مجلة التجارة والتمويل

/https://caf.journals.ekb.eg

كلية التجارة – جامعة طنطا

العدد : الرابع

ديسمبر ٢٠٢١

Characterization and Estimation of Some Probability Distributions by Minimum χ^2-Divergence Principle

Mohamed Abdelsalam Agamy

Statistics Department, Faculty of Commerce
Al-Azhar University (Boys' Branch), Egypt.

Corresponding author e-mail: dr.muhammad_alagamy@azhar.edu.eg

Characterization and Estimation of Some Probability Distributions by Minimum χ^2−Divergence Principle

Abstract

A probability distribution can be characterized through various methods, given a prior probability distribution g(x) and some available information on moments of the random variable X, the original probability distribution f(x) is determined such that the χ2- divergence measure of the distance between f(x) and g(x) is a minimum. The minimum chi-squared density function f(x) is determined. The expressions of the non-central moments as well as the cumulative distribution function F(x), survival function S(x), and hazard function h(x) are also determined under different available information on moments. In this paper we discussed the characterization of the exponential and exponentiated exponential (EExp) distribution. The available information on moments included: first moment, second moment and the first two moments. Some illustrative examples are included for special values of the parameters. Moreover, we have considered the principle of minimizing chi square divergence and used it for characterizing the probability distributions given a prior distribution as the exponential and the partial information in terms of averages and variance. It is observed that the probability distributions which minimize the χ2 -distance also minimize the Kullback's measure of the directed divergence. It is shown that by applying the minimum chi square divergence principle, new probability distributions are obtained. Hence, the probability models can be revised to find the best estimated probability models given the new information on moments.

Keywords: *Available Information, Characterization, Exponential Distribution, Exponentiated Exponential Distribution, Hazard Function,* χ^2*−Divergence Principle.*

1. Introduction

A characterization theorem states that a certain object, such as a function or a space, is the only one that has the features described in the statement. Accordingly, a characterization of a probability distribution asserts that it is the only probability distribution that meets certain criteria. Zolotarev (1976). On the probability space we define the space $x = \{X\}$ of random variables with values in measurable metric space (U, d_u) and the space $y = \{Y\}$ of random variables with values in measurable metric space (V, d_u). We grasp broad difficulties of description of some set C by characterizations of probability distributions in the space x by extracting the sets $A \subseteq x$ and $B \subseteq y$ which describe the properties of random variables $X \in A$ and their images $Y = FX \in B$, obtained by means of a specially chosen mapping $F: x \to y$. The description of the properties of the random variables X and of their images $Y = FX$ is equivalent to the indication of the set $A \subseteq x$ from which X must be taken and of the set $B \subseteq y$ into which its image must fall, see Ord (1975).So, the set which interests us appears therefore in the following form:

$$X \in A, FX \in B \Leftrightarrow X \in C, i.e. C = F^{-1}B,$$

Where $F^{-1}B$ denotes the complete inverse image of B in A. The problem of characterizing a distribution is an important problem which has recently attracted the attention of many researchers. Thus, various characterizations have been established in many different directions .The minimum χ2- divergence principle states: When a prior probability density function (PDF) of X , g(x), which estimates the underlying PDF f(x) is given in addition to some constraints, then among all the density functions f(x) which satisfy the given constraints we should select that probability density function which minimizes the χ^2- divergence, see Yanushkevichius (2014).

The rest of the paper is organized as follows: section 2 presents minimum χ^2-Divergence probability distributions. In section 3, the exponential distribution, and the Available information of the moments for the exponential distribution is presented. While section 4 address the exponentiated exponential distribution, and the Available information of the moments for this distribution. The results of the numerical application have been presented in Section 5. Finally, Section 6 offers the concluding remarks.

1. Minimum χ^2-Divergence Probability Distributions

The maximum entropy principle and the minimum discrimination (minimum cross-entropy) principle are used in characterization of many univariate and multivariate probability distributions. Kawamura and Iwase (2003) applied the maximum entropy principle in characterizing the power inverse Gaussian, and generalized gumbel probability distributions. The equivalence of minimum discrimination information principle and the maximum likelihood principle and Gauss's principle have been discussed in Shore and Johnson (1980). Here, we make use of and prove the results of Kumar (2005, 2006) and apply them to characterize some exponential family distribution, under different available information on moments.

Let the random variable X be a continuous variable with PDF f(x), defined on the interval (0, ∞). According to Kullback (1997), the minimum cross entropy principle states that: When a prior PDF of X, g(x), which estimates the underlying PDF f(x) is given in addition to some constraints, then among all the density functions f(x) which satisfy the given constraints, then we can choose the PDF that has the least Kullback and Leibler divergence:

$$K(f,g) = \int_0^\infty f(x) \ln \tfrac{f(x)}{g(x)} \, dx. \qquad (1)$$

Kumar and Taneja (2004) have considered the minimum χ^2 – divergence principle as follows:

when a prior PDF X, g(x) of the random variable is the prior PDF estimating the underlying PDF f(x) is given with some available information on moments of f(x), then f(x) is called the minimum χ^2 – divergence PDF if it satisfies the given partial available information and minimizes the χ^2 – divergence:

$$\chi^2(f,g) = \int_0^\infty \frac{f^2(x)}{g(x)} dx - 1 \ . \qquad (2)$$

The minimum cross-entropy principle and the minimum χ^2 – divergence principle applies to both the discrete and continuous random variables. Kumar and Taneja (2004) defined the minimum χ^2 – divergence probability distribution for continuous random variable as:

Definition 2.1 f(x) is the probability density of the minimum χ^2 – divergence continuous probability distribution of random variable X if it minimizes the χ^2 – divergence:

$$\chi^2(f,g) = \int_0^\infty \frac{f^2(x)}{g(x)} dx - 1$$

That is, f(x) is the minimum χ^2 – divergence PDF of the random variable X if ;

1. Probability density function constraints: f(x) ≥ 0, and ∫f(x) d x = 1,
2. a prior probability density function: g(x) ≥ 0, and ∫g(x) d x = 1,

3. It minimizes χ^2 (f, g) defined in Eq. (2), and

3. Partial information in terms of averages:

$$\int_0^\infty [h(x)]^t f(x) dx = \mu'_{h(X),t,f}$$

Where $h(x)$ is any real-valued function of X and $t = 0,1,2,\ldots,r$, see Ahsanullah (2017). We present the main result in the following lemma:

Lemma 2.1 Let the given prior PDF be g(x) and the constraints are:

$$f(x) \geq 0, \int_0^\infty f(x) dx = 1, \int_0^\infty [h(x)]^t f(x) dx = \mu'_{h(X),t,f} \qquad (3)$$

Then the minimum χ^2 – divergence probability distribution of the random variable x has the PDF;

$$f(x)=\frac{g(x)}{2}\left(\alpha_0+\sum_{t=1}^{r}\alpha_t[h(x)]^t\right). \tag{4}$$

The coefficients $\alpha_t, t=0,1,2,\ldots,r$ are determined from;

$$\int_0^\infty \frac{g(x)}{2}\left(\alpha_0+\sum_{t=1}^{r}\alpha_t[h(x)]^t\right)dx=1, \tag{5}$$

$$\int_0^\infty [h(x)]^t\,\frac{g(x)}{2}\left(\alpha_0+\sum_{t=1}^{r}\alpha_t[h(x)]^t\right)dx=\mu'_{h(X),t,f}\,. \tag{6}$$

