ᲡᲐᲥᲐᲠᲗᲕᲔᲚᲝᲡ ᲛᲔᲪᲜᲘᲔᲠᲔᲑᲐᲗᲐ ᲔᲠᲝᲕᲜᲣᲚᲘ ᲐᲙᲐᲓᲔᲛᲘᲘᲡ ᲛᲝᲐᲛᲑᲔ, ტ. 9, №2, 2015
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 9, no. 2, 2015

Mathematics

On Estimation of Unknown Parameters of Exponential-Logarithmic Distribution by Censored Data

Alex Pijyan

Department of Mathematics, Faculty of Exact and Natural Sciences, Tbilisi State University, Tbilisi

(Presented by Academy Member Elizbar Nadaraya)

ABSTRACT. The problem of estimation of parameters of Exponential-Logarithmic distribution in the case of censored data is considered. We used pseudo maximum likelihood method and constructed a procedure to solve this problem. Theorem of consistency is proved. Simulation is used to study the properties of estimators derived. *© 2015 Bull. Georg. Natl. Acad. Sci.*

Key words: *Exponential-Logarithmic distribution, pseudo maximum likelihood estimators, consistent estimators, partly censored data.*

The Exponential-Logarithmic (EL) distribution is a family of lifetime distributions with decreasing failure rate, defined on the interval $[0,\infty)$, see [1]. This distribution is parameterized by two parameters: $p\in(0,1)$ and β>0. (EL) distribution is used in the study of lengths of organisms, devices, materials, etc., which is of major importance in the biological and engineering sciences. The given work studies the problem of parameters estimation of (EL) distribution with a specific observation pattern. The observations are assumed to group and full non-observations of individual realization values are possible. Application of a modified in a certain sense, method of maximum likelihood is suggested and the offered procedure is shown to lead to consistent estimators.

Let X be a random variable with a distribution function $F(x)=F(x,\theta)$, where $\theta\in\Theta$ is an unknown vector parameter in a finite dimensional Euclidean space $\Theta\subset R^q$. Suppose Θ is compact. We have to construct a consistent estimator θ based on observations of a random variable X. The experiment is designed in such a way that the actual number of realizations is unknown to us, we know only a part of those realizations.

Let fixed points $0\le t_1\le t_2\le\cdots\le t_n\le\infty$ be given on the line R (including the case when the last point takes an infinite value). These points make three categories of intervals:

0) The interval (t_i,t_{i+1}) belongs to zero category if we do not know either the individual values of the sample, or the number of sample values of the random variable appearing in this interval.

1) The interval (t_i,t_{i+1}) belongs to the first category if individual sample values are unknown, but we

know the number of sample values of the random variable X in this interval. As usual, we denote this number by n_i.

2) The interval (t_i, t_{i+1}) belongs to the second category if we know individual sample value $x_{i1}, x_{i2}, \cdots, x_{in_i}$.

Further we denote summation and integration in the intervals of the zero, first and second categories by (0),(1) and (2), respectively (see [2]).

We call this type of sample-partially grouped sample with censoring. Evidently, censored samples of both types as well as truncated samples make a special case of the stated problem.

Absence of information in the zero category interval creates certain difficulties, which we try to overcome based on our knowledge of the distribution type of $F(x,\theta)$ and the number of sample members that do not appear in the zero category interval: $n = \sum_{(1)(2)} n_i$.

Let $A_i = (t_i, t_{i+1})$ be zero category interval. We denote by m_i the number of sample members appearing in A_i. Then, $r = n + \sum_{(0)} m_i$ in the total number of observations. Note that $\frac{m_i}{n + \sum_{(0)} m_i}$ is, relative frequency of X appearing in A_i. If $\widehat{F}_r(x) = \widehat{F}_r(x,\theta)$ denotes an empirical distribution function, then

$$\frac{m_i}{n + \sum_{(0)} m_i} = \widehat{F}_r(t_{i+1}) - \widehat{F}_r(t_i) \tag{1}$$

