International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 20(2), 2020, pp 257-266

Record Values

M. Ahsanullah

Rider University, Lawrenceville, New Jersey, USA.
Email: ahsan@rider.edu

[Received July 01, 2020; Accepted November 29, 2020]

Abstract

Records occur naturally in various fields of studies such as science, sports, engineering, medicine, and industry among others. An observation is called is called a record if it is greater than (or less than) all the preceding observations. In the case of discrete distributions an observation is called a weak record if it is equal to the preceding record. In this paper some basic properties of records including weak records are presented. Examples are given for the use of records to estimate parameters and characterizations of continuous distributions.

Keywords: Record value, Probability distributions, Parameter estimation.

AMS Classification: 62G30, 62E10.

1. Introduction

Record values are the local maxima or minima of a sequence of random variables.

Consider the following observations:

6, 10, 5, 6, 11, 11, 14, 4, 18, 4, 1, …

6,10,11,14,18 - upper record values

6,5,4,1 - lower record values

6,10,11,11,14.18 - upper weak records

6, 5,4,4,1 -lower weak record

One can go from lower records to upper records by replacing the original sequence of random variables $\{X_j\}$ by $\{-X_j,\ j \geq 1\}$ or if $P(X_j > 0) = 1$ by $\{1/X_j,\ i \geq 1\}$, j=1,2,…

Chandler (1952) introduced the record values, record times and inter record times. He proved the interesting result that for any given distribution of a random variable the expected value of the inter record time is infinite. Feller (1966) gave some examples of record values with respect to gambling problems.

2. Definition of Record Values and Record Times

Suppose that X_1, X_2 .... is a sequence of independent and identically distributed random variables with cumulative distribution function $F(x)$. Let Y_n = max (min)$\{X_1, X_2, ..., X_n\}$ for $n \geq 1$. We say X_j is an upper (lower) record value of $\{X_n, n \geq 1\}$, if $Y_j > (<) Y_{j-1}$, $j > 1$. By definition X_1 is an upper as well as a lower record value.

The indices at which the upper record values occur are given by the record times $\{U(n)\}$, $n > 0$, where $U(n) = \min\{j | j > U(n-1), X_j > X_{U(n-1)}, n > 1\}$ and $U(1)=1$. The record times of the sequence $\{X_n, n \geq 1\}$ are the same as those for the sequence $\{F(X_n), n \geq 1\}$. Since $F(X)$ has an uniform distribution, it follows that the distribution of $U(n)$, $n \geq 1$ does not depend on F. We will denote $L(n)$ as the indices where the lower record values occur. By our assumption $U(1) = L(1) = 1$. The distribution of $L(n)$ also does not depend on F.

3. The Exact Distribution of Record Values of Continuous Random Variables

Many properties of the record value sequence can be expressed in terms of $R(x)$, where $R(x) = -\ln \bar{F}(x)$, $0 < \bar{F}(x) < 1$ and $\bar{F}(x) = 1 - F(x)$. Here 'ln' is used for the natural logarithm. If we define $F_n(x)$ as the distribution function of $X_{U(n)}$ for $n \geq 1$, then we have

$$F_1(x) = P[X_{U(1)} \leq x] = F(x) \tag{3.1}$$

The cdf $F_n(x)$ of $X_{U(n)}$ is

$$F_n(x) = P(X_{U(n)} \leq x) = \int_{-\infty}^{x} \frac{R^{n-1}(x)}{\Gamma(n)} dF(x), \quad -\infty < x < \infty \tag{3.2}$$

This can be expressed as

$$F_n(x) = \int_{-\infty}^{R(x)} \frac{u^{n-1}}{\Gamma(n)} e^{-u} du, \quad -\infty < x < \infty$$

The pdf $f_n(x)$ of $X_{U(n)}$ is

$$f_n(x) = \frac{R^{n-1}(x)}{\Gamma(n)} f(x), \; -\infty < x < \infty \tag{3.3}$$

The joint pdf $f(x_1, x_2, \ldots, x_n)$ of the n record values $(X_{U(1)}, X_{U(2)}, \ldots, X_{U(n)})$ is given by

$$f(x_1, x_2, \ldots, x_n) = r(x_1)\, r(x_2) \ldots r(x_{n-1})\, f(x_n)$$

for $-\infty < x_1 < x_2 < \ldots < x_{n-1} < x_n < \infty$.

The function $r(x)$ is known as hazard rate.

