J. Appl. Prob. **47,** 146–156 (2010)
Printed in England

STEIN'S METHOD FOR COMPOUND GEOMETRIC APPROXIMATION

FRASER DALY,* *Universität Zürich*

Abstract

We apply Stein's method for probabilistic approximation by a compound geometric distribution, with applications to Markov chain hitting times and sequence patterns. Bounds on our Stein operator are found using a complex analytical approach based on generating functions and Cauchy's formula.

Keywords: Stein's method; compound geometric approximation; Markov chain hitting times

2000 Mathematics Subject Classification: Primary 62E17
Secondary 60E15; 60J10

1. Introduction

Stein's method for probabilistic approximation has been greatly developed in recent years and applied in a wide variety of situations. See, for example, [1], [5], and [17] for further detail. In this work we apply Stein's techniques to calculate error bounds in approximations by compound geometric distributions. The compound geometric distribution generalises the geometric distribution, to which Stein's method has already been applied in [2], [13], and [14].

We wish to bound the error in approximating some (discrete) random variable W by a compound geometric random variable. We say that Y has a compound geometric distribution if

$$Y \stackrel{\mathrm{D}}{=} \sum_{i=1}^{N} X_i, \tag{1.1}$$

where $X_1, X_2, \ldots$ are independent and identically distributed (i.i.d.) positive integer-valued random variables and $N \sim \mathrm{Geom}(p)$ has a geometric distribution that satisfies $\mathrm{P}(N = k) = p(1-p)^k$ for $k \geq 0$. Here, '$\stackrel{\mathrm{D}}{=}$' denotes equality in distribution. Such distributions arise naturally in many situations. See [11] and the references therein.

We are mainly concerned with the approximation in the total variation distance, defined by

$$d_{\mathrm{TV}}(W, Y) = \sup_{B \subseteq \mathbb{Z}^+} |\mathrm{P}(W \in B) - \mathrm{P}(Y \in B)|,$$

although our results may also be used to give bounds in other probability metrics.

In Section 2 we outline how Stein's method may be applied in the compound geometric case, and state bounds on the resulting Stein operator which we will need in the work that follows. The proof of these bounds is deferred until Section 4. In Section 3 we consider some applications of our results. In particular, we derive in Section 3.1 a compound geometric approximation for Markov chain hitting times which generalises a result of [13].

Received 7 October 2008; revision received 2 July 2009.
* Postal address: Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.
Email address: fraser.daly@math.uzh.ch

2. The Stein equation

We consider the problem of approximating our random variable W by our compound geometric distribution Y defined by (1.1). We write $p = 1 - q$, $\mu_i = \mathrm{P}(X = i)$, and $q_i = q\mu_i$. Denote this distribution by $Y \sim \mathrm{CG}(p, \boldsymbol{\mu})$, where $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots)$.

Following Stein's method, we need a linear operator A such that

$$W \stackrel{\mathrm{D}}{=} Y \quad \Longleftrightarrow \quad \mathrm{E}[Ag(W)] = 0 \quad \text{for all } g \in \mathcal{F},$$

for some suitable class of functions $\mathcal{F}$. In this case we let $\mathcal{F}$ be the class of bounded functions $h\colon \mathbb{Z}^+ \mapsto \mathbb{R}$. We call A our *characterising operator*. When applying Stein's method to the compound geometric distribution, we can choose the characterising operator A defined by

$$Ag(k) = q\,\mathrm{E}[g(k + X)] - g(k). \tag{2.1}$$

This choice can be established in the same way as the characterising operator for the compound Poisson distribution considered in [4]. We denote by S the so-called *Stein operator*, defined such that $f = Sh$ solves the *Stein equation*

$$h(k) - \mathrm{E}[h(Y)] = Af(k). \tag{2.2}$$

We may then write $\mathrm{E}[h(W)] - \mathrm{E}[h(Y)] = \mathrm{E}[Af(W)]$. Bounding $|\mathrm{E}[Af(W)]|$ thus yields a bound on the error in our approximation.

In Section 4 we derive a representation of our Stein operator from which we establish the bounds in Theorem 2.1, below. Note that this representation satisfies $Sh(0) = 0$ for all h bounded. Furthermore, these bounds are applicable only when there is some $l \in \mathbb{N}$ such that $\mathrm{P}(X \le l) = 1$.

In the sequel we use $\|\cdot\|_\infty$ to denote the supremum norm, $\|g\|_\infty = \sup_{i\in\mathbb{Z}^+} |g(i)|$. We write Δ for the forward difference operator, so that $\Delta g(k) = g(k+1) - g(k)$.

