Stochastic Models in Probability and Statistics
Vol. -, No. -, 2024, pp. -. Research Paper

On the use of sampling costs in quality control to select an estimator for the variance and standard deviation

Ayda Amniattalab[†] **and J. B. G. Frenk**[‡*]

[†]*Industrial Engineering Department, Faculty of Engineering and Natural Sciences, 34956 Istanbul, Sabancı University, Istanbul, Turkey*
[‡]*Industrial Engineering Department, Faculty of Engineering, Gebze Technical University, 41400, Kocaeli, Gebze,Turkey*
Email(s): ayda@sabanciuniv.edu, frenk@gtu.edu.tr

Abstract. In R and S-charts in quality control applied to a normal population one mostly uses a linear transformation of the range or the sample deviation as unbiased estimators of the unknown process standard deviation. In this paper, we propose related statistics as alternative estimators of the unknown standard deviation and variance having a smaller mean squared error. At the same time, we give a theoretical explanation for the rules of thumb recommended in quality control which unbiased estimators to use. Since obtaining samples from different independent subgroups is costly, we propose a mathematical model for selecting these samples to satisfy the budget constraint. The used estimator for the variance or standard deviation has a minimal mean squared error within a certain class of estimators. It is shown that selecting the whole sample from one particular chosen subgroup is a good strategy for linear sampling costs.

Keywords: Mean squared error; Pooled statistics; Quality control; R-estimator; S-estimator.

1 Introduction.

In R and S-charts (see Section 4.7 and the Appendix of Chapter 4 of Ryan (2011) or Montgomery (2009)) applied to a normal population, one uses a linear transformation of the range or the sample standard deviation as unbiased estimators of the unknown process standard deviation. Within quality control, there is a tradition to use only unbiased estimators. However, if one uses as a quality measure of an estimator its mean squared error, there is no need to restrict to unbiased estimators. In this paper we will consider a class of estimators containing both

[*]Corresponding author

Received: / Accepted:

 https://smps.um.ac.ir

biased and unbiased estimators from which it is easy to derive a (biased) estimator having a minimal mean squared error among this class. A similar set-up was also followed in Woodall and Montgomery (2000), Mahmoud et al. (2010) and Vardeman (1999). Our analysis extends their approach. Since the theoretical properties of the R and S estimators in R and S-charts are already known for a long time most of the proofs of these results can be found in papers written more than thirty years ago. Also these results coincide with classical results in statistics and as a consequence of this no papers appeared on this topic recently. After the introduction of these estimators in quality control the focus shifted on how to apply these estimators to different industrial applications. For a recent overview on the practical use of these so-called univariate charts to all kind of different settings like healthcare and environmental control the reader is referred to Suman and Prajapati (2018), Zwetsloot et al. (2023) and Arciszewski (2023).

As in statistics it is mostly assumed in quality control models that gathering data does not involve any costs. However, in practice data gathering might be costly and subject to budget restrictions. Due to these restrictions we need to decide beforehand from which available resources we will generate our data. These available resources should be selected in such a way that our constructed estimator will satisfy the quality measure imposed on an estimator. In this paper this can be either an estimator within the class of unbiased estimators having a minimal mean squared error or a biased estimator within a given class of estimators satisfying this property. In this paper we will propose a simple static optimization problem dealing with the selection of the available resources from which the data are generated.

The outline of this paper is now as follows. In the first subsection of Section 2, we present a compact overview of the classical unbiased estimators of the standard deviation used in S and R-charts for single groups. Computing the mean squared error of these estimators we also confirm theoretically some empirical rules of thumb applied in the literature which estimator to use. In the past these rules of thumb were advised without any theoretical explanation.

In the second subsection of Section 2 we propose related biased estimators having a smaller mean squared error than the classical used estimators for both the standard deviation and the variance. At the same time we extend these results for a normal population consisting of either single or multiple independent subgroups to a population having a location-scale family of distributions.

In section 3 of this paper we formulate in the presence of a budget constraint on collecting data a simple static model constructing an estimator for both the variance and standard deviation having the smallest mean squared error. In this set-up we are allowed to sample from different independent subgroups each having different sampling costs. Due to the separable structure of the mean squared error objective function this optimization problem can be solved by a dynamic programming procedure. As a special case this model confirms the intuitively clear result that for linear sampling costs it is a good strategy to generate the whole sample from a single subgroup.

Finally in Section 4 we list our findings in this paper and give in the Appendix the proofs of our main results. Observe that the first subsection of Section 2 can be regarded as an elementary and unifying note on the property of the classical estimators used in R and S-charts discussed in standard textbooks on quality control.

2 On classical and alternative estimators of the standard deviation and variance used in quality control.

In the first subsection we will discuss the main properties of the classical unbiased estimators for the standard deviation used in R and S-charts. The result in Lemma 1 about the normalisation constant used in a S-chart seems to be new. In the second subsection we propose for single and multiple subgroups some alternative biased estimators for both the standard deviation and variance having a smaller mean squared error than the classical unbiased estimators. We will also analyze in this subsection the differences between these estimators and compute the ratios of their mean squared error. In the same subsection encountering multiple independent subgroups of data having different sample sizes the relation of these estimators to so-called pooled estimators (see page 360 of Irwin and Miller (2004) or Arnold (1990)) is also discussed.

2.1 On the main properties of the classical unbiased estimators in R and S-charts.

In this section we first give for completeness an overview of the estimators used in quality control to estimate the unknown standard deviation for normal populations. Similar results can be found in Mahmoud et al. (2010) and Woodall and Montgomery (2000). Introducing in a normal sample $\mathbf{X}^\top = (X_1, \ldots, X_n)$, the largest order statistic $X_{n:n} := \max\{X_1, \ldots, X_n\}$ and the smallest order statistic $X_{1:n} := \min\{X_1, \ldots, X_n\}$, one uses in R-charts within quality control as an unbiased estimator of the unknown standard deviation $\sigma > 0$ in a sample of size n the statistic

$$T_R(\mathbf{X}) := \theta_n R_n(\mathbf{X}), \tag{1}$$

with R_n the so-called sample range given by

$$R_n(\mathbf{X}) := X_{n:n} - X_{1:n},$$

(see page 229 of Montgomery (2009)). Of course, it is assumed this sample is taken from a system in control. To compute the constant θ_n in this unbiased estimator, we observe that

$$R_n(\mathbf{X}) \stackrel{d}{=} \sigma R_{n,s}(\mathbf{Y}), R_{n,s}(\mathbf{Y}) := Y_{n:n} - Y_{1:n}, \tag{2}$$

with $Z_1 \stackrel{d}{=} Z_2$ meaning the random variables Z_1 and Z_2 have the same distribution and $R_{n,s}(\mathbf{Y}) := Y_{n:n} - Y_{1:n}$ the range of the sample $\mathbf{Y}^\top = (Y_1, \ldots, Y_n)$ from a standard normal population. As shown in the remainder of this paper it is easy to generalize the above observations for a normal population to location-scale families of distributions. By the symmetry of a standard normal distribution around zero, we obtain by relations (1) and (2) that

$$\sigma = \mathbb{E}(T_R(\mathbf{X})) = \mathbb{E}(\theta_n \sigma R_{n,s}(\mathbf{Y})) = \theta_n \sigma \mathbb{E}(R_{n,s}(\mathbf{Y})) = 2\sigma\theta_n \mathbb{E}(Y_{n:n}),$$

with $Y_{n:n}$ the largest order statistic of a sample of size n coming from a standard normal distribution. This shows

$$\theta_n = (\mathbb{E}(R_{n,s}(\mathbf{Y})))^{-1} = (2\mathbb{E}(Y_{n:n}))^{-1}. \tag{3}$$

Introducing the mean squared error of an estimator T of some unknown constant b defined by

$$\mathrm{MSE}(T) := \mathbb{E}((T(\mathbf{X}) - b)^2),$$

and the bias of an estimator given by

$$\mathrm{bias}(T) := \mathbb{E}(T(\mathbf{X})) - b,$$

yields by relation (2) and (3) that the mean squared error of the unbiased estimator $T_R(\mathbf{X}) = \theta_n R_n(\mathbf{X})$ is given by

$$\mathrm{MSE}(T_R) = \mathrm{Var}(\theta_n \sigma R_{n,s}(\mathbf{Y})) = \frac{\sigma^2 \mathrm{Var}(R_{n,s}(\mathbf{Y}))}{(\mathbb{E}(R_{n,s}(\mathbf{Y})))^2} = \frac{\sigma^2 \mathrm{Var}(R_{n,s}(\mathbf{Y}))}{4(\mathbb{E}(Y_{n:n}))^2}. \quad (4)$$

