Stein's Method: Expository Lectures and Applications
Institute of Mathematical Statistics
Lecture Notes – Monograph Series
Vol. 46 (2004) 93–132

6. Stein's method for the bootstrap

Susan Holmes[1] **and Gesine Reinert**[2]

Stanford University, INRA-Montpellier, and University of Oxford

Abstract: This paper gives new proofs for many known results about the convergence in law of the bootstrap distribution to the true distribution of smooth statistics, whether the samples studied come from independent realizations of a random variable or dependent realizations with weak dependence. Moreover it suggests a novel bootstrap procedure and provides a proof that this new bootstrap works under uniform local dependence. The techniques employed are based on Stein's method for empirical processes as developed by Reinert [19]. The last section provides some simulations and applications for which the relevant `matlab` functions are available from the first author.

6.1. Overview

Stein's method proves weak-convergence results through approximations to expectations without direct use of characteristic functions, allowing it to be used in complex problems with dependence. In this work we show how consistency can be proved and even some error terms provided for any bootstrap with exchangeable weights using Stein's method. We say that the bootstrap works if the distance between the bootstrap empirical measure and a Gaussian measure centred around the true empirical measure, or the true mean measure, tends to zero as sample size tends to infinity. Our results also provide an explicit error bound for the difference to Gaussianity of the bootstrap for any finite sample size.

Many of the results themselves are known, see for instance Bickel and Freedman (1981), Singh (1981) for the consistency results in the multinomial case or Praestgaard and Wellner (1993) for the case of exchangeable weights. However this paper proposes a new way of bounding error terms for the bootstrap that does not rely on Edgeworth expansions as does the other theoretical work on convergence rates to date. In the independent case see [10, 12, 11, 14, 13]. Examples of proofs for dependent variables using Edgeworth expansions can be found in Lahiri [15] and the book by Politis et al. [17].

Instead of comparing two distributions directly, in this approach we compare their Stein operators on certain test functions and the expectation of their difference is used to bound the actual distance between distributions.

After defining the operators and notations, we start with the simple case of the consistency of the bootstrap distribution for the mean following Stein's proof of the central limit theorem closely. We then pass to the use of empirical processes to prove the general case of consistency for exchangeable weights in Section 6.3. This approach does not depend strongly on the hypothesis of independence, and we show in Section 6.4 that a weak neighborhood dependency structure does not invalidate the bootstrap procedure as long as a block-type bootstrap similar to Carlstein et al. [3] is used. Section 6.5 presents some examples; we give various

[1]Statistics, Sequoia Hall, Stanford University, Stanford, CA 94305, USA, and Unité de Biométrie, INRA-Montpellier, France. e-mail: `susan@stat.stanford.edu`

[2]Department of Statistics, University of Oxford, Oxford, OX1 3TG, UK. e-mail: `reinert@stats.ox.ac.uk`

dependency structures and the results that the neighborhood bootstrap provides in these cases.

6.2. Notation

Suppose we have a probability space $(\Omega, \mathcal{B}, P)$. We will call $E : \mathcal{X} \to \mathbb{R}$ the expectation associated to P on $\mathcal{X}$, the space of real-valued random variables defined on Ω that have finite expectation.

Our first theorem is a simple application of Stein's normal approximation theorems as developed in Stein (1986). We can show that smooth functions of means have bootstrap distributions that are approximately normal. Later a generalization of this result will be provided through the empirical process approach.

6.2.1. Exchangeable variables

Stein has introduced exchangeable variables in a randomization scheme that enables a characterization of the null space *ker* E of the expectation operator E. By definition, (X, X') is a pair of exchangeable variables if and only if the joint distribution of the pair (X, X') is identical to the distribution of (X', X), written sometimes $(X, X') \stackrel{d}{=} (X', X)$. In what follows (X, X') is always used to denote an exchangeable pair.

6.2.2. Operators of antisymmetric functions

Call $\mathcal{F}$ the set of bounded measurable antisymmetric functions defined on Ω^2.

We will denote by $\mathbf{T}$ the operator $\mathbf{T} : \mathcal{F} \longrightarrow \mathcal{X}$ which associates to every antisymmetric F in $\mathcal{F}$ the function:

$$\mathbf{T}F \quad \text{such that} \quad \mathbf{T}F(x) = E\big(F\big(X, X'\big)|x\big)$$

where $E(A|x)$ is the conditional expectation given $X = x$.

6.2.3. Bootstrap of the mean

To illustrate the method and for first results, let $\mathbf{X} = (x_1, x_2, \ldots, x_n)$ be a sample of independent identically distributed observations from a distribution P. Denote by $P_n = \frac{1}{n}\sum_1^n \delta_{x_i}$ the empirical distribution.

The bootstrap replaces the unknown distribution P by the empirical distribution P_n in the computation of statistical functionals of P. A bootstrap sample is characterized by a n-vector $\mathbf{k} = (k_1, k_2, \ldots, k_n)$ of the simplex

$$C_n = \big\{\mathbf{k} = (k_1, \ldots, k_n),\ k_1 + \cdots + k_n = n,\ k_i \geq 0,\ k_i \in \mathbb{N}\big\}$$

The original multinomial bootstrap uses $\xi_n = \frac{1}{n}\sum_1^n k_i \delta_{x_i}$, where the $\mathbf{k}$ is distributed as a multinomial, in particular $Ek_i = 1$. Later we will also consider more generally weighted bootstraps where the k_i's are simply exchangeable as in Praestgaard and Wellner (1993) for instance.

Firstly, here we study the original bootstrap with multinomial weights $\mathbf{k} = (k_1, \ldots, k_n)$. As a first step, Stein's method provides a straightforward way of showing that the bootstrap sum $W = \sum k_i X_i$ is asymptotically normal.

To this purpose, we employ the classical result from Stein (1986, Chapter 2). Assume that W is mean zero, variance 1, and that (W, W') is an exchangeable pair such that there is a $0 < \lambda < 1$ with

$$E\big(W'|W\big) = (1 - \lambda)W. \tag{6.1}$$

Stein (1986) proved the following theorem, with an improvement on the bounds by Baldi, Rinott and Stein (1989).

Theorem 6.1. *Let* (W, W') *be an exchangeable pair satisfying* (6.1) *and assume* $EW = 0$, $Var(W) = 1$. *For any continuous, bounded function* $h : \mathbb{R} \to \mathbb{R}$ *with bounded, piecewise continuous derivatives, we have*

$$\begin{aligned} &\big|Eh(W) - \Phi h\big| \\ &\qquad \le (\sup h - \inf h)\sqrt{E\Big(1 - \frac{1}{2\lambda}E((W-W')^2|W)\Big)^2} + \frac{1}{4\lambda}\big\|h'\big\|E\big|W - W'\big|^3. \end{aligned}$$

Following the classical Stein procedure we build **the exchangeable pair** $(\mathbf{k}, \mathbf{k}')$, using the auxiliary random variables (I, J), independent of X, as follows:

- Choose I according to the weight k_I, that is : $P(I = i) = k_I/n$, if $I = i$, decrease k_i by 1, $k_i' = k_i - 1$.
- Choose J uniformly between 1 and n,$P(J = j) = \frac{1}{n}$, if $J = j$, increase the component k_j by 1, and $k_j' = k_j + 1$.

Thinking of $\mathbf{k}$ representing the count vector when n balls are thrown into n urns, this exchangeable pair corresponds to choosing one of the balls at random, taking it out of the urn where it landed, and throwing it again; $\mathbf{k}'$ gives the new count vector.

$$\text{Put } W = \sum_i \frac{(k_i - 1)X_i}{\sqrt{\sum_i (X_i - \bar{X}_i)^2}},$$

the recentered bootstrap sum of $\mathbf{X}$, and its exchangeable counterpart

$$W' = \sum_i \frac{(k_i' - 1)X_i}{\sqrt{\sum_i (X_i - \bar{X}_i)^2}}.$$

To apply Stein's method, various conditional moments are needed. Throughout all calculations are conditional on the sample $\mathbf{X}$, so we will write $E(W)$ to mean $E^{\mathbf{X}}(W)$ throughout this section, and replace the X_i's by the observed values x_i. To simplify the calculations we can rescale $\mathbf{X}$ so that $\sum_i x_i = 0$ and we use $s_2^2 := \sum_i x_i^2$ and $s_4 := \sum_i x_i^4$. Observe first

$$\begin{aligned} E^{\mathbf{k}}\big(W' - W\big) = \frac{1}{s_2}E\big(W' - W|\mathbf{k}\big) &= E\Big(\sum k_i' x_i - \sum k_i x_i |\mathbf{k}\Big) \\ &= \sum E\big(k_i' - k_i|\mathbf{k}\big)\frac{x_i}{s_2} \\ &= \frac{1}{n^2}\sum E\big(n - k_i - k_i(n-1)|\mathbf{k}\big)\frac{x_i}{s_2} \\ &= \frac{1}{n}\sum(1 - k_i)\frac{x_i}{s_2} \\ &= -\frac{1}{n}W. \end{aligned}$$

This gives a version of Stein's contraction property $E(W' - W|W) = -\lambda W$ with $\lambda = \frac{1}{n}$.

For the computation of the variance we use Stein's clever trick:

$$\begin{aligned}
E\big(W'^2 - W^2\big) &= E\big(W'-W\big)\big(W'+W\big) = 0 \text{ by antisymmetry}\\
E\big(\big(W'-W\big)^2|W\big) &= E\big(W'^2 - W^2|W\big) + 2WE\big(W-W'|W\big)\\
&= E\big(W'^2 - W^2|W\big) + 2\lambda W^2.\\
\text{This implies that } E\big(W^2\big) &= \frac{1}{2\lambda}E\big(W'-W\big)^2.
\end{aligned}$$

Moreover,

$$\begin{aligned}
E^k\big(W'-W\big)^2 &= \frac{1}{n^2}\sum_i k_i \sum_j (x_i - x_j)^2 \frac{1}{s_2^2}\\
&= \frac{1}{n^2}\left(\sum_i k_i n \frac{x_i^2 + s_2^2}{s_2^2}\right)\\
&= \frac{1}{n}\left(\sum_i k_i x_i^2 + s_2\right)\\
&= \frac{1}{n}\left(\sum_i (k_i - 1)x_i^2 + 2s_2\right).\\
\text{Thus } E\big(W^2\big) &= \frac{1}{2\lambda}E\big(W'-W\big)^2 = s_2.
\end{aligned}$$

Next, an auxiliary computation gives

$$\begin{aligned}
E\left(\sum (k_i - 1)x_i^2\right)^2 &= \sum_{i,j} E(k_i-1)(k_j-1)x_i^2x_j^2\\
&= \sum_{i,j,i\neq j} Ex_i^2x_j^2E(k_i-1)(k_j-1) + \sum x_i^4 E(k_i-1)^2\\
&= \left(-\frac{1}{n}\right)\left(\sum_{i,j,i\neq j} x_i^2y_j^2\right) + \sum x_i^4\left(1-\frac{1}{n}\right)\\
&= \left(-\frac{1}{n}\right)(s_2^2 - s_4) + s_4\left(1-\frac{1}{n}\right)\\
&= s_4 - \frac{1}{n}s_2^2.
\end{aligned}$$

This yields

$$\begin{aligned}
E\left\{1 - E^k\left(\frac{n}{2}\big(W'-W\big)\right)\right\}^2 &= E\left\{1 - \frac{1}{2}\sum_i (k_i-1)x_i^2 - s_2\right\}^2\\
&= (1-s_2)^2 + \frac{1}{4}E\left(\sum_i (k_i-1)x_i^2\right)\\
&= (1-s_2)^2 + \frac{1}{4}\left(s_4 - \frac{1}{n}s_2^2\right).
\end{aligned}$$

and finally

$$E^k\big|W'-W\big|^3 = \sum_i\sum_j \left(\frac{k_i}{n^2}\right)|x_j - x_i|^3$$

thus

$$E\left|W'-W\right|^3 = \frac{1}{n^2}\sum_i\sum_j |x_j - x_i|^3$$

The sums will typically be small because of the weak law of large numbers.

Stein's theorem (Theorem 6.1) implies that if h is a piecewise continuously differentiable function whose derivative is bounded we will have the following result.

Proposition 6.1. *Let $x_i, 1 \leq i \leq n$ be fixed real numbers. let W have the bootstrap distribution of the centered mean. For all functions h that are piecewise continuously differentiable with bounded derivative, E_0 is the expectation with regards to the standard normal distribution.*

$$\begin{aligned}
&|Eh(W) - E_0 h| \\
&\quad \leq (\sup h - \inf h)\sqrt{(1-s_2)^2 + \frac{1}{4}\left(s_4 - \frac{1}{n}s_2^2\right)} + \frac{1}{4n}\sup\left|h'\right| \ \sum_i\sum_j |x_j - x_i|^3.
\end{aligned}$$

Remarks. Had we rescaled x_i so that $s_2 = 1$, this would have looked simpler with the right hand side being

$$(\sup h - \inf h)\frac{1}{2}\sqrt{s_4 - \frac{1}{n}} + \frac{1}{4n}\sup\left|h'\right|\sum_i\sum_j |x_j - x_i|^3.$$

Stein (1986, Chapter 3) gives a variety of bounds on the normal approximation to W in the presence of a bound on moments of $W' - W$. It is possible to include the dependent case by using neighborhoods of dependence as in Stein (1972) and Rinott and Rotar (1997).

We now expand this approach using the empirical processes. Stein's method for empirical measures was used by Reinert (1995) to prove a weak law of large numbers in a general setting.

6.2.4. Bootstrap of the empirical process

In this section, we use empirical processes to prove that the bootstrap works in the sense of having weak convergence of the bootstrap empirical process to the true underlying process. Instead of taking W to be a function of the bootstrap mean, we can use the same exchangeable pair (k, k') to create a pair of exchangeable processes. Of course a pair of processes (ξ, ξ') is said to be exchangeable when $(\xi, \xi) \stackrel{d}{=} (\xi', \xi)$.

The bootstrap is founded on an extension of von Mises plug-in principle. Let

$$P_n = \frac{1}{n}\sum_1^n \delta_{x_i}$$

be the empirical measure of the i.i.d. sample $\mathcal{X}$ drawn from the original measure P.

We would like to have the distribution of a plug-in estimate of a statistical functional $h(P)$. The estimator will be $h(P_n, n) = h_n(P_n)$, a function of both the empirical distribution and the sample size. This approach to the bootstrap is clearly laid out in Efron [8] and more recently in Lehmann (1998).

Ideally, if we knew what the process $G_n = \sqrt{n}(P_n - P)$ is, we would be able to establish measures of performance of the estimator.

This being unavailable, Efron [7] proposed the bootstrap *plug-in principle*, replacing G_n by its non parametric estimate: $\xi_n = \widehat{G}_n = \sqrt{n}(\hat{P}_n - P_n)$ where $\hat{P}_n$ is the bootstrap empirical distribution given that the sample was $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$, thus $\hat{P}_n = \frac{1}{n}\sum_1^n k_i \delta_{x_i}$, and we call

$$\xi_n = \widehat{G}_n = \frac{1}{\sqrt{n}} \sum_1^n (k_i \delta_{x_i} - \delta_{x_i}) \text{ the bootstrap empirical measure.}$$

In the case of a sample $\mathbf{X}$ consisting of i.i.d. observations from P, take G_P to be the P Brownian bridge. This is the Gaussian process with mean zero and covariance functions:

$$\mathrm{Cov}\big(\langle G_P, f\rangle, \langle G_P, g\rangle\big) = Ef(X)g(X) - Ef(X)Eg(X) = Pfg - PfPg$$

Then the distribution of G_P is the limiting distribution of G_n and should be the limiting distribution of ξ_n. We will prove this later. When the sample $\mathbf{X}$ consists of i.i.d. observations, this result is well-known; however, we will give explicit bounds in a metric we will specify in the next section. It will follow as a corollary from the case of general exchangeable weights. We turn to this now.

Præstgaard and Wellner [18] generalize the weighting scheme from multinomial weights to sequences w_i such that

1. $(w_1, w_2, \ldots, w_n)$ exchangeable for each n.
2. All the weights are positive, and each vector sums to n.

