Assessing the distance to the Gaussian distribution for random measures via Stein's method

Running title: Distance to Gaussian for random measures

Gesine Reinert
Department of Statistics
University of Oxford
1 South Parks Road
UK - Oxford OX1 3TG

April 20, 2001

Abstract

Stein's approach for assessing distributional distances is developed here for comparing random measures on a locally compact Hausdorff space E with Gaussian random measures. The result is applied to the (nth) empirical process $\xi_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\delta_{X_i} - \mu_i)$ of a family $(X_i)_{i=1,2,\ldots,n}$ of random elements on E, with $\mathcal{L}(X_i) = \mu_i, i = 1, 2, \ldots, n$, where δ_x denotes the Dirac measure at x. Due to Stein's method, the $(X_i)_{i=1,2,\ldots,n}$ may be allowed to be weakly dependent; in particular, we study local dependence. Moreover, we introduce size-biasing for random measures, and we obtain a coupling result. As an example, an immigration-death process is considered.

Key words and phrases. Exchangeable pair, Gaussian distribution, Palm distributions, random measures, size biasing, Stein's method
MSC 1991 classification. 60F05, 60G57, 62G30

Introduction

In 1972, Stein [25] published a very elegant method for proving convergence of random variables $(W_n)_{n\in\mathbf{N}}$ towards a standard normal variable N. It is based on the fact that a random variable W is standard normal distributed if and only if, for all "smooth" functions $f : \mathbf{R} \to \mathbf{R}$, we have (see Stein [26], p. 21, Lemma 1)

$$\mathbb{E}f'(W) = \mathbb{E}Wf(W).$$

For every smooth function g, there exists a function f that solves the so-called Stein equation

$$g(x) - \mathbb{E}g(N) = f'(x) - xf(x), \quad x \in \mathbf{R}. \tag{1}$$

Replacing x by W_n and taking expectations, (1) yields

$$\mathbb{E}g(W_n) - \mathbb{E}g(N) = \mathbb{E}f'(W_n) - \mathbb{E}W_n f(W_n). \tag{2}$$

Thus estimating $\mathbb{E}g(W_n) - \mathbb{E}g(N)$ is equivalent to estimating $\mathbb{E}f'(W_n) - \mathbb{E}W_n f(W_n)$ for f being a solution of (1), and an estimate of this quantity gives a bound on the distance to the normal distribution. Taking the special structure of W_n into account, the estimation of the right hand side of (2) often is considerably easier than the estimation of the left hand side of (2). Stein's method has been particularly useful for dependent observations; typically, local dependence as well as weak global dependence can be covered by this approach. The method was generalized for many other distributions; for an overview see Reinert [20], e.g..

Barbour [3] rewrote the Stein equation (1) as

$$g(x) - \mathbb{E}g(N) \quad = \quad \mathcal{A}f(x), \quad x \in \mathbf{R},$$

where $\mathcal{A}f(x) = f''(x) - xf'(x)$ is the generator of an Ornstein-Uhlenbeck process, with the standard normal distribution as equilibrium distribution. Hence, if $(T_t)_{t\geq 0}$ is the transition semigroup of the Ornstein-Uhlenbeck process, then, (see Barbour [3],) formally,

$$g(x) - \mathbb{E}g(N) = -\mathcal{A}\left(\int_0^\infty T_u g \, du\right)$$

would be a solution of the Stein equation if it the right-hand side exists. Thus this generator approach provides us again with solutions of the Stein equation. Moreover, the generator approach can be extended to assessing the distance to the Gaussian distribution for more general random objects. In the path-valued case, Barbour [3] developed Stein's method for diffusion approximations, and obtained a bound on the rate of convergence. In the present paper, we use a similar approach for Gaussian measure-valued random elements.

Measure-valued random elements, in particular empirical measures and empirical processes, have become a powerful tool in statistics, see, for example, van der Geer [9], van der Vaart and Wellner [28]. Hence it is desirable to extend the Stein approach to these objects. For empirical measures this has been carried out in Reinert [21]; there also, as an example, an immigration-death process is considered, illustrating the connection with point processes.

Moreover, usually in the literature almost-sure convergence results are derived. Yet, in applications, only a finite number of observations is available. Here we focus our attention on the distance to a Gaussian random measure for a random measure that depends only on a finite number of observations; the distance is related to the weak topology on the space of measures. The bounds that we obtain will be valid for any number n of observations, in contrast to almost-sure convergence results.

The setup is that E is a locally compact space with a countable basis, and η_n is a random measure on E. A general theorem is derived that bounds the distance, in law, of η_n to a Gaussian random measure ζ. If $(X, \mathcal{B})$ is a measurable space, then ζ is Gaussian with respect to a linear space $\mathcal{C} \subset \{g : X \to \mathbf{R} \text{ measurable}\}$, if for all $g \in \mathcal{C}$, the projection $\mathbb{P}_\zeta \circ g^{-1}$ is a Gaussian measure on $\mathbf{R}$ (see Vakhania, Tarieladze and Chobanyan [29], p. 210). In our results, the linear space $\mathcal{C}$ will be a convergence-determining set for vague convergence in the space of measures on E.

In particular, we are interested in the (nth) empirical process

$$\xi_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\delta_{X_i} - \mu_i)$$

of a family $(X_i)_{i\in\mathbf{N}}$ of random elements on E, with laws $\mathcal{L}(X_i) = \mu_i, i = 1, 2, \ldots n$. Having derived a general bound for random measures, we will conclude a bound on the distance to a Gaussian random measure ζ for the empirical process of locally dependent random elements which may also be weakly globally dependent. In the case of m-dependent random elements, the bound simplifies to $20\frac{m^2}{\sqrt{n}}$. The independent case is well studied (see, for example, Dembo and Zajic [6], Einmahl, Mason and Ruymgaart [7], and Ledoux and Talagrand

[16]), but few dependent structures have been investigated (see Ginè [10], Arcones and Yu [1], Shao and Yu [24], and Philipp [17]). Götze [11] gives a general expansion which might break down in non-i.i.d. settings. Those results involve mostly mixing conditions or associate variates (Shao and Yu [24]), whereas our results allow for dependence structures that include local dependence and weak global dependence.

To treat the empirical process of weakly globally dependent variates, in connection with Stein's method coupling techniques have shown to be very useful; see, for example, Bolthausen and Götze [5], Goldstein and Rinott [14]. Here, as coupling techniques, we employ exchangeable pairs as well as size-biasing for random measures. For a nonnegative random variable W with positive mean, W^* is said to have the W size biased distribution, if

$$\mathbb{E}Wg(W) = \mathbb{E}W\mathbb{E}g(W^*)$$

for all functions g for which the expectations exist. This coupling has been applied by Baldi, Rinott and Stein [2] and by Stein [26], e.g., to prove normal approximations, and in Barbour, Holst and Janson [4], e.g., to prove Poisson approximations. Goldstein and Rinott [14] generalized this approach for nonnegative random vectors and obtained multivariate normal approximations. For random measures, size biasing can be defined similarly, but now in "direction" of a function; this is connected to Palm distributions. With conditions phrased in terms of this coupling, another bound to a Gaussian measure is derived.

As an example, we put the general theorem to use in case of complex dependence. We consider an immigration-death process as in Reinert [21] with population size n, arrival times $(A_i^n)_{i=1,\ldots,n}$ and life spans $(Z_i)_{i\in\mathbf{N}}$; the life spans are assumed to be i. i. d. and independent of the arrival times, whereas dependence between the arrival times is allowed. With $X_i = (A_i^n, A_i^n + Z_i)$, a bound on the distance to a Gaussian random measure with appropriate covariance structure is derived. This illustrates how the general framework can be applied to stochastic processes.

The approach is related to multivariate central limit theorems. Multivariate central limit theorems using Stein's method have been obtained already by, for example, Bolthausen and Götze [5], Götze [12], Goldstein and Rinott [14], Rinott and Rotar [23]. Indeed, in case $E = \mathbf{R}^k$, the results from Götze [12] for general multivariate statistics based on i.i.d. observations could be transferred directly. The advantage of the approach presented here is, firstly, that it allows for dependence between the observations. Moreover it is valid for general random measures, not only for empirical measures, and it is valid on general locally compact Hausdorff spaces E, not only $\mathbf{R}^k$. Thus, for example, it could be applied to weighted empirical processes as occur in bootstrapping, see Holmes and Reinert [15]. Furthermore, it fits into the framework of empirical processes, which have emerged as an important tool in probability and statistics, and thus provides a unified treatment of these empirical processes.

The paper is organized as follows. As empirical processes are random elements taking values in the space $M^f(E)$ of finite Radon measures on E, to formulate a result on distances for empirical processes, some topological considerations about weak convergence in that space are needed. This is addressed in Section 1. Section 2 contains a general theorem assessing the distance to Gaussian for random measures, and the derivation of this theorem via Stein's method. The general result is applied to empirical processes of n locally strongly dependent and globally weakly dependent random elements. In Section 3, couplings are employed. Firstly a result is given using exchangeable pairs, in the spirit of Stein [26]. Then the notion of size biasing for random measures is introduced, and, related to Goldstein and Rinott [14], we assess the distance to the normal distribution for empirical measures in terms of size-bias couplings. Lastly, in Section 4, as an example, we treat the above immigration-death process. In the presentation, proofs that would disturb the line of argumentation are postponed until the appendix.

Parts of this work are also presented in Reinert [22].

1 Topological structure on the space of measures

Throughout, let E be a locally compact Hausdorff space with a countable basis. Thus E is metrizable as a Polish space; let d be a metric on E that makes E Polish. If $E = \mathbf{R}$, d shall denote the absolute value. Let $C_c(E) = \{f : E \to \mathbf{R}$ continuous with compact support$\}$, denote by $\mathcal{E} = \mathcal{B}(E)$ the Borel-σ-field of E, and let $M(E)$ be the space of all (real-valued, not necessarily positive) Radon measures on E, equipped with the vague topology: For $(\nu_n)_{n\in\mathbf{N}}, \nu \in M(E)$,

$$\nu_n \stackrel{v}{\Rightarrow} \nu \text{ iff for all } f \in C_c(E), \langle \nu_n, f\rangle \to \langle \nu, f\rangle,$$

where we use the notation $\langle \nu, f\rangle = \int f d\nu$. If $(\nu_n)_{n\in\mathbf{N}}, \nu \in M_1(\mathcal{U})$ are laws of finite random measures, vague convergence implies weak convergence. Note that, as in Rachev [18], p. 209, vague convergence could also be described via Lipschitz functions.

