STEIN'S METHOD FOR DISCRETE GIBBS MEASURES

PETER EICHELSBACHER AND GESINE REINERT

Abstract. Stein's method provides a way of bounding the distance of a probability distribution to a target distribution μ. Here we develop Stein's method for the class of *discrete Gibbs measures* with a density e^V, where V is the energy function. Using size bias couplings, we treat an example of Gibbs convergence for strongly correlated random variables due to Chayes and Klein [6]. We obtain estimates of the approximation to a grand-canonical Gibbs ensemble. As side results, we slightly improve on the Barbour, Holst and Janson [4] bounds for Poisson approximation to the sum of independent indicators, and in the case of the geometric distribution we derive better non-uniform Stein bounds than Brown and Xia [5].

Introduction

Stein [17] introduced an elegant method for proving convergence of random variables towards a standard normal variable. Barbour [2], [3] and Götze [10] developed a dynamical point of view of Stein's method using time-reversible Markov processes. If μ is the stationary distribution of a homogeneous Markov process with generator $\mathcal{A}$, then $X \sim \mu$ if and only if $E\mathcal{A}g(X) = 0$ for all functions g in the domain $\mathcal{D}(\mathcal{A})$ of the operator $\mathcal{A}$. For any random variable W and for any suitable function f, to assess the distance $|Ef(W) - \int f d\mu|$ we first find a solution g of the equation

$$\mathcal{A}g(x) = f(x) - \int f d\mu.$$

If g is in the domain $\mathcal{D}(\mathcal{A})$ of $\mathcal{A}$, then we obtain

$$\left| Ef(W) - \int f d\mu \right| = |E\mathcal{A}g(W)|. \tag{0.1}$$

Bounding the right-hand side of (0.1) for a sufficiently large class of functions f leads to bounds on distances between the distributions. Here we will mainly focus on the *total variation distance*, where a point indicator $f(x) = f_k(x) = 1_{\{x=k\}}$ is the type of test function to consider.

2000 *Mathematics Subject Classification.* 60E05 (primary); 60F05, 60E15, 82B05 (secondary).

Key words and phrases. Stein's method, Gibbs measures, birth and death processes, size bias coupling.

Research partially supported by the DAAD Foundation and the British Council, ARC Project, Contract No. 313-ARC-00/27749 and by the RiP program at Oberwolfach, Germany.

Such distributional bounds cannot only be employed to assess the limiting behaviour of the quantity considered, but they are also of particular interest for example when one has only few observations, or when there is considerable dependence in the data slowing down the convergence. In such a case even a relatively large bound can be informative for lack of a better approximation.

In Section 1 we introduce Stein's method using the generator approach for *discrete Gibbs measures*, which are probability measures on $\mathbb{N}_0$ with probability weights $\mu(k)$ proportional to $\exp(V(k))$ for some function $V : \mathbb{N}_0 \to \mathbb{R}$. The discrete Gibbs measures include the classical distributions Poisson, binomial, geometric, negative binomial, hypergeometric and the discrete uniform, to name but a few. One can construct simple birth-death processes, which are time-reversible, and which have a discrete Gibbs measure as its equilibrium measure. In the context of spatial Gibbs measures this connection was introduced by Preston [15].

In Section 2 we not only recall bounds for the increments of the solution of the Stein equation from [5], but we also derive bounds on the solution itself, in terms of potential function of the Gibbs measure, see Lemma 2.9, Lemma 2.15, and Lemma 2.19. The bounds, which to our knowledge are new, are illustrated for the Poisson, the binomial, and the geometric distribution.

For nonnegative random variables, the size bias coupling is a very useful approach to disentangle dependence. Its formulation and its application to assess the distance to Gibbs measures is described in Section 3. We compare the distributions by comparing their respective generators, an idea also used in [11], while paying special attention to the case that the domains of the two generators are not identical. The size bias coupling then naturally leads to Theorem 3.9 and Corollary 3.13.

Section 4 applies these theoretical results to assess the distance to a Gibbs distribution for the law of a sum of possibly strongly correlated random variables. The main results are Theorem 4.5 and Proposition 4.7, where we give general bounds for the total variation distance between the distribution of certain sums of strongly correlated random variables and discrete Gibbs distributions of a grand-canonical form, see [16, Section 1.2.3]. In particular Theorem 4.5 gives a bound on the rate of convergence for the qualitative results in [6] by bounding the rate of convergence.

Considering two examples with non-trivial interaction we obtain bounds to limiting non-classical Gibbs distributions. Our bound on the approximation error is phrased in terms of the particle number and the average density of the particles.

Summarizing, the main advantage of our considerations is the application of Stein's method to models with interaction described by Gibbs measures. When applying our bounds on the solution of the Stein equation to the Poisson distribution and the geometric distribution, surprisingly we obtain improved bounds for these well-studied distributions. Thus our investigation of discrete Gibbs measures serves also as a vehicle for obtaining results for classical discrete distributions.

The results presented here will also provide a foundation for introducing Stein's method for spatial Gibbs measures and Gibbs point processes, in forthcoming work.

1. Gibbs measures and birth-death processes

1.1. **Birth-death processes.** A birth–death process $\{X(t), t \in \mathbb{R}\}$ is a Markov process on the state space $\{0, 1, \ldots, N\}$, where $N \in \mathbb{N}_0 \cup \{\infty\}$, characterized by (non-negative) birth rates $\{b_j, j \in \{0, 1, \ldots, N\}\}$ and (non-negative) death rates $\{d_j, j \in \{1, \ldots, N\}\}$ and has a generator

$$(\mathcal{A}h)(j) = b_j \left(h(j+1) - h(j)\right) - d_j \left(h(j) - h(j-1)\right) \quad \text{with} \quad j \in \{0, 1, \ldots, N\}. \tag{1.1}$$

It is well known that for any given probability distribution μ on $\mathbb{N}_0$ one can construct a birth–death process which has this distribution as its stationary distribution. For $N = \infty$, recurrence of the birth–death process $X(\cdot)$ is equivalent to $\sum_{n\geq 1} \frac{d_1 \cdots d_n}{b_1 \cdots b_n} = \infty$. The process $X(\cdot)$ is ergodic if and only if the process is recurrent and $c := 1 + \sum_{n\geq 1} \frac{b_0 \cdots b_{n-1}}{d_1 \cdots d_n} < \infty$. For $N < \infty$, irreducibility and hence ergodicity holds if $b_j > 0$, $j = 0, 1, \ldots, N-1$, $b_N = 0$ and $d_j > 0$, $j = 1, \ldots, N$.

In either case the stationary distribution of the ergodic process is given by $\mu(0) = 1/c$ and

$$\mu(n) = \mu(0) \frac{b_0 \cdots b_{n-1}}{d_1 \cdots d_n}. \tag{1.2}$$

For any given probability distribution μ on $\mathbb{N}_0$ these recursive formulas give the corresponding class of birth–death processes which have μ as the stationary distribution. For the choice of a unity per capita death rate $d_j = j$ one simply obtains that

$$b_j = \frac{\mu(j+1)}{\mu(j)} (j+1), \tag{1.3}$$

for $j \leq N - 1$. Here and throughout, if $N = \infty$ then by $j \leq N - 1$ we mean $j = 0, 1, \ldots$. The choice of these rates corresponds to the case where the *detailed balance condition*

$$\mu(j) b_j = \mu(j+1) d_{j+1}, \quad j = 0, 1, \ldots N-1, \tag{1.4}$$

holds, see for example, [1] and [12]. We will apply these well known facts to the discrete Gibbs measure introduced in the next subsection.

Now might be a good time to note that not all probability distributions on $\mathbb{N}_0$ are given in closed form expressions; notable exceptions occur for example in compound Poisson distributions.

1.2. **Gibbs measures as stationary distributions of birth-death processes.** Gibbs measures can be viewed as stationary measures of birth-death processes, as follows.

We start with a discrete Gibbs measure μ; for convenience we assume that μ has support $\mathrm{supp}(\mu) = \{0, \ldots, N\}$, where $N \in \mathbb{N}_0 \cup \{\infty\}$, so that μ is given by

$$\mu(k) = \frac{1}{\mathcal{Z}} e^{V(k)} \frac{\omega^k}{k!}, \quad k = 0, 1, \ldots, N, \tag{1.5}$$

for some function $V : \mathbb{N}_0 \to \mathbb{R}$. Here $\mathcal{Z} = \sum_{k=0}^N \exp(V(k)) \frac{\omega^k}{k!}$, and $\omega > 0$ is fixed. We assume that $\mathcal{Z}$ exists; $\mathcal{Z}$ is known as the partition function in models of statistical mechanics. We set $V(k) = -\infty$ for $k > N$. In terms of statistical mechanics, μ is a grand-canonical ensemble, ω is the activity and V is the potential energy, see [16, Chapter 1.2].

The class of discrete Gibbs-measures in (1.5) is equivalent to the class of all discrete probability distributions on $\mathbb{N}_0$ by the following simple identification: For a given probability distribution $(\mu(k))_{k \in \mathbb{N}_0}$ we have

$$V(k) = \log \mu(k) + \log k! + \log \mathcal{Z} - k \log \omega, \quad k = 0, 1, \ldots, N, \tag{1.6}$$

with $V(0) = \log \mu(0) + \log \mathcal{Z}$. Hence $\mathcal{Z} = \frac{e^{V(0)}}{\mu(0)}$. The latter formula gives the possibility of proving convergence for a sequence of partition functions $\mathcal{Z}_n$ by using the convergence of the corresponding sequence $\mu_n(0)$.

However, the representation of a probability measure as a Gibbs measure is not unique. For example, the Poisson distribution with parameter λ, $Po(\lambda)$, can be written in the form (1.5) with $\omega = \lambda$, $V(k) = -\lambda$, $k \geq 0$, $\mathcal{Z} = 1$. Alternatively, we could have chosen $V(k) = 0$, $\omega = \lambda$, $\mathcal{Z} = e^{\lambda}$.

From (1.3), if we choose a *unit per-capita death rate* $d_k = k$, and if we choose the birth rate

$$b_k = \omega e^{V(k+1)-V(k)} = (k+1) \frac{\mu(k+1)}{\mu(k)}, \tag{1.7}$$

for $k, k+1 \in \mathrm{supp}(\mu)$, then

$$(\mathcal{A}h)(k) = (h(k+1) - h(k)) \omega e^{V(k+1)-V(k)} + k(h(k-1) - h(k)), \tag{1.8}$$

for $k, k+1, k-1 \in \mathrm{supp}(\mu)$, $k \in \mathbb{N}$ (set $h(-1) = 0$), is the generator of the time-reversible birth-death process with invariant measure μ.

Note that this choice of d_k and b_k ensures that the detailed balance condition (1.4) is satisfied. Hence we have chosen a birth-death process with generator of the type (1.1) which is easily seen to be an ergodic process. Namely, if $N = \infty$ then the recurrence of the corresponding process is given since

$$\sum_{n \geq 1} \frac{d_1 \cdots d_n}{b_1 \cdots b_n} = \sum_{n \geq 1} \frac{\mu(1)}{(n+1)\mu(n+1)} \geq \sum_{n \geq 1} \frac{\mu(1)}{n+1} = \infty.$$

For $N < \infty$ we have that $V(k) > -\infty$ for $k \in \operatorname{supp}(\mu)$, so that $b_k > 0$ for $k = 0, \dots, N-1$, and as $V(N+1) = -\infty$, we obtain $b_N = 0$. Due to the unit capita death rate the ergodicity follows. From $0 \in \operatorname{supp}(\mu)$ we obtain $c = 1/\mu(0) < \infty$. Hence μ is indeed the unique stationary distribution of the birth-death process.

