Electronic Journal of Probability

Vol. 9 (2004), Paper no. 2, pages 18–36.

Journal URL
http://www.math.washington.edu/~ejpecp/

A Non-uniform Bound for Translated Poisson Approximation

A. D. Barbour [1]
Angewandte Mathematik, Universität Zürich
Winterthurerstrasse 190, CH-8057, Switzerland
textttadb@amath.unizh.ch

Kwok Pui Choi [2]
Department of Mathematics, National University of Singapore
Singapore 117543
matckp@nus.edu.sg

Abstract: Let $X_1, \ldots, X_n$ be independent, integer valued random variables, with p^{th} moments, $p > 2$, and let W denote their sum. We prove bounds analogous to the classical non-uniform estimates of the error in the central limit theorem, but now, for approximation of $\mathcal{L}(W)$ by a translated Poisson distribution. The advantage is that the error bounds, which are often of order no worse than in the classical case, measure the accuracy in terms of total variation distance. In order to have good approximation in this sense, it is necessary for $\mathcal{L}(W)$ to be sufficiently smooth; this requirement is incorporated into the bounds by way of a parameter α, which measures the average overlap between $\mathcal{L}(X_i)$ and $\mathcal{L}(X_i+1)$, $1 \le i \le n$.

Keywords and phrases: Stein's method, non-uniform bounds, translated Poisson approximation, total variation.

AMS subject classification (2000): 62E17; 60F05.

Submitted to EJP on September 1, 2003. Final version accepted on January 6, 2004.

[1] Work supported in part by Schweizerischer Nationalfonds Projekt Nr. 20-67909.02, and by the Institute for Mathematical Sciences of the National University of Singapore.

[2] Work supported in part by BMRC Research Grant BMRC01/1/21/19/140. Part of the work was done while the author was visiting the Institute for Mathematical Sciences, National University of Singapore in 2003.

1 Introduction

Let $X_1, \ldots, X_n$ be independent, integer valued random variables. For each $1 \leq i \leq n$, we centre X_i by requiring that $0 \leq \mathbb{E}X_i = \mu_i < 1$, and we assume that $\mathbb{E}|X_i|^p < \infty$, where $p \in (2, \infty)$ is the same for all i; we write $\mathrm{Var}\, X_i = \sigma_i^2$. The Berry–Esseen theorem then bounds the error in the approximation to $W := \sum_{i=1}^n X_i$ by a normal distribution with the same mean and variance as W:

$$\sup_{x \in \mathbb{R}} |\mathbb{P}[W - \mu \leq x\sigma] - \Phi(x)| \leq C\Gamma_{3 \wedge p}, \tag{1.1}$$

where $\mu := \sum_{i=1}^n \mu_i = \mathbb{E}W$, $\sigma^2 := \sum_{i=1}^n \sigma_i^2 = \mathrm{Var}\, W$, $\Gamma_s := \sigma^{-s} \sum_{i=1}^n \mathbb{E}\{|X_i - \mu_i|^s\}$ for any $0 < s \leq p$, and Φ denotes the standard normal distribution function. However, for integer valued random variables, approximation by a distribution which, like the normal, has a smooth probability density is not necessarily the most natural choice. It was shown using Stein's method in Barbour and Xia (1999) and in Čekanavičius and Vaitkus (2001) that there are probability distributions on the integers, the simplest of which is the translated Poisson distribution, which allow an approximation whose error, with respect to the stronger, total variation distance, is of essentially the same order (Barbour and Xia 1999, Corollary 4.5) as in (1.1); in such results, an extra condition is required to exclude the possibility that W is nearly concentrated on a lattice of span greater than 1 and this is also reflected in the bounds obtained. Analogous asymptotic expansions, with error again measured in total variation norm, were established in Barbour and Čekanavičius (2002), Theorem 5.1.

In this paper, we also consider approximation in total variation, but in a non–uniform sense. For sets $A \subset [\mu + x\sigma, \infty) \cap \mathbb{Z}$, $x > 0$, we show in Theorem 2.1 that, under mild conditions, the error in approximating the probability $\mathbb{P}[W \in A]$ by the corresponding probability for a suitably chosen translated Poisson distribution is of order $O(\{\Gamma_{3 \wedge p} + \Gamma_p\}(1 + x^p)^{-1})$, becoming smaller as x increases. This result is a natural analogue of the non–uniform bounds for the error in the usual normal approximation (Bikelis (1966), Petrov (1975, Theorem 13, p.125) , Chen and Shao (2001)), but now once again with respect to total variation distance. The translated Poisson distribution is chosen to have the same mean as W, and also to have variance close to that of W; because only translations by integer amounts are appropriate, an exact match of both moments cannot usually be achieved.

We prove our result using Stein's method. We are able to make use of much of the standard theory associated with Poisson approximation by way of the Stein–Chen approach, but there are a number of significant differences. First, we show that the solutions to the Stein equation for sets A as above take very small values for arguments $k \ll \mu + x\sigma$, as do their first and second differences. For values of k comparable to $\mu + x\sigma$, there are no better bounds than the standard bounds used in the Stein–Chen approach, and in order to get results of the required accuracy, it is necessary instead to use some smoothness of the distribution of W, expressed in the form of a bound on $d_{TV}(\mathcal{L}(W), \mathcal{L}(W + 1))$. Here, the procedure is much more delicate than for uniform bounds, since it must be shown that this smoothness is preserved well enough into the tails of the distribution of W. We do so using an argument based on the Mineka coupling, in Lemma 3.5.

There is another approach to total variation approximation for sums of integer valued random variables, that of Chen and Suan (2003). Their argument, although based on Stein's method, is entirely different. They take Stein's method for normal approximation as their starting point, and give bounds on approximation with a discretized normal distribution. It seems likely that their approach will yield results comparable to ours in the setting of this paper; in the context of sums of dependent random variables, the existence of the two different methods may prove to be very useful.

2 Notation and results

We begin by specifying the translated Poisson distribution to be used in our approximation. To do this, we first define $0 \leq \delta < 1$ by

$$\delta := (\mu - \sigma^2) - \lfloor \mu - \sigma^2 \rfloor.$$

We then set $\gamma := \lfloor \mu - \sigma^2 \rfloor = \mu - \sigma^2 - \delta$, and note that $W - \gamma$ is an integer valued random variable whose mean and variance are almost equal: indeed, $\mathbb{E}(W - \gamma) - \mathrm{Var}\,(W - \gamma) = \delta$. This makes $W - \gamma$ a good candidate for Poisson approximation; our choice is to approximate it by the Poisson distribution $\mathrm{Po}\,(\lambda)$ with mean $\lambda := \sigma^2 + \delta$, having the same mean as $W - \gamma$, and variance larger by the amount δ.

We also need some way to ensure that the distribution of W has sufficient smoothness; clearly, if the distribution of W is concentrated on the odd integers, *total variation* approximation by a translated Poisson distribution has no hope of success, even though approximation with respect to Kolmogorov distance may be good. We therefore define

$$d_i := d_{TV}(\mathcal{L}(X_i), \mathcal{L}(X_i + 1)), \tag{2.1}$$

and set

$$\alpha := n^{-1} \sum_{i=1}^{n} (1 - d_i). \tag{2.2}$$

The quantity $1 - d_i$ measures the overlap of the distribution of X_i and its translate by 1, and the average value α of this overlap appears as an ingredient in the bounds; the larger the value of α, the smoother the distribution of W. In order to get errors of approximation of the same order as the classical results for normal approximation, the quantity α should not become small with n. Note that, by combining successive summands, it can typically be arranged that $d_i < 1$ for all i, though the value of n is then correspondingly reduced. For neatness, we assume that $\lambda \geq n$, which can again be achieved by grouping summands if need be.

