



A goodness-of-fit test for the zeta distribution with unknown parameter

Bruno Ebner¹ · Daniel Hlubinka²

Received: 24 December 2025 / Accepted: 8 September 2026
© The Author(s) 2026

Abstract

We introduce a new goodness-of-fit test for count data on $\mathbb{N}$ for the Zeta distribution with unknown parameter. The test is built on a Stein-type characterization that uses, as Stein operator, the infinitesimal generator of a birth–death process whose stationary distribution is Zeta. The resulting L^2 -type statistic is shown to be omnibus consistent, and we establish the limit null behavior as well as the validity of the associated parametric bootstrap procedure. In a Monte Carlo simulation study, we compare the proposed test with the only existing Zeta-specific consistent procedure of Meintanis (2009), as well as with more general competitors based on empirical distribution functions, kernel Stein discrepancies and other Stein-type characterizations.

Keywords Goodness-of-fit · Zeta distribution · Discrete data · Hilbert-space valued random elements · Parametric bootstrap

1 Introduction

In this paper we study goodness-of-fit testing for the Zeta (also called Riemann–Zeta, Zipf or discrete Pareto) family of distributions. This one-parameter family emerges as a canonical model for heavy-tailed count data on $\mathbb{N}$, with probability mass function

$$p_s(k) = \frac{k^{-s}}{\zeta(s)}, \quad k \in \mathbb{N}, \quad s > 1,$$

✉ Bruno Ebner
Bruno.Ebner@kit.edu

Daniel Hlubinka
daniel.hlubinka@matfyz.cuni.cz

¹ Institute of Stochastics, Karlsruhe Institute of Technology (KIT), Englerstr. 2, 76133 Karlsruhe, Germany

² Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, Praha 18675, Czech Republic

where $\zeta(s) = \sum_{k=1}^{\infty} k^{-s}$, $s > 1$, denotes the Riemann zeta function. Zeta and related Zipf-type laws are widely used in applications where empirical frequencies decay approximately like a power of the rank, such as word counts in natural language, city-size distributions, and degree distributions in large networks; see, for example, Pareto (1896), Lotka (1926), Zipf (1949), Mandelbrot (1961), Gabaix (1999), Barabási and Albert (1999), Clauset et al. (2009), Newman (2005). Despite this broad range of applications, methodological work on formal goodness-of-fit tests specifically tailored to the Zeta family is rather limited. General omnibus procedures based on empirical distribution functions or tail fitting can be adapted to the Zeta model, but only a few tests exploit its particular structure, and the behavior of different procedures under realistic alternatives has, to date, not been systematically investigated. The aim of this paper is to fill this gap by developing a new procedure including its theory, and by comparing several Stein-based, Mellin-transform-based, and characterization-based tests for the Zeta law.

To be specific, we denote the Zeta family of distributions by $\mathcal{Z} = \{\text{Zeta}(s) : s > 1\}$, where s is a shape parameter. Note that the moments of the Zeta law only exist for $s > 2$ (mean) and $s > 3$ (variance), so we potentially have a heavy tailed discrete distribution. This fact is one reason inference must be tail-aware (Johnson et al. 2005; Newman 2005). Let $X_1, \dots, X_n$ be independent and identically distributed copies of a random variable X taking values in $\mathbb{N}$ and denote the distribution of X by $\mathbb{P}^X$. The testing problem of interest is to test the composite hypothesis

$$H_0 : \mathbb{P}^X \in \mathcal{Z} \quad (1)$$

against general alternatives based on the sample $X_1, \dots, X_n$.

The existing literature on this testing problem is relatively limited. The earliest goodness-of-fit procedure for the Zeta family was proposed in Luong and Doray (1996), but their fixed-dimensional tests are not consistent to all fixed alternatives. General omnibus procedures based on the empirical distribution function, such as Kolmogorov–Smirnov and Cramér–von Mises type tests proposed in Henze (1996), can be adapted to the present setting. Kernelized Stein discrepancy tests for discrete sample spaces are developed in Yang et al. (2018), but they are formulated only for simple null hypotheses on finite sample spaces, and therefore require suitable modifications to address the composite hypothesis in (1). In Betsch et al. (2022), the authors propose general tests derived from characterizations of probability mass functions, which are also applicable to this problem. To the best of our knowledge, the only consistent procedure specifically tailored to the Zeta family of distributions is that of Meintanis (2009), where the empirical Mellin transform of inverse moments is incorporated into a weighted L^2 -type statistic.

We propose a new competitor to Meintanis (2009) based on a Stein characterization of the Zeta law using the generator approach, see (Barbour 1988, 1990), by providing a birth and death process whose stationary distribution is precisely the $\text{Zeta}(s)$ distribution and which then induces using the infinitesimal generator a suitable Stein operator. Note that this is the first goodness-of-fit test that uses the generator approach to provide the Stein characterization and then we follow (Anastasiou et al. 2023), Section 5.2, to propose a weighted L^2 -type test statistic. First, consider a birth–death

process $(Y_t)_{t \geq 0}$ on $\mathbb{N}$ with birth and death rates

$$\lambda_k = 1, \quad \mu_k = \left(\frac{k}{k-1} \right)^s, \quad k \geq 2,$$

and $\mu_1 = 0$. These rates satisfy the detailed balance condition

$$p_s(k)\lambda_k = p_s(k+1)\mu_{k+1}, \quad k \geq 1,$$

so that $P_s = \text{Zeta}(s)$ is the unique stationary (and reversible) distribution of this Markov chain. The infinitesimal generator L_s of the process acts on functions $f : \mathbb{N} \rightarrow \mathbb{R}$ as

$$L_s f(k) = \lambda_k (f(k+1) - f(k)) + \mu_k (f(k-1) - f(k)), \quad k \geq 1, \quad (2)$$

with the convention that the second term vanishes for $k = 1$, i.e.

$$L_s f(1) = f(2) - f(1), \quad L_s f(k) = (f(k+1) - f(k)) + \left(\frac{k}{k-1} \right)^s (f(k-1) - f(k)), \quad k \geq 2.$$

Following the generator approach (Barbour 1988, 1990), this infinitesimal generator is a Stein operator for the stationary distribution, see also Section 1 in Eichelsbacher and Reinert (2008). In our case this leads to the following characterization of the Zeta distribution.

Theorem 1 *Let $X \sim P_s = \text{Zeta}(s)$, for some $s > 1$. Then we have for all functions $f : \mathbb{N} \rightarrow \mathbb{R}$ such that the expectations below exist,*

$$\mathbb{E}_{P_s}[L_s f(X)] = 0.$$

Conversely, if a probability measure P on $\mathbb{N}$ satisfies $\mathbb{E}_P[L_s f(X)] = 0$ for all bounded f , then $P = P_s$.

Proof The forward direction follows from the reversibility condition $p_s(k) = p_s(k+1)\mu_{k+1}$, which ensures that for all f ,

$$\begin{aligned} \mathbb{E}_{P_s}[L_s f(X)] &= \sum_{k=1}^{\infty} p_s(k) L_s f(k) = \sum_{k=1}^{\infty} p_s(k) \lambda_k (f(k+1) - f(k)) \\ &\quad + \sum_{k=2}^{\infty} p_s(k) \mu_k (f(k-1) - f(k)) = 0. \end{aligned}$$

The converse follows from the fact that the stationary distribution of a positive recurrent birth–death process is unique. $\square$

Corollary 2 For the family of test functions $\{f_t(x) = t^x : t \in [0, 1]\}$ and $s > 1$ we have

$$L_s f_t(1) = t^2 - t, \quad L_s f_t(k) = (t^{k+1} - t^k) + \left(\frac{k}{k-1}\right)^s (t^{k-1} - t^k), \quad k \geq 2, \quad (3)$$

and the class of test functions $\{f_t : t \in [0, 1]\}$ is characterizing, i.e. $\mathbb{E}[L_s f_t(X)] = 0$ for all $t \in [0, 1]$ if and only if $X \sim P_s$.

