Supplementary materials to Robust and Efficient Parameter Estimation for Discretely Observed Stochastic Processes

Rohan Hore · Abhik Ghosh

1 Theoretical Robustness of MDPDEs under IIP Models

Let us consider the notation of Section 2.1 and let $G(\cdot; t_i, t_{i-1})$ denote the true distribution for the datum Y_i having density $g(\cdot; t_i, t_{i-1})$ for all i. We can define the minimum DPD functional $\boldsymbol{T}_\alpha(G(\cdot; t_1, t_0), \ldots, G(\cdot; t_n, t_{n-1}))$ for discrete data (with finite sample size n) observed from an IIP as the minimizer of $\frac{1}{n}\sum_{i=1}^n d_\alpha(g(\cdot; t_i, t_{i-1}), f_i(\cdot; \boldsymbol{\theta}, t_i, t_{i-1}))$, with respect to $\boldsymbol{\theta} \in \Theta$. Note that, as in the definition of MDPDE, $\boldsymbol{T}_\alpha(G(\cdot; t_1, t_0), \ldots, G(\cdot; t_n, t_{n-1}))$ can equivalently be defined as the minimizer of the simpler objective function

$$\sum_{i=1}^n \left[\int f^{1+\alpha}(y; \boldsymbol{\theta}, t_i, t_{i-1}) dy - \left(1 + \frac{1}{\alpha}\right) \int f^\alpha(y; \boldsymbol{\theta}, t_i, t_{i-1}) dG(y; t_i, t_{i-1}) \right].$$

Under appropriate differentiability conditions, it leads to the estimating equation

$$\sum_{i=1}^n \Bigg[\int f^{1+\alpha}(y; \boldsymbol{\theta}, t_i, t_{i-1}) \boldsymbol{u}(y; \boldsymbol{\theta}, t_i, t_{i-1}) dy - \int f^\alpha(y; \boldsymbol{\theta}, t_i, t_{i-1}) \boldsymbol{u}(y; \boldsymbol{\theta}, t_i, t_{i-1}) g(y; t_i, t_{i-1}) dy \Bigg] = \mathbf{0}.$$

To derive the Influence function (IF) for IIP set-up, we will follow the approach used by Huber (1983) in the context of the influence function for the non-IID fixed-carriers linear models. We consider the contaminated density $g_{i,\epsilon} = (1-\epsilon) g(\cdot; t_i, t_{i-1}) + \epsilon \delta_{r_i}$ where δ_{r_i} is the degenerate distribution at the the point of the contamination point r_i for $i = 1, 2, \ldots, n$. Now, since the

Rohan Hore
Department of Statistics, University of Chicago, US
E-mail: rohanhere@uchicago.edu

Abhik Ghosh
Interdisciplinary Statistical Research Department,
Indian Statistical Institute, Kolkata, India
E-mail: abhik.ghosh@isical.ac.in

setup here is similar to that one of non-homogeneous independent setup, we set $\boldsymbol{\theta} = \boldsymbol{T}_\alpha(G(\cdot;t_1,t_0),\ldots,G(\cdot;t_n,t_{n-1}))$ to follow the process of Ghosh and Basu (2013) and our required IF of the MDPDE under the present IIP set-up come out to be

$$
\begin{aligned}
&IF(r_1,r_2,\ldots;r_n,G(\cdot,t_1,t_0),\ldots,G(\cdot,t_n,t_{n-1}))\\
&\quad= \boldsymbol{\Psi}_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\left[f^\alpha(r_i;\boldsymbol{\theta},t_i,t_{i-1})\boldsymbol{u}(r_i;\boldsymbol{\theta},t_i,t_{i-1}) - \boldsymbol{\zeta}_i\right],
\end{aligned}
$$

where $\boldsymbol{\zeta}_i = \int f(y;\boldsymbol{\theta},t_i,t_{i-1})\boldsymbol{u}(y;\boldsymbol{\theta},t_i,t_{i-1})g(y;\boldsymbol{\theta},t_i,t_{i-1})dy$.

In particular, when the increment distributions come from a single increment distribution family, as per our notation in Theorem 2, This IF can simply be expressed as

$$
\begin{aligned}
IF(r_1,r_2,\ldots,r_n;&G(\cdot,t_1,t_0),\ldots,G(\cdot,t_n,t_{n-1}))\\
&=\boldsymbol{\Psi}_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\left[f(r_i;\boldsymbol{\lambda}_i)^\alpha\boldsymbol{u}(r_i;\boldsymbol{\lambda}_i) - \boldsymbol{\zeta}_i\right],
\end{aligned}
$$

with $\boldsymbol{\zeta}_i = \int f(y;\boldsymbol{\lambda}_i)\boldsymbol{u}(y;\boldsymbol{\lambda}_i)g(y;\boldsymbol{\lambda}_i)dy$. In the following, we illustrate the form and the nature of this IF for the examples of Poisson process and the drifted Brownian motion to theoretically justify the claimed robustness of the MDPDEs for these set-ups.

1.1 Example 1: Poisson Process

Let us consider the Poisson process model and assume that the true distributions belong the same Poisson family, i.e., $g(\cdot;t_i,t_{i-1})$ is the Poisson(λ_i) density. In this case, we can easily derive that $\boldsymbol{u}(y,\lambda_i) = \boldsymbol{\Lambda}_i^T\frac{y-\lambda_i}{\lambda_i}$, and thus, $\boldsymbol{\psi}_n = \frac{1}{n}\sum_{i=1}^{n}(1+\alpha)\boldsymbol{\Lambda}_i^T\boldsymbol{C}_\alpha^{(2)}(\lambda_i)\boldsymbol{\Lambda}_i$, and $\boldsymbol{\zeta}_i = \sum_{X=1}^{\infty}\boldsymbol{\Lambda}_i^T\left(e^{-\lambda_i}\frac{\lambda_i^X}{X!}\right)^2\frac{X-\lambda_i}{\lambda_i}$. Hence, using our general form of influence function for IIP process, we have

$$
\begin{aligned}
&IF(r_1,r_2,\ldots,r_n;G(\cdot,t_1,t_0),\ldots,G(\cdot,t_n,t_{n-1}))\\
&\quad= \boldsymbol{\Psi}_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\left[\boldsymbol{\Lambda}_i^T e^{-\alpha\lambda_i}\frac{\lambda_i^{\alpha r_i}}{(r_i!)^\alpha}\frac{r_i-\lambda_i}{\lambda_i} - \boldsymbol{\zeta}_i\right].
\end{aligned}
$$

A close form precise expression of $\boldsymbol{\Psi}_n^{-1}$ or of $\boldsymbol{\zeta}_i$ or of $\boldsymbol{C}_\alpha^{(2)}(\lambda_i)$ is difficult to get without imposing any further structure; we illustrate it for a specific example below.