For $\nu \in M(E)$, let $\|\mu\| = |\mu(E)|$, let $M^f(E) = \{\mu \in M(E) : \|\mu\| < \infty\}$ be the space of finite Radon measures, and let $M_1(E) = \{\mu \in M(E) : \mu \text{ positive}, \|\mu\| \le 1\}$ be the space of all positive Radon measures with total mass smaller or equal to 1. As E has a countable basis, $M_1(E)$ is Polish with respect to the vague topology. Moreover, $M^f(E)$ is locally compact with a countable basis. As $M^f(E)$ is again Polish, we can define vague convergence in $M_1(M^f(E))$, which will be the space of the laws of the random measures we consider. Similarly as in Reinert [21], we define the following classes of functions.

Let $\mathcal{C}$ be a convergence-determining class for vague convergence on $M^f(E)$. We think of $\mathcal{C} = C_c(E)$ or $\mathcal{C} = {C_c}^+(E)$, the latter being the space of all nonnegative functions in $C_c(E)$. If $E = \mathbf{R}^k$ we may also choose $\mathcal{C} = C_b^\infty(\mathbf{R}^k)$. Denote by $C_0(E)$ the space of all continuous real-valued functions f on E that vanish at infinity (for all $\epsilon > 0$, there is a compact set K outside of which $|f(x)| < \epsilon$).

Let $\mathcal{C} \subset C_0(E)$ be a convergence-determining class for vague convergence in $M^f(E)$, and let $\mathcal{C}$ be closed under multiplication with scalars. Put

$$\mathcal{C}^* = \{\phi \in \mathcal{C} : \|\phi\| \le 1\},$$

if $\mathcal{C}$ is a class of bounded functions, and

$$\mathcal{C}^* = \{\phi \in \mathcal{C} : \|\phi\| \le 1, \|\phi'\| \le 1, \|\phi''\| \le 1\}$$

if $E = \mathbf{R}^k$ and $\mathcal{C} = C_b^\infty(\mathbf{R}^k)$. Here, $\| \phi \| = \sup_x |\phi(x)|$, $\| \phi' \| = \sum_{j=1}^k \| \phi_{(j)} \|, \| \phi'' \| = \sum_{j,l=1}^k \| \phi_{(j,l)} \|$, and $\| \phi''' \| = \sum_{j,l,m=1}^k \| \phi_{(j,l,m)} \|$, where $\phi_{(j)}$ is the partial derivative of ϕ in direction x_j, and similarly $\phi_{(i,j)}, \phi_{(i,j,k)}$ denote higher partial derivatives. Define

$$\begin{aligned} \mathcal{F}_{\mathcal{C}} \quad = \{ \quad & F \in C_{\mathcal{C}}(M^f(E)) : F \text{ has the form} \\ & F(\mu) = f(\langle \mu, \phi_1 \rangle, \ldots, \langle \mu, \phi_m \rangle) \\ & \text{ for an } m \in \mathbf{N}, f \in C_b^\infty(\mathbf{R}^m) \text{ with } \|f'\| \le 1, \|f''\| \le 1, \|f'''\| \le 1; \\ & \phi_1, \ldots, \phi_m \in \mathcal{C}^* \}. \end{aligned} \tag{3}$$

Using the Stone-Weierstrass Theorem the following proposition is easy to show.

Proposition 1 *The class $\mathcal{F}_{\mathcal{C}}$ is convergence-determining for vague convergence in the space $M_1(M^f(E))$.*

In the following, we will also use the abbreviation

$$\underline{\langle \mu, \phi \rangle} = (\langle \mu, \phi_1 \rangle, \ldots, \langle \mu, \phi_m \rangle).$$

Remark 2 *In view of the above convergence-determining class, the approach is much reminiscent of convergence of finite-dimensional distributions. This raises the question of tightness conditions. Here, as the class $\mathcal{F}_{\mathcal{C}}$ is convergence-determining, no tightness condition for weak convergence in the space of measures $M_1(M^f(E))$ is needed. However, if one wanted to translate the results into convergence in $D[0,1]$, equipped with the Skorohod topology, say, then more work would be necessary. The goal of this paper is to present the general framework, so this special case is not treated.*

Note also that the class $\mathcal{F}_{\mathcal{C}}$ is a class of smooth test functions. The treatment of nonsmooth test functions is a technically involved issue; for $\mathbf{R}^k$, this can be found in Götze [12], in Rinott and Rotar [23], and in Raij [19], for example.

2 The distance to the Gaussian distribution

2.1 Stein's method for the distance to Gaussianity of random measures

Throughout this section, we assume that $b : \mathcal{C}^2 \to \mathbf{R}$ is an operator such that, for any $m \in \mathbf{N}$ and for all $\phi_1, \dots, \phi_m \in \mathcal{C}$,

$$B = B(\phi_1, \dots, \phi_m) = (b(\phi_i, \phi_j))_{i,j=1,\dots,m}$$

is a symmetric, positive definite matrix. For $F \in C_{\mathcal{C}}(M^f(E))$ with representation (3), we define the generator

$$\mathcal{A}F(\nu) = -\sum_{j=1}^{m} f_{(j)}(\underline{\langle \nu, \phi \rangle})\langle \nu, \phi_j \rangle + \sum_{j,k=1}^{m} f_{(j,k)}(\underline{\langle \nu, \phi \rangle}) b(\phi_j, \phi_k). \tag{4}$$

We say that $\mathcal{A}$ is the generator associated with the operator b, or, slightly abusing notation, associated with the matrix B. For a heuristic derivation of this generator, follow the lines of the heuristics given in Reinert [21] for the law of large numbers. It follows from the Stein equation for the multivariate normal distribution (see Götze [12], eq. (1.7)), that, if ζ is a random measure taking values in $M^f(E)$ almost surely such that, for all $m \in \mathbf{N}, \phi_1, \dots, \phi_m \in \mathcal{C}$,

$$\mathcal{L}(\langle \zeta, \phi_1 \rangle, \dots, \langle \zeta, \phi_m \rangle) = \mathcal{MVN}_m(0, B), \tag{5}$$

and if $\mathcal{A}$ is the generator associated with B, then $\mathcal{L}(\zeta)$ is stationary for $\mathcal{A}$. Thus, for $G \in C_{\mathcal{C}}(M^f(E))$ of the form

$$G(\nu) = g(\underline{\langle \nu, \psi \rangle}), \tag{6}$$

the Stein equation corresponding to the Gaussian random measure ζ with (5) is

$$g(\underline{\langle \nu, \psi \rangle}) - \mathbb{E}g(\underline{\langle \zeta, \phi \rangle}) \quad = \quad -\sum_{j=1}^{m} \mathbb{E}f_{(j)}(\underline{\langle \nu, \phi \rangle})\langle \nu, \phi_j \rangle + \sum_{j,k=1}^{m} \mathbb{E}[f_{(j,k)}(\underline{\langle \nu, \phi \rangle})] b(\phi_j, \phi_k). \tag{7}$$

This equation can be solved using the semigroup technique as in Barbour [3] (the proof is in the appendix).

Lemma 3 *Let ζ be as in (5). Then, for each $G = g(\underline{\langle \nu, \psi \rangle}) \in C_{\mathcal{C}}(M^f(E))$ there is a function $F \in C_{\mathcal{C}}(M^f(E))$ that solves (7) for G. Moreover, there is a function $f \in C_b^\infty(\mathbf{R}^m)$ such that $F(\nu) = f(\underline{\langle \nu, \psi \rangle})$, and $\|f^{(k)}\| \le \|g^{(k)}\|, k \in \mathbf{N}$.*

The generator approach now immediately yields the following general theorem.

Theorem 4 *Let η_n be a random measure taking values in $M^f(E)$ almost surely. Let ζ be a random measure taking values in $M^f(E)$ almost surely such that, for all $m \in \mathbf{N}, \phi_1, \dots, \phi_m \in \mathcal{C}$, (5) is satisfied with B. Let $\mathcal{A}$ be the generator associated with B. Then for each $G \in C_{\mathcal{C}}(M^f(E))$ of the form (6) we have*

$$|\mathbb{E}G(\eta_n) - \mathbb{E}G(\zeta_n)| \quad \le \quad |\mathbb{E}\mathcal{A}F_G(\eta_n)|\,,$$

where F_G, given in (18), solves (7) for G.

Remark 5 *1. If $(\eta_n)_{n\in\mathbf{N}}$ is a family of random measures taking values in $M^f(E)$ almost surely, and if, for all $G \in C_{\mathcal{C}}(M^f(E))$ we have $|\mathbb{E}\mathcal{A}F_G(\eta_n)| \to 0$ as $n \to \infty$, then it follows that $\mathcal{L}(\eta_n) \overset{w}{\Rightarrow} \mathcal{L}(\zeta)$. Thus Theorem 4 is also a Gaussian approximation theorem.*

2. Theorem 4 assumes the existence of a Gaussian random measure ζ that is finite almost surely. In general, this existence is not guaranteed. However, if E is compact, then, with $\mathcal{C} = C_c(E)$ or $\mathcal{C} = C_c^+(E)$, we have

$$\mathcal{L}(\langle\zeta, 1_E\rangle) = \mathcal{N}(0, b(1_E, 1_E)).$$

Hence, $|\zeta(E)| < \infty$ almost surely, and thus ζ takes values in $M^f(E)$ almost surely.

