

11-1-2016

A Comparison Of Usual t-Test Statistic and Modified t-Test Statistics on Skewed Distribution Functions


Wooi K. Lim

William Paterson University of New Jersey, limw@wpunj.edu

Alice W. Lim

University of California, Los Angeles, canalice@gmail.com

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Lim, Wooi K. and Lim, Alice W. (2016) "A Comparison Of Usual t-Test Statistic and Modified t-Test Statistics on Skewed Distribution Functions," *Journal of Modern Applied Statistical Methods*: Vol. 15: Iss. 2, Article 8.

DOI: 10.22237/jmasm/1478001960

Available at: <http://digitalcommons.wayne.edu/jmasm/vol15/iss2/8>

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized administrator of DigitalCommons@WayneState.

A Comparison of Usual t -Test Statistic and Modified t -Test Statistics on Skewed Distribution Functions

Wooi K. Lim

William Paterson University
Wayne, New Jersey

Alice W. Lim

University of California at Los Angeles
Los Angeles, California

When the sample size n is small, the random variable $T = \sqrt{n}(\bar{X} - \mu)/S$ is said to follow a central t distribution with degrees of freedom $(n - 1)$, where \bar{X} is the sample mean and S is the sample standard deviation, provided that the data $X \sim N(\mu, \sigma^2)$. The random variable T can be used as a test statistic to hypothesize the population mean μ . Some argue that the t -test statistic is robust against the normality of the distribution and claim that the normality assumption is not necessary. In this article we will use simulation to study whether the t -test is really robust if the population distribution is not normally distributed. In particular, we will study how the skewness of a probability distribution will affect the confidence interval as well as the t -test statistic.

Keywords: Skewness, t -test, confidence interval, Edgeworth expansion

Introduction

The effect of skewness, denoted by γ from here on, of a random variable X on t -test have been investigated by Johnson (1978), Hall (1992), Abramovitch and Singh (1985) and many others; but, those are more on the theoretic investigation and concentrated on the t -test. Very little has been studied on the confidence interval. Two independent samples t -test are studied by Sawilowsky and Blair (1992). Their studies are based on several skewed distributions and various sample sizes. Their simulation results show that the proportions of rejection in the upper tail or lower tail are affected by the skewness of the distribution when samples sizes are small. Blair and Sawilowsky (1993) comparing the performance

Wooi K. Lim is an Associate Professor of Statistics in the Department of Mathematics at William Paterson University. Email him at: limw@wpunj.edu. Alice W. Lim is a student majoring in Mathematics at UCLA. Email her at: canalice@gmail.com.

T-TEST ON SKEWED DISTRIBUTION FUNCTION

usual independent samples t -test and modified t -tests under different distribution functions and various samples sizes. For further details on the performance and recommendation of which t -test under various distributions, see Blair and Sawilowsky (1993).

Consider the one sample t -test. Based on simulation studies, skewness of the distribution does not affect the t -test as much as the confidence interval. It can be shown that the coverage error is larger than the pre-determined coverage error, α , if the data follow a skewed distribution function.

Intuitively, if X is a random variable with mean μ but is positively skewed, $\gamma > 0$, then the population median is less than the population mean μ . A sample of size n from X is likely to have more than 50% of values to be less than μ ; hence most likely $(\bar{X} - \mu) < 0$. If $\gamma > 0$, then a $(1 - \alpha) \times 100\%$ confidence interval for μ

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

will miss the mean μ more on the upper side than the lower side. This effect is reported by Boos and Hughes-Oliver (2000). Define the missed right and missed left as given in Boos and Hughes-Oliver (2000, p. 122), where miss right occurs when the population mean μ is above the upper confidence limit, i.e.,

$$\mu > \bar{X} + t_{\alpha/2} \frac{S}{\sqrt{n}},$$

and miss left occurs when the population mean μ is below the lower confidence limit, i.e.,

$$\mu < \bar{X} - t_{\alpha/2} \frac{S}{\sqrt{n}},$$

and miss = (miss right + miss left). Tables 1, 2, 3 and 4 are the simulated results of missed right, missed left and missed of usual confidence interval given in equation (1). Four types of population distributions, namely normal distribution ($\gamma = 0$), Laplace distribution ($\gamma = 0$), Gamma distribution ($\gamma > 0$) and Gumbel distribution ($\gamma < 0$) were selected for the simulation study.

LIM & LIM

Table 1. Table of miss right, miss left and miss of $(1 - \alpha) \times 100\%$ confidence interval for μ with $X_1, X_2, \dots, X_n \sim \text{Normal}(1, 2)$, $n = 10$ and skewness = 0.0.

α	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
miss right	0.005	0.011	0.015	0.021	0.025	0.029	0.033	0.041	0.043	0.052
miss left	0.004	0.009	0.017	0.022	0.023	0.033	0.034	0.038	0.049	0.06
miss	0.009	0.02	0.032	0.044	0.048	0.061	0.067	0.079	0.092	0.111
α	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.2
miss right	0.053	0.057	0.067	0.073	0.07	0.082	0.088	0.1	0.098	0.101
miss left	0.056	0.061	0.063	0.071	0.082	0.082	0.087	0.093	0.107	0.101
miss	0.109	0.118	0.129	0.144	0.152	0.164	0.175	0.192	0.205	0.201

Table 2. Table of miss right, miss left and miss of $(1 - \alpha) \times 100\%$ confidence interval for μ with $X_1, X_2, \dots, X_n \sim \text{Laplace}(1, 2)$, $n = 10$ and skewness = 0.0.

α	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
miss right	0.003	0.007	0.012	0.015	0.02	0.025	0.03	0.037	0.041	0.047
miss left	0.004	0.005	0.01	0.014	0.02	0.026	0.031	0.035	0.039	0.047
miss	0.007	0.012	0.022	0.029	0.04	0.051	0.061	0.072	0.08	0.094
α	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.2
miss right	0.052	0.059	0.058	0.07	0.075	0.081	0.086	0.095	0.095	0.101
miss left	0.054	0.058	0.07	0.07	0.079	0.079	0.083	0.089	0.1	0.1
miss	0.106	0.117	0.128	0.14	0.154	0.16	0.169	0.184	0.195	0.201

Table 3. Table of miss right, miss left and miss of $(1 - \alpha) \times 100\%$ confidence interval for μ with $X_1, X_2, \dots, X_n \sim \text{Gamma}(1, 2)$, $n = 10$ and skewness = 2.

α	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
miss right	0.051	0.061	0.08	0.093	0.101	0.108	0.121	0.122	0.125	0.13
miss left	0.0	0.001	0.001	0.002	0.005	0.004	0.007	0.009	0.009	0.015
miss	0.051	0.062	0.081	0.095	0.106	0.112	0.128	0.132	0.134	0.145
α	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.2
miss right	0.144	0.149	0.155	0.157	0.167	0.174	0.177	0.175	0.185	0.192
miss left	0.015	0.017	0.02	0.025	0.03	0.029	0.037	0.044	0.047	0.047
miss	0.159	0.166	0.175	0.182	0.197	0.203	0.215	0.219	0.232	0.239

T-TEST ON SKEWED DISTRIBUTION FUNCTION

Table 4. Table of miss right, miss left and miss of $(1 - \alpha) \times 100\%$ confidence interval for μ with $X_1, X_2, \dots, X_n \sim \text{Gumbel}(1, 2)$, $n = 10$ and skewness = -1.14.

α	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
miss right	0.001	0.003	0.005	0.008	0.01	0.014	0.016	0.02	0.021	0.029
miss left	0.019	0.03	0.036	0.044	0.054	0.066	0.068	0.082	0.084	0.089
miss	0.02	0.033	0.041	0.052	0.064	0.08	0.084	0.102	0.105	0.118
α	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.2
miss right	0.031	0.031	0.039	0.042	0.047	0.049	0.055	0.062	0.063	0.063
Miss left	0.099	0.094	0.103	0.117	0.118	0.118	0.129	0.137	0.142	0.149
miss	0.13	0.125	0.142	0.159	0.165	0.167	0.184	0.199	0.206	0.212

It is shown in Table 1 if X is normally distributed, the nominal coverage error α is close to the simulated missed coverage error. Results in Tables 1 and 2 show that if the probability distributions are symmetrically distributed, then the missed left \approx missed right. Tables 3 and 4 show that if X is skewed, such as in the Gamma distribution (Table 3) or Gumbel distribution (Table 4), the missed coverage error is more than the nominal coverage error α . Interestingly, the results show that for a right skewed distribution, the missed right coverage errors are substantially greater than the missed left coverage errors (see Table 3). The opposite is true for the left skewed, Gumbel, distribution (see Table 4).

