isid/ms/2007/08
June 20, 2007
http://www.isid.ac.in/~statmath/eprints

Sensitivity Analysis of the T - Distribution Under Truncated Normal Populations

DALE BOROWIAK and ASHISH DAS

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi-110 016, India

Sensitivity Analysis of the T - Distribution

Under Truncated Normal Populations

Dale Borowiak
Department of Statistics, University of Akron, Akron, Ohio 44325

and

Ashish Das
Stat-Math Division, Indian Statistical Insitute, New Delhi 110019

Abstract

In recent literature the truncated normal distribution has been used to model the stochastic structure for variety of random structures. In this paper the sensitivity of the t-random variable under a left truncated normal population is explored. Simulation results are used to assess the errors associated with estimating the reliabilities, over a range of locations, with the standard t distribution. The maximum errors are modeled as a linear function of the magnitude of the truncation and sample size. In the case of a left truncated normal population, adjustments to standard inferences for the mean, namely confidence intervals and observed significance levels, based on the t random variable are introduced.

Key Words and Phrases : Left Truncated Normal, t-Distribution, Skewness, Simulation

1. Introduction

The statistical modeling of physical and economic systems often involves nonnegative random variables. Statistical distributions with nonnegative supports such as the Gamma, Lognormal and the left-truncated normal have been used to model the stochastic structure associated with nonnegative random variables. Recently the left truncated normal random variable has been applied to these problems. Applications in the literature have been diverse and include management science (Johnson (2001)), forecasting (Thomopoules (1980)) and inventory systems (Sinha (1991)). Further, examples of stochastic modeling using the truncated normal random variable include the modeling of demand functions in the determination of safety stocks by Johnson (1978) and Johnson and Thomopoulos (2002) and the assessment of the technical efficiency of fishing vessels by Flores-Lagunes, Horrace and Schnier (2006).

In this paper the tail probabilities of the standard t-random variable when the underlying distribution is a left-truncated normal are explored through simulation experiments. Conservative inference procedures, namely confidence intervals and hypothesis tests for the mean, utilizing the standard t-distribution are constructed. The magnitude of the truncation and the sample size are varying parameters.

2. Left-Truncated Normal Distribution

Many statistical techniques in both the applied and theoretical settings are based on the classical normal distribution. The normal random variable X takes on all real values and has mean μ and variance $\sigma^2 > 0$. The standard normal random variable $Z = (X - \mu)/\sigma$ has probability density function $\phi(z) = (2\pi\sigma^2)^{-1/2} \exp(- z^2/2)$ and associated $\Phi(z) = P(Z \leq z)$.

The normal random variable can be left-truncated at any fixed quantity and the general distribution is given in Schneider (1986) and Cohen (1991). By a linear translation the truncation point can be shifted to zero and the probability density function can be written as

$$f(x, \mu, \sigma^2) = (2\pi\sigma^2)^{-½} \exp(- (x - \mu)^2/\sigma^2)/\Phi(\delta) \text{ for } x > 0 \text{ , } = 0 \text{ for } x \leq 0 \qquad (1)$$

for fixed $\mu > 0$, $\sigma^2 > 0$ and $\delta = \mu/\sigma$. The parameter δ is the distance between μ and the truncation point zero and measures the magnitude of the truncation. As δ increases the effects of truncation decrease and (1) approaches the classical normal distribution. The left-truncated normal random variable with probability density function (1) is denoted by LTNRV.

The mean and variance for general truncated normal random variables are given by Barr and Sherrill (1999). The probability density function of a LTNRV is skewed to the right and the central moments are functions of μ, $\sigma > 0$ and $\Delta(\delta) = \phi(\delta)/\Phi(\delta)$. Utilizing (1) the mean and variance are computed in Appendix A as

$$\mu(\delta) = \mu + \sigma\,\Delta(\delta) \quad \text{and} \quad \sigma^2(\delta) = \sigma^2\,(1 - \delta\Delta(\delta) - \Delta(\delta)^2) \qquad (2)$$

Further, in Appendix A an explicit form of the skewness, denoted $Sk(\delta) = E\{(X - \mu(\delta))^3\}/[\sigma^2(\delta)]^{3/2}$, is derived and takes the form

$$Sk(\delta) = (\delta^2\Delta(\delta) - \Delta(\delta) + 3\,\delta\,\Delta(\delta)^2 + 2\Delta(\delta)^3)/(1 - \delta\Delta(\delta) - \Delta(\delta)^2)^{3/2} \qquad (3)$$

As the magnitude of truncation decreases, manifested by an increasing δ, $\Delta(\delta)$ decreases and so does the skewness as given by (3).