Theorem 2.1. *Let $Y \sim CG(p, \boldsymbol{\mu})$, where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_l, 0, 0, \dots)$ for some $l \in \mathbb{N}$. Let S be the Stein operator corresponding to the characterising operator (2.1). For $h\colon \mathbb{Z}^+ \mapsto \mathbb{R}$ bounded and $n \in \mathbb{N}$,*

$$\|Sh\|_\infty \le \frac{1}{p}\|h - \mathrm{E}[h(Y)]\|_\infty, \tag{2.3}$$

$$\|\Delta^n Sh\|_\infty \le \frac{1}{p}\|\Delta^n h\|_\infty, \tag{2.4}$$

$$\sup_{j,k\in\mathbb{Z}^+} |Sh(j) - Sh(k)| \le \frac{1}{p}\Big(\sup_{i\in\mathbb{Z}^+} h(i) - \inf_{i\in\mathbb{Z}^+} h(i)\Big). \tag{2.5}$$

The proof of Theorem 2.1 is given in Section 4, before which we consider some applications of the compound geometric approximation.

3. Applications

Throughout this section, we let $Y \sim \mathrm{CG}(p, \boldsymbol{\mu})$ and W be a random variable supported on $\mathbb{Z}^+$. We write $f = Sh$, where S is defined as in Theorem 2.1. We assume that W satisfies

$$\mathrm{P}(W = 0) = \mathrm{P}(Y = 0) = p, \tag{3.1}$$

https://doi.org/10.1239/jap/1269610822 Published online by Cambridge University Press

and define the random variable V such that

$$V + X \stackrel{\mathrm{D}}{=} W \mid W > 0. \tag{3.2}$$

Such coupling constructions are used in [5] for the Poisson approximation and in [13] for the geometric case.

Assuming that W and X are independent, we use (2.1) and (2.2) to write

$$\begin{aligned} \mathrm{E}[h(W)] - \mathrm{E}[h(Y)] &= \mathrm{E}[\mathcal{A} f(W)] \\ &= q\,\mathrm{E}[f(W+X)] - q\,\mathrm{E}[f(W) \mid W > 0] \\ &= q\,\mathrm{E}[f(W+X) - f(V+X)], \end{aligned} \tag{3.3}$$

since $f(0) = 0$. Combining (3.3) with Theorem 2.1, we obtain the bounds in Propositions 3.1 and 3.2, below. In the following, we use $W \le_{\mathrm{st}} V$ to denote that the random variable W is stochastically smaller than V.

Proposition 3.1. *Let the random variables V, W, and Y be as above. Suppose that either $W \le_{\mathrm{st}} V$ or $V \le_{\mathrm{st}} W$. For $h\colon \mathbb{Z}^+ \mapsto \mathbb{R}$ bounded,*

$$|\mathrm{E}[h(W)] - \mathrm{E}[h(Y)]| \le \|\Delta h\|_\infty |\mathrm{E}[W] - \mathrm{E}[Y]|.$$

Proof. From (3.3) we immediately obtain the bound

$$|\mathrm{E}[h(W)] - \mathrm{E}[h(Y)]| \le q\|\Delta f\|_\infty\,\mathrm{E}\,|W - V| \le \frac{q}{p}\|\Delta h\|_\infty\,\mathrm{E}\,|W - V|,$$

using Theorem 2.1. If, say, $V \le_{\mathrm{st}} W$ then we can construct random variables V^* and W^* on the same probability space such that $W^* \stackrel{\mathrm{D}}{=} W$, $V^* \stackrel{\mathrm{D}}{=} V$, and $V^* \le W^*$ almost surely. See [16, p. 5]. In this case $\mathrm{E}\,|W - V| = \mathrm{E}[W - V]$. Using the definition of V to compute this expectation yields the desired result. A similar argument holds if $W \le_{\mathrm{st}} V$.

Proposition 3.2. *Let the random variables V, W, and Y be as above. For $h\colon \mathbb{Z}^+ \mapsto \mathbb{R}$ bounded,*

$$|\mathrm{E}[h(W)] - \mathrm{E}[h(Y)]| \le \frac{q}{p}\,\mathrm{P}(W \ne V)\Big[\sup_{i\in\mathbb{Z}^+} h(i) - \inf_{i\in\mathbb{Z}^+} h(i)\Big],$$

and in particular $d_{\mathrm{TV}}(W, Y) \le (q/p) d_{\mathrm{TV}}(W, V)$.

Proof. From (3.3) we have

$$|\mathrm{E}[h(W)] - \mathrm{E}[h(Y)]| \le q\,\mathrm{P}(W \ne V) \sup_{j,k\in\mathbb{Z}^+} |f(j) - f(k)|,$$

from which the first part of the result follows on applying Theorem 2.1. The final part follows using the definition of the total variation distance and taking the supremum over all random variables V satisfying the required condition.

