ROBUST'2000, 298 – 307

ON SOME EXACT DENSITIES IN AN EXPONENTIAL FAMILY

Milan Stehlík

Abstract. The aim of this paper is to give some results on the exact density of the I-divergence in the exponential family with exponentially distributed observations. It is shown in particular that the I-divergence can be decomposed as a sum of two independent variables with known distributions. Since the considered I-divergence is related to the likelihood ratio statistics, we present the method of computing the exact distribution of the likelihood ratio test. One numerical example is provided to illustrate the methods discussed.

Резюме: Уважается проблема найти распределение И-дивергенции для ешпоненциаљного семейства.

1 Introduction. Let's consider a statistical model with N independent observations $y_1, ..., y_N$ which are distributed according to the exponential densities

$$f(y_i|\vartheta) = \begin{cases} \gamma_i(\vartheta) \exp\{-\gamma_i(\vartheta) y_i\}, & \text{for } y_i > 0, \\ 0, & \text{for } y_i \leq 0. \end{cases} \tag{1}$$

Here $\vartheta := (\vartheta_1, ..., \vartheta_p)$ is the vector of unknown scale parameters, which are the parameters of interest. The parameter space Θ is an open subset of $\mathbb{R}^p$, $\gamma_i \in C^2(\Theta)$ and the matrix of first order derivatives of the mapping $\gamma := (\gamma_1, ..., \gamma_N)$ has full rank on Θ.

This model is motivated e.g. by a situation when we observe the time intervals between $(N+1)$ succesive random events in a Poisson process, which is commonly used in queueing systems (c.f. Kaufmann and Cruon, [5] and Kleinrock [6]), then the parameters $\gamma_i(\vartheta)$ are equal to the (usually parametrized) intensity γ.

Another example is the software-reliability model of Moranda developed further by Gaudoin and Soler [4]. Here the objective is to assess the failure rate based on the observation of the successive failure times,

$$0 < t_1 < t_2 < \cdots < t_n < \ldots,$$

1991 *Mathematics Subject Classification.* 62E15, 62F10, 60D05, 60F10.

Key words and phrases. Exponential family, I-divergence, likelihood ratio, maximum likelihood, GLIM, Lambert-W function.

Research was supported by the VEGA grant(Slovak Grant Agency) No 1/7295/20.

which are observations of a random point-process.

Gaudoin and Soler(1992) suppose, that

$$\gamma_i(\vartheta_1, \vartheta_2) = \vartheta_1 \lambda_i(\vartheta_2);\ i = 1, ..., N$$

with $\vartheta_1 \in \mathbb{R}^+, \vartheta_2 \in \mathbb{R}$. The parameter ϑ_1 is a parameter of scale and the parameter ϑ_2 represents the change of the reliability of the software, resulting from the correction after its i-th failure. In particular, when

$$\lambda_i(\vartheta_2) = \exp\{-(i-1)\vartheta_2\}$$

(cf. Gaudoin and Soler (1992)), then $\vartheta_2 = 0$ if the reliability is stable and $\vartheta_2 > 0$ if there is a growth of reliability. Here is proposed a rather complicated expression for the distribution function of the maximum likelihood estimator (MLE). In Pázman [8] the probability density of the parameter estimators is considered.

The other one interesting problem is a test on proportion of the exponential distribution which can be used for constructing a statistical quality control sampling plan (c.f. Rublík, [9]).

The last but not least example is GLM with exponential distribution (c.f. McCullagh and Nelder, [3]).

Such set-ups motivate interesting statistical problems. In the present paper we put stress on the testing problem and consider the likelihood ratio (LR) test. We derive exact distribution function and the density of the LR test of the canonical parameter. The table of critical constants is also given and we compare the exact distribution of the LR with the asymptotic one.