Examples for the exchangeable weight bootstrap include the case that $Y_1, Y_2, \ldots, Y_n$ are i.i.d. positive random variables, and we define the bootstrap weights by $w_{n,i} = \frac{Y_i}{\bar{Y}_n}$. If, for instance, the weights are exponentially distributed with mean 1, then we recover the Bayesian bootstrap of Rubin and Lo, see Præstgaard and Wellner [18].

We will derive bounds that involve the quantity

$$\lambda_n = \left\{ E\left(\frac{1}{n} \sum_i (w_i - 1)^2 - c^2 \right)^2 \right\}^{\frac{1}{2}}.$$

This corresponds to a weakening of the third assumption in Præstgaard and Wellner [18], that $\frac{1}{n}\sum_{j=1}^n (w_j - 1)^2 \longrightarrow c^2$ in probability.

We then generalize the bootstrap empirical measure to

$$\xi_n = \frac{1}{\sqrt{n}} \left(\sum w_j \delta_{x_j} - \sum_j \delta_{x_j} \right) = \frac{1}{\sqrt{n}} \sum (w_j - 1) \delta_{x_j},$$

and let $\zeta = cG_P$, where G_P is the P-Brownian bridge. Then we will prove the convergence of ξ_n to ζ. Moreover we will also be able to treat the case where the observations exhibit weak dependence. These results will follow from a more general result, assessing the distance to a Gaussian random measure with covariance matrix

$$\begin{aligned}
\mathrm{Cov}\big(\langle G_w, f\rangle, \langle G_w, g\rangle\big) &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n E(w_i - 1)(w_j - 1) Ef(X_i)g(X_j) \\
&= \frac{c^2}{n} \sum_{i=1}^n Ef(X_i)g(X_i) \\
&\quad + \frac{1}{n} \sum_{i=1}^n \sum_{j \neq i} E(w_i - 1)(w_j - 1) Ef(X_i)g(X_j).
\end{aligned}$$

Due to exchangeability, we have, for $i \neq j$, that $E(w_i-1)(w_j-1) = E(w_1-1)(w_2-1) := c_{1,1}$, and

$$0 \quad = \quad E\left(\sum_{i=1}^n (w_i-1)\right)^2 = nc^2 + n(n-1)c_{1,1},$$

hence

$$c_{1,1} = -\frac{1}{n-1}c^2,$$

and

$$\begin{aligned} \operatorname{Cov}\big(\langle G_w, f\rangle, \langle G_w, g\rangle\big) \quad &= \quad \frac{c^2}{n}\sum_{i=1}^n Ef(X_i)g(X_i) &(6.2)\\ &\quad - \frac{c^2}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i} Ef(X_i)g(X_j). &(6.3)\end{aligned}$$

We will condition on the sample; this corresponds to X_i having point mass distribution δ_{x_i}, if X_i is observed as x_i. In this case, the covariance given by (6.2) becomes

$$\begin{aligned} &\operatorname{Cov}\big(\langle G_{samp}, f\rangle, \langle G_{samp}, g\rangle\big)\\ &\quad = \frac{c^2}{n}\sum_{i=1}^n f(x_i)g(x_i) - \frac{c^2}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i} f(x_i)g(x_j)\\ &\quad = \frac{c^2}{n}\sum_{i=1}^n f(x_i)g(x_i) - \frac{c^2}{n(n-1)}\sum_{i=1}^n\sum_{j=1}^n f(x_i)g(x_j) + \frac{c^2}{n(n-1)}\sum_{i=1}^n f(x_i)g(x_i)\\ &\quad = \frac{c^2}{n-1}\sum_{i=1}^n f(x_i)g(x_i) - \frac{c^2}{n(n-1)}\sum_{i=1}^n\sum_{j=1}^n f(x_i)g(x_j). &(6.4)\end{aligned}$$

6.3. Gaussian approximations for empirical measures using Stein's method

We equip the space $M^f(\mathbb{R})$ of real-valued bounded Radon measures on $\mathbb{R}$ with the topology of vague convergence. Let $C_c(\mathbb{R})$ be the space of real-valued continuous functions on $\mathbb{R}$ with support contained in a compact set. Let $(\nu_n)_n$ be a family of measures in $M^f(\mathbb{R})$, and let ν be a measure in $M^f(\mathbb{R})$. We say that ν_n converges vaguely to ν, in short, $\nu_n \overset{v}{\Rightarrow} \nu$, if and only if for all functions $f \in C_c(\mathbb{R})$ we have

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

In words, if the projection of ν_n on any continuous function converges in $\mathbb{R}$ to the projection on ν. Here we use the notation

$$\langle \nu, f\rangle = \int f\, d\nu.$$

We will mainly consider cylinder-type functions $F \in C_b(M^f(\mathbb{R}))$ of the form

$$\begin{aligned} &F(\mu) = f\big(\langle \mu, \phi_1\rangle, \ldots, \langle \mu, \phi_m\rangle\big) &(6.5)\\ &\qquad \text{for an } m \in \mathbb{N}, f \in C_b^\infty\big(\mathbb{R}^m\big), \phi_i \in C_b^\infty(\mathbb{R}), i = 1, \ldots, m. &(6.6)\end{aligned}$$

For $\phi \in C_b^\infty(\mathbb{R})$, let $\|\phi\| = \sup_x |\phi(x)|$ and $\Delta\phi = \sum_{x.y} |\phi(x) - \phi(y)|$. Define the sets

$$\mathcal{C} \quad = \quad \left\{ \phi \in C_b^\infty(\mathbb{R}) \quad \text{with} \quad \|\phi\| \le 1, \sum_{j=1}^m \| \phi_{(j)} \| \le 1, \Delta\phi \le 1 \right\} \tag{6.7}$$

and

$$\begin{aligned} \mathcal{F} \quad := \quad & \Big\{ F \in C_b\big(M^f(\mathbb{R})\big) : F \text{ has the form (6.5) for an } m \text{ and } f \\ & \quad \text{with } \sum_{i=1}^m \| f_{(i)} \| \le 1, \sum_{i,j=1}^m \| f_{(i.j)} \| \le 1, \sum_{i,j,k=1}^m \| f_{(i,j,k)} \| \le 1, \\ & \quad \text{and } \phi_i \in \mathcal{C}, i = 1, \ldots, m \Big\} \end{aligned} \tag{6.8}$$

Here $f_{(j)}$ is the partial derivative of f in direction x_j, and similarly $f_{(i,j)}, f_{(i,j,k)}$ denote higher partial derivatives.

This construction is similar to the algebra of polynomials used by Dawson [4]. It is shown in Reinert [20] that this class of functions is convergence-determining for vague convergence.

It might be easier to understand this set by thinking of ϕ_i to be the indicator function of a convex set A_i. Then the class of functions reduces to functions of the type

$$\begin{aligned} & F(\mu) = f(\mu(A_1), \ldots, \mu(A_m)) \text{ for an } m \in \mathbb{N}, \text{ where} \\ & \quad A_1, \ldots, A_m \text{ are convex sets in } \mathbb{R}. \end{aligned}$$

From this it is easy to see that the convergence corresponds to convergence of finite-dimensional distributions, but now on the measure-valued level. However, for technical reasons we require the functions ϕ_i to be continuous and infinitely often differentiable (continuity is needed for the proof of Lemma 11 in Reinert [20]; differentiability is an assumption for convenience).

The following is based on Reinert [20]. We assume that $b : C_b^\infty(\mathbb{R}) \times C_b^\infty(\mathbb{R}) \to \mathbb{R}$ is a quadratic form such that, for any $m \in \mathbb{N}$ and for all $\phi_1, \ldots, \phi_m \in C_b^\infty(\mathbb{R})$,

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

is a symmetric, positive definite matrix. Similarly to the real-valued case, for $F \in \mathcal{F}$ with representation (6.5), we define the generator

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

We say that $\mathcal{A}$ is the generator associated with the operator b, or, the generator associated with the matrix B. (Note that $\mathcal{A}$ can be seen as a second order differential operator in terms of Gateaux derivatives.) Often we will abbreviate

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

Then the generator $\mathcal{A}$ reads

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

Let ζ be a random measure taking values in $M^f(\mathbb{R})$ almost surely such that, for all $m \in \mathbb{N}$, and for all $\phi_1, \ldots, \phi_m \in C_b^\infty(\mathbb{R})$,

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

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 $H \in \mathcal{F}$, we know we can write it

$$H(\nu) = h\big(\langle \nu, \psi_1\rangle, \ldots, \langle \nu, \psi_m\rangle\big), \tag{6.9}$$

the Stein equation corresponding to the Gaussian random measure ζ is

$$\begin{aligned} h\big(\underline{\langle \nu, \phi\rangle}\big) - Eh\big(\underline{\langle \zeta, \phi\rangle}\big) &= -\sum_{j=1}^{m} f_{(j)}\big(\underline{\langle \nu, \phi\rangle}\big)\langle \nu, \phi_j\rangle \\ &\quad + \sum_{j,k=1}^{m} f_{(j,k)}\big(\underline{\langle \nu, \phi\rangle}\big) b(\phi_j, \phi_k). \end{aligned} \tag{6.10}$$

This equation can be solved using the semigroup technique as in Barbour [1].

Lemma 6.1. *For each $H \in \mathcal{F}$ has the form* (6.9)*, there is a function $F \in \mathcal{F}$, and there is a function $f \in C_b^\infty(\mathbb{R}^m)$ such that $F(\nu) = f(\langle \nu, \psi_1\rangle, \ldots, \langle \nu, \psi_m\rangle)$, and $\|f^{(k)}\| \le \|h^{(k)}\|, k \in \mathbb{N}$.*

To prove a Gaussian approximation, we may employ the following result.

Proposition 6.2. *Let $(\eta_n)_{n\in\mathbb{N}}$ be a family of random measures taking values in $M^f(\mathbb{R})$ almost surely. Let ζ be a random measure taking values in $M^f(\mathbb{R})$ almost surely such that, for all $m \in \mathbb{N}, \phi_1, \ldots, \phi_m \in C_b^\infty(\mathbb{R})$,*

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

Let $\mathcal{A}$ be the generator associated with B. Let H be of the form (6.9) *and let F be the solution of the Stein equation* (6.10) *from Lemma* 6.1*. Then*

$$\big|EH(\eta_n) - EH(\zeta)\big| = \big|E\mathcal{A}F(\eta_n)\big|.$$

In particular, if for all $F \in \mathcal{F}$, we have that $E\mathcal{A}F(\eta_n) \to 0 \quad (n \to \infty)$, where $\mathcal{F}$ is given in (6.8), then Proposition 6.2 gives that $\mathcal{L}(\eta_n) \overset{w}{\Rightarrow} \mathcal{L}(\zeta) \quad (n \to \infty)$.

Proposition 6.2 assumes the existence of a Gaussian random measure ζ that is finite almost surely. In general, the almost sure finiteness is not guaranteed. Note also that Proposition 6.2 describes the distributional distance in terms of finite-dimemsional projections.

Let

$$B_n = \big(b_n(\phi_j, \phi_k)\big)_{j,k=1,\ldots,m} = \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j\ne i} \mathrm{Cov}(\phi_j(X_i), \phi_k(X_l))\right)_{j,k=1,\ldots,m} \tag{6.11}$$

In particular, Taylor expansion around $\sum_{j\in\Gamma_n(i)} \delta_{X_j}$ easily yields the following result.

Proposition 6.3. *For all $i, n \in \mathbb{N}$ let $\Gamma_n(i) \subset \{1, \ldots, n\}$ be a set such that for each $l \notin \Gamma_n(i)$, X_l is independent of X_i. Let $\gamma_n = \max_{i=1\ldots,n} |\Gamma_n(i)|$. Let ζ_n be a random measure taking values in $M^f(E)$ almost surely such that, for all $m \in \mathbb{N}, \phi_1, \ldots, \phi_m \in \mathcal{C}$,*

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

Then, for each $G \in C_{\mathcal{C}}(M^f(E))$ of the form (6.9) we have

$$\big| EG(\xi_n) - EG(\zeta_n) \big| \leq \frac{20}{\sqrt{n}} \gamma_n^2.$$

In the independent case this gives $\gamma = 1$, and

$$B_n = \left(\frac{1}{n} \sum_{i=1}^{n} \mathrm{Cov}\big(\phi_j(X_i), \phi_k(X_i)\big) \right)_{j,k=1,\ldots,m}. \tag{6.12}$$

Often we are interested in estimating the variance, in which case $\phi_j(x) = x$, for $j = 1, \ldots, m$, giving $\frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}(X_i)$ as variance in the Gaussian distribution. Note, though, that these test functions are not members of the family $\mathcal{F}$.

6.3.1. Application to the bootstrap: A general theorem

Let us consider the case of general weights, and not necessarily i.i.d. observations. Assume that $(w_1, \ldots, w_n)$ is a vector of weights such that

1. $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ is exchangeable.
2. All the weights are positive, and they sum to n.

Put

$$c^2 = c_n^2 = E(w_1 - 1)^2. \tag{6.13}$$

Define

$$\lambda_n = \left\{ E \left(\frac{1}{n} \sum_i (w_i - 1)^2 - c^2 \right)^2 \right\}^{\frac{1}{2}}. \tag{6.14}$$

Let

$$\xi_n = \frac{1}{\sqrt{n}} \Big(\sum w_j \delta_{x_j} - \sum_j \delta_{x_j} \Big) = \frac{1}{\sqrt{n}} \sum (w_j - 1) \delta_{x_j}.$$

Note that, due to exchangeability,

$$\begin{aligned} \lambda_n^2 &= \mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^{n} (w_i - 1)^2 \right) \\ &= \frac{1}{n} \mathrm{Var}\big((w_1 - 1)^2\big) + 2 \frac{n(n-1)}{n^2} \mathrm{Cov}\big((w_1 - 1)^2, (w_2 - 1)^2\big). \end{aligned}$$

Hence

$$\mathrm{Cov}\big((w_1 - 1)^2, (w_2 - 1)^2\big) = \frac{n-1}{2n} \left\{ \lambda_n^2 + \frac{1}{n} \mathrm{Var}\big((w_1 - 1)^2\big) \right\}. \tag{6.15}$$

In this case the exchangeable pair is naturally provided by the weights so that I and J are both picked uniformly at random from $\{1, 2, 3, \ldots, n\}$, and given that the current weights are w_I and w_J the new weights are $w_I{}' = w_J$ and $w_J{}' = w_I$, with $w'_k = w_k$ for $k \neq I, J$; call

$$\xi_n{}' = \frac{1}{\sqrt{n}} \sum_i \big(w_i{}' - 1\big)\delta_{x_i}.$$

Then $(\xi_n{}', \xi_n)$ is also an exchangeable pair and

$$\begin{aligned} \xi_n{}' - \xi_n &= \frac{1}{\sqrt{n}}(w_I\delta_{x_J} - w_J\delta_{x_J} + w_J\delta_{x_I} - w_I\delta_{x_I}) \\ &= \frac{1}{\sqrt{n}}(w_I - w_J)(\delta_{x_J} - \delta_{x_I}), \end{aligned}$$

and for all $\phi \in C_n^\infty(\mathbb{R})$,

$$\begin{aligned} E^{\mathcal{X},\mathbf{w}}\langle \xi_n{}' - \xi_n, \phi\rangle) &= E\big(\langle \xi_n{}' - \xi_n, \phi\rangle | \mathcal{X}, \mathbf{w}\big) \\ &= \frac{1}{\sqrt{n}}\frac{1}{n}\sum_i \frac{1}{n}\sum_j (w_i - w_j)\big(\phi(x_J) - \phi(x_I)\big) \\ &= 2\bigg(\frac{1}{\sqrt{n}}\frac{1}{n}\sum_i \frac{1}{n}\sum_j w_i\phi(x_j) - \frac{1}{n}\sum_j w_j\phi(x_j)\bigg) \\ &= \frac{2}{n}\bigg(\frac{1}{\sqrt{n}}\sum_j \phi(x_j) - \frac{1}{\sqrt{n}}\sum_j w_j\phi(x_j)\bigg) \\ &= -\frac{2}{n}\langle \xi_n, \phi\rangle. \end{aligned}$$

Thus we have, for all functions f and ϕ,

$$E f(\xi_n)\langle \xi_n, \phi\rangle = \frac{n}{4} E\big(f\big(\xi_n{}'\big) - f(\xi_n)\big)\big\langle \xi_n{}' - \xi_n, \phi\big\rangle. \tag{6.16}$$

In particular,

$$E\big\langle \xi_n{}' - \xi_n, \phi\big\rangle\big\langle \xi_n{}' - \xi_n, \psi\big\rangle = \frac{n}{4} E\langle \xi_n, \phi\rangle\langle \xi_n, \psi\rangle. \tag{6.17}$$

First we show a general result.

