ON UNBIASED ESTIMATES OF THE POPULATION MEAN BASED ON THE SAMPLE STRATIFIED BY MEANS OF ORDERING

KOITI TAKAHASI AND KAZUMASA WAKIMOTO

(Received July 6, 1967)

1. Introduction

We assume throughout this paper that the population under consideration has the distribution function $F(x)$ and the density function $f(x)$ with finite mean μ and finite variance σ^2. It should be noted that we assume nothing about the distribution except the above existence assumption, that is, we shall consider a non-parametric problem. When we are concerned with estimation of the population mean: we often encounter the situations where the measurement of the quantity of each element drawn from the population is very laborious but several elements can easily be arranged in the order of magnitude, for example, the case where the elements can be arranged without the measurement of each quantity. In practice the number of elements which are easily arranged will possibly be two or three, but we shall consider the general case.

The following three examples will give us a better understanding of the situations:

Example 1. Let us suppose that the quantity under consideration is the length of a kind of bacterial cells and the length of the cells in a microscopic field is measured by using a micrometer. While the operation for the measurement will be laborious, the order of magnitude of two or three cells in the same microscopic field may be found by a glance in most cases.

Example 2. Let us suppose that the quantity under consideration is the height of trees. We can find by a glance the order of height of two or three trees standing nearly each other.

Example 3. Let us suppose that the quantity under consideration is the number of a kind of bacterial cells per unit volume. If there are several test tubes containing the cell suspension, we can rearrange these tubes in order of concentration by using an optical instrument without knowing the exact values.

In the situation mentioned above, we can obtain an unbiased estimator of the population mean based on the sample which is stratified by means of the order of magnitude as follows.

Let $X_{11}, X_{12}, \cdots, X_{1n}; X_{21}, X_{22}, \cdots, X_{2n}; \cdots\cdots; X_{n1}, X_{n2}, \cdots, X_{nn}$ be independent random variables all having the cdf $F(x)$ and $X_{i(1)}, X_{i(2)}, \cdots, X_{i(n)}$ be the order statistics of $X_{i1}, X_{i2}, \cdots, X_{in}$ $(i=1, 2, \cdots, n)$.

Let us define $\bar{X}_{[n]}$ by $\bar{X}_{[n]}=\sum_{i=1}^{n} X_{i(i)}/n$. We shall consider the statistic $\bar{X}_{[n]}$ as an estimator for μ. To obtain an observed value of the $\bar{X}_{[n]}$, we need an observed value of $(X_{1(1)}, X_{2(2)}, \cdots, X_{n(n)})$. To obtain an observed value $X_{i(i)}$, we need the ordering of the sample of $(X_{i1}, X_{i2}, \cdots, X_{in})$. Thus we shall be able to expect that the variance of the unbiased (see (3.3)) estimator $\bar{X}_{[n]}$ of μ will be considerably smaller than that of the usual estimator $\bar{X}_n$, the sample mean of a simple random sample of size n. The reason why we compare the variance of $\bar{X}_{[n]}$ not with $\bar{X}_{n^2}$ but with $\bar{X}_n$ is that in our situation the cost of ordering need not be taken into account and we have only to take into account the cost of measurement.

In general the above procedure will be repeated m times. Then, we have m observed values of $\bar{X}_{[n]}$. The total number, say N, of elements whose quantities are measured, is mn, while the total number of elements which are drawn from the population is mn^2, whether they are measured or not. The estimator of μ is the arithmetic mean of m observed values of $\bar{X}_{[n]}$.

Let us here interpret our procedures by an example in the case $N=6$;

Simple random sampling procedure:

Draw 6 elements from the population.
Measure the quantity of each element.
Make the sample mean as an estimate of μ.

Our procedure, $n=2$ (thus $m=3$):

Draw 6 pairs of elements from the population.
Find the order of magnitude in each pair.
Measure the quantity of the smaller element in the first, second and third pair and that of the larger element in the fourth, fifth and sixth pair.
Make the arithmetic mean of these quantities.

Our procedure, $n=3$ (thus $m=2$):

Draw 6 triplets of elements from the population.
Find the order of magnitude in each triplet.

Measure the quantity of the least element in the first and second triplet and that of the middle element in the third and fourth triplet and that of the largest element in the fifth and sixth triplet. Make the arithmetic mean of these quantities.

Let us denote the variance of the kth least order statistic of a sample of size n from the population by $\sigma^2_{n,k}$ $(k=1, 2, \cdots, n)$. If the variances $\sigma^2_{n,k}$ are known (or have been estimated), then we can apply the so-called Neyman allocation to our problem. For simplicity we assume that $\sigma_{n,1}:\sigma_{n,2}:\cdots:\sigma_{n,n}=N_1:N_2:\cdots:N_n$, where $N_1, N_2, \cdots, N_n$ are positive integers. Let $X_{11}, X_{12}, \cdots, X_{1n}; X_{21}, X_{22}, \cdots, X_{2n}; \cdots\cdots; X_{N1}, X_{N2}, \cdots, X_{Nn}$ be a random sample of size nN from the population, where $N=N_1+N_2+\cdots+N_n$ and $X_{i(1)}, X_{i(2)}, \cdots, X_{i(n)}$ be the order statistics of $X_{i1}, X_{i2}, \cdots, X_{in}$, $(i=1, 2, \cdots, N)$. We now define $\bar{X}_{\langle n\rangle N}$ by $\bar{X}_{\langle n\rangle N}=\frac{1}{n}\left(\frac{1}{N_1}\sum_{i=1}^{N_1}X_{i(1)}+\frac{1}{N_2}\sum_{i=N_1+1}^{N_1+N_2}X_{i(2)}+\cdots+\frac{1}{N_n}\sum_{i=N_1+N_2+\cdots+N_{n-1}+1}^{N}X_{i(n)}\right)$. This $\bar{X}_{\langle n\rangle N}$ is an unbiased estimator of the population mean.

In this paper we shall also consider this estimator, but our main purpose is to study the properties of the estimator $\bar{X}_{[n]}$ which seems more practical.

2. Notation and preliminary

Let $X_{n,k}$ be the kth least order statistic in a sample of size n drawn from a continuous population with the pdf $f(x)$, the cdf $F(x)$, the mean μ and the variance σ^2 (We shall use the abbreviations 'the pdf' and 'the cdf' throughout this paper for the probability density function and the cumulative distribution function, respectively.). The pdf, the cdf, the mean and the variance of the distribution of $X_{n,k}$ will be denoted by $f_{n,k}(x)$, $F_{n,k}(x)$, $\mu_{n,k}$ and $\sigma^2_{n,k}$, respectively. Let us denote $\Gamma(n+1)/(\Gamma(k)\cdot\Gamma(n-k+1))$ by $a_{n,k}$. Let us denote the expected value and the variance of a random variable X by $\mathrm{E}(X)$ and $\sigma^2(X)$, respectively.

Some well-known results will be shown below for the latter use. We have, in the first place,

$$f_{n,k}(x)=a_{n,k}F^{k-1}(x)(1-F(x))^{n-k}f(x)\ , \qquad k=1, 2, \cdots, n\ . \tag{2.1}$$

Suppose in the next place that $f(x)$ satisfies the relation

$$f(x)=\sum_{i=1}^{n}\alpha_i f_i(x)\ , \qquad i=1, 2, \cdots, n\ , \tag{2.2}$$

where $f_i(x)$ is a pdf and α_i is a positive constant. Then $\sum_{i=1}^{n}\alpha_i=1$ must be satisfied. Let us denote the mean and the variance of $f_i(x)$ by μ_i

and σ_i^2. If t_i is an unbiased estimate of μ_i and $t_1, t_2, \cdots, t_n$ are independent, then

$$t=\sum_{i=1}^{n} \alpha_i t_i \tag{2.3}$$

is an unbiased estimate of μ and has the variance

$$\sigma^2(t)=\sum_{i=1}^{n} \alpha_i^2 \sigma^2(t_i) . \tag{2.4}$$

If t_i is the sample mean of a simple random sample of size N_i drawn from $f_i(x)$, then

$$\sigma^2(t)=\sum_{i=1}^{n} \alpha_i^2 \sigma_i^2 \Big/ N_i . \tag{2.5}$$

If $N_i=N\alpha_i$ (the proportional allocation), then

$$\sigma_p^2(t)=\left(\sum_{i=1}^{n} \alpha_i \sigma_i^2\right) \Big/ N . \tag{2.6}$$

If $N_i=N\alpha_i\sigma_i \Big/ \left(\sum_{j=1}^{n} \alpha_j \sigma_j\right)$ (the Neyman allocation), then

$$\sigma_N^2(t)=\left(\sum_{i=1}^{n} \alpha_i \sigma_i\right)^2 \Big/ N . \tag{2.7}$$

We also have

$$\begin{aligned} \mu &= \sum_{i=1}^{n} \alpha_i \mu_i \\ \sigma^2 &= \sum_{i=1}^{n} \alpha_i \sigma_i^2 + \sum_{i=1}^{n} \alpha_i (\mu_i - \mu)^2 \end{aligned} \tag{2.8}$$

and

$$\sigma_p^2(t)=\left\{\sigma^2-\sum_{i=1}^{n} \alpha_i(\mu_i-\mu)^2\right\} \Big/ N . \tag{2.9}$$

If $\alpha_1=\alpha_2=\cdots=\alpha_n=\dfrac{1}{n}$, then we have

$$\sigma_p^2(t)=\left(\sum_{i=1}^{n} \sigma_i^2\right) \Big/ nN , \tag{2.10}$$

$$\sigma_N^2(t)=\left(\sum_{i=1}^{n} \sigma_i\right)^2 \Big/ n^2N , \tag{2.11}$$

and

$$\mu=\frac{1}{n}\sum_{i=1}^{n}\mu_i \tag{2.12}$$
$$\sigma^2=\left(\sum_{i=1}^{n}\sigma_i^2\right)\Big/n+\sum_{i=1}^{n}(\mu_i-\mu)^2\Big/n=\left(\sum_{i=1}^{n}\sigma_i^2\right)\Big/n+\sum_{i<j}(\mu_i-\mu_j)^2\Big/n^2$$

corresponding to (2.6), (2.7) and (2.8), respectively.

