Stein's Method in Application to Empirical Measures

V Simposio De Probabilidad Y Procesos Estocasticos

Marzo 29 - Abril 3, 1998, CIMAT, A.C., Guanajuato, Mexico

Gesine Reinert *

April 11, 2001

Contents

*Department of Mathematics, University of California, Los Angeles CA 90095-1555, reinert@stat.ucla.edu

Overview

In the following lecture notes we give an overview on Stein's method, focussing on application to random measures. Stein's method, based on a paper by Stein [73], has been applied in numerous settings during the last ten or fifteen years. Besides describing the method in general, here we will consider three main cases, namely Gaussian approximations, laws of large numbers, and Poisson approximations. Together with the theory we will introduce couplings as a set of tools that have proved to be useful in connection with Stein's method.

We will see that in particular if the random elements underlying the empirical measure are dependent, Stein's method is an excellent way of bounding distances to the target distribution. It must be emphasized that Stein's method is not only useful in proving weak convergence, but moreover provides an explicit bound on the distance to the limiting distribution.

In Section 1 we describe Stein's method for real-valued random elements. For historical reasons, we first introduce Stein's method for normal approximations, before describing the approach for the weak law of large numbers and for Poisson approximations. As main couplings, the local approach, exchangeable pairs, size biasing and zero biasing are introduced. To keep this introduction short, the only example discussed is the sum of i.i.d. variates.

Section 2 gives the setting for expanding Stein's method to measure-valued random elements. The theoretical results for laws of large numbers, for Gaussian approximations, and for Poisson (point process) approximation, are given. Here we treat the law of large numbers first, because it is the easiest case. The normal case can be investigated by related means. As the Poisson case will turn out to be of different nature, we put it at the end of the section.

Stein's method reveals much of its power when investigating dependent random elements. The last sections give some examples that illustrate how to apply Stein's method to sums of dependent variates.

In Section 3, as examples for the law of large numbers we first give a dissociated array, illustrating the local coupling. Furthermore an immigration-death process and the general stochastic epidemic are considered. The latter two examples show how Stein's method can be applied to stochastic processes.

Section 4 discusses iterates of random maps, and simple random sampling, as examples for Gaussian approximations.

Lastly, in Section 5 Poisson (point process) approximation is illustrated by the example of counting joint occurrences of multiple words in DNA sequences. We will see that Poisson approximations for measure-valued elements can be treated as Poisson process approximations, and hence the well-developed theory on Poisson process approximations can be applied here (see Section 1 for references). Thus many more examples can be found in the literature.

1 Stein's Method for Real-Valued Random Elements

1.1 Stein's Method for Normal Approximations

1.1.1 Stein's Equation

In 1972, Stein [73] published a very elegant method to prove normal approximations. It is based on the fact that a random variable Z is standard normal if and only if for all smooth, real-valued functions f,

$$\mathbf{E}\left\{f''(Z) - Zf'(Z)\right\} = 0.$$

(This is easily seen using dominated convergence and integration by parts.) Stein [73] then showed that for any smooth, real-valued function h there is a function f solving the now-called "Stein equation"

$$f''(x) - xf'(x) = h(x) - \Phi h, \tag{1}$$

Φh denoting the expectation of h with respect to the standard normal density. Moreover, there is a solution f of the Stein equation (1) satisfying

$$\|f'\| \le \sqrt{\frac{\pi}{2}}\|h - \Phi h\|; \quad \|f''\| \le (\sup h - \inf h); \quad \|f^{(3)}\| \le 2\|h'\|; \tag{2}$$

where $\|\cdot\|$ denotes the supremum norm and $f^{(k)}$ denotes the kth derivative of f (see Stein [74], p.25 and Baldi *et al.* [8]). Now, for any random variable W, taking expectations in (1) gives

$$\mathbf{E}h(W) - \Phi h = \mathbf{E}f''(W) - \mathbf{E}Wf'(W). \tag{3}$$

Thus the distance of W from the normal, in terms of a test function h, can be bounded by bounding the right-hand side of (3); the immediate bound on the distance is one of the key advantages of Stein's method compared to moment-generating functions or characteristic functions. Typical classes of test functions h are $C^4(\mathbf{R})$ (for weak convergence), or the class of indicator functions of half-lines (giving the Kolmogorov distance).

Nearly twenty years later, Barbour [11] and Götze [42] derived alternative bounds for the solution f of (1), namely

$$\|f^{(j)}\| \le j^{-1}\|h^{(j)}\|, \quad j = 1, 2, \dots \tag{4}$$

The main advantage of these bounds is that they can easily be extended to more general Gaussian approximations; Barbour [11] considered diffusion approximations, Götze [42] multivariate normal approximations.

Evidently, Stein's method is in place when the right-hand side of (3) is easier to bound than the left-hand side of (3). A typical approach for evaluating the right-hand side of (3) is to employ couplings, and often the success of the method is connected with finding an effective coupling for the right-hand side of (3).

1.1.2 The Local Approach

Suppose $X_1, \dots X_n$ are independent, mean zero, and let $\sigma_i^2 = \mathrm{Var}(X_i)$. Put $W = \sum_{i=1}^n X_i$, and assume $\mathrm{Var}(W) = \sum_{i=1}^n \sigma_i^2 = 1$. For each $i = 1, \dots, n$ put

$$W_i = W - X_i = \sum_{j \ne i} X_j.$$

Thus W and W_i are defined on the same probability space. For any smooth function f we have, by Taylor expansion

$$\begin{aligned} \mathbf{E}Wf'(W) &= \sum_{i=1}^n \mathbf{E}X_i f'(W) \\ &= \sum_{i=1}^n \mathbf{E}X_i f'(W_i) + \sum_{i=1}^n \mathbf{E}X_i^2 f''(W_i) + R, \end{aligned}$$

where

$$|R| \le \frac{1}{2}\|f^{(3)}\| \sum_{i=1}^n \mathbf{E}|X_i^3|.$$

Using the independence we obtain for the right-hand side of (3) that

$$\begin{aligned}
\mathbf{E}Wf'(W) - \mathbf{E}f''(W) &= \sum_{i=1}^n \mathbf{E}X_i^2 \mathbf{E}f''(W_i) - \mathbf{E}f''(W) + R \\
&= \sum_{i=1}^n \sigma_i^2 \mathbf{E}\{f''(W_i) - f''(W)\} + R.
\end{aligned}$$

Taylor expansion and the bounds (2) now give the following theorem (see Barbour [9]).

Theorem 1 *Suppose $X_1, \ldots X_n$ are independent, mean zero, and let $\sigma_i^2 = \mathrm{Var}(X_i)$. Put $W = \sum_{i=1}^n X_i$, and assume $\mathrm{Var}(W) = 1$. For any continuous, bounded function h with piecewise continuous, bounded first derivative, we have*

$$|\mathbf{E}h(W) - \Phi h| \le \|h'\| \left(2\sum_{i=1}^n \sigma_i^3 + \sum_{i=1}^n \mathbf{E}|X_i^3| \right).$$

This approach can be generalized for locally dependent variables; we will see examples later.

1.1.3 Exchangeable Pair Couplings

A pair (W, W') of random variables defined on the same probability space is called *exchangeable* if for all measurable sets B and B',

$$\mathbf{P}(W \in B, W' \in B') = \mathbf{P}(W \in B', W' \in B).$$

Following Stein [74] we assume that W is mean zero, variance 1, and that there is a $0 < \lambda < 1$ such that

$$\mathbf{E}(W'|W) = (1-\lambda)W. \tag{5}$$

This assumption can be related to regression; if (W, W') is bivariate normal with correlation ρ, then $1 - \lambda = \rho$. Under (5) it is easy to see that

$$\mathbf{E}Wf'(W) = \frac{1}{2\lambda}\mathbf{E}(W - W')(f'(W) - f'(W')). \tag{6}$$

Taylor expansion gives the following result (see Stein [74]).

Theorem 2 *Let (W, W') be an exchangeable pair satisfying (5) and assume $\mathbf{E}W = 0, \mathrm{Var}(W) = 1$. For any continuous, bounded function $h : \mathbf{R} \to \mathbf{R}$ with bounded, piecewise continuous derivatives, we have*

$$
\begin{aligned}
&|\mathbf{E}h(W) - \Phi h| \\
&\leq (\sup h - \inf h)\sqrt{\mathbf{E}\left(1 - \frac{1}{2\lambda}\mathbf{E}((W - W')^2|W)\right)} + \frac{1}{4\lambda}\|h'\|\mathbf{E}|W - W'|^3.
\end{aligned}
$$

Example 1 *Let us assume that $X_1, \ldots, X_n$ are independent, mean zero; let $\sigma_i^2 = \mathrm{Var}(X_i)$ and $\sum_{i=1}^n \sigma_i^2 = 1$. As usual, we consider $W = \sum_{i=1}^n X_i$. To construct W' such that (W, W') is an exchangeable pair, pick an index I uniformly from $\{1, \ldots, n\}$. If $I = i$, we replace X_i by an independent copy X_i^*, and we put*

$$W' = W - X_I + X_I^*. \tag{7}$$

Then (W, W') is exchangeable, and

$$\mathbf{E}(W'|W) = W - \frac{1}{n}W + \mathbf{E}X_I^* = \left(1 - \frac{1}{n}\right)W,$$

so (5) is satisfied with $\lambda = \frac{1}{n}$. Theorem 2 thus gives

$$
\begin{aligned}
&|\mathbf{E}h(W) - \Phi h| \\
&\leq (\sup h - \inf h)\sqrt{\mathbf{E}\left(1 - \frac{n}{2}\mathbf{E}((W - W')^2|W)\right)} + \frac{n}{4}\|h'\|\mathbf{E}|W - W'|^3.
\end{aligned}
$$

Bounding the expectations gives (see Stein [74], p.37)

$$
\begin{aligned}
&|\mathbf{E}h(W) - \Phi h| \\
&\leq (\sup h - \inf h)\sqrt{\sum_{i=1}^n (\mathbf{E}X_i^4 - \sigma_i^4)} + \frac{1}{2}\|h'\|\sum_{i=1}^n \left(\mathbf{E}|X_i|^3 + 3\sigma_i^3\right).
\end{aligned}
$$

The construction used in Example 1 is typical for the exchangeable pair approach.

Construction 1 *Let $X_1, \ldots, X_n$ be possibly dependent, mean zero variates with existing variances, and let $W = \sum_{i=1}^n X_i$. Pick an index I uniformly from $\{1, \ldots, n\}$. If $I = i$, replace X_i by an independent copy X_i^*. If $X_i^* = x$, construct $\hat{X}_j, j \neq i$ such that*

$$\mathcal{L}(\hat{X}_j, j \neq i) = \mathcal{L}(X_j, j \neq i | X_i = x).$$

Then put

$$W' = \sum_{j \neq I} \hat{X}_j + X_I^*.$$

The pair (W, W') is exchangeable. If, in addition, (5) holds, then Theorem 2 can be applied. However, not always will the exchangeable pair (W, W') constructed above satisfy (5); see, for example, Rinott and Rotar [68].

1.1.4 Size Bias Couplings

It is not always easy to construct an exchangeable pair for a given problem. The so-called size bias coupling provides an alternative, and sometimes an improvement on the exchangeable pair coupling.

Let W be a nonnegative random variable, $\mathrm{Var}(W) = 1$, and $\mathbf{E}W = \mu > 0$. W^* is said to have the W-size biased distribution if, for all functions g for which the expectation exists,

$$\mathbf{E}Wg(W) = \mu \mathbf{E}g(W^*).$$

Thus, if W is discrete, say, then using $g(x) = \mathbf{1}_w(x)$ yields that for all w we have $\mathbf{P}(W^* = w) = \frac{w}{\mu}\mathbf{P}(W = w)$. This illustrates that size biasing corresponds to sampling proportional to size; the larger a subpopulation, the more likely it is to be in the sample.

If (W, W^*) are defined on the same probability space, with W^* having the W-size biased distribution, then with $g(x) = f'(x - \mu)$ the right-hand side of equation (3) becomes

$$\begin{aligned}
\mathbf{E}(W-\mu)f'(W-\mu) &= \mu\mathbf{E}(f'(W^*-\mu) - f'(W-\mu)) \\
&= \mu\mathbf{E}\left\{f''(W-\mu)(W^*-W)\right\} + R,
\end{aligned}$$

using Taylor expansion, where

$$|R| \le \mu \|f^{(3)}\| \mathbf{E}(W^* - W)^2.$$

Note that using $g(x) = x$ yields $\mu\mathbf{E}(W^* - W) = \mathbf{E}W^2 - \mu^2 = \mathrm{Var}(W) = 1$. Thus

$$\begin{aligned}
&\mu\mathbf{E}f''(W-\mu)(W^*-W) - \mathbf{E}f''(W-\mu) \\
&= \mathbf{E}f''(W-\mu)(\mu\mathbf{E}(W^*-W|W) - 1) \\
&\le \|f''\|\sqrt{\mathbf{E}(\mu\mathbf{E}(W^*-W|W)-1)^2} \\
&= \|f''\|\mu\sqrt{\mathrm{Var}\mathbf{E}(W^*-W|W)}.
\end{aligned}$$

This gives (see Goldstein and Rinott [44])

Theorem 3 *Let W be nonnegative, $\mathbf{E}W = \mu > 0$, and $\mathrm{Var}(W) = 1$. Let W^* have the W-size biased distribution. For any continuous, bounded function h with piecewise continuous, bounded first derivative, we have*

$$\begin{aligned}
&|\mathbf{E}h(W-\mu) - \Phi h| \\
&\le (\sup h - \inf h)\mu\sqrt{\mathrm{Var}\mathbf{E}(W^*-W|W)} + \|h'\|\mu\mathbf{E}(W^*-W)^2.
\end{aligned}$$

Example 2 *let $X_1, \ldots X_n$ be independent, nonnegative, $\mathbf{E}X_i = \mu_i > 0$, $W = \sum_{i=1}^n X_i$, $\mathbf{E}W = \mu$, $\mathrm{Var}(W) = 1$. Choose an index I from $\{1, \ldots, n\}$ according to*

$$\mathbf{P}(I = i) = \frac{\mu_i}{\mu}, \tag{8}$$

that is, choose index i proportionally to its expectation. If $I = i$, replace X_i by X_i^* having the X_i-size bias distribution, and put

$$W^* = W - X_I + X_I^*. \tag{9}$$

Then W^* has the W-size biased distribution. (Note the similarity to (7)). Using Theorem 3 gives

$$|\mathbf{E}h(W-\mu) - \Phi h| \quad \le \quad (\sup h - \inf h)\frac{\mu}{n} + \|h'\| \sum_{i=1}^n \mathbf{E}|X_i|^3.$$

In the usual scaling , $X_i \asymp \frac{1}{\sqrt{n}}$, we have $\mu \asymp \sqrt{n}$ and the bound is of order $\frac{1}{\sqrt{n}}$. Depending on $\sigma_i, i = 1, \ldots, n$, this bound may be better than the one obtained in Theorem 1 by the local approach.

Goldstein and Rinott [44] also give a general construction.

Construction 2 *Let $X_1, \ldots X_n$ be nonnegative, $\mathbf{E}X_i = \mu_i > 0$, $W = \sum_{i=1}^n X_i$, $\mathbf{E}W = \mu$, $\mathrm{Var}(W) = 1$. Choose an index I from $\{1, \ldots, n\}$ according to (8). If $I = i$, replace X_i by a variate X_i^* having the X_i-size bias distribution. If $X_i^* = x$, construct $\hat{X}_j, j \neq i$ such that*

$$\mathcal{L}(\hat{X}_j, j \neq i) = \mathcal{L}(X_j, j \neq i | X_i = x).$$

Then

$$W^* = \sum_{j \neq I} \hat{X}_j + X_I^*$$

has the W-size biased distribution. (The difference from Construction 1 are the choice of I and the distribution of X_i^*.)

1.1.5 Zero Bias Couplings

The size bias coupling above is only defined for nonnegative random variables. Random variables that could also assume negative values could be truncated and shifted, but this procedure does not appear very natural. Instead the so-called zero bias coupling is much better adapted to this situation.

Let W be mean zero, variance one. We say that a random variable W^* has the W-zero biased distribution if, for all g for which the expectation exists,

$$\mathbf{E}Wg(W) = \mathbf{E}g'(W^*). \tag{10}$$

This notion was introduced in Goldstein and Reinert [43]. Related analytical ideas, without coupling constructions, appear in Ho and Chen [47], Chen [30], and in Cacoullos *et al.* [26]. Bolthausen [24] uses a related coupling, without formalizing it. As the standard normal distribution is the unique fixed point of (10), it seems to be another natural approach for normal approximations.

Using (10), the right-hand side of Equation (3) can be written as

$$\begin{aligned}\mathbf{E}Wf'(W) - \mathbf{E}f'(W) &= \mathbf{E}(f''(W^*) - f''(W)) \\ &= R,\end{aligned}$$

where

$$|R| \le \|f^{(3)}\| \mathbf{E}|W^* - W|.$$

Thus we save a step in the Taylor expansion. Expanding one step further, though, gives the following, sharper result, see Goldstein and Reinert [43], where we use the bounds (4), as derivatives higher than second order occur.

Theorem 4 *Let W be mean zero, variance 1, and let W^* have the W-zero biased distribution. For any bounded, continuous function h with bounded derivatives up to order 4, we have*

$$|\mathbf{E}h(W) - \Phi h| \quad \le \quad \frac{1}{3}\|h^{(3)}\|\sqrt{\mathbf{E}(\mathbf{E}(W^* - W|W))^2} + \frac{1}{8}\|h^{(4)}\|\mathbf{E}(W - W^*)^2.$$

Example 3 *Let $X_1, \ldots X_n$ be independent, mean zero, $\mathrm{Var}(X_i) = \sigma_i^2$, $W = \sum_{i=1}^n X_i$, and assume $\mathrm{Var}(W) = 1$. Choose an index I from $\{1, \ldots, n\}$ according to*

$$\mathbf{P}(I = i) = \sigma_i^2,$$

that is, choose index i proportionally to its variance. (This resembles (8), where i is drawn proportionally to its expectation.) If $I = i$, replace X_i by X_i^* having the X_i-zero biased distribution, and put

$$W^* = W - X_I + X_I^*. \tag{11}$$

Then W^* has the W-zero biased distribution; note the similarity to (7) and (9). Using Theorem 4 gives

$$\begin{aligned}&|\mathbf{E}h(W - \mu) - \Phi h| \\ &\le \quad \frac{1}{3}\|h^{(3)}\|\sqrt{\mathbf{E}\left(\mathbf{E}X_I^3 - \frac{W}{n}\right)^2} + \frac{1}{8}\|h^{(4)}\|\left(\frac{\mathbf{E}X_I^4}{3} + \mathbf{E}X_I^2\right).\end{aligned}$$

In the special case that $\mathbf{E}X_i^3 = 0$ we obtain

$$|\mathbf{E}h(W - \mu) - \Phi h| \quad \le \quad \frac{1}{3}\|h^{(3)}\|\frac{1}{n} + \frac{1}{8}\|h^{(4)}\|\left(\frac{\mathbf{E}X_I^4}{3} + \mathbf{E}X_I^2\right). \tag{12}$$

With the scaling $X_i \asymp \frac{1}{\sqrt{n}}$ we thus obtain a $\frac{1}{n}$ - bound on the rate of convergence for smooth test functions.

