International J. of Pure & Engg. Mathematics (IJPEM)
ISSN 2348-3881, Vol. 3 No. III (December, 2015), pp. 39-44

STATISTICAL APPROXIMATIONS FOR ASSESSING IN AN INTERVAL AND IMPLEMENTATION IN R

LAZIM KAMBERI[1] AND DASHMIR IBISHI[2]
[1,2] State University of Tetovo
1200, Tetovo Macedonia
E-mail: [1] lazim.kamberi@unite.edu.mk, [2] dashmir.ibishi@unite.edu.mk

Abstract

One of the main tasks of mathematical Statistics is the estimate of the unknown parameters in statistical models. In this paper we have described some methods for the estimate of the parameters.
Real data previously analyzed with different methods have already been analyzed by other methods as well as contemporary classics.
Special attention is paid to the selection of the method for the estimate of the parameters for some specific distribution.
Distribution of estimates interval are taken specific connection, like the discrete type, binomial distribution and the continual type with normal distribution.
During the work with these problems are taken classic examples with detailed explanations about these distribution methods.
The study reveals that the Bernoulli estimation and least squares methods are highly competitive with the maximum likelihood method and product estimating methods in small and large samples. Also each example is executed in R, to make the estimate of parameters with different distribution, where was taken algorithm for generating numbers distributions.

Key Words : *Estimate, Distribution methods, The Bernoulli estimation, R code.*

2010 AMS Subject Classification : 62N02.

1. Introduction

With a point estimate $\hat{\theta}$ of the parameter θ, we take the estimated value of the parameter, and that this approximation depends on the volume of the sample. So we need to find $|\theta - \hat{\theta}|$, so that this value should be as small as possible.
Let be M-sample (model) with volume n, X-a random variable.
Let be $p(x, \theta)$-the rule of the density for X, where θ-parameter. In the general case, when X have m-parameters, then receives $p(x, \theta_1, \theta_2, \cdots, \theta_m)$.
When need to apply a point evaluation $\hat{\theta}$ of the parameter θ, than $\hat{\theta}$ should be unbiased.
Let be $\alpha \in (0,1)$ such that

$$p(\theta_1 < \theta < \theta_2) = 1 - \alpha = \gamma. \qquad (*)$$

Exactly, the interval $I_\theta = (\theta_1, \theta_2)$ called a confidence interval for θ, α-called Level of importance, while $\gamma = 1$ called level of confidence (coefficient of confidence).
Usually, for α are used small values (for example $\alpha = 0.1; \alpha = 0.05; \alpha = 0.01$). So, if $\alpha = 0.05$, for I_p the parameter of θ belongs I_p in 95% of the empirical possibilities (because, $\gamma = 1 - \alpha = 0.95$) therefore, 5% does not belong to the interval I_p. We called $|I_p| = |\theta_2 - \theta_1| = 2\epsilon$-length of the interval.
ϵ is called assessing of the trust in an interval.

2.

2.1 Statistical Approximations for Assessing in an Interval

Suppose that population has characteristic $X \approx N(\mu, \sigma)$. Let be M-sample (model) with volume - n. We have :

$$\begin{aligned}
\overline{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i,\ \overline{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2,\ \tilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \quad \text{and} \\
S^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \\
(n-1)\tilde{S}^2 &= n\overline{S}^2,\ m(\overline{X}) = \mu,\ m(\tilde{S}^2) = \sigma^2,\ m(\overline{S}^2) = \frac{n-1}{n}\sigma^2.
\end{aligned}$$

Respectively the realized values are:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Theorem 1 :

(i) $\frac{nS^2}{\sigma^2} \sim \chi^2(n)$;

(ii) $\frac{n\overline{S}^2}{\sigma^2} \sim \chi^2(n-1)$;

(iii) $\frac{(n-1)\tilde{S}^2}{\sigma^2} \sim \chi^2(n-1)$.

Proof : (i) By that

$$\frac{nS^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 .$$

And since $X_i \approx N(0,1)$ and $X_i^2 \approx \chi^2(1)$, from the other side

$$D_k \sim \chi^2(n_k) \Rightarrow \sum_{i=1}^{n} X_k \sim \chi^2(n_1 + n_2 + \cdots + n_m),$$

than,

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(1) + \cdots + \chi^2(1) \sim \chi^2(n)$$

therefore:

$$\frac{nS^2}{\sigma^2} \sim \chi^2(n).$$

(ii) μ-unknown. Deals relevant assessment on point (unbiased) : $\hat{\mu} = \overline{X}$. Than :

$$\overline{S}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{1}{n} \sum_{i=1}^{n} [(X_i - \mu) - (\overline{X} - \mu)]^2 = S^2 - (\overline{X} - \mu)^2$$

therefore

$$S^2 = \overline{S}^2 + (\overline{X} - \mu)^2.$$

If we say:

$$U = \frac{nS^2}{\sigma^2}, ‘ \quad V = \frac{n\overline{S}^2}{\sigma^2}, \quad W = \left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right).$$

Then, is $U = V + W$. From (i) we see that $U \sim \chi^2(n), W \approx \chi^2(1)$.
Since $g_{ij}(t) = (1-2t)^{-\frac{n}{2}}$, and $g_W(t) = (1-2t)^{-\frac{1}{2}}$ $(n = 1)$ then:

$$g_U(t) = g_{V+W}(t) = g_V(t) \cdot g_W(t)‘ \quad \text{(from funksion } g).$$

Therefore: $V \sim \chi^2(n-1)$. So, $\frac{n\overline{S}^2}{\sigma^2} \sim \chi^2(n-1)$.

