Quality Control with Unequal Sample Sizes (Motivation)

Chanseok PARK[1] Linhan OUYANG[2] Min WANG[3]

[1]Applied Statistics Laboratory
Department of Industrial Engineering, Pusan National University

[2]College of Economics and Management
Nanjing University of Aeronautics and Astronautics

[3]Department of Management Science and Statistics
The University of Texas at San Antonio

December 3, 2020

Hosted by NUAA

Overview

Overview

Overview

1. Overview
2. Motivation from Shewhart-type $\bar{X}$ and S charts
3. Location parameter estimation
4. Scale parameter estimation
5. Inequalities and their proofs
6. Summary

Overview

Overview

Overview

Talk-1

- Motivation from Control Charts
- We will look at how to pool (combine) location and scale estimators.
- Which methods are preferred? We will provide some proofs (time permitting).

Talk-2

- In Talk-1, we provided some theoretical works under nice conditions (normality and no contamination).
- Some other applications using the method considered in Talk-1.
- What if these conditions are violated?
  Some current works
  Discussions

Motivation from Shewhart-type $\bar{X}$ and S charts

Shewhart-type $\bar{X}$ and S charts

In general, the Shewhart-type control charts are made up of

- **UCL**: the upper control limit
- **CL**: the center line
- **LCL**: the lower control limit

This control charts are widely used and well developed when sample sizes are **equal**.

- However, when sample sizes are **not equal**, there exist only *ad hoc* methods.
- For the location estimator, two estimators are used.
  Then which one is better?
- Especially for the scale estimator, they are all **biased**.

Basics on the Shewhart-type charts

Assumptions

- m samples but each subgroup has a **different sample size**.
- Let X_{ij} denote the ith subgroup of size n_i,
  where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n_i$.
- Assume that X_{ij} are IID normal normal with mean μ and variance σ^2.

$\bar{X}$ chart

For the ith $(i = 1, 2, \ldots, m)$ subgroup, we have

$$\frac{\bar{X}_i - E(\bar{X}_i)}{\mathrm{SE}(\bar{X}_i)} \overset{\bullet}{\sim} N(0, 1),$$

where

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}.$$

$\bar{X}$ chart — continued

Since $E(\bar{X}_i) = \mu$ and $\mathrm{SE}(\bar{X}_i) = \sigma/n_i$, the $\mathrm{CL} \pm 3 \cdot \mathrm{SE}$ control limits are obtained by setting

$$\frac{\bar{X}_i - \mu}{\sigma/n_i} = \pm 3$$

and solving for $\bar{X}_i$.

Note that we can obtain the CI if we solve the above for μ.

Basics on the Shewhart-type charts

$\bar{X}$ chart — continued

Solving for $\bar{X}_i$, the $\bar{X}$ control chart is constructed as follows

$$\begin{aligned} \mathrm{UCL} &= \mu + \frac{3\sigma}{\sqrt{n_i}} \\ \mathrm{CL} &= \mu \\ \mathrm{LCL} &= \mu - \frac{3\sigma}{\sqrt{n_i}}. \end{aligned}$$

Issues

- The population parameters μ and σ are **not** known in general.
- Especially, conventional *ad hoc* scale estimators for σ are **biased** when the sample sizes are **not** equal.

S chart

Similarly, we also have

$$\frac{S_i - E(S_i)}{\mathrm{SE}(S_i)} \overset{\bullet}{\sim} N(0,1),$$

where S_i is the sample standard deviation with a sample of size n_i. For the $\mathrm{CL} \pm 3 \cdot \mathrm{SE}$ control limits, we can set up

$$\frac{S_i - E(S_i)}{\mathrm{SE}(S_i)} = \pm 3,$$

where $E(S_i) = c_4(n_i)\sigma$, $\mathrm{Var}(S_i) = (1 - c_4(n_i)^2)\sigma^2$, and

$$c_4(n_i) = \sqrt{\frac{2}{n_i - 1}} \frac{\Gamma(n_i/2)}{\Gamma((n_i - 1)/2)}$$