Unfortunately, for sums of dependent random variables the construction is more involved. The following construction of such a coupling when W is the sum of the dependent variates $X_1, \ldots, X_n$ is given in Goldstein and Reinert [43].

Given the vector $\mathbf{X} = (X_1, \ldots, X_n)$ with distribution $dF(\mathbf{x})$, for every i one constructs a vector $\mathbf{X}_i = (X_1, \ldots, X_{i-1}, X_i', X_i'', X_{i+1}, \ldots, X_n)$ with distribution $dF_{n,i}$ such that $v_i^2 = E(X_i' - X_i'')^2 > 0$, and where removing either X_i' or X_i'' results in a vector with the original distribution $dF(\mathbf{x})$. Now consider an $n+1$ vector $\hat{\mathbf{X}}_i = (\hat{X}_1, \ldots, \hat{X}_{i-1}, \hat{X}_i', \hat{X}_i'', \hat{X}_{i+1}, \ldots, \hat{X}_n)$ with distribution

$$d\hat{F}_{n,i}(\hat{\mathbf{x}}_i) = \frac{(\hat{x}_i' - \hat{x}_i'')^2}{v_i^2} dF_{n,i}(\hat{\mathbf{x}}_i).$$

Suppose that there is a ρ such that for all f for which $\mathbf{E}Wf(W)$ exists ,

$$\sum_{i=1}^n \mathbf{E}X_i' f(W_i + X_i'') \quad = \quad \rho \mathbf{E}Wf(W),$$

where $W_i = W - X_i$. Let U be an independent uniform variate on $[0,1]$, and set

$$\mathbf{X}_{ii}^* = (\hat{X}_1, \ldots, \hat{X}_{i-1}, U\hat{X}_i' + (1-U)\hat{X}_i'', \hat{X}_{i+1}, \ldots, \hat{X}_n), \tag{13}$$

and

$$W_i^* = U\hat{X}_i' + (1-U)\hat{X}_i'' + \sum_{j \neq i} \hat{X}_j.$$

With I an independent random index taking on the value i with probability proportional to v_i^2, it is shown in Goldstein and Reinert [43] that the mixture

$$W^* = W_I^* \tag{14}$$

has the W-zero bias distribution. This construction agrees with the one given previously when the X-variates are independent.

This particular coupling of W and W^* requires the construction of a vector of $n+1$ elements for each of the n variates of $\mathbf{X}$, which may be difficult to achieve in certain cases.

1.2 Stein's Method in General

The general procedure is: Find a good characterization of the desired distribution in terms of an equation, that is of the type

$$\mathcal{L}(X) = \mu \iff \mathbf{E}\mathcal{A}f(X) = 0, \text{ for all smooth functions } f,$$

where $\mathcal{A}$ is an operator associated with the distribution μ. (Thus, in the standard normal case, $\mathcal{A}f(x) = f'(x) - xf(x)$.) Then assume X to have distribution μ, and consider the Stein equation

$$h(x) - \mathbf{E}h(X) = \mathcal{A}f(x), \quad x \in \mathbf{R}. \tag{15}$$

For every smooth h, find a corresponding solution f of this equation. For any random element W,

$$\mathbf{E}h(W) - \mathbf{E}h(X) = \mathbf{E}\mathcal{A}f(W).$$

Hence, to estimate the proximity of W and X, it is sufficient to estimate $\mathbf{E}\mathcal{A}f(W)$ for all possible solutions f of (15).

However, in this procedure it is not completely clear which characterizing equation for the distribution to choose (one could think of a whole set of possible equations). The aim is to be able to solve (15) for a sufficiently large class of functions g, to obtain convergence in a known topology, and to obtain rates of convergence in a known metric.

1.2.1 The Generator Method

A key tool for solving the higher-dimensional case is the generator method developed by Barbour [10], [11], [12]. Observe that the left-hand side of (1) can be written as $Af(x)$, where A is the generator of an Ornstein-Uhlenbeck process. Thus semigroup theory can be applied to solve the generator equation. Note that the target distribution, here the standard normal, is the stationary distribution of this Markov process.

In general, Barbour [11] suggested employing as operator $\mathcal{A}$ in equation (15) the generator of a Markov process, which then provides a way to look for solutions of (15). This is what in the following will be called the generator method. Suppose we can find a Markov process $(X(t))_{t\geq 0}$ with generator $\mathcal{A}$ and unique stationary distribution μ, such that $\mathcal{L}(X(t)) \stackrel{w}{\Rightarrow} \mu \ (t \to \infty)$. Then, if a random variable X has distribution μ,

$$\mathbf{E}\mathcal{A}f(X) = 0$$

for all $f \in \mathcal{D}(\mathcal{A})$. Now a method for solving equation (15) is provided by Proposition 1.5 of Ethier and Kurtz ([38], p. 9; for the argument, see Barbour [11]). Let $(T_t)_{t\geq 0}$ be the transition semigroup of the Markov process $(X(t))_{t\geq 0}$. Then

$$T_t h - h = \mathcal{A}\left(\int_0^t T_u h\, du\right).$$

As $(T_t)_{t\geq 0}$ is a strongly continuous contraction semigroup, $\mathcal{A}$ is closed (Ethier and Kurtz [38], Corollary 1.6), and we could formally take limits:

$$h(x) - \mathbf{E}h(X) = -\mathcal{A}\left(\int_0^\infty T_u h\, du\right).$$

Thus $f = -\int_0^\infty T_u h\, du$ would be a solution of (15), if this expression exists and if $f \in \mathcal{D}(\mathcal{A})$. This will in general be the case only for a certain class of functions g. However, the latter conditions can usually be checked.

This generator method has proved to be very useful for convergence towards Wiener measure (Barbour [11]) and for Poisson process approximations (see, e. g., Barbour *et al.* [19]).

However, for a given distribution μ, there may be various Markov processes with μ as stationary distribution, and it is still not completely clarified which process to take to obtain good results (though many persons have a good intuition on it).

1.2.2 Further Examples

Stein's method has been generalized to many other distributions, foremost the Poisson distribution (see Chen [27], Arratia *et al.* [2], Barbour *et al.* [19], Aldous [4], to cite but a few). Other distributions include the uniform distribution (Diaconis [33]), the binomial distribution (Ehm [39]), the compound Poisson distribution (Barbour *et al.* [13], Barbour and Utev [21], Roos [69]), the multinomial distribution (Loh [52]), the gamma distribution (Luk [54]; for the χ^2 distribution see also Mann [55]), the geometric distribution (Peköz [56]) and, more generally, Pearson curves (Diaconis and Zabell [34], Loh [53]).

1.3 Stein's Method for the Law of Large Numbers

For the law of large numbers, Stein's method has been worked out by Reinert [61]. Although the weak law of large numbers has been very well studied by many methods, for the real-valued case, we give the Stein approach here to motivate the weak law of large numbers for empirical measures in Section 2.

1.3.1 Stein's Equation and the Generator Method

Intuitively, point mass can be seen as an extreme case of the normal distribution with zero variance. Hence we put

$$(\mathcal{A}f)(x) = -xf'(x), \quad x \in \mathbf{R}.$$

Then $\mathcal{A}$ is a good candidate for the weak law of large numbers - generator. Note that $\mathcal{A}$ is the generator of the deterministic Markov process $(Y(t))_{t\geq 0}$ that is given by

$$\mathbf{P}[Y(t) = xe^{-t} \mid Y(0) = x] = 1, \quad x \in \mathbf{R}.$$

The corresponding transition semigroup is

$$T_t h(x) = h(xe^{-t}),$$

and the unique stationary distribution is δ_0.
According to the general equation (15), the Stein equation in this context is

$$h(x) - h(0) = -xf'(x), \quad x \in \mathbf{R}. \tag{16}$$

Let $C_b^2(\mathbf{R})$ be the space of all bounded, twice continuously differentiable real-valued functions on $\mathbf{R}$ with bounded first and second derivatives, and let $D_b^2(\mathbf{R})$

be the space of all twice continuously differentiable functions $f : \mathbf{R} \to \mathbf{R}$ with bounded first and second derivatives. Using the semigroup approach the following proposition is easy to derive.

Proposition 1 *For any $h \in C_b^2(\mathbf{R})$, there is a function $f = \phi(h) \in D_b^2(\mathbf{R})$ that solves the Stein equation (16) for h. Furthermore, for the derivatives, $\|f'\| \leq \|h'\|$, and $\|f''\| \leq \|h''\|$.*

Now we have all the ingredients to derive weak laws of large numbers.

Theorem 5 *Let $(X_i)_{i\in\mathbf{N}}$ be a family of random variables on $\mathbf{R}$, defined on the same probability space, with finite variances. Put*

$$Y_n = \sum_{i=1}^{n} (X_i - \mathbf{E}X_i).$$

Then, for all $h \in C_b^2(\mathbf{R})$

$$|\mathbf{E}h(Y_n) - h(0)| \leq \|h''\| \mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big).$$

As $C_b^2(\mathbf{R})$ is convergence-determining for weak convergence of the laws of real-valued random variables, a weak law of large numbers follows from Theorem 5 provided that

$$\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) \to 0 \quad (n \to \infty).$$

Proof of Theorem 5 Suppose without loss of generality that $\mathbf{E}X_i = 0$ for all $i \in \mathbf{N}$. Let $h \in C_b^2(\mathbf{R})$. As the solution $f = \phi(h)$ of the Stein equation (16) satisfies $f \in D_b^2(\mathbf{R})$, it is enough to bound $\mathbf{E}[\mathcal{A}f(Y_n)]$ for all $f \in D_b^2(\mathbf{R})$. For such f, we have

$$\begin{aligned}
&\mathbf{E}\mathcal{A}f(Y_n) \\
&= -\mathbf{E}\Big(\sum_{i=1}^{n} X_i\Big) f'\Big(\sum_{i=1}^{n} X_i\Big) \\
&= -\mathbf{E}\Big(\sum_{i=1}^{n} X_i\Big) f'(0) - \mathbf{E}\int_0^{\sum_{i=1}^n X_i} \left(\sum_{i=1}^{n} X_i - t\right)^2 f''(t)dt,
\end{aligned}$$

where we used Taylor's expansion. As we assumed the X_i to have zero mean, we hence get

$$\mathbf{E}\mathcal{A}f(Y_n) = -\mathbf{E}\int_0^{\sum_{i=1}^n X_i} \left(\sum_{i=1}^{n} X_i - t\right)^2 f''(t)dt$$

Thus

$$|\mathbf{E}\mathcal{A}f(Y_n)| \quad \leq \quad \|f''\|\mathbf{E}\Big[\Big(\sum_{i=1}^{n} X_i\Big)^2\Big].$$

This proves the theorem.

1.3.2 The Local Approach

The method for proving Theorem 5 can be generalized to local dependence. Suppose $(X_i)_{i\in I}$ is a family of random variables with finite first and second moments. In addition assume that for each $i \in I$ there is a set (neighborhood) $B_i \subset I$ such that X_i is independent of $\sigma(X_j, j \notin B_i)$. Then, using Taylor expansion, we have the following proposition.

Proposition 2 *We have that*

$$|\mathbf{E}h(W) - h(\mu)| \leq \|h''\| \sum_{i\in I} \sum_{j\in B_i} \mathbf{E}|(X_i - \mathbf{E}X_i)(X_j - \mathbf{E}X_j)|.$$

This result can be generalized to allow for weak dependence between X_i and $\sigma(X_j, j \notin B_i)$; we will encounter related theorems later.

1.3.3 Exchangeable Pair Couplings

Theorem 5 can also be obtained using exchangeable pairs. Let W have mean zero, and suppose (W, W') is an exchangeable pair satisfying (5). Recall (6)

$$\mathbf{E}Wf(W) = \frac{1}{2\lambda}\mathbf{E}(W - W')(f(W) - f(W')),$$

giving in particular that

$$\mathbf{E}(W - W')^2 = 2\lambda \mathrm{Var}(W).$$

With Taylor expansion we thus recover Theorem 5. In particular, if W is the sum of n i.i.d. random variables, then $\lambda = \frac{1}{n}$ as before. This result may be generalized to cases of global weak dependence.

1.3.4 Size Bias Couplings

The size bias coupling gives a slightly differently flavoured result. Let W be nonnegative with positive mean, and let W^s have the W-size bias distribution. Then, for all smooth f,

$$\mathbf{E}(W - \mu)f'(W) = \mu\mathbf{E}(f'(W) - f'(W^*)),$$

so that we obtain the following proposition.

Proposition 3 *For all smooth functions g,*

$$|\mathbf{E}h(W) - h(\mu)| \le \|h''\| \mu \mathbf{E}|W - W^*|.$$

In particular, if W is the sum of n i.i.d. random variables $X_1, X_2, \ldots, X_n$ having the same distribution as X, then $\mathbf{E}|W - W^*| = \mathbf{E}|X_I - X_I^*| \le \sqrt{\mathrm{Var}(X)} - 2(\sigma^2 + \mu^2) + \frac{1}{\mu}\mathbf{E}X^3$.

1.3.5 Zero Bias Couplings

Using the zero bias coupling, Theorem 5 is immediate. For Y having mean zero and finite variance, a random variable Y^* having the Y-zero bias distribution exists, and satisfies

$$\mathbf{E}Yg(Y) = \mathrm{Var}(Y)\mathbf{E}g'(Y^*)$$

for all functions g for which both sides of the equation exist. Putting $Y = W - \mu$ we have

$$\begin{aligned}
\mathbf{E}(W-\mu)f'(W) &= \mathbf{E}Yf'(Y+\mu) \\
&= \mathrm{Var}(Y)\mathbf{E}f''(Y^* + \mu) \\
&= \mathrm{Var}(W)\mathbf{E}f''(Y^* + \mu).
\end{aligned}$$

Thus we immediately obtain Theorem 5.

1.4 Stein's Method for Poisson Approximation

In the context of Poisson approximation Stein's method is called the Chen-Stein method, in honor of Louis Chen [27], who in 1975 published this method as part of his Ph.D. thesis under Charles Stein. It phrases the approximation in terms of the total variation distance. A friendly exposition is in Arratia *et al.* [2] and a description with many examples can be found in Arratia *et al.* [3] and Barbour *et al.* [19].

For any two probability measures μ_1 and μ_2 on the same measurable space E, the total variation distance is defined to be

$$\begin{aligned}
d_{TV}(\mu_1, \mu_2) &= \sup_{B \subset E, measurable} |\mu_1(B) - \mu_2(B)| \\
&= \sup_{h: E \to [0,1], measurable} \left| \int h d\mu_1 - \int h d\mu_2 \right|.
\end{aligned}$$

(Note that there are two different definitions of the total variation distance in use, differing from each other by a factor of 2; the definition of the total variation distance in Arratia *et al.* [2] is twice as large as the one used here.)

For the Poisson distribution with parameter λ the corresponding Stein equation is

$$h(x) - Po(\lambda)h = \lambda f(x+1) - xf(x). \tag{17}$$

It is easy to show that for any indicator function h the Stein equation (17) has a solution f; Barbour *et al.* [19] show that this solution satisfies

$$\begin{aligned} \|f\| &= \sup_{x \in \mathbf{Z}^+} |f(x)| \le \min(1, \lambda^{-1/2}) \\ \Delta f &= \sup_{x \in \mathbf{Z}^+} |f(x+1) - f(x)| \le \min(1, \lambda^{-1}). \end{aligned} \tag{18}$$

To see how the Stein equation (17) fits into the generator approach, write $f(x) = g(x) - g(x-1)$. Then the operator $\mathcal{A}f(x) = \lambda f(x+1) - xf(x)$ in equation (17) becomes

$$\mathcal{A}g(x) = \lambda g(x+1) + xg(x-1) = (\lambda + x)g(x).$$

This is the generator of an immigration-death process with immigration rate λ and unit per-capita death rate.

Example 4 *Let $I_1, \dots, I_n$ be i.i.d. Bernoulli indicator random variables with $\mathbf{E}I_i = p_i, i = 1, \dots, n$, and put $W = \sum_{i=1}^n I_i$. Let $\lambda = \sum_{i=1}^n p_i = \mathbf{E}W$. Put $W_i = W - X_i$. Then we have for any function f*

$$\begin{aligned} \mathbf{E}Wf(W) &= \sum_{i=1}^n \mathbf{E}X_i f(W) \\ &= \sum_{i=1}^n \mathbf{E}I_i f(W_i + I_i) \\ &= \sum_{i=1}^n p_i \mathbf{E}f(W_i + 1), \end{aligned}$$

where we used the independence. Thus

$$\begin{aligned} \lambda \mathbf{E}f(W+1) - \mathbf{E}Wf(W) &= \sum_{i=1}^n p_i \mathbf{E}(f(W+1) - f(W_i+1)) \\ &= \sum_{i=1}^n p_i^2 \mathbf{E}(f(W_i+2) - f(W_i+1)), \end{aligned}$$

where we conditioned on $I_i = 1$. Employing the bounds (18) we obtain that

$$d_{TV}(\mathcal{L}(W), Po(\lambda)) \le \min(1, \lambda^{-1}) \sum_{i=1}^n p_i^2.$$

1.4.1 The Local Approach

Taylor expansion leads to a very convenient result for proving Poisson approximations.

Theorem 6 (Chen [27], Arratia *et al.* **[3], Barbour** *et al.* **[19])** *Take any index set I. For each $\alpha \in I$, let I_α be a Bernoulli random variable with $\pi_\alpha =$*

$\mathbb{P}(I_\alpha = 1) > 0$. *Suppose that, for each* $\alpha \in I$, *we have chosen* $B_\alpha \subset I$ *with* $\alpha \in B_\alpha$. *Let* $W = \sum_{\alpha \in I} I_\alpha$ *and* $\lambda = \sum_{\alpha \in I} \pi_\alpha < \infty$. *Then*

$$d_{TV}(\mathcal{L}(W), \mathrm{Po}(\lambda)) \leq \min\left(1, \lambda^{-1}\right)(b_1 + b_2) + \min\left(1, \lambda^{-\frac{1}{2}}\right) b_3,$$

where

$$\begin{aligned} b_1 &= \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} \pi_\alpha \pi_\beta \\ b_2 &= \sum_{\alpha \in I} \sum_{\beta \in B_\alpha, \beta \neq \alpha} \mathbf{E}(I_\alpha I_\beta) \\ b_3 &= \sum_{\alpha \in I} \mathbf{E}\left|\mathbf{E}\left\{I_\alpha - \pi_\alpha | \sigma(I_\beta, \beta \notin B_\alpha)\right\}\right|. \end{aligned}$$

Note that $b_3 = 0$ if I_α is independent of $\sigma(I_\beta, \beta \notin B_\alpha\}$. We think of B_α as a neighborhood of strong dependence of I_α.