Proof We apply (2) to $f_t(x) = t^x$, $t \in [0, 1]$, which yields directly (3). To show that $\{f_t : t \in [0, 1]\}$ is characterizing, let q be any probability mass function (pmf) on $\mathbb{N}$ satisfying $\mathbb{E}_q[(L_s f_t)(X)] = 0$ for all $t \in [0, 1]$. Substituting the above expression gives

$$0 = \sum_{k \geq 1} q(k) \left(t^{k+1} - t^k + \left(\frac{k}{k-1}\right)^s (t^{k-1} - t^k) \mathbf{1}_{\{k > 1\}} \right).$$

Define the generating functions $G_q(t) = \sum_{k \geq 1} q(k)t^k$, $H_q(t) = \sum_{k \geq 1} q(k)\mu_k t^k$. Then for all $t \in [0, 1]$,

$$0 = (1-t)(H_q(t) - tG_q(t)),$$

implying $H_q(t) = tG_q(t)$ as an identity of power series. Matching coefficients gives $q(k)\mu_k = q(k-1)$ for $k \geq 2$, hence

$$q(k) = q(1) \prod_{j=2}^k \frac{1}{\mu_j} = q(1) \prod_{j=2}^k \left(\frac{j-1}{j}\right)^s = q(1)k^{-s}.$$

Normalization yields $q(1) = 1/\zeta(s)$, i.e. $q(k) = k^{-s}/\zeta(s)$. Thus the family $\{f_t : t \in [0, 1]\}$ uniquely characterizes the Zeta(s) law. $\square$

Given i.i.d. data $X_1, \dots, X_n$ from an unknown distribution P , we define the empirical Stein process as

$$Z_n(t; s) = \frac{1}{\sqrt{n}} \sum_{i=1}^n L_s f_t(X_i), \quad t \in [0, 1],$$

and to test the composite hypothesis in (1) we propose the test statistic

$$T_n = \int_0^1 |Z_n(t; \hat{s}_n)|^2 w(t) dt, \quad (4)$$

where $\hat{s}_n$ is a consistent estimator of s and w is an almost everywhere positive and integrable weight function satisfying $\int_0^1 w(t)dt < \infty$. Since under the null hypothesis Z_n is close to zero, T_n should produce small values and hence we reject H_0 for 'large' values of the test statistic. Since the limit null distribution of T_n clearly depends on

the true but unknown shape parameter s_0 (say) of the Zeta distribution, a resampling technique is necessary to compute suitable critical values.

For the weight function $w(t) = (1 - t)^\beta$, $\beta \geq 0$, we have after straightforward evaluation of integrals the explicit and integration-free representation

$$\begin{aligned}
 T_{n,\beta} = & \frac{1}{n} \sum_{j,k=1}^n \left\{ \frac{2\mathbf{1}\{X_j = X_k = 1\}}{(3 + \beta)(4 + \beta)(5 + \beta)} \right. \\
 & - 2 \left[\left(\frac{X_j}{X_j - 1} \right)^{\hat{s}_n} - \frac{X_j + 1}{\beta + X_j + 4} \right] B(X_j + 1, 3 + \beta) \mathbf{1}\{X_j \geq 2, X_k = 1\} \\
 & + \left[B(X_{j,k}^+ - 1, 3 + \beta) \left(\frac{X_j}{X_j - 1} \right)^{\hat{s}_n} \left(\frac{X_k}{X_k - 1} \right)^{\hat{s}_n} \right. \\
 & - B(X_{j,k}^+, 3 + \beta) \left(\left(\frac{X_j}{X_j - 1} \right)^{\hat{s}_n} + \left(\frac{X_k}{X_k - 1} \right)^{\hat{s}_n} \right) \\
 & \left. \left. + B(X_{j,k}^+ + 1, 3 + \beta) \right] \mathbf{1}\{X_j, X_k \geq 2\} \right\}, \quad (5)
 \end{aligned}$$

where $\mathbf{1}$ denotes the indicator function, $B(\cdot, \cdot)$ is the beta function and $X_{j,k}^+ = X_j + X_k$.

The remainder of the paper is structured as follows. In the next section we derive the asymptotic properties of the test statistic by applying a central limit theorem for triangular arrays of Hilbert-space-valued random elements, and we establish both the consistency of the test and the validity of the proposed parametric bootstrap scheme. Section 3 presents a comprehensive Monte Carlo study for the testing problem in (1), and Sect. 4 concludes with an outlook and several directions for future research.

2 Asymptotics

In this section we derive asymptotic properties of the proposed tests. Let $L_w^2 = L^2([0, 1], \mathcal{B}_{[0,1]}, w(t)dt)$ be the separable Hilbert space (of equivalence classes) of Borel-measurable functions $g : [0, 1] \rightarrow \mathbb{R}$ satisfying $\|g\|_{L_w^2}^2 = \int_0^1 g^2(t)w(t)dt < \infty$ with respect to the measurable positive weight function $w(\cdot)$. The scalar product on L_w^2 is defined by $\langle g, h \rangle = \int_0^1 g(t)h(t)w(t)dt$. This is a suitable setting, since the test statistic admits the representation

$$T_n = \|Z_n(\cdot; \hat{s}_n)\|_{L_w^2}^2.$$

2.1 Limit null distribution

Let $X_{n,1}, \dots, X_{n,n}$ be a triangular array of row-wise independent and identically distributed copies of $X_n \sim P_{s_n}$, where (s_n) is a sequence of numbers with $s_n > 1$ and $\lim_{n \rightarrow \infty} s_n = s_0 > 1$. We choose as consistent estimator the maximum-likelihood

estimator (MLE) $\widehat{s}_n$, which satisfies

$$\frac{\zeta'(\widehat{s}_n)}{\zeta(\widehat{s}_n)} = -\frac{1}{n} \sum_{i=1}^n \log X_{n,i}. \quad (6)$$

Note that there exists no closed form formula for $\widehat{s}_n$, so (6) has to be solved numerically, see (Johnson et al. 2005, Section 11.2.20) for more details on the estimation of the shape parameter of the Zeta law. Since it produces dependencies in the sum of the empirical Stein process, we provide the Bahadur representation of the estimator and related helping processes that share the same limit distribution. Denote by $u(x; s) = \partial_s \log p_s(x) = -\log(x) - \zeta'(s)/\zeta(s)$, $s > 1$, the Zeta score function and by $I(s) = \text{Var}(u(X; s)) = \zeta''(s)/\zeta(s) - (\zeta'(s)/\zeta(s))^2$, $s > 1$, the Fisher information. Then applying the Bahadur representation, see Corollary 10.16 in Henze (2024), we get

$$\begin{aligned} \sqrt{n}(\widehat{s}_n - s_n) &= \frac{I(s_n)^{-1}}{\sqrt{n}} \sum_{j=1}^n u(X_{n,j}; s_n) + o_{\mathbb{P}}(1) \\ &= -\frac{1}{\sqrt{n} \left(\frac{\zeta''(s_n)}{\zeta(s_n)} - \left(\frac{\zeta'(s_n)}{\zeta(s_n)} \right)^2 \right)} \sum_{j=1}^n \left(\log(X_{n,j}) + \frac{\zeta'(s_n)}{\zeta(s_n)} \right) \\ &\quad + o_{\mathbb{P}}(1). \end{aligned} \quad (7)$$

Define the functions

$$h_s(x, t) = -t(1-t)\mathbf{1}\{x=1\} + (1-t)t^{x-1} \left[\left(\frac{x}{x-1} \right)^s - t \right] \mathbf{1}\{x \geq 2\}, \quad t \in [0, 1],$$

and

$$\frac{\partial h_s(x, t)}{\partial s} = (1-t)t^{x-1} \left(\frac{x}{x-1} \right)^s \log \left(\frac{x}{x-1} \right) \mathbf{1}\{x \geq 2\} =: g_s(x, t), \quad t \in [0, 1].$$

Now, consider the helping processes

$$\widetilde{Z}_n(t; s_n) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \{h_{s_n}(X_{n,j}, t) + g_{s_n}(X_{n,j}, t)(\widehat{s}_n - s_n)\}, \quad t \in [0, 1],$$

and

$$\widehat{Z}_n(t; s_n) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left\{ h_{s_n}(X_{n,j}, t) + a_n(t) I(s_n)^{-1} u(X_{n,j}; s_n) \right\}, \quad t \in [0, 1],$$

where $a_n(t) = \mathbb{E}[g_{s_n}(X_n, t)] = \frac{1-t}{\zeta(s_n)} \sum_{k=2}^{\infty} t^{k-1} (k-1)^{-s_n} \log \left(\frac{k}{k-1} \right)$, $t \in [0, 1]$, $n \in \mathbb{N}_0$. The next lemma shows that the helping processes are asymptotically equivalent.