Now consider the Stein equation for Poisson approximation with $\mathrm{Po}\,(\lambda)$. For each $A \subset \mathbb{Z}_+$, there exists a unique function $g_A : \mathbb{Z}_+ \to \mathbb{R}$ such that

$$\lambda g_A(w + 1) - w g_A(w) = \mathbf{1}_A(w) - \mathrm{Po}\,(\lambda)(A), \quad w \geq 0 \tag{2.3}$$

with $g_A(0) = 0$, which satisfies $\|\Delta g_A\| \leq \min\{1, \lambda^{-1}\}$, and which, for $w \geq 1$, is given by

$$\begin{aligned} g_A(w) &= -\frac{\mathbb{E}\{[\mathbf{1}_A(Z) - \mathrm{Po}\,(\lambda)(A)] I[Z \geq w]\}}{\lambda \mathbb{P}[Z = w - 1]} && (2.4) \\ &= \frac{\mathbb{E}\{[\mathbf{1}_A(Z) - \mathrm{Po}\,(\lambda)(A)] I[Z \leq w - 1]\}}{\lambda \mathbb{P}[Z = w - 1]}, && (2.5) \end{aligned}$$

where $Z \sim \mathrm{Po}\,(\lambda)$, $\mathrm{Po}\,(\lambda)(A) = P(Z \in A)$ and $\mathrm{Po}\,(\lambda)\{m\} = P(Z = m)$; see Barbour, Holst and Janson (1992, Chapter 1, (1.18)). We extend the definition of g_A to the whole of $\mathbb{Z}$ by setting $g_A(w) = 0$ for $w < 0$; hence

$$\lambda g_A(w+1) - w g_A(w) = \{\mathbf{1}_A(w) - \mathrm{Po}\,(\lambda)(A)\} I[w \geq 0], \quad w \in \mathbb{Z}.$$

We also define h_A by

$$h_A(w) = \begin{cases} g_A(w-\gamma) & w \geq \gamma, \\ 0 & w < \gamma, \end{cases} \tag{2.6}$$

so that

$$\lambda h_A(w+1) - (w-\gamma) h_A(w) = [\mathbf{1}_A(w-\gamma) - \mathrm{Po}\,(\lambda)(A)] I[w \geq \gamma], \quad w \in \mathbb{Z},$$

noting that $\|\Delta h_A\| \leq \min\{1, \lambda^{-1}\}$. For $A \subset \mathbb{Z}_+$, it then follows that

$$\begin{aligned} & \mathbb{P}[W-\gamma \in A] - \mathrm{Po}\,(\lambda)(A) \\ &= \mathbb{E}\{\mathbf{1}_A(W-\gamma) - \mathrm{Po}\,(\lambda)(A)\} \\ &= \mathbb{E}\{[\mathbf{1}_A(W-\gamma) - \mathrm{Po}\,(\lambda)(A)] I[W \geq \gamma]\} - \mathrm{Po}\,(\lambda)(A)\mathbb{P}[W < \gamma] \\ &= \mathbb{E}\{[\lambda h_A(W+1) - (W-\gamma) h_A(W)]\} - \mathrm{Po}\,(\lambda)(A)\mathbb{P}[W < \gamma]. \end{aligned}$$

Recalling the definitions of λ and γ, we now have

$$\begin{aligned} & \mathbb{E}[\lambda h_A(W+1) - (W-\gamma) h_A(W)] \\ &= \mathbb{E}[\sigma^2 \Delta h_A(W) - (W-\mu) h_A(W)] + \delta \mathbb{E}\Delta h_A(W) \\ &= \sum_{i=1}^n \left[\sigma_i^2 \mathbb{E}\Delta h_A(W) - \mathbb{E}\{(X_i - \mu_i)(h_A(W) - h_A(W_i))\}\right] + \delta \mathbb{E}\Delta h_A(W), \end{aligned} \tag{2.7}$$

where $W_i := W - X_i$ is independent of X_i. We write Newton's expansion in the following form: for $l \in \mathbb{Z}$,

$$f(w+l) = f(w) + \sum_{j \in \mathbb{Z}} K(j,l) \Delta f(w+j) \tag{2.8}$$

$$= f(w) + l\Delta f(w) + \sum_{j \in \mathbb{Z}} (l-1-j) K(j,l) \Delta^2 f(w+j), \tag{2.9}$$

where

$$K(j,l) := I[0 \leq j \leq l-1] - I[l \leq j \leq -1]. \tag{2.10}$$

By taking $f = \Delta h_A$, $w = W_i$ and $l = X_i$ in (2.8), and by the independence of W_i and X_i, we have

$$\sigma_i^2 \mathbb{E}\Delta h_A(W) = \sigma_i^2 \mathbb{E}\Delta h_A(W_i) + \sigma_i^2 \sum_{j \in \mathbb{Z}} \mathbb{E}K(j, X_i)\ \mathbb{E}\Delta^2 h_A(W_i + j),$$

and similarly, with $f = h_A$ in (2.9) and (2.8),

$$\begin{aligned}
&(X_i-\mu_i)\{h_A(W)-h_A(W_i)\} \\
&\quad = (X_i-\mu_i)X_i\Delta h_A(W_i) + \sum_{j\in\mathbb{Z}}(X_i-\mu_i)(X_i-1-j)K(j,X_i)\Delta^2 h_A(W_i+j) \qquad (2.11)\\
&\quad = (X_i-\mu_i)X_i\Delta h_A(W_i) + (X_i-\mu_i)\left(\sum_{j\in\mathbb{Z}}K(j,X_i)\Delta h_A(W_i+j) - X_i\Delta h_A(W_i)\right),
\end{aligned}$$

so that, in particular,

$$\begin{aligned}
&\mathbb{E}\{(X_i-\mu_i)[h_A(W)-h_A(W_i)]\} \\
&\quad = \mathbb{E}\{(X_i-\mu_i)X_i\}\,\mathbb{E}\Delta h_A(W_i) + \sum_{j\in\mathbb{Z}}\mathbb{E}[(X_i-\mu_i)(X_i-1-j)K(j,X_i)]\,\mathbb{E}\Delta^2 h_A(W_i+j)\\
&\quad = \sigma_i^2\mathbb{E}\Delta h_A(W_i) + \sum_{j\in\mathbb{Z}}\mathbb{E}[(X_i-\mu_i)(X_i-1-j)K(j,X_i)]\,\mathbb{E}\Delta^2 h_A(W_i+j).
\end{aligned}$$

After putting all the pieces together, we thus have

$$\begin{aligned}
&\mathbb{P}[W-\gamma\in A] - \mathrm{Po}\,(\lambda)(A)\\
&\quad = \sum_{i=1}^n\sum_{j\in\mathbb{Z}}\mathbb{E}\{[\sigma_i^2-(X_i-\mu_i)(X_i-1-j)]K(j,X_i)\}\,\mathbb{E}\Delta^2 h_A(W_i+j) \qquad (2.12)\\
&\qquad +\delta\mathbb{E}\Delta h_A(W) - \mathrm{Po}\,(\lambda)(A)\mathbb{P}[W<\gamma].
\end{aligned}$$

This, and a similar expression based on (2.11), are used to prove the following theorem.

Theorem 2.1 *With the assumptions and definitions preceding (2.2), and supposing also that* $\Gamma_p \le 1$ *and that* $\sqrt{n\alpha}\exp\{-\frac{1}{2}n\alpha(1-\log 2)\} \le (1+2n)^{-p}$, *we have*

$$\sup_{A\subset[\lambda+x\sqrt{\lambda},\infty)}|\mathbb{P}[W-\gamma\in A]-\mathrm{Po}\,(\lambda)(A)| \le C_{2.1}(p)\frac{\Gamma_{3\wedge p}+\Gamma_p}{(1+x^p)}\left\{\sqrt{\frac{\lambda}{n\alpha}}\right\},$$

for some universal constant $C_{2.1}(p)$, *uniformly in* $x \ge 8$ *and* $\lambda \ge n$.

Remark. In proving Theorem 2.1 we have not attempted to take advantage of a situation where the random variables X_i take mostly the value 0 and only occasionally other values, as was done, for instance in Barbour and Xia (1999).

3 Bounding the differences

Our main tool for proving Theorem 2.1 is expression (2.12). In order to exploit it, we first need to bound the differences $\mathbb{E}\Delta h_A(W)$ and $\mathbb{E}\Delta^2 h_A(W_i+j)$, $1 \le i \le n$. We devote this section to doing so. We shall from now on consider only subsets $A \subset [\lambda + x\sqrt{\lambda},\infty)$ with $x \ge 8$, setting

$$\lambda(x) := \left\lfloor \lambda + \tfrac{1}{2}x\sqrt{\lambda}\right\rfloor; \qquad (3.1)$$

when bounding $\mathbb{E}\Delta^2 h_A(W_i+j)$, we restrict attention to the range $-1 \le j \le \sqrt{\lambda}+1$.

