Electronic Journal of Probability

Vol. 14 (2009), Paper no. 81, pages 2352–2370.

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

A functional combinatorial central limit theorem

A. D. Barbour* and S. Janson†
Universität Zürich and Uppsala Universitet

Abstract

The paper establishes a functional version of the Hoeffding combinatorial central limit theorem. First, a pre-limiting Gaussian process approximation is defined, and is shown to be at a distance of the order of the Lyapounov ratio from the original random process. Distance is measured by comparison of expectations of smooth functionals of the processes, and the argument is by way of Stein's method. The pre-limiting process is then shown, under weak conditions, to converge to a Gaussian limit process. The theorem is used to describe the shape of random permutation tableaux.

Key words: Gaussian process; combinatorial central limit theorem;permutation tableau; Stein's method.

AMS 2000 Subject Classification: Primary 60C05, 60F17, 62E20, 05E10.

Submitted to EJP on July 2, 2009, final version accepted October 16, 2009.

*Angewandte Mathematik, Universität Zürich, Winterthurertrasse 190, CH-8057 ZÜRICH; a.d.barbour@math.uzh.ch. Work supported in part by Schweizerischer Nationalfonds Projekte Nr. 20–107935/1 and 20–117625/1.

†Matematiska institutionen, Uppsala universitet, Box 480, SE–751 06 UPPSALA; Svante.Janson@math.uu.se.
Research done while both authors visited the Institut Mittag–Leffler, Djursholm, Sweden.

1 Introduction

Let $a_0^{(n)} := (a_0^{(n)}(i,j),\, 1 \le i,j \le n)$, $n \ge 1$, be a sequence of real matrices. Hoeffding's (1951) combinatorial central limit theorem asserts that if π is a uniform random permutation of $\{1,2,\ldots,n\}$, then, under appropriate conditions, the distribution of the sum

$$S_0^{(n)} \;:=\; \sum_{i=1}^{n} a_0^{(n)}(i,\pi(i)),$$

when centered and normalized, converges to the standard normal distribution. The centering is usually accomplished by replacing $a_0^{(n)}(i,j)$ with

$$\tilde{a}^{(n)}(i,j) \;\;:=\;\; a_0^{(n)}(i,j) - \bar{a}_0^{(n)}(+,j) - \bar{a}_0^{(n)}(i,+) + \bar{a}_0^{(n)}(+,+),$$

where

$$\begin{aligned} \bar{a}_0^{(n)}(+,j) \;\;&:=\;\; n^{-1}\sum_{i=1}^{n} a_0(i,j); \quad \bar{a}_0^{(n)}(i,+) \;:=\; n^{-1}\sum_{j=1}^{n} a_0(i,j); \\ \bar{a}_0^{(n)}(+,+) \;\;&:=\;\; n^{-2}\sum_{i=1}^{n} a_0(i,j). \end{aligned}$$

This gives $\widetilde{S}^{(n)} = S_0^{(n)} - \mathbb{E}S_0^{(n)}$, and the variance $\operatorname{Var}\widetilde{S}^{(n)} = \operatorname{Var} S_0^{(n)}$ is then given by

$$\{\tilde{s}^{(n)}(a)\}^2 \;:=\; (n-1)^{-1}\sum_{i,j=1}^{n}\{\tilde{a}^{(n)}(i,j)\}^2.$$

Bolthausen (1984) proved the analogous Berry–Esseen theorem: that, for any $n \times n$ matrix a,

$$\sup_{x\in\mathbb{R}} |\mathbb{P}[S_0 - m(a) \le x\tilde{s}(a)] - \Phi(x)| \;\le\; C\widetilde{\Lambda}(a),$$

for a universal constant C, where Φ denotes the standard normal distribution function,

$$\begin{aligned} S_0 := \sum_{i=1}^{n} a_0(i,\pi(i)), \qquad m(a) \;&:=\; n^{-1}\sum_{i,j=1}^{n} a_0(i,j) \;=\; \mathbb{E}S_0, \\ \tilde{s}^2(a) := (n-1)^{-1}\sum_{i,j=1}^{n}\tilde{a}^2(i,j) \;&=\; \operatorname{Var} S, \end{aligned} \tag{1.1}$$

(we tacitly assume $n \ge 2$ when necessary) and

$$\widetilde{\Lambda}(a) \;:=\; \frac{1}{n\tilde{s}^3(a)}\sum_{i,j=1}^{n}|\tilde{a}(i,j)|^3$$

is the analogue of the Lyapounov ratio.

In this paper, we begin by proving a functional version of Bolthausen's theorem, again with an error expressed in terms of a Lyapounov ratio. When centering the functional version $S_0(t) :=$

$\sum_{i=1}^{\lfloor nt \rfloor} a_0(i,\pi(i))$, $0 \le t \le 1$, it is however no longer natural to make the double standardization that is used to derive $\tilde{a}$ from a_0. Instead, we shall at each step center the random variables $a_0(i,\pi(i))$ individually by their means $\bar{a}_0(i,+)$. Equivalently, in what follows, we shall work with matrices a satisfying $\bar{a}(i,+) = 0$ for all i, but with no assumption as to the value of $\bar{a}(+,j)$. For example, if we have $a_0(i,j) = b(i) + c(j)$, then $\tilde{a}(i,j) = 0$ for all i,j, and hence $S_0 = \mathbb{E}S_0 = n\bar{a}_0(+,+) = n(\bar{b}+\bar{c})$ is a.s. constant. However, we are interested instead in

$$S(t) \ := \ \sum_{i=1}^{\lfloor nt \rfloor} \{a_0(i,\pi(i)) - \bar{a}_0(i,+)\},$$

giving $S(t) = \sum_{i=1}^{\lfloor nt \rfloor}\{c(\pi(i)) - \bar{c}\}$, a non-trivial process with a Brownian bridge as natural approximation.

We thus, throughout the paper, define the matrix a by

$$a(i,j) \ := \ a_0(i,j) - \bar{a}_0(i,+), \tag{1.2}$$

so that $\bar{a}(i,+) = 0$. Correspondingly, we define

$$S(t) \ := \ \sum_{i=1}^{\lfloor nt \rfloor} a(i,\pi(i)) \ = \ S_0(t) - \mathbb{E}S_0(t).$$

We then normalize by a suitable factor $s(a) > 0$, and write

$$Y(t) \ := \ s(a)^{-1}S(t) = s(a)^{-1}\big(S_0(t) - \mathbb{E}S_0(t)\big); \tag{1.3}$$

this can equivalently be expressed as

$$Y \ := \ Y(\pi) \ := \ \frac{1}{s(a)}\sum_{i=1}^{n} a(i,\pi(i))J_{i/n}, \tag{1.4}$$

where $J_u(t) := \mathbf{1}_{[u,1]}(t)$. In Theorem 2.1, we approximate the random function Y by the Gaussian process

$$Z \ := \ \sum_{i=1}^{n} W_i J_{i/n}, \tag{1.5}$$

in which the jointly Gaussian random variables $(W_i, 1 \le i \le n)$ have zero means and covariances given by

$$\begin{aligned} \operatorname{Var} W_i \ &= \ \frac{1}{ns^2(a)}\sum_{l=1}^{n} a^2(i,l) \ =: \ \sigma_{ii}; \\ \operatorname{Cov}(W_i, W_j) \ &= \ -\frac{1}{n(n-1)s^2(a)}\sum_{l=1}^{n} a(i,l)a(j,l) \ =: \ \sigma_{ij}, \qquad i \ne j. \end{aligned} \tag{1.6}$$

A simple calculation shows that $\operatorname{Cov}\big(a(i,\pi(i)), a(j,\pi(j))\big) = s^2(a)\sigma_{ij}$ for all i,j, and thus the covariance structures of the processes Y and Z are identical. The error in the approximation is expressed in terms of a probability metric defined in terms of comparison of expectations of certain smooth functionals of the processes, and it is bounded by a multiple of the Lyapounov ratio

$$\Lambda(a) \ := \ \frac{1}{ns^3(a)}\sum_{i,j=1}^{n} |a(i,j)|^3. \tag{1.7}$$

The normalization factor $s(a)$ may be chosen in several ways. One obvious possibility is to choose $s(a) = \tilde{s}(a)$ defined in (1.1), which makes $\operatorname{Var} Y(1) = \operatorname{Var} Z(1) = 1$. At other times this is inappropriate; for example, as seen above, $\tilde{s}(a)$ may vanish, although we have a non-trivial Brownian bridge asymptotic. A canonical choice of normalization is

$$s^2(a) \;:=\; \frac{1}{n-1}\sum_{i,j=1}^{n} a^2(i,j), \tag{1.8}$$

or, for simplicity, $n^{-1}\sum_{i,j=1}^{n} a^2(i,j)$, which makes no difference asymptotically. In the special case where $\bar{a}(+,j) = 0$ for each j, as with the matrix $\tilde{a}$, this gives $s^2(a) = \tilde{s}^2(a)$, so $\operatorname{Var} Y(1) = \operatorname{Var} Z(1) = 1$, but in general this does not hold. In specific applications, some other choice may be more convenient. We thus state our main results for an arbitrary normalization.

