Minimum χ^2 –Divergence Continuous Probability Distributions

Pranesh Kumar
Mathematics Department
University of Northern British Columbia
Prince George, BC V2N 5P8,Canada
(http://web.unbc.ca/~kumarp)
and
I. J. Taneja
Departamento de Matemática
Universidade Federal de Santa Catarina
88.040-900 Florianópolis, SC, Brasil
(http://www.mtm.ufsc.br/~taneja)

Abstract

The *minimum* χ^2 *–divergence principle* and its application to characterize the continuous probability distributions, given (i) a *prior* distribution and (ii) partial information in the form of average or (iii) partial information in the form of average and variance, are discussed. Uniform probability distributions are studied to illustrate the properties of the *minimum* χ^2 *–divergence probability distributions.*

Key Words: Divergence measure, Chi square divergence, Minimum cross-entropy principle, Minimum chi square divergence principle, Continuous probability distributions.

1. Introduction

The maximum entropy principle (MEP) due to Jayne (1957) and the minimum discrimination information principle (MDIP) or minimum cross-entropy principle of Kullback(1959) are well known to provide a methodology for identifying and characterizing the most unbiased univariate and multivariate probability distributions [Kagan et al.(1975), Kapur(1989), Kapur and Kesavan(1989;1992)]. Kapur(1982) used MEP and MDIP to characterize univariate distributions: *uniform, geometric, Gibb's, discrete normal, gamma, exponential, beta, Cauchy, Laplace, normal, lognormal, and Pareto* distributions. Kesavan and Kapur(1989) described generalizations of MEP and MDIP and presented a formalism, one in the MEP version and another in the MDIP version. Recently, Kawamura and Iwase(2003) applied MEP to characterize distributions of the power inverse Gaussian, power Birnaum-Saunders and generalized Gumbel. The equivalence of MDIP and statistical principles like, maximum likelihood principle and Gauss's principle, has been discussed by Campbell(1970) and Shore and Johnson(1980). Minimizing cross entropy is equivalent to maximizing the likelihood function [Kapur(1983)] and the distribution produced by an application of Gauss principle is also the distribution which minimizes the cross entropy. In literature on statistics, the χ^2 –divergence due to Pearson (1900) is well known. Kumar and Taneja(2004) have used the minimum χ^2 –divergence principle and have studied the discrete probability distributions. In this paper, we consider the minimum χ^2 –divergence principle and discuss a methodology to derive the continuous probability distributions using the minimum χ^2 –divergence principle when given is: (i) a *prior* distribution and (ii) partial information in the form of average or (iii) partial information in the form of average and variance. Uniform probability distributions are considered to elaborate the properties of the *minimum* χ^2 *–divergence probability distributions.*

2. Minimum χ^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 intreval $(-\infty,+\infty)$ or finite closed interval $[a,b]$, denoted by $\mathbb{I}$. In what follows henceforth, integral $\int$ is considered over $\mathbb{I}$. The minimum cross-entropy principle of Kullback(1959) is:

"*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 χ^2 –divergence*

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

We state the *minimum* χ^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 χ^2 –divergence*

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

The minimum cross-entropy principle and the *minimum* χ^2 *–divergence principle* applies to both the discrete and continuous random variables. We present the case of continuous random variable and define the *minimum* χ^2 *–divergence probability distribution 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*

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

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:* $\int x^t f(x)dx = m_{tf},\ t = 1,2,3,..,r.$

We present the main result in the following lemma:

