A central limit theorem for sums of correlated products

G. Hooghiemstra*

Delft University of Technology, Department of Technical Mathematics, Mekelweg 4, 2628 CD Delft, The Netherlands

M. Keane

CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

Consider a sequence of random points placed on the nonnegative integers with i.i.d. geometric (1/2) interpoint spacings y_i. Let x_i denote the number of points placed at integer i. We prove a central limit theorem for the partial sums of the sequence $x_0 y_0, x_1 y_1, \ldots$. The problem is connected with a question concerning different bootstrap procedures.

Key Words & Phrases: martingale central limit theorem, simple random walk.

1 Introduction

Consider a simple random placement of particles on the nonnegative integers as follows. Begin by flipping a fair coin. If the outcome of the toss is head, place the first particle at 0. If not, move one position to the right, but do not place a particle. At each successive step, toss a fair coin, place a particle at the current position if the outcome of the toss is head, and otherwise move one position to the right. After completion of this infinite procedure a possible picture is something like:
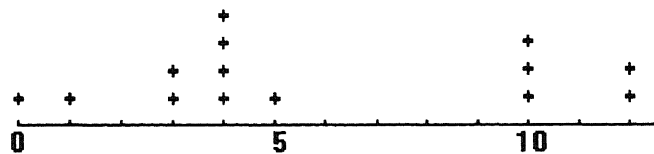

Fig. 1.

Now denote by y_i, $i \geq 1$, the spacing between the ith and $(i + 1)$th particle (with y_0 being the place of the first particle), and by x_i the number of particles at position i ($i \geq 0$). In this article, we investigate the asymptotic behavior of

$$\sum_{i=0}^{m} x_i y_i$$

as m becomes large.

* G.Hooghiemstra@twi.tudelft.nl

Let X_m denote the total number of particles in positions 0 up to m and Y_m the position of the $(m + 1)$th particle. That is

$$X_m := x_0 + \cdots + x_m \tag{1}$$

and

$$Y_m := y_0 + \cdots + y_m \tag{2}$$

Our theorem states that

$$\frac{1}{\sqrt{6m}} \left(\sum_{i=0}^{m} x_i y_i - \min (X_m, Y_m) \right)$$

converges in distribution to a standard normal variable.

WELLNER (1992a) states the following problem (in modified notation). Denote by $E_1, E_2, \ldots$, a sequence of independent, exponentially distributed random variables with mean 1, and by N_i the number of Poisson points contained in $(i - 1, i]$. So

$$N_i := \# \{ j : E_1 + \cdots + E_j \in (i - 1, i] \}$$

Show that

$$\frac{1}{m} \sum_{i=1}^{m} N_i E_i \overset{P}{\to} 1$$

(where $\overset{P}{\to}$ denotes convergence in probability) and secondly, show that

$$\frac{1}{\sqrt{m}} \sum_{i=1}^{m} (N_i E_i - 1) \tag{3}$$

converges in distribution and find the limiting distribution. The first question was solved by the authors together with Serguei Foss (see problems and solutions Statistica Neerlandica, 1994, vol 48, 2, p. 187–200). The second question is much harder. We do think the limit distribution of (3) is normal, and simulation supports this claim; however this seems difficult to prove, as the sequence $N_i E_i$ is a non-stationary sequence of random variables which does not satisfy ordinary mixing conditions. By simplifying the problem (and of course at the same time trying to retain the essential structure of the problem) we came across the variables x_i and y_i introduced above. These variables are exactly the discretized versions of the variables in Wellner's problem, because the y_i introduced above are geometrically distributed with expectation 1. Originally Wellner's problem is motivated by the bootstrap. In trying to understand different bootstrap procedures with correlated multipliers, Wellner formulated the above problem. We refer the interested reader to Section 2 of the survey article WELLNER (1992b).