In most circumstances, such an approximation by $Z = Z^{(n)}$ depending on n is in itself not particularly useful; one would prefer to have some fixed, and if possible well-known limiting approximation. This requires making additional assumptions about the sequence of matrices $a^{(n)}$ as $n \to \infty$. In extending Bolthausen's theorem, it is enough to assume that $\widetilde{\Lambda}^{(n)}(a) \to 0$, since the approximation is already framed in terms of the standard normal distribution. For functional approximation, even if we had standardized to make $\operatorname{Var} Y(1) = 1$, we would still have to make some further assumptions about the $a^{(n)}$, in order to arrive at a limit. A natural choice would be to take $a^{(n)}(i,j) := \alpha(i/n, j/n)$ for a continuous function $\alpha\colon [0,1]^2 \to \mathbb{R}$ which does not depend on n. We shall make a somewhat weaker assumption, enough to guarantee that the covariance function of $Z^{(n)}$ converges to a limit, which itself determines a limiting Gaussian process. The details are given in Theorem 3.3. Note that we require that $\Lambda^{(n)}(a)\log^2 n \to 0$ for process convergence, a slightly stronger condition than might have been expected. This is as a result of the method of proof, using the approach in Barbour (1990), in which the probability metric used for approximation is perhaps not strong enough to metrize weak convergence in the Skorohod topology. Requiring the rate of convergence of $\Lambda^{(n)}(a)$ to zero to be faster than $1/\log^2 n$ is however enough to ensure that weak convergence also takes place: see Proposition 3.1.

The motivation for proving the theorems comes from the study of permutation tableaux. In Section 5, we show that the boundary of a random permutation tableau, in the limit as its size tends to infinity, has a particular shape, about which the random fluctuations are approximately Gaussian. The main tool in proving this is Theorem 3.3, applied to the matrices $a_0^{(n)}(i,j) := \mathbf{1}_{\{i \le j\}}$.

2 The pre-limiting approximation

We wish to show that the distributions of the processes Y and Z of (1.4) and (1.5) are close. To do so, we adopt the approach in Barbour (1990). We let M denote the space of all twice Fréchet differentiable functionals $f : D := D[0,1] \to \mathbb{R}$ for which the norm

$$\begin{aligned}\|f\|_M \;:=\; &\sup_{w\in D}\{|f(w)|/(1+\|w\|^3)\} + \sup_{w\in D}\{\|Df(w)\|/(1+\|w\|^2)\} \\ &+ \sup_{w\in D}\{\|D^2 f(w)\|/(1+\|w\|)\} + \sup_{w,h\in D}\{\|D^2 f(w+h) - D^2 f(w)\|/\|h\|\}\end{aligned} \tag{2.1}$$

is finite; here, $\|\cdot\|$ denotes the supremum norm on D, and the norm of a (symmetric) k-linear form B on function in D is defined to be $\|B\| := \sup_{h\in D\colon \|h\|=1} |B[h^{(k)}]|$, where $h^{(k)}$ denotes the k-

tuple $(h, h, \ldots, h)$. Our aim is to show that $|\mathbb{E}g(Y) - \mathbb{E}g(Z)|$ is small for all $g \in M$. We do this by Stein's method, observing that, for any $g \in M$, there exists a function $f \in M$ satisfying

$$g(w) - \mathbb{E}g(Z) \;=\; (\mathscr{A} f)(w) \;:=\; -Df(w)[w] + \sum_{i,j=1}^{n} \sigma_{ij} D^2 f(w)[J_{i/n}, J_{j/n}], \tag{2.2}$$

and that

$$\|f\|_M \le C_0 \|g\|_M, \tag{2.3}$$

where C_0 does not depend on the choice of g: see, for example, Barbour (1990, (2.24), Remark 7 after Theorem 1 and the remark following Lemma 3.1). Hence it is enough to prove that $|\mathbb{E}(\mathscr{A} f)(Y)| \le \varepsilon \|f\|_M$ for all $f \in M$ and for some small ε.

Theorem 2.1. *Let $Y = Y(\pi)$ and Z be defined as in* (1.4) *and* (1.5)*, with π a uniform random permutation of $\{1, 2, \ldots, n\}$, and $\Lambda(a)$ as in* (1.7)*, for some $n \times n$ matrix $a(i, j)$ with $\bar{a}(i, +) = 0$ and some $s(a) > 0$. Then there exists a universal constant K such that, for all $f \in M$,*

$$|\mathbb{E}(\mathscr{A} f)(Y)| \;\le\; K\Lambda(a)\|f\|_M.$$

Thus, for all $g \in M$,

$$|\mathbb{E}g(Y) - \mathbb{E}g(Z)| \;\le\; C_0 K \Lambda(a)\|g\|_M,$$

with C_0 as in (2.3).

Proof. We begin by noting that

$$\mathbb{E}Df(Y)[Y] \;=\; \frac{1}{s(a)} \sum_{i=1}^{n} \mathbb{E}\{X_i Df(Y)[J_{i/n}]\}, \tag{2.4}$$

where $X_i := a(i, \pi(i))$. We then write

$$\mathbb{E}\{X_i Df(Y)[J_{i/n}]\} \;=\; \frac{1}{n} \sum_{l=1}^{n} a(i, l)\mathbb{E}\{Df(Y(\pi))[J_{i/n}] \,|\, \pi(i) = l\}. \tag{2.5}$$

Now realize π' with the distribution $\mathscr{L}(\pi \,|\, \pi(i) = l)$ by taking π to be a uniform random permutation, and setting

$$\begin{aligned} &\pi' \;=\; \pi, && \text{if} \quad \pi(i) \;=\; l;\\ &\pi'(i) \;=\; l; \quad \pi'(\pi^{-1}(l)) \;=\; j; \quad \pi'(k) \;=\; \pi(k),\ k \notin \{i, \pi^{-1}(l)\}, && \\ & && \text{if} \quad \pi(i) \;=\; j \;\ne\; l. \end{aligned}$$

This gives

$$Y(\pi') \;=\; Y(\pi) + \Delta_{il}(\pi) \;=:\; Y'(\pi), \tag{2.6}$$

where

$$s(a)\Delta_{il}(\pi) \;:=\; \{a(i, l) - a(i, \pi(i))\}J_{i/n} + \{a(\pi^{-1}(l), \pi(i)) - a(\pi^{-1}(l), l)\}J_{\pi^{-1}(l)/n}, \tag{2.7}$$

and $Y'(\pi)$ has the distribution $\mathscr{L}(Y(\pi)\,|\,\pi(i)=l)$. Hence, putting (2.6) into (2.5), it follows that

$$\frac{1}{s(a)}\mathbb{E}\{X_i Df(Y)[J_{i/n}]\} \;=\; \frac{1}{ns(a)}\sum_{l=1}^{n} a(i,l)\mathbb{E}\{Df(Y(\pi)+\Delta_{il}(\pi))[J_{i/n}]\}. \tag{2.8}$$

Using Taylor's expansion, and recalling the definition (2.1) of $\|\cdot\|_M$, we now have

$$\begin{aligned} &|\mathbb{E}\{Df(Y+\Delta_{il})[J_{i/n}]\} - \mathbb{E}\{Df(Y)[J_{i/n}]\} - \mathbb{E}\{D^2 f(Y)[J_{i/n},\Delta_{il}]\}| \\ &\quad\le\; \|f\|_M \mathbb{E}\|\Delta_{il}\|^2, \end{aligned} \tag{2.9}$$

where, from (2.7),

$$\|\Delta_{il}(\pi)\| \;\le\; \{s(a)\}^{-1}\{|a(i,l)|+|a(i,\pi(i))|+|a(\pi^{-1}(l),\pi(i))|+|a(\pi^{-1}(l),l)|\}, \tag{2.10}$$

and thus

$$\|\Delta_{il}(\pi)\|^2 \;\le\; \frac{4}{\{s(a)\}^2}\{a^2(i,l)+a^2(i,\pi(i))+a^2(\pi^{-1}(l),\pi(i))+a^2(\pi^{-1}(l),l)\}.$$