Lemma *2.2. Given a prior probability density function $g(x)$ of the continuous random variable X, and the constraints*

$$f(x) \geq 0, \int f(x)dx = 1,\ \int x^t f(x)dx = m_{tf},\ t = 1,2,3,\ldots,r, \quad (2.3)$$

the minimum χ^2 –divergence probability distribution of X has the probability density function

$$f(x) = \frac{g(x)}{2}\left(\alpha_0 + \sum_{t=1}^{r} x^t \alpha_t\right), \quad (2.4)$$

and the $(r+1)$ constants, α_0 and $\alpha_t,\ t = 1,2,3,\ldots,r,$ are determined from

$$\int \frac{g(x)}{2}\left(\alpha_0 + \sum_{t=1}^{r} x^t \alpha_k\right)dx = 1, \quad (2.5)$$

and

$$\int \frac{x^t g(x)}{2}\left(\alpha_0 + \sum_{t=1}^{r} x^t \alpha_t\right)dx = m_{tf}, \quad (2.6)$$

where

$$m_{tf} = \int x^t f(x)dx. \quad (2.7)$$

(Please assign the equation numbers in the proof)

Expressions for the $(r+1)$ constants, α_0 and $\alpha_t, t = 1,2,3,\ldots,r,$ are obtained from the $n+r+1$ equations (2.5) and (2.6) .

Using the results of the lemma, the minimum χ^2 –divergence measure is given by:

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

3. Minimum χ^2 –Divergence Probability Distributions: Specific Cases

We present the minimum χ^2 –divergence probability distributions when *a prior* probability density function and the partial information in the form of arithmetic/geometric means and/or variance are given.

3.1. Given a *prior* Distribution and Partial Information in the Form of Arithmetic Mean of X

Suppose that *a prior* probability density function $g(x)$ and the partial information in the form of arithmetic mean, i.e., $\int xf(x)dx = m_{1,f}$ is available. Then we have from Lemma 2.2:

Theorem *3.1.1. Given a prior probability density function $g(x)$ of the continuous random variable X, and the constraints*

$$f(x) \geq 0, \int f(x)dx = 1, \quad \int xf(x)dx = m_{1,f}, (3.1.1)$$

the minimum χ^2 –divergence probability distribution of X has the probability density function

$$f(x) = g(x)\left[\frac{(m_{2,g} - m_{1,f}\, m_{1,g}) + x(m_{1,f} - m_{1,g})}{\sigma_g^2} \right], (3.1.2)$$

for $m_{1,f}$ between $[m_{1,g}, \frac{m_{2,g}}{m_{1,g}}]$,

where

$$m_{t,g} = \int x^t g(x)dx, \quad t = 1,2,3,\ldots, (3.1.3)$$

$$\sigma_g^2 = m_{2,g} - m_{1,g}^2. (3.1.4)$$

The t^{th} moment of X $(t = 1,2,3,\ldots)$ with density function $f(x)$ is

$$m_{t,f} = \frac{(m_{2,g} - m_{1,f}\, m_{1,g})\, m_{t,g} + (m_{1,f} - m_{1,g})m_{t+1,g}}{\sigma_g^2}. (3.1.4)$$

The minimum χ^2 –divergence measure is

$$\chi^2_{min}(f,g) = \frac{(m_{1,f} - m_{1,g})^2}{\sigma_g^2}, (3.1.5)$$

and

$$0 \leq \chi^2_{\min}(f,g) \leq \frac{\sigma_g^2}{m_{1,g}^2}. (3.1.6)$$

The mean (μ_f) and variance (σ_f^2) of the minimum χ^2 –divergence probability distribution are $\mu_f = m_{1,f}$ and

$$\sigma_f^2 = \frac{(m_{2,g} - m_{1,f}\, m_{1,g})m_{2,g} + (m_{1,f} - m_{1,g})\, m_{3,g} - \mu_f^2\, \sigma_g^2}{\sigma_g^2}. (3.1.7)$$

Following are some interesting cases.