3. The unbiased estimates of the population mean

From (2.1) the following relation can be obtained easily, but it is fundamental to our discussion:

$$f(x)=\frac{1}{n}\sum_{k=1}^{n}f_{n,k}(x)\ . \tag{3.1}$$

Therefore, we can apply the results of section 2 to our following discussion. Let Y_k be a random variable with the pdf $f_{n,k}(x)$, $(k=1,2,\cdots,n)$ and $Y_1, Y_2, \cdots, Y_n$ be independent. In order to obtain $Y_1, Y_2, \cdots, Y_n$ from a sample drawn from the population with the pdf $f(x)$ we should take $X_{k(k)}$ as Y_k, that is $Y_k=X_{k(k)}$, $(k=1,2,\cdots,n)$, where the $X_{k(k)}$'s are the random variables explained in the introduction.

Let us define $\bar{Y}_{[n]}$ by

$$\bar{Y}_{[n]}=\frac{1}{n}\sum_{k=1}^{n}Y_k\ , \tag{3.2}$$

that is, $\bar{X}_{[n]}$ in the introduction. From (2.3) and (2.10) we have

$$\mathrm{E}(\bar{Y}_{[n]})=\mu \tag{3.3}$$

and

$$\sigma^2(\bar{Y}_{[n]})=\frac{1}{n^2}\sum_{k=1}^{n}\sigma_{n,k}^2\ . \tag{3.4}$$

It is our purpose to compare $\sigma^2(\bar{Y}_{[n]})$ with $\sigma^2(\bar{X}_n)=\frac{\sigma^2}{n}$, that is, the variance of the sample mean of the simple random sample of size n, because of the situation explained in the introduction. It is, therefore, convenient to define $\sigma_{[n]}^2$ by

$$\sigma_{[n]}^2=\frac{1}{n}\sum_{k=1}^{n}\sigma_{n,k}^2\ . \tag{3.5}$$

Then, while $\sigma^2(\bar{X}_n)=\frac{\sigma^2}{n}$, $\sigma^2(\bar{Y}_{[n]})=\frac{\sigma_{[n]}^2}{n}$. Suppose that the actual sample size (strictly speaking, the number of observations whose values are measured) is N. For the simplicity let $N=mn$. In this case m independent $\bar{Y}_{[n]1}, \bar{Y}_{[n]2}, \cdots, \bar{Y}_{[n]m}$, each with the same distribution as that of $\bar{Y}_{[n]}$, are available. If we define $\bar{\bar{Y}}_{[n]m}$ by

$$\bar{\bar{Y}}_{[n]m}=\frac{1}{m}\sum_{j=1}^{m}\bar{Y}_{[n]j}\ , \tag{3.6}$$

then $\mathrm{E}(\bar{\bar{Y}}_{[n]m})=\mu$ and

$$\sigma^2(\bar{\bar{Y}}_{[n]m})=\frac{\sigma^2_{[n]}}{mn}=\frac{\sigma^2_{[n]}}{N}\ . \tag{3.7}$$

On the other hand, $\sigma^2(\bar{X}_N)=\dfrac{\sigma^2}{N}$. Thus $\sigma^2_{[n]}$ in such a sense corresponds to σ^2. Hence our problem will be the comparison between σ^2 and $\sigma^2_{[n]}$.

Now we define $\tau_{[n]}$ by

$$\tau_{[n]}=\frac{\sigma^2(\bar{X}_n)-\sigma^2(\bar{Y}_{[n]})}{\sigma^2(\bar{X}_n)}=\frac{\sigma^2-\sigma^2_{[n]}}{\sigma^2}\ . \tag{3.8}$$

The $\tau_{[n]}$ will represent in a sense the efficiency of the stratification by means of ordering. It should be noted that $\tau_{[n]}$ is invariant under the linear transformation of the variables. Since the covariance of any two of order statistics in a given sample is positive [7], it is obvious that $\tau_{[n]}>0$.

In order that we may be justified in restricting our consideration to the estimator $\bar{Y}_{[n]}$ in (3.2) as the unbiased estimator of the mean of any populations it will be necessary that we state here the following theorem.

THEOREM 1. *A linear combination of Y_k's*

$$\sum_{k=1}^{n}\alpha_k Y_k$$

is an unbiased estimator of population mean whatever the distribution of the population is if and only if $\alpha_1=\alpha_2=\cdots=\alpha_n=\dfrac{1}{n}$.

PROOF. The "if part" is established by (3.3). Next suppose that $\sum \alpha_k Y_k$ is an unbiased estimator of all the population means. We have

$$\sum_{k=1}^{n}\alpha_k\mu_{n,k}=\mu\ .$$

From this and the "if part" we have

$$\sum_{k=1}^{n}\left(\frac{1}{n}-\alpha_k\right)\mu_{n,k}=0 .$$

If we can find n distributions each of which has $\mu_{n,k}^{\langle i\rangle}$ as $\mu_{n,k}$ $(i=1, 2, \cdots, n)$ satisfying

$$\det D_n\neq 0 ,$$

where D_n is the matrix having $\mu_{n,k}^{\langle i\rangle}$ as (i, k)-element, then the proof of the "only if part" will be completed.

In fact, we can take the distributions with the density functions

$$f_l(x)=\begin{cases}\dfrac{1}{l}x^{1/l-1}\exp(-x^{1/l}) , & x>0 ,\\ 0 , & x\leqq 0\end{cases}\qquad (l=1, 2, \cdots, n)$$

as n distributions satisfying the above condition.

We have

$$\mu_{n,k}^{\langle l\rangle}=\frac{n!}{(n-k)!}\sum_{j=0}^{k-1}\frac{(-1)^j\, l!}{j!(k-1-j)!\,(n-k+j+1)^{l+1}} .$$

After some calculations we obtain the relation

$$D_n=C\begin{vmatrix} n^{n-1} & (n-1)^{n-1} & \cdots & 2^{n-1} & 1^{n-1}\\ n^{n-2} & (n-1)^{n-2} & \cdots & 2^{n-2} & 1^{n-2}\\ \vdots & \vdots & & \vdots & \vdots\\ n & n-1 & \cdots & 2 & 1\\ 1 & 1 & \cdots & 1 & 1\end{vmatrix}=C\prod_{i>j}(i-j)\neq 0 ,$$

where C is a non-zero constant. Thus our proof is completed.

Now we have

THEOREM 2.

$$\sigma_{[n]}^2>\sigma_{[n+1]}^2 , \tag{3.9}$$

hence,

$$\tau_{[n]}<\tau_{[n+1]} , \qquad \textit{for all} \quad n\geqq 1 ,$$

where $\sigma_{[n]}^2=\dfrac{1}{n}\sum_{k=1}^{n}\sigma_{n,k}^2$ *and* $\tau_{[n]}=\dfrac{\sigma^2-\sigma_{[n]}^2}{\sigma^2}$.

PROOF. From (2.1) we have

$$f_{n,k}=\frac{n+1-k}{n+1}f_{n+1,k}+\frac{k}{n+1}f_{n+1,k+1} . \tag{3.10}$$

If we denote $\int_{-\infty}^{\infty}x^{\nu}f_{n,k}(x)\,dx$ by $\alpha_{\nu,n,k}$, then from (3.10) we have

$$\alpha_{\nu,n,k}=\frac{n+1-k}{n+1}\alpha_{\nu,n+1,k}+\frac{k}{n+1}\alpha_{\nu,n+1,k+1}\,. \tag{3.11}$$

From this

$$\begin{aligned}\sigma^2_{n,k}&=\alpha_{2,n,k}-\alpha^2_{1,n,k}\\ &=\left(\frac{n+1-k}{n+1}\sigma^2_{n+1,k}+\frac{k}{n+1}\sigma^2_{n+1,k+1}\right)+\left(\frac{n+1-k}{n+1}\mu^2_{n+1,k}+\frac{k}{n+1}\mu^2_{n+1,k+1}\right)\\ &\quad-\left(\frac{n+1-k}{n+1}\mu_{n+1,k}+\frac{k}{n+1}\mu_{n+1,k+1}\right)^2.\end{aligned} \tag{3.12}$$

Then, from (3.12) it follows that

$$\begin{aligned}&\sigma^2_{[n]}-\sigma^2_{[n+1]}\\ &=\frac{1}{n}\sum_{k=1}^{n}\sigma^2_{n,k}-\frac{1}{n+1}\sum_{k=1}^{n+1}\sigma^2_{n+1,k}\\ &=\frac{1}{n(n+1)^2}\left\{n(n+1)\sum_{k=1}^{n+1}\mu^2_{n+1,k}-\sum_{k=1}^{n}((n+1-k)\mu_{n+1,k}+k\mu_{n+1,k+1})^2\right\}\\ &=\frac{1}{n(n+1)^2}\left\{\sum_{k=1}^{n}k(n+1-k)(\mu_{n+1,k+1}-\mu_{n+1,k})^2\right\}>0\,.\end{aligned} \tag{3.13}$$

This completes the proof of the theorem.

COROLLARY 1. *Let $N=ab=cd$, where a, b, c, d are positive integers and $N>a>c>1$, then the following inequalities hold.*

$$\sigma^2(\bar{X}_N)>\sigma^2(\bar{\bar{Y}}_{[c]d})>\sigma^2(\bar{\bar{Y}}_{[a]b})>\sigma^2(\bar{\bar{Y}}_{[N]1})\,. \tag{3.14}$$

This corrollary can easily be obtained from (3.6), (3.7), (3.9).

According to the result of theorem 1 or corrollary 1, the variance of the estimate $\bar{\bar{Y}}_{[n]m}$ decreases as n increases under the condition $N=mn$. The large n, however, will be impractical. In most practical cases n will be two or three. If we can practically take both two and three as the value of n, in such a case we had better take $n=3$ from the viewpoint of the variance of the estimate apart from the other problem.

4. The Neyman allocation

Let us suppose that the variances $\sigma^2_{n,k}$ are known (or have been estimated). Let us denote $\sigma_{n,k}\Big/\sum_{j=1}^{n}\sigma_{n,j}$ by $\gamma_{n,k}$. Let $N_k=N\gamma_{n,k}$, $(k=1,2,\cdots,n)$. Further let us assume for the simplicity that every N_k is positive integer. Let $Y_{11}, Y_{12},\cdots,Y_{1N_1}; Y_{21}, Y_{22},\cdots,Y_{2N_2};\cdots; Y_{n1}, Y_{n2},\cdots,Y_{nN_n}$ be independent random variables and Y_{kj}, $(j=1,2,\cdots,N_k)$ be drawn from the distribution with the pdf $f_{n,k}(x)$.

