NECESSARY CONDITIONS FOR THE CONVERGENCE OF KULLBACK-LEIBLER'S MEAN INFORMATION

By Sadao Ikeda

(Received March 2, 1962)

Summary

Two types of necessary conditions are given for the convergence of Kullback-Leibler's mean information, one of which is connected with an asymptotic equivalence of two sequences of probability measures, and in special cases, with convergence of a sequence of probability distributions. The other is given in terms of the generalized probability density functions.

1. Introduction

Let (R, S, m) be a σ-finite measure space, and consider the family, $V(R, S, m)$, of all probability distributions defined on (R, S, m), which are absolutely continuous with respect to m. For any X and Y belonging to $V(R, S, m)$ with *gpdf.'s* $f(z)$ and $g(z)$, the Kullback-Leibler mean information is defined, in a generalized form, by

$$(1) \qquad I(X:Y)=\int_R f(z)\log\frac{f(z)}{g(z)}\,dm,$$

where no restriction is imposed on the inclusion relation between their carriers $D(X)$ and $D(Y)$ [1], i.e., for the following expression of (1)

$$(2) \qquad I(X:Y)=\left\{\int_{R-D(X)-D(Y)}+\int_{D(X)-D(Y)}+\int_{D(Y)-D(X)}+\int_{D(X)\frown D(Y)}\right\}f(z)\log\frac{f(z)}{g(z)}\,dm,$$

the first and the third integrals are always assumed to be zero, while the second one is assumed to be infinity if $m(D(X)-D(Y))>0$ and to be zero otherwise. It will easily be shown, for this generalized definition, that $I(X:Y)\geqq 0$ with equality when and only when $f(z)=g(z)$ $(a.e.\,m)$ on R.

Sufficient conditions for the convergence of the type as

$$I(X_i:Y_i)\rightarrow I(X:Y),\ (i\rightarrow\infty),$$

have been investigated under the situation that $D(X_i)=D(Y_i)$ up to a set of measure (m) zero for all i, and $D(X)=D(Y)$ up to a set of measure (m) zero [2], [3]. These results may easily be extended to the case of generalized definition of that information. As far as the present author is aware, any necessary condition has not been found yet, except the one by Kullback [3] in some special case, but his result seems to be invalidated by a false argument.

In the present paper, we shall derive two types of necessary conditions for the convergence of the information measure such as

$$(3) \qquad I(X_i : Y_i) \to 0 \qquad (i \to \infty),$$

with the generalized definition (1). The first of these, guarantees equivalence in the limit of two sequences $\{X_i\}(i=1, 2, \cdots)$ and $\{Y_i\}(i=1, 2, \cdots)$. Hence, in particular, if we take $X_i=X$ for all i, or $Y_i=Y$ for all i, then it relates to convergence of probability distributions, as will be seen in Corollaries 1 and 2, and in some examples in the final section. The second condition will be given by a sort of in-measure convergence of probability density functions, which will be shown in Theorem 2 and its corollaries.

2. Necessary conditions

We shall begin with the following

LEMMA 1. *Let $\{p_i(z)\}$ $(i=1, 2, \cdots)$ be a sequence of nonnegative, integrable (m) functions defined over a σ-finite measure space (R, S, m), and let $\{E_i\}$ $(i=1, 2, \cdots)$ be a sequence of measurable subsets of R such that*

$$\int_{E_i} p_i(z)\, dm \to 0 \qquad (i \to \infty).$$

Then it holds that, for any $\varepsilon>0$,

$$m(E_i \frown \{z ; p_i(z) \geqq \varepsilon\}) \to 0 \qquad (i \to \infty).$$

PROOF. The result follows immediately from the inequality

$$\int_{E_i} p_i(z)\, dm = \left\{\int_{E_i \frown \{z;p_i \geqq \varepsilon\}} + \int_{E_i \frown \{z;p_i < \varepsilon\}}\right\} p_i(z)\, dm$$
$$\geqq \varepsilon \cdot m(E_i \frown \{z ; p_i \geqq \varepsilon\}).$$

LEMMA 2. *Let $p(z)$ be a non-negative measurable function defined over R. Then it holds that*

$$(4) \qquad p(z)\log p(z)=p(z)-1+\frac{1}{2}(p(z)-1)^2/h(z),$$

where $h(z)$ *is a measurable function such that*

$$(5) \qquad \min\{1, p(z)\}\leqq h(z)\leqq\max\{1, p(z)\},$$

and, in particular, if $p(z)=0$, *then* $h(z)=1/2$.

