PLOS ONE

RESEARCH ARTICLE

# Statistical methods for comparing two independent exponential-gamma means with application to single cell protein data

**Jia Wang**[1], **Lili Tian**[1], **Li Yan**[2]*

**1** Department of Biostatistics, University at Buffalo, Buffalo, NY, United States of America, **2** Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, NY, United States of America

* Li.Yan@RoswellPark.org



OPEN ACCESS

**Citation:** Wang J, Tian L, Yan L (2024) Statistical methods for comparing two independent exponential-gamma means with application to single cell protein data. PLoS ONE 19(12): e0314705. https://doi.org/10.1371/journal.pone.0314705

**Editor:** Umair Khalil, Abdul Wali Khan University Mardan, PAKISTAN

**Received:** May 15, 2024

**Accepted:** November 14, 2024

**Published:** December 13, 2024



**Data Availability Statement:** All relevant data for this study are within the paper and its Supporting information files.

**Funding:** This work was partially supported by National Cancer Institute (NCI) grant P30CA016056 involving the use of Roswell Park Comprehensive Cancer Center's Biostatistics and Statistical Genomics Shared Resource. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

In genomic study, log transformation is a common prepossessing step to adjust for skewness in data. This standard approach often assumes that log-transformed data is normally distributed, and two sample t-test (or its modifications) is used for detecting differences between two experimental conditions. However, recently it was shown that two sample t-test can lead to exaggerated false positives, and the Wilcoxon-Mann-Whitney (WMW) test was proposed as an alternative for studies with larger sample sizes. In addition, studies have demonstrated that the specific distribution used in modeling genomic data has profound impact on the interpretation and validity of results. The aim of this paper is three-fold: 1) to present the Exp-gamma distribution (exponential-gamma distribution stands for log-transformed gamma distribution) as a proper biological and statistical model for the analysis of log-transformed protein abundance data from single-cell experiments; 2) to demonstrate the inappropriateness of two sample t-test and the WMW test in analyzing log-transformed protein abundance data; 3) to propose and evaluate statistical inference methods for hypothesis testing and confidence interval estimation when comparing two independent samples under the Exp-gamma distributions. The proposed methods are applied to analyze protein abundance data from a single-cell dataset.

## 1 Introduction

Recent investigations of physical models [1–4] of individual cells have demonstrated that the protein copy number (or abundance) distribution can be approximated by a gamma distribution. These studies claimed that the shape parameter of the gamma distribution can be interpreted as the number of mRNA produced per cell cycle, and the scale parameter as the protein molecules produced per mRNA within individual cells. Although studies at single-cell level were costly and scarce a decade ago, recent technology advances make large scale of protein abundance data at single-cell level proliferate [5, 6].

In practice, up-regulated and down-regulated genes between samples are assessed using fold change which represents a proportional rather than additive changes from a reference

**Competing interests:** The authors have declared that no competing interests exist.

(e.g. healthy) to an alternative (e.g. tumor) state. Hence log-transformed abundance level is more biologically relevant, and expression (or concentration) of genes is usually pre-processed using log-transformation before statistical modeling. Additionally, log-transformation is used to adjust for skewness and for variance stabilization [7–11]. Such transformation is widely used as a preprocessing step for many types of molecular markers. Therefore, the exponential-gamma distribution (Exp-gamma), derived from the gamma distribution by applying a logarithmic transformation, is an ideal candidate for modeling log-transformed protein abundance data.

However, researchers often resort to two sample t-test or the Wilcoxon-Mann-Whitney (WMW) test in differential analysis of log-transformed protein abundance and other molecular data [12–14] to detect the difference between two experimental conditions, often with some pre- and post-model adjustment to reduce the false positive rate [15, 16]. Fay and Proschan [17] argued that two sample t-test decision rules are asymptotically valid under quite general conditions even if the normality assumption is rejected. Recently, Li et al. [14] pointed out that two sample t-test often results in exaggerated false positive rate, and recommended using the WMW test for comparing two sets of expression levels measured under two conditions for a gene in population-level RNA-seq studies with large sample sizes. Hao et al. [5] analyzed the differential abundance of cell types across experimental conditions using the WMW test after log-normalization of the protein data. However, some researchers pointed out [17, 18] that although the WMW test does not require parametric assumptions, it assumes that the two distributions are equal under the null hypothesis; hence it could result in inflated type I errors when testing the equality of means. Recently, Torrente et al. [19] studied the shape of gene expression and discovered that the gamma distribution was the predominant non-normal category of genes in both microarray and RNA-seq datasets.

Although there exist some research on the appropriateness of two sample t-test and the WMW test in the differential analysis of log-transformed protein abundance data [5, 12–14], there does not exist such an investigation under the Exp-gamma distribution. Furthermore, accurate statistical inference methods for comparing two Exp-gamma means are of particular interest since identifying differences in log transformed protein abundance data under two different experiment conditions is a fundamental research question in genomics study. Despite the existence of rich statistical research on gamma means [20–29], to our knowledge, there does not exist literature on inference of the Exp-gamma means. Therefore, the aim of this paper is three-fold: 1) to present the Exp-gamma distribution as a proper biological and statistical model for the analysis of log-transformed protein abundance data from single cell experiments; 2) to demonstrate the inappropriateness of using two sample t-test and the WMW test in analyzing log-transformed protein abundance data; 3) to propose and evaluate statistical inference methods for hypothesis testing and confidence interval estimation for comparing two independent samples under Exp-gamma distributions.

This paper is organized as follows. In Section 2, we provide some preliminary results on features of the Exp-gamma distribution, along with its characteristics. In Section 3, the motivation for this research is addressed by a more detailed description of the molecular process of protein production and its critical role in human traits and disease, as given in Section 3.1, followed by an investigation into the inappropriateness of two sample t-test and the WMW test for testing the equality of two Exp-gamma means in Section 3.2. In Section 4, methods for hypothesis testing for the equality of two independent Exp-gamma means and confidence interval estimation for mean difference are proposed. In Section 5, we present the simulation studies on the type I error control and power of the proposed tests, as well as the coverage probability of proposed confidence intervals. In Section 6, a subset of Seurat data used in

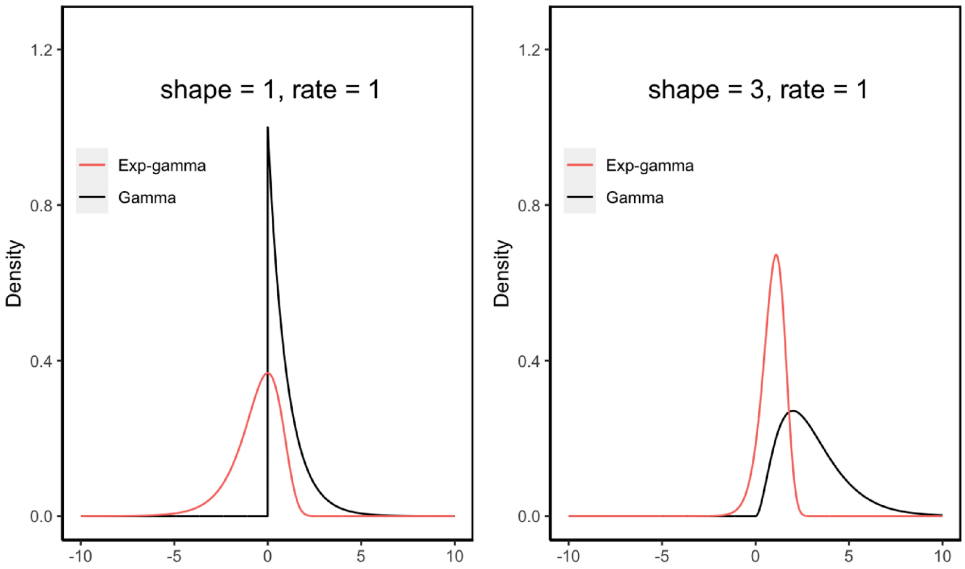


**Fig 1. Probability density of $Y \sim$ *Exp-gamma*$(\alpha, \beta)$ and $X = e^Y \sim$ *gamma*$(\alpha, \beta)$, for $(\alpha, \beta) = (1, 1)$ and $(3, 1)$, respectively.**

https://doi.org/10.1371/journal.pone.0314705.g001

scRNA-seq studies is analyzed using the proposed methods. Finally, concluding remarks are provided in Section 7.

## 2 The setting

Let $Y_1$ and $Y_2$ denote two independent random variables from log-transformed gene expression/protein abundance, where $Y_i \sim Exp\text{-}gamma(\alpha_i, \beta_i)$, i.e.

$$Y_i \sim f_i(y; \alpha_i, \beta_i) = \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} e^{\alpha_i y} e^{-\beta_i e^y}, \quad i = 1, 2,$$

where $y \in (-\infty, \infty)$, and $\alpha_i, \beta_i > 0$. Note that $X_i = e^{Y_i}$ following a gamma distribution, i.e. $X_i \sim gamma(\alpha_i, \beta_i)$ where $\alpha_i$ and $\beta_i$ stand for the shape parameter and rate parameter, respectively. Fig 1 contains two graphs of the probability density functions of $Y_i$ and $X_i$ at $(\alpha_i, \beta_i) = (1, 1)$ and $(\alpha_i, \beta_i) = (3, 1)$, respectively. The Exp-gamma distribution is skewed to the left (negatively skewed), with its both tails extending indefinitely.

Let $\delta_i$ and $\sigma_i^2$ denote the population mean and variance for $Y_i$, respectively. It can be proved that

$$\delta_i = \psi(\alpha_i) - \ln \beta_i, \tag{1}$$

$$\sigma_i^2 = \psi^{(1)}(\alpha_i), \tag{2}$$

for $i = 1, 2$, where $\psi()$ is the digamma function and $\psi^{(1)}()$ is the trigamma function. The details of the proof are presented in S1 Appendix.

Skewness and excess kurtosis are the other two measures which describe the distributional properties of a probability distribution. Skewness measures the asymmetry of the probability distribution, and excess kurtosis measures how much the distribution deviates from a normal distribution in terms of tails. Both the skewness (*skew*) and the excess kurtosis (*ex-kurt*) of

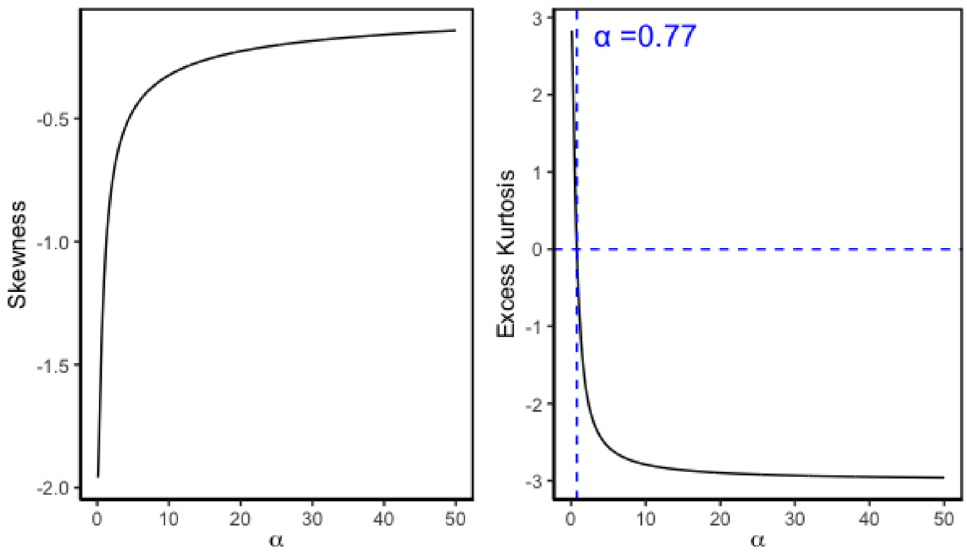


**Fig 2. Plot of skewness and excess kurtosis for Exp-gamma distribution as $\alpha$ ranges from 0.1 to 50. When $\alpha$ = 0.7689, the excess kurtosis is 0.**

https://doi.org/10.1371/journal.pone.0314705.g002

Exp-gamma distribution only depend on its shape parameter $\alpha_i$,

$$\begin{aligned} skew_i &= \psi^{(2)}(\alpha_i)/[\psi^{(1)}(\alpha_i)]^{3/2}, \\ ex\text{-}kurt_i &= \psi^{(3)}(\alpha_i)/[\psi^{(1)}(\alpha_i)]^2 - 3, \end{aligned} \tag{3}$$

for $i = 1, 2$, where $\psi^{(k-1)}()$ is $k$th derivative of the log gamma function. The detailed proof is shown in S1 Appendix.

Fig 2 shows the skewness and excess kurtosis of Exp-gamma distribution as the shape parameter ($\alpha$) ranges from 0.1 to 50. The negative skewness confirms the appearance of Exp-gamma distribution is left skewed. The excess kurtosis of Exp-gamma distribution can be positive and negative, whereas the positive value means that the Exp-gamma distribution is thin-tailed and has fewer outliers, and the negative value means that the Exp-gamma distribution is fat-tailed and has many outliers. When $\alpha = 0.7689$, the Exp-gamma distribution has the same kurtosis as the normal distribution. As $\alpha$ tends to infinity, the value of *skew* converges to 0, and the value of *ex-kurt* converges to −3.

We are interested in testing the hypothesis $H_0 : \delta_1 = \delta_2$, vs. $H_1 : \delta_1 \neq \delta_2$, as well as constructing confidence interval for mean difference $\delta_1 - \delta_2$. The mean difference of two independent Exp-gamma distributions is given by

$$\eta = \delta_1 - \delta_2 = \psi(\alpha_1) - \ln \beta_1 - (\psi(\alpha_2) - \ln \beta_2).$$

Let $\hat{\alpha}_i$ and $\hat{\beta}_i$ stand for the maximum likelihood estimates for $\alpha_i$ and $\beta_i$, respectively. The maximum likelihood estimator (MLE) of $\delta_i$ is

$$\hat{\delta}_i = \psi(\hat{\alpha}_i) - \ln(\hat{\beta}_i).$$

Then

$$\hat{\eta} = \hat{\delta}_1 - \hat{\delta}_2 = \psi(\hat{\alpha}_1) - \ln(\hat{\beta}_1) - (\psi(\hat{\alpha}_2) - \ln(\hat{\beta}_2)).$$

The variance of $\hat{\eta}$ is

$$Var(\hat{\eta}) = Var(\hat{\delta}_1 - \hat{\delta}_2) = \frac{\psi^{(1)}(\hat{\alpha}_1)}{n_1} + \frac{\psi^{(1)}(\hat{\alpha}_2)}{n_2}. \quad (4)$$

## 3 Motivation

In this section, we provide detailed arguments about the compelling importance of the exponential-gamma (Exp-gamma) distribution in analyzing log-transformed protein abundance data from single-cell experiments, as well as the paramount significance of developing statistical inference procedures under the Exp-gamma distribution.

### 3.1 Justification of using the Exp-gamma distribution for cellular protein abundance measurements

The central dogma of molecular biology is a fundamental theory developed by Francis Crick in 1958 that explains how genetic information flows within a biological system. The core idea can be simply stated as: "DNA makes (messenger) RNA, and RNA makes protein". The abundance of cellular protein is intimately linked to all biological functions in living cells. Since then, this theory has withstood the test of time and intensive investigations, with only minor exceptions and enrichment. The expression levels of messenger RNAs (mRNAs) and proteins are essential measurements of an organism's genetic makeup (genotypes), and are often directly related to many observable characteristics or traits (phenotypes), including morphology, development, biochemical, and physiological properties. Common phenotypes in humans include height and blood type, as well as disease related characteristics, e.g. cancer subtypes. Understanding the differences in genotypes (e.g. protein abundance) and their relationships with phenotypes (e.g. cancer progressions) is the focus of molecular biology.

