ISSN 1407-8015 TRANSPORT AND ENGINEERING MAŠĪNZINĀTNE UN TRANSPORTS 2008-8015

INTELLIGENT TRANSPORT SYSTEMS
INTELEKTUĀLĀS TRANSPORTA SISTĒMAS

STATISTICAL TASKS SOLUTION OF PUNCTUAL AND UNBIASED ESTIMATION WITH THE SOFTWARE APPLICATION

Victor Lyumkis. *Department of Computer Technologies,*
Transport and Telecommunication Institute,
Lomonosov str. 1, LV-1019, Riga, Latvia,
Asoc. professor , Dr. Sc. Ing.,
e-mail: vlyumkis@yahoo.com

Keywords: parameters, unbiased estimation, Weibull distribution, Matchcad package

1.Introduction

The classical approach for the punctual estimation, as it is known, is the method of moments and maximum-likelihood method [1,2,3,4]. However, their application leads to the considerable calculation difficulties that make us cosider this problem as a separate, independent task. A number of concrete practical tasks of such kind is considered in the given paper. In the class of punctual estimates a peculiar importance belongs to the unbiased estimates. The idea of the "good" punctual estimates calculation with the further liquidation of the bias of this estimate [1] is being realized in the work for the whole groups of tasks.

2. Composition of normal and equal distributions

Let a random value (r.v.) X has a normal distribution N(m,σ), and a random value Y has an equal distribution with the density like

$$f_Y(u) = \begin{cases} 1/2b, & -b \le u \le b \\ 0, & u \notin [-b;b] \end{cases} \quad (1)$$

There is a samping $z_1, z_2, ..., z_n$ of random value observations. Z=X+Y, being a composition of normal and equal distributions [2,3]. According to the metioned selected data it is necessary to calculate the parameters estimates m, σ, b.

For this purpose, preliminary, we will find the expression of density of r.v. Z, taking into account, that

$$\varphi(z) = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y)dy = \frac{1}{\sqrt{2\pi}\sigma 2b} \int_{-b}^{+b} e^{-1/2(z-y-m/\sigma)^2} dy$$

(2)

After replacement t=(z-y-m)/σ, we have

$$\varphi(z)=\frac{1}{\sqrt{2\pi}2b}\int_{(z-b-m)/\sigma}^{(z+b-m)/\sigma}e^{-t^2/2}dy=\frac{1}{2b}\left[\phi(\frac{b+z-m}{\sigma})-\phi\left(\frac{-b+z-m}{\sigma}\right)\right], \quad (3)$$

where $\phi(u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u}e^{-t^2/2}dt$.

Further EZ=EX+EY=m, because EY= $\int_{-b}^{b}\frac{y}{2b}dy=0$ and similar to

$$DZ=DX+DY=\sigma^2+\int_{-b}^{b}\frac{y^2}{2b}dy=\sigma^2+b^2/3 \quad (4)$$

It is obvious that due to the summetry of distributions $\mu_3(z)=\mu_3(x)+\mu_3(y)=0$, where $\mu_3(.)$ – a central moment of the third order.

It is easy to see, that $\mu_4(z)=\mu_4(x)+\mu_4(y)+6\mu_2(x)\mu_2(y)$.

Now we will calculate the expression of the corresponding central moments.

$\mu_4(x)=E[(X-EX)^4]=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-m)^4e^{-1/2(\frac{x-m}{\sigma})^2}dx=3\sigma^4$.

Similarly $\mu_4(y)=\frac{1}{2b}\int_{-b}^{b}y^4dy=b^4/5$.

Further $\mu_2(x)=\sigma^2, a \ \ \mu_2(y)=\frac{b^2}{3}$, that gives the expression for $\mu_4(z)$ like

$\mu_4(z)=3\sigma^2+b^4/5+2\sigma^2b^2$.

From the all said above the equation set of the method of moments will look like:

$$\begin{cases}\hat{m}=\bar{z}\\ \sigma^2+b^2/3=\sum_{i=1}^{n}(z_i-\bar{z})^2/n\\ 3\sigma^4+b^4/5+2\sigma^2b^2=\sum_{i=1}^{n}(z_i-\bar{z})^4/n\end{cases} \quad (5)$$

The solution of the equation set is being realized in MathCAD 2001 package with the help of the Given, Find block.

