On the Shrinkage Estimation of Variance and Pitman Closeness Criterion for Large Samples

Gérard Biau[1]
Université Pierre et Marie Curie[2] *& Ecole Normale Supérieure*[3]*, France*
gerard.biau@upmc.fr

Yannis G. Yatracos
Cyprus University of Technology, Cyprus[4]
yannis.yatracos@cut.ac.cy

Abstract

For a large class of distributions and large samples, it is shown that estimates of the variance σ^2 and of the standard deviation σ are more often Pitman closer to their target than the corresponding shrinkage estimates which improve the mean squared error. Our results indicate that Pitman closeness criterion, despite its controversial nature, should be regarded as a useful and complementary tool for the evaluation of estimates of σ^2 and of σ.

Index Terms — Variance estimation, standard deviation, shrinkage, Pitman closeness, mean squared error.

2010 Mathematics Subject Classification: 62F10, 62C15.

1 Introduction

Given two estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ of an unknown parameter θ, Pitman [13] has suggested in 1937 that $\hat{\theta}_1$ should be regarded as a "closer" estimate of θ if

$$\mathbb{P}\left(|\hat{\theta}_2 - \theta| > |\hat{\theta}_1 - \theta|\right) > 1/2.$$

---

[1]Corresponding author.

[2]Research partially supported by the French National Research Agency under grant ANR-09-BLAN-0051-02 "CLARA".

[3]Research carried out within the INRIA project "CLASSIC" hosted by Ecole Normale Supérieure and CNRS.

[4]Research supported by a start-up grant from Cyprus University of Technology.

This criterion, which is often called Pitman closeness, has an intuitive appeal and is in accordance with statistical tradition that preference should be expressed on a probability scale.

Much attention has been given to Pitman closeness criterion (PCC) properties in the 90's. It has been sharply criticized by some and vigorously defended by others on various counts. A good illustration of the debate is the paper by Robert, Hwang, and Strawderman [14] and the subsequent discussion by Blyth; Casella and Wells; Ghosh, Keating, and Sen; Peddada; and Rao [14], in which different views in PCC's favor or against it are presented.

Leaving the controversy behind, the object of this communication is to compare PCC with the familiar concept of mean squared error for variance estimation purposes. For a large class of distributions and large samples, it is shown herein that estimates of the variance σ^2 and of the standard deviation σ are more often "closer" to their target than the corresponding shrinkage estimates which improve the mean squared error. The same phenomenon is also observed for small and moderate sample sizes. Our results modestly indicate that PCC should be regarded as a useful and complementary tool for the evaluation of estimates of σ^2 and of σ, in agreement with Rao's comment [14]: "I believe that the performance of an estimator should be examined under different criteria to understand the nature of the estimator and possibly to provide information to the decision maker. I would include PCC in my list of criteria, except perhaps in the rare case where the customer has a definite loss function".

To go straight to the point, suppose that $X_1, \ldots, X_n$ $(n \geq 2)$ are independent, identically distributed (i.i.d.) real-valued random variables, with unknown mean and unknown finite positive variance σ^2. We consider here the estimation problem of the variance σ^2. Set

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

The sample variance estimate

$$S_{\text{SV},n}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}_n \right)^2$$

and the unbiased estimate

$$S_{\text{U},n}^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n \right)^2$$

2

are both standard statistical procedures to estimate σ^2. However, assuming squared error loss, more general estimates of the form

$$\delta_n \sum_{i=1}^{n} \left(X_i - \bar{X}_n \right)^2,$$

where $(\delta_n)_n$ is a positive sequence, are often preferred. For example, if $X_1, \ldots, X_n$ are sampled from a normal distribution, Goodman [3] proved that we can improve upon $S^2_{\text{SV},n}$ and $S^2_{\text{U},n}$ uniformly by taking $\delta_n = 1/(n+1)$. This means, setting

$$S^2_{\text{M},n} = \frac{1}{n+1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n \right)^2,$$

that for all n and all values of the parameter,

$$\mathbb{E} \left[S^2_{\text{M},n} - \sigma^2 \right]^2 < \mathbb{E} \left[S^2_{\text{SV},n} - \sigma^2 \right]^2 \quad \text{and} \quad \mathbb{E} \left[S^2_{\text{M},n} - \sigma^2 \right]^2 < \mathbb{E} \left[S^2_{\text{U},n} - \sigma^2 \right]^2.$$

To see this, it suffices to note that, in the normal setting,

$$\mathbb{E} \left[\delta_n \sum_{i=1}^{n} \left(X_i - \bar{X}_n \right)^2 - \sigma^2 \right]^2 = \sigma^4 \left[\left(n^2 - 1 \right) \delta_n^2 - 2(n-1)\delta_n + 1 \right], \quad (1.1)$$

and that the right-hand side is uniformly minimized by $\delta_n^\star = 1/(n+1)$ (Lehmann and Casella [11, Chapter 2]). Since the values $\delta_n = 1/n$ and $\delta_n = 1/(n-1)$, corresponding to $S^2_{\text{SV},n}$ and $S^2_{\text{U},n}$, respectively, lie on the same side of $1/(n+1)$, it is often referred to $S^2_{\text{M},n}$ as a shrinked version of $S^2_{\text{SV},n}$ and $S^2_{\text{U},n}$, respectively. Put differently, $S^2_{\text{M},n} = c_n S^2_{\text{SV},n}$ (respectively, $S^2_{\text{M},n} = \tilde{c}_n S^2_{\text{U},n}$) where, for each n, c_n (respectively, $\tilde{c}_n$) belongs to $(0, 1)$.

Under different models and assumptions, inadmissibility results in variance and standard deviation estimation were proved using such estimates, among others, by Goodman [3, 4], Stein [16], Brown [2], Arnold [1] and Rukhin [15]. For a review of the topic, we refer the reader to Maatta and Casella [12], who trace the history of the problem of estimating the variance based on a random sample from a normal distribution with unknown mean. More recently, Yatracos [18] provided shrinkage estimates of U-statistics based on artificially augmented samples and generalized, in particular, the variance shrinkage approach to non-normal populations by proving that, for all probability models with finite second moment, all values of σ^2 and all sample sizes $n \geq 2$, the estimate

$$S^2_{\text{Y},n} = \frac{n+2}{n(n+1)} \sum_{i=1}^{n} \left(X_i - \bar{X}_n \right)^2$$

3

ameliorates the mean squared error of $S_{\text{U},n}^2$. [Note that $S_{\text{Y},n}^2 = c_n S_{\text{U},n}^2$ for some $c_n \in (0,1)$, so that $S_{\text{Y},n}^2$ is in fact a shrinked version of $S_{\text{U},n}^2$.]

