The Annals of Probability
1995, Vol. 23, No. 1, 334–354

A WEAK LAW OF LARGE NUMBERS FOR EMPIRICAL MEASURES VIA STEIN'S METHOD[1]

BY GESINE REINERT

University of Zürich

Let E be a locally compact Hausdorff space with countable basis and let $(X_i)_{i \in \mathbb{N}}$ be a family of random elements on E with $(1/n)\sum_{i=1}^{n} \mathscr{L}(X_i) \overset{v}{\Rightarrow} \mu(n \to \infty)$ for a measure μ with $\|\mu\| \leq 1$. Conditions are derived under which $\mathscr{L}((1/n)\sum_{i=1}^{n} \delta_{X_i}) \overset{w}{\Rightarrow} \delta_\mu(n \to \infty)$, where δ_x denotes the Dirac measure at x. The proof being based on Stein's method, there are generalisations that allow for weak dependence between the X_i's. As examples, a dissociated family and an immigration–death process are considered. The latter illustrates the possible applications in proving convergence of stochastic processes.

Introduction. Stein [17], [16] developed a very elegant method for proving convergence in distribution of random variables toward a standard normal variable. This method was generalized by Chen [4] for the Poisson distribution, by Loh [8] for the multinomial distributions and by Barbour, Chen and Loh [3] for the compound Poisson distribution. The general procedure could be described as follows: Find a good characterization of the desired distribution in terms of an equation, that is, of the type

$$\mathscr{L}(X) = \mu \iff \mathbb{E}[\mathscr{A} f(X)] = 0 \quad \text{for all smooth functions } f,$$

where $\mathscr{A}$ is an operator associated with the distribution μ. Assume X to have distribution μ and consider the Stein equation

$$g(x) - \mathbb{E}[g(X)] = \mathscr{A} f(x), \qquad x \in \mathbb{R}. \tag{1}$$

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

$$\mathbb{E}[g(W)] - \mathbb{E}[g(X)] = \mathbb{E}[\mathscr{A} f(W)].$$

Hence, to estimate the proximity of W and X, it is sufficient to estimate $\mathbb{E}[\mathscr{A} f(W)]$ for all possible solutions f of (1). The aim is hence to be able to solve (1) for a class of functions g that is sufficiently large to obtain convergence in a known topology and rates of convergence in a known metric.

Barbour [1] suggested employing as operator $\mathscr{A}$ in (1) the generator of a Markov process, which then provides a way to look for solutions of (1). In

Received November 1992; revised May 1994.

[1]This work was supported in part by the DAAD and by the Schweizer Nationalfonds Grant 20-31262.91.

AMS 1991 *subject classifications*. 60F05, 60G57, 60K25, 62G30.

Key words and phrases. Weak law of large numbers, empirical measures, Stein's method, dissociated family, immigration–death process.

the following, this will be called the generator method. Suppose we find a Markov process $(X(t))_{t\geq 0}$ with generator $\mathscr{A}$ and unique stationary distribution μ, such that $\mathscr{L}(X(t)) \overset{w}{\Rightarrow} \mu$ as $t \to \infty$. Then, if a random variable X has distribution μ,

$$\mathbb{E}[\mathscr{A}f(X)] = 0$$

for all $f \in \mathscr{D}(\mathscr{A})$. Thus, a method for solving the Stein equation (1) is provided by Proposition 1.5 of Ethier and Kurtz ([6], page 9; for the argument, see [1]). Let $(T_t)_{t\geq 0}$ be the transition semigroup of the Markov process $(X(t))_{t\geq 0}$. Then, formally,

$$g(x) - \mathbb{E}[g(X)] = -\mathscr{A}\Big(\int_0^\infty T_u g\, du\Big).$$

Thus, $f = -\int_0^\infty T_u g\, du$ would be a solution of (1) if this expression exists and if $f \in \mathscr{D}(\mathscr{A})$. 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 toward Wiener measure [1] and for Poisson process approximations (see, e.g., [2]).

For a given distribution μ, there may be various Markov processes with μ as stationary distribution, and it is still not completely clear which process to take to obtain good results (though many people have a good intuition on it). So, there is a need for further examples.

One basic example is a degenerate distribution δ_a. In this context, the question naturally occurs as to whether laws of large numbers can be proven via Stein's method. This is the main topic of this paper. It will turn out that it is indeed possible to show a weak law of large numbers for empirical measures by much the same method one would apply for the weak law of large numbers for random variables (for the latter, see [13]). Thus we simultaneously extend the range of Stein's method and provide a class of new results for empirical measures.

For didactic reasons, we first show the theorem for independent random elements, though, in this case, there are Glivenko–Cantelli type results, which even give almost sure convergence. However, as is in general a great advantage of Stein's method, the proof is easily carried over to (arrays of) dependent random elements, and in these cases the assertions seem to be essentially new (Horowitz [7] gives examples of Glivenko–Cantelli type results). Another advantage of Stein's method is that we, quasiautomatically, also get rates of convergence.

Let E throughout this paper be a locally compact Hausdorff space with countable basis (l.c.c.b.) and let $(X_i)_{i\in\mathbb{N}}$ be a family of random elements with values in E. Assume that there is a measure μ with total mass $\|\mu\| \leq 1$ and $(1/n)\sum_{i=1}^n \mathscr{L}(X_i) \overset{v}{\Rightarrow} \mu(n \to \infty)$. Let $\xi_n = (1/n)\sum_{i=1}^n \delta_{X_i}$. Then we derive conditions under which $\mathscr{L}(\xi_n) \overset{w}{\Rightarrow} \delta_\mu(n \to \infty)$ holds: this is a weak law of large numbers (w.l.l.n.).

Examples are then given to illustrate the range of applications for weakly dependent random arrays, with special reference to stochastic processes. First, we consider a "uniformly" weakly dependent array $(X_{i_1,\dots,i_k})$ of random elements on E, where $k \in \mathbb{N}$ and

$$X_{i_1,\dots,i_k} = \psi_{i_1,\dots,i_k}(Y_{i_1},\dots,Y_{i_k}),$$

with $(Y_i)_i$ being independent random elements on a space $\mathscr{X}$ and $\psi_{i_1,\dots,i_k}$: $\mathscr{X}^k \to E$ measurable, that is, $(X_{i_1,\dots,i_k})_{i_1,\dots,i_k}$ is a dissociated family. The second example deals with an immigration–death process with population size n, arrival times $(A_i^n)_{i=1,\dots,n}$ and life spans $(Z_i)_{i\in\mathbb{N}}$; the life spans are assumed to be i.i.d. and independent of the arrival times, whereas dependence between the arrival times is allowed. With

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

under certain conditions, we obtain the convergence of $(1/n)\sum_{i=1}^n \delta_{X_i}$; this describes the asymptotic path behavior of the process. Thus, the w.l.l.n. is a new tool to prove process convergence, and this tool does not employ any Markov structure or martingales. A more powerful, but also much more complex example, is the general stochastic epidemic in [12]. There the method leads to a widening of the class of epidemic models for which a deterministic approximation can be proven.

The outline of the paper is as follows. In Section 1, the topological structure of the underlying spaces is described. Section 2 concerns several formulations of the w.l.l.n. First we treat the independent case and we give a heuristic argument for deriving the appropriate generator. Then we extend this result for various kinds of dependences. Afterward we derive an estimate for the rate of convergence in terms of a Zolotarev metric. Section 3 gives the two examples to illustrate the power of the results. Finally, Section 4 contains most of the detail of the proofs of the results of Section 1 and 2.

