J. King Saud Univ., Vol. 11, *Admin. Sci.* (2), pp. 51-58 (A.H. 1419/1999)

Estimation in the Doubly Truncated Logistic Distribution with Unknown Truncation Points

M. H. Al Yousef
Associate Professor. Department of Quantitative Methods
College of Administrative Sciences, King Saud University
Riyadh, Saudi Arabia

(Received 14/7/1417H; accepted for publication 6/11/1418H)

Abstract. In this article we discuss the problem of estimating the parameters of the doubly truncated logistic distribution when truncation points are unknown. The parameters of the distribution and the truncation points will be estimated using the method of maximum likelihood estimation. Asymptotic sampling errors of estimates have been carried out. Finally a numerical example is given to illustrate the results.

Introduction

Truncated distributions arise when sample selection and / or observation is not possible in some sub-region of the sample space. This can occur as a consequence of actual elimination of part of the original data .

Charernkavanich and Cohen [1] discussed this problem for complete samples with a variety of estimation problems involving truncated normal, gamma, Weibull, lognormal and various other truncated distributions. Bain and Weeks [2], and Deemer and Votaw [3] gave the main results in the case of truncated exponential distribution using censored data, Shalaby [4] and Al Yousef [5] have discussed the problem in the case of Weibull and Gompertz distributions .

The present study is concerned with a logistic distribution that is doubly truncated at unknown truncation points. The truncation points then become additional parameters which must be estimated from sample data along with the primary distribution parameters.

The Logistic Distribution

The logistic distribution is an appropriate model for the distribution of time - to - failure of creatures or systems with increasing hazard rate. Bennett [6] has used this model in his studies and Johnson and Johnson [7, p.63] have shown that the logistic distribution is used in survival analysis.

The probability density function (p.d.f) of this distribution can be written as:

$$f(x;a,b) = b\exp(a-bx)\,[1+\exp(a-bx)]^{-2} \tag{2.1}$$

for $-\infty < x < \infty$, $b > 0$. The corresponding cumulative distribution function (c.d.f) is:

$$F(x;a,b) = [1+\exp(a-bx)]^{-1} \tag{2.2}$$

and the characteristic function is:

$$\Phi(t) = 2\frac{\pi}{b}t\left[e^{\frac{t}{b}(\pi-ai)} - e^{-\frac{t}{b}(\pi+ai)}\right]^{-1}, t \neq 0$$

$$\Phi(0) = 1 \tag{2.3}$$

The mean μ_1' and the variance μ_2 of this distribution are

$$\mu_1' = \frac{a}{b} \text{ and } \mu_2 = \mu_2' - \mu_1'^2 = \frac{\pi^2}{3b^2}$$

The distribution of X is symmetrical about the mean and all odd central moments are equal to zero. The even central moments can be obtained using the following equation.

$$\mu_{2k} = 2(2k)!\,b^{-2k}(1-2^{1-2k})\varsigma(2k) \tag{2.4}$$

where $\varsigma(Z) = \sum_{k=1}^{\infty} k^{-2}$ is the Riemann - Zeta function [8, p.1073].

The shape factors α_1 and α_2 are obtained as:

$$\alpha_1 = \mu_3/\mu_2^{3/2} = 0 \text{ and } \alpha_2 = \mu_4/\mu_2^2 - 3 = 1.2$$

Doubly Trucated Logistic Distribution

When the distibution with p.d.f. (2.1) is doubly truncated over the interval [t_1, t_2,], the resulting truncated distribution becomes

$$f_T(x;t_1,t_2,a,b) = f(x;a,b).[F(t_2;a,b) - F(t_1;a,b)]^{-1} \tag{3.1}$$

for $t_1 \leq x \leq t_2$ (0 otherwise) . And it follows from (2.1) and (2.2) that

$$f_T(x;t_1,t_2,a,b) = b(v-u)^{-1}\exp(a-bx)[1+\exp(a-bx)]^{-2} \tag{3.2}$$

for $t_1 \leq x \leq t_2$, zero elsewhere . where

$$u = [1+\exp(a-bt_1)]^{-1} \text{ and } v = [1+\exp(a-bt_2)]^{-1} \tag{3.3}$$

The likelihood function of a random sample of size n from a truncated distribution with p.d.f. (3.2) becomes

$$L(x_1,\ldots,x_n;t_1,t_2,a,b) = b^n(v-u)^{-n}\prod_{j=1}^{n}\exp(a-bx_j)[1+\exp(a-bx_j)]^{-2} \tag{3.4}$$

