Second derivative of the log-likelihood in the model given by a Lévy driven SDE’s

By means of the Malliavin calculus, integral representation for the second derivative of the log-likelihood function are given for a model based on discrete time observations of the solution to equation $dX_t = a_\theta(X_t)dt + dZ_t$ with a Lévy process $Z$.

If we have a logarithm of transition kernel for Markov chain and can calculate two its derivatives w.r.t. parameter, we can obtain a precise formula for the logarithm of joint density and its derivatives.

The likelihood function in our model is highly implicit. In this paper, we develop an approach which makes it possible to control the properties of the likelihood and log-likelihood functions only in the terms of the objects involved in the model: the function $a_\theta(x)$, its derivatives, and the Lévy measure of the Lévy process $Z$.

**Key Words: MLE, Likelihood function, Lévy driven SDE, Regular statistical experiment, LAN.**

**Introduction**

Let $Z$ be a Lévy process without a diffusion component; that is,

$$Z_t = c + t \int_0^t \nu(ds,du) + \int_0^t \int_{|s|>1} u\nu(ds,du)$$

where $\nu$ is a Poisson point measure with the intensity measure $d\mu(du)$, and $\tilde{\nu}(ds,du) = \nu(ds,du) - d\mu(du)$ is respective compensated Poisson measure. In the sequel, we assume the Lévy measure $\mu$ to satisfy the following:

**H.**

(i) for some $\kappa > 0$,

$$\int_{|u|>1} u^{2+\kappa} \mu(du) < \infty;$$

(ii) for some $u_0 > 0$, the restriction of $\mu$ on $[-u_0, u_0]$ has a positive density

$$\sigma \in C^2([-u_0, 0) \cup (0, u_0]);$$

(iii) there exists $C_0$ such that

$$|\sigma'(u)| \leq C_0 |u|^{-1} \sigma(u),$$

$$|\sigma''(u)| \leq C_0 u^{-2} \sigma(u), |u| \in (0, u_0];$$

(iv) \((\log 1/\varepsilon)^{-1} \mu \{ u : |u| \geq \varepsilon \} \to \infty, \varepsilon \to 0.\)

Consider stochastic equation of the form

$$dX^\theta_t = a_\theta(X^\theta_t)dt + dZ_t, \quad (1)$$

where $a : \Theta \times \mathbb{R} \to \mathbb{R}$ is a measurable function,

$\Theta \subset \mathbb{R}$ is a parametric set.

In [1] it was proved that under conditions of smoothness and growth of $a_\theta$ the Markov process $X$ given by (1) has a transition probability density $p^\theta_t$ w.r.t. the Lebesgue measure. Besides, according to [1] this density has a derivative $\partial p^\theta_t(x,y)$. The extension of the asymptotic
methods of mathematical statistics is used as a key tool the second derivative of the log-likelihood ratio w.r.t. parameter. The purpose of this paper is to give a Malliavin-type integral representation of this derivative.

1 Main results

We denote by $P_x^\theta$ the distribution of this process in $\mathcal{D}([0, \infty))$ with $X_0 = x$, and by $E_x^\theta$ the expectation w.r.t. this distribution. Respectively finite-dimensional distribution for given time moments $t_1 < \cdots < t_n$ is denoted by $P_{x,\{t_k\}_{k=1}^n}^\theta$. On the other hand, solution $X$ to Eq. (1) is a random function defined on the same probability space $\Omega, \mathcal{F}, P$ with the process $Z$, which depends additionally on the parameter $\theta$ and the initial value $x = X(0)$. We do not indicate this dependence in the notation, i.e. write $X_t$ instead of e.g. $X_{x,t}^\theta$, but it will be important in the sequel that, under certain conditions, $X_t$ is $L_2$-differentiable w.r.t. $\theta$ and is $L_2$-continuous w.r.t. $(t, x, \theta)$.

