CHI-SQUARE DIVERGENCE AND MINIMIZATION PROBLEM

PRANESH KUMAR AND INDER JEET TANEJA

Abstract. The minimum discrimination information principle for the Kullback-Leibler cross-entropy is well known in the literature. In this paper we have developed a minimum χ^2-divergence discrete probability distribution given a *prior* distribution and partial information in the form of average or partial information in the form of average and variance.The probability distributions, given a *prior* as *uniform, Poisson, binomial, logarithmic* and *geometric* distributions are discussed.

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 the literature on statistics, the χ^2-divergence due to Pearson (1900) is well known. In this paper, we present a methodology to derive the 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. The discrete probability distribuions considered are: Uniform, Poisson, Binomial, Logarithmic and Geometric.

2. Minimum χ^2-divergence for Discrete Probability Distributions

Let

$$\Gamma_n = \left\{ P = (p_1, p_2, ..., p_n) | \ p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}, \quad n \geqslant 2,$$

be the set of all complete finite discrete probability distributions.Given a *prior* probability distribution $Q \in \Gamma_n$, aim is to estimate, using the principle of minimum directed divergence, a probability distribution $P \in \Gamma_n$, when this underlying probability distribution P satisfies the usual probability constraints and the partial information in terms of averages.

Definition 2.1. *A probability distribution $P \in \Gamma_n$ is termed as the minimum χ^2-divergence probability distribution if it minimizes the χ^2-divergence measure*

$$\chi^2(P||Q) = \sum_{i=1}^{n} \frac{p_i^2}{q_i} - 1, \tag{2.1}$$

given:

(i) *a prior probability distribution: $Q \in \Gamma_n$,*

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

Key words and phrases. Divergence measure, Relative information, Minimum discrimination information principle, Pearson's chi-square, Characterization, Discrete probability distributions.

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

(ii) *probability constraints:* $p_i \geq 0,\ \sum_{i=1}^{n} p_i = 1,$

(iii) *partial information:* $\sum_{i=1}^{n} i^k p_i = m_k,\ k = 1, 2, 3, ..., r.$

We present the main result in the following lemma:

Lemma 2.1. *Given a prior probability distribution*

$$Q = \left\{ (q_1, q_2, ..., q_n) \middle| q_i \geq 0, \sum_{i=1}^{n} q_i = 1 \right\},\ n \geqslant 2,$$

and the constraints

$$p_i \geq 0,\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} i^k p_i = m_k,\ k = 1, 2, 3, ..., r, \tag{2.2}$$

the minimum χ^2-*divergence probability distribution is*

$$P = \left\{ (p_1, p_2, ..., p_n) \middle| p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\},\ n \geqslant 2, \tag{2.3}$$

where

$$p_i = \frac{q_i}{2} \left(\alpha_0 + \sum_{k=1}^{r} i^{\ k}\ \alpha_k \right), \tag{2.4}$$

and $(r+1)$ *constants,* α_0 *and* $\alpha_k,\ k = 1, 2, 3, ..., r,$ *are determined from*

$$\sum_{i=1}^{n} \frac{q_i}{2} \left(\alpha_0 + \sum_{k=1}^{r} i^{\ k}\ \alpha_k \right) = 1, \tag{2.5}$$

and

$$\sum_{i=1}^{n} \frac{i^{\ k} q_i}{2} \left(\alpha_0 + \sum_{k=1}^{r} i^{\ k}\ \alpha_k \right) = m_k. \tag{2.6}$$

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

$$f = \left(\sum_{i=1}^{n} \frac{p_i^2}{q_i} - 1 \right) - \alpha_0 \left(\sum_{i=1}^{n} p_i - 1 \right) - \sum_{k=1}^{r} \alpha_k \left(\sum_{i=1}^{n} i^k p_i - m_k \right). \tag{2.7}$$

Differentiating f with respect to p_i and equating the result to zero, we get the following n equations:

$$\frac{\partial f}{\partial p_i} = \frac{2p_i - \alpha_0 q_i - \sum_{k=1}^{r} \alpha_k i^k q_i}{q_i} = 0. \tag{2.8}$$

¿From (2.8), we obtain the desired value of p_i given by (2.4) that minimizes $\chi^2(P||Q)$, since for all s

$$\frac{\partial^2 f}{\partial p_i^2} = \frac{2}{q_i} > 0.$$

□

Expressions for the $(r+1)$ constants, α_0 and $\alpha_k, k = 1, 2, 3, ..., r,$ given above, are obtained from the $n + r + 1$ equations (2.2) and (2.8).

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

$$\chi^2(P||Q)_{\min} = \left[\sum_{i=1}^{n} \frac{q_i}{4} \left(\alpha_0 + \sum_{k=1}^{r} i^{\ k}\ \alpha_k \right)^2 - 1 \right]. \tag{2.9}$$

2.1. **Given a *prior* Distribution and Partial Information in the Form of Average.** Suppose that a *prior* probability distribution Q and the partial information in the form of average (m) , i.e., $\sum_{i=1}^{n} ip_i = m$ is available. Then we have from Lemma 2.1 :

Theorem 2.1. *Given a prior probability distribution*

$$Q = \left\{ (q_1, q_2, ..., q_n) \middle| q_i \geq 0, \sum_{i=1}^{n} q_i = 1 \right\}, \; n \geqslant 2, \tag{2.10}$$

and the constraints

$$p_i \geq 0, \;\; \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} i \; p_i = m, \tag{2.11}$$

the minimum χ^2-divergence probability distribution is

$$P = \left\{ (p_1, p_2, ..., p_n) \middle| p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}, \; n \geqslant 2, \tag{2.12}$$

where

$$p_i = \frac{\left[\sum\limits_{i=1}^{n} i^2 q_i - m \sum\limits_{i=1}^{n} i q_i + i \left(m - \sum\limits_{i=1}^{n} i q_i \right) \right] q_i}{\sum\limits_{i=1}^{n} i^2 q_i - \sum\limits_{i=1}^{n} i q_i^2}. \tag{2.13}$$

2.2. **Given a *prior* Distribution and Partial Information in the Form of Average and Variance.** When a *prior* probability distribution Q and the partial information in the form of average (m) and variance (σ^2) ,i.e., $\sum_{i=1}^{n} i \; p_i = m$ and $\sum_{i=1}^{n} i^2 \; p_i = m^2 + \sigma^2$, are given, we get from Lemma 2.1:

Theorem 2.2. *Given a prior probability distribution*

$$Q = \left\{ (q_1, q_2, ..., q_n) \middle| q_i \geq 0, \sum_{i=1}^{n} q_i = 1 \right\}, \; n \geqslant 2, \tag{2.14}$$

and the constraints

$$p_i \geq 0, \;\; \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} i \; p_i = m, \;\; \sum_{i=1}^{n} i^2 p_i = m^2 + \sigma^2, \tag{2.15}$$

the minimum χ^2-divergence probability distribution is

$$P = \left\{ (p_1, p_2, ..., p_n) \middle| p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}, \; n \geqslant 2, \tag{2.16}$$

where

$$p_i = \frac{q_i}{2} \left(\alpha_0 + i \;\; \alpha_1 + i^2 \;\; \alpha_2 \right) \tag{2.17}$$

