Advances and Applications in Statistics
http://www.pphmj.com
http://dx.doi.org/10.17654/AS054020199
Volume 54, Number 2, 2019, Pages 199-236 **ISSN: 0972-3617**

ON SIGMA-ESTIMATORS FROM A NORMAL UNIVERSE

Saeed Maghsoodloo

ISE Department
Auburn University
3301D Shelby Center, AL 36849, U.S.A.
e-mail: maghsood@eng.auburn.edu

Abstract

This article examines the Central Limit Theorem from the standpoint of skewness and kurtosis, and then investigates the relative and absolute efficiencies of all standard deviation estimators based on one simple random sample of size n from a normal population. We obtained the most precise estimator of the normal population standard deviation σ using Cramer-Rao's lower bound for the variance of an unbiased estimator; however, this last estimator suffers from large negative bias. As a result, we used Lindgren-Cramer-Rau's inequality to develop another estimator of σ that is more accurate than all others for sample sizes $n \geq 2$. The case of $M > 1$ subgroups will be discussed in a subsequent paper, and for $M > 1$, slight modifications will be recommended for some Quality Control Charts.

1. Historical Background

Throughout this entire article, unless otherwise stated, we are assuming that the underlying distribution (or population) is Laplace-Gaussian with the

Received: October 10, 2018; Accepted: December 2, 2018

Keywords and phrases: the central limit theorem (CLT), Fisher's information matrix (IM), mean square error (MSE), relative-efficiencies (REL-EFFs).

pdf (probability density function) given by the probability law:

$$f(x;\ \mu,\ \sigma) = \frac{1}{\sigma\sqrt{2\pi}} Exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad -\infty < x < \infty.$$

The above density was first discovered by Abraham de Moivre (a French mathematician) in 1738 as the limiting distribution of the binomial pmf (probability mass function) and as such he did not study all its properties. During 1809, Marquis P. S. de Laplace published the Central Limit Theorem that stated the limiting sampling distribution (SMD) of sample arithmetic mean $\bar{x}$ is of the form $f(\bar{x};\ \mu,\ \sigma/\sqrt{n})$, n being the sample size. In 1810 Carl F. Gauss published his *Theoria Motus* in which he discussed the statistical properties of the above $f(x;\ \mu,\ \sigma)$. For more complete historical details, the reader should refer to Johnson et al. [10], Vol. 1, pp. 85-88, where we have extracted nearly all of the above historical background. As stated by Kendall and Stuart [12], second edition, Vol. 1, p. 135 footnote, the description of $f(x;\ \mu,\ \sigma)$ as the "normal," was due to Karl Pearson, and we surmise around 1894.

Historical Summary and the CLT

In Statistical literature the designation $X \sim N(\mu,\ \sigma^2)$ implies that a variate (or random variable, *rv*) X is normally distributed with population mean μ and population variance $\sigma^2 = \sigma_X^2$, where μ is the location-parameter and σ is the scale of $f(x;\ \mu,\ \sigma)$.

(1) Because the normal (or bell-shaped) density is the most important of all Statistical distributions, and unique in developing Statistical theory starting in the mid 1700's, below we summarize some of its practical properties, almost all of which are fairly well known. Generally, (nearly) the dimension of every part that is manufactured can be approximately modeled by a normal underlying distribution (provided μ is to the right of zero by at least $6 \times$ STDEV).

(2) The standardized-value is given by $Z = (x - \mu)/\sigma,$ so that $f(x; \mu, \sigma)$ $= \text{Exp}(-Z^2/2)/(\sigma\sqrt{2\pi}) = e^{-Z^2/2}/(\sigma\sqrt{2\pi});$ further, the standardized normal density is universally denoted by $\phi(z) = \text{Exp}(-z^2/2)/\sqrt{2\pi},$ and the corresponding cumulative (cdf) is denoted by $\Phi(z) = \int_{-\infty}^{z} \phi(u)du;$ further, because all normal pdfs have identical bell-shape, no shape parameter exits.

(3) Due to symmetry, the skewness of a normal density is identically equal to zero, i.e., $\alpha_3 = E[(X - \mu)/\sigma]^3 \equiv E(Z^3) = \mu_3/\sigma^3 \equiv 0,$ where $\mu_3 = E[(X - \mu)^3]$ is the 3rd central moment, and E is the linear expected-value operator. The forth standardized central moment of X is given by $\alpha_4 = \mu_4/\sigma^4 \equiv E(Z^4) = 3,$ where $\mu_4 = E[(X - \mu)^4],$ and hence the kurtosis of all normal distributions is identically equal to $\beta_4 = \alpha_4 - 3 \equiv 0.$ Because the normal kurtosis $\beta_4 = \alpha_4 - 3 \equiv 0,$ then the kurtosis of all other statistical distributions is compared against 0 in order to assess their tail thickness. Kendall and Stuart ([12], pp. 85-86) denote kurtosis by γ_2 and name curves with zero γ_2 as Mesokurtic; curves with $\gamma_2 < 0$ as Platykurtic, and those with $\gamma_2 = \beta_4 > 0$ as Leptokurtic. However, they do emphasize that Leptokurtic curves are not necessarily more sharply peaked in the middle than the normal curve, and vice a versa for Platykurtic curves. In fact, $\alpha_4 \geq \alpha_3^2,$ or $\beta_4 \geq \alpha_3^2 - 3$ for all continuous underlying random variables (see also Han and Shapiro [5]). Before discussing the CLT for the general linear combination $Y_n = \sum_{i=1}^{n} a_i X_i,$ where a_i's are any constants, we must refer to the article by Chen et al. [3] that showed in general the 3rd central moment of $Y_n = \sum_{i=1}^{n} a_i X_i$ is given by

$$\mu_3(Y_n) = E\{[Y_n - E(Y_n)]^3\}$$

$$= \sum_{i=1}^{n} a_i^3 \mu_{3i} + 3\sum_{i \neq j}\sum a_i^2 a_j E[(X_i - \mu_i)^2 \times (X_j - \mu_j)]$$

$$+ 6\sum_{i=1}^{n-2}\sum_{j>i}^{n-1}\sum_{k>j}^{n}\{a_i a_j a_k E[(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)]\},\ n > 2, \quad (1a)$$

and its 4th central moment is given by

$$\mu_4(Y_n) = E\{[Y_n - E(Y_n)]^4\} = \sum_{i=1}^{n} a_i^4 \mu_{4i} + 4\sum_{i \neq j}\sum a_i^3 a_j a_{iiij}$$

$$+ 6\sum_{i=1}^{n-1}\sum_{j>i}^{n} a_i^2 a_j^2 a_{iijj} + 12\sum_{i \neq k > j} a_i^2 a_j a_k \sigma_{iijk}$$

$$+ 24\sum_{i=1}^{n-3}\sum_{j>i}^{n-2}\sum_{k>j}^{n-1}\sum_{L>k}^{n} a_i a_j a_k a_L \sigma_{ijkL},\ n > 3, \quad (1b)$$

Where $\sigma_{ijkL} = E[(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)(X_L - \mu_L)]$. We had to provide the above general expressions for the 3rd and 4th central moments of a linear combination in order to illustrate that without the independence assumption amongst the variates $X_1, X_2, ..., X_n$, none of the expectations such as $E[(X_i - \mu_i)^2 \times (X_j - \mu_j)]$, σ_{ijk}, σ_{iiij}, σ_{iijk} and σ_{ijkL} will vanish, while the independence assumption forces each of the last 5 expectations identically to zero. Further, assuming pairwise independence amongst the *n* variates, then the $E[(X_i - \mu_i)^2 \times (X_j - \mu_j)^2] = \sigma_{iijj} = E[(X_i - \mu_i)^2] \times E[(X_j - \mu_j)^2] = \sigma_{ii} \times \sigma_{jj} = V(X_i) \times V(X_j) = \sigma_i^2 \times \sigma_j^2, i \neq j$. Moreover, in the field of Statistics expectations such as $E[(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)]$, $i \neq j \neq k$ have no statistical/and or practical meaning whatsoever, as (zero-order) correlation is always defined pairwise between only two variates.

Fortunately, by the definition of a simple random sample, the sample vector $[X_1\ X_2\ ...\ X_n]^T$ is random iff: **(1)** X_i's follow precisely the same identical statistical distribution as their parent *X*, **(2)** X_i and X_j, for all $i \neq j$, are stochastically pairwise independent so that covariances such as $\text{COV}(X_i, X_j) \equiv 0$ for all $i \neq j$. Although CLT for correlated variates exist in the field of Statistics (but well outside the scope of this article), the above developments clearly show that if X_i's were correlated, then $\alpha_3(S_n) = \mu_3(S_n)/\sigma^3$ will differ from zero and $\alpha_4(S_n) = \mu_4(S_n)/\sigma^4$ will not equal to 3, and hence $Y_n = \sum_{i=1}^{n} a_i X_i$ will not approach normality with increasing *n* from the standpoints of its skewness and kurtosis.

Consequently, henceforth the assumption of pairwise independence is inherent amongst the X_i's in the linear combination $\sum_{i=1}^{n} a_i X_i$, i.e., X_i's are iid (independently and identically distributed).

For example, if all a_i's are equal to $1/n$, then $Y_n = \sum_{i=1}^{n} a_i X_i = \bar{x}$, and for an $N(\mu, \sigma^2)$ equation (1a) shows that $\mu_3(Y_n) = \mu_3(\bar{x}) = \sum_{i=1}^{n} a_i^3 \mu_{3i} = \sum_{i=1}^{n} (1/n)^3 \mu_{3i} \equiv 0$ because $\mu_{3i} = \mu_3(X_i) \equiv 0$ for all *i*. Equation (1b) shows that

$$\mu_4(Y_n) = \mu_4(\bar{x}) = \sum_{i=1}^{n} a_i^4 \mu_{4i} + 6\sum_{i=1}^{n-1}\sum_{j>i}^{n} (a_i^2 a_j^2 \sigma_{ii} \times \sigma_{jj})$$

$$= \sum_{i=1}^{n} (1/n)^4 (3\sigma^4) + 6\sum_{i=1}^{n-1}\sum_{j>i}^{n} (1/n)^2 (1/n)^2 \sigma^4$$

$$= 3\sigma^4/n^3 + 6\sum_{i=1}^{n-1}\sum_{j>i}^{n}\sigma^4/n^4 = 3\sigma^4/n^3 + (6\sigma^4/n^4)({}_nC_2)$$

$$= 3\sigma^4/n^3 + (\sigma^4/n^4)6 \times n(n-1)/2$$

$$= 3\sigma^4/n^3 + 3(n-1)(\sigma^4/n^3) = 3\sigma^4/n^2,$$

which yields $\alpha_4(\bar{x}) = \mu_4(\bar{x})/V^2(\bar{x}) = (3\sigma^4/n^2)/(\sigma^2/n)^2 = 3$ leading to the kurtosis $\beta_4(\bar{x}) = \alpha_4(\bar{x}) - 3 \equiv 0$, as expected. As has been known in the field of Statistics since the mid 1800's, we have again established that the sample arithmetic mean $\bar{x}$ has identical first origin moment and identical 3rd and 4th standardized central moments like its parent $X \sim N(\mu, \sigma^2)$; only the scale of $\bar{x}$ differs as given by $\sigma_{\bar{x}} = \sigma_X/\sqrt{n}$, $n > 1$. Further, it is well known that the ordinate (or height) of the mode (MO) for the $f(\bar{x} = MO; \mu, \sigma/\sqrt{n}) = \sqrt{n}/(\sigma\sqrt{2\pi})$ while that of $f(x = MO; \mu, \sigma) = 1/(\sigma\sqrt{2\pi})$ so that for $n > 1$ the SMD of $\bar{x}$ is more peaked in the middle than its parent *X*, yet both have identical kurtosis of $\beta_4(\bar{x}) = \beta_4(X) \equiv 0$ showing that both curves are Mesokurtic. Put differently, the fact that the two tail-thicknesses

$$Pr(|\bar{x} - \mu| > 3\sigma_{\bar{x}}) \equiv Pr(|X - \mu| > 3\sigma_X) = 0.0026997960633$$

is congruent with $\beta_4(\bar{x}) = \beta_4(X) \equiv 0$.

