



# Stein's identities and the related topics: an instructive explanation on shrinkage, characterization, normal approximation and goodness-of-fit

Tatsuya Kubokawa<sup>1</sup>

Received: 14 May 2023 / Revised: 21 December 2023 / Accepted: 22 December 2023 /

Published online: 31 January 2024

© The Author(s) 2024

## Abstract

The Stein-type identities are widely recognized for their substantial utility and potency in deriving shrinkage estimators improving on crude estimators in normal, gamma, Poisson, and negative binomial distributions. Additionally, these identities serve to characterize these distributions themselves. The Stein identities are also used to demonstrate normal approximation. Moreover, they are instrumental in constructing statistical tests to assess the goodness-of-fit for normality, exponentiality, and Poissonity of distributions. This article offers an instructive and comprehensive explanation of the applications of Stein-type identities in the aforementioned contexts.

**Keywords** Characterization of distributions · Exponentiality · Goodness-of-fit tests · Negative binomial · Normal approximation · Normality · Poissonity · Stein identity · Stein's method

## 1 Introduction

Stein (1973, 1981) introduced the Stein identity, also known as the Stein equation, to derive unbiased estimators for risk functions of shrinkage estimators in the simultaneous estimation of means within normal distributions. This innovative method was employed to enhance the performance of unbiased estimators. The simplicity and potency of this technique have led to significant developments in the field of shrinkage estimation, as extensively documented in the literature. For a comprehensive exploration of this subject, refer to Stein (1981), Strawderman (1971), Shinozaki (1984), Berger (1985), Brandwein and Strawderman (1990), Robert (2007), and Fourdrinier et al. (2018). Komaki (2001) made an intriguing contribution by extending the Stein

---

✉ Tatsuya Kubokawa  
tatsuya@e.u-tokyo.ac.jp

<sup>1</sup> Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

phenomenon to the prediction of predictive distributions. For further insights into this topic, see Ghosh et al. (2020).

It is crucial to note that Stein identities offer utility not only in the context of shrinkage estimation within decision-theoretic frameworks but also in normal approximations, such as the central limit theorem. Their relevance in normal approximation stems from the fact that Stein identities provide a characterization of the normal distribution. Consequently, these identities can be employed in constructing goodness-of-fit test statistics for normality. Hudson (1978) extended Stein-type identities to gamma, Poisson, and negative binomial distributions. Thus, the threefold applications of shrinkage estimation, distribution characterization, and goodness-of-fit testing can be extended to these alternative distributions.

In this paper, we offer an instructive exposition and review of these expanded applications of Stein identities. While many of the results presented herein are well-established in the literature, readers will appreciate the versatile utility of Stein identities in both statistical theory and practical applications.

In Sect. 2, we explain that the normal distribution is equivalent to the Stein identity or the differential equation based on the moment-generating function. Although many characterizations of normal distributions were given in the literature, some of which are summarized there. Two applications of the Stein identity are provided in Sect. 3. Especially, we construct a goodness-of-fit test statistic for normality based on the Stein identity and investigate numerically the performance.

Another important application of the Stein identity is the normal approximation of sum of independent random variables. An instructive explanation is provided in Sect. 4 based on Chen et al. (2011).

In Sect. 5, we describe that the gamma distribution is equivalent to the Stein-type identity or the differential equation of the moment-generating function. Some characterizations of gamma and exponential distributions and shrinkage estimation in decision-theoretic frameworks are summarized. A goodness-of-fit test statistic for exponentiality is constructed based on the Stein-type identity and the performance is investigated numerically.

In Sect. 6, we explain that the Poisson distribution is equivalent to the Stein-type identity or the differential equation of the moment-generating function. Shrinkage estimation and goodness-of-fit test for Poissonity are demonstrated. The Stein-type identity in a negative binomial distribution is briefly described. Some remarks and extensions are given in Sect. 7 as concluding remarks.

## 2 Stein identity and characterization of normal distributions

In normal distributions, Stein (1973, 1981) developed the so-called Stein identity which is not only useful for calculating higher moments, but also powerful for developing shrinkage estimators improving on the minimax estimator in the simultaneous estimation of normal means. The Stein identity also provides the characterization of normal distribution, which means that the central limit theorem can be shown by using the Stein identity. For recent developments and review on the Stein identity, see Bellec and Zhang (2021), Chen (2021), Fathi et al. (2022) and Anastasiou (2023).

**Theorem 2.1** *Let  $X$  be a random variable with mean  $E[X] = \mu$  and variance  $\text{Var}(X) = \sigma^2$ . Then, the following four conditions are equivalent.*

- (a)  $X \sim \mathcal{N}(\mu, \sigma^2)$ .
- (b) For any differentiable function  $h(\cdot)$  with  $E[|(X - \mu)h(X)|] < \infty$  and  $E[|h'(X)|] < \infty$ , it holds that

$$E[(X - \mu)h(X)] = \sigma^2 E[h'(X)], \tag{2.1}$$

where  $h'(x)$  is the derivative of  $h(x)$ .

- (c) For any real constant  $t$  satisfying  $E[|X|e^{tX}] < \infty$ , it holds that

$$E[(X - \mu) \exp\{tX\}] = t\sigma^2 E[\exp\{tX\}]. \tag{2.2}$$

- (d) Let  $g(t) = E[e^{t(X-\mu)}]$ . Then for any  $t$  in the interval  $(-c, c)$  for positive constant  $c$ ,  $g(t)$  satisfies the differential equation

$$\frac{d}{dt} \log g(t) = \frac{g'(t)}{g(t)} = \sigma^2 t \quad \text{or} \quad \frac{d^2}{dt^2} \log g(t) = \frac{g''(t)}{g(t)} - \left(\frac{g'(t)}{g(t)}\right)^2 = \sigma^2, \tag{2.3}$$

where  $g(0) = 1$ ,  $g'(0) = 0$  and  $g''(0) = \sigma^2$ .

**Proof** The proof from (a) to (b) can be done by integration by parts as seen from Stein (1981) and Fourdrinier et al. (2018). We here introduce another approach. Making the transformation  $Y = X - \mu$  gives

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/\sigma^2} dx = \int_{-\infty}^{\infty} h(y + \mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/\sigma^2} dy.$$

Differentiating both sides with respect to  $\mu$  and using Lebesgue’s dominated convergence theorem, we can demonstrate that

$$\int_{-\infty}^{\infty} h(x) \frac{x - \mu}{\sigma^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} dx = \int_{-\infty}^{\infty} h'(y + \mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/(2\sigma^2)} dy,$$

which is rewritten as in (2.1) by turning back with  $X = Y + \mu$ .

Clearly, one gets (c) from (b). Also, it is easy to get (d) from (c). For the proof from (d) to (a), the solution of the differential equation in (2.3) is  $g(t) = \exp\{\sigma^2 t^2/2\}$ , which implies that  $X - \mu \sim \mathcal{N}(0, \sigma^2)$ . □

We briefly provide some conditions for characterizing normality. The study of characterizations of normality has had a long history as explained in Kagan et al. (1973) and Kotz (1974). For a good review of the book of Kagan et al. (1973), see Diaconis et al. (1977). In Theorem 2.2, sufficient condition (b) was given by Cramér (1936), and we provide a simple proof by using the Stein identity. Condition (d) was shown by Kac (1939), Bernstein (1941) and Lukacs (1942).

**Theorem 2.2** Assume that independent random variables  $X_1$  and  $X_2$  are identically distributed with  $E[X_i] = \mu$  and  $\text{Var}(X_i) = \sigma^2$  for  $i = 1, 2$ . Then the following four conditions are equivalent.

- (a)  $X_i \sim \mathcal{N}(\mu, \sigma^2)$  for  $i = 1, 2$ .
- (b)  $X_1 + X_2 \sim \mathcal{N}(2\mu, 2\sigma^2)$ .
- (c) For  $i = 1, 2$ , the density function of  $X_i - \mu$  is symmetric, and  $(X_1 - X_2)^2 / (2\sigma^2) \sim \chi_1^2$ .
- (d)  $X_1 + X_2$  and  $X_1 - X_2$  are independent.

**Proof** Since one gets clearly (b), (c) and (d) from (a), it is sufficient to demonstrate the opposite directions. For the proof from (b) to (a), the condition in (b) and Theorem 2.1(c) implies that

$$E[(X_1 + X_2 - 2\mu) \exp\{t(X_1 + X_2)\}] = 2\sigma^2 t E[\exp\{t(X_1 + X_2)\}]. \quad (2.4)$$

From the independence of  $X_1$  and  $X_2$ , we have

$$\begin{aligned} E[(X_1 - \mu) \exp\{tX_1\}]E[\exp\{tX_2\}] + E[(X_2 - \mu) \exp\{tX_2\}]E[\exp\{tX_1\}] \\ = 2\sigma^2 t E[\exp\{tX_1\}]E[\exp\{tX_2\}]. \end{aligned}$$

Since  $X_1$  and  $X_2$  have the same distribution, we have  $E[(X_1 - \mu) \exp\{tX_1\}]E[\exp\{tX_2\}] = E[(X_2 - \mu) \exp\{tX_2\}]E[\exp\{tX_1\}]$ , so that we can see that for  $i = 1, 2$ ,

$$E[(X_i - \mu) \exp\{tX_i\}] = \sigma^2 t E[\exp\{tX_i\}],$$

which, from Theorem 2.1, shows that  $X_i \sim \mathcal{N}(\mu, \sigma^2)$ .

For the proof from (c) to (b), let  $Y_i = X_i - \mu$  for simplicity. Since  $(Y_1 - Y_2)^2 / (2\sigma^2) \sim \chi_1^2$ , we have  $Y_1 - Y_2 \sim \mathcal{N}(0, 2\sigma^2)$ . From Theorem 2.1(c), it follows that

$$E[(Y_1 - Y_2) \exp\{t(Y_1 - Y_2)\}] = 2\sigma^2 t E[\exp\{t(Y_1 - Y_2)\}]. \quad (2.5)$$

Note that  $Y_1$  and  $Y_2$  are independent and  $-Y_i$  has the same distribution as  $Y_i$ . Thus, equality (2.5) can be rewritten as

$$E[(Y_1 + Y_2) \exp\{t(Y_1 + Y_2)\}] = 2\sigma^2 t E[\exp\{t(Y_1 + Y_2)\}],$$

which is identical to equality (2.4). Hence one gets (b).

Finally, we provide the proof from (d) to (a) along with the proof of Lukacs (1942). The independence of  $X_1 + X_2$  and  $X_1 - X_2$  is equivalent to the independence of  $Y_1 + Y_2$  and  $Y_1 - Y_2$  for  $Y_i = X_i - \mu$ , which implies that

$$E[\exp\{s(Y_1 + Y_2) + t(Y_1 - Y_2)\}] = E[\exp\{s(Y_1 + Y_2)\}]E[\exp\{t(Y_1 - Y_2)\}]. \quad (2.6)$$

Letting  $g(t) = E[\exp\{tY_i\}]$ , we can see that LHS of (2.6) is written as

$$\begin{aligned} & E[\exp\{s(Y_1 + Y_2) + t(Y_1 - Y_2)\}] \\ &= E[\exp\{(s + t)Y_1\}]E[\exp\{(s - t)Y_2\}] = g(s + t)g(s - t). \end{aligned}$$

On the other hand, RHS of (2.6) is written as  $\{g(s)\}^2g(t)g(-t)$ , so that Eq. (2.6) is expressed as

$$g(s + t)g(s - t) = \{g(s)\}^2g(t)g(-t),$$

equivalently rewritten as

$$\log g(s + t) + \log g(s - t) = 2 \log g(s) + \log g(t) + \log g(-t). \tag{2.7}$$

Let  $\psi(t) = (d/dt) \log g(t)$ . Differentiating the both sides in (2.7) with respect to  $s$  and  $t$ , we have

$$\psi(s + t) + \psi(s - t) = 2\psi(s), \tag{2.8}$$

$$\psi(s + t) - \psi(s - t) = \psi(t) - \psi(-t). \tag{2.9}$$

Note that  $\psi(0) = 0$ . Substituting  $s = 0$  in (2.8) gives  $\psi(t) + \psi(-t) = 0$ , or  $\psi(-t) = -\psi(t)$ , which is used to rewrite (2.9) as  $\psi(s + t) - \psi(s - t) = 2\psi(t)$ . Combining this equality and (2.8) gives

$$\psi(s + t) = \psi(s) + \psi(t). \tag{2.10}$$

Equation (2.10) implies that  $\psi(t)$  is written as  $\psi(t) = ct$  for constant  $c$ , namely  $(d/dt) \log g(t) = ct$ . From Theorem 2.1(d), the solution is  $g(t) = \exp\{ct^2/2\}$ . Since  $g''(0) = \sigma^2$ , we have  $c = \sigma^2$ . Thus,  $Y_i = X_i - \mu \sim \mathcal{N}(0, \sigma^2)$ .  $\square$

The normality can be characterized by a random sample. Conditions (b), (e) and (f) in Theorem 2.3 were derived by Kagan et al. (1965), Lukacs (1942) and Ruben (1974), respectively.

**Theorem 2.3** *Let  $X_1, \dots, X_n$  be a random sample from a population with  $E[X_i] = \mu$  and  $\text{Var}(X_i) = \sigma^2$ . Let  $\bar{X} = n^{-1} \sum_{i=1}^n X_i$  and  $S^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ . Then, the following six conditions are equivalent.*

- (a)  $X_i \sim \mathcal{N}(\mu, \sigma^2)$  for  $i = 1, \dots, n$ .
- (b)  $E[\bar{X}|X_1 - \bar{X}, \dots, X_n - \bar{X}] = \mu$  for  $n \geq 3$ .
- (c)  $\bar{X}$  and  $(X_1 - \bar{X}, \dots, X_n - \bar{X})$  are independent.
- (d)  $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ .
- (e)  $\bar{X}$  and  $S^2$  are independent.
- (f) For  $i = 1, \dots, n$ , the density function of  $X_i - \mu$  is symmetric, and  $nS^2/\sigma^2 \sim \chi_{n-1}^2$ .

**Proof** Using well-known properties of a normal distribution, one gets (b), (c), (d), (e) and (f) from (a). For the proof from (b) to (a), let  $Z_i = (X_i - \mu)/\sigma$  for simplicity. Then, (b) is rewritten as  $E[\bar{Z}|Z_1 - \bar{Z}, \dots, Z_n - \bar{Z}] = 0$  for  $\bar{Z} = n^{-1} \sum_{i=1}^n Z_i$ . This implies that

$$E\left[\sum_{i=1}^n Z_i \exp\{t_1(Z_1 - \bar{Z}) + \dots + t_n(Z_n - \bar{Z})\}\right] = 0.$$

Let  $s_i = t_i - \bar{t}$  for  $\bar{t} = n^{-1} \sum_{i=1}^n t_i$ . Then the above equality is expressed as

$$E\left[\sum_{i=1}^n Z_i \exp\left\{\sum_{j=1}^n s_j Z_j\right\}\right] = \sum_{i=1}^n E[Z_i \exp\{s_i Z_i\}] \prod_{j=1, j \neq i}^n E[\exp\{s_j Z_j\}] = 0,$$

equivalently rewritten as

$$\sum_{i=1}^n \frac{E[Z_i \exp\{s_i Z_i\}]}{E[\exp\{s_i Z_i\}]} = 0, \quad \text{or} \quad \sum_{i=1}^n \psi(s_i) = 0$$

for  $\psi(t) = (d/dt) \log E[\exp\{t Z_i\}]$ . Since  $\sum_{i=1}^n s_i = 0$ , we have  $s_n = -\sum_{i=1}^{n-1} s_i$ . Thus,

$$\sum_{i=1}^{n-1} \psi(s_i) = -\psi\left(-\sum_{i=1}^{n-1} s_i\right).$$

Substituting  $s_2 = \dots = s_{n-1} = 0$  gives  $\psi(s_1) = -\psi(-s_1)$ . Hence, the above equality is expressed as

$$\sum_{i=1}^{n-1} \psi(s_i) = \psi\left(\sum_{i=1}^{n-1} s_i\right).$$

Since this equality holds for  $n \geq 3$ , we can see that the solution is  $\psi(t) = ct$ . Since  $\text{Var}(Z_i) = 1$ , we have  $c = 1$ . Thus, from Theorem 2.1(d), it follows that  $Z_i \sim \mathcal{N}(0, 1)$ .

For the proof from (c) to (a), the case of  $n = 2$  follows from Theorem 2.2(d). When  $n \geq 3$ , the independence between  $\bar{X}$  and  $(X_1 - \bar{X}, \dots, X_n - \bar{X})$  implies that

$$E[\bar{X}|X_1 - \bar{X}, \dots, X_n - \bar{X}] = E[\bar{X}] = \mu,$$

which results in (b) and leads to (a).

The proof from (d) to (a) can be done by using the same arguments as in the proof of Theorem 2.2(d).

For the proof from (e) to (a), we provide the proof given by Lukacs (1942). Let  $Y_i = X_i - \mu$  and  $\bar{Y} = n^{-1} \sum_{i=1}^n Y_i$ . Note that  $\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2$

and  $\bar{X} = \bar{Y} + \mu$ . Since  $\sum_{i=1}^n (Y_i - \bar{Y})^2$  and  $\bar{Y}$  are independent, we have

$$E\left[e^{s \sum_{i=1}^n Y_i + t \sum_{i=1}^n (Y_i - \bar{Y})^2}\right] = E\left[e^{s \sum_{i=1}^n Y_i}\right] E\left[e^{t \sum_{i=1}^n (Y_i - \bar{Y})^2}\right]. \tag{2.11}$$

For  $g(s) = E[e^{sY_i}]$ , we can see that  $E[e^{s \sum_{i=1}^n Y_i}] = \{g(s)\}^n$ . Differentiating (2.11) with respect to  $t$  and putting  $t = 0$  gives

$$E\left[\sum_{i=1}^n (Y_i - \bar{Y})^2 e^{s \sum_{i=1}^n Y_i}\right] = \{g(s)\}^n E\left[\sum_{i=1}^n (Y_i - \bar{Y})^2\right] = (n - 1)\sigma^2 \{g(s)\}^n. \tag{2.12}$$

Noting that

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2 = \frac{n-1}{n} \sum_{i=1}^n Y_i^2 - \frac{1}{n} \sum_{i=1}^n \sum_{j=1, j \neq i}^n Y_i Y_j,$$

we can express (2.12) as

$$E\left[\left\{\frac{n-1}{n} \sum_{i=1}^n Y_i^2 - \frac{1}{n} \sum_{i=1}^n \sum_{j=1, j \neq i}^n Y_i Y_j\right\} e^{s \sum_{k=1}^n Y_k}\right] = (n - 1)\sigma^2 \{g(s)\}^n. \tag{2.13}$$

Since  $Y_1, \dots, Y_n$  are independently and identically distributed, the terms in LHS of (2.13) are evaluated as

$$\begin{aligned} E\left[\sum_{i=1}^n Y_i^2 e^{s \sum_{k=1}^n Y_k}\right] &= nE\left[Y_1^2 e^{s \sum_{k=1}^n Y_k}\right] = nE\left[Y_1^2 e^{tY_1}\right] \{g(s)\}^{n-1} \\ &= ng''(s) \{g(s)\}^{n-1}, \\ E\left[\sum_{i=1}^n \sum_{j=1, j \neq i}^n Y_i Y_j e^{s \sum_{k=1}^n Y_k}\right] &= n(n-1)E\left[Y_1 Y_2 e^{s(Y_1+Y_2)}\right] \{g(s)\}^{n-2} \\ &= n(n-1) \{g'(s)\}^2 \{g(s)\}^{n-2}. \end{aligned}$$

Substituting these quantities into (2.13) yields

$$g''(s)g(s) - \{g'(s)\}^2 = \sigma^2 \{g(s)\}^2 \quad \text{or} \quad \frac{d^2}{ds^2} \log g(s) = \sigma^2. \tag{2.14}$$

Thus, from Theorem 2.1(d), it follows that  $X_i - \mu \sim \mathcal{N}(0, \sigma^2)$ , and one gets (a). For the proof of (f) to (a), see Ruben (1974). □

### 3 Applications to shrinkage estimation and goodness-of-fit test

We now provide two applications of the Stein identity to shrinkage estimation and goodness-of-fit tests for normality.

