Check for updates

Adv. Appl. Probab. **57,** 1510–1538 (2025)
*doi:*10.1017/apr.2025.20

STEIN'S METHOD FOR DISTRIBUTIONS MODELLING COMPETING AND COMPLEMENTARY RISK PROBLEMS

ANUM FATIMA,* *University of Oxford, UK & Lahore College for Women University, Pakistan*
GESINE REINERT,** *University of Oxford & The Alan Turing Institute, London, UK*

Abstract

Competing and complementary risk (CCR) problems are often modelled using a class of distributions of the maximum, or minimum, of a random number of independent and identically distributed random variables, called the CCR class of distributions. While CCR distributions generally do not have an easy-to-calculate density or probability mass function, two special cases, namely the Poisson–exponential and exponential–geometric distributions, can easily be calculated. Hence, it is of interest to approximate CCR distributions with these simpler distributions. In this paper, we develop Stein's method for the CCR class of distributions to provide a general comparison method for bounding the distance between two CCR distributions, and we contrast this approach with bounds obtained using a Lindeberg argument. We detail the comparisons for Poisson–exponential, and exponential–geometric distributions.

Keywords: Stein's method; competing and complementary risk; Poisson–exponential distribution; Poisson–geometric distribution; exponential–geometric distribution; distributional distance

2020 Mathematics Subject Classification: Primary 60F05
Secondary 60E05

1. Introduction

Competing and complementary risk (CCR) problems typically focus on the failure of a system composed of multiple components or a system with several, sometimes even a countably infinite number of, risk factors that can cause its failure. Here we think of components and risk factors as having random lifetimes, denoted by a sequence of independent and identically distributed (i.i.d.) positive random variables $X_1, X_2, \ldots$. At the end of its lifetime the component fails, or the risk occurs.

CCR systems are either sequential or parallel; in sequential systems, the whole system fails at the occurrence of the first among an unknown number N of risk factors; risks compete with each other to cause failure. In this case, the observed failure time is the minimum of the lifetimes of the risk factors. In parallel systems, the system fails after all of an unknown number N of risk factors occur, and the observed failure time is the maximum of the lifetimes of these

Received 28 March 2025; accepted 28 March 2025.
* Postal address: Department of Statistics, University of Oxford. Email: anumfatimam@gmail.com
** Postal address: Department of Statistics, University of Oxford. Email: reinert@stats.ox.ac.uk

© The Author(s), 2025. Published by Cambridge University Press on behalf of Applied Probability Trust. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

risk factors. In both settings, the lifetimes of the components or risks are modelled as random, as is the number of risks N, which is assumed to be independent of the lifetimes. These two CCR settings are frequently studied together, as $\max(X_1, \ldots, X_N) = -\min(-X_1, \ldots, -X_N)$; see [6].

CCR settings arise in many fields, such as industrial reliability, demography, biomedical studies, public health, and actuarial sciences. For example, in a study of vertical transmission of HIV from an infected mother to her newborn child, several factors increase the risk of transmission, such as high maternal virus load and low birth weight. Exactly which factors determine the timing of transmission is an ongoing area of research. Even with a list of known risk factors, there are possibly many unknown risk factors involved. The problem is therefore modelled as a CCR problem in [8], [20], [21], and [35]. The number of successive failures of the air conditioning system of each member of a fleet of 13 Boeing 720 jet aeroplanes was modelled as a CCR problem in [17], with potential risk factors including defective components in the air conditioning system and errors during the production process. In [3] the period between successive coal-mining disasters is modelled as a CCR problem, with construction faults or human errors committed by inexperienced miners as examples of risks. The daily ozone concentrations in New York during May–September 1973 is treated as a CCR problem in [16].

We call the distributions used to model the lifetimes in CCR problems the *CCR family of distributions*. Examples from this family that have been studied in the literature include the exponential–Poisson distribution [17], the Poisson–exponential (PE) lifetime distribution [8], the Weibull–Poisson distribution [22], the extended Weibull–Poisson distribution, and the extended generalised extreme value Poisson distribution [29]. In [34] a detailed review is given of CCR distributions, and additional ones are proposed therein.

The probability mass function of a CCR distribution can be quite unwieldy. Hence, it is of interest to approximate a CCR distribution by a simpler CCR distribution such as the PE distribution or the exponential–geometric distribution. For these two distributions, the number of risks or components follows a zero-truncated Poisson or geometric distribution, respectively. To assess such approximations, we develop Stein's method for CCR distributions.

The seminal work of Charles Stein [32] derives bounds on the approximation error for normal approximations. In [10], Stein's method is adapted to Poisson approximation; see also [4] and [5]. Generalisations to many other distributions and dimensions are available; see, for example, [9], [24], and [25]. The Stein operators in this paper are based on the density approach; see [18]. A main difficulty in distributional comparison problems is comparing a discrete and a continuous distribution; related works on Stein's method include [15], which bounds the Wasserstein distance between a beta distribution and the distribution of the number of white balls in a Pólya–Eggenberger urn, and [14], where standardisations related to the ones we propose are used. Maxima of a fixed number of random variables were treated with Stein's method in [13]; Stein's method applied to a random sum of random variables can be found in [28]. Here we propose a comparison of the distributions of a maximum (or minimum) of discrete random variables and a maximum (or minimum) of continuous random variables when the number of these random variables is itself an independent random variable. In future work, the Stein characterisations derived in the present paper could be employed to construct Stein-based goodness-of-fit statistics as in [7].

The remainder of this paper is organised as follows. Section 2 gives a brief introduction to Stein's method using the density approach and applies it to obtain a general representation for the CCR class of distributions through a Stein operator. Section 3 develops Stein's method for

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

comparing CCR distributions; as an alternative approach, it also provides a comparison based on a Lindeberg argument. As the main illustration of our results, Section 4 details Stein's method for the PE distribution and uses it to bound the total variation distance between a PE distribution and a distribution from the CCR family, as well as between a PE distribution and a distribution not from the CCR family. As a second illustration, in Section 5 we develop Stein's method for the exponential–geometric distribution and perform distributional comparisons in total variation distance. Section 6 gives a bound on the bounded Wasserstein distance between the distribution of the maximum waiting time of sequence patterns in Bernoulli trials and a PE distribution. Proofs that are standard but which would disturb the flow of the argument are postponed to Appendix.

2. Stein's method for CCR distributions

We use the notation $\mathbb{R}^+_{>0} = (0, \infty)$ and $\mathbb{N} = \{1, 2, 3, \ldots\}$. The backward difference operator Δ^- operates on a function $g : \mathbb{R} \to \mathbb{R}$ by $\Delta^- g(x) = g(x) - g(x-1)$. We note that $\Delta^-(gq)(y) = q(y)\Delta^- g(y) + g(y-1)\Delta^- q(y)$. For a function $h \in \mathrm{Lip}_b(1)$, its derivative is denoted by h' and exists almost everywhere (by Rademacher's theorem).

For two probability distributions q and p on $\mathbb{R}^+_{>0}$, we seek bounds on distances of the form

$$d_{\mathcal{H}}(p, q) := \sup_{h \in \mathcal{H}} \big|\mathbb{E}h(X) - \mathbb{E}h(Z)\big|, \tag{2.1}$$

where $\mathcal{H}$ is a set of test functions, $Z \sim q$, and $X \sim p$. The sets of functions $\mathcal{H}$ in (2.1) are, for the total variation distance d_{TV}, $\mathcal{H} = \{\mathbb{I}[\cdot \in A] : A \in \mathcal{B}(\mathbb{R})\}$ and, for the bounded Wasserstein distance d_{BW},

$$\mathcal{H} = \mathrm{Lip}_b(1) := \{h : \mathbb{R}^+_{>0} \to \mathbb{R} : |h(x) - h(y)| \le |x - y| \text{ for all } x, y \in \mathbb{R};\ \|h\| \le 1\}.$$

Here $\mathcal{B}(\mathbb{R})$ denotes the Borel sets of $\mathbb{R}$ and $\|\cdot\|$ is the supremum norm in $\mathbb{R}$. We note here the alternative formulation for total variation distance based on Borel-measurable functions $h : \mathbb{R} \to \mathbb{R}$,

$$d_{\mathrm{TV}}(p, q) = \frac{1}{2} \sup_{\|h\| \le 1} \big|\mathbb{E}h(X) - \mathbb{E}h(Z)\big|. \tag{2.2}$$

2.1. Stein's method for distributional comparisons

To obtain explicit bounds on the distance between a probability distribution $\mathcal{L}(X)$ of interest and a usually well-understood approximating distribution $\mathcal{L}_0(Z)$, often called the *target* distribution, Stein's method connects the test function $h \in \mathcal{H}$ in (2.1) to the distribution of interest through a *Stein equation*

$$h(x) - \mathbb{E}h(Z) = \mathcal{T}g(x). \tag{2.3}$$

In (2.3), $\mathcal{T}$ is a *Stein operator* for the distribution $\mathcal{L}_0(Z)$, with an associated *Stein class* $\mathcal{F}(\mathcal{T})$ of functions such that $\mathbb{E}[\mathcal{T}g(Z)] = 0$ for all $g \in \mathcal{F}(\mathcal{T})$ if and only if $Z \sim \mathcal{L}_0(Z)$; thus, a Stein operator characterises the distribution. The distance (2.1) can then be bounded by $d_{\mathcal{H}}(\mathcal{L}(X), \mathcal{L}_0(Z)) \le \sup_{g \in \mathcal{F}(\mathcal{H})} |\mathbb{E}\mathcal{T}g(X)|$ where $\mathcal{F}(\mathcal{H}) = \{g_h : h \in \mathcal{H}\}$ is the set of solutions of the Stein equation (2.3) for the set of test functions $h \in \mathcal{H}$.

The Stein operator $\mathcal{T}$ for a probability distribution is not unique; see e.g. [18]. In this paper, we employ the so-called *density method*, which uses the score function of a probability distribution to construct a Stein operator, called a *score Stein operator*. Following [19, 33],

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

a score Stein operator for a continuous distribution with probability density function (PDF) p and support $[a, b] \in \mathbb{R}$ acts on functions g for which the derivative exists by

$$\mathcal{T}_p g(x) = \frac{(gp)'(x)}{p(x)}, \tag{2.4}$$

where we take $0/0 = 0$. For differentiable g and p, (2.4) simplifies to $\mathcal{T}_p g(x) = g'(x) + g(x)\rho(x)$, where $\rho = p'/p$ is the score function of p. The Stein class $\mathcal{F}(\mathcal{T}_p)$ is the collection of functions $g : \mathbb{R} \to \mathbb{R}$ such that $g(x)\ p(x)$ is differentiable with integrable derivative and $\lim_{x \to a,b} g(x)p(x) = 0$. It is straightforward to see that for a PDF p on $[a, b] \subset \mathbb{R}$,

$$g(x) = g_h(x) = \frac{1}{p(x)} \int_0^x [h(t) - \mathbb{E}h(X)]p(t)\, \mathrm{d}t \tag{2.5}$$

solves (2.3) for h and that if h is bounded then $g \in \mathcal{F}(\mathcal{T}_p)$.

To use the Stein equation for a distributional comparison, let X and Y be two random variables with PDFs p_X and p_Y, defined on the same probability space and with nested supports $\mathrm{supp}(p_Y) \subset \mathrm{supp}(p_X) = [a, b] \subset \mathbb{R}$, score functions ρ_X and ρ_Y, and corresponding score Stein operators $\mathcal{T}_X$ and $\mathcal{T}_Y$. Then

$$\mathbb{E}h(Y) - \mathbb{E}h(X) = \mathbb{E}g_{X,h}(Y)(\rho_Y(Y) - \rho_X(Y)), \tag{2.6}$$

where $g_{X,h}(x)$ is the solution of the Stein equation for h and $\mathcal{T}_X$. Equation (2.6) is a special case of Equation 23 in Section 2.5 of [18].

For a discrete distribution with probability mass function (PMF) q having support $\mathcal{I} = [a, b] \subset \mathbb{N}$, the discrete backward score function is $\frac{\Delta^- q(y)}{q(y)}$, and, as in Remark 3.2 and Example 3.13 of [18], a discrete backward score Stein operator is

$$\mathcal{T}_q g(y) = \Delta^- g(y) + \frac{\Delta^- q(y)}{q(y)} g(y-1), \quad y \in \mathcal{I}. \tag{2.7}$$

With an abuse of notation, we often refer to a Stein operator for the distribution of X as the Stein operator for X, and similarly refer to the score function of the distribution of X as the score function of X. Further, if $X \sim p$, we also write $\mathcal{T}_X$ for $\mathcal{T}_p$.

2.2. CCR distributions

Let $N \in \mathbb{N}$ be a random variable with finite second moment and let $\mathbf{Y} = (Y_1, Y_2, \dots)$ be a sequence of positive i.i.d. random variables, with cumulative distribution function (CDF) F_Y independent of N. Then

$$W_\alpha = W_\alpha(N, \mathbf{Y}) = \begin{cases} \min\{Y_1, Y_2, \dots, Y_N\} & \text{if } \alpha = -1, \\ \max\{Y_1, Y_2, \dots, Y_N\} & \text{if } \alpha = 1 \end{cases} \tag{2.8}$$

is called a CCR random variable with type indicator $\alpha \in \{-1, 1\}$. Setting

$$U_Y^\alpha(\cdot) = \begin{cases} 1 - F_Y(\cdot) & \text{if } \alpha = -1, \\ F_Y(\cdot) & \text{if } \alpha = 1 \end{cases} \tag{2.9}$$

and writing $G_N(x) = \mathbb{E}\, x^N$, the probability generating function of N, $W_\alpha > 0$ has CDF

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

$$F_{W_\alpha}(w) = \mathbb{P}(W_\alpha \le w) \tag{2.10}$$

$$= \begin{cases} 1 - \sum_n \mathbb{P}(N=n) U_Y^\alpha(w)^n = 1 - (G_N \circ U_Y^\alpha)(w) & \text{if } \alpha = -1, \\ \sum_n \mathbb{P}(N=n) U_Y^\alpha(w)^n = (G_N \circ U_Y^\alpha)(w) & \text{if } \alpha = 1. \end{cases} \tag{2.11}$$

If the Y_i have a continuous distribution with PDF f_Y, then W_α has PDF

$$f_{W_\alpha}(w) = \sum_n \mathbb{P}(N=n)\, n(U_Y^\alpha(w))^{n-1} f_Y(w) = f_Y(w)(G_N' \circ U_Y^\alpha)(w); \tag{2.12}$$

see also [34]. Here we have used that $(U_Y^\alpha)'(w) = \alpha f_Y(w)$.

If the Y_i are discrete with PMF p_Y on $\mathbb{N}$, the resulting random variable W_α has PMF p_{W_α}, which can be expressed in terms of $G_N(\cdot)$ as

$$p_{W_\alpha}(x) = \alpha\big(G_N(U_Y^\alpha(x)) - G_N(U_Y^\alpha(x-1))\big) = \alpha \Delta^-(G_N \circ U_Y^\alpha)(x). \tag{2.13}$$

Example 1.

(i) *If N is a zero-truncated Poisson random variable with parameter θ, then (2.12) simplifies to $f_{W_\alpha}(w) = f_Y(w)\frac{\theta}{(1-e^{-\theta})} e^{-\theta(1-U_Y^\alpha(w))}$ for $w > 0$. This is the PDF of the extended Poisson family of distributions given in Equation 3 of [29], where it is called the G-Poisson class of distributions [34].*

(ii) *If N is a* Geometric(p) *random variable with PMF $\mathbb{P}(N=n) = (1-p)^{(n-1)}p$ for $n \in \mathbb{N}$, then (2.12) yields $f_{W_\alpha}(w) = f_Y(w)p/(1-(1-p)U_Y^\alpha(w))^2$ for $w > 0$, which gives the distributions in Equations 5.2 and 5.3 of [23].*

2.3. Stein's method for the CCR class of distributions

To obtain a Stein operator for the CCR random variable $W_\alpha = W_\alpha(N, \mathbf{Y})$ in (2.8), we use the density method. First, we assume that the Y_i are continuous with differentiable PDF f_Y. From (2.12), the score function for the distribution of W_α is

$$\rho_{W_\alpha}(w) = \alpha f_Y(w) \frac{(G_N'' \circ U_Y^\alpha)(w)}{(G_N' \circ U_Y^\alpha)(w)} + \rho_Y(w), \tag{2.14}$$

where $\rho_Y = f_Y'/f_Y$ is the score function of Y. Hence $\mathcal{T}_{W_\alpha}$ given by

$$\mathcal{T}_{W_\alpha} g(w) = g'(w) + \rho_{W_\alpha}(w) g(w) \tag{2.15}$$

for differentiable g is a Stein operator acting on the functions $g \in \mathcal{F}(\mathcal{T}_{W_\alpha})$. For a test function $h \in \mathcal{H}$, the corresponding Stein equation is

$$g'(w) + \rho_{W_\alpha}(w) g(w) = h(w) - \mathbb{E}h(W_\alpha). \tag{2.16}$$

Thus, for any random variable X, the distance $d_{\mathcal{H}}$ from (2.1) between the distributions of X and W_α can be bounded by bounding the expectation of the left-hand side of (2.16). For Y_i taking values in $\mathbb{N}$ with PMF p_Y, the backward score function is

$$\rho_{W_\alpha}(w) = \frac{\Delta^- p_{W_\alpha}(w)}{p_{W_\alpha}(w)} = \frac{\Delta^-(\Delta^-(G_N \circ U_Y^\alpha))(w)}{\Delta^-(G_N \circ U_Y^\alpha)(w)}, \tag{2.17}$$

with corresponding discrete backward score Stein operator $\mathcal{T}_{W_\alpha}$ operating as

$$\mathcal{T}_{W_\alpha} g(w) = \Delta^- g(w) + \rho_{W_\alpha}(w) g(w-1). \tag{2.18}$$

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

3. A general comparison approach

To illustrate the use of Stein's method for CCR distributions, we compare the distributions of two maxima or two minima of a random number of i.i.d. random variables $W_\alpha(N, \mathbf{Y})$ and $W_\alpha(M, \mathbf{Z})$.