Nevertheless, the variance shrinkage approach, which is intended to improve the mean squared error of estimates, should be carefully considered when performing point estimation. The rationale behind this observation is that the mean squared error is the average of the parameter estimation error over all samples whereas, in practice, we use an estimate's value based on one sample only and we care for the distance from its target. To understand this remark, just consider the following example, due to Yatracos [19]. Suppose again that $X_1, \ldots, X_n$ are independently normally distributed, with finite variance σ^2. Then, an easy calculation reveals that

$$\mathbb{P}\left(\left|S_{\text{M},n}^2 - \sigma^2\right| > \left|S_{\text{U},n}^2 - \sigma^2\right|\right) = \mathbb{P}\left(\chi_{n-1}^2 < n - \frac{1}{n}\right), \qquad (1.2)$$

where χ_{n-1}^2 is a (central) chi-squared random variable with $n-1$ degrees of freedom (for a rigorous proof of this equality, see Lemma 4.1 in Section 4). Figure 1 depicts the values of probability (1.2) for sample sizes ranging from 2 to 200. It is seen on this example that the probability slowly decreases towards the value $1/2$, and that it may be significantly larger than $1/2$ for small and even for moderate values of n.
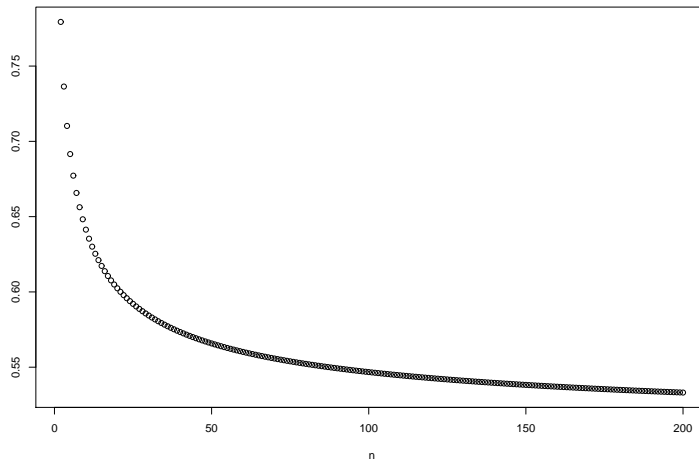

Figure 1: Plot of $\mathbb{P}(\chi_{n-1}^2 < n - 1/n)$ as a function of n, $n = 2, \ldots, 200$.

Thus, Figure 1 indicates that, for a normal population, the standard unbiased estimate $S_{\text{U},n}^2$ is Pitman closer to the target σ^2 than the shrinkage estimate

4

$S^2_{\mathrm{M},n}$, despite the fact that, for all n,

$$\mathbb{E}\left[S^2_{\mathrm{M},n} - \sigma^2\right]^2 < \mathbb{E}\left[S^2_{\mathrm{U},n} - \sigma^2\right]^2.$$

Moreover, the advantage of $S^2_{\mathrm{U},n}$ with this respect becomes prominent for smaller values of n, and a similar phenomenon may be observed by comparing the probability performance of $S^2_{\mathrm{SV},n}$ vs $S^2_{\mathrm{M},n}$. In fact, our main Theorem 2.1 reveals (in the particular case of normal distribution) that

$$\mathbb{P}\left(\left|S^2_{\mathrm{M},n} - \sigma^2\right| > \left|S^2_{\mathrm{U},n} - \sigma^2\right|\right) = \frac{1}{2} + \frac{5}{6\sqrt{\pi n}} + \mathrm{o}\left(\frac{1}{\sqrt{n}}\right)$$

and

$$\mathbb{P}\left(\left|S^2_{\mathrm{M},n} - \sigma^2\right| > \left|S^2_{\mathrm{SV},n} - \sigma^2\right|\right) = \frac{1}{2} + \frac{13}{12\sqrt{\pi n}} + \mathrm{o}\left(\frac{1}{\sqrt{n}}\right),$$

that is, $S^2_{\mathrm{U},n}$ and $S^2_{\mathrm{SV},n}$ are both asymptotically Pitman closer to σ^2 than $S^2_{\mathrm{M},n}$. It is therefore clear, at least on these Gaussian examples, that we should be cautious when choosing to shrink the variance for point estimation purposes. As we recently discovered, the examples in Yatracos [19] are similar to those studied in Khattree [7, 8, 9, 10].

In the present paper, we generalize this discussion to a large class of distributions. Taking a more general point of view, we let $X_1, \ldots, X_n$ be a sample drawn according to some unknown distribution with finite variance σ^2, and consider two candidates to estimate σ^2, namely

$$S^2_{1,n} = \alpha_n \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2 \quad \text{and} \quad S^2_{2,n} = \beta_n \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2.$$

Assuming mild moment conditions on the sample distribution, our main result (Theorem 2.1) offers an asymptotic development of the form

$$\mathbb{P}\left(\left|S^2_{2,n} - \sigma^2\right| \geq \left|S^2_{1,n} - \sigma^2\right|\right) = \frac{1}{2} + \frac{\Delta}{\sqrt{n}} + \mathrm{o}\left(\frac{1}{\sqrt{n}}\right),$$

where the quantity Δ depends both on the moments of the distribution and the ratio of the sequences $(\alpha_n)_n$ and $(\beta_n)_n$. It is our belief that this probability should be reported in priority before deciding whether to use $S^2_{2,n}$ instead of $S^2_{1,n}$, depending on the sign and values of Δ. Standard distribution examples together with classical variance estimates are discussed, and similar results pertaining to the estimation of the standard deviation σ are also reported.

2 Main results

As for now, we let $X_1, \ldots, X_n$ $(n \geq 2)$ be independent and identically distributed real-valued random variables, with unknown finite variance $\sigma^2 > 0$. Throughout the document, we let X be a generic random variable distributed as X_1 and make the following assumption on the distribution of X:

Assumption [A] Let $m = \mathbb{E}X$. Then

(*i*) $\mathbb{E}X^6 < \infty$ and $\tau > 0$, where

$$\tau^2 = \mathbb{E}\left[\frac{X-m}{\sigma}\right]^4 - 1,$$

(*ii*) and

$$\limsup_{|u|+|v|\to\infty} \left|\mathbb{E}\exp\left(iuX + ivX^2\right)\right| < 1.$$

The latter restriction, often called Cramér's condition, holds if the distribution of X is nonsingular or, equivalently, if that distribution has a nondegenerate absolutely continuous component—in particular, if X has a proper density function. A proof of this fact is given in Hall [5, Chapter 2].

On the basis of the given sample $X_1, \ldots, X_n$, we wish to estimate σ^2. In this context, suppose that we are given two estimates $S_{1,n}^2$ and $S_{2,n}^2$, respectively defined by

$$S_{1,n}^2 = \alpha_n \sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2 \quad \text{and} \quad S_{2,n}^2 = \beta_n \sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2, \qquad (2.1)$$

where $(\alpha_n)_n$ and $(\beta_n)_n$ are two positive sequences. Examples of such sequences have already been reported in the introduction section, and various additional illustrations will be discussed below. As a leading example, the reader should keep in mind the normal case, with $\alpha_n = 1/(n-1)$ (unbiased estimate) and $\beta_n = 1/(n+1)$ (minimum quadratic risk estimate). We first state our main result, whose proof relies on the technique of Edgeworth expansion (see, e.g., Hall [5, Chapter 2]).