Since in general, it seems impossible to derive an elementary expression for the above ratio divided by σ^2 (Arnold et al. (2008)), tables are used within quality control (Montgomery (2009), Harter (1960)). In our computational experiments, we generate sample paths of the stochastic range process $\mathbf{R} = \{R_{n,s} : n \in \mathbb{N}\}$ and use Monte Carlo simulation to compute both $\mathrm{Var}(R_{n,s}(\mathbf{Y}))$ and $\mathbb{E}(R_{n,s}(\mathbf{Y}))$ for $n \leq 50$. The density of the random variable $R_{n,s}(\mathbf{Y})$ is given by David (1970)

$$f_{R_{n,s}(\mathbf{Y})}(t) = n(n-1)\int_{-\infty}^{\infty} \varphi(u)(\Phi(u+t) - \Phi(u))^{n-2}\varphi(u+t)du,\ t \geq 0,$$

with φ the standard normal density and Φ the standard normal cdf. Using relation (2) this implies that the density of the unbiased estimator $T_R(\mathbf{X}) = \theta_n R_n(\mathbf{X})$ is given by

$$f_{T_R(\mathbf{X})}(t) = \frac{1}{\sigma\theta_n} f_{R_{n,s}(\mathbf{Y})}\left(\frac{t}{\sigma\theta_n}\right), t \geq 0,$$

and so it is possible to give a plot of the density of the cdf of this estimator. On page 112 of Montgomery (2009), as a rule of thumb, the unbiased estimator T_R is recommended for samples from a normal population of size n with $n \leq 6$ instead of the unbiased estimator T_S used in a S-chart. In an S-chart, the unbiased estimator $T_S(\mathbf{X}) := \gamma_n S_n(\mathbf{X}), n \geq 2$ is applied with S_n^2 denoting the well-known unbiased sample variance estimator of σ^2 given by

$$S_n^2(\mathbf{X}) := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2, \overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i,$$

with $\overline{X}_n$ representing the sample mean. The constant γ_n is selected in such a way that the estimator $\gamma_n S_n$ is an unbiased estimator of the unknown standard deviation σ. By the properties of a normal distribution it is well known (Casella and Berger (2002)) that

$$S_n(\mathbf{X}) \stackrel{d}{=} \sigma S_{n,c}(\mathbf{Y}) \text{ and } (n-1)S_{n,c}^2(\mathbf{Y}) \stackrel{d}{=} Z, \quad (5)$$

with $S_{n,c}$ the sample standard deviation of a sample $\mathbf{Y}^\top = (Y_1, \ldots, Y_n)$ from a standard normal population and Z a random variable having a chi-square distribution with $n-1$ degrees of freedom. This yields

$$\sigma = \mathbb{E}(\gamma_n S_n(\mathbf{X})) = \mathbb{E}(\gamma_n \sigma S_{n,c}(\mathbf{Y})) = \frac{\gamma_n \sigma}{\sqrt[2]{n-1}}\mathbb{E}(\sqrt[2]{Z}) = \gamma_n \sigma \sqrt[2]{\frac{2}{n-1}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})},$$

with $\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}dx$ the well known gamma function. Hence we obtain for every $n\geq 2$ that

$$\gamma_n=(\mathbb{E}(S_{n,c}(\mathbf{Y})))^{-1}=\sqrt[2]{\frac{n-1}{2}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}. \tag{6}$$

Using $S_{n,c}^2$ is an unbiased estimator of the known variance of a standard normal population and so $\mathbb{E}(S_{n,c}^2(\mathbf{Y}))=1$, it follows by relation (5) and (6) that the mean squared error of the estimator $\gamma_n S_n, n\geq 2$ is given by

$$\text{MSE}(\gamma_n S_n)=\text{Var}(\gamma_n\sigma S_{n,c}(\mathbf{Y}))=\sigma^2\frac{\text{Var}(S_{n,c}(\mathbf{Y}))}{(\mathbb{E}(S_{n,c}(\mathbf{Y})))^2}=\sigma^2(\gamma_n^2-1). \tag{7}$$

Since MSE $(\gamma_n S_n)\geq 0$, we obtain for every $n\geq 2$ that $\gamma_n\geq 1$. In the next result, we list some global properties of the sequence $\gamma_n, n\geq 2$. Before discussing these properties, we introduce the following definition.

Definition 1. *The non-negative sequence δ_n, $n\geq 2$ is called log-convex if the first order differences $\Delta\delta_n:=\delta_{n+1}-\delta_n$, $n\geq 2$ are increasing.*

For the sequence $\gamma_n:=(\mathbb{E}(S_{n,c}(\mathbf{Y}))^{-1}$, $n\geq 2$ one can show the following result. Its proof is given in the Appendix.

Lemma 1. *The sequence $\gamma_n, n\geq 2$ is decreasing and log-convex and it satisfies $\gamma_2=\sqrt[2]{\frac{\pi}{2}}$ and for every $n\geq 2$ the first order non-linear recurrence relation*

$$\gamma_{n+1}=\gamma_n^{-1}\sqrt[2]{\frac{n}{n-1}}. \tag{8}$$

Moreover $\lim_{n\uparrow\infty}n(\gamma_n-1)=4^{-1}$ *and for every $n\geq 3$*

$$1\leq\left(\frac{n}{n-1}\right)^{\frac{1}{4}}\leq\gamma_n\leq\left(\frac{n-1}{n-2}\right)^{\frac{1}{4}}. \tag{9}$$

By relation (8), it is easy and numerically stable to compute the constants γ_n, $n\geq 2$. Hence it is not necessary to use tables for the sequence γ_n, $n\in\mathbb{N}$, $n\geq 2$. Since

$$\frac{(n-1)S_n^2(\mathbf{X})}{\sigma^2}\overset{d}{=}Z,$$

with Z a random variable having a chi-square distribution with $n-1$ degrees of freedom (Casella and Berger (2002)) having density $f_{\mathbf{Z}}(z)=\frac{e^{-\frac{z}{2}}z^{\frac{n-1}{2}-1}}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}$ this yields

$$T_S(\mathbf{X})=\gamma_n S_n(\mathbf{X})\overset{d}{=}\frac{\gamma_n\sigma\sqrt[2]{Z}}{\sqrt[2]{n-1}}. \tag{10}$$

Since the density of the non-negative random variable $\sqrt[2]{Z}$ with Z having a chi-square distribution having $n-1$ degrees of freedom is given by

$$f_{\sqrt[2]{Z}}(z)=\frac{z^{\frac{n-1}{2}}e^{-\frac{z^2}{2}}}{2^{\frac{n-3}{2}},\Gamma(\frac{n-1}{2})},\quad z>0.$$

This yields by relation (10) that the density $f_{T_S}(t)$, $t>0$ of the estimator $T_S(\mathbf{X})=\gamma_n S_n(\mathbf{X})$ equals

$$f_{T_S(\mathbf{X})}(t)=\frac{\sqrt[2]{n-1}}{\gamma_n\sigma}f_{\sqrt[2]{Z}}\left(\frac{\sqrt[2]{n-1}t}{\gamma_n\sigma}\right)=\frac{\sigma\sqrt[2]{2}\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}f_{\sqrt[2]{Z}}\left(\frac{\sigma\sqrt[2]{2}\Gamma(\frac{n}{2})t}{\Gamma(\frac{n-1}{2})}\right),$$

and so it is possible to give a plot of the density of this estimator. In Figure 1, we give a plot of the functions $n\to\mathrm{MSE}(\gamma_n S_{n,c})=\gamma_n^2-1$ and $n\to\mathrm{MSE}(\theta_n R_{n,c})$ for $2\leq n\leq 50$. As this figure shows, comparing the mean squared error of both the R and S estimators defined in relations (4) and (7), it is always better to use the estimator $T_S=\gamma_n S_n$ instead of the estimator $T_R=\theta_n R_n$ for $n\geq 7$ while there is a slight preference for the S-estimator when $n\leq 6$. This confirms the rule of thumb given in Montgomery (2009) or Vardeman (1999) and the recommendation of always using the S-estimator in Mahmoud et al. (2010). An additional advantage of the S-estimator is that the constant γ_n can be numerically computed in a stable way applying relation (8) without making use of tables, while θ_n needs to be computed by using tables or (if not available) by simulation or numerical integration. This means computing θ_n can be less accurate.

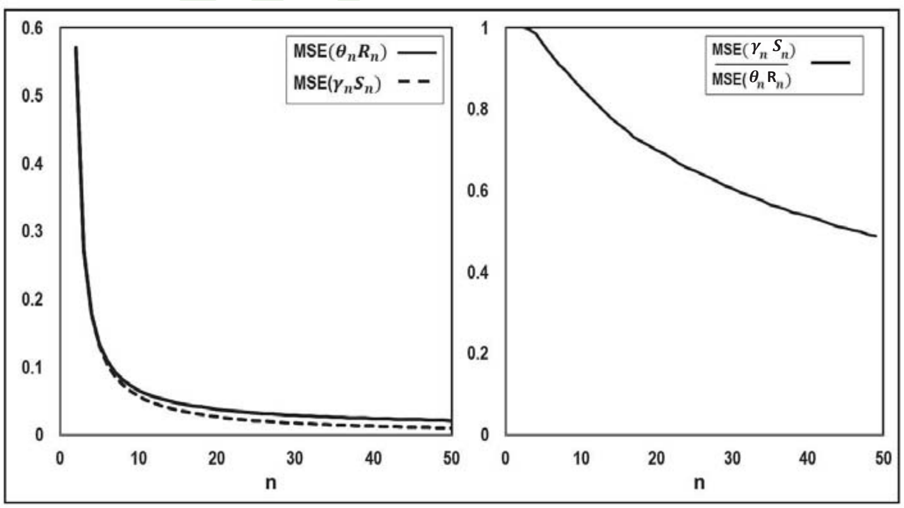

Figure 1: Plot of the mean squared errors of R and S estimators (left panel) for $\sigma=1$ and their ratio (right panel).

In the next subsection we introduce related unbiased estimators having a smaller mean squared error.

2.2 On some alternative biased estimators for the standard deviation and the variance

It is well known within statistics that biased estimators of a unknown parameter having a smaller mean squared error (Woodall and Montgomery (2000)) can under certain conditions be constructed using a linear transformation of the unbiased estimator of the same parameter. The construction of these alternative estimators for both the range estimator and sample standard deviation estimator used in R and S-charts is discussed in the following result. Its proof is given in the Appendix.

