# MINIMUM $\chi^2$ DISTANCE PROBABILITY DISTRIBUTIONS GIVEN EXPONENTIAL DISTRIBUTION AND MOMENTS

PRANESH KUMAR

Abstract. Given an exponential distribution $g(x)$ and the information in terms of moments of the random variable $X$, the probability distribution $f(x)$ is estimated such that the $\chi^2-$ distance between $f(x)$ and $g(x)$ is minimum. The expressions for the density function $f(x)$, minimum $\chi^2-$ measure and moments are derived. When the available information is in the form of observed frequency distribution as the exponential and on mean and/or variance of random variable, results on the minimum $\chi^2-$ distance distributions are discussed in detail. An application to the data from "Relation of serum retinol to acute phase proteins and malarial morbidity in Papua New Guinea" shows the usefulness of the results in practice.

## 1. Introduction

In repeated surveys or experiments, information on an observed probability distribution generally through sampling and sometimes on the mean and/or variance of some random variables is quite often available. Thus, the objective is to estimate the best probability distribution utilizing the available information. There are several estimation procedures like the maximum likelihood, the least squares which result in the best expected probability distribution. By maximizing the entropy subject to these constraints, we find the most random or most unbiased probability distribution [4-9]. However there are different opinions about which particular method to use [2]. Guiasu [3] has analyzed the weighted deviations subject to the given mean value of the random variable and determined the best prediction probability distribution by considering measures of deviations such as the Pearson's chi-square [11], Neyman's reduced chi-square [2], the Kullback-Leibler divergence [10], the Kolmogorov' index [1]. Kumar and Taneja [13] has considered the minimum chi-square divergence principle and the information available on moments of the random variables to determine the best probability distribution. The objective of this paper is to estimate the best probability distribution using the minimum chi-square divergence principle given observed probability distribution as an exponential distribution and the information on moments of the random variable. Section 2 briefly summarizes the minimum chi-square divergence principle and probability distributions. We derive in sections 3 through 6 the minimum chi-square distance probability distributions given exponential distribution as the observed probability distribution and information on the arithmetic mean, geometric mean, second moment and variance. An application to the data from "Relation of serum retinol to acute phase proteins and malarial morbidity in Papua New Guinea" is given in section 7. Section 8 provides the concluding remarks.

## 2. Minimum $\chi^2-$Divergence Principle and Probability Distributions

Let the random variable $X$ be a continuous variable with probability density function $f(x)$ defined over the open interval $(-\infty, +\infty)$ or the closed interval $[\theta, \phi]$. Then, the minimum cross-entropy principle of Kullback(1959) states:

2000 *Mathematics Subject Classification.* 94A17; 26D15.

*Key words and phrases.* Divergence measure, Minimum chi-square divergence principle, Characterization, Moments.

This research is partially supported by the Natural Sciences and Engineering Research Council's Discovery Grant to Pranesh Kumar.

*When a prior probability density function of $X$, $g(x)$, which estimates the underlying probability density function $f(x)$ 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 Kullback - Leibler divergence*

$$K(f,g)=\int f(x)\ln\frac{f(x)}{g(x)}dx. \tag{2.1}$$

Now, we state the *minimum $\chi^2-$divergence principle* as:

*When a prior probability density function of $X$, g(x), which estimates the underlying probability density function $f(x)$ 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 $\chi^2-$divergence*

$$\chi^2(f,g)=\int\frac{f^2(x)}{g(x)}dx-1. \tag{2.2}$$

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

**Definition 2.1.** *$f(x)$ is the probability density of the minimum $\chi^2-$divergence continuous probability distribution of random variable $X$ if it minimizes the $\chi^2-$divergence*

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

*given:*
*(i) a prior probability density function: $g(x)\geq 0,\int g(x)dx$=1,*
*(ii) probability density function constraints: $f(x)\geq 0$, $\int f(x)dx$=1,and*
*(iii) partial information in terms of averages: $\int x^t f(x)dx=m_{t,f}$, $t=1,2,3,..,r$.*

Now, given a prior probabilty distribution $g(x)$ as the exponential, we prove the following lemma. In what follows henceforth, integral $\int$ is over 0 to $\infty$ .