The minimum χ^2 – divergence measure is:

$$\chi^2_{\min}(f,g)=\int_0^\infty \frac{g(x)}{4}\left(\alpha_0+\sum_{t=1}^{r}\alpha_t[h(x)]^t\right)^2 dx-1. \tag{7}$$

3. Exponential Distribution

A continuous non-negative random variable X (X ≥ 0) is called to have an exponential distribution with parameter λ; λ > 0, see Balakrishnan (2019). The PDF is given by;

$$g(x;\lambda)=\lambda\exp(-\lambda x), x>0, \lambda>0 \tag{8}$$

Or equivalently, if its distribution functions is given by;

$$G(x)=1-\exp(-\lambda x), x>0$$

$$\mu'_{k,g}=\int_0^\infty x^k\cdot\lambda e^{-\lambda x}\,dx=\frac{\Gamma(k+1)}{\lambda^k}=\frac{k!}{\lambda^k}$$

$$g_n(x)=\frac{1}{n!}x^n\lambda^{n+1}\exp(-\lambda x), x>0, \lambda>0.$$

$$G_n(x)=\int_0^x \frac{1}{n!}\lambda^{n+1}u^n\exp(-\lambda u)\,du, x>0$$

if follows that the survival function S(x) is given by;

$$S(x)=1-G(x)=\exp(-\lambda x), \quad for\ x>0 \tag{9}$$

3.1 The Available Information is the n^{th} Moment for Exponential Distribution

Let the prior is the exponential distribution with parameter λ, and the available information is $\mu'_{n,f}$, n^{th} non-central moment of f(x; λ), see Ahsanullah and Hamedani (2010). Using equations (8) – (9) under minimum χ^2– divergence, we present the following theorem:

Theorem 3.1 Suppose that the observed estimated prior is the exponential distribution with parameter λ and the available information be:

$$f(x) \geq 0,\ \int_0^\infty f(x)dx = 1,\ E(X^n) = \int_0^\infty x^n f(x)dx = \mu'_{n,f}$$

Then the corresponding minimum χ^2– divergence PDF f(x; λ) is;

$$f(x;\lambda) = A(n)g(x;\lambda) + B(n)\ g_n(x;\lambda).$$

Equivalently;

$$f(x;\lambda) = w_1(n)\ g(x;\lambda) + w_2(n)\ g_n(x;\lambda), \quad (10)$$

$$\mu'_{n,f} = t\frac{n!}{\lambda^n} + (1-t)\frac{(2n)!}{n!\lambda^n},\ 0 \leq t \leq 1. \quad (11)$$

$$w_1(n) = A(n) = \frac{(2n)! - \mu'_{n,f}(n!)\lambda^n}{(2n)! - (n!)^2},$$

$$w_2(n) = \lambda^n B(n) = \frac{\lambda^n(\lambda^n \mu'_{n,f} - n!)}{(2n)! - (n!)^2}.$$

Eq.(10) expresses the PDF f(x; λ) as a weighted mixture of two PDFs g(x; λ) and g_n (x; λ) with respective weights $w_1(n)$ and $w_2(n)$ and the minimum χ^2– divergence measure between f(x; λ) and g(x; λ) is;

$$\chi^2_{\min}(f,g) = \frac{\left(\lambda^n \mu'_{n,f} - n!\right)^2}{(2n)! - (n!)^2}.$$

We have the following corollaries;

Corollary 3.1 When t = 1, Eq. (10) implies $\mu'_{n,f} = n!/\lambda^n$, and $w_1(n)$ = 1 and $w_2(n)$ = 0. Hence, f(x; λ) = g(x; λ) with zero minimum χ^2– divergence measure between f(x; λ) and g(x; λ).

Corollary 3.2 When t = 0, Eq. (11) implies $\mu'_{n,f} = (2n)!/(n!\lambda^{n})$, $w_1(n) = 0$ and $w_2(n) = 1$. Hence, f(x; λ) = g_n (x; λ) defined in Eq. (8) and

$$\chi^2_{\min}(f,g) = \frac{(2n)! - (n!)^2}{(n!)^2}.$$

Corollary 3.3 When 0< t <1, $\mu'_{n,f} = t\, n!/\lambda^{n} + (1-t)(2n)!/(n!\lambda^{n})$,

and f(x; λ) is the proper mixture described in Eq. (9). For;

t = ½, $\mu'_{n,f} = \frac{1}{2}\left(n!/\lambda^{n} + (2n)!/(n!\lambda^{n})\right)$, w_1 (n) = ½, w_2 (n) = ½

And;

$$f(x;\lambda) = \tfrac{1}{2} g(x;\lambda) + \tfrac{1}{2} g_n(x;\lambda), \tag{12}$$

$$\chi^2_{\min}(f,g) = \frac{(2n)! - (n!)^2}{4(n!)^2}.$$

3.2 Some Properties of the Minimum χ^2 – divergence in this Case

Using equations (10) – (12), we explore some properties of the minimum χ2 – divergence in available information is the n^{th} moment for exponential distribution;

Property 3.1 The r^{th} non-central moment, mean, and variance of f(x; λ) are;

$$\mu'_{r,f} = A(n)\frac{r!}{\lambda^{r}} + B(n)\frac{(n+r)!}{\lambda^{n+r}},$$

$$\mu'_{1,f} = A(n)\frac{1}{\lambda} + B(n)\frac{(n+1)!}{\lambda^{n+1}},$$

$$\sigma^2_f = \frac{2}{\lambda^2}A(n) + \frac{(n+2)!}{\lambda^{n+2}}B(n) - \mu'^2_{1,f}$$

Property 3.2 The CDF of f(x) has the form;

$$F(x) = w_1(n)G(x) + w_2(n)G_n(x).$$

The functions G(x) and G_n (x) are the CDFs of g(x; λ) and g_n (x; λ) respectively. That is in equations (8) and (9).

Property 3.3 The survival function of f(x) is;

$$S(x) = w_1(n)[1 - G(x)] + w_2(n)[1 - G_n(x)].$$

Property 3.4 The hazard function of f(x; λ) is:

$$h(x) = \frac{w_1(n)\, g(x;\lambda) + w_2(n)\, g_n(x;\lambda)}{w_1(n)[1 - G(x)] + w_2(n)[1 - G_n(x)]}.$$

3.3 Available Information is the First Two Moments

When the prior distribution has a PDF g(x) and the available information is $\mu'_{k,f}, k = 1, 2$, then we provided the following theorem;

Theorem 3.2 Suppose that the observed estimated distribution has a PDF g(x). Let the available information be;

$$f(x) \ge 0,\ \int_0^\infty f(x)dx = 1, E(X) = \int_0^\infty xf(x)dx = \mu'_{1,f},\ E(X^2) = \int_0^\infty x^2 f(x)dx = \mu'_{2,f}$$