The joint pdf of $X_{U(i)}$ and $X_{U(j)}$ is

$$f(x_i, x_j) = \frac{(R(x))^{i-1}}{\Gamma(i)} r(x_i) \quad \frac{(R(x_j) - R(x_i))^{j-i-1}}{\Gamma(j-i)} f(x_j) \tag{3.4}$$

for $-\infty < x_i < x_j < \infty$.

Example 3.1

Consider the exponential distribution with pdf $f(x)$ as $f(x) = e^{-x}, 0 \le x < \infty$ and cdf $F(x)$ as $F(x) = 1 - e^{-x}$, $0 \le x < \infty$. Then $R(x) = x$ and

$$f_n(x) = \frac{x^{n-1}}{\Gamma(n)} e^{-x}, \; x \ge 0$$

$$= 0, \text{ otherwise.}$$

Thus the nth record from exponential distribution has gamma distribution.

Example 3.2

Suppose that the random variable X has the Gumbel distribution with pdf

$$f(x) = e^{-x} e^{-e^{-x}}, -\infty < x < \infty .$$

Let $F_{(n)}(x)$ and $f_{(n)}(x)$ be the cdf and pdf of $X_{L(n)}$ respectively. It is easy to see that

$$F_{(n)}(x) = \int_{-\infty}^{x} \frac{e^{-nu}}{\Gamma(n)} e^{-e^{-u}} \, du$$

and

$$f_{(n)}(x) = \frac{e^{-nx}}{\Gamma(n)} e^{-e^{-x}}, \; -\infty < x < \infty .$$

Let $f_{(m,n)}(x,y)$ be the joint pdf of $X_{L(m)}$ and $X_{L(n)}$, $m<n$. Using (3.4), we obtain for the Gumbel distribution

$$f_{(m,n)}(x,y)=\frac{\left(e^{-y}-e^{-x}\right)^{n-m-1}}{\Gamma(n-m)}\frac{e^{-mx}}{\Gamma(m)}e^{-y}e^{-e^{-y}}, \quad -\infty<y<x<\infty .$$

4. Number of Records in a Sequence of Observations

We will consider here the number of upper records among the sequence of observations $X_1, X_2, ..., X_n$, the result for the lower records are identical.

Let M_n be the number of upper records among the sequence $X_1, ..., X_n$.

$$P[M_n< 2] = P[U(2) > n] = P[\Delta_1 > n\text{-}1] = \frac{1}{n}.$$

$$P[M_n < 3] = P[U(3) > 3] = \sum_{m=1}^{n-1} P[\Delta_1 = n-m, \Delta_2 > m-1] = \frac{1}{n}\sum_{m=1}^{n-1}\frac{1}{n-m}$$

In general

$$P[M_n < k+1] = P[U(k+1) > n] = \frac{1}{n}\sum_{1<m_1<....<m_k<n}\frac{1}{n-m_k}\cdot\frac{1}{n-m_{k-1}}\cdots\frac{1}{m_1} .$$

Theorem 4.1

$$E(M_n) = \sum_{i=1}^{n}\frac{1}{i} \text{ and } Var(M_n) = \sum_{i=1}^{n}\frac{i-1}{i^2}, \quad n\geq 1 .$$

5. Representation of Records

Let $Y_1, Y_2, \ldots, Y_n$, be a sequence of independent and identically distributed random variables with the cdf as $F_0(x) = 1\text{-}\exp(\text{-}x)$, $x>0$. Further suppose that $X_1, X_2, \ldots$ be a sequence i.i.d. r.v.'s with continuous cdf F. Then one has

$$-\ln(1-F(x)) \stackrel{d}{=} Y_1 .$$

The following theorem gives the representation of the nth record as a function of n independent random variables.

Theorem 5.1

$X_{U(n)} \overset{d}{=} g_F^{-1}(g_F(X_1)+g_F(X_2)+...+g_F(X_n))$, where $g_F(x)$ = -ln(1-$F(x)$) and

$g_F^{-1}(x) = F^{-1}(1-e^{-x})$.