The central limit theorem below contains an item of interest which allows the treatment of a wider class of normality problems. Essentially, what happens

is that when the sum of interest is conditioned on the correct filtration, the compensator involves the sum itself multiplied by a factor 1/2 and some additional terms which are easy to analyze. Thus the martingale central limit theorem is of relevance.

2 A random path

Let $\xi_1, \xi_2, \ldots$ be an i.i.d. sequence of zeros and ones, with

$$P(\xi_i = 0) = P(\xi_i = 1) = \tfrac{1}{2}$$

The ξ's represent the outcomes of the coin tosses of the previous section. A nice way to visualize our placement is given by the following random path Π. The vertices of Π are the points

$$(s_k, k - s_k), \quad k \geq 0$$

where $s_0 = 0$ and $s_k = \xi_1 + \cdots + \xi_k$, $k \geq 1$, and the edges are the straight line segments connecting successive vertices. The following figure represents the placement of Fig. 1.
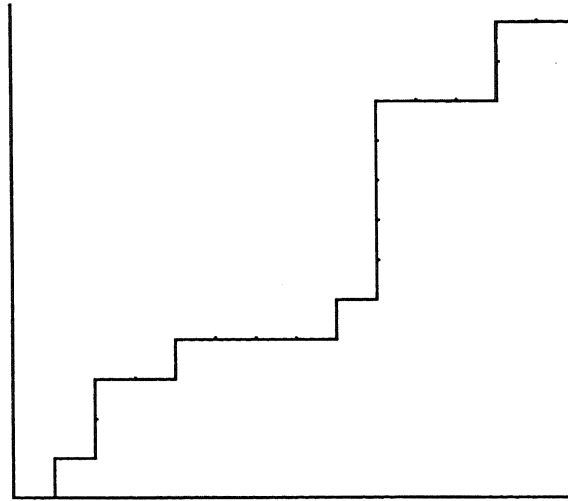

Fig. 2. A path Π.

If we set

$$x_i := \text{length} \left(\Pi \cap \{(x, y) : y = i\} \right) \tag{4}$$

and

$$y_i := \text{length} \left(\Pi \cap \{(x, y) : x = i\} \right) \tag{5}$$

then it is easy to see that these definitions coincide with those of the previous section. The x_i and y_i are a kind of occupation times (or local times) for each direction. Note

that the path Π visits the diagonal infinitely often, but that the expectation of the number of steps between two visits is not finite, since the random walk $\{s_k\}$ is null-recurrent. This phenomenon is exactly what makes the process $\Sigma\, x_i y_i$ a difficult one to study. Although the process has natural times (the visits to the diagonal) at which it regenerates, the expected length of the regeneration cycles is infinite, and this infinite expectation is the cause that mixing conditions such as strong mixing or ρ-mixing are not fulfilled.

The backbone of the construction above is the imbedding of the simple random walk $\{s_k\}$. In order to avoid problems with the mixing we construct a martingale with respect to the filtration generated by the simple random walk. To calculate conditional expectations given $s_0, \ldots, s_m$, or equivalently given $\xi_1, \ldots, \xi_m$, we denote by Π_m the initial part of the random path Π for $0 \le k \le m$, and we set for fixed m,

$$\tilde{x}_i := \text{length} \left(\Pi_m \cap \{(x, y): y = i\} \right) \tag{6}$$

and

$$\tilde{y}_i := \text{length} \left(\Pi_m \cap \{(x, y): x = i\} \right) \tag{7}$$

for $0 \le i \le \min (s_m, m - s_m)$; these random variables are the same as the original x_i and y_i except perhaps for the last value of i, and we have omitted the dependence on m for ease of notation.