The minimum χ^2 – divergence PDF f(x) is:

$$\mathrm{f}(\mathrm{x}) = (A + Bx + Cx^2)g(x)$$

Equivalently;

$$\mathrm{f}(\mathrm{x}) = w_1 g(x) + w_2 g_1(x) + w_3 g_2(x)$$

$$A = \frac{\mu'_{1,f}(\mu'_{2,g}\mu'_{3,g} - \mu'_{1,g}\mu'_{4,g}) + \mu'_{2,f}(\mu'_{1,g}\mu'_{3,g} - \mu'^2_{2,g}) + \mu'_{2,g}\mu'_{4,g} - \mu'^2_{3,g}}{\mu'_{2,g}\mu'_{4,g} + 2\mu'_{1,g}\mu'_{2,g}\mu'_{3,g} - \mu'^2_{1,g}\mu'_{4,g} - \mu'^2_{3,g} - \mu'^3_{2,g}}, \quad (13)$$

$$B = \frac{\mu'_{1,f}(\mu'_{4,g} - \mu'^2_{2,g}) + \mu'_{2,f}(\mu'_{1,g}\mu'_{2,g} - \mu'_{3,g}) + \mu'_{2,g}\mu'_{3,g} - \mu'_{1,g}\mu'_{4,g}}{\mu'_{2,g}\mu'_{4,g} + 2\mu'_{1,g}\mu'_{2,g}\mu'_{3,g} - \mu'^2_{1,g}\mu'_{4,g} - \mu'^2_{3,g} - \mu'^3_{2,g}} \quad (14)$$

$$C = \frac{\mu'_{1,f}(\mu'_{1,g}\mu'_{2,g} - \mu'_{3,g}) + \mu'_{2,f}(\mu'_{2,g} - \mu'^2_{1,g}) + \mu'_{1,g}\mu'_{3,g} - \mu'^2_{2,g}}{\mu'_{2,g}\mu'_{4,g} + 2\mu'_{1,g}\mu'_{2,g}\mu'_{3,g} - \mu'^2_{1,g}\mu'_{4,g} - \mu'^2_{3,g} - \mu'^3_{2,g}}, \quad (15)$$

$$w_1 = A,\ w_2 = B\,\mu'_{1,g},\ w_3 = C\,\mu'_{2,g}$$

Proof Applying the available information on moments;

$$f(x) \ge 0,\ \int_0^\infty f(x)dx = 1, E(X) = \int_0^\infty xf(x)dx = \mu'_{1,f},\ E(X^2) = \int_0^\infty x^2 f(x)dx = \mu'_{2,f},$$

We get the following system of three linear equations in the unknowns A, B, and C:

$$A + \mu'_{1,g} B + \mu'_{2,g} C = 1$$
$$\mu'_{1,g} A + \mu'_{2,g} B + \mu'_{3,g} C = \mu'_{1,f}$$
$$\mu'_{2,g} A + \mu'_{3,g} B + \mu'_{4,g} C = \mu'_{2,f}$$

Solving the above system of linear equations in the unknowns A, B, and C, using Cramer's rule we get;

$$\Delta = \begin{vmatrix} 1 & \mu'_{1,g} & \mu'_{2,g} \\ \mu'_{1,g} & \mu'_{2,g} & \mu'_{3,g} \\ \mu'_{2,g} & \mu'_{3,g} & \mu'_{4,g} \end{vmatrix}$$
$$= \mu'_{2,g}\mu'_{4,g} + 2\mu'_{1,g}\mu'_{2,g}\mu'_{3,g} - \mu'^{2}_{1,g}\mu'_{4,g} - \mu'^{2}_{3,g} - \mu'^{3}_{2,g}$$

$$\Delta_A = \begin{vmatrix} 1 & \mu'_{1,g} & \mu'_{2,g} \\ \mu'_{1,f} & \mu'_{2,g} & \mu'_{3,g} \\ \mu'_{2,f} & \mu'_{3,g} & \mu'_{4,g} \end{vmatrix}$$
$$= \mu'_{1,f}(\mu'_{2,g}\mu'_{3,g} - \mu'_{1,g}\mu'_{4,g}) + \mu'_{2,f}(\mu'_{1,g}\mu'_{3,g} - \mu'^{2}_{2,g}) + \mu'_{2,g}\mu'_{4,g} - \mu'^{2}_{3,g}$$

$$\Delta_B = \begin{vmatrix} 1 & 1 & \mu'_{2,g} \\ \mu'_{1,g} & \mu'_{1,f} & \mu'_{3,g} \\ \mu'_{2,g} & \mu'_{2,f} & \mu'_{4,g} \end{vmatrix}$$
$$= \mu'_{1,f}(\mu'_{4,g} - \mu'^{2}_{2,g}) + \mu'_{2,f}(\mu'_{1,g}\mu'_{2,g} - \mu'_{3,g}) + \mu'_{2,g}\mu'_{3,g} - \mu'_{1,g}\mu'_{4,g}$$

$$\Delta_C = \begin{vmatrix} 1 & \mu'_{1,g} & 1 \\ \mu'_{1,g} & \mu'_{2,g} & \mu'_{1,f} \\ \mu'_{2,g} & \mu'_{3,g} & \mu'_{2,f} \end{vmatrix}$$
$$= \mu'_{1,f}(\mu'_{1,g}\mu'_{2,g} - \mu'_{3,g}) + \mu'_{2,f}(\mu'_{2,g} - \mu'^{2}_{1,g}) + \mu'_{1,g}\mu'_{3,g} - \mu'^{2}_{2,g}.$$

Hence, the values of A, B, and C obey equations (13) – (15). Normalizing the values of A, B, and C we obtain the values of w_1, w_2, and w_3 as in Eq. (15).

Eq. (13) expresses the PDF f(x) as a weighted mixture of the three PDFs g(x), $g_1(x)$, and $g_2(x)$ with respective weights w_1, w_2, and w_3. Note that, $w_1 + w_2 + w_3 = 1$. See Tavangar and Asadi (2010).

The minimum χ^2-divergence measure of the distance between f(x) and g(x) is:

$$\chi^2_{\min}(f, g) = A^2 + B^2 \mu'_{2,g} + C^2 \mu'_{4,g} + 2AB\, \mu'_{1,g} + 2AC\, \mu'_{2,g} + 2BC\, \mu'_{3,g} - 1$$

Here, A, B, and C are as given in equations (13) – (15).

We have the following corollary:

Corollary 3.4 When $\mu'_{1,f} = \mu'_{1,g}$, $\mu'_{2,f} = \mu'_{2,g}$, $w_1 = 1$, $w_2 = 0$, and $w_3 = 0$. Therefore, f(x) = g(x) and the minimum χ^2 – divergence measure between f(x) and g(x) equals zero.

3.4 Some Properties of the Minimum χ^2– divergence in this Case

Property 3.5 The rth non-central moment, mean, and variance of f(x) are:

$$\mu'_{r,f} = A\, \mu'_{r,g} + B\, \mu'_{r+1,g} + C\, \mu'_{r+2,g} ,$$

$$\mu'_{1,f} = A\, \mu'_{1,g} + B\, \mu'_{2,g} + C\, \mu'_{3,g} ,$$

$$\sigma^2_f = A\, \mu'_{2,g} + B\, \mu'_{3,g} + C\, \mu'_{4,g} - \mu'^2_{1,f} .$$

Property 3.6 The cumulative distribution function of f(x) has the form;

$$F(x) = w_1 G(x) + w_2 G_1(x) + w_3 G_2(x)$$

Here, G(x), G_1(x), and G_3(x) are respectively, the CDFs of g(x), g_1(x), and g_2(x) defined in Eq. (11).

