ON THE EXPECTATIONS OF CERTAIN ORDER STATISTICS

ON THE EXPECTATIONS OF
CERTAIN ORDER STATISTICS

By

Edith M. Guntley

A Thesis

Submitted to the Faculty of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Science

McMaster University

September 1968

MASTER OF SCIENCE (1968)
(Mathematics)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: On the Expectations of certain Order Statistics

AUTHOR: Edith Mae Guntley, B.Sc. (Edinburgh University)

SUPERVISOR: Dr. S. K. Banerjee

NUMBER OF PAGES: iii, 57

SCOPE AND CONTENTS: This thesis deals with order statistics and the asymptotic distributions of certain functions of the first and second quartiles of a sample drawn at random from a bivariate population, whose distribution function is specified by its truncated bivariate Edgeworth Series.

ACKNOWLEDGMENTS

The author has pleasure in expressing sincere gratitude to her supervisor, Professor S. K. Banerjee for his advice and encouragement during the preparation of this thesis. Also gratitude is expressed to McMaster University for financial support.

INTRODUCTION

It is appropriate at the outset to define the fundamental elements of the study presented here, and attempt to remove all possible ambiguities which could arise from the now accepted name assigned to them. If the set of values $x_1, x_2, \ldots x_n$ occuring in a random sample of size n drawn from population having (known or unknown) cumlative probability function is ordered according to magnitude so that $x_{(1)}, x_{(2)} \cdot x_{(n)}$ is a permutation of $x_1, x_2, \ldots, x_n$ with $x_{(i)} \leq x_{(i+1)}$ $1 \leq i \leq n-1$, then the elements $x_{(i)}$, as well as functions of such variables, are known as order statistics, and, in particular, $x_{(r)}$, the value not exceeded by r members of the sample, is termed the r^{th} order statistic.

In non-parametric statistical inferences it is being found that order statistics are playing a significant role. The importance attached to work on non-parametric problems and order statistics is justified by recognition of the advantages to be obtained from the possible development of methods of statistical inferences which are applicable to broad classes of probability distribution functions, and the knowledge that considerable amount of new statistical information theory can be derived using order statistics, assuming no stronger conditions than that of continuity of the derivative of the cumulative distribution function. For the statistician interested in paractical applications it is advantageous to make statistical procedures simple and as broadly applicable as possible - which is the case of statistical inference theory based on order statistics.

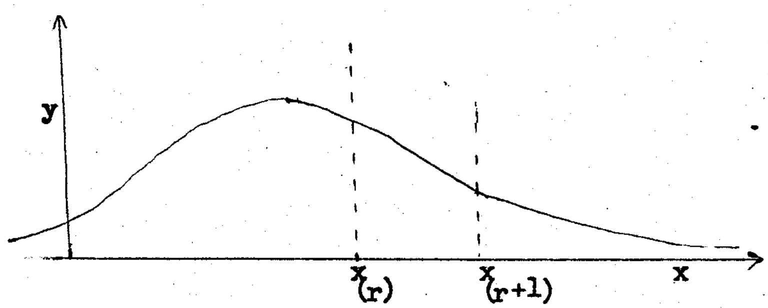
Among the earliest problems on the sampling theory of order statistics was that of finding the mean value of the difference between the r^{th} and $r+1^{th}$ order statistic in a sample of n values from a population having continuous probability density function. No other assumption was made about the probability distribution. This, the Galton's difference problem was studied in 1902 by Karl Pearson 11 using a deferred integration technique:

A random sample of n individuals is drawn from a population of N members, which when N is large may be taken to obey any law of frequency expressed by the curve

$$y = N\,\phi(x)$$

$y\delta x$ being the total frequency of individuals with the measured random variable taking a value somewhere in the interval between x and $x + \delta x$. We seek the expected value of $x_{(r+1)} - x_{(r)}$.

Consider the graph with ordinate y and corresponding abscissa x, the measured random variable.



Then the area between the curve and the x-axis, by the definition of y, is the number of individuals in the population so that

$$\int_{-\infty}^{\infty} y\,dx = N \qquad \int_{-\infty}^{\infty} \frac{y}{N}\,dx = \int_{-\infty}^{\infty} \phi(x)\,dx = 1.$$

The probability of an element at $x_{(r)}$ is $\frac{y}{N}(x_r)\,\delta x_r$ and at x_{r+1} is $\frac{y}{N}(x_{r+1})\,\delta x_{r+1}$. Also the probability of having an individual fall below $x_{(r)}$ is $\frac{A}{N}(x_{(r)})$ where A is the area beneath the y-curve, and to the left of the abscissa $x_{(r)}$.

$$\text{i.e. Prob } (x < x_{(r)}) = \frac{1}{N} \cdot \int_{-\infty}^{x_r} y\,dx = \int_{-\infty}^{x_r} \emptyset(x)\,dx = \frac{A}{N}(x_r)$$

$$\text{and } \frac{dA(x_r)}{dx_r} = N\,\emptyset\,(x_r) = y(x_r) \qquad \text{(i)}$$

Thus the joint probability of $x_{(r)}$ and $x_{(r+1)}$ is

$$f(x_{(r)}, x_{(r+1)}) = \frac{n!}{(r-1)!(n-r-1)!}\, \frac{y}{N}(x_{(r)})\, \frac{y}{N}(x_{(r+1)}) \left[\frac{A(x_{(r)})}{N}\right]^{r-1} \left[1-\frac{A(x_{(r+1)})}{N}\right]^{n-r-1} \delta x_{(r)}\,\delta x_{(r+1)}$$

and

$$\chi_r = E(x_{(r+1)} - x_{(r)}) = \int_{x_{(r+1)}=-\infty}^{\infty} dx_{(r+1)} \int_{x_{(r)}=-\infty}^{x_{(r+1)}} dx_{(r)}\; f(x_{(r)}, x_{(r+1)})\; (x_{(r+1)} - x_{(r)})$$

Integrating w.r.t. variable $x_{(r)}$. One needs consider only

$$I = \int_{x_{(r)}=-\infty}^{x_{(r+1)}} dx_{(r)}\; y(x_{(r)})\; A(x_{(r)})^{r-1} \cdot (x_{(r+1)} - x_{(r)})$$

and using integration by parts and (i)

$$= \left[(x_{(r+1)} - x_{(r)})\, \frac{A^r(x_{(r)})}{r}\right]_{x_{(r)}=-\infty}^{x_{(r+1)}} + \int_{x_{(r)}=-\infty}^{x_{(r+1)}} dx_{(r)}\, \frac{A^r(x_{(r)})}{r}$$

where the first bracket vanishes at both limits and we define a function U by

$$U = \int_{x=-\infty}^{x_{(r+1)}} dx\; A^r(x)$$

Then $\chi_r = k \int_{x_{(r+1)}=-\infty}^{\infty} (n-r)dx_{(r+1)} \left\{ \frac{y}{N}(x_{(r+1)}) \left[1 - \frac{A(x_{(r+1)})}{N}\right]^{n-r-1} \frac{U}{N^{r+1}} \right., \; k = \frac{n!}{(r!)\,(n-r)!}$

Employing (i) and integration by parts

$$X_r = \left[-k \; N \left(1 - \frac{A(x_{(r+1)})}{N}\right)^{n-r} \frac{U}{N^{4+1}} \right]_{x_{r+1}=-\infty}^{\infty}$$

$$+ k \int_{-\infty}^{\infty} \left[1 - \frac{A(x_{r+1})}{N}\right]^{n-r} \frac{1}{N^r} \frac{dU}{dx_{(r+1)}} \; dx_{(r+1)}$$

where the first bracket again vanishes at both limit points to give:

$$\chi_r = k \int_{-\infty}^{+\infty} 1 - \frac{A(x_{r+1})}{N}^{\;n-r} A^r(x_{(r+1)})dx_{(r+1)}$$

Thus if $F(x)$ = Prob. (random variable $< x$).

$$\chi_r = \frac{n!}{r!(n-r)!} \int_{-\infty}^{\infty} F^r(x) \left[1 - F(x)\right]^{n-r} dx.$$

N.B. : were the sample ordered from greatest occuring, to least, as K. Pearson did, the form of X_r would be the same as above except for an interchange of r and n-r .

Pearson's work was later extended by Tippett 13 who found the mean values of the sample range R, the difference between the least and greatest order statistic of sample.

This derivation involved summing χ_r for $1 \leq r \leq n-1$. Thus

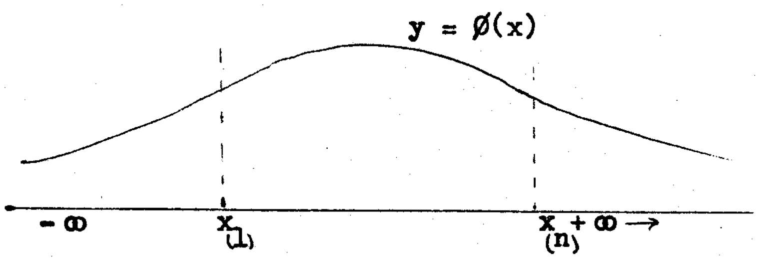
$$E(R) = \int_{-\infty}^{\infty} \sum_{r=1}^{n-1} \frac{n!}{r!(n-r)!} F^r(x) \left[1 - F(x)\right]^{n-r} dx$$

$$= \int_{-\infty}^{\infty} \sum_{r=0}^{n} \frac{n!}{r!(n-r)!} F^r(x) \left[1-F(x)\right]^{n-r} dx$$

$$- \int_{-\infty}^{\infty} \left[1 - F(x)\right]^n dx - \int_{-\infty}^{\infty} F^n(x)\, dx$$

$$= \int_{-\infty}^{\infty} \left\{1 - F^n(x) - (1 - F(x))^n\right\} dx.$$

But Tippett demonstrated an alternative technique which allowed the extension to obtain the moments of the range's distribution. The alternative method employed the definition of a function $\theta = \frac{1-\alpha^{n-s-1}}{n-s-1}$ so that $\frac{d\theta}{d\alpha} = \alpha^{n-s-2}$.



Let $x_{(1)}$ be the first order statistic and $x_{(n)}$ the last in a sample of size n from a population with curve of distribution $y = \phi(x)$. Then, supposing the population to be infinite, the chance of getting one individual at $x_{(1)}$, one at $x_{(n)}$ and n-2 between $x_{(1)}$ and $x_{(n)}$ is:

$$\frac{(-1)^n n!}{(n-2)!} (\alpha_1 - \alpha_n)^{n-2} d\alpha_1\, d\alpha_n$$

where $\alpha_1 = \int_{-\infty}^{x_{(1)}} \phi(x)\, dx$

and $\int_{-\infty}^{\infty} \phi(x)\, dx = 1.$

$$\text{Then } E(R) = \int_{x_{(n)}=-\infty}^{\infty} d\alpha_n \int_{x_{(1)}=-\infty}^{x_{(n)}} \frac{(-1)^n n!}{(n-2)!} (\alpha_1 - \alpha_n)^{n-2} \, d\alpha_1 \, (x_{(n)} - x_{(1)})$$

$$= \int_{x_{(n)}=-\infty}^{\infty} d\alpha_n \int_{x_1=-\infty}^{x_{(n)}} \frac{n!}{(n-2)!} \sum_{s=0}^{n-2} \frac{(n-2)!}{s!(n-s-2)!} (-1)^{n-s} \alpha_1^s \, \alpha_n^{n-s-2} \, d\alpha_1 (x_{(n)} - x_{(1)})$$

$$= \sum_{s=0}^{n-2} \int_{x_n=-\infty}^{\infty} d\alpha_n \frac{n!}{s!(n-s-2)!} (-1)^{n-s} \alpha_n^{n-s-2} \int_{-\infty}^{x_n} \alpha_1^s \, d\alpha_1 \, (x_{(n)} - x_{(1)}).$$

and using integration by parts

$$\sum_{s=0}^{n-2} \int_{x_n=-\infty}^{\infty} \frac{(-1)^s \; n!}{(s+1)!(n-s-2)!} \; \alpha_n^{n-s-2} \, U \, d\alpha_n$$

