Ann. Inst. Statist. Math.
34 (1982), Part A, 209-224

SOME STRONG ε-EQUIVALENCE OF RANDOM VARIABLES

T. MATSUNAWA

(Received Jan. 14, 1982; revised Apr. 28, 1982)

Summary

Strongness and related error evaluations are investigated on type $(\boldsymbol{B})_d$, type (A) and type (I) ε-equivalence of random variables, which are based on Kolmogorov-Smirnov distance, a difference of random variables and Kullback-Leibler information number, respectively. As an application the Prohorov-LeCam type binomial-Poisson approximation problem is discussed and is given the best possible constant for the problem. Similar discussions are made on the negative binomial-Poisson approximation.

1. Introduction

Let $(\Omega, \boldsymbol{F}, P)$ be an underlying basic probability space and let R be any abstract space and $\boldsymbol{B}$ a σ-field of subsets of R. Denote the family of all $(R, \boldsymbol{B})$-random variables defined over $(\Omega, \boldsymbol{F}, P)$ by $\mathcal{F}(R, \boldsymbol{B})$. Further, designate by $\mathcal{P}(R, \boldsymbol{B}, \mu)$ the family of all $(R, \boldsymbol{B})$-random variables whose probability distributions are absolutely continuous with respect to a σ-finite measure μ on the measurable space $(R, \boldsymbol{B})$. Let X and Y be any two random variables belonging to $\mathcal{F}(R, \boldsymbol{B})$ and P^X and P^Y be their respective probability distributions on $(R, \boldsymbol{B})$.

Let us consider the following measures of discrepancy between the two distributions:

$$(1.1) \qquad D(X, Y; \boldsymbol{B}) \equiv \sup_{E \in \boldsymbol{B}} |P^X(E) - P^Y(E)|$$

and

$$(1.2) \qquad \Delta(X, Y; \boldsymbol{F}) \equiv \mathrm{P}\,(X \neq Y)\,.$$

Moreover, if X and Y are members of $\mathcal{P}(R, \boldsymbol{B}, \mu)$, the following quantities are also familiar measures of discrepancy:

$$(1.3) \qquad V(X, Y; R) = \int_R |f-g|\,d\mu \qquad [=2D(X, Y; \boldsymbol{B})]\,,$$

(1.4) $$I_f(X, Y; R)=\int_R f \ln \frac{f}{g} d\mu ,$$

and

(1.4)′ $$I_g(X, Y; R)=\int_R g \ln \frac{g}{f} d\mu ,$$

where f and g are gpdf's $[\mu]$ of P^X and P^Y, respectively.

Now, we define the three types of ε-equivalence of the random variables:

DEFINITION 1.1. The two random variables X and Y belonging to $\mathcal{F}(R, \boldsymbol{B})$ are said to be *ε-equivalent in the sense of type $(\boldsymbol{B})_d$* [or *type $(\boldsymbol{B})_d$ ε-equivalent*] and are denoted by

(1.5) $$X \dot{\approx} Y \quad (\boldsymbol{B})_d ,$$

if it holds that

(1.6) $$D(X, Y; \boldsymbol{B}) \leqq \varepsilon .$$

DEFINITION 1.2. The random variables X and Y belonging to $\mathcal{F}(R, \boldsymbol{B})$ are said to be *ε-equivalent with respect to the difference $\Delta(X, Y; \boldsymbol{F})$* [or *type (A) ε-equivalent*] and are denoted by

(1.7) $$X \dot{\approx} Y \quad (A) ,$$

if it holds that

(1.8) $$\Delta(X, Y; \boldsymbol{F}) \leqq \varepsilon .$$

DEFINITION 1.3. The random variables X and Y belonging to $\mathcal{P}(R, \boldsymbol{B}, \mu)$ are said to be *ε-equivalent with respect to the K-L information* [or *type (I) ε-equivalent*] and are denoted by

(1.9) $$X \dot{\approx} Y \quad (I) ,$$

if it holds that

(1.10) $$I(X, Y; R) \equiv \min \{I_f(X, Y; R),\ I_g(X, Y; R)\} \leqq \varepsilon .$$

It is of interest to compare the strongness of the above approximate equivalences and to give quantitative error evaluations for them.

Remark 1.1. We can also define other types of ε-equivalence based on the affinity, the W-divergence and so on. However, the above three types of approximate equivalence seem to be more important from the practical point of view.

Remark 1.2. The concept $[X \dot{\approx} Y \quad (\boldsymbol{B})_d]$ is closely related to that of $[X \overset{\varepsilon}{\sim} Y \ \ (\boldsymbol{B})_d]$ defined in Matsunawa [11], which was introduced to give necessary and sufficient conditions for uniform approximate equivalence between X and Y based on some information type measures of discrepancy. It should be noted, however, that the subject in the former is to find the bound ε as small as possible, whereas in the latter to estimate $\phi=\phi(\varepsilon\,;\,\delta^*)$ sharply becomes main concern under the situation that $\delta^*\leqq\varepsilon$ for any given $\varepsilon>0$. It should be also remarked that the type $(\boldsymbol{B})_d$ ε-equivalence is a stronger notion than that of ε-coincidence between two cumulative distribution functions defined by Meshalkin [12].

2. Inclusion relation of type $(\boldsymbol{B})_d$ and type (A) ε-equivalence

The concept of the type $(\boldsymbol{B})_d$ ε-equivalence is meaningful in such cases where we wish to approximate the distribution P^X of a discrete (resp. a continuous) type random variable by another distribution P^Y of a discrete (resp. a continuous) type random variable Y. On the other hand the type (A) ε-equivalence of random variables would be useful in various stochastic approximation problems. Two types of ε-equivalence stated above are of different natures. The former evaluates the closeness of two probability measures, while the latter that of two random variables as measurable transformations.