Proposition 6.4. *Let cG_{samp} be a Gaussian random measure given by* (6.4) *that is finite almost surely. We have that, for all $H \in \mathcal{F}$ of the form* (6.9),

$$\begin{aligned} \big|EH(\xi_n) - EH(cG_{samp})\big| &\leq R_{samp} \\ &:= \frac{1}{\sqrt{n}}\Big\{E\big(w_1^3 + 3w_1^2\big) + 2\big(E\big((w_1 - 1)^4\big)\big)^{\frac{1}{2}}\Big\} + 3\lambda_n, \end{aligned}$$

where c is given in (6.13).

Remark. It follows that, with $\mathcal{X}_n = \{x_1, x_2, \ldots, x_n\}$,

$$E\big|E^{\mathcal{X}_n} H(\xi_n) - EH(cG_{samp})\big| \leq R_{samp}.$$

So, using Markov's inequality:

$$P\left(\left|E^{\mathcal{X}_n} H(\xi_n) - EH(cG_{samp})\right| > \frac{\log n}{\sqrt{n}}\right) \quad \leq \quad \frac{\sqrt{n}}{\log n} R_{samp}.$$

Thus we get finite n results that are related to

$$\mathcal{L}(\xi_n) \stackrel{\mathcal{D}}{\Longrightarrow} \mathcal{L}(G_{samp}) \qquad \text{for almost all } \mathcal{X}_n,$$

provided that M is sufficiently large.

Proof. We abbreviate $G = G_{samp}$. To apply Proposition 6.2, all we need to bound is:

$$-\sum_{j=1}^{m} Ef_{(j)}\big(\underline{\langle \xi_n, \phi \rangle}\big)\langle \xi_n, \phi_j\rangle + c^2 \sum_{j,k=1}^{m} Ef_{(j,k)}\big(\underline{\langle \xi_n, \phi \rangle}\big) E\big(\langle G, \phi_j\rangle \langle G, \phi_k \rangle\big),$$

where F of the form (6.5) is the solution of the Stein equation for H from Lemma 6.1. From exchangeability (6.16) and Taylor expansion we have

$$\begin{aligned}
&\sum_{j=1}^{m} Ef_{(j)}\big(\underline{\langle \xi_n, \phi \rangle}\big)\langle \xi_n, \phi_j\rangle \\
&\qquad = \frac{n}{4} \sum_{j,k=1}^{m} Ef_{(j,k)}\big(\underline{\langle \xi_n, \phi \rangle}\big)\langle {\xi_n}' - \xi_n, \phi_j\rangle \langle {\xi_n}' - \xi_n, \phi_k\rangle + R_1,
\end{aligned}$$

where the remainder term R_1 can be bounded as

$$\begin{aligned}
|R_1| &\leq \frac{n}{4} \sum_{j,k,l=1}^{m} \|f_{(j,k,l)}\| \left(\frac{1}{\sqrt{n}}\right)^3 E\big|(w_I - w_J)^3\big| \\
&\leq \frac{1}{2\sqrt{n}} \sum_{j,k,l=1}^{m} \|f_{(j,k,l)}\| \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j \neq i} E\big|(w_i - w_j)^3\big| \\
&= \frac{1}{\sqrt{n}} \sum_{j,k,l=1}^{m} \|f_{(j,k,l)}\| \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j \neq i} E\big|w_i^3 - 3w_i^2 w_j\big| \\
&= \frac{1}{\sqrt{n}} \sum_{j,k,l=1}^{m} \|f_{(j,k,l)}\| \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j \neq i} E\big(w_i^3 + 3w_i^2 w_j\big),
\end{aligned}$$

as the weights are nonnegative. Using that $\sum_{j \neq i} w_j = n - w_i$ we obtain

$$\begin{aligned}
|R_1| &\leq \frac{1}{\sqrt{n}} \sum_{j,k,l=1}^{m} \|f_{(j,k,l)}\| \frac{1}{n^2} \sum_{i=1}^{n} E\big((n-1)w_i^3 + 3w_i^2 n - 3w_i^3\big) \\
&\leq \frac{1}{\sqrt{n}} \sum_{j,k,l=1}^{m} \|f_{(j,k,l)}\| E\big(w_1^3 + 3w_1^2\big),
\end{aligned}$$

where we used exchangeability of the weights for the last step. Hence, due to the assumptions on the function f,

$$|R_1| \quad \leq \quad \frac{1}{\sqrt{n}} E\big(w_1^3 + 3w_1^2\big).$$

Furthermore

$$
\begin{aligned}
&\frac{n}{4}\sum_{j,k=1}^{m} Ef_{(j,k)}\big(\underline{\langle \xi_n,\phi\rangle}\big)\big\langle {\xi_n}'-\xi_n,\phi_j\big\rangle\big\langle {\xi_n}'-\xi_n,\phi_k\big\rangle \\
&\quad= \frac{n}{4}\sum_{j,k=1}^{m} Ef_{(j,k)}\big(\underline{\langle \xi_n,\phi\rangle}\big)E^{\mathcal{X},\mathbf{w}}\big\langle {\xi_n}'-\xi_n,\phi_j\big\rangle\big\langle {\xi_n}'-\xi_n,\phi_k\big\rangle \\
&\quad= \frac{1}{4}\sum_{j,k=1}^{m} Ef_{(j,k)}\big(\underline{\langle \xi_n,\phi\rangle}\big)E^{\mathcal{X},\mathbf{w}}(w_I-w_J)^2\big(\phi_j(x_J)-\phi_j(x_I)\big)\big(\phi_k(x_J)-\phi_k(x_I)\big) \\
&\quad= \frac{1}{4n^2}\sum_{j,k=1}^{m} Ef_{(j,k)}\big(\underline{\langle \xi_n,\phi\rangle}\big) \\
&\qquad\times\sum_{i=1}^{n}\sum_{l\neq i}(w_i-w_l)^2\big(\phi_j(x_l)-\phi_j(x_i)\big)\big(\phi_k(x_l)-\phi_k(x_i)\big).
\end{aligned}
$$

It is an easy manipulation to write

$$
(w_i-w_l)^2 \quad=\quad (w_i-1)^2-2(w_i-1)(w_l-1)+(w_l-1)^2.
$$

Using symmetry, and that $\sum_{l\neq i}(w_l-1)=-(w_i-1)$, a straightfroward calculation shows that

$$
\begin{aligned}
&\frac{1}{4n^2}\sum_{i=1}^{n}\sum_{l\neq i}(w_i-w_l)^2\big(\phi_j(x_l)-\phi_j(x_i))(\phi_k(x_l)-\phi_k(x_i)\big) \\
&\quad= \frac{1}{2n^2}\sum_{l=1}^{n}\phi_j(x_l)\phi_k(x_l)\sum_i(w_i-1)^2+\frac{1}{2n}\sum_{i=1}^{n}\phi_j(x_i)\phi_k(x_i)(w_i-1)^2 \\
&\qquad-\frac{1}{2n^2}\sum_{i=1}^{n}\sum_{l\neq i}\phi_j(x_i)\phi_k(x_l)(w_i-1)^2-\frac{1}{2n^2}\sum_{i=1}^{n}\sum_{l\neq i}\phi_j(x_l)\phi_k(x_i)(w_i-1)^2 \\
&\qquad+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{l\neq i}\phi_k(x_l)\phi_j(x_i)(w_i-1)(w_l-1).
\end{aligned}
$$

This gives

$$
\begin{aligned}
&\sum_{j,k=1}^{m} Ef_{(j,k)}\big(\underline{\langle \xi_n,\phi\rangle}\big)Ec^2\langle G,\phi_j\rangle\langle G,\phi_k\rangle \\
&\qquad-\frac{n}{4}\sum_{j,k=1}^{m} Ef_{(j,k)}\big(\underline{\langle \xi_n,\phi\rangle}\big)\big\langle {\xi_n}'-\xi_n,\phi_j\big\rangle\big\langle {\xi_n}'-\xi_n,\phi_k\big\rangle \\
&\quad= c^2\sum_{j,k=1}^{m} Ef_{(j,k)}\big(\underline{\langle \xi_n,\phi\rangle}\big)E^{\mathbf{w}} \\
&\qquad\times\Bigg\{\langle G,\phi_j\rangle\langle G,\phi_k \\
&\qquad-\frac{1}{n}\sum_{i=1}^{n}\phi_j(x_i)\phi_k(x_i)+\frac{1}{n(n-1)}\sum_i\sum_{l\neq i}\phi_j(x_i)\phi_k(x_l)\Bigg\} \\
&\qquad+R_2+R_3+R_4 \\
&\quad= R_2+R_3+R_4,
\end{aligned}
$$

where

$$
R_2=\sum_{j,k=1}^{m} Ef_{(j,k)}\big(\underline{\langle \xi_n,\phi\rangle}\big)\frac{1}{n}\sum_{i=1}^{n}\phi_j(x_i)\phi_k(x_i)\Bigg\{c^2-\frac{1}{2}(w_i-1)^2-\frac{1}{2n}\sum_l(w_l-1)^2\Bigg\}
$$

and

$$R_3 = \sum_{j,k=1}^{m} Ef_{(j,k)}\big(\underline{\langle \xi_n, \phi \rangle}\big)\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}\phi_j(x_i)\phi_k(x_l)\frac{n-1}{n}\big\{(w_i-1)^2-c^2\big\}$$

and

$$\begin{aligned} R_4 &= -\sum_{j,k=1}^{m} Ef_{(j,k)}\big(\underline{\langle \xi_n, \phi \rangle}\big) \\ &\quad \times \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}\phi_j(x_i)\phi_k(x_l)\left\{\frac{1}{n}c^2 + (w_i-1)(w_l-1)\right\}. \end{aligned}$$

For R_2, we have

$$|R_2| \leq \frac{1}{2}\lambda_n + \frac{1}{2}\tau_n,$$

where

$$\tau_n = \sup_{\phi:\|\phi\|\leq 1} E\left|\frac{1}{n}\sum_{i=1}^{n}\phi(x_i)\big((w_i-1)^2-c^2\big)\right|. \tag{6.18}$$

Similarly, with $\phi(x_i) = \frac{1}{n-1}\sum_{l\neq i}\phi_j(x_i)\phi_k(x_l)$, we obtain

$$|R_3| \leq \tau_n.$$

To bound τ_n, from the Cauchy-Schwarz inequality,

$$\tau_n^2 \leq \sup_{\phi:\|\phi\|\leq 1} E\left(\frac{1}{n}\sum_{i=1}^{n}\phi(x_i)\big((w_i-1)^2-c^2\big)\right)^2,$$

and, due to the exchangeability,

$$\begin{aligned} & E\left(\frac{1}{n}\sum_{i=1}^{n}\phi(x_i)\big((w_i-1)^2-c^2\big)\right)^2 \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\phi^2(x_i)E\big((w_i-1)^2-c^2\big)^2 \\ &\quad + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\phi(x_i)\phi(x_j)E\big((w_i-1)^2-c^2\big)\big((w_j-1)^2-c^2\big) \\ &= \mathrm{Var}\big((w_1-1)^2\big)\frac{1}{n^2}\sum_{i=1}^{n}\phi^2(x_i) \\ &\quad + \mathrm{Cov}\big((w_1-1)^2,(w_2-1)^2\big)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\phi(x_i)\phi(x_j). \end{aligned}$$

Using (6.15),

$$
\begin{aligned}
\tau_n^2 &\le \frac{1}{n}\mathrm{Var}\big((w_1-1)^2\big) + \mathrm{Cov}\big((w_1-1)^2,(w_2-1)^2\big) \\
&= \frac{1}{n}\mathrm{Var}\big((w_1-1)^2\big) + \frac{n-1}{2n}\left\{\lambda_n^2 + \frac{1}{n}\mathrm{Var}\big((w_1-1)^2\big)\right\} \\
&\le \frac{1}{2}\lambda_n^2 + \frac{3}{2n}\mathrm{Var}\big((w_1-1)^2\big) \\
&\le \frac{1}{2}\lambda_n^2 + \frac{3}{2n}E\big((w_1-1)^4\big).
\end{aligned}
$$

Finally we need to bound R_4. From the Cauchy–Schwarz inequality,

$$
\begin{aligned}
|R_4|^2 &\le E\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{l\neq i}\phi_j(x_i)\phi_k(x_l)\left\{\frac{1}{n-1}c^2 + (w_i-1)(w_l-1)\right\}\right)^2 \\
&= \frac{1}{n^4}\sum_{i=1}^n\sum_{l\neq i}\sum_{s=1}^n\sum_{t\neq s}\phi_j(x_i)\phi_k(x_l)\phi_j(x_s)\phi_k(x_t) \\
&\quad \times E\left\{\frac{1}{n-1}c^2 + (w_i-1)(w_l-1)\right\}\left\{\frac{1}{n-1}c^2 + (w_s-1)(w_t-1)\right\}.
\end{aligned}
$$

Distinguishing concerning the overlap between i, l and s, t, we obtain

$$
\begin{aligned}
|R_4|^2 &\le \frac{n(n-1)}{n^4}\left\{E\left((w_1-1)(w_2-1)+\frac{c^2}{n-1}\right)^2\right\} + \frac{4n(n-1)(n-2)}{n^4} \\
&\quad \times E\left((w_1-1)(w_2-1)+\frac{c^2}{n-1}\right)\left((w_1-1)(w_3-1)+\frac{c^2}{n-1}\right) \\
&\quad + \frac{3n(n-1)(n-2)(n-3)}{n^4}E\left((w_1-1)(w_2-1)+\frac{c^2}{n-1}\right) \\
&\quad \times \left((w_3-1)(w_4-1)+\frac{c^2}{n-1}\right) \\
&= \frac{1}{n^4}\mathrm{Var}\left(\sum_i\sum_{\ell\neq i}(w_i-1)(w_\ell-1)\right) \\
&= \frac{1}{n^4}\mathrm{Var}\left(\sum_i (w_i-1)^2\right) \\
&= \frac{\lambda_n^2}{n^2};
\end{aligned}
$$

for the second to last step, we used that $\sum_{\ell\neq i}(w_\ell-1) = -(w_i-1)$. Hence

$$
|R_4| \le \frac{\lambda_n}{n},
$$

and

$$
\begin{aligned}
|R_2|+|R_3|+|R_4| &\le \left(\frac{1}{2}+\frac{1}{\sqrt{2}}\right)\lambda_n + \left(\frac{3}{2n}E\big((w_1-1)^4\big)\right)^{\frac{1}{2}} + \frac{\lambda_n}{n} \\
&\le \frac{2}{\sqrt{n}}\big(E\big((w_1-1)^4\big)\big)^{\frac{1}{2}} + 3\lambda_n.
\end{aligned}
$$

Collecting the bounds yields the assertion. □

Note that, in Proposition 6.4, the sample does not occur in the bound on the approximation. This is due to having uniformly bounded test functions – even if the sample is rather skewed, the test function will smoothen it, as, for $\phi \in \mathcal{C}^*$, $\phi(X) \leq 1$.

Now let us consider special cases in Proposition 6.4.

