Electron. Commun. Probab. **17** (2012), no. 29, 1–10.
DOI: 10.1214/ECP.v17-2009
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Compound Poisson approximation for triangular arrays with application to threshold estimation*

Pavel Chigansky† Fima C. Klebaner‡

Abstract

We prove weak convergence of sums over triangular arrays to the compound Poisson limit using Tikhomirov's method. The result is applied to statistical estimation of the threshold parameter in autoregressive models.

Keywords: compound Poisson; weak convergence; Tikhomirov's method; threshold estimation.
AMS MSC 2010: 60F05; 62F12; 62M10.
Submitted to ECP on May 6, 2012, final version accepted on July 13, 2012.
Supersedes arXiv:1110.5381v2.

1 Introduction and main result

This paper is concerned with weak convergence of sums over triangular arrays with certain dependence structure to the compound Poisson distribution. It is motivated by the threshold estimation problem, described in details in Section 2. We consider triangular arrays of random variables $Y_{n,j}$, $j = 1, ..., n$, $n \in \mathbb{N}$ with rows, adapted to a filtration $(\mathcal{F}_j)$, $j \in \mathbb{N}$. $Y_{n,j}$'s are asymptotically negligible and satisfy a weak dependence (mixing) condition made precise by the following assumptions.

A1. there is a constant $C_1 > 0$, such that

$$\mathbb{P}(Y_{n,j} \neq 0) \le \frac{C_1}{n}, \quad \text{and} \quad \mathbb{E}|Y_{n,j}| \le \frac{C_1}{n}, \quad j = 1, ..., n$$

and

$$\mathbb{E}|Y_{n,j}|\mathbf{1}_{\{Y_{n,i}\neq 0\}} \le \left(\frac{C_1}{n}\right)^2, \quad i \neq j.$$

A2. there is an integer $\ell \ge 1$, such that

$$\Big|\mathbb{E}\big(Y_{n,j}\big|\mathcal{F}_i\big) - \mathbb{E}Y_{n,j}\Big| \le \frac{C_1}{n}\alpha(j-i), \quad i \le j - \ell$$

where $\alpha(n) \ge 0$ is a decreasing sequence with $\lim_{n\to\infty} \alpha(n) = 0$

A3. for a measurable function $|v(x)| \le 1$, $x \in \mathbb{R}^{n-j+1}$

$$\Big|\mathbb{E}\big(v(Y_{n,j}, ..., Y_{n,n})\big|\mathcal{F}_i\big) - \mathbb{E}v(Y_{n,j}, ..., Y_{n,n})\Big| \le \alpha(j-i), \quad i < j$$

*Supported by the Australian Research Council Grant DP0988483
†The Hebrew University, Israel E-mail: pchiga@mscc.huji.ac.il
‡Monash University, Australia. E-mail: fima.klebaner@monash.edu

The following condition on the individual characteristic functions $\phi_{n,j}(t) = \mathbb{E}e^{\mathbf{i}tY_{n,j}}$ together with the above assumptions, will assure convergence of the sums

$$S_n = \sum_{j=1}^{n} Y_{n,j}, \quad n \in \mathbb{N},$$

to the compound Poisson law (hereafter we shall abbreviate $\dot{\varphi}(t) = \frac{d}{dt}\varphi(t)$, etc.):

A4. There exists a characteristic function $\varphi(t)$ and positive constants C_2 and μ such that
$$\Big|\dot{\phi}_{n,j}(t) - n^{-1}\mu\dot{\varphi}(t)\Big| \le C_2 n^{-2}, \quad t \in \mathbb{R}.$$

Note that the mixing in A2 and A3 can be arbitrarily weak. Further assumptions on the rate of convergence of $\alpha(k)$ to zero, such as:

A5. $\alpha(k) \le C_3 r^k$ for some $r \in (0,1)$ and $C_3 > 0$.

allow to obtain rates of convergence in an appropriate metric. Below we shall work with the Lévy distance, defined for a pair of distribution functions F and G by (see e.g. [10])
$$L(G,F) = \inf\big\{h > 0 : G(x-h) - h \le F(x) \le G(x+h) + h, \ \forall\ x\big\}.$$

Our main result is the following:

Theorem 1.1. *Let $Y_{n,j}$, $j = 1,...,n$, $n \in \mathbb{N}$ be a triangular array of random variables, whose rows are adapted to a filtration $(\mathcal{F}_j)$, $j \in \mathbb{N}$ and satisfy the assumptions A1-A4. Then*
$$S_n = \sum_{j=1}^{n} Y_{n,j} \xrightarrow[n\to\infty]{d} S, \tag{1.1}$$
where S has the compound Poisson distribution, with intensity μ and i.i.d. jumps with characteristic function $\varphi(t)$.

Moreover, if the assumption A5 holds then there is a constant $C > 0$, such that for all n large enough,
$$L\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) \le C n^{-1/2}\log n, \tag{1.2}$$
where $L\big(\mathcal{L}(S_n), \mathcal{L}(S)\big)$ is the Lévy distance between the distribution functions of S_n and S.