We proceed by way of a series of lemmas, starting with the simple observation that, if $r \ge m := \lambda + \xi\sqrt{\lambda}$, $\xi \ge 1$, then

$$\frac{\mathrm{Po}\,(\lambda)(r)}{\mathrm{Po}\,(\lambda)(m)} \le (1+\xi/\sqrt{\lambda})^{-(r-m)}. \tag{3.2}$$

Throughout the section, the constant C may change from one use to another.

Lemma 3.1 *We have*

$$\mathrm{Po}\,(\lambda)(A) \le \sqrt{\lambda}\, q_0(x,\lambda)\mathrm{Po}\,(\lambda)\{\lambda(x)\},$$

where

$$\begin{aligned} q_0(x,\lambda) &:= x^{-1}(2+x/\sqrt{\lambda})\left(1-\frac{x}{2\sqrt{\lambda}+x}\right)^{x\sqrt{\lambda}/2} \\ &\le \begin{cases} 3x^{-1}e^{-x^2/6}, & 0 < x \le \sqrt{\lambda}; \\ 3\lambda^{-1/2}e^{-x\sqrt{\lambda}/6}, & x > \sqrt{\lambda}. \end{cases} \end{aligned}$$

Note that there exist constants C_p, $p \in (2,\infty)$, such that

$$(1+x^2)q_0(x,\lambda) \le C_p/(1+x^p), \tag{3.3}$$

uniformly in $x \ge 0$ and $\lambda \ge 1$.

Proof. With $\nu(x) := \lceil \lambda + x\sqrt{\lambda}\rceil$, we have

$$\begin{aligned} \mathrm{Po}\,(\lambda)(A) \le \mathrm{Po}\,(\lambda)([\nu(x),\infty)) &= \sum_{r=\nu(x)}^{\infty} \mathrm{Po}\,(\lambda)\{r\} \\ &\le \mathrm{Po}\,(\lambda)\{\lambda(x)\} \sum_{r\ge\nu(x)} [1+x/(2\sqrt{\lambda})]^{-(r-\lambda(x))} \\ &\le x^{-1}(2\sqrt{\lambda}+x)\mathrm{Po}\,(\lambda)\{\lambda(x)\}\left(1+\frac{x}{2\sqrt{\lambda}}\right)^{-x\sqrt{\lambda}/2}, \end{aligned}$$

which is the statement of the lemma. □

Lemma 3.2 *Let* $G(j) := \mathrm{Po}\,(\lambda)([0,j])/\mathrm{Po}\,(\lambda)\{j\}$. *Then, for any* $r \ge 1$,

$$\Delta^r G(j) = \frac{\mathbb{E}\{I[Z \le j](j+1-Z)\cdots(j+r-Z)\}}{\lambda^r \mathrm{Po}\,(\lambda)\{j\}} > 0,$$

where $\Delta^r G$ *is the* r*th forward difference of* G *and* $Z \sim \mathrm{Po}\,(\lambda)$.

Proof. For $r = 1$, we calculate

$$\begin{aligned}
\Delta G(j) &= \frac{\mathrm{Po}\,(\lambda)([0,j+1])}{\mathrm{Po}\,(\lambda)\{j+1\}} - \frac{\mathrm{Po}\,(\lambda)([0,j])}{\mathrm{Po}\,(\lambda)\{j\}} \\
&= \frac{(j+1)\mathrm{Po}\,(\lambda)([0,j+1]) - \lambda\mathrm{Po}\,(\lambda)([0,j])}{\lambda\mathrm{Po}\,(\lambda)\{j\}} \\
&= \frac{\mathbb{E}\{I[Z \le j+1](j+1-Z)\}}{\lambda\mathrm{Po}\,(\lambda)\{j\}} \\
&= \frac{\mathbb{E}\{I[Z \le j](j+1-Z)\}}{\lambda\mathrm{Po}\,(\lambda)\{j\}},
\end{aligned}$$

where we use the Stein identity $\lambda\mathbb{E}f(Z+1) = \mathbb{E}Zf(Z)$ in the third equality. Induction combined with analogous manipulations finishes the proof. □

Lemma 3.3 *If $-1 \le w - \gamma \le \lambda(x)$, then*

$$|h_A(w)| \le \frac{1}{\sqrt{\lambda}} q_0(x,\lambda); \tag{3.4}$$

$$|\Delta h_A(w)| \le \frac{1}{\lambda}(1+x)q_0(x,\lambda); \tag{3.5}$$

$$|\Delta^2 h_A(w)| \le \frac{1}{\lambda^{3/2}}(1+x^2)q_0(x,\lambda), \tag{3.6}$$

uniformly in $\lambda \ge 1$ and $x \ge 4$.

Proof. Recalling (2.6) and (2.5), and writing $k = w - \gamma$ for $w - \gamma \le \lambda(x)$, we have $h_A(w) = -\mathrm{Po}\,(\lambda)(A)G(k-1)/\lambda$, and hence, for $r \ge 0$,

$$\begin{aligned}
|\Delta^r h_A(w)| &= \frac{\mathrm{Po}\,(\lambda)(A)}{\lambda}\Delta^r G(k) = \frac{\mathrm{Po}\,(\lambda)(A)}{\lambda}\left[\Delta^r G(k+1) - \Delta^{r+1}G(k)\right] \\
&\le \frac{\mathrm{Po}\,(\lambda)(A)}{\lambda}\Delta^r G(k+1),
\end{aligned}$$

from Lemma 3.2. To deduce (3.4)–(3.6), we thus only need to bound $\Delta^r G(\lambda(x))$ for $r = 0, 1$ and 2.

For $r = 0$, we apply Lemma 3.1 to get

$$\frac{\mathrm{Po}\,(\lambda)(A)}{\lambda}G(\lambda(x)) \le \frac{q_0(x,\lambda)\mathrm{Po}\,(\lambda)([0,\lambda(x)])}{\sqrt{\lambda}} \le \frac{1}{\sqrt{\lambda}}q_0(x,\lambda).$$

For $r = 1$, from Lemma 3.2, we have

$$\begin{aligned}
&\frac{\mathrm{Po}\,(\lambda)(A)}{\lambda}\Delta G(\lambda(x)) \\
&\quad \le \frac{q_0(x,\lambda)\mathbb{E}\{I[Z \le \lambda(x)](\lambda(x)+1-Z)\}}{\lambda^{3/2}} \\
&\quad = \lambda^{-3/2}q_0(x,\lambda)\left\{\mathbb{E}(\lambda(x)+1-Z) + \mathbb{E}\left\{(Z-\lambda(x)-1)I[Z>\lambda(x)]\right\}\right\} \\
&\quad = \frac{1}{\lambda}q_1(x,\lambda),
\end{aligned}$$

where

$$q_1(x,\lambda) = \frac{xq_0(x,\lambda)}{2}\left\{1 + 2\frac{1+\mathbb{E}\{(Z-\lambda(x)-1)I[Z>\lambda(x)]\}}{x\sqrt{\lambda}}\right\}.$$

Since, for $m \geq \lambda$,

$$\begin{aligned}\mathbb{E}\{(Z-m)I[Z\geq m]\} &= \lambda \mathrm{Po}\,(\lambda)([m-1,\infty)) - m\mathrm{Po}\,(\lambda)([m,\infty)) \\ &\leq \lambda \mathrm{Po}\,(\lambda)\{m-1\} \leq \sqrt{\lambda/2e},\end{aligned}$$

where we use a result of Barbour and Jensen (1989, p.78) in the last inequality, it follows that

$$q_1(x,\lambda) \leq \tfrac{1}{2}q_0(x,\lambda)\{x+3\} \leq (1+x)q_0(x,\lambda)$$

in $x \geq 1$. Finally, for $r = 2$, again from Lemma 3.2, we have

$$\begin{aligned}\frac{\mathrm{Po}\,(\lambda)(A)}{\lambda}\Delta^2 G(\lambda(x)) &\leq \frac{q_0(x,\lambda)\mathbb{E}\{I[Z\leq\lambda(x)](\lambda(x)+1-Z)(\lambda(x)+2-Z)\}}{\lambda^{5/2}} \\ &\leq \lambda^{-5/2}q_0(x,\lambda)\mathbb{E}\{(\lambda(x)+1-Z)(\lambda(x)+2-Z)\} \\ &= \frac{1}{\lambda^{3/2}}q_2(x,\lambda),\end{aligned}$$

where we have used $(\lambda(x)+1-Z)(\lambda(x)+2-Z) > 0$ with probability 1, and thus

$$q_2(x,\lambda) \leq q_0(x,\lambda)[\lambda + (2+x\sqrt{\lambda})(4+x\sqrt{\lambda})/4]/\lambda \leq (1+x^2)q_0(x,\lambda).$$

□

Lemma 3.4 *If $\Gamma_p \leq 1$, there is a constant $C_1 < \infty$ such that*

$$\max\left\{|\mathbb{E}\Delta h_A(W)|, \max_{1\leq l\leq n}\sup_{j\leq\sqrt{\lambda}+1}|\mathbb{E}\Delta h_A(W_l+j)|\right\} \leq \frac{C_1}{\lambda(1+x^p)},$$

uniformly for all $A \subset [\lambda + x\sqrt{\lambda}, \infty)$ and $x \geq 8$.

