Journal of the Korean
Statistical Society
Vol. 24, No. 2, 1995

A Sanov-Type Proof of the Joint Sufficiency of the Sample Mean and the Sample Variance

Chul Eung Kim[1] and ByoungSeon Choi[1]

ABSTRACT

It is well-known that the sample mean and the sample variance are jointly sufficient under normality assumption. In this paper a proof of the joint sufficiency is given without using the factorization criterion. It is related to a finite Sanov-type conditional theorem, i.e., the conditional probability density of Y_1 given sample mean μ and sample variance σ^2, where $Y_1, Y_2, \ldots, Y_n$ are independently and identically distributed (*i.i.d.*) normal random variables with mean m and variance δ^2, equals that of Y_1 given sample mean μ and sample variance σ^2, where $Y_1, Y_2, \ldots, Y_n$ are *i.i.d.* normal random variables with mean μ and variance σ^2.

KEYWORDS: Conditional probability, Sufficiency, Sanov Theorem, *I*-Projection.

1. INTRODUCTION

Let $Y_1, Y_2, \ldots, Y_n$ be independently and identically distributed (*i.i.d.*) random variables with probability distribution Q. Sanov(1957) presented a large deviation theorem for the empirical distribution $\hat{P}_n$ as

$$\lim_{n\to\infty} \frac{1}{n} \log \Pr\left(\hat{P}_n \in \Pi\right) = - \inf_{P\in\Pi} D(P \parallel Q),$$

[1]Department of Applied Statistics, Yonsei University, Seoul, 120–749, Korea.

where Π is a given set of probability distributions satisfying some regularity conditions and $D(P \parallel Q)$ is the Kullback-Leibler information divergence of P from Q. The probability distribution minimizing $D(P \parallel Q)$ subject to $P \in \Pi$ is called the I-projection of Q on Π. The Sanov theorem has been studied by Bartfai(1972), Vincze(1972), Vasicek(1980) and Van Campenhout and Cover(1981). More detailed references can be found on Csiszár, Cover and Choi(1987).

A special case of the Sanov theorem is about

$$\Pi^* = \Big\{P \Big| E_P h_j(Y) \geq \alpha_j, j = 1, 2, \ldots, k\Big\},$$

where $h_1, h_2, \ldots, h_k$ are given functions defined on the range of the Y_i's and $\alpha_1, \alpha_2, \ldots, \alpha_k$ are given constants. If the probability density function ($p.d.f.$)of Q is denoted by q, then the $p.d.f.$ of the I-projection of Q on Π^* is

$$p^*(y) = cq(y) \exp\Big[\sum_{j=1}^{k} \lambda_j h_j(y)\Big].$$

Van Campenhout and Cover(1981) presented a conditional limit theorem related to the special case. Under some regularity conditions the conditional $p.d.f.$ of Y_1 given

$$\frac{1}{n}\sum_{i=1}^{n} h_1(Y_i) = \alpha_1, \frac{1}{n}\sum_{i=1}^{n} h_2(Y_i) = \alpha_2, \ldots, \frac{1}{n}\sum_{i=1}^{n} h_k(Y_i) = \alpha_k$$

tends to $p^*(y)$ as $n \to \infty$. The constants $c, \lambda_1, \lambda_2, \ldots, \lambda_k$ are determined by the equations

$$\int p^*(y)dy = 1 \text{ and } \int h_j(y)p^*(y)dy = \alpha_j, j = 1, 2, \ldots, k.$$

The purpose of this paper is to present a finite sample property of the conditional problem on the normal random variables, where the condition is about the sample mean and the sample variance. The property implies the joint sufficiency of the sample mean and the sample variance.

2. THEOREM AND PROOF

Theorem 1. The conditional *p.d.f.* of Y_1 given

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = \mu \text{ and } \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sigma^2,$$

where $Y_1, Y_2, \ldots, Y_n$ are *i.i.d.* normal random variables with mean m and variance δ^2, is the same as that of Y_1 given

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = \mu \text{ and } \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sigma^2,$$

where $Y_1, Y_2, \ldots, Y_n$ are *i.i.d. normal* random variables with mean μ and variance σ^2.