In the sequel we will show that, under appropriate conditions, Markov process $X$ admits a transition probability density $p_t^\theta(x, y)$ w.r.t. Lebesgue measure, which is continuous w.r.t. $(t, x, y) \in (0, \infty) \times \mathbb{R} \times \mathbb{R}$. Then (see [2]), for every $t > 0$, $x, y \in \mathbb{R}$ such that

$$p_t^\theta(x, y) > 0, \quad \text{(2)}$$

there exists a weak limit in $\mathcal{D}([0, t])$

$$P_{x,y}^{t, \theta} = \lim_{\varepsilon \to 0} P_{x, y - \varepsilon}^\theta \left( \cdot \mid X_1 - y \leq \varepsilon \right),$$

which can be interpreted naturally as a bridge of the process $X$ started at $x$ and conditioned to arrive to $y$ at time $t$. We denote by $E_{x, y}^{t, \theta}$ the expectation w.r.t. $P_{x, y}^{t, \theta}$.

In what follows, $C$ denotes a constant which is not specified explicitly and may vary from place to place. By $C^{k,m}(\mathbb{R} \times \Theta)$, $k, m \geq 0$ we denote the class of functions $f : \mathbb{R} \times \Theta \to \mathbb{R}$ which has continuous derivatives

$$\frac{\partial^i}{\partial x^i} \frac{\partial^j}{\partial \theta^j} f, \quad i \leq k, \quad j \leq m.$$

In [1] it was proved that under the conditions of following Theorem $\partial_\theta p_t^\theta(x, y)$ has a Malliavin-type integral representation

$$\partial_\theta p_t^\theta(x, y) = g_t^\theta(x, y) p_t^\theta(x, y) \quad \text{(3)}$$

with

$$g_t^\theta(x, y) = \left\{ \begin{array}{ll} \partial_\theta \log p_t^\theta(x, y) = E_{x,y}^{t, \theta} 1, & p_t^\theta(x, y) > 0, \\ 0, & \text{otherwise.} \end{array} \right. \quad \text{(4)}$$

The goal of this section is to obtain the same representation for second derivative, i.e.

$$\partial^2_{\theta \theta} p_t^\theta(x, y) = G_t^\theta(x, y) p_t^\theta(x, y) \quad \text{(5)}$$

with

$$G_t^\theta(x, y) = \left\{ \begin{array}{ll} \partial_\theta^2 \log p_t^\theta(x, y) + g_t^\theta(x, y)^2 = \\ = E_{x,y}^{t, \theta} 2, & p_t^\theta(x, y) > 0, \\ 0, & \text{otherwise.} \end{array} \right. \quad \text{(6)}$$

The functional $\Xi_t^1$ and $\Xi_t^2$, involved in expressions for $g$ and $G$, will be introduced explicitly in the proof below; see formulas (19) and (21).

**Theorem 1.** Let $a \in C^{3,2}(\mathbb{R} \times \Theta)$ have bounded derivatives $\partial_x a, \partial_x^2 a, \partial_x^3 a, \partial_x^4 a, \partial_x^5 a$ and for all $\theta \in \Theta, x \in \mathbb{R}$

$$|a_0(x)| + |\partial_\theta a_0(x)| + |\partial^2_{\theta \theta} a_0(x)| \leq C(1 + |x|). \quad \text{(7)}$$

Then the transition probability density has a second derivative $\partial^2_{\theta \theta} p_t^\theta(x, y)$, which is continuous w.r.t. $(t, x, y, \theta) \in (0, \infty) \times \mathbb{R} \times \mathbb{R} \times \Theta$, and (5) holds true.

**Remark 1.** By statement of Theorem, the logarithm of the transition probability density has a second continuous derivative w.r.t. $\theta$ on the open subset of $(0, \infty) \times \mathbb{R} \times \Theta$ defined by inequality $p_t^\theta(x, y) > 0$ and, on this subset, admits the integral representation

$$\partial^2_{\theta \theta} \log p_t^\theta(x, y) = E_{x,y}^{t, \theta} 2 \left( E_{x,y}^{t, \theta} 1 \right)^2. \quad \text{(8)}$$

