Abstract:
The distribution of the range in samples from Pearson-type populations is considered in this thesis. Explicit density functions of the range, together with the cumulative distributions and certain moments, are given for four types. The difficulties precluding exact distributions have been pointed out in the other cases. An asymptotic distribution is suggested as a means of approximating the distribution of the range for large samples from particular populations.

# DISTRIBUTION OF THE SAMPLE RANGE FOR PARENT POPULATIONS ASSOCIATED WITH PEARSON'S DIFFERENTIAL EQUATION

by

GLENN R. INGRAM

A THESIS

Submitted to the Graduate Faculty

in

partial fulfillment of the requirements

for the degree of

Master of Science in Applied Mathematics

at

Montana State College

Approved:

John W. Hurst

Head, Major Department

B. Ostle

Chairman, Examining Committee

J. A. Nelson

Dean, Graduate Division

Bozeman, Montana
June, 1954

## ACKNOWLEDGMENT

The writer wishes to express his thanks to Dr. Bernard Ostle for suggesting this problem and for his help and encouragement in the completion of this thesis. Thanks are also due Dr. Robert Lowney for his helpful suggestions.

ABSTRACT

The distribution of the range in samples from Pearson-type populations is considered in this thesis. Explicit density functions of the range, together with the cumulative distributions and certain moments, are given for four types. The difficulties precluding exact distributions have been pointed out in the other cases. An asymptotic distribution is suggested as a means of approximating the distribution of the range for large samples from particular populations.

## I. INTRODUCTION AND STATEMENT OF THE PROBLEM

In a consideration of sample[1] statistics, one of the most obvious and easily obtained is the range; i.e., the difference between the largest and smallest sample values. Hence, if the distribution of this statistic could be obtained, it would be useful in situations where a minimum of calculation is expedient.

Formally, the density function of range is obtained as an integral from the joint distribution of largest and smallest sample values. However, evaluation of this integral for arbitrary sample size is difficult in most cases and impossible by exact methods in some.

Practical applications of the range suggest that further study of its properties would be worthwhile. Statistical quality control utilizes the sample range to a considerable extent because of the ease with which it is computed, in contrast with the more time consuming calculation of other measures of variation.

Most of the research concerning the sample range has been concerned with parent normal populations because of their wide application. One other population, the rectangular, has also been exhaustively studied.

A broad class of probability density functions, including the two mentioned above as special cases, is the Pearson System. This system is generated by specializing constants in the differential equation

$$\frac{dy}{dx} = \frac{(x-a)y}{b_0 + b_1 x + b_2 x^2} \qquad (1)$$

1. All samples referred to in this thesis will be random samples.

Many of the distributions important in sampling theory are special cases of different members of the family of solutions of the above differential equation. Among these distributions are: the normal, chi-square, Student's $t$, certain correlation coefficients, and $F$.

This thesis is concerned with the distribution of the range in samples of size $n$ from parent populations that are members of the Pearson system. Each of the twelve types will be considered, and the difficulties pointed out in cases where an explicit result has not been obtained.

## II. THE FORMAL SOLUTION FOR DISTRIBUTION OF THE RANGE

The distribution of range in integral form can be obtained at once from the joint distribution of the largest and smallest sample values. This frequency function is given by

(2) $$h(u,v) = n(n-1)\left[F(v) - F(u)\right]^{n-2} f(v)f(u), \qquad a \leq u < v \leq b,$$

where $f(x)$ is the probability density function specifying the population sampled, with $a \leq x \leq b$,

$\underline{u}$ is the smallest sample value

$\underline{v}$ is the largest sample value

$\underline{n}$ is the sample size

and $F(x) = \int_a^x f(x)dx$ is the cumulative distribution function.[2]

By the transformation $v = u + R$, where $\underline{R}$ is sample range, a joint function of $\underline{u}$ and $\underline{R}$ is obtained,

(3) $$h(u,R) = n(n-1)\left[F(u+R) - F(u)\right]^{n-2} f(u+R)f(u).$$

Then by integrating over the range of $\underline{u}$, the desired density function is obtained,

(4) $$g(R) = n(n-1)\int_a^{b-R} \left[F(u+R) - F(u)\right]^{n-2} f(u+R)f(u)du, \quad 0 \leq R \leq b-a.$$

Gumbel (4) has shown that the cumulative distribution takes an elegant, if not particularly useful, form when the upper limit is independent of R;

2. With appropriate changes of the argument, this notation will be followed throughout the thesis. That is, a lower case letter will denote a probability density function, and the corresponding upper case letter will denote the cumulative distribution.