Because sample X_i's, $i = 1, 2, ..., n$ in a simple random sample are iid rvs with means μ and variances σ^2, then one form of the CLT states that the limiting, in terms of *n*, SMD of a general linear combination, $Y_n = \sum_{i=1}^{n} a_i X_i$, where a_i's are constants, approaches normality with mean $E(Y_n) = \mu\sum_{i=1}^{n} a_i$, and variance $V(Y_n) = \sigma^2 \sum_{i=1}^{n} a_i^2$, V being the nonlinear variance-operator. The rate of approach to normality depends strictly on the skewness and kurtosis of the individual X_i's [see Hool and Maghsoodloo [9]], each of which is identically distributed like the parent *X*. Equation (1) clearly shows that in the case of iid variates $\mu_3\left(S_n = \sum_{i=1}^{n} X_i\right) = n\mu_3(X)$, $V(S_n) = nV(X) = n\sigma_X^2$ leading to the skewness of $S_n = \sum_{i=1}^{n} X_i$ as $\alpha_3(S_n) = \mu_3(S_n)/[V(S_n)]^{3/2}$ $= n\mu_3(X)/[n\sigma^2]^{3/2} = \alpha_3(X)/\sqrt{n}$, and similarly the kurtosis of S_n is given by $\beta_4(S_n) = [\alpha_4(X_i) - 3]/n = \beta_4(X)/n$. For example, if X_i's are uniformly and independently distributed over the same interval, then $\alpha_3(S_n) \equiv 0$ and its kurtosis $\beta_4(S_n) = [\mu_4(X_i)/\sigma^4 - 3]/n = [(1/80)/(1/12)^2 - 3]/n =$ $(144/80 - 3)/n = (1.80 - 3)/n = -1.20/n$; because the $\beta_4(X) = -1.20$, then the Uniform distribution is Platykurtic, while no MO exits. Thus, for any symmetrical distribution, such as the Uniform, we have a perfect fit of the first 3 moments of S_n with those of the corresponding normal, and for the Uniform underlying distribution an $n = 10$ forces the kurtosis of $S_{10} = \sum_{i=1}^{10} X_i$ to the value of –0.120, which is sufficiently close to zero for an adequate normal approximation of $\sum_{i=1}^{10} X_i$ SMD. The exact Uniform $[-1/2, 1/2]$ convolutions for $n = 2, 3, 4, 5, 6, 8$ were obtained by Maghsoodloo and Hool ([17], pp. 1-13); unfortunately there are typos in the 8th convolution. The author will report the corrected version on his website. That article concluded that the 4-fold convolution of the Uniform density is fairly well approximated by the corresponding normal only from the standpoint of skewness and kurtosis, as has been well known, but the quintiles are poorly approximated unless $n > 8$. The exact 7 expressions for

7-fold convolution, $f_{(7)}(x)$, of the $U[a, b]$ was derived by Maghsoodloo and Helvaci ([16], p. 6), where $c = b - a$ was the base.

As another symmetrical example, the Laplace (1774) pdf, or double-exponential, is given by $g(x; \theta, \lambda) = \lambda \exp(-\lambda | x - \theta |)/2, -\infty < x < \infty, \lambda$ being positive-definite, has an $E(X) = \mu = \theta, V(X) = 2/\lambda^2$, the scale $\sigma_L = \lambda^{-1}\sqrt{2}$, and all odd central moments are zero because the 1st law of Laplace $g(x; \mu, \lambda)$ is symmetrical about $\theta = \mu$, and no shape parameter exists. However, $\alpha_4(X) = 6$ so that the kurtosis $\beta_4(X) \equiv 3$, implying that Laplace density is Leptokurtic; the tail probability $Pr(| X - \mu | > 3\lambda^{-1}\sqrt{2}) = e^{-3\sqrt{2}} = 0.0143696$, as compared to 0.0026998 of the $N(\mu, \sigma^2)$, confirms its tail-thickness. Although calculus cannot be used to obtain the modal point (the Laplace density has a corner at $x = \mu$), it can be argued that $MO = \mu = x_{0.50}$. The height of MO is $g(MO; \mu, \lambda) = \lambda/2 = 1/(\sigma_L\sqrt{2}) > 1/(\sigma_X\sqrt{2\pi})$, relative to their scales, again verifying that the Laplace density is Leptokurtic. As stated by Johnson et al. ([11], p. 164), the Laplace first law, $g(x; \mu, \lambda)$, is used to represent the statistical distributions of measurement errors. If the individual X_i's in $S_n = \sum_{i=1}^{n} X_i$ are iid like Laplace, then equation (1a) shows that $\alpha_3(S_n) \equiv 0$, and equation (1b) shows that $\beta_4(S_n) = [\mu_4(X_i)/\sigma^4 - 3]/n = (24\lambda^{-4}/\sigma_L^4 - 3)/n = (6 - 3)/n = 3/n$. Thus, an $n = 10$ forces the kurtosis of the Laplace linear sum $\sum_{i=1}^{10} X_i$ to 0.30, much farther from zero than the corresponding uniform of -0.12. An $n = 25$ is needed to attain the same normal approximation of Laplace $S_n = \sum_{i=1}^{n} X_i$ from the standpoint of kurtosis. The 8-fold convolution of the Laplace 1st law was derived by Maghsoodloo and Hool [[17], pp. 1-13, equation (15)].

On the other hand, if X_i's are independently and Exponentially distributed (all at the same identical rate-parameter λ), then it is well known that the n-fold convolution of $\sum_{i=1}^{n} X_i$ (which is the Erlang density, n being its shape-parameter) has a skewness of $2/\sqrt{n}$ and a kurtosis of $\beta_4 = 6/n$, where $\alpha_3(X_i) = 2$ and $\beta_4(X_i) = 6$ for each identical exponential variate in the $\sum_{i=1}^{n} X_i$. Thus, in this latter case an $n = 50$ is needed to force the kurtosis of S_n to 0.12. Unfortunately, this last $n = 50$ is not sufficient for approximate normality of the exponential S_n because the skewness of $\sum_{i=1}^{50} X_i$ equals to $2/\sqrt{50} = 0.282843,$ far from the normal-skewness of zero. It is well known that $\alpha_3(S_n)$ plays a far more important role than $\beta_4(S_n)$ in the case of asymmetrical (or skewed) distributions in the limiting approach to normality. Further, our experience indicates that both skewness and kurtosis of Y_n should be within 0.10 of zero for adequate normal approximation.

The symmetrical distributions Uniform and Laplace can also be used as underlying distributions in a Manufacturing setting, while the exponential density, that describes useful life, plays an enormous role in describing all Stochastic Processes in at least the following 2 cases. Consider a Poisson process at the rate of $\lambda = 1$ event/unit of time; then the interoccurrence time has the standard exponential with density e^{-x} and hence $\int_0^\infty e^{-x} \equiv 1;$ now make the transformation $x = \left(\dfrac{t-\delta}{\theta-\delta}\right)^{\beta}$ in this last integral. This will lead to the well-known Weibull density with guaranteed-life δ, characteristic-life θ $(\text{scale} = \theta - \delta)$ and shape (or slope) β. Second, the sum of identical exponentials $\sum_{i=1}^{n} X_i,$ as stated above, has the Erlang SMD (special case of

Gamma) which is used to describe the distribution of time to *n*th Poisson event (measured from the last).

2. Introduction

Consider one simple random sample of size *n* from a $N(\mu, \sigma^2)$, where both process mean μ and process standard deviation $\sigma = \sigma_X$ are unknown. There are numerous estimators of μ such as the sample arithmetic mean, sample median, the sample mode, mid-range, Geometric and Harmonic means, trimmed mean, etc. Because all Laplace-Gaussian distributions possess uncountably infinite quantiles, then there exists uncountable estimators of population mean μ for large sample sizes based on quantiles nearly all having been discussed in statistical literature since the mid 1700's [see Maghsoodloo and Huang [24] as to why $21 \leq n \leq 60$ constitutes a moderate sample size]. For example, similar to the point-estimator "Trimean = (the first sample quartile $+ 2 \times$ the sample median + the 3rd sample quartile)/4" given by Hogg and Tanis [8], the statistic $(\hat{x}_{0.10} + \hat{x}_{0.20} + \hat{x}_{0.30} + \hat{x}_{0.40} + 8\hat{x}_{0.50} + \hat{x}_{0.60} + \hat{x}_{0.70} + \hat{x}_{0.80} + \hat{x}_{0.90})/16, (n > 60),$ is also a point estimator of μ, where $\hat{x}_p$ is the *p*th sample-quantile from a $N(\mu, \sigma^2)$ whose population *p*th quantile is dented by x_p; however, estimating μ using either of these last two estimators must be avoided.

Before discussing all estimators of process standard deviation σ, we remind the reader about the concepts of efficiency and relative efficiency (REL-EFF) from statistical theory, e.g., see Bernard W. Lindgren ([15], 4th Ed., pp. 240-265).

The REL-EFF of an estimator $\hat{\theta}_1$ relative to another estimator $\hat{\theta}_2$, of the same parameter θ, is defined as $\text{REL-EFF}(\hat{\theta}_1, \hat{\theta}_2) = \text{MSE}(\hat{\theta}_2)/\text{MSE}(\hat{\theta}_1)$, where $\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = V(\hat{\theta}) + [B_{\hat{\theta}}(\theta)]^2$, and the parameter-function

$B_{\hat{\theta}}(\theta) = E(\hat{\theta}) - \theta,$ only a function of θ, is the amount of bias in $\hat{\theta}$ as a point-estimator of the parameter θ; when $B_{\hat{\theta}}(\theta) = 0, \hat{\theta}$ is said to be an unbiased estimator. According to Bernard W. Lindgren (2nd Ed., [14], pp. 274-5), an estimator of σ is absolutely (or 100%) efficient only if its variance attains the lower bound given by:

$$\frac{[1 + B'(\sigma)]^2}{I(\sigma)} \leq V(\hat{\sigma}), \tag{2}$$

where $B'(\sigma) = \partial B_{\hat{\sigma}}(\sigma)/\partial\sigma$ and $I(\sigma)$ is the amount of information in a simple random sample of size *n* about the parameter σ, given by $I(\sigma) = (2n)/\sigma^2$. A more lucid and easily-understood proof of Lindgren's inequality (2) is given in Appendix A for Applied Statisticians. Note that only for convenience we will use the notation $B(\hat{\theta})$ for $B_{\hat{\theta}}(\theta)$. [Chou [4, p.38] states that the concept of efficiency should be confined to *sufficient* statistics that are unbiased, a concept that is well outside the scope of this article, which is primarily written for Applied Statistics and QC Literatures. Fortunately, if a random sample of size *n* does not violate the normality assumption (say, a GOF-test p-vale > 0.15), then the 2×1 sample vector $[\bar{x} \quad S]^T$ is sufficient for parameter-estimation if the underlying distribution is normal; intuitively this means that practitioners can now throw away the original random sample of size *n* and no statistical information whatsoever will be lost about the corresponding $N(\mu, \sigma^2)$; also see Casella and Berger [2] (pp. 272-289). Further, we recommend to Quality Engineers to compute the sample range, *R*, before discarding the entire normal-data.