Proposition 1. *Let $W_{\alpha_1}(N, \mathbf{Y})$ and $W_{\alpha_2}(M, \mathbf{Z})$ for $\alpha_1, \alpha_2 \in \{-1, 1\}$ be CCR random variables with PDFs f_Y and f_Z and score functions ρ_Y and ρ_Z. Then for any test function h such that the $W_{\alpha_1}(N, \mathbf{Y})$ Stein equation (2.16) for h has a solution $g = g_h$,*

$$\begin{aligned}&\left|\mathbb{E}h(W_{\alpha_1}(N, \mathbf{Y})) - \mathbb{E}h(W_{\alpha_2}(M, \mathbf{Z}))\right| \\ &= \left|\mathbb{E}g(W)\left(\alpha_1 f_Y(W)\frac{(G_N'' \circ U_Y^{\alpha_1})(W)}{(G_N' \circ U_Y^{\alpha_1})(W)} - \alpha_2 f_Z(W)\frac{(G_M'' \circ U_Z^{\alpha_2})(W)}{(G_M' \circ U_Z^{\alpha_2})(W)}\right) + \mathbb{E}g(W)(\rho_Y - \rho_Z)(W)\right|,\end{aligned} \tag{3.1}$$

where $W = W_{\alpha_2}(M, \mathbf{Z})$ and $U_\cdot^\alpha$ is as in (2.9).

If the Y_i are discrete with PMF f_Y and the Z_i are discrete with PMF f_Z, then

$$\begin{aligned}&\left|\mathbb{E}h(W_{\alpha_1}(N, \mathbf{Y})) - \mathbb{E}h(W_{\alpha_2}(M, \mathbf{Z}))\right| \\ &= \left|\mathbb{E}g(W-1)\left(f_Y(W)\frac{\Delta^-(\Delta^-(G_N \circ U_Y^{\alpha_1}))(W)}{\Delta^-(G_N \circ U_Y^{\alpha_1})(W)} - f_Z(W)\frac{\Delta^-(\Delta^-(G_N \circ U_Z^{\alpha_2}))(W)}{\Delta^-(G_N \circ U_Z^{\alpha_2})(W)}\right)\right|.\end{aligned} \tag{3.2}$$

Proof. Substitute the score functions (2.14) and (2.17) into (2.6); simplifying then gives (3.1) and (3.2) respectively. □

To compare a discrete random variable and a continuous random variable, we use the concept of *standardised Stein equations* as in [14] and [18]. For a continuous random variable W with score function ρ_W and a differentiable function $c : \mathbb{R} \to \mathbb{R}$, define a c-standardised Stein operator $\mathcal{T}_W^{(c)}$ by

$$\mathcal{T}_W^{(c)}(g(w)) = \mathcal{T}_W(cg)(w) = c(w)g'(w) + [c(w)\rho_W(w) + c'(w)]\, g(w). \tag{3.3}$$

For a random variable $V \in \mathbb{N}$ with discrete backward score function ρ_V and a function $d : \mathbb{N} \to \mathbb{R}$, define a d-standardised Stein operator for V by

$$\begin{aligned}\mathcal{T}_V^{(d)}(g(w)) &= \mathcal{T}_V(dg)(w) = \Delta^-(dg)(w) + \rho_V(w)(dg)(w-1) \\ &= d(w-1)\Delta^- g(w) + g(w)\Delta^- d(w) + d(w-1)g(w-1)\rho_V(w).\end{aligned} \tag{3.4}$$

For CCR random variables $W_\alpha(N, \mathbf{Y})$ and $W_\alpha(M, \mathbf{Z})$, when the Y_i are continuous on $\mathbb{R}_{>0}^+$ with differentiable PDF f_y and the Z_i take values in $\mathbb{N}$, we rescale $W_\alpha(M, \mathbf{Z})$ by dividing it by n, to obtain $W_n = W_{\alpha,n} = \frac{1}{n}W_\alpha(M, \mathbf{Z})$. If $W_\alpha(M, \mathbf{Z}) \in \mathbb{N}$ has PMF p and backward score function ρ, then W_n has PMF $\mathbb{P}(W_n = z) = p(nz)$ and backward score function $\tilde{\rho}_n(z) = \rho(nz)$. We note here that the ratio $\rho(nz) = \frac{p(nz)-p(nz-1)}{p(nz)}$ is the score function of $W_\alpha(M, \mathbf{Z})$ evaluated at nz, which for $n \neq 1$ does not equal the score function of W_n. With $\Delta^{-n}f(x) := f(x) - f(x - 1/n)$, we obtain the Stein operator $\mathcal{T}_n^{(d)}$ given by

$$\mathcal{T}_n^{(d)}g(z) = \mathcal{T}_n(dg)(z) = \Delta^{-n}(dg)(z)d(z) + \tilde{\rho}_n(z)(dg)\,(z - 1/n)\,. \tag{3.5}$$

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

Proposition 2. *Let $W_\alpha(M, \mathbf{Z})$ be a discrete CCR random variable with discrete backward score function ρ_W; for $n \in \mathbb{N}$ set $W_n = W_\alpha(M, \mathbf{Z})/n$ and $\tilde{\rho}_n(z) = \rho_W(nz)$. Let $W = W_\alpha(N, \mathbf{Y})$ be a continuous CCR random variable with score function ρ. Let $h \in \mathcal{H}$ be a test function such that the $\mathcal{L}(W)$ Stein equation (2.3) has solution $g = g_h$. Then for any differentiable function $c : \mathbb{R}^+_{>0} \to \mathbb{R}$ and any function $d : \mathbb{N} \to \mathbb{R}$,*

$$\begin{aligned}\big|\mathbb{E}h(W_n) - \mathbb{E}h(W)\big| \le \big|&\mathbb{E}[n\Delta^{-n}(dg)(W_n) - (cg)'(W_n)] \\ &+ \mathbb{E}[n\tilde{\rho}_n(W_n)(dg)(W_n - 1/n) - (cg)(W_n)\rho(W_n)]\big|.\end{aligned} \tag{3.6}$$

Proof. To compare the two distributions, for a given test function h we have

$$\mathbb{E}h(W_n) - \mathbb{E}h(W) = \mathbb{E}\mathcal{T}_W(cg)(W_n) = \mathbb{E}[(cg)'(W_n) + (cg)(W_n)\rho(W_n)] \tag{3.7}$$

with g solving the continuous Stein equation (3.3) for h. Next, we note that for the Stein operator given in (3.5), $\mathbb{E}\mathcal{T}_n^{(d)}(g)(W_n) = 0$ by construction. Hence, also $n\mathbb{E}\mathcal{T}_n^{(d)}(g)(W_n) = 0$. Thus, (3.7) yields

$$\begin{aligned}\mathbb{E}h(W_n) - \mathbb{E}h(W) = \mathbb{E}\big[&(cg)'(W_n) + (cg)(W_n)\rho(W_n) \\ &- n\Delta^{-n}(dg)(W_n) - n\tilde{\rho}_n(W_n)(dg)(W_n - 1/n)\big].\end{aligned}$$

Rearranging gives the assertion. □

Adaptation to other deterministic scaling functions should be straightforward; as in our examples, we concentrate on the case where we scale by dividing by n. As an aside, while such standardisations and scalings could perhaps be used for a 'bespoke derivative' as in [14], the connection is not obvious.

An alternative comparison of CCR distributions can be achieved using a Lindeberg argument, to arrive at the following result.

Proposition 3. *Let $W_\alpha(M, \mathbf{X})$ and $W_\alpha(N, \mathbf{E})$ be given in (2.8). Then for functions $h \in \mathrm{Lip}_b(1)$,*

$$\begin{aligned}&\big|\mathbb{E}h(W_\alpha(N, \mathbf{E})) - \mathbb{E}h(W_\alpha(M, \mathbf{X}))\big| \\ &\quad\le \|h'\| \sum_{i=1}^{\infty} \mathbb{E}|E_i - X_i|\, \mathbb{P}(M \ge i) \\ &\qquad+ \sum_{m,n=1}^{\infty} \mathbb{P}(M = m, N = n)\, \big|\mathbb{E}h(W_\alpha(n, \mathbf{E})) - \mathbb{E}h(W_\alpha(m, \mathbf{E}))\big|.\end{aligned} \tag{3.8}$$

If M and N are identically distributed random variables, then

$$\big|\mathbb{E}h(W_\alpha(N, \mathbf{E})) - \mathbb{E}h(W_\alpha(N, \mathbf{X}))\big| \le \|h'\| \sum_{i=1}^{\infty} \mathbb{E}|E_i - X_i|\, \mathbb{P}(N \ge i). \tag{3.9}$$

Proof. We employ a Lindeberg argument, as follows. Defining $X_1, X_2, \ldots$ and $E_1, E_2, \ldots$ on the same probability space, we have

$$\big|\mathbb{E}h(W_\alpha(N, \mathbf{E})) - \mathbb{E}h(W_\alpha(M, \mathbf{X}))\big| \le \big|\mathbb{E}h(W_\alpha(N, \mathbf{E})) - \mathbb{E}h(W_\alpha(M, \mathbf{E}))\big| \tag{3.10}$$

$$+ \big|\mathbb{E}h(W_\alpha(M, \mathbf{E})) - \mathbb{E}h(W_\alpha(M, \mathbf{X}))\big|. \tag{3.11}$$

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

To bound (3.10), we simply note that

$$\begin{aligned}&|\mathbb{E}h(W_\alpha(N,\mathbf{E}))-\mathbb{E}h(W_\alpha(M,\mathbf{E}))|\\&\quad\le\|h'\|\sum_{m,n=1}^{\infty}\mathbb{P}(M=m,N=n)\,\mathbb{E}\big|W_\alpha(n,\mathbf{E})-W_\alpha(m,\mathbf{E})\big|.\end{aligned}\tag{3.12}$$

Now, to bound (3.11), if $\alpha=1$, then

$$\begin{aligned}&\big|\mathbb{E}h(\max\{E_1,\dots,E_M\})-\mathbb{E}h(\max\{X_1,\dots,X_M\})\big|\\&\le\sum_{m=1}^{\infty}\mathbb{P}(M=m)\sum_{i=1}^{m}\mathbb{E}\big|h(\max\{E_1,\dots,E_i,X_{i+1},\dots,X_m\})\\&\qquad\qquad-h(\max\{E_1,\dots,E_{i-1},X_i,\dots,X_m\})\big|\\&\le\sum_{m=1}^{\infty}\mathbb{P}(M=m)\sum_{i=1}^{m}\mathbb{E}|E_i-X_i|\|h'\|=\sum_{i=1}^{\infty}\mathbb{E}|E_i-X_i|\mathbb{P}(M\ge i)\|h'\|.\end{aligned}\tag{3.13}$$

Similarly, the minimum case ($\alpha=-1$) follows as $\min(X_i)\le\max(X_i)\le\sum X_i$. Adding (3.12) and (3.13) and taking the supremum over all $h\in\mathrm{Lip}_b(1)$ gives the first assertion. The second assertion follows from coupling M and N so that $M=N$ almost surely, in which case (3.12) vanishes. □

Propositions 1, 2 and 3 complement each other; the first two yield bounds in d_{TV} distance, for a general variable W, while the last result can be translated into a bound in d_{BW} distance, for comparing CCR distributions; see Remark 4 for more details. We note that Proposition 1 can be used to compare a maximum and a minimum, whereas Propositions 2 and 3 require the same value of α.

4. Application to the Poisson–exponential distribution

Cancho et al. [8] introduced the PE distribution as a distribution of the maximum of N i.i.d. exponential random variables from an infinite sequence $\mathbf{E}=(E_1,E_2,\dots)$ such that $E_i\sim\mathrm{Exp}(\lambda)$ with parameter λ (having mean $1/\lambda$) and N follows a zero-truncated Poisson distribution with parameter θ, independently of $\mathbf{E}$. This maximum has the PE distribution with parameters $\theta,\lambda>0$, denoted by $\mathrm{PE}(\theta,\lambda)$, which has the differentiable PDF

$$p(w|\theta,\lambda)=\frac{\theta\lambda\mathrm{e}^{-\lambda w-\theta\mathrm{e}^{-\lambda w}}}{1-\mathrm{e}^{-\theta}},\qquad w>0.\tag{4.1}$$

To obtain a Stein operator for $\mathrm{PE}(\theta,\lambda)$, we use (2.14) with $G_N(u)=\frac{\mathrm{e}^{-\theta}}{1-\mathrm{e}^{-\theta}}(\mathrm{e}^{\theta u}-1)$ and $\frac{G_N''(u)}{G_N'(u)}=\theta$, yielding the score function

$$\rho(w)=\lambda(\theta\mathrm{e}^{-\lambda w}-1).\tag{4.2}$$

Equation (2.15) gives

$$\mathcal{T}f(w)=f'(w)+\lambda(\theta\mathrm{e}^{-\lambda w}-1)f(w).\tag{4.3}$$

For a bounded test function $h\in\mathcal{H}$, the score $\mathrm{PE}(\theta,\lambda)$ Stein equation is

$$f'(w)+\lambda(\theta\mathrm{e}^{-\lambda w}-1)f(w)=h(w)-\mathbb{E}h(W),\tag{4.4}$$

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

with the solution f_h, given by (2.5), satisfying the following bounds.

Lemma 1. *Let $h : \mathbb{R}^+_{>0} \to \mathbb{R}$ be bounded and let f denote the solution (2.5) of the Stein equation (4.4) for h. Let $\tilde{h}(w) = h(w) - \mathbb{E}h(W)$ for $W \sim \mathrm{PE}(\theta, \lambda)$. Then for all $w > 0$,*

$$\left|\mathrm{e}^{-\lambda w} f(w)\right| \leq \frac{\|\tilde{h}\|}{\theta\lambda}\left(1 - \mathrm{e}^{-\theta+\theta\mathrm{e}^{-\lambda w}}\right) \tag{4.5}$$

$$\leq \frac{\|\tilde{h}\|}{\theta\lambda}; \tag{4.6}$$

$$\left|\lambda(\theta\mathrm{e}^{-\lambda w} - 1) f(w)\right| \leq \|\tilde{h}\|; \tag{4.7}$$

$$|f(w)| \leq \frac{2\|\tilde{h}\|}{\lambda}; \tag{4.8}$$

$$|f'(w)| \leq 2\|\tilde{h}\|. \tag{4.9}$$

If in addition $h \in Lip_b(1)$, then at all points w at which h' exists,

$$|f''(w)| \leq \|h'\| + 2\lambda\theta\|\tilde{h}\| + 3\lambda\|\tilde{h}\|. \tag{4.10}$$

Proof. We write p for the PDF of $\mathrm{PE}(\theta, \lambda)$. To prove (4.5) and (4.6), we bound

$$\left|\mathrm{e}^{-\lambda w} f(w)\right| \leq \|\tilde{h}\| \frac{\mathrm{e}^{-\lambda w}}{p(w)} \int_0^w p(t)\,\mathrm{d}t = \|\tilde{h}\| \frac{1 - \mathrm{e}^{-\theta(1-\mathrm{e}^{-\lambda w})}}{\theta\lambda},$$

and (4.5) follows. From $1 - \mathrm{e}^{-y} < 1$ for all $y > 0$, we get (4.6).