**Remark 2.** For every $\gamma < 1 + \kappa/2$ there exists constant $C$ which depends on $t$ and $\gamma$ only, such that

$$E_{x,y}^\theta \left| \partial_\theta g_t^\theta (x, X_t^\theta) \right|^\gamma \leq C(1 + |x|)^\gamma. \quad \text{(9)}$$
2 Proof of Theorem \[1\]

We need to repeat some notations and statements defined in Section 3 \[1\]. Fix \(u_1 \in (0, u_0)\), where \(u_0\) comes from \(H\) (ii), and introduce a \(C^2\)-function \(\varphi : \mathbb{R} \to \mathbb{R}^+\) with bounded derivative, such that

\[
\varphi(u) = \begin{cases} u^2, & |u| \leq u_1; \\ 0, & |u| > u_0. \end{cases}
\]

Denote by \(Q_c(x), c \in \mathbb{R}\) the value at the time moment \(s = c\) of the solution to Cauchy problem

\[
q'(s) = \varphi(q(s)), \quad q(0) = x.
\]

Then \(\{Q_c, c \in \mathbb{R}\}\) is a group of transformations of \(\mathbb{R}\), and \(\partial_c Q_c(x)|_{c=0} = \varphi(x)\).

**Definition 1.** A functional \(F \in L_2(\Omega, F, P)\) is called stochastically differentiable, if there exists an \(L_2(\Omega, F, P)\)-limit

\[
\hat{D}F = \lim_{c \to 0} \frac{1}{c} \left( Q_c(F - F) \right). \tag{10}
\]

The closure \(\hat{D}\) of the operator \(\hat{D}\) defined by (10) is called the stochastic derivative. The adjoint operator \(\delta = \hat{D}^*\) is called the divergence operator or the extended stochastic integral.

**Remark 3.** \(\text{dom}(\hat{D})\) is dense in \(L_2(\Omega, F, P)\), hence \(\delta = \hat{D}^\ast\) is well defined. In addition, \(\text{dom}(\delta)\) is dense in \(L_2(\Omega, F, P)\), hence \(\hat{D}\) is closable. The operator \(\delta\) itself is closed as an adjoint one; e.g. Theorem VIII.1 in \[3\].

Denote \(\chi(u) = -\frac{\sigma(u)\varphi(u)}{\sigma(u)}, u \neq 0\).

**Proposition 1.** 1. Let \(\varphi \in C^1(\mathbb{R}^d, \mathbb{R})\) have bounded derivatives and \(F_k \in \text{dom}(\hat{D}), k = 1, d\). Then \(\varphi(F_1, \ldots, F_d) \in \text{dom}(\hat{D})\) and

\[
\hat{D}[\varphi(F_1, \ldots, F_d)] = \sum_{k=1}^d [\partial_{x_k} \varphi](F_1, \ldots, F_d) \hat{D}F_k. \tag{11}
\]

2. The constant function \(1\) belongs to \(\text{dom}(\delta)\) and

\[
\delta(1) = \int_0^T \int_\Omega \chi(u)\bar{v}(ds, du). \tag{12}
\]

3. Let \(G \in \text{dom}(\hat{D})\) and

\[
E(\delta(1)G) < \infty. \tag{13}
\]

Then \(G \in \text{dom}(\delta)\) and \(\delta(G) = \delta(1)G - DG\).

The proofs of this Proposition and Remark \[4\] can be found in \[1\].

**Lemma 1.** Under the conditions of Theorem \[1\]. \(X^\theta_t\) is thrice stochastically differentiable and

\[
\hat{D}^j X^\theta_t = \sum_{i=0}^{j-1} \frac{(i + 1)^{j-i-1}}{i!} \int_0^t \hat{D}^{i-1}(E^s \mathcal{E}_s^{-1}) g(u) \hat{D}^s \nu(ds, du), \quad j = 1, 2, 3; \tag{14}
\]

where \(E_t := \exp \left\{ \int_0^t \partial_x a_\theta(X^\theta_s) d\tau \right\}\)

\[
D^n \mathcal{E}_t = \sum_{k=0}^{n-1} \sum_{j=0}^{k} C^k_{n-k-1} \partial_x^j \mathcal{E}_t \times \int_0^t \hat{D}^j \left( \partial_x^2 a_\theta(X^\theta_s) \right) D^{n-k-j} X^\theta_s d\tau, \quad n = 1, 2. \tag{15}
\]

**Remark 4.** The expressions for \(D^n (\partial_x^2 a_\theta(X^\theta_s))\) and \(D^n (E_t \mathcal{E}_s^{-1})\) can be found by the first statement of Proposition \[1\] (and formula (15) respectively).