#### 3.1 Shrinkage estimation

The Stein identity is very powerful for deriving unbiased risk estimators of shrinkage estimators. Let  $X_1, \dots, X_p$  be independent random variables such that  $X_i \sim \mathcal{N}(\theta_i, \sigma^2)$ ,  $i = 1, \dots, p$ . Consider the problem of estimating  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$  simultaneously for known  $\sigma^2$ . When estimator  $\hat{\boldsymbol{\theta}}$  is evaluated with the risk function relative to the quadratic loss  $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2/\sigma^2 = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})/\sigma^2$ , Stein (1956) established the inadmissibility of  $\mathbf{X} = (X_1, \dots, X_p)^\top$  in the case of  $p \geq 3$ , and James and Stein (1961) suggested the shrinkage estimator

$$\hat{\boldsymbol{\theta}}^{\text{JS}} = \mathbf{X} - \frac{(p-2)\sigma^2}{\|\mathbf{X}\|^2} \mathbf{X}.$$

The improvement over  $\mathbf{X}$  was proved using a somewhat complicated properties of noncentral chi-squares distribution. Stein (1973) provided a new technique based on the Stein identity for the proof. Because of a simple integration-by-part, the Stein identity enabled us to develop innovated results and great contributions to this research area. The Stein identity was extended to the identity in the chi-square and Wishart distributions and those identities were unified by Konno (2009) which enables us to handle the high-dimensional cases. For some developments and extensions, see Berger (1985), Brandwein and Strawderman (1990), Fourdrinier et al. (2018), Ghosh et al. (2020), Tsukuma and Kubokawa (2020) and Maruyama et al. (2023) and the reference therein.

**Theorem 3.1** *Let  $\mathbf{h}(\mathbf{X}) = (h_1(\mathbf{X}), \dots, h_p(\mathbf{X}))^\top$ , where  $h_i(\mathbf{X})$  is differentiable and satisfies  $\mathbb{E}[|(X_i - \theta_i)h_i(\mathbf{X})|] < \infty$  and  $\mathbb{E}[|(\partial/\partial X_i)h_i(\mathbf{X})|] < \infty$ . Then, the shrinkage estimator  $\hat{\boldsymbol{\theta}}_{\mathbf{h}} = \mathbf{X} - \mathbf{h}(\mathbf{X})$  has the unbiased risk estimator*

$$R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{\mathbf{h}}) = p - 2\nabla^\top \mathbf{h}(\mathbf{X}) + \{\mathbf{h}(\mathbf{X})\}^\top \mathbf{h}(\mathbf{X})/\sigma^2 \quad (3.1)$$

where  $\nabla = (\partial/\partial X_1, \dots, \partial/\partial X_p)^\top$ . Especially,  $\hat{\boldsymbol{\theta}}_\phi = \mathbf{X} - W^{-1}\phi(W)\mathbf{X}$  for  $W = \|\mathbf{X}\|^2/\sigma^2$  has the risk unbiased estimator

$$R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_\phi) = p + \frac{\{\phi(W) - (p-2)\}^2 - (p-2)^2}{W} - 4\phi'(W). \quad (3.2)$$

**Proof** The risk function of  $\hat{\boldsymbol{\theta}}_{\mathbf{h}}$  is  $R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{\mathbf{h}}) = p - 2\mathbb{E}[(\mathbf{X} - \boldsymbol{\theta})^\top \mathbf{h}(\mathbf{X})]/\sigma^2 + \mathbb{E}\{\{\mathbf{h}(\mathbf{X})\}^\top \mathbf{h}(\mathbf{X})\}/\sigma^2$ , and the Stein identity gives

$$\mathbb{E}[(\mathbf{X} - \boldsymbol{\theta})^\top \mathbf{h}(\mathbf{X})] = \sum_{i=1}^p \mathbb{E}[(X_i - \theta_i)h_i(\mathbf{X})] = \sum_{i=1}^p \mathbb{E}\left[\sigma^2 \frac{\partial}{\partial X_i} h_i(\mathbf{X})\right] = \sigma^2 \mathbb{E}[\nabla^\top \mathbf{h}(\mathbf{X})].$$



This provides the unbiased estimator of the risk function given in (3.1). (3.2) can be derived from (3.1). □

From (3.1) or (3.2), we can derive conditions on  $h(\cdot)$  or  $\phi(\cdot)$  for improvement over  $X$ . For example, Baranchik’s (1970) condition is (a)  $\phi(w)$  is nondecreasing and (b)  $0 \leq \phi(w) \leq 2(p - 2)$ .

The James–Stein estimator  $\widehat{\theta}^{JS}$  corresponds to the case of  $\phi(w) = p - 2$ , and the risk unbiased estimator suggests the equation

$$E[\|X - \theta\|^2] = E[\|\widehat{\theta}^{JS} - \theta\|^2] + E[\|X - \widehat{\theta}^{JS}\|^2],$$

which is interpreted as the Pythagorean triangle among  $X$ ,  $\theta$  and  $\widehat{\theta}^{JS}$ . Kubokawa (1994) constructed a class of estimators improving on the James–Stein estimator. See also Kubokawa (1991).

**Theorem 3.2** *The estimator  $\widehat{\theta}_\phi = X - W^{-1}\phi(W)X$  improves on the James–Stein estimator if (a)  $\phi(w)$  is nondecreasing, and (b)  $\lim_{w \rightarrow \infty} \phi(w) = p - 2$  and*

$$\phi(w) \geq \phi_0(w) \equiv \frac{\int_0^w y^{p/2-1} e^{-y/2} dy}{\int_0^w y^{p/2-2} e^{-y/2} dy}. \tag{3.3}$$

**Proof** The risk difference is  $\Delta = R(\theta, \widehat{\theta}^{JS}) - R(\theta, \widehat{\theta}_\phi) = -E[\{\phi(W) - (p - 2)\}^2/W] + 4E[\phi'(W)]$ , and from condition (b), it is noted that

$$\begin{aligned} -\{\phi(W) - (p - 2)\}^2 &= \left[ \{\phi(tW) - (p - 2)\}^2 \right]_{t=1}^\infty = \int_1^\infty \left\{ \frac{d}{dt} \{\phi(tW) - (p - 2)\}^2 \right\} dt \\ &= 2W \int_1^\infty \{\phi(tW) - (p - 2)\} \phi'(tW) dt, \end{aligned}$$

so that, after making the transformation, the first term is written as

$$\begin{aligned} &-E\left[\{\phi(W) - (p - 2)\}^2/W\right] \\ &= 2 \int_0^\infty \int_1^\infty \{\phi(tw) - (p - 2)\} \phi'(tw) dt f_p(w, \lambda) dw, \end{aligned} \tag{3.4}$$

where  $f_p(w, \lambda)$  denotes the density function of noncentral chi-square distribution with  $p$  degrees of freedom and noncentrality  $\lambda = \|\theta\|^2/\sigma^2$ . Thus,

$$\Delta = 2 \int_0^\infty \phi'(w) \left[ \{\phi(w) - (p - 2)\} \int_0^w \frac{1}{y} f_p(y, \lambda) dy + 2f_p(w, \lambda) \right] dw.$$

Since  $\phi'(w) \geq 0$  from condition (a), we have  $\Delta \geq 0$  if  $\phi(w)$  satisfies  $\phi(w) \geq \phi_\lambda(w)$ , where

$$\phi_\lambda(w) = p - 2 - \frac{2f_p(w, \lambda)}{\int_0^w y^{-1} f_p(y, \lambda) dy}.$$

We here show that  $\phi_0(w) \geq \phi_\lambda(w)$ , which is written as

$$\frac{2f_p(w, \lambda)}{\int_0^w y^{-1} f_p(y, \lambda) dy} \geq \frac{2f_p(w, 0)}{\int_0^w y^{-1} f_p(y, 0) dy},$$

or

$$\int_0^w \frac{f_p(y, 0)f_p(w, 0)}{y} \left\{ \frac{f_p(w, \lambda)}{f_p(w, 0)} - \frac{f_p(y, \lambda)}{f_p(y, 0)} \right\} dy \geq 0.$$

Since the noncentral chi-squared distribution can be expressed as a mixture of Poisson and central chi-squared distributions, it is noted that

$$f_p(y, \lambda) = \sum_{k=0}^{\infty} P_\lambda(k) \frac{1}{\Gamma(p/2 + k) 2^{p/2+k}} y^{p/2+k-1} e^{-y/2}$$

for  $P_\lambda(k) = (\lambda/2)^k e^{-\lambda/2} / k!$ . Hence,

$$\frac{f_p(y, \lambda)}{f_p(y, 0)} = \sum_{k=0}^{\infty} P_\lambda(k) \frac{\Gamma(p/2)}{\Gamma(p/2 + k)} \frac{1}{2^k} y^k$$

is increasing in  $y$ , so that for  $w > y$ ,

$$\frac{f_p(w, \lambda)}{f_p(w, 0)} - \frac{f_p(y, \lambda)}{f_p(y, 0)} \geq 0,$$

which implies that  $\phi_0(w) \geq \phi_\lambda(w)$ . Using integration by parts, we can see that

$$\phi_0(w) = p - 2 - \frac{2w^{p/2-1} e^{-w/2}}{\int_0^w y^{p/2-2} e^{-y/2} dy} = \frac{\int_0^w y^{p/2-1} e^{-y/2} dy}{\int_0^w y^{p/2-2} e^{-y/2} dy},$$

which is given in (3.3). Hence, it is proved that  $\Delta \geq 0$  under condition  $\phi(w) \geq \phi_0$ .  $\square$

It is interesting to note that the estimator  $\widehat{\theta}^{GB} = \widehat{\theta}_{\phi_0}$  with  $\phi_0(W)$  is the generalized Bayes estimator against the prior distribution  $\pi(\theta) = \|\theta\|^{2-p}$ . Since  $\phi_0(w)$  satisfies the above conditions (a), the generalized Bayes estimator  $\widehat{\theta}_{\phi_0}$  improves on the James–Stein estimator. It is also interesting to note that the prior distribution  $\pi(\theta) = \|\theta\|^{2-p}$  is a harmonic function, namely  $\nabla^\top \nabla \pi(\theta) = \sum_{i=1}^p (\partial^2 / \partial \theta_i^2) \|\theta\|^{2-p} = 0$ . The positive-part Stein estimator  $\widehat{\theta}^{S+} = \max\{1 - (p - 2)/W, 0\} X$  also satisfies the conditions (a) and (b).

The risk performances of the shrinkage estimators  $\widehat{\theta}^{JS}$ ,  $\widehat{\theta}^{S+}$  and  $\widehat{\theta}^{GB}$ , denoted by JS, PS and GB, respectively, are investigated by simulation with 10,000 replications and the average values of the risks are reported in Table 1 for  $p = 6$ ,  $\sigma^2 = 1$  and  $\theta = (k/3)I$ ,  $k = 0, \dots, 9$ . As distributions of  $X_i$ 's, we treat normal, double exponential and t-distributions with 5 degrees of freedom. From Table 1, it is seen

**Table 1** Risks of three shrinkage estimators for  $p = 6$  and  $\theta = (k/3)I$

$k$	Normal distribution $R(\theta, X) = 6$									
	0	1	2	3	4	5	6	7	8	9
JS	2.03	2.39	3.31	4.15	4.75	5.13	5.36	5.52	5.62	5.70
PS	1.35	1.84	2.94	4.01	4.72	5.12	5.36	5.52	5.62	5.70
GB	2.00	2.28	2.98	3.84	4.57	5.06	5.35	5.52	5.62	5.70
$k$	$t_5$ distribution $R(\theta, X) = 10$									
	0	1	2	3	4	5	6	7	8	9
JS	4.92	5.28	6.10	6.99	7.70	8.23	8.61	8.88	9.07	9.21
PS	4.47	4.87	5.84	6.88	7.68	8.23	8.61	8.88	9.07	9.21
GB	4.98	5.24	5.92	6.78	7.58	8.18	8.60	8.88	9.07	9.21
$k$	Double exponential distribution $R(\theta, X) = 12.0$									
	0	1	2	3	4	5	6	7	8	9
JS	6.97	7.22	7.90	8.71	9.44	10.02	10.45	10.78	11.01	11.19
PS	6.40	6.77	7.66	8.63	9.42	10.01	10.45	10.78	11.01	11.19
GB	6.83	7.08	7.72	8.55	9.34	9.98	10.44	10.77	11.01	11.19

that the minimality of the three shrinkage estimators is robust for the t- and double exponential distributions.

### 3.2 Goodness-of-fit tests for normality

Goodness-of-fit tests for normality have been studied in a lot of articles. For references and explanations including omnibus test procedures, see Madansky (1988) and Thode (2002). An idea of using the Stein identity for testing normality is interesting and reasonable, because the Stein identity characterizes normal distributions. Henze and Visagie (2020) and Betsch and Ebner (2020) constructed test statistics based on the Stein identity.

Let  $X_1, \dots, X_n$  be a random sample from a population with distribution function  $F(\cdot)$ , where the mean and variance are denoted by  $E[X_i] = \mu$  and  $\text{Var}(X_i) = \sigma^2$ . The problem is to test the normality of the underlying distribution under the null hypothesis  $H_0 : F = \mathcal{N}(\mu, \sigma^2)$ . From Theorem 2.1, the characterization of a normal distribution of random variable  $X$  is

$$E\left[\left(\frac{X - \mu}{\sigma} - t\right)e^{t(X - \mu)/\sigma}\right] = 0,$$

and the sample counterpart of the LHS is expressed by  $w_t/\sqrt{n}$ , where

$$w_t = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - t)e^{tY_i},$$

for  $Y_i = (X_i - \bar{X})/S$  and  $S^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ . It is noted that  $w_t$  is invariant under the transformation of location and scale. Then, the normality can be tested based on  $ST_t = |w_t|$ . Since it depends on  $t$ , however, it is better to take a weighted  $L^2$  distance and integrate over  $t$ . Henze and Visagie (2020) considered the test statistic  $\int_{-\infty}^{\infty} w_t^2 K(t) dt$  for a weight function  $K(t)$  and suggested the use of  $K(t) = e^{-\gamma t^2}$  for positive  $\gamma$ . The resulting test statistic is

$$HV_{\gamma} = \int_{-\infty}^{\infty} w_t^2 e^{-\gamma t^2} dt = \frac{\sqrt{\pi}}{n\sqrt{\gamma}} \sum_{i=1}^n \sum_{j=1}^n \left( Y_i Y_j - \frac{A_{ij}^2}{2\gamma} + \frac{1}{2\gamma} + \frac{A_{ij}^2}{4\gamma^2} \right) e^{A_{ij}/(4\gamma)},$$

where  $A_{ij} = Y_i + Y_j$ . Taking  $K(t) = 1$  for  $-c < t < c$  and otherwise  $K(t) = 0$  with positive constant  $c$ , one gets another test statistic

$$\begin{aligned} IST_c &= \int_{-c}^c w_t^2 dt \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left\{ \frac{Y_i Y_j + c^2}{A_{ij}} (e^{cA_{ij}} - e^{-cA_{ij}}) \right. \\ &\quad \left. - \left( 1 + \frac{2}{A_{ij}} \right) \left( (c - A_{ij}^{-1}) e^{cA_{ij}} + (c + A_{ij}^{-1}) e^{-cA_{ij}} \right) \right\}. \end{aligned}$$

Based on  $w_t$ , we also suggest the test statistic

$$MST_c = \sup_{-c < t < c} |w_t|$$

for positive constant  $c$ .

The following lemma is helpful for investigating asymptotic properties of these test statistics.

