UW Biostatistics Working Paper Series

9-14-2004

Nonparametric Confidence Intervals for the One- and Two-Sample Problems

Xiao-Hua Zhou
University of Washington, azhou@u.washington.edu

Phillip Dinh
University of Washington, ispvd@u.washington.edu

1. Introduction

1.1. Motivating Example

Researchers are often interested in comparing the difference of some measures between two groups, e.g., drug effect between treatment and control groups, a health outcome between intervention A and intervention B. For health services researchers, interest is also on the comparison of cost between two groups, e.g., cost incurs from diagnostic testing between depressed patients and non-depressed patients. Diagnostic testing is a costly and discretionary practice that is largely driven by the physician's judgments and patient's demands; some patients may equate quality of care with the intensity and novelty of diagnostic testing. The overuse of diagnostic testing could lead to inappropriately high diagnostic charges among older adults with depression and ill-defined symptoms (Callahan et al., 1997). One question of interest from Callahan's study is to compare medical charges between depressed and non-depressed patients. The focus of the statistical analysis is on the mean of diagnostic charges because the mean can be used to recover the total charge, which reflects the entire diagnostic expenditure in a given patient population.

We have patients level data from this study. Summary statistics of the two samples are presented in table 1. It can be seen from the table that the two samples are highly skewed with skewness coefficient 5.41 and 2.55. The 95% confidence interval for the difference in means based on the t-statistic is (-552.37, 1156.27) (interval width 1708.64) and based on bootstrap-t is (-338.57, 1476.24) (interval width 1864.81). Given that the two samples are highly skewed, one could ask whether the two above confidence intervals cover the true parameters at the specified level and that they are as narrow as possible. In the remaining of this paper, we will try to answer this question.

1.2. Existing Methods

Let $X_1, \ldots, X_n$ be an i.i.d. sample from a population with mean M and variance V. The commonly used interval for M is based on the one-sample t-statistic, proposed by "Student" (1908) and is given by

$$t = \frac{\bar{X} - M}{S/\sqrt{n}} \tag{1.1}$$

where $\bar{X} = \sum_{i=1}^{n} X_i/n$, and $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$.

The corresponding t-statistic based (t-based) confidence interval for the mean M is

$$\left(\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right) \tag{1.2}$$

and for large sample, the corresponding confidence interval based on central limit theorem (CLT) is

$$\left(\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right) \tag{1.3}$$

It is well known that the above interval has exact $1-\alpha$ coverage when the data come from a normal distribution and approximate $1-\alpha$ coverage for nonnormal data. Several authors have investigated the effect of skewness and sample size on the coverage accuracy of the above interval. These include, among many others, Gayen (1949), Barrett and Goldsmith (1976), Johnson (1978),

Table 1. *Descriptive statistics for the data set*

Group	n	mean	std. dev.	skewness coef.	$\hat{A}_m$ coef.	$\hat{A}_m/\sqrt{N}$
Non-depressed	108	1646.53	4103.84	5.41	5.52	0.38
Depressed	103	1344.58	1785.54	2.55		

All units are in U.S. dollars

Chen (1995), Boos and Hughes-Oliver (2000). They found that the coverage accuracy of the t-interval: (1) can be poor with skewed data; (2) depends on the magnitude of the population skewness; and (3) improves with increasing n (Boos and Hughes-Oliver, 2000).

When dealing with skewed data, several nonparametric solutions have been proposed for testing the mean of a distribution. The first relies on asymptotic results providing that the sample size n is sufficiently large. The central limit theorem (CLT) states that for a random sample from a distribution with mean M and finite variance V, the distribution of the sample mean $\bar{X}$ is approximately normal with mean M and variance V/n for sufficiently large n. This theorem can be used to justify the confidence intervals (1.3). The second approach is to transform the observed data. The logarithm is typically used. Inferences then will be made on the mean of the transformed data. The third approach is to use standard nonparametric methods like the Wilcoxon test.

Similarly, for the two-sample case, the ordinary-t statistic is given by

$$T = \frac{\bar{Y}_1 - \bar{Y}_2 - (M_1 - M_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}. \tag{1.4}$$

The corresponding t-based confidence interval for $M_1 - M_2$ is

$$\left(\bar{Y}_1 - \bar{Y}_2 - t_{\alpha/2,\nu}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, \bar{Y}_1 - \bar{Y}_2 + t_{\alpha/2,\nu}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right), \tag{1.5}$$

and for large samples, the corresponding CLT based confidence interval for $M_1 - M_2$ is

$$\left(\bar{Y}_1 - \bar{Y}_2 - z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, \bar{Y}_1 - \bar{Y}_2 + z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right) \tag{1.6}$$

where M_1 and M_2 are the population means of the two samples, $\{Y_{11}, \ldots, Y_{1n_1}\}$ and $\{Y_{21}, \ldots, Y_{2n_2}\}$. Here $\bar{Y}_1$ and $\bar{Y}_2$ are their corresponding sample means, and S_1^2 and S_2^2 are their corresponding sample variances. The degree of freedom, ν, in the t-based confidence interval (1.5) can be approximated (see, for example, Scheffé, (1970)).

Similarly nonparametric approaches are also available for the two-sample case. The first approach involves the use of the CLT based on large sample theory to justify the confidence interval given in equation (1.6). The second approach involves transformation of observations to reduce the effect of skewness; inference then will be made on the means of transformed data. The third approach uses standard nonparametric methods like the Wilcoxon test.

1.3. Limitations of existing methods

Each of the aforementioned methods have their own weaknesses. The t-based approach is not very robust under extreme deviations from normality (Boos and Hughes-Oliver, 2000). For the two-sample problem, our simulations indicate that coverage of confidence intervals given in equation (1.5) depends on the relative skewness of the two samples, and may be different from the true coverage by a substantial amount.

The CLT does not give any indication on how large n has to be for approximations in equation (1.3) and (1.6) to be reasonable. How large n has to be depends on the skewness, and to less extent, the kurtosis of the distribution of the observations (Barrett and Goldsmith, 1976; Boos and Hughes-Oliver, 2000). Gayen (1949), citing Pearson's work, stated that "the effect of universal 'excess' and of 'skewness' on 'Student's' ratio z (which is related to t by t = $z\sqrt{n-1}$) may be considerable." (Gayen, 1949, p.353).

The transformation of observations approach can be inappropriate since testing the mean (for the one-sample problem) and difference in means (for the two-sample problem) on transformed-scale is not always equivalent to testing on the original scale (Zhou et al., 1997).

The standard nonparametric Wilcoxon test is not the test for means. For one sample, Wilcoxon test can be used as a test for median. For two sample, the Wilcoxon test is a test for equality of distributions, and is not the test for equality of means unless the two distributions have the same shapes. In addition, it is not easy to construct confidence intervals based on the Wilcoxon test.

1.4. Proposed methods

Another approach is to modify the t-statistic to remove the effect of skewness. The method is based on the Edgeworth expansion (Hall, 1992a). For one sample, this method has been investigated by Johnson (1978), Hall (1992b), and Chen (1995). They showed that when the sample size is small and the parent distribution is asymmetrical, the t-statistic should be replaced by (Johnson, 1978; Chen, 1995):

$$t_1 = \left\{ (\bar{X} - M) + \frac{\hat{\mu}_3}{6nS^2} + \frac{\hat{\mu}_3}{3S^4}(\bar{X} - M)^2 \right\} (S^2/n)^{-1/2}$$

where $\hat{\mu}_3$ is an estimate of the population third central moment. This is the approach that we will pursue in this paper.

The remaining of this paper will be organized as following: in Section 2, we will revisit the one-sample problem; in Section 3, we will derive an Edgeworth expansion for a two-sample t-statistic; in Section 4, we will demonstrate the method via a simulation study; in Section 5, we apply our method to existing cost data sets; in Section 6, we will summarize the methods and provide our recommendation.