3. Associated to $\mathcal{F}_{\mathcal{C}}$ one could also introduce a Zolotarev type metric on the space of measures by

$$\zeta_{\mathcal{F}_{\mathcal{C}}}(\nu, \eta) = \sup_{f\in\mathcal{F}_{\mathcal{C}}} \big|\langle\nu, f\rangle - \langle\eta, f\rangle\big|.$$

From Proposition 1 it follows that $\zeta_{\mathcal{F}_{\mathcal{C}}}$ is a metric on the space $M_1(M^f(E))$, and for all $(\nu_N)_{n\in\mathbf{N}}, \nu \in M_1(M^f(E))$, if $\zeta_{\mathcal{F}_{\mathcal{C}}}(\nu_n, \nu) \to 0 \ \ (n \to \infty)$, then $\nu_n \overset{v}{\Rightarrow} \nu \ \ (n \to \infty)$. Theorem 4 and later results could be formulated in terms of this metric.

2.2 Application to empirical processes

In particular we are interested in empirical processes. Throughout the remainder of this paper, $(X_i)_{i\in\mathbf{N}}$ is family of random elements on E with laws $\mathcal{L}(X_i) = \mu_i$, and $\xi_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n(\delta_{X_i} - \mu_i)$ denotes the corresponding (nth) empirical process. For all $m \in \mathbf{N}$ and $\phi_1, \ldots, \phi_m \in \mathcal{C}$, let

$$B_n = \big(b_n(\phi_j, \phi_k)\big)_{j,k=1,\ldots,m} = \left(\frac{1}{n}\sum_{i,l=1}^n \mathrm{Cov}\,(\phi_j(X_i), \phi_k(X_l))\right)_{j,k=1,\ldots,m} \tag{8}$$

be the corresponding covariance matrix, and let $\mathcal{A}_n$ be the generator associated with B_n.

To illustrate the use of Proposition 4, we consider the case of strong local dependence and weak global dependence .

Corollary 6 *For all $i, n \in \mathbf{N}$ let $\Gamma_n(i) \subset \{1, \ldots, n\}$ be an index set. Let $\gamma_n = \max_{i=1,\ldots,n}|\Gamma_n(i)|$. Let ζ_n be a random measure taking values in $M^f(E)$ almost surely such that, for all $m \in \mathbf{N}, \phi_1, \ldots, \phi_m \in \mathcal{C}$, (5) is satisfied with B_n for B, where*

$$B_n = \big(b_n(\phi_j, \phi_k)\big)_{j,k=1,\ldots,m} = \left(\frac{1}{n}\sum_{i=1}^n\sum_{j\in\Gamma_n(i)} \mathrm{Cov}\,(\phi_j(X_i), \phi_k(X_l))\right)_{j,k=1,\ldots,m}$$

Then, for each $G \in C_{\mathcal{C}}(M^f(E))$ we have

$$|\mathbb{E}G(\xi_n) - \mathbb{E}G(\zeta_n)| \le \frac{20}{\sqrt{n}}\gamma_n^2 + S_1 + S_2,$$

where

$$S_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \max_{\phi \in \mathcal{C}} \mathbb{E} | E\{(\phi(X_i) - \mathbb{E}\phi(X_i)) | X_j, j \notin \Gamma_n(i)\}|$$

and

$$S_2 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j \in \Gamma_n(i)} \max_{\phi, \psi \in \mathcal{C}} \mathbb{E} | E\{(\phi(X_i) - \mathbb{E}\phi(X_i))(\psi(X_j) - \mathbb{E}\psi(X_j)) | X_\ell, \ell \notin \Gamma_n(i) \cup \Gamma_n(j)\}|.$$

Remark 7 *1. In Corollary 6, the index set $\Gamma_n(i)$ is thought of as the set of all indices j such that X_j and X_i are strongly dependent.*

2. In the case of local dependence only, such as m-dependence, the terms S_1 and S_2 vanish.

Proof: Let $F(\nu) = f(\underline{\langle \nu, \phi \rangle})$ be the solution for G of the Stein equation with generator $\mathcal{A}_n$ associated with B_n. Put $\xi_n^i = \frac{1}{\sqrt{n}} \sum_{j \notin \Gamma_n(i)} (\delta_{X_j} - \mu_j)$ and $\xi_n^{i,l} = \frac{1}{\sqrt{n}} \sum_{j \notin \{\Gamma_n(i) \cup \Gamma_n(l)\}} (\delta_{X_j} - \mu_j)$. By Taylor expansion,

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbb{E} f_{(j)} (\underline{\langle \xi_n, \phi \rangle}) \langle \delta_{X_i} - \mu_i, \phi_j \rangle \\ = \; & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j,k=1}^{m} \mathbb{E} f_{(j,k)} \left(\underline{\langle \xi_n^i, \phi \rangle} \right) \langle \xi_n - \xi_n^i, \phi_k \rangle \langle \delta_{X_i} - \mu_i, \phi_j \rangle + R_1 + R_2, \end{aligned}$$

where

$$\begin{aligned} |R_1| &\le \frac{1}{2} n^{-\frac{3}{2}} \sum_{i=1}^{n} \sum_{s,t \in \Gamma_n(i)} \sum_{j,k,r=1}^{m} \| f_{(j,k,r)} \| \, E | \langle \delta_{X_s} - \mu_s, \phi_r \rangle \langle \delta_{X_t} - \mu_t, \phi_k \rangle \langle \delta_{X_i} - \mu_i, \phi_j \rangle | \\ &\le \frac{4}{\sqrt{n}} \| f''' \| \max_{1 \le i \le m} \| \phi_i \|^3 \gamma_n^2 \end{aligned}$$

and

$$\begin{aligned} |R_2| &= \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbb{E} f_{(j)} \left(\underline{\langle \xi_n^i, \phi \rangle} \right) \langle \delta_{X_i} - \mu_i, \phi_j \rangle \right| \\ &= \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbb{E} f_{(j)} \left(\underline{\langle \xi_n^i, \phi \rangle} \right) E\{(\phi_j(X_i) - \mathbb{E}\phi_j(X_i)) | X_j, j \notin \Gamma_n(i)\} \right| \\ &\le S_1. \end{aligned}$$

Moreover

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j,k=1}^{m} \mathbb{E} f_{(j,k)} \left(\underline{\langle \xi_n^i, \phi \rangle} \right) \langle \xi_n - \xi_n^i, \phi_k \rangle \langle \delta_{X_i} - \mu_i, \phi_j \rangle \\ = \; & \frac{1}{n} \sum_{i=1}^{n} \sum_{l \in \Gamma_n(i)} \sum_{j,k=1}^{m} \mathbb{E} f_{(j,k)} \left(\underline{\langle \xi_n^{i,l}, \phi \rangle} \right) \langle \delta_{X_l} - \mu_l, \phi_k \rangle \langle \delta_{X_i} - \mu_i, \phi_j \rangle + R_3 \\ = \; & \frac{1}{n} \sum_{i=1}^{n} \sum_{l \in \Gamma_n(i)} \sum_{j,k=1}^{m} \mathbb{E} f_{(j,k)} \left(\underline{\langle \xi_n^{i,l}, \phi \rangle} \right) \mathrm{Cov}\left(\phi_j(X_i), \phi_k(X_l) \right) + R_3 + R_4, \end{aligned}$$

where

$$|R_3| \leq \frac{8}{\sqrt{n}} \|f'''\| \max_{1\leq i\leq m} \|\phi_i\|^3 \gamma_n^2,$$

and

$$\begin{aligned}
&|R_4| \\
&= \Bigg| \frac{1}{n} \sum_{i=1}^{n} \sum_{l\in\Gamma_n(i)} \sum_{j,k=1}^{m} \mathbb{E} f_{(j,k)} \left(\underline{\langle \xi_n^{i,l}, \phi \rangle} \right) \{ \mathbb{E}((\phi_k(X_l) - \mathbb{E}\phi_k(X_l))(\phi_j(X_i) - \mathbb{E}\phi_j(X_i)) | X_t, t \notin \Gamma_n(i) \cup \Gamma_n(l)) \\
&\quad - \mathrm{Cov}\,(\phi_j(X_i), \phi_k(X_l)) \} \Big| \\
&\leq S_2.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\mathbb{E}\mathcal{A}F(\xi_n) \\
&= -\frac{1}{n} \sum_{i=1}^{n} \sum_{l\in\Gamma_n(i)} \sum_{j,k=1}^{m} \mathbb{E} f_{(j,k)} \left(\underline{\langle \xi_n^{i,l}, \phi \rangle} \right) \mathrm{Cov}\,(\phi_j(X_i), \phi_k(X_l)) \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} \sum_{l\in\Gamma_n(i)} \sum_{j,k=1}^{m} \mathbb{E} f_{(j,k)} \left(\underline{\langle \xi_n, \phi \rangle} \right) \mathrm{Cov}\,(\phi_j(X_i), \phi_k(X_l)) + R_1 + R_2 + R_3 + R_4.
\end{aligned}$$

By similar calculations as above, this results in

$$|\mathbb{E}\mathcal{A}F(\xi_n)| \quad \leq \quad \frac{20}{\sqrt{n}} \|f'''\| \max_{1\leq i\leq m} \|\phi_i\|^3 \gamma_n^2 + S_1 + S_2.$$

Using Lemma 3, Proposition 4 and that, in $\mathcal{F}_{\mathcal{C}}$, we have $\|f'''\| \leq 1$ and $\|\phi_i\| \leq 1$, the assertion follows. □

3 Coupling approaches

To treat global weak dependence with Stein's method, coupling approaches have proved very useful.

