Adv. Appl. Prob. **22,** 957–960 (1990)
Printed in N. Ireland

LETTERS TO THE EDITOR

ON THE JOINT DISTRIBUTION OF THE NUMBER OF UPPER AND LOWER RECORDS AND THE NUMBER OF INVERSIONS IN A RANDOM SEQUENCE

W. KATZENBEISSER,* *Institute of Statistics, Vienna*

1. Introduction

Let $X_1, X_2, \cdots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with continuous distribution function $F(x)$. To this sequence the random variables U_n, L_n and I_n can be associated, where

$$U_n = [\text{number of upper records of the } X\text{'s}],$$
$$L_n = [\text{number of lower records of the } X\text{'s}],$$
$$I_n = [\text{number of inversions among the } X\text{'s}].$$

The random variable X_i is called an upper record, if $X_i > \max\{X_1, \cdots, X_{i-1}\}$; equivalently, X_j is called a lower record, if $X_j < \min\{X_1, \cdots, X_{j-1}\}$. By convention, X_1 is an upper record only. Therefore the support of the random variable U_n is the integers $1, 2, \cdots, n$ whereas the support of the random variable L_n is the integers $0, 1, \cdots, n-1$. The pair (X_i, X_j) constitutes an inversion of the X's if $X_i > X_j$ for $i < j$; the support of this random variable is the integers $0, 1, \cdots, \binom{n}{2}$.

Distributional properties of the random variables U_n, L_n, and I_n are extensively studied in the literature. For the record variables see, for example, Sparre Anderson (1954), Rényi (1962), Haghighi-Talab and Wright (1973), Resnick (1973), the series of papers by Shorrock (1972), (1973), (1974) and Pfeifer (1989). The application of the record variables as test statistics for tests against trend is discussed in Foster and Stuart (1954) and Brunk (1960). Some properties of the random variable I_n are given in, for example, Comtet (1972). Of course, I_n is related to Kendall's τ, i.e., $\tau = 1 - 2I_n \Big/ \binom{n}{2}$; distributional properties for τ can be found in any textbook on non-parametric statistical methods. Moreover, this random variable has some importance in the analysis of algorithms; see, for example, Knuth (1969), Kemp (1984) and Panny (1986). The bivariate distribution of (U_n, L_n) and some related distributions are discussed in Foster and Stuart (1954) and in David and Barton (1962); those of (U_n, I_n) are derived in Katzenbeisser (1988), respectively. However, the joint distribution of all three random variables U_n, L_n, and I_n has not so far been discussed in the literature.

The purpose of this paper is therefore to derive the joint probability generating function of U_n, L_n and I_n. This function will then be used to derive the first two moments, the covariances and the correlation coefficients between these random variables.

Received 26 February 1990; revision received 17 September 1990.

* Postal address: Institut für Statistik der Wirtschaftsuniversität Wien, A-1090 Wien, Augasse 2–6, Austria.

2. The joint distribution of U_n, L_n, I_n

In this section the joint probability generating function (p.g.f.) of the random variable (U_n, L_n, I_n), denoted by $G_n(x, y, z)$, will be derived, where

$$G_n(x, y, z) := \sum_{k\geqq 0}\sum_{l\geqq 0}\sum_{m\geqq 0} x^k y^l z^m p_n(k, l, m),$$

with $p_n(k, l, m) := P\{U_n = k, L_n = l, I_n = m\}$. In order to derive this p.g.f. our considerations will be based on the following well-known fact. Let $X_1, X_2, \cdots, X_{n-1}$ be a sequence of i.i.d. random variables with continuous distribution function $F(x)$ and denote by $X_{1,n-1}, X_{2,n-1}, \cdots, X_{n-1,n-1}$ the $n-1$ order statistics of the X's. Let X_n be a further random variable independent of $X_1, X_2, \cdots, X_{n-1}$, with the same distribution as the X's. Then it holds that the probability that X_n is included in any of the n intervals generated by the $n-1$ order statistics of the X's is $1/n$. Using this fact the joint p.g.f. can be derived via the following recurrence relation.