In the development of Stein's method unit per capita is a common and useful choice for the death rate, see [2]. It is worth noting that there are modifications of this choice in [5] and [11].

To compare with the approach in Barbour [2] we reformulate the generator: let $h(k+1) - h(k) =: g(k+1)$, then (1.8) yields

$$(\mathcal{A}g)(k) = g(k+1)\omega e^{V(k+1)-V(k)} - kg(k), \quad k = 0, 1, \dots, N. \tag{1.9}$$

The generalisation to case of arbitrary death rates d_k is straightforward; we omit it here to streamline the paper.

2. Stein identity for Gibbs-measures and bounds

In view of the generator approach to Stein's method, for a test function f: $\operatorname{supp}(\mu) \to \mathbb{R}$ the appropriate *Stein equation* for μ given in (1.5) is

$$(\mathcal{A}g)(j) = f(j) - \mu(f) \tag{2.1}$$

for $j \in \{0, \dots, N\}$ and $\mathcal{A}$ the generator given by (1.8). Here,

$$\mu(f) := \sum_{k=0}^{N} f(k)\mu(k)$$

is the expectation of f under μ. We are interested in indicator functions $f(j) = I_{[j\in A]}$ for some $A \subset \operatorname{supp}(\mu)$. Thus if W is a random variable on $\operatorname{supp}(\mu)$, we obtain

$$E\mathcal{A}g(W) = P(W \in A) - \mu(A). \tag{2.2}$$

The right-hand side of Equation (2.2) links in nicely with the total variation distance. Recall that for P and Q being probability distributions on $\mathbb{N}_0$, we define the *total variation distance (metric)* by

$$\begin{aligned} d_{\mathrm{TV}}(P,Q) &:= \sup_{A\subset\mathbb{N}_0} |P(A) - Q(A)| \\ &= \sup_{f\in\mathcal{B}_1} |P(f) - Q(f)| \\ &= \frac{1}{2}\sum_{k\in\mathbb{N}_0} |P(\{k\}) - Q(\{k\})|, \end{aligned}$$

where $\mathcal{B}_1$ denotes the set of measurable functions f with $0 \le f \le 1$. Hence bounding the left hand side of (2.2) uniformly in $A \subset \mathbb{N}_0$ gives a bound on the total variation distance.

In the following we give a Stein characterisation for μ given in (1.5) and a solution of the corresponding Stein equation.

Let Z be a random variable distributed according to the Gibbs measure μ defined in (1.5) and for a function $g\colon \mathbb{N}_0 \to \mathbb{R}$ assume $E|Zg(Z)| < \infty$. From Equation (1.9) we obtain the Stein characterisation for μ that if Z is distributed according to μ, given by (1.5), then for every function $g\colon \mathbb{N}_0 \to \mathbb{R}$ with $E|Z\,g(Z)| < \infty$

$$E\Big\{\omega e^{V(Z+1)-V(Z)}\, g(Z+1) - Z\, g(Z)\Big\} = 0. \tag{2.3}$$

If $f\colon \operatorname{supp}(\mu) \to \mathbb{R}$ is an arbitrary function, and μ is given by (1.5), then there exists a solution $g_{f,V}\colon \mathbb{N}_0 \to \mathbb{R}$ for (2.1) with operator as in (1.9),

$$g_{f,V}(k+1)\omega e^{V(k+1)-V(k)} - k g_{f,V}(k) = f(k) - \mu(f),\ k \le N, \tag{2.4}$$

see [5]. This solution $g_{f,V}$ is such that $g_{f,V}(0) = 0$, and for $j = 0, \ldots, N-1$ the solution $g_{f,V}$ can be represented by recursion as

$$\begin{aligned} g_{f,V}(j+1) &= \frac{j!}{\omega^{j+1}} e^{-V(j+1)} \sum_{k=0}^{j} e^{V(k)} \frac{\omega^k}{k!} \left(f(k) - \mu(f)\right) && (2.5)\\ &= -\frac{j!}{\omega^{j+1}} e^{-V(j+1)} \sum_{k=j+1}^{N} e^{V(k)} \frac{\omega^k}{k!} \left(f(k) - \mu(f)\right). && (2.6) \end{aligned}$$

We may set $g_{f,V}(N+1) = 0$.

Having a suitable Stein equation for Gibbs measures and its solution at our disposal, the next step in Stein's method is to bound the increments of the solutions; it will turn out advantageous to bound the solutions themselves as well. For any function $g : \mathbb{N}_0 \to \mathbb{R}$ we define

$$\Delta g(j) := g(j+1) - g(j).$$

In applications often only bounds on the increments are needed, hence we start with these. Uniform bounds on the increments are also called *Stein factors* or *magic factors*. Non-uniform Stein factor bounds may yield better overall bounds on distributional distances and are therefore of particular interest. Lemma 2.9 gives such a non-uniform bound. The proof is given in [5], Lemma 2.4 and Theorem 2.1. We introduce the class of functions

$$\mathcal{B} := \{f\colon \operatorname{supp}(\mu) \to [0,1]\,\} \tag{2.7}$$

and we define

$$F(k) := \sum_{i=0}^{k} \mu(i), \quad \overline{F}(k) := \sum_{i=k}^{N} \mu(i). \tag{2.8}$$

Lemma 2.9 (Non-uniform bounds for increments)**.** *Assume that the death rates are unit per capita and assume that the birth rates in* (1.7) *fulfill, for each* $k = 1, 2, \ldots, N-1$,

$$k \, \frac{F(k)}{F(k-1)} \geq \omega \, e^{V(k+1)-V(k)} \geq k \, \frac{\overline{F}(k+1)}{\overline{F}(k)}. \tag{2.10}$$

Let $f \in \mathcal{B}$ *and let* $g_{f,V}$ *be its solution to the Stein equation* (2.4)*. Then, for every* $j \in \{0, \ldots, N\}$,

$$\sup_{f \in \mathcal{B}} |\Delta g_{f,V}(j)| = \frac{1}{\omega} e^{V(j)-V(j+1)} \overline{F}(j+1) + \frac{1}{j} F(j-1). \tag{2.11}$$

Moreover, for every $j \in \{0, \ldots, N\}$,

$$\sup_{f \in \mathcal{B}} |\Delta g_{f,V}(j)| \leq \frac{1}{j} \wedge \frac{e^{V(j)}}{\omega e^{V(j+1)}}.$$

Remark 2.12. *Condition* (2.10) *is Condition (C2) in* [5, Lemma 2.4]*. In this paper, three more conditions are formulated, which are all equivalent to (C2). For example, if the death rates are unit per capita and the birth rates are non-increasing:*

$$e^{V(k+1)-V(k)} \leq e^{V(k)-V(k-1)}, \quad k = 0, 1, \ldots, N, \tag{2.13}$$

Condition (C4) in [5] *is satisfied and* (2.11) *holds.*

Remark 2.14. [11] *gives an elegant recursive proof of* (2.11) *for a choice of birth rates and death rates which make the method of exchangeable pairs work. In particular unit per-capita death rates are not used in her results.*

Under a slightly weaker condition than (2.10) it is possible to derive non-uniform bounds on the solution $g_{f,V}$ of the Stein equation itself, as follows.

Lemma 2.15. *Consider the solution* $g_{f,V}$ *of the Stein equation* (2.4)*, where* $f \in \mathcal{B}$*, given in* (2.7)*. Assume that, for each* $k = 1, 2, \ldots, N-1$,

$$\omega \, e^{V(k+1)-V(k)} \geq k \, \frac{\overline{F}(k+1)}{\overline{F}(k)} \tag{2.16}$$

is satisfied. Then we obtain for every $j \in \{1, \ldots, N\}$ *that*

$$|g_{f,V}(j)| \leq \left(\min\left\{ \ln(j), \sum_k k\mu(k) \right\} + \frac{1}{\omega} e^{V(0)-V(1)} \right) \frac{1}{\overline{F}(j)}.$$

The proof is related to ideas of [4], pp. 7-8, used in Poisson approximation.

Proof. For $j \in \{0, \ldots, N\}$ let $U_j = \{0, 1, \ldots, j\}$. We use the notation $f\mathbf{1}_A(x) = f(x)\mathbf{1}(x \in A)$. It is easy to see from (2.5) that for $f \in \mathcal{B}$ and $A \subset \{0, \ldots, N\}$, and for $j \in \{0, \ldots, N-1\}$,

$$g_{f,V}(j+1) \quad = \quad \frac{j!}{\omega^{j+1}} e^{-V(j+1)} \{ \mu(f\mathbf{1}_{U_j}) \mu(U_j^c) - \mu(f\mathbf{1}_{U_j^c}) \mu(U_j) \}$$

and hence

$$|g_{f,V}(j+1)| \quad \leq \quad \frac{j!}{\omega^{j+1}} e^{-V(j+1)} \mu(U_j)\mu(U_j^c). \tag{2.17}$$

From (2.16) we have that

$$\begin{aligned}
\frac{j!}{\omega^{j+1}} e^{-V(j+1)} \mu(U_j) &= \frac{1}{\omega} \sum_{k=0}^{j} e^{V(k)-V(j+1)} \frac{\omega^{k-j} j!}{k!} \\
&\leq \sum_{k=1}^{j} \frac{\bar{F}(0)}{\bar{F}(j+1)} \frac{1}{k} + \frac{1}{\omega} e^{V(0)-V(1)} \frac{\bar{F}(0)}{\bar{F}(j+1)} \\
&\leq \left(\ln(j) + \frac{1}{\omega} e^{V(0)-V(1)} \right) \frac{1}{\bar{F}(j+1)}.
\end{aligned}$$

Alternatively,

$$\begin{aligned}
\frac{j!}{\omega^{j+1}} e^{-V(j+1)} \mu(U_j) &\leq \sum_{k=1}^{j} \frac{\bar{F}(k)}{\bar{F}(j+1)} \frac{1}{k} + \frac{1}{\omega} e^{V(0)-V(1)} \frac{\bar{F}(0)}{\bar{F}(j+1)} \\
&\leq \left(\sum_{k=1}^{N} \bar{F}(k) + \frac{1}{\omega} e^{V(0)-V(1)} \right) \frac{1}{\bar{F}(j+1)} \\
&= \left(\sum_{k=1}^{N} k\mu(k) + \frac{1}{\omega} e^{V(0)-V(1)} \right) \frac{1}{\bar{F}(j+1)},
\end{aligned}$$

proving the assertion. □

We now prove a crude but usable bound for the supremum norm $\| g_{f,V} \|$ which does not require (2.10) or (2.16) to be satisfied. To this purpose we introduce the quantities

$$\begin{aligned}
\lambda_1 &:= \omega \inf_{0 \leq k \leq N-1} e^{V(k+1)-V(k)} = \inf_{0 \leq k \leq N-1} b_k \text{ and} \\
\lambda_2 &:= \omega \sup_{0 \leq k \leq N-1} e^{V(k+1)-V(k)} = \sup_{0 \leq k \leq N-1} b_k.
\end{aligned} \tag{2.18}$$

Note that $\frac{\lambda_2}{\lambda_1} \geq 1$ by construction.