Property 3.7 The Survival Function of f(x) is;

$$S(x) = w_1[1 - G(x)] + w_2[1 - G_1(x)] + w_3[1 - G_2(x)]$$

Property 3.8 The hazard function of f(x) is;

$$h(x) = \frac{w_1\, g(x) + w_2\, g_1(x) + w_3\, g_2(x)}{w_1[1-G(x)] + w_2[1-G_1(x)] + w_3[1-G_2(x)]}$$

4. Exponentiated Exponential Distribution

The Exponentiated Exponential (EExp) distribution is a member of the exponentiated weibull (EW) distribution. In Gupta and Kundu (2001), the two-parameter EExp distribution is quite effectively, used in analyzing lifetime's data, particularly in place of two-parameter gamma or two-parameter weibull distribution. If the shape parameter θ =1, then all the three distributions coincide with the one-parameter exponential distribution. Therefore, all the three distributions are generalizations of the exponential distribution in different ways. The PDF of the EExp distribution has different shapes.

It is unimodal for θ >1, and is reversed J-shaped for θ ≤ 1. It is log-convex if θ < 1 and log-concave if θ >1. The hazard function is non-decreasing when θ >1, non-increasing when θ < 1, and constant when θ =1.

4.1 The Available information is the n^{th} moment

For the case, when the prior is the EExp (θ, λ) distribution and the available information is $\mu'_{n,f}$, n^{th} non-central moment of f(x), we present the following theorem:

Theorem 4.1 Suppose that the observed estimated prior is the exponentiated exponential distribution with parameters θ and λ and the available information be;

$$g(x;\theta,\lambda)=\lambda\theta\exp(-\lambda x)\left(1-\exp(-\lambda x)\right)^{\theta-1}, x>0, \theta>0, \lambda>0 \qquad (16)$$

$$f(x)\geq 0, \int_0^\infty f(x)dx=1, E(X^n)=\int_0^\infty x^n f(x)dx=\mu'_{n,f}$$

Then the corresponding minimum χ^2 – divergence PDF f(x) is:

$$f(x;\theta,\lambda)=w_1(n)g(x;\theta,\lambda)+w_2(n)\,g_n(x;\theta,\lambda), \qquad (17)$$

$$\mu'_{n,f}=t\,\mu'_{n,g}+(1-t)\frac{\mu'_{2n,g}}{\mu'_{n,g}},\ 0\leq t\leq 1, \qquad (18)$$

$$w_1(n)=\frac{1}{\sigma^2_{n,g}}(\mu'_{2n,g}-\mu'_{n,f}\,\mu'_{n,g}),$$

$$w_2(n)=\frac{1}{\sigma^2_{n,g}}(\mu'_{n,f}-\mu'_{n,g})\mu'_{n,g},$$

$$\chi^2_{\min}(f,g)=\frac{\left(\mu'_{n,f}-\mu'_{n,g}\right)^2}{\mu'_{2n,g}-\mu'^2_{n,g}}=\frac{\left(\mu'_{n,f}-\mu'_{n,g}\right)^2}{\sigma^2_{n,g}},$$

For non-negative integral values of (θ – 1), see Nadarajah (2011).

$$\mu'_{k,g}=E(X^k)=\frac{\theta}{\lambda^k}\sum_{i=0}^{\theta-1}\frac{(-1)^i\,\Gamma(\theta)k!}{\Gamma(i+1)\Gamma(\theta-i)(i+1)^{k+1}},\ k=1,2,\ldots \tag{19}$$

$$g_n(x;\theta,\lambda)=\frac{1}{\mu'_{n,g}}x^n\lambda\theta\exp(-\lambda x)\left(1-\exp(-\lambda x)\right)^{\theta-1},x>0,\ \theta,\lambda>0 \tag{20}$$

Proof Eq. (19) implies:

$$\mu'_{k,g}=\int_0^\infty x^k\lambda\theta e^{-\lambda x}(1-e^{-\lambda x})^{\theta-1}dx$$

$$=\int_0^\infty \lambda\theta x^k e^{-\lambda x}\sum_{i=0}^{\theta-1}{}^{\theta-1}C_i\,(-e^{-\lambda x})^i\,dx$$

$$=\lambda\theta\sum_{i=0}^{\theta-1}(-1)^i\,{}^{\theta-1}C_i\int_0^\infty x^k e^{-\lambda x}e^{-\lambda ix}\,dx$$

$$=\lambda\theta\sum_{i=0}^{\theta-1}(-1)^i\cdot\frac{(\theta-1)!}{i!(\theta-i-1)!}\int_0^\infty x^k e^{-(i+1)\lambda x}\,dx$$

Let $(i+1)\lambda x=y$

$$\Rightarrow x=\frac{y}{(i+1)\lambda},\quad dx=\frac{dy}{(i+1)\lambda}$$

$$\Rightarrow \mu'_{k,g}=\lambda\theta\sum_{i=0}^{\theta-1}(-1)^i\frac{(\theta-1)!}{i!(\theta-i-1)!}\int_0^\infty\frac{y^k}{(i+1)^k\lambda^k}e^{-y}\frac{dy}{\lambda(i+1)}$$

$$=\lambda\theta\sum_{i=0}^{\theta-1}(-1)^i\cdot\frac{(\theta-1)!}{i!(\theta-i-1)!}\cdot\frac{1}{(i+1)^{k+1}\lambda^{k+1}}\cdot\Gamma(k+1)$$

$$=\frac{\theta}{\lambda^k}\sum_{i=0}^{\theta-1}(-1)^i\frac{\Gamma(\theta)k!}{\Gamma(i+1)\Gamma(\theta-1)(i+1)^{k+1}}$$

Otherwise, the sum runs to ∞. We have the following corollaries:

Corollary 4.1 When t = 1, equation (18) implies $\mu'_{n,f}=\mu'_{n,g}$, w_1 (n) = 1, w_2 (n) = 0, and f(x; θ, λ) = g (x; θ, λ) with zero minimum χ^2 – divergence measure.

Corollary 4.2 When t = 0, equation (18) implies $\mu'_{n,f} = \mu'_{2n,g} / \mu'_{n,g}$, w_1 (n) = 0 and w_2 (n) =1. Hence f(x; θ,λ) = g_n(x; θ, λ) defined in equation (20), with $\chi^2_{\min}(f,g) = \sigma^2_{n,g} \Big/ \mu'^2_{n,g}$.

Corollary 4.3 For 0< t <1, $\mu'_{n,f} = t\mu'_{n,g} + (1-t)\mu'_{2n,g} / \mu'_{n,g}$, and f(x) is the proper mixture described in equation (17), i.e. w_1 (n) ≠ 0, w_2 (n) ≠ 0 and sum to 1. For example, t = ½, $\mu'_{n,f} = \frac{1}{2}\left[\mu'_{n,g} + \mu'_{2n,g} / \mu'_{n,g}\right]$, w_1 (n) = ½, and w_2 (n) = ½.