Proof. Since $\Delta h_A(w) = 0$ for $w - \gamma < 0$ and since $\|\Delta h_A\| \leq \lambda^{-1}$, it follows from (3.5) that

$$\begin{aligned}|\mathbb{E}\Delta h_A(W)| &\leq \mathbb{E}|\Delta h_A(W)I[0\leq W-\gamma\leq\lambda(x)]| + \mathbb{E}|\Delta h_A(W)I[W-\gamma>\lambda(x)]| \\ &\leq \lambda^{-1}(1+x)q_0(x,\lambda) + \lambda^{-1}\mathbb{P}[W-\gamma>\lambda(x)] \\ &\leq \lambda^{-1}(1+x)q_0(x,\lambda) + \lambda^{-1}\mathbb{P}[W-\mu\geq\tfrac{1}{2}x\sqrt{\lambda}] \\ &\leq \frac{1}{\lambda}(1+x)q_0(x,\lambda) + \frac{2^p\mathbb{E}|W-\mu|^p}{x^p\lambda^{1+p/2}},\end{aligned}$$

the last term coming from Chebyshev's inequality. Applying Rosenthal's inequality to the final term, recalling (3.3) and remembering that $\Gamma_p \leq 1$, completes the bound for $|\mathbb{E}\Delta h_A(W)|$.

For the remaining elements, note that, in $x \geq 8$ and $\lambda \geq 1$, it follows that $\sqrt{\lambda}+1 \leq \frac{1}{4}x\sqrt{\lambda}$. The previous argument can thus be used, with $W_l + j$ replacing W throughout, and hence $\mathbb{P}[W_l - \mathbb{E}W_l \geq \frac{1}{4}x\sqrt{\lambda}]$ replacing $\mathbb{P}[W-\mu\geq\frac{1}{2}x\sqrt{\lambda}]$ leading to the bound

$$|\mathbb{E}\Delta h_A(W_l+j)| \leq \lambda^{-1}(1+x)q_0(x,\lambda) + 4^p x^{-p}\lambda^{-(1+p/2)}c_p\sigma^p(1+\Gamma_p),$$

uniformly in $1 \leq l \leq n$ and in $j \leq \sqrt{\lambda}+1$; c_p is a constant implied by the Rosenthal inequality. The lemma follows. □

Lemma 3.5 *If* $\Gamma_p \le 1$ *and*

$$\sqrt{n\alpha}e^{-\frac{1}{2}n\alpha(1-\log 2)} \le (1+2n)^{-p},$$

then there is a constant $C_2 < \infty$ *such that*

$$|\mathbb{E}\Delta^2 h_A(W_l + j)| \le \frac{C_2}{\lambda\sqrt{n\alpha}(1+x^p)},$$

uniformly for all $A \subset [\lambda + x\sqrt{\lambda}, \infty)$, $8 \le x \le 2n$, $j \le 1 + \sqrt{\lambda}$ *and* $1 \le l \le n$.

Proof. Observing that $\Delta h_A(w) = 0$ for $w - \gamma < -1$ and that $j \le \sqrt{\lambda} + 1 \le \frac{1}{4}x\sqrt{\lambda}$, we have

$$\begin{aligned} |\mathbb{E}\Delta^2 h_A(W+j)| &\le \mathbb{E}|\Delta^2 h_A(W+j) I[-1 \le W + j - \gamma \le \lambda(x)]| + |R| \\ &\le \lambda^{-3/2}(1+x^2)q_0(x,\lambda) + |R|, \end{aligned} \tag{3.7}$$

where

$$\begin{aligned} R &:= \mathbb{E}\{[\Delta h_A(W+j+1) - \Delta h_A(W+j)]I[W+j-\gamma > \lambda(x)]\} \\ &= \mathbb{E}\{\Delta h_A(W+j+1)I[W+j+1 > 1+\gamma+\lambda(x)] \\ &\qquad -\Delta h_A(W+j)I[W+j > \gamma + \lambda(x)]\} \\ &= \mathbb{E}\big\{\Delta h_A(W+j+1)I[W+j+1 > 1+\gamma+\lambda(x)] \\ &\qquad -\Delta h_A(W+j)I[W+j > 1+\gamma+\lambda(x)]\big\} \\ &\qquad -\Delta h_A(1+\gamma+\lambda(x))\mathbb{P}[W+j = 1+\gamma+\lambda(x)]. \end{aligned} \tag{3.8}$$

In order to estimate the first term in (3.8), we use the Mineka coupling (Lindvall 1992, Section II.14). We construct two partial sum processes $S = S^{(j)}$ and $S' = S'^{(j)}$ having $S_0 = j$, $S_0^{'} = j+1$ and, for $1 \le k \le n$, $S_k = j + \sum_{i=1}^k Y_i$ and $S_k^{'} = j + 1 + \sum_{i=1}^k Y_i^{'}$, where the pairs $\{(Y_i, Y_i'), 1 \le i \le n\}$ form an independent sequence, and $\mathcal{L}(Y_i) = \mathcal{L}(Y_i^{'}) = \mathcal{L}(X_i)$; thus

$$\mathcal{L}(W+j+1) = \mathcal{L}(S_n^{'}) \quad \text{and} \quad \mathcal{L}(W+j) = \mathcal{L}(S_n).$$

Letting

$$T = \min\{k : S_k = S_k^{'}\}$$

be the first time that S and S' meet, it is then immediate that

$$\mathcal{L}(S_n{}^{(j)} \,|\, T \le n) = \mathcal{L}(S_n'{}^{(j)} \,|\, T \le n),$$

implying that

$$\begin{aligned} &\mathbb{E}\big\{\Delta h_A(W+j+1)I[W+j+1 > 1+\gamma+\lambda(x)] - \Delta h_A(W+j)I[W+j > 1+\gamma+\lambda(x)]\big\} \\ &= \mathbb{E}\Big\{\Big(\Delta h_A(S_n'{}^{(j)})I[S_n'{}^{(j)} > 1+\gamma+\lambda(x)] \\ &\qquad -\Delta h_A(S_n{}^{(j)})I[S_n{}^{(j)} > 1+\gamma+\lambda(x)]\Big)\, I[T > n]\Big\}, \end{aligned} \tag{3.9}$$

of essential importance when bounding (3.8).

The random variables Y_i and Y_i' are coupled as follows. Setting

$$\theta_{is} := \tfrac{1}{2}\min\{\mathbb{P}[X_i = s], \mathbb{P}[X_i = s+1]\},$$

we want to arrange each pair (Y_i, Y_i') in such a way that, for each $s \in \mathbf{Z}$,

$$\begin{aligned} \mathbb{P}[(Y_i, Y_i') = (s, s+1)] &= \mathbb{P}[(Y_i, Y_i') = (s+1, s)] = \theta_{is}; \\ \mathbb{P}[(Y_i, Y_i') = (s, s)] &= \mathbb{P}[X_i = s] - (\theta_{i,s-1} + \theta_{i,s}). \end{aligned}$$

We do this by first defining independent indicator random variables I_i with

$$\mathbb{P}[I_i = 1] = 1 - \mathbb{P}[I_i = 0] = 1 - d_i,$$

where d_i is as in (2.1): note also that

$$d_i := 1 - 2\sum_{s\in\mathbf{Z}} \theta_{is}.$$

We then define three mutually independent sequences of independent random variables Z_i, $\widetilde{Z}_i$ and ε_i, $1 \le i \le n$, independent also of the I_i, with

$$\begin{aligned} \mathbb{P}[Z_i = s + \tfrac{1}{2}] &= 2\theta_{is}/(1 - d_i); \\ \mathbb{P}[\widetilde{Z}_i = s] &= \{\mathbb{P}[X_i = s] - (\theta_{i,s-1} + \theta_{i,s})\}/d_i, \end{aligned}$$

for each $s \in \mathbb{Z}$ (take $\mathbb{P}[Z_i = 0] = 1$ if $d_i = 1$), and with each ε_i taking the values ± 1 with probability one half. If $I_i = 1$, we set

$$(Y_i, Y_i') = (Z_i - \tfrac{1}{2}\varepsilon_i, Z_i + \tfrac{1}{2}\varepsilon_i),$$

and if $I_i = 0$, we set

$$(Y_i, Y_i') = (\widetilde{Z}_i, \widetilde{Z}_i).$$

With this construction, the event $\{T > n\}$ is realized by the event

$$E := \left\{ \min_{1\le i\le n} \sum_{l=1}^{i} I_l \varepsilon_l \ge 0 \right\}.$$

We write $J := \{1 \le i \le n : I_i = 1\}$, and note that, conditional on J, the event E is independent of $\{(Z_i, \widetilde{Z}_i),\, 1 \le i \le n\}$.