2. One-sample Problem

Let $U = (\bar{X} - M)/S$. The distribution of a statistic U admits the Edgeworth expansion (Hall, 1992b),

$$P(n^{1/2}U \le x) = \Phi(x) + n^{-1/2}\gamma(ax^2 + b)\phi(x) + O(n^{-1}) \quad (2.1)$$

where a = 1/3 and b = 1/6, γ is the population skewness that needs to be estimated, and Φ and ϕ are the standard normal cumulative distribution function and density function. Hall (1992b) proposed two transformations:

$$T_1 = T_1(U) = U + a\hat{\gamma}U^2 + \frac{1}{3}a^2\hat{\gamma}^2U^3 + n^{-1}b\hat{\gamma} \quad (2.2)$$

$$T_2 = T_2(U) = (2an^{-1/2}\hat{\gamma})^{-1}\{exp(2an^{-1/2}\hat{\gamma}U) - 1\} + n^{-1}b\hat{\gamma} \tag{2.3}$$

Skewness can be thought of as produced by a reshaping function of a normal random variable that affects positive values differently from negative values. In addition, the appearance of skewness is often greater away from the median (Hoaglin, 1985). Therefore, to reduce skewness, we need to find a transformation with $T(U) \approx U$ for U near zero and T(0) = 0 (except for a shifting factor of $n^{-1}b\hat{\gamma}$). See Hoaglin (1985) for a more detailed discussion on this idea. Following this idea, we introduce a new, simpler transformation:

$$T_3 = T_3(U) = U + U^2 + \frac{1}{3}U^3 + n^{-1}b\hat{\gamma} \tag{2.4}$$

The $(1-\alpha)100\%$ confidence interval for the mean M is given by

$$\bar{X} - T_i^{-1}(n^{-1/2}\xi_{1-\alpha/2})S \leq M \leq \bar{X} - T_i^{-1}(n^{-1/2}\xi_{\alpha/2})S \tag{2.5}$$

where $\xi_\alpha = \Phi(\alpha)$ and $T_i^{-1}(\cdot)$, i=1, 2, 3, is the inverse function of $T_i(\cdot)$, can be solved analytically, and has the following expressions:

$$T_1^{-1}(t) = \left(\frac{1}{3}\hat{\gamma}\right)^{-1}\left\{1 + 3\frac{1}{3}\hat{\gamma}(t - n^{-1}\frac{1}{6}\hat{\gamma})\right\}^{1/3} - \left(\frac{1}{3}\hat{\gamma}\right)^{-1},$$
$$T_2^{-1}(t) = \left(2\frac{1}{3}n^{-1/2}\hat{\gamma}\right)^{-1}\log\left\{2\frac{1}{3}n^{-1/2}\hat{\gamma}\left(t - n^{-1}\frac{1}{6}\hat{\gamma}\right) + 1\right\},$$
$$T_3^{-1}(t) = \left\{1 + 3(t - n^{-1}\frac{1}{6}\hat{\gamma})\right\}^{1/3} - 1\cdot$$

The validity of the transformation method has been investigated by several authors (Hall (1992b), Zhou and Gao (2000)). Here we report a simulation study on the finite-sample performance of the transformation methods in comparison with other methods. Table 2 summarizes our results based on 10,000 simulations from 3 log-normal distribution LN(μ, σ^2) where μ and σ^2 are the mean and variance of the log-transformed observations, respectively. In the table, the ordinary-t interval is denoted by "Ord_t." The bootstrap-t percentile interval is denoted by "Boot_t." BCa denotes biased corrected acceleration confidence interval. $T_1(\hat{\gamma}), T_2(\hat{\gamma})$, and $T_3(\hat{\gamma})$ denote confidence intervals based on three transformations T_1, T_2, and T_3 given in equation (2.5), respectively. The details of the bootstrap methods are described in Efron and Tibshirani (1993). For the bootstrap resampling, we used 1,000 bootstrap samples for each generated data set. From the table, it can be seen that the bootstrap-t intervals give good results. Our method using T_3 transformation or Hall's T_1 transformation is comparable with the bootstrap-t interval and sometimes better, but requires less computing in term of bootstrap resampling. For sample size greater than 100, our interval based on T_3 transformation gives tighter coverage in term of average confidence interval length compared to the bootstrap-t interval and the transformed interval based on T_1. We also found that the ordinary-t interval is inadequate when the coeffcient $\hat{\gamma}/\sqrt{n}$ is greater than 0.3. Thus for data coming from highly skewed distribution and the sample size is relatively small ($\hat{\gamma}/\sqrt{n} \geq 0{\cdot}3$), confidence intervals based on T_1 or T_3 transformation or ones based on the bootstrap-t interval are recommended over the ordinary-t interval.

Table 2. *Coverage of 95% 2-sided confidence intervals for the mean of the log-normal $M = exp(\mu + \frac{1}{2}\sigma^2)$*

n	σ^2	$\frac{\hat{\gamma}}{\sqrt{n}}$	Ord_t	Boot_t	BCa	$T_1(\hat{\gamma})$	$T_2(\hat{\gamma})$	$T_3(\hat{\gamma})$
25	0.5	0.326	0.9186 (0.81)	0.9382 (0.97)	0.9067 (0.79)	0.9255 (1.07)	0.9121 (0.77)	0.9576 (1.84)
25	1.0	0.432	0.8823 (1.56)	0.9284 (2.30)	0.8865 (1.60)	0.9159 (2.51)	0.8813 (1.50)	0.9732 (3.59)
25	1.5	0.501	0.8380 (2.59)	0.9125 (5.03)	0.8583 (2.73)	0.9038 (4.56)	0.8459 (2.51)	0.9774 (5.97)
25	2.0	0.548	0.7933 (4.01)	0.8940 (9.84)	0.8310 (4.31)	0.8913 (7.35)	0.8042 (3.89)	0.9827 (9.25)
100	0.5	0.224	0.9368 (0.40)	0.9455 (0.43)	0.9333 (0.41)	0.9403 (0.44)	0.9368 (0.40)	0.9498 (0.43)
100	1.0	0.325	0.9180 (0.81)	0.9363 (0.95)	0.9201 (0.85)	0.9324 (1.08)	0.9193 (0.81)	0.9442 (0.88)
100	1.5	0.395	0.8919 (1.39)	0.9311 (1.83)	0.9070 (1.49)	0.9286 (2.13)	0.8965 (1.38)	0.9354 (1.51)
100	2.0	0.454	0.8749 (2.26)	0.9240 (3.47)	0.8955 (2.48)	0.9215 (3.86)	0.8815 (2.24)	0.9232 (2.46)
500	0.5	0.120	0.9496 (0.18)	0.9501 (0.18)	0.9475 (0.18)	0.9516 (0.18)	0.9501 (0.18)	0.9517 (0.18)
500	1.0	0.200	0.9380 (0.37)	0.9443 (0.39)	0.9371 (0.38)	0.9419 (0.40)	0.9387 (0.37)	0.9463 (0.38)
500	1.5	0.271	0.9282 (0.67)	0.9409 (0.74)	0.9284 (0.69)	0.9406 (0.81)	0.9298 (0.66)	0.9391 (0.68)
500	2.0	0.332	0.9118 (1.11)	0.9281 (1.34)	0.9144 (1.18)	0.9272 (1.53)	0.9135 (1.11)	0.9247 (1.13)

μ is chosen to be 0

$T_i(\hat{\gamma})$ denotes $T_i(\cdot)$ transformation intervals given in equation (2.5), for i = 1, 2, 3.

Values in the parenthesis are average confidence interval lengths

3. Edgeworth Expansion for the Two-sample t-Statistic

In this section we extend the three transformation methods T_1, T_2, and T_3 presented above to the two-sample problem. We show that the confidence interval based on the two-sample t-statistic can be modified to obtain better coverage when observations come from skewed distributions.

Let $Y_{11}, Y_{12}, \cdots, Y_{1n_1}$ and $Y_{21}, Y_{22}, \cdots, Y_{2n_2}$ be independently and identically distributed from some distributions F with mean M_1, variance V_1, skewness γ_1 and G with mean M_2, variance V_2, skewness γ_2, respectively. Let $\bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$ and $S_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2$ for i = 1, 2. We are interested in constructing confidence intervals for the difference $M_1 - M_2$.

Proposition 1. Let $\lambda_N = n_1/(n_1 + n_2) = n_1/N$. Assume $\lambda_N = \lambda + O(N^{-r})$ for some $r \geq 0$. Under regularity conditions (Hall, 1992a), the distribution of the t-statistic given in equation (1.4) has the following expansion,

$$P(T \leq x) = P(N^{1/2}U \leq x) = \Phi(x) + \frac{1}{\sqrt{N}}\frac{A}{6}(2x^2 + 1)\phi(x) + O(N^{-min(1,r+1/2)}) \qquad (3.1)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density function and cumulative distribution function of the standard normal variable, and

$$A = \left\{\frac{V_1}{\lambda} + \frac{V_2}{1-\lambda}\right\}^{-3/2}\left\{\frac{V_1^{3/2}\gamma_1}{\lambda^2} - \frac{V_2^{3/2}\gamma_2}{(1-\lambda)^2}\right\}$$

For a proof, see appendix A.