Lemma 2. *Let $k > 0$ and $\mathbf{X}^\top = (X_1, ..., X_n), n \geq 2$ be a random vector of size n satisfying*

$$X_i \stackrel{d}{=} a + bY_i, 1 \leq i \leq n,$$

with unknown location parameter $a \in \mathbb{R}$ and scale parameter $b > 0$ and the random vector $\mathbf{Y}^\top = (Y_1, ..., Y_n)$ consisting of independent and identically distributed random variables has a known cdf F having a finite variance. If the statistic $T : \mathbb{R}^n \to \mathbb{R}$ is an estimator of the unknown scale parameter $b > 0$ and

$$T(\mathbf{X}) \stackrel{d}{=} bf(\mathbf{Y}).$$

for some function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying $\mathbb{E}(f^k(\mathbf{Y})) > 0$ and $\mathbb{E}(f^{2k}(\mathbf{Y}))$ finite, then among the class of statistics $\alpha T^k, \alpha \in \mathbb{R}$, used as an estimator of the unknown parameter b^k the estimator $\alpha^\star T^k$ with

$$\alpha^\star = \frac{\mathbb{E}(f^k(\mathbf{Y}))}{\mathbb{E}(f^{2k}(\mathbf{Y}))},$$

is an estimator of b^k with the smallest mean squared error. Its bias is given by

$$bias(\alpha^\star T^k) = \frac{-b^k\, Var(f^k(\mathbf{Y}))}{\mathbb{E}(f^{2k}(\mathbf{Y}))},$$

and its mean squared error by

$$MSE\left(\alpha^\star T^k\right) = -bias(\alpha^\star T^k) b^k.$$

Since $\mathrm{Var}(f^k(\mathbf{Y})) \geq 0$, we obtain by Lemma 2 that on average for every $k \in \mathbb{N}$ we underestimate the unknown parameter b^k using the estimator $\alpha^\star T^k$ for b^k. The result in Lemma 2 can also be applied to a non-normal population. Since we know the cdf of the random variable Y_1 it is possible by simulation to determine the value of $\alpha^\star$. Observe in a normal population it follows for the standard sample deviation estimator that the constant $\alpha^\star$ is given by an elementary expression. As shown in Corollary 3 this constant is given by γ_n^{-1} with γ_n listed in relation (6). As an example of a non-normal population we mention a nonnegative sample $\mathbf{X}^\top = (X_1, ..., X_n)$ satisfying

$$X_i \stackrel{d}{=} aY_i^b, a > 0,\ b > 0, \tag{11}$$

with Y_i, $i = 1, ..., n$, nonnegative independent and identically distributed random variables having a known cdf F and the random variables $\ln(Y_i)$ having a finite variance. Well known examples covering this case are Y_1 exponentially distributed with parameter 1. In this case the random variable X_i has a Weibull distribution. Another example is given by $\mathbf{Y}_i$ has a gamma distribution with scale parameter 1 and known shape parameter $\alpha > 0$. In this case the random variable X_i has a so-called generalised gamma distribution (Abbaszadehpeivasti and Frenk (2023)). Since by relation (11) it follows

$$\ln(X_i) \stackrel{d}{=} \ln(a) + b\ln(Y_i), \tag{12}$$

we can easily apply Lemma 2 replacing in the used statistic the random variable X_i by the random variable $\ln(X_i)$. In case the nonnegative independent random variables $X_i,\ i=1,2,..,n$ are log-normal distributed, it follows by definition that the random variables $\ln(Y_i)$ are standard normal distributed. Another example is given by

$$X_i \stackrel{d}{=} (a+bY_i)^\beta, a>0,\ b>0,$$

with $\beta>0$ known and $Y_i,\ i=1,...,n$ a sequence of non-negative and identically distributed random variables having a known cdf F and finite variance. In this case it follows that

$$X_i^{\beta^{-1}} \stackrel{d}{=} a+bY_i,$$

and we replace in the used statistic the random variable X_i by the random variable $X_i^{\beta^{-1}}$. The above transformations resemble the Box-Cox transformations used in statistical Process Control transforming non-normal charts into charts which are approximately normal distributed (see section 3 of Figueiredo and Gomes (2006)). If we compare the mean squared error of the unbiased estimator $\alpha_0 T^k,\ \alpha_0 := \mathbb{E}(f^k(\mathbf{Y}))^{-1}$ of the parameter b^k with the minimum mean squared error estimator within the class of estimators $\alpha T^k,\ \alpha\in\mathbb{R}$ the next corollary follows immediately.

Corollary 1. *If the conditions of Lemma 2 hold and the unbiased estimator $\alpha_0 T^k$ of b^k with $\alpha_0 = \mathbb{E}(f^k(\mathbf{Y}))^{-1}$ is considered, then*

$$\frac{MSE(\alpha^\star T^k)}{MSE(\alpha_0 T^k)} = \frac{\mathbb{E}(f^k(\mathbf{Y}))^2}{\mathbb{E}(f^{2k}(\mathbf{Y}))} \leq 1.$$

Proof. It follows by relation (32) that

$$\mathrm{MSE}(\alpha_0 T^k) = b^{2k} p_k(\alpha_0) = \frac{b^{2k}\mathrm{Var}(f^k(\mathbf{Y}))}{\mathbb{E}(f^k(\mathbf{Y}))^2}.$$

Applying Lemma 2 implies the desired result. □

We now apply Lemma 2 (for $k=1$) to the classical unbiased estimators used in a R and S-chart for a normal population. The next result is an easy application of Lemma 2.

Corollary 2. *If $\mathbf{X}^\top = (X_1,....,X_n),\ n\geq 2$ is a random sample of size n from a normal population then the estimator $T_R(\alpha^\star) := \alpha^\star R_n$ of the standard deviation $\sigma>0$ with*

$$\alpha^\star = \frac{\mathbb{E}(R_{n,s}(\mathbf{Y}))}{\mathbb{E}(R_{n,s}^2(\mathbf{Y}))},$$

and $\mathbf{Y}^\top = (Y_1,...,Y_n)$ a random sample from a standard normal population satisfies

$$\mathbb{E}(\alpha^\star R_n(\mathbf{X})) = \frac{\sigma\mathbb{E}(R_{n,s}(\mathbf{Y}))^2}{\mathbb{E}(R_{n,s}^2(\mathbf{Y}))}.$$

Among the class of estimators $T_R(\alpha) := \alpha R_n, \alpha>0$ of the standard deviation $\sigma>0$ in a R-chart applied to a normal population the biased estimator $T_R(\alpha^\star)$ has the smallest mean squared error and this is given by

$$MSE(\alpha^\star R_n) = \sigma^2 \frac{Var(R_{n,s}(\mathbf{Y}))}{E(R_{n,s}(\mathbf{Y})^2)}.$$

Proof. Since in a normal population we know that

$$(X_1,...,X_n) = (\mu + \sigma Y_1,...,\mu + \sigma Y_n),$$

with $\mathbf{Y}^\top = (Y_1,...,Y_n)$ consisting of independent and standard normal distributed random variables we obtain

$$R_n(\mathbf{X}) = \max\{\mu + \sigma Y_1,...,\mu + \sigma Y_n\} - \min\{\mu + \sigma Y_1,...,\mu + \sigma Y_n\} = \sigma R_{n,s}(\mathbf{Y}).$$

Since $f(\mathbf{Y}) = Y_{n:n} - Y_{1:n} = R_{n,s}(\mathbf{Y})$ and $\mathbb{E}(R_{n,s}(\mathbf{Y})) > 0$ for $n \geq 2$ we may apply Lemma 2 substituting $k = 1$. This shows the result. □

As already observed, one uses in a S-chart in quality control the unbiased estimator $T_S = \gamma_n S_n$ with γ_n listed in relation (6). The next result is again an application of Lemma 2 for $k = 1$.

Corollary 3. *If $\mathbf{X}^\top = (X_1,....,X_n)$, $n \geq 2$ is a random sample of size n from a normal population then the estimator $T_S(\alpha^{\star\star}) = \alpha^{\star\star} S_n$ of the standard deviation $\sigma > 0$ with*

$$\alpha^{\star\star} = \gamma_n^{-1} = \sqrt[2]{\frac{2}{n-1}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})},$$

satisfies

$$\mathbb{E}(\alpha^{\star\star} S_n(\mathbf{X})) = \frac{2}{n-1} \frac{\Gamma^2(\frac{n}{2})\sigma}{\Gamma^2(\frac{n-1}{2})} = \gamma_n^{-2}\sigma.$$

Among the class of estimators $T_S(\alpha) = \alpha S_n$, $\alpha > 0$ of the standard deviation $\sigma > 0$ in a normal population the biased estimator $T_S(\alpha^{\star\star})$ has the smallest mean squared error and this is given by

$$MSE(\alpha^{\star\star} S_n) = \sigma^2 \left(1 - \frac{2\Gamma^2(\frac{n}{2})}{(n-1)\Gamma^2(\frac{n-1}{2})}\right) = \sigma^2(1 - \gamma_n^{-2}). \tag{13}$$