It is well known that the random variable $T = \sqrt{n}(\bar{X} - \mu)/S$ is a ratio of the normal random variable and $\chi_{(n-1)}$ random variable with $(\bar{X} - \mu)$ and S statistically independent. Will the random variable $T = \sqrt{n}(\bar{X} - \mu)/S$ be affected by the skewness of the probability distribution? Simulated empirical distribution of T for the same four chosen population distributions are under studied. Our results are summarized in Figures 1, 2, 3 and 4. In this simulation, a sample of $n = 10$ is drawn from the population distribution with replications of $M = 5000$. Each figure contains figures (a), (b) and (c), with the exception of Figure 1. Figures (a) are histograms of the t -test statistics, $t^* = \sqrt{n}(\bar{X} - \mu_0)/S$, under the assumption that $H_0 : \mu = \mu_0$ is true. Figures (b) are the plots of $\bar{X} - \mu_0$ versus S .

In Figure 1, $X_1, X_2, \dots, X_{10} \stackrel{iid}{\sim} N(1, 2)$ were sampled. The histogram in Figure 1(a) is an empirical distribution of t^* under the assumption of $H_0 : \mu = 1$. The histogram is quite symmetric and the plot of $\bar{X} - \mu_0$ versus S does not seem

to have any correlation. This is what we expected from a t -test statistic. What happens if X is not normal?

Figure 2(a) is the distribution of t^* with $X_1, X_2, \dots, X_{10} \stackrel{iid}{\sim} Laplace(1, 2)$. The histogram shows that t^* is symmetric. The plot of $(\bar{X} - \mu_0)$, where $\mu_0 = 1$, versus S , see Figure 2(b), does not show any correlation. Figure 2(c) is the empirical distribution of t^* based on $X \sim N(1, 2)$ versus $X \sim Laplace(1, 2)$. It can be seen that the distribution of t^* based on $X \sim Laplace(1, 2)$ has shorter tails than the t^* computed from $X \sim N(1, 2)$. Clearly the variability of $(\bar{X} - \mu_0)$ plays a role in the distribution of t^* . This may suggest that t^* generated from X following a *Laplace* distribution may not be as sensitive as the t^* obtained from a normal distribution.

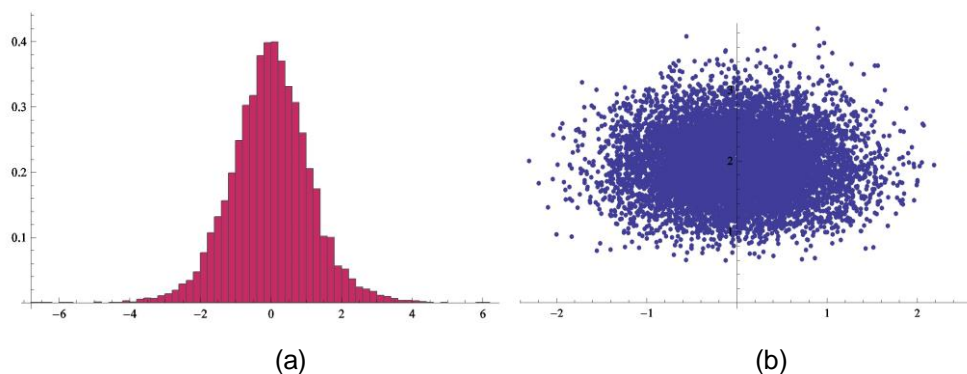


Figure 1. (a) histogram of $t^* = \sqrt{n}(\bar{X} - \mu_0)/S$, where $\mu_0 = 1$, $X_1, X_2, \dots, X_{10} \sim Normal(1, 2)$. (b) plot of $(\bar{X} - \mu_0)$ versus S .

T-TEST ON SKEWED DISTRIBUTION FUNCTION

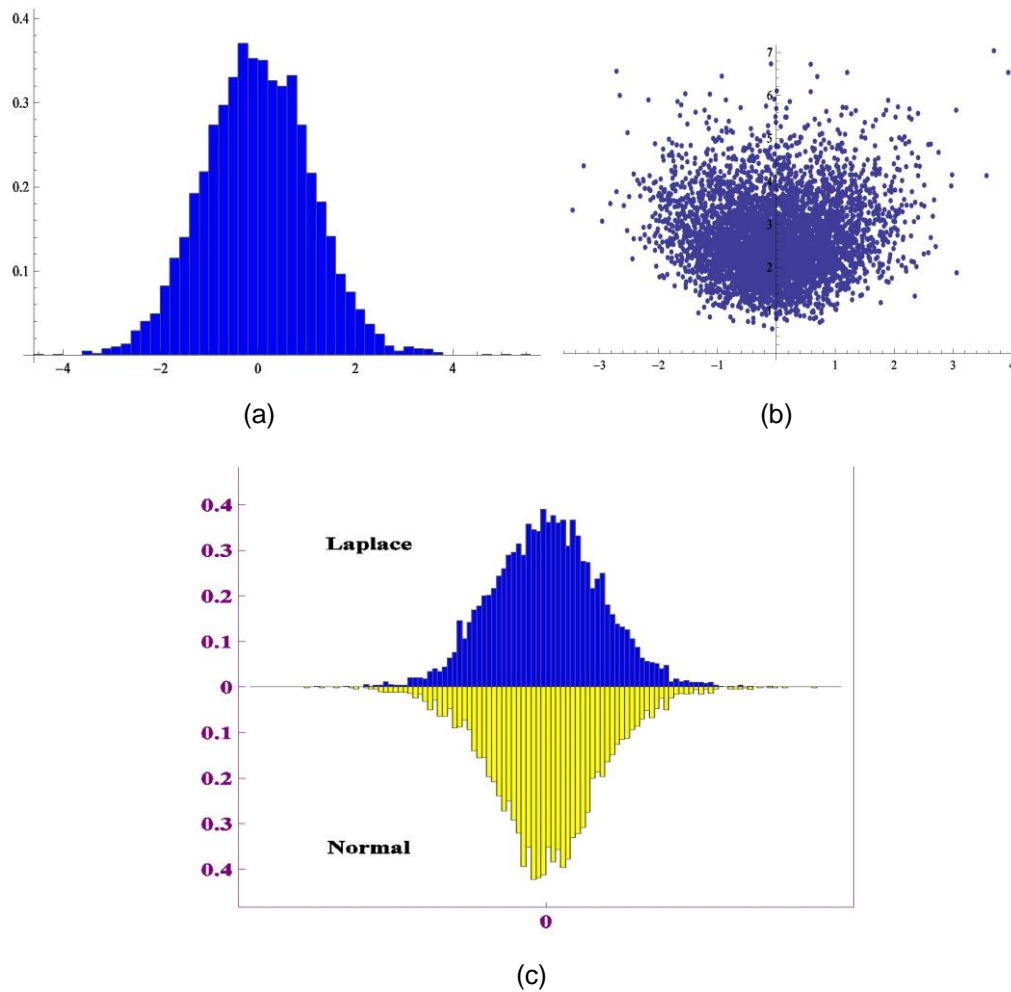


Figure 2. (a) histogram of $t^* = \sqrt{n}(\bar{X} - \mu_0)/S$, where $\mu_0 = 1$, $X_1, X_2, \dots, X_{10} \sim Laplace(1, 2)$. (b) plot of $(\bar{X} - \mu_0)$ versus S . (c) Fig. 2a over Fig. 1a