Since its introduction in 2008 [30], cost effective and rapid mRNA quantification of whole genome (transcriptome) has become a standard tool in the life sciences research community. Initially developed for bulk samples, this method evolved to quantify mRNA levels in single cells, and revolutionized the field of cancer research. Numerous analysis methods and pipelines have been developed for mRNA quantification, based on the organism under study, platform characteristics, and researcher's goals [31]. Due to its wide-spread usage, the mRNA quantification is often used as a synonym for gene expression in many studies. However, in fact, the protein abundance data is a more accurate measurement for gene activity. It is well established that mRNA transcript level only partially correlates with protein abundances [32], and transcriptomics alone is often incapable of distinguishing between categories of cells that are molecularly similar, but functionally distinct. Due to the high cost and experimental complexities, studies that access protein abundance remain scarce, especially at single-cell level because of the low abundance of proteins in cells. Only in the past a few years, genome-wide analysis of protein abundance at single-cell level became practical [5]. Unfortunately, this belated development also means lack of investigation of protein specific statistical analysis method. Most of the methods that were adopted from RNA-seq analysis [7] overlooked sample distribution, except for some preprocessing and normalization steps to compensate for the obvious skewness of the protein data. It becomes clear that single-cell protein abundance specific statistical method for accurate assessment of such data is in great need. Consistent with the two-stage model of gene expression described in the central dogma of molecular biology, the intriguing physical models [1–4] unveiled intrinsic association between gamma distribution parameters and biological process of protein synthesis. Based on these observations, we

propose to use the Exp-gamma distribution for modeling single-cell protein levels in molecular biology and cancer research, since log-transformed protein abundances are often biologically more relevant to their cellular functions.

It is worth mentioning that many molecular biology data can be modeled by gamma distribution. For example, microRNA sequencing data often align closely with a gamma distribution due to the stochastic nature of exponential PCR amplification [19, 33, 34]. Additionally, the absolute abundance levels of metabolic [35] and microbiome [36] data exhibit characteristics that align with gamma distributions. Furthermore, log-transformation is a standard preprocessing step in the statistical analysis of these data. Hence, the Exp-gamma distribution is a good candidate for modeling log-transformed cellular protein abundance measurements.

## 3.2 Two sample t-test and the Wilcoxon-Mann-Whitney (WMW) test could be misleading

The two sample t-test and the WMW test are widely used in differential analysis for log-transformed protein abundance data in proteomics [5, 7, 14, 37–39]. However, the appropriateness of using these two tests in differential analysis under the Exp-gamma has not been investigated. Hence, in this section, we aim to use a simulation study to demonstrate their limitations in differential analysis.

Assume two samples of protein abundance obtained under different experimental conditions are from gamma distributions, i.e. $X_1 \sim gamma(\alpha_1, \beta_1)$, and $X_2 \sim gamma(\alpha_2, \beta_2)$. The differential analysis is based on log-transformed data from $Y_1$ and $Y_2$, where $Y_1 = \log(X_1) \sim Exp\text{-}gamma(\alpha_1, \beta_1)$ and $Y_2 = \log(X_2) \sim Exp\text{-}gamma(\alpha_2, \beta_2)$. We are interested in testing the equality of two Exp-gamma means.

We carried out simulations to evaluate the type I error control of two sample t-test and the WMW test under $H_0 : \delta_1 - \delta_2 = 0$. Four parameter settings for $(\alpha_1, \beta_1)$ vs. $(\alpha_2, \beta_2)$ are considered: A) (0.2, 0.005) vs. (5, 4.509); B) (0.5, 0.14) vs. (10, 9.504); C) (1, 0.561) vs. (5, 4.509); and D)(5, 0.048) vs. (5, 0.048). In settings A, B, and C, the two Exp-gamma distributions differ, while in setting D, they are identical. Fig 3 presents the density plots under these four settings. It can be seen that these settings vary considerably despite the fact that they are all under $H_0 : \delta_1 - \delta_2 = 0$. Under the null hypothesis of equal population means, the probability that $Y_1$ is greater than $Y_2$ (i.e. $P(Y_1 > Y_2)$), a measure for the difference between two populations, is 0.621, 0.593, 0.556, and 0.5, for settings A, B, C, and D, respectively, indicating setting A has the largest difference between two populations and setting D has the smallest difference. Note that generally speaking, $P(Y_1 > Y_2) = 0.5$ does not necessarily imply two populations are identical. In this simulation study, we deliberately design setting D to have two identical populations for the purpose of checking the applicability of two sample t-test and the WMW test under two identical Exp-gamma distributions. For each setting, we considered sample sizes from small (10) to large (75). For a given set of sample sizes and parameter configuration, 2000 observed datasets are generated. The simulated type I errors by two sample t-test and the WMW test are reported in Figs 4 and 5, respectively.

As shown in Fig 4, the type I errors of two sample t-test (or Welch's test for unequal variances) converge to nominal level as sample sizes increase, as guaranteed by the central limit theorem. In addition, when two Exp-gamma distributions are identical (setting D), the two sample t-test maintains controlled type I errors even when sample sizes are small. Note that the type I errors for setting D lie completely between two dashed lines in Fig 4, which indicate boundaries for satisfactory coverage given 2000 simulation runs. However, if two Exp-gamma distributions are different (i.e. settings A, B and C), the type I errors for testing the equality of means can be as high as 0.1, particularly when sample sizes are small (e.g. less than (50, 50) for

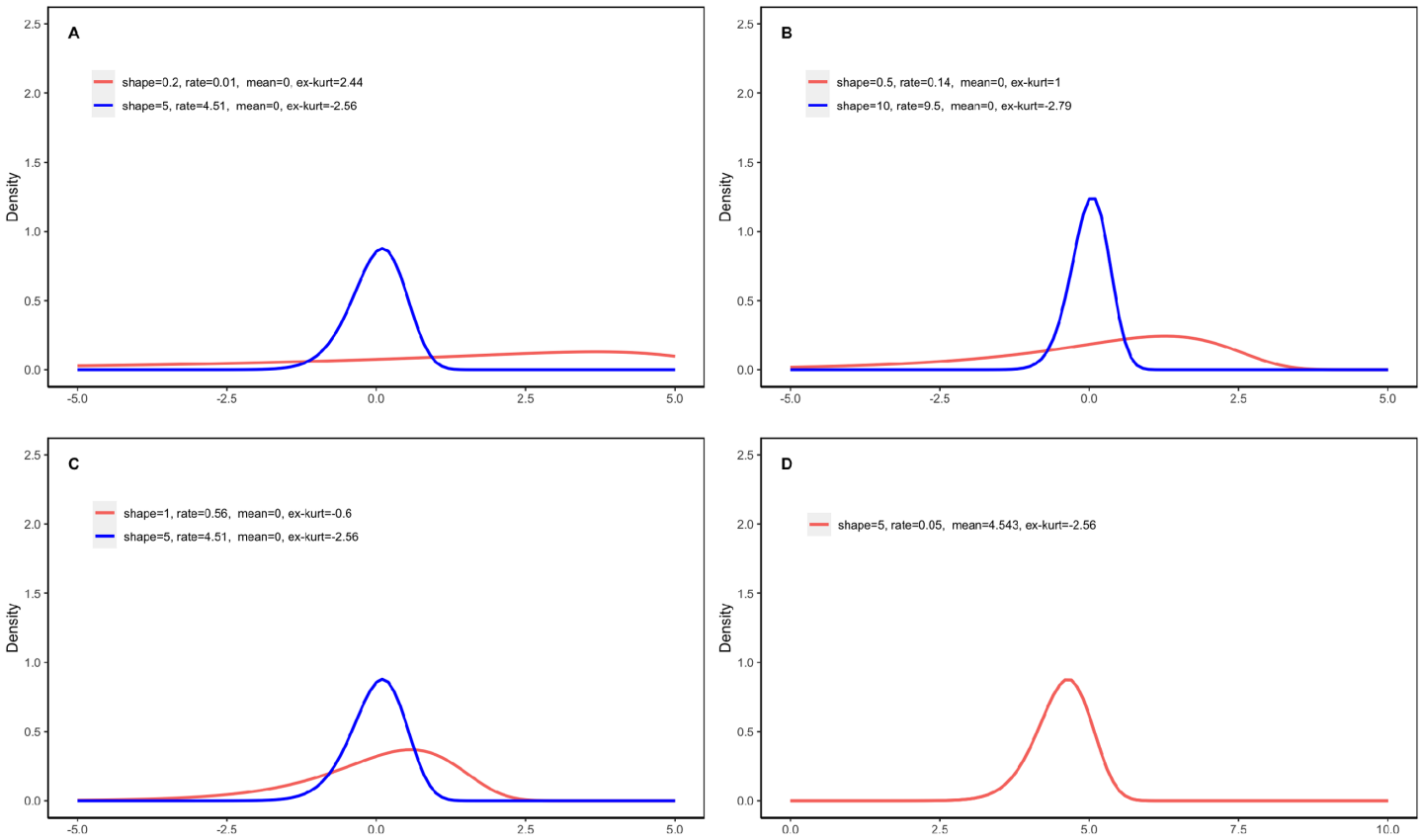


**Fig 3. Density plots of samples from four pair of comparisons $Y_1$ vs $Y_2$ where $Y_1 \sim$ *Exp-gamma*$(\alpha_1, \beta_1)$ vs. $Y_2 \sim$ *Exp-gamma*$(\alpha_2, \beta_2)$.** (A : (0.2, 0.005) vs. (5, 4.509); B: (0.5, 0.14) vs. (10, 9.504); C: (1, 0.561) vs. (5, 4.509); D: (5, 0.048) vs. (5, 0.048)).

https://doi.org/10.1371/journal.pone.0314705.g003

settings A and B, and less than (30, 30) for setting C). Thus, two sample t-test is appropriate for testing the equality of two means of log-transformed protein abundance data when sample sizes are larger than (50, 50). When dealing with small to medium sample sizes, we should exercise caution with two sample t-test, especially when two underlying distributions are very different.

When the assumption of normality is in doubt, it is a common practice that the WMW test is used as an alternative as it is a non-parametric test. However, while non-parametric tests such as the WMW test do not require normality, they test the null hypothesis that two populations are identical. Hence, when two populations have the same mean but not identical, the WMW test does not guarantee that the significance level will be preserved. More details can be found in the paper by Pratt [18] which thoroughly investigated the effect of differences between two populations on the level of the WMW test for normal, double exponential, and rectangular distributions. In this simulation study, we investigate the effect of the difference between two Exp-gamma distributions on the significance level of the WMW test under null hypothesis of equality of two Exp-gamma means. In Fig 5, we observe inflated type I errors for settings A, B, and C in the WMW test, and the magnitude of inflation increases as sample sizes grow. Furthermore, given sample sizes, as the disparity measured by $P(Y_1 > Y_2)$ increases, the inflation of type I error becomes worse; and setting A has the worst type I error control among all settings. It is also notable that the type I errors are well controlled for setting D in which two distributions are identical. Hence, for testing equality of two Exp-gamma means, the

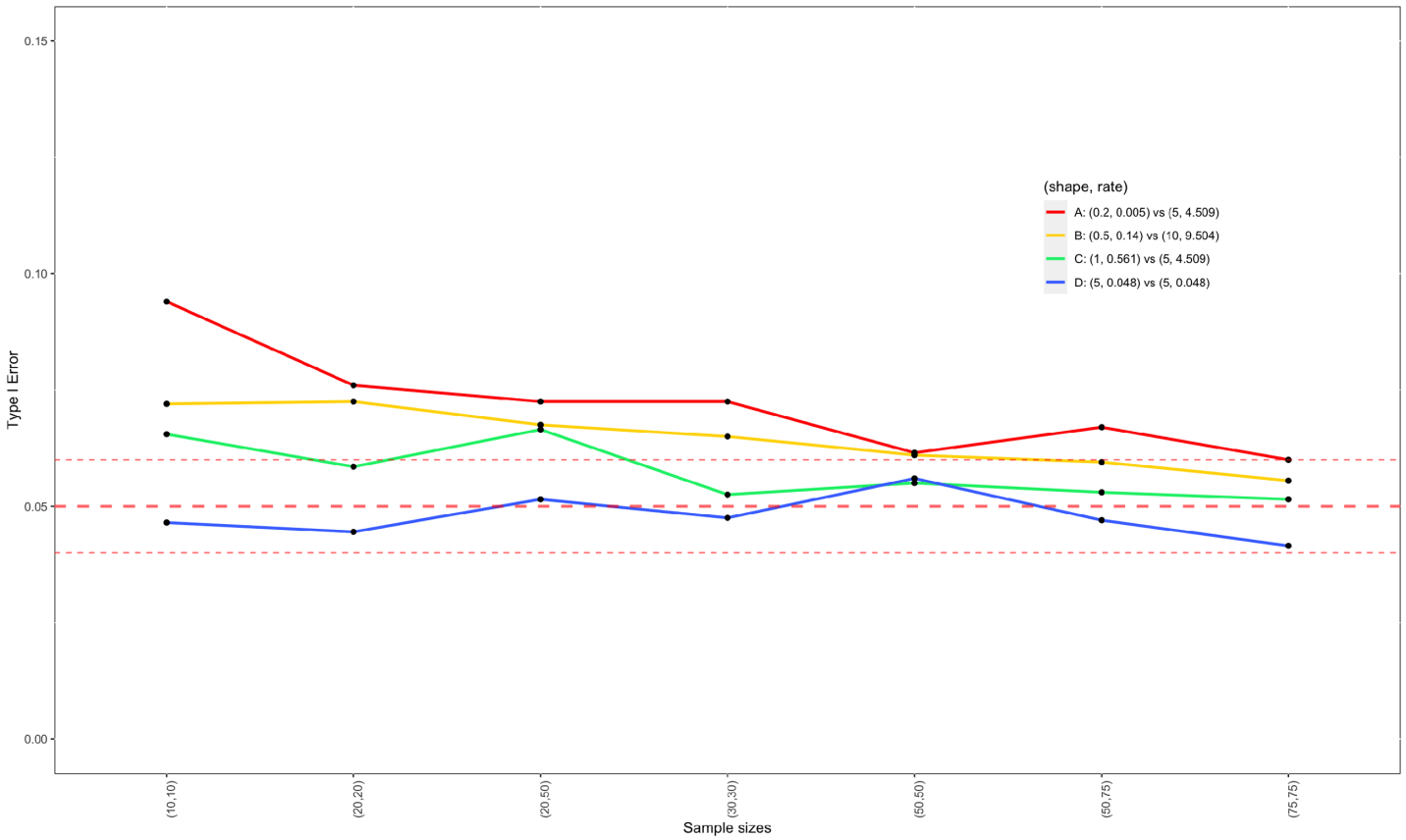


**Fig 4. Estimated type I errors of two sample t-test for testing the equality of mean of two Exp-gamma distributions as a function of sample sizes.** The middle dashed line represents the nominal significance level at $\alpha = 0.05$; and upper and lower dashed lines are upper and lower limits for satisfactory type I error rates, which are 0.06 and 0.04 with 2000 simulations runs, respectively.

https://doi.org/10.1371/journal.pone.0314705.g004

WMW test can control type I error only when two distributions are exactly the same, and the type I error can be severely out of control when two distributions are not the same.

In summary, both two sample t-test and the WMW test have limitations in testing of equality of two independent Exp-gamma means. While two sample t-test is not ideal when sample sizes are below medium, the limitations for the WMW test are more severe as it requires the two distributions to be exactly the same under the null hypothesis. In practice, small to medium sizes are common in genomics studies, and scenarios with identical populations under the null hypothesis could be rare. Therefore, accurate procedures for statistical inference for the mean difference of two independent Exp-gamma distributions are desirable.