S chart — continued

Using $E(S_i) = c_4(n_i)\sigma$ and $\mathrm{Var}(S_i) = (1 - c_4(n_i)^2)\sigma^2$, we have

$$\frac{S_i - c_4(n_i)\sigma}{\sqrt{1 - c_4(n_i)^2}\,\sigma} = \pm 3.$$

Solving for S_i,

$$\begin{aligned}
\mathrm{UCL} &= c_4(n_i)\sigma + 3\sqrt{1 - c_4(n_i)^2}\sigma \\
\mathrm{CL} &= c_4(n_i)\sigma \\
\mathrm{LCL} &= c_4(n_i)\sigma - 3\sqrt{1 - c_4(n_i)^2}\sigma.
\end{aligned}$$

Issue again

- Especially, conventional *ad hoc* scale estimators for σ are **biased**.

Location parameter estimation

Two conventional estimators for μ

Montgomery (2013) provided the following location estimators for μ in Equations (6.2) and (6.30) of his book

- $\bar{\bar{X}}_A = \dfrac{\bar{X}_1 + \bar{X}_2 + \cdots + \bar{X}_m}{m} = \dfrac{1}{m} \sum_{i=1}^{m} \bar{X}_i$
- $\bar{\bar{X}}_B = \dfrac{n_1\bar{X}_1 + n_2\bar{X}_2 + \cdots + n_m\bar{X}_m}{N} = \dfrac{1}{N} \sum_{i=1}^{m} n_i\bar{X}_i.$

where $N = \sum_{i=1}^{m} n_i$.

Note that these estimators for μ are unbiased
while the *ad hoc* estimates for σ are **biased**.

Issue

Which one of $\bar{\bar{X}}_A$ or $\bar{\bar{X}}_B$ is selected?
The one with the smaller variance!

Location parameter estimation

Theorem 1

If X_{ij} are normal with mean μ and variance σ^2, then we have

$$\mathrm{Var}(\bar{\bar{X}}_A) \geq \mathrm{Var}(\bar{\bar{X}}_B).$$

Proof.

It is immediate from $\mathrm{Var}(\bar{X}_i) = \sigma^2/n_i$ that we have

- $\mathrm{Var}(\bar{\bar{X}}_A) = \frac{1}{m^2}\sum_{i=1}^m \mathrm{Var}(\bar{X}_i) = \frac{\sigma^2}{m^2}\sum_{i=1}^m \frac{1}{n_i} = \frac{\sigma^2}{m}\frac{1}{\bar{n}_h}$
- $\mathrm{Var}(\bar{\bar{X}}_B) = \frac{1}{N^2}\sum_{i=1}^m n_i^2 \mathrm{Var}(\bar{X}_i) = \frac{\sigma^2}{N^2}\sum_{i=1}^m n_i = \frac{\sigma^2}{N} = \frac{\sigma^2}{m}\frac{1}{\bar{n}_a}$

where

$$\bar{n}_h = \left(\frac{1}{m} \sum_{i=0}^{m} n_i^{-1} \right)^{-1} \quad \text{and} \quad \bar{n}_a = \frac{1}{m} \sum_{i=1}^{m} n_i.$$

Then we have

$$\frac{\mathrm{Var}(\bar{\bar{X}}_A)}{\mathrm{Var}(\bar{\bar{X}}_B)} = \frac{\bar{n}_a}{\bar{n}_h}.$$

- $\bar{n}_h$ and $\bar{n}_a$ are harmonic mean and arithmetic mean, respectively.
- $\bar{n}_a \geq \bar{n}_h$.
- Thus, $\mathrm{Var}(\bar{\bar{X}}_A)/\mathrm{Var}(\bar{\bar{X}}_B) \geq 1$ which results in $\mathrm{Var}(\bar{\bar{X}}_A) \geq \mathrm{Var}(\bar{\bar{X}}_B)$.

$\square$

Which one of $\bar{\bar{X}}_A$ or $\bar{\bar{X}}_B$ is selected? **$\bar{\bar{X}}_B$!**

Scale parameter estimation

$\bar{X}$ chart — Recalled

Substituting $\mu = \bar{\bar{X}}_B$, we have

$$\text{UCL} = \bar{\bar{X}}_B + \frac{3\sigma}{\sqrt{n_i}}$$
$$\text{CL} = \bar{\bar{X}}_B$$
$$\text{LCL} = \bar{\bar{X}}_B - \frac{3\sigma}{\sqrt{n_i}}.$$

Thus, we need to estimate σ.