Example 3.1. We consider an example from [14]. Suppose that m balls are placed randomly in $d \ge 2$ compartments, with all assignments equally likely. Let W be the number of balls in the first compartment. Then $W \sim \mathrm{Pya}(m, d)$ has a Pólya distribution with

$$\mathrm{P}(W = k) = \binom{d+m-k-2}{m-k} \Big/ \binom{d+m-1}{m}, \qquad 0 \le k \le m.$$

In line with (3.1), we compare Y to $W \sim \mathrm{Geom}(p)$, where

$$p = \mathrm{P}(W = 0) = \frac{d-1}{d+m-1}.$$

In this case we have $\mathrm{P}(X = 1) = 1$, and so using (3.2) we can check that $V \sim \mathrm{Pya}(m-1, d)$ and that $V \leq_{\mathrm{st}} W$. Hence, by Proposition 3.1,

$$d_{\mathrm{TV}}(W, Y) \leq \frac{m}{d(d-1)}. \tag{3.4}$$

In many cases this performs better than other bounds available. In [14], W is compared to $Z \sim \mathrm{Geom}(r)$, where $r = d/(d+m)$. This approximating distribution is chosen such that $\mathrm{E}[W] = \mathrm{E}[Z]$, and is therefore different to ours. It was shown in [14] that

$$d_{\mathrm{TV}}(W, Z) \leq \frac{2m(d+2m)}{d(d+1)(d+m-1)}. \tag{3.5}$$

Comparing the bounds (3.4) and (3.5), our bound is sharper whenever $m > (d-1)^2/(3d-5)$.

We turn now to our main application, the approximation of Markov chain hitting times by a compound geometric distribution. In this work we generalise a bound of [13]. For further applications of Stein's method to Markov chains, see, for example, [6].

3.1. Markov chain hitting times

Suppose that $\{\xi_i : i \geq 0\}$ is an ergodic discrete-time Markov chain started (at time zero) according to its stationary distribution. Define $W = \min\{i : \xi_i \in B\}$, the time of the first visit to some set of states B. We denote the stationary distribution of this Markov chain by π, and write $\pi_i = \mathrm{P}(\pi = i)$. We also write

$$P_{ij}^{(n)} = \mathrm{P}(\xi_n = j \mid \xi_0 = i).$$

We let B_i denote the set of states from which a move to B requires a minimum time i. That is,

$$B_i = \{j : \ \mathrm{P}(\xi_i \in B \mid \xi_0 = j) > 0 \text{ and } \mathrm{P}(\xi_k \in B \mid \xi_0 = j) = 0 \text{ for } k = 0, 1, \dots, i-1\}$$

for $i = 0, 1, \dots$. Note that $B_0 = B$. In order to apply Theorem 2.1, we assume that there is some $l \in \mathbb{N}$ such that $B_i = \varnothing$ for $i > l$.

We approximate W by $Y \sim \mathrm{CG}(p, \boldsymbol{\mu})$, choosing

$$p = \mathrm{P}(W = 0) = \sum_{j \in B} \pi_j \tag{3.6}$$

and

$$\mu_i = \mathrm{P}(\xi_0 \in B_i \mid \xi_0 \notin B) = \frac{\sum_{j \in B_i} \pi_j}{\sum_{j \notin B} \pi_j} = \frac{\sum_{j \in B_i} \pi_j}{1-p}. \tag{3.7}$$

We then obtain the following bound.

Theorem 3.1. *Let $\{\xi_i : i \geq 0\}$ be an ergodic Markov chain started according to its stationary distribution π, and let $W = \min\{i : \xi_i \in B\}$ be the time of its first visit to a set of states B. Let $Y \sim CG(p, \boldsymbol{\mu})$, where p is defined by (3.6) and μ_i is defined by (3.7). Let X be the random*

https://doi.org/10.1239/jap/1269610822 Published online by Cambridge University Press

variable with $\mathrm{P}(X = i) = \mu_i$*, and assume that there is some* $l \in \mathbb{N}$ *such that* $\mu_i = 0$ *for all* $i > l$*. Then*

$$d_{\mathrm{TV}}(W, Y) \leq \frac{1}{p} \mathrm{E} \sum_{i,j \in B} \pi_i \sum_{n \geq 0} |P_{ij}^{(n+X)} - \pi_j|.$$

Proof. The result follows from Proposition 3.2. Fix some $k \geq 1$, and define a Markov chain $\{\hat{Z}_1^{(k)}, \dots, \hat{Z}_k^{(k)}, \hat{\xi}_0^{(k)}, \hat{\xi}_1^{(k)}, \dots\}$ with the same transition probabilities as our original chain, but started according to π restricted to B_k and rescaled so as to sum to 1. Let

$$V^{(k)} = \min\{i : \hat{\xi}_i^{(k)} \in B\}.$$

Given that $X = k$, letting $V \stackrel{\mathrm{D}}{=} V^{(k)}$, we find that V satisfies (3.2).

We further define the Markov chain $\{Z_1^{(k)}, \dots, Z_k^{(k)}, \xi_0^{(k)}, \xi_1^{(k)}, \dots\}$, again with the same transition probabilities as our original, and again started according to π. We define $W^{(k)} = \min\{i : \xi_i^{(k)} \in B\}$ so that, for each $k \geq 1$, $W^{(k)} \stackrel{\mathrm{D}}{=} W$. Hence, we obtain

$$\mathrm{P}(W \neq V) = \sum_{k=1}^{l} \mu_k \, \mathrm{P}(W^{(k)} \neq V^{(k)}). \tag{3.8}$$