Theorem 2.1 *Assume that Assumption* [**A**] *is satisfied, and that the sequences* $(\alpha_n)_n$ *and* $(\beta_n)_n$ *in (2.1) satisfy the constraints*

$$(i)\ \beta_n < \alpha_n \quad and \quad (ii)\ \frac{2}{\alpha_n + \beta_n} = n + a + \mathrm{o}(1) \quad as\ n \to \infty,$$

where $a \in \mathbb{R}$. Then, for the estimates $S_{1,n}^2$ and $S_{2,n}^2$ in (2.1),

$$\mathbb{P}\left(\left|S_{2,n}^2 - \sigma^2\right| \geq \left|S_{1,n}^2 - \sigma^2\right|\right)$$
$$= \frac{1}{2} + \frac{1}{\sqrt{2\pi n}}\left[\frac{a+1}{\tau} - \frac{1}{\tau^3}\left(\gamma^2 - \frac{\lambda}{6}\right)\right] + o\left(\frac{1}{\sqrt{n}}\right)$$

as $n \to \infty$, where

$$\gamma = \mathbb{E}\left[\frac{X-m}{\sigma}\right]^3 \quad and \quad \lambda = \mathbb{E}\left[\left(\frac{X-m}{\sigma}\right)^2 - 1\right]^3.$$

Some comments are in order to explain the meaning of the requirements of Theorem 2.1. Condition (i) may be interpreted by considering that $S_{2,n}^2$ is a shrinked version of $S_{1,n}^2$. For example, in the normal population context, we typically have the ordering

$$\frac{1}{n+1} < \frac{1}{n} < \frac{1}{n-1},$$

which corresponds to the successive shrinked estimates $S_{\mathrm{M},n}^2$, $S_{\mathrm{SV},n}^2$ and $S_{\mathrm{U},n}^2$. To understand condition (ii), it is enough to note that an estimate of σ^2 of the form $\delta_n \sum_{i=1}^n (X_i - \bar{X}_n)^2$ is (weakly or strongly) consistent if, and only if, $\delta_n \sim 1/n$ as $n \to \infty$. Therefore, for consistent estimates $S_{1,n}^2$ and $S_{2,n}^2$, it holds $2/(\alpha_n + \beta_n) \sim n$, and condition (ii) just specifies this asymptotic development.

Finally, it is noteworthy to mention that all presented results may be adapted without too much effort to the known mean case, by replacing $\sum_{i=1}^n (X_i - \bar{X}_n)^2$ by $\sum_{i=1}^n (X_i - m)^2$ in the corresponding estimates. To see this, it suffices to observe that the proof of Theorem 2.1 starts with the following asymptotic normality result (see Proposition 4.1):

$$\sqrt{n}\,\frac{\frac{1}{n}\sum_{i=1}^n (Z_i - \bar{Z}_n)^2 - 1}{\tau} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \text{as } n \to \infty, \tag{2.2}$$

where

$$Z_i = \frac{X_i - m}{\sigma} \quad and \quad \tau^2 = \mathbb{E}\left[\frac{X-m}{\sigma}\right]^4 - 1.$$

When the mean m is known, (2.2) has to be replaced by

$$\sqrt{n}\,\frac{\frac{1}{n}\sum_{i=1}^n Z_i^2 - 1}{\tau} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \text{as } n \to \infty,$$

and the subsequent developments are similar. We leave the interested reader the opportunity to adapt the results to this less interesting situation.

We are now in a position to discuss some application examples.

Example 1   Suppose that $X_1, \ldots, X_n$ are independently normally distributed with unknown positive variance σ^2. Elementary calculations show that, in this setting, $\tau^2 = 2$, $\gamma = 0$ and $\lambda = 8$.

The sample variance (maximum likelihood) $S_{\text{SV},n}^2$ has $\alpha_n = 1/n$, whereas the unbiased (jacknife) estimate $S_{\text{U},n}^2$ has $\alpha_n = 1/(n-1)$. The minimum risk estimate $S_{\text{M},n}^2$, which minimizes the mean squared error uniformly in n and σ^2, has $\beta_n = 1/(n+1)$ (Lehmann and Casella [11, Chapter 2]). Thus, $S_{\text{M},n}^2$ is a shrinked version of both $S_{\text{SV},n}^2$ and $S_{\text{U},n}^2$ (that is, $\beta_n < \alpha_n$), with

$$\frac{2}{\alpha_n + \beta_n} = \frac{2n^2 + 2n}{2n+1} = n + \frac{1}{2} + \text{o}(1) \quad \text{for } S_{\text{M},n}^2 \text{ vs } S_{\text{SV},n}^2, \qquad (2.3)$$

and

$$\frac{2}{\alpha_n + \beta_n} = \frac{n^2 - 1}{n} = n - \frac{1}{n} \quad \text{for } S_{\text{M},n}^2 \text{ vs } S_{\text{U},n}^2. \qquad (2.4)$$

Therefore, in this context, Theorem 2.1 asserts that

$$\mathbb{P}\left(\left|S_{\text{M},n}^2 - \sigma^2\right| > \left|S_{\text{SV},n}^2 - \sigma^2\right|\right) = \frac{1}{2} + \frac{13}{12\sqrt{\pi n}} + \text{o}\left(\frac{1}{\sqrt{n}}\right)$$

and

$$\mathbb{P}\left(\left|S_{\text{M},n}^2 - \sigma^2\right| > \left|S_{\text{U},n}^2 - \sigma^2\right|\right) = \frac{1}{2} + \frac{5}{6\sqrt{\pi n}} + \text{o}\left(\frac{1}{\sqrt{n}}\right).$$

Put differently, $S_{\text{SV},n}^2$ and $S_{\text{U},n}^2$ are both asymptotically Pitman closer to σ^2 than $S_{\text{M},n}$. It is also interesting to note that, according to (1.1), the maximum likelihood estimate has uniformly smaller risk than the unbiased estimate, i.e., for all n and all values of the parameter,

$$\mathbb{E}\left[S_{\text{SV},n}^2 - \sigma^2\right]^2 < \mathbb{E}\left[S_{\text{U},n}^2 - \sigma^2\right]^2.$$

Clearly, $S_{\text{SV},n}^2$ may be regarded as a shrinkage estimate of $S_{\text{U},n}^2$ and, with $\alpha_n = 1/(n-1)$ and $\beta_n = 1/n$, we obtain

$$\frac{2}{\alpha_n + \beta_n} = \frac{2n^2 - 2n}{2n-1} = n - \frac{1}{2} + \text{o}(1),$$

so that

$$\mathbb{P}\left(\left|S_{\text{SV},n}^2 - \sigma^2\right| > \left|S_{\text{U},n}^2 - \sigma^2\right|\right) = \frac{1}{2} + \frac{7}{12\sqrt{\pi n}} + \text{o}\left(\frac{1}{\sqrt{n}}\right).$$

