Süleyman Demirel Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
Cilt 21, Sayı 3, 843-846, 2017

Süleyman Demirel University
Journal of Natural and Applied Sciences
Volume 21, Issue 3, 843-846, 2017

DOI: 10.19113/sdufbed.18244

On Distributional Properties of Record Ranges and Record Range Times

Agah KOZAN*[1], Halil TANIL[1]

[1]Ege University, Faculty of Science, Department of Statistics, 35100, İzmir

(Alınış / Received: 10.03.2017, Kabul / Accepted: 19.09.2017, Online Yayınlanma / Published Online: 13.10.2017)

Keywords
Record range,
Record range times,
Record coverage,
Current record,
Extreme value

Abstract: In this paper, some distributional results concerning the number of record ranges in the first n independent and identically distributed (iid) continuous random observations are obtained. The probability of changing the mth record range by a new observation is also given. Furthermore, we introduce the times of record ranges and derive their joint probability mass function. One of the main results found in this paper is that record range times have finite expected values although usual record times have infinite expected values.

Rekor Açıklıkları ve Rekor Açıklık Zamanlarının Dağılımsal Özellikleri Üzerine

Anahtar Kelimeler
Rekor açıklığı,
Rekor açıklık zamanları,
Rekor kapsamı,
Geçerli rekor,
Uç değer

Özet: Bu çalışmada, bağımsız ve aynı dağılımlı sürekli ilk n adet rassal gözlem içerisinde ortaya çıkacak rekor açıklıklarının sayısına dayalı bazı dağılımsal sonuçlar elde edilmiştir. Ayrıca, m. rekor açıklığının yeni bir gözlem tarafından değiştirilmesi olasılığı da verilmiştir. Bunun yanında, rekor açıklıklarının zamanları ortaya atılmış ve bunların ortak olasılık fonksiyonları elde edilmiştir. Bu çalışmada elde edilen ana sonuçlardan birisi, genel rekor zamanlarının beklenen değerlerinin sonsuz olmasına rağmen rekor açıklıklarının zamanlarının beklenen değerlerinin sonlu olmasıdır.

1. Introduction

The quantities of extreme events are widely observed in meteorological data analysis, athletic events, industrial stress testing, and other similar situations. For instance, in industrial stress testing, the current highest (or lowest) value among time ordered observations is recorded since this value is considered to determine breaking points or safe usage limits of a technical system or a material. In statistical literature, there are many models related to these quantities like upper records (lower records), current records, record coverages, and record ranges. In an iid continuous random sequence, upper (or lower) records which are larger (or smaller) than all previous observations were first introduced as a model of successive extremes by Chandler in 1952 [1]. Afterwards, various similar papers and books such as Foster and Stuart [2], Shorrock [3], Resnick [4], Galambos [5], Glick [6], Dunsmore [7], Nagaraja [8], Nevzorov [9], Ahsanullah [10], Borovkov and Pfeifer [11], Feuerverger and Hall [12], Arnold et al. [13] have been published based on record values and record times. Furthermore, Houchens [14] was the first to introduce the concept of current records which is defined as the current values of the upper and lower records in any sequence when the record of any kind (upper or lower record) is observed. It is well known in the literature that record coverage is defined as an interval constructed by the pair of current records and record range is defined as a distance of current records. These two concepts have been studied by many authors. For instance, an application of record range was considered and some characterization results were derived by Basak [15], Ahmadi and Balakrishnan [16] obtained confidence intervals for quantiles in terms of record range, Ahmadi and Balakrishnan [17] gave distribution-free confidence intervals based on current records, Raqab [18] found inequalities for expected current record statistics, Raqab [19] and Ahmadi and Balakrishnan [20] considered distribution-free prediction intervals based on current records and record coverage.

The present paper is organized as follows. In Section 2, some distributional results about the number of record ranges in finite and iid continuous random observations are obtained. In addition, the probability of changing the mth record range by a new observation is given. Also, some distributional results of record range times are derived.