Lemma 3 *Under the standing assumptions, we have*

$$\|Z_n(\cdot; \widehat{s}_n) - \widetilde{Z}_n(\cdot; s_n)\|_{L_w^2} = o_{\mathbb{P}}(1) \quad \text{and} \quad \|\widetilde{Z}_n(\cdot; s_n) - \widehat{Z}_n(\cdot; s_n)\|_{L_w^2} = o_{\mathbb{P}}(1).$$

Proof Let $W = \int_0^1 w(t) dt < \infty$, and $\Delta_n := \widehat{s}_n - s_n$. Note that the Bahadur representation (7) implies $\sqrt{n} \Delta_n = O_{\mathbb{P}}(1)$, and hence $\Delta_n = o_{\mathbb{P}}(1)$. Next, let $\eta := (s_0 - 1)/4$, and define the compact interval $K := [s_0 - 2\eta, s_0 + 2\eta] \subset (1, \infty)$. For all sufficiently large n , $|s_n - s_0| \leq \eta$, and on the event $E_n := \{|\Delta_n| \leq \eta\}$, whose probability tends to one, the line segment joining s_n and $\widehat{s}_n$ is contained in K . We next establish uniform bounds for the derivatives of h_s . Since $x/(x-1) \leq 2$, and $\log(x/(x-1)) \leq \log 2$ for $x \geq 2$, we obtain, uniformly in $s \in K$ and $x \in \mathbb{N}$,

$$\begin{aligned} \|g_s(x, \cdot)\|_{L_w^2}^2 &= \left(\frac{x}{x-1}\right)^{2s} \left[\log\left(\frac{x}{x-1}\right)\right]^2 \int_0^1 (1-t)^2 t^{2x-2} w(t) dt \mathbf{1}_{\{x \geq 2\}} \\ &\leq 2^{3s_0-1} (\log 2)^2 W =: M_1^2. \end{aligned}$$

Thus, we have

$$\sup_{s \in K} \sup_{x \in \mathbb{N}} \|g_s(x, \cdot)\|_{L_w^2} \leq M_1.$$

Furthermore,

$$\partial_s^2 h_s(x, t) = (1-t)t^{x-1} \left(\frac{x}{x-1}\right)^s \left[\log\left(\frac{x}{x-1}\right)\right]^2 \mathbf{1}_{\{x \geq 2\}}.$$

Therefore

$$\begin{aligned} \|\partial_s^2 h_s(x, \cdot)\|_{L_w^2}^2 &= \left(\frac{x}{x-1}\right)^{2s} \left[\log\left(\frac{x}{x-1}\right)\right]^4 \int_0^1 (1-t)^2 t^{2x-2} w(t) dt \mathbf{1}_{\{x \geq 2\}} \\ &\leq 2^{3s_0-1} (\log 2)^4 W =: M_2^2. \end{aligned}$$

Hence, we have

$$\sup_{s \in K} \sup_{x \in \mathbb{N}} \|\partial_s^2 h_s(x, \cdot)\|_{L_w^2} \leq M_2.$$

Taylor's formula with integral remainder gives

$$h_{\widehat{s}_n}(x, t) - h_{s_n}(x, t) - g_{s_n}(x, t) \Delta_n = \Delta_n^2 \int_0^1 (1-\lambda) \partial_s^2 h_{s_n + \lambda \Delta_n}(x, t) d\lambda.$$

Consequently, on E_n , $\|Z_n(\cdot; \widehat{s}_n) - \widetilde{Z}_n(\cdot; s_n)\|_{L_w^2} \leq M_2 \sqrt{n} \Delta_n^2 / 2 = o_{\mathbb{P}}(1)$. Since $\mathbb{P}(E_n) \rightarrow 1$, the first assertion follows.

For the second assertion, define

$$G_n := \frac{1}{n} \sum_{j=1}^n g_{s_n}(X_{n,j}, \cdot), \quad \text{and} \quad U_n := \frac{1}{\sqrt{n}} \sum_{j=1}^n u(X_{n,j}, s_n).$$

By row-wise independence and centering,

$$\mathbb{E} \|G_n - a_n\|_{L_w^2}^2 = \frac{1}{n} \mathbb{E} \|g_{s_n}(X_{n,1}, \cdot) - a_n\|_{L_w^2}^2 \leq \frac{M_1^2}{n}.$$

Thus $\|G_n - a_n\|_{L_w^2} = o_{\mathbb{P}}(1)$. Moreover, $\|a_n\|_{L_w^2} \leq M_1$, and hence $\|G_n\|_{L_w^2} = O_{\mathbb{P}}(1)$. Let $\rho_n := \sqrt{n} \Delta_n - I(s_n)^{-1} U_n = o_{\mathbb{P}}(1)$. From the definitions of the helping processes,

$$\tilde{Z}_n - \hat{Z}_n = \rho_n G_n + I(s_n)^{-1} U_n (G_n - a_n).$$

Therefore, $\|\tilde{Z}_n - \hat{Z}_n\|_{L_w^2} \leq |\rho_n| \|G_n\|_{L_w^2} + |I(s_n)^{-1} U_n| \|G_n - a_n\|_{L_w^2} = o_{\mathbb{P}}(1)$. This proves both assertions. $\square$

Theorem 4 *Under the triangular array stated at the beginning of this section, we have as $n \rightarrow \infty$*

$$T_n \xrightarrow{d} \|Z\|_{L_w^2}^2,$$

where $Z(\cdot)$ is a centered Gaussian process in L_w^2 with covariance kernel

$$\begin{aligned} C(r, t) &= \frac{rt(1-r)(1-t)}{\zeta(s_0)} - \frac{a_0(r)a_0(t)}{I(s_0)} \\ &+ \frac{(1-r)(1-t)}{\zeta(s_0)} \sum_{k=2}^{\infty} k^{-s_0} (rt)^{k-1} \left(\left(\frac{k}{k-1} \right)^{s_0} - r \right) \left(\left(\frac{k}{k-1} \right)^{s_0} - t \right), \quad r, t \in [0, 1]. \end{aligned}$$

Proof By Lemma 3, it suffices to prove that $\hat{Z}_n(\cdot; s_n) \xrightarrow{\mathcal{D}} Z$ in L_w^2 . To show this we use Theorem 17.30 in Henze (2024). Define

$$w_n(x) := h_{s_n}(x, \cdot) + a_n(\cdot) I(s_n)^{-1} u(x; s_n), \quad Y_{n,j} := \frac{1}{\sqrt{n}} w_n(X_{n,j}).$$

and

$$w_0(x) := h_{s_0}(x, \cdot) + a_0(\cdot) I(s_0)^{-1} u(x; s_0), \quad W_0 := w_0(X),$$

where $X \sim P_{s_0}$. Then

$$\hat{Z}_n(\cdot; s_n) = \sum_{j=1}^n Y_{n,j}, \quad \mathbb{E}[Y_{n,j}] = 0.$$

Since $s_n \rightarrow s_0 > 1$, the sequence (s_n) is eventually contained in a compact subset of $(1, \infty)$. Thus, after putting $m = k - 1$,

$$a_n = \sum_{m=1}^{\infty} c_{n,m} \phi_m, \quad c_{n,m} := \frac{m^{-s_n}}{\zeta(s_n)} \log \left(1 + \frac{1}{m} \right), \quad \phi_m(t) := (1-t)t^m.$$

Likewise,

$$a_0 = \sum_{m=1}^{\infty} c_{0,m} \phi_m, \quad c_{0,m} := \frac{m^{-s_0}}{\zeta(s_0)} \log \left(1 + \frac{1}{m} \right).$$

For each fixed m , $c_{n,m} \rightarrow c_{0,m}$. Moreover, writing $W = \int_0^1 w(t) dt$, we have

$$\|\phi_m\|_{L_w^2}^2 = \int_0^1 (1-t)^2 t^{2m} w(t) dt \leq W.$$

Choose $s_* \in (1, s_0)$. Since $s_n \rightarrow s_0$, eventually $s_n \geq s_*$. Using $\zeta(s) > 1$ and $\log(1 + 1/m) \leq 1/m$, we obtain

$$(|c_{n,m}| + |c_{0,m}|) \|\phi_m\|_{L_w^2} \leq 2\sqrt{W} m^{-(s_*+1)}.$$

The right-hand side is summable. Hence, by dominated convergence for series,

$$\|a_n - a_0\|_{L_w^2} \leq \sum_{m=1}^{\infty} |c_{n,m} - c_{0,m}| \|\phi_m\|_{L_w^2} \rightarrow 0$$

as $n \rightarrow \infty$. Consequently, $w_n(x) \rightarrow w_0(x)$ in L_w^2 for every fixed $x \in \mathbb{N}$. Since $s_n \rightarrow s_0 > 1$, there are $1 < s_* < s^* < \infty$ such that $s_n \in [s_*, s^*]$ for all sufficiently large n . Moreover,

$$\sup_n \|a_n\|_{L_w^2} < \infty, \quad \sup_n I(s_n)^{-1} < \infty, \quad \|h_{s_n}(x, \cdot)\|_{L_w^2} \leq C,$$

and $|u(x; s_n)| \leq C(1 + \log x)$, where C is a positive constant. Hence

$$\|w_n(x)\|_{L_w^2} \leq C(1 + \log x).$$

Since $p_{s_n}(x) \leq x^{-s_*}$, it follows that, for $q \in \{2, 4\}$,

$$p_{s_n}(x) \|w_n(x)\|_{L_w^2}^q \leq C x^{-s_*} (1 + \log x)^q, \quad x \in \mathbb{N}. \quad (8)$$

The right-hand side is summable over x .