1.4.2 Size Bias Couplings

Size biasing in the Poisson context has been discussed in detail in Barbour *et al.* [19], where it is called the *coupling approach*. Observe that the size bias distribution of a nontrivial indicator random variable is point mass at 1. Recalling Construction 2, for $W = \sum_{i=1}^n I_i$ being the sum of indicators with $\mathbf{E}I_i = p_i$, a random variable W^s having the W-size bias distribution can be constructed as follows.

Choose an index I from $\{1, \ldots, n\}$ according to $\mathbf{P}(I = i) = p_i$. If $I = i$, replace I_i by 1, and construct $\hat{I}_j, j \neq i$ such that

$$\mathcal{L}(\hat{I}_j, j \neq i) = \mathcal{L}(I_j, j \neq i | I_i = 1).$$

Then

$$W^s = \sum_{j \neq I} \hat{I}_j + 1.$$

has the W-size biased distribution. If $I = i$, put

$$\hat{W}_i = \sum_{j \neq i} \hat{I}_j.$$

Then Barbour *et al.* [19] prove (Theorem 1.B in Barbour *et al.* [19], with different notation) the following proposition.

Proposition 4 *We have*

$$d_{TV}(\mathcal{L}(W), Po(\lambda)) \leq \min(1, \lambda^{-1}) \sum_{i=1}^n p_i \mathbf{E}|W - \hat{W}_i|.$$

2 Stein's Method for Measure-Valued Random Elements

Now we turn our attention ot measure-valued random elements. To make use of the bounds on convergence that Stein's method provides we need to construct a metric on the space of measures that is defined in terms of smooth test functions. Most of what follows can be found in Reinert [61], [62].

2.1 The Space of Measures

Let E be a locally compact Hausdorff space with a countable basis (for instance, $E = \mathbf{R}^d$), let $\mathcal{E} = \mathcal{B}(E)$ be the Borel-σ-field of E, and let $M^{\mathrm{b}}(E)$ the space of all bounded Radon measures on E, equipped with the vague topology. Let $C_c(E)$ be the space of real-valued continuous functions on E with support contained in a compact set. Convergence in the vague topology means the following.

$$\nu_n \Rightarrow_v \nu \iff \text{for all } f \in C_c(E) : \int_E f d\nu_n \to \int_E f d\nu \quad (n \to \infty).$$

For $\mu \in M^{\mathrm{b}}(E)$, set $\|\mu\| = \sup_{A \in \mathcal{E}} |\mu(A)|$. Let

$$M^f(E) = \{\mu \in M^{\mathrm{b}}(E) : \|\mu\| < \infty\}$$

be the space of finite Radon measures, and let

$$M_1(E) = \{\mu \in M^{\mathrm{b}}(E) : \mu \text{ positive}, \|\mu\| \leq 1\}$$

be the space of all positive Radon measures with total mass smaller or equal to 1. The space $M^f(E)$ will be needed for Gaussian approximations, whereas the space $M_1(E)$ is sufficient to describe laws of large numbers.

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 (see Reinert [64]). Furthermore, $M^{\mathrm{b}}(E)$ is a topological linear space over $\mathbf{R}$. Denote by $D_{\mathcal{G}}(A; G)$ for the set of all Gâteaux-differentiable functions $f : A \to G$, and denote by $f'(a)[\mu]$ the Gâteaux derivative of f at the point a in direction μ. In general, denote by $f^{(k)}(a)[\nu^{(k)}]$ the kth derivative of f in a, as a linear form, applied to the vector $\nu^{(k)} = (\nu, \dots, \nu) \in (M^{\mathrm{b}}(E))^k$; and let $D^k_{\mathcal{G}}(A; G)$ denote the set of all k times Gâteaux-differentiable functions $f : A \to G$. For $f : A \to \mathbf{R}$, we have Taylor's theorem. We need some more notation. Put

$$\begin{aligned}
\|f'(\nu)\| &= \sup\{|f'(\nu)[\eta]| : \eta \in M^{\mathrm{b}}(E), \|\eta\| \leq 1\}, \\
\|f''(\nu)\| &= \sup\{|f''(\nu)[\eta, \mu]| : \eta, \mu \in M^{\mathrm{b}}(E), \|\eta\| \leq 1, \|\mu\| \leq 1\}; \\
\|f'\| &= \sup_{\nu \in M^{\mathrm{b}}(E)} \|f'(\nu)\|, \\
\|f''\| &= \sup_{\nu \in M^{\mathrm{b}}(E)} \|f''(\nu)\|.
\end{aligned}$$

Moreover we abbreviate the integral

$$\langle \phi, \mu \rangle = \int_E \phi d\mu.$$

Vague convergence can be described via Lipschitz functions. Fix $a \in E$ and define $S_m = \{x : d(x,a) < m\}, m = 1, 2, \ldots$. For $m = 1, 2, \ldots$, put

$$\begin{aligned} \mathcal{FL}_m(E) \quad = \{ \quad & f : E \to \mathbf{R}; |f(x) - f(y)| \leq k_f d(x,y), x, y \in E, \\ & \text{for some constant } k_f \text{ depending only on } f; f(x) = 0, x \notin S_m\}, \end{aligned}$$

and

$$\mathcal{FL}(E) = \cup_m \mathcal{FL}_m(E).$$

For $(\nu_n)_{n\in\mathbf{N}}, \nu \in M(E)$,

$$\nu_n \Rightarrow_v \nu \text{ iff for all } f \in \mathcal{FL}(E), \langle \nu_n, f \rangle \to \langle \nu, f \rangle.$$

This is shown in Rachev [59], p. 209, Corollary 10.2.1 and equation (10.2.5), for nonnegative measures; using the Hahn decomposition theorem, it is easily extended to signed measures. Note that the definition of $\mathcal{FL}_m(E)$ differs slightly to the one in Rachev [59] in not requiring the Lipschitz constant to be 1; our definition ensures that $\mathcal{FL}(E)$ is closed under multiplication with scalars.

Moreover, the space $\mathcal{FL}^+(E)$ of nonnegative functions in $\mathcal{FL}(E)$ is also convergence-determining for vague convergence in $M^{\mathrm{b}}(E)$; this can be seen by decomposing functions into their positive and negative part.

Let $\mathcal{C}$ be a convergence-determining class for vague convergence on $M^f(E)$. For simplicity, from here on we will assume that $\mathcal{C} = \mathcal{FL}(E)$, $\mathcal{C} = \mathcal{FL}^+(E)$, $\mathcal{C} = C_c(E)$ or $\mathcal{C} = {C_c}^+(E)$, the latter being the space of all nonnegative functions in $C_c(E)$. Put

$$\begin{aligned} \mathcal{F}_{\mathcal{C}}(M^f(E)) \quad := \quad & \{F \in C_b(M^f(E)) : F \text{ has the form} \\ & F(\mu) = f(\langle \mu, \phi_1 \rangle, \ldots, \langle \mu, \phi_m \rangle) \\ & \text{for an } m \in \mathbf{N}, f \in C_b^\infty(\mathbf{R}^m) \\ & \text{with } \|f'\| \leq 1, \|f''\| \leq 1, \|f'''\| \leq 1, \\ & \text{and for } \phi_i \in \mathcal{C} \text{ with } \|\phi_i\| \leq 1, i = 1, \ldots, m\}. \end{aligned} \tag{19}$$

Similarly we define $\mathcal{F}_{\mathcal{C}}(M_1(E))$ by replacing $M^f(E)$ with $M_1(E)$ in (19). This construction is similar to the algebra of polynomials used by Dawson, see for example Dawson [31]. We abbreviate $\mathcal{F}_{\mathcal{C},f} = \mathcal{F}_{\mathcal{C}}(M^f(E))$, and $\mathcal{F}_{\mathcal{C}} = \mathcal{F}_{\mathcal{C}}(M_1(E))$.

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 $|f(x)| < \epsilon$). The following proposition can be shown by the Stone-Weierstrass Theorem.

Proposition 5 *$C_{\mathcal{C}}(M^f(E))$ is dense in $C_0(M^f(E))$ with respect to the topology of uniform convergence.*

Now let $\mathcal{F}$ be a dense subset of $C_0(M^f(E))$. The following proposition shows that $\mathcal{F}$ is convergence-determining.

Proposition 6 *Let $\mathcal{U}$ be a locally compact Hausdorff space with a countable basis, and let $\mathcal{F} \subset C_0(\mathcal{U})$ be a class of functions that is dense in $C_0(\mathcal{U})$ with respect to the norm of uniform convergence. Let $(\nu_n)_{n\in\mathbf{N}}, \nu \in M_1(\mathcal{U})$. If, for all $f \in \mathcal{F}$,*

$$\langle \nu_n, f\rangle \to \langle \nu, f\rangle \quad (n \to \infty),$$

then

$$\nu_n \Rightarrow_v \nu \quad (n \to \infty).$$

If $(\nu_n)_{n\in\mathbf{N}}, \nu \in M_1(\mathcal{U})$ are laws of finite random measures, we obtain weak convergence:

Corollary 1 *Let $(\xi_n)_{n\in\mathbf{N}}, \xi$ be random measures taking values in $M^f(E)$ almost surely, where E is a locally compact Hausdorff space with a countable basis. Let $\mathcal{F} \subset C_0(M^f(E))$ be a class of functions that is dense in $C_0(M^f(E))$ with respect to the norm of uniform convergence. Suppose that for all $f \in \mathcal{F}$,*

$$\langle \mathcal{L}(\xi_n), f\rangle \to \langle \mathcal{L}(\xi), f\rangle \quad (n \to \infty).$$

Then

$$\mathcal{L}(\xi_n) \overset{w}{\Rightarrow} \mathcal{L}(\xi) \quad (n \to \infty).$$

The proof of Corollary 1 is immediate.

The above considerations justify introducing the following metric. In general, if $(\mathcal{U}, d)$ is a separable metric space, and if $C_b(\mathcal{U})$ is the set of all bounded continuous functions on $\mathcal{U}$, then, for each subset $\mathcal{F}$ of $C_b(\mathcal{U})$, the functional $\zeta_\mathcal{F}$ on $M^f(\mathcal{U}) \times M^f(\mathcal{U})$ given by

$$\zeta_\mathcal{F}(\nu, \eta) = \sup_{f\in\mathcal{F}} \left|\langle \nu, f\rangle - \langle \eta, f\rangle\right|$$

defines a semimetric on $M^f(\mathcal{U})$ (see Rachev [59], p.72) and is called the Zolotarev $\zeta_\mathcal{F}$-metric. (Rachev [59] uses this definition only for probability measures, but the notion can obviously be generalized for measures in $M^f(\mathcal{U})$.) Here $\zeta_\mathcal{F}$ is actually a metric.

Proposition 7 *Let $\mathcal{U}$ be a separable metric space and $\mathcal{F} \subset C_b(\mathcal{U})$ a class of functions that is convergence-determining for vague convergence in $\mathcal{U}$. Then $\zeta_\mathcal{F}$ is a metric on $M^f(\mathcal{U})$.*

Proof. We only have to show that, if $\zeta(\nu, \eta) = 0$, then $\nu = \eta$. As $\mathcal{F}$ is convergence-determining, it is also measure-determining. Thus, for all $\nu \neq \mu \in M^f(\mathcal{U})$, there is an $f \in \mathcal{F}$ such that $\langle \nu, f\rangle \neq \langle \eta, f\rangle$. Therefore, if $\nu \neq \eta$, then $\zeta_\mathcal{F}(\nu, \eta) \neq 0$. This proves the assertion.

Note that the above results remain true of $M^f(E)$ is replaced by $M_1(E)$.

2.2 Couplings for Random Measures

Similarly as in the real-valued case, we can define exchangeable pairs as well as size bias couplings and zero bias couplings for measure-valued random elements.

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

$$\mathbf{P}(\xi \in B, \xi' \in B') = \mathbf{P}(\xi \in B', \xi' \in B).$$

Similarly to Condition (5) we define Condition (20) by requiring that there is a $0 < \lambda < 1$ such that for all $\phi \in \mathcal{C}$

$$\mathbf{E}(\langle \xi', \phi \rangle | \langle \xi, \phi \rangle) = (1 - \lambda) \langle \xi, \phi \rangle. \tag{20}$$

For example, if $\xi = \sum_{i=1}^n \delta_{X_i}$, with $X_1, \dots, X_n$ being i.i.d, then it is easy to check that, just like in the real-valued case, picking an index I uniformly form $\{1, \dots, n\}$, replacing X_I by an independent copy X_I' and putting $\xi' = \xi - \delta_{X_I} + \delta_{X_I'}$ gives an exchangeable pair, and Condition (20) is satisfied with $\lambda = \frac{1}{n}$.

Size-biasing can also be defined for random measures. In this case the size bias measure is well-known as the Palm measure. Let η be a nonnegative random measure, let $\mathcal{C}$ be a convergence-determining class on E, let $\phi \in \mathcal{C}$ with $\mathbf{E}\langle \eta, \phi \rangle > 0$ and $\phi \geq 0$. We say that η^ϕ has the η-size biased distribution in direction ϕ if, for all f for which the expectations exist,

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

Moreover we will define zero-biasing for random measures, as indicated in Goldstein and Reinert [43]. Firstly we introduce zero-biasing for vectors of random variables. Given a mean zero vector $\mathbf{X} = (X_1, \dots, X_p)$ with covariance matrix Σ, we say the collection of vectors $\mathbf{X}^* = (\mathbf{X}_{ij}^*)$ has the $\mathbf{X}$-zero bias distribution if

$$\mathbf{E} \sum_{i=1}^p X_i f_i(\mathbf{X}) = \mathbf{E} \sum_{i,j=1}^p \sigma_{ij} f_{ij}(\mathbf{X}_{ij}^*), \tag{21}$$

for all smooth f. Here f_i and f_{ij} denote the partial coordinates of f with respect to the indicated coordinates.

Now let $\mathcal{H}$ be the class of functions ϕ from $\mathcal{E}$ to the reals such that, with $\langle \mu, \phi \rangle = \int \phi d\mu$, we have $\mathbf{E}\langle \xi, \phi \rangle = 0$ and $0 < \mathbf{E}\langle \xi, \phi \rangle^2 < \infty$. Given a collection $\{\xi(\phi), \phi \in \mathcal{H}\}$ of real valued mean zero random variables with nontrivial finite second moment, we say the collection $\{\xi_{\phi\psi}^*, \phi, \psi \in \mathcal{H}\}$ has the ξ-zero biased distribution if for all $p = 1, 2, \dots$ and $(\phi_1, \phi_2, \dots, \phi_p) \in \mathcal{H}^p$, the collection of p-vectors $(\mathbf{X}_{ij}^*)$ has the $\mathbf{X}$-zero bias distribution, where, for $1 \leq i, j \leq p$,

$$(\mathbf{X}_{ij}^*) = (\xi_{\phi_i \phi_j}^*(\phi_1), \dots, \xi_{\phi_i \phi_j}^*(\phi_p)),$$

and

$$\mathbf{X} = (\xi(\phi_1), \ldots, \xi(\phi_p)).$$

Note that, by choosing other sets $\mathcal{H}$, this definition can be extended to include the case of real-valued random variables, as well as the case of random processes.

2.3 The Law of Large Numbers for Measure-Valued Random Elements

Throughout this section, let $(X_i)_{i\in\mathbf{N}}$ be a family of random elements on E, defined on the same probability space, let $\mu_i = \mathcal{L}(X_i), i \in \mathbf{N}$, put $\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^n \mu_i$, and assume that

$$\text{there is a } \mu \in M_1(E) \text{ such that } \bar{\mu}_n \Rightarrow_v \mu \quad (n \to \infty).$$

Let $\xi_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ be the empirical measure of $(X_1, \ldots, X_n)$.

Definition 1 *We say that the weak law of large numbers for empirical measures holds if*

$$\mathcal{L}(\xi_n) \overset{w}{\Rightarrow} \delta_\mu \quad (n \to \infty).$$

This definition may require some remarks.

1. $\mathcal{L}(\xi_n) \overset{w}{\Rightarrow} \delta_\mu \quad (n \to \infty)$ means that, for every $f \in C_c(M_1(E))$,

$$\int f(\nu)\mathbf{P}[\xi_n \in d\nu] \to \int f(\nu)\delta_\mu(d\nu) = f(\mu) \quad (n \to \infty).$$

2. The name "weak law of large numbers" is based on the following fact (see Dudley [36], p.305, Proposition 11.1.3., e.g.)

   If $(\mathcal{S}, d)$ is a metric space, $p \in \mathcal{S}$, and $(Y_i)_{i\in\mathbf{N}}$ is a family of random elements on E, defined on the same probability space, with $\mathcal{L}(Y_n) \overset{w}{\Rightarrow} \delta_p \quad (n \to \infty)$, then $Y_n \overset{\mathbf{P}}{\to} p$. As $M_1(E)$ is Polish, we can find a metric d on $M_1(E)$ to make it a metric space. The weak law of large numbers holds iff, for all $\epsilon > 0$,

$$\mathbf{P}[d(\mathcal{L}(\xi_n), \delta_\mu) \geq \epsilon] \to 0 \quad (n \to \infty).$$

   Thus the situation is analogous to the real-valued case.

3. Take $\mathcal{C}$ to be a convergence-determining class for the vague convergence on E. Then the weak law of large numbers holds iff

$$\mathbf{E}G(\xi_n) \to G(\mu) \quad (n \to \infty) \text{ for all } G \in C_{\mathcal{C}}(M_1(E)).$$

This is equivalent to having for all $m \in \mathbf{N}$, for all $f \in C_b^\infty(\mathbf{R}^m)$, and for all $\phi_1, \ldots, \phi_m \in \mathcal{C}$,

$$\mathbf{E}[f(\langle \xi_n, \phi_1 \rangle, \ldots, \langle \xi_n, \phi_m \rangle)] \to f(\langle \mu, \phi_1 \rangle, \ldots, \langle \mu, \phi_m \rangle) \quad (n \to \infty);$$

that is,

$$\begin{aligned} &\mathbf{E} f\left(\sum_{i=1}^n \phi_1(X_i), \ldots, \sum_{i=1}^n \phi_m(X_i) \right) \\ &\quad \to \quad f(\langle \mu, \phi_1 \rangle, \ldots, \langle \mu, \phi_m \rangle) \quad (n \to \infty). \end{aligned}$$

To describe the connection with the classical weak law of large numbers for random variables, we thus have, for $n \to \infty$,

$$\mathcal{L}(\xi_n) \stackrel{w}{\Rightarrow} \delta_\mu \iff \sum_{i=1}^n \phi(X_i) \stackrel{\mathbf{P}}{\to} \int \phi d\mu \text{ for all } \phi \in \mathcal{C}.$$

Observe that now $(\phi(X_i))_{i \in \mathbf{N}}$ are real-valued random variables. Therefore we could prove the weak law of large numbers for empirical measures to hold by proving the weak law of large numbers for random variables to hold for all $(\phi(X_i))_{i \in \mathbf{N}}, \phi \in \mathcal{C}$. From this viewpoint, the standing assumption of having a measure μ with $\bar{\mu}_n \stackrel{w}{\Rightarrow} \mu \quad (n \to \infty)$ means that, for all $\phi \in \mathcal{C}$,

$$\sum_{i=1}^n \mathbf{E}\phi(X_i) \to \int \phi d\mu.$$

This is the classical assumption on the average of the expectations in the weak law of large numbers for the real-valued case. However, we would like to have a result that takes the structure of the space of empirical measures more into account, and that also describes convergence to δ_μ for more general random measures.