Proof of (4.7). Case 1: $\theta\mathrm{e}^{-\lambda w} - 1 > 0$. In this case $0 < w < \frac{\ln\theta}{\lambda}$ and we have

$$\left|\lambda(\theta\mathrm{e}^{-\lambda w} - 1) f(w)\right| \leq \|\tilde{h}\| \frac{\lambda(\theta\mathrm{e}^{-\lambda w} - 1)}{p(w)} \int_0^w p(t)\,\mathrm{d}t.$$

As $p'(t) = \lambda(\theta\mathrm{e}^{-\lambda t} - 1)p(t) \geq \lambda(\theta\mathrm{e}^{-\lambda w} - 1)p(t)$ for $0 < t < w < \frac{\ln\theta}{\lambda}$, it follows that

$$0 < \frac{\lambda(\theta\mathrm{e}^{-\lambda w} - 1)}{p(w)} \int_0^w p(t)\,\mathrm{d}t \leq \frac{1}{p(w)} \int_0^w p'(t)\,\mathrm{d}t = \left(1 - \mathrm{e}^{-\theta(1-\mathrm{e}^{-\lambda w})+\lambda w}\right) \leq 1.$$

Hence we obtain the bound (4.7) for $0 < w < \frac{\ln\theta}{\lambda}$.

Case 2: $\theta\mathrm{e}^{-\lambda w} - 1 \leq 0$. In this case $w \geq \frac{\ln\theta}{\lambda}$ and

$$0 < \lambda(1 - \theta\mathrm{e}^{-\lambda w})p(t) < \lambda(1 - \theta\mathrm{e}^{-\lambda t})p(t) = p'(t).$$

Using (2.5) gives

$$\left|\lambda(\theta\mathrm{e}^{-\lambda w} - 1) f(w)\right| \leq \|\tilde{h}\| \frac{\lambda(1 - \theta\mathrm{e}^{-\lambda w})}{p(w)} \int_w^\infty p(t)\,\mathrm{d}t \leq \|\tilde{h}\| \frac{1}{p(w)} \int_w^\infty p'(t)\,\mathrm{d}t \leq \|\tilde{h}\|.$$

Hence the bound (4.7) follows for all $w > 0$.

Proof of (4.8)–(4.10). As $\lambda(\theta\mathrm{e}^{-\lambda w} - 1)f(w) = \lambda\theta\mathrm{e}^{-\lambda w} f(w) - \lambda f(w)$, the triangle inequality along with (4.7) and (4.6) gives (4.8). To show (4.9), using the triangle inequality, from (4.4)

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

we obtain $|f'(w)| \le |h(w) - \mathbb{E}h(W)| + |\lambda(\theta e^{-\lambda w} - 1)f(w)|$, and using (4.7) yields the bound (4.9).

Now, for h differentiable at w, taking the first-order derivative in (4.4) gives $|f''(w)| \le |h'(w)| + |\lambda(\theta e^{-\lambda w} - 1)f'(w)| + |\theta\lambda^2 e^{-\lambda w} f(w)|$. Using (4.6) and (4.9), we obtain the bound (4.10) through

$$|f''(w)| \le \|h'\| + \lambda|\theta e^{-\lambda w} - 1|2\|\tilde{h}\| + \theta\lambda^2 \frac{\|\tilde{h}\|}{\theta\lambda} \le \|h'\| + 2\lambda\theta e^{-\lambda w}\|\tilde{h}\| + 2\lambda\|\tilde{h}\| + \lambda\|\tilde{h}\|.$$

This completes the proof. □

Remark 1. If $\theta \to 0$, the PE distribution converges to the exponential distribution Exp(λ); when $\lambda = 1$, (4.4) reduces to the Stein equation (4.2) in [27]. For this simplified version, the bound in [27] is only one-half of the bound (4.9); this discrepancy arises through our use of the triangle inequality for $\theta > 0$.

In the following subsections, we compare a PE distribution with a distribution of a maximum of a random number of i.i.d. random variables, with a generalised Poisson–exponential distribution, and with a Poisson–geometric distribution.

4.1. Approximating the distribution of the maximum of a random number of i.i.d. random variables by a PE distribution

Let $M \in \mathbb{N}$ be independent of $\mathbf{X} = (X_1, X_2, \dots)$, a sequence of i.i.d. random variables, and let $W = W_1(M, \mathbf{X})$ have a PDF of the form (2.12) for $\alpha = 1$. Our first comparison result employs Stein's method.

Corollary 1. *Assume that the X_i have differentiable PDF p_X, CDF F_X, and score function ρ_X. Let $W = W_1(M, \mathbf{X}) = \max\{X_1, \dots, X_M\}$. Then*

$$d_{\mathrm{TV}}\big(\mathrm{PE}(\theta, \lambda), \mathcal{L}(W_1(M, \mathbf{X}))\big) \le \frac{4}{\lambda}\left(\mathbb{E}\left|\theta\lambda e^{-\lambda W} - \frac{(G_M'' \circ F_X)(W)}{(G_M' \circ F_X)(W)} p_X(W)\right| + \mathbb{E}\big|-\lambda - \rho_X(W)\big|\right). \tag{4.11}$$

If M is a zero-truncated Poisson(θ_m) random variable, then (4.11) reduces to

$$d_{\mathrm{TV}}\big(\mathrm{PE}(\theta, \lambda), \mathcal{L}(W_1(M, \mathbf{X}))\big) \le \frac{4}{\lambda}\Big(\mathbb{E}\big|\theta\lambda e^{-\lambda W} - \theta_m p_X(W)\big| + \mathbb{E}\big|-\lambda - \rho_X(W)\big|\Big).$$

Proof. We employ Proposition 1. For a zero-truncated Poisson random variable N with parameter θ such that $G_N''(\cdot)/G_N'(\cdot) = \theta$ and for $Y \sim \mathrm{Exp}(\lambda)$ with PDF $f_Y(y) = \lambda e^{-\lambda y}$, using (3.1) along with (4.8) and taking h to be an indicator function so that $\|\tilde{h}\| \le 2\|h\| \le 2$ gives (4.11). The simplification when M is a zero-truncated Poisson random variable follows from $G_M''(\cdot)/G_M'(\cdot) = \theta_M$. □

Remark 2. As $-\lambda$ is the score function of the exponential distribution Exp(λ), for M being a zero-truncated Poisson(θ) random variable, the bound in Corollary 1 is close to zero if the density and the score function of X are close to those of Exp(λ).

The next result is based on the Lindeberg argument.

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

Corollary 2. *Let* $W_1(N, \mathbf{E}) \sim \mathrm{PE}(\theta, \lambda)$ *and let* $W_1(M, \mathbf{X}) = \max(X_1, \ldots, X_M)$ *have a CCR distribution. Then*

$$d_{\mathrm{BW}}\big(\mathrm{PE}(\theta, \lambda), \mathcal{L}(W_1(M, \mathbf{X})\big) \leq \mathbb{E}M\, \mathbb{E}|E_1 - X_1| + \sum_{m,n=1}^{\infty} \mathbb{P}(M = m, N = n) \left(H_{\max\{m,n\}} - H_{\min\{m,n\}}\right),$$

where H_n *is the nth harmonic number.*

If M *is also a zero-truncated* Poisson(θ) *random variable, then*

$$d_{\mathrm{BW}}(\mathrm{PE}\big(\theta, \lambda), \mathcal{L}(W_1(M, \mathbf{X})\big) \leq \frac{\theta}{1 - \mathrm{e}^{-\theta}}\, \mathbb{E}|E_1 - X_1|.$$

Proof. Proposition 3 gives

$$\begin{aligned}&\left|\mathbb{E}h(W_1(N, \mathbf{E})) - \mathbb{E}h(W_1(M, \mathbf{X}))\right| \\ &\leq \sum_{i=1}^{\infty} \mathbb{E}|E_i - X_i|\, \mathbb{P}(M \geq i)\, \|h'\| \\ &\quad + \sum_{m,n=1}^{\infty} \mathbb{P}(M = m, N = n) \left|\mathbb{E}h(W_1(n, \mathbf{E})) - \mathbb{E}h(W_1(m, \mathbf{E}))\right| \|h'\|.\end{aligned}$$

Since the random variables in $\mathbf{X}$ are i.i.d., those in $\mathbf{E}$ are i.i.d. exponential with parameter λ, and the expectation of the maximum of n exponential random variables is $\sum_{i=1}^{n} \frac{1}{i} = H_n$, the assertion follows. □

4.2. Approximating the generalised Poisson–exponential distribution

The PE distribution has an increasing or constant failure rate. To model decreasing failure rate as well, Fatima & Roohi [12] introduced the family of generalised Poisson–exponential (GPE) distributions. The differentiable PDF of a GPE distribution with parameters $\beta, \theta, \lambda > 0$, denoted by GPE($\theta, \lambda, \beta$), is

$$p(x|\theta, \lambda, \beta) = \frac{\beta\theta\lambda \mathrm{e}^{-\lambda x - \theta\beta \mathrm{e}^{-\lambda x}}}{(1 - \mathrm{e}^{-\theta})^{\beta}} \left(1 - \mathrm{e}^{-\theta + \theta \mathrm{e}^{-\lambda x}}\right)^{\beta - 1}, \quad x > 0. \tag{4.12}$$

For $\beta = 1$ the density of the GPE distribution simplifies to that of the PE distribution given in (4.1). For $0 < \beta < 1$ the PDF of the GPE distribution is monotonically decreasing, while for $\beta \geq 1$ it is unimodal positively skewed with skewness depending on the shape parameters β and θ. The shape of the hazard function also depends on these two shape parameters. For example, for $\theta = 1$ and $\lambda = 2$, the failure rate is decreasing for $0 < \beta < 1$ and increasing for $\beta \geq 1$, as in Figure 2 of [12].

For a data set from [1] which consists of 50 observations on the time to first failure of devices, in [12] it was shown that the GPE distribution provides a better fit than the PE and some other candidate distributions. However, a GPE distribution is not as easy to manipulate and interpret as a PE distribution. Therefore, a natural question is how to quantify the sacrifice when approximating a GPE distribution with a PE distribution. Here we note that GPE distributions are not from the CCR family, and hence Proposition 3 cannot be applied. Instead we use Stein's method to bound the approximation error.

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

For such an approximation to be intuitive, the failure rate of the approximating distribution should be qualitatively similar. Hence we restrict attention to the case of $\beta \geq 1$, for which both the GPE and the PE distributions have an increasing failure rate. As an aside, we note that for $\beta \geq 1$, the limit of the PDF at 0 is 0, while for $0 < \beta < 1$ it is undefined; the latter condition leads to a Stein class for the GPE Stein score operator which differs from the Stein class for the PE Stein score operator.

We note here that if $X \sim \mathrm{GPE}(\theta, \lambda, \beta)$, then for $\beta \geq 1$ we have

$$\mathbb{E}X \leq \frac{\beta\theta}{\lambda(1-\mathrm{e}^{-\theta})^{\beta}}. \tag{4.13}$$

The proof of (4.13) is given in the appendix.

The GPE random variable is not of the CCR form (2.8), but its score function, given in (4.14) below, can be derived from its PDF (4.12) with parameters $\beta, \theta, \lambda > 0$ as

$$\rho(x) = \frac{p'(x)}{p(x)} = \lambda\theta\mathrm{e}^{-\lambda x}\left[\beta + \frac{(\beta-1)\,\mathrm{e}^{-\theta+\theta\mathrm{e}^{-\lambda x}}}{1-\mathrm{e}^{-\theta+\theta\mathrm{e}^{-\lambda x}}}\right] - \lambda, \quad x > 0. \tag{4.14}$$

In the following theorem we bound the distance between a GPE with $\beta \geq 1$ and a PE distribution, using their corresponding score Stein operators in (2.6).

Theorem 1. *Let $W \sim \mathrm{PE}(\theta_1, \lambda_1)$ and $X \sim \mathrm{GPE}(\theta_2, \lambda_2, \beta)$, let $h : \mathbb{R}^{+}_{>}0 \to \mathbb{R}$ be bounded, and let $\tilde{h}(x) = h(x) - \mathbb{E}h(W)$. Then for $\lambda_1 \leq \lambda_2$ and $\beta \geq 1$,*

$$\left|\mathbb{E}h(X) - \mathbb{E}h(W)\right| \leq \|\tilde{h}\| \left\{\frac{\lambda_2}{\lambda_1}|\beta-1| + \left|\frac{\lambda_2\theta_2}{\lambda_1\theta_1}\beta - 1\right| + \left(\frac{\lambda_2}{\lambda_1} - 1\right)(\lambda_1\mathbb{E}X + 2)\right\}. \tag{4.15}$$

Proof. Let p_W and p_X denote the PDFs of W and X, and let $\mathcal{T}_X$ denote a score Stein operator for a GPE distribution. To employ (2.6), we first check that f_W as in (2.5) for the PE distribution, for h bounded, is in the Stein class of $\mathcal{T}_X$. Invoking Lemma 1, f_W is bounded. Now $\mathbb{E}[\mathcal{T}_X f_W(X)] = \int_0^{\infty} \frac{(f_W p_X)'(x)}{p_X(x)} p_X(x)\,\mathrm{d}x = \lim_{x\to\infty}(f_W p_X)(x) - \lim_{x\to 0}(f_W p_X)(x)$, and for $\beta \geq 1$ we have that $(f_W p_X)(x) \to 0$ as $x \to 0$ and as $x \to \infty$, showing that f_W is in the Stein class for $\mathcal{T}_X$. Applying (2.6) with the score functions from (4.2) and (4.14), we have

$$\begin{aligned}
&\mathbb{E}h(X) - \mathbb{E}h(W) \\
&= \mathbb{E}\left[\lambda_2\theta_2\mathrm{e}^{-\lambda_2 X}\left(\beta + \frac{(\beta-1)\mathrm{e}^{-\theta_2+\theta_2\mathrm{e}^{-\lambda_2 X}}}{1-\mathrm{e}^{-\theta_2+\theta_2\mathrm{e}^{-\lambda_2 X}}}\right) - \lambda_2 - \lambda_1(\theta_1\mathrm{e}^{-\lambda_1 X} - 1)\right] f_W(X) \\
&\leq \mathbb{E}\left[\lambda_2\theta_2(\beta-1)\mathrm{e}^{(\lambda_1-\lambda_2)X}\frac{\mathrm{e}^{-\theta_2+\theta_2\mathrm{e}^{-\lambda_2 X}}}{1-\mathrm{e}^{-\theta_2+\theta_2\mathrm{e}^{-\lambda_2 X}}}\,\mathrm{e}^{-\lambda_1 X}|f_W(X)|\right] \\
&\quad + \mathbb{E}\left|\lambda_2\theta_2\beta\mathrm{e}^{-\lambda_2 X} - \lambda_1\theta_1\mathrm{e}^{-\lambda_1 X}\right| |f_W(X)| + \mathbb{E}|\lambda_1 - \lambda_2|\,|f_W(X)|.
\end{aligned}$$

Next,

$$|\theta_2\lambda_2\mathrm{e}^{-\lambda_2 x} - \theta_1\lambda_1\mathrm{e}^{-\lambda_1 x}| \leq |\theta_2\lambda_2\beta - \theta_1\lambda_1|\,\mathrm{e}^{(\lambda_1-\lambda_2)x}\mathrm{e}^{-\lambda_1 x} + (\lambda_2 - \lambda_1)x\theta_1\lambda_1\mathrm{e}^{-\lambda_1 x}, \tag{4.16}$$

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

since $|e^{-x}-1| \le x$ for all $x \ge 0$. With $e^{(\lambda_1-\lambda_2)x} \le 1$ for $\lambda_1 \le \lambda_2$, we write

$$\begin{aligned}
&|\mathbb{E}h(X) - \mathbb{E}h(W)| \\
&\le \lambda_2\theta_2|\beta-1|\,\mathbb{E}\left|\frac{e^{-\theta_2+\theta_2e^{-\lambda_2X}}}{1-e^{-\theta_2+\theta_2e^{-\lambda_2X}}}e^{-\lambda_1X}f_W(X)\right| + (\lambda_2-\lambda_1)\,\mathbb{E}|f_W(X)| \\
&\quad + \lambda_1\theta_1\left|\frac{\lambda_2\theta_2}{\lambda_1\theta_1}\beta - 1\right|\mathbb{E}|e^{-\lambda_1X}f_W(X)| + \lambda_1\theta_1(\lambda_2-\lambda_1)\,\mathbb{E}X|e^{-\lambda_1X}f_W(X)|. \qquad (4.17)
\end{aligned}$$

Now, using (4.5) as well as $e^x - 1 \ge x$ and $1 - e^{-x} \le x$ for $x \ge 0$, we have

$$\begin{aligned}
\left|\frac{e^{-\theta_2+\theta_2e^{-\lambda_2x}}}{1-e^{-\theta_2+\theta_2e^{-\lambda_2x}}}e^{-\lambda_1x}f_W(x)\right| &\le \frac{\|\tilde{h}\|}{\lambda_1\theta_1}\frac{e^{-\theta_2+\theta_2e^{-\lambda_2x}}}{1-e^{-\theta_2+\theta_2e^{-\lambda_2x}}}\,(1-e^{-\theta_1(1-e^{-\lambda_1x})}) \\
&\le \frac{\|\tilde{h}\|}{\lambda_1\theta_2}. \qquad (4.18)
\end{aligned}$$

Using (4.6), (4.8), (4.13) and (4.18) in (4.17) gives the bound (4.15). □

Remark 3. For $\lambda_1 = \lambda_2$ and $\theta_1 = \theta_2$ in Theorem 1, the bound (4.15) can be improved by a factor of 2, using that with f_W being the solution of the PE score Stein equation,

$$\big|\mathbb{E}h(X) - \mathbb{E}h(W)\big| = \mathbb{E}\left(\lambda\theta|\beta-1|\frac{e^{-\lambda X}}{1-e^{-\theta+\theta e^{-\lambda X}}}f_W(X)\right).$$

With (4.5) and $\tilde{h}(x) = h(x) - \mathbb{E}h(W)$ we obtain

$$\big|\mathbb{E}h(X) - \mathbb{E}h(W)\big| \le \|\tilde{h}\|\,|\beta-1|. \qquad (4.19)$$

This bound depends solely on the parameter β and tends to 0 as $\beta \to 1$, which is in line with the fact that for $\beta \to 1$, GPE(θ, λ, β) converges to PE(θ, λ); see [12].