The proof of this lemma will be omitted, since it is found in [4].

A necessary condition for (3) is given by the following

THEOREM 1. *For any sequences* $\{X_i\}$ $(i=1, 2, \cdots)$ *and* $\{Y_i\}$ $(i=1, 2, \cdots)$, *of the members of* $V(R, S, m)$, *with corresponding gpds.'s* $\{f_i(z)\}$, $(i=1, 2, \cdots)$, *and* $\{g_i(z)\}$ $(i=1, 2, \cdots)$, *the condition*

$$(6) \qquad I(X_i : Y_i)=\int_R f_i(z)\log\frac{f_i(z)}{g_i(z)}\,dm \to 0, \quad (i\to\infty),$$

implies that

$$(7) \qquad \int_R |f_i(z)-g_i(z)|\,dm \to 0 \quad (i\to\infty).$$

PROOF. Put, for each i,

$$A_i=\{z\,;f_i(z)=0,\ g_i(z)=0\},$$
$$B_i=\{z\,;f_i(z)>0,\ g_i(z)=0\},$$
$$C_i=\{z\,;f_i(z)=0,\ g_i(z)>0\},$$
$$D_i=\left\{z\,;f_i(z)>0,\ g_i(z)>0,\ 0<\frac{f_i(z)}{g_i(z)}<1\right\},$$
$$E_i=\left\{z\,;f_i(z)>0,\ g_i(z)>0,\ \frac{f_i(z)}{g_i(z)}=1\right\},$$
$$F_i=\left\{z:f_i(z)>0,\ g_i(z)>0,\ \frac{f_i(z)}{g_i(z)}>1\right\}.$$

Then, these subsets constitute an m-partition of R. Here we can assume, without any loss of generality, that $m(B_i)=0$ for $i=1, 2, \cdots$, because the condition (6) requires the finiteness of $I(X_i : Y_i)$'s for, at least, sufficiently large values of i.

By the remark on the definition (1) in the preceding section, $I(X_i : Y_i)$ becomes

$$I(X_i : Y_i)=\int_{C_i+D_i+E_i+F_i} f_i(z)\log\frac{f_i(z)}{g_i(z)}\,dm.$$

Hence, setting $d\mu_i=g_i dm$, and

$$p_i(z)=\begin{cases} f_i(z)/g_i(z), & \text{if } z\in C_i+D_i+E_i+F_i\,, \\ 0, & \text{otherwise,}\end{cases}$$

we get, by Lemma 2,

$$(8) \qquad I(X_i : Y_i)=\int_{C_i+D_i+F_i}\Big(p_i(z)-1+\frac{1}{2}(p_i(z)-1)^2/h_i(z)\Big)\,d\mu_i\,,$$

$$=\int_{C_i} d\mu_i+\Big\{\int_{D_i}+\int_{F_i}\Big\}\frac{1}{2}(p_i(z)-1)^2/h_i(z)\cdot d\mu_i\,.$$

Therefore, the condition (6) implies that

$$(9) \qquad \int_{C_i} d\mu_i \to 0\,,$$

$$(9)' \qquad \int_{D_i}(p_i(z)-1)^2/h_i(z)\cdot d\mu_i \to 0\,,$$

and

$$(9)'' \qquad \int_{F_i}(p_i(z)-1)^2/h_i(z)\cdot d\mu_i \to 0\,.$$

as $i\to\infty$.

Using the relation (5) in Lemma 2, we have the following:

$$(10) \qquad \int_{C_i} d\mu_i=\int_{C_i} g_i(z)\,dm\,,$$

$$(10)' \qquad \int_{D_i}(p_i(z)-1)^2/h_i(z)\cdot d\mu_i \geqq \int_{D_i}(p_i(z)-1)^2 g_i(z)\,dm,$$

and

$$(10)'' \qquad \int_{F_i}(p_i(z)-1)^2/h_i(z)\cdot d\mu_i \geqq \int_{F_i}(p_i(z)-1)^2\,\frac{g_i^2(z)}{f_i(z)}\,dm.$$

Now, we shall examine the integral of (7). Under the condition (6), it is easy to see that

$$(11) \qquad \int_R |f_i(z)-g_i(z)|\,dm=\int_{C_i+D_i+F_i}|f_i(z)-g_i(z)|\,dm.$$

