{S}$ of smooth functionals, as those in the form $F = \phi\left(X_{t_1}, \dots, X_{t_n}\right)$ for some $t_1, \dots, t_n \in [0, T]$ and $\phi \in \mathcal{C}_b^\infty(\mathbb{R}^n)$, $n \geq 1$. The Malliavin derivative DF is then defined as the $L^2(0, T)$-valued random variable

$$D_t F = \sum_{i=1}^n \frac{\partial \phi}{\partial x_i}\left(X_{t_1}, \dots, X_{t_n}\right) \mathbf{1}_{[0,t_i]}(t), \quad \text{for a.e. } t \in [0, T].$$

For $h \in L^2(0, T)$, we let $D_h F = \int_0^T D_t F h(t) dt$ (in the classical Wiener space framework, this corresponds to differentiation along the direction in H_0^1 given by $\tilde{h}(t) = \int_0^t h(s)ds$, $t \in [0, T]$: differently from the previous sections, we prefer to focus on the space $L^2(0, T)$ instead of H_0^1). The Cameron–Martin theorem entails the following integration by parts formula for smooth functionals.

Proposition 3.1. *Let $F \in \mathcal{S}$ and $h \in L^2(0, T)$. Then*

$$\mathrm{E}(D_h F) = \mathrm{E}\left(F h^*\right), \tag{7}$$

where we let $h^* = \int_0^T h(s)dX_s$ be the Itô(–Wiener) integral.

A straightforward consequence of the integration by parts formula above is closability for the operator $D : \mathcal{S} \subset L^2(\Omega) \to L^2(\Omega \times [0, T])$. The domain of its closure defines the Sobolev–Malliavin space $\mathbb{D}^{1,2}$, on which the operator D extends continuously.

Proposition 3.2 (*Chain Rule*)**.** *Let $F_1, \dots, F_n \in \mathbb{D}^{1,2}$ and $\phi \in C_b^1(\mathbb{R}^n)$. Then $\phi(F_1, \dots, F_n) \in \mathbb{D}^{1,2}$ with*

$$D_t \phi(F_1, \dots, F_n) = \sum_{i=1}^n \frac{\partial \phi}{\partial x_i}(F_1, \dots, F_n) D_t F_i, \quad \text{for a.e. } t \in [0, T].$$

3.2. Stein's shrinkage estimators in fractional Sobolev spaces

In [10], Privault and Réveillac consider an estimator $\xi_t = D_{\mathbf{1}_{[0,t]}} \ln F$, $t \in [0, T]$, where F is any P-a.s. non-negative random variable in $\mathbb{D}^{1,2}$ such that $\sqrt{F}$ is Δ-superharmonic with respect to a suitable "Laplacian" operator, actually related to the structure of the risk considered (which is not, in the Gaussian case, the usual Gross–Malliavin Laplacian). We show that a similar approach leads to super-efficient estimators also in fractional Sobolev spaces $W^{\alpha,2}$, for $\alpha \in [0, 1/2)$ (of course, this perturbative approach does not provide any information for larger values of α). Indeed, for every $\xi = (\xi_t)_{t\in[0,T]}$, with $\mathrm{E}^u(\|\xi\|^2_{W_0^{2,\alpha}}) < \infty$, we write

$$\mathrm{E}^u(\|X + \xi - u\|^2_{W_0^{\alpha,2}}) = \mathrm{E}^u\left(\|X - u\|^2_{W_0^{\alpha,2}} + \|\xi\|^2_{W_0^{\alpha,2}}\right) + 2 \int \mathrm{E}^u\left[(\xi_t - \xi_s)\{(X_t - u_t) - (X_s - u_s)\}\right] \mu_\alpha(ds, dt),$$

where we introduce the Borel measure $\mu_\alpha(ds, dt) = 2(t-s)^{-2\alpha-1}\mathbf{1}_{\{s<t\}}dsdt$ on $[0, T]^2$. If $\xi_t - \xi_s \in \mathbb{D}^{1,2}$, for every s, $t \in [0, T]$, with $s < t$, the integration by parts (7) for the Malliavin derivative (to be rigorous, we should write in what follows D^u, because the derivative is built with respect to the probability P^u, not P), entails