Scale parameter estimation

Section 6.3.2 of Montgomery (2013) recommended the followings to estimate σ:

- S_p
  where $S_p^2 = \sum_{i=1}^{m}(n_i - 1)S_i^2/(N - m)$ (pooled sample variance).

- $\bar{S}^*$
  where $\bar{S}^* = \bar{S}/c_4(\bar{n})$, $\bar{S} = \sum_{i=1}^{m} S_i/m$ and $\bar{n} = \sum_{i=1}^{m} n_i/m$.

Issues

- Are they unbiased?
- Which one is better?

Scale parameter estimation

Note that S_p^2 is unbiased for σ^2 so S_p should underestimate σ. This is easily proved using the Jensen's inequality as follows.

Since $\sqrt{T}$ is concave, we have

$$E(\sqrt{T}) < \sqrt{E(T)}.$$

With $T = S_p^2$, we have

$$E(\sqrt{S_p^2}) = \boldsymbol{E(S_p)} < \sqrt{E(S_p^2)} = \sqrt{\sigma^2} = \boldsymbol{\sigma} \implies E(S_p) < \sigma.$$

- S_p always **underestimate** σ,
- so it is also **biased**.

Note: whenever S^2 is unbiased for σ^2, S underestimates σ regardless of an underlying distribution.

Scale parameter estimation

Then how about $\bar{S}^*$?

Recall $\bar{S}^*$

$$\bar{S}^* = \frac{\bar{S}}{c_4(\bar{n})},$$

where $\bar{S} = \sum_{i=1}^m S_i/m$ and $\bar{n} = \sum_{i=1}^m n_i/m$.

It **underestimates** $\boldsymbol{\sigma}$ since

$$E[\bar{S}^*] = \frac{\frac{1}{m}\sum_{i=1}^m E[S_i]}{c_4(\bar{n})\sigma} = \frac{\frac{1}{m}\sum_{i=1}^m c_4(n_i)}{c_4(\bar{n})} \cdot \sigma < \sigma.$$

How to prove the above? The claim is to show (the proof is tricky)

$$\frac{1}{m}\sum_{i=1}^m c_4(n_i) < c_4(\bar{n}) \qquad \text{(the proof will be provided later)}$$

Scale parameter estimation

As aforementioned, both S_p and $\bar{S}^*$ underestimate σ — **biased!**

We propose

- $\bar{S}_A = \dfrac{S_1/c_4(n_1) + S_2/c_4(n_2) + \cdots + S_m/c_4(n_m)}{m} = \dfrac{1}{m}\displaystyle\sum_{i=1}^{m} \dfrac{S_i}{c_4(n_i)}$
- $\bar{S}_B = \dfrac{S_1 + S_2 + \cdots + S_m}{c_4(n_1) + c_4(n_2) + \cdots + c_4(n_m)} = \dfrac{\sum_{i=1}^{m} S_i}{\sum_{i=1}^{m} c_4(n_i)}$

Using $E(S_i) = c_4(n_i)\sigma$ under the normality assumption, we see that these are **unbiased** for σ.

Issue

Which one is used?

That is, which one has a smaller variance?

Answer: $\bar{S}_B$ (the proof will be provided later).

Scale parameter estimation

It is immediate upon using

$$\mathrm{Var}(S_i) = E(S_i^2) - E(S_i)^2 = \sigma^2\{1 - c_4(n_i)^2\}$$

that we have

$$\mathrm{Var}(\bar{S}_A) = \frac{\sigma^2}{m^2}\sum_{i=1}^{m}\left\{\frac{1}{c_4(n_i)^2} - 1\right\}$$

and

$$\mathrm{Var}(\bar{S}_B) = \sigma^2 \cdot \frac{\sum_{i=1}^{m}\{1 - c_4(n_i)^2\}}{\left\{\sum_{i=1}^{m} c_4(n_i)\right\}^2}.$$

In the following section, we will provide several inequalities which are needed for proving it.