1. Topological structure. Let $\mathscr{E} = \mathscr{B}(E)$ be the Borel σ-field of E and let $M^b(E)$ be the space of all bounded Radon measures on E, equipped with the vague topology. For $\mu \in M^b(E)$, set $\|\mu\| = \sup_{A\in\mathscr{E}} |\mu(A)|$ and $M_1(E) = \{\mu \in M^b(E)\colon \mu \text{ positive}, \|\mu\| \le 1\}$. Then $M_1(E)$ is vaguely compact. Furthermore, $M^b(E)$ is a topological linear space over $\mathbb{R}$. Following Yamamuro [18], Gâteaux differentiability of a function $f\colon M^b(E) \to \mathbb{R}$ can be defined as follows.

DEFINITION 1.1. Let G be a topological linear space. Let A be open in $M^b(E)$ and $a \in A$. A function $f\colon A \to G$ is Gâteaux differentiable in a, if there exists a linear function $u_a\colon M^b(E) \to G$ such that

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}[f(a+\varepsilon\mu) - f(a) - u_a(\varepsilon\mu)] = 0,$$

for every direction $\mu \in M^b(E)$. If f is Gâteaux differentiable in every point $a \in A$, then f is called Gâteaux differentiable. We use the notation $D_{\mathscr{G}}(A;G)$

for the set of all Gâteaux-differentiable functions $f: A \to G$, and

$$f'(a)[\mu] = u_a(\mu)$$

denotes the derivative of f at the point a in direction μ.

Higher derivatives can be defined in the obvious way: we denote by $f^{(k)}(a)[\nu^{(k)}]$ the kth derivative of f in a, as a linear form, applied to the vector $\nu^{(k)} = (\nu, \dots, \nu) \in (M^b(E))^k$ and let $D^k_{\mathscr{G}}(A;G)$ denote the set of all k times Gâteaux-differentiable functions $f: A \to G$. For $f: A \to \mathbb{R}$, we have Taylor's Theorem. Gâteaux differentiability for functions $f: M^b(E) \to G$ can be defined in the same way, where A in the definition is replaced by $M^b(E)$. We need some more notations. Put

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

Then we define the following sets of functions:

$$C_b(M^b(E)) = \{f: M^b(E) \to \mathbb{R} \text{ continuous}, \sup_{\nu \in M^b(E)} |f(\nu)| < \infty\},$$

$$C_b^2(M^b(E)) = \{f \in C_b(M^b(E)): f \in D_{\mathscr{G}}(M^b(E);\mathbb{R}), \text{ and } \|f'\| < \infty, \|f''\| < \infty\},$$

$$C_f(M^b(E)) = \{F \in C_b(M^b(E)): F \text{ has the form } F(\mu) = f(\langle\mu, \phi_1\rangle, \dots, \langle\mu, \phi_m\rangle) \text{ for an } m \in \mathbb{N}, f \in C_b^\infty(\mathbb{R}^m), \phi_1, \dots, \phi_m \in C_c(E)\}, \tag{2}$$

$$C_t(M^b(E)) = \{F \in C_f(M^b(E)), \text{ where the function } f \text{ has the form } f(x) = \int f_0(z) \exp\left\{-\sum_{i=1}^m \frac{(x_i - z_i)^2}{2\sigma_i^2}\right\} dz \text{ for an } m \in \mathbb{N}, f_0 \in C_c(\mathbb{R}^m), (\sigma_1, \dots, \sigma_m) \in \mathbb{R}^m\} \tag{3}$$

(where $\langle\mu, \phi\rangle = \int \phi\, d\mu$). Then functions in $C_f(M^b(E))$ and $C_t(M^b(E))$ are infinitely often Gâteaux differentiable on every open subset of $M^b(E)$. As $M_1(E)$ is compact, the restrictions of functions in the above classes to $M_1(E)$ are again continuous and bounded. Furthermore, $C_t(M^b(E))$ is an algebra, and for $C_t(M_1(E)) = \{f|_{M_1(E)}, f \in C_t(M^b(E))\}$ we have the following theorem.

THEOREM 1.2. *$C_t(M_1(E))$ is dense in $C(M_1(E))$ with respect to the topology of uniform convergence.*

(The proof is essentially based on the Stone–Weierstrass Theorem.) An important special case is $E = \mathbb{R}^k$, for a $k \in \mathbb{N}$. Then we may replace $C_c(E)$ by

$C_b^\infty(\mathbb{R}^k)$ in (2) and (3). Set

$$C_f^\infty(M^b(\mathbb{R}^k)) = \{F \in C_b(M^b(\mathbb{R}^k)) \colon F \text{ has the form } F(\mu) = f(\langle\mu,\phi_1\rangle,\dots,\langle\mu,\phi_m\rangle) \text{ for an } m \in \mathbb{N}, f \in C_b^\infty(\mathbb{R}^m), \phi_1,\dots,\phi_m \in C_b^\infty(\mathbb{R}^k)\} \tag{4}$$

and define $C_t^\infty(M^b(\mathbb{R}^k))$ analogously. Let $C_f^\infty(M_1(\mathbb{R}^k))$ and $C_t^\infty(M_1(\mathbb{R}^k))$ be the restrictions to $M_1(\mathbb{R}^k)$. We have the following corollary.

COROLLARY 1.3. *$C_t^\infty(M_1(\mathbb{R}^k))$ is dense in $C(M_1(\mathbb{R}^k))$ with respect to the topology of uniform convergence.*

Thus, $C_t(M_1(E))$ and $C_t^\infty(M_1(\mathbb{R}^k))$ are convergence-determining classes for $M_1(E)$ and $M_1(\mathbb{R}^k)$, respectively, for vague convergence.

2. The weak law of large numbers for empirical measures.

2.1. *The main theorem.* Throughout this section, let $(X_i)_{i\in\mathbb{N}}$ be a family of random elements on E, $\mu_i = \mathscr{L}(X_i), i \in \mathbb{N}$, $\bar{\mu}_n = (1/n)\sum_{i=1}^n \mu_i$ and assume:

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

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

THEOREM 2.1 (Weak law of large numbers). *If the $(X_i)_{i\in\mathbb{N}}$ are independent, we have*

$$\mathscr{L}(\xi_n) \overset{w}{\Rightarrow} \delta_\mu \qquad (n \to \infty)$$

REMARK 2.2.

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

$$\int_{M_1(E)} f(\nu)\,\mathbb{P}[\xi_n \in d\nu] \to \int_{M_1(E)} f(\nu)\delta_\mu(d\nu) = f(\mu) \qquad (n \to \infty).$$

2. The name "weak law of large numbers" is based on the following fact (see [5], page 305, Proposition 11.1.3.). If $(\mathscr{S},\rho)$ is a metric space, $p \in \mathscr{S}$ and $(Y_n)_{n\in\mathbb{N}}$ is a family of E-valued random elements defined on the same probability space, with $\mathscr{L}(Y_n) \overset{w}{\Rightarrow} \delta_p(n \to \infty)$, then $Y_n \overset{\mathbb{P}}{\to} p$. Let d denote the metric of the Polish space $M_1(M_1(E))$. Then, the weak law of large numbers holds if and only if, for all $\varepsilon > 0$,

$$\mathbb{P}[d(\mathscr{L}(\xi_n),\delta_\mu) \ge \varepsilon] \to 0 \qquad (n \to \infty).$$

As already mentioned, the proof of Theorem 2.1 is based on Stein's method. In this context, the method can briefly be sketched as follows (see [2], pages 205–206). Let $\Phi_n : E^n \to E$ be a measurable mapping and let $(W_n)_{n\in\mathbb{N}}$ be a family of random elements on E^n such that $\mathscr{L}(\Phi_n(W_n)) = \mathscr{L}(\xi_n)$. Let W be a Markov process on E with generator $\mathscr{A}$ and stationary distribution δ_μ. Suppose M is a class of functions such that for all $g \in M$ there is a solution f of the "Stein equation"

$$g(x) - \langle \delta_\mu, g\rangle = (\mathscr{A} f)(x), \qquad x \in E. \tag{5}$$

Then

$$\mathbb{E}g(\xi_n) - \langle \delta_\mu, f\rangle = \mathbb{E}(\mathscr{A} f)(\Phi_n(W_n)).$$

So, if M is convergence-determining in $M_1(E)$ and if

$$\mathbb{E}(\mathscr{A} f)(\Phi_n(W_n)) \to 0 \qquad (n \to \infty)$$

for all possible solutions f of (5), then it follows that

$$\mathscr{L}(\xi_n) \overset{w}{\Rightarrow} \delta_\mu \qquad (n \to \infty)$$

and $|\mathbb{E}(\mathscr{A} f)(\Phi_n(W_n))|$ gives the rate of convergence. To construct such W_n and an approximating W, we start with a heuristic argument.