The next result follows immediately from Theorem 1 and (4.13).

Corollary 3. *Let* $W_1 \sim \mathrm{PE}(\theta_1, \lambda_1)$ *and* $W_2 \sim \mathrm{PE}(\theta_2, \lambda_2)$ *with* $\lambda_1 \le \lambda_2$. *Let* $\mathcal{H} = \{h : \mathbb{R}^+_{>0} \to \mathbb{R},\ \|h\| \le 1\}$. *Then for all* $h \in \mathcal{H}$, *letting* $\tilde{h}(w) = h(w) - \mathbb{E}h(W)$, *we have from (4.15) with* $\beta = 1$ *that*

$$\big|\mathbb{E}h(W_2) - \mathbb{E}h(W_1)\big| \le \|\tilde{h}\|\left\{\left|\frac{\lambda_2\theta_2}{\lambda_1\theta_1} - 1\right| + \left(\frac{\lambda_2}{\lambda_1} - 1\right)\left(\frac{\lambda_1\theta_2}{\lambda_2(1-e^{-\theta_2})} + 2\right)\right\}. \qquad (4.20)$$

Remark 4.

(i) For $\|h\| \le 1$ such that $\|\tilde{h}\| \le 2$, the bounds can easily be converted into bounds in total variation distance using (2.2).

(ii) To bound $d_{\mathcal{H}}(\mathrm{PE}(\theta_1, \lambda_1), \mathrm{GPE}(\theta_2, \lambda_2, \beta))$ when $\lambda_1 > \lambda_2$, we can use

$$\begin{aligned}
d_{\mathcal{H}}\big(\mathrm{GPE}(\theta_2, \lambda_2, \beta), \mathrm{PE}(\theta_1, \lambda_1)\big) &\le d_{\mathcal{H}}\big(\mathrm{GPE}(\theta_2, \lambda_2, \beta), \mathrm{PE}(\theta_2, \lambda_2)\big) \\
&\quad + d_{\mathcal{H}}\big(\mathrm{PE}(\theta_2, \lambda_2), \mathrm{PE}(\theta_1, \lambda_1)\big)
\end{aligned}$$

and apply Remark 3 and Corollary 3.

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

4.3. Approximating the Poisson–geometric distribution

Next we consider the distribution of $W_G = \max\{T_1, \ldots, T_M\}$, where $T_1, T_2, \ldots \in \{1, 2, \ldots\}$ are i.i.d. Geometric(p) random variables and M is an independent zero-truncated Poisson(θ) random variable; the distribution of W_G is the *Poisson–geometric (PG) distribution*; it has PMF

$$\mathbb{P}(W_G = w) = \frac{\mathrm{e}^{-(1-p)^w\theta} - \mathrm{e}^{-(1-p)^{w-1}\theta}}{(1-\mathrm{e}^{-\theta})}, \quad w \in \mathbb{N} \tag{4.21}$$

and the discrete backward score function

$$\rho_{W_G}(w) = 1 - \frac{\mathrm{e}^{-(1-p)^{w-1}\theta} - \mathrm{e}^{-(1-p)^{w-2}\theta}}{\mathrm{e}^{-(1-p)^w\theta} - \mathrm{e}^{-(1-p)^{w-1}\theta}}, \quad w \in \mathbb{N}.$$

For $T \sim$ Geometric(λ/n), the distribution of $n^{-1}T$ converges to Exp(λ) in probability, and hence it is plausible to approximate the distribution of $Z_n = W_{G,n}/n$, for $W_{G,n} \sim \mathrm{PG}(\theta, \lambda/n)$, by a corresponding PE distribution. With $q_n = 1 - \lambda/n$ and $\tilde{\rho}_n(z) = \rho_{W_G}(nz)$,

$$\tilde{\rho}_n(z) = 1 - \frac{\mathrm{e}^{-q_n^{nz-1}\theta} - \mathrm{e}^{-q_n^{nz-2}\theta}}{\mathrm{e}^{-q_n^{nz}\theta} - \mathrm{e}^{-q_n^{nz-1}\theta}}, \quad nz \in \mathbb{N}.$$

This function is the ratio of two exponential functions, complicating the comparison using (2.6). To simplify the comparison, we use Proposition 2. From (3.4), a standardised PG Stein operator for Z_n is

$$\mathcal{T}^{(d)}_{Z_n}(g(z)) = g(z)d(z) - g\left(z - \frac{1}{n}\right)d\left(z - \frac{1}{n}\right) + \tilde{\rho}_n(z)g\left(z - \frac{1}{n}\right)d\left(z - \frac{1}{n}\right); \tag{4.22}$$

here we choose

$$d(z) = \mathrm{e}^{-q_n^{nz+1}\theta} - \mathrm{e}^{-q_n^{nz}\theta} = \mathrm{e}^{-\theta q_n^{nz}}\left(\mathrm{e}^{-\theta(q_n-1)q_n^{nz}} - 1\right) \tag{4.23}$$

so that $\tilde{\rho}_n(z)d\left(z - \frac{1}{n}\right) = \mathrm{e}^{-q_n^{nz}\theta} - 2\mathrm{e}^{-q_n^{nz-1}\theta} + \mathrm{e}^{-q_n^{nz-2}\theta}$. As $nd(z) \to \lambda\theta\mathrm{e}^{-\lambda z - \theta\mathrm{e}^{-\lambda z}}$ as $n \to \infty$, for the approximating PE distribution we choose

$$c(z) = \frac{\lambda\theta}{n}\mathrm{e}^{-\lambda z - \theta\mathrm{e}^{-\lambda z}}, \quad z > 0$$

as a standardisation function in (3.3), giving rise to the standardised PE Stein equation

$$c(w)g'(w) + \left[c(w)\lambda(\theta\mathrm{e}^{-\lambda w} - 1) + c'(w)\right]g(w) = h(w) - \mathbb{E}h(W) \tag{4.24}$$

for $W \sim \mathrm{PE}(\theta, \lambda)$. Again we can bound the solution of this Stein equation as follows.

Lemma 2. *For $W \sim \mathrm{PE}(\theta, \lambda)$, the solution $g(w)$ of the Stein equation (4.24), bounded differentiable h such that $\|h\| \le 1$ and $\|h'\| \le 1$ with $\tilde{h}(x) = h(x) - \mathbb{E}h(W)$, and $c(w) = \frac{\lambda\theta}{n}\,\mathrm{e}^{-\lambda w - \theta\mathrm{e}^{-\lambda w}}$, we have*

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

$$\left|\mathrm{e}^{-2\lambda w-\theta\mathrm{e}^{-\lambda w}}g(w)\right| \le \frac{n}{\lambda^2\theta^2}\|\tilde{h}\|, \tag{4.25}$$

$$\left|\mathrm{e}^{-\lambda w-\theta\mathrm{e}^{-\lambda w}}(\theta\mathrm{e}^{-\lambda w}-1)g(w)\right| \le \frac{n}{\lambda^2\theta}\|\tilde{h}\|, \tag{4.26}$$

$$\left|\mathrm{e}^{-\lambda w-\theta\mathrm{e}^{-\lambda w}}g(w)\right| \le 2\frac{n}{\lambda^2\theta}\|\tilde{h}\|, \tag{4.27}$$

$$\left|\mathrm{e}^{-\lambda w-\theta\mathrm{e}^{-\lambda w}}g'(w)\right| \le 3\frac{n}{\lambda\theta}\|\tilde{h}\|, \tag{4.28}$$

$$\left|\frac{\lambda\theta}{n}\mathrm{e}^{-\lambda w-\theta\mathrm{e}^{-\lambda w}}g''(w)\right| \le \|h'\| + 9\lambda\theta\|\tilde{h}\| + 11\lambda\|\tilde{h}\|, \tag{4.29}$$

and

$$\frac{c(x)}{c\left(x+\frac{\rho}{n}\right)} \le \begin{cases} \mathrm{e}^{\theta(\mathrm{e}^{\frac{\lambda}{n}}-1)}, & \rho=-1, \\ \mathrm{e}^{\frac{\lambda}{n}}, & 0<\rho<1. \end{cases} \tag{4.30}$$

Proof. In (4.24), $cg = f$ is the solution of the Stein equation (4.4), so we use the bounds for f to bound g. The bound (4.6) in Lemma 1 immediately gives (4.25). Also $c'(w) = c(w)\lambda(\theta\mathrm{e}^{-\lambda w}-1)$ so that $c'(w)g(w) = \lambda(\theta\mathrm{e}^{-\lambda w}-1)f(w)$; using (4.7) we get (4.26). Combining (4.25), (4.26), and the triangle inequality gives (4.27).

Since $(cg)' = cg' + c'g$, rearranging and using (4.9) and (4.7) with the triangle inequality gives (4.28). For $(cg)'' - c''g - 2c'g' = cg''$, using (4.28) we have

$$|c'(w)g'(w)| \le \lambda|\theta\mathrm{e}^{-\lambda w}-1|c(w)g'(w) \le 3\lambda(\theta+1)\|\tilde{h}\|.$$

For $c''(w)g(w) = \lambda^2(\theta\mathrm{e}^{-\lambda w}-1)^2c(w)g(w) - \lambda^2\theta\mathrm{e}^{-\lambda w}c(w)g(w)$, using the triangle inequality, (4.26), and (4.25) yields

$$|c''(w)g(w)| \le \lambda|\theta\mathrm{e}^{-\lambda w}-1|\|\tilde{h}\| + \lambda\|\tilde{h}\| \le \lambda\theta\|\tilde{h}\| + 2\lambda\|\tilde{h}\|.$$

These two results along with (4.10) give (4.29). Now, for any $0<\rho<1$,

$$\frac{c(x)}{c\left(x+\frac{\rho}{n}\right)} = \mathrm{e}^{\lambda\frac{\rho}{n}-\theta\mathrm{e}^{-\lambda x}(1-\mathrm{e}^{-\lambda\frac{\rho}{n}})} \le \mathrm{e}^{\lambda\frac{\rho}{n}};$$

for $\rho=-1$ we have $\frac{c(x)}{c\left(x-\frac{1}{n}\right)} = \mathrm{e}^{-\frac{\lambda}{n}+\theta\mathrm{e}^{-\lambda x}(\mathrm{e}^{\frac{\lambda}{n}}-1)} \le \mathrm{e}^{\theta(\mathrm{e}^{\frac{\lambda}{n}}-1)}$, yielding (4.30). □

Theorem 2. *Let $W_{G,n} \sim \mathrm{PG}(\theta, p_n)$ with $p_n = \lambda/n$ where $0<\lambda<n$, and let $W \sim \mathrm{PE}(\theta,\lambda)$. Then for the scaled PG random variable $Z_n = W_{G,n}/n$ and any bounded function h with bounded first derivative, we have*

$$\begin{aligned} &\left|\mathbb{E}h(Z_n) - \mathbb{E}h(W)\right| \\ &\le \frac{\mathrm{e}^{\frac{\lambda}{n}}}{n}\left(1+\frac{\mathrm{e}^{\frac{\theta\lambda}{n\mathrm{e}}}}{n}B_1(\theta,\lambda)\right)\|h'\| \\ &\quad + \frac{1}{n}\left(\left(1-\frac{\lambda}{n}\right)^{-6}\mathrm{e}^{\theta\mathrm{e}^{\frac{\lambda}{n}}+\frac{\theta\lambda}{n}+2\frac{\theta\lambda}{n\mathrm{e}}+\frac{\lambda}{n}}B_2(\theta,\lambda) + \frac{1}{2n}\mathrm{e}^{\frac{\lambda}{n}+\frac{\theta\lambda}{n\mathrm{e}}}B_3(\theta,\lambda)\right)\|\tilde{h}\|, \end{aligned} \tag{4.31}$$

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

where

$$B_1(\theta,\lambda) = \lambda \max(1,\theta) + \frac{e^{\lambda\theta+\frac{\theta}{e}}}{2\theta\lambda} + \frac{\theta\lambda}{2e}\frac{(1+e-e^{-\theta})}{(1-e^{-\theta})},$$

$$B_2(\theta,\lambda) = 11\lambda + 9\lambda\theta + 6\lambda\max(1,\theta) + \frac{3e^{\lambda\theta+\frac{\theta}{e}}}{\theta\lambda} + 3\theta\lambda\frac{(1+e-e^{-\theta})}{(1-e^{-\theta})e} + 4\lambda e^{-\theta}\left(1 + 2\max(1,\theta) + \frac{(3e+1)\theta}{e} + \frac{(5e+3)\theta^2}{3e} + \frac{\theta(2\theta+1)}{(1-e^{-\theta})}\right),$$

$$B_3(\theta,\lambda) = (9\lambda\theta + 11\lambda)\left(2\lambda\max(1,\theta) + \frac{e^{\lambda\theta+\frac{\theta}{e}}}{\theta\lambda} + \theta\lambda\frac{(1+e-e^{-\theta})}{(1-e^{-\theta})e}\right).$$

Remark 5.

(i) For fixed λ and θ, the bound (4.31) is of $O(n^{-1})$. As $n\to\infty$, for $\lambda = \lambda(n)$ and $\theta = \theta(n)$ the bound decreases to 0 as long as $\lambda(n)\theta(n)/n \to 0$ and $\lambda(n)/n \to 0$.

(ii) Equation (4.31) can be translated into a bound in the bounded Wasserstein distance using $\mathrm{Lip}_b(1)$ as the class of test functions.