Theorem 1. The joint p.g.f. of the random variables (U_n, L_n, I_n) is given by

$$G_n(x, y, z) = \begin{cases} x, & \text{if } n = 1, \\ \dfrac{1}{n!}x(x+yz)(x+z+yz^2)\cdots(x+z+z^2+\cdots+yz^{n-1}), & \text{if } n\geqq 2. \end{cases}$$

Proof. Consider the decomposition of the event exactly k upper records, l lower records, and m inversions among the n X's into disjoint events:

$$\begin{aligned} &\{U_n = k, L_n = l, I_n = m\} \\ &\quad = \{U_{n-1} = k-1, L_{n-1} = l, I_{n-1} = m \wedge X_n > X_{n-1,n-1}\} \\ &\quad \cup \{U_{n-1} = k, L_{n-1} = l, I_{n-1} = m-1 \wedge X_{n-2,n-1} < X_n < X_{n-1,n-1}\} \cup \cdots \\ &\quad \cup \{U_{n-1} = k, L_{n-1} = l-1, I_{n-1} = m-(n-1) \wedge X_n < X_{1,n-1}\}. \end{aligned}$$

Because of the i.i.d. property of the random variables $X_1, X_2, \cdots, X_{n-1}$ and X_n we obtain the recursion

$$\begin{aligned} p_n(k, l, m) = {} & p_{n-1}(k-1, l, m)\frac{1}{n} + p_{n-1}(k, l, m-1)\frac{1}{n} + \cdots \\ & + p_{n-1}(k, l-1, m-(n-1))\frac{1}{n}, \end{aligned} \tag{1}$$

with initial conditions $p_1(k, l, m) = 1$ if $k = 1$, $l = m = 0$ and zero elsewhere, and $p_n(k, l, m) = 0$ if $k \notin \{1, 2, \cdots, n\}$ or $l \notin \{0, 1, \cdots, n-1\}$ or $m \notin \left\{0, 1, \cdots, \binom{n}{2}\right\}$, for $n \geqq 2$. An application of the p.g.f. $G_n(k, l, m)$ leads to the recursion

$$\begin{aligned} G_n(k, l, m) = {} & \frac{x}{n}G_{n-1}(k, l, m) + \frac{z}{n}G_{n-1}(k, l, m) + \cdots \\ & + \frac{z^{n-2}}{n}G_{n-1}(k, l, m) + \frac{yz^{n-1}}{n}G_{n-1}(k, l, m) \\ = {} & \frac{x+z+\cdots+z^{n-2}+yz^{n-1}}{n}G_{n-1}(k, l, m), \end{aligned}$$

for $n \geqq 2$, with initial condition $G_1(k, l, m) = x$. Iteration of this formula yields finally the expression given in Theorem 1.

https://doi.org/10.2307/1427570 Published online by Cambridge University Press

There seems to be no easy way to derive a 'closed form' expression for the $p_n(k, l, m)$'s; however, they can be calculated recursively using (1). From the p.g.f. one can also see that the joint distribution may be interpreted as the distribution of the sum of n independent random variables (V_i, W_i, Z_i), $i = 1, 2, \cdots, n$, with the individual p.g.f.

$$G_i(x, y, z) = \begin{cases} x, & \text{if } i = 1, \\ \frac{1}{i}(x + z + z^2 + \cdots + yz^{i-1}), & \text{otherwise.} \end{cases}$$

The p.g.f. can now be used to derive the (joint) moments of the random variables U_n, L_n, and I_n. Denote by $H_n := \sum_{i=1}^{n} 1/i$, the nth harmonic number and by $H_n^{(2)} = \sum_{i=1}^{n} 1/i^2$, and further let the vector $\boldsymbol{Y}$ be defined by $\boldsymbol{Y} = (U_n, L_n, I_n)'$; then we have the following.

Theorem 2. The vector of expectations and the variance–covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{Y}$ are given by

$$E\{\boldsymbol{Y}\} = \begin{pmatrix} H_n \\ H_n - 1 \\ \frac{1}{2}\binom{n}{2} \end{pmatrix},$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} H_n - H_n^{(2)} & \cdot & \cdot \\ 1 - H_n^{(2)} & H_n - H_n^{(2)} & \cdot \\ \frac{1}{2}(H_n - n) & \frac{1}{2}(H_n - n) & \frac{1}{36}\binom{n}{2}(2n+5) \end{pmatrix}.$$

Proof. These expressions follow by straightforward differentiation of the joint p.g.f.

