

## SMALL SAMPLE PROPERTIES OF A "SOMETIMES POOL ESTIMATE" IN THE NONPARAMETRIC CASE

Tamura, Ryoji  
Shimane University

<https://doi.org/10.5109/13027>

---

出版情報：統計数理研究. 12 (3/4), pp.75-83, 1967-03. Research Association of Statistical Sciences

バージョン：

権利関係：



# SMALL SAMPLE PROPERTIES OF A "SOMETIMES POOL ESTIMATE" IN THE NONPARAMETRIC CASE.

By

Ryoji TAMURA

(Received June 7th, 1965)

§ 1. **Introduction and summary.** The author has discussed in [2], [3] and [4] some large sample properties of the "sometimes pool estimate" ("s.p.e." in brief) for the population median, shift and spread parameters of the population in the nonparametric circumstances where the type of the underlying distribution is unknown. Saying in more detail, the "s.p.e." for the above parameters have been defined and their asymptotic distributions, expected values and mean square errors have been derived. And the asymptotic efficiency against the "never pool estimate" ("n.p.e." in brief) has been also discussed under the normal assumption.

In order that these procedures may be applied for the practical purpose in the nonparametric cases, it needs in addition to investigate their properties when the sample size is small. Thus this paper is concerned with the "s.p.e." for the population median in the small sample cases. Unfortunately we have no techniques to develop the theory in the general forms available for arbitrary sample sizes and/ or level of significance in the preliminary nonparametric test. Therefore we are in the situation that we must try to derive the results respectively for each value of the sample size ( $m, n$ ) and/ or the level of significance  $\alpha$ .

In this paper, we shall first derive the distribution of the "s.p.e." in the case that  $m=3, n=2, \alpha=0.2$ . The methods of such derivation are also applied in the other cases. We investigate in detail under the uniform distribution the effects of  $\alpha$  on the efficiency with respect to the "n.p.e." when  $m=3, n=4$ . We shall also compute the efficiency in the cases (a) normal distribution and  $\Delta=0, m=3, n=4$  (b) the double exponential distribution,  $m=3, n=2$ . All the results will be compared with those in the large sample cases.

Though we only deal with some properties of the "s.p.e." under the very special circumstances where the sample size is extremely small, we may conclude that the behaviour of the "s.p.e." in the small sample cases are similar to those in the large sample cases.

§ 2. **Procedures.** Let  $\mathbf{X}=(X_1, \dots, X_m)$  and  $\mathbf{Y}=(Y_1, \dots, Y_n)$  be respectively the random samples of size  $m, n$  from the continuous distributions  $F(x)$  and  $F(x-\Delta)$ , where  $F(\xi)=0.5$  and  $\Delta \geq 0$ . Our main object is to estimate the value of the median  $\xi$  and for this purpose, we first test the hypothesis

$d=0$ . The preliminary test is done by the Wilcoxon two sample test and the sample medians are used as the estimator of  $\xi$ . There exists another more efficient estimator than the sample median, say the median of the average of pair of the observations. However we adopt the sample median for the sake of simplicity, for our main purpose is to compare the "s.p.e." with the "n.p.e."

Let now the critical region of size  $\alpha$  be  $R_\alpha$ . Then the "s.p.e." is defined as follows,

$$(1) \quad \begin{aligned} (i) \quad \hat{\xi} &= \text{median}(\mathbf{X}) && \text{when } (\mathbf{X}, \mathbf{Y}) \in R_\alpha \\ (ii) \quad \hat{\xi} &= \text{median}(\mathbf{X}, \mathbf{Y}) && \text{when } (\mathbf{X}, \mathbf{Y}) \in R_\alpha^c \end{aligned}$$

**§ 3. The distributions of the "s.p.e."** Since we cannot derive the distributions of the "s.p.e." in the forms available for arbitrary values of  $m$ ,  $n$  and  $\alpha$ , we shall in 3.1 deal with the derivation of the distribution in the case  $m=3$ ,  $n=2$  and  $\alpha=0.2$ . The similar methods may be used in the other cases of  $m$ ,  $n$  and  $\alpha$ , so that we only denote in 3.2 the densities of  $\hat{\xi}$  in the case  $m=3$ ,  $n=4$  and uniform  $F(x)$ .