**Lemma 3.1** *Let  $Z_i = (X_i - \mu)/\sigma$  for  $i = 1, \dots, n$ . For  $g(t) = E[e^{tZ_1}]$ , let  $h_0(t) = g'(t) - tg(t)$ ,  $h_1(t) = tg'(t) + (1 - t^2)g(t)$  and  $h_2(t) = tg''(t) + (1 - t^2)g'(t)$ . Assume that  $E[Z_1^2 e^{tZ_1}] < \infty$  for  $t$  around zero. Then,  $w_t$  is approximated as  $w_t = W_n(t) + \sqrt{n}h_0(t) + o_p(1)$ , where*

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ (Z_i - t)e^{tZ_i} - h_0(t) - Z_i h_1(t) - \frac{1}{2}(Z_i^2 - 1)h_2(t) \right\}. \tag{3.5}$$

**Proof** Let  $\bar{Z} = n^{-1} \sum_{i=1}^n Z_i$ . Note that  $\{1 + (S/\sigma - 1)\}^{-1} = 1 - (S/\sigma - 1) + O_p(n^{-1})$  and  $S/\sigma - 1 = \sqrt{S^2/\sigma^2} - 1 = 2^{-1}(S^2/\sigma^2 - 1) + O_p(n^{-1})$ . Then,  $(X_i - \bar{X})/S$  is approximated as

$$\frac{X_i - \bar{X}}{S} = \frac{Z_i - \bar{Z}}{1 + (S/\sigma - 1)} = (Z_i - \bar{Z}) \left\{ 1 - \left( \frac{S}{\sigma} - 1 \right) + O_p(n^{-1}) \right\}$$

$$\begin{aligned}
 &= (Z_i - \bar{Z}) \left\{ 1 - \frac{1}{2} \left( \frac{S^2}{\sigma^2} - 1 \right) \right\} + O_p(n^{-1}) \\
 &= (Z_i - \bar{Z}) \left\{ 1 - \frac{1}{2n} \sum_{j=1}^n (Z_j^2 - 1) \right\} + O_p(n^{-1}).
 \end{aligned}$$

Using this approximation, we evaluate  $w_t$  as

$$\begin{aligned}
 w_t &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S} - t \right) e^{t(X_i - \bar{X})/S} \\
 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ Z_i - t - \bar{Z} - \frac{Z_i - \bar{Z}}{2n} \sum_{j=1}^n (Z_j^2 - 1) + O_p(n^{-1}) \right\} \\
 &\quad \times \exp \left[ t Z_i - t \bar{Z} - t \frac{Z_i - \bar{Z}}{2n} \sum_{j=1}^n (Z_j^2 - 1) + O_p(n^{-1}) \right].
 \end{aligned}$$

Since  $e^x = 1 + x + O(x^2)$  and  $\bar{Z} = O_p(1/\sqrt{n})$ ,  $w_t$  is approximated as

$$\begin{aligned}
 w_t &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ Z_i - t - \bar{Z} - \frac{Z_i}{2n} \sum_{j=1}^n (Z_j^2 - 1) + O_p(n^{-1}) \right\} \\
 &\quad \times \left\{ 1 - t \bar{Z} - t \frac{Z_i}{2n} \sum_{j=1}^n (Z_j^2 - 1) + O_p(n^{-1}) \right\} e^{t Z_i} \\
 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ Z_i - t - \bar{Z} - \frac{Z_i}{2n} \sum_{j=1}^n (Z_j^2 - 1) \right. \\
 &\quad \left. - t(Z_i - t) \left\{ \bar{Z} + \frac{Z_i}{2n} \sum_{j=1}^n (Z_j^2 - 1) \right\} \right] e^{t Z_i} + o_p(1),
 \end{aligned}$$

which can be rewritten as

$$\begin{aligned}
 w_t &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (Z_i - t) e^{t Z_i} - \bar{Z} \frac{1}{\sqrt{n}} \sum_{i=1}^n e^{t Z_i} - \frac{1}{2n} \sum_{j=1}^n (Z_j^2 - 1) \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i e^{t Z_i} \\
 &\quad - t \bar{Z} \frac{1}{\sqrt{n}} \sum_{i=1}^n (Z_i - t) e^{t Z_i} - \frac{t}{2n} \sum_{j=1}^n (Z_j^2 - 1) \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i (Z_i - t) e^{t Z_i} + o_p(1).
 \end{aligned}$$

Each term can be evaluated as

$$n^{-1/2} \sum_{i=1}^n (Z_i - t) e^{t Z_i} = \sqrt{n} \{ n^{-1} \sum_{i=1}^n (Z_i - t) e^{t Z_i} - h_0(t) \} + \sqrt{n} h_0,$$

$$\begin{aligned} \bar{Z}n^{-1/2} \sum_{i=1}^n e^{tZ_i} &= \sqrt{n}\bar{Z}g(t) + o_p(1), \\ n^{-1} \sum_{j=1}^n (Z_j^2 - 1)n^{-1/2} \sum_{i=1}^n Z_i e^{tZ_i} &= n^{-1/2} \sum_{j=1}^n (Z_j^2 - 1)g'(t) + o_p(1), \\ \bar{Z}n^{-1/2} \sum_{i=1}^n (Z_i - t)e^{tZ_i} &= \sqrt{n}\bar{Z}\{g'(t) - tg(t)\} + o_p(1), \\ n^{-1} \sum_{j=1}^n (Z_j^2 - 1)n^{-1/2} \sum_{i=1}^n Z_i(Z_i - t)e^{tZ_i} &= n^{-1/2} \sum_{j=1}^n (Z_j^2 - 1)\{g''(t) - tg'(t)\} + o_p(1). \end{aligned}$$

Thus, it can be verified that  $w_t$  is approximated as  $w_t = W_n + \sqrt{nh_0(t)} + o_p(1)$ , which proves Lemma 3.1 □

From (3.5), the central limit theorem shows that  $W_n(t)$  is asymptotically distributed as the normal distribution with mean zero and variance  $\text{Var}(W_n(t))$  under the assumption of  $E[Z_1^4] < \infty$ , where the variance can be evaluated as

$$\begin{aligned} \text{Var}(W_n(t)) &= E[(Z_1 - t)^2 e^{2tZ_1}] - h_0(t)^2 + h_1(t)^2 \\ &\quad + \frac{1}{4}h_2(t)^2(E[Z_1^4] - 1) - 2h_1(t)E[Z_1(Z_1 - t)e^{tZ_1}] \\ &\quad - h_2(t)\{E[Z_1^2(Z_1 - t)e^{tZ_1}] - h_0(t)\} + h_1(t)h_2(t)E[Z_1^3] + o(1). \end{aligned}$$

Note that  $h_1(t) = g(t) + th_0(t)$  and  $h_2(t) = 2tg(t) + h_0(t) + th_0'(t)$ . Under the normality hypothesis  $H_0$ , we have  $h_0(t) = 0$ ,  $h_1(t) = g(t)$ ,  $h_2(t) = 2tg(t)$  and  $E[(Z_1 - t)^2 e^{2tZ_1}] = (1 + t^2)e^{2t^2}$ . Thus, the asymptotic variance of  $w_t$  under the normality is  $\text{Var}(W_n(t)) = V(t)^2 + o(1)$ , where

$$V(t)^2 = (1 + t^2)e^{2t^2} - (1 + 2t^2)e^{t^2}.$$

Henze and Visagie (2020) showed the consistency of  $HV_\gamma$ , namely  $P_F(HV_\gamma > d) \rightarrow 1$  as  $n \rightarrow \infty$  for positive constant  $d$  and non-normal distributions  $F$ . Using Lemma 3.1, we can verify the consistency of  $IST_c$  and  $MST_c$ .

**Theorem 3.3** *Assume that  $E[Z_1^4] < \infty$ . Then, the test statistics  $IST_c$  and  $MST_c$  are consistent.*

**Proof** Concerning the consistency of  $MST_c$ , it can be observed that for all  $t$  in the interval  $(-c, c)$ ,

$$\begin{aligned} P_F(\sup_{-c < t < c} |w_t| > d) &\geq P_F(|w_t| > d) = P_F(w_t > -d) + P_F(w_t > d) \\ &= P_F\{W_n(t) < -\sqrt{nh_0(t)} - d + o_p(1)\} \\ &\quad + P_F\{W_n(t) > -\sqrt{nh_0(t)} + d + o_p(1)\}. \end{aligned}$$

Note that  $W_n(t)$  converges in distribution to the normal distribution. When  $F$  is not a normal distribution, from Theorem 2.1, there is some  $t_0$  in  $(-c, c)$  such that  $h_0(t_0) \neq 0$ . Hence,  $P_F\{W_n(t_0) < -\sqrt{n}h_0(t_0) - d + o_p(1)\} \rightarrow 1$  when  $h_0(t_0) < 0$ , and  $P_F\{W_n(t_0) > -\sqrt{n}h_0(t_0) + d + o_p(1)\} \rightarrow 1$  when  $h_0(t_0) > 0$ . This shows the consistency of  $MST_c$ .

For  $IST_c$ , it is observed that for large  $n$ ,

$$\begin{aligned} P_F\left(\int_{-c}^c w_t^2 dt > d\right) &= P_F\left(\int_{-c}^c W_n(t)^2 dt \right. \\ &\quad \left. + 2\sqrt{n} \int_{-c}^c W_n(t)h_0(t)dt + n \int_{-c}^c h_0(t)^2 dt > d\right) + o(1) \\ &\geq P_F\left(-2\sqrt{n} \left| \int_{-c}^c W_n(t)h_0(t)dt \right| + n \int_{-c}^c h_0(t)^2 dt > d\right) + o(1) \\ &= P_F\left(\left| \int_{-c}^c W_n(t)h_0(t)dt \right| < \frac{\sqrt{n}}{2} \left( \int_{-c}^c h_0(t)^2 dt - d/n \right)\right) + o(1) \\ &\geq 1 - \frac{E\left[\left\{ \int_{-c}^c W_n(t)h_0(t)dt \right\}^2\right]}{(n/2)\left(\int_{-c}^c h_0(t)^2 dt - d/n\right)^2} + o(1). \end{aligned}$$

From (3.5), we have  $\int_{-c}^c W_n(t)h_0(t)dt = n^{-1/2} \sum_{i=1}^n Y_i^*$  for

$$\begin{aligned} Y_i^* &= \int_{-c}^c \{(Z_i - t)e^{tZ_i} - h_0(t)\}h_0(t)dt - Z_i \int_{-c}^c h_1(t)h_0(t)dt - \frac{1}{2}(Z_i^2 - 1) \\ &\quad \int_{-c}^c h_2(t)h_0(t)dt. \end{aligned}$$

Since  $Y_1^*, \dots, Y_n^*$  are independently and identically distributed with zero mean and a finite variance, it can be seen that  $E\left[\left\{ \int_{-c}^c W_n(t)h_0(t)dt \right\}^2\right]$  converges to a positive constant. Thus, it is concluded that  $P_F\left(\int_{-c}^c w_t^2 dt > d\right) \rightarrow 1$ . □

We investigate the performances of powers of the test statistics  $HV_\gamma$  with  $\gamma = 3$  given in Henze and Visagie (2020),  $IST_c$  with  $c = 1$  and  $MST_c$  with  $c = 1$ . We also treat the test statistic  $ST_t = w_t/V(t)$  with  $t = 0.5$ , which converges in distribution to  $\mathcal{N}(0, 1)$  under  $H_0$ . From the proof of Theorem 3.3, this test can be seen to be consistent in the sense that  $P_F(ST_{t_0} > d) \rightarrow 1$  for distributions with  $h_0(t_0) \neq 0$  for  $t_0 = 0.5$ . As another competitor, we add the test statistic suggested by De Wet and Ventner (1972), who modified the Shapiro–Francia (1972) and the Shapiro–Wilk (1965) test statistics, as

$$DWV = \frac{\sum_{i=1}^n a_i X_{(i)}}{\sqrt{\sum_{i=1}^n a_i^2 \sqrt{n} S^2}}, \quad a_i = \Phi^{-1}\left(\frac{i}{n+1}\right),$$

where  $X_{(1)} \leq \dots \leq X_{(n)}$  are order statistics of  $X_i$ 's. The idea of this test is simple, but powerful. When data are distributed as a normal distribution, the points  $(x_{(i)}, a_i)$

**Table 2** Power of the six tests for  $n = 50$  and  $w = 0.2, 0.5, 0.8, 1.0$ 

$w$	$M_1 : Ex(1)$					$M_2 : DE(0, 2)$				$M_3 : \text{mixed}$			
	0.0	0.2	0.5	0.8	1.0	0.2	0.5	0.8	1.0	0.2	0.5	0.8	1.0
$ST_{0.5}$	5.0	6.0	43.0	93.7	97.0	9.2	31.9	36.6	36.7	5.4	11.4	33.4	37.4
$HV_3$	5.0	6.0	34.6	87.2	92.6	10.9	41.6	48.8	49.3	5.6	11.6	34.5	39.0
$IST_1$	5.0	5.9	35.4	90.1	95.1	10.7	40.7	47.9	48.2	5.5	11.1	33.7	38.4
$MST_1$	5.0	6.0	33.0	81.7	87.1	10.6	42.2	49.8	50.3	5.7	11.3	34.3	39.0
$DWV$	5.0	5.5	35.1	95.8	99.7	10.6	49.1	62.8	63.8	5.5	11.2	40.3	48.2

are plotted on or near the line, so that DWV is close to one. Thus,  $H_0$  is rejected when DWV is smaller than a critical value.

The powers of the five test statistics  $ST_{0.5}$ ,  $HV_3$ ,  $IST_1$ ,  $MST_1$  and  $DWV$  are investigated by simulation, where their critical values are adjusted so that their type I errors are  $\alpha = 5\%$ . We consider the three alternative distributions

$$M_1 : X_i = (1 - w)\mathcal{N}(0, 1) + wEx(1),$$

$$M_2 : X_i = (1 - w)\mathcal{N}(0, 1) + wDE(0, 2),$$

$$M_3 : X_i = (1 - w)\mathcal{N}(0, 1) + wEx(1)/2 + wDE(0, 0.5)/2,$$

where  $\mathcal{N}(0, 1)$ ,  $Ex(1)$  and  $DE(0, \sigma)$  denote random variables having standard normal distribution  $\mathcal{N}(0, 1)$ , exponential distribution  $Ex(1)$  and double exponential distribution  $DE(0, \sigma)$  with scale parameter  $\sigma$ , respectively.

The values of the powers for  $w = 0.2, 0.5, 0.8, 1.0$  and  $n = 50$  are obtained based on simulation data with 10,000 replications by using the Ox by Doornik (2007) and reported in Table 2. The model  $M_1$  has skewness, and the tests  $ST_{0.5}$  and  $DWV$  are more powerful, but  $MST_1$  is less powerful. The model  $M_2$  has kurtosis, and the test  $DWV$  is more powerful, but  $ST_{0.5}$  is less powerful. In the model  $M_3$  with mixed distributions as an alternative, the test  $DWV$  is good, and the other four tests are similarly performed. Through Table 2, it is seen that the performances of  $HV_3$  and  $IST_1$  are not bad, but such tests based on moments are less powerful than  $DWV$  based on quantiles.

#### 4 Stein's methods for normal approximations

An important application of the Stein identity is the normal approximation. This approach is called Stein's method, and it has been studied in the literature including Ho and Chen (1978), Stein (1986), Goldstein and Reinert (1997), Shorack (2000), Barbour and Chen (2005), Diaconis and Holmes (2004), Chen and Shao (2005), Chen et al. (2011), Chen et al. (2013), Lehmann and Romano (2022) and the reference therein. Of these, Chen et al. (2011) gives us a good explanation for the Stein method. For instructive purposes, we here provide a simple introduction based on Chen et al. (2011).



### 4.1 A basic concept and a simple method

Let  $X_1, \dots, X_n$  be independent random variables with  $E[X_i] = 0$  and  $\text{Var}(X_i) = 1$ . Let

$$X = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i, \quad \xi_i = \frac{X_i}{\sqrt{n}}, \quad X^{(i)} = X - \xi_i \text{ and } \mu_k = \frac{1}{n} \sum_{i=1}^n E[|X_i|^k] \text{ for } k = 1, 2, 3.$$

Note that  $X^{(i)}$  is independent of  $\xi_i$ . Let  $Y$  be a random variable having  $\mathcal{N}(0, 1)$ . Then, we want to show that for any real  $z$ ,

$$|P(X \leq z) - \Phi(z)| = |E[I(X \leq z)] - E[Y \leq z]| \rightarrow 0$$

as  $n \rightarrow \infty$ .

For any nonnegative function  $h(\cdot)$  satisfying  $E[h(X)] < \infty$ , in general, the solution of the equation

$$h(x) - E[h(Y)] = f'_h(x) - xf_h(x) \tag{4.1}$$

is given by

$$f_h(x) = \int_{-\infty}^x [h(y) - \mu_h] \phi(y) dy / \phi(x) = - \int_x^{\infty} [h(y) - \mu_h] \phi(y) dy / \phi(x), \tag{4.2}$$

where  $\mu_h = E[h(Y)]$  and  $\phi(x)$  is the probability density function of  $\mathcal{N}(0, 1)$ . In particular, for  $h_z(x) = I(x \leq z) - \Phi(z)$ , the solution of the equation  $I(x \leq z) - \Phi(z) = f'_z(x) - xf_z(x)$  is written as

$$\begin{aligned} f_z(x) &= \int_{-\infty}^x [I(y \leq z) - \Phi(z)] \phi(y) dy / \phi(x) \\ &= \begin{cases} \Phi(x) \bar{\Phi}(z) / \phi(x), & x \leq z, \\ \Phi(z) \bar{\Phi}(x) / \phi(x), & x > z, \end{cases} \end{aligned} \tag{4.3}$$

where  $\bar{\Phi}(x) = 1 - \Phi(x)$ . Then, we get the equality

$$E[I(X \leq z) - \Phi(z)] = E[f'_z(X) - Xf_z(X)]. \tag{4.4}$$

For the central limit theorem (CLT), it is sufficient to show that  $\lim_{n \rightarrow \infty} E[f'_z(X) - Xf_z(X)] = 0$ . It is noted that the RHS of (4.4) is exactly zero from the Stein identity if  $X \sim \mathcal{N}(0, 1)$ . To this end, we prepare the following lemma.

**Lemma 4.1** *For any nonnegative function  $h(\cdot)$  satisfying  $E[h(X)] < \infty$ , it holds that*

$$E[f'_h(X) - Xf_h(X)] = \sum_{i=1}^n E \left[ \xi_i^2 \int_0^1 \{f'_h(X^{(i)}) - f'_h(X^{(i)} + u\xi_i)\} du \right]$$

$$+ \frac{1}{n} \sum_{i=1}^n E[f'_h(X^{(i)} + \xi_i) - f'_h(X^{(i)})]. \tag{4.5}$$

**Proof** The proof is from Chen et al. (2011). From the Taylor series expansion with integral remainder, it follows that

$$\begin{aligned} f_h(X^{(i)} + \xi_i) &= f_h(X^{(i)}) + \int_{X^{(i)}}^{X^{(i)}+\xi_i} f'_h(t) dt \\ &= f_h(X^{(i)}) + \xi_i \int_0^1 f'_h(X^{(i)} + u\xi_i) du. \end{aligned} \tag{4.6}$$

We first write  $E[Xf_h(X)] = \sum_{i=1}^n E[\xi_i f_h(X^{(i)} + \xi_i)]$ . From (4.6),  $E[\xi_i] = 0$  and independence of  $X^{(i)}$  and  $\xi_i$ , we observe that

$$\begin{aligned} E[\xi_i f_h(X^{(i)} + \xi_i)] &= E\left[\xi_i f_h(X^{(i)}) + \xi_i^2 \int_0^1 f'_h(X^{(i)} + u\xi_i) du\right] \\ &= E\left[\xi_i^2 \int_0^1 f'_h(X^{(i)} + u\xi_i) du\right]. \end{aligned} \tag{4.7}$$

Since  $E[\xi_i^2 f'_h(X^{(i)})] = E[\xi_i^2]E[f'_h(X^{(i)})] = n^{-1}E[f'_h(X^{(i)})]$ , on the other hand, it can be seen that

$$\begin{aligned} E[f'_h(X)] &= \sum_{i=1}^n E[n^{-1} f'_h(X^{(i)})] + \frac{1}{n} \sum_{i=1}^n E[f'_h(X) - f'_h(X^{(i)})] \\ &= \sum_{i=1}^n E[\xi_i^2 \int_0^1 f'_h(X^{(i)}) du] + \frac{1}{n} \sum_{i=1}^n E[f'_h(X) - f'_h(X^{(i)})]. \end{aligned} \tag{4.8}$$

Combining (4.7) and (4.8) yields (4.5) in Lemma 4.1. □

Hereafter, we consider the specific function  $h_{z,\alpha}(x)$ , defined by

$$h_{z,\alpha}(x) = \begin{cases} 1, & x \leq z, \\ 1 + (z - x)/\alpha, & z < x \leq z + \alpha, \\ 0, & x > z + \alpha, \end{cases}$$

for positive constant  $\alpha$ . It is noted that  $h_{z,\alpha}(x)$  is absolutely continuous and bounded as  $|h_{z,\alpha}(x)| \leq 1$ . Let  $f_{z,\alpha}(x)$  be the function given in (4.2) for  $h(x) = h_{z,\alpha}(x)$ .