In consistent with our previous illustrations, e.g., as in Section 4.1, let us consider the Poisson process with the intensity function $\lambda(t) = \frac{\theta}{2\sqrt{t}}$ and a time stamp vector as $\boldsymbol{t} = (0, 1, 2, 3, \ldots, 50)$. Here θ is an uni-dimensional parameter and $\lambda_i = \theta(\sqrt{t_i} - \sqrt{t_{i-1}})$. Thus, we have

$$\psi_n = \frac{1}{n}\sum_{i=1}^{n}(1+\alpha)C_\alpha^{(2)}(\lambda_i)(\sqrt{t_i} - \sqrt{t_{i-1}})^2, \quad \text{where}$$
$$C_\alpha^{(2)}(\lambda_i) = \sum_{k\in\mathcal{N}} \frac{e^{-\lambda_i(1+\alpha)}\lambda_i^{k(1+\alpha)}(k-\lambda_i)^2}{(k!)^{1+\alpha}\lambda_i^2}.$$

Using these simplified formulas, we can then numerically compute the IF for different α, which are plotted in Figure 1.1 over the contamination points $r_1 = r_2 = \cdot = r_n$. It can be seen that, in terms of the IF analysis, the effect of contamination is linearly increasing at $\alpha = 0$, i.e., for the MLE; this unbounded nature of the IF theoretically proves the non-robust nature of the MLE. But, as α increases towards 1, the robustness of our proposed MDPDEs is more visible via their bounded IFs. The absolute value of the IF dips down significantly as we increase contamination point for high values of α. Further, the maximum of these absolute IF values decreases as α increases indicating, theoretically, the increasing robustness of the corresponding MDPDEs, in consistent with our earlier empirical illustrations.

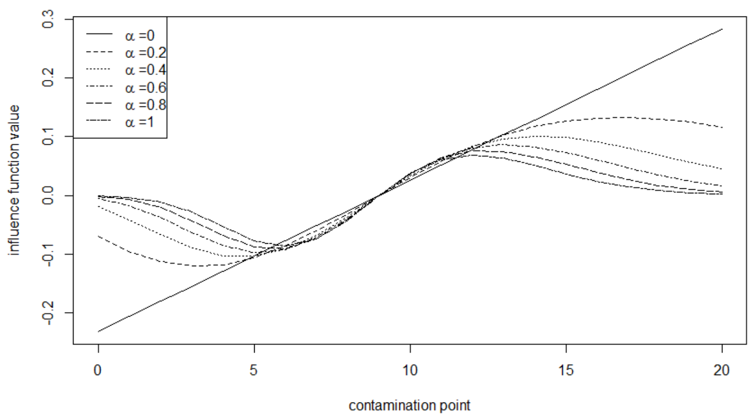

Fig. 1.1 Influence Function plot for Poisson Process for $\theta = 9$

1.2 Example 2: Drifted Brownian Motion

Let us now also derive the IF for the drifted Brownian motion assuming that the true distributions are coming from the same normal family, i.e., $g(\cdot; t_i, t_{i-1})$ is an univariate Normal(μ_i, σ_i) density. We have already derived the form of

$\boldsymbol{\Psi}_n(\boldsymbol{t})$ in Corollary 3 with the notation $\mu_i = \mu(t_i) - \mu(t_{i-1})$ and $\sigma_i = \sigma(t_i - t_{i-1})$. Following calculations in 2.4, we also have $\boldsymbol{u}(y, \boldsymbol{\lambda}_i) = \boldsymbol{\Lambda}_i^T \left(\frac{y-\mu_i}{\sigma_i^2}, -\frac{1}{\sigma_i} + \frac{(y-\mu_i)^2}{\sigma_i^3} \right)$, and $\boldsymbol{\zeta}_i = \boldsymbol{\Lambda}_i^T \frac{(2\pi)^{-\frac{1}{2}}}{\sqrt{2}\sigma_i^2} \times \left(0, -\frac{1}{2}\right)$. Hence, we have the required IF as given by

$$IF(r_1, r_2, \ldots, r_n, G(\cdot, t_1, t_0), \ldots, G(\cdot, t_n, t_{n-1}))$$
$$= \boldsymbol{\Psi}_n^{-1} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Lambda}_i^T \left(\frac{1}{(2\pi)^{\alpha/2} \sigma_i^{\alpha}} e^{-\frac{\alpha(r_i - \mu_i)^2}{2\sigma_i^2}} \begin{bmatrix} \frac{y-\mu_i}{\sigma_i^2} \\ -\frac{1}{\sigma_i} + \frac{(y-\mu_i)^2}{\sigma_i^3} \end{bmatrix} - \frac{(2\pi)^{-\frac{1}{2}}}{\sqrt{2}\sigma_i^2} \begin{bmatrix} 0 \\ -\frac{1}{2} \end{bmatrix} \right).$$

To get a better visualization of the above IF, let us consider the an example with a specific mean and sigma function as given by $\mu(t;\theta) = \theta\sqrt{t}$ and $\sigma(\theta) = 3$ (constant), along with the where time stamp vector being $\boldsymbol{t} = (0, 1, 2, 3, \ldots, 50)$. Note that, clearly θ is again an uni-dimensional parameter and $\boldsymbol{\lambda}_i = (\mu_i, \sigma_i)^T = (\theta(\sqrt{t_i} - \sqrt{t_{i-1}}), 3)^T$. Thus, we get

$$\boldsymbol{\Psi}_n(\boldsymbol{t}) = \frac{1}{n} \sum_{i=1}^{n} \frac{(2\pi)^{-\frac{\alpha}{2}}}{n(1+\alpha)^{1/2} \sigma_i^{2+\alpha}} (\sqrt{t_i} - \sqrt{t_{i-1}})^2.$$