where

$$\alpha_0 = -2\frac{\begin{array}{c} m\left[(\sum_{i=1}^n iq_i)(\sum_{i=1}^n i^4q_i)-(\sum_{i=1}^n i^2q_i)(\sum_{i=1}^n i^3q_i)\right]+\ (m^2+\sigma^2)\left[(\sum_{i=1}^n i^2q_i)^2-(\sum_{i=1}^n iq_i)(\sum_{i=1}^n i^3q_i)\right] \\ +(\sum_{i=1}^n i^3q_i)^2-(\sum_{i=1}^n i^4q_i)(\sum_{i=1}^n i^2q_i)\end{array}}{\begin{array}{c} 2(\sum_{i=1}^n i^2q_i)(\sum_{i=1}^n iq_i)(\sum_{i=1}^n i^3q_i)-(\sum_{i=1}^n i^2q_i)^3-(\sum_{i=1}^n i^3q_i)^2 \\ -(\sum_{i=1}^n i^4q_i)(\sum_{i=1}^n iq_i)^2+(\sum_{i=1}^n i^4q_i)(\sum_{i=1}^n i^2q_i)\end{array}},$$

$$\alpha_1 = 2\frac{\begin{array}{c} m\left[\ (\sum_{i=1}^n i^4q_i)-(\sum_{i=1}^n i^2q_i)^2\right]+(m^2+\sigma^2)\left[(\sum_{i=1}^n i^2q_i)(\sum_{i=1}^n iq_i)-\sum_{i=1}^n i^3q_i\right] \\ -(\sum_{i=1}^n iq_i)(\sum_{i=1}^n i^4q_i)+(\sum_{i=1}^n i^3q_i)(\sum_{i=1}^n i^2q_i)\end{array}}{\begin{array}{c} 2(\sum_{i=1}^n i^2q_i)(\sum_{i=1}^n iq_i)(\sum_{i=1}^n i^3q_i)-(\sum_{i=1}^n i^2q_i)^3-(\sum_{i=1}^n i^3q_i)^2 \\ -(\sum_{i=1}^n i^4q_i)(\sum_{i=1}^n iq_i)^2+(\sum_{i=1}^n i^4q_i)(\sum_{i=1}^n i^2q_i)\end{array}},$$

$$\alpha_2 = 2\frac{\begin{array}{c} m\left[(\sum_{i=1}^n i^2q_i)(\sum_{i=1}^n iq_i)-\sum_{i=1}^n i^3q_i\right]+(m^2+\sigma^2)\left[(\sum_{i=1}^n i^2q_i)-(\sum_{i=1}^n iq_i)^2\right] \\ -(\sum_{i=1}^n i^2q_i)^2+(\sum_{i=1}^n i^3q_i)(\sum_{i=1}^n iq_i)\end{array}}{\begin{array}{c} 2(\sum_{i=1}^n i^2q_i)(\sum_{i=1}^n iq_i)(\sum_{i=1}^n i^3q_i)-(\sum_{i=1}^n i^2q_i)^3-(\sum_{i=1}^n i^3q_i)^2 \\ -(\sum_{i=1}^n i^4q_i)(\sum_{i=1}^n iq_i)^2+(\sum_{i=1}^n i^4q_i)(\sum_{i=1}^n i^2q_i)\end{array}}.$$

In what follows now, we discuss special cases of the minimum χ^2-divergence probability distributions and their properties.

3. χ^2-Divergence and Discrete Probability Distributions

In this section we shall consider the problem done in previous section the minimization of χ^2-divergence considering *a priori* some well known discrete probability distributions such as, uniform, Poisson, binomial, log series, etc.

3.1. Given Uniform Distribution A *Prior* and Partial Information in the Form of Average.

Proposition 3.1. *The probability distribution P which minimizes the χ^2-divergence between P and Q given a prior distribution Q as uniform probability distribution, i.e., $q_i = \frac{1}{n}, i = 1,2,...,n$, and the constraints*

$$p_i \geq 0,\ \sum_{i=0}^{\infty} p_i = 1,\ \sum_{i=1}^{n} ip_i = m,$$

is

$$p_i = \frac{2}{n(n-1)}[(2n-3m+1)+3(\frac{2m-n-1}{n+1})\ i]\ ,\frac{n+1}{2} \leq m \leq \frac{2n+1}{3}. \tag{3.1}$$

The r^{th} moment $(r = 1,2,3,...)$ about origin of the minimum χ^2-divergence probability distribution P is given by

$$M_r = \sum_{i=1}^{n} \frac{2}{n(n-1)}[(2n-3m+1)i^r+3(\frac{2m-n-1}{n+1})\ i^{r+1}]\ ,\ \frac{n+1}{2} \leq m \leq \frac{2n+1}{3}. \tag{3.2}$$

In particular, the first four moments about origin are:

$$M_1 = m,$$
$$M_2 = \frac{(n+1)(6m-n-2)}{6},$$
$$M_3 = \frac{(3n+4)\,(3n+1)\,m-(2n+1)\,(n+2)\,(n+1)}{10},$$

and

$$M_4 = \frac{(n+1)\{6\left(4n^2+5n-1\right)m-(n+2)\left(6n^2+6n-1\right)\}}{30},$$

where

$$\frac{n+1}{2} \leq m \leq \frac{2n+1}{3}.$$

The mean (μ) and variance (σ^2) of the minimum χ^2-divergence probability distribution P for $\frac{n+1}{2} \leq m \leq \frac{2n+1}{3}$ are:

$$\frac{n+1}{2} \leq \mu \leq \frac{2n+1}{3},$$

and

$$\frac{(11n+13)(n-1)}{144} \leq \sigma^2 \leq \frac{\left(n^2-1\right)}{12},$$

respectively, where $\mu = m$ and $\sigma^2 = \frac{(n+1)(6m-n-2)}{6} - m^2$.

Here below are some interesting particular cases.

Case 3.1.1. *For $m = \frac{n+1}{2}$, $P = Q$ and the probability distribution which minimizes the χ^2-divergence between P and Q is the uniform distribution*

$$p_i = \frac{1}{n}, i = 1, 2, ..., n.$$

Case 3.1.2. *For $m = \frac{2n+1}{3}$, the probability distribution which minimizes the χ^2-divergence between P and Q is*

$$p_i = \frac{2i}{n(n+1)}, i = 1, 2, ..., n.$$

In this case, the r^{th} moment about origin of P is given by

$$M_r = \sum_{i=1}^{n} \frac{2i^{r+1}}{n(n+1)},\ r = 1, 2, 3, ...$$

In particular, the first four moments about origin are:

$$M_1 = \frac{2n+1}{3},\ M_2 = \frac{n(n+1)}{2},$$

$$M_3 = \frac{(2n+1)\left(3n^2+3n-1\right)}{15},\ M_4 = \frac{n\,(n+1)\left(2n^2+2n-1\right)}{6},$$

and the mean (μ) and variance (σ^2) are

$$\mu = \frac{2n+1}{3},\ \sigma^2 = \frac{(n+2)(n-1)}{18}.$$

Case 3.1.3. *Consider a prior uniform distribution Q with $n = 6$. Suppose the partial information about average (m) is such that*

$$\frac{7}{2} \leq m \leq \frac{13}{3}.$$

Thus, for $m = 3.5$, the probability distribution which minimizes the χ^2-divergence between P and Q is the uniform distribution

$$p_i = \frac{1}{6},\ i = 1, 2, ..., 6.$$

However, for $\frac{7}{2} < m \leq \frac{13}{3}$, the probability distribution which minimizes the χ^2-divergence between P and Q is not the uniform distribution. This distribution is

$$p_i = \frac{13}{15} - \frac{m(7-i)}{35} - \frac{i(7-m)}{35},\quad \frac{7}{2} \leq m \leq \frac{13}{3},\ i = 1, 2, ..., 6.$$