Let C_n be the covariance operator of $\widehat{Z}_n(\cdot; s_n) = \sum_{j=1}^n Y_{n,j}$. Since $Y_{n,j} = w_n(X_{n,j})/\sqrt{n}$ and the summands are row-wise independent and centered,

$$C_n = \text{Cov}(w_n(X_{n,1})).$$

Let $(e_k)_{k \geq 1}$ be an orthonormal basis of L_w^2 . By (8) and dominated convergence, for every $k, \ell \geq 1$,

$$\lim_{n \rightarrow \infty} \langle C_n e_k, e_\ell \rangle_{L_w^2} = \mathbb{E} \left[\langle W_0, e_k \rangle_{L_w^2} \langle W_0, e_\ell \rangle_{L_w^2} \right] =: \gamma_{k\ell}.$$

Thus condition (a) of Theorem 17.30 holds. Likewise, by Parseval's identity and dominated convergence,

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} \langle C_n e_k, e_k \rangle_{L_w^2} = \mathbb{E} \|W_0\|_{L_w^2}^2 = \sum_{k=1}^{\infty} \gamma_{kk} < \infty,$$

which verifies condition (b). Finally, for every $\varepsilon > 0$ and $k \geq 1$,

$$\begin{aligned} L_n(\varepsilon, e_k) &= \mathbb{E} \left[|\langle w_n(X_{n,1}), e_k \rangle_{L_w^2}|^2 \mathbf{1} \left\{ |\langle w_n(X_{n,1}), e_k \rangle_{L_w^2}| > \varepsilon \sqrt{n} \right\} \right] \\ &\leq \frac{1}{\varepsilon^2 n} \mathbb{E} \|w_n(X_{n,1})\|_{L_w^2}^4 \longrightarrow 0, \end{aligned}$$

because (8) implies $\sup_n \mathbb{E} \|w_n(X_{n,1})\|_{L_w^2}^4 < \infty$. Hence condition (c) holds, and Theorem 17.30 in Henze (2024) yields

$$\widehat{Z}_n(\cdot; s_n) \xrightarrow{\mathcal{D}} Z \quad \text{in } L_w^2.$$

It remains to identify the covariance kernel of Z . Since the limiting covariance matrix satisfies $\gamma_{k\ell} = \mathbb{E} \left[\langle W_0, e_k \rangle_{L_w^2} \langle W_0, e_\ell \rangle_{L_w^2} \right]$, the covariance operator of Z is the covariance operator of W_0 . Thus

$$C(r, t) = \mathbb{E}[W_0(r)W_0(t)], \quad r, t \in [0, 1].$$

Differentiating the identity $\mathbb{E}[h_s(X, t)] = 0$ with respect to s at $s = s_0$ gives

$$0 = \mathbb{E}[g_{s_0}(X, t)] + \mathbb{E}[h_{s_0}(X, t)u(X; s_0)].$$

We obtain $\mathbb{E}[h_{s_0}(X, t)u(X; s_0)] = -a_0(t)$. Also, $\mathbb{E}[u(X; s_0)^2] = I(s_0)$. Expanding the covariance therefore yields

$$C(r, t) = \mathbb{E}[h_{s_0}(X, r)h_{s_0}(X, t)] - \frac{a_0(r)a_0(t)}{I(s_0)}.$$

Using the definition of h_{s_0} , we have

$$\begin{aligned} \mathbb{E}[h_{s_0}(X, r)h_{s_0}(X, t)] &= \frac{rt(1-r)(1-t)}{\zeta(s_0)} \\ &\quad + \frac{(1-r)(1-t)}{\zeta(s_0)} \sum_{x=2}^{\infty} x^{-s_0} (rt)^{x-1} \left[\left(\frac{x}{x-1} \right)^{s_0} - r \right] \left[\left(\frac{x}{x-1} \right)^{s_0} - t \right]. \end{aligned}$$

This is the covariance kernel stated in the theorem. Finally, Lemma 3 gives

$$\|Z_n(\cdot; \hat{s}_n) - \hat{Z}_n(\cdot; s_n)\|_{L_w^2} \leq \|Z_n(\cdot; \hat{s}_n) - \tilde{Z}_n(\cdot; s_n)\|_{L_w^2} + \|\tilde{Z}_n(\cdot; s_n) - \hat{Z}_n(\cdot; s_n)\|_{L_w^2} = o_{\mathbb{P}}(1).$$

Hence, by Slutsky's theorem, $Z_n(\cdot; \hat{s}_n) \xrightarrow{\mathcal{D}} Z$ in L_w^2 . Since the map $f \mapsto \|f\|_{L_w^2}^2$ is continuous, the continuous mapping theorem gives $T_n = \|Z_n(\cdot; \hat{s}_n)\|_{L_w^2}^2 \xrightarrow{\mathcal{D}} \|Z\|_{L_w^2}^2$. This completes the proof. $\square$

2.2 Consistency of the testing procedure

We now state that the test which rejects the hypothesis H_0 for large values of T_n is consistent against general alternatives. Hereafter, we consider an i.i.d. sequence $(X_n)_{n \in \mathbb{N}}$ of copies of X , where X is a non-degenerate positive integer-valued random variable satisfying $\mathbb{E}[X^2] < \infty$. Moreover, we assume that there is $s_0 > 1$ such that

$$\hat{s}_n \xrightarrow{\text{a.s.}} s_0, \quad \text{as } n \rightarrow \infty,$$

where $\hat{s}_n$ is the maximum likelihood estimator as in the previous section. The following result is a direct consequence of a Taylor expansion and Fatou's lemma.

Theorem 5 *Under the stated conditions, we have*

$$\liminf_{n \rightarrow \infty} \frac{T_n}{n} \geq \Lambda_{s_0, w} \quad \mathbb{P}\text{-a.s.},$$

where

$$\Lambda_{s_0, w} = \int_0^1 \mathbb{E}[h_{s_0}(X, t)]^2 w(t) dt.$$

Note that by Corollary 2 we have $\Lambda_{s_0, w} = 0$ if and only if $X \sim P_{s_0}$. This implies that T_n is consistent for each non-Zeta alternative satisfying the stated assumptions, since T_n is diverging almost surely to infinity as $n \rightarrow \infty$.

Remark 1 Under a fixed alternative, the MLE $\hat{s}_n$ converges almost surely to s_0 if the limit

$$s_0 = \arg \max_{s > 1} \mathbb{E}[\log p_s(X)] = \arg \max_{s > 1} \mathbb{E}[-s \log X - \log \zeta(s)]$$

exists and is unique, where X follows the true distribution P . This holds when $\mathbb{E}[|\log X|] < \infty$ and the Kullback–Leibler divergence between P and $\text{Zeta}(s)$ is minimized at a unique $s_0 > 1$.

2.3 Parametric bootstrap procedure

Since the limit null distribution of T_n in Theorem 4 depends on the unknown parameter $s_0 > 1$ of the underlying Zeta distribution, we propose a parametric bootstrap procedure to obtain critical values. For a sample $X_1, \dots, X_n$ satisfying the assumptions above, we compute the MLE $\hat{s}_n = \hat{s}_n(X_1, \dots, X_n)$. We then generate a bootstrap sample of size n , say $X_1^*, \dots, X_n^*$, following the $\text{Zeta}(\hat{s}_n)$ distribution, estimate the parameter s from $X_1^*, \dots, X_n^*$ (denote it $\hat{s}_n^*$), and calculate the test statistic T_n^* . By repeating this procedure b times, we obtain $T_{n,1}^*, \dots, T_{n,b}^*$ and compute the empirical distribution function

$$H_{n,b}^*(t) = \frac{1}{b} \sum_{i=1}^b \mathbf{1}\{T_{n,i}^* \leq t\}, \quad t \geq 0.$$

