MOMENTS OF RECORD STATISTICS

By

A. Nagabhushanam* and M. L. Aggarwal**

(Received in December, 1974)

1. Introduction

Let $X_1, X_2, \ldots\ldots$ be a sequence of independent observations from a continuous population with density function $f(x)$ and distribution function $F(x)$. With $N_0 = M_0 = 1$, we successively define

$$N_q = \text{Min}\left\{ j : j > N_{q-1},\ X_j < X_{N_{q-1}} \right\}$$

$$M_q = \text{Min}\left\{ j : j > M_{q-1},\ X_j > X_{M_{q-1}} \right\}$$

and call

$$X_{N_q} = L_q$$

the qth lower record and

$$X_{M_q} = U_q$$

the qth upper record where it is customary to treat X_1 as the zeroth lower as well as upper record. In a given sequence we thus distinguish those elements which are less than (or exceeding) all the preceding observations. This problem arises in many practical situations such as weather reporting, sports events etc.

The statistical behaviour of records relative to others is of interest in its own apart from its applications to forecasting and decision making. Chandler[4] initiated the study of records in a stationary time series by deriving the distributions of (*i*) the rth record (*ii*) its place of occurrence. Foster and Stuart[5] developed test procedures based on number of records for testing the trend of a time series. For other distribution free tests for stochastic processes see Bell, Woodrofe and Avdhani[3]. For the study of coverage, exceedenees

* Department of Mathematics, I.I.T., Hauz Khas, New Delhi-110029.

** Faculty of Mathematics, University of Delhi, Delhi-110007.

in Records refer to the authors[1,6]. Barton and Mallows[2] have given the connection between record value problem and other problems such as (*i*) Amalgamation and (*ii*) Simon Newcomb problems. Renyi[9] investigated the index N_q of the qth record and its limiting distribution showing that log N_q is asymptotically normal. The waiting time Δ_q between q—1th and qth record obeys the weak law of large numbers and a central limit theorem; see Neuts[8]. Resnick[10] has given necessary and sufficient conditions under which a record value has asymptotically normal distribution. A number of papers dealing with various aspects of records have appeared recently; see for example[6,7,11,12,13].

In this paper, we obtain the moments of the sampling distribution of the lower (upper) record statistics where the parent population has a specified form say with a given *c.d.f.* $F(.)$. The joint distribution of two consecutive records alongwith their marginal distributions is given below to be used in the sequel. The probability that there are exactly $\Delta_q = N_q - N_{q-1} - 1$ observations from the occurrence of q—1th to that of the qth record and that the qth lower record has value in (y_q, y_q+dy_q) for $q=1, 2, \ldots, r$ is

$$\left[\int_{y_1}^{\infty} (1-F(y_o))^{N_1-2}\, dF(y_o)\right] dF(y_1) \prod_{q=2}^{r} (1-F(y_{q-1}))^{\Delta_q}\, dF(y_q)$$

so that on summing over all possible N's we have the joint density of the records $L_1, \ldots, L_r$ to be

$$\left[\int_{y_1}^{\infty} \frac{dF(y_o)}{F(y_o)}\right] \frac{f(y_1)\ldots f(y_r)}{F(y_1)\ldots F(y_{r-1})} \tag{1}$$

$$-\infty \angle y_r < y_{r-1} \ldots < y_1 < \infty$$

from which we easily have the joint density of y_{r-1}, y_r as

$$g_{r-1,r}(x, y) = \frac{[-\log F(x)]^{r-1}}{(r-1)!} f(x) \frac{f(y)}{F(x)} \tag{2}$$

$$-\infty < y < x < \infty,$$

and the density of L_r as

$$g_r(y) = f(y)[-\log F(y))]^r/_r!, \quad -\infty < y < \infty \tag{3}$$

The results in upper records are easily given by replacing F by $1-F$ in the above results.

Remark—. If the parent population is symmetric then U_r and $-L_r$ have the same distribution. Assuming that the respective moments exist, U_r+L_r and U_r-L_r are uncorrelated.