We shall now prove the following

THEOREM 2.1. *Type (A) ε-equivalence is stronger than type $(\boldsymbol{B})_d$, i.e., for any random variables X and Y belonging to $\mathscr{F}(R, \boldsymbol{B})$, and for any $\varepsilon>0$, it holds that*

$$X \overset{\varepsilon}{\approx} Y \quad (A) \Rightarrow X \overset{\varepsilon}{\approx} Y \quad (\boldsymbol{B})_d\,. \tag{2.1}$$

PROOF. It suffices to show that

$$D(X, Y; \boldsymbol{B}) \leqq \Delta(X, Y; \boldsymbol{F})\,. \tag{2.2}$$

Let $A=\{\omega;\, X(\omega)\neq Y(\omega)\}$. Then, for any X and Y, belonging to $\mathscr{F}(R, \boldsymbol{B})$, for any measurable set $E\in\boldsymbol{B}$, it holds that

$$\begin{aligned} P^X(E)-P^Y(E) &= \mathrm{P}\,(X^{-1}(E))-\mathrm{P}\,(Y^{-1}(E)) \\ &= \mathrm{P}\,((X^{-1}(E)\cap A)\cup(X^{-1}(E)\cap A^c)) \\ &\quad -\mathrm{P}\,((Y^{-1}(E)\cap A)\cup(Y^{-1}(E)\cap A^c)) \\ &= \mathrm{P}\,(X^{-1}(E)\cap A)-\mathrm{P}\,(Y^{-1}(E)\cap A) \\ &\leqq \max\,(\mathrm{P}\,(X^{-1}(E)\cap A),\ \mathrm{P}\,(Y^{-1}(E)\cap A)) \\ &\leqq \mathrm{P}\,(A)\,, \end{aligned} \tag{2.3}$$

from which we get (2.2).

Remark 2.1. Weaker versions of (2.1) are useful, too. Let $F(x)$ and $G(x)$ be cdf.'s of random variables X and Y, respectively. One of the versions is

$$\sup_x |F(x)-G(x)| \leqq \mathrm{P}(X \neq Y) , \tag{2.4}$$

which is found in Hodges and LeCam [2]. They evaluated the right hand probability for the Poisson approximation problem to Poisson-binomial distribution.

Instead of (2.4) we are sometimes required a directed evaluation as

$$F(x)-G(x) \leqq \mathrm{P}(X<Y) \quad [\leqq \mathrm{P}(X \neq Y)] ,$$

for any real x, which is obtained in a similar manner as in the proof of the above theorem. To estimate the probability $\mathrm{P}(X<Y)$ often occurs. For instance, let X be a stress loaded to a material with strength Y. Then, the material has no failure whenever $X<Y$, and the estimation of the probability is of interest in the field of reliability theory. Here, assume that the two random variables are independent and that $a \leqq X \leqq b$ and $c \leqq Y \leqq d$, where a, b, c and d are extended real numbers. Then, we can represent the probability as

$$\begin{aligned} \mathrm{P}(X<Y) &= \int_c^d \left[\int_a^{y_-} dF(x) \right] dG(y) = \int_c^d [F(y_-)-F(a)] dG(y) \\ &= \int_a^b \left[\int_{x_+}^d dG(y) \right] dF(x) = \int_a^b [1-G(x_+)] dF(x) . \end{aligned}$$

Let us consider an example. Suppose X and Y are independently distributed according to $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$, respectively, then

$$\mathrm{P}(X<Y) = \Phi((\mu_Y - \mu_X)/\sqrt{\sigma_Y^2+\sigma_X^2}) ,$$

Φ being the distribution function of the standard normal distribution $N(0, 1)$ (cf. Church and Harris [1]). Of course, in this case the RHS term of (2.5) in the brackets is meaningless as a bound because of $\mathrm{P}(X \neq Y)=1$.

Next, we investigate the implication relation between two types of ε-equivalence, type (A) and type (I). Let X and Y be in $\mathcal{P}(R, \boldsymbol{B}, \mu)$. The following theorem shows that the above two types of ε-equivalence are incomparable.

THEOREM 2.2. (i) *type* (A) *is not necessarily stronger than type* (I), *and* (ii) *type* (I) *is not stronger than type* (A).

PROOF. (i) To prove the statement we shall consider the following

Example 2.1. Assume that $(\Omega, \boldsymbol{F}, P)$ is the Wiener probability space, namely, $\Omega=[0, 1)$, $\boldsymbol{F}$ is the σ-field of Borel subsets of Ω and P is the usual Lebesgue measure defined only for sets belonging to $\boldsymbol{F}$. Let τ be any given number such that $0<\tau<0.8$ and let J_i $(i=1, 2, 3)$ be the intervals defined by $J_1=[0, 1-\tau^2\{1+\exp(-1/\tau^3)\})$, $J_2=[1-\tau^2\{1+\exp(-1/\tau^3)\}, 1-\tau^2\exp(-1/\tau^3))$ and $J_3=[1-\tau^2\exp(-1/\tau^3), 1)$. Further, let $p_1=1-\tau^2\{1+\exp(-1/\tau^3)\}$, $p_2=\tau^2$, $p_3=\tau^2\exp(-1/\tau^3)$ and a_i $(i=1, 2, 3)$ be some positive constants.

