Responses to Reviewer # 1

This paper presented robust methods for estimation of geometric distribution parameters and a g chart for statistical monitoring. The paper is well written and the derivations are sound. I have following suggestions:

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

---

1. All the case studies are simulation based. Some real dataset and application are needed.

   > **Response**
   >
   > Thank you for your valuable suggestion. In Section 4, we added one more example with real data set analyzed by Jacquez et al. (1996) and Benneyan (2001). Please refer to the red paragraphs on page 11 and Figure 2 on page 12 for details.

2. The influence of sample size on these estimation methods are needed.

   > **Response**
   >
   > Thank you for your great comments. We fully agree with your comments and added more simulation results with increased sample sizes in Section 5. Numerical results from simulations showed that as the sample size increases, the values of the bias, the variance and the MSE get smaller, whereas the RE values do not change much for both uncontaminated and contaminated cases. We here refer to the reviewer to the red paragraphs on page 18, Figures 6 and 7 on page 19, Table 8 on page 20, Table 9 on page 21, Figure 8 on page 22 for details. In addition, we have added some remarks (red parts) in the Section of concluding remarks.

3. What if we first use some outlier detection technique to remove outliers and then use the conventional ML method for the estimation? Comparing with the proposed robust method, what if we use such a two step procedure?

   > **Response**
   >
   > Thank you for your constructive comment. It is know that this is a long-standing issue about data screening. The references below can be an answer to this issue.
   >
   > https://www.itl.nist.gov/div898/handbook/eda/section3/eda35h.htm
   >
   > https://en.wikipedia.org/wiki/Robust_statistics
   >
   > Traditionally, statisticians would manually screen data for outliers, and remove them, usually checking the source of the data to see whether the outliers were erroneously recorded. For the case of the Newcomb's speed-of-light example in Table 5 of Stigler (1977), it is easy to see and detect the outliers prior to proceeding with any further analysis. However, in modern times, real-data sets often consist of large numbers of variables

being measured on large numbers of experimental units. Therefore, manual screening for outliers may be impractical. Thus, it seems beneficial to provide practitioners with some robust estimators for the in-control parameter of the control charts.

Responses to Reviewer # 2

The author(s) proposed two robust estimators for the geometric distribution and developed g-type quality control charts to monitor the number of conforming cases between the two consecutive appearances of nonconformities. The manuscript is well-written and smooth to read. The methods are interesting and of good value. However, there are two big concerns that need to be addressed.

> **Response**
>
> Thank you so much for your constructive and positive comments. According to your valuable comments, we have carefully addressed your comments point by point in the revised manuscript as below.

---

1. The author(s) proposed general estimators that work for any location shift a (as stated in line 17, page 4). Yet only the regular cases that $a = 1$ and $a = 0$ are considered in Section 4 and Section 5 respectively. It is suggested to consider values except for 0 and 1 to validate the effectiveness and generality of the proposed methods.

   > **Response**
   >
   > Thank you for your valuable comment. In this paper, we mainly focus on the regular cases that $a = 1$ and $a = 0$, since to our knowledge, the current existing control charts for the geometric distribution are based only on two regular cases: a = 1 (for example, Quesenberry 1995) and a = 0 (for example, Yang et al. 2002). The main objective of this paper is develop two robust estimators for the process parameter and then construct robust g-type quality control charts to monitor the number of occurrences of a given event in production processes, so we did not pursue such study in this direction. However, we fully agree with your comment that there are several cases of using different values of a. Thus, in the revision, we briefly discuss theses cases for the sake of completeness. We here refer the reviewer to the red parts on page 5 for details. Again, thank you for your great comment.

2. The application to real data is very important to show the effectiveness of the proposed methods. Although it is stated in the abstract that "a real-data application" is illustrated, data of the illustrative example in Section 4 are actually artificially simulated data instead of from a real case. Those data are "generated from the geometric distribution with $p = 0.2$ and $a = 1$" as stated in page 7 and 8. The author(s) are suggested to apply the proposed methods to a "real" case.

   > **Response**
   >
   > Thanks for your great suggestion about a real-data application, which is raised by another reviewer. Thus, in Section 4, we added one more example with real data set analyzed by Jacquez et al. (1996) and Benneyan (2001) for illustrative purposes. Please refer to the red paragraphs on page 11 and Figure 2 on page 12 for details.
   >
   > It deserves mentioning that in the revision, we still keep the analysis of the data set (the first data in Section 4 of the revision) provided by Kaminsky et al. (1992), since this data was firstly studied by these authors who originally developed the Shewhart-type

statistical control chart based on the geometric distribution.

Responses to Reviewer # 3

This manuscript develops two estimators for the process parameters of a g-type quality control chart. First of all, there is a lack of in-depth discussion in Literature Review Section. The authors failed to summarize the main gaps between state-of-the-art and engineering expectations. In addition, the proposed analysis is based on some simple manipulation of geometric probability equations. The work is considered as an simply application of some technique to an engineering problem.

Response

Thank you for your constructive comments. According to your valuable comments, we have significantly rewritten the Introduction section in the revision to provide a complete recent literature review and summarize some gaps between state-of-the-art and engineering expectations. We here refer the reviewer to the red parts on pages 2 and 3 for details.

It is worth noting that there exist good robust methods for estimating *location* and *scale* estimators under *symmetric* distributions. However, to the best of our knowledge, we are unable to find an appropriate robust estimator (especially Bernoulli probability parameter) when a distribution is not symmetric. Therefore, it seems beneficial to provide practitioners with some robust estimators for the in-control parameter and to study the effect of parameter estimation on the performance of the resulting control charts. This observation motivated us to develop two robust estimators for the process parameter and then construct robust g-type quality control charts to monitor the number of occurrences of a given event in production processes, given that accurate parameter estimation is usually difficult and may require a larger sample size than the one available in practical engineering applications.

To be more specific, we first developed a new method for estimating the unknown parameter based on the memoryless property of the geometric distribution. We acknowledge that the second method can be considered as some simple manipulation of the existing method. However, it deserves mentioning that we carried our extensive simulation studies and real-data applications to investigate the performance of these methods under considerations. Numerical results showed somewhat surprising results. For example, the conventional g chart is extremely vulnerable to data contamination (even with a single outlier), so that its performance degrades seriously. To further investigate this phenomenon, we used the relative efficiency and influence ideas, etc. from our humble viewpoint, we believe that the proposed methods are simple but very effective and useful for field engineers and practitioners. In addition, we are planning to add the proposed methods in the existing `rQCC` R package (Park and Wang, 2020) can be easily and conveniently implemented by field engineers and practitioners to monitor the state of the process.

References

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

Jacquez, G. M., Waller, L. A., Grimson, R., and Wartenberg, D. (1996). The analysis of disease clusters, Part I: State of the art. *Infection Control and Hospital Epidemiology*, 17(5):319–327.

Kaminsky, F. C., Benneyan, J. C., and Davis, R. D. (1992). Statistical control charts based on a geometric distribution. *Journal of Quality Technology*, 24:63–69.

Park, C. and Wang, M. (2020). `rQCC`: Robust quality control chart. `https://CRAN.R-project.org/package=rQCC`. R package version 1.20.7 (published on July 5, 2020).

Quesenberry, C. P. (1995). Geometric q charts for high quality processes. *Journal of Quality Technology*, 27(4):304–315.

Stigler, S. M. (1977). Do robust estimators work with real data? *Annals of Statistics*, 5:1055–1098.

Yang, Z., Xie, M., Kuralmani, V., and Tsui, K.-L. (2002). On the performance of geometric charts with estimated control limits. *Journal of Quality Technology*, 34:448–458.