### 3.1. The case $m=3$ , $n=2$ and $\alpha=0.2$ .

Now we arrange three  $X$ 's and two  $Y$ 's in ascending order and then we may get ten different orderings. For example, we write one of them as  $XXYXY$  which means that any two  $X$ 's among three  $X$ 's are less than  $Y$ 's and the remaining  $X$  is put between two  $Y$ 's. Our preliminary test is to reject the hypothesis  $d=0$  when the rank sum of  $Y$ 's in the combined sample  $(\mathbf{X}, \mathbf{Y})$  is too large. Thus the critical region  $R_\alpha$  of size 0.2 is the region determined by the following orderings

$$(2) \quad XXXYY, XXYXY.$$

In this case, the estimate for  $\xi$  is respectively given by the value of the underlined  $X$  in (2). If  $(\mathbf{X}, \mathbf{Y})$  is in the region of the following orderings

$$(3) \quad \begin{array}{cccc} XYXYX & XXYXY & YXXXY & XYXYX \\ YXXYX & XYXYX & YXYXX & YXXXX \end{array}$$

, the hypothesis  $d=0$  is accepted and the estimate of  $\xi$  is respectively given by the underlined  $X$  or  $Y$  in (3). Denote the density of the estimator by  $g(z)$  and we get the following form from (2) and (3)

$$\begin{aligned} g(z)dz &= Pr\left[z < X < z+dz, \quad X' < X < \underset{Y'}{X''} < Y\right] \\ &+ Pr\left[z < X < z+dz, \quad \underset{Y}{X'} < X < \underset{Y'}{X''}\right] + Pr\left[z < X < z+dz, \quad \underset{Y'}{Y} < X < \underset{X''}{X'}\right] \\ &+ Pr\left[z < Y < z+dz, \quad \underset{Y'}{X'} < Y < \underset{X''}{X}\right] + Pr\left[z < Y < z+dz, \quad \underset{X}{X'} < Y < Y' < X''\right]. \end{aligned}$$

The above probabilities are easily calculated and the density  $g(z)$  may be expressed as follows,

$$(4) \quad g(z) = 12F(z)f(z) \int_z^\infty \{F(x) - F(z)\} \{F(x-D) - F(z-D)\} dF(x-D) \\ + 12F(z)F(z-D) \{1 - F(z)\} \{1 - F(z-D)\} f(z) + 3F(z-D)^2 \{1 - F(z)\}^2 f(z) \\ + 6F(z)F(z-D) \{1 - F(z)\}^2 f(z-D) + 6F(z)^2 f(z-D) \int_z^\infty \{1 - F(x)\} dF(x-D)$$

, where  $f(z) = F'(z)$ .

### 3.1. a Uniform distribution.

Assume that  $F(x)$  be uniform over  $[0, 1]$  and we get from (4)

$$(5) \quad g(z) = \begin{cases} 6z(1-z) - 2(1-D)^3 z, & 0 \leq z \leq D \\ 4z(1-z)^3 + 24(z-D)^2 + 36D(z-D)(1-z)^2 + 12D^2(1-z)^2 \\ + 12D(z-D)^2(1-z) + 18D^2(z-D)(1-z) + 6D^3(1-z) \\ + 3D^2(1-z)^2 \varepsilon(z), & D \leq z \leq 1 \end{cases}$$

, where  $\varepsilon(z) = 0(1)$  for  $z = D(>D)$ .

### 3.1. b Double exponential distribution.

Assume that  $F(x)$  be double exponential, that is

$$f(x) = \frac{1}{2} \exp(-|x|), \quad -\infty < x < \infty.$$

Then we may get from (4)

$$(6) \quad g(z) = \begin{cases} \frac{1}{2} e^{2z} (3 - 1.25e^{-D} + 0.25e^{-2D}) + \frac{3}{4} e^{3z} (-1 + 3e^{-2D} + 0.75De^{-2D}) \\ - \frac{9}{4} e^{4z-2D} + \frac{5}{8} e^{5z-2D}, & z \leq 0 \\ \frac{3}{4} e^{-2D} - \frac{5}{4} e^{-z-D} - \frac{3}{8} (1 + 0.5D - 0.5z) e^{-z-2D} + \frac{3}{2} e^{-2z} + \frac{5}{8} e^{-2z-D} \\ - \frac{3}{4} e^{-3z} + \frac{3}{4} (0.5 + D - z) e^{z-2D}, & 0 \leq z \leq D \\ \frac{3}{4} (0.5 + 3e^D + 0.5e^{2D}) e^{-3z} - \frac{3}{2} e^{-4z+D} - \frac{5}{4} e^{-4z+2D} + \frac{5}{8} e^{-5z+2D}, & D \leq z. \end{cases}$$

