Privacy Preserving Inference on the Ratio of Two Gaussians Using (Weighted) Sums

Jingang Miao, Yiming Paul Li
Facebook

October 28, 2021

Abstract

The ratio of two Gaussians is useful in many contexts of statistical inference. We discuss statistically valid inference of the ratio estimator under Differential Privacy (DP). We use the delta method to derive the asymptotic distribution of the ratio estimator and use the Gaussian mechanism to provide (ϵ, δ) privacy guarantees. Like many statistics, the quantities needed here can be re-written as functions of sums, and sums are easy to work with for many reasons. In the DP case, the sensitivity of a sum can be easily obtained. We focus on the coverage of 95% confidence intervals (CIs). Our simulations shows that the no correction method, which ignores the noise mechanism, gives CIs that are too narrow to provide proper coverage for small samples. We propose two methods to mitigate the under-coverage issue, one based on Monte Carlo simulations and the other based on analytical correction. We show that the CIs of our methods have the right coverage with proper privacy budget. In addition, our methods can handle weighted data, where the weights are fixed and bounded.

Some Key Words: Differential privacy; Ratio of two Gaussians; Delta method.

1 Introduction

Ratio estimation is useful in many contexts. In randomized experiments, one may be interested in the percent difference of the outcome metric between two experimental arms, which involves a ratio. Other examples include the ratio of regression coefficients (Hirschberg and Lye, 2007) and the therapeutic safety ratio (Dunlap and Silver, 1986).

The motivating example here is in the context of supervised machine learning, where a model is said to be calibrated if its average score is close to the average label. Equivalently, the ratio of the two should be close to 1. Further, one may choose to bucketize the scores into (usually 10) groups and check the calibration ratio within each bucket.

Differential Privacy (DP) has become one of the more popular formal definitions of privacy (Dwork et al., 2006b). DP can be achieved by adding noise to each unit (known as local DP (Kasiviswanathan et al., 2008)), or to intermediate/final summary statistics (known as central or global DP).

There is a relatively small literature on valid statistical inference under DP (Brawner and Honaker, 2018; Covington et al., 2021; D'Orazio et al., 2015; Du et al., 2020; Evans et al., 2019; Ferrando et al., 2020; Karwa and Vadhan, 2017; Movahedi et al., 2021). To the best of the authors' knowledge, there is no existing practical work on differentially private statistical inference on the ratio estimator of two Gaussians. This work is an attempt to fill this gap.

2 Definitions and methodology

We define the quantity of interest and the privacy semantics. We use n for sample size, y for the label, and s for the score. Both y and s are non-negative. Further, l_y, u_y, l_s, u_s are the lower and upper bounds on. We focus on the binary classification models, where the bounds on y and s are [0, 1].

When the data is weighted, we use l_w, u_w for the lower and upper bounds of w, the sample weights, which are assumed to be fixed (e.g., design weights). Also, we also assume that u_w is known, which is the case for example when the bounds are specified in the weight calibration step.

2.1 Calibration Ratio

Given a model, calibration ratio is simply $r = \mu_s/\mu_y$, where μ_s and μ_y are the true means of s and y. An estimator of r is $\hat{r} = \bar{s}/\bar{y}$. Note this estimator is statistically biased, but its bias is of order $1/n$ and vanishes quickly as sample size increases. What's more interesting is its variance.

A fact we will use is that $\bar{s}/\bar{y} = \sum s/\sum y$, where the ratio of sums is easier to work with for inference. We use the same $\bar{s}$ and $\bar{y}$ to denote the weighted means when data is weighted.

2.2 Differential Privacy

A randomized algorithm satisfies the requirement of **Differential Privacy** (DP) (Dwork et al., 2006b) if for every two neighboring datasets that differ on exactly one record, and for every possible output, the probabilities of the output is close up to a multiplicative factor of $e^\epsilon \approx 1 + \epsilon$ whether the randomized algorithm is applied on one dataset or the other. This is often called ϵ-DP or pure DP.

When we say that two **neighboring datasets** differ on exactly one record, we mean one of the dataset can be obtained by adding or removing one record from the other dataset. This definition of neighboring is known as add/remove-one, as opposed to the alternative definition based on changing one record.