If $\hat{\sigma}$ is an unbiased estimator, then $B'(\sigma) = \partial B_{\hat{\sigma}}(\sigma)/\partial\sigma \equiv 0,$ and Lindgren's inequality in equation (2) reduces to the well-known Cramer-Rao Inequality that every unbiased estimator $\hat{\theta}$ is bounded below by the reciprocal of its information, i.e., $1/I(\theta) \leq V(\hat{\theta})$. It is well documented in

Statistical Literature [e.g., see Lindgren ([15], pp. 240-245)] that the Fisher Information Matrix (1912-1922) for a normal distribution is given by $IM(\mu, \sigma) = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & (2n)/\sigma^2 \end{bmatrix}$. This last Fisher's IM reveals the following when random-sampling a $N(\mu, \sigma^2)$. (1) the amount of information in a simple random sample improves as n increases and σ decreases (see the diagonal elements in the IM that also shows the information-amount about σ increases at twice the rate than about μ for $n > 1$). (2) the sample mean, $\bar{x}$, is 100% efficient because its variance $V(\bar{x}) = \sigma^2/n = 1/I(\mu)$, where $I(\mu) = n/\sigma^2$, $n \geq 1$. (3) any estimator of σ, denoted $\hat{\sigma}$, and the corresponding sample mean $\bar{x}$ are stochastically independent because the normal $\text{IM}(\mu, \sigma)$ is diagonal, i.e., the covariance between $\bar{x}$ and $\hat{\sigma}$ is identically zero. Consequently, two separate control charts S and $\bar{x}$ are used in the field of QC to monitor process variation and then process mean; it should be emphasized that getting the process mean on target in any process, while variation is out of statistical control, is a waste of resources [see Maghsoodloo et al. [18], pp. 73-126 and also Tatum ([22], p. 127)]. (4) the expression for the $V(S)$ has been documented both in Statistical and QC Literatures, but its variance does not attain the lower bound $[1 + B'(\sigma)]^2/I(\sigma)$, where $I(\sigma) = (2n)/\sigma^2$, $n \geq 2$, and hence S, unlike $\bar{x}$, is not an (absolutely or 100%) efficient estimator of σ. Chou ([4], p. 38) used the Lingren-Cramer-Rao Inequality and the MSE of a biased estimator to obtain a lower bound for the MSE of all σ-estimators for $M >$ subgroups. We used Chou's inequality on his p. 38,

$$\frac{[1 + B'(\sigma)]^2}{(2N/\sigma^2)} + [B_{\hat{\sigma}}(\sigma)]^2 \leq MSE(\hat{\sigma}),\ N = \sum_{i=1}^{M} n_i,$$

for $M > 1$ subgroups, to obtain the following (accuracy) efficiency for one simple random sample of size n (i.e., $M = 1$), in Chou's exact notation.

$$Eff(\hat{\sigma}) = \{[1 + B'(\sigma)]^2 \sigma^2/(2n)] + [B_{\hat{\sigma}}(\sigma)]^2\}/MSE(\hat{\sigma}) \le 1. \tag{3a}$$

The accuracy-efficiency defined in (3a) is somewhat misleading due the fact that if an estimator is heavily biased, its $Eff(\hat{\sigma})$ will be inflated because the MSE in the denominator will not compensate for the size of bias-squared in the numerator. We may use Lindgren's inequality (2) to define the precision-efficiency of $\hat{\sigma}$ as follows:

$$eff(\hat{\sigma}) = [1 + B'(\sigma)]^2 \times I^{-1}(\sigma)/V(\hat{\sigma}). \tag{3b}$$

The $eff(\hat{\sigma})$ in (3b) somewhat ignores the amount of bias in $\hat{\sigma}$. Perhaps, we should define the absolute efficiency of an estimator as:

$$EFF(\hat{\sigma}) = [Eff(\hat{\sigma}) + eff(\hat{\sigma})]/2. \tag{3c}$$

It is not clear to the author that the constant 0.50, depending on *n*, in the above convex combination in equation (3c) is optimal; perhaps 0.40 and 0.60 would be more optimal. All three efficiencies in equation (3) should be used with caution for a biased estimator. Nonetheless, if $\hat{\sigma}$ is unbiased, then $B_{\hat{\sigma}}(\sigma) = B'(\sigma) \equiv 0$, and equation (3c) reduces to that of Lindgren's equation (5), (1993, p. 262), for absolute (or 100%) efficiency defined as $e(\hat{\sigma}) = I^{-1}(\sigma)/V(\hat{\sigma}) = \sigma^2/[2n\,V(\hat{\sigma})]$.

It is widely known that the statistic $S^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)$ is an unbiased estimator of σ^2 iff (if and only if) the corresponding underlying population is infinite. However, $S = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)}$ is always a biased estimator of σ from any infinite population. In fact, both the Statistical and QC Literatures show that for $E(S) = c_4(n) \times \sigma$, where the unbiasing-factor $c_4(n) = c_4 = \sqrt{\frac{2}{n-1}} \times \frac{\Gamma(n/2)}{\Gamma[(n-1)/2]} = \sqrt{\frac{n-1}{2}} \times \frac{\Gamma(n/2)}{\Gamma[(n+1)/2]}$ is a function of n, $0.79788456 < c_4 < 1$ for all $2 \le n < 344$ [Matlab 2018 runs out of memory at $n > 343$], and the symbol $\Gamma(\cdot)$ stands for the Gamma function. We have verified that a good approximation for $c_4 \cong (4n^2 - 8n + 3.8775)/(4n^2 - 7n + 3)$, to 4 decimals for $n \geq 13$, the approximation improving with increasing *n*. It is well known (e.g., see Boyles [1], p. 391) that the amount of bias in *S* as an estimator of σ is given by $B(S) = E(S) - \sigma = (c_4 - 1)\sigma < 0$ for all $n \geq 2$, $V(S) = (1 - c_4^2) \times \sigma^2$ so that the Mean Square Error of *S* is $MSE(S) = E[(S - \sigma)^2] = V(S) + [B_S(\sigma)]^2 = 2(1 - c_4)\sigma^2$, where MSE measures the accuracy of an estimator, while variance *V* measures the precision of an estimator. The relationship $E(S) = c_4\sigma$ clearly shows that an unbiased estimator of σ is given by $\hat{\sigma}_{ub} = S/c_4$ so that the MSE of $\hat{\sigma}_{ub}$ is given by $MSE(\hat{\sigma}_{ub}) = V(\hat{\sigma}_{ub}) = (c_4^{-2} - 1)\sigma^2 > MSE(S)$ for all $n \geq 2$, showing that $\hat{\sigma}_{ub}$ is a less accurate estimator of σ than *S*. The REL-EFF $(\hat{\sigma}_{ub}, S) = 2c_4^2/(1 + c_4)$, which is equal to 70.819%, 91.091%, 98.141%, and 99.377% at $n = 2, 5, 21,$ and 61, respectively.

The exact SMD of *S* was first derived by Kendall and Stuart [12], Vol. 1, pp. 255-6 where their "*s*" is the scale-estimator from the $N(0, 1)$, and is also given by Ostle and Malone [19], pp. 80-81 for the general $N(\mu, \sigma^2)$, and referenced by Chou ([4], p. 10). Further, Chou [4] computed the specific quantiles of *S* from a $N(\mu, 1)$-population for $Q = 0.00135, 0.50000,$ and 0.99865 and are given in his Table 2.2 on p. 19; Chou ([4], p.24) used these last three quantiles to obtain the exact 3-sigma control limits for an S-chart. However, because the exact SMD of $S^2 \sim \sigma^2\chi^2_{n-1}/(n-1)$, nearly all Statistical properties of the variate *S* can be studied using the (Central) Chi-squared χ^2_{n-1}.

From Statistical Literature, the maximum likelihood estimator (*mle*) of σ is given by the square-root of the sample variance $m_2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/n$, this last being the sample 2nd central moment, i.e., $\hat{\sigma}_{mle} = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/n}$ $= \sqrt{m_2}$, the so called *root mean square*. Some Statistical authors refer to S^2 as the sample variance for convenience because it is an unbiased estimator of σ^2 only for infinite populations, but Kendall and Stuart's ([13], Vol. 2, designation on their pages 4-5 clearly states that the sample variance should be defined as the 2nd central moment of the sample $m_2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/n$, because the population second central moment is $\sigma^2 = \mu_2 = E[(X - \mu)^2]$; also see Lindgren ([15], p. 204). The amount of bias in $\hat{\sigma}_{mle}$ is $B(\hat{\sigma}_{mle})$ $= (C_4 - 1)\sigma < 0$, $C_4 = c_4\sqrt{(n-1)/n}$, for $n \geq 2$, its variance is given by $V(\hat{\sigma}_{mle}) = C_4^2(c_4^{-2} - 1)\sigma^2$, resulting in $MSE(\hat{\sigma}_{mle}) = (C_4^2 c_4^{-2} - 2C_4 + 1)\sigma^2$ $< MSE(S)$. This last inequality shows that $\hat{\sigma}_{mle}$ is a more accurate point estimator of σ than S, and that the REL-EFF $(\hat{\sigma}_{mle}, S) = 2(1 - c_4)/$ $(C_4^2 c^{-2} - 2C_4 + 1) > 1$ for all finite $n \geq 2$. The REL-EFF $(\hat{\sigma}_{mle}, S) =$ 108.775%, 101.288%, 100.071%, and 100.0084% at $n = 2, 5, 21,$ and 61, respectively.

Another estimator of σ from the field of QC is $\hat{\sigma} = R/d_2$, where $R = x_{(n)} - x_{(1)}$ is the sample range and $d_2 = E(W = R/\sigma)$, and $W = R/\sigma$ is the sample relative-range. The values of d_2 are reproduced in Table 1 for convenience and are also tabulated in every text on QC. Clearly, $\hat{\sigma} = R/d_2$ is an unbiased estimator of σ and hence its MSE is equal to its variance. The SMD of the sample range R is extremely complicated and is provided by Kendall and Stuart ([12], Vol. 1), pp. 337-341. Numerical integration must be used to compute both $d_2 = E(W)$, and the population standard deviation of relative-range $d_3 = \sigma_w$, at different values of $n \geq 2$. Nevertheless, $V(\hat{\sigma}) = V(R/d_2) = V(R)/d_2^2 = \sigma^2 V(R/\sigma)/d_2^2 = \sigma^2 \times d_3^2/d_2^2$, and $d_3^2 =$

$V(R/\sigma)$. The values of d_2 and d_3 were obtained by Tippett [23], Pearson [20], Pearson [21] and Harter [7]. Their results are also reproduced below in Tables 1 and 2. We used their values in Tables 1 and 2 in order to obtain the REL-EFF$(R/d_2, S) = MSE(S)/V(R/d_2)$ at $n = 2(1)20(5)35$ and are given in Table 3 in percent. Not surprisingly, Table 3 clearly shows that the estimator $\hat{\sigma} = R/d_2$ attains its maximum REL-EFF at the Nominal Sample Size $n = 5$ in the field of QC for R- & $\bar{x}$-charts. Because R/d_2 is not an efficient estimator, Harter ([6], pp. 1980-1999) discusses the use of Quasi-ranges in lieu of *R*, which is the zero-quasi range; the 1st quasi-range is $R_1 = x_{(n-1)} - x_{(2)}, n > 17$. Further, as Kendall and Stuart [12], p. 339, mention the SMD of sample range *R* diverges from normality as *n* increases, and that the SMD of *R* is unstable for moderate to large $n > 20$. Hence, the QC Literature uses R/d_2 as an estimator of σ_X only when subgroup sizes are identical and only for small sample sizes $2 \le n \le 15$. The REL-EFF $(R/d_2, S)$ monotonically decreases as it uses less and less sample information after $\mathbf{n = 5}$, and is a mere 44.2255% at $n = 61$. Only at $n = 2$, $V(R/d_2) = V(\hat{\sigma}_{ub})$, and $V(R/d_2) > V(\hat{\sigma}_{ub})$ only by 0.00821% at $n = 3$.

