Malaysian Journal of Science **25** (2): 153 – 157 (2006)

On the Convexity of a Rate Function

Choon Peng Tan
Faculty of Information and Communication Technology. Universiti Tunku Abdul Rahman.
13 Jalan 13/6, 46200 Petaling Jaya, Selangor, Malaysia
tancp@mail.utar.edu.my

ABSTRACT Consider a sequence of independent, identically distributed random variables from a common distribution Q. Suppose that there are k rare events associated with Q, for example events involving the k sample moments. Then the dominant rate at which the probability of the intersection of the k events converges to zero can be expressed as the relative entropy of a certain distribution and Q. Considering the intersection of the k events as a function of $\boldsymbol{\alpha}$ in a k-dimensional Euclidean space, we shall show that the dominant rate of convergence is convex in $\boldsymbol{\alpha}$. This result is the consequence of the convex property of a certain function associated with the dominant rate of convergence which shall be shown.

ABSTRAK Pertimbangkan satu jujukan pembolehubah rawak yang tertabur secara secaman dan tak bersandar daripada satu taburan sepunya Q. Katalah terdapat k peristiwa kejarangan yang bersekutu dengan Q, misalnya peristiwa-peristiwa yang melibatkan k momen sample. Maka kadar dominant yang mana kebarangkalian persilangan k peristiwa menumpu kepada sifar boleh diungkapkan sebagai entropi relatif bagi sesuatu taburan dan Q. Dengan mempertimbangkan persilangan k peristiwa sebagai satu fungsi $\boldsymbol{\alpha}$ dalam ruang Euklidan berdimensi k, kami akan tunjukkan bahawa kadar penumpuan dominant adalah cembung dalam $\boldsymbol{\alpha}$. Hasil ini adalah akibat sifat kecembungan sesuatu fungsi lain yang bersekutu dengan kadar penumpuan dominant yang akan ditunjukkan.

(Convex property, dominant rate of convergence, large deviation theory, relative entropy)

INTRODUCTION

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent, identically distributed random variables from a discrete distribution $Q(x)$. Consider k linearly independent functions of X, namely $g_1(X), g_2(X), \ldots, g_k(X)$ and define k events $\left\{ \frac{1}{n} \sum_{i=1}^{n} g_1(X_i) \geq \alpha_1 \right\}$,

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} g_2(X_i) \geq \alpha_2 \right\}, \ldots, \left\{ \frac{1}{n} \sum_{i=1}^{n} g_k(X_i) \geq \alpha_k \right\}$$

associated with these functions, where $\alpha_1, \alpha_2, \ldots, \alpha_k$ do not depend on n. For example, if $g_j(X_i) = X_i^r$ for some j and integer $r \geq 1$, then the jth event

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} X_i^r \geq \alpha_j \right\}$$

represents the rth sample moment larger than the constant α_j. We are interested in the probability of

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} X_i^r \geq \alpha_j \right\}$$

where $\alpha_j > E_Q(X^r)$. In general, we are interested in the probability of the intersection of the k rare events, namely,

$$\Pr\left\{\frac{1}{n}\sum_{i=1}^{n} g_j(X_i) \geq \alpha_j \ \text{ for } j = 1,2,\ldots,k\right\}.$$

This probability is equivalent to

$$\Pr\left\{\sum_x P_{X^n}(x) g_j(x) \geq \alpha_j \ \text{ for } j = 1,2,\ldots,k\right\},$$

where P_{X^n} is the type of $X^n = (X_1, X_2, \ldots, X_n)$ and $P_{X^n}(x)$ is the relative frequency of x in x^n, for x in the alphabet of X. The evaluation of this probability, denoted by $Q^n(E)$, is done by defining a convex set of probability distributions

$$E = \left\{P : \sum_x P(x) g_j(x) \geq \alpha_j\ ,\ j = 1,2,\ldots,k\right\} \tag{1}$$