3.1 Exchangeable pair coupling

Exchangeable pairs can be defined analogously to the definition for real-valued random elements. Let η be a random measure with expectation measure $\mathbb{E}\eta = \mu$, and let η' be another random measure with the same distribution as η. We say that (η, η') is an exchangeable pair if for all sets $B, B' \in \mathcal{E}$,

$$\mathbb{P}(\eta \in B, \eta' \in B') = \mathbb{P}(\eta \in B', \eta' \in B).$$

Similarly as in Stein [26], where the real-valued case is treated, we define Condition (9) by requiring that there is a $0 < \lambda < 1$ such that for all $\phi \in \mathcal{C}$

$$\mathbb{E}(\langle \eta' - \mu, \phi \rangle | \langle \eta - \mu, \phi \rangle) = (1 - \lambda) \langle \eta - \mu, \phi \rangle. \tag{9}$$

From (9) it follows immediately that for all functions f and ϕ,

$$\mathbb{E}f(\eta-\mu)\langle\eta-\mu,\phi\rangle = \frac{1}{2\lambda}\mathbb{E}(f(\eta'-\mu)-f(\eta-\mu))\langle\eta'-\eta,\phi\rangle. \tag{10}$$

In particular,

$$\mathbb{E}\langle\eta'-\eta,\phi\rangle\langle\eta'-\eta,\psi\rangle \quad = \quad 2\lambda\mathbb{E}\langle\eta-\mu,\phi\rangle\langle\eta-\mu,\psi\rangle. \tag{11}$$

Similarly as in Stein [26], we obtain the following theorem.

Theorem 8 *Let $(X_i)_{i\in\mathbf{N}}$ be a family of random elements on E with laws $\mathcal{L}(X_i)=\mu_i$, and let ξ_n be the corresponding empirical process. Let ζ_n be a random measure taking values in $M^f(E)$ such that, for all $m\in\mathbf{N},\phi_1,\dots,\phi_m\in\mathcal{C}$, (5) is satisfied with B_n for B, where B_n is given in (8), and let $\mathcal{A}_n$ be the generator associated with B_n. Let $\eta_n=\frac{1}{\sqrt{n}}\sum_{i=1}^n\delta_{X_i}$, so that $\xi_n=\eta-\mathbb{E}\eta$. Let η_n,η_n' be an exchangeable pair satisfying (9). Then, for each $G\in C_{\mathcal{C}}(M^f(E))$,*

$$\begin{aligned}|\mathbb{E}G(\xi_n)-\mathbb{E}G(\zeta_n)| \quad \leq \quad & \frac{1}{2\lambda}\left\{\sup_{\phi_1,\phi_2,\phi_3\in\mathcal{C}}\mathbb{E}|\langle\eta_n'-\eta,\phi_1\rangle\langle\eta_n'-\eta,\phi_2\rangle\langle\eta_n'-\eta,\phi_3\rangle|\right.\\ &\left.+\sup_{\phi_1,\phi_2\in\mathcal{C}}\{\,Var(\mathbb{E}(\langle\eta_n'-\eta,\phi_1\rangle\langle\eta_n'-\eta,\phi_2\rangle|\xi_n)\}^{\frac{1}{2}}\right\}.\end{aligned}$$

PROOF: Let $F(\nu)=f(\underline{\langle\nu,\phi\rangle})$ be the solution for G of the Stein equation with generator $\mathcal{A}_n$ associated with B_n. Put $\xi_n'=\eta_n'-\mu$. From (10) we have

$$\begin{aligned}\sum_{j=1}^m\mathbb{E}f_{(j)}(\underline{\langle\xi_n,\phi\rangle})\langle\eta-\mu,\phi_j\rangle \quad &= \quad \frac{1}{2\lambda}\sum_{j=1}^m\mathbb{E}(f_{(j)}(\underline{\langle\xi_n,\phi\rangle})-f_{(j)}(\underline{\langle\xi_n',\phi\rangle})\langle\eta_n-\eta_n',\phi_j\rangle\\ &= \quad \frac{1}{2\lambda}\sum_{j,k=1}^m\mathbb{E}f_{(j,k)}(\underline{\langle\xi_n,\phi\rangle})\langle\eta_n-\eta_n',\phi_j\rangle\langle\eta_n-\eta_n',\phi_k\rangle+R_1\\ &= \quad \frac{1}{2\lambda}\sum_{j,k=1}^m\mathbb{E}f_{(j,k)}(\underline{\langle\xi_n,\phi\rangle})\mathbb{E}\{\langle\eta_n-\eta_n',\phi_j\rangle\langle\eta_n-\eta_n',\phi_k\rangle|\xi_n\}+R_1\\ &= \quad \frac{1}{2\lambda}\sum_{j,k=1}^m\mathbb{E}f_{(j,k)}(\underline{\langle\xi_n,\phi\rangle})b_n(\phi_j,\phi_k)+R_1+R_2,\end{aligned}$$

where

$$|R_1| \quad \leq \quad \frac{1}{2\lambda}\sup_{\phi_1,\phi_2,\phi_3\in\mathcal{C}}\mathbb{E}|\langle\eta_n'-\eta,\phi_1\rangle\langle\eta_n'-\eta,\phi_2\rangle\langle\eta_n'-\eta,\phi_3\rangle|$$

and

$$|R_2| \quad \leq \quad \sup_{\phi_1,\phi_2\in\mathcal{C}}\mathbb{E}\left|\frac{1}{2\lambda}\mathbb{E}f_{(j,k)}(\underline{\langle\xi_n,\phi\rangle})\mathbb{E}\{\langle\eta_n-\eta_n',\phi_1\rangle\langle\eta_n-\eta_n',\phi_2\rangle|\xi_n\}-\frac{1}{n}\sum_{i,\ell=1}^n\mathrm{Cov}\,(\phi_1(X_i),\phi_2(X_\ell))\right|.$$

From (11) it follows that

$$\begin{aligned}\mathbb{E}(\mathbb{E}f_{(j,k)}(\underline{\langle\xi_n,\phi\rangle})\mathbb{E}\{\langle\eta_n-\eta'_n,\phi_1\rangle\langle\eta_n-\eta'_n,\phi_2\rangle|\xi_n\}) &= 2\lambda\mathbb{E}\langle\xi_n,\phi_1\rangle\langle\xi_n,\phi_2\rangle \\ &= 2\lambda\frac{1}{n}\sum_{i,\ell=1}^n \mathrm{Cov}\,(\phi_1(X_i),\phi_2(X_\ell)).\end{aligned}$$

Using the Cauchy-Schwarz inequality we obtain

$$|R_2| \;\le\; \frac{1}{2\lambda}\sup_{\phi_1,\phi_2\in\mathcal{C}}\{\mathrm{Var}(\mathbb{E}(\langle\eta'_n-\eta,\phi_1\rangle\langle\eta'_n-\eta,\phi_2\rangle|\xi_n)\}^{\frac{1}{2}}.$$

□

For example, if $\eta_n=\frac{1}{\sqrt{n}}\sum_{i=1}^n\delta_{X_i}$, with $X_1,\dots,X_n$ being i.i.d, then it is easy to check that picking an index I uniformly form $\{1,\dots,n\}$, replacing X_I by an independent copy X'_I and putting $\eta'_n=\eta_n-\frac{1}{\sqrt{n}}\delta_{X_I}+\frac{1}{\sqrt{n}}\delta_{X'_I}$ gives an exchangeable pair, and Condition (9) is satisfied with $\lambda=\frac{1}{n}$. Moreover,

$$|R_1| \;\le\; \frac{n}{2}n^{-\frac{3}{2}}\sup_{\phi_1,\phi_2,\phi_3\in\mathcal{C}}\mathbb{E}|(\phi_1(X)-\mathbb{E}\phi_1(X))(\phi_2(X)-\mathbb{E}\phi_2(X))(\phi_3(X)-\mathbb{E}\phi_3(X))|\le\frac{4}{\sqrt{n}},$$

and

$$|R_2| \;\le\; \frac{n}{2}n^{-\frac{3}{2}}\sup_{\phi_1,\phi_2\in\mathcal{C}}\{\mathrm{Var}((\phi_1(X)-\mathbb{E}\phi_1(X))(\phi_2(X)-\mathbb{E}\phi_2(X))\}^{\frac{1}{2}}\le\frac{2}{\sqrt{n}},$$

and Theorem 8 yields a bound of $\frac{6}{\sqrt{n}}$. This compares favourably to the bound $\frac{20}{\sqrt{n}}$ from Theorem 6. However, in neither case optimality of the bounds is claimed.

3.2 Size biasing and Palm distributions for random measures

Size-biasing for random variables is a very useful tool for Poisson and for normal approximations, see, e.g., Baldi, Rinott and Stein [2], Barbour, Holst and Janson [4], Stein [26], and Goldstein and Rinott [14]. For a nonnegative random variable W with positive mean, W^* is said to have the W size biased distribution if

$$\mathbb{E}Wg(W)=\mathbb{E}W\,\mathbb{E}g(W^*)$$

for all functions g for which the expectations exist. Goldstein and Rinott [14] generalized this approach for general nonnegative random vectors $\mathbf{W}$. Here we introduce a generalization for random measures.

Definition 9 *Let η be a finite nonnegative random measure on E, let $\mathcal{C}$ be a convergence-determining class on E, let $\phi\in\mathcal{C}$ with $\langle\mathbb{E}\eta,\phi\rangle>0$ and $\phi\ge 0$. We say that η^ϕ has the η size biased distribution in direction ϕ if, for all f for which the expectations exist,*

$$\mathbb{E}f(\eta)\langle\eta,\phi\rangle=\langle\mathbb{E}\eta,\phi\rangle\mathbb{E}f(\eta^\phi).$$

The size bias distribution in direction ϕ is related to Palm distributions for random measures as follows. Let $\mathcal{M}$ be the smallest $\sigma-$algebra making all mappings $\eta \to \eta(B), B \in \mathcal{E}$, measureable. For a random measure η, the Campbell measure on $\mathcal{E} \times \mathcal{M}$ is given by (see Stoyan *et al.*, p.229) $\mathcal{C}(B \times Y) = \mathbb{E}\eta(B)\mathbf{1}_Y(\eta)$ for $B \in \mathcal{E}, Y \in \mathcal{M}$. If the intensity measure Λ given by $\Lambda(B) = \mathbb{E}\eta(B)$ is locally finite, then there is a family $(P_x)_x$ of distributions on $M^f(E)$ such that

$$\mathcal{C}(B \times Y) \quad = \quad \int_B P_x(Y)\Lambda(dx); \tag{12}$$

the P_x are called *Palm distributions.* Now, for $f(\eta) = \mathbf{1}_Y(\eta)$ and $B = \{a\}$ being a singleton, we have $\mathbb{E}f(\eta)\eta(\{a\}) = \mathbb{E}\eta(\{a\})\mathbb{E}f(\eta^{\mathbf{1}_{\{a\}}})$, recovering (12) and the Palm distributions. For a general bounded set $B \in \mathcal{E}$, if there are only a finite number of points in B with nonzero mass, then the r.h.s. of (12) gives

$$\int_B P_x(Y)\Lambda(dx) \quad = \quad \mathbb{E}\eta(B) \sum_{a \in B} \frac{\mathbb{E}\eta(\{a\})}{\mathbb{E}\eta(B)} P_a(Y).$$

Thus we can write $\eta^{\mathbf{1}_B}$ as a mixture: Pick an element $A \in B$ at random according to

$$P(A = a) \quad = \quad \frac{\mathbb{E}\eta(\{a\})}{\mathbb{E}\eta(B)};$$

if $A = a$ then put $\eta^{\mathbf{1}_B} = \eta^{\mathbf{1}_{\{a\}}}$.