Maximum Likelihood Estimation (MLE)

On taking logarithms of (3.4) and differentiating we have

$$\frac{\partial \ln L}{\partial t_1} = nbu(1-u)/(v-u)$$
$$\frac{\partial \ln L}{\partial t_2} = -nbv(1-v)/(v-u)$$
$$\frac{\partial \ln L}{\partial a} = n(1-u-v) + \sum_{j=1}^{n}[1-\exp(a-bx_j)][1+\exp(a-bx_j)]^{-1}$$
$$\frac{\partial \ln L}{\partial b} = n/b - n[v(1-v)t_2 - u(1-u)t_1]/(v-u) -$$
$$\sum_{j=1}^{n} x_j[1-\exp(a-bx_j)][1+\exp(a-bx_j)]^{-1} \tag{3.5}$$

The function (3.4) attains its maximum value when t_1 is as large as possible and t_2 is as small as possible.

Since $t_1 \leq x \leq t_2$ we then have

$$\hat{t}_1 = y_1 \text{ and } \hat{t}_2 = y_n \tag{3.6}$$

where y_1 is the smallest sample observation and y_n is the largest sample observation; i.e. y_1 is the first order statistic and y_n is the n th order statistic in a random sample of size n. Estimators for a and b follow as solutions of equations:

$$n(1-\hat{u}-\hat{v})+\sum_{j=1}^{n}[1-\exp(\hat{a}-\hat{b}x_j)][1+\exp(\hat{a}-\hat{b}x_j)]^{-1}=0 \tag{3.7}$$

$$n/\hat{b}-n[\hat{v}(1-\hat{v})y_n-\hat{u}(1-\hat{u})y_1]/(\hat{v}-\hat{u})-$$
$$\sum_{j=1}^{n}x_j[1-\exp(\hat{a}-\hat{b}x_j)][1+\exp(\hat{a}-\hat{b}x_j)]^{-1}=0$$

where: $\hat{u}=[1+\exp(\hat{a}-\hat{b}y_1)]^{-1}$ and
$\hat{v}=[1+\exp(\hat{a}-\hat{b}y_n)]^{-1}$

The solution of the set of equations given in (3.7) can be obtained using iterative techniques for solving a pair of simultaneous non-linear equations in two unknowns.

Sampling errors

The p.d.f of the r th order statistic in random sample of size n from the truncated distribution (3.2) is

$$g(y_r;t_1,t_2,a,b)=br\binom{n}{r}(v-u)^{-n}\{[1+\exp(a-by_r)]^{-1}-u\}^{r-1}\cdot \tag{3.8}$$
$$\{v-[1+\exp(a-by_r)]^{-1}\}^{n-r}\cdot\exp(a-by_r)[1+\exp(a-by_r)]^{-2}$$

$t_1 \leq y_r \leq t_2$, zero elsewhere.

if $r = 1$ and n, the p.d.f. of the first and last order statistic may be obtained respectively as

$$g(y_1;t_1,t_2,a,b)=bn(v-u)^{-n}\{v-[1+\exp(a-by_1)]^{-1}\}^{n-1}\cdot \tag{3.9}$$
$$\exp(a-by_1)[1+\exp(a-by_1)]^{-2}$$

and

$$g(y_n;t_1,t_2,a,b)=bn(v-u)^{-n}\{[1+\exp(a-by_n)]^{-1}-u\}^{n-1}\cdot \tag{3.10}$$
$$\exp(a-by_n)[1+\exp(a-by_n)]^{-2}$$

it can be shown that the first two moments of (3.9) and (3.10) are

$$E(Y_1) = \frac{a}{b} + J_1, E(Y_1^2) = \frac{a^2}{b^2} + \frac{2a}{b} J_1 + J_2 \tag{3.11}$$