2 The I-divergence. In this section we consider some structural properties of the model to be used later. Model (1) is a regular exponential family which is a subfamily of the family

$$\{h(y|\gamma) = \prod_{i=1}^{N} \gamma_i e^{-\gamma_i y_i}\}_{\gamma \in \Gamma}. \tag{2}$$

To relate this to the general theory of exponential families we can write densities (2) in the form

$$\{\exp(-\psi(y) + t(y)^T\gamma - \kappa(\gamma))\}_{\gamma \in \Gamma} \tag{3}$$

where

$$\psi(y) = 0$$

$$t_i(y) = -y_i;\ i = 1, ..., N$$

$$\kappa(\gamma) = -\sum_{i=1}^{N} \ln(\gamma_i)$$

$$\Gamma = \{(\gamma_1, ..., \gamma_N), \gamma_i > 0,\ i = 1, ..., N\}.$$

The family (1) can be considered as a particular case of a nonlinear family with densities

$$\{\exp(-\psi(y) + t^T(y)\gamma(\vartheta) - \kappa(\gamma(\vartheta)))\}_{\vartheta \in \Theta}$$

which is "covered" by the family (3) in the sence that it is a subfamily of the family (3), and

$$\{t(y) : y \in Y\} \subseteq \{\mathbb{E}_\gamma\{t(y)\} : \gamma \in \mathit{\Gamma}\} \text{ (c.f. Pázman, [7], chpt.9).} \tag{4}$$

Since in regular exponential families we have the relation (c.f. Barndorf-Nielsen, [1])

$$\mathbb{E}_\gamma\{t(y)\} = \frac{\partial \kappa(\gamma)}{\partial \gamma},$$

by the "covering" property (4) we can associate to each value of $t(y)$ a value of γ, denoted $\hat{\gamma}_y$, which satisfies

$$\frac{\partial \kappa(\gamma)}{\partial \gamma}|_{\gamma=\hat{\gamma}_y} = t(y). \tag{5}$$

From (5) it follows that $\hat{\gamma}_y$ is the MLE of the canonical parameter γ in the covering family.

The *Kullback-Leibler* divergence

$$I(\gamma^\star, \gamma) = \mathbb{E}_{\gamma^\star}\{\frac{h(y|\gamma^\star)}{h(y|\gamma)}\};\ \gamma, \gamma^\star \in \mathit{\Gamma}$$

in the model (2) is equal to

$$I(\gamma^\star, \gamma) = -N + \sum_{i=1}^{N}\{\frac{\gamma_i}{\gamma_i^\star} - \ln(\frac{\gamma_i}{\gamma_i^\star})\}.$$

By the use of (5) we can define the I-divergence of the observed vector y

$$I_N(y, \gamma) := I(\hat{\gamma}_y, \gamma) = -N + \sum_{i=1}^{N}\{y_i\gamma_i - \ln(y_i\gamma_i)\}.$$

For $v > 0$ let us introduce a function

$$G_v(x) = \begin{cases} x - v\ln(x), & \text{for } x > 0, \\ 0, & \text{for } x \leq 0. \end{cases}$$

Further $G := G_1$ is used. This allows to write

$$I_N(y, \gamma) = \sum_{i=1}^{N}\{G(\gamma_i y_i) - G(1)\}. \tag{6}$$

3 The geometric approach to the I-divergence.

In this Section we give our main results, Theorems 3.2.1. and 3.2.2..

Let us introduce notation $y \sim \exp(\gamma_1, ..., \gamma_N),\ \gamma \in \mathit{\Gamma}$ when $y = (y_1, ..., y_N)$ is distributed according to the density $h(y|\gamma)$ (see (2)).

3.1 The case of one observation.

Let's have $y \sim \exp(\bar{\gamma})$. Then the I-divergence I_1 has the form $I_1(y,\gamma) = -1 + G(\gamma y)$ (see (6)). The function $G : (0,+\infty) \to < 1,+\infty)$ has as an inverse multifunction $G^{-1} :< 1,+\infty) \to 2^{\mathbb{R}}$ with two differrentiable selectors $\min G^{-1}$ and $\max G^{-1}$. In the Appendix A is given that

$$\min G^{-1}(u) = -\mathcal{LW}(0,-e^{-u}),\ u > 1$$

and

$$\max G^{-1}(u) = -\mathcal{LW}(-1,-e^{-u}),\ u > 1$$

where $\mathcal{LW}(k,w),\ k \in \mathbb{Z},\ w \in \mathbb{C}$ is the value of the $k-$th branch of the Lambert-W function at the point w (see Appendix A).