The Jacobian of the transformation from X to $-X$ being unity, the density function of $-L_r$ is

$$f(-y)\,[-\log F(-y)]^r/r!$$
$$=f(y)\,[-\log\{1-F(y)\}]^r/r!$$

as the population is symmetric; the density function of $-L_r$ is thus the same as that of U_r. Now

$$E(U_r+L_r)=E(-L_r)+E(L_r)=0,$$

it follows that cov $(U_r+L_r, U_r-L_r)=E(U^2_r)-E(L^2_r)$ and is zero.

2. $f(.)$ Concentrated Over A Finite Range :

Here we consider the case when f vanishes outside the interval (α, β). We assume that $F^{-1}(x)$ is well defined, possesses continuous derivatives at $x=o$ which are bounded. The moment generating function of L_r is

$$E(e^{tL_r}) = \int_{\alpha}^{\beta} e^{tx} f(x) \frac{[-\log F(x))]^r}{r!}\,dx$$

The transformation $Z=-\log F(x)$ reduces it to

$$\int_0^{\infty} e^{-z} \frac{z^r}{r!} \exp\left\{ tF^{-1}(e^{-z}) \right\} dz$$

Expanding $G(x)=F^{-1}(x)$ about $x=o$, we have

$$G(x)=\alpha+xa_1+\frac{x^2}{2!}a_2+\ldots$$

where

$$a_k=[G^{(k)}(x)]_{x=o}$$

and by our assumption $|a_k|<M$. Now substituting

$$G(e^{-z})=\alpha+a_1e^{-z}+\frac{a_2}{2!}e^{-2z}+\ldots$$

we have the moment generating function to be

$$M_r(t) = e^{\alpha t}\left[\ 1+\ \frac{a_1 t}{2^{r+1}}+\frac{a_2 t+a_1 t^2}{2(3r+1)}+\ldots\right]\ldots \quad (4)$$

from which on differentiation and setting $t=o$,

$$E(L_r) = \alpha+\frac{a_1}{2^{r+1}} \quad (5)$$

$$E(L^2_r) \cong \alpha^2+\frac{a_1\alpha}{2^r}+\frac{a_1^2}{3^{r+1}}$$

and Var

$$(L_r) \cong a_1^2\,(3^{-r-1}-4^{-r-1})$$

so that

$$\sigma(L_r) \cong a_1/(\sqrt{3})^{r+1} \quad (6)$$

It follows that as $r\to\infty$,

$$E(L_r)\to\alpha,\ \sigma(U_r)\to o$$

In a similar manner one can show that

$$E(U_r)\to\beta,\ \sigma(U_r)\to 0,$$

Thus L_r is a consistent estimate for α and U_r for β. If the sample size is not fixed and observations are taken until a record of specified order occurs then it can be used in estimating the parameter α or β of the distribution. From the moment generating function of L_r, it can be shown that

$$\operatorname{lng} E\left[exp\left\{t\ \frac{(L_n-\alpha)(\sqrt{3})^{r+1}}{a_1}\right\}\right]$$

$$=-\frac{\alpha t}{a_1}(\sqrt{3})^{r+1}+\log M\!\left(\frac{t}{a_1}(\sqrt{3})^{r+1}\right)$$

$$\to t^2/2 \text{ as } r\to\infty. \quad (7)$$

Thus L_r is asymptotically normal and a similar result holds for U_r.

3. Logistic Population

Here we have

$$F(x)=[1+e^{-gx}]^{-1}, -\infty<x<\infty$$

Resnick[10] has shown that the class of limit laws for record values is of the form $N(-\log(-\log G(x))$ where $G(x)$ is an extreme value distribution and $N(x)$ is the standard normal distribution. In particular, the limiting distribution is normal *iff* the minimum in a random sample of size n from the population with

$$H(x)=1-exp\,[-\{-\log F(x)\}^{\frac{1}{2}}\,]$$

as the distribution function has the limiting distribution

$$\wedge(x)=exp\,\{-e^{-x}].$$

With $F(x)=[1+e^{-gx}]^{-1}$ it can be shown that the limit law for the rth record is the standard normal distribution. We proceed to establish this result directly by computing the moments of L_r. For this we define a function $G(\theta)$ related to the moment generating function of L_r by

$$G(\theta)=M_r(g\theta)=E(e^{g\theta L_r}).$$

For $\theta>-1$, $g>o$ it is

$$\int_{-\infty}^{\infty}\frac{\log^r(1+e^{-gx})}{r\,!}\,\frac{ge^{g\theta x-gx}}{(1+e^{-gx})^2}dx.$$