**Lemma 2.1.** *Let given be a prior exponential probability distribution of $X$ with the density function*

$$g(x)=\frac{e^{-x/a}}{a},\quad a>0,x>0, \tag{2.3}$$

*and the constraints*

$$f(x)\geq 0,\int f(x)dx=1,\ \int x^t f(x)dx=m_{t,f},\ t=1,2,3,...,r. \tag{2.4}$$

*Then, the minimum $\chi^2-$divergence probability distribution of $X$ has the probability density function*

$$f(x)=\frac{e^{-x/a}}{2a}\left(\alpha_0+\sum_{t=1}^{r}x^{\ t}\ \alpha_t\right), \tag{2.5}$$

*and the $(r+1)$ constants, $\alpha_0$ and $\alpha_t$, $t=1,2,3,...,r,$ are determined from*

$$\int\frac{e^{-x/a}}{2a}\left(\sum_{t=1}^{r}x^{\ t}\ \alpha_k\right)dx=1-\frac{\alpha_0}{2}, \tag{2.6}$$

*and*

$$\int\ \frac{x^t e^{-x/a}}{2a}\left(\sum_{t=1}^{r}x^{\ t}\ \alpha_t\right)dx=m_{t,f}-\frac{t!a^t\alpha_0}{2}. \tag{2.7}$$

*Further, the minimum $\chi^2-$divergence:*

$$\chi^2_{\min}(f,g) = \int \frac{e^{-x/a}}{4a}\left(\alpha_0 + \sum_{t=1}^{r} x^{\,t}\,\alpha_t\right)^2 dx - 1. \tag{2.8}$$

*Proof.* We apply the familiar method of finding extrema of a function by introducing Lagrangian multipliers, one for each constraint. Thus we minimize the Lagrangian function

$$L = \left(\int \frac{f^2(x)}{g(x)}\,dx - 1\right) - \alpha_0\left(\int f(x)\,dx - 1\right) - \sum_{t=1}^{r}\alpha_t\left(\int x^t\,f(x)\,dx - m_{t,f}\right)$$

$$= \left(\int a\,e^{\,x/a}\,f^2(x)\,dx - 1\right) - \alpha_0\left(\int f(x)\,dx - 1\right) - \sum_{t=1}^{r}\alpha_t\left(\int x^t\,f(x)\,dx - m_{t,f}\right). \tag{2.9}$$

Minimizing $L$ with respect to $f(x)$, i.e., differentiating the integrand with respect to $f(x)$ and setting the derivative equal to zero, we get the following equation

$$\frac{\partial L}{\partial f(x)} = 2ae^{\,x/a}\,f(x) - \alpha_0 - \sum_{t=1}^{r}\alpha_t x^t = 0. \tag{2.10}$$

Solving (2.10), we get the desired expression for $f(x)$ given by (2.5) which minimizes $\chi^2(f,g)$ since

$$\frac{\partial^2 L}{\partial f^2(x)} = 2ae^{\,x/a} > 0. \tag{2.11}$$

The $(r+1)$ constants, $\alpha_0$ and $\alpha_t, t = 1, ..., r$, are determined from the $n+r+1$ equations (2.6) and (2.7). The minimum $\chi^2$ divergence measure given by (2.8) follows on substituting $f(x)$ and $g(x)$ in (2.2).

## 3. Probability Distributions Given Arithmetic Mean

We apply lemma 2.1 to the exponential distributions when the information on arithmetic mean is given.

**Theorem 3.1.** *Let given be a prior exponential probability distribution of $X$ with the density function*

$$g(x) = \frac{1}{a}e^{-x/a}, \quad a > 0, x > 0, \tag{3.1}$$

*and the constraints*

$$f(x) \geq 0, \int f(x)dx = 1, \int xf(x)dx = b > 0\ ,\ . \tag{3.2}$$

*Then, the minimum $\chi^2-$divergence probability distribution of $X$ has the density function*

$$f(x) = \frac{2a-b}{a}\left(\frac{e^{-x/a}}{a}\right) + \frac{b-a}{a}\left(\frac{xe^{-x/a}}{a^2}\right), \tag{3.3}$$

*for $b \in [a, 2a]$.*

*The minimum $\chi^2-$divergence measure is*

$$\chi^2_{\min}(f,g) = \frac{(b-a)^2}{a^2}, \tag{3.4}$$

*and*

$$0 \leq \chi^2_{\min}(f,g) \leq 1. \tag{3.5}$$

This may be noted from (3.3) that the density function $f(x)$ is the weighted mixture of the $Gamma(1,a)$ and $Gamma(2,a)$ density functions with respective weights being $\frac{2a-b}{a}$ and $\frac{b-a}{a}$, i.e.,

$$X\tilde{\ } \left[\frac{2a-b}{a}\{X\tilde{\ }Gamma(1,a)\}\right] + \left[\frac{b-a}{a}\{X\tilde{\ }Gamma(2,a)\}\right], \tag{3.6}$$

where $X\tilde{\ }Gamma(p,q)$, $p,q>0$, is a *gamma* variable with density function

$$\begin{aligned} f(x) &= \frac{x^{p-1}e^{-x/q}}{q^p\Gamma(p)}, \quad p,q>0, x>0, \\ \Gamma(p) &= \int_0^\infty x^{p-1}e^{-x}dx. \end{aligned}$$

Properties of the *Minimum* $\chi^2-$*Divergence Probability Distribution* $f(x)$:

i) $t^{th}$ *Moment* of $X$ $(t=1,2,3,...)$:

$$m_{t,f} = \left[(2a-b) + (b-a)2^t\right]\frac{\Gamma(a+t)}{\Gamma(a+1)}. \tag{3.7}$$