We now use these considerations to bound (3.9). Recalling that $\|\Delta h_A\| \le \lambda^{-1}$ and noting that $\sqrt{\lambda} + 1 \le \frac{1}{4}x\sqrt{\lambda}$ in $x \ge 8$ and $\lambda \ge 1$, it follows for any $j \le \sqrt{\lambda} + 1$ that, again writing S_n' for $S_n'^{(j)}$,

$$\begin{aligned} &\left|\mathbb{E}\left\{\Delta h_A(S_n')\, I[S_n' > 1 + \gamma + \lambda(x)]\, I[T > n] \,|\, J\right\}\right| \\ &\quad\le \lambda^{-1}\mathbb{E}\left\{ I[E]\, I\left[\sum_{i\in J}(Z_i + \tfrac{1}{2}\varepsilon_i) + \sum_{j\notin J} \widetilde{Z}_i > 1 + \gamma + \lambda + \tfrac{1}{4}x\sqrt{\lambda} \right] \,\middle|\, J \right\} \\ &\quad\le \lambda^{-1}\mathbb{E}\left\{ I[E]\, I\left[\sum_{i\in J}(Z_i + \tfrac{1}{2}\tilde{\varepsilon}_i) + \sum_{j\notin J} \widetilde{Z}_i > 1 + \gamma + \lambda + \tfrac{1}{8}x\sqrt{\lambda} \right] \,\middle|\, J \right\} \\ &\qquad + \lambda^{-1}\mathbb{E}\left\{ I[E]\, I\left[\tfrac{1}{2}\sum_{i\in J}(\varepsilon_i - \tilde{\varepsilon}_i) > \tfrac{1}{8}x\sqrt{\lambda} \right] \,\middle|\, J \right\}, \end{aligned} \tag{3.10}$$

where the $(\tilde\varepsilon_i, 1 \le i \le n)$ have the same distribution as the ε_i's, but are independent of everything else. They are introduced so that conditional expectation in the first term of (3.10) factorizes. As a result, we find that

$$\begin{aligned}
&\left| \mathbb{E}\left\{ \Delta h_A(S_n')\, I[S_n' > 1+\gamma+\lambda(x)]\, I[T>n] \,\middle|\, J \right\} \right| \\
&\quad \le \lambda^{-1}\mathbb{P}[E \,|\, J] \left\{ \mathbb{P}\left[\sum_{i=1}^n X_i > 1+\gamma+\lambda+\tfrac18 x\sqrt{\lambda} \,\middle|\, J \right] \right. \\
&\qquad \left. + \mathbb{P}\left[\sum_{i\in J} \tilde\varepsilon_i > \tfrac18 x\sqrt{\lambda} \,\middle|\, J \right] + \mathbb{P}\left[\sum_{i\in J} \varepsilon_i > \tfrac18 x\sqrt{\lambda} \,\middle|\, E, J \right] \right\}. \qquad (3.11)
\end{aligned}$$

We now want to take expectations over J in (3.11), in order to bound

$$\left| \mathbb{E}\left\{ \Delta h_A(S_n')\, I[S_n' > 1+\gamma+\lambda(x)]\, I[T>n] \right\} \right|.$$

Before doing so, it is convenient to collect a few useful facts. The first is that

$$\begin{aligned}
&\left| \mathbb{E}\left\{ \Delta h_A(S_n')\, I[S_n' > 1+\gamma+\lambda(x)]\, I[T>n]\, I[|J| < n\alpha/2] \right\} \right| \\
&\quad \le \lambda^{-1}\mathbb{P}\left[\sum_{i=1}^n I_i < n\alpha/2 \right] \;\le\; \lambda^{-1}\exp\{-\tfrac12 n\alpha(1-\log 2)\}, \qquad (3.12)
\end{aligned}$$

the last inequality following from a simple Chernoff–type bound. Then, using the reflection principle, we note that

$$\mathbb{P}[E \,|\, |J|=s] \le \mathrm{Bi}\,(s,\tfrac12)\{s/2,(s+1)/2\} \;\le\; C/\sqrt{n\alpha}, \qquad s \ge n\alpha/2. \qquad (3.13)$$

Hence it follows from (3.11) that

$$\begin{aligned}
&\left| \mathbb{E}\left\{ \Delta h_A(S_n')\, I[S_n' > 1+\gamma+\lambda(x)]\, I[T>n] \right\} \right| \\
&\quad \le \lambda^{-1}\exp\{-\tfrac12 n\alpha(1-\log 2)\} + C\lambda^{-1}(n\alpha)^{-1/2}\,\mathbb{P}\left[W > \mathbb{E}W + \tfrac18 x\sqrt{\lambda} \right] \\
&\qquad + \lambda^{-1}(n\alpha)^{-1/2}\left(\mathbb{P}\left[\sum_{i\in J} \tilde\varepsilon_i > \tfrac18 x\sqrt{\lambda} \right] \right. \\
&\qquad\quad \left. + \mathbb{E}\left\{ \mathbb{P}\left[\sum_{i\in J} \varepsilon_i > \tfrac18 x\sqrt{\lambda} \,\middle|\, E, J \right] I[|J| \ge n\alpha/2] \right\} \right). \qquad (3.14)
\end{aligned}$$

The inequality

$$\mathbb{P}[W - \mathbb{E}W > \tfrac18 x\sqrt{\lambda}] \le C\,\frac{1+\Gamma_p}{1+x^p} \qquad (3.15)$$

is an easy consequence of Chebyshev's and Rosenthal's inequalities. Then, from the reflection principle, we find that, on $\{|J|=s\}$,

$$\begin{aligned}
\mathbb{P}\left[\sum_{i\in J} \varepsilon_i \ge r \,\middle|\, E, J \right] &= \frac{\mathrm{Bi}\,(s,1/2)\{\frac12(s+r),\frac12(s+r+1)\}}{\mathrm{Bi}\,(s,1/2)\{s/2,(s+1)/2\}} \\
&\le C\exp\{-r^2/(s+1/2)\}, \qquad (3.16)
\end{aligned}$$

and since $|J| = \sum_{i=1}^n I_i$, it follows that

$$\begin{aligned}
&\mathbb{E}\left\{\mathbb{P}\left[\sum_{i\in J} \varepsilon_i > \tfrac{1}{8}x\sqrt{\lambda}\,\middle|\, E, J\right]\right\} \\
&\quad\le\ C\{\exp\{-x^2\lambda/(128n\alpha)\} + \mathbb{P}[|J| \ge 2n\alpha]\} \\
&\quad\le\ C\{\exp\{-x^2\lambda/(128n\alpha)\} + \exp\{-n\alpha(2\log 2 - 1)\}\},
\end{aligned} \tag{3.17}$$

again from Chernoff bounds. The same bound can also be used for $\mathbb{P}\left[\sum_{i\in J}\tilde{\varepsilon}_i > \frac{1}{8}x\sqrt{\lambda}\right]$. Putting these bounds into (3.14), it follows that, for $\Gamma_p \le 1$,

$$\begin{aligned}
&\left|\mathbb{E}\left\{\Delta h_A(S_n')\, I[S_n' > 1 + \gamma + \lambda(x)]\, I[T > n]\right\}\right| \\
&\qquad \le \frac{C}{\lambda}\left\{\frac{1}{\sqrt{n\alpha}(1 + x^p)} + e^{-\frac{1}{2}n\alpha(1-\log 2)}\right\}.
\end{aligned} \tag{3.18}$$

We actually require a bound for each W_l, $1 \le l \le n$, but not for W itself. However, the argument above applies essentially unchanged, if all sums are taken with the index l omitted; (3.15) is still true as a bound for $\mathbb{P}[W_l - \mathbb{E}W_l > \frac{1}{8}x\sqrt{\lambda}]$, and the only alteration to the bounds arises in (3.12), where

$$\mathbb{P}\left[\sum_{\substack{i=1\\ i\ne l}}^n I_i < n\alpha/2\right] \le \mathbb{P}\left[\sum_{i=1}^n I_i < 1 + n\alpha/2\right] \le 2e\exp\{-\tfrac{1}{2}n\alpha(1 - \log 2)\}$$

replaces the corresponding estimate, which lacked the factor $2e$.