$$f(x;\theta,\lambda) = \frac{1}{2} g(x;\theta,\lambda) + \frac{1}{2} g_n(x;\theta,\lambda),$$

$$\chi^2_{\min}(f,g) = \frac{\sigma^2_{n,g}}{4\mu'^2_{n,g}}.$$

4.2 Some Properties of the Minimum χ^2– divergence in this Case

Property 4.1 The r^{th} non-central moment, mean, and variance of f(x) are;

$$\mu'_{r,f} = A(n)\mu'_{r,g} + B(n)\mu'_{n+r,g},$$

$$\mu'_{1,f} = A(n)\mu'_{1,g} + B(n)\mu'_{n+1,g},$$

$$\sigma^2_f = A(n)\mu'_{2,g} + B(n)\mu'_{n+2,g} - \mu'^2_{1,f}.$$

Property 4.2 The cumulative distribution function of f(x) has the form:

$$F(x) = w_1(n)G(x) + w_2(n)G_n(x).$$

The functions G(x) and G_n (x) are the CDFs of g(x; θ, λ) and g_n(x; θ, λ) respectively. That is;

$$G(x) = (1-\exp(-\lambda x))^{\theta},$$

$$G_n(x) = \int_0^x \frac{1}{m_{n,g}} u^n \lambda\theta \exp(-\lambda u)(1-\exp(-\lambda u))^{\theta-1} du, x > 0$$

Property 4.3 The Survival Function of f(x) is;

$$S(x) = w_1(n)[1 - G(x)] + w_2(n)[1 - G_n(x)]$$

Property 4.4 The hazard function of f(x) is;

$$h(x) = \frac{w_1(n)g(x;\theta,\lambda) + w_2(n)g_n(x;\theta,\lambda)}{w_1(n)[1-G(x)] + w_2(n)[1-G_n(x)]}$$

4.3 Available Information is the First two Moments

When the prior distribution is the EExp distribution with parameters θ and λ, and the available information is $\mu'_{k,f}, k = 1,2$. We present the following theorem:

Theorem 4.2 Suppose that the observed estimated prior is the EExp with parameters θ, λ. Let the available information be:

$$E(X) = \int_0^\infty xf(x)dx = \mu'_{1,f}, \quad E(X^2) = \int_0^\infty x^2 f(x)dx = \mu'_{2,f},$$

$$f(x) \geq 0, \; and \; \int_0^\infty f(x)\,dx = 1.$$

The minimum χ^2 – divergence PDF f(x; θ, λ) is;

$$\text{f}(x;\ \theta,\ \lambda) = (A + Bx + Cx^2)g(x;\theta,\lambda).$$

$$\text{f}(x;\ \theta,\ \lambda) = w_1 g(x;\theta,\lambda) + w_2 g_1(x;\theta,\lambda) + w_3 g_2(x;\theta,\lambda), \tag{21}$$

$$A = \frac{\mu'_{1,f}(\mu'_{2,g}\mu'_{3,g} - \mu'_{1,g}\mu'_{4,g}) + \mu'_{2,f}(\mu'_{1,g}\mu'_{3,g} - \mu'^2_{2,g}) + \mu'_{2,g}\mu'_{4,g} - \mu'^2_{3,g}}{\mu'_{2,g}\mu'_{4,g} + 2\mu'_{1,g}\mu'_{2,g}\mu'_{3,g} - \mu'^2_{1,g}\mu'_{4,g} - \mu'^2_{3,g} - \mu'^3_{2,g}}, \tag{22}$$

$$B = \frac{\mu'_{1,f}(\mu'_{4,g} - \mu'^2_{2,g}) + \mu'_{2,f}(\mu'_{1,g}\mu'_{2,g} - \mu'_{3,g}) + \mu'_{2,g}\mu'_{3,g} - \mu'_{1,g}\mu'_{4,g}}{\mu'_{2,g}\mu'_{4,g} + 2\mu'_{1,g}\mu'_{2,g}\mu'_{3,g} - \mu'^2_{1,g}\mu'_{4,g} - \mu'^2_{3,g} - \mu'^3_{2,g}}, \tag{23}$$

$$C = \frac{\mu'_{1,f}(\mu'_{1,g}\mu'_{2,g} - \mu'_{3,g}) + \mu'_{2,f}(\mu'_{2,g} - \mu'^2_{1,g}) + \mu'_{1,g}\mu'_{3,g} - \mu'^2_{2,g}}{\mu'_{2,g}\mu'_{4,g} + 2\mu'_{1,g}\mu'_{2,g}\mu'_{3,g} - \mu'^2_{1,g}\mu'_{4,g} - \mu'^2_{3,g} - \mu'^3_{2,g}}, \tag{24}$$

$$w_1 = A,\ w_2 = B\,\mu'_{1,g},\ w_3 = C\,\mu'_{2,g}$$

Eq. (21) expresses the PDF f(x; θ, λ) as a weighted mixture of the three PDFs g(x; θ, λ), g_1(x; θ, λ), and g_2(x; θ, λ) with respective weights w_1, w_2, and w_3. Notice that, $w_1 + w_2 + w_3 = 1$,

$$g_i(x;\theta,\lambda)=\frac{1}{\mu'_{i,g}}x^i\lambda\theta\exp(-\lambda x)(1-\exp(-\lambda x))^{\theta-1}, x>0, i=1,2$$

The minimum χ^2-divergence measure between f(x; θ, λ) and g(x; θ, λ) is;

$$\chi^2_{\min}(f,g)=A^2+B^2\mu'_{2,g}+C^2\mu'_{4,g}+2AB\,\mu'_{1,g}+2AC\,\mu'_{2,g}+2BC\,\mu'_{3,g}-1$$

Where A, B, and C are as given in equations (21) – (24). We have the following corollary;

Corollary 4.4 When $\mu'_{k,f}, k=1,2$, and $\mu'_{k,f}, k=1,2$, $w_1 = 1$, $w_2 = 0$, and $w_3 = 0$. Therefore,

f(x; θ, λ) = g(x; θ, λ) and the minimum χ^2– divergence measure between f(x) and g(x; θ, λ) equals zero. Some Properties of the minimum χ^2–divergence PDF f(x; θ, λ), see Hamedani Ahsanullah (2011).

4.4 Some properties of the Minimum χ^2– divergence in this Case

Property 4.5 The r^{th} non-central moment, mean, and variance of f(x) are;

$$\mu'_{r,f}=A\,\mu'_{r,g}+B\,\mu'_{r+1,g}+C\,\mu'_{r+2,g}$$

$$\mu'_{1,f}=A\,\mu'_{1,g}+B\,\mu'_{2,g}+C\,\mu'_{3,g},$$

$$\sigma_f^2=A\,\mu'_{2,g}+B\,\mu'_{3,g}+C\,\mu'_{4,g}-\mu'^2_{1,f}.$$

Property 4.6 The cumulative distribution function of f(x; θ, λ) has the form

$$F(x)=w_1G(x)+w_2G_1(x)+w_3G_2(x).$$

Where G(x), $G_1(x)$, and $G_3(x)$ are respectively, the CDFs of g(x; θ, λ), $g_1(x; \theta, \lambda)$, and $g_2(x; \theta, \lambda)$ defined in Eq. (22).

Property 4.7 The Survival Function of f(x; θ, λ) is;

$$S(x)=w_1[1-G(x)]+w_2[1-G_1(x)]+w_3[1-G_2(x)].$$

Property 4.8 The hazard function of f(x; θ, λ) is;

$$h(x) = \frac{w_1\, g(x;\theta,\lambda) + w_2\, g_1(x;\theta,\lambda) + w_3\, g_2(x;\theta,\lambda)}{w_1[1-G(x)] + w_2[1-G_1(x)] + w_3[1-G_2(x)]} \ .$$

5. Numerical Application

In this section, some illustrative examples are included for initial values of the parameters, by using software "Mathcad version 15". We tabulated the results and the corresponding characteristics of the exponential and exponentiated exponential distribution are graphically, compared.