Next, define for $m \ge 1$,

$$W_m := \sum_{i=0}^{s_m \wedge (m - s_m)} \tilde{x}_i \tilde{y}_i \tag{8}$$

and let $\mathcal{F}_m$ denote the σ-field generated by $\xi_1, \ldots, \xi_m$. Our goal is to make a martingale out of the sequence W_m by subtracting the quantity (compensator)

$$C_{m-1} = \sum_{i=1}^{m-1} E(W_{i+1} - W_i | \mathcal{F}_i)$$

and after that, to apply the martingale central limit theorem to the stopped sequence $W_{t_m} - C_{t_m - 1}$ where

$$t_m := \inf \left\{ k \ge 2m : s_k \ge m, (k - s_k) \ge m \right\} \tag{9}$$

Note that by choice of t_m, the absolute difference between W_{t_m} and $\Sigma_{i=0}^{m-1} x_i y_i$ is at most $x_m y_m$. As a consequence of Lemma 2 below we see that

$$C_{t_m} - \frac{1}{2} \left(\sum_{i=0}^{m-1} x_i y_i + \min (X_{m-1}, Y_{m-1}) \right) = \frac{1}{2}$$

Hence if

$$\frac{1}{\sqrt{m}} (W_{t_m} - C_{t_m - 1})$$

has a normal limit so does

$$\frac{1}{\sqrt{m}} \left(\sum_{i=0}^{m} x_i y_i - \min (X_m, Y_m) \right)$$

LEMMA 1.

$$E(W_{i+1} - W_i | \mathcal{F}_i) = \begin{cases} \frac{1}{2}\tilde{y}_l, & s_i = k > l = i - s_i \\ \frac{1}{2}(\tilde{y}_k + \tilde{x}_k), & s_i = k = i - s_i \\ \frac{1}{2}\tilde{x}_k, & s_i = k < l = i - s_i \end{cases} \tag{10}$$

PROOF: We will only treat the case where $s_i = k > l = i - s_i$. If $\xi_{i+1} = 1$ then length $(\Pi_{i+1} \cap \{(x, y): x = k\}) = 1 + \text{length}(\Pi_i \cap \{(x, y): x = k\})$, however $W_{i+1} - W_i = 0$, because since $l < k$ the product $\tilde{x}_k \tilde{y}_k$ is not included in the sum W_{i+1}. On the other hand if $\xi_{i+1} = 0$ then length $(\Pi_{i+1} \cap \{(x, y): y = l\}) = 1 + \text{length}(\Pi_i \cap \{(x, y): y = l\})$ and so $\tilde{x}_l$ increases by 1, and this has the effect that $W_{i+1} - W_i = \tilde{y}_l$. $\quad\square$

The random variable

$$\mu_i := E(W_{i+1} - W_i | \mathcal{F}_i) \tag{11}$$

is clearly $\mathcal{F}_i$-measurable, and if we define the compensator

$$C_m := \sum_{i=1}^{m} \mu_i$$

then $W_1 = C_0 = 0$ and

$$W_m - C_{m-1}, \quad m \geq 1$$

is a martingale with respect to the filtration $(\mathcal{F}_m)$.

LEMMA 2. *With probability one:*

$$C_{t_m} = \begin{cases} \frac{1}{2}\left(1 + \sum_{i=0}^{m-1} (x_i + 1)y_i \right), & s_{t_m} > t_m - s_{t_m} \\ \frac{1}{2} \sum_{i=0}^{m-1} x_i y_i + \frac{1}{2}m, & t_m = 2m \\ \frac{1}{2}\left(1 + \sum_{i=0}^{m-1} x_i(y_i + 1) \right), & s_{t_m} < t_m - s_{t_m} \end{cases} \tag{12}$$

PROOF: Let k_1 be the smallest positive integer with $s_{k_1} = 2k_1 - s_{k_1}$, and suppose that $2k_1 \leq t_m$. Since the problem is symmetric we can restrict ourselves to the case where the path Π_{2k_1} is below the diagonal $y = x$. Then:

$$\sum_{i=0}^{2k_1} \mu_i = 0 + \tfrac{1}{2}(\tilde{y}_0 + \cdots + \tilde{y}_0) + \cdots + \tfrac{1}{2}(\tilde{y}_{k_1-1} + \cdots + \tilde{y}_{k_1-1}) + \tfrac{1}{2}(\tilde{x}_{k_1} + \tilde{y}_{k_1})$$