Approximate DP Dwork et al. (2006a) relaxes the DP requirement by allowing for the violation of ϵ-DP with a (cryptographically) small probability δ. This is often called (ϵ, δ)-DP. Formally, a randomized algorithm $M : \mathcal{X}^n \to \mathcal{Y}$ is (ϵ, δ)-DP if for all neighboring datasets $X, X' \in \mathcal{X}^n$ and all outcomes $T \subseteq \mathcal{Y}$ we have $\Pr(M(X) \in T) \leq e^\epsilon \Pr(M(X') \in T) + \delta$.

We use a few important properties of DP algorithms (Dwork et al., 2014):

1. **Closure under composition**: the composition of K differential private mechanisms, where the kth mechanism is (ϵ_k, δ_k)-DP, for $1 \leq k \leq K$, is $(\sum_{k=1}^K \epsilon_k, \sum_{k=1}^K \delta_k)$-DP. This is known as basic composition, which we use in this paper. There are more advanced theorems that have tighter composition bounds than the basic one.

2. **Immune to post-processing**: If an algorithm is (ϵ, δ)-DP, then any post-processing

of its outputs (i.e., without going back and looking at the raw data) is still (ϵ, δ)-DP.

DP makes intuitive sense for robust predictive modeling or statistical inference (Dwork and Lei, 2009). The ultimate goal of a predictive model is to have accurate predictions out of sample, not in sample. Similarly, the ultimate goal of statistical inference is to generalize the conclusion beyond the sample at hand. As a result, a small change in the sample, or one observation in the DP case, should not change the model or the inference much.

DP provides strong privacy guarantee for the worst-case scenario, at the cost of utility degradation. The privacy guarantee holds no matter how the data is distributed and what type of attack happens, but the added noise makes the statistical inference less precise.

2.3 Inference

For inference, the point estimate of the ratio is simply the ratio of the two (weighted) means, which is biased but the bias goes away quickly as sample size increases. So, we instead focus on the confidence interval (CI), usually at the 95% confidence level. Due to the Central Limit Theorem, both the numerator and the denominator of $\hat{r}$ are means of independent and identically distributed variables and are thus asymptotic Gaussians. For a ratio of two Gaussians, the delta method shows that the asymptotic distribution of $\hat{r}$ is itself a Gaussian with variance

$$\mathrm{Var}(\hat{r}) = \frac{1}{\mu_{\bar{y}}^2}\sigma_{\bar{s}}^2 - 2\frac{\mu_{\bar{s}}}{\mu_{\bar{y}}^3}\sigma_{\bar{y}\bar{s}} + \frac{\mu_{\bar{s}}^2}{\mu_{\bar{y}}^4}\sigma_{\bar{y}}^2, \tag{1}$$

where $\mu_{\bar{s}}$ and $\mu_{\bar{y}}$ are the means of $\bar{s}$ and $\bar{y}$, $\sigma_{\bar{s}}^2$ and $\sigma_{\bar{y}}^2$ are their variances, and $\sigma_{\bar{y}\bar{s}}$ is their covariance. See Seltman for a derivation.

2.3.1 DP mechanism

In statistics, many quantities of interest can be written as functions of sums, a fact we make use of here. In particular, for the DP context, sums are attractive because their sensitivity can be easily calculated. It is straightforward to re-write the plug-in estimator of equation (1) in terms of sums, where x is a placeholder for either s or y:

$$\hat{\mu}_{\bar{x}} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \tag{2}$$

$$\hat{\sigma}^2_{\bar{x}} = \frac{\sum_{i=1}^n w_i^2}{(\sum_{i=1}^n w_i)^2} \left\{ \frac{\sum_{i=1}^n w_i x_i^2}{\sum_{i=1}^n w_i} - \left[\frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \right]^2 \right\} \tag{3}$$

$$\hat{\sigma}^2_{\bar{y}\bar{s}} = \frac{\sum_{i=1}^n w_i^2}{(\sum_{i=1}^n w_i)^2} \left\{ \frac{\sum_{i=1}^n w_i y_i s_i}{\sum_{i=1}^n w_i} - \frac{\sum_{i=1}^n w_i y_i \sum_{i=1}^n w_i s_i}{(\sum_{i=1}^n w_i)^2} \right\} \tag{4}$$

