NOTES

ON THE DISTRIBUTION OF THE NUMBER OF EXCEEDANCES

By K. Sarkadi

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

0. Introduction. In paper [1] E. J. Gumbel and H. von Schelling deal with the distribution discussed below, called by them "distribution of the number of exceedances."

Suppose that we have $n + N$ independent observations, regarded as two samples of sizes n and N, respectively, from a population with a continuous distribution function. Let us denote by ξ the number of those elements of the sample of size N which surpass (are larger than) at least $n - m + 1$ elements of the sample of size $n(1 \leqq m \leqq n)$. Thus ξ shows how many elements of the second sample exceed a given order statistic of the first sample; ξ is called by the authors the "number of exceedances." The distribution of ξ is given by the following formula:

$$p_x = P(\xi = x) = \frac{\binom{n}{m} m \binom{N}{x}}{(n + N)\binom{N + n - 1}{m + x - 1}} \qquad (x = 0, 1, \cdots, N) \tag{1}$$

Papers [2], [3], and [4] deal also with the above distribution.

The aim of the present paper is to show that the distribution of the number of exceedances is a special case of the Pólya-distribution. In addition relationships to other distributions as to Laplace's law of succession, etc., are mentioned.

1. Comparison of the distributions. The formula defining the Pólya-distribution is as follows:

$$P(\xi = x) = \binom{N}{x} \frac{\prod_{i=0}^{x-1} (m + iR) \prod_{j=0}^{N-x-1} (n - m + jR)}{\prod_{k=0}^{N-1} (n + kR)} \tag{2}$$

(see, e.g., [5], p. 12, and [6], p. 128.)

If in Eq. (2) we put $n + 1^1$ and 1^1 instead of n and R, respectively, we obtain formula (1).

L. Takács called my attention to the fact that the formula of the number of exceedances agrees with that of Laplace's law of succession.

Suppose that an event with a priori rectangular probability distribution in

Received September 25, 1956; revised May 27, 1957.

interval (0, 1) occurs m^* times in n^* trials. Then the probability that it will occur x times in a following set of N trials is as follows:

$$\binom{N}{x} \frac{\int_0^1 z^{m^*+x}(1 - z)^{n^*+N-m^*-x}\, dz}{\int_0^1 z^{m^*}(1 - z)^{n^*-m^*}\, dz} \tag{3}$$

(see [7], p. 31; [8], pp. 68–69). In most textbooks the law of succession appears for the case $m^* = n^*$, $N = 1$ only (see, e.g., [6], pp. 83–85).

Putting $n^* = n - 1$, $m^* = m - 1$, formula (3) goes into (1).

The classical inverse problem of the sampling without replacement is a further example which leads to the same distribution. An urn contains L balls the number of red ones out of them being a random variable equally distributed on the numbers 0, 1, 2, $\cdots$, L. We take n^* random drawings without replacement. Suppose that m^* out of the n^* balls turn out to be red. Then the a posteriori probability that the urn contained r balls is as follows ([9], pp. 109–110):

$$\frac{\binom{r}{m^*}\binom{L - r}{n^* - m^*}}{\binom{L + 1}{n + 1}} \tag{4}$$

Taking $r = x + m + 1$, $n^* = n - 1$, $m^* = m - 1$, $L = N + n + 1$, formula (4) goes into (1).

Similarly, it can be shown that some distributions treated in papers [10], [11], [12], [13], and [14] are of type Pólya.

It is not difficult to find appropriate models illustrating the fact that the above different problems lead to identical formulae. The author wishes to give these models elsewhere.

2. The moments. The moments of the number of exceedances can be derived from that of Pólya's distribution (see, e.g., [15]) too.

Kozniewska [16] determined also the mean deviation (the first absolute central moment) of the Pólya-distribution. Applying her results we obtain for the mean deviation of the number of exceedances

$$2rp_r\left(\frac{n - m + 1}{n + 1} + \frac{N - r}{N}\right),$$

where $r - 1$ denotes the greatest integer $\leqq Nm/(n + 1)$.

3. Limiting forms. The limiting forms of the Pólya distribution are treated by Bricas [15] in detail. The limiting forms of the distribution of the number of exceedances are special cases of the distributions derived by Bricas. Particularly, the "law of rare exceedances," the formula given in [1] for the limiting case $n = N \to \infty$, m remaining finite,

$$w = \binom{x + m - 1}{x}\left(\frac{1}{2}\right)^{m+x}$$

provides a distribution of Pascal type.

4. The discrete case. It is known that Laplace's law of succession is valid in the case of a finite population too (see [8], p. 72; [9], pp. 110–111). The answer of the problem is given by formula (3) independently of the size of the population in case of sampling without replacement.

Similarly the problem of the number of exceedances permits the same generalization.

An urn contains L balls numbered with different real numbers. A group of n balls is chosen at random without replacement, and following that a second group of N balls is chosen without replacement too. We define the number of exceedances as in the continuous case (see the Introduction). Provided that $L \geq n + N$, we obtain the same distribution as there, independently of L.

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, *Statistical theory of extreme values and some practical applications*, National Bureau of Standards, Washington, 1954.
[3] E. J. Gumbel, "Elementare Ableitung der Momente für die Zahl der Überschreitungen," *Mitteilungsblatt für Math. Stat.*, Vol. 6 (1954), pp. 164–169.
[4] B. Epstein, "Tables for the distribution of the number of exceedances," *Ann. of Math. Stat.*, Vol. 25 (1954), pp. 762–768.
[5] O. Lundberg, *On Random Processes and Their Applications to Sickness and Accident Statistics*, Almqvist, Uppsala, 1940.
[6] W. Feller, *Probability Theory and Its Applications*, John Wiley and Sons, New York, 1950.
[7] P. S. Laplace, *Oeuvres completes*, t. VIII., Gauthier-Villars, Paris, 1891.
[8] J. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill Book Co., New York, 1937.
[9] H. Jeffreys, *Theory of Probability*, 2nd. ed., Oxford University Press, Oxford, 1948.
[10] J. G. Skellam, "A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials," *Jour. Roy. Stat. Soc.*, Series B., Vol. 10 (1948), pp. 257–261.
[11] J. O. Irwin, "A distribution arising in the study of infectious diseases," *Biometrika*, Vol. 41 (1954), pp. 266–268.
[12] J. W. Hopkins, "An instance of negative hypergeometric sampling in practice," *Bull. Int. Stat. Inst.*, Vol. 34 (1955), pp. 298–306.
[13] C. D. Kemp and A. W. Kemp, "Generalized hypergeometric distributions," *J. Roy. Stat. Soc.*, Series B, Vol. 18 (1956), pp. 202–212.
[14] K. Sarkadi, "On the a priori beta-distribution of fraction defective," *Magyar Tudományos Akadémia Alkalmazott Matematikai Intézetének Közleményei*, Vol. II (1956), pp. 287–293.
[15] M. A. Bricas, *Le systeme de courbes de Pearson et le schéma d'urne de Pólya*, Cristou, Athens, 1949.
[16] J. Kozniewska, "The first absolute central moment for Pólya's distribution," *Zastosowania Matematyki*, Vol. I (1954), pp. 206–211.