### 3.2. The cases $m=3, n=4$ .

By the same methods as 3.1., we may derive the density  $h_\alpha(z)$  where  $\alpha$  means the level of significance in the preliminary test. We write only the forms of the density in the case of uniform  $F(x)$  over  $[0, 1]$ .

$$\begin{aligned}
 h_{.057}(z) = & \left\{ \begin{aligned} & 2.4(1-D^5)z - 6(1-D)D^4z - 6z(z-D), \quad 0 \leq z \leq D \\ & 132(z-D)^3(1-z)^3 + 216D(z-D)^3(1-z)^2 + 96D^2(z-D)^3(1-z) \\ & + 8D^3(z-D)^3 + 288D^2(z-D)^2(1-z)^2 + 156D(z-D)^2(1-z)^3 \\ & + 144D^3(z-D)^2(1-z) + 96D^3(z-D)(1-z)^2 + 60D^2(z-D)(1-z)^3 \\ & + 12D^4(z-D)^2 + 42D^4(z-D)(1-z) + 2.4(z-D)(1-z)^5 \\ & + 12D(z-D)(1-z)^4 + 2.4D(1-z)^5 + 12D^2(1-z)^4 + 24D^3(1-z)^3 \\ & + 24D^4(1-z)^2 + 6D^5(1-z), \quad D \leq z \leq 1 \end{aligned} \right. \\
 (7)
 \end{aligned}$$

$$\begin{aligned}
 h_{.114}(z) = & \left\{ \begin{aligned} & \{6(1-D) - 2.4(1-D)^5 - 6D(1-D)^4\}z - 6z(z-D), \quad 0 \leq z \leq D \\ & 124(z-D)^3(1-z)^3 + 192D(z-D)^3(1-z)^2 + 72D^2(z-D)^3(1-z) \\ & + 4D^3(z-D)^3 + 3(z-D)^2(1-z)^4 + 144D(z-D)^2(1-z)^3 \\ & + 234D^2(z-D)^2(1-z)^2 + 84D^3(z-D)^2(1-z) + 24D(z-D)(1-z)^4 \\ & + 84D^2(z-D)(1-z)^3 + 96D^3(z-D)(1-z)^2 + 30D^4(z-D)(1-z) \\ & + 3.6(z-D)(1-z)^5 + 3.6D(1-z)^4 + 18D^2(1-z)^4 + 36D^3(1-z)^3 \\ & + 24D^4(1-z)^2 + 6D^5(1-z), \quad D \leq z \leq 1 \end{aligned} \right. \\
 (8)
 \end{aligned}$$

$$\begin{aligned}
 h_{.2}(z) = & \left\{ \begin{aligned} & 6(1-z)z - 1.2(1-D)^5z - 6z(z-D), \quad 0 \leq z \leq D \\ & 112(z-D)^3(1-z)^3 + 156D(z-D)^3(1-z)^2 + 48D^2(z-D)^3(1-z) \\ & + 9(z-D)^2(1-z)^4 + 144D(z-D)^2(1-z)^3 + 198D^2(z-D)^2(1-z)^2 \\ & + 60D^3(z-D)^2(1-z) + 4.8(z-D)(1-z)^5 + 36D(z-D)(1-z)^4 \\ & + 96D^3(z-D)(1-z)^2 + 30D^4(z-D)(1-z) + 4.8D(1-z)^5 \\ & + 24D^2(1-z)^4 + 96D^2(z-D)(1-z)^3 + 36D^3(1-z)^3 + 24D^4(1-z)^2 \\ & + 6D^5(1-z), \quad 0 \leq z \leq 1 \end{aligned} \right. \\
 (9)
 \end{aligned}$$

$$\begin{aligned}
 h_{.429}(z) = & \left\{ \begin{aligned} & 6z(1-z), \quad 0 \leq z \leq D \\ & 92(z-D)^3(1-z)^3 + 120D(z-D)^3(1-z)^2 + 21(z-D)^2(1-z)^4 \\ & + 144D(z-D)^2(1-z)^3 + 180D^2(z-D)^2(1-z)^2 \\ & + 96D^3(z-D)^2(1-z) + 6(z-D)(1-z)^5 + 48D(z-D)(1-z)^4 \\ & + 108D^2(z-D)(1-z)^3 + 96D^3(z-D)(1-z)^2 + 30D^4(z-D)(1-z) \\ & + 6D(1-z)^5 + 36D^3(1-z)^3 + 24D^2(1-z)^4 + 24D^4(1-z)^2 \\ & + 6D^5(1-z), \quad D \leq z \leq 1 \end{aligned} \right. \\
 (10)
 \end{aligned}$$

#### § 4. The expected values and mean square errors.

##### 4.1. The case $m=3$ , $n=4$ and uniform $F(x)$ .

We shall first calculate the expected values and mean square errors about 0.5 of the "s.p.e." in the uniform  $F(x)$  over  $[0, 1]$  when  $m=3$ ,  $n=4$ . The computations are elementary by using (7)~(10) in the previous section.