5.1 The First Two Moments are Available for Exponential Distribution

In this case;

$$g(x;\lambda) = \lambda \exp(-\lambda x), x > 0, \lambda > 0,$$

$$f(x) = (A + Bx + Cx^2) g(x).$$

Or, equivalently;

$$f(x) = w_1 g(x) + w_2 g_1(x) + w_3 g_2(x).$$

Using Equations (13) – (15) express A, B, C, w_1, w_2, and w_3, we have three possibilities. Table 1, displays the different cases of Prior Exponential distribution according to the values of the first two moments $\mu'_{1,f}$, $\mu'_{2,f}$, the parent distribution is determined. The PDF f(x) of the parent distribution is characterized as a single PDF in cases (1) and (2). In cases (3) – (6), it is determined as a proper weighted binary mixture of two PDFs, while in cases (7) – (9), f(x) is determined as a proper weighted mixture of three PDFs. The values in the fourth column, indicates how close are the different forms of f(x) and the prior estimated PDF g(x).

Table 1. Prior is Exponential Distribution, and Available is $\mu'_{k,f}, k = 1, 2$

Case	$\mu'_{1,f}$	$\mu'_{2,f}$	w_1	w_2	w_3	f(x)	$\chi^2_{\min}(f,g)$
1	0.333	0.222	1	0	0	$f_1(x) = g(x)$	0
2	0.667	0.667	0	1	0	$f_2(x) = g_1(x)$	1
3	0.5	0.444	0.5	0.5	0	$f_3(x) = 0.5g(x) + 0.5g_1(x)$	0.25
4	0.667	0.778	0.5	0	0.5	$f_4(x) = 0.5g(x) + 0.5g_2(x)$	1.25
5	0.556	0.519	0.333	0.667	0	$f_5(x) = 0.333g(x) + 0.667g_1(x)$	0.444
6	0.556	0.593	0.667	0	0.333	$f_6(x) = 0.667g(x) + 0.333g_2(x)$	0.556
7	0.667	0.741	0.333	0.333	0.333	$f_7(x) = 0.333g(x) + 0.333g_1(x) + 0.333g_2(x)$	1.111
8	0.533	0.533	0.6	0.2	0.2	$f_8(x) = 0.6g(x) + 0.2g_1(x) + 0.2g_2(x)$	0.4
9	0556	0.568	0.556	0.222	0.222	$f_9(x) = 0.556g(x) + 0.222g_1(x) + 0.222g_2(x)$	0.494

From the Figures (1) - (2), there is evidence that the new (minimum $\chi2$– divergence) distribution, is not the exponential distribution. However according to the minimum cross-entropy principle the new distribution remains the same as the given prior distribution. Thus, given prior information about the underlying distribution as exponential distribution.

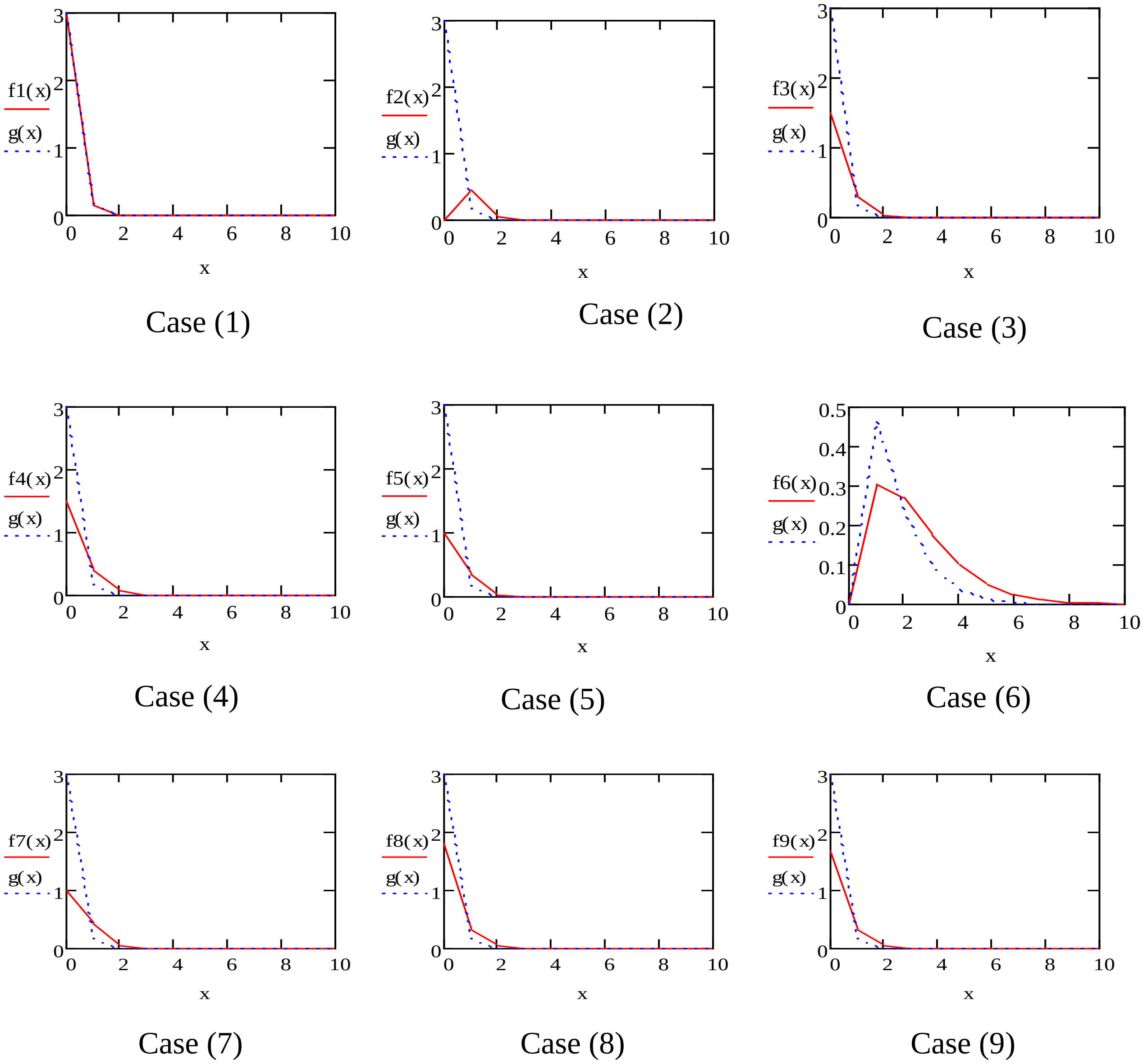
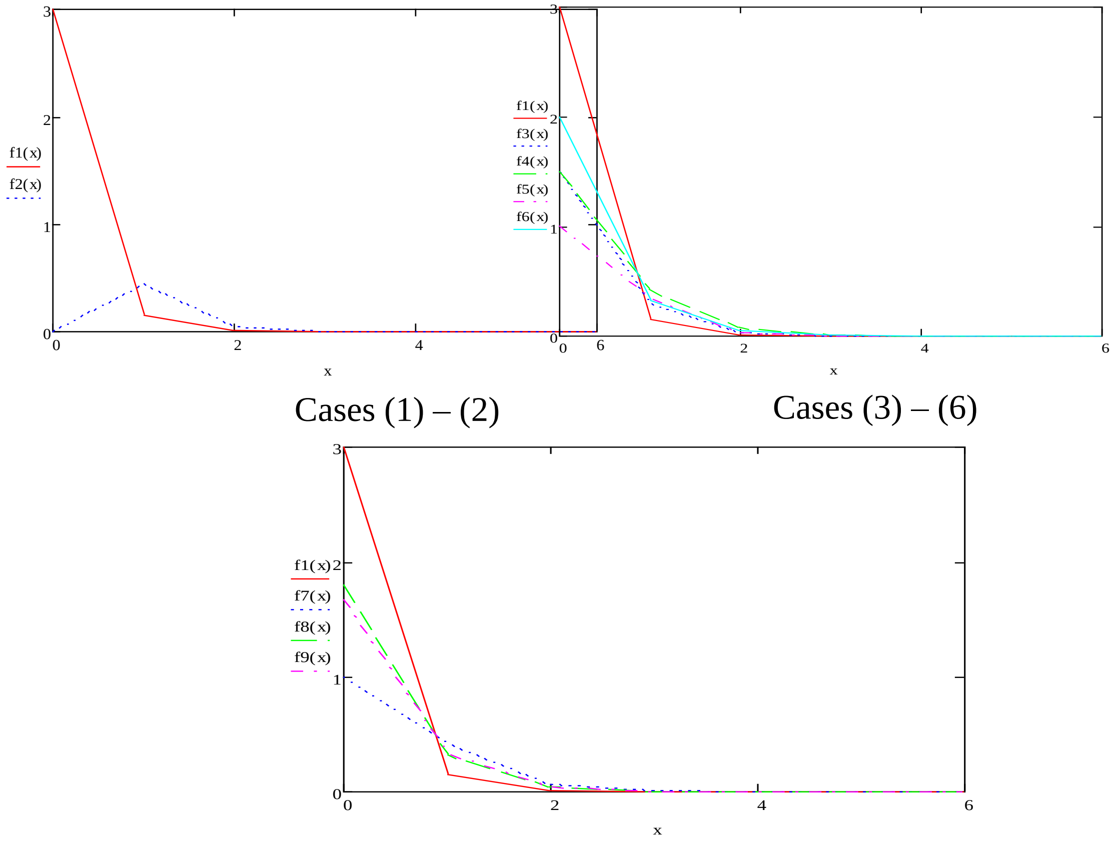