and by virtue of Bernoulli's law of large numbers it converges to $p_i(\theta) = F(t_{i+1},\theta) - F(t_i,\theta)$ with probability 1. By summation of Equalities (1) over all zero category intervals we obtain

$$\sum_{(0)} m_i = n \frac{\sum_{i \in (0)} \left[\widehat{F}_r(t_{i+1}) - \widehat{F}_r(t_i) \right]}{1 - \sum_{i \in (0)} \left[\widehat{F}_r(t_{i+1}) - \widehat{F}_r(t_i) \right]},$$

which leads to

$$m_i = n \frac{\widehat{F}_r(t_{i+1}) - \widehat{F}_r(t_i)}{1 - \sum_{i \in (0)} \left[\widehat{F}_r(t_{i+1}) - \widehat{F}_r(t_i) \right]}. \tag{2}$$

We intend to apply the method of pseudo maximum likelihood. Assume that the random variable X has a distribution density with respect to Lebesgue measure $f(x) = f(x,\theta)$ .Then the likelihood function has the following form:

$$L_n(x;\theta) = \prod_{i \in (0)} \left[F(t_{i+1}) - \mathrm{F}(\mathrm{t}_i) \right]^{m_i} \prod_{i \in (1)} \left[F(t_{i+1}) - \mathrm{F}(\mathrm{t}_i) \right]^{n_i} \prod_{j=1}^{n_i} f(x_{ij}), \tag{3}$$

where $m_i, i \in (0)$ are defined by Formulas (2). Finding maximum points of the function $L(x;\theta)$ becomes complicated since it is difficult to study smoothness properties of empirical functions. Therefore, we consider a modified likelihood function:

$$\overline{L_n}(x;\theta) = \prod_{i \in (0)} \left[F(t_{i+1}) - \mathrm{F}(\mathrm{t}_i) \right]^{n \frac{F(t_{i+1}) - F(t_i)}{1 - [F(t_{i+1}) - F(t_i)]}} \prod_{i \in (1)} \left[F(t_{i+1}) - \mathrm{F}(\mathrm{t}_i) \right]^{n_i} \prod_{j=1}^{n_i} f(x_{ij}). \tag{4}$$

Lemma:

Let the following conditions be satisfied:

a) The distribution function $F(x,\theta)$ is continuous in both variable and has a continuous derivative $f(x,\theta)=\frac{\partial F(x,\theta)}{\partial x}$;

b) The function $\overline{L_n}(x;\theta)$ has an absolute maximum $\theta=\overline{\theta}_n$.

Then $\overline{\theta}_n$ is consistent and asymptotically efficient estimator of the true value of the parameter $\theta=\theta_0$.

The proof results from the corresponding theorems of [3-4].

Estimation of the parameters *p* and β of (EL) Distribution

Let X be a random variable distributed with respect to (EL) distribution, with density

$$f(x,p,\beta)=\left(\frac{1}{-\ln p}\right)\frac{\beta(1-p)e^{-\beta x}}{1-(1-p)e^{-\beta x}}$$

and distribution function

$$F(x)=1-\frac{\ln(1-(1-p)e^{-\beta x})}{\ln p},$$

$p\in(0,1),\ \beta>0,\ x\in[0,\infty)$. Let $[1,\infty)$ be a zero category interval. However, we have observations $X_1,X_2,\cdots,X_n$ in the interval [0,1). We have to estimate β and *p* parameters with respect to these observations. We apply pseudo maximum likelihood estimates. In order to construct likelihood functions note that if we denote by ***k*** the number of general sample members that appear in $[1,\infty)$, then $\frac{k}{k+n}$ will be the frequency of them appearing in the interval $[1,\infty)$. Therefore by Bernoulli-Kolmogorov theorem we have

$$\hat{k}=n\frac{\left[F(\infty)-F(1)\right]}{1-\left[F(\infty)-F(1)\right]}$$

Since the probability of ***k*** elements of the sample appearing in the "black hole" is $[\mathrm{F}(\infty)-\mathrm{F}(1)]^k$, we can write the pseudo likelihood function as

$$L=\left[\frac{\ln(1-(1-p)e^{-\beta x})}{\ln p}\right]^{\frac{n\ln(1-(1-p)e^{-\beta x})}{\ln p-\ln(1-(1-p)e^{-\beta x})}}\prod_{i=1}^{n}(\frac{1}{-\ln p})\frac{\beta(1-p)e^{-\beta x}}{1-(1-p)e^{-\beta x}} \tag{5}$$

