AN APPLICATION OF A BALLOT THEOREM IN ORDER STATISTICS[1]

By Lajos Takács

Columbia University

1. Introduction. By making use of two simple combinatorial theorems the author [6], [7] arrived at the following extension of the classical ballot theorem:

Theorem 1. *Let $\nu_1, \nu_2, \cdots, \nu_{n+1}$ be interchangeable random variables that assume nonnegative integer values and write $N_r = \nu_1 + \nu_2 + \cdots + \nu_r$ for $r = 1, 2, \cdots, n+1$. Denote by $\Delta_{n+1}^{(c)}$ the number of subscripts $r = 1, 2, \cdots, n+1$ for which $N_r < r + c$ where c is a nonnegative integer. Then*

$$\mathbf{P}\{\Delta_{n+1}^{(0)} = j \mid N_{n+1} = n\} = 1/(n+1) \tag{1}$$

for $j = 1, 2, \cdots, n+1$ and

$$\mathbf{P}\{\Delta_{n+1}^{(c)} = n+1 \mid N_{n+1} = n\} = 1 - \sum_{i=1}^{n-c} \frac{c+1}{n+1-i} \mathbf{P}\{N_i = i + c \mid N_{n+1} = n\} \tag{2}$$

provided that the conditional probabilities are defined.

Now we shall give two examples for the application of this theorem in order statistics.

2. Two distribution-free statistics. Let $\xi_1, \xi_2, \cdots, \xi_m, \eta_1, \eta_2, \cdots, \eta_n$ be mutually independent random variables having a common continuous distribution function. Denote by $F_m(x)$ and $G_n(x)$ the empirical distribution functions of the samples $(\xi_1, \xi_2, \cdots, \xi_m)$ and $(\eta_1, \eta_2, \cdots, \eta_n)$ respectively. It is supposed that $F_m(x)$ and $G_n(x)$ are continuous on the right. Denote by $\eta_1^*, \eta_2^*, \cdots, \eta_n^*$ the random variables $\eta_1, \eta_2, \cdots, \eta_n$ arranged in increasing order. Let $\gamma(m, n)$ be the number of subscripts $r = 1, 2, \cdots, n$ for which $F_m(\eta_r^*) \leqq G_n(\eta_r^* - 0)$, i.e., $\gamma(m, n)$ is equal to the number of positive jumps of $G_n(x)$ relative to $F_m(x)$. Further let

$$\delta^+(m, n) = \sup_{-\infty < x < \infty} [F_m(x) - G_n(x)]. \tag{3}$$

It is easy to see that $\gamma(m, n)$ and $\delta^+(m, n)$ are distribution-free statistics. The distribution of the random variable $\gamma(m, n)$ for $n = m$ was found by B. V. Gnedenko and V. S. Mihalevič [3] and for $n = mp$, where p is a positive integer, by B. V. Gnedenko and V. S. Mihalevič [4]. The distribution of the random variable $\delta^+(m, n)$ for $n = m$ was found by B. V. Gnedenko and V. S. Koroljuk [2] and for $n = mp$, where p is a positive integer, by V. S. Koroljuk [5]. In this paper we shall show that if $n = mp$, where p is a positive integer, then the distributions of $\gamma(m, n)$ and $\delta^+(m, n)$ can easily be obtained by using Theorem 1.

Received 23 January 1964.

[1] This research was supported by the Office of Naval Research under Contract Number Nonr-266(59).

THEOREM 2. *If* $n = mp$, *where* p *is a positive integer, and* c *is a nonnegative integer, then*

$$\mathbf{P}\{\gamma(m, n) = j\} = 1/(n + 1) \tag{4}$$

for $j = 0, 1, \cdots, n$, *and*

$$\mathbf{P}\left\{\delta^+(m, n) \leqq \frac{c}{n}\right\} = 1 - \sum_{(c+1)/p \leqq s \leqq m} \frac{c + 1}{n + c + 1 - sp} \cdot \binom{sp + s - c - 1}{s}\binom{m + n + c - sp - s}{m - s}\Big/\binom{m + n}{m}. \tag{5}$$