**Lemma 4.2** *The function  $f_{z,\alpha}(x)$  satisfies that  $|f_{z,\alpha}(x)| \leq \sqrt{\pi/2}$ ,  $|f'_{z,\alpha}(x)| \leq 2$  and*

$$|f'_{z,\alpha}(w + x) - f'_{z,\alpha}(w)| \leq x(\sqrt{\pi/2} + 2|w|) + d(w, x), \tag{4.9}$$

where  $d(w, x) = |h_{z,\alpha}(w + x) - h_{z,\alpha}(w)|$ .

**Proof** For  $x > 0$ , from RHS of (4.2), it follows that

$$|f_{z,\alpha}(x)| \leq \int_x^\infty |h_{z,\alpha}(y) - \mu_{h_{z,\alpha}}| \phi(y) dy / \phi(x) \leq \frac{1 - \Phi(x)}{\phi(x)}.$$

Since  $\{1 - \Phi(x)\} / \phi(x)$  is decreasing, we have  $\{1 - \Phi(x)\} / \phi(x) \leq \{1 - \Phi(0)\} / \phi(0) = \sqrt{\pi/2}$ . For  $x < 0$ , it follows from (4.2) that

$$|f_{z,\alpha}(x)| = \int_{-\infty}^x |h_{z,\alpha}(y) - \mu_{h_{z,\alpha}}| \phi(y) dy / \phi(x) \leq \frac{\Phi(x)}{\phi(x)}.$$

Since  $\Phi(x) / \phi(x)$  is increasing, we have  $\Phi(x) / \phi(x) \leq \Phi(0) / \phi(0) = \sqrt{\pi/2}$ . Thus,  $|f_{z,\alpha}(x)| \leq \sqrt{\pi/2}$ .

We next show that  $|f'_{z,\alpha}(x)| \leq 2$ . Note that  $f'_{z,\alpha}(x) = x f_{z,\alpha}(x) + h_{z,\alpha}(y) - \mu_{h_{z,\alpha}}$ . For  $x > 0$ , it can be demonstrated that

$$\frac{x}{x^2 + 1} < \frac{1 - \Phi(x)}{\phi(x)} < \frac{1}{x},$$

which is called Mills' ratio. Then from RHS of (4.2), it follows that

$$\begin{aligned} |f'_{z,\alpha}(x)| &\leq x |f_{z,\alpha}(x)| + |h_{z,\alpha}(y) - \mu_{h_{z,\alpha}}| \leq x \int_x^\infty \phi(y) dy / \phi(x) + 1 \\ &= x \frac{1 - \Phi(x)}{\phi(x)} + 1 < 2. \end{aligned}$$

For  $x < 0$ , Mills' ratio implies that

$$\frac{1 - \Phi(-x)}{\phi(x)} < -\frac{1}{x}, \quad \text{or} \quad \frac{1}{x} < \frac{\Phi(x)}{\phi(x)},$$

namely we have  $x \Phi(x) / \phi(x) < 1$ . Then from RHS of (4.2), it follows that

$$|f'_{z,\alpha}(x)| \leq x \int_{-\infty}^x \phi(y) dy / \phi(x) + 1 = x \frac{\Phi(x)}{\phi(x)} + 1 < 2.$$

Thus, the inequality  $|f'_{z,\alpha}(x)| \leq 2$  is proved.

Finally, it is noted that  $f'_{z,\alpha}(w + x) - f'_{z,\alpha}(w) = (w + x) f_{z,\alpha}(w + x) - w f_{z,\alpha}(w) + h_{z,\alpha}(w + x) - h_{z,\alpha}(w)$ . Since  $|f'_{z,\alpha}(x)| < 2$ , it can be observed that  $|f_{z,\alpha}(w + x) - f_{z,\alpha}(w)| < 2|x|$ . Since  $|f_{z,\alpha}(x)| < \sqrt{\pi/2}$ , we have

$$\begin{aligned} &|f'_{z,\alpha}(w + x) - f'_{z,\alpha}(w)| \\ &\leq |x| |f_{z,\alpha}(w + x)| + |w| \{ |f_{z,\alpha}(w + x) - f_{z,\alpha}(w)| \} + |h_{z,\alpha}(w + x) - h_{z,\alpha}(w)| \\ &\leq |x| \sqrt{\pi/2} + 2|w| + d(w, x), \end{aligned}$$

which shows (4.9). □

We now show the central limit theorem using Lemmas 4.1 and 4.2. It is first noted that for  $Y \sim \mathcal{N}(0, 1)$ ,

$$\begin{aligned}
 |\mathbb{P}(X \leq z) - \Phi(z)| &= |\mathbb{E}[h_{z,\alpha}(X)] - \mathbb{E}[h_{z,\alpha}(Y)] + \mathbb{E}[h_{z,\alpha}(Y)] - \Phi(z)| \\
 &\leq \Delta + |\Phi(z + \alpha) - \Phi(z)|,
 \end{aligned}$$

where  $\Delta = |\mathbb{E}[h_{z,\alpha}(X)] - \mathbb{E}[h_{z,\alpha}(Y)]|$ . Since  $|\Phi(z + \alpha) - \Phi(z)| < \alpha/\sqrt{2\pi}$ , we have

$$|\mathbb{P}(X \leq z) - \Phi(z)| \leq \Delta + \alpha/\sqrt{2\pi}. \tag{4.10}$$

From (4.1), (4.2) and Lemma 4.1, it follows that

$$\begin{aligned}
 \Delta &= \sum_{i=1}^n \mathbb{E} \left[ \xi_i^2 \int_0^1 \{f'_{z,\alpha}(X^{(i)}) - f'_{z,\alpha}(X^{(i)} + u\xi_i)\} du \right] \\
 &\quad + \frac{1}{n} \sum_{i=1}^n \mathbb{E}[f'_{z,\alpha}(X^{(i)} + \xi_i) - f'_{z,\alpha}(X^{(i)})].
 \end{aligned} \tag{4.11}$$

From (4.9) in Lemma 4.2, it can be seen that

$$|f'_{z,\alpha}(X^{(i)} + \xi_i) - f'_{z,\alpha}(X^{(i)})| \leq |\xi_i|(\sqrt{\pi/2} + 2|X^{(i)}|) + d(X^{(i)}, \xi_i). \tag{4.12}$$

Note that  $d(X^{(i)}, \xi_i) = |\xi_i|/\alpha$  and  $(\mathbb{E}[|X^{(i)}|])^2 \leq \mathbb{E}[(X^{(i)})^2] = (n - 1)/n < 1$ . Then,

$$\begin{aligned}
 \frac{1}{n} \sum_{i=1}^n \mathbb{E}[f'_{z,\alpha}(X^{(i)} + \xi_i) - f'_{z,\alpha}(X^{(i)})] &\leq \frac{1}{n} \sum_{i=1}^n \mathbb{E}[|\xi_i|](\sqrt{\pi/2} + 2 + 1/\alpha) \\
 &= \left\{ \frac{\sqrt{\pi/2} + 2}{\sqrt{n}} + \frac{1}{\alpha\sqrt{n}} \right\} \frac{\sum_{i=1}^n \mathbb{E}[|X_i|]}{n}.
 \end{aligned}$$

Similarly,

$$\begin{aligned}
 \sum_{i=1}^n \mathbb{E} \left[ \xi_i^2 \int_0^1 \{f'_{z,\alpha}(X^{(i)}) - f'_{z,\alpha}(X^{(i)} + u\xi_i)\} du \right] \\
 \leq \left\{ \frac{\sqrt{\pi/2} + 2}{\sqrt{n}} + \frac{1}{\alpha\sqrt{n}} \right\} \frac{\sum_{i=1}^n \mathbb{E}[|X_i|^3]}{n}.
 \end{aligned}$$

Combining (4.10), (4.11) and these observations gives the following theorem.

**Theorem 4.1** For  $X = \sum_{i=1}^n X_i/\sqrt{n}$ , it holds that

$$|\mathbb{P}(X \leq z) - \Phi(z)| \leq \left\{ \frac{\sqrt{\pi/2} + 2}{\sqrt{n}} + \frac{1}{\alpha\sqrt{n}} \right\} (\mu_1 + \mu_3) + \frac{\alpha}{\sqrt{2\pi}}. \tag{4.13}$$

Assume that  $\sum_{i=1}^n E[|X_i|^3]/n$  converges to a positive constant. Let  $\alpha = n^{-1/4}$ . Then,  $|\mathbb{P}(X \leq z) - \Phi(z)| \rightarrow 0$  as  $n \rightarrow \infty$ .

The inequality in (4.17) is improved if we use the following concentration inequality due to Chen et al. (2011).

**Lemma 4.3** For any  $a$  and  $b$  satisfying  $a < b$ ,

$$\mathbb{P}(a \leq X^{(i)} \leq b) \leq \sqrt{2}(b - a) + \frac{2(\sqrt{2} + 1)\mu_3}{\sqrt{n}}. \tag{4.14}$$

In the evaluation of  $E[d(X^{(i)}, \xi_i)]$  in (4.12), we can see that for  $\xi_i > 0$ ,

$$E[d(X^{(i)}, \xi_i)|\xi_i] \leq E[I(z - \xi_i \leq X^{(i)} < z + \alpha)|\xi_i] = \mathbb{P}(z - \xi_i \leq X^{(i)} < z + \alpha)|\xi_i.$$

From the inequality (4.14), it follows that for all  $\xi_i$ ,

$$E[d(X^{(i)}, \xi_i)|\xi_i] \leq \sqrt{2}(|\xi_i| + \alpha) + \frac{2(\sqrt{2} + 1)\mu_3}{\sqrt{n}},$$

which gives

$$E[|f'_{z,\alpha}(X^{(i)} + \xi_i) - f'_{z,\alpha}(X^{(i)})||\xi_i] \leq |\xi_i|(\sqrt{\pi/2} + 2) + \sqrt{2}(|\xi_i| + \alpha) + \frac{2(\sqrt{2} + 1)\mu_3}{\sqrt{n}}.$$

Thus, the same arguments used above provide the inequality

$$|\mathbb{P}(X \leq z) - \Phi(z)| \leq \frac{C(\alpha)}{\sqrt{n}}, \tag{4.15}$$

where

$$C(\alpha) = (\sqrt{\pi/2} + 2 + \sqrt{2})(\mu_1 + \mu_3) + \{\sqrt{2n}\alpha + 2(\sqrt{2} + 1)\mu_3\}(1 + \mu_2) + \alpha\sqrt{n/(2\pi)}.$$

When  $\alpha = 1/\sqrt{n}$ , we have

$$C(1/\sqrt{n}) = (\sqrt{\pi/2} + 2 + \sqrt{2})(\mu_1 + \mu_3) + \{\sqrt{2} + 2(\sqrt{2} + 1)\mu_3\}(1 + \mu_2) + 1/\sqrt{2\pi}.$$

The inequality provides a Berry–Esseen-type bound.

### 4.2 A method based on the K-function

Another method based on the K-function is useful for evaluating  $E[f'_h(X) - Xf_h(X)]$  for  $f_h(\cdot)$  in (4.2). Define  $K_i(t)$  by

$$K_i(t) = E[\xi_i \{I(0 \leq t \leq \xi_i) - I(\xi_i \leq t < 0)\}].$$

It can be seen that  $K_i(t) \geq 0$  for  $t \in \mathfrak{R}$  and that

$$\int_{-\infty}^{\infty} K_i(t)dt = E[\xi_i^2] \quad \text{and} \quad \int_{-\infty}^{\infty} |t|K_i(t)dt = \frac{1}{2}E[|\xi_i|^3].$$

**Lemma 4.4** For  $f_h(\cdot)$  in (4.2), it holds that

$$E[f'_h(X) - Xf_h(X)] = \sum_{i=1}^n \int_{-\infty}^{\infty} E[f'_h(X) - f'_h(X^{(i)} + t)]K_i(t)dt. \quad (4.16)$$

**Proof** Since  $\xi_i$  and  $X^{(i)}$  are independent and  $E[\xi_i] = 0$ , it is observed that  $E[Xf_h(X)] = \sum_{i=1}^n E[\xi_i f_h(X)] = \sum_{i=1}^n E[\xi_i \{f_h(X) - f(X^{(i)})\}]$ , which is rewritten as

$$\begin{aligned} & \sum_{i=1}^n E\left[\xi_i \int_0^{\xi_i} f'_h(X^{(i)} + t)dt\right] \\ &= \sum_{i=1}^n E\left[\int_{-\infty}^{\infty} f'_h(X^{(i)} + t)\xi_i\{I(0 \leq t \leq \xi_i) - I(\xi_i \leq t < 0)\}dt\right] \\ &= \sum_{i=1}^n E\left[\int_{-\infty}^{\infty} f'_h(X^{(i)} + t)K_i(t)dt\right]. \end{aligned}$$

Since  $\sum_{i=1}^n \int_{-\infty}^{\infty} K_i(t)dt = \sum_{i=1}^n E[\xi_i^2] = 1$ , we have  $E[f'_h(X)] = \sum_{i=1}^n E[\int_{-\infty}^{\infty} f'_h(X)K_i(t)dt]$ . Combining these observations gives the expression in (4.16).  $\square$

We treat  $h_z(x) = I(x \leq z) - \Phi(z)$  and the function  $f_z(x)$  given in (4.3). From (4.1), it follows that

$$\begin{aligned} f'_z(X) - f'_z(X^{(i)} + t) &= \{Xf_z(X) - (X^{(i)} + t)f_z(X^{(i)} + t)\} \\ &\quad + I(z - \xi_i \vee t < X^{(i)} < z - \xi_i \wedge t), \end{aligned}$$

where  $a \vee b = \max(a, b)$  and  $a \wedge b = \min(a, b)$ . Similarly to Lemma 4.2, it can be shown that

$$\begin{aligned} & (w + u)f_z(w + u) - (w + v)f_z(w + v) \\ & \leq |w||f_z(w + u) - f_z(w + v)| \\ & \quad + uf_z(w + u) - vf_z(w + v) \\ & \leq (|w| + \sqrt{2\pi}/4)(|u| + |v|), \end{aligned}$$

because  $|f_z(x)| \leq \sqrt{2\pi}/4$ . Thus,

$$E[Xf_z(X) - (X^{(i)} + t)f_z(X^{(i)} + t)|\xi_i] \leq 1 + (\sqrt{2\pi}/4)(|\xi_i| + t).$$

From Lemma 4.3, it follows that

$$P(z - \xi_i \vee t < X^{(i)} < z - \xi_i \wedge t | \bar{x}_i) \leq \sqrt{2}|\xi_i - t| + 2(\sqrt{2} + 1) \frac{\mu_3}{\sqrt{n}}.$$

Hence from Lemma 4.4, we get

$$\begin{aligned} & \sum_{i=1}^n \int_{-\infty}^{\infty} E[f'_h(X) - f'_h(X^{(i)} + t)]K_i(t)dt \\ & \leq \sum_{i=1}^n \int_{-\infty}^{\infty} \left\{ (1 + \sqrt{2\pi}/4 + \sqrt{2})(E[|\xi_i|] + |t|) + \frac{2(\sqrt{2} + 1)\mu_3}{\sqrt{n}} \right\} K_i(t)dt, \end{aligned}$$

which yields the following bound.

**Theorem 4.2** For  $X = \sum_{i=1}^n X_i/\sqrt{n}$ , it holds that

$$\begin{aligned} & |P(X \leq z) - \Phi(z)| \\ & \leq \frac{(1 + \sqrt{2} + \sqrt{2\pi}/4)(\mu_1\mu_2 + \mu_3/2) + 2(1 + \sqrt{2})\mu_3}{\sqrt{n}}. \end{aligned} \tag{4.17}$$

Chen and Shao (2001) derived a more refined concentration inequality and obtained the improved bound given by

$$|P(X \leq z) - \Phi(z)| \leq 4.1(\beta_2 + \beta_3), \tag{4.18}$$

where  $\beta_2 = \sum_{i=1}^n E[\xi_i^2 I(|\xi_i| > 1)]$  and  $\beta_3 = \sum_{i=1}^n E[\xi_i^3 I(|\xi_i| \leq 1)]$ . This corresponds to the Lindeberg’s condition. In fact, let  $X_1, \dots, X_n$  be independent random variables with  $E[X_i] = 0$  and  $\text{Var}(X_i) = \sigma_i^2$ . Let  $S_n = \sum_{i=1}^n X_i$  and  $B_n^2 = \sum_{i=1}^n \sigma_i^2$ . Then,  $\xi_i$  and  $X$  correspond to  $\xi_i = X_i/B_n$  and  $X = S_n/B_n$ . It is observed that for any  $\varepsilon > 0$ ,

$$\begin{aligned} \beta_2 + \beta_3 &= \frac{1}{B_n^2} \sum_{i=1}^n E[X_i^2 I(|X_i| > B_n)] + \frac{1}{B_n^3} \sum_{i=1}^n E[X_i^3 I(|X_i| \leq B_n)] \\ &\leq \frac{1}{B_n^2} \sum_{i=1}^n E[X_i^2 I(|X_i| > B_n)] + \frac{1}{B_n^3} \sum_{i=1}^n B_n E[X_i^2 I(\varepsilon B_n \leq |X_i| \leq B_n)] \\ &\quad + \frac{1}{B_n^3} \sum_{i=1}^n \varepsilon B_n E[X_i^2 I(|X_i| \leq \varepsilon B_n)] \\ &\leq \frac{1}{B_n^2} \sum_{i=1}^n E[X_i^2 I(|X_i| > \varepsilon B_n)] + \varepsilon. \end{aligned}$$

Thus from (4.18), it follows that

$$|P(S_n/B_n \leq z) - \Phi(z)| \leq 4.1 \left\{ \frac{1}{B_n^2} \sum_{i=1}^n E[X_i^2 I(|X_i| > \varepsilon B_n)] + \varepsilon \right\},$$

which converges to zero if the Lindeberg condition

$$\lim_{n \rightarrow \infty} \frac{1}{B_n^2} \sum_{i=1}^n E[X_i^2 I(|X_i| > \varepsilon B_n)] = 0$$

is satisfied.