## 4 Inferences on the mean difference of two independent Exp-gamma distributions

Let $Y_1$ and $Y_2$ be two independent Exp-gamma random variables, i.e. $Y_1 \sim Exp\text{-}gamma(\alpha_1, \beta_1)$ and $Y_2 \sim Exp\text{-}gamma(\alpha_2, \beta_2)$. Note that $Y_i = \ln X_i$ where $X_i \sim gamma(\alpha_i, \beta_i)$, $i = 1, 2$, and $X_1$ and $X_2$ are independent. Then the population means for $Y_1$ and $Y_2$ are given as follows:

$$\delta_1 = \psi(\alpha_1) - \ln \beta_1 \text{ and } \delta_2 = \psi(\alpha_2) - \ln \beta_2.$$

Thus, the research interest is to perform hypothesis testing with satisfactory type I error

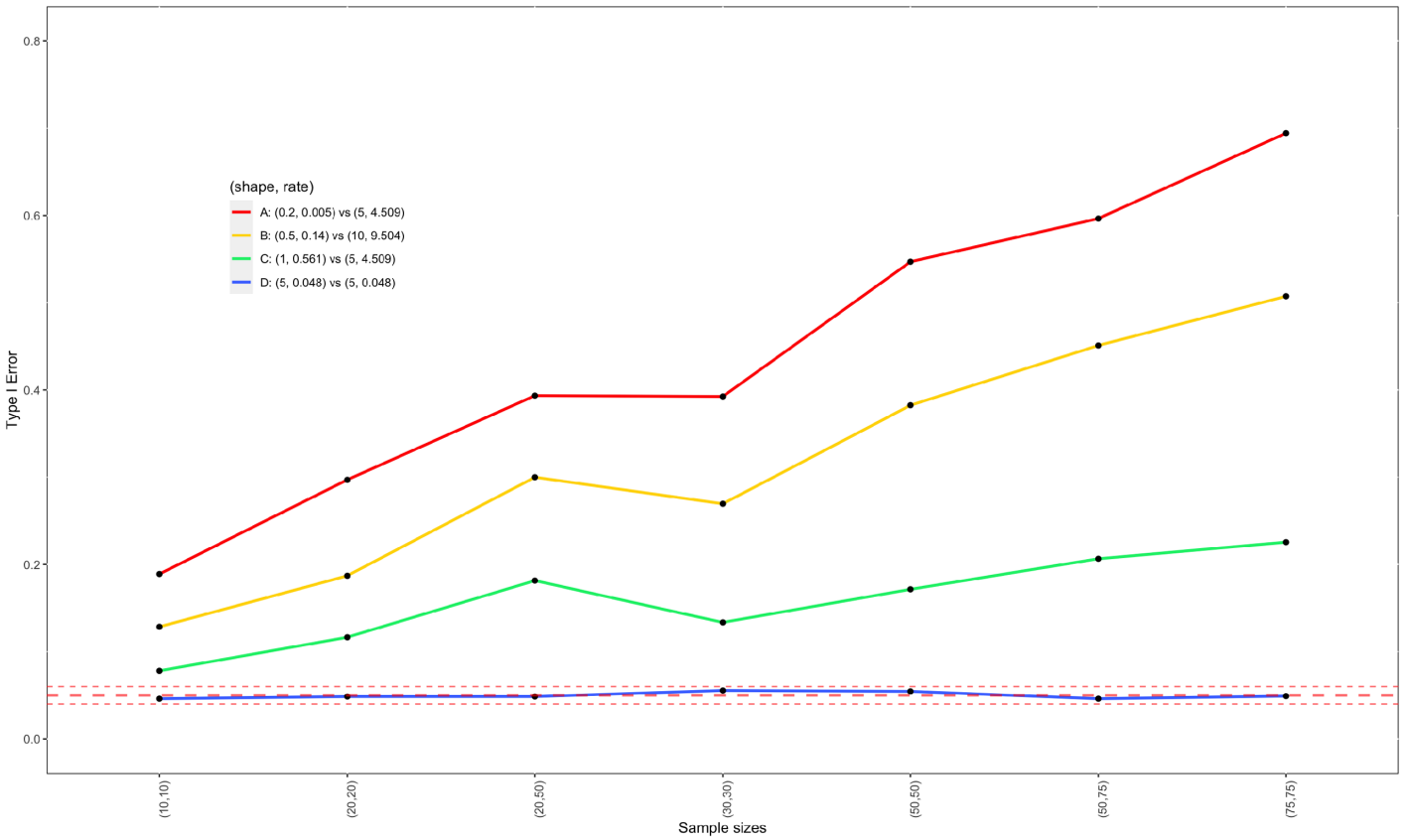


**Fig 5. Estimated type I errors of the Wilcoxon-Mann-Whitney (WMW) test for testing the equality of mean of two Exp-gamma distributions as a function of sample sizes.** The middle dashed line represents the nominal significance level, which is set to $\alpha = 0.05$; and upper and lower dashed lines are upper and lower limits for the type I error rates, which are 0.06 and 0.04, respectively.

https://doi.org/10.1371/journal.pone.0314705.g005

control under $H_0 : \delta_1 = \delta_2$ vs. $H_1 : \delta_1 \neq \delta_2$, and estimate the confidence interval for the mean difference $\eta = \delta_1 - \delta_2$ with satisfactory coverage probability. jpeg.

## 4.1 The method based on generalized inference

The concepts of generalized variables and generalized pivots were introduced by Tsui and Weerahandi [40] and Weerahandi [41]. More details can be found in the book of Werrahandi [42]. In S2 Appendix, a brief summary of the core concepts is presented. The concepts of generalized pivotal quantity and generalized confidence interval have been successfully applied to a variety of practical problems when standard exact solutions do not exist, and it has been shown that generalized inference methods generally have good performance, even when sample sizes are small; see e.g. [43–45].

Although there does not exist exact generalized pivots for gamma parameters, approximates generalized pivots have been proposed [22–26]. These approximate pivots have been utilized to make inference for gamma distributions, including single gamma means and difference between two gamma means under different scenarios [27–29]. Utilizing the existing approximate generalized pivots for gamma parameters, we will develop the generalized inference methods for hypothesis testing and confidence interval estimation for mean difference of two independent Exp-gamma distributions.

**4.1.1 Generalized pivots for population parameters: A review.** Assume $X \sim gamma(\alpha, \beta)$. In the following, we will first briefly review the existing approximate generalized pivots for gamma parameters $\alpha$ and $\beta$.

Krishnamoorthy and Wang's method: [25, 26] By applying the Wilson-Hilferty normal approximation, i.e. $W = X^{1/3} \sim N(\mu, \sigma^2)$. Generalized pivotal quantities for normal mean and variance, $R_\mu$ and $R_{\sigma^2}$ can be obtained for transformed data. Let $\bar{w}$ and $s_i^2$ be the observed sample mean and sample variance based on the transformed data $W$. The generalized pivotal quantities for $\alpha$ and $\beta$ can be further expressed as:

$$R_\alpha = \frac{1}{9}\left\{\left(1 + 0.5R_\mu^2/R_{\sigma^2}\right) + [(1 + 0.5R_\mu^2/R_{\sigma^2})^2 - 1]^{\frac{1}{2}}\right\},$$
$$R_\beta = \frac{1}{27(R_\alpha)^{\frac{1}{2}}(R_{\sigma^2})^{\frac{3}{2}}}, \tag{5}$$

where $R_\mu = \bar{w} - \frac{Z}{\sqrt{U_1}}\sqrt{\frac{(n-1)s^2}{n}}$, and $R_{\sigma^2} = \frac{(n-1)s^2}{U_2} \sim \frac{(n-1)s^2}{\chi^2_{n-1}}$, with $Z \sim N(0, 1)$, $U_1 \sim \chi^2_{n-1}$, $U_2 \sim \chi^2_{n-1}$, and $Z$, $U_1$, and $U_2$ are independent.

Chen and Ye's method: [22, 23] It is known that $2n\alpha \log(\bar{X}/\tilde{X}) \sim c\chi^2_\nu$ approximately, where $\nu = 2E^2(V_1)/\mathrm{Var}(V_1)$ and $c = E(V_1/\nu)$. The detailed formulas for $E(V_1)$ and $\mathrm{Var}(V_1)$ can be found in Chen and Ye [22]. Using this result, an approximate generalized pivotal quantity for $\alpha$ can be written as

$$R_\alpha = V_1/[2n \log(\bar{x}/\tilde{x}))],$$

where $V_1 \sim \hat{c}\chi^2_{\hat{\nu}}$, $\bar{x}$ and $\tilde{x}$ are observed values of $\bar{X}$ and $\tilde{X}$. Furthermore, utilizing a well-known result regarding gamma distribution, i.e. $2n\beta\bar{X} \sim \chi^2_{2n\alpha}$, the generalized pivot quantity for $\beta$ can be written as

$$R_\beta = V_2/(2n\bar{x}), \tag{6}$$

where $V_2 \sim \chi^2_{2nR_\alpha}$.

Wang and Wu's method: [24] Let $T = \log(\tilde{X}/\bar{X})$. Note that $U = F(.) \sim U(0, 1)$, where $F(.)$ is the c.d.f of $T$. On the basis of Cornish-Fisher expansion, the $U$th percentile of $T$ can be approximated by $\kappa_1(\alpha) + [\kappa_2(\alpha)]^{1/2}Q(\alpha, U)$, where $\kappa_j(\alpha)$ is the $j$th cumulant of $T$ and $Q(\alpha, U)$ is a function of $\kappa_j(\alpha)$'s. The detailed formulas can be found in Wang and Wu [24]. Let $t$ denote the observed value of $T$. An approximate generalized pivotal quantity for $\alpha$, i.e. $R_\alpha$, can be obtained by solving $t = \kappa_1(\alpha) + [\kappa_2(\alpha)]^{1/2}Q(\alpha, U)$. Similar to Chen and Ye's method, the approximate generalized pivotal quantity for rate parameter, $R_\beta$, can be obtained by Eq (6). This method improves Chen and Ye's method and can work well even when the shape parameter $\alpha$ is small.

**4.1.2 The generalized inference methods for hypothesis testing and confidence interval estimation for two independent Exp-gamma means.** For the $i$th ($i = 1, 2$) sample, the generalized pivotal quantities $R_{\alpha_i}$ and $R_{\beta_i}$ can be obtained by one of the three approximate generalized inference methods for gamma parameters, i.e. Krishnamoorthy and Wang's method [25, 26], Chen and Ye's method [22, 23], and Wang and Wu's method [24], as reviewed in Section 4.1.1. Replacing $\alpha_i$ with $R_{\alpha_i}$ and $\beta_i$ with $R_{\beta_i}$ in Eq (2), the generalized pivotal quantity for $\delta_i$ can be expressed as

$$R_{\delta_i} = \psi(R_{\alpha_i}) - \ln R_{\beta_i}, \quad i = 1, 2. \tag{7}$$

The generalized pivotal quantity we propose for the mean difference ($\eta$) of two independent Exp-gamma distributions can be expressed as

$$R_\eta = R_{\delta_1} - R_{\delta_2} = \psi(R_{\alpha_1}) - \ln R_{\beta_1} - (\psi(R_{\alpha_2}) - \ln R_{\beta_2}). \tag{8}$$

It is easy to verify that $R_\eta$ is a *bona fide* generalized pivotal quantity for $\eta$ approximately. For a given data set $Y_{11}, Y_{12}, \ldots, Y_{1n_1}$ and $Y_{21}, Y_{22}, \ldots, Y_{2n_2}$, the following holds: 1) the distribution of $R_\eta$ is independent of any unknown parameters; 2) the value of $R_\eta$ is $\eta$ approximately when the statistics used in the definitions of $R_{\alpha_i}$ and $R_{\beta_i}$ ($i = 1, 2$) are equal their observed value (e.g. in $\bar{X}_i = \bar{x}_i$ and $\tilde{X}_i = \tilde{x}_i$ in Chen and Ye's method).

For testing the hypothesis of equality of two Exp-gamma means,

$$H_0 : \delta_1 - \delta_2 = \eta \text{ vs. } H_1 : \delta_1 - \delta_2 \neq \eta, \tag{9}$$

where $\eta = 0$. The generalized test variable is defined as

$$T_\eta = R_\eta - \eta \tag{10}$$

where $R_\eta$ is the generalized pivotal quantity defined in Eq (8). Note that $T_\eta$ satisfies the three conditions to be a *bona fide* generalized test variables: 1) the distribution of $T_\eta$ is free of nuisance parameters; (2) $t_\eta$, the observed value of $T_\eta$, is 0, and hence is free of any unknown parameters; and (3) $T_\eta$ is stochastically decreasing in $\eta$.

The generalized $p$-value for testing the hypothesis of equality of two Exp-gamma means is given by

$$2 \times \min\{P(R_\eta \leq 0), P(R_\eta \geq 0)\}. \tag{11}$$

**4.1.3 Computing algorithm.** Consider a given data set $Y_{ij}$'s ($i = 1, 2, j = 1, 2, \ldots, n_i$) where the $i$th sample $Y_i \sim$ *Exp-gamma*($\alpha_i, \beta_i$). The generalized $p$-value for testing equality of two Exp-gamma means, and estimated confidence interval of the mean difference of two Exp-gamma distributions, can be computed by the following steps:

1. Use one of the three methods presented above, generate $R_{\alpha_i}$ and $R_{\beta_i}$ for $i = 1, 2$, then compute generalized pivot $R_{\delta_i}$ for $\delta_i$ following (7) for $i = 1, 2$.
2. Compute generalized pivot $R_\eta = R_{\delta_1} - R_{\delta_2}$ for $\eta$ following (8).
3. Repeat steps 1-2 a total $B$ ($B = 2000$) times and obtain array of $R_\eta^b$'s for $b = 1, 2, \ldots, B$.

Let $R_{\eta : p}$ denote the $100p$ percentile of the $B$ $R_\eta$'s generated in the preceding steps. Then ($R_{\eta : p/2}, R_{\eta : 1 - p/2}$) is a $100(1 - p)\%$ confidence interval for the mean difference of two independent Exp-gamma distributions.

Under the $H_0 : \delta_1 = \delta_2$, the generalized $p$-value can be obtained by (11), i.e.

$$p\text{−value} = 2 \times \min\{\frac{\sum_{i=1}^{B} I_{\{R_\eta^b \leq 0\}}}{B}, \frac{\sum_{i=1}^{B} I_{\{R_\eta^b \geq 0\}}}{B}\}. \tag{12}$$

The $H_0$ can be rejected if the $p$-value is less than a given significant level $a$.

We refer the three methods based on the generalized pivotal quantity of Exp-gamma mean difference as $\mathbf{G_K}$, $\mathbf{G_C}$, and $\mathbf{G_W}$, corresponding to the methods used for gamma parameters, i.e. Krishnamoorthy and Wang's method, [25, 26], Chen and Ye's method [22, 23], and Wang and Wu's method [24], respectively.

### 4.2 The parametric bootstrap method

Parametric bootstrap (**PB**) method has been widely used in estimating confidence intervals when the parametric model is justified, e.g. [21, 46]. In this section, we propose a **PB** method for hypothesis testing and confidence interval estimation for mean different of two independent Exp-gamma distributions.

Let $\bar{Y}_i$ denotes the mean based on a sample of size $n_i$ from a $Exp\text{-}gamma(\alpha_i, \beta_i)$ distribution, $i = 1, 2$. Let $\hat{\alpha}_i$ and $\hat{\beta}_i$ denote the MLEs of $\alpha_i$ and $\beta_i$, respectively. Similarly, let $\bar{Y}_i^*$ denotes the mean based on a bootstrap sample of size $n_i$ from the $Exp-gamma(\hat{\alpha}_i, \hat{\beta}_i)$. Let $(\hat{\alpha}_i^*, \hat{\beta}_i^*)$ denote the MLEs based on a boostrap sample, $i = 1, 2$. The **PB** pivot to estimate the difference between two means $\delta_1 = \psi(\alpha_1) - \ln(\beta_1)$ and $\delta_2 = \psi(\alpha_2) - \ln(\beta_2)$ is given by

$$Q_\eta = \frac{(\bar{Y}_1^* - \bar{Y}_2^*) - (\bar{Y}_1 - \bar{Y}_2)}{\sqrt{\frac{\psi^{(1)}(\hat{\alpha}_1^*)}{n_1} + \frac{\psi^{(1)}(\hat{\alpha}_2^*)}{n_2}}}. \tag{13}$$

The following steps can be used to obtain the $p$-values for hypothesis testing in Eq (9), decision rules, and confidence interval for $\eta$ based on **PB** method:

1. For a given sample of size $n_i$, calculate the MLEs $\hat{\alpha}_i$ and $\hat{\beta}_i$, $i = 1, 2$.
2. Generate bootstrap samples of size $n_i$ from gamma$(\hat{\alpha}_i, \hat{\beta}_i)$. Then calculate the $\bar{Y}_i^*$, and MLEs $(\hat{\alpha}_i^*, \hat{\beta}_i^*)$ based on the bootstrap samples for $i = 1, 2$.
3. Calculate $Q_\eta$ as in Eq (13).
4. Repeat steps 2–3 a total $B$ ($B = 2000$) times and obtain array of $Q_\eta^b$'s for $b = 1, 2, \ldots, B$.
5. The p-value can be obtained by

$$p-\text{value} = 2 \times \min\{\frac{\sum_{i=1}^B I_{\{Q_\eta^b \le 0\}}}{B}, \frac{\sum_{i=1}^B I_{\{Q_\eta^b \ge 0\}}}{B}\}, H_a : \delta_1 \neq \delta_2.$$

6. The $H_0$ can be rejected if the $p$-values is less than a given significant level $a$.
7. The $100(1-p)\%$ **PB** confidence interval can be obtained as

$$\{(\bar{Y}_1 - \bar{Y}_2) - Q_{\eta;1-p/2}\sqrt{\frac{\psi^{(1)}(\hat{\alpha}_1)}{n_1} + \frac{\psi^{(1)}(\hat{\alpha}_2)}{n_2}}, (\bar{Y}_1 - \bar{Y}_2) - Q_{\eta;p/2}\sqrt{\frac{\psi^{(1)}(\hat{\alpha}_1)}{n_1} + \frac{\psi^{(1)}(\hat{\alpha}_2)}{n_2}}\},$$

where $Q_{\eta;p}$ denotes the $100p$ percentile of $Q_\eta$.