Under the above set-up consider the following random variables defined by

$$X(\omega)\equiv\begin{cases} a_1/p_1\cdot\omega\,, & \text{if } \omega\in J_1\,,\\ a_1+a_2/p_2\cdot(\omega-p_1)\,, & \text{if } \omega\in J_2\,,\\ a_1+a_2+a_3/p_3\cdot(\omega-p_1-p_2)\,, & \text{if } \omega\in J_3\,,\end{cases}$$

and

$$Y(\omega)\equiv\begin{cases} a_1/p_1\cdot\omega\,, & \text{if } \omega\in J_1\,,\\ a_1+a_2+a_3/p_2\cdot(\omega-p_1)\,, & \text{if } \omega\in J_2\,,\\ a_1+a_2/p_3\cdot(\omega-p_1-p_2)\,, & \text{if } \omega\in J_3\,.\end{cases}$$

Thus, the pdf.'s of X and Y are respectively given by

$$f(x)=\begin{cases} [1-\tau^2\{1+\exp(1-1/\tau^3)\}]/a_1\,, & \text{if } 0\leqq x<a_1\,,\\ \tau^2/a_2\,, & \text{if } a_1\leqq x<a_1+a_2\,,\\ \tau^2\exp(-1/\tau^3)/a_3\,, & \text{if } a_1+a_2\leqq x<a_1+a_2+a_3\,,\\ 0\,, & \text{otherwise}\,,\end{cases}$$

and

$$g(x)=\begin{cases} f(x)\,, & \text{if } 0\leqq x<a_1\,,\\ \tau^2\exp(-1/\tau^3)/a_2\,, & \text{if } a_1\leqq x<a_1+a_2\,,\\ \tau^2/a_3\,, & \text{if } a_1+a_2\leqq x<a_1+a_2+a_3\,,\\ 0\,, & \text{otherwise}\,.\end{cases}$$

Then, it is easily seen that

$$\mathrm{P}(X\neq Y)\leqq\mathrm{P}(J_2)+\mathrm{P}(J_3)=p_2+p_3=\tau^2\{1+\exp(-1/\tau^3)\}<2\tau^2\,,$$

$$\begin{aligned} I_f(X, Y; R)&=\int_{-\infty}^{\infty} f\ln(f/g)dx\\ &=\int_{a_1}^{a_1+a_2}\tau/a_1\cdot\frac{1}{\tau^3}dx+\int_{a_1+a_2}^{a_1+a_2+a_3}\tau\exp(-1/\tau^3)/a_3\cdot(-1/\tau^3)dx\end{aligned}$$

$$=(1/\tau^2)\{1-\exp(-1/\tau^3)\}>1/(2\tau^2)$$

and

$$I_g(X, Y; R)=\int_{-\infty}^{\infty} g \ln (g/f)dx=(1/\tau^2)\{1-\exp(-1/\tau^3)\}>1/(2\tau^2)\,.$$

Since τ is an arbitrary positive number such that $0<\tau<0.8$, the above results prove the assertion (i).

(ii) To prove the statement (ii) we shall use the following

Example 2.2. Suppose that $(\Omega, \boldsymbol{F}, P)$ is the Wiener probability space as in the preceding example. Let τ be any given number such that $0<\tau<1$. Let I_i, J_i $(i=1, 2)$ be subintervals of Ω respectively defined by $I_1=[0, 1-\tau^2)$, $I_2=[1-\tau^2, 1)$, $J_1=[0, 1-\tau^3)$ and $J_2=[1-\tau^3, 1)$. Then, $\mathrm{P}(I_1)=1-\tau^2\equiv p_1$, $\mathrm{P}(I_2)=\tau^2\equiv p_2$, $\mathrm{P}(J_1)=1-\tau^3\equiv q_1$, $\mathrm{P}(J_2)=\tau^3\equiv q_2$.

Under the above situation let us consider the following real random variables defined by

$$X(\omega)\equiv\begin{cases} (1-\tau)/p_1\cdot\omega\,, & \text{if } \omega\in I_1\,,\\ 1-\tau+\tau/p_2\cdot(\omega-p_1)\,, & \text{if } \omega\in I_2\,,\end{cases}$$

and

$$Y(\omega)\equiv\begin{cases} (1-\tau)/q_1\cdot\omega\,, & \text{if } \omega\in J_1\,,\\ 1-\tau+\tau/q_2\cdot(\omega-q_1)\,, & \text{if } \omega\in J_2\,.\end{cases}$$

Thus, the pdf.'s of X and Y are respectively given by

$$f(x)=\begin{cases} 1+\tau\,, & \text{if } 0\leqq x<1-\tau\,,\\ \tau\,, & \text{if } 1-\tau\leqq x<1\,,\\ 0\,, & \text{otherwise}\,,\end{cases}$$

and

$$g(x)=\begin{cases} 1+\tau+\tau^2\,, & \text{if } 0\leqq x<1-\tau\,,\\ \tau^2\,, & \text{if } 1-\tau\leqq x<1\,,\\ 0\,, & \text{otherwise}\,.\end{cases}$$