We again numerically compute the IF of this particular example of drifted Brownian motion, using the simplified formulas, at various contamination points $r_1 = \cdots = r_n$, which is presented in Figure 1.2 for different values of α. The nature of these IFs are again the same as in the case of Poisson process example — the unbounded IF of the MLE (at $\alpha = 0$) theoretically

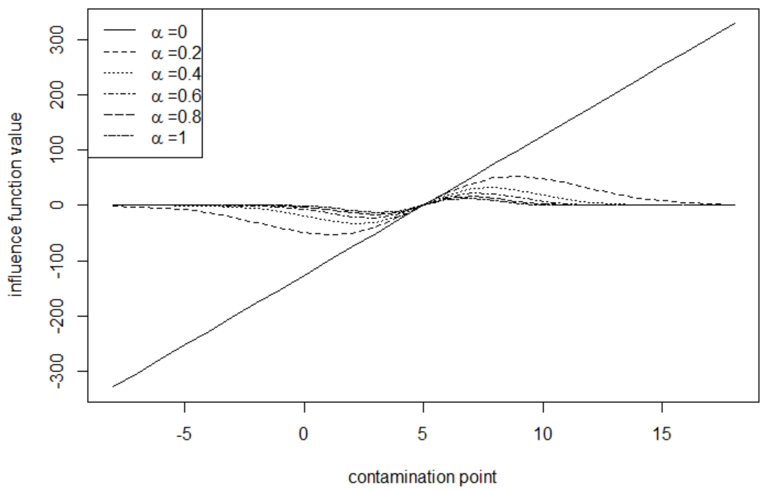

Fig. 1.2 Influence Function plot for Drifted Brownian Motion for $\theta = 5$

justify its non-robust nature and the bounded redescending nature of the IFs of the MDPDEs with $\alpha > 0$ further justify the claimed increasing robustness of the proposed MDPDEs with increasing α. These are again in line with our empirical findings from the simulation study illustrating the concurrence of the numerical and theoretical results derived in the paper.

2 Proofs of theorems and corollaries

2.1 Proof of Theorem 2

We will need the following lemma to prove the theorem.

Lemma 1 *Suppose that A and B are two non-negative random variables such that $A \leq B$ a.e.. Then, for any $N > 0$, we have*

$$AI(A > N) \leq BI(B > N) \qquad a.e..$$

Proof Suppose $\omega \in \{A > N\}$. Since, $A \leq B$ we also have $\omega \in \{B > N\}$. Hence, $\{A > N\} \subseteq \{B > N\}$. Also, since A is non-negative random variable, we have $AI(A > N) \leq AI(B > N)$ a.e.. Now, again using $A \leq B$ a.e., we will have $AI(A > N) \leq BI(B > N)$ a.e.. Hence, proved.

Proof of the theorem:
Remember, we are working under the setup that the increment distributions come from a single family of distribution. Also, we have assumed that the true distributions belong to the model family. It is easy to observe that 11 imply 1-3. Thus, as per Theorem 1, it is enough to show that 11-13 implies 6 and 7.

First Condition of 6: Now, first recall that

$$\begin{aligned}\boldsymbol{\nabla}_j V_i(X;\boldsymbol{\theta},\boldsymbol{t}) =&(1+\alpha)\left(\frac{\partial\boldsymbol{\lambda}_i}{\partial\theta_j}\right)^T\Big[\boldsymbol{C}_\alpha(\boldsymbol{\lambda}_i) - f^\alpha(X,\boldsymbol{\lambda}_i)\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)\Big]\\ =&(1+\alpha)\sum_{m=1}^k\left(\frac{\partial\boldsymbol{\lambda}_i}{\partial\theta_j}\right)_m\Big[(\boldsymbol{C}_\alpha(\boldsymbol{\lambda}_i))_m - f^\alpha(X,\boldsymbol{\lambda}_i)(\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i))_m\Big].\end{aligned}$$

Thus, we have following inequalities

$$\begin{aligned}&|\boldsymbol{\nabla}_j V_i(X;\boldsymbol{\theta},\boldsymbol{t})|\\ &\quad\leq(1+\alpha)\sum_{m=1}^k\left|\left(\frac{\partial\boldsymbol{\lambda}_i}{\partial\theta_j}\right)_m\right|\left[\Big|(\boldsymbol{C}_\alpha(\boldsymbol{\lambda}_i))_m\Big| + f^\alpha(X,\boldsymbol{\lambda}_i)\Big|(\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i))_m\Big|\right]\\ &\quad\leq(1+\alpha)\sum_{m=1}^k\left|\left(\frac{\partial\boldsymbol{\lambda}_i}{\partial\theta_j}\right)_m\right|\left[\Big|(\boldsymbol{C}_\alpha(\boldsymbol{\lambda}_i))_m\Big| + M^\alpha\Big|(\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i))_m\Big|\right]\quad\text{[using 11]}\\ &\quad\leq(1+\alpha)\sup_{i\geq1}\sup_{m\leq k}\left|\left(\frac{\partial\boldsymbol{\lambda}_i}{\partial\theta_j}\right)_m\right|\left(\sum_{m=1}^k\Big[|(\boldsymbol{C}_\alpha(\boldsymbol{\lambda}_i))_m|\Big] + kM^\alpha\|\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)\|_\infty\right).\end{aligned}$$

Now, observe that, by 11 we have, for all $m \leq k$, $|(\boldsymbol{C}_\alpha(\boldsymbol{\lambda}_i))_m|$ a continuous function of $\lambda_i \in \Lambda$. Also, Λ being a compact space we have $|(\boldsymbol{C}_\alpha(\boldsymbol{\lambda}_i))_m|$ bounded for all m. Hence, there exists a $B > 0$ such that

$$\sum_{m=1}^{k} |(\boldsymbol{C}_\alpha(\boldsymbol{\lambda}_i))_m| \leq B.$$

Thus, we have

$$|\boldsymbol{\nabla}_j V_i(X;\boldsymbol{\theta},\boldsymbol{t})| \leq (1+\alpha) \sup_{i\geq 1} \sup_{m \leq k} \left| \left(\frac{\partial \boldsymbol{\lambda}_i}{\partial \theta_j}\right)_m \right| \left[B + kM^\alpha \|\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)\|_\infty\right].$$

