Reprinted from the JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
March 1963, Vol. 58
pp. 13-30

PROBABILITY INEQUALITIES FOR SUMS OF BOUNDED RANDOM VARIABLES[1]

WASSILY HOEFFDING
University of North Carolina

Upper bounds are derived for the probability that the sum S of n independent random variables exceeds its mean ES by a positive number nt. It is assumed that the range of each summand of S is bounded or bounded above. The bounds for $\Pr\{S-ES\geq nt\}$ depend only on the endpoints of the ranges of the summands and the mean, or the mean and the variance of S. These results are then used to obtain analogous inequalities for certain sums of dependent random variables such as U statistics and the sum of a random sample without replacement from a finite population.

1. INTRODUCTION

LET $X_1, X_2, \cdots, X_n$ be independent random variables with finite first and second moments,

$$S = X_1 + \cdots + X_n, \qquad \overline{X} = S/n, \tag{1.1}$$

$$\mu = E\overline{X} = ES/n, \qquad \sigma^2 = n\operatorname{var}(\overline{X}) = (\operatorname{var} S)/n. \tag{1.2}$$

(Thus if the X_i have a common mean then its value is μ and if they have a common variance then its value is σ^2.) In section 2 upper bounds are given for the probability

$$\Pr\{\overline{X} - \mu \geq t\} = \Pr\{S - ES \geq nt\}, \tag{1.3}$$

where $t>0$, under the additional assumption that the range of each random variable X_i is bounded (or at least bounded from above). These upper bounds depend only on t, n, the endpoints of the ranges of the X_i, and on μ, or on μ and σ. We assume $t>0$ since for $t\leq 0$ no nontrivial upper bound exists under our assumptions. Note that an upper bound for Pr $\{\overline{X}-\mu\geq t\}$ implies in an obvious way an upper bound for Pr $\{-\overline{X}+\mu\geq t\}$ and hence also for

$$\Pr\{\,|\overline{X} - \mu|\, \geq t\} = \Pr\{\overline{X} - \mu \geq t\} + \Pr\{-\overline{X} + \mu \geq t\}. \tag{1.4}$$

Known upper bounds for these probabilities include the Bienaymé-Chebyshev inequality

$$\Pr\{\,|\overline{X} - \mu|\, \geq t\} \leq \frac{\sigma^2}{nt^2}, \tag{1.5}$$

Chebyshev's[2] inequality

$$\Pr\{\overline{X} - \mu \geq t\} \leq \frac{1}{1 + \dfrac{nt^2}{\sigma^2}} \tag{1.6}$$

[1] This research was supported by the Air Force Office of Scientific Research.

[2] Inequality (1.6) has been attributed to various authors. Chebyshev [14] seems to be the first to have announced an inequality which implies (1.6) as an illustration of a general class of inequalities.

(which do not require the assumption of bounded summands) and the inequalities of Bernstein and Prohorov (see formulas (2.13) and (2.14)). Surveys of inequalities of this type have been given by Godwin [6], Savage [13], and Bennett [2]. Bennett also derived new inequalities, in particular inequality (2.12), and made instructive comparisons between different bounds.

The method employed to derive the inequalities, which has often been used (apparently first by S. N. Bernstein), is based on the following simple observation. The probability Pr $\{S-ES \geq nt\}$ is the expected value of the function which takes the values 0 and 1 according as $S-ES-nt$ is <0 or ≥ 0. This function does not exceed exp $\{h(S-ES-nt)\}$, where h is an arbitrary positive constant. Hence

$$\Pr\{\overline{X} - \mu \geq t\} = \Pr\{S - ES \geq nt\} \leq Ee^{h(S-ES-nt)}. \tag{1.7}$$

If, as we here assume, the summands of S are independent, then

$$Ee^{h(S-ES-nt)} = e^{-hnt} \prod_{i=1}^{n} Ee^{h(X_i - EX_i)}. \tag{1.8}$$

It remains to obtain an upper bound for the expected value in (1.8) and to minimize this bound with respect to h. The bounds (2.1) and (2.8) of Theorems 1 and 3 are the best that can be obtained by this method under the assumptions of the theorems. They are not the best possible[3] bounds for the probability in (1.7). The bounds derived in this paper are better than the Chebyshev bounds (1.5) and (1.6) except for small values of t or small values of n. Typically, if t is held fixed, they tend to zero at an exponential rate as n increases.

The bounds of Theorems 1 and 3 are compared in section 3. The proofs of the theorems are given in section 4.

In section 5 the results of the preceding sections are used to obtain probability bounds for certain sums of dependent random variables such as U statistics and sums of m-dependent random variables. In section 6 a relation between samples with and without replacement from a finite population is established which implies probability bounds for the sum of a sample without replacement.

The following facts about convex functions will be used; for proofs see reference [7]. A continuous function $f(x)$ is convex in the interval I if and only if $f(px+(1-p)y) \leq pf(x)+(1-p)f(y)$ for $0<p<1$ and all x and y in I. If this is true for all real x and y, the function is simply called convex. A continuous function is convex in I if it has a nonnegative second derivative in I. If $f(x)$ is continuous and convex in I then for any positive numbers $p_1, \cdots, p_N$ such that $p_1 + \cdots + p_N = 1$ and any numbers $x_1, \cdots, x_N$ in I

$$f\left(\sum_{i=1}^{N} p_i x_i\right) \leq \sum_{i=1}^{N} p_i f(x_i). \tag{1.9}$$

This is known as Jensen's inequality.

[3] By the best possible bound for the probability in (1.7) is meant the least upper bound which depends only on t, n, the endpoints of the ranges of the X_i, and μ (or μ and σ). Approximations for the probability in (1.7) which involve the upper bound in (1.7) (minimized with respect to h) have been considered by several authors; see, in particular, Bahadur and Rao [1], where references to earlier work can be found. The present paper is concerned with exact bounds, not with approximations for the probability.