Proof. Again using $(X_1,...,X_n) = (\mu + \sigma Y_1,...,\mu + \sigma Y_n)$ with $\mathbf{Y}^\top = (Y_1,...,Y_n)$ consisting of independent and standard normal distributed random variables we obtain $S_n(\mathbf{X}) = \sigma S_{n,c}(\mathbf{Y})$ and so

$$f(\mathbf{Y}) := S_{n,c}(\mathbf{Y}) = \sqrt[2]{\frac{\sum_{i=1}^n (Y_i - \overline{Y}_n)^2}{n-1}}.$$

Since $\mathbb{E}(f(\mathbf{Y})) > 0$ for every $n \geq 2$, the conditions of Lemma 2 are satisfied for $k = 1$ and this shows the result observing $\mathbb{E}\left(f^2(\mathbf{Y})\right) = 1$ and

$$E(f(\mathbf{Y})) = \frac{1}{\sqrt[2]{n-1}} \mathbb{E}\left(\sqrt[2]{Z}\right) = \sqrt[2]{\frac{2}{n-1}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}.$$

with the random variable Z having a chi-square distribution with $n-1$ degrees of freedom. □

Observe the above results also hold for any sequence of independent and identically distributed random variables $\mathbf{X}_i$, $i = 1, \ldots, n$, for which the normalized random variable $Y_i = b^{-1}(X_i - a)$ has a known cdf. If this holds we cannot compute the optimal α but need to approximate this value by simulation. By relations (7) and (13) we obtain for a normal population

$$\gamma_n^2 \mathrm{MSE}(\alpha^{\star\star} S_n) = \mathrm{MSE}(\gamma_n S_n).$$

Since by relation (9) we know $(\frac{n}{n-1})^{\frac{1}{2}} \leq \gamma_n^2 \leq (\frac{n-1}{n-2})^{\frac{1}{2}}, n \geq 3$, this implies that both the optimal biased and unbiased estimators of the standard variance in a normal population have almost the same mean squared error for already relatively small values of the sample size. Observe the mean squared error also depends linearly on the unknown variance. Hence, if it is suspected that the variance is large and the mean squared error is regarded as a more important quality measure of an estimator than whether an estimator is biased or unbiased, one should apply the biased estimator $\alpha^{\star\star} S_n$. In Figure 2, we now list the mean squared error of the most important three different estimators excluding the classical range estimator for $\sigma = 1$. As expected, the estimator $\alpha^{\star\star} S_n$ has the smallest mean squared error.

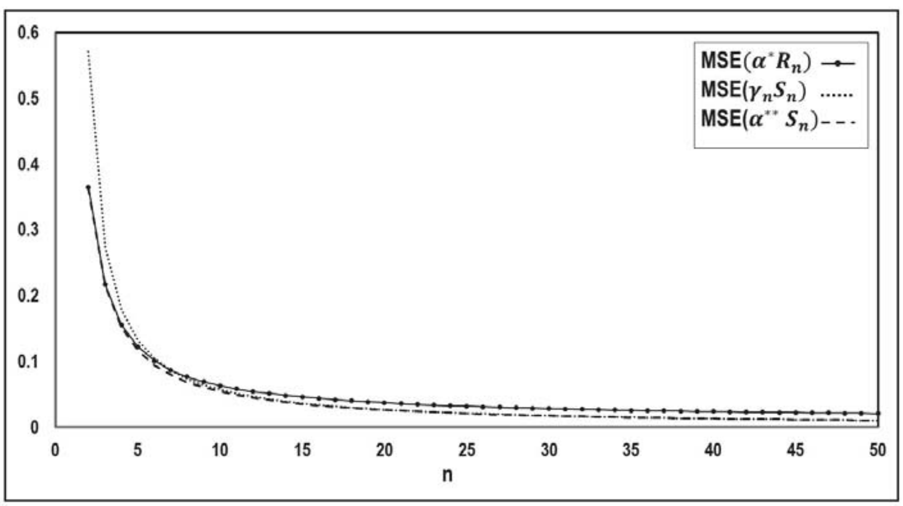

Figure 2: Plot of the mean squared errors of the three estimators of the standard deviation.

In case we like to give an estimation of the variance we identify in the next lemma the best possible estimator with respect to the mean squared error within the class of linear transformations of the sample variance. The next lemma is actually a special case of Lemma 2 for $k = 2$. The proof is listed in the Appendix.

Lemma 3. *If* $\mathbf{X}^\top = (X_1, \ldots, X_n), n \geq 2$, *is a random sample of size* n *from a population satisfying*

$$X_i \stackrel{d}{=} \mu + \sigma Y_i, \tag{14}$$

having unknown expectation μ *and standard deviation* σ *and the random vector* $\mathbf{Y}^\top = (Y_1, \ldots, Y_n)$ *consists of independent and identically distributed random variables having a known cdf* F *and* *a finite 4th moment, then the estimator* $T_S^2(\alpha^{\star\star\star}) = \alpha^{\star\star\star} S_n^2$ *of the variance* $\sigma^2 > 0$ *with*

$$\alpha^{\star\star\star} = \frac{n}{\mathbb{E}(Y_1^4) - \frac{n-3}{n-1} + n},$$

satisfies

$$\mathbb{E}(\alpha^{\star\star\star} S_n^2) = \frac{n\sigma^2}{\mathbb{E}(Y_1^4) - \frac{n-3}{n-1} + n}.$$

Among the class of estimators $T_S^2(\alpha) = \alpha S_n^2,\ \alpha > 0$ *of the variance* σ^2 *the biased estimator* $T_S^2(\alpha^{\star\star\star})$ *has the smallest mean squared error and this is given by*

$$MSE(\alpha^{\star\star\star} S_n^2) = b^4(1 - \alpha^{\star\star\star}) = \frac{b^4((n-1)\mathbb{E}(Y_1^4) - n + 3)}{(n-1)\mathbb{E}(Y_1^4) - n + 3 + n^2}.$$

Suppose we consider M independent samples $\mathbf{X}_m = (X_{1,m}, \ldots, X_{n_m,m})$ of different or equal sizes $n_m \geq 2$, $m = 1, \ldots, M$ from location-scale families having possibly different unknown location parameters but the same unknown scale parameter. This set-up often occurs in quality control. Typically, an initial series of independent samples or independent subgroups is used to estimate the mean and standard deviation of a process. During this initial phase, the process should be in control. These estimations are then used to produce control limits in R and S-charts. Since the means might be different in each of the samples we consider for the estimation of the scale parameter the following class of statistics

$$T(\mathbf{X}_1, \ldots, \mathbf{X}_M) = \sum_{m=1}^{M} \alpha_m T_m^k(\mathbf{X}_m), \alpha_m \in \mathbb{R}, \tag{15}$$

with $T_m : \mathbb{R}^{n_m} \to \mathbb{R}$ the statistic used to estimate the scale parameter in sample subgroup $m = 1, \ldots, M$. This means we use a weighted combination of estimators for the estimation of the unknown scale parameter. One can now show the following generalization of Lemma 2. This result is also mentioned in Woodall and Montgomery (2000) without proof for the special case $k = 1$ and the normal population applying the sample standard deviation statistic to each subgroup. For completeness its elementary proof is listed in the Appendix.

Lemma 4. *Let* $k > 0$ *and for* $m = 1, \ldots, M$ *the random vectors*

$$\mathbf{X}_m = (X_{1,m}, \ldots, X_{n_m,m}),$$

$n_m \geq 2$ *are independent random samples of size* $n_m, m = 1, \ldots, M$*, satisfying*

$$X_{i,m} \stackrel{d}{=} a_m + bY_{i,m}, \tag{16}$$

for every $1 \leq i \leq n_m$ *with the random vector* $\mathbf{Y}_m = (Y_{1,m}, \ldots, Y_{n_m,m})$ *consisting of independent and identically distributed random variables having a known cdf* $F_m, m = 1, \ldots, M$*. If for* $m = 1, \ldots, M$ *the statistic* $T_m : \mathbb{R}^{n_m} \to \mathbb{R}$ *is an estimator of the unknown scale parameter* b *and*

$$T_m(\mathbf{X}_m) \stackrel{d}{=} b f_m(\mathbf{Y}_m), \tag{17}$$

for some function $f_m : \mathbb{R}^{n_m} \to \mathbb{R}$ *satisfying* $\mathbb{E}(f_m^k(\mathbf{Y}_m)) > 0$ *and* $\mathbb{E}(f_m^{2k}(\mathbf{Y}_m))$ *is finite then among the class of statistics* $\sum_{m=1}^{M} \alpha_m T_m^k, \alpha_m \in \mathbb{R}$ *used as an estimator of the parameter* b^k *the estimator* $\sum_{m=1}^{M} \alpha_m^{\star} T_m^k$ *with*

$$\alpha_m^{\star} = \left(1 + \sum_{m=1}^{M} \frac{\left(\mathbb{E}(f_m^k(\mathbf{Y}_m))\right)^2}{Var\left(f_m^k(\mathbf{Y}_m)\right)}\right)^{-1} \frac{\mathbb{E}(f_m^k(\mathbf{Y}_m))}{Var(f_m^k(\mathbf{Y}_m))},$$

is an estimator of b^k *having the smallest mean squared error among this class. Its bias equals*

$$bias\left(\sum_{m=1}^{M} \alpha_m^{\star} T_m^k\right) = -b^k \left(1 + \sum_{m=1}^{M} \frac{\left(\mathbb{E}(f_m^k(\mathbf{Y}_m))\right)^2}{Var(f_m^k(\mathbf{Y}_m))}\right)^{-1},$$

and its mean squared error

$$MSE\left(\sum_{m=1}^{M} \alpha_m^{\star} T_m^k\right) = -bias\left(\sum_{m=1}^{M} \alpha_m^{\star} T_m^k\right) b^k.$$

If we restrict ourselves to the class of unbiased estimators within the class of estimators $T = \sum_{k=1}^{K} \alpha_m T_m^k, \alpha_k \in \mathbb{R}$ and select that unbiased estimator with the smallest mean squared error one can easily show the following result. Its proof is listed in the Appendix.