The independent case. The classical case is that the observations in the sample are independent, in which case $\gamma_n = 1$. From Proposition 6.3 thus gives

$$\left| EG\big(\sqrt{n}P_n\big) - EG(\zeta_n) \right| \leq \frac{20}{\sqrt{n}},$$

where ζ_n is a random measure taking values in $M^f(E)$ almost surely such that, for all $m \in \mathbb{N}, \phi_1, \ldots, \phi_m \in \mathcal{C}$,

$$\mathcal{L}\big(\langle \zeta_n, \phi_1 \rangle, \ldots, \langle \zeta_n, \phi_m \rangle\big) = \mathcal{M}VN_m(0, B_n)$$

with

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

see (6.12). From Proposition 6.4 we obtain also a limiting Gaussian measure, but now with covariance structure (6.4),

$$\mathrm{Cov}\big(\langle G_{samp}, f \rangle, \langle G_{samp}, g \rangle\big) = \frac{c^2}{n} \sum_{i=1}^n f(x_i) g(x_i) - \frac{c^2}{n(n-1)} \sum_{i=1} \sum_{j \neq i} f(x_i) g(x_j).$$

A law of large numbers argument shows that, when taking many bootstrap samples, this covariance approaches

$$\begin{aligned} \mathrm{Cov}\big(\langle G, f \rangle, \langle G, g \rangle\big) &= \frac{c^2}{n} \sum_{i=1}^n Ef(X_i) g(X_i) - \frac{c^2}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} Ef(X_i) g(X_j) \\ &= c^2 \Bigg\{ \frac{1}{n} \sum_{i=1}^n \mathrm{Cov}\big(f(X_i), g(X_i)\big) \\ &\qquad - \frac{1}{n} \sum_{i=1}^n Ef(X_i) \bigg(\frac{1}{n-1} \sum_{j \neq i} Eg(X_j) - Eg(X_i) \bigg) \Bigg\}. \end{aligned}$$

With B_n given in (6.12), we hence obtain that the limiting Gaussian measures, suitably scaled by c^2, differ by at most

$$\sup_{g \in \mathcal{C}} c^2 \frac{1}{n} \sum_{i=1}^n \left| Eg(X_i) - \frac{1}{n-1} \sum_{j \neq i} Eg(X_j) \right|. \tag{6.19}$$

Thus the independent bootstrap works provided that (6.19) is small – the Gaussian approximation for the bootstrap empirical measure, suitably scaled, is close to the Gaussian approximation for the empirical measure. In particular, in the i.i.d. case (6.19) vanishes.

Multinomial weights. Another classical case is that the weights are multinomial $\mathbf{k} = (k_1, \ldots, k_n)$, as in the introduction, the probability of cell i being $\frac{1}{n}$, for

$i = 1, \ldots, n$. More generally, we consider the M-out-of-n bootstrap. where M samples are drawn with replacement. Then each k_i is $\mathrm{Bin}(M, 1/n)$ distributed, so that $Ek_i = \frac{M}{n}$. Assigning the weights

$$w_i = \frac{n}{M} k_i$$

thus gives exchangeable weights satisfying $\sum w_i = 1$. In this case we have

$$c^2 \quad = \quad \frac{n^2}{M^2} E\left(k_1 - \frac{M}{n}\right)^2 = \frac{n^2}{M} \frac{M(n-1)}{n^2} = \frac{n-1}{M}.$$

Proposition 6.4 yields the following proposition.

Proposition 6.5. *With the notation from Proposition* 6.4*, for all $H \in \mathcal{F}$, with covariance operator* (6.4)*, conditioned on the sample we have*

$$\left| EH(\xi_n) - EH\left(\sqrt{\frac{n-1}{M}} G_{samp}\right) \right| \quad \leq \quad R_{mult},$$

where

$$|R_{mult}| \quad = \quad \frac{1}{\sqrt{n}} \left\{ 4 + 8\frac{n}{M} + \frac{n^2}{M^2} + 2\frac{n^{3/2}}{M^{3/2}} + 9\frac{n^3}{M^3} + 3\frac{n^{7/2}}{M^{7/2}} + \frac{6\sqrt{n}}{M} \right\}. \tag{6.20}$$

Proof. In view of Proposition 6.4 we need to bound Ew_1^2, Ew_1^3, $\mathrm{Var}((w_1 - 1)^2)$, and λ_n. Note that $\frac{M}{n} w_1$ is $\mathrm{Bin}(M, 1/n)$-distributed. Hence

$$Ew_1^2 = \left(\frac{n}{M}\right)^2 \left\{ \frac{M(n-1)}{n^2} + \frac{M^2}{n^2} < \frac{n}{M} + 1. \right.$$

Following Stuart and Ord [24], p. 76, we have

$$\begin{aligned} Ew_1^3 \quad &= \quad \left(\frac{n}{M}\right)^3 \left\{ \frac{M(M-1)(M-2)}{n^3} + 3\frac{M(M-1)}{n^2} + \frac{M}{n} \right\} \\ &< \quad 1 + 3\frac{n}{M} + \frac{n^2}{M^2}. \end{aligned}$$

Moreover,

$$\begin{aligned} \mathrm{Var}\big((w_1 - 1)^2\big) \quad &\leq \quad E(w_1 - 1)^4 \\ &= \quad \left(\frac{n}{M}\right)^4 \left\{ 3\frac{M^2(n-1)^2}{n^4} + \frac{M(n-1)}{n^2}\left(1 - 6\frac{n-1}{n^2}\right) \right\} \\ &< \quad 3\frac{n^2}{M^2} + \frac{n^3}{M^3}, \end{aligned} \tag{6.21}$$

where we used Stuart and Ord [24], p. 76, (3.13).

Now we bound λ_n. Due to exchangeability

$$\begin{aligned} \lambda_n^2 \quad &= \quad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n (w_i - 1)^2\right) \\ &= \quad \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}\big((w_i - 1)^2\big) + \frac{1}{n^2}\sum_{i=1}^n \sum_{j \neq i} \mathrm{Cov}\big((w_i - 1)^2, (w_i - 1)^2\big) \\ &= \quad \frac{1}{n}\mathrm{Var}\big((w_1 - 1)^2\big) + \frac{n-1}{n}\mathrm{Cov}\big((w_1 - 1)^2, (w_2 - 1)^2\big) \\ &\leq \quad 3\frac{n^5}{M^6} + \frac{n^6}{M^7} + \frac{n-1}{n}\mathrm{Cov}\big((w_1 - 1)^2, (w_2 - 1)^2\big), \end{aligned}$$

where we used (6.21) for the first term. To calculate the covariance, note that

$$\left(\frac{M}{n}w_1, \frac{M}{n}w_2, \frac{M}{n}(n-w_1-w_2)\right) \sim \mathcal{M}\left(M; \frac{1}{n}, \frac{1}{n}, \frac{n-2}{n}\right)$$

is multinomial. We use Stuart and Ord [24], p. 196, (5.126) to obtain

$$\begin{aligned}
&\frac{n-1}{n}\mathrm{Cov}\big((w_1-1)^2, (w_2-1)^2\big) \\
&\quad = \frac{n-1}{n}\left(\frac{n}{M}\right)^4 \left\{-\frac{M}{n^2}\left(\frac{(n-2)^2}{n^2} + \frac{2}{n^2}\right) + \frac{M^2}{n^4} + 2\frac{M^2}{n^4}\right\} \\
&\quad \leq \frac{3}{M^2}.
\end{aligned}$$

Using that, for $a, b \geq 0$, we have $(a+b)^{1/2} \leq a^{1/2} + b^{1/2}$, we obtain

$$\lambda_n \quad \leq \quad 3\frac{n^{5/2}}{M^3} + \frac{n^3}{M^{7/2}} + \frac{2}{M}. \tag{6.22}$$

Adding the bounds yields the assertion. □

We have

$$\xi_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n w_i \delta_{x_i} - \frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_{x_i}$$

as our bootstrap measure. This can be rewritten as

$$\xi_n = \frac{\sqrt{n}}{M}\sum_{i=1}^n k_i \delta_{x_i} - \frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_{x_i}.$$

Sometimes one might instead want to consider

$$\zeta_n = \frac{1}{\sqrt{M}}\sum_{i=1}^n k_i \delta_{x_i} - \frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_{x_i}.$$

We have

$$\zeta_n = \xi_n + \left(\frac{1}{\sqrt{M}} - \frac{\sqrt{n}}{M}\right)\sum_{i=1}^n k_i \delta_{x_i}.$$

Thus, for all $H \in \mathcal{F}$,

$$\big|EH(\xi_n) - EH(\zeta_n)\big| \leq \frac{|\sqrt{n} - \sqrt{M}|}{M}. \tag{6.23}$$

Note that the difference between two independent Gaussian random measures η and ρ with same mean and different covariance operators b and c can be bounded by

$$\zeta_{\mathcal{F}}\big(\mathcal{L}(\eta), \mathcal{L}(\rho)\big) \quad \leq \quad \sup_{\phi,\psi \text{ bounded as in (6.8)}} \big|b(\phi,\psi) - c(\phi,\psi)\big|.$$

Hence we could also approximate by $\sqrt{\frac{n}{M}}G$ instead of $\sqrt{\frac{n-1}{M}}G$, with an additional error of the order $\frac{1}{M}$.

Note that, in the multinomial case, another straightforward way of constructing an exchangeable pair is given in Section 2 from auxiliary random variables (I, J). I is chosen proportionally to the k_I's and J chosen uniformly between 1 and n. Then subtract one from that k_I and add one to the k_J.

- Pick I such that $P(I=i)=\frac{k_i}{n}$, $k'_I=k_I-1$.
- Pick J such that $P(J=j)=\frac{1}{n}$, $k'_J=k_J+1$.

Let $\xi'=\frac{1}{\sqrt{n}}\sum_1^n(k'_i\delta_{x_i}-\delta_{x_i})$. Note that we have a contraction property: For any ϕ,

$$\begin{aligned} E\big(\langle\xi'_n-\xi_n,\phi\rangle|\xi_n\big)=E^{\xi_n}\big(\langle\xi'_n-\xi_n,\phi\rangle\big)= &= \frac{1}{n}\sum_i\sum_j\frac{k_i}{n}\frac{1}{n}\big(\phi(x_j)-\phi(x_i)\big)\\ &= -\frac{1}{n}\langle\xi_n,\phi\rangle. \end{aligned}$$

Thus we could apply the exchangeable pair method directly with this construction, yielding slightly different bounds. For simplicity's sake, we omit this approach here; however, it will reappear later.

6.4. Dependency graphs

Here we will consider situations where there is some dependency between the variables, thus showing that this method has advantages over methods that use independence of the random variables. Let the vertex set V be of cardinal N, and $\{X_i, i\in V\}$ be the random variables of interest, say for simplicity's sake, with mean 0.

The dependency graph has edges (i,j) when X_i and X_j are dependent, we will denote this by $i\sim j$. We define $S_i=\{j: i\sim j, j\neq i\}$ the neighborhood of dependence for the random variable X_i. Put

$$\gamma_i=|S_i|,\qquad i\in V. \tag{6.24}$$

Let $n\leq N$, sample $X_1,X_2,\ldots X_n$, obtaining $\mathcal{X}=\{x_1,x_2,\ldots,x_n\}$ The dependency graph given this sample is deterministic. When sampling according to exchangeable weights as before, we again obtain the covariance structure (6.4). From Proposition (6.4) we immediately obtain the following proposition.

Proposition 6.6. *For the dependency graph bootstrap, with* $\xi_n=\frac{1}{\sqrt{n}}\sum_i(w_i-1)\delta_{X_i}$*, let* G_{samp} *be a Gaussian random measure with covariance structure given in* (6.4)*. Suppose that the sample is fixed. Then we have that, for all* $H\in\mathcal{F}$*,*

$$\big|EH(\xi_n)-EH(G_{samp})\big| \;\leq\; \frac{1}{\sqrt{n}}\left\{E\big(w_1^3+3w_1^2\big)+2\Big(E\big((w_1-1)^4\big)\Big)^{\frac{1}{2}}\right\}+3\lambda_n,$$

where c is given in (6.13)*.*

Conditioning on the sample, we have, from (6.4),

$$\begin{aligned} \mathrm{Cov}\big(\langle G_{samp},f\rangle,\langle G_{samp},g\rangle\big) &= \frac{c^2}{n}\sum_{i=1}^n f(x_i)g(x_i)\\ &\quad -\frac{c^2}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i, j\not\sim i} f(x_i)g(x_j) \end{aligned}$$

where $\mathrm{Cov}(\langle G_{samp},f\rangle,\langle G_{samp},g\rangle)$ is given in (6.4). Using a law of large numbers argument again, if we draw many bootstrap samples then this will approach

$$
\begin{aligned}
&\operatorname{Cov}(\langle G_{dep,boot}, f\rangle, \langle G_{dep,boot}, g\rangle)\\
&\quad= \frac{c^2}{n}\sum_{i=1}^n Ef(X_i)g(X_i) - \frac{c^2}{n(n-1)}\sum_{i=1}^n \sum_{j\neq i, j\sim i} Ef(X_i)g(X_j)\\
&\qquad - \frac{c^2}{n(n-1)}\sum_{i=1}^n \sum_{j\neq i, j\not\sim i} Ef(X_i)Eg(X_j)\\
&\quad= c^2\Bigg\{\frac{1}{n}\sum_{i=1}^n \operatorname{Cov}\big(f(X_i)g(X_i)\big) + \frac{1}{n}\sum_{i=1}^n Ef(X_i)\\
&\qquad \times\Bigg(Eg(X_i) - \frac{1}{n-1}\sum_{j\neq i, j\sim i} g(X_j) - \frac{1}{n-1}\sum_{j\not\sim i} Eg(X_j)\Bigg)\Bigg\}. \qquad (6.25)
\end{aligned}
$$

In the case that all observations are identically distributed like X, and that all neighorhoods are of the same size γ, this reduces to

$$
\begin{aligned}
&c^2\Bigg\{\operatorname{Cov}\big(f(X), g(X)\big) + Ef(X)\Bigg(\frac{\gamma-1}{n-1} - \frac{1}{n-1}\sum_{j\neq i, j\sim i} g(X_j)\Bigg)\Bigg\}\\
&\quad= c^2\Bigg\{\operatorname{Cov}\big(f(X), g(X)\big) + \frac{\gamma-1}{n-1}Ef(X)Eg(X)\\
&\qquad - \frac{1}{n(n-1)}\sum_{i=1}^n \sum_{j\neq i, j\sim i} Ef(X_i)g(X_j)\Bigg\}\\
&\quad= c^2\Bigg\{\operatorname{Cov}\big(f(X), g(X)\big) - \frac{1}{n(n-1)}\sum_{i=1}^n \sum_{j\neq i, j\sim i} \operatorname{Cov}\big(f(X_i), g(X_j)\big)\Bigg\}.
\end{aligned}
$$

If, for all $i \sim j$, we had that

$$\operatorname{Cov}\big(f(X_i), g(X_j)\big) = c(f,g)$$

was independent of i, j, then we obtain

$$c^2\Bigg\{\operatorname{Cov}\big(f(X), g(X)\big) - c(f,g)\frac{\gamma-1}{n-1}\Bigg\}.$$

The last summand goes to zero as $n \to \infty$.

From Proposition (6.3), however, the limiting covariance in the Gaussian approximation for $\sqrt{n}P_n$ is given by

$$b_n(f,g) = \frac{1}{n}\sum_{i=1}^n \operatorname{Cov}\big(f(X_i), g(X_i)\big) + \frac{1}{n}\sum_{i=1}^n \sum_{j\neq i, j\sim i} \operatorname{Cov}\big(f(X_i), g(X_l)\big). \qquad (6.26)$$

In the very homogeneous case this reduces to

$$\frac{1}{n}\sum_{i=1}^n \operatorname{Cov}\big(f(X_i), g(X_i)\big) + (\gamma-1)c(f,g).$$

Thus we do not recover the Gaussian distribution that reflects how $\sqrt{n}P_n$ relates to the true distribution. In this sense, this ordinary bootstrap does not work in presence of dependence, unless the neighbourhood size γ goes to zero as n tends to ∞. This is no surprise, as the dependence should bias the bootstrap.

6.4.1. Special procedure for bootstrapping dependent data

It is intuitively obvious that multinomial weights might not be the best choice; instead, the bootstrap should perform better if the neighbourhood dependence is taken into account. We now introduce a novel bootstrap procedure, based on bootstrapping whole neighborhoods of dependence (an idea related to the blockwise bootstrap). For simplicity, from now on we assume that

Assumption A. The neighborhood size is constant, that is, $\gamma_i = \gamma$ for all $i \in V$, where γ_i is given in (6.24). Put

$$\kappa = \frac{M}{\gamma} \tag{6.27}$$

and assume that κ is integer.