Inequalities and their proofs

In the previous sections, the followings are left to prove

- $\mathrm{Var}(\bar{S}_A) \geq \mathrm{Var}(\bar{S}_B)$
- $E[\bar{S}^*] = \dfrac{\frac{1}{m}\sum_{i=1}^{m} E[S_i]}{c_4(\bar{n})\sigma} = \dfrac{\frac{1}{m}\sum_{i=1}^{m} c_4(n_i)}{c_4(\bar{n})} \cdot \sigma < \sigma$

  Note: $\bar{S}^*$ is recommended for the estimator of σ in Section 6.3.2 of Montgomery (2013).

- To prove the above, it is essential to figure out the behaviors of the c_4 function which is again given by

$$c_4(x) = \sqrt{\frac{2}{(x-1)}}\frac{\Gamma(x/2)}{\Gamma((x-1)/2)}$$

- To this end, we will use the Wallis (1656) and Watson (1959) representations.

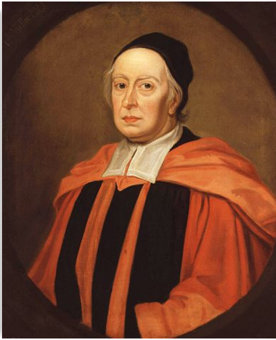
Inequalities and their proofs

Wallis' production formula (Wallis, 1656)

$$\frac{2}{1}\cdot\frac{4}{3}\cdot\frac{6}{5}\cdot\frac{8}{7}\cdot\frac{10}{9}\cdot\frac{12}{11}\cdots\to\infty \quad \text{(diverges to } \boldsymbol{\infty}\text{)}$$

$$\times\ \frac{2}{3}\cdot\frac{4}{5}\cdot\frac{6}{7}\cdot\frac{8}{9}\cdot\frac{10}{11}\cdot\frac{12}{13}\cdots\to 0 \quad \text{(trivially goes to } \mathbf{0}\text{)}$$

$$=\frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdot\frac{8}{7}\cdot\frac{8}{9}\cdot\frac{10}{9}\cdot\frac{10}{11}\cdots\to\frac{\boldsymbol{\pi}}{\mathbf{2}}$$



Another version is also given by

$$\frac{1}{\sqrt{\pi(n+\frac{1}{2})}} < \boxed{\frac{1}{2}\cdot\frac{3}{4}\cdot\frac{5}{6}\cdot\cdots\cdot\frac{2n-1}{2n}} < \frac{1}{\sqrt{\pi n}}$$

Inequalities and their proofs

The Wallis' production formula can also be rewritten as

$$\boxed{\frac{1}{2}\cdot\frac{3}{4}\cdot\frac{5}{6}\cdot\cdots\cdot\frac{2n-1}{2n}} = \frac{1}{\sqrt{\pi(n+\theta)}},$$

where and $n \geq 1$ and $0 < \theta < \frac{1}{2}$.
This formula will play a pivotal role in proving the inequalities in this talk.

Note: it is provided in his book, *Arithmetica Infinitorum* in the year of 1656, written in Latin. Recently, Stedall (2004) translated into English.

- Extension of the Wallis' production formula by Watson (1959)
- $c_4(x)$ function written by the Watson's $\theta(x)$ function
- Using the Watson's $\theta(x)$, we can provide several useful results

Inequalities and their proofs

Recall

$$\boxed{\frac{1}{2}\cdot\frac{3}{4}\cdot\frac{5}{6}\cdot\cdots\cdot\frac{2n-1}{2n}} = \frac{1}{\sqrt{\pi(n+\theta)}},$$

where and $n \geq 1$ and $0 < \theta < \frac{1}{2}$. We have

$$\begin{aligned}\frac{\Gamma(n+\frac{1}{2})}{\Gamma(n+1)} &= \frac{\sqrt{\pi}\cdot\frac{1}{2}}{1}\cdot\frac{1+\frac{1}{2}}{2}\cdot\frac{2+\frac{1}{2}}{3}\cdot\cdots\cdot\frac{n-\frac{1}{2}}{n}\\ &= \sqrt{\pi}\cdot\boxed{\frac{1}{2}\cdot\frac{3}{4}\cdot\frac{5}{6}\cdot\cdots\cdot\frac{2n-1}{2n}} = \frac{1}{\sqrt{n+\theta}}\end{aligned}$$

Motivated by the above, Watson (1959) extended to

$$\frac{\Gamma(x+\frac{1}{2})}{\Gamma(x+1)} = \frac{1}{\sqrt{x+\theta(x)}}, \tag{1}$$

where $x > -\frac{1}{2}$ and $0 < \theta(x) < \frac{1}{2}$.