8

The take-home message here is that even if shrinkage improves the risk of squared error loss, it should nevertheless be carefully considered from a point estimation perspective. In particular, the unbiased estimate $S_{\mathrm{U},n}^2$ is asymptotically Pitman closer to the target σ^2 than the shrinked (and mean squared optimal) estimate $S_{\mathrm{M},n}^2$. We have indeed

$$\lim_{n\to\infty} \sqrt{n} \left[\mathbb{P}\left(\left|S_{\mathrm{M},n}^2 - \sigma^2\right| > \left|S_{\mathrm{U},n}^2 - \sigma^2\right| \right) - \frac{1}{2} \right] = \frac{5}{6\sqrt{\pi}},$$

despite the fact that, for all n,

$$\mathbb{E}\left[S_{\mathrm{M},n}^2 - \sigma^2 \right]^2 < \mathbb{E}\left[S_{\mathrm{U},n}^2 - \sigma^2 \right]^2.$$

This clearly indicates a potential weakness for any estimate obtained by minimizing a risk function, because extreme estimate's values that have small probability can drastically increase the risk function's value.

To continue the discussion, we may denote by ℓ a real number less than 1 and consider variance estimates of the general form

$$S_{\ell,n}^2 = \frac{1}{n+\ell} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2, \quad n > -\ell. \tag{2.5}$$

Clearly, $S_{\mathrm{M},n}^2$ is a shrinked version of $S_{\ell,n}^2$ and, in the normal setting, for all $n > -\ell$,

$$\mathbb{E}\left[S_{\mathrm{M},n}^2 - \sigma^2 \right]^2 < \mathbb{E}\left[S_{\ell,n}^2 - \sigma^2 \right]^2.$$

Next, applying Theorem 2.1 with

$$\alpha_n = \frac{1}{n+\ell} \quad \text{and} \quad \beta_n = \frac{1}{n+1},$$

we may write

$$\frac{2}{\alpha_n + \beta_n} = n + \frac{\ell+1}{2} + o(1)$$

and, consequently,

$$\mathbb{P}\left(\left|S_{\mathrm{M},n}^2 - \sigma^2\right| > \left|S_{\ell,n}^2 - \sigma^2\right| \right) = \frac{1}{2} + \frac{1}{4\sqrt{\pi n}} \left(\ell + \frac{13}{3} \right) + o\left(\frac{1}{\sqrt{n}} \right).$$

The multiplier of the $1/\sqrt{n}$ term is positive for all $\ell > -13/3 \approx -4.33$. Thus, for $\ell \in (-13/3, 1)$, the estimate (2.5) is asymptotically Pitman closer to σ^2 than $S_{\mathrm{M},n}^2$, the minimum quadratic risk estimate. Note that this result is in accordance with Pitman's observation that, in the Gaussian case, the

9

best variance estimate with respect to PCC should have approximately $\alpha_n \approx 1/(n - 5/3)$ (Pitman [13, Paragraph 6]).

Example 2   If $X_1, \ldots, X_n$ follow a Student's t-distribution with $\nu > 6$ degrees of freedom and unknown variance σ^2, then it is known (see, e.g., Yatracos [18, Remark 7]) that $S^2_{\mathrm{M},n}$ improves both $S^2_{\mathrm{SV},n}$ and $S^2_{\mathrm{U},n}$ in terms of quadratic error. In this case, $m = 0$, $\gamma = 0$, whereas, for $0 < k < \nu$, even,

$$\mathbb{E}X^k = \nu^{k/2} \prod_{i=1}^{k/2} \frac{2i - 1}{\nu - 2i}.$$

Therefore,

$$\sigma^2 = \frac{\nu}{\nu - 2}, \quad \mathbb{E}X^4 = \frac{3\nu^2}{(\nu - 2)(\nu - 4)} \quad \text{and} \quad \mathbb{E}X^6 = \frac{15\nu^3}{(\nu - 2)(\nu - 4)(\nu - 6)}.$$

Consequently,

$$\tau^2 = \left(\frac{\nu - 2}{\nu}\right)^2 \times \frac{3\nu^2}{(\nu - 2)(\nu - 4)} - 1 = \frac{2\nu - 2}{\nu - 4}$$

and

$$\lambda = \left(\frac{\nu - 2}{\nu}\right)^3 \times \frac{15\nu^3}{(\nu - 2)(\nu - 4)(\nu - 6)}$$
$$- 3\left(\frac{\nu - 2}{\nu}\right)^2 \times \frac{3\nu^2}{(\nu - 2)(\nu - 4)} + 2$$
$$= \frac{8\nu(\nu - 1)}{(\nu - 4)(\nu - 6)}.$$

Hence, using identities (2.3)-(2.4), Theorem 2.1 takes the form

$$\mathbb{P}\left(\left|S^2_{\mathrm{M},n} - \sigma^2\right| > \left|S^2_{\mathrm{SV},n} - \sigma^2\right|\right)$$
$$= \frac{1}{2} + \frac{1}{6\sqrt{\pi n}}\left(\frac{\nu - 4}{\nu - 1}\right)^{1/2}\left(\frac{13\nu/2 - 27}{\nu - 6}\right) + \mathrm{o}\left(\frac{1}{\sqrt{n}}\right)$$

and

$$\mathbb{P}\left(\left|S^2_{\mathrm{M},n} - \sigma^2\right| > \left|S^2_{\mathrm{U},n} - \sigma^2\right|\right)$$
$$= \frac{1}{2} + \frac{1}{6\sqrt{\pi n}}\left(\frac{\nu - 4}{\nu - 1}\right)^{1/2}\left(\frac{5\nu - 18}{\nu - 6}\right) + \mathrm{o}\left(\frac{1}{\sqrt{n}}\right).$$

10

We see that all the constants in front of the $1/\sqrt{n}$ terms are positive for $\nu > 6$, despite the fact that

$$\mathbb{E}\left[S_{\mathrm{M},n}^2 - \sigma^2\right]^2 < \mathbb{E}\left[S_{\mathrm{SV},n}^2 - \sigma^2\right]^2 \quad \text{and} \quad \mathbb{E}\left[S_{\mathrm{M},n}^2 - \sigma^2\right]^2 < \mathbb{E}\left[S_{\mathrm{U},n}^2 - \sigma^2\right]^2.$$