Further, note the following implication

$$\begin{aligned}
&(1+\alpha) \sup_{i\geq 1} \sup_{m \leq k} \left| \left(\frac{\partial \boldsymbol{\lambda}_i}{\partial \theta_j}\right)_m \right| \left[B + kM^\alpha \|\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)\|_\infty\right] > N. \\
\implies &\|\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)\|_\infty > \frac{N}{kM^\alpha(1+\alpha)\sup_{i\geq 1}\sup_{m\leq k}\left|\left(\frac{\partial \boldsymbol{\lambda}_i}{\partial \theta_j}\right)_m\right|} - \frac{B}{kM^\alpha} := N_0.
\end{aligned}$$

Now, since by 12, the terms in denominator of N_0 are finite, we have

$$N \to \infty \implies N_0 \to \infty$$

Then, using Lemma 1, we get

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} E_{\boldsymbol{\lambda}_i}\left[|\boldsymbol{\nabla}_j V_i(X;\boldsymbol{\theta},\boldsymbol{t})| I(|\boldsymbol{\nabla}_j V_i(X;\boldsymbol{\theta},\boldsymbol{t})| > N)\right] \\
\leq &\frac{1}{n}\sum_{i=1}^{n} E_{\boldsymbol{\lambda}_i}\left[(1+\alpha) \sup_{i\geq 1} \sup_{m \leq k} \left| \left(\frac{\partial \boldsymbol{\lambda}_i}{\partial \theta_j}\right)_m \right| \left[B + kM^\alpha \|\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)\|_\infty\right]\right. \\
&\quad \left. I\left((1+\alpha) \sup_{i\geq 1} \sup_{m \leq k} \left| \left(\frac{\partial \boldsymbol{\lambda}_i}{\partial \theta_j}\right)_m \right| \left[B + kM^\alpha \|\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)\|_\infty\right] > N\right)\right] \\
&\quad = (1+\alpha) \sup_{i\geq 1} \sup_{m \leq k} \left| \left(\frac{\partial \boldsymbol{\lambda}_i}{\partial \theta_j}\right)_m \right| B \left(\frac{1}{n}\sum_{i=1}^{n} P_{\boldsymbol{\lambda}_i}(\|\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)\|_\infty > N_0)\right) \\
&\quad + (1+\alpha) \sup_{i\geq 1} \sup_{m \leq k} \left| \left(\frac{\partial \boldsymbol{\lambda}_i}{\partial \theta_j}\right)_m \right| kM^\alpha \times \\
&\qquad\qquad \left(\frac{1}{n}\sum_{i=1}^{n} E_{\lambda_i}\left(\|\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)\|_\infty I(\|\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)\|_\infty > N_0)\right)\right).
\end{aligned}$$

Now, finally by Assumptions 12-13 we can see that first condition of 6 holds true.

Second condition of 6: It will follow in the same way as the first condition did, and thus, skipped here for brevity.

Assumption 7: To begin with, we will state two elementary results on norms we are gonna use. Firstly, for any given matrix $\boldsymbol{M}$ and vector $\boldsymbol{x}$, we have

$$\|\boldsymbol{M}\boldsymbol{x}\| \leq \|\boldsymbol{M}\|\|\boldsymbol{x}\|, \tag{1}$$

which directly follows from the definition. Our second result is as follows. For any two given vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^p$,we have

$$\|\boldsymbol{a} - \boldsymbol{b}\|^2 \leq 2(\||\boldsymbol{a}|\|^2 + \||\boldsymbol{b}|\|^2). \tag{2}$$

where $|\boldsymbol{a}| \in \mathbb{R}^p$ is defined as $|\boldsymbol{a}|_i = |a_i|$ for $i \leq p$. This second result (2) can be proved easily as follows:

$$\|\boldsymbol{a} - \boldsymbol{b}\|^2 = \sum_{i=1}^{p}(a_i - b_i)^2 \leq \sum_{i=1}^{p} 2(|a_i|^2 + |b_i|^2) = 2(\||\boldsymbol{a}|\|^2 + \||\boldsymbol{b}|\|^2).$$

Now, by observation (1) and Assumption 12, there exists $B > 0$ such that

$$\|\boldsymbol{\Omega}_n^{-1/2}\boldsymbol{\nabla} V_i(Y;\boldsymbol{\theta},\boldsymbol{t})\| \leq \|\boldsymbol{\Omega}_n^{-1/2}\|\|\boldsymbol{\nabla} V_i(Y;\boldsymbol{\theta},\boldsymbol{t})\| \leq B\|\boldsymbol{\nabla} V_i(Y;\boldsymbol{\theta},\boldsymbol{t})\|.$$

Recall, we have for single family of increment distributions and hence

$$\boldsymbol{\nabla}_j V_i(X;\boldsymbol{\theta},\boldsymbol{t}) = (1+\alpha)\left(\frac{\partial \boldsymbol{\lambda}_i}{\partial \theta_j}\right)^T [\boldsymbol{C}_\alpha(\boldsymbol{\lambda}_i) - f^\alpha(X,\boldsymbol{\lambda}_i)\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)].$$

In a more compact matrix notations, we can write

$$\boldsymbol{\nabla} V_i(X;\boldsymbol{\theta},\boldsymbol{t}) = (1+\alpha)\boldsymbol{\Lambda}_i^T[\boldsymbol{C}_\alpha(\boldsymbol{\lambda}_i) - f^\alpha(X,\boldsymbol{\lambda}_i)\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)].$$

Now, note that

$$\|\boldsymbol{\nabla} V_i(X;\boldsymbol{\theta},\boldsymbol{t})\| \leq (1+\alpha)\|\boldsymbol{\Lambda}_i^T\| \times \|\boldsymbol{C}_\alpha(\boldsymbol{\lambda}_i) - f^\alpha(X,\boldsymbol{\lambda}_i)\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)\|.$$

Then, by 12, one can show that there exists K satisfying $\|\boldsymbol{\Lambda}_i^T\| \leq K$ for all i.