Remark 1.2. *Both the constant C and the smallest n for which* (1.2) *holds, can be found explicitly in terms of the C_i's and $\alpha(\cdot)$, mentioned in the assumptions above. Also bounds on the Lévy distance can be obtained similarly for e.g. polynomially decreasing $\alpha(\cdot)$, by replacing $b\log n$ with n^δ for some $\delta > 0$ in the proof of Theorem 1.1 and optimizing the right hand side of the corresponding inequality, analogous to* (3.5) *below.*

In application to threshold estimation, $Y_{n,j}$ is derived from an autoregressive stationary process X_j, generated by the recursion

$$X_j = h(X_{j-1}) + \varepsilon_j, \quad j \ge 1, \tag{1.3}$$

where $h(\cdot)$ is a given measurable function and (ε_j) is a sequence of i.i.d. random variables, with continuous positive probability density $q(\cdot)$. As explained in Section 2, in this context
$$Y_{n,j} := f(\varepsilon_j)\mathbf{1}_{\{X_{j-1}\in B_n\}},$$

where $B_n := [0, 1/n]$, $f(\cdot)$ is a measurable function and

$$S_n := \sum_{j=1}^{n} f(\varepsilon_j)\mathbf{1}_{\{X_{j-1}\in B_n\}}.$$

Theorem 1.1 implies that under appropriate conditions, S_n converges weakly to the compound Poisson random variable with i.i.d. jumps, distributed as $f(\varepsilon_1)$, and the intensity $\mu := p(0)$, where $p(\cdot)$ is the unique invariant density of (X_j).

Somewhat surprisingly, we were not able to find in the literature a general result, from which this limit could be deduced. In this regard, one naturally thinks of Stein's method or martingale convergence results. Stein's method appears to be particularly well suited to the compound Poisson distribution with *integer* valued jumps (see e.g. [3], [2]). The results such as [4], [18, 19], [5] or [17] come close, but apparently do not quite fit our setting.

In the particular case, when $\mathbb{E}f(\varepsilon_j) = 0$, S_n becomes a sum over the array of martingale differences $Y_{n,j} := f(\varepsilon_j)\mathbf{1}_{\{X_{j-1}\in B_n\}}$, $j = 1, ..., n$ with the quadratic variation sequence

$$V_{n,m} = \sum_{j=1}^{m} \mathbf{1}_{\{X_{j-1}\in B_n\}}\mathbb{E}f^2(\varepsilon_1), \quad m = 1, ..., n.$$

A typical martingale limit result such as e.g. [6] or Theorem 2.27 Ch. VIII §2c in [13] requires that $V_{n,n}$ converges in probability. However in our case $V_{n,n}$ converges only in distribution (to a Poisson random variable), but not in probability (since e.g. $V_{n,n}$ is uniformly integrable, but is not a Cauchy sequence in L_1). It is known that S_n may have a different limit or no limit at all, if the convergence in probability of quadratic variation is replaced with convergence in distribution (see [1] and the references therein), so that the martingale results also do not appear applicable[1].

The objective of this paper is to give a proof of Theorem 1.1, using Tikhomirov's method from [20]. Originally applied to CLT in the dependent case, it turns to be remarkably suitable to the setting under consideration. Before proceeding to the proof in Section 3, we shall discuss in more details the application, in which the aforementioned convergence arises.

2 Application to threshold estimation

Suppose one observes a sample $X^n = (X_1, ..., X_n)$ from a threshold autoregressive (TAR) time series, generated by the recursion

$$X_j = g_+(X_{j-1})\mathbf{1}_{\{X_{j-1}\ge\theta\}} + g_-(X_{j-1})\mathbf{1}_{\{X_{j-1}<\theta\}} + \varepsilon_j, \quad j \in \mathbb{Z}_+, \tag{2.1}$$

where $g_+(\cdot)$ and $g_-(\cdot)$ are known functions and (ε_j) is a sequence of i.i.d. random variables with known probability density $q(\cdot)$. The unknown *threshold* parameter θ, taking values in an open interval $\Theta := (a, b) \subset \mathbb{R}$, is to be estimated from the sample X^n. TAR models, such as (2.1), have been the subject of extensive research in statistics and econometrics (see e.g. [21] and the recent surveys [22], [11], [8]).

From the statistical analysis point of view, this estimation problem classifies as "singular", since the corresponding likelihood function

$$L_n(X^n; \theta) = \prod_{j=1}^{n} q\Big(X_j - g_+(X_{j-1})\mathbf{1}_{\{X_{j-1}\ge\theta\}} - g_-(X_{j-1})\mathbf{1}_{\{X_{j-1}<\theta\}}\Big) \tag{2.2}$$

[1] in this connection, it is interesting to note, that in the analogous continuous time setting, the quadratic variation does converge in probability, essentially due to the continuity of the sample paths, see [14]

is discontinuous in θ. Typically in such problems, the sequence of the Bayes estimators

$$\tilde{\theta}_n := \frac{\int_\Theta \theta L_n(X^n;\theta)\pi(\theta)d\theta}{\int_\Theta L_n(X^n;\theta)\pi(\theta)d\theta}, \quad n \ge 1$$

is asymptotically efficient in the minimax sense for an arbitrary continuous prior density $\pi(\cdot)$ (see [12]). The asymptotic distribution of these estimators is determined by the weak limit of the likelihood ratios as follows. Let $\theta_0 \in \Theta$ be the true unknown value of the parameter and r_n an increasing sequence of numbers. The change of variables $u = r_n(\theta - \theta_0) \in r_n(\Theta - \theta_0) =: \mathbb{U}_n$ gives

$$r_n(\tilde{\theta}_n - \theta) = \frac{\int_{\mathbb{U}_n} u Z_n(u)\pi(\theta_0 + u/r_n)du}{\int_{\mathbb{U}_n} Z_n(u)\pi(\theta_0 + u/r_n)du},$$

where $Z_n(u)$, $n \ge 1$ are the rescaled likelihood ratios

$$Z_n(u) = \frac{L_n(X^n;\theta_0 + u/r_n)}{L_n(X^n;\theta_0)}, \quad u \in \mathbb{U}_n.$$