Table 1. The Expected-Value of Relative Range $(W = R/\sigma)$ of a $N(\mu, \sigma^2)$ from Tippett [23]

n	2	3	4	5	6	7	8	9	10	11
d_2	1.128379	1.692569	2.058751	2.325929	2.534413	2.704357	2.847201	2.970026	3.077505	3.172873
n	12	13	14	15	16	17	18	19	20	21
d_2	3.258455	3.335980	3.406763	3.471827	3.531983	3.587884	3.640064	3.688963	3.734950	3.778336
n	22	23	24	25	26	27	28	29	30	35
d_2	3.819385	3.858323	3.895348	3.930629	3.964316	3.996539	4.027414	4.057044	4.085522	4.213219

Table 2. The Standard Deviation of $W = R/\sigma$ of a Normal Universe from Table 2 of Harter [7]. Harter's values were truncated from 10 to 6 decimal accuracies

N	2	3	4	5	6	7	8	9	10	11
d_3	0.852502	0.888368	0.879808	0.864082	0.848040	0.833205	0.819831	0.807834	0.797051	0.787315
N	12	13	14	15	16	17	18	19	20	21
d_3	0.778478	0.770416	0.763023	0.756211	0.749908	0.744052	0.738591	0.733481	0.728686	0.724173
N	22	23	24	25	26	27	28	29	30	35
d_3	0.719915	0.715887	0.712068	0.708441	0.704988	0.701697	0.698553	0.695546	0.692665	0.679871

Table 3. The REL-EFF $(R/d_2, S)$ in percent to two decimals

n	2	3	4	5*	6	7	8	9	10	11	12
REL-EFF	70.82	82.60	86.17	**86.97***	86.58	85.61	84.35	82.96	81.52	80.07	78.63
n	13	14	15	16	17	18	19	20	25	30	35
REL-EFF	77.23	75.87	74.55	73.28	72.05	70.88	69.74	68.65	63.78	59.71	56.26

3. Estimates of σ Based on Quantiles

There are uncountable infinite estimators of σ [See also Kendall and Stewart ([12], p. 239)] because of the fact that the *p*th quantile of a $N(\mu, \sigma^2)$ is given by $x_p = \mu + Z_{1-p} \times \sigma$, where Z_{1-p} is the $(1-p) \times 100$ percentage point, or the *p*th quantile of the $N(0, 1)$ density. For example, the 0.90 quantile of X is given by $x_{0.90} = \mu + Z_{0.10} \times \sigma$, where $Z_{0.10} =$ 1.28155157 is the 0.90-quantile of $Z \sim N(0, 1)$. All Statistical Packages, such as Minitab, also Matlab and Microsoft Excel, provide the inverse (or quantile) functions of nearly all Statistical distributions that are not directly invertible to 15 decimal accuracy, albeit, Minitab18 [in their Calc Menu → Probability Distributions] has a slight discrepancy starting in the 11th decimal place in the $\Phi^{-1}(1-\alpha) = Z_\alpha$, $0 < \alpha < 1$. At first glance, $x_p = \mu + Z_{1-p} \times \sigma$ leads to the biased estimator $\hat{\sigma}_Q = (\hat{x}_p - \bar{x})/Z_{1-p}$; however, computing the $V[(\hat{x}_p - \bar{x})/Z_{1-p}]$ requires the knowledge of $COV(\hat{x}_p, \bar{x})$, which is beyond our grasp; also, the variance of the estimator $(\hat{x}_p - \hat{x}_{0.50})/Z_{1-p}$ shows that it has very poor REL-EFFs.

Consequently, we use the fact that $x_{1-p} = \mu + Z_p \times \sigma$ resulting in $x_p - x_{1-p} = (Z_p - Z_{1-p}) \times \sigma$, $0.50 < p < 1$, $Z_p > Z_{1-p}$, which covers $(2p-1)$ proportion of the population. Statistical Literature refers to $x_p - x_{1-p}$ as the population IPR (Inter-Percentile Range). Thus, the percentile estimator $\hat{\sigma}_p = (\hat{x}_p - \hat{x}_{1-p})/(Z_{1-p} - Z_p) = ipr_p/(Z_{1-p} - Z_p)$, where $ipr_p = \hat{x}_p - \hat{x}_{1-p} > 0$ is the corresponding sample inter-percentile range $0.50 < p < 1$, and $\hat{x}_p$ is the pth sample quantile. Sample quantiles can also be obtained from sample order-statistics $X_{(i)}$, whose sampling distributions depend on the parental density $f(x)$. It is well known that a maximum of 2 order-statistics are used in computing $\hat{x}_p$ such that, as also stated by Kendall and Stuart ([12], near the bottom p. 236), $\hat{x}_p$ is biased and the amount of bias $B(\hat{x}_p) = E(\hat{x}_p) - x_p$ diminishes as n increases to order of $1/n$. Kendall and Stuart ([12], Vol. 1) further obtained the approximate variance of a sample quantile for any continuous underlying distribution, for $n > 20$, as $V(\hat{x}_p) \cong pq/[nf^2(x_p)]$, where $q = 1 - p$ and their $f(x_p)$ is the ordinate (or height) of the parent density at x_p. This last approximation, given in their equation (10.29) on p. 237, leads to the well-known formula for $V(\hat{x}_{0.50}) = V$ (the sample median) $\cong \pi\sigma^2/(2n) = \pi V(\bar{x})/2$, $p = 0.50$.

Kendall and Stuart ([12], pp. 236-9) proceed to obtain the variance of difference of two distinct sample quantiles, x_1 and x_2, and their equation in their notation is repeated below: "$\text{var}\,\delta = \dfrac{1}{n}\left\{\dfrac{p_1 q_1}{f_1^2} + \dfrac{p_2 q_2}{f_2^2} - \dfrac{2p_2 q_1}{f_1 f_2}\right\}$",

where $\delta = x_1 - x_2$, f_1 is the ordinate of density at the quantile x_1, $q_2 =$

$1 - p_2$. We used their above result to obtain $V(ipr_p) = V(\hat{x}_p - \hat{x}_{1-p}) \cong 2q(p-q)/[f(x_p)\sqrt{n}]^2 = 4\pi(3p - 2p^2 - 1)e^{Z_p^2} n^{-1} \times \sigma^2$, $p > q = 1 - p$, where $f_1 = f(x_p) = f(x_{1-p}) = Exp(-Z_p^2/2)/(\sigma\sqrt{2\pi})$, and $Z_p = -Z_{1-p}$; e.g., $Z_{0.75} = -0.67448975$. As a result, $V(\hat{\sigma}_p) = V[(\hat{x}_p - \hat{x}_{1-p})/(2Z_{1-p})] = [\pi(3p - 2p^2 - 1)e^{Z_p^2}](nZ_p^2)^{-1} \times \sigma^2$. For examples, based on sample IQR the estimator $\hat{\sigma}_{0.75} = (\hat{x}_{0.75} - \hat{x}_{0.25})/(2Z_{0.25})$, $p = 0.75$, has the well-known $V(\hat{\sigma}_{0.75}) \equiv V[(\hat{x}_{0.75} - \hat{x}_{0.25})/(2Z_{0.25})] \cong 1.360460n^{-1} \times \sigma^2$, the population $IPR_{0.80} = x_{0.80} - x_{0.20} = 2Z_{0.20\times\sigma}$ leads to the estimator $\hat{\sigma}_{0.80} = (\hat{x}_{0.80} - \hat{x}_{0.20})/(2Z_{0.20})$ having an approximate variance of $(1.080737/n) \times \sigma^2$, the interdecile estimator $\hat{\sigma}_{0.90} = (\hat{x}_{0.90} - \hat{x}_{0.10})/(2Z_{0.10})$ has an approximate variance of $(0.790755/n)\sigma^2$, while $\hat{\sigma}_{0.95} = (\hat{x}_{0.95} - \hat{x}_{0.05})/(2Z_{0.10})$ has an approximate variance of $(0.781827/n)\sigma^2$. As the sample-coverage increases from 50%, 60% 80%, to 90%, the corresponding $V(\hat{\sigma}_p)$ decreases, and then starts increasing for all coverages (0.90, 1]; at 99% coverage the $V(\hat{\sigma}_{0.99}) = (1.784103/n)\sigma^2$. Although both estimators $\hat{x}_p$ and $\hat{x}_{1-p}$ of x_p and x_{1-p}, respectively, are biased, the sign of bias in them should be the same, and thus the sample $ipr_p = \hat{x}_p - \hat{x}_{1-p}$, $p > 0.50 > 1 - p$, should suffer from less bias than individual sample quantiles. If sample order-statistics are used to compute $\hat{x}_p$, then more sample-information (due to ordering) is used than just 2 order-statistics to compute $\hat{x}_p$, then $MSE(\hat{\sigma}_p) \approx V(\hat{\sigma}_p)$ for moderate to large samples ($n > 20$). The REL-EFF$(\hat{\sigma}_{0.95}, S) = V(S)/V(\hat{\sigma}_{0.95})$ =92.957%, 74.458%, 66.291%, 64.746% at $n = 2, 5, 21, 61$, respectively.

Minitab 18 provides another estimator of σ^2 from Boyles ([1], p. 383) in their "Display Descriptive Statistics" menu called the Mean of Successive

Squared Differences, which they denote by MSSD, and is defined as: Boyles's MSSD $= \frac{1}{2(n-1)}\sum_{i=1}^{n-1}(x_{i+1}-x_i)^2$, where x_i is the ith random observation (not the ith order-statistic $x_{(i)}$) in the sample of size $n \geq 2$. It can be shown that Boyles's MSSD is an unbiased estimator of σ^2, and that the point-estimator of σ given by

$$\hat{\sigma}_{SSD} = \sqrt{\sum_{i=1}^{n-1}(x_{i+1}-x_i)^2/2(n-1)},$$

where SSD stands for Successive-Squared-Differences, is negatively biased. Clearly, the successive differences $SD_i = (x_{i+1}-x_i)$, have $E(SD_i) \equiv 0$ and $V(SD_i) \equiv 2\sigma^2$, but

$$(x_2-x_1)^2/(2\sigma^2), (x_3-x_2)^2/(2\sigma^2), ..., (x_n-x_{n-1})^2/(2\sigma^2),$$

except in the trivial case $n = 2$, are obviously heavily and hopelessly auto-correlated, and hence the $\sum_{i=1}^{n-1}(x_{i+1}-x_i)^2/(2\sigma^2)$ does not have a χ^2 distribution, so that the exact variance of $\hat{\sigma}_{SSD}$ is unknown to the author; we doubt the SMD of $\hat{\sigma}_{SSD}$ can be obtained using Noncentral χ^2.