Example 5.1 Exponential distribution

For the two parameter exponential distribution, $E(\mu, \sigma)$, with cdf $F(x)$ as

$F(x) = 1-e^{-(x-\mu)/\sigma}$, $x \geq 0$, we obtain

$$X_{U(n)} \overset{d}{=} X_1 + X_2 + ... + X_n - (n-1)\mu$$

where X's are distributed as $E(\mu, \sigma)$,

Example 5.2 Weibull Distribution

For the Weibull distribution with $F(x)$ as

$$F(x) = 1 - e^{-\beta x^\alpha}, \quad x > 0,\ \alpha > 0,\ \beta > 0$$
$$= 0, \quad \text{for } x < 0.$$

Using Theorem 5.1, we obtain

$$\boldsymbol{X_{U(n)} \overset{d}{=} (X_1^\alpha + X_2^\alpha + \cdots + X_n^\alpha)^{\frac{1}{\alpha}}},$$

where X_1, X_2, $\cdots$ X_n are independent identically distributed with as Weibull cdf F(x) as given above.

6. Records of the Discrete Distributions

Let X_1, X_2, …, X_n be a sequence independent and identically distributed random variables taking values on 0, 1, 2, … such that $F(n)$ <1 for all n, n = 0, 1, 2, ….

We define the upper record times, $U(n)$ as $U(1)$ =1, $U(n+1)$ = min{$j > U(n)$, $X_j > X_{U(n)}$ }, n=1, 2, …. The nth upper record value is defined as $X_{U(n)}$. Let p_k = $P(X_1=k)$,

$$P(k) = \sum_{j=0}^{k} p(j), \bar{P}(k) = 1 - P(k) \text{ and } P(\infty) = 1.$$

The joint probability mass function (pmf) of the $X_{U(1)}$, $X_{U(2)}$,…, $X_{U(n)}$ is defined as

$$P_{1,2,\ldots,n}(x_1,x_2,..x_n) = P(X_{U(1)}=x_1, X_{U(2)}=x_2,\ldots, X_{U(n)}=x_n)$$
$$= \frac{p(x_1)}{\bar{P}(x_1)}\frac{p(x_2)}{\bar{P}(x_2)}\cdots\frac{p(x_{n-1})}{\ldots\bar{P}Ix_{n-1})}\,p(x_n),\ -\infty<x_1 < x_2 < \ldots <x_n$$
$$= 0, \text{ otherwise.} \tag{6.1}$$

The marginal pmfs of the upper record values are given as

$p_1(x_1) = P(X_{U(1)}= x_1) = p(x_1)$, $x_1 = 0,1, 2,\ldots,$

$p_2(x_2) = P(X_{U(2)}= x_2) = R_1(x_2)\, p(x_2),$

where

$$R_1(k) = \sum_{0\le x_1<x_2} B(x_1),\ B(x) = \frac{p(x)}{\bar{P}(x)},\ x_2 = 1, 2, \ldots,$$

$P_n(x_n) = P(X_{U(n)}=x_n) = R_{n-1}(x_n)p(x_n),$

$$R_{n-1}(x_n) = \sum_{0\le x_1>x_2<\ldots<x_n} B(x_1)B(x_2)\ldots.B(x_{n-1}),\ x_n = n-1,\ n,\ \ldots \tag{6.2}$$

The joint pmf of $X_{U(m)}$ and $X_{U(n)}$, $m < n$ is given by

$$P_{m,n}(x_m,x_n) = P(X_{U(m)} = x_m, X_{U(n)} = x_n)$$
$$= R_{m-1}(x_m)A(x_m)\, R_{m+1,n}(x_m,x_n)\, p(x_n),\ m\le x_m \le x_n - n + m < \infty, \tag{6.3}$$

where

$$R_{m+1,n}(x,y) = \sum_{x_m<x_{m+1}<x_{m+2}<\ldots<x_n} B(x_{m+1})\ldots.B(x_{n-1}),\ m<n-1$$
$$=1,\quad \text{if}\quad m=n-1$$

The sequence of upper record values $X_{U(1)}, X_{U(2)} \ldots$ forms a Markov chain.

Example 6.1

Consider the geometric distribution with probability mass function:

$p(k) = P(X = k) = pq^{k-1},\quad 0 < p < 1,\ q = 1-p\ ,\ k \in A_o$

$= 0$, otherwise, (6.4)

where A_n = is the set of integers $n+1$, $n+2$, ..., and $n \ge 0$. We say $X \in$ GE(p), if the pmf of X is as given in (6.4). For $k > 0$, we define $r(k) = P[X = k|\ X \ge k\ ]$.

If $X \in$ GE(p), then $\bar{P}(x) = q^x$ and $p(x) = pq^{x-1}$, for $x \in A_O$.