Example 3 For non-normal populations, $S_{\mathrm{U},n}^2$ may have a smaller mean squared error than either $S_{\mathrm{SV},n}^2$ or $S_{\mathrm{M},n}^2$. In this general context, Yatracos [18] proved that the estimate

$$S_{\mathrm{Y},n}^2 = \frac{n+2}{n(n+1)} \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2$$

ameliorates the mean squared error of $S_{\mathrm{U},n}^2$ for all probability models with finite second moment, all values of σ^2 and all sample sizes $n \geq 2$. Here,

$$\alpha_n = \frac{1}{n-1} \quad \text{and} \quad \beta_n = \frac{n+2}{n(n+1)},$$

so that, for all $n \geq 2$, $\beta_n/\alpha_n < 1$ and

$$\frac{2}{\alpha_n + \beta_n} = \frac{n^3 - n}{n^2 + n - 1} = n - 1 + \mathrm{o}(1).$$

It follows, assuming that Assumption [**A**] is satisfied, that

$$\mathbb{P}\left(\left|S_{\mathrm{Y},n}^2 - \sigma^2\right| \geq \left|S_{\mathrm{U},n}^2 - \sigma^2\right|\right) = \frac{1}{2} + \frac{1}{\sqrt{2\pi n}}\left[-\frac{1}{\tau^3}\left(\gamma^2 - \frac{\lambda}{6}\right)\right] + \mathrm{o}\left(\frac{1}{\sqrt{n}}\right).$$

For example, if X follows a normal distribution,

$$\mathbb{P}\left(\left|S_{\mathrm{Y},n}^2 - \sigma^2\right| > \left|S_{\mathrm{U},n}^2 - \sigma^2\right|\right) = \frac{1}{2} + \frac{1}{3\sqrt{\pi n}} + \mathrm{o}\left(\frac{1}{\sqrt{n}}\right).$$

3   Standard deviation shrinkage

Section 2 was concerned with the shrinkage estimation problem of the variance σ^2. Estimating the standard deviation σ is more involved, since, for example, it is not possible to find an estimate of σ which is unbiased for all population distributions (Lehmann and Casella [11, Chapter 2]). Nevertheless, interesting results may still be reported when the sample observations $X_1, \ldots, X_n$ follow a normal distribution $\mathcal{N}(m, \sigma^2)$.

The most common estimates used to assess the standard deviation parameter σ typically have the form

$$\sqrt{S_{\mathrm{SV},n}^2} = \frac{1}{\sqrt{n}} \left[\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \right]^{1/2} \text{ or } \sqrt{S_{\mathrm{U},n}^2} = \frac{1}{\sqrt{n-1}} \left[\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \right]^{1/2}.$$

In all generality, both $\sqrt{S_{\mathrm{SV},n}^2}$ and $\sqrt{S_{\mathrm{U},n}^2}$ are biased estimates of σ. However, when the random variable X is normally distributed, a minor correction exists to eliminate the bias. To derive the correction, just note that, according to Cochran's theorem, $\sum_{i=1}^{n}(X_i - \bar{X}_n)^2/\sigma^2$ has a chi-squared distribution with $n-1$ degrees of freedom. Consequently, $[\sum_{i=1}^{n}(X_i - \bar{X}_n)^2]^{1/2}/\sigma$ has a chi distribution with $n-1$ degrees of freedom (Johnson, Kotz, and Balakrishnan [6, Chapter 18]), whence

$$\mathbb{E} \left[\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \right]^{1/2} = \frac{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \sigma,$$

where $\Gamma(.)$ is the gamma function. It follows that the quantity

$$\hat{\sigma}_{\mathrm{U},n} = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)} \left[\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \right]^{1/2}$$

is an unbiased estimate of σ. Besides, still assuming normality and letting $(\delta_n)_n$ be some generic positive normalization sequence, we may write

$$\mathbb{E} \left[\delta_n \left[\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \right]^{1/2} - \sigma \right]^2$$

$$= \delta_n^2 \,\mathbb{E} \left[\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \right] - 2\sigma\delta_n \,\mathbb{E} \left[\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \right]^{1/2} + \sigma^2,$$

hence

$$\mathbb{E} \left[\delta_n \left[\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \right]^{1/2} - \sigma \right]^2 = \sigma^2 \left[(n-1)\delta_n^2 - 2\delta_n \frac{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} + 1 \right].$$

Solving this quadratic equation in δ_n, we see that the right-hand side is uniformly minimized for the choice

$$\delta_n^\star = \frac{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)}{(n-1)\Gamma\left(\frac{n-1}{2}\right)} = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)}.$$

(see Goodman [3]). Put differently, the estimate

$$\hat{\sigma}_{\mathrm{M},n} = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)}\left[\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right]^{1/2}$$

improves uniformly upon $\sqrt{S_{\mathrm{SV},n}^2}$, $\sqrt{S_{\mathrm{U},n}^2}$ and $\hat{\sigma}_{\mathrm{U},n}$, which have, respectively,

$$\delta_n = \frac{1}{\sqrt{n}}, \quad \delta_n = \frac{1}{\sqrt{n-1}} \quad \text{and} \quad \delta_n = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)}.$$

Using the expansion

$$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = \sqrt{\frac{n}{2}}\left[1 - \frac{1}{4n} + \mathrm{o}\left(\frac{1}{n}\right)\right],$$

we may write

$$\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)} = \frac{1}{\sqrt{n}}\left[1 + \frac{1}{4n} + \mathrm{o}\left(\frac{1}{n}\right)\right] \tag{3.1}$$

and

$$\frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)} = \frac{\sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)}{(n-1)\Gamma\left(\frac{n}{2}\right)} = \frac{1}{\sqrt{n-1}}\left[1 + \frac{1}{4n} + \mathrm{o}\left(\frac{1}{n}\right)\right]. \tag{3.2}$$

The relative positions of the estimates $\sqrt{S_{\mathrm{SV},n}^2}$, $\hat{\sigma}_{\mathrm{M},n}$, $\sqrt{S_{\mathrm{U},n}^2}$ and $\hat{\sigma}_{\mathrm{U},n}$ together with their coefficients are shown in Figure 2.
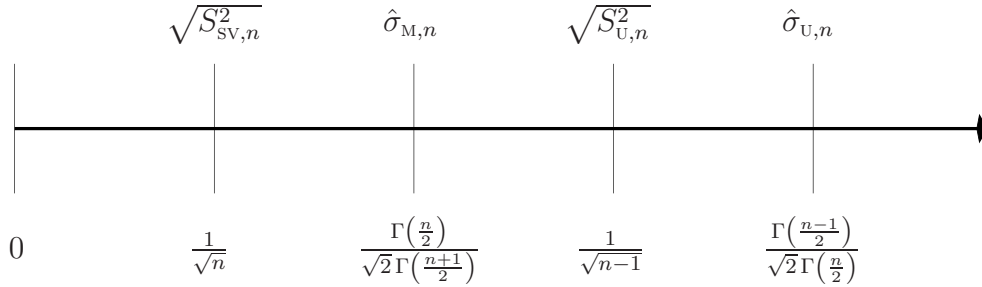
Figure 2: Relative positions of the estimates $\sqrt{S_{\mathrm{SV},n}^2}$, $\hat{\sigma}_{\mathrm{M},n}$, $\sqrt{S_{\mathrm{U},n}^2}$ and $\hat{\sigma}_{\mathrm{U},n}$, and their coefficients.