## 5 Simulation studies

In previous section, we presented several methods for hypothesis testing and confidence interval estimation for the mean difference between two independent Exp-gamma distributions: three methods based on the generalized pivots (i.e.$\mathbf{G_C}$, $\mathbf{G_W}$, and $\mathbf{G_K}$), and a parametric bootstrap method (i.e. **PB**). Simulation studies are carried out to evaluate the performance of the proposed methods for hypothesis testing and confidence interval estimation.

## 5.1 Hypothesis testing

Sample sizes are set from small (10) to large (75), including both balanced and unbalanced settings. The parameter settings for type I error control include scenarios of equal/unequal shape parameters, with the common mean of two samples ranging from −1.369 to 4.634. The parameter settings for power study include scenarios with equal/unequal shape parameters with the mean difference ranging from 0.5 to 1.386. For each parameter setting, 2000 random samples are generated with given sample sizes. For the type I error and power based on generalized inference methods ($\mathbf{G_C}$, $\mathbf{G_W}$, and $\mathbf{G_K}$), 2000 values of generalized pivots are obtained for each random sample. For the type I error and power obtained by **PB** method, 2000 bootstrap samples are generated for each random sample.

Table 1 presents the type I error rate estimates of hypothesis testing based on the proposed methods ($\mathbf{G_C}$, $\mathbf{G_W}$, $\mathbf{G_K}$, and **PB**), in comparison with t-test and the WMW test, for testing the equality of means of two Exp-gamma distributions. Note that for the first three scenarios, the two Exp-gamma distributions are identical. The remaining scenarios are ranked using $P(Y_1 > Y_2)$ in ascending order, indicating a larger disparity between two Exp-gamma distributions under the null hypothesis.

Out of the three proposed methods based on the generalized pivots, $\mathbf{G_C}$ and $\mathbf{G_W}$ have excellent type I error control regardless of shapes, rates, sample sizes, and the value of $P(Y_1 > Y_2)$. In contract, $\mathbf{G_K}$ can have inflated type I errors when the shape parameter(s) are small (e.g. scenarios 10, 15–18). The reason is that $\mathbf{G_K}$ obtains approximate generalized pivotal quantities based on the normal approximation of the distribution with a cube root transformation, and such approximation can be very inaccurate when shape parameter is small [22]. For all scenarios, the **PB** method has inflated type I errors when sample sizes are less than (50, 50). As sample sizes increase, the **PB** method shows improved type I error control. The inflation of type I errors in the **PB** method with small sample sizes is primarily due to the instability in estimating parameters from limited data. Furthermore, this improvement in type I error control does not occur when the shape parameter(s) are small (e.g. scenarios 10, 15–18). Small shape parameters in Exp-gamma distributions lead to highly left-skewed data, which exacerbates the difficulty of accurately estimating the parameters. The type I error of two sample t-test converges to nominal level as sample sizes increase, as guaranteed by the center limit theorem. However, it can have inflated type I errors when sample sizes are less than (50, 50), especially when the disparity between two distributions is obvious (e.g. when $P(Y_1 > Y_2)$ is larger than 0.555). For the first three scenarios for which two Exp-gamma distributions are identical, the WMW test has excellent type I error control. However, as the value of $P(Y_1 > Y_2)$ deviates from 0.5, the WMW test tends to have more severely inflated type I errors as sample sizes increase. Moreover, the magnitude of type I error inflation increases as the value of $P(Y_1 > Y_2)$ becomes larger.

Note that scenarios 2, 12, 16, and 18 in Table 1 are the scenario D, C, B and A, respectively, discussed in Section 3.2 and the type I errors of two sample t-test and the WMW test are presented in Figs 4 and 5. To help to visualize the performance of the proposed methods, Fig 6 presents the type I errors obtained $\mathbf{G_C}$, $\mathbf{G_W}$, $\mathbf{G_K}$, and **PB**, for these four scenarios.

Table 2 presents estimated power of hypothesis testing based on proposed methods ($\mathbf{G_C}$, $\mathbf{G_W}$, $\mathbf{G_K}$, and **PB**), in comparison with t-test and the WMW test.

Reflecting on the type I error control presented in Table 1, caution should be exercised while interpreting estimated power and making comparisons between methods. Note the following: 1) the power of the WMW method when the value of $P(Y_1 > Y_2)$ deviates from 0.5 is not interpretable due to its inflated type I error for these cases, 2) the power of two sample t-test might be inflated when sample sizes less than 50 due to its inflated type I error for these

**Table 1. Estimated type I errors for testing the equality of means of two independent Exp-gamma distributions (2000 simulations).**

| Scenario | $(\alpha_1, \beta_1)$ $(\alpha_2, \beta_2)$ | Mean | $P(Y_1 > Y_2)$ | $(skew_1, ex\text{-}kurt_1)^{\dagger}$ $(skew_2, ex\text{-}kurt_2)$ | Sample size | Type I Error | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $G_C$ | $G_W$ | $G_K$ | PB | t-test | WMW |
| 1 | (1, 1.5) (1, 1.5) | -0.983 | 0.500 | (-1.140, -0.600) (-1.140, -0.600) | (10,10) | 0.036 | 0.057 | 0.044 | 0.099 | 0.058 | 0.056 |
| | | | | | (20,20) | 0.038 | 0.042 | 0.040 | 0.066 | 0.049 | 0.052 |
| | | | | | (20,50) | 0.041 | 0.048 | 0.036 | 0.062 | 0.050 | 0.048 |
| | | | | | (30,30) | 0.038 | 0.042 | 0.031 | 0.054 | 0.039 | 0.044 |
| | | | | | (50,50) | 0.042 | 0.041 | 0.042 | 0.052 | 0.046 | 0.045 |
| | | | | | (50,75) | 0.042 | 0.049 | 0.043 | 0.061 | 0.053 | 0.047 |
| | | | | | (75,75) | 0.033 | 0.037 | 0.037 | 0.044 | 0.038 | 0.041 |
| 2* | (5, 0.048) (5, 0.048) | 4.541 | 0.500 | (-0.469, -2.563) (-0.469, -2.563) | (10,10) | 0.036 | 0.035 | 0.037 | 0.084 | 0.047 | 0.046 |
| | | | | | (20,20) | 0.039 | 0.028 | 0.036 | 0.057 | 0.045 | 0.049 |
| | | | | | (20,50) | 0.045 | 0.042 | 0.042 | 0.064 | 0.052 | 0.049 |
| | | | | | (30,30) | 0.049 | 0.043 | 0.044 | 0.062 | 0.048 | 0.056 |
| | | | | | (50,50) | 0.057 | 0.045 | 0.057 | 0.057 | 0.056 | 0.055 |
| | | | | | (50,75) | 0.046 | 0.040 | 0.044 | 0.052 | 0.047 | 0.046 |
| | | | | | (75,75) | 0.041 | 0.037 | 0.041 | 0.040 | 0.042 | 0.049 |
| 3 | (50, 0.048) (50, 0.048) | 4.634 | 0.500 | (-0.142, -2.960) (-0.142, -2.960) | (10,10) | 0.043 | 0.034 | 0.040 | 0.085 | 0.052 | 0.048 |
| | | | | | (20,20) | 0.057 | 0.047 | 0.051 | 0.077 | 0.057 | 0.057 |
| | | | | | (20,50) | 0.045 | 0.046 | 0.046 | 0.057 | 0.051 | 0.054 |
| | | | | | (30,30) | 0.046 | 0.047 | 0.049 | 0.061 | 0.055 | 0.055 |
| | | | | | (50,50) | 0.050 | 0.045 | 0.049 | 0.050 | 0.047 | 0.048 |
| | | | | | (50,75) | 0.052 | 0.044 | 0.048 | 0.054 | 0.050 | 0.043 |
| | | | | | (75,75) | 0.056 | 0.048 | 0.055 | 0.055 | 0.053 | 0.056 |
| 4 | (5, 0.048) (10, 0.101) | 4.541 | 0.512 | (-0.469, -2.563) (-0.324, -2.790) | (10,10) | 0.028 | 0.030 | 0.031 | 0.073 | 0.043 | 0.046 |
| | | | | | (20,20) | 0.048 | 0.041 | 0.045 | 0.064 | 0.051 | 0.058 |
| | | | | | (20,50) | 0.054 | 0.051 | 0.047 | 0.068 | 0.057 | 0.071 |
| | | | | | (30,30) | 0.044 | 0.046 | 0.046 | 0.061 | 0.048 | 0.053 |
| | | | | | (50,50) | 0.045 | 0.041 | 0.047 | 0.047 | 0.048 | 0.051 |
| | | | | | (50,75) | 0.054 | 0.056 | 0.056 | 0.060 | 0.056 | 0.072 |
| | | | | | (75,75) | 0.053 | 0.046 | 0.055 | 0.051 | 0.051 | 0.064 |
| 5 | (1.5, 4.077) (2, 6) | -1.369 | 0.513 | (-0.917, -1.388) (-0.780, -1.812) | (10,10) | 0.022 | 0.030 | 0.029 | 0.078 | 0.042 | 0.045 |
| | | | | | (20,20) | 0.034 | 0.032 | 0.033 | 0.053 | 0.038 | 0.045 |
| | | | | | (20,50) | 0.049 | 0.044 | 0.045 | 0.061 | 0.052 | 0.049 |
| | | | | | (30,30) | 0.045 | 0.040 | 0.041 | 0.058 | 0.050 | 0.052 |
| | | | | | (50,50) | 0.039 | 0.037 | 0.042 | 0.052 | 0.042 | 0.052 |
| | | | | | (50,75) | 0.048 | 0.038 | 0.046 | 0.054 | 0.047 | 0.052 |
| | | | | | (75,75) | 0.044 | 0.041 | 0.041 | 0.054 | 0.047 | 0.057 |
| 6 | (2, 0.200) (4, 0.460) | 2.032 | 0.522 | (-0.780, -1.812) (-0.529, -2.443) | (10,10) | 0.028 | 0.034 | 0.029 | 0.086 | 0.046 | 0.047 |
| | | | | | (20,20) | 0.040 | 0.038 | 0.040 | 0.057 | 0.047 | 0.053 |
| | | | | | (20,50) | 0.050 | 0.050 | 0.041 | 0.070 | 0.053 | 0.087 |
| | | | | | (30,30) | 0.048 | 0.044 | 0.046 | 0.059 | 0.050 | 0.060 |
| | | | | | (50,50) | 0.050 | 0.041 | 0.048 | 0.054 | 0.051 | 0.066 |
| | | | | | (50,75) | 0.051 | 0.048 | 0.054 | 0.061 | 0.059 | 0.087 |
| | | | | | (75,75) | 0.051 | 0.046 | 0.050 | 0.052 | 0.049 | 0.082 |

(*Continued*)

**Table 1.** (Continued)

| Scenario | $(\alpha_1, \beta_1)$ $(\alpha_2, \beta_2)$ | Mean | $P(Y_1 > Y_2)$ | $(skew_1, ex\text{-}kurt_1)^{\dagger}$ $(skew_2, ex\text{-}kurt_2)$ | Sample size | Type I Error | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $G_C$ | $G_W$ | $G_K$ | PB | t-test | WMW |
| 7 | (2, 0.300) (6, 1.083) | 1.627 | 0.531 | (-0.780, -1.812) (-0.425, -2.640) | (10,10) | 0.034 | 0.040 | 0.038 | 0.080 | 0.054 | 0.054 |
| | | | | | (20,20) | 0.048 | 0.047 | 0.046 | 0.071 | 0.053 | 0.078 |
| | | | | | (20,50) | 0.050 | 0.051 | 0.045 | 0.078 | 0.064 | 0.117 |
| | | | | | (30,30) | 0.049 | 0.047 | 0.050 | 0.064 | 0.053 | 0.089 |
| | | | | | (50,50) | 0.056 | 0.049 | 0.054 | 0.060 | 0.058 | 0.100 |
| | | | | | (50,75) | 0.048 | 0.050 | 0.044 | 0.052 | 0.047 | 0.106 |
| | | | | | (75,75) | 0.055 | 0.052 | 0.055 | 0.055 | 0.054 | 0.124 |
| 8 | (3, 2.516) (20, 19.502) | 0 | 0.532 | (-0.621, -2.237) (-0.226, -2.898) | (10,10) | 0.041 | 0.041 | 0.037 | 0.085 | 0.056 | 0.065 |
| | | | | | (20,20) | 0.042 | 0.037 | 0.042 | 0.058 | 0.046 | 0.076 |
| | | | | | (20,50) | 0.048 | 0.050 | 0.044 | 0.069 | 0.052 | 0.126 |
| | | | | | (30,30) | 0.047 | 0.047 | 0.046 | 0.062 | 0.048 | 0.082 |
| | | | | | (50,50) | 0.054 | 0.058 | 0.051 | 0.063 | 0.057 | 0.117 |
| | | | | | (50,75) | 0.055 | 0.054 | 0.050 | 0.061 | 0.058 | 0.129 |
| | | | | | (75,75) | 0.054 | 0.051 | 0.062 | 0.057 | 0.055 | 0.116 |
| 9 | (1, 1.500) (2, 4.077) | -0.983 | 0.534 | (-1.140, -0.600) (-0.780, -1.812) | (10,10) | 0.037 | 0.048 | 0.038 | 0.095 | 0.062 | 0.059 |
| | | | | | (20,20) | 0.040 | 0.043 | 0.036 | 0.060 | 0.047 | 0.067 |
| | | | | | (20,50) | 0.046 | 0.055 | 0.041 | 0.075 | 0.060 | 0.104 |
| | | | | | (30,30) | 0.049 | 0.050 | 0.051 | 0.068 | 0.059 | 0.085 |
| | | | | | (50,50) | 0.044 | 0.048 | 0.047 | 0.066 | 0.052 | 0.097 |
| | | | | | (50,75) | 0.047 | 0.050 | 0.046 | 0.057 | 0.051 | 0.104 |
| | | | | | (75,75) | 0.035 | 0.039 | 0.044 | 0.043 | 0.043 | 0.114 |
| 10 | (0.5, 0.14) (1, 0.621) | 0 | 0.553 | (-1.535, 1) (-1.140, -0.600) | (10,10) | 0.035 | 0.051 | 0.035 | 0.098 | 0.060 | 0.069 |
| | | | | | (20,20) | 0.045 | 0.059 | 0.047 | 0.072 | 0.060 | 0.097 |
| | | | | | (20,50) | 0.035 | 0.053 | 0.027 | 0.060 | 0.056 | 0.117 |
| | | | | | (30,30) | 0.041 | 0.062 | 0.056 | 0.071 | 0.063 | 0.125 |
| | | | | | (50,50) | 0.030 | 0.051 | 0.049 | 0.050 | 0.045 | 0.141 |
| | | | | | (50,75) | 0.034 | 0.054 | 0.052 | 0.059 | 0.056 | 0.186 |
| | | | | | (75,75) | 0.041 | 0.063 | 0.086 | 0.062 | 0.053 | 0.218 |
| 11 | (1, 0.207) (5, 1.659) | 1 | 0.555 | (-1.140, -0.600) (-0.469, -2.563) | (10,10) | 0.046 | 0.054 | 0.044 | 0.108 | 0.072 | 0.090 |
| | | | | | (20,20) | 0.038 | 0.046 | 0.039 | 0.066 | 0.055 | 0.107 |
| | | | | | (20,50) | 0.043 | 0.051 | 0.035 | 0.068 | 0.052 | 0.167 |
| | | | | | (30,30) | 0.044 | 0.047 | 0.041 | 0.064 | 0.054 | 0.133 |
| | | | | | (50,50) | 0.047 | 0.056 | 0.046 | 0.066 | 0.059 | 0.175 |
| | | | | | (50,75) | 0.044 | 0.057 | 0.046 | 0.057 | 0.055 | 0.221 |
| | | | | | (75,75) | 0.038 | 0.042 | 0.048 | 0.042 | 0.046 | 0.246 |
| 12* | (1, 0.561) (5, 4.509) | 0 | 0.556 | (-1.140, -0.600) (-0.469, -2.563) | (10,10) | 0.046 | 0.050 | 0.038 | 0.106 | 0.066 | 0.078 |
| | | | | | (20,20) | 0.044 | 0.052 | 0.042 | 0.079 | 0.059 | 0.117 |
| | | | | | (20,50) | 0.048 | 0.055 | 0.041 | 0.082 | 0.067 | 0.182 |
| | | | | | (30,30) | 0.050 | 0.056 | 0.041 | 0.068 | 0.053 | 0.134 |
| | | | | | (50,50) | 0.043 | 0.043 | 0.044 | 0.056 | 0.055 | 0.172 |
| | | | | | (50,75) | 0.043 | 0.051 | 0.041 | 0.056 | 0.053 | 0.207 |
| | | | | | (75,75) | 0.043 | 0.045 | 0.053 | 0.052 | 0.052 | 0.226 |