LIM & LIM

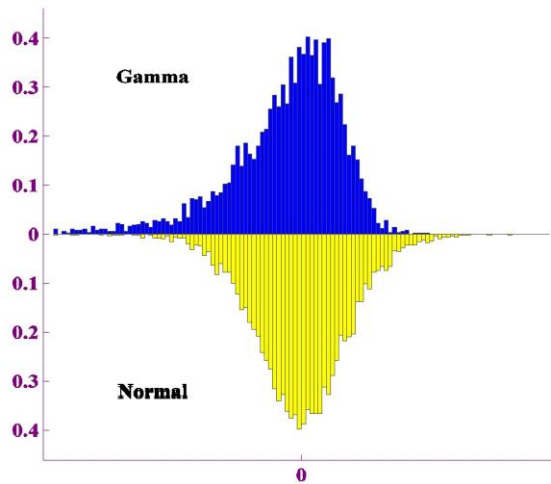
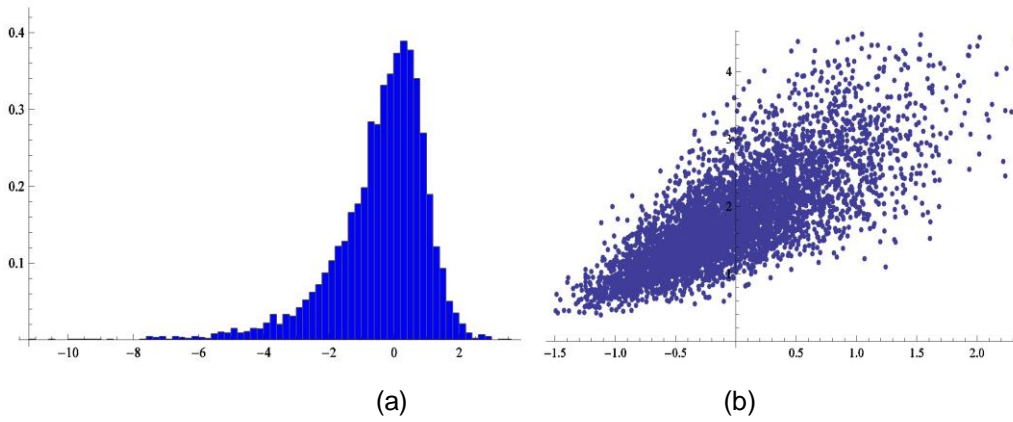


Figure 3. (a) histogram of $t^* = \sqrt{n}(\bar{X} - \mu_0)/S$, where $\mu_0 = 2$, $X_1, X_2, \dots, X_{10} \sim \text{Gamma}(1, 2)$. (b) plot of $(\bar{X} - \mu_0)$ versus S . (c) Fig 3a over Fig 1a

T-TEST ON SKEWED DISTRIBUTION FUNCTION

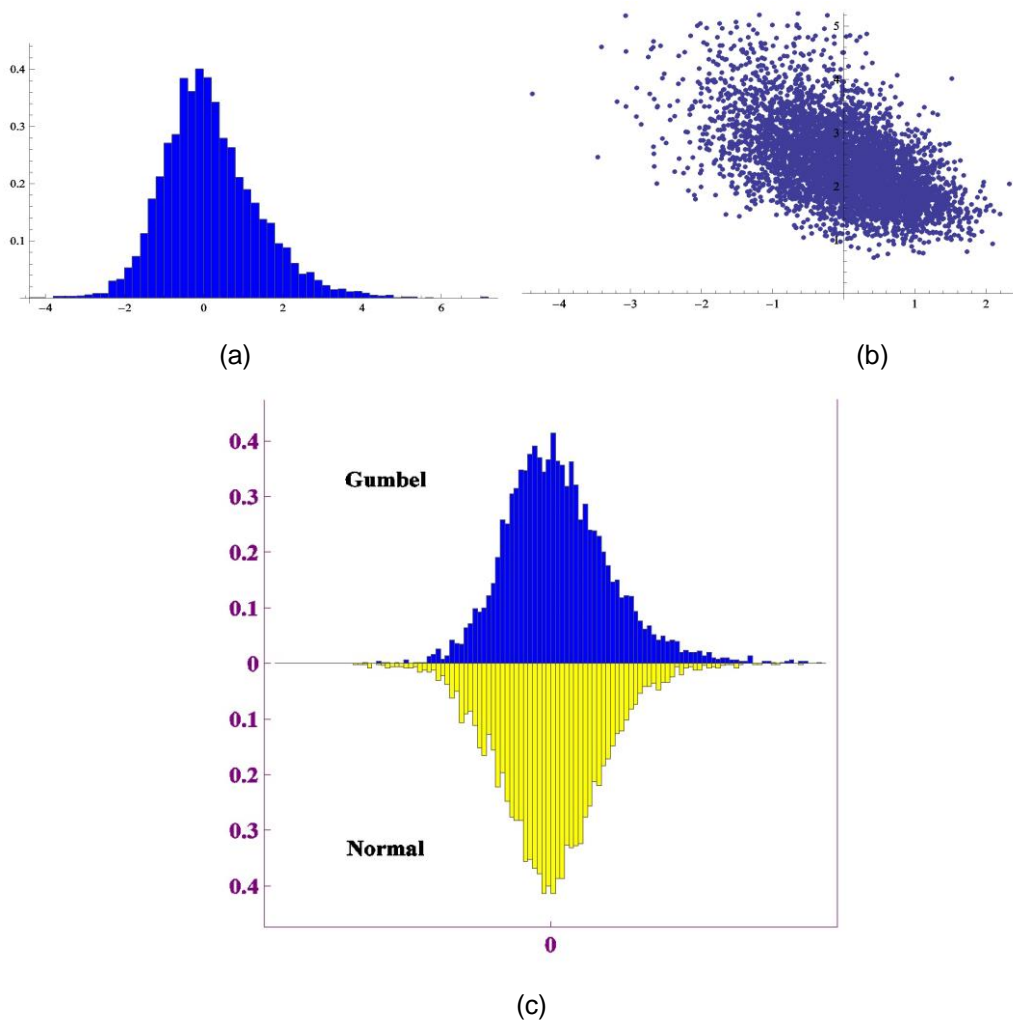


Figure 4. (a) histogram of $t^* = \sqrt{n}(\bar{X} - \mu_0)/S$, where $\mu_0 \approx -0.1544$, $X_1, X_2, \dots, X_{10} \sim Gumbel(1, 2)$. (b) plot of $(\bar{X} - \mu_0)$ versus S . (c) Fig 4a over Fig 1a

Figures in 3 and 4 are simulation results from a skewed probability distributions. Figure 3(a) is the distribution of t^* with $X_1, X_2, \dots, X_{10} \stackrel{iid}{\sim} Gamma(1, 2)$. Interestingly, Gamma distribution is a right skewed distribution but the distribution of t^* is left skewed. One can see in Figure 3(b) that there is a positive correlation between $(\bar{X} - \mu_0)$, where $\mu_0 = 2$, and S .

Further, Figure 3(c) shows that t^* 's constructed from $X \sim \text{Gamma}(1, 2)$ lie below 0 more often than fall above 0. One can see in Figure 3(b) that $(\bar{X} - \mu_0)$ versus S is more disperse when $(\bar{X} - \mu_0) > 0$ while it is less varied when $(\bar{X} - \mu_0) < 0$. Thus, when $(\bar{X} - \mu_0) > 0$ and large, it tends to counter by large S making t^* more concentrated on the right hand side. On the other hand, when $(\bar{X} - \mu_0) < 0$ and S is small, t^* tends to stretch further towards the negative side making t^* skewed negatively. Similar arguments can explain why left skewed distributions will have \bar{X} overestimate μ more often and making the distribution of t^* positively skewed, see Figure 4 where $\mu_0 \approx -0.1544$. In the next section we will compare the two transformation methods, proposed by Hall (1992), with the usual test statistics T .