3. Distributions with the shift and scale parameter

Among the discussed tasks of the unbiased estimation we will mark out the task of the unbiased estimates calculation of the parameter degrees θ(k,j)= $\theta_0^k*\theta_1^k$ of the distribution class with the shift and scale parameters $F_X(x)=F((x-\theta_0)/\theta_1)$. The application of the so-called regular estimates
[1] leads to the use of formulas of the unbiased estimates calculation θ(k,j) like

$$\overset{*}{\theta}(0,j)=\overset{\wedge}{\theta}(0,j)/\beta(0,j), \ [\ 1\ ],$$

where $\beta(0,j)=E\hat{\hat{\theta}}(0,j)$ - is a coefficient of the sclae parameter regular estimate smoothing $\hat{\hat{\theta}}(0,j)$, obtained according to the sampling with the true value of parameter $\theta_1=1$ and $\theta_0=0$ from the initial distribution.

As the illustration of the mentioned method let us consider a concrete distribution. Let independent r.v. $X_{i,\ i=1,2,...n,}$ have the exponential distribution like

$$f_{X_i}(u) = \begin{cases} \frac{1}{\lambda} * e^{-\frac{u}{\lambda}}, & u \geq 0 \\ 0, & u < 0 \end{cases} \tag{6}$$

By sampling $x_1, x_{2,...} x_n$ from the distribution (6) we calculate the unbiased estimate of the parameter λ^k, basing on the maximum-likelihood method

$$\overset{*}{\lambda} = \frac{(\sum_{i=1}^{n} x_i / n)^k}{\beta(0,k)}, \tag{7}$$

where β(0,k)= $E([\dot{\lambda}]^k)$, at that $\dot{\lambda} = \sum_{i=1}^{n} \dot{x}_i / n$, a $\dot{x}_1, \dot{x}_2, ... \dot{x}_n$ - r.v. from (6) with λ=1.

Taking into account all said above, we see that the distribution of the statistics $\dot{\lambda}$ is built on the statistics T=$\sum_{i=1}^{n} \dot{x}_i$, which has gamma-distribution with the density like

$$f_T(t) = \frac{t^{n-1} * e^{-t}}{(n-1)!}$$

In its turn the statistics $T' = T / n$ has density $\frac{n^n * t^{n-1} * e^{-t}}{(n-1)!}$ that allows calculating the coefficient liquidating the bias like

$$\beta(0,k) = E([\dot{\lambda}]^k) = \int_0^{\infty} \frac{t^k n^n t^{n-1} e^{-nt} dt}{(n-1)!} = \frac{(k+n-1)!}{n^k (n-1)!}$$

For the comparison of the approximate* and exact values of the mentioned coefficients in Table 1 the corresponding values with some n и k are shown.

---

*Let us pay attention to the fact that usually when considering the distributions with parameters of shift and scale we cannot calculate the similar expressions of the coefficient for liquidation of bias. We are forced to calculate this coefficient numerically. Its calculation algorithm does not differ from the said in the given paper. Basically, such approach is easily may be realized for the calculation of the biased estimate of the degrees of the scale parameter of the Weibull distribution even in the terms of censoring. In particular, let here it is said about the Weibull distribution (double exponential distribution) like

$$F_{X_i}(u) = 1 - \exp(-\exp(-\exp((u - \theta_0)/\theta_1)), i = 1,...,n.$$

In this case the equation set of the maximum-likelihood method for the parameters estimation is the following (according to sampling of volume n with the use of the first r observations, i.e. in the terms of censoring of the II type)

$$\begin{cases} \dfrac{\sum_{i=1}^{n} v_i x_{(i)}}{\sum_{i=1}^{n} v_i} - r\hat{\theta}_1 - \sum_{i=1}^{r} x_{(i)} = 0, \\ \hat{\theta}_{0X} = \hat{\theta}_1 (\ln(\frac{1}{r} \sum_{i=1}^{n} v_i)), \end{cases}$$

where $v_i = \begin{cases} \exp(x_{(i)} / \hat{\theta}_1), 1 \le i \le r, \\ \exp(x_{(r)} / \hat{\theta}_1), r < i \le n, \end{cases} \quad x_{(i)} = \begin{cases} x_{(i)}, i \le r, \\ x_{(r)}, i > r. \end{cases}$

The author has developed the programme of solution of the above written-out equation set that is a key moment for the further calcualtion of the coefficients of estimate bias liquidation of the scale parameter degrees.

Table 1.

Comparison of approximate and exact values of coefficients β(0,k) for the number of degrees k of parameter θ with various volumes of sampling n

n=10; k=2		n=10; k=4		n=20; k=2		n=20; k=4		N=30; k=2	
1.103	1.1	1.72	1.716	1.052	1.05	1.327	1.328	1.034	1.033
n=30 ; k=4		n=40; k=2		n=40; k=4		n=50; k=2		n=50; k=4	
1.218	1.212	1.025	1.025	1.158	1.157	1.02	1.02	1.125	1.124

Below, it is a listing of program in the Matchcad environment for the calculation of the approximate values β(0,k).

```
biasexpo(m, n, k1) := | sum ← 0
                      | for j1 ∈ 1..m
                      |   | s ← 0
                      |   | for j ∈ 1..n
                      |   |   | u ← −ln[1 − (rnd(1))]
                      |   |   | s ← s + u
                      |   | s ← s/n
                      |   | sum ← sum + s^k1
                      |   | sum1 ← sum/j1
```

Fig. 1. The calculation of coefficients β(0,k) in the Matchcad environment.