Following the argument of [13], for each $k \geq 1$, we have

$$\mathrm{P}(W^{(k)} \neq V^{(k)}) \leq \sum_{j \in B} \sum_{n \geq 0} [\mathrm{P}(\xi_n^{(k)} = j, \, \hat{\xi}_n^{(k)} \neq j) + \mathrm{P}(\xi_n^{(k)} \neq j, \, \hat{\xi}_n^{(k)} = j)]. \tag{3.9}$$

We couple our Markov chains using the maximal coupling of [8] and [15], so that $\mathrm{P}(\xi_n^{(k)} = \hat{\xi}_n^{(k)} = j) = \pi_j \wedge \mathrm{P}(\hat{\xi}_n^{(k)} = j)$. Following calculations on [13, p. 709], we then obtain

$$\mathrm{P}(\xi_n^{(k)} = j, \, \hat{\xi}_n^{(k)} \neq j) = \biggl(\pi_j - \sum_{i \notin B} \frac{\pi_i}{q} P_{ij}^{(n+k)}\biggr)_+ = \biggl(\frac{1}{q} \sum_{i \in B} \pi_i [P_{ij}^{(n+k)} - \pi_j]\biggr)_+.$$

This is bounded to obtain

$$\mathrm{P}(\xi_n^{(k)} = j, \, \hat{\xi}_n^{(k)} \neq j) \leq \frac{1}{q} \sum_{i \in B} \pi_i (P_{ij}^{(n+k)} - \pi_j)_+.$$

A similar argument yields

$$\mathrm{P}(\xi_n^{(k)} \neq j, \, \hat{\xi}_n^{(k)} = j) \leq \frac{1}{q} \sum_{i \in B} \pi_i (\pi_j - P_{ij}^{(n+k)})_+.$$

Combining these with (3.9) we obtain

$$\mathrm{P}(W^{(k)} \neq V^{(k)}) \leq \frac{1}{q} \sum_{i,j \in B} \pi_i \sum_{n \geq 0} |P_{ij}^{(n+k)} - \pi_j|.$$

Applying (3.8) gives us

$$\mathrm{P}(W \neq V) \leq \frac{1}{q} \sum_{k=1}^{l} \mu_k \sum_{i,j \in B} \pi_i \sum_{n \geq 0} |P_{ij}^{(n+k)} - \pi_j| = \frac{1}{q} \mathrm{E} \sum_{i,j \in B} \pi_i \sum_{n \geq 0} |P_{ij}^{(n+X)} - \pi_j|.$$

The result then follows on application of Proposition 3.2.

https://doi.org/10.1239/jap/1269610822 Published online by Cambridge University Press

Example 3.2. We again follow [13] and apply our Theorem 3.1 to sequence patterns. Let $\eta_0, \eta_1, \dots$ be a sequence of i.i.d. Bernoulli trials, and let I_j be the indicator of the event that a given k-digit pattern appears in $\eta_j, \dots, \eta_{j+k-1}$. We let W be the number of trials before the first occurrence of our pattern, so that $W = \min\{j : I_j = 1\}$.

Define $c_j = \mathrm{P}(I_j = 1 \mid I_0 = 1)$ for $j = 1, \dots, k-1$. Following [13] we define a 2^k state Markov chain such that at time n our Markov chain gives the outcome of the k Bernoulli trials starting at time n. Let B denote the state corresponding to our given k-digit pattern. We define our approximating compound geometric distribution Y using (3.6) and (3.7).

Since

$$P_{BB}^{(n)} = \begin{cases} c_n & \text{if } 1 \le n \le k-1, \\ p & \text{otherwise,} \end{cases}$$

Theorem 3.1 gives

$$d_{\mathrm{TV}}(W, Y) \le \sum_{i=1}^{k-1} \sum_{n=1}^{i} \mu_n |c_i - p|. \tag{3.10}$$

This gives a sharper bound than the geometric approximation considered in [13]. Theorem 3 of [13] shows that

$$d_{\mathrm{TV}}(W, Z) \le \sum_{i=1}^{k-1} |c_i - p|, \tag{3.11}$$

where $Z \sim \mathrm{Geom}(p)$. Since $\sum_{n=1}^{i} \mu_n < 1$ for each $1 \le i \le k-1$, (3.10) does indeed give a tighter bound.

For example, let $k = 3$ and consider the time until the appearance of the pattern 010 in a sequence of i.i.d. Bernoulli trials with $\mathrm{P}(\eta_i = 1) = 1 - \mathrm{P}(\eta_i = 0) = r$ for each i. The state space of our Markov chain consists of all eight three-digit binary patterns. Following the construction of our approximating compound geometric distribution, we partition this state space into the disjoint sets B, B_1, B_2, and B_3 given by

$$B = \{010\}, \qquad B_1 = \{001, 101\}, \qquad B_2 = \{000, 100, 110\}, \qquad B_3 = \{011, 111\}.$$

Using definition (3.6), we have $p = \mathrm{P}(W = 0) = r(1-r)^2$. We also note that

$$\sum_{i \in B_1} \pi_i = r(1-r)^2 + r^2(1-r) = r(1-r).$$

Similar calculations for B_2 and B_3 then give, using (3.7),

$$\mu_1 = \frac{r(1-r)}{1-p}, \qquad \mu_2 = \frac{(1-r)(1-r+r^2)}{1-p}, \qquad \mu_3 = \frac{r^2}{1-p}.$$

We may also check, from the definition above, that $c_1 = 0$ and $c_2 = r(1-r)$. Letting $Y \sim \mathrm{CG}(p, (\mu_1, \mu_2, \mu_3))$, bound (3.10) may be evaluated to give

$$d_{\mathrm{TV}}(W, Y) \le \frac{r^2(1-r)^2(2-r+r^2)}{1-r(1-r)^2}.$$

For comparison, (3.11) gives $d_{\mathrm{TV}}(W, Z) \le r - r^2$, where $Z \sim \mathrm{Geom}(p)$.