The corresponding Stein equation for the weak law of large numbers for random measures is

$$h(\nu) - \langle \delta_\mu, h \rangle = f'(\nu)[\mu - \nu], \quad \nu \in M_1(E). \tag{22}$$

This equation can easily be solved using the generator method, see Reinert [62].

Proposition 8 *For any function $H \in \mathcal{F}_\mathcal{C}$, there is a function $\psi(H) \in \mathcal{F}_\mathcal{C}$ that solves the Stein equation (22) for H. If*

$$H(\nu) = h(\langle \nu, \phi_1 \rangle, \ldots, \langle \nu, \phi_m \rangle),$$

then

$$\psi(H)(\nu) = f(\langle \nu, \phi_1 \rangle, \ldots, \langle \nu, \phi_m \rangle)$$

for a function $f \in C_b^\infty(\mathbf{R}^m)$ with $\|f^{(k)}\| \leq \|h^{(k)}\|$ for all $k \in \mathbf{N}$.

Thus we have established the ingredients for Stein's method, yielding the following theorem, which is our basic formulation of the weak law of large numbers.

Theorem 7 *Let $(\eta_n)_{n\in\mathbf{N}}$ be a family of random elements with values in $M_1(E)$, defined on the same probability space. Let $\mu \in M_1(E)$. Then*

$$\mathcal{L}(\eta_n) \stackrel{w}{\Rightarrow} \delta_\mu \quad (n \to \infty)$$

iff for all $f \in \mathcal{F}_\mathcal{C}$ we have

$$\mathbf{E}f'(\eta_n)[\mu - \eta_n] \to 0 \quad (n \to \infty).$$

Instead of using $\mathcal{F}_\mathcal{C}$, other classes of functions are of course possible. The next corollary gives an example for such a result.

Corollary 2 *If for all $f \in D^2_\mathcal{G}(M_1(E), \mathbf{R})$ with bounded first and second derivatives,*

$$\mathbf{E}f'(\xi_n)[\mu - \xi_n] \to 0 \quad (n \to \infty),$$

then

$$\mathcal{L}(\xi_n) \stackrel{w}{\Rightarrow} \delta_\mu \quad (n \to \infty).$$

The assertion of Corollary 2 is not an if-and-only-if-statement, because Gâteaux-differentiability does not imply continuity (see Yamamuro [78], p.7). Thus, if for $f \in D^2_\mathcal{G}(M_1(E), \mathbf{R})$, we put $h(\nu) = f'(\nu)[\mu - \nu], \nu \in M^b(E)$, then f solves (22) for h, but it is not necessarily true that $h \in C(M_1(E))$.

Almost immediately we get a (rather formal) estimate on the rate of convergence.

Corollary 3 *If η is a random measure on $M_1(E)$,*

$$\zeta_{\mathcal{F}_\mathcal{C}}(\mathcal{L}(\eta), \delta_{\mathbf{E}[\eta]})) : f \in \mathcal{F}_\mathcal{C}\} \quad \leq \quad \sup\{|\mathbf{E}f'(\eta)[\mu - \eta]| : f \in \mathcal{F}_\mathcal{C}\}..$$

In particular

$$\zeta_{\mathcal{F}_\mathcal{C}}(\mathcal{L}(\xi_n), \delta_\mu)) \quad \leq \quad \sup\{|\mathbf{E}f'(\xi_n)[\mu - \xi_n]| : f \in \mathcal{F}_\mathcal{C}\}.$$

Similarly to the real-valued case we can obtain a criterion for the weak law of large numbers in terms of the variance.

Proposition 9 *We have*

$$\begin{aligned}\zeta_{\mathcal{F}_\mathcal{C}}(\mathcal{L}(\xi_n), \delta_\mu) \quad \leq \quad & \sup_{\phi\in\mathcal{C}, \|\phi\|\leq 1} |\langle \mu - \bar{\mu}_n, \phi\rangle| + \sup_{\phi\in\mathcal{C}, \|\phi\|\leq 1} \langle \mu - \bar{\mu}_n, \phi\rangle^2 \\ & + \sup_{\phi\in\mathcal{C}, \|\phi\|\leq 1} \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n \phi(X_i)\right).\end{aligned}$$

In particular, if $F \in \mathcal{F}_\mathcal{C}$ such that $F(\nu) = f(\langle \nu, \phi_1 \rangle, \ldots, \langle \nu, \phi_m \rangle)$, we have

$$
\begin{aligned}
|\mathbf{E}F'(\xi_n)[\mu - \xi_n]| \quad &\leq \quad \max_{1 \leq j \leq m} |\langle \mu - \bar{\mu}_n, \phi_j \rangle| + \max_{1 \leq j \leq m} \langle \mu - \bar{\mu}_n, \phi_j \rangle^2 \\
&\quad + \max_{1 \leq j \leq m} \mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^{n} \phi_j(X_i) \right).
\end{aligned}
$$

Proof: With the Stein equation, we have that for all $g \in \mathcal{F}_\mathcal{C}$, if $\psi(h)$ denotes the solution of the Stein equation (22) for h, that

$$
\begin{aligned}
\mathbf{E}h(\xi_n) - \langle \mu, h \rangle \quad &= \quad \mathbf{E}\mathcal{A}\psi(h)(\xi_n) \\
&= \quad \mathbf{E}\psi'(h)(\xi_n)[\mu - \xi_n] \\
&= \quad \mathbf{E}\psi'(h)(\mu)[\mu - \xi_n] + R_1,
\end{aligned}
$$

where R_1 is the remainder term in the Taylor expansion. Thus

$$
\mathbf{E}h(\xi_n) - \langle \mu, h \rangle = \psi'(h)(\mu)[\mu - \bar{\mu}_n] + R_1.
$$

Now, if $h \in \mathcal{F}_\mathcal{C}$, we have

$$
\psi(h)(\nu) = f(\langle \nu, \phi_1 \rangle, \ldots, \langle \nu, \phi_m \rangle)
$$

for an $m \in \mathbf{N}, f \in C_b^\infty(\mathbf{R}^m)$ with $\sum_{i=1}^m \|f_{(i)}\| \leq 1, \sum_{i,j=1}^m \|f_{(i,j)}\| \leq 1$, and $\phi_1, \ldots, \phi_m \in \mathcal{C}$ with $\|\phi_i\| \leq 1, i = 1, \ldots, m$. Thus

$$
\begin{aligned}
|\psi'(h)(\mu)[\mu - \bar{\mu}_n]| \quad &= \quad |\sum_{i=1}^{m} f_{(i)}(\langle \mu, \phi_1 \rangle, \ldots, \langle \mu, \phi_m \rangle) \langle \mu, \phi_i \rangle| \\
&\leq \quad \max_{1 \leq j \leq m} |\langle \mu - \bar{\mu}_n, \phi_j \rangle|.
\end{aligned}
$$

Furthermore, by Taylor's expansion,

$$
\begin{aligned}
|R_1| \quad &\leq \quad \sum_{i,j=1}^{m} \|f_{(i,j)}\| \mathbf{E} |\langle \mu - \xi_n, \phi_i \rangle \langle \mu - \xi_n, \phi_j \rangle| \\
&\leq \quad \sum_{i,j=1}^{m} \|f_{(i,j)}\| \max_{1 \leq j \leq m} \mathbf{E} |\langle \mu - \xi_n, \phi_j \rangle^2| \\
&\leq \quad \max_{1 \leq j \leq m} \mathbf{E} \langle \mu - \xi_n, \phi_j \rangle^2.
\end{aligned}
$$

As $\langle \xi_n, \phi \rangle = \frac{1}{n}\sum_{i=1}^n \phi(X_i)$, we have

$$\begin{aligned}
&\mathbf{E}\langle \mu - \xi_n, \phi\rangle^2 \\
&= \mathbf{E}\langle \mu - \bar{\mu}_n + \bar{\mu}_n - \xi_n, \phi_j\rangle^2 \\
&= \langle \mu - \bar{\mu}_n, \phi_j\rangle^2 + 2\langle \mu - \bar{\mu}_n, \phi_j\rangle[\langle \bar{\mu}_n, \phi_j\rangle - \frac{1}{n}\sum_{i=1}^n \mathbf{E}\phi_j(X_i)] \\
&\quad + \mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^n (\phi_j(X_i) - \mathbf{E}\phi_j(X_i))\right\}^2 \\
&= \langle \mu - \bar{\mu}_n, \phi_j\rangle^2 + \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n \phi_j(X_i)\right),
\end{aligned}$$

and the assertion follows.

2.3.1 The Local Approach

The above results can be specified for random elements having local dependence. Using Taylor expansion, the following result is not hard to show (see Reinert [61]).

Corollary 4 *Assume that for all $i, n \in \mathbf{N}$ there is a $\Gamma_s^n(i) \subset \{1, \ldots, n\}$ such that, for each i, X_i is independent of $\sigma(X_j, j \notin \Gamma_s^n(i))$. Then*

$$\zeta_{\mathcal{F}_C}(\mathcal{L}(\xi_n), \delta_\mu)) \leq \frac{|\Gamma_s^n(i)|}{n}\{1 + \zeta_C(\bar{\mu_n}, \mu)\} + \frac{1}{n^2}\sum_{i=1}^n \sum_{j \in \Gamma_s^n(i)} \mathbf{E}|X_i X_j|.$$

2.3.2 Size Bias Couplings

Moreover the size bias coupling can be applied to the law of large numbers for measure-valued elements. For ξ as above, similarly to Section 1, we construct ξ^ϕ as follows. Pick an index $V \in \{1, \ldots, n\}$ according to

$$\mathbf{P}[V = v] = \frac{\mathbf{E}\phi(X_v)}{\langle \bar{\mu}_n, \phi\rangle}.$$

If $V = v$, take $\delta^*_{X_v}$ to have the δ_{X_v} size biased distribution in direction ϕ. Thus, considering $f(\nu) = 1_{d\eta}(\nu)$, we have

$$\mathbf{P}[\delta^*_{X_v} \in d\eta] = \frac{1}{\langle \bar{\mu}_n, \phi\rangle}\langle \eta, \phi\rangle \mathbf{P}[\delta_{X_v} \in d\eta].$$

If $\delta^*_{X_v} = \eta$, choose $\delta^v_{X_u}, u \neq v$, such that

$$\mathcal{L}(\delta^v_{X_u}, u \neq v) = \mathcal{L}(\delta_{X_u}, u = 1, \ldots, n | \delta_{X_v} = \eta).$$

Put

$$\xi_n^\phi = \frac{1}{n}\sum_{u=1}^n \delta^v_{X_u}.$$

Then we have

$$\begin{aligned}
\mathbf{E}f(\xi_n^\phi) &= \sum_{v=1}^n \frac{\mathbf{E}\phi(X_v)x]}{\langle\bar\mu_n,\phi\rangle}\mathbf{E}f\left(\frac{1}{n}\sum_{u=1}^n \delta^v_{X_u}\right)\\
&= \sum_{v=1}^n \frac{\mathbf{E}\phi(X_v)}{\langle\bar\mu_n,\phi\rangle}\int_{M_1(E)}\mathbf{E}\left\{f\left(\frac{1}{n}\sum_{u=1}^n \delta_{X_u}\right)\middle|\delta_{X_v}=\eta\right\}\mathbf{P}[\delta^*_{X_v}\in d\eta]\\
&= \sum_{v=1}^n \frac{1}{\langle\bar\mu_n,\phi\rangle}\int_{M_1(E)}\langle\eta,\phi\rangle\mathbf{E}\left\{f\left(\frac{1}{n}\sum_{u=1}^n \delta_{X_u}\right)\middle|\delta_{X_v}=\eta\right\}\mathbf{P}[\delta_{X_v}\in d\eta]\\
&= \frac{1}{n\langle\bar\mu_n,\phi\rangle}\sum_{v=1}^n \mathbf{E}f(\xi_n)\langle\delta_{X_v},\phi\rangle\\
&= \frac{1}{\langle\bar\mu_n,\phi\rangle}\mathbf{E}f(\xi_n)\langle\xi_n,\phi\rangle.
\end{aligned}$$

Thus ${\xi_n}^\phi$ has the ξ_n size biased distribution in direction ϕ.

Corollary 5 *Let $\mathcal{C} = C_c(E,\mathbf{R}_+)$ or $C^\infty_{0,b}(\mathbf{R}^k)$, if $E=\mathbf{R}^k$. Then*

$$\zeta_{\mathcal{F}_\mathcal{C}}(\mathcal{L}(\xi_n),\delta_\mu)) \quad\leq\quad \zeta_\mathcal{C}(\bar{\mu_n},\mu) + \max_{\phi,\psi\in\mathcal{C}}\mathbf{E}|\langle\xi_n^\phi-\xi_n,\phi\rangle|.$$

Finally, similarly to the real-valued case, zero biasing provides a fast proof for Proposition 9, restricted to those functions ϕ for which $\mathbf{E}\phi(X_i)=0, i=1,\dots,n$.

2.4 Gaussian Approximations for Measure-Valued Random Elements

This section is based on Reinert [64]. 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. Similarly to the real-valued case, for $F\in\mathcal{F}_{\mathcal{C},f}$ with representation (19), we define the generator

$$\begin{aligned}
\mathcal{A}F(\nu) &= -\sum_{j=1}^m f_{(j)}(\langle\nu,\phi_1\rangle,\dots,\langle\nu,\phi_m\rangle)\langle\nu,\phi_j\rangle\\
&\quad+\sum_{j,k=1}^m f_{(j,k)}(\langle\nu,\phi_1\rangle,\dots,\langle\nu,\phi_m\rangle)b(\phi_j,\phi_k).
\end{aligned}$$

We say that $\mathcal{A}$ is the generator associated with the operator b, or, slightly abusing notation, associated with the matrix B.

Let ζ be a random measure taking values in $M^f(E)$ almost surely such that, for all $m\in\mathbf{N},\phi_1,\dots,\phi_m\in\mathcal{C}$,

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

where $\mathcal{MVN}_m(0,B)$ denotes the multivariate normal law with mean vector 0 and covariance matrix B. Let $\mathcal{A}$ be the generator associated with B. Then $\mathcal{L}(\zeta)$ is stationary for $\mathcal{A}$. Thus, for $H \in \mathcal{F}_{\mathcal{C},f}$ of the form

$$H(\nu) = h(\langle \nu, \psi_1 \rangle, \dots, \langle \nu, \psi_m \rangle), \tag{23}$$

the Stein equation corresponding to the Gaussian random measure ζ is

$$\begin{aligned} & h(\langle \nu, \psi_1 \rangle, \dots, \langle \nu, \psi_m \rangle) - \mathbf{E}[h(\langle \zeta, \psi_1 \rangle, \dots, \langle \zeta, \psi_m \rangle) \\ = & \quad - \sum_{j=1}^m f_{(j)}(\langle \nu, \phi_1 \rangle, \dots, \langle \nu, \phi_m \rangle) \langle \nu, \phi_j \rangle \\ & \quad + \sum_{j,k=1}^m f_{(j,k)}(\langle \nu, \phi_1 \rangle, \dots, \langle \nu, \phi_m \rangle) b(\phi_j, \phi_k). \end{aligned} \tag{24}$$

This equation can be solved using the semigroup technique as in Barbour [10].

Lemma 8 *For $F(\nu) \in \mathcal{F}_{\mathcal{C},f}$, and for $t \geq 0$, define the operator T_t by*

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

Then, if $H \in \mathcal{F}_{\mathcal{C},f}$ has the form (23), the function

$$F(\nu) = - \int_0^\infty \{T_t H(\nu) - \mathbf{E} H(\zeta)\} dt$$

solves (24) for H. Moreover, $F \in \mathcal{F}_{\mathcal{C},f}$, and there is a function $f \in C_b^\infty(\mathbf{R}^m)$ such that $F(\nu) = f(\langle \nu, \psi_1 \rangle, \dots, \langle \nu, \psi_m \rangle)$, $\|f^{(k)}\| \leq \|h^{(k)}\|, k \in \mathbf{N}$, and $\|F^{(k)}\| \leq \|H^{(k)}\|, k \in \mathbf{N}$.

Theorem 9 *Let $(\eta_n)_{n \in \mathbf{N}}$ be a family of random measures 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 $m \in \mathbf{N}, \phi_1, \dots, \phi_m \in \mathcal{C}$,*

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

Let $\mathcal{A}$ be the generator associated with B. Suppose that for all $F \in \mathcal{F}_{\mathcal{C},f}$,

$$\mathbf{E}\mathcal{A}F(\eta_n) \to 0 \quad (n \to \infty).$$

Then

$$\mathcal{L}(\eta_n) \stackrel{w}{\Rightarrow} \mathcal{L}(\zeta) \quad (n \to \infty).$$

PROOF. By the generator method, $\mathbf{E}\mathcal{A}F(\eta_n) \to 0 \quad (n \to \infty)$ for all $F \in \mathcal{F}_{\mathcal{C},f}$ implies that, for all $G \in \mathcal{F}_{\mathcal{C},f}$,

$$\mathbf{E}G(\eta_n) - \mathbf{E}G(\zeta) \to 0 \quad (n \to \infty).$$

Furthermore, $M^f(E)$ is locally compact. Corollary 1 gives the assertion.

Theorem 9 assumes the existence of a Gaussian random measure ζ that is finite almost surely. In general, the almost sure finiteness is not guaranteed. However, if E is compact, then, with $\mathcal{C} = C_c(E), C_c^+(E), \mathcal{FL}(E)$ or $\mathcal{C} = \mathcal{FL}^+(E)$, 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.

Stein's method also allows us to compare random measures with Gaussian random measures.

