# ON THE SCALE PARAMETER OF EXPONENTIAL DISTRIBUTION

**Anca Ileana LUPAŞ**

Military Technical Academy, Bucharest, Romania (lupas_a@yahoo.com)

DOI: 10.19062/1842-9238.2017.15.2.16

***Abstract:*** *Exponential distribution is one of the widely used continuous distributions in various fields for statistical applications. In this paper we study the exact and asymptotical distribution of the scale parameter for this distribution. We will also define the confidence intervals for the studied parameter as well as the fixed length confidence intervals.*

## 1. INTRODUCTION

Exponential distribution is used in various statistical applications. Therefore, we often encounter exponential distribution in applications such as: life tables, reliability studies, extreme values analysis and others.

In the following paper, we focus our attention on the exact and asymptotical repartition of the exponential distribution scale parameter estimator.

## 2. SCALE PARAMETER ESTIMATOR OF THE EXPONENTIAL DISTRIBUTION

We will consider the random variable $X$ with the following cumulative distribution function:

$$\mathrm{F}(x;\theta)=1-e^{-\frac{x}{\theta}} \quad (x>0\ ,\ \theta>0) \tag{1}$$

where θ is an unknown scale parameter

Using the relationships between $M(X)=\theta$ ; $\sigma^2(X)=\theta^2$ ; $\sigma(X)=\theta$ , we obtain a theoretical variation coefficient $\gamma=\dfrac{\sigma(X)}{M(X)}=1$. This is a useful indicator, especially if you have observational data which seems to be exponential and with variation coefficient of the selection closed to 1.

If we consider $x_1,x_2,...x_n$ as a part of a population that follows an exponential distribution, then by using the maximum likelihood estimation method we obtain the following estimate

$$\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}x_i \tag{2}$$

Since $M\left(\hat{\theta}\right) = \theta$, it follows that $\hat{\theta}$ is an unbiased estimator for $\theta$. Similarly, because $D^2(\hat{\theta}) = \frac{\theta^2}{n}$ and $\lim_{n\to\infty} \frac{\theta^2}{n} = 0$, we obtain that $\hat{\theta}$ is an absolutely correct estimator.

The efficiency of the estimator $\hat{\theta}$ can be calculated using the following formula:

$$e_n(\hat{\theta}) = \frac{\dfrac{1}{n\int_0^\infty \left[\frac{\partial \ln f(x,\theta)}{\partial \theta}\right]^2 f(x,\theta)\,dx}}{D(\hat{\theta})} \tag{3}$$

where $f(x;\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$ $(x > 0,\ \theta > 0)$ is the probability density function of the exponential distribution. After the calculation we obtain $n\int_0^\infty \left[\frac{\partial \ln f(x;\theta)}{\partial \theta}\right]^2 \cdot f(x;\theta)\,dx = \frac{n}{\theta^2}$ and thus $e_n(\hat{\theta}) = 1$ which implies that $\hat{\theta}$ is also an efficient estimator.

Taking into account the reproducibility property of the exponential distribution we can calculate the exact distribution of the random variable $\hat{\theta}$.

A sum $X_1 + X_2 + ... + X_n$, consisting of $n$ randomly selected variables, all exponentially distribuited, has a distribution density function equal to

$$f_n(x) = \frac{1}{\theta\,\Gamma(n)}\left(\frac{x}{\theta}\right)^{n-1} \cdot e^{-\frac{x}{\theta}}. \tag{4}$$

This leads us to the cumulative distribution function

$$H_n(x) = P(\hat{\theta} \le x) = P(X_1 + ... + X_n \le nx) = \frac{1}{\theta\,\Gamma(n)}\int_0^{nx} \left(\tfrac{u}{\theta}\right)^{n-1} e^{-\frac{u}{\theta}}\,du = F_n(nx) \tag{5}$$

from which we obtain that the distribution density function of the estimator $\hat{\theta}$ is equal to

$$h_n(x) = f_n(nx)\cdot n = \frac{n}{\theta\,\Gamma(n)}\left(\frac{nx}{\theta}\right)^{n-1} e^{-\frac{nx}{\theta}} \tag{6}$$

From the above we obtain that $M(\hat{\theta}) = \int_0^\infty x\,h_n(x)\,dx = \theta$ and $D^2(\hat{\theta}) = \int_0^\infty (x-\theta)^2\,h_n(x) = \frac{\theta^2}{n}$ which is consistent with previously known results.