Lemma 2.19. *Consider the solution $g_{f,V}$ of the Stein equation* (2.4)*, where $f \in \mathcal{B}$, given in* (2.7)*. Assume that $\lambda_1 > 0$ and $\lambda_2 < \infty$. Then*

$$\| g_{f,V} \| \leq 2 + \frac{1}{2} \left(\frac{\lambda_2}{\lambda_1 + 1} \right)^{\lambda_2 - \lambda_1 - 2} \mathbf{1}(\lambda_2 - 2 \geq \lambda_1).$$

Proof. As the proof follows the ideas of [4], pp. 7-8, used in Poisson approximation, we only sketch it here. With the notation as for (2.17) we obtain the bounds

$$\begin{aligned}
|g_{f,V}(j+1)| &\leq \frac{j!}{\omega^{j+1}} e^{-V(j+1)} \mu(U_j) \\
&= \frac{1}{\omega} \sum_{k=0}^{j} e^{V(k)-V(j+1)} \frac{\omega^{k-j} j!}{k!} \\
&\leq \lambda_1^{-1} \sum_{\ell=0}^{j} \lambda_1^{-\ell} \frac{j!}{(j-\ell)!}.
\end{aligned}$$

Similarly we have

$$\begin{aligned}
|g_{f,V}(j+1)| &\leq \frac{j!}{\omega^{j+1}} e^{-V(j+1)} \mu(U_j^c) \\
&= \frac{1}{\omega} \sum_{k=j+1}^{N} e^{V(k)-V(j+1)} \frac{\omega^{k-j} j!}{k!} \\
&\leq \sum_{k=j+1}^{N} \lambda_2^{k-j-1} \frac{j!}{k!} \\
&= j!\, e^{\lambda_2} \lambda_2^{-j-1} Po(\lambda_2)(U_j^c).
\end{aligned}$$

This puts us in the situation of (1.20) and (1.21) in Chapter 1 of [4]. We obtain that for $j < \lambda_1$

$$|g_{f,V}(j+1)| \leq 2 \min\left(1, \lambda_1^{-\frac{1}{2}}\right) \leq 2,$$

and for $j > \lambda_2 - 2$

$$|g_{f,V}(j+1)| \leq \frac{j+2}{(j+1)(j+2-\lambda_2)} \leq \frac{5}{4} < 2.$$

Thus we have proved the assertion for $j < \lambda_1$ and for $j > \lambda_2 - 2$. If $\lambda_1 > \lambda_2 - 2$ then our bound covers the whole domain.

Now assume that $S = \{\lfloor \lambda_1 \rfloor + 1, \ldots, \lfloor \lambda_2 \rfloor - 2\}$ is non-empty. For $j \in S$ we have

$$\begin{aligned}
|g_{f,V}(j+1)| &\leq \frac{j!}{\omega^{j+1}} e^{-V(j+1)} (\mu(U_j^c \cap S) + \mu(U_j^c \setminus S)) \mu(U_j) \\
&\leq 2 + \frac{j!}{\omega^{j+1}} e^{-V(j+1)} \mu(U_j^c \cap S) \\
&\leq 2 + j! \sum_{k=j+1}^{\lfloor \lambda_2 \rfloor - 2} \lambda_2^{k-j-1} \frac{1}{k!} \\
&\leq 2 + j! \lambda_2^{-j-1} e^{\lambda_2} Po(\lambda_2)\{0, \ldots, \lfloor \lambda_2 \rfloor - 2\}.
\end{aligned}$$

From [4], Proposition A.2.3 (iii) , p.259, the Poisson probabilites can be bounded as

$$Po(\lambda_2)\{0, \ldots, \lfloor \lambda_2 \rfloor - 2\} \leq \frac{\lambda_2}{\lambda_2 + 2 - \lfloor \lambda_2 \rfloor} Po(\lambda_2)(\lfloor \lambda_2 \rfloor - 2)$$

and so, as $\lfloor \lambda_2 \rfloor - 1 - j \geq 1$ for $j \leq \lfloor \lambda_2 \rfloor - 2$,

$$
\begin{aligned}
|g_{f,V}(j+1)| &\leq 2 + \frac{1}{2} \frac{j!}{(\lfloor \lambda_2 \rfloor - 2)!} \lambda_2^{\lfloor \lambda_2 \rfloor - 2 - j} \\
&\leq 2 + \frac{1}{2} (j+1)^{-(\lfloor \lambda_2 \rfloor - 2 - j)} \lambda_2^{\lfloor \lambda_2 \rfloor - 2 - j} \\
&\leq 2 + \frac{1}{2} \left(\frac{\lambda_2}{\lambda_1 + 1} \right)^{\lfloor \lambda_2 \rfloor - 2 - j} \\
&\leq 2 + \frac{1}{2} \left(\frac{\lambda_2}{\lambda_1 + 1} \right)^{\lambda_2 - \lambda_1 - 2}.
\end{aligned}
$$

This finishes the proof. □

Remark 2.20. *As λ_1 and λ_2 stay invariant under the reparametrisation $\omega \to \tilde{\omega} = \alpha\omega$, we argue that these are reasonable quantities to employ.*

Remark 2.21. *While the examples below will show that the bound in Lemma 2.19 may not be informative, in particular examples better bounds may be obtainable in a straightforward manner. Lemma 2.19 is nevertheless useful as it gives conditions on the birth rates so that $\| g_{f,V} \|$ is bounded, and these conditions do not involve monotonicity of the birth rates.*

Remark 2.22. *Note that neither the last bound in Lemma 2.9 nor the bound in Lemma 2.19 use the normalizing constant $\mathcal{Z}$ explicitly.*

Again, the generalisation of the bounds to the case of arbitrary death rates d_k would be straightforward.

A complication arises when we compare two distributions with non-identical supports. Therefore it will turn out to be useful to consider the following extension from finite to infinite support for a generator $\mathcal{A}$. For convenience assume that the corresponding measure μ has $\text{supp}(\mu) = \{0, 1, \ldots, n\}$, for some finite n, so that $\mathcal{A}$ is only defined for functions with support $\{0, 1, \ldots, n\}$ (recall that we set $g(n+1) = 0$). We extend $\mathcal{A}$ to be defined for functions on $\{0, 1, 2, \ldots\}$ as follows;

$$
\begin{aligned}
(\tilde{\mathcal{A}}g)(k) &:= g(k+1)\omega e^{V(k+1)-V(k)} - kg(k), \quad k = 0, 1, \ldots, n \\
(\tilde{\mathcal{A}}g)(k) &:= -kg(k), \quad k \geq n+1.
\end{aligned}
\tag{2.23}
$$

Thus when there are more than n individuals the process is pure death. Now if $X \sim \mu$, then still $E\tilde{\mathcal{A}}f(X) = 0$ for all functions $f \in \mathcal{D}(\tilde{\mathcal{A}})$, the domain of $\tilde{\mathcal{A}}$, since the operator $\tilde{\mathcal{A}}$ represents a birth-death process with the same invariant distribution as $\mathcal{A}$, see (1.2). Next we extend the solution of the Stein equation, so that for $f : \{0, 1, \ldots, n\} \to \mathbb{R}$, the solution $g_{f,V}$ is defined by Equation (2.5) for $k \in \{0, 1, \ldots, n\}$, and by

$$
g_{f,V}(k) := \frac{\mu(f)}{k}, \quad k \geq n+1. \tag{2.24}
$$

(Note that our formula (2.5) would have yielded $g_{f,V}(n+1) = 0$.) For a related suggestion see [4], Chapter 9.2. The above definition ensures that the Stein equation (2.4) is still satisfied. However, the bounds on the solution of Equation (2.4) change slightly. In contrast to (2.7), let

$$\mathcal{B}_0 := \{f : \mathbb{N}_0 \to [0,1], f(x) = 0 \text{ for } x \notin \operatorname{supp}(\mu)\}. \tag{2.25}$$

Lemma 2.26. *Let $\tilde{A}$, associated with μ, V, and ω be given as in* (2.23)*, where μ has supp$(\mu) = \{0, 1, \dots, n\}$. Let $f \in \mathcal{B}_0$ and let $g_{f,V}$ be the solution of the Stein equation* (2.4) *given in* (2.5) *for $k \le n$, and as in* (2.24) *for $k \ge n+1$. Defining the sum $\sum_{l=j+1}^n$ as* 0 *for $j \ge n$, with λ_1 and λ_2 being given in* (2.18) *for μ,*

$$\begin{aligned} |g_{f,V}(j)| &\le 2 + \frac{1}{2}\left(\frac{\lambda_2}{\lambda_1+1}\right)^{\lambda_2-\lambda_1-2} \mathbf{1}(\lambda_2 - 2 \ge \lambda_1), \quad j = 0, 1, \dots, n; \\ |g_{f,V}(j)| &\le \frac{1}{j}, \quad j \ge n+1. \end{aligned}$$

In the case that (2.16) *is satisfied we also have*

$$|g_{f,V}(j)| \le \left(\min\left\{\ln(j), \sum_k k\mu(k)\right\} + \frac{1}{\omega}e^{V(0)-V(1)}\right)\frac{1}{\bar{F}(j)}, \quad j = 0, 1, \dots, n,$$

and if (2.10) *is satisfied, then*

$$\sup_{f\in\mathcal{B}_0} |\Delta g_{f,V}(j)| \le \begin{cases} \frac{1}{j} \wedge \frac{e^{V(j)}}{\omega e^{V(j+1)}}, & j \le n, \\ \frac{1}{j}, & j > n. \end{cases}.$$

Proof. The first assertion follows directly from Lemma 2.19 and (2.24). The second assertion follows from Lemma 2.9 for $j \le n-1$, and for $j \ge n+1$ it follows from (2.24) that

$$g_{f,V}(j+1) - g_{f,V}(j) = \mu(f)\left\{\frac{1}{j+1} - \frac{1}{j}\right\},$$

so that

$$|g_{f,V}(j+1) - g_{f,V}(j)| < \frac{1}{j}.$$

For $j = n$ we obtain

$$\begin{aligned} g_{f,V}(n+1) - g_{f,V}(n) &= \mu(f)\frac{1}{n+1} - \left\{-\frac{1}{n}(f(n) - \mu(f))\right\} \\ &= \frac{1}{n}\left(f(n) - \mu(f)\frac{1}{n+1}\right), \end{aligned}$$

so that

$$|g_{f,V}(n+1) - g_{f,V}(n)| < \frac{1}{n}.$$

□

We conclude this section with some examples.