Here, we have

$$\int_{C_i}|f_i(z)-g_i(z)|\,dm=\int_{C_i} g_i(z)\,dm,$$

and, using the Schwarz inequality,

$$\int_{D_i}|f_i(z)-g_i(z)|\,dm=\int_{D_i}|p_i(z)-1|\,g_i(z)\,dm$$

$$\leqq \sqrt{\int_{D_i} (p_i(z)-1)^2 g_i(z)\,dm \cdot \int_{D_i} g_i(z)\,dm}$$

$$\leqq \sqrt{\int_{D_i} (p_i(z)-1)^2 g_i(z)\,dm}.$$

Analogously, we get

$$\int_{F_i} |f_i(z)-g_i(z)|\,dm \leqq \sqrt{\int_{F_i} (p_i(z)-1)^2 \frac{g_i^2(z)}{f_i(z)}\,dm}.$$

Hence, it follows from (9), (9)′, (9)″, (10), (10)′, (10)″ and (11), that

$$\int_R |f_i(z)-g_i(z)|\,dm \to 0 \quad (i \to \infty),$$

which proves (7).

As immediate consequences of this result we obtain the following corollaries.

COROLLARY 1. *Let* $\{X_i\}$ $(i=1, 2, \cdots)$ *and* Y *be the members of* $V(R, S, m)$ *with gpdf.'s* $\{f_i(z)\}$ $(i=1, 2, \cdots)$ *and* $g(z)$. *Then, the condition*

(12) $$I(X_i : Y) \to 0 \quad (i \to \infty)$$

implies that

(13) $$\int_R |f_i(z)-g(z)|\,dm \to 0 \quad (i \to \infty).$$

COROLLARY 2. *Let* X *and* $\{Y_i\}$ $(i=1, 2, \cdots)$ *be the members of* $V(R, S, m)$ *with gpdf.'s* $f(z)$ *and* $\{g_i(z)\}$ $(i=1, 2, \cdots)$. *Then, the condition*

(14) $$I(X : Y_i) \to 0 \quad (i \to \infty)$$

implies that

(15) $$\int_R |f(z)-g_i(z)|\,dm \to 0 \quad (i \to \infty).$$

In these corollaries we are concerned with a sort of convergence of probability distributions. In fact, the conditions (12) and (14) can be regarded as criterions for convergences (13) and (15) of the corresponding probability distributions, respectively. Relation of these results to other convergence theorems, particularly to a useful convergence theorem due to Scheffé [5], will be investigated in the following section. The statistical meaning and application of the result of Theorem 1 will be given in another place.

Now, we shall state another necessary condition.

THEOREM 2. *Under the same situation as in Theorem 1, the condition*

(6) implies that, for any $\varepsilon>0$,

(16) $$m(\{z\,;|f_i(z)-g_i(z)|\geqq\varepsilon\})\to 0\ \ (i\to\infty).$$

The proof of this theorem is straight-forward from Theorem 1 and Lemma 1, and will be omitted. In fact, (16) is a necessary condition for (7).

Similarly, from corollaries 1 and 2 we obtain:

COROLLARY 3. *Under the same situation as in Corollary 1, the condition (12) implies that, for any* $\varepsilon>0$,

(17) $$m(\{z\,;|f_i(z)-g(z)|\geqq\varepsilon\})\to 0\ \ (i\to\infty).$$

COROLLARY 4. *Under the same situation as in Corollary 2, the condition (14) implies that, for any* $\varepsilon>0$,

(18) $$m(\{z\,;|f(z)-g_i(z)|\geqq\varepsilon\})\to 0\ \ (i\to\infty).$$

3. Convergence of probability distributions

In the present section we shall mainly be concerned with the result of Corollary 1. For the sake of convenience, we shall denote by (I) the condition $I(f_i:f)\to 0$ as $i\to\infty$.

A useful convergence theorem has been given by Scheffé [5], which states that convergence of *gpdf.*'s

(S) $$f_i(z)\to f(z) \text{ almost everywhere } (m),\ \ (i\to\infty),$$

implies the convergence of corresponding probability measures

(P) $$\int_R|f_i(z)-f(z)|dm\to 0,\ \ (i\to\infty),$$

which is called the "mean convergence", and is equivalent to

(U) $$\int_E f_i(z)\,dm\to\int_E f(z)\,dm \text{ uniformly in } E,\ \ (i\to\infty).$$

Corollary 1 of the present paper also states that (I) implies (P). As for the implication relation between the conditions (I) and (S) we can show the following:

THEOREM 3.

