DOI: 10.1214/EJP.v13-479

Electronic Journal of Probability

Vol. 13 (2008), Paper no. 20, pages 566–587.

Journal URL
http://www.math.washington.edu/~ejpecp/

Upper bounds for Stein-type operators

Fraser Daly
School of Mathematical Sciences
University of Nottingham
University Park
Nottingham, NG7 2RD
England
Email: fraser.daly@maths.nottingham.ac.uk

Abstract

We present sharp bounds on the supremum norm of $\mathcal{D}^j Sh$ for $j \geq 2$, where $\mathcal{D}$ is the differential operator and S the Stein operator for the standard normal distribution. The same method is used to give analogous bounds for the exponential, Poisson and geometric distributions, with $\mathcal{D}$ replaced by the forward difference operator in the discrete case. We also discuss applications of these bounds to the central limit theorem, simple random sampling, Poisson-Charlier approximation and geometric approximation using stochastic orderings .

Key words: Stein-type operator; Stein's method; central limit theorem; Poisson-Charlier approximation; stochastic ordering.

AMS 2000 Subject Classification: Primary 60F05, 60J80, 62E17.

Submitted to EJP on March 7, 2007, final version accepted April 1, 2008.

1 Introduction and results

Introduction and main results

Stein's method was first developed for normal approximation by Stein (1972). See Stein (1986) and Chen and Shao (2005) for more recent developments. These powerful techniques were modified by Chen (1975) for the Poisson distribution and have since been applied to many other cases. See, for example, Peköz (1996), Brown and Xia (2001), Erhardsson (2005), Reinert (2005) and Xia (2005).

We consider approximation by a random variable Y and write $\Phi_h := E[h(Y)]$. Following Stein's method, we assume we have a linear operator A such that

$$X \stackrel{d}{=} Y \iff E[Ag(X)] = 0 \text{ for all } g \in \mathcal{F}\,,$$

for some suitable class of functions $\mathcal{F}$. From this we construct the so-called Stein equation

$$h(x) - \Phi_h = Af(x)\,,$$

whose solution we denote $f := Sh$. We call S the Stein operator. If $Y \sim \mathrm{N}(0,1)$, Stein (1986) shows that we may choose $Ag(x) = g'(x) - xg(x)$ and

$$Sh(x) = \frac{1}{\varphi(x)} \int_{-\infty}^{x} (h(t) - \Phi_h)\,\varphi(t)\,dt\,, \tag{1.1}$$

where φ is the density of the standard normal random variable. It is this Stein operator we employ when considering approximation by the standard normal distribution. We also consider approximation by the exponential distribution, Poisson distribution and geometric distribution (starting at zero). In the case of the exponential distribution with mean λ^{-1} we use the Stein operator given, for $x \geq 0$, by

$$Sh(x) = e^{\lambda x} \int_0^x [h(t) - \Phi_h] e^{-\lambda t}\,dt\,. \tag{1.2}$$

When Y has the Poisson distribution with mean λ we let

$$Sh(k) = \frac{(k-1)!}{\lambda^k} \sum_{i=0}^{k-1} \left[h(i) - \Phi_h\right] \frac{\lambda^i}{i!}\,, \tag{1.3}$$

for $k \geq 1$. See, for example, Erhardsson (2005). We discuss possible choices of $Sh(0)$ following the proof of Theorem 1.3. For the geometric distribution with parameter $p = 1 - q$ we define $Sh(0) = 0$ and, for $k \geq 1$,

$$Sh(k) = \frac{1}{q^k} \sum_{i=0}^{k-1} \left[h(i) - \Phi_h\right] q^i\,. \tag{1.4}$$

Essential ingredients of Stein's method are so-called Stein factors giving bounds on the derivatives (or forward differences) of the solutions of our Stein equation. Theorems 1.1–1.4 present a selection of such bounds. Throughout we will let $\mathcal{D}$ and Δ denote the differential and forward difference operators respectively. As usual, the supremum norm of a real-valued function g is given by $\|g\|_\infty := \sup_x |g(x)|$.

Theorem 1.1. *Let $k \geq 0$ and S be the Stein operator given by (1.1). For $h : \mathbb{R} \mapsto \mathbb{R}$ $(k+1)$-times differentiable with $\mathcal{D}^k h$ absolutely continuous,*

$$\|\mathcal{D}^{k+2}Sh\|_\infty \leq \sup_{t\in\mathbb{R}} \mathcal{D}^{k+1}h(t) - \inf_{t\in\mathbb{R}} \mathcal{D}^{k+1}h(t) \leq 2\|\mathcal{D}^{k+1}h\|_\infty .$$

Theorem 1.2. *Let $k \geq 0$ and S be the Stein operator given by (1.2). For $h : \mathbb{R}^+ \mapsto \mathbb{R}$ $(k+1)$-times differentiable with $\mathcal{D}^k h$ absolutely continuous,*

$$\|\mathcal{D}^{k+2}Sh\|_\infty \leq \sup_{t\geq 0} \mathcal{D}^{k+1}h(t) - \inf_{t\geq 0} \mathcal{D}^{k+1}h(t) \leq 2\|\mathcal{D}^{k+1}h\|_\infty .$$

Theorem 1.3. *Let $k \geq 0$ and S be the Stein operator given by (1.3). For $h : \mathbb{Z}^+ \mapsto \mathbb{R}$ bounded,*

$$\sup_{i\geq 1} \left|\Delta^{k+2}Sh(i)\right| \leq \frac{1}{\lambda}\left(\sup_{i\geq 0} \Delta^{k+1}h(i) - \inf_{i\geq 0} \Delta^{k+1}h(i)\right) \leq \frac{2}{\lambda}\|\Delta^{k+1}h\|_\infty .$$

Theorem 1.4. *Let $k \geq 0$ and S be the Stein operator given by (1.4). For $h : \mathbb{Z}^+ \mapsto \mathbb{R}$ bounded,*

$$\|\Delta^{k+2}Sh\|_\infty \leq \frac{1}{q}\left(\sup_{i\geq 0} \Delta^{k+1}h(i) - \inf_{i\geq 0} \Delta^{k+1}h(i)\right) \leq \frac{2}{q}\|\Delta^{k+1}h\|_\infty .$$

Our work is motivated by that of Stein (1986, Lemma II.3). Stein proves that for $h : \mathbb{R} \mapsto \mathbb{R}$ bounded and absolutely continuous and S the Stein operator for the standard normal distribution

$$\begin{aligned} \|Sh\|_\infty &\leq \sqrt{\frac{\pi}{2}}\|h - \Phi_h\|_\infty , \\ \|\mathcal{D}Sh\|_\infty &\leq 2\|h - \Phi_h\|_\infty , \\ \|\mathcal{D}^2Sh\|_\infty &\leq 2\|\mathcal{D}h\|_\infty . \end{aligned} \tag{1.5}$$

Our work extends this result, though relies on the differentiability of h. Some of Stein's bounds above may be applied when h is not differentiable.

Barbour (1986, Lemma 5) shows that for $j \geq 0$

$$\|\mathcal{D}^{j+1}Sh\|_\infty \leq c_j\|\mathcal{D}^j h\|_\infty$$

for some universal constant c_j depending only on j. Rinott and Rotar (2003, Lemma 16) employ this formulation of Barbour's work. We show later that our Theorem 1.1 is sharp, so that $c_j = 2$ for $j \geq 2$. The proof we give in the normal case, however, relies more heavily on Stein's work than Barbour's. For the other cases considered in Theorems 1.2–1.4 we give proofs analogous to that for the standard normal distribution.

Goldstein and Reinert (1997, page 943) employ a similar result, derived from Barbour (1990) and Götze (1991), again in the case of the standard normal distribution. For a fixed $j \geq 1$ and $h : \mathbb{R} \mapsto \mathbb{R}$ with j bounded derivatives

$$\|\mathcal{D}^{j-1}Sh\|_\infty \leq \frac{1}{j}\|\mathcal{D}^j h\|_\infty . \tag{1.6}$$

It is straightforward to verify that this bound is also sharp. We take $h = H_j$, the jth Hermite polynomial, and get that $Sh = -H_{j-1}$.

Bounds of a similar type to ours have also been established for Stein-type operators relating to the chi-squared distribution and the weak law of large numbers. See Reinert (1995, Lemma 2.5) and Reinert (2005, Lemmas 3.1 and 4.1).

The central limit theorem and simple random sampling

Motivated by bounds in the central limit theorem proved, for example, by Ho and Chen (1978), Bolthausen (1984) and Barbour (1986) we consider applications of our Theorem 1.1.

Using the bound (1.6) Goldstein and Reinert (1997) give a proof of the central limit theorem. We could instead use our Theorem 1.1 in an analogous proof which relaxes the differentiability conditions of Goldstein and Reinert's result.

Goldstein and Reinert (1997, Theorem 3.1) also establish bounds on the difference $E[h(W)] - \Phi_h$ for a random variable W with zero mean and variance 1, where $\Phi_h := E[h(Y)]$ and $Y \sim \mathrm{N}(0,1)$. These bounds are given in terms of the random variable W^* with the W-zero biased distribution, defined such that for all differentiable functions f for which $E[Wf(W)]$ exists $E[Wf(W)] = E[\mathcal{D}f(W^*)]$. We can again use our Theorem 1.1 in an analogous proof to derive the following.