Let us define $Z_{N,n}$ by

(4.1) $$Z_{N,n}=\frac{1}{n}\sum_{k=1}^{n}\left(\frac{1}{N_k}\sum_{j=1}^{N_k} Y_{k,j}\right) .$$

This $Z_{N,n}$ has been appeared in the introduction as $\bar{X}_{\langle n\rangle N}$. From (2.3) and (2.7), we have

(4.2) $$\mathrm{E}(Z_{N,n})=\mu$$

and

(4.3) $$\sigma^2(Z_{N,n})=\frac{1}{Nn^2}\left(\sum_{k=1}^{n}\sigma_{n,k}\right)^2 .$$

Let us define $\sigma^2_{\langle n\rangle}$ and $\tau_{\langle n\rangle}$ by

(4.4) $$\sigma^2_{\langle n\rangle}=\left(\frac{1}{n}\sum_{k=1}^{n}\sigma_{n,k}\right)^2 .$$

and

(4.5) $$\tau_{\langle n\rangle}=\frac{\sigma^2-\sigma^2_{\langle n\rangle}}{\sigma^2} , \qquad \text{respectively.}$$

The variance $\sigma^2(Z_{N,n})$ can be written in terms of $\sigma^2_{\langle n\rangle}$ as

(4.6) $$\sigma^2(Z_{N,n})=\frac{\sigma^2_{\langle n\rangle}}{N} .$$

Therefore, when we use $Z_{N,n}$ as the unbiased estimate of μ, the efficiency of $Z_{N,n}$ relative to $\bar{X}_N$ will be expressed by $\tau_{\langle n\rangle}$ independently of N.

From (3.5) and (4.4) it follows that

(4.7) $$\sigma^2_{[n]}-\sigma^2_{\langle n\rangle}=\frac{1}{n^2}\sum_{1\leqq l<k\leqq n}\sum(\sigma_{n,l}-\sigma_{n,k})^2\geqq 0 .$$

The last inequality in (4.7) will be obvious since $\sigma^2_{[n]}$ corresponds to proportional allocation and $\sigma^2_{\langle n\rangle}$ corresponds to Neyman allocation.

5. Examples of $\tau_{[n]}$ and $\tau_{\langle n\rangle}$

In this section the numerical values of $\tau_{[n]}$ and $\tau_{\langle n\rangle}$ are shown for several distributions, especially for $n=2$ and 3, and the moments of order statistics which are necessary for the calculation of $\tau_{[n]}$ and $\tau_{\langle n\rangle}$ are shown. It should be noted again that $\tau_{[n]}$ and $\tau_{\langle n\rangle}$ are invariant under the linear transformation of the random variables.

It may be useful to introduce another representation of the efficiencies of estimators, denoted by $e_{[n]}$ and $e_{\langle n\rangle}$, which are defined by

$$e_{[n]}=\frac{\sigma^2}{\sigma^2_{[n]}}\times 100 \tag{5.1}$$

and

$$e_{\langle n\rangle}=\frac{\sigma^2}{\sigma^2_{\langle n\rangle}}\times 100 . \tag{5.2}$$

5.1 The moments of order statistics

[R] Rectangular distribution;

$$\sigma^2_{n,k}=\frac{12k(n-k+1)}{(n+1)^2(n+2)}\sigma^2 , \qquad \text{([11] p. 383)} . \tag{5.3}$$

From this we have the simple form

$$\tau_{[n]}=\frac{n-1}{n+1} . \tag{5.4}$$

[E] Exponential distribution;

$$\sigma^2_{n,k}=\sigma^2\sum_{j=1}^{k}\frac{1}{(n-j+1)^2} , \qquad \text{([11] p. 343)} . \tag{5.5}$$

Thus we have

$$\tau_{[n]}=1-\frac{1}{n}\sum_{k=1}^{n}\frac{1}{k} . \tag{5.6}$$

[S] Symmetrical distributions;

Let us put

$$f(x)=\begin{cases}\frac{p}{2}|x|^{p-1} , & -1<x<1 , \\ 0 , & \text{otherwise} ,\end{cases} \tag{5.7}$$

where $p>0$. We have

(5.8) $E(X^{\nu}_{n,k})$

$$=2^{-n}\Gamma(n+1)\left\{(-1)^{\nu}\sum_{j=0}^{n-k}\frac{\Gamma(j+1+(\nu/p))}{\Gamma(j+1)\Gamma(n+1-k-j)\Gamma(k+1+j+(\nu/p))}\right.$$
$$\left.+\sum_{j=0}^{k-1}\frac{\Gamma(j+1+(\nu/p))}{\Gamma(j+1)\Gamma(k-j)\Gamma(n+2-k+j+(\nu/p))}\right\} .$$

Thus we have

(5.9)
$$\sigma^2_{2,1}=\sigma^2_{2,2}=\frac{(3p^2+2p+1)}{(2p+1)^2}\sigma^2 ,$$
$$\tau_{[2]}=\frac{p(p+2)}{(2p+1)^2} ,$$
$$\sigma^2_{3,1}=\sigma^2_{3,3}=\frac{3(7p^3+8p^2+8p+4)}{4(2p+1)^2(3p+2)}\sigma^2 ,$$
$$\sigma^2_{3,2}=\frac{3p}{(3p+2)}\sigma^2 ,$$
$$\tau_{[3]}=\frac{3}{2}\,\frac{p(p+2)}{(2p+1)^2} .$$

[J] J-shaped distributions;

Let us put

(5.10)
$$f(x)=\begin{cases} px^{p-1} , & 0<x<1 , \\ 0 , & \text{otherwise} , \end{cases}$$

where $p>0$. We have

(5.11)
$$\mathrm{E}(X^{\nu}_{n,k})=\frac{\Gamma(n+1)\Gamma(k+(\nu/p))}{\Gamma(n+(\nu/p)+1)\Gamma(k)} , \qquad \text{([2] p. 305)}$$

Thus we have

(5.12)
$$\sigma^2=\frac{p}{(p+1)^2(p+2)} ,$$
$$\sigma^2_{2,1}=\frac{p(5p+1)}{(2p+1)^2}\sigma^2 ,$$
$$\sigma^2_{2,2}=\frac{(p+1)(p+2)}{(2p+1)^2}\sigma^2 ,$$
$$\tau_{[2]}=\frac{p(p+2)}{(2p+1)^2} ,$$
$$\tau_{\langle 2\rangle}=\frac{5p^2+6p+1-\sqrt{p(p+1)(p+2)(5p+1)}}{2(2p+1)^2} ,$$
$$\sigma^2_{3,1}=\frac{3p^2(49p^3+47p^2+11p+1)}{(3p+2)(2p+1)^2(3p+1)^2}\sigma^2 ,$$
$$\sigma^2_{3,2}=\frac{3p(p+1)(p+2)(13p^2+10p+1)}{(3p+2)(2p+1)^2(3p+1)^2}\sigma^2 ,$$
$$\sigma^2_{3,3}=\frac{3(p+1)^2(p+2)}{(3p+2)(3p+1)^2}\sigma^2 ,$$

and

$$\tau_{[3]}=\frac{2p(p+2)(7p^2+4p+1)}{(2p+1)^2(3p+1)^2}\ .$$

Put $p=\frac{1}{2}$, then we have

$$(5.13)\qquad \sigma^2=\frac{4}{45}\ ,\qquad \sigma^2_{[n]}=\frac{3n+5}{15(n+1)(n+2)}\ ,\qquad \tau_{[n]}=\frac{(n-1)(4n+7)}{4(n+2)(n+1)}\ .$$

[B] Compound exponential (Burr's) distributions;

Let us put

$$(5.14)\qquad f(x)=\begin{cases}\dfrac{p}{(1+x)^{1+p}}\ , & x>0\ ,\\ 0\ , & \text{otherwise,}\end{cases}$$

where $p>2$.

$$(5.15)\qquad \mathrm{E}(X^{\nu}_{n,k})=\frac{\Gamma(n+1)}{\Gamma(n-k+1)}\sum_{j=0}^{\nu}\binom{\nu}{j}(-1)^{\nu-j}\frac{\Gamma(n+1-k-(j/p))}{\Gamma(n+1-(j/p))}\ .$$

Thus we have

$$\sigma^2=\frac{p}{(p-2)(p-1)^2}\ ,\qquad \sigma^2_{2,1}=\frac{p}{(2p-1)^2(p-1)}\ ,$$

$$\sigma^2_{2,2}=\frac{p^2(5p-1)}{(p-1)^2(p-2)(2p-1)^2}\ ,$$

$$\tau_{[2]}=\frac{p(p-2)}{(2p-1)^2}\ ,$$

$$(5.16)\qquad \tau_{\langle 2\rangle}=\frac{(p-1)(5p-1)}{2(2p-1)^2}\left(1-\sqrt{\frac{p(p-2)}{(p-1)(5p-1)}}\right)\ ,$$

$$\sigma^2_{3,1}=\frac{3(p-2)(p-1)^2}{(3p-2)(3p-1)^2}\sigma^2\ ,$$

$$\sigma^2_{3,2}=\frac{3p(13p^2-10p+1)(p-1)(p-2)}{(3p-2)(3p-1)^2(2p-1)^2}\sigma^2\ ,$$

$$\sigma^2_{3,3}=\frac{3p^2(49p^3-47p^2+11p-1)}{(3p-2)(3p-1)^2(2p-1)^2}\sigma^2$$

and

$$\tau_{[3]}=\frac{2p(p-2)(7p^2-4p+1)}{(3p-1)^2(2p-1)^2}\ .$$

[T] Triangular distributions;