$$\text{where } U = \int_{-\infty}^{x_n} \alpha_1^{s+1} \, dx_1 = \int_{-\infty}^{x_n} \alpha(x)^{s+1} \, dx$$

$$\text{and defining } \theta = \frac{1-\alpha_n^{n-s-1}}{n-s-1} \quad \text{so that } \frac{d\theta}{d\alpha_n} = -\alpha_n^{n-s-2}.$$

$$E(R) = \sum_{s=0}^{n-2} \int_{x_n=-\infty}^{\infty} (-1)^s \frac{n!}{(s+1)!(n-s-2)!} \, U \, d\theta.$$

$$= \sum_{s=0}^{n-2} \frac{(-1)^s \; n!}{(s+1)! \; (n-s-2)!} \left\{ \Big[- U \, \theta \Big]_{x_n=-\infty}^{\infty} + \int \frac{dU}{d\alpha_n} \cdot \theta \cdot d\alpha_n \right\}.$$

and U vanishes at the lower limit $x = -\infty$ while θ vanishes at the upper limit $x = \infty$, so that $\Big[-U\theta\Big]_{x_n=-\infty}^{\infty} = 0.$

$$\text{Thus } E(R) = \sum_{s=0}^{n-2} \frac{(-1)^{n-s} \; n!}{(s+1)!(n-s-2)!} \int_{-\infty}^{\infty} \theta \, \alpha_n^{s+1} \, dx_n.$$

and using integration by parts

$$= \sum_{s=0}^{n-2} \frac{(-1)^s \; n!}{(s+1)!(n-s-1)!} \int_{-\infty}^{\infty} \alpha_n^{s+1} (1-\alpha_n^{n-s-1}) \, d\alpha_n.$$

and splitting the summation into two parts

$$E(R) = \int_{-\infty}^{\infty} 1 - (1-\alpha)^n - \alpha^n \; dx. \qquad (2)$$

agreeing with the earlier solution.

This method of solution also yields expressions for other moments of R, and is therefore more general in application.

In the same paper Tippett tabulated the mean range for a standardized, normal distribution for samples from two to one thousand, these being evaluated by finding a framework of values by direct computation of equation (2) using quadrature and filling this in by interpolation, using first Lagrangian Formulae and finally a difference formula. In addition, using the functional relation

$$\int_{-\infty}^{x_p} f(x)\, dx = \alpha_p^n$$

where f(x) is the distribution of the largest individual in samples of n (where $\int_{-\infty}^{\infty} f(x)\, dx = 1$) and $\alpha_p^* = \int_{-\infty}^{x_p} \phi(x)\, dx$, $y = \phi(x)$ being the graphical representation of an infinite population's distribution, he tabulated the probability integral or cumulative distribution function of the largest order statistic in a sample from a normal population having zero mean and unit variance.

Later R. A. Fisher and L. H. C. Tippett 3 determined by a method of functional equations, and for specified regularity conditions on the population distribution, the asymptotic distribution of the greatest (and also the least) values in a sample as the sample size tended to infinity.

It appears that a particular set of distributions provides the limiting distribution in all cases and the case derived for the normal curve is peculiar for the extreme slowness with which the limiting form is derived.

The possible limiting forms are deduced from the functional relations they must satisfy:

$$P^n(x) = P(a_n x + b_n).$$

The solutions of this functional equation will give all possible limiting forms; and consequently these fall into 3 classes,

i) $a = 1$ $\quad P^n(x) = P(x+b_n)$

ii) $P = 0$ when $x = 0$ $\quad P^n(x) = P(a_n x)$

iii) $P = 1$ when $x = 0$ $\quad P^n(x) = P(a_n x)$,

which show that the only possible limiting curves are such that

i') $dP = e^{-x-e^{-x}} dx.$

ii') $dP = \frac{k}{x^{k+1}} e^{-x^{-k}} dx.$

iii') $dP = k(-x)^{k-1} e^{-(-x)^k} dx.$

Further studies of the limiting distributions was made by Grumbel 5 , who made several applications to such problems as flood flows, where the random variable often is the annual rainfall, and the sample size n is the number of years for which the records of the annual rainfall are available; and papers on order statistics continued to appear. In 1932 A. T. Craig 2 gave general expressions for the exact distribution functions of the median, quartiles and range of a sample of size n.

Suppose a variable x to obey a law of probability given by f(x) which, initially is assumed to vanish outside of the interval from 0 to some positive real number A; and consider a sample consisting of $n = 2m + 1$ (m, an integer) values of x with median ξ be drawn.

The probability that m of the 2m + 1 items be in the interval from 0 to ξ is $\frac{(2m+1)!}{m!(m+1)!} \left[\int_0^{\xi} f(t)\, dt\right]^m$.

The probability that of the remaining elements m lie in the interval from 0 to A and one lies in $[\xi, \xi + d\xi]$ is $(m+1)\left[\int_{\xi}^{a} f(t)\, dt\right]^m \cdot f(\xi)\, d\xi$.

Thus the probability distribution $\phi(\xi)$ of the median in samples of size $n = 2m+1$ is given by the equation

$$\phi(\xi) = \frac{(2m+1)!}{(m!)^2} \left[\int_0^{\xi} f(t)\, dt\right]^m \left[\int_{\xi}^{A} f(t)\, dt\right]^m f(\xi)\, d\xi,$$

and $\phi(\xi)$ has same form when the range of x is the entire real line.

Similarly it may be shown that the probability function of the lower quartile $\bar{x}_1$ of samples of $n = 4m+1$ elements, drawn from a universe represented by f(x) is

$$\phi(\bar{x}_1) = \frac{(4m+1)!}{m! \cdot (3m)!} \left[\int_0^{\bar{x}_1} f(t)\, dt\right]^m \left[\int_{\bar{x}_1}^{\infty} f(t)\, dt\right]^{3m} f(\bar{x}_1)\, d\bar{x}_1$$

and abviously any statistic which is defined as the value of the variate which exceeds and is exceeded by specified numbers of elements in the sample may have its distribution determined in like manner. Still studying the median, Thomson [12] in 1936 showed how confidence limits for the median (and also for other quantiles) of a population having a continuous cumulative distribution function could be established from order statistics in a sample from such a population.

In recent times the probability behaviour of order statistics has been significantly developed and unified by S. S. Wilks, his associates, and students at Princeton 14 , 15 and the posthumous publication of collected papers 17 by him provides good evidence of his involvement with the study of order statistics, and their applications. The accumulation of theoretical knowledge of order statistics had stimulated the development of areas of their application, in particular their application to non-parametric statistical inference. [Inferences from samples about distribution functions, under normal assumptions - e.g. continuity of the cumulative distribution function - are referred to as non-parametric inferences, in contrast to parametric inferences which are concerned with inferences about values of parameters of distribution functions of known functional form, depending on one or more unknown parameters]. The probability theory underlying such inference consists essentially of the probability theory of certain functions of order statistics. Wilks 16 gives a survey of some of the basic ideas and results of non-parametric statistical inference. By their usefulness in this field one is prompted to ask if order statistics may not be used in the estimation of parameters.

Order statistics often permit very simple inefficient solutions of some of the more important parametric problems of statistical estimation. R. A. Fisher introduced the concept of efficient statistics, or estimates of efficiency. They serve as a measure of the information a statistic draws from a sample so that if statistics $\hat{\theta}'$ and $\hat{\theta}''$, unbiased estimates of a population parameter θ, with variance $\hat{\theta}'$ less than variance of $\hat{\theta}''$, then the efficiency of $\hat{\theta}''$ relative to $\hat{\theta}'$ is the ratio of the smaller variance to the larger; and if there exists an unbiased estimate $\hat{\theta}_o$ for which the variance is minimum, then the latter is called the most efficient unbiased estimate and "the" efficiency of all other estimations may be taken as their efficiency, relative to $\hat{\theta}_o$. Mosteller 8 has investigated the efficiency of various linear combinations of several order statistics in large samples for estimating the mean and variance of a normal distribution function and he obtained efficiencies as great as 0.87 by using the average of 10 properly spaced order statistics to estimate the mean.

As an attempt to achieve further usage of order statistics in parameter estimation one may consider their application in estimating parameters of multivariate populations. In particular, can order statistics be used to estimate the correlation coefficient p of some bivariate population? S. K. Banerjee 1 derived the asymptotic approximations to the joint distribution of certain sample quartiles, which I shall use in this study to observe the efficiency of certain functions of these order statistics when used to estimate p; for bivariate populations whose distributions satisfy certain specified conditions.

Banerjee's derivation of the asymptotic distribution considers two variates x_1 and x_2 with probability density function $f(x_1,x_2)$ which satisfies the following conditions,

(i) $$\int_{-\infty}^{\infty} f(x_1, 1/N)\, dx_1 = \int_{-\infty}^{\infty} f(x_1, 0)\, dx_1 + O(1/N)$$

(ii) $$\int_{-\infty}^{\infty} f(1/N), x_2)\, dx_2 = \int_{-\infty}^{\infty} f(0, x_2)\, dx_2 + O(1/N)$$

(iii) The following equations:

(a) $$\int_{-\infty}^{\xi_1} \int_{-\infty}^{\infty} f(x_1,x_2)\, dx_2\, dx_1 = \frac{1}{4}$$

(b) $$\int_{-\infty}^{\xi_2} \int_{-\infty}^{\infty} f(x_1,x_2)\, dx_1\, dx_2 = \frac{1}{4}$$

have unique real roots.

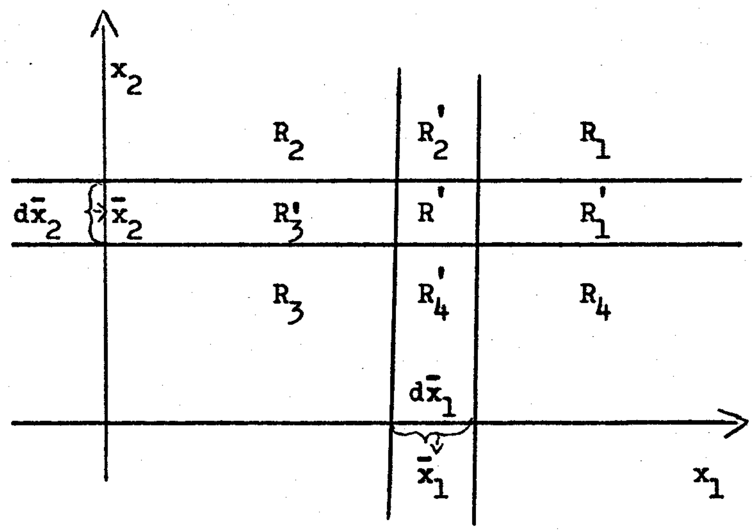
[In particular $f(x_1,x_2)$ may be the bivariate normal density function.] Let a sample of $(4n + 1)$ elements (x_{1r},x_{2r}) $(r = 1,2,\ldots,4n+1)$ be drawn from such a population. Let $\bar{x}_1$, $\bar{x}_2$ designate the first quartiles (corresponding to ξ_1, ξ_2 in population) of the two variables. $(\bar{x}_1, \bar{x}_2)$ will be referred to as the Quartile. Let us divide the plane into 9 zones R_1, R_2, R_3, R_4, R_1', ..., R_4', R'' by the straight lines:

$$x_2 = \bar{x}_2 - \frac{1}{2}\, d\bar{x}_2$$

$$x_2 = \bar{x}_2 + \frac{1}{2}\, dx_2$$

$$x_1 = \bar{x}_1 - \frac{1}{2}\, d\bar{x}_1$$

$$x_1 = \bar{x}_1 + \frac{1}{2}\, d\bar{x}_1$$



Let the probability that an element falls in the region $R_i^{(j)}$ be

$$p_i^j = \iint_{R_i^j} f(x_1, x_2)\, dx_1\, dx_2$$

For example $\quad p_1 = \int_{\bar{x}_1}^{\infty} \int_{\bar{x}_2}^{\infty} f(x_1, x_2)\, dx_2\, dx_1$

$$p'' = f(x_1, x_2)\, d\bar{x}_1\, d\bar{x}_2 \quad \text{etc.}$$

We shall consider now that the sample is drawn from a multinomial population with probabilities $p_1, \ldots, p''$ and pick out those terms which give rise to a sample quartile $(\bar{x}_1, \bar{x}_2)$. This can be done in the following five manners:-

(i) If the quartile is an element of the sample, then that element may fall in R'' and the other elements must fall in regions R_1, R_2, R_3 and R_4 with frequencies n_1, n_2, n_3, n_4 with the conditions

$$\left.\begin{aligned} n_1 + n_2 &= 3n \\ n_1 + n_4 &= 3_n \end{aligned}\right\} \text{ and } \left.\begin{aligned} n_2 + n_3 &= n \\ n_3 + n_4 &= n \end{aligned}\right\}$$

We have therefore $n_4 = n_2$ and $n_1 = 2n_2 + 3n_3$.