Construction. Suppose the random measure has the form $\eta = \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i$. Let $\phi \in \mathcal{C}$ be nonnegative such that $\langle \mathbb{E}\eta, \phi \rangle = n^{-\frac{1}{2}} \sum_{i=1}^n \mathbb{E}\langle \eta_i, \phi \rangle > 0$. Pick an index $V \in \{1, \ldots, n\}$ proportional to the mean, such that $\mathbb{P}[V = v] \propto \mathbb{E}\langle \eta_v, \phi \rangle$. If $V = v$, independently of $\eta_1 \ldots, \eta_n$, choose $\eta_v^{v,\phi}$ having the η_v size biased distribution in direction ϕ. If $\eta_v^{v,\phi} = \eta$, choose $\eta_u^{v,\phi}, u \neq v$, such that

$$\mathcal{L}(\eta_u^{v,\phi}, u = 1, \ldots, n, u \neq v) = \mathcal{L}(\eta_u, u = 1, \ldots, n, u \neq v | \eta_v = \eta).$$

Put

$$\eta^\phi = \frac{1}{\sqrt{n}} \sum_{u=1}^n \eta_u^{V,\phi}.$$

Then we have, for all f for which the expectations exist,

$$\begin{aligned}
\mathbb{E}f(\eta^\phi) &= \sum_{v=1}^n \frac{\mathbb{E}\langle \eta_v, \phi \rangle}{\sum_{i=1}^n \mathbb{E}\langle \eta_i, \phi \rangle} \mathbb{E}f\left(\frac{1}{\sqrt{n}} \sum_{u=1}^n \eta_u^{v,\phi}\right) \\
&= \sum_{v=1}^n \frac{1}{\sum_{i=1}^n \mathbb{E}\langle \eta_i, \phi \rangle} \int_{M_1(E)} \langle \rho, \phi \rangle \mathbb{E}\left[f\left(\frac{1}{\sqrt{n}} \sum_{u=1}^n \eta_u\right) \middle| \eta_v = \rho \right] \mathbb{P}[\eta_v \in d\rho] \\
&= \sum_{v=1}^n \frac{1}{\sum_{i=1}^n \mathbb{E}\langle \eta_i, \phi \rangle} \mathbb{E}\left[\mathbb{E}[f(\eta)|\eta_v]\langle \eta_v, \phi \rangle\right] \\
&= \frac{1}{\langle \mathbb{E}\eta, \phi \rangle} \mathbb{E}f(\eta)\langle \eta, \phi \rangle.
\end{aligned}$$

Thus η^{ϕ} has the η size biased distribution in direction ϕ. In particular, if $\eta = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_{X_i}$, where $(X_i)_{i \in \mathbf{N}}$ is a family of random elements on E with laws $\mathcal{L}(X_i) = \mu_i$, and if $\phi \in \mathcal{C}$ is nonnegative such that $\sum_{i=1}^{n} \mathbb{E}\phi(X_i) > 0$, then, in the above construction, for any function $\psi \in \mathcal{C}$,

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \mathbb{E}\phi(X_i) \mathbb{E}\langle \eta^{\phi} - \eta, \psi \rangle = n^{-1} \sum_{i,v=1}^{n} \mathrm{Cov}\,(\phi(X_v), \psi(X_j)). \tag{13}$$

Moreover, if $V = v$ (which excludes indices i with $\mathbb{E}\phi(X_i) = 0$), $\delta_{X_v}^{v,\phi}$ can be constructed by choosing $X_v^{v,\phi}$, independently of the other variates, according to the distribution

$$\mathbb{P}[X_v^{v,\phi} \in dx] = \frac{\phi(x)}{\mathbb{E}\phi(X_v)} \mathbb{P}[X_v \in dx].$$

Theorem 10 *Let $(X_i)_{i \in \mathbf{N}}$ be a family of random elements on E with laws $\mathcal{L}(X_i) = \mu_i$, and let ξ_n be the corresponding empirical process. Choose $\mathcal{C} = C_c^{+}(E)$. Let ζ_n be a random measure taking values in $M^f(E)$ such that, for all $m \in \mathbf{N}, \phi_1, \ldots, \phi_m \in \mathcal{C}$, (5) is satisfied with B_n for B, where B_n is given in (8), and let $\mathcal{A}_n$ be the generator associated with B_n. Let $\eta = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_{X_i}$, so that $\xi_n = \eta - \mathbb{E}\eta$. Let $\eta^{\phi} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_{X_i^{V,\phi}}$ have the size-bias distribution of η in direction ϕ as constructed above. Then, for each $G \in C_{\mathcal{C}}(M^f(E))$,*

$$|\mathbb{E}G(\xi_n) - \mathbb{E}G(\zeta_n)| \quad \leq \quad \sqrt{n} \max_{\phi, \psi, \tau \in \mathcal{C}^*} \left\{ \left\{ \mathit{Var}(\mathbb{E}(\langle \eta^{\phi} - \eta, \psi \rangle | \eta)) \right\}^{\frac{1}{2}} + \frac{1}{2} \mathbb{E} |\langle \eta^{\phi} - \eta, \psi \rangle \langle \eta^{\phi} - \eta, \tau \rangle| \right\}.$$

PROOF: Let $F, f, \phi_1, \ldots, \phi_m$ be as in the assertion. For each $j = 1, \ldots, m$, as $\phi_j \geq 0$, if $\langle \mathbb{E}\eta, \phi_j \rangle = 0$, then $\langle \eta, \phi_j \rangle = 0$ almost surely, and the assertion is trivial. Hence we may restrict our attention to those ϕ_j with $\langle \mathbb{E}\eta, \phi_j \rangle > 0$. Using size-biasing, with Taylor expansion we obtain

$$\begin{aligned}
&\sum_{j=1}^{m} \mathbb{E} f_{(j)}\left(\underline{\langle \xi_n, \phi \rangle}\right) \langle \xi_n, \phi_j \rangle \\
&\quad = \sum_{j=1}^{m} \left(\langle \mathbb{E}\eta, \phi_j \rangle \mathbb{E} \left\{ f_{(j)}\left(\underline{\langle \eta^{\phi_j} - \mathbb{E}\eta, \phi \rangle}\right) - f_{(j)}\left(\underline{\langle \eta - \mathbb{E}\eta, \phi \rangle}\right) \right\} \right) \\
&\quad = \sum_{j,k=1}^{m} \langle \mathbb{E}\eta, \phi_j \rangle \mathbb{E} f_{(j,k)}\left(\langle \eta - \mathbb{E}\eta, \phi_1 \rangle, \ldots, \langle \eta - \mathbb{E}\eta, \phi_m \rangle\right) \langle \eta^{\phi_j} - \eta, \phi_k \rangle + R,
\end{aligned}$$

where

$$|R| \quad \leq \quad \frac{1}{2} \sum_{j,k,l=1}^{m} \langle \mathbb{E}\eta, \phi_j \rangle \| f_{(j,k,l)} \| \mathbb{E} | \langle \eta^{\phi_j} - \eta, \phi_k \rangle \langle \eta^{\phi_j} - \eta, \phi_l \rangle |.$$

Hence, using (13),

$$\begin{aligned}
&\mathbb{E}\mathcal{A}_n F(\xi_n) \\
&\quad = \sum_{j,k=1}^{m} \mathbb{E} f_{(j,k)}\left(\underline{\langle \xi_n, \phi \rangle}\right) \left(\frac{1}{n} \sum_{i,l=1}^{n} \mathrm{Cov}\,(\phi_j(X_i), \phi_k(X_l)) - \langle \mathbb{E}\eta, \phi_j \rangle \langle \eta^{\phi_j} - \eta, \phi_k \rangle \right) + R
\end{aligned}$$

$$= \quad -\sum_{j,k=1}^{m} \mathbb{E} f_{(j,k)}\left(\underline{\langle \xi_n, \phi \rangle}\right)\langle \mathbb{E}\eta, \phi_j \rangle \left(\langle \eta^{\phi_j} - \eta, \phi_k \rangle - \mathbb{E}\langle \eta^{\phi_j} - \eta, \phi_k \rangle \right) + R$$

$$= \quad -\sum_{j,k=1}^{m} \langle \mathbb{E}\eta, \phi_j \rangle \mathbb{E} f_{(j,k)}\left(\underline{\langle \xi_n, \phi \rangle}\right) \mathbb{E}\left\{ \langle \eta^{\phi_j} - \eta, \phi_k \rangle - \mathbb{E}\langle \eta^{\phi_j} - \eta, \phi_k \rangle | \eta \right\} + R.$$

Applying the Cauchy-Schwarz inequality and using the boundedness of the derivatives of f and of $\phi_1, \ldots, \phi_m$, we obtain the assertion. $\square$

As described in Goldstein and Reinert [13], for random measures also a zero-bias coupling can be defined. Details can be found in Reinert [20].