Corollary 1.5. *Let W be a random variable with mean zero and variance 1, and assume (W, W^*) is given on a joint probability space, such that W^* has the W-zero biased distribution. Furthermore, let $h : \mathbb{R} \mapsto \mathbb{R}$ be twice differentiable, with h and $\mathcal{D}h$ absolutely continuous. Then*

$$|E[h(W)] - \Phi_h| \leq 2\|\mathcal{D}h\|_\infty \sqrt{E\left[E[W^* - W|W]^2\right]} + \|\mathcal{D}^2 h\|_\infty E\left[W^* - W\right]^2 .$$

We can further follow Goldstein and Reinert (1997) in applying our Corollary 1.5 to the case of simple random sampling, obtaining the following analogue of their Theorem 4.1.

Corollary 1.6. *Let $X_1, \ldots, X_n$ be a simple random sample from a set $\mathcal{A}$ of N (not necessarily distinct) real numbers satisfying*

$$\sum_{a \in \mathcal{A}} a = \sum_{a \in \mathcal{A}} a^3 = 0 .$$

Set $W := X_1 + \cdots + X_n$ and suppose $Var(W) = 1$. Let $h : \mathbb{R} \mapsto \mathbb{R}$ be twice differentiable, with h and $\mathcal{D}h$ absolutely continuous. Then

$$|E[h(W)] - \Phi_h| \leq 2C\|\mathcal{D}h\|_\infty + \left(11 \sum_{a \in \mathcal{A}} a^4 + \frac{45}{N}\right) \|\mathcal{D}^2 h\|_\infty ,$$

where $C := C_1(N, n, \mathcal{A})$ is given by Goldstein and Reinert (1997, page 950).

Applications of our Theorem 1.1 to other CLT-type results come in combining our work with Proposition 4.2 of Lefèvre and Utev (2003). This gives us Corollary 1.7.

Corollary 1.7. *Let $X_1, \ldots, X_n$ be independent random variables each with zero mean such that $Var(X_1) + \cdots + Var(X_n) = 1$. Suppose further that for a fixed $k \geq 0$*

$$E[X_j^i] = Var(X_j)^{\frac{i}{2}} E[Y^i] ,$$

for all $i = 1, \ldots, k+2$ and $j = 1, \ldots, n$. Write $W = X_1 + \cdots + X_n$. If $h : \mathbb{R} \mapsto \mathbb{R}$ is $(k+1)$-times differentiable with $\mathcal{D}^k h$ absolutely continuous then

$$
\begin{aligned}
|E[h(W)] - \Phi_h| &\leq p_{k+2}\left(\sup_t \mathcal{D}^{k+1}h(t) - \inf_t \mathcal{D}^{k+1}h(t)\right) L_{n,k+3} \\
&\leq 2p_{k+2}\|\mathcal{D}^{k+1}h\|_\infty L_{n,k+3}\,,
\end{aligned}
$$

where $L_{n,j} := E[|X_1|^j] + \cdots + E[|X_n|^j]$ is the Lyapunov characteristic, $p_2 = 0.5985269$, $p_3 = \frac{1}{3}$ and $p_j \leq \frac{1}{(j-1)!}$ for $j \geq 4$.

We need show only the value of the universal constant p_3 to establish Corollary 1.7. This proof is given in Section 2. The remainder of the result follows immediately from Lefèvre and Utev's (2003) work.

As noted by the referee, the bounds in our Theorem 1.1 may also be applied to Edgeworth expansions. See, for example, Barbour (1986), Rinott and Rotar (2003) and Rotar (2005).

Corollary 1.7 may be used to improve the constant in Theorem 3.2 of Chen and Shao (2005), yielding

$$
\left| E\,|W| - \sqrt{\frac{2}{\pi}} \right| \leq 2p_2 L_{n,3}\,,
$$

for suitable random variables $X_1, \ldots, X_n$.

In the large deviation case we may employ our Corollary 1.7 to give an improvement on a constant of Lefèvre and Utev (2003, page 364). We consider a sequence $X, X_1, X_2, \ldots$ of iid random variables with zero mean and unit variance and write $W := X_1 + \cdots + X_n$. We let $Y \sim \mathrm{N}(0,1)$. Define $t \in \mathbb{R}$ and the random variable U, with expected value α, as in Lefèvre and Utev (2003, page 364). Following an analogous argument, we apply our Corollary 1.7 to the function $h(x) := x_+ e^{-rx}$, where $r := t\sqrt{n\mathrm{Var}(U)}$, to obtain

$$
E[(W - n\alpha)_+] = \sqrt{n\mathrm{Var}(U)}E^n[e^{t(X-\alpha)}]\left(E\left[Y_+ e^{-rY}\right] + \nu\frac{E|U-\alpha|^3}{\sqrt{n}[\mathrm{Var}(U)]^{\frac{3}{2}}}\right),
$$

where $|\nu| \leq p_2\left(\sup_t \mathcal{D}h(t) - \inf_t \mathcal{D}h(t)\right) = p_2\left(1 + e^{-2}\right)$.

By modifying the representation in Lefèvre and Utev (2003, page 361) used in deriving Corollary 1.7 we may also prove the following bound.

Corollary 1.8. *Let $X_1, \ldots, X_n$ be independent random variables with zero mean and with $Var(X_j) = \sigma_j^2$ for $j = 1, \ldots, n$. Let $\sigma_1^2 + \cdots + \sigma_n^2 = 1$. Suppose $h : \mathbb{R} \mapsto \mathbb{R}$ is twice differentiable with h and $\mathcal{D}h$ absolutely continuous. Then*

$$
\left| E[h(W)] - \Phi_h + \frac{1}{2}\sum_{j=1}^n E\left[X_j^3\right] E\left[\mathcal{D}^2 Sh(Y_j)\right] \right|
$$

$$
\leq 2p_2\|\mathcal{D}^2 h\|_\infty L_{n,3}\sum_{j=1}^n \frac{E[|X_j|^3]}{1-\sigma_j^2} + 2p_3\|\mathcal{D}^2 h\|_\infty L_{n,4}\,,
$$

where $Y_j \sim N(0, 1-\sigma_j^2)$ and S is given by (1.1).

We postpone the proof of Corollary 1.8 until Section 2.

Poisson-Charlier approximation

We suppose $X_1, \dots, X_n$ are independent random variables with $P(X_i = 1) = \theta_i = 1 - P(X_i = 0)$ for $1 \le i \le n$ and $W := X_1 + \cdots + X_n$. Define $\lambda := \sum_{i=1}^n \theta_i$, $\lambda_m := \sum_{i=1}^n \theta_i^m$ and let $\kappa_{[j]}$ be the jth factorial cumulant of W. We follow Barbour (1987) and Barbour *et al* (1992) in considering the approximation of W by the Poisson-Charlier signed measures $\{Q_l : l \ge 1\}$ on $\mathbb{Z}^+$ defined by

$$Q_l(m) := \left(\frac{e^{-\lambda}\lambda^m}{m!}\right)\left(1 + \sum_{s=1}^{l-1}\sum_{[s]}\prod_{j=1}^{s}\left[\frac{1}{r_j!}\left(\frac{\kappa_{[j+1]}}{(j+1)!}\right)^{r_j}\right] C_{R+s}(\lambda, m)\right),$$

where $C_n(\lambda, x)$ is the nth Charlier polynomial, $\sum_{[s]}$ denotes the sum over all s-tuples $(r_1, \dots, r_s) \in (\mathbb{Z}^+)^s$ with $\sum_{j=1}^s j r_j = s$, and we let $R := \sum_{j=1}^s r_j$.

Barbour (1987) shows that for $h : \mathbb{Z}^+ \mapsto \mathbb{R}$ bounded and $l \ge 1$ fixed such that $E[X_i^{l+1}] < \infty$ for all $1 \le i \le n$ we have

$$\left|E[h(W)] - \int h \, dQ_l\right| \le \eta_l \,,$$

where

$$|\eta_l| \le \lambda_{l+1} \sum_{(l)} \lambda^k \left\| \left(\prod_{j=1}^{k+1} \Delta^{s_j} T\right) h \right\|_\infty . \tag{1.7}$$

We let T be the operator such that $Th(k) = Sh(k+1)$, S the Stein operator for the Poisson distribution with mean λ and $\sum_{(l)}$ denote the sum over

$$\left\{(s_1, \dots, s_{k+1}) \in \mathbb{N}^{k+1} : k \in \{0, \dots, l-1\}, \sum_{j=1}^{k+1} s_j = l\right\}.$$

Now, we use our Theorem 1.3 and a result of Barbour (1987, Lemma 2.2) to note that $\|\Delta^s Th\|_\infty \le \frac{2}{\lambda}\|\Delta^{s-1}h\|_\infty$ for all $s \ge 1$, so that

$$\left\| \left(\prod_{j=1}^{k+1} \Delta^{s_j} T\right) h \right\|_\infty \le \left(\frac{2}{\lambda}\right)^{k+1} \left\|\Delta^{l-k-1} h\right\|_\infty .$$

Hence, using (1.7),

$$|\eta_l| \le \frac{\lambda_{l+1}}{\lambda} \sum_{(l)} 2^{k+1} \|\Delta^{l-k-1} h\|_\infty = \frac{\lambda_{l+1}}{\lambda} \sum_{k=0}^{l-1} \binom{l-1}{k} 2^{k+1} \|\Delta^{l-k-1} h\|_\infty \,, \tag{1.8}$$

for $l \ge 2$. This provides an alternative bound to that established in Barbour's work. Barbour (1987, page 756) further considers the case in which

$$P(X_i = m) = \binom{k_i + m - 1}{m} \theta_i^m (1 - \theta_i)^{k_i} \,,$$

for all $m \geq 0$. We can prove a bound analogous to (1.8) in this situation, with λ and λ_{l+1} replaced by $\sum_{i=1}^n k_i \left(\frac{\theta_i}{1-\theta_i}\right)$ and $\sum_{i=1}^n k_i \left(\frac{\theta_i}{1-\theta_i}\right)^{l+1}$ respectively.