Proof. Denote the *p.d.f.* of the normal random variable Y with mean α and variance ω^2 by $\phi(y \mid \alpha, \omega^2)$. Also, let

$$G = \Big\{(y_1, \ldots, y_n)^t \ ; \ \frac{1}{n}\sum_{i=1}^{n} y_i = \mu \text{ and } \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sigma^2\Big\}$$

and

$$G_\epsilon = \Big\{(y_1, \ldots, y_n)^t \ ; \ \Big|\frac{1}{n}\sum_{i=1}^{n} y_i - \mu\Big| < \epsilon_1 \text{ and } \Big|\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 - \sigma^2\Big| < \epsilon_2\Big\},$$

where $\epsilon = \max(\epsilon_1, \epsilon_2)$.

The following equality can be easily shown.

$$\frac{1}{2\delta^2}\sum_{i=1}^{n}(y_i - m)^2 = \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2 + C_1 + s(y),$$

where

$$C_1 = \frac{n(\mu - m)^2}{2\delta^2} + \frac{1}{2}\Big(\frac{1}{\delta^2} - \frac{1}{\sigma^2}\Big)(n-1)\sigma^2$$

and

$$s(y) = \frac{1}{2}\Big(\frac{1}{\delta^2} - \frac{1}{\sigma^2}\Big)\Big(\Big[\sum_{i=1}^{n}(y_i - \bar{y})^2 - (n-1)\sigma^2\Big] + n(\bar{y} - \mu)^2\Big) + \frac{(\mu - m)n}{\delta^2}(\bar{y} - \mu).$$

If $\epsilon < 1$, then, for any $(y_1, y_2, \ldots, y_n)^t \in G_\epsilon$, the following holds.

$$\left|s(y)\right| < \frac{1}{2}\Big(\frac{1}{\delta^2} - \frac{1}{\sigma^2}\Big)\Big((n-1)\epsilon_2 + n\epsilon_1^2\Big) + \frac{(\mu - m)}{\delta^2} n\epsilon_1 < C_2\epsilon,$$

where

$$C_2 = \Big\{\Big(\frac{1}{\delta^2} - \frac{1}{\sigma^2}\Big) + \frac{(\mu - m)}{\delta^2}\Big\} n\epsilon.$$

Thus, if any $(y_1, y_2, \ldots, y_n)^t \in G_\epsilon$, then

$$\exp\Big[-C_1 - C_2\epsilon\Big]\Big(\frac{\sigma^2}{\delta^2}\Big)^{n/2} \prod_{i=1}^{n} \phi(y_i \mid \mu, \sigma^2) < \prod_{i=1}^{n} \phi(y_i \mid m, \delta^2)$$

$$< \exp\Big[-C_1 + C_2\epsilon\Big]\Big(\frac{\sigma^2}{\delta^2}\Big)^{n/2} \prod_{i=1}^{n} \phi(y_i \mid \mu, \sigma^2).$$

These inequalities imply that

$$\lim_{\epsilon \to 0} \frac{\int \cdots \int_{G_\epsilon} \prod_{i=1}^{n} \phi(y_i \mid m, \delta^2) dy_1 \cdots dy_n}{\int \cdots \int_{G_\epsilon} \prod_{i=1}^{n} \phi(y_i \mid \mu, \sigma^2) dy_1 \cdots dy_n} = \Big(\frac{\sigma^2}{\delta^2}\Big)^{n/2} \exp\Big[-C_1\Big]$$

and

$$\lim_{\epsilon \to 0} \frac{\int \cdots \int_{G_\epsilon} \prod_{i=1}^{n} \phi(y_i \mid m, \delta^2) dy_2 \cdots dy_n}{\int \cdots \int_{G_\epsilon} \prod_{i=1}^{n} \phi(y_i \mid \mu, \sigma^2) dy_2 \cdots dy_n} = \Big(\frac{\sigma^2}{\delta^2}\Big)^{n/2} \exp\Big[-C_1\Big].$$