By using the central limit theorem we can state that the random variable $\overline{X}$ follows a normal distribution with the parameters $\theta$ and $\frac{\theta^2}{n}$, which means that $\overline{X} \approx N\left(\theta; \frac{\theta^2}{n}\right)$.

Based on this we can calculate the asymptotical cumulative distribution function for the estimator $\hat{\theta}$,

$$G_n(x) = \lim_{n\to\infty} P(\hat{\theta} \le x) = \lim_{n\to\infty} P(\overline{X} \le x) . \quad (7)$$

This shows that the estimator $\hat{\theta}$ has the same asymptotical distribution as $\overline{X}$ and a distribution density function equal to:

$$g_n(x) = \frac{\sqrt{n}}{\theta\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{(x-\theta)\sqrt{n}}{\theta}\right]^2} \quad (8)$$

Using the asymptotical distribution of the estimator $\hat{\theta}$ we can determine the confidence intervals for the scale parameter $\theta$. To do this we need to consider the reduced normal random variable $Z = \frac{(\hat{\theta}-\theta)\sqrt{n}}{\theta}$ and the significance level α.

By definition we will get

$$P(|Z| \le Z_{1-\frac{\alpha}{2}}) = 1-\alpha \quad (9)$$

where $Z_{1-\frac{\alpha}{2}}$ is the $1-\frac{\alpha}{2}$ quartile of the standard normal distribution. Equation (9) can be rewritten such as

$$P(-Z_{1-\frac{\alpha}{2}} \le \frac{(\hat{\theta}-\theta)\sqrt{n}}{\theta} \le Z_{1-\frac{\alpha}{2}}) = 1-\alpha . \quad (10)$$

We may assume that $1-\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} > 0$ because we deal with asymptotical distribution and thus *n* is sufficiently high. After carrying out the calculation we obtain for the parameter $\theta$ the following confidence interval which depends on the maximum likelihood estimation $\hat{\theta}$

$$\frac{\hat{\theta}}{1+\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}} \le \theta \le \frac{\hat{\theta}}{1-\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}} . \quad (11)$$

## 3. SET LENGTH CONFIDENCE INTERVALS FOR THE SCALE PARAMETER

Using a similar method to that which Stein proposed for the double selection method we find a set length confidence interval $\delta$ for the parameter $\theta$.

Let *p* be a system of independent selections, each with a volume *n*, taken from a population which has the same exponential distribution as the random variable *X*

$$\begin{cases} x_{11},\ldots,x_{1n} \\ \ldots \\ x_{p1},\ldots,x_{pn} \end{cases} \tag{12}$$

These selections allow us to obtain independent and identical assigned estimators

$$\hat{\theta}_i \approx N\left(\theta;\frac{\theta^2}{n}\right) \quad i=1,\ldots,p. \tag{13}$$

Let $\bar{\hat{\theta}} = \frac{1}{p}\sum_{i=1}^{p}\hat{\theta}_i$ and $s_{\hat{\theta}}^2 = \frac{1}{p-1}\sum_{i=1}^{p}(\hat{\theta}_i - \bar{\hat{\theta}})^2$. Because $\bar{\hat{\theta}}$ is the selection mean for the random variables $\hat{\theta}_i$ ; $\mathrm{i}=1,\ldots,\mathrm{p}$ we know that $\bar{\hat{\theta}} \approx N(\theta;\frac{\theta^2}{np})$.