(*Continued*)

**Table 1.** (Continued)

| Scenario | $(\alpha_1, \beta_1)$ $(\alpha_2, \beta_2)$ | Mean | $P(Y_1 > Y_2)$ | $(skew_1, ex\text{-}kurt_1)^{\dagger}$ $(skew_2, ex\text{-}kurt_2)$ | Sample size | Type I Error | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $G_C$ | $G_W$ | $G_K$ | PB | t-test | WMW |
| 13 | (1, 0.561) (10, 9.504) | 0 | 0.564 | (-1.140, -0.600) (-0.324, -2.790) | (10,10) | 0.047 | 0.052 | 0.038 | 0.103 | 0.063 | 0.091 |
| | | | | | (20,20) | 0.036 | 0.045 | 0.038 | 0.068 | 0.055 | 0.122 |
| | | | | | (20,50) | 0.052 | 0.057 | 0.042 | 0.075 | 0.058 | 0.213 |
| | | | | | (30,30) | 0.051 | 0.052 | 0.044 | 0.068 | 0.060 | 0.160 |
| | | | | | (50,50) | 0.042 | 0.052 | 0.046 | 0.057 | 0.054 | 0.215 |
| | | | | | (50,75) | 0.037 | 0.045 | 0.040 | 0.056 | 0.050 | 0.243 |
| | | | | | (75,75) | 0.044 | 0.050 | 0.058 | 0.053 | 0.052 | 0.277 |
| 14 | (1, 0.561) (50, 49.501) | 0 | 0.569 | (-1.140, -0.600) (-0.142, -2.960) | (10,10) | 0.051 | 0.053 | 0.041 | 0.112 | 0.061 | 0.110 |
| | | | | | (20,20) | 0.051 | 0.056 | 0.051 | 0.082 | 0.065 | 0.152 |
| | | | | | (20,50) | 0.040 | 0.046 | 0.032 | 0.065 | 0.048 | 0.222 |
| | | | | | (30,30) | 0.045 | 0.049 | 0.050 | 0.067 | 0.059 | 0.191 |
| | | | | | (50,50) | 0.040 | 0.046 | 0.042 | 0.055 | 0.050 | 0.251 |
| | | | | | (50,75) | 0.042 | 0.049 | 0.042 | 0.055 | 0.049 | 0.288 |
| | | | | | (75,75) | 0.039 | 0.045 | 0.046 | 0.043 | 0.046 | 0.326 |
| 15 | (0.5, 0.052) (5, 1.659) | 1 | 0.587 | (-1.535, 1) (-0.469, -2.563) | (10,10) | 0.042 | 0.057 | 0.042 | 0.106 | 0.073 | 0.119 |
| | | | | | (20,20) | 0.045 | 0.062 | 0.051 | 0.072 | 0.067 | 0.193 |
| | | | | | (20,50) | 0.042 | 0.061 | 0.042 | 0.076 | 0.064 | 0.299 |
| | | | | | (30,30) | 0.044 | 0.055 | 0.054 | 0.076 | 0.068 | 0.255 |
| | | | | | (50,50) | 0.042 | 0.061 | 0.061 | 0.067 | 0.064 | 0.334 |
| | | | | | (50,75) | 0.036 | 0.054 | 0.065 | 0.066 | 0.058 | 0.409 |
| | | | | | (75,75) | 0.042 | 0.057 | 0.112 | 0.064 | 0.061 | 0.480 |
| 16* | (0.5, 0.140) (10, 9.504) | 0 | 0.593 | (-1.535, 1) (-0.324, -2.790) | (10,10) | 0.040 | 0.056 | 0.039 | 0.113 | 0.081 | 0.131 |
| | | | | | (20,20) | 0.035 | 0.055 | 0.048 | 0.066 | 0.061 | 0.211 |
| | | | | | (20,50) | 0.047 | 0.053 | 0.039 | 0.074 | 0.068 | 0.293 |
| | | | | | (30,30) | 0.043 | 0.056 | 0.061 | 0.078 | 0.073 | 0.274 |
| | | | | | (50,50) | 0.034 | 0.051 | 0.061 | 0.056 | 0.052 | 0.390 |
| | | | | | (50,75) | 0.037 | 0.055 | 0.066 | 0.063 | 0.058 | 0.436 |
| | | | | | (75,75) | 0.044 | 0.065 | 0.111 | 0.068 | 0.061 | 0.508 |
| 17 | (0.5, 0.031) (10, 2.121) | 1.500 | 0.595 | (-1.535, 1) (-0.324, -2.790) | (10,10) | 0.039 | 0.056 | 0.041 | 0.100 | 0.074 | 0.140 |
| | | | | | (20,20) | 0.039 | 0.057 | 0.051 | 0.071 | 0.062 | 0.196 |
| | | | | | (20,50) | 0.041 | 0.053 | 0.034 | 0.070 | 0.067 | 0.299 |
| | | | | | (30,30) | 0.035 | 0.047 | 0.049 | 0.059 | 0.058 | 0.269 |
| | | | | | (50,50) | 0.029 | 0.040 | 0.056 | 0.053 | 0.047 | 0.382 |
| | | | | | (50,75) | 0.043 | 0.060 | 0.074 | 0.072 | 0.066 | 0.441 |
| | | | | | (75,75) | 0.030 | 0.042 | 0.102 | 0.052 | 0.049 | 0.519 |
| 18* | (0.2, 0.005) (5, 4.509) | 0 | 0.621 | (-1.868, 2.440) (-0.469, -2.563) | (10,10) | 0.034 | 0.054 | 0.076 | 0.113 | 0.094 | 0.189 |
| | | | | | (20,20) | 0.032 | 0.050 | 0.250 | 0.074 | 0.076 | 0.297 |
| | | | | | (20,50) | 0.037 | 0.056 | 0.174 | 0.072 | 0.073 | 0.394 |
| | | | | | (30,30) | 0.034 | 0.057 | 0.391 | 0.069 | 0.073 | 0.393 |
| | | | | | (50,50) | 0.035 | 0.055 | 0.676 | 0.059 | 0.062 | 0.547 |
| | | | | | (50,75) | 0.036 | 0.060 | 0.667 | 0.068 | 0.067 | 0.597 |
| | | | | | (75,75) | 0.042 | 0.059 | 0.908 | 0.066 | 0.060 | 0.695 |

* Scenarios discussed in Section 3.2. Scenarios 12, 16, and 18 are the scenarios C, B, and A, respectively.

† $skew_i$ and $ex\text{-}kurt_i$ are defined in Eq (3), $i = 1, 2$.

https://doi.org/10.1371/journal.pone.0314705.t001

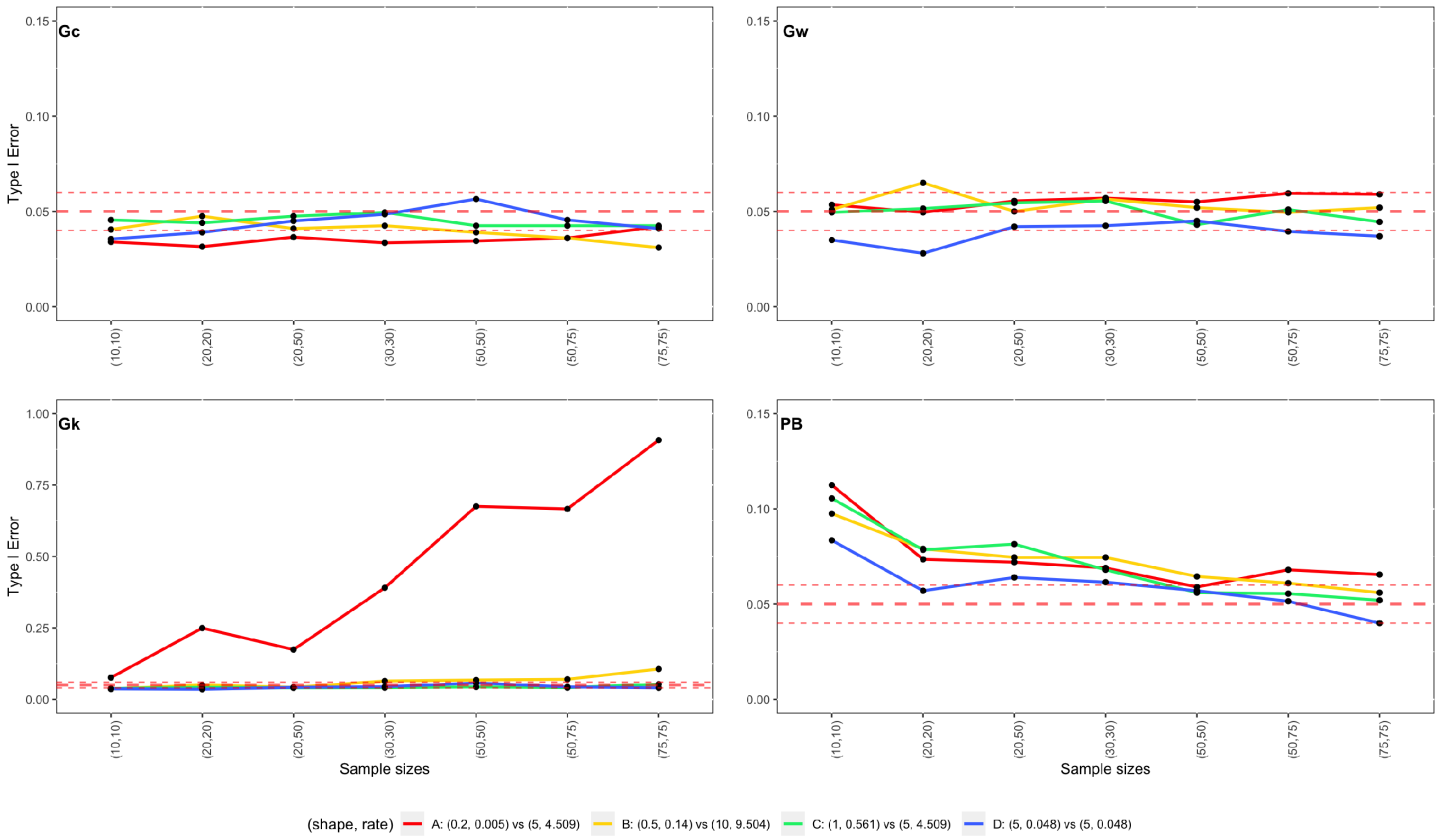


**Fig 6. Estimated type I errors of the hypothesis testing based on generalized pivots ($G_C$, $G_W$, and $G_K$) and parametric bootstrap method (PB) for testing the equality of mean of two Exp-gamma distributions as a function of sample sizes.** The middle dashed line represents the nominal significance level, which is set to $\alpha$ = 0.05; and upper and lower dashed lines are upper and lower limits for the type I error rates, which are 0.06 and 0.04, respectively.

https://doi.org/10.1371/journal.pone.0314705.g006

cases; 3) the power of the PB method could be inflated due to its poor type I error control, especially at small sample sizes; 4) $\mathbf{G_K}$ can have inflated power due to inflated type I error when the shape parameter ($\alpha$) is small. For example, for scenarios 26 and 27 where the value of $P(Y_1 > Y_2)$ has a larger deviation from 0.5, the WMW test and two sample t-test have higher power than that of $\mathbf{G_C}$, $\mathbf{G_W}$ and $\mathbf{G_K}$ when sample sizes are small. Such observations are due to the inflated type I error for the WMW test and the two sample t-test; hence they should not be interpreted as an evidence that t-test and the WMW test are more powerful.

Overall speaking, the two generalized inference methods with good type I error control, i.e. $\mathbf{G_C}$ and $\mathbf{G_W}$, have comparable power. When sample sizes exceed (50, 50), the powers by two sample t-test and the PB test are comparable to those of $\mathbf{G_C}$ and $\mathbf{G_W}$.

In summary, we recommend both $\mathbf{G_C}$ and $\mathbf{G_W}$ methods for hypothesis testing of two independent Exp-gamma distributions due to their ability to provide decent power with excellent type I error control, even when sample sizes are small. The $\mathbf{G_K}$ method is not recommended because it has inflated type I errors for certain scenarios, such as when the shape parameter is less than 0.5. The **PB** method has inflated type I errors when sample sizes are small or when shape parameter(s) are small, leading to incorrect rejection of the null hypothesis. Two sample t-test may exhibit inflated type I errors at small sample sizes. The WMW test only maintains controlled type I errors when the two distributions are identical, hence it is not a reliable choice.