$$= \tfrac{1}{2}(\tilde{x}_1\tilde{y}_1 + \cdots + \tilde{x}_{k_1}\tilde{y}_{k_1}) + \tfrac{1}{2}(\tilde{y}_1 + \cdots + \tilde{y}_{k_1})$$

$$= \frac{1}{2}\sum_{i=0}^{k_1-1} \tilde{x}_i\tilde{y}_i + \tfrac{1}{2}k_1 = \frac{1}{2}\sum_{i=0}^{k_1-1} x_iy_i + \tfrac{1}{2}k_1$$

The first equality follows from (10); the second because $\tilde{y}_0 = 0$, and because the number of vertices of the graph on the line $y = l$ is equal to $\tilde{x}_l + 1$. The third equality is evident from $\tilde{x}_{k_1}\tilde{y}_{k_1} = 0$, and the fact that $\tilde{y}_1 + \cdots + \tilde{y}_{k_1} = k_1$. The final equality follows since $x_i = \tilde{x}_i$ and $y_i = \tilde{y}_i$ for $i = 0, 1, \ldots, k_1 - 1$. By induction the formula is also true if there are more excursions with final endpoint on the diagonal. This proves the middle expression of Formula (12).

Now suppose that the random path ends at (k, m) with $k > m$, and let (k_e, k_e) be the last visit of the path to the diagonal. Then following the same reasoning as above

$$C_{l_m} = \left(\frac{1}{2}\sum_{i=0}^{k_e-1} x_iy_i + \tfrac{1}{2}k_e \right) + \tfrac{1}{2}(\tilde{y}_{k_e} + \cdots + \tilde{y}_{k_e})$$

$$+ \cdots + \tfrac{1}{2}(\tilde{y}_{m-1} + \cdots + \tilde{y}_{m-1}) + \tfrac{1}{2}\tilde{y}_m$$

$$= \frac{1}{2}\sum_{i=0}^{m-1} x_iy_i + \frac{1}{2}\sum_{i=0}^{m-1} y_i + \tfrac{1}{2}$$

because $\tilde{y}_m = 1$. This completes the proof of Lemma 2. □

3 Asymptotic normality

We now present the result.

THEOREM.

$$\frac{1}{\sqrt{6m}}\left(\sum_{i=0}^{m} x_iy_i - (X_m \wedge Y_m) \right) \xrightarrow{\mathscr{D}} N(0, 1)$$

PROOF. As indicated above we intend to apply the central limit theorem for maringales (cf. LÉVY, 1935; for a modern treatment see POLLARD, 1984). Let

$$\sigma_i^2 := E((W_{i+1} - W_i - \mu_i)^2 | \mathscr{F}_i), \quad i \geq 1$$

then

$$\sigma_i^2 = \begin{cases} \tfrac{1}{4}\tilde{y}_i^2 & s_i = k > l = i - s_i \\ \tfrac{1}{4}(\tilde{y}_k^2 + \tilde{x}_k^2), & s_i = i - s_i = k \\ \tfrac{1}{4}\tilde{x}_k^2, & s_i = k < l = i - s_i \end{cases}$$

Clearly the contribution to the conditional variance for an excursion of the path below the diagonal is of the form

$$\frac{1}{4}\sum (\tilde{x}_i + 1)\tilde{y}_i^2$$

whereas above the diagonal the contribution is

$$\frac{1}{4}\sum (\tilde{y}_i + 1)\tilde{x}_i^2$$

Denote by $\tilde{Y}_m := \tilde{y}_0 + \cdots + \tilde{y}_m$ and by $\tilde{X}_m := \tilde{x}_0 + \cdots + \tilde{x}_m$. A more detailed analysis, similar to the proof of Lemma 2, shows that