Then,

$$\mathrm{P}(X\neq Y)=1-\mathrm{P}(X=Y)=1-0=1\,,$$

$$I_f(X, Y; R)=\int_{[0,1-\tau)}(1+\tau)\ln\frac{1+\tau}{1+\tau+\tau^2}d\mu+\int_{[1-\tau,1)}\tau\ln\frac{\tau}{\tau^2}d\mu$$
$$=(1-\tau^2)\ln\left(1-\frac{\tau^2}{1+\tau+\tau^2}\right)+\tau^2\ln\frac{1}{\tau}<\tau^2\ln\frac{1}{\tau}<\tau-\tau^2$$

and

$$I_g(X, Y; R)=\int_{[0,1-\tau)}(1+\tau+\tau^2)\ln\frac{1+\tau+\tau^2}{1+\tau}d\mu+\int_{[1-\tau,1)}\tau^2\ln\frac{\tau^2}{\tau}d\mu$$
$$=(1-\tau^3)\ln\left(1+\frac{\tau^2}{1+\tau}\right)+\tau^2\ln\tau<2\tau^3+\tau^4 .$$

Since τ is an arbitrary positive number, such that $0<\tau<1$, it holds that $X\dot{\approx}Y$ (I) with $\varepsilon=\min(\tau-\tau^2, 2\tau^3+\tau^4)$ but $X\dot{\approx}Y$ (A) does not hold, which completes the proof of the theorem.

3. A quantitative evaluation of the type $(\boldsymbol{B})_d$ ε-equivalence by the K-L information number and its applications

In the practical situations direct calculations of the distance $D(X, Y; \boldsymbol{B})$ are often difficult, in which cases it is sometimes required to approximate the quantity as accurately as possible. It is seen that $(\boldsymbol{B})_d$ ε-equivalence is a weaker concept than type (I) approximate one (cf. Matsunawa [11]). But, the K-L information number $I_f(X, Y; R)$ (or $I_g(X, Y; R)$) is easy to calculate for a fairly wide class of distributions. Moreover, the information number is useful to give a considerably good upper bound for $D(X, Y; \boldsymbol{B})$ as follows;

THEOREM 3.1. *Let* $I=\min[I_f(X, Y; R), I_g(X, Y; R)]$ *then it holds that*

(3.1) $$D(X, Y; \boldsymbol{B})\leqq\min[\eta(I), \zeta(I)]$$

where

(3.2) $$\eta(I)=[1-\exp(-I)]^{1/2}$$

and

(3.3) $$\zeta(I)=[a(I)+b(I)-175/264]^{1/2}\ (\leqq\sqrt{I/2}) ,$$

where

(3.4) $$a(I)=[q(I)+\sqrt{q^2(I)+r^3}]^{1/3} ,\qquad b(I)=[q(I)-\sqrt{q^2(I)+r^3}]^{1/3}$$

with

(3.5) $$r=177275/69696 ,\qquad q(I)=1575I/704+49214375/18399744 .$$

The equality in (3.1) *holds if and only if* $f=g$ *(a.e.* μ*), in which case* $D(X, Y; \boldsymbol{B})=0$.

PROOF. The bound $\eta(I)$ is the same one given in Lemma 2.1 in Matsunawa [8]. To derive the upper bound $\zeta(I)$ we can use the Kraft-

Schmitz inequality [6],

$$x \ln \frac{x}{y} + (1-x) \ln \frac{1-x}{1-y} \geqq 2(x-y)^2 + \frac{4}{9}(x-y)^4 + \frac{352}{1575}(x-y)^6$$

for $0<x<1$ and $0<y<1$. By the same manner as in [8], it can be seen that

$$I(X, Y; R) \geqq 2[D(X, Y; \boldsymbol{B})]^2 + \frac{4}{9}[D(X, Y; \boldsymbol{B})]^4 + \frac{352}{1575}[D(X, Y; \boldsymbol{B})]^6 .$$

Therefore, it remains to solve the inequality in $D(X, Y; \boldsymbol{B})$:

$$D^6 + (175/88)D^4 + (1575/176)D^2 - (1575/352)I \leqq 0 , \qquad (0 \leqq D < 1) ,$$

which is equivalent to

$$h(d) = d^3 + 3rd - 2q(I) \leqq 0 , \qquad (175/264 \leqq d \equiv D^2 - 175/264 < 439/264) ,$$

where r and $q(I)$ are quantities given in (3.5). Since $r^3+q^2>0$, the equation $h(d)=0$ has only one real root $d_0 \equiv d_0(I) = a(I) + b(I)$ which is easily seen less than 175/264. Then, $d \leqq \min(d_0(I), 175/264)$ and hence $D \leqq \min(d_0(I) - 175/264, 1)$. Here, noticing the fact that $\eta(I) \leqq 1$ we have $D \leqq \min[\eta(I), \zeta(I)]$, which completes the proof of the theorem.

The above theorem is useful to the problems of error evaluation in the sense of type $(\boldsymbol{B})_d$ approximation between two probability distributions, if a related K-L information number is evaluated sharply. Concerning this kind of problems Vervaat [15] along the line of Ikeda's work [3] gave some error evaluation essentially in the above $(\boldsymbol{B})_d$ sense between the binomial or negative binomial distribution and the Poisson distribution by estimating the information numbers accurately. He, however, resorted to a less inferior bound than ours in (3.1), so his results can be improved as shown below.

Let for $k, n = 0, 1, 2, \cdots$, $a>0$, $0<p<1$, $p+q=1$, $\lambda>0$,

$$b_k(n, p) = \binom{n}{k} p^k q^{n-k} ,$$

$$c_k(a, q) = \binom{a-1+k}{k} p^a q^k ,$$

$$p_k(\lambda) = e^{-\lambda} \frac{\lambda^k}{k!} .$$

In this case the underlying measure space $(R, \boldsymbol{B}, \mu)$ should be understood such that R is the set of all nonnegative integers, $\boldsymbol{B}$ is the σ-field of all subsets of R, and μ is the counting measure on R. For comparisons with the well known results [2], [5], [7], [13], [14], [15] we shall evaluate the variational distance $V(\cdot, \cdot; R)$ defined by (1.3) which is equivalent to $2D(\cdot, \cdot; \boldsymbol{B})$.