Correction and Transformation

Johnson (1978) and others noticed some undesired effects on skewed distributions on the t -test. Hall (1992) proposed to modify the t -test statistic T , say $g(T)$, so that $g(T)$ is less skew and less bias. The transformed test statistic $g(T)$ must be invertible to obtain a unique modified confidence interval for μ . He suggested g been a monotonic function to achieve the invertibility. The two monotonic transformations of T proposed by Hall (1992) are:

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

and

$$T_2 = \frac{\sqrt{n}}{2a\hat{\gamma}} \left(\text{Exp} \left(\frac{2a\hat{\gamma}T}{\sqrt{n}} \right) - 1 \right) + \frac{1}{n} b \hat{\gamma}, \quad (3)$$

where $\hat{\gamma} = \left\{ \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{n} \right\} / S^3$ is an estimate of γ . Note that as the estimated skewness $\hat{\gamma} \rightarrow 0$ both $T_1 \rightarrow 0$ and $T_2 \rightarrow 0$. The test statistic T_1 is a direct consequence of the Edgeworth expansion of T given below, see for example, A. DasGupta (2008) page 191.

T-TEST ON SKEWED DISTRIBUTION FUNCTION

Theorem 1. Let X_1, X_2, \dots, X_n be iid with CDF F having mean μ , variance σ^2 , and $E(X_1 - \mu)^4 < \infty$. Assume that F satisfies Cramér's condition. Let $\gamma = E(X_1 - \mu)^3 / \sigma^3$, $P_1(x) = (2x^2 + 1) / 6$, then the CDF of t -statistic $T = \sqrt{n}(\bar{X} - \mu) / S$ admits the expansion

$$P(T \leq x) = \Phi(x) + \frac{\gamma P_1(x)}{\sqrt{n}} \phi(x) + O(n^{-1}) \quad (4)$$

uniformly in x , where $\Phi(x)$ and $\phi(x)$ are standard normal distribution and density function, respectively.

From the above theorem, the skewness of the distribution F has significant effect on T especially when the sample size n is small. One term Edgeworth expansion for T is (see Hall 1987)

$$P(T \leq x) = \Phi(x) + \frac{\gamma}{6\sqrt{n}} (2x^2 + 1) \phi(x) + O(n^{-1}). \quad (5)$$

From (5) a modified test statistic is

$$T_0 = T + \frac{\hat{\gamma}}{3\sqrt{n}} T^2 + \frac{\hat{\gamma}}{6\sqrt{n}}, \quad (6)$$

which may be used to correct the skewness of T . One may use $a = 1/3$ and $b = 1/6$ in equations (2) and (3). As indicated by Hall (1992), T_0 is not a monotonic function and hence is not invertible to construct a confidence interval for μ . Hall (1992) modified T_0 to T_1 as given in (2) so that it can be inverted to construct the confidence interval as well as to correct the bias and skewness. We are not sure why the last term of (2) and (3) is $(b\hat{\gamma})/n$ rather than $(b\hat{\gamma})/\sqrt{n}$. Zhou and Gao (2000) uses

$$T_1 = T + \frac{\hat{\gamma}}{\sqrt{n}} (aT^2 + b) + \frac{1}{3n} a^2 \hat{\gamma}^2 T^3 \quad (7)$$

which is slightly different from Hall's T_1 and we will called (7) the T_1 from now on.

The simulation will justify the $P_1(x)$ in the one term of Edgeworth expansion of T is indeed a polynomial function of order 2. Consider the Edgeworth expansion of T ,

$$P(T \leq t_\alpha) = \Phi(t_\alpha) + \frac{\gamma P_1(t_\alpha)}{\sqrt{n}} \phi(t_\alpha) + O(n^{-1}),$$

and one can show that

$$P(\text{miss left}) = P\left(\mu < \bar{X} - t_{\alpha/2} \frac{S}{\sqrt{n}}\right) = \frac{\alpha}{2} - \frac{\gamma}{\sqrt{n}} P_1(t_{\alpha/2}) \phi(t_{\alpha/2}) + O(n^{-1}). \quad (8)$$

Similarly,

$$P(\text{miss right}) = P\left(\mu > \bar{X} + t_{\alpha/2} \frac{S}{\sqrt{n}}\right) = \frac{\alpha}{2} + \frac{\gamma}{\sqrt{n}} P_1(-t_{\alpha/2}) \phi(t_{\alpha/2}) + O(n^{-1}). \quad (9)$$

If $P_1(t) > 0$ for all t , one can see then, a positively skewed distribution ($\gamma > 0$) $P(\text{miss right}) > P(\text{miss left})$. It can be seen in (9) that, with $\gamma > 0$ and $P_1(t) > 0$, $P(\text{miss right}) > \frac{\alpha}{2}$ and in equation (8) one obtains $P(\text{miss left}) < \frac{\alpha}{2}$. The opposite is true for a negatively skewed distribution. Let

$$\begin{aligned} P(\text{miss}) &= P(\text{miss left}) + P(\text{miss right}) \\ &= \alpha + \frac{\gamma}{\sqrt{n}} \{P_1(-t_{\alpha/2}) - P_1(t_{\alpha/2})\} \phi(t_{\alpha/2}) + O(n^{-1}). \end{aligned} \quad (10)$$

Let $k(\alpha) = (P(\text{miss}) - \alpha)$ and $g(t_{\alpha/2}) = P_1(-t_{\alpha/2}) - P_1(t_{\alpha/2})$. Rewrite Equation (10) as

$$k(\alpha) = \frac{\gamma}{\sqrt{n}} g(t_{\alpha/2}) \phi(t_{\alpha/2}) + O(n^{-1})$$

and

T-TEST ON SKEWED DISTRIBUTION FUNCTION

$$\frac{\sqrt{nk}(\alpha)}{\gamma\phi(t_{\alpha/2})} \approx g(t_{\alpha/2}).$$

A plot of $\sqrt{nk}(\alpha)/\hat{\gamma}\phi(t_{\alpha/2})$ versus $t_{\alpha/2}$ will review the structure of $P_1(x)$ if the random variable X is skewed.

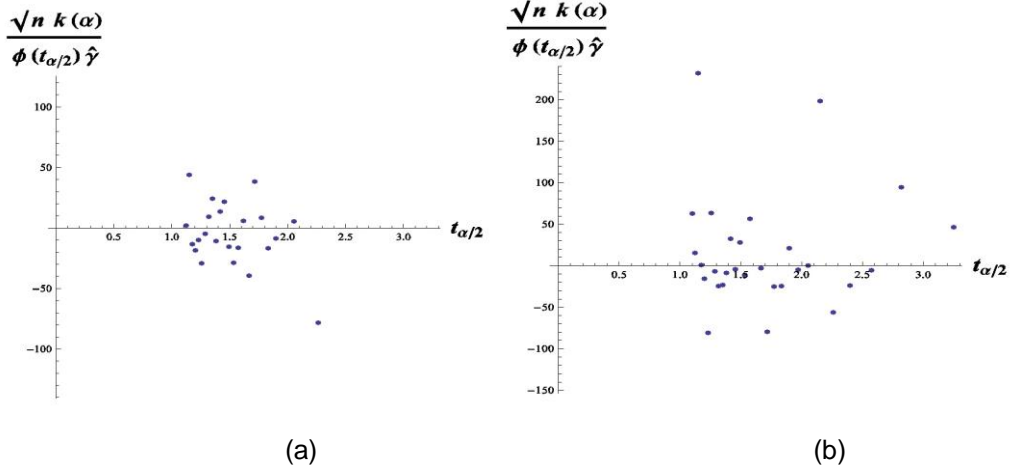


Figure 5. Plot of $\sqrt{nk}(\alpha)/\hat{\gamma}\phi(t_{\alpha/2})$ versus $t_{\alpha/2}$ (a) $X_1, X_2, \dots, X_n \sim Normal(1, 2)$ and (b) $X_1, X_2, \dots, X_n \sim Laplace(1, 2)$.