Let the pdf be

$$f(x)=\begin{cases} \dfrac{2}{\rho}x\ , & 0\leqq x\leqq 1\ , \\ \dfrac{2}{\rho(\rho-1)}(\rho-x)\ , & 1\leqq x\leqq \rho\ , \\ 0\ , & \text{otherwise,} \end{cases} \tag{5.18}$$

where $\rho\geqq 1$. Then we have

$$\begin{aligned} \mathrm{E}(X_{n,k}^{\nu})=&2a_{n,k}\sum_{j=0}^{n-k}(-1)^{j}\binom{n-k}{j}\frac{1}{(2k+2j+\nu)\rho^{k+j}} \\ &+2\rho^{\nu}a_{n,k}\sum_{l=0}^{\nu}\sum_{j=0}^{k-1}(-1)^{j+l}\binom{\nu}{l}\binom{k-1}{j}\left(1-\frac{1}{\rho}\right)^{n+j+l+1-k}\frac{1}{2n+2j+l+2-2k}\ . \end{aligned} \tag{5.19}$$

Therefore,

$$\begin{aligned} \sigma^2&=\frac{1}{18}(\rho^2-\rho+1)\ , \\ \sigma_{2,1}^2&=\frac{1}{225\rho^2}(6\rho^4-6\rho^3+15\rho^2-3\rho-1)\ , \\ \sigma_{2,2}^2&=\frac{1}{225\rho^2}(11\rho^4-11\rho^3+7\rho-1)\ , \\ \tau_{[2]}&=\frac{2(2\rho^2-\rho+1)^2}{25\rho^2(\rho^2-\rho+1)}\ , \\ \sigma_{3,1}^2&=\frac{1}{4900\rho^4}(75\rho^6-75\rho^5+350\rho^4-140\rho^3-45\rho^2+40\rho-4)\ , \\ \sigma_{3,2}^2&=\frac{1}{4900\rho^4}(146\rho^6-146\rho^5+308\rho^3-194\rho^2+48\rho-16)\ , \\ \sigma_{3,3}^2&=\frac{1}{4900\rho^4}(201\rho^6-201\rho^5+95\rho^2-16\rho-4) \end{aligned} \tag{5.20}$$

and

$$\tau_{[3]}=\frac{4}{1225}\cdot\frac{(148\rho^6-148\rho^5+175\rho^4-63\rho^3+54\rho^2-27\rho+9)}{\rho^4(\rho^2-\rho+1)}\ .$$

[SP] A special distribution [5];

The means and variances of order statistics from the distribution of the random variable X defined by

$$X=(1-U)^{-1/10}-U^{-1/10}\ , \tag{5.21}$$

where U has the uniform distribution on the interval $[0, 1]$ were tabulated in [5]. From the table we can obtain the numerical values of $\sigma^2_{n,k}$, $\tau_{[n]}$ and $\tau_{\langle n\rangle}$ for $n \leqq 10$.

[EV] An extreme-value distribution;

Liebelein [8] evaluated the moments of order statistics in sample from the extremal value distribution with the cdf

$$F(x)=\exp(-\exp(-x)) . \tag{5.22}$$

By using the results we can obtain the numerical values of $\tau_{[2]}$, $\tau_{\langle 2\rangle}$, $\tau_{[3]}$ and $\tau_{\langle 3\rangle}$.

[W] Weibull distributions;

Let us denote the cdf of Weibull distribution by

$$F(x)=\begin{cases} 1-\exp(-x^b) , & x>0 , \\ 0 , & x\leqq 0 , \end{cases} \tag{5.23}$$

where $b>0$. Then we have

$$\mathrm{E}(X^{\nu}_{n,k})=a_{n,k}\Gamma\left(1+\frac{\nu}{b}\right)\sum_{l=0}^{k-1}(-1)^l\binom{k-1}{l}(n+l-k+1)^{-1-(\nu/b)} , \qquad ([9]) . \tag{5.24}$$

Thus

$$\sigma^2=\Gamma\left(1+\frac{2}{b}\right)-\Gamma^2\left(1+\frac{1}{b}\right) ,$$

$$\sigma^2_{2,1}=2^{-2/b}\left(\Gamma\left(1+\frac{2}{b}\right)-\Gamma^2\left(1+\frac{1}{b}\right)\right) ,$$

$$\sigma^2_{2,2}=(2-2^{-2/b})\Gamma\left(1+\frac{2}{b}\right)-(2-2^{-1/b})^2\Gamma^2\left(1+\frac{1}{b}\right) ,$$

$$\tau_{[2]}=\frac{\Gamma^2\left(1+\frac{1}{b}\right)(1-2^{-1/b})^2}{\Gamma\left(1+\frac{2}{b}\right)-\Gamma\left(1+\frac{1}{b}\right)^2} , \tag{5.25}$$

$$\sigma^2_{3,1}=3^{-2/b}\left(\Gamma\left(1+\frac{2}{b}\right)-\Gamma^2\left(1+\frac{1}{b}\right)\right) ,$$

$$\sigma^2_{3,2}=(3\cdot 2^{-2/b}-2\cdot 3^{-2/b})\Gamma\left(1+\frac{2}{b}\right)-(3\cdot 2^{-1/b}-2\cdot 3^{-1/b})^2\Gamma^2\left(1+\frac{1}{b}\right) ,$$

$$\sigma^2_{3,3}=(3-3\cdot 2^{-2/b}+3^{-2/b})\Gamma\left(1+\frac{2}{b}\right)-(3-3\cdot 2^{-1/b}+3^{-1/b})^2\Gamma^2\left(1+\frac{1}{b}\right)$$

and

$$\tau_{[3]}=\frac{2}{\Gamma\left(1+\frac{2}{b}\right)-\Gamma^2\left(1+\frac{1}{b}\right)}$$
$$\times(3^{-2/b}+3\cdot 2^{-2/b}-3\cdot 2^{-1/b}3^{-1/b}+3^{-1/b}-3\cdot 2^{-1/b}+1)\Gamma^2\left(1+\frac{1}{b}\right).$$

[G] Gamma distribution;

We consider the Gamma distribution with the pdf

$$f(x)=\begin{cases}\dfrac{e^{-x}}{\Gamma(p)}x^{p-1}, & x>0, \\ 0, & x\leqq 0, \end{cases} \tag{5.26}$$

where $p=1, 2, 3, 4, 5$. The tables of $\mu_{n,k}$ and $\sigma^2_{n,k}$ have been obtained by Gupta ([4], [11] p. 439) for $n=1, 2, \cdots, 10$.

[DE] Double exponential distribution;

Let $f(x)$ be the pdf of the double exponential distribution:

$$f(x)=\frac{1}{2}e^{-|x|}, \qquad -\infty<x<\infty. \tag{5.27}$$

Then we have

$$\mathrm{E}(X^{\nu}_{n,k})=a_{n,k}\sum_{l=0}^{n-k}\frac{(-1)^{\nu+l}}{2^{k+l}}\binom{n-k}{l}\frac{\nu!}{(k+l)^{\nu+1}} + a_{n,k}\sum_{l=0}^{k-1}\frac{(-1)^l}{2^{n+l+1-k}}\cdot\frac{\nu!}{(n+l+1-k)^{\nu+1}}. \tag{5.28}$$

The numerical values of $\sigma^2_{n,k}$ for $n=2, 3, 4, 5$ have been tabulated by Sarhan [10].

[N] Normal distribution;

In order to calculate $\tau_{[n]}$ and $\tau_{\langle n\rangle}$ we have used the table of $\sigma^2_{n,k}$'s represented in ([11] pp. 200–205).

5.2 The table of $\tau_{[n]}$, $\tau_{\langle n\rangle}$, $e_{[n]}$ and $e_{\langle n\rangle}$ for $n=2, 3$

Since, in the most practical situations, n will be two or three, we shall now show as Table 1 the values of $\tau_{[n]}$, $\tau_{\langle n\rangle}$, $e_{[n]}$ and $e_{\langle n\rangle}$ of the distributions mentioned in 5.1 for $n=2, 3$. It should be noted that in symmetrical distribution $\tau_{[2]}$ and $\tau_{\langle 2\rangle}$ are identical because of (4.7). It may be worth-while to mention that the efficiency $\tau_{[2]}$ can be expressed in terms of the mean difference Δ_1 ([6] p. 46) of the parent distribution as $\Delta_1^2/4\sigma^2$.

Table 1. Efficiencies for $n=2, 3$

Distribution	$\tau_{[2]}$	$\tau_{[3]}$	$\tau_{\langle 2\rangle}$	$\tau_{\langle 3\rangle}$	$e_{[2]}$	$e_{[3]}$	$e_{\langle 2\rangle}$	$e_{\langle 3\rangle}$
[R]	.333	.500	.333	.502	150	200	150	201
[N]	.318	.477	.318	.479	147	191	147	192
[E]	.250	.389	.345	.510	133	164	153	204
[D E]	.281	.422	.281	.439	139	173	139	178
[E V]	.292	.442	.323	.484	141	179	148	194
[S P]	.286	.429	.286	.442	140	175	140	179
[G] $p=$ 2	.281	.430	.334	.497	139	175	150	199
3	.293	.445	.329	.491	141	180	149	197
4	.299	.453	.327	.488	143	183	149	195
5	.303	.458	.325	.487	143	184	148	195
[W] $b=1/2$	.113	.191	.393	.559	113	124	165	227
2	.314	.473	.327	.489	146	190	149	196
3	.322	.484	.323	.485	147	194	148	194
4	.322	.483	.322	.484	147	193	147	194
5	.319	.479	.322	.483	147	192	147	193
10	.310	.467	.322	.483	145	188	147	193
[T] $\rho=$ 1	.320	.483	.335	.502	147	193	150	201
2	.327	.490	.327	.490	149	196	149	196
3	.325	.488	.329	.493	148	195	149	197
4	.323	.487	.331	.497	148	195	149	199
[J] $p=1/5$	.224	.365	.350	.523	129	157	154	210
1/2	.313	.475	.336	.507	145	190	151	203
3	.306	.465	.337	.503	144	187	151	201
4	.296	.452	.338	.504	142	183	151	202
[S] $p=1/2$	.313	.469	.313	.471	146	.188	146	189
3	.306	.459	.306	.475	144	.185	144	190
4	.296	.444	.296	.463	142	.180	142	186
[B] $p=$ 3	.120	.195	.377	.538	114	124	161	216
4	.163	.262	.364	.526	119	136	157	211
5	.185	.295	.358	.521	123	142	156	209
10	.222	.348	.351	.514	129	153	154	206