We used the concept inherent in the above Boyles's MSSD in order to obtain another σ-estimator whose Statistical properties can easily be studied. Let n be an even integer and define $\hat{\sigma}^2_{even} = \frac{1}{n}\sum_{i=1}^{n/2}(x_{2i}-x_{2i-1})^2$, so that the successive differences $x_2 - x_1$, $x_4 - x_3$, ..., $x_n - x_{n-1}$ are now mutually independent, although this last estimator is not an average. Therefore, $(x_{2i}-x_{2i-1})^2/(2\sigma^2)$ follows a χ_1^2 and $\sum_{i=1}^{n/2}(x_{2i}-x_{2i-1})^2/(2\sigma^2)$ follows the Central $\chi^2_{n/2}$. We must state that for odd n this last estimator cannot be defined unless either the 1st or the nth-order statistic (the one that is closest to $\bar{x}$ and/or the median) is truncated from the sample. It can easily be shown that $E(\hat{\sigma}^2_{even}) = \sigma^2$ so that $\hat{\sigma}^2_{even}$ is an unbiased estimator of σ^2. However, $\hat{\sigma}_{even} = \sqrt{\frac{1}{n}\sum_{i=1}^{n/2}(x_{2i} - x_{2i-1})^2}$ is a biased estimator whose amount is given by $B(\hat{\sigma}_{even}) = (1 - \varsigma_4) \times \sigma < 0$, $\varsigma_4 = \frac{2 \times \Gamma[(n+2)/4]}{\sqrt{n} \times \Gamma(n/4)}$, for $n = 2, 4, 6, 8, 10, \ldots$.

Consequently, the $V(\hat{\sigma}_{even}) = E(\hat{\sigma}^2_{even}) - [E(\hat{\sigma}_{even})]^2 = (1 - \varsigma_4^2) \times \sigma^2$ resulting in $MSE(\hat{\sigma}_{even}) = 2(1 - \varsigma_4) \times \sigma^2$ (see Appendix B). At $n = 2$, the value of $\hat{\sigma}_{even}$ is identical to S, and for $n > 2$, the estimator $\hat{\sigma}_{even}$ monotonically uses less sample-information than S, and hence its REL-EFFs decrease with increasing n; the REL-EFFs$(\hat{\sigma}_{even}, S) = 100\%, 61.599\%$, 52.716%, and 50.927% at $n = 2, 6, 22,$ and 62, respectively. Further, for all $n > 343$, REL-EFF$(\hat{\sigma}_{even}, S) \approx 50\%$, but never less than 50%.

Another point estimator of σ, proposed by Kendall and Stuart ([12], p. 239) is $\hat{\sigma}_{md} = MD \times \sqrt{\pi/2}$, where the sample mean-deviation (or average-deviation) is given by $AD = MD = \sum_{i=1}^{n} |x_i - \bar{x}|/n = \sum_{i=1}^{n} \text{abs}(x_i - \bar{x})/n$. In Appendix C, we have provided a proof for their recommendation showing that $V(\hat{\sigma}_{md}) \cong (0.50\pi - 1)n^{-1} \times \sigma^2$, to the order of $1/\sqrt{n}$. The three well-known authors Johnson et al. ([10], Vol. 1, also clearly state atop their p. 91 that $E(MD) = \sigma\sqrt{2/\pi}$, also leading to the above estimator $\hat{\sigma}_{md}$. [Further, these last 3 authors provide the values of $E(MD)$ for nearly all underlying distributions; e.g., for all Uniform (or Rectangular) distributions they give $E(MD) = \sigma\sqrt{3}/2$ on p. 279 of their *Volume* 2 (1995), while for the *Laplace* distribution they give $\sigma/\sqrt{2}$]. The Appendix C shows that a rough approximation for the amount of bias in $\hat{\sigma}_{md}$ is given by $B(\hat{\sigma}_{md}) \approx \omega n^{-0.40}\sigma$, $\omega = \sqrt{2/\pi} - 1$, $\text{MSE}(\hat{\sigma}_{md}) \approx (0.50\pi - 1 + \omega^2 n^{0.20})n^{-1} \times \sigma^2$ and its $\text{REL-EFF}(\hat{\sigma}_{md}, S) \cong \text{MSE}(S)/\text{MSE}(\hat{\sigma}_{md})$ is maximum at $n = 2$ equaling 130.878%, diminishes to 95.692% at $n = 5$, and monotonically decreases with n; its limiting value is roughly 71.40%. As long as data contains no mild outliers, i.e., $R < 4 \times S$, then $\hat{\sigma}_{md}$ generally exceeds the corresponding S by at most 5%.

Finally, because the ordinate of normal mode, $f(MO)$, is equal to $1/(\sigma\sqrt{2\pi})$, and if a sample histogram $(n > 60)$ unimodal frequency per unit of X, $\hat{f}(MO)$, can be obtained, then a rough point estimate of σ_X for grouped data is given by $\hat{\sigma} \approx [\hat{f}(MO) \times \sqrt{2\pi}]^{-1}$; this last estimator must be avoided, as other superior estimators can easily be computed.

4. Obtaining the Most Accurate Estimator of σ

Because the Cramer-Rao lower bound for an unbiased estimator of σ is $1/I(\sigma) = 1/[(2n)/\sigma^2] = \sigma^2/(2n)$, $n \geq 2$, we used this last information to set $\sigma^2/(2n)$ equal to $k \times V(S/c_4) = k(c_4^{-2} - 1)\sigma^2$, which led to the multiplier $k = c_4^2/[2n(1 - c_4^2)]$ for $V(S/c_4)$. Therefore, $k^{1/2}$ will force the following estimator to become most precise (i.e., smallest variance)

$$\hat{\sigma}_{sv} = \frac{c_4}{\sqrt{2n(1 - c_4^2)}} \times (S/c_4) = \frac{S}{\sqrt{2n(1 - c_4^2)}} = k_n \times S, \qquad (4a)$$

where $0.82944 < k_n = [2n(1 - c_4^2)]^{-1/2} < 1$ for all $n \geq 2$; one can apply the variance-operator to (4a) and verify that $V(\hat{\sigma}_{sv}) = \sigma^2/(2n) = 1/I(\sigma)$. Clearly, $\hat{\sigma}_{sv}$ is a biased estimator and its negative bias-amount is given by

$$B_{\hat{\sigma}_{sv}}(\sigma) = (c_4 k_n - 1) \times \sigma < 0, \text{ for all } n \geq 2. \tag{4b}$$

Unfortunately, equation (4b) shows that $\hat{\sigma}_{sv}$ underestimates σ in repeated sampling each of size *n*, and its use as an estimator of σ must be avoided because it is not a conservative estimator of σ, that is, $E(\hat{\sigma}_{sv}) << \sigma$, nevertheless, it is statistically efficient. Further, at $n = 5$, the contribution of bias to the $\text{MSE}(\hat{\sigma}_{sv})$ is 14.24%, and hence unacceptable. Note that the alternative of setting $\frac{[1 + B'(\sigma)]^2}{I(\sigma)}$ equal to $V(\hat{\sigma} = b_n \times S)$, where b_n is a function of $n > 1$, will lead to a contradiction and no solution for b_n. It seems no σ-estimator will ever exit whose Variance attains the Lindgren lower bound $\frac{[1 + B'(\sigma)]^2}{I(\sigma)}$ and also has minimum MSE. So, we had to modify equations (4a and b) in order to obtain the most possible accurate estimator of σ as shown below. We used the expressions in equations (4a and b) as follows. Because the bias-parameter $B_{\hat{\sigma}_{sv}}(\sigma)$ in equation (4b) is large, we first reduced its size such that our estimator will have the amount of bias given by $(c_4 k_n - c_4) \times \sigma$, i.e., the $B_{\hat{\sigma}_X}(\sigma) = c_4(k_n - 1) \times \sigma$, but no longer with minimum variance. Note that the alternative of letting $B(\hat{\sigma}_X) = (c_4 k_n - c) \times \sigma$, where the constant $0 < c < 1$, will lead to the same near-optimum obtained below. By definition, $B(\hat{\sigma}_X) = E(\hat{\sigma}_X) - \sigma = c_4(k_n - 1) \times \sigma$, leading to $E(\hat{\sigma}_X) = c_4(k_n - 1) \times \sigma + \sigma = [c_4(k_n - 1) + 1] \times \sigma = C_5(n) \times \sigma$, where the QC constant $C_5 = c_4(k_n - 1) + 1$ is only a function of *n*, and $0.8639 < C_5 < 1$ for $2 \leq n < \infty$. Comparing $E(\hat{\sigma}_X) = C_5 \times \sigma$ with $E(S) = c_4 \times \sigma$ leads to our estimator

$$\hat{\sigma}_X = C_5 \times (S/c_4) = C_5 \times (\hat{\sigma}_{ub}). \tag{5a}$$

The estimator in equation (5a) has a larger REL-EFF than others for all $n > 5$. The amount of bias is $B(\hat{\sigma}_X) = (C_5 - 1) \times \sigma, V(\hat{\sigma}_X) = C_5^2(c_4^{-2} - 1)\sigma^2$ leading to the $\text{MSE}(\hat{\sigma}_X) = (1 - 2C_5 + C_5^2 c_4^{-2}) \times \sigma^2$. It is widely known that the $n = 5$ is the Nominal Sample Size in the field of QC; this required that we optimize C_5 such that $\text{MSE}(\hat{\sigma}_X)$ will be less than all other estimators for $n \geq 2$. Clearly, as C_5 increases toward 1, the size of $B(\hat{\sigma}_X)$ decreases but the $V(\hat{\sigma}_X)$ increases. Thereby, we redefined $C_5 = c_4(k_n - 1) + 1$ as

$$C_5(n) = c_4(k_n - 1) + c_5(n). \tag{5b}$$

Differentiating $\text{MSE}(\hat{\sigma}_X) = (1 - 2C_5 + C_5^2 c_4^{-2})\sigma^2$ with respect to C_5 and setting it to zero yields a fairly accurate minimum point $C_5^0 = c_4^2$, the 2nd derivative being positive-definite; it must be stated that this last optimum is not exact because minimizing $\text{MSE} = V + B^2$ will not necessarily simultaneously minimize both V and bias-squared. Further, it must be stated that this last optimum $C_5^0 = c_4^2$ is consistent with that of Vardeman's [25] who used a completely different approach to arrive at the optimum value of C_5. Substituting this last near-optimal C_5^0 into (5b) results in $c_4^2 = c_4(k_n - 1) + c_5^0$; hence, $c_5^0 = c_4(c_4 - k_n + 1)$. The corresponding $B(\hat{\sigma}_X) = (c_2^4 - 1) \times \sigma$, $V(\hat{\sigma}_X) = c_4^2(1 - c_4^2)\sigma^2$, $\text{MSE}(\hat{\sigma}_X) = (1 - c_4^2) \times \sigma^2$ $= V(S)$. The near optimum $\text{REL-EFF}(\hat{\sigma}_X, \hat{\sigma}_{mle}) = \dfrac{\text{MSE}(\hat{\sigma}_{mle})}{\text{MSE}(\hat{\sigma}_X)} = 102.268$, 101.783, 100.553, and 100.200% at $n = 2, 5, 21$, and 61, respectively, and the corresponding $\text{REL-EFF}(\hat{\sigma}_X, S)$ are 111.242, 103.094, 100.625, and 100.208%.