PROOF. Let ν_r, $r = 1, 2, \cdots, n + 1$, be p times the number of variables $\xi_1, \xi_2, \cdots, \xi_m$ falling in the interval $(\eta^*_{r-1}, \eta^*_r]$ where $\eta^*_0 = -\infty$ and $\eta^*_{n+1} = +\infty$, and write $N_r = \nu_1 + \nu_2 + \cdots + \nu_r$ for $r = 1, 2, \cdots, n + 1$. Now $\nu_1, \nu_2, \cdots, \nu_{n+1}$ are interchangeable random variables for which $N_{n+1} = mp$ and

$$\mathbf{P}\{N_i = sp\} = \binom{i + s - 1}{s}\binom{m + n - i - s}{m - s}\Big/\binom{m + n}{m}. \tag{6}$$

Evidently $F_m(\eta^*_r) = N_r/mp$ and $G_n(\eta^*_r - 0) = (r - 1)/n$ for $r = 1, \cdots, n$. If $n = mp$, then $N_{n+1} = n$ and $\gamma(m, n)$ equals the number of subscripts $r = 1, \cdots, n$ for which $N_r < r$. Since $N_{n+1} < n + 1$ also holds, we have $\gamma(m, n) = \Delta^{(0)}_{n+1} - 1$ where by (1) $\mathbf{P}\{\Delta^{(0)}_{n+1} = j\} = 1/(n + 1)$ for $j = 1, \cdots, n + 1$. This proves (4). To prove (5) we note that if $n = mp$, then

$$\delta^+(m, n) = \max_{1 \leqq r \leqq n} [F_m(\eta^*_r) - G_n(\eta^*_r - 0)] = n^{-1} \max_{1 \leqq r \leqq n+1} (N_r - r + 1).$$

Thus

$$\mathbf{P}\{\delta^+(m, n) \leqq c/n\} = \mathbf{P}\{N_r < r + c \text{ for } r = 1, \cdots, n + 1\}$$

and the right hand side is given by (2) where N_i has the distribution (6). This proves (5). It should be noted that if $p = 1$, then (5) reduces to

$$\mathbf{P}\left\{\delta^+(n, n) \leqq \frac{c}{n}\right\} = 1 - \binom{2n}{n + 1 + c}\Big/\binom{2n}{n}. \tag{7}$$

Finally, we mention that E. F. Drion [1] has considered a related problem. He found that the probability that $\inf_{0<G_n(x)<1} [F_m(x) - G_n(x)] > 0$ is $1/(4n - 2)$ if $m = n$, and $1/(m + n)$ if $(m, n) = 1$.

REFERENCES

[1] DRION, E. F. (1952). Some distribution-free tests for the difference between two empirical cumulative distribution functions. *Ann. Math. Statist.* **23** 563–574.

[2] GNEDENKO, B. V. and KOROLJUK, V. S. (1951). On the maximum discrepancy between two empirical distribution functions. *Dokl. Akad. Nauk SSSR.* **80** 525–528. (English translation: Selected Translations in Math. Statist. and Prob., IMS and AMS, **1** (1961) 13–16.)

[3] GNEDENKO, B. V. and MIHALEVIČ, V. S. (1952). On the distribution of the number of excesses of one empirical distribution function over another. *Dokl. Akad. Nauk*

SSSR. **82** 841–843. (English translation: Selected Translations in Math. Statist. and Prob., IMS and AMS, **1** (1961) 83–85.)

[4] GNEDENKO, B. V. and MIHALEVIČ, V. S. (1952). Two theorems on the behavior of empirical distribution functions. *Dokl. Akad. Nauk SSSR.* **85** 25–27. (English translation: Selected Translations in Math. Statist. and Prob., IMS and AMS, **1** (1961) 55–57.)

[5] KOROLJUK, V. S. (1955). On the discrepancy of empirical distribution functions for the case of two independent samples. *Izv. Akad. Nauk SSSR. Ser. Math.* **19** 81–96.

[6] TAKÁCS, L. (1962). Ballot problems. *Z. Wahrscheinlichkeitstheorie* **1** 154–158.

[7] TAKÁCS, L. (1963). The distribution of majority times in a ballot. *Z. Wahrscheinlichkeitstheorie* **2** 118–121.