$$\begin{aligned}\mathrm{E}^u\left[(\xi_t - \xi_s)\{(X_t - u_t) - (X_s - u_s)\}\right] &= \mathrm{E}^u\{(\xi_t - \xi_s)(X_t^u - X_s^u)\}\\ &= \mathrm{E}^u\{(\xi_t - \xi_s)\mathbf{1}^*_{[s,t]}\} = \mathrm{E}^u\{\tilde{D}_{s,t}(\xi_t - \xi_s)\},\end{aligned}$$

where $\tilde{D}_{s,t}F = D_{\mathbf{1}_{[s,t]}}\int_s^t D_rFdr$. Hence, if we let $\rho = \mathrm{E}^u(\|X-u\|^2_{W_0^{\alpha,2}})$ denote the Cramér–Rao lower bound, we deduce that

$$\mathrm{E}^u\big(\|X + \xi - u\|^2_{W_0^{\alpha,2}}\big) = \rho + \int \mathrm{E}^u\big\{|\xi_t - \xi_s|^2 + 2\tilde{D}_{s,t}(\xi_t - \xi_s)\big\}\mu_\alpha(ds, dt).$$

It is then convenient to introduce the following notion of Laplacian,

$$\Delta_\alpha F = \int_{[0,T]^2}(\tilde{D}_{s,t})^2 F\mu_\alpha(ds, dt), \tag{8}$$

initially defined on $\mathcal{S}$. Arguing as in [10, Proposition 4.5], it is possible to show that $\Delta_\alpha : \mathcal{S} \subseteq L^2(\Omega, P^u) \to L^2(\Omega, P^u)$ is closable and that the random variables $G \in \mathbb{D}^{1,2}$, with

$$\tilde{D}_{s,t}G \in \mathbb{D}^{1,2}, \quad \text{for a.e. } s, t \in [0, T] \quad \text{and} \quad \tilde{D}^2_{s,t}G \in L^2\left(\Omega \times [0, T]^2, P \times \mu_\alpha\right), \tag{9}$$

belong to the domain of the closure, so that $\Delta_\alpha G$ is well-defined (actually, by the same expression as in (8)). Moreover, the operator Δ_α is of diffusion type, i.e., for every $F_1, \ldots, F_n \in \mathcal{S}$, $\phi \in C_b^2(\mathbb{R}^n)$, the function $\phi \circ \mathbf{F}$ (we write $\mathbf{F} = (F_1, \ldots, F_n)$) belongs to the domain of Δ_α, and one has

$$\Delta_\alpha(\phi \circ \mathbf{F}) = \sum_{i=1}^n \frac{\partial\phi}{\partial x_i}(\mathbf{F})\Delta_\alpha F_i + \sum_{i,j=1}^n \frac{\partial^2\phi}{\partial x_i\partial x_j}(\mathbf{F})\Gamma_\alpha(F_i, F_j), \quad P\text{-a.e. in } \Omega, \tag{10}$$

with $\Gamma_\alpha(F_i, F_j) = \int_{[0,T]^2}\tilde{D}_{s,t}F_i\tilde{D}_{s,t}F_j\mu_\alpha(ds, dt)$, for all $i, j \in \{1, \ldots, n\}$ (the Malliavin matrix associated to $(F_i)_{i=1}^n$). This identity, by density, extends under natural integrability assumptions on $\mathbf{F}$ as well as on ϕ.

The operator Δ_α enters in the picture if we assume that the process ξ is of the form $\xi_t = \tilde{D}_{0,t}\ln F^2$, $t \in [0, T]$, for some P-a.e. positive random variable $F \in \mathbb{D}^{1,2}$, with $G = \ln F^2$ satisfying (9). If we are in a position to apply the chain rule (10), one then gets

$$\Delta_\alpha \ln F^2 = 2\,\frac{\Delta_\alpha F}{F} - \frac{2}{F^2}\,\Gamma_\alpha(F, F) = \frac{2\Delta_\alpha F}{F} - \frac{1}{2}\,\Gamma_\alpha(\ln F^2, \ln F^2),$$

which can be explicitly written in terms of ξ as

$$\frac{4\Delta F}{F} = \int_{[0,T]^2}\big\{2\tilde{D}_{s,t}(\xi_t - \xi_s) + |\xi_t - \xi_s|^2\big\}\mu_\alpha(ds, dt).$$