If r_n can be chosen so that $Z_n(u)$, $u \in \mathbb{R}$ converges weakly to a random process $Z(u)$, $u \in \mathbb{R}$ in an appropriate topology, then

$$r_n(\tilde{\theta}_n - \theta) \xrightarrow[n\to\infty]{d} \frac{\int_{\mathbb{R}} u Z(u)du}{\int_{\mathbb{R}} Z(u)du}, \tag{2.3}$$

holds (a comprehensive account of this approach can be found in [12]).

For the likelihoods as in (2.2), a simple calculation (see eq. (4) in [9]) reveals that

$$\log Z_n(u) = \sum_{j=1}^n \mathbf{1}_{\{X_{j-1}\in B_n\}} \log \frac{q\big(\varepsilon_j + \delta(X_{j-1})\big)}{q\big(\varepsilon_j\big)}, \quad u \ge 0 \tag{2.4}$$

where $B_n := [\theta_0, \theta_0 + u/n]$ and $\delta(x) := g_+(x) - g_-(x)$, and a similar expression is obtained for $u < 0$. It can be shown that (2.3) indeed holds with $r_n := n$, if (X_j) is a sufficiently fast mixing with the unique invariant probability density $p(x;\theta_0)$, and the sequence $\log Z_n(u)$ converges weakly to the compound Poisson process

$$\log Z(u) := \begin{cases} \sum_{j=1}^{\Pi^+(u)} \log \frac{q\left(\varepsilon_j^+ + \delta(\theta_0)\right)}{q\left(\varepsilon_j^+\right)} & u \ge 0 \\ \sum_{j=1}^{\Pi^-(-u)} \log \frac{q\left(\varepsilon_j^- - \delta(\theta_0)\right)}{q\left(\varepsilon_j^-\right)} & u < 0 \end{cases}, \tag{2.5}$$

where $(\varepsilon_j^\pm)$ are i.i.d. copies of ε_1 and $\Pi^+(u)$ and $\Pi^+(u)$ are independent Poisson processes with the same intensity $p(\theta_0;\theta_0)$.

The rate $r_n = n$ and the Poisson behavior is typical for discontinuous likelihoods (see e.g. Ch. 5, [12]). For the linear TAR model, i.e. when $g_\pm(x) = \rho_\pm x$ with constants $\rho_- \ne \rho_+$, this asymptotic appeared in [7] and the aforementioned generalization is taken from [9].

One particularly interesting ingredient in the proof, which is the main focus of this article, is the convergence of the finite dimensional distributions of $Z_n(u)$ to those of $Z(u)$. In its prototypical form, the problem can be restated as follows. Consider the stationary Markov sequence (X_j), generated by the recursion (1.3) and let (cf. (2.4))

$$S_n := \sum_{j=1}^n f(\varepsilon_j)\mathbf{1}_{\{X_{j-1}\in B_n\}}, \tag{2.6}$$

where $B_n := [0, 1/n]$ and $f(\cdot)$ is a measurable function. It is required to show that, the sums (S_n) converge weakly to the compound Poisson random variable with i.i.d. jumps, distributed as $f(\varepsilon_1)$, and the intensity $p(0)$, where $p(\cdot)$ is the unique invariant density of (X_j).

This convergence is not hard to prove using the blocks technique: S_n is partitioned into, say, $n^{1/2}$ blocks of $n^{1/2}$ consecutive summands, $n^{1/4}$ of which are discarded. Removing total of $n^{1/2} \cdot n^{1/4}$ out of n terms in the sum does not alter its limit, but the residual blocks become nearly independent, if the mixing is fast enough. Moreover, a single event $\{X_j \in B_n\}$ occurs within each block with probability of order $n^{-1/2}$ and hence the sum over approximately independent $n^{1/2}$ blocks yields the claimed compound Poisson behavior. This approach dates back to at least [15] in the Poisson case, and the details for the compound Poisson setting can be found in [9].

An alternative proof now can be given by applying Theorem 1.1:

Corollary 2.1. *Let (X_j) be defined by* (1.3) *and S_n by* (2.6)*. Assume that*

i. ε_1 has positive Lipschitz continuous bounded probability density $q(x)$, $x \in \mathbb{R}$ with the finite first absolute moment $\int_{\mathbb{R}} |x|q(x)dx < \infty$

ii. for some $r \in (0, 1)$ and $C > 0$,

$$|h(x)| \le r|x|, \quad \forall\, |x| \ge C$$

iii. $\mathbb{E}|f(\varepsilon_1)| < \infty$ and for some constant C',

$$\sup_{z,x\in[0,n^{-1}]} \big|f\big(z - h(x)\big)\big| \le C'$$

for all n large enough.