Geometric approximation using stochastic orderings

We present here an application of our results based on unpublished work of Utev. Suppose Y is a random variable on $\mathbb{Z}^+$ with characterising linear operator A. For $k \geq 0$ and X another random variable on $\mathbb{Z}^+$ define $m_k^{(0)} := E[AI(X = k)]$ and $m_k^{(l)} := \sum_{j=k}^\infty m_j^{(l-1)}$ for $l \geq 1$. For $g : \mathbb{Z}^+ \mapsto \mathbb{R}$ bounded we can write

$$E[Ag(X)] = \sum_{k=0}^\infty g(k)[m_k^{(1)} - m_{k+1}^{(1)}] .$$

Rearranging this sum we get

$$E[Ag(X)] = \sum_{k=0}^\infty \Delta g(k) m_{k+1}^{(1)} + g(0) m_0^{(1)} .$$

A similar argument yields

$$E[Ag(X)] = \sum_{k=0}^\infty \Delta^l g(k) m_{k+l}^{(l)} + \sum_{j=0}^{l-1} \Delta^j g(0) m_j^{(j+1)} ,$$

for any $l \geq 1$. We can take, for example, $g = Sh$, where S is the Stein operator for the random variable Y, so that $E[ASh(X)] = E[h(X)] - E[h(Y)]$. Supposing that $Sh(0) = 0$, we obtain the following.

Proposition 1.9. *Let $l \geq 1$ and suppose $m_j^{(j+1)} = 0$ for $j = 1, \ldots, l-1$. Then*

$$\left|E[h(X)] - E[h(Y)]\right| \leq \|\Delta^l Sh\|_\infty \sum_{k=0}^\infty |m_{k+l}^{(l)}| .$$

Furthermore, if $m_{k+l}^{(l)}$ has the same sign for each $k \geq 0$ we have

$$\left|E[h(X)] - E[h(Y)]\right| \leq \|\Delta^l Sh\|_\infty \Big| \sum_{k=0}^\infty m_{k+l}^{(l)} \Big| .$$

Consider the case $l = 2$ and $P(Y = k) = pq^k$ for $k \geq 0$, so that Y has the geometric distribution starting at zero with parameter $p = 1 - q$. It is well-known that in this case we can choose $Ag(k) = qg(k+1) - g(k)$. See, for example, Reinert (2005, Example 5.3). With this choice we may also define $Sh(0) = 0$. It is straightforward to verify that, by the linearity of A,

$$m_k^{(1)} = qP(X \geq k-1) - P(X \geq k) , \tag{1.9}$$

$$m_k^{(2)} = E[A(X - k + 1)_+] , \tag{1.10}$$

so that $m_1^{(2)} = 0$ if and only if $E[X] = \frac{q}{p} = E[Y]$. Furthermore,

$$\sum_{k=2}^{\infty} m_k^{(2)} = \frac{p}{2}\left(\frac{q(1+q)}{p^2} - E[X^2]\right) = \frac{p}{2}\big(E[Y^2] - E[X^2]\big)\,,$$

assuming $E[X] = E[Y]$.

So, combining the above with Theorem 1.4, we assume that X is a random variable on $\mathbb{Z}^+$ with $E[X] = \frac{q}{p}$, such that $E[A(X-k+1)_+]$ has the same sign for each $k \geq 2$. Then, for all $h : \mathbb{Z}^+ \mapsto \mathbb{R}$ bounded

$$\big|E[h(X)] - E[h(Y)]\big| \leq \frac{p}{q}\|\Delta h\|_\infty \left|E[Y^2] - E[X^2]\right|\,. \tag{1.11}$$

We consider an example from Phillips and Weinberg (2000). Suppose m balls are placed randomly in d compartments, with all assignments equally likely, and let X be the number of balls in the first compartment. Then X has a Pólya distribution with

$$P(X = k) = \frac{\binom{d+m-k-2}{m-k}}{\binom{d+m-1}{m}}\,,$$

for $0 \leq k \leq m$. We compare X to $Y \sim \text{Geom}\left(\frac{d}{d+m}\right)$. Then $E[X] = E[Y] = \frac{m}{d}$, $E[X^2] = \frac{m(d+2m-1)}{d(d+1)}$ and $E[Y^2] = \frac{m(d+2m)}{d^2}$. It can easily be checked that $m_k^{(2)} \geq 0$ for all $k \geq 2$, so that our bound (1.11) becomes

$$\big|E[h(X)] - E[h(Y)]\big| \leq \|\Delta h\|_\infty \frac{2(d+m)}{d(d+1)}\,,$$

and in particular $d_{TV}(X,Y) \leq \frac{2(d+m)}{d(d+1)}$. In many cases this performs better than the analogous bound found by Phillips and Weinberg (2000, page 311).

We consider a further application of (1.11). Suppose $X = X(\xi) \sim \text{Geom}(\xi)$ for some random variable ξ taking values in $[0,1]$. Let $Y \sim \text{Geom}(p)$ with p chosen such that $E[X] = E[Y]$, that is

$$\frac{1}{p} = E\left[\frac{1}{\xi}\right]\,. \tag{1.12}$$

Using (1.10) we get that

$$m_k^{(2)} = E\left[(1-\xi)^{k-1}\frac{(\xi - p)}{\xi}\right]\,,$$

So that

$$m_k^{(2)} \leq 0 \iff \text{Cov}\left(\frac{1}{\xi}, (1-\xi)^{k-1}\right) \geq 0\,. \tag{1.13}$$

For example, suppose $\xi \sim \text{Beta}(\alpha, \beta)$ for some $\alpha > 2$, $\beta > 0$. It can easily be checked that the criterion (1.13) is satisfied for all $k \geq 2$ and the correct choice of p, using (1.12), is

$$p = \frac{\alpha - 1}{\alpha + \beta - 1}\,.$$

We get that $E[X^2] = \frac{\beta(\alpha+2\beta)}{(\alpha-1)(\alpha-2)}$ and $E[Y^2] = \frac{\beta(\alpha+2\beta-1)}{(\alpha-1)^2}$, so our bound (1.11) becomes

$$\big|E[h(X)] - E[h(Y)]\big| \leq \|\Delta h\|_\infty \frac{2(\alpha+\beta-1)}{(\alpha-1)(\alpha-2)}\,,$$

and in particular $d_{TV}(X,Y) \leq \frac{2(\alpha+\beta-1)}{(\alpha-1)(\alpha-2)}$.

2 Proofs

Proof of Theorem 1.1

In order to prove our theorem we introduce a variant of Mills' ratio and exploit several of its properties. We define the function $Z : \mathbb{R} \mapsto \mathbb{R}$ by

$$Z(x) := \frac{\Phi(x)}{\varphi(x)}\,,$$

where Φ and φ are the standard normal distribution and density functions, respectively. The function Z has previously been used in a similar context by Lefèvre and Utev (2003). Note that Lemma 5.1 of Lefèvre and Utev (2003) gives us that $\mathcal{D}^j Z(x) > 0$ for each $j \geq 0$, $x \in \mathbb{R}$. The properties of our function which we require are given by Lemma 2.1 below. Several of these use inductive proofs, in which the following easily verifiable expressions will be useful for establishing the base cases. Note that throughout we will take $\mathcal{D}^j Z(-x)$ to mean the function $\frac{d^j}{dt^j} Z(t)$ evaluated at $t = -x$.

$$\mathcal{D}Z(x) = 1 + xZ(x)\,, \tag{2.1}$$
$$\mathcal{D}^2 Z(x) = x + (1+x^2)Z(x)\,. \tag{2.2}$$

Also

$$Z(-x) = \frac{1}{\varphi(x)} - Z(x)\,, \tag{2.3}$$
$$\mathcal{D}Z(-x) = \mathcal{D}Z(x) - \frac{x}{\varphi(x)}\,, \tag{2.4}$$
$$\mathcal{D}^2 Z(-x) = \frac{(1+x^2)}{\varphi(x)} - \mathcal{D}^2 Z(x)\,. \tag{2.5}$$

Lemma 2.1. *Let $k \geq 0$. For all $x \in \mathbb{R}$*

i.
$$\mathcal{D}^{k+2} Z(x) = (k+1)\mathcal{D}^k Z(x) + x\mathcal{D}^{k+1} Z(x)\,,$$

ii.
$$\varphi(x)\Big[\mathcal{D}^{k+1} Z(-x)\mathcal{D}^k Z(x) + \mathcal{D}^{k+1} Z(x)\mathcal{D}^k Z(-x)\Big] = k!\,,$$

iii.
$$\varphi(x)\Big[\mathcal{D}^{k+2} Z(x)\mathcal{D}^k Z(-x) - \mathcal{D}^{k+2} Z(-x)\mathcal{D}^k Z(x)\Big] = (k!)x\,,$$

iv.

$$\alpha_k(x) := \int_{-\infty}^{x} \varphi(t)\mathcal{D}^k Z(t)\,dt = \frac{1}{k+1}\varphi(x)\mathcal{D}^{k+1}Z(x)\,.$$

Proof. (i) and (ii) are straightforward. The case $k = 0$ is established using (2.1)–(2.4). A simple induction argument completes the proof. (iii) follows directly from (i) and (ii).