Example 2.27. Poisson-distribution *$Po(\lambda)$ with parameter $\lambda > 0$: We use $\omega = \lambda, V(k) = -\lambda, \mathcal{Z} = 1$. The Stein-operator is*

$$(\mathcal{A}g)(k) = g(k+1)\,\lambda - kg(k).$$

Applying Lemma 2.9 ($b_k = \lambda$ is non-increasing in k) gives the non-uniform bound

$$\sup_{f\in\mathcal{B}} |\Delta g_{f,\mathrm{pos}}(k)| \le \frac{1}{k} \wedge \frac{1}{\lambda}, \tag{2.28}$$

which leads to the well known uniform bound $1 \wedge 1/\lambda$, see [4, Lemma 1.1.1]. *Yet* [4, Lemma 1.1.1] *give the better bound $\lambda^{-1}\big(1-e^{-\lambda}\big) \le \min\big(1, 1/\lambda\big)$. In the Poisson case the right hand side of* (2.11) *gives*

$$1/k \sum_{l=0}^{k-1} e^{-\lambda}\frac{\lambda^l}{l!} + 1/\lambda \sum_{l=k+1}^{\infty} e^{-\lambda}\frac{\lambda^l}{l!}.$$

This sum can be bounded by

$$e^{-\lambda}\left(\frac{1}{\lambda}\sum_{l=1}^{\infty}\frac{\lambda^l}{l!}\right) = \frac{e^{-\lambda}}{\lambda}\big(e^{\lambda}-1\big) = \frac{1}{\lambda}\big(1-e^{-\lambda}\big),$$

and therefore we obtain for any $k \ge 1$

$$\sup_{f\in\mathcal{B}} |\Delta g_{f,\mathrm{pos}}(k)| \le \frac{1}{\lambda}\big(1-e^{-\lambda}\big) \tag{2.29}$$

as in [4, Lemma 1.1.1]. *The bounds* (2.28) *and* (2.29) *lead to*

$$\sup_{f\in\mathcal{B}} |\Delta g_{f,\mathrm{pos}}(k)| \le \frac{1}{k} \wedge \frac{1}{\lambda}\big(1-e^{-\lambda}\big), \tag{2.30}$$

which is a slight improvement of (2.28) *and of* [5].

For the Poisson distribution, $\lambda_1 = \lambda_2 = \lambda$ in Lemma 2.19, and the bound 2 from Lemma 2.19 is poor compared to the bound $\| g \| \le \min(1, \lambda^{-\frac{1}{2}})$ from Lemma 1.1.1. on p.7 in [4]. *The bound in Lemma 2.15 is cumbersome to compute; but for $\lambda \ge \sqrt{2}$ and for $j = 1$, say, the non-uniform bound $|g(1)| \le \big(\lambda(1-e^{-\lambda})\big)^{-1}$ is slightly more informative than the uniform bound $\| g \| \le \lambda^{-\frac{1}{2}}$. The non-uniform bound improves with increasing λ, and deteriorates with increasing j.*

Example 2.31. Binomial distribution *with parameters n and $0 < p < 1$: We use $\omega = \frac{p}{1-p}, V(k) = -\log((n-k)!)$, and $\mathcal{Z} = (n!(1-p)^n)^{-1}$. The Stein-operator is*

$$(\mathcal{A}g)(k) = g(k+1)\,\frac{p(n-k)}{(1-p)} - kg(k).$$

In [4] *and* [7] *we find $(\mathcal{A}g)(k) = g(k+1)\,p(n-k) - (1-p)kg(k)$ as the operator which is equivalent to our formulation. The birth rates $(b_k)_k$ are non-increasing and we obtain from Lemma 2.9 that*

$$\sup_{f\in\mathcal{B}} |\Delta g_{f,\mathrm{bin}}(k)| \le \frac{1}{(1-p)k} \wedge \frac{1}{p(n-k)}.$$

The proof of [4, Lemma 9.2.1] *implicitly contains the same non-uniform bound. Formula (18) in* [7] *gives a bound of the form*

$$\sup_{f\in\mathcal{B}} |\Delta g_{f,\mathrm{bin}}(k)| \le \left(1 - p^{n+1} - (1-p)^{n+1}\right)/((n+1)\,(1-p)\,p)$$

for every $0 < k < n$*. Simple calculations show that the bound (18) in* [7] *is for some cases better and for some cases worse in comparison to our non-uniform result. From Lemma 2.19 with* $\lambda_1 = 1, \lambda_2 = n$ *we obtain a bound on* $\| \, g_{f,\mathrm{bin}} \, \|$ *which is of order* $O\left(e^{n\ln n}\right)$ *and therefore in most applications not useful. The non-uniform bound Lemma 2.15*

$$|g(j)| \le \left(1 - \sum_{k=0}^{j-1} \binom{n}{k} p^k (1-p)^{n-k}\right)^{-1} \left(\min\{\ln(j), np\} + \frac{1-p}{np}\right)$$

will still be informative for j small and np large.

Example 2.32. Geometric distribution*: Consider* $\mu(k) = p(1-p)^k$ *for* $k = 0, 1, \ldots$*. The Stein-operator is*

$$(\mathcal{A}g)(k) = g(k+1)\,(1-p)(k+1) - kg(k).$$

The birth rates $b_k = (1-p)(k+1)$ *fulfill* $b_k - b_{k-1} \le k - (k-1) = d_k - d_{k-1}$*. This is condition (C4) in* [5]*, which is sufficient for* (2.10)*. Hence applying* [5, Theorem 2.10]*, one obtains*

$$\sup_{f\in\mathcal{B}} |\Delta g_{f,\mathrm{geo}}(k)| \le \frac{1}{k} \wedge \frac{1}{(1-p)(k+1)},$$

which leads to the uniform bound $1 \wedge \frac{1}{(1-p)}$*.*

We can improve their bounds calculating the right hand side of (2.11) *explicitly:*

$$\begin{aligned} |\Delta g_{f,V}(j)| &= \frac{p}{(1-p)(j+1)}\left(\frac{(1-p)^{j+1}}{p}\right) + \frac{1}{j}\left(1-(1-p)^j\right) \\ &= \frac{(1-p)^j}{j+1} + \frac{1}{j}\left(1-(1-p)^j\right) = \frac{j+1-(1-p)^j}{j(j+1)}. \end{aligned}$$

Obviously $|\Delta g_{f,V}(j)| \le \frac{1}{j}$*. Using Bernoulli's inequality* $(1-x)^n \ge 1 - n\,x$ *for* $x < 1$ *and* $n \in \mathbb{N}$ *it follows that*

$$|\Delta g_{f,V}(j)| \le \frac{j+pj}{j(j+1)} = \frac{1+p}{j+1}.$$

We obtain

$$\sup_{f\in\mathcal{B}} |\Delta g_{f,\mathrm{geo}}(k)| \le \frac{1}{k} \wedge \frac{1+p}{(k+1)},$$

which leads to the uniform bound $1 \wedge (1+p)$*.*

In Lemma 2.19 we obtain $\lambda_1 = 1-p$*,* $\lambda_2 = \infty$*, and hence the lemma does not give informative bounds. Lemma 2.15 gives, with* $\overline{F}(j) = (1-p)^j$*, the bound*

$|g_{f,\mathrm{geo}}(j)| \le \frac{1}{(1-p)^j}\left(\min\left\{\ln(j), \frac{1}{p}\right\} + \frac{1}{1-p}\right)$. *However, with* (2.6) *and* $f \in \mathcal{B}$ *we may bound directly*

$$\begin{aligned} |g_{f,\mathrm{geo}}(j+1)| &\le \frac{j!}{(1-p)^{j+1}(j+1)!}\sum_{k=j+1}^{\infty}(1-p)^k \\ &= \frac{1}{j+1}\sum_{k=0}^{\infty}(1-p)^k = \frac{1}{p\,(j+1)}, \end{aligned}$$

and using that $g_{f,\mathrm{geo}}(0) = 0$ *we obtain the improved bound* $\| g_{f,\mathrm{geo}} \| \le \frac{1}{p}$. *To our knowledge this bound is new.*

In [13] *the author considered another Stein-operator. Hence the results cannot be compared. In* [14] *the authors considered the same Stein-operator. They obtain* $\sup_{f\in\mathcal{B}} |\Delta g_{f,\mathrm{geo}}(k)| \le \frac{1}{k}$. *Hence* [5] *already improved this result.*

3. Size-Bias couplings and Gibbs measures

3.1. **Size-bias couplings and Stein characterisations.** Stein's method is most powerful in presence of dependence. To treat global dependence, couplings have proved a useful tool. For discrete, non-negative random variables, so-called *size bias* couplings fit nicely into our framework as they link in with unit per-capita death rate generators.

For any random variable $X \ge 0$ with $EX > 0$ we say that a random variable X^*, defined on the same probability space as X, has the X-*size biased* distribution if

$$EXf(X) = EXEf(X^*) \tag{3.1}$$

for all functions f such that both sides of (3.1) exists. Size-bias couplings (X, X^*) have been studied in connection with Stein's method for Poisson approximation, see [4], e.g., and for normal approximations, see [9], e.g.; a general framework is give in [8],

If X is discrete then for all x we have $P(X^* = x) = \frac{x}{EX}P(X = x)$. This illustrates that size biasing corresponds to sampling proportional to size; the larger a subpopulation, the more likely it is to be contained in the sample.

Example 3.2. Poisson distribution. *If* $X \sim Po(\lambda)$*, then from the Stein-operator in Example 2.27 we read off that* $X^* = X+1$ *has the* $Po(\lambda)$*-size biased distribution. In* [4] *a related coupling is used, namely the reduced size-biased coupling* (X, X_*)*, with* $X_* = X^* - 1$.

Example 3.3. Bernoulli distribution. *If* $X \sim Be(p)$*, it is easy to see that* $X^* = 1$ *has the* $Be(p)$*-size biased distribution, for all* $p \in (0, 1]$*. As an aside, this shows that* X^* *does not uniquely define* X.

Example 3.4. Sum of nonnegative random variables. *Let $X_1, \dots, X_n$ be nonnegative, $EX_i = \mu_i > 0$, $W = \sum_{i=1}^n X_i$, $EW = \mu$, $\mathrm{Var}(W) = 1$. Goldstein and Rinott* [9] *give the following construction, valid also in the presence of dependence. Choose an index I from $\{1, \dots, n\}$ according to $P(I = i) = \frac{\mu_i}{\mu}$. If $I = i$, replace X_i by a variate X_i^* having the X_i-size bias distribution. If $X_i^* = x$, construct $\hat{X}_j$, $j \neq i$, such that*

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

Then

$$W^* = \sum_{j \neq I} \hat{X}_j + X_I^* \tag{3.5}$$

has the W-size biased distribution; see [9]*, construction after Lemma 2.1.*

As in the Poisson case, the size bias coupling can be used to derive a characterisation of a Gibbs measure, as follows.

Lemma 3.6. *Let $X \geq 0$ be such that $0 < E(X) < \infty$, and let μ be a discrete Gibbs measure given in* (1.5)*. If $X \sim \mu$ and X^* having the X-size biased distribution is defined on the same probability space as X, then for every function $g \colon \mathbb{N}_0 \to \mathbb{R}$ such that $E|X\, g(X)|$ exists,*

$$\omega\, E e^{V(X+1)-V(X)} g(X+1) \quad = \quad \omega\, E e^{V(X+1)-V(X)} E g(X^*). \tag{3.7}$$

Proof. In view of (2.3), for any $X \geq 0$ with $EX > 0$ and (X, X^*) a size-biased coupling we have, for any bounded function $g \colon \mathbb{N}_0 \to \mathbb{R}$

$$\begin{aligned} & E\Big\{\omega\, e^{V(X+1)-V(X)}\, g(X+1) - X\, g(X)\Big\} \\ = \quad & E\Big\{\omega\, e^{V(X+1)-V(X)}\, g(X+1) - EX E g(X^*)\Big\}. \end{aligned}$$

For X having distribution (1.5), the expectation is given by

$$EX = \omega E e^{V(X+1)-V(X)}.$$

From $EX > 0$ it follows that $\omega \neq 0$. The result now follows from the Stein characterisation (2.3) of (1.5) for $\omega \neq 0$. $\square$

Indeed in (3.7) the factor ω cancels. However, in a moment we shall relate (3.7) to the Stein equation. In order not to get confused about the solutions for the Stein equation with and without the ω involved, we have decided to keep the ω.