(i) *(S) is not necessarily stronger than (I), and*
(ii) *(I) is not necessarily stronger than (S).*

These results will be shown in turn by the following examples.

Example 1. Let (R, S, m) be a σ-finite measure space such that $m(R)=\infty$, and the range of m-measure, $M(S)=[0,\infty]$. Choose a sequence of measurable subsets $\{E_n\}$ $(n=2, 3, \cdots)$ with $m(E_n)=1/(n \log n)$ $(n=2, 3, \cdots)$. We shall define a simple function as follows:

$$f(z)=\begin{cases} 1/(\alpha\, n \log n) \text{ on } E_n \ (n=2, 3, \cdots), \\ 0 \text{ elsewhere,} \end{cases} \tag{19}$$

where $\alpha=\sum_{n=2}^{\infty} 1/(n \log n)^2$. Then this determines a *gpdf.* of a certain probability distribution belonging to $V(R, S, m)$.

On the other hand, we consider a sequence of *gpdf.*'s which are defined for $i \geqq 2$ by

$$f_i(z)=\begin{cases} 1/(\alpha\, 2 \log 2)-2\beta_i \log 2, \text{ on } E_2, \\ f(z) \text{ on } E_n \text{ for } 3 \leqq n \leqq i-1, \\ 1/(\alpha \log n) \text{ on } E_n \text{ for } n \geqq i, \\ 0 \text{ elsewhere,} \end{cases} \tag{20}$$

where $\beta_i=(1/\alpha)\sum_{n=i}^{\infty}(1-1/n)/(n(\log n)^2)$. It will be easy to see that $f_i(z)$ is a *gpdf.* of a certain member of $V(R, S, m)$ for each i, and β_i tends to zero as $i \to \infty$.

Since

$$\sup_R |f(z)-f_i(z)| \leqq \max\left(\beta_i,\ \left(1-\frac{1}{i}\right)/(\alpha \log i)\right),$$

it holds that $f_i(z) \to f(z)$ uniformly on R, as $i \to \infty$. Hence, the condition (S) is satisfied.

Now we have

$$\begin{aligned} I(f_i : f) &= \frac{1}{2 \log 2}\left(\frac{1}{\alpha\, 2 \log 2}-\beta_i\right)\log \frac{(1/\alpha\, 2 \log 2)-\beta_i}{1/\alpha\, 2 \log 2} \\ &\quad + \sum_{n=i}^{\infty} \frac{1}{n \log n} \frac{1}{\alpha \log n} \log \frac{1/\alpha \log n}{1/\alpha n \log n} \\ &= \frac{1-\alpha\beta_i\, 2 \log 2}{\alpha(2 \log 2)^2}\log(1-\alpha\beta_i\, 2 \log 2)+\frac{1}{\alpha}\sum_{n=i}^{\infty}\frac{1}{n \log n}. \end{aligned}$$

Hence, remembering that $\sum_{n=2}^{\infty} 1/(n \log n)$ is a divergent series, we obtain

$$I(f_i : f)=\infty, \ (i=2, 3, \cdots),$$

which means that the condition (I) is not satisfied.

Example 2. Let (R, S, m) be a finite measure space such that $M(S)=[0, 1]$, and consider a sequence of (m)-partitions, $\{Z_n\}$ $(n=1, 2, \cdots)$, such that

$$Z_n=\{A_{nk}\}, \ m(A_{nk})=1/n^2 \ (k=1, 2, \cdots, n^2;\ n=1, 2, \cdots).$$

For each A_{nk}, let $f_{nk}(z)$ be a function defined by

$$f_{nk}(z)=\begin{cases} n, \text{ if } z\in A_{nk}, \\ \dfrac{1-1/n}{1-1/n^2}, \text{ otherwise.} \end{cases} \tag{21}$$

Then we obtain a sequence of *gpdf.*'s, $\{f_{11}, f_{21}, f_{22}, f_{23}, f_{24}, f_{31}, f_{32}, \cdots\}$. Renumbering this sequence such that $f_1=f_{11}$, $f_2=f_{21}$, $f_3=f_{22}$, and so on, we have a sequence $\{f_i(z)\}$ $(i=1, 2, \cdots)$.

On the other hand, let $f(z)$ be a function which takes unity everywhere on R. Then it is obvious that the condition (S) is not fulfilled.

On the other hand, it holds, if i corresponds to some n, that

$$I(f_i : f)=\frac{\log n}{n}+\left(1-\frac{1}{n}\right)\log\frac{1-1/n}{1-1/n^2},$$

and, since $i\to\infty$ implies that $n\to\infty$, the condition (I) is satisfied.