Laborious calculation now shows that

$$\frac{1}{ns(a)}\sum_{i=1}^{n}\sum_{l=1}^{n}|a(i,l)|\mathbb{E}\|\Delta_{il}\|^2 \;\le\; C_1\frac{1}{ns^3(a)}\sum_{i=1}^{n}\sum_{l=1}^{n}|a(i,l)|^3 \;=\; C_1\Lambda(a), \tag{2.11}$$

for a universal constant C_1; for instance,

$$\begin{aligned} &\frac{1}{ns(a)}\sum_{i=1}^{n}\sum_{l=1}^{n}|a(i,l)|\frac{4}{s^2(a)}\mathbb{E}\{a^2(\pi^{-1}(l),\pi(i))\} \\ &\le\;\; \frac{4}{ns^3(a)}\sum_{i=1}^{n}\sum_{l=1}^{n}|a(i,l)|\Big\{\frac{1}{n}a^2(i,l)+\frac{1}{n(n-1)}\sum_{j\neq l}\sum_{k\neq i}a^2(k,j)\Big\} \\ &\le\;\; \frac{4}{ns^3(a)}\sum_{i=1}^{n}\sum_{l=1}^{n}\Big\{\frac{1}{n}|a(i,l)|^3+\frac{1}{n(n-1)}\sum_{j\neq l}\sum_{k\neq i}\tfrac{1}{3}\{|a(i,l)|^3+2|a(k,j)|^3\}\Big\} \\ &=\;\; \frac{4}{ns^3(a)}\sum_{i=1}^{n}\sum_{l=1}^{n}|a(i,l)|^3. \end{aligned}$$

Thus, in view of (2.8), when evaluating the right hand side of (2.4), we have

$$\begin{aligned} &\mathbb{E}Df(Y)[Y] \\ &=\;\; \frac{1}{ns(a)}\sum_{i=1}^{n}\sum_{l=1}^{n}a(i,l)\big(\mathbb{E}\{Df(Y)[J_{i/n}]\}+\mathbb{E}\{D^2 f(Y)[J_{i/n},\Delta_{il}]\}\big)+\eta_1, \end{aligned} \tag{2.12}$$

where $|\eta_1|\le C_1\Lambda(a)\|f\|_M$.

Now, because $\bar{a}(i,+)=0$, the first term on the right hand side of (2.12) is zero, so we have only the second to consider. We begin by writing

$$D^2 f(Y)[J_{i/n},\Delta_{il}] \;=\; D^2 f(Y)[J_{i/n},\mathbb{E}\Delta_{il}]+D^2 f(Y)[J_{i/n},\Delta_{il}-\mathbb{E}\Delta_{il}]. \tag{2.13}$$

From (2.7), it follows easily that

$$\begin{aligned}\mathbb{E}\{D^2 f(Y)[J_{i/n}, \mathbb{E}\Delta_{il}]\} &= \{s(a)\}^{-1} a(i,l)\mathbb{E}\{D^2 f(Y)[J_{i/n}^{(2)}]\} \\ &\quad - \frac{1}{(n-1)s(a)} \sum_{r\neq i} a(r,l)\mathbb{E}\{D^2 f(Y)[J_{i/n}, J_{r/n}]\}.\end{aligned} \tag{2.14}$$

Substituting this into (2.12) gives a contribution to $\mathbb{E}Df(Y)[Y]$ of

$$\begin{aligned}\phi_1 &:= \frac{1}{ns^2(a)} \sum_{i=1}^n \sum_{l=1}^n a^2(i,l)\mathbb{E}\{D^2 f(Y)[J_{i/n}^{(2)}]\} \\ &\quad - \frac{1}{n(n-1)s^2(a)} \sum_{i=1}^n \sum_{l=1}^n a(i,l) \sum_{r\neq i} a(r,l)\mathbb{E}\{D^2 f(Y)[J_{i/n}, J_{r/n}]\} \\ &= \sum_{i=1}^n \sigma_{ii}\mathbb{E}\{D^2 f(Y)[J_{i/n}^{(2)}]\} + \sum_{i=1}^n \sum_{r\neq i} \sigma_{ir}\mathbb{E}\{D^2 f(Y)[J_{i/n}, J_{r/n}]\},\end{aligned} \tag{2.15}$$

from (1.6). Thus, from (2.2), (2.12) and (2.13), and noting that (2.15) cancels the second term in (2.2), we deduce that

$$|\mathbb{E}(\mathscr{A} f)(Y)| \leq |\eta_1| + |\eta_2|, \tag{2.16}$$

where

$$|\eta_2| \leq \frac{1}{ns(a)} \sum_{i=1}^n \sum_{l=1}^n |a(i,l)|\,|\mathbb{E}\{D^2 f(Y)[J_{i/n}, \Delta_{il} - \mathbb{E}\Delta_{il}]\}|. \tag{2.17}$$

It thus remains to find a bound for this last expression.

To address this last step, we write

$$\begin{aligned}&\mathbb{E}\{D^2 f(Y)[J_{i/n}, \Delta_{il} - \mathbb{E}\Delta_{il}]\} \\ &\quad = \sum_{j,k=1}^n p_{jk}\mathbb{E}\{D^2 f(Y)[J_{i/n}, \Delta_{il} - \mathbb{E}\Delta_{il}] \,|\, \pi(i) = j, \pi^{-1}(l) = k\},\end{aligned}$$

where $p_{jk} := \mathbb{P}[\pi(i) = j, \pi^{-1}(l) = k]$; note that $p_{li} = 1/n$, and that $p_{jk} = 1/n(n-1)$ for $j \neq l$, $k \neq i$. We then observe that, much as for (2.6),

$$Y''(\pi) := Y(\pi) + \Delta'_{il;jk}(\pi) \sim \mathscr{L}(Y(\pi) \,|\, \pi(i) = j, \pi^{-1}(l) = k), \tag{2.18}$$

where, for $j \neq l$, $k \neq i$,

$$\begin{aligned}s(a)\Delta'_{il;jk}(\pi) := &\big\{[a(i,j) - a(i,\pi(i))]J_{i/n} + [a(k,l) - a(k,\pi(k))]J_{k/n} \\ &\quad + [a(\pi^{-1}(l),\pi(k)) - a(\pi^{-1}(l),l)]J_{\pi^{-1}(l)/n} \\ &\quad + [a(\pi^{-1}(j),\pi(i)) - a(\pi^{-1}(j),j)]J_{\pi^{-1}(j)/n}\big\}\mathbf{1}_{\{\pi(i)\neq l,\, \pi(k)\neq j\}} \\ &+ \big\{[a(i,j) - a(i,\pi(i))]J_{i/n} + [a(k,l) - a(k,j)]J_{k/n} \\ &\quad + [a(\pi^{-1}(l),\pi(i)) - a(\pi^{-1}(l),l)]J_{\pi^{-1}(l)/n}\big\}\mathbf{1}_{\{\pi(i)\neq l,\, \pi(k)=j\}} \\ &+ \big\{[a(i,j) - a(i,l)]J_{i/n} + [a(k,l) - a(k,\pi(k))]J_{k/n} \\ &\quad + [a(\pi^{-1}(j),\pi(k)) - a(\pi^{-1}(j),j)]J_{\pi^{-1}(j)/n}\big\}\mathbf{1}_{\{\pi(i)=l\}},\end{aligned} \tag{2.19}$$

and

$$s(a)\Delta'_{il;li}(\pi) := [a(i,l) - a(i,\pi(i))]J_{i/n} + [a(\pi^{-1}(l),\pi(i)) - a(\pi^{-1}(l),l)]J_{\pi^{-1}(l)/n}. \tag{2.20}$$

Then $\Delta_{il} = \Delta_{il}(\pi(i),\pi^{-1}(l))$ is measurable with respect to $\sigma(\pi(i),\pi^{-1}(l))$, and

$$\begin{aligned}
&\sum_{j=1}^{n}\sum_{k=1}^{n} p_{jk}\mathbb{E}\{D^2 f(Y)[J_{i/n}, \Delta_{il} - \mathbb{E}\Delta_{il}] \,|\, \pi(i) = j, \pi^{-1}(l) = k\} \\
&\quad = \sum_{j=1}^{n}\sum_{k=1}^{n} p_{jk}\mathbb{E}\{D^2 f(Y + \Delta'_{il;jk})[J_{i/n}, \Delta_{il}(j,k) - \mathbb{E}\Delta_{il}]\} \\
&\quad = \sum_{j=1}^{n}\sum_{k=1}^{n} p_{jk}\mathbb{E}\{D^2 f(Y)[J_{i/n}, \Delta_{il}(j,k) - \mathbb{E}\Delta_{il}]\} \\
&\qquad + \sum_{j=1}^{n}\sum_{k=1}^{n} p_{jk}\mathbb{E}\{D^2 f(Y + \Delta'_{il;jk})[J_{i/n}, \Delta_{il}(j,k) - \mathbb{E}\Delta_{il}] \\
&\qquad\qquad - D^2 f(Y)[J_{i/n}, \Delta_{il}(j,k) - \mathbb{E}\Delta_{il}]\}.
\end{aligned} \tag{2.21}$$