Construction of a Markov process $(W_n)_{n\in\mathbb{N}}$ *with* $\mathscr{L}(\Phi_n(W_n)) = \mathscr{L}(\xi_n)$. Let $n \in \mathbb{N}$ be fixed and set

$$Z_n^0 = (Z_{n,1}^0, \ldots, Z_{n,n}^0) = (X_1, \ldots, X_n).$$

Choose randomly, according to the uniform distribution and independently of the X_i's, an index $M \in \{1, \ldots, n\}$, and replace X_M by $\widetilde{X}_M$, where $\widetilde{X}_M$ is defined on the same probability space as X_M, independently of the X_i's and with the same distribution as X_M. Set

$$Z_n^1 = (Z_{n,1}^1, \ldots, Z_{n,n}^1) = (X_1, \ldots, X_{M-1}, \widetilde{X}_M, X_{M+1}, \ldots, X_n).$$

Again, choose an index $M_2 \in \{1, \ldots, n\}$ as before and replace Z_{n,M_2}^1 by $\widetilde{Z}_{n,M_2}^1$ as above; set

$$Z_n^2 = (Z_{n,1}^1, \ldots, Z_{n,M_2-1}^1, \widetilde{Z}_{n,M_2}^1, Z_{n,M_2+1}^1, \ldots, Z_{n,n}^1).$$

Iteration of this procedure yields a family $(Z_n^m)_{m\in\mathbb{N}}$ of random elements on E^n. For all $m \in \mathbb{N}$, let

$$U_n^m = (\delta_{Z_{n,1}^m}, \ldots, \delta_{Z_{n,n}^m});$$

then, $(U_n^m)_m$ is a Markov chain on $(M_1(E))^n$, with transition probabilities

$$K_n(\nu, \Gamma) = \frac{1}{n} \sum_{M=1}^{n} I[\delta_{x_i} \in \Gamma_i, i \neq M]\mathbb{P}[\delta_{X_M} \in \Gamma_M],$$

where $\Gamma = \Gamma_1 \times \cdots \times \Gamma_n \in \mathscr{B}(M_1(E))^n, \nu \in (M_1(E))^n$. [Because $M_1(E)$ is Polish, $\mathscr{B}((M_1(E))^n) = (\mathscr{B}(M_1(E)))^n$.]

Furthermore, let $(N(t))_{t\geq 0}$ be a Poisson process with parameter 1, independent of all other random elements of the system, and set

$$W_n(t) = U_n^{N(t)}.$$

Then $(W_n(t))_{t\geq 0}$ is a (pseudo-Poisson type) Markov process with generator

$$(\mathscr{A}_n f)(\nu) = \int_{(M_1(E))^n} f(\eta) K_n(\nu, d\eta) - f(\nu),$$

for all f for which the right side exists. Let

$$\Phi_n : (M_1(E))^n \to M_1(E); \qquad \Phi_n(\nu_1 \times \ldots \times \nu_n) = \bar{\nu}.$$

Then $\mathscr{L}(\Phi_n(W_n(t))) = \mathscr{L}(\xi_n)$ for all $t \geq 0$.

Construction of an approximating process. Consider, for $\nu = \nu_1 \times \cdots \times \nu_n \in (M_1(E))^n$,

$$\begin{aligned}
&\mathscr{A}_n(f \circ \Phi_n)(\nu)\\
&\quad = \frac{1}{n}\sum_{M=1}^{n}\int_{(M_1(E))^n} f(\bar{\eta}) I[\delta_{x_i} \in d\eta_i, i \neq M]\mathbb{P}[\delta_{X_M} \in d\eta_M] - f(\bar{\nu})\\
&\quad = \frac{1}{n}\sum_{M=1}^{n}\int_{M_1(E)} f\Big(\bar{\nu} - \frac{1}{n}(\nu_M - \eta_M)\Big)\mathbb{P}[\delta_{X_M} \in d\eta_M] - f(\bar{\nu})\\
&\quad = \frac{1}{n}\sum_{M=1}^{n}\int_{M_1(E)} \bigg[\frac{1}{n} f'(\bar{\nu})[\eta_M - \nu_M]\\
&\qquad\qquad + \frac{1}{n^2} f''(\bar{\nu} + \theta_M(\eta_M - \nu_M))[(\eta_M - \nu_M)^{(2)}]\bigg]\mathbb{P}[\delta_{X_M} \in d\eta_M]\\
&\qquad\qquad\qquad \text{(for some } 0 < \theta_M < 1, \text{ by Taylor expansion)}\\
&\quad = \frac{1}{n^2}\sum_{M=1}^{n}\int_{M_1(E)} f'(\bar{\nu})[\eta_M]\mathbb{P}[\delta_{X_M} \in d\eta_M] - \frac{1}{n^2}\sum_{M=1}^{n} f'(\bar{\nu})[\nu_M] + R(\theta),
\end{aligned}$$

where

$$\begin{aligned}
R(\theta) = \frac{1}{n^3}\sum_{M=1}^{n}\int_{M_1(E)} & f''(\bar{\nu} + \theta_M(\eta_M - \nu_M))\\
&\times [(\eta_M - \nu_M)^{(2)}]\mathbb{P}[\delta_{X_M} \in d\eta_M].
\end{aligned}$$

We now use a technical lemma which follows easily from Theorem 3.27 in Rudin [15], pages 74–75.

LEMMA 2.3. *Let $\Lambda : M_1(E) \to \mathbb{R}$ be continuous and linear, and let η be a probability measure on $M_1(E)$. Then*

$$\Lambda\Big(\int_{M_1(E)} \nu\, \eta(d\nu)\Big) = \int_{M_1(E)} \Lambda(\nu)\, \eta(d\nu).$$

Hence we have

$$\mathscr{A}_n(f\circ\Phi_n)(\nu)=R(\theta)+\frac{1}{n}f'(\bar\nu)[\bar\mu-\bar\nu].$$

Furthermore,

$$|R(\theta)|\le\frac{1}{n^2}\sup_{\nu\in M_1(E)}\|f''(\nu)\|.$$

Therefore we choose

$$(\mathscr{A}f)(\nu)=f'(\nu)[\mu-\nu],\qquad \nu\in M_1(E),$$

as approximating generator. $\mathscr{A}$ is the generator of the deterministic Markov process $(Y(t))_{t\ge0}$ that is given by

$$\mathbb{P}[Y(t)=(\nu-\mu)e^{-t}+\mu\mid Y(0)=\nu]=1,\qquad \nu\in M_1(E),$$

and has as stationary distribution δ_μ. Now we can prove Theorem 2.1.