Proposition 10 *Let ζ be a random measure taking values in $M^f(E)$ almost surely such that, for all $m \in \mathbf{N}, \phi_1, \ldots, \phi_m \in \mathcal{C}$,*

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

where $B = (b_{i,j})_{i,j=1,\ldots,m} = (b(\phi_i, \phi_j))_{i,j=1,\ldots,m}$ is symmetric, positive definite, and let $\mathcal{A}$ be the generator associated with B. Then, for any random element η taking values in $M^f(E)$ almost surely,

$$\zeta_{\mathcal{F}_{\mathcal{C},f}}(\mathcal{L}(\eta), \mathcal{L}(\zeta)) \quad \leq \quad \sup_{F \in \mathcal{F}_{\mathcal{C},f}} |\mathbf{E}\mathcal{A}F(\eta)|\,.$$

Proof. By Lemma 8, we have

$$\zeta_{\mathcal{F}_{\mathcal{C},f}}(\mathcal{L}(\eta), \mathcal{L}(\zeta)) = \sup_{f \in \mathcal{F}_{\mathcal{C},f}} \left| \langle \mathcal{L}(\eta), f \rangle - \langle \mathcal{L}(\zeta), f \rangle \right| \leq \sup_{f \in \mathcal{F}_{\mathcal{C},f}} |\mathbf{E}\mathcal{A}F(\eta)|\,.$$

This proves the assertion.

We now consider the case of empirical processes. Throughout this section, $(X_i)_{i \in \mathbf{N}}$ is a family of (possibly dependent) 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 for all $\phi_1, \ldots, \phi_m \in \mathcal{C}$, let

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

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

$$\mathcal{C}_* = \{\phi \in \mathcal{C} : \|\phi\| \leq 1\}.$$

From Proposition 10 we obtain the following corollary.

Corollary 6 *Let ζ be a random measure taking values in $M^f(E)$ almost surely such that, for all $m \in \mathbf{N}, \phi_1, \ldots, \phi_m \in \mathcal{C}_,$,*

$$\mathcal{L}(\langle \zeta_n, \phi_1 \rangle, \ldots, \langle \zeta_n, \phi_m \rangle) = \mathcal{MVN}_m(0, C),$$

where $C = (c(\phi_k, \phi_l))_{k,l=1,\dots,m}$ is symmetric and positive definite, and let $\mathcal{A}$ be the generator associated with c. Then

$$\zeta_{\mathcal{F}_{\mathcal{C},f}}(\mathcal{L}(\xi_n), \mathcal{L}(\zeta)) \quad \leq \quad \sup_{F \in \mathcal{F}_{\mathcal{C},f}} |\mathbf{E}\mathcal{A}_n F(\xi_n)| + \sup_{\phi,\psi \in \mathcal{C}^*} |b_n(\phi,\psi) - c(\phi,\psi)|.$$

PROOF. Let ζ_n be a random measure taking values in $M^f(E)$ almost surely such that, for all $m \in \mathbf{N}, \phi_1, \dots, \phi_m \in \mathcal{C}$, $\mathcal{L}(\langle \zeta_n, \phi_1 \rangle, \dots, \langle \zeta_n, \phi_m \rangle) = \mathcal{MVN}_m(0, B_n)$. From Proposition 10,

$$\begin{aligned} \zeta_{\mathcal{F}_{\mathcal{C},f}}(\mathcal{L}(\xi_n), \mathcal{L}(\zeta)) &\leq \zeta_{\mathcal{F}_{\mathcal{C},f}}(\mathcal{L}(\xi_n), \mathcal{L}(\zeta_n)) + \zeta_{\mathcal{F}_{\mathcal{C},f}}(\mathcal{L}(\zeta_n), \mathcal{L}(\zeta)) \\ &\leq \sup_{F \in \mathcal{F}_{\mathcal{C},f}} |\mathbf{E}\mathcal{A}_n F(\xi_n)| + \sup_{F \in \mathcal{F}_{\mathcal{C},f}} |\mathbf{E}\mathcal{A} F(\zeta_n)|. \end{aligned}$$

For the second summand, let $F \in \mathcal{F}_{\mathcal{C},f}$ have the representation (19). Then

$$\begin{aligned} &\mathbf{E}\mathcal{A}F(\zeta_n) \\ &= -\sum_{j=1}^m \mathbf{E} f_{(j)}(\langle \zeta_n, \phi_1 \rangle, \dots, \langle \zeta_n, \phi_m \rangle) \langle \zeta_n, \phi_j \rangle \\ &\quad + \sum_{j,k=1}^m \mathbf{E} f_{(j,k)}(\langle \zeta_n, \phi_1 \rangle, \dots, \langle \zeta_n, \phi_m \rangle) c(\phi_j, \phi_k) \\ &= \sum_{j,k=1}^m \mathbf{E} f_{(j,k)}(\langle \zeta_n, \phi_1 \rangle, \dots, \langle \zeta_n, \phi_m \rangle) \left(c(\phi_j, \phi_k) - b_n(\phi_j, \phi_k) \right), \end{aligned}$$

using that ζ_n is Gaussian. Taking absolute values proves the assertion.

To illustrate the use of Proposition 10, we derive a Gaussian approximation in the independent case (see also Ledoux and Talagrand [51]).

Corollary 7 *Assume that $(X_i)_{i \in \mathbf{N}}$ is a family of independent random elements on E. Let ζ_n be a random measure taking values in $M^f(E)$ such that, for all $m \in \mathbf{N}, \phi_1, \dots, \phi_m \in \mathcal{C}$,*

$$\mathcal{L}(\langle \zeta_n, \phi_1 \rangle, \dots, \langle \zeta_n, \phi_m \rangle) = \mathcal{MVN}_m(0, B_n).$$

Then

$$\zeta_{\mathcal{F}_{\mathcal{C},f}}(\mathcal{L}(\xi_n), \mathcal{L}(\zeta_n)) \leq \frac{8}{\sqrt{n}}.$$

PROOF. Let $F \in \mathcal{F}_{\mathcal{C},f}$, so that

$$\begin{aligned} &\mathbf{E}\mathcal{A}_n F(\xi_n) \\ &= -\sum_{j=1}^m \mathbf{E} f_{(j)}(\langle \xi_n, \phi_1 \rangle, \dots, \langle \xi_n, \phi_m \rangle) \langle \xi_n, \phi_j \rangle \\ &\quad + \frac{1}{n} \sum_{i=1}^n \sum_{j,k=1}^m \mathbf{E} f_{(j,k)}(\langle \xi_n, \phi_1 \rangle, \dots, \langle \xi_n, \phi_m \rangle) \mathrm{Cov}\,(\phi_j(X_i), \phi_k(X_i)). \end{aligned}$$

Put $\xi_n^i = \frac{1}{\sqrt{n}} \sum_{j \neq i} (\delta_{X_j} - \mu_j)$. By Taylor expansion, we obtain for the first term

$$
\begin{aligned}
&\sum_{j=1}^m \mathbf{E} f_{(j)}(\langle \xi_n, \phi_1 \rangle, \dots, \langle \xi_n, \phi_m \rangle) \langle \xi_n, \phi_j \rangle \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=1}^m \mathbf{E} f_{(j)}(\langle \xi_n, \phi_1 \rangle, \dots, \langle \xi_n, \phi_m \rangle) \langle \delta_{X_i} - \mu_i, \phi_j \rangle \\
&= \sum_{i=1}^n \sum_{j,k=1}^m \mathbf{E} f_{(j,k)}\big(\langle \xi_n^i, \phi_1 \rangle, \dots, \langle \xi_n^i, \phi_m \rangle\big) \langle \delta_{X_i} - \mu_i, \phi_k \rangle \langle \delta_{X_i} - \mu_i, \phi_j \rangle \\
&\quad + R,
\end{aligned}
$$

using the independence. Here,

$$
|R| \quad \leq \quad \frac{4}{\sqrt{n}} \|f'''\| \max_{1 \leq i \leq m} \|\phi_i\|^3.
$$

Hence, using independence and Taylor expansion again,

$$
|\mathbf{E}[\mathcal{A}_n F(\xi_n)]| \quad \leq \quad \frac{8}{\sqrt{n}} \|f'''\| \max_{1 \leq i \leq m} \|\phi_i\|^3.
$$

As in $\mathcal{F}_{\mathcal{C},f}$, we have $\|f'''\| \leq 1$ and $\|\phi_i\| \leq 1$, the assertion follows using Proposition 10.

2.4.1 The Local Approach

Taylor expansion gives the following result for the sum of locally dependent random measures.

Corollary 8 *For all $i, n \in \mathbf{N}$ let $\Gamma_n(i) \subset \{1, \dots, n\}$ be a set such that for each $l \notin \Gamma_n(i)$, X_l is independent of X_i. Let $\gamma_n = \max_{i=1,\dots,n} |\Gamma_n(i)|$. Let ζ_n be a random measure taking values in $M^f(E)$ almost surely such that, for all $m \in \mathbf{N}, \phi_1, \dots, \phi_m \in \mathcal{C}$,*

$$
\mathcal{L}(\langle \zeta_n, \phi_1 \rangle, \dots, \langle \zeta_n, \phi_m \rangle) = \mathcal{MVN}_m(0, B_n).
$$

Then, for each $H \in \mathcal{F}_{\mathcal{C},f}$ of the form (23) we have

$$
|\mathbf{E}H(\xi_n) - \mathbf{E}H(\zeta_n)| \leq \frac{20}{\sqrt{n}} \|h'''\| \max_{1 \leq i \leq m} \|\phi_i\|^3 \gamma_n^2.
$$

Moreover,

$$
\zeta_{\mathcal{F}_{\mathcal{C},f}}(\mathcal{L}(\xi_n), \mathcal{L}(\zeta_n)) \leq \frac{20}{\sqrt{n}} \gamma_n^2.
$$

2.4.2 Size Bias Couplings

Similarly as in the context of random measures, we obtain a coupling result using size biasing for random measures.

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 $\mathcal{C} = C_c^+(E)$ or $\mathcal{C} = \mathcal{FL}^+(E)$. Let ζ_n be a random measure taking values in $M^f(E)$ such that, for all $m \in \mathbf{N}, \phi_1, \dots, \phi_m \in \mathcal{C}$,*

$$\mathcal{L}(\langle \zeta_n, \phi_1\rangle, \dots, \langle \zeta_n, \phi_m\rangle) = \mathcal{MVN}_m(0, B_n),$$

where

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

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 $\xi_n = \eta - \mathbf{E}\eta$. Let $\eta^\phi = \frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_{X_i^V}$ have the size-bias distribution of η in direction ϕ as constructed in Subsubsection 2.3.2. Then, for all $F \in \mathcal{F}_{\mathcal{C},f}$ of the form (19),

$$\begin{aligned}
|\mathbf{E}\mathcal{A}_nF(\xi_n)| \quad &\le \quad \sum_{j,k=1}^{m} \|f_{(j,k)}\| |\langle \mathbf{E}\eta, \phi_j\rangle| \big\{\mathrm{Var}(\mathbf{E}(\langle \eta^{\phi_j} - \eta, \phi_k\rangle|\eta))\big\}^{\frac{1}{2}} \\
&\quad + \frac{1}{2}\sum_{j,k,l=1}^{m} \|f_{(j,k,l)}\| \langle \mathbf{E}\eta, \phi_j\rangle \mathbf{E}|\langle \eta^{\phi_j} - \eta, \phi_k\rangle\langle \eta^{\phi_j} - \eta, \phi_l\rangle|
\end{aligned}$$

and

$$\begin{aligned}
\zeta_{\mathcal{F}_\mathcal{C}}(\mathcal{L}(\xi_n), \mathcal{L}(\zeta_n)) \quad &\le \quad \max_{\phi,\psi\in\mathcal{C}^*}\left(\langle \mathbf{E}\eta, \phi\rangle \big\{\mathrm{Var}(\mathbf{E}(\langle \eta^\phi - \eta, \psi\rangle|\eta))\big\}^{\frac{1}{2}}\right) \\
&\quad + \frac{1}{2}\max_{\phi,\psi,\tau\in\mathcal{C}^*}\left(\langle \mathbf{E}\eta, \phi\rangle \mathbf{E}|\langle \eta^\phi - \eta, \psi\rangle\langle \eta^\phi - \eta, \tau\rangle|\right).
\end{aligned}$$

2.4.3 Zero Bias Coupings

Using Taylor expansion, the zero bias coupling immediately yields the following theorem.

Theorem 11 *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 $\mathcal{C} = C_c(E)$ or $\mathcal{C} = \mathcal{FL}(E)$. Let $\mathcal{H} = \{\phi \in \mathcal{C} : \mathbf{E}\langle \xi_n, \phi\rangle = 0\}$. Let ζ_n be a random measure taking values in $M^f(E)$ such that, for all $m \in \mathbf{N}, \phi_1, \dots, \phi_m \in \mathcal{H}$,*

$$\mathcal{L}(\langle \zeta_n, \phi_1\rangle, \dots, \langle \zeta_n, \phi_m\rangle) = \mathcal{MVN}_m(0, B_n),$$

where

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

and let $\mathcal{A}_n$ be the generator associated with B_n. Let $\xi^*_{n,\phi,\psi}$ have the ξ_n-zero biased distribution as described in Subsection 2.2 and assume that $\xi_n, \xi^*_{n,\phi,\psi}$ are constructed on the same probability space. Then, for all $F \in \mathcal{F}_{\mathcal{C},f}$ of the form (19),

$$|\mathbf{E}\mathcal{A}_n F(\xi_n)| \quad \le \quad \sum_{j,k,l=1}^{m} b_n(\phi_j,\phi_k)\|f_{(j,k,l)}\|\mathbf{E}|\langle \xi^*_{n,\phi_j,\phi_k} - \xi_n, \phi_l\rangle|.$$

2.5 Poisson Approximations for Point Processes

In the case that we want to approximate the distribution of a point process

$$\zeta = \sum_{\alpha\in\Gamma} \delta_\alpha I_\alpha$$

with $(I_\alpha, \alpha\in\Gamma)$ a collection of indicators, by the Poisson measure $Po(\pi)$ on Γ with intensity $(\pi_\alpha, \alpha\in\Gamma)$, we may conveniently use the total variation distance. Let $\xi = \sum_{\alpha\in\Gamma}\delta_\alpha x_\alpha$ be an element of the space $\mathcal{Z}$ of configurations of point processes over Γ, so that $x_\alpha \in \mathbf{Z}^+$ for each $\alpha\in\Gamma$. Then, following Barbour *et al.* [19], p. 207, the generator is

$$\mathcal{A}f(\xi) = \sum_{\alpha\in\Gamma}\pi_\alpha(f(\xi+\delta_\alpha)-f(\xi)) + \sum_{\alpha\in\Gamma} x_\alpha(f(\xi-\delta_\alpha)-f(\xi)),$$

for $f:\mathcal{Z}\to\mathbf{R}$ bounded. The Chen-Stein equation becomes

$$h(\xi) - Po(\pi)h = \mathcal{A}f(\xi). \tag{25}$$

In Barbour *et al.* [19], Lemma 10.1.31, it is shown that for f being the solution of equation (25) for $h(\xi) = \mathbf{I}(\xi\in A)$, the following bounds hold.

$$\begin{aligned}\Delta_1 f &= \sup_{\xi\in\mathcal{Z},\alpha\in\Gamma} |f(\xi+\delta_\alpha) - f(\xi)| \le 1\\ \Delta_2 f &= \sup_{\xi\in\mathcal{Z},\alpha,\beta\in\Gamma} |f(\xi+\delta_\alpha+\delta_\beta) - f(\xi+\delta_\alpha) - f(\xi+\delta_\beta) + f(\xi)| \le 1.\end{aligned}$$

2.5.1 The Local Approach

Assume I_α and B_α, $\alpha\in\Gamma$, are as in Theorem 6. Barbour *et al.* [19] give the following theorem.

Theorem 12 *For each $\alpha\in\Gamma$, put $Z_\alpha = \sum_{\beta\in B_\alpha} I_\beta$. We have*

$$d_{TV}(\mathcal{L}(\zeta), \mathrm{Po}(\pi)) \quad\le\quad \sum_{\alpha\in\Gamma}\left\{\pi_\alpha^2 + \pi_\alpha\mathbf{E}Z_\alpha + \mathbf{E}(I_\alpha Z_\alpha)\right\} + b_3,$$

where b_3 is as in Theorem 6.

2.5.2 Size Bias Couplings

Again, size biasing of point mass is particularly easy. Barbour *et al.* [19] obtain the following theorem (Theorem 10.B in Barbour *et al.* [19]).

Theorem 13 *Suppose that, for each $\alpha \in \Gamma$, a random element $\hat{\zeta}_\alpha$ can be realized on the same probability space as ζ in such a way that $\mathcal{L}(\hat{\zeta}_\alpha + \delta_\alpha) = \mathcal{L}(\zeta | I_\alpha = 1)$. Then*

$$d_{TV}(\mathcal{L}(\zeta), \mathrm{Po}(\pi)) \quad \leq \quad \sum_{\alpha, \beta \in \Gamma} \pi_\alpha \mathbf{E} |\hat{\zeta}_\alpha\{\beta\} - \zeta\{\beta\}|.$$

3 Examples for the Law of Large Numbers for Measure-Valued Random Elements

Most of the following examples can be found in Reinert [61]. We start with an example for the local approach.