5.3 $\tau_{[n]}$ and $\tau_{\langle n\rangle}$ for $n \geqq 4$

In Table 2 the values of $\tau_{[n]}$ and $\tau_{\langle n\rangle}$ for $n=2, 3, 4, \cdots, 20$ are given for a rectangular [R], a normal [N], an exponential [E], a compound exponential [B], and a J-shaped [J] distribution, and the values of $\tau_{[n]}$ and $\tau_{\langle n\rangle}$ for $n=2, 3, 4, \cdots, 10$ are given for a gamma [G] and a special

Table 2. $\tau_{[n]}$ and $\tau_{\langle n\rangle}$ for $2 \leqq n \leqq 20$

	n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\tau_{[n]}$	[R], Sup $\tau_{[n]}$		.333	.500	.600	.667	.714	.750	.778	.800	.818	.833	.846	.857	.867	.875	.882	.889	.895	.900	.905
	[N]		.318	.477	.574	.639	.686	.722	.750	.773	.791	.807	.821	.832	.842	.851	.859	.866	.873	.879	.884
	[E]		.250	.389	.479	.543	.592	.630	.660	.686	.707	.725	.741	.755	.768	.779	.789	.798	.806	.813	.820
	[B] $p=2.1$		.021	.034	.045	.053	.060	.066	.071	.076	.080	.083	.087	.090	.093	.096	.098	.101	.103	.105	.107
	[B] $p=3$		.120	.195	.248	.289	.322	.350	.373	.393	.410	.426	.440	.453	.464	.475	.485	.494	.503	.511	.518
	[J] $p=1/1000$		.002	.004	.006	.008	.010	.012	.014	.016	.018	.020	.021	.023	.025	.027	.029	.031	.033	.035	.037
	[J] $p=1/2$		.313	.475	.575	.643	.692	.729	.758	.782	.801	.817	.831	.843	.853	.862	.870	.877	.884	.889	.894
	[G] $p=4$		.299	.453	.547	.612	.660	.696	.725	.749	.768										
	[S P]		.286	.429	.517	.578	.623	.658	.686	.709	.728										
$\tau_{\langle n\rangle}$	Sup $\tau_{\langle n\rangle}$		.500	.667	.750	.800	.833	.857	.875	.889	.900	.909	.917	.923	.924	.933	.938	.941	.944	.947	.950
	[R]		.333	.502	.604	.672	.720	.756	.784	.806	.824	.839	.852	.863	.872	.880	.887	.894	.900	.905	.909
	[N]		.318	.479	.576	.642	.690	.726	.755	.778	.797	.813	.826	.838	.848	.857	.865	.872	.878	.884	.889
	[E]		.345	.510	.606	.670	.715	.749	.776	.797	.815	.830	.842	.853	.862	.870	.877	.884	.890	.895	.900
	[B] $p=2.1$		.438	.601	.688	.742	.779	.807	.828	.844	.858	.869	.879	.887	.894	.900	.905	.910	.915	.919	.922
	[B] $p=3$		.377	.538	.630	.689	.731	.762	.786	.806	.822	.835	.847	.856	.865	.873	.879	.885	.891	.895	.900
	[J] $p=1/1000$		.479	.646	.732	.783	.818	.842	.861	.876	.888	.897	.905	.912	.918	.923	.928	.932	.935	.938	.941
	[J] $p=1/2$		.336	.507	.610	.678	.726	.762	.790	.812	.830	.845	.857	.868	.877	.885	.892	.898	.904	.909	.913
	[G] $p=4$		.327	.488	.586	.651	.698	.734	.762	.784	.802										
	[S P]		.286	.442	.541	.608	.658	.696	.726	.750	.771										

[SP] distribution. (Refer to 7.1 and 7.2 for Sup $\tau_{[n]}$ and Sup $\tau_{\langle n\rangle}$ in Table 2.)

5.4 The relation between efficiency and parameter

There are the families of distributions with the parameter which

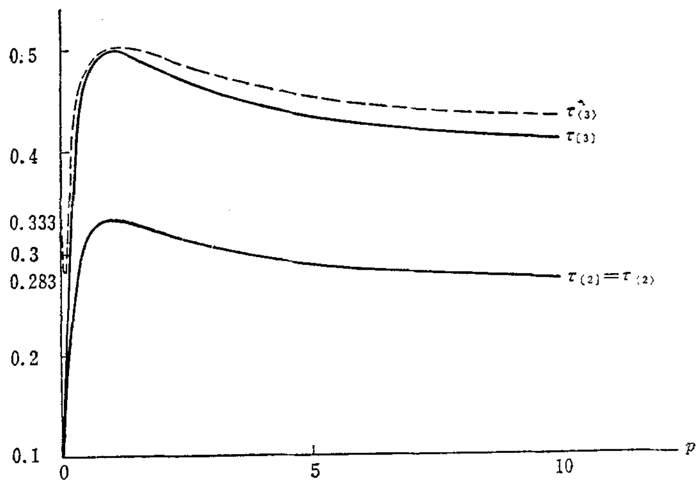

Fig. 1a. [S]

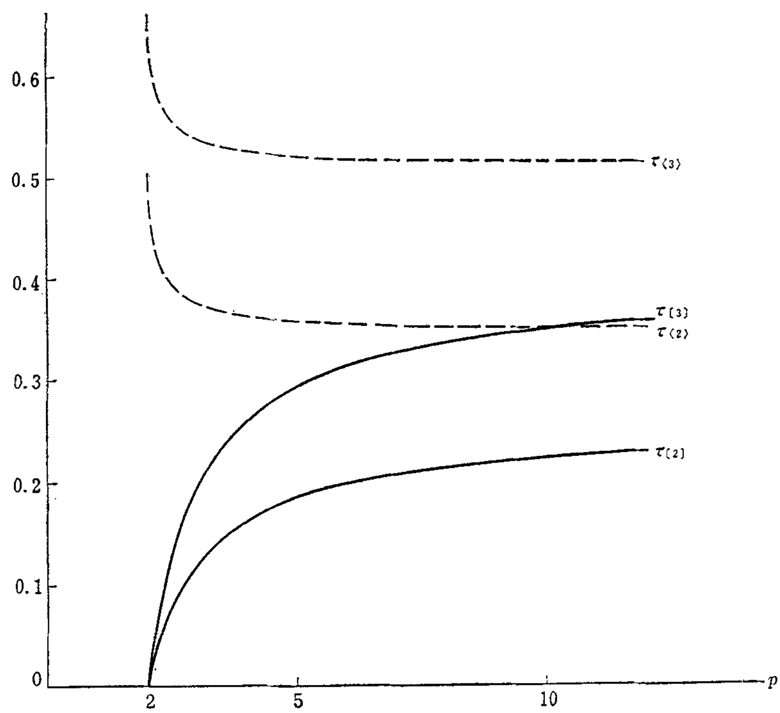

Fig. 1b. [B]

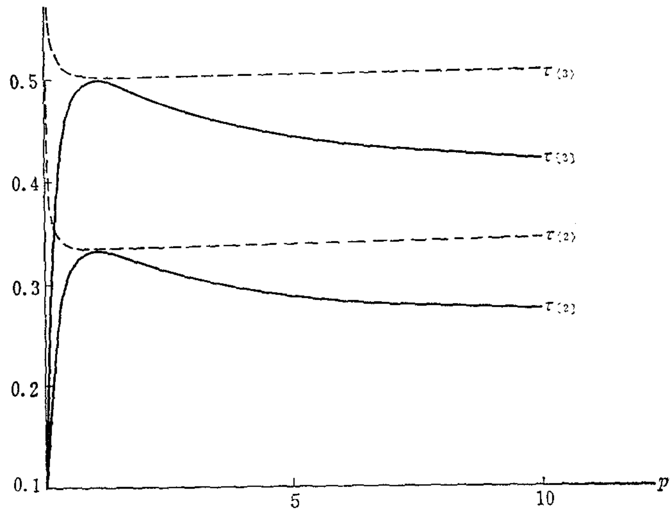

Fig. 1c. [J]

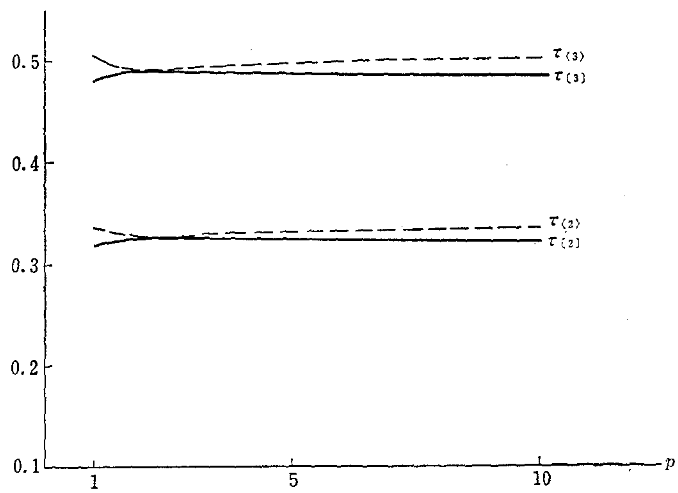

Fig. 1d. [T]

Fig. 1. The variation of values of $\tau_{[2]}$, $\tau_{\langle 2\rangle}$, $\tau_{[3]}$ and $\tau_{\langle 3\rangle}$ with the parameter

has its effect on $\tau_{[n]}$ and $\tau_{\langle n\rangle}$. Here we shall only show several examples. In Fig. 1 the values of $\tau_{[2]}$, $\tau_{\langle 2\rangle}$, $\tau_{[3]}$ and $\tau_{\langle 3\rangle}$ are traced against the parameter of the distribution for the families of the compound exponential [B], the J-shaped [J], the symmetric [S] and the triangular [T] distributions.

There are some limits which are easily obtained.

In [S] $\lim_{p\to\infty}\tau_{[2]}=\frac{1}{4}$, $\lim_{p\to\infty}\tau_{[3]}=\frac{3}{8}$, $\lim_{p\to\infty}\tau_{\langle 3\rangle}=\frac{25-4\sqrt{7}}{36}$ and $\lim_{p\to 0}\tau_{\langle 3\rangle}=\frac{1}{3}$.