and

$$E(Y_n) = \frac{a}{b} + L_1, E(Y_n^2) = \frac{a^2}{b^2} + \frac{2a}{b} L_1 + L_2 \tag{3.12}$$

where

$$J_r = nb^{-r}(v-u)^{-n} \int_u^v (v-z)^{n-1} \left[\ln\left(\frac{z}{1-z} \right) \right]^r dz; r = 1,2 \tag{3.13}$$

and

$$L_r = nb^{-r}(v-u)^{-n} \int_u^v (z-u)^{n-1} \left[\ln\left(\frac{z}{1-z} \right) \right]^r dz; r = 1,2 \tag{3.14}$$

The variance of the MLE (3.6) can be found as

$$V(\hat{t}_1) = J_2 - J_1^2 \tag{3.15}$$

and

$$V(\hat{t}_2) = L_2 - L_1^2 \tag{3.16}$$

The asymptotic variance - covariance matrix of the MLE of $\hat{a}$ and $\hat{b}$ can be expressed as

$$V(\hat{a}, \hat{b}) = [Z_{ij}]^{-1}; i, j = 1,2 \tag{3.17}$$

where:

$$Z_{11} = -E\left(\frac{\partial^2 \ln L}{\partial a^2} \right)$$

$$Z_{12} = Z_{21} = -E\left(\frac{\partial^2 \ln L}{\partial a \partial b} \right)$$

$$Z_{22} = -E\left(\frac{\partial^2 \ln L}{\partial b^2} \right)$$

The exact expressions for the expectations in the above matrix are difficult to obtain. However an approximate estimate for variance - covariance matrix can be obtained using the approximation of Cohen [9], namely:

$$Z_{11} = -\frac{\partial^2 \ln L}{\partial a^2} \bigg|_{a=\hat{a}, b=\hat{b}}$$

$$Z_{12} = -\frac{\partial^2 \ln L}{\partial a \partial b} \bigg|_{a=\hat{a}, b=\hat{b}}$$

$$Z_{22} = -\frac{\partial^2 \ln L}{\partial b^2} \bigg|_{a=\hat{a}, b=\hat{b}}$$

These approximations are derived to be:

On inverting the matrix (3.17), it follows that

$$Z_{11} = -n[\hat{u}(1-\hat{u}) + \hat{v}(1-\hat{v})] + 2\sum_{j=1}^{n} \exp(\hat{a} - \hat{b}x_j)/[1+\exp(\hat{a} - \hat{b}x_j)]^2$$

$$Z_{12} = n[\hat{u}(1-\hat{u})y_1 + \hat{v}(1-\hat{v})y_n] - 2\sum_{j=1}^{n} x_j \exp(\hat{a} - \hat{b}x_j)/[1+\exp(\hat{a} - \hat{b}x_j)]^2$$

$$Z_{22} = (n/\hat{b}^2)[(n/\hat{b}^2) + n[\hat{v}(1-\hat{v})(1-2\hat{v})y_n^2 - \hat{u}(1-\hat{u})(1-2\hat{u})y_1^2]/(\hat{v}-\hat{u}) - n[\hat{v}(1-\hat{v})y_n - \hat{u}(1-\hat{u})y_1]^2/(\hat{v}-\hat{u})^2 + 2\sum_{j=1}^{n} x_j^2 \exp(\hat{a} - \hat{b}x_j)/[1+\exp(\hat{a} - \hat{b}x_j)]^2 \quad (3.18)$$

On inverting the matrix (3.17), it follows that

$$V(\hat{a}) = Z_{22}/(Z_{11}Z_{22} - Z_{12}^2)$$
$$V(\hat{b}) = Z_{11}/(Z_{11}Z_{22} - Z_{12}^2) \quad (3.19)$$
$$\text{cov}(\hat{a}, \hat{b}) = -Z_{12}/(Z_{11}Z_{22} - Z_{12}^2)$$

Illustrative example

The practical application of estimators resulting in this work are illustrated with simulated data from the doubly truncated logistic distribution. The numerical techniques and iteration processes of IMSL routines are used. In this example we have a random sample consisting of 20 observations from a doubly truncated logistic population in which $t_1=10$, $t_2=50$, a=1.8 and b=0.06. The individual observations are listed in Table 1, for this sample n=20, $\overline{X}=30$, S=10.4; $y_1=11.6$ and $y_n=48.4$. For the simulated data of Table 1, MLE's and asymptotic variances are calculated and summarized in Table 2.