Let us define the empirical measure

$$Q_n = \frac{1}{\kappa} \sum_{i=1}^{\kappa} k_i \frac{1}{\gamma} \sum_{j \in S_i} \delta_{x_j}$$

and

$$P_n = \frac{1}{n} \sum_{j=1}^{n} \delta_{x_j}.$$

Then Q_n is the bootstrap measure from the following procedure: Choose indices according to the multinomial distribution $\mathcal{M}(\kappa; \frac{1}{n}, \ldots, \frac{1}{n})$. If index i is chosen, then we sample the whole dependency neighborhood of X_i. Similarly to the blockwise bootstrap, if the neighborhoods of dependence are small, then the bootstrap should work. To prove a Gaussian approximation, put

$$\xi_n = \sqrt{n}(Q_n - P_n) = \frac{\sqrt{n}}{M} \sum_{j=1}^{n} \delta_{x_j} \left(\sum_{i \in S_j} k_i - \frac{M}{n} \right).$$

Put weights

$$w_i = \frac{n}{M} \sum_{j \sim i} k_j. \tag{6.28}$$

Then we can treat this similarly to the section of multinomial weights. Note that the weights are generally not exchangeable any more, and that now the observations may be dependent. Yet, the k_i's are still independent of $\mathcal{X}$, and that $Ew_i = 1$.

Recall that the multinomial vector $\mathbf{k}$ can be viewed as resulting from κ independent ball tosses into n urns, where each urn has probability $\frac{1}{n}$ of being hit. Thus, writing

$$\mathbf{1}(i \to j) = \mathbf{1}(\text{ ball } i \text{ lands in urn } j),$$

we have that $((\mathbf{1}(i \to j))_{i=1,\ldots,\kappa})$ are independent Bernoulli $\frac{1}{n}$-variables, and

$$k_j = \sum_{i=1}^{\kappa} \mathbf{1}(i \to j).$$

To determine the approximating Gaussian measure here, we write

$$\begin{aligned} w_i &= \frac{n}{M} \sum_{j \sim i} \sum_{s=1}^{\kappa} \mathbf{1}(s \to j) \\ &= \frac{n}{M} N_i, \end{aligned}$$

where $N_i = \sum_{s=1}^{\kappa} \mathbf{1}(s \to S_i)$ is the number of balls that hit the neighborhood S_i. Thus N_i is a $\mathrm{Bin}(\kappa, \frac{\gamma_i}{n})$-random variable. We have

$$\begin{aligned} c_i^2 &= E(w_i - 1)^2 \\ &= \frac{n^2}{M^2} \mathrm{Var}(N_i) \\ &= \frac{n^2}{M^2} \frac{M}{\gamma} \frac{\gamma(n-\gamma)}{n^2} \\ &= \frac{n-\gamma}{M}. \end{aligned}$$

To obtain $\mathrm{Cov}(w_i, w_j)$ we use the following decomposition. Define, for any subset $A \subset \{1, \dots, n\}$, that

$$N_A = \sum_{s=1}^{\kappa} \mathbf{1}(s \to A).$$

Then $N_A \sim Bin(\kappa, \frac{|A|}{n})$. Moreover, for two sets A, B, the covariance $\mathrm{Cov}(N_A, N_B)$ can be calculated using a decomposition according to the dependence structure. Denote by $C_A = N_A - EN_A$ the centered counts. Then

$$\begin{aligned} \mathrm{Cov}(N_A, N_B) &= \mathrm{Cov}(C_{A\cap B} + C_{A\setminus B}, C_{A\cap B} + C_{B\setminus A}) \\ &= \mathrm{Var}(C_{A\cap B}) + \mathrm{Cov}(C_{A\cap B}, C_{B\setminus A}) \\ &\quad + \mathrm{Cov}(C_{A\setminus B}, C_{A\cap B}) + \mathrm{Cov}(C_{A\setminus B}, C_{B\setminus A}). \end{aligned}$$

From the binomial distribution, we obtain that

$$\mathrm{Var}(C_{A\cap B}) = \frac{M}{\gamma} \frac{|A\cap B|(n - |A\cap B|)}{n^2}.$$

Moreover, for C, D mutually disjoint, the triple $(N_C, N_D, N_{\{1,\dots n\}\setminus(C\cup D)})$ is trinomially distributed with parameters $(\frac{M}{\gamma}; \frac{|C|}{n}, \frac{|D|}{n}, \frac{n-|C\cup D|}{n})$. From Stuart and Ord [24], p. 105 and p. 196, we obtain that

$$\mathrm{Cov}(N_C, N_D) = -\frac{M}{\gamma} \frac{|C||D|}{n^2}.$$

This gives that

$$\begin{aligned} &\mathrm{Cov}(N_A, N_B) \\ &\quad = \frac{M}{\gamma} \frac{|A\cap B|(n - |A\cap B|)}{n^2} - \frac{M}{\gamma} \frac{|A\cap B||B\setminus A|)}{n^2} - \frac{M}{\gamma} \frac{|A\cap B||A\setminus B|)}{n^2} \\ &\qquad - \frac{M}{\gamma} \frac{|B\setminus A||A\setminus B|)}{n^2} \end{aligned}$$

$$
\begin{aligned}
&= \frac{M}{\gamma n^2}\big\{|A\cap B|(n-|A\cap B|) - |A\cap B||B\setminus A| - |A\cap B||A\setminus B| \\
&\qquad - |B\setminus A||A\setminus B|\big\} \\
&= \frac{M}{\gamma n^2}\big\{n|A\cap B| - (|A\cap B| + |B\setminus A|)(|A\cap B| + |A\setminus B|)\big\} \\
&= \frac{M}{\gamma n^2}\big\{n|A\cap B| - |A||B|\big\}. \qquad (6.29)
\end{aligned}
$$

Now we decompose the weights w_i and w_j according to their dependence structure,

$$
\begin{aligned}
w_i &= \frac{n}{M}(N_{S(i)\cap S(j)} + N_{S(i)\setminus S(j)}) \\
w_j &= \frac{n}{M}(N_{S(i)\cap S(j)} + N_{S(j)\setminus S(i)}).
\end{aligned}
\qquad (6.30)
$$

Hence

$$
\mathrm{Cov}(w_i, w_j) = \frac{n|S(i)\cap S(j)|}{\gamma M} - \frac{\gamma}{M}. \qquad (6.31)
$$

As we do not have exchangeable weights, we cannot apply Proposition 6.4 directly. However, note that in this case an exchangeable pair is again provided by the weights so that we choose a ball I at random and take it out of its bin and throw it again. Formally, choose an index $I \in \{1,\dots,n\}$ according to

$$
P(I=i) = \frac{k_i}{\kappa}, \qquad i = 1,\dots,n
$$

and choose an index $J \in \{1,\dots,n\}$ according to the discrete uniform distribution,

$$
P(J=j) = \frac{1}{n}, \qquad j = 1,\dots,n.
$$

If $I=i$ and $J=j$, put

$$
\begin{aligned}
k_i' &= k_i - 1 \\
k_j' &= k_j + 1 \\
k_\ell' &= k_\ell, \quad \ell \neq i, j.
\end{aligned}
$$

Define weights

$$
w_j' = \frac{n}{M}\sum_i \mathbf{1}(i\in S_j)k_i' \qquad (6.32)
$$

and

$$
{\xi_n}' = \frac{\sqrt{n}}{M}\sum_{i=1}^n \left(\sum_{i\in S_j}\frac{k_i'}{\gamma} - \frac{M}{n}\right)\delta_{x_j}.
$$

Then $({\xi_n}', \xi_n)$ is an exchangeable pair and

$$
\begin{aligned}
{\xi_n}' - \xi_n &= \frac{\sqrt{n}}{M}\sum_{i=1}^n\sum_{j\sim i}(k_j' - k_j)\delta_{x_i} \\
&= \frac{\sqrt{n}}{M}\sum_{\ell=1}^n\big\{\mathbf{1}(J\sim\ell) - \mathbf{1}(I\sim\ell)\big\}\delta_{x_\ell}, \qquad (6.33)
\end{aligned}
$$

and for all $\phi \in C_n^\infty(\mathbb{R})$,

$$\begin{aligned}
E^{\mathcal{X},\mathbf{w}}\big\langle \xi_n{}' - \xi_n, \phi \big\rangle\big) &= \frac{\sqrt{n}}{M}\sum_{\ell=1}^n \phi(x_\ell) E^{\mathcal{X},\mathbf{w}}\big\{\mathbf{1}(J\sim\ell) - \mathbf{1}(I\sim\ell)\big\} \\
&= \frac{\sqrt{n}}{M}\sum_{\ell=1}^n \phi(x_\ell)\sum_{i,j=1}^n \frac{k_i}{\kappa n}\big\{\mathbf{1}(j\sim\ell) - \mathbf{1}(i\sim\ell)\big\} \\
&= \frac{\sqrt{n}}{M^2 n}\sum_{\ell=1}^n \phi(x_\ell)\big\{\gamma M - n\gamma\sum_{i\sim\ell} k_i\big\} \\
&= \frac{\sqrt{n}\gamma}{M\sqrt{n}}\sum_{\ell=1}^n \phi(x_\ell)\bigg\{\frac{n}{M} - \frac{n}{M}\sum_{i\sim\ell} k_i\bigg\} \\
&= \frac{\gamma}{M\sqrt{n}}\sum_{\ell=1}^n \phi(x_\ell)(1-w_\ell) \\
&= -\frac{\gamma}{M}\langle \xi_n, \phi\rangle.
\end{aligned}$$

Thus we have, for all functions f and ϕ,

$$Ef(\xi_n)\langle \xi_n, \phi\rangle = \frac{\kappa}{2}E\big(f\big(\xi_n{}'\big) - f(\xi_n)\big)\big\langle \xi_n{}' - \xi_n, \phi\big\rangle.$$

In particular,

$$\frac{\kappa}{2}E\big\langle \xi_n{}' - \xi_n, \phi\big\rangle\big\langle \xi_n{}' - \xi_n, \psi\big\rangle \quad = \quad E\langle \xi_n, \phi\rangle\langle \xi_n, \psi\rangle.$$

We are in a similar situation as with exchangeable weights, hence we mimick the proof of Proposition 6.4. Let

$$\begin{aligned}
&\mathrm{Cov}\big(\langle G_{mult,dep,samp}, f\rangle, \langle G_{mult,dep,samp}, g\rangle\big) \\
&\quad = \frac{1}{nM}\sum_{i=1}^n (n-\gamma) f(x_i) g(x_i) \\
&\quad\quad + \frac{\gamma}{nM}\sum_{i=1}^n\sum_{j\neq i}\bigg\{\frac{n|S(i)\cap S(j)|}{\gamma^2} - 1\bigg\} f(x_i) g(x_j).
\end{aligned} \tag{6.34}$$

Proposition 6.7. *In the dependency graph bootstrap, suppose that* $(k_1,\dots,k_n)$ *are multinomial* $\mathcal{M}(M, 1/n, \dots, 1/n)$*-distributed, with* $M \le n$. *Let*

$$w_i = \frac{n}{M}\sum_{j\sim i} k_j.$$

Let $G_{mult,dep,samp}$ *be the centered Gaussian random measure given by* (6.34)*, that is, we condition on the sample. Then, for all* $H \in \mathcal{F}$,

$$\big|EH(\xi_n) - EH(G_{mult,dep,samp})\big| \quad \le \quad \frac{n}{M^{3/2}}\bigg\{\frac{7\gamma^2\sqrt{n}}{\sqrt{M}} + \frac{1}{2} + 11\sqrt{\gamma} + 4\gamma\bigg\}.$$

Remark. The above bound seems to improve with M, but not with n, which appears to be odd at first sight. However, note that

$$\frac{n}{M^{3/2}} \quad = \quad \bigg\{\frac{n}{M}\bigg\}^{\frac{3}{2}}\frac{1}{\sqrt{n}},$$

so that with M of the order of n, the bound improves with n.

Proof. Similarly to the proof of Proposition 6.4, we obtain

$$\begin{aligned}
&\sum_{j=1}^{m} Ef_{(j)}\big(\underline{\langle \xi_n, \phi\rangle}\big)\langle \xi_n, \phi_j\rangle \\
&\qquad = \frac{M}{2\gamma}\sum_{j,k=1}^{m} Ef_{(j,k)}\big(\underline{\langle \xi_n, \phi\rangle}\big)\big\langle {\xi_n}' - \xi_n, \phi_j\big\rangle\big\langle {\xi_n}' - \xi_n, \phi_k\big\rangle + R_1,
\end{aligned}$$

where

$$\begin{aligned}
|R_1| &\le \frac{\kappa}{2}\left(\frac{\sqrt{n}}{M}\right)^3 \sum_{r,s,t} E\big|\{\mathbf{1}(J\sim r)-\mathbf{1}(I\sim r)\}\{\mathbf{1}(J\sim s)-\mathbf{1}(I\sim s)\} \\
&\qquad\qquad \times\{\mathbf{1}(J\sim t)-\mathbf{1}(I\sim t)\}\big| \\
&\le \frac{\kappa}{2}\left(\frac{\sqrt{n}}{M}\right)^3 \sum_{r,s,t} E\{\mathbf{1}(J\sim r)+\mathbf{1}(I\sim r)\}\{\mathbf{1}(J\sim s)+\mathbf{1}(I\sim s)\} \\
&\qquad\qquad \times\{\mathbf{1}(J\sim t)+\mathbf{1}(I\sim t)\} \\
&= \frac{\kappa}{2}\left(\frac{\sqrt{n}}{M}\right)^3 \sum_{r,s,t}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{Ek_i}{\kappa n}\{\mathbf{1}(j\sim r)+\mathbf{1}(i\sim r)\}\{\mathbf{1}(j\sim s)+\mathbf{1}(i\sim s)\} \\
&\qquad\qquad \times\{\mathbf{1}(j\sim t)+\mathbf{1}(i\sim t)\} \\
&= \frac{\kappa}{2}\left(\frac{\sqrt{n}}{M}\right)^3 \frac{1}{n^4}\sum_{r,s,t}\sum_{i=1}^{n}\sum_{j=1}^{n}\{\mathbf{1}(j\sim r)+\mathbf{1}(i\sim r)\}\{\mathbf{1}(j\sim s)+\mathbf{1}(i\sim s)\} \\
&\qquad\qquad \times\{\mathbf{1}(j\sim t)+\mathbf{1}(i\sim t)\} \\
&= \kappa\left(\frac{\sqrt{n}}{M}\right)^3 \frac{1}{n^4}\sum_{r,s,t}\sum_{i=1}^{n}\sum_{j=1}^{n}\big\{\mathbf{1}(j\sim r)\mathbf{1}(j\sim s)\mathbf{1}(j\sim t) \\
&\qquad\qquad +\mathbf{1}(j\sim r)\mathbf{1}(j\sim s)\mathbf{1}(i\sim t) \\
&\qquad\qquad +\mathbf{1}(j\sim r)\mathbf{1}(i\sim s)\mathbf{1}(j\sim t) \\
&\qquad\qquad +\mathbf{1}(j\sim r)\mathbf{1}(i\sim s)\mathbf{1}(i\sim t)\big\},
\end{aligned}$$

where the last step followed by symmetry. Evaluating the sums gives

$$\begin{aligned}
|R_1| &\le \kappa\left(\frac{\sqrt{n}}{M}\right)^3 \frac{1}{n^2}n\sum_{r,s,t}\big|S(r)\cap S(s)\cap S(t)\big| + 3n\gamma\sum_{r,s}\big|S(r)\cap S(s)\big| \\
&= \kappa\left(\frac{\sqrt{n}}{M}\right)^3 \frac{1}{n}\sum_{r\ne s}\sum_{t\ne r,s}\big|S(r)\cap S(s)\cap S(t)\big| \\
&\qquad + 6n\gamma\sum_{r\ne s}\big|S(r)\cap S(s)\big| + n^2\gamma + 3n^2\gamma^2.
\end{aligned}$$

As

$$\begin{aligned}
\sum_{r\ne s}\big|S(r)\cap S(s)\big| &= \sum_{u}\sum_{r\ne s}\mathbf{1}(u\sim r)\mathbf{1}(u\sim s) \\
&= n\gamma(\gamma-1),
\end{aligned}$$

and

$$\begin{aligned}
\sum_{r\ne s}\sum_{t\ne r,s}\big|S(r)\cap S(s)\cap S(t)\big| &= \sum_{u}\sum_{r\ne s}\sum_{t\ne r,s}\mathbf{1}(u\sim r)\mathbf{1}(u\sim s)\mathbf{1}(u\sim t) \\
&= n\gamma(\gamma-1)(\gamma-2),
\end{aligned}$$

we obtain

$$\begin{aligned}
|R_1| &\le \kappa\left(\frac{\sqrt{n}}{M}\right)^3 \frac{1}{n^2}\big\{n^2\gamma(\gamma-1)(\gamma-2)+6n^2\gamma^2(\gamma-1)+n^2\gamma+3n^2\gamma^2\big\} \\
&= M\left(\frac{\sqrt{n}}{M}\right)^3 \big\{(\gamma-1)(\gamma-2)+6\gamma(\gamma-1)+1+3\gamma\big\} \\
&\le 7\gamma^2\frac{n^{3/2}}{M^2}. \qquad (6.35)
\end{aligned}$$

