Responses to the Reviewer's Comments

Review

The discussed problem in the paper "A note on the g and h control charts" has as it's main objective to use unbiased estimators (minimum variance unbiased - MVU) and parameter estimation with unequal sample sizes in "g-chart" and "h-chart." In the literature, biased estimators are used. I enjoyed the paper and have not found any problems in relation to the math used in it. However, I understand that it is necessary to discuss the real usefulness of MVU estimators in the context of statistical process control (SPC). In my opinion, the work would be publishable if the following points were strengthened prior to further evaluation.

We would like to express our sincere thanks for your constructive and positive comments. According to your valuable comments, we have carefully addressed your comments point by point in the revision as below.

In the area of statistical process control, the unbiasedness of parameter estimation is of significant importance. For instance, it is known that the control limits are obtained by solving the below for $\bar{X}_k$

$$\frac{\bar{X}_k - \mu}{\sqrt{\sigma^2/n_k}} = \pm g,$$

whereas the confidence limits (or confidence interval) are obtained by solving the above for μ with $g = z_{\alpha/2}$. As shown in the paper, we obtain that

$$\text{UCL} = \mu + g\sqrt{\frac{\sigma^2}{n_k}} = \frac{1-p}{p} + a + g\sqrt{\frac{1-p}{n_k p^2}},$$

$$\text{CL} = \mu = \frac{1-p}{p} + a,$$

$$\text{LCL} = \mu - g\sqrt{\frac{\sigma^2}{n_k}} = \frac{1-p}{p} + a - g\sqrt{\frac{1-p}{n_k p^2}}.$$

To construct the control limits, we need to estimate the unknown parameter, which is usually obtained in Phase-I, where there are m samples, In Phase-II, we often have a large single sample with a sample size, which can provide an estimator for the parameter with a small variance. Under this scenario, the unbiasedness issue is of particular importance. Another example is the famous unbiasing constant for the standard deviation under normality. That is $E(S/c_4) = \sigma$ with $c_4 = \sqrt{2/(n-1)}\Gamma(n/2)/\Gamma((n-1)/2)$. The unbiased estimator (S/c_4) for σ is a de facto standard method in the $\bar{X}$ control chart. Thus, it is reasonable to select the best estimator among all the unbiased estimators, which is the MVU estimator.

Thanks so much for your great suggestion. For clarification purposes, We briefly introduced Phase I and Phase II at the beginning of Section 3. Please refer to the red parts on pages 4 and 5 in detail. It deserves mentioning that in this paper, we mainly focused on the bias issue of other estimators in Phase II, given that we usually use the control charts based on the estimators obtained in Phase-I to monitor the manufacturing process in Phase II. To further validate the theoretical results, we conducted Monte Carlo simulations to study the empirical biases and MSEs of the estimators under consideration, which are provided Figure 1, Tables 1 and 2. Thus, we do not use m samples in the Monte Carlo simulations since the biases and MSEs of the considered estimators do not depend on the value of m. We again greatly appreciate your constructive comments in this direction.

Thank you for your great comments and suggestions. From our humble experience, the performance of the control chart depends on the sample size and the Bernoulli probability. Indeed, our results in Figure 1 showed that when a sample size is small with a large Bernoulli probability, the existing estimators used in the construction of the control charts in well-known software (including Minitab and SPC-excel) can be disastrous in terms of the bias and MSE. This observation motivates us to point out the commonly used estimator for p (Benneyan, 2001) below

$$\hat{p}_{\mathrm{b}} = \frac{(N-1)/N}{\bar{\bar{X}} - a + 1},\tag{1}$$

is a biased estimator (See the highlighted part on page 8 of this reply for Minitab and on page 12 for SPC-excel). These software used this biased estimator in (1) maybe

because they believe that this is unbiased, whereas it is actually biased as we proved in this paper. To the best of knowledge, we are the first to point out that his formula is biased (of course, it is not MVU or MLE).

Also, our response to your comment 4) is related to this comment. For your convenience, we have

$$S_{\mathrm{b}} > S_{\mathrm{ml}} > S_{\mathrm{mvu}},$$

where

$$S_{\mathrm{b}} = \frac{N}{N-1}\sqrt{\frac{1}{n_k}\left(\bar{\bar{X}} - a + \frac{1}{N}\right)(\bar{\bar{X}} - a + 1)},$$

$$S_{\mathrm{ml}} = \sqrt{\frac{(\bar{\bar{X}} - a)(\bar{\bar{X}} - a + 1)}{n_k}},$$

$$S_{\mathrm{mvu}} = \sqrt{\frac{N}{N+1}\frac{(\bar{\bar{X}} - a)(\bar{\bar{X}} - a + 1)}{n_k}},$$

which shows that the ARL1 of the proposed method can be the smallest.