4 Example: An immigration-death process

We consider the following immigration-death process with total population size n. Let A_i^n be the (positive) arrival time of the ith individual and Z_i its life span, and assume the $(Z_i)_{i \in \mathbf{N}}$ are positive, i. i. d. and independent of the $(A_i^n)_{i,n \in \mathbf{N}}$. Denote the distribution function of Z_i by G. Suppose that $A_i^n = \sum_{j=1}^n E_i^n$, where $(E_i^n)_{i=1,\ldots,n}$ are i. i. d. positive random variables. Start at time $t = 0$, and let

$$X_i^n = (A_i^n, A_i^n + Z_i).$$

Then $\delta_{X_i^n}$ can be regarded as a measure on $\mathbf{R}_+^2$, where the half-open interval $[a, b) \subset [0, \infty)$ is represented by the point $(a, b) \in [0, \infty)^2$, and

$$\delta_{X_i^n}([0, t] \times [t, \infty)) = I[A_i^n \leq t < A_i^n + Z_i].$$

Thus $\delta_{X_i^n}$ describes the temporal evolution of the ith individual. Graphically, this point process corresponds to points on the positive quadrant $\mathbf{R}_+^2$, with points (immigration time, time of death) of an individual. In Reinert [21] conditions are given under which the law of the empirical measure converges weakly to δ_μ, for some $\mu \in M_1(\mathbf{R}^2)$. In the following we assume $(E_i^n)_{i=1,\ldots,n}$ to be such that $E_i^n = \frac{1}{n} Y_i$, where the Y_i have finite moments up to fourth order.

Here we will restrict attention to the behaviour for finite n. Let $\mu_i^n = \mathcal{L}(X_i)$, and let $\mu_n = \frac{1}{n}\sum_{i=1}^n \mathcal{L}(X_i)$. Firstly note that, with $\mathcal{C} = C_b^\infty(\mathbf{R}^2)$, the covariance matrix (8) is finite, and hence the Gaussian random measure with covariance matrix (8) is finite almost surely. Using Taylor expansion, we obtain the following approximation for the covariance matrix (the proof is in the appendix).

Lemma 11 *Abbreviate*

$$\bar{\phi}_x(y) = \phi(y, y + x) - \mathbb{E}\phi(A_i^n, A_i^n + Z_i). \tag{14}$$

Suppose that $E_i^n = \frac{1}{n}Y_i$, *where* $Y, Y_1, \dots, Y_n$ *are i.i.d. with* $\mathbb{E}Y^4 < \infty$. *Then, for* $\phi, \psi \in \mathcal{C}^*$,

$$\frac{1}{n}\sum_{i=1}^n Cov(\phi(A_i^n, A_i^n + Z_i), \psi(A_i^n, A_i^n + Z_i)) \;=\; \frac{1}{n}\sum_{i=1}^n \int \bar{\phi}_x(\mathbb{E}A_i^n)\bar{\psi}_x(\mathbb{E}A_i^n)G(dx) + R_{1,c}, \tag{15}$$

where

$$|R_{1,c}| \;\le\; \frac{3}{n}VarY,$$

and

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^n \sum_{l\ne i} Cov(\phi(A_i^n, A_i^n + Z_i), \psi(A_l^n, A_l^n + Z_l)) \\ &= \frac{1}{n}\sum_{i=1}^n \sum_{l>i} iVar(E_1)\int \bar{\phi}_x'(\mathbb{E}A_i^n)G(dx)\int \bar{\psi}_z'(\mathbb{E}A_l^n)G(dz) + R_{2,c}, \end{aligned} \tag{16}$$

where

$$|R_{2,c}| \;\le\; \frac{3}{n}(VarY)^2 + \frac{2}{n}(VarY)^{\frac{1}{2}} + \frac{2}{\sqrt{n}}(VarY)^{\frac{3}{2}}.$$

Note that, from this lemma, also a limiting expression for $n \to \infty$ could easily be obtained. Thus an approximation with a Gaussian random measure that does not depend on n could also be derived.

Proposition 12 *In the above immigration-death process, suppose that* $E_i^n = \frac{1}{n}Y_i$, *where* $Y, Y_1, \dots, Y_n$ *are i.i.d. with* $\mathbb{E}Y^4 < \infty$. *Let* ζ_n *be a random measure taking values in* $M^f(E)$ *almost surely such that, for all* $m \in \mathbf{N}, \phi_1, \dots, \phi_m \in \mathcal{C}$, *(5) is satisfied with* B_n *for* B, *where* B_n *is given in (8). Then, for each* $G \in C_{\mathcal{C}}(M^f(E))$ *we have*

$$|\mathbb{E}G(\xi_n) - \mathbb{E}G(\zeta_n)| \;\le\; \frac{1}{\sqrt{n}}\left[4 + \max\{(\mathbb{E}Y^2)^{\frac{1}{2}}, (\mathbb{E}Y^2)^{\frac{3}{2}}\}(11 + 2(\mathbb{E}Y^4)^{\frac{1}{2}}) + \frac{9}{\sqrt{n}}\max\{(\mathbb{E}Y^2)^{\frac{1}{2}}, (\mathbb{E}Y^2)^2\}\right].$$

Remark 13 *The order of the bound,* $\frac{1}{\sqrt{n}}$, *is to be expected from one-dimensional approximations. There is no claim that the constant is optimal.*

PROOF. Put $\xi_n^i = \frac{1}{\sqrt{n}}\sum_{j\ne i}(\delta_{X_j} - \mu_j^n)$. With the notation (14), by Taylor expansion

$$\begin{aligned} \mathbb{E}\sum_{j=1}^m f_{(j)}(\underline{\langle \xi_n, \phi\rangle})\langle \xi_n, \phi_j\rangle \;&=\; \frac{1}{\sqrt{n}}\sum_i \sum_{j=1}^m \mathbb{E}f_{(j)}(\underline{\langle \xi_n^i, \phi\rangle})\langle \delta_{X_i} - \mu_i^n, \phi_j\rangle \\ &\quad + \frac{1}{n}\sum_i \sum_{j,k=1}^m \mathbb{E}f_{(j,k)}(\underline{\langle \xi_n^i, \phi\rangle})\langle \delta_{X_i} - \mu_i^n, \phi_j\rangle\langle \delta_{X_i} - \mu_i^n, \phi_k\rangle + R_1 \\ &= \frac{1}{\sqrt{n}}\sum_i \sum_{j=1}^m \mathbb{E}f_{(j)}(\underline{\langle \xi_n^i, \phi\rangle})\int (\bar{\phi}_j)_z(A_i^n)G(dz) \\ &\quad + \frac{1}{n}\sum_i \sum_{j,k=1}^m \mathbb{E}f_{(j,k)}(\underline{\langle \xi_n^i, \phi\rangle})\int (\bar{\phi}_j)_z(A_i^n)(\bar{\phi}_k)_z(A_i^n)G(dz) + R_1 \end{aligned}$$

$$= \frac{1}{\sqrt{n}}\sum_i\sum_{j=1}^m \mathbb{E}f_{(j)}(\underline{\langle\xi_n^i,\phi\rangle})\int(\bar\phi_j)_z(A_i^n)G(dz) \tag{17}$$
$$+\frac{1}{n}\sum_i\sum_{j,k=1}^m \mathbb{E}f_{(j,k)}(\underline{\langle\xi_n^i,\phi\rangle})\int(\bar\phi_j)_z(\mathbb{E}A_i^n)(\bar\phi_k)_z(\mathbb{E}A_i^n)G(dz)+R_1+R_2,$$

where

$$|R_1|\le\frac{4}{\sqrt{n}}$$

and

$$|R_2| \le \frac{1}{n}\sum_i(Var(A_i^n))^{\frac12}\le n^{\frac12}(Var(E_1))^{\frac12}.$$

For (17), we have, from the independence, and by Taylor expansion around $\mathbb{E}A_i^n$,

$$\frac{1}{\sqrt{n}}\sum_i\sum_{j=1}^m \mathbb{E}f_{(j)}(\underline{\langle\xi_n^i,\phi\rangle})\int(\bar\phi_j)_z(A_i^n)G(dz)$$
$$= \frac{1}{\sqrt{n}}\sum_i\sum_{j=1}^m \mathbb{E}f_{(j)}(\underline{\langle\xi_n^i,\phi\rangle})\int(A_i^n-\mathbb{E}A_i^n)(\bar\phi_j)'_z(\mathbb{E}A_i^n)G(dz)+R_3$$
$$= \frac{1}{\sqrt{n}}\sum_i\int(\bar\phi_j)'_z(\mathbb{E}A_i^n)G(dz)\sum_{j=1}^m\sum_{l=1}^i\mathbb{E}(E_l-\mathbb{E}E_l)f_{(j)}(\underline{\langle\xi_n^i,\phi\rangle})+R_3,$$

where

$$|R_3| \le 2n^{\frac32}Var(E_1).$$

Put $(A_i^n)^l=\sum_{j=1,j\neq l}^i E_j$, let $\mu_j^{n,l}=\mathcal{L}((A_j^n)^l,(A_j^n)^l+Z_j))$, and put

$$\xi_n^{i,l}=\frac{1}{\sqrt{n}}\sum_{j\neq i}(\delta_{(A_j^n)^l,(A_j^n)^l+Z_j)}-\mu_j^{n,l}).$$

Then $\xi_n^{i,l}$ is independent of E_l, and we obtain, by Taylor expansion,

$$\frac{1}{\sqrt{n}}\sum_i\int(\bar\phi_j)'_z(\mathbb{E}A_i^n)G(dz)\sum_{j=1}^m\sum_{l=1}^i\mathbb{E}(E_l-\mathbb{E}E_l)f_{(j)}(\underline{\langle\xi_n^i,\phi\rangle})$$
$$= \frac{1}{\sqrt{n}}\sum_i\int(\bar\phi_j)'_z(\mathbb{E}A_i^n)G(dz)\sum_{j,k=1}^m\sum_{l=1}^i\mathbb{E}(E_l-\mathbb{E}E_l)f_{(j,k)}(\underline{\langle\xi_n^{i,l},\phi\rangle})\langle\xi_n^i-\xi_n^{i,n},\phi_k\rangle+R_4,$$

where

$$|R_4| \le \frac{2}{\sqrt{n}}\sum_i\sum_{l=1}^i\frac{1}{n}\sum_{t,s\ge l}\mathbb{E}|E_l-\mathbb{E}E_l|E_l^2\le n^{\frac52}(VarE_1)^{\frac12}(\mathbb{E}E_1^4)^{\frac12}.$$