Now, since $\sum_{j=1}^{n}\sum_{k=1}^{n} p_{jk}\Delta_{il}(j,k) = \mathbb{E}\Delta_{il}$, the first term in (2.21) is zero, by bilinearity. For the remainder, we have

$$\begin{aligned}
&\|D^2 f(Y + \Delta'_{il;jk})[J_{i/n}, \Delta_{il}(j,k) - \mathbb{E}\Delta_{il}] - D^2 f(Y)[J_{i/n}, \Delta_{il}(j,k) - \mathbb{E}\Delta_{il}]\| \\
&\quad \le \|f\|_M \|\Delta'_{il;jk}\|\{\|\Delta_{il}(j,k)\| + \|\mathbb{E}\Delta_{il}\|\},
\end{aligned} \tag{2.22}$$

so that, from (2.17),

$$|\eta_2| \le \|f\|_M \frac{1}{ns(a)} \sum_{i=1}^{n}\sum_{l=1}^{n} |a(i,l)| \sum_{j=1}^{n}\sum_{k=1}^{n} p_{jk}\mathbb{E}\|\Delta'_{il;jk}\|\{\|\Delta_{il}(j,k)\| + \|\mathbb{E}\Delta_{il}\|\}. \tag{2.23}$$

Here, from (2.7), (2.10) and (2.19), each of the norms can be expressed as $1/s(a)$ times a sum of elements of $|a|$. Another laborious calculation shows that indeed

$$|\eta_2| \le C_2\Lambda(a)\|f\|_M,$$

and the theorem is proved. □

3 A functional limit theorem

The pre-limiting approximation is simpler than the original process, inasmuch as it involves only jointly Gaussian random variables with prescribed covariances. However, if the matrix a can be naturally imbedded into a sequence $a^{(n)}$ exhibiting some regularity as n varies, and if n is large, it may be advantageous to look for an n-independent limiting approximation, in the usual sense of weak convergence. Unfortunately, the approximation given in Theorem 2.1 is not naturally compatible with weak convergence with respect to the Skorohod metric, and something extra is needed.

With this in mind, we prove the following extension of Theorem 2 of Barbour (1990). To do so, we introduce the class of functionals $g \in M^0 \subset M$ for which

$$\|g\|_{M^0} \;:=\; \|g\|_M + \sup_{w\in D}|g(w)| + \sup_{w\in D}\|Dg(w)\| + \sup_{w\in D}\|D^2 g(w)\| \;<\; \infty.$$

Proposition 3.1. *Suppose that, for each $n \ge 1$, the random element Y_n of $D := D[0,1]$ is piecewise constant, with intervals of constancy of length at least r_n. Let Z_n, $n \ge 1$, be random elements of D converging weakly in D to a random element Z of $C[0,1]$. Then, if*

$$|\mathbb{E}g(Y_n) - \mathbb{E}g(Z_n)| \;\le\; C\tau_n\|g\|_{M^0} \tag{3.1}$$

for each $g \in M^0$, and if $\tau_n \log^2(1/r_n) \to 0$ as $n \to \infty$, then $Y_n \to Z$ in D.

Proof. First note that, by Skorohod's representation theorem, we may assume that the processes Z_n and Z are all defined on the same probability space, in such a way that $Z_n \to Z$ in D a.s. as $n \to \infty$. Since Z is continuous, this implies that $\|Z_n - Z\| \to 0$ a.s.

As in the proof of Barbour (1990, Theorem 2), it is enough to show that

$$\mathbb{P}[Y_n \in B] \to \mathbb{P}[Z \in B] \tag{3.2}$$

for all sets B of the form $\bigcap_{1\le l\le L} B_l$, where $B_l = \{w \in D\colon \|w - s_l\| < \gamma_l\}$ for $s_l \in C[0,1]$, and $\mathbb{P}[Z \in \partial B_l] = 0$. To do so, we approximate the indicators $I[Y_n \in B_l]$ from above and below by functions from a family $g := g\{\varepsilon,p,\rho,\eta,s\}$ in M^0, and use (3.1). We define

$$g\{\varepsilon,p,\rho,\eta,s\}(w) \;:=\; \phi_{\rho,\eta}(h_{\varepsilon,p}(w-s)),$$

where

$$h_{\varepsilon,p}(y) \;:=\; \left(\int_0^1 (\varepsilon^2 + y^2(t))^{p/2}\,dt\right)^{1/p} \;=:\; \|(\varepsilon^2+y^2)^{1/2}\|_p,$$

and $\phi_{\rho,\eta}(x) := \phi((x-\rho)/\eta)$, for $\phi\colon \mathbb{R}^+ \to [0,1]$ non-increasing, three times continuously differentiable, and such that $\phi(x) = 1$ for $x \le 0$ and $\phi(x) = 0$ for $x \ge 1$. Note that each such function g is in M^0, and that $\|g\|_{M^0} \le C'p^2\varepsilon^{-2}\eta^{-3}$ for a constant C' not depending on $\varepsilon,p,\rho,\eta,s$, and that the same is true for finite products of such functions, if the largest of the p's and the smallest of the ε's and η's is used in the norm bound.

Now, if $x \in B_l$, it follows that $g_l(x) = 1$, for

$$g_l \;:=\; g\{\varepsilon\gamma_l, p, \gamma_l(1+\varepsilon^2)^{1/2}, \eta, s_l\},$$

for all ε,p,η. Hence, for all ε,p,η,

$$\mathbb{P}\Big[Y_n \in \bigcap_{1\le l\le L} B_l\Big] \;\le\; \mathbb{E}\Big\{\prod_{i=1}^{L} g_l(Y_n)\Big\} \;\le\; \mathbb{E}\Big\{\prod_{i=1}^{L} g_l(Z_n)\Big\} + C\tau_n C_B' p^2(\varepsilon\gamma)^{-2}\eta^{-3}, \tag{3.3}$$

where $\gamma := \min_{1\le l\le L}\gamma_l$. Then, by Minkowski's inequality,

$$h_{\varepsilon,p}(Z - s_l) \;\le\; h_{\varepsilon,p}(Z_n - s_l) + \|Z_n - Z\|_p \;\le\; h_{\varepsilon,p}(Z_n - s_l) + \|Z_n - Z\|.$$

Hence, if $p_n \to \infty$ as $n \to \infty$ and ε is fixed,

$$\liminf_{n\to\infty} h_{\varepsilon,p_n}(Z_n - s_l) \;\geq\; \liminf_{n\to\infty}\{h_{\varepsilon,p_n}(Z - s_l) - \|Z_n - Z\|\} \;=\; \|(\varepsilon^2 + |Z - s_l|^2)^{1/2}\|$$

a.s. It thus follows that, if $\|Z - s_l\| > \gamma_l$, and if $\eta_n \to 0$ as $n \to \infty$, then

$$\liminf_{n\to\infty}\{h_{\varepsilon\gamma_l,p_n}(Z_n - s_l) - \eta_n\} \;\geq\; \|(\varepsilon^2\gamma_l^2 + |Z - s_l|^2)^{1/2}\| \;>\; \gamma_l(1+\varepsilon^2)^{1/2}$$

a.s., and so $g_{ln}(Z_n) = 0$ for all n sufficiently large, where

$$g_{ln} := g\{\varepsilon\gamma_l, p_n, \gamma_l(1+\varepsilon^2)^{1/2}, \eta_n, s_l\}.$$

Applying Fatou's lemma to $1 - \prod_{l=1}^L g_{ln}(Z_n)$, and because $\mathbb{P}[Z \in \partial B_l] = 0$ for each l, we then have,

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{E}\Big\{\prod_{i=1}^L g_{ln}(Z_n)\Big\} &\leq \mathbb{E}\Big\{\limsup_{n\to\infty} \prod_{i=1}^L g_{ln}(Z_n)\Big\} \\ &\leq \mathbb{E}\Big(\prod_{i=1}^L \mathbf{1}\{\|Z - s_l\| \leq \gamma_l\}\Big) \;=\; \mathbb{P}[Z \in B].\end{aligned}$$

Thus, letting $p_n \to \infty$ and $\eta_n \to 0$ in such a way that $\tau_n p_n^2 \eta_n^{-3} \to 0$, it follows from (3.3) that $\limsup_{n\to\infty} \mathbb{P}[Y_n \in B] \leq \mathbb{P}[Z \in B]$, and we have proved one direction of (3.2).