PROOF OF THEOREM 2.1. Consider the Stein equation

$$g(\nu)-\langle\delta_\mu,g\rangle=f'(\nu)[\mu-\nu],\qquad \nu\in M_1(E). \tag{6}$$

As sketched before, the proof of Theorem 2.1 consists of two steps:

1. To solve (6) for all $g\in C_b^2(M_1(E))$ and to show that these solutions are in $C_b^2(M_1(E))$.
2. To prove that $\mathbb{E}(\mathscr{A}f)(\xi_n)\to 0(n\to\infty)$, for all $f\in C_b^2(M_1(E))$.

Step 1. *Solution of the Stein equation* (6). Let $g\in C_b^2(M_1(E))$ be fixed. We may assume that $g(\mu)=0$, as the equation (6) remains the same, if g would be replaced by $g-g(\mu)$. As in [1], we use the fact that if $(T_t)_t$ is a semigroup for the generator $\mathscr{A}$ and if $\int_0^t T_s f ds\in\mathscr{D}(\mathscr{A})$, then

$$T_tf-f=\mathscr{A}\int_0^t T_sfds$$

(see [6], page 9). Therefore,

$$\psi(g)(\nu)\doteq-\int_0^\infty T_sg(\nu)\,ds,\qquad \nu\in M_1(E),$$

is a solution of the Stein equation if it is an element of $\mathscr{D}(\mathscr{A})$. To ensure the existence of $\psi(g)$, observe that for all $\nu\in M_1(E)$,

$$\begin{aligned}\left|\int_0^\infty T_sg(\nu)\,ds\right|&=\left|\int_0^\infty g((\nu-\mu)e^{-s}+\mu)\,ds\right|\\&\le\int_0^\infty|g'(\theta_u(\nu-\mu)e^{-s}+\mu)[\nu-\mu]|\,ds\\&\le\sup_{\nu\in M_1(E)}\|g'(\nu)\|\int_0^\infty e^{-s}\,ds\\&<\infty\quad\text{as } g\in C_b^2(M_1(E)).\end{aligned}$$

Thus $\psi(g)$ exists. The fact that $\psi(g)$ is twice continuously Gâteaux differentiable and bounded follows by straightforward calculation of the derivatives. Thus $\psi(g) \in C_b^2(M_1(E))$.

Step 2. Asymptotics $(n \to \infty)$ *of* $\mathbb{E}(\mathscr{A}f)(\xi_n)$. Let $f \in C_b^2(M_1(E))$ be fixed. Then

$$\begin{aligned}\mathbb{E}(\mathscr{A}f)(\xi_n) &= \mathbb{E}\left[f'\left(\frac{1}{n}\sum_{i=1}^n \delta_{X_i}\right)\left[\mu - \frac{1}{n}\sum_{i=1}^n \delta_{X_i}\right]\right]\\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[f'\left(\frac{1}{n}\sum_{j\neq i}\delta_{X_j}\right)[\mu - \delta_{X_i}]\right] + R_1,\end{aligned}$$

where $R_1 = (1/n)\sum_{i=1}^n \mathbb{E}[(f'((1/n)\sum_{j=1}^n \delta_{X_j}) - f'((1/n)\sum_{j\neq i}\delta_{X_j}))[\mu - \delta_{X_i}]]$. Thus, by independence and Lemma 2.3,

$$\begin{aligned}\mathbb{E}(\mathscr{A}f)(\xi_n) &= R_1 + \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[f'\left(\frac{1}{n}\sum_{j\neq i}\delta_{X_j}\right)[\mu - \mu_i]\right]\\ &= R_1 + \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[f'\left(\frac{1}{n}\sum_{j=1}^n \delta_{X_j}\right)[\mu - \mu_i]\right] + R_2,\end{aligned}$$

where $R_2 = (1/n)\sum_{i=1}^n \mathbb{E}[(f'((1/n)\sum_{j\neq i}\delta_{X_j}) - f'((1/n)\sum_{j=1}^n \delta_{X_j}))[\mu - \mu_i]]$. From $\bar{\mu}_n \overset{v}{\Rightarrow} \mu (n \to \infty)$ and the boundedness of f' follows (by dominated convergence)

$$\left|\mathbb{E}\left[f'\left(\frac{1}{n}\sum_{j=1}^n \delta_{X_j}\right)[\mu - \bar{\mu}_n]\right]\right| \to 0 \qquad (n \to \infty).$$

Furthermore, $|R_1|$ and $|R_2|$ can be dominated simultaneously by

$$\begin{aligned}|R_1|, |R_2| &\le \frac{1}{n}\sup_{\nu\in M_1(E)} \|f''(\nu)\|\\ &\to 0 \qquad (n \to \infty).\end{aligned}$$

This completes the proof. □

Observe that in the above considerations we could have replaced δ_{X_i} by $g(X_i)$, with any $g: E \to M_1(E)$ being a bounded measurable function.

2.2. *The w.l.l.n. for dependent random elements.* In the proof of Theorem 2.1, the assumption of independence was only used to obtain $\mathbb{E}[f'(\xi_n)[\mu - \xi_n]] \to 0 (n \to \infty)$. This leads to the obvious next corollary.

COROLLARY 2.4. *If, for all* $f \in C_b^2(M_1(E))$, *we have*

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

then

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

We can weaken the assumption of this corollary by observing that to prove Theorem 2.1 it would have been sufficient to consider only f's that are solutions of the Stein equation for $g \in C_f(M_1(E))$ [or $g \in C_f^\infty(M_1(\mathbb{R}^k))$, if $E = \mathbb{R}^k$]. These solutions have a fairly simple form.

LEMMA 2.5. *For every $g \in C_f(M_1(E))$ [or $g \in C_f^\infty(M_1(\mathbb{R}^k))$, if $E = \mathbb{R}^k$], the solution $\psi(g)$ in the proof of Theorem 2.1 is of the form (2) [or (4), if $E = \mathbb{R}^k$]. Furthermore, $\|\psi'(g)\| \le \|g'\|$ and $\|\psi''(g)\| \le \|g''\|$.*

This yields a much more easily applicable version of the w.l.l.n.

COROLLARY 2.6. *Assume that for all functions F of the form (2) [or (4), if $E = \mathbb{R}^k$] and for all $\phi \in C_c(E)$ [or $\phi \in C_b^\infty(\mathbb{R}^k)$, if $E = \mathbb{R}^k$] we have*

$$\mathbb{E}[F(\xi_n)\langle \mu - \xi_n, \phi\rangle] \to 0 \qquad (n \to \infty).$$

Then

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

This corollary follows directly from calculating the derivative of functions F of the form (2) [or (4), respectively]. To be more concrete, we can use this corollary to prove the next corollary.

COROLLARY 2.7. *Assume that for all $i, n \in \mathbb{N}$ there is a $\Gamma_s^n(i) \subset \{1, \dots, n\} \setminus \{i\}$ such that, with $\Gamma_w^n(i) = \{1, \dots, n\} \setminus [\{i\} \cup \Gamma_s^n(i)]$, we have:*

(i) $(1/n^2)\sum_{i=1}^n |\Gamma_s^n(i)| \to 0 (n \to \infty)$.

