On the Ratioes between Subranges

by Ryoji TAMURA

(*Received Nov* 30. 1956)

田村亮二 ： 部分範囲間の比について

§ 1. Introduction.

In recent years the statistic relating to range have been used as the test functions or estimaters for the theories of the statistical inference. Especially in the processes of mass production, they are indispensable for simplicity and rapidity of computation in spite of less efficiency. In [1], [2] Dixon proposed the use of the ratio involving extreme values to test the outlying observation in normal type. On the whole the rejection tests in normal type have been investigated by many research workers, and good results have been realized. On the other hand in exponential type it was only M. Matsuyama in [3] who argued by making use of χ^2-distribution.

This paper investigates the properties of the statistic which have suggested by Dixon as test functions ahd here in paticular the statistic coming from exponential distribution are treated in detail.

§ 2. R-statistic and its distribution function.

(1) Exponential case.

Let $X_1 < X_2 < \cdots < X_n$ be the ordered sample of size n from the density function :

$$f(x)=\begin{cases} \frac{1}{\sigma} exp\{-\frac{1}{\sigma}(x-a)\} & for \quad x \geqq a \\ o & for \quad x < a \end{cases} \qquad (1)$$

and $F(x)=\int_o^x f(t)dt$ *is the distribution function.*

Then we call the ratio between subranges by name of R-statistic for a time. There will be no loss in generality by considering X to have drawn from a standart exponential distribution :

$$f(x)=\begin{cases} exp(-x) & for \quad x \geqq o \\ o & for \quad x < o \end{cases} \qquad (1)'$$

,since the R-statistic is the ratio of two difference.

Thus we may formulate our problem as follows : Let X be a random variable with the probability density function (1)' and $X_1 < X_2 < \cdots\cdots < X_n$ be the ordered sample, then it is what are the sampling distribution of R- statistic. Moreover we proceed to research their asymptotic properties and calculate the probability tables that are necessary for the test. Now we consider the R-statistic in general form

$$R=\frac{X_n-X_m}{X_n-X_i} \qquad (n>m>i)$$

and if we accept i=1,2,3; $m=u-1, n-2$ as special case, they are ultilized for practical application.

The joint probability density function (henceforth we set the j. p. d. f. for brevity) of X_i, X_m and X_n is given by

$$\frac{n\,!}{(i-1)\,!\,(m-i-1)\,!\,(n-m-1)\,!}\left(\int_o^{xi} dF(x)\right)^{i-1}\cdot\left(\int_{xi}^{xm} dF(x)\right)^{m-i-1}\cdot$$

$$\left(\int_{xm}^{xn} dF(x)\right)^{n-m-1}\cdot\quad f(x_i)f(x_m)f(x_n) = k\cdot exp\{-(x_i+x_m+x_n)\}\{1-exp(-x_i)\}^{i-1}$$

$$\{exp(-x_i)-exp(-x_m)\}^{m-i-1}\cdot\quad\{exp(-x_m)-exp(-x_n)\}^{n-m-1}$$

,where k means $\frac{n\,!}{(i-1)\,!\,(m-i-1)\,!\,(n-m-1)\,!}$

And transforming $(X_i, X_m, X_n,)$ into (U, V, W) by means of the linear relation $U=X_n-X_m$, $V=X_n-X_i$, $W=X_i$, then we have the j. p. d. f. of (U, V, W). That is

$$k\cdot exp\ \{-3w+2v+u)\{1-exp(-w)\}^{i-1}\ \{exp(-w)-exp(-w-v+u)\}^{m-i-1}$$

$$\times\{exp(-w-v+u)-exp(-w-v)\}^{n-m-1}\quad .$$

Furthermore integrate out the above function with regards to w over its range of definition , we have the j. p d f. f(u, v) of (U, V),

$$f(u,v)=\ k\cdot B(n-i+1,i)\cdot exp\{u-v(n-i)\}\{exp(v)-exp(u)\}^{m-i-1}\cdot\ \{exp(u)-1\}^{n-m-1}$$