3.2. Given Uniform Distribution A *Prior* and Partial Information in the Form of Average and Variance.

Proposition 3.2. *The probability distribution P which minimizes the χ^2-divergence between P and Q given a prior distribution Q as uniform probability distribution, i.e., $q_i = \frac{1}{n}$, $i = 1, 2, ..., n$, and the constraints*

$$p_i \geq 0,\ \sum_{i=0}^{\infty} p_i = 1,\ \sum_{i=1}^{n} i\,p_i = m \text{ and } \sum_{i=1}^{n} i^2 p_i = m^2 + \sigma^2,$$

is

$$p_i = \left[\begin{array}{c} \{10m^2 - 6m(2n+1)\ + 3n\,(n+1) + 2(5\sigma^2+1)\} \\ -2\frac{30m^2(n+1)-2m(8n+11)(2n+1)\ +3(2n+1)(n+2)(n+1)+30\sigma^2(n+1)}{(n+1)(n+2)}i \\ +10\frac{6\ m^2-6m(n+1)+(n+2)(n+1)+6\sigma^2}{(n+1)(n+2)}i^2 \end{array}\right]\frac{3}{n\,(n-1)\,(n-2)}, \tag{3.3}$$

subject to the simultaneous sufficient restrictions on m *and* σ^2 *in terms of* n *so that* $p_i \geq 0$,*i.e.,*

$$\frac{1}{10}\left[3(n+1) - \sqrt{6n(n+1) - 11 - 100\sigma^2}\right] \leq m \leq \frac{1}{10}\left[3(n+1) + \sqrt{6n(n+1) - 11 - 100\sigma^2}\right], \tag{3.4}$$

$$\frac{1}{6}\left[3(n+1) - \sqrt{3(n^2-1) - 36\sigma^2}\right] \leq m \leq \frac{1}{6}\left[3(n+1) + \sqrt{3(n^2-1) - 36\sigma^2}\right], \tag{3.5}$$

$$\frac{(8n+11)(2n+1) - \sqrt{(2n+1)(38n^3+56n^2-32n-59) - 900\sigma^2(n+1)^2}}{30(n+1)} \leq m \leq \frac{(8n+11)(2n+1) + \sqrt{(2n+1)(38n^3+56n^2-32n-59) - 900\sigma^2(n+1)^2}}{30(n+1)}. \tag{3.6}$$

The r^{th} moment $(r = 1, 2, 3, ...)$ about origin of the minimum χ^2-divergence probability distribution P is given by

$$M_r = \sum_{i=0}^{\infty} \left[\begin{array}{c} \{10m^2 - 6m(2n+1) + 3n(n+1) + 2(5\sigma^2+1)\} i^r \\ -2\frac{30m^2(n+1) - 2m(8n+11)(2n+1) + 3(2n+1)(n+2)(n+1) + 30\sigma^2(n+1)}{(n+1)(n+2)} i^{r+1} \\ +10\frac{6\ m^2 - 6m(n+1) + (n+2)(n+1) + 6\sigma^2}{(n+1)(n+2)} i^{r+2} \end{array} \right] \frac{3}{n(n-1)(n-2)}, \tag{3.7}$$

Now we discuss some special cases of the minimum χ^2-divergence probability distribution P that arise for specific values of average (m) and variance (σ^2):

Case 3.2.1. *For* $m = \frac{n+1}{2}$ *and* $\sigma^2 = \frac{n^2-1}{12}$. *Note that*

$$m - \sigma^2 = \frac{(n+1)(7-n)}{12},$$

implying that

$$m = \begin{cases} \geq \sigma^2, & n \leq 7, \\ < \sigma^2, & n > 7. \end{cases}$$

In this case, $\alpha_0 = 1$, $\alpha_1 = \alpha_2 = 0$, *and* $P = Q$, *that is,*

$$p_i = \frac{1}{n}, \quad i = 1, 2, ..., n.$$

Case 3.2.2. *For* $m = \frac{2n+1}{3}$ *and* $\sigma^2 = \frac{(n+2)(n-1)}{18}$. *Since*

$$m - \sigma^2 = \frac{8 + 11n - n^2}{18},$$

$$m = \begin{cases} \geq \sigma^2, & n \leq 11, \\ < \sigma^2, & \textit{else.} \end{cases}$$

In this case, $\alpha_1 = \frac{2}{n+1}$, $\alpha_0 = \alpha_2 = 0$, *and* P *becomes*

$$p_i = \frac{2i}{n(n+1)}, \quad i = 1, 2, ..., n.$$

Case 3.2.3. *For* $m = \frac{3n(n+1)}{2(2n+1)}$ *and* $\sigma^2 = \frac{(n-1)(n+2)(3n^2+3n+2)}{20(2n+1)^2}$.
We have

$$m - \sigma^2 = \frac{4 - n(n+1)\left(3n^2 - 57n - 34\right)}{20(2n+1)^2},$$

implying that

$$m = \begin{cases} \geq \sigma^2, & n \geq 20, \\ < \sigma^2, & \textit{else.} \end{cases}$$

In this case, $\alpha_0 = \alpha_1 = 0$, $\alpha_2 = \frac{6}{(n+1)(2n+1)}$, *and* P *is*

$$p_i = \frac{6i^2}{n(n+1)(2n+1)}, \quad i = 1, 2, ..., n.$$

The r^{th} moment about origin for $r = 1, 2, 3, ...$ is given by

$$M_r = \sum_{i=1}^{n} \frac{6\ i^{r+2}}{n\,(n+1)\,(2n+1)},\ \ i = 1, 2, ..., n.$$

3.3. **Given Poisson Distribution A *Prior* and Partial Information in the Form of Average.**

Proposition 3.3. *The probability distribution P which minimizes the χ^2-divergence between P and Q given a prior distribution Q as Poisson probability distribution, i.e, $q_i = \frac{a^i e^{-a}}{i!}$, $a > 0,\ i = 0, 1, 2, ...,$ and the constraints*

$$p_i \geq 0,\ \ \sum_{i=0}^{\infty} p_i = 1,\ \ \sum_{i=0}^{\infty} i p_i = m,$$

is

$$p_i = [(1 + a - m) + (\frac{m-a}{a})\ i]\ q_i, \quad a \leq m \leq 1 + a. \tag{3.8}$$

The r^{th} moment $(r = 1, 2, 3, ...)$ about origin of the minimum χ^2-divergence probability distribution P is given by

$$M_r = \frac{1+a-m}{e^a} \sum_{i=0}^{\infty} \frac{a^i i^r}{i!} + \frac{m-a}{ae^a} \sum_{i=0}^{\infty} \frac{a^i\ i^{r+1}}{i!},\ \ a \leq m \leq 1 + a. \tag{3.9}$$

In particular, the first four moments about origin for $a \leq m \leq 1 + a$ are:

$$M_1 = m,$$
$$M_2 = m(1 + 2a) - a^2,$$
$$M_3 = m(1 + 6a + 3a^2) - a^2(3 + 2a),$$

and

$$M_4 = m\{1 + 2a\,(2a + 7)\,(1 + a)\} - a^2(7 + 12a + 3a^2).$$

The mean (μ) and variance (σ^2) of the minimum χ^2-divergence probability distribution P are:

$$a \leq \mu \leq 1 + a \text{ and } a \leq \sigma^2 \leq \frac{1}{4} + a$$

where $\mu = m$ and $\sigma^2 = m - (m - a)^2$.