Table 1. A random sample of 20 obsevations from f_T $(x;t_1,t_2,a,b) = f_T(x;10,50,1.8,0.06)$

45.7	48.4	32.4	18.8	20.8
41.2	35.7	14.3	29.2	11.6
39.2	26.0	24.3	27.6	16.7
34.0	37.4	22.6	43.3	30.8

Table 2. Estimates from doubly truncated logistic distribution

Parameter	MLE	Asymptotic variance
t_1	11.6	7.113412
t_2	48.4	7.113412
a	1.462741	0.941550
b	0.048758	0.009518

It is of interest to note the following special cases:

(i) if $t_1 \to -\infty$, the doubly truncated logistic distribution reduces to the right truncated logistic distribution.

(ii) if $t_2 \to +\infty$, the doubly truncated logistic distribution reduces to the left truncated logistic distribution.

(iii) the likelihood function when the parent distribution is the two parameters logistic can be obtained as a special case from (3.4) if $v = 1$ and $u = 0$.

References

[1] Charernkavanich, D. and Cohen, A.C. "Estimation in the Singly Truncated Weibull Distribution with an Unknown Truncation Point." *Commun. Stat.Theor. Math,* 13(1984), 843-857.

[2] Bain, L.J. and Weeks, D.L. "A Note on the Truncated Exponential Distribution." *Ann. Math. Stat,* 35 (1964), 1366-1367.

[3] Deemer, W. L. and Votaw, D.F. "Estimation of Parameters of Truncated or Censored Exponential Distribution." *Ann. Math. Stat,* 26 (1955), 498-504.

[4] Shalaby, O. A. "Estimation in Truncated Weibull Distribution." *J. King Saud Univ.* Vol 2, *Admin. Sci.* (1), (1990), 25-40.

[5] Al Yousef, M.H. "Estimation of Parameters and Truncation Point for the Truncated Gompertz Distribution." *J. King Saud Univ.* Vol. 5, *Admin. Sci.* (2), (1993), 35-48.

[6] Bennett, S. "Log-Logistic Regression Models for Survival Data." *App. Stat.,* 32, No. 2 (1983), 165-171.

[7] Johnson, E.R.C. and Johnson, N.L. *Survival Models and Data Analysis.* J. Wiley: New York, (1980).

[8] Gradshteyn, I.S. and Ryshik, T.M. *Tables of Integrals, Series, and Products.* New York: Academic Press, (1980).

[9] Cohen, A.C. "Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and Censored Sample." *Tech,* 7 (1965), 579-588.

تقدير معالم التوزيع اللوجستي المبتور من الطرفين عندما تكون نقطتا البتر غير معلومتين

محمود حسين اليوسف

أستاذ مشارك، قسم الأساليب الكمية، كلية العلوم الإدارية، جامعة الملك سعود، الرياض، المملكة العربية السعودية

(قدم للنشر في ١٤/ ٧/ ١٤١٧هـ، وقبل للنشر في ٦/ ١١/ ١٤١٨هـ)

ملخص البحث . إن هذا البحث يناقش مشكلة تقدير معالم التوزيع اللوجستي المبتور من الطرفين عندما تكون نقطتا البتر غير معلومتين، إذ يمكن اعتبارهما عندئذ معلمتين إضافيتين يجب تقديرهما . ولقد استخدم الباحث في دراسته طريقة الإمكان الأعظم في التقدير للوصول إلى المقدرات اللازمة لتقدير المعالم الأساسية للتوزيع اللجستي بالإضافة إلى نقطتي البتر، كما تناولت الدراسة أيضاً خصائص هذه المقدرات، وأخيراً قام الباحث باختبار النتائج على بيانات مأخوذة من التوزيع اللوجستي المبتور باستخدام الحاسب الآلي .