The probability that this occurs is:

$$S_1 = \sum \frac{(4n+1)!}{(n_1)!(n_2)!(n_3)!(n_4)!} \; p'' \cdot p_1^{n_1} \cdot p_2^{n_2} \cdot p_3^{n_3} \cdot p_4^{n_4}$$

$$= \sum_{n_2+n_3=n} \frac{(4n+1)!}{(3n_3+2n_2)!(n_2!)^2(n_3!)} \; p'' \cdot p_1^{3n_3+2n_2} \cdot p_2^{n_2} \cdot p_3^{n_3} \cdot p_4^{n_4}$$

$$= \sum_{n_1+n_2=n} \frac{(4n+1)!}{(3n_1+2n_2)!(n_2!)^2(n_1!)} \; p'' \cdot p_1^{3n_1+2n_2} \cdot p_2^{n_2} \cdot p_3^{n_1} \cdot p_4^{n_2} \cdot$$

(ii) Now let us suppose that the quartile is determined by two different elements of the sample, for example, one in R_1' and one in R_2' and n_i elements in R_i $(i = 1, 2, 3, 4)$ with:

$$\left.\begin{aligned} n_3 + n_4 &= n \\ n_2 + n_3 &= n \end{aligned}\right\} \text{ and } \left.\begin{aligned} n_1 + n_2 + 1 &= 3n \\ n_1 + n_4 + 1 &= 3n \end{aligned}\right\}$$

Therefore

$$n_2 = n_4 \quad \text{and} \quad n_1 = 2n_2 + 3n_3 - 1$$

The probability in this case is:

$$S_2 = \sum_{n_1+\cdots+n_4=4n-1} \frac{(4n+1)!}{(n_1)!(n_2!)(n_3!)(n_4!)} \cdot$$

$$\cdot\, p_1' \cdot p_2' \cdot p_1^{n_1} \cdot p_2^{n_2} \cdot p_3^{n_3} \cdot p_4^{n_4}$$

$$= \sum_{n_2+n_3=n} \frac{(4n+1)!}{(2n_2+3n_3-1)!(n_2!)^2(n_3!)} \cdot$$

$$\cdot\, p_1' \cdot p_2' \cdot p_1^{2n_2+3n_3-1} \cdot p_2^{n_3} \cdot p_4^{n_2}$$

$$= p_1' \cdot p_2' \cdot \sum_{n_1+n_2=n-1} \frac{(4n+1)!}{(n_1+1)!(3n_1+2n_2+2)!(n_2!)^2} \cdot$$

$$\cdot\, p_1^{3n_1+2n_2+2} \cdot p_2^{n_2} \cdot p_3^{n_1+1} \cdot p_4^{n_2}$$

(iii) Similarly considering R_2' and R_3':

$$\left.\begin{aligned} n_1 + n_2 + 1 &= 3_n \\ n_1 + n_4 &= 3n \end{aligned}\right\} \text{ and } \left.\begin{aligned} n_2 + n_3 + 1 &= n \\ n_3 + n_4 &= n \end{aligned}\right\} \text{; therefore}$$

$$n_4 = n_2+1 \quad \text{and} \quad n_1 = 2n + n_3 = 2n_2 + 3n_3 + 2$$

$$\text{Prob: } S_3 = \sum_{n_1+\cdots+n_4=4n-1} \frac{(4n+1)!}{n_1!\, n_2!\, n_3!\, n_4!} \cdot$$

$$\cdot\, p_2' \circ p_3' \cdot p_1^{n_1} \cdot p_2^{n_2} \cdot p_3^{n_3} \cdot p_4^{n_4}$$

$$= p_2' \cdot p_3' \cdot \sum_{n_2+n_3=n-1} \frac{(4n+1)!}{(2n_2+3n_3+2)!n_2!n_3!(n_2+1)!} \cdot$$

$$\cdot\, p_1^{2n_2+3n_3+2} \cdot p_2^{n_2} \cdot p_3^{n_3} \cdot p_4^{n_2+1}$$

$$= p_2' \cdot p_3' \cdot \sum_{n_1+n_2=n-1} \frac{(4n+1)!}{(3n_1+2n_2+2)!n_2!n_1!(n_2+1)!} \cdot$$

$$\cdot\ p_1^{3n_1+2n_2+2} \cdot p_2^{n_2} \cdot p_3^{n_1} \cdot p_4^{n_2+1}$$

(iv) Considering R_3' and R_4'

$$n_1 + n_2 = 3n = n_1 + n_4 , \quad \text{and}$$

$$n_2 + n_3 + 1 = n = n_4 + n_3 + 1$$

$$\text{Prob.} = S_4 = \sum_{n_1+\cdots+n_4=4n-1} \frac{(4n+1)!}{n_1!\ n_2!\ n_3!\ n_4!} \cdot$$

$$\cdot\ p_3' \cdot p_4' \cdot p_1^{n_1} \cdot p_2^{n_2} \cdot p_3^{n_3} \cdot p_4^{n_4}$$

$$= p_3' \cdot p_4' \cdot \sum_{n_2+n_3=n-1} \frac{(4n+1)!}{(2n_2+3n_3+3)!(n_2!)^2(n_3!)}$$

$$\cdot p_1^{2n_2+3n_3+3} \cdot p_2^{n_2} \cdot p_3^{n_3} \cdot p_4^{n_2}$$

$$= p_3' \cdot p_4' \cdot \sum_{n_1+n_2=n-1} \frac{(4n+1)!}{(3n_1+2n_2+3)!(n_2!)^2(n_1!)}$$

$$\cdot p_1^{3n_1+2n_2+3} \cdot p_2^{n_2} \cdot p_3^{n_1} \cdot p_4^{n_2}$$

(v) Lastly, considering R_1', R_4'

$$\left.\begin{array}{l} n_2 + n_3 = n \\ n_1 + n_2 = 3n \end{array}\right\} \text{ and } \left.\begin{array}{l} n_3 + n_4 + 1 = n \\ n_1 + n_4 + 1 = 3n \end{array}\right\} \text{ and therefore}$$

$$n_2 = n_4 + 1 \quad \text{and} \quad n_1 = 2n + n_3 = 3n_3 + 2n_4 + 2$$

$$\text{Prob.} = S_5 = \sum_{n_1+\cdots+n_4=4n-1} \frac{(4n+1)!}{(n_1!\ n_2!\ n_3!\ n_4!)} \cdot p_1' \cdot p_4' \cdot p_1^{n_1} \cdot p_2^{n_2} \cdot p_3^{n_3} \cdot p_4^{n_4}$$

$$= p_1' \cdot p_4' \cdot \sum_{n_3+n_4=n-1} \frac{(4n+1)!}{(3n_3+2n_4+2)!(n_4+1)!n_3!n_4!} \cdot p_1^{3n_3+2n_4+2} \cdot p_2^{n_4+1} \cdot p_3^{n_3} \cdot p_4^{n_4}$$

$$= p_1' \cdot p_4' \cdot \sum_{n_1+n_2=n-1} \frac{(4n+1)!}{(3n_1+2n_2+2)!(n_2+1)!n_1!n_2!} \cdot p_1^{3n_1+2n_2+2} \cdot p_2^{n_2+1} \cdot p_3^{n_1} \cdot p_4^{n_2} .$$

Therefore the elemental probability corresponding to $\bar{x}_1$ and $\bar{x}_2$ will be

$$D(\bar{x}_1, \bar{x}_2)\ d\bar{x}_1\ d\bar{x}_2 = S_1 + S_2 + S_3 + S_4 + S_5 .$$

Asymptotic Distribution. In order to get an approximation to the distribution for large n, we shall assume:

(a) (i) If $A = B[1 + O(1/n^{1/2})]$, we shall write $A = . B$ where $O(1/n^{1/2})$ represents any function such that

$$\lim_{N \to \infty} . N.O(1/N) = L < \infty$$

(ii) We know the following result (Multinomial distribution tends to Normal in the limit):

$$\frac{m!}{\prod_{i=1}^{r} m_i!} \prod_{i=1}^{r} p_i^{m_i} = \cdot (|A|/2\pi)^{1/2(r-1)} \cdot \exp(-1/2 \sum_1^{r-1} \sum_1^{r-1} A_{ij} Z_i Z_j) \prod_{i=1}^{r-i} dZ_i$$

where $Z_i = (m_i - mp_i)/m^{1/2}$; $i = 1, 2, \ldots, r-1$

$A_{ii} = 1/p_i + 1/p_r$; $A_{ij} = A_{ji} = 1/p_r$

(b).(i) We see that S_1 has one more factor in the denominator than the corresponding fractions in other sums. This may therefore be neglected in the asymptotic form as it is of order $(1/n)$ in comparison with others.

(ii) Therefore we have:

$$S_2 = 4n(4n+1) \cdot p_1' \cdot p_2' \cdot \sum_{n_1+n_2=n-1} \frac{(4n-1)!}{(3n_1+2n_2+2)!(n_2!)^2(n_1+1)!}$$

$$\cdot p_1^{3n_1+2n_2+2} \cdot p_2^{n_2} \cdot p_3^{n_1+1} \cdot p_4^{n_2} \cdot$$

$$= \cdot 4n(4n+1) \cdot p_1' \circ p_2' \cdot \sum_{n_1+n_2=n-1} \left(\frac{|A|}{(2\pi)^3}\right)^{1/2} \cdot \exp-1/2(\sum_1^3 \sum_1^3 A_{ij} Z_i Z_j).$$

$$\cdot dZ_1 \cdot dZ_2 \cdot dZ_3$$

where $Z_1 = \dfrac{(n_1+1) - (4n-1)p_3}{(4n-1)^{1/2}}$

$$Z_2 = \frac{(n_2) - (4n-1)p_2}{(4n-1)^{1/2}}$$

$$Z_3 = \frac{n_2 - (4n-1)p_4}{(4n-1)^{1/2}}$$

Now $Z_1 + Z_2 = \dfrac{n - (4n-1)(p_2+p_3)}{(4n-1)^{1/2}} =. (4n)^{1/2}(\frac{1}{4} - p_2 - p_3) = U_1$ (say)

Similarly

$$Z_1 + Z_3 = \frac{n - (4n-1)(p_4+p_3)}{(4n-1)^{1/2}} = . (4n)^{1/2}(\frac{1}{4} - p_4 - p_3) = U_2 \quad \text{(say)}$$

$A_{11} = 1/p_3 + 1/p_1$; $A_{22} = 1/p_2 + 1/p_1$; $A_{33} = 1/p_4 + 1/p_1$

$A_{12} = A_{21} = A_{13} = A_{31} = A_{23} = A_{32} = 1/p_1$

$/A/ = (p_1 + p_2 + p_3 + p_4)/p_1\, p_2\, p_3\, p_4$

$$\therefore \quad S_2 = . 4n . p_1' \circ p_2' . \frac{/A/^{1/2}}{(2\pi)^{3/2}} \circ \sum_{n_1+n_2=n-1} \exp. -1/2 \Big\{(1/p_1 +$$

$$+ 1/p_2 + 1/p_3 + 1/p_4)Z_1^2 - 2(U_1/p_1 + U_2/p_2 + U_1/p_2) +$$

$$+ U_2 /p_4)Z_1 + (U_1+U_2)^2/p_1 + U_1^2/p_2 + U_2^2/p_4\Big\} . dZ_1$$

(Since in the approximate relation, Multinomial Distribution - Multivariate Normal Law, the factors dZ_i correspond to factors $1/m^{1/2}$ and we therefore let dZ_2 and dZ_3 cancel the factor 4n-1 in the coefficient of exponential terms in S_2).