Proof. We employ Proposition 2. First we note that $c(w)\rho(w) = c'(w)$ and that $\tilde{\rho}_n(w)d\left(w-\frac{1}{n}\right) = d\left(w-\frac{1}{n}\right) - d\left(w-\frac{2}{n}\right)$. Thus, for (3.6), with $h\in \mathrm{Lip}_b(1)$ we have

$$\begin{aligned}
&\left|\mathbb{E}h(Z_n) - \mathbb{E}h(W)\right| \\
&\le \left|\mathbb{E}[n\Delta^{-n}(dg)(Z_n) - (cg)'(Z_n)] + \mathbb{E}\left[n\tilde{\rho}_n(Z_n)(dg)\left(Z_n - \frac{1}{n}\right) - (cg)(Z_n)\,\rho(Z_n)\right]\right| \\
&\le \mathbb{E}\left|n\left\{g(Z_n) - g\left(Z_n - \frac{1}{n}\right)\right\}c(Z_n) - g'(Z_n)c(Z_n)\right| \qquad (4.32)\\
&\quad + \mathbb{E}\left|n\left\{g(Z_n) - g\left(Z_n - \frac{1}{n}\right)\right\}\left(d(Z_n) - c(Z_n) - \frac{2}{n}c'(Z_n)\right)\right| \qquad (4.33)\\
&\quad + \mathbb{E}\left|d(Z_n) - d\left(Z_n - \frac{2}{n}\right) - \frac{2}{n}c'(Z_n)\right|\left|ng\left(Z_n - \frac{1}{n}\right)\right|. \qquad (4.34)
\end{aligned}$$

To bound the term (4.32), for some $0<\rho<1$ we write

$$\begin{aligned}
\mathbb{E}\left|n\left\{g(Z_n) - g\left(Z_n - \frac{1}{n}\right)\right\}c(Z_n) - g'(Z_n)c(Z_n)\right| &\le \frac{1}{n}\,\mathbb{E}\left|c(Z_n)g''\left(Z_n + \frac{\rho}{n}\right)\right| \\
&\le \frac{1}{n}\|cg''\|\,\mathbb{E}\left|\frac{c(Z_n)}{c\left(Z_n+\frac{\rho}{n}\right)}\right|.
\end{aligned}$$

Then (4.29) and (4.30) give the bound

$$\mathbb{E}\left|n\{g(Z_n) - g(Z_n - n^{-1})\}c(Z_n) - g'(Z_n)c(Z_n)\right| \le \frac{e^{\frac{\lambda}{n}}}{n}\left\{\|h'\| + 9\lambda\theta\|\tilde{h}\| + 11\lambda\|\tilde{h}\|\right\}. \qquad (4.35)$$

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

To bound (4.33), we let $\tau(z) = \mathrm{e}^{\lambda z + \theta \mathrm{e}^{-\lambda z}}$, so that $\tau^{-1} g = \frac{n}{\theta\lambda} cg$ can be bounded as in Lemma 2, and write

$$\begin{aligned}&\mathbb{E}\left|n\left\{g(Z_n) - g\left(Z_n - \frac{1}{n}\right)\right\}\left(d(Z_n) - c(Z_n) - \frac{2}{n}c'(Z_n)\right)\right| \\ &\quad = \mathbb{E}\left|l(Z_n) n\left\{g(Z_n) - g\left(Z_n - \frac{1}{n}\right)\right\}\tau^{-1}(Z_n)\right|,\end{aligned}$$

where $l(Z_n) = \tau(Z_n)\big(d(Z_n) - c(Z_n) - \frac{2}{n}c'(Z_n)\big)$. We show in Appendix that

$$|l(Z_n)| \le \frac{2\lambda^2\theta}{n^2}\max(1,\theta) + \frac{1}{n^2}\mathrm{e}^{\lambda\theta+\frac{\theta}{\mathrm{e}}} + \frac{\theta^2\lambda^2}{n^2}\mathrm{e}^{\frac{\theta\lambda}{n\mathrm{e}}-1} + \frac{\theta\lambda^3 Z_n}{n^2}\mathrm{e}^{\frac{\theta\lambda}{n\mathrm{e}}}. \tag{4.36}$$

By Taylor expansion, for some $0 < \epsilon < 1$ we have

$$\begin{aligned}n\left\{g(Z_n) - g\left(Z_n - \frac{1}{n}\right)\right\}\tau^{-1}(Z_n) &= g'(Z_n)\tau^{-1}(Z_n) + \frac{1}{2n}g''\left(Z_n + \frac{\epsilon}{n}\right)\tau^{-1}(Z_n) \\ &= \frac{n}{\theta\lambda}\|cg'\| + \frac{1}{2\theta\lambda}\|cg''\|\frac{c(Z_n)}{c\left(Z_n + \frac{\epsilon}{n}\right)}.\end{aligned} \tag{4.37}$$

Using (4.36) and (4.37) along with (4.28), (4.29) and (4.30) yields

$$\begin{aligned}&\mathbb{E}\left|n\left\{g(Z_n) - g\left(Z_n - \frac{1}{n}\right)\right\}\left(d(Z_n) - c(Z_n) - \frac{2}{n}c'(Z_n)\right)\right| \\ &\le \left(\frac{2\lambda^2\theta}{n^2}\max(1,\theta) + \frac{1}{n^2}\mathrm{e}^{\lambda\theta+\frac{\theta}{\mathrm{e}}} + \frac{\theta^2\lambda^2}{n^2}\mathrm{e}^{\frac{\theta\lambda}{n\mathrm{e}}-1} + \frac{\theta\lambda^3\mathbb{E}Z_n}{n^2}\mathrm{e}^{\frac{\theta\lambda}{n\mathrm{e}}}\right) \\ &\quad \times\left\{\frac{n}{\theta\lambda}3\|\tilde{h}\| + \frac{1}{2\theta\lambda}\{\|h'\| + 9\lambda\theta\|\tilde{h}\| + 11\lambda\|\tilde{h}\|\}\mathrm{e}^{\frac{\lambda}{n}}\right\}.\end{aligned} \tag{4.38}$$

Finally, to bound (4.34), we show in Appendix that

$$\begin{aligned}&\left|\frac{2}{n}c'(z) - d(z) + d\left(z - \frac{2}{n}\right)\right|\left|ng\left(z - \frac{1}{n}\right)\right| \\ &\le 2\lambda\left\{\left(1 - \frac{\lambda}{n}\right)^{-2} - 1\right\}\max(1,\theta)c(z)\left|g\left(z - \frac{1}{n}\right)\right| \\ &\quad + \frac{\theta\lambda^2}{n}\left(1 - \frac{\lambda}{n}\right)^{-2}\mathrm{e}^{\frac{\theta\lambda}{n\mathrm{e}}}\left\{\frac{1}{3}\left(1 - \frac{\lambda}{n}\right)^{-1}\left[\theta + \theta\mathrm{e}^{\frac{\theta\lambda}{n}} + 6\mathrm{e}^{\frac{\theta\lambda}{n}} + 12\left(1 - \frac{\lambda}{n}\right)^{-1}\right.\right. \\ &\qquad \left.\left. + 8\theta\left(1 - \frac{\lambda}{n}\right)^{-3}\right] + 2\left(\frac{\theta}{\mathrm{e}} + 2\lambda z\right)\right\}c(z)\left|g\left(z - \frac{1}{n}\right)\right| \\ &\quad + \frac{2\lambda^2}{n}\left(1 - \frac{\lambda}{n}\right)^{-2}\mathrm{e}^{\frac{\theta\lambda}{n\mathrm{e}}}\left[\left(1 - \frac{\lambda}{n}\right)^{-1}\mathrm{e}^{\frac{\theta\lambda}{n}} + \lambda z + \frac{\theta}{\mathrm{e}}\right]c(z)\left|g\left(z - \frac{1}{n}\right)\right|.\end{aligned} \tag{4.39}$$

Note that

$$0 \le \left(1 - \frac{\lambda}{n}\right)^{-2} - 1 = \frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right)^{-2}\left(2 - \frac{\lambda}{n}\right) \tag{4.40}$$

and that, using (4.27) and (4.30),

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

$$c(z)\left|g\left(z-\frac{1}{n}\right)\right| = \left|c(z)g\left(z-\frac{1}{n}\right)\right| \le \|cg\|\left|\frac{c(z)}{c\left(z-\frac{1}{n}\right)}\right| \le \frac{2}{\lambda}\mathrm{e}^{\theta(\mathrm{e}^{\frac{\lambda}{n}}-1)}\|\tilde{h}\|.$$

Combining this with (4.39) and (4.40), we bound (4.34) as

$$\begin{aligned}
&\mathbb{E}\left|d(Z_n)-d\left(Z_n-\frac{2}{n}\right)-\frac{2}{n}c'(Z_n)\right|\left|ng\left(Z_n-\frac{1}{n}\right)\right| \qquad (4.41)\\
&\le \frac{4\lambda}{n}\max(1,\theta)\left(1-\frac{\lambda}{n}\right)^{-2}\left|\frac{\lambda}{n}-2\right|\mathrm{e}^{\theta(\mathrm{e}^{\frac{\lambda}{n}}-1)}\|\tilde{h}\| + \frac{2\theta\lambda}{n}\left(1-\frac{\lambda}{n}\right)^{-2}\mathrm{e}^{\frac{\theta\lambda}{n\mathrm{e}}}\mathrm{e}^{\theta(\mathrm{e}^{\frac{\lambda}{n}}-1)}\|\tilde{h}\|\\
&\quad\times\left\{\frac{1}{3}\left(1-\frac{\lambda}{n}\right)^{-1}\left[\theta+\theta\mathrm{e}^{\frac{\theta\lambda}{n}}+6\mathrm{e}^{\frac{\theta\lambda}{n}}+12\left(1-\frac{\lambda}{n}\right)^{-1}+8\theta\left(1-\frac{\lambda}{n}\right)^{-3}\right]\right.\\
&\quad\left.+2\left(\frac{\theta}{\mathrm{e}}+2\lambda\mathbb{E}Z_n\right)+\frac{2}{\theta}\left[\left(1-\frac{\lambda}{n}\right)^{-1}\mathrm{e}^{\frac{\theta\lambda}{n}}+\lambda\mathbb{E}Z_n+\frac{\theta}{\mathrm{e}}\right]\right\}.
\end{aligned}$$

To bound $\mathbb{E}Z_n$, we argue as for (4.13); $W_{G,n}=\max\{T_1,\dots,T_M\}\le\sum_{i=1}^M T_i$ and so $\mathbb{E}W_{G,n}\le\mathbb{E}\sum_{i=1}^M T_i=\frac{n\theta}{\lambda(1-\mathrm{e}^{-\theta})}$, giving that $\mathbb{E}Z_n=\frac{1}{n}\mathbb{E}W_{G,n}\le\frac{\theta}{\lambda(1-\mathrm{e}^{-\theta})}$. Adding (4.35), (4.38), and (4.41) and simplifying gives (4.31). □

The next result instead uses Proposition 3 to bound the distance between a PG and a PE distribution.

Corollary 4. *Let* $W=W_1(M,\boldsymbol{E})\sim\mathrm{PE}(\theta,\lambda)$ *and* $W_{G,n}=W_1(M,\boldsymbol{G})\sim\mathrm{PG}(\theta,\frac{\lambda}{n})$ *with* $G_i\sim$ Geometric(λ/n) *be two CCR random variables, and let* $Z_n=\frac{W_{G,n}}{n}$. *Then, for all bounded Lipschitz functions* $h:\mathbb{R}^+_{>0}\to\mathbb{R}$,

$$\left|\mathbb{E}h(W)-\mathbb{E}h(Z_n)\right|\le\frac{\theta}{1-\mathrm{e}^{-\theta}}\frac{2}{n}\left(\frac{n-\mathrm{e}^{\frac{\lambda}{n}}(n-\lambda)}{\lambda(\mathrm{e}^{\frac{\lambda}{n}}-1)}\right)\|h'\|. \qquad (4.42)$$

Proof. Using (3.9) in Proposition 3 gives

$$\left|\mathbb{E}h(W)-\mathbb{E}h(Z_n)\right|=\left|\mathbb{E}h(W)-\mathbb{E}h\left(\frac{W_{G,n}}{n}\right)\right|\le\frac{\theta}{1-\mathrm{e}^{-\theta}}\,\mathbb{E}\left|E-\frac{G}{n}\right|\|h'\|.$$

With the coupling $\tilde{G}=\lceil nE\rceil\sim$ Geometric$(1-\mathrm{e}^{-\lambda/n})$, $\tilde{G}$ is stochastically greater than or equal to G and $\mathbb{E}|nE-G|\le\mathbb{E}|nE-\tilde{G}|+\mathbb{E}(\tilde{G}-G)$. Moreover,

$$\mathbb{E}|nE-\tilde{G}|=\sum_{k=0}^{\infty}\int_{\frac{(k-1)}{n}}^{\frac{k}{n}}\lambda\mathrm{e}^{-\lambda x}\,(k-nx)\,\mathrm{d}x=\frac{n-\mathrm{e}^{\frac{\lambda}{n}}(n-\lambda)}{\lambda(\mathrm{e}^{\frac{\lambda}{n}}-1)}$$

and $\mathbb{E}(\tilde{G}-G)=(1-\mathrm{e}^{-\frac{\lambda}{n}})^{-1}-n/\lambda$. Hence the bound (4.42) follows. □

Remark 6. To compare (4.31) and (4.42), first note that as no fixed bound on $\|h\|$ is assumed, through rescaling h we can make $\|\tilde{h}\|$ as small as desired. Therefore, for a comparison we focus on the terms involving $\|h'\|$. For continuous $\|h'\|$, (4.31) outperforms the bound (4.42) for large n, since $\lim_{n\to\infty}\mathrm{e}^{\frac{\lambda}{n}}\left(1+\frac{\mathrm{e}^{\frac{\theta\lambda}{n\mathrm{e}}}}{n}B_1(\theta,\lambda)\right)=1$ and $\lim_{n\to\infty}\frac{\theta}{1-\mathrm{e}^{-\theta}}2\left(\frac{n-\mathrm{e}^{\frac{\lambda}{n}}(n-\lambda)}{\lambda(\mathrm{e}^{\frac{\lambda}{n}}-1)}\right)=\frac{\theta}{(1-\mathrm{e}^{-\theta})}$. In

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

TABLE 1. Values of n above which the coefficient of $\|h'\|$ is smaller in (4.31) than in (4.42)

λ	1	1	2	2	2	3	3	3	5
θ	1	2	1	2	3	2	3	5	3
n_0	9	7	14	16	35	49	374	67 775	87 731

particular, for any $n \geq n_0$ the right-hand side of (4.42) is larger than the coefficient of $\|h'\|$ in the bound (4.31), where $n_0 = n_0(\theta, \lambda)$ solves $\mathrm{e}^{\frac{\lambda}{n}} \leq \frac{2n\theta}{\lambda(1-\mathrm{e}^{-\theta})(\mathrm{e}^{\frac{\lambda}{n}}-1)}\left(\frac{n-\mathrm{e}^{\frac{\lambda}{n}}(n-\lambda)}{n+\mathrm{e}^{\frac{\theta\lambda}{n\mathrm{e}}}B_1(\theta,\lambda)}\right)$. Table 1 shows such values of n_0.

5. Application to the exponential–geometric distribution

The *exponential–geometric (EG) distribution* $\mathrm{EG}(\lambda, p)$ introduced in [3] is the distribution of the minimum of N i.i.d. random variables $\mathbf{E} = (E_1, E_2, \dots)$ with $E_i \sim \mathrm{Exp}(\lambda)$ for $i \in \mathbb{N}$, where N is a Geometric(p) random variable and independent of all the E_i. We set $q = 1 - p$. An EG random variable $W = W_{-1}(N, \mathbf{E})$ is thus a CCR random variable. As $G_N(u) = \frac{up}{(1-qu)}$, $G_N'(u) = \frac{p}{(1-qu)^2}$, and $G_N''(u) = \frac{2pq}{(1-qu)^3}$, the PDF of $W \sim \mathrm{EG}(\lambda, p)$ is

$$f(w;\lambda, p) = \frac{\lambda p \mathrm{e}^{-\lambda w}}{(1 - q\mathrm{e}^{-\lambda w})^2} \quad \text{for } w, \lambda > 0,\ 0 \leq p \leq 1, \tag{5.1}$$

with score function

$$\rho(w) = -\lambda\left(\frac{1 + q\mathrm{e}^{-\lambda w}}{1 - q\mathrm{e}^{-\lambda w}}\right). \tag{5.2}$$

This gives the score Stein operator

$$\mathcal{T}g(w) = g'(w) - \lambda\left(\frac{1 + q\mathrm{e}^{-\lambda w}}{1 - q\mathrm{e}^{-\lambda w}}\right)g(w) \tag{5.3}$$

and the Stein equation

$$g'(w) - \lambda\left(\frac{1 + q\mathrm{e}^{-\lambda w}}{1 - q\mathrm{e}^{-\lambda w}}\right)g(w) = h(w) - \mathbb{E}h(W). \tag{5.4}$$

The next lemma bounds the solution (2.5) of this EG Stein equation.