(ii) *For all $m \in \mathbb{N}$, $f \in C_b^\infty(\mathbb{R}^m)$ and $\phi_1, \dots, \phi_m, \phi \in C_c(E)$ [or $C_b^\infty(\mathbb{R}^k)$, respectively] we have*

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\Bigg[f\Bigg(\Bigg\langle \frac{1}{n}\sum_{j\in\Gamma_w^n(i)} \delta_{X_j}, \phi_1\Bigg\rangle, \dots, \Bigg\langle \frac{1}{n}\sum_{j\in\Gamma_w^n(i)} \delta_{X_j}, \phi_m\Bigg\rangle\Bigg) \times (\phi(X_i) - \langle \mu, \phi\rangle)\Bigg] \to 0 \qquad (n \to \infty).$$

Then

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

Of course, the above results can be generalized to arrays of random elements. Let, from now on, $(X_{i,n})_{i=1,\dots,r(n);n\in\mathbb{N}}$ be an array of random elements on E, where $r(n) \in \mathbb{N}$ and $r(n) \to \infty (n \to \infty)$. Put $\mu_{in} = \mathscr{L}(X_{i,n})$, $i, n \in \mathbb{N}$, $\bar{\mu}_n = (1/r(n))\sum_{i=1}^{r(n)} \mu_{in}$ and suppose

$$\bar{\mu}_n \overset{v}{\Rightarrow} \mu \qquad (n \to \infty) \text{ for a } \mu \in M_1(E).$$

The corresponding generalizations then are, with $\xi_n = (1/r(n))\sum_{i=1}^{r(n)} \delta_{X_{i,n}}$, given in the next theorem.

THEOREM 2.8. *Suppose the* $(X_{i,n})_{i=1,\dots,r(n);n\in\mathbb{N}}$ *are independent. Then,*

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

COROLLARY 2.9. *Suppose that for every function* F *of the form* (2) [*or* (4), *if* $E = \mathbb{R}^k$],

$$\mathbb{E}[F'(\xi_n)[\mu - \xi_n]] \to 0 \qquad (n \to \infty).$$

Then

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

COROLLARY 2.10. *Assume that for all functions* F *of the form* (2) [*or* (4), *if* $E = \mathbb{R}^k$] *and for all* $\phi \in C_c(E)$ [*or* $C_b^\infty(\mathbb{R}^k)$, *if* $E = \mathbb{R}^k$], *we have*

$$\mathbb{E}[F(\xi_n)\langle \mu - \xi_n, \phi\rangle] \to 0 \qquad (n \to \infty).$$

Then

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

COROLLARY 2.11. *Assume that, for all* $i, n \in \mathbb{N}$, *there is a subset* $\Gamma_s^n(i)$ *of* $\{1,\dots,r(n)\}\setminus\{i\}$ *such that, with* $\Gamma_w^n(i) = \{1,\dots,r(n)\}\setminus[\{i\}\cup\Gamma_s^n(i)]$, *we have:*

(i) $1/(r(n)^2)\sum_{i=1}^{r(n)} |\Gamma_s^n(i)| \to 0 (n \to \infty)$.

(ii) *For all* $m \in \mathbb{N}, f \in C_b^\infty(\mathbb{R}^m), \phi_1,\dots,\phi_m, \phi \in C_c(E)$ [*or* $C_b^\infty(\mathbb{R}^k)$, *if* $E = \mathbb{R}^k$] *we have*

$$\frac{1}{r(n)}\sum_{i=1}^{r(n)} \mathbb{E}\Bigg[f\Bigg(\Bigg\langle\frac{1}{r(n)}\sum_{j\in\Gamma_w^n(i)} \delta_{X_{j,n}}, \phi_1\Bigg\rangle,\dots,\Bigg\langle\frac{1}{r(n)}\sum_{j\in\Gamma_w^n(i)} \delta_{X_{j,n}}, \phi_m\Bigg\rangle\Bigg) \times (\phi(X_{i,n}) - \langle\mu,\phi\rangle)\Bigg] \to 0 \qquad (n \to \infty).$$

Then

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

The assumptions of Corollary 2.8 are obviously satisfied for strong ϕ-mixing sequences of real random variables. For this case, under some additional assumptions, Rama Krishnaiah [11] proved even a Glivenko–Cantelli theorem. However, the usefulness of the w.l.l.n. for other kinds of dependent random elements will be illustrated in the next section.

2.3. *The rate of convergence.* With help of Corollary 2.7, we can give an estimate on the rate of convergence. Define, for $\mathscr{C} = C_c(E)$ or $\mathscr{C} = C_b^\infty(\mathbb{R}^k)$, if $E = \mathbb{R}^k$, the set of functions

$$\mathscr{F}_{\mathscr{C}} = \{F \in C_{\mathscr{C}}(M_1(E)) \colon F \text{ has the form } F(\mu) = f(\langle \mu, \phi_1\rangle, \dots, \langle \mu, \phi_m\rangle) \text{ for an } m \in \mathbb{N}, f \in C_b^\infty(\mathbb{R}^m) \text{ with } \|f'\| \le 1, \|f''\| \le 1; \text{ and for } \phi_1, \dots, \phi_m \in \mathscr{C} \text{ with } \|\phi_i\| \le 1, i = 1, \dots, m\}.$$

Then we can define the following Zolotarev semimetric (see [10]) on $M_1(M_1(E))$:

$$\zeta_{\mathscr{F}_{\mathscr{C}}}(\nu, \eta) = \sup_{f \in \mathscr{F}_{\mathscr{C}}} \left| \int_{M_1(E)} f\, d\nu - \int_{M_1(E)} f\, d\eta \right|.$$

Because $\mathscr{F}_{\mathscr{C}}$ is convergence-determining for vague convergence in $M_1(M_1(E))$, we have that $\zeta_{\mathscr{F}_{\mathscr{C}}}$ is a metric on $M_1(M_1(E))$ and for all $(\nu_n)_{n \in \mathbb{N}}$, $\nu \in M_1(M_1(E))$,

$$\zeta_{\mathscr{F}_{\mathscr{C}}}(\nu_n, \nu) \to 0 \qquad (n \to \infty) \quad \Rightarrow \quad \nu_n \overset{v}{\Rightarrow} \nu \qquad (n \to \infty).$$

(The converse need not be true, though.) We get, without any independence assumptions, the following proposition.

PROPOSITION 2.12. (i) *For all $m \in \mathbb{N}, f \in C_b^\infty(\mathbb{R}^m)$ and $\phi_1, \dots, \phi_m \in \mathscr{C}$, where $\mathscr{C} = C_c(E)$ or $\mathscr{C} = C_b^\infty(\mathbb{R}^k)$, if $E = \mathbb{R}^k$, we have*

$$\begin{aligned} &\left| \mathbb{E}\left[\sum_{j=1}^m f_{(j)}(\langle \xi_n, \phi_1\rangle, \dots \langle \xi_n, \phi_m\rangle) \langle \mu - \xi_n, \phi_j\rangle \right] \right| \\ &\quad \le \sum_{j=1}^m \|f_{(j)}\| |\langle \mu - \bar{\mu}_n, \phi_j\rangle| \\ &\qquad + \sum_{j,k=1}^m \|f_{(j,k)}\| \left\{ \max_{1 \le j \le m} \langle \mu - \bar{\mu}_n, \phi_j\rangle^2 + \operatorname{Var}\left(\frac{1}{n} \sum_{i=1}^n \phi_j(X_i) \right) \right\}. \end{aligned}$$

(ii) *We have*

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

Note that, in the i.i.d. case, we get that $\zeta_{\mathscr{F}_{\mathscr{C}}}(\mathscr{L}(\xi_n), \delta_\mu) \le 2/n$. Thus, we get the expected order of magnitude.

REMARK 2.13. The method we used can of course also be applied to recover "classical" weak law of large numbers for random variables. Moreover, the approach also provides a general result for convergence toward the Dirac measure; the special structure of the empirical measures is not required. This (and a more detailed treatment of the above) can be found in Reinert [13].