Given the nominal level $\alpha \in (0, 1)$, we use the empirical $(1 - \alpha)$ -quantile:

$$c_{n,b}^*(\alpha) = H_{n,b}^{*-1}(1 - \alpha) = \begin{cases} T_{b(1-\alpha):b}^*, & b(1 - \alpha) \in \mathbb{N}, \\ T_{\lfloor b(1-\alpha) \rfloor + 1:b}^*, & \text{otherwise,} \end{cases}$$

where $T_{1:b}^*, \dots, T_{b:b}^*$ are the order statistics. We reject H_0 if $T_n > c_{n,b}^*(\alpha)$. Next we show the validity of the bootstrap procedure. Denote by

$$H_{n,s}(t) := \mathbb{P}_s(T_n \leq t), \quad H_s(t) := \mathbb{P}\left(\|Z_s\|_{L_w^2}^2 \leq t\right), \quad t \geq 0,$$

where Z_s denotes the centered Gaussian element from Theorem 4 corresponding to the parameter value $s > 1$. Moreover, let $\mathcal{X}_n := \sigma(X_1, \dots, X_n)$ and define the conditional bootstrap distribution by

$$G_n^*(t) := \mathbb{P}(T_n^* \leq t \mid \mathcal{X}_n), \quad t \geq 0.$$

The corresponding bootstrap critical value is

$$c_n^*(\alpha) := \inf\{t \geq 0 : G_n^*(t) \geq 1 - \alpha\}, \quad \alpha \in (0, 1).$$

Conditionally on $\mathcal{X}_n$, the bootstrap sample $X_1^*, \dots, X_n^*$ is i.i.d. $\text{Zeta}(\hat{s}_n)$, and the bootstrap statistic T_n^* is computed from that sample in exactly the same way as T_n . Hence, for every $t \geq 0$,

$$G_n^*(t) = H_{n,\hat{s}_n}(t) \quad \mathbb{P}_{s_0}\text{-a.s.}$$

Assume now that the null hypothesis holds, i.e. $X_1, X_2, \dots$ are i.i.d. $\text{Zeta}(s_0)$ with $s_0 > 1$, and that $\hat{s}_n \rightarrow s_0$ almost surely. Then there exists a set A with $\mathbb{P}_{s_0}(A) = 1$ such

that, for every $\omega \in A$, the deterministic sequence $s_n(\omega) = \widehat{s}_n(\omega)$ satisfies $s_n(\omega) \rightarrow s_0$. Fix $\omega \in A$. Conditional on the data, the bootstrap row $X_{n,1}^*(\omega), \dots, X_{n,n}^*(\omega)$ is i.i.d. $\text{Zeta}(s_n(\omega))$. Therefore, Theorem 4 applied to the deterministic triangular array with row parameter sequence $s_n(\omega)$ yields

$$G_n^*(\omega, t) = H_{n,s_n(\omega)}(t) \rightarrow H_{s_0}(t)$$

for every $t \geq 0$. Since H_{s_0} is continuous, Pólya's theorem implies

$$\sup_{t \geq 0} |G_n^*(\omega, t) - H_{s_0}(t)| \rightarrow 0, \quad \omega \in A.$$

Thus, the parametric bootstrap is valid since the conditional law of T_n^* given the sample converges to the same null limit law as T_n , for $\mathbb{P}_{s_0}$ -almost every sample sequence. In particular, by continuity and strict monotonicity of H_{s_0} , see also Remark 2,

$$c_n^*(\alpha) \rightarrow H_{s_0}^{-1}(1 - \alpha) \quad \mathbb{P}_{s_0}\text{-a.s.},$$

and therefore

$$\lim_{n \rightarrow \infty} \mathbb{P}_{s_0}(T_n > c_n^*(\alpha)) = \alpha.$$

Next, let $T_{n,1}^*, \dots, T_{n,b}^*$ be conditionally i.i.d. copies of T_n^* given $\mathcal{X}_n$, and define $H_{n,b}^*$ as above. Then, for each fixed n , the Glivenko–Cantelli theorem yields

$$\sup_{t \geq 0} |H_{n,b}^*(t) - G_n^*(t)| \rightarrow 0, \quad \text{a.s.}$$

with respect to the bootstrap resampling, for $\mathbb{P}_{s_0}$ -almost every realization of the original sample. Hence $H_{n,b}^*$ is a Monte Carlo approximation to the ideal conditional bootstrap law G_n^* , and $c_{n,b}^*(\alpha)$ is the corresponding Monte Carlo approximation to $c_n^*(\alpha)$. Consequently,

$$\lim_{n \rightarrow \infty} \lim_{b \rightarrow \infty} \mathbb{P}_{s_0}(T_n > c_{n,b}^*(\alpha)) = \alpha.$$

If $X_1, \dots, X_n$ do not follow a Zeta distribution, then by Theorem 5, $\Lambda_{s_0,w} > 0$, so

$$\lim_{n \rightarrow \infty} \lim_{b \rightarrow \infty} \mathbb{P}(T_n > c_{n,b}^*(\alpha)) = 1.$$

Thus, the bootstrap test is consistent against any fixed alternative distribution satisfying the stated assumptions.

Remark 2 Since the limit random element in Theorem 4 is a centered Gaussian process taking values in L_w^2 with known covariance kernel $C(s, t)$, $s, t \in [0, 1]$, it is well-known by the Karhunen–Loève expansion (see Duan and Wang (2014), Chapter 3), that the distribution of $\|Z\|_{L_w^2}^2$ corresponds to $\sum_{j=1}^{\infty} \lambda_j(s_0) N_j^2$, where N_j are i.i.d. standard normal random variables and $(\lambda_j(s_0))_{j \in \mathbb{N}}$ is a sequence of decaying positive

eigenvalues defined by the (linear homogeneous Fredholm of the second kind) integral equation

$$\lambda f(s) = \mathcal{C}f(s), \quad s \in [0, 1],$$

where $\mathcal{C} : L_w^2 \mapsto L_w^2$ is defined as

$$\mathcal{C}f(s) = \int_0^1 C(s, t) f(t) w(t) dt,$$

corresponding to the covariance kernel C of Z in Theorem 4. The eigenvalues obviously depend on the true unknown parameter $s_0 > 1$ as well as the weight function $w(\cdot)$. By the complexity of the kernel C it seems hopeless to find an analytical solution to the eigenvalue problem, although numerical procedures such as the Rayleigh-Ritz method could be applied (Ebner et al. 2025). The limiting distribution is non-degenerate.

Indeed, for any $q \in (0, 1)$, $C(q, q) = \mathbb{E} \left[h_{s_0}(X, q) + \frac{a_0(q)}{I(s_0)} u(X; s_0) \right]^2 \geq 0$. If $C(q, q) = 0$, then, since the Zeta distribution has full support on $\mathbb{N}$, $h_{s_0}(k, q) + \frac{a_0(q)}{I(s_0)} u(k; s_0) = 0$, $k \in \mathbb{N}$. For $k \geq 2$, $h_{s_0}(k, q) \rightarrow 0$ as $k \rightarrow \infty$ exponentially fast, whereas $u(k; s_0) = -\log k - \frac{\xi'(s_0)}{\xi(s_0)} \rightarrow -\infty$. Hence necessarily the constant $a_0(q)/I(s_0) = 0$. But then $h_{s_0}(k, q) = 0$ for all k , which is a contradiction since $h_{s_0}(1, q) = -q(1 - q) \neq 0$. Thus $C(q, q) > 0$ for all $q \in (0, 1)$. Since w is positive, the covariance operator has positive trace, $\text{tr}(C) = \int_0^1 C(q, q) w(q) dq > 0$, and therefore at least one eigenvalue $\lambda_j(s_0)$ is strictly positive. Consequently, the limit law $\|Z\|_{L_w^2}^2 \stackrel{\mathcal{D}}{=} \sum_{j=1}^{\infty} \lambda_j(s_0) N_j^2$ is non-degenerate.

Remark 3 Defining the feature map $\Phi_s(x)(t) = h_s(x, t)$ and the Stein kernel $k_s(x, y) = \langle \Phi_s(x), \Phi_s(y) \rangle_{L_w^2}$, we have for T_n in (4) a V -statistic $T_n = n^{-1} \sum_{j,k=1}^n k_{\hat{s}_n}(X_j, X_k)$, which shows that T_n is a special maximum mean discrepancy (MMD)-type statistic with a Stein-type kernel, and because k_s is defined as an inner product of features, it is automatically positive definite. For more details on this type of test statistic see (Key et al. 2025) and for related theory (Brueck et al. 2025).