The summation can now be performed to within terms of order $1/m^{1/2}$ by integration with respect to Z_1 between, $-\infty$ to ∞, which gives

$$S_2 =. 4n\, p_1' \cdot p_2' \cdot \frac{/A/^{1/2}}{2\pi(1/p_1+1/p_2+1/p_3+1/p_4)^{1/2}} \cdot \exp.-1/2 \left\{ (U_1+U_2)^2/p_1 \right.$$

$$\left. + U_1^2/p_2 + U_2^2/p_4 - \frac{(U_1/p_1 + U_2/p_1 + U_1/p_2 + U_2/p_4)^2}{1/p_1 + 1/p_2 + 1/p_3 + 1/p_4} \right\}.$$

Now let us define:

$$q_1 = \int_{\xi_1}^{\infty} \int_{\xi_2}^{\infty} f(x_1,x_2)dx_1dx_2 \qquad q_1' = \int_{\xi_1}^{\infty} f(x_1,0)dx_1$$

$$q_2 = \int_{-\infty}^{\xi_1} \int_{\xi_2}^{\infty} f(x_1,x_2)dx_1dx_2 \qquad q_2' = \int_{\xi_2}^{\infty} f(0,x_2)dx_2$$

$$q_3 = \int_{-\infty}^{\xi_1} \int_{-\infty}^{\xi_2} f(x_1,x_2)dx_1dx_2 \qquad q_3' = \int_{-\infty}^{\xi_1} f(x_1,0)dx_1$$

$$q_4 = \int_{\xi_1}^{\infty} \int_{-\infty}^{\xi_2} f(x_1,x_2)dx_1dx_2 \quad \text{and} \quad q_4' = \int_{-\infty}^{\xi_2} f(0,x_2)dx_2$$

We have $p_i =. q_i \qquad (i = 1, 2, 3, 4)$

$p_i' =. q_i'\, d\bar{x}_2 \; (i = 1, 3)$

$p_i' =. q_i'\, d\bar{x}_1 \; (i = 2, 4)$

Now let $U_2 = (4n)^{1/2} (\frac{1}{4} - p_3 - p_4)$

where

$$p_3 + p_4 = \int_{-\infty}^{\infty} \int_{-\infty}^{\overline{x_2}} f(x_1,x_2)dx_2dx_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{0} f(x_1,x_2)\, dx_2dx_1 +$$

$$\int_{-\infty}^{\infty} \int_{0}^{\overline{x_2}} f(x_1,x_2)dx_2dx_1$$

$$= \frac{1}{2} + \overline{x_2} \int_{-\infty}^{\infty} f(x_1,\theta x_2)d\bar{x}_1 \qquad 0 < \theta < 1$$

$$\doteq \frac{1}{2} + a_2\bar{x}_2 \quad (\text{say, where } a_2 = \int_{-\infty}^{\infty} f(x_1,\theta x_2)dx_1$$

$$\doteq \int_{-\infty}^{\infty} f(x_1,0)dx_1$$

$$= q_1' + q_3')$$

$$\therefore \quad U_2 = -(4n)^{1/2} \left(\frac{1}{4} + a_2\, \bar{x}_2\right)$$

Similarly let

$$U_1 \doteq -(4n)^{1/2} \left(\frac{1}{4} + a_1\bar{x}_1\right), \text{ where } a_1 = \int_{-\infty}^{\infty} f(0,x_2)dx_2$$

$$= q_2' + q_4'$$

Therefore

$$S_2 \doteq \frac{4n\, q_1'q_2'}{2\pi} \frac{(q_1 + q_2 + q_3 + q_4)^{1/2}}{(q_1q_2q_3q_4)^{1/2}} (1/q_1 + 1/q_2 + 1/q_3 + 1/q_4)^{-1/2}$$

$$\exp. -1/2 \left\{ (U_1 + U_2)^2/q_1 + (U_1^2 + U_2^2)/q_2 - \frac{\left(\frac{U_1+U_2}{q_1} + \frac{U_1+U_2}{q_2}\right)^2}{(1/q_1 + 1/q_2 + 1/q_3 + 1/q_4)} \right\} d\bar{x}_1 d\bar{x}_2$$

or

$$S_2 =. \frac{q_1' \cdot q_2'}{2\pi a_1 a_2} \circ (q_1+q_2+q_3+q_4)^{\frac{1}{2}} / (q_1q_2q_3+q_1q_2q_4+q_1q_3q_4+q_2q_3q_4)^{\frac{1}{2}} \cdot$$

$$\exp\circ - \left\{ \frac{(U_1^2 - 2\beta U_1 U_2 + U_2^2}{2\sigma^2} \right\} = q_1' \cdot q_2' \cdot M \quad \text{(say)}$$

Where $1/\sigma^2 = (1/q_1+1/q_2)(1/q_3+1/q_4)/(1/q_1+1/q_2+1/q_3+1/q_4)$

and $-\beta/\sigma^2 = 1/q_1 - (1/q_1+1/q_2)^2/(1/a_1+1/q_2+1/q_3+1/q_4)$

(iii) $$S_3 = p_2' \cdot p_3' \sum_{n_1+n_2=n-1} \frac{(4n+1)!}{(3n_1+2n_2+2)!n_2!n_1!(n_2+1)!} \cdot$$

$$\cdot p_1^{3n_1+2n_2+2} \cdot p_2^{n_2} \cdot p_3^{n_1} \cdot p_4^{n_1+1}$$

$$=. 4n(4n+1)\ p_2' \cdot p_3' \cdot \sum \frac{/A/^{\frac{1}{2}}}{(2\pi)^{3/2}} \exp. -\tfrac{1}{2}\left(\sum_1 \sum_1 A_{ij} Z_i Z_j \right) dZ_1 dZ_2 dZ_3$$

where $$Z_1 = \frac{n_1 - (4n-1)p_3}{(4n-1)^{1/2}}$$

$$Z_2 = \frac{n_2 - (4n-1)p_2}{(4n-1)^{1/2}}$$

$$Z_3 = \frac{(n_2+1) - (4n-1)p_4}{(4n-1)^{1/2}}$$

and A_{ij}'s are exactly as in S_2.

Here also $Z_1 + Z_2 =. (4n)^{\frac{1}{2}} (\frac{1}{4} - p_2 - p_3) = U_1$ (say)

$Z_1 + Z_3 =. (4n)^{\frac{1}{2}} (\frac{1}{4} - p_4 - p_3) = U_2$ (say)

$\therefore$ S_3 is exactly equal to S_2 except that p_1' p_2' are replace by p_2' p_3'

$\therefore$ $S_3 \doteq q_2' \cdot q_3' \cdot M$

(iv) $$S_4 = p_3' p_4' \sum_{n_1+n_2=n-1} \frac{(4n+1)!}{(3n_1+2n_2+3)!(n_2!)^2(n_1!)} \cdot$$

$$\cdot\, p_1^{3n_1+2n_2+3} \cdot p_2^{n_2} \cdot p_3^{n_1} \cdot p_4^{n_2} \cdot$$

$$\doteq 4n(4n+1)\, p_3' p_4' \cdot \sum \frac{/A/^{1/2}}{(2\pi)^{3/2}} \exp. -\frac{1}{2}\left(\sum_1^3 \sum_1^3 A_{ij} Z_i Z_j\right) \cdot$$

$$\cdot\, dZ_1 dZ_2 dZ_3.$$

where $Z_1 = \dfrac{n_1-(4n-1)p_3}{(4n-1)^{1/2}}$

$$Z_2 = \frac{n_2-(4n-1)p_2}{(4n-1)^{1/2}}$$

$$Z_3 = \frac{n_2-(4n-1)p_4}{(4n-1)^{1/2}}$$

and A_{ij}'s are exactly as in S_2 and S_3.

Here also $Z_1 + Z_2 \doteq (4n)^{1/2} \left(\frac{1}{4} - p_2 - p_3\right) = U_1$ (say)

and $Z_1 + Z_3 \doteq (4n)^{1/2} \left(\frac{1}{4} - p_4 - p_3\right) = U_2$ (say)

$\therefore$ $S_4 \doteq q_3' \, q_4' \cdot M$

(v) $$S_5 = p_1' p_4' \sum_{n_1+n_2=n-1} \frac{(4n+1)!}{(3n_1+2n_2+2)!(n_2+1)!n_1!n_4!} \cdot$$

$$\cdot\, p_2^{n_2+1} \cdot p_1^{3n_1+2n_2+2} \cdot p_3^{n_1} \cdot p_4^{n_2} \cdot$$

$$=. \ 4n(4n+1)\ p_1' p_4' \sum \frac{/A/^{1/2}}{(2\pi)^{3/2}} \exp. -\frac{1}{2}\left(\sum_1^3 \sum_1^3 A_{ij} Z_i Z_j\right) dZ_1 dZ_2 dZ_3$$

where $Z_1 = \dfrac{n_1-(4n-1)p_3}{(4n-1)^{1/2}}$;

$$Z_2 = \frac{(n_2+1)-(4n-1)^{1/2} p_2}{(4n-1)^{1/2}} ;$$

$$Z_3 = \frac{n_2-(4n-1)p_4}{(4n-1)^{1/2}}$$

and A_{ij}'s same as in previous cases.

Here also $Z_1 + Z_2 =. (4n)^{1/2}(\frac{1}{4} - p_2-p_3) = U_1$

and $Z_1 + Z_3 =. (4n)^{1/2} (\frac{1}{4} - p_4-p_3) = U_2$

$\therefore \quad S_5 =. q_1' \, q_4' \, M.$

Therefore Distribution of (U_1, U_2) is

$$dF(U_1, U_2) =. \frac{(q_1'q_2'+q_2'q_3'+q_3'q_4'+q_4'q_1')}{2\pi q_1 q_2} \circ \frac{(q_1+q_2+q_3+q_4)^{1/2}}{(q_1q_2q_3+\dots)^{1/2}} \ .$$

$$\exp. - (U_1{}^2-2\beta U_1 U_2+U_2{}^2)/2\sigma^2 . \ dU_1 \ . \ dU_2$$

or

$$dF(U_1, U_2) =_\circ (q_1+q_2+q_3+q_4)^{1/2}/(2\pi)(q_1q_2q_3+q_1q_2q_4+q_1q_3q_4+q_2q_3q_4)^{1/2}{}_\circ$$

$$. \ \exp. -(U_1{}^2-2\beta U_1 U_2+U_2{}^2)/2)^2 . dU_1 dU_2 .$$

The constant term in this integral can be finally chosen as to make the total integral unity.

In Chapter I, I have tabulated values of certain constants defining the asymptotic joint distribution of the first sample quartiles of a bivariate normal population with zero mean and unit variance and correlation coefficient p.

The application of the Edgeworth Series expansion for a population distribution function considered in 1 has prompted the derivation in Chapter 2, of the expectation and variance of the variable $\bar{y}_1 - \bar{x}_1$ when the population distribution is specified to a sufficient degree of accuracy by the first terms of its Bivariate Edgeworth Series.

Finally Chapter 3 embodies a consideration of the use of the "straight product interquartile range", $(\bar{x}_2 - \bar{x}_1)(\bar{y}_2 - \bar{y}_1)$ and the "crossproduct interquartile range" $(\bar{x}_2 - \bar{y}_1)$ $(\bar{x}_2 - \bar{x}_1)$ as estimators of the population correlation coefficient.