3. Skewed T Random Variable and Simulation Results

The classical t- random variable introduced by Student (1908) has been used to construct exact inference procedures such as confidence intervals and hypothesis tests in the case of normal populations. The robustness of the t-random variable has been classically explored by Neyman and Pearson (1928). Nair (1941) observed that a positive population skewness resulted in a negatively skewed t-random variable. Johnson (1978) using a Cornish-Fisher expansion proposed some modified inference procedures concentrating to the case of a skewed population.

In this paper we explore conditions on the LTNRV under which the standard t distribution yields a good approximation. Let X_i be independent and identically distributed with pdf given by (1) for $1 \leq i \leq n$. We define the skewed t-random variable as

$$T(\delta) = n^{1/2} (\overline{X} - \mu(\delta)) / S \qquad (4)$$

where the sample mean and standard deviation are denoted by $\overline{X}$ and S, respectively. As δ increases and the magnitude of the truncation decreases and the skewed T random variable (4) approaches the classical t-random variable with degrees of freedom n – 1 denoted by T_{n-1}. This suggests that the classical t-distribution is an approximate distribution for T(δ) under suitable conditions. A measure of the efficiency of this approximation is found to be the skewness of the sample mean computed as

$$Sk(\delta,n) = Sk(\delta)/ n^{1/2} \qquad (5)$$

for truncation parameter δ and sample size n. Small values of (5) arise from either minor truncation or large sample size n and indicate closeness of the skewed T and classical t distributions.

We now look into the distribution of the skewed t-random variable for chosen values of δ and n. Through simulation the distribution of T(δ) in terms of their right tailed probabilities are computed. The resulting simulated tail probabilities or reliabilities are compared to the corresponding values for T_{n-1}.

Probabilities associated with the skewed T random variable, T(δ), are functions of both δ and n. Let R(t, δ, n) denote the reliability associated with t > 0, i.e.

$$R(t, \delta, n) = P(T(\delta) > t) \qquad (6)$$

These values are computed using simulation procedures consisting of 100,000 repetitions for chosen values t between -3.5 to 3.5. The simulated reliabilities R(t, δ, n) are listed in Appendix B. The graphs of the probability density functions for the standard t-distribution and the skewed t-distribution, resulting from midpoint approximations from the simulated reliabilities, are given in Fig. 1 for $\delta = 1$ and $n = 3$. As noted by Johnson (1978) the skewed t distribution is skewed to the left.

The simulated reliabilities are compared to the right tailed probabilities based on T_{n-1} denoted by R(t, n). The observed errors associated with chosen t values and parameters δ and n are defined as

$$e(t, \delta, n) = R(t, n) - R(t, \delta, n) \tag{7}$$

Consistent with the left skewness property of $T(\delta)$ as demonstrated in Fig. 1 the errors (7) are observed to be nonnegative. As a measure of the closeness of the two distributions we use the maximum error over the chosen t values in the range of -3.5 to 3.5 denoted by

$$e(\delta, n) = \max \{e(t, \delta, n)\} \tag{8}$$

Simulation results that relate the sample size n, skewness as measured by (3) and (5) and maximum observed errors are given in Table 1. We observe that the maximum error occurs at locations $-1.5 \leq t \leq 1.5$. Further, the observed errors at tail locations, namely $|t| \geq 2$ are 10% of the maximum error quantities.

From the quantities of Table 1 we observe that the skewness measures and maximum errors are directly related as they decrease together. A least squares fit is applied to the observed values of $(Sk(\delta,n), e(\delta,n))$ for values of δ and n given in Table 2. The least squares line and corresponding correlation coefficient are

$$e(\delta,n) = .002 + .109\, Sk(\delta,n) \quad \text{and} \quad r = .9862 \tag{9}$$

The large correlation indicates an efficient linear fit. Thus for a given magnitude of truncation δ and sample size n a maximum error is approximated by (9). In the next section relation (9) is used to obtain robust inference procedures based on the skewed T random variable.

In the case of a LTNRV the approximation of the skewed t distribution by the t distribution is efficient if the maximum errors are minimal. To explore the usefulness of (9) for given δ and n we can fix a desired maximum error at $\gamma > 0$ so that $e(\delta, n) \leq \gamma$. Putting (5) into (9) and solving for n yields

$$n^{1/2} \geq .109\ Sk(\delta) / (\gamma - .002) \quad (10)$$

For demonstration, fixing $\gamma = .01$, we find the smallest integers, denoted by n* that satisfy (10) for various values of δ. For these sample sizes simulations are run and the results in terms of the maximum errors are listed in Table 2. From Table 2 we observe that as desired (10) yields maximum errors of about 1%. The tail errors, $|t| \geq 2$ are observed to be considerably less than 1%.