Let $F(x)$ is the cummulative distribution function (c.d.f.) of $I_1(y,\gamma)$. Then $F(x) = 0$ for $x \leq 0$. For $x > 0$ we have

$$F(x) = P\{G(\gamma y) < 1 + x\} = P\{\min G^{-1}(1+x) < \gamma y < \max G^{-1}(1+x)\} =$$

$$= P\{-\frac{\bar{\gamma}}{\gamma}\mathcal{LW}(0,-e^{-1-x}) < \bar{\gamma} y < -\frac{\bar{\gamma}}{\gamma}\mathcal{LW}(-1,-e^{-1-x})\} =$$

$$= e^{\frac{\bar{\gamma}}{\gamma}\mathcal{LW}(0,-e^{-1-x})} - e^{\frac{\bar{\gamma}}{\gamma}\mathcal{LW}(-1,-e^{-1-x})}.$$

To obtain the density $f(x) = \frac{dF(x)}{dx}$ we follow the rules for derivatives explained in Appendix A. Thus we have proved the following proposition.

Propostion.

Let $y \sim \exp(\bar{\gamma})$. Then the c.d.f. of $I_1(y,\gamma)$ has the form

$$F(x) = \begin{cases} e^{\frac{\bar{\gamma}}{\gamma}\mathcal{LW}(0,-e^{-1-x})} - e^{\frac{\bar{\gamma}}{\gamma}\mathcal{LW}(-1,-e^{-1-x})}, & \text{for } x > 0, \\ 0, & \text{for } x \leq 0, \end{cases}$$

and the density of $I_1(y,\gamma)$ has the form

$$f(x) = \begin{cases} q(1,x,\frac{\bar{\gamma}}{\gamma}) - q(0,x,\frac{\bar{\gamma}}{\gamma}), & \text{for } x > 0, \\ 0, & \text{for } x \leq 0. \end{cases}$$

Here

$$q(k,r,s) = \frac{s\mathcal{LW}(-k,-e^{-1-r})}{1+\mathcal{LW}(-k,-e^{-1-r})} e^{s\mathcal{LW}(-k,-e^{-1-r})} \text{ for } k \in \mathbb{Z};\ r,s > 0.$$

3.2 The case of N observations.

Let us consider the $(N-1)$-dimensional Euclidean simplex

$$\sigma_{N-1} = \{u \in \mathbb{R}^{N-1} : 0 < u_1 < 1, ...,\ 0 < u_{N-1} < 1 - u_1 - ... - u_{N-2}\}$$

and the positive cone $\mathcal{E}_N = \{x \in \mathbb{R}^N, x_1 > 0, ..., x_N > 0\}$ of the N-dimensional Euclidean space $\mathbb{E}^N$. It holds $\mathcal{E}_N = \sigma_{N-1} \times \mathcal{E}_1$. For $\bar{\gamma} \in \Gamma$ we define the map $\nu_{\bar{\gamma}} : \mathcal{E}_1 \times \sigma_{N-1} \to \mathcal{E}_N$

$$y_i = \begin{cases} \dfrac{r}{\bar{\gamma}_i} u_i, & \text{for } 1 \le i \le N-1, \\ \dfrac{r}{\bar{\gamma}_N}(1 - u_1 - u_2 - ... - u_{N-1}), & \text{for } i = N. \end{cases} \tag{7}$$

This map is regular and $dy = \dfrac{r^{N-1}}{\bar{\gamma}_1 ... \bar{\gamma}_N} dr du.$

Theorem 3.2.1.