*Corresponding Author: agahkozan@gmail.com

2. Results

Let $\mathbf{X} = \{X_1, X_2, \ldots\}$ be a sequence of iid random variables with distribution function $F(x)$ and corresponding probability density function $f(x)$. In addition, let U_i' and L_i' denote the current values of the upper and lower records in the sequence when the ith record of any kind is observed. Denote $L_0' = U_0' = X_1$. In the statistical literature, it is well known that mth record coverage and mth record range are defined as the interval (L_m' , U_m') and the distance $R_m = U_m' - L_m'$, respectively. In statistical literature [13], the pdf of mth record range in $\mathbf{X}$ is well known as

$$f^{R_m}(r) = \frac{2^m \int_{-\infty}^{\infty} f(r+z)\times\left[-log\left(1-F(r+z)+F(z)\right)\right]^{m-1} dF(z)}{(m-1)!} \quad (1)$$

In this section, some distributional properties of the number of record ranges in finite iid samples are discussed. Also, the probabilities of changing record ranges by a new observation are derived. Furthermore, distribution and moments of record range times are obtained.

2.1. Number of record ranges

Let N_n' indicates the number of record ranges in $X_1, X_2, \ldots, X_n$. Before the distributional properties of N_n' are given, an indicator random variable can be defined as follows:

Definition 1. An indicator random variable can be defined in order to determine the observations at which upper or lower record occur in the sequence $\mathbf{X}$ as follows: $\xi_1 = \xi_2 = 1$ and

$$\xi_i = \begin{cases} 1, & X_i < X_{1:i-1} \text{ or } X_i > X_{i-1:i-1} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

for $i = 3,4, \ldots$.□

Also, the distributions of ξ_i's are given by the following lemma.

Lemma 1. In the sequence $\mathbf{X}$,

(i) $\xi_i \sim$ Bernoulli$(2/i)$ for $i \in \{3,4,5,\ldots\}$.
(ii) $\xi_3, \xi_4, \xi_5, \ldots$ are mutually independent.

Proof. (i) First of all, let us examine how many different ways the event $\{\xi_i = j_i\}$ can occur for $\xi_i \in \{0,1\}$. The first $(i-1)$ observations of $\mathbf{X}$ can be arranged in $(i-1)!$ different ways and afterwards, the ith observation can be arranged in $2^{j_i}(i-2)^{1-j_i}$ different ways. Thus, this event can occur in $(i-1)!\, 2^{j_i}(i-2)^{1-j_i}$ different ways. Since these observations are iid and continuous, all permutations of the i observations have equal probabilities. As a result, the probability can be obtained as

$$P(\xi_i = j_i) = (2^{j_i}(i-2)^{1-j_i})/i \quad (3)$$

which is the pmf of bernoulli distribution with parameter $2/i$.

(ii) For $i \in \{3,4,5,\ldots\}$ and $j_3, j_4, \ldots, j_i \in \{0,1\}$, let us present how many different ways the event $A_i = \{\xi_3 = j_3, \xi_4 = j_4, \ldots, \xi_i = j_i\}$ can occur. The first 2 observations of $\mathbf{X}$ can be arranged in 2 different ways. In general, it is clear that the ith observation can be arranged $2^{j_i}(i-2)^{1-j_i}$ different ways considering the $(i-1)$ observations previously arranged. As a result, the event A_i can occur in

$$2^{1+j_3+j_4+\ldots+j_i} \prod_{k=4}^{i} (k-2)^{1-j_k} \quad (4)$$

different ways. In iid case, since all arrangements of the i observations are equal probabilities, the probability of event A_i can be written as

$$P(A_i) = \frac{2^{1+j_3+j_4+\ldots+j_i} \prod_{k=4}^{i} (k-2)^{1-j_k}}{i!} \quad (5)$$

From this point of view, one can easily show that

$$P(A_i) = \prod_{k=3}^{i} P(\xi_k = j_k). \; \square \quad (6)$$

From Definition 1, it is clear that $N_n' = \xi_2 + \xi_3 + \ldots + \xi_n$. Thus, the expected value of N_n' can easily be obtained by using Lemma 1(i) as follows:

$$E(N_n') = \sum_{i=2}^{n} E(\xi_i) = \sum_{i=2}^{n} \frac{2}{i}. \quad (7)$$

Also, the variance of N_n' is

$$\begin{aligned} V(N_n') &= \sum_{i=2}^{n} V(\xi_i) = \sum_{i=2}^{n} \frac{2}{i}\left(1 - \frac{2}{i}\right) \\ &= E(N_n') - \sum_{i=2}^{n} \frac{4}{i^2}. \end{aligned} \quad (8)$$