Lemma 5. *Under the same conditions as in Lemma 4 the unbiased estimator of* b^k *with the smallest mean squared error among the class of estimators* $T = \sum_{m=1}^{M} \alpha_m T_m^k, \alpha_m \in \mathbb{R}$ *is given by* $\sum_{m=1}^{M} \alpha_m^{\star\star} T_m^k$ *with*

$$\alpha_m^{\star\star} = \left(\sum_{m=1}^{M} \frac{\mathbb{E}(f_m^k(\mathbf{Y}_m))^2}{Var(f_m^k(\mathbf{Y}_m))}\right)^{-1} \frac{\mathbb{E}(f_m^k(\mathbf{Y}_m))}{Var(f_m^k(\mathbf{Y}_m))}.$$

Its mean squared error is given by

$$MSE\left(\sum_{m=1}^{M} \alpha_m^{\star\star} T_m^k\right) = b^{2k} \left(\sum_{m=1}^{M} \frac{\left(\mathbb{E}\left(f_m^k(\mathbf{Y}_m)\right)\right)^2}{Var\left(f_m^k(\mathbf{Y}_m)\right)}\right)^{-1}.$$

The next result is a generalization of Corollary 1 for independent samples.

Lemma 6. *If the conditions of Lemma 4 hold and* $\sum_{m=1}^{M} \alpha_m^{\star\star} T_m^k$ *is the unbiased minimum mean squared error estimator of* b^k *and* $\sum_{m=1}^{M} \alpha_m^{\star} T_m^k$ *the biased minimum mean squared error estimator of* b^k *then*

$$\frac{MSE\left(\sum_{m=1}^{M} \alpha_m^{\star} T_m^k\right)}{MSE\left(\sum_{m=1}^{M} \alpha_m^{\star\star} T_m^k\right)} = \frac{\sum_{m=1}^{M} \frac{(\mathbb{E}(f_m^k(\mathbf{Y}_m)))^2}{Var(f_m^k(\mathbf{Y}_m))}}{1 + \sum_{m=1}^{M} \frac{\left(\mathbb{E}(f_m^k(\mathbf{Y}_m))\right)^2}{Var(f_m^k(\mathbf{Y}_m))}}.$$

If the independent random variables $Y_i^{(m)}$, $1 \le m \le M$, $1 \le i \le m$ mentioned in Lemma 4 and 5 have a standard normal distribution then for the estimation of the standard deviation $\sigma > 0$ we apply Lemma 4 with $k = 1$ and $f_m(\mathbf{Y}_m) = S_{n_m,c}(\mathbf{Y}_m)$. By relation (6) it follows that

$$\mathbb{E}(f_m(\mathbf{Y}_m)) = \frac{1}{\gamma_{n_m}}, \mathrm{Var}(f_m(\mathbf{Y}_m) = 1 - \frac{1}{\gamma_{n_m}^2}. \tag{18}$$

By Lemma 4 for $k=1$ it follows using relation (18) that the weights $\alpha_m^{\star}$, $m=1,\ldots,M$ of the minimum mean squared error estimator of the standard deviation within the class $\sum_{m=1}^{M}\alpha_m S_{n_m}$, $\alpha_m \in \mathbb{R}$ are given by

$$\alpha_m^{\star} = \left(1+\sum_{m=1}^{M}\frac{1}{\gamma_{n_m}^2-1}\right)^{-1}\frac{\gamma_{n_m}}{\gamma_{n_m}^2-1}, 1\le m\le M. \tag{19}$$

and this estimator has mean squared error

$$\mathrm{MSE}\left(\sum_{m=1}^{M}\alpha_m^{\star}S_{n_m}\right) = \sigma^2\left(1+\sum_{m=1}^{M}\frac{1}{\gamma_{n_m}^2-1}\right)^{-1}. \tag{20}$$

Applying Lemma 5 and relation (18) it follows that the weights $\alpha_m^{\star\star}$, $m=1,\ldots,M$ of the minimum mean squared error unbiased estimator of the standard deviation within the class $\sum_{m=1}^{M}\alpha_m S_{n_m}$, $\alpha_m \in \mathbb{R}$ equal

$$\alpha_m^{\star\star} = \left(\sum_{m=1}^{M}\frac{1}{\gamma_{n_m}^2-1}\right)^{-1}\frac{1}{\gamma_{n_m}^2-1}, 1\le m\le M \tag{21}$$

and this estimator has mean squared error

$$\mathrm{MSE}\left(\sum_{m=1}^{M}\alpha_m^{\star\star}S_{n_m}\right) = \sigma^2\left(\sum_{m=1}^{M}\frac{1}{\gamma_{n_m}^2-1}\right)^{-1}. \tag{22}$$

This shows for a normal population that

$$\frac{\mathrm{MSE}\left(\sum_{m=1}^{M}\alpha_m^{\star}S_{n_m}\right)}{\mathrm{MSE}\left(\sum_{m=1}^{M}\alpha_m^{\star\star}S_{n_m}\right)} = \frac{\sum_{m=1}^{M}\frac{1}{\gamma_{n_m}^2-1}}{1+\sum_{m=1}^{M}\frac{1}{\gamma_{n_m}^2-1}}. \tag{23}$$

In order to illustrate the relations we consider the following example.

Example 1. Devor et al. (1992) provides on page 165 a data set made up of measurements of an engine block's cylinder bore diameter. The inside of the cylinder bore was measured after the boring operation and the units of measurement are 1/10,000 of an inch. Approximately every 30 minutes, samples of size $n=5$ are collected. The first 35 samples are displayed in Table 1. The precise dimensions are in the range of 3.5205, 3.5202, 3.5204, and so forth. The final three numbers in the measurements are provided by the entries in Table 1. Since we assume that the above non-negative measurements follow the multiplicative model (see relation (11)) satisfying $\ln(X_i)$ is normally distributed we need to compute first as shown in relation (12) the sample $\ln(X_{ij})$, $i=1,\ldots,35$, $j=1,\ldots,5$ before evaluating the value of the estimator of b. For this data set every subgroup $m=1,\ldots,35$ has sample size $n_m=5$ and so

$$\gamma_{n_m} = \gamma_5 = \sqrt[2]{\frac{5-1}{2}}\,\frac{\Gamma\left(\frac{5-1}{2}\right)}{\Gamma\left(\frac{5}{2}\right)} = 1.0638.$$

Applying relations (20) and (22) we obtain using Excel that the mean squared error of the biased estimator of the parameter b is given by

$$\mathrm{MSE}\left(\sum_{m=1}^{M}\alpha_m^{\star}S_{n_m}\right) = 0.0000883\times b^2,$$

Table 1: Cylinder diameter data.

Sample i	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}	Sample i	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}
1	205	202	204	207	205	19	207	206	194	197	201
2	202	196	201	198	202	20	200	204	198	199	199
3	201	202	199	197	196	21	203	200	204	199	200
4	205	203	196	201	197	22	196	203	197	201	194
5	199	196	201	200	195	23	197	199	203	200	196
6	203	198	192	217	196	24	201	197	196	199	207
7	202	202	198	203	202	25	204	196	201	199	197
8	197	196	196	200	204	26	206	206	199	200	203
9	199	200	204	196	202	27	204	203	199	199	197
10	202	196	204	195	197	28	199	201	201	194	200
11	205	204	202	208	205	29	201	196	197	204	200
12	200	201	199	200	201	30	203	206	201	196	201
13	205	196	201	197	198	31	203	197	199	197	201
14	202	199	200	198	200	32	197	194	199	200	199
15	200	200	201	205	201	33	200	201	200	197	200
16	201	187	209	202	200	34	199	199	201	201	201
17	202	202	204	198	203	35	200	204	197	197	199
18	201	198	204	201	201						

and the mean squared error of the unbiased estimator of the parameter b by

$$\text{MSE}\left(\sum_{m=1}^{M} \alpha_m^{\star\star} S_{n_m}\right) = 0.0000936 \times b^2.$$

This shows using relation (23) that the mean squared error of the unbiased estimator is approximately 5% bigger than the mean squared error of the biased estimator.