3. Some examples.

3.1. *A dissociated family.* Let $(Y_i)_{i\in\mathbb{N}}$ be a family of independent random elements on a space $\mathscr{X}$, let $k \in \mathbb{N}$ be fixed and set

$$\Gamma = \{(j_1,\dots,j_k) \in \mathbb{N}^k\colon j_r \neq j_s \text{ for } r \neq s\},$$
$$\Gamma^{(n)} = \{(j_1,\dots,j_k) \in \Gamma\colon j_1,\dots,j_k \in \{1,\dots,n\}\}.$$

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

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

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

$$\frac{1}{n(n-1)\cdots(n-k+1)} \sum_{(j_1,\dots,j_k)\in\Gamma^{(n)}} \mathscr{L}(\psi_{j_1,\dots,j_k}(Y_{j_1},\dots,Y_{j_k})) \overset{v}{\Rightarrow} \mu$$

for a $\mu \in M_1(E)$. Let

$$\xi_n = \frac{1}{n(n-1)\cdots(n-k+1)} \sum_{(j_1,\dots,j_k)\in\Gamma^{(n)}} \delta_{X_{j_1,\dots,j_k}}.$$

THEOREM 3.1. *For the above dissociated family, we have*

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

PROOF. The proof is based on Corollary 2.11. For $n \in \mathbb{N}$ fixed, the set $\Gamma^{(n)}$ has $n(n-1)\cdots(n-k+1)$ elements. Fix a counting for $\Gamma^{(n)}$. If $(j_1,\dots,j_k)$ is the ith element, identify $(j_1,\dots,j_k)$ with i and set $X_{j_1,\dots,j_k} = X_{i,n}$. Let $r(n) = n(n-1)\cdots(n-k+1)$. Then

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

that is, ξ_n has the required form for Corollary 2.11. For $i = (j_1,\dots,j_k)$, define

$$\Gamma_s^n(i) = \{(l_1,\dots,l_k) \in \Gamma^{(n)}\colon (l_1,\dots,l_k) \neq i;$$
$$\{l_1,\dots,l_k\} \cap \{j_1,\dots,j_k\} \neq \varnothing\}.$$

Then, for all $i \leq r(n)$,

$$|\Gamma_s^n(i)| = k[k(n-1)(n-2)\cdots(n-k+1) - 1]$$
$$\leq \frac{r(n)}{n}k^2$$

and thus the first condition of Corollary 2.11 is satisfied. For the second condition, let $d \in \mathbb{N}$ be arbitrary, $f \in C_b^\infty(\mathbb{R}^d)$, $\phi_1, \dots, \phi_d, \phi \in C_c(E)$. 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

$$\frac{1}{r(n)} \sum_{i=1}^{r(n)} \mathbb{E}\Bigg[f\Bigg(\Bigg\langle \frac{1}{r(n)} \sum_{m \in \Gamma_w^n(i)} \delta_{X_{m,n}}, \phi_1 \Bigg\rangle, \dots, \Bigg\langle \frac{1}{r(n)} \sum_{m \in \Gamma_w^n(i)} \delta_{X_{m,n}}, \phi_d \Bigg\rangle \Bigg) \times (\phi(X_{i,n}) - \langle \mu, \phi \rangle) \Bigg]$$

$$= \frac{1}{r(n)} \sum_{i=1}^{r(n)} \mathbb{E}\Bigg[f\Bigg(\Bigg\langle \frac{1}{r(n)} \sum_{m \in \Gamma_w^n(i)} \delta_{X_{m,n}}, \phi_1 \Bigg\rangle, \dots, \Bigg\langle \frac{1}{r(n)} \sum_{m \in \Gamma_w^n(i)} \delta_{X_{m,n}}, \phi_d \Bigg\rangle \Bigg) \Bigg] \times \mathbb{E}[\phi(X_{i,n}) - \langle \mu, \phi \rangle]$$

$$= \mathbb{E}\Bigg[f(\langle \xi_n, \phi_1 \rangle, \dots, \langle \xi_n, \phi_d \rangle) \frac{1}{r(n)} \sum_{i=1}^{r(n)} [\langle \mathscr{L}(X_{i,n}), \phi \rangle - \langle \mu, \phi \rangle] \Bigg] + R_1,$$

where

$$R_1 = \frac{1}{r(n)} \sum_{i=1}^{r(n)} \mathbb{E}\Bigg[f\Bigg(\Bigg\langle \frac{1}{r(n)} \sum_{m \in \Gamma_w^n(i)} \delta_{X_{m,n}}, \phi_1 \Bigg\rangle, \dots, \Bigg\langle \frac{1}{r(n)} \sum_{m \in \Gamma_w^n(i)} \delta_{X_{m,n}}, \phi_d \Bigg\rangle \Bigg) - f(\langle \xi_n, \phi_1 \rangle, \dots, \langle \xi_n, \phi_d \rangle) \Bigg] \mathbb{E}[\phi(X_{i,n}) - \langle \mu, \phi \rangle].$$

Due to the assumption, we only need to show that $|R_1| \to 0 (n \to \infty)$. This follows easily by Taylor's expansion:

$$|R_1| \le 2\|\phi\| \|f'\| \sup_{1 \le l \le d} \|\phi_l\| \frac{1}{(r(n))^2} \sum_{i=1}^{r(n)} |\Gamma_s^n(i)| \to 0 \qquad (n \to \infty),$$

and by Corollary 2.11 the assertion follows. □

Note that in the exchangeably dissociated case, that is, $(Y_i)_i$ 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^n be the (positive) arrival time of the ith individual and Z_i its life span, and assume the $(Z_i)_{i \in \mathbb{N}}$ are positive, i.i.d. and independent of the $(A_i^n)_{i,n \in \mathbb{N}}$ [but allow for dependence between the $(A_i^n)_{i,n \in \mathbb{N}}$]. Suppose:

1. $(1/n^2) \sum_{i,j=1}^n \mathbb{E}[|(A_i^n - \mathbb{E}A_i^n)(A_j^n - \mathbb{E}A_j^n)|] \to 0 \qquad (n \to \infty).$

2. There is a measure $\mu \in M_1(\mathbb{R}^2)$ with

$$\frac{1}{n}\sum_{i=1}^{n}\mathscr{L}((A_i^n, A_i^n + Z_i)) \overset{v}{\Rightarrow} \mu \qquad (n \to \infty).$$

Start at time $t = 0$ and let

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

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

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

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 3.2. *In the above setting,*

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

PROOF. We employ Proposition 2.12. Due to the second assumption, we only have to show that the variance term tends to 0 as $n \to \infty$. We have, for all $\phi \in C_b^\infty(\mathbb{R}^2)$,

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

Hence, using Taylor's expansion, we get

$$\begin{aligned}
\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\phi(A_i^n, A_i^n + Z_i)\right) &\le \|\phi'\|^2\frac{1}{n^2}\sum_{i,j=1}^{n}\mathbb{E}[|(A_i^n - \mathbb{E}A_i^n)(A_j^n - \mathbb{E}A_j^n)|]\\
&\to 0 \qquad (n \to \infty). \quad \square
\end{aligned}$$

As an example, in the following case it is easily checked that the conditions are satisfied. Let $(E_i^n)_{i=1,\ldots,n}$ be i.i.d. $\exp(n)$ random variables and put $A_i^n =$

$\sum_{j=1}^{n} E_i^n$. If G is the distribution function of Z_1, then the limiting measure μ is such that, for real rectangles,

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

This covers the well-known results for this case; see, for example, Ross [14], page 214.