Finding the structure of g is a matter of regression. However, we are interested in whether the structure of g agrees with the quadratic function given in Theorem 1. Note that $P_1(x) = 2x^2 + 1$ is an even function. Hence, $P(-t_{\alpha/2}) - P(t_{\alpha/2}) = 0$. However, if X is a right skewed distributed function, we have seen in Table 3 that $P(\text{miss right}) > P(\text{miss left})$; we expect $g(t_{\alpha/2}) > 0$ (see Figure 6(a)). Similarly, one can see that $g(t_{\alpha/2}) < 0$ when X has left skewed distribution (see Figure 6(b)). If X is a symmetrically distributed function with skewness $\gamma = 0$, then the plot of $\sqrt{nk}(\alpha)/\hat{\gamma}\phi(t_{\alpha/2})$ versus $t_{\alpha/2}$ does not show any pattern as seen in Figures 5 (a) and (b). Figures 6 (a) and (b) show that $g(t_{\alpha/2})$ resembles a quadratic function, confirming that the use of the second term in (5) is necessary if skewness appeared in the data. From the simulation and the

equation (10) one can see that if X is skewed, then $P(\text{miss}) > \alpha$; because $\left\{ \frac{\hat{\gamma}}{\sqrt{n}} g(t_{\alpha/2}) \phi(t_{\alpha/2}) \right\} > 0$. Thus, it explains that when X is a skewed distribution the coverage error will be larger than the nominal coverage error α .

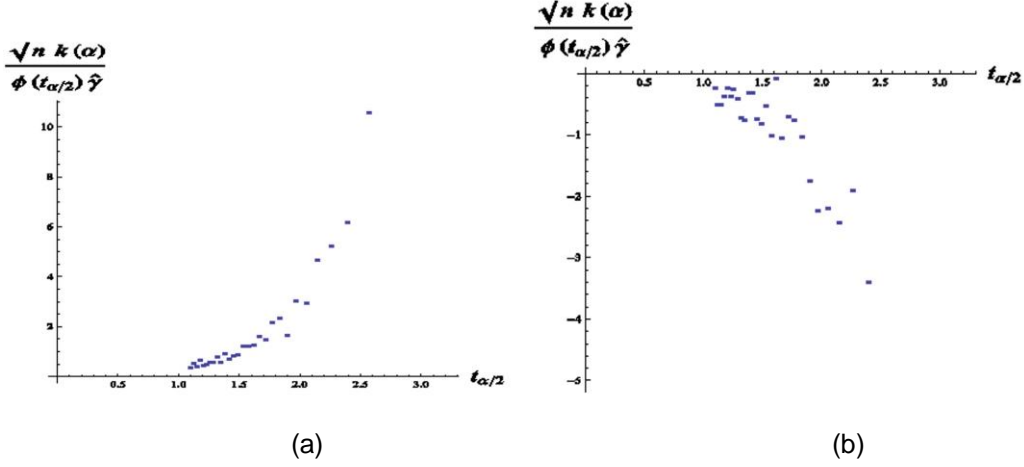


Figure 6: Plot of $\sqrt{nk}(\alpha)/\hat{\gamma}\phi(t_{\alpha/2})$ versus $t_{\alpha/2}$ (a) $X_1, X_2, \dots, X_n \sim \text{Gamma}(1, 2)$ and (b) $X_1, X_2, \dots, X_n \sim \text{Gumbel}(1, 2)$.

Comparison of T , T_1 and T_2

The objective is to compare the test statistics T , T_1 , and T_2 . The modified test statistic T_2 given in Hall (1992) has not been paid attention to as far as we know. As mentioned earlier, $\lim_{\hat{\gamma} \rightarrow 0} T_2 = T$ and if X is positively skewed, most likely $(\bar{X} - \mu) < 0$. We modify the T in T_2 by adding the term $b\bar{\gamma}/\sqrt{n}$ to \bar{X} so that it shifts to the correct direction, i.e.,

$$T_b = \sqrt{n} \left(\bar{X} + \frac{b\bar{\gamma}}{\sqrt{n}} - \mu \right) / S. \quad (11)$$

Unlike T_2 in (3), our modified T_2 is

T-TEST ON SKEWED DISTRIBUTION FUNCTION

$$T_2^* = \frac{\sqrt{n}}{2a\hat{\gamma}} \left(\text{Exp} \left(\frac{2a\hat{\gamma}T_b}{\sqrt{n}} \right) - 1 \right) + (b\hat{\gamma})/\sqrt{n}. \quad (12)$$

The simulation study is repeated on the four chosen distributions but this time we compare the empirical distributions of three test statistics, i.e., T , T_1 and T_2^* . Both T_1 and T_2^* are less skewed than T if X is simulated from a skewed distribution, which are shown in Figures 9 and 10. Figures 7 and 8 show that if X is a symmetric distribution, the distributions of T_1 and T_2^* remain symmetric.

Confidence Interval

A simulation study of confidence intervals derived from T , T_1 and T_2^* was conducted. The $(1 - \alpha) \times 100\%$ confidence interval for μ derived from T_1 is

$$\left(\bar{X} - \frac{S}{\sqrt{n}} h_1^{-1}(-t_{\alpha/2}), \bar{X} + \frac{S}{\sqrt{n}} h_1^{-1}(t_{\alpha/2}) \right), \quad (13)$$

where $h_1^{-1}(t) = \sqrt{n}(a\hat{\gamma})^{-1} \left(1 + \frac{3a\hat{\gamma}}{\sqrt{n}} \left(t - \frac{b\hat{\gamma}}{\sqrt{n}} \right) \right)^{1/3} - \sqrt{n}(a\hat{\gamma})^{-1}$. One can see that $h_1^{-1}(t)$ may produce complex values for some $\hat{\gamma}$ and t . If the first 4 terms of $h_1^{-1}(t)$ are expanded and the expansion is simplified,

$$h_1^{-1}(t) \approx \left(t - \frac{b\hat{\gamma}}{\sqrt{n}} \right) - \frac{a\hat{\gamma}}{\sqrt{n}} \left(t - \frac{b\hat{\gamma}}{\sqrt{n}} \right)^2 + \frac{5(a\hat{\gamma})^2}{3n} \left(t - \frac{b\hat{\gamma}}{\sqrt{n}} \right)^3 = h_{1*}^{-1}(t). \quad (14)$$

Then, replace $h_1^{-1}(t)$ by $h_{1*}^{-1}(t)$ in (13), the approximation confidence interval of T_1 , called T_{1*}^* , will guarantee to produce a real valued confidence interval. The confidence interval of T_{1*}^* is

$$\left(\bar{X} - \frac{S}{\sqrt{n}} h_{1*}^{-1}(-t_{\alpha/2}), \bar{X} + \frac{S}{\sqrt{n}} h_{1*}^{-1}(t_{\alpha/2}) \right), \quad (15)$$

LIM & LIM

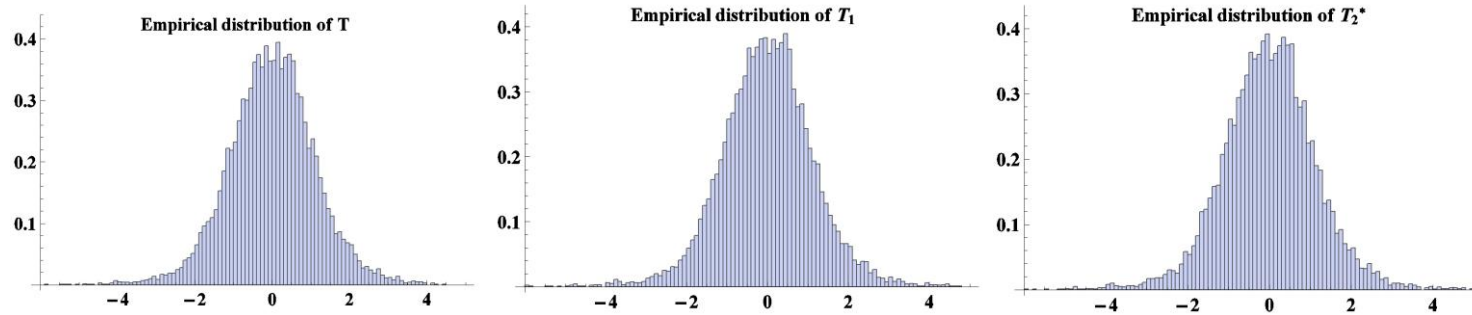


Figure 7. Left figure is the histogram of $T = \sqrt{n}(\bar{X} - \mu)/S$, $X_1, X_2, \dots, X_n \sim \text{Normal}(1, 2)$. The graph in the middle is the histogram of T_1 and the histogram on the right is the empirical distribution of T_2^* .