Lemma 3. *For any bounded test function $h : \mathbb{R}^+_{>0} \to \mathbb{R}$, let $\tilde{h}(w) = h(w) - \mathbb{E}h(W)$ for $W \sim \mathrm{EG}(\lambda, p)$. Then the solution $g = g_h$ of the EG Stein equation (5.4) satisfies*

$$|g(w)| \leq \frac{\|\tilde{h}\|}{\lambda}, \tag{5.5}$$

$$|g'(w)| \leq \frac{2\|\tilde{h}\|}{p}. \tag{5.6}$$

Proof. For the solution (2.5) of the EG Stein equation (5.4),

$$|g(w)| \leq \|\tilde{h}\|\left|\frac{-1}{f(w)}\int_w^\infty f(t)\,\mathrm{d}t\right| = \|\tilde{h}\|\left|\frac{1 - F(w)}{f(w)}\right| = \|\tilde{h}\|\left|\frac{(1 - q\mathrm{e}^{-\lambda w})}{\lambda}\right| \leq \frac{\|\tilde{h}\|}{\lambda}.$$

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

The last inequality follows since with $0 \le \mathrm{e}^{-\lambda w} \le 1$ and $0 \le q \le 1$, we have $1 - q \le 1 - q\mathrm{e}^{-\lambda w} \le 1$ and $1 \le 1 + q\mathrm{e}^{-\lambda w} \le 1 + q$. For a bound on $|g'(w)|$, we use the Stein equation (5.4) and the triangle inequality to obtain

$$|g'(w)| \le \|\tilde{h}\| + \left|\lambda\left(\frac{1+q\mathrm{e}^{-\lambda w}}{1-q\mathrm{e}^{-\lambda w}}\right)\right| \frac{\|\tilde{h}\|}{\lambda} \le \|\tilde{h}\|\left(1 + \frac{1+q}{1-q}\right).$$

Simplifying the last inequality gives (5.6). □

5.1. The minimum of a geometric number of i.i.d. random variables

Let $N \in \mathbb{N}$ be a random variable which is independent of $\mathbf{X} = (X_1, X_2, \dots)$, a sequence of i.i.d. random variables, and let $W = W_{-1}(N, \mathbf{X})$ have PDF of the form (2.12) for $\alpha = -1$. In this subsection we approximate its distribution by an EG distribution. First we apply Proposition 1.

Corollary 5. *Assume that the X_i have CDF F_X, differentiable PDF p_X, and score function ρ_X, and let $N \sim$ Geometric(p). Then*

$$d_{\mathrm{TV}}\big(\mathrm{EG}(\lambda, p), \mathcal{L}(W)\big) \le \frac{2}{\lambda}\left|\mathbb{E}(-\lambda - \rho_X(W)) - \mathbb{E}\left(\frac{2(1-p)\lambda\mathrm{e}^{-\lambda W}}{1-(1-p)\mathrm{e}^{-\lambda W}} - \frac{2(1-p)f_X(W)}{1-(1-p)(1-F_X(W))}\right)\right|.$$

Proof. Substituting $\frac{G_N''(1-F(w))}{G_N'(1-F(w))} = \frac{2(1-p)f(w)}{1-(1-p)(1-F(w))}$ and the score function of the exponential distribution into (3.1), taking h an indicator function so that $\|\tilde{h}\| \le 2\|h\| \le 2$, and using (5.5) gives the bound. □

Next, we instead use Proposition 3; the corollary follows immediately from (3.9).

Corollary 6. *For $W = W_{-1}(N, \boldsymbol{X})$ with $N \sim$ Geometric(p),*

$$d_{\mathrm{BW}}\big(\mathrm{EG}(\lambda, p), \mathcal{L}(W)\big) \le \mathbb{E}|E - X|/p.$$

5.2. Approximating the extended exponential–geometric distribution

Motivated by population heterogeneity, Adamidis et al. [2] developed the extended exponential–geometric (EEG) distribution by assuming that individual units in a population have increasing failure rates that depend on a random scale parameter A. Their lifetimes X/A are modelled by a modified extreme value distribution; if $A = \alpha$, then the PDF is $f(x|\alpha;\beta) = \alpha\beta\mathrm{e}^{\beta x + \alpha(1-\mathrm{e}^{\beta x})}$, where $x, \alpha, \beta \in \mathbb{R}^+_{>0}$; it is assumed that A has an Exp(γ) distribution. Then the unconditional lifetime distribution X has PDF

$$f(x;\beta,\gamma) = \frac{\beta\gamma\mathrm{e}^{-\beta x}}{(1-(1-\gamma)\mathrm{e}^{-\beta x})^2}, \quad x > 0, \tag{5.7}$$

with $\beta, \gamma \in \mathbb{R}^+_{>0}$; we use the notation $X \sim \mathrm{EEG}(\beta, \gamma)$. Its score function is

$$\rho(x) = \frac{f'(x)}{f(x)} = -\beta\left(\frac{1+(1-\gamma)\mathrm{e}^{-\beta x}}{1-(1-\gamma)\mathrm{e}^{-\beta x}}\right). \tag{5.8}$$

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

This distribution is not in the CCR family. However, the EG distribution EG(β, γ) is a special case when $\gamma \in (0, 1)$. To assess the total variation distance between the distributions, we use the general approach developed in Section 2.

Theorem 3. *For $X \sim \mathrm{EEG}(\beta, \gamma)$ and $W \sim \mathrm{EG}(\lambda, p)$, with $\beta, \gamma, \lambda \in \mathbb{R}^+_{>0}$ and $p \in (0, 1)$, and for a bounded test function h, we have*

$$\left|\mathbb{E}h(X) - \mathbb{E}h(W)\right| \leq \begin{cases} \frac{\|\tilde{h}\|}{\lambda}\left[|\lambda-\beta|\left(\frac{2-p}{p} + \frac{2(1-\gamma)}{\mathrm{e}\gamma^2}\right) + \frac{2\beta}{\min(\gamma,p)^2}|p-\gamma|\right], & 0<\gamma<1, \\ \frac{\|\tilde{h}\|}{\lambda p}\left[|\lambda-\beta|\,(1+(\gamma-1)(1-p)) + (\lambda+\beta)\,(\gamma-p)\right], & \gamma \geq 1. \end{cases} \tag{5.9}$$

Proof. Using the score functions (5.2) and (5.8) in (2.6) yields

$$\begin{aligned} |\mathbb{E}h(X) - \mathbb{E}h(W)| &= |\mathbb{E}(\rho_X(X) - \rho_Y(X))g(X)| \\ &= \left|\mathbb{E}\left(-\beta\left(\frac{1+(1-\gamma)\mathrm{e}^{-\beta X}}{1-(1-\gamma)\mathrm{e}^{-\beta X}}\right) + \lambda\left(\frac{1+(1-p)\mathrm{e}^{-\lambda X}}{1-(1-p)\mathrm{e}^{-\lambda X}}\right)\right) g(X)\right| \\ &\leq \frac{\|\tilde{h}\|}{\lambda}\mathbb{E}\left|-\beta\left(\frac{1+(1-\gamma)\mathrm{e}^{-\beta X}}{1-(1-\gamma)\mathrm{e}^{-\beta X}}\right) + \lambda\left(\frac{1+(1-p)\mathrm{e}^{-\lambda X}}{1-(1-p)\mathrm{e}^{-\lambda X}}\right)\right|. \end{aligned} \tag{5.10}$$

To bound the expectation in the above equation, we let

$$R(x) = -\beta\left(\frac{1+(1-\gamma)\mathrm{e}^{-\beta x}}{1-(1-\gamma)\mathrm{e}^{-\beta x}}\right) + \lambda\left(\frac{1+(1-p)\mathrm{e}^{-\lambda x}}{1-(1-p)\mathrm{e}^{-\lambda x}}\right).$$

Case 1: $0 < \gamma < 1$. In this case we decompose $|R(x)|$ as

$$|R(x)| \leq \left|(\lambda-\beta)\frac{1+(1-p)\mathrm{e}^{-\lambda x}}{1-(1-p)\mathrm{e}^{-\lambda x}}\right| \tag{5.11}$$

$$+ \left|\beta\left(\frac{1+(1-p)\mathrm{e}^{-\lambda x}}{1-(1-p)\mathrm{e}^{-\lambda x}} - \frac{1+(1-\gamma)\mathrm{e}^{-\lambda x}}{1-(1-\gamma)\mathrm{e}^{-\lambda x}}\right)\right| \tag{5.12}$$

$$+ \left|\beta\left(\frac{1+(1-\gamma)\mathrm{e}^{-\lambda x}}{1-(1-\gamma)\mathrm{e}^{-\lambda x}} - \frac{1+(1-\gamma)\mathrm{e}^{-\beta x}}{1-(1-\gamma)\mathrm{e}^{-\beta x}}\right)\right| \tag{5.13}$$

and bound these terms separately. For (5.11) we use that $\frac{1+\alpha q}{1-\alpha q} \leq \frac{1+q}{1-q}$ when $\alpha \in [0, 1]$ and $q \in (0, 1)$. Hence

$$\left|(\lambda-\beta)\frac{1+(1-p)\mathrm{e}^{-\lambda x}}{1-(1-p)\mathrm{e}^{-\lambda x}}\right| \leq |\lambda-\beta|\frac{2-p}{p}.$$

For (5.12), Taylor expansion about $1-\gamma$ of the function $f(q) = \frac{1+aq}{1-aq}$ with $a = \mathrm{e}^{-\lambda x} \in (0, 1)$ and $q = 1-p$ gives $f'(q) = \frac{a}{1-aq} + \frac{a(1+aq)}{(1-aq)^2} = \frac{2a}{(1-aq)^2} > 0$. Moreover, for $0 < a < 1$ we have $\frac{2a}{(1-aq)^2} \leq \frac{2}{(1-q)^2}$, and hence for $\theta \in (0, 1)$ we have $0 < f'(\theta(1-p) + (1-\theta)(1-\gamma)) \leq \frac{2}{(1-\max(1-\gamma,1-p))^2} = \frac{2}{(\min(\gamma,p))^2}$. Therefore,

$$\left|\beta\left(\frac{1+(1-p)\mathrm{e}^{-\lambda x}}{1-(1-p)\mathrm{e}^{-\lambda x}} - \frac{1+(1-\gamma)\mathrm{e}^{-\lambda x}}{1-(1-\gamma)\mathrm{e}^{-\lambda x}}\right)\right| \leq \beta\frac{2|p-\gamma|}{(\min(\gamma,p))^2}.$$

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

For (5.13), first-order Taylor expansion of the function $f(\beta) = \frac{1+(1-\gamma)e^{-\beta x}}{1-(1-\gamma)e^{-\beta x}}$ gives

$$\left|\beta\left(\frac{1+(1-\gamma)e^{-\lambda x}}{1-(1-\gamma)e^{-\lambda x}} - \frac{1+(1-\gamma)e^{-\beta x}}{1-(1-\gamma)e^{-\beta x}}\right)\right| \leq \beta|\lambda-\beta|\,|f'(\theta\lambda+(1-\theta)\beta)|$$

for some $\theta \in (0,1)$. Now the function $f'(\beta) = \frac{2x(1-\gamma)e^{-\beta x}}{(1-(1-\gamma)e^{-\beta x})^2}$ is positive; moreover, $xe^{-\beta x} \leq (e\beta)^{-1}$ for $x \geq 0$, so that $f'(\beta) \leq \frac{1}{\beta e}\frac{2(1-\gamma)}{(1-(1-\gamma)e^{-\beta x})^2} \leq \frac{2(1-\gamma)}{e\beta\gamma^2}$. Hence we can bound

$$\left|\beta\left(\frac{1+(1-\gamma)e^{-\lambda x}}{1-(1-\gamma)e^{-\lambda x}} - \frac{1+(1-\gamma)e^{-\beta x}}{1-(1-\gamma)e^{-\beta x}}\right)\right| \leq \frac{2(1-\gamma)}{e\gamma^2}|\lambda-\beta|.$$

This gives as overall bound

$$|R(x)| \leq |\lambda-\beta|\left(\frac{2-p}{p} + \frac{2(1-\gamma)}{e\gamma^2}\right) + \frac{2\beta}{\min(\gamma,p)^2}|p-\gamma|.$$

Substituting this bound into (5.10) gives the final bound for the case where $0 < \gamma < 1$.

Case 2: For $\gamma \geq 1$, we have

$$\begin{aligned}|R(x)| &\leq \left|\frac{(\lambda-\beta)(1-(1-\gamma)(1-p)e^{-(\lambda+\beta)x})}{(1-(1-\gamma)e^{-\beta x})(1-(1-p)e^{-\lambda x})}\right| \\ &\quad + \left|\frac{(\lambda+\beta)((1-p)e^{-\lambda x}-(1-\gamma)e^{-\beta x})}{(1-(1-\gamma)e^{-\beta x})(1-(1-p)e^{-\lambda x})}\right| \\ &\leq \frac{|\lambda-\beta|\,(1+(\gamma-1)(1-p))}{p} + \frac{(\lambda+\beta)(\gamma-p)}{p}.\end{aligned}$$

Substituting this result into (5.10) gives the bound for $\gamma \geq 1$ in (5.9). □

Remark 7.

(i) The bounds are not optimised for numerical value, but for $\beta = \lambda$ and $\gamma = p$, (5.9) reduces to zero, as it should.

(ii) For $\gamma \in (0,1)$, (5.9) gives a bound on the distance between an EG(β, γ) and an EG(λ, p) distribution.

6. The maximum waiting time of sequence patterns in Bernoulli trials

This application is motivated by the results in Section 4 of [26], which gives bounds on the distribution of the number of trials preceding the first appearance of a pattern in dependent Bernoulli trials. Here we are interested in the distribution of the maximum of such random variables.

Consider M independent parallel systems $(X_1^{(i)}, X_2^{(i)}, \ldots)$, for $i = 1, 2, \ldots, M$, of possibly dependent Bernoulli(a) trials, which are jointly independent of $M \in \mathbb{N}$. For each sequence i, let $I_j^{(i)}$ be the indicator function that a fixed non-overlapping binary sequence pattern of length k occurs starting at $X_j^{(i)}$; the pattern may be specific to sequence i. Let $V_i = \min\{j : I_j^{(i)} = 1\}$ denote the first occurrence of the pattern of interest in the ith system; assume that $\mathbb{P}(V_i = 1) = p$ for all

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

$i \in \mathbb{N}$. We denote the maximum waiting time for the occurrence of a corresponding sequence pattern in all M parallel systems by $W = W_1(M, \mathbf{V})$ where $\mathbf{V} = (V_1, V_2 \dots)$.

Example 2. *If the Bernoulli trials are independent and the pattern of interest is a run of ones of length k, starting with a zero and followed by k ones, then* $\mathbb{P}(V_i = 1) = p = (1-a)a^k$, *as given in Corollary 2 of [26]. Intuitively, the waiting time for the occurrence of this pattern in an individual sequence is approximately geometric with parameter p. For an approximation by a PE distribution we are particularly interested in instances where we can write the probability* $\mathbb{P}(V_i = 1) = p$ *as* $p = \lambda/n$; *here this leads to scaling the run length k as* $k \sim a \log n$.

Corollary 7. *In the above setting, let* $U_n = W/n$ *with* $p = \frac{\lambda}{n}$, $W_1(M', \mathbf{E}) \sim \mathrm{PE}(\theta, \lambda)$ *with* M' *being a zero-truncated Poisson random variable, and* $E_i \sim \mathrm{Exp}(\lambda)$. *Then we have*

$$\begin{aligned} &d_{\mathrm{BW}}\big(\mathcal{L}(U_n), \mathrm{PE}(\theta, \lambda)\big) \\ &\le \frac{2\lambda(k-1)}{n}\,\mathbb{E}M + \frac{1}{\ln(p^{-1})}\,\mathbb{E}\sum_{k=\min(M,M')}^{\max(M,M')}\frac{1}{k} + \frac{\mathrm{e}^{\frac{\lambda}{n}}}{n}\left(1 + \frac{\mathrm{e}^{\frac{\theta\lambda}{n\mathrm{e}}}}{n}B_1(\theta,\lambda)\right) \\ &\quad + \frac{1}{n}\left(\left(1-\frac{\lambda}{n}\right)^{-6}\mathrm{e}^{\theta\mathrm{e}^{\frac{\lambda}{n}}+2\frac{\theta\lambda}{n}+\frac{\theta\lambda}{n\mathrm{e}}+\frac{\lambda}{n}}B_2(\theta,\lambda) + \frac{1}{2n}\mathrm{e}^{\frac{\lambda}{n}+\frac{\theta\lambda}{n\mathrm{e}}}B_3(\theta,\lambda)\right), \end{aligned}$$

where $B_1(\theta, \lambda)$, $B_2(\theta, \lambda)$, *and* $B_3(\theta, \lambda)$ *are as given in Theorem 2.*

Example 3. *In the above example of independent Bernoulli trials and the pattern of interest being a run of ones of length* $k \sim a \log n$, *the bound in Corollary 7 is of order* $(\log n)^{-1}$.