4. Proof of Theorem 2.1

Throughout this section, we use the notation in the statement of Theorem 2.1 and let $\hat{h} = h - \mathrm{E}[h(Y)]$. We begin by stating a representation of our operator S. Writing $f = Sh$, we have

$$f(j) = -\,\mathrm{E}\biggl[\sum_{r=0}^{\infty} q^r \hat{h}(W_r) \biggm| W_0 = j\biggr], \tag{4.1}$$

where $W_r = W_0 + X_1 + \cdots + X_r$ and $X_1, X_2, \ldots$ are i.i.d. random variables as in (1.1). To check that this satisfies our Stein equation, we note that

$$\mathrm{E}\,\mathrm{E}\biggl[\sum_{r=0}^{\infty} q^{r+1} \hat{h}(W_r) \biggm| W_0 = j + X\biggr] = \mathrm{E}\biggl[\sum_{r=1}^{\infty} q^r \hat{h}(W_r) \biggm| W_0 = j\biggr],$$

and, hence, with f as above, and recalling that we define $Af(j) = q\,\mathrm{E}[f(j+X)] - f(j)$,

$$\begin{aligned} Af(j) &= \mathrm{E}\biggl[\sum_{r=0}^{\infty} q^r \hat{h}(W_r) \biggm| W_0 = j\biggr] - \mathrm{E}\,\mathrm{E}\biggl[\sum_{r=0}^{\infty} q^{r+1} \hat{h}(W_r) \biggm| W_0 = j + X\biggr] \\ &= \mathrm{E}[\hat{h}(W_0) \mid W_0 = j] \\ &= \hat{h}(j), \end{aligned}$$

so that f satisfies (2.2).

We next use (4.1) to find another representation of f that we may bound to prove Theorem 2.1. From (4.1) we may write

$$f(j) = -\sum_{k \geq 0} \hat{h}(k) u_k(j), \quad \text{where} \quad u_k(j) = \sum_{r=0}^{\infty} q^r\, \mathrm{P}(W_r = k \mid W_0 = j). \tag{4.2}$$

Consider a defective renewal process in discrete time, such that the first renewal occurs at time W_0. Given a renewal at time m, there is a renewal at time $m + k$ with probability q_k for $k = 1, 2, \ldots$. With probability p, there are no further renewals after time m. With this process in mind we can write

$$u_k(j) = \mathrm{P}(\text{renewal occurs at time } k \mid \text{ first renewal occurs at time } j).$$

We define the generating functions $U_j(t)$ and $Q(t)$ for $t \in \mathbb{C}$ by

$$U_j(t) = \sum_{k \geq 0} u_k(j) t^{-k} \quad \text{and} \quad Q(t) = \sum_{k \geq 0} q_k t^{-k}.$$

Clearly, $u_k(j) = 0$ for $k < j$ and $u_j(j) = 1$. For $k > j$, conditioning on the time of the second renewal gives $u_k(j) = \sum_{s \geq 1} u_{k-s}(j) q_s$. Hence, we have $U_j(t) = t^{-j} + Q(t)U_j(t)$, from which we obtain $U_j(t) = t^{-j}[1 - Q(t)]^{-1}$. See also, for example, Section 13.4 of [7].

We apply Cauchy's formula to invert $U_j(t)$. That is, we note that

$$u_k(j) = \frac{1}{2\pi \mathrm{i}} \int_{\Gamma} U_j(t) t^{k-1}\, \mathrm{d}t = \frac{1}{2\pi \mathrm{i}} \int_{\Gamma} \frac{t^{k-j-1}}{1 - Q(t)}\, \mathrm{d}t, \tag{4.3}$$

where Γ is a circular contour in $\mathbb{C}$ centred on the origin. Thus, combining (4.2) and (4.3),

$$f(j) = -\frac{1}{2\pi \mathrm{i}} \sum_{k=0}^{\infty} \hat{h}(k) \int_{\Gamma} \frac{t^{k-j-1}}{1-Q(t)} \, \mathrm{d}t = \frac{1}{2\pi \mathrm{i}} \sum_{k=0}^{\infty} \hat{h}(k) \int_{\Gamma} \frac{t^{l+k-j-1}}{-t^l + \sum_{m=0}^{l} q_m t^{l-m}} \, \mathrm{d}t, \quad (4.4)$$

where we choose Γ to be a circle in $\mathbb{C}$ centred on the origin and with radius r for some $q^{1/l} < r < 1$. Cauchy's formula is also employed in [3], in combination with Fourier techniques, to derive a representation of a Stein operator for compound Poisson approximation.