As a result, we obtain

$$\mathrm{E}^u\big(\|X + \xi - u\|^2_{W_0^{\alpha,2}}\big) = \rho + 4\mathrm{E}^u\left(\frac{\Delta_\alpha F}{F}\right).$$

Therefore, in order to find super-efficient estimators, it is enough to prove the existence of some ξ (independent of u) that can be written in terms of some F (possibly depending on u), with $\Delta_\alpha F \le 0$ (i.e., super-harmonic) with strict inequality on a set of positive P^u (or equivalently P) measure. In the case of shifted Brownian motion, we provide the following example.

Example 3.3. Let F be a random variable of the form of increments $F = \phi\big(X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}\big)$, for some $0 = t_0 < \cdots < t_n \le T$ (with $\phi : \mathbb{R}^n \to \mathbb{R}^n$ sufficiently regular, in order to perform all the computations below). Then, by (10), we can express $\Delta_\alpha F$ in terms of $\nabla\phi$, $\nabla^2\phi$, $\Delta_\alpha(\delta_i X)$ and, for $i, j \in \{1, \ldots, n\}$,

$$\Gamma_\alpha(\delta_i X, \delta_j X) = \int_{[0,T]^2}\tilde{D}_{s,t}\delta_i X\tilde{D}_{s,t}\delta_j X\mu_\alpha(ds, dt),$$

with the notation $\delta_i X = X_{t_i} - X_{t_{i-1}}$.

Before we proceed further, we have to take into account that, with different probabilities P^u, the random variables may have different derivatives $DF = D^uF$ and Laplacians $\Delta_\alpha F = \Delta_\alpha^u F$, since the calculus with respect to P^u is "modeled" on the process $X^u = X - u$. Thus, for $h \in L^2(0, T)$, $t \in [0, T]$, one has

$$D_h X_t = D_h X_t^u + D_h u_t = \int_0^t h(s)ds + D_h u_t$$

and $\Delta_\alpha X_t = \Delta_\alpha X_t^u + \Delta_\alpha u_t = \Delta_\alpha u_t$, provided that u_t is sufficiently regular. To proceed further with computations, we assume that the process u is deterministic, i.e., we restrict the space of parameters Θ to H_0^1 only, so that $D_h u_t = \Delta_\alpha u_t = 0$, ruling out the problem of possible dependence upon u of the Malliavin calculus that we consider. Then (10) reduces to

$$\Delta_\alpha F = \sum_{i,j=1}^n \frac{\partial^2 \phi}{\partial x_i \partial x_j} a_{i,j},$$

where, for $i, j \in \{1, \ldots, n\}$, with $t_0 = 0$,

$$a_{i,j} = \int_{[0,T]^2} \int_s^t \mathbf{1}_{[t_{i-1},t_i]}(r)dr \int_s^t \mathbf{1}_{[t_{j-1},t_j]}(r)dr \mu_\alpha(dt, ds).$$

To prove that the symmetric matrix $A = (a_{ij})_{i,j=1}^n$ is well-defined and invertible, we argue as follows: for every $v = (v_i)_{i=1}^n$, one has, using the notation $\langle Av, v\rangle = \sum_{i,j}^n a_{i,j} v_i v_j$,

$$\begin{aligned}
\langle Av, v\rangle &= \int_{[0,T]^2} \sum_{i,j}^n v_i v_j \int_s^t \mathbf{1}_{[t_{i-1},t_i]}(r)dr \int_s^t \mathbf{1}_{[t_{j-1},t_j]}(r)dr \mu_\alpha(dt, ds) \\
&= \int_{[0,T]^2} \left\{ \int_s^t \sum_{i=1}^n v_i \mathbf{1}_{[t_{i-1},t_i]}(r)dr \right\}^2 \mu_\alpha(dt, ds) \\
&= \int_{[0,T]^2} |\tilde{v}(t) - \tilde{v}(s)|^2 \mu_\alpha(dt, ds) = \|\tilde{v}\|^2_{W_0^{\alpha,2}},
\end{aligned}$$

where we let

$$\tilde{v}(t) = \int_0^t \sum_{i=1}^n \mathbf{1}_{[t_{i-1},t_i]}(s) v_i ds.$$

From this identity and (6) we deduce that A is well-defined, while non-degeneracy follows from the fact that, if $\|\tilde{v}\|_{W_0^{\alpha,2}} = 0$, then $\tilde{v}$ is constant, which cannot happen except when $v = 0$.