Thus, the proof of Theorem 3 is complete.

An analogous result would be obtained on the implication relation between the condition (S) and $(I)'$ defined by

$$I(f : f_i)\to 0 \quad (i=\infty). \tag{$(I)'$}$$

The above theorem means that the conditions (S) and (I) are incomparable with each other. Any interesting example, to which the present result is effectively applicable, has not been found yet, but the usual convergence of probability distributions familiar to statistical analysis seems to be criticized by the condition (I) or $(I)'$, too, as will be seen in the following examples.

Example 3. It is well-known that the t-distribution with degrees of freedom n, tends to the standard normal distribution as $n\to\infty$. This is an in-the-mean convergence of the type (P), and the Scheffé theorem is easily applicable.

We shall show below that the condition $(I)'$ is also satisfied. The probability density function of the t-distribution with D.F. n is given by

$$f_n(x)=\frac{1}{\sqrt{n\pi}}\,\frac{\Gamma((n+1)/2)}{\Gamma(n/2)}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}, \quad (-\infty<x<\infty). \tag{22}$$

Denote by $f(x)$ the *pdf*. of the standard normal distribution. Then it is easily seen that

$$I(f : f_n)=-\log\frac{\sqrt{2}\,\Gamma((n+1)/2)}{\sqrt{n}\,\Gamma(n/2)}+\int_{-\infty}^{\infty} f(x)\left[\frac{n+1}{2}\log\left(1+\frac{x^2}{n}\right)-\frac{x^2}{2}\right]dx. \tag{23}$$

Using Stirling's formula, we have

$$\log\frac{\sqrt{2}\,\Gamma((n+1)/2)}{\sqrt{n}\,\Gamma(n/2)}\to 0\quad(n\to\infty).$$

It will be seen also that

$$\frac{n+1}{2}\log\left(1+\frac{x^2}{n}\right)\leqq\frac{n+1}{2n}x^2,$$

hence, from (23), it follows that

$$\lim_{n\to\infty} I(f:f_n)\leqq\lim_{n\to\infty}\int_{-\infty}^{\infty} f(x)\left[\frac{1}{2n}x^2\right]dx \tag{24}$$

$$=\lim_{n\to\infty}\frac{1}{2n}=0.$$

Since $I(f:f_n)\geqq 0$ for all n, (24) implies $(I)'$.

Example 4. The chi-square distribution with D.F. n has the *pdf*. such that

$$f_n(x)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma(n/2)}x^{(n/2)-1}\exp\left[-\dfrac{x}{2}\right] & \text{for } x\geqq 0,\\ 0 & \text{otherwise.}\end{cases} \tag{25}$$

It has been known that the mean is n and the variance is $2n$, and the standardized variable is distributed according to the standard normal distribution in the limit as $n\to\infty$, or the original chi-square variable is asymptotically normally distributed with mean n and variance $2n$.

In the present example, we shall examine the applicability of the result of Theorem 1.

Put, for the asymptotic distribution,

$$g_n(x)=\frac{1}{2\sqrt{\pi n}}\exp\left[-\frac{(x-n)^2}{4n}\right],\quad(-\infty<x<\infty). \tag{26}$$

Then the mean information becomes

$$I(f_n:g_n)=\int_0^{\infty} f_n(x)\log\frac{f_n(x)}{g_n(x)}\,dx \tag{27}$$

$$=\log\frac{2\sqrt{\pi n}}{2^{n/2}\Gamma(n/2)}+\int_0^{\infty} f_n(x)\left[\left(\frac{n}{2}-1\right)\log x-\frac{x}{2}+\frac{(x-n)^2}{4n}\right]dx.$$

Applying the Stirling formula, the first term is

$$\log\frac{2\sqrt{\pi n}}{2^{n/2}\Gamma(n/2)}\sim-\frac{n-2}{2}\log n+\frac{n}{2},$$

for sufficiently large values of n. It will easily be seen that

$$\left(\frac{n}{2}-1\right)\int_0^\infty f_n(x)\log x\,dx=\left(\frac{n}{2}-1\right)\left[\frac{\Gamma'(n/2)}{\Gamma(n/2)}+\log 2\right],$$

and, since it holds [6] that, for sufficiently large n,

$$\Gamma'\left(\frac{n}{2}\right)\Big/\Gamma\left(\frac{n}{2}\right)\sim\log(n/2)-1/n,$$