Figure 1. Represent PDFs with Prior Exponential Distribution in Cases (1) – (9)



Cases (1) – (2) Cases (3) – (6)

Case (7) – (9)

Figure 2. PDFs with Prior Exponential Distribution, When $\mu'_{k,f}\ , k = 1, 2$ in all Cases, Collectively.

5.2 The First Two Moments are Available for EExp Distribution

In this case

$$f(x) = A\, g(x) + B\, g_1(x) + C\, g_2(x)$$

Or, equivalently

$$f(x) = w_1\, g(x) + w_2\, g_1(x) + w_3\, g_2(x),$$

$$\mu'_{1,f} \in [1.5, 2.33],$$

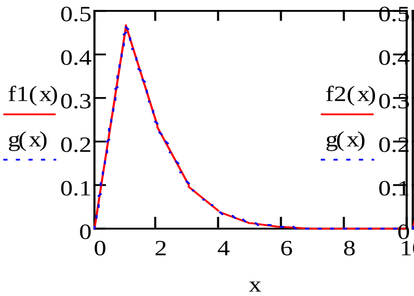
$$\mu'_{2,f} \in [3.5\,, 13.286],$$

Where A, B, C, w_1, w_2, and w_3 are expressed in equations (22) – (24).

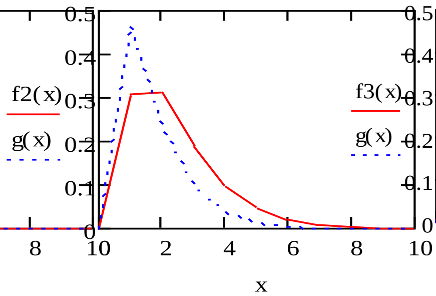
Table 2. Prior is EExp distribution and available is $\mu'_{k,f}, k=1,2$

Case	$\mu'_{1,f}$	$\mu'_{2,f}$	w_1	w_2	w_3	f(x)	$\chi^2_{\min}(f,g)$
1	1.5	3.5	1	0	0	$f_1(x) = g(x)$	0
2	2.333	7.5	0	1	0	$f_2(x) = g_1(x)$	0.556
3	1.917	5.5	0.5	0.5	0	$f_3(x) = 0.5g(x) + 0.5g_1(x)$	0.139
4	2.056	6.167	0.333	0.667	0	$f_4(x) = 0.333g(x) + 0.667g_1(x)$	0.247
5	2.071	6.762	0.667	0	0.333	$f_5(x) = 0.667g(x) + 0.333g_2(x)$	0.311
6	2.01	6.257	0.333	0.333	0.333	$f_6(x) = 0.333g(x) + 0.333g_1(x) + 0.333g_2(x)$	0.225
7	1.711	4.666	0.851	0.051	0.098	$f_7(x) = 0.851g(x) + 0.051g_1(x) + 0.098g_2(x)$	0.04
8	1.7	4.605	0.858	0.048	0.093	$f_8(x) = 0.858g(x) + 0.048g_1(x) + 0.093g_2(x)$	0.034

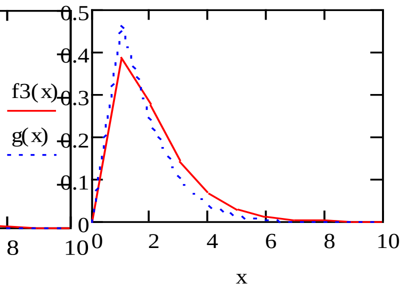
Table 2. Previous, displays the different cases of Prior is EExp distribution, According to the values of the first two moments $\mu'_{1,f}, \mu'_{2,f}$, the parent distribution is determined. The PDF f(x) of the parent distribution is characterized as a single PDF in cases (1) and (2). In cases (3) – (5), it is determined as a proper weighted binary mixture of two PDFs, while in cases (6) – (8), f(x) is determined as a proper weighted mixture of three PDFs. The values in the final column, indicates how close are the different forms of f(x) and the prior estimated PDF g(x). As can be seen from Figures (3) - (4).It is shown that by applying the minimum chi square divergence principle, new probability distributions are obtained. There is evidence that the new (minimum χ2– divergence) distribution, is not the EExp distribution.



