SOME LIMIT THEOREMS ASSOCIATED WITH MULTINOMIAL TRIALS

D. A. DARLING
UNIVERSITY OF MICHIGAN

1. Introduction

Let $X_1, X_2, \cdots$ be independent random variables, each having the same distribution $\Pr\{X_i = k\} = p_k$, $k = 1, 2, \cdots$. We assume without loss of generality that $p_i > 0$ and $p_1 \geq p_2 \geq p_3 \geq \cdots$.

Let $N_n(k)$ be the number of those X_j which equal k, $j = 1, 2, \cdots, n$. In this paper we are going to study certain limiting properties of the random variables

$$R_n = \sum_{N_n(k)>0} 1, \tag{1.1}$$

$$L_n = \sum_{N_n(k) \equiv 1 \pmod 2} 1. \tag{1.2}$$

Thus R_n is the number of distinct values assumed by the sequence

$$\{X_1, X_2, \cdots, X_n\}, \tag{1.3}$$

or the "range" of this sequence, while L_n is the number of values assumed an odd number of times. In principle, other random variables of the form $\sum_{k=1}^{\infty} \phi(N_n(k))$, where ϕ has a finite range, could be studied by the methods of this paper. But the important case of the "coverage" C_n,

$$C_n = \sum_{N_n(k)>0} p_k, \tag{1.4}$$

cannot apparently be so studied.

The random variable R_n is related to the "coupon collector's problem" (cf. Feller [1], p. 102) and has been studied in the case of finitely many equal $p_i > 0$ by Békéssy [2] among others. The random variable L_n is related to a random walk on a simple Abelian group, as described in section 3. It turns out that the studies of the random variables R_n and L_n are almost identical.

The main results of this paper are given in (2.9), (2.11), (3.10), (3.11), and (4.2).

2. The generating functions

As is well known and easily proved, if in the definition of $N_n(k)$ of section 1 we replace n by a random variable Λ which is independent of the $\{X_i\}$ and has a Poisson distribution with parameter λ, the random variables $N_\Lambda(k) = \Lambda_k$ are independent Poisson random variables, $k = 1, 2, \cdots$, Λ_k having a parameter λp_k,