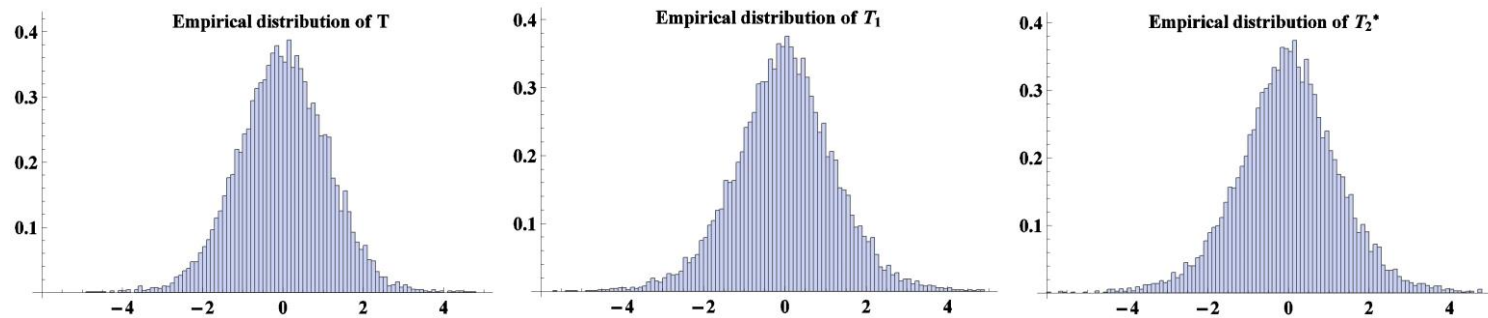


Figure 8. Left figure is the histogram of $T = \sqrt{n}(\bar{X} - \mu)/S$, $X_1, X_2, \dots, X_n \sim \text{Laplace}(1, 2)$. The graph in the middle is the histogram of T_1 and the histogram on the right is the empirical distribution of T_2^* .

T-TEST ON SKEWED DISTRIBUTION FUNCTION

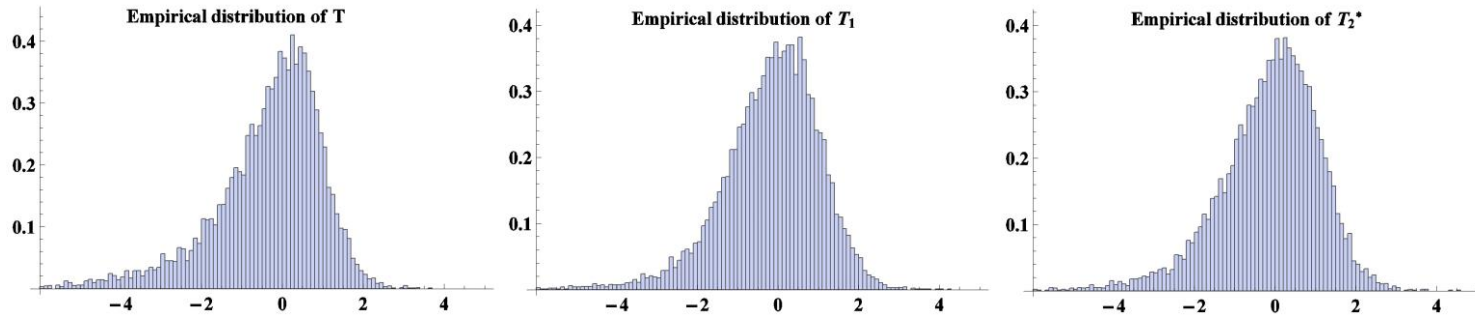


Figure 9. Left figure is the histogram of $T = \sqrt{n}(\bar{X} - \mu)/S$, $X_1, X_2, \dots, X_n \sim \text{Gamma}(1, 2)$. The graph in the middle is the histogram of T_1 and the histogram on the right is the empirical distribution of T_2^* .

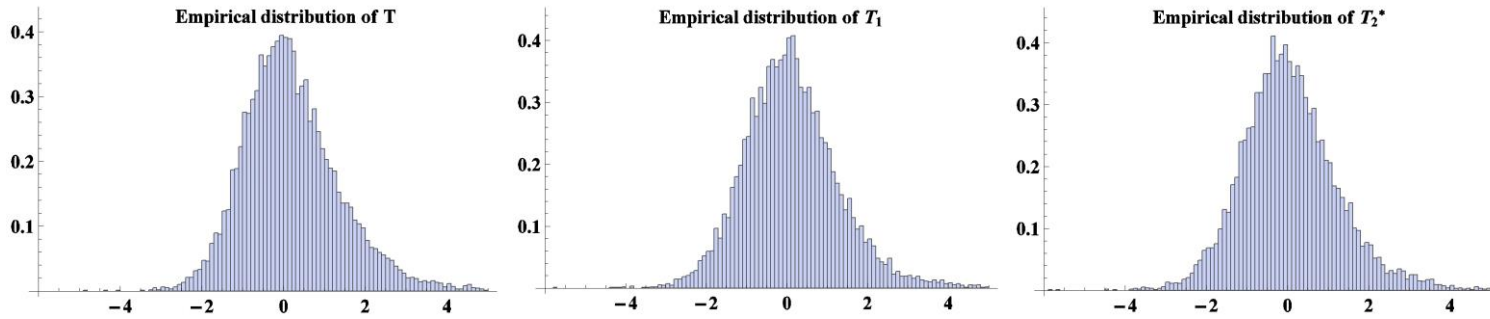


Figure 10. Left figure is the histogram of $T = \sqrt{n}(\bar{X} - \mu)/S$, $X_1, X_2, \dots, X_n \sim \text{Gumbel}(1, 2)$. The graph in the middle is the histogram of T_1 and the histogram on the right is the empirical distribution of T_2^* .

where $h_*^{-1}(t)$ is given in (14). The $(1 - \alpha) \times 100\%$ confidence interval for μ derived from T_2^* is

$$\left[\left(\bar{X} + \frac{b\hat{\gamma}}{\sqrt{n}} \right) - \frac{S}{\sqrt{n}} \left\{ \frac{\sqrt{n}}{2a\hat{\gamma}} \ln \left(1 - \frac{2a\hat{\gamma}}{\sqrt{n}} \left(t_{\alpha/2} + \frac{b\hat{\gamma}}{\sqrt{n}} \right) \right) \right\}, \right. \\ \left. \left(\bar{X} + \frac{b\hat{\gamma}}{\sqrt{n}} \right) + \frac{S}{\sqrt{n}} \left\{ \frac{\sqrt{n}}{2a\hat{\gamma}} \ln \left(1 + \frac{2a\hat{\gamma}}{\sqrt{n}} \left(t_{\alpha/2} - \frac{b\hat{\gamma}}{\sqrt{n}} \right) \right) \right\} \right].$$

It is not surprising that the logarithm function may produce a complex number. Expand the logarithm function and keep the first 3 terms of the Taylor series expansions, the approximation confidence interval for μ is

$$\left[\left(\bar{X} + \frac{b\hat{\gamma}}{\sqrt{n}} \right) - \frac{S}{\sqrt{n}} \left\{ \left(t_{\alpha/2} - \frac{b\hat{\gamma}}{\sqrt{n}} \right) - \frac{a\hat{\gamma}}{\sqrt{n}} \left(t_{\alpha/2} - \frac{b\hat{\gamma}}{\sqrt{n}} \right)^2 + \frac{4}{3} \frac{(a\hat{\gamma})^2}{\sqrt{n}} \left(t_{\alpha/2} - \frac{b\hat{\gamma}}{\sqrt{n}} \right)^3 \right\}, \right. \\ \left. \left(\bar{X} + \frac{b\hat{\gamma}}{\sqrt{n}} \right) + \frac{S}{\sqrt{n}} \left\{ \left(t_{\alpha/2} + \frac{b\hat{\gamma}}{\sqrt{n}} \right) + \frac{a\hat{\gamma}}{\sqrt{n}} \left(t_{\alpha/2} + \frac{b\hat{\gamma}}{\sqrt{n}} \right)^2 + \frac{4}{3} \frac{(a\hat{\gamma})^2}{\sqrt{n}} \left(t_{\alpha/2} + \frac{b\hat{\gamma}}{\sqrt{n}} \right)^3 \right\} \right]. \quad (16)$$

The above confidence interval (16) may be called a confidence interval from T_2^{**} . The confidence interval in (16) is different from that of in (15). The confidence interval in (15) subtract $(b\hat{\gamma})/\sqrt{n}$ from $t_{\alpha/2}$ on upper and lower confidence limit. Unlike the confidence interval in (15), the confidence interval in (16) tends to subtract $(b\hat{\gamma})/\sqrt{n}$ from $t_{\alpha/2}$ on the lower confidence limit but add $(b\hat{\gamma})/\sqrt{n}$ on the upper confidence limit.