THEOREM 3.2 (*Binomial-Poisson approximation*). *Let*

$$V_1(n, p) \equiv \sum_{k=0}^{\infty} |b_k(n, p) - p_k(np)| . \tag{3.6}$$

Then, for $0<p<1$ *it holds that*

$$V_1(n, p) \leqq 2 \min [\eta(I_1^*), \zeta(I_1^*)] , \tag{3.7}$$

where

$$\begin{aligned} I_1 &= \sum_{k=0}^{\infty} b_k(n, p) \ln [b_k(n, p)/p_k(np)] \\ &\leqq \min\left\{\frac{p^2}{2(1-p)},\ -\frac{1}{2}\ln(1-p) - \frac{p(1-p)(71-98p+32p^2)}{6(3-2p)^3}\right\} \\ &\equiv I_1^* , \end{aligned} \tag{3.8}$$

and asymptotically

$$I_1 \sim -\frac{1}{2}\ln(1-p) - \frac{p}{2} \equiv I_{1A} , \tag{3.9}$$

provided that $nq \to \infty$ *as* $n \to \infty$.

From this result, we can give a positive answer to a conjecture due to Vervaat [15]:

COROLLARY 3.1. *For* $0 \leqq p \leqq 1$ *it holds that*

$$V_1(n, p) \leqq 2p , \tag{3.10}$$

where the equality sign hold for $p=0$ *and* $p=1$ *with* $n=k \to \infty$. *The constant* 2 *in* (3.10) *is the best possible one.*

PROOF OF THEOREM 3.2. The bound $p^2/\{2(1-p)\} = I_{1V}^*$ in (3.8) was given by Vervaat [15]. Thus, we shall prove another bound in the inequality. We have

$$I_1 = nq \ln q + np + \sum_{k=2}^{n} \binom{n}{k} p^k q^{n-k} \ln \frac{n!}{(n-k)!\,n^k} . \tag{3.11}$$

Now let us evaluate the summation term sharply.

Using a modified Stirling formula of the form (cf. Matsunawa [9], [10]):

$$\ln \Gamma(x+1)=\frac{1}{2}\ln 2\pi+\left(x+\frac{1}{2}\right)\ln(x+1)-(x+1)+\frac{1}{12(x+1)}-R(x) \qquad (x>-1) \tag{3.12}$$

with

$$0<\frac{1}{360(x+1)(x+2)(x+3)}<R(x) \tag{3.13}$$
$$<\frac{1}{360(x+1)(x+2)(x+3)}+\frac{1}{32(x+1)^2(x+2)(x+3)}<\frac{49}{8640(x+1)}$$

we can see that

$$\ln\frac{n!}{(n-k)!n^k}<\left(n+\frac{1}{2}\right)\ln\left(1+\frac{1}{n}\right)-\left(n-k+\frac{1}{2}\right)\ln\left(1-\frac{k-1}{n}\right)-k\equiv u(k),$$

for $2\leqq k\leqq n$. Next, we must evaluate the quantity $u(k)$. To this end let us introduce the function $u(t)$ defined on the interval $2\leqq t\leqq n$ by changing k of $u(k)$ into t. It is easily calculated that

$$u'(t)=\ln\left(1-\frac{t-1}{n}\right)-\frac{1}{2}(n-t+1)^{-1}$$

$$u''(t)=-(n-t+1)^{-1}-\frac{1}{2}(n-t+1)^{-2}<0$$

$$u'''(t)=-(n-t+1)^{-2}-(n-t+1)^{-3}<0$$

$$u''''(t)=-2(n-t+1)^{-3}-3(n-t+1)^{-4}<0,$$

then by the generalized Jensen inequality (cf. Vervaat [15]) we have

$$\mathrm{E}\,[u(k)]\equiv\sum_{k=2}^{n}\binom{n}{k}p^kq^{n-k}u(k) \tag{3.14}$$
$$\leqq u(\mathrm{E}\,(k))+\frac{u''(\mathrm{E}\,(k))}{2}\,\mathrm{Var}\,(k)+\frac{u'''(\mathrm{E}\,(k))}{6}\,\mathrm{E}\,(x-\mathrm{E}\,(k))^3$$
$$=\left(n+\frac{1}{2}\right)\ln\left(1+\frac{1}{n}\right)-\left(nq+\frac{1}{2}\right)\ln\left(q+\frac{1}{n}\right)-np$$
$$+\frac{npq}{2}\left\{-\frac{1}{nq+1}-\frac{1}{2(nq+1)^2}\right\}$$
$$+\frac{npq(q-p)}{6}\left\{-\frac{1}{(nq+1)^2}-\frac{1}{(nq+1)^3}\right\}$$
$$\leqq -nq\ln q-np-\frac{1}{2}\ln q-\left\{\frac{1}{2}-w(z;q)\right\}p$$

where we have put $z=nq$ and

$$w(z;q)=\frac{1}{2(z+1)}-\frac{\{3+2(q-z)\}z}{12(z+1)^2}-\frac{(q-p)z}{6(z+1)^3}=\frac{(5-4q)z^2+(13-8q)z+6}{12(z+1)^3}.$$