6. Efficiencies of σ-Estimators

As stated before, the Chou's $\mathit{Eff}(\hat{\sigma}_{mle})$ is inflated by bias-squared because $|B(\hat{\sigma}_{mle})|/\sigma > |B(\hat{\sigma}_X)|/\sigma > |B(S)|/\sigma$ for all $n \geq 2$; for example, at $n = 2, 5, 21$, and 61, the respective 3 relative-biases are (−0.43581, −0.36608, −0.20212), (−0.15925, −0.11643, −0.06001), (−0.03622, −0.02468, −0.01242), and at $n = 61$ they are (−0.01235, −0.00830,

-0.00416). Consequently, for comparative purposes we provide both absolute EFFs from equations (3a and b). Inserting $B'_S(\sigma) = c_4 - 1$, $\text{MSE}(S) = 2(1 - c_4)\sigma^2$ into equation (3a) yields the (absolute) efficiency $Eff(S) = [c_4^2/2n + (c_4 - 1)^2]/[2(1 - c_4)]$. Similarly, $Eff(\hat{\sigma}_{ub}) = c_4^2/[2n/(1 - c_4^2)]$, $Eff(\hat{\sigma}_{mle}) = \dfrac{C_4^2/(2n) + (1 - C_4)^2}{C_4^2 c_4^{-2} - 2C_4 + 1}$, $Eff(R/d_2) = \dfrac{d_2^2/2n}{d_3^2}$,

$Eff(\hat{\sigma}_{md}) \approx \dfrac{(n^{0.40} + \omega)^2 + 2n\omega^2}{(\pi - 2)n^{0.80} + 2n\omega^2}$, and $Eff(\hat{\sigma}_X) = \dfrac{C_5^2/2n + (1 - C_5)^2}{1 - 2C_5 + C_5^2 c_4^{-2}}$.

From equation (3b) the Lindgren precision-efficiencies are $eff(S) = eff(\hat{\sigma}_{ub}) = eff(\hat{\sigma}_{mle}) = eff(\hat{\sigma}_X) = c_4^2/[2n(1 - c_4^2)]$, $eff(R/d_2) = \dfrac{d_2^2/2n}{d_3^2}$ as expected, and $eff(\hat{\sigma}_{md}) = \dfrac{(1 + \omega n^{-0.40})^2}{\pi - 2}$.

7. Ranking the Six Estimators of σ

We rank only the estimators of σ that are highly pertinent to Applied Statistics and the field of QC. We will rank them from the standpoint of their precision, accuracy, and then the absolute efficiency $Eff(\hat{\sigma})$ from equation (3a).

(i) From the standpoint of Precision (Smaller variance):

(1) $\hat{\sigma}_{mle} = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/n}$, (2) $\hat{\sigma}_X = C_5 \times (S/c_4)$,

(3) $S = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)}$, (4) $\hat{\sigma}_{ub} = S/c_4$,

(5) $\hat{\sigma}_{md} = \sqrt{\pi/2} \times \sum_{i=1}^{n}|x_i - \bar{x}|/n$,

(6) $\hat{\sigma}_x = R/d_2$; note that for $2 \le n < 10$, the above rankings of S, $\hat{\sigma}_{ub}$ and $\hat{\sigma}_{md}$ change somewhat.

(ii) From the standpoint of accuracy (Smaller MSE):

(1) $\hat{\sigma}_X = (C_5/c_4) \times S$, (2) $\hat{\sigma}_{mle}$,

(3) S, (4) $\hat{\sigma}_{ub}$,

(5) $\hat{\sigma}_{md}$,

(6) $\hat{\sigma}_x = R/d_2$. For $n = 2, 3, 4$, $\hat{\sigma}_{md}$ has a smaller MSE than S.

(iii) (1) $\hat{\sigma}_{mle}$, (2) $\hat{\sigma}_X = (C_5/c_4) \times S$,

(3) S, (4) $\hat{\sigma}_{ub}$,

(5) $\hat{\sigma}_{md}$,

(6) $\hat{\sigma}_x = R/d_2$. All three above rankings show that $\hat{\sigma}_{ub}$ ranks only 4th relative to the others.

An Excel file has also been developed that computes all REL-EFFs and absolute efficiencies discussed in this article and will be available on request.

7. Summary and Conclusions

We discussed the CLT from the standpoint of skewness and kurtosis, and determined that both must be within 0.10 of zero for the adequate quantile approximation of $S_n = \sum_{i=1}^{n} X_i$ *SMD*. Further, in great detail, we examined the statistical properties of all possible estimators of the normal standard deviation σ and determined that only 6, with bias ≤ 0, have practical applications. We also obtained the estimator $\hat{\sigma}_X = (C_5/c_4) \times (S)$ that is the most accurate of all others. As has been known in Statistical Literature the *mle* is the most precise estimator (smallest variance), but our Section 6 shows that its Lindgren's precision-efficiency is the same as $\hat{\sigma}_X$, S, and $\hat{\sigma}_{ub}$. The size of bias in the estimator $S < 0.001$ for $n > 250$ and its Lindgren's absolute efficiency $e(S) = I^{-1}(\sigma)/V(S) = 99.7815\%$ at $n = 343$.

Appendix A

The likelihood function $L(\mathbf{X};\, \boldsymbol{\theta}) = \prod_{i=1}^{n} f(X_i;\, \theta)dx_i$, where the $n \times 1$ observation-vector $\mathbf{X} = [X_1\ X_2 \cdots X_n]^T$, X_i's are prior random components of the sample vector **X**, and the $m \times 1$ parameter-vector $\boldsymbol{\theta} = [\theta_1\ \theta_2 \cdots \theta_m]^T$, e.g., the Uniform has at most 2, the normal has 2, the Weibull has at most 3, and the Laplace distribution has at most $m = 2$ parameters. The *Pr* element $f(X_i;\, \boldsymbol{\theta}) \times dx_i$ gives the prior *Pr* that the random observation X_i will lie within the random interval of length dx_i around X_i. In order to obtain the *mle* of the vector parameter θ, the $\frac{\partial}{\partial \boldsymbol{\theta}} \prod_{i=1}^{n} f(X_i;\, \boldsymbol{\theta})dx_i$ must be set to zero and the resulting *m* equations in *m* unknowns must be solved simultaneously in order to obtain the $m \times 1$ *mle* estimator $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1\ \hat{\theta}_2 \cdots \hat{\theta}_m]^T$. This last procedure is extremely awkward and cumbersome to carry out. As a result, Statistical Theory has used the fact that if $y(x)$ is any function of *x*, then $\frac{\partial}{\partial x}\log[y(x)] = \frac{1}{y(x)} \times [\partial y(x)/\partial(x)] = 0$ iff $\partial y(x)/\partial(x) = 0,\ y(x) \neq \infty;$ the most common base values for the log-function are *e* and 10. Statistical Theory always uses the natural-log, denoted by $Ln = \log_e$, i.e., the function $Ln\prod_{i=1}^{n} f(X_i;\, \boldsymbol{\theta})dx_i$ is always maximized, which can also be expressed as

$$Ln\prod_{i=1}^{n} f(X_i;\, \boldsymbol{\theta})dx_i = \sum_{i=1}^{n} Ln[f(X_i;\, \boldsymbol{\theta})] + \sum_{i=1}^{n} Ln(dx_i).$$

Only because $\sum_{i=1}^{n} Ln(dx_i)$ is completely free and independent of θ and will always dropout after differentiation with respect to θ, Statistical Theory makes further simplification by referring to log-likelihood as

$\sum_{i=1}^{n} Ln[f(X_i;\, \boldsymbol{\theta})] = Ln[L(\mathbf{X};\, \theta)],$ where $L(\mathbf{X};\, \boldsymbol{\theta}) = \prod_{i=1}^{n} f(X_i;\, \boldsymbol{\theta})$ is the joint density of the sample vector $\mathbf{X} = [X_1\ X_2 \cdots X_n]^T$. Further, because the sample vector **X** plays no role in partial-differentiation process, then $L[(\mathbf{X};\, \boldsymbol{\theta})]$ is denoted by $L(\boldsymbol{\theta})$, and hence the log-likelihood function simplifies to $Ln[L(\boldsymbol{\theta})]$. In summary, the *mle* estimator $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1\ \ \hat{\theta}_2\ \ \hat{\theta}_m]^T$ is obtained by setting $\frac{d}{d\boldsymbol{\theta}} Ln[L(\boldsymbol{\theta})]$ to zero and solving the resulting m equations in m unknowns parameters for the estimator $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1\ \ \hat{\theta}_2\ \ \hat{\theta}_m]^T$. From a practical standpoint, $\frac{d}{d\theta_i} Ln[L(\boldsymbol{\theta})]$ indicates how fast the likelihood function changes with respect to θ_i. Lindgren ([15], atop p. 241) refers to $\frac{\partial}{\partial\theta_i} Ln[L(\boldsymbol{\theta})]$ as the *score function*, i.e.,

$$SF_{\mathbf{X}}(\theta_i) = \frac{\partial}{\partial\theta_i} Ln[L(\boldsymbol{\theta})] = \frac{\partial L/\partial\theta_i}{L(\boldsymbol{\theta})} = \frac{L'(\theta_i)}{L(\boldsymbol{\theta})}, \tag{A1}$$

where prime denotes partial-differentiation with respect to any one parameter θ_i. For notational convenience let θ denote any one of the θ_i's. As was done in his Theorem 17 on Lindgren's p. 241, we also first show that $E[SF_{\mathbf{X}}(\theta)]$ is always zero for any vector parameter **θ**, dropping **X**.

Proof.

$$E[(SF(\theta))] = \iint \cdots \int \frac{L'(\boldsymbol{\theta})}{L(\boldsymbol{\theta})} \prod_{i=1}^{n} f(X_i;\, \boldsymbol{\theta})\, dx_1 \cdots dx_n$$

$$= \iint \cdots \int \frac{L'(\boldsymbol{\theta})}{L(\boldsymbol{\theta})} L(\boldsymbol{\theta})\, d\mathbf{X} = \iint \cdots \int L'(\boldsymbol{\theta})\, d\mathbf{X}$$

$$= \frac{\partial}{\partial\theta} \iint \cdots \int L(\boldsymbol{\theta})\, d\mathbf{X};$$

however $L(\boldsymbol{\theta})$ is the joint density of vector **X**, and its *n*-tuple integral over the entire *n*-dimensional space must identically equal to 1. That is, $E[(SF(\theta))] = \frac{\partial}{\partial\theta}(1) \equiv 0,$ as expected, i.e., the average rate of change in the likelihood function is zero.