Remark 1. Let N_n indicates the number of upper records among iid and continuous random variables $X_1, X_2, \ldots, X_n$, it is well known that the expected value of N_n for large n is

$$E(N_n) = \sum_{i=1}^{n} \frac{1}{i} \cong log\, n + \gamma \quad (9)$$

where γ is Euler's constant and equals to 0.5772 … [13]. Thus, using (9), (7) can be rewritten as

$$\mathrm{E}(N_n') = 2(E(N_n) - 1) \cong 2(\log n + \gamma - 1) \quad (10)$$

In addition, using (8) one can write

$$\mathrm{E}(N_n') - \mathrm{V}(N_n') \cong (2/3)\pi^2 - 4 \quad (11)$$

for large n. □

Hereby, the probability mass function (pmf) of the number of record range in iid case is given by the following theorem.

Theorem 1. For $k = 1,2,3,\ldots,n-1$, the pmf of N_n' in any iid continuous finite random sequence is

$$P(N_n' = k) = \text{coefficient of } s^k \text{ in } \frac{\prod_{i=0}^{n-2}(2s+i)}{n!}. \quad (12)$$

Proof. From Lemma 1(ii), the probability generating function of N_n', $E\left(s^{N_n'}\right)$, is equal to the product of the $(n-1)$ probability generating functions of the Bernoulli random variables $\xi_2, \xi_3, \dots, \xi_n$. Using Lemma 1(i), one can obtain $E\left(s^{\xi_i}\right) = (2s+i-2)/i$ and $E\left(s^{N_n'}\right) = (\prod_{i=0}^{n-2}(2s+i))/n!$. Thus, the proof is completed. □

2.2. Probability of changing by a new observation

Let T_m, which will be defined in Section 2.3, indicates the time of the mth record range in **X**. For m and $i \in \{1,2,3,\dots\}$, it is clear that $B_i = \left\{X_{T_m+i} \notin (L_m', U_m')\right\}$ is an event that the ith new observation changes the mth record range. In addition, for iid case, the joint density of L_m' and U_m' is known as follows [14]: For $l < u$,

$$f^{L_m', U_m'}(l,u) = \frac{2^m f(l) f(u)\left[-log\left(1-F(u)+F(l)\right)\right]^{m-1}}{(m-1)!}. \quad (13)$$

Before the probability of event B_i is derived, the following Ψ_a function which simplify its proof is given:

$$\Psi_a(b,c) = \int_a^1 x^b (logx)^c \, dx \,. \quad (14)$$

It is true that

$$\Psi_a(b,c) = \frac{(-1)^c c!}{(b+1)^{c+1}} - a^{b+1} \sum_{j=0}^{c} \frac{(-1)^c c! \left(log\frac{1}{a}\right)^{c-j}}{(b+1)^{1+j}(c-j)!} \quad (15)$$

for $a \in [0,1] \subset \mathbb{R}$ and $b, c \in \{0,1,2,\dots\}$ [21].

The probability of the event B_i is given with the following theorem.

Theorem 2 It is true that

$$P\left(X_{T_m+i} \notin (L_m', U_m')\right) = \left(\frac{2}{3}\right)^m. \quad (16)$$

Proof. Recall that the event B_i is $\left\{X_{T_m+i} \notin (L_m', U_m')\right\}$. Let the complement of B_i is indicated by ${B_i}^c$.Then, one can write

$$P({B_i}^c) = \int_{-\infty}^{\infty}\int_{-\infty}^{u}\left(F(u)-F(l)\right) f^{L_m', U_m'}(l,u)\, dl du \quad (17)$$