## 5 Stein-type identities in gamma and exponential distributions

### 5.1 Stein-type identities and characterization of gamma and exponential distributions

We treat the gamma distribution  $Ga(\alpha, \beta)$  with the density function

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad x > 0,$$

where  $\alpha$  and  $\beta$  are positive parameters. Hudson (1978) derived the Stein-type identity in the gamma distribution, and Betsch and Ebner (2019) showed that the identity characterizes the gamma distribution.

**Theorem 5.1** *Let  $X$  be a positive random variable with  $E[X] = \alpha\beta$  and  $\text{Var}(X) = \alpha\beta^2$ . Then, the following four conditions are equivalent.*

- (a)  $X \sim Ga(\alpha, \beta)$ .
- (b) For any differentiable function  $h(\cdot)$  with  $E[|Xh(X)|] < \infty$  and  $E[|Xh'(X)|] < \infty$ , it holds that

$$E[(X - \alpha\beta)h(X)] = \beta E[Xh'(X)]. \tag{5.1}$$

- (c) For real constant  $t$  with  $t < 1/\beta$ , it holds that

$$E[(X - \alpha\beta) \exp\{tX\}] = t\beta E[X \exp\{tX\}]. \tag{5.2}$$

- (d)  $g(t) = E[\exp\{tX\}]$  satisfies the differential equation

$$\frac{d}{dt} \log\{g(t)\} = \frac{\alpha\beta}{1 - t\beta} \quad \text{or} \quad \frac{d^2}{dt^2} \log\{g(t)\} = \frac{\alpha\beta^2}{(1 - t\beta)^2}. \tag{5.3}$$



**Proof** For the proof from (a) to (b), the identity (5.1) can be derived by integration by parts, because  $\beta(d/dx)f(x|\alpha, \beta) = -(x - \alpha\beta)f(x|\alpha, \beta)$ . We here provide another approach. Making the transformation  $y = x/\beta$  gives the expression

$$E[h(X)] = \int_0^\infty h(x) \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx = \int_0^\infty h(\beta y) \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-y} dy.$$

Differentiating both sides with respect to  $\beta$ , we have

$$\begin{aligned} \int_0^\infty h(x) \left( \frac{x}{\beta^2} - \frac{\alpha}{\beta} \right) \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx &= \int_0^\infty y h'(\beta y) \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-y} dy \\ &= \int_0^\infty \frac{x}{\beta} h'(x) \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx, \end{aligned}$$

which leads to (5.1). Clearly, one gets (c) from (b). For the proof from (c) to (d), the identity (5.2) is written as  $g'(t) - \alpha\beta g(t) = t\beta g'(t)$ , which is expressed as (5.3). For the proof from (d) to (a), the solution of the differential equation in (5.3) is  $\log g(t) = -\alpha \log(1 - t\beta)$ , namely  $g(t) = (1 - t\beta)^{-\alpha}$ , which implies that  $X \sim Ga(\alpha, \beta)$ .  $\square$

We here provide some conditions for characterizing gamma distributions. Condition (c) in Theorem 5.2 is due to Lukacs (1955). See also Kagan et al. (1973) and Kotz (1974).

**Theorem 5.2** *Assume that independent positive random variables  $X_1$  and  $X_2$  are identically distributed with  $E[X_i] = \alpha\beta$  and  $\text{Var}(X_i) = \alpha\beta^2$ . Then the following three are equivalent.*

- (a)  $X_i \sim Ga(\alpha, \beta)$  for  $i = 1, 2$ .
- (b)  $X_1 + X_2 \sim Ga(2\alpha, \beta)$ .
- (c)  $X_1 + X_2$  and  $X_1/(X_1 + X_2)$  are independent.

**Proof** Since one gets clearly (b) and (c) from (a), it is sufficient to demonstrate the opposite directions. For the proof from (b) to (a), the condition in (b) and Theorem 5.1(c) implies that

$$E[(X_1 + X_2 - 2\alpha\beta) \exp\{t(X_1 + X_2)\}] = \beta t E[(X_1 + X_2) \exp\{t(X_1 + X_2)\}]. \tag{5.4}$$

From the independence of  $X_1$  and  $X_2$ , equality (5.4) is rewritten as

$$E[(X_1 - \alpha\beta) \exp\{t X_1\}] E[\exp\{t X_2\}] = \beta t E[X_1 \exp\{t X_1\}] E[\exp\{t X_2\}],$$

which, from Theorem 5.1, shows that  $X_i \sim Ga(\alpha, \beta)$ .

The proof from (c) to (a) can be done along with the proof of Lukacs (1955). From the independence of  $X_1 + X_2$  and  $X_1/(X_1 + X_2)$ , it follows that

$$E\left[ \exp \left\{ s(X_1 + X_2) + t \frac{X_1}{X_1 + X_2} \right\} \right]$$

$$= E[\exp\{s(X_1 + X_2)\}]E\left[\exp\left\{t\frac{X_1}{X_1 + X_2}\right\}\right]. \tag{5.5}$$

Differentiating both sides twice with respect to  $s$  and  $t$  and putting  $t = 0$ , we have

$$\begin{aligned} & E\left[X_1^2 \exp\left\{s(X_1 + X_2)\right\}\right] \\ &= E[(X_1 + X_2)^2 \exp\{s(X_1 + X_2)\}]E\left[\left(\frac{X_1}{X_1 + X_2}\right)^2\right]. \end{aligned} \tag{5.6}$$

Let  $a = E[X_1^2/(X_1 + X_2)^2]$  and  $g(s) = E[\exp\{sX_1\}]$ . Then,

$$\begin{aligned} E[(X_1 + X_2)^2 \exp\{s(X_1 + X_2)\}] &= E[(X_1^2 + X_2^2 + 2X_1X_2) \exp\{s(X_1 + X_2)\}] \\ &= 2g''(s)g(s) + 2\{g'(s)\}^2, \end{aligned}$$

which rewrite (5.6) as

$$g''(s)g(s) = 2a[g''(s)g(s) + \{g'(s)\}^2] \quad \text{or} \quad (1 - 2a)g''(s)g(s) = \{g'(s)\}^2. \tag{5.7}$$

Let  $\psi(s) = g'(s)/g(s)$ , and we have  $\psi'(s) = g''(s)/g(s) - \{\psi(s)\}^2$ . The equality (5.7) is expressed as

$$(1 - 2a)[\psi'(s) + \{\psi(s)\}^2] = \{\psi(s)\}^2 \quad \text{or} \quad (1 - 2a)\psi'(s) = 2a\{\psi(s)\}^2.$$

The solution of the differential equation is

$$-\frac{1}{\psi(s)} = \frac{2a}{1 - 2a}s - c_0 \quad \text{or} \quad \psi(s) = \frac{1}{c_0 - bs}$$

for  $b = 2a/(1 - 2a)$  and constant  $c_0$ . Since  $\psi(s) = (d/ds) \log g(s)$ , we have

$$\log g(s) = -\frac{1}{b} \log(c_0 - bs) + \log c_1 \quad \text{or} \quad g(s) = \frac{c_1}{(c_0 - bs)^{1/b}}.$$

Since  $g(0) = 1$ ,  $g'(0) = \alpha\beta$  and  $g''(0) = \alpha\beta(1 + \beta)$ , constants satisfy  $c_1/c_0^{1/b} = 1$ ,  $c_1/c_0^{1+1/b} = \alpha\beta$  and  $c_1(1 + b)/c_0^{2+1/b} = \alpha(\alpha + 1)\beta^2$ , which gives  $c_0 = (\alpha\beta)^{-1}$ ,  $c_1 = (\alpha\beta)^{-\alpha}$  and  $b = 1/\alpha$ . This yields  $g(s) = (1 - \beta s)^{-\alpha}$ , which means that  $X_i \sim Ga(\alpha, \beta)$  for  $i = 1, 2$ . □

Condition (b) can be easily extended to the case of a sample with size  $n$ . Such an extension of condition (c) was done by Khatri and Rao (1968).

The exponential distribution  $Ex(\lambda)$  corresponds to the case of  $\alpha = 1$  and  $\beta = 1/\lambda$ . The characterization problem of the exponential distribution has been studied in the literature. Of these, Shanbhag (1970a) showed that memoryless property characterizes the exponential distribution:  $X \sim Ex(\lambda)$  if and only if for any  $x, y > 0$ ,

$$P(X > x + y|X > y) = P(X > x) \quad \text{or} \quad P(X > x + y) = P(X > x)P(X > y).$$

In fact,  $\log P(X > x + y) = \log P(X > x) + \log P(X > y)$  implies that  $\log P(X > x) = -cx$  for  $c > 0$ , namely  $P(X > x) = \exp(-cx)$ , which means that  $X$  has an exponential distribution. For independently and identically distributed positive random variables  $X_1$  and  $X_2$ , Ferguson (1964) proved that  $X_i$  has an exponential distribution if and only if  $\min(X_1, X_2)$  is independent of  $X_1 - X_2$ .

### 5.2 Shrinkage estimation

The Stein identity is useful for obtaining improved shrinkage estimators in simultaneous estimation in gamma distributions. Let  $X_1, \dots, X_p$  be independent random variables such that  $X_i \sim Ga(\alpha_i, \beta_i)$ ,  $i = 1, \dots, p$ . We first consider the simultaneous estimation of  $\alpha = (\alpha_1, \dots, \alpha_p)^T$  in the case of  $\beta_1 = \dots = \beta_p = 1$ . When estimator  $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_p)^T$  is evaluated by the risk relative to the quadratic loss  $\sum_{i=1}^p (\hat{\alpha}_i - \alpha_i)^2$ , Hudson (1978) suggested the shrinkage estimator

$$\hat{\alpha}_i^S = X_i - \frac{p - 2}{\sum_{j=1}^p (\log X_j)^2} \log X_i. \tag{5.8}$$

**Theorem 5.3** For  $p \geq 3$ , the risk functions have the relation

$$E[\|X - \alpha\|^2] = E[\|\hat{\alpha}^S - \alpha\|^2] + E[\|X - \hat{\alpha}^S\|^2],$$

which is interpreted as the Pythagorean triangle among  $X$ ,  $\alpha$  and  $\hat{\alpha}^S$ .

**Proof** The estimator in (5.8) has the risk

$$\begin{aligned} E[\|\hat{\alpha}^S - \alpha\|^2] &= E[\|X - \alpha\|^2] - 2 \sum_{i=1}^p E\left[ (X_i - \alpha_i) \frac{p - 2}{\sum_{j=1}^p (\log X_j)^2} \log X_i \right] \\ &\quad + E\left[ \frac{(p - 2)^2}{\sum_{j=1}^p (\log X_j)^2} \right]. \end{aligned}$$

The Stein-type identity in (5.1) gives  $E[(X_i - \alpha_i)h(X_i)] = E[X_i h'(X_i)]$  for

$$h(X_i) = \frac{p - 2}{\sum_{j=1}^p (\log X_j)^2} \log X_i.$$

Noting that

$$X_i \frac{\partial h(X_i)}{\partial X_i} = \frac{p - 2}{\sum_{j=1}^p (\log X_j)^2} - \frac{2(p - 2)(\log X_i)^2}{\{\sum_{j=1}^p (\log X_j)^2\}^2},$$

we can see that

$$\sum_{i=1}^p X_i \frac{\partial h(X_i)}{\partial X_i} = \frac{(p - 2)p}{\sum_{j=1}^p (\log X_j)^2} - \frac{2(p - 2) \sum_{i=1}^p (\log X_i)^2}{\{\sum_{j=1}^p (\log X_j)^2\}^2} = \frac{(p - 2)^2}{\sum_{j=1}^p (\log X_j)^2}.$$

**Table 3** Risks of the estimators  $\mathbf{X}$  and  $\hat{\alpha}^S$  for  $p = 3, 6$  and  $\alpha = (k/3)\mathbf{I}$

$k$	$p = 3$									
	1	2	3	4	5	6	7	8	9	10
$\mathbf{X}$	1.00	2.00	2.98	4.01	4.99	6.00	6.98	8.02	9.00	9.96
$\hat{\alpha}^S$	0.94	1.61	2.16	2.92	3.87	5.03	6.23	7.44	8.57	9.63
$k$	$p = 6$									
	1	2	3	4	5	6	7	8	9	10
$\mathbf{X}$	2.01	3.99	5.99	7.96	10.0	12.0	14.3	16.0	18.3	20.0
$\hat{\alpha}^S$	1.74	2.79	3.21	3.77	5.30	7.59	10.3	13.0	15.5	17.9

Thus,

$$\sum_{i=1}^p E\left[(X_i - \alpha_i) \frac{p - 2}{\sum_{j=1}^p (\log X_j)^2} \log X_i\right] = E\left[\frac{(p - 2)^2}{\sum_{j=1}^p (\log X_j)^2}\right],$$

which is used to rewrite the risk as

$$E[\|\hat{\alpha}^S - \alpha\|^2] = E[\|\mathbf{X} - \alpha\|^2] - E\left[\frac{(p - 2)^2}{\sum_{j=1}^p (\log X_j)^2}\right].$$

This shows that the estimator  $\hat{\alpha}^S$  improves on  $\mathbf{X}$  for  $p \geq 3$ . Since  $(p - 2)^2 / \sum_{j=1}^p (\log X_j)^2 = \|\mathbf{X} - \hat{\alpha}^S\|^2$ , the above risk function expresses the Pythagorean triangle.  $\square$

The risk performances of the estimators  $\mathbf{X}$  and  $\hat{\alpha}^S$  are investigated by simulation with 10, 000 replications and the average values of the risks are reported in Table 3 for  $p = 3, 6, \beta = 1$  and  $\alpha = (k/3)\mathbf{I}, k = 1, \dots, 10$ . Table 3 shows that the improvement of the shrinkage estimator  $\hat{\alpha}^S$  over  $\mathbf{X}$  is significant in the case of  $p = 6$ .

We next consider the simultaneous estimation of  $\beta = (\beta_1, \dots, \beta_p)$  for known common  $\alpha_1 = \dots = \alpha_p = \alpha$ , where estimator  $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$  is evaluated by the risk relative to the quadratic loss  $\sum_{i=1}^p (\hat{\beta}_i - \beta_i)^2$ . This estimation problem was studied by Berger (1980), Das Gupta (1986) and others. Das Gupta (1986) suggested the shrinkage estimator  $\hat{\beta}^S = (\hat{\beta}_1^S, \dots, \hat{\beta}_p^S)$  with

$$\hat{\beta}_i^S = \frac{1}{\alpha + 1} X_i + cV, \quad V = \left(\prod_{j=1}^p X_j\right)^{1/p}, \tag{5.9}$$

and derived a condition for improving on  $\hat{\beta}_0 = \mathbf{X}/(\alpha + 1)$ .

**Theorem 5.4** *When  $p \geq 2$ , the estimator  $\hat{\beta}^S$  improves on  $\hat{\beta}_0$  relative to the quadratic loss if*

$$0 < c \leq \frac{2(p - 1)}{(\alpha + 1)(\alpha p + 1)}.$$

**Table 4** Risks of the estimators  $\widehat{\beta}_0$  and  $\widehat{\beta}^S$  for  $p = 2, 6, \alpha = 1$  and  $\beta = (k/3)\mathbf{I}$

$k$	$p = 2$									
	1	2	3	4	5	6	7	8	9	10
$\widehat{\beta}_0$	0.11	0.45	1.00	1.78	2.78	4.01	5.44	7.10	8.99	11.15
$\widehat{\beta}^S$	0.10	0.41	0.93	1.65	2.57	3.71	5.03	6.55	8.29	10.33
$k$	$p = 6$									
	1	2	3	4	5	6	7	8	9	10
$\widehat{\beta}_0$	0.33	1.33	3.00	5.34	8.35	12.0	16.3	21.3	27.0	33.3
$\widehat{\beta}^S$	0.25	1.00	2.24	4.01	6.26	8.99	12.2	15.9	20.2	25.0

**Proof** The risk function of the estimator (5.9) is

$$E[\|\widehat{\beta}^S - \beta\|^2] = E[\|\widehat{\beta}_0 - \beta\|^2] + 2 \sum_{i=1}^p E\left[\left(\frac{X_i}{\alpha + 1} - \beta_i\right)cV\right] + pE[c^2V^2].$$

The Stein-type identity in (5.1) gives

$$E[X_i V] = \beta_i E\left[\alpha V + X_i \frac{\partial V}{\partial X_i}\right] = \beta_i \frac{\alpha p + 1}{p} E[V],$$

because  $\partial V / \partial X_i = V / (pX_i)$ . Thus, the risk difference is written as

$$\begin{aligned} \Delta &= E[\|\widehat{\beta}^S - \beta\|^2] - E[\|\widehat{\beta}_0 - \beta\|^2] = \sum_{i=1}^p E\left[2\frac{c}{\alpha + 1}VX_i - 2\frac{cp}{\alpha p + 1}VX_i + pc^2V^2\right] \\ &= cE\left[-2\frac{(p - 1)p}{(\alpha + 1)(\alpha p + 1)}\bar{X}V + pcV^2\right] \leq cE\left[-2\frac{(p - 1)p}{(\alpha + 1)(\alpha p + 1)}V^2 + pcV^2\right], \end{aligned}$$

because  $\bar{X} \geq V$ . This shows that  $\Delta \leq 0$  under the condition in Theorem 5.4. □

The risk performances of the estimators  $\widehat{\beta}_0$  and  $\widehat{\beta}^S$  are investigated by simulation with 10, 000 replications and the average values of the risks are reported in Table 4 for  $p = 2, 6, \alpha = 1$  and  $\beta = (k/3)\mathbf{I}, k = 1, \dots, 10$ . From the table, the improvement of the shrinkage estimator  $\widehat{\beta}^S$  over  $\widehat{\beta}_0$  is numerically confirmed.