CHAPTER I

Let us consider a bivariate population in which the random variables, denoted x and y, have joint probability distribution function f(x,y), such that the conditions (A), specified on page 37 of 1 are satisfied.

i.e. i) $\int_{-\infty}^{\infty} f(x,\frac{1}{N})\, dx = \int_{-\infty}^{\infty} f(x,0)dx + O(\frac{1}{N})$

ii) $\int_{-\infty}^{\infty} f(\frac{1}{N},y)\, dy = \int_{-\infty}^{\infty} f(0,y)dy + O(\frac{1}{N})$

iii) Equations a) $\int_{x=-\infty}^{\xi_1} \int_{y=-\infty}^{\infty} f(x,y)\, dx\, dy = \frac{1}{4}$

b) $\int_{y=-\infty}^{\eta_1} \int_{x=-\infty}^{\infty} f(x,y)\, dx\, dy = \frac{1}{4}$,

have unique roots.

On page 52 of 1 the asymptotic joint distribution function of two linear functions, U_1 and U_2, of the first quartiles, $\bar{x}_1$; $\bar{y}_1$ is given to be

$$dF(U_1,U_2) = \frac{(q_1 + q_2 + q_3 + q_4)^{1/2}}{2\pi(q_1q_2q_3 + q_1q_2q_4 + q_1q_3q_4 + q_2q_3q_4)^{1/2}} \tag{1.1}$$

$$\cdot\exp - \frac{(U_1^2 - 2\beta U_1U_2 + U_2^2}{2\sigma^2}\, dU_1\, dU_2$$

where $U_1 = (4n)^{1/2} \ (a_2\bar{y}_1 - \frac{1}{4})$ and $U_2 = (4n)^{1/2} \ (a_1\bar{x}_1 - \frac{1}{4})$ (1.2)

$$a_1 = \int_{-\infty}^{\infty} f(0,y)\, dy \qquad a_2 = \int_{-\infty}^{\infty} f(x,0)\, dx \tag{1.3}$$

$$\frac{1}{\sigma^2} = \left(\frac{1}{q_1} + \frac{1}{q_2}\right)\left(\frac{1}{q_3} + \frac{1}{q_4}\right) \Big/ \left(\frac{1}{q_1} + \frac{1}{q_2} + \frac{1}{q_3} + \frac{1}{q_4}\right) \tag{1.4}$$

$$\frac{-\beta}{\sigma^2} = \frac{1}{q_1} - \left(\frac{1}{q_1} + \frac{1}{q_2}\right)^2 \Big/ \left(\frac{1}{q_1} + \frac{1}{q_2} + \frac{1}{q_3} + \frac{1}{q_4}\right) \tag{1.5}$$

and $q_1 = \int_{\xi_1}^{\infty} \int_{\eta_1}^{\infty} f(x,y)\, dx\, dy$ (1.6)

$$q_2 = \int_{-\infty}^{\xi_1} \int_{\eta_1}^{\infty} f(x,y)\, dy\, dx \tag{1.7}$$

$$q_3 = \int_{-\infty}^{\xi_1} \int_{-\infty}^{\eta_1} f(x,y)\, dy\, dx \tag{1.8}$$

$$q_4 = \int_{\xi_1}^{\infty} \int_{-\infty}^{\eta_1} f(x,y)\, dx\, dy \tag{1.9}$$

ξ_1, η_1 being the population first quartiles corresponding to $\bar{x}_1$, $\bar{y}_1$, respectively, and • indicating accuracy to order $\frac{1}{\sqrt{n}}$ when $4n + 1$ is the sample size

i.e. We write $A = \cdot B$ (1.10)

if $A = B\left[1 + O(1/n^{\frac{1}{2}})\right]$, where $O(\frac{1}{n})$ represents a function such that $\lim_{N \to \infty} N(\frac{1}{N}) = L < \infty$.

It is observed that if $f(x,y)$ is such as to allow commutativity of integration w.r.t. variable x and that w.r.t. variable y then

$$q_2 = \frac{3}{4} - q_1 \qquad (1.11)$$

$$q_3 = -\frac{1}{2} + q_1 \qquad (1.12)$$

$$q_4 = \frac{3}{4} - q_1 = q_2. \qquad (1.13)$$

In the case where the samples is drawn from a bivariate normal population of known means, μ_x, μ_y; variances, σ_x^2 σ_y^2 and correlation coefficient p.

$$\xi_1 = -.6745\ \sigma_x + \mu_x \qquad (1.14)$$

$$\eta_1 = -.6745\ \sigma_y + \mu_y \qquad \text{(page 136 of 10)} \qquad (1.15)$$

and for $\mu_x = \mu_y = 0$ and $\sigma_x^2 = \sigma_y^2 = 1$, the standard bivariate normal population, we have

$$q_1 = \int_{-.6745}^{\infty} \int_{-.6745}^{\infty} \frac{1}{2\pi(1-p^2)^{1/2}} \exp \frac{-1}{2(1-p^2)} \quad x^2-2pxy+y^2 \quad dx\, dy. \qquad (1.16)$$

Tables for the function

$$\int_{h}^{\infty} \int_{k}^{\infty} \frac{1}{2\pi(1-p^2)^{1/2}} \exp \frac{-1}{2(1-p^2)} \left\{ x^2-2pxy+y \right\} dx\, dy$$

have been given in 11 for p = 0 (.1) 1 , whence q_1 may be approximated from tabled values of p by using repeated applications of Everett's interpolation formula.

TABLE OF VALUES FOR q_1, $\frac{1}{\sigma^2}$, $\frac{-\beta}{\sigma^2}$, β

p	q_1	$q_2 = q_4 = .75 - q_1$	$q_3 = -\frac{1}{2} + q_1$	$\frac{1}{\sigma^2}$	$\frac{-\beta}{\sigma^2}$	β
.1	.572 832 49	.177 167 51	.072 832 49	5.349 578 59	0.294 797 16	.055 106 61
.2	.583 674 14	.166 325 86	.083 674 14	5.402 227 40	0.610 066 76	.112 928 74
.3	.595 101 22	.154 898 78	.095 101 22	5.499 596 46	0.956 232 29	.173 873 17
.4	.607 238 22	.142 761 78	.107 238 22	5.655 299 67	1.349 376 22	.238 603 84
.5	.620 273 28	.129 726 72	.120 273 28	5.871 359 66	1.803 242 36	.307 125 17
.6	.634 503 05	.115 496 95	.134 503 05	6.255 878 27	2.402 359 01	.384 016 26
.7	.650 453 67	.099 546 33	.150 453 67	6.837 974 16	3.207 599 58	.469 086 23
.8	.669 080 19	.080 919 81	.169 080 19	7.879 173 19	4.478 740 14	.568 427 68
.9	.693 042 61	.056 947 39	.193 042 61	10.350 608 51	7.206 375 74	.696 227 25

Correct to 6 places of decimal.

CHAPTER II

Suppose the joint probability distribution of the two random variables satisfies the specified conditions and may be approximated by the terms of its Edgeworth series containing fourth and lower order derivatives w.r.t. each variable, so that

$$f(x,y) = \exp\left\{ \sum{}' (-1)^{r+s} K_{rs} \frac{D_1^r}{r!} \frac{D_2^s}{s!} \right. \alpha(x)\, \alpha(y) \qquad (2.1)$$

[where operator $D_1 \equiv \frac{\partial}{\partial x}$ and $D_2 \equiv \frac{\partial}{\partial y}$;

$$\alpha(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}; \qquad (2.2)$$

the summation $\sum{}'$ extends over all values of $r+s \geq 3$ together with the term K_{11}; and the K's are the bivariate cumulants of the distribution in standard measure so that $K_{11} = \mu_{11} = p$ and terms K_{01}, K_{10}, K_{20}, K_{02} do not appear.]

may be approximated by

$$f(x,y) = \left\{ 1 + K_{11}D_1D_2 - \frac{K_{30}}{3!} D_1^3 - \frac{K_{21}}{2!} D_1^2 D_2 - \frac{K_{12}}{2!} D_1 D_2^2 \right.$$

$$- K_{03} \frac{D_2^3}{3!} + \frac{K_{40}}{4!} D_1^4 + \frac{K_{31}}{3!} D_1^3 D_2 + \left(\frac{K_{22}}{2!2!} + \frac{K_{11}^2}{2!}\right) D_1^2 D_2^2$$

$$\left. + \frac{K_{13}}{3!} D_1 D_2^3 + \frac{K_{04}}{4!} D_2^4 \right\} \alpha(x)\, \alpha(y) \qquad (2.3)$$

Then the population first quartiles are ξ_1, η_1, where these are the solutions of

$$\int_{\xi_1}^{\infty} \alpha(x)\,dx + \frac{K_{30}}{3!}\left(-\alpha(\xi_1) + \xi_1^2\,\alpha(\xi_1)\right) - \frac{K_{40}}{4!}\left[3\alpha(\xi_1) - 6\xi_1^2\,\alpha(\xi_1) + \xi_1^4\alpha(\xi_1)\right] = \frac{3}{4} \quad (2.4)$$

(1)

and

$$\int_{\eta_1}^{\infty} \alpha(y)\,dy + \frac{K_{03}}{3!}\left[-\alpha(\eta_1) + \eta_1^2\,\alpha(\eta_1)\right] - \frac{K_{40}}{4!}\left[3\alpha(\eta_1) - 6\eta_1^2\,\alpha(\eta_1) + \eta_1^4\alpha(\eta_1)\right] = \frac{3}{4}$$

(2.5)

Utilizing result: $\lim_{x\to\infty} x^a e^{-\frac{1}{2}x^2} = 0$ for all real a.

and notation $p(x) = \int_x^{\infty} \alpha(t)\,dt$ the equations for evaluating ξ_1 and η_1 may be written (2.6)

$$p(\xi_1) - \frac{K_{30}}{3!}(1-\xi_1^2)\,\alpha(\xi_1) + \frac{K_{40}}{4!}\left[-3 + 6\,\xi_1^2 - \xi_1^4\right]\alpha(\xi_1) = \frac{3}{4} \quad (2.7)$$

(1)

$$p(\eta_1) - \frac{K_{03}}{2!}(1-\eta_1^2)\,\alpha(\eta_1) + \frac{K_{04}}{4!}\left[-3 + 6\eta_1^2 - \eta_1^4\right]\alpha(\eta_1) = \frac{3}{4} \quad (2.8)$$

and having solved for ξ_1 and η_1 one can determine q_1 by

$$q_1 = \int_{\xi_1}^{\infty}\int_{\eta_1}^{\infty} f(x,y)\,dx\,dy$$

$$= p(\xi_1)\,p(\eta_1) + K_{11}\,\alpha(\xi_1)\,\alpha(\eta_1) - \frac{K_{30}}{3!}\left[\alpha(\xi_1) - \xi_1^2\alpha(\xi_1)\right]p(\eta_1)$$

$$+ \frac{K_{21}}{2!}\,\xi_1\,\alpha(\xi_1)\,\alpha(\eta_1) + \frac{K_{12}}{2!}\,\alpha(\xi_1)\,\eta_1\,\alpha(\eta_1) - \frac{K_{03}}{3!}\,p(\xi_1)\left[\alpha(\eta_1) - \eta_1^2\,\alpha(\eta_1)\right]$$

$$+ \frac{K_{40}}{4!}\left[-3\,\xi_1\,\alpha(\xi_1) + \xi_1^3\,\alpha(\xi_1)\right] p(\eta_1) + \frac{K_{31}}{3!}\left[-\alpha(\xi_1) + \xi_1^2\,\alpha(\xi_1)\right]\alpha(\eta_1)$$

$$+ \frac{K_{22}}{2!2!}\,\xi_1\,\eta_1\,\alpha(\xi_1)\,\alpha(\eta_1) + \frac{K_{13}}{3!}\,\alpha(\xi_1)\left[-\alpha(\eta_1) + \eta_1^2\,\alpha(\eta_1)\right]$$

$$+ \frac{K_{04}}{4!}\,p(\xi_1)\left[-3\,\eta_1\alpha(\eta_1) + \eta_1^3\,\alpha(\eta_1)\right] + \frac{K_{11}^2}{2!}\,\xi_1\eta_1\,\alpha(\xi_1)\,\alpha(\eta_1) \qquad (2.9)$$