(iii) Now is evidently that :

$$\frac{(n-1)\tilde{S}^2}{\sigma^2} = \frac{n\overline{S}^2}{\sigma^2} \sim \chi^2(n-1) \quad \text{from (ii)).}$$

2.2 Specific Interval Estimates

I_p interval for $B(n,p)$: If n-the numbers of the recurring test, and m-the number of the realization of events $A(m \leq n)$, then, for $X \sim B(n,p)$ we know that:

$$p^* = p(X=m) = \begin{pmatrix} n \\ \\ m \end{pmatrix} p^n(1-p)^{n-m}$$

where: $\hat{p} = \frac{m}{n}$ is a point estimate. (X-the number of the realization of the event of A in n-tests.

From the theorem of Laplace, for a grate n-we have:

$$T = \frac{X-np}{\sqrt{npq}} \sim N(0,1) \ \ (p+q=1).$$

According to this we can specify the confidence interval: $I_p = (p_1, p_2)$, who τ is so that $p(-\tau \leq T \leq \tau) = 2\Phi_0(\tau) = 1-\alpha$ where α-level of importance. So,

$$-\gamma \leq \frac{X-np}{\sqrt{np(1-p)}} < \tau$$

which is equivalent to

$$\frac{(X-np)^2}{np(1-p)} \leq \tau^2$$

and we win the inequality of p :

$$(n+\tau^2)p^2 - (2m+\tau^2)p + \frac{m^2}{n} \leq 0 \qquad (ku \ X = m)$$

where : $\frac{m}{n} = \hat{p}$ i.e. : $m = n\hat{p}$, $\frac{m^2}{n} = n\hat{p}^2$, or:

$$(n+\tau^2)p^2 - (2n\hat{p}+\tau^2)p + n\hat{p}^2 \leq 0$$

since $1+\tau^2 > 0$ then exist : p_1, p_2 $(p_1 < p_2)$ so that : $p \in [p_1, p_2]$. So

$$p_{1/2} = \frac{\left(n\hat{p} + \frac{\tau^2}{2}\right) \pm r\sqrt{n\hat{p}(1-\hat{p}) + \frac{\tau^2}{4}}}{n+\tau^2}. \tag{1}$$

Or for great n- : $n\hat{p}^2 + \frac{\tau^2}{2} \approx n\hat{p}, n + \tau^2 \approx n$ so, we win :

$$p_{1/2} = \hat{p} \pm \tau\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

3. Main Results

1. Numerical example, in R code :

Find I_p for $p = p(A)$, if A in 100 tests is realized 30 times. And the important level is $\alpha = 0.05$.

$$\tau^{1')} = \Phi_0^{-1}\left(\frac{1-0.05}{2}\right) = \Phi_0^{-1}(0.475) = 1.96; \quad \hat{p} = \frac{30}{100} = \frac{3}{10} = 0.30.$$

From (1) we have:

$$p_{1/2} = \frac{3}{10} \mp 1.96 \cdot \frac{6}{125} = 0.30 \mp 0.09408, \; i_p = (0.20; 0.36).$$

The same example can be realize by using the code in R :

1-sample proportions test with continuity correction data: 30 out of 100, null probability 0.5.

X-squared = 7.29, df = 1, p-value = 0.006934

alternative hypothesis: true p is not equal to 0.5

95 percent confidence interval:

0.2021721 0.3627255

sample estimates:

p

0.36

If we say that the important level should be 90%:

In R conf.level :

> prop.test(30,100,conf.level=0.90)

1-sample proportions test with continuity correction

data: 30 out of 100, null probability 0.5

X-squared = 7.29, df = 1, p-value = 0.006934

alternative hypothesis: true p is not equal to 0.5

90 percent confidence interval:

0.2210610 0.3467754

sample estimates:

p

0.361.

References

[1] Abdul-Moniem, Ibrahim B., L-moments and TL-moments estimation for the exponential distribution. Far East Journal of Theoretical Statistics, 23(1) (2007), 51.

[2] Barreto-Souza W., Santos A., Cordeiro GM (2011) The beta generalized exponential distribution. J Stat Comput Simul., 81, 645-657.

[3] Breslow NE, Day NE., Statistical Methods in Cancer Research: II - The Design and Analysis of Cohort Studies. Lyon: International Agency for Research on Cancer, (1987).

[4] Swain J. S., Venkatraman S. and Wilson J., Least squares estimation of distribution function in Johnsons translation system. Journal of Statistical Computation and Simulation, 29 (1988), 271-297.

[4] Tytler J. A., Seeley H.F., The Nellcor N-101 pulse oximeter - a clinical-evaluation in anesthesia and intensive-care. Anaesthesia, 41 (1986), 302-305.

[5] James F., Statistical Methods in Experimental Physics, Singapore: World Scientific., 7(4) (2006).

[6] Shahzad M. N. and Asghar Z., Comparing TL-Moments, L-Moments and Conventional Moments of Dagum Distribution by Simulated data. Revista Colombiana de Estadistica, 36(1) (2013), 79-93.

[7] Ling Dan, Hong-Zhong Huang and Yu Liu, A method for parameter estimation of mixed Weibull distribution. In Reliability and Maintainability Symposium, (2009). RAMS 2009. Annual, pp. 129-133. IEEE, 2009.