Further, using 11 again along with Result (2), we get

$$\|\boldsymbol{C}_\alpha(\boldsymbol{\lambda}_i) - f^\alpha(X,\boldsymbol{\lambda}_i)\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)\|^2 \leq 2\|\boldsymbol{C}_\alpha(\boldsymbol{\lambda}_i)\|^2 + 2M^{2\alpha}\|\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)\|^2.$$

Next, using 11 again, $\|\boldsymbol{C}_\alpha(\boldsymbol{\lambda}_i)\|^2$ is a continuous function of $\boldsymbol{\lambda}_i$, and hence, is bounded by some $L > 0$. Thus, using Lemma 1, we get

$$\begin{aligned}
&\sum_{i=1}^{n} E_{\boldsymbol{\lambda}_i}[\|\boldsymbol{\Omega}_n(\boldsymbol{t})^{-1/2}\boldsymbol{\nabla} V_i(Y;\boldsymbol{\theta},\boldsymbol{t})\|^2 I(\|\boldsymbol{\Omega}_n(\boldsymbol{t})^{-1/2}\boldsymbol{\nabla} V_i(Y;\boldsymbol{\theta},\boldsymbol{t})\| > \epsilon\sqrt{n})] \\
&\leq \sum_{i=1}^{n} 2(B(1+\alpha)K)^2 E_{\boldsymbol{\lambda}_i}\Bigg[\left(L + M^{2\alpha}\|\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)\|^2\right) \\
&\qquad\qquad I\left(L + M^{2\alpha}\|\boldsymbol{u}_\lambda(X,\boldsymbol{\lambda}_i)\|^2 > \frac{\epsilon^2 n}{2(B(1+\alpha)K)^2}\right)\Bigg].
\end{aligned}$$

Finally, proceeding like what we showed above for assumption 6, similar results for 7 will follow too. □

2.2 Proof of Corollary 1

First observe that, here $\ell = 1$. Also, just by elementary calculus one can verify for $f(X, \lambda_i) = e^{-\lambda_i}\lambda_i^X/X!$ that

$$u_\lambda(X, \lambda_i) = \frac{X - \lambda_i}{\lambda_i},\ \nabla_\lambda u_\lambda(X, \lambda_i) = -\frac{X}{\lambda_i^2} \quad \text{and}$$
$$u_\lambda(X, \lambda_i) u_\lambda(X, \lambda_i)^T = \frac{(X - \lambda_i)^2}{\lambda_i^2}.$$

Thus, following our notation from Section 3.2, we have

$$C_\alpha(\lambda_i) = \sum_{k\in\mathcal{N}} \frac{e^{-\lambda_i(1+\alpha)}\lambda_i^{k(1+\alpha)}(k - \lambda_i)}{(k!)^{1+\alpha}\lambda_i},$$
$$C_\alpha^{(1)}(\lambda_i) = \sum_{k\in\mathcal{N}} -\frac{e^{-\lambda_i(1+\alpha)}\lambda_i^{k(1+\alpha)}k}{(k!)^{1+\alpha}\lambda_i^2} \text{ and,}$$
$$C_\alpha^{(2)}(\lambda_i) = \sum_{k\in\mathcal{N}} \frac{e^{-\lambda_i(1+\alpha)}\lambda_i^{k(1+\alpha)}(k - \lambda_i)^2}{(k!)^{1+\alpha}\lambda_i^2}.$$

Accordingly, we get

$$\Psi_n(\boldsymbol{t}) = \sum_{i=1}^n \frac{(1+\alpha)}{n}\boldsymbol{\Lambda}_i^T C_{\alpha,i}^{(2)}\boldsymbol{\Lambda}_i, \text{ and } \Omega_n(\boldsymbol{t}) = \frac{1}{n}\sum_{i=1}^n (1+\alpha)^2\boldsymbol{\Lambda}_i^T\left[C_{2\alpha,i}^{(2)} - C_{\alpha,i}^2\right]\boldsymbol{\Lambda}_i.$$

Further, in terms of notation of 13 from the main paper, with our defined notations, we also have

$$W_{2i} = \frac{|X - \lambda_i|}{\lambda_i} + \frac{|X|}{\lambda_i^2} + \frac{(X - \lambda_i)^2}{\lambda_i^2},$$

and additionally define $W_{1i} = W_{3i} = |X - \lambda_i|/\lambda_i$. To study the tail bounds of these quantities, let us note the following lemma on Poisson Distribution; its proof is given in section 3.4

Lemma 2 *Suppose $X \sim Poi(\lambda)$. Then following holds for large enough $k \in \mathbb{N}$*

$$P_\lambda(|X| > k) = O\left(\frac{1}{\sqrt{k}}\right), \quad E_\lambda(|X|I(|X| > k) = O\left(\frac{\lambda}{\sqrt{k}}\right) \quad \textit{and}$$
$$E_\lambda(X^2 I(|X| > k) = O\left(\frac{\lambda^2}{\sqrt{k}}\right).$$

Now, in order to show that 13 follows directly from 11-13 under the Poisson process model, we observe that, for large N and any i, using above Lemma 2 we have

$$P_{\lambda_i}(W_{1i} > N) = P_{\lambda_i}(X - \lambda_i > N\lambda_i) = O\left(\frac{1}{(\sqrt{(N+1)\lambda_i}}\right)$$
$$= O\left(\frac{1}{\sqrt{N\lambda_i}}\right) = O\left(\frac{1}{\sqrt{N}}\right),$$

where the last step follows using 11. Thus, one can note that, for all n,

$$\frac{1}{n}\sum_{i=1}^{n} P_{\lambda_i}(W_{1i} > N) = O\left(\frac{1}{\sqrt{N}}\right).$$

Now, similarly observe that

$$\begin{aligned}
E_{\lambda_i}[W_{1i}I(W_{1i} > N)] &= E_{\lambda_i}[XI(X > (N+1)\lambda_i)] - \lambda_i P_{\lambda_i}(X > (N+1)\lambda_i)\\
&\le E_{\lambda_i}[XI(X > (N+1)\lambda_i)]\\
&= O\left(\frac{\lambda_i}{\sqrt{(N+1)\lambda_i}}\right)\\
&= O\left(\sqrt{\frac{\lambda_i}{N}}\right) = O\left(\frac{1}{\sqrt{N}}\right) \quad \text{(Using 11).}
\end{aligned}$$

Hence, the third condition of 13 follows directly for Poisson Process if 11 holds. Also, the second condition of 13 will follow similarly as above.

