

# Ignatov's theorem and correlated record values

Erol A. Peköz \*

*Statistics Department, University of California, Berkeley, CA 94720, USA*

Received May 1998

---

## Abstract

For a sequence of i.i.d. random variables an upper (lower)  $k$ -record is a value which is  $k$ th largest (smallest) at its appearance. A consequence of Ignatov's theorem is that the event a value  $x$  ever appears as an upper  $k$ -record is independent of the event a value  $y$  ever appears as an upper  $j$ -record, when either  $j \neq k$  or  $x \neq y$ . We study the correlation between the event a value ever appears as an upper  $k$ -record and the event a value ever appears as a lower  $j$ -record, and show that it can be strictly positive or negative depending on simple criteria. We also show that these events are, as either  $j$  or  $k$  increases, asymptotically independent. © 1999 Published by Elsevier Science B.V. All rights reserved

*Keywords:* Random variables; Ignatov's theorem;  $k$ -record

---

## 1. Introduction and summary

Let  $X_1, X_2, \dots$ , be real-valued independent and identically distributed random variables, and call a value  $x$  an upper (lower)  $k$ -record if there is an  $n$  where  $X_n = x$  and  $X_n \leq (\geq) X_i$  for exactly  $k$  of the values  $i, i = 1, \dots, n$ . Let  $U_k$  and  $L_k$ , respectively be the set of upper and of lower  $k$ -records. It is a consequence of Ignatov's theorem (see Engelen et al. (1988), Samuels (1992), and Adler and Ross (1997)) that the events  $\{x \in U_k\}$  and  $\{y \in U_j\}$  are independent when either  $x \neq y$  or  $j \neq k$ . In Section 2 we show that the events  $\{x \in U_k\}$  and  $\{x \in L_j\}$ , when non-trivial, are either positively correlated, negatively correlated, or uncorrelated for all  $k, j$  depending, respectively, on whether  $P(X_1 = x)/(P(X_1 \geq x)P(X_1 \leq x))$  is less than, greater than, or equal to 1. We also show that for all  $j, k$  the events  $\{x \in U_k\}$  and  $\{y \in L_j\}$ , when non-trivial, are negatively correlated when  $x < y$  and independent when  $x > y$ . In Section 3 we show that these events are, as either  $j$  or  $k$  increases, asymptotically independent.

Non-triviality of the events  $\{x \in U_k\}$  and  $\{x \in L_j\}$  essentially implies  $x$  is an atom of the distribution, and is not the supremum or infimum of the range. Thus part of the main theorem essentially says that for such values  $x$  which are sufficiently rare and sufficiently in the center of the range of the random variable, a value being an upper record is surprisingly strictly positively correlated with that value being a lower record.

---

\* Current address: Dept. of Finance, Marketing, and Decision Science, Western Washington University, Bellingham, WA 98225-9077, USA. Tel.: +1-360-650-7309; fax: +1-360-650-4844; e-mail: pekoze@cc.wvu.edu.

## 2. Correlated records

In this section we prove the following theorem.

**Theorem 2.1.** *With the above definitions and any  $j = 1, 2, \dots$ , and  $k = 1, 2, \dots$ , the events  $\{x \in U_k\}$  and  $\{y \in L_j\}$ , when non-trivial, are*

1. *independent when  $x > y$ ,*
2. *strictly negatively correlated when  $x < y$ , and*
3. *either strictly positively correlated, strictly negatively correlated, or uncorrelated depending respectively on whether*

$$P(X_1 = x)/(P(X_1 \geq x)P(X_1 \leq x))$$

*is less than, greater than, or equal to 1 when  $x = y$ .*

The theorem follows immediately from the two lemmas below. Let  $Y_n = \sum_{i=1}^n \mathbf{1}_{X_i \geq x}$ ,  $Z_n = \sum_{i=1}^n \mathbf{1}_{X_i \leq y}$ , and note that  $(Y_n, Z_n)$  is a Markov random walk. Also let  $T(i) = \min\{n: Y_n = i\}$ , and  $V(i) = \min\{n: Z_n = i\}$ . In what follows  $\{x \in U_k\}$  and  $\{y \in L_j\}$  are assumed non-trivial.

The idea behind the proof is that the event  $x$  is an upper  $k$ -record is conditionally independent of the event  $y$  is a lower  $j$ -record given that the records happen at different times in the sequence of random variables. In the rare event that the records occur at the same time, a positive or negative correlation can be obtained depending on the criteria of the theorem.