**Table 2. Estimated powers for testing the equality of means of two independent Exp-gamma distributions under $H_1 : \delta_1 \neq \delta_2$ (2000 simulations).**

| Scenario | $(\alpha_1, \beta_1)$ $(\alpha_2, \beta_2)$ | $\delta_1, \delta_2$ $\eta = \delta_1 - \delta_2$ | $P(Y_1 > Y_2)$ | $(skew_1, ex\text{-}kurt_1)^{\dagger}$ $(skew_2, ex\text{-}kurt_2)$ | Sample size | Power | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $G_C$ | $G_W$ | $G_K$ | PB | t-test | WMW |
| 19 | (5, 2.735) (0.5, 0.140) | 0.500, 0 0.5 | 0.509 | (-0.469, -2.563) (-1.535, 1) | (10,10) | 0.134 | 0.135 | 0.080 | 0.144 | 0.065 | 0.081 |
| | | | | | (20,20) | 0.177 | 0.211 | 0.093 | 0.167 | 0.093 | 0.081 |
| | | | | | (20,50) | 0.215 | 0.243 | 0.094 | 0.193 | 0.109 | 0.133 |
| | | | | | (30,30) | 0.242 | 0.292 | 0.130 | 0.234 | 0.167 | 0.080 |
| | | | | | (50,50) | 0.376 | 0.457 | 0.180 | 0.375 | 0.317 | 0.086 |
| | | | | | (50,75) | 0.374 | 0.454 | 0.188 | 0.381 | 0.316 | 0.102 |
| | | | | | (75,75) | 0.517 | 0.568 | 0.286 | 0.500 | 0.441 | 0.080 |
| | | | | | (75,75) | 0.932 | 0.934 | 0.924 | 0.943 | 0.935 | 0.856 |
| 20 | (5, 12.257) (1, 2.516) | -1, -1.5 0.500 | 0.607 | (-0.469, -2.563) (-1.140, -0.600) | (10,10) | 0.218 | 0.226 | 0.173 | 0.245 | 0.138 | 0.125 |
| | | | | | (20,20) | 0.434 | 0.424 | 0.362 | 0.417 | 0.321 | 0.227 |
| | | | | | (20,50) | 0.459 | 0.485 | 0.390 | 0.450 | 0.337 | 0.305 |
| | | | | | (30,30) | 0.573 | 0.561 | 0.508 | 0.554 | 0.479 | 0.298 |
| | | | | | (50,50) | 0.779 | 0.809 | 0.756 | 0.786 | 0.757 | 0.478 |
| | | | | | (50,75) | 0.804 | 0.839 | 0.774 | 0.808 | 0.772 | 0.538 |
| | | | | | (75,75) | 0.921 | 0.935 | 0.901 | 0.916 | 0.902 | 0.613 |
| 21 | (5, 1.659) (0.5, 0.14) | 1,0 1 | 0.620 | (-0.469, -2.563) (-1.535, 1) | (10,10) | 0.376 | 0.376 | 0.252 | 0.363 | 0.179 | 0.163 |
| | | | | | (20,20) | 0.620 | 0.650 | 0.487 | 0.602 | 0.449 | 0.264 |
| | | | | | (20,50) | 0.656 | 0.681 | 0.478 | 0.623 | 0.467 | 0.374 |
| | | | | | (30,30) | 0.777 | 0.821 | 0.666 | 0.769 | 0.685 | 0.376 |
| | | | | | (50,50) | 0.950 | 0.963 | 0.880 | 0.952 | 0.933 | 0.560 |
| | | | | | (50,75) | 0.947 | 0.966 | 0.889 | 0.950 | 0.933 | 0.620 |
| | | | | | (75,75) | 0.995 | 0.996 | 0.969 | 0.990 | 0.985 | 0.705 |
| 22 | (2, 1.5) (1, 1) | 0.017, -0.577 0.594 | 0.640 | (-0.469, -2.563) (-1.140,-0.600) | (10,10) | 0.171 | 0.221 | 0.175 | 0.285 | 0.181 | 0.167 |
| | | | | | (20,20) | 0.438 | 0.426 | 0.408 | 0.471 | 0.416 | 0.349 |
| | | | | | (20,50) | 0.545 | 0.549 | 0.493 | 0.534 | 0.443 | 0.440 |
| | | | | | (30,30) | 0.603 | 0.573 | 0.557 | 0.618 | 0.565 | 0.474 |
| | | | | | (50,50) | 0.810 | 0.799 | 0.810 | 0.831 | 0.810 | 0.696 |
| | | | | | (50,75) | 0.864 | 0.870 | 0.863 | 0.889 | 0.858 | 0.769 |
| | | | | | (75,75) | 0.932 | 0.934 | 0.924 | 0.943 | 0.935 | 0.856 |
| 23 | (3, 5) (2, 5) | -0.687,-1.187 0.500 | 0.688 | (-0.621, -2.237) (-0.780, -1.812) | (10,10) | 0.224 | 0.261 | 0.251 | 0.384 | 0.290 | 0.257 |
| | | | | | (20,20) | 0.562 | 0.517 | 0.561 | 0.616 | 0.562 | 0.545 |
| | | | | | (20,50) | 0.739 | 0.710 | 0.714 | 0.745 | 0.675 | 0.695 |
| | | | | | (30,30) | 0.754 | 0.720 | 0.745 | 0.781 | 0.758 | 0.729 |
| | | | | | (50,50) | 0.931 | 0.924 | 0.933 | 0.939 | 0.931 | 0.910 |
| | | | | | (50,75) | 0.967 | 0.966 | 0.968 | 0.971 | 0.967 | 0.952 |
| | | | | | (75,75) | 0.992 | 0.987 | 0.990 | 0.992 | 0.990 | 0.988 |
| 24 | (2, 5) (2, 10) | -1.187,-1.880 0.693 | 0.741 | (-0.780, -1.812) (-0.780, -1.812) | (10,10) | 0.313 | 0.354 | 0.353 | 0.572 | 0.459 | 0.439 |
| | | | | | (20,20) | 0.658 | 0.691 | 0.696 | 0.776 | 0.749 | 0.770 |
| | | | | | (20,50) | 0.752 | 0.790 | 0.761 | 0.885 | 0.868 | 0.896 |
| | | | | | (30,30) | 0.862 | 0.871 | 0.873 | 0.910 | 0.893 | 0.906 |
| | | | | | (50,50) | 0.987 | 0.984 | 0.987 | 0.988 | 0.989 | 0.994 |
| | | | | | (50,75) | 0.994 | 0.992 | 0.994 | 0.996 | 0.995 | 0.998 |
| | | | | | (75,75) | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

(*Continued*)

**Table 2.** (Continued)

| Scenario | $(\alpha_1, \beta_1)$ $(\alpha_2, \beta_2)$ | $\delta_1, \delta_2$ $\eta = \delta_1 - \delta_2$ | $P(Y_1 > Y_2)$ | $(skew_1, ex\text{-}kurt_1)^\dagger$ $(skew_2, ex\text{-}kurt_2)$ | Sample size | Power | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $G_C$ | $G_W$ | $G_K$ | PB | t-test | WMW |
| 25 | (5, 1.5)<br>(3, 1.5) | 1.101,0.517<br>0.584 | 0.773 | (-0.469, -2.563)<br>(-0.621, -2.237) | (10,10) | 0.523 | 0.540 | 0.562 | 0.681 | 0.591 | 0.561 |
| | | | | | (20,20) | 0.912 | 0.890 | 0.906 | 0.932 | 0.912 | 0.893 |
| | | | | | (20,50) | 0.975 | 0.975 | 0.972 | 0.980 | 0.967 | 0.964 |
| | | | | | (30,30) | 0.981 | 0.975 | 0.978 | 0.985 | 0.981 | 0.974 |
| | | | | | (50,50) | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 |
| | | | | | (50,75) | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | | | | (75,75) | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 26 | (3, 5)<br>(3, 10) | -0.687,-1.380<br>0.693 | 0.790 | (-0.621, -2.237)<br>(-0.621, -2.237) | (10,10) | 0.485 | 0.538 | 0.530 | 0.728 | 0.646 | 0.617 |
| | | | | | (20,20) | 0.875 | 0.892 | 0.895 | 0.929 | 0.917 | 0.921 |
| | | | | | (20,50) | 0.947 | 0.956 | 0.950 | 0.975 | 0.975 | 0.983 |
| | | | | | (30,30) | 0.978 | 0.981 | 0.982 | 0.990 | 0.987 | 0.987 |
| | | | | | (50,50) | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | | | | (50,75) | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | | | | (75,75) | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 27 | (1, 1)<br>(1, 4) | -0.577,-1.964<br>1.386 | 0.800 | (-1.140,-0.600)<br>(-1.140,-0.600) | (10,10) | 0.375 | 0.490 | 0.493 | 0.716 | 0.646 | 0.669 |
| | | | | | (20,20) | 0.804 | 0.869 | 0.884 | 0.926 | 0.914 | 0.952 |
| | | | | | (20,50) | 0.874 | 0.913 | 0.907 | 0.951 | 0.953 | 0.979 |
| | | | | | (30,30) | 0.954 | 0.971 | 0.977 | 0.983 | 0.981 | 0.993 |
| | | | | | (50,50) | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | | | | (50,75) | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | | | | (75,75) | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

† $skew_i$ and $ex\text{-}kurt_i$ are defined in Eq (3), $i = 1, 2$.

https://doi.org/10.1371/journal.pone.0314705.t002

## 5.2 Confidence intervals

The proposed three methods based on the generalized pivots (i.e. $\mathbf{G_C}$, $\mathbf{G_W}$, and $\mathbf{G_K}$), and parametric bootstrap (**PB**) method can provide estimated confidence interval for the mean difference between two Exp-gamma distributions. Additionally, the estimated confidence intervals by two sample t-test are also provided for comparison purpose. Note that theoretically, the WMW method can not yield estimated confidence interval for the mean difference.

Simulation studies are carried out to evaluate the performances of proposed methods regarding coverage probabilities and the average lengths of proposed confidence intervals for mean difference of two independent Exp-gamma distributions. The sample sizes are set as (10, 10), (20, 20), (30, 30), (20, 50), (50, 50), (50, 75), and (75, 75). We considered settings with equal means (i.e. $\eta = 0$), as well as different means (i.e. $\eta \neq 0$), and with equal/unequal shape parameters. For each parameter setting, 2000 samples are simulated. For generalized confidence intervals, 2000 $R_\eta$'s are obtained. For **PB** method, $B = 2000$ bootstrap samples are used.

Table 3 presents the coverage probabilities and average lengths of proposed confidence intervals. Overall speaking, the $\mathbf{G_C}$ and $\mathbf{G_W}$ methods that based on the generalized pivots maintain satisfactory coverage probabilities for all settings except that they might be slightly conservative at small sample sizes such as (10, 10), while the $\mathbf{G_K}$ method is not recommended when the shape parameter is less than 0.5, due to the fact that this normal-based method does not work well when shape parameter is small [22]. The confidence intervals obtained by the **PB** method are liberal when sample sizes are small, although its coverage probabilities converge to nominal level when sample sizes reach (50, 50). The coverage probabilities of the two sample t-

**Table 3. Coverage probabilities and average lengths of proposed 95% confidence intervals† for mean difference of two independent Exp-gamma distributions (2000 simulations).**

| Scenario* | $(\alpha_1, \beta_1)$ $(\alpha_2, \beta_2)$ | $\eta$ | Sample size | Coverage probability(Average length) | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | $G_C$ | $G_W$ | $G_K$ | PB | t-test |
| 2 | (5, 0.048) (5, 0.048) | 0 | (10,10) | 0.965 (1.021) | 0.965 (0.994) | 0.963 (0.970) | 0.916 (0.773) | 0.954 (0.872) |
| | | | (20,20) | 0.962 (0.627) | 0.972 (0.652) | 0.965 (0.628) | 0.943 (0.569) | 0.956 (0.599) |
| | | | (20,50) | 0.955 (0.523) | 0.958 (0.531) | 0.958 (0.527) | 0.936 (0.476) | 0.949 (0.499) |
| | | | (30,30) | 0.952 (0.496) | 0.958 (0.517) | 0.956 (0.503) | 0.939 (0.468) | 0.953 (0.485) |
| | | | (50,50) | 0.944 (0.375) | 0.955 (0.391) | 0.943 (0.375) | 0.943 (0.370) | 0.944 (0.373) |
| | | | (50,75) | 0.955 (0.341) | 0.961 (0.353) | 0.957 (0.341) | 0.949 (0.335) | 0.953 (0.339) |
| | | | (75,75) | 0.959 (0.304) | 0.963 (0.313) | 0.960 (0.305) | 0.960 (0.303) | 0.959 (0.303) |
| 3 | (50, 0.481) (50, 0.481) | 0 | (10,10) | 0.958 (0.282) | 0.966 (0.296) | 0.960 (0.284) | 0.915 (0.234) | 0.949 (0.265) |
| | | | (20,20) | 0.944 (0.184) | 0.954 (0.194) | 0.949 (0.188) | 0.923 (0.171) | 0.943 (0.181) |
| | | | (20,50) | 0.956 (0.158) | 0.955 (0.157) | 0.955 (0.157) | 0.943 (0.145) | 0.949 (0.152) |
| | | | (30,30) | 0.954 (0.151) | 0.954 (0.153) | 0.952 (0.151) | 0.940 (0.142) | 0.945 (0.146) |
| | | | (50,50) | 0.951 (0.112) | 0.955 (0.115) | 0.951 (0.113) | 0.950 (0.112) | 0.953 (0.113) |
| | | | (50,75) | 0.949 (0.102) | 0.956 (0.105) | 0.952 (0.103) | 0.947 (0.102) | 0.951 (0.103) |
| | | | (75,75) | 0.945 (0.091) | 0.952 (0.093) | 0.946 (0.091) | 0.945 (0.092) | 0.948 (0.092) |
| 4 | (5, 0.048) (10, 0.101) | 0 | (10,10) | 0.973 (0.864) | 0.971 (0.850) | 0.970 (0.841) | 0.927 (0.664) | 0.957 (0.756) |
| | | | (20,20) | 0.953 (0.536) | 0.959 (0.554) | 0.955 (0.539) | 0.936 (0.490) | 0.949 (0.516) |
| | | | (20,50) | 0.946 (0.481) | 0.950 (0.486) | 0.954 (0.493) | 0.932 (0.439) | 0.943 (0.465) |
| | | | (30,30) | 0.956 (0.435) | 0.954 (0.441) | 0.954 (0.430) | 0.940 (0.404) | 0.952 (0.418) |
| | | | (50,50) | 0.956 (0.323) | 0.959 (0.330) | 0.953 (0.321) | 0.953 (0.318) | 0.953 (0.321) |
| | | | (50,75) | 0.947 (0.307) | 0.945 (0.308) | 0.944 (0.305) | 0.941 (0.298) | 0.944 (0.302) |
| | | | (75,75) | 0.948 (0.259) | 0.954 (0.265) | 0.945 (0.259) | 0.949 (0.262) | 0.950 (0.261) |
| 1–96 | (2, 0.200) (4, 0.460) | 0 | (10,10) | 0.973 (1.607) | 0.967 (1.457) | 0.972 (1.474) | 0.914 (1.108) | 0.955 (1.271) |
| | | | (20,20) | 0.961 (0.953) | 0.963 (0.948) | 0.961 (0.925) | 0.943 (0.822) | 0.954 (0.872) |
| | | | (20,50) | 0.951 (0.852) | 0.951 (0.841) | 0.960 (0.857) | 0.930 (0.743) | 0.948 (0.792) |
| | | | (30,30) | 0.952 (0.737) | 0.957 (0.744) | 0.954 (0.727) | 0.941 (0.675) | 0.951 (0.700) |
| | | | (50,50) | 0.951 (0.544) | 0.959 (0.563) | 0.953 (0.546) | 0.947 (0.538) | 0.949 (0.540) |
| | | | (50,75) | 0.949 (0.523) | 0.953 (0.532) | 0.947 (0.520) | 0.939 (0.509) | 0.942 (0.512) |
| | | | (75,75) | 0.950 (0.440) | 0.955 (0.455) | 0.950 (0.438) | 0.949 (0.438) | 0.951 (0.440) |
| 7 | (2, 0.300) (6, 1.083) | 0 | (10,10) | 0.967 (1.521) | 0.961 (1.387) | 0.963 (1.414) | 0.920 (1.055) | 0.947 (1.219) |
| | | | (20,20) | 0.953 (0.897) | 0.954 (0.887) | 0.955 (0.871) | 0.929 (0.772) | 0.947 (0.823) |
| | | | (20,50) | 0.950 (0.827) | 0.950 (0.817) | 0.956 (0.834) | 0.922 (0.720) | 0.936 (0.769) |
| | | | (30,30) | 0.951 (0.699) | 0.953 (0.705) | 0.951 (0.690) | 0.936 (0.641) | 0.948 (0.664) |
| | | | (50,50) | 0.944 (0.515) | 0.951 (0.529) | 0.947 (0.517) | 0.940 (0.506) | 0.942 (0.509) |
| | | | (50,75) | 0.952 (0.502) | 0.951 (0.502) | 0.956 (0.502) | 0.948 (0.489) | 0.953 (0.491) |
| | | | (75,75) | 0.946 (0.413) | 0.949 (0.424) | 0.945 (0.411) | 0.945 (0.413) | 0.947 (0.415) |
| 19 | (5, 2.735) (0.5, 0.140) | 0.5 | (10,10) | 0.955 (4.470) | 0.944 (3.624) | 0.961 (3.282) | 0.880 (2.580) | 0.916 (3.020) |
| | | | (20,20) | 0.962 (2.458) | 0.939 (2.172) | 0.941 (1.975) | 0.920 (1.936) | 0.928 (2.044) |
| | | | (20,50) | 0.955 (2.407) | 0.942 (2.150) | 0.967 (2.083) | 0.931 (1.920) | 0.940 (2.028) |
| | | | (30,30) | 0.970 (1.932) | 0.952 (1.711) | 0.949 (1.601) | 0.938 (1.592) | 0.936 (1.657) |
| | | | (50,50) | 0.968 (1.443) | 0.949 (1.291) | 0.948 (1.238) | 0.941 (1.230) | 0.949 (1.280) |
| | | | (50,75) | 0.968 (1.415) | 0.944 (1.288) | 0.931 (1.205) | 0.934 (1.211) | 0.944 (1.262) |
| | | | (75,75) | 0.961 (1.101) | 0.951 (1.034) | 0.894 (0.942) | 0.944 (1.010) | 0.949 (1.029) |