Lastly, observe the following set inequality:

$$\begin{aligned}
\{W_{2i} > N\} &\subseteq \left\{\frac{|X-\lambda_i|}{\lambda_i} > N/3\right\} \cup \left\{\frac{|X|}{\lambda_i^2} > N/3\right\} \cup \left\{\frac{(X-\lambda_i)^2}{\lambda_i^2} > N/3\right\}\\
&\subseteq \{X > (N/3+1)\lambda_i\} \cup \{X > N\lambda_i^2/3\} \cup \left\{X > (\sqrt{N/3}+1)\lambda_i\right\}.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
P(V_{2i} > N) &\le P(X > (N/3+1)\lambda_i) + P(X > N\lambda_i^2/3) + P(X > (\sqrt{N/3}+1)\lambda_i)\\
&= O\left(\frac{1}{\sqrt{N}}\right) + O\left(\frac{1}{\sqrt{N}}\right) + O\left(\frac{1}{N^{1/4}}\right) \quad \text{(Proceeding as before)}\\
&= O\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}$$

Hence, first condition of 13 can be proved following a similar path. □

2.3 Proof of Corollary 2

Suppose that $\boldsymbol{\mu}_i$ is the i-th location parameter and σ_i is the i-th scale parameter and the corresponding location-scale family is

$$f(X; \boldsymbol{\mu}_i, \sigma_i) = \frac{1}{\sigma_i} f\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right).$$

Thus, one can check that

$$\begin{aligned}
\boldsymbol{u}_\lambda(X, \boldsymbol{\lambda}_i) &= \left[-\frac{1}{\sigma_i}\frac{f'\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right)}{f\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right)}\nabla_\lambda \boldsymbol{\mu}_i, -\frac{1}{\sigma_i}\left(1 + \frac{f'\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right)^T\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right)}{f\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right)}\right)\nabla_\lambda \sigma_i\right]\\
&=: \frac{1}{\sigma_i} G\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right),
\end{aligned}$$

for some suitable function G, where $f'(\cdot)$ is the derivative of $f(\cdot)$. Similarly, one can find suitable functions $G^{(1)}$ and $G^{(2)}$ such that

$$\nabla_{\boldsymbol{\lambda}} \boldsymbol{u}_\lambda(X, \boldsymbol{\lambda}_i) =: \frac{1}{\sigma_i^2} G^{(1)}\left(\frac{X-\mu_i}{\sigma_i}\right), \text{ and}$$
$$\boldsymbol{u}_\lambda(Y, \boldsymbol{\lambda}_i)\boldsymbol{u}_\lambda(Y, \boldsymbol{\lambda}_i)^T =: \frac{1}{\sigma_i^2} G^{(2)}\left(\frac{X-\mu_i}{\sigma_i}\right).$$

Hence, we can do the following simplifications

$$\begin{aligned}
\boldsymbol{C}_\alpha(\boldsymbol{\lambda}_i) =& E_{\boldsymbol{\mu}_i,\sigma_i}\left(\frac{1}{\sigma_i^{1+\alpha}} f^\alpha\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right) G\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right)\right)\\
=& \frac{1}{\sigma_i^{1+\alpha}} E_{\mathbf{0},1}\left(f^\alpha(X)G(X)\right),\\
\boldsymbol{C}_\alpha^{(1)}(\boldsymbol{\lambda}_i) =& E_{\boldsymbol{\mu}_i,\sigma_i}\left(\frac{1}{\sigma_i^{2+\alpha}} f^\alpha\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right) G^{(1)}\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right)\right)\\
=& \frac{1}{\sigma_i^{2+\alpha}} E_{\mathbf{0},1}\left(f^\alpha(X)G^{(1)}(X)\right),\\
\boldsymbol{C}_\alpha^{(2)}(\boldsymbol{\lambda}_i) =& E_{\boldsymbol{\mu}_i,\sigma_i}\left(\frac{1}{\sigma_i^{2+\alpha}} f^\alpha\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right) G^{(2)}\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right)\right)\\
=& \frac{1}{\sigma_i^{2+\alpha}} E_{\mathbf{0},1}\left(f^\alpha(X)G^{(2)}(X)\right).
\end{aligned}$$

Using these simplified results, one can easily observe that

$$\begin{aligned}
\boldsymbol{\Psi}_n(\boldsymbol{t}) =& \sum_{i=1}^n \frac{(1+\alpha)}{n\sigma_i^{2+\alpha}} \boldsymbol{\Lambda}_i^T \boldsymbol{C}_\alpha^{(2)}(\mathbf{0},1)\boldsymbol{\Lambda}_i, \text{ and}\\
\boldsymbol{\Omega}_n(\boldsymbol{t}) =& \frac{1}{n}\sum_{i=1}^n \frac{(1+\alpha)^2}{\sigma_i^{2+2\alpha}} \boldsymbol{\Lambda}_i^T \left[\boldsymbol{C}_{2\alpha}^{(2)}(\mathbf{0},1) - \boldsymbol{C}_\alpha(\mathbf{0},1)\boldsymbol{C}_\alpha(\mathbf{0},1)^T\right]\boldsymbol{\Lambda}_i.
\end{aligned}$$

But, in terms of the notation from 13, we have

$$W_{2i}(Y) = \|\boldsymbol{u}_\lambda(Y,\boldsymbol{\lambda}_i)\|_\infty + \|\nabla_{\boldsymbol{\lambda}}\boldsymbol{u}_\lambda(Y,\boldsymbol{\lambda}_i)\|_\infty, + \|\boldsymbol{u}_\lambda(Y,\boldsymbol{\lambda}_i)\boldsymbol{u}_\lambda(Y,\boldsymbol{\lambda}_i)^T\|_\infty.$$

For simplicity of notation, we further define

$$W_{1i}(Y) := \|\boldsymbol{u}_\lambda(Y,\boldsymbol{\lambda}_i)\|_\infty, \quad W_{3i}(Y) := \|\boldsymbol{u}_\lambda(Y,\boldsymbol{\lambda}_i)\|_2.$$