Then the Markov process (X_j) has unique invariant density $p(x)$, $x \in \mathbb{R}$, which is positive, Lipschitz continuous and bounded; for stationary (X_j), the sums (S_n) converge weakly to the compound Poisson random variable with intensity $p(0)$ and i.i.d. jumps with the same distribution as $f(\varepsilon_1)$.

Remark 2.2. *The Corollary 2.1 verifies the weak convergence of the one-dimensional distributions of the processes $\log Z_n(u)$ from* (2.4) *to those of $\log Z(u)$, $u \in \mathbb{R}$ defined in* (2.5)*. The convergence of finite dimensional distributions of higher orders can be treated along the same lines. The limit* (2.3) *then follows from the tightness of the sequence of processes $\log Z_n(u)$ (see [9] for further details).*

Remark 2.3. *The assumption iii holds if e.g. $f(\cdot)$ and $h(\cdot)$ are continuous at 0.*

Proof. Under the assumptions i and ii, the standard ergodic theory of Markov chains (see e.g. Theorem 16.0.2 in [16]) implies that (X_j) is irreducible, aperiodic and positive recurrent Markov chain with the unique invariant measure. Due to the additive structure of the recursion (1.3), the invariant measure has density $p(\cdot)$, which is positive and continuous with the same Lipschitz constant L_q as the density $q(\cdot)$ and $\|p\|_\infty \le \|q\|_\infty := \sup_{x\in\mathbb{R}} q(x)$. Moreover, (X_j) is geometrically mixing, i.e. there exist positive constants R and $\rho < 1$, such that for any measurable function $|g(x)| \le 1$

$$\left|\mathbb{E}\big(g(X_j)|\mathcal{F}_i\big) - \int_{\mathbb{R}} g(x)p(x)dx\right| \le R\rho^{j-i}, \quad j > i, \tag{2.7}$$

where $\mathcal{F}_j = \sigma\{\varepsilon_i, i \le j\}$. Define $Y_{n,j} := f(\varepsilon_j)\mathbf{1}_{\{X_{j-1}\in B_n\}}$, then

$$\mathbb{E}|Y_{n,j}| = \mathbb{E}|f(\varepsilon_j)|\mathbb{P}(X_{j-1} \in B_n) = \mathbb{E}\big|f(\varepsilon_1)\big| \int_0^{1/n} p(x)dx \le \mathbb{E}|f(\varepsilon_1)|\|q\|_\infty n^{-1},$$

and similarly

$$\mathbb{P}(Y_{n,j} \ne 0) \le \mathbb{P}(X_{j-1} \in B_n) \le \|q\|_\infty n^{-1}.$$

Further, for $i < j-1$,

$$\begin{aligned}\mathbb{E}\big|Y_{n,i}\mathbf{1}_{\{Y_{n,j}\ne 0\}}\big| = \mathbb{E}\big|f(\varepsilon_i)\big|\mathbf{1}_{\{X_{i-1}\in B_n\}}\mathbb{P}(Y_{n,j} \ne 0|\mathcal{F}_i) \le \\ \mathbb{E}\big|f(\varepsilon_i)\big|\mathbf{1}_{\{X_{i-1}\in B_n\}}\mathbb{P}(X_{j-1} \in B_n|\mathcal{F}_i) \le \\ \mathbb{E}\big|f(\varepsilon_i)\big|\mathbf{1}_{\{X_{i-1}\in B_n\}}\|q\|_\infty n^{-1} \le \mathbb{E}\big|f(\varepsilon_1)\big|\|q\|^2_\infty n^{-2},\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\big|\mathbf{1}_{\{Y_{n,i}\ne 0\}}Y_{n,j}\big| = \mathbb{E}\mathbf{1}_{\{Y_{n,i}\ne 0\}}\mathbb{E}\big(|Y_{n,j}|\big|\mathcal{F}_i\big) \le \\ \mathbb{E}\mathbf{1}_{\{X_{i-1}\in B_n\}}\mathbb{E}\Big(\big|f(\varepsilon_j)\big|\mathbf{1}_{\{X_{j-1}\in B_n\}}\big|\mathcal{F}_i\Big) \le \mathbb{E}\big|f(\varepsilon_1)\big|\|q\|^2_\infty n^{-2}.\end{aligned}$$

Similarly,

$$\begin{aligned}\mathbb{E}|Y_{n,j-1}|\mathbf{1}_{\{Y_{n,j}\ne 0\}} \le \mathbb{E}\mathbf{1}_{\{X_{j-2}\in B_n\}}|f(\varepsilon_{j-1})|\mathbf{1}_{\{X_{j-1}\in B_n\}} = \\ \mathbb{E}\mathbf{1}_{\{X_{j-2}\in B_n\}}\mathbb{E}\Big(|f(\varepsilon_{j-1})|\mathbf{1}_{\{h(X_{j-2})+\varepsilon_{j-1}\in B_n\}}|\mathcal{F}_{j-2}\Big) = \\ \mathbb{E}\mathbf{1}_{\{X_{j-2}\in B_n\}}\int_{\mathbb{R}}|f(y)|\mathbf{1}_{\{h(X_{j-2})+y\in B_n\}}q(y)dy = \\ \mathbb{E}\mathbf{1}_{\{X_{j-2}\in B_n\}}\int_0^{n^{-1}}\big|f\big(z-h(X_{j-2})\big)\big|q\big(z-h(X_{j-2})\big)dz \le \\ \|q\|_\infty\mathbb{E}\mathbf{1}_{\{X_{j-2}\in B_n\}}n^{-1}\sup_{z,x\in[0,n^{-1}]}|f\big(z-h(x)\big)| \le \|q\|^2_\infty n^{-2}C'\end{aligned}$$

and

$$\mathbb{E}\mathbf{1}_{\{Y_{n,j-1}\ne 0\}}|Y_{n,j}| \le \mathbb{E}\mathbf{1}_{\{X_{j-2}\in B_n\}}|f(\varepsilon_j)|\mathbf{1}_{\{X_{j-1}\in B_n\}} \le \mathbb{E}\big|f(\varepsilon_1)\big|\|q\|^2_\infty n^{-2}.$$