Lemma 3.6 provides a new formulation of the Stein approach (2.2): Let $W \geq 0$ with $0 < EW < \infty$. If W has distribution μ, then for all $f \in \mathcal{B}$

$$\begin{aligned} & Ef(W) - \mu(f) \\ = \quad & \omega\{E e^{V(W+1)-V(W)} g_{f,V}(W+1) - E e^{V(W+1)-V(W)} E g_{f,V}(W^*)\}, \end{aligned} \tag{3.8}$$

where $g_{f,V}$, given in (2.5), is the solution of the Stein equation (2.4).

3.2. **Comparison of distributions via their generators.** From Equation (2.2), for any random variable W and any measurable set A we can find a function $f = f_\mu$ such that $P(W \in A) - \mu(A) = E\mathcal{A}f(W)$, where $\mathcal{A}$ is the generator associated with the target distribution μ. Applications of Stein's method usually continue by bounding the right-hand side $E\mathcal{A}f(W)$. Nevertheless, Stein's method can also be employed to compare two distributions by comparing their generators. While generators may in general not be available, for any discrete distribution we have implicitly constructed a generator in (1.9). It is from this angle that we shall bound distances between Gibbs distributions. Assume that μ_1 and μ_2 are two distributions with generators $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively. Then, for $W \sim \mu_2$, $E\mathcal{A}_2 f_{\mu_1}(W) = 0$, and therefore

$$\mu_1(A) - P(W \in A) = E\mathcal{A}_1 f_{\mu_1}(W) - E\mathcal{A}_2 f_{\mu_1}(W) = E(\mathcal{A}_1 - \mathcal{A}_2) f_{\mu_1}(W).$$

Hence we can compare two discrete Gibbs distributions by comparing their birth rates and their death rates, as follows.

Theorem 3.9. *Let μ_1 have generator $\mathcal{A}_1$ as in* (1.9) *and corresponding (ω_1, V_1), and let μ_2 have generator $\mathcal{A}_2$, and corresponding (ω_2, V_2), both described in terms of unit per-capita death rates. Suppose that $\mathcal{D}(\mathcal{A}_1) = \mathcal{D}(\mathcal{A}_2)$. Then, for $X_2 \sim \mu_2$, $f \in \mathcal{B}$, if g_{f,V_1} is the solution of the Stein equation for μ_1,*

$$\begin{aligned} &\left| Ef(X_2) - \int f d\mu_1 \right| \le \\ &\quad \min\Bigg\{ \| g_{f,V_1} \| \, E(X_2) \left(\frac{|\omega_1 - \omega_2|}{\omega_2} \right. \\ &\quad \left. + \frac{\omega_1}{\omega_2} E\left| e^{(V_1(X_2^*) - V_1(X_2^*-1)) - (V_2(X_2^*) - V_2(X_2^*-1))} - 1 \right| \right), \\ &\quad \| g_{f,V_2} \| \, E(X_1) \left(\frac{|\omega_2 - \omega_1|}{\omega_1} \right. \\ &\quad \left. + \frac{\omega_2}{\omega_1} E\left| e^{(V_2(X_1^*) - V_2(X_1^*-1)) - (V_1(X_1^*) - V_1(X_1^*-1))} - 1 \right| \right) \Bigg\}. \end{aligned} \tag{3.10}$$

Proof. For $X_2 \sim \mu_2$, $f \in \mathcal{B}$, if g_{f,V_1} is the solution of the Stein equation for μ_1, we have $Ef(X_2) - \mu_1(f) = E\mathcal{A}_1 g_{f,V_1}(X_2) = E(\mathcal{A}_1 - \mathcal{A}_2) g_{f,V_1}(X_2)$. Using the size biasing and (3.7) we obtain

$$\begin{aligned}
& Ef(X_2) - \mu_1(f) \\
= \;& Eg_{f,V_1}(X_2+1)\left(\omega_1 e^{V_1(X_2+1)-V_1(X_2)} - \omega_2 e^{V_2(X_2+1)-V_2(X_2)}\right) \\
= \;& \omega_1 Eg_{f,V_1}(X_2+1)e^{V_2(X_2+1)-V_2(X_2)}e^{V_1(X_2+1)-V_1(X_2)-(V_2(X_2+1)-V_2(X_2))} \\
& -E(X_2)Eg_{f,V_1}(X_2^*) \\
= \;& \frac{\omega_1}{\omega_2}E(X_2)Eg_{f,V_1}(X_2^*)e^{(V_1(X_2^*)-V_1(X_2^*-1))-(V_2(X_2^*)-V_2(X_2^*-1))} \\
& -E(X_2)Eg_{f,V_1}(X_2^*) \\
= \;& \frac{\omega_1-\omega_2}{\omega_2}E(X_2)Eg_{f,V_1}(X_2^*) \\
& +\frac{\omega_1}{\omega_2}E(X_2)Eg_{f,V_1}(X_2^*)\left\{e^{(V_1(X_2^*)-V_1(X_2^*-1))-(V_2(X_2^*)-V_2(X_2^*-1))} - 1\right\}.
\end{aligned}$$

Taking absolute values together with the triangle inequality, and observing that the argument is symmetric in the indices 1 and 2, we obtain (3.10). □

Remark 3.11. *Theorem 3.9 has the advantage that the partition function $\mathcal{Z}$ may not be needed to assess the distance.*

Remark 3.12. *Note that it is not surprising that $\| g_{f,V_1} \|$ appears in the bound - even if we compare two Poisson distributions, with different means λ_1 and λ_2, the bound would be of the type $|\lambda_1 - \lambda_2| \, \| g_{f,V_1} \|$.*

When $\mathcal{D}(\mathcal{A}_1) \neq \mathcal{D}(\mathcal{A}_2)$, we first extend $\mathcal{A}_1$ to $\tilde{\mathcal{A}}_1$, defined for functions on $\{0,1,2,\ldots\}$ as in (2.23), and apply then Lemma 2.26. For convenience assume that $\mathrm{supp}(\mu_1) = \{0,1,\ldots,n\}$ and $\mathrm{supp}(\mu_2) = \{0,1,2,\ldots\}$. Using similar calculations as for Theorem 3.9, we derive the following result.

Corollary 3.13. *Let μ_1 have generator $\mathcal{A}_1$ and corresponding (ω_1, V_1), and let μ_2 have generator $\mathcal{A}_2$, and corresponding (ω_2, V_2), both described in terms of unit per-capita death rates. Suppose that* $\mathrm{supp}(\mu_1) = \{0,1,\ldots,n\}$ *and that* $\mathrm{supp}(\mu_2) = \{0,1,2,\ldots\}$*. Then, for $X_2 \sim \mu_2$ and $f \in \mathcal{B}_0$ as in* (2.25)*, if g_{f,V_1} is the solution of* *the Stein equation for* μ_1 *as in* (2.23),

$$\begin{aligned}
&\left|Ef(X_2) - \int f d\mu_1\right| &(3.14)\\
\leq\ & \min\Bigg\{\| g_{f,V_1} \| \, E(X_2)\left(\frac{|\omega_1-\omega_2|}{\omega_2}\right.\\
&\left.+\frac{\omega_1}{\omega_2}E\left|e^{(V_1(X_2^*)-V_1(X_2^*-1))-(V_2(X_2^*)-V_2(X_2^*-1))}-1\right|\right),\\
&\| g_{f,V_2} \| \, E(X_1)\left(\frac{|\omega_2-\omega_1|}{\omega_1}\right.\\
&\left.+\frac{\omega_2}{\omega_1}E\left|e^{(V_2(X_1^*)-V_2(X_1^*-1))-(V_1(X_1^*)-V_1(X_1^*-1))}-1\right|\right)\Bigg\}\\
&+\mu_1(f)\sum_{k=n+1}^{\infty}\mu_2(k).
\end{aligned}$$

The extra term $\mu_1(f)\sum_{k=n+1}^{\infty}\mu_2(k)$ in the bound (3.14) as compared to (3.10) accounts for the extension $\tilde{\mathcal{A}}_1$ of the generator $\mathcal{A}_1$.

Proof. We proceed as in the proof of Theorem 3.9. Let $\tilde{\mathcal{A}}_1$ be the extension of $\mathcal{A}_1$ as in (2.23). For $X_2 \sim \mu_2$, $f \in \mathcal{B}_0$, if g_{f,V_1} as in (2.24) is the solution of the Stein equation for μ_1, we have

$$\begin{aligned}
&Ef(X_2) - \mu_1(f)\\
&= E\tilde{\mathcal{A}}_1 g_{f,V_1}(X_2)\\
&= E(\tilde{\mathcal{A}}_1 - \mathcal{A}_2) g_{f,V_1}(X_2)\\
&= Eg_{f,V_1}(X_2+1)\left(\omega_1 e^{V_1(X_2+1)-V_1(X_2)} - \omega_2 e^{V_2(X_2+1)-V_2(X_2)}\right)\\
&= \sum_{k=0}^{n-1} P(X_2=k) g_{f,V_1}(k+1)\left(\omega_1 e^{V_1(k+1)-V_1(k)} - \omega_2 e^{V_2(k+1)-V_2(k)}\right)\\
&\quad -\sum_{k=n}^{\infty} P(X_2=k)\frac{\mu_1(f)}{k+1}\omega_2 e^{V_2(k+1)-V_2(k)},
\end{aligned}$$

where we used (2.24). Now the summand

$$\sum_{k=0}^{n-1} P(X_2=k) g_{f,V_1}(k+1)\left(\omega_1 e^{V_1(k+1)-V_1(k)} - \omega_2 e^{V_2(k+1)-V_2(k)}\right)$$

can be treated exactly as in Theorem 3.9; from Lemma 2.26 we have the same bounds on the solution of the Stein equation when $k \leq n$. For the second summand we have

$$\begin{aligned}
&\sum_{k=n}^{\infty} P(X_2 = k)\frac{\mu_1(f)}{k+1}\omega_2 e^{V_2(k+1)-V_2(k)} \\
&= \omega_2\mu_1(f)\sum_{k=n}^{\infty}\frac{1}{\mathcal{Z}}\, e^{V_2(k)}\frac{\omega_2^k}{k!(k+1)}e^{V_2(k+1)-V_2(k)} \\
&= \mu_1(f)\sum_{k=n+1}^{\infty}\mu_2(k).
\end{aligned}$$

Thus we obtain (3.14). □

Note that the bound in Corollary 3.13 contains the measure μ_2 explicitly; hence examples will require the calculation of the normalising constant $\mathcal{Z}_2$. We now treat the instructive example that the two generators $\mathcal{A}_1$ and $\mathcal{A}_2$ have the same birth and death rates, but live on different domains.