Now,

$$\frac{1}{\sqrt{n}}\sum_i \int (\bar\phi_j)'_z(\mathbb{E}A_i^n)G(dz)\sum_{j,k=1}^m\sum_{l=1}^i \mathbb{E}(E_l-\mathbb{E}E_l)f_{(j,k)}(\underline{\langle \xi_n^{i,l},\phi\rangle})\langle \xi_n^i-\xi_n^{i,l},\phi_k\rangle$$

$$= \frac{1}{n}\sum_{i,t=}^n\sum_{j,k=1}^m\sum_{l=1}^i \int (\bar\phi_j)'_z(\mathbb{E}A_i^n)G(dz)\mathbb{E}(E_l-\mathbb{E}E_l)f_{(j,k)}(\underline{\langle \xi_n^{i,l},\phi\rangle})(\phi_k(A_t^n,A_t^n+Z_t)-\phi_k((A_t^n)^l,(A_t^n)^l+Z_t))$$

$$= \frac{1}{n}\sum_i\sum_{j,k=1}^m\sum_{l=1}^i\sum_{t\ge l}\int (\bar\phi_j)'_z(\mathbb{E}A_i^n)G(dz)\mathbb{E}(E_l-\mathbb{E}E_l)E_lf_{(j,k)}(\underline{\langle \xi_n^{i,l},\phi\rangle})\phi_k'((A_t^n)^l,(A_t^n)^l+Z_t)+R_5,$$

where

$$|R_5| \le \frac{1}{n}\sum_i\sum_{l=1}^i\sum_{t\ge l}\mathbb{E}|E_l-\mathbb{E}E_l|E_l^2\le n^2(VarE_1)^{\frac12}(\mathbb{E}E_1^4)^{\frac12}.$$

Put

$$\xi_n^{i,l,t}=\frac{1}{\sqrt{n}}\sum_{j\ne i,t}(\delta_{(A_j^n)^l,(A_j^n)^l+Z_j)}-\mu_j^{n,l}),$$

so that $\xi_n^{i,l,t}$ is independent of Z_t. Then

$$\frac{1}{n}\sum_i\sum_{j,k=1}^m\sum_{l=1}^i\sum_{t\ge l}\int (\bar\phi_j)'_z(\mathbb{E}A_i^n)G(dz)\mathbb{E}(E_l-\mathbb{E}E_l)E_lf_{(j,k)}(\underline{\langle \xi_n^{i,l},\phi\rangle})\phi_k'((A_t^n)^l,(A_t^n)^l+Z_t)$$

$$= \frac{1}{n}\sum_i\sum_{j,k=1}^m\sum_{l=1}^i\sum_{t\ge l}\int (\bar\phi_j)'_z(\mathbb{E}A_i^n)G(dz)\mathbb{E}(E_l-\mathbb{E}E_l)E_lf_{(j,k)}(\underline{\langle \xi_n^{i,l,t},\phi\rangle})\int(\bar\phi_k)'_z((A_t^n)^l)G(dz)+R_5,$$

where

$$|R_5| \le \frac{1}{n\sqrt{n}}\sum_i\sum_{l=1}^i\sum_{t\ge l}\mathbb{E}|(E_l-\mathbb{E}E_l)E_l|\le n^{\frac32}(VarE_1)^{\frac12}(\mathbb{E}E_1^2)^{\frac12}.$$

Approximate $(A_t^n)^l$ by $\mathbb{E}A_t^n$ again:

$$\frac{1}{n}\sum_i\sum_{j,k=1}^m\sum_{l=1}^i\sum_{t\ge l}\int (\bar\phi_j)'_z(\mathbb{E}A_i^n)G(dz)\mathbb{E}(E_l-\mathbb{E}E_l)E_lf_{(j,k)}(\underline{\langle \xi_n^{i,l,t},\phi\rangle})\int(\bar\phi_k)'_z((A_t^n)^l)G(dz)$$

$$= \frac{1}{n}\sum_i\sum_{t=1}^i Var(E_1)\int (\bar\phi_j)'_z(\mathbb{E}A_i^n)G(dz)\int(\bar\phi_k)'_z(\mathbb{E}A_t^n)G(dz)\sum_{l\le t}\sum_{j,k=1}^m\mathbb{E}f_{(j,k)}(\underline{\langle \xi_n^{i,l,t},\phi\rangle})+R_6+R_7,$$

where

$$|R_6| \le \frac{1}{n}\sum_i\sum_{l=1}^i\sum_{t\ge l}\mathbb{E}|(E_l-\mathbb{E}E_l)E_1|\mathbb{E}|E_l|\le n^2(\mathbb{E}E_1^2)^{\frac32}$$

and

$$|R_7| \le \frac{1}{n}\sum_i\sum_{l=1}^i\sum_{t\ge l}\mathbb{E}|(E_1-\mathbb{E}E_1)E_1|(VarA_l^n)^{\frac12}\le n^{\frac52}(\mathbb{E}E_1^2)^{\frac32}.$$

Taylor expansion again approximates

$$\sum_{l\leq t}\sum_{j,k=1}^{m}\mathbb{E}f_{(j,k)}(\underline{\langle\xi_n^{i,l,t},\phi\rangle}) = t\sum_{j,k=1}^{m}\mathbb{E}f_{(j,k)}(\underline{\langle\xi_n,\phi\rangle}) + R_8.$$

The term R_8 can be bounded as before.

Using the crude bound that $\mathrm{Var}(E_1)\leq\mathbb{E}(E_1^2)$, comparing with (15) and (16) and bounding the remainder term R_8 as before gives the assertion. □

Remark 14 *Note that the immigration-death process treated above is not necessarily a Markovian process. Moreover, the result could easily be generalized to weakly dependent or locally dependent inter-arrival times.*

Acknowledgements: Part of this work is contained in my Ph.D. thesis, and I am very grateful to my advisor A.D. Barbour for many discussions, and for his support during and after the thesis. Furthermore, I would like to thank L. Goldstein for his help concerning the appropriate formulation of size biasing for random measures, and L. Gorostiza for pointing out a possible connection to Palm measures. P. Eichelsbacher deserves thanks for emphasizing it. This work has been supported in part by ARC grant No. 1161.

Appendix

Proof of Lemma 3

For $F(\nu)\in C_{\mathcal{C}}(M^f(E))$, and for $t\geq 0$, define the semigroup $(T_t)_t$ corresponding to the generator $\mathcal{A}$ is given by

$$T_tF(\nu) \quad = \quad \mathbb{E}F(\nu e^{-t}+\sqrt{1-e^{-2t}}\zeta).$$

We claim that if $G\in C_{\mathcal{C}}(M^f(E))$ has the form $G(\nu)=g(\langle\nu,\psi_1\rangle,\ldots,\langle\nu,\psi_m\rangle)$, the function

$$F(\nu)=F_G(\nu)=-\int_0^\infty\{T_tG(\nu)-\mathbb{E}G(\zeta)\}dt \tag{18}$$

solves (7) for G. From the generator method it follows that F given in (18) is a solution of the Stein equation (7) if F exists. But the existence of F follows easily using Taylor expansion.

It remains to show that $F\in C_{\mathcal{C}}(M^f(E))$, that $\|f^{(k)}\|\leq\|g^{(k)}\|$, and $\|F^{(k)}\|\leq\|G^{(k)}\|$, for all $k\in\mathbf{N}$. We have that $F(\nu)=f(\langle\nu,\psi_1\rangle,\ldots,\langle\nu,\psi_m\rangle)$, where

$$\begin{aligned}
&f(x_1,\ldots,x_m)\\
&= \int_0^\infty[\mathbb{E}g(\langle\zeta,\psi_1\rangle,\ldots,\langle\zeta,\psi_m\rangle)\\
&\quad -g(x_1e^{-t}+\sqrt{1-e^{-2t}}\langle\zeta,\psi_1\rangle,\ldots,x_me^{-t}+\sqrt{1-e^{-2t}}\langle\zeta,\psi_m\rangle)]\,dt.
\end{aligned}$$

By dominated convergence, $f \in C_b^\infty(\mathbf{R}^m)$. Furthermore, obviously $\|f^{(k)}\| \leq \|g^{(k)}\|, k \in \mathbf{N}$. Hence $\|F^{(k)}\| \leq \|G^{(k)}\|, k \in \mathbf{N}$. $\square$

Proof of Lemma 11

For all $\phi, \psi \in \mathcal{C}^*$, we have

$$\begin{aligned} \frac{1}{n}\sum_{i,l=1}^n Cov(\phi(X_i),\psi(X_l)) &= \frac{1}{n}\sum_{i=1}^n Cov(\phi(A_i^n, A_i^n+Z_i),\psi(A_i^n, A_i^n+Z_i)) \\ &+\frac{1}{n}\sum_{i=1}^n\sum_{l\neq i} Cov(\phi(A_i^n, A_i^n+Z_i),\psi(A_l^n, A_l^n+Z_l)) \end{aligned}$$

With the notation (14),

$$\frac{1}{n}\sum_{i=1}^n Cov(\phi(A_i^n, A_i^n+Z_i),\psi(A_i^n, A_i^n+Z_i)) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\int \bar{\phi}_x(A_i^n)\bar{\psi}_x(A_i^n)G(dx).$$

Taylor expansion around $\mathbb{E}A_i^n$ yields

$$\frac{1}{n}\sum_{i=1}^n Cov(\phi(A_i^n, A_i^n+Z_i),\psi(A_i^n, A_i^n+Z_i)) = \frac{1}{n}\sum_{i=1}^n \int \bar{\phi}_x(\mathbb{E}A_i^n)\bar{\psi}_x(\mathbb{E}A_i^n)G(dx) + R_{1,c},$$

where, for $\phi, \psi \in \mathcal{C}^*$,

$$|R_{1,c}| \leq \frac{3}{n}\sum_{i=1}^n Var(A_i^n) \leq 3nVar(E_1).$$

Using that $E_i^n = \frac{1}{n}Y_i$ gives (15).