For (iv) we note firstly that using Lemma 5.1 of Lefèvre and Utev (2003) the required integrals exist. Now, for the case $k = 0$,

$$\int_{-\infty}^{x} \Phi(t)\,dt = \int_{-\infty}^{x} (x-t)\varphi(t)\,dt = \varphi(x)\mathcal{D}Z(x)\,,$$

by (2.1). For the inductive step let $k \geq 1$ and assume α_{k-1} has the required form. Integrating by parts,

$$\alpha_k(x) = \varphi(x)\mathcal{D}^{k-1}Z(x) + \int_{-\infty}^{x} t\varphi(t)\mathcal{D}^{k-1}Z(t)\,dt\,.$$

Integrating by parts again, and using the inductive hypothesis, we get

$$\alpha_k(x) = \varphi(x)\mathcal{D}^{k-1}Z(x) + \frac{x}{k}\varphi(x)\mathcal{D}^k Z(x) - \frac{1}{k}\alpha_k(x)\,.$$

Rearranging and applying (i) gives us the result. □

We are now in a position to establish the key representation of $\mathcal{D}^{k+2}Sh$, with S given by (1.1).

Lemma 2.2. *Let $k \geq 0$ and let $h : \mathbb{R} \mapsto \mathbb{R}$ be $(k+1)$-times differentiable with $\mathcal{D}^k h$ absolutely continuous. Then for all $x \in \mathbb{R}$*

$$\mathcal{D}^{k+2}Sh(x) = \mathcal{D}^{k+1}h(x) - \frac{\mathcal{D}^{k+2}Z(-x)}{k!}\gamma_k(x) - \frac{\mathcal{D}^{k+2}Z(x)}{k!}\delta_k(x)\,,$$

where

$$\begin{aligned} \gamma_k(x) &:= \int_{-\infty}^{x} \mathcal{D}^{k+1}h(t)\varphi(t)\mathcal{D}^k Z(t)\,dt\,, \\ \delta_k(x) &:= \int_{x}^{\infty} \mathcal{D}^{k+1}h(t)\varphi(t)\mathcal{D}^k Z(-t)\,dt\,. \end{aligned}$$

Proof. Again, we proceed by induction on k. The case $k = 0$ was established by Stein (1986, (58) on page 27). The required form can be seen using (2.2) and (2.5). Now let $k \geq 1$. We assume firstly that h satisfies the additional restriction that $\mathcal{D}^k h(0) = 0$. Using this and the absolute continuity of $\mathcal{D}^k h$ we may write, for $t \in \mathbb{R}$,

$$\mathcal{D}^k h(t) = \begin{cases} \int_0^t \mathcal{D}^{k+1}h(y)\,dy & \text{if } t \geq 0 \\ -\int_t^0 \mathcal{D}^{k+1}h(y)\,dy & \text{if } t < 0 \end{cases}$$

We firstly consider the case $x \geq 0$. Using the above, and interchanging the order of integration, we get that

$$\gamma_{k-1}(x) = \int_0^x \mathcal{D}^{k+1}h(y)\left(\int_y^x \varphi(t)\mathcal{D}^{k-1}Z(t)\,dt\right)dy \\ - \int_{-\infty}^0 \mathcal{D}^{k+1}h(y)\left(\int_{-\infty}^y \varphi(y)\mathcal{D}^{k-1}Z(t)\,dt\right)dy\,.$$

Applying Lemma 2.1(iv) we obtain

$$\gamma_{k-1}(x) = \frac{1}{k}\Big[\mathcal{D}^k h(x)\varphi(x)\mathcal{D}^k Z(x) - \gamma_k(x)\Big]\,. \tag{2.6}$$

In a similar way we get

$$\delta_{k-1}(x) = \frac{1}{k}\Big[\mathcal{D}^k h(x)\varphi(x)\mathcal{D}^k Z(-x) + \delta_k(x)\Big]\,. \tag{2.7}$$

In the case $x < 0$ a similar argument also yields (2.6) and (2.7). Now, by the inductive hypothesis we assume the required form for $\mathcal{D}^{k+1}Sh$. Differentiating this we get

$$\mathcal{D}^{k+2}Sh(x) = \mathcal{D}^{k+1}h(x) + \frac{\mathcal{D}^{k+2}Z(-x)}{(k-1)!}\gamma_{k-1}(x) - \frac{\mathcal{D}^{k+2}Z(x)}{(k-1)!}\delta_{k-1}(x) \\ + \frac{\mathcal{D}^k h(x)\varphi(x)}{(k-1)!}\Big[\mathcal{D}^{k+1}Z(x)\mathcal{D}^{k-1}Z(-x) - \mathcal{D}^{k+1}Z(-x)\mathcal{D}^{k-1}Z(x)\Big]\,.$$

Using (2.6) and (2.7) in the above we obtain the desired representation along with the additional term

$$\frac{\mathcal{D}^k h(x)\varphi(x)}{(k-1)!}\Big[\mathcal{D}^{k+1}Z(x)\mathcal{D}^{k-1}Z(-x) - \mathcal{D}^{k+1}Z(-x)\mathcal{D}^{k-1}Z(x)\Big] \\ + \frac{\mathcal{D}^k h(x)\varphi(x)}{k!}\Big[\mathcal{D}^{k+2}Z(-x)\mathcal{D}^k Z(x) - \mathcal{D}^{k+2}Z(x)\mathcal{D}^k Z(-x)\Big]\,,$$

which is zero by Lemma 2.1(iii).

The proof is completed by removing the condition that $\mathcal{D}^k h(0) = 0$. We do this by applying our result to $g(x) := h(x) - \frac{B}{k!}x^k$, where $B := \mathcal{D}^k h(0)$. Clearly $\mathcal{D}^{k+1}g = \mathcal{D}^{k+1}h$. Also, by the linearity of S, $(Sg)(x) = (Sh)(x) - \frac{B}{k!}(Sp_k)(x)$, where $p_k(x) = x^k$. Finally, it is easily verified that Sp_k is a polynomial of degree $k-1$ and thus $\mathcal{D}^{k+2}Sp_k \equiv 0$. Hence $\mathcal{D}^{k+2}Sg = \mathcal{D}^{k+2}Sh$. □

We now use the representation established in Lemma 2.2 to prove Theorem 1.1. Fix $k \geq 0$ and let h be as in Lemma 2.2, with the additional assumption that $\mathcal{D}^{k+1}h \geq 0$. Since $\varphi(x)$, $\mathcal{D}^j Z(x) > 0$ for each $j \geq 0$, $x \in \mathbb{R}$, we get that for all $x \in \mathbb{R}$

$$|\mathcal{D}^{k+2}Sh(x)| \leq \left|\mathcal{D}^{k+1}h(x) - \frac{\|\mathcal{D}^{k+1}h\|_\infty}{k!}\rho_k(x)\right|\,, \tag{2.8}$$

where

$$\rho_k(x) := \mathcal{D}^{k+2}Z(-x)\int_{-\infty}^x \varphi(t)\mathcal{D}^k Z(t)\,dt + \mathcal{D}^{k+2}Z(x)\int_x^\infty \varphi(t)\mathcal{D}^k Z(-t)\,dt\,.$$

Now, $\int_x^\infty \varphi(t)\mathcal{D}^k Z(-t)\,dt = \alpha_k(-x)$ and so

$$\rho_k(x) = \mathcal{D}^{k+2}Z(-x)\alpha_k(x) + \mathcal{D}^{k+2}Z(x)\alpha_k(-x)\,.$$

Applying Lemma 2.1(iv) and (ii) we get that $\rho_k(x) = k!$ for all x. Combining this with (2.8) we obtain

$$\begin{aligned} \left|\mathcal{D}^{k+2}Sh(x)\right| &\leq \left|\mathcal{D}^{k+1}h(x) - \|\mathcal{D}^{k+1}h\|_\infty\right| \\ &\leq \max\{\mathcal{D}^{k+1}h(x), \|\mathcal{D}^{k+1}h\|_\infty\} = \|\mathcal{D}^{k+1}h\|_\infty\,, \end{aligned} \tag{2.9}$$

which gives us that $\|\mathcal{D}^{k+2}Sh\|_\infty \leq \|\mathcal{D}^{k+1}h\|_\infty$ for all such h.