Example 3.15. *Suppose that the distributions μ_1 and μ_2 have supports* $\mathrm{supp}(\mu_1) = \{0, 1, \ldots, n\}$ *and* $\mathrm{supp}(\mu_2) = \{0, 1, 2, \ldots\}$*, respectively. Let μ_1 have generator $\mathcal{A}_1$ and corresponding (ω, V), and let μ_2 have generator $\mathcal{A}_2$, and corresponding same (ω, V), both described in terms of unit per-capita death rates, so that the two generators have the same birth rates and death rates for $k = 0, \ldots, n$. Then, for $X_2 \sim \mu_2$ and $f \in \mathcal{B}_0$, if g_{f,V_1} is the solution of the Stein equation for μ_1, it follows from* (3.14) *that*

$$\left| Ef(X_2) - \int f d\mu_1 \right| \leq \mu_1(f) \sum_{k=n+1}^{\infty} \mu_2(k).$$

In particular it follows that

$$d_{TV}(\mu_1, \mu_2) \leq \sum_{k=n+1}^{\infty} \mu_2(k),$$

stating that the difference between the two distributions can be bounded by the total mass of the domain, under the respective distribution, which is in the support of the one distribution, but not of the other distribution.

Indeed the bound is sharp, as from (1.2) *we have that $\mu_1(k) = \alpha\mu_2(k)$ for $k = 0, 1, \ldots, n$, with $\alpha = (\sum_{k=0}^{n}\mu_2(k))^{-1} > 1$. Hence*

$$d_{TV}(\mu_1, \mu_2) = \frac{1}{2}\sum_{k=0}^{n}(\alpha - 1)\mu_2(k) + \frac{1}{2}\sum_{k=n+1}^{\infty}\mu_2(k) = \sum_{k=n+1}^{\infty}\mu_2(k).$$

Theorem 3.9 and Corollary 3.13 invite to the study of the total variation distance of discrete probability distributions with different supports of the corresponding operators, for example considering the distance between a binomial and a hypergeometric distribution. As this leads away from the flow of the paper, we do not pursue it here. Instead we turn to the application of our approach to Gibbs measures arising in interacting particle systems.

4. Application to lattice approximation in statistical physics

Now we apply our results to the model studied by Chayes and Klein [6]. Their setup is as follows. Assume $A \subset \mathbb{R}^d$ is a rectangle; denote its volume by $|A|$. Consider the intersection of A with the d-dimensional lattice $n^{-1}\mathbb{Z}^d$. For each site m in this intersection we associate a Bernoulli random variable X_m^n which takes value 1, with probability p_m^n, if a particle is present at m and 0 otherwise. If the collection $(X_m^n)_m$ is independent, the joint distribution can be interpreted as the Gibbs distribution for an ideal gas on the lattice. The Poisson convergence theorem states that, for n going to infinity, when preserving the average density of particles the lattice ideal gas distribution converge weakly to a Poisson distribution, which is the standard Gibbs distribution for an ideal gas in the continuum. On physical grounds one expects that a similar result might hold for interacting particles. The model is as follows. Pick $n \in \mathbb{N}$ and suppose that A can be partitioned into a regular array of $d(n)$ sub-rectangles $\{S_1^n, \dots, S_{d(n)}^n\}$ with volumes $v(S_m^n) = \frac{z_m^n}{z_n}$. Here, $z_m^n > 0$ for each m, n. Thus

$$|A| = \frac{1}{z_n} \sum_{m=1}^{d(n)} z_m^n.$$

As a guideline motivated by [6] we choose $z_m^n > 0$ and z_n such that $z_m^n \to 0$ and $z_n \to z > 0$ for $n \to \infty$. For each m, n choose a point $q_m^n \in S_m^n$. Now the following class of functions is considered.

Assumption 4.1. *Let $(f_k)_k$ be a sequence of functions satisfying $f_0 \equiv 1$ and for each $k \geq 1$, $f_k(x_1, \dots, x_k)$ is a nonnegative function, Riemann integrable on A^k, such that f_k is a symmetric function for each k, that is, for any permutation σ, $f_k(x_{\sigma(1)}, \dots, x_{\sigma(k)}) = f_k(x_1, \dots, x_k)$.*

Let $X_m^n, 1 \leq m \leq d(n)$, be 0-1 random variables with joint density function

$$P(X_1^n = a_1, \dots, X_{d(n)}^n = a_{d(n)}) \propto f_k(q_{i_1}^n, \dots, q_{i_k}^n) \prod_{m=1}^{d(n)} (z_m^n)^{a_m}$$

where each $a_i \in \{0, 1\}$ and the indices on the right-hand side are determined by

$$k = \sum_{m=1}^{d(n)} a_m.$$

Define

$$S_n = \sum_{m=1}^{d(n)} X_m^n.$$

Then, due to the symmetry of f_k,

$$P(S_n = k) = \frac{\frac{(z_n)^k}{k!} \sum_{i_1=1}^{d(n)} \cdots \sum_{i_k=1}^{d(n)} f_k(q_{i_1}^n, \ldots, q_{i_k}^n) \prod_{m=1}^k v(S_{i_m}^n)}{\sum_{k=0}^{d(n)} \frac{(z_n)^k}{k!} \sum_{i_1=1}^{d(n)} \cdots \sum_{i_k=1}^{d(n)} f_k(q_{i_1}^n, \ldots, q_{i_k}^n) \prod_{m=1}^k v(S_{i_m}^n)}. \tag{4.2}$$

Note that we can write (4.2) as a Gibbs measure μ_n of the form (1.5) with

$$\omega_n \quad = \quad z_n; \quad V_n(k) = \log\left(W_n(k)\right), \quad k \in \{0, 1, \ldots\ d(n)\}, \tag{4.3}$$

where

$$W_n(k) \quad = \quad \sum_{i_1=1}^{d(n)} \cdots \sum_{i_k=1}^{d(n)} f_k(q_{i_1}^n, \ldots, q_{i_k}^n) \prod_{m=1}^k v(S_{i_m}^n), \quad k \in \{0, \ldots, d(n)\}.$$

Let S be a nonnegative integer valued random variable defined by

$$P(S = k) \quad = \quad \frac{\frac{z^k}{k!} \int_{A^k} f_k(x_1, \ldots, x_k) dx_1 \cdots dx_k}{\sum_{k=0}^{\infty} \frac{z^k}{k!} \int_{A^k} f_k(x_1, \ldots, x_k) dx_1 \cdots dx_k}. \tag{4.4}$$

We write (4.4) as a Gibbs measure μ of the form (1.5) with

$$\omega = z, \quad V(k) = \log\left(W(k)\right), \quad k \in \{0, 1, \ldots\},$$

where

$$W(k) \quad = \quad \int_{A^k} f_k(x_1, \ldots, x_k) dx_1 \cdots dx_k, \quad k \in \{0, 1, \ldots\}.$$

In [6] the following class of functions is considered. Let $(f_k)_k$ be a sequence of functions satisfying Assumption 4.1 and, in addition,

(a) There exists a constant C such that $f_k(x_1, \ldots, x_k) \le C^k$ for all $k \ge 0$.
(b) $f_k(x_1, \ldots, x_k) = 0$ if $x_i = x_j$ for some $i \ne j$. This is not a necessary condition, see [6].

In [6] it is shown that under these additional conditions, $S_n \Rightarrow S$, that is, S_n converges weakly to S, under the conditions that $\lim_{n\to\infty} z_n = z$ and that

$$\lim_{n\to\infty} \Big(\max_{1 \le m \le d(n)} z_m^n \Big) = 0.$$

For the proof of this convergence, Condition (a) on $(f_k)_k$ enters to ensure that the Riemann sum converges, so that the normalising constants $\mathcal{Z}_n$ and $\mathcal{Z}$ for μ_n and for μ, respectively, are finite. Condition (b) avoids some measure-theoretic considerations. In [6] a bound on the rate of convergence is given for the special case that $f_k = 1$, all $k \ge 0$, resulting in a Poisson approximation.

Here we give a general bound on the distance, not only in the case $f_k = 1$, $k \ge 0$.

Note that in [6] the generalisation of Poisson convergence allows the authors to develop a lattice-to-continuum theory of classical statistical mechanics; similarly our results could be applied to obtain bounds on such lattice-to-continuum theory.

Using Corollary 3.13 with $\mu_1 = \mu_n$ and $\mu_2 = \mu$, we obtain

$$\begin{aligned}
&\left| Ef(S_n) - \int f d\mu \right| \\
&\leq \min\Bigg\{ E(S_n) \parallel g_{f,V} \parallel \left(\frac{|z_n - z|}{z_n} + \frac{z}{z_n} E \left| \frac{W(S_n^*) W_n(S_n^* - 1)}{W(S_n^* - 1) W_n(S_n^*)} - 1 \right| \right), \\
&\quad E(S) \parallel g_{f,V_n} \parallel \left(\frac{|z - z_n|}{z} + \frac{z_n}{z} E \left| \frac{W_n(S^*) W(S^* - 1)}{W_n(S^* - 1) W(S^*)} - 1 \right| \right) \Bigg\} \\
&\quad + \mu_n(f) \sum_{k=d(n)+1}^{\infty} \mu(k).
\end{aligned}$$

Writing out the size bias distribution gives the following result.

Theorem 4.5. *Let S_n, μ, z_n be as above. Let us assume that $(f_k)_k$ satisfies Assumption 4.1 and Condition (a). Then*

$$\begin{aligned}
&d_{TV}(\mathcal{L}(S_n), \mu) \\
&\leq \max(\parallel g_{f,V} \parallel, \parallel g_{f,V_n} \parallel) \min\Bigg\{ \Bigg(E(S_n) \frac{|z_n - z|}{z_n} \\
&\quad + \frac{z}{z_n} \sum_k k \mu_n(k) \left| \frac{W(k) W_n(k-1)}{W(k-1) W_n(k)} - 1 \right| \Bigg), \Bigg(E(S) \frac{|z - z_n|}{z} \\
&\quad + \frac{z_n}{z} \sum_k k \mu(k) \left| \frac{W_n(k) W(k-1)}{W_n(k-1) W(k)} - 1 \right| \Bigg) \Bigg\} + \mu_n(f) \sum_{k=d(n)+1}^{\infty} \mu(k).
\end{aligned}$$

Remark 4.6. *Note that for Theorem 4.5, no monotonicity assumptions on the birth and death rates in the corresponding birth-death process are needed. For the bound to be informative, though, we need that* $\max(\parallel g_{f,V} \parallel, \parallel g_{f,V_n} \parallel) < \infty$*; conditions to ensure this behaviour are given in Lemma 2.19.*

For the case of non-increasing birth rates and unit per-capita death rates, we also obtain bounds based on the increments Δg. Recall the notation of Example 3.4.

Proposition 4.7. *Let S_n be the sum of $X_1, \ldots, X_n$, where $X_i \sim Be(p_i)$, and construct S_n^* via* (3.5)*, and let μ be given by* (1.5)*. Assume that the birth rates in* (1.7) *are non-increasing, and assume that the death rates are unit per-capita. Then*

$$\begin{aligned}
&d_{TV}(\mathcal{L}(S_n), \mu) \\
&\leq \omega E \Big\{ e^{V(S_n+1) - V(S_n)} \min\Big\{ S(S_n, S_n^* - 1), \qquad (4.8) \\
&\quad (|S_n - S_n^* + 1|/\omega)\, e^{V(\min(S_n, S_n^* - 1)) - V(\min(S_n, S_n^* - 1) + 1)} \Big\} \Big\} \\
&\quad + \parallel g_{f,V} \parallel \omega E \left| e^{V(S_n+1) - V(S_n)} - E e^{V(S_n+1) - V(S_n)} \right|,
\end{aligned}$$

where we put $S(x,y) = \sum_{\ell=\min(x,y)+1}^{\max(x,y)} \frac{1}{\ell}$.