Then by Sanov's Theorem [1] [2],

$$\frac{1}{n}\log Q^n(E) \to -D(P^* \| Q) \quad \text{as } n \to \infty, \tag{2}$$

where $D(\cdot \| \cdot)$ denotes the relative entropy of two probability distributions, P^* is the distribution that minimizes $D(P \| Q)$ over all P in E. By using the method of Lagrange multipliers [1], the distribution P^* is shown to be

$$P^*(x) = cQ(x)2^{\sum_{j=1}^{k}\lambda_j g_j(x)} \tag{3}$$

where the normalizing constant

$$c = \left[\sum_x Q(x) 2^{\sum_{j=1}^{k}\lambda_j g_j(x)}\right]^{-1} \tag{4}$$

$\lambda_1, \lambda_2, \ldots, \lambda_k$ are positive solutions to

$$\sum_x P^*(x) g_j(x) = \alpha_j, \quad j = 1,2,\ldots,k, \tag{5}$$

if $\alpha_j > E_Q[g_j(X)]$, $j = 1,2,\ldots,k$ and logarithms are to base 2. For large n,

$$Q^n(E) \approx 2^{-nD(P^* \| Q)} \tag{6}$$

and hence $D(P^* \| Q)$ is the dominant rate of convergence of $Q^n(E)$ to 0. In this paper, we shall assume that $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)$ is a variable and consider $D(P^* \| Q)$ as a function of α. We shall show that $D(P^* \| Q)$ is a convex function of α. In a previous paper [3], we have studied the properties of $D(P^* \| Q)$ as a function of the scalar α.

MAIN RESULTS

Theorem

Let $g_1(x), g_2(x), \ldots, g_k(x)$ be k linearly independent functions of x and

$$E = \left\{P : \sum_{x \in A} P(x) g_j(x) \geq \alpha_j, \quad j = 1,2,\ldots,k\right\}$$

be a convex set of probability distributions, where A is the alphabet set of the random variable X. For $Q \notin E$, let

$$P^*_\alpha(x) = cQ(x)2^{\sum_{j=1}^{k}\lambda_j(\alpha) g_j(x)} \qquad \text{for } x \in A$$

be the distribution that minimizes $D(P\|Q)$ over all $P \in E$, where $\lambda = (\lambda_1(\alpha), \lambda_2(\alpha), \ldots, \lambda_k(\alpha))$ is the solution to the k equations

$$\sum_{x \in A} P_\alpha^*(x) g_j(x) = \alpha_j, \quad j = 1, 2, \ldots, k,$$

the normalizing constant

$$c = \left[\sum_{x \in A} Q(x) 2^{\sum_{j=1}^{k} \lambda_j(\alpha) g_j(x)} \right]^{-1}$$

.

Then,

(i)

$$h(\lambda) = \log \left[\sum_{x \in A} Q(x) 2^{\sum_{j=1}^{k} \lambda_j(\alpha) g_j(x)} \right]$$

is convex in λ.

(ii)

$D(P_\alpha^* \| Q)$ is convex in α.

Proof

(i)

$$\frac{\partial h}{\partial \lambda_i} = \frac{\left[\sum_x Q(x) g_i(x) 2^{\sum_{j=1}^{k} \lambda_j(\alpha) g_j(x)} \right]}{\left[\sum_x Q(x) 2^{\sum_{j=1}^{k} \lambda_j(\alpha) g_j(x)} \right]} = \alpha_i, \tag{7}$$

and,

$$\frac{\partial^2 h}{\partial \lambda_i \partial \lambda_j} = (\ln 2) \left\{ E_{P_\alpha^*}[g_i(X) g_j(X)] - E_{P_\alpha^*}[g_i(X)] E_{P_\alpha^*}[g_j(X)] \right\}$$

for $i = 1, 2, \ldots, k; j = 1, 2, \ldots, k$. **(8)**

The Hessian $\left(\frac{\partial^2 h}{\partial\lambda_i \partial\lambda_j}\right)$ of $h(\lambda)$ is $(\ln 2)\Sigma$, where Σ is the covariance matrix of the random vector $\mathbf{g}(X) = g_1(X), g_2(X), \ldots, g_k(X))$ and X has the distribution P_α^*. Since $\left(\frac{\partial^2 h}{\partial\lambda_i \partial\lambda_j}\right)$ is positive semidefinite, therefore $h(\lambda)$ is convex in λ.