We let $B = (b_{i,j})_{i,j=1}^n$ be the inverse matrix of A, and consider the function defined, for all $x \in \mathbb{R}^n$, by

$$\phi(x) = \langle Bx, x\rangle^a$$

for a suitable choice of $a \in \mathbb{R}$. Then, by formally applying the chain rule in $\mathbb{R}^n$, one gets

$$\sum_{i,j}^n \frac{\partial^2 \phi}{\partial x^i \partial x^j} a_{i,j} = 2a\{2(a-1) + n\} \langle Bx, x\rangle^{a-1},$$

which suggests the choice $a \in (1-n/2, 0)$ (and $n \geq 3$). However, for a in this range, ϕ is not $C_b^2(\mathbb{R}^n)$ and in order to rigorously conclude super-efficiency for an estimator in the form $X_t + \tilde{D}_{0,t} \ln F^2$, $t \in [0, T]$, we have to justify all the applications of the chain rule above. Indeed, the only non-trivial step is to prove the following estimate, for every $u \in H_0^1$:

$$\mathrm{E}^u \left\{ \langle B(\delta X), (\delta X)\rangle^{-1} \right\} < \infty.$$

In turn, this holds true because we may pass to the joint law of $\delta X = (\delta_i X)_{i=1}^n$, which is Gaussian non-degenerate (possibly non-centered) and the integrand can then be estimated from above by some constant times the function $x \mapsto |x|^{-2}$ (here the assumption $n \geq 3$ plays a role, too).

Next, to prove, e.g., that $\ln F^2 \in \mathbb{D}^{1,2}$, with

$$D_t \ln F^2 = 2a \frac{\sum_{i,j=1}^n b_{i,j} \delta_i X \mathbf{1}_{[t_{j-1},t_j]}(t)}{\langle B(\delta X), (\delta X)\rangle}, \quad \text{for a.e. } t \in [0, T],$$

it is sufficient to notice that, assuming this identity true, then we could estimate, by the Cauchy–Schwarz inequality,

$$\int_0^T \mathrm{E}^u(|D_t \ln F^2|^2)dt \leq 4a^2 T \operatorname{trace}(B) \mathrm{E}^u \left\{ \langle B(\delta X), (\delta X)\rangle^{-1} \right\}.$$

This a priori estimate entails $\ln F^2 \in \mathbb{D}^{1,2}$, by suitably approximating the function $z \mapsto \ln z$ with smooth functions.

Similarly, to estimate $\mathrm{E}(\|\xi\|^2_{W_0^{\alpha,2}})$, we apply the Cauchy–Schwarz inequality and deduce, for $s, t \in [0, T]$, with $s < t$,

$$\mathrm{E}^u(|\tilde{D}_{s,t} \ln F^2|^2) \leq 4a^2(t-s) \operatorname{trace}(B) \mathrm{E}^u \left\{ \langle B(\delta X), (\delta X)\rangle^{-1} \right\},$$

which can be integrated with respect to μ_α (recall that $\alpha \in (0, 1/2)$).

In conclusion, the example above shows that, in the case of deterministic shifts, i.e., $\Theta = H_0^1$, we are able to explicitly build super-efficient Stein-type estimators. Although it seems plausible, we do not know whether this technique can actually be extended to stochastic shifts; it would be even more interesting to provide super-efficient *adapted* estimators; see also Remark 2.3.

4. Intensity estimation for the Cox process

In this section, we study the problem of Cramér–Rao lower bounds in the case of Cox processes (i.e., doubly stochastic Poisson processes), as it is quite interesting to compare similarities and differences between the continuous and the jump cases, the latter being in general less developed.