3.1 A Dissociated Array

Let $(Y_i)_{i \in \mathbf{N}}$ be a family of independent random elements on a space $\mathcal{X}$, let $k \in \mathbf{N}$ be fixed, and set

$$\begin{aligned} \Gamma &= \{(j_1, \ldots, j_k) \in \mathbf{N}^k : j_r \neq j_s \;\text{ for }\; r \neq s\}; \\ \Gamma^{(n)} &= \{(j_1, \ldots, j_k) \in \Gamma : j_1, \ldots, j_k \in \{1, \ldots, n\}\}. \end{aligned}$$

Suppose, $(\psi_{j_1, \ldots, j_k})_{(j_1, \ldots, j_k) \in \Gamma}$ is a family of measurable functions $\mathcal{X}^k \to E$, and put, for $(j_1, \ldots, j_k) \in \Gamma$,

$$X_{j_1, \ldots, j_k} = \psi_{j_1, \ldots, j_k}(Y_{j_1}, \ldots, Y_{j_k}).$$

Then, $(X_{j_1, \ldots, j_k})_{(j_1, \ldots, j_k) \in \Gamma}$ is a dissociated family. Assume furthermore that

$$\frac{1}{n_{(k)}} \sum_{(j_1, \ldots, j_k) \in \Gamma^{(n)}} \mathcal{L}(X_{j_1, \ldots, j_k}) \Rightarrow_v \mu$$

for a $\mu \in M_1(E)$, where

$$n_{(k)} = n(n-1) \cdots (n-k+1).$$

Let

$$\xi_n = \frac{1}{n_{(k)}} \sum_{(j_1, \ldots, j_k) \in \Gamma^{(n)}} \delta_{X_{j_1, \ldots, j_k}}.$$

Using the local approach, we obtain

Theorem 14 *For the above dissociated family, we have*

$$\begin{aligned}\zeta_{\mathcal{F}_C}(\mathcal{L}(\xi_n),\delta_\mu) &\leq \zeta_{C_C(E)}\left(\frac{1}{n_{(k)}}\sum_{(j_1,\dots,j_k)\in\Gamma^{(n)}}\mathcal{L}(X_{j_1,\dots,j_k}),\mu\right)\\ &\quad+\frac{2}{n_{(k)}^2}\sum_{i=1}^{n_{(k)}}|\Gamma_s^n(i)|.\end{aligned}$$

Proof: For $n\in\mathbf{N}$ fixed, the set $\Gamma^{(n)}$ has $n_{(k)}$ elements. Fix a counting for $\Gamma^{(n)}$. If $(i_1,\dots,i_k)$ is the ith element, identify $(i_1,\dots,i_k)$ with i, and set $X_{i_1,\dots,i_k}=X_{i,n}$. Then

$$\xi_n=\frac{1}{n_{(k)}}\sum_{i=1}^{n_{(k)}}\delta_{X_{i,n}}.$$

For $i=(i_1,\dots,i_k)$, define

$$\Gamma_s^n(i)=\{(l_1,\dots,l_k)\in\Gamma^{(n)}:(l_1,\dots,l_k)\neq i;\{l_1,\dots,l_k\}\cap\{i_1,\dots,i_k\}\neq\emptyset\}.$$

Then, for all $i\leq n_{(k)}$,

$$\begin{aligned}|\Gamma_s^n(i)| &= k[k(n-1)(n-2)\cdots(n-k+1)-1]\\ &\leq \frac{n_{(k)}}{n}k^2.\end{aligned}$$

Now let $d\in\mathbf{N}$ be arbitrary, $f\in C_b^\infty(\mathbf{R}^d),\phi_1,\dots,\phi_d,\phi\in C_c(E)$. Put

$$\xi_{n,w,i}=\frac{1}{n_{(k)}}\sum_{m\in\Gamma_w^n(i)}\delta_{X_{m,n}}.$$

Then, by the independence of $X_{i,n}$ and $X_{l_1,\dots,l_k}$ for $(l_1,\dots,l_k)\in\Gamma_w^n(i)$, we have

$$\begin{aligned}&\frac{1}{n_{(k)}}\sum_{i=1}^{n_{(k)}}\mathbf{E}f\big(\big\langle\xi_{n,w,i},\phi_1\big\rangle,\dots,\big\langle\xi_{n,w,i},\phi_d\big\rangle\big)(\phi(X_{i,n})-\langle\mu,\phi\rangle)\\ &= \frac{1}{n_{(k)}}\sum_{i=1}^{n_{(k)}}\mathbf{E}f\big(\big\langle\xi_{n,w,i},\phi_1\big\rangle,\dots,\big\langle\xi_{n,w,i},\phi_d\big\rangle\big)\mathbf{E}\big[\phi(X_{i,n})-\langle\mu,\phi\rangle\big]\\ &= \frac{1}{n_{(k)}}\sum_{i=1}^{n_{(k)}}\mathbf{E}f(\langle\xi_n,\phi_1\rangle,\dots,\langle\xi_n,\phi_d\rangle)\big[\langle\mathcal{L}(X_{i,n}),\phi\rangle-\langle\mu,\phi\rangle\big]+R_1,\end{aligned}$$

where

$$\begin{aligned}R_1 &= \frac{1}{n_{(k)}}\sum_{i=1}^{n_{(k)}}\mathbf{E}\big[f\big(\big\langle\xi_{n,w,i},\phi_1\big\rangle,\dots,\big\langle\xi_{n,w,i},\phi_d\big\rangle\big)\\ &\quad -f(\langle\xi_n,\phi_1\rangle,\dots,\langle\xi_n,\phi_d\rangle)\big]\mathbf{E}\big[\phi(X_{i,n})-\langle\mu,\phi\rangle\big].\end{aligned}$$

By Taylor expansion,

$$|R_1| \quad \leq \quad 2\|\phi\|\|f'\| \sup_{1\leq l\leq d} \|\phi_l\| \frac{1}{n_{(k)}^2} \sum_{i=1}^{n_{(k)}} |\Gamma_s^n(i)|$$

and the assertion follows.

Note that in the exchangeably dissociated case, i.e. $(Y_i)_i$ being i. i. d. and $\psi_{j_1,\dots,j_k} = \psi$ for all $(j_1, \dots, j_k) \in \Gamma$, the $X_{j_1,\dots,j_k}$'s are identically distributed, and thus the assumption about vague convergence is trivially satisfied.

3.2 An Immigration-Death Process

We consider the following immigration-death process with total population size n. Let A_i^K be the (positive) arrival time of the ith individual, and let Z_i its life span. Assume the $(Z_i)_{i\in\mathbf{N}}$ are positive, i. i. d. and independent of the $(A_i^K)_{i,n\in\mathbf{N}}$ (but allow for dependence between the $(A_i^K)_{i,n\in\mathbf{N}}$). Start at time $t = 0$, and let

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

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

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

Thus $\delta_{X_i^n}$ describes the temporal evolution of the ith individual, and

$$\xi_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i^n}$$

gives the "average" path behavior of the process.

Theorem 15 *Let $\mu_i = \mathcal{L}([A_i^K, A_i^K + Z_i))$. In the above setting,*

$$\begin{aligned}\zeta_{\mathcal{F}_C}(\mathcal{L}(\xi_n), \delta_\mu) \quad \leq \quad & \max_{1\leq j\leq m} |\langle \mu - \bar{\mu}_n, \phi_j\rangle| + \max_{1\leq j\leq m} \langle \mu - \bar{\mu}_n, \phi_j\rangle^2 \\ & + \frac{1}{n^2} \sum_{i,j=1}^n \mathbf{E}|(A_i^K - \mathbf{E}A_i^K)(A_j^K - \mathbf{E}A_j^K)|.\end{aligned}$$

PROOF: We employ Proposition 9. Due to the second assumption, we only have to bound the variance term. We have for all $\phi \in C_b^\infty(\mathbf{R}^2)$

$$
\begin{aligned}
&\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n \phi(A_i^K, A_i^K + Z_i)\right) \\
&= \frac{1}{n^2}\sum_{i,j=1}^n \mathbf{E}\{\mathbf{E}[(\phi(A_i^K, A_i^K + Z_i) - \mathbf{E}\phi(A_i^K, A_i^K + Z_i)) \\
&\quad (\phi(A_j^K, A_j^K + Z_j) - \mathbf{E}\phi(A_j^K, A_j^K + Z_j))|Z_i, Z_j]\} \\
&= \frac{1}{n^2}\sum_{i,j=1}^n \mathbf{E}\Big\{\mathbf{E}\Big[\int (\phi(A_i^K, A_i^K + Z_i) - \phi(x, x + Z_i))\,\mathbf{P}[A_i^K \in dx] \\
&\quad \int (\phi(A_j^K, A_j^K + Z_j) - \phi(y, y + Z_j))\,\mathbf{P}[A_j^K \in dy]|Z_i, Z_j\Big]\Big\}.
\end{aligned}
$$

Hence, using Taylor's expansion, we get

$$
\begin{aligned}
&\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n \phi(A_i^K, A_i^K + Z_i)\right) \\
&\le \|\phi'\|^2 \frac{1}{n^2}\sum_{i,j=1}^n \mathbf{E}[|(A_i^K - \mathbf{E}A_i^K)(A_j^K - \mathbf{E}A_j^K)|],
\end{aligned}
$$

and the assertion follows.

Note that under the conditions

1. $\frac{1}{n^2}\sum_{i,j=1}^n \mathbf{E}[|(A_i^K - \mathbf{E}A_i^K)(A_j^K - \mathbf{E}A_j^K)|] \to 0 \quad (n \to \infty)$
2. there is a measure $\mu \in M_1(\mathbf{R}^2)$ with

$$
\frac{1}{n}\sum_{i=1}^n \mathcal{L}([A_i^K, A_i^K + Z_i]) \Rightarrow_v \mu \quad (n \to \infty)
$$

the weak convergence of the empirical measures follows. As an example, in the following case it is easily checked that the conditions are satisfied. Let $(E_i^n)_{i=1,\dots,n}$ be i. i. d. exp(n) random variables, and put $A_i^K = \sum_{j=1}^n E_i^n$. If G is the distribution function of Z_1, then the limiting measure μ is such that, for real rectangles,

$$
\mu([\beta_{11}, \beta_{12}) \times [\beta_{21}, \beta_{22})) = \int_{\beta_{11}}^{\beta_{12}} 1_{[0,1]}(x)[G(\beta_{22} - x) - G(\beta_{21} - x)]dx.
$$

3.3 The General Stochastic Epidemic

The General Stochastic Epidemic (GSE) is perhaps the best-known stochastic epidemic disease model (for a survey see Bailey [5]). Based on Sellke's [71] approach, we construct a generalization of the GSE as follows. For more detail see Reinert [63].

A population with total size K is considered. At time $t = 0$, a fraction aK of these individuals are infected by a certain disease (and are infectious - the infectious period and the period of being infected are assumed to coincide); the remaining bK individuals are susceptible to that disease; $aK + bK = K$. Infectious individuals will get removed after some time, e.g., by lifelong immunity or death, and are then no longer affected by the disease. (Thus, we have a so-called SIR model, where the abbreviation stands for susceptible - infected - removed.)

Let $(l_i, r_i)_{i \in \mathbf{N}}$ be a family of positive i.i.d. random vectors, and let $(\hat{r}_i)_{i \in \mathbf{N}}$ be a family of positive, independent random variables. Assume that the families $(l_i, r_i)_{i \in \mathbf{N}}$ and $(\hat{r}_i)_{i \in \mathbf{N}}$ are mutually independent.

An initially infected individual i stays infectious for a period of length $\hat{r}_i$, then is removed. (That the $\hat{r}_i$ need not be identically distributed reflects the possibility that an infected individual has already been infectious for a certain period before, at time $t = 0$, it is observed.) An initially susceptible individual i, once infected, stays infectious for a period of length r_i, until removal. Furthermore, an initially susceptible individual i accumulates exposure to infection with a rate that depends of the evolution of the epidemic; if the total exposure reaches l_i, the individual i becomes infected. The possible dependence between l_i and r_i for each fixed i reflects the fact that both the resistance to infection and the duration of the infection may, for a fixed individual, depend on its physical constitution.

An initially susceptible individual i gets infected as soon as a certain functional, depending on the course of the epidemic, exceeds the individual's level l_i of resistance; denote its infection time by A_i^K. If $Z_K(t)$ denotes the proportion of infected individuals present in the population at time $t \in \mathbf{R}_+$, then A_i^K is given by

$$A_i^K = \inf \Big\{ t \in \mathbf{R}_+ : \int_{(0,t]} \lambda(s, Z_K)\, ds = l_i \Big\}, \tag{26}$$

for a certain function λ. Thus A_i^K takes values in $(0, \infty]$.

Since, for epidemics, the length of the infectious period of an individual is usually very small compared to its life length, we neglect births and removals that are not caused by the disease, as well as any age-dependence of the infectivity or the susceptibility. Furthermore, the population is idealized to be homogeneously mixing. Despite its simplicity, there are many useful applications of the model (cf. Berard *et al.* [22], Bailey [5]).

In special cases, there are already some asymptotic results for the proportion of susceptible and infectious individuals. However, the previous results were obtained for cases where l_i and r_i are independent and where the transition behaviour is "Markovian", i.e. $\mathcal{L}(l_1) = \exp(1)$. This case, in the special form $\lambda(t, x) = \lambda(x(t))$, was analyzed by Wang [76], [77]. For general λ, Solomon [72] has discussed a related, age-dependent population model that deals only with one class of individuals. The very special case $\lambda(t, x) = x(t)$ and $(\hat{r}_i), (r_i)$ being i.i.d. $\exp(\rho)$ yields the classical GSE, as constructed by Sellke [71].

Employing Stein's method we will not only investigate a more general model, but also describe the asymptotic evolution in a more detailed form. As the GSE is a birth-death process, as in the previous example the behaviour of an individual i can be described by the indicator $1_{[0,\hat{r}_i)}(t)$, if i is initially infected, or $1_{[A_i^K,\, A_i^K+r_i)}(t)$, if i is initially susceptible. A typical feature of investigation for the GSE is the asymptotic behaviour of the proportion of infected individuals and the proportion of susceptible individuals, as the population size tends to infinity. More generally, we would like to obtain the asymptotic average path behaviour, described via the limit behaviour of the empirical measure

$$\xi_K = \frac{1}{K}\sum_{i=1}^{aK}\delta_{(0,\hat{r}_i)} + \frac{1}{K}\sum_{i=1}^{bK}\delta_{(A_i^K,\, A_i^K+r_i)},$$

considered as a substochastic measure on $[0,\infty)^2$. The proportion of infected individuals and the proportion of susceptible individuals can be reconstructed via ξ_K. If $t \geq 0$, then $\xi_K([0,t]\times(t,\infty))$ describes the proportion of individuals, with respect to the total population, that are infected at time t, and $\xi_K(\mathbf{R}_+ \times [0,t])$ gives the proportion of individuals that are removed at time t. However, ξ_K contains even more information. For instance,

$$\xi_K([0,s]\times(t,\infty)),\ \ t > s,$$

gives the proportion of individuals that were infected up to time s and are not removed up to time t, that is the infectivity at time t in the population resulting from individuals that were infected before time s. Thus, by investigating ξ_K we gain new insights concerning the behaviour of the epidemic.

We need some more notation. Let Ψ be the common distribution function of the $(l_i)_{i\in\mathbf{N}}$, let Φ be the common distribution function of the $(r_i)_{i\in\mathbf{N}}$, and let $(\hat{\Phi}_i)_{i\in\mathbf{N}}$ be the distribution functions of $(\hat{r}_i)_{i\in\mathbf{N}}$. Let D_+ be the space of all functions $x : [0,\infty) \to [-1,1]$ that are right continuous with left-hand limits, and let $\lambda : \mathbf{R}_+ \times D_+ \to \mathbf{R}_+$ be the "accumulation" function determining the infection mechanism. Recall, for an initially susceptible individual i, its infection time A_i^K is given by (26), with

$$Z_K(t) = \frac{1}{K}\sum_{j=1}^{aK} 1_{[0,\hat{r}_j)}(t) + \frac{1}{K}\sum_{j=1}^{bK} 1_{[A_j^K, A_j^K+r_j)}(t)$$

being the proportion of infected individuals present in the population at time $t \in \mathbf{R}_+$. (We use the notation $1_C(t)$ to denote the indicator function on the set C; the notation $I[t \in C]$ refers to the indicator of a set, not considered as a function.) This completes the description of the model. Furthermore, we make some technical assumptions.

1. There is a probability measure $\hat{\mu}$ on $\mathbf{R}_+$ such that for all $T \in \mathbf{R}_+$,

$$\sup_{0\leq t\leq T}\left|\frac{1}{aK}\sum_{i=1}^{aK}\mathbf{P}[\hat{r}_i \leq t] - \hat{\mu}([0,t])\right| \to 0\ \ (K\to\infty);$$

denote its distribution function by $\hat{\Phi}$.

2. The function $\lambda : \mathbf{R}_+ \times D_+ \to \mathbf{R}_+$ satisfies for all $t \in \mathbf{R}_+$, $x, y \in D_+$
   a) $\lambda(t,x) = \lambda(t, x_t)$, where, for $t, u \in \mathbf{R}_+$, $x \in D_+$, $x_t(u) = x(t \wedge u)$;
   b) there is a positive constant α such that
   $$|\lambda(t,x) - \lambda(t,y)| \le \alpha \sup_{0 \le s \le t} |x(s) - y(s)|;$$
   c) there is a positive constant γ such that $\sup_{0 \le s \le t} \lambda(s,x) \le \gamma$.
3. There is a positive constant β such that, for each $x \in \mathbf{R}_+$, $\Psi_x(t) := \mathbf{P}[l_1 \le t | r_1 = x]$ fulfills, for all $s, t \in \mathbf{R}_+$,
   $$|\Psi_x(t) - \Psi_x(s)| \le \beta |t - s|.$$

The first task is to determine the limiting measure. We define, for $f \in C(\mathbf{R}_+, \mathbf{R}), t \in \mathbf{R}_+$, an operator $\mathcal{Z}$ and an operator L:

$$\begin{aligned} \mathcal{Z}f(t) &= a(1 - \hat{\Phi}(t)) + b\Psi(f(t)) - b\int_{(0,t]} \Psi_x(f(t-x))\mathbf{P}[r_1 \in dx], \\ Lf(t) &= \int_{(0,t]} \lambda(s, \mathcal{Z}f)ds \end{aligned}$$

(as $\mathcal{Z}f \in D_+$, the latter expression is well defined). Let $\|f\|_T = \sup_{s \le T} |f(s)|$ denote the supremum norm on $C([0,T])$. A contraction argument can be employed to prove the following assertion (see Reinert [63]).

Theorem 16 *For $T \in \mathbf{R}_+$, the equation*

$$f(t) = \int_{(0,t]} \lambda(s, \mathcal{Z}f)ds, \quad 0 \le t \le T, \tag{27}$$

has a unique solution G_T.

Now we restrict the observations to finite intervals $[0,T] \times [0,T]$ for a $T \in \mathbf{R}_+$ arbitrary, fixed. This leads to some notation. For $T \in \mathbf{R}_+$, put $[0,T]^2 = [0,T] \times [0,T]$, and $\mathcal{B}_T = \mathcal{B}([0,T]^2)$. Let $\nu \in M_1(\mathbf{R}^2)$, then

$$\nu^T \quad = \quad \nu|_{\mathcal{B}_T}$$

is the restriction of ν on $\mathcal{B}_T$ (hence, $\nu^T \in M_1([0,T]^2)$). For $A \in \mathcal{B}(\mathbf{R}^2)$, put

$$\nu^T(A) = \nu(A \cap [0,T]^2);$$

this defines ν^T also on $\mathcal{B}(\mathbf{R}^2)$. If in addition X is a random element with $\mathcal{L}(X) = \nu$, then, for all $T \in \mathbf{R}_+, f \in L_1(\nu), A \in \mathcal{B}(\mathbf{R}^2)$,

$$\begin{aligned} \mathbf{E}^T f(X) &= \int f(x)\nu^T(dx) \\ \mathbf{P}^T[X \in A] &= \int 1_A(x)\nu^T(dx) \\ \mathcal{L}^T f(X) &= \mathcal{L}(f(X))|_{\mathcal{B}_T} \end{aligned}$$

are the corresponding restrictions. With this notation we identify the following limiting measure.

Theorem 17 *For $T \in \mathbf{R}_+$, let G_T be the solution of (27) and $\tilde{\mu}^T \in M_1(E)$ be given for $r, s \in (0, T]$ by*

$$\tilde{\mu}^T([0,r] \times [0,s]) \quad = \quad \mathbf{P}^T[l_1 \leq G_T(r), l_1 \leq G_T(s - r_1)].$$

Put

$$\mu^T = a(\delta_0 \times \hat{\mu})^T + b\tilde{\mu}^T.$$

Then

$$\frac{1}{K}\sum_{i=1}^{aK} \mathcal{L}^T((0, \hat{r}_i)) + \frac{1}{K}\sum_{i=1}^{bK} \mathcal{L}^T((A_i^K, A_i^K + r_i)) \Rightarrow_v \mu^T \quad (K \to \infty).$$

Using Stein's method it can be seen that the following holds. The complete proof is in Reinert [63]; below we give a brief sketch.