*Mean* $(\mu_f)$ and *Variance* $(\sigma_f^2)$:

$$\begin{aligned} \mu_f &= b, \\ \sigma_f^2 &= (a+1)(3b-2a)-b^2, \end{aligned}$$

ii) *Probability Distribution Function*:

$$F(x) = 1-e^{-\frac{x}{a}}\left[\frac{(b-a)x}{a^2}+1\right]. \tag{3.8}$$

iii) *Survival Function:*

$$S(x) = e^{-\frac{x}{a}}\left[\frac{(b-a)x}{a^2}+1\right]. \tag{3.9}$$

iv) *Hazard Function*:

$$h(x) = \frac{a(2a-b)+(b-a)x}{a[a^2+(b-a)x]}. \tag{3.10}$$

v) *Mills Ratio*:

$$m(x) = \frac{a[a^2+(b-a)x]}{a(2a-b)+(b-a)x}. \tag{3.11}$$

vi) *Memoryless Property*: For $m,n>0$,

$$P(X>m+n|X>m) = \frac{(m+n)(b-a)+a^2}{m(b-a)+a^2}e^{-n/a} \neq P(X>n). \tag{3.12}$$

Thus, the *minimum* $\chi^2-$*divergence* probability distribution of $X$ does not have the *memoryless* property except the case when $b=a$. In this case

$$P(X>m+n|X>m) = e^{-n/a} = P(X>n), \tag{3.13}$$

and hence the *memoryless* property.

Following are some interesting corollaries:

**Corollary 3.1.** *For $b = a$, the density function $f(x) = g(x)$, that is, the minimum $\chi^2-$divergence probability distribution of $X$ has an exponential distribution with mean parameter $a$.*

**Corollary 3.2.** *For $b = 2a$, the probability distribution which minimizes the $\chi^2-$divergence is*

$$f(x) = \frac{xe^{-x/a}}{a^2}. \tag{3.14}$$

*that is, $X$ has a gamma distribution with parameters $(2, a)$.*

**Corollary 3.3.** *For $b \in [a, 2a]$, the probability distribution which minimizes the $\chi^2-$divergence between $f(x)$ and $g(x)$ is not the exponential. This distribution is as given in (3. 3).*

*For example, if $b = \frac{a+2a}{2} = \frac{3a}{2}$, then the probability distribution has the density function*

$$f(x) = \frac{(a + x)e^{-x/a}}{2a^2}. \tag{3.15}$$

*The mean ($\mu_f$) and variance ($\sigma_f^2$)*

$$\begin{aligned} \mu_f &= 1.5a^3, \\ \sigma_f^2 &= 0.25\ a^4\ \left(16 - 9a^2\right), \end{aligned}$$

*The minimum $\chi^2-$divergence*

$$\chi^2_{\min}(f, g) = 0.25\ .$$

The *minimum $\chi^2-$divergence* probability distributions when a *prior* distribution as *exponential* with *mean* $a = 0.5, 0.75, 1$ and new mean values as $b = a, 1.5a, 2a$, are shown in table 1 and figures 1.a - 1.c.

[INSERT TABLE 1]
[INSERT FIGURES 1.a - 1.c]

## 4. Probability Distributions Given $E(X^2)$

Suppose that *a prior* exponential density function $g(x)$ and the information on $E(X^2)$=c>0, is available. Then from lemma 2.1, we have:

**Theorem 4.1.** *Let given be a prior exponential probability distribution of $X$ with the density function*

$$g(x) = \frac{1}{a}e^{-x/a}, \quad a >, x > 0, \tag{4.1}$$

*and the constraints*

$$f(x) \geq 0, \int f(x)dx = 1, \int x^2 f(x)dx = c > 0\ . \tag{4.2}$$

*Then, the minimum $\chi^2-$divergence probability distribution of $X$ has the density function*

$$f(x) = \frac{12a^2 - c}{10a^2}\left(\frac{e^{-x/a}}{a}\right) + \frac{c - 2a^2}{10a^2}\left(\frac{x^2 e^{-x/a}}{2a^3}\right), \tag{4.3}$$

*for $c \in [2a^2, 12a^2]$.*

*The minimum $\chi^2-$divergence measure is*

$$\chi^2_{\min}(f, g) = \frac{(c - 2a^2)^2}{20a^4}, \tag{4.4}$$

*and*

$$0 \leq \chi^2_{\min}(f, g) \leq 5. \tag{4.5}$$

Thus, we note from (4.3) that the density function $f(x)$ is the weighted mixture of the $Gamma(1,a)$ and $Gamma(2,a)$ density functions with respective weights being $\frac{12a^2-c}{10a^2}$ and $\frac{c-2a^2}{10a^2}$, i.e.,

$$X\tilde{}\left[\frac{12a^2-c}{10a^2}\{X\tilde{}Gamma(1,a)\}\right]+\left[\frac{c-2a^2}{10a^2}\{X\tilde{}Gamma(3,a)\}\right], \tag{4.6}$$