Similar to the one-sample case with a = 1/3, b = 1/6, and $\gamma = A$, we can define the three transformations T_i, i = 1, 2, 3, given by Equation (2.2), (2.3), and (2.4), respectively. Hence, we can derive three transformation-based confidence intervals for $M_1 - M_2$ as following: Let $\xi_\alpha = \Phi(\alpha)$ and $\hat{\sigma} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$, and the $(1-\alpha)100\%$ confidence interval for the difference $M_1 - M_2$ is given by:

$$\bar{Y}_1 - \bar{Y}_2 - N^{1/2} T_i^{-1}(N^{-1/2}\xi_{1-\alpha/2})\hat{\sigma} \leq M_1 - M_2 \leq \bar{Y}_1 - \bar{Y}_2 - N^{1/2} T_i^{-1}(N^{-1/2}\xi_{\alpha/2})\hat{\sigma}, \quad (3.2)$$

where $T_i^{-1}(t)$ is the inverse function of T_i, can be solved analytically, and has the following expressions:

$$T_1^{-1}(t) = \left(\frac{1}{3}\hat{A}\right)^{-1}\left\{1 + 3\frac{1}{3}\hat{A}(t - N^{-1}\frac{1}{6}\hat{A})\right\}^{1/3} - \left(\frac{1}{3}\hat{A}\right)^{-1},$$

$$T_2^{-1}(t) = \left(2\frac{1}{3}N^{-1/2}\hat{A}\right)^{-1}\log\left\{2\frac{1}{3}N^{-1/2}\hat{A}\left(t - N^{-1}\frac{1}{6}\hat{A}\right) + 1\right\},$$

$$T_3^{-1}(t) = \left\{1 + 3(t - N^{-1}\frac{1}{6}\hat{A})\right\}^{1/3} - 1\cdot$$

Here $\hat{A}$ is a moment estimator for the coefficient A and is defined as follows:

$$\hat{A} \equiv \hat{A}_m = \frac{(N/n_1)^2 S_1^3 \hat{\gamma}_1 - (N/n_2)^2 S_2^3 \hat{\gamma}_2}{\left\{(N/n_1)S_1^2 + (N/n_2)S_2^2\right\}^{3/2}}, \quad (3.3)$$

where, for $i = 1, 2$,

$$S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2, \hat{\gamma}_i = \frac{n_i}{(n_i-1)(n_i-2)}\sum_{j=1}^{n_i}\left\{\frac{Y_{ij} - \bar{Y}_i}{S_i}\right\}^3 \quad (3.4)$$

4. A Simulation Study

In this section, we conduct a simulation study to assess the coverage accuracy of two-sided confidence intervals given in section 3 for the difference in means of two positively skewed distributions. The two families that we considered are the log-normal family and the gamma family. To keep the sampling variation small, we used 10,000 simulated samples for each parameter setting and each sample size. For the bootstrap resampling, we used 1,000 bootstrap samples for each generated data set.

Table 3 and 4 summarize the design for our simulations. The two log-normal distributions are LN(μ_1, σ_1^2) and LN(μ_2, σ_2^2) where $\mu_1(\mu_2)$ and $\sigma_1^2(\sigma_2^2)$ are the mean and variance of the log-transformed sample 1 (2), accordingly. For convenience, we set $\mu_1 = \mu_2 = 0$. The gamma family G(s, r) has two parameters: shape (s) and rate (r). Its mean is given by s/r, variance is given by s/r^2. Of course, when s=1, it reduces to an exponential family, and when r = 1/2, it reduces to a χ^2 family.

Figure 1 and 2 summarize the distributions that we conduct for our simulations. Figure 1 has seven panels representing seven pairs of log-normal densities. In this figure, first panel represents simulation design L1a to L6a. The second panel is design L1b to L6b. The third panel is design

Table 3. *Setup parameters for the lognormal simulations*

Design	n_1	n_2	σ_1^2	σ_2^2	γ_1	γ_2
L1a	25	25	0.51	0.5	2.99	2.94
L1b	25	25	1.00	0.5	6.18	2.94
L1c	25	25	1.50	0.5	12.09	2.94
L1d	25	25	2.00	0.5	23.73	2.94
L1e	25	25	2.50	0.5	47.43	2.94
L1f	25	25	3.00	0.5	96.49	2.94
L2a	50	50	0.51	0.5	2.99	2.94
L2b	50	50	1.00	0.5	6.18	2.94
L2c	50	50	1.50	0.5	12.09	2.94
L2d	50	50	2.00	0.5	23.73	2.94
L2e	50	50	2.50	0.5	47.43	2.94
L2f	50	50	3.00	0.5	96.49	2.94
L3a	100	100	0.51	0.5	2.99	2.94
L3b	100	100	1.00	0.5	6.18	2.94
L3c	100	100	1.50	0.5	12.09	2.94
L3d	100	100	2.00	0.5	23.73	2.94
L3d	100	100	2.50	0.5	47.43	2.94
L3f	100	100	3.00	0.5	96.49	2.94
L4a	500	500	0.51	0.5	2.99	2.94
L4b	500	500	1.00	0.5	6.18	2.94
L4c	500	500	1.50	0.5	12.09	2.94
L4d	500	500	2.00	0.5	23.73	2.94
L4e	500	500	2.50	0.5	47.43	2.94
L4f	500	500	3.00	0.5	96.49	2.94
L5b	100	25	1.00	0.5	6.18	2.94
L5c	100	25	1.50	0.5	12.09	2.94
L5d	100	25	2.00	0.5	23.73	2.94
L5e	100	25	2.50	0.5	47.43	2.94
L5f	100	25	3.00	0.5	96.49	2.94
L6a	25	100	0.51	0.5	2.99	2.94
L6b	25	100	1.00	0.5	6.18	2.94
L6c	25	100	1.50	0.5	12.09	2.94
L6d	25	100	2.00	0.5	23.73	2.94
L6e	25	100	2.50	0.5	47.43	2.94
L6f	25	100	3.00	0.5	96.49	2.94
L7a	25	25	2.01	2.0	24.06	23.73
L7b	100	100	2.01	2.0	24.06	23.73
L7c	25	100	2.01	2.0	24.06	23.73