Hence A1 is satisfied for all n large enough with

$$C_1 := \Big(\|q\|^2_\infty \vee 1\Big)\Big(\mathbb{E}\big|f(\varepsilon_1)\big| \vee C' \vee 1\Big).$$

Further, by the Markov property,

$$\begin{aligned}\mathbb{E}(Y_{n,j}|\mathcal{F}_{j-2}) = \mathbb{E}\Big(\mathbf{1}_{\{X_{j-1}\in B_n\}}\mathbb{E}\big(f(\varepsilon_j)|\mathcal{F}_{j-1}\big)\big|\mathcal{F}_{j-2}\Big) = \\ \mathbb{E}f(\varepsilon_1)\mathbb{P}\big(X_{j-1} \in B_n\big|X_{j-2}\big) =: H(X_{j-2}),\end{aligned}$$

and

$$\big|H(X_{j-2})\big| \le \big|\mathbb{E}f(\varepsilon_1)\big|\|q\|_\infty n^{-1} \le C_1 n^{-1}.$$

Hence by (2.7), for $i < j-1$,

$$\begin{aligned}\Big|\mathbb{E}\big(Y_{n,j}\big|\mathcal{F}_i\big) - \mathbb{E}Y_{n,j}\Big| = \Big|\mathbb{E}\Big(\mathbb{E}\big(Y_{n,j}\big|\mathcal{F}_{j-2}\big)|\mathcal{F}_i\Big) - \mathbb{E}\mathbb{E}\big(Y_{n,j}\big|\mathcal{F}_{j-2}\big)\Big| = \\ \Big|\mathbb{E}\big(H(X_{j-2})|\mathcal{F}_i\big) - \mathbb{E}H(X_{j-2})\Big| \le C_1 n^{-1}R\rho^{j-i-2},\end{aligned}$$

and A2 holds with $\ell = 2$ and

$$\alpha(k) := R\rho^{k-2}. \tag{2.8}$$

The assumption A3 is checked similarly. Finally,

$$\dot{\phi}_{n,j}(t) = \mathbb{E}\mathbf{i}f(\varepsilon_j)\mathbf{1}_{\{X_{j-1}\in B_n\}}e^{\mathbf{i}tf(\varepsilon_j)\mathbf{1}_{\{X_{j-1}\in B_n\}}} = \mathbb{E}\mathbf{i}f(\varepsilon_j)\mathbf{1}_{\{X_{j-1}\in B_n\}}e^{\mathbf{i}tf(\varepsilon_j)} = \mathbb{E}\mathbf{1}_{\{X_{j-1}\in B_n\}}\mathbb{E}\Big(\mathbf{i}f(\varepsilon_j)e^{\mathbf{i}tf(\varepsilon_j)}\big|\mathcal{F}_{j-1}\Big) = \mathbb{P}(X_1\in B_n)\dot{\varphi}(t),$$

where $\varphi(t) = \mathbb{E}e^{\mathbf{i}tf(\varepsilon_1)}$ and interchanging derivative and the expectation is valid by the dominated convergence and iii.

Since the invariant density is Lipschitz, it follows that

$$\Big|\dot{\phi}_{n,j}(t) - p(0)\frac{1}{n}\dot{\varphi}(t)\Big| \le L_q n^{-2},$$

which verifies A4 and the claim now follows from Theorem 1.1. In fact, the assumption A5 holds by virtue of (2.8) and the Lévy distance to the limit distribution converges at the rate, claimed in (1.2). □

3 Proof of Theorem 1.1

Tikhomirov's approach [20] is applicable, when the characteristic function of the limit distribution uniquely solves an ordinary differential equation. Roughly, the idea is then to show that the characteristic functions of the prelimit distributions satisfy the same equation in the limit.

The characteristic function of the compound Poisson distribution with intensity μ and characteristic function of the jumps $\varphi(t)$ is given by

$$\psi(t) = e^{\mu(\varphi(t)-1)}, \quad t\in\mathbb{R}$$

which solves uniquely the initial value problem

$$\dot{\psi}(t) = \mu\dot{\varphi}(t)\psi(t), \quad \psi(0)=1, \quad t\in\mathbb{R}.$$

Since $\mathbb{E}|S_n|<\infty$, the characteristic function $\psi_n(t) := \mathbb{E}e^{\mathbf{i}tS_n}$ is continuously differentiable and $\Delta_n(t) := \psi(t)-\psi_n(t)$ satisfies

$$\dot{\Delta}_n(t) = \mu\dot{\varphi}(t)\Delta_n(t) + r_n(t), \quad t\in\mathbb{R},$$

subject to $\Delta_n(0)=0$, where $r_n(t) := \mu\dot{\varphi}(t)\psi_n(t) - \dot{\psi}_n(t)$. Solving for $\Delta_n(t)$ gives

$$\Delta_n(t) = \int_0^t \exp\Big(\mu\big(\varphi(t)-\varphi(s)\big)\Big)r_n(s)ds, \quad t\ge 0. \tag{3.1}$$