$$0 \le \frac{1}{4}\left\{(\tilde{x}_0 + 1)\tilde{y}_0^2 + (\tilde{y}_0 + 1)\tilde{x}_0^2 + \sum_{i=1}^{s_m \wedge (m - s_m)} [(\tilde{x}_i + 1)\tilde{y}_i^2 1(\tilde{Y}_{i-1} < i)\right.$$

$$\left. + (\tilde{y}_i + 1)\tilde{x}_i^2 1(\tilde{X}_{i-1} < i)]\right\} - (\sigma_1^2 + \cdots + \sigma_m^2) \tag{13}$$

$$\le \frac{1}{4}\left\{\sum_{i=1}^{s_m \wedge (m - s_m)} [\tilde{y}_i^2 1(\tilde{Y}_{i-1} < i < \tilde{Y}_i) + \tilde{x}_i^2 1(\tilde{X}_{i-1} < i < \tilde{X}_i)]\right\}$$

By conditioning on Y_{i-1} it is seen that

$$Ey_i^2 1(Y_{i-1} < i < Y_i) = E(E(y_i^2 1(Y_{i-1} < i < Y_{i-1} + y_i)|Y_{i-1}))$$

$$= E\left(\sum_{l=1}^{\infty} l^2 (\tfrac{1}{2})^{l+1} 1(i - l < Y_{i-1} < i)\right)$$

$$= \sum_{l=1}^{\infty} l^2 (\tfrac{1}{2})^{l+1} P(i - l < Y_{i-1} < i) = O\left(\frac{1}{\sqrt{i}}\right)$$

Hence

$$E\left\{\sum_{i=1}^{m} y_i^2 1(Y_{i-1} < i < Y_i)\right\} = O(\sqrt{m}), \quad m \to \infty$$

which fact implies, by nonnegativity of the summand and because s_m has a binomial distribution with parameters m and $1/2$,

$$\frac{1}{m}\sum_{i=1}^{s_m \wedge (m - s_m)} \tilde{y}_i^2 1(\tilde{Y}_{i-1} < i < \tilde{Y}_i) \overset{P}{\to} 0, \quad m \to \infty$$

Using symmetry we conclude from (13) that

$$\frac{1}{m}(\sigma_1^2 + \cdots + \sigma_m^2) - \frac{1}{4m}\left\{\sum_{i=1}^{s_m \wedge (m - s_m)} [(\tilde{x}_i + 1)\tilde{y}_i^2 1(\tilde{Y}_{i-1} < i)\right.$$

$$\left. + (\tilde{y}_i + 1)\tilde{x}_i^2 1(\tilde{X}_{i-1} < i)]\right\} \overset{P}{\to} 0$$

In the lemma following this proof we will show that

$$\frac{1}{m}\left\{\sum_{i=1}^{m} [(x_i + 1)y_i^2 1(Y_{i-1} < i) + (y_i + 1)x_i^2 1(X_{i-1} < i)]\right\}\xrightarrow{P} 2\mu_2, \quad m\to\infty$$

where $\mu_2 = Ey_i^2 = 3$. From the previous two statements and the law of large numbers for the sequence s_m it is immediate that the conditional variance

$$\frac{1}{m}(\sigma_1^2 + \cdots + \sigma_m^2)\xrightarrow{P}\mu_2/4$$

To finish the proof of the theorem we verify Lindeberg's condition. This condition is trivial because $\sigma_1^2 + \cdots + \sigma_m^2 \sim m\mu_2/4$, and $|\xi_j| \leq 1$, almost surely. Applying the martingale central limit theorem, with stopping time t_m we obtain:

$$\frac{1}{\sqrt{2m}}(W_{t_m} - C_{t_m-1})\xrightarrow{\mathcal{D}}N(0, \tfrac{1}{4}\mu_2)$$

because

$$\frac{t_m}{2m}\to 1, \quad \text{a.s.}$$

Now use the remarks preceeding Lemma 1 to obtain the result.   □

LEMMA 3.