**Remark 5.** Under additional conditions about smoothness and growth of \(a_\theta\) the formulas (14) and (15) are equitable if \(j\) is more than 3 and \(n\) is more than 2.

The case \(j = 1, 2\) and \(n = 1\) was considered in \[1\]. The proof of (14) as \(j \geq 3\) provides by induction using the argument of proof of relation (27) \[1\] and based on Theorem II.2.8.5 \[1\]. The same arguments that in Section 3.2 \[1\] give (see details in proof of relations (27), (31) and (32) \[1\]):

\[
\hat{D}^2 \partial_\theta X^\theta_t = \int_0^t \hat{D}^2 (E^s \mathcal{E}_s^{-1}) \partial_\theta a_\theta(X^\theta_s) ds + 2 \int_0^t \hat{D} (E^s \mathcal{E}_s^{-1}) D^2 X^\theta_s ds + E_t \int_0^t \mathcal{E}_s^{-1} \left( \partial_x^3 a_\theta(X^\theta_s)^2 + \partial_x^2 a_\theta(X^\theta_s) D^2 X^\theta_s \right) ds, \tag{16}
\]

\[
\partial_\theta X^\theta_t = E_t \int_0^t \mathcal{E}_s^{-1} \left[ \partial_\theta^2 a_\theta(X^\theta_s)^2 \right] ds + 2[\partial_\theta^2 a_\theta(X^\theta_s)^2] \partial_\theta X^\theta_s + [\partial_\theta^2 a_\theta(X^\theta_s)^2] \partial_\theta X^\theta_s \tag{17}
\]
This function is continuous w.r.t. \((t, x, y, \theta)\) because \(p^\theta_t, g^\theta_t\), and \(\partial_\theta \Xi^1_t\) depend continuously (in \(L_2\)) on \(x, t, \theta\), and relation

\[
P^\theta_{x}(X_t = y) = 0, \quad x, y \in \mathbb{R}, \quad t > 0, \quad \theta \in \Theta
\]

holds true (by representation (19)).

To prove (19), we use moment bounds for \(\partial_\theta X^\theta_t, \partial_{\theta \theta} X^\theta_t, \partial_\theta \partial_\theta X^\theta_t, D(\partial_\theta X^\theta_t), D^2(\partial_\theta X^\theta_t), D^3(\partial_\theta X^\theta_t), D^4(\partial_\theta X^\theta_t), \) and \(D^5(\partial_\theta X^\theta_t)\) to get, similarly to the proof of (37) \([1]\) (integral representation for \(p^\theta_t\)), that

\[
\frac{(\partial_\theta X^\theta_t)^2}{DX^\theta_t} \cdot \frac{1}{DX^\theta_t} \left( \delta \left( \frac{(\partial_\theta X^\theta_t)^2}{DX^\theta_t} \right) + \frac{\partial_{\theta \theta} X^\theta_t}{DX^\theta_t} \right)
\]

belong to \(\operatorname{dom}(\delta)\) and

\[
\Xi^1_t := \delta \left( \frac{(\partial_\theta X^\theta_t)^2}{DX^\theta_t} \right) + \frac{\partial_{\theta \theta} X^\theta_t}{DX^\theta_t} + \frac{\partial_\theta \Xi^1_t}{DX^\theta_t},
\]

(19)