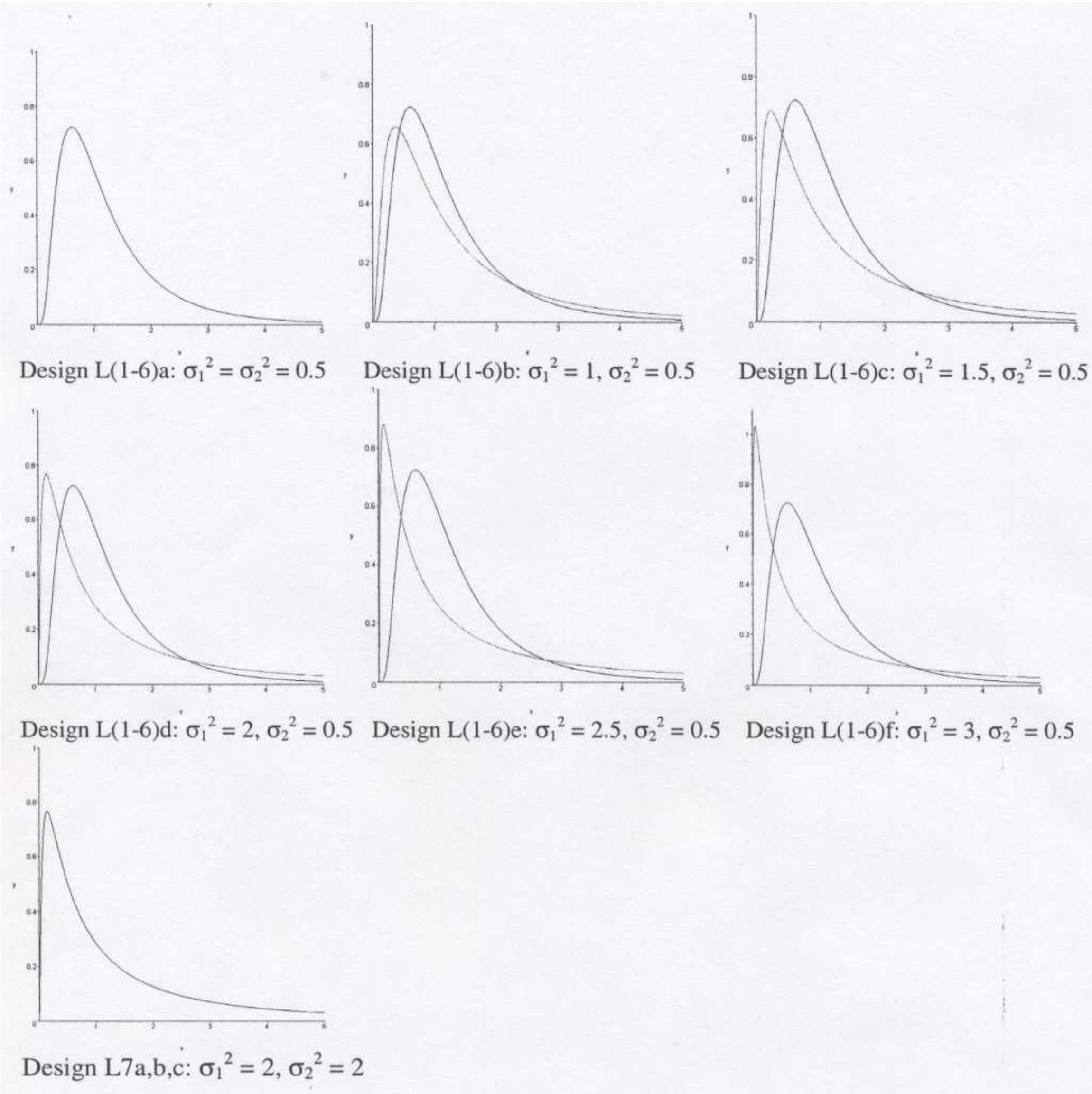

Fig. 1. Distribution of lognormal simulations

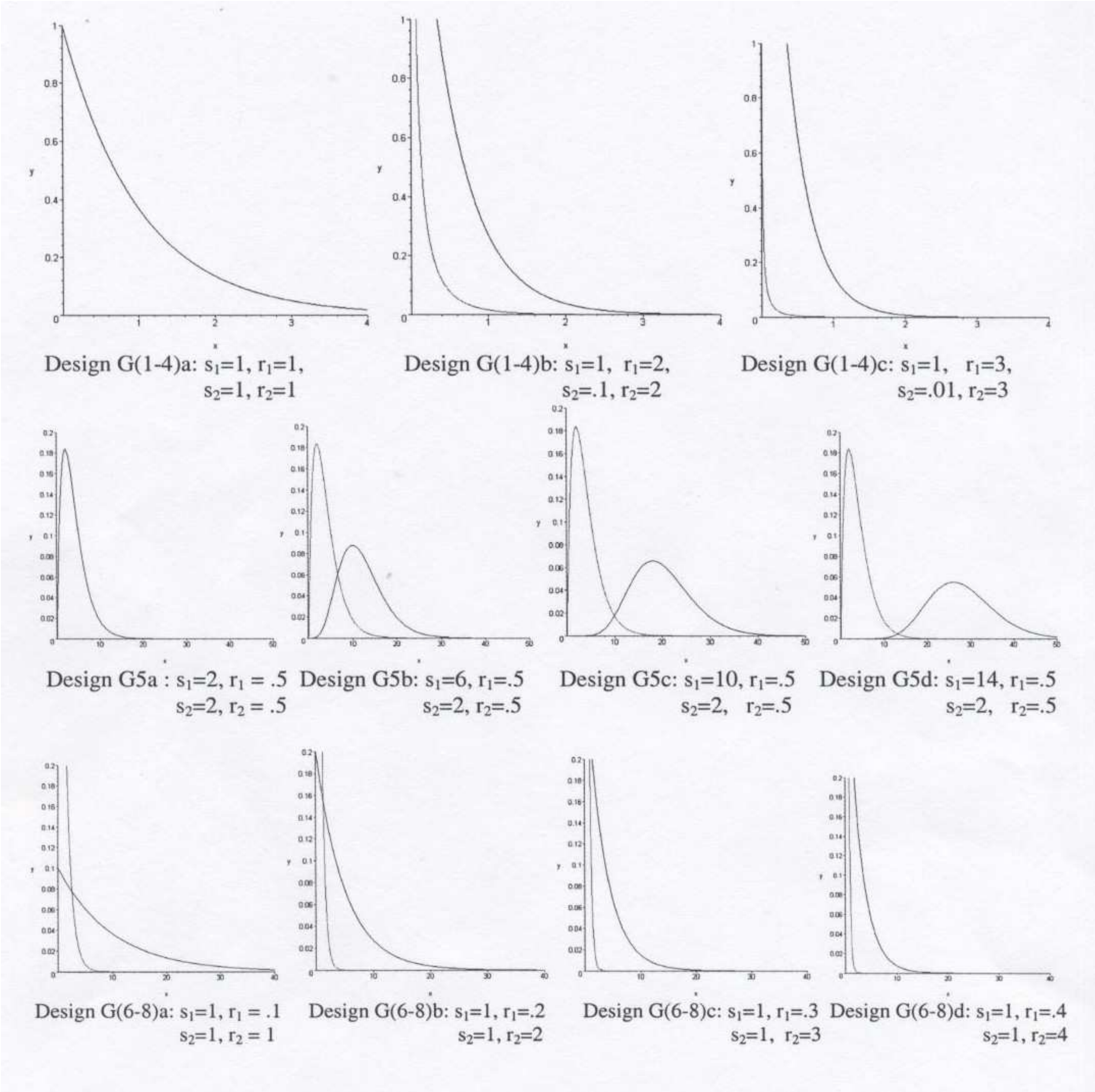

Fig. 2. Distribution of gamma simulations

Table 4. *Setup parameters for the Gamma simulations*

Design	n_1	n_2	s_1	r_1	s_2	r_2	γ_1	γ_2
G1a	25	25	1.01	1.0	1.00	1.0	1.99	2.00
G1b	25	25	1.00	2.0	0.10	2.0	2.00	6.33
G1c	25	25	1.00	3.0	0.01	3.0	2.00	20.00
G2a	50	50	1.01	1.0	1.00	1.0	1.99	2.00
G2b	50	50	1.00	2.0	0.10	2.0	2.00	6.33
G2c	50	50	1.00	3.0	0.01	3.0	2.00	20.00
G3a	100	100	1.01	1.0	1.00	1.0	1.99	2.00
G3b	100	100	1.00	2.0	0.10	2.0	2.00	6.33
G3c	100	100	1.00	3.0	0.01	3.0	2.00	20.00
G4a	25	100	1.01	1.0	1.00	1.0	1.99	2.00
G4b	25	100	1.00	2.0	0.10	2.0	2.00	6.33
G4c	25	100	1.00	3.0	0.01	3.0	2.00	20.00
G5a	25	25	2.10	0.5	2.00	0.5	1.38	1.41
G5b	25	25	6.00	0.5	2.00	0.5	0.82	1.41
G5c	25	25	10.00	0.5	2.00	0.5	0.63	1.41
G5d	25	25	14.00	0.5	2.00	0.5	0.54	1.41
G6a	25	25	1.00	0.1	1.00	1.0	2.00	2.00
G6b	25	25	1.00	0.2	1.00	2.0	2.00	2.00
G6c	25	25	1.00	0.3	1.00	3.0	2.00	2.00
G6d	25	25	1.00	0.4	1.00	4.0	2.00	2.00
G7a	50	50	1.00	0.1	1.00	1.0	2.00	2.00
G7b	50	50	1.00	0.2	1.00	2.0	2.00	2.00
G7c	50	50	1.00	0.3	1.00	3.0	2.00	2.00
G7d	50	50	1.00	0.4	1.00	4.0	2.00	2.00
G8a	25	50	1.00	0.1	1.00	1.0	2.00	2.00
G8b	25	50	1.00	0.2	1.00	2.0	2.00	2.00
G8c	25	50	1.00	0.3	1.00	3.0	2.00	2.00
G8d	25	50	1.00	0.4	1.00	4.0	2.00	2.00

L1c to L6c. The fourth panel is design L1d to L6d. The fifth panel is design L1e to L6e. The sixth panel is design L1f to L6f. The last panel is design L7a-L7c.

Figure 2 presents the gamma distribution for the simulation. In figure 2, the first panel is simulation design G1a to G4a. The second panel is design G1b to G4b. The third panel is design G1c to G4c. The next 4 panels are designs G5a-G5d (χ^2 case). The last 4 panels in figure 2 are designs G6a-G6d (exponential case). This setup is repeated for designs G7a-G7d and G8a-G8d where the sample sizes will change.