Substituting the values of $\bar{P}(x_i)$ and $p(x_i)$ in (6.4), we get the joint pmf of the n record values as

$$P(x_1, x_2 \dots, x_n) = p^n q^{x_n - n},\ 1 < x_1 < x_2 < \cdots < x_n < \infty$$

$$= 0, \text{ Otherwise.} \qquad (6.5)$$

The conditional pmf of $X_{U(n)} \mid X_{U(n-1)} = x_{n-1}$ is

$$P(X_{U(n)} = x_n \mid X_{U(n-1)} = x_{n-1}) = pq^{x_n - x_{n-1} - 1},\ n-1 \leq x_{n-1} < x_n < \infty,$$

$$= 0, \text{ otherwise.}$$

Thus $X_{U(n)} - X_{U(n-1)}$ is independent of $X_{U(n-1)}$ and $X_{U(n)} - X_{U(n-1)} \in \mathrm{GE}(p)$, $n=2,3,\dots$.

Let $V_1 = X_{U(1)}$

$V_2 = X_{U(2)} - X_{U(1)}$

$V_n = X_{U(n)} - X_{U(n-1)}$.

Then V_i's are independent and $V_i \in \mathrm{GE}(p)$.

We have

$$X_{U(n)} = V_1 + V_2 + \dots + V_n.$$

Thus the marginal pmf of $X_{U(m)}$ can be written as

$$p_m(x) = p[X_{U(m)} = x] = \binom{x-1}{m-1} p^m q^{x-m},\ x \in A_{m-1},\ m \geq 1$$

$$= 0, \text{ otherwise.} \qquad (6.6)$$

We see that $X_{U(m)}$ has a negative binomial distribution with parameters m and p.

7. Weak Records

Vervaat (1973) introduced the concept of weak records of discrete distribution. Let X_1, X_2, … be a sequence of independent and identically distributed random variables taking values on 0, 1,…. with distribution function F such that $F(n)<1$ for any n. The weak record times $U_w(n)$ and weak upper record values $X_{U_W(n)}$ are defined as follows:

$U_w(1) = 1$

$U_w(n+1) = \min\{j > L_w(n),\ X_j \geq \max(X_1, X_2, \dots, X_{j-1})\}$

and the corresponding weak upper record value is defined as $X_{U_W(n+1)}$. If in the above expression if we replace $\geq$ by >, then we obtain record times and record values instead of weak record times and weak record values.

The joint pmf of $X_{U_W(1)}, X_{U_W(2)}, \cdots, X_{U_W(n)}$ is given by

$$P_{w,1,2,...,n}(x_1, x_2, ..., x_n) = \left(\prod_{i=1}^{n-1} \frac{p(x_i)}{\vec{P}(x_i - 1)}\right) p(x_n) \tag{7.1}$$

for $0 \leq x_1 \leq x_2 \leq \ldots \leq x_n < \infty$.

For any $m>1$ and $n>m$, we can write

$$P(X_{U_W(n)} = x_n, ...X_{U_W(m+1)} = x_{m+1} \mid X_{U_W(m)} = x_m, ...$$
$$X_{U_W(1)} = x_1)$$

$$= \left(\prod_{i=m}^{n-1} \frac{p(x_i)}{\vec{P}(x_i - 1)}\right) \frac{p(x_n)}{\vec{F}(x_m - 1)} \tag{7.2}$$

It follows easily from (7.1) and (7.2) that the weak records, $X_{U_W(1)}$, $X_{U_W(2)}$, … form a Markov chain.

The marginal pmf's of the upper weak records are given by

$P(X_{U_W(1)} = x_1) = P_{w,1}(x_1) = p(x_1)$, $x_1 = 0,1,2,\ldots$,….