Proof. With the notation as in Example 3.4, denote the conditional expectation given $I = i$ by E_i. Let $f \in \mathcal{B}$. If $X_i \sim Be(p_i)$, $i = 1, \ldots, n$, then $X_i^* = 1$. Put

$$\hat{S}_{n,i} = \sum_{j \neq i} \hat{X}_j.$$

By (3.8) and (3.5), we have

$$\begin{aligned}
&|Ef(S_n) - \mu(f)| \\
&= \omega \Bigg| \sum_{i=1}^n \frac{p_i}{\sum_{j=1}^n p_j} \Bigg\{ E_i \Big[e^{V(S_n+1)-V(S_n)} \, g_{f,V}(S_n+1) \Big] \\
&\quad - Ee^{V(S_n+1)-V(S_n)} \, Eg_{f,V}\big(\sum_{j \neq i} \hat{X}_j + 1\big) \Bigg\} \Bigg| \\
&= \omega \Bigg| \sum_{i=1}^n \frac{p_i}{\sum_{j=1}^n p_j} E_i \Big[e^{V(S_n+1)-V(S_n))} \big(g_{f,V}(S_n+1) - g_{f,V}(\hat{S}_{n,i}+1) \big) \Big] \\
&\quad + E_i \, g_{f,V}\big(\hat{S}_{n,i}+1\big) \Big(e^{V(S_n+1)-V(S_n)} - E\, e^{V(S_n+1)-V(S_n)} \Big) \Bigg| \qquad (4.9) \\
&\leq \frac{\omega}{\sum_{j=1}^n p_j} \sum_{i=1}^n p_i \, E_i \Bigg\{ e^{V(S_n+1)-V(S_n)} \min \Bigg\{ S(S_n, \hat{S}_{n,i}), \\
&\quad (|S_n - \hat{S}_{n,i}|/\omega) \, e^{V(\min(S_n,\hat{S}_{n,i}))-V(\min(S_n,\hat{S}_{n,i})+1)} \Bigg\} \Big| S_n \neq \hat{S}_{n,i} \Bigg\} P(S_n \neq \hat{S}_{n,i}) \\
&\quad + \parallel g_{f,V} \parallel \, \omega \, E \Big| e^{V(S_n+1)-V(S_n)} - Ee^{V(S_n+1)-V(S_n)} \Big|,
\end{aligned}$$

where we used Lemma 2.9 for the last inequality. □

Note that in contrast to Theorem 4.5, we did not employ the difference between generators in the proof.

4.1. **Poisson approximation for a sum of indicators.** In the Poisson case we can use the bound (2.30) on Δg instead of Lemma 2.9 to obtain improved overall bounds, as follows. Let $X_i \sim Be(p_i)$, $i = 1, \ldots, n$, then as in Example 3.4 we have for $S_n = \sum_{j=1}^n X_j$ that $S_n^* = \hat{S}_{n,I} + 1$, where as above $\hat{S}_{n,i} = \sum_{j \neq i} \hat{X}_j$. Let $\lambda = ES_n = \sum_{j=1}^n p_j$. Recall that if $\mu = Po(\lambda)$ then $\omega = \lambda$ and $V(k) = -\lambda$. Hence using Equality (4.9) we obtain

$$|Ef(S_n) - Po(\lambda)(f)| = \Bigg| \sum_{i=1}^n p_i \, E_i \Big(g_{f,V}(S_n+1) - g_{f,V}(\hat{S}_{n,i}) \Big) \Bigg|.$$

Calculating as in the proof of Proposition 4.7 and using (2.30) we get

$$d_{TV}(\mathcal{L}(S_n), Po(\lambda)) \qquad (4.10)$$
$$\leq \sum_{i=1}^{n} p_i E_i \left\{ \min\left\{ S(S_n, \hat{S}_{n,i}), \frac{|S_n - \hat{S}_{n,i}|}{\lambda}\left(1 - e^{-\lambda}\right) \right\} |S_n \neq \hat{S}_{n,i} \right\} P(S_n \neq \hat{S}_{n,i}).$$

In particular we recover the bound for couplings in [4], Theorem 1.B,

$$d_{TV}(\mathcal{L}(S_n), Po(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{i=1}^{n} p_i E_i |S_n - \hat{S}_{n,i}|. \qquad (4.11)$$

Example 4.12. *Assume that the variables S_n and S_n* can be coupled such that*

$$|S_n - \hat{S}_{n,i}| \leq 1, \quad i = 1, \ldots, n. \qquad (4.13)$$

Then we can improve the bound (4.11) , *using*

$$\begin{aligned} & E_i \left\{ S(S_n, \hat{S}_{n,i}) | S_n \neq \hat{S}_{n,i} \right\} P(S_n \neq \hat{S}_{n,i}) \\ \leq \; & E_i \left\{ \frac{S_n - \hat{S}_{n,i}}{1 + \hat{S}_{n,i}} | S_n \neq \hat{S}_{n,i} \right\} P(S_n \neq \hat{S}_{n,i}) \\ \leq \; & P(S_n \neq \hat{S}_{n,i}) \left[P\left\{ \hat{S}_{n,i} = 0 | S_n \neq \hat{S}_{n,i} \right\} + \frac{1}{2} P\left\{ \hat{S}_{n,i} \geq 1 | S_n \neq \hat{S}_{n,i} \right\} \right] \\ = \; & P(S_n \neq \hat{S}_{n,i}) \frac{1}{2}\left(1 + P\left\{ \hat{S}_{n,i} = 0 | S_n \neq \hat{S}_{n,i} \right\}\right). \end{aligned}$$

Example 4.14. *In case that the X_i's are independent and we have $\hat{S}_{n,i} = S_n - X_i$, so that* (4.13) *is satisfied, and* (4.11) *yields that*

$$d_{TV}(\mathcal{L}(S_n), Po(\lambda)) \leq \sum_{i=1}^{n} p_i^2 \frac{1 - e^{-\lambda}}{\lambda}. \qquad (4.15)$$

The bound (4.15) *coincides with the bound (1.23) given in* [4]*. We can improve on it by using that*

$$P\left\{ \hat{S}_{n,i} = 0 | S_n \neq \hat{S}_{n,i} \right\} = P\left(\sum_{j \neq i} X_j = 0 | X_i = 1 \right) = \prod_{j \neq i} (1 - p_j).$$

This results in the bound for the independent case

$$d_{TV}(\mathcal{L}(S_n), Po(\lambda)) \leq \sum_{i=1}^{n} p_i^2 \min\left\{ \frac{1}{2}\left(1 + \prod_{j \neq i}(1 - p_j)\right), \frac{1 - e^{-\lambda}}{\lambda} \right\}.$$

To see when the above bound is an improvement on the bound (1.23) given in [4], *consider the inequality*

$$\frac{1}{2}(1 + x) < \frac{1 - e^{-\lambda}}{\lambda}.$$

We re-arrange this inequality to give

$$\lambda x < 2 - \lambda - 2e^{-\lambda}.$$

Numerical calculations yield that the right-hand side is positive when $0 < \lambda < 1.59362$. *For such* λ *and some* y *such that* $0 < y < 2 - \lambda - 2e^{-\lambda}$ *we can construct strongly inhomogeneous cases such that*

$$\prod_{j \neq i} (1 - p_j) \leq \lambda^{-1} y$$

for some indices i. *Examples are, for* $\lambda = 1$, *the vector* $\mathbf{p} = (p_1, \dots, p_n) = (1, 0, 0, \dots, 0)$, *or the vector* $\mathbf{p} = (p_1, \dots, p_n) = (1 - 1/(n-1), 1/(n-1)^2, 1/(n-1)^2, \dots, 1/(n-1)^2)$ *for* $n \geq 3$. *In such cases, our bound provides an improvement.*

Example 4.14 can only be an improvement on (4.11) when the underlying random variables are not identically distributed; in the independent and *identically distributed* case, we recover exactly the bound given in [4, (1.23), page 8].

4.2. **An example with repelling interaction.** Our first example with interaction partitions $A = [0, 1]$ into n intervals, the *ith* interval given by $S_i^n = [\frac{i-1}{n}, \frac{i}{n}]$, and we choose $q_i^n = \frac{2i-1}{2n}$, the midpoint of the interval S_i^n. Then each S_i^n has volume $\frac{1}{n} = \frac{\lambda}{\lambda n}$; we choose $z_i^n = \frac{\lambda}{n}$ and $z_n = \lambda$. Note the freedom of choice in $\lambda > 0$. We consider the set of functions f_k given by $f_0 = 1, f_1(x) = 1$ for all x, and for $k \geq 2$, $f_k(x_1, \dots, x_k) = 0$ if $x_i = x_j$ for some $i \neq j$, and otherwise

$$f_k(x_1, \dots, x_k) = \sum_{1 \leq i \neq j \leq k} (x_i - x_j)^2,$$

so that values $x_1, \dots, x_n$ which are far away from each other are preferred. Then clearly f_k satisfies Assumption 4.1 as well as Conditions (a) and (b). To avoid trivialities we assume that $n \geq 2$.

In our set-up we choose 0-1 random variables $X_m^n, 1 \leq m \leq n$, with density function

$$\begin{aligned} P(X_1^n = a_1, \dots, X_n^n = a_n) &\propto n^{-k} \lambda^k \sum_{1 \leq i \neq j \leq k} (q_i^n - q_j^n)^2 \quad \text{if } k = \sum_{m=1}^n a_m \geq 2; \\ P(X_1^n = a_1, \dots, X_n^n = a_n) &\propto \frac{\lambda}{n} \quad \text{if } k = 1; \\ P(X_1^n = 0, \dots, X_n^n = 0) &\propto 1. \end{aligned}$$

With $S_n = \sum_{m=1}^n X_m^n$ we have with (4.2) that

$$P(S_n = k) \propto \frac{\lambda^k}{k!} W_n(k),$$

where

$$W_n(k) = n^{-k} \sum_{i_1=1}^n \cdots \sum_{i_k=1}^n f_k(q_{i_1}^n, \dots, q_{i_k}^n), \quad k = 0, \dots, n.$$

Let S be a nonnegative integer valued random variable defined as in (4.4) by

$$P(S=k)=\frac{\frac{\lambda^k}{k!}\int_{[0,1]^k} f_k(x_1,\ldots,x_k)dx_1\cdots dx_k}{\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\int_{[0,1]^k} f_k(x_1,\ldots,x_k)dx_1\cdots dx_k} \propto \frac{\lambda^k}{k!}W(k),$$

with

$$W(k) \;=\; \int_{[0,1]^k} f_k(x_1,\ldots,x_k)dx_1\cdots dx_k, \quad k\geq 0.$$

It is immediate that $W(0)=W(1)=1$, and for $k\geq 2$,

$$W(k) \;=\; \sum_{i=1}^{k}\sum_{j\neq i}\int_0^1\int_0^1 (x_i-x_j)^2\,dx_i\,dx_j=\frac{k(k-1)}{6}.$$