Since it is shown in Lemma 1 that γ_n satisfies some tight upper and lower bounds we can approximate $\alpha_m^{\star}$ in relation (19) and $\alpha_m^{\star\star}$ in relation (21), $1 \leq m \leq M$ by some simpler expressions. By Lemma 1 it follows that

$$\gamma_n^2 \simeq \left(\frac{n}{n-1}\right)^{\frac{1}{2}}.$$

This shows using $x^{\frac{1}{2}} - 1 \simeq \frac{1}{2}(x-1)$ (first order Taylor approximation around $x=1$) that by relation (19)

$$\alpha_m^{\star} \simeq 2\left(1 + 2\sum_{m=1}^{M}(n_m - 1)\right)^{-1}(n_m - 1)$$

and by relation (21)

$$\alpha_m^{\star\star} \simeq \left(\sum_{m=1}^{M}(n_m-1)\right)^{-1}(n_m-1) = \left(\sum_{m=1}^{M} n_m - M\right)^{-1}(n_m-1).$$

This shows that the minimal mean squared error unbiased estimator of the standard deviation is approximately the pooled standard deviation estimator. At the same time the mean squared error of both estimators listed in relation (20) and (22) satisfy

$$\mathrm{MSE}(\textstyle\sum_{m=1}^{M} \alpha_m^{*} S_{n_m}) \simeq \sigma^2 \left(1 + 2\sum_{m=1}^{M}(n_m - 1)\right)^{-1},$$

and

$$\mathrm{MSE}(\textstyle\sum_{m=1}^{M} \alpha_m^{**} S_{n_m}) \simeq \sigma^2 \left(2\sum_{m=1}^{M}(n_m - 1)\right)^{-1}.$$

Finally we analyze in this section the estimation of the variance if there exist independent samples from the location-scale families m, $m = 1, \ldots, M$ with possibly different means but the same scale parameter. This means that the independent random vectors

$$\mathbf{Y}_m = (Y_{1,m}, \ldots., Y_{n_m,m}),$$

consist of independent and identically distributed random variables having mean zero and variance 1. It is assumed that additionally the random variables $Y_{1,m}$, $m = 1, \ldots, M$ have a finite fourth moment. It follows introducing $f_m^2(\mathbf{Y}_m) = S_{n_m}^2(\mathbf{Y}_m)$ that by relation (33)

$$\mathbb{E}(f_m^2(\mathbf{Y}_m)) = 1, \mathrm{Var}(f_m^2(\mathbf{Y}_m)) = \frac{1}{n_m}\left(\mathbb{E}\left(Y_{1,m}^4\right) - \frac{n_m - 3}{n_m - 2}\right). \tag{24}$$

By Lemma 4 and using relation (24) we obtain that the weights $\alpha_m^{\star}$ of the minimum mean squared error estimator of the variance within the class $\sum_{m=1}^{M} \alpha_m S_{n_m}^2$, $\alpha_m \in \mathbb{R}$ are given by

$$\alpha_m^{\star} = \left(1 + \sum_{m=1}^{M} \frac{n_m}{\mathbb{E}\left(Y_{1,m}^4\right) - \frac{n_m-3}{n_m-1}}\right)^{-1} \frac{n_m}{\mathbb{E}\left(Y_{1,m}^4\right) - \frac{n_m-3}{n_m-1}}, 1 \le m \le M,$$

and this estimator has mean squared error

$$\mathrm{MSE}\left(\textstyle\sum_{m=1}^{M} \alpha_m^{\star} S_{n_m}^2\right) = \sigma^4 \left(1 + \sum_{m=1}^{M} \frac{n_m}{\mathbb{E}\left(Y_{1,m}^4\right) - \frac{n_m-3}{n_m-1}}\right)^{-1}. \tag{25}$$

By Lemma 5 and again relation (24) the weights $\alpha_m^{\star\star}$ of the minimum mean squared error unbiased estimator of the variance within the same class equal

$$\alpha_m^{\star\star} = \left(\sum_{m=1}^{M} \frac{n_m}{\mathbb{E}\left(Y_{1,m}^4\right) - \frac{n_m-3}{n_m-1}}\right)^{-1} \frac{n_m}{\mathbb{E}\left(Y_{1,m}^4\right) - \frac{n_m-3}{n_m-1}}, 1 \le m \le M, \tag{26}$$

and this estimator has mean squared error

$$\mathrm{MSE}\left(\textstyle\sum_{m=1}^{M} \alpha_m^{\star\star} S_{n_m}^2\right) = \sigma^4 \left(\sum_{m=1}^{M} \frac{n_m}{\mathbb{E}\left(Y_{1,m}^4\right) - \frac{n_m-3}{n_m-1}}\right)^{-1}.$$

Since for every $m = 1, \ldots, M$ the cdf of the random variable $Y_{1,m}$ is known we can estimate by simulation the fourth moment or calculate it. Depending on which estimator we prefer to use, we can thus identify for every m the weights $\alpha_m^{\star\star}$ or $\alpha_m^{\star}$. If the random variable $Y_{1,m}$ for some m has a standard normal cdf it follows that its fourth moment equals 3. In some particular cases it is also possible to approximate the optimal weights by a simpler expression. If in each subgroup the random variable $Y_{1,m}$, $m = 1,\ldots,M$, have the same cdf and the sample size n_m in each subgroup is relatively large implying $\frac{n_m-3}{n_m-1} \simeq 1$ it follows by relation (26) that for every $1 \leq m \leq M$ the optimal weights $\alpha_m^{\star\star}$ approximates

$$\alpha_m^{\star\star} \simeq n_m \left(\sum_{m=1}^{M} n_m\right)^{-1}. \tag{27}$$

Hence by relation (27) we observe that for $Y_{1,m}$, $m = 1,\ldots,M$ having the same cdf that the pooled sample variance estimator has a mean squared error close to the minimum mean squared error unbiased estimator of the variance within the class $\sum_{m=1}^{M} \alpha_m S_{n_m}^2$, $\alpha_m \in \mathbb{R}$. Observe in quality control (see section Section 6.3.2 of Montgomery (2009) for identically distributed samples the biased estimator

$$\left(\frac{\sum_{m=1}^{M}(n_m-1)S_{n_m}^2}{\sum_{m=1}^{M} n_m - M}\right)^{\frac{1}{2}},$$

of the standard deviation is recommended without any explanation. By the previous analysis this estimator of the standard deviation is approximately for n_m large and the number of samples small the square root of the minimum mean squared error unbiased estimator (within the class $\sum_{m=1}^{M} \alpha_m S_{n_m}^2$) of the variance. In the next section we introduce sampling costs.

3 On sampling costs in quality control.

Suppose there exists M different independent subgroups from which we can generate samples needed to estimate the unknown standard deviation or variance . It is assumed that all these subgroups are in control and to collect a sample from each of these subgroups, we need to pay sampling costs. It is assumed that generating a sample of size n from subgroup $m = 1,\ldots,M$ has sampling costs $g_m(n)$ with $g_m : \mathbb{Z}_+ \to \mathbb{N}$ denoting the increasing sampling cost function of generating a sample of size n from this subgroup m. We assume for simplicity that the range of the function g_m is an integer and since we only consider samples with sample sizes bigger or equal to 2 we assume without loss of generality for any of the independent subgroups $m = 1,\ldots M$ that $g_m(0) = g_m(1) = 0$. Since in practice we only deal with cost measured in certain units this integer range assumption is not restrictive and simplifies our proposed dynamic solution procedure for the general sampling cost case. We now like to determine, given our available budget $B \in \mathbb{N}$, from which of the independent subgroup m, $m = 1,\ldots,M$ we should generate a sample and how large this sample should be under the condition that our budget constraint is satisfied and the used estimator for the unknown standard deviation or variance has minimum mean squared error over all possible samples satisfying our budget constraint. Clearly, this is a static allocation optimization problem. To formulate this static allocation problem for both the estimation of the variance and the standard deviation introduce for every $1 \leq m \leq M$ the sequence of random vectors $\mathbf{Y}_m(n) = (\mathbf{Y}_{1,m},\ldots,\mathbf{Y}_{n,m})$, $n \in \mathbb{N}$ and the functions

$$f_m(\mathbf{Y}_m(n)) := S_{n,c}(\mathbf{Y}_m(n)) = \sqrt[2]{\frac{1}{n-1}\sum_{i=1}^{n}(Y_{i,m} - \overline{Y}_{n,m})^2},\ n \geq 2,$$

and let the function $h_{m,k}:\mathbb{N}\to(0,\infty)$, $m=1,\dots,m$ be given by

$$h_{m,k}(n)=\begin{cases}0 & \text{if } n=0,1\\ \frac{\left(\mathbb{E}(f_m^k(\mathbf{Y}_m(n))\right)^2}{\operatorname{Var}\left(f_m^k(\mathbf{Y}_m(n))\right)} & n\geq 3\end{cases}$$

By Lemma 4 the static allocation problem for the standard deviation estimation problem is given by

$$\min\left\{b^2\left(1+\sum_{m=1}^M h_{m,1}(n_m)\right)^{-1}:\sum_{m=1}^M g_m(n_m)\leq B, n_j\in\mathbb{N}, 1\leq j\leq m\right\},$$

and for the variance estimation problem by

$$\min\left\{b^4\left(1+\sum_{m=1}^M h_{m,2}(n_m)\right)^{-1}:\sum_{m=1}^M g_m(n_m)\leq B, n_j\in\mathbb{N}, 1\leq j\leq m\right\}.$$

Since $b>0$ it is equivalent to consider for both $k=1$ or $k=2$ the optimization problems

$$\upsilon(P_k):=\max\left\{\sum_{m=1}^M h_{m,k}(n_m):\sum_{m=1}^M g_m(n_m)\leq B, n_j\in\mathbb{N}\right\}.$$

For $k=2$ it follows by relation (25) that the objective functions $h_{m,2}$ for $m=1,\dots,M$ are given by

$$h_{m,2}(n)=\frac{n}{\mathbb{E}\left(Y_{1,m}^4\right)-\frac{n-3}{n-1}}=\frac{n}{\mathbb{E}\left(Y_{1,m}^4\right)-1+\frac{2}{n-1}.}, n\geq 2 \tag{28}$$