For the other direction, fix $\theta > 0$ small, and let $\delta > 0$ be such that, if $\|Y_n - s_l\| \geq \gamma_l$, then

$$\text{leb}\big\{t\colon\, |Y_n(t) - s_l| \geq \gamma_l(1-\theta)\big\} \geq \big(\delta \wedge \tfrac{1}{2} r_n\big), \tag{3.4}$$

where leb$\{\cdot\}$ denotes Lebesgue measure. Such a δ exists, because the collection $(s_l,\, 1 \leq l \leq L)$ is uniformly equicontinuous, and because the functions Y_n are piecewise constant on intervals of length at least r_n. Hence, for such Y_n,

$$h_{\varepsilon\gamma_l,p}(Y_n - s_l) \;\geq\; \gamma_l\{\varepsilon^2 + (1-\theta)^2\}^{1/2}\big(\delta \wedge \tfrac{1}{2} r_n\big)^{1/p},$$

and thus $g_l^*(Y_n) = 0$, where, for any p and η,

$$g_l^* \;:=\; g\big\{\varepsilon\gamma_l, p, \gamma_l(\varepsilon^2 + (1-\theta)^2)^{1/2}\big(\delta \wedge \tfrac{1}{2} r_n\big)^{1/p} - \eta, \eta, s_l\big\}.$$

Thus, for any p and h, $I[Y_n \in B_l] \geq g_l^*(Y_n)$, and hence

$$\mathbb{P}\Big[Y_n \in \bigcap_{1\leq l\leq L} B_l\Big] \;\geq\; \mathbb{E}\Big\{\prod_{i=1}^L g_l^*(Y_n)\Big\} \;\geq\; \mathbb{E}\Big\{\prod_{i=1}^L g_l^*(Z_n)\Big\} - C\tau_n C_B' p^2 (\varepsilon\gamma)^{-2}\eta^{-3}. \tag{3.5}$$

Now suppose that $\|Z - s_l\| < \gamma_l(1-\theta)$. Then there exists an $\alpha > 0$ such that a.s. $\|Z_n - s_l\| < \gamma_l(1-\theta) - \alpha$ for all n sufficiently large. This in turn implies that

$$\begin{aligned}h_{\varepsilon\gamma_l,p_n}(Z_n - s_l) &\leq \{\varepsilon^2\gamma_l^2 + \|Z_n - s_l\|^2\}^{1/2} \;\leq\; \gamma_l\{\varepsilon^2 + (1-\theta-\alpha\gamma_l^{-1})^2\}^{1/2} \\ &< \gamma_l\{\varepsilon^2 + (1-\theta)^2\}^{1/2}\big(\delta \wedge \tfrac{1}{2} r_n\big)^{1/p_n} - \eta_n\end{aligned}$$

for all n large enough, if $\eta_n \to 0$ and $p_n \to \infty$ in such a way that $r_n^{1/p_n} \to 1$. This in turn implies that $g^*_{ln}(Z_n) = 1$ for all n large enough, where

$$g^*_{ln} \;:=\; g\{\varepsilon\gamma_l, p_n, \gamma_l(\varepsilon^2 + (1-\theta)^2)^{1/2}(\delta \wedge \tfrac{1}{2}r_n)^{1/p_n} - \eta_n, \eta_n, s_l\}. \tag{3.6}$$

Hence

$$\mathbb{E}\Big\{\liminf_{n\to\infty}\prod_{i=1}^{L} g^*_{ln}(Z_n)\Big\} \;\geq\; \mathbb{P}\Big[\bigcap_{1\leq l\leq L}\{\|Z - s_l\| < \gamma_l(1-\theta)\}\Big]. \tag{3.7}$$

Applying Fatou's lemma, and recalling (3.5), we now have a.s.

$$\liminf_{n\to\infty}\mathbb{P}\Big[Y_n \in \bigcap_{1\leq l\leq L} B_l\Big] \;\geq\; \liminf_{n\to\infty}\mathbb{E}\Big\{\prod_{i=1}^{L} g^*_{ln}(Z_n)\Big\} \;\geq\; \mathbb{E}\Big\{\liminf_{n\to\infty}\prod_{i=1}^{L} g^*_{ln}(Z_n)\Big\}, \tag{3.8}$$

provided that also $\tau_n p_n^2 \eta_n^{-3} \to 0$: this can be arranged by judicious choice of $p_n \to \infty$ and $\eta_n \to 0$ if, as assumed, $\tau_n \log^2(1/r_n) \to 0$. Hence, since θ was chosen arbitrarily, it follows from (3.7) and (3.8) that

$$\liminf_{n\to\infty}\mathbb{P}[Y_n \in B] \geq \mathbb{P}[Z \in B],$$

and the theorem is proved. □

Note that, in Barbour (1990, Theorem 2), restricting to functions g satisfying (2.32) of that paper is not permissible: the bound (3.1) is needed for functions in M_0 that do not necessarily satisfy (2.32).

Remark 3.2. The assumption that Y_n is piecewise constant can be relaxed to Y_n being piecewise linear, with intervals of linearity of length at least r_n; in particular, this allows processes Y_n obtained by linear interpolation. The only difference in the proof is that, if $\|Y_n - s_l\| \geq \gamma_l$, then $|Y_n(t_0) - s_l(t_0)| > (1-\theta/4)\gamma_l$ for some t_0. Thus, by the assumption on Y_n and the continuity of s_l, there exists an interval I_0 of length at least $l_n := \frac{1}{2}r_n \wedge \delta$, with t_0 as an endpoint, on which Y_n is linear and $|s_l(t) - s_l(t_0)| < \theta\gamma_l/4$. A simple geometrical argument now shows that $|Y_n(t) - s_l(t_0)| > (1-\theta/2)\gamma_l$ in a subinterval of length at least $\theta l_n/8$, at one or other end of I_0. Hence, (3.4) can be replaced by

$$\mathrm{leb}\{t\colon\, |Y_n(t) - s_l| \geq \gamma_l(1-\theta)\} \geq \tfrac{\theta}{16}(\delta \wedge r_n),$$

and the rest of the proof is the same.

We now turn to proving a functional limit theorem for the sums derived from a sequence of matrices $a^{(n)}$, $n \geq 1$. Supposing that $s^{(n)}(a) > 0$, we define functions

$$\begin{aligned} f_n(t) &:= \frac{1}{n(s^{(n)}(a))^2}\sum_{i=1}^{\lfloor nt\rfloor}\sum_{l=1}^{n}(a^{(n)}(i,l))^2; \\ g_n(t,u) &:= \frac{1}{(ns^{(n)}(a))^2}\sum_{i=1}^{\lfloor nt\rfloor}\sum_{j=1}^{\lfloor nu\rfloor}\sum_{l=1}^{n} a^{(n)}(i,l)a^{(n)}(j,l), \end{aligned} \tag{3.9}$$

for $0 \leq t,u \leq 1$. Note that if we choose $s^{(n)}(a)$ by (1.8), then $f_n(1) = (n-1)/n \to 1$. Conversely, if $f_n(1)$ converges to a limit $c > 0$, then $s^{(n)}(a)$ differs from the value in (1.8) only by a factor $c^{-1/2} + o(1)$.

Theorem 3.3. *Suppose that $f_n \to f$ and $g_n \to g$ pointwise, with f continuous, and that $\Lambda^{(n)}(a)\log^2 n \to 0$. Then there exists a zero mean continuous Gaussian process Z on $[0,1]$ with covariance function given by*

$$\mathrm{Cov}(Z(t),Z(u)) \;=\; \sigma(t,u) \;:=\; f(t\wedge u)-g(t,u), \tag{3.10}$$

and $Y_n \to Z$ in $D[0,1]$.

Proof. Fix $n \geq 2$. We begin by realizing the random variables $W_i^{(n)}$ as functions of a collection $(X_{il},\, i,l \geq 1)$ of independent standard normal random variables. Writing $\overline{X}_l := n^{-1}\sum_{i=1}^n X_{il}$, we set

$$W_{il}^{(n)} \;:=\; \frac{1}{s^{(n)}(a)\sqrt{n-1}}\, a^{(n)}(i,l)(X_{il}-\overline{X}_l); \qquad W_i^{(n)} \;:=\; \sum_{l=1}^n W_{il}^{(n)}. \tag{3.11}$$

Direct calculation shows that, with δ_{ij} the Kronecker delta,

$$\begin{aligned}\mathrm{Cov}(W_i^{(n)},W_j^{(n)}) \;&=\; \sum_{l=1}^n \mathrm{Cov}(W_{il}^{(n)},W_{jl}^{(n)}) \\ &=\; \sum_{l=1}^n \frac{1}{(n-1)(s^{(n)}(a))^2}\, a^{(n)}(i,l)a^{(n)}(j,l)(\delta_{ij}-n^{-1}),\end{aligned} \tag{3.12}$$

in accordance with (1.6), so we can set

$$Z_n \;:=\; \sum_{i=1}^n W_i^{(n)} J_{i/n}. \tag{3.13}$$

Now Theorem 2.1 shows that $|\mathbb{E}\{g(Y_n)-g(Z_n)\}| \leq C\Lambda^{(n)}(a)\|g\|_{M^0}$ for any $g \in M^0$; furthermore, the process Y_n is piecewise constant on intervals of lengths $1/n$, and, by assumption, $\Lambda^{(n)}(a)\log^2 n \to 0$. Hence, in order to apply Proposition 3.1, it is enough to show that $Z_n \to Z$ for a continuous Gaussian process.