Theorem 3.1 below is the standard deviation counterpart of Theorem 2.1 for normal populations. Let

$$\hat{\sigma}_{1,n}^2 = \alpha_n\left[\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2\right]^{1/2} \quad \text{and} \quad \hat{\sigma}_{2,n}^2 = \beta_n\left[\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2\right]^{1/2} \tag{3.3}$$

be two candidates to the estimation of σ.

Theorem 3.1 *Assume that X has a normal distribution, and that the sequences $(\alpha_n)_n$ and $(\beta_n)_n$ in (3.3) satisfy the constraints*

$$(i)\ \beta_n < \alpha_n \quad and \quad (ii)\ \left[\frac{2}{\alpha_n + \beta_n}\right]^2 = n + b + o(1) \quad as\ n \to \infty,$$

where $b \in \mathbb{R}$. Then, for the estimates $\hat{\sigma}_{1,n}$ and $\hat{\sigma}_{2,n}$ in (3.3)

$$\mathbb{P}\left(|\hat{\sigma}_{2,n} - \sigma| > |\hat{\sigma}_{1,n} - \sigma|\right) = \frac{1}{2} + \frac{1}{2\sqrt{\pi n}}\left(b + \frac{5}{3}\right) + o\left(\frac{1}{\sqrt{n}}\right)$$

as $n \to \infty$.

As expressed by Figure 2, $\hat{\sigma}_{\mathrm{M},n}$ is a shrinked version of both $\sqrt{S^2_{\mathrm{U},n}}$ and $\hat{\sigma}_{\mathrm{U},n}$. Thus, continuing our discussion, we may first compare the performance, in terms of Pitman closeness, of $\hat{\sigma}_{\mathrm{U},n}$ vs $\hat{\sigma}_{\mathrm{M},n}$. These estimates have, respectively,

$$\alpha_n = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)} \quad and \quad \beta_n = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)}.$$

Using (3.1) and (3.2), we easily obtain

$$\left[\frac{2}{\alpha_n + \beta_n}\right]^2 = n - 1 + o(1),$$

so that

$$\mathbb{P}\left(|\hat{\sigma}_{\mathrm{M},n} - \sigma| > |\hat{\sigma}_{\mathrm{U},n} - \sigma|\right) = \frac{1}{2} + \frac{1}{3\sqrt{\pi n}} + o\left(\frac{1}{\sqrt{n}}\right).$$

Similarly, with

$$\alpha_n = \frac{1}{\sqrt{n-1}} \quad and \quad \beta_n = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)},$$

we conclude

$$\mathbb{P}\left(|\hat{\sigma}_{\mathrm{M},n} - \sigma| > \left|\sqrt{S^2_{\mathrm{U},n}} - \sigma\right|\right) = \frac{1}{2} + \frac{11}{24\sqrt{\pi n}} + o\left(\frac{1}{\sqrt{n}}\right).$$

Remark 1 The methodology developed in the present paper can serve as a basis for analyzing other types of estimates. Suppose, for example, that $X_1, \ldots, X_n$ $(n \geq 2)$ are independent identically distributed random variables with common density $f(x; \mu, \sigma) = \sigma^{-1} e^{-(x-\mu)/\sigma}$ $(x > \mu)$, where $-\infty < \mu < \infty$ and $\sigma > 0$. On the basis of the given sample we wish to estimate the standard deviation σ. Denoting the order statistics associated with $X_1, \ldots, X_n$ by

14

$X_{(1)}, \ldots, X_{(n)}$, one may write the maximum likelihood estimate of σ (which turns out to be minimum variance unbiased) in the form

$$T_{\mathrm{ML},n} = \frac{1}{n-1} \sum_{i=2}^{n} (X_{(i)} - X_{(1)}).$$

By sacrificing unbiasedness, we can consider as well the estimate

$$T_{\mathrm{M},n} = \frac{1}{n} \sum_{i=2}^{n} (X_{(i)} - X_{(1)})$$

which improves upon $T_{\mathrm{ML},n}$ uniformly (Arnold [1]) in terms of mean squared error. $T_{\mathrm{M},n}$ is a shrinkage estimate of $T_{\mathrm{ML},n}$ and, by an application of Lemma 4.1, we have

$$\mathbb{P}\left(|T_{\mathrm{M},n} - \sigma| > |T_{\mathrm{ML},n} - \sigma|\right) = \mathbb{P}\left(\Gamma_{n-1} < \frac{2n(n-1)}{2n-1}\right),$$

since $\sum_{i=2}^{n} (X_{(i)} - X_{(1)})/\sigma$ is distributed as a gamma random variable with $n-1$ degrees of freedom, denoted Γ_{n-1}. Recalling that $\Gamma_{n-1} \sim \sum_{i=1}^{n-1} Y_i$, where $Y_1, \ldots, Y_{n-1}$ are independent standard exponential random variables, we easily obtain, using the same Edgeworth-based methodology as was used to prove Theorem 2.1,

$$\mathbb{P}\left(|T_{\mathrm{M},n} - \sigma| > |T_{\mathrm{ML},n} - \sigma|\right) = \frac{1}{2} + \frac{5}{6\sqrt{2\pi n}} + \mathrm{o}\left(\frac{1}{n}\right).$$

4 Proofs

4.1 Some preliminary results

Recall that $X_1, \ldots, X_n$ ($n \geq 2$) denote independent real-valued random variables, distributed as a generic random variable X with finite variance $\sigma^2 > 0$. Let $\Phi(x)$ be the cumulative distribution function of the standard normal distribution, that is, for all $x \in \mathbb{R}$,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} \mathrm{d}u.$$

We start by stating the following lemma, which is but a special case of Proposition 2.1 in Yatracos [19].

Lemma 4.1 *Let T be a $\mathbb{P}$-a.s. nonnegative real-valued random variable, and let $(\theta, c) \in \mathbb{R}^+ \times (-1, 1)$ be two real numbers. Then*

$$\mathbb{P}\left(|cT - \theta| \geq |T - \theta|\right) = \mathbb{P}\left(T \leq \frac{2\theta}{1 + c}\right).$$

Proof of Lemma 4.1   Just observe that

$$
\begin{aligned}
\mathbb{P}\left(|cT - \theta| \geq |T - \theta|\right) &= \mathbb{P}\left((cT - \theta)^2 \geq (T - \theta)^2\right) \\
&= \mathbb{P}\left(\left[(1 + c)T - 2\theta\right](c - 1)T \geq 0\right) \\
&= \mathbb{P}\left((1 + c)T - 2\theta \leq 0\right) \\
&\quad \text{(since } T \text{ is } \mathbb{P}\text{-a.s. nonnegative, } c < 1 \text{ and } \theta \geq 0) \\
&= \mathbb{P}\left(T \leq \frac{2\theta}{1 + c}\right) \\
&\quad \text{(since } c > -1).
\end{aligned}
$$