$\alpha$	$E(\hat{\xi})$
0.057	$0.4857 + 0.4571\Delta - 0.4\Delta^2 - 0.6\Delta^3 + 0.8\Delta^5 + 0.4\Delta^6 - 0.8857\Delta^7 + 0.2428\Delta^8$
0.114	$0.475 + 0.3714\Delta - 0.6\Delta^2 - 1.4\Delta^3 + 5.75\Delta^4 - 8.6\Delta^5 + 7.2\Delta^6 - 3.4\Delta^7 + 0.7036\Delta^8$
0.2	$0.4607 + 0.2857\Delta - 0.6\Delta^2 + 1.75\Delta^4 - 2.8\Delta^5 + 2\Delta^6 - 0.6857\Delta^7 + 0.0893\Delta^8$
0.429	$0.4393 + 0.2571\Delta - 1.5\Delta^2 + 8.8\Delta^3 - 26.25\Delta^4 + 41.4\Delta^5 - 35.9\Delta^6 - 16.2857\Delta^7 - 3.0321\Delta^8$
$\alpha$	M.S.E. ( $\hat{\xi}$ )
0.057	$0.0310 + 0.0143\Delta + 0.2857\Delta^2 - 0.6667\Delta^3 + 3.8\Delta^4 - 14.6\Delta^5 + 25.5333\Delta^6 - 22.7143\Delta^7 + 10.2857\Delta^8 - 1.9190\Delta^9$
0.114	$0.0329 + 0.0214\Delta + 0.2286\Delta^2 + 0.6\Delta^3 - 7.25\Delta^4 + 20.4\Delta^5 - 28.8867\Delta^6 + 22.9143\Delta^7 - 9.7964\Delta^8 + 1.7659\Delta^9$
0.2	$0.0353 + 0.0214\Delta + 0.1714\Delta^2 - 0.4667\Delta^3 - 0.05\Delta^4 + 1.6\Delta^5 - 2.5333\Delta^6 + 1.8286\Delta^7 - 0.6464\Delta^8 + 0.0897\Delta^9$
0.429	$0.0385 + 0.0071\Delta + 0.4429\Delta^2 - 4.3333\Delta^3 + 19.55\Delta^4 - 47.8\Delta^5 + 67.4333\Delta^6 - 55.0857\Delta^7 + 24.2678\Delta^8 - 4.4706\Delta^9$

After some computations, we get the following Table I. The mean square error of the median of size 3 in the uniform distribution over  $[0, 1]$ , that is the "n.p.e." in our case, is given by 0.05. Therefore we may also compute the efficiency  $e(s, n)$  of the "s.p.e." relative to the "n.p.e."

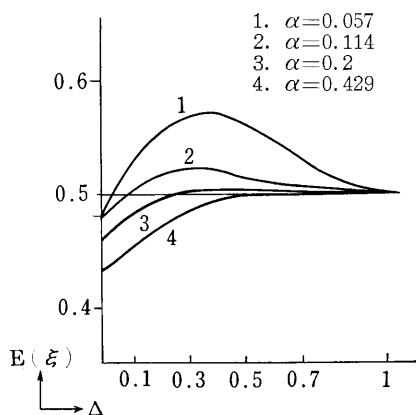
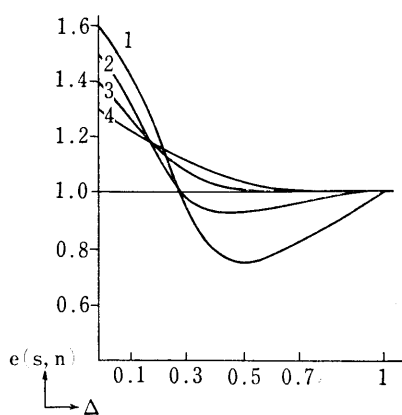
Table I.  $E(\hat{\xi})$ 