Case *3.1.2. For $m_{1,f} = m_{1,g}$, the density functions $f(x) = g(x)$.*

Case *3.1.3. For $m_{1,f} = \frac{m_{2,g}}{m_{1,g}}$, the probability distribution which minimizes the χ^2 –divergence is*

$$f(x) = \frac{x}{m_{1,g}} g(x),\ x \in \mathbb{I}. (3.1.8)$$

In this case, the t^{th} moment is given by

$$m_{t,f} = \int \frac{x^{t+1}}{m_{1,g}} g(x)dx, \quad t = 1,2,3,\ldots, (3.1.9)$$

and the mean (μ_f) and variance (σ_f^2) are

$$\mu_f = \frac{m_{2,g}}{m_{1,g}},\ \sigma_f^2 = \frac{m_{1,g}\, m_{3,g} - m_{2,g}^2}{m_{1,g}^2}, m_{1,g} \neq 0. (3.1.10)$$

Example 3.1.4. Consider a *prior* uniform distribution of *X* defined over *[-1,3]*, i.e., $g(x) = \frac{1}{4},\ x \in [-1,3]$. Then, $m_{1,g} = 1$, $m_{2,g} = \frac{7}{3}, \frac{m_{2,g}}{m_{1,g}} = \frac{7}{3}$ and $\sigma_g^2 = \frac{4}{3}$. Suppose the partial information about average ($m_{1,f}$) is such that $1 \leq m_{1,f} \leq \frac{7}{3}$. Then, for $m_{1,f} = 1$, the probability distribution which minimizes the χ^2 –divergence is the *uniform* distribution $f(x) = \frac{1}{4},\ x \in [-1,3]$, which is given in Figure 1.a. However, for $1 < m_{1,f} \leq \frac{7}{3}$, the probability distribution which minimizes the χ^2 –divergence between $f(x)$ and $g(x)$ is *not* the *uniform* distribution. This distribution is

$$f(x) = \frac{(7 - 3m_{1,f}) + 3(m_{1,f} - 1)x}{16},\ x \in [-1,3], (3.1.11)$$

and the minimum χ^2 –divergence measure

$$\chi^2_{\min}(f,g) = \frac{3(m_{1,f} - 1)^2}{4}. (3.1.12)$$

As a particular case, the distribution by taking $m_{1,f} = \frac{1+\frac{7}{3}}{2} = \frac{5}{3}$ results in

$$f(x) = \frac{1+x}{8}, \ x \in [-1,3], (3.1.13)$$

$$\chi^2_{\min}(f,g) = \frac{1}{3}, (3.1.14)$$

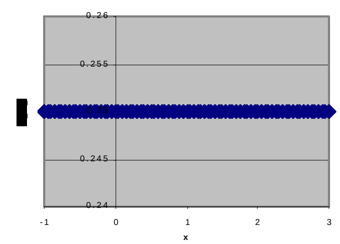

Figure 1.a. f(x) for X~U[-1,3], AM=1.

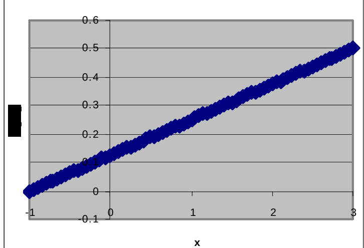

Figure 1.b. f(x) for X~U[-1,3], AM=5/3.

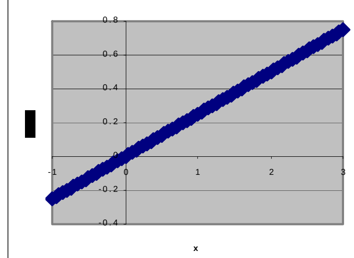

Figure 1.c. f(x) for X~U[-1,3], AM=7/3.

as shown in Figure 1.b. The distribution for $m_{1,f} = \frac{7}{3}$ with $\chi^2_{\min}(f,g) = \frac{4}{3}$ is given in Figure 1.c.