Before proceeding, we state three results from complex analysis we will need in our proof. Proofs of Lemmas 4.1–4.3 may be found in [9], [10], and [12], respectively. In the sequel, when applying Lemmas 4.1 and 4.2, we need the fact that the integrands of (4.4) may be written as ratios of polynomials (of finite degree). Without the existence of $l \in \mathbb{N}$ such that $\mu_j = 0$ for all $j > l$ (as specified in the statement of Theorem 2.1), this would not be the case.

Lemma 4.1. *Let $R(t) = a_n t^n + a_{n-1} t^{n-1} + \cdots + a_0$ be a polynomial of degree n, and let $r > 0$. If*

$$|a_m r^m| > \sum_{\substack{j=1 \\ j \neq m}}^{n} |a_j r^j|$$

then R has m roots inside the circle of centre 0 and radius r in $\mathbb{C}$.

Lemma 4.2. *Let P and Q be polynomials, of degree m and n, respectively. Let P have leading coefficient a and Q have leading coefficient b. Then*

$$\int_{\Gamma} \frac{P(t)}{Q(t)} \, \mathrm{d}t = \begin{cases} 2\pi \mathrm{i} a/b & \text{if } n = m+1, \\ 0 & \text{if } n \geq m+2, \end{cases}$$

where Γ is a smooth, closed contour in $\mathbb{C}$ with all the zeros of Q in its interior.

Lemma 4.3. *Let $z \in \mathbb{C}$. Suppose that $g(z) = 0$ and $g'(z) \neq 0$. The residue of $1/g$ at z is $1/g'(z)$.*

Defining $R(t) = -t^l + \sum_{k=1}^{l} q_k t^{l-k}$, we can show that $R((q+\varepsilon)^{1/l}) < 0$ for any $\varepsilon > 0$. Applying Lemma 4.1, we note that in (4.4) all the poles of our integrands lie within our chosen contour Γ. We may then immediately apply Lemma 4.2 to find that many of the integrals in (4.4) are 0. We thus obtain the representation of f that we employ:

$$f(j) = \frac{1}{2\pi \mathrm{i}} \sum_{k=j-l+1}^{\infty} \hat{h}(k) \int_{\Gamma} \frac{t^{l+k-j-1}}{-t^l + \sum_{m=1}^{l} q_m t^{l-m}} \, \mathrm{d}t. \quad (4.5)$$

We will also need the following result.

Lemma 4.4. *Let $l \in \mathbb{N}$ and $m \in \mathbb{Z}^+$. Let $q_1, \ldots, q_l$, and let the contour Γ be as above. We have*

$$\int_{\Gamma} \frac{t^m}{-t^l + \sum_{k=1}^{l} q_k t^{l-k}} \, \mathrm{d}t = -\mathrm{i}c$$

for some $c \geq 0$ depending on l and m.

https://doi.org/10.1239/jap/1269610822 Published online by Cambridge University Press

Proof. When $m < l$, we apply Lemma 4.2 directly to obtain the result. When $m \geq l$, we write the integrand as the sum of a polynomial and a rational function to which Lemma 4.2 may be applied. The former is analytic and so has integral 0. We can check that Lemma 4.2 applied to the latter gives the required result.

Combining (4.5) and Lemma 4.4, we immediately obtain

$$
\begin{aligned}
|f(j)| &\leq -\frac{1}{2\pi \mathrm{i}} \|\hat{h}\|_\infty \sum_{k=0}^{\infty} \int_\Gamma \frac{t^k}{-t^l + \sum_{m=1}^{l} q_m t^{l-m}} \, \mathrm{d}t \\
&= -\frac{1}{2\pi \mathrm{i}} \|\hat{h}\|_\infty \int_\Gamma \frac{1}{(1-t)(-t^l + \sum_{m=1}^{l} q_m t^{l-m})} \, \mathrm{d}t. \qquad (4.6)
\end{aligned}
$$

Using Lemma 4.2, if Γ were extended to include the point 1 in its interior, the integral in (4.6) would be 0. Since Γ includes all other poles of the integrand, we use Cauchy's residue formula to note that (4.6) gives the bound $|f(j)| \leq C\|\hat{h}\|_\infty$, where C is the residue of $(1-t)^{-1}(-t^l + \sum_{m=1}^{l} q_m t^{l-m})^{-1}$ at $t = 1$. Using Lemma 4.3, this residue is p^{-1}. This completes the proof of (2.3).

For (2.4), we need a generalisation of the representation (4.5), as in Lemma 4.5, below.