Of course, we fully agree with your suggestion for investigating if the MVU estimator could present a smaller ARL1 once the ARL0 is fixed. Thus, we have added the suggestion into Section 6 as a promising future research point. In ongoing work, we plan to conduct extensive simulation studies to compare the performance of control charts based on the MVU and the one in (1) used in the Minitab and SPC-excel. We here refer the reviewer to the red parts in Section 6 (Concluding remarks) in detail.

> **Review**
>
> *3) The authors state "It deserves mentioning that these conventional g and h charts are developed based on the underlying assumption that samples from the process should be balanced so that the samples have the same size, whereas such an assumption can be restrictive and may not be satisfied in many practical applications. To overcome this issue, we propose the method of how to construct the g and h charts with unbalanced samples." Please exemplify cases where such a situation occurs.*

Thank you so much for this great comment! We can think the situation where we can carry out experiments with enough time in Phase I so that we have a large sample size, but we have only a small sample size due to time limitation in Phase II. Thus, it is beneficial to provide our formula which can handle different sample sizes.

This issue can be related to the control chart based on the negative binomial distribution. For example, we can merge two samples into one to increase sample size

in Phase II. Then the underlying distribution is not geometric any more. It is the negative binomial distribution. In fact, we are currently working on this issue, As shown in Section 4, we mainly provide a method of constructing the g and h charts with unbalanced samples from a methodological perspective. One clear benefit of using the negative binomial distribution is that we can increase the statistical power and thus we can shorten ARL1. Frankly, we do not find a real-data to illustrate the occurrence of this unbalanced case under consideration. Accordingly to your valuable comment, we have revised such statements in the revision. Please refer to the red parts on page 2 in detail.

> **Review**
>
> *4) In Table 2, the MSE's presented for MVU are, in most cases, higher when $p < 0.9$. I suggest that the authors discuss this point in more detail, as it seems to be a serious disadvantage of the estimator and can harm the performance of a control chart;*

Thanks for your critical comments about the behavior of the MVU estimator with respect to the MSE. We fully agree with your comment since the proposed MVU method can have a disadvantage around with $p < 0.75$ in terms of the MSE. This phenomenon occurs for the MVU estimator becomes it comes at a tradeoff cost between unbiasedness and variance.

Let S_{b}, S_{ml}, and S_{mvu} be the standard deviation parts in the h charts. These are given by

$$S_{\mathrm{b}} = \frac{N}{N-1} \sqrt{\frac{1}{n_k} \left(\bar{\bar{X}} - a + \frac{1}{N} \right) (\bar{\bar{X}} - a + 1)},$$

$$S_{\mathrm{ml}} = \sqrt{\frac{(\bar{\bar{X}} - a)(\bar{\bar{X}} - a + 1)}{n_k}},$$

and

$$S_{\mathrm{mvu}} = \sqrt{\frac{N}{N+1} \frac{(\bar{\bar{X}} - a)(\bar{\bar{X}} - a + 1)}{n_k}}.$$

Then we can easily show that we have the following results regardless of the value of p,

$$S_{\mathrm{b}} > S_{\mathrm{ml}} > S_{\mathrm{mvu}}.$$

This result clearly show that the interval length of the charts based on Benneyan (2001) and ML are wider than the control chart based on the MVU. Thus, the performance of the proposed estimate does not degrade even with $p < 0.75$. We added

the above inequality relation at the end of Section 4 of the revised version. Please refer to the red parts on page 16 in detail. According to your valuable comments, in the revised version, we added two real-data examples for illustrative purposes. The results clearly show that the chart based on Benneyan (2001) is the widest, while the chart based on the proposed estimate is the narrowest.

However, your great insight can bring us a new topic about constructing the control chart based on the minimum MSE estimator, which is currently under investigation and will be reported elsewhere. We have added this research direction in the concluding remark. Please refer to the red part on page in detail.

> **Review**
>
> *5) As in Kaminsky et. al. (1992) – Section Development of Trial Control Charts: "Standards Not Given" – the authors could present an example of the use of the control chart using the MVU estimator. Additionally it could also show the use of the rQCC R package developed by the authors. I believe the paper ended very abruptly in section 4.*

Thank you so much for your comment. We already developed the rQCC package and will add this part to this package after this paper is finished. The current version can handle $\bar{X}$ chart, S chart and some other robust control charts. By using the URL provided in the reference, one can easily install and use the rQCC package. The `rQCC` R package can be easily installed as below:

```
install.packages("rQCC")  # Install from R server
library("rQCC")           # Add rQCC package in user's R
help(package="rQCC")      # Help files
```

References

Benneyan, J. C. (2001). Number-between g-type statistical quality control charts for monitoring adverse events. *Health Care Management Science 4*, 305–318.