$\Delta$						
$\alpha \backslash \Delta$	0	0.1	0.3	0.5	0.7	1
0.057	0.4857	0.5268	0.5727	0.5645	0.5283	0.5
0.114	0.475	0.5052	0.5249	0.5150	0.5036	0.5
0.2	0.4607	0.4834	0.5011	0.5017	0.5002	0.5
0.429	0.4393	0.4565	0.4842	0.5004	0.5001	0.5

Table II.  $e(s, n)$ 

$\Delta$						
$\alpha \backslash \Delta$	0	0.1	0.3	0.5	0.7	1
0.057	1.616	1.435	0.952	0.765	0.837	1
0.114	1.518	1.337	0.994	0.938	0.979	1
0.2	1.416	1.292	1.074	1.005	0.999	1
0.429	1.299	1.225	1.129	1.029	1.001	1

These values are also shown in Figure I and II to compare with the results in the large sample cases which are discussed later.

Figure I.  $E(\hat{\xi})$ Figure II.  $e(s, n)$ 

We may consider from these Tables or Figures as follows,

(1) If there exist informations that the value of  $\Delta$  is very small, it is preferable to choose the value of  $\alpha$  small,

(2) In the cases that we have not such the informations as the above, it is wise that we adopt about 0.1~0.4 as the value of  $\alpha$ .

#### 4.2. The case $m=3$ , $n=4$ , $\Delta=0$ and normal $F(x)$ .

It is difficult to investigate the behaviour of  $\hat{\xi}$  in the case that  $F(x)$  is the standard normal  $\Phi(x)$  with density  $\varphi(x)$ , but it is easy if  $\Delta=0$ . Their densities are respectively expressed as follows,

$$(11) \quad h_{.057}(z) = \frac{33}{35} p(z; 4, 7) + \frac{2}{35} p(z; 2, 7)$$

$$(12) \quad h_{.114}(z) = \frac{31}{35} p(z; 4, 7) + \frac{3}{35} p(z; 2, 7) + \frac{1}{35} p(z; 3, 7)$$

$$(13) \quad h_{.2}(z) = \frac{28}{35} p(z; 4, 7) + \frac{4}{35} p(z; 2, 7) + \frac{3}{35} p(z; 3, 7)$$

$$(14) \quad h_{.429}(z) = \frac{23}{35} p(z; 4, 7) + \frac{5}{35} p(z; 2, 7) + \frac{7}{35} p(z; 3, 7)$$

, where  $p(z; i, 7)$  denotes the density of the  $i$  th smallest of the sample of size 7 from the standard normal distribution, that is

$$p(z; i, 7) = \frac{7!}{(i-1)!(7-i)!} \Phi^{i-1}(z) \{1 - \Phi(z)\}^{7-i} \varphi(z).$$

These  $h(z)$  are respectively expressed as some contaminations of the densities  $p(z; i, 7)$ , so that we may easily compute the expected values and mean square errors by using the table of the order statistics from the normal distribution, for example [1].

**Table III.** Normal case,  $\Delta=0$ 

$\alpha$	0.057	0.114	0.2	0.429
$E(\hat{\xi})$	-0.0433	-0.0750	-0.1168	-0.1787
$e(s, n)$	1.822	1.675	1.530	1.375

**4.3. The case  $m=3$ ,  $n=2$ ,  $\alpha=0.2$  and double exponential  $F(x)$ .**

In this case, we may easily get from (6) the following

$$(15) \quad E(\hat{\xi}) = -2.7708e^{-\Delta} + 1.25\Delta e^{-\Delta} + 2.7563e^{-2\Delta} + 1.4375\Delta e^{-2\Delta} \\ + 0.375\Delta^2 e^{-2\Delta} - 0.1e^{-3\Delta},$$

$$(16) \quad \text{M.S.E.}(\hat{\xi}) = 0.6389 + 2.7778e^{-\Delta} - 3.6667\Delta e^{-\Delta} + 1.25\Delta^2 e^{-\Delta} - 2.4848e^{-2\Delta} \\ + 2.2604\Delta e^{-2\Delta} + 0.9375\Delta^2 e^{-2\Delta} + 0.25\Delta^3 e^{-2\Delta} - 0.4848e^{-3\Delta} \\ - 0.2\Delta e^{-3\Delta}.$$

**Table IV.**  $E(\hat{\xi})$ 

$\Delta$	0	0.5	1	1.5	2	3	4
$E(\hat{\xi})$	-0.1146	-0.011	0.054	0.086	0.094	0.075	0.046

The numerical values of the mean square errors are given in Figure IV.