,*where* $B(m,n)$ *means Beta-function.*

There the probability that the random variable R is less than r, that is written as $Pr(R\leqq r)$, is calculated by

$$Pr\ (P\leqq r) =\int_o^\infty du\ \int_{u/r}^\infty f(u,v)\ dv$$

After some calculation, we have the sampling distribution F(r) of R,

$$F(r) = k\ B(n-i+1,\ i)\ \sum_{j=o}^{m-i-1}\binom{m-i-1}{j}(-1)^j\ \frac{B(n-m,\ \frac{n-m+1+(m-n-j)r}{r})}{n-m+1+j} \qquad (2)$$

Now we can propose the following R-statistic for the test function

$$R_i = \frac{X_n - X_{n-1}}{X - X_i}, \qquad R'_i = \frac{X_n - X_{n-2}}{X_n - X_i} \qquad (i-1,\ 2,\ 3)$$

and denoting their distribution functions $F_i^{(m)}(r)$, $F_i'^{(n)}(r)$, respectively, we may easily see the following recurrence relations (3)

$$F_i^{(n-1)}(r) = F_{i+1}^{(n)}(r) \quad , \quad F_i'^{(n-1)}(r) = F_{i+1}'^{(n)}(r) \qquad (3)$$

The recurrence relations leave us to pursue only the statistic R_1 and R'_1 Again we replace R_1 and $R'_1$by R_1 and R_2, and their distributions to be $F_1(r)$, $F_2(r)$, respectively. Then

$$F_1(r) = (n-1)(n-2) \sum_{j=0}^{n-3} \binom{n-3}{j} (-1)^j \frac{B\left(1, \frac{j+3-(j+1)r}{r}\right)}{j+2}$$

$$F_2(r) = (n-1)(n-2)(n-3) \sum_{j=0}^{n-4} \binom{n-4}{j} (-1)^j \frac{B\left(2, \frac{j+3-(j+2)r}{r}\right)}{j+3} \qquad (4)$$

Lemma.

For any c. we have the following combinational relation.

$$\sum_{j=0}^{n} \binom{n}{j} (-1)^j \frac{1}{j+c} = c \left(\frac{n+c}{n}\right)^{-1} \qquad (5)$$

For, in the identity $(1+x)^{n+c}(1+x)^{-c}=(1+x)^n$,comparing the coefficient of x^n in each member, we have

$$\sum_{j=0}^{n} \binom{n+c}{n-i}\binom{-c}{j} = 1$$

Moreover it is easy to see that the first member equals to

$$\sum_{j=0}^{n} \binom{n+c}{n-j}\binom{-c}{j} = \sum_{j=0}^{n} \binom{n}{j}(-1)^j c\binom{n+c}{n}(c+j)^{-1}$$

This proves our lemma.

By subtituting this relaton into (4), we have Theorem I

Theoreim I.

The R-ststistic from the exponential distribution (1)

$$R_1 = \frac{X_n - X_{n-1}}{X^n - X_1} \quad , \quad R_2 = \frac{X_n - X_{n-2}}{X_n - X_1}$$

are distributed respectively according to $F_1(r)$, $F_2(r)$

$$F_1(r)= 1-(n-1)!\ \Gamma(2+\frac{r}{1-r})\ \Gamma(n+\frac{r}{1-r})^{-1} \tag{6}$$

$$F_2(r)= 1-(n-1)!\ \left\{\Gamma(3+\frac{r}{1-r})\Gamma(n+\frac{r}{1-r})^{-1}+\Gamma(3+\frac{2r}{1-r})\Gamma(n+\frac{2r}{1-r})^{-1}\right\} \tag{7}$$

,where $\Gamma(s)$ means Gamma-function.