(ii) $$D(P_\alpha^* \| Q) = \sum_x P_\alpha^*(x) \log \frac{P_\alpha^*(x)}{Q(x)}$$

$$= \sum_x P_\alpha^*(x) \left[\log c + \sum_{j=1}^{k} \lambda_j(\alpha) g_j(x) \right]$$

$$= -h(\lambda) + \sum_{j=1}^{k} \lambda_j(\alpha)\alpha_j \quad (9)$$

and hence,

$$\frac{\partial D(P_\alpha^* \| Q)}{\partial \alpha_i} = -\sum_{j=1}^{k} \frac{\partial h}{\partial \lambda_j} \frac{\partial \lambda_j}{\partial \alpha_i} + \sum_{j=1}^{k} \frac{\partial \lambda_j}{\partial \alpha_i} \alpha_j + \lambda_i$$

and from (7),

$$\frac{\partial D(P_\alpha^* \| Q)}{\partial \alpha_i} = \lambda_i \quad \text{for } i = 1,2,\ldots,k. \quad \textbf{(10)}$$

From (9), the function $D(P_\alpha^* \| Q)$ is convex in α if and only if, for any α_1 and α_2,

$$D(P_{\alpha_2}^* \| Q) - D(P_{\alpha_1}^* \| Q) \geq \nabla D(P_\alpha^* \| Q)|_{\alpha=\alpha_1} (\alpha_2 - \alpha_1)$$

This is equivalent to:

$$h(\lambda_1) - h(\lambda_2) + \sum_{j=1}^{k} \lambda_j(\alpha_2)\alpha_j(2) - \sum_{j=1}^{k} \lambda_j(\alpha_1)\alpha_j(1) \geq \sum_{i=1}^{k} \left. \frac{\partial D(P_\alpha^* \| Q)}{\partial \alpha_i} \right|_{\alpha=\alpha_1} [\alpha_i(2) - \alpha_i(1)]$$

where $\alpha_i = (\alpha_1(i), \alpha_2(i), \ldots, \alpha_k(i))$ and

$$\lambda_i = (\lambda_1(\alpha_i), \lambda_2(\alpha_i), \ldots, \lambda_k(\alpha_i))$$

for $i = 1,2$. From (7) and (10), this reduces to

$$h(\lambda_1) - h(\lambda_2) \geq \sum_{i=1}^{k} \alpha_i(2)[\lambda_i(\alpha_1) - \lambda_i(\alpha_2)]$$

$$= \sum_{i=1}^{k} \left. \frac{\partial h}{\partial \lambda_i} \right|_{\lambda=\lambda_2} [\lambda_i(\alpha_1) - \lambda_i(\alpha_2)]$$

$$= \nabla h|_{\lambda=\lambda_2} (\lambda_1 - \lambda_2).$$

Hence $D(P_\alpha^* \| Q)$ is convex in α if and only if $h(\lambda)$ is convex in λ. The result in (i) implies that $D(P_\alpha^* \| Q)$ is convex in α.

REFERENCES

1. Cover, T.M. and Thomas, J.A. (1991). *Elements of Information Theory*. New York: John Wiley.

2. Sanov, I.N. (1957). On the probability of large deviations of random variables. *Mat. Sbornik* **42**: 11 - 44.

3. Tan, C.P. (2002). The behaviour of the error probabilities in hypothesis testing for large sample sizes (Malay). *Proceedings of the Tenth National Symposium on the Mathematical Sciences*, Skudai, Malaysia: Penerbit Jabatan Matematik, Universiti Teknologi Malaysia: 436 - 442.