In this argument, the first step was to replace $\Delta h_A(\cdot)$ by the upper bound λ^{-1} for its norm, and it was in fact $\lambda^{-1}\mathbb{E}\{I[S_n' > 1+\gamma+\lambda(x)]\, I[T > n]\}$ that was then treated. Hence the same bound (3.18) can also be used for

$$\left|\mathbb{E}\left\{\Delta h_A(S_n)\, I[S_n > 1 + \gamma + \lambda(x)]\, I[T > n]\right\}\right|,$$

because, on $T > n$, $S_n < S_n'$. To complete the estimate of $|R|$ in (3.8), it therefore remains only to bound the term

$$|\Delta h_A(1 + \gamma + \lambda(x))|\mathbb{P}[W + j = 1 + \gamma + \lambda(x)].$$

However, it follows from Lemma 3.3 that $|\Delta h_A(1+\gamma+\lambda(x))| \le \lambda^{-1}(1+x)q_0(x,\lambda)$. For the second factor, we use the representation of $W + j$ as ${S_n}^{(j)}$. This shows that, conditional on the I_i's, Z_i's and $\widetilde{Z}_i$'s, the largest point probability taken by ${S_n}^{(j)}$ cannot exceed the largest point probability of the distribution of $\sum_{i\in J}\varepsilon_i$, which is itself at most $C(n\alpha)^{-1/2}$ whenever $|J| \ge n\alpha/2$, from the properties of the binomial distribution. Hence, once again using the bound $\mathbb{P}[|J| \le n\alpha/2] \le \exp\{-\frac{1}{2}n\alpha(1 - \log 2)\}$ (with an extra factor of $2e$ if W is replaced by W_l) to deal with the possibility of small values of $|J|$, it follows that

$$|\Delta h_A(1 + \gamma + \lambda(x))|\mathbb{P}[W + j = 1 + \gamma + \lambda(x)] \le C\,\frac{(1+x)q_0(x,\lambda)}{\lambda\sqrt{n\alpha}},$$

of no larger order than the bound in (3.18). Collecting the various terms, it thus follows that

$$|\mathbb{E}\Delta^2 h_A(W+j)| = O\left(\frac{1}{\lambda}\left\{\frac{1}{\sqrt{n\alpha}(1+x^p)} + e^{-\frac{1}{2}n\alpha(1-\log 2)}\right\}\right), \tag{3.19}$$

uniformly in $j \le \sqrt{\lambda}+1$, and the same is true for $|\mathbb{E}\Delta^2 h_A(W_l+j)|$, for each $1 \le l \le n$. Finally, if $8 \le x \le 2n$, it follows by assumption that

$$\sqrt{n\alpha}e^{-\frac{1}{2}n\alpha(1-\log 2)} \le (1+2n)^{-p} \le (1+x^p)^{-1},$$

completing the proof of the lemma. □

4 Proof of Theorem 2.1

The proof is based on (2.12). All its constituents have been bounded in the previous section, if only values of $j \le \sqrt{\lambda}+1$ are allowed in the sum. This is all that is necessary if we suppose that $X_i \le \sqrt{\lambda}+1$ a.s. for all $1 \le i \le n$, so we begin by assuming this to be true. In that case, we have

$$|\delta\mathbb{E}\Delta h_A(W)| \quad \le \quad C_1\lambda^{-1}(1+x^p)^{-1},$$

from Lemma 3.4. Then, easily,

$$\mathrm{Po}\,(\lambda)(A)\mathbb{P}[W<\gamma] = \mathrm{Po}\,(\lambda)(A)\mathbb{P}[W-\mu<-\lambda] \le \lambda^{-1}q_0(x,\lambda),$$

from Chebyshev's inequality and Lemma 3.1. For the remainder, if $p \ge 3$, we have the uniform bound for $|\mathbb{E}\Delta^2 h_A(W_i+j)|$ given by Lemma 3.5 whenever $8 \le x \le 2n$, since $X_i \le \sqrt{\lambda}+1$ a.s. for all i, and in (2.12) this multiplies at most

$$\begin{aligned}
&\sum_{i=1}^n \sum_{j\in\mathbf{Z}} \mathbb{E}\left\{\sigma_i^2|K(j,X_i)| + |X_i-\mu_i|(X_i-1-j)K(j,X_i)\right\} \\
&\quad \le \sum_{i=1}^n \mathbb{E}\left\{\sigma_i^2|X_i| + |X_i-\mu_i|\,\tfrac{1}{2}X_i(X_i-1)\right\} \\
&\quad \le \sum_{i=1}^n \mathbb{E}\left\{\sigma_i^2(|X_i-\mu_i|+|\mu_i|) + |X_i-\mu_i|\,\tfrac{1}{2}|X_i-\mu_i|\{|X_i-\mu_i|+1\}\right\} \qquad (4.1)\\
&\quad = O\left(\lambda^{3/2}\Gamma_3\right),
\end{aligned}$$

since $0 \le \mu_i < 1$ for all i and $\sqrt{\lambda}\Gamma_3 \ge \sqrt{\lambda/n} \ge 1$, this last because $\lambda \ge n$. Note also that, because $X_i \le \sqrt{\lambda}+1$ for all i, $P[W > n(\sqrt{\lambda}+1)] = 0$, so we are only interested in x satisfying $\lambda + x\sqrt{\lambda} \le n(\sqrt{\lambda}+1)$, and such values of x indeed satisfy $x \le 2n$. Combining these results, we find under the conditions of Theorem 2.1 that, for $p \ge 3$,

$$|\mathbb{P}[W-\gamma\in A] - \mathrm{Po}\,(\lambda)(A)| = O\left(\frac{\Gamma_3}{(1+x^p)}\sqrt{\frac{\lambda}{n\alpha}}\right),$$

as required, provided that $X_i \le \sqrt{\lambda}+1$ for all i. We note that (4.1) is a wasteful bound if the X_i take only the values 0 or 1.

For $2 < p < 3$, we bound

$$|\delta \mathbb{E}\Delta h_A(W)| + \mathrm{Po}\,(\lambda)(A)\mathbb{P}[W<\gamma] = O(\lambda^{-1}(1+x^p))$$

as before, and then use the more detailed expression in (2.11), combined with Lemmas 3.4 and 3.5. We write

$$|\mathbb{E}\{\sigma_i^2 \Delta h_A(W) - (X_i-\mu_i)(h_A(W)-h_A(W_i))\}| = |\mathbb{E}(\eta_{1i}-\eta_{2i})|,$$

where

$$\eta_{1i} = \sigma_i^2 \sum_{j\in\mathbb{Z}} K(j,X_i)\Delta^2 h_A(W_i+j)$$

and

$$\begin{aligned}\eta_{2i} &= \sum_{j\in\mathbb{Z}} (X_i-\mu_i)(X_i-1-j)K(j,X_i)\Delta^2 h_A(W_i+j) &\quad (4.2)\\ &= \sum_{j\in\mathbb{Z}} (X_i-\mu_i)\{K(j,X_i)\Delta h_A(W_i+j) - X_i\Delta h_A(W)\}. &\quad (4.3)\end{aligned}$$

As before, from Lemma 3.5, we have

$$\mathbb{E}|\eta_{1i}| \le \frac{C_2\sigma_i^2(\mathbb{E}|X_i-\mu_i|+|\mu_i|)}{\lambda\sqrt{n\alpha}(1+x^p)} \tag{4.4}$$

and

$$|\mathbb{E}(\eta_{2i}\,|\,X_i)| \le \frac{C_2}{\lambda\sqrt{n\alpha}(1+x^p)}\,\tfrac{1}{2}(X_i-\mu_i)^2\{|X_i-\mu_i|+1\},$$

this latter derived from (4.2); but then, taking (4.3), we can use Lemma 3.4 to give

$$|\mathbb{E}(\eta_{2i}\,|\,X_i)| \le |X_i|\,|X_i-\mu_i|\frac{2C_1}{\lambda(1+x^p)}.$$