4. Robust Inference Procedures

Applying the standard t distribution to inferences involving the skewed t random variable result in robust techniques when the maximum error (8) is minimal. For fixed constant $\gamma > 0$ we assume δ and n are such that $e(\delta, n) \leq \gamma$ so that

$$0 \leq R(t_j,n) - R(t_j,\delta,n) \leq \gamma \quad (11)$$

For $p > 0$, let $t_{p,n-1}$ be such that $R(t_{p,n-1}, n) = p$. If (11) holds where $p > \gamma$ then

$$p - \gamma \leq R(t_p, \delta, n) \leq p \quad (12)$$

In the case of LTNRV, inequality (12) is used to construct robust inference procedures. In the case of hypothesis testing for $\mu(\delta)$ using test statistic $T(\delta)$ the robust p-value for right tailed tests applies the t-distribution directly while the left tail p-value is increased by γ.

A confidence interval for the mean of a LTNRV, $\mu(\delta)$, is constructed using $T(\delta)$ as the pivot. For confidence coefficient $1 - \alpha$ we assume (12) holds where $\alpha/2 > \gamma$. The robust confidence interval takes the form

$$\bar{X} - t_{\alpha/2,n-1}\, S/n^{1/2} \leq \mu(\delta) \leq \bar{X} + t_{\alpha/2 - \gamma,\, n-1}\, S/n^{1/2} \qquad (13)$$

Note that the confidence interval (13) is not symmetric about the sample mean and is slightly wider than the confidence interval assuming no truncation.

Illustration : The inference techniques presented in this section are demonstrated using data taken from the fishing industry. Statistical frontier models used to estimate the efficiency of fishing vessels result in nonnegative technical efficiency ratings (see Aigner, Lovell and Schmidt (1977) and Battesse and Coelli (1988)). The technical efficiencies of 39 vessels in the North Atlantic from 2000 and 2003 list in Table 2 of Flores-Lagunes, Horrace and Schnier (2006)) are utilized. The sample has a mean of .393 and standard deviation of .301. Using $\delta = 1$ from (3) we find Sk(1) = .592 and using (5) Sk(1,39) = .095. Applying the least squares line (9) we have maximum error e(1, 39) = .012. A 95% confidence interval for the mean given by (13) is computed using $\alpha/2 = .025$ and $\gamma = .012$ as $.2805 \leq \mu(1) \leq .5186$. We remark that this confidence interval is approximately 5% wider than the confidence interval ignoring truncation.

5. Conclusion

For a LTNRV right tail probabilities associated with the t-random variable are a function of the magnitude of truncation and sample size. Based on simulation results the effects of approximating tail probabilities with the standard t distribution are measured. The effects are found to be minimal (maximum errors of 1% and tail errors of about .1%) for sample sizes of 30 or more and $\delta \geq 1.5$. If the magnitude of the truncation is smaller, say $\delta \geq 2$, then the minimal sample size for 1% errors or less reduces to 10.

Reference

Aigner, D.J., Lovell, A.K. and Schmidt, P. (1977) Formulation and Estimation of Stochastic Frontier Production Function Models, *Journal of Econometrics*, 6, 21-37.

Barr, D.R. and Sherrill S.A. (1999). Mean and Variance of Truncated Normal Distributions, *The American Statistician*, 53, 357-361.

Battesse, G.E. and Coelli, T.J. (1988). Production of Firm-Level Technical Efficiencies with a Generalized Frontier production Function and Panel Data. *Journal of Econometrics*, 38, 387-399.

Cohen, A.C. (1991) **Truncated and Censored Samples**, Marcel Dekker, New York, NY.

Flores-Lagunes, A.F., Horrace, W.C. and Schnier, K. E. (2006) Identifying Technically Efficient Fishing Vessels : A Non-Empty Minimal Subset Approach, Working Paper, University of Arizona, Arizona, USA.

Johnson, A.C. (2001) On the Truncated Normal Distribution : Characteristics of Singly and Doubly Truncated Populations of Application in Management Science, PhD Dissertation, Stuart School of Buisness, Illinois Institute of Technology, Illinois.

Johnson, N.J. (1978) Modified t Test and Confidence Intervals for Asymmetrical Populations. *JASA*, Vol 73, No. 363, 536-544.

Johnson, A.C. and Thomopoulos, N.T. (2002) Use of the Left-Truncated Normal Distribution for Improving Achieved Service Levels, Decision Sciences Institute 2002 Annual Meeting

Nair, A.K.N. (1941). Distribution of the Student's t in the Correlation Coefficient in Samples from Non-Normal Populations, *Sankhya*, 5, 383-400.