Properties of the *Minimum* $\chi^2$*−Divergence Probability Distribution* $f(x)$:

i) $t^{th}$ *Moment* of $X$ $(t = 1, 2, 3, ...)$:

$$m_{t,f} = \frac{a^{t+2}}{10}\left[(12a^2-c)\Gamma(1+t)+(c-2a^2)\Gamma(3+t)\right]. \tag{4.7}$$

*Mean* $(\mu_f)$ and *Variance* $(\sigma_f^2)$:

$$\begin{aligned} \mu_f &= 0.5a^3c\ , (3.2.8) \\ \sigma_f^2 &= 0.05a^4\left(44c-48a^2-5c^2a^2\right), \end{aligned}$$

ii) *Probability Distribution Function*:

$$F(x) = 1-\frac{e^{-\frac{x}{a}}[\ 20a^4+x\,(x+2a)\left(c-2a^2\right)]}{20a^4}. \tag{4.8}$$

iii) *Survival Function:*

$$S(x) = \frac{e^{-\frac{x}{a}}[\ 20a^4+x\,(x+2a)\left(c-2a^2\right)]}{20a^4}. \tag{4.9}$$

iv) *Hazard Function*:

$$h(x) = \frac{2a^2(12a^2-c)+x^2(c-2a^2)}{a[\ 20a^4+x\,(x+2a)\,(c-2a^2)]}. \tag{4.10}$$

v) *Mills Ratio*:

$$m(x) = \frac{a[\ 20a^4+x\,(x+2a)\left(c-2a^2\right)}{2a^2(12a^2-c)+x^2(c-2a^2)}. \tag{4.11}$$

vi) *Memoryless Property*: For $m, n > 0$,

$$P(X > m+n|X > m) = \frac{20a^4+(m+n)\,(m+n+2a)\left(c-2a^2\right)}{20a^4+m\,(m+2a)\,(c-2a^2)}e^{-\frac{n}{a}} \neq P(X > n). \tag{4.12}$$

Thus, the *minimum* $\chi^2$*−divergence* probability distribution of $X$ does not have the *memoryless* property except the case when $c = 2a^2$. In this case,

$$P(X > m+n|X > m) = e^{-n/a} = P(X > n), \tag{4.13}$$

and hence the *memoryless* property.

Following are some interesting corollaries:

**Corollary 4.1.** *For* $c = 2a^2$*, the density function* $f(x) = g(x)$*, that is,* $X$ *has an exponential distribution with mean parameter* $a$*.*

**Corollary 4.2.** *For* $c = 12a^2$*, the probability distribution which minimizes the* $\chi^2$*−divergence is*

$$f(x) = \frac{x^2e^{-x/a}}{2a^3}, \tag{4.14}$$

*that is,* $X$ *has a gamma distribution with parameters* $(3, a)$*.*

**Corollary 4.3.** *For $c \in [2a^2, 12a^2]$, the probability distribution which minimizes the $\chi^2-$divergence between $f(x)$ and $g(x)$ is not the exponential. This distribution is as given in (4.3).*

*For an example, if $c = \frac{2a^2+12a^2}{2} = 7a^2$, then the probability distribution has the density function*

$$f(x) = \frac{(2a^2 + x^2)e^{-x/a}}{4a^3}. \tag{4.15}$$

*Mean ($\mu_f$) and Variance ($\sigma_f^2$):*

$$\begin{aligned} \mu_f &= 3.5a^5, \\ \sigma_f^2 &= 0.25a^6\left(52 - 49a^4\right), \end{aligned}$$

*The minimum $\chi^2-$divergence*

$$\chi^2_{\min}(f, g) = \frac{5}{4} \ .$$

The *minimum $\chi^2-$divergence* probability distributions when a *prior* distribution $g(x)$ is *exponential* with *mean* $a = 0.5, 0.75, 1$ and given the *new* information about $E(X^2)$,i.e., $c = 2a^2, 7a^2, 12a^2$ are, respectively, shown in table 2 and figures 2.a - 2.c.