Proof. We couple $U_n = \frac{1}{n}\max\{V_1, \cdots, V_M\}$ and $Z_n = \frac{1}{n}\max\{T_1, \cdots, T_M\}$, where $T_i \sim \mathrm{Geometric}(\frac{\lambda}{n})$ for $i = 1, \dots, M$, by using the same random variable M. Then

$$d_{\mathrm{BW}}\big(\mathcal{L}(U_n), \mathcal{L}(\mathrm{PE}(\theta,\lambda))\big) \le d_{\mathrm{BW}}\big(\mathcal{L}(U_n), \mathcal{L}(Z_n)\big) + d_{\mathrm{BW}}\big(\mathcal{L}(Z_n), \mathcal{L}(\mathrm{PE}(\theta,\lambda))\big). \tag{6.1}$$

Taking a union bound, we have

$$d_{\mathrm{TV}}\big(\mathcal{L}(U_n), \mathcal{L}(Z_n)\big) \le \sum_{m=1}^{\infty}\sum_{i=1}^{m} d_{\mathrm{TV}}\left(\mathcal{L}\left(\frac{V_i}{n}\right), \mathcal{L}\left(\frac{T_i}{n}\right)\right)\mathbb{P}(M=m) \le \frac{\lambda(k-1)}{n}\,\mathbb{E}M,$$

where in the last step we have used Corollary 1 from [26]. With (2.2),

$$d_{\mathrm{BW}}\big(\mathcal{L}(U_n), \mathcal{L}(Z_n)\big) \le 2d_{\mathrm{TV}}\big(\mathcal{L}(U_n), \mathcal{L}(Z_n)\big) \le \frac{2\lambda(k-1)}{n}\,\mathbb{E}M. \tag{6.2}$$

Now

$$d_{\mathrm{BW}}\big(\mathcal{L}(Z_n), \mathrm{PE}(\theta,\lambda)\big) \le d_{\mathrm{BW}}\left(\mathcal{L}(Z_n), \mathcal{L}\left(\frac{W_1(M',\mathbf{T})}{n}\right)\right) \tag{6.3}$$

$$+ d_{\mathrm{BW}}\left(\mathcal{L}\left(\frac{W_1(M',\mathbf{T})}{n}\right), \mathrm{PE}(\theta,\lambda)\right). \tag{6.4}$$

The term (6.4) is bounded in (4.31). To bound (6.3), (3.12) gives

$$d_{\mathrm{BW}}\left(\mathcal{L}(Z_n), \mathcal{L}\left(\frac{W_1(M',\mathbf{T})}{n}\right)\right) \le \sum_{m,m'=1}^{\infty}\mathbb{P}(M=m, M'=m')\,\mathbb{E}\left|W_1(m',\mathbf{T}) - W_1(m,\mathbf{T})\right|.$$

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

The expectation of the maximum of m Geometric(p) variables satisfies

$$\frac{1}{\ln(p^{-1})}\sum_{k=1}^{m}\frac{1}{k} \le \mathbb{E}W_1(m, \mathbf{T}) \le 1 + \frac{1}{\ln(p^{-1})}\sum_{k=1}^{m}\frac{1}{k},$$

as given in [11, p. 136]. Hence

$$\begin{aligned} &d_{\mathrm{BW}}\left(\mathcal{L}(Z_n), \mathcal{L}\left(\frac{W_1(M', \mathbf{T})}{n}\right)\right) \\ &\quad\le \sum_{m,m'=1}^{\infty}\mathbb{P}(M=m, M'=m')\frac{1}{\ln(p^{-1})}\sum_{k=\min(m,m')}^{\max(m,m')}\frac{1}{k} \\ &\quad= \frac{1}{\ln(p^{-1})}\left|\mathbb{E}\sum_{k=1}^{M}\frac{1}{k} - \mathbb{E}\sum_{k=1}^{M'}\frac{1}{k}\right|. \end{aligned}$$

Combining this result with (4.31) and (6.2) in (6.1), we obtain the assertion. □

Remark 8. For $M = M'$, the bound in Corollary 7 reduces to

$$\begin{aligned} d_{\mathrm{BW}}\big(\mathcal{L}(U_n), \mathrm{PE}(\theta, \lambda)\big) \le{}& \frac{2\theta\lambda(k-1)}{n(1-\mathrm{e}^{-\theta})} + \frac{\mathrm{e}^{\frac{\lambda}{n}}}{n}\left(1 + \frac{\mathrm{e}^{\frac{\theta\lambda}{n\mathrm{e}}}}{n}B_1(\theta, \lambda)\right) \\ &+ \frac{1}{n}\left(\left(1-\frac{\lambda}{n}\right)^{-6}\mathrm{e}^{\theta\mathrm{e}^{\frac{\lambda}{n}}+2\frac{\theta\lambda}{n}+\frac{\theta\lambda}{n\mathrm{e}}+\frac{\lambda}{n}}B_2(\theta, \lambda) + \frac{1}{2n}\mathrm{e}^{\frac{\lambda}{n}+\frac{\theta\lambda}{n\mathrm{e}}}B_3(\theta, \lambda)\right). \end{aligned}$$

The assumption of i.i.d. sequences can be weakened to that of a Markov chain by applying Theorem 5.5 from [30] with $M = M'$. This theorem gives a Poisson process approximation for the number of 'declumped' counts of each pattern, which in turn yields that the waiting time for each pattern to occur is approximately exponentially distributed. The theorem also gives an explicit bound on the approximation, but this result requires considerable notation and hence we do not pursue it here.

Acknowledgements

We thank Christina Goldschmidt, David Steinsaltz, and Tadas Temcinas for helpful discussions. We would also like to thank the editor and the anonymous reviewers for their suggestions, which have led to overall improvements of the paper.

Funding information

AF is supported by the Commonwealth Scholarship Commission, United Kingdom, and in part by EPSRC grant EP/X002195/1. GR is supported in part by EPSRC grants EP/T018445/1, EP/R018472/1, EP/X002195/1, and EP/Y028872/1. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

References

[1] AARSET, M. V. (1987). How to identify a bathtub hazard rate. IEEE Transactions on Reliability R-36, 106–108.

[2] ADAMIDIS, K., DIMITRAKOPOULOU, T. AND LOUKAS, S. (2005). On an extension of the exponential–geometric distribution. *Statistics & Probability Letters* **73**, 259–269.

[3] ADAMIDIS, K. AND LOUKAS, S. (1998). A lifetime distribution with decreasing failure rate. *Statistics & Probability Letters* **39**, 35–42.

[4] ARRATIA, R., GOLDSTEIN, L. AND GORDON, L. (1989). Two moments suffice for Poisson approximations: the Chen-Stein method. *The Annals of Probability* 9–25.

[5] BARBOUR, A. D., HOLST, L. AND JANSON, S. (1992). *Poisson Approximation*. The Clarendon Press, Oxford, UK.

[6] BASU, A. P. AND KLEIN, J. P. (1982). Some recent results in competing risks theory. *Lecture Notes-Monograph Series* **2**, 216–229.

[7] BETSCH, S. AND EBNER, B. (2019). A new characterization of the Gamma distribution and associated goodness-of-fit tests. *Metrika* **82**, 779–806.

[8] CANCHO, V. G., LOUZADA-NETO, F. AND BARRIGA, G. D. C. (2011). The Poisson-exponential lifetime distribution. *Computational Statistics & Data Analysis* **55**, 677–686.

[9] CHEN, L. H., GOLDSTEIN, L. AND SHAO, Q.-M. (2011). *Normal Approximation by Stein's Method*. Springer, Heidelberg, London.

[10] CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *The Annals of Probability* **3**, 534–545.

[11] EISENBERG, B. (2008). On the expectation of the maximum of IID geometric random variables. *Statistics & Probability Letters* **78**, 135–143.

[12] FATIMA, A. AND ROOHI, A. (2015). The generalized Poisson-exponential distribution. *Journal of ISOSS* **1**, 103–118.

[13] FEIDT, A. (2013). Stein's method for multivariate extremes. arXiv:1310.2564.

[14] GERMAIN, G. AND SWAN, Y. (2025). Bounding the L^1-distance between one-dimensional continuous and discrete distributions via Stein's method. *Journal of Theoretical Probability* **38**, 9.

[15] GOLDSTEIN, L. AND REINERT, G. (2013). Stein's method for the Beta distribution and the Pólya-Eggenberger urn. *Journal of Applied Probability* **50**, 1187–1205.

[16] JAYAKUMAR, K., BABU, M. G. AND BAKOUCH, H. S. (2021). General classes of complementary distributions via random maxima and their discrete version. *Japanese Journal of Statistics and Data Science* **4**, 797–820.

[17] KUS, C. (2007). A new lifetime distribution. *Computational Statistics & Data Analysis* **51**, 4497–4509.

[18] LEY, C., REINERT, G. AND SWAN, Y. (2017). Stein's method for comparison of univariate distributions. *Probability Surveys* **14**, 1–52.

[19] LEY, C. AND SWAN, Y. (2013). Stein's density approach and information inequalities. *Electronic Communications in Probability* **18**, 1–14.

[20] LOUZADA, F., BERETA, E. AND FRANCO, M. (2012). On the distribution of the minimum or maximum of a random number of i.i.d. lifetime random variables. *Applied Mathematics* **3**, 350–353.

[21] LOUZADA, F., RAMOS, P. L. AND PERDONÁ, G. S. C. (2016). Different estimation procedures for the parameters of the extended exponential geometric distribution for medical data. *Computational and Mathematical Methods in Medicine* **2016**, 8727951.

[22] LU, W. AND SHI, D. (2012). A new compounding life distribution: the Weibull–Poisson distribution. *Journal of Applied Statistics* **39**, 21–38.

[23] MARSHALL, A. W. AND OLKIN, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika* **84**, 641–652.

[24] MIJOULE, G., RAIČ, M., REINERT, G. AND SWAN, Y. (2023). Stein's density method for multivariate continuous distributions. *Electronic Journal of Probability* 28, 1–40.

[25] NOURDIN, I. AND PECCATI, G. (2012). *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality*. Cambridge University Press, Cambridge, UK.

[26] PEKÖZ, E. A. (1996). Stein's method for geometric approximation. *Journal of Applied Probability* **33**, 707–713.

[27] PEKÖZ, E. A. AND RÖLLIN, A. (2011). New rates for exponential approximation and the theorems of Rényi and Yaglom. *The Annals of Probability* **39**, 587–608.

[28] PEKÖZ, E. A., RÖLLIN, A. AND ROSS, N. (2013). Total variation error bounds for geometric approximation. *Bernoulli* **19**, 610–632.

[29] RAMOS, P. L., DEY, D. K., LOUZADA, F. AND LACHOS, V. H. (2020). An extended Poisson family of life distribution: a unified approach in competitive and complementary risks. *Journal of Applied Statistics* **47**, 306–322.

[30] REINERT, G., SCHBATH, S. AND WATERMAN, M. S. (2000). Probabilistic and statistical properties of words: an overview. *Journal of Computational Biology* **7**, 1–46.

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

[31] SALAS, A. H. (2012). The exponential function as a limit. *Applied Mathematical Sciences* **6**, 4519–4526.
[32] STEIN, C. (1986). *Approximate Computation of Expectations*. Institute of Mathematical Statistics, Hayward, California.
[33] STEIN, C., DIACONIS, P., HOLMES, S. AND REINERT, G. (2004). Use of exchangeable pairs in the analysis of simulations. *Lecture Notes-Monograph Series* **46**, 1–26.
[34] TAHIR, M. H. AND CORDEIRO, G. M. (2016). Compounding of distributions: a survey and new generalized classes. *Journal of Statistical Distributions and Applications* **3**, 13.
[35] TOJEIRO, C., LOUZADA, F., ROMAN, M. AND BORGES, P. (2014). The complementary Weibull geometric distribution. *Journal of Statistical Computation and Simulation* **84**, 1345–1362.

Appendix Further proofs

Proof of (4.13)

For $X \sim \text{GPE}(\theta, \lambda, \beta)$ with $\lambda, \theta > 0$ and $\beta \geq 1$, we have $(1 - \mathrm{e}^{-\theta+\theta\mathrm{e}^{-\lambda x}})^{\beta-1} \leq 1$, and hence

$$\begin{aligned}\mathbb{E}(X) &= \frac{\beta\theta\lambda}{(1-\mathrm{e}^{-\theta})^{\beta}} \int_0^{\infty} x\mathrm{e}^{-\lambda x-\theta\beta\mathrm{e}^{-\lambda x}} \left(1-\mathrm{e}^{-\theta+\theta\mathrm{e}^{-\lambda x}}\right)^{\beta-1} \mathrm{d}x \\ &\leq \frac{\beta\theta\lambda}{(1-\mathrm{e}^{-\theta})^{\beta}} \int_0^{\infty} x\mathrm{e}^{-\lambda x}\,\mathrm{d}x.\end{aligned}$$

Proof of the inequality (4.36)

First note that for $n \in \mathbb{N}$ and $|x| < n$,

$$\mathrm{e}^{x}\left(1-\frac{x^2}{n}\right) \leq \left(1+\frac{x}{n}\right)^{n}, \qquad 0 \leq \mathrm{e}^{x} - \left(1+\frac{x}{n}\right)^{n} \leq \frac{x^2}{n}\mathrm{e}^{x}, \tag{A.1}$$

and $x\mathrm{e}^{-x} \leq \mathrm{e}^{-1}$ for $x > 0$. Hence, for $0 < \lambda < n$ and $z > 0$,

$$0 \leq \mathrm{e}^{-\lambda z} - \left(1-\frac{\lambda z}{nz}\right)^{nz} \leq \frac{(\lambda z)^2}{nz}\mathrm{e}^{-\lambda z} = \frac{\lambda^2 z}{n}\mathrm{e}^{-\lambda z} \tag{A.2}$$

$$\leq \frac{\lambda}{ne}. \tag{A.3}$$

Also, we can write $\mathrm{e}^{-\theta\left(1-\frac{\lambda}{n}\right)^{nz}} = \mathrm{e}^{\theta\left[\mathrm{e}^{-\lambda z}-\left(1-\frac{\lambda}{n}\right)^{nz}\right]}\mathrm{e}^{-\theta\mathrm{e}^{-\lambda z}}$, and so

$$\mathrm{e}^{-\theta\left(1-\frac{\lambda}{n}\right)^{nz}} \leq \mathrm{e}^{\frac{\theta\lambda}{ne}}\mathrm{e}^{-\theta\mathrm{e}^{-\lambda z}}. \tag{A.4}$$

Now, to bound $l = \tau(d - c - \frac{1}{n}c')$, we have

$$\tau(z)c(z) = \mathrm{e}^{\lambda z+\theta\mathrm{e}^{-\lambda z}}c(z) = \frac{\lambda\theta}{n}, \qquad \frac{2}{n}\tau(z)c'(z) = \frac{2\lambda^2\theta}{n^2}\left(\theta\mathrm{e}^{-\lambda z}-1\right),$$

and

$$\begin{aligned}\mathrm{e}^{\lambda z+\theta\mathrm{e}^{-\lambda z}}d(z) &= \mathrm{e}^{\lambda z+\theta\mathrm{e}^{-\lambda z}-\theta\left(1-\frac{\lambda}{n}\right)^{nz}}\left(\mathrm{e}^{\frac{\lambda\theta}{n}\left(1-\frac{\lambda}{n}\right)^{nz}}-1\right) \\ &= \mathrm{e}^{\lambda z+\theta\mathrm{e}^{-\lambda z}-\theta\left(1-\frac{\lambda}{n}\right)^{nz}}\left(\mathrm{e}^{\frac{\lambda\theta}{n}\left(1-\frac{\lambda}{n}\right)^{nz}}-1-\frac{\lambda\theta}{n}\left(1-\frac{\lambda}{n}\right)^{nz}\right) \\ &\quad + \mathrm{e}^{\lambda z+\theta\mathrm{e}^{-\lambda z}-\theta\left(1-\frac{\lambda}{n}\right)^{nz}}\frac{\lambda\theta}{n}\left(1-\frac{\lambda}{n}\right)^{nz}.\end{aligned}$$

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

Thus,

$$l(z) = \frac{2\lambda^2\theta}{n^2}\left(\theta e^{-\lambda z} - 1\right)$$
$$+ e^{\lambda z + \theta\left(e^{-\lambda z} - \left(1-\frac{\lambda}{n}\right)^{nz}\right)}\left(e^{\frac{\lambda\theta}{n}\left(1-\frac{\lambda}{n}\right)^{nz}} - 1 - \frac{\lambda\theta}{n}\left(1-\frac{\lambda}{n}\right)^{nz}\theta\right) \tag{A.5}$$
$$+ \frac{\lambda\theta}{n}e^{\lambda z + \theta e^{-\lambda z}}\left(e^{-\theta\left(1-\frac{\lambda}{n}\right)^{nz}}\left(1-\frac{\lambda}{n}\right)^{nz} - e^{-\lambda z - \theta e^{-\lambda z}}\right). \tag{A.6}$$