4. Proofs.

PROOF OF THEOREM 1.2. Because $M_1(E)$ is compact, by the Stone–Weierstrass Theorem it is sufficient to show that:

(i) For all $\mu \in M_1(E)$, there is a $F \in C_t(M_1(E))$ with $F(\mu) \neq 0$.
(ii) For all $\mu, \nu \in M_1(E), \mu \neq \nu$, there is a $F \in C_t(M_1(E))$ with $F(\mu) \neq F(\nu)$.

For item (i), let $\mu \in M_1(E)$ be arbitrary, fixed. Let $\phi \in C_c(E)$ be arbitrary, fixed and set $c = \langle \mu, \phi \rangle$,

$$f_0(z) = \begin{cases} 0, & z \notin [c-2, c+2], \\ z - c + 2, & z \in [c-2, c-1], \\ 1, & z \in [c-1, c+1], \\ c + 2 - z, & z \in [c+1, c+2], \end{cases}$$

$$f(x) = \int f_0(z) \exp\left\{ -\frac{(z-x)^2}{2} \right\} dz.$$

Then $f_0 \in C_c(\mathbb{R}), f \in C_b^\infty(\mathbb{R})$ and a simple calculation gives $f(c) > 0$. Set

$$F(\nu) = f(\langle \nu, \phi \rangle), \qquad \nu \in M_1(E).$$

Then $F \in C_t(M_1(E))$ and item (i) is satisfied.

For item (ii), let $\mu \neq \nu \in M_1(E)$ be arbitrary, fixed. Because $C_c(E)$ is convergence-determining in $M_1(E)$, there is a $\phi \in C_c(E)$ with $\langle \mu, \phi \rangle \neq \langle \nu, \phi \rangle$. Keep this Φ and let $b = \langle \nu, \phi \rangle, c = \langle \mu, \phi \rangle, d = c - b$ and without loss of generality assume $d > 0$. Define

$$f_0(z) = \begin{cases} 0, & z \notin [c-d/4, c+d/4], \\ z - c + d/4, & z \in [c-d/4, c-d/8], \\ 1, & z \in [c-d/8, c+d/8], \\ c + d/4 - z, & z \in [c+d/8, c+d/4], \end{cases}$$

$$f(x) = \int f_0(z) \exp\left\{ -\frac{(z-x)^2}{2} \right\} dz.$$

Then $f_0 \in C_c(\mathbb{R}), f \in C_b^\infty(\mathbb{R})$ and $f(c) - f(b) > 0$. Set

$$F(\nu) = f(\langle \nu, \phi \rangle), \qquad \nu \in M_1(E).$$

Then $F \in C_t(M_1(E))$ and item (ii) is satisfied. □

PROOF OF COROLLARY 1.3. From the proof of Theorem 1.2, Corollary 1.3 is immediate, if it is shown that Lemma 4.1 holds.

LEMMA 4.1. *For each* $k \in \mathbb{N}$, $C_t^\infty(M_1(\mathbb{R}^k))$ *is convergence-determining for the vague convergence in* $M_1(\mathbb{R}^k)$.

PROOF. The proof follows the lines of Pollard [9], pages 48–49, where it is shown for probability measures. As [9], the following is easily seen to be true:

Let $(\nu_m)_{n\in\mathbb{N}}$, $\nu, \mu \in M_1(\mathbb{R}^k)$ with $\mu(\mathbb{R}^k) = 1$, define for all $\sigma > 0, A \in \mathscr{B}(\mathbb{R}^k)$, the measure $\mu_\sigma(A) = \mu(\sigma A)$, and assume

$$\nu_n * \mu_\sigma \overset{v}{\Rightarrow} \nu * \mu_\sigma \qquad (n \to \infty)$$

for all $\sigma > 0$. Then

$$\nu_n \overset{v}{\Rightarrow} \nu \qquad (n \to \infty).$$

Take $\mu = \mathscr{N}(0, I_k)$. Then for all $f_0 \in C_c(\mathbb{R})$ and $\sigma > 0$, we have for all $\eta \in M_1(\mathbb{R}^k)$,

$$(\eta * \mu_\sigma)(f_0) = \int f_\sigma(x)\eta(dx),$$

where

$$f_\sigma(x) = (2\pi\sigma^2)^{-k/2} \int f_0(z) \exp\left\{-\frac{(z-x)^2}{2}\right\} dz.$$

Thus, if we have $\nu_n(f_\sigma) \to \nu(f_\sigma)(n \to \infty)$ for all f_σ as above, then, for all $f_0 \in C_c(\mathbb{R})$,

$$\nu_n * \mu_\sigma(f_0) \to \nu * \mu_\sigma(f_0) \qquad (n \to \infty)$$

and hence

$$\nu_n \overset{v}{\Rightarrow} \nu \qquad (n \to \infty). \qquad \square$$

PROOF OF LEMMA 2.6. Recall from the proof of Theorem 2.1 that

$$\psi(g)(\nu) = -\int_0^\infty T_u(g(\nu) - g(\mu))\, du$$

exists. If $G \in C_f(M_1(E))$ is of the form $G(\nu) = f(\langle \nu, \phi_1\rangle, \dots, \langle \nu, \phi_m\rangle)$ for an $m \in \mathbb{N}$, $f \in C_b^\infty(\mathbb{R}^m)$, $\phi_1, \dots, \phi_m \in \mathscr{C}$, we furthermore have

$$\psi(G)(\nu) = \int_0^\infty [f(\langle(\nu-\mu)e^{-u} + \mu, \phi_1\rangle, \dots, \langle(\nu-\mu)e^{-u} + \mu, \phi_m\rangle) \\ - f(\langle \mu, \phi_1\rangle, \dots, \langle \mu, \phi_m\rangle)]\, du.$$

Put

$$h(x_1,\dots,x_m) = \int_0^\infty [f(e^{-u}x_1 + (1-e^{-u})\langle\mu,\phi_1\rangle,\dots,e^{-u}x_m + (1-e^{-u})\langle\mu,\phi_m\rangle) - f(\langle\mu,\phi_1\rangle,\dots,\langle\mu,\phi_m\rangle)]\,du.$$

Then

$$\psi(G)(\nu) = h(\langle\nu,\phi_1\rangle,\dots,\langle\nu,\phi_m\rangle).$$

It remains to show that $h \in C_b^\infty(\mathbb{R}^m)$, but this is easily done via Taylor's expansion and the theorem of dominated convergence. □

PROOF OF COROLLARY 2.7. In view of Corollary 2.6, it suffices to prove

$$\mathbb{E}(\mathscr{A}f)(\xi_n) \to 0 \qquad (n\to\infty)$$

for all f of the form (4), $\phi \in C_c(E)$. For such f and ϕ, we have

$$\mathbb{E}[f(\xi_n)\langle\mu-\xi_n,\phi\rangle] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[f\left(\frac{1}{n}\sum_{j\in\Gamma_w^n(i)}\delta_{X_j}\right)\langle\mu-\delta_{X_i},\phi\rangle\right] + R_1,$$

where $R_1 = \mathbb{E}[(f(\xi_n) - f((1/n)\sum_{j\in\Gamma_w^n(i)}\delta_{X_j}))\langle\mu-\xi_n,\phi\rangle]$, and from the assumption it follows that

$$\left|\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[f\left(\frac{1}{n}\sum_{j\in\Gamma_w^n(i)}\delta_{X_j}\right)\langle\mu-\delta_{X_i},\phi\rangle\right]\right| \to 0 \qquad (n\to\infty).$$