### 5.3 Goodness-of-fit tests for exponentiality

We consider to construct a statistic for testing exponentiality using the Stein identity. Goodness-of-fit tests for exponentiality have been studied in the literature. For some good reviews, see Henze and Meintanis (2005) and Ossai et al. (2022). The idea of constructing test statistics for exponentiality based on the Stein identity appears in Betsch and Ebner (2019) and Henze et al. (2012).

Let  $X_1, \dots, X_n$  be a positive random sample from a distribution function  $F(\cdot)$  with  $E[X_i] = \sigma$ . Consider the problem of testing the exponentiality of the underlying distribution  $H_0 : F = Ex(1/\sigma) = Ga(1, \sigma)$ . From Theorem 5.1, the characterization of exponential distributions is

$$E\left[\{(1 - t)X/\sigma - 1\}e^{tX/\sigma}\right] = 0,$$

and the sample counterpart is  $w_t/\sqrt{n}$ , where

$$w_t = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{(1 - t)Y_i - 1\}e^{tY_i} \tag{5.10}$$

for  $Y_i = X_i/\bar{X}$ . It is noted that  $w_t$  is invariant under the transformation of scale. Henze et al. (2012) suggested a couple of test statistics based on  $w_t$ , one of which is

$$HME_\gamma = \int_{-\infty}^0 w_t^2 e^{\gamma t} dt = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left\{ \frac{B_{ij} - A_{ij} + 1}{A_{ij} + \gamma} + \frac{2B_{ij} - A_{ij}}{(A_{ij} + \gamma)^2} + \frac{2B_{ij}}{(A_{ij} + \gamma)^3} \right\},$$

where  $A_{ij} = Y_i + Y_j$  and  $B_{ij} = Y_i Y_j$ . Similarly to the problem of testing normality, we can consider the test statistics  $IST_c^+ = \int_0^c w_t^2 dt$ ,  $IST_c^- = \int_{-c}^0 w_t^2 dt$ ,  $IST_c = \int_{-c}^c w_t^2 dt$  and  $MST_c = \sup_{-c < t < c} |w_t|$  for positive constant  $c$ , where

$$\begin{aligned} IST_c^+ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left[ \left\{ c \left( c - 2 - \frac{2}{A_{ij}} \right) e^{cA_{ij}} - \left( 1 + \frac{2}{A_{ij}} + \frac{2}{A_{ij}^2} \right) (1 - e^{cA_{ij}}) \right\} \frac{B_{ij}}{A_{ij}} \right. \\ &\quad \left. + (c - 1)e^{cA_{ij}} + 1 \right], \\ IST_c^- &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left[ \left\{ -c \left( c + 2 + \frac{2}{A_{ij}} \right) e^{-cA_{ij}} + \left( 1 + \frac{2}{A_{ij}} + \frac{2}{A_{ij}^2} \right) (1 - e^{-cA_{ij}}) \right\} \frac{B_{ij}}{A_{ij}} \right. \\ &\quad \left. + (c + 1)e^{-cA_{ij}} - 1 \right], \\ IST_c &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left[ \left\{ \left( c^2 + 1 + \frac{2}{A_{ij}} + \frac{2}{A_{ij}^2} \right) (e^{cA_{ij}} - e^{-cA_{ij}}) \right. \right. \\ &\quad \left. \left. - \frac{2c}{A_{ij}} \left( 1 + \frac{1}{A_{ij}} \right) (e^{cA_{ij}} + e^{-cA_{ij}}) \right\} \frac{B_{ij}}{A_{ij}} \right. \\ &\quad \left. - e^{cA_{ij}} + e^{-cA_{ij}} + c(e^{cA_{ij}} + e^{-cA_{ij}}) \right]. \end{aligned}$$

The following lemma is helpful for investigating asymptotic properties of these test statistics.

**Lemma 5.1** *Let  $Z_i = X_i/\sigma$  for  $i = 1, \dots, n$ . Let  $h_0(t) = E[\{(1 - t)Z_1 - 1\}e^{tZ_1}]$  and  $h_1(t) = (1 - 2t)E[Z_1 e^{tZ_1}] + t(1 - t)E[Z_1^2 e^{tZ_1}]$ . Then,  $w_t$  is approximated as*

$w_t = W_n(t) + \sqrt{nh_0(t)} + o_p(1)$ , where

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \{(1-t)Z_i - 1\} e^{tZ_i} - h_0(t) - (Z_i - 1)h_1(t) \right]. \tag{5.11}$$

**Proof** Letting  $\bar{Z} = n^{-1} \sum_{i=1}^n Z_i$ , we rewrite  $w_t$  as

$$w_t = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{(1-t)Z_i/\bar{Z} - 1\} e^{tZ_i/\bar{Z}}.$$

Since  $Z_i/\bar{Z} = Z_i/\{1 + (\bar{Z} - 1)\} = 1 - (\bar{Z} - 1) + O_p(n^{-1})$ ,  $w_t$  can be approximated as

$$\begin{aligned} w_t &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{(1-t)Z_i - (1-t)(\bar{Z} - 1)Z_i - 1 + O_p(n^{-1})\} e^{tZ_i - t(\bar{Z} - 1)Z_i + O_p(n^{-1})} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{(1-t)Z_i - 1 - (1-t)(\bar{Z} - 1)Z_i\} \{1 - t(\bar{Z} - 1)Z_i\} e^{tZ_i} + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{(1-t)Z_i - 1 - (1-2t)(\bar{Z} - 1)Z_i - t(1-t)(\bar{Z} - 1)Z_i^2\} e^{tZ_i} + o_p(1), \end{aligned}$$

which leads to the approximation  $w_t = W_n + \sqrt{nh_0(t)} + o_p(1)$ . Hence, Lemma 5.1 is proved. □

From (5.11), the central limit theorem shows that  $W_n(t)$  is asymptotically distributed as the normal distribution with mean zero and variance  $\text{Var}(W_n(t))$  under the assumption of  $E[Z_1^2 e^{2tZ_1}] < \infty$ , where the variance can be evaluated as

$$\begin{aligned} \text{Var}(W_n(t)) &= E[\{(1-t)Z_1 - 1\}^2 e^{2tZ_1}] - h_0(t)^2 + h_1(t)^2 (E[Z_1^2] - 1) \\ &\quad - 2h_1(t)E[(Z_1 - 1)\{(1-t)Z_1 - 1\} e^{tZ_1}] + o(1). \end{aligned}$$

Note that  $h_0(t) = (1-t)g'(t) - g(t)$  for  $g(t) = E[e^{tZ_1}]$ . Then,  $h_1(t) = (1-2t)g'(t) + t(1-t)g''(t) = th_0'(t) + \{h_0(t) + g(t)\}/(1-t)$  and

$$E[(Z_1 - 1)\{(1-t)Z_1 - 1\} e^{tZ_1}] = h_0'(t) + \frac{t}{1-t} h_0(t) + \frac{1}{1-t} g(t).$$

Under the exponentiality hypothesis  $H_0$ , we have  $g(t) = 1/(1-t)$ ,  $h_0(t) = 0$ ,  $h_1(t) = g(t)/(1-t) = 1/(1-t)^2$  and  $E[Z_1^2] = 2$ . Also note that

$$\begin{aligned} E[\{(1-t)Z_1 - 1\}^2 e^{2tZ_1}] &= (1-t)^2 E[Z_1^2 e^{2tZ_1}] - 2(1-t)E[Z_1 e^{2tZ_1}] + E[e^{2tZ_1}] \\ &= \frac{2(1-t)^2}{(1-2t)^3} - 2\frac{1-t}{(1-2t)^2} + \frac{1}{1-2t} = \frac{(1-t)^2 + t^2}{(1-2t)^3}. \end{aligned}$$

Thus, the asymptotic variance of  $w_t$  under the exponentiality is  $\text{Var}(W_n(t)) = V(t)^2 + o(1)$ , where

$$V(t)^2 = \frac{(1-t)^2 + t^2}{(1-2t)^3} - \frac{1}{(1-t)^4}$$

for  $0 < t < 1/2$ .

Henze et al. (2012) showed the consistency of  $\text{HME}_\gamma$ . Using Lemma 5.1 and the same arguments as in the proof of Theorem 3.3, we can verify the consistency of the suggested test statistics.

**Theorem 5.5** *Assume that  $E[Z_1^2 e^{tZ_1}] < \infty$  for  $t$  around zero. Then, the test statistics  $\text{IST}_c^+$ ,  $\text{IST}_c^-$ ,  $\text{IST}_c$  and  $\text{MST}_c$  given below (5.10) are consistent.*

We investigate the performances of powers of the suggested estimators  $\text{IST}_c^+$ ,  $\text{IST}_c^-$ ,  $\text{IST}_c$  and  $\text{MST}_c = \sup_{-c < t < c} |w_t|$  for  $c = 0.1$ . As competitors, we treat the test  $\text{HME}_\gamma$  of Henze et al. (2012) for  $\gamma = 1$  and two more simple test statistics. One of them is the Cox and Oakes (1984) test

$$\text{CO} = \frac{\sqrt{6}}{\sqrt{n\pi}} \left\{ n + \sum_{i=1}^n \left( 1 - \frac{X_i}{\bar{X}} \right) \log \left( \frac{X_i}{\bar{X}} \right) \right\},$$

and the hypothesis  $H_0$  is rejected when  $|\text{CO}| > z_{\alpha/2}$ . Another test is based on the coefficient of variation  $\text{HS} = nS^2/\bar{X}^2$ , given in Hahn and Shapiro (1967), and the hypothesis  $H_0$  is rejected when  $\text{HS} < \chi_{n-1, 1-\alpha/2}^2$  or  $\text{HS} > \chi_{n-1, \alpha/2}^2$ . This is also derived from a likelihood ratio for testing the homogeneity  $H_0^* : \lambda_1 = \dots = \lambda_n$  for  $X_i \sim \text{Ex}(\lambda_i)$ , and the null hypothesis  $H_0^*$  is rejected when  $\text{HS} > \chi_{n-1, \alpha}^2$ .

The powers of those test statistics are examined by simulation with 10,000 replication. For  $n = 50$ , we adjust the type-I errors before calculating their powers under the following three alternatives for  $w = 0.2, 0.5, 0.8, 1.0$ .

- $M_1 : X_i = (1-w)\text{Ex}(1) + w\text{Ga}(1.2, 0.8),$
- $M_2 : X_i = (1-w)\text{Ex}(1) + w\text{LogN}(0, 1),$
- $M_3 : X_i = (1-w)\text{Ex}(1) + w\text{InvG}(1, 1),$

where  $\text{Ex}(1)$ ,  $\text{Ga}(a, b)$ ,  $\text{LogN}(0, 1)$  and  $\text{InvG}(0, 1)$  denote random variables having exponential distribution  $\text{Ex}(1)$ , gamma distribution  $\text{Ga}(a, b)$ , log normal distribution  $\text{LogN}(0, 1)$  and inverse Gaussian distribution  $\text{InvG}(1, 1)$ , respectively. The values of their powers are reported in Table 5. From the table, it is observed that the tests  $\text{IST}_{0.1}$ ,  $\text{HME}_1$  and  $\text{CO}$  are more powerful for  $w = 0.2, 0.5$  and  $0.8$  in  $M_1, M_2$  and  $M_3$ . In the case of  $w = 1$ , the test  $\text{IST}_{0.1}$  is most powerful for  $\text{Ga}(1.2, 0.8)$ , the tests  $\text{IST}_{0.1}^+, \text{IST}_{0.1}^-, \text{MST}_{0.1}$  and  $\text{HS}$  are more powerful for  $\text{LN}(0, 1)$  and the tests  $\text{IST}_{0.1}, \text{HME}_1$  and  $\text{CO}$  are more powerful for  $\text{InvG}$ . Overall, the tests  $\text{IST}_{0.1}, \text{HME}_1$  and  $\text{CO}$  have similar performances and the tests  $\text{IST}_{0.1}^+, \text{IST}_{0.1}^-, \text{MST}_{0.1}$  and  $\text{HS}$  are similarly performed.



**Table 5** Power of the five tests for  $n = 50$  and  $w = 0.2, 0.5, 0.8, 1.0$

$w$	$M_1 : Ga(1.2, 0.8)$					$M_2 : LogN(0, 1)$				$M_3 : InvG(1, 1)$			
	0.0	0.2	0.5	0.8	1.0	0.2	0.5	0.8	1.0	0.2	0.5	0.8	1.0
$IST_{0.1}^+$	5.0	34.4	71.7	24.3	4.9	25.1	33.7	24.2	34.9	19.0	62.5	33.0	12.3
$IST_{0.1}^-$	5.0	50.5	84.8	38.8	8.1	39.4	44.5	24.8	32.3	28.4	75.2	43.3	12.9
$IST_{0.1}$	5.0	88.4	98.9	79.7	22.8	77.7	78.6	37.1	7.2	69.5	98.2	87.7	35.5
$MST_{0.1}$	5.0	32.2	69.2	22.4	4.5	24.0	32.8	24.2	35.1	16.7	59.0	30.2	11.9
$HME_1$	5.0	84.0	97.9	72.3	14.2	76.3	85.0	49.2	18.3	64.6	98.5	93.1	40.7
HS	5.0	53.3	86.1	41.3	9.9	41.9	46.0	26.2	32.3	31.1	76.3	44.5	14.0
CO	5.0	86.1	98.4	76.0	14.3	79.7	80.7	43.4	14.6	70.2	98.9	90.6	41.9

## 6 Stein-type identities in Poisson and negative binomial distributions

### 6.1 Stein-type identity in Poisson distributions

In the Poisson distribution  $Po(\lambda)$ , Hudson (1978) provided the Stein-type identity, which is also characterizes the Poisson distribution as seen below.

**Theorem 6.1** *Let  $X$  be a non-negative and discrete random variable with  $E[X] = \lambda$ . Then, the following four conditions are equivalent.*

- (a)  $X \sim Po(\lambda)$ .
- (b) For any function  $h(\cdot)$  with  $E[|Xh(X)|] < \infty$ , it holds that

$$E[\lambda h(X)] = E[Xh(X - 1)] \text{ or } E[Xh(X)] = E[\lambda h(X + 1)] \tag{6.1}$$

- (c) For any real constant  $t$ , it holds that

$$E[(X - \lambda e^t)e^{tX}] = 0. \tag{6.2}$$

- (d)  $g(t) = E[\exp\{tX\}]$  satisfies the differential equation

$$\frac{d}{dt} \log\{g(t)\} = \lambda e^t. \tag{6.3}$$

**Proof** For the proof from (a) to (b), it is noted that

$$\lambda \frac{\lambda^x}{x!} e^{-\lambda} = (x + 1) \frac{\lambda^{x+1}}{(x + 1)!} e^{-\lambda},$$

which produces the identity (6.1). Clearly, one gets (c) from (b). For the proof from (c) to (d), the identity (6.2) is written as  $g'(t) = \lambda e^t g(t)$ , which is given in (6.3). For the proof from (d) to (a), the solution of the differential equation in (6.3) is  $g(t) = \exp\{\lambda(e^t - 1)\}$ , which implies that  $X \sim Po(\lambda)$ . □

The characterization of the Poisson distribution has been studied in a lot of papers. A feature of this distribution is that the sample mean and the unbiased sample variance have the same expectation. Shanbhag (1970b) derived the related condition for characterizing the Poisson distribution.

**Theorem 6.2** *Assume that nonnegative and discrete random variables  $X_1$  and  $X_2$  are independently and identically distributed. Then, the following three conditions are equivalent.*

- (a)  $X_i \sim Po(\lambda)$  for  $i = 1, 2$ .
- (b)  $X_1 + X_2 \sim Po(2\lambda)$ .
- (c) *The conditional expectation of  $(X_1 - X_2)^2$  given  $X_1 + X_2$  is equal to  $X_1 + X_2$ , namely  $E[(X_1 - X_2)^2|X_1 + X_2] = X_1 + X_2$ .*

**Proof** The proof from (a) to (b) is trivial. For the proof from (b) to (a), from Theorem 6.1, we have

$$E[(X_1 + X_2 - 2\lambda e^t)e^{t(X_1+X_2)}] = 0,$$

which easily leads to

$$2E[(X_1 - \lambda e^t)e^{tX_1}]E[e^{tX_2}] = 0.$$

Thus, we get (a) by using Theorem 6.1 again.

For the proof from (a) to (c), it is noted that  $E[(X_1 - X_2)^2] = 2\lambda$  and  $E[X_1 + X_2] = 2\lambda$ , namely

$$E[(X_1 - X_2)^2 - (X_1 + X_2)] = 0.$$

Since  $X_1 + X_2$  is complete and sufficient, from  $E[E[(X_1 - X_2)^2|X_1 + X_2] - (X_1 + X_2)] = 0$ , it follows that  $E[(X_1 - X_2)^2|X_1 + X_2] - (X_1 + X_2) = 0$ , and we get (c).

For the proof from (c) to (a), it is noted that condition (c) implies

$$E[(X_1 - X_2)^2 e^{t(X_1+X_2)}] = E[(X_1 + X_2)e^{t(X_1+X_2)}]. \tag{6.4}$$

It is observed that

$$\begin{aligned} E[(X_1 - X_2)^2 e^{t(X_1+X_2)}] &= E[(X_1^2 + X_2^2 - 2X_1X_2)e^{t(X_1+X_2)}] \\ &= 2E[X_1^2 e^{tX_1}]E[e^{tX_2}] - 2E[X_1 e^{tX_1}]E[X_2 e^{tX_2}], \\ E[(X_1 + X_2)e^{t(X_1+X_2)}] &= 2E[X_1 e^{tX_1}]E[e^{tX_2}]. \end{aligned}$$

Then from (6.4), for  $g(t) = E[e^{tX_1}]$ , we have

$$g''(t)g(t) - \{g'(t)\}^2 = g'(t)g(t) \quad \text{or} \quad \frac{g''(t)g(t) - \{g'(t)\}^2}{\{g(t)\}^2} = \frac{g'(t)}{g(t)}.$$

Let  $\psi(t) = g'(t)/g(t)$ , and this equality is expressed as  $\psi'(t) = \psi(t)$ . The solution of this differential equation is  $\log \psi(t) = t + \log c_0$ , namely  $\psi(t) = c_0 e^t$ . This implies that  $\log g(t) = c_0 e^t + \log c_1$ , or  $g(t) = c_1 \exp\{c_0 e^t\}$ . Since  $g(0) = c_1 e^{c_0} = 1$  or  $c_1 = e^{-c_0}$ , we get  $g(t) = \exp\{c_0(e^t - 1)\}$ , which leads to the Poisson distribution.  $\square$

Theorem 6.2 can be extended to the case of a random sample with size  $n$ , where condition (c) is replaced by  $E[(n - 1)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2 | \bar{X}] = \bar{X}$ .