Moreover, again by Taylor expansion,

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^n\sum_{l\neq i} Cov(\phi(A_i^n, A_i^n+Z_i),\psi(A_l^n, A_l^n+Z_l)) \\ &= \frac{2}{n}\sum_{i=1}^n\sum_{l>i}\mathbb{E}\int \bar{\phi}_x(A_i^n)G(dx)\int\bar{\psi}_z(A_l^n)G(dz) \\ &= \frac{2}{n}\sum_{i=1}^n\sum_{l>i}\mathbb{E}\Bigg\{\int \bar{\phi}_x(\mathbb{E}A_i^n)G(dx)\int\bar{\psi}_z(\mathbb{E}A_i^n+\sum_{j=i+1}^l E_j)G(dz) && (19) \\ &+(A_i^n-\mathbb{E}A_i^n)\int\bar{\phi}'_x(\mathbb{E}A_i^n)G(dx)\int\bar{\psi}_z(eA_i^n+\sum_{j=i+1}^l E_j)G(dz) && (20) \\ &+(A_i^n-\mathbb{E}A_i^n)\int\bar{\phi}_x(\mathbb{E}A_i^n)G(dx)\int\bar{\psi}'_z(eA_i^n+\sum_{j=i+1}^l E_j)G(dz) && (21) \\ &+\frac{1}{2}(A_i^n-\mathbb{E}A_i^n)^2\int\bar{\phi}''_x(\mathbb{E}A_i^n)G(dx)\int\bar{\psi}_z(eA_i^n+\sum_{j=i+1}^l E_j)G(dz) && (22) \end{aligned}$$

$$+\frac{1}{2}(A_i^n - \mathbb{E}A_i^n)^2 \int \bar{\phi}_x(\mathbb{E}A_i^n)G(dx) \int \bar{\psi}_z''(eA_i^n + \sum_{j=i+1}^{l} E_j)G(dz) \tag{23}$$

$$+\frac{1}{2}(A_i^n - \mathbb{E}A_i^n)^2 \int \bar{\phi}_x'(\mathbb{E}A_i^n)G(dx) \int \bar{\psi}_z'(eA_i^n + \sum_{j=i+1}^{l} E_j)G(dz) \Bigg\} \tag{24}$$

$$+R_{2,c}, \tag{25}$$

where

$$|R_{2,c}| \leq \frac{6}{3n} \sum_{i=1}^{n} \sum_{l>i} \mathbb{E}|A_i^n - \mathbb{E}A_i^n|^3 \leq 2n^2 \mathbb{E}|E_1 - \mathbb{E}E_1|^3 + n^{\frac{5}{2}}(VarE_1)^{\frac{3}{2}}.$$

Due to the independence, (20) and (21) vanish. For (19), note that

$$\left| \int \bar{\psi}_z(\mathbb{E}A_l^n)G(dz) \right| = |\mathbb{E}\{\psi(\mathbb{E}A_i^n, \mathbb{E}A_i^n + Z_i) - \psi(A_i^n, A_i^n + Z_i)\}| \leq \frac{1}{2} Var(A_i^n).$$

Hence we have that

$$\frac{2}{n} \sum_{i=1}^{n} \sum_{l>i} \mathbb{E} \int \bar{\phi}_x(\mathbb{E}A_i^n)G(dx) \int \bar{\psi}_z(\mathbb{E}A_i^n + \sum_{j=i+1}^{l} E_j)G(dz) = R_{3,c},$$

and

$$|R_{3,c}| \leq \frac{1}{n} \sum_{i=1}^{n} \sum_{l>i} Var(A_i^n) Var(A_l^n) \leq n^3 (VarE_1)^2.$$

Similarly we may bound (22) and (23). This gives

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} \sum_{l \neq i} Cov(\phi(A_i^n, A_i^n + Z_i), \psi(A_l^n, A_l^n + Z_l)) \\ &= \frac{1}{n} \sum_{i=1}^{n} \sum_{l>i} \mathbb{E}(A_i^n - \mathbb{E}A_i^n)^2 \int \bar{\phi}_x'(\mathbb{E}A_i^n)G(dx) \int \bar{\psi}_z'(\mathbb{E}A_i^n + \sum_{j=i+1}^{l} E_j)G(dz) + R_{4,c}, \end{aligned}$$

where

$$|R_{4,c}| \leq 3n^3 (VarE_1)^2 + 2n^2 \mathbb{E}|E_1 - \mathbb{E}E_1|^3 + n^{\frac{5}{2}}(VarE_1)^{\frac{3}{2}}.$$

Expanding around $\mathbb{E}\sum_{j=i+1}^{l} E_j$ yields

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} \sum_{l \neq i} Cov(\phi(A_i^n, A_i^n + Z_i), \psi(A_l^n, A_l^n + Z_l)) \\ &= \frac{1}{n} \sum_{i=1}^{n} \sum_{l>i} iVar(E_1) \int \bar{\phi}_x'(\mathbb{E}A_i^n)G(dx) \int \bar{\psi}_z'(\mathbb{E}A_l^n)G(dz) + R_{5,c}, \end{aligned}$$

where

$$|R_{5,c}| \leq 3n^3 (VarE_1)^2 + 2n^2 \mathbb{E}|E_1 - \mathbb{E}E_1|^3 + 2n^{\frac{5}{2}}(VarE_1)^{\frac{3}{2}}.$$

Using that $E_i^n = \frac{1}{n} Y_i$ gives the second assertion. □

References

[1] Arcones, M. A. and Yu. B. (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *Journal Theor. Probab.,* **7**, 47 – 71.

[2] Baldi, P., Rinott, Y., and Stein, C. (1989). A normal approximation for the number of local maxima of a random function on a graph. In *Probability, Statistics and Mathematics. Papers in Honor of Samuel Karlin.* (ed. Anderson, T.W., Athreya, K.B., and Iglehart, D.L.), 59 – 81. Academic Press, Boston.

[3] Barbour, A.D. (1990). Stein's method for diffusion approximations. *Probab. Theory Rel. Fields,* **84**, 297 – 322.

[4] Barbour, A.D., Holst, L., and Janson, S. (1992). *Poisson approximation.* Clarendon, Oxford.

[5] Bolthausen, E., and Götze, F. (1993). The rate of convergence for multivariate sampling statistics. *Ann. Statist.* **21**, 1692–1710.

[6] Dembo, A. and Zajic, T. (1997). Uniform large and moderate deviations for functional empirical processes. *Stoch. Proc. Applic.* **67**, 195–211.

[7] Einmahl, J.H.J., Mason, D.N., and Ruymgaart, F.H. (1988) Processus empiriques multidimensionnels: Aperçu de quelques résultats récents. *Intern. Statistical Review,* **56**, 139 – 151.

[8] Ethier, S.N. and Kurtz, T.G. (1986). *Markov processes: characterization and convergence.* Wiley, New York.

[9] van de Geer, S. (1999). *Applications of empirical process theory.* Cambridge University Press.

[10] Giné, E. (1996). Empirical processes and applications: an overview. *Bernoulli,* **2**, 1 – 28.

[11] Götze, F. (1985). Asymptotic expansions in functional limit theorems. *J. Multivariate Analysis.,* **16**, 1 – 20.

[12] Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.,* **11**, 724 – 739.

[13] Goldstein, L., and Reinert, G. (1997). Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.,* **7**, 935–952.

[14] Goldstein, L. and Rinott, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Prob.,* **33**, 1–17.

[15] Holmes, S. and Reinert, G. (2000). Stein's method for the bootstrap. To appear in *Stein's method and Monte Carlo Markov chains.* P. Diaconis and S. Holmes, eds.

[16] Ledoux, M. and Talagrand, M. (1989). Comparison theorems, random geometry and some limit theorems for empirical processes. *Ann. Probab.,* **17**, 596 – 631.

[17] Philipp, W. (1982). Invariance principles for sums of mixing random elements and the multivariate empirical process. *Colloquia Mathematica Societatis János Bolyai; 36. Limit Theorems in Probability and Statistics, Veszprém,* 843 – 873.

[18] Rachev, S. T. (1991). *Probability metrics and the stability of stochastic models.* Wiley, Chichester.

[19] Raij, M. (2000). On the rate of convergence in the multivariate CLT for dependent random vectors. Preprint.

[20] Reinert, G. (1999). An introduction to Stein's method and application to empirical measures. In *Modelos Estocasticos.* M.Gonzalez Barrios and L.G.Gorostiza, eds., Sociedad Matematica Mexicana, 65 - 120. (Proceedings of the Symposium on Probability and Stochastic Processes, Guanajuato, Mexico).

[21] Reinert, G. (1995). A weak law of large numbers for empirical measures via Stein's method. *Ann. Probab.,* **23**, 334 – 354.

[22] Reinert, G. (1994). *A weak law of large numbers for empirical measures via Stein's method, and applications.* Ph. D. thesis, Zürich.

[23] Rinott, Y. and Rotar, V. (1996) A multivariate CLT for local dependence with $n^{-1/2}\log n$ rate and applications to multivariate graph related statistics. *J. Multivariate Analysis.* **56**, 333–350.

[24] Shao, Q.-M. and Yu, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *Ann. Probab.* **24**, 2098–2127.

[25] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2**, 583 – 602.

[26] Stein, C. (1986). *Approximate computation of expectations.* Institute of Mathematical Statistics Lecture Notes - Monograph Series, Vol. **7**, Hayward, California.

[27] Stoyan, D., Kendall, W.S., and Mecke, J. (1995). *Stochastic Geometry and its Applications.* 2^{nd} ed. Wiley, New York.

[28] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes.* Springer, New York.

[29] Vakhania, N.N., Tarieladze, V.I. and Chobanyan, S.A. (1987). *Probability distributions on Banach spaces.* D. Reidel, Dordrecht.