4. Liquidation of the bias by the direct taking of mathematical expectation

Let us notice that the method used for the obtaining of the unbiased estimates in the previous case is based on the idea of the direct taking of the mathematical expectation with the further correction for the bias liquidation. It is rather easy to replicate this idea from the whole number of tasks. For example, in a number of situations it is necessary to obtain the unbiased estimates of the parameter degrees g(θ)= θ^k , k=1,2,3,... in the assumption that a random value X has the Poisson's distribution like $P(X = m) = \theta^m * e^{-\theta} / m!$, m=0,1,2,....

The unbiased estimates is obvious and unbiased T(X)=X for parameter θ. It is easy to generalize this result for the arbitrary k, asserting that the unbiased estimate for the function g(θ)= θ^k the estimate T(X)=X(X-1)(X-2)...(X-k+1) will be. Really, let k=2. Then, from the mathematical expectation definition we have

$$E(X(X-1))=\sum_{k=0}^{\infty}\frac{k(k-1)\theta^k e^{-\theta}}{k!}=\sum_{k=2}^{\infty}\frac{k(k-1)\theta^k e^{-\theta}}{k!}=\sum_{k=2}^{\infty}\frac{\theta^k e^{-\theta}}{(k-2)!}=\theta^2$$

Similar calculations have a place for the arbitrary k, and the estimates themselves quite easily are realized in the calculation aspect in the environment of the Matchcad package.

The following task is considered in the similar way. Let r.v. X_i, $i=1,2,\ldots n$, have a normal distribution N(a,σ), at that the parameter is known. Using the above said idea, we will show that the unbiased and consistent estimate σ is the estimate $T=\sqrt{\frac{\pi}{2}}*\sum_{i=1}^{n}|X_i-a|/n$. Really, r.v. $\dot{X}_i=X_i-a$ has a normal distribution with the density $f_{\dot{X}_i}(u)=\frac{e^{-\frac{u^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$.

Therefore $$f_{|\dot{X}_i|}(u)=f_{\dot{X}_i}(u)+f_{\dot{X}_i}(-u)=\frac{2*e^{-\frac{u^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}.$$

Hence, after some transformations

$M(|\dot{X}_i|)=\sigma\sqrt{\frac{2}{\pi}}$, that proves the unbias of the estimate T.

By the easy calculations it is possible to prove the consistence of the estimate T. For this purpose it is enough to make sure in the tendency of the estimate dispersion T to 0 with n→∝. Really

$$D(|\dot{X}_i|)=M(|\dot{X}_i|^2)-M^2(|\dot{X}_i|)=\frac{2*\int_0^{\infty}u^2 e^{-\frac{u^2}{2\sigma^2}}du}{\sigma\sqrt{2\pi}}-\frac{2\sigma^2}{\pi}=\sigma^2\left(\frac{\pi-2}{\pi}\right)$$

It is seen that the dispersion of the estimate $D(T)=\frac{\sigma^2(\pi-2)}{2n}$, i.e. D(T)→0 with n→∞.

Let us consider at last a random value having the density of gamma-distribution like

$$f_X(x)=\begin{cases}\theta^{\lambda}x^{\lambda-1}e^{-\theta x}/\Gamma(\lambda), & x>0,\\ 0, & x\le 0,\end{cases} \qquad (8)$$

where parameter $\theta>0,\lambda-$ is some known parameter, $\Gamma(\lambda)$ -- gamma-function in point λ. It is easy to calculate the parameter estimate θ by the maximum-likelihood method, showing at that, that it is biased. Really, likelihood function by sampling for our case will look like

$$L(\theta)=\frac{\theta^{n\lambda}.(x_1x_2..x_n)^{\lambda-1}e^{-\theta\sum_{i=1}^{n}x_i}}{(\Gamma(\lambda))^n} \qquad (9)$$

By taking the derivative from logarithms of likelihood function and equating it to zero, we will obtain the equation

$$\frac{n\lambda}{\theta} - \sum_{i=1}^{n} x_i = 0.$$

Hence, we obtain the esatime of the maximum likelihood of the parameter θ like $\hat{\theta} = \lambda / \bar{x}$, but at that due to $\frac{d^2 \ln L}{d\theta^2} = -\frac{n\lambda}{\theta^2} < 0$, the calculated extremum of the likelihood function is a maximum.