To be explicit, up to 7 sums are needed: $\sum_{i=1}^n w_i$, $\sum_{i=1}^n w_i y_i$, $\sum_{i=1}^n w_i s_i$, $\sum_{i=1}^n w_i^2$, $\sum_{i=1}^n w_i y_i^2$, $\sum_{i=1}^n w_i s_i^2$, and $\sum_{i=1}^n w_i y_i s_i$. However, for a binary classification model, y is either 0 or 1, so $\sum_{i=1}^n w_i y_i = \sum_{i=1}^n w_i y_i^2$, leading to only to 6 sums needed. Further, when the data is not weighted, aka $w_i = 1$ for all i, then $\sum_{i=1}^n w_i = \sum_{i=1}^n w_i^2$, leading to only 5 sums needed. Also, you may recognize the inverse of Kish's effective sample size $(\sum_{i=1}^n w_i)^2/(\sum_{i=1}^n w_i^2)$ (Kish, 1965) in equations (3) and (4). Without weights, they would become $1/n$. The effective sample size indicates the loss of efficiency due to weighting.

Recall that one reason we use the sums is that their sensitivity can be easily obtained. Under the add/remove-one definition of neighboring datasets, the sensitivity of each sum is simply the summand with s, y, and w replaced by their (positive) upper bounds. In the binary classification case, the bounds for s and y are [0, 1], so the sensitivity for all sums is simply u_w.

We use the Gaussian mechanism to achieve (ϵ, δ)-DP (Dwork et al., 2006a), which uses Gaussian noise with standard deviation [1]

$$\sigma = \frac{\Delta\sqrt{2\log(1.25/\delta)}}{\epsilon}. \tag{5}$$

For example, $\sum_{i=1}^n w_i y_i$ will be released as $(\sum_{i=1}^n w_i y_i)_{\text{dp}} = \sum_{i=1}^n w_i y_i + e$, where we use a subscript dp to indicate the noisy quantity that can be released. Here, e is the noise term coming from a Gaussian distribution $e \sim \text{Gaussian}(0, \sigma^2_{\sum_{i=1}^n w_i y_i})$, where σ is obtained by plugging $\Delta = u_w$ into (5) (upper bound $u_y = 1$ for the binary case). Due to composition,

[1] Alternatively, we can use an improved method described in Balle and Wang (2018), so that smaller scale noise is used. The smaller variance of the noise has no closed-form expression and has to be solved numerically.

the global budget is split among quantities released. For example, if 6 sums are released, then each one would get ($\epsilon/6$, $\delta/6$). Tighter composition theorems can be used for large number of composition rounds, but here we use the basic composition for simplicity.

2.4 CI calculation

Now the DP version of the up to 7 sums are released, all calculation based on them are post-processing, therefore the privacy guarantees remain the same. The point estimate is simply $\hat{r} = \frac{\left(\sum_{i=1}^{n} w_i s_i\right)_{\text{dp}}}{\left(\sum_{i=1}^{n} w_i y_i\right)_{\text{dp}}}$. What's more interesting is the standard error. Instead of ignoring the DP mechanism, we propose two methods that appropriately account for it in the CI calculation.

2.4.1 No correction

One option is to ignore the DP noise added to the sums without applying any correction. To be explicit, we just plug in the DP version of the sums into equations 2 to 4 to get the mean and variance/covariance estimates and then plus those into 1 to get the final variance estimate. We call the variance obtained this way $\sigma^2_{\text{no\_correction}}$, which ignore some uncertainty and the resulting CIs are expected to be too narrow in finite sample settings.

2.4.2 Monte Carlo

How much additional variance is injected by the DP mechanism to the ratio estimate? We can estimate that via Monte Carlo simulations. Note that the ratio of means is the same as the ratio of sums, which we'll use here for convenience. The procedure is straightforward, where we

1. calculate point estimate $\hat{r} = \frac{\left(\sum_{i=1}^{n} w_i s_i\right)_{\text{dp}}}{\left(\sum_{i=1}^{n} w_i y_i\right)_{\text{dp}}}$.
2. for $b = 1, ..., B$, where B is a large integer:
   (a) generate independent Gaussian noises $e_{s,b}$ for $\sum_{i=1}^{n} w_i s_i$ and $e_{y,b}$ for $\sum_{i=1}^{n} w_i y_i$ from distributions with the same variances as in the original DP process, again according to Equation (5).