3.2. Given a *prior* Distribution and Partial Information in the Form of Geometric Mean of X

When *a prior* probability density function $g(x)$ and the partial information in the form of geometric mean of X ,i.e., $\int \ln x f(x)dx = m_{\ln x,1,f}$, are given, we get from Lemma 2.2:

Theorem *Given a prior probability density function $g(x)$ of the continuous random variable X, and the constraints*

$$f(x) \geq 0, \int f(x)dx = 1, \ \int \ln x f(x)dx = m_{\ln x,1,f}, (3.2.1)$$

the minimum χ^2 –divergence probability distribution of X is

$$f(x) = g(x)\left[\frac{(m_{\ln x,2,g} - m_{\ln x,1,f}\, m_{\ln x,1,g}) + (m_{\ln x,1,f} - m_{\ln x,1,g}) \ln x}{\sigma^2_{\ln x,g}}\right], (3.2.2)$$

for $m_{\ln x,1,f}$ between [$m_{\ln x,1,g}, \frac{m_{\ln x,2,g}}{m_{\ln x,1,g}}$],

where for $t = 1,2,3,\ldots,$

$$m_{\ln x,t,g} = \int (\ln x)^t g(x)dx, (3.2.3)$$

and

$$\sigma^2_{\ln x,g} = m_{\ln x,2,g} - m^2_{\ln x,1,g}. (3.2.4)$$

The t^{th} moment $(t = 1,2,3,..)$ of X with density function $f(x)$ is

$$m_{t,f} = \frac{(m_{\ln x,2,g} - m_{\ln x,1,f}\, m_{\ln x,1,g})\, m_{t,g} + (m_{\ln x,1,f} - m_{\ln x,1,g})\, m_{x\ln x,t,g}}{\sigma^2_{\ln x,g}}, (3.2.5)$$

where

$$m_{x\ln x,t,g} = \int (x^t \ln x) g(x) dx. (3.2.6)$$

The minimum χ^2 –divergence measure is

$$\chi^2_{min}(f,g) = \frac{(m_{\ln x,1,f} - m_{\ln x,1,g})^2}{\sigma^2_{\ln x,g}}, (3.2.7)$$

and

$$0 \leq \chi^2_{\min}(f,g) \leq \frac{\sigma^2_{\ln x,g}}{m^2_{\ln x,1,g}}. (3.2.8)$$

The mean (μ_f) and variance (σ^2_f) of the minimum χ^2 –divergence probability distribution are:

$$\mu_f = \frac{(m_{\ln x,2,g} - m_{\ln x,1,f}\, m_{\ln x,1,g})\, m_{1,g} + (m_{\ln x,1,f} - m_{\ln x,1,g})\, m_{x\ln x,1,g}}{\sigma^2_{\ln x,g}}, (3.2.9)$$

and

$$\sigma^2_f = \frac{(m_{\ln x,2,g} - m_{\ln x,1,f}\, m_{\ln x,1,g})\, m_{2,g} + (m_{\ln x,1,f} - m_{\ln x,1,g})\, m_{x\ln x,2,g} - \sigma^2_{\ln x,g}\, \mu^2_f}{\sigma^2_{\ln x,g}}. (3.2.10)$$

We have the following interesting cases:

Case *3.2.2. For $m_{\ln x,1,f} = m_{\ln x,1,g}$, the density functions $f(x) = g(x)$.*

Case *3.2.3. For $m_{\ln x,1,f} = \frac{m_{\ln x,2,g}}{m_{\ln x,1,g}}$, the probability distribution which minimizes the χ^2 –divergence is*

$$f(x) = \frac{\ln x}{m_{\ln x,1,g}} g(x),\ x \in \mathbb{I}. (3.2.11)$$

In this case, the t^{th} moment is given by

$$m_{t,f} = \int \frac{x^t \ln x}{m_{\ln x,1,g}} g(x) dx, \quad t = 1,2,3,\ldots, (3.2.12)$$

and the mean (μ_f) and variance (σ^2_f) are

$$\mu_f = \frac{m_{x\ln x,1,g}}{m_{\ln x,1,g}},\ \sigma^2_f = \frac{m_{\ln x,1,g}\, m_{x\ln x,2,g} - m^2_{x\ln x,1,g}}{m^2_{\ln x,1,g}}. (3.2.13)$$