Further, we will show that the obtained estimate $\hat{\theta} = \lambda / \bar{X}$ is biased, taking into account the opportunity of the bias liquidation. Really, the density of r.v. $Y = \frac{1}{\bar{X}}$ has the distribution like $F_Y(u) = P(\frac{1}{\bar{X}} < u) = 1 - P(\bar{X} < \frac{1}{u})$, therefore the density is $f_y(u) = f_{\bar{X}}(1/u)\frac{1}{u^2}$.

However, the density of r.v. $X = \sum_{i=1}^{n} X_i$ possesses the simulated property, therefore r.v. X has the initial distribution (8) with parameter $\lambda = n\lambda$, and density of r.v.$Y = \frac{1}{\bar{X}}$ $(\bar{X} = \sum_{i=1}^{n} X_i / n)$ looks like

$$f_Y(x) = f_{\bar{X}}(1/x)\frac{1}{x^2} = \begin{cases} (n\theta)^{n\lambda}(1/x)^{n\lambda+1} e^{-n\theta/x} / \Gamma(n\lambda), & x > 0, \\ 0, & x \le 0, \end{cases}.$$

Now, the mathematical expectation of the maximum likelihood $\hat{\theta} = \lambda / \bar{X}$ is defined by the integral

$$M(\frac{\lambda}{\bar{X}}) = \lambda \int_0^{\infty} \frac{y(n\theta)^{n\lambda}(1/y)^{n\lambda+1} e^{-n\theta/y} dy}{\Gamma(n\lambda)} = \frac{(n\theta)^{n\lambda}\lambda}{\Gamma(n\lambda)} \int_0^{\infty} t^{n\lambda-2} e^{-n\theta t} dt$$

By using the known integral [3] $\int_0^{\infty} x^{\lambda-1} e^{-\beta x} dx = \Gamma(\lambda)/\beta^{\lambda}$ and corrrelation $\Gamma(\lambda+1) = \lambda\Gamma(\lambda)$, we finally obtain

$$M(\frac{\lambda}{\bar{X}}) = \frac{n\lambda\theta}{n\lambda - 1} \neq \theta.$$

Therefore, the unbiased estimate of our parameter θ is the value

$$\hat{\theta} = \frac{n\lambda - 1}{n\bar{X}}.$$

References

1. Yu. M. Paramonov. Methods of the mathematical statistics in problems, which is connected with estimator and ensuring relability fatigue durability of the aviation constructions, Riga, RKII GA,1994.

2. Ya. B. Shor. Statistical methods of analyse and control quality and relability, M., 1972.

3. M. Kendall, A. Stuart. The advanced theory of statistics. Inference and relationship.M.:Nauka, 1973.

4. L.J. Bain, M. Engelhardt. Statistical Analysis of Reliability and Life-testing Models.New York, 1991.

5. Андронов А.М.,Копытов Е.А.,Гринглаз Л.Я.Теория вероятностей и математическая статистика: Учебник для вузов.Спб.:Питер, 2004, 461с.

V. Ļumkis. Punktveida un nenobīdītā novērtējuma statistisko uzdevumu risināšana, izmantojot pakešlīdzekļus

Darbā aplūkota punktveida un nenobīdītā novērtējuma uzdevumu virkne. Akcentēta mūsdienu paketes Matchad 2001 pielietošanas aktualitāte, kur samērā sarežģīti skaitļošanas algoritmi var tikt realizēti kompaktu programmu veidā. Veselai virknei uzdevumu tiek parādīta iespēja atrast nenobīdītos novērtējumus analītiski.
Pēdējā gadījumā izmantota pieeja, kas balstīta uz nobīdes likvidēšanu, izmantojot aplūkojamā parametra zināmā punktveida novērtējuma tiešo matemātisko gaidīšanu.

V. Lyumkis. Statistical Tasks Solution of Punctual and Unbiased Estimation with the Software Application

In the paper a number of tasks of the unbiased and punctual estimation is discussed. The importance of the application of the modern Matchcad 2001 package is stressed where rather complicated calculation algorithms may be realized in the form of compact programs. For the whole class of tasks the opportunity of analytical calculation of the unbiased estimates. In the latter case the approach based on the liquidation of the bias by the direct taking of the mathematical expectation of the known punctual estimate of the considered parameter is used.

V. Lyumkis. Решение статистических задач точечного и несмещенного оценивания с использованием пакетных средств

В работе обсуждается ряд задач несмещенного и точечного оценивания. Подчеркивается актуальность применения современного пакета Matchcad 2001, где достаточно сложные вычислительные алгоритмы могут быть реализованы в виде компактных программ. Для целого класса задач показывается возможность нахождения несмещенных оценок аналитически. В последнем случае используется подход, основанный на ликвидации смещения путем непосредственного взятия математического ожидания известной точечной оценки рассматриваемого параметра.