CONTRIBUTION TO THE THEORY OF THE χ^2 TEST

J. NEYMAN

UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction

In the present paper several alternative definitions of the familiar symbol χ^2 are discussed. The body of the paper is divided into two parts. In the first part (sec. 3) a class of estimates is defined, termed best asymptotically normal estimates (BAN estimates, for short), all having the same asymptotic properties as the maximum likelihood estimates but varying in the ease with which they can be computed. In the second part (sec. 4) a class of tests is developed which are all equivalent in the limit to λ-tests. Both the computation of BAN estimates and the application of the statistical tests considered involve the minimization of the alternatively defined χ^2's.

Some of the results given below were announced in 1940 [**8**].*

2. General conditions

The problems considered refer to the following situation. Consider s sequences of independent trials and let n_i denote the number of trials in the ith sequence. Each trial of the ith sequence is capable of producing one of the ν_i mutually exclusive results, say

$$R_{i,1}, R_{i,2}, \cdots, R_{i,\nu_i}, \tag{1}$$

with probabilities

$$p_{i,1}, p_{i,2}, \cdots, p_{i,\nu_i}, \tag{2}$$

where

$$\sum_{j=1}^{\nu_i} p_{i,j} = 1 . \tag{3}$$

Denote by $n_{i,j}$ the number of occurrences of $R_{i,j}$ in the course of the n_i trials forming the ith sequence. Also let $q_{i,j} = n_{i,j}/n_i$. Finally let $N = n_1 + n_2 + \cdots + n_s$ and $Q_i = n_i/N$. The symbols $n_{i,j}$ and $q_{i,j}$ will be treated as random variables. The Q_i's will be considered as constants. N, the total number of observations, will be assumed to increase without limit.

The problems treated below arise when the values of the probabilities $p_{i,j}$ are unknown but it is given that each $p_{i,j}$ $(i = 1, 2, \cdots, s; j = 1, 2, \cdots, \nu_i)$ is a specified function of several parameters $\theta_1, \theta_2, \cdots, \theta_k$. The reasoning which follows does not depend very much on the value of k, provided $k \geqq 2$. In order

* **Boldface numbers in brackets refer to references at the end of the paper (p. 273).**