Then, following the discussion above, we can observe that there exists a suitable non-negative functions G_1 and G_3 independent of i such that

$$W_{1i}(Y) = \frac{1}{\sigma_i} G_1\left(\frac{Y-\boldsymbol{\mu}_i}{\sigma_i}\right) \quad \text{and} \quad W_{3i}(Y) = \frac{1}{\sigma_i} G_3\left(\frac{Y-\boldsymbol{\mu}_i}{\sigma_i}\right).$$

and further we can find some non-negative functions G_{21} and G_{22} independent of i, such that

$$W_{2i} = \frac{1}{\sigma_i} G_{21}\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right) + \frac{1}{\sigma_i^2} G_{22}\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right).$$

Now, using 11 we also have $\frac{1}{\sigma_i}$ bounded for all i. Then we have, for some $B > 0$,

$$W_{1i} \leq BG_1\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right) \quad \text{and} \quad W_{2i} \leq BG_{21}\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right) + B^2 G_{22}\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right)$$
$$\implies W_{2i} \leq G_2\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right) \quad \text{for some function } G_2 \ .$$

Thus, we have the following for all i:

$$\begin{aligned} P_{\lambda_i}[W_{1i} > N] \leq & P_{\boldsymbol{\mu}_i,\sigma_i}\left(BG_1\left(\frac{X-\boldsymbol{\mu}_i}{\sigma_i}\right) > N\right) \\ = & P_{\mathbf{0},1}(BG_1(X) > N). \end{aligned}$$

Hence, we have

$$\lim_{N\to\infty} \sup_{n>1} \frac{1}{n}\sum_{i=1}^{n} P_{\boldsymbol{\lambda}_i}[W_{1i} > N] \leq \lim_{N\to\infty} P_{\mathbf{0},1}(G_1(X) > N/B) = 0.$$

Hence, the first part of the third condition of 13 holds directly for location scale family. The second condition also follows directly, if we first observe

$$\begin{aligned} &\lim_{N\to\infty} \sup_{n>1} \frac{1}{n}\sum_{i=1}^{n} E_{\boldsymbol{\lambda}_i}[W_{1i} I(W_{1i} > N)] \\ &\qquad \leq \lim_{N\to\infty} E_{\mathbf{0},1}[BG_1(X) I(BG_1(X) > N)] = 0. \end{aligned}$$

In the above, the first inequality follows from Lemma 1 and the second equality follows by an application of the dominated convergence theorem (DCT), assuming that $\boldsymbol{u}_\lambda(X, \boldsymbol{\lambda}_i)$ is integrable. Hence, the third condition of 13 follows.

Similarly we can also show similar tail behavior for W_{2i} and W_{3i} as required in the first and second conditions of 13. □

2.4 Proof of Corollary 3

Recall that the drifted Brownian motion falls into category of location-scale family of increment distribution, and thus, the result follows for this IIP readily from Theorem 2. So, we only need to simplify the expressions for $\boldsymbol{\Psi}_n$ and $\boldsymbol{\Omega}_n$ as follows.

Here, as per the notation for the location scale family of increment distributions, we have $f(\cdot)$ to be the standard normal density, i.e.,

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Hence, with the notation from the proof of Theorem 2, one can check that

$$G(x) = \left(x, x^2 - 1\right), \quad G^{(1)}(x) = \begin{bmatrix} -1 & -2x \\ -2x & 1-3x^2 \end{bmatrix} \quad \text{and}$$
$$G^{(2)}(x) = \begin{bmatrix} x^2 & x^3 - x \\ x^3 - x & (x^2-1)^2 \end{bmatrix}.$$

Note, in this case, $f^{1+\alpha}(x) = (2\pi)^{-\frac{\alpha}{2}} \frac{1}{\sqrt{1+\alpha}}$ times the pdf of Normal distribution with mean 0 variance $\frac{1}{1+\alpha}$. Thus, one can compute that

$$\boldsymbol{C}_\alpha(\boldsymbol{\lambda}_i) = \frac{(2\pi)^{-\frac{\alpha}{2}}}{\sqrt{1+\alpha}\sigma_i^{1+\alpha}} \times \left(0, \frac{1}{1+\alpha} - 1\right),$$
$$\boldsymbol{C}_\alpha^{(1)}(\boldsymbol{\lambda}_i) = \frac{(2\pi)^{-\frac{\alpha}{2}}}{\sqrt{1+\alpha}\sigma_i^{2+\alpha}} \times \begin{bmatrix} -1 & 0 \\ 0 & 1 - \frac{3}{1+\alpha} \end{bmatrix},$$

and

$$\boldsymbol{C}_\alpha^{(2)}(\boldsymbol{\lambda}_i) = \frac{(2\pi)^{-\frac{\alpha}{2}}}{\sqrt{1+\alpha}\sigma_i^{2+\alpha}} \times \begin{bmatrix} \frac{1}{1+\alpha} & 0 \\ 0 & \frac{3}{(1+\alpha)^2} - \frac{2}{1+\alpha} + 1 \end{bmatrix},$$

and thus, it is easy to verify that

$$\boldsymbol{\Psi}_n(\boldsymbol{t}) = \sum_{i=1}^n \frac{(2\pi)^{-\frac{\alpha}{2}}(1+\alpha)^{1/2}}{n\sigma_i^{2+\alpha}} \boldsymbol{\Lambda}_i^T \begin{bmatrix} \frac{1}{1+\alpha} & 0 \\ 0 & \frac{3}{(1+\alpha)^2} - \frac{2}{1+\alpha} + 1 \end{bmatrix} \boldsymbol{\Lambda}_i,$$

and $\boldsymbol{\Omega}_n(\boldsymbol{t})$ can be equivalently expressed as

$$\frac{1}{n}\sum_{i=1}^n \frac{(1+\alpha)^2(2\pi)^{-\alpha}}{\sigma_i^{2+2\alpha}\sqrt{1+2\alpha}} \boldsymbol{\Lambda}_i^T \begin{bmatrix} \frac{1}{1+2\alpha} & 0 \\ 0 & \frac{3}{(1+2\alpha)^2} - \frac{2}{1+2\alpha} + 1 - \frac{\sqrt{1+2\alpha}}{1+\alpha}\left(\frac{1}{1+\alpha} - 1\right)^2 \end{bmatrix} \boldsymbol{\Lambda}_i.$$