Obviously, the correlation matrix $\boldsymbol{R}$ is given by

$$\boldsymbol{R} = \begin{pmatrix} 1 & \cdot & \cdot \\ \dfrac{1 - H_n^{(2)}}{H_n - H_n^{(2)}} & 1 & \cdot \\ \dfrac{\frac{1}{2}(H_n - n)}{\sqrt{(H_n - H_n^{(2)})\frac{1}{36}\binom{n}{2}(2n+5)}} & \dfrac{\frac{1}{2}(H_n - n)}{\sqrt{(H_n - H_n^{(2)})\frac{1}{36}\binom{n}{2}(2n+5)}} & 1 \end{pmatrix}.$$

Asymptotic approximations for the moments of U_n and L_n can now easily be derived. An application of the well-known facts (i) $H_n = \ln n + \gamma + o(n^{-1})$, as $n \to \infty$, where γ denotes Euler's constant, and (ii) $\lim_{n\to\infty} H_n^{(r)} = \zeta(r)$, where $\zeta(r)$ denotes Riemann's zeta-function. If r is an even integer, then we have $\zeta(r) = \frac{1}{2}|B_r|(2\pi)^r/r!$, where B_r denotes the rth Bernoulli number. Taking these facts into account we have for example, as $n \to \infty$

$$E\{U_n\} = \ln n + \gamma + o(n^{-1}),$$

$$\operatorname{var}\{U_n\} = \operatorname{var}\{L_n\} = \ln n + \gamma - \frac{\pi^2}{6} + o(n^{-1}),$$

$$\operatorname{cov}\{U_n, L_n\} = 1 - \frac{\pi^2}{6},$$

$$\operatorname{cov}\{U_n, I_n\} = \operatorname{cov}\{L_n, I_n\} = \tfrac{1}{2}\ln n + \tfrac{1}{2}\gamma - n + o(n^{-1}).$$

The first two expressions above are well known and are given for example in Rényi (1962) and in Knuth (1969) and Kemp (1984), respectively. Moreover, from the asymptotic approximations we immediately have the following result.

Theorem 3. The random variables U_n, L_n, and I_n are asymptotically uncorrelated.

References

ANDERSON, E. SPARRE (1954) On the fluctuation of sums of random variables. *Math. Scand.* **2,** 171–181.

BRUNK, H. D. (1960) On a theorem of E. Sparre Andersen and its application to tests against trend. *Math. Scand.* **4,** 305–326.

COMTET, L. (1972) *Advanced Combinatorics.* D. Reidel, Dordrecht.

DAVID, F. N. AND BARTON, D. E. (1962) *Combinatorial Chance.* Griffin, London.

FOSTER, F. G. AND STUART, A. (1954) Distribution-free tests in time-series based on the breaking of records. *J. R. Statist. Soc.* **16,** 1–22.

HAGHIGHI-TALAB, D. AND WRIGHT, C. (1973) On the distribution of records in a finite sequence of observations, with an application to road traffic problems. *J. Appl. Prob.* **10,** 556–571.

KATZENBEISSER, W. (1988) On the joint distribution of the random variables number of inversions and number of outstanding variables in a randomly arranged sequence. *Statistical Papers* **29,** 133–144.

KEMP, R. (1984) *Fundamentals of the Average Case Analysis of Particular Algorithms.* Wiley, New York; Teubner, Stuttgart.

KNUTH, D. E. (1969) *The Art of Computer Programming,* Vol. 1. Addison-Wesley, Reading, Mass.

PANNY, W. (1986) A note on the higher moments of the expected behavior of straight insertion sort. *Inf. Proc. Lett.* **22,** 175–177.

PFEIFER, D. (1989) Extremal processes, secretary problems and the $1/e$ law. *J. Appl. Prob.* **26,** 722–733.

RÉNYI, A. (1962) Théorie des éléments saillants d'une suite d'observations. *Proc. Coll. Comb. Methods in Prob. Th.* Aarhus University, 104–115.

RESNICK, S. I. (1973) Record values and maxima. *Ann. Prob.* **1,** 650–662.

SHORROCK, R. (1972) On record values and record times. *J. Appl. Prob.* **9,** 316–326.

SHORROCK, R. (1973) Record values and inter-record times. *J. Appl. Prob.* **10,** 543–555.

SHORROCK, R. (1974) On discrete time extremal processes. *Adv. Appl. Prob.* **6,** 580–592.