Inequalities and their proofs

Solving (1) for $\theta(x)$, we have

$$\theta(x) = -x + \frac{\Gamma(x+1)^2}{\Gamma(x+\frac{1}{2})^2} = -x + x\frac{\Gamma(x)\Gamma(x+1)}{\Gamma(x+\frac{1}{2})^2}. \tag{2}$$

Since

$$c_4(x)^2 = \frac{1}{(x-1)/2}\frac{\Gamma(x/2)^2}{\Gamma((x-1)/2)^2} = \frac{\Gamma(x/2)^2}{\Gamma((x-1)/2)\Gamma((x+1)/2)},$$

we have

$$c_4(2x+1)^2 = \frac{\Gamma(x+\frac{1}{2})^2}{\Gamma(x)\Gamma(x+1)}. \tag{3}$$

Substituting (3) into (2), we have

$$\boxed{\theta(x) = -x + x \cdot \frac{1}{c_4(2x+1)^2}} \quad \text{or} \quad \boxed{c_4(2x+1) = \sqrt{\frac{x}{\theta(x)+x}}}$$

- We succeeded in rewriting c_4 function using the Watson's $\theta(x)$ function (extended version of Wallis product).
- Using the nice properties of $\theta(x)$, we can get several useful results easily.

Properties of $\theta(x)$

- $0 < \theta < \frac{1}{2}$.
- $\theta(x)$ is monotonically decreasing.

Inequalities and their proofs

Lemma 1

- $\sqrt{\dfrac{x-1}{x}} < c_4(x) < 1$

Proof.

Recall $\theta(x) = -x + x/c_4(2x+1)^2$ and $0 < \theta(x) < \frac{1}{2}$. Let $x = (t-1)/2$. Then we have

$$0 < -x + \frac{x}{c_4(2x+1)^2} < \frac{1}{2} \implies 0 < -\left(\frac{t-1}{2}\right) + \frac{(t-1)/2}{c_4(t)^2} < \frac{1}{2}.$$

Note: $c_4(t)$ is defined on $t > 1$. Solving the above for $c_4(t)^2$, we have

$$\frac{t-1}{t} < c_4(t)^2 < 1.$$

□

Inequalities and their proofs

Lemma 2

- $c_4(x)$ is monotonically increasing
- The function

$$\frac{1}{c_4(x)^2} - 1$$

is monotonically decreasing.

Proof.

Proofs are omitted. Please refer to the manuscript. □

Inequalities and their proofs

- Recall: we claim $\mathrm{Var}(\bar{S}_A) \geq \mathrm{Var}(\bar{S}_B)$.
- To prove this, we need Lemma 3 below.

Lemma 3 (Chebyshev's sum inequality)

If $a_1 \geq a_2 \geq \cdots \geq a_m$ and $b_1 \geq b_2 \geq \cdots \geq b_m$, then

$$m \sum_{i=1}^{m} a_i b_i \geq \sum_{i=1}^{m} a_i \cdot \sum_{i=1}^{m} b_i.$$

Similarly, if $a_1 \leq a_2 \leq \cdots \leq a_m$ and $b_1 \geq b_2 \geq \cdots \geq b_m$, then

$$m \sum_{i=1}^{m} a_i b_i \leq \sum_{i=1}^{m} a_i \cdot \sum_{i=1}^{m} b_i.$$

- The original formula is in an integral form (Chebyshev, 1882).
- Its proof can be found in Chebyshev (1883).

Inequalities and their proofs

Proof of Chebyshev's sum inequality.