(b) calculate $\hat{r}_b = \frac{\left(\sum_{i=1}^n w_i s_i\right)_{\text{dp}} + e_{s,b}}{\left(\sum_{i=1}^n w_i y_i\right)_{\text{dp}} + e_{y,b}}$

3. the extra variance due to DP is then $\sigma^2_{\text{extra}} = 1/B \sum_{b=1}^B (\hat{r}_b - \hat{r})^2$

4. the final variance is then $\sigma^2_{\text{sim}} = \sigma^2_{\text{no\_correction}} + \sigma^2_{\text{extra}}$

Note that we are not looking at the raw data beyond the released sums and thus not consuming additional privacy budget due to the post-processing property of DP. The Monte Carlo is easy to implement. In addition, the computation is fairly cheap since it can be vectorized.

2.4.3 Analytical correction

Recall that the variance of $\hat{r}$ depends on the means and variance/covariance of $\bar{s}$ and $\bar{y}$. Here, for convenience we again use the ratio of sums instead of means. To get the corresponding terms for the sums version, we multiply the right hand side of equation (2) by $\sum_{i=1}^n w_i$ and the right hand side of equations (3) and (4) by $(\sum_{i=1}^n w_i)^2$.

How do the Gaussian noises added do to $\sum_{i=1}^n w_i s_i$ and $\sum_{i=1}^n w_i y_i$ change their variance? It's actually simple. The noise is coming from independent Gaussians, so the variance of the sum is simply the sum of the variance. Further, the independent noises do not change the covariance term. As a result, all we need is to add the variance of the noise term to the variance of $\sum_{i=1}^n w_i s_i$ and $\sum_{i=1}^n w_i y_i$ before plugging into equation (1).

Once the point estimates and CIs are obtained via any of the three methods above for two groups, hypothesis testing of the equality of the two ratios can be easily carried out since $\hat{r}_1 - \hat{r}_2 \xrightarrow{d} \text{Normal}(r_1 - r_2, \sigma^2_{r_1} + \sigma^2_{r_2})$.

3 Simulations

With a sample size of $5,000$ or $10,000$, we simulated $s \sim \text{Beta}(2, 2)$, $y \sim \text{Bernoulli}(s/1.1)$ (so that true calibration ratio was 1.1), and w as Exponential(1) clipped to the range of [0.2, 5.0]. Values of ϵ used included $\{0.5, 1.0, 4.0\}$, δ = 1e-6, and both weighted and unweighted data were analyzed.

For each simulated dataset, we generated the 95% Wald confidence intervals, calculated the width of the intervals, and checked whether each covered the true calibration ratio for the following methods

- Public: the public method without DP
- No_correction: the method without correction for DP
- Monte Carlo: the Monte Carlo simulation based method
- Analytical correction: the correction based on modified variance terms

We also calculated the effective sample size, which gave us a rough idea of how variable the weights are, using the Kish formula $(\sum_{i=1}^{n} w_i)^2)/(\sum_{i=1}^{n} w_i^2)$ (Kish, 1965). Recall that the inverse of Kish's effective sample size appeared in equations (3) and (4). We repeated the simulation $1,000$ times. The python code for the simulation can be found at https://github.com/miaojingang/private_ratio.

4 Results

	Public		No Correction		Monte Carlo		Analytical	
ϵ	width	coverage	width	coverage	width	coverage	width	coverage
No weights; $n = 5,000$, effective $n = 5,000$								
0.5	0.061	0.947	0.061	0.575	0.156	0.949	0.156	0.949
1.0	0.061	0.947	0.061	0.784	0.094	0.947	0.094	0.947
4.0	0.061	0.947	0.061	0.936	0.064	0.948	0.064	0.948
With weights; $n = 5,000$, effective $n = 2,622$								
0.5	0.084	0.947	0.086	0.142	0.989	0.949	0.903	0.946
1.0	0.084	0.947	0.083	0.312	0.447	0.947	0.441	0.947
4.0	0.084	0.947	0.084	0.776	0.136	0.946	0.136	0.949
No weights; $n = 10,000$, effective $n = 10,000$								
0.5	0.043	0.955	0.043	0.665	0.083	0.934	0.084	0.935
1.0	0.043	0.955	0.043	0.874	0.056	0.951	0.056	0.947
4.0	0.043	0.955	0.043	0.951	0.044	0.955	0.044	0.955
With weights; $n = 10,000$, effective $n = 5,323$								
0.5	0.059	0.946	0.058	0.178	0.442	0.932	0.435	0.925
1.0	0.059	0.946	0.059	0.369	0.222	0.924	0.222	0.922
4.0	0.059	0.946	0.059	0.836	0.080	0.947	0.080	0.948

Table 1: Average width and coverage of 95% confidence intervals. Public: no noise added and thus Non-DP; the average width and coverage does not change as a function of ϵ. No correction: ignoring the face that DP noise was added. Monte Carlo: correction via Monte Carlo simulation. Analytical: correction via modified variance terms.

