ON THE MEDIAN OF THE DISTRIBUTION OF EXCEEDANCES

By K. Sarkadi

Mathematical Institute of the Hungarian Academy of Sciences, Budapest, Hungary

The distribution of exceedances may be defined by the following formula (see, e.g. [2]), where x corresponds to the number of exceedances

$$(1)\qquad w(n_1, m, n_2, x) = \frac{\binom{n_1 + n_2 - m - x}{n_1 - m}\binom{x + m - 1}{m - 1}}{\binom{n_1 + n_2}{n_1}},$$

$$x = 0, 1, \cdots, n_2;\ n_1, n_2, m \leqq n$$

given natural numbers.

There are known—among others—the following two fundamentally equivalent models or representations of this distribution [2], [3], [4]:

A. Exceedances. We have two random samples of sizes n_1 and n_2, respectively, from the same continuous distribution. The number of exceedances is defined as the number of elements of the second sample which surpass at least $n_1 - m + 1$ elements of the first, for a fixed natural number $m \leqq n_1$. The distribution of the number of exceedances is given by formula (1).

B. Pascal model without replacement. An urn contains n_1 black and n_2 red balls. We draw balls from the urn until we have drawn m black balls. The distribution of the number of the red balls drawn is given by (1).

In [1], Gumbel proved that, for $n_1 = n_2$, the median of the number of exceedances is $m - 1$, more precisely that

$$(2)\qquad W(n,m,n,m - 1) = \tfrac{1}{2},$$

where W is the cumulated form of w,

$$W(n_1, m,n_2, x) = \sum_{y=0}^{x} w(n_1, m, n_2, y)$$

In this paper a simple proof of this result is given.

We shall, in fact, prove the following more general result:

$$(3)\qquad W(n_1, m_1, n_2, m_2 - 1) + W(n_2, m_2, n_1, m_1 - 1) = 1.$$

In terms of model A, the m_1th element of the first sample exceeds the m_2th element of the second sample if and only if the number of exceedances (y) takes one of the values $0, 1, \cdots, m_2 - 1$, these possibilities being mutually exclusive.

Received May 20, 1959; revised August 10, 1959.

Therefore, the first term of the left side in (3) denotes the probability that the m_1th element of the first sample (from above) is larger than the m_2th element of the second one; the second term denotes the probability of the opposite inequality. Since either the inequality or its opposite must hold, equation (3) is proved.

This proof can also be formulated with the notions of the Pascal model (B).

The author is indebted to Prof. E. J. Gumbel for drawing his attention to this problem.

REFERENCES

[1] E. J. Gumbel and H. Von Schelling, "The distribution of the number of exceedances," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 247–262.

[2] E. J. Gumbel, *Statistics of Extremes*, Columbia University Press, New York, 1958.

[3] K. Sarkadi, "On the distribution of the number of exceedances," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 1021–1023.

[4] Károly Sarkadi, "Generalized hypergeometric distributions," *Publ. Math. Inst. Hung. Ac. Sc.*, Vol. II (1957), pp. 59–69.