Let $T \geq 0$ and let $X = (X_t)_{t\in[0,T]}$ be a Poisson process defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$, with jump times $(T_k)_{k\geq 1}$ (for $k \geq 1$, we let $T_k(\omega) = T$ in the eventuality that no kth jump occurs). For the parameter space Θ, we consider the set of all absolutely continuous, (strictly) increasing, $\mathcal{F}_0$-measurable processes $u = (u_t)_{t\in[0,T]}$ such that their a.e. derivatives $(\dot{u}_t)_{t\in[0,T]}$ satisfy the assumptions of Girsanov's theorem for the Poisson process (the proofs work also for slightly smaller sets). Given $u \in \Theta$, we define the probability $P^u = L^u P$, where

$$L^u = \prod_{k=1}^{X_T} \dot{u}_{T_k} \exp\left\{-\int_0^T (\dot{u}_s - 1)ds\right\}.$$

Girsanov's theorem entails that, with respect to the probability P^u, the process X is a Cox process with intensity $(\dot{u}_t)_{t\in[0,T]}$; see, e.g., [6, Section 8.4] for details on related doubly stochastic Poisson processes. Notice that $P^u(A)$ does not depend on u for $A \in \mathcal{F}_0$; thus, in particular, for $t \in [0, T]$, $v \in \Theta$, u_t is integrable with respect to P^v and its expectation $\mathrm{E}^v(u_t)$ actually does not depend on v.

We address the problem of estimating u, or equivalently the intensity of X with respect to P^u, based on a single observation of X. In the case of a deterministic intensity, i.e., when X is an inhomogeneous Poisson process, this is investigated, e.g., in [11]. By analogy with the case of shifted Brownian motion, we introduce the following definition.

Definition 4.1. Any measurable stochastic process $\xi : \Omega \times [0, T] \to \mathbb{R}$ is called an estimator of the intensity u. An estimator of the intensity u is said to be unbiased if, for every $u \in \Theta$, $t \in [0, T]$, ξ_t is integrable and it holds $\mathrm{E}^u(\xi_t) = \mathrm{E}(u_t)$.

As in the previous section, we omit to specify "of the intensity u" and simply refer to estimators.

Privault and Révelliac studied the estimation problem, in the case of deterministic intensities, w.r.t. the risk in $L^2(\mu)$, defined as in (1), for any finite Borel measure on $[0, T]$. Their set of parameters Θ consists of all the spaces of deterministic absolutely continuous, increasing processes u, see [11, Definition 2.1]. We briefly show how a similar argument indeed applies as well to the case of stochastic intensities.

Theorem 4.2 (*Cramér–Rao Inequality in $L^2(\mu)$*). *For any unbiased estimator ξ, it holds*

$$\mathrm{E}^u(\|\xi - u\|^2_{L^2(\mu)}) \geq \int_0^T \mathrm{E}^u(u_t)\mu(dt), \quad \textit{for every } u \in \Theta,$$

and equality is attained by the (efficient) estimator $\xi = X$.

Proof. For every process $v \in \Theta$, since ξ is unbiased we have

$$\mathrm{E}^{u+\varepsilon v}(\xi_t) = \mathrm{E}^{u+\varepsilon v}(u_t + \varepsilon v_t) = \mathrm{E}^{u+\varepsilon v}(u_t) + \varepsilon \mathrm{E}^{u+\varepsilon v}(v_t), \quad \text{for } t \in [0, T].$$

Differentiating with respect to ε, as in [11, Proposition 2.3], we obtain the identity

$$\begin{aligned} \mathrm{E}^u(v_t) &= \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \mathrm{E}^{u+\varepsilon v}(\xi_t - u_t) \\ &= \mathrm{E}^u\left\{(\xi_t - u_t)\int_0^T \frac{\dot{v}_s}{\dot{u}_s}(dX_s - \dot{u}_s ds)\right\}. \end{aligned} \tag{11}$$

By the Cauchy–Schwarz inequality and the fact that X is a Cox process with intensity $\dot{u}$, we get, for $t \in [0, T]$,

$$\mathrm{E}^u(v_t)^2 \leq \mathrm{E}^u\{(\xi_t - u_t)^2\}\mathrm{E}^u\left(\int_0^T \frac{\dot{v}_s^2}{\dot{u}_s} ds\right).$$

Thus $\mathrm{E}^u\{(\xi_t - u_t)^2\} \geq \mathrm{E}^u(u_t)$ once we let $\dot{v} = \dot{u}\mathbf{1}_{[0,t]}$. The thesis follows by integration with respect to μ. $\square$

In contrast to the case of Brownian motion, the lower bound depends on the parameter $u \in \Theta$. This is quite natural in view of the classical, finite-dimensional, Cramér–Rao lower bound, where the inverse of the Fisher information appears, measuring the local regularity of the densities: when u is small, the density becomes very peaked and the bound becomes trivial.