$\square$

3 Proofs of the Lemmas

3.1 Proof of Lemma 3

Fix $\epsilon > 0$ and $n \in \mathbb{N}$. Using Chebychev's Inequality, we have

$$P\left[\left|\frac{Sn - E(S_n)}{n}\right| > \epsilon\right] \le \frac{Var(S_n/n)}{\epsilon^2}.$$

Now, first observe that, for all i, j, by Cauchy-Schwartz inequality, we get

$$|Cov(X_i, X_j)| \le \sqrt{Var(X_i)Var(X_j)} = \sigma^2.$$

Therefore,

$$Var\left(\frac{S_n}{n}\right) = \frac{Var(S_n)}{n^2} \leq \frac{\sum_i^n \sum_j^n |Cov(X_i, X_j)|}{n^2}$$
$$\leq \frac{\sum_i^n \sum_{j=max(i-m,0)}^{min(n,i+m)} |Cov(X_i, X_j)|}{n^2}$$
$$\leq \frac{\sum_{i=1}^n 2m\sigma^2}{n^2} \leq \frac{nm\sigma^2}{n^2} \to 0,$$

and the result follows. □

3.2 Proof of Lemma 4

To show the result, let us first observe that

$$\frac{Var(S_n)}{n} = \frac{Var(S_n)}{n} \leq \frac{\sum_i^n \sum_j^n |Cov(X_i, X_j)|}{n}$$
$$= \frac{\sum_i^n \sum_{j=max(i-m,0)}^{min(n,i+m)} |Cov(X_i, X_j)|}{n}$$
$$\underset{\text{By stationarity}}{\leq} \frac{\sum_{i=1}^n \sum_{j=i-m}^{i+m} |Cov(X_1, X_j)|}{n}$$
$$= \frac{\sum_{i=1}^n A}{n} = A.$$

Similarly, also observe that

$$\frac{Var(S_n)}{n} \geq \frac{\sum_{i=m+1}^{n-m} \sum_j^n |Cov(X_i, X_j)|}{n} \underset{\text{By stationarity}}{\geq} \frac{\sum_{i=m+1}^{n-m} A}{n} \geq \frac{(n-2m)A}{n}.$$

Hence, by Sandwich theorem, we have $\frac{Var(S_n)}{n} \to A$. □

3.3 Proof of Lemma 1

Fix $\epsilon > 0$ and $n \in$ N. Then, using Chebychev's Inequality, we have

$$P\left[\left|\frac{Sn - E(S_n)}{n}\right| > \epsilon\right] \leq \frac{Var(S_n/n)}{\epsilon^2}.$$

Now, first observe that

$$Var\left(\frac{S_n}{n}\right) = \frac{Var(S_n)}{n^2} \leq \sigma^2 \frac{\sum_i^n \sum_j^n |Cov(X_i, X_j)|}{n^2} \leq \sigma^2 \frac{\sum_i^n \sum_j^n c^{|i-j|}}{n^2}.$$

But, we can do the following calculations:

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} c^{|i-j|} = \sum_{i=1}^{n} 1 + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1} c^{|i-j|} &= n + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1} c^{i-j} \\ &= n + 2c\sum_{i=1}^{n}\frac{1-c^{i-1}}{1-c} \\ &\leq n + \frac{2c}{1-c}\sum_{i=1}^{n} 1 = n\frac{1+c}{1-c}.\end{aligned}$$

Thus, finally we have

$$\frac{Var(S_n/n)}{\epsilon^2} \leq \frac{n\left(\frac{1+c}{1-c}\right)}{n^2\epsilon^2} \to 0 \text{ as } n \to \infty.$$

Since ϵ is arbitrary, we can conclude that

$$\frac{S_n - E(S_n)}{n} \xrightarrow{P} 0.$$

□

3.4 Proof of Lemma 2

Observe firstly that, for $k \in \mathbb{N}$, we have

$$P_\lambda(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Thus, we can see that

$$\begin{aligned}P_\lambda(|X| > k) = &\sum_{i=k+1}^{\infty} e^{-\lambda}\frac{\lambda^i}{i!} \\ &\approx \sum_{i=k+1}^{\infty} e^{-\lambda}\frac{\lambda^i e^i}{(\sqrt{2\pi i})i^i} &&\left(\text{Stirling Approximation } i! \approx \sqrt{2\pi i}(i/e)^i\right) \\ &= \sum_{i=k+1}^{\infty} e^{-\lambda}\left(\frac{\lambda e}{i}\right)^i \frac{1}{\sqrt{2\pi i}} \\ &\leq \sum_{i=k+1}^{\infty} e^{-\lambda}\left(\frac{\lambda e}{k}\right)^i \frac{1}{\sqrt{2\pi k}} &&(i > k) \\ &= e^{-\lambda}\frac{1}{\sqrt{2\pi k}}\frac{1}{1-\frac{\lambda e}{k}}\left(\frac{\lambda e}{k}\right)^k &&\left(k \text{ is big, } \left(\frac{\lambda e}{k}\right) < 1\right) \\ &\leq 1 \times \frac{1}{\sqrt{k}} \times \frac{k}{k-\lambda e} \times 1 \\ &= O\left(\frac{1}{\sqrt{k}}\right) &&\left(\frac{k}{k-\lambda e} \approx 1\right).\end{aligned}$$

Next, similarly we can observe that

$$E_\lambda(|X|I(|X|>k)) = \sum_{i=k+1}^{\infty} ie^{-\lambda}\frac{\lambda^i}{i!} = \sum_{i=k}^{\infty} \lambda e^{-\lambda}\frac{\lambda^i}{i!} = \lambda P_\lambda(X>k-1)$$
$$= O\left(\frac{\lambda}{\sqrt{k}}\right).$$

Similarly, we also can show

$$E_\lambda(X^2 I(|X|>k)) = O\left(\frac{\lambda^2}{\sqrt{k}}\right).$$

□

References

Ghosh A, Basu A (2013) Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. Electronic Journal of statistics 7:2420–2456

Huber PJ (1983) Minimax aspects of bounded-influence regression. Journal of the American Statistical Association 78(381):66–72