Thus employing methods of Numerical Analysis ξ_1 and η_1 may be determined from equations (1) and hence q_1 may be evaluated for any specific distribution satisfying the specified conditions, and given by its Edgeworth series truncated at terms of order four in derivatives. In particular if the population under consideration follows the standard bivariate normal distribution $\xi_1 = \eta_1 = -.6745$ and by p.82 of 6

$$K_{11} = \rho$$

$$K_{21} = 0 = K_{12}$$

$$K_{30} = K_{03} = 0$$

$$K_{04} = K_{40} = 0$$

$$K_{31} = K_{13} = 0$$

$$K_{22} = 0$$

so q_1 reduces to form:

$$q_1 = \left[p(-0.6745)\right]^2 + \rho.\left[\alpha(-0.6745)\right]^2 + \tfrac{1}{2}\rho^2.\left[-0.6745\right]^2.\left[\alpha(-0.6745)\right]^2 \qquad (2.10)$$

$$= 0.5625 + 0.100\,980\rho + 0.015314\rho^2$$

and $\frac{1}{q_1} = 1.\dot{7} - 0.31\ 9\ 147p - 0.088\ 903p^2$

$q_4 = q_2 = .1825 - 0.100\ 980p - 0.045\ 941p^2$

$\frac{1}{q_4} = \frac{1}{q_2} = 5.479\ 452 + 3.031\ 886p + 2.969\ 908p^2$

$q_3 = 0.0625 + 0.100\ 980p + 0.045\ 94p^2$

$\frac{1}{q_3} = 16 - 2.585\ 088p + 3.000\ 611p^2$

$\frac{1}{q_1} + \frac{1}{q_2} = 7.25\ 7229 + 2.712\ 739p + 2.88\ 1005p^2$

$\frac{1}{q_3} + \frac{1}{q_4} = 21.479\ 452 + .446\ 798p + 5.97\ 0519p^2$

$(\frac{1}{q_1} + \frac{1}{q_2} + \frac{1}{q_3} + \frac{1}{q_4}) = 28.736\ 681 + 3.159\ 537p + 8.851\ 524p^2$

$(\frac{1}{q_1} + \frac{1}{q_2} + \frac{1}{q_3} + \frac{1}{q_4})^{-1} = 0.034799 - 0.000\ 383p - 0.001\ 033p^2$

$(\frac{1}{q_1} + \frac{1}{q_2})(\frac{1}{q_3} + \frac{1}{q_4}) = 155.881\ 302 + 61.5106\ 63p + 106.423\ 879p^2$

$$\frac{(\frac{1}{q_1} + \frac{1}{q_2})(\frac{1}{q_3} + \frac{1}{q_4})}{(\frac{1}{q_1} + \frac{1}{q_2} + \frac{1}{3} + \frac{1}{4})} = 5.424\ 513 + -1.543\ 484p + 3.518\ 861p^2$$

$(\frac{1}{q_1} + \frac{1}{q_2})^2 = 52.667\ 373 + 39.373\ 936p + 28.267\ 066p^2$

$$\frac{\left(\frac{1}{q_1} + \frac{1}{q_2}\right)^2}{\left(\frac{1}{q_1} + \frac{1}{q_2} + \frac{1}{q_3} + \frac{1}{q_4}\right)} = 1.832\ 772 + 1.350002p + 0.914\ 180p^2$$

$$\frac{1}{q_1} = 1.777\ 778 - 0.319\ 147p - 0.088\ 903p^2$$

$$\frac{\left(\frac{1}{q_1} + \frac{1}{q_2}\right)^2}{\left(\frac{1}{q_1} + \frac{1}{q_2} + \frac{1}{q_3} + \frac{1}{q_4}\right)} - \frac{1}{q_1} = 0.054\ 994 + 1.669\ 149p + 1.003\ 083p^2.$$

Thus $$\frac{1}{\sigma^2} = 5.424513 - 1.543\ 484p + 3.518\ 861p^2 \qquad (2.11)$$

$$\frac{\beta}{\sigma^2} = 0.054994 + 1.669\ 149p + 1.003\ 083p^2 \qquad (2.12)$$

Continuing our consideration of the standardized Bivariate normal population.

If U_1 and U_2 have joint probability distribution function $k \exp - \frac{1}{2} \left(\frac{U_1^2}{\sigma^2} - 2\beta \frac{U_1}{\sigma} \frac{U_2}{\sigma} + \frac{U_2^2}{\sigma^2}\right) dU_1\, dU_2$ where the constant k may be determined by integration.

$$k \exp - \frac{1}{2} \left(\frac{U_1^2}{\sigma^2} - 2\beta \frac{U_1}{\sigma} \frac{U_2}{\sigma} + \frac{U_2^2}{\sigma^2}\right) dU_1\, dU_2$$

$$= \frac{k\, 2\pi}{\left(\frac{1}{\sigma^4} - \left(\frac{\beta}{\sigma^2}\right)^2\right)} \frac{1}{2\pi\left(1-\beta^2\right)^{1/2}} \exp - \frac{1}{2(1-\beta^2)} \left\{ \left(U_1 \frac{\sqrt{1-\beta^2}}{\sigma}\right)^2 \right.$$

$$\left. - 2\beta\left(U_1 \frac{\sqrt{1-\beta^2}}{\sigma}\right)\left(U_2 \frac{\sqrt{1-\beta^2}}{\sigma}\right) + \left(U_2 \frac{\sqrt{1-\beta^2}}{\sigma}\right)^2 \right\} . \; dU_1 dU_2 \left(\frac{1}{\sigma^4} - \left(\frac{\beta}{\sigma^2}\right)^2\right) . \; \sigma^2 \quad (2.13)$$

Thus variables $V_1 = U_1 \frac{\sqrt{1-\beta^2}}{\sigma}$ and $V_2 = U_2 \frac{\sqrt{1-\beta^2}}{\sigma}$ follow the standardized bivariate normal distribution, with correlation coefficient β

$$\text{and} \quad k = \frac{1}{2\pi} \left\{ \frac{1}{\sigma^4} - \left(\frac{\beta}{\sigma^2}\right)^2 \right\} . \; \sigma^2$$

$$\bar{x}_1 = - \frac{1}{a_1} \left\{ V_1 \frac{\sigma}{\sqrt{4n(1-\beta^2)}} + \frac{1}{4} \right\} \quad \text{where } a_1 = \int_{-\infty}^{\infty} f(0,y_2)\, dy = \frac{1}{\sqrt{2\pi}} \quad (2.14)$$

for Bivarate normal

$$\bar{y}_2 = - \frac{1}{a_2} \left\{ V_2 \frac{\sigma}{\sqrt{4n(1-\beta^2)}} + \frac{1}{4} \right\} \quad \text{and } a_2 = \frac{1}{\sqrt{2\pi}} . \quad (2.15)$$

$$\bar{y}_1 - \bar{x}_1 = -\frac{1}{\sqrt{2\pi}} \cdot \frac{\sigma}{\sqrt{4n(1-\beta^2)}} \; (V_2 - V_1). \qquad (2.16)$$

Thus Exp. $(\bar{y}_1 - \bar{x}) = 0$

$$\text{Exp.}\ (\bar{y}_1 - \bar{x}_1)^2 = \frac{\text{Exp}\ (V_2 - V_1)^2}{8\pi n(1/_\sigma 2 - \beta^2/_\sigma 2)} = \frac{\frac{1}{\sigma^2}}{8\pi n\left[(1/_\sigma 2)^2 - (\beta/_\sigma 2)^2\right]}\ \text{Exp.}(V_2 - V_1)^2$$

$$= \frac{\frac{1}{\sigma^2}}{8\pi n \quad (1/_\sigma 2)^2 - (\beta/_\sigma 2)^2} \quad (2 - 2\beta) \qquad (2.17)$$

$$= \frac{1}{4\pi n\left(1/_\sigma 2 + \beta/_\sigma 2\right)} = \frac{1}{4\pi n}\left\{5.479\ 507 + 0.125\ 665\rho + 4.521\ 944\rho^2\right\}^{-1}$$

$$= \frac{1}{4\pi n}\left\{0.182\ 498 - 0.004\ 185\ 3\rho - 0.150\ 510\rho^2\right\}$$

$$\text{Exp.}(\bar{y}_1 - \bar{x}_1)^4 = \frac{\frac{1}{\sigma^4}}{(8\pi n)^2\left\{(\frac{1}{\sigma^2})^2 - (\frac{\beta}{\sigma^2})^2\right\}^2}\ \text{Exp}\left\{V_2^4 - 4V_2^3V_1 + 6V_2^2V_1^2 - 4V_2V_1^3 + V_1^4\right\}$$

$$= \frac{\frac{1}{\sigma^4}}{(8\pi n)^2\left\{(\frac{1}{\sigma^2})^2 - (\frac{\beta}{\sigma^2})\right\}^2} \cdot 12(1-\beta^2)$$

$$= \frac{3}{(4\pi n)^2\left\{\frac{1}{\sigma^2} + \frac{\beta}{\sigma^2}\right\}^2} \qquad (2.18)$$

$$\therefore \text{Var.}\ (\bar{y}_1-\bar{x}_1)^2 = \text{Exp}(y_1-\bar{x}_1)^4 - [\text{Exp.}(\bar{y}_1-\bar{x}_1)^2]^2$$

$$= \frac{3}{(4\pi n)^2}\ \frac{1}{\left\{\frac{1}{\sigma^2}+\frac{\beta}{\sigma^2}\right\}^2} - \frac{1}{(4\pi n)^2\left\{\frac{1}{\sigma^2}+\frac{\beta}{\sigma^2}\right\}^2}$$

$$= \frac{2}{(4\pi n)^2}\ \frac{1}{\left\{\frac{1}{\sigma^2}+\frac{\beta}{\sigma^2}\right\}^2} \qquad (2.19)$$

$$\simeq \frac{2}{(4\pi n)^2}\left\{0.033306 - 0.001\ 528\rho - .027\ 450\rho^2\right\} \qquad (2.20)$$

Thus the linear function of ρ which approximates $\text{Exp.}(\bar{y}_1-\bar{x}_1)^2$ is

$$\frac{1}{4\pi n}\left[0.182\ 498 - 0.004\ 185\rho\right]$$

so that to this degree of approximation

$$p' = \frac{1}{0.004\ 185}\left[(\bar{y}_1-\bar{x}_1)^2 4n\pi - 0.182\ 498\right] \text{ is an}$$

estimate of the correlation coefficient ρ and

$$\text{Var.}\ p' = \left(\frac{4n\pi}{.004\ 185}\right)^2 \text{var}\ (\bar{y}_1-\bar{x}_1)^2.$$

Now the variance of the minimum variance estimator ρ. is for a sample of size 4n+1 is

$$\frac{1-\rho^2}{(4n+1)(1+\rho^2)}$$

Hence using R. A. Fisher's definition the efficiency of p', defined above, is

$$\frac{1-\rho^2}{(4n+1)(1+\rho^2)} \cdot \left[\frac{0.004185}{4n\pi}\right]^2 \cdot \frac{(4\pi n)^2}{2} \cdot \left\{\frac{1}{\sigma^2} \mp \frac{\beta}{\sigma^2}\right\}^2$$