As we show below, for any constant $b>0$, such that $b\log n$ is a positive integer,

$$|r_n(t)| \le C_2 n^{-1} + 3C_1\alpha(b\log n) + 8C_1^2 b\frac{\log n}{n}, \quad t\ge 0 \tag{3.2}$$

and, since $|\varphi(t)|\le 1$, it follows from (3.1) that

$$|\Delta_n(t)| \le e^{2\mu}\int_0^t |r_n(s)|ds \le e^{2\mu}\left(C_2 n^{-1} + 3C_1\alpha(b\log n) + 8C_1^2 b\frac{\log n}{n}\right)t, \quad t\ge 0. \tag{3.3}$$

Similar bound holds for $t<0$ and the claimed weak limit (1.1) follows, once we check

(3.2). To this end, we have

$$\dot{\psi}_n(t) := \frac{d}{dt}\mathbb{E}e^{\mathbf{i}tS_n} = \mathbb{E}\frac{d}{dt}\exp\Big(\mathbf{i}t\sum_{j=1}^n Y_{n,j}\Big) = \sum_{k=1}^n \mathbb{E}\mathbf{i}Y_{n,k}\exp\Big(\mathbf{i}t\sum_{j=1}^n Y_{n,j}\Big) =$$
$$\sum_{k=1}^n \mathbb{E}\mathbf{i}Y_{n,k}e^{\mathbf{i}tY_{n,k}}\exp\Big(\mathbf{i}t\sum_{j\neq k} Y_{n,j}\Big) = \sum_{k=1}^n \mathbb{E}\mathbf{i}Y_{n,k}e^{\mathbf{i}tY_{n,k}}\exp\Big(\mathbf{i}t\sum_{|j-k|>b\log n} Y_{n,j}\Big) +$$
$$\sum_{k=1}^n \mathbb{E}\mathbf{i}Y_{n,k}e^{\mathbf{i}tY_{n,k}}\bigg(\exp\Big(\mathbf{i}t\sum_{j\neq k} Y_{n,j}\Big) - \exp\Big(\mathbf{i}t\sum_{|j-k|>b\log n} Y_{n,j}\Big)\bigg) := J_1 + J_2$$

where we used $\mathbb{E}|S_n| < \infty$ and the dominated convergence to interchange the derivative and the expectation. Note that $\big|e^{\mathbf{i}x} - e^{\mathbf{i}(x+y)}\big| \le 2\mathbf{1}_{\{y\neq 0\}}$ for any $x, y \in \mathbb{R}$, and hence by the assumption A1

$$|J_2| \le \sum_{k=1}^n \mathbb{E}|Y_{n,k}|\bigg|\exp\Big(\mathbf{i}t\sum_{j\neq k} Y_{n,j}\Big) - \exp\Big(\mathbf{i}t\sum_{|j-k|>b\log n} Y_{n,j}\Big)\bigg| \le \tag{3.4}$$
$$2\sum_{k=1}^n \mathbb{E}|Y_{n,k}|\mathbf{1}_{\left\{\sum_{|j-k|\le b\log n, j\neq k} Y_{n,j}\neq 0\right\}} \le 2\sum_{k=1}^n \mathbb{E}|Y_{n,k}|\sum_{|j-k|\le b\log n, j\neq k}\mathbf{1}_{\{Y_{n,j}\neq 0\}} \le 4C_1^2 b\frac{\log n}{n}.$$

Further, by the triangle inequality

$$\bigg|\mathbb{E}\mathbf{i}Y_{n,k}e^{\mathbf{i}tY_{n,k}}\exp\Big(\mathbf{i}t\sum_{|j-k|>b\log n} Y_{n,j}\Big) - \dot{\phi}_{n,k}(t)\psi_n(t)\bigg| =$$
$$\bigg|\mathbb{E}\mathbf{i}Y_{n,k}e^{\mathbf{i}tY_{n,k}}\exp\Big(\mathbf{i}t\sum_{|j-k|>b\log n} Y_{n,j}\Big) - \mathbb{E}\mathbf{i}Y_{n,k}e^{\mathbf{i}tY_{n,k}}\mathbb{E}\exp\Big(\mathbf{i}t\sum_{j=1}^n Y_{n,j}\Big)\bigg| \le$$
$$\bigg|\mathbb{E}Y_{n,k}e^{\mathbf{i}tY_{n,k}}\exp\Big(\mathbf{i}t\sum_{|j-k|>b\log n} Y_{n,j}\Big) - \mathbb{E}Y_{n,k}e^{\mathbf{i}tY_{n,k}}\mathbb{E}\exp\Big(\mathbf{i}t\sum_{|j-k|>b\log n} Y_{n,j}\Big)\bigg| +$$
$$\Big|\mathbb{E}Y_{n,k}e^{\mathbf{i}tY_{n,k}}\Big|\bigg|\mathbb{E}\exp\Big(\mathbf{i}t\sum_{|j-k|>b\log n} Y_{n,j}\Big) - \mathbb{E}\exp\Big(\mathbf{i}t\sum_{j=1}^n Y_{n,j}\Big)\bigg| =: J_3 + J_4.$$