Table 5 summarizes our simulation results for the lognormal family. Values presented in the

Table 5. *Coverage of 95% 2-sided confidence intervals for $M_1 - M_2$ for lognormal family*

Design	$\frac{\hat{A}_m}{\sqrt{N}}$	Ord_t	Boot_t	BCa	$T_1(\hat{A}_m)$	$T_2(\hat{A}_m)$	$T_3(\hat{A}_m)$
L1a	0.005	0.9553 (1.15)	0.9242 (1.20)	0.9016 (1.12)	0.9120 (1.27)	0.9383 (1.12)	0.9342 (1.35)
L1b	0.233	0.9283 (1.77)	0.9124 (2.09)	0.8874 (1.77)	0.9011 (2.38)	0.9193 (1.72)	0.9634 (2.13)
L1c	0.377	0.8693 (2.77)	0.8899 (4.44)	0.8597 (2.86)	0.8805 (4.43)	0.8679 (2.68)	0.9483 (3.38)
L1d	0.462	0.8151 (4.04)	0.8641 (8.49)	0.8321 (4.26)	0.8586 (7.04)	0.8178 (3.90)	0.9121 (4.96)
L1e	0.531	0.7722 (6.07)	0.8536 (18.27)	0.8062 (6.55)	0.8511 (11.13)	0.7799 (5.87)	0.8821 (7.51)
L1f	0.573	0.7146 (8.43)	0.8273 (34.80)	0.7654 (9.24)	0.8375 (15.79)	0.7272 (8.15)	0.8346 (10.4)
L2a	0.004	0.9531 (0.81)	0.9307 (0.83)	0.9207 (0.81)	0.9271 (0.85)	0.9456 (0.80)	0.9445 (0.87)
L2b	0.226	0.9321 (1.27)	0.9182 (1.41)	0.9029 (1.29)	0.9123 (1.59)	0.9263 (1.25)	0.9499 (1.36)
L2c	0.366	0.8873 (2.01)	0.9069 (2.70)	0.8853 (2.12)	0.9013 (3.10)	0.8894 (1.97)	0.9326 (2.16)
L2d	0.455	0.8486 (3.12)	0.8907 (5.49)	0.8642 (3.40)	0.8871 (5.39)	0.8519 (3.07)	0.9011 (3.37)
L2e	0.510	0.8023 (4.63)	0.8706 (10.70)	0.8327 (5.18)	0.8728 (8.53)	0.8113 (4.55)	0.8598 (5.01)
L2f	0.552	0.7628 (6.87)	0.8639 (26.36)	0.8151 (7.80)	0.8692 (13.02)	0.7739 (6.77)	0.8307 (7.45)
L3a	0.006	0.9543 (0.58)	0.9373 (0.58)	0.9303 (0.57)	0.9350 (0.59)	0.9501 (0.57)	0.9489 (0.59)
L3b	0.204	0.9354 (0.91)	0.9265 (0.98)	0.9180 (0.93)	0.9245 (1.07)	0.9330 (0.90)	0.9465 (0.94)
L3c	0.336	0.9058 (1.47)	0.9221 (1.83)	0.9041 (1.56)	0.9195 (2.11)	0.9069 (1.46)	0.9321 (1.52)
L3d	0.418	0.8656 (2.30)	0.9075 (3.38)	0.8835 (2.51)	0.9062 (3.76)	0.8722 (2.28)	0.9010 (2.38)
L3e	0.481	0.8441 (3.54)	0.9010 (6.44)	0.8722 (3.97)	0.9000 (6.37)	0.8515 (3.51)	0.8805 (3.68)
L3f	0.520	0.8056 (5.18)	0.8873 (11.57)	0.8499 (5.90)	0.8909 (9.70)	0.8153 (5.15)	0.8475 (5.39)
L4a	0.003	0.9526 (0.26)	0.9487 (0.26)	0.9450 (0.26)	0.9477 (0.26)	0.9513 (0.26)	0.9536 (0.26)
L4b	0.135	0.9446 (0.42)	0.9435 (0.43)	0.9380 (0.42)	0.9428 (0.43)	0.9443 (0.42)	0.9501 (0.42)
L4c	0.240	0.9358 (0.69)	0.9394 (0.75)	0.9318 (0.71)	0.9381 (0.81)	0.9353 (0.69)	0.9418 (0.70)
L4d	0.314	0.9145 (1.12)	0.9300 (1.32)	0.9155 (1.19)	0.9281 (1.51)	0.9171 (1.12)	0.9238 (1.13)
L4e	0.373	0.8932 (1.78)	0.9237 (2.34)	0.9026 (1.93)	0.9218 (2.69)	0.8965 (1.78)	0.9076 (1.80)
L4f	0.418	0.8760 (2.78)	0.9182 (4.58)	0.8969 (3.10)	0.9192 (4.55)	0.8825 (2.78)	0.8923 (2.81)
L5b	0.010	0.9527 (1.15)	0.9247 (1.20)	0.9100 (1.15)	0.9196 (1.27)	0.9438 (1.13)	0.9465 (1.20)
L5c	0.204	0.9350 (1.64)	0.9171 (1.88)	0.9005 (1.69)	0.9114 (2.11)	0.9282 (1.63)	0.9514 (1.74)
L5d	0.331	0.8906 (2.42)	0.8978 (3.25)	0.8754 (2.59)	0.8926 (3.64)	0.8912 (2.40)	0.9279 (2.57)
L5e	0.425	0.8539 (3.60)	0.8931 (6.07)	0.8672 (3.98)	0.8902 (6.12)	0.8592 (3.58)	0.9046 (3.85)
L5f	0.489	0.8139 (5.27)	0.8773 (10.83)	0.8435 (5.98)	0.8772 (9.66)	0.8231 (5.24)	0.8698 (5.65)
L6a	0.197	0.9296 (0.91)	0.9195 (0.98)	0.9032 (0.89)	0.9126 (1.06)	0.9228 (0.88)	0.9416 (0.94)
L6b	0.363	0.8870 (1.60)	0.9105 (2.12)	0.8833 (1.62)	0.9011 (2.38)	0.8840 (1.54)	0.9187 (1.65)
L6c	0.460	0.8398 (2.61)	0.8952 (4.51)	0.8546 (2.72)	0.8876 (4.44)	0.8392 (2.50)	0.8845 (2.69)
L6d	0.525	0.7969 (4.00)	0.8784 (9.26)	0.8261 (4.25)	0.8775 (7.24)	0.7994 (3.83)	0.8444 (4.13)
L6e	0.571	0.7571 (5.98)	0.8636 (20.23)	0.8020 (6.47)	0.8691 (11.15)	0.7628 (5.73)	0.8125 (6.18)
L6f	0.602	0.7230 (8.55)	0.8528 (41.27)	0.7787 (9.37)	0.8630 (16.20)	0.7286 (8.20)	0.7779 (8.85)
L7a	0.010	0.9688 (6.17)	0.8931 (10.35)	0.8375 (6.54)	0.8690 (9.67)	0.9417 (6.00)	0.9346 (7.25)
L7b	0.009	0.9590 (3.37)	0.8975 (3.95)	0.8670 (3.62)	0.8888 (4.67)	0.9474 (3.35)	0.9453 (3.46)
L7c	0.198	0.9163 (4.88)	0.8645 (7.67)	0.8343 (5.14)	0.8569 (7.60)	0.9012 (4.75)	0.9470 (5.07)

$T_i(\hat{A}_m)$ denotes $T_i(\cdot)$ transformation intervals given in equation (3.2), for i = 1,2,3

Values in the parenthesis are average confidence interval lengths

Table 6. *Coverage of 95% 2-sided confidence intervals for $M_1 - M_2$ for Gamma family*