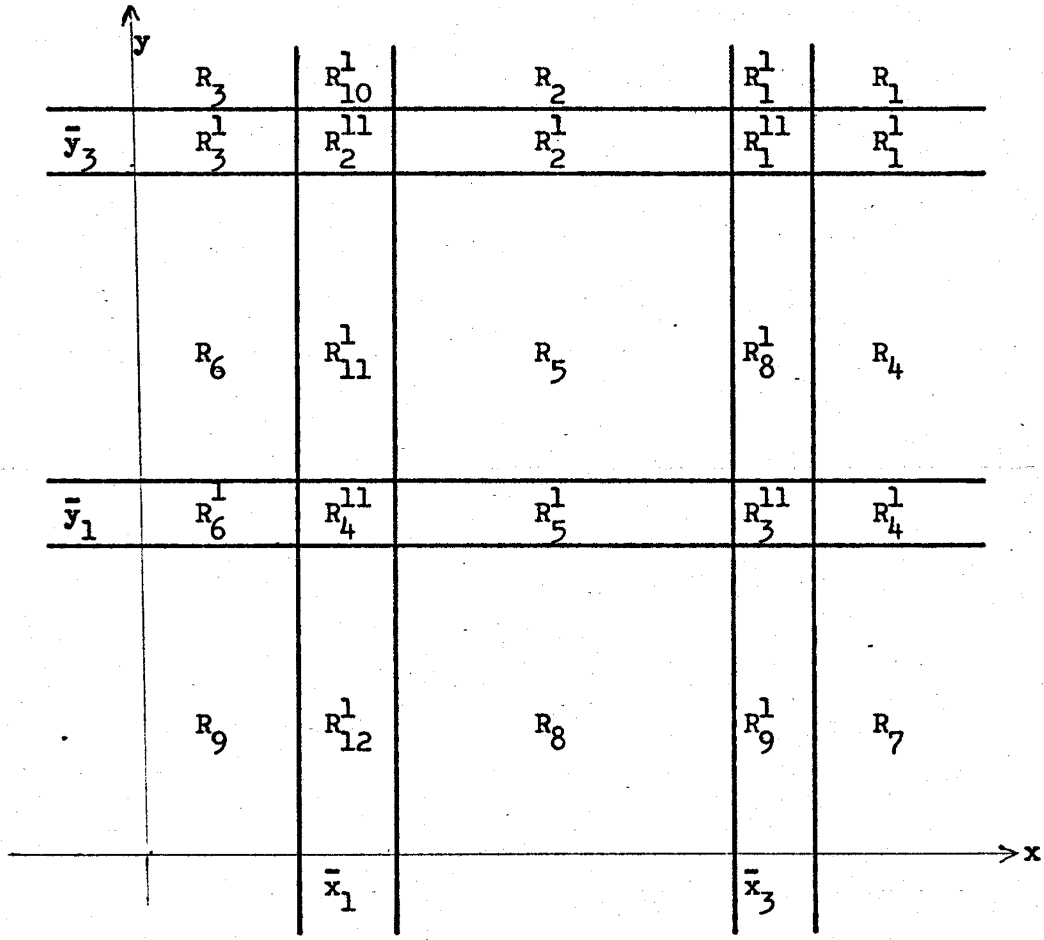
$$= \frac{1}{4n+1} \left[.000\,263 + .0000\,12\rho \dot{+} .000\,092\rho^2\right]$$

CHAPTER III

INTRODUCTION:

In this section I shall use Banerjee's derivation of the joint distribution of first quartiles and second quartiles in the case of bivariate normal population 1 to obtain the mean value of the statistic $(\bar{x}_3 - \bar{x}_1)(\bar{y}_3 - \bar{y}_1)$ in terms of the population correlation coefficient ρ where $\bar{x}_1$, $\bar{x}_3$ are defined as the first and third quartiles of the x variate, and $\bar{y}_1$, $\bar{y}_3$ the corresponding quartiles of the other variate y in a bivariate normal population, approximated by the first few terms of its Edgeworth series.

Let f(x,y) be the joint probability density function of two variables from which a sample of 4n + 2 observations (x_r, y_r) $1 \leq r \leq n$ is drawn. Assume that f(x,y) satisfies the conditions specified on page 13; $\bar{x}_1$, $\bar{x}_3$, $\bar{y}_1$; $\bar{y}_3$ are as defined above, and ξ_1 ξ_3, η_1 η_3 are the corresponding population quartiles. $(\bar{x}_1, \bar{x}_3, \bar{y}_1, \bar{y}_3)$ will be referred to as the quartile.



The x - y plane is divided into 25 regions by the lines

$x = \bar{x}_1 \pm \frac{1}{2}\, d\bar{x}_1$ $\qquad$ $y = \bar{y}_1 \pm \frac{1}{2}\, d\bar{y}_1$

$x = \bar{x}_3 \pm \frac{1}{2}\, d\bar{x}_3$ $\qquad$ $y = \bar{y}_3 \pm \frac{1}{2}\, d\bar{y}_3$

and we define p_{ij} to be the probability that an element falls in the region R_i

$$\text{i.e.} \quad p_{ij} = \iint_{R_i^j} f(x,y)\, dx\, dy. \qquad (3.1)$$

$$p_1 = \int_{x=\bar{x}_3}^{\infty} \int_{y=\bar{y}_3}^{\infty} f(x,y)\, dx\, dy \qquad (3.2)$$

$$p_2 = \int_{x=\bar{x}_1}^{\bar{x}_3} \int_{y=\bar{y}_3}^{\infty} f(x,y)\, dx\, dy \qquad (3.3)$$

$$p_3 = \int_{x=-\infty}^{\bar{x}_1} \int_{y=\bar{y}_3}^{\infty} f(x,y)\, dx\, dy \qquad (3.4)$$

$$p_4 = \int_{x=\bar{x}_3}^{\infty} \int_{y=\bar{y}_1}^{\bar{y}_3} f\, dx\, dy \qquad (3.5)$$

$$p_5 = \int_{x=\bar{x}_1}^{\bar{x}_3} \int_{y=\bar{y}_1}^{\bar{y}_3} f\, dx\, dy \qquad (3.6)$$

$$p_6 = \int_{x=-\infty}^{\bar{x}_1} \int_{y=\bar{y}_1}^{\bar{y}_3} f\, dx\, dy \qquad (3.7)$$

$$p_7 = \int_{x=\bar{x}_3}^{\infty} \int_{y=-\infty}^{\bar{y}_1} f\, dx\, dy \qquad (3.8)$$

$$p_8 = \int_{x=\bar{x}_1}^{\bar{x}_3} \int_{y=-\infty}^{\bar{y}_1} f\, dx\, dy \qquad (3.9)$$

$$p_9 = \int_{x=-\infty}^{\bar{x}_1} \int_{y=-\infty}^{\bar{y}_1} f\, dx\, dy \qquad (3.10)$$

The technique of considering the sample as drawn from a multinomial population with probabilities $p_1, \ldots, p_4''$ is then employed and those terms which give rise to the observed sample quartile $(\bar{x}_1\ \bar{x}_3\ \bar{y}_1\ \bar{y}_3)$ are picked out. For example, the quartile may be determined by one pair falling in R_1'' and the other in R_4'', the remaining 4n pairs being distributed in zones R_i with frequency $n_i, 1 \leq i \leq 9$ with

$$\sum_{i=1}^{9} n_i = 4n$$

$$\begin{aligned} n_1 + n_2 + n_3 &= n & \qquad n_1 + n_4 + n_9 &= n \\ n_4 + n_5 + n_6 &= 2n & \qquad n_2 + n_5 + n_8 &= 2n \\ n_7 + n_8 + n_9 &= n & \qquad n_3 + n_6 + n_9 &= n. \end{aligned} \tag{3.10}$$

The probability that this occurs is

$$s_1 = p_1''\ p_4''\ \sum (4n+2)!\ \prod_{i=1}^{9} \frac{p_i^{n_i}}{n_i!} \quad \text{where the summation} \tag{3.11}$$

is taken over all possible arrays n_i with $\sum_i n_i = 4n$.

Another possibility is that in which the quartile $(\bar{x}_1\ \bar{x}_3\ \bar{y}_1\ \bar{y}_3)$ is determined by four different elements in the sample, for example, one in R_6', one in R_{11}', one in R_2' and the other in R_7' the other 4n-2 elements falling in the regions R_i with frequences n_i $1 \leq i \leq a$ such that

$$\sum_{i=1}^{8} n_i = 4n - 2$$

$$\begin{array}{ll} n_1 + n_2 + n_3 + 1 = n & n_1 + n_4 + n_7 = 4 \\ n_4 + n_5 + n_6 + 1 = 2n & n_2 + n_5 + n_8 + 1 = 2n \\ n_7 + n_8 + n_9 = n & n_3 + n_6 + n_9 + 1 = n \end{array} \tag{3.12}$$

and the probability that this occurs is

$$s_2 = p_2' \, p_6' \, p_7' \, p_{11}' \sum_{\Sigma n_i = 4n-2} (4n+2)! \prod_{i=1}^{9} \frac{p_i^{n_i}}{n_i!} \tag{3.13}$$

Banerjee lists the other 32 ways in which $(\bar{x}_1 \; \bar{x}_3 \; \bar{y}_1 \; \bar{y}_3)$ may be obtained and derives the asymptotic joint distribution, to approximation of order $n^{-1/2}$ by using the normal approximation for the multinomial distribution when the sample size 4n+2, and hence n is large; and computing the sums involved by integration.

For example,

$$\text{defining} \qquad z_i = \frac{n_i - (4n-2)p_i}{(4n-2)^{1/2}} \tag{3.14}$$

in the normal approximation

$$s_2 = .(4n)^4 \, p_2' \, p_6' \, p_7' \, p_{11}' \sum \frac{|A|^{\frac{1}{2}}}{(2\pi)^4} \exp - \frac{1}{2} \left\{ \sum_1^8 \sum_1^8 A_{ij} z_i z_j \right\} \prod dz_i \tag{3.15}$$

where $A = \begin{matrix} A_{11}, \ldots, A_{18} \\ \vdots \\ A_{81}, \ldots, A_{88} \end{matrix}$

$$\text{with } A_{ij} = A_{ji} = \frac{1}{pa} \qquad i \neq j$$

$$A_{ii} = \frac{1}{pi} + \frac{1}{pa} \qquad (*) \qquad (3.16)$$

By the conditions which f(x,y) is assumed to satisfy

$$p_1 + p_2 + p_3 = \int_{-\infty}^{\infty} \int_{\bar{y}_2}^{\infty} f(x,y)\, dx\, dy = .\frac{1}{2} - b(\bar{y}_2) \qquad (3.17)$$

$$p_4 + p_5 + p_6 = \int_{-\infty}^{\infty} \int_{\bar{y}_1}^{\bar{y}_2} f(x,y)\, dx\, dy = .b(\bar{y}_3 - \bar{y}_1) \qquad (3.18)$$

$$p_1 + p_4 + p_7 = \int_{-\infty}^{\infty} \int_{\bar{x}_2}^{\infty} f(x,y)\, dx\, dy = .\frac{1}{2} - a\bar{x}_3 \qquad (3.19)$$

$$p_2 + p_5 + p_8 = \int_{-\infty}^{\infty} \int_{\bar{x}_1}^{\bar{x}_2} f(x,y)\, dx\, dy = .a(\bar{x}_2 - \bar{x}_1) \qquad (3.20)$$

$$\text{where } a = \int_{-\infty}^{\infty} f(x\,,0)\, dx \quad \text{and } b = \int_{-\infty}^{\infty} f(0,x)\, dx_3 \qquad (3.21)$$

$$\text{and } z_1 + z_2 + z_3 = .(4n)^{\frac{1}{2}} \left(-\frac{1}{4} + b\bar{y}_3\right) \qquad (3.22)$$

$$z_4 + z_5 + z_6 = .(4n)^{\frac{1}{2}} \left(\frac{1}{2} - b(\bar{y}_3 - \bar{y}_1)\right). \qquad (3.23)$$

$$z_1 + z_4 + z_7 = .(4n)^{\frac{1}{2}} \left(- \frac{1}{4} + a\bar{x}_2\right) \tag{3.24}$$

$$z_2 + z_5 + z_8 = .(4n)^{\frac{1}{2}} \left(\frac{1}{2} - a(\bar{x}_2 - \bar{x}_1\right). \tag{3.25}$$