Here below are some particular cases:

Case 3.3.1. *For $m = a$, $P = Q$. Thus the probability distribution which minimizes the χ^2-divergence between P and Q is the Poisson distribution ,i.e.,*

$$p_i = \frac{a^i e^{-a}}{i!},\quad a > 0,\ i = 0, 1, 2, ...$$

Case 3.3.2. *For $m = 1 + a$, the probability distribution which minimizes the χ^2-divergence between P and Q is*

$$p_i = \frac{a^{i-1} e^{-a}}{(i-1)!}, a > 0,\ i = 1, 2, ...$$

The r^{th} moment about origin of P for $r = 1, 2, 3, ...$is given by

$$M_r = \sum_{i=1}^{\infty} \frac{i^r a^{i-1} e^{-a}}{(i-1)!}\ .$$

Case 3.3.3. *Let given be a prior Poisson distribution Q with $a = 2$ and the partial information about average (m) be such that $a = 2 \leq m \leq\ 1 + a =\ 3$. Thus, for $m = 2$, the probability distribution which minimizes the χ^2-divergence between P and Q is a Poisson distribution*

$$p_i = \frac{2^i e^{-2}}{i!},\quad i = 0, 1, 2, ...$$

However, for $2 < m \leq 3$, the probability distribution which minimizes the χ^2-divergence between P and Q is not the Poisson distribution. This distribution is

$$p_i = [(6-m)+(m-2)i]\frac{2^{i-1}e^{-2}}{i!}, \quad 2 < m \leq 3, i = 1,2,....$$

3.4. **Given Poisson Distribution A *Prior* and Partial Information in the Form of Average and Variance.**

Proposition 3.4. *The probability distribution P which minimizes the χ^2-divergence between P and Q given a prior distribution Q as Poisson probability distribution, i.e, $q_i = \frac{a^i e^{-a}}{i!}$, $a > 0$, $i = 0,1,2,...$, and the constraints*

$$p_i \geq 0, \ \sum_{i=0}^{\infty} p_i = 1, \ \sum_{i=0}^{\infty} i\ p_i = m \text{ and } \sum_{i=0}^{\infty} i^2 p_i = m^2+\sigma^2,$$

is

$$p_i = \left[\begin{array}{c} a^2\{m^2-(3+2a)m+(2+2a+a^2+\sigma^2)\} - i\{(1+2a)m^2 \\ -(1+6a+4a^2)m+(\sigma^2+2a\sigma^2+3a^2+2a^3)\} + i^2\{m^2-(1+2a)m+(a^2+\sigma^2)\end{array}\right]\frac{q_i}{2a^2}, \tag{3.10}$$

subject to the simultaneous sufficient restrictions on m and σ^2 in terms of a so that $p_i \geq 0$, i.e.,

$$a+\tfrac{3}{2}-\sqrt{1+4(a-\sigma^2)} \leq m \leq a+\tfrac{3}{2}+\sqrt{1+4(a-\sigma^2)}, \tag{3.11}$$

$$a+\tfrac{1}{2}-\tfrac{1}{2}\sqrt{1+4(a-\sigma^2)} \leq m \leq a+\tfrac{1}{2}+\tfrac{1}{2}\sqrt{1+4(a-\sigma^2)}, \tag{3.12}$$

$$\frac{2a+(\ 2a+1)^2-\sqrt{1+4a(2a+3)(2a+1)-\{2\sigma(2a+1)\}^2}}{2(\ 2a+1)} \tag{3.13}$$

$$\leq m \leq \frac{2a+(\ 2a+1)^2+\sqrt{1+4a(2a+3)(2a+1)-\{2\sigma(2a+1)\}^2}}{2(2a+1)}.$$

The r^{th} moment about origin of the minimum χ^2-divergence probability distribution P for $r = 1,2,3,...$ is given by

(3.14)

$$M_r = \sum_{i=0}^{\infty}\left[\begin{array}{c} a^2\{m^2-(3+2a)m+(2+2a+a^2+\sigma^2)\}i^{r-1} - \ \{(1+2a)m^2- \\ (1+6a+4a^2)m+(\sigma^2+2a\sigma^2+3a^2+2a^3)\}i^r + \ \{m^2-(1+2a)m+(a^2+\sigma^2)\}i^{r+1}\end{array}\right]\frac{e^{-a}a^i}{2a^2(i-1)!},$$

Now we discuss some special cases of the minimum χ^2-divergence probability distribution P that arise for specific values of average (m) and variance (σ^2) :

Case 3.4.1. *For $m = \ \sigma^2 = a$. In this case, $\alpha_0 = 1$, $\alpha_1 = \alpha_2 = 0$, and $P = Q$, that is,*

$$p_i = \frac{a^i\ e^{-a}}{i\ !}, \quad a > 0,\ i = 0,1,2,....$$

Case 3.4.2. *For $m = 1+a$, $\sigma^2 = a$. Note that $m-\sigma^2 = 1$, implying that the average is larger than variance by one.*

Further, $\alpha_1 = \frac{1}{a}$, $\alpha_0 = \alpha_2 = 0$, and P becomes

$$p_i = \frac{a^{i-1}e^{-a}}{(i-1)!}, \quad a > 0,\ i = 1,2,....$$

Case 3.4.3. *For $m = \frac{1+3a+a^2}{1+a}$, $\sigma^2 = \frac{a(2+2a+a^2)}{(1+a)^2}$. It is noted that*

$$m-\sigma^2 = 1+\left(\frac{a}{1+a}\right)^2,$$

ascertaining that the average is larger than the variance by more than one.

In this case, $\alpha_0 = \alpha_1 = 0$, $\alpha_2 = \frac{1}{a(1+a)}$, and P is

$$p_i = \frac{ia^{i-1}e^{-a}}{(i-1)!(1+a)}, \quad a > 0,\ i = 1,2,....$$

The r^{th} moment about origin for $r = 1,2,3,...$ is given by

$$M_r = \sum_{i=1}^{\infty}\frac{i^{r+1}a^{i-1}e^{-a}}{(i-1)!(1+a)}, \ a > 0.$$

3.5. **Given Binomial Distribution A *Prior* and Partial Information in the Form of Average.**

Proposition 3.5. *The probability distribution P which minimizes the χ^2-divergence between P and Q given a prior distribution Q as binomial probability distribution, i.e, $q_i = \binom{n}{i}a^i(1-a)^{n-i},\ 0<a<1,\ i = 0,1,2,...,n$ and the constraints*

$$p_i \geq 0,\ \sum_{i=0}^{n} p_i = 1 \text{ and } \sum_{i=0}^{n} ip_i = m,$$

is

$$p_i = \left[\frac{na(1+na-m-a)+(m-na)\ i}{na(1-a)}\right] q_i,\quad na \leq m \leq na+1-a. \tag{3.15}$$

The r^{th} moment $(r = 1,2,3,...)$ about origin of the minimum χ^2-divergence probability distribution P is given by

$$M_r = \frac{\sum_{i=0}^{n} \binom{n}{i}a^i(1-a)^{n-i}i^r}{1-a}\left[(1+na-m-a)+\frac{(m-na)}{na}\ i\ \right],\quad na \leq m \leq na+1-a. \tag{3.16}$$

In particular, the first four moments about origin for $na \leq m \leq na+1-a$, are:

$$\begin{aligned}
M_1 &= m,\\
M_2 &= m(1+2na\ -2a\ )-n(n-1)a^2,\\
M_3 &= m\{1-6a(1-a)+3n\,(na+2-3a)\,a\}-n\,(n-1)\,(2na+3-4a)\,a^2,\\
M_4 &= m[1+2\,(n-1)\,\{7+a\,(n-2)\,(2na-6a+9)\}a]+na^2[7(1-n)-6a(4-3a)\\
&\quad -3\ n\left(n^2a-6na+4n+11a-12\right)a].
\end{aligned}$$

and the mean (μ) and variance (σ^2) of the minimum χ^2-divergence probability distribution P are:

$$na \leq \mu \leq na+(1-a) \text{ and } (n-1)a(1-a) \leq \sigma^2 \leq \frac{(1-a)[(1+a)+4(n-1)a]}{4}$$

respectively, where $\mu = m$ and $\sigma^2 =\ m(1-m)+(n-1)(2m-na)a$.

