ON THE CHI-SQUARE DISTRIBUTION FOR SMALL SAMPLES[1]

By Paul G. Hoel

1. **Introduction.** The use of what is known as the χ^2 distribution function for testing goodness of fit involves two types of error. One arises from the fact that the derivation of this function is based upon rough approximations, while the other arises from using the integral of this continuous function in place of summing the proper terms of a discrete set. Both of these errors become increasingly important as the sample becomes small. The purpose of this paper is to investigate the nature of this first type of error by finding a better approximation than the customary one to what might be called the exact continuous χ^2 distribution function.

The method employed is that of generating or characteristic functions, and consists in expressing successively in expanded form the generating function of the multinomial, the distribution function of the multinomial, the generating function of χ^2, and the distribution function of χ^2. Only the first and second order terms of this final distribution function are evaluated explicitly because of the increasingly heavy algebra involved. By means of these second order terms, the nature of the error involved in the use of the customary first order approximation is investigated.

2. **The Generating Function of the Multinomial.** Consider $k + 1$ cells into which observations can fall, and let p_i be the probability that an observation will fall in cell i. If n observations are made, the probability that cell i will contain α_i of these observations is given by the multinomial

$$P = \frac{n!}{\alpha_1!\,\alpha_2!\cdots\alpha_{k+1}!}\,p_1^{\alpha_1}p_2^{\alpha_2}\cdots p_{k+1}^{\alpha_{k+1}},$$

where $\sum_{i=1}^{k+1}\alpha_i = n$. The generating function of this multinomial can be written as[2]

$$M = [p_1e^{t_1} + \cdots + p_ke^{t_k} + p]^n = \left[1 + \sum_{i=1}^{k} p_i(e^{t_i} - 1)\right]^n,$$

where α_{k+1} is chosen as the dependent variable and p_{k+1} is written as p.

---

[1] Presented to the American Mathematical Society, April 9, 1938.

[2] Cf. Darmois, *Statistique Mathematique*, pp. 237–242, for the methods used in this and the next two paragraphs.