Lindgren [15] on his p. 241 further states that the (Sir R. A. Fisher 1912-1922) amount of information in a random sample of size *n* about a single parameter θ as the variance of its *SF*, i.e., $I_{\mathbf{X}}(\theta;\, n) = V[SF(\theta)] = E(SF^2)$ $- E[(SF)]^2 = E(SF^2) - 0 = E(SF^2) = E\left\{\left[\frac{L'(\theta)}{L(\theta)}\right]^2\right\}.$ As Lindgren states after his equation (3) on p. 241, the notation for Fisher's $I_{\mathbf{X}}(\theta;\, n)$ could be misleading because the expected-value and variance operators have already averaged-out all the X_i's, i.e., $I_{\mathbf{X}}(\theta;\, n)$ is not a function of **X**; the subscript **X** in $I_{\mathbf{X}}(\theta;\, n)$ simply says the amount of information about the parameter θ in the observable vector $\mathbf{X} = [X_1\ X_2 \cdots X_n]^T$. As Lindgren states on his pp. 241-2, the Fisher sample information is additive, i.e., $I_{\mathbf{X}}(\theta;\, n) = n \times$ $I_{\mathbf{X}}(\theta;\, n = 1).$ As an example, the log-likelihood for a normal distribution, well documented in Statistical Literature in the past century, is given by

$$Ln[L(\mu,\, \sigma)] = Ln\prod_{i=1}^{n}\frac{Exp[-0.50(x_i - \mu)^2/\sigma^2]}{\sigma\sqrt{2\pi}}$$

$$= Ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n + \sum_{i=1}^{n}[-0.50(x_i - \mu)^2/\sigma^2]$$

$$= -nLn(\sigma) - nLn(\sqrt{2\pi}) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2;$$

thus the $SF(\sigma) = \frac{\partial}{\partial\sigma}Ln[L(\mu,\, \sigma)] = -n/\sigma + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2$ which is a function of the vectors $\mathbf{X}$ and $\boldsymbol{\theta} = [\mu \quad \sigma]^T$; further, because the $E\sum_{i=1}^{n}(x_i - \mu)^2 = n\sigma^2$, then it follows that this last $E[SF(\sigma)] \equiv 0$, as shown above for any estimator. In order to obtain the amount of information about σ in a simple random sample of size $n > 1$, we compute the variance of its *score function*:

$$V[SF(\sigma)] = V\left[-n/\sigma + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2\right] = \sigma^{-6}V\left[\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

$$= \sigma^{-6}\sum_{i=1}^{n}[V(x_i - \mu)^2] = \sigma^{-6}\sum_{i=1}^{n}\{E(x_i - \mu)^4 - [E(x_i - \mu)^2]^2\}$$

$$= \sigma^{-6}\sum_{i=1}^{n}\{3\sigma^4 - [\sigma^2]^2\} = \sigma^{-6}\sum_{i=1}^{n}2\sigma^4 = \sigma^{-6}2n\sigma^4$$

$$= 2n/\sigma^2 = I_{\mathbf{X}}(\sigma),$$

as was given in our Section 2. In order to arrive at the *information inequality* (see p. 262 of Lindgren) for the variance of any estimator, as given in equation (2), we use the fact that the correlation coefficient (of order zero) between any two variates is given by

$$-1 \le \rho(SF, \hat{\theta}) = \frac{COV(SF, \hat{\theta})}{\sqrt{V(SF) \times V(\hat{\theta})}} \le 1$$

and hence

$$0 \le \rho^2(SF, \hat{\theta}) = \left[\frac{COV(SF, \hat{\theta})}{\sqrt{V(SF) \times V(\hat{\theta})}}\right]^2 = \frac{COV^2(SF, \hat{\theta})}{V(SF) \times V(\hat{\theta})} \le 1. \qquad \text{(A2)}$$

Thus equation (A2) provides the variance lower bound for any estimator in the Universe, called the

Information Inequality:

$$\frac{COV^2(SF, \hat{\theta})}{V[SF(\theta)]} \le V(\hat{\theta}) < \infty. \qquad \text{(A3)}$$

For the $SF(\sigma)$, we already have $E[SF(\sigma)] = 0$, $V[SF(\sigma)] = 2n/\sigma^2$, and letting $\hat{\sigma} = \hat{\theta}$ be any point-estimator of σ, equation (A3) requires that we compute the $COV[SF(\sigma), \hat{\sigma}]$.

$$
\begin{aligned}
COV[SF(\sigma), \hat{\sigma}] &= E[\hat{\sigma} \times SF(\sigma)] - E[SF(\sigma)] \times E(\hat{\sigma}) \\
&= E[SF(\sigma) \times \hat{\sigma}] - 0 \times E(\hat{\sigma}) \\
&= E\left\{\left[-n/\sigma + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2\right] \times \hat{\sigma}\right\} \\
&= E\left\{\left[\frac{\partial}{\partial\sigma} LnL(\mu, \sigma)\right] \times \hat{\sigma}\right\} = E\left[\frac{L'(\mu, \sigma)}{L(\mu, \sigma)}\right] \times \hat{\sigma} \\
&= \iint \cdots \int \frac{L'(\sigma)}{L(\sigma)} \times \hat{\sigma} \times L(\mathbf{X}; \mu, \sigma)\, d\mathbf{X} \\
&= \iint \cdots \int L'(\sigma) \times \hat{\sigma} d\mathbf{X} = \frac{d}{d\sigma}\iint \cdots \int \hat{\sigma} \times L(\sigma)\, d\mathbf{X} \\
&= \frac{d}{d\sigma} E(\hat{\sigma}) = \frac{d}{d\sigma}[\sigma + B_{\hat{\sigma}}(\sigma)] = 1 + B'_{\hat{\sigma}}(\sigma) = 1 + B'(\sigma). \qquad \text{(A4)}
\end{aligned}
$$

Note that the expectation in the 4th step leading to (A4) cannot be carried out directly because it does not allow the use of $\int \frac{d}{dx} h(x) = \frac{d}{dx} \int h(x)$ iff $\int h(x)$ exists. Equation (A4) is identical to that of Lindgren ([14], p. 273). Substituting equation (A4) into the *Information Inequality* (A3) and using the fact that $V[SF(\sigma)] = 2n/\sigma^2$ we obtain the well-known *Information Inequality*:

$$
\frac{[1 + B'(\sigma)]^2}{I(\sigma)} = \frac{[1 + B'(\sigma)]^2}{2n/\sigma^2} \le V(\hat{\sigma}) < \infty.
$$

Lindgren ([15], p. 242) proceeded to derive a simpler formula for the $I_{\mathbf{X}}(\theta)$ as follows:

$$I_{\mathbf{X}}(\theta) = V(SF) = E\left[\frac{L'(\theta)}{L(\theta)}\right]^2 = E\left[\left(\frac{L'}{L}\right)^2\right]. \qquad \text{(A5)}$$

Now consider the rate of change of the *score function*:

$$\frac{\partial(SF)}{\partial\theta} = \frac{\partial}{\partial\theta}\left[\frac{\partial}{\partial\theta}LnL(\theta)\right] = \frac{\partial}{\partial\theta}\left[\frac{L'(\theta)}{L(\theta)}\right] = \frac{LL'' - L'L'}{L^2} = \frac{L''}{L} - \left(\frac{L'}{L}\right)^2. \quad \text{(A6)}$$

Applying the liner-operator E to equation (A6) and using A(5) results in $I_{\mathbf{X}}(\theta) = E\left[\left(\frac{L'}{L}\right)^2\right] = E\left(\frac{L''}{L}\right) - E\left[\frac{\partial(SF)}{\partial\theta}\right]$. However,

$$E\left(\frac{L''}{L}\right) = \iint\cdots\int \frac{L''(\theta)}{L(\theta)}L(\theta)d\mathbf{X} = \iint\cdots\int L''(\boldsymbol{\theta})d\mathbf{X} \equiv \frac{d^2}{d\theta^2}(1) \equiv 0.$$

Inserting zero for $E\left(\frac{L''}{L}\right) = 0$, the previous equation reduces to $I_{\mathbf{X}}(\theta) = E\left[\left(\frac{L'}{L}\right)^2\right] = 0 - E\left[\frac{\partial(SF)}{\partial\theta}\right]$. Hence, $I_{\mathbf{X}}(\theta) = V(SF) = -E\left[\frac{\partial(SF)}{\partial\theta}\right]$; this last is Lindgren's [15] equation (6) on his p. 242. For the $N(\mu, \sigma^2)$, the

$$SF(\sigma) = -n/\sigma + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2; \ \frac{\partial(SF)}{\partial\sigma} = n/\sigma^2 - 3\sigma^{-4}\sum_{i=1}^{n}(x_i - \mu)^2;$$

$$E\left[n/\sigma^2 - 3\sigma^{-4}\sum_{i=1}^{n}(x_i - \mu)^2\right] = n/\sigma^2 - 3\sigma^{-4}n\sigma^2 = -2n/\sigma^2$$

so that $I_{\mathbf{X}}(\theta) = -E\left[\frac{\partial(SF)}{\partial\theta}\right] = 2n/\sigma^2$, as before. As has been well known, the efficiency of an estimator is directly proportional to the amount of sample information used by the estimator.

In general, if $\hat{\theta}_i$ $(i = 1, 2, ..., m)$ are the MLEs of m parameters with the log-likelihood function $LnL(\mathbf{X}; \theta_1, \theta_2, ..., \theta_m)$, then the (i, j)th element of

Fisher's information matrix is given by

$$IM_{ij} = -E[\partial^2 LnL(x;\, \theta_1,\, \theta_2,\, ...,\, \theta_m)/\partial\theta_i \partial\theta_j].$$

Statistical theory, as repeated above, clearly shows that the asymptotic covariance matrix of the vector-estimator $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1\ \hat{\theta}_2\ \cdots\ \hat{\theta}_m]^T$ is given by the inverse of the information matrix IM, i.e.,

$$COV(\hat{\boldsymbol{\theta}}) = COV[\hat{\theta}_1\ \hat{\theta}_2\ \cdots\ \hat{\theta}_m]^T \cong IM^{-1};$$

we do not know exactly how large an n is needed for adequacy of this last approximation for estimators that are not 100% efficient from different underlying distributions.

Appendix B

For even n, consider the estimator $\hat{\sigma}_{even} = \sqrt{\sum_{i=1}^{n/2} (x_{2i} - x_{2i-1})^2 / n}$ whose bias-amount is given by $E(\hat{\sigma}_{even}) - \sigma,$ where

$$E(\hat{\sigma}_{even}) = E\sqrt{\frac{1}{n}\sum_{i=1}^{n/2}(x_{2i} - x_{2i-1})^2} = E\sqrt{\frac{2\sigma^2}{n}\sum_{i=1}^{n/2}[(x_{2i} - x_{2i-1})/\sigma\sqrt{2}]^2}$$

$$= \frac{\sigma}{\sqrt{n/2}} E\sqrt{\sum_{i=1}^{n/2} Z_i^2} = \frac{\sigma}{\sqrt{n/2}} E\sqrt{\chi^2_{n/2}}.$$

It is widely known that $E\sqrt{\chi^2_\nu} = \frac{\sqrt{2}\Gamma[(\nu + 1)/2]}{\Gamma(\nu/2)}$, where ν is the degrees of freedom of the χ^2_ν. Thus,

$$E(\hat{\sigma}_{even}) = \frac{\sigma}{\sqrt{n/2}} E\sqrt{\chi^2_{n/2}} = \frac{\sigma}{\sqrt{n/2}} \times \frac{\sqrt{2}\Gamma[(n/2 + 1)/2]}{\Gamma(n/4)}$$

$$= \frac{2\Gamma[(n + 2)/4]}{\sqrt{n}\Gamma(n/4)} \times \sigma = \varsigma_4(n) \times \sigma,$$

$$\varsigma_4(n) = 2\Gamma[(n+2)/4]/\sqrt{n}\Gamma(n/4).$$

Hence, the amount of bias is $B(\hat{\sigma}_{even}) = (\varsigma_4 - 1) \times \sigma$ for $n = 2, 4, 6, 8, 10, \ldots$. Clearly, the SMD of $\hat{\sigma}_{even}$ follows that of a $\frac{\sigma}{\sqrt{n/2}}\sqrt{\chi^2_{n/2}}$, or $\hat{\sigma}_{even}\sqrt{n}/\sqrt{2\sigma^2}$ follows a $\sqrt{\chi^2_{n/2}}$.