Theorem 18 *Let μ^T be as in Theorem 17. Then, for all $T \in \mathbf{R}_+$,*

$$\mathcal{L}(\xi_K^T) \overset{w}{\Rightarrow} \delta_{\mu^T} \quad (K \to \infty).$$

Sketch of Proof of Theorem 18

It suffices to show that for all $T \in \mathbf{R}_+$, for all $m \in \mathbf{N}$, for all $f \in C_{\mathrm{b}}^\infty(\mathbf{R}^m)$, and for all $\phi_1, \ldots, \phi_m, \psi \in C_{\mathrm{b}}^\infty([0,T]^2)$

$$\mathbf{E}f(\langle \xi_K^T, \phi_1 \rangle, \ldots \langle \xi_K^T, \phi_m \rangle)\langle \mu^T - \xi_K^T, \psi \rangle \to 0 \quad (K \to \infty).$$

Let $f, \phi_1, \ldots, \phi_m, \psi$ be as above and $\mu^T, \hat{\mu}^T, \tilde{\mu}^T$ as in Theorem 17. Then

$$\begin{aligned} &\mathbf{E}f(\langle \xi_K^T, \phi_1 \rangle, \ldots \langle \xi_K^T, \phi_m \rangle)\langle \mu^T - \xi_K^T, \psi \rangle \\ &= \quad \mathbf{E}f(\langle a(\delta_0 \times \hat{\mu})^T + \frac{1}{K}\sum_{i=1}^{bK} \delta^T{}_{(A_i^K, A_i^K + r_i)}, \phi_l \rangle, l = 1, \ldots, m) \\ &\qquad \langle \mu^T - \xi_K^T, \psi \rangle + R_1 \end{aligned}$$

with

$$\begin{aligned} R_1 \quad = \quad & \mathbf{E}\Big[(f(\langle \xi_K^T, \phi_l \rangle, l = 1, \ldots, m) - f(\langle a(\delta_0 \times \hat{\mu})^T \\ & + \frac{1}{K}\sum_{i=1}^{bK} \delta^T{}_{(A_i^K, A_i^K + r_i)}, \phi_l \rangle, l = 1, \ldots, m))\langle \mu^T - \xi_K^T, \psi \rangle\Big]. \end{aligned}$$

Now let $\delta > 0$ be arbitrary, fixed, and put

$$H_T(s,t) = \tilde{\mu}^T([0,s] \times [0,t])$$

and

$$\begin{aligned}\Delta_{k,m}H_T &= H_T((k+1)\delta,(k+m+1)\delta)-H_T((k+1)\delta,(k+m)\delta)\\ &\quad -H_T(k\delta,(k+m+1)\delta)+H_T(k\delta,(k+m)\delta)\\ &= \tilde{\mu}^T([k\delta,(k+1)\delta)\times[(k+m)\delta,(k+m+1)\delta)).\end{aligned}$$

Then

$$\begin{aligned}&\mathbf{E}f(\langle\xi_K^T,\phi_l\rangle,l=1,\dots,m)\langle\mu^T-\xi_K^T,\psi\rangle\\ &= \mathbf{E}f(\langle a(\delta_0\times\hat{\mu})^T+b\sum_{k,m=0}^{\infty}\Delta_{k,m}H_T\delta^T{}_{(k\delta,(k+m)\delta)},\phi_l\rangle,l=1,\dots,m)\\ &\quad \langle\mu^T-\xi_K^T,\psi\rangle+R_1+R_2,\end{aligned}$$

where

$$\begin{aligned}R_2 &= \mathbf{E}\Big\{f(\langle a(\delta_0\times\hat{\mu})^T+\frac{1}{K}\sum_{i=1}^{bK}\delta^T{}_{(A_i^K,A_i^K+r_i)},\phi_l\rangle,l=1,\dots,m)\\ &\quad -f(\langle a(\delta_0\times\hat{\mu})^T+b\sum_{k,m=0}^{\infty}\Delta_{k,m}H_T\delta^T{}_{(k\delta,(k+m)\delta)},\phi_l\rangle,l=1,\dots,m)\Big\}\\ &\quad \langle\mu^T-\xi_K^T,\psi\rangle.\end{aligned}$$

Using similar approximations for ξ_K^T in $\langle\mu^T-\xi_K^T,\psi\rangle$, we obtain

$$\begin{aligned}&\mathbf{E}f(\langle\xi_K^T,\phi_l\rangle,l=1,\dots,m)\langle\mu^T-\xi_K^T,\psi\rangle\\ &= \mathbf{E}f(\langle a(\delta_0\times\hat{\mu})^T+b\sum_{k,m=0}^{\infty}\Delta_{k,m}H_T\delta^T{}_{(k\delta,(k+m)\delta)},\phi_l\rangle,l=1,\dots,m)\\ &\quad \langle b\tilde{\mu}^T-b\sum_{k,m=0}^{\infty}\Delta_{k,m}H_T\delta^T{}_{(k\delta,(k+m)\delta)},\psi\rangle+R_1+R_2+R_3+R_4,\end{aligned}$$

where

$$\begin{aligned}R_3 &= \mathbf{E}f(\langle a(\delta_0\times\hat{\mu})^T+b\sum_{k,m=0}^{\infty}\Delta_{k,m}H_T\delta^T{}_{(k\delta,(k+m)\delta)},\phi_l\rangle,l=1,\dots,m)\\ &\quad \langle a(\delta_0\times\hat{\mu})^T-\frac{1}{K}\sum_{i=1}^{aK}\delta^T_{(0,\hat{r}_i)},\psi\rangle,\\ R_4 &= \mathbf{E}f(\langle a(\delta_0\times\hat{\mu})^T+b\sum_{k,m=0}^{\infty}\Delta_{k,m}H_T\delta^T{}_{(k\delta,(k+m)\delta)},\phi_l\rangle,l=1,\dots,m)\\ &\quad \langle b\tilde{\mu}^T-\frac{1}{K}\sum_{i=1}^{bK}(\sum_{k,m=0}^{\infty}\Delta_{k,m}H_T\delta^T{}_{(k\delta,(k+m)\delta)}-\delta^T{}_{(A_i^K,A_i^K+r_i)}),\psi\rangle.\end{aligned}$$

The last term is deterministic: put

$$
\begin{aligned}
R_5 &= f(\langle a(\delta_0 \times \hat{\mu})^T + b \sum_{k,m=0}^{\infty} \Delta_{k,m} H_T {\delta^T}_{(k\delta,(k+m)\delta)}, \phi_l \rangle, l = 1, \dots, m) \\
& \langle b\tilde{\mu}^T - b \sum_{k,m=0}^{\infty} \Delta_{k,m} H_T {\delta^T}_{(k\delta,(k+m)\delta)}, \psi \rangle.
\end{aligned}
$$

It remains to show that the remainders tend to zero, as $\delta \to 0$ and $K \to \infty$. Here the rate of convergence will depend on the rate of convergence in the Glivenko-Cantelli Theorem.

4 Examples for Gaussian Approximations

4.1 Iterates of Random Maps

This example has been treated in Barbour *et al.* [16]. Let $h : I = [0,1] \to I$ be piecewise monotone C^1 and uniformly expanding: that is, there is a finite set $U = U(h)$ of points

$$0 = u_0 < u_1 < \cdots < u_{m_1} < u_{m_1+1} = 1$$

in I such that, for each interval $J_i = J_i(h) = (u_{i-1}, u_i)$, both h restricted to J_i and its continuous extension to $[u_{i-1}, u_i]$ are C^1 and monotone, satisfying

$$1 < c(h) \le |h'(x)| \le C(h) < \infty$$

for all x. We denote the rth iterate of h by h_r. A (measurable) set A is said to have period r if $h_r(A) \doteq A$, where $A \doteq B$ means that $\lambda(A_\triangle B) = 0$ and λ denotes Lebesgue measure; if A has period 1, it is called invariant. An invariant measure is a measure μ such that $\mu(h^{-1}(A)) = \mu(A)$ for all A. We assume the following.

1. There are no periodic sets A with $0 < \lambda(A) < 1$.
2. There exists an $r = r_1 \ge 1$ for which $|h_r^{-1}(x)| \ge 4$ for all $x \notin h_r(U(h_r))$.
3. h' is piecewise Hölder continuous with exponent ζ, for some $0 < \zeta \le 1$.

First in Lasota and Yorke [50] and then, more generally, in Keller [49], it is shown that under these assumptions, there exists at least one invariant probability measure μ which is absolutely continuous with respect to λ, and the density f of μ is of bounded variation.

Our interest lies mainly in the extent to which the properties of the h–sequence $\{h_r(x_0), r \ge 0\}$ mimic those of a more conventional stochastic process, when x_0 is suitably chosen at random. If x_0 is exactly known, the whole future of the h–sequence is completely determined, and randomness does not enter at all. However, in practice, x_0 can never be known without error, and the small uncertainty in the value of x_0 has an enormous effect on the later values in the sequence. A Gaussian approximation for the empirical process of these iterates is one example of the parallels to conventional stochastic processes.

First observe that if x_0 is chosen at random according to the invariant measure μ, then the h–sequence is a stationary Markov chain taking values in I, as is its time–reversal; see, for instance, the references given in Isham (1993, Section 3.6.3). An advantage of considering the time-reversed process is that randomness enters progressively at each step, and not only when setting the initial state x_0, making the analogy with classical stochastic processes clearer. Secondly observe that the time reversal of the h–sequence of a uniformly expanding map has an induced contraction property, which enables coupling methods to be introduced. It is shown in Barbour *et al.* [16] that the first steps in a time-reversed chain starting in x_0 and in one starting in x_0' can typically be realized in such a way that, with high probability, after one time step the chains take the values x_1 and x_1', respectively, such that $x_1 = \phi(x_0)$ and $x_1' = \phi(x_0')$ for the same branch ϕ of h^{-1}. If this is the case, then

$$|x_1 - x_1'| = |\phi(x_0) - \phi(x_0')| \leq |x_0 - x_0'| \sup_{y \in I}\{1/|h'(y)|\} = c(h)^{-1}|x_0 - x_0'|,$$

and the positions of the two chains after one step are closer than they were initially, at least by a geometric factor of $c(h)^{-1} < 1$. With some effort, Barbour *et al.* [16] prove that, however two time-reversed chains $(Y_n,\ n \geq 0)$ and $(Y_n',\ n \geq 0)$ are started, they can be realized simultaneously in such a way that $|Y_n - Y_n'| \leq Zc(h)^{-n}$ for all n, where Z is a random variable with Pareto tail.

Theorem 19 *There is a constant $K > 0$ and there are constants $0 < \alpha, \beta < 1$, (related to the function h,) such that, if $m = \max\{(2+\frac{3}{\beta})/\log c, 2/\log\frac{1}{\alpha}\}\log N$, then*

$$\begin{aligned}\zeta_{\mathcal{F}_{C,f}}(\xi_N, \zeta_N) &\leq 20m^2N^{-\frac{1}{2}} + KN^{-3/2} + N^{-2}\left\{\frac{K}{1-\alpha} + 4(m+1)K\right\} \\ &= O(N^{-\frac{1}{2}}\log^2 N).\end{aligned}$$

A proof of this theorem can be found in Barbour *et al.* [16]. The main tools used are, firstly, the exponential decay of correlations. Suppose that u_1 and u_2 are smooth integrable functions. For any $g : [0,1] \to \mathbb{R}$, we define

$$\begin{aligned}\bar{m}(g,\eta) &= \int_0^1 \sup_{\{z:|x-z|<\eta\}\cap T} |g(z) - g(x)|\, dx; \\ m_\gamma(g) &= \sup_{0<\eta\leq 1} \eta^{-\gamma}\bar{m}(g,\eta); \\ m_0(g) &= \sup_{x,y\in T}|g(x) - g(y)| = \bar{m}(g,1).\end{aligned}$$

Here T is a set of Lebesgue measure 0 that accounts for the jumps in h. Suppose that X_0 has a density g_0 satisfying $m_\gamma(g_0) < \infty$. Then, for some $K < \infty$, under smoothness conditions on the function g we have (see Barbour *et al.* [16])

$$\begin{aligned}
&\Big|\mathbf{E}\{u_1(X_0)u_2(h_n(X_0))\} - \int_I u_1(x)g_0(x)\,dx \int_I u_2(x)f(x)\,dx\Big| \\
&\leq \; K\alpha^n \int_I |u_2(x)|\,dx\Big\{\int_I |u_1(x)|g_0(x)\,dx + m_{\gamma'}(g_0)m_{\gamma'}(|u_1|)\Big\}.
\end{aligned}$$

The second ingredient is a coupling of the time reversal $(Y_1, \ldots, Y_n)$ of $(X_1, \ldots, X_n)$ with an m-dependent process $(Y_1', \ldots, Y_n')$. In particular, in Barbour *et al.* [16] it is shown that for each $n \geq 0$, we have

$$\mathbf{P}[|Y_{n+m}' - Y_{n+m}| \geq xc^{-m}] \leq Kx^{-\beta}, \quad x > 0,$$

so that, for instance,

$$\mathbf{P}\Big[\bigcup_{j=m}^{N+m}\{|Y_j' - Y_j| \geq N^{-2}\}\Big] \leq KN^{-2}$$

whenever $m \geq (2 + 3/\beta)\log N/\log c$, where $c = c(h)$.

Once this m-dependent approximation has been proved, all that remains is to apply the results for the local approach. As this gives an explicit bound on the distance, it is possible to optimize the choice of m.

4.2 Simple Random Sampling

Let $\mathcal{A} = \{a_1, \ldots, a_N\}$ be a set of nonnegative numbers such that $\sum_{a\in\mathcal{A}} a > 0$. Let us assume that the elements of $\mathcal{A}$ are distinct. Consider the random vector $\mathbf{X} = (X_1, \ldots, X_n)$ obtained by a simple random sample of size n from $\mathcal{A}$, that is, $\mathbf{X}$ is a realization of one of the equally likely $N_{(n)}$ n-dimensional vectors of distinct elements of $\mathcal{A}$ ($n \leq N$). Let ξ_n be the corresponding empirical measure. The construction of the zero bias coupling for real-valued random elements is given in Goldstein and Reinert ([43]). In the construction from Section 1, because of exchangeability we may choose $I = 1$. Independently of $X_1, \ldots, X_n$, pick a pair $(\hat{X}_1', \hat{X}_1'')$ from the distribution

$$q(u, v) = \frac{(u - v)^2}{2N}\mathbf{1}(\{u, v\} \subset \mathcal{A}).$$

Firstly then, independently of the chosen sample $\mathbf{X}$, pick $(\hat{X}_1', \hat{X}_1'')$ from the distribution $q(u, v)$. The random variables $(\hat{X}_1', \hat{X}_1'')$ are now placed as the first two components in the vector $\hat{\mathbf{X}}$. The remaining $n - 1$ random variables $\hat{\mathbf{X}}$ are sampled by rejection. If the two sets $\{X_2, \ldots, X_n\}$ and $\{\hat{X}_1', \hat{X}_1''\}$ do not intersect, fill in the remaining $n - 1$ components of $\hat{\mathbf{X}}$ with $(X_2, \ldots, X_n)$. If the sets have an intersection, remove from the vector $(X_2, \ldots, X_n)$ the two random variables (or single random variable) that intersect and replace them (or it) with values obtained by a simple random sample of size two (one) from

$\mathcal{A} \setminus \{\hat{X}_1', \hat{X}_1'', X_2, \ldots, X_n\}$. This new vector now fills in the remaining $n-1$ positions in $\hat{\mathbf{X}}$. In Goldstein and Reinert [43] it is shown that this construction yields a bound of order $\frac{1}{n}$ for the normal approximation of $W = \sum_{i=1}^n X_i$, provided the elements of $\mathcal{A}$ are scaled such that $a_i = \frac{c_i}{\sqrt{n}}$, where $c_1, \ldots, c_N$ do not depend on n.

Now we apply this construction to empirical measures. Pick a function $\phi \in \mathcal{H}$, where $\mathcal{H}$ is the set of all function s $\phi \in \mathcal{C}$ such that $\mathbf{E}\langle \xi, \phi \rangle = 0$. Instead of considering the vector $(X_1, \ldots, X_n)$, we now consider the vector $(\phi(X_1), \ldots, \phi(X_n))$. For simplicity assume that $\phi(x), x \in \mathcal{A}$, are all distinct values. Then we may use the above to construct

$$\mathbf{X}_{ii}^* = (\phi_i(X_1), \ldots, \phi_i(X_n))^*$$

The empirical measure of this vector is ξ_{n,ϕ_i,ϕ_i}^*. One can show that $(\mathbf{X}_{ij}^*)$ for $i \neq j$ can be attained simply by taking the equal mixture of $\mathbf{X}_{ii}^*$ and $\mathbf{X}_{jj}^*$. This completes the construction of $\{\xi_{n,\phi,\psi}^*, \psi, \phi \in \mathcal{H}\}$. As in each step, at most two of the underlying random elements are changed, Theorem 11 yields a bound of $6n^{-1/2}$ on the distance to the corresponding Gaussian random measure.

5 Joint Occurrences of Multiple Words in DNA Sequences: An Example for Poisson Approximation

A Poisson point process approximation can be seen as a multivariate Poisson process approximation for the process of indicators. There are a multitude of examples for this approach, see for example Arratia *et al.* [3]. Here we give only a recent one, for which the analysis has been carried out by Reinert and Schbath [65].

Consider a stationary Markov chain $\mathcal{M} = \{X_i\}_{i \in \mathbf{Z}}$ on a finite alphabet $\mathcal{A}$, with transition matrix Π such that $\Pi(x,y) > 0$ for all $x, y \in \mathcal{A}$. This implies that the Markov chain has a unique stationary distribution μ defined by

$$\mu(x) = \sum_{y \in \mathcal{A}} \mu(y)\Pi(y,x) \text{ for all } x \in \mathcal{A}.$$

Let $\underline{u} = u_1 u_2 \cdots u_\ell$ be a word of length ℓ on $\mathcal{A}$. Say that an occurrence of $\underline{u}$ starts at position i in the infinite sequence $\mathcal{M}$ if $X_i X_{i+1} \cdots X_{i+\ell-1} = u_1 u_2 \cdots u_\ell$, and denote the indicator random variable of this event by $\mathbf{I}_i(\underline{u})$. Thus the expectation of $\mathbf{I}_i(\underline{u})$ is given by $\mu(\underline{u}) = \mu(u_1)\Pi(u_1,u_2)\cdots\Pi(u_{\ell-1},u_\ell)$. Note that in reality we only observe a finite segment $S = X_1 X_2 \cdots X_n$ of the infinite sequence $\mathcal{M}$. In what follows we will be concerned with counting the joint number of occurrences of words $\underline{u}$ and $\underline{v}$ in S; for simplicity we assume that they have the same length. A more general case is treated in Reinert and Schbath [65]. Through this section the example $S =$ `TAAGAAGAAGAAGAAGT` and $\underline{u} =$ `AAGAAGAA` is used. In this case, the word $\underline{u}$ occurs in S at positions 2, 5 and 8.