We now remove the assumption that $\mathcal{D}^{k+1}h \geq 0$. We use a method analogous to that in the last part of the proof of Lemma 2.2. Consider the function $H : \mathbb{R} \mapsto \mathbb{R}$ given by $H(x) := h(x) - \frac{C}{(k+1)!}x^{k+1}$, where $C := \inf_t \mathcal{D}^{k+1}h(t)$. As before, $\mathcal{D}^{k+2}SH = \mathcal{D}^{k+2}Sh$. Clearly $\mathcal{D}^{k+1}H(x) = \mathcal{D}^{k+1}h(x) - C \geq 0$. Then, by (2.9),

$$\|\mathcal{D}^{k+2}Sh\|_\infty \leq \sup_t \left|\mathcal{D}^{k+1}h(t) - C\right| = \sup_t \mathcal{D}^{k+1}h(t) - \inf_t \mathcal{D}^{k+1}h(t)\,,$$

which gives Theorem 1.1. The proofs of Theorems 1.2, 1.3 and 1.4 below are analogous to our proof of Theorem 1.1.

Remark. We note that the bound we have established in Theorem 1.1 is sharp. Let $a > 0$ and suppose h_a is the function with $\mathcal{D}^{k+1}h_a$ continuous such that $\mathcal{D}^{k+1}h_a(x) = 1$ for $|x| > a$ and $\mathcal{D}^{k+1}h_a(0) = -1$. Using Lemma 2.2 we get

$$\begin{aligned} \mathcal{D}^{k+2}Sh_a(0) = \frac{\mathcal{D}^{k+2}Z(0)}{k!}\Bigg[\int_{-a}^0 &\left(1 - \mathcal{D}^{k+1}h_a(t)\right)\varphi(t)\mathcal{D}^k Z(t)\,dt \\ &+ \int_0^a \left(1 - \mathcal{D}^{k+1}h_a(t)\right)\varphi(t)\mathcal{D}^k Z(-t)\,dt\Bigg] - 1 - \frac{1}{k!}\rho_k(0)\,. \end{aligned}$$

Since each of the integrands in the above are bounded, letting $a \to 0$ gives us that $\mathcal{D}^{k+2}Sh_a(0) \to -1 - \frac{1}{k!}\rho_k(0) = -2$.

Proof of Theorem 1.2

We suppose now that $Y \sim \mathrm{Exp}(\lambda)$, the exponential distribution with mean λ^{-1}. It can easily be checked that a Stein equation for Y is given by

$$\mathcal{D}Sh(x) = \lambda Sh(x) + h(x) - \Phi_h\,, \tag{2.10}$$

for all $x \geq 0$, where $\Phi_h := E[h(Y)]$. The corresponding Stein operator is given by (1.2). We establish an analogue of Lemma 2.2 in this case.

Lemma 2.3. *Let $k \geq 0$ and let $h : \mathbb{R}^+ \mapsto \mathbb{R}$ be $(k+1)$-times differentiable with $\mathcal{D}^k h$ absolutely continuous. Then for all $x \geq 0$*

$$\mathcal{D}^{k+2}Sh(x) = \mathcal{D}^{k+1}h(x) - \lambda e^{\lambda x}\int_x^\infty \mathcal{D}^{k+1}h(t)e^{-\lambda t}\,dt\,.$$

Proof. We proceed by induction on k. For $k=0$ we follow the argument of Stein (1986, Lemma II.3) used to establish (1.5). From (2.10) we have that

$$\mathcal{D}^2 Sh(x) = \lambda^2 Sh(x) + \lambda[h(x) - \Phi_h] + \mathcal{D}h(x)\,, \tag{2.11}$$

for all $x \geq 0$. Now, we write

$$h(x) - \Phi_h = \int_0^x [h(x) - h(t)] f(t)\, dt - \int_x^\infty [h(t) - h(x)] f(t)\, dt\,,$$

where f is the density function of Y. Using the absolute continuity of h to write, for example, $h(x) - h(t) = \int_t^x \mathcal{D}h(y)\, dy$ and interchanging the order of integration we obtain

$$h(x) - \Phi_h = \int_0^x \mathcal{D}h(y) F(y)\, dy - \int_x^\infty \mathcal{D}h(y)[1 - F(y)]\, dy\,, \tag{2.12}$$

where F is the distribution function of Y. We substitute (2.12) into (1.2), interchange the order of integration and rearrange to get

$$\begin{aligned} Sh(x) = -\frac{1}{f(x)}\Bigg[(1 - F(x)) \int_0^x \mathcal{D}h(y) F(y)\, dy \\ + F(x) \int_x^\infty \mathcal{D}h(y)(1 - F(y))\, dy\Bigg]. \end{aligned}$$

Combining this with (2.11) and (2.12) we establish our lemma for $k=0$. Now let $k \geq 1$ and assume for now that h also satisfies $\mathcal{D}^k h(0) = 0$. We proceed as in the proof of Lemma 2.2. We use the absolute continuity of $\mathcal{D}^k h$ to write $\mathcal{D}^k h(t) = \int_0^t \mathcal{D}^{k+1} h(y)\, dy$ and hence show that

$$\int_x^\infty \mathcal{D}^k h(t) e^{-\lambda t}\, dt = \frac{1}{\lambda} \mathcal{D}^k h(x) e^{-\lambda x} + \frac{1}{\lambda} \int_x^\infty \mathcal{D}^{k+1} h(y) e^{-\lambda y}\, dy\,.$$

Using the above with our inductive hypothesis we obtain the required representation. The restriction that $\mathcal{D}^k h(0) = 0$ is removed as in the proof of Lemma 2.2, noting that S applied to a polynomial of degree k returns a polynomial of degree k in this case. $\square$

Suppose h is as in the statement of Theorem 1.2, with the additional condition that $\mathcal{D}^{k+1} h \geq 0$. Noting that $\lambda e^{\lambda x} \int_x^\infty e^{-\lambda t}\, dt = 1$, we use Lemma 2.3 to obtain

$$\left|\mathcal{D}^{k+2} Sh(x)\right| \leq \left|\mathcal{D}^{k+1} h(x) - \|\mathcal{D}^{k+1} h\|_\infty\right| \leq \|\mathcal{D}^{k+1} h\|_\infty\,,$$

for $k \geq 0$. The restriction that $\mathcal{D}^{k+1} h \geq 0$ is lifted and the proof completed as in the proof of Theorem 1.1.

Proof of Theorems 1.3 and 1.4

Suppose we have a birth-death process on $\mathbb{Z}^+$ with constant birth rate λ and death rates μ_k, with $\mu_0 = 0$. Let this have an equilibrium distribution π with $P(\pi = k) = \pi_k = \pi_0 \lambda^k \left(\prod_{i=1}^k \mu_i\right)^{-1}$ and $F(k) := \sum_{i=0}^k \pi_i$. It is well-known that in this case a Stein equation is given by

$$\Delta Sh(k) = \frac{1}{\lambda}\left[(\mu_k - \lambda) Sh(k) + h(k) - \Phi_h\right]\,, \tag{2.13}$$

for $k \geq 0$, where $\Phi_h := E[h(\pi)]$ and

$$Sh(k) := \frac{1}{\lambda \pi_{k-1}} \sum_{i=0}^{k-1} \big[h(i) - \Phi_h\big]\pi_i \,, \tag{2.14}$$

for $k \geq 1$. See, for example, Brown and Xia (2001) or Holmes (2004). $Sh(0)$ is not defined by (2.13), and we leave this undefined for now. We consider particular choices later. With appropriate choices of λ and μ_k our Stein operator (2.14) gives us (1.3) and (1.4). We define Z_1^* and Z_2^* analogously to our function Z in the proof of Theorem 1.1. Let

$$Z_1^*(k) := \frac{F(k-1)}{\pi_{k-1}}\,, \qquad Z_2^*(k) := \frac{1-F(k-1)}{\pi_{k-1}}\,,$$

for $k \geq 1$. We note that for any functions f and g on $\mathbb{Z}^+$ we have

$$\Delta(fg)(k) = \Delta f(k) g(k) + f(k+1)\Delta g(k)\,. \tag{2.15}$$

The following easily verifiable identities will prove useful.