When n is large enough, we get the following asymptotic distribution of R_1 and R_2 after some computation by making use of Stirling's formula

$$\Gamma(s) \sim \sqrt{2\pi}\ s^{s-\frac{1}{2}}\ exp(-s)$$

That is

$$F_1(r) \sim 1-\Gamma(2+\frac{r}{1-r})\ n^{-\frac{r}{1-r}} \tag{8}$$

$$F_2(r) \sim 1-\Gamma(3+\frac{r}{1-r})\ n^{-\frac{r}{1-r}}+\frac{1}{2}\Gamma(3+\frac{2r}{1-r})\ n^{-\frac{2r}{1-r}} \tag{9}$$

Furthermore, if we define T_i by $R_i\ logn = T_i$ $(i = 1,2)$ and consider the limiting distribution function of T_i as $n\to\infty$, we have

$$P_r(T_1\leqq t) \sim 1-\Gamma\left[2+\left(\frac{t}{logn}\right)\left(1-\frac{t}{logn}\right)^{-1}\right]\cdot exp\left[-\left(\frac{t}{logn}\right)\left(1-\frac{t}{logn}\right)^{-1}\times logn\right] \to 1-e^{-t}\quad (as\ n\to\infty) \tag{10}$$

Similarly,

$$\lim_{n\to\infty} P_r(T_2\leqq t) = 1-2e^{-t}+e^{-2t} = (1-e^{-t})^2 \tag{11}$$

The moments of R_1 may be given from (6) after considerably complicated computation, and especially its expected value $E(R_1)$ is given by

$$E(R_1) = (n-1)\left\{1+\sum_{j=1}^{n-2}\binom{n-2}{j}(-1)^j\frac{log(1+j)}{j}\right\} \tag{12}$$

It may be assumed that the evaluation of E (R_1) for large n can be computed from (12), but as for myself, I could not show a satisfactory result. However, considering the limiting form (10), we can suggest the order of the expected value and variance of R_1, respectively, $\frac{1}{logn}$ *and* $\left(\frac{1}{logn}\right)^2$,since $E(T_1)=1$ and $E(T_1^2)=1$. These are the same as for R_2.

Theorem II.

The asymptotic distribution function of R_i $(i=1,2)$ are given by (8), (9) and their limiting forms are

$$\lim_{n\to\infty} P_r\ (R_1 \leq \frac{r}{\log n}) = 1-e^{-r}$$

$$\lim_{n\to\infty} R_r\ (R_2 \leq \frac{r}{\log n}) = (1-e^{-r})^2$$

Furthermore,

$$E\ (R_i) = O\ (\frac{1}{\log n})\ , \quad V(R_i) = O\ (\frac{1}{(\log n)^2})$$

In §3, we shall calculate the probability for each n from the exact distribution (6) and (7).

(II) Normal case

In this case, Dixon gave the density function of R-statistic. Without loss of generality, we set the density function as follows :

$$f(x) = \frac{1}{\sqrt{2\pi}}\ exp\ (-\frac{x^2}{2}) \qquad for\ -\infty<x<\infty$$

Then the sampling density function $g_n(r)$ of R_1 are given, which are

$$g_3(r) = \frac{3\sqrt{3}}{\sqrt{2\pi}}\ \frac{1}{r_2-r+1}$$

$$g_4(r) = \frac{3}{\pi}\ \frac{1}{r_2-r+1}\left(\frac{1-2r}{\sqrt{4r_2-4r+3}} - \frac{r-2}{\sqrt{3r_2-4r+4}}\right)$$

$$g_5(r) = \frac{15\ h(r)+h(\frac{1}{r})}{\pi^2\ (r_2-r+1)}$$

$$where \qquad h(r) = \frac{2-r}{\sqrt{3r^2-4r+4}}\ tan^{-1}\frac{(1-r)\sqrt{5(3r^2-4r+4)}}{3r^2-4r+4}$$

§ 3. The probability tables.

Table I and II give the values r such as Pr $(R \geq r) = \alpha$ for each n and α. Making use of this tables, we may perform the rejection test.