[INSERT TABLE 2]
[INSERT FIGURES 2.a - 2.c]

## 5. Probability Distributions Given Geometric Mean

When *a prior* exponential probability density function $g(x)$ and the geometric mean of $X$ ,i.e., $\int \ln x f(x)dx = m_f$, are given, we get from lemma 2.1:

**Theorem 5.1.** *Let given be a prior exponential probability distribution of $X$ with the density function*

$$g(x) = \frac{1}{a}e^{-x/a}, \quad a > 0, x > 0, \tag{5.1}$$

*and the constraints*

$$f(x) \geq 0, \int f(x)dx = 1, \int \ln x f(x)dx = m_f > 0 \quad . \tag{5.2}$$

*Then, the minimum $\chi^2-$divergence probability distribution of $X$ has the density function*

$$f(x) = \frac{e^{-x/a}\left[(\sigma_g^2 + m_g^2 - m_f\ m_g) + (m_f - m_g)\ln x\right]}{a\sigma_g^2}, \tag{5.3}$$

*for*

$$m_f \in [m_g, \frac{\sigma_g^2 + m_g^2}{m_g}], \tag{5.4}$$

*where $m_g$ and $\sigma_g^2$ are, respectively, the mean and variance of $\ln X$ using $g(x)$.*

*The minimum $\chi^2-$divergence measure is*

$$\chi^2_{\min}(f, g) = \frac{(m_f - m_g)^2}{\sigma_g^2}, \tag{5.5}$$

*and*

$$0 \leq \chi^2_{\min}(f, g) \leq \frac{\sigma_g^2}{m_g^2} \ . \tag{5.6}$$

Properties of the *minimum* $\chi^2-$*divergence* probability distribution are presented below.

i) $t^{th}$ *Moment* of $X$ $\quad(t=1,2,3,...)$:

$$m_{t,f}=\frac{(\sigma_g^2+m_g^2-m_f\ m_g)\ \ t!a^t+(m_f-m_g)\ m_{\ x\ln x,t,g}}{\sigma_g^2}, \tag{5.7}$$

where

$$m_{\ x\ln x,t,g}=\int\frac{(x^t\ln x)e^{-x/a}}{a}dx. \tag{5.8}$$

*Mean* $(\mu_f)$ and *Variance* $(\sigma_f^2)$:

$$\begin{aligned}\mu_f &= \frac{(\sigma_g^2+m_g^2-m_f\ m_g)\ a+(m_f-m_g)\ m_{\ x\ln x,1,g}}{\sigma_g^2},\\ \sigma_f^2 &= \frac{2(\sigma_g^2+m_g^2-m_f\ m_g)\ a^2+(m_f-m_g)\ m_{x\ln x,2,g}-\sigma_g^2\ \mu_f^2}{\sigma_g^2}.\end{aligned}$$

ii) *Probability Distribution Function*: For $x>0$,

$$F(x) = \frac{\left[(\sigma_g^2+m_g^2-m_f\ m_g)\left(1-e^{-\frac{x}{a}}\right)+(m_f-m_g)\left[\ln a-0.5772157-(\ln x)\,e^{-\frac{x}{a}}-Ei(1,\frac{x}{a})\right]\right]}{\sigma_g^2},$$

where $Ei(1,\frac{x}{a})$ is the exponential integral.

iii) *Survival Function:* For $x>0$,

$$S(x) = \frac{\left[\sigma_g^2-(\sigma_g^2+m_g^2-m_f\ m_g)\left(1-e^{-\frac{x}{a}}\right)+(m_f-m_g)\left[\ln a-0.5772157-(\ln x)\,e^{-\frac{x}{a}}-Ei(1,\frac{x}{a})\right]\right]}{\sigma_g^2},$$

iv) *Hazard Function*: For $x>0$,

$$h(x) = \frac{e^{-x/a}\left[(\sigma_g^2+m_g^2-m_f\ m_g)+(m_f-m_g)\ln x\right]}{a\left[\sigma_g^2-(\sigma_g^2+m_g^2-m_f\ m_g)\left(1-e^{-\frac{x}{a}}\right)+(m_f-m_g)\left[\ln a-0.5772157-(\ln x)\,e^{-\frac{x}{a}}-Ei(1,\frac{x}{a})\right]\right]}.$$

v) *Mills Ratio*:

$$m(x) = \frac{a\left[\sigma_g^2-(\sigma_g^2+m_g^2-m_f\ m_g)\left(1-e^{-\frac{x}{a}}\right)+(m_f-m_g)\left[\ln a-0.5772157-(\ln x)\,e^{-\frac{x}{a}}-Ei(1,\frac{x}{a})\right]\right]}{e^{-x/a}\left[(\sigma_g^2+m_g^2-m_f\ m_g)+(m_f-m_g)\ln x\right]}.$$

Following are some interesting corollaries:

**Corollary 5.1.** *For $m_f=m_g$, the density function $f(x)=g(x)$, that is, $X$ has an exponential distribution with mean parameter $a$.*