(*Continued*)

**Table 3.** (Continued)

| Scenario* | $(\alpha_1, \beta_1)$ $(\alpha_2, \beta_2)$ | $\eta$ | Sample size | Coverage probability(Average length) | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | $G_C$ | $G_W$ | $G_K$ | PB | t-test |
| 20 | (5, 12.257) (1, 2.516) | 0.5 | (10,10) | 0.955 (2.472) | 0.947 (2.084) | 0.956 (2.091) | 0.891 (1.545) | 0.928 (1.796) |
| | | | (20,20) | 0.963 (1.401) | 0.954 (1.318) | 0.961 (1.302) | 0.933 (1.160) | 0.946 (1.234) |
| | | | (20,50) | 0.958 (1.341) | 0.949 (1.258) | 0.965 (1.307) | 0.932 (1.120) | 0.948 (1.199) |
| | | | (30,30) | 0.957 (1.083) | 0.953 (1.043) | 0.959 (1.039) | 0.935 (0.952) | 0.946 (0.997) |
| | | | (50,50) | 0.953 (0.816) | 0.944 (0.786) | 0.954 (0.787) | 0.935 (0.743) | 0.942 (0.764) |
| | | | (50,75) | 0.957 (0.805) | 0.944 (0.767) | 0.955 (0.775) | 0.941 (0.730) | 0.946 (0.752) |
| | | | (75,75) | 0.962 (0.634) | 0.959 (0.630) | 0.952 (0.607) | 0.958 (0.615) | 0.954 (0.620) |
| 25 | (5, 1.5) (3, 1.5) | 0.583 | (10,10) | 0.957 (1.243) | 0.957 (1.174) | 0.956 (1.168) | 0.906 (0.906) | 0.942 (1.030) |
| | | | (20,20) | 0.956 (0.756) | 0.963 (0.769) | 0.960 (0.747) | 0.940 (0.670) | 0.951 (0.709) |
| | | | (20,50) | 0.961 (0.667) | 0.961 (0.669) | 0.963 (0.673) | 0.934 (0.594) | 0.949 (0.629) |
| | | | (30,30) | 0.950 (0.593) | 0.958 (0.610) | 0.951 (0.594) | 0.932 (0.553) | 0.944 (0.574) |
| | | | (50,50) | 0.948 (0.438) | 0.959 (0.459) | 0.949 (0.441) | 0.947 (0.435) | 0.950 (0.438) |
| | | | (50,75) | 0.951 (0.420) | 0.950 (0.427) | 0.950 (0.417) | 0.948 (0.409) | 0.949 (0.414) |
| | | | (75,75) | 0.943 (0.355) | 0.952 (0.371) | 0.941 (0.357) | 0.944 (0.356) | 0.944 (0.357) |
| 26 | (3, 5) (3, 10) | 0.693 | (10,10) | 0.966 (1.459) | 0.963 (1.356) | 0.962 (1.332) | 0.923 (1.033) | 0.953 (1.166) |
| | | | (20,20) | 0.954 (0.862) | 0.961 (0.873) | 0.957 (0.846) | 0.932 (0.759) | 0.952 (0.800) |
| | | | (20,50) | 0.958 (0.706) | 0.960 (0.707) | 0.961 (0.710) | 0.941 (0.634) | 0.955 (0.665) |
| | | | (30,30) | 0.953 (0.675) | 0.964 (0.692) | 0.958 (0.675) | 0.936 (0.622) | 0.952 (0.644) |
| | | | (50,50) | 0.946 (0.500) | 0.960 (0.523) | 0.948 (0.503) | 0.943 (0.489) | 0.949 (0.497) |
| | | | (50,75) | 0.950 (0.455) | 0.958 (0.474) | 0.953 (0.458) | 0.948 (0.448) | 0.955 (0.454) |
| | | | (75,75) | 0.942 (0.402) | 0.954 (0.421) | 0.949 (0.407) | 0.940 (0.398) | 0.949 (0.404) |

* Scenarios in Table 3 are subset of scenarios in Tables 1 and 2.

† The confidence interval for the WMW test is not applicable.

https://doi.org/10.1371/journal.pone.0314705.t003

test converges to nominal level as sample sizes increase. However, for some scenarios, it can be liberal when sample sizes are small, such as scenario 19 as sample sizes being less than (30, 30), and scenario 20 at (10, 10). In terms of the length of confidence intervals, the **PB** method appears to provide shortest confidence intervals among the proposed methods when sample sizes are small. However, this observation is due to the fact that the **PB** method is liberal at small sizes, hence it should not be interpreted. As sample sizes reach (50, 50), all four methods are generally comparable in terms of length.

In summary, generally we recommend the proposed $\mathbf{G_C}$ and $\mathbf{G_W}$ methods over $\mathbf{G_K}$, **PB** method, and two sample t-test, due to the fact that $\mathbf{G_C}$ and $\mathbf{G_W}$ methods maintain satisfactory coverage probabilities even at small sample sizes and when the shape parameters are small.

## 6 Data examples

In this section, we illustrate the proposed method using publicly accessible data from a recent study that measured mRNA expression and protein abundance at single cell level simultaneously by Hao et.al. [5]. In this study, peripheral blood mononuclear cell (PBMC) samples from eight volunteers were collected at pre (day 0) and post HIV vaccination (day 3 and 7), yielding a total of 210,911 cells. The CITE-seq method was used to simultaneously quantify RNA and surface protein abundance in single cells via the sequencing of antibody-derived tags (ADTs). Analyses identified 57 clusters of different types of cells, encapsulated all major and minor immune cell types and revealed striking cellular diversity.

For demonstration purpose without delving deeply into the biological details of immune cells functions, we focus on protein abundance data in the cluster of plasmacytoid dendritic Cell (pDC) cells. The pDC releases type 1 interferon in response to viral infection [47], thus could serve as an indicator of immune response to vaccination. Although pDC cell counts are usually low in PBMC samples, as shown in Hao's study, they may play a critical role in regulating gene expression and innate immune responses [48]. The data used in this manuscript were attached as S1 Table. Based on the physical model [1–4], it is reasonable to assume the measured protein abundances of individual cells follow Gamma distribution. This assumption is further verified by goodness-of-fit test [49] for Gamma distribution using method implemented in R goft package. The biological effects were estimated by comparing the log-transformed protein abundances.

In this section, different analyses are performed to investigate the protein abundance variation between donors and across time points within pDC cells. According to the simulation results in Section 5, the **PB** method is not suitable, as it yields inflated type I errors when sample sizes are small, which is very common for protein abundance data. Furthermore, $\mathbf{G}_K$, one of the methods based on generalized pivots, generates inaccurate results when the shape parameter is small, potentially leading to unreliable testing outcomes. Thus, we use two recommended testing approaches based on the generalized pivots (i.e.$\mathbf{G}_C$ and $\mathbf{G}_W$) on the log-transformed data. For comparison purpose, we also analyze data using two sample t-test and the WMW test, which are commonly used in the differential analysis of protein abundance data and other genomic studies. More details described in **Example 1** and **Example 2** below. To enable direct comparisons between different donors regardless of differences in sample size, we use relative counts (RC) of protein abundance. In the settings of this single-cell studies, the sample sizes refer to the counts of pDC cells, making our proposed methods ideal choices for modeling them.

**Example 1.** Comparison of log-transformed protein abundance data for two different donors at same time point.

Analyzing protein abundance data across different donors at given time points allows us to perceive variations in the immune response to vaccination among different individuals. Since the consistency of immune response is crucial for vaccine success, accurately assessing protein levels at fixed time related to vaccination is essential for evaluating its quality and effectiveness.

Table 4 lists summary statistics for four genes: Rat-IgG1–2 (donor P1 vs. P3 at day 7), CD3–2 (donor P3 vs. P8 at day 0), CD226 (donor P1 vs. P6 at day 0), and CD44–2 (donor P1 vs. P3 at day 3). The estimated $p$-values as well as confidence intervals by the proposed methods ($\mathbf{G}_C$ and $\mathbf{G}_W$) and two sample t-test and the WMW test are also presented. For these four proteins, the $\mathbf{G}_C$ and $\mathbf{G}_W$ methods yield different conclusions in terms of significance, in contrast to two sample t-test and the WMW test.

As observed in simulation studies, the two sample t-test is unreliable as the disparity between two distributions is large, especially when sample sizes are less than (50, 50). Moreover, the WMW test is very sensitive to difference between the shapes of distributions. For this data set, the sample sizes (counts of pDC cells) are generally small, and the data exhibit different distributions for different donors at same time point. Therefore, two sample t-test and the WMW test should not be trusted for their testing results.

For instance, when comparing the protein abundance of Rat-IgG1–2 between donor P1 and donor P3 at day 7, our proposed approaches ($\mathbf{G}_C$ and $\mathbf{G}_W$) reveal significant mean difference. However, two sample t-test and the WMW test fail to identify this difference. On the other hand, when examining the protein abundance of CD226 between donor P1 and donor P6 at day 0, our $\mathbf{G}_C$ and $\mathbf{G}_W$ methods indicate there are no significance between two donors, whereas two sample t-test and the WMW test state otherwise. These erroneous conclusions

**Table 4. Testing the equality of protein abundance data from different donors at the same time point ($p$-value and estimated confidence interval for mean difference).**

| Protein | Time | Donor 1<br>Donor 2 | $n_1$<br>$n_2$ | $(\hat{\alpha}_1, \hat{\beta}_1)$<br>$(\hat{\alpha}_2, \hat{\beta}_2)$ | $\hat{\delta}_1$<br>$\hat{\delta}_2$ | $\hat{\eta}$ | Methods | $p$-value | Est. CI*<br>(lower, upper) |
|---|---|---|---|---|---|---|---|---|---|
| Rat-IgG1–2 | 7 | P1 | 22 | (9.800, 0.366) | 3.236 | 0.339 | $G_C$ | 0.039 | (0.020, 0.717) |
| | | P3 | 17 | (3.440, 0.163) | 2.897 | | $G_W$ | 0.033 | (0.037, 0.719) |
| | | | | | | | t-test | 0.055 | (-0.007, 0.684) |
| | | | | | | | WMW | 0.092 | † |
| CD38–2 | 0 | P3 | 19 | (5.520, 0.075) | 4.212 | 0.524 | $G_C$ | 0.027 | (0.062, 1.290) |
| | | P8 | 12 | (2.220, 0.044) | 3.688 | | $G_W$ | 0.034 | (0.061, 1.190) |
| | | | | | | | t-test | 0.051 | (-0.003, 1.060) |
| | | | | | | | WMW | 0.064 | † |
| CD226 | 0 | P1 | 27 | (5.240, 0.190) | 3.219 | -0.379 | $G_C$ | 0.075 | (-0.737, 0.038) |
| | | P6 | 17 | (3.130, 0.072) | 3.598 | | $G_W$ | 0.065 | (-0.719, 0.021) |
| | | | | | | | t-test | 0.046 | (-0.755, -0.007) |
| | | | | | | | WMW | 0.018 | † |
| CD44–2 | 3 | P1 | 8 | (9.410, 0.104) | 4.451 | 0.324 | $G_C$ | 0.069 | (-0.034, 0.606) |
| | | P3 | 42 | (6.641, 0.099) | 4.127 | | $G_W$ | 0.067 | (-0.038, 0.620) |
| | | | | | | | t-test | 0.041 | (0.016, 0.637) |
| | | | | | | | WMW | 0.044 | † |

* Estimated confidence interval.

† The Est. CI for the WMW test is not applicable.

https://doi.org/10.1371/journal.pone.0314705.t004

based on t-test and the WMW test may lead us to mis-characterize the nature of vaccine response related to these genes.

Furthermore, the estimated 95% confidence intervals for the mean difference ($\eta$) are also presented in Table 4, and $\mathbf{G_C}$ and $\mathbf{G_K}$ methods generally have the comparable lengths.

**Example 2.** Comparison of log-transformed protein abundance data at two different times for same donor.

One important aspect of the study by Hao et al. [5] is to characterize the response to vaccination for each of previously identified cell types, with particular interests in identifying cell populations that contribute most strongly to the innate immune response. This response is expected to be highly activated at the first vaccinated time point (day 3), and subsequently dampen at the second time point (day 7), as observed with another non-replicating viral vectored HIV vaccine [50]. Therefore, donor with strong innate immune response to vaccination and the activated genes can be identified by comparing the gene expression at time 0 to 3 and 7, as described in Hao et al. [5] The ability to identification individual with strong innate immune response is critical in our understanding of antibody production related to vaccine and other factors. Such insight may also help us developing more personalized treatment. Finally, we would like to point out that, due to the nature of single cell experiments, the gene protein levels at different time points are extracted from independent sets of cells. Therefore, t-test and WMW test of two independent samples were used when we investigate the changes of means of log-transformed protein abundance between two times for a given donor in this example.

Table 5 lists summary statistics for three genes (CD48, CD45–1, and CD337) of donor P8 for day 0 vs. day 3, and day 0 vs. day 7. The estimated $p$-values as well as confidence intervals by the proposed methods ($\mathbf{G_C}$ and $\mathbf{G_W}$), and the commonly used two sample t-test and the

**Table 5. Testing the equality of protein abundance data from same donor at different time points ($p$-value and estimated confidence interval for mean difference).**

| Protein | Donor | Time 1<br>Time 2 | $n_1$<br>$n_2$ | $(\hat{\alpha}_1, \hat{\beta}_1)$<br>$(\hat{\alpha}_2, \hat{\beta}_2)$ | $\hat{\delta}_1$<br>$\hat{\delta}_2$ | $\hat{\eta}$ | Methods | $p$-value | Est. CI*<br>(lower, upper) |
|---|---|---|---|---|---|---|---|---|---|
| CD48 | P8 | 0 | 12 | (13.700, 0.080) | 5.112 | -0.190 | $G_C$ | 0.045 | (-0.411, -0.005) |
| | | 3 | 34 | (17.700, 0.086) | 5.302 | | $G_W$ | 0.042 | (-0.413, -0.006) |
| | | | | | | | t-test | 0.054 | (-0.392, 0.004) |
| | | | | | | | WMW | 0.097 | † |
| CD48 | P8 | 0 | 12 | (13.700, 0.080) | 5.112 | 0.040 | $G_C$ | 0.761 | (-0.207, 0.281) |
| | | 7 | 20 | (9.750, 0.058) | 5.072 | | $G_W$ | 0.702 | (-0.195, 0.268) |
| | | | | | | | t-test | 0.710 | (-0.187, 0.270) |
| | | | | | | | WMW | 0.526 | † |
| CD45–1 | P8 | 0 | 12 | (0.730, 0.174) | 0.611 | -0.914 | $G_C$ | 0.049 | (-3.100, -0.005) |
| | | 3 | 34 | (1.610, 0.249) | 1.525 | | $G_W$ | 0.037 | (-2.460, -0.071) |
| | | | | | | | t-test | 0.124 | (-2.120, 0.286) |
| | | | | | | | WMW | 0.101 | † |
| CD45–1 | P8 | 0 | 12 | (0.730, 0.174) | 0.611 | 0.139 | $G_C$ | 0.923 | (-2.150, 1.800) |
| | | 7 | 20 | (0.615, 0.141) | 0.472 | | $G_W$ | 0.900 | (-1.520, 1.480) |
| | | | | | | | t-test | 0.832 | (-1.260, 1.550) |
| | | | | | | | WMW | 0.953 | † |
| CD337 | P8 | 0 | 12 | (3.190, 0.608) | 1.493 | 0.860 | $G_C$ | 0.026 | (0.108, 1.870) |
| | | 3 | 34 | (0.648, 0.134) | 0.633 | | $G_W$ | 0.028 | (0.135, 1.670) |
| | | | | | | | t-test | 0.023 | (0.124, 1.590) |
| | | | | | | | WMW | 0.506 | † |
| CD337 | P8 | 0 | 12 | (3.190, 0.608) | 1.493 | 0.127 | $G_C$ | 0.699 | (-0.516, 0.831) |
| | | 7 | 20 | (1.510, 0.267) | 1.366 | | $G_W$ | 0.689 | (-0.493, 0.788) |
| | | | | | | | t-test | 0.685 | (-0.489, 0.735) |
| | | | | | | | WMW | 0.833 | † |

* Estimated confidence interval.