In order to study the occurrences of a word $\underline{u}$ the concept of clumps of a word $\underline{u}$ is introduced. A *clump* of $\underline{u}$ in a sequence is a maximal set of overlapping occurrences of $\underline{u}$ in this sequence; no two clumps of $\underline{u}$ overlap in the sequence. Say that a clump of $\underline{u}$ starts at position i in the infinite sequence $\mathcal{M}$ if an occurrence of $\underline{u}$ starts at position i in $\mathcal{M}$ and if this occurrence is not overlapped by a preceding occurrence of $\underline{u}$. Denote the corresponding indicator random variable by $\widetilde{\mathbf{I}}_i(\underline{u})$; i.e.

$$\widetilde{\mathbf{I}}_i(\underline{u}) \quad = \quad \mathbf{I}_i(\underline{u}) \prod_{j=i-\ell+1}^{i-1} (1 - \mathbf{I}_j(\underline{u})).$$

Denote by $\widetilde{\mu}(\underline{u})$ the probability that a clump of $\underline{u}$ starts at a given position in $\mathcal{M}$.

Now let $\underline{u} = u_1 u_2 \cdots u_\ell$ and $\underline{v} = v_1 v_2 \cdots v_\ell$ be two different words of length ℓ on $\mathcal{A}$. To describe the possible overlaps between $\underline{u}$ and $\underline{v}$, we define

$$\mathcal{P}(\underline{u}, \underline{v}) \quad := \quad \{p \in \{1, \ldots, \ell - 1\} : v_i = u_{i+p}, \text{ for all } i = 1, \ldots, \ell - p\}.$$

Thus $\mathcal{P}(\underline{u}, \underline{v}) \neq \emptyset$ means that an occurrence of $\underline{v}$ can overlap an occurrence of $\underline{u}$ from the right, and $\mathcal{P}(\underline{v}, \underline{u}) \neq \emptyset$ means that $\underline{v}$ can overlap $\underline{u}$ from the left. Note the lack of symmetry; for example, if $\underline{u} = \texttt{AAGAAGAA}$ and $\underline{v} = \texttt{AAGAATCA}$, we have $\mathcal{P}(\underline{u}, \underline{v}) = \{3, 6, 7\}$ and $\mathcal{P}(\underline{v}, \underline{u}) = \{7\}$.

Firstly, we define the Bernoulli and the Poisson processes that will be used to apply Theorem 6. We form the random vector $\underline{\widetilde{Y}} = (\widetilde{Y}_i)_{i \in I}$, with index set $I = \{1, 2, \ldots, 2(n - \ell + 1)\}$, as

$$\widetilde{Y}_i = \begin{cases} \widetilde{\mathbf{I}}_i(\underline{u}) & \text{if } 1 \le i \le n - \ell + 1\,, \\ \widetilde{\mathbf{I}}_{\check{\imath}}(\underline{v}) & \text{if } n - \ell + 1 < i \le 2(n - \ell + 1)\,, \end{cases}$$

with the notation

$$\check{\imath} = \begin{cases} i & \text{if } 1 \le i \le n - \ell + 1\,, \\ i - n + \ell - 1 & \text{if } n - \ell + 1 < i \le 2(n - \ell + 1)\,. \end{cases} \tag{28}$$

Thus $\underline{\widetilde{Y}}$ is the concatenated vector of the occurrence indicators for clumps of $\underline{u}$ and clumps of $\underline{v}$. Note that we do not consider "mixed" clumps of $\underline{u}$ and $\underline{v}$; each clump consists only in occurrences of $\underline{u}$, or only in occurrences of $\underline{v}$. However, a clump of $\underline{u}$ may overlap a clump of $\underline{v}$ in the sequence.

Let $\underline{Z} = (Z_i)_{i \in I}$ be a random vector of independent Poisson variables such that

$$Z_i \sim \begin{cases} \mathrm{Po}\,(\widetilde{\mu}(\underline{u})) & \text{if } 1 \le i \le n - \ell + 1\,, \\ \mathrm{Po}\,(\widetilde{\mu}(\underline{v})) & \text{if } n - \ell + 1 < i \le 2(n - \ell + 1). \end{cases}$$

To apply Theorem 6, we choose the following neighborhood of $i \in I$:

$$B_i := \{j \in I : |\check{\jmath} - \check{\imath}| \le 3\ell - 3\}\,,$$

where the notation $\check{\jmath}$ is defined in (28). Define the quantities

$$\begin{aligned}
\Omega &= \sum_{s=1}^{2\ell-2} \Pi^s, \\
M(\underline{u},\underline{v}) &= \sum_{p \in \mathcal{P}(\underline{u},\underline{v})} \frac{1}{\mu(\underline{v}^{(\ell-p)})} \mathbf{1}(\underline{u} \neq \underline{v}), \\
T_1(\underline{u},\underline{v}) &= 2(n-\ell+1)\mu(\underline{u})\widetilde{\mu}(\underline{v}) \left(\frac{\Omega(v_\ell, u_1)}{\mu(u_1)} + M(\underline{v},\underline{u}) \right).
\end{aligned}$$

The quantities $M(\underline{u},\underline{v})$ and $M(\underline{v},\underline{u})$ can be seen as measures of the overlapping structure between $\underline{u}$ and $\underline{v}$. If $\underline{u}$ and $\underline{v}$ cannot overlap, these quantities are equal to zero; otherwise, the more they can overlap, the larger are $M(\underline{u},\underline{v})$ and $M(\underline{v},\underline{u})$.

Moreover, we need some quantities related to the transition matrix. Let $(\alpha_t)_{t=1,\ldots,|\mathcal{A}|}$ be the eigenvalues of Π such that $|\alpha_1| \geq |\alpha_2| \geq \cdots \geq |\alpha_{|\mathcal{A}|}|$. Then $\alpha_1 = 1$ and $|\alpha_2| < 1$; abbreviate α_2 by α. Define matrices $Q_t, t = 1, \ldots, |\mathcal{A}|$ satisfying the decomposition

$$\Pi = \sum_{t=1}^{|\mathcal{A}|} \alpha_t Q_t.$$

Define $\gamma_1(\ell, a)$ for any $a \in \mathcal{A}$ by

$$\begin{aligned}
&\gamma_1(\ell, a) \\
&= \sum_{x,y \in \mathcal{A}} \mu(x) \max_{b \in \mathcal{A}} \left| \frac{1}{\mu(b)} \sum_{(t,t') \neq (1,1)} \frac{\alpha_t^\ell \alpha_{t'}^\ell}{\alpha^\ell} Q_t(x,b) Q_{t'}(a,y) \right. \\
&\qquad \left. - \sum_{t=2}^{|\mathcal{A}|} \frac{\alpha_t^{4\ell-2}}{\alpha^\ell} Q_t(x,y) \right|.
\end{aligned}$$

Note that $\gamma_1(\ell, a)$ depends on ℓ and is bounded by a constant. Then Reinert and Schbath [65] showed the following theorem.

Theorem 20 *With the previous notation, we have*

$$\begin{aligned}
&d_{TV}\left(\mathcal{L}(\widetilde{\underline{Y}}), \mathcal{L}(\underline{Z})\right) \\
&\leq (n-\ell+1)(6\ell-5)\big(\widetilde{\mu}(\underline{u}) + \widetilde{\mu}(\underline{v})\big)^2 \\
&\quad + T_1(\underline{u},\underline{u}) + T_1(\underline{u},\underline{v}) + T_1(\underline{v},\underline{u}) + T_1(\underline{v},\underline{v}) \\
&\quad + (n-\ell+1)|\alpha|^\ell \Big(\gamma_1(\ell, u_\ell)\widetilde{\mu}(\underline{u}) + \gamma_1(\ell, v_\ell)\widetilde{\mu}(\underline{v}) \Big).
\end{aligned}$$

If the X_i's are independent, then the bound simplifies to

$$\begin{aligned}
& d_{TV}\left(\mathcal{L}(\underline{\widetilde{Y}}), \mathcal{L}(\underline{Z})\right) \\
& \quad \leq \quad (n-\ell+1)(6\ell-5)\left(\widetilde{\mu}(\underline{u})+\widetilde{\mu}(\underline{v})\right)^2 \\
& \qquad\quad +4(n-\ell+1)(\ell-1)(\mu(\underline{u})+\mu(\underline{v}))(\widetilde{\mu}(\underline{u})+\widetilde{\mu}(\underline{v})) \\
& \qquad\quad +2(n-\ell+1)\left(\mu(\underline{u})\widetilde{\mu}(\underline{v})M(\underline{v},\underline{u})+\mu(\underline{v})\widetilde{\mu}(\underline{u})M(\underline{u},\underline{v})\right).
\end{aligned}$$

6 Open Problems

Stein's method for empirical measures can be a powerful tool in many examples where there is a dependence structure between the underlying random elements. In particular this applies to bootstrap procedures and to interacting particle systems (unpublished works with Larry Goldstein). Many more examples would be desirable, to illustrate the use of the method and of different couplings. Note that the couplings described above serve as examples, not as a complete list - in specific problems, other couplings might be more appropriate.

Moreover it would we very interesting to apply the method to processes requiring a time structure. The above results do not include tightness, so there might be additional work needed. In particular it might be interesting to investigate the quality of discrete approximations to measure-valued diffusion processes, as in Donnelly and Kurtz [35], for example. Finally it would be interesting but challenging to investigate the empirical measure process of superprocesses as described in Gorostiza [45].

Acknowledgements. I would like to thank the organizers for the opportunity to present these lectures. In particular I would like to thank Luis Gorostiza for many fruitful discussions. Moreover, the referee has made many helpful remarks. Lastly, the zero bias example for a simple random sample of random measures is based on unpublished work with Larry Goldstein.

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] Arratia, R., Goldstein, L., and Gordon, L. (1989). Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.* **17** 9–25.

[3] Arratia, R., Goldstein, L., and Gordon, L. (1990). Poisson approximation and the Chen-Stein method. *Statistical Science* **5** 403–434.

[4] Aldous, D. (1989). Stein's method in a two-dimensional coverage problem. *Statist. Probab. Lett.* **8**, 307 - 314.

[5] Bailey, N.T.J. (1975). *The mathematical theory of infectious diseases and its applications.* (2nd edn). Griffin, London.

[6] Baldi, P., and Rinott, Y. (1989). Asymptotic normality of some graph-related statistics. *J. Appl. Prob.* **26**, 171–175.

[7] Baldi, P., and Rinott, Y. (1989). On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.* **17**, 1646–1650.

[8] 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.* T. W. Anderson, K.B. Athreya and D. L. Iglehart, eds., Academic Press, 59–81.

[9] Barbour, A.D. (1986). Asymptotic expansions based on smooth functions in the Central Limit Theorem. *Probab. Theory Relates Fields* **72**, 289–303.

[10] Barbour, A.D. (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* **25 (A)**, 175–184.

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

[12] Barbour, A.D. (1997). Stein's method. *Encyclopaedia of Statistical Science,* Update, Vol.1, Wiley, New York, 513–521.

[13] Barbour, A.D., Chen, L.H.Y., and Loh, W.-L. (1992). Compound Poisson approximation for nonnegative random variables via Stein's method. *Ann. Probab.* **20**, 1843 - 1866.

[14] Barbour, A.D., and Eagleson, G.K. (1985). Multiple comparisons and sums of dissociated random variables. *Adv. Appl. Prob.* **17**, 147–162.

[15] Barbour, A.D., and Eagleson, G.K. (1986). Random association of symmetric assays. *Stoch. Analysis Applics* **4**, 239–281.

[16] Barbour, A.D., Gerrard, R., and Reinert, G. (1997). Iterates of expanding maps. Preprint.

[17] Barbour, A.D., and Hall, P. (1983). On bounds to the rate of convergence in the central limit theorem. *Bull. London Math. Soc.* **17**, 151–156.

[18] Barbour, A.D., and Hall, P. (1984). Stein's method and the Berry-Esséen theorem. *Austral. J. Statist.* **26 (1)**, 8–15.

[19] Barbour, A.D., Holst, L., and Janson, S. (1992). *Poisson Approximation.* Oxford Science Publications.

[20] Barbour, A.D., Karoński, M., and Ruciński, A. (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Comb. Theory*, Ser. B, **47**, 125–145.

[21] Barbour, A.D., and Utev, S. (1997). Compound Poisson approximation in total variation. Preprint.

[22] Berard, D.R., Burgess, J.W. and Coss, R.G. (1981). Plasticity of dendritic spine formation: a state-dependent stochastic process. *Intern. J. Neuroscience,* **13**, 93 - 98.

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

[24] Bolthausen, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrscheinlichkeitstheor. Verw. Geb.* **66**, 379–386.

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

[26] Cacoullos, T., Papathanasiou, V., and Utev, S. (1994). Variational inequalities with examples and an application to the central limit theorem. *Ann. Probab.* **22**, 1607–1618.

[27] Chen, L.H.Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3**, 534–545.

[28] Chen, L.H.Y. (1978). Two central limit problems for dependent random variables. *Z. Wahrscheinlichkeitstheor. Verw. Geb.* **43**, 223–243.

[29] Chen, L.H.Y. (1986). The rate of convergence in a central limit theorem for dependent random variables with arbitrary index set. *IMA Preprint Series* **243**, Univ. Minnesota.

[30] Chen, L.H.Y. (1996). Stein's method: some perspectives with applications. Preprint.

[31] Dawson, D.A. (1993). Measure-valued Markov processes. In *Ecole d'Eté de Probabilitéa de Saint-Flour 1991*, LNM 1541, Hennequin, P.L., eds., Springer-Verlag, 1–260.

[32] Dembo, A., and Rinott, Y. (1996). Some examples of normal approximations by Stein's method. In *Random Discrete Structures*, IMA volume 76, Aldous, D. and Pemantle, R., eds., Springer-Verlag, 25–44.

[33] Diaconis, P. (1989). An example for Stein's method. Stanford Stat. Dept. Tech. Rep.

[34] Diaconis, P., and Zabell, S. (1991). Closed form summation for classical distributions: variations on a theme of de Moivre. *Statistical Science* **6**, 284-302.

[35] Donnelly, P. and Kurtz, T. (1996). A countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Probab.* **24**, 698–742.

[36] Dudley, R.M. (1989). *Real analysis and probability.* Wadsworth & Brooks/ Cole, Belmont.

[37] 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.

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

[39] Ehm, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.* **11**, 7–16.

[40] Fulman, J. (1997). A Stein's method proof for the asymptotic normality of descents and inversions in the symmetric group. Preprint.

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

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

[43] Goldstein, L., and Reinert, G. (1997). Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.,* in print.

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

[45] Gorostiza, L. G. (1996). Asymptotic fluctuations and critical dimension for a two-level branching system. *Bernoulli* **2**, 109–132.

Springer-Verlag.

[46] Hall, P., and Barbour, A.D. (1984). Reversing the Berry-Esséen theorem. *Proc. AMS* **90**, 107–110.

[47] Ho, S.-T., and Chen, L.H.Y. (1978). An L_p bound for the remainder in a combinatorial central limit theorem. *Ann. Probab.* **6**, 231–249.

[48] Horowitz, J. (1990). A uniform law of large numbers and empirical central limit theorem for limits of finite populations. *Statistics & Probability Letters,* **10**, 159 - 166.

[49] Keller, G. (1985). Generalized bounded variation and applications to piecewise monotonic transformations. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **69,** 461–478.

[50] Lasota, A. and Yorke, J. A. (1973). On the existence of invariant measures for piecewise monotonic transformations. *Trans. Amer. Math. Soc.* **186,** 481–488.

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

[52] Loh, W.-L. (1992). Stein's method and multinomial approximation. *Ann. Appl. Probab.* **2**, 536–554.

[53] Loh, W.-L. (1997). Stein's method and Pearson Type 4 approximations. Research report.

[54] Luk, H.M. (1994). Stein's method for the gamma distribution and related statistical applications. Ph.D. thesis. University of Southern California, Los Angeles, USA.

[55] Mann, B. (1997). Stein's method for χ^2 of a multinomial. Preprint.

[56] Peköz, E. (1995). Stein's method for geometric approximation. Tech. Report #225, Stat. Dept. U. California. Berkeley.

[57] 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.

[58] Pollard, D. (1984). *Convergence of stochastic processes.* Springer, New York.

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

[60] Rama Krishnaiah, Y. S. (1990). On the Glivenko-Cantelli theorem for generalized empirical processes based on strong mixing sequences.

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

[62] Reinert, G. (1999). A weak law of large numbers for empirical measures via Stein‘s method. *Ann. Prob.,* **23**, 334 – 354.

[63] Reinert, G. (1995). The asymptotic evolution of the General Stochastic Epidemic. *Ann. Appl. Prob.,* **5**, 1061 – 1086.

[64] Reinert, G. (1997). A Gaussian approximation for random measures via Stein's method. Preprint.

[65] Reinert, G. and Schbath, S. (1998). Compound Poisson and Poisson process approximations for occurrences of multiple words in Markov chains. Preprint.

[66] Rinott, Y. (1994). On normal approximation rates for certain sums of dependent random variables. *J. Comp. Appl. Math.* **55**, 135–143.

[67] 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.

[68] Rinott, Y., and Rotar, V. (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the anti-voter model and weighted U-statistics. Preprint.

[69] Roos, M. (1994). Stein's method for compound Poisson approximation: the local approach. *Ann. Appl. Probab.* **4**, 1177–1187.

[70] Rudin, W. (1973). *Functional analysis.* McGraw-Hill, New York.

[71] Sellke, T. (1983). On the asymptotic distribution of the size of a stochastic epidemic. *J. Appl. Prob.,* **20**, 390 - 394.

[72] Solomon. W. (1987). Representation and approximation of large population age distributions using Poisson random measures. *Stoch. Proc. Appl.,* **26**, 237 - 255.

[73] 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. Univ. California Press, Berkeley.

[74] Stein, C. (1986). *Approximate Computation of Expectations.* IMS, Hayward, California.

[75] Stein, C. (1992). A way of using auxiliary randomization. In *Probability Theory,* L. H. Y. Chen, K. P. Choi, K. Hu and J.-H. Lou, eds., Walter de Gruyter, Berlin - New York, 159–180.

[76] Wang, F.S.J. (1975) Limit theorems for age and density dependent stochastic population models. *J. Math. Bio.,* **2**, 373 - 400.

[77] Wang, F.S.J. (1977) A central limit theorem for age-and-density-dependent population processes. *Stoch. Proc. Appl.,* **5**, 173 - 193.

[78] Yamamuro, S. (1974). *Differential calculus in topological linear spaces.* Springer, Berlin (= Lecture Notes in Mathematics 374).