In [J] $\lim_{p\to\infty}\tau_{[2]}=\frac{1}{4}$, $\lim_{p\to\infty}\tau_{\langle 2\rangle}=\frac{5-\sqrt{5}}{8}$, $\lim_{p\to\infty}\tau_{[3]}=\frac{7}{18}$, $\lim_{p\to 0}\tau_{[2]}=0$ and $\lim_{p\to 0}\tau_{\langle 2\rangle}=\frac{1}{2}$.

In [T] $\lim_{\rho\to\infty}\tau_{[2]}=\frac{8}{25}$, $\lim_{\rho\to\infty}\tau_{\langle 2\rangle}=\frac{33-2\sqrt{66}}{50}$, $\lim_{\rho\to\infty}\tau_{[3]}=\frac{592}{1225}$, $\lim_{\rho\to\infty}\tau_{\langle 3\rangle}=\frac{9018-90\sqrt{438}}{11025}$ and $\max_{\rho}\tau_{[2]}=\frac{49}{150}$ when $\rho=2$.

In [W] $\lim_{b\to\infty}\tau_{[2]}=6\left(\frac{\log 2}{\pi}\right)^2$ and $\lim_{b\to 0}\tau_{[2]}=0$.

6. Examples of the distribution of $\bar{Y}_{[n]}$

In order to compare the distribution of our estimate $\bar{Y}_{[n]}$ with that of the usual sample mean $\bar{X}_n$ we shall give several examples.

Let $h_n(x)$ be the pdf of $\bar{Y}_{[n]}$ and let $g_n(x)$ be the pdf of $\bar{X}_n$.

(i) Let the pdf of the population be

$$f(x)=1\,,\qquad 0\leqq x\leqq 1\,. \tag{6.1}$$

Then, we have

$$h_2(x)=\begin{cases}\dfrac{16}{3}x^2(3-2x)\,, & 0\leqq x\leqq\dfrac{1}{2}\,,\\[2mm] \dfrac{16}{3}(1-x)^2(1+2x)\,, & \dfrac{1}{2}\leqq x\leqq 1\,,\end{cases} \tag{6.2}$$

and

(6.3) $h_3(x)$

$$=\begin{cases}\dfrac{9}{280}(3x)^5\{84-56(3x)+12(3x)^2-(3x)^3\}\,, & 0\leqq x\leqq\dfrac{1}{3}\,,\\[2mm] \dfrac{9}{280}\{1191-3888(3x)+4536(3x)^2-2268(3x)^3+630(3x)^4 & \\ \qquad -252(3x)^5+112(3x)^6-24(3x)^7+2(3x)^8\}\,, & \dfrac{1}{3}\leqq x\leqq\dfrac{2}{3}\,,\end{cases}$$

$$\left\{ \begin{array}{l} \frac{9}{280}(3(1-x))^5\{84-56(3(1-x)) \\ \qquad +12(3(1-x))^2-(3(1-x))^3\}\ , \qquad \frac{2}{3}\leqq x\leqq 1\ . \end{array} \right.$$

The pdf's g_2, g_3 are well known (see, for example, [1] p. 245). The pdf's h_2, g_2 are shown in Fig. 2 a. The pdf's h_3, g_3 are shown in Fig. 2 b.

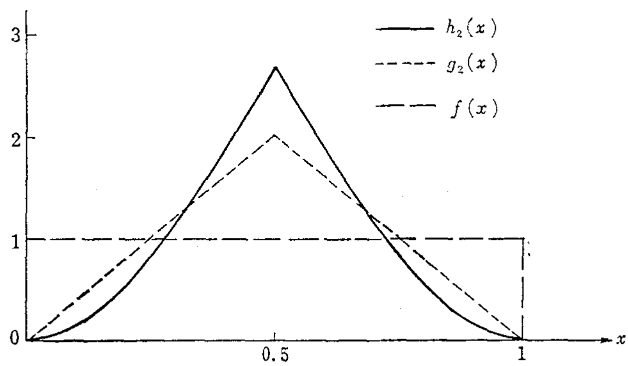

Fig. 2 a. h_2 and g_2 for the rectangular distribution

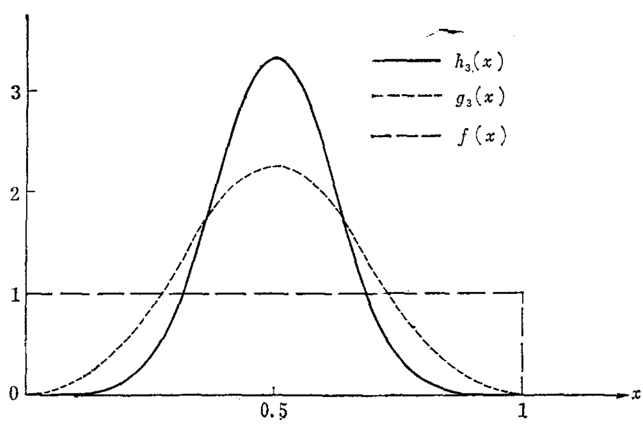

Fig. 2 b. h_3 and g_3 for the rectangular distribution

(ii) Let the pdf of the population be

(6.4) $$f(x)=e^{-x}\ , \qquad x>0\ .$$

Then

(6.5) $$h_2(x)=8(e^{-2x}-e^{-4x}(1+2x))\ , \qquad x>0\ ,$$

and

(6.6) $$g_2(x)=4x\,e^{-2x}\ , \qquad x>0\ .$$

The pdf's are shown in Fig. 2 c.

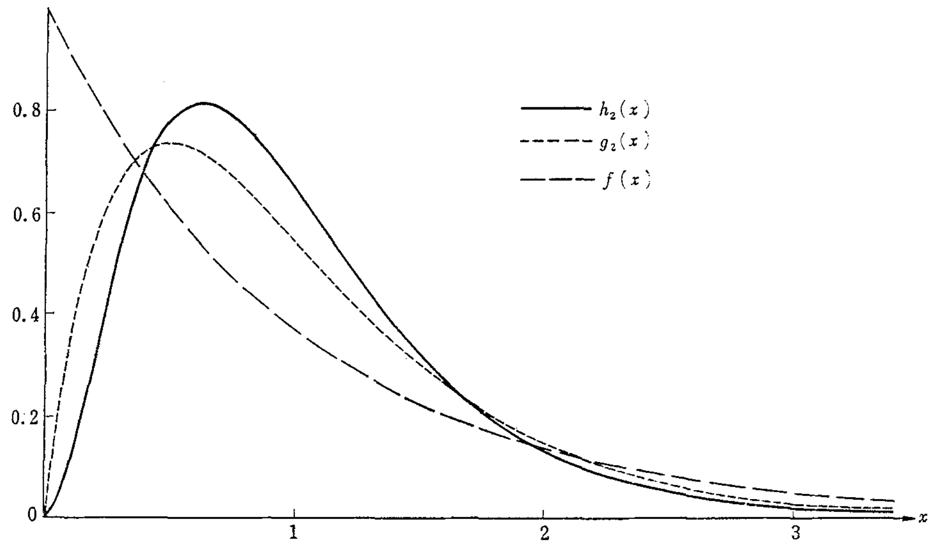

Fig. 2 c. h_2 and g_2 for the exponential distribution

(iii) Let the pdf of the population be

$$f(x)=2x\,, \qquad 0\leqq x\leqq 1\,. \tag{6.7}$$

Then

$$h_2(x)=\begin{cases} \dfrac{256}{35}(7x^5-4x^7)\,, & 0\leqq x\leqq \dfrac{1}{2}\,, \\ \dfrac{128}{35}(-272x^7+700x^6-644x^5+245x^4-35x^3+7x^2-1)\,, & \\ & \dfrac{1}{2}\leqq x\leqq 1\,, \end{cases} \tag{6.8}$$

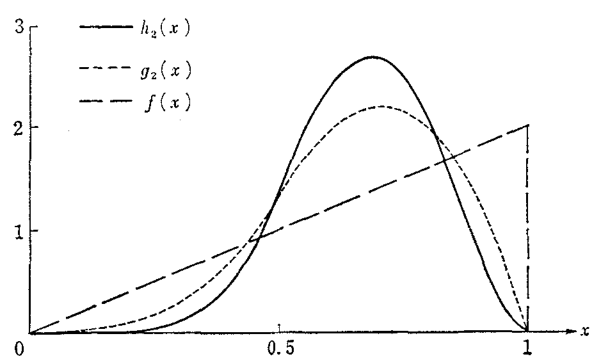

Fig. 2 d. h_2 and g_2 for the right-triangular distribution

and

$$g_2(x)=\begin{cases} \dfrac{32}{3}x^3\,, & 0\leqq x\leqq\dfrac{1}{2}\,, \\ -\dfrac{32}{3}x^3+16x-\dfrac{16}{3}\,, & \dfrac{1}{2}\leqq x\leqq 1\,. \end{cases} \tag{6.9}$$

The pdf's h_2, g_2 are shown in Fig. 2 d.

(iv) Let the pdf of the population be

$$f(x)=\frac{3}{2}x^2\,, \qquad -1<x<1\,. \tag{6.10}$$

Then

$$h_2(x)=\begin{cases} \dfrac{1}{770}(-2560x^{11}+40656x^5+79200x^4+69300x^3 \\ \qquad\qquad +40040x^2+13860x+2016)\,, & -1<x\leqq 0\,, \\ \dfrac{1}{770}(2560x^{11}-40656x^5+79200x^4-69300x^3 \\ \qquad\qquad +40040x^2-13860x+2016)\,, & 0\leqq x<1\,, \end{cases} \tag{6.11}$$

and

$$g_2(x)=\begin{cases} \dfrac{3}{20}(32x^5+80x^2+60x+12)\,, & -1<x\leqq 0\,, \\ \dfrac{3}{20}(-32x^5+80x^2-60x+12)\,, & 0\leqq x<1\,. \end{cases}$$

The pdf's h_2, g_2 are shown in Fig. 2 e.