Furthermore, from (6.33),

$$\begin{aligned}
&\frac{M}{2\gamma}\sum_{j,k=1}^m Ef_{(j,k)}\big(\underline{\langle\xi_n,\phi\rangle}\big)\big\langle{\xi_n}'-\xi_n,\phi_j\big\rangle\big\langle{\xi_n}'-\xi_n,\phi_k\big\rangle \\
&\quad= \frac{M}{2\gamma}\frac{n}{M^2}\sum_{j,k=1}^m Ef_{(j,k)}\big(\underline{\langle\xi_n,\phi\rangle}\big) \\
&\quad\quad\times\sum_{s=1}^n\sum_{t=1}^n \phi_j(x_s)\phi_k(x_t)\big\{\mathbf{1}(J\sim s)-\mathbf{1}(I\sim s)\big\}\big\{\mathbf{1}(J\sim t)-\mathbf{1}(I\sim t)\big\} \\
&\quad= \frac{n}{2M\gamma}\sum_{j,k=1}^m Ef_{(j,k)}\big(\underline{\langle\xi_n,\phi\rangle}\big) \\
&\quad\quad\times\sum_{s=1}^n\sum_{t=1}^n \phi_j(x_s)\phi_k(x_t)\sum_{i,j}\frac{\gamma k_i}{Mn}\big\{\mathbf{1}(j\sim s)-\mathbf{1}(i\sim s)\big\}\big\{\mathbf{1}(j\sim t)-\mathbf{1}(i\sim t)\big\} \\
&\quad= \frac{1}{2M^2}\sum_{j,k=1}^m Ef_{(j,k)}\big(\underline{\langle\xi_n,\phi\rangle}\big)\sum_{s=1}^n\sum_{t\ne s}\phi_j(x_s)\phi_k(x_t) \\
&\quad\quad\times\sum_{i,j}k_i\big\{\mathbf{1}(j\sim s)-\mathbf{1}(i\sim s)\big\}\big\{\mathbf{1}(j\sim t)-\mathbf{1}(i\sim t)\big\} \\
&\quad\quad+\frac{1}{2M^2}\sum_{j,k=1}^m Ef_{(j,k)}\big(\underline{\langle\xi_n,\phi\rangle}\big)\sum_{s=1}^n\phi_j(x_s)\phi_k(x_s) \\
&\quad\quad\times\sum_{i,j}k_i\big\{\mathbf{1}(j\sim s)-\mathbf{1}(i\sim s)\big\}^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\sum_{j,k=1}^m Ef_{(j,k)}\big(\underline{\langle\xi_n,\phi\rangle}\big)E\langle G_{mult,dep,samp},\phi_j\rangle\langle G_{mult,dep,samp},\phi_k\rangle \\
&\qquad-\frac{M}{2\gamma}\sum_{j,k=1}^m Ef_{(j,k)}\big(\underline{\langle\xi_n,\phi\rangle}\big)\big\langle{\xi_n}'-\xi_n,\phi_j\big\rangle\big\langle{\xi_n}'-\xi_n,\phi_k\big\rangle
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j,k=1}^{m} Ef_{(j,k)}\big(\underline{\langle \xi_n, \phi\rangle}\big)E^{\mathbf{w}}\Big\{\langle G_{mult,dep,samp}, \phi_j\rangle\langle G_{mult,dep,samp}, \phi_k\rangle \\
&\qquad\qquad - \frac{1}{n}\sum_{i=1}^{n}(n-\gamma)\phi_j(x_i)\phi_k(x_i) \\
&\qquad\qquad - \frac{\gamma}{n}\sum_{i}\sum_{l\neq i}\left\{\frac{n|S(i)\cap S(j)|}{\gamma^2} - 1\right\}\phi_j(x_i)\phi_k(x_l)\Big\} \\
&\quad + R_2 + R_3 \\
&= R_2 + R_3,
\end{aligned}
$$

where

$$
\begin{aligned}
R_2 \quad = \quad & \sum_{j,k=1}^{m} Ef_{(j,k)}(\underline{\langle \xi_n, \phi\rangle}) \\
& \times \frac{1}{n}\sum_{s=1}^{n}\phi_j(x_s)\phi_k(x_s)\Big\{\frac{n-\gamma}{M} - \frac{n}{2M^2}\sum_{i,j}k_i\{\mathbf{1}(j\sim s) - \mathbf{1}(i\sim s)\}^2\Big\} \qquad (6.36)
\end{aligned}
$$

and

$$
\begin{aligned}
R_3 \quad = \quad & \sum_{j,k=1}^{m} Ef_{(j,k)}\big(\underline{\langle \xi_n, \phi\rangle}\big)\frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\phi_j(x_s)\phi_k(x_t) \\
& \times \Big\{n\mathrm{Cov}(w_s, w_t) - \frac{n^2}{2M^2}\sum_{i,j}k_i\big\{\mathbf{1}(j\sim s) - \mathbf{1}(i\sim s)\big\} \\
& \qquad\qquad \times \big\{\mathbf{1}(j\sim t) - \mathbf{1}(i\sim t)\big\}\Big\}. \qquad (6.37)
\end{aligned}
$$

We first bound R_2. We have that, for all $\phi, \psi \in \mathcal{C}$,

$$
\begin{aligned}
|R_2| \quad &\le \quad E\left|\frac{1}{n}\sum_{s=1}^{n}\phi(x_s)\psi(x_s)\Big\{n-\gamma - \frac{n}{2M^2}\sum_{i,j}k_i\{\mathbf{1}(j\sim s) - \mathbf{1}(i\sim s)\}^2\Big\}\right| \\
&= \quad E\Bigg|\frac{1}{n}\sum_{s=1}^{n}\phi(x_s)\psi(x_s)\Big\{n-\gamma - \frac{n}{2M^2}\sum_{i,j}k_i\{\mathbf{1}(j\sim s) \\
&\qquad\qquad - 2\mathbf{1}(j\sim s)\mathbf{1}(i\sim s) + \mathbf{1}(i\sim s)\}\Big\}\Bigg| \\
&= \quad E\left|\frac{1}{n}\sum_{s=1}^{n}\phi(x_s)\psi(x_s)\left(\frac{n-2\gamma}{2M} - \frac{n(n-2\gamma)}{2M^2}\sum_{i\sim s}k_i\right)\right| \\
&= \quad E\left|\frac{1}{n}\sum_{s=1}^{n}\phi(x_s)\psi(x_s)\left(\frac{n-2\gamma}{2M} - \frac{(n-2\gamma)}{2M}w_s\right)\right| \\
&= \quad \frac{(n-2\gamma)}{2M}E\left|\frac{1}{n}\sum_{s=1}^{n}\phi_j(x_s)\phi_k(x_s)(1-w_s)\right| \\
&\le \quad \frac{(n-2\gamma)}{2M}\left\{\mathrm{Var}\left(\frac{1}{n}\sum_{s=1}^{n}\phi(x_s)\psi(x_s)(1-w_s)\right)\right\}^{\frac{1}{2}},
\end{aligned}
$$

where we used the Cauchy–Schwarz inequality. Now,

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{n}\sum_{s=1}^{n}\phi(x_s)\psi(x_s)(1-w_s)\right) &\leq \frac{1}{n^2}\sum_{s=1}^{n}\mathrm{Var}(w_s)+\frac{1}{n^2}\sum_{s\neq t=1}^{n}\big|\mathrm{Cov}(w_s,w_t)\big| \\
&\leq \frac{n-\gamma}{Mn}+\frac{\gamma}{Mn^2}\sum_{s\neq t=1}^{n}\left\{1+\frac{n}{\gamma^2}\big|S(s)\cap S(t)\big|\right\} \\
&= \frac{n-\gamma}{Mn}+\frac{\gamma}{Mn^2}\left\{n(n-1)+\frac{n}{\gamma^2}n\gamma(\gamma-1)\right\} \\
&\leq \frac{n-\gamma}{Mn}+2\frac{\gamma}{M}.
\end{aligned}$$

Hence

$$|R_2|\leq\frac{(n-2\gamma)}{2M}\left\{\frac{n-\gamma}{Mn}+\frac{2\gamma}{M}\right\}^{\frac{1}{2}}\leq\frac{n}{M^{3/2}}\left\{\frac{1}{2}+\sqrt{\gamma}\right\}. \tag{6.38}$$

For R_3 we have, for all $\phi,\psi\in\mathcal{C}$,

$$\begin{aligned}
|R_3| &\leq E\Bigg|\frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\phi(x_s)\psi(x_t)\Bigg\{n\mathrm{Cov}(w_s,w_t) \\
&\qquad -\frac{n^2}{2M^2}\sum_{i,j}k_i\{\mathbf{1}(j\sim s)-\mathbf{1}(i\sim s)\}\{\mathbf{1}(j\sim t)-\mathbf{1}(i\sim t)\}\Bigg\}\Bigg| \\
&= E\Bigg|\frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\phi(x_s)\psi(x_t)\Bigg\{n\mathrm{Cov}(w_s,w_t)-\frac{n^2}{2M^2}\sum_{i,j}k_i\{\mathbf{1}(j\sim s)\mathbf{1}(j\sim t) \\
&\qquad -\mathbf{1}(i\sim s)\mathbf{1}(j\sim t)\}-\mathbf{1}(i\sim t)\mathbf{1}(j\sim s)+\mathbf{1}(i\sim s)\mathbf{1}(i\sim t)\}\Bigg\}\Bigg| \\
&= E\Bigg|\frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\phi(x_s)\psi(x_t)\Bigg\{\frac{n^2|S(s)\cap S(t)|}{\gamma M}-\frac{n\gamma}{M}-\frac{n^2}{2M^2}\Bigg\{\frac{M}{\gamma}\big|S(s)\cap S(t)\big| \\
&\qquad -\gamma\Bigg\{\sum_{i\sim s}k_i+\sum_{i\sim t}k_i\Bigg\}+n\sum_{i\in S(s)\cap S(t)}k_i\Bigg\}\Bigg\}\Bigg| \\
&\leq E\Bigg|\frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\phi(x_s)\psi(x_t)\Bigg\{\frac{n^2|S(s)\cap S(t)|}{\gamma M} \\
&\qquad -\frac{n^2}{2M^2\gamma}\Bigg\{M|S(s)\cap S(t)|-n\sum_{i\in S(s)\cap S(t)}k_i\Bigg\}\Bigg\}\Bigg| \\
&\quad +E\Bigg|\frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\phi(x_s)\psi(x_t)\Bigg\{-\frac{n\gamma}{M}+\frac{n^2\gamma}{2M^2}\Bigg\{\sum_{i\sim s}k_i+\sum_{i\sim t}k_i\Bigg\}\Bigg\}\Bigg|
\end{aligned}$$

so that

$$
\begin{aligned}
|R_3| \quad \leq \quad & E\left|\frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\phi(x_s)\psi(x_t)\right. \\
& \qquad \left. \times \frac{n^2|S(s)\cap S(t)|}{2\gamma M}\left\{1-\frac{n}{\gamma}M|S(s)\cap S(t)|\sum_{i\in S(s)\cap S(t)}k_i\right\}\right| && (6.39) \\
& + E\left|\frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\phi(x_s)\psi(x_t)\left\{-\frac{n\gamma}{M}+\frac{n^2\gamma}{M^2}\sum_{i\sim s}k_i\right\}\right|. && (6.40)
\end{aligned}
$$

by symmetry. We consider the two summands separately. Firstly, using the Cauchy-Schwarz inequality, for (6.39) we have

$$
\begin{aligned}
& E\left|\frac{1}{n}\sum_{s=1}^{n}\sum_{t\neq s}\phi(x_s)\psi(x_t)\frac{n|S(s)\cap S(t)|}{2\gamma M}\left\{1-\frac{n\gamma}{M|S(s)\cap S(t)|}\sum_{i\in S(s)\cap S(t)}k_i\right\}\right| \\
& \quad \leq \frac{1}{2}\left\{\frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\sum_{u=1}^{n}\sum_{v\neq u}\frac{n^2|S(s)\cap S(t)||S(u)\cap S(v)|}{\gamma^2M^2}\right. \\
& \qquad \left.\times\left|\mathrm{Cov}\left(\frac{n\gamma}{M|S(s)\cap S(t)|}\sum_{i\in S(s)\cap S(t)}k_i,\frac{n\gamma}{M|S(u)\cap S(v)|}\sum_{j\in S(u)\cap S(v)}k_j\right)\right|\right\}^{\frac{1}{2}}
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\sum_{u=1}^{n}\sum_{v\neq u}\frac{n^2|S(s)\cap S(t)||S(u)\cap S(v)|}{\gamma^2M^2} \\
& \qquad \times\left|\mathrm{Cov}\left(\frac{n\gamma}{M|S(s)\cap S(t)|}\sum_{i\in S(s)\cap S(t)}k_i,\frac{n\gamma}{M|S(u)\cap S(v)|}\sum_{j\in S(u)\cap S(v)}k_j\right)\right| \\
& \quad \leq \frac{2}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\frac{n^2|S(s)\cap S(t)|^2}{\gamma^2M^2}\mathrm{Var}\left(\frac{n}{\gamma}M|S(s)\cap S(t)|\sum_{i\in S(s)\cap S(t)}k_i\right) \\
& \qquad + 4\frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\sum_{u\neq s,t}\frac{n^2|S(s)\cap S(t)||S(u)\cap S(s)|}{\gamma^2M^2} \\
& \qquad \times\left|\mathrm{Cov}\left(\frac{n\gamma}{M|S(s)\cap S(t)|}\sum_{i\in S(s)\cap S(t)}k_i,\frac{n\gamma}{M|S(u)\cap S(s)|}\sum_{j\in S(u)\cap S(s)}k_j\right)\right| \\
& \qquad + \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\sum_{u\neq s,t}\sum_{v\neq u,s,t}\frac{n^2|S(s)\cap S(t)||S(u)\cap S(v)|}{\gamma^2M^2} \\
& \qquad \times\left|\mathrm{Cov}\left(\frac{n\gamma}{M|S(s)\cap S(t)|}\sum_{i\in S(s)\cap S(t)}k_i,\frac{n\gamma}{M|S(u)\cap S(v)|}\sum_{j\in S(u)\cap S(v)}k_j\right)\right| \\
& \quad \leq \frac{2}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\frac{n^2|S(s)\cap S(t)|^2}{\gamma^2M^2}\frac{(n-|S(s)\cap S(t)|)\gamma}{|S(s)\cap S(t)|M} \\
& \qquad + 4\frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\sum_{u\neq s,t}\frac{n^2|S(s)\cap S(t)||S(u)\cap S(s)|}{\gamma^2M^2}
\end{aligned}
$$

$$
\begin{aligned}
&\quad \times \frac{(n|S(s)\cap S(t)\cap S(u)| - \gamma|S(s)\cap S(t)||S(u)\cap S(s)|)}{M|S(s)\cap S(t)||S(u)\cap S(s)|} \\
&\quad + \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\sum_{u\neq s,t}\sum_{v\neq u,s,t}\frac{n^2|S(s)\cap S(t)||S(u)\cap S(v)|}{\gamma^2 M^2} \\
&\quad \times \frac{(n|S(s)\cap S(t)\cap S(u)\cap S(v)| - \gamma|S(s)\cap S(t)||S(u)\cap S(v)|)}{M|S(s)\cap S(t)||S(u)\cap S(v)|} \\
&\leq \frac{2}{n^2}\frac{n^3}{M^3}\frac{n\gamma(\gamma-1)}{\gamma} \\
&\quad + 4\frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\sum_{u\neq s,t}\frac{n^2}{\gamma M^2}\frac{n|S(s)\cap S(t)\cap S(u)| - \gamma|S(s)\cap S(t)||S(u)\cap S(s)|}{M} \\
&\quad + \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\sum_{u\neq s,t}\sum_{v\neq u,s,t}\frac{n^2}{\gamma M^2} \\
&\qquad\qquad \times \frac{n|S(s)\cap S(t)\cap S(u)\cap S(v)| - \gamma|S(s)\cap S(t)||S(u)\cap S(v)|}{M} \\
&\leq \frac{2}{n}\frac{n^3\gamma}{M^3} \\
&\quad + 4\frac{1}{n^2}\frac{n^2}{\gamma M^3}\left\{n\gamma(\gamma-1)(\gamma-2) - \sum_{s=1}^{n}\sum_{t\neq s}\sum_{u\neq s,t}\big|S(s)\cap S(t)\big|\big|S(u)\cap S(s)\big|\right\} \\
&\quad + \frac{1}{\gamma M^3}\Bigg\{n\gamma(\gamma-1)(\gamma-2)(\gamma-3) \\
&\qquad\qquad - \sum_{s=1}^{n}\sum_{t\neq s}\sum_{u\neq s,t}\sum_{v\neq u,s,t}\big|S(s)\cap S(t)\big|\big|S(u)\cap S(v)\big|\Bigg\} \\
&\leq \frac{2}{n}\frac{n^3\gamma}{M^3} + 8\frac{1}{n^2}\frac{n^2}{\gamma M^3}n\gamma(\gamma-1)(\gamma-2) + 8\frac{1}{n^2}\frac{n^2}{\gamma M^3}n\gamma(\gamma-1)(\gamma-2)(\gamma-3) \\
&< 2\frac{n^2\gamma}{M^3} + 8\frac{n\gamma^2}{M^3} + 8\frac{n\gamma^3}{M^3} \\
&= 2\frac{n^2\gamma}{M^3}\left\{1 + 4\frac{\gamma}{n} + 4\frac{\gamma^2}{n}\right\}.
\end{aligned}
$$