Design	$\frac{\hat{A}_m}{\sqrt{N}}$	Ord_t	Boot_t	BCa	$T_1(\hat{A}_m)$	$T_2(\hat{A}_m)$	$T_3(\hat{A}_m)$
G1a	-0.001	0.9523 (1.12)	0.9287 (1.15)	0.9125 (1.08)	0.9202 (1.14)	0.9378 (1.09)	0.9347 (1.32)
G1b	0.236	0.9337 (0.42)	0.9341 (0.46)	0.9120 (0.40)	0.9242 (0.48)	0.9242 (0.40)	0.9494 (0.50)
G1c	0.292	0.9252 (0.27)	0.9471 (0.30)	0.9172 (0.26)	0.9347 (0.32)	0.9187 (0.25)	0.9478 (0.32)
G2a	0.002	0.9508 (0.79)	0.9357 (0.80)	0.9290 (0.78)	0.9327 (0.79)	0.9442 (0.78)	0.9451 (0.84)
G2b	0.185	0.9405 (0.29)	0.9389 (0.31)	0.9268 (0.29)	0.9324 (0.31)	0.9350 (0.29)	0.9443 (0.31)
G2c	0.231	0.9331 (0.19)	0.9462 (0.20)	0.9301 (0.18)	0.9400 (0.20)	0.9304 (0.18)	0.9467 (0.20)
G3a	0.001	0.9486 (0.56)	0.9413 (0.56)	0.9362 (0.55)	0.9403 (0.56)	0.9457 (0.55)	0.9460 (0.57)
G3b	0.141	0.9440 (0.21)	0.9470 (0.21)	0.9390 (0.20)	0.9425 (0.21)	0.9433 (0.20)	0.9489 (0.21)
G3c	0.177	0.9435 (0.13)	0.9480 (0.14)	0.9400 (0.13)	0.9456 (0.14)	0.9422 (0.13)	0.9473 (0.14)
G4a	0.188	0.9342 (0.89)	0.9317 (0.94)	0.9128 (0.86)	0.9247 (0.95)	0.9264 (0.86)	0.9413 (0.92)
G4b	0.281	0.9242 (0.40)	0.9417 (0.45)	0.9134 (0.39)	0.9311 (0.47)	0.9185 (0.38)	0.9326 (0.41)
G4c	0.301	0.9201 (0.27)	0.9439 (0.31)	0.9157 (0.26)	0.9330 (0.32)	0.9131 (0.25)	0.9330 (0.27)
G5a	0.002	0.9536 (3.23)	0.9422 (3.27)	0.9260 (3.10)	0.9341 (3.19)	0.9441 (3.14)	0.9383 (3.81)
G5b	0.053	0.9532 (4.54)	0.9481 (4.60)	0.9342 (4.32)	0.9404 (4.45)	0.9444 (4.39)	0.9418 (5.35)
G5c	0.058	0.9481 (5.57)	0.9449 (5.66)	0.9281 (5.27)	0.9352 (5.43)	0.9375 (5.37)	0.9371 (6.54)
G5d	0.060	0.9487 (6.47)	0.9479 (6.57)	0.9308 (6.10)	0.9390 (6.29)	0.9395 (6.22)	0.9389 (7.57)
G6a	0.295	0.9267 (8.00)	0.9446 (9.12)	0.9174 (7.71)	0.9347 (9.52)	0.9212 (7.63)	0.9500 (9.50)
G6b	0.294	0.9259 (4.02)	0.9447 (4.58)	0.9147 (3.87)	0.9311 (4.78)	0.9194 (3.84)	0.9494 (4.77)
G6c	0.294	0.9224 (2.67)	0.9448 (3.04)	0.9176 (2.57)	0.9329 (3.18)	0.9176 (2.54)	0.9482 (3.16)
G6d	0.295	0.9209 (2.00)	0.9439 (2.28)	0.9153 (1.93)	0.9311 (2.39)	0.9171 (1.91)	0.9476 (2.37)
G7a	0.235	0.9373 (5.59)	0.9508 (5.99)	0.9353 (5.52)	0.9447 (5.98)	0.9347 (5.46)	0.9521 (5.95)
G7b	0.233	0.9367 (2.80)	0.9474 (3.00)	0.9343 (2.77)	0.9430 (2.99)	0.9342 (2.74)	0.9471 (2.98)
G7c	0.236	0.9317 (1.87)	0.9464 (2.01)	0.9285 (1.85)	0.9385 (2.01)	0.9298 (1.83)	0.9426 (1.99)
G7d	0.233	0.9345 (1.40)	0.9490 (1.50)	0.9334 (1.38)	0.9434 (1.49)	0.9319 (1.37)	0.9497 (1.49)
G8a	0.295	0.9246 (7.98)	0.9460 (9.08)	0.9218 (7.68)	0.9356 (9.43)	0.9172 (7.60)	0.9413 (8.58)
G8b	0.296	0.9206 (3.98)	0.9428 (4.54)	0.9148 (3.84)	0.9331 (4.73)	0.9123 (3.79)	0.9395 (4.29)
G8c	0.297	0.9215 (2.66)	0.9466 (3.03)	0.9187 (2.56)	0.9360 (3.17)	0.9173 (2.53)	0.9430 (2.86)
G8d	0.296	0.9247 (1.99)	0.9492 (2.27)	0.9186 (1.92)	0.9365 (2.36)	0.9185 (1.90)	0.9443 (2.15)

$T_i(\hat{A}_m)$ denotes $T_i(\cdot)$ transformation intervals given in equation (3.2), for i = 1,2,3

Values in the parenthesis are average confidence interval lengths table are confidence intervals based on the ordinary t statistic (denoted by Ord_t), the bootstrap-t interval (denoted by Boot_t), the bias-corrected accelerated confidence interval (BCa), the three transformation intervals (denoted by T_1, T_2, T_3). Values in the parenthesis are the average lengths of the corresponding intervals. Here we also saw that the bootstrap-t intervals give good results. The T_1 and T_3 transformation intervals also give consistent results. The T_1 intervals, in few cases, outperform T_3 intervals, while for other cases, the reverse is true. The ordinary-t intervals are certainly inadequate when the coefficient $\hat{A}_m/\sqrt{N}$ is large ($\geq 0{\cdot}3$). Here we also found that the intervals based on T_3 transformation gives tighter coverage in term of interval lengths compared to the bootstrap-t and T_1 intervals.

Table 6 shows our simulation results for the gamma family. Our simulation indicates that the ordinary-t intervals are relatively good. Similar to our observation previously, the ordinary-t intervals can be improved upon by the bootstrap-t, the T_1, or the T_3 intervals. The tightness of these intervals measured in term of interval lengths are relatively comparable. The ordinary-t intervals give very good coverage for the chi-square family that we considered in this simulation study and so are the bootstrap-t and the three transformation intervals. For the exponential family that we considered, the 95% ordinary-t intervals give coverage above 92% in all cases considered. However, they can be improved upon by using the bootstrap-t, the T_1, or the T_3 intervals.

It is clear from Proposition 1 (equation (3.1)) that the coefficient $A/\sqrt{N}$ (in absolute value) plays an important role in determining how good the normal approximation will be. In our simulation, when $\hat{A}_m/\sqrt{N}$ is small ($< 0{\cdot}3$), the ordinary t-interval will be quite satisfactory. On the contrary, when $\hat{A}_m/\sqrt{N} \geq 0{\cdot}3$, intervals based on bootstrap-t, T_1, or T_3 should be recommended. Our simulation also shows that skewness alone is not a big factor. It is the relative skewness that affects the ordinary t-interval. In fact, if both samples are skewed, but their relative skewness cancels each other and yields small coefficient A (like in design L7a and L7b), the ordinary-t interval is quite good.

In summary of our simulation, when dealing with data from skewed distributions, confidence intervals based on T_1 or T_3 transformation or ones based on the bootstrap-t interval are recommended over the ordinary-t interval. Intervals based on T_3 transformation have several advantages including tighter coverage compared to T_1 and the bootstrap-t intervals and require less computing than bootstrap-t intervals.

5. Application to a Cost Data

5.1. Medical charges in mental health patients

In this section we revisited the motivating application presented in the introductory section. As mentioned previously, we are interested in comparing the mean of diagnostic charges between depressed and non-depressed patients.

Figure 3 represents the histograms and the Q-Q Plots of the two samples. It is clear that both samples are positively skewed with the estimated coefficient $\hat{A}_m/\sqrt{N}$ of 0.38.