Appendix C

Applying the expected-value operator, we obtain $E(MD) = E(AD)$

$$\frac{1}{n}E\sum_{i=1}^{n}|x_i - \bar{x}| = \frac{1}{n}\sum_{i=1}^{n}E|x - \bar{x}| = \frac{1}{n}nE|x - \bar{x}|$$

$$= E|x - \bar{x}| = \int_{-\infty}^{\infty}|x - \bar{x}|f(x; \mu, \sigma)dx.$$

However, carrying out the exact integration of this last integral is beyond our grasp, and as has been done by numerous other authors, we will first make the following identity in the integrand:

$$E(AD) \equiv \int_{-\infty}^{\infty}|(x - \mu) - (\bar{x} - \mu)|f(x; \mu, \sigma)dx.$$

The exact expected value of $|(x - \mu) - (\bar{x} - \mu)|$ is out of our reach, and therefore, we argue as others that in simple random samples of moderate to large sizes $n > 20$, the value of $(x - \mu)$ is often much larger than $\bar{x} - \mu$, and hence $|(x - \mu) - (\bar{x} - \mu)| \cong |x - \mu|$ to the order of $n^{-1/2}$. In fact the values of $|x - \mu|$ is expected to exceed those of $|x - \bar{x}|$. Applying this approximation, we obtain

$$E(MD) = \int_{-\infty}^{\infty}|x - \bar{x}|f(x; \mu, \sigma)dx \le \int_{-\infty}^{\infty}|x - \mu|\phi(z)dx/\sigma,$$

so that $E(MD) \approx \int_{0}^{\infty}|x - \mu|\phi(z)dx/\sigma$, where $\approx$ denotes rough approximation. Using the definition of absolute values and making the transformation of $Z = (x - \mu)/\sigma$ results in

$$E(AD) \leq \int_{-\infty}^{0} -(x-\mu)\phi(z)\,dZ + \int_{0}^{\infty} (x-\mu)\phi(z)\,dZ$$

$$= 2\sigma \int_{0}^{\infty} Z \frac{e^{-Z^2/2}}{\sqrt{2\pi}}\,dZ = \frac{2\sigma}{\sqrt{2\pi}}\left[-e^{-Z^2/2}\right]_0^{\infty} = \frac{\sigma}{\sqrt{\pi/2}}.$$

This last result is also provided by other authors; for example, see Johnson et al. [10], Vol. 1, 2nd edition, p. 91. Setting this last approximate-expectation to the corresponding sample statistic $\frac{1}{n}\sum_{i=1}^{n} | x_i - \bar{x} |$ results in $\hat{\sigma}_{md} = \sqrt{\pi/2} \times MD = \sqrt{\pi/2} \times \sum_{i=1}^{n} | x_i - \bar{x} |/n$. Kendall and Stuart [12], p. 240, give an approximation for the variance of *MD* as $V(MD) \cong \sigma^2 (1 - 2/\pi)/n$ which is illustrated below.

$$V(MD) = V\left(\sum_{i=1}^{n} | x_i - \bar{x} |/n\right) = \frac{1}{n^2} V \sum_{i=1}^{n} | x_i - \bar{x} |$$

$$= \frac{1}{n^2}\left[E\left(\sum_{i=1}^{n} | x_i - \bar{x} |\right)^2 - \left(E\sum_{i=1}^{n} | x_i - \bar{x} |\right)^2\right]. \qquad \text{(C1)}$$

We first compute the 1st expectation on the RHS inside brackets in equation (C1):

$$E\left[\left(\sum_{i=1}^{n} | x_i - \bar{x} |\right)^2\right] = E\left[\sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum\sum_{i \neq j}^{n} | x_i - \bar{x} || x_j - \bar{x} |\right]$$

$$= E[(n-1)S^2] + \sum\sum_{i \neq j}^{n} [E| x_i - \bar{x} | \times E| x_j - \bar{x} |]$$

$$\approx (n-1)\sigma^2 + \sum\sum_{i \neq j}^{n} [\sigma/\sqrt{\pi/2} \times \sigma/\sqrt{\pi/2}]$$

$$= (n-1)\sigma^2 + n(n-1)2\sigma^2/\pi.$$

The 2nd expectation on the RHS of equation (C1) is equal to

$$E\sum_{i=1}^{n} |\, x_i - \bar{x}\,| = nE|\, x_i - \bar{x}\,| \approx \frac{n\sigma}{\sqrt{\pi/2}};$$

thus, insertion into (C1) yields

$$V(MD) \cong \frac{1}{n^2}\left[(n-1)\sigma^2 + 2n(n-1)\sigma^2/\pi - \left(\frac{n\sigma}{\sqrt{\pi/2}}\right)^2\right]$$

$$= \frac{1}{n^2}[(n-1)\sigma^2 - 2n\sigma^2/\pi] \cong [(n-1)/n - 2/\pi]\sigma^2/n. \qquad \text{(C2)}$$

Clearly the variance expression in (C2) is negative at $n = 2$ and roughly 0.03 at $n = 3$ because $E(MD) \le \frac{\sigma}{\sqrt{\pi/2}}$ which led to $V(MD)$ to be smaller than its actual value; as a result, $(n-1)/n$ on the RHS of (C2) was increased to 1 so that $V(MD) > 0,$ for all $n \ge 2,$ leading to $V(MD) \approx (1 - 2/\pi)n^{-1}\sigma^2,$ $n \ge 20.$ Thus,

$$V(\hat{\sigma}_{md}) = V(\sqrt{\pi/2} \times MD) \cong (\pi/2) \times \sigma^2 (1 - 2/\pi)/n = (0.50\pi - 1)n^{-1} \times \sigma^2.$$

Clearly, the estimator $\hat{\sigma}_{md}$ must be biased, and as shown above, the amount of bias is approximately given by $B(\hat{\sigma}_{md}) = \frac{\sigma}{\sqrt{\pi/2}} - \sigma = \omega \times \sigma,$ where $\omega = (\sqrt{2/\pi} - 1) < 0$ to order of $n^{-1/2}$. This last expression for the $B(\hat{\sigma}_{md})$ is disturbing because it is impossible for the bias-amount in a σ-estimator not to diminish with increasing sample size. Clearly $B(\hat{\sigma}_{md}) = \omega$ $\times \sigma/n^a$, and the exponent a should lie within [0.25, 0.75]. A simulationstudy may better determine a good approximation for the constant a. Perhaps, $B(\hat{\sigma}_{md}) \cong \omega \times \sigma/\sqrt{n}$, but we choose the conservative value, compared to $\omega \times \sigma n^{-0.50}$, of $B(\hat{\sigma}_{md}) \approx \omega n^{-0.40} \times \sigma$.

References

[1] R. A. Boyles, Estimating common-cause sigma in the presence of special causes, JQT 29(4) (1997), 381-395.

[2] G. Casella and R. L. Berger, Statistical Inference, Second Edition, Duxbury/ Thompson Learning, 2002.

[3] C. L. Chen, S. Maghsoodloo and C. S. Park, A method for approximating semivariance in project portfolio analysis, The Engineering Economist 37(1) (1991), 33-59.

[4] (Thomas) Chao-Yu Chou, Exact Control Limits for Monitoring Process Variation, Auburn University dissertation, directed by Saeed Maghsoodloo, 1992.

[5] G. J. Han and S. S. Shapiro, Statistical Models in Engineering, John Wiley & Sons, New York, 1967.

[6] H. L. Harter, The use of sample quasi-ranges in estimating population standard deviation, The Annals of Mathematical Statistics 30(4) (1959), 980-999.

[7] H. L. Harter, Tables of range and studentized range, The Annals of Mathematical Statistics 31(4) (1960), 1122-1147.

[8] V. H. Hogg and E. A. Tanis, Probability and Statistical Inference, 8th ed., Prentice Hall, 2010.

[9] J. N. Hool and S. Maghsoodloo, Normal approximation to linear combinations of independently distributed random variables, AIIE Transactions 12(1) (1980), 140-144.

[10] N. L. Johnson, S. Kotz and N. Balakrishnan, Continuous Univariate Distributions, Vol. 1, 2nd ed., John Wiley & Sons, Inc., 1994.

[11] N. L. Johnson, S. Kotz and N. Balakrishnan, Continuous Univariate Distributions, Vol. 2, 2nd ed., John Wiley & Sons, Inc., 1995.

[12] M. G. Kendall and A. Stuart, The Advanced Theory of Statistics, Vol. 1, Charles Griffin & Company Limited, 1963.

[13] M. G. Kendall and A. Stuart, The Advanced Theory of Statistics, Vol. 2, Charles Griffin & Company Limited, 1967.

[14] B. W. Lindgren, Statistical Theory, 2nd ed., The Macmillan Company, New York, 1968.

[15] Bernard W. Lindgren, Statistical Theory, 4th ed., Chapman and Hall/CRC, London and New York, 1993.

[16] S. Maghsoodloo and D. Helvaci, Renewal and renewal-intensity functions with minimal repair, Journal of Quality and Reliability Engineering 2014 (2014), 1-10.

[17] S. Maghsoodloo and J. N. Hool, On forth limits required to achieve 0.01 accuracy of normal approximation for symmetrical linear combinations, The Journal of the Alabama Academy of Science 54(1) (1983), 1-13.

[18] S. Maghsoodloo, G. Ozdemir, V. Jordan and C.-H. Huang, Strengths and limitations of Taguchi's contributions to quality, manufacturing, and process engineering, Journal of Manufacturing Systems 23(2) (2004), 73-126.

[19] B. Ostle and L. C. Malone, Statistics in Research: Basic Concepts and Techniques for Research Workers, 4th ed., Iowa State University Press/Ames, 1988, pp. 80-81.

[20] E. S. Pearson, A further note on the distribution of range in samples taken from a normal population, Biometrika 18 (1926), 173-194.

[21] E. S. Pearson, The percentage limits for the distribution of range in samples from a normal population, Biometrika 24 (1932), 404-417.

[22] L. G. Tatum, Robust estimation of the process standard deviation for control charts, Technometrics 39(2) (1997), 127-141.

[23] L. H. C. Tippett, On the extreme individuals and the range of samples taken from a normal population, Biometrika 17 (1925), 364.

[24] S. Maghsoodloo and Ching-Ying Huang, Comparing the overlapping of two independent confidence intervals with a single confidence interval for two normal population parameters, Journal of Statistical Planning and Inference 140(11) (2010), 1-10.

[25] S. B. Vardeman, A brief tutorial on the estimation of the process standard deviation, IIE Transactions 31 (1999), 503-507.