When a_i and b_i are both decreasing, $(a_i - a_j)$ and $(b_i - b_j)$ have the **same sign** for any i and j. Thus, we have

$$\sum_{i=1}^{m}\sum_{j=1}^{m}(a_i - a_j)(b_i - b_j) \geq 0,$$

which results in

$$2m\sum_{i=1}^{m} a_i b_i - 2\sum_{i=1}^{m} a_i \sum_{j=1}^{m} b_j \geq 0.$$

Next, when a_i are increasing and b_i are decreasing, $(a_i - a_j)$ and $(b_i - b_j)$ have **different signs** or zero for any i and j, Thus, we have $\sum_{i=1}^{m}\sum_{j=1}^{m}(a_i - a_j)(b_i - b_j) \leq 0$. $\square$

Inequalities and their proofs

Recall: we propose

- $$\bar{S}_A = \frac{S_1/c_4(n_1) + S_2/c_4(n_2) + \cdots + S_m/c_4(n_m)}{m} = \frac{1}{m}\sum_{i=1}^{m}\frac{S_i}{c_4(n_i)}$$
- $$\bar{S}_B = \frac{S_1 + S_2 + \cdots + S_m}{c_4(n_1) + c_4(n_2) + \cdots + c_4(n_m)} = \frac{\sum_{i=1}^m S_i}{\sum_{i=1}^m c_4(n_i)}$$

Theorem 2

We have $\mathrm{Var}(\bar{S}_A) \geq \mathrm{Var}(\bar{S}_B)$.

Recall: we showed

$$\mathrm{Var}(\bar{S}_A) = \frac{\sigma^2}{m^2}\sum_{i=1}^m\left\{\frac{1}{c_4(n_i)^2} - 1\right\} \quad \text{and} \quad \mathrm{Var}(\bar{S}_B) = \sigma^2 \cdot \frac{\sum_{i=1}^m\left\{1 - c_4(n_i)^2\right\}}{\left\{\sum_{i=1}^m c_4(n_i)\right\}^2}$$

Inequalities and their proofs

Sketch proof.

It suffices to show

$$\frac{1}{m^2}\sum_{i=1}^{m}\left\{\frac{1}{c_4(n_i)^2}-1\right\} \geq \frac{\sum_{i=1}^{m}\left\{1-c_4(n_i)^2\right\}}{\left\{\sum_{i=1}^{m} c_4(n_i)\right\}^2}$$

- First, let $a_i = c_4(n_i)$ (increasing) and $b_i = 1/c_4(n_i)^2 - 1$ (decreasing)
- Apply the Chebyshev's sum inequality with a_i and b_i. Then we have an intermediate result.
- Next, let $a_i = c_4(n_i)$ (increasing) and $b_i = 1/c_4(n_i) - c_4(n_i)$ (decreasing)
- Apply the Chebyshev's sum inequality again to the above intermediate result with a_i and b_i. Then we have a result.

□

Inequalities and their proofs

Recall

Section 6.3.2 of Montgomery (2013) uses

$$\bar{S}^* = \frac{\bar{S}}{c_4(\bar{n})},$$

where $\bar{S} = \sum_{i=1}^{m} S_i/m$ and $\bar{n} = \sum_{i=1}^{m} n_i/m$.

Claim: $\bar{S}^*$ **underestimates** $\boldsymbol{\sigma}$ since

$$E[\bar{S}^*] = \frac{\frac{1}{m}\sum_{i=1}^{m} E[S_i]}{c_4(\bar{n})\sigma} = \frac{\frac{1}{m}\sum_{i=1}^{m} c_4(n_i)}{c_4(\bar{n})} \cdot \sigma < \sigma.$$

It suffices to show

$$\frac{1}{m}\sum_{i=1}^{m} c_4(n_i) < c_4(\bar{n})$$

Lemma 4

The $c_4(x)$ function is log-concave.

Sketch proof.

Recall the c_4 function which is again given by

$$c_4(x) = \sqrt{\frac{2}{(x-1)}}\frac{\Gamma(x/2)}{\Gamma((x-1)/2)}$$

Thus, we have

$$c_4(x)^2 = \frac{1}{(x-1)/2}\frac{\Gamma(x/2)^2}{\Gamma((x-1)/2)^2} = \frac{\Gamma(x/2)^2}{\Gamma((x-1)/2)\Gamma((x+1)/2)}.$$

Sketch proof – continued.