Example 3.2.4. Consider a *prior* uniform distribution of X defined over *[2,12]*, i.e., $g(x) = \frac{1}{10}$, $x \in [2,12]$. Then $m_{\ln x,1,g} = 1.8433$, $m_{\ln x,2,g} = 3.6271$, $\frac{m_{\ln x,2,g}}{m_{\ln x,1,g}} = 1.9677$ and $\sigma^2_{\ln x,g} = 0.22935$. Suppose the partial information about average $(m_{\ln x,1,f})$ is such that $1.8433 \leq m_{\ln x,1,f} \leq 1.9677$. Then, for $m_{\ln x,1,f} = 1.8433$, the probability distribution which minimizes the χ^2 –divergence is the *uniform* distribution $f(x) = \frac{1}{10}$, $x \in [2,12]$, which is shown in Figure 1.a. However, for $1.8433 < m_{\ln x,1,f} \leq 1.9677$, the probability distribution which minimizes the χ^2 –divergence between $f(x)$ and $g(x)$ is *not* the *uniform* distribution. This distribution is

$$f(x) = (15.815 - 8.0371 m_{\ln x,1,f}) + 4.3601(m_{\ln x,1,f} - 1.8433)\ln x \,, x \in [2,12]. \quad (3.2.14)$$

and the minimum χ^2 –divergence measure

$$\chi^2_{\min}(f,g) = \frac{(m_{\ln x,1,f} - 1.8433)^2}{0.22935} \,. \quad (3.2.15)$$

As a particular case, the probability distribution $f(x)$, for $m_{\ln x,1,f} = \frac{1.8433+1.9677}{2} =1.9055$, is

$$f(x) = 0.50007 + 0.2712\ln x \,, x \in [2,12], \quad (3.2.16)$$

$$\chi^2_{\min}(f,g) = 0.016869, \quad (3.2.17)$$

and is shown in Figure 1.b.

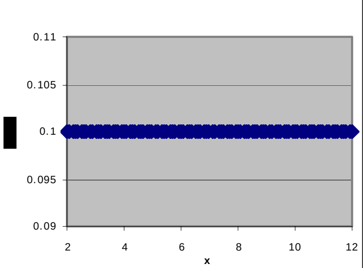

Figure 2.a. f(x) for X~U[2,12],GM=1.8433.

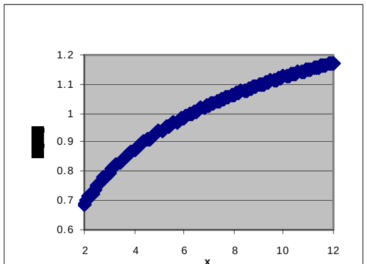

Figure 2.b. f(x) for X~U[2,12], GM=1.9055.

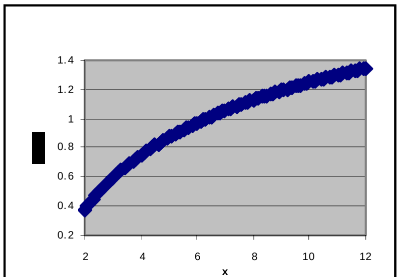

Figure 2.c. f(x) for X~U[2,12], GM=1.9677.

The uniform probability distribution for $m_{\ln x,1,f} = 1.9677$ with $\chi^2_{\min}(f,g) = 0.067475$ is given in Figure 2.c.

3.3. Given a *prior* Distribution and Partial Information in the Form of Arithmetic and Geometric Means

When *a prior* probability density function $g(x)$ and the partial information in the form of arithmetic and geometric means of X and ,i.e., $\int xf(x)dx = m_{1,f}$ and $\int \ln x f(x)dx = m_{\ln x,1,f}$ are given, we get from Lemma 2.2:

Theorem *3.3.1. Given a prior probability density function $g(x)$ of the continuous random variable X, and the constraints*

$$f(x) \geq 0, \int f(x)dx = 1, \int xf(x)dx = m_{1,f}, \int \ln x f(x)dx = m_{\ln x,1,f}, (3.3.1)$$

the minimum χ^2 –divergence probability distribution of X has the probability density function

$$f(x) = g(x)\left(\alpha_0 + \alpha_1 x + \alpha_2 \ln x\right), (3.3.2)$$

where

$$\alpha_0 = -\frac{-m_{1,g}m_{1,f}m_{(\ln x)^2,1,g}+m_{1,g}m_{\ln x,2,g}m_{\ln x,1,f}+m_{\ln x,1,g}m_{\ln x,2,g}m_{1,f}-m_{\ln x,1,g}m_{\ln x,1,f}m_{2,g}-m^2_{\ln x,2,g}+m_{(\ln x)^2,1,g}m_{2,g}}{-2m_{\ln x,1,g}m_{1,g}m_{\ln x,2,g}+m^2_{\ln x,1,g}m_{2,g}+m^2_{\ln x,2,g}+m_{(\ln x)^2,1,g}m^2_{1,g}-m_{(\ln x)^2,1,g}m_{2,g}}$$

(3.3.3)

$$\alpha_1 = -\frac{m_{\ln x,1,g}m_{1,g}m_{\ln x,1,f}-m_{1,g}m_{(\ln x)^2,1,g}-m_{1,f}m^2_{\ln x,1,g}+m_{1,f}m_{(\ln x)^2,1,g}+m_{\ln x,1,g}m_{\ln x,2,g}-m_{\ln x,2,g}m_{\ln x,1,f}}{-2m_{\ln x,1,g}m_{1,g}m_{\ln x,2,g}+m^2_{\ln x,1,g}m_{2,g}+m^2_{\ln x,2,g}+m_{(\ln x)^2,1,g}m^2_{1,g}-m_{(\ln x)^2,1,g}m_{2,g}}, (3.3.4$$

$$\alpha_2 = \frac{-m_{\ln x,1,g}m_{1,g}m_{1,f}+m_{\ln x,1,g}m_{2,g}-m_{\ln x,2,g}m_{1,g}+m_{\ln x,2,g}m_{1,f}+m_{\ln x,1,f}m^2_{1,g}-m_{\ln x,1,f}m_{2,g}}{-2m_{\ln x,1,g}m_{1,g}m_{\ln x,2,g}+m^2_{\ln x,1,g}m_{2,g}+m^2_{\ln x,2,g}+m_{(\ln x)^2,1,g}m^2_{1,g}-m_{(\ln x)^2,1,g}m_{2,g}}, (3.3.5)$$

and for $t = 1,2,3,\ldots,$

$$m_{\ln x,t,g} = \int x^{t-1}\ln x\, g(x)dx, (3.3.6)$$

$$m_{(\ln x)^2,1,g} = \int (\ln x)^2 g(x)dx. (3.3.7)$$

The t^{th} moment ($t = 1,2,3,\ldots$) about origin of the minimum χ^2 –divergence continuous probability distribution with probability density function $f(x)$ is given by

$$M_t(f) = \alpha_0 m_{t,g} + \alpha_1 m_{t+1,g} + \alpha_2 m_{\ln x,t+1,g}. (3.3.8)$$

The minimum χ^2 –divergence measure is given by

$$\chi^2_{min}(f,g) = \int g(x)\left(\alpha_0 + \alpha_1 x + \alpha_2 \ln x\right)^2 dx - 1. (3.3.9)$$

It may be noted if $m_{1,f} = m_{1,g}$ and $m_{\ln x,1,f} = m_{\ln x,1,g}$, then $f(x) = g(x),\ x \in \mathbb{I}$.

Example 3.3.2. Consider a *prior* uniform distribution of X defined over *[3,8]*, i.e., $g(x) = \frac{1}{5},\ x \in [3,8]$. Then, $m_{1,g} = \frac{11}{2}$, $m_{2,g} = \frac{97}{3}$, $m_{\ln x,1,g} = 1.6679$, $m_{\ln x,2,g} = 9.5697$, $m_{(\ln x)^2,1,g} = 2.8585$. In case $m_{1,f} = m_{1,g} = \frac{11}{2}$ and $m_{\ln x,1,f} = m_{\ln x,1,g} = 1.6679$, then $f(x) = g(x) = \frac{1}{5},\ x \in [3,8]$, which is given in Figure 3.a.