### 6.2 Two applications of the Stein-type identity in Poisson distributions

We here provide two applications of the Stein-type identity in Poisson distributions. One of them is to obtain improved shrinkage estimators in simultaneous estimation in Poisson distributions. Let  $X_1, \dots, X_p$  be independent random variables such that  $X_i \sim Po(\lambda_i)$ ,  $i = 1, \dots, p$ . Consider the problem of simultaneously estimating  $\lambda = (\lambda_1, \dots, \lambda_p)$  relative to the loss  $\sum_{i=1}^p (\hat{\lambda}_i - \lambda_i)^2 / \lambda_i$ . Clevenson and Zidek (1975) constructed a class of estimators improving on  $X$ , given by  $\hat{\lambda}_\phi = (\hat{\lambda}_{\phi,1}, \dots, \hat{\lambda}_{\phi,p})$  for

$$\hat{\lambda}_{\phi,i} = X_i - \frac{\phi(Z)}{Z + p - 1} X_i, \quad Z = \sum_{j=1}^p X_j. \tag{6.5}$$

**Theorem 6.3** *The unbiased risk estimator of the shrinkage estimator  $\hat{\lambda}_\phi$  is*

$$R(\lambda, \hat{\lambda}_\phi) = p - \frac{2Z}{Z + p - 1} \{\phi(Z + 1) - \phi(Z)\} - \frac{2(p - 1)\phi(Z + 1)}{Z + p - 1} + \frac{\phi^2(Z + 1)}{Z + p}.$$

Thus, for  $p \geq 2$ ,  $\hat{\lambda}_\phi$  improves on  $X$  if  $\phi(\cdot)$  satisfies the conditions (a)  $\phi(z)$  is nondecreasing in  $z$ , and (b)  $0 \leq \phi(z) \leq 2(p - 1)$ .

**Proof** The risk function of  $\hat{\lambda}_\phi$  is

$$R(\lambda, \hat{\lambda}_\phi) = p - 2 \sum_{i=1}^p E \left[ \left( \frac{X_i}{\lambda_i} - 1 \right) \frac{X_i \phi(Z)}{Z + p - 1} \right] + \sum_{i=1}^p E \left[ \frac{X_i^2}{\lambda_i} \frac{\phi^2(Z)}{(Z + p - 1)^2} \right].$$

From the Stein identity in Theorem 6.1, we use the identity  $E[(X_i/\lambda_i)h(X_i)] = E[h(X_i + 1)]$  to write

$$\begin{aligned} \sum_{i=1}^p E \left[ \frac{X_i}{\lambda_i} \frac{X_i \phi(Z)}{Z + p - 1} \right] &= \sum_{i=1}^p E \left[ \frac{(X_i + 1)\phi(Z + 1)}{Z + p} \right] = E[\phi(Z + 1)], \\ \sum_{i=1}^p E \left[ \frac{X_i^2}{\lambda_i} \frac{\phi^2(Z)}{(Z + p - 1)^2} \right] &= \sum_{i=1}^p E \left[ \frac{(X_i + 1)\phi^2(Z + 1)}{(Z + p)^2} \right] = E \left[ \frac{\phi^2(Z + 1)}{Z + p} \right]. \end{aligned}$$

These observations give the expression in the unbiased risk estimator.  $\square$

In the case of  $\phi(z) = p - 1$ , the estimator  $\widehat{\lambda}^S = X - (p - 1)(Z + p - 1)^{-1}X$  has the risk unbiased estimator  $\widehat{R}(\lambda, \widehat{\lambda}^S) = p - (p - 1)^2(Z + p + 1)/\{(Z + p - 1)(Z + p)\}$ . Clevenson and Zidek (1975) showed that the Bayes estimator  $\widehat{\lambda}^{GB} = X - (p - 1 + \beta)(Z + p - 1 + \beta)^{-1}X$  satisfies the conditions of Theorem 6.3 for  $0 \leq \beta \leq p - 1$  and it is admissible for  $\beta > 1$ . Numerical investigation of the estimator  $\widehat{\lambda}^{CZ}$  is given in Table 7.

We next consider to derive goodness-of fit test statistics for Poissonity based on the Stein identity. The problem of testing Poissonity has been studied in the literature, and one can see Mijburgh and Visagie (2020) for an overview.

Let  $X_1, \dots, X_n$  be a discrete and nonnegative random sample from a population with distribution function  $F(\cdot)$  with mean  $E[X_i] = \lambda$ . Consider the problem of testing the Poissonity of the underlying distribution  $H_0 : F = Po(\lambda)$ . From Theorem 6.1, the characterization of the Poisson distribution is

$$E[(X - \lambda e^t)e^{tX}] = 0,$$

and the sample counterpart is  $w_t/\sqrt{n}$ , where

$$w_t = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \bar{X}e^t)e^{tX_i}. \tag{6.6}$$

The idea of Henze et al. (2012) is used to construct the test statistic

$$BHT_\gamma = \int_{-\infty}^0 w_t^2 e^{\gamma t} dt = \frac{1}{n1} \sum_{i=1}^n \sum_{j=1}^n \left\{ \frac{B_{ij}}{A_{ij} + \gamma} - \frac{\bar{X}A_{ij}}{A_{ij} + 1_\gamma} + \frac{\bar{X}^2}{A_{ij} + 2 + \gamma} \right\}$$

for  $A_{ij} = X_i + X_j$ ,  $B_{ij} = A_i X_j$  and positive constant  $\gamma$ . This was suggested by Treutler (1995) and the related test was proposed by Baringhaus and Henze (1992). For other test statistics, see Gurtler and Henze (2000). Similarly to the problem of testing exponentiality, we can consider the test statistics

$$IST_c = \int_{-c}^0 w_t^2 e^t dt = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left\{ \frac{B_{ij}}{A_{ij} + 1} (1 - e^{-cA_{ij}}) - \frac{\bar{X}A_{ij}}{A_{ij} + 2} (1 - e^{-cA_{ij}-2c}) + \frac{\bar{X}^2}{A_{ij} + 3} (1 - e^{-cA_{ij}-3c}) \right\},$$

and  $MST_c = \sup_{-c < t < 0} |w_t|$  for positive constant  $c$ .

It is observed that  $w_t$  can be approximated as

$$w_t = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \lambda e^t)e^{tX_i} - (\bar{X} - \lambda e^t) \frac{1}{\sqrt{n}} \sum_{i=1}^n e^{tX_i}$$

$$= \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n (X_i - \lambda e^t) e^{tX_i} - h_0(t) \right\} + \sqrt{nh_0(t)} + \sqrt{n}(\bar{X} - \lambda) e^t g(t) + o_p(1),$$

where  $h_0(t) = E[(X_1 - \lambda e^t) e^{tX_1}]$  and  $g(t) = E[e^{tX_1}]$ . This gives the following lemma.

**Lemma 6.1**  $w_t$  is approximated as  $w_t = W_n(t) + \sqrt{nh_0(t)} + o_p(1)$ , where

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ (X_i - \lambda e^t) e^{tX_i} - h_0(t) - (X_i - \lambda) e^t g(t) \right\}. \tag{6.7}$$

From the central limit theorem, it follows that  $W_n(t)$  converges in distribution to the normal distribution with mean zero and the variance

$$\begin{aligned} \text{Var}(W_n(t)) &= E[(X_1 - \lambda e^t)^2 e^{2tX_1}] - h_0(t)^2 + (E[X_1^2] - \lambda^2) e^{2t} g(t)^2 \\ &\quad - 2\{E[X_1(X_1 - \lambda e^t) e^{tX_1}] - \lambda h_0(t)\} e^t g(t) + o(1). \end{aligned}$$

Since  $h_0(t) = g'(t) - \lambda e^t g(t)$ , we have  $h'_0(t) = g''(t) - \lambda e^t g'(t) - \lambda e^t g(t)$ , so that  $E[X_1(X_1 - \lambda e^t) e^{tX_1}] = g''(t) - \lambda e^t g'(t) = h'_0(t) + \lambda e^t g(t)$ . Under the null hypothesis  $H_0$  of Poissonity, it can be seen that  $g(t) = \exp\{\lambda(e^t - 1)\}$ ,  $h_0(t) = 0$  and

$$E[(X_1 - \lambda e^t)^2 e^{2tX_1}] = \lambda\{\lambda(e^t - 1)^2 + 1\} e^{2t} \exp\{\lambda(e^{2t} - 1)\},$$

so that  $\text{Var}(W_n(t)) = V(t)^2 + o(1)$ , where

$$V(t)^2 = \lambda e^{2t} \left[ \{\lambda(e^t - 1)^2 + 1\} \exp\{\lambda(e^{2t} - 1)\} - \exp\{2\lambda(e^t - 1)\} \right].$$

Using Lemma 6.1 and the same arguments as in the proof of Theorem 3.3, we can verify the consistency of the suggested test statistics.

**Theorem 6.4** Assume that  $E[X_1^2 e^{tX_1}] < \infty$  for  $t$  around zero. Then, the test statistics  $BHT_\gamma$ ,  $IST_c$  and  $MST_c$  given below (6.6) are consistent.

We investigate the performances of powers of the test statistics  $BHT_\gamma$  for  $\gamma = 1$ ,  $IST_c$  for  $c = 1$  and  $MST_c$  for  $c = 1$ . As a competitor, we employ the test based on Fisher’s index, given by  $FI = (n - 1)S^2/\bar{X}$ . The null hypothesis  $H_0$  is rejected when  $FI > \chi^2_{n-1, 1-\alpha/2}$  or  $FI > \chi^2_{n-1, \alpha/2}$ . The simulation experiments are conducted under the following two alternatives for  $\alpha = 0, 2, 3, 4, 5$  and  $\lambda = 1, 3$ .

- $M_1 : X_i = Po(\lambda) + \alpha Nbin(10, 10/10.2),$
- $M_2 : X_i = Po(\lambda) + \alpha Po(0.2),$
- $M_3 : X_i = Po(\lambda) + \alpha Bin(10, 0.1),$

where  $Po(\lambda)$ ,  $Po(0.2)$ ,  $Nbin(10, 10/10.2)$  and  $Bin(10, 0.1)$  denote independent random variables having Poisson distributions  $Po(\lambda)$ ,  $Po(1.5)$ , negative binomial

**Table 6** Sizes and powers of the four tests for  $n = 50$  and  $\alpha = 0, 2, 3, 4, 5$

$\alpha$	$M_1 (\lambda = 1)$					$M_1 (\lambda = 3)$				
	0	2	3	4	5	0	2	3	4	5
BHT <sub>1</sub>	3.7	38.8	83.2	97.2	99.5	4.9	9.1	24.9	52.8	79.3
IST <sub>1</sub>	4.1	39.8	85.2	97.7	99.7	4.8	9.4	26.1	54.3	80.3
MST <sub>1</sub>	3.9	40.4	85.9	98.1	99.8	4.3	10.5	30.5	64.7	85.6
FI	4.8	27.0	77.5	97.8	99.9	3.9	9.0	30.8	67.9	88.2
$\alpha$	$M_2 (\lambda = 1)$					$M_2 (\lambda = 3)$				
	0	2	3	4	5	0	2	3	4	5
BHT <sub>1</sub>	5.4	36.5	82.1	97.0	99.7	4.9	9.6	27.2	56.7	79.9
IST <sub>1</sub>	5.1	38.1	83.8	97.5	99.7	5.1	9.9	27.8	58.4	81.2
MST <sub>1</sub>	5.4	38.4	84.6	97.9	99.7	5.2	9.9	32.2	64.6	87.3
FI	5.6	27.3	77.6	97.4	99.9	5.1	9.3	31.8	69.5	89.6
$\alpha$	$M_3 (\lambda = 1)$					$M_3 (\lambda = 3)$				
	0	2	3	4	5	0	2	3	4	5
BHT <sub>1</sub>	3.7	86.5	99.9	100	100	4.9	32.2	87.7	99.3	100
IST <sub>1</sub>	4.1	87.9	99.9	100	100	4.8	33.5	88.7	99.3	100
MST <sub>1</sub>	3.9	88.4	99.9	100	100	4.3	41.0	92.7	100	100
FI	4.8	59.5	98.5	100	100	3.9	29.5	87.3	99.6	100

distribution  $Nbin(10, 10/10.2)$  and binomial distribution  $Bin(10, 0.1)$ , respectively. It is here noted that the type-I errors of the tests depend on the unknown parameter  $\lambda$ . To fix this problem, the parametric bootstrap method is useful for critical values of the tests. For example, we explain how to obtain the critical value of  $BHT_\gamma$  in model  $M_2$ . We first generate  $K$  samples of size  $n$  from  $M_2$  and each sample consists of  $u_1, \dots, u_n$  from  $Po(\lambda)$  and  $v_1, \dots, v_n$  from  $Po(0.2)$  for  $K = 1,000$  and  $n = 50$ . Then, we generate the  $B$  bootstrap samples  $(u_1^{(b)}, \dots, u_n^{(b)})$ ,  $b = 1, \dots, B$  from  $Po(\bar{u})$  for  $\bar{u} = \sum_{i=1}^n u_i/n$  and calculate the value of  $BHT_\gamma^{(b)}$  for  $b = 1, \dots, B$  and  $B = 1,000$ . We obtain the critical value  $q_{BHT}$  of the test  $BHT_\gamma$  as an upper 5% quantile of the histogram of  $BHT_\gamma^{(b)}$ 's. We then calculate  $K$  values of  $BHT_\gamma$  based on  $x_i = u_i + \alpha v_i$  for  $i = 1, \dots, n$  and count the number such that  $BHT_\gamma > q_{BHT}$ . In this way, we can obtain approximated values of the size and the power of the test  $BHT_\gamma$ . These values are reported in Table 6.

From the table, the sizes of the tests have small variations for  $n = 50$ ,  $B = 1,000$  and  $K = 1,000$  and the variations will lower for large  $n$ ,  $B$  and  $K$ . The tests  $BHT_1$ ,  $IST_1$  and  $MST_1$  have similar performances in the powers and are more powerful than  $FI$  in the case of  $\lambda = 1$  for  $M_1$ ,  $M_2$  and  $M_3$ . In the case of  $\lambda = 3$ , the test  $MST_1$  is more powerful than  $BHT_1$  and  $IST_1$  and  $FI$  is more powerful for  $\alpha = 4, 5$  in  $M_1$  and  $M_2$ .

### 6.3 Stein-type identity in negative binomial distributions

Consider the negative binomial distribution  $Nbin(\alpha, p)$  with the probability function

$$f(x|\alpha, p) = \frac{\Gamma(x + \alpha)}{\Gamma(\alpha)x!} q^x p^\alpha, \quad q = 1 - p,$$

where  $\alpha > 0$  and  $0 < p < 1$ . Although  $\alpha$  is a natural number in the negative binomial distribution, we here treat  $\alpha$  as a positive real number. Hudson (1978) provided the Stein-type identity, which is also characterizes the negative binomial distribution as seen below.

**Theorem 6.5** *Let  $X$  be a non-negative and discrete variable. Then, the following four conditions are equivalent.*

- (a)  $X \sim Nbin(\alpha, p)$ .
- (b) For any function  $h(\cdot)$  with  $E[|Xh(X)|] < \infty$ , it holds that

$$E[Xh(X)] = qE[(X + \alpha)h(X + 1)]. \tag{6.8}$$

- (c) For any real constant  $t$ , it holds that

$$E[Xe^{tX}] = qE[(X + \alpha)e^{t(X+1)}]. \tag{6.9}$$

- (d)  $g(t) = E[\exp\{tX\}]$  satisfies the differential equation

$$\frac{d}{dt} \log\{g(t)\} = \frac{\alpha q e^t}{1 - q e^t}. \tag{6.10}$$

**Proof** For the proof from (a) to (b), it is noted that

$$x \frac{\Gamma(x + \alpha)}{\Gamma(\alpha)x!} q^x p^\alpha = q(x - 1 + \alpha) \frac{\Gamma(x - 1 + \alpha)}{\Gamma(\alpha)(x - 1)!} q^{x-1} p^\alpha,$$

which shows the identity (6.8). Clearly, one gets (c) from (b). For the proof from (c) to (d), the identity (6.9) is written as  $g'(t) = q\{g'(t) + \alpha g(t)\}e^t$ , which is (6.10). For the proof from (d) to (a), the solution of the differential equation in (6.10) is  $\log g(t) = \alpha \log p - \alpha \log(1 - qe^t)$ , namely,  $g(t) = p^\alpha / (1 - qe^t)^\alpha$ , which implies that  $X \sim Nbin(\alpha, p)$ . □

**Theorem 6.6** *Assume that nonnegative and discrete random variables  $X_1$  and  $X_2$  are independently and identically distributed with  $E[|X|e^{tX}] < \infty$ . The following two are equivalent.*

- (a)  $X_i \sim Nbin(\alpha, p)$  for  $i = 1, 2$ .
- (b)  $X_1 + X_2 \sim Nbin(2\alpha, p)$ .

**Proof** It is easy to see (b) from (a). For the proof from (b) to (a), from Theorem 6.5, it follows that

$$E[(X_1 + X_2)e^{t(X_1+X_2)}] = qE[(X_1 + X_2 + 2\alpha)e^{t(X_1+X_2)}].$$

This equality leads to

$$2E[X_1e^{tX_1}]E[e^{tX_2}] = 2qE[(X_1 + \alpha)e^{tX_1}]E[e^{tX_2}],$$

which, from Theorem 6.5, shows (a). □

We briefly describe the Stein problem of simultaneous estimation of means of  $k$  negative binomial distributions. This is due to Tsui (1984). Let  $X_1, \dots, X_p$  be independent random variables such that  $X_i \sim Nbin(\alpha_i, \eta_i)$  for  $i = 1, \dots, p$ . The mean of  $X_i$  is denoted by  $\theta_i = \alpha_i(1 - \eta_i)/\eta_i$ . Let  $\mathbf{X} = (X_1, \dots, X_p)$  and  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$ , and we consider the estimation of  $\boldsymbol{\theta}$  for known  $\alpha_i$ 's relative to the loss  $\sum_{i=1}^p (\hat{\theta}_i - \theta_i)^2/\theta_i$ . Tsui (1984) suggested the shrinkage estimator  $\hat{\boldsymbol{\theta}}_\phi = (\hat{\theta}_{\phi,1}, \dots, \hat{\theta}_{\phi,p})$ , where

$$\hat{\theta}_{\phi,i} = X_i - \frac{\phi(Z)}{Z + p - 1} X_i,$$

for  $Z = \sum_{i=1}^p X_i$  and nonnegative function  $\phi(\cdot)$ . Let  $\Delta = R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_\phi) - R(\boldsymbol{\theta}, \mathbf{X})$ . The conditions on  $\phi(\cdot)$  for improving on  $\mathbf{X}$  was derived by Tsui (1984).