**Lemma 2.1.** *The following hold:*

1.  $P(x \in U_k) = P(X_1 = x)/P(X_1 \geq x)$ ,
2.  $P(y \in L_j) = P(X_1 = y)/P(X_1 \leq y)$ .

**Proof.** To prove (1) first note that  $\{x \in U_k\} = \{X_{T(k)} = x\}$  and  $\{y \in L_j\} = \{X_{V(j)} = y\}$ . Thus

$$\begin{aligned} P(x \in U_k) &= P(X_{T(k)} = x) \\ &= \sum_{i=1}^{\infty} P(X_{T(k-1)+1} < x, \dots, X_{T(k-1)+i-1} < x, X_{T(k-1)+i} = x) \\ &= \sum_{i=1}^{\infty} (P(X_1 < x))^{i-1} P(X_1 = x) \\ &= P(X_1 = x)/P(X_1 \geq x), \end{aligned}$$

where the third line follows from the strong Markov property. Non-triviality of the event  $\{x \in U_k\}$  implies  $0 < P(X_1 = x) < 1$ ,  $0 < P(X_1 > x) < 1$ , and  $0 < P(X_1 > x) < 1$ . Using a similar argument, (2) follows.  $\square$

**Lemma 2.2.** *Let  $H(m, n) = \{\exists i \geq 1: Z_i = m, Y_i = n\}$ , and put  $p = P(H(j-1, k-1))$ . Then when  $x \leq y$ ,*

$$P(x \in U_k, y \in L_j) = \frac{P(X_1 = x) P(X_1 = y)}{P(X_1 \geq x) P(X_1 \leq y)} (1 - p) + p P(X_1 = x) \mathbf{1}_{x=y}$$

and  $0 < p < 1$ .

**Proof.** For  $n = 0, 1, \dots, j$ , let

$$E_n = \{Z_{T(k-1)} = j - 1 - n\}$$

be the event that at the time we have observed  $k - 1$  variables with values greater than or equal to  $x$  there have also been  $j - 1 - n$  variables with values less than or equal to  $y$ . Letting

$$R_i = \bigcap_{l=1}^{i-1} \{X_{T(k-1)+l} < x\} \cap \{X_{T(k-1)+i} = x\}$$

and

$$S_i = \bigcap_{l=1}^{i-1} \{X_{V(j-1)+l} > y\} \cap \{X_{V(j-1)+i} = y\},$$

$$\begin{aligned} P(x \in U_k, x \in L_j | E_n) &= P(X_{T(k)} = x, X_{V(j)} = y | E_n) \\ &= \sum_{i=1}^n \sum_{h=1}^{\infty} P(R_i \cap S_h | E_n) + P(R_{n+1} | E_n) 1_{x=y} \\ &= \sum_{i=1}^n P(R_i | E_n) \sum_{h=1}^{\infty} P(S_h | E_n) + P(R_{n+1} | E_n) 1_{x=y} \\ &= \frac{P(X_1 = x)}{P(X_1 \geq x)} \frac{P(X_1 = x)}{P(X_1 \leq x)} (1 - P(X_1 < x)^n) + P(X_1 < x)^n P(X_1 = x = y) \\ &= \frac{P(X_1 = x)}{P(X_1 \geq x)} \frac{P(X_1 = y)}{P(X_1 \leq y)} (1 - P(H(j-1, k-1) | E_n)) + P(H(j-1, k-1) | E_n) P(X_1 = x = y), \end{aligned}$$

where the third equality follows from the strong Markov property. Also, letting  $F_n = \{Y_{V(j-1)} = k - 1 - n\}$ , the same argument gives

$$\begin{aligned} P(x \in U_k, x \in L_j | F_n) &= \frac{P(X_1 = x)}{P(X_1 \geq x)} \frac{P(X_1 = y)}{P(X_1 \leq y)} (1 - P(H(j-1, k-1) | F_n)) \\ &\quad + P(H(j-1, k-1) | F_n) P(X_1 = x = y), \end{aligned}$$

and unconditioning then gives the first assertion of the lemma. The non-triviality of the events  $\{x \in U_k\}$  and  $\{y \in L_j\}$  imply  $0 < p < 1$ .  $\square$

We now assemble the pieces to prove the main theorem.