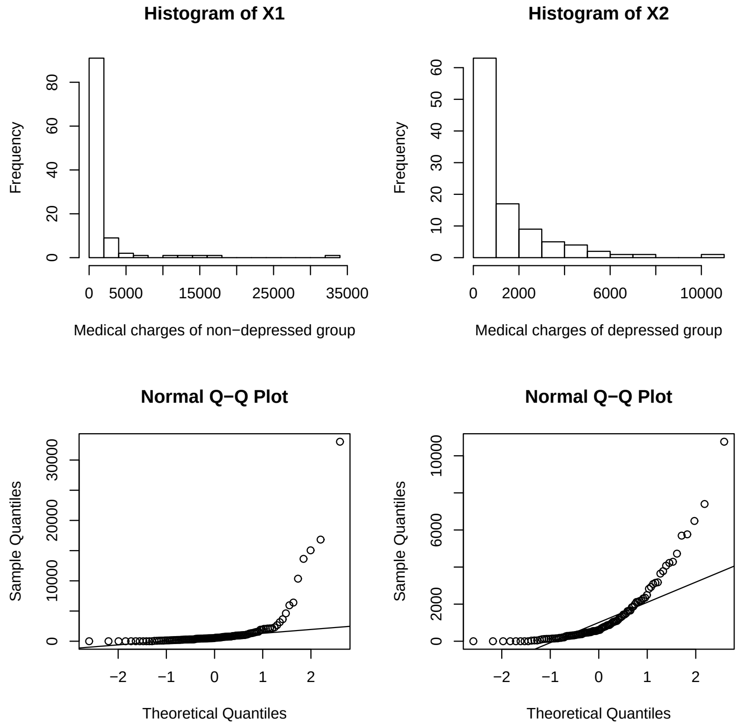

Fig. 3. Histograms and Q-Q Plots of the two samples

Table 7. *95% Confidence intervals for the difference in average costs between depressed and non-depressed groups*

	Interval	Interval length
Ordinary-t interval	(-552.37, 1156.27)	1708.64
T_1 interval	(-374.99, 1619.49)	1994.48
T_2 interval	(-504.75, 1192.41)	1697.16
T_3 interval	(-429.51, 1338.22)	1767.72
Bootstrap-t interval	(-388.57, 1476.24)	1864.81
BCA interval	(-338.64, 1593.15)	1931.79

All units are in U.S. dollars

The resulting confidence intervals for the difference in average medical charges between the depressed and non-depressed patients are given in table 7.

It can be seen that the T_1, T_3, and the bootstrap-t interval are relatively similar. T_2 interval resembles the ordinary-t interval the most. As anticipated, T_3 interval has shortest interval length compared to T_1 and the bootstrap-t intervals. All intervals include zero indicating that the difference in average costs between depressed and non-depressed patients are not statistically

Table 8. *Descriptive statistics for the data set*

Group	n	mean	std. dev.	skewness coef.	$\hat{A}_m$ coef.	$\hat{A}_m/\sqrt{N}$
Caucasian	66	2431.95	2188.58	1.16	3.03	0.31
others	28	4047.01	4351.32	1.95		

All units are in U.S. dollars

significant. Based on our simulation study, either T_1, T_3, or the bootstrap-t interval should be reported.

5.2. *Medical charges from computer-based interventions study*

In 2001, the Institute of Medicine documented the gap between recommended and actual practice of medicine in the United States. Many proven interventions were not routinely being used for common diseases that are morbid and costly. Reactive air ways diseases, asthma and chronic obstructive pulmonary disease (COPD), are examples. It has also been demonstrated that computer-based interventions can increase preventive care and reduce costs. Recently, Tierney et al. (2004) conducted a randomized, controlled trial to assess whether guideline-based care suggestions delivered via physicians' and pharmacists' computer workstations could improve the outpatient management and outcomes among patients with asthma or COPD. In this section, we reanalyze a subset of the real data set in Tierney's study to compare indirect health care charges between Caucasian patients and others. As in previous example, the focus will be on the average costs since the means can be used to recover the total charges. We are interested in comparing average costs between female Caucasian patients and female patients of other races with chronic obstructive pulmonary disease in this study.

Summary statistics of the two samples are presented in the table 8. Figure 4 represents the histograms and the Q-Q Plots of the two samples. It can be seen that both samples are positively skewed with the estimated coefficient $\hat{A}_m/\sqrt{N}$ of 0.31.

The resulting confidence intervals for the difference in average indirect medical charges between the Caucasian and other patients are given in table 9.

It can be seen that the T_1 and T_3 are very similar. The bootstrap-t interval is also relatively similar. The ordinary-t interval includes zero indicating the difference in average costs between the two groups is insignificant. However, all T_1, T_3, and bootstrap-t intervals indicate a significant difference between Caucasian patients and other patients with Caucasian patients having less average indirect costs. Based on our simulation, the conclusion of insigifincant difference based on the ordinary-t interval would be incorrect. Instead, the significant difference based on T_1, T_3, or the bootstrap-t interval should be reported.

6. Discussion

Our study shows that the coefficient $\gamma/\sqrt{n}$ (for the one-sample case) and coefficient $A/\sqrt{N}$ (for two-sample case) play an important role in the normal approximation for constructing confidence intervals. In our simulation study, we found that when $\hat{\gamma}/\sqrt{n}$ (respectively, $\hat{A}_m/\sqrt{N}$) is small ($< 0{\cdot}3$), confidence interval based on ordinary t is quite good. On the contrary, when $\hat{\gamma}/\sqrt{n}$ (respectively, $\hat{A}_m/\sqrt{N}$) is large ($\geq 0{\cdot}3$), the ordinary-t intervals can be improved upon by the bootstrap-t, T_1, or T_3 intervals. When dealing with confidence intervals for the means of skewed

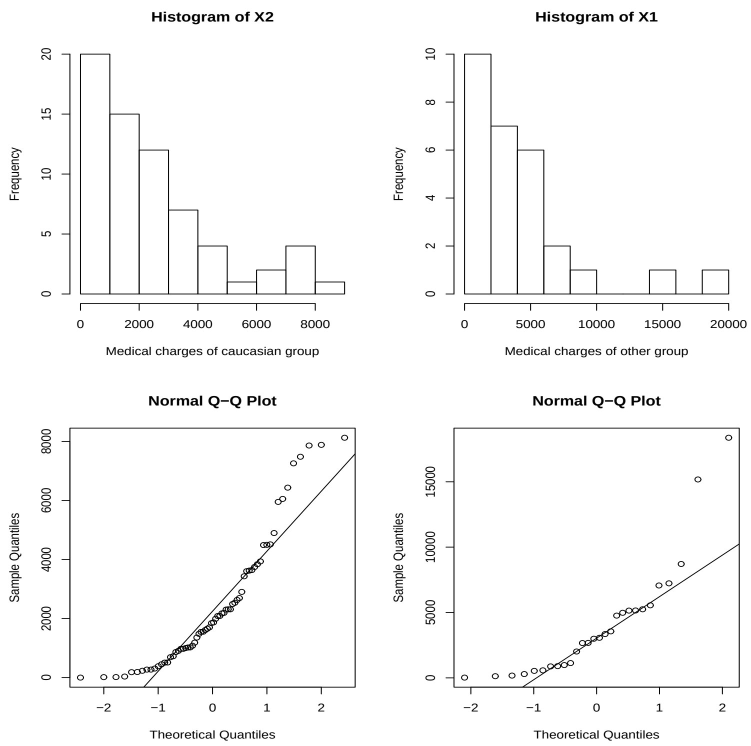

Fig. 4. Histograms and Q-Q plots of the two samples

Table 9. *95% Confidence intervals for the difference in average costs between Caucasian patients and other patients*

	Interval	Interval length
Ordinary-t interval	(-145.56, 3375.68)	3521.24
T_1 interval	(213.61, 3950.95)	3737.34
T_2 interval	(-2.99, 3394.83)	3397.82
T_3 interval	(211.86, 3941.67)	3729.81
Bootstrap-t interval	(175.94, 4082.93)	3906.99
BCA interval	(251.99, 3649.84)	3397.86

All units are in U.S. dollars

data, our simulations show that the bootstrap-t interval gives consistent and best coverage. Confidence intervals based on T_1 and T_3 transformations are comparable to the bootstrap-t intervals but require much less computing in term of bootstrap resampling. Among the bootstrap-t, the T_1, and the T_3 intervals, intervals based on T_3 transformation give tightest coverage measured in term of interval lengths, and should be recommended over the ordinary-t interval for skewed data. Standard textbook recommendation of sample size 30 is apparently inadequate for highly skewed data.

In our extensive simulation, we also found that our transformations intervals work best when coefficient A is positive. This won't be a problem in practice since we can always arrange the two samples to yield positive value of A.