Let's have $y \sim \exp(\bar{\gamma}_1, ..., \bar{\gamma}_N)$ and $\delta > 0$.
Then $I_N(y, \delta\bar{\gamma})$ is the sum of two independent variables R_N and S_N where

$$R_N = G_N(\delta r) - G_N(N) \tag{8}$$

and

$$S_N = -\ln\{N^N u_1 ... u_{N-1}(1 - u_1 - ... - u_{N-1})\}. \tag{9}$$

Here r is $\Gamma(N,1)$-distributed and $(u_1, ..., u_{N-1})$ is uniformly distributed on the simplex σ_{N-1}.

Proof.

Into the characteristic function ψ of the variable $I_N(y, \delta\bar{\gamma})$ we introduce the coordinates $(r, u_1, ..., u_{N-1})$ corresponding to the map $\nu_{\bar{\gamma}}$ (see (7)).

After the substitution we obtain $\psi(\mu) = \psi_1(\mu)\psi_2(\mu)$ where

$$\psi_1(\mu) = \int_0^{+\infty} \exp\{i\ \mu\{G_N(\delta r) - G_N(N)\}\} \frac{r^{N-1}}{\Gamma(N)} e^{-r} dr$$

and

$$\psi_2(\mu) = \int_{\sigma_{N-1}} \exp\{-i\mu \ln\{N^N u_1 ... u_{N-1}(1 - u_1 - .. - u_{N-1})\}\} du$$

are the characteristic functions of the random variables R_N and S_N from the Theorem 3.2.1.. This completes the proof.

$\square$

Theorem 3.2.2.

The c.d.f. of the variable R_N from the Theorem 3.2.1. has the form

$$F_N(\rho) = \begin{cases} \mathcal{F}_N(-N\delta^{-1}\ \mathcal{LW}(-1, -e^{-\frac{\rho+N}{N}})) - \mathcal{F}_N(-N\delta^{-1}\ \mathcal{LW}(-e^{-\frac{\rho+N}{N}})), & \rho > 0, \\ 0, & \rho \le 0 \end{cases}$$

and the density of the R_N has the form

$$f_N(\rho) = \begin{cases} h(N, 1, \rho, \delta^{-1}) - h(N, 0, \rho, \delta^{-1}), & \textit{for } \rho > 0, \\ 0, & \textit{for } \rho \le 0. \end{cases}$$

Here $\mathcal{F}_N(x)$ is the c.d.f. of the $\Gamma(N,1)$-distribution and for $N \in \mathbb{N},\ k \in \mathbb{Z};\ r,s > 0$ we define

$$h(N,k,r,s) = \frac{(-N)^{N-1}s^N}{\Gamma(N)} \frac{\{\mathcal{LW}(-k,-e^{-\frac{r+N}{N}})\}^N}{1+\mathcal{LW}(-k,-e^{-\frac{r+N}{N}})} e^{Ns\mathcal{LW}(-k,e^{-\frac{r+N}{N}})}.$$

Proof. For $\rho \leq 0$ we obtain $F_N(\rho) = 0$. Let us have $\rho > 0$. Then

$$F_N(\rho) = P\{-N\delta^{-1}\mathcal{LW}(-e^{-\frac{\rho+N}{N}}) < r < -N\delta^{-1}\mathcal{LW}(-1,-e^{-\frac{\rho+N}{N}})\} =$$

$$= \mathcal{F}_N(-N\delta^{-1}\mathcal{LW}(-1,-e^{-\frac{\rho+N}{N}})) - \mathcal{F}_N(-N\delta^{-1}\mathcal{LW}(-e^{-\frac{\rho+N}{N}})).$$

The density of the R_N follows after the differentiation of $F_N(\rho)$ (see Appendix A).

□

3.3 The comparison to the normal model.

The components R_N and S_N (see Theorem 3.2.1.) are independent variables (it means that our decomposition is also deconvolution), R_N is the "radial" component depending only on the "radial" coordinate r and S_N is the simplectic component depending only on the simplectic coordinates $(u_1, ..., u_{N-1})$.