Finally, by Taylor's expansion and the form of f,

$$\begin{aligned}|R_1| &\le \frac{1}{n}\sum_{i=1}^n \|Df\| \sum_{j\in\Gamma_s^n(i)}\sum_{l=1}^m \mathbb{E}\left|\left\langle\frac{1}{n}\delta_{X_j},\phi_l\right\rangle\langle\mu-\delta_{X_i},\phi\rangle\right| \\ &\le 2m\|Df\| \sup_{1\le l\le m}\|\phi_l\|\|\phi\|\frac{1}{n^2}\sum_{i=1}^n |\Gamma_s^n(i)| \\ &\to 0 \qquad (n\to\infty)\end{aligned}$$

due to the assumption. □

PROOF OF THEOREM 2.8. We can do exactly the same heuristic considerations as for Theorem 2.1; so, we obtain the same generator $\mathscr{A}$. Because the Stein equation has already been solved in the proof of Theorem 2.1, it is sufficient to show that for all $f \in C_b^2(M_1(E))$, we have $\mathbb{E}(\mathscr{A}f)(\xi_n) \to 0 (n\to\infty)$. Similarly to the proof of Theorem 2.1, we have

$$\mathbb{E}(\mathscr{A}f)(\xi_n) = \frac{1}{r(n)}\sum_{i=1}^{r(n)} \mathbb{E}\left[f'\left(\frac{1}{r(n)}\sum_{j\ne i}\delta_{X_{j,n}}\right)\langle\mu-\delta_{X_{i,n}},\phi\rangle\right] + R_1,$$

where

$$R_1 = \frac{1}{r(n)} \sum_{i=1}^{r(n)} \mathbb{E}\left[\left(f'(\xi_n) - f'\left(\frac{1}{r(n)} \sum_{j \neq i} \delta_{X_{j,n}}\right)\right) \langle \mu - \delta_{X_{i,n}}, \phi \rangle\right];$$

thus, due to the independence and Lemma 2.3,

$$\begin{aligned}\mathbb{E}(\mathscr{A}f)(\xi_n) &= R_1 + \frac{1}{r(n)} \sum_{i=1}^{r(n)} \mathbb{E}\left[f'\left(\frac{1}{r(n)} \sum_{j \neq i} \delta_{X_{j,n}}\right) \langle \mu - \mu_{i,n}, \phi \rangle\right] \\ &= R_1 + \mathbb{E}[f'(\xi_n)\langle \mu - \bar{\mu}_n, \phi \rangle] + R_2,\end{aligned}$$

where

$$R_2 = \frac{1}{r(n)} \sum_{i=1}^{r(n)} \mathbb{E}\left[\left(f'\left(\frac{1}{r(n)} \sum_{j \neq i} \delta_{X_{j,n}}\right) - f(\xi_n)\right) \langle \mu - \mu_{i,n}, \phi \rangle\right].$$

As $\bar{\mu}_n \overset{v}{\Rightarrow} \mu (n \to \infty)$, we have $\mathbb{E}[f'(\xi_n)\langle \mu - \bar{\mu}_n, \phi \rangle] \to 0 (n \to \infty)$ by the same argument as before, and, again with Taylor's expansion,

$$\begin{aligned}|R_1|, |R_2| &\leq \frac{K}{r(n)} \quad \text{for a constant } K > 0 \\ &\to 0 \qquad (n \to \infty)\end{aligned}$$

by assumption. □

Corollaries 2.9 and 2.11 can be proven in exactly the same way as Corollaries 2.6 and 2.7, respectively, so we skip the proofs.

PROOF OF PROPOSITION 2.12. For the first part, observe that

$$\begin{aligned}&\mathbb{E}\left[\sum_{j=1}^{m} f_{(j)}(\langle \xi_n, \phi_1 \rangle, \dots \langle \xi_n, \phi_m \rangle) \langle \mu - \xi_n, \phi_j \rangle\right] \\ &\quad = \mathbb{E}\left[\sum_{j=1}^{m} f_{(j)}(\langle \mu, \phi_1 \rangle, \dots \langle \mu, \phi_m \rangle) \langle \mu - \xi_n, \phi_j \rangle\right] + R_1,\end{aligned}$$

where, by Taylor's expansion,

$$\begin{aligned}|R_1| &\leq \sum_{j,k=1}^{m} \mathbb{E}[|f_{(j,k)}(\langle \xi_n + \theta(\mu - \xi_n), \phi_1 \rangle, \dots, \langle \xi_n + \theta(\mu - \xi_n), \phi_m \rangle) \\ &\qquad \times \langle \mu - \xi_n, \phi_j \rangle \langle \mu - \xi_n, \phi_k \rangle|] \\ &\leq \sum_{j,k=1}^{m} \|f_{(j,k)}\| \max_{1 \leq j \leq m} \mathbb{E}[\langle \mu - \xi_n, \phi_j \rangle^2].\end{aligned}$$

Furthermore, for all $1 \leq j \leq m$,

$$\mathbb{E}[\langle \mu - \xi_n, \phi_j \rangle^2] = \langle \mu - \bar{\mu}_n, \phi_j \rangle^2 + \operatorname{Var}\biggl(\frac{1}{n}\sum_{i=1}^n \phi_j(X_i)\biggr).$$

With Lemma 2.3, we hence get

$$\begin{aligned}
&\left|\mathbb{E}\left[\sum_{j=1}^m f_{(j)}(\langle \xi_n, \phi_1 \rangle, \dots \langle \xi_n, \phi_m \rangle)\langle \mu - \xi_n, \phi_j \rangle\right]\right| \\
&\quad \leq \sum_{j=1}^m \|f_{(j)}\| |\langle \mu - \bar{\mu}_n, \phi_j \rangle| \\
&\qquad + \sum_{j,k=1}^m \|f_{(j,k)}\| \left\{\max_{1 \leq j \leq m} \langle \mu - \bar{\mu}_n, \phi_j \rangle^2 + \operatorname{Var}\biggl(\frac{1}{n}\sum_{i=1}^n \phi_j(X_i)\biggr)\right\}.
\end{aligned}$$

This proves item (i). Item (ii) follows immediately, employing the Stein equation and Lemma 2.6. □

Acknowledgments. I am very grateful to A. D. Barbour, my advisor, for very many fruitful discussions. I also want to thank him for his permanent encouragement and support. Furthermore, I would like to thank S. Utev for discussions that lead to a considerable improvement in the treatment of metrics.

REFERENCES

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

[2] BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation.* Clarendon, Oxford.

[3] 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.

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

[5] DUDLEY, R. M. (1989). *Real Analysis and Probability.* Wadsworth & Brooks/Cole, Belmont, CA.

[6] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York.

[7] HOROWITZ, J. (1990). A uniform law of large numbers and empirical central limit theorem for limits of finite populations. *Statist. Probab. Lett.* **10** 159–166.

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

[9] POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

[10] RACHEV, S. T. (1991). *Probability Metrics and the Stability of Stochastic models.* Wiley, New York.

[11] RAMA KRISHNAIAH, Y. S. (1990). On the Glivenko–Cantelli theorem for generalized empirical processes based on strong mixing sequences. *Statist. Probab. Lett.* **10** 439–447.

[12] REINERT, G. (1994). The asymptotic evolution of the general stochastic epidemic. Submitted.

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

[14] ROSS, S. R. (1985). *Introduction to Probability Models*, 3rd ed. Academic Press, San Diego.

[15] RUDIN, W. (1973). *Functional Analysis.* McGraw-Hill, New York.

[16] STEIN, C. (1986). *Approximate Computation of Expectations*. IMS, Hayward, CA.
[17] 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.
[18] YAMAMURO, S. (1974). *Differential Calculus in Topological Linear Spaces. Lecture Notes in Math.* **374**. Springer, Berlin.

DEPARTMENT OF MATHEMATICS
DRB-155
UNIVERSITY OF SOUTHERN CALIFORNIA
LOS ANGELES, CALIFORNIA 90089-1113