Lemma 4.5. *Let $n \geq 1$. With the notation above,*

$$
\Delta^n f(j) = \frac{1}{2\pi \mathrm{i}} \sum_{k=j-l+1}^{\infty} \Delta^n h(k) \int_\Gamma \frac{t^{l+k-j-1}}{-t^l + \sum_{m=1}^{l} q_m t^{l-m}} \, \mathrm{d}t.
$$

Proof. We proceed by induction on n. From (4.5) we can show that

$$
\Delta f(j) = \frac{1}{2\pi \mathrm{i}} \sum_{k=j-l+1}^{\infty} \hat{h}(k) \int_\Gamma \frac{(1-t)t^{j+k-j-2}}{-t^l + \sum_{m=1}^{l} q_m t^{l-m}} \, \mathrm{d}t. \qquad (4.7)
$$

We note that

$$
\sum_{k=-1}^{\infty} \int_\Gamma \frac{(1-t)t^k}{-t^l + \sum_{m=1}^{l} q_m t^{l-m}} \, \mathrm{d}t = \int_\Gamma \frac{1}{t(-t^l + \sum_{m=1}^{l} q_m t^{l-m})} \, \mathrm{d}t = 0, \qquad (4.8)
$$

by Lemma 4.2. Hence, (4.7) gives

$$
\Delta f(j) = \frac{1}{2\pi \mathrm{i}} \sum_{k=j-l+1}^{\infty} h(k) \int_\Gamma \frac{(1-t)t^{j+k-j-2}}{-t^l + \sum_{m=1}^{l} q_m t^{l-m}} \, \mathrm{d}t. \qquad (4.9)
$$

We write $h(k) = h(0) + \sum_{m=0}^{k-1} \Delta h(m)$. Substituting this into (4.9) and using (4.8) to note that the term involving $h(0)$ is 0, we interchange the order of summation in the resulting expression to obtain

$$
\begin{aligned}
\Delta f(j) = \frac{1}{2\pi \mathrm{i}} \bigg[\sum_{m=0}^{j-l} \Delta h(m)\bigg] \sum_{k=j-l+1}^{\infty} \int_\Gamma \frac{(1-t)t^{l+k-j-2}}{-t^l + \sum_{s=1}^{l} q_s t^{l-s}} \, \mathrm{d}t \\
+ \frac{1}{2\pi \mathrm{i}} \sum_{m=j-l+1}^{\infty} \sum_{k=m+1}^{\infty} \Delta h(m) \int_\Gamma \frac{(1-t)t^{l+k-j-2}}{-t^l + \sum_{s=1}^{l} q_s t^{l-s}} \, \mathrm{d}t. \qquad (4.10)
\end{aligned}
$$

https://doi.org/10.1239/jap/1269610822 Published online by Cambridge University Press

Again, using (4.8), the first term of (4.10) is 0. Interchanging the integration and summation over k in the second term gives us the required representation when $n = 1$.

For the inductive step, we proceed similarly to the above. We assume our representation of $\Delta^{n-1} f$, and, thus, obtain an analogue of (4.7):

$$\Delta^n f(j) = \frac{1}{2\pi \mathrm{i}} \sum_{k=j-l+1}^{\infty} \Delta^{n-1} h(k) \int_{\Gamma} \frac{(1-t)t^{l+k-j-2}}{-t^l + \sum_{m=1}^{l} q_m t^{l-m}} \, \mathrm{d}t. \tag{4.11}$$

Writing $\Delta^{n-1} h(k) = \Delta^{n-1} h(0) + \sum_{m=0}^{k-1} \Delta^n h(m)$ and proceeding as we did for $n = 1$ gives us our lemma.

To complete the proof of (2.4), we bound the representation of Lemma 4.5 in the same way as (4.5) above. To see (2.5), we define

$$c_k(j) = -\frac{1}{2\pi \mathrm{i}} \int_{\Gamma} \frac{t^{l+k-j-1}}{-t^l + \sum_{m=1}^{l} q_m t^{l-m}} \, \mathrm{d}t.$$

By Lemma 4.4, $c_k(j) \in \mathbb{R}^+$ for all $j \in \mathbb{Z}^+$ and $k \geq j - l + 1$. Using (4.5), we can show that $f(r) - f(s) = \sum_{k \geq s-l+1} h(k) c_k(s) - \sum_{k \geq r-l+1} h(k) c_k(r)$ for any $r, s \in \mathbb{Z}^+$. Thus, for any bounded, nonnegative function $h \colon \mathbb{Z}^+ \mapsto \mathbb{R}$, we have

$$\begin{aligned} |f(r) - f(s)| &\leq \max\biggl\{ \sum_{k \geq r-l+1} h(k) c_k(r), \sum_{k \geq s-l+1} h(k) c_k(s) \biggr\} \\ &\leq \|h\|_\infty \sum_{k \geq j-l+1} c_k(j) \\ &= \frac{1}{p} \|h\|_\infty, \end{aligned}$$

where the integrals are evaluated as above. It remains only to remove the restriction that h must be nonnegative. To this end, we define, for any $h \colon \mathbb{Z}^+ \mapsto \mathbb{R}$, the function H given by $H(k) = h(k) - \inf_{m \in \mathbb{Z}^+} h(m)$. Since H is nonnegative, we apply the above to note that, for $r, s \in \mathbb{Z}^+$,

$$|SH(r) - SH(s)| \leq \frac{1}{p} \|H\|_\infty = \frac{1}{p} \Bigl(\sup_{m \in \mathbb{Z}^+} h(m) - \inf_{m \in \mathbb{Z}^+} h(m) \Bigr).$$

Finally, it is straightforward to check that $SH(r) - SH(s) = Sh(r) - Sh(s)$. This completes the proof of Theorem 2.1.