Hence it follows that

$$\begin{aligned}&|\mathbb{E}\eta_{2i}| \le \mathbb{E}|\mathbb{E}(\eta_{2i}\,|\,X_i)|\\ &\le \frac{1}{\lambda(1+x^p)}\mathbb{E}\left\{|X_i-\mu_i|(|X_i-\mu_i|+1)\left(\frac{C_2|X_i-\mu_i|}{2\sqrt{n\alpha}}\wedge 2C_1\right)\right\}. \quad (4.5)\end{aligned}$$

Since

$$\left(\frac{C_2|X_i-\mu_i|}{4C_1\sqrt{n\alpha}}\wedge 1\right) \le \left(\frac{C_2|X_i-\mu_i|}{4C_1\sqrt{n\alpha}}\right)^{(p-2)}$$

and since, by Hölder's inequality,

$$\sigma_i^2\mathbb{E}|X_i-\mu_i| \le \sigma_i^{(3-p)/2}\mathbb{E}\{|X_i-\mu_i|^p\} \le \lambda^{(3-p)/2}\mathbb{E}\{|X_i-\mu_i|^p\},$$

we deduce from (4.4) and (4.5) that

$$
\begin{aligned}
&|\mathbb{P}[W-\gamma \in A] - \mathrm{Po}\,(\lambda)(A)| \\
&\quad\le \left(\sqrt{\lambda}\Gamma_p + 3/2\right) \frac{C_2}{\sqrt{n\alpha}(1+x^p)} + \frac{2C_1\Gamma_p}{1+x^p}\left(\frac{C_2\sqrt{\lambda}}{4C_1\sqrt{n\alpha}}\right)^{(p-2)} + O\left(\frac{1}{\lambda(1+x^p)}\right), \qquad (4.6)
\end{aligned}
$$

proving the required estimate when $2 < p < 3$, again provided that $X_i \le \sqrt{\lambda}+1$ for all i, since $\lambda \ge n$ and $\sqrt{\lambda}\Gamma_p \ge \lambda^{1/2} n^{-(p-2)/2} \ge 1$.

To remove the restriction on the distributions of the X_i, we argue as in Chen and Shao (2003). We begin by defining

$$X_i^* := \mu_i + (X_i - \mu_i) I[X_i - \mu_i \le \sqrt{\lambda}]$$

and $W^* = \sum_{i=1}^n X_i^*$, so that W^* satisfies the previous assumptions. We now note that, for any $A \subset [z, \infty)$,

$$
\begin{aligned}
0 \le \mathbb{P}[W \in A] - \mathbb{P}[W^* \in A] &\le \mathbb{P}[\{W \ge z\} \cap \{\max_{1\le i\le n} X_i - \mu_i > \sqrt{\lambda}\}] \\
&\le \sum_{i=1}^n \mathbb{P}[\{W \ge z\} \cap \{X_i - \mu_i > \sqrt{\lambda}\}].
\end{aligned}
$$

Taking $z = \mathbb{E}W + y$, it follows that

$$
\begin{aligned}
&\mathbb{P}[\{W \ge \mathbb{E}W + y\} \cap \{X_i - \mu_i > \sqrt{\lambda}\}] \\
&\quad \le \mathbb{P}[\{W_i \ge \mathbb{E}W_i + y/2\} \cap \{\sqrt{\lambda} < X_i - \mu_i \le y/2\}] + \mathbb{P}[X_i - \mu_i > y/2].
\end{aligned}
$$

By Chebyshev's and Rosenthal's inequalities and by the independence of W_i and X_i, we have

$$
\begin{aligned}
&\mathbb{P}[\{W_i \ge \mathbb{E}W_i + y/2\} \cap \{\sqrt{\lambda} < X_i - \mu_i \le y/2\}] \\
&\quad \le 2^p y^{-p} \mathbb{E}|W_i - \mathbb{E}W_i|^p \times \lambda^{-p/2}\mathbb{E}|X_i - \mu_i|^p \\
&\quad \le C y^{-p} \lambda^{p/2}(1+\Gamma_p) \times \lambda^{-p/2}\mathbb{E}|X_i - \mu_i|^p
\end{aligned}
$$

and

$$\mathbb{P}[X_i - \mu_i > y/2] \le 2^p y^{-p} \mathbb{E}|X_i - \mu_i|^p.$$

Summing over i and replacing y by $x\sqrt{\lambda}$, it follows that, for any $A \subset [\mathbb{E}W + x\sqrt{\lambda}, \infty)$,

$$|\mathbb{P}[W \in A] - \mathbb{P}[W^* \in A]| \le C\Gamma_p (1+x^p)^{-1},$$

uniformly in x, so that probabilities of sets $A \subset [\mathbb{E}W + x\sqrt{\lambda}, \infty)$, calculated using W^* in place of W, are equal up to the required order.

This is not, however, the full story, since the mean and variance of W^* are not exactly the same as those of the original W; we need also to show that probabilities calculated with the translated Poisson distribution appropriate to W^* are equivalent to those for the translated Poisson distribution appropriate to W. The parameters for the translated

Poisson distributions are calculated from the respective first and second moments. Their differences can be bounded as follows. For the first moments,

$$\begin{aligned} 0 \ \le\ & \mu - \mu^* = \sum_{i=1}^n \mathbb{E}\{(X_i - \mu_i) I[X_i - \mu_i > \sqrt{\lambda}]\} \\ \le\ & \lambda^{-(p-1)/2} \sum_{i=1}^n \mathbb{E}|X_i - \mu_i|^p = \lambda^{1/2}\Gamma_p; \end{aligned} \tag{4.7}$$

and then, writing $X_i' = X_i - \mu_i$,

$$\begin{aligned} \mathrm{Var}\, W - \mathrm{Var}\, W^* &= \sum_{i=1}^n (\mathrm{Var}\, X_i - \mathrm{Var}\, X_i^*) \\ &= \sum_{i=1}^n \left\{ \mathbb{E}\{(X_i')^2 I[X_i' > \sqrt{\lambda}]\} + (\mathbb{E}\{X_i' I[X_i' > \sqrt{\lambda}]\})^2 \right\} \ge 0, \end{aligned}$$

so that

$$|\mathrm{Var}\, W - \mathrm{Var}\, W^*| \ \le\ 2\sum_{i=1}^n \lambda^{1-p/2}\mathbb{E}|X_i - \mu_i|^p = 2\lambda\Gamma_p. \tag{4.8}$$

For the random variable W and for any $A \subset [\mu + x\sqrt{\lambda}, \infty)$, our current approximation shows that

$$|\mathbb{P}[W \in A] - \mathrm{Po}\,(\lambda^*)\{\lambda^* - \mu^* + A\}| = O\left(\frac{\Gamma_{3\wedge p} + \Gamma_p}{(1 + x^p)} \sqrt{\frac{\lambda}{n\alpha}} \right),$$

where $\mu^* := \mathbb{E}W^*$ and $\lambda^* := \mathrm{Var}\, W^* + \delta^*$, with

$$\delta^* := (\mu^* - \mathrm{Var}\, W^*) - \lfloor \mu^* - \mathrm{Var}\, W^* \rfloor;$$

this is because $\mu^* \le \mu$ and $\lambda^* \le \lambda + 1$, and hence also $A \subset [\mathbb{E}W^* + x'\sqrt{\lambda^*}, \infty)$ with $x' = x\sqrt{\lambda/(\lambda+1)} \ge x/\sqrt{2}$. We thus need to compare $\mathrm{Po}\,(\lambda^*)\{\lambda^* - \mu^* + A\}$ with $\mathrm{Po}\,(\lambda)\{\lambda - \mu + A\}$. However, for each $k \in A$,

$$\frac{\mathrm{Po}\,(\lambda^*)\{\lambda^* - \mu^* + k\}}{\mathrm{Po}\,(\lambda)\{\lambda - \mu + k\}} = \frac{\mathrm{Po}\,(\lambda^*)\{\lambda^* - \mu^* + k\}}{\mathrm{Po}\,(\lambda)\{\lambda - \mu^* + k\}} \frac{\mathrm{Po}\,(\lambda)\{\lambda - \mu^* + k\}}{\mathrm{Po}\,(\lambda)\{\lambda - \mu + k\}},$$