Write $Z_n = Z_n^{(1)} - Z_n^{(2)}$, where

$$\begin{aligned} Z_n^{(1)}(t) \;&:=\; \frac{1}{s^{(n)}(a)\sqrt{n-1}} \sum_{i=1}^{\lfloor nt\rfloor}\sum_{l=1}^n a^{(n)}(i,l)X_{il}, \\ Z_n^{(2)}(t) \;&:=\; \frac{1}{s^{(n)}(a)\sqrt{n-1}} \sum_{i=1}^{\lfloor nt\rfloor}\sum_{l=1}^n a^{(n)}(i,l)\overline{X}_l. \end{aligned} \tag{3.14}$$

The process $Z_n^{(1)}$ is a Gaussian process with independent increments, and can be realized as $W(\tilde{f}_n(\cdot))$, where W is a standard Brownian motion and $\tilde{f}_n(t) := nf_n(t)/(n-1)$. Now f is continuous, by assumption, and each $\tilde{f}_n$ is non-decreasing, so $\tilde{f}_n \to f$ uniformly on $[0,1]$, and hence $W(\tilde{f}_n(\cdot)) \to W(f(\cdot))$ in $D[0,1]$. Since the latter process is continuous, it follows that the sequence $Z_n^{(1)}$ is C-tight in $D[0,1]$.

To show that $Z_n^{(2)}$ is also C-tight, we use criteria from Billingsley (1968). For $0 \le t \le u \le 1$, it follows from (3.14) and Hölder's inequality that

$$\begin{aligned}
\mathbb{E}|Z_n^{(2)}(u) - Z_n^{(2)}(t)|^2 &= \frac{1}{(n-1)(s^{(n)}(a))^2} \sum_{l=1}^{n} \Big(\sum_{i=\lfloor nt \rfloor+1}^{\lfloor nu \rfloor} a^{(n)}(i,l) \Big)^2 \frac{1}{n} \\
&\le \frac{1}{n(n-1)(s^{(n)}(a))^2} (\lfloor nu \rfloor - \lfloor nt \rfloor) \sum_{l=1}^{n} \sum_{i=\lfloor nt \rfloor+1}^{\lfloor nu \rfloor} (a^{(n)}(i,l))^2 \\
&\le f_n(1) \frac{\lfloor nu \rfloor - \lfloor nt \rfloor}{n-1}.
\end{aligned}$$

Hence, since $Z_n^{(2)}$ is Gaussian, we have

$$\mathbb{E}|Z_n^{(2)}(u) - Z_n^{(2)}(t)|^4 = 3(\mathbb{E}|Z_n^{(2)}(u) - Z_n^{(2)}(t)|^2)^2 \le 3\Big(f_n(1) \frac{\lfloor nu \rfloor - \lfloor nt \rfloor}{n-1} \Big)^2. \qquad (3.15)$$

Thus, if $0 \le t \le v \le u \le 1$ and $u - t \ge 1/n$, it follows that

$$\begin{aligned}
&\mathbb{E}\left\{ |Z_n^{(2)}(v) - Z_n^{(2)}(t)|^2 |Z_n^{(2)}(u) - Z_n^{(2)}(v)|^2 \right\} \\
&\le \sqrt{\mathbb{E}|Z_n^{(2)}(v) - Z_n^{(2)}(t)|^4 \, \mathbb{E}|Z_n^{(2)}(u) - Z_n^{(2)}(v)|^4} \\
&\le 3 f_n^2(1) \Big(\frac{\lfloor nv \rfloor - \lfloor nt \rfloor}{n-1} \, \frac{\lfloor nu \rfloor - \lfloor nv \rfloor}{n-1} \Big) \le 12 f_n^2(1)(u-t)^2; \qquad (3.16)
\end{aligned}$$

the inequality is immediate for $u - t < 1/n$, since then $\lfloor nv \rfloor \in \{\lfloor nt \rfloor, \lfloor nu \rfloor\}$.

Now, for any $0 \le t \le u \le 1$, we have

$$\mathrm{Cov}(Z_n^{(2)}(t), Z_n^{(2)}(u)) = \frac{n}{n-1} g_n(t,u) \to g(t,u).$$

Hence there exists a zero mean Gaussian process $Z^{(2)}$ with covariance function g, and the finite dimensional distributions of $Z_n^{(2)}$ converge to those of $Z^{(2)}$. By (3.15) and Fatou's lemma, $\mathbb{E}|Z^{(2)}(u) - Z^{(2)}(t)|^4 \le 3 f_n^2(1)(u-t)^2$ for any $0 \le t \le u \le 1$, so that, from Billingsley (1968, Theorem 12.4), we may assume that $Z^{(2)} \in C[0,1]$. From (3.16) and Billingsley (1968, Theorem 15.6), it now follows that $Z_n^{(2)} \to Z^{(2)}$ in $D[0,1]$. Thus $Z_n^{(2)}$ is C-tight also.

Now, since both $\{Z_n^{(1)}\}$ and $\{Z_n^{(2)}\}$ are C-tight, so is their difference $\{Z_n\}$. From (3.9) and (3.12), for $t, u \in [0,1]$,

$$\mathrm{Cov}(Z_n(t), Z_n(u)) = \frac{n}{n-1} f_n(t \wedge u) - \frac{n}{n-1} g_n(t,u) \to f(t \wedge u) - g(t,u),$$

so that the finite dimensional distributions of Z_n converge to those of a random element Z of $C[0,1]$ with covariance function $\sigma(t,u)$, as required. □

4 Rate of convergence

Under more stringent assumptions, the approximation of Z_n by Z can be made sharper. To start with, note that it follows from the representation (3.11) and (3.13) that Z_n can be written as a two dimensional stochastic integral

$$Z_n(t) \;=\; \frac{n}{s^{(n)}(a)\sqrt{n-1}} \int_{I_n(t)\times I} \alpha_n(v,w)K(dv,dw) \tag{4.1}$$

with respect to a Kiefer process K, where $I_n(t) := [0, n^{-1}\lfloor nt \rfloor]$, $I := [0,1]$ and $\alpha_n(v,w) := a^{(n)}(\lceil nv \rceil, \lceil nw \rceil)$. Recall that the Kiefer process K has covariance function $\mathrm{Cov}(K(v_1,w_1),K(v_2,w_2)) = (v_1 \wedge v_2 - v_1v_2)(w_1 \wedge w_2)$ and can be represented in the form $K(v,w) = W(v,w) - vW(1,w)$, where W is the two-dimensional Brownian sheet (Shorack & Wellner 1986, (5) p. 30 and Exercise 12, p. 32). Thus K is like a Brownian bridge in v, and a Brownian motion in w.

In this section, we let $s(a^{(n)})$ be given by (1.8). Hence if, for example, the functions α_n converge in L_2 to a square integrable limit α (not a.e. 0), then,

$$\frac{n-1}{n^2}\{s^{(n)}(a)\}^2 \;=\; \|\alpha_n\|_2^2 \;\to\; \sigma_a^2 \;:=\; \int_0^1 dv \int_0^1 dw\, \alpha^2(v,w) \;=\; \|\alpha\|_2^2,$$

and the limiting process Z can be represented as

$$Z(t) \;=\; \sigma_a^{-1} \int_{[0,t]\times I} \alpha(v,w)K(dv,dw), \tag{4.2}$$

enabling a direct comparison between Z_n and Z to be made. Since $\alpha_n \to_{L_2} \alpha$, it follows that

$$\begin{aligned} f_n(t) \to f(t) \;&:=\; \sigma_a^{-2} \int_0^t dv \int_0^1 dw\, \alpha^2(v,w); \\ g_n(t,u) \to g(t,u) \;&:=\; \sigma_a^{-2} \int_0^t dv \int_0^u dx \int_0^1 dw\, \alpha(v,w)\alpha(x,w), \end{aligned} \tag{4.3}$$

with f continuous, as required for Theorem 3.3, and that Z has covariance function $\sigma(t,u)$ as defined in (3.10). For the following lemma, we work under silghtly stronger assumptions.