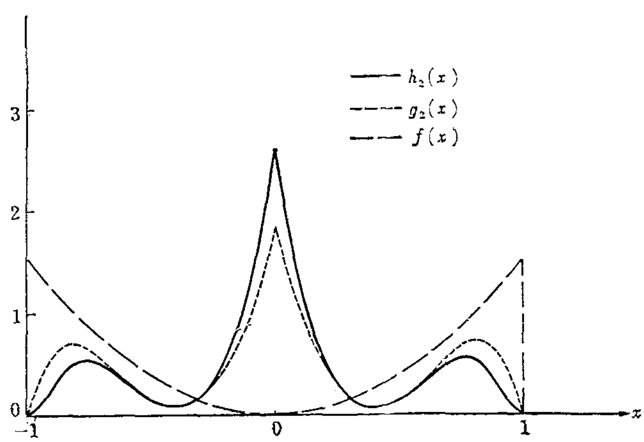

Fig. 2 e. h_2 and g_2 for the symmetric distribution

7. Sup $\tau_{[n]}$ and Sup $\tau_{\langle n\rangle}$

7.1 Sup $\tau_{[n]}$

From the tables in section 5, it will be found that the values of $\tau_{[n]}$ are, for each fixed n, concentrated in rather small range. It is known that $\tau_{[2]} \leqq 1/3$ for all continuous distributions with finite variances, and this supremum is attained by the rectangular distribution [3]. Thus we are led to the consideration of the supremum of $\tau_{[n]}$ for general n. We shall give the value of Sup $\tau_{[n]}$ which is attained by the rectangular distribution for general n.

The $\tau_{[n]}$ which we are going to maximize was given by

$$\tau_{[n]}=\left(\sigma^2-\frac{1}{n}\sum_{k=1}^{n}\sigma_{n,k}^2\right)\Big/\sigma^2=\frac{1}{n\sigma^2}\sum_{k=1}^{n}(\mu_{n,k}-\mu)^2 . \tag{7.1}$$

We can without loss of generality assume that

$$\mu=0 \qquad \text{and} \qquad \sigma^2=1 . \tag{7.2}$$

Then

$$\tau_{[n]}=\frac{1}{n}\sum_{k=1}^{n}\left(\int_0^1 a_{n,k}G(u)\,u^{k-1}(1-u)^{n-k}\,du\right)^2 , \tag{7.3}$$

where $G=G(u)=F^{-1}(u)$, i.e., the inverse function of the cdf $F(x)$.

In order to maximize (7.3) under the conditions (7.2), we put the first variation of

$$\frac{1}{n}\sum_{k=1}^{n}\left(\int_0^1 a_{n,k}G\cdot u^{k-1}(1-u)^{n-k}\,du\right)^2-2\lambda_1\int_0^1 G\,du-\lambda\left(\int_0^1 G^2\,du-1\right) \tag{7.4}$$

equal to zero. Then we obtain as the characteristic equation

$$\frac{1}{n}\sum_{k=1}^{n}a_{n,k}^2\gamma_{n,k}u^{k-1}(1-u)^{n-k}-\lambda_1-\lambda G=0 , \tag{7.5}$$

where

$$\gamma_{n,k}=\int_0^1 Gu^{k-1}(1-u)^{n-k}\,du . \tag{7.6}$$

Integrating the left-side of (7.5), it turns out to be

$$\lambda_1=0 . \tag{7.7}$$

Multiplying (7.5) by $G(u)$ and integrating, it turns out to be

$$\tau_{[n]}=\lambda . \tag{7.8}$$

From (7.5), $G(u)$ must be a polynomial of degree at most $n-1$. Let

$$G(u)=\sum_{j=0}^{n-1} a_j u^j \; . \tag{7.9}$$

Then from (7.5), (7.6), (7.7), (7.8) and (7.9), we have

$$\gamma_{n,k}=\sum_{j=0}^{n-1} a_j B(j+k,\, n-k+1) \; . \tag{7.10}$$

and

$$\lambda G=\sum_{j=0}^{n-1}\binom{n-1}{j}\left[\sum_{l=0}^{j}(-1)^{j-l}\binom{j}{l}\left\{\sum_{s=0}^{n-1}\frac{\binom{s+l}{l}}{\binom{n+s}{n}}a_s\right\}\right]u^j \; . \tag{7.11}$$

Hence we have the following equations;

$$\lambda a_j=\binom{n-1}{j}\sum_{s=0}^{n-1}\frac{a_s}{\binom{n+s}{n}}\left\{\sum_{l=0}^{j}(-1)^{j-l}\binom{j}{l}\binom{s+l}{l}\right\}\, , \tag{7.12}$$

$$(j=0,1,2,\cdots,n-1)\; .$$

Now it is easy to prove the following lemma (see, [12] p. 62).

LEMMA 1. *If a, b are non-negative integers, then*

$$\sum_{l=0}^{a}(-1)^{a-l}\binom{a}{l}\binom{b+l}{l}=\begin{cases}\binom{b}{a}\, , & a\leqq b\, ,\\ 0\;\;\, , & a>b\; .\end{cases} \tag{7.13}$$

PROOF. Let us consider the polynomial

$$\begin{aligned} p(x)&=(1+x)^b x^a=(1+x)^b\{(1+x)-1\}^a \\ &=(1+x)^b\left\{\sum_{l=0}^{a}\binom{a}{l}(-1)^{a-l}(1+x)^l\right\} \\ &=\sum_{l=0}^{a}(-1)^{a-l}\binom{a}{l}\left\{\sum_{h=0}^{b+l}\binom{b+l}{h}x^h\right\}\; . \end{aligned} \tag{7.14}$$

The coefficient of x^b is

$$\sum_{l=0}^{a}(-1)^{a-l}\binom{a}{l}\binom{b+l}{b}\; . \tag{7.15}$$

On the other hand, from the definition of $p(x)$ it must be equal to $\binom{b}{b-a}$, if $b\geqq a$, and 0, if $a>b$. This completes the proof of lemma 1.

By this lemma (7.12) becomes

$$\left(\lambda-\frac{\binom{n-1}{j}}{\binom{n+j}{n}}\right)a_j-\binom{n-1}{j}\sum_{s=j+1}^{n-1}\binom{s}{j}a_s\Big/\binom{n+s}{s}=0\,, \tag{7.16}$$

$$j=0,1,2,\cdots,n-1\,.$$

Let us denote $\binom{n-1}{j}\Big/\binom{n+j}{n}$ by ω_j, $(j=0,1,2,\cdots,n-1)$. Since $\omega_j=\binom{n-1}{j}\Big/\binom{n+j}{n}=\binom{2n-1}{n+j}\Big/\binom{2n-1}{n}$, we have

$$1=\omega_0>\omega_1>\cdots>\omega_{n-1}>0\,. \tag{7.17}$$

If $a_2=a_3=\cdots=a_{n-1}=0$, then $a_0<0$ and $a_1a_0\neq0$, since $G(u)$ must be the inverse function of the cdf $F(x)$ with zero mean. From (7.16) we can show that for a solution satisfying $a_{n-1}=a_{n-2}=\cdots=a_{j+1}=0$ and $a_j\neq0$ the corresponding λ must be ω_j. Hence the solution which maximizes the corresponding λ must be of the form

$$a_{n-1}=a_{n-2}=\cdots=a_2=0 \quad \text{and} \quad a_1a_0\neq0\,. \tag{7.18}$$

Thus we can conclude that

$$G(u)=a_0-2a_0u \tag{7.19}$$

and

$$\lambda=\frac{n-1}{n+1}\,. \tag{7.20}$$

From the condition (7.2), $a_0=-\sqrt{3}$. Thus, $G(u)=-\sqrt{3}+2\sqrt{3}\,u$ independently of n. Of course, $G(u)$ is the inverse function of the distribution function of the rectangular distribution.

Summarizing the above discussions, we have the following theorem

THEOREM 3.

$$\operatorname{Sup}\tau_{[n]}=\frac{n-1}{n+1}\,, \qquad \textit{for} \quad n\geqq2.$$

This supremum is attained by the rectangular distribution, where 'Sup' *is taken for all continuous distributions with the finite variance.*

7.2 $\operatorname{Sup}\tau_{\langle n\rangle}$

We shall start with proving the following lemmas.

LEMMA 2.

$$\operatorname{Sup}\tau_{\langle n\rangle}\geqq\frac{n-1}{n}$$

where 'Sup' *is taken for all continuous distributions with the finite variance.*

PROOF. Let $f(x)=px^{p-1}$, $0<x<1$, where $p>0$. This distribution has been discussed in 5.3 [J]. In this case, we have

$$\frac{\sigma_{n,k}^2}{\sigma^2}=(p+1)^2(p+2)\,p^{n-k}\left\{\frac{n!}{(k-1)!(np+2)((n-1)p+2)\cdots(kp+2)}\right.$$
$$\left.-p^{n-k+1}\left(\frac{n!}{(k-1)!(np+1)((n-1)p+1)\cdots(kp+1)}\right)^2\right\}.$$

Hence

$$\lim_{p\to 0}\frac{\sigma_{n,k}}{\sigma}=\begin{cases}0\ , & 1\leqq k\leqq n-1\ ,\\ \sqrt{n}\ , & k=n\ .\end{cases}$$

On the other hand, by the definition

$$\tau_{\langle n\rangle}=\frac{\sigma^2-\sigma_{\langle n\rangle}^2}{\sigma^2}=1-\left(\frac{\sigma_{n,1}+\sigma_{n,2}+\cdots+\sigma_{n,n}}{n\sigma}\right)^2.$$

Thus we have

$$\lim_{p\to 0}\tau_{\langle n\rangle}=1-\frac{1}{n}=\frac{n-1}{n}\ .$$

This completes the proof of Lemma 2.

LEMMA 3.

$$\tau_{\langle n\rangle}\leqq\frac{n-1}{n}\ ,\qquad n\geqq 2\ .$$

PROOF. Let $X_1, X_2, \cdots, X_n$ be a random sample of size n drawn from the population. Let $(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ be their order statistics. Since $X_1+X_2+\cdots+X_n=X_{(1)}+X_{(2)}+\cdots+X_{(n)}$, we have

$$\begin{aligned}n\sigma^2&=\sigma_{n,1}^2+\sigma_{n,2}^2+\cdots+\sigma_{n,n}^2+2\operatorname{Cov}(X_{(1)}, X_{(2)})+\cdots+2\operatorname{Cov}(X_{(n-1)}, X_{(n)})\\ &\leqq\sigma_{n,1}^2+\sigma_{n,2}^2+\cdots+\sigma_{n,n}^2+2\sigma_{n,1}\sigma_{n,2}+\cdots+2\sigma_{n,n-1}\sigma_{n,n}\\ &=(\sigma_{n,1}+\sigma_{n,2}+\cdots+\sigma_{n,n})^2=n^2\sigma_{\langle n\rangle}^2\ .\end{aligned}$$

Thus we have

$$\tau_{\langle n\rangle}=1-\frac{\sigma_{\langle n\rangle}^2}{\sigma^2}\leqq 1-\frac{1}{n}=\frac{n-1}{n}\ .$$

From Lemma 2 and Lemma 3 the following theorem has been proved.