It can be seen in Table 7 that if X is severely skewed, the modified confidence intervals T_1^* and T_2^{**} perform substantially better than the usual confidence interval derived from T . If the skewness is not severe, T performs better than the modified T .

T-TEST ON SKEWED DISTRIBUTION FUNCTION

Table 5. Table of nominal coverage error α and the simulated missed for $(1 - \alpha) \times 100\%$ confidence interval for μ with $X_1, X_2, \dots, X_n \sim Normal(1, 2)$.

$n = 10$	$M = 10000$	<i>Normal</i> (1,2)		<i>skewness</i> = 0,		$a = 1/3,$		$b = 1/6$		
α	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
miss T	0.01	0.02	0.03	0.04	0.049	0.06	0.069	0.079	0.089	0.101
miss T_1^*	0.011	0.022	0.031	0.04	0.05	0.06	0.069	0.08	0.09	0.101
miss T_2^{**}	0.013	0.025	0.035	0.045	0.055	0.064	0.074	0.084	0.094	0.104
α	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.2
miss T	0.109	0.122	0.13	0.14	0.15	0.16	0.168	0.177	0.189	0.202
miss T_1^*	0.109	0.12	0.128	0.138	0.149	0.158	0.167	0.176	0.187	0.2
miss T_2^{**}	0.114	0.125	0.132	0.142	0.152	0.162	0.171	0.181	0.191	0.203

Table 6. Table of nominal coverage error α and the simulated missed for $(1 - \alpha) \times 100\%$ confidence interval for μ with $X_1, X_2, \dots, X_n \sim Laplace(1, 2)$.

$n = 10$	$M = 10000$	<i>Laplace</i> (1,2)		<i>skewness</i> = 0,		$a = 1/3,$		$b = 1/6$		
α	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
miss T	0.005	0.014	0.023	0.032	0.041	0.051	0.061	0.073	0.084	0.095
miss T_1^*	0.01	0.023	0.037	0.049	0.061	0.074	0.086	0.1	0.113	0.125
miss T_2^{**}	0.015	0.032	0.049	0.064	0.076	0.092	0.104	0.119	0.132	0.144
α	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.2
miss T	0.105	0.116	0.127	0.14	0.151	0.161	0.171	0.182	0.194	0.202
miss T_1^*	0.135	0.148	0.159	0.172	0.186	0.194	0.205	0.216	0.23	0.237
miss T_2^{**}	0.154	0.167	0.178	0.191	0.204	0.213	0.224	0.236	0.249	0.256

Table 7. Table of nominal coverage error α and the simulated missed for $(1 - \alpha) \times 100\%$ confidence interval for μ with $X_1, X_2, \dots, X_n \sim Gamma(1, 2)$.

$n = 10$	$M = 10000$	<i>Gamma</i> (1,2)		<i>skewness</i> = 2,		$a = 1/3,$		$b = 1/6$		
α	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
miss T	0.047	0.062	0.077	0.089	0.1	0.109	0.119	0.13	0.139	0.148
miss T_1^*	0.024	0.035	0.048	0.058	0.068	0.077	0.088	0.098	0.108	0.118
miss T_2^{**}	0.022	0.033	0.046	0.056	0.065	0.074	0.086	0.097	0.107	0.117
α	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.2
miss T	0.156	0.166	0.176	0.186	0.19	0.198	0.206	0.22	0.229	0.238
miss T_1^*	0.128	0.138	0.149	0.16	0.167	0.178	0.186	0.2	0.21	0.221
miss T_2^{**}	0.127	0.137	0.149	0.159	0.168	0.178	0.186	0.201	0.211	0.223

Table 8. Table of nominal coverage error α and the simulated missed for $(1 - \alpha) \times 100\%$ confidence interval for μ with $X_1, X_2, \dots, X_n \sim \text{Gumbel}(1, 2)$.

$n = 10$	$M = 10000$	<i>Gumbel</i> (1,2) <i>skewness</i> ≈ -1.14 , $a = 1/3$, $b = 1/6$									
α	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	
miss T	0.019	0.033	0.043	0.055	0.064	0.075	0.085	0.094	0.105	0.115	
miss T_1^*	0.015	0.027	0.037	0.048	0.058	0.068	0.079	0.088	0.099	0.109	
miss T_2^{**}	0.017	0.028	0.039	0.050	0.061	0.071	0.081	0.091	0.101	0.111	
α	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.2	
miss T	0.125	0.134	0.142	0.153	0.162	0.173	0.182	0.193	0.201	0.212	
miss T_1^*	0.12	0.129	0.138	0.149	0.158	0.169	0.18	0.19	0.199	0.208	
miss T_2^{**}	0.121	0.133	0.141	0.152	0.163	0.174	0.182	0.191	0.200	0.210	

Hypothesis Testing

The three test statistics T, T_1 and T_2^* are compared in terms of the power of their tests. A Computer Approach Technique (CAT), given in Pal, Lim and Ling (2007), will be used. For a normal distribution all three test statistics perform relatively well. However, T_1 and T_2^* lost some power on the Laplace distribution, more on the T_1 than T_2^* . If X is a positively skewed distribution, such as Gamma, T_1 and T_2^* perform slightly better than T on the right side of μ_0 while T performs better than other two on the left side of μ_0 . The opposite is true for negatively skewed distribution. In terms of modified test statistics, T_2^* performs slightly better than T_1 from the point of view of power of the test. The simulation results for the power of the tests are summarized in Figures 11, 12, 13 and 14.

T-TEST ON SKEWED DISTRIBUTION FUNCTION

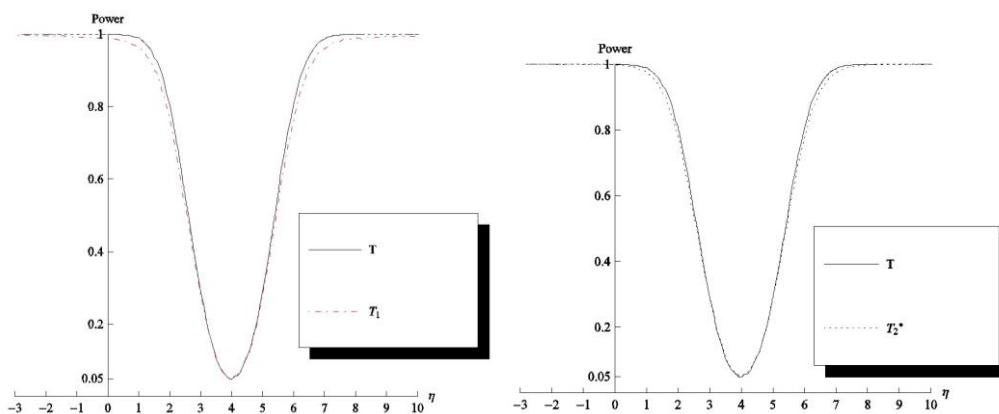


Figure 11. Plot of η versus power of the test for where $X_1, X_2, \dots, X_n \sim Normal(\eta, 2)$, with $n = 10$. The hypothesis testing is $H_0 : \eta = 4 (\mu = 4)$ versus $H_1 : \eta \neq 4 (\mu \neq 4)$