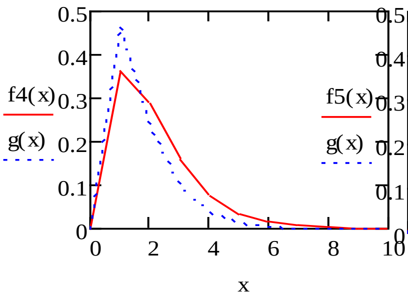
Case (1)



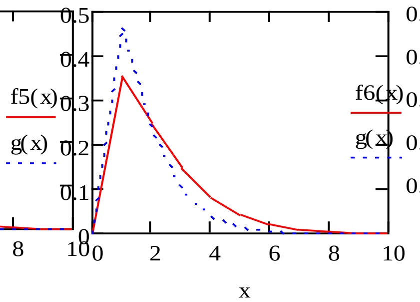
Case (2)



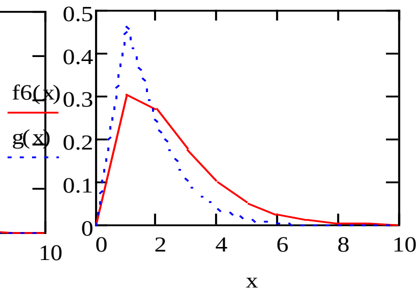
Case (3)



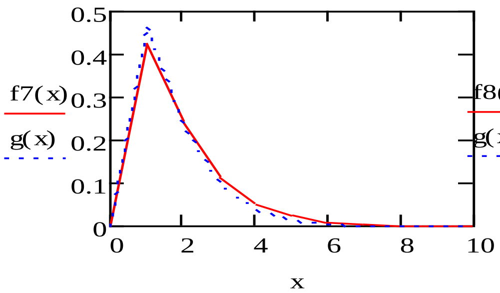
Case (4)



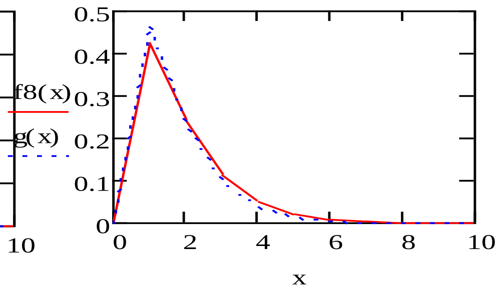
Case (5)



Case (6)

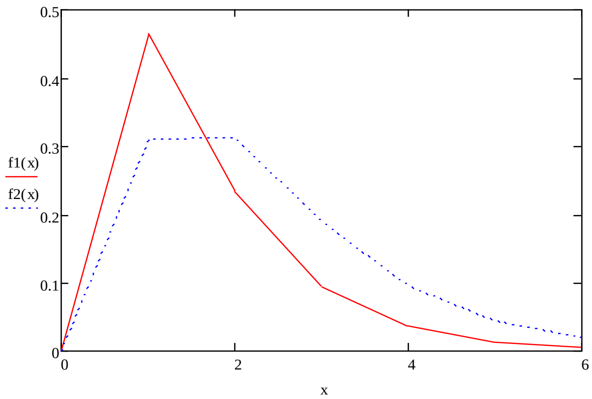


Case (7)

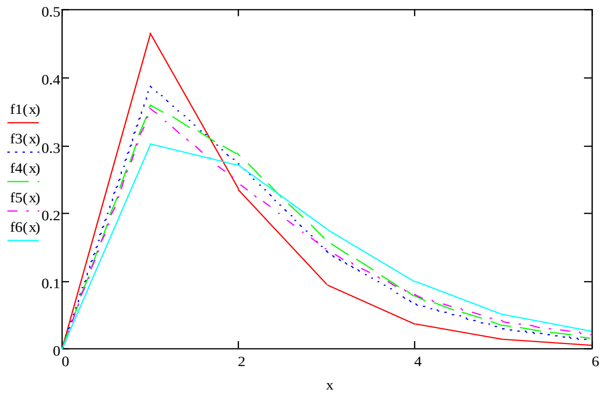


Case(8)

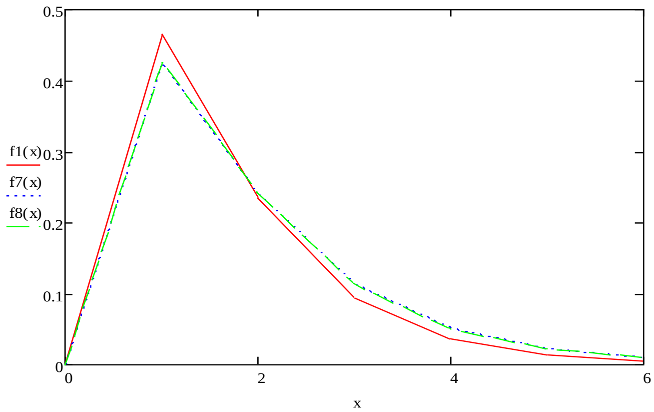
Figure 3. Represent PDFs with prior EExp in cases (1) – (8).



Cases (1) – (2)



Cases (3) – (6)



Cases (7) – (8)

Figure 4. PDFs with prior EExp, $\mu'_{k,f}$, k = 1, 2. is available in all cases

6. Conclusion

This paper discusses the characterization of some probability distributions conditioned on available a certain prior probability distribution and available information on moments. More precisely, we obtained characterizations when the prior is the exponential (λ) and EExp (θ, λ) with different available information on moments. Namely, the following available information on moments is considered for different cases, the first moment is available, the second moment is available and the first two moments are available. As the special case EExp (1, λ). The properties of the exponential (λ) and EExp (θ, λ) are determined, given a prior distribution as EExp and the new (current) information on moments, and from these tables and figures, there is evidence that the new (minimum χ^2– divergence) distribution, is not the exponential and EExp distribution. However according to the minimum cross-entropy principle the new distribution remains the same as the given prior distribution. Thus, given prior information about the underlying distribution as exponential and EExp distribution, in addition to the partial information in terms of the moments, the minimum χ^2– divergence principle provides a useful methodology for characterizing exponential and EExp probability distributions.

7. References

1. Ahsanullah, M., Hamedani, G. Exponential Distribution: Theory and Methods, *NOVA Science* 2010, New York.
2. Ahsanullah, M. (2017). *Characterizations of univariate continuous distributions*.Springer.
3. Balakrishnan, K. (2019). *Exponential distribution: theory, methods and applications*. Routledge.
4. Gupta, R. D., & Kundu, D. (2001). Exponentiated exponential family: an alternative to gamma and Weibull distributions. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *43*(1), 117-130.

5. Hamedani, G., & Ahsanullah, M. (2011). Characterizations of Weibull geometric distribution. *Journal of Statistical Theory and Applications*, *10*(4), 581-590.
6. Kawamura, T., & Iwase, K. (2003). Characterizations of the distributions of power inverse Gaussian and others based on the entropy maximization principle. *Journal of the Japan Statistical Society*, *33*(1), 95-104.
7. Kullback, S. (1997). *Information theory and statistics*. Courier Corporation. New York, John Wiley.
8. Kumar, P., & Taneja, I.(2004). Chi square divergence and minimization problem. *J. Comb. Inf. Syst. Sci*, *28*, 181-207.
9. Kumar, P. (2005). Characterizations of Beta Probability Distributions Based on the Minimum $\chi 2-$ Divergence Principle.
10. Kumar, P. (2006). Probability distributions conditioned by the available information: Gamma distribution and moments. *Computers & Mathematics with Applications*, *52*(3-4), 289-304.
11. Nadarajah, S. (2011). The exponentiated exponential distribution: a survey, *AStA Advances in Statistical Analysis*, 95, $219-251$.
12. Ord, J. K. (1975). Characterization Problems in Mathematical Statistics. *Journal of the Royal Statistical Society: Series A (General)*, *138*(4), 576-577.
13. Shore, J., & Johnson, R. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on information theory*, *26*(1), 26-37.
14. Tavangar, M., & Asadi, M. (2010). Some new characterization results on exponential and related distributions.
15. Yanushkevichius, R. (2014). Stability characterizations of some probability distributions. LAP LAMBERT Academic Publishing.
16. Zolotarev, V. M. (1976). Metric distances in spaces of random variables and their distributions. *Mathematics of the USSR-Sbornik*, *30*(3), 373.