Here

$$\frac{2\lambda^2\theta}{n^2}\left(\theta e^{-\lambda z} - 1\right) \le \frac{2\lambda^2\theta}{n^2}\max(1,\theta). \tag{A.7}$$

To bound (A.5), we use (A.3) and series expansion, recalling that $0 < \lambda < n$, to get

$$e^{\lambda z + \theta e^{-\lambda z} - \left(1-\frac{\lambda}{n}\right)^{nz}\theta}\left(e^{\frac{\lambda\theta}{n}\left(1-\frac{\lambda}{n}\right)^{nz}} - 1 - \frac{\lambda\theta}{n}\left(1-\frac{\lambda}{n}\right)^{nz}\theta\right) \tag{A.8}$$
$$\le e^{\lambda z + \frac{\theta\lambda}{ne}}\sum_{k=2}^{\infty}\frac{\left(\lambda\theta\left(1-\frac{\lambda}{n}\right)^{nz}\right)^k}{n^k k!}$$
$$\le e^{\lambda z + \frac{\theta\lambda}{ne}}\sum_{k=2}^{\infty}\frac{(\lambda\theta)^k e^{-\lambda z k}}{n^k k!} \le \frac{1}{n^2}e^{-\lambda z + \frac{\theta\lambda}{ne} + \lambda\theta} \le \frac{1}{n^2}e^{\frac{\theta}{e} + \lambda\theta}. \tag{A.9}$$

To bound (A.6), we first bound

$$e^{-\lambda z - \theta e^{-\lambda z}} - e^{-\theta\left(1-\frac{\lambda}{n}\right)^{nz}}\left(1-\frac{\lambda}{n}\right)^{nz}$$
$$= \left(e^{-\lambda z} - \left(1-\frac{\lambda}{n}\right)^{nz}\right)e^{-\theta\left(1-\frac{\lambda}{n}\right)^{nz}} + e^{-\lambda z - \theta e^{-\lambda z}}\left(1 - e^{\theta\{e^{-\lambda z} - \left(1-\frac{\lambda}{n}\right)^{nz}\}}\right)$$
$$\le \left(e^{-\lambda z} - \left(1-\frac{\lambda}{n}\right)^{nz}\right)e^{-\theta\left(1-\frac{\lambda}{n}\right)^{nz}}$$
$$+ \theta e^{-\lambda z - \theta e^{-\lambda z}}\left(e^{-\lambda z} - \left(1-\frac{\lambda}{n}\right)^{nz}\right)e^{\theta\left\{e^{-\lambda z} - \left(1-\frac{\lambda}{n}\right)^{nz}\right\}}$$
$$\le \frac{\lambda^2 z}{n}e^{\frac{\theta\lambda}{ne}}e^{-\lambda z - \theta e^{-\lambda z}} + \frac{\theta\lambda}{ne}e^{\frac{\theta\lambda}{ne}}e^{-\lambda z - \theta e^{-\lambda z}}. \tag{A.10}$$

Here we have used Property 4 from [31], that for all $x > 0$ we have $\left(1+\frac{x}{n}\right)^n - 1 \le xe^x$ and $e^x - 1 \le xe^x$, along with (A.2) and (A.4). Thus

$$\frac{\lambda\theta}{n}e^{\lambda z + \theta e^{-\lambda z}}\left(e^{-\theta\left(1-\frac{\lambda}{n}\right)^{nz}}\left(1-\frac{\lambda}{n}\right)^{nz} - e^{-\lambda z - \theta e^{-\lambda z}}\right) \le \frac{\theta\lambda^3 z}{n^2}e^{\frac{\theta\lambda}{ne}} + \frac{\theta^2\lambda^2}{n^2}e^{\frac{\theta\lambda}{ne} - 1}. \tag{A.11}$$

Combining (A.7), (A.9), and (A.11), we get

$$l(z) \le \frac{2\lambda^2\theta}{n^2}\max(1,\theta) + \frac{1}{n^2}e^{\lambda\theta + \frac{\theta}{e}} + \frac{\theta^2\lambda^2}{n^2}e^{\frac{\theta\lambda}{ne} - 1} + \frac{\theta\lambda^3 z}{n^2}e^{\frac{\theta\lambda}{ne}}.$$

Replacing z by Z_n gives (4.36).

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

Bounding (4.39)

For $z > 0$,

$$\begin{aligned}
&2\frac{\lambda^2\theta}{n^2}\mathrm{e}^{-\lambda z-\theta\mathrm{e}^{-\lambda z}}(\theta\mathrm{e}^{-\lambda z}-1)\\
&-\left(\mathrm{e}^{-\theta(1-\frac{\lambda}{n})^{nz+1}}-\mathrm{e}^{-\theta(1-\frac{\lambda}{n})^{nz}}-\mathrm{e}^{-\theta(1-\frac{\lambda}{n})^{nz-1}}+\mathrm{e}^{-\theta(1-\frac{\lambda}{n})^{nz-2}}\right)\\
=&2\frac{\lambda^2\theta}{n^2}\mathrm{e}^{-\lambda z-\theta\mathrm{e}^{-\lambda z}}(\theta\mathrm{e}^{-\lambda z}-1)-k\left(\frac{n}{n-\lambda}\right),
\end{aligned}$$

where, with $a=\theta\left(1-\frac{\lambda}{n}\right)^{nz}$,

$$k(b)=\mathrm{e}^{-\frac{a}{b}}-\mathrm{e}^{-a}-\mathrm{e}^{-ab}+\mathrm{e}^{-ab^2}=2(b-1)^2a\mathrm{e}^{-a}(a-1)+R_1$$

with R_1 being the remainder term from the Taylor expansion for $k(b)$ about 1,

$$R_1=\frac{1}{3}(b-1)^3a\left(a^2\mathrm{e}^{-a\xi}+\frac{a^2\mathrm{e}^{-\frac{a}{\xi}}}{\xi^6}-\frac{6a\mathrm{e}^{-\frac{a}{\xi}}}{\xi^5}+\frac{6\mathrm{e}^{-\frac{a}{\xi}}}{\xi^4}+12a\xi\mathrm{e}^{-a\xi^2}-8a^2\xi^3\mathrm{e}^{-a\xi^2}\right)$$

for some $1<\xi\le b=\frac{n}{n-\lambda}$. To bound R_1, we use that $1<\xi\le\frac{n}{n-\lambda}$ so that $\mathrm{e}^{-a\xi}\le\mathrm{e}^{-a}$ and $\mathrm{e}^{-a\xi^2}\le\mathrm{e}^{-a}$; also $\mathrm{e}^{-\frac{a}{\xi}}\le\mathrm{e}^{-a\frac{n-\lambda}{n}}$. Hence, with the crude bounds $a\le\theta$ and $\mathrm{e}^{-a}\le1$,

$$\begin{aligned}
|R_1|\le&\frac{1}{3}(b-1)^3a\left(a^2\mathrm{e}^{-a}+\left(\frac{n-\lambda}{n}\right)^6a^2\mathrm{e}^{-a\frac{n-\lambda}{n}}+\left(\frac{n-\lambda}{n}\right)^5 6a\mathrm{e}^{-a\frac{n-\lambda}{n}}\right.\\
&\left.+\left(\frac{n-\lambda}{n}\right)^4 6\mathrm{e}^{-a\frac{n-\lambda}{n}}+12a\left(\frac{n}{n-\lambda}\right)\mathrm{e}^{-a}+8a^2\left(\frac{n}{n-\lambda}\right)^3\mathrm{e}^{-a}\right)\\
\le&\frac{1}{3}(b-1)^3a^2\mathrm{e}^{-a}\left(\theta+\theta\mathrm{e}^{\theta\frac{\lambda}{n}}+6\mathrm{e}^{\theta\frac{\lambda}{n}}+6a^{-1}\mathrm{e}^{\theta\frac{\lambda}{n}}+12\left(\frac{n}{n-\lambda}\right)+8\theta\left(\frac{n}{n-\lambda}\right)^3\right).
\end{aligned}$$

Substituting the expressions for a and b gives

$$\begin{aligned}
|R_1|\le&\frac{1}{3}\left(\frac{\lambda}{n-\lambda}\right)^3\theta^2\left(1-\frac{\lambda}{n}\right)^{2nz}\mathrm{e}^{-\theta\left(1-\frac{\lambda}{n}\right)^{nz}}\left[\theta+\theta\mathrm{e}^{\frac{\lambda\theta}{n}}+6\mathrm{e}^{\frac{\lambda\theta}{n}}\right.\\
&\left.+6\mathrm{e}^{\frac{\lambda\theta}{n}}\frac{1}{\theta}\left(1-\frac{\lambda}{n}\right)^{-nz}+12\left(\frac{n}{n-\lambda}\right)+8\theta\left(\frac{n}{n-\lambda}\right)^3\right]\\
\le&\frac{\theta^2\lambda^3}{3n^3}\left(1-\frac{\lambda}{n}\right)^{-3}\mathrm{e}^{-2\lambda z}\mathrm{e}^{-\theta\left(1-\frac{\lambda}{n}\right)^{nz}}\left[\theta+\theta\mathrm{e}^{\frac{\lambda\theta}{n}}+6\mathrm{e}^{\frac{\lambda\theta}{n}}+12\left(1-\frac{\lambda}{n}\right)^{-1}\right.\\
&\left.+8\theta\left(1-\frac{\lambda}{n}\right)^{-3}\right]+\frac{2\theta\lambda^3}{n^3}\left(1-\frac{\lambda}{n}\right)^{-3}\mathrm{e}^{\frac{\lambda\theta}{n}}\mathrm{e}^{-\lambda z}\mathrm{e}^{-\theta\left(1-\frac{\lambda}{n}\right)^{nz}},
\end{aligned}$$

where we have used $\mathrm{e}^{-\theta\left(1-\frac{\lambda}{n}\right)^{nz}}=\mathrm{e}^{\theta\left[\mathrm{e}^{-\lambda z}-\left(1-\frac{\lambda}{n}\right)^{nz}\right]}\mathrm{e}^{-\theta\mathrm{e}^{-\lambda z}}$. Hence, with (A.4),

$$\begin{aligned}
|R_1|\le&\frac{\theta^2\lambda^3}{3n^3}\mathrm{e}^{\frac{\theta\lambda}{ne}}\left(1-\frac{\lambda}{n}\right)^{-3}\left[\theta+\theta\mathrm{e}^{\frac{\lambda\theta}{n}}+6\mathrm{e}^{\frac{\lambda\theta}{n}}+12\left(1-\frac{\lambda}{n}\right)^{-1}\right.\\
&\left.+8\theta\left(1-\frac{\lambda}{n}\right)^{-3}\right]\mathrm{e}^{-2\lambda z-\theta\mathrm{e}^{-\lambda z}}+2\frac{\theta\lambda^3}{n^3}\left(1-\frac{\lambda}{n}\right)^{-3}\mathrm{e}^{\frac{\theta\lambda}{ne}(\mathrm{e}+1)}\mathrm{e}^{-\lambda z-\theta\mathrm{e}^{-\lambda z}}.
\end{aligned}\tag{A.12}$$

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press

Thus,

$$\begin{aligned}&\left|2\frac{\lambda^2\theta}{n^2}\mathrm{e}^{-\lambda z-\theta\mathrm{e}^{-\lambda z}}(\theta\mathrm{e}^{-\lambda z}-1)\right.\\&\left.-\left(\mathrm{e}^{-\theta(1-\frac{\lambda}{n})^{nz+1}}-\mathrm{e}^{-\theta(1-\frac{\lambda}{n})^{nz}}-\mathrm{e}^{-\theta(1-\frac{\lambda}{n})^{nz-1}}+\mathrm{e}^{-\theta(1-\frac{\lambda}{n})^{nz-2}}\right)\right|\\&\quad\le\left|2\frac{\lambda^2\theta}{n^2}\mathrm{e}^{-\lambda z-\theta\mathrm{e}^{-\lambda z}}(\theta\mathrm{e}^{-\lambda z}-1)\left[1-\left(1-\frac{\lambda}{n}\right)^{-2}\right]\right|+|R_1|+|R_2|,\end{aligned}\tag{A.13}$$

with

$$\begin{aligned}R_2=2\left(\frac{\lambda}{n-\lambda}\right)^2\Bigg\{&\theta^2\left[\left(1-\frac{\lambda}{n}\right)^{2nz}\mathrm{e}^{-\theta\left(1-\frac{\lambda}{n}\right)^{nz}}-\mathrm{e}^{-2\lambda z-\theta\mathrm{e}^{-\lambda z}}\right]\\&-\theta\left[\left(1-\frac{\lambda}{n}\right)^{nz}\mathrm{e}^{-\theta\left(1-\frac{\lambda}{n}\right)^{nz}}-\mathrm{e}^{-\lambda z-\theta\mathrm{e}^{-\lambda z}}\right]\Bigg\}.\end{aligned}$$

Using (A.10), we obtain the bound

$$\begin{aligned}|R_2|\le{}&4\frac{\theta^2\lambda^4}{n^3}\left(1-\frac{\lambda}{n}\right)^{-2}z\mathrm{e}^{\frac{\theta\lambda}{ne}}\mathrm{e}^{-2\lambda z-\theta\mathrm{e}^{-\lambda z}}+2\frac{\theta^3\lambda^3}{n^3\mathrm{e}}\left(1-\frac{\lambda}{n}\right)^{-2}\mathrm{e}^{\frac{\theta\lambda}{ne}}\mathrm{e}^{-2\lambda z-\theta\mathrm{e}^{-\lambda z}}\\&+2\frac{\theta\lambda^4}{n^3}\left(1-\frac{\lambda}{n}\right)^{-2}z\mathrm{e}^{\frac{\theta\lambda}{ne}}\mathrm{e}^{-\lambda z-\theta\mathrm{e}^{-\lambda z}}+2\frac{\theta^2\lambda^3}{n^3\mathrm{e}}\left(1-\frac{\lambda}{n}\right)^{-2}\mathrm{e}^{\frac{\theta\lambda}{ne}}\mathrm{e}^{-\lambda z-\theta\mathrm{e}^{-\lambda z}}.\end{aligned}\tag{A.14}$$

Combining (A.12), (A.13), and (A.14) and simplifying then gives the bound

$$\begin{aligned}&\left|2\frac{\lambda^2\theta}{n^2}\mathrm{e}^{-\lambda z-\theta\mathrm{e}^{-\lambda z}}(\theta\mathrm{e}^{-\lambda z}-1)\right.\\&\left.-\left(\mathrm{e}^{-\theta(1-\frac{\lambda}{n})^{nz+1}}-\mathrm{e}^{-\theta(1-\frac{\lambda}{n})^{nz}}-\mathrm{e}^{-\theta(1-\frac{\lambda}{n})^{nz-1}}+\mathrm{e}^{-\theta(1-\frac{\lambda}{n})^{nz-2}}\right)\right|\\&\le\frac{2\lambda}{n}\max(1,\theta)\left\{\left(1-\frac{\lambda}{n}\right)^{-2}-1\right\}c(z)+\frac{\theta\lambda^2}{n^2}\left(1-\frac{\lambda}{n}\right)^{-2}\mathrm{e}^{\frac{\theta\lambda}{ne}}\left\{\frac{1}{3}\left(1-\frac{\lambda}{n}\right)^{-1}\right.\\&\quad\left.\times\left[\theta+\theta\mathrm{e}^{\frac{\theta\lambda}{n}}+6\mathrm{e}^{\frac{\theta\lambda}{n}}+12\left(1-\frac{\lambda}{n}\right)^{-1}+8\theta\left(1-\frac{\lambda}{n}\right)^{-3}\right]+2\left(\frac{\theta}{\mathrm{e}}+2\lambda z\right)\right\}\mathrm{e}^{-\lambda z}c(z)\\&\quad+\frac{2\lambda^2}{n^2}\left(1-\frac{\lambda}{n}\right)^{-2}\mathrm{e}^{\frac{\theta\lambda}{ne}}\left[\left(1-\frac{\lambda}{n}\right)^{-1}\mathrm{e}^{\frac{\theta\lambda}{n}}+\lambda z+\frac{\theta}{\mathrm{e}}\right]c(z).\end{aligned}$$

Multiplying the above inequality by $n\left|g\left(z-\frac{1}{n}\right)\right|$ gives (4.39).

https://doi.org/10.1017/apr.2025.20 Published online by Cambridge University Press