we get

$$\left(\frac{n}{2}-1\right)\int_0^\infty f_n(x)\log x\,dx\sim\left(\frac{n}{2}-1\right)[\log n-1/n],$$

for large n. Summarizing these results, we obtain from (27) the following

$$I(f_n:g_n)\sim 1/n,\ \text{for sufficiently large } n,$$

which shows that the condition (6) of Theorem 1 holds. Hence it follows that

$$\int_{-\infty}^{\infty}|f_n(x)-g_n(x)|\,dx\to 0\quad(n\to\infty).$$

Example 4. As an example for the discrete case, we shall consider the well-known result concerning convergence of binomial distributions to a limiting Poisson distribution. Put

$$f_n(x)=\binom{n}{x}p^x(1-p)^{n-x}\quad(x=0,1,2,\cdots,n),$$

and

$$f(x)=e^{-\lambda}\frac{\lambda^x}{x!}\quad(x=0,1,2,\cdots).$$

In this case, of course, the basic measure (R, S, m) is taken such that R is the set of all non-negative integers, S is the σ-field consisting of all subsets of R, and m is the counting measure on R.

It is known that $f_n(x)$ converges pointwise to $f(x)$ according to the limiting process such as $n\to\infty$, $np\to\lambda$ (fixed), i.e., that the Scheffé criterion (S) is satisfied. We shall show that the condition (I) is also satisfied.

Without any loss of generality, we can assume that $np=\lambda$ for all n. By the Chebycheff inequality we have, for any fixed θ $(0<\theta<1-p)$ such that $(\theta+p)n$ becomes an integer,

$$\sum_{x=(\theta+p)n}^{n} f_n(x) \leqq \lambda/\theta^2 n^2 .$$

Hence, using the Stirling formula, we obtain

$$\sum_{x=(\theta+p)n}^{n} f_n(x) \log \frac{n!}{(n-x)!} \leqq \sum_{x=(\theta+p)n}^{n} f_n(x) \log n!$$

$$\sim \sum_{x=(\theta+p)n}^{n} f_n(x)(\log \sqrt{2\pi}+(n+1/2)\log n-n)$$

$$\leqq \lambda(\log\sqrt{2\pi}+(n+1/2)\log n)/\theta^2 n^2 \to 0 \quad (n\to\infty).$$

On the other hand, using again Stirling's formula, we obtain for sufficiently large values of n,

$$\sum_{x=0}^{(\theta+p)n-1} f_n(x)\log\frac{n!}{(n-x)!} \leqq \sum_{x=0}^{n} f_n(x)\left[\left(n+\frac{1}{2}\right)\log n - x-\left(n-x+\frac{1}{2}\right)\log(n-x)\right]. \tag{28}$$

Since

$$(n-x+1/2)\log(n-x)=(n-x+1/2)\log n-\left(x-\frac{x^2}{n}+\frac{x}{2n}\right)\log(1-x/n)^{-n/x}$$

$$\geqq (n+1/2)\log n - x\log n-(x-x^2/n+x/2n),$$

the right-hand member of (28) is bounded to the above by $\lambda\log n+p/2$.

Now, the Kullback-Leibler mean information is evaluated as follows: for sufficiently large n, it holds that

$$I(f_n:f)=\sum_{x=0}^{n} f_n(x)\left[\log\frac{n!}{(n-x)!}+\lambda-x\log\lambda+x\log p+(n-x)\log(1-p)\right]$$

$$\leqq \lambda\log n+p/2+\lambda-np\log\lambda+np\log p+(n-np)\log(1-p),$$

from which, using the approximation $(n-\lambda)\log(1-p)\sim -\lambda+\lambda^2/n$, we get the condition (I).

DEPT. OF MATH., COLLEGE OF SCI. AND ENG., NIHON UNIV.

REFERENCES

[1] S. Ikeda, "An application of the discrimination information measure to the theory of testing hypotheses, Part II," *Ann. Inst. Stat. Math.*, Vol. 13 (1961), pp. 61–89.

[2] S. Ikeda, "A remark on the convergence of Kullback-Leibler's mean information," *Ann. Inst. Stat. Math.*, Vol. 12 (1960), pp. 81–88.

[3] S. Kullback, *Information Theory and Statistics*, John Wiley and Sons, 1959.

[4] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Camb. Univ. Press, 1934.

[5] H. Scheffé, "A useful convergence theorem for probability distributions," *Ann. Math. Stat.*, Vol. 18 (1947). pp. 434–438.

[6] H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946.