$$\frac{1}{m}\left\{\sum_{i=1}^{m} [(x_i + 1)y_i^2 1(Y_{i-1} < i) + (y_i + 1)x_i^2 1(X_{i-1} < i)]\right\}\xrightarrow{P} 2\mu_2, \quad m\to\infty$$

$$(14)$$

PROOF: Instead of (14) we show that

$$\frac{1}{m}\left\{\sum_{i=1}^{m} [(x_i + 1)y_i^2 1(Y_i < i) + (y_i + 1)x_i^2 1(X_i < i)]\right\}\xrightarrow{P} 2\mu_2, \quad m\to\infty \quad (15)$$

That is we show the statement of the lemma with Y_{i-1} and X_{i-1} replaced by Y_i and X_i, respectively. Note that $1(Y_{i-1} < i) - 1(Y_i < i) = 1(Y_{i-1} < i \leq Y_i)$. It follows from

$$\frac{1}{m}E\sum_{i=1}^{m}(x_i + 1)y_i^2 1(Y_{i-1} < i \leq Y_i) = O(m^{-1/2}), \quad m\to\infty$$

and

$$\frac{1}{m}E\sum_{i=1}^{m}(y_i + 1)x_i^2 1(X_{i-1} < i \leq X_i) = O(m^{-1/2}), \quad m\to\infty$$

that (15) implies (14) (here we used that L_1 convergence of nonnegative random variables implies convergence in probability). We next verify that the expectation of the left-hand side of (15) converges to $2\mu_2$ as $m\to\infty$.

$$E(x_i + 1)y_i^2 1(Y_i < i) = 2Ey_i^2 1(Y_i < i)$$
$$= 2E(E(y_i^2 1(y_i + Y_{i-1} < i)|Y_{i-1}))$$
$$= 2 \sum_{l=1}^{x} P(\tfrac{1}{2})^{l+1} P(Y_{i-1} < i - l) \to 2 \cdot \tfrac{1}{2}Ey_1^2 = \mu_2$$

for $i \to \infty$. This follows by dominated convergence since $Ey_1^2 < \infty$, and the fact that for fixed l, according to the central limit theorem, $P(Y_{i-1} < i - l) \to 1/2$. By symmetry the same result holds when y and x are interchanged:

$$\lim_{i \to x} E(y_i + 1)x_i^2 1(X_i < i) = \mu_2$$

Hence the statement of the lemma is true if we show that the variance (Var) of the left-hand side of (15) converges to 0 as $m \to \infty$.

Now consider the reflected random walk $\xi_1' = 1, \xi_2', \xi_3', \ldots$, where ξ_{2k+1}' is Bernoulli with probability $1/2$, when $\xi_1' + \cdots + \xi_{2k}' \neq k$, and $\xi_{2k+1}' = 1$ with probability 1 when $\xi_1' + \cdots + \xi_{2k}' = k$. Formally $s_k' = \xi_1' + \cdots + \xi_k'$ forms a Markov chain, starting from $s_1' = 1$ and with transition probabilities

$$P(s_{2k+1}' = s_{2k}' + 1|s_{2k}') = 1, \quad s_{2k}' = k$$
$$P(s_{2k+1}' = s_{2k}' + 1|s_{2k}') = P(s_{2k+1}' = s_{2k}'|s_{2k}') = \tfrac{1}{2}, \quad s_{2k}' > k$$

As before we define by Π' the random path with vertices $(s_k', k - s_k')$, $k \geq 0$. Note that all vertices satisfy $s_k' \geq k - s_k'$. Let x' and y' be defined as in (4) and (5) but with Π replaced by Π', and let X' and Y' be the partial sums. A probabilistic replica of the path Π can be obtained from Π' in a pathwise manner by randomizing the variables ξ_{2k+1}' (choosing probability $1/2$ to each of the possibilities 0 and 1) for which $s_{2k}' = k$. It follows from the correspondence between the paths Π and Π' that for $i \geq 1$,

$$1(Y_i' = i) = 1(Y_i = i) + 1(X_i = i)$$

almost surely. Moreover by symmetry of the random walk $\xi_1, \xi_1 + \xi_2 \cdots$, we obtain the almost sure identity

$$\sum_{i=1}^{m} (x_i' + 1)(y_i')^2 1(Y_i' < i) = \sum_{i=1}^{m} [(x_i + 1)y_i^2 1(Y_i < i) + (y_i + 1)x_i^2 1(X_i < i)]$$