Thus we obtain for (6.39) as upper bound

$$
\frac{1}{2}\left\{2\frac{n^2\gamma}{M^3}\left\{1 + 4\frac{\gamma}{n} + 4\frac{\gamma^2}{n}\right\}\right\}^{\frac{1}{2}} \quad \leq \quad \frac{n\gamma}{M^{3/2}}\left\{3 + 2\frac{\gamma}{\sqrt{n}}\right\}.
$$

Next, for (6.40),

$$
\begin{aligned}
&E\left|\frac{1}{n^2}\sum_{s=1}^{n}\sum_{t\neq s}\phi(x_s)\psi(x_t)\left\{-\frac{n\gamma}{M} + \frac{n^2\gamma}{M^2}\sum_{i\sim s}k_i\right\}\right| \\
&\qquad = E\left|\frac{\gamma}{nM}\sum_{s=1}^{n}\sum_{t\neq s}\phi(x_s)\psi(x_t)(w_s-1)\right| \\
&\qquad \leq \frac{(n+1)\gamma}{nM}\left\{\sum_{s=1}^{n}\sum_{t=1}^{n}\big|\mathrm{Cov}(w_s,w_t)\big|\right\}^{\frac{1}{2}}
\end{aligned}
$$

$$
\begin{aligned}
&= 2\frac{(n+1)\gamma}{nM}\left\{\frac{n(n-\gamma)}{M} + \frac{\gamma}{M}\sum_{s=1}^{n}\sum_{t\neq s}\left|\frac{n|S(s)\cap S(t)|}{\gamma^2} - 1\right|\right\}^{\frac{1}{2}} \\
&\leq \frac{(n+1)\gamma}{nM}\left\{\frac{n(n-\gamma)}{M} + \frac{n^2(\gamma+2))}{M}\right\}^{\frac{1}{2}} \\
&\leq \frac{(n+1)\gamma}{nM}\left\{\frac{n}{\sqrt{M}} + \frac{n\sqrt{\gamma+2}}{\sqrt{M}}\right\} \\
&\leq \frac{n}{M^{3/2}}\left\{2\gamma + 6\sqrt{\gamma}\right\}.
\end{aligned}
$$

Hence

$$
|R_3| \quad \leq \quad \frac{n}{M^{3/2}}\left\{4\gamma + 10\sqrt{\gamma}\right\}.
$$

With (6.27), the assertion follows. □

Now, again using a law of large numbers argument, if we draw many samples then, approximately, the Gaussian covariance is given by

$$
\begin{aligned}
&\mathrm{Cov}\big(\langle G_{mult,depboot}, f\rangle, \langle G_{mult,depboot}, g\rangle\big) \\
&\quad = \frac{1}{nM}\sum_{i=1}^{n}(n-\gamma)Ef(X_i)g(X_i) \\
&\qquad + \frac{\gamma}{nM}\sum_{i=1}^{n}\sum_{j\neq i}\left\{\frac{n|S(i)\cap S(j)|}{\gamma^2} - 1\right\}Ef(X_i)g(X_j) \\
&\quad = \frac{n-\gamma}{M}\Bigg\{\frac{1}{n}\sum_{i=1}^{n}Ef(X_i)g(X_i) + \frac{1}{\gamma(n-\gamma)}\sum_{i=1}^{n}\sum_{j\neq i}\big|S(i)\cap S(j)\big|Ef(X_i)g(X_j) \\
&\qquad\qquad - \frac{\gamma}{n(n-\gamma)}\sum_{i=1}^{n}\sum_{j\neq i}Ef(X_i)g(X_j)\Bigg\} \\
&\quad = \frac{n-\gamma}{M}\Bigg\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{Cov}\big(f(X_i), g(X_i)\big) + \frac{1}{n}\sum_{i=1}^{n}Ef(X_i)Eg(X_i) \\
&\qquad\qquad + \frac{1}{\gamma(n-\gamma)}\sum_{i=1}^{n}\sum_{j\neq i}\big|S(i)\cap S(j)\big|\mathrm{Cov}\big(f(X_i), g(X_j)\big) \\
&\qquad\qquad + \frac{1}{\gamma(n-\gamma)}\sum_{i=1}^{n}\sum_{j\neq i}\big|S(i)\cap S(j)\big|Ef(X_i)Eg(X_j) \\
&\qquad\qquad - \frac{\gamma}{n(n-\gamma)}\sum_{i=1}^{n}\sum_{j\neq i}\mathrm{Cov}\big(f(X_i), g(X_j)\big) \\
&\qquad\qquad - \frac{\gamma}{n(n-\gamma)}\sum_{i=1}^{n}\sum_{j\neq i}Ef(X_i)Eg(X_j)\Bigg\} \\
&\quad = \frac{n-\gamma}{M}\Bigg\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{Cov}\big(f(X_i), g(X_i)\big) + \frac{1}{n}\sum_{i=1}^{n}Ef(X_i)Eg(X_i) \\
&\qquad\qquad + \frac{1}{\gamma(n-\gamma)}\sum_{i=1}^{n}\sum_{j\neq i, j\sim i}\big|S(i)\cap S(j)\big|\mathrm{Cov}\big(f(X_i), g(X_j)\big)
\end{aligned}
$$

$$
\begin{aligned}
&\qquad + \frac{1}{\gamma(n-\gamma)} \sum_{i=1}^{n} \sum_{j \neq i} \big|S(i) \cap S(j)\big| Ef(X_i) Eg(X_j) \\
&\qquad - \frac{\gamma}{n(n-\gamma)} \sum_{i=1}^{n} \sum_{j \neq i, j \sim i} \operatorname{Cov}\big(f(X_i), g(X_j)\big) \\
&\qquad - \frac{\gamma}{n(n-\gamma)} \sum_{i=1}^{n} \sum_{j \neq i} Ef(X_i) Eg(X_j) \Bigg\} \\
&= \frac{n-\gamma}{M} \Bigg\{ \frac{1}{n} \sum_{i=1}^{n} \operatorname{Cov}\big(f(X_i), g(X_i)\big) + \frac{1}{n} \sum_{i=1}^{n} Ef(X_i) Eg(X_i) \\
&\qquad + \frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i, j \sim i} \operatorname{Cov}\big(f(X_i), g(X_j)\big) \left(\frac{n}{\gamma(n-\gamma)} \big|S(i) \cap S(j)\big| - \frac{\gamma}{n-\gamma} \right) \\
&\qquad + \frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i} Ef(X_i) Eg(X_j) \left(\frac{n}{\gamma(n-\gamma)} \big|S(i) \cap S(j)\big| - \frac{\gamma}{n-\gamma} \right) \Bigg\}.
\end{aligned}
$$

In the very homogeneous case, where all X_i are identically distributed and all covariances are identical $Ef(X_i)g(X_j) = c(f,g)$ for $i \neq j$, this reduces to

$$
\begin{aligned}
&\frac{n-\gamma}{M} \Bigg\{ \operatorname{Cov}\big(f(X), g(X)\big) + Ef(X) Eg(X) \\
&\qquad + c(f,g) \frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i, j \sim i} \left(\frac{n}{\gamma(n-\gamma)} \big|S(i) \cap S(j)\big| - \frac{\gamma}{n-\gamma} \right) \\
&\qquad + Ef(X) Eg(X) \left(\frac{n}{\gamma(n-\gamma)} \gamma(n-\gamma) - \frac{\gamma(n-1)}{n-\gamma} \right) \Bigg\} \\
&= \frac{n-\gamma}{M} \Bigg\{ \operatorname{Cov}\big(f(X), g(X)\big) \\
&\qquad + c(f,g) \frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i, j \sim i} \left(\frac{n}{\gamma(n-\gamma)} |S(i) \cap S(j)| - \frac{\gamma}{n-\gamma} \right) \Bigg\}.
\end{aligned}
$$

For the last term, note that $\sum_{i=1}^{n} \sum_{j \neq i, j \sim i} |S(i) \cap S(j)|$ depends on the graph we consider.

Example 6.4.1. In one extreme very homogeneous case, case, $|S(i) \cap S(j)| = \gamma$ for all $i \sim j$,

$$
\sum_{i=1}^{n} \sum_{j \neq i, j \sim i} \big|S(i) \cap S(j)\big| \quad = \quad \sum_{i=1}^{n} \sum_{j \neq i} \big|S(i) \cap S(j)\big| = n\gamma(\gamma - 1),
$$

that is, the neighbors capture all the overlap. In this case, we obtain

$$
\begin{aligned}
\frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i, j \sim i} \left(\frac{n}{\gamma(n-\gamma)} \big|S(i) \cap S(j)\big| - \frac{\gamma}{n-\gamma} \right) &= \frac{n}{\gamma(n-\gamma)} \gamma(\gamma-1) - \frac{\gamma(\gamma-1)}{n-\gamma} \\
&= \gamma - 1,
\end{aligned}
$$

and the asymptotic covariance approximates

$$
\frac{n-\gamma}{M} \big\{ \operatorname{Cov}\big(f(X), g(X)\big) + c(f,g)(\gamma - 1) \big\},
$$

which is the correct distribution; scaled by $\frac{n-\gamma}{M}$; this bootstrap works.

Example 6.4.2. In the other extreme case, $|S(i) \cap S(j)| = 0$ for all $i \sim j$, no neighbours are shared, and we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i, j\sim i}\left(\frac{n}{\gamma(n-\gamma)}|S(i)\cap S(j)| - \frac{\gamma}{n-\gamma}\right) \quad = \quad -\frac{\gamma^2}{n-\gamma}.$$

This gives as approximating variance

$$\frac{n-\gamma}{M}\mathrm{Cov}\big(f(X), g(X)\big) - \frac{\gamma^2}{M}c(f,g).$$

For $M = M(n) \to \infty$, the last term approaches 0.

Example 6.4.3. In the k-nearest neighbour graph, n vertices are placed on a circle, and each vertex is connected to its k nearest neighbours to the left and to the right, so that each vertex has degree $d = 2k$. This is a Cayley graph, and we assume that $n > 2k$. If a vertex j has distance $\ell < k$ from vertex i, then it is easy to see that $|S_i) \cap S_j| = k + (k-\ell) - 1$. Assuming that the X_i are identically distributed, we obtain as asymptotic covariance

$$\begin{aligned}\frac{n-2k}{M}\Bigg\{&\mathrm{Cov}\big(f(X), g(X)\big) + Ef(X)Eg(X)\left(1 + \frac{n(3k-1)}{2(n-2k)}\right)\\ &+ \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i, j\sim i}\mathrm{Cov}\big(f(X_i), g(X_j)\big)\left(\frac{n}{2k(n-2k)}\big|S(i)\cap S(j)\big| - \frac{2k}{n-2k}\right)\Bigg\}.\end{aligned}$$

In the very homogeneous case where the covariance between $f(X_i)$ and $g(X_j)$ depends only on the distance $|i-j|$ as well as on f and g, this simplifies to

$$\begin{aligned}\frac{n-2k}{M}\Bigg\{&\mathrm{Cov}\big(f(X), g(X)\big) + Ef(X)Eg(X)\left(1 + \frac{n(3k-1)}{2(n-2k)}\right)\\ &+ \sum_{\ell=1}^{k}\mathrm{Cov}\big(f(X_1), g(X_{1+\ell})\big)\frac{n(2k-\ell-1) - 4k^2}{2k(n-2k)}\Bigg\}.\end{aligned}$$

In the last two cases, the approximating Gaussian measure has to be adjusted not only by $\frac{n-\gamma}{M}$, but also the covariance has to be adjusted separately. In Example 6.4.3 the covariance does not vanish for $n \to \infty$; we can still claim that the bootstrap, with this modification, works. In general, we will need to adjust the approximating Gaussian measure accordingly.

6.5. Applications of the bootstrap for dependent variables

The original sample is $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$, a set of n real numbers, and we suppose that the dependency graph given this sample is known. We remark that there is a simple relation between the size of the neighborhood S_i and the degree of the dependency graph $|S_i| = d_i + 1$. Following the theoretical development in section 4 we develop a

$\boxed{\text{Bootstrap algorithm for known dependency graphs}}$

Input arguments:
– B is the number of bootstrap resamples,
– n is the original sample size,
– M is the bootstrap sample size.
`Repeat` B `times:`

- `Pick` I_j `from` $\{1, 2, .., n\}$ `uniformly with replacement.`
- `Suppose` $I_j = \ell$ `then take all the observations` x_i `for` $i \in S_\ell = S_{I_j}$ `add these to the current bootstrap sample.`
- `Continue until the desired sample size` M `is attained.`

Note that the bootstrap sample may be larger than M in general.

6.5.1. Spatial process

Consider data which may be modeled by a marked spatial process, for instance different trees in a forest. Here it is plausible that close trees interfere with each other, whereas trees far away do not.

The variables of interest may be measurements of growth or fertility, we will call them X_i. The radius of interference can depend on the species type. Suppose we are in the simple instance where this is not the case and that the number of trees influenced is constant then we will define the dependency graph of each observation as it's d nearest neighbors, thus we have a fixed degree for the all the dependency graphs as a first approximation. This is the case when the dependency graph has a spatial meaning, and we can ignore second order effects, ie if one tree dwarfs a neighbor, this is to the advantage of the dwarfs' neighbors on the other side.

Simulating a particular dependence structure

We start with a simulation study in which the points are the vertices of a regular graph such as the 60 points of a bucky ball, this has a weak dependency structure since the degree of this regular graph is 3. Here is the projection of the front of the bucky ball as provided by a `matlab` function `bucki`.