- Taking the logarithm and dividing by two, we have

$$\log c_4(x) = \log\Gamma\Big(\frac{x}{2}\Big) - \frac{1}{2}\log\Gamma\Big(\frac{x-1}{2}\Big) - \frac{1}{2}\log\Gamma\Big(\frac{x+1}{2}\Big).$$

- From Section 11.14 (iv) of Schilling (2005) and Merkle (1996), the second derivative of $\log\Gamma(x)$ can be written as the sum of the series so that we have

$$\frac{d^2}{dx^2}\log\Gamma(x) = \sum_{k=0}^{\infty}\frac{1}{(x+k)^2}. \tag{4}$$

Sketch proof — continued.

After tedious algebra, we have

$$\begin{aligned}\frac{d^2}{dx^2}\log c_4(x) &= \sum_{k=0}^{\infty}\frac{1}{(\frac{x}{2}+k)^2} - \frac{1}{2}\sum_{k=0}^{\infty}\frac{1}{(\frac{x-1}{2}+k)^2} - \frac{1}{2}\sum_{k=0}^{\infty}\frac{1}{(\frac{x+1}{2}+k)^2} \\ &= \sum_{k=0}^{\infty}\frac{-12(x+2k)^2+4}{(x+2k)^2(x-1+2k)^2(x+1+2k)^2} < 0.\end{aligned}$$

- Thus, we have $\frac{d^2}{dx^2}\big(-\log c_4(x)\big) > 0$ so $-\log c_4(x)$ is convex due to Theorem 6.4.6 of Bartle and Sherbert (2011).
- Then $\log c_4(x)$ is concave and $c_4(x)$ is thus log-concave, which completes the proof.

□

Inequalities and their proofs

Lemma 5

We have the following inequality

$$\frac{1}{m}\sum_{i=1}^{m} c_4(n_i) \leq c_4(\bar{n}).$$

Proof.

The log-concavity of $c_4(x)$ from Lemma 4 guarantees that $c_4(x)$ is concave.
Thus, we can apply the Jensen's inequality to $c_4(x)$ and we have

$$\frac{1}{m}\sum_{i=1}^{m} c_4(n_i) \leq c_4(\bar{n}),$$

where $\bar{n} = \sum_{i=1}^{m} n_i/m$. This completes the proof. $\square$

Inequalities and their proofs

Burr (1969) proposed

$$\bar{S}_C = \frac{\displaystyle\sum_{i=1}^{m} \frac{c_4(n_i)S_i}{1-c_4(n_i)^2}}{\displaystyle\sum_{i=1}^{m} \frac{c_4(n_i)^2}{1-c_4(n_i)^2}} \quad \text{and} \quad \bar{S}_D = \frac{S_p}{c_4(N-m+1)}, \tag{5}$$

where $S_p^2 = \sum_{i=1}^{m}(n_i-1)S_i^2/(N-m)$. Park and Wang (2020) considered

$$\bar{S}_E = \frac{S_N}{c_4(N)},$$

where $S_N^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(X_{ij}-\bar{\bar{X}})^2/(N-1)$ and $\bar{\bar{X}} = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i} X_{ij}$. This $\bar{S}_E$ is the UMVUE (uniformly minimum variance unbiased estimator). See Theorem 7.3.23 of Casella and Berger (2002) and Definition 1.6 of Lehmann and Casella (1998).

Inequalities and their proofs

In summary, we have five choices

- $\bar{S}_A = \frac{1}{m} \sum_{i=1}^{m} \frac{S_i}{c_4(n_i)}$
- $\bar{S}_B = \frac{\sum_{i=1}^{m} S_i}{\sum_{i=1}^{m} c_4(n_i)}$
- $\bar{S}_C = \frac{\displaystyle\sum_{i=1}^{m} \frac{c_4(n_i) S_i}{1 - c_4(n_i)^2}}{\displaystyle\sum_{i=1}^{m} \frac{c_4(n_i)^2}{1 - c_4(n_i)^2}}$
- $\bar{S}_D = S_p / c_4(N - m + 1)$ Note: S_i^2 are pooled, but $\bar{X}_i$ are not pooled.
- $\bar{S}_E = S_N / c_4(N)$ Note: S_i^2 are kind-of pooled and $\bar{X}_i$ are too.