and using (13), (17) can be rewritten as

$$\begin{aligned} P({B_i}^c) &= \int_0^1\int_0^y \frac{(-1)^{m-1}2^m}{(m-1)!}(y-x) \\ &\quad \times (\log(1-y+x))^{m-1}\, dxdy \\ &= \frac{(-1)^{m-1}2^m}{(m-1)!}\left[\int_0^1\int_0^y y(log(1-y+x))^{m-1} dxdy \right. \\ &\quad \left. - \int_0^1\int_0^y x(log(1-y+x))^{m-1} dxdy\right] \end{aligned} \quad (18)$$

and considering (14) one can obtain

$$\begin{aligned} \pi_1 &:= \int_0^1\int_0^y y(\log(1-y+x))^{m-1} dxdy \\ &= \int_0^1 y\psi_{1-y}(0,m-1)dy \end{aligned} \quad (19)$$

and

$$\begin{aligned} \pi_2 &:= \int_0^1\int_0^y x(\log(1-y+x))^{m-1} dxdy \\ &= \int_0^1 \psi_{1-y}(1,m-1)dy \\ &\quad - \int_0^1 (1-y)\psi_{1-y}(0,m-1)dy \end{aligned} \quad (20)$$

Consequently, one has

$$\begin{aligned} \pi_1 - \pi_2 &= \int_0^1 \psi_{1-y}(0,m-1)dy \\ &\quad - \int_0^1 \psi_{1-y}(1,m-1)dy \end{aligned} \quad (21)$$

As a result, using (15) into (21), one can reach

$$\pi_1 - \pi_2 = (-1)^{m-1}(m-1)!\left[\frac{1}{2^m} - \frac{1}{3^m}\right] \quad (22)$$

which completes the proof. □

Remark 2. Although, the probability of the mth change of the upper record by a new observation in **X** is equal to $1/2^m$ ([22], Lemma 2.1., p.17), such changing of record range is $(2/3)^m$.

2.3. Times of record ranges

The times of record ranges (or the times of record coverages) can be defined recursively as the following definition. Formally, let $X_{i:n}$ denote the ith order statistics from a random sample of size n.

Definition 2. In an iid continuous random sequence as $X_1, X_2, X_3, \dots$, the times of record ranges are defined recursively as the following: $T_1 = 2$ and for $j = 2,3,4, \dots$.

$$T_j = min\{i : X_i < X_{1:i-1} \text{ or } X_i > X_{i-1:i-1}, i > T_{j-1}\}. \, □(23)$$

After Definition 2, the joint pmf of $T_2, T_3, \dots, T_m$ is derived in the following theorem:

Theorem 3. For iid case, the joint pmf of $T_2, T_3, \dots, T_m$ is $p_{2,3,\dots,m}(t_2, t_3, \dots, t_m) = \frac{2^m}{t_m(t_m-1)\prod_{i=2}^{m}(t_i-2)}$ where $m \in \{2,3,4,\dots\}$ and $2 = t_1 < t_2 < t_3 < \dots < t_m$.

Proof. Using the indicator random variables in Definition 1, the following relation can be written:

$$\{T_2 = t_2, T_3 = t_3, \dots, T_m = t_m\} \equiv \left\{\begin{array}{c} \xi_3 = 0, \xi_4 = 0, \dots, \xi_{t_2-1} = 0, \xi_{t_2} = 1, \\ \xi_{t_2+1} = 0, \xi_{t_2+2} = 0, \dots, \xi_{t_3-1} = 0, \xi_{t_3} = 1, \\ \dots, \xi_{t_{m-1}+1} = 0, \xi_{t_{m-1}+2} = 0, \dots \xi_{t_m-1} = 0, \xi_{t_m} = 1 \end{array}\right\} \quad (24)$$

Thus, the theorem can be proved by considering this relation and Lemma 1 in Section 2.1. □

Remark 3. It's clear that $E(T_1) = 2$ since $T_1 = 2$. For $m = 2,3,4,\ldots$, using the joint pmf of $T_2, T_3, \ldots, T_m$ in Theorem 3, one can easily write the expected value of the time of mth record range as

$$E(T_m) = 2^m \sum_{t_2=3}^{\infty} \frac{1}{t_2-2} \sum_{t_3=t_2+1}^{\infty} \frac{1}{t_3-2} \sum_{t_4=t_3+1}^{\infty} \frac{1}{t_4-2} \cdots \sum_{t_{m-1}=t_{m-2}+1}^{\infty} \frac{1}{t_{m-1}-2} \sum_{t_m=t_m+1}^{\infty} \frac{1}{(t_m-1)(t_m-2)} \quad (25)$$

and using the equation

$$\sum_{i=j+1}^{\infty} \frac{1}{(i-1)(i-2)} = \frac{1}{(j-1)}, \quad (26)$$

it can be shown that the expected value of T_m is

$$E(T_m) = 2^m, \quad m = 1,2,3,\ldots. \quad (27)$$

Although the expected times of upper (or lower) records are infinite in classical record theory [13], the times of record ranges have finite means.