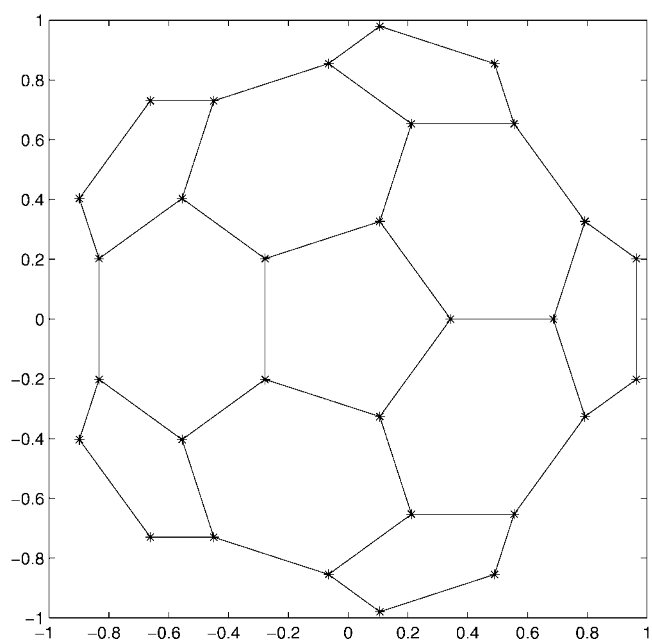

Figure 1: Part of the dependency graph for the Bucky ball

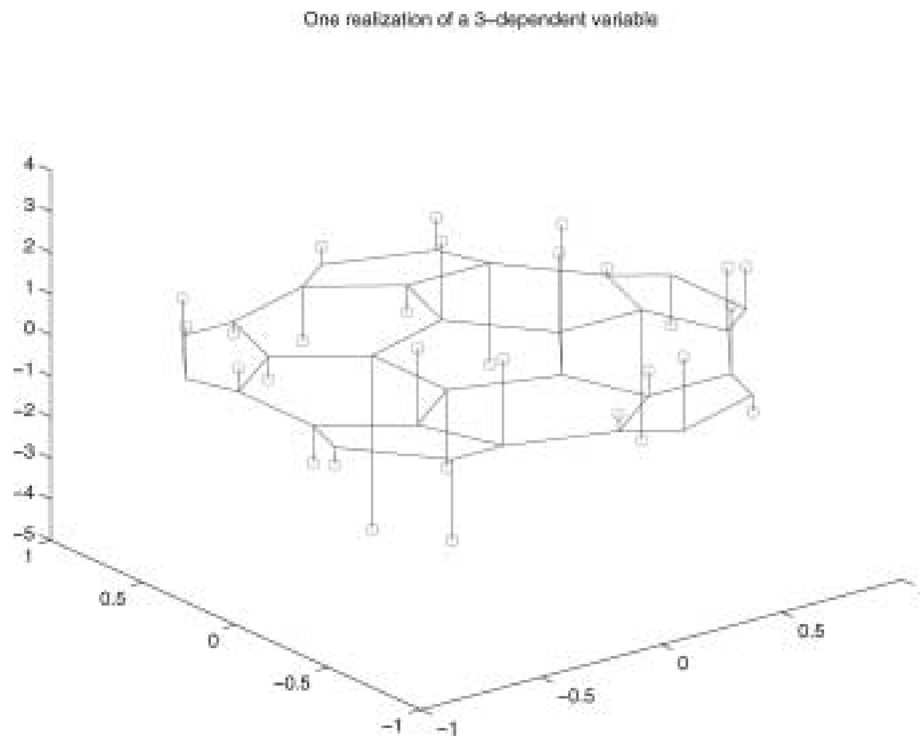

Figure 2: Values simulated at the vertices of the bucky ball

Here are some practical details on our implementation. We code dependency graphs as symmetric binary matrices A where a one at $A(i, j)$ means that the vertices i and j are dependent. Saying that the graph is regular, just means that we are forcing the row and column sums to be all equal to the same number d.

To simulate the spatial process with this dependence structure, we use a scheme proposed by Diaconis and Evans [5]. Let (V, E) be the dependency graph with incidence matrix A. Suppose that for each $e \in E$ we are given an iid random variable Z_e. We use these to form variables X_v as follows:

Step 1: Fix an orientation of the edges of the graph, $O(E)$.

Step 2: For each vertex v add the Z_e's for e going into the vertex and substract Z_g for v_g going out from the vertex. Call this sum X_v.

$$X_v = \sum_{(e,v)\in O(E)} Z_e - \sum_{(v,g)\in O(E)} Z_g$$

It is easy to see the random variables X_v have the desired dependency structure. These random variables have a Normal distribution if the Z_e's are Normal. This will be useful in checking the bootstrap empirically, but we have also used Poisson distributions for the Z_i's.

Suppose we have the regular 3-degree, 60 vertices Bucky ball as the dependency graph. We simulate first according to the scheme above, with the Z's independent standard normal. Figure 2 shows one realization of the process

We can check how this simulation scheme works by comparing the variance-covariances. Figure 3 shows the true covariances (which are known for this example)

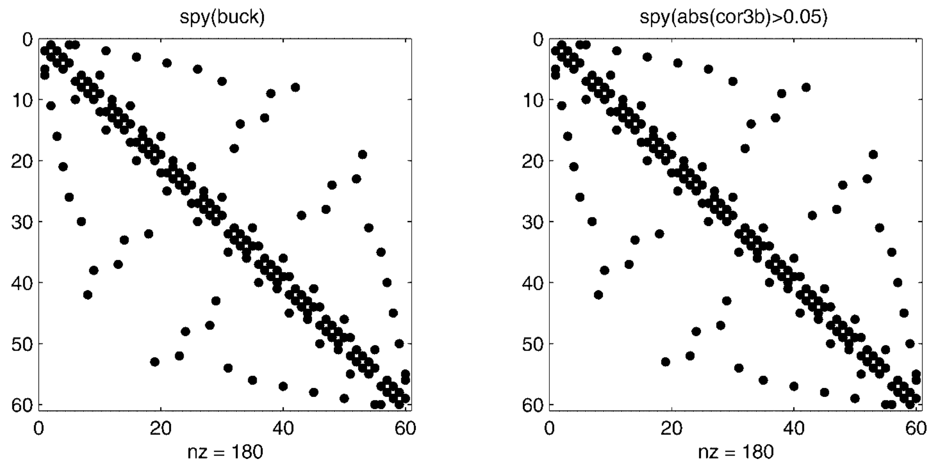

Figure 3: Comparison of simulated and known correlation structures

and the covariances simulated from a simulation of 1,000 realizations. There is a perfect fit if we ignore any correlation smaller than .05.

Now we have a simulation procedure for generating samples with the correct dependency structure, we actually know what the variance of a sample of size 60 should be, it should be around 3, the simulation with matlab gives:

```
simul2=simuldep(10000,buck,'normal');
mean(var(simul2))
    3.0428
 var(var(simul2))
    0.4054
```

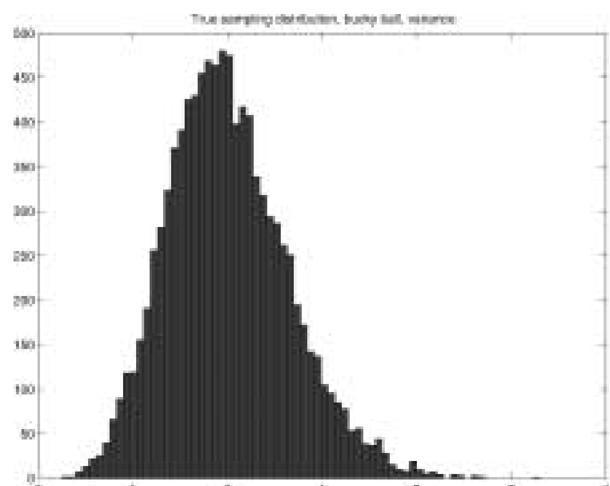

Figure 4: True sampling distribution of the variance

Here are the bootstrap results, starting with a sample $\mathcal{X}_1$ such that the variance of the original sample is 3.38, the bootstrap estimate was 3.33

```
 s1=simuldep(1,buck,'normal');
btvarres=simboot(s1,60,10000,'var');

 var(s1)
    3.3788
 mean(btvarres)
    3.3299
```

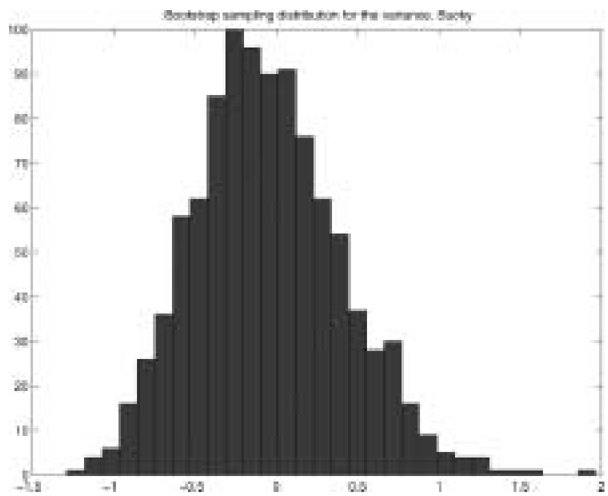

Figure 5: Bootstrap distribution of the variance

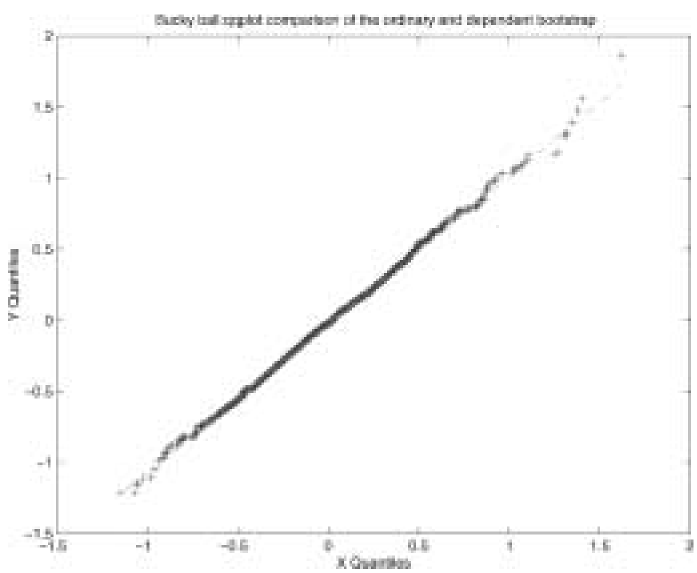

Figure 6: Two types of bootstrap

Figures 6 and 7 show the QQplots comparing the dependent bootstrap and the ordinary bootstrap and the bootstrap to the true sampling distribution for 1000 bootstrap samples.

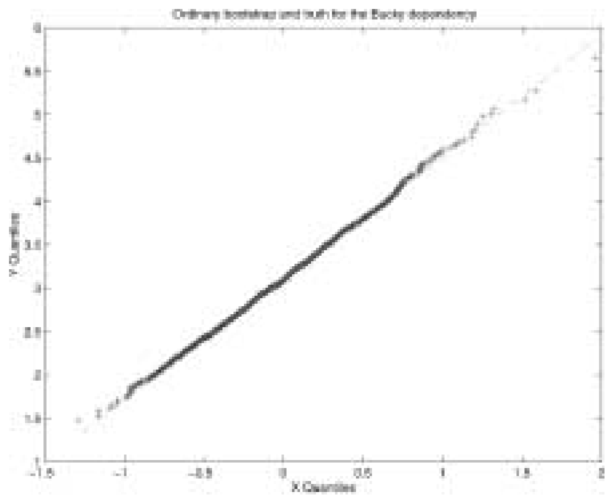

Figure 7: The bootstrap and sampling distributions

Let's look closer at cases where the graph is denser, thus building a stronger dependence.

Denser dependency graph

On a suggestion of Persi Diaconis we use the Cayley graph for a group generated by a small generating set for integers modulo n to generate denser graphs.

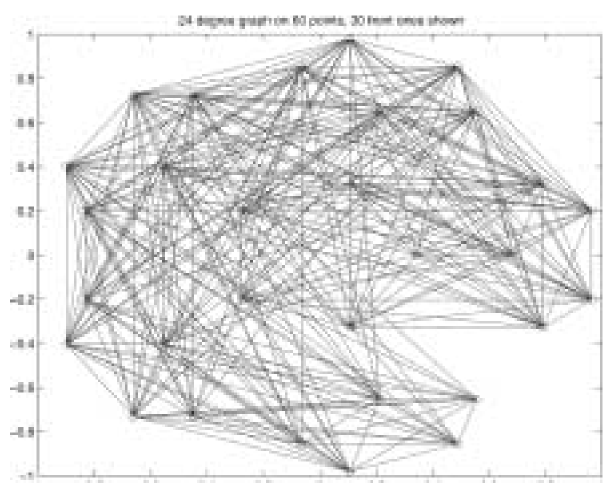

Figure 8: Dependency graph for 24 degrees on 60 points

S is the number of simulations from the truth. B is the number of bootstrap resamples, n is the sample size.

Note that it was unnecessary to do the simulations for the second column, it is simple to see that the variance is equal to the degree.

Table 1: Table of results for $B = 1000$ simulations on 9 different samples:

d	σ^2 (simul)	$\hat{s}^2$	$var(\hat{s}^2 - \sigma^2)$	$ave(var(\hat{s}^{*2} - \hat{s}^2))$	$ave(var(\hat{s}^{*2} - \hat{s}^2))$
				dependent	ordinary
	$S = 1000$	$n = 60$	$S = 1000$	$B = 1000,\ S = 100$	$B = 1000,\ S = 100$
2	2.0	1.77	0.185	0.18	0.12
3	3.0	2.70	0.431	0.30	0.28
4	4.1	5.37	0.706	0.69	0.54
6	6.2	6.23	1.506	1.11	1.32
8	8.2	9.46	2.321	2.05	2.31
10	10.1	9.00	3.726	2.80	3.23
12	12.3	9.26	5.754	4.01	4.60
14	14.3	9.39	6.883	6.01	6.36
16	16.3	18.98	8.643	6.32	7.82
24	24.6	19.99	22.41	17.87	19.12

Acknowledgements

We would like to thank Rabi Bhattacharya and Persi Diaconis for reading the manuscript carefully and to Charles Stein for many helpful discussions. This work was partially supported by NSF grant DMS-0241246 to SH and by an EPSRC grant no. GT/R52183/01 to GR.

References

[1] Barbour, A. D. (1988). Stein's method and Poisson process convergence. In *A Celebration of Applied Probability (Journal of Applied Probability, Volume 25A)*, pages 175–184. MR974580

[2] Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9: 1196–1217. MR630103

[3] Carlstein, E., Do, K.-A., Hall, P., Hesterberg, T., and Künsch, H. R. (1998). Matched-block bootstrap for dependent data. *Bernoulli*, 4(3): 305–328. MR1653268

[4] Dawson, D. A. (1993). *St. Flour School on Probability XXI (French)*. Springer-Verlag.

[5] Diaconis, P. and Evans, S. (2002). *Institut Henri Poincaré*, B38, 863–878. MR1955341

[6] Diaconis, P. and Zabell, S. (1991). Closed form summation for classical distributions: Variations on a theme of De Moivre. *Statistical Science*, 6: 284–302. MR1144242

[7] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7: 1–26. MR515681

[8] Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* Society for Industrial and Applied Mathematics. MR659849

[9] Götze, F. (1989). Edgeworth expansions in functional limit theorems. *The Annals of Probability*, 17: 1602–1634. MR1048948

[10] Hall, P. (1986). On the bootstrap and confidence intervals. *The Annals of Statistics*, 14: 1431–1452. MR868310

[11] Hall, P. (1988a). On symmetric bootstrap confidence intervals. *Journal of the Royal Statistical Society, Series B, Methodological*, 50: 35–45. MR954730

[12] Hall, P. (1988b). Rate of convergence in bootstrap approximations. *The Annals of Probability*, 16: 1665–1684. MR958209

[13] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag. MR1145237

[14] Hall, P. and Martin, M. A. (1991). On the error incurred using the bootstrap variance estimate when constructing confidence intervals for quantiles. *Journal of Multivariate Analysis*, 38: 70–81. MR1128937

[15] Lahiri, S. N. (1995). On the asymptotic behaviour of the moving block bootstrap for normalized sums of heavy-tail random variables. *The Annals of Statistics*, 23: 1331–1349. MR1353508

[16] Lehmann, E. (1998). *Elements of Large Sample Theory*. Springer-Verlag, New York. MR1663158

[17] Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling*. Springer, New York. MR1707286

[18] Præstgaard, J. and Wellner, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *The Annals of Probability*, 21: 2053–2086. MR1245301

[19] Reinert, G. (1994). *A weak law of large numbers for empirical measures via Stein's method and applications*. PhD thesis, Zurich. MR1330773

[20] Reinert, G. (1995). A weak law of large numbers for empirical measures via Stein's method. *The Annals of Probability*, 23: 334–354.
MR1330773

[21] 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 Anal.*, 56(2): 333–350. MR1379533

[22] Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *The Annals of Statistics*, 9: 1187–1195. MR630102

[23] Stein, C. (1986). *Approximate computation of expectations*. Institute of Mathematical Statistics. MR882007

[24] Stuart, A. and Ord, J. K. (1994). *Kendall's advanced theory of statistics. Volume 1. Distribution theory (Sixth edition)*. Edward Arnold Publishers Ltd. MR1280717