Similarly to (3.4), we have

$$|J_4| \le \mathbb{E}|Y_{n,k}|\mathbb{E}\bigg|\exp\Big(\mathbf{i}t\sum_{|j-k|>b\log n} Y_{n,j}\Big) - \exp\Big(\mathbf{i}t\sum_{j=1}^n Y_{n,j}\Big)\bigg| \le$$
$$2\mathbb{E}|Y_{n,k}|\mathbb{E}\mathbf{1}_{\left\{\sum_{|j-k|\le b\log n} Y_{n,j}\neq 0\right\}} \le 2\mathbb{E}|Y_{n,k}|\sum_{|j-k|\le b\log n}\mathbb{P}(Y_{n,j}\neq 0) \le 4C_1^2 b\frac{\log n}{n^2}.$$

For brevity, define

$$U := \exp\Big(\mathbf{i}t\sum_{j<k-b\log n} Y_{n,j}\Big), \quad V := Y_{n,k}e^{\mathbf{i}tY_{n,k}}, \quad W := \exp\Big(\mathbf{i}t\sum_{j>k+b\log n} Y_{n,j}\Big).$$

By the triangle inequality,

$$\big|\mathbb{E}UVW - \mathbb{E}V\mathbb{E}UW\big| \le \big|\mathbb{E}UVW - \mathbb{E}UV\mathbb{E}W\big| +$$
$$\big|\mathbb{E}UV\mathbb{E}W - \mathbb{E}U\mathbb{E}V\mathbb{E}W\big| + \big|\mathbb{E}U\mathbb{E}V\mathbb{E}W - \mathbb{E}V\mathbb{E}UW\big|.$$

Since U and V are $\mathcal{F}_k$-measurable, $|U| \le 1$, $|W| \le 1$ and $\mathbb{E}|V| \le C_1 n^{-1}$, A3 implies

$$\big|\mathbb{E}UVW - \mathbb{E}UV\mathbb{E}W\big| \le \mathbb{E}|UV|\big|\mathbb{E}(W|\mathcal{F}_k) - \mathbb{E}W\big| \le C_1 n^{-1}\alpha(b\log n),$$

and, since U is measurable with respect to $\mathcal{F}_{k-b\log n}$,

$$\big|\mathbb{E}U\mathbb{E}V\mathbb{E}W \;-\; \mathbb{E}V\mathbb{E}UW\big| \;\le\; |\mathbb{E}V|\mathbb{E}|U|\big|\mathbb{E}W \;-\; \mathbb{E}(W|\mathcal{F}_{k-b\log n})\big| \;\le\; C_1 n^{-1}\alpha(2b\log n).$$

Further, by A2 for $b\log n \ge \ell$,

$$\big|\mathbb{E}UV\mathbb{E}W - \mathbb{E}U\mathbb{E}V\mathbb{E}W\big| \le |\mathbb{E}W|\mathbb{E}|U|\big|\mathbb{E}(V|\mathcal{F}_{k-b\log n}) - \mathbb{E}V\big| \le C_1 n^{-1}\alpha(b\log n).$$

Hence

$$|J_3| = \Big|\mathbb{E}UVW - \mathbb{E}V\mathbb{E}UW\Big| \le 3C_1 n^{-1}\alpha(b\log n),$$

and consequently, by A4

$$\Big|J_1 - \mu\dot\varphi(t)\psi_n(t)\Big| = \Big|\sum_{k=1}^n \mathbb{E}\mathbf{i}Y_{n,k}e^{\mathbf{i}tY_{n,k}}\exp\Big(\mathbf{i}t\sum_{|j-k|>b\log n} Y_{n,j}\Big) - \mu\dot\varphi(t)\psi_n(t)\Big| \le$$
$$C_2 n^{-1} + 3C_1\alpha(b\log n) + 4C_1^2 b\frac{\log n}{n}.$$

Assembling all parts together, we obtain (3.2).

The bound (1.2) for the Lévy metric is obtained by means of Zolotorev's inequality [23],

$$L\Big(\mathcal{L}(S_n), \mathcal{L}(S)\Big) \le \frac{1}{\pi}\int_0^T \frac{|\psi_n(t) - \psi(t)|}{t}dt + 2e\frac{\log T}{T}, \quad T > 1.3,$$

which in view of the bound in (3.3) gives

$$L\Big(\mathcal{L}(S_n), \mathcal{L}(S)\Big) \le \frac{e^{2\mu}}{\pi}\Big(C_2 n^{-1} + 3C_1\alpha(b\log n) + 2C_1^2 b\frac{\log n}{n}\Big)T + 2e\frac{\log T}{T}. \tag{3.5}$$

If $\alpha(k)$ decays geometrically as in A5, the bound (1.2) is obtained by choosing $T = n^{1/2}$ and $b \ge \frac{1}{\log 1/r}$. $\square$