Neyman, J. and Pearson, E.S. (1928). On the Use and Interpertation of Certain Test Criteria for Purposes of Statistical Inference Part I, *Biometrika*, A20, 175-240.

Schneider, H. (1986) **Truncated and Censored Samples From Normal Populations**, Marcel Dekker, New York NY.

Sinha, D. (1991). Optimal Policy Estimation for Continuous Review Inventory Systems, *Computers in Operations Research*, 18, 487-496.

“Student” (1908), The Probable Error of the Mean, *Biometrika*, 6, 1-25.

Thomopoulos, N.T. (1980). **Applied Forcasting Methods**, Prentice – Hall, Englewood Cliffs, New Jersey.

Appendix A : Moment Computations of the LTNRV.

Direct integration using (1) gives $E\{X - \mu\} = \sigma\,\Delta(\delta)$ so that the mean $\mu(\delta)$ in (2) holds. To compute the variance we applying a binomial expansion

$$E\{(X - \mu(\delta))^2\} = E\{(X - \mu)^2\} + \sigma^2\,\Delta(\delta)^2 - 2\,\sigma\,\Delta(\delta)\,E\{X - \mu\}$$

Integration by parts applied gives $E\{(X - \mu)^2\} = \sigma^2 - \delta\,\sigma^2\,\Delta(\delta)$ and the variance in (2) follows. To find the third central moment we again expand the expectation as

$$E\{(X - \mu(\delta))^3\} = E\{(X - \mu)^3\} - 3\,\sigma\,\Delta(\delta)\,E\{(X - \mu)^2\} + 3\,\sigma^2\,\Delta(\delta)^2 - \sigma^3\,\Delta(\delta)^3$$

Applying integration by parts gives $E\{(X - \mu)^3\} = \sigma^3\,\delta^2\,\Delta(\delta) + 2\,\sigma^3\,\Delta(\delta)$ and the third central moment is

$$E\{(X - \mu(\delta))^3\} = \delta^2\Delta(\delta) - \Delta(\delta) + 3\,\delta\,\Delta(\delta)^2 + 2\Delta(\delta)^3$$

From these the skewness formula (3) is verified.

Appendix B : LTNRV Simulated Reliabilities Based On 100,000 Repetitions.

Simulated Reliabilities R(t,δ,n)