**Corollary 5.2.** *For $m_f=\frac{\sigma_g^2+m_g^2}{m_g}$, the probability distribution which minimizes the $\chi^2-$divergence is*

$$f(x)=\frac{(\ln x)e^{-x/a}}{a\ m_g}. \tag{5.9}$$

*The mean ($\mu_f$) and variance ($\sigma_f^2$) are*

$$\mu_f = \frac{m_{x\ln x,1,g}}{m_g},$$

$$\sigma_f^2 = \frac{m_g\ m_{x\ln x,2,g} - m^2_{x\ln x,1,g}}{m_g^2}.$$

**Corollary 5.3.** *For $m_f \in [m_g, \frac{\sigma_g^2+m_g^2}{m_g}]$, the probability distribution which minimizes the $\chi^2-$divergence between $f(x)$ and $g(x)$ is not the exponential. This distribution is as given in (5.3).*

*For an example, if $m_f$ is in the center of the interval $[m_g, \frac{\sigma_g^2+m_g^2}{m_g}]$, then the probability distribution of $X$ has the density function*

$$f(x) = \frac{(m_g + \ln x)e^{-x/a}}{2am_g}. \tag{5.10}$$

*The mean ($\mu_f$) and variance ($\sigma_f^2$)*

$$\mu_f = \frac{m_g^2 + m_{x\ln x,1,g}}{2m_g},$$

$$\sigma_f^2 = \frac{2m_g(m_g m_{2,g} + m_{x\ln x,2,g}) - (m_g^2 + m_{x\ln x,1,g})^2}{4m_g^2}.$$

*The minimum $\chi^2-$divergence*

$$\chi^2_{\min}(f,g) = \left(\frac{\sigma_g}{2m_g}\right)^2.$$

## 6. Probability Distributions Given Arithmetic and Geometric Means

When *a prior* exponential probability density function $g(x)$ and the information on arithmetic and geometric means of $X$ ,i.e., $\int xf(x)dx = b$ and $\int \ln xf(x)dx = m_f$ are given, we get from lemma 2.1:

**Theorem 6.1.** *Let given be a prior exponential probability distribution of $X$ with the density function*

$$g(x) = \frac{1}{a}e^{-x/a}, \quad a > 0, x > 0, \tag{6.1}$$

*and the constraints*

$$f(x) \geq 0, \int f(x)dx = 1, \int xf(x)dx = b, \int \ln xf(x)dx = m_f. \tag{6.2}$$

*Then, the minimum $\chi^2-$divergence probability distribution of $X$ has the probability density function*

$$f(x) = g(x)\left(\alpha_0 + \alpha_1 x + \alpha_2 \ln x\right), \tag{6.3}$$

*where*

$$\alpha_0 = -\frac{a(2a-b)(\sigma_g^2+m_g^2) + a(m_{x\ln x,2,g} - 2am_{x\ln x,1,g})m_f + bm_{x\ln x,1,g}m_{x\ln x,2,g} - m^2_{x\ln x,2,g}}{-2am_{x\ln x,1,g}\ m_{x\ln x,2,g} + 2a^2m^2_{x\ln x,1,g} + m^2_{x\ln x,2,g} + a(1-2a)(\sigma_g^2+m_g^2)},$$

$$\alpha_1 = -\frac{(am_{x\ln x,1,g} - m_{x\ln x,2,g})m_f - bm^2_{x\ln x,1,g} + (b-a)(\sigma_g^2+m_g^2) + m_{x\ln x,1,g}m_{x\ln x,2,g}}{-2am_{x\ln x,1,g}\ m_{x\ln x,2,g} + 2a^2m^2_{x\ln x,1,g} + m^2_{x\ln x,2,g} + a(1-2a)(\sigma_g^2+m_g^2)},$$

$$\alpha_2 = \frac{a(2a-b)m_{x\ln x,1,g} + (b-a)m_{x\ln x,2,g} + a(1-2a)m_f}{-2am_{x\ln x,1,g}\ m_{x\ln x,2,g} + 2a^2m^2_{x\ln x,1,g} + m^2_{x\ln x,2,g} + a(1-2a)(\sigma_g^2+m_g^2)},$$

*$m_g$ and $\sigma_g^2$ respectively are the mean and variance of $\ln X$ using $g(x)$ and $m_{x\ln x,t,g}$ is defined in (5.7).*

*The minimum $\chi^2-$divergence measure is given by*

$$\chi^2_{\min}(f,g) = \left[(\alpha_0 + a\ \alpha_1)^2 + a^2\ \alpha_1^2 - 1\right] + \int\ \left(2\alpha_0\alpha_2 +\ 2\alpha_1\alpha_2 x +\ \alpha_2^2 \ln x\ \right) \frac{\ln x e^{-\frac{x}{a}}}{a} dx\ . \tag{6.4}$$

The $t^{th}$ moment $(t = 1, 2, 3, ...)$ about origin:

$$M_t(f) = [\alpha_0\ + (t+1)a\alpha_1]\ t!a^t\ + \alpha_2 m_{\ln x, t, g}. \tag{6.5}$$