$$\Delta Z_1^*(k) = \frac{(\mu_k - \lambda)}{\lambda} Z_1^*(k) + 1\,, \tag{2.16}$$

$$\Delta^2 Z_1^*(k) = \left[\frac{(\mu_{k+1}-\lambda)(\mu_k - \lambda)}{\lambda^2} + \frac{(\mu_{k+1}-\mu_k)}{\lambda}\right] Z_1^*(k) + \frac{(\mu_{k+1}-\lambda)}{\lambda}\,, \tag{2.17}$$

$$\Delta Z_2^*(k) = \frac{(\mu_k - \lambda)}{\lambda} Z_2^*(k) - 1\,, \tag{2.18}$$

$$\Delta^2 Z_2^*(k) = \left[\frac{(\mu_{k+1}-\lambda)(\mu_k - \lambda)}{\lambda^2} + \frac{(\mu_{k+1}-\mu_k)}{\lambda}\right] Z_2^*(k) - \frac{(\mu_{k+1}-\lambda)}{\lambda}\,, \tag{2.19}$$

for $k \geq 1$. Now, we follow the proof of Lemma 2.3 and use (2.13) to get that

$$\Delta^2 Sh(k) = \frac{1}{\lambda}\Delta h(k) + Sh(k)\left[\frac{(\mu_{k+1}-\lambda)(\mu_k - \lambda)}{\lambda^2} + \frac{(\mu_{k+1}-\mu_k)}{\lambda}\right] + \frac{(\mu_{k+1}-\lambda)}{\lambda^2}\big[h(k) - \Phi_h\big], \tag{2.20}$$

for $k \geq 0$. We obtain the discrete analogue of (2.12), for $h : \mathbb{Z}^+ \mapsto \mathbb{R}$ bounded and $k \geq 0$,

$$h(k) - \Phi_h = \sum_{l=0}^{k-1} \Delta h(l) F(l) - \sum_{l=k}^{\infty} \Delta h(l)[1 - F(l)]\,, \tag{2.21}$$

and combining this with (2.14) we get that

$$Sh(k) = -\frac{1}{\lambda \pi_{k-1}}\Big[[1-F(k-1)]\sum_{l=0}^{k-1} \Delta h(l) F(l) + F(k-1)\sum_{l=k}^{\infty} \Delta h(l)[1-F(l)]\Big], \tag{2.22}$$

for $k \geq 1$. Now, combining this with (2.17), (2.19), (2.20) and (2.21) we get

$$\Delta^2 Sh(k) = \frac{1}{\lambda}\Delta h(k) - \frac{1}{\lambda}\Delta^2 Z_2^*(k)\sum_{i=0}^{k-1}\Delta h(i)F(i)$$

$$- \frac{1}{\lambda}\Delta^2 Z_1^*(k)\sum_{i=k}^{\infty}\Delta h(i)(1-F(i))\,, \quad (2.23)$$

for $k \geq 1$. To prove our Theorems 1.3 and 1.4 we first generalise (2.23) in both the geometric and Poisson cases.

The geometric case

Suppose $\mu_k = 1$ for all $k \geq 1$ and $\lambda = q = 1 - p$. Then $\pi_k = pq^k$ for $k \geq 0$ and $\pi \sim \mathrm{Geom}(p)$, the geometric distribution starting at zero. We now define $Sh(0) = 0$. In this case we have the following representation.

Lemma 2.4. *Let $j \geq 0$. For $h : \mathbb{Z}^+ \mapsto \mathbb{R}$ bounded we have*

$$\Delta^{j+2} Sh(k) = \frac{1}{q}\Delta^{j+1}h(k) - \frac{p}{q^{k+1}}\sum_{i=k}^{\infty}\Delta^{j+1}h(i)q^i\,,$$

for all $k \geq 0$.

Proof. We proceed by induction on j. It is easily verified that the case $j = 0$ is given by (2.23) for $k \geq 1$. For $k = 0$ we can combine (2.20) and (2.21) to get our result. Now let $j \geq 1$, and assume additionally that $\Delta^j h(0) = 0$. Then, writing $\Delta^j h(i) = \sum_{l=0}^{i-1}\Delta^{j+1}h(l)$ it can be shown that

$$\sum_{i=k}^{\infty}\Delta^j h(i)q^i = \frac{q^k}{p}\Delta^j h(k) + \frac{q}{p}\sum_{l=k}^{\infty}\Delta^{j+1}h(l)q^l\,. \quad (2.24)$$

Using (2.15) together with (2.24) and our representation of $\Delta^{j+1}Sh$ we can show the desired result. Finally, we remove the condition that $\Delta^j h(0) = 0$ as in the final part of the proof of Lemma 2.2, noting that S applied to a polynomial of degree j gives a polynomial of degree j in this case. □

We now complete the proof of Theorem 1.4. Let $h : \mathbb{Z}^+ \mapsto \mathbb{R}$ be bounded with $\Delta^{j+1}h \geq 0$. Since $\frac{p}{q^k}\sum_{i=k}^{\infty} q^i = 1$, we can use Lemma 2.4 to obtain

$$|\Delta^{j+2}Sh(k)| \leq \frac{1}{q}\left|\Delta^{j+1}h(k) - \|\Delta^{j+1}h\|_\infty\right| \leq \frac{1}{q}\|\Delta^{j+1}h\|_\infty\,,$$

for all $j \geq 0$. The restriction that $\Delta^{j+1}h \geq 0$ is removed and the proof completed as in the final part of the proof of Theorem 1.1.

The Poisson case

We turn our attention now to the case where $\mu_k = k$, so that $\pi \sim \mathrm{Pois}(\lambda)$. We begin with some properties of our functions Z_1^* and Z_2^* in this case. Lemma 2.6 gives us an analogue to Lemma 2.1 for the Poisson distribution.

Lemma 2.5. *Let $j \geq 0$. For all $k \geq 1$*

i.

$$\Delta^j Z_1^*(k) \geq 0$$

ii.

$$\Delta^j Z_2^*(k) \begin{cases} \geq 0 & \text{if } j \text{ is even} \\ \leq 0 & \text{if } j \text{ is odd} \end{cases}$$

Proof. We note that $\frac{d}{d\lambda} P(\pi \leq k) = -\pi_k$ and hence

$$P(\pi \leq k) = \int_\lambda^\infty \frac{e^{-v} v^k}{k!}\, dv = 1 - \int_0^\lambda \frac{e^{-v} v^k}{k!}\, dv\,.$$

So, we can write

$$Z_1^*(k) = e^\lambda \int_\lambda^\infty e^{-v} \left(\frac{v}{\lambda}\right)^{k-1} dv\,, \qquad Z_2^*(k) = e^\lambda \int_0^\lambda e^{-v} \left(\frac{v}{\lambda}\right)^{k-1} dv\,,$$

which implies our result. □

Lemma 2.6. *Let $j \geq 0$. For all $k \geq 1$ and $l \in \{1, 2\}$*

i.

$$(j+1)\Delta^j Z_l^*(k) + (k+j+1-\lambda)\Delta^{j+1} Z_l^*(k) = \lambda \Delta^{j+2} Z_l^*(k)\,,$$

ii.

$$\pi_{k-1}\big[\Delta^{j+1} Z_2^*(k) \Delta^j Z_1^*(k) - \Delta^{j+1} Z_1^*(k) \Delta^j Z_2^*(k)\big] = \frac{(-1)^{j-1} j!}{\lambda^j}\,,$$

iii.

$$\pi_{k-1}\big[\Delta^{j+2} Z_2^*(k) \Delta^j Z_1^*(k) - \Delta^{j+2} Z_1^*(k) \Delta^j Z_2^*(k)\big] = \frac{(-1)^{j-1}(k+j+1-\lambda) j!}{\lambda^{j+1}}\,,$$

iv.

$$\alpha_j^*(k) := \sum_{i=0}^{k-1} \pi_i \Delta^j Z_1^*(i+1) = \frac{\lambda}{j+1} \pi_{k-1} \Delta^{j+1} Z_1^*(k)\,,$$

v.

$$\beta_j^*(k) := \sum_{i=k}^{\infty} \pi_i \Delta^j Z_2^*(i+1) = -\frac{\lambda}{j+1} \pi_{k-1} \Delta^{j+1} Z_2^*(k)\,.$$

Proof. (i) and (ii) are proved by a simple induction argument. The case $j = 0$ follows from (2.16)–(2.19). (iii) follows directly from (i) and (ii).

To prove (iv) we again use induction on j. For the case $j = 0$ we check the result directly for $k = 1$. For $k \geq 2$ note that $\sum_{i=0}^{k-1} F(i) = kF(k-1) - \lambda F(k-2)$ and use (2.16). For the inductive step we will use that for all functions f and g on $\mathbb{Z}^+$

$$\sum_{i=m}^{n-1} f(i+1)\Delta g(i) = f(n)g(n) - f(m)g(m) - \sum_{i=m}^{n-1} \Delta f(i)g(i)\,. \tag{2.25}$$

Since $\Delta\pi_{i-1} = \frac{(\lambda-i)}{\lambda}\pi_i$ we apply (2.25) to $\alpha_j^*(k)$ to get

$$\alpha_j^*(k) = \pi_{k-1}\Delta^{j-1}Z_1^*(k+1) - \pi_0\Delta^{j-1}Z_1^*(1) - \frac{1}{\lambda}\sum_{i=1}^{k-1}(\lambda-i)\pi_i\Delta^{j-1}Z_1^*(i+1)\,.$$

Applying (2.25) once more and using our representation for α_{j-1}^* we obtain

$$\alpha_j^*(k) = \pi_{k-1}\Delta^{j-1}Z_1^*(k) + \frac{(k+j-\lambda)}{j}\pi_{k-1}\Delta^j Z_1^*(k) - \frac{1}{j}\alpha_j^*(k)\,.$$

Rearranging and applying (i) completes the proof. (v) is proved analogously to (iv). $\Box$

We may now prove our key representation in the Poisson case.