If q_i and $q_i^{'}$ be the integrals represented in the equations on page 28 with $\bar{x}_1\ \bar{x}_3\ \bar{y}_1\ \bar{y}_3$ replaced by the population quartiles $\xi_1\ \xi_3\ \eta_1\ \eta_3$ respectively

$$\text{i.e. } q_1 = \int_{\eta_3}^{\infty} \int_{\xi_3}^{\infty} f(x,y)\, dx\, dy \tag{3.26}$$

$$q_3^{'} = \int_{-\infty}^{\xi_1} f(x,y_2)\, dx \qquad \text{etc.}$$

and we have

$$\begin{aligned} p_i &= .q_i & 1 \leq i \leq 9 \\ p_i^{'} &= .q_i^{'} d\bar{x}_1 & i = 10, 11, 12 \\ p_i^{'} &= .q_i^{'}\, d\bar{x}_2 & i = 7, 8, 9 \\ p_i &= .q_i^{'}\, d\bar{y}_1 & i = 4, 5, 6 \\ p_i^{'} &= .q_i\, d\bar{y}_2 & i = 1, 2, 3 \end{aligned} \tag{3.27}$$

Thus if p_i and $p_i^{'}$ are replaced by the corresponding q_i and $q_i^{'}$, after the transformation:

$$z_1 + z_2 + z_3 = U_1 \tag{3.28}$$

$$z_4 + z_5 + z_6 = U_2 \tag{3.29}$$

$$z_1 + z_4 + z_7 = U_3 \tag{3.30}$$

$$z_2 + z_5 + z_8 = U_4 \tag{3.31}$$

and $$z_1 = z_1,\ z_2 = z_2,\ z_4 = z_4,\ z_5 = z_5. \tag{3.32}$$

yielding

$$z_3 = U_1 - (z_1 + z_2)$$

$$z_6 = U_2 - (z_4 + z_5)$$

$$z_7 = U_3 - (z_1 + z_4)$$

$$z_8 = U_4 - (z_2 + z_5)$$

the exponent in the probability s_2 is

$$= -\frac{1}{2}\Big(\sum_{i,j=1,2,4,5}\sum A_{ij}\, z_i\, z_j + A_{33}(U_1 - z_1 - z_2)^2 + \text{etc.} \tag{3.33}$$

with A calculated from equations * on page 31 of this section, but with q_i replacing p_i $\quad 1 \leq i \leq 8$.

$$= -\frac{1}{2}\big(Q_1(U) + Q_2(z) - 2L(z)\big) \text{ say} \tag{3.34}$$

where $Q_1(U) = A_{33}U_1^2 + A_{66}U_2^2 + A_{77}U_3^2 + A_{88}U_4^2 + 2A_{36}U_1U_2 + 2A_{37}U_1U_3$

$$+ 2A_{38}U_1U_4 + 2A_{67}U_2U_3 + 2A_{68}U_2U_4 + 2A_{78}U_3U_4 \tag{3.35}$$

$Q_2(z) = z'\,C\,z$ with $\underline{z}' = (z_1, z_2, z_4, z_5)$ and matrix C having elements. (3.36)

$$c_{11} = A_{11} + A_{33} + A_{77} + 2A_{37}$$

$$c_{12} = A_{12} + A_{33} + A_{37} + A_{38} + A_{78} = c_{21}$$

$$c_{13} = A_{14} + A_{73} + A_{36} + A_{37} + A_{67} = c_{31}$$

$$c_{14} = A_{15} + A_{36} + A_{38} + A_{67} + A_{78} = c_{41} \qquad (3.37)$$

$$c_{22} = A_{22} + A_{33} + A_{88} + 2A_{38}$$

$$c_{23} = A_{24} + A_{36} + A_{37} + A_{68} + A_{78} = c_{32}$$

$$c_{24} = A_{25} + A_{88} + A_{36} + A_{38} + A_{68} = c_{42}$$

$$c_{33} = A_{44} + A_{66} + A_{77} + 2A_{67}$$

$$c_{34} = A_{43} + A_{66} + A_{67} + A_{68} + A_{78} = c_{43}$$

$$c_{44} = A_{55} + A_{66} + A_{88} + 2A_{68}$$

and $L(z) = (U_1, U_2, U_4, U_5)\ B(z_1, z_2, z_4, z_5)'$ (3.38)

where

$$B_{11} = A_{33} + A_{37} \quad ; \quad B_{12} = A_{33} + A_{38}$$

$$B_{13} = A_{36} + A_{37} \qquad B_{14} = A_{36} + A_{38}$$

$$B_{21} = A_{36} + A_{67} \qquad B_{22} = A_{36} + A_{68} \qquad (3.39)$$

$$B_{23} = A_{66} + A_{67} \qquad B_{24} = A_{66} + A_{68}$$

$$B_{31} = A_{77} + A_{37} \qquad B_{32} = A_{37} + A_{78}$$

$$B_{33} = A_{67} + A_{77} \qquad B_{34} = A_{67} + A_{78}$$

$B_{41} = A_{38} + A_{78}$ $\qquad B_{42} = A_{38} + A_{88}$

$B_{43} = A_{68} + A_{78}$ $\qquad B_{44} = A_{68} + A_{88}$

If $f(x,y)$ satisfies the conditions stated earlier and specified by the terms of order four (in derivatives) of its bivariate Edgeworth series, using the same method as in Chapter I it can be shown that

$$p_1 = \int_{\xi_3}^{\infty}\int_{\eta_3}^{\infty} = \int_{\xi_3}^{\infty} \alpha(x)\,dx \int_{\eta_3}^{\infty} \alpha(y)\,dy + K_{11}\Big[\alpha(x)\Big]_{\xi_3}^{\infty}\Big[\alpha(y)\Big]_{\eta_3}^{\infty}$$

$$- \frac{K_{30}}{3!}\Big[-\alpha(x) + x^2\alpha(x)\Big]_{\xi_3}^{\infty} \int_{\eta_3}^{\infty} \alpha(y)\,dy - \frac{K_{21}}{2!}\Big[-x\alpha(x)\Big]_{\xi_3}^{\infty}\Big[\alpha(y)\Big]_{\eta_3}^{\infty}$$

$$- \frac{K_{12}}{2!}\Big[\alpha(x)\Big]_{\xi_3}^{\infty}\Big[-y\alpha(y)\Big]_{\eta_3}^{\infty} - \frac{K_{03}}{3!}\int_{\xi_3}^{\infty} \alpha(x)\,dx \cdot \Big[-\alpha(y) + y^2\alpha(y)\Big]_{3}^{\infty}$$

$$+ \frac{K_{40}}{4!}\Big[3x\alpha(x) - x^3\alpha(x)\Big]_{\xi_3}^{\infty} \int_{\eta_3}^{\infty} \alpha(y)\,dy + \frac{K_{31}}{3!}\Big[-\alpha(x) + x^2\alpha(x)\Big]_{\xi_3}^{\infty}\Big[\alpha(y)\Big]_{\eta_3}^{\infty}$$

$$+ \left(\frac{K_{22}}{2!2!} + \frac{K_{11}^2}{2!}\right)\Big[-x\alpha(x)\Big]_{\xi_3}^{\infty}\Big[-y(\alpha(y)\Big]_{\eta_3}^{\infty} + \frac{K_{13}}{3!}\Big[\alpha(x)\Big]_{\xi_3}^{\infty}\Big[-\alpha(y) + y^2\alpha(y)\Big]_{\eta_3}^{\infty}$$

$$+ \frac{K_{04}}{4!}\Big[3y\alpha(y) - y^3\alpha(y)\Big] \cdot \int_{\xi_3}^{\infty} \alpha(x)\,dx.$$

For the standardized normal where $K_{11} = p$, $K_{21} = 0 = K_{12}$, $K_{30} = K_{03} = 0$, $K_{40} = 0$, $K_{31} = K_{13} = 0$ and $K_{22} = 0$. This reduces to the quadratic approximation in terms of p.

$$p_1 = \frac{1}{16} + \rho.(0.317774)^2 + \frac{1}{2}\rho^2(0.6745 \times 0.217774)^2$$

$$= .0625000 + .100980\rho + .022970\rho^2$$

and $p_q = p_1$.

Similarly it can be shown that

$$p_2 = .125 - .045941\rho^2 - .045941\rho^2$$

and $p_2 = p_4 = p_6 = p_8$;

$$p_3 = .062500 - .100980\rho + .022970\rho^2$$

and $p_3 = p_7$;

$$p_5 = .25 + .27564606.$$

Hence $\frac{1}{p_1} = 16 - 25.850958\rho + 35.886555\rho^2$

$$\frac{1}{p_2} = 8 + 2.940224\rho^2$$

$$\frac{1}{p_3} = 16 + 25.850958\rho + 35.886555\rho^2$$

$$\frac{1}{p_5} = 4 - 4.410336\rho^2$$

All the above being approximated to second power in ρ.

Thus for the standardized bivariate normal distribution, equalities exist between the elements of the matrix A such that

$A_{ij} = \frac{1}{p_1} \quad i \neq j \quad 1 \leq i \leq 8, \quad 1 \leq j \leq 8.$

$A_{11} = \frac{2}{p_1}$

$$A_{22} = A_{44} = A_{66} = A_{88} = \frac{1}{p_2} + \frac{1}{p_1}$$

$$A_{33} = A_{77} = \frac{1}{p_3} + \frac{1}{p_1}$$

$$A_{55} = \frac{1}{p_5} + \frac{1}{p_1}$$

and the symmetric matrix C has elements C_{ij} i,j = 1,2,3,4. :

$$C_{11} = \frac{6}{p_1} + \frac{2}{p_3} = 128 - 103.403834p + 287.092440p^2$$

$$C_{12} = C_{13} = \frac{5}{p_1} + \frac{1}{p_3} = 96 - 103.403834p + 215.319330p^2$$

$$C_{14} = C_{23} = \frac{5}{p_1} = 80 - 129.254792p + 179.432775p^2$$

$$C_{22} = C_{33} = \frac{5}{p_1} + \frac{2}{p_2} + \frac{1}{p_3} = 112 - 103.403834p + 220.553576p^2$$

$$C_{24} = C_{34} = \frac{5}{p_1} + \frac{1}{p_2} = 88 - 129.254792p + 182.372999p^2$$

$$C_{44} = \frac{5}{p_1} + \frac{2}{p_2} + \frac{1}{p_5} = 100 - 129.254792p + 180.902886p^2$$

and the inverse of C is the matrix C^{-1} with elements, denoted C^{ij}, given by:

$$C^{11} = .034856 + .025536p - 3.244710p^2$$

$$C^{12} = C^{13} = - .024038 - .024640p + 10.186090p^2$$

$$C^{14} = .014423 + .024490p - 6.816902p^2$$

$$C^{33} = C^{22} = .045673 + .001188p - 1.194586p^2$$

$$C^{23} = .014423 + .026433p - 7.540494p^2$$

$$C^{24} = C^{34} - .033654 - .001486p + 2.382654p^2$$

$$C^{44} = .057692 - .010764p + 2.245257p^2$$

McMASTER UNIVERSITY LIBRARY

$$B_{11} = B_{12} = B_{31} = \frac{2}{p_1} + \frac{1}{p_3} = 48 - 25.850958p + 107.659666p^2$$

$$B_{13} = B_{14} = B_{21} = B_{22} = B_{32} = B_{34} = B_{41} = B_{43} = \frac{2}{p_1}$$

$$= 32 - 51.701917p + 71.773110p^2$$

$$B_{23} = B_{24} = B_{42} = B_{44} = \frac{2}{p_1} + \frac{1}{p_2} = 40 - 51.701917p + 74.713334p^2$$

Now to find the marginal joint distribution of $U_1\ U_2\ U_3\ U_4$ one integrates the joint distribution; (which turns out to be just a multiple of s_2) of the U's and z_1, z_2, z_4, z_5 over the (z_1, z_2, z_4, z_5) region to obtain

$$f(U_1, U_2, U_3, U_4) = k \iiiint_{-\infty}^{\infty} \exp.- \frac{1}{2}\, Q_1(U)+Q_2(z)-2L(z))\ dz_1 dz_2 dz_4 dz_5$$

where k is the normalizing constant.