*Mean* $(\mu_f)$ and *Variance* $(\sigma_f^2)$:

$$\begin{aligned} \mu_f &= (\alpha_0\ + 2a\alpha_1)\ a + \alpha_2 m_{\ln x, 1, g}, \\ \sigma_f^2 &= 2(\alpha_0\ + 3a\alpha_1)a^2\ + \alpha_2 m_{\ln x, 2, g} - \mu_f^2. \end{aligned}$$

## 7. Probability Distributions Given Arithmetic Mean and Variance of $X$

When *a prior* exponential probability density function $g(x)$ and the information on average $(m_{1,f})$ and variance $(\sigma_f^2)$ ,i.e., $\int x f(x) dx = b$ and $\int x^2 f(x) dx = b^2 + \sigma_f^2 = c$, are given, we use lemma 2.1 to get the following theorem:

**Theorem 7.1.** *Let given be a prior exponential probability distribution of $X$ with the density function*

$$g(x) = \frac{1}{a} e^{-x/a}, \quad a > 0, x > 0, \tag{7.1}$$

*and the constraints*

$$f(x) \geq 0, \int f(x) dx = 1,\ \int x f(x) dx = b\ , \int x^2 f(x) dx\ = m_{1,f}^2 + \sigma_f^2 = c. \tag{7.2}$$

*Then, the minimum $\chi^2-$divergence probability distribution of $X$ has the density function*

$$f(x) = \frac{\left(\alpha_0 + \alpha_1 x\ + \alpha_2 x^2\ \right) e^{-x/a}}{a}\ , \tag{7.3}$$

*where*

$$\alpha_0 = \frac{3a^2 c - c - 3a - 6a^3 b + 24a^4}{36a^4 - 1 - 3a - 12a^3}$$

$$\alpha_1 = -\frac{6a^3 + 2ac + b - 12a^2 b}{36a^4 - 1 - 3a - 12a^3}$$

$$\alpha_2 = \frac{-4a^2 b - 2a + 6a^3 - c + 2ac}{2a(36a^4 - 1 - 3a - 12a^3)}$$

*The minimum $\chi^2-$divergence measure is given by*

$$\chi^2_{\min}(f,g) = (\alpha_0^2 + a\alpha_1)^2 +\ a^2(\alpha_1^2 + 4\alpha_0\alpha_2) + 6a^3\alpha_2(2\alpha_1 + \alpha_2) - 1\ . \tag{7.4}$$

The $t^{th}$ moment $(t = 1, 2, 3, ...)$ about the origin:

$$M_t(f) = \left[\alpha_0 + (t+1)a\alpha_1\ + (t+2)a^2\alpha_2\right] t!a^t\ . \tag{7.5}$$

*Mean* $(\mu_f)$ and *Variance* $(\sigma_f^2)$:

$$\begin{aligned} \mu_f &= \frac{a\left[c\left(4a^2 - 2 - 3a\right) + 2a\left(-3 + 12a^2 b + 21a^3 - 2b - 3a\right)\right]}{2(36a^4 - 1 - 3a - 12a^3)}, \\ \sigma_f^2 &= \frac{2a^2\left[c\left(a^2 - 1 - 2a\right) + a\left(-3 + 22a^2 b + 18a^3 - 3b - 4a\right)\right]}{36a^4 - 1 - 3a - 12a^3} - \mu_f^2. \end{aligned}$$

## 8. An Application

The C-reactive protein (CRP) is a substance that can be measured in the blood. Values increase substantially within 6 hours of an infection and reach a peak within 24 to 48 hours after. In adults, chronically high values have been linked to an increased risk of cardiovascular disease. In a study of apparently healthy children aged 6 to 60 months in Papua New Guinea, CRP was measured in 90 children [12]. The units are milligrams per liter (mg/l). Data from a random sample of 40 of these children [14] are given in table 3. We consider finding the 95% confidence interval for the mean CRP.

INSERT TABLE 3

A preliminary examination of data suggest that the distribution is skewed and can be approximated by the exponential distribution. First we consider data (excluding 73.20) from 39 children and obtain the maximum likelihood estimate of the mean parameter as $a = 8.41$. Thus, we have

$$g(x) = \frac{e^{-x/8.41}}{8.41}, x > 0. \tag{8.1}$$

Now based on data in the table for all 40 children, we calculate new mean as $b = 10.28692$. Thus, using this information, one would like to use the model as

$$f^{\,*}(x) = \frac{e^{-x/10.28692}}{10.28692}, x > 0. \tag{8.2}$$

However, as proved in theorem 3.1, $f^{\,*}(x)$ is not the best estimated model in terms of the minimum chi square distance between $g(x)$ and $f^{\,*}(x)$. The minimum chi square distance distribution from (3.3) is

$$f(x) = (0.092369 + 0.0031554x)e^{-x/8.41}, x > 0. \tag{8.3}$$

The graphs for $g(x)$, $f^{\,*}(x)$ and $f(x)$ are shown in figure 3 and the 95% confidence interval for the mean CRP using $g(x)$, $f^{\,*}(x)$, $f(x)$ and normal distribution are presented in table 4.