Assume now that q be temporarily fixed number. Since

$$\frac{\partial w(z;q)}{\partial z}=-\frac{(5-4q)z^2+(16-8q)z+(5+8q)}{12(z+1)^3}<0$$

then, $w(z;q)$ is a monotone decreasing function of z ($\geqq 2q>0$). Hence,

(3.15) $$w(z,q)\leqq w(2q,q)=\frac{3+13q+2q^2-8q^3}{6(2q+1)^3}$$

Combining (3.11), (3.14) and (3.15), we have the desired result

$$I_1\leqq -\frac{1}{2}\ln(1-p)-\frac{(1-p)(71-98p+32p^2)}{6(3-2p)^3}p\equiv I_{1M}^*, \qquad (0<p<1).$$

As for (3.9), since $w(z;q)\to 0$ under the condition $z=nq\to\infty$ as $n\to\infty$, then from (3.11) and (3.14) we have

$$I_1\lesssim -\frac{1}{2}\ln(1-p)-\frac{p}{2}+O\left(\frac{1}{nq}\right).$$

The reverse side inequality can be similarly shown. Thus, we complete the proof of the theorem.

PROOF OF COROLLARY 3.1. To prove the inequality $V_1(n,p)<2p$ $(0<p<1)$ we use Figure 3.1, where the curves

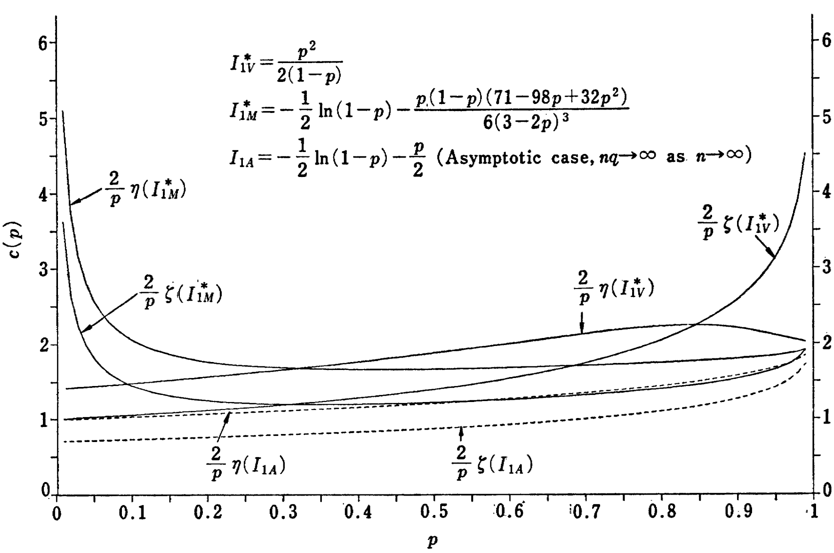

Fig. 3.1. $c(p)$ for $V(n,p)\leqq c(p)\cdot p$

$$c(p)\equiv\frac{1}{p}\min\{\eta(I_1),\ \zeta(I_1)\}\quad\left[\geqq\frac{V_1(n,\,p)}{p}\right]$$

are plotted on the p-$c(p)$ plane. Noticing the facts that $\lim_{p\to 0_+} c(p)=1$ and $\lim_{p\to 1_-} c(p)=2$, we can see from the figure at once that $c(p)<2$ for $0<p<1$. For $p=0$ it is directly calculated that $V_1(n,\,0)=0=2p$. Further, since it is always true that $V_1(n,\,p)\leqq 2$ for any p $(0\leqq p\leqq 1)$, then $V_1(n,\,1)<2=2p$ automatically holds. In this case we can also see that $V_1(n,\,1)=2(1-e^{-n}n^n/n!)\to 2$ as $n\,(=k)\to\infty$. So, the constant 2 cannot be improved without extra condition on p. Thus, the proof of the corollary is completed.

We can also evaluate the Poisson-approximation to the negative binomial distribution along the same line as the approximation to the binomial distribution discussed above. We have the following result:

THEOREM 3.3 (*Negative binomial-Poisson approximation*). *Let*

$$V_2(a,\,q)\equiv\sum_{k=0}^{\infty}|c_k(a,\,q)-p_k(aq/p)|\,. \tag{3.16}$$

Then, for $0<p<1$ *it holds that*

$$V_2(a,\,q)\leqq 2\min[\eta(I_2),\ \zeta(I_2)]\,, \tag{3.17}$$

where $I_2=\min(I_c;\,I_p)$ *with*

$$\begin{aligned} I_c=I_c(c_k,\,p_k)&=\sum_{k=0}^{\infty}c_k(a,\,q)\ln[c_k(a,\,q)/p_k(aq/p)]\\ &\leqq\frac{q^2}{4p^2}\left\{1+\frac{2}{3}aq-\frac{1}{3(a+2)}\right\}<\frac{q^2}{4p^2}\left(1+\frac{2}{3}aq\right) \end{aligned} \tag{3.18}$$

and

$$\begin{aligned} I_p=I_p(c_k,\,p_k)&=\sum_{k=0}^{\infty}p_k(aq/p)\ln[p_k(aq/p)/c_k(a,\,q)]\\ &\leqq-\frac{1}{2}\ln(1-q)-\frac{q}{2}+\frac{61q^2}{768a}<\frac{q^2}{p^2}\left(1+\frac{61}{768a}\right). \end{aligned} \tag{3.19}$$

COROLLARY 3.2. *For* $0<q<1$ *it holds that*

$$V_2(a,\,q)<\sqrt{2}\,\frac{q}{p}\,. \tag{3.20}$$

PROOF OF THEOREM 3.3. We have

$$I_c=\frac{a}{p}\ln p+\frac{aq}{p}+\sum_{k=2}^{\infty}\binom{a-1+k}{k}p^a q^k\ln\frac{(a-1+k)!}{(a-1)!a^k}\,. \tag{3.21}$$