$\blacksquare$

Proposition 4.1 *Assume that Assumption* [**A**] *is satisfied. Then, as $n \to \infty$,*

$$
\mathbb{P}\left(\sum_{i=1}^n (X_i - \bar{X}_n)^2 \leq (n + t)\sigma^2\right)
$$

$$
= \Phi\left(\frac{t}{\tau\sqrt{n}}\right) + \frac{1}{\sqrt{2\pi n}}\, p_1\left(\frac{t}{\tau\sqrt{n}}\right) e^{-\frac{t^2}{2\tau^2 n}} + \mathrm{o}\left(\frac{1}{\sqrt{n}}\right),
$$

uniformly in $t \in \mathbb{R}$, where

$$
p_1(x) = \frac{1}{\tau} + \frac{1}{\tau^3}\left(\gamma^2 - \frac{\lambda}{6}\right)(x^2 - 1),
$$

with

$$
\gamma = \mathbb{E}\left[\frac{X - m}{\sigma}\right]^3 \quad \text{and} \quad \lambda = \mathbb{E}\left[\left(\frac{X - m}{\sigma}\right)^2 - 1\right]^3.
$$

Proof of Proposition 4.1   Set

$$
Z = \frac{X - m}{\sigma} \quad \text{and} \quad Z_i = \frac{X_i - m}{\sigma}, \quad i = 1, \ldots, n,
$$

and observe that, by the central limit theorem and Slutsky's lemma (van der Vaart [17, Chapter 2]),

$$
\sqrt{n}\,\frac{\frac{1}{n}\sum_{i=1}^n (Z_i - \bar{Z}_n)^2 - 1}{\tau} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } n \to \infty,
$$

16

where

$$\bar{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i \quad \text{and} \quad \tau^2 = \operatorname{Var} Z^2 = \mathbb{E}\left[\frac{X-m}{\sigma}\right]^4 - 1.$$

The result will be proved by making this limit more precise using an Edgeworth expansion (see, e.g., Hall [5, Chapter 2]). To this aim, we first need some additional notation. Set $\mathbf{Z} = (Z, Z^2)$, $\mathbf{m} = \mathbb{E}\mathbf{Z} = (0, 1)$ and, for $\mathbf{z} = (z^{(1)}, z^{(2)}) \in \mathbb{R}^2$, let

$$A(\mathbf{z}) = \frac{z^{(2)} - (z^{(1)})^2 - 1}{\tau}.$$

Clearly, $A(\mathbf{m}) = 0$ and

$$\sqrt{n}\,\frac{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z}_n)^2 - 1}{\tau} = \sqrt{n}\,A(\bar{\mathbf{Z}}_n).$$

For $j \geq 1$ and $i_j \in \{1, 2\}$, put

$$a_{i_1 \dots i_j} = \frac{\partial^j A(\mathbf{z})}{\partial z^{(i_1)} \dots \partial z^{(i_j)}}\Big|_{\mathbf{z}=\mathbf{m}}.$$

For example,

$$a_2 = \frac{\partial^1 A(\mathbf{z})}{\partial z^{(2)}}\Big|_{\mathbf{z}=\mathbf{m}} = \frac{1}{\tau}$$

and

$$a_{11} = \frac{\partial^2 A(\mathbf{z})}{\partial z^{(1)} \partial z^{(1)}}\Big|_{\mathbf{z}=\mathbf{m}} = -\frac{2}{\tau}.$$

Let also

$$\mu_{i_1 \dots i_j} = \mathbb{E}\left[(\mathbf{Z}-\mathbf{m})^{(i_1)} \dots (\mathbf{Z}-\mathbf{m})^{(i_j)}\right],$$

where $(\mathbf{Z}-\mathbf{m})^{(i)}$ denotes the i-th component of the vector $(\mathbf{Z}-\mathbf{m})$. Thus, with this notation, according to Hall [5, Theorem 2.2], under the condition

$$\limsup_{|u|+|v|\to\infty} \left|\mathbb{E}\exp\left(iuX + ivX^2\right)\right| < 1,$$

we may write, as $n \to \infty$,

$$\mathbb{P}\left(\sqrt{n}\,A(\bar{\mathbf{Z}}_n) \leq x\right) = \Phi(x) + \frac{1}{\sqrt{2\pi n}}\,p_1(x)e^{-x^2/2} + o\left(\frac{1}{\sqrt{n}}\right),$$

uniformly in $x \in \mathbb{R}$, where

$$p_1(x) = -A_1 - \frac{1}{6}A_2(x^2 - 1).$$

17

The coefficients A_1 and A_2 in the polynomial p_1 are respectively given by the formulae

$$A_1 = \frac{1}{2} \sum_{i=1}^{2} \sum_{j=1}^{2} a_{ij} \mu_{ij}$$

and

$$A_2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{2} a_i a_j a_k \mu_{ijk} + 3 \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{2} \sum_{\ell=1}^{2} a_i a_j a_{k\ell} \mu_{ik} \mu_{j\ell}.$$

Elementary calculations show that

$$a_2 = \tau^{-1}, \quad a_{11} = -2\tau^{-1}, \quad \text{and} \quad a_1 = a_{22} = a_{12} = a_{21} = 0.$$

Similarly,

$$\mu_{11} = 1, \quad \mu_{22} = \tau^2, \quad \mu_{12} = \mu_{21} = \mathbb{E}[X - m]^3 \sigma^{-3}, \quad \text{and}$$

$$\mu_{222} = \mathbb{E} \left[(X - m)^2 \sigma^{-2} - 1 \right]^3.$$

Consequently,

$$A_1 = -\frac{1}{\tau} \quad \text{and} \quad A_2 = \frac{1}{\tau^3} \left(\lambda - 6\gamma^2 \right),$$

with

$$\lambda = \mathbb{E} \left[\left(\frac{X - m}{\sigma} \right)^2 - 1 \right]^3 \quad \text{and} \quad \gamma = \mathbb{E} \left[\frac{X - m}{\sigma} \right]^3.$$

Therefore

$$p_1(x) = \frac{1}{\tau} + \frac{1}{\tau^3} \left(\gamma^2 - \frac{\lambda}{6} \right) (x^2 - 1).$$

The conclusion follows by observing that, for all $t \in \mathbb{R}$,

$$\mathbb{P} \left(\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \leq (n + t)\sigma^2 \right) = \mathbb{P} \left(\sqrt{n} \, A(\bar{\mathbf{Z}}_n) \leq \frac{t}{\tau\sqrt{n}} \right).$$