Remark 3.1. *The rate in* (1.2) *is not as sharp as the one, obtained by Tikhomirov in [20] in the CLT case. Apparently, the deficiency originates in the specific form of the compound Poisson characteristic function* $\psi(t) = e^{\mu(\varphi(t)-1)}$*, which does not vanish as* $t \to \infty$*. More specifically, the integration kernel* $K(s,t) := e^{\mu(\varphi(t)-\varphi(s))}$ *in* (3.1) *does not decay when* t *is fixed and* s *decreases, which contributes the linear growth in* t *of the right hand side of* (3.3) *and the corresponding linear growth in* T *in* (3.5)*. In the Gaussian case, this kernel has the form* $K(s,t) := e^{s^2/4-t^2/4}$ *(see eq. (3.25) page 809 in [20]), which yields better balance between growth in* t *and the decrease in* n*. It seems that in the compound Poisson setting under consideration the rate cannot be essentially improved within the framework of Tikhomirov's method.*

References

[1] M. Alvo, P. Cabilio, and P. D. Feigin, *A class of martingales with nonsymmetric limit distributions*, Z. Wahrsch. Verw. Gebiete **58** (1981), no. 1, 87–93. MR-635273

[2] R. Arratia, L. Goldstein, and L. Gordon, *Poisson approximation and the Chen-Stein method*, Statist. Sci. **5** (1990), no. 4, 403–434, With comments and a rejoinder by the authors. MR-1092983

[3] A. D. Barbour and O. Chryssaphinou, *Compound Poisson approximation: a user's guide*, Ann. Appl. Probab. **11** (2001), no. 3, 964–1002. MR-1865030

[4] S. M. Berman, *A compound Poisson limit for stationary sums, and sojourns of Gaussian processes*, Ann. Probab. **8** (1980), no. 3, 511–538. MR-573291

[5] Michael V. Boutsikas and Eutichia Vaggelatou, *A new method for obtaining sharp compound Poisson approximation error estimates for sums of locally dependent random variables*, Bernoulli **16** (2010), no. 2, 301–330. MR-2668903

[6] B. M. Brown and G. K. Eagleson, *Martingale convergence to infinitely divisible laws with finite variances*, Trans. Amer. Math. Soc. **162** (1971), 449–453. MR-0288806

[7] K. S. Chan, *Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model*, Ann. Statist. **21** (1993), no. 1, 520–533. MR-1212191

[8] Ngai Hang Chan and Yury A. Kutoyants, *Recent developments of threshold estimation for nonlinear time series*, J. Japan Statist. Soc. **40** (2010), no. 2, 277–303. MR-2830136

[9] P. Chigansky and Y. A. Kutoyants, *On nonlinear tar processes and threshold estimation*, Mathematical Methods of Statistics **21** (2012), no. 2, 142–152.

[10] A. L. Gibbs and F. E. Su, *On choosing and bounding probability metrics*, International Statistical Review **70** (2002), no. 3, 419–435.

[11] Bruce E. Hansen, *Threshold autoregression in economics*, Stat. Interface **4** (2011), no. 2, 123–127. MR-2812805

[12] I. A. Ibragimov and R. Z. Has′minskiĭ, *Statistical estimation*, Applications of Mathematics, vol. 16, Springer-Verlag, New York, 1981, Asymptotic theory, Translated from the Russian by Samuel Kotz. MR-620321

[13] J. Jacod and A. N. Shiryaev, *Limit theorems for stochastic processes*, second ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 288, Springer-Verlag, Berlin, 2003. MR-1943877

[14] Yury A. Kutoyants, *On identification of the threshold diffusion processes*, Ann. Inst. Statist. Math. **64** (2012), no. 2, 383–413. MR-2878912

[15] R. M. Meyer, *A Poisson-type limit theorem for mixing sequences of dependent "rare" events*, Ann. Probability **1** (1973), 480–483. MR-0350816

[16] Sean Meyn and Richard L. Tweedie, *Markov chains and stochastic stability*, second ed., Cambridge University Press, Cambridge, 2009, With a prologue by Peter W. Glynn. MR-2509253

[17] Bero Roos, *Kerstan's method for compound Poisson approximation*, Ann. Probab. **31** (2003), no. 4, 1754–1771. MR-2016599

[18] R. F. Serfozo, *Compound Poisson approximations for sums of random variables*, Ann. Probab. **14** (1986), no. 4, 1391–1398. MR-866359

[19] ______, *Correction to: "Compound Poisson approximations for sums of random variables" [Ann. Probab.* **14** *(1986), no. 4, 1391–1398; MR0866359 (88d:60019)]*, Ann. Probab. **16** (1988), no. 1, 429–430. MR-920283

[20] A. N. Tihomirov, *Convergence rate in the central limit theorem for weakly dependent random variables*, Teor. Veroyatnost. i Primenen. **25** (1980), no. 4, 800–818. MR-595140

[21] H. Tong, *Threshold models in nonlinear time series analysis*, Lecture Notes in Statistics, vol. 21, Springer-Verlag, New York, 1983. MR-717388

[22] ______, *Threshold models in time series analysis—30 years on*, Stat. Interface **4** (2011), no. 2, 107–118. MR-2812802

[23] V. M. Zolotarev, *Estimates of the difference between distributions in the Lévy metric*, Trudy Mat. Inst. Steklov. **112** (1971), 224–231, 388, Collection of articles dedicated to Academician Ivan Matveevič Vinogradov on his eightieth birthday, I. MR-0321156

Acknowledgments. The authors are grateful to Y. Kutoyants and R. Liptser for the enlightening discussions on the subject of this article. We also appreciate referee's suggestion, which led to the ultimate improvement of the rate in (1.2).