The above summation part can be evaluated as

$$(3.22)\quad \underset{c}{\mathrm{E}}\left[\sum_{\nu=1}^{k-1}\ln\left(1+\frac{\nu}{a}\right)\right]\leqq \underset{c}{\mathrm{E}}\left[\sum_{\nu=1}^{k-1}\frac{\nu}{a}\left\{1-\frac{1}{2(a+1)}-\frac{1}{6(a+1)(a+2)}\right\}\right]$$
$$=\frac{1}{2a}\left\{1-\frac{1}{2(a+1)}-\frac{1}{6(a+1)(a+2)}\right\}\underset{c}{\mathrm{E}}\,[k(k-1)]$$
$$=\frac{(a+1)q^2}{2p^2}\left\{1-\frac{1}{2(a+1)}-\frac{1}{6(a+1)(a+2)}\right\}.$$

Hence, from (3.21) and (3.22), it follows that

$$I_c\leqq\frac{q^2}{2p^2}-\frac{q^2}{4p^2}-\frac{q^2}{12(a+2)p^2}+\frac{a}{p}\left(\ln p+q+\frac{q^2}{2p}\right)$$
$$=\frac{q^2}{4p^2}-\frac{q^2}{12(a+2)p^2}+\frac{a}{p}\sum_{\nu=3}^{\infty}\left(\frac{1}{2}-\frac{1}{\nu}\right)q^\nu$$
$$\leqq\frac{q^2}{4p^2}-\frac{q^2}{12(a+2)p^2}+\frac{a}{6p}\sum_{\nu=3}^{\infty}q^\nu\leqq\frac{q^2}{4p^2}\left\{1+\frac{2}{3}aq-\frac{1}{3(a+2)}\right\}$$
$$\leqq\frac{q^2}{4p^2}\left(1+\frac{2}{3}aq\right),$$

which proves (3.18) and is an improvement of Vervaat's result. Now, we proceed to prove (3.19). We have

$$(3.23)\quad I_p=-\frac{a}{p}\ln p-\frac{aq}{p}-\sum_{k=2}^{\infty}e^{-aq/p}\frac{(aq/p)^k}{k!}\ln\frac{(a-1+k)!}{(a-1)!a^k}.$$

Applying the formula (3.12) with (3.13) again it can be seen that

$$\ln\frac{(a-1+k)!}{(a-1)!a^k}\geqq\left(a-\frac{1}{2}+k\right)\ln\left(1+\frac{k}{a}\right)-k+\frac{1}{12}\left(\frac{1}{a+k}-\frac{1}{a}\right)$$
$$-\frac{1}{32(a+k)^2(a+k+1)(a+k+2)}$$
$$\geqq\left(a-\frac{1}{2}+k\right)\ln\left(1+\frac{k}{a}\right)-k+\frac{61}{768}\left(\frac{1}{a+k}-\frac{1}{a}\right).$$

Let us put

$$v(t)=-\left(a-\frac{1}{2}+t\right)\ln\left(1+\frac{t}{a}\right)+t-\frac{61}{768}\left(\frac{1}{a+t}-\frac{1}{a}\right)\qquad(a>0,\ t\geqq2),$$

then

$$v'(t)=-\ln\left(1+\frac{t}{a}\right)+\frac{1}{2}(a+t)^{-1}+\frac{61}{768}(a+t)^{-2}$$

$$v''(t)=-(a+t)^{-1}-\frac{1}{2}(a+t)^{-2}-\frac{61}{384}(a+t)^{-3}<-(a+t)^{-1}-\frac{1}{2}(a+t)^{-2}<0$$

$$v'''(t)=(a+t)^{-2}+(a+t)^{-3}+\frac{61}{128}(a+t)^{-4}>(a+t)^{-2}>0$$

$$v''''(t)=-2(a+t)^{-3}-3(a+t)^{-4}-\frac{61}{32}(a+t)^{-5}<0\,.$$

Thus, by the generalized Jensen inequality,

$$(3.24)\quad \underset{p}{\mathrm{E}}\left[-\ln\frac{(a-1+k)!}{(a-1)!a^k}\right]$$
$$\leqq -\left(a-\frac{1}{2}+\frac{aq}{p}\right)\ln\left(1+\frac{aq/p}{a}\right)+\frac{aq}{p}-\frac{61}{768}\left(\frac{1}{a+aq/p}-\frac{1}{a}\right)$$
$$+\frac{aq}{2p}\left[-\frac{1}{a+aq/p}-\frac{1}{2(a+aq/p)^2}\right]+\frac{aq}{6p}\left[\frac{1}{(a+aq/p)^2}\right]$$
$$\leqq \frac{a}{p}\ln p+\frac{aq}{p}-\frac{1}{2}\ln(1-q)-\frac{q}{2}+\frac{61q}{768a}-\frac{pq}{12a}$$
$$\leqq \frac{a}{p}\ln p+\frac{aq}{p}-\frac{1}{2}\ln(1-q)-\frac{q}{2}+\frac{61q^2}{768a}\,.$$

Hence, combining (3.23) and (3.24), we obtain

$$I_p\leqq -\frac{1}{2}\ln(1-q)-\frac{q}{2}+\frac{61q^2}{768a}<\frac{q^2}{p}-\frac{q^3}{3p}+\frac{61q^2}{768a}<\frac{q^2}{p^2}\left\{p-\frac{pq}{3}+\frac{p^2}{768a}\right\},$$

which completes the proof of the theorem.