Note that \(X_t\) is twice \(L_2\)-differentiable w.r.t. parameter \(\theta\), see (17) for its derivative. In addition, \(DX^\theta_t, D^2X^\theta_t, \) and \(D^\theta_\theta X^\theta_t\), are \(L_2\)-differentiable w.r.t. \(\theta\), and all these derivatives satisfy moment bounds similar to (35) \([1]\) (moment bounds for \(DX^\theta_t\)). Now it is easy to prove that \(\Xi^1_t\) is \(L_2\)-differentiable w.r.t. \(\theta\) (the explicit formula of the derivative is omitted). One can just replace \(DX_t\) in the denominator in the formula (19) by \(DX_t+\varepsilon\), prove that this new functional is \(L_2\)-differentiable w.r.t. \(\theta\) using the chain rule, and then show using (36) \([1]\) (negative order moment bounds for \(DX^\theta_t\)) that both this functional and its derivative w.r.t. \(\theta\) converge (locally uniformly) in \(L_2\) as \(\varepsilon \to 0\), respectively, to \(\Xi^1_t\) and to the functional \(\partial_\theta \Xi^1_t\) which comes from the formal differentiation of (19). This argument also shows that \(\Xi^1_t\) and \(\partial_\theta \Xi^1_t\) depend continuously (in \(L_2\)) on \(x, t, \theta\).

Therefore, we can take a derivative at the right hand side in (20), which gives

\[
\partial_{\theta \theta} p^\theta_t(x, y) = p^\theta_t(x, y) E^\theta_{x, y} \partial_\theta \Xi^1_t + p^\theta_t(x, y) g^\theta_t(x, y).
\]
\( C^2(\mathbb{R}) \) with bounded derivatives we have

\[
\frac{\partial^2}{\partial^2 \theta} \mathbb{E}_x f(X^\theta_t) = \\
\mathbb{E}_x \left( f''(X^\theta_t) (\partial^2 \theta X^\theta_t) + f'(X^\theta_t) (\partial^2 \theta X^\theta_t)^2 \right) = \\
\mathbb{E}_x \left( f'(X^\theta_t) \frac{Df(X^\theta_t)}{DX^\theta_t} + f'(X^\theta_t) (\partial^2 \theta X^\theta_t)^2 \right) = \\
\mathbb{E}_x \left( f(X^\theta_t) \frac{D(Df(X^\theta_t))}{DX^\theta_t} \right) + \frac{\partial^2}{\partial^2 \theta} \mathbb{E}_x f(X^\theta_t) \Xi^2_t = \mathbb{E}_x f(X^\theta_t) G^\theta_t(x, X^\theta_t); \tag{22}
\]

see (6) for the definition of \( G^\theta_t(x, y) \). Because the test function \( f \) is arbitrary, the integral identity (22) proves (5).

**Remark 6.** From (22) with \( f \equiv 1 \) it follows that

\[
\frac{\partial^2}{\partial^2 \theta} \mathbb{E}_x g^\theta_t(x, X^\theta_t) = 0.
\]

**Proof of Remark 2** By the moment bounds and formula (24), we have

\[
\mathbb{E}_x \| \Xi^2_t \|^p \leq C(1 + |x|^p) \tag{23}
\]

for every \( p \in [1, 2 + \kappa) \), with the constants \( C \) depending on \( t, p \) only.

Combining relations (3) – (6) we get

\[
\frac{\partial}{\partial \theta} g^\theta_t(x, X^\theta_t) = \mathbb{E}_x \left[ \Xi^2_t X^\theta_t - g^\theta_t(x, X^\theta_t)^2 \right],
\]

Moreover, inequality (9) follows directly from (23).

(45) (moment bounds for \( g^\theta_t \)) and Jensen’s inequality.

---

**References**

1. IVANENKO, D. and KULIK, A. (2014) Malliavin calculus approach to statistical inference for Levy driven SDE’s. [Online] – Available from [arXiv:1301.5141](arXiv:1301.5141).

2. CHAUMONT, L., URIBE BRAVO, G. (2011) Markovian bridges: Weak continuity and pathwise constructions, *Ann. Probab.*, 39(2). p. 609-647.

3. SIMON, B. READ, M. (1972) Methods of Modern Mathematical Physics. San Diego. Functional Analysis, Academic Press.

4. GIHMAN, I. and SKOROHOD, A. (1972) Stochastic differential equations and their applications.. New York, Springer-Verlag.

Received: 23.12.2013