**Proof of Theorem 2.1.** When  $x \leq y$  the lemmas above together immediately give parts (2) and (3) of the theorem. For the case  $x > y$  the independence claimed in part (1) follows immediately from an argument identical to the one used for Lemma 2.1 in [2].  $\square$

### 3. Asymptotically independent records

In this section we show that the correlation between the events  $\{x \in U_k\}$  and  $\{y \in L_j\}$  converges to zero as either  $k$  or  $j$  increases. The essential idea is that the chance the records happen at the same time in the sequence goes to zero asymptotically. This is established by computing the mean and variance of a related quantity, and using Chebyshev’s inequality.

First it may be interesting to note the following formula for  $P(H(j, k))$  which appears in the expression for the correlation in Lemma 2.2.

**Proposition 3.1.** *For  $x \leq y$  let  $a = P(X_1 > y)$ ,  $b = P(X_1 < x)$ , and  $c = 1 - a - b$ , then*

$$P(H(j, k)) = \sum_{i=0}^{j \wedge k} \frac{(j+k-i)!}{(i)!(j-i)!(k-i)!} a^{k-i} b^{j-i} c^i.$$

**Proof.** Let  $A_n = \sum_{i=1}^n 1_{X_i < y}$ ,  $B_n = \sum_{i=1}^n 1_{X_i > x}$ , and  $C_n = \sum_{i=1}^n 1_{x \leq X_i \leq y}$ . Clearly,  $(A_n, B_n, C_n)$  has a multinomial distribution with  $n$  trials and probabilities  $(a, b, c)$ . Next,

$$\begin{aligned} P(H(j, k)) &= P(A_n + C_n = k, C_n + B_n = j \text{ for some } n) \\ &= \sum_n P(A_n + C_n = k, C_n + B_n = j) \\ &= \sum_n \sum_i P(A_n = k - i, B_n = j - i, C_n = i) \\ &= \sum_{i=0}^{j \wedge k} P(A_{j+k-i} = k - i, B_{j+k-i} = j - i, C_{j+k-i} = i) \end{aligned}$$

and the result follows.  $\square$

**Theorem 3.1.** *For any  $x, y$ ,*

$$\lim P(x \in U_k, y \in L_j) - P(x \in U_k)P(y \in L_j) = 0,$$

where the limit is taken as either  $k \rightarrow \infty$  with  $j$  fixed, or as  $j \rightarrow \infty$  with  $k$  fixed.

**Proof.** We will show

$$\lim_{k \rightarrow \infty} P(H(j, k)) = 0 \quad \text{and} \quad \lim_{j \rightarrow \infty} P(H(j, k)) = 0$$

and the theorem will then follow from Lemma 2.2. First note

$$\begin{aligned} P(H(j, k)) &\leq P(Z_{T(k)} \leq j) \\ &= P(E[Z_{T(k)}] - Z_{T(k)} \geq E[Z_{T(k)}] - j) \\ &\leq \frac{\text{Var}(Z_{T(k)})}{(E[Z_{T(k)}] - j)^2}, \end{aligned}$$

where the last line follows from Chebyshev’s inequality.

Next, using the notation from Proposition 3.1, since  $T(k) = \min\{n: B(n) + C(n) = k\}$ ,  $T(k)$  has a negative-binomial distribution with parameters  $k, 1 - a$ . Also note that  $C_{T(k)}$  has a binomial distribution with

parameters  $k, c/(c + b)$ , and is independent of  $T(k)$ . Next, since

$$Z_{T(k)} = A_{T(k)} + C_{T(k)} = T(k) - k + C_{T(k)},$$

we obtain that

$$E[Z_{T(k)}] = kc/(c + b) + (k/a - k) \equiv k\alpha$$

with  $\alpha > 0$ , and

$$\text{Var}(Z_{T(k)}) = kcb/(c + b)^2 + k/a^2 \equiv k\beta$$

with  $\beta > 0$ . Thus,

$$P(H(j, k)) \leq \frac{k\alpha}{(k\beta - j)^2} \rightarrow 0,$$

where the limit is either  $j \rightarrow \infty$  or  $k \rightarrow \infty$ . This establishes the theorem.  $\square$

## References

- Adler, I., Ross, S.M., 1997. Distribution of the time of the first  $k$ -record. *Probab. Eng. Inform. Sci.* 11, 273–278.  
 Engelen, R., Thommassen, P., Vervaat, W., 1988. Ignatov's theorem – a new and short proof. *J. Appl. Probab.* 25 (a), 229–236.  
 Samuels, S., 1992. All at once proof of Ignatov's theorem. *Contemp. Math.* 125, 231–237.