Here below are some particular cases.

Case 3.5.1. *For $m = na$, $P = Q$. Thus the probability distribution which minimizes the χ^2-divergence between P and Q is the binomial distribution, i.e.,*

$$p_i = \binom{n}{i}a^i(1-a)^{n-i},\quad 0<a<1,\ i = 0,1,2,...,n.$$

Case 3.5.2. *For $m = na+1-a$, the probability distribution which minimizes the χ^2-divergence between P and Q is*

$$p_i = \binom{n-1}{i-1}a^{i-1}(1-a)^{n-i},\ 0<a<1,\ i = 1,2,...,n.$$

The r^{th} moment about origin of P for $r = 1,2,3,...$ is given by

$$M_r = \sum_{i=1}^{n}\binom{n-1}{i-1}i^r a^{i-1}(1-a)^{n-i}.$$

Thus, the mean (μ) and variance (σ^2) are

$$\mu = na+(1-a\,) \text{ and } \sigma^2 = (n-1)a(1-a).$$

Case 3.5.3. *Consider a prior binomial distribution Q with $n = 10$, $a = 0.2$. Suppose the partial information about average (m) is such that $2 \leq m \leq 2.8$. Thus, for $m = 2$, the probability distribution which minimizes the χ^2-divergence between P and Q is a binomial distribution*

$$p_i = \binom{10}{i}(0.2)^i(0.8)^{10-i},\ i = 0,1,2,...,10.$$

However, for $2 < m \leq 2.8$, the probability distribution which minimizes the χ^2-divergence between P and Q is not the binomial distribution. This distribution is

$$p_i = \binom{10}{i}(0.2)^i(0.8)^{10-i}\left[\frac{(2.8-m)+(0.5m-1)\ i}{0.8}\right],\ i = 1,2,...,10.$$

3.6. **Given Binomial Distribution A *Prior* and Partial Information in the Form of Average and Variance.**

Proposition 3.6. *The probability distribution P which minimizes the χ^2-divergence between P and Q given a prior distribution Q as binomial probability distribution, i.e, $q_i = \binom{n}{i}a^i(1-a)^{n-i},\ 0 < a < 1,\ i = 0,1,2,...,n$ and the constraints*

$$p_i \geq 0,\ \sum_{i=0}^{n} p_i = 1,\ \sum_{i=0}^{n} ip_i = m \text{ and } \sum_{i=0}^{n} i^2 p_i = m^2 + \sigma^2,$$

is

$$p_i = (\alpha_0 + \alpha_1 i + \alpha_2 i^2)q_i, \tag{3.17}$$

where

$$\begin{aligned}
\alpha_0 &= \frac{(n-1)(n-2)a^2 - 2(m-1)(n-2)a + (2 - 3m + m^2 + \sigma^2)}{2(1-a)^2},\\
\alpha_1 &= \frac{(n-1)a\left\{2(3m - m^2 - \sigma^2)\ + (4mn - 3n - 6m)a\ - 2n(n-2)a^2\right\} - m(m-1) - \sigma^2}{2n(n-1)a^2(1-a)^2},\\
\alpha_2 &= \frac{(n-1)\ (na - 2m)a +\ (m^2 + \sigma^2 - m)}{2n(n-1)a^2(1-a)^2},
\end{aligned}$$

subject to the simultaneous sufficient restrictions on m and σ^2 in terms of a so that $p_i \geq 0$,i.e.,

$$\begin{aligned}
(n-2)a - \frac{3}{2} - \frac{1}{2}\sqrt{\ 1\ - 4\sigma^2 - 4(n-2)a(a-1)} &\leq m \leq (n-2)a - \frac{3}{2} + \frac{1}{2}\sqrt{1\ - 4\sigma^2 - 4(n-2)a(a-1)},\\
(n-1)a + \frac{1}{2} - \frac{1}{2}\sqrt{1 - 4\sigma^2 + 4(n-1)a(1-a)} &\leq m \leq (n-1)a + \frac{1}{2} + \frac{1}{2}\sqrt{1 - 4\sigma^2 + 4(n-1)a(1-a)}\ ,\\
L &\leq m \leq U,
\end{aligned}$$

where

$$\begin{aligned}
L &= \frac{1 + 2(n-1)(2an - 3a + 3)a - \sqrt{(1\ - 4\sigma^2) - (n-1)\left\{\begin{array}{l} 4\left(-3 + 4\sigma^2\right)a - 16\left(-2n + n\sigma^2 - \sigma^2 + 3\right)a^2 \\ +8\left(2n^2 - 10n + 9\right)a^3 - 4(4n-9)(n-1)a^4 \end{array}\right\}}}{2\{1 + 2(n-1)a\}},\\
U &= \frac{1 + 2(n-1)(2an - 3a + 3)a + \sqrt{(1\ - 4\sigma^2) - (n-1)\left\{\begin{array}{l} 4\left(-3 + 4\sigma^2\right)a - 16\left(-2n + n\sigma^2 - \sigma^2 + 3\right)a^2 \\ +8\left(2n^2 - 10n + 9\right)a^3 - 4(4n-9)(n-1)a^4 \end{array}\right\}}}{2\{1 + 2(n-1)a\}}.
\end{aligned}$$

The r^{th} moment about origin of the minimum χ^2-divergence probability distribution P is given by

$$M_r = \sum_{i=0}^{n} \binom{n}{i} a^i(1-a)^{n-i}(\alpha_0 + \alpha_1 i + \alpha_2 i^2)i^r\ ,\ r = 1,2,3,...$$

Now we discuss some special cases of the minimum χ^2-divergence probability distribution P that arise for specific values of average (m) and variance (σ^2) :

Case 3.6.1. *For $m = na$ and $\sigma^2 = na(1-a)$. Note that $m - \sigma^2 = \ na^2 > 0$.*

In this case, $\alpha_0 = 1,\ \alpha_1 = \alpha_2 = 0,$ and $P = Q$, that is,

$$p_i = \binom{n}{i}a^i(1-a)^{n-i},\quad 0 < a < 1,\ i = 0,1,2,...,n.$$

Case 3.6.2. *For $m = 1 + (n-1)a$ and $\sigma^2 = (n-1)a(1-a)$. Note that*

$$m - \sigma^2 = 1 + (n-1)a^2 > 0.$$

In this case, $\alpha_0 = \alpha_2 = 0, \alpha_1 = \frac{1}{na}$, and P becomes

$$p_i = \binom{n-1}{i-1} a^{i-1}(1-a)^{n-i}, \quad 0 < a < 1,\ i = 1, 2, ..., n.$$

Case 3.6.3. *For $m = \frac{(1-a)(1-2a)+na(3-3a+na)}{1+(n-1)a}$ and $\sigma^2 = \frac{(n-1)a(1-a)[2(1-a)^2+na(2-3a+na)]}{[1+(n-1)a]^2}$. It is noted that*

$$m - \sigma^2 = \frac{1 + 2(n-1)a + (n-1)(2n+1)a^2 + 2(n-1)(n-2)a^3 + (n-2)(n-1)^2a^4}{\{1+(n-1)a\}^2} > 0,$$

ascertaining that the average is larger than the variance for $n \geq 2$.