Lemma 2.7. *Let $j \geq 0$. For $h : \mathbb{Z}^+ \mapsto \mathbb{R}$ bounded we have*

$$\begin{aligned}\Delta^{j+2}Sh(k) = \frac{1}{\lambda}\Delta^{j+1}h(k) + {}& \\ (-1)^{j+1}\frac{\lambda^{j-1}}{j!}\Bigg[&\Delta^{j+2}Z_2^*(k)\sum_{i=0}^{k-1}\Delta^{j+1}h(i)\pi_i\Delta^j Z_1^*(i+1) \\ &+ \Delta^{j+2}Z_1^*(k)\sum_{i=k}^{\infty}\Delta^{j+1}h(i)\pi_i\Delta^j Z_2^*(i+1)\Bigg]\,,\end{aligned}$$

for all $k \geq 1$.

Proof. We proceed by induction on j. The case $j = 0$ is established by (2.23). Now, we let $j \geq 1$ and assume in addition that $\Delta^j h(0) = 0$. Hence, we write $\Delta^j h(i) = \sum_{l=0}^{i-1}\Delta^{j+1}h(l)$. Combining this with Lemma 2.6(iv) we obtain

$$\begin{aligned}\sum_{i=0}^{k}\Delta^j h(i)\pi_i\Delta^{j-1}Z_1^*(i+1) = {}& \frac{\lambda}{j}\Delta^j h(k)\pi_k\Delta^j Z_1^*(k+1) \\ &- \frac{\lambda}{j}\sum_{l=0}^{k-1}\Delta^{j+1}h(l)\pi_l\Delta^j Z_1^*(l+1)\,.\end{aligned} \tag{2.26}$$

By a similar argument employing Lemma 2.6(v) we get

$$\begin{aligned}\sum_{i=k+1}^{\infty}\Delta^j h(i)\pi_i\Delta^{j-1}Z_2^*(i+1) = {}& -\frac{\lambda}{j}\pi_k\Delta^j Z_2^*(k+1)\Delta^j h(k) \\ &- \frac{\lambda}{j}\sum_{l=k}^{\infty}\Delta^{j+1}h(l)\pi_l\Delta^j Z_2^*(l+1)\,.\end{aligned} \tag{2.27}$$

Now, our inductive hypothesis gives us our representation of $\Delta^{j+1}Sh$. Using (2.15) to take forward differences and employing (2.26) and (2.27) we obtain the required representation along with the additional terms

$$\Delta^j h(k)\pi_k\left[\Delta^{j+1}Z_2^*(k)\Delta^{j-1}Z_1^*(k+1)-\Delta^{j+1}Z_1^*(k)\Delta^{j-1}Z_2^*(k+1)\right] \\ +\frac{\lambda}{j}\Delta^j h(k)\pi_k\left[\Delta^{j+2}Z_2^*(k)\Delta^j Z_1^*(k+1)-\Delta^{j+2}Z_1^*(k)\Delta^j Z_2^*(k+1)\right]. \quad (2.28)$$

Writing, for example $\Delta^j Z_1^*(k+1)=\Delta^{j+1}Z_1^*(k)+\Delta^j Z_1^*(k)$, rearranging and applying Lemma 2.6(ii) and (iii) gives that (2.28) is zero for all $k\geq 1$, and hence our representation.

Finally, the restriction that $\Delta^j h(0)=0$ is lifted as in the proof of Lemma 2.2. It can easily be checked that in the Poisson case if $h(k)$ is a polynomial of degree j then $Sh(k)$ is a polynomial of degree $j-1$ for $k\geq 1$. □.

We now complete the proof of Theorem 1.3. Let $h:\mathbb{Z}^+\mapsto\mathbb{R}$ be bounded with $\Delta^{j+1}h\geq 0$. Then we can use Lemmas 2.5 and 2.7 to write

$$\Delta^{j+2}Sh(k)=\frac{1}{\lambda}\Delta^{j+1}h(k)-\frac{\lambda^{j-1}}{j!}\left(\left|\Delta^{j+2}Z_2^*(k)\right|\sum_{i=0}^{k-1}\Delta^{j+1}h(i)\pi_i\Delta^j Z_1^*(i+1)\right. \\ \left.+\Delta^{j+2}Z_1^*(k)\left|\sum_{i=k}^{\infty}\Delta^{j+1}h(i)\pi_i\Delta^j Z_2^*(i+1)\right|\right),$$

for $k\geq 1$. Hence

$$\left|\Delta^{j+2}Sh(k)\right|\leq\frac{1}{\lambda}\left|\Delta^{j+1}h(k)-\|\Delta^{j+1}h\|_\infty\frac{\lambda^j}{j!}\rho_j^*(k)\right|,$$

where

$$\rho_j^*(k):=\left|\Delta^{j+2}Z_2^*(k)\right|\sum_{i=0}^{k-1}\pi_i\Delta^j Z_1^*(i+1)+\Delta^{j+2}Z_1^*(k)\left|\sum_{i=k}^{\infty}\pi_i\Delta^j Z_2^*(i+1)\right|.$$

Applying Lemma 2.6(ii), (iv) and (v) we get that $\rho_j^*(k)=\frac{j!}{\lambda^j}$ for all $k\geq 1$, and so

$$\left|\Delta^{j+2}Sh(k)\right|\leq\frac{1}{\lambda}\left|\Delta^{j+1}h(k)-\|\Delta^{j+1}h\|_\infty\right|\leq\frac{1}{\lambda}\|\Delta^{j+1}h\|_\infty.$$

We remove our condition that $\Delta^{j+1}h\geq 0$ and complete the proof as in the standard normal case.

Remark. The value of $Sh(0)$ is not defined by the Stein equation (2.13) and so may be chosen arbitrarily. In the geometric case it was convenient to choose $Sh(0)=0$ so that the representation established in Lemma 2.4 holds at $k=0$. We now consider possible choices of $Sh(0)$ in the Poisson case. Common choices are $Sh(0)=0$, as in Barbour (1987) and Barbour *et al* (1992), and $Sh(0)=Sh(1)$, as in Barbour and Xia (2006). However, with neither of these choices can we use the above methods to obtain a representation directly analogous to our Lemma 2.7 for $k=0$ and all bounded h. Our proof relies on the fact that if $h(k)=k^j$ for a fixed $j\geq 1$ and all $k\geq 0$ then Sh is a polynomial of degree $j-1$ for $k\geq 1$. Taking $h(k)=k^2$, for example, shows that with neither of the choices of $Sh(0)$ outlined above is Sh a polynomial for all $k \geq 0$.

Despite these limitations, there are some cases where useful bounds can be obtained. Suppose we choose $Sh(0) = Sh(1)$, so that $\Delta^2 Sh(0) = \Delta Sh(1)$. Using (2.13), (2.21) and (2.22) we can write

$$\Delta^2 Sh(0) = \frac{1}{\lambda}\Delta h(0) - \frac{1}{\lambda^2}\sum_{l=0}^{\infty}\Delta h(l)[1-F(l)]\,.$$

Assuming $\Delta h \geq 0$, we can then proceed as in the proof of Theorem 1.3 and use Lemma 2.6(v) to get that

$$|\Delta^2 Sh(0)| \leq \frac{1}{\lambda}\|\Delta h\|_\infty\,. \tag{2.29}$$

With our choice of $Sh(0)$ here we are able to remove the condition that $\Delta h \geq 0$. Setting $H(k) = h(k) - g(k)$ for $k \geq 0$, where $g(k) := Ck$ and $C := \inf_{i\geq 0}\Delta h(i)$, we have $\Delta H(k) = \Delta h(k) - C \geq 0$. It can easily be checked that $Sg(k) = -C$ for all $k \geq 0$ and so $\Delta^2 SH = \Delta^2 Sh$. Applying (2.29) to H and combining with Theorem 1.3 we have

$$\|\Delta^2 Sh\|_\infty \leq \frac{1}{\lambda}\left(\sup_{i\geq 0}\Delta h(i) - \inf_{i\geq 0}\Delta h(i)\right) \leq \frac{2}{\lambda}\|\Delta h\|_\infty\,,$$

and so we obtain an analogous result to that of Barbour and Xia (2006, Theorem 1.1). We note that this argument is heavily dependent on our choice of $Sh(0)$.

Of course, if we wish to estimate $\|\Delta^{j+2}Sh\|_\infty$ for a single, fixed $j \geq 0$ when $Sh(0)$ may be chosen arbitrarily, we can always choose such that $\Delta^{j+2}Sh(0) = 0$ and apply Theorem 1.3.

Proof of Corollaries 1.7 and 1.8

We let $Y \sim \mathrm{N}(0,1)$ and $\Phi_h := E[h(Y)]$. We begin by establishing the bound in Corollary 1.8. Using (4.1) of Lefèvre and Utev (2003) gives us that

$$E[h(W)] - \Phi_h = -E\sum_{j=1}^{n}\int_0^1 \mathcal{D}^2 Sh(W - tX_j)P_2(X_j,t)\,dt\,, \tag{2.30}$$

where $P_k(X,t) := X^{k+1}\frac{t^{k-1}}{(k-1)!} - \sigma_j^2 X^{k-1}\frac{t^{k-2}}{(k-2)!}$ and S is given by (1.1).