3 Simulations

In this section we provide (to the best of our knowledge, the first) comparative Monte Carlo Simulation study for testing the fit to the Zeta family of distributions. In each simulation run, we simulate a data set $X_1, \dots, X_n$ of sample size $n = 100$, and fix the significance level to 5%. Every entry in Table 1 is based on 10000 replications. To ensure an 'apples-to-apples' comparison, we employ for all considered testing procedures the same resampling procedure. Since the parametric bootstrap in Sect. 2.3 is computationally intensive, we employ the warp-speed Monte Carlo method of Giacomini et al. (2013). Hence, for each simulated dataset we compute the test statistic on the original sample and on one bootstrap resample, and then use the resulting pairs to estimate rejection probabilities as the number of Monte Carlo replications grows.

Table 1 Empirical rejection percentages for the different test statistics for a sample size of $n = 100$ and a significance level of 5%

Alt	$T_{n,0}$	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	$T_{n,4}$	$T_{n,5}$	$Z_{n,1}$	$Z_{n,2}$	$Z_{n,3}$	$Z_{n,4}$	$Z_{n,5}$	$Z_{n,6}$	C_n^e	S_{KSD}	S_{BEN}
Zeta(1.5)	5	5	5	5	5	5	5	4	4	4	5	5	5	5	5
Zeta(1.75)	5	5	5	5	5	5	4	4	4	4	4	4	5	5	5
Zeta(2)	5	5	5	4	4	5	4	4	4	4	4	4	5	5	4
Zeta(2.25)	5	5	5	5	5	5	4	4	3	3	3	4	4	5	4
Geom(2.5)	93	93	92	91	91	90	99	99	99	99	99	99	100	100	99
Geom(3)	77	76	75	75	75	74	73	69	66	64	64	67	80	0	73
Geom(3.5)	48	48	48	48	48	48	37	34	30	29	27	27	45	0	39
ZG(3,10,0.2)	4	4	5	4	4	5	3	3	3	3	3	3	4	6	4
ZG(3,10,0.4)	4	5	5	5	5	5	3	3	2	2	2	3	4	7	4
ZG(3,10,0.6)	5	5	5	5	5	5	3	3	2	2	2	3	4	7	5
ZG(3,10,0.8)	5	5	5	5	5	5	3	3	3	2	3	3	4	8	4
ZG(2.5,0.2)	5	5	5	5	5	5	12	14	18	22	26	32	9	42	7
ZG(2.5,0.4)	6	5	6	5	5	5	25	29	33	41	47	51	18	90	24
ZG(2.5,0.6)	6	6	6	6	6	6	32	36	41	47	54	59	27	99	49
ZG(2.5,0.8)	6	6	6	6	6	5	38	42	47	53	60	66	35	100	74
ZG(2,30,0.4)	5	5	5	5	5	5	4	4	5	7	9	12	4	17	5
GZ(2.5,5)	93	92	91	91	90	90	99	99	99	99	99	100	100	96	99
GZ(2.5,10)	93	92	91	91	90	90	99	99	99	99	99	100	100	100	99

Table 1 continued

Alt	$T_{n,0}$	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	$T_{n,4}$	$T_{n,5}$	$Z_{n,1}$	$Z_{n,2}$	$Z_{n,3}$	$Z_{n,4}$	$Z_{n,5}$	$Z_{n,6}$	C_n^e	S_{KSD}	S_{BEN}
GZ(2.5,20)	93	92	92	91	91	90	99	99	99	99	99	99	99	100	98
GZ(2.5,50)	94	93	92	92	91	91	99	99	99	99	99	99	100	100	99
Zipf(1.5,10)	28	27	26	25	24	24	84	87	90	93	95	97	84	100	75
Zipf(1.75,10)	18	17	17	16	16	16	58	61	65	70	75	79	54	100	42
Zipf(2,10)	13	13	13	13	13	13	33	35	36	41	46	52	28	64	22
Zipf(2.25,10)	8	8	8	8	8	8	16	16	16	18	23	26	14	20	12
Zipf(2.5,10)	7	7	7	7	7	7	8	8	8	8	10	11	8	11	7
Zipf(3,10)	5	5	5	5	5	5	4	3	3	3	3	3	4	8	5
Zipf(1.5,20)	15	14	14	13	13	13	70	78	85	90	95	98	67	100	45
Zipf(1.75,20)	10	10	10	9	9	9	36	42	48	55	61	69	30	99	20
Zipf(2,5)	33	31	31	30	30	29	61	61	61	62	66	69	59	28	53
Zipf(2.25,5)	21	20	20	20	20	20	40	40	39	39	41	45	39	8	33
Zipf(2.5,5)	14	14	14	14	13	13	23	22	21	21	22	25	22	6	18
Zipf(3,5)	8	8	8	7	7	7	7	7	7	7	7	7	8	5	7
Zigzag(1.75,0.1)	12	12	12	12	12	12	8	8	10	11	9	5	7	5	11
Zigzag(1.75,0.5)	96	96	96	96	96	96	89	85	75	56	27	12	93	20	*
Zigzag(2,0.1)	11	12	12	12	12	12	8	8	9	10	11	12	9	7	11
Zigzag(2,0.5)	96	96	97	97	97	97	87	84	83	78	66	45	93	25	98
Zigzag(2.5,0.1)	10	10	10	10	10	11	7	7	6	6	6	8	10	6	9
Zigzag(2.5,0.5)	95	95	95	95	95	95	81	76	74	75	76	74	91	26	95
Zigzag(3,0.1)	9	9	9	9	9	9	6	5	5	5	5	5	8	5	7
Zigzag(3,0.5)	90	90	90	90	90	90	70	64	59	59	61	63	84	5	88

Every entry is based on 10000 replications. The entry with * could not be computed due to numerical instability.

There are relatively few goodness-of-fit tests specifically designed for positive integer-valued random variables. We adopt the approaches mentioned in the introduction to compare the finite sample performance of the test statistics. Henze (1996) proposed a test statistic for parametric families of discrete distributions, based on a comparison between the empirical distribution function and the cumulative distribution function with estimated parameters.

Let us briefly describe the test of Henze (1996). We present it here in its specific form for the Zeta distribution with unknown parameter s . Let $X_1, \dots, X_n$ be an i.i.d. sequence of random variables following the Zeta(s) distribution, and let $\widehat{F}_n$ denote the empirical distribution function based on this sample. Furthermore, let $\widehat{s}_n$ be a consistent estimator of the parameter $s > 1$, and let $F(t; \widehat{s}_n)$ be the cumulative distribution function of the Zeta($\widehat{s}_n$) distribution.

We consider two asymptotically equivalent versions of the Cramér–von Mises test statistic, namely

$$C_n^a = n \sum_{k=1}^{\ell} (\widehat{F}_n(k) - F(k; \widehat{s}_n))^2 \mathbb{P}[X = k; \widehat{s}_n] + R(\ell),$$

where $\mathbb{P}[X = k; \widehat{s}_n]$, $k \geq 1$, is the Zeta($\widehat{s}_n$) probability mass function, and $R(\ell)$ is a remainder term, and

$$C_n^e = n \sum_{k=1}^{\infty} (\widehat{F}_n(k) - F(k; \widehat{s}_n))^2 (\widehat{F}_n(k) - \widehat{F}_n(k-1)),$$

where

$$\widehat{F}_n(k) - \widehat{F}_n(k-1) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}\{X_j = k\}, \quad k \geq 1,$$

is the empirical probability mass function. The remainder term $R(\ell)$ of C_n^a depends on ℓ and on the true value of s . Henze (1996) recommends choosing ℓ sufficiently large, in particular at least so large that $n(1 - F(\ell; \widehat{s}_n)) \leq 10^{-4}$. Note that, for C_n^e , there are at most finitely many nonzero summands, namely at most $M = \max\{X_1, \dots, X_n\}$, and there is no need to choose any upper limit ℓ for the sum, in contrast to C_n^a .

The test of Yang et al. (2018) cannot be directly compared with our procedure, since it is developed for a finite sample space and a simple null hypothesis with fully specified probabilities. In order to obtain a method that is applicable to the Zeta distribution with unknown parameter s , we adapt their kernelized Stein discrepancy by plugging in an estimate $\widehat{s}_n$ of s into the discrete Stein operator and working on the finite support $\{1, \dots, K_n\}$, where $K_n = \max_{1 \leq i \leq n} X_i$. Define the forward and backward neighbours

$$\tau(x) = \begin{cases} x+1, & x < K_n, \\ 1, & x = K_n, \end{cases} \quad \rho(x) = \begin{cases} x-1, & x > 1, \\ K_n, & x = 1, \end{cases}$$

and the discrete score

$$s_{\widehat{s}_n}(x) = 1 - \left(\frac{x}{\tau(x)} \right)^{\widehat{s}_n}, \quad x \in \{1, \dots, K_n\}.$$

Let k be a positive definite kernel on $\{1, \dots, K_n\}$. The corresponding Stein kernel is

$$\begin{aligned} \kappa_{\widehat{s}_n}(x, x') &= s_{\widehat{s}_n}(x) k(x, x') s_{\widehat{s}_n}(x') - s_{\widehat{s}_n}(x) (k(x, x') - k(x, \rho(x'))) \\ &\quad - (k(x, x') - k(\rho(x), x')) s_{\widehat{s}_n}(x') \\ &\quad + (k(x, x') - k(\rho(x), x') - k(x, \rho(x')) + k(\rho(x), \rho(x'))), \end{aligned}$$

and the Yang-type kernel Stein discrepancy with estimated parameter is given by a plug-in pairwise KSD statistic of U-statistic form

$$S_{\text{KSD}} = \frac{1}{n(n-1)} \sum_{\substack{i, j=1 \\ j \neq i}}^n \kappa_{\widehat{s}_n}(X_i, X_j).$$

We implement a Gaussian kernel $k(x, x') = \exp(-(x - x')^2/2)$.