There is an analogy between the radial component of the I-divergence in the normal linear regression and in the model (1). To see this we note that in the normal regression with $y \sim \mathcal{N}_N(\vartheta, \sigma^2 I)$ we have

$$I_N(t(y), \vartheta) = \frac{1}{2\sigma^2}||y - \vartheta||^2$$

where $\vartheta \in \Theta$ is the unknown parameter of the interest and σ is a known variance parameter. Here we must use the spherical coordinates $(r(\vartheta), \varphi_1, ..., \varphi_{N-1})$ of $\mathbb{E}^N - \eta(\vartheta)$ and the decomposition contains only the radial component

$$I_N(t(y), \vartheta) = R_N^*(r, \vartheta) = \frac{1}{2\sigma^2} r(\vartheta)^2.$$

The LR test of the hypothesis $H_0 : \vartheta = \vartheta_0$ versus $H_1 : \vartheta \neq \vartheta_0$ based on the statistics $-2\ln\lambda = 2R_N^*(r, \vartheta_0)$ has χ^2_N-distribution under the null hypothesis.

To make the comparison to the exponential model let us consider the LR test of the hypothesis $H_0 : \gamma = \gamma_0$ versus $H_1 : \gamma \neq \gamma_0$ in the simple GLM $y \sim \exp(\gamma, ..., \gamma)$ (Poisson process, see Section 4). The LR statistics $-2\ln\lambda = 2R_N(r, \gamma_0)$ is asymptotically χ^2_1-distributed under the null hypothesis (see Wilks, [10]).

4 Illustrative example.

In this Section we show some applications of our results to the LR test of the parameter of interest in the model (1).

We consider a statistical model with N independent observations $y_1, ..., y_N$ which are distributed according to the exponential density

$$f(y_i|\gamma) = \begin{cases} \gamma \exp(-\gamma y_i), & \text{for } y_i > 0, \\ 0, & \text{for } y_i \le 0. \end{cases} \qquad (10)$$

Here γ is an unknown scale parameter (intensity) and the parameter space $\Gamma = \mathcal{E}_1$ is the open subset of $\mathbb{R}$. This model is a special case of the model (1).

4.1. The exact LR test of the intensity.

We consider the test

$$H_0 : \gamma = \gamma_0 \text{ versus } H_1 : \gamma \neq \gamma_0 \qquad (11)$$

in the model (10).

Theorem 4.1.1.

The test statistics $-\ln\lambda$ of the LR test to the hypothesis (11) has the form

$$\tau_N(y) = G_N(\gamma_0 \sum_{i=1}^{N} y_i) - G_N(N) \qquad (12)$$

and $\tau_N = R_N$, where R_N is the "radial" component of the $I_N(y, \gamma_0)$ (see (8)) .

Under the null hypothesis the c.d.f. and the density of τ_N is given in the Theorem 3.2.2. where we take $\delta = 1$ and the Wilks statistics $-2\ln\lambda$ has the c.d.f. of the form

$$F_N(\tau) = \begin{cases} \mathcal{F}_N(-N\mathcal{LW}(-1, -e^{-1-\frac{\tau}{2N}})) - \mathcal{F}_N(-N\mathcal{LW}(-e^{-1-\frac{\tau}{2N}})), & \tau > 0, \\ 0, & \tau \le 0, \end{cases} \qquad (13)$$

and the density of the form

$$f_N(\tau) = \begin{cases} \frac{1}{2}\{h(N, 1, \frac{\tau}{2}, 1) - h(N, 0, \frac{\tau}{2}, 1)\}, & \textit{for } \tau > 0 \ , \\ 0, & \textit{for } \tau \le 0. \end{cases}$$

The proof is presented in Appendix B.

Table 3 contains the critical constants $c_{\alpha,N}$ of the LR test $-2\ln\lambda$ on the level of significance α in small dimensions N obtained from its exact distribution (13).

Table 3. Critical constants $c_{\alpha,N}$.