Table Ⅰ $Pr\ (R_1 \geq r) = \alpha$

$n \backslash \alpha$	0.005	0.01	0.02	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
3	0.997	0.995	0.989	0.974	0.974	0.889	0.824	0.750	0.667	0.571	0.462	0.333	0.182	0.095
4	0.970	0.960	0.937	0.895	0.840	0.750	0.667	0.585	0.500	0.412	0.320	0.221	0.115	0.059
5	0.933	0.917	0.885	0.830	0.766	0.667	0.581	0.500	0.423	0.341	0.260	0.176	0.090	0.046
6	0.900	0.875	0.844	0.782	0.714	0.612	0.527	0.448	0.373	0.300	0.226	0.152	0.077	0.039
7	0.873	0.846	0.810	0.746	0.675	0.573	0.489	0.413	0.341	0.272	0.204	0.136	0.069	0.034
8	0.850	0.822	0.784	0.717	0.646	0.543	0.460	0.387	0.318	0.252	0.188	0.125	0.065	0.031
9	0.831	0.800	0.762	0.694	0.621	0.519	0.438	0.367	0.300	0.237	0.177	0.117	0.058	0.029
10	0.812	0.783	0.744	0.675	0.602	0.500	0.420	0.350	0.286	0.225	0.167	0.110	0.055	0.027
11	0.800	0.769	0.731	0.658	0.585	0.484	0.406	0.337	0.275	0.216	0.165	0.105	0.052	0.026
12	0.788	0.755	0.715	0.644	0.571	0.470	0.393	0.326	0.265	0.208	0.153	0.101	0.050	0.025
13	0.776	0.744	0.703	0.631	0.558	0.459	0.382	0.316	0.256	0.201	0.148	0.097	0.048	0.024
14	0.767	0.734	0.692	0.621	0.547	0.448	0.372	0.306	0.249	0.195	0.144	0.094	0.047	0.023
15	0.758	0.729	0.683	0.611	0.538	0.439	0.365	0.300	0.243	0.190	0.140	0.091	0.045	0.022

Table Ⅱ $P_r\ (R_2 \geq r) = \alpha$

$n \backslash \alpha$	0.005	0.01	0.02	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
5	0.984	0.977	0.966	0.944	0.914	0.865	0.818	0.769	0.717	0.658	0.588	0.500	0.375	0.277
6	0.963	0.948	0.932	0.899	0.862	0.802	0.748	0.695	0.640	0.580	0.511	0.427	0.313	0.223
7	0.937	0.922	0.901	0.866	0.820	0.755	0.698	0.643	0.593	0.528	0.461	0.381	0.276	0.199
8	0.915	0.898	0.875	0.833	0.787	0.718	0.660	0.604	0.552	0.490	0.426	0.349	0.250	0.179
9	0.897	0.877	0.853	0.808	0.760	0.689	0.630	0.574	0.519	0.460	0.398	0.325	0.235	0.165
10	0.881	0.860	0.834	0.788	0.737	0.665	0.609	0.500	0.495	0.438	0.373	0.307	0.219	0.153
11	0.867	0.846	0.819	0.769	0.724	0.646	0.585	0.530	0.475	0.420	0.360	0.292	0.201	0.145
12	0.854	0.832	0.803	0.753	0.701	0.627	0.566	0.510	0.459	0.404	0.346	0.279	0.196	0.138
13	0.843	0.820	0.790	0.740	0.687	0.612	0.552	0.498	0.445	0.391	0.333	0.269	0.188	0.132
14	0.832	0.808	0.779	0.730	0.674	0.599	0.539	0.485	0.433	0.379	0.323	0.260	0.181	0.127
15	0.823	0.799	0.769	0.718	0.662	0.587	0.528	0.473	0.422	0.369	0.314	0.252	0.175	0.122

References.

〔1〕 W. J. Dixon : "Analysis of Extreme Values" Ann. Math. Stat. Vol. 21. 1950.

〔2〕 〃 〃 "Ratio Involving Extreme Values." 〃 〃 Vol22. 1951.

〔3〕 M. Matsuyama : "The Rejection Limit in Exponential Distribution". Res. Mem. Inst. Math. Stat.