In this case, $\alpha_0 = \alpha_1 = 0$, $\alpha_2 = \frac{1}{na\{1+(n-1)a\}}$, and P is

$$p_i = \binom{n-1}{i-1} \frac{ia^{i-1}(1-a)^{n-i}}{1+(n-1)a}, \quad 0 < a < 1,\ i = 1, 2, ..., n.$$

The r^{th} moment about origin for $r = 1, 2, 3, ...$ is given by

$$M_r = \sum_{i=1}^{n} \binom{n-1}{i-1} \frac{i^{r+1}a^{i-1}(1-a)^{n-i}}{1+(n-1)a},\ 0 < a < 1\ .$$

3.7. Given Logarithmic Series Distribution A *Prior* and Partial Information in the Form of Average.

Proposition 3.7. *The probability distribution P which minimizes the χ^2-divergence between P and Q given a prior distribution Q as logarithmic series probability distribution, i.e., $q_i = \frac{Ka^i}{i}$, $0 < a < 1$, $K = \frac{-1}{\log(1-a)} > 0$, $i = 1, 2, 3, ...$, and the constraints*

$$p_i \geq 0,\ \sum_{i=1}^{\infty} p_i = 1,\ \sum_{i=1}^{\infty} ip_i = m,$$

is

$$p_i = \left[\frac{aK\{1-(1-a)m\} + (1-a)\{(1-a)m - aK\}\ i}{aK(1-aK)}\right]\ q_i,\ \frac{aK}{1-a} \leq m \leq \frac{1}{1-a}. \tag{3.18}$$

The r^{th} moment about origin of the minimum χ^2-divergence probability distribution P for $r = 1, 2, 3, ...$is given by

$$M_r = \frac{K}{aK(1-aK)}\left[\ aK\{1-(1-a)m\}\sum_{i=1}^{\infty} i^{r-1}a^i + (1-a)\{(1-a)m - aK\}\sum_{i=1}^{\infty} i^r a^i\right],\ \frac{aK}{1-a} \leq m \leq \frac{1}{1-a}.$$

The mean (μ) and variance (σ^2) of the minimum χ^2-divergence probability distribution P are:

$$\frac{aK}{1-a} \leq \mu \leq \frac{1}{1-a},$$

$$\frac{a}{(1-a)^2} \leq \sigma^2 \leq \frac{aK}{(1-a)^2}, \quad \text{if} \quad 0 < a \leq 0.9,$$

$$\frac{aK}{(1-a)^2} \leq \sigma^2 \leq \frac{a}{(1-a)^2}, \quad \text{if} \quad 0.9 \leq a < 1.$$

where $\mu = m$ and $\sigma^2 = \frac{(1-a)(1+a-aK)m - Ka^2 - m^2(1-a)^2(1-aK)}{(1-aK)(1-a)^2}$.

Equality holds if $a = 0.9$. Then, we have $K = 1$ and variance $\sigma^2 = 90$.

Here below are some particular cases

Case 3.7.1. *For* $m = \frac{aK}{1-a}$, $P = Q$ *. Thus the probability distribution which minimizes the* χ^2-*divergence between* P *and* Q *is the logarithmic series distribution*

$$p_i = \frac{-a^i}{i\log(1-a)} \ , \ 0 < a < 1, \ i = 1, 2, 3, ...$$

Case 3.7.2. *For* $m = \frac{1}{1-a}$, *the probability distribution which minimizes the* χ^2-*divergence between* P *and* Q *is*

$$p_i = (1-a)a^{i-1}, \ 0 < a < 1, \ i = 1, 2, 3, ...$$

The r^{th} *moment about origin of* P *for* $r = 1, 2, 3, ...$ *is given by*

$$M_r = \sum_{i=1}^{\infty}(1-a)a^{i-1}i^r.$$

Thus, the mean (μ) *and variance* (σ^2) *are* $\mu = \frac{1}{1-a}$ *and* $\sigma^2 = \frac{a}{(1-a)^2}$ *respectively.*

Case 3.7.3. *Let there be a prior logarithmic series distribution* Q *with* $a = 0.8$. *Suppose the partial information about average* (m) *is such that* $2.4853 \leq m \leq 5$. *Then, for* $m = 2.4853$, *the probability distribution which minimizes the* χ^2-*divergence between* P *and* Q *is a logarithmic series distribution*

$$p_i = \frac{(0.62133)(0.8)^i}{i}, \ i = 1, 2, 3, ...$$

However, for $2.4853 < m \leq 5$, *the probability distribution which minimizes the* χ^2-*divergence between* P *and* Q *is not the logarithmic series distribution. This distribution is*

$$p_i = (0.\,151\,63 + 0.\,218\,95i)\,(m - 4.\,649\,6 - 0.\,870\,83i)\,, \quad i = 1, 2, 3, ...$$

3.8. Given Logarithmic Series Distribution A *Prior* and Partial Information in the Form of Average and Variance.

Proposition 3.8. *The probability distribution* P *which minimizes the* χ^2-*divergence between* P *and* Q *given a prior distribution* Q *as logarithmic series probability distribution, i.e.,* $q_i = \frac{Ka^i}{i}$, $0 < a < 1$, $K = \frac{-1}{\log(1-a)} > 0$, $i = 1, 2, 3, ...$, *and the constraints*

$$p_i \geq 0, \ \sum_{i=1}^{\infty} p_i = 1, \sum_{i=1}^{\infty} ip_i = m, \ \sum_{i=1}^{\infty} i^2 p_i = m^2 + \sigma^2,$$

is

$$p_i = (\alpha_0 + \alpha_1 i + \alpha_2 i^2)\ q_i, \tag{3.19}$$

where

$$\begin{aligned}
\alpha_0 &= \frac{m(3-m) - \sigma^2 - 2 + 2(m^2+\sigma^2-m)a - (m+m^2+\sigma^2)a^2}{Ka(a+2)-2}, \\
\alpha_1 &= \frac{(1-a)\left[\begin{array}{c} K(m^2+\sigma^2)a^3 + \left\{3(m+K) - Km + 2(K-1)(m^2+\sigma^2)\right\}a^2 - \{m(3-K) \\ +(K+1)(m^2+\sigma^2)\}a + m(m-1) + \sigma^2 \end{array}\right]}{\{Ka(a+2)-2\}\,Ka^2}, \\
\alpha_2 &= \frac{(1-a)^2\left[\begin{array}{c} \left\{m + K - (K-1)(m^2+\sigma^2)\right\}a^3 + \left\{m(1-K) + K + (2K+1)(m^2+\sigma^2)\right\}a^2 \\ -\{Km - (K+2)(m^2+\sigma^2)\}a - m(m-1) - \sigma^2 \end{array}\right]}{\{Ka(a+2)-2\}\,Ka^2},
\end{aligned}$$

subject to the simultaneous sufficient restrictions on m *and* σ^2 *in terms of* a *so that* $p_i \geq 0$,*i.e.,*

$$\begin{aligned}
\frac{a+3-\ \sqrt{1+\ \ a^2+6a-4\sigma^2\,(a-1)^2}}{2(1-a)} &\leq\ m \leq \frac{a+3+\ \sqrt{1+\ \ a^2+6a-4\sigma^2\,(a-1)^2}}{2(1-a)}, \\
L_1, L_2 &\leq\ m \leq U_1, U_2,
\end{aligned}$$