$\alpha \setminus N$	1	2	3	4	5
0.005	8.852997810	8.460579550	8.287166100	8.192107640	8.132601599
0.01	7.498403700	7.136930670	6.983699006	6.901147440	6.849915290
0.02	6.15478803	5.831756370	5.700796220	5.631435842	5.588746670
0.05	4.407670803	4.149077148	4.050520530	3.999436000	3.968318015

4.2 Comparison with the χ^2-asymptotics.

The MLE $\hat{\gamma}$ of the parameter γ in the model (10) is consistent and $-2\ln\lambda$ has asymptotically χ_1^2-distribution (c.f. Wilks, [10]). In this Section we present how the exact distribution of the LR test differs from the asymptotic one.

Let us present the Table 4 of the critical constants k_α obtained from the asymptotic χ_1^2-distribution of the $-2\ln\lambda$ on the level of significance $0.005,\ 0.01,\ 0.02$ and 0.05.

Table 4. Critical constants k_α .

α	k_α
0.005	7.879438577
0.01	6.634896601
0.02	5.411894431
0.05	3.841458821

The following Table 5 shows that the exact levels of significance $\alpha_{e,N}$ of the critical constants k_α are higher in small dimensions N than the asymptotic ones.

Table 5. The exact levels of significance $\alpha_{e,N}$.

$\alpha \setminus N$	1	2	3	4	5
0.005	0.008224735	0.006771825	0.006204483	0.005908783	0.005728607
0.01	0.015599286	0.013037809	0.012058871	0.011552053	0.011244013
0.02	0.029448482	0.025065314	0.023424550	0.022579936	0.022067611
0.05	0.067701923	0.059361294	0.056314364	0.054754992	0.053810812

$\alpha \setminus N$	10	20	30	40	50
0.005	0.005364607	0.005182040	0.005121267	0.005090910	0.005072708
0.01	0.010622626	0.010311006	0.010207222	0.010155366	0.010124268
0.02	0.021035190	0.00517344	0.020344780	0.020258532	0.020206799
0.05	0.051909321	0.050954881	0.050636560	0.050477398	0.050381907

Appendix A. The Lambert-W function satisfies

$$\mathcal{LW}(z)e^{\mathcal{LW}(z)} = z.$$

As the equation $y\exp(y) = z$ has an infinite number of solutions y_k for each (non-zero) value of $z \in \mathbb{C}$, Lambert-W has an infinite number of branches. Exactly one of these branches is analytic at 0. Usually this branch is referred to as the principal branch of Lambert-W, and is denoted by $\mathcal{LW}(z)$ or $\mathcal{LW}(0, z)$. The other branches $\mathcal{LW}(k, z)$, $k \in \mathbb{Z} \setminus \{0\}$ all have a branch point at 0. The principal branch and the pair of branches $\mathcal{LW}(-1, z)$ and $\mathcal{LW}(1, z)$ share an order 2 branch point at point $-\exp(-1)$. The branch cut dividing these branches is the subset of the real line from $-\infty$ to $-\exp(-1)$, and the values of the branches of Lambert-W on this branch cut are assigned using the rule of counter-clockwise continuity around the branch point. This means that $\mathcal{LW}(z)$ is real-valued for $z \in (-\exp(-1), \infty)$ and similarly, the branch corresponding to -1, $\mathcal{LW}(-1, z)$, is real-valued on the interval $(-\exp(-1), 0)$. For all the branches other than the principal branch, the branch cut dividing them is the negative real axis. The branches are numbered up and down from the real axis (this is very similar to the way the branches of the logarithm are indexed by the multiple of $2\pi i$ which must be subtracted from the imaginary part to recover the principal branch).

