MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS OF A NORMAL DISTRIBUTION FROM DOUBLY TRUNCATED SAMPLES: A COMPUTER APPROACH

Ahmed Saeed Akhter
Department of Statistics, Islamia University, Bahawalpur

Abstract

In this paper the parameters of the truncated normal distribution are estimated by the method of maximum likelihood. The maximum likelihood equations require the use of special tables. Here these equations are solved by the computer.

Introduction

The problem estimation from a truncated normal distribution has been treated by several authors. Two different kinds of problem arise, depending on whether the population from which the sample is drawn is truncated or sample itself is truncated, the population being complete. To distinguish between the two cases Hald (1949) called them truncated and censored respectively. The population can be singly truncated at known truncation point or doubly truncated about two known truncation points. Hald (1949) Cohen (1949) and Douglas (1983) studied singly truncated normal samples with a known truncation point. Cohen (1950) considered doubly truncated normal samples. In three latter papers Cohen (1955, 1959, 1961) extended the results given in his 1950 paper. In these studies equations were derived with which maximum likelihood estimates of the population mean and variance could be obtained. Practical application of the various equations involves use of special tables which may frequently be unavailable. Cohen (1957) used graphic solution to estimate the equations for doubly truncated samples.

The maximum likelihood estimating equations are derived here to estimate the mean and variance of a normal distribution from doubly truncated samples, and are solved simultaneously by computer.

Maximum Likelihood Estimates of the parameters in the truncated normal distribution for ungrouped data

A random variable has a doubly truncated normal distribution if its probability density funcation is

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} \left[\frac{1}{\sigma\sqrt{2\pi}} \int_A^B -1/2\left(\frac{t-\mu}{\sigma}\right)^2 dt\right]^{-1}$$

$$= \sigma^{-1} z\left(\frac{x-\mu}{\sigma}\right)\left[\phi\left(\frac{B-\mu}{\sigma}\right) - \phi\left(\frac{A-\mu}{\sigma}\right)\right]^{-1} \quad \ldots\ldots (1)$$

The lower and upper truncation points are A,B respectively; the degree of truncation are $\phi\left(\frac{A-\mu}{\sigma}\right)$ (from below) and $\left(1 - \phi\left(\frac{b-\mu}{\sigma}\right)\right)$ (from above)

Let $\xi = \frac{x-\mu}{\sigma}$

$$\phi(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2}$$

$$\phi(\xi) = \int_{-\infty}^{\xi} \phi(\xi)\, d\xi$$

and

$$\xi_1 = \left(\frac{A-\mu}{\sigma}\right), \xi_2 = \left(\frac{B-\mu}{\sigma}\right)$$

Then the logarithm of the likelihood function is

$$L = -n \ln[\phi(\xi_2) - \phi(\xi_1)] - n \ln\sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 + C$$

$$\frac{\partial L}{\partial \mu} = -\frac{n}{\sigma} \frac{[\phi(\xi_2) - \phi(\xi_1)]}{[\phi(\xi_2) - \phi(\xi_1)]} + \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 = 0 \quad \ldots\ldots (2)$$

$$\frac{\partial L}{\partial \mu} = -\frac{n}{\sigma} \frac{[\phi(\xi_2) - \phi(\xi_1)]}{[\phi(\xi_2) - \phi(\xi_1)]} - \frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - \mu)^3 = 0 \quad \ldots\ldots (3)$$

The maximum likelihood (ML) equations (2) and (3) are solved are here by computer.

Numerical Example

The following random sample of size 20 is taken from normal distribution with $\mu = 0.0$, $\sigma = 1.0$. The sample is doubly truncated at -2.0 and 2.0.

-1.8742	-0.2910	-1.2939	0.4677	-0.2225
1.4371	-0.0627	-1.1778	0.3557	1.3508
0.4168	0.3828	1.4894	-0.3982	0.3722
-0.9462	0.0308	0.9902	0.0816	0.2671

By using the given computer program we obtain $\hat{\mu} = 0.900$, $\hat{\sigma} = 1.0123$.

Computer Program to Solve the ML Equations

Description

The subroutine SOLU(START1, START2, EMEAN, SIGMA) is designed to solve the ML equaitons. Its parameters are defined as under:
START1: Guess value for mean
START2: Guess value for standard deviation
EMEAN: Estimated value of populaiton mean
SIGMA: Estimated value of population standard deviation
NAG routines CO5NCF and S15ABF are used.