Let a *priori* information available be: $m_{1,f} = 5.6$ and $m_{\ln x,1,f} = 1.6679$. Thus, $\alpha_0 = 10.252$, $\alpha_1 = 2.9588$ and $\alpha_3 = -15.304$ and the minimum χ^2 –divergence probability distribution of X is

$$f(x) = \frac{1}{5}\left(10.252 + 2.9588x - 15.304\ln x\right), x \in [3,8], (3.3.10)$$

which is shown in Figure 3.b.

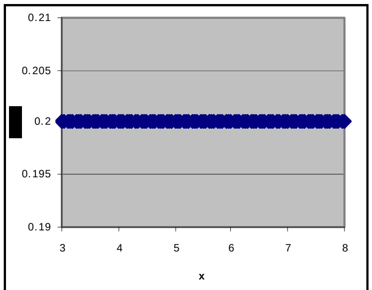

Figure 3.a. f(x) for X~U[3,8], AM=5.5, GM=1.6679.

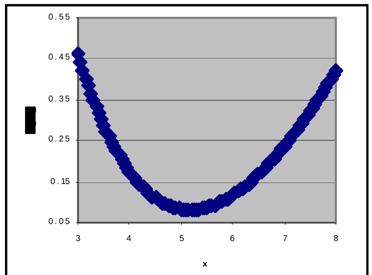

Figure 3.b. f(x) for X~U[3,8], AM=5.6, GM=1.6679.

3.4. Given a *prior* Distribution and Partial Information in the Form of Arithmetic Mean and Variance of X

When *a prior* probability density function $g(x)$ and the partial information in the form of average $(m_{1,f})$ and variance (σ_f^2) ,i.e., $\int xf(x)dx = m_{1f}$ and $\int x^2 f(x)dx = m_{2f} = m_{1,f}^2 + \sigma_f^2$, are given, we get from Lemma 2.2:

Theorem *3.4.1. Given a prior probability density function $g(x)$ of the continuous random variable X, and the constraints*

$$f(x) \geq 0, \int f(x)dx = 1,\ \int xf(x)dx = m_{1,f}\ , \int x^2 f(x)dx = m_{2,f} = m_{1,f}^2 + \sigma_f^2, (3.4.1)$$

the minimum χ^2 –divergence probability distribution of X has the probability density function

$$f(x) = g(x)\left(\alpha_0 + \alpha_1 x + \alpha_2 x^2\right), x \in \mathbb{I}, (3.4.2)$$

where

$$\alpha_0 = \frac{m_{1,g}m_{1,f}m_{4,g} - m_{1,g}m_{3,g}m_{2,f} - m_{2,g}m_{3,g}m_{1,f} + m_{2,g}^2 m_{2,f} + m_{3,g}^2 - m_{4,g}m_{2,g}}{-2m_{1,g}m_{2,g}m_{3,g} + m_{2,g}^3 + m_{3,g}^2 + m_{4,g}m_{1,g}^2 - m_{4,g}m_{2,g}}, (3.4.3)$$

$$\alpha_1 = \frac{m_{1,g}m_{4,g} - m_{1,g}m_{2,f}m_{2,g} + m_{2,g}^2 m_{1,f} - m_{1,f}m_{4,g} - m_{2,g}m_{3,g} + m_{3,g}m_{2,f}}{-2m_{1,g}m_{2,g}m_{3,g} + m_{2,g}^3 + m_{3,g}^2 + m_{4,g}m_{1,g}^2 - m_{4,g}m_{2,g}}, (3.4.4)$$

$$\alpha_2 = \frac{-m_{1,g}m_{2,g}m_{1,f} + m_{2,g}^2 - m_{3,g}m_{1,g} + m_{3,g}m_{1,f} + m_{2,f}m_{1,g}^2 - m_{2,f}m_{2,g}}{-2m_{1,g}m_{2,g}m_{3,g} + m_{2,g}^3 + m_{3,g}^2 + m_{4,g}m_{1,g}^2 - m_{4,g}m_{2,g}}. (3.4.5)$$

The t^{th} moment $(t = 1,2,3,\ldots)$ about origin of X is

$$m_{t,f} = \alpha_0 m_{t,g} + \alpha_1 m_{t+1,g} + \alpha_2 m_{t+2,g}\ . (3.4.6)$$

The minimum χ^2 –divergence measure is given by

$$\chi_{min}^2(f,g) = \int g(x)\left(\alpha_0 + \alpha_1 x + \alpha_2 x^2\right)^2 dx - 1. (3.4.7)$$

Example 3.4.2. Consider a *prior* uniform distribution of *X* defined over *[0,1]*, i.e., $g(x) = 1,\ x \in [0,1]$. Then, $m_{1,g} = \frac{1}{2}$, $m_{2,g} = \frac{1}{3}$, $m_{3,g} = \frac{1}{4}$ and $m_{4,g} = \frac{1}{5}$.

In case $m_{1,f} = m_{1,g} = \frac{1}{2}$ and $m_{2,f} = m_{2,g} = \frac{1}{3}$, then $f(x) = g(x),\ x \in [0,1]$, which is given in Figure 4.a. Let a *priori* information available be: $m_{1,f} = \frac{3}{5}$ and $\sigma_f^2 = \frac{1}{10}$, i.e., $m_{2,f} = \frac{23}{50}$. Thus, $\alpha_0 = \frac{6}{5}$, $\alpha_1 = -\frac{18}{5}$ and $\alpha_3 = \frac{24}{5}$ and the minimum χ^2 –divergence probability distribution of *X* is

$$f(x) = \frac{1}{5}\left(6 - 18x + 24x^2\right), x \in [0,1], (3.4.8)$$

which is shown in Figure 4.b and and the minimum χ^2 –divergence measure is

$$\chi^2_{\min}(f,g) = \int_0^1 \left(\frac{6}{5} - \frac{18}{5}x + \frac{24}{5}x^2\right)^2 dx - 1 = \frac{31}{125}. (3.4.9)$$

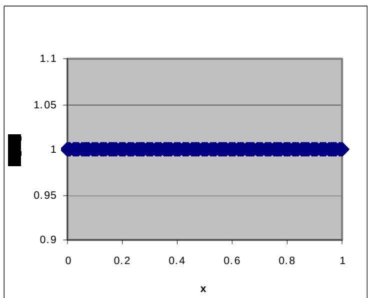

Figure 4.a. f(x) for X~U[0,1], AM=0.5.

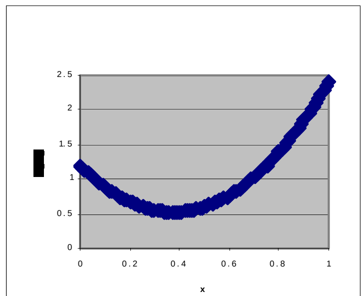

Figure 4.b. f(x) for X~U[0,1], AM=3/5, Var=1/10.

4. 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 methodology for characterizing the continuous probability distributions given a *prior* distribution and the partial information in terms of averages and variance. It is seen that the probability distributions which minimize the χ^2-distance also minimize the Kullback's measure of the directed divergence. However, with conditions on *averages* and *variance*, the minimum χ^2 –divergence principle results in the new probability distributions.

REFERENCES

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

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.

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

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

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

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.

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

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.

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

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

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

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.

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.

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

Taneja, I.J. and Kumar, Pranesh. Relative Information of Type *s*, Csiszár $f-$Divergence, and Information Inequalities. *Information Sciences*, 2004 (to appear).