Thus our limiting Gibbs measure μ has normalising constant

$$\mathcal{Z}=1+\lambda+\frac{1}{6}\sum_{k\geq 2}\frac{\lambda^k}{(k-2)!}=1+\lambda+\frac{\lambda^2}{6}e^{\lambda}$$

and is therefore given by

$$\begin{aligned}
\mu(0) &= \left(1+\lambda+\frac{\lambda^2}{6}e^{\lambda}\right)^{-1}\\
\mu(1) &= \frac{\lambda}{1+\lambda+\frac{\lambda^2}{6}e^{\lambda}}\\
\mu(k) &= \left(1+\lambda+\frac{\lambda^2}{6}e^{\lambda}\right)^{-1}\frac{\lambda^k}{(k-2)!}, \quad k\geq 2.
\end{aligned}$$

To assess the distance between the distributions of S_n and of S we note that $W_n(0)=W_n(1)=1=W(0)=W(1)$, and some algebra yields, for $k\geq 2$, that

$$\begin{aligned}
W_n(k) &= n^{-k}\sum_{i_1=1}^{n}\cdots\sum_{i_k=1}^{n}\sum_{l\neq s}\left(\frac{2i_s-1}{2n}-\frac{2i_j-1}{2n}\right)^2\\
&= n^{-k-2}\sum_{1\leq l\neq s\leq k}\sum_{i=1}^{n}\sum_{j=1}^{n}(i-j)^2 n^{k-2}\\
&= 2n^{-4}k(k-1)\left\{\sum_{i=1}^{n} i^2-\left(\sum_{i=1}^{n} i\right)^2\right\}\\
&= \frac{k(k-1)(n+1)(n-1)}{6n^2}.
\end{aligned}$$

Thus

$$\frac{W_n(k)}{W_n(k-1)}=\frac{W(k)}{W(k-1)}$$

for $k = 0, \ldots, n$, and we are in the situation of Example 3.15. Therefore it follows that

$$\begin{aligned} d_{TV}(\mu_n, \mu) &= \sum_{k=n+1}^{\infty} \mu(k) \\ &= \left(1 + \lambda + \frac{\lambda^2}{6} e^{\lambda}\right)^{-1} \sum_{k=n+1}^{\infty} \frac{\lambda^k}{k!} \\ &\leq \frac{\lambda^{n+1} e^{\lambda}}{(n+1)! \left(1 + \lambda + \frac{\lambda^2}{6} e^{\lambda}\right)}. \end{aligned}$$

Summarising,

Corollary 4.16. *Let S_n and S be as constructed above. Then*

$$d_{TV}(\mathcal{L}(S_n), \mu) \leq \frac{\lambda^{n+1} e^{\lambda}}{(n+1)! \left(1 + \lambda + \frac{\lambda^2}{6} e^{\lambda}\right)}.$$

4.3. **A second example with interaction.** Using the same partition of $A = [0, 1]$ and the same notation as in the previous example, and choosing for simplicity $\omega = \lambda = 1$, we now consider the set of functions $(f_k)_k$ given by $f_0(x) = f_1(x) = 1$ for all x, and for $k \geq 2$,

$$f_k(x_1, \ldots, x_k) = \prod_{1 \leq i \neq j \leq k} x_i \, x_j.$$

Clearly $(f_k)_k$ satisfy Assumption 4.1 as well as Condition (a) and (b). We take $q_i^n = \frac{i-1}{n}$, the left endpoint of the interval S_i^n. Assume that $n \geq 3$. Along the lines of the calculations in the previous example we obtain now $W(0) = W(1) = 1$ and for $k \geq 2$

$$W(k) = k^{-k},$$

so that the distribution of S given in (4.4) is

$$\mu(k) = \frac{1}{\mathcal{Z}} \frac{1}{k!} k^{-k}, \quad k \geq 0.$$

We note that

$$1 \leq \mathcal{Z} = 1 + \sum_{k=1}^{\infty} \frac{1}{k!} e^{-k \ln k} \leq e^{\frac{1}{e}}.$$

For the distribution of S_n given in (4.2) we obtain $W_n(0) = W_n(1) = 1$, and for $k \geq 2$,

$$W_n(k) = n^{-k^2} \left(\sum_{i=0}^{n-1} i^{k-1} \right)^k.$$

We employ an integral approximation, obtaining

$$\left(\frac{n-1}{n}\right)^{k^2} k^{-k} < W_n(k) < k^{-k}, \; k \geq 2.$$

Observe that for $k = 0, \dots, n$,

$$\left(\frac{n-1}{n}\right)^{k^2} < \frac{W_n(k)}{W(k)} < 1.$$

For Theorem 4.5, we may thus bound for $k = 2, \dots, n$,

$$\begin{aligned}
&\left| \frac{W(k)W_n(k-1)}{W(k-1)W_n(k)} - 1 \right| \\
&\leq \max\left\{ 1 - \left(\frac{n-1}{n}\right)^{k^2} ; \left(\frac{n}{n-1}\right)^{k^2} - 1 \right\} \\
&= \left(\frac{n}{n-1}\right)^{k^2} - 1.
\end{aligned}$$

Now we calculate

$$\sum_k k\mu_n(k) \left| \frac{W(k)W_n(k-1)}{W(k-1)W_n(k)} - 1 \right| \leq \sum_k k\mu_n(k) \left\{ \left(\frac{n}{n-1}\right)^{k^2} - 1 \right\}.$$

Taylor expansion gives the inequality

$$0 \leq \left(\frac{n}{n-1}\right)^{k^2} - 1 \leq \frac{k^2}{n-1}\left(1 + \frac{1}{n-1}\right)^{k^2} < \frac{k^2}{n-1} e^{\frac{k^2}{n-1}} < \frac{k^2}{n-1} e^k,$$

for $k \leq n$ and $n \geq 3$. Applying this inequality, we obtain that

$$\begin{aligned}
\sum_k k\mu_n(k) \left| \frac{W(k)W_n(k-1)}{W(k-1)W_n(k)} - 1 \right| &\leq \frac{1}{(n-1)\mathcal{Z}_n} \sum_{k=1}^{n-1} \frac{1}{k!} e^{-k\log(k)+k} \\
&< \frac{e^e}{(n-1)\mathcal{Z}_n} < \frac{2e^e}{n},
\end{aligned}$$

where we used $\mathcal{Z}_n = 1 + \sum_{k=1}^{\infty} \frac{1}{k!} W_n(k) \geq 1$ in the last inequality. Next we estimate

$$\sum_{k=n+1}^{\infty} \mu(k) = \frac{1}{\mathcal{Z}} \sum_{k=n+1}^{\infty} \frac{1}{k!} k^{-k} \leq \sum_{k=n+1}^{\infty} \frac{1}{k!} n^{-k} \leq e^{\frac{1}{n}} \frac{n^{-(n+1)}}{(n+1)!},$$

where we used that $\mathcal{Z} > 1$.

Note that $\lambda_2 = \sup_k \frac{(k+1)^{k+1}}{k^k} = \infty$, so that we cannot apply Lemma 2.19; instead we find an alternative bound on $\| g_{f,V} \|$ as follows. From Equation (2.6) we have

$$|g_{f,V}(j+1)| \leq \frac{j!}{\omega^{j+1}} e^{-V(j+1)} \sum_{k=j+1}^{N} e^{V(k)} \frac{\omega^k}{k!} \leq \sum_{k=0}^{\infty} \frac{1}{k!} k^{-k} = \mathcal{Z} \leq e^{\frac{1}{e}}.$$

Corollary 4.17. *Let S_n and S be as constructed above, and let $n \geq 3$. Then*

$$d_{TV}(\mathcal{L}(S_n), \mu) \leq 2\frac{e^{e+1/e}}{n} + e^{\frac{1}{n}} \frac{n^{-(n+1)}}{(n+1)!}.$$

Acknowledgement. The authors would like to thank Hans Zessin for pointing out the work of Preston as well as the work of Chayes and Klein. We would also like to thank Andrew Barbour and two anonymous referees for helpful comments.

References

1. S. Asmussen, *Applied probability and queues*, John Wiley & Sons Ltd., Chichester, 1987. MR 89a:60208
2. A. D. Barbour, *Stein's method and Poisson process convergence*, J. Appl. Probab. (1988), no. Special Vol. 25A, 175–184, A celebration of applied probability. MR 90a:60064
3. ______, *Stein's method for diffusion approximations*, Probab. Theory Related Fields **84** (1990), no. 3, 297–322. MR 91d:60081
4. A. D. Barbour, L. Holst, and S. Janson, *Poisson Approximation*, Oxford University Press, Oxford, 1992.
5. T. C. Brown and A. Xia, *Stein's method and birth-death processes*, Ann. Probab. **29** (2001), no. 3, 1373–1403. MR 1 872 746
6. L. Chayes and D. Klein, *A generalization of Poisson convergence to "Gibbs convergence" with applications to statistical mechanics*, Helv. Phys. Acta **67** (1994), no. 1, 30–42. MR 95i:82015
7. W. Ehm, *Binomial approximation to the Poisson binomial distribution*, Statist. Probab. Lett. **11** (1991), no. 1, 7–16. MR 92b:60024
8. L. Goldstein and G. Reinert, *Distributional transformations, orthogonal polynomials, and Stein characterizations*, J. Theoret. Probab. **18** (2005), no. 1, 237–260. MR MR2132278 (2006b:60026)
9. L. Goldstein and Y. Rinott, *Multivariate normal approximations by Stein's method and size bias couplings*, J. Appl. Probab. **33** (1996), no. 1, 1–17. MR 97d:60038
10. F. Götze, *On the rate of convergence in the multivariate CLT*, Ann. Probab. **19** (1991), no. 2, 724–739. MR MR1106283 (92g:60028)
11. S. Holmes, *Stein's method for birth and death chains*, Stein's method: expository lectures and applications, IMS Lecture Notes Monogr. Ser., vol. 46, Inst. Math. Statist., Beachwood, OH, 2004, pp. 45–67. MR MR2118602
12. S. Karlin and J. McGregor, *The classification of birth and death processes*, Trans. Amer. Math. Soc. **86** (1957), 366–400. MR 20 #1363
13. E. A. Peköz, *Stein's method for geometric approximation*, J. Appl. Probab. **33** (1996), no. 3, 707–713. MR 97g:60034
14. M. J. Phillips and G. V. Weinberg, *Non-uniform bounds for geometric approximation*, Statist. Probab. Lett. **49** (2000), no. 3, 305–311. MR 2001i:60041
15. C. Preston, *Spatial birth-and-death processes*, Proceedings of the 40th Session of the International Statistical Institute (Warsaw, 1975), Vol. 2. Invited papers, vol. 46, 1975, With discussion, pp. 371–391, 405–408 (1975). MR 57 #14170
16. D. Ruelle, *Statistical mechanics*, World Scientific Publishing Co. Inc., River Edge, NJ, 1999, Rigorous results, Reprint of the 1989 edition. MR MR1747792 (2002f:82001)
17. C. Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory (Berkeley, Calif.), Univ. California Press, 1972, pp. 583–602. MR 53 #6687

(P. Eichelsbacher) Fakultät für Mathematik, Ruhr-Universität Bochum, NA 3/68, D-44780 Bochum, Germany

E-mail address: peter.eichelsbacher@ruhr-uni-bochum.de

(G. Reinert) Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK

E-mail address: reinert@stats.ox.ac.uk