The goodness-of-fit test proposed by Meintanis (2009) is based on the Mellin transform and was originally developed for both continuous and discrete Pareto-type laws under a simple null hypothesis. In the discrete case, the null hypothesis is equivalent to (1) and the test statistic involves the difference between the empirical Mellin transform and its theoretical counterpart. For a sample $X_1, \dots, X_n$ and a weight function $w_\beta(t) = \exp(-\beta t)$ with $\beta > 0$, the empirical Mellin transform is

$$M_n(t) = \frac{1}{n} \sum_{j=1}^n X_j^{-t}, \quad t \geq 0.$$

Adapting the discrete-Zeta version of Meintanis (2009) to the composite hypothesis with unknown parameter s , we plug in an estimator $\widehat{s}_n$ of s into the test statistic. Using the integral representation (cf. formulas (1.3) and (3.1) in Meintanis (2009)), the resulting Meintanis statistic is

$$Z_{n,\beta} = n \int_0^\infty \left[\zeta^2(\widehat{s}_n) M_n^2(t) + \zeta^2(\widehat{s}_n + t) - 2 \zeta(\widehat{s}_n) \zeta(\widehat{s}_n + t) M_n(t) \right] e^{-\beta t} dt, \quad (9)$$

where $\zeta(\cdot)$ denotes the Riemann zeta function. In practice, the integral in (9) is evaluated numerically for each bootstrap sample and for each chosen value of the tuning parameter β .

Finally, we adapt the characterization-based approach of Betsch et al. (2022) to the Zeta family. For $s > 2$ the Zeta pmf $p_s(k) = k^{-s}/\zeta(s)$ satisfies the Stein-type identity

$$\mathbb{P}(X = k) = \mathbb{E}\left[\left(1 - \left(\frac{X}{X+1}\right)^s\right)\mathbf{1}_{\{X \geq k\}}\right], \quad k \in \mathbb{N},$$

and deviations from this identity can be used to measure lack of fit. Given a sample $X_1, \dots, X_n$ and an estimator $\hat{s}_n$ of s , we define the empirical pmf $\rho_n(k) = n^{-1} \sum_{j=1}^n \mathbf{1}_{\{X_j=k\}}$ and

$$e_n(k; \hat{s}_n) = \frac{1}{n} \sum_{j=1}^n \left(1 - \left(\frac{X_j}{X_j+1}\right)^{\hat{s}_n}\right) \mathbf{1}_{\{X_j \geq k\}}, \quad k \in \mathbb{N}.$$

The BEN test statistic is the squared ℓ^2 -distance

$$S_{\text{BEN}}(\hat{s}_n) = \sum_{k=1}^{M_n} (e_n(k; \hat{s}_n) - \rho_n(k))^2, \quad M_n = \max_{1 \leq j \leq n} X_j,$$

which vanishes under the model and becomes large when the characterization is violated. Critical values are obtained by a parametric bootstrap from the fitted Zeta distribution $p_{\hat{s}_n}$, exactly in analogy with the procedures proposed by Betsch et al. (2022).

Remark 4 The general characterization in Theorem 3.1 of Betsch et al. (2022) requires a finite first moment (cf. condition (C3)), so its assumptions are rigorously satisfied for the Zeta family only in the parameter range $s > 2$. For $1 < s \leq 2$ the Zeta law has infinite mean, so condition (C3) is violated and the formal applicability of the theorem is no longer guaranteed. Nevertheless, the above identity continues to hold at a heuristic level and the corresponding statistic remains well defined, although a rigorous asymptotic theory in this heavy-tailed regime would require additional arguments beyond those of Betsch et al. (2022).

We consider several alternatives to explore the power of the test, all supported on $\mathbb{N}$. First, for $s > 2$ we write $\text{Geom}(s)$ for the geometric distribution on $\{1, 2, \dots\}$ with success probability p_s chosen such that its mean coincides with that of $\text{Zeta}(s)$, that is $1/p_s = \mathbb{E}[Z] = \zeta(s-1)/\zeta(s)$, $Z \sim \text{Zeta}(s)$. The notation $\text{Zipf}(s, N)$ refers to the truncated Zipf distribution on $\{1, \dots, N\}$ with pmf

$$\mathbb{P}(X = k) = \frac{k^{-s}}{\sum_{j=1}^N j^{-s}}, \quad k = 1, \dots, N.$$

For $s > 1$, $N \in \mathbb{N}$ and $p_g \in (0, 1)$, the Zeta/Geometric splice $\text{ZG}(s, N, p_g)$ coincides with $\text{Zeta}(s)$ on $\{1, \dots, N\}$ and has a shifted geometric tail beyond N , i.e. for $k > N$ the probabilities are proportional to $(1 - p_g)^{k-N-1} p_g$, with a global normalizing constant chosen so that the pmf sums to one. Conversely, for $s > 1$ and $N \in \mathbb{N}$ the Geometric/Zeta splice $\text{GZ}(s, N)$ has a geometric probability mass on $\{1, \dots, N\}$ and a shifted Zeta tail beyond N , so that for $k > N$ the probabilities are proportional to $(k - N)^{-s}$ and the overall pmf is properly normalized. Finally, the zigzag alternatives $\text{Zigzag}(s, \varepsilon)$, with $s > 1$ and $|\varepsilon| < 1$, are defined by the pmf

$$\mathbb{P}(X = k) \propto k^{-s} (1 + \varepsilon(-1)^k), \quad k \in \mathbb{N},$$

which corresponds to alternating up- and down-perturbations of the Zeta(s) pmf.

The results of the simulation study are presented in Table 1. The first result we can see is that no test statistic uniformly dominates the others, which is not surprising, since there cannot be an omnibus best testing procedure as is shown by Janssen (2000). Overall, all considered tests maintain the nominal significance level reasonably well, keeping the empirical size close to 0.05. For benchmark alternatives such as the Poisson, binomial, discrete uniform, and negative binomial distributions, all procedures achieve rejection rates close to 100%, so these cases are not reported in Table 1. In general, most tests exhibit good power when deviations from the Zeta law occur in the left tail, where the probabilities are relatively large. Somewhat surprisingly, the test of Yang et al. (2018) is the strongest procedure against the Geometric alternative when $s = 2.5$, but its power drops essentially to zero when $s = 3$ or $s = 3.5$. A decline in power for the Geometric alternative is visible for all competitors as s increases; in contrast, our statistics $T_{n,\beta}$, where β is the tuning parameter of the weight function as in (5), retain comparatively high power because, for larger s (and thus larger p_s), the probabilities at 1 and 2 are quite substantial, and our test is particularly sensitive to deviations from the Zeta model at small values of X . All tests fail to detect the Geometric right-tail alternatives if the mixture split between the Zeta and Geometric components occurs at $k = 10$. When the split is at $k = 5$, the competing procedures perform visibly better, while our tests still achieve rejection rates only slightly above the nominal level; this is consistent with the fact that our test gives relatively small weight to discrepancies in $p(k) = \mathbb{P}(X = k)$ for large k , and suggests that alternative weightings might be explored in future work. On the other hand, all tests perform very well when the split is reversed, that is, when the Geometric component governs the left part of the distribution and the Zeta component the right tail. The truncated Zeta (Zipf) distribution is well detected if the truncation point is close to zero; however, all of our statistics show a relatively rapid loss of power as the ratio N/s decreases, that is, when the truncation is so close to one that observations with substantial probability mass are no longer visible. Finally, the new statistic $T_{n,\beta}$ clearly outperforms the other procedures for the Zigzag alternatives, which again reflects the particular strength of our test in situations where the relative frequencies of small count values deviate from those implied by the Zeta distribution.