Lemma 4.1. *Suppose that $\alpha_n \to \alpha$ in L_2, where α is bounded and not a.e. 0, and that, for some $0 < \beta \le 2$,*

$$|g(t,t) + g(u,u) - 2g(t,u)| \;\le\; C_g^2 |u-t|^\beta, \qquad 0 \le t \le u \le 1. \tag{4.4}$$

Define $\alpha^+ := \|\alpha\|_\infty / \|\alpha\|_2 < \infty$ and $\varepsilon_n(v,w) := \|\alpha\|_2^{-1}\{\alpha_n(v,w) - \alpha(v,w)\}$. Then, for any $r > 0$, there is a constant $c(r)$ such that

$$\mathbb{P}\big[\sup_{t\in I} |Z_n(t) - Z(t)| > c(r)\{\|\varepsilon_n\|_2 + (\alpha^+ + C_g)n^{-(\beta\wedge 1)/2}\}\sqrt{\log n}\big] \;\le\; n^{-r},$$

where Z is as defined in (4.2).

Proof. Define $\tilde{\varepsilon}_n(v,w) := \frac{n}{s^{(n)}(a)\sqrt{n-1}}\alpha_n(v,w) - \sigma_a^{-1}\alpha(v,w)$. We start by considering t of the form i/n, $1 \le i \le n$, so that

$$Z_n(t) - Z(t) \;=\; \int_{[0,t]\times I} \tilde{\varepsilon}_n(v,w) K(dv,dw).$$

From this and the representation $K(v,w) = W(v,w) - vW(1,w)$, it follows that $\max_{t\in I} \mathbb{E}\{Z_n(t) - Z(t)\}^2 \le \|\tilde{\varepsilon}_n\|_2^2$, and hence, from the Borell–TIS maximal inequality for Gaussian processes (Adler and Taylor 2007, Theorem 2.1.1), we have

$$\mathbb{P}\Big[\max_{t\in n^{-1}\{1,2,\ldots,n\}} \Big|\int_{[0,t]\times I} \tilde{\varepsilon}_n(v,w)K(dv,dw)\Big| > c_1(r)\|\tilde{\varepsilon}_n\|_2\sqrt{\log n}\Big] \;\le\; \tfrac{1}{2}n^{-r},$$

if $c_1(r)$ is chosen large enough. However,

$$\tilde{\varepsilon}_n \;=\; \frac{\alpha_n}{\|\alpha_n\|_2} - \frac{\alpha}{\|\alpha\|_2},$$

from which it follows that

$$\|\tilde{\varepsilon}_n\|_2 \;\le\; 2\|\varepsilon_n\|_2.$$

It thus remains to consider the differences $Z_n(t) - Z(t)$ for t not of the form i/n. Between $n^{-1}\lfloor nt\rfloor$ and t, the process Z_n remains constant, whereas Z changes; hence it is enough to control the maximal fluctuation of Z over intervals of the form $[(i-1)/n, i/n]$, $1 \le i \le n$. Here, we use the Fernique–Marcus maximal inequality for Gaussian processes (Leadbetter *et al.* 1983, Lemma 12.2.1), together with the inequality

$$|\sigma(u,u) + \sigma(t,t) - 2\sigma(t,u)| \;\le\; C_g^2|t-u|^\beta + (\alpha^+)^2|t-u|,$$

to give the bound

$$\mathbb{P}\Bigg[\max_{1\le i\le n}\sup_{\frac{i-1}{n}\le v\le\frac{i}{n}} |Z(v) - Z((i-1)/n)| > c_2(r)(C_g + \alpha^+)n^{-(\beta\wedge 1)/2}\sqrt{\log n}\Bigg] \;\le\; \tfrac{1}{2}n^{-r},$$

if $c_2(r)$ is chosen large enough, and the proof is now complete. □

Note that, under the conditions of Lemma 4.1, the requirements for Theorem 3.3 are fulfilled, provided that $\Lambda^{(n)}(a) \to 0$ fast enough. This is true if also, for instance, for some $c < \infty$, $\|\alpha_n\|_\infty \le c\|\alpha\|_\infty$ for all n, since then $\Lambda^{(n)}(a) \;\le\; 2c\alpha^+ n^{-1/2}$ for all n large enough. Combining Theorems 2.1 and 3.3 with Lemma 4.1 then easily gives the following conclusions.

Theorem 4.2. *Under the conditions of Lemma 4.1, and if also $\|\alpha_n\|_\infty/\|\alpha\|_\infty$ is bounded, then $Y_n \to_d Z$ in $D[0,1]$, for Z as defined in* (4.2)*, and, for any functional $g \in M_0$,*

$$\begin{aligned}&|\mathbb{E}g(Y_n) - \mathbb{E}g(Z)| \\ &\quad\le\; C\big\{\Lambda^{(n)}(a) + n^{-1} + \{\|\varepsilon_n\|_2 + (\alpha^+ + C_g)n^{-(\beta\wedge 1)/2}\}\sqrt{\log n}\big\}\,\|g\|_{M^0},\end{aligned}$$

for some constant C.

Proof. We note that

$$|\mathbb{E}g(Y_n) - \mathbb{E}g(Z)| \;\le\; |\mathbb{E}g(Y_n) - \mathbb{E}g(Z_n)| + \mathbb{E}|g(Z_n) - g(Z)|.$$

The first term is bounded using Theorem 2.1, whereas, for any $a > 0$,

$$\begin{aligned}\mathbb{E}|g(Z_n) - g(Z)| &\le 2\sup_{w\in D}|g(w)|\mathbb{P}[\|Z_n - Z\|_\infty > a] + a\sup_{w\in D}\|Dg(w)\| \\ &\le \|g\|_{M_0}\{2\mathbb{P}[\|Z_n - Z\|_\infty > a] + a\},\end{aligned}$$

and the theorem follows by taking $a = c(1)\{\|\varepsilon_n\|_2 + (\alpha^+ + C_g)n^{-(\beta\wedge 1)/2}\}\sqrt{\log n}$ and applying Lemma 4.1 with $r = 1$. □

5 The shape of permutation tableaux

We begin by studying the number of *weak exceedances* in a uniform random permutation π on $\{1,2,\ldots,n\}$; we shall suppress the index n where possible. The number of weak exceedances is defined to be the sum $\sum_{i=1}^n I_i$, where $I_i := \mathbf{1}_{\{\pi(i)\ge i\}}$. The process $S_0(t) := \sum_{i=1}^{\lfloor nt\rfloor} I_i$ is thus of the kind studied in the introduction, with $a_0(i,j) := \mathbf{1}_{\{i\le j\}}$. Simple calculations show that $\mathbb{E}I_i = \bar{a}_0(i,+) = (n-i+1)/n$, and thus

$$a(i,j) \;=\; a^{(n)}(i,j) \;=\; \mathbf{1}_{\{i\le j\}} - 1 + (i-1)/n, \tag{5.1}$$

$$\mathbb{E}S_0(k/n) \;=\; \frac{k(2n-k+1)}{2n}. \tag{5.2}$$

Hence, as $n\to\infty$,

$$\mathbb{E}S_0(t) \;=\; nt(1-t/2) + O(1). \tag{5.3}$$

Further, although we will not need it, for $i < j$,

$$\mathbb{E}\{I_i \,|\, I_j = 1\} \;=\; \frac{n-i}{n-1}, \qquad \mathbb{E}\{I_iI_j\} \;=\; \frac{(n-i)(n-j+1)}{(n-1)n},$$

which makes it possible to calculate variances and covariances exactly. Higher moments can be computed exactly, too.

We now turn to the approximation of $S(t) := S_0(t) - \mathbb{E}S_0(t)$. We first note that

$$|a(i,j) - \alpha(i/n, j/n)| \;\le\; n^{-1},$$

where $\alpha(t,u) := \mathbf{1}_{\{t\le u\}} - 1 + t$, so that $|\alpha_n(t,u) - \alpha(t,u)| \le 2n^{-1}$ for $|t-u| > n^{-1}$, and that $|\alpha_n(t,u) - \alpha(t,u)| \le 1$ for all $t,u\in I$. Thus $\alpha_n \to \alpha$ in L_2, with

$$\|\alpha\|_2^2 \;=\; 1/6; \qquad \|\varepsilon_n\|_2^2 \;\le\; 18/n; \qquad \alpha^+ \;=\; \sqrt{6},$$

and $\|\alpha_n\|_\infty/\|\alpha\|_\infty$ is bounded. Calculation based on (4.3) shows also that, for $0 \le t \le u \le 1$,

$$\begin{aligned} f(t) &= 6\int_0^t x(1-x)\,dx \;=\; 3t^2 - 2t^3; \\ g(t,u) &= 6\int_0^t \int_0^u \{(1-x\vee y) - (1-x)(1-y)\}\,dx\,dy \\ &= 3t^2u - t^3 - \tfrac{3}{2}t^2u^2, \end{aligned}$$

and that we can take $\beta = 2$ in (4.4). Hence we can apply Theorem 4.2, and defining Y_n by (1.3) with (1.8), conclude that $Y_n \to Z$ in $D[0,1]$, with convergence rate $O\big(n^{-1/2}\sqrt{\log n}\big)$ as measured by M_0-functionals, where Z is the Gaussian process given by (4.2):

$$Z(t) \;=\; \sqrt{6}\int_{[0,t]\times I} \{\mathbf{1}_{\{v\le w\}} - 1 + v\}\,K(dv,dw).$$

Note also that

$$Y_n(t) \;=\; \sqrt{6/n}\{S_0(t) - nt(1-t/2)\} + O(n^{-1/2}), \tag{5.4}$$

indicating that the approximation can be simplified, as in the following theorem.