A. Proof of Proposition 1

The two-sample t-statistic is given by:

$$T = \frac{\bar{Y_1} - \bar{Y_2} - (M_1 - M_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Let $Y_{ij}^* = \frac{Y_{ij} - M_i}{V_i^{1/2}}$, $\bar{Y}_i^* = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}^*$ and $S_i^{*2} = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(Y_{ij}^* - \bar{Y}_i^*)^2$, for i=1,2 and j=1,...,$n_i$. Then,

$$T = \frac{V_1^{1/2}\bar{Y}_1^* - V_2^{1/2}\bar{Y}_2^*}{\sqrt{\frac{V_1 S_1^{*2}}{n_1} + \frac{V_2 S_2^{*2}}{n_2}}} = \sqrt{N}\frac{V_1^{1/2}\bar{Y}_1^* - V_2^{1/2}\bar{Y}_2^*}{\sqrt{\frac{V_1 S_1^{*2}}{\lambda_N} + \frac{V_2 S_2^{*2}}{1-\lambda_N}}}$$

where $\lambda_N = n_1/N = n_1/(n_1 + n_2)$. Let $X \equiv (X_1, X_2, X_3, X_4)$ where

$$X_1 = \bar{Y}_1^*, X_2 = n_1^{-1}\sum_{j=1}^{n_1} Y_{1i}^{*2}, X_3 = \bar{Y}_2^*, X_4 = n_2^{-1}\sum_{j=1}^{n_2} Y_{2j}^{*2}$$

$$h(X) = \frac{V_1 S_1^{*2}}{\lambda_N} + \frac{V_2 S_2^{*2}}{1 - \lambda_N} = \frac{V_1}{\lambda_N}(X_2 - X_1^2) + \frac{V_2}{1 - \lambda_N}(X_4 - X_3^2)$$

$$g(X) = \frac{V_1^{1/2}X_1 - V_2^{1/2}X_3}{h(X)^{1/2}}$$

Then, $T = \sqrt{N}g(X)$.

By Taylor expansion, with $EX \equiv U \equiv (U_1, U_2, U_3, U_4) = (0, 1, 0, 1)$, we obtain

$$g(X) = g(U) + \frac{\partial g(U)}{\partial U}(X - U) + \frac{1}{2}\frac{\partial^2 g(U)}{\partial U^2}(X - U)^2 + \cdots$$

$$T = \sqrt{N}\left\{\frac{\partial g(U)}{\partial U}(X - U)' + \frac{1}{2}(X - U)'\frac{\partial^2 g(U)}{\partial U^2}(X - U) + \cdots\right\}$$

Note that $T = \sqrt{N}g(X)$ and g(U) = 0. Let

$$W_N = \sqrt{N}\left\{\frac{\partial g(U)}{\partial V}(X - U)' + \frac{1}{2}(X - U)'\frac{\partial^2 g(U)}{\partial U^2}(X - U)\right\}$$

We can show under some regularity conditions that

$$T = W_N + O(N^{-1})$$

If we assume $EY_{ij}^6 < \infty$, we can show that the first three moments of W_N are given as follows:

$$EW_n = -\frac{1}{2}AN^{-1/2} + O(N^{-min(1,r+1/2)}), EW_n^2 = 1 + O(N^{-1})$$
$$EW_n^3 = -\frac{7}{2}AN^{-1/2} + O(N^{-min(1,r+1/2)}),$$

where

$$A = h_0(V)^{-3/2}\left\{\frac{V_1^{3/2}\gamma_1}{\lambda^2} - \frac{V_2^{3/2}\gamma_2}{(1-\lambda)^2}\right\},$$

and

$$h_0(V) = \left\{\frac{V_1^2}{\lambda} + \frac{V_2^2}{(1-\lambda)}\right\}$$

Let K_{1N}, K_{2N}, K_{3N} be the first three cumulants of W_n. Then,

$$K_{1N} = -\frac{1}{2}AN^{-1/2} + O(N^{-min(1,r+1/2)})$$
$$K_{2N} = EW_n^2 - (EW_n)^2 = 1 + O(N^{-min(1,r+1/2)})$$
$$K_{3N} = E(W_n - EW_n)^3 = -2AN^{-1/2} + O(N^{-min(1,r+1/2)}).$$

Let $\chi_N(t)$ be the characteristic function of W_n. Then

$$\begin{aligned}\chi_N(t) &= exp\{K_{1N}(it) + K_{2N}\frac{(it)^2}{2} + K_{3N}\frac{(it)^3}{6} + \cdots\} \\ &= exp\{(-\frac{1}{2}AN^{-1/2})(it) - \frac{t^2}{2} + (-2A)\frac{(it)^3}{6}N^{-1/2} + O(N^{-min(1,r+1/2)})\} \\ &= exp(-\frac{t^2}{2})exp\left\{N^{-1/2}\left(-\frac{1}{2}A(it) - \frac{2A}{6}(it)^3\right) + O(N^{-min(1,r+1/2)})\right\}.\end{aligned}$$

By Taylor expansion, we obtain

$$\chi_N(t) = exp(-\frac{t^2}{2})\left\{1 + N^{-1/2}(-\frac{1}{2}A(it) - \frac{2A}{6}(it)^3) + O(N^{-min(1,r+1/2)})\right\}$$

Letting $r_1(it) = (-\frac{1}{2}A(it) - \frac{2A}{6}(it)^3)$, we can write

$$\chi_N(t) = exp(-\frac{t^2}{2})\left\{1 + N^{-1/2}r_1(it) + O(N^{-min(1,r+1/2)})\right\}(*)$$

Since $\chi_N(t) = \int_{-\infty}^{\infty} e^{itx}dp(W_n \le x)$ and $e^{-t^2/2} = \int_{-\infty}^{\infty} e^{itx}d\Phi(x)$, expression (*) suggests that

$$P(W_n \le x) = \Phi(x) + N^{-1/2}R_1(X) + O(N^{-min(1,r+1/2)}),$$

where $R_1(X)$ is such a function that is Fourier-Stieltjes transform equals to $r_1(it)e^{-t^2/2}$,

$$\int_{-\infty}^{\infty} e^{itx} dR_1(x) = r_1(it)e^{-t^2/2}$$

This idea of inverting an expansion of characteristic function was first proposed by Hall (1992b) for a one-sample i.i.d. mean. Applying integration by part to the identity (characteristic function): $e^{-t^2/2} = \int e^{itx}\phi(x)dx$, we obtain

$$R_1(x) = \left[\frac{A}{2} + \frac{2A}{6}(x^2-1)\right]\phi(x) = \frac{A}{6}(2x^2+1)\phi(x)$$

Therefore,

$$P(W_n \le x) = \Phi(x) + N^{-1/2}q(x)\phi(x) + O(N^{-min(1,r+1/2)})$$

where

$$q(x) = \frac{A}{6}(2x^2+1), A = \left\{\frac{V_1}{\lambda} + \frac{V_2}{1-\lambda}\right\}^{-3/2}\left\{\frac{V_1^{3/2}\gamma_1}{\lambda^2} - \frac{V_2^{3/2}\gamma_2}{(1-\lambda)^2}\right\}$$

Since $T = W_N + O(N^{-1})$, Proposition 1 follows.

References

Barrett, J. and Goldsmith, L. (1976). When is n sufficiently large? *American Statistician* **30** 67–70.

Boos, D. and Hughes-Oliver, J. (2000). How large does n have to be for z and t intervals. *American Statistician* **54** 121–128.

Callahan, C. Kesterson, J.and Tierney, W. (1997). Association of symptoms of depression with diagnostic test charges among older adults. *Annals of Internal Medicine* **126** 426–432.

Chen, L (1995). Testing the mean of skewed distributions *Journal of the American Statistical Association* **90** 767–72.

Efron, B. and Tibshirani, R. (1993). *An introduction to the bootstrap.* New York: Chapman & Hall.

Gayen, A. (1949). The distribution of students t in random samples of any size drawn from non-normal universes. *Biometrika* **36**, 353-369.

Hall, P. (1992a). *The bootstrap and edgeworth expansion.* New York: Springer.

Hall, P. (1992b). On the removal of skewness by transformation. *Journal of the Royal Statistical Society. Ser. B* **54**, 221-228.

Hoaglin, D. (1985). Summarizing shape numerically: The g-and-h distributions. in *Exploring data, tables, trends, and shapes* (Hoaglin, D. et al., eds.) pages 461-511.

Johnson, N. (1978). Modified t tests and confidence intervals for asymmetrical populations. *Journal of the American Statistical Association* **73**, 536-544.

Student (1908). The probable error of a mean. *Biometrika* **6**, 1-25.

Scheffé, H. (1970). Practical Solutions of the Behrens-Fisher Problem. *Journal of the American Statistical Association* **65**, 1501-1508.

Zhou, X.-H. and Gao, S. (2000). One-sided confidence intervals for means of positively skewed distributions. *American Statistician* **54**, 100-104.

Zhou, X.-H., Gao, S. and Hui, S. (1997). Methods for comparing the means of two independent log-normal samples. *Biometrics* **53**, 1129-1135.

[*Received*]