**§ 5. Comparison with the large sample cases.** We have derived in [2] the asymptotic efficiency of the "s.p.e." when the sample size are large. The results are given by the following,

$$(17) \quad E(\hat{\xi}) = \xi + (1-\lambda)r\Phi(k)/\sqrt{N} - \sqrt{3(1-\lambda)}\sigma\varphi(k)/2\sqrt{m}$$

$$(18) \quad \text{M.S.E.}(\hat{\xi}) = \frac{\sigma^2}{m}\{1-\Phi(k)\} + \frac{\sigma^2}{N}\Phi(k) + \frac{3(1-\lambda)}{4m}\sigma^2 k\varphi(k) \\ + \frac{r^2}{N}(1-\lambda)^2\Phi(k)$$

, where  $\Delta = r/\sqrt{N}$  ( $r \geq 0$ ),  $m+n=N$ ,  $m/N \rightarrow \lambda$

$$k = z_\alpha - \sqrt{12\lambda(1-\lambda)}r \int_{-\infty}^{\infty} f^2(x)dx, \quad \sigma^2 = 1/4f^2(\xi), \\ 1-\Phi(z_\alpha) = \alpha.$$

First we investigate the effects of  $\alpha$  on the asymptotic efficiency in the case  $\lambda=3/7$  and uniform  $F(x)$  which corresponds to the case  $m=3$ ,  $n=4$  in the small sample case. Since the mean square error of the "n.p.e." is given by  $\sigma^2/m$ , the relative asymptotic efficiency may be expressed as fol-



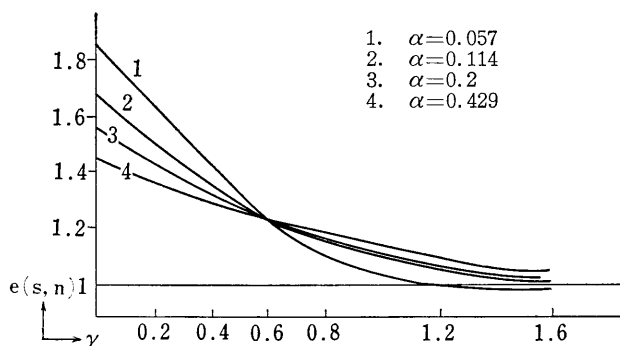
lows,

$$(19) \quad e(s, n) = [1 + (0.5598r^2 - 0.5714)\Phi(k) + 0.4286k\varphi(k)]^{-1},$$

$$k = z_\alpha - 1.7143r.$$

We get the following Figure III.

Figure III. Efficiency



From Figures II and III, we may see that the behaviour of the efficiency in the small and large sample cases is similar.

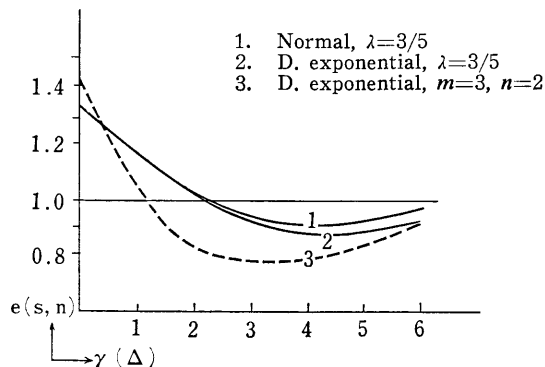
Lastly we compute the efficiency in the cases  $\lambda=3/5$ ,  $\alpha=0.2$  and double exponential or normal  $F(x)$ . In this case, the relative efficiency is given by the following

$$(20) \quad e(s, n) = [1 + (0.096r^2 - 0.4)\Phi(k) + 0.3k\varphi(k)]^{-1}$$

, where  $k = \begin{cases} 0.8416 - 0.4243r & \text{for d. exponential} \\ 0.8416 - 0.4787r & \text{for normal.} \end{cases}$

These results are denoted in Figure IV.

Figure IV. Efficiency



**References**

- [1] Sarhan, A. E. and Greenberg, B. G. *Contributions to order statistics*. John Wiley & Sons, Ins. (1962).
- [2] Tamura, R. *Nonparametric inferences with a preliminary test*. Bull. Math. Stat. Vol. 11 38–61 (1965).
- [3] Tamura, R. *Some estimate procedures with a nonparametric preliminary test. I*. Bull. Math. Stat. Vol. 63–7 1(1965).
- [4] Tamura, R. *Some estimate procedures with a nonparametric preliminary test. II*. (submitted to Bull. Math. Stat.)

Shimane University