Theorem 5.1. *Let $S_0^{(n)}(t) := \sum_{i=1}^{\lfloor nt\rfloor} I_i^{(n)}$, where $I_i^{(n)} := \mathbf{1}_{\{\pi^{(n)}(i)\ge i\}}$ and $\pi^{(n)}$ is a uniform random permutation on $\{1,2,\ldots,n\}$. Write $\mu(t) := t(1-t/2)$. Then*

$$\widehat{Y}_n \;:=\; n^{-1/2}\,\{S_0^{(n)} - n\mu\} \;\to_d\; \widehat{Z} \quad \text{in} \quad D[0,1],$$

where $\widehat{Z}$ is a zero mean Gaussian process with covariance function $\widehat{\sigma}$ given by

$$\widehat{\sigma}(t,u) \;=\; \tfrac{1}{6}\sigma(t,u) \;=\; \tfrac{1}{6}(f(t) - g(t,u)) \;=\; \tfrac{1}{2}t^2(1 - u + \tfrac{1}{2}u^2) - \tfrac{1}{6}t^3, \qquad 0 \le t \le u \le 1.$$

The number of weak exceedances of a permutation is one of a number of statistics that can be deduced from the permutation tableaux introduced by Steingrímsson and Williams (2007). Such a tableau is a Ferrers diagram (a representation of a partition of an integer $n = r_1 + \cdots + r_m$ with parts in decreasing order, in which the i'th row consists of r_i cells; here, rows of length 0 are permitted) with elements from the set $\{0,1\}$ assigned to the cells, under the following restrictions:

1. Each column of the rectangle contains at least one 1;

2. There is no 0 that has a 1 above it in the same column *and* a 1 to its left in the same row.

The length of a tableau is defined to be the sum of the numbers of its rows and columns, and the set of possible tableaux of length n is in one-to-one correspondence with the permutations of n objects. In particular, under the bijection between tableaux and permutations defined by Steingrímsson and Williams (2007, Lemma 5), the lower right boundary, which consists of a sequence of n unit steps down or to the left, has its i-th step down if $I_i^{(n)} = 1$ and to the left if $I_i^{(n)} = 0$. Hence the Theorem 5.1 above, together with (5.3), provides information about the asymptotic shape of the lower right boundary Γ_n of the tableau corresponding to a randomly chosen permutation. Let the upper left corner of the Ferrers diagram represent the origin with the x-axis to the right and the y-axis vertically *downward*, so that the lower right boundary runs from $(n-S_0(1),0)$ to $(0,S_0(1))$: then Γ_n consists of the set $\{(n-S_0(1)-l+S_0(l),S_0(l)),\ 0\le l\le n\}$, linearly interpolated. Hence, $n^{-1}\Gamma_n$ is approximated within $O(n^{-1})$ by the curve

$$\{(\tfrac12[1-t^2]+n^{-1/2}(\widehat{Y}_n(t)-\widehat{Y}_n(1)),\ \tfrac12[1-(1-t)^2]+n^{-1/2}\,\widehat{Y}_n(t)),\ 0\le t\le 1\},$$

where $\widehat{Y}_n$ is as defined in Theorem 5.1.

Corollary 5.2. *As $n\to\infty$, $n^{-1}\Gamma_n$ can be approximated in distribution by*

$$\{(\tfrac12[1-t^2]+n^{-1/2}(\widehat{Z}_n(t)-\widehat{Z}_n(1)),\ \tfrac12[1-(1-t)^2]+n^{-1/2}\,\widehat{Z}_n(t)),\\ 0\le t\le 1\},$$

with an error $o(n^{-1/2})$.

In particular, as can also be seen more directly, $n^{-1}\Gamma_n$ converges in probability to the deterministic curve

$$\{(\tfrac12[1-t^2],\tfrac12[1-(1-t)^2],\ 0\le t\le 1\}=\{(x,y)\in[0,\infty)^2 : x+y=\tfrac34-(x-y)^2\},$$

an arc of a parabola.

Another statistic of interest is the area A_n of such a tableau, which is given by the formula $A_n:=\sum_{i=1}^n I_i\sum_{j=i+1}^n(1-I_j)$, again because of the bijection above. Direct computation yields the expression

$$\begin{aligned}
A_n &= \sum_{i=1}^n S_0(i/n)-\tfrac12 S_0^2(1)-\tfrac12 S_0(1)\\
&= \sum_{i=1}^n\{i(1-i/2n)+\sqrt{n}\,\widehat{Y}_n(i/n)\}-\tfrac12\{(n/2)+\sqrt{n}\,\widehat{Y}_n(1)\}^2\\
&\qquad -\tfrac12\{(n/2)+\sqrt{n}\,\widehat{Y}_n(1)\}\\
&= \frac{5n^2-2}{24}+n^{3/2}\Big\{n^{-1}\sum_{i=1}^n\widehat{Y}_n(i/n)-\tfrac12\widehat{Y}_n(1)\Big\}\\
&\qquad -\tfrac12\{\sqrt{n}\,\widehat{Y}_n(1)+n\widehat{Y}_n(1)^2\}.
\end{aligned}$$

This leads to the following limiting approximation.

Corollary 5.3. *As $n\to\infty$,*

$$n^{-3/2}\left(A_n-\frac{5n^2}{24}\right)\ \to_d\ \mathcal{N}(0,\tfrac{1}{144}).$$

Proof. By the continuous mapping theorem and Slutsky's lemma, it is immediate from Theorem 5.1 that

$$n^{-3/2}\Big(A_n-\frac{5n^2}{24}\Big)\ \to_d\ \int_0^1\widehat{Z}(t)\,dt-\tfrac12\widehat{Z}(1).$$

Now the random variable $\{\int_0^1 \widehat{Z}(t)\,dt - \frac{1}{2}\widehat{Z}(1)\}$ has mean zero and variance

$$\int_0^1 \int_0^1 \widehat{\sigma}(t,u)\,du\,dt - \int_0^1 \widehat{\sigma}(t,1)\,dt + \tfrac{1}{4}\widehat{\sigma}(1,1),$$

with $\widehat{\sigma}$ as in Theorem 5.1, and this gives the value $1/144$. The corollary follows. $\square$

Note also that the number of rows in the permutation tableau $R_n = S_0(1)$; hence Theorem 5.1 implies also, using $\widehat{\sigma}(1,1) = 1/12$,

$$n^{-1/2}\left(R_n - \tfrac{1}{2}n\right) \to_d \mathcal{N}(0, \tfrac{1}{12}).$$

This, however, does not require the functional limit theorem; it follows by the arguments above from Hoeffding's (1951) combinatorial central limit theorem, and it can also be shown in other ways, see Hitczenko and Janson (2009+).

Acknowledgement. We are grateful to the referee for a number of helpful suggestions.

References

[1] R. J. Adler & J. E. Taylor (2007) *Random Fields and Geometry.* Springer, New York. MR2319516

[2] A. D. Barbour (1990) Stein's method for diffusion approximation. *Prob. Theory Rel. Fields* **84**, 297–322. MR1035659

[3] P. Billingsley (1968) *Convergence of Probability Measures*. Wiley, New York. MR0233396

[4] E. Bolthausen (1984) An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrscheinlichkeit verw. Geb.* **66**, 379–386. MR0751577

[5] P. Hitczenko & S. Janson (2009+) Asymptotic normality of statistics on permutation tableaux. Preprint, `arXiv:0904.1222`

[6] W. Hoeffding (1951) A combinatorial central limit theorem. *Ann. Math. Stat.* **22**, 558–566. MR0044058

[7] M. R. Leadbetter, G. Lindgren & H. Rootzén (1983) *Extremes and Related Properties of Random Sequences and Processes.* Springer, New York. MR0691492

[8] G. R. Shorack & J. A. Wellner (1986) *Empirical Processes with Applications to Statistics.* Wiley, New York. MR0838963

[9] E. Steingrímsson and L. K. Williams (2007) Permutation tableaux and permutation patterns. *J. Comb. Theory, Ser. A* **114**, 211–234. MR2293088