Hence,

$$\ln L=\frac{n\ln(1-(1-p)e^{-\beta x})}{\ln p-\ln(1-(1-p)e^{-\beta x})}\ln\frac{\ln(1-(1-p)e^{-\beta x})}{\ln p}+n\ln(\frac{1}{-\ln p})+n\ln(1-p)+$$

$$+n\ln\beta-\beta\sum_{i=1}^{n}x_i-\ln\prod_{i=1}^{n}(1-(1-p)e^{-\beta x_i})$$

Computing $\frac{\partial\ln L}{\partial\beta}$ we get

$$\frac{(1-p)ne^{-\beta}}{(1-(1-p)e^{-\beta})(\ln p-\ln(1-(1-p)e^{-\beta x}))}+\frac{n}{\beta}-\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\frac{(1-p)x_i e^{-\beta x_i}}{1-(1-p)e^{-\beta x_i}}+$$

$$+\frac{e^{-\beta}(1-p)n\ln p}{1-(1-p)e^{-\beta})(\ln p-\ln(1-(1-p)e^{-\beta x}))^2}\ln\frac{\ln(1-(1-p)e^{-\beta x})}{\ln p} \qquad (6)$$

Similarly $\frac{\partial \ln L}{\partial p}=$,

$$\left(\frac{n\ln\frac{\ln(1-(1-p)e^{-\beta x})}{\ln p}}{(\ln p-\ln(1-(1-p)e^{-\beta x}))^2}+\frac{n}{(\ln p-\ln(1-(1-p)e^{-\beta x}))\ln p}\right)\cdot$$

$$\left(\frac{ne^{-\beta}\ln p}{1-(1-p)e^{-\beta}}-\frac{n\ln(1-(1-p)e^{-\beta x})}{p}\right)-$$

$$-\frac{n}{p\ln p}-\frac{n}{1-p}-\sum_{i=1}^{n}\frac{e^{-\beta x_i}}{1-(1-p)e^{-\beta x_i}} \qquad (7)$$

Studying (6), when $\beta\to 0$ and $\beta\to\infty$ we get

$$\lim_{\beta\to 0}\frac{\partial \ln L}{\partial \beta}=+\infty \quad \text{and} \quad \lim_{\beta\to\infty}\frac{\partial \ln L}{\partial \beta}=-\sum_{i=1}^{n}x_i\ .$$

The continuous function $\frac{\partial \ln L}{\partial \beta}$ changes its sign and hence there exists a point $\widehat{\beta}$, such that

$\frac{\partial \ln L}{\partial \beta}(\beta=\widehat{\beta})=0.$

We come to the same conclusion related to parameter p:

$$\lim_{p\to 0}\frac{\partial \ln L}{\partial p}=+\infty \ \text{and} \ \lim_{p\to\infty}\frac{\partial \ln L}{\partial p}=-\left(\frac{n}{2}+\sum_{i=1}^{n}e^{-\beta x_i}\right).$$

The continuous function $\frac{\partial \ln L}{\partial p}$ changes its sign and hence there exists a point $\widehat{p}$, such that

$$\frac{\partial \ln L}{\partial p}(p=\widehat{p})=0.$$

On the basis of above conclusions and taking into consideration ***Lemma*** we can state that the following theorems of consistency are true:

Theorem 1: Let $X_1,X_2,\cdots,X_n$ be random variables of a sample distributed with respect to (EL) distribution with unknown parameter β and known p. The observations are made on the interval $[0,1)$ and in $[1,\infty)$ neither sample members, nor their number are recorded. Then the pseudo maximum likelihood estimator for β parameter exists and is a unique root of the equation:

$$\frac{(1-p)ne^{-\beta}}{(1-(1-p)e^{-\beta})(\ln p-\ln(1-(1-p)e^{-\beta x}))}+\frac{n}{\beta}-\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\frac{(1-p)x_i e^{-\beta x_i}}{1-(1-p)e^{-\beta x_i}}+$$

$$+\frac{e^{-\beta}(1-p)n\ln p}{1-(1-p)e^{-\beta})(\ln p-\ln(1-(1-p)e^{-\beta x}))^2}\ln\frac{\ln(1-(1-p)e^{-\beta x})}{\ln p}=0,$$

and the estimator is consistent.

Theorem 2: Let $X_1, X_2, \cdots, X_n$ be random variables of a sample distributed with respect to (EL) distribution with unknown parameter *p* and known β. The observations are made on the interval $[0,1)$ and in $[1,\infty)$ neither sample members nor their number are recorded. Then the pseudo maximum likelihood estimator for *p* parameter exists and is a unique root of the equation:

$$\left(\frac{n\ln\dfrac{\ln(1-(1-p)e^{-\beta x})}{\ln p}}{(\ln p-\ln(1-(1-p)e^{-\beta x}))^2}+\frac{n}{(\ln p-\ln(1-(1-p)e^{-\beta x}))\ln p}\right)\cdot$$

$$\left(\frac{ne^{-\beta}\ln p}{1-(1-p)e^{-\beta}}-\frac{n\ln(1-(1-p)e^{-\beta x})}{p}\right)-$$

$$-\frac{n}{p\ln p}-\frac{n}{1-p}-\sum_{i=1}^{n}\frac{e^{-\beta x_i}}{1-(1-p)e^{-\beta x_i}}=0,$$

and the estimator is consistent.

Table. The result of the experiment

		n=100	n=250	n=500	n=1000
β	*p*=0.2	4.58168887	4.44823007	4.35079164	4.2762195
	p=0.5	4.38954193	4.22439165	4.17349481	4.0995178
	p=0.7	4.53690541	4.39316409	4.19231157	4.0816133
p	β=5	17.2190903	2.48308392	1.31192655	0.8477539
	β=6	65.2284317	12.0087317	2.95259175	0.5622895
	β=7	225.707934	47.3803005	13.8997548	3.7100971

Simulation. The objective of our simulation is to check results received in preceding section and to compare the actual parameters values with values, which were gotten via experiments. In this study, the sample size chosen is n = 100, 250, 500 and 1000. We consider several cases for both β and *p* parameters. Estimation results are shown in the table below. First column shows estimated parameters and in the second column we consider several cases for parameter estimating. The results shown in the Table imply that using this method estimators have their error decreasing as the sample size increases.

მათემატიკა

ექსპონენციალურ-ლოგარითმული განაწილების უცნობი პარამეტრების შეფასება ცენზურირებული მონაცემებით

ა. პიჯიანი

ი. ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტის ზუსტ და საბუნებისმეტყველო მეცნიერებათა ფაკულტეტის მათემატიკის მიმართულება

(წარმოდგენილია აკადემიის წევრის ე. ნადარაიას მიერ)

სტატიაში წარმოდგენილია ექსპონენციალურ-ლოგარითმული განაწილების უცნობი პარამეტრების შეფასება ცენზურირებული მონაცემებით.

REFERENCES

1. *Tahmasbi R., Rezaei S.,* (2008) Computational Statistics and Data Analysis, **52**(8): 3889-3901.
2. *Kulldorff G.,* Contributions to the theory of estimation from grouped and partially grouped samples. ALMQVIST&WIKSELL, STOCKHOLM, GOTEBORG UPPSALA.
3. *Ferguson T.S.* (1996) A Course in large sample theory. Chapman&Hall.
4. *Dillon J.V., Lebanon G.* (2010) Statistical and Computational Tradeoffs in Stochastic Compose Likelihood. arXiv: 1003.0691,v1:29 p.

Received July, 2015