Remark 4.1. We note that the bounds of Theorem 2.1 are sharp. To see this for (2.4), take $h \colon \mathbb{Z}^+ \mapsto \mathbb{R}$ such that $\Delta^n h(j) = 1$ for all j. Using the representation in Lemma 4.5 and evaluating the integrals as above gives us our result.

For (2.3) and (2.5), define $h_n(j) = I_{\{j \geq n\}}$, so that $\sup_i h_n(i) - \inf_i h_n(i) = 1$ and $\|\hat{h}_n\|_\infty = \max\{\mathrm{P}(Y \geq n), \mathrm{P}(Y < n)\} \to 1$ as $n \to \infty$. By (4.5) we have

$$Sh_n(l + n - 1) = \frac{1}{2\pi \mathrm{i}} \sum_{k=n}^{\infty} \mathrm{P}(Y < n) \int_{\Gamma} \frac{t^{k-n}}{-t^l + \sum_{m=1}^{l} q_m t^{l-m}} \, \mathrm{d}t.$$

We interchange the summation and integration and evaluate the remaining integral as we did in (4.6) to obtain $Sh_n(l+n-1) = -p^{-1}\,\mathrm{P}(Y < n) \to -p^{-1}$ as $n \to \infty$. This means that (2.3) is sharp. Recalling that $Sh(0) = 0$ for all bounded functions h (and in particular for h_n), we also find that (2.5) is sharp.

Remark 4.2. The techniques employed above may also be used to prove other bounds for our Stein operator S. For example, we can rearrange (4.11) to obtain, for $n \geq 1$,

$$\Delta^n f(j) = \frac{1}{2\pi \mathrm{i}} \sum_{k=j-l+1}^{\infty} \Delta^{n-1} h(k) \bigg[\int_\Gamma \frac{pt^{l+k-j-2}}{q-t}\,\mathrm{d}t + \int_\Gamma t^{l+k-j-2}\,\mathrm{d}t \bigg].$$

Applying Lemmas 4.2, 4.3, and 4.4, we proceed as in the proof of Theorem 2.1, bounding this representation to obtain $|\Delta^n f(j)| \leq 2\|\Delta^{n-1}h\|_\infty$.

Acknowledgements

The author wishes to thank Sergey Utev and Andrew Barbour for many useful discussions, and an anonymous referee for suggestions which greatly improved the presentation of the work. This work was partially completed while the author was at the University of Nottingham.

References

[1] Barbour, A. D. and Chen, L. H. Y. (eds) (2005). *An Introduction to Stein's Method*. World Scientific, Singapore.
[2] Barbour, A. D. and Grübel, R. (1995). The first divisible sum. *J. Theoret. Prob.* **8,** 39–47.
[3] Barbour, A. D. and Utev, S. (1998). Solving the Stein equation in compound Poisson approximation. *Adv. Appl. Prob.* **30,** 449–475.
[4] Barbour, A. D., Chen, L. H. Y. and Loh, W.-L. (1992). Compound Poisson approximation for nonnegative random variables using Stein's method. *Ann. Prob.* **20,** 1843–1866.
[5] Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation*. Oxford University Press.
[6] Erhardsson, T. (1999). Compound Poisson approximation for Markov chains using Stein's method. *Ann. Prob.* **27,** 565–596.
[7] Feller, W. (1950). *An Introduction to Probability Theory and Its Applications*, Vol. 1. John Wiley, New York.
[8] Griffeath, D. (1975). A maximal coupling for Markov chains. *Z. Wahrscheinlichkeitsth* **31,** 95–106.
[9] Jury, E. I. (1964). *Theory and Application of the z-Transform Method*. John Wiley, New York.
[10] Just, E. and Schaumberger, N. (1964). Contour integration for rational functions. *Amer. Math. Monthly* **71,** 546–547.
[11] Kalashnikov, V. (1997). *Geometric Sums: Bounds for Rare Events with Applications*. Kluwer, Dordrecht.
[12] Lang, S. (1999). *Complex Analysis*, 4th edn. Springer, New York.
[13] Peköz, E. A. (1996). Stein's method for geometric approximation. *J. Appl. Prob.* **33,** 707–713.
[14] Phillips, M. J. and Weinberg, G. V. (2000). Non-uniform bounds for geometric approximation. *Statist. Prob. Lett.* **49,** 305–311.
[15] Pitman, J. W. (1976). On coupling of Markov chains. *Z. Wahrscheinlichkeitsth* **35,** 315–322.
[16] Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic Orders and Their Applications*. Academic Press, Boston, MA.
[17] Stein, C. (1986). *Approximate Computation of Expectations* (IMS Lecture Notes Monogr. Ser. **7**). Institute of Mathematical Statistics, Hayward, CA.