Remark 4. For $\alpha \in \{2,3,4,\ldots\}$, it is easy to prove that the αth moment of T_m and the product moments of T_u and T_v $(u \neq v)$ are infinite since $\sum_{i=j+1}^{\infty} i/((i-1)(i-2)) = \infty$ for any integer j.

Acknowledgment

The authors thank the anonymous referee for valuable comments which resulted in the improvement of the presentation of this paper.

References

[1] Chandler, K.N., 1952. The distribution and frequency of record values. Journal of the Royal Statistical Society, Series B, 14, 220-228.

[2] Foster, F.G. and Stuart, A., 1954. Distribution-free tests in time-series based on the breaking of records (with discussions). Journal of the Royal Statistical Society, Series B., 16, 2-22.

[3] Shorrock, R.W., 1972. On record values and record times. Journal of Applied Probability, 9, 316-326.

[4] Resnick, S.I., 1973. Record values and Maxima. The Annals of Probability, 1(4), 650-662.

[5] Galambos, J., 1978. The Asymptotic Theory of Extreme Order Statistics. John Wiley, New York.

[6] Glick, N., 1978. Breaking records and breaking boards. American Mathematical Monthly, 85, 2-26.

[7] Dunsmore, I.R., 1983. The future occurence of records. Annals of the Institute of Statistical Mathematics, 35, 267-277.

[8] Nagaraja, H.N., 1988. Record Values and Related Statistics. A review, Communication in Statistics, Theory and Methods, 17, 2223-2238.

[9] Nevzorov, V., 1988. Records. Theor. Prob. Appl. 32, 201-228.

[10] Ahsanullah, M., 1995. Record Statistics. Nova Science Publishers Inc. Commack. NY.

[11] Borovkov, K., Pfeifer, D., 1995. On record indices and record times. Journal of Statistical Planning and Inference, 45, 65-80.

[12] Feuerverger, A., Hall, P., 1998. On statistical inference based on record values. Extremes. 1(2), 169-190.

[13] Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N., 1998. Records. John Wiley. New York.

[14] Houchens, R.L., 1984. Record value, theory and inference. Ph.D. Dissertation, University of California, Riverside.

[15] Basak, P., 2000. An application of record range and some characterization results. In: Balakrishnan, N. (eds.) Advances on Theoretical and Methodological Aspects of Probability and Statistics, pp. 83-95. Newark, NJ, Gordon and Breach.

[16] Ahmadi, J., Balakrishnan, N., 2004. Confidence intervals for quantiles in terms of record range. Statistics and Probability Letters, 68, 395-405.

[17] Ahmadi, J., Balakrishnan, N., 2005. Distribution-free confidence intervals for quantile intervals based on current records. Statistics and Probability Letters 75, 190-202.

[18] Raqab, M.Z., 2007. Inequalities for expected current record statistics. Communication in Statistics, Theory and Methods, 36, 1367-1380.

[19] Raqab, M.Z., 2009. Distribution-free prediction intervals for the future current record statistics. Stat Papers, 50, 429-439.

[20] Ahmadi, J., Balakrishnan, N., 2011. Distribution-free prediction intervals for orders statistics based on record coverage. Journal of the Korean Statistical Society, 40, 181-192.

[21] Tanil, H., 2009. An order statistics model based on the list of top m scores after ℓth change. Journal of Statistical Planning and Inference, 139, 2189-2195.

[22] Bairamov, I.G., 1997. Some distribution free properties of statistics based on record values and characterizations of the distribution through a record. Journal of Applied Statistical Science. Sci, 5(1), 17-25.