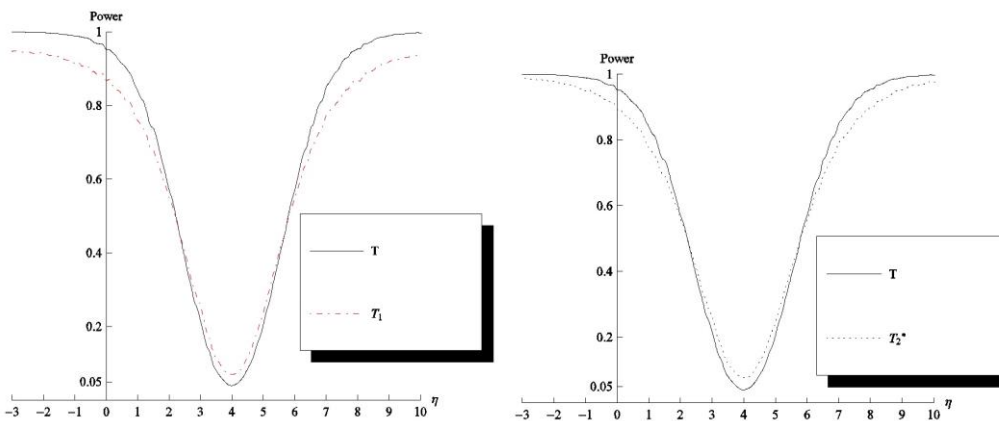


Figure 12. Plot of η versus power of the test for where $X_1, X_2, \dots, X_n \sim Laplace(\eta, 2)$, with $n = 10$. The hypothesis testing is $H_0 : \eta = 4 (\mu = 4)$ versus $H_1 : \eta \neq 4 (\mu \neq 4)$

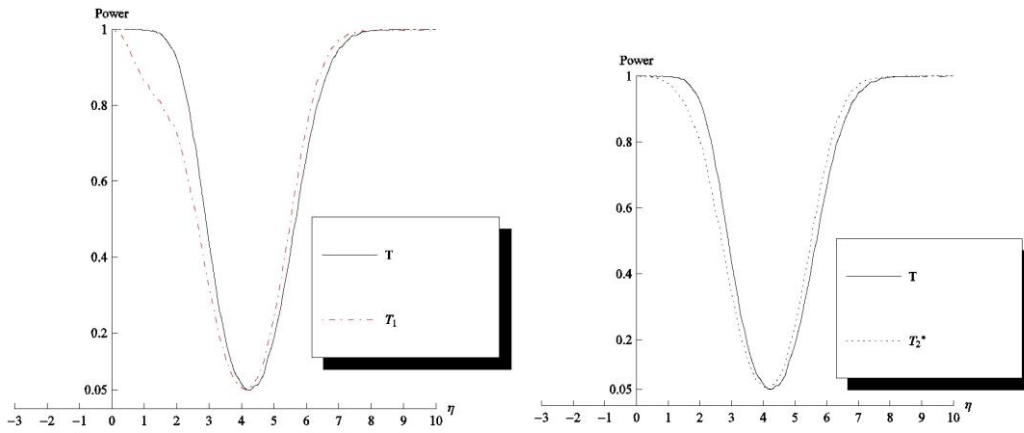


Figure 13. Plot of η versus power of the test for where $X_1, X_2, \dots, X_n \sim \text{Gamma}(\eta, 2)$, with $n = 10$. The hypothesis testing is $H_0 : \eta = 4 (\mu = 8)$ versus $H_1 : \eta \neq 4 (\mu \neq 8)$

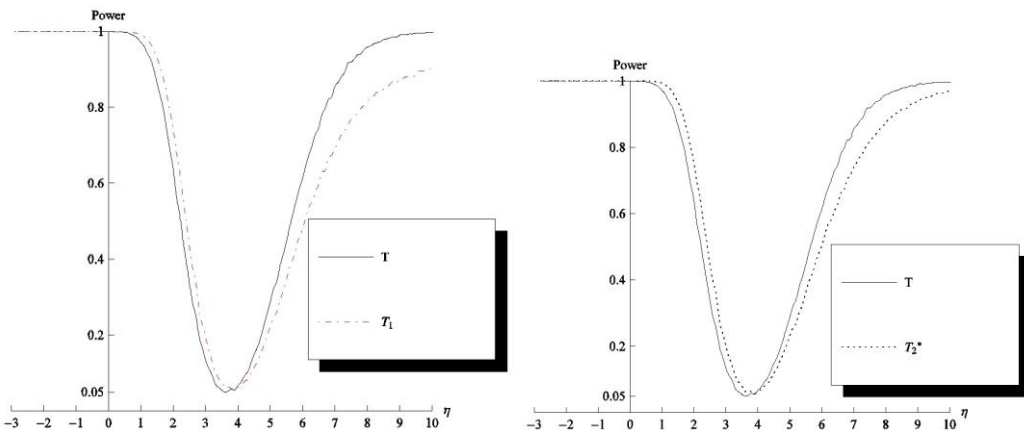


Figure 14. Plot of η versus power of the test for where $X_1, X_2, \dots, X_n \sim \text{Gumbel}(\eta, 2)$, with $n = 10$. The hypothesis testing is $H_0 : \eta = 4 (\mu = 2.84557)$ versus $H_1 : \eta \neq 4 (\mu \neq 2.84557)$

Conclusion

Based on these results, it appears that the usual t-test statistic, T , is quite robust regardless of the skewness of the distribution. The modified t-test statistics T_1 and

T-TEST ON SKEWED DISTRIBUTION FUNCTION

T_2^* can improve the power on one side of the $\mu = \mu_0$ only, but not on both sides. From the results, it appears the modified confidence intervals perform much better than the usual confidence interval derived from T when X is simulated from a skewed distribution.

Acknowledgements

The authors thank Eswar Phadia, Professor (Ret'd), William Paterson University of New Jersey for many helpful suggestions. Also, the authors thank the editor and referees for suggestions to improve the presentation of this article.

References

- Abramovitch, L. & Singh, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *The Annals of Statistics*, 13(1), 116-132.
doi: 10.1214/aos/1176346580
- Blair, R. C. & Sawilowsky, S. S. (1993). Comparison of two tests useful in situations where treatment is expected to increase variability relative to controls. *Statistics in Medicine*, 12(23), 2233-2243. doi: 10.1002/sim.4780122308
- Boos, D. D. & Hughes-Oliver, J. M. (2000). How large does n have to be for Z and t intervals. *The American Statistician*, 54(2), 121-128.
doi: 10.1080/00031305.2000.10474524
- DasGupta, A. (2008). *Asymptotic theory of statistics and probability*. Springer Science & Business Media.
- Hall, P. (1987). Edgeworth expansion for student's t statistic under minimal moment conditions. *Annals of Probability*, 15(3), 920-931.
doi: 10.1214/aop/1176992073
- Hall, P. (1992). On the removal of skewness by transformation. *Journal of Royal Statistical Society, Series B (Methodological)*, 54(1), 221-228.
- Johnson, N. J. (1978). Modified t -test and confidence intervals for asymmetrical population. *Journal of the American Statistical Association*, 73(363), 536-544. doi: 10.1080/01621459.1978.10480051
- Pal, N., Lim, W. K., & Ling, C. (2007). A computational approach to statistical inferences. *Journal of Applied Probability & Statistics*, 2(1), 13-35.
- Sawilowsky, S. S. & Blair, R. C. (1992). A more realistic look at the robustness and type II error properties of the t -test to departures from population

normality. *Psychological Bulletin*, 111(2), 352-360.

doi: 10.1037/0033-2909.111.2.352

Zhou, X.-H. & Gao S. (2000). One-sided confidence intervals for mean of positively skewed distributions. *The American Statistician*, 54(2), 100-104.

doi: 10.1080/00031305.2000.10474519