and it follows from Lemmas A1 and A2 that thus

$$\begin{aligned} &\left| \log\left(\frac{\mathrm{Po}\,(\lambda^*)\{\lambda^* - \mu^* + k\}}{\mathrm{Po}\,(\lambda)\{\lambda - \mu + k\}} \right) \right| \\ &\quad = O\{\lambda^{-1}|\lambda - \lambda^*|(1 + \lambda^{-1}|k - \mu^*|^2) + \lambda^{-1/2}|\mu - \mu^*|(1 + \lambda^{-1/2}|k - \mu^*|)\} \\ &\quad = O\{\Gamma_p(1 + \lambda^{-1}(k - \mu)^2)\}, \end{aligned}$$

from (4.7) and (4.8). Hence it follows, for $x \le \sqrt{\lambda}$, that

$$\begin{aligned} \sum_{k=\mu+\lceil x\sqrt{\lambda}\rceil}^{\lfloor \mu+\lambda \rfloor} |\mathrm{Po}\,(\lambda^*)\{\lambda^* - \mu^* + k\} - \mathrm{Po}\,(\lambda)\{\lambda - \mu + k\}| &= O\left(\Gamma_p x^2 \exp\{-x^2/4\}\right) \\ &= O(\Gamma_p(1 + x^p)^{-1}); \end{aligned}$$

and the sum over all larger k is exponentially small in λ, which, for such x, is also of small enough order. For $x \geq \sqrt{\lambda}$, both $\mathrm{Po}\,(\lambda^*)\{\lambda^* - \mu^* + A\}$ and $\mathrm{Po}\,(\lambda)\{\lambda - \mu + A\}$ are of order $O(\exp\{-cx\sqrt{\lambda}\}) = O(\Gamma_p(1+x^p)^{-1})$. Hence, whatever the value of $x \geq 8$, we can replace $\mathrm{Po}\,(\lambda^*)\{\lambda^* - \mu^* + A\}$ with $\mathrm{Po}\,(\lambda)\{\lambda - \mu + A\}$ with an error of at most $O(\Gamma_p(1+x^p)^{-1})$. Thus the theorem remains true without the restriction that $X_i \leq \sqrt{\lambda} + 1$ a.s. for all i. □

For sets $A \subset (-\infty, \lambda - x\sqrt{\lambda}]$ in the lower range, we can simply replace W by $-W$, centre anew, and use the theorem as it stands. However, there is a difference between the distributions obtained by translating a Poisson and translating the negative of a Poisson. To show that this difference is also of small enough order, so that the same translated Poisson distribution is good for both ranges at once, Lemma A3 can be used in an argument similar to that above.

5 Appendix

We collect some useful estimates concerning the Poisson distribution.

Lemma A1. *For $1 \leq |s| \leq \varepsilon n$ and $j \geq 0$, we have*

$$\left|\log \frac{\mathrm{Po}\,(n+s)\{n+j+s\}}{\mathrm{Po}\,(n)\{n+j\}}\right| = O\left(\varepsilon(1+j^2/n)\right),$$

uniformly in $0 \leq \varepsilon \leq 1/4$.

Proof. Direct calculation using Stirling's formula shows that

$$\begin{aligned}
&\left|\log \frac{\mathrm{Po}\,(n+s)\{n+j+s\}}{\mathrm{Po}\,(n)\{n+j\}}\right| \\
&\quad \leq |-s\log(1+j/n) - \tfrac{1}{2}\log(1+s/(n+j)) \\
&\qquad\quad + (n+j+s)\log\{(n+s)(n+j)/(n[n+j+s])\}| + \frac{1}{9n}.
\end{aligned}$$

We now use simple bounds for the Taylor remainders in the logarithmic series (with $|s|/n \leq \varepsilon \leq 1/4$) to give

$$\begin{aligned}
&\left|\log \frac{\mathrm{Po}\,(n+s)\{n+j+s\}}{\mathrm{Po}\,(n)\{n+j\}}\right| \\
&\quad \leq \left\{\frac{|s|j^2}{2n^2} + \frac{2|s|}{3(n+j)} + \frac{8j^2s^2}{9n^2(n+j+s)}\right\} + \frac{1}{9n} \\
&\quad \leq \varepsilon\left\{\frac{2}{3} + \frac{1}{9} + \frac{j^2}{n}\left(\frac{1}{2} + \frac{8}{27}\right)\right\},
\end{aligned}$$

completing the proof. □

Lemma A2. *For $1 \leq |m| \leq \varepsilon\sqrt{n}$ and $j \geq 0$, we have*

$$\left|\log \frac{\mathrm{Po}\,(n)\{n+j+m\}}{\mathrm{Po}\,(n)\{n+j\}}\right| = O\left(\varepsilon(1+j/\sqrt{n})\right),$$

uniformly in $0 \le \varepsilon \le 1/4$.

Proof. Once again, direct calculation using Stirling's formula yields

$$\left|\log \frac{\mathrm{Po}\,(n)\{n+j+m\}}{\mathrm{Po}\,(n)\{n+j\}}\right|$$
$$= |-m\log(1+j/n) + m - (n+j+m+\tfrac{1}{2})\log(1+m/(n+j))| + \frac{1}{9n},$$

and expansion of the logarithms then gives

$$\begin{aligned}
&\left|\log \frac{\mathrm{Po}\,(n)\{n+j+m\}}{\mathrm{Po}\,(n)\{n+j\}}\right| \\
&\le \frac{4|m|j}{3n} + \frac{8m^2}{9(n+j)} + \frac{4m(m+\frac{1}{2})}{3(n+j)} + \frac{1}{9n} \\
&\le 2\varepsilon(1+j/\sqrt{n}),
\end{aligned}$$

as required. □

Lemma A3. *We have*

$$\left|\log \frac{\mathrm{Po}\,(n)\{n+j\}}{\mathrm{Po}\,(n)\{n-j\}}\right| = O\left(n^{-1}|j|(1+j^2/n)\right),$$

uniformly in $1 \le |j| \le n/4$.

Proof. The argument is much as before, giving

$$\begin{aligned}
&\left|\log \frac{\mathrm{Po}\,(n)\{n+j\}}{\mathrm{Po}\,(n)\{n-j\}}\right| \\
&= |(n-j+\tfrac{1}{2})\log(1-j/n) - (n+j+\tfrac{1}{2})\log(1+j/n) + 2j| + \frac{1}{9n} \\
&\le \frac{|j|}{n} + \frac{|j|^3}{n^2} + \frac{64|j|^3}{27n^2} + \frac{1}{9n},
\end{aligned}$$

as required. □

References

[1] Bikelis, A. (1966). Estimates of the remiander in the central limit theorem. *Litovsk. Mat. Sb.*, **6,** 323-346 (in Russian).

[2] Barbour, A. D. and Čekanavičius, V. (2002). Total variation asymptotics for sums of independent integer random variables. *Ann. Probab.*, **30,** 509-545.

[3] Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation.* Oxford Studies in Probability 2, Clarendon Press, Oxford.

[4] Barbour, A. D. and Jensen, J. L. (1989). Local and tail approximations near the Poisson limit. *Scand. J. Statist.*, **16,** 75-87.

[5] Barbour, A. D. and Xia, A. (1999). Poisson perturbations. *ESAIM: P & S*, **3,** 131-150.

[6] Čekanavičius, V. and Vaitkus, P. (2001). Centered Poisson approximation by Stein method. (Russian) *Liet. Mat. Rink.*, **41,** 409-423; translation in *Lith. Math. J.*, **41,** (2001), 319-329.

[7] Chen, L. H. Y. and Suan, I. (2003). Nonuniform bounds in discretized normal approximations. Manuscript.

[8] Chen, L. H. Y. and Shao, Q. (2001). A non-uniform Berry-Esseen bound via Stein's method. *Probab. Theory Related Fields*, **120,** 236-254.

[9] Chen, L. H. Y. and Shao, Q. (2003). Normal approximation under local dependence. Manuscript.

[10] Lindvall, T. (1992). *Lectures on the coupling method.* Wiley, New York.

[11] Petrov, V. V. (1975). *Sums of independent random variables.* Springer-Verlag, Berlin.