The asymptotic behavior of the Lambert-W at complex infinity and at 0 (for the non-principal branches) is given by

$$\mathcal{LW}(k,z) \sim \ln(k,z) - \ln(\ln(k,z)) + \sum_{m,n=0}^{+\infty} \frac{\{c(m,n)\ln(\ln(k,z)\}^{m+1}}{\ln(k,z)^{m+n+1}},$$

where $\ln z$ denotes the principal branch of the logarithm,

$$\ln(k,z) = \ln z + 2k\pi i$$

and the $c(m,n)$ are constants independent of k. The expansion for $\mathcal{LW}(-1,z)$ is not valid for $z \to 0$ along the negative real axis (the effect of the branch point at $-\exp(-1)$ must be considered), but holds otherwise.

Without the proof (which is rather technical) we present the following Lemma:

Lemma.

Let $v > 0$. Then for $u > G_v(v)$ we have the following expressions in terms of Lambert-W function:

$$\min G_v^{-1}(u) = -v\mathcal{LW}(0, -\frac{e^{-\frac{u}{v}}}{v})$$

and

$$\max G_v^{-1}(u) = -v\mathcal{LW}(-1, -\frac{e^{-\frac{u}{v}}}{v}).$$

Here $\min G_v^{-1}$ and $\max G_v^{-1}$ are selectors of the multifunction $G_v^{-1} :< G_v(v), +\infty) \to 2^{\mathbb{R}}$. Further let us define $f_k(x) = \mathcal{LW}(-k, -e^{-x}),\ x > 1,\ k = 0, 1$.

$$\textit{Then } f_k \in C^1((1,+\infty),\mathbb{R}) \textit{ and } f_k'(x) = -\frac{f_k(x)}{1+f_k(x)} \textit{ holds.}$$

Previous Lemma shows that all densities and c.d. functions in this paper are real-valued functions and gives the formula for computing of derivatives.

The Lambert-W function has a good implementation in mathematical software such as Matlab 5.3 or Maple V Release 5.1. For more information see Corless [2].

Appendix B.

Proof of the Theorem 4.1.1. The LR to the hypothesis (11) is

$$\lambda(y) = \frac{f(y|\gamma_0)}{f(y|\hat{\gamma})} \tag{14}$$

where $f(y|\gamma) = \gamma^N \exp\{-\gamma \sum_{i=1}^{N} y_i\}$ and $\hat{\gamma} = N\{\sum_{i=1}^{N} y_i\}^{-1}$ is the MLE of the parameter γ. Putting it into the (14) we obtain the formula (12). Application of Theorem 3.2.1. completes the proof.

□

References

1. Barndorff-Nielsen OE, *Information and Exponential Families In Statistical Theory*, John Wiley & Sons, Ltd., New York, part 3, 1978.
2. Corless RM, Gonnet GH, Hare DEG, Jeffrey DJ , Knuth DE, *On The Lambert W Function, Maple Share Library.*
3. McCullagh P, Nelder JA, *Generalized Linear Models, (285-322)*, CHAPMAN & HALL, 1989.
4. Gaudoin O, Soler JL, *Statistical Analysis of the Geometric De-Eutrophication Software Reliability Model*, IEEE Transactions on Reliability **41** (1992 Dec).
5. Kaufmann A, Cruon R, *Queueing Systems*, Mir, Moscow, 1965. (Russian)
6. Kleinrock L, *Queueing Systems*, vol. 1:Theory, John Wiley & Sons, Toronto, 1975.
7. Pázman A, *Nonlinear statistical Models.*, Kluwer Acad. Publ., Dordrecht, chapters 9.1, 9.2, 1993.
8. Pázman A, *The Density of the Parameter Estimators when the Observations are Distributed Exponentially*, Metrika **44** (1996), Physica-Verlag, Heidelberg, 9–26.
9. Rublík F, *Some tests on exponential populations*, Tatra Mountains Math. Publ. **7** (1996), 229–235.
10. Wilks SS, *Mathematical Statistics, (416-420)*, Nauka, Moscow, 1967. (Russian)

Department of Probability and Statistics, Faculty of Mathematics, Physics and Informatics, Comenius University Bratislava, Mlynska Dolina, SK – 842 15 Bratislava, Slovakia
E-mail: Stehlik@fmph.uniba.sk