The substitution $e^u=1+e^{-gx}$ reduces to it

$$\int_0^{\infty}\frac{u^r}{r!}(1-e^{-u})^{-\theta}\,exp[-u(1+\theta)]du$$

$$=\sum_{k=0}^{\infty}(^{-\theta}{}_k)\frac{(-1)^k}{(k+1+\theta)^{r+1}} \tag{8}$$

Differentiating (8) *w.r.t.* θ,

$$gM'_r(g\theta)=-(r+1)\sum_{k=0}^{\infty}(k^{-\theta})\;\frac{(-1)^k}{(k+1+\theta)^{r+2}}$$

$$+\sum_{k=1}^{\infty}(k^{-\theta})\;\frac{(-1)^k}{(k+1+\theta)^{r+1}}\sum_{j=0}^{k-1}\frac{1}{(\theta+j)} \tag{9}$$

From which we have

$$gM'_r(0)=-(r+1)+\sum_{k=1}^{\infty}\frac{1}{k(K+1)^{r+1}} \tag{10}$$

so that $E(L_{r-1})-E(L_r)$

$$=g^{-1}[1+\sum_{k=1}^{\infty}\frac{1}{(k+1)^{r+1}}]$$

$$=g^{-1}\sum_{k=1}^{\infty}\frac{1}{k^{r+1}} \tag{10'}$$

We now express the result in terms of the Riemann Zeta function $\zeta(z,q)$ defined by

$$\zeta(Z,q)=\frac{1}{\Gamma(Z)}\int_0^{\infty} e^{-qt}\ \frac{t^{z-1}}{(1-e^{-t})}dt \tag{11}$$

$\zeta\,(Z,q)$ has the series representation:

$$\zeta(Z,q)=\sum_{k=0}^{\infty}\frac{1}{(q+k)^z},\ Re\ (Z)>1 \tag{11^1}$$

We denote $\zeta\,(Z,1)$ by $\zeta\,(Z)$. Thus from (10′),(11′)

$$E\,(L_{r-1})-E\,(L_r)=g^{-1}\,\zeta\,(r+1) \tag{12}$$

We will now make use of the following property of $\zeta\,(Z)$ in deriving the limiting distribution of L_r :

$$\zeta\,(Z)=1+0(Z^{-1}) \tag{13}$$

using (13), we have from (12)

$$E\,(L_r)=-g^{-1}r+0(1) \tag{14}$$

Now further differentiating (9) *w.r.t.* θ we get

$$g^2M''_r(g\theta)=(r+1)(r+2)\sum_{k=0}^{\infty}\binom{-\theta}{k}\frac{(-1)^k}{(k+1+\theta)^{r+3}}$$

$$-2(r+1)\sum_{k=1}^{\infty}\binom{-\theta}{k}\frac{(-1)^k}{(\theta+k+1)^{r+2}}\sum_{j=0}^{k-1}\frac{1}{(\theta+j)}$$

$$+\sum_{k=2}^{\infty}\binom{-\theta}{k}\ \frac{(-1)^k}{(k+1+\theta)^{r+1}}\sum_{\substack{i,j=0\\ i\neq j}}^{k-1}\frac{1}{(\theta+i)(\theta+j)}$$

so that

$$g^2M''_r(0)=(r+1)(r+2)-2(r+1)\sum_{k=1}^{\infty}\frac{1}{k(k+1)^{r+2}}$$

$$+2\sum_{k=2}^{\infty}\frac{1}{k(k+1)^{r+1}}\sum_{j=1}^{k-1}\frac{1}{j} \tag{15}$$