Implementation

The user should supply a main program with the following declarations and values of AX, NT, T1, T2.
DIMENSION AX(1000)
COMMON /BLK1/AX,NT,T1,T2
where
AX: Data points up to maximum 1000 observations.
(for more than 1000 observations make changes for dimension of AX in the given subroutine accordingly)
NT: Total number of observations
T1: Left truncation point
T2: Right truncation point

```
C***** SUBROUTINE TO SOLVE THE LH EQUATIONS
      SUBROUTINE  SOLU(START1, START2, EMEAN, SIGMA)
      PARAMETER (KD1 = 1000)
      DIMENSION AX(KD1),XX(2),FVEC(2),
1     FJAC(2,2),R(3),QTF(2),W(2,4)
      EXTERNAL FCN
```

```
      COMMON /BLK1/AX,NT,T1,T2
      N = 2
      XTOL = SQRT(X02AAF(0.0))
      MAXFEV = 200*(N + 1)
      ML = N-1
      MU = N-1
      EPSFCN = 0.0
      MODE = 1
      FACTOR = 100.0
      NPRINT = 1
      LDFJAC = N
      LR = 3
      DELA = 0.01
      DELB = 0.01
      ILOOP = 0
30 IFAIL = 1
      DEL1 = DELA*ILOOP*(-1)**ILOOP
      DEL2 = DELB*ILOOP*(-1)**(ILOOP + 1)
      XX1OLD = START1 + DEL1
      XX2OLD = START2 + DEL2
      ILOOP = ILOOP + 1
      IF(ILOOP.EQ.5)GO TO 50
      XX(1) = XX1OLD
      XX(2) = XX2OLD
C***** NAG ROUTINE C05NCF TO SOLVE THE EQUATIONS
      CALL C05NCF (FCN,N,XX,FVEC,XTOL,MAXFEV,ML,MU,EPSFCN,DIAG,
1 MODE, FACTOR,NPRINT,NFEV,FJAC,LDFJAC,R,LR,QTF,W,IFAIL)
      IF(IFAIL.EQ.4.OR.IFAIL.EQ.5)GO TO 30
      IF(XX(1).EQ.XX1OLD.AND.XX(2).EQ.XXOLD) GO TO 30
      IF(XX(2).LT.0.0) GO TO 30
      IF(NFEV.LE.5)GO TO 30
      DIFF1 = ABS(FVEC(1))
      DIFF2 = ABS(FVEC(2))
      IF(DIFF1.GT.0.01.OR.DIFF2.GT.0.01)GO TO 30
      EMEAN = XX(1)
      SIGMA = XX(2)
50 RETURN
      END
      SUBROUTINE FCN(N,XX,FVEC,IFLAG)
      PARAMETER (KD1 = 1000)
      DIMENSION AX(KD1),XX(2),FVEC(2)
      COMMON /BLK1/AX,NT,T1,T2
      A = (T1-XX(1))/XX(2)
      B = (T2-XX(1))/XX(2)
C***** NAG ROUTINE S15ABF TO CALCULATE THE AREA
C***** OF NORMAL CURVE
      PHIA = S15ABF(A,IFAIL)
      PHIB = S15ABF(B,IFAIL)
```

CamScanner

```
      BOT1 = DENS(A)
      BOT2 = -A*BOT1
      TOP1 = DENS(B)
      TOP2 = -B*TOP1
      DIFF1 = PHIB-PHIA
      DIFF2 = TOP1-BOT1
      DIFF3 = TOP2-BOT2
      RDF1 = DIFF2/DIFF1
      RDF2 = DIFF3/DIFF1
      SUM1 = 0.0
      SUM2 = 0.0
      C1 = XX(2)*XX(2)
      C2 = C1*XX(2)
      C3 = NT/XX(2)
      DO 10 I = 1,NT
      SUM1 = SUM1 + (AX(I)-XX(1))
      SUM2 = SUM2 + (AX(I)-XX(1))**2
10 CONTINUE
      FVEC(1) = C3*RDF1 + SUM1/C1
      FVEC(2) = -C3*RDF2-C3 + SUM2/C2
      RETURN
      END
C***** THE DENSITY OF A NORMAL CURVE
      FUNCTION DENS(X)
      DENOM = SQRT(8.0*ATAN(1.0))
      DENS = (EXP(-X*X*0.5))/DENOM
      RETURN
      END
```

Conclusions

The solution of maximum likelihood equations by conventional techniques, using special tables, is laborious and time-consuming, particularly where analysis of large number of samples is involved.

The use of computer makes it possible to undertake the difficult numerical calculations involved in solving ML equations, as examplified in this study.

References

COHEN, A.C. JR. (1949) On estimating the mean and standard deviation of truncated normal distribution. Jour.Am.Stat.Assn.,44:518-525.

__,(1950). Estimating the mean and variance of normal populations from singly and doubly truncated samples.Ann. Math. Statist., 21: 557-69.

CamScanner

__(1955). Truncated and censored samples from normal populations. Trans. Amer. Soc. Qual. Cont. 9: 27-36.

__(1957). On the solution of estimating equations for truncated and censored samples from normal populations. Biometrika, 44:225-36.

__(1959). Simplified estimators for the normal distribution when samples are singly censored or truncated. Technometrics, 1: 217- 37.

__(1961). Tables for maximum likelihood estimates; singly truncated and singly censored samples. Technometrics, 3: 535-41.

DOUBLAS, J.(1983). Using the singly truncated normal distribution to analyze satellite data. Commun. Statist. -Theor. Meth., 12: 263-272.

HALD, A.(1949). Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. Skandinavisk Aktuarietidskrift, 32: 119-134.

CamScanner