4 Comments and Outlook

We have proposed a new family of test statistics for assessing the composite hypothesis that an integer-valued data set on $\mathbb{N}$ arises from an unspecified member of the zeta family. We derived the asymptotic distribution of the corresponding test and established the consistency of a parametric bootstrap approximation. Moreover, we conducted what appears to be the first comprehensive Monte Carlo study for this testing problem: although (Meintanis 2009) outlines a bootstrap procedure, simulation results are only reported for the simple null hypothesis. Our experiments demonstrate that the proposed tests are strong competitors to existing

methods. In addition, we extended the kernelized Stein discrepancy approach of Yang et al. (2018) to the composite setting and introduced, in the spirit of Betsch et al. (2022), a previously unavailable characterization-based test tailored to the zeta distribution.

We conclude the article by pointing out some open research directions. In Remark 2 we pointed out the connection of the limiting null distribution to a weighted sum of independent χ_1^2 distributed random variables. To use the Rayleigh-Ritz method one needs an orthonormal basis of L_w^2 , which depends on the underlying weight function. In the considered case $w(t) = (1-t)^\beta$, $\beta \geq 0$, a starting point would be the normalized shifted Jacobi polynomials

$$\varphi_{n,\beta}(t) = \sqrt{2n + \beta + 1} \sum_{k=0}^n (-1)^k \binom{n + \beta}{n - k} \binom{n}{k} (1 - t)^k t^{n-k}, \quad t \in [0, 1],$$

which could lead to a precise approximation of the eigenvalues of the covariance operator, see (Ebner et al. 2025) for the methodology and examples of a variety of weighted L^2 spaces, covariance kernels of Gaussian processes, and other families of orthogonal polynomials.

In the simulation study we exclusively employed the maximum likelihood estimator for the shape parameter of the Zeta distribution. Since (Drost et al. 1990) indicate that the choice of estimation procedure can substantially affect the power of goodness-of-fit tests, it would be of interest to explore alternative estimators, such as minimum distance estimators or generalized method-of-moments type procedures inspired by Stein's method as in Betsch et al. (2021); Ebner et al. (2025).

Obviously, the Zeta distribution is the infinite version of the truncated Zipf distribution, so it would be interesting to consider the testing problem of fit to the truncated Zipf family, either for fixed truncation parameter N or unknown N . The corresponding rates of the birth- and death process in analogy to the derivation in the introduction would then stay the same (with $\lambda_N = 0$) but the index k is restricted to $\{1, \dots, N\}$. Another interesting generalization of the Zeta distribution is the so-called Lerch distribution, which can also be expressed as stationary distribution of a birth- and death process, for details see (Klar et al. 2010).

Acknowledgements The authors thank the editors for organizing this special issue in honor of Simos Meintanis. We remember with deep gratitude Simos' exceptional kindness, and the sincere goodwill he extended to everyone. We also thank him for revealing to us the beauty of working with empirical characteristic functions, for his continued support, and for his unwavering encouragement of young scientists. Furthermore, we are very thankful for the helpful comments of two anonymous reviewers, which led to an improvement of the paper. The work of DH was supported by the Czech Science Foundation project GAČR No. 25-15844 S.

Author Contributions The authors contributed equally to this work.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anastasiou A, Barp A, Briol F-X, Ebner B, Gaunt RE, Ghaderinezhad F, Gorham J, Gretton A, Ley C, Liu Q, Mackey L, Oates CJ, Reinert G, Swan Y (2023) Stein's method meets computational statistics: a review of some recent developments. *Stat Sci* 38(1):120–139. <https://doi.org/10.1214/22-STS863>
- Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512. <https://doi.org/10.1126/science.286.5439.509>
- Barbour AD (1988) Stein's method and Poisson process convergence. *J Appl Probab* 25:175–184
- Barbour AD (1990) Stein's method for diffusion approximations. *Probab Theory Relat Fields* 84(3):297–322. <https://doi.org/10.1007/BF01197887>
- Betsch S, Ebner B, Klar B (2021) Minimum L^q -distance estimators for non-normalized parametric models. *Can J Stat* 49(2):514–548. <https://doi.org/10.1002/cjs.11574>
- Betsch S, Ebner B, Nestmann F (2022) Characterizations of non-normalized discrete probability distributions and their application in statistics. *Electron J Stat* 16(1):1303–1329. <https://doi.org/10.1214/22-EJS1983>
- Brueck F, Reimoser V, Baier F (2025) Composite goodness-of-fit test with the Kernel Stein Discrepancy and a bootstrap for degenerate U-statistics with estimated parameters. [arXiv:2510.22792](https://arxiv.org/abs/2510.22792)
- Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. *SIAM Rev* 51(4):661–703. <https://doi.org/10.1137/070710111>
- Drost FC, Kallenberg WCM, Oosterhoff J (1990) The power of edf tests of fit under non-robust estimation of nuisance parameters. *Stat Risk Model* 8(2):167–182. <https://doi.org/10.1524/strm.1990.8.2.167>
- Duan J, Wang W (2014) *Effective dynamics of stochastic partial differential equations*. Academic Press, Boston
- Ebner B, Jiménez-Gamero MD, Milošević B (2025) Efficient eigenvalue approximation in covariance operators via Rayleigh–Ritz with statistical applications. *Stat Pap* 66(6):135. <https://doi.org/10.1007/s00362-025-01751-5>
- Ebner B, Fischer A, Gaunt RE, Picker B, Swan Y (2025) Stein's method of moments. *Scand J Stat* 52(4):1594–1624. <https://doi.org/10.1111/sjos.70003>
- Eichelsbacher P, Reinert G (2008) Stein's method for discrete Gibbs measures. *Ann Appl Probab* 18(4):1588–1618. <https://doi.org/10.1214/07-AAP0498>
- Gabaix X (1999) Zipf's law for cities: an explanation. *Q J Econ* 114(3):739–767. <https://doi.org/10.1162/003353599556133>
- Giacomini R, Politis DN, White H (2013) A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Economet Theor* 29(3):567–589. <https://doi.org/10.1017/S0266466612000655>
- Henze N (2024) *Asymptotic Stochastics. An Introduction with a View Towards Statistics*. Math. Study Resour., vol 10. Springer, Berlin. <https://doi.org/10.1007/978-3-662-68923-3>
- Henze N (1996) Empirical-distribution-function goodness-of-fit tests for discrete models. *Can J Stat* 24(1):81–93
- Janssen A (2000) Global power functions of goodness of fit tests. *Ann Stat* 28(1):239–253
- Johnson NL, Kemp AW, Kotz S (2005) *Univariate discrete distributions*, 3rd edn. Wiley, Hoboken, NJ

- Key O, Gretton A, Briol F-X, Fernandez T (2025) Composite goodness-of-fit tests with kernels. *J Mach Learn Res* 26(51):1–60
- Klar B, Parthasarathy PR, Henze N (2010) Zipf and Lerch limit of birth and death processes. *Probab Eng Inf Sci* 24(1):129–144. <https://doi.org/10.1017/S0269964809990179>
- Lotka AJ (1926) The frequency distribution of scientific productivity. *J Wash Acad Sci* 16:317–323
- Luong A, Doray LG (1996) Goodness of fit test statistics for the zeta family. *Insurance Math Econom* 19(1):45–53. [https://doi.org/10.1016/S0167-6687\(96\)00007-8](https://doi.org/10.1016/S0167-6687(96)00007-8)
- Mandelbrot B (1961) On the theory of word frequencies and on related Markovian models of discourse. In: Jakobson R (ed) *Structure of Language and Its Mathematical Aspects*, vol 12. *Proceedings of Symposia in Applied Mathematics*. American Mathematical Society, Providence, RI, pp 190–219
- Meintanis SG (2009) A unified approach of testing for discrete and continuous Pareto laws. *Stat Pap* 50(3):569–580. <https://doi.org/10.1007/s00362-007-0103-2>
- Newman MEJ (2005) Power laws, Pareto distributions and Zipf's law. *Contemp Phys* 46(5):323–351. <https://doi.org/10.1080/00107510500052444>
- Pareto V (1896) *Cours d'Économie Politique*. Rouge, Lausanne. Two volumes. pp 1896–1897
- Yang J, Liu Q, Rao V, Neville J (2018) Goodness-of-fit testing for discrete distributions via Stein discrepancy. In: Dy J, Krause A (eds) *Proceedings of the 35th International Conference on Machine Learning*. *Proceedings of Machine Learning Research*, vol 80. PMLR, Stockholm, Sweden, pp 5561–5570
- Zipf GK (1949) *Human behavior and the principle of least effort*. Addison-Wesley, Cambridge, MA

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.