The results were summarized in Table (1). The public version, as expected, had coverages fairly close to the nominal level of 95%. Also, we were able to verify that the estimated variance agreed with the sampling variance.

The no correction method under-covered, and the width of its CIs were practically the same as the public method. This is because the no correction method did not account for the extra variability introduced by the DP mechanism. As a result, the CIs were too narrow, especially for cases with small sample sizes and/or small privacy budget and/or with

weighted sample. For example, on the weighted data with $n = 10,000$, $\epsilon = 0.5$, its CIs only covered the true value 17.8% of the time, which is grossly lower than the nominal coverage level.

Both correction methods gave the right coverage. For a fixed sample size and weighting scenario, as ϵ got smaller, more noise was injected by the DP mechanism, and both correction methods correctly accounted for that by giving wider CIs, which had the right coverage. With a large sample size and a big privacy budget, the DP CIs were only slightly wider than the public ones; for example, with $n = 10,000, \epsilon = 4.0$ and no weights, both correction methods had an mean CI width of 0.044, which is barely large than the public method's 0.043. On the other hand, the increase in CI width was more pronounced for smaller sample sizes, smaller privacy budgets, and weighted data.

5 Concluding remarks and future work

We explored the ratio estimation problem and proposed a DP mechanism on summary statistics and two inference methods that gave the valid CIs under DP. The proposal has a few nice features:

1. Simple. Calculation of the sums themselves is simple, calculation of their sensitivity is simple, and the correction needed to get valid CIs are again simple.

2. Flexible. Suppose the data has a hierarchical structure. For example, if the inference is done at the state level and later on one wants to aggregate to national level. The sums can be trivially added up.

3. Extensible. It can be extended for inference on other quantities. Sums are the building blocks of many statistics, including the moments and in turn some more complex quantities that depend on the moments. DP mechanisms based noising sums can be applied to other statistics.

This work represents an early effort on ratio estimation under DP. We hope future research will lead to better methods. We list a few possible optimizations and potential future directions here:

1. smaller Gaussian noise variance using analytical Gaussian Mechanism by Balle and Wang (2018)
2. more accurate DP mechanisms such as truncated Laplace for ϵ, δ-DP (Geng et al., 2020).
3. smarter budget allocation
4. advanced composition theorems, potentially leverage other relaxed DP definitions such as Concentrated DP (Bun and Steinke, 2016) and Gaussian DP (Dong et al., 2019)
5. Fieller's interval (Sherman et al., 2011) instead of the delta method for small samples
6. more conservative approaches to account for the DP noise uncertainty are available (e.g., Karwa and Vadhan (2017)), which will result in wider CIs and likely over-cover.
7. releasing fewer intermediary quantities, or conducting DP hypothesis testing of ratios without releasing both ratios. Intuitively, releasing fewer quantities might be more efficient both in terms of privacy budget usage and in terms of less noise needed. However, it takes work to derive the correct sensitivity of more complex quantities.
8. generic Monte Carlo based simulations to directly measure the combined uncertainty from sampling and DP (Du et al. (2020); Ferrando et al. (2020)).
9. other generic methods that work for many statistics. One example is bootstrapping (Brawner and Honaker, 2018).
10. methods that work with more generic survey weights that are not necessarily fixed. One example is calibration weights (Deville and Särndal, 1992) that depend on the sample at hand.

Instead of r, one may consider using $\log(r)$, which we briefly discuss in the Appendix.

Acknowledgments

The authors are grateful to Imanol Arrieta Ibarra, Ilya Mironov, Jonathan Tannen, Brian Karrer, and James Honaker for many discussions and for reviewing the manuscript.

References

Borja Balle and Yu-Xiang Wang. Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In *International Conference on Machine Learning*, pages 394–403. PMLR, 2018.

Thomas Brawner and James Honaker. Bootstrap inference and differential privacy: Standard errors for free. *Unpublished Manuscript*, 2018.

Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. *Lecture Notes in Computer Science*, 9985 LNCS:635–658, 2016. ISSN 16113349. doi: 10.1007/978-3-662-53641-4_24.

Christian Covington, Xi He, James Honaker, and Gautam Kamath. Unbiased statistical estimation and valid confidence intervals under differential privacy. *Presentation to Joint Statistical Meetings of the American Statistical Association*, 2021.

Jean-Claude Deville and Carl-Erik Särndal. Calibration estimators in survey sampling. *Journal of the American statistical Association*, 87(418):376–382, 1992.

Jinshuo Dong, Aaron Roth, and Weijie J Su. Gaussian Differential Privacy. Technical report, 2019.

Vito D'Orazio, James Honaker, and Gary King. Differential privacy for social science inference. *Sloan Foundation Economics Research Paper*, 2015.

Wenxin Du, Canyon Foot, Monica Moniot, Andrew Bray, and Adam Groce. Differentially private confidence intervals. *arXiv preprint arXiv:2001.02285*, 2020.

William P Dunlap and N Clayton Silver. Confidence intervals and standard errors for ratios of normal variables. *Behavior Research Methods, Instruments, & Computers*, 18(5):469–471, 1986.

Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proceedings of the Annual ACM Symposium on Theory of Computing*, pages 371–380, 2009. ISBN 978-1-60558-506-2. doi: 10.1145/1536414.1536466.

Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 486–503. Springer, 2006a.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006b.

Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.

Georgina Evans, Gary King, Margaret Schwenzfeier, and Abhradeep Thakurta. Statistically valid inferences from privacy protected data, 2019.

Cecilia Ferrando, Shufan Wang, and Daniel Sheldon. General-purpose differentially-private confidence intervals. *arXiv preprint arXiv:2006.07749*, 2020.

Quan Geng, Wei Ding, Ruiqi Guo, and Sanjiv Kumar. Tight analysis of privacy and utility tradeoff in approximate differential privacy. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 89–99. PMLR, August 2020.

JG Hirschberg and Jenny N Lye. Providing intuition to the fieller method with two geometric representations using stata and eviews. *https://minerva-access.unimelb.edu.au/handle/11343/34613*, 2007.

Vishesh Karwa and Salil Vadhan. Finite sample differentially private confidence intervals. *arXiv preprint arXiv:1711.03908*, 2017.

Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What Can We Learn Privately? In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 531–540, October 2008. doi: 10.1109/FOCS.2008.27.

Leslie Kish. *Survey sampling*. Number 04; HN29, K5. in Wiley Classics Library. New York : John Wiley & Sons, 1965.

Mahnush Movahedi, Benjamin M. Case, Andrew Knox, James Honaker, Li Li, Yiming Paul Li, Sanjay Saravanan, Shubho Sengupta, and Erik Taubeneck. Privacy-Preserving Randomized Controlled Trials: A Protocol for Industry Scale Deployment. In *Proceedings of the 2021 Cloud Computing Security Workshop (CCSW '21)*, Republic of Korea, November 2021. ACM.

Howard Seltman. Approximations for mean and variance of a ratio. https://www.stat.cmu.edu/~hseltman/files/ratio.pdf, Accessed: 10-30-2020.

Michael Sherman, Arnab Maity, and Suojin Wang. Inferences for the ratio: Fieller's interval, log ratio, and large sample based confidence intervals. *AStA Advances in Statistical Analysis*, 95(3):313–323, 2011.

Appendix

5.1 Consider $\log(r)$

As a result of the fact that both y and s are non-negative, the distribution of r is skewed. However, the CIs constructed using the method presented here is symmetric. There are two remedies. First, $\log(r)$ might be a better quantity to use. Second, if the quantity of interest is indeed r, one can construct CI of $\log(r)$ and exponentiate the limits to get back a CI of

r. For both remedies, the asymptotic variance of $\log(\hat{r})$ is needed. The good news is that it can be constructed again using the delta method:

$$\text{Var}(\log(\hat{r})) = \frac{1}{\mu_{\bar{s}}^2}\sigma_{\bar{s}}^2 - 2\frac{1}{\mu_{\bar{s}}\mu_{\bar{y}}}\sigma_{\bar{y}\bar{s}} + \frac{1}{\mu_{\bar{y}}^2}\sigma_{\bar{y}}^2.$$

The even better news is that it depends on the same quantities as equation 1, which means what we've proposed for r can be trivially adapted for the log version.