Since $Y_1, Y_2, \ldots, Y_n$ are *i.i.d.* normal random variables with mean m and variance δ^2, the conditional *p.d.f.* of y_1 given G is

$$\begin{aligned}
f(y_1 \mid G) &= \lim_{\epsilon \to 0} f(y_1 \mid G_\epsilon) \\
&= \lim_{\epsilon \to 0} \frac{\int \cdots \int_{G_\epsilon} \prod_{i=1}^{n} \phi(y_i \mid m, \delta^2) dy_2 \cdots dy_n}{\int \cdots \int_{G_\epsilon} \prod_{i=1}^{n} \phi(y_i \mid m, \delta^2) dy_1 \cdots dy_n} \\
&= \lim_{\epsilon \to 0} \frac{\int \cdots \int_{G_\epsilon} \prod_{i=1}^{n} \phi(y_i \mid \mu, \sigma^2) dy_2 \cdots dy_n}{\int \cdots \int_{G_\epsilon} \prod_{i=1}^{n} \phi(y_i \mid \mu, \sigma^2) dy_1 \cdots dy_n} \\
&= \lim_{\epsilon \to 0} \phi(y_1 \mid G_\epsilon) = \phi(y_1 \mid G).
\end{aligned}$$

Here $\phi(y_1 \mid G)$ is the conditional *p.d.f.* of Y_1 given G, where $Y_1, Y_2, \ldots, Y_n$ are *i.i.d.* normal random variables with mean μ and variance σ^2.

3. COMMENTS

Theorem 1 means that if $Y_1, Y_2, \ldots, Y_n$ are *i.i.d.* normal random variables with mean m and variance δ^2, the conditional *p.d.f.* of Y_1 given the sample mean and the sample variance does not depend on m and δ^2. Thus, even for finite sample cases, the sample mean and variance are more loyal to the underlying probability structure than the originally assumed mean and variance. Also, Theorem 1 implies the joint sufficiency of the sample mean and the sample variance under normality assumption, which is usually shown through the factorization theorem.

Using the law of large numbers and a local limit theorem, one can easily show that Theorem 1 implies the conditional *p.d.f.* of Y_1 given the sample mean μ and the sample variance σ^2 tends to $\phi(y \mid \mu, \sigma^2)$ regardless of the original mean m and variance δ^2. Since the limiting *p.d.f.* is the I-projection of the original *p.d.f.* $\phi(y \mid m, \delta^2)$ subject to $E(Y_1) = \mu$ and $\mathrm{Var}(Y_1) = \sigma^2$, Theorem 1 is a finite version of the Van Campenhout-Cover conditional limit theorem related to the Sanov large deviation problem. It supports the minimum discrimination information principle, i.e., to update the prior (or original) *p.d.f.* based on new information, the posterior (or updated) *p.d.f.* should be the closest in the Kullback-Leibler sense to the original one (Csiszár, Cover and Choi[1987, p. 801]).

REFERENCES

(1) Bartfai, P. (1972). On a Conditional Limit Theorem. *Progress in Statistics: Colloquia Mathematica Societatis János Bolyai*, **9**, 85–91.

(2) Csiszár, I., Cover, T.M., and Choi, B.S. (1987). Conditional Limit Theorems under Markov Conditioning. *IEEE Transactions on Information Theory*, **IT–33**, 788–801.

(**3**) Sanov, I.N. (1957). On the Probability of Large Deviations of Random Variables, in *IMS and AMS Selected Transactions in Mathematical Statistics and Probability*, **1**, 1961, 213–244.

(**4**) Van Campenhout, J.M. and Cover, T.M. (1981). Maximum Entropy and Conditional Probability. *IEEE Transactions on Information Theory*, **IT–27**, 483–489.

(**5**) Vasicek, O.A. (1980). A Conditional Law of Large Numbers. *The Annals of Probability*, **8**, 142–147.

(**6**) Vincze, I. (1972). On the Maximum Probability Principle in Statistical Physics. *Progress in Statistics: Colloquia Mathematica Societatis János Bolyai*, **9**, 869–893.