INSERT FIGURE 3

INSERT TABLE 4

It may be interpreted from results in table 4 that:

(i) In case, there is no new information available on mean parameter and it is assumed that the population of children under study continues to have the same location parameter mean, the 95% confidence interval for the mean CRP will be (2.4194,31.023), width of the confidence interval being 28.604.

(ii) If the mean based on new data is observed to be 10.28692, the 95% confidence interval for the mean CRP using $f^{\,*}(x)$ is estimated to be (2.9594,37.947) and the width of the confidence interval is 34.938.

(iii) Using the best estimated model $f(x)$, the 95% confidence interval for the mean CRP is (3.0802,36.748) and the width of this confidence interval is 33.668.

(iv) The 95% confidence interval for the mean CRP using the approximation of the normal distribution is estimated as (5.153732,15.42011) and the width of the confidence interval is 10.266.

It may be remarked that the estimate of the confidence interval in (iv) should not be used because this is incorrect being based on the normal approximation while the distribution is actually skewed. The best estimated confidence interval is given in (iii) which is calculated using the minimum chi square distance distribution $f(x)$.

## 9. Concluding Remarks

The minimum cross entropy principle (MDIP) of Kullback and the maximum entropy principle (MEP) due to Jayne have been often used to characterize univariate and multivariate probability distributions. Minimizing cross entropy is equivalent to maximizing the likelihood function and the distribution produced by an application of Gauss principle is also the distribution which minimizes the cross entropy. Thus, given a *prior* information about the underlying distribution, in addition to the partial information in terms of the expected values, MDIP provides a useful methodology for characterizing probability distributions. 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 $\chi^2$-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.

## REFERENCES

[1] Campbell, L.L. Equivalence of Gauss's principle and minimum discrimination information estimation of probabilities. *Annals of Math. Statist.* ,1970,41,1,1011-1015.

[2] Gokhle, D.V. Maximum entropy characterization of some distributions. In *Statistical Distribution in Scientific Work*, Vol. III, Patil, Golz and Old (eds.), 1975, 292-304, A. Riedel, Boston.

[3] Iwase, K. and Hirano, K. Power inverse Gaussian distribution and its applications. *Japan. Jour. Applied Statist.*, 1990, 19, 163-176.

[4] Jaynes, E. T. Information theory and statistical mechanics. *Physical Reviews*, 1957,106,620-630.

[5] Kagan, A.M., Linnik, V. Ju and Rao, C. R. *Characterization Problems in Mathematical Statistics.*1975, New York: John Wiley.

[6] Kapur, J.N. Maximum entropy probability distributions of a continuous random variate over a finite interval. *Jour. Math. and Phys. Sciences*, 1982, 16,1,97-103.

[7] Kapur, J.N. Twenty-five years of maximum entropy principle. *Jour. Math. and Phys. Sciences*, 1983, 17,2,103-156.

[8] Kawamura, T. and Iwase, K. Characterizations of the distributions of power inverse Gaussian and others based on the entropy maximization principle. *Jour. Japan Statist. Soc.*, 2003, 33,1,95-104.

[9] Kesavan, H.K. and Kapur, J. The generalized maximum entropy principle. *IEEE Trans. Systems, Man and Cybernetics*, 1989, 19,5,1042-1052.

[10] Kullback, S. *Information Theory and Statistics.* 1959, New York: John Wiley.

[11] Kumar, Pranesh and Taneja, I.J. Chi square divergence and minimization problem. (Communicated),2004.

[12] Moore, D.S. and Mccabe, G.P. *Introduction to the Practice of Statistics.* 2004, Ex. 7.2, p.472, Freeman, NY.

[13] Pearson, K. On the Criterion that a given system of deviations from the probable in the case of correlated system of variables is such that it can be reasonable supposed to have arisen from random sampling, *Phil. Mag.*,1900, 50, 157-172.

[14] Rosales, F. et.al. Relation of serum retinol to acute phase protiens and malarial morbidity in papua New Guinea children. *American Jour. Clinical Nutrition*, 71, 2000, 1580-1588.

[15] Shore, J.E. and Johnson, R.W. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross entropy. *IEEE Trans. Information Theory*, 1980, IT-26,1,26-37.

[16] Taneja, I. J. *Generalized Information Measures and their Applications.* On line book: *http://www.mtm.ufsc.br/∼taneja/book/book.html*, 2001.

[17] Taneja, I.J. and Kumar, Pranesh. Relative Information of Type $s$, Csiszár $f-$Divergence, and Information Inequalities. *Information Sciences*,166, 105-125, 2004.

*E-mail address*: `kumarp@unbc.ca`
*URL*: `http://web.unbc.ca/∼kumarp`

Mathematics Department, College of Science and Management, University of Northern British Columbia, Prince George BC V2N4Z9, Canada.