† The Est. CI for the WMW test is not applicable.

https://doi.org/10.1371/journal.pone.0314705.t005

WMW test, are presented. For these three genes, the proposed methods may or may not yield different conclusions in terms of significance, comparing to the two sample t-test and the WMW test.

For example, when comparing the protein abundance of CD48 between day 0 and 3 for donor P8, our proposed approaches ($\mathbf{G_C}$ and $\mathbf{G_W}$) identify significant mean difference in log-transformed samples. Furthermore, the immune response is dampens at day 7, and the $\mathbf{G_C}$ and $\mathbf{G_W}$ methods yield insignificant difference from day 0 to 7. This pattern aligns with the characteristics of innate immune response stated above. However, both two sample t-test and the WMW test fail to generate significant differences between day 0 and day 3, indicating that the changes of CD48 abundance do not fit the pattern. Similar patterns are observed for CD337 and CD45–1 by the $\mathbf{G_C}$ and $\mathbf{G_W}$ methods, while such discoveries would have been missed by either the WMW test or t-test.

Interestingly, genes CD48 [51], CD45–1 [52] and CD337 [48] are all playing important roles in human's immune system. The observed multiple protein abundance modifications in donor P8 may indicate a different innate immune response compared to other donors, which do not show the patterns of changes described above. This discovery may point to potential existence of minority subtypes with different responses to the vaccine. A closer examination of changes in immune-related gene profiles in response to the vaccine might have clinical value.

Furthermore, the estimated 95% confidence intervals for the mean difference ($\eta$) by proposed methods are presented in Table 5.

## 7 Summary and discussion

In genomics, the two sample t-test and the Wilcoxon-Mann-Whitney (WMW) test are commonly used to identify proteins that can differentiate between different experiment conditions, and the comparison is usually applied on log-transformed protein abundance data [39]. However, the protein abundance data could be modeled by gamma distribution [3, 53, 54], and the shape of protein abundance distribution needs to be taken into consideration in differential analysis [19].

In this paper, we demonstrated the inappropriateness of using two sample t-test and the WMW test for testing the equality of means of two log-transformed protein abundance samples. Several methods for two-sample hypothesis testing and confidence interval estimation for mean difference of two independent Exp-gamma distributions are proposed.

Through comprehensive simulation studies, we demonstrated that two proposed methods (i.e. $\mathbf{G_C}$ and $\mathbf{G_W}$) based on the concepts of generalized inference can have excellent type I error control for testing the equality of two Exp-gamma means. Additionally, these methods can provide satisfactory confidence intervals for the Exp-gamma mean difference, with consistent performance across different parameter settings and sample sizes. Furthermore, the $\mathbf{G_C}$ and $\mathbf{G_W}$ methods work well even when the data are highly skewed. On the other hands, the $\mathbf{G_K}$ method is not recommended when the shape parameter(s) are less than 0.5, as the cube root transformation is not accurate with small shape parameter(s). The **PB** method is not advised when sample sizes are small or when the shape parameter(s) are less than 0.5, because highly skewed data increases the risk of parameter estimation instability. The possible inflation of type-I error when using parametric bootstrap analysis have been observed by many researchers in different settings; e.g. *Golzarri-Arroyo et.al.* [55]. Hence, we advise practitioners use caution when applying parametric bootstrappjng method in practice.

We expect the proposed methods have broad applicability to differential analysis in genomics studies and other applied fields. The proposed approaches for hypothesis testing and confidence interval estimation are easy to implement and their running time of these methods is quite feasible on standard computer platforms.

The R program is available at request from Dr. Yan at li.yan@roswellpark.org.

## Supporting information

**S1 Table. Real data example.** The data used in Tables 4 and 5.
(XLSX)

**S1 Appendix. The characteristics of exponential-gamma (Exp-gamma) distribution.**
(PDF)

**S2 Appendix. Generalized pivots and generalized test variables.**
(PDF)

## Author Contributions

**Conceptualization:** Lili Tian, Li Yan.

**Data curation:** Jia Wang, Li Yan.

**Formal analysis:** Jia Wang, Li Yan.

**Methodology:** Jia Wang, Lili Tian, Li Yan.

**Software:** Jia Wang, Li Yan.

**Supervision:** Lili Tian, Li Yan.

**Writing – original draft:** Jia Wang, Li Yan.

**Writing – review & editing:** Jia Wang, Lili Tian, Li Yan.

## References


1. Friedman N, Cai L, Xie XS. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. Physical Review Letter. 2006; 97:168302. https://doi.org/10.1103/PhysRevLett.97.168302 PMID: 17155441
2. Shahrezaei V, Swain PS. Analytical distributions for stochastic gene expression. Proceedings of the National Academy of Sciences. 2008; 105(45):17256–17261. https://doi.org/10.1073/pnas.0803850105 PMID: 18988743
3. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, et al. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science. 2010; 329(5991):533–538. https://doi.org/10.1126/science.1188308 PMID: 20671182
4. Li GW, Xie XS. Central dogma at the single-molecule level in living cells. Nature (London). 2011; 475 (7356):308–315. https://doi.org/10.1038/nature10315 PMID: 21776076
5. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021; 184(13):3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048 PMID: 34062119
6. Xie H, Ding X. The intriguing landscape of single-cell protein analysis. Advanced Science. 2022; 9 (12):2105932. https://doi.org/10.1002/advs.202105932 PMID: 35199955
7. Kammers K, Cole RN, Tiengwe C, Ruczinski I. Detecting significant changes in protein abundance. EuPA Open Proteomics. 2015; 7(C):11–19. https://doi.org/10.1016/j.euprot.2015.02.002 PMID: 25821719
8. Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, et al. A gene expression database for the molecular pharmacology of cancer. Nature Genetics. 2000; 24(3):236–244. https://doi.org/10.1038/73439 PMID: 10700175
9. Tauman R, Ivanenko A, O'Brien LM, Gozal D. Plasma C-reactive protein levels among children with sleep-disordered breathing. Pediatrics. 2004; 113(6):e564–e569. https://doi.org/10.1542/peds.113.6.e564 PMID: 15173538
10. Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP. Statistical analysis of membrane proteome expression changes in saccharomyces cerevisiae. Journal of Proteome Research. 2006; 5(9):2339–2347. https://doi.org/10.1021/pr060161n PMID: 16944946
11. Taylor SC, Nadeau K, Abbasi M, Lachance C, Nguyen M, Fenrich J. The ultimate qPCR experiment: producing publication quality, reproducible data the first time. Trends in Biotechnology (Regular ed). 2019; 37(7):761–774. https://doi.org/10.1016/j.tibtech.2018.12.002 PMID: 30654913
12. Thomas JG, Olson JM, Tapscott SJ, Zhao LP. An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. Genome Research. 2001; 11(7):1227–1236. https://doi.org/10.1101/gr.165101 PMID: 11435405
13. Gontarz P, Fu S, Xing X, Liu S, Miao B, Bazylianska V, et al. Comparison of differential accessibility analysis strategies for ATAC-seq data. Scientific Reports. 2020; 10(1):10150. https://doi.org/10.1038/s41598-020-66998-4 PMID: 32576878
14. Li Y, Ge X, Peng F, Li W, Li JJ. Exaggerated false positives by popular differential expression methods when analyzing human population samples. Genome Biology. 2022; 23(1):79. https://doi.org/10.1186/s13059-022-02648-4 PMID: 35292087
15. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biology. 2014; 15(2):R29. https://doi.org/10.1186/gb-2014-15-2-r29 PMID: 24485249
16. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014; 15(12):550. https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281

17. Fay MP, Proschan MA. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. Statistics Surveys. 2010; 4:1. https://doi.org/10.1214/09-SS051 PMID: 20414472

18. Pratt JW. Robustness of some procedures for the two-sample location problem. Journal of the American Statistical Association. 1964; 59(307):665–680. https://doi.org/10.1080/01621459.1964.10480721

19. de Torrenté L, Zimmerman S, Suzuki M, Christopeit M, Greally JM, Mar JC. The shape of gene expression distributions matter: how incorporating distribution shape improves the interpretation of cancer transcriptomic data. BMC Bioinformatics. 2020; 21(21):1–18. https://doi.org/10.1186/s12859-020-03892-w PMID: 33371881

20. Fraser DAS, Reid N, Wong A. Simple and accurate inference for the mean of the gamma model. Canadian Journal of Statistics. 1997; 25(1):91–99. https://doi.org/10.2307/3315359

21. Krishnamoorthy K, León-Novelo L. Small sample inference for gamma parameters: one-sample and two-sample problems. Environmetrics (London, Ont). 2014; 25(2):107–126. https://doi.org/10.1002/env.2261

22. Chen P, Ye ZS. Approximate statistical limits for a gamma distribution. Journal of Quality Technology. 2017; 49(1):64–77. https://doi.org/10.1080/00224065.2017.11918185

23. Chen P, Ye ZS. Estimation of field reliability based on aggregate lifetime data. Technometrics. 2017; 59 (1):115–125. https://doi.org/10.1080/00401706.2015.1096827

24. Wang BX, Wu F. Inference on the Gamma distribution. Technometrics. 2018; 60(2):235–244. https://doi.org/10.1080/00401706.2017.1328377

25. Krishnamoorthy K, Wang XG. Fiducial confidence limits and prediction limits for a Gamma distribution: censored and uncensored cases. Environmetrics. 2016; 27:479–493. https://doi.org/10.1002/env.2408

26. Krishnamoorthy K, Mathew T, Mukherjee S. Normal–Based methods for a Gamma distribution: prediction and tolerance intervals and stress-strength reliability. Technometrics. 2008; 50(1):69–78. https://doi.org/10.1198/004017007000000353

27. Wang X, Zou C, Yi L, Wang J, Li X. Fiducial inference for gamma distributions: two-sample problems. Communications in Statistics—Simulation and Computation. 2021; 50(3):811–821. https://doi.org/10.1080/03610918.2019.1568471

28. Wang X, Li X, Zhang L, Liu Z, Li M. Fiducial inference on gamma distributions: two-sample problems with multiple detection limits. Environmental and Ecological Statistics. 2022; 29(3):453–475. https://doi.org/10.1007/s10651-022-00528-5

29. Gao Y, Tian L. Confidence interval estimation for the difference and ratio of the means of two gamma distributions. Communications in Statistics—Simulation and Computation. 2022; 0(0):1–14. https://doi.org/10.1080/03610918.2022.2116646

30. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. Cell. 2008; 133(3):523–536. https://doi.org/10.1016/j.cell.2008.03.029 PMID: 18423832

31. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. Genome Biology. 2016; 17(1):13. https://doi.org/10.1186/s13059-016-0881-8 PMID: 26813401

32. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. Nature. 2011; 473(7347):337–342. https://doi.org/10.1038/nature10098 PMID: 21593866

33. Wang J, Huang M, Torre E, Dueck H, Shaffer S, Murray J, et al. Gene expression distribution deconvolution in single-cell RNA sequencing. Proceedings of the National Academy of Sciences. 2018; 115(28): E6437–E6446. https://doi.org/10.1073/pnas.1721085115 PMID: 29946020

34. Qin LX, Tuschl T, Singer S. Empirical insights into the stochasticity of small RNA sequencing. Scientific Reports. 2016; 6(1):24061–24061. https://doi.org/10.1038/srep24061 PMID: 27052356

35. Stopka SA, Khattar R, Agtuca BJ, Anderton CR, Pasa-Tolic L, Stacey G, et al.; WA (United States) Pacific Northwest National Lab. (PNNL). Metabolic noise and distinct subpopulations observed by single cell LAESI mass spectrometry of plant cells in situ. Frontiers in Plant Science. 2018; 9:1646–1646. https://doi.org/10.3389/fpls.2018.01646 PMID: 30498504

36. Cappellato M, Baruzzo G, Di Camillo B. Investigating differential abundance methods in microbiome data: A benchmark study. PLOS Computational Biology. 2022; 18:e1010467. https://doi.org/10.1371/journal.pcbi.1010467 PMID: 36074761

37. Devore JL, Peck R. Statistics: The exploration and analysis of data. Taylor & Francis; 1997.

38. Pan W. A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. Bioinformatics. 2002; 18(4):546–554. https://doi.org/10.1093/bioinformatics/18.4.546 PMID: 12016052

39. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics* S. Molecular & Cellular Proteomics. 2005; 4(10):1487–1502. https://doi.org/10.1074/mcp.M500084-MCP200

40. Tsui KW, Weerahandi S. Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. Journal of the American Statistical Association. 1989; 84(406):602–607. https://doi.org/10.2307/2289949

41. Weerahandi S. Generalized confidence intervals. Journal of the American Statistical Association. 1993; 88(423):899–905. https://doi.org/10.1080/01621459.1993.10476355

42. Weerahandi S. Exact statistical methods for data analysis. Springer-Verlag; 1995.

43. Tian L, Cappelleri JC. A new approach for interval estimation and hypothesis testing of a certain intraclass correlation coefficient: the generalized variable method. Statistics in Medicine. 2004; 23 (13):2125–2135. https://doi.org/10.1002/sim.1782 PMID: 15211607

44. Lin SH, Lee JC, Wang RS. Generalized inferences on the common mean vector of several multivariate normal populations. Journal of Statistical Planning and Inference. 2007; 137(7):2240–2249. https://doi.org/10.1016/j.jspi.2006.07.005

45. Lai CY, Tian L, Schisterman EF. Exact confidence interval estimation for the Youden index and its corresponding optimal cut-point. Computational Statistics and Data Analysis. 2012; 56(5):1103–1114. https://doi.org/10.1016/j.csda.2010.11.023 PMID: 27099407

46. Yan L. Confidence interval estimation of the common mean of several gamma populations. PloS One. 2022; 17(6):1–13. https://doi.org/10.1371/journal.pone.0269971 PMID: 35714130

47. Collin M, McGovern N, Haniffa M. Human dendritic cell subsets. Immunology. 2013; 140(1):22–30. https://doi.org/10.1111/imm.12117 PMID: 23621371

48. Xi Y, Troy NM, Anderson D, Pena OM, Lynch JP, Phipps S, et al. Critical role of plasmacytoid dendritic cells in regulating gene expression and innate immune responses to human rhinovirus-16. Frontiers in Immunology. 2017; 8:1351. https://doi.org/10.3389/fimmu.2017.01351 PMID: 29118754

49. Villaseñor JA, González-Estrada E. A variance ratio test of fit for Gamma distributions. Statistics & Probability Letters. 2015; 96:281–286. https://doi.org/10.1016/j.spl.2014.10.001

50. Zak DE, Andersen-Nissen E, Peterson ER, Sato A, Hamilton MK, Borgerding J, et al. Merck Ad5/HIV induces broad innate immune activation that predicts CD8+ T-cell responses but is attenuated by preexisting Ad5 immunity. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109(50):E3503–E3512. https://doi.org/10.1073/pnas.1208972109 PMID: 23151505

51. McArdel S, Terhorst C, Sharpe A. Roles of CD48 in regulating immunity and tolerance. Clinical Immunology. 2016; 164. https://doi.org/10.1016/j.clim.2016.01.008 PMID: 26794910

52. Donovan JA, Koretzky GA. CD45 and the immune response. Journal of the American Society of Nephrology. 1993; 4(4):976–985. https://doi.org/10.1681/ASN.V44976 PMID: 8286719

53. Fisher RA, Corbet AS, Williams CB. The relation between the number of species and the number of individuals in a random sample of an animal population. The Journal of Animal Ecology. 1943; p. 42–58. https://doi.org/10.2307/1411

54. Koziol J, Griffin N, Long F, Li Y, Latterich M, Schnitzer J. On protein abundance distributions in complex mixtures. Proteome Science. 2013; 11:1–9. https://doi.org/10.1186/1477-5956-11-5 PMID: 23360617

55. Golzarri-Arroyo L, Dickinson SL, Jamshidi-Naeini Y, Zoh RS, Brown AW, Owora AH, et al. Evaluation of the type I error rate when using parametric bootstrap analysis of a cluster randomized controlled trial with binary outcomes and a small number of clusters. Computer Methods and Programs in Biomedicine. 2022; 215:106654. https://doi.org/10.1016/j.cmpb.2022.106654 PMID: 35093646