It is easy to check that the functions $h_{m,2}$ in relation (28) are increasing. Also for the standard deviation problem we obtain by relation (22) that for normal independent subgroups $m=1,\dots,M$ it follows for every $m=1,\dots M$ that

$$h_{m,1}(n)=(\gamma_n^2-1)^{-1},\ n\geq 2. \tag{29}$$

By Lemma 1 this function is also increasing. Due to the separability of both the objective function and the restriction it is well known that for any increasing sampling cost functions g_m, $1\leq m\leq M$ both optimization problems can be solved by the following dynamic programming recursion formulas. Introduce for $j=1,\dots,M$ the sequences $w_{j,k}:\mathbb{N}\to\mathbb{R}_+$ given by

$$w_{j,k}(y)=\max\left\{\sum_{m=j}^M h_{m,k}(n_m):\sum_{m=j}^M g_m(n_m)\leq y, n_j\in\mathbb{N}, j=m,\dots,M\right\}.$$

Clearly $w_{1,k}(B)=\upsilon(P_k)$ and

$$w_{M,k}(y)=\max\left\{h_{M,k}(n_M):g_M(n_M)\leq y, n_j\in\mathbb{N}\right\}.$$

Also it follows for every $y\in\{0,\dots,B\}$ that

$$w_{j,k}(y)=\max\{h_{j,k}(n_j)+w_{j+1,k}(y-g_j(n_j)):n_j\in\mathbb{N}, g_j(n_j)\in\{0,\dots,y\}\}.$$

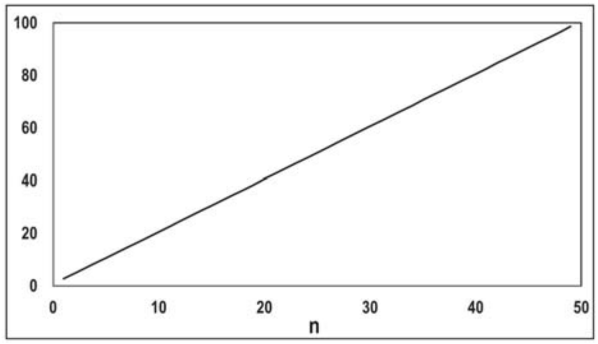

Figure 3: Plot of the function $n \to (\gamma_n^2 - 1)^{-1}$.

Although both allocation problems can be solved by dynamic programming for any set of sampling cost functions by computing iteratively the sequences $w_{j,k}$ from $j = M$ up to 1 it is easy to identify a close to optimal allocation for linear sampling cost functions given by

$$g_m(n) = \begin{cases} 0 & \text{if } n = 0,1 \\ c_m n & \text{if } n \geq 2 \end{cases}, c_m \in \mathbb{N}.$$

Since for $k = 2$ the functions in relation (28) are approximately linear a good heuristic is to spend the whole budget on generating a sample from subgroup m_* satisfying

$$m_* = \arg\min \left\{ c_m(\mathbb{E}(Y_{1,m}^4) - 1) : 1 \leq m \leq M \right\}.$$

In case we consider the standard deviation estimation allocation problem for independent normal subgroups we observe a similar behavior. As already noticed we conjectured using Lemma 1 that $(\gamma_n^2 - 1)^{-1} \approx 2n$ and so the objective function in (29) is almost linear. In Figure 3 we listed a plot of the function $n \to (\gamma_n^2 - 1)^{-1}$ confirming our approximation.

Hence the functions h_m listed in relation (29) are the same for every m and approximately a linear function. This means that we should generate the samples from the cheapest subgroup to obtain the best possible estimator of the process standard deviation. Since all subgroups generate the same type of probabilistic information and give the same additional increase to the objective function, one should, given the available budget, generate a sample having the largest possible size. Such a sample has on average the smallest mean squared error. This means for linear sampling costs it is a good heuristic to generate samples from the cheapest subgroup. In this particular case one can generate the biggest sample size.

4 Conclusion.

In this paper, we give partly an overview and extension of the classical unbiased estimators and their biased extensions used within R and S-charts for estimating the standard variation for single and multiple independent subgroups. We also consider the case of estimating the variance for single and multiple subgroups and unify both approaches for location-scale parameter families. We also propose a simple mathematical model in case we need to pay sampling costs for generating samples from different independent subgroups in our effort to construct an estimator within a certain class of estimators having minimal mean squared error. Analyzing this static model (contrary to for example using dynamic Bayesian type control charts (Makis (2008))) confirms the intuition that, given the available budget and linear sampling cost functions, it is a good strategy to generate the whole sample from one particular subgroup.

5 Acknowledgement

The authors like to thank the constructive comments of the referees which greatly improved the presentation of the paper.

Disclosure statement

The authors report there are no competing interests to declare

References

Abbaszadehpeivasti, H. and Frenk, J. B. G. (2023). On the method of moment approach applied to a generalized gamma distribution. *Communications in Statistics-Theory and Methods*, **52**, 3685–3708.

Alzer, H. (1997). On some inequalities for the gamma and psi functions. *Mathematics of Computation*, **66**, 373–389.

Arciszewski, T. J. (2023). A review of control charts and exploring their utility for regional environmental monitoring programs. *Environments*, **10**, 78.

Arnold, S. F. (1990). *Mathematical Statistics*. Prentice Hall, Englewood Cliffs.

Arnold, B. C. and Balakrishnan, N. and Nagaraja, H. N. (2008). *A First Course on Order statistics*. SIAM, Philadelphia.

Casella, G. and Berger, R. L. (2002). *Statistical Inference (Second Edition)*. Duxbury, Pacific Grove.

Cho, E., and Cho, M. J. (2008). Variance of sample variance. *Section on Survey Research Methods-JSM*, **2**, 1291–1293.

David, H. A. (1970). *Order statistics*. Wiley, New York.

Devor, R., Sutherland, J. and Chang, T-H. (1992). *Statistical Quality Design and Control: Contemporary Concepts and Methods*. Prentice Hall, Upper Saddle River.

Figueiredo, F. and Gomes, M. I. (2006). Box-Cox transformations and robust control charts in SPC. In *Advanced Mathematical and Computational Tools in Metrology* VII, (pp. 35-46), World Scientific, New Jersey.

Harter, H. L. (1960). Tables of range and studentized range. *The Annals of Mathematical Statistics*, **31**, 1122–1147.

Irwin, M. and Miller, M. (2004). *John E. Freund's Mathematical Statistics: With Applications*. Prentice Hall, New York.

Mahmoud, M. A., Henderson, G. R., Epprecht, E. K., and Woodall, W. H. (2010). Estimating the standard deviation in quality-control applications. *Journal of Quality Technology*, **42**, 348–357.

Makis, V. (2008). Multivariate Bayesian control chart. *Operations Research*, **56**, 487–496.

Montgomery, D. C. (2009). *Introduction to Statistical Quality Control*. Wiley, New York.

Ryan, T. P. (2011). *Statistical Methods of Quality Improvement*. Wiley, New York.

Suman, G. and Prajapati, D. (2018). Control chart applications in healthcare: a literature review. *International Journal of Metrology and Quality Engineering*, **9**, 5.

Vardeman, S. B. (1999). A brief tutorial on the estimation of the process standard deviation. *IIE transactions*, **31**, 503–507.

Woodall, W. H., and Montgomery, D. C. (2000). Using ranges to estimate variability. *Quality Engineering*, **13**, 211–217.

Wilks, S. S. (1962). Mathematical Statistics (Second Edition). Wiley, New York.

Zwetsloot, I. M. Jones-Farmer, A. and Woodall, W. H. (2023). Monitoring Univariate processes using control charts: some practical related issues and advice. *Quality Engineering*, **36**, 487–499.

6 Appendix

In the first section of this appendix we give the proofs of the main results mentioned in this paper. In the second subsection we list some useful properties of the gamma function needed to show the log convexity of the sequence γ_n, $n \geq 2$.

6.1 Proof of the main results

We start this subsection with a proof of Lemma 1 discussing the global behaviour of the sequence $\gamma_n, n \geq 2$ occurring in a normal population.

Proof. Since it is well known that $\Gamma(\frac{1}{2}) = \sqrt[2]{\pi}, \Gamma(1) = 1$ and $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ for every $\alpha > 0$, this shows $\gamma_2 = \sqrt[2]{\frac{\pi}{2}}$. Also by the definition of γ_n listed in relation (6), it follows for every $n \geq 2$

$$\gamma_n \gamma_{n+1} = \sqrt[2]{\frac{n-1}{2}} \sqrt[2]{\frac{n}{2}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n+1}{2})} = \frac{\sqrt[2]{\frac{n-1}{2}}}{\frac{n-1}{2}} \sqrt[2]{\frac{n}{2}} = \sqrt[2]{\frac{n}{n-1}}. \tag{30}$$

Hence for every $n \geq 2$ we obtain

$$\gamma_{n+1} = \gamma_n^{-1} \sqrt[2]{\frac{n}{n-1}},$$

and we have verified relation (8). By Lemma 8, the non-negative sequence γ_n, $n \geq 2$ is decreasing and log-convex. This implies using relation (30) that for every $n \geq 2$

$$\gamma_{n+1}^2 \leq \gamma_n \gamma_{n+1} = \sqrt[2]{\frac{n}{n-1}} \leq \gamma_n^2.$$

and we have verified relation (9). Using the inequalities in relation (9), it is easy to verify that the limit relation holds. □

Next we present a proof of Lemma 2.