Thus

$$g^2[M''_r(0)-M''_{r-1}(0)]=2(r+1)\sum_{k=0}^{\infty}\frac{1}{(k+1)^{r+2}}$$

$$-2\sum_{k=1}^{\infty}\frac{1}{(k+1)^{r+1}}\sum_{j=1}^{K}\frac{1}{j}$$

$$=2(r+1)\zeta(r+2)-2\sum_{j=1}^{\infty}\frac{1}{j}\zeta(r+1,j+1) \tag{16}$$

From (13) and (16) it follows that

$$g^2M''_r(0)\cong(r+1)\,(r+2) \tag{17}$$

From (14) and (17) we have

$$g^2\ Var(L_r)=r+0(1).$$

Asymptotic distribution. We will now show that $\{V_r\}$ where

$$V_r=(gL_r+r)/\sqrt{r}$$

converges in distribution to the standard normal distribution. For this, we consider the moment generating function of V_r and show that it tends to $e^{t^2/2}$. We have

$$E\,(e^{\theta V}r)=e^{\,\theta\sqrt{r}}\sum_{k=0}^{\infty}(k^{-\theta'\sqrt{r}})\ \frac{(-1)^k}{(k+1+\theta/\sqrt{r})^{r+1}}$$

$$=e^{\theta}\sqrt{r}\quad[(1+\theta/\sqrt{r})^{-r-1}+\sum_{k=1}^{\infty}\frac{(-1)^k}{(k+1+\theta/\sqrt{r})^{r+1}}\binom{-\theta/\sqrt{r}}{k}]$$

(18)

The first term on the right of (18) is

$$exp[\theta\sqrt{r} -(r+1)\log(1+\theta/\sqrt{r})]$$

$$= exp[\theta\sqrt{r} -(r+1)\left\{\frac{\theta}{\sqrt{r}} - \frac{\theta^2}{2r} + \frac{\theta^3}{3r^{3/2}} + 0(r^{-2})\right\}]$$

$$= et^{2/2}[1+0(1/\sqrt{r})] \tag{19}$$

The second term on the right of (18) is numerically less then

$$\sum_{k=1}^{\infty} e^{\theta\sqrt{r}} \binom{-\theta/\sqrt{r}}{k}(1+\theta/\sqrt{r})^{-r-1} \tag{20}$$

for sufficiently large r. As $r \to \infty$ the expression given by (20) is easily seen to approach 0. Thus we have

$$\lim_{r\to\infty} E(e^{tV_r}) = e^{t^2/2},$$

showing that L_r is asymptotically normally distributed.

Remark. Since the logistic population is symmetric, U_r is also asymptotically normal.

Asymptotic equivalence of L_r, L_{r-1} : We recall that two statistics S_n and T_n used in estimating a parameter θ are asymptotically equivalent if

(*i*) They are of equal efficiency

(*ii*) $\rho(S_n, T_n) \to 1$ as $n \to \infty$.

Now L_r and L_{r-1} are of equal efficiency for large r. Consider the function $G(\theta_1, \theta_2)$ defined by

$$G(\theta_1, \theta_2) = E[exp\{g(\theta_1 L_r + \theta_2 L_{r-1})\}]$$

From (2), we have

$$G(\theta_1,\theta_2) = \int_{-\infty}^{\infty}\int_{y}^{\infty} \frac{[-\log F(x)]^{r-1}}{(r-1)!} f(x)\frac{f(y)}{F(x)} \; e^{g(\theta_1 x+\theta_2 y)}\, dxdy \tag{21}$$

The transformation $e^u = 1+e^{-gx}$, $e = 1+e^{-gy}$ reduces (21) to

$$\int_0^{\infty}\int_0^{V} \frac{u^{r-1}}{(r-1)!} \frac{e^{-v}\, dudv}{(e^u-1)^{\theta_1} (e^{v}-1)^{\theta}} \tag{22}$$

Assuming that $\theta_1 > -1$, $\theta_2 > -1$ and $g > 0$. We express (22) in the form

$$G(\theta_1,\theta_2) = \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} (k_1^{-\theta_1})(k_2^{-\theta_2}) \frac{(-1)^{k_1+k_2}}{(\theta_2+k_2+1)(1+\theta_1+\theta_2+k_1+k_2)^r}$$

Differentiating partially *w.r.t.* θ_1 and θ_2 and then letting $\theta_1, \theta_2 \to 0$ we get

$$\left(\frac{\partial^2 G}{\partial\theta_1\partial\theta_2}\right)_{\theta_1=\theta_2=0} = r + r(r+1) - r\sum_{k_2=1}^{\infty}\frac{1}{(1+k_2)^{r+2}k_2}$$

$$-\sum_{k_1=1}^{\infty}\frac{1}{k_1(1+k_1)^r} - r\sum_{k_1=1}^{\infty}\frac{1}{k_1(1+k_1)^{r+1}}$$

$$+\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty}\frac{1}{(k_1k_2(k_2+1)(1+k_1+k_2)^r} = P_r \text{ say.}$$