Hence the statement of the lemma follows if we show that

$$\text{Var}\left(\frac{1}{m} \sum_{i=1}^{m} (x_i' + 1)(y_i')^2 1(Y_i' < i)\right) \to 0, \quad m \to \infty \tag{16}$$

To this end put

$$Z_i = (x_i' + 1)(y_i')^2 1(Y_i' < i)$$

Since

$$\frac{1}{m^2} \sum_1^m \mathrm{Var}\,(Z_i) \leq \frac{1}{m^2} \sum_1^m E(x_i' + 1)^2(y_i')^4 = O\!\left(\frac{1}{m}\right)$$

it is sufficient to show that

$$\frac{1}{m^2} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} (EZ_iZ_j - EZ_iEZ_j) \to 0, \quad m \to \infty \tag{17}$$

Since $Z_i = 0$ on the set $\{Y_i' = i\}$ it is no restriction to prove (17) on the set $\{Y_i' < i\}$. Then condition (17) is equivalent to

$$\frac{1}{m^2} \sum_{i=1}^{m-1} E\!\left(Z_i E\!\left(\sum_{j=i+1}^{m} Z_j - \sum_{j=i+1}^{m} EZ_j | y_i', x_i' \right) \right) \to 0, \quad m \to \infty$$

For $k, l \in \mathbb{N}$ we consider a random path Π' starting from vertex $(i + 1, Y_i')$ to (m, Y_m'). The section of this path from $(i + 1, Y_i')$ to (p, p), where $p > i + 1$ is the first index for which the path hits the diagonal can serve as a realisation of a path starting from $(i + 1, Y_i' - 1)$ and ending at $(p, p - 1)$, by translation of the section over the vector $(0, -1)$. At $(p, p - 1)$ we move either to (p, p) or to $(p + 1, p - 1)$, each possibility has probability $1/2$. In the first case the lower path couples with the original path; in the second case we end at $(p + 1, p - 1)$, while the original path is at $(p + 1, p)$; from these two points the above procedure can be repeated. Note that if coupling occurs at (p, p) then $\sum_{j=i+1}^{p} (y_j')^2 1(Y_j' < j)$ is the same for both paths, while in case of no coupling the absolute difference is $(y_p')^2$. Hence for $k, l \in \mathbb{N}$,

$$\left| E\!\left(\sum_{j=i+1}^{m} Z_j | x_i' = k, y_i' = l + 1 \right) - E\!\left(\sum_{j=i+1}^{m} Z_j | x_i' = k, y_i' = l \right) \right|$$

$$\leq E(x_i' + 1) \left| E\!\left(\sum_{j=i+1}^{m} (y_j')^2 1(Y_j' < j) | x_i' = k, y_i' = l + 1 \right) \right.$$

$$\left. - E\!\left(\sum_{j=i+1}^{m} (y_j')^2 1(Y_j' < j) | x_i' = k, y_i' = l \right) \right|$$

$$\leq 2\mu_2 \sum_{n=0}^{\infty} n(\tfrac{1}{2})^{n+1} = 2\mu_2$$

A similar coupling argument which we leave to the reader shows that

$$\left| E\!\left(\sum_{j=i+1}^{m} Z_j | x_i' = k + 1, y_i' = l \right) - E\!\left(\sum_{j=i+1}^{m} Z_j | x_i' = k, y_i' = l \right) \right| \leq 2\mu_2$$