Proof. It follows for any $k > 0$

$$\mathrm{MSE}(\alpha T^k) = \mathrm{Var}(\alpha T^k) + \left(\mathbb{E}(\alpha T^k) - b^k\right)^2 = \alpha^2 \mathrm{Var}(T^k) + \left(\alpha \mathbb{E}(T^k) - b^k\right)^2. \tag{31}$$

Since $T(\mathbf{X}) \stackrel{d}{=} bf(\mathbf{Y})$ we obtain for every $k > 0$ that $T^k(\mathbf{X}) \stackrel{d}{=} b^k f^k(\mathbf{Y})$. This implies by relation (31) that

$$MSE(\alpha T^k) = \alpha^2 b^{2k} \mathrm{Var}\left(f^k(\mathbf{Y})\right) + \left(\alpha b^k \mathbb{E}(f^k(\mathbf{Y})) - b^k\right)^2 = b^{2k} p_k(\alpha), \tag{32}$$

with

$$p_k(\alpha) = \alpha^2 \mathbb{E}(f^{2k}(\mathbf{Y})) - 2\alpha \mathbb{E}(f^k(\mathbf{Y})) + 1.$$

Since the function p_k is a convex quadratic function and by Lyapunovs inequality and $\mathbb{E}(f^k(\mathbf{Y})) > 0$ we obtain $\mathbb{E}(f^{2k}(\mathbf{Y})) > 0$ the desired result follows by applying the first order conditions to the function p_k. □

The next proof is a proof of Lemma 3 and is actually a special case of the above result for $k = 2$.

Proof. It follows by relation (14) that $S_n(\mathbf{X}) = \sigma f(\mathbf{Y})$ with

$$f^2(\mathbf{Y}) := \frac{1}{n-1} \sum_{i=1}^n (Y_i - \overline{Y}_n)^2.$$

Since S_n^2 is an unbiased estimator of σ^2 and $\mathrm{Var}(Y_i) = 1$ we obtain $\mathbb{E}(f^2(\mathbf{Y})) = 1$. Also it can be shown ((Cho and Cho, 2008),(Wilks, 1962)) that

$$0 \leq \mathrm{Var}\left(\frac{1}{n-1} \sum_{i=1}^n (Y_i - \overline{Y}_n)^2\right) = \frac{1}{n}\left(\mathbb{E}(Y_1^4) - \frac{n-3}{n-1}\right). \tag{33}$$

Since

$$\mathbb{E}(f^4(\mathbf{Y})) = \mathrm{Var}\left(\frac{1}{n-1} \sum_{i=1}^n (Y_i - \overline{Y}_{n)^2}\right) + 1,$$

this yields

$$\mathbb{E}(f^4(\mathbf{Y})) = \frac{1}{n}\left(\mathbb{E}(Y_1^4) - \frac{n-3}{n-1}\right) + 1,$$

and applying Lemma 2 for $k = 2$ we obtain the desired result. □

In the next proof we verify the generalization of the above result to subgroups as expressed in Lemma 4.

Proof. Since the subgroups are independent and hence the random variables T_m, $m = 1,...,M$ are also independent it follows using relations (15), (16) and (17) and introducing the random vector $\mathbf{Y}_m = (Y_{1,m},...,Y_{n_m,m})$ that

$$T(\mathbf{X}_1,...,\mathbf{X}_M) \stackrel{d}{=} b^k \sum_{m=1}^{M} \alpha_m f_m^k(\mathbf{Y}_m).$$

This implies

$$\begin{aligned} \mathrm{MSE}(T) &= \mathrm{Var}(T) + \mathrm{bias}(T)^2 \\ &= b^{2k}\left(\sum_{m=1}^{M} \alpha_m^2 \mathrm{Var}\left(f_m^k(\mathbf{Y}_m)\right) + \left(\sum_{m=1}^{M} \alpha_k \mathbb{E}(f_m(\mathbf{Y}_m)) - 1\right)^2\right). \end{aligned} \tag{34}$$

To determine the optimal mean squared error we need to solve the unconstrained strictly convex quadratic minimization problem

$$\min\left\{\sum_{m=1}^{M} \alpha_m^2 \mathrm{Var}\left(f_m^k(\mathbf{Y}_m)\right) + \left(\sum_{m=1}^{M} \alpha_m \mathbb{E}(f_m^k(\mathbf{Y}_m)) - 1\right)^2 : \alpha_m \in \mathbb{R}, 1 \leq m \leq M\right\}.$$

This optimal solution must be the unique solution of the first order conditions applied to the above objective function and by substituting one can verify that this optimal solution is given by $\alpha^\star = (\alpha_1^\star,....,\alpha_M^\star)$ with

$$\alpha_m^\star = \left(1 + \sum_{j=1}^{M} \frac{\mathbb{E}(f_m^k(\mathbf{Y}_m))}{\mathrm{Var}(f_m^k(\mathbf{Y}_m))}\right)^{-1} \frac{\mathbb{E}(f_m^k(\mathbf{Y}_m))}{\mathrm{Var}(f_m^k(\mathbf{Y}_m))}.$$

Substituting this into relation (34) yields the expression for the bias and optimal mean squared error. □

Finally we list the proof of Lemma 5.

Proof. By relation (34) we need to solve the strict convex optimization problem

$$\min\left\{\sum_{m=1}^{M} \alpha_m^2 \mathrm{Var}(f_m^k(\mathbf{Y}^{(m)})) : \sum_{m=1}^{M} \alpha_m \mathbb{E}(f_m^k(\mathbf{Y}^{(m)})) = 1, \alpha_m \in \mathbb{R}, 1 \leq m \leq M\right\}.$$

Substituting now the suggested solution into the necessary and sufficient KKT conditions, we obtain the desired result. □

6.2 Some useful properties of gamma functions

First, we mention the following well-known definitions.

Definition 2. *The function* $\psi : (0,\infty) \to \mathbb{R}$ *given by*

$$\psi(\alpha) = \frac{\Gamma^{(1)}(\alpha)}{\Gamma(\alpha)},$$

with $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt, \alpha > 0$ *the well-known Gamma function is called the Digamma function.*

Definition 3. *The function $f:(0,\infty)\to\mathbb{R}$ is called completely monotone if all its derivatives exist and for every $x\geq 0$ and $m\in\mathbb{Z}_+$ it holds that*

$$(-1)^m f^{(m)}(x)\geq 0,$$

with $f^{(m)}$ denoting the mth derivative of the function f.

We now list the following result.

Lemma 7. *The function $p:(0,\infty)\to\mathbb{R}$ given by*

$$p(\alpha)=\ln\Gamma\left(\frac{\alpha+1}{2}\right)-\ln\Gamma\left(\frac{\alpha}{2}\right), \tag{35}$$

is non-negative and satisfies $(-1)^{m-1}p^{(m)}(\alpha)\geq 0$ for every $m\in\mathbb{N}$.

Proof. It follows for every $m\in\mathbb{N}$ and $\alpha\geq 0$ that

$$p^{(m)}(\alpha)=\frac{1}{2^{m-1}}\left(\psi^{(m-1)}\left(\frac{\alpha+1}{2}\right)-\psi^{(m-1)}\left(\frac{\alpha}{2}\right)\right). \tag{36}$$

In (Alzer, 1997) it is shown that the function $h_0:(0,\infty)\to\mathbb{R}$ given by $h_0(\alpha)=\ln(\alpha)-\psi(\alpha)$ is completely monotone on $(0,\infty)$. By the definition of h_0 we obtain $\psi(\alpha)=\ln(\alpha)-h_0(\alpha)$ and so for every $m\in\mathbb{N}$

$$\psi^{(m)}(\alpha)=\frac{(-1)^{m-1}(m-1)!}{\alpha^m}-h_0^{(m)}(\alpha).$$

This shows for every $m\in\mathbb{N}$ that

$$(-1)^{m-1}\psi^{(m)}(\alpha)=\frac{(-1)^{2m-2}(m-1)!}{\alpha^m}+(-1)^m\psi^{(m)}(\alpha)\geq 0. \tag{37}$$

Hence for m even it follows by relation (37) that $\psi^{(m)}(\alpha)\leq 0$, and so the function $\alpha\to\psi^{(m-1)}(\alpha)$ is decreasing. For m odd $\psi^{(m)}(\alpha)\geq 0$, and so the function $\alpha\to\psi^{(m-1)}(\alpha)$ is increasing. This shows by relation (36) the result. □

An easy application of Lemma 7 is given by the following result.

Lemma 8. *The function $\gamma_0:(1,\infty)\to\mathbb{R}$ given by $\gamma_0(x)=\ln(\gamma(x))$ with*

$$\gamma(x):=\sqrt[2]{\frac{x-1}{2}}\frac{\Gamma(\frac{x-1}{2})}{\Gamma(\frac{x}{2})},$$

is completely monotone.

Proof. Since $\alpha\Gamma(\alpha)=\Gamma(\alpha+1)$ for every $\alpha\geq 0$ we obtain

$$2\gamma_0(x)=\ln(\gamma(x)^2)=\ln\left(\frac{x-1}{2}\frac{\Gamma^2(\frac{x-1}{2})}{\Gamma^2(\frac{x}{2})}\right)=\ln\left(\frac{\Gamma(\frac{x+1}{2})\Gamma(\frac{x-1}{2})}{\Gamma^2(\frac{x}{2})}\right)=p(x)-p(x-1),$$

with the function p listed in relation (35). This implies

$$\gamma_0^{(m(}(x)=\frac{1}{2}\left(p^{(m)}(x)-p^{(m)}(x-1)\right),$$

and since by Lemma 7, we know that $(-1)^m p^{(m+1)}(x)\geq 0$ we obtain the desired result. □