Since the intensity $u \in \Theta$ is absolutely continuous, also in this case we investigate lower bounds for the H_0^1 risk and no unbiased estimators exist. In the next result, we also collect the case of fractional Sobolev spaces $W^{\alpha,2}$, for $\alpha \in (0, 1)$.

Theorem 4.3. *For any unbiased estimator* ξ, $\alpha \in (0, 1)$, *it holds*

$$\mathrm{E}^u(\|\xi - u\|^2_{W_0^{\alpha,2}}) \geq 2\int_0^T \mathrm{E}^u(\dot{u}_r)\int_r^T\int_0^r \frac{1}{(t-s)^{2\alpha+1}}\,dsdtdr,$$

for every $u \in \Theta$. *There exists no unbiased estimator* ξ *with finite risk in* $W^{\alpha,2}$ *for* $\alpha \in [1/2, 1)$, *as well as in* H_0^1.

Proof. We subtract (11) for two different times $s, t \in [0, T]$, and apply the Cauchy–Schwarz inequality, which yields

$$\mathrm{E}^u(|\Delta_t - \Delta_s|^2) \geq \frac{\mathrm{E}^u(|v_t - v_s|)^2}{\mathrm{E}^u\left(\int_0^T \frac{\dot{v}_s^2}{\dot{u}_s}\,ds\right)}.$$

Hence, taking $\dot{v}_r = \mathbf{1}_{[s\wedge t, s\vee t]}(r)\dot{u}_r$, we find

$$\mathrm{E}^u(|\Delta_t - \Delta_s|^2) \geq \mathrm{E}^u(|u_t - u_s|), \quad \text{for every } s, t \in [0, T].$$

If $s < t$, then the right-hand side above coincides with $\mathrm{E}^u(\int_s^t \dot{u}_r dr)$. Integrating with respect to $s, t \in [0, T]$, with measure $|t-s|^{-2\alpha-1}dtds$, we obtain the required inequality. To deduce that no unbiased estimators with finite risk exist, it is sufficient to observe that the double integral equals $+\infty$, for $\alpha \in [1/2, 1)$, and $\mathrm{E}(\dot{u}_r) > 0$ for a.e. $r \in [0, T]$. The case of H_0^1 follows at once from inequality (6). □

We end this section with some remark on the possibility of Stein-type super-efficient estimators in the case of Cox processes.

Remark 4.4 (*Malliavin Calculus for a Cox Process*)**.** It seems reasonable to develop a theory of differential calculus for Cox processes, akin to that for Poisson processes introduced in [11]. In the setting of Section 4, we let $(X_t)_{t\in[0,T]}$ be a Cox process on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$, with intensity $\lambda = (\lambda_t)_{t\in[0,T]}$ and jump times $(T_k)_{k\geq 1}$. We then denote by $\mathcal{S}$ the space of random variables F of the form

$$F = f_0\mathbf{1}_{\{X_T=0\}} + \sum_{n=1}^{\infty}\mathbf{1}_{\{X_T=n\}}f_n(T_1, \ldots, T_n),$$

where, for $n \geq 0$, $f_n : \Omega \times \mathbb{R}^n \to \mathbb{R}$ is bounded, measurable with respect to $\mathcal{F}_0 \times \mathcal{B}(\mathbb{R}^n)$ (i.e., its randomness depends only on λ) and for every $\omega \in \Omega$, $f_n(\omega; \cdot)$ is $\mathcal{C}_b^\infty(\mathbb{R}^n)$ and symmetric, i.e., $f_n(\omega; t_1, \ldots, t_n)$ is left unchanged by any permutation of the coordinates $(t_1, \ldots, t_n)$ and that, for every $n \geq 0$, one has $f_n(\omega; t_1, \ldots, t_n) = f_{n+1}(\omega; t_1, \ldots, t_n, T)$, for $\omega \in \Omega$, $t_1, \ldots, t_n \in \mathbb{R}$.