Observed Locations T = t

n	δ	-3.5	-3	-2.5	-2	-1.5	-1	-.5	0	.5	1.0	1.5	2	2.5	3	3.5
5	3	.987	.980	.966	.941	.894	.811	.676	.499	.320	.185	.103	.058	.034	.020	.012
5	2.5	.986	.978	.964	.938	.892	.808	.676	.496	.317	.182	.100	.057	.032	.019	.012
5	2.25	.986	.977	.962	.937	.889	.806	.672	.493	.315	.180	.099	.055	.031	.019	.012
5	2	.984	.976	.960	.933	.887	.802	.670	.493	.311	.175	.095	.053	.030	.017	.011
5	1.75	.982	.972	.957	.929	.881	.800	.668	.490	.308	.170	.091	.049	.028	.016	.010
5	1.5	.980	.970	.953	.925	.877	.795	.666	.488	.304	.165	.088	.047	.026	.015	.010
5	1.25	.977	.966	.948	.919	.871	.789	.661	.486	.299	.161	.083	.044	.024	.014	.008
5	1	.973	.961	.943	.912	.864	.783	.657	.480	.296	.156	.077	.039	.021	.012	.007
10	3	.996	.992	.982	.960	.915	.826	.683	.496	.313	.171	.084	.037	.016	.007	.003
10	2.5	.996	.992	.981	.958	.913	.825	.683	.496	.309	.167	.080	.036	.016	.007	.003
10	2.25	.996	.991	.980	.957	.910	.823	.678	.493	.309	.165	.078	.035	.015	.006	.003
10	2	.995	.990	.979	.955	.907	.819	.678	.494	.306	.163	.076	.033	.014	.006	.003
10	1.75	.994	.988	.976	.952	.904	.817	.677	.493	.303	.160	.073	.032	.013	.006	.002
10	1.5	.993	.987	.975	.950	.900	.814	.675	.490	.302	.155	.070	.029	.012	.005	.002
10	1.25	.992	.985	.972	.946	.898	.811	.673	.489	.300	.151	.065	.027	.011	.004	.002
10	1	.990	.983	.968	.942	.893	.807	.671	.486	.296	.148	.064	.025	.009	.003	.001
15	3	.998	.995	.986	.967	.920	.831	.685	.497	.312	.165	.076	.032	.012	.004	.002
15	2.5	.998	.994	.986	.965	.919	.830	.684	.497	.310	.162	.075	.031	.012	.004	.001
15	2.25	.998	.994	.986	.965	.917	.828	.684	.496	.305	.161	.073	.030	.011	.004	.001
15	2	.998	.993	.984	.961	.915	.830	.683	.494	.310	.159	.071	.030	.011	.004	.001
15	1.75	.997	.992	.982	.960	.912	.823	.679	.493	.304	.157	.069	.027	.010	.003	.001
15	1.5	.996	.991	.980	.957	.910	.820	.679	.493	.304	.154	.066	.024	.009	.003	.001
15	1.25	.996	.991	.979	.954	.905	.818	.677	.492	.299	.151	.063	.023	.008	.003	.001
15	1	.994	.988	.976	.952	.904	.815	.675	.489	.300	.148	.062	.022	.007	.002	.001
20	3	.999	.996	.988	.969	.924	.833	.686	.497	.310	.163	.074	.029	.011	.004	.001
20	2.5	.999	.996	.988	.968	.923	.832	.685	.497	.306	.160	.071	.028	.010	.003	.001
20	2.25	.998	.996	.987	.966	.921	.830	.683	.494	.305	.160	.071	.027	.009	.003	.001
20	2	.998	.995	.987	.965	.919	.828	.683	.494	.304	.158	.069	.026	.009	.003	.001
20	1.75	.998	.994	.985	.963	.916	.826	.680	.493	.303	.154	.066	.025	.008	.003	.001
20	1.5	.997	.994	.984	.962	.914	.824	.678	.492	.301	.154	.065	.023	.008	.002	.001
20	1.25	.997	.993	.982	.960	.911	.821	.677	.492	.300	.150	.062	.022	.007	.002	.001
20	1	.996	.991	.981	.957	.909	.818	.677	.489	.300	.149	.059	.020	.006	.002	.000
25	3	.999	.997	.990	.970	.924	.834	.686	.496	.309	.163	.073	.029	.010	.003	.001
25	2.5	.999	.996	.989	.970	.925	.832	.687	.497	.307	.159	.070	.027	.009	.003	.001
25	2.25	.999	.996	.989	.969	.922	.831	.685	.496	.306	.158	.069	.026	.009	.002	.001
25	2	.999	.996	.988	.967	.921	.830	.684	.495	.305	.157	.068	.025	.008	.002	.001
25	1.75	.998	.995	.987	.966	.919	.829	.682	.494	.304	.157	.066	.023	.007	.002	.001
25	1.5	.998	.995	.986	.964	.917	.826	.682	.492	.302	.154	.065	.023	.007	.002	.000
25	1.25	.998	.994	.985	.963	.914	.823	.680	.493	.300	.150	.062	.021	.006	.002	.000
25	1	.997	.993	.984	.960	.912	.821	.678	.491	.299	.149	.060	.020	.006	.001	.000

Table 1 : Maximum Error Simulation Results For Values of Skewness and Sample Size

	n = 5		n = 10		n = 15		n = 20		n = 25	
δ	Sk(δ,n)	e(δ,n)	Sk(δ,n)	e(δ,n)	Sk(δ,n)	e(δ,n)	Sk(δ,n)	e(δ,n)	Sk(δ,n)	e(δ,n)
1	.265	.031	.187	.024	.153	.019	.132	.017	.118	.015
1.25	.219	.023	.155	.018	.126	.015	.109	.017	.098	.013
1.5	.175	.022	.124	.017	.101	.013	.087	.011	.078	.011
1.75	.134	.015	.095	.012	.078	.010	.067	.010	.060	.008
2	.099	.012	.070	.009	.057	.009	.049	.007	.044	.007
2.25	.069	.007	.049	.007	.040	.006	.034	.005	.031	.006
2.5	.046	.005	.032	.006	.026	.005	.023	.006	.020	.005
3	.016	.003	.012	.002	.009	.002	.008	.003	.007	.003

Table 2 : Sample Size and Simulations for 1% Maximum Errors

	δ = 1.0	δ = 1.25	δ = 1.5	δ = 1.75	δ = 2.0	δ = 2.25	δ = 2.5
Sk(δ)	.592	.489	.391	.301	.221	.154	.102
n*	66	45	29	17	10	5	2
e(δ,n)	.009	.009	.011	.011	.009	.007	.008

Fig 1 : Probability Density Functions For T(δ =1) —— and T -----