Next let us consider a second system with *m* independent selections, each of volume *n*, taken from a population which has the same exponential distribution as the random variable $X$

$$\begin{cases} x_{p+1;1},\ldots,x_{p+1;n} \\ \ldots \\ x_{p+m;1},\ldots,x_{p+m;n} \end{cases} \tag{14}$$

These new selections allow us to calculate *m* independent and identically assigned estimators

$$\hat{\theta}_{p+j} \approx N(\theta;\frac{\theta^2}{n}) \quad j=1,\ldots,m. \tag{15}$$

Let $\bar{\bar{\hat{\theta}}}$ be the calculated average of both selection systems:

$$\bar{\bar{\hat{\theta}}} = \frac{1}{p+m}\sum_{i=1}^{p+m}\hat{\theta}_i \tag{16}$$

It is know that $\bar{\bar{\hat{\theta}}} \approx N\left(\theta;\frac{\theta^2}{n(m+p)}\right)$. Note that for *m*=0 we obtain $\bar{\bar{\hat{\theta}}} = \bar{\hat{\theta}}$. Due to the fact that $\hat{\theta} \approx N(\theta;\frac{\theta^2}{n})$ and $s_{\theta}^2$ is the variance of random variable $\hat{\theta}$ considered in equation 2, we deduce the following equation

$$\frac{s_{\hat{\theta}}^2}{\frac{\theta^2}{n}} \approx \chi_{p-1}^2 / (p-1) \tag{17}$$

Also the random variable $\bar{\bar{\hat{\theta}}}$ matches with the reduced random variable through the following equation:

$$\frac{\sqrt{n(m+p)}}{\theta}(\bar{\bar{\hat{\theta}}}-\theta) \approx N(0,1) \tag{18}$$

The random variables from equations (17) and (18) are independent and thus we can consider the following ratio

$$\frac{\frac{\sqrt{n(m+p)}}{\theta}(\bar{\bar{\hat{\theta}}}-\theta)}{\frac{s_{\hat{\theta}}\cdot\sqrt{n}}{\theta}} = \frac{\sqrt{(m+p)}}{s_{\hat{\theta}}}(\bar{\bar{\hat{\theta}}}-\theta) \approx t_{p-1} \tag{19}$$

where $t_{p-1}$ is a Student random variable with *p-1* degrees of freedom.

For the significance level α we have the following equation:

$$P\left(\left|\frac{\sqrt{m+p}}{s_{\hat{\theta}}}(\bar{\bar{\hat{\theta}}}-\theta)\right| \le t_{\frac{\alpha}{2};p-1}\right) = 1-\alpha \tag{20}$$

from which we deduce the following confidence interval for $\theta$

$$\bar{\bar{\hat{\theta}}} - \frac{s_{\hat{\theta}}t_{\frac{\alpha}{2};p-1}}{\sqrt{m+p}} \le \theta \le \bar{\bar{\hat{\theta}}} + \frac{s_{\hat{\theta}}t_{\frac{\alpha}{2};p-1}}{\sqrt{m+p}} \tag{21}$$

The length of this interval can be easily calculated and is equal to

$$l = \frac{2\cdot s_{\hat{\theta}}\cdot t_{\frac{\alpha}{2};p-1}}{\sqrt{m+p}}. \tag{22}$$

The length of this interval must not be greater than the considered length δ.

We will start with *m*=0 and compare the length *l* with *δ*. If $l \le \delta$ then the interval $\left(\bar{\hat{\theta}} - \frac{1}{2}\delta; \bar{\hat{\theta}} + \frac{1}{2}\delta\right)$ is a confidence interval for $\theta$ of size *δ*, that has a confidence coefficient greater or at least equal to $1-\alpha$, build only with the help of the first selection system.

If $l > \delta$, we need to carry out the second selection system, where *m* is equal with the smallest integer for which $\frac{2\cdot s_{\hat{\theta}}\cdot t_{\frac{\alpha}{2};p-1}}{\sqrt{m+p}} \le \delta$. In this case $\left(\bar{\bar{\hat{\theta}}} - \frac{1}{2}\delta; \bar{\bar{\hat{\theta}}} + \frac{1}{2}\delta\right)$ will be the confidence interval for $\theta$ of size *δ*, build on both selection systems and having a confidence coefficient greater then 1-*α*.

## REFERENCES

[1] Gheorghe, F.C. , Obreja , G. Teste statistice pentru selecţii de volum mic ( n ≤ 50) *Ed. Univ. Piteşti 1999*.

[2] Margolin,H.B., Maurer,W. Tests of the Kolmogorov Smirnov type for exponential data with unknow scale and related problems. *Biometrica (1976), 63,1 pg.149-160*