**Theorem 6.7** *The risk difference  $\Delta$  is decomposed as  $\Delta = \Delta_1 + \Delta_2$ , where*

$$\begin{aligned} \Delta_1 &= E\left[\frac{(Z + p)\phi^2(Z + 1)}{(Z + p)^2} - 2\frac{(Z + p)\phi(Z + 1)}{Z + p} + 2\frac{Z\phi(Z)}{Z + p - 1}\right], \\ \Delta_2 &= \sum_{i=1}^p E\left[\frac{X_i}{\alpha_i} \left\{ \frac{(X_i + 1)\phi(Z + 1)}{Z + p} \left(\frac{\phi(Z + 1)}{Z + p} - 2\right) + \frac{X_i\phi(Z)}{Z + p - 1} \left(2 - \frac{\phi(Z)}{Z + p - 1}\right) \right\}\right]. \end{aligned}$$

Then, it holds that  $\Delta \leq 0$  if the following conditions are satisfied. (a)  $\phi(z)$  is nondecreasing, (b)  $0 < \phi(z) \leq 2(p - 1)$  and (c)  $\phi(z)/z$  is nonincreasing.

**Proof** More generally, we consider the estimator  $\hat{\theta}_i = X_i + f_i(\mathbf{X})$ . Let  $\mathbf{e}_i$  be a  $p$ -variate vector whose  $i$ -th coordinate is one and whose other coordinates are zero. Note that  $\alpha_i/\theta_i = 1/q_i - 1$  and  $E[h(X_i)/q_i] = E[(X_i + \alpha_i)h(X_i + 1)/(X_i + 1)]$  from Theorem 6.5. Then,

$$\begin{aligned} \Delta &= R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) - R(\boldsymbol{\theta}, \mathbf{X}) = \sum_{i=1}^p E[f_i^2(\mathbf{X})/\theta_i + 2X_i f_i(\mathbf{X})/\theta_i - 2f_i(\mathbf{X})] \\ &= \sum_{i=1}^p E\left[-\frac{f_i^2(\mathbf{X})}{\alpha_i} - 2\frac{X_i f_i(\mathbf{X})}{\alpha_i} - 2f_i(\mathbf{X}) + \frac{f_i^2(\mathbf{X})}{\alpha_i q_i} + \frac{2X_i f_i(\mathbf{X})}{\alpha_i q_i}\right] \end{aligned}$$



**Table 7** Risks of  $X$  and  $\hat{\theta}^{CZ}$  for  $p = 6$  and  $\theta_{ik} = k/2$

$k$	Poisson distribution $Po(\theta)$									
	1	2	3	4	5	6	7	8	9	10
$X$	5.98	6.01	6.00	5.99	5.99	6.01	5.98	5.98	6.00	6.00
$\hat{\theta}^{CZ}$	2.33	3.38	3.98	4.36	4.63	4.83	4.96	5.07	5.17	5.24
$k$	Geometric distribution $Geo(\theta)$									
	1	2	3	4	5	6	7	8	9	10
$X$	9.02	12.05	14.89	18.06	21.03	23.94	26.92	29.83	32.94	35.96
$\hat{\theta}^{CZ}$	3.32	6.21	8.99	11.96	14.84	17.68	20.58	23.44	26.46	29.42
$k$	Negative binomial distribution $Nbin(5, \theta)$									
	1	2	3	4	5	6	7	8	9	10
$X$	6.60	7.17	7.80	8.41	8.99	9.58	10.17	10.80	11.42	12.00
$\hat{\theta}^{CZ}$	2.51	3.89	4.93	5.80	6.56	7.28	7.98	8.66	9.32	9.97

$$\begin{aligned}
 &= \sum_{i=1}^p E \left[ -\frac{f_i^2(X)}{\alpha_i} - 2\frac{X_i f_i(X)}{\alpha_i} - 2f_i(X) \right. \\
 &\quad \left. + \frac{(X_i + \alpha) f_i^2(X + e_i)}{\alpha_i(X_i + 1)} + \frac{2(X_i + \alpha_i) f_i(X + e_i)}{\alpha_i} \right] \\
 &= \Delta_1 + \Delta_2,
 \end{aligned}$$

where

$$\begin{aligned}
 \Delta_1 &= \sum_{i=1}^p E \left[ \frac{X_i f_i^2(X + e_i)}{X_i + 1} + 2f_i(X + e_i) - 2f_i(X) \right], \\
 \Delta_2 &= \sum_{i=1}^p E \left[ \frac{X_i f_i^2(X + e_i)}{\alpha_i(X_i + 1)} - \frac{f_i^2(X)}{\alpha_i} + 2\frac{X_i f_i(X + e_i)}{\alpha_i} - 2\frac{X_i f_i(x)}{\alpha_i} \right].
 \end{aligned}$$

Substituting  $f_i(X) = -X_i \phi(Z)/(Z + p - 1)$  yields the expressions in Theorem 6.7. It can be easily checked that  $\Delta_1 \leq 0$  and  $\Delta_2 \leq 0$  under the conditions (a), (b) and (c). □

We investigate the risk performances of the estimators  $X$  and  $\hat{\theta}^{CZ}$  under the three distributions of  $X_i$ : the Poisson distribution  $Po(\theta_i)$ , geometric distribution  $Geo(1/(\theta_i + 1))$  and negative binomial distribution  $Nbin(5, 5/(\theta_i + 5))$ . The simulation experiment has been conducted with  $p = 6$  and  $\theta_i = k/2$  for  $i = 1, \dots, p$  and  $k = 1, \dots, 10$ , and the average losses based on simulation with 10,000 replications are reported Table 7. From Table 7, it is seen that the improvements of  $\hat{\theta}^{CZ}$  and  $\hat{\theta}^{TS}$  over  $X$  are significant and robust in the Poisson, geometric and negative binomial distributions.

## 7 Concluding remarks

We conclude the paper with some remarks and extensions. Through the paper, we have used the moment-generating functions for characterizing and testing the normal, exponential and Poisson distributions. As well known, however, the use of the moment generating functions is limited to the existence of their expectations. To avoid this point, it may be better to use the characteristic functions, and the results given in the paper can be extended to the arguments based on the characteristic functions.

The test of normality has been treated in Sect. 3.2 in the uni-variate case. For testing multivariate normal distributions, Ebner (2021) provided a test statistic based on the Stein method. Many other statistics for testing multivariate-normality have been suggested in the literature. For example, see Mardia (1970), Mecklin and Mundfrom (2004) and Ebner and Henze (2020).

Although the Stein method in this paper is limited to normal distributions, Ley and Swan (2013) suggested the general density approach which is applicable to general distributions. Barbour (1988) and Götze (1991) introduced the generator approach which can adapt the method to many other distributions. For the details, see Reinert (2005).

Finally, we give some recent developments related to the Stein method. Betsch et al. (2021) suggested new techniques based on the Stein method for estimating parameters. In addition, Betsch et al. (2022) discussed the estimation of parameters in negative-binomial distributions and the test of Poissonity based on the Stein method. Betsch and Ebner (2021) provided characterization of continuous distributions based on indicator functions, and Betsch et al. (2022) applied a similar argument to discrete distributions.

**Acknowledgements** I would like to thank the Editor, the Associate Editor and the reviewer for many valuable comments and helpful suggestions which led to an improved version of this paper. I am also grateful to Dr. Ryo Imai for his valuable comments. This research was supported in part by Grant-in-Aid for Scientific Research (18K11188, 22K11928) from Japan Society for the Promotion of Science.

**Funding** Open Access funding provided by The University of Tokyo.

## Declarations

**Conflict of interest** The author declares that he has no conflict of interest.

**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., et al. (2023). Stein's method meets computational statistics: A review of some recent developments. *Statistical Science*, 38, 120–139.
- Baranchik, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Annals of Mathematical Statistics*, 41, 642–645.
- Barbour, A. D. (1988). Stein's method and Poisson process convergence. *Journal of Applied Probability*, 25, 175–184.
- Barbour, A. D., & Chen, L. H. Y. (2005). *Stein's method and applications*. World Scientific.
- Baringhaus, L., & Henze, N. (1992). A goodness of fit test for the Poisson distribution based on the empirical generating function. *Statistics and Probability Letters*, 13, 269–274.
- Bellec, P. C., & Zhang, C.-H. (2021). Second-order Stein: URE for SURE and other applications in high-dimensional inference. *Annals of Statistics*, 49, 1864–1903.
- Berger, J. O. (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma problem. *Annals of Statistics*, 8, 545–571.
- Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer.
- Bernstein, S. N. (1941). Sur une propriété caractéristique de la loi de Gauss. *Leningrad. Polytechnic Institute. Transactions*, 3, 21–22.
- Betsch, S., & Ebner, B. (2019). A new characterization of the gamma distribution and associated goodness-of-fit tests. *Metrika*, 82, 779–806.
- Betsch, S., & Ebner, B. (2020). Testing normality via a distributional fixed point property in the Stein characterization. *TEST*, 29, 105–138.
- Betsch, S., & Ebner, B. (2021). Fixed point characterizations of continuous univariate probability distributions and their applications. *Annals of the Institute of Statistical Mathematics*, 73, 31–59.
- Betsch, S., Ebner, B., & Klar, B. (2021). Minimum  $L^q$ -distance estimators for non-normalized parametric models. *Canadian Journal of Statistics*, 49, 514–548.
- Betsch, S., Ebner, B., & Nestmann, F. (2022). Characterizations of non-normalized discrete probability distributions and their application in statistics. *Electronic Journal of Statistics*, 16, 1303–1329.
- Brandwein, A. C., & Strawderman, W. E. (1990). Stein estimation: The spherically symmetric case. *Statistical Science*, 5, 356–369.
- Chen, H. Y. (2021). Stein's method of normal approximation: Some recollections and reflections. *Annals of Statistics*, 49, 1850–1863.
- Chen, L. H. Y., Fang, X., & Shao, Q.-M. (2013). From Stein identities to moderate deviations. *Annals of Statistics*, 41, 262–293.
- Chen, L. H. Y., Goldstein, L., & Shao, Q.-M. (2011). *Normal approximation by Stein's method*. Springer.
- Chen, L. H. Y., & Shao, Q.-M. (2001). A non-uniform Berry–Esseen bound via Stein's method. *Probability Theory and Related Fields*, 120, 236–254.
- Chen, L. H. Y., & Shao, Q.-M. (2005). Stein's method for normal approximation. In: A. D. Barbour & L. H. Y. Chen (Eds.), *An introduction to Stein's method*, Lecture Notes Series No. 4 (pp. 1–59), Institute for Mathematical Sciences, National University of Singapore, Singapore University Press and World Scientific.
- Clevenson, M. L., & Zidek, J. V. (1975). Simultaneous estimation of the means of independent Poisson laws. *Journal of the American Statistical Association*, 70, 698–705.
- Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. Chapman and Hall.
- Cramér, H. (1936). Über eine Eigenschaft der normalen verteilungsfunktion. *Mathematische Zeitschrift*, 41, 405–414.
- Das Gupta, A. (1986). Simultaneous estimation in the multiparameter gamma distribution under weighted quadratic losses. *Annals of Statistics*, 14, 206–219.
- De Wet, T., & Ventner, J. H. (1972). Asymptotic distributions of certain test criteria of normality. *South African Statistical Journal*, 6, 135–149.
- Diaconis, P., & Holmes, S. (2004). *Stein's method: Expository lectures and applications*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 46, IMS, Hayward, CA.
- Diaconis, P., Olkin, I., & Ghurye, S. G. (1977). Review of the book of Kagan, Linnik and Rao (1973). *Annals of Statistics*, 5, 583–592.
- Doornik, J. A. (2007). *Object-oriented matrix programming using Ox* (3rd ed.). Timberlake Consultants Press.

- Ebner, B. (2021). On combining the zero bias transform and the empirical characteristic function to test normality. *Latin American Journal of Probability and Mathematical Statistics*, 18, 1029–1045.
- Ebner, B., & Henze, N. (2020). Tests for multivariate normality—A critical review with emphasis on weighted  $L^2$ -statistics. *TEST*, 29, 845–892.
- Efron, B., & Morris, C. (1973). Stein's estimation rule and its competitors—An empirical Bayes approach. *Journal of the American Statistical Association*, 68, 117–130.
- Fathi, M., Goldstein, L., Reinert, G., & Saumard, A. (2022). Relaxing the Gaussian assumption in shrinkage and SURE in high dimension. *Annals of Statistics*, 50, 2737–2766.
- Fourdrinier, D., Strawderman, W. E., & Wells, M. T. (2018). *Shrinkage estimation*. Springer Nature.
- Ferguson, T. S. (1964). A characterization of the exponential distribution. *Annals of Mathematical Statistics*, 35, 1199–1207.
- Ghosh, M., Kubokawa, T., & Datta, G. S. (2020). Density prediction and the Stein phenomenon. *Sankhya*, 82–A, 330–352.
- Goldstein, L., & Reinert, G. (1997). *Annals of Applied Probability*, 7, 935–952.
- Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Annals of Probability*, 19, 724–739.
- Gürtler, N., & Henze, N. (2000). Recent and classical goodness-of-fit tests for the Poisson distribution. *Journal of Statistical Planning and Inference*, 90, 207–225.
- Hahn, G. J., & Shapiro, S. S. (1967). *Statistical models in engineering*. Wiley.
- Henze, N., & Meintanis, G. (2005). Recent and classical tests for exponentiality: A partial review with comparisons. *Metrika*, 61, 29–45.
- Henze, N., Meintanis, G., & Ebner, B. (2012). Goodness-of-fit tests for the gamma distribution based on the empirical Laplace transform. *Communications in Statistics-Theory and Methods*, 41, 1543–1556.
- Henze, N., & Visagie, J. (2020). Testing for normality in any dimension based on a partial differential equation involving the moment generating function. *Annals of the Institute of Statistical Mathematics*, 72, 1109–1136.
- Ho, S.-T., & Chen, L. H. (1978). An  $L_p$  bound for the remainder in a combinatorial central limit theorem. *Annals of Probability*, 6, 231–249.
- Hudson, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Annals of Statistics*, 6, 473–484.
- James, W., & Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 361–379). University of California Press, Berkeley.
- Kac, M. (1939). On a characterization of the normal distribution. *American Journal of Mathematics*, 61, 726–728.
- Kagan, A. M., Linnik, Y. V., & Rao, C. R. (1965). On a characterization of the normal law based on a property of the sample average. *Sankhya*, A27, 405–406.
- Kagan, A. M., Linnik, Y. V., & Rao, C. R. (1973). *Characterization problems in mathematical statistics*. Wiley.
- Khatri, C. G., & Rao, C. R. (1968). Some characterizations of the gamma distribution. *Sankhya*, A30, 157–166.
- Komaki, F. (2001). A shrinkage predictive distribution for multivariate normal observables. *Biometrika*, 88, 859–864.
- Konno, Y. (2009). Shrinkage estimators for large covariance matrices in multivariate real and complex normal distributions under an invariant quadratic loss. *Journal of Multivariate Analysis*, 100, 2237–2253.
- Kotz, S. (1974). Characterizations of statistical distributions: A supplement to recent surveys. *International Statistical Review*, 42, 39–65.
- Kubokawa, T. (1991). An approach to improving the James–Stein estimator. *Journal of Multivariate Analysis*, 36, 121–126.
- Kubokawa, T. (1994). A unified approach to improving equivariant estimators. *Annals of Statistics*, 22, 290–299.
- Lehmann, E. L., & Romano, J. P. (2022). *Testing statistical hypotheses* (4th ed.). Springer.
- Ley, C., & Swan, Y. (2013). Stein's density approach and information inequalities. *Electronic Communications in Probability*, 18, 1–14.
- Lukacs, E. (1942). A characterization of the normal distribution. *Annals of Mathematical Statistics*, 13, 91–93.

- Lukacs, E. (1955). A characterization of the gamma distribution. *Annals of Mathematical Statistics*, 26, 319–324.
- Madansky, A. (1988). *Prescriptions for working statisticians*. Springer.
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519–530.
- Maruyama, Y., Kubokawa, T., & Strawderman, W. E. (2023). *Stein estimation. Springer briefs in statistics*. Springer.
- Mecklin, C. J., & Mundfrom, D. J. (2004). An appraisal and Bibliography of tests for multivariate normality. *International Statistical Review*, 72, 123–138.
- Mijburgh, P. A., & Visagie, I. J. (2020). An overview of goodness-of-fit tests for the Poisson distribution. *South African Statistical Journal*, 54, 207–230.
- Ossai, E. O., Madukaife, M. S., & Oladugba, A. V. (2022). A review of tests for exponentiality with Monte Carlo comparisons. *Journal of Applied Statistics*, 49, 1277–1304.
- Reinert, G. (2005). Three general approaches to Stein's method. In A. D. Barbour & L. H. Y. Chen (Eds.), *An introduction to Stein's method*. Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore (Vol. 4). Singapore University Press.
- Robert, C. P. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation* (2nd ed.). Springer.
- Rubén, H. (1974). A new characterization of the normal distribution through the sample variance. *Sankhya*, A36, 379–388.
- Shanbhag, D. N. (1970). The characterizations for exponential and geometric distributions. *Journal of the American Statistical Association*, 65, 1256–1259.
- Shanbhag, D. N. (1970). Another characteristic property of the Poisson distribution. *Proceedings of the Cambridge Philosophical Society*, 68, 167–169.
- Shapiro, S. S., & Francia, R. S. (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67, 215–216.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality. *Biometrika*, 52, 591–611.
- Shinozaki, N. (1984). Simultaneous estimation of location parameters under quadratic loss. *Annals of Statistics*, 12, 322–335.
- Shorack, G. R. (2000). *Probability for statisticians*. Springer.
- Stein, C. (1973). Estimation of the mean of a multivariate normal distribution. In *Proceedings of the Prague symposium on asymptotic statistics* (pp. 345–381).
- Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, 9, 1135–1151.
- Stein, C. (1986). *Approximate computation of expectations*. Institute of Mathematics Lecture Notes – Monograph Series (Vol. 7). IMS, Hayward, CA.
- Strawderman, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Annals of Mathematical Statistics*, 42, 385–388.
- Thode, H. C. (2002). *Testing for normality*. Marcel Dekker.
- Treutler, B. (1995). Tests for the Poisson distribution. Diploma Thesis. University of Karlsruhe (in German).
- Tsui, K.-W. (1984). Robustness of Clevenson–Zidek-type estimators. *Journal of the American Statistical Association*, 79, 152–157.
- Tsukuma, H., & Kubokawa, T. (2020). *Shrinkage estimation for mean and covariance matrices. Springer briefs in statistics*. Springer.