From the above two estimates we obtain quite easily

$$\left| E\!\left(\sum_{j=i+1}^{m} Z_j | x_i' = k, y_i' = l \right) - E \sum_{j=i+1}^{m} Z_j \right| \leq 2\mu_2 E(|x_i' - k| + |y_i' - l|)$$

$$\leq 2((k + 1)\mu_2 + (l + 1)\mu_2)$$

This finally shows that

$$\frac{1}{m^2}\left|\sum_{i=1}^{m-1} E\left(Z_i E\left(\sum_{j=i+1}^{m} Z_j | x_i', y_i'\right) - E \sum_{j=i+1}^{m} Z_j\right)\right|$$

$$\leq \frac{1}{m^2} \sum_{i=1}^{m-1} 2\mu_2 E((x_i' + 1)(y_i')^2(x_i' + y_i' + 2)) = O\left(\frac{1}{m}\right) \qquad \square$$

4 Some concluding remarks

In the paper we proved a central limit theorem for $\sum x_i y_i$ using the random centering (min (X_m, Y_m)). It is conceivable that the proof, albeit less elegant, can be pushed through for exponential random variables. We would then obtain a result in the spirit of Wellner, though again with random centering.

To obtain a limit result for $\sum x_i y_i$ with deterministic centering we need the joint asymptotic behaviour of $(\sum x_i y_i, Y_m)$. It is well-known (see IGLEHART and WHITT, 1971), that the asymptotic behaviour of X_m determines that of Y_m in the sense that if either

$$m^{-1/2}(X_m - m) \overset{\mathcal{D}}{\to} N \quad \text{or} \quad m^{-1/2}(Y_m - m) \overset{\mathcal{D}'}{\to} -N,$$

then

$$m^{-1/2}(X_m - m, Y_m - m) \overset{\mathcal{D}}{\to} (N, -N)$$

We tried the martingale method developed in this paper on linear combinations of the form

$$\alpha \sum_{i=0}^{s_m \wedge (m - s_m)} x_i y_i + \beta \sum_{i=0}^{s_m} y_i$$

The compensator for this expression stopped at time t_m equals

$$\alpha C_{t_m} + \beta(\tfrac{1}{2}(X_m \vee Y_m) + \tfrac{1}{2}m)$$

However, for α and β both unequal to 0 the proof of the convergence in probability of the conditional variance breaks down. For $\alpha = 0$ we obtain the curious one dimensional central limit theorem:

$$\frac{1}{\sqrt{2m}}((Y_m - m)1(Y_m \geq m) + (m - X_m)1(Y_m < m)) \overset{\mathcal{D}}{\to} N(0, 1)$$

which is a consequence of the above cited result of Iglehart and Whitt, while for $\beta = 0$ we obtain the contents of our theorem. Simulation indicates that the joint limit of

$$\frac{1}{\sqrt{m}}\left(\sum x_i y_i - (X_m \wedge Y_m)\right), (Y_m - m)1(Y_m \geq m) + (m - X_m)1(Y_m < m)$$

is indeed not normal. However this does not contradict the possibility of a normal limit for $1/\sqrt{m} \Sigma (x_iy_i - 1)$.

Note that joint asymptotic normality (with deterministic centering) of the pair Σx_iy_i, Σy_i is not possible, because it conflicts with the result in this paper.

References

IGLEHART, D. L. and W. WHITT (1971), The equivalence of functional central limit theorems for counting processes and associated partial sums, *Annals of Mathematical Statistics* **42**, 1372–1378.

LÉVY, P. (1935), Propriétés asymptotiques des sommes de variables aléatoires indépendantes ou enchainées, *Journal de Mathématiques Pures et Appliquées* **14** *(Ser. 9)*, 347–402.

POLLARD, D. (1984), *Convergence of stochastic processes*, Springer-Verlag, New York.

WELLNER, J. A. (1992a), *Problem section, Statistica Neerlandica* **46**, 299.

WELLNER, J. A. (1992b), Bootstrap limit theorems: a partial survey, in: A. Saleh (editor), *Nonparametric statistics and related topics*, Elseviers Science Publ., Amsterdam, 313–329.