For $F \in \mathcal{S}$, we may let $DF(\omega) \in L^2(0, T)$

$$D_tF = -\sum_{n=1}^{\infty}\mathbf{1}_{\{X_T=n\}}\sum_{k=1}^{n}\mathbf{1}_{[0,T_k]}(t)\frac{1}{\lambda_{T_k}}\,\partial_k f_n(T_1, \ldots, T_n)\lambda_t,$$

for a.e. $t \in [0, T]$.

One can prove the validity of the chain rule and an integration-by-parts formula, providing some notion of divergence, thus defining Sobolev–Malliavin spaces in this setting. However, it is at present unclear how to use effectively such calculus to produce super-efficient Stein-type estimators; see Remark 4.5.

Remark 4.5 (*Stein Estimators for Cox Processes*)**.** In the case of Cox processes, nothing prevents us from performing similar arguments as in Section 3.2 using, in place of Malliavin calculus, the calculus sketched in Remark 4.4. The case of Poisson processes and $L^2(\mu)$-type risks is investigated in [11]. However, here we currently face a strong limitation to provide explicit examples, due to the possible dependence upon u (i.e., λ) of the Malliavin calculus. Let us remark that a similar limitation is also present in [11] and perhaps, at least in the one-dimensional parametric cases considered in [11, Section 5] (or in the recent paper [1] on spatial Poisson point processes) one might similarly provide explicit examples of super-efficient estimators also with respect to Sobolev risks, but the general, infinite-dimensional parametric problem would remain open.

Acknowledgments

The second and third authors are members of the GNAMPA group of the Istituto Nazionale di Alta Matematica (INdAM). This work is partially supported by the Università degli Studi di Pisa, Project PRA_2016_41. All authors thank the anonymous referees and the editor C. Genest for remarks and suggestions that led to an overall improvement of the article.

References

[1] M. Clausel, J.-F. Coeurjolly, J. Lelong, Stein estimation of the intensity of a spatial homogeneous Poisson point process, Ann. Appl. Probab. 26 (3) (2016) 1495–1534.
[2] J.M. Corcuera, A. Kohatsu-Higa, Statistical inference and Malliavin calculus, in: Seminar on Stochastic Analysis, Random Fields and Applications VI, in: Progr. Probab., vol. 63, Birkhäuser/Springer, Basel AG, Basel, 2011, pp. 59–82.
[3] E. Di Nezza, G. Palatucci, E. Valdinoci, Hitchhiker's guide to the fractional Sobolev spaces, Bull. Sci. Math. 136 (5) (2012) 521–573.
[4] E. Gobet, Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach, Bernoulli 7 (6) (2001) 899–912.
[5] W. James, C. Stein, Estimation with quadratic loss, in: Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I, Univ. California Press, Berkeley, Calif., 1961, pp. 361–379.
[6] M. Jeanblanc, M. Yor, M. Chesney, Mathematical Methods for Financial Markets, in: Springer Finance, Springer-Verlag London, Ltd., London, 2009, p. xxvi+732.
[7] J. Liu, Remarks on parameter estimation for the drift of fractional Brownian sheet, Acta Math. Vietnam. 38 (2) (2013) 241–253.
[8] I. Nourdin, G. Peccati, Normal Approximations with Malliavin Calculus, in: Cambridge Tracts in Mathematics, vol. 192, Cambridge University Press, Cambridge, 2012, p. xiv+239. From Stein's method to universality.
[9] D. Nualart, The Malliavin Calculus and Related Topics, second ed., in: Probability and its Applications (New York), Springer-Verlag, Berlin, 2006.
[10] N. Privault, A. Réveillac, Stein estimation for the drift of Gaussian processes using the Malliavin calculus, Ann. Statist. 36 (5) (2008) 2531–2550.
[11] N. Privault, A. Réveillac, Stein estimation of Poisson process intensities, Stat. Inference Stoch. Process. 12 (1) (2009) 37–53.
[12] N. Privault, A. Réveillac, Sure shrinkage of Gaussian paths and signal identification, ESAIM Probab. Stat. 15 (2011) 180–196.
[13] C.M. Stein, Estimation of the mean of a multivariate normal distribution, Ann. Statist. 9 (6) (1981) 1135–1151.