Now, integrating by parts we get

$$\begin{aligned} E[h(W)] - \Phi_h = -\frac{1}{2}\sum_{j=1}^{n} E\left[\mathcal{D}^2 Sh(W - X_j)\right] E\left[X_j^3\right] \\ - E\sum_{j=1}^{n}\int_0^1 \mathcal{D}^3 Sh(W - tX_j)P_3(X_j,t)\,dt\,, \end{aligned} \tag{2.31}$$

by independence and since $E[X_j] = 0$ for each j. Now, define $X_{i,j} := \sqrt{\frac{1}{1-\sigma_j^2}}X_i$ and let $T_{n,j} := \sum_{\substack{i=1\\i\neq j}}^{n} X_{i,j}$. Then $W - X_j = \left(\sqrt{1-\sigma_j^2}\right)T_{n,j}$ and we may write $\mathcal{D}^2 Sh(W - X_j) = G_j(T_{n,j})$, where $G_j(x) := \mathcal{D}^2 Sh\left(\left[\sqrt{1-\sigma_j^2}\right] x\right)$. We apply (2.30) to $G_j(T_{n,j})$ for each j and combine this with (2.31) to obtain

$$E\left[h(W)\right] - \Phi_h = \frac{1}{2}\sum_{j=1}^{n} E[X_j^3] E \sum_{\substack{i=1\\ i\neq j}}^{n} \int_0^1 \mathcal{D}^2 SG_j(T_{n,j} - tX_{i,j}) P_2(X_{i,j}, t)\, dt$$
$$- E\sum_{j=1}^{n}\int_0^1 \mathcal{D}^3 Sh(W - tX_j) P_3(X_j, t)\, dt - \frac{1}{2}\sum_{j=1}^{n} E\left[X_j^3\right] E\left[\mathcal{D}^2 Sh(Y_j)\right] .$$

We bound the above as in Lefèvre and Utev (2003, page 361) and as was modified to give Corollary 1.7 above. We can then write

$$\left|E[h(W)] - \Phi_h + \frac{1}{2}\sum_{j=1}^{n} E\left[X_j^3\right] E\left[\mathcal{D}^2 Sh(Y_j)\right]\right|$$
$$\leq \frac{p_2}{2}\sum_{j=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n} \|\mathcal{D}^2 SG_j\|_\infty E|X_j^3| E|X_{i,j}^3| + p_3 \|\mathcal{D}^3 Sh\|_\infty L_{n,4} ,$$

from which our bound follows using Theorem 1.1.

It remains now only to establish the value of the universal constant p_3. We already have, from Lefèvre and Utev (2003, Proposition 4.1), that $p_3 \leq \frac{1}{2}$. Now, by Lefèvre and Utev's definition,

$$p_3 = \sup_X \left\{ E\left[\int_0^1 \frac{|\frac{1}{2}X^4 t^2 - \mathrm{Var}(X) X^2 t|}{E[X^4]}\, dt\right] : E[X] = 0 \right\} .$$

Without loss we may restrict the supremum to be taken over random variables X with $\mathrm{Var}(X) = 1$. We further restrict our attention to nonnegative random variables, since the transformation $X \mapsto |X|$ leaves invariant the function over which the supremum is taken. We must then remove our previous conditions imposed on $E[X]$. Hence we obtain the representation of p_3 which we employ.

$$\begin{aligned} p_3 &= \sup_{X\geq 0}\left\{U(X) : E[X^2] = 1\right\} \\ &= \sup_{X\geq 0}\left\{ E\left[\int_0^1 \frac{|\frac{1}{2}X^4 t^2 - X^2 t|}{E[X^4]}\, dt\right] : E[X^2] = 1 \right\} . \end{aligned}$$

Now, $U(X)$ is continuous and weakly convex. That is, for all $\lambda \in [0,1]$, $U(\lambda X + (1-\lambda)Y) \leq \max\{U(X), U(Y)\}$. Thus, by a result of Hoeffding (1955) we need only consider random variables with at most two points in their support. We suppose that X takes the value a with probability $p \in [0,1]$ and b with probability $1-p$ such that $0 \leq a \leq 1 \leq b < \infty$. We use our condition $E[X^2] = 1$ to obtain $p = \frac{b^2-1}{b^2-a^2}$. This allows us to express $U(X)$ in terms of a and b. Using elementary techniques we maximise the resulting expression to give $p_3 = \frac{1}{3}$.

Acknowledgements: The author wishes to thank both Sergey Utev and the referee for useful comments and suggestions. Much of this work was supported by the University of Nottingham.

References

[1] A. D. Barbour, Asymptotic expansions based on smooth functions in the central limit theorem, Probab. Theory Relat. Fields **72** (1986), no. 2, 289–303. MR0836279 (87k:60060) MR0836279

[2] A. D. Barbour, Asymptotic expansions in the Poisson limit theorem, Ann. Probab. **15** (1987), no. 2, 748–766. MR0885141 (88k:60036) MR0885141

[3] A. D. Barbour, Stein's method for diffusion approximations, Probab. Theory Related Fields **84** (1990), no. 3, 297–322. MR1035659 (91d:60081) MR1035659

[4] A. D. Barbour, L. Holst and S. Janson, *Poisson approximation*, Oxford Univ. Press, New York, 1992. MR1163825 (93g:60043) MR1163825

[5] A. D. Barbour and A. Xia, On Stein's factors for Poisson approximation in Wasserstein distance, Bernoulli **12** (2006), no. 6, 943–954. MR2274850 MR2274850

[6] E. Bolthausen, An estimate of the remainder in a combinatorial central limit theorem, Z. Wahrsch. Verw. Gebiete **66** (1984), no. 3, 379–386. MR0751577 (85j:60032) MR0751577

[7] T. C. Brown and A. Xia, Stein's method and birth-death processes, Ann. Probab. **29** (2001), no. 3, 1373–1403. MR1872746 (2002k:60039) MR1872746

[8] L. H. Y. Chen, Poisson approximation for dependent trials, Ann. Probability **3** (1975), no. 3, 534–545. MR0428387 (55 #1408) MR0428387

[9] L. H. Y. Chen and Q.-M. Shao, Stein's method for normal approximation, in *An introduction to Stein's method*, 1–59, Singapore Univ. Press, Singapore, 2005. MR2235448 MR2235448

[10] T. Erhardsson, Stein's method for Poisson and compound Poisson approximation, in *An introduction to Stein's method*, 61–113, Singapore Univ. Press, Singapore, 2005. MR2235449 MR2235449

[11] L. Goldstein and G. Reinert, Stein's method and the zero bias transformation with application to simple random sampling, Ann. Appl. Probab. **7** (1997), no. 4, 935–952. MR1484792 (99e:60059) MR1484792

[12] F. Götze, On the rate of convergence in the multivariate CLT, Ann. Probab. **19** (1991), no. 2, 724–739. MR1106283 (92g:60028) MR1106283

[13] S. T. Ho and L. H. Y. Chen, An L_p bound for the remainder in a combinatorial central limit theorem, Ann. Probability **6** (1978), no. 2, 231–249. MR0478291 (57 #17775) MR0478291

[14] W. Hoeffding, The extrema of the expected value of a function of independent random variables, Ann. Math. Statist. **26** (1955), 268–275. MR0070087 (16,1128g) MR0070087

[15] S. Holmes, Stein's method for birth and death chains, in *Stein's method: expository lectures and applications*, 45–67, Inst. Math. Statist., Beachwood, OH, 2004. MR2118602 MR2118602

[16] C. Lefèvre and S. Utev, Exact norms of a Stein-type operator and associated stochastic orderings, Probab. Theory Related Fields **127** (2003), no. 3, 353–366. MR2018920 (2004i:60028) MR2018920

[17] E. A. Peköz, Stein's method for geometric approximation, J. Appl. Probab. **33** (1996), no. 3, 707–713. MR1401468 (97g:60034) MR1401468

[18] M. J. Phillips and G. V. Weinberg, Non-uniform bounds for geometric approximation, Statist. Probab. Lett. **49** (2000), no. 3, 305–311. MR1794749 (2001i:60041) MR1794749

[19] G. Reinert, A weak law of large numbers for empirical measures via Stein's method, Ann. Probab. **23** (1995), no. 1, 334–354. MR1330773 (96e:60056) MR1330773

[20] G. Reinert, Three general approaches to Stein's method, in *An introduction to Stein's method*, 183–221, Singapore Univ. Press, Singapore, 2005. MR2235451 MR2235451

[21] Y. Rinott and V. Rotar, On Edgeworth expansions for dependency-neighborhoods chain structures and Stein's method, Probab. Theory Related Fields **126** (2003), no. 4, 528–570. MR2001197 (2004h:62030) MR2001197

[22] V. Rotar, Stein's method, Edgeworth's expansions and a formula of Barbour, in *Stein's method and applications*, 59–84, Singapore Univ. Press, Singapore, 2005. MR2201886 MR2201886

[23] C. Stein, A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, 583–602, Univ. California Press, Berkeley, Calif, 1972. MR0402873 (53 #6687) MR0402873

[24] C. Stein, *Approximate computation of expectations*, Inst. Math. Statist., Hayward, CA, 1986. MR0882007 (88j:60055) MR0882007

[25] A. Xia, Stein's method and Poisson process approximation, in *An introduction to Stein's method*, 115–181, Singapore Univ. Press, Singapore, 2005. MR2235450 MR2235450