THEOREM 4.

$$\operatorname{Sup} \tau_{\langle n \rangle} = \frac{n-1}{n}\,, \qquad n \geqq 2\,.$$

8. Related problems

8.1 A modification of sampling

The method mentioned above may be applied with various modifications. We shall here simply discuss only one example.

Let $X_1, X_2, \cdots, X_{10}$ be an independent random sample from a population. Let us define $Y_{(1)}$, $Y_{(2\,1)}$, $Y_{(2\,2)}$ and $\bar{Y}^*_{[3]}$ by

$$Y_{(1)} = \min\{X_1, X_2\}\,,$$

$$Y_{(2\,1)} = \min\left\{\max\{X_3, X_4\},\ \max\{X_5, X_6\}\right\}\,,$$

$$Y_{(2\,2)} = \max\left\{\max\{X_7, X_8\},\ \max\{X_9, X_{10}\}\right\}$$

and

$$\bar{Y}^*_{[3]} = \frac{Y_{(1)} + \dfrac{Y_{(2\,1)} + Y_{(2\,2)}}{2}}{2}\,.$$

Then $\bar{Y}^*_{[3]}$ is obviously an unbiased estimate of the population mean. Let us denote the pdf of the distribution of $Y_{(2\,i)}$ by $f_{(2\,i)}(x)$, $i=1, 2$, then

$$f_{(2\,1)}(x) = 2f_{2,2}(x)F_{2,2}(x)$$

and

$$f_{(2\,2)}(x) = f_{4,4}(x)\,.$$

The variance of $\bar{Y}^*_{[3]}$ is given by

$$\sigma^2(\bar{Y}^*_{[3]}) = \frac{1}{16}(4\sigma^2_{2,1} + \sigma^2(Y_{(2\,1)}) + \sigma^2_{4,4})\,.$$

Let us define $\sigma^{*2}_{[3]}$ and $\tau^*_{[3]}$ by $\sigma^{*2}_{[3]} = 3\sigma^2(\bar{Y}^*_{[3]})$ and $\tau^*_{[3]} = (\sigma^2 - \sigma^{*2}_{[3]})/\sigma^2$. It should be noted that $\bar{Y}^*_{[3]}$ is based on three measured observations. If $\sigma^2_{2,2}$ is considerably larger than $\sigma^2_{2,1}$, then it will be expected that such an estimate may be useful.

Suppose that the pdf of the distribution of the population be e^{-x}, $(x>0)$. Then we have

$$\mu_{2,2} = \frac{3}{2}\,, \qquad \mu_{4,4} = \frac{25}{12}\,, \qquad \mathrm{E}(Y_{(2\,1)}) = 2\mu_{2,2} - \mu_{4,4} = \frac{11}{12}\,,$$

$$\sigma_{2,1}^2=\frac{1}{4} \qquad \text{and} \qquad \sigma_{2,2}^2=\frac{5}{4} .$$

Hence

$$\sigma_{[3]}^{*2}=\frac{3}{16}\left(4\sigma_{2,1}^2+2\sigma_{2,2}^2-\frac{(\mu_{4,4}-\mathrm{E}(Y_{(2\,1)}))^2}{2}\right)=\frac{203}{384}$$

and

$$\tau_{[3]}^{*}=\frac{181}{384}=0.471 .$$

Thus our estimates in the exponential population satisfy the following inequalities;

$$\tau_{[2]}(=0.250)<\tau_{\langle 2\rangle}(=0.345)<\tau_{[3]}(=0.389)$$
$$<\tau_{[3]}^{*}(=0.471)<\tau_{[4]}(=0.479)<\tau_{\langle 3\rangle}=0.510 .$$

8.2 Note on Neyman-allocation

When we want to apply the so-called Neyman-allocation we may have to estimate the variances $\sigma_{n,1}^2, \sigma_{n,2}^2, \cdots, \sigma_{n,n}^2$ and sometimes we may have to use the approximate values in practical cases. For this reason we shall simply consider the quantity $\tau_{\langle n\rangle}^{*}$ defined in the following. Let us define $\sigma_{\langle n\rangle}^{*2}$ by

$$\sigma_{\langle n\rangle}^{*2}=\frac{1}{n}\left(\frac{\sigma_{n,1}^2}{n\beta_1}+\frac{\sigma_{n,2}^2}{n\beta_2}+\cdots+\frac{\sigma_{n,n}^2}{n\beta_n}\right) ,$$

where $\beta_1+\beta_2+\cdots+\beta_n=1$ and each $\beta_i>0$. Let us define $\tau_{\langle n\rangle}^{*}$ by

$$\tau_{\langle n\rangle}^{*}=\frac{\sigma^2-\sigma_{\langle n\rangle}^{*2}}{\sigma^2} .$$

(i) For simplicity suppose that $n=2$ and $\sigma_{2,1}^2=\sigma_{2,2}^2$. Then we have

$$\tau_{\langle 2\rangle}^{*}=1-\frac{1}{4\beta_1}\left(\frac{\sigma_{2,1}}{\sigma}\right)^2-\frac{1}{4\beta_2}\left(\frac{\sigma_{2,2}}{\sigma}\right)^2 .$$

Let $\frac{\sigma_{2,1}}{\sigma}=\frac{\sigma_{2,2}}{\sigma}=\gamma$ and let $\beta_1=p=1-\beta_2$. Then

$$\tau_{\langle 2\rangle}^{*}=\frac{1}{p(1-p)}\left\{p(1-p)-\frac{(1-p)\gamma^2}{4}-\frac{p\gamma^2}{4}\right\}$$
$$=\frac{1}{p(1-p)}\left(p(1-p)-\frac{\gamma^2}{4}\right) .$$

Hence $\tau_{\langle 2\rangle}^{*}\geqq 0$ is equivalent to

$$\frac{1}{2}-\frac{1}{2}\sqrt{1-\gamma^2} \leqq p \leqq \frac{1}{2}+\frac{1}{2}\sqrt{1-\gamma^2} \ .$$

It should be noted that $\frac{2}{3}\leqq\gamma^2\leqq1$. For example in the normal population $\gamma^2=0.68169$. Hence $\tau^*_{\langle 2\rangle}\geqq0$ is equivalent to $0.218\leqq p\leqq0.782$.

(ii) Suppose that $n=2$ and the population has the exponential distribution. Let

$$\beta_1=\frac{1}{1+\rho}=1-\beta_2 \, .$$

Then

$$\tau^*_{\langle 2\rangle}=\frac{10\rho-\rho^2-5}{16\rho} \ .$$

Hence $\tau^*_{\langle 2\rangle}\geqq0$ is equivalent to

$$5-2\sqrt{5} \leqq\rho\leqq5+2\sqrt{5} \ ,$$

and

$$\tau^*_{\langle 2\rangle}\geqq\tau_{[2]}$$

is equivalent to $1\leqq\rho\leqq5$.

In Fig. 3 this $\tau^*_{\langle 2\rangle}$ is traced against ρ.

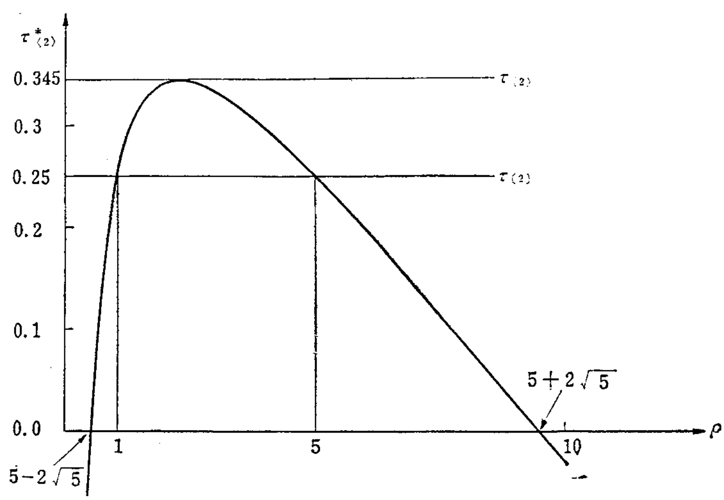

Fig. 3. $\tau^*_{\langle 2\rangle}$ for the exponential distribution

REFERENCES

[1] H. Cramer, *Mathematical Methods of Statistics*, Princeton, 1961.

[2] F. Downton, "Least-squares estimates using ordered observations," *Ann. Math. Statist.*, 25 (1954), 303-316.

[3] E. J. Gumbel, "The maxima of the mean largest value and of the range," *Ann. Math. Statist.*, 25 (1954), 76-84.

[4] S. S. Gupta, "Order statistics from the gamma distribution," *Technometrics*, 2 (1960), 243-262.

[5] C. Hastings, Jr., F. Mosteller, J. W. Tukey and C. P. Winsor, "Low moments for small samples: a comparative study of order statistics," *Ann. Math. Statist.*, 18 (1947), 413-426.

[6] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Charles Griffin, London, 1958.

[7] E. L. Lehmann, "Some concepts of dependence," *Ann. Math. Statist.*, 37 (1966), 1137-1153.

[8] J. Liebelein, "On the exact evaluation of the variances and covariances of order statistics in samples from the extreme-value distribution," *Ann. Math. Statist.*, 24 (1953), 282-287.

[9] J. Liebelein, "On moments of order statistics from the Weibull distribution," *Ann. Math. Statist.*, 26 (1955), 330-333.

[10] A. E. Sarhan, "Estimation of the mean and standard deviation by order statistics," *Ann. Math. Statist.*, 25 (1954), 317-328.

[11] A. E. Sarhan and B. E. Greenberg, *Contributions to Order Statistics*, John Wiley, 1962.

[12] W. Feller, *An Introduction to Probability Theory and Its Applications*, 1, John Wiley, 1957.