PROOF OF COROLLARY 3.2. From (3.18) and (3.19)

$$(3.25)\quad I_2\leqq\frac{q^2}{p^2}\min\left(\frac{1}{4}+\frac{1}{6}aq,\ 1-\frac{4}{3}q+\frac{1}{3}q^2+\frac{61(1-q)^2}{768a}\right)\equiv\frac{q^2}{p^2}\varphi_q(a)$$
$$\leqq\frac{q^2}{p^2}\varphi_q(a_0)\left[a_0=\frac{4(9-16q+4q^2)+\sqrt{16(9-16q+4q^2)^2+122q(1-q)^2}}{16q}\right]$$
$$=\frac{q^2}{p^2}\left[\frac{1}{4}+\frac{1}{6}a_0q\right]\equiv I_2^*$$

and thus from (3.17)

$$(3.26)\qquad V_2(a,q)\leqq 2\min\left[\eta(I_2^*),\ \zeta(I_2^*)\right]\equiv c^*(q)\frac{q}{p}\,.$$

In Figure 3.2, $2(q/p)^{-1}\eta(I_2^*)$ and $2(q/p)^{-1}\zeta(I_2^*)$ are plotted. It can be seen that $c^*(q)$ is a continuous and monotone decreasing function of q. So,

$$c^*(q)<\lim_{q\to 0+}2(q/p)^{-1}\zeta(I_2^*)=\sqrt{2}\,,$$

which completes the proof of the inequality (3.20).

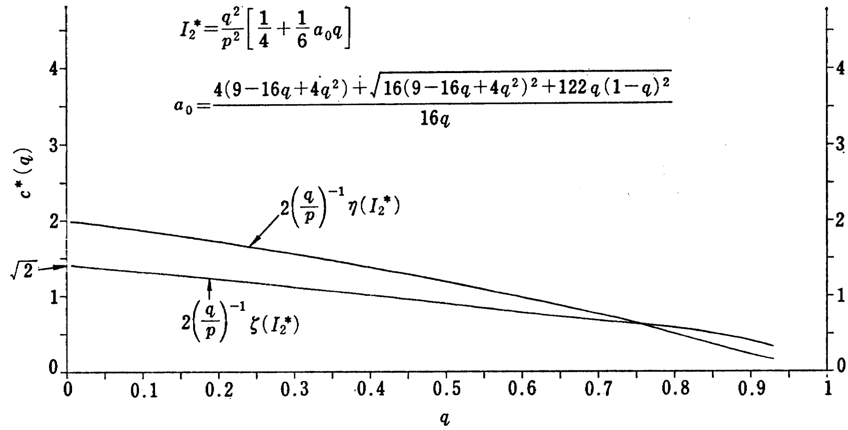

Fig. 3.2. $c^*(q)$ for $V(a, q)\leqq c^*(q)\frac{q}{p}$

Acknowledgements

The author wishes to thank the referee for his comments and helpful suggestions.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Church, J. D. and Harris, B. (1970). The estimation of reliability from stress-strength relationships, *Technometrics*, **12**, 49-54.
[2] Hodges, J. L., Jr. and LeCam, L. (1960). The Poisson approximation to the Poisson binomial distribution, *Ann. Math. Statist.*, **31**, 737-740.
[3] Ikeda, S. (1963). Asymptotic equivalence of probability distributions with applications to some problems of asymptotic independence, *Ann. Inst. Statist. Math.*, **15**, 87-116.
[4] Ikeda, S. (1969). A remark on the incomparability of two criteria for a uniform convergence of probability measures, *Ann. Inst. Statist. Math.*, **21**, 533-536.
[5] Kerstan, J. (1964). Verallgemeinerung eines Satzes von Prochorow und LeCam, *Zeit. Wahrscheinlichkeitsth.*, **2**, 173-179.
[6] Kraft, O. and Schmitz, N. (1969). A note on Hoeffding's inequality, *J. Amer. Statist. Ass.*, **64**, 907-912.
[7] LeCam, L. (1965). On the distribution of sums of independent random variables, "*Bernoulli, Bays, Laplace*", proceedings of an international research seminar (Berkeley 1963), Springer Berlin, 179-202.
[8] Matsunawa, T. (1975). On the error evaluation of the joint normal approximation for sample quantiles, *Ann. Inst. Statist. Math.*, **27**, 189-199.
[9] Matsunawa, T. (1976). Some inequalities based on inverse factorial series, *Ann. Inst. Statist. Math.*, **28**, A, 291-305.
[10] Matsunawa, T. (1977). Approximations to the probabilities of binomial and multinomial random variables and chi-squarety pe statistics, *Ann. Inst. Statist. Math.*, **29**, A, 333-358.
[11] Matsunawa, T. (1982). Uniform ϕ-equivalence of probability distributions based on

information and related measures of discrepancy, *Ann. Inst. Statist. Math.*, **34**, A, 1-17.

[12] Meshalkin, L. D. (1968). On the robustness of some characterizations of the normal distribution, *Ann. Math. Statist.*, **39**, 1747-1750.

[13] Prohorov, Yu. V. (1953). Asymptotic behavior of the binomial distribution, *Uspehi Matematicheskiikh Nauk*, 8, 135-142.

[14] Romanovska, M. (1977). A note on the upper bound for the distance in total variation between the binomial and the Poisson distribution, *Statist. Neerlandica*, **31**, 127-130.

[15] Vervaat, W. (1969). Upper bounds for the distance in total variation between the binomial or negative binomial and the Poisson distribution, *Statist. Neerlandica*, **23**, 79-86.