$P(X_{U_W(2)} = x_2) = P_{w,2}(x_2) = R_{w,1}(x_2)\, p(x_2)$, $x_2 = 0,1,2,\ldots$.

where

$$R_{w,1(x_2)} = \sum_{0 \leq x_1 \leq x_2} \frac{p(x_1)}{\bar{P}(x_1 - 1)} \tag{7.3}$$

$P(X_{U_W(n)} = x_n) = P_{w,n}(x_n) = R_{w,n-1}(x_n)\, p(x_n)$,

where

$$R_{w,n-1}(x_n) = \sum_{0 \leq x_1 \leq x_2 \leq ... x_{n-1}} \prod_{i=1}^{n-1} \frac{p(x_i)}{\vec{P}(x_i - 1)} p(x_n) \tag{7.4}$$

The joint pmf of $X_{U_W(m)}$ and $X_{U_W(n)}$ $m<n$ nd $X_{U(n)}$, $m<n$ is given by

$$P_{w,m,n}(x_m, x_n) = R_{w,m}(x_m)\, A_w(x_m)\, R_{wm,+1,n}(x_m, x_n)\, p(x_n),$$
$$m \leq x_m \leq x_n - n + m < \infty,$$

where

$$R_{w,m,n}(x,y) = \sum_{x_m \le x_{m+1} \le x_{m+2} \cdots \le n} A_w(x_{m+1}) \ldots\ldots A_w(x_{n-1}),\ m < n-1$$
$$= 1 \text{ if } m = n-1,$$

and $A_w(x) = \dfrac{p(x)}{\bar{P}(x-1)}$.

8. Estimation

It is possible to use record values to estimate the parameters of distributions.

Example 8.1

Suppose $X_{U(1)}, X_{U(2)}, \ldots, X_{U(m)}$ are the m record values from an i.i.d. sequence from an exponential distribution with cdf as $F(x) = 1-e^{-\frac{x-\mu}{\sigma}}$, $-\infty < \mu < x < \infty,\ \sigma > 0$. It can be shown that $x(m)$ and $x(n)$ are sufficient statistics for of μ and σ . The minimum variance linear unbiased estimates (MVLUE) $\hat{\mu}, \hat{\sigma}$ of μ and σ respectively are

$$\hat{\mu} = \left(mX_{U(1)} - X_{U(m)}\right)/(m-1)$$
$$\hat{\sigma} = \left(X_{U(m)} - X_{U(1)}\right)/(m-1) \tag{8.1}$$

with

$$Var(\hat{\mu}) = m\sigma^2/(m-1), Var(\hat{\sigma}) = \sigma^2/(m-1) \text{ and } Cov(\hat{\mu}, \hat{\sigma}) = \sigma^2/(m-1). \tag{8.2}$$

9. Characterizations

Record values can be used to characterize various distributions.

Theorem 9.1

Let F be a continuous cumulative distribution function. If for some constants, a and b, $E(X(n)|X(n-1)=x) = ax+b$, then except for a change of location and scale parameters,

(i) $F(x) = 1-(-x)^{\theta}$, $-1<x<0$, if $0<a<1$

(ii) $F(x) = 1-e^{-x}$, $x \ge 0$, if $a=1$

(iii) $F(x) = 1-x^{\theta}$, $x >1$ if $a >1$,

where $\theta = a/(1-a)$.

Acknowledgement: The author thanks the referees for helpful suggestions which improved the presentation of the paper.

References

[1] Ahsanullah, M. (1988). Introduction to Record Statistics, Ginn Press, Needham Heights, MA.USA

[2] Ahsanullah, M. (1995). Record Statistics. Nova Science Publishers Inc, New York, NY, USA.

[3] Ahsanullah, M. (2004). Record Values- Theory and Applications. University Press of America, Lanham, MD, USA.

[4] Ahsanullah, M. and Nevzorov, V. B. (2017). Records via Probability Theory. Atlantis-Press, Paris, France.

[5] Ahsanullah, M. and Raqab, M. Z. (2006). Bounds and Characterizations of Records. Nova Science Publishers Inc, New York, NY, USA.

[6] Ahsanullah, M. and Yanev, G. P. (2008).Records and Branching Processes. Nova Science Publishers Inc, New York, NY, USA.

[7] Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1998). Records. John Wiley & Sons Inc. New York. NY, USA.

[8] Chandler, K. N. (1952). The Distribution and Frequency of Record Values. J. R. Statist. Soc. B 14, 220-228.

[9] Feller, W. (1966). An Introduction to Probability Theory and its Applications. Vol. II, Wiley, New York.

[10] Gulati, S. and Padgett, W. (2003). Parametric and nonparametric inference from record breaking data. Springer Verlag, New York, USA.

[11] Nevzorov, V. B. (2001). Records: Mathematical Theory. Translation of Mathematical Monographs, Volume 194. American Mathematical Society. Providence, RI, USA.

[12] Vervaat, W. (1973). Limit Theorems for Records from Discrete Distributions. Stoch. Proc. and Their Appl. 1, 317-334.