International Journal of Information Sciences and Application.
ISSN 0974-2255 Volume 17, Number 1 (2025), pp. 1-3

http://www.irphouse.com

Minimum Cross-Entropy Probability Distribution

P. Jha*
* Principal, Govt. College, Komakhan, Mahasamund (Chhattisgarh) INDIA
purush.jha@gmail.com

Anjali Chandravanshi**
**Deptt. of Mathematics, Govt J. Y. Chhattisgarh College, Raipur
(Chhattisgarh), INDIA.
chandravanshi_anjali@rediffmail.com

Abstract: In this paper we study the increasing or decreasing nature of probabilities obtained when measure of cross-entropy (or relative entropy) is minimized subject to moment constraints.
Key Words: Cross-Entropy, Moment**,** Entropy, Directed divergence.
Subject Classification No.: 94A17.

Introduction

Minimizing the Kullback-Leibler's [3] cress-entropy (or relative entropy) with respect to a set of moment constraints finds its importance in the celebrated Kullback's minimum cross-entropy principle. This principle, also known as the minimum directed divergence principle [4] is an entropy optimization principle similar to Jayne's [1] maximum entropy principle. Minimum cross-entropy principle more general compared to Jayne's maximum entropy principle, in the sense that minimizing cross-entropy is equivalent to maximizing Shannon entropy [5] when the prior is a uniform distribution.

Let $P=\left(p_1,p_2,...,p_n\right)$ are the probabilities obtained by minimizing Kullback-Leibler's measure of cross-entropy

$$D(P:Q)=\sum_{i=1}^{n} p_i \ln\frac{p_i}{q_i}, \tag{1}$$

where $Q=\left(q_1,q_2,...,q_n\right)$ is any given probability distribution, subject to the natural constraint and moment constraint

$$\sum_{i=1}^{n} p_i = 1, \quad \forall p_i \geq 0 \text{ and} \tag{2}$$

$$\sum_{i=1}^{n} p_i g\left(i\right) = M \tag{3}$$

where $g(i)$ is a positive integral-valued increasing function of i for which $g(0)=0$, from this, Kapur [2] showed that $\frac{p_i}{q_i}$ is an increasing, constant, decreasing function of i according as M is greater than, equal to, less than M_0, where M_0 is defined by

$$\sum_{i=1}^{n} q_i g(i) = M_0$$

In this communication, next section we minimize cross-entropy due to Fermi Dirac's Measure of entropy and we discuss the increasing or decreasing nature of probability distribution.

Minimum Cross-Entropy Probability Distribution

Theorem: The probability distribution obtained by minimizing the measure of cross-entropy due to Fermi Dirac

$$D_f(P:Q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} + \sum_{i=1}^{n} (1-p_i) \ln \frac{(1-p_i)}{(1-q_i)} \tag{4}$$

subject to (2) and (3) is such that each p_i and $p_i(1-q_i)/q_i(1-p_i)$ always increase with i or always decrease with i or remains equal to unity.

Proof: Minimizing (4) subject to the constraint (2) and (3), we get

$$\frac{p_i}{(1-p_i)} = \frac{q_i}{(1-q_i)} e^{-\lambda-\mu g(i)} \tag{5}$$

where λ and μ are determined by using

$$\sum_{i=1}^{n} \frac{1}{\left(\frac{1-q_i}{q_i}\right) e^{\lambda+\mu g(i)}} = 1, \tag{6}$$

and
$$\sum_{i=1}^{n} \frac{g(i)}{\left(\frac{1-q_i}{q_i}\right) e^{\lambda+\mu g(i)}} = M \tag{7}$$

(i) If $\mu = 0$, then $\lambda = 0$ and $p_i(1-q_i)/q_i(1-p_i) = 1$ for each i.

(ii) If $\mu < 0$, $p_i(1-q_i)/q_i(1-p_i)$ increase with i.

(iii) If $\mu > 0$, $p_i(1-q_i)/q_i(1-p_i)$ decrease with i.

However, in the last two cases p_i/q_i may sometimes increase and sometimes decrease with i.

If, $q_i = 1/n$, we get the maximum entropy probability distribution and $p_i/(1-p_i)$ will always increase with i or always decrease with i or remain constant with value $1/(n-1)$ for all i . Now,

$$\frac{p_i}{(1-p_i)} = k_i \Rightarrow p_i = \frac{k_i}{(1+k_i)} = \frac{1}{\frac{1}{k_i}+1} \tag{8}$$

so that if k_i always increase or decrease with i ,then p_i always increase or decrease with i .
In the general case,

$$\frac{p_i}{(1-p_i)} = k_i \frac{q_i}{(1-q_i)} \tag{9}$$

(i) If k_i and q_i both increase with i , p_i always increases with i .

(ii) If both k_i and q_i decrease with i , p_i also decreases with i .

(iii) If k_i increases and q_i decreases with i , then p_i will increase or decrease with i according as $k_i\, q_i/(1-q_i)$ increases or decreases.

REFERENCES

[1]. E.T.Jaynes(1957), Information Theory and Statiatical Mechanics, Physical Reviews,. Vol. 106, pp. 620 - 630.

[2]. J.N. Kapur(1994), Measure of Information and Their Applications, Wiley Eastern, New Delhi.

[3]. S. Kullback and R.LA. Leibler(1951), On Information and Sufficiency, Ann. Math. Stat., Vol. 22, pp. 79 - 88.

[4]. S. Kullback(1959), Information Theory and Statistics, John Wiley, New York.

[5]. C.E.Shannon, A Mathematical Theory of Communication, Bell System Tech. J. 27, (1948), 379-423, 623-659.