Inequalities and their proofs

Park and Wang (2020) proved that

- $\bar{S}_A$, $\bar{S}_B$, $\bar{S}_C$, $\bar{S}_D$, and $\bar{S}_E$ are all unbiased
- $\boxed{\mathrm{Var}(\bar{S}_A) \geq \mathrm{Var}(\bar{S}_B) \geq \mathrm{Var}(\bar{S}_C) \geq \mathrm{Var}(\bar{S}_D) \geq \mathrm{Var}(\bar{S}_E)}$

Then, is $\mathrm{Var}(\bar{S}_E)$ the best? **Maybe not.**

Recall: $\bar{S}_D$ and $\bar{S}_E$

- $\bar{S}_D = S_p/c_4(N-m+1)$ Note: S_i^2 are pooled, but $\bar{X}_i$ are not pooled.
  Weak (less powerful) under heteroscedasticity (Burr, 1969)
  (due to pooled S_i^2).
- $\bar{S}_E = S_N/c_4(N)$ Note: S_i^2 are kind-of pooled and $\bar{X}_i$ are too.
  Weak either under heteroscedasticity or alternative H_1 (out of control)
  (due to pooled S_i^2 and $\bar{\bar{X}}$).

Summary

Summary

- The c_4 function plays a pivotal role in constructing the $\bar{X}$ and S charts.
- With the useful inequalities, we showed that the conventional *ad hoc* methods are **biased**, which are used for the $\bar{X}$ and S charts.
- We also proposed new **unbiased** estimators for σ which can be used for various control charts.
- Thus, we can construct **proper** $\bar{X}$ **and** S **charts** based on the new **unbiased** estimators.
- The behaviors of the c_4 function are much understood thanks to the connection with Wallis' product and Watson representation.
- $\bar{S}_D$ or $\bar{S}_E$ seem to be the best mathematically, but they may not be preferred over $\bar{S}_C$. (because we need to consider statistical powers under heteroscedasticity or out-of-control).

  $\Rightarrow$ We need to investigate these more thoroughly.

References

Bartle, R. G. and Sherbert, D. R. (2011). Introduction to Real Analysis. Wiley, 4th edition.

Burr, I. W. (1969). Control charts for measurements with varying sample sizes. Journal of Quality Technology, 1:163–167.

Casella, G. and Berger, R. L. (2002). Statistical Inference. Duxbury, Pacific Grove, CA, second edition.

Chebyshev, P. L. (1882). Sur les expressions approximatives des intégrales définies par les autres prises entre les même limites. In Markov, A. A. and Sonin, N., editors, Oeuvres de P. L. Tchebychef I–II, Vol. 2, pages 716–719. Imprimerie de l'Academie imperiale des sciences, St. Petersbourg.

References

Chebyshev, P. L. (1883). Sur une série qui fournit les valeurs extrêmes des intégrales, lorsque la fonction sous le signe est décomposée en deux facteurs. In Markov, A. A. and Sonin, N., editors, Oeuvres de P. L. Tchebychef I–II, Vol. 2, pages 405–419. Imprimerie de l'Academie imperiale des sciences, St. Petersbourg.

Lehmann, E. L. and Casella, G. (1998). Theory of Point Estimation. Springer-Verlag, New York, second edition.

Merkle, M. (1996). Logarithmic convexity and inequalities for the gamma function. Journal of Mathematical Analysis and Applications, 203:369–380.

Montgomery, D. C. (2013). Statistical Quality Control: An Modern Introduction. John Wiley & Sons, Singapore, 7th edition.

Park, C. and Wang, M. (2020). A study on the X-bar and S control charts with unequal sample sizes. Mathematics, 8(5):698.

References

Schilling, R. L. (2005). Measures, Integrals and Martingales. Cambridge University Press, Cambridge, UK.

Stedall, J. A. (2004). Arithmetica Infinitorum: John Wallis 1656. Springer, New York. (English translation from the original book written in Latin).

Wallis, J. (1656). Arithmetica Infinitorum. Oxford, England.

Watson, G. N. (1959). A note on gamma functions. Edinburgh Mathematical Notes, 42:7–9.