After certain algebraic manipulations, we find

$$P_r - P_{r-1} = 2r\sum_{k=0}^{\infty}\frac{1}{(1+k)^{r+1}} - \sum_{k=1}^{\infty}\frac{1}{(k+1)^r}$$

$$-2\sum_{k=1}^{\infty}\sum_{j=1}^{k}\frac{1}{j(k+1)^{r+1}}$$

$$= 2r\zeta(r+1) - \zeta(r) + 1 - 2\sum_{j=1}^{\infty}\frac{1}{j}\zeta(j+1, r+1)$$

Once again making use of (13), we have

$$P_r \simeq r(r+2) \tag{23}$$

Thus

$$Cov\,(L_r, L_{r-1}) = E(L_r\, L_{r-1}) - E\,(L_r)E\,(L_{r-1})$$
$$= g^{-2}\, r + 0(1),$$

and

$$\rho(L_r, L_{r-1}) = \frac{g^{-2}\, r + 0(1)}{g^{-2}\, r + 0(1)} \to 1.$$

SUMMARY

In this paper we consider the moments of the sampling distribution of $L_r\,(U_r)$, the rth lower (upper) record statistic assuming that the observations are drawn independently from a continuous population. In particular if the population is symmetric we show that $U_r + L_r$ and $U_r - L_r$ are uncorrelated. For the logistic population,

we consider the asymptotic distribution of L_r, U_r and show that are asymptotically independent and have normal distributions. Also we establish that the limiting correlation coefficient between L_r and L_{r-1} is unity.

ACKNOWLEDGEMENT

The authors are grateful to Dr. H.C. Gupta, Professor and Head of the Department of Mathematical Statistics, University of Delhi for guidance and supervision in the preparation of this paper. The authors are also grateful to the referee for valuable suggestions.

REFERENCES

1. Aggarwal, M. L. and Nagabhushanam (1971) : Coverage of a record value and related distribution problems. Bulletin of Calcutfa Statistical Association, 20, 99-103.

2. Barton, D.E. and Mallows, C.L. (1965) : Some aspects of the random sequence. Annals Math. Stats. 36, 236-260.

3. Bell, C.B., Woodroofe, M. and Avdhani, T. V. (1970) : Some non-parametric tests for Stochastic Processes in Nonparametric techniques in Statistical inference. Cambridge Univ. Press, 215-258.

4. Chandler, K. N. (1952) : The distribution and frequency of record values. J. Roy. Statist. Soc. Ser. 'B' 14, 220-228.

5. Foster, F.G. and Stuart, A. (1954) : Distribution free tests in time series based on the breaking of records. J Roy. Statist. Soc. Ser. 'B' 16, 1-22.

6. Nagabhushanam, A. and Aggarwal, M.L. (1974) : The distribution of record values and the problem of number of excedenees. Journal of Mathematical Sc. To appear.

7. Nagabhushanam, A., Aggarwal, M.L. and Gupta, H.C. (1975) : Mean of outstanding elements. Studia Sci. Math. To appear.

8. Neuts, M.F. (1967) : Waiting times between record observations. J. Appl Prob. 4, 206-208.

9. Renyi, A. (1962) : Theorie des elements sailiants d'une suite d' observations. Colloquium on combinatorial methods in Probability Theory, August 1—10, Mathematisk Institute, Aarhus Universitet, Denmark.

10. Resnick, S.I. (1973) : Limit laws for record values. Stochastic Processes and its Appl., 1, 67-82.

11. Shorrock, R.W. (1972) A limit theorem for inter-record times. J. Appl. Prob. 9, 219-223.

12. Strawderman, W.E. and Holmes, P.T. (1970) On the law of the iterated logarithm for inter-record times. J. Appl. Prob., 7, 432-439.

13. Tata, M.N. (1969) : On outstanding values in a sequence of random variables. Z. Wahrscheinlichkeitsth. 12, 9-20.