where

$$L_1 = \frac{1+(a+4-K)a-\sqrt{\begin{array}{c}\left(1+4a+a^2\right)^2-2Ka(a+1)(9a+1)+K^2a^2\left(1+4a^2+12a\right)\\ -4\sigma^2(a-1)^2(Ka-a-1)^2\end{array}}}{2(1-a)(1+a-Ka)},$$

$$U_1 = \frac{1+(a+4-K)a+\sqrt{\begin{array}{c}\left(1+4a+a^2\right)^2-2Ka(a+1)(9a+1)+K^2a^2\left(1+4a^2+12a\right)\\ -4\sigma^2(a-1)^2(Ka-a-1)^2\end{array}}}{2(1-a)(1+a-Ka)},$$

$$L_2 = \frac{1+(1-K)a-\sqrt{(1+a)^2-2Ka(3a+1)+K^2(1+4a)a^2-4\sigma^2(a-1)^2(Ka-1)^2}}{2(1-a)(1-Ka)},$$

$$U_2 = \frac{1+(1-K)a-\sqrt{(1+a)^2-2Ka(3a+1)+K^2(1+4a)a^2-4\sigma^2(a-1)^2(Ka-1)^2}}{2(1-a)(1-Ka)}.$$

The r^{th} moment about origin of the minimum χ^2-divergence probability distribution P is given by

$$M_r = \sum_{i=1}^{\infty}\frac{Ka^i}{i!}(1-a)^{n-i}(a_0+a_1i+a_2i^2)i^r\ ,\ r=1,2,3,... \tag{3.20}$$

Now we discuss some special cases of the minimum χ^2-divergence probability distribution P that arise for specific values of average (m) and variance (σ^2) :

Case 3.8.1. *For* $m=\frac{Ka}{1-a}$, $\sigma^2=\frac{Ka(1-Ka)}{(1-a)^2}$. *Note that*

$$m-\sigma^2=\frac{K(K-1)a^2}{(1-a)^2},$$

implying that

$$m\ = \begin{cases}\geq\sigma^2, & \text{if } K\geq 1,\\ <\sigma^2, & \text{else.}\end{cases}$$

In this case, $\alpha_0=1$, $\alpha_1=\alpha_2=0$, *and* $P=Q$, *that is,*

$$p_i=\frac{-a^i}{i\log(1-a)}\quad ,0<a<1,i=1,2,3,...$$

Case 3.8.2. *For* $m=\frac{1}{1-a}$, $\sigma^2=\frac{a}{(1-a)^2}$. *Note that*

$$m-\sigma^2=\frac{1-2a}{(1-a)^2}$$

implying that

$$m\ = \begin{cases}\geq\sigma^2, & \text{if } 0<a\leq 0.5,\\ <\sigma^2, & \text{if }\ 0.5<a<1.\end{cases}$$

In this case, $\alpha_0=\alpha_2=0,\alpha_1=\frac{1-a}{Ka}$, *and* P *becomes*

$$p_i=(1-a)a^{i-1},0<a<1,i=1,2,3,....$$

The r^{th} *moment about origin for* $r=1,2,3,...$ *is given by*

$$M_r=\sum_{i=1}^{\infty}(1-a)i^ra^{i-1},\ 0<a<1.$$

Case 3.8.3. *For* $m=\frac{1+a}{1-a}$, $\sigma^2=\frac{2a}{(1-a)^2}$. *It is noted that*

$$m-\sigma^2=\frac{1-2a-a^2}{(1-a)^2},$$

implying that

$$m\ = \begin{cases}\geq\sigma^2, & \text{if } 0<a\leq\sqrt{2}-1,\\ <\sigma^2, & \text{if }\ \sqrt{2}-1<a<1.\end{cases}$$

In this case, $\alpha_0=\alpha_1=0$, $\alpha_2=\frac{(1-a)^2}{Ka}$, *and* P *is*

$$p_i = \quad (1-a)^2 i a^{i-1}, 0 < a < 1, i = 1, 2, 3, ...$$

The r^{th} moment about origin for $r = 1, 2, 3, ...$ is given by

$$M_r = \sum_{i=1}^{\infty} (1-a)^2 i^{r+1} a^{i-1},\ 0 < a < 1.$$

3.9. **Given Geometric Distribution A *Prior* and Partial Information in the Form of Average.**

Proposition 3.9. *The probability distribution P which minimizes the χ^2-divergence between P and Q given a prior distribution Q as geometric probability distribution, i.e., $q_i = (1-a)a^i,\ 0 < a < 1,\ i = 0, 1, 2, ...$ and the constraints*

$$p_i \geq 0,\ \sum_{i=0}^{\infty} p_i = 1,\ \sum_{i=0}^{\infty} i p_i = m,$$

is

$$p_i = [a\,(1 + a - m + ma) + (1-a)(m - ma - a)\ i]\ (1-a)a^{i-1},\ \frac{a}{1-a} \leq m \leq \frac{1+a}{1-a}. \tag{3.21}$$

The r^{th} moment $(r = 1, 2, 3, ...)$ about origin of the minimum χ^2-divergence probability distribution P is :

$$M_r = \sum_{i=0}^{\infty} [a\,(1 + a - m + ma) + (1-a)(m - ma - a)\ i] i^r\ (1-a)a^{i-1}, \quad \frac{a}{1-a} \leq m \leq \frac{1+a}{1-a}. \tag{3.22}$$

In particular, the first four moments about origin are:

$$M_1 = m,$$

$$M_2 = \frac{m\,(3a+1)\,(1-a) - 2a^2}{(1-a)^2},$$

$$M_3 = \frac{m\left(7a^2 + 10a + 1\right)(1-a) - 6a^2(1+a)}{(1-a)^3},$$

and

$$M_4 = \frac{m\left(15a^3 + 55a^2 + 25a + 1\right)(1-a) - 2a^2(7a^2 + 22a + 7)}{(1-a)^4},\ \frac{a}{1-a} \leq m \leq \frac{1+a}{1-a}.$$

The mean (μ) and variance (σ^2) of the minimum χ^2-divergence probability distribution P are:

$$\frac{a}{1-a} \leq \mu \leq \frac{1+a}{1-a},$$

and

$$\frac{a}{(1-a)^2} \leq \sigma^2 \leq \frac{1+6a}{4\,(1-a)^2},$$

respectively, where $\mu = m$ and $\sigma^2 = \frac{m^2(a-1)^2 + m(1+2a-3a^2) - 2a^2}{(1-a)^2}$.