∎

4.2 Proof of Theorem 2.1

Observe that $S_{2,n}^2 = c_n S_{1,n}^2$, where $c_n = \beta_n/\alpha_n \in (0,1)$ by assumption (i). Consequently, by Lemma 4.1,

$$
\begin{aligned}
\mathbb{P}\left(\left|S_{2,n}^2 - \sigma^2\right| \geq \left|S_{1,n}^2 - \sigma^2\right|\right) &= \mathbb{P}\left(S_{1,n}^2 \leq \frac{2\sigma^2}{1 + \beta_n/\alpha_n}\right) \\
&= \mathbb{P}\left(\sum_{i=1}^n (X_i - \bar{X}_n)^2 \leq \frac{2\sigma^2}{\alpha_n + \beta_n}\right) \\
&= \mathbb{P}\left(\sum_{i=1}^n (X_i - \bar{X}_n)^2 \leq (n + a + \zeta_n)\sigma^2\right)
\end{aligned}
$$

(by assumption (ii)),

where $\zeta_n \to 0$ as $n \to \infty$. Let $\Phi(x)$ be the cumulative distribution function of the standard normal distribution, that is, for all $x \in \mathbb{R}$,

$$
\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2}\mathrm{d}u.
$$

Thus, assuming $[\mathbf{A}]$ and using Proposition 4.1, we may write

$$
\begin{aligned}
&\mathbb{P}\left(\left|S_{2,n}^2 - \sigma^2\right| \geq \left|S_{1,n}^2 - \sigma^2\right|\right) \\
&= \Phi\left(\frac{a + \zeta_n}{\tau\sqrt{n}}\right) + \frac{1}{\sqrt{2\pi n}} p_1\left(\frac{a + \zeta_n}{\tau\sqrt{n}}\right) e^{-\frac{(a+\zeta_n)^2}{2\tau^2 n}} + \mathrm{o}\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

where

$$
p_1(x) = \frac{1}{\tau} + \frac{1}{\tau^3}\left(\gamma^2 - \frac{\lambda}{6}\right)(x^2 - 1),
$$

with

$$
\tau^2 = \mathbb{E}\left[\frac{X - m}{\sigma}\right]^4 - 1,
$$

$$
\gamma = \mathbb{E}\left[\frac{X - m}{\sigma}\right]^3 \quad \text{and} \quad \lambda = \mathbb{E}\left[\left(\frac{X - m}{\sigma}\right)^2 - 1\right]^3.
$$

Using finally the Taylor series expansions, valid as $x \to 0$,

$$
\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\left(x + \mathrm{o}(x^2)\right) \quad \text{and} \quad e^x = 1 + \mathrm{o}(1),
$$

we obtain

$$
\begin{aligned}
&\mathbb{P}\left(\left|S_{2,n}^2 - \sigma^2\right| \geq \left|S_{1,n}^2 - \sigma^2\right|\right) \\
&= \frac{1}{2} + \frac{1}{\sqrt{2\pi n}}\left[\frac{a + 1}{\tau} - \frac{1}{\tau^3}\left(\gamma^2 - \frac{\lambda}{6}\right)\right] + \mathrm{o}\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

as desired.

4.3 Proof of Theorem 3.1

By assumption (i), we may write $\hat{\sigma}_{2,n} = c_n \hat{\sigma}_{1,n}$, where $c_n = \beta_n/\alpha_n \in (0,1)$. Consequently, by Lemma 4.1,

$$\mathbb{P}\left(|\hat{\sigma}_{2,n} - \sigma| > |\hat{\sigma}_{1,n} - \sigma|\right)$$
$$= \mathbb{P}\left(|\hat{\sigma}_{2,n} - \sigma| \geq |\hat{\sigma}_{1,n} - \sigma|\right)$$
$$= \mathbb{P}\left(\hat{\sigma}_{1,n} \leq \frac{2\sigma}{1 + \beta_n/\alpha_n}\right)$$
$$= \mathbb{P}\left(\left[\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right]^{1/2} \leq \frac{2\sigma}{\alpha_n + \beta_n}\right)$$
$$= \mathbb{P}\left(\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \leq (n + b + \zeta_n)\sigma^2\right)$$
$$\text{(by assumption } (ii)),$$

where $\zeta_n \to 0$ as $n \to \infty$. The end of the proof is similar to the one of Theorem 2.1, recalling that, in the Gaussian setting, $\tau^2 = 2$, $\gamma = 0$ and $\lambda = 8$.

References

[1] B.C. Arnold. Inadmissibility of the usual scale estimate for a shifted exponential distribution. *Journal of the American Statistical Association*, 65:1260–1264, 1970.

[2] L. Brown. Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. *The Annals of Mathematical Statistics*, 39:29–48, 1968.

[3] L.A. Goodman. A simple method for improving some estimators. *The Annals of Mathematical Statistics*, 22:114–117, 1953.

[4] L.A. Goodman. A note on the estimation of variance. *Sankhyā*, 22:221–228, 1960.

[5] P. Hall. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York, 1992.

[6] N.L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions. Vol. 1. Second Edition*. John Wiley & Sons, New York, 1994.

[7] R. Khattree. On comparison of estimates of dispersion using generalized Pitman nearness criterion. *Communications in Statistics-Theory and Methods*, 16:263–274, 1987.

[8] R. Khattree. Comparing estimators for population variance using Pitman nearness. *The American Statistician*, 46:214–217, 1992.

[9] R. Khattree. Estimation of guarantee time and mean life after warranty for two-parameter exponential failure model. *Australian Journal of Statistics*, 34:207–215, 1992.

[10] R. Khattree. On the estimation of σ and the process capability indices C_p and C_{pm}. *Probability in the Engineering and Informational Sciences*, 13:237–250, 1999.

[11] E.L. Lehmann and G. Casella. *Theory of Point Estimation. Second Edition.* Springer-Verlag, New York, 1998.

[12] J.M. Maatta and G. Casella. Developments in decision-theoretic variance estimation. *Statistical Science*, 5:90–120, 1990.

[13] E.J.G. Pitman. The "closest" estimates of statistical parameters. *Proceedings of the Cambridge Philosophical Society*, 33:212–222, 1937.

[14] C.P. Robert, J.T.G. Hwang, and W.E. Strawderman. Is Pitman closeness a reasonable criterion? With comment and rejoinder. *Journal of the American Statistical Association*, 88:57–76, 1993.

[15] A.L. Rukhin. How much better are better estimators of a normal variance. *Journal of the American Statistical Association*, 82:925–928, 1987.

[16] C. Stein. Inadmissibility of the usual estimate for the variance of a normal distribution with unknown mean. *Annals of the Institute of Statistical Mathematics*, 16:155–160, 1964.

[17] A.W. van der Vaart. *Asymptotic Statistics.* Cambridge University Press, Cambridge, 1999.

[18] Y.G. Yatracos. Artificially augmented samples, shrinkage, and mean squared error reduction. *Journal of the American Statistical Association*, 100:1168–1175, 2005.

[19] Y.G. Yatracos. *Pitman's closeness criterion and shrinkage estimates of the variance and the s.d.* Technical Report, Cyprus University of Technology, Cyprus, 2011. `http://ktisis.cut.ac.cy/handle/10488/5009`.