Case 3.9.1. *For $m = \frac{a}{1-a}$, $P = Q$ and the probability distribution which minimizes the χ^2-divergence between P and Q is the geometric distribution*

$$p_i = \ (1-a)a^i, \quad 0 < a < 1,\ i = 0, 1, 2, ...$$

Case 3.9.2. *For $m = \frac{1+a}{1-a}$, the probability distribution which minimizes the χ^2-divergence between P and Q is*

$$p_i = \ (1-a)^2 i a^{i-1}, 0 < a < 1, i = 1, 2, ...$$

The r^{th} moment about origin of P for $r = 1, 2, 3, ...$is given by

$$M_r = \sum_{i=1}^{\infty} \quad (1-a)^2 i a^{i-1} i^r \quad .$$

Thus, the mean (μ) *and variance* (σ^2) *are*

$$\mu = \frac{1+a}{1-a} \text{ and } \sigma^2 = \frac{2a}{(1-a)^2}.$$

Case 3.9.3. *Consider a geometric distribution Q with $a = 0.8$. Suppose the partial information about average (m) is such that $4 \le m \le 9$. Thus, for $m = 4$, the probability distribution which minimizes the χ^2-divergence between P and Q is a geometric distribution*

$$p_i = 0.8(0.2)^i, \ i = 0, 1, 2, ...$$

However, for $4 < m \le 9$, the probability distribution which minimizes the χ^2-divergence between P and Q is not the geometric distribution. This distribution is

$$p_i = (0.008)[4\,(9-m) + (m-4)\ i]\ (0.8)^{i-1}, \ 4 < m < 9, \ i = 1, 2, ...$$

3.10. **Given Geometric Distribution A *Prior* and Partial Information in the Form of Average and Variance.**

Proposition 3.10. *The probability distribution P which minimizes the χ^2-divergence between P and Q given a prior distribution Q as geometric probability distribution, i.e., $q_i = (1-a)a^i$, $0 < a < 1$, $i = 0, 1, 2, ...$ and the constraints*

$$p_i \ge 0, \ \sum_{i=0}^{\infty} p_i = 1, \sum_{i=0}^{\infty} i p_i = m, \ \sum_{i=0}^{\infty} i^2 p_i = m^2 + \sigma^2,$$

is

$$p_i = (\alpha_0 + \alpha_1 i + \alpha_2 i^2)\ q_i, \tag{3.23}$$

where

$$\begin{aligned}
\alpha_0 &= \frac{2(1+a+a^2)\ + 3m(1-a^2) + (m^2+\sigma^2)\,(1-a)^2}{2}, \\
\alpha_1 &= \frac{-6a^2\left(1-a^2\right)\ + m\,(1+a)\,(1+9a)\,(1-a)^2 - (m^2+\sigma^2)\,(3a+1)\,(1-a)^3}{4a^2}, \\
\alpha_2 &= \frac{2a^2\,(1-a)^2\ - m\,(1+3a)\,(1-a)^3 + (m^2+\sigma^2)\ (1-a)^4}{4a^2},
\end{aligned}$$

subject to the simultaneous sufficient restrictions on m and σ^2 in terms of a so that $p_i \ge 0$,i.e.,

$$\frac{3(a+1) - \sqrt{\left(a^2+10a+1-4\sigma^2\,(a-1)^2\right)}}{2(a-1)} \le m \le \frac{3(a+1) + \sqrt{\left(a^2+10a+1-4\sigma^2\,(a-1)^2\right)}}{2(a-1)},$$

$$\frac{-\,(1+a)\,(1+9a) - \sqrt{\left((1+a)\,(9a^3+75a^2+19a+1) - 4\sigma^2\,(3a+1)^2\,(a-1)^2\right)}}{2(3a+1)(a-1)} \le m$$

$$\le \frac{-\,(1+a)\,(1+9a) + \sqrt{\left((1+a)\,(9a^3+75a^2+19a+1) - 4\sigma^2\,(3a+1)^2\,(a-1)^2\right)}}{2(3a+1)(a-1)},$$

$$\frac{-3(a+1) - \sqrt{\left(a^2+6a+1-4\sigma^2\,(a-1)^2\right)}}{2(a-1)} \le m \le \frac{-3(a+1) + \sqrt{\left(a^2+6a+1-4\sigma^2\,(a-1)^2\right)}}{2(a-1)}.$$

The r^{th} moment about origin of the minimum χ^2-divergence probability distribution P is given by

$$M_r = \sum_{i=0}^{\infty} (1-a)a^i(a_0 + a_1 i + a_2 i^2) i^r\ , \ r = 1, 2, 3, ... \tag{3.24}$$

Now we discuss some special cases of the minimum χ^2-divergence probability distribution P that arise for specific values of average (m) and variance (σ^2) :

Case 3.10.1. *For* $m = \frac{a}{1-a}$, $\sigma^2 = \frac{a}{(1-a)^2}$. *It is noted that*

$$m - \sigma^2 = -\frac{a^2}{(1-a)^2},$$

implying that $m < \sigma^2$.

In this case, $\alpha_1 = \alpha_2 = 0$, $\alpha_1 = 1$, *and* $P = Q$, *i.e.,*

$$p_i = (1-a)a^i , \quad 0 < a < 1,\ i = 0, 1, 2, ...$$

Case 3.10.2. *For* $m = \frac{1+a}{1-a}$, $\sigma^2 = \frac{2a}{(1-a)^2}$. *It is noted that*

$$m - \sigma^2 = \frac{1 - 2a - a^2}{(1-a)^2}.$$

Further it may be seen that

$$m = \begin{cases} \geq \sigma^2, & \textit{if } 0 < a \leq \sqrt{2} - 1, \\ < \sigma^2, & \textit{if } \sqrt{2} - 1 < a < 1. \end{cases}$$

In this case, $\alpha_0 = \alpha_2 = 0$, $\alpha_1 = \frac{1-a}{a}$, *and* P *is*

$$p_i = (1-a)^2\ i\ a^{i-1}, \quad 0 < a < 1,\ i = 1, 2, ...$$

The r^{th} *moment about origin for* $r = 1, 2, 3, ...$ *is given by*

$$M_r = \sum_{i=1}^{\infty} i^{r+1}\ a^{i-1}(1-a)^2,\ 0 < a < 1.$$

Case 3.10.3. *For* $m = \frac{1+4a+a^2}{1-a^2}$, $\sigma^2 = \frac{4a\left(a+a^2+1\right)}{(1+a)^2(1-a)^2}$. *It is noted that*

$$m - \sigma^2 = \frac{1 - a^2\left(a^2 + 8a + 4\right)}{(1+a)^2(1-a)^2}.$$

Further

$$m = \begin{cases} \geq \sigma^2, & \textit{if } 0 < a \leq 0.3743, \\ < \sigma^2, & \textit{if } 0.3743 < a < 1. \end{cases}$$

In this case, $\alpha_0 = \alpha_1 = 0$, $\alpha_2 = \frac{(1-a)^2}{a(1+a)}$, *and* P *is*

$$p_i = \frac{i^2\ a^{i-1}(1-a)^3}{1+a}, \quad 0 < a < 1,\ i = 1, 2, ...$$

The r^{th} *moment about origin for* $r = 1, 2, 3, ...$ *is given by*

$$M_r = \sum_{i=1}^{\infty} \frac{i^{r+2}\ a^{i-1}(1-a)^3}{1+a},\ 0 < a < 1.$$

4. Concluding Remarks

The minimum discrimination information or 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 methodology for characterizing the discrete probability distributions based on the Pearson χ^2- distance given a *prior* distribution and the partial information in terms of moments. We have shown by considering the discrete probability distributions like *uniform, Poisson, binomial, logarithmic* and *geometric* which minimize the Kullback's measure of the directed divergence (1959)

$$K(P||Q) = \sum_{i=1}^{n} p_i \ln(\frac{p_i}{q_i})$$

that these distributions also minimize the Pearson's χ^2-distance. Further, with conditions on *mean* and *variance*, we have obtained the new probability distributions that minimize the χ^2- distance. The work on minimum χ^2- divergence *continuous* probability distributions is in progress and would be reported elsewhere.

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.

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, Csiszá r $f-$Divergence, and Information Inequalities. *Information Sciences*, 2004 (to appear).

E-mail address: kumarp@unbc.ca
URL: http://web.unbc.ca/∼kumarp
E-mail address: taneja@mtm.ufsc.br
URL: http://www.mtm.ufsc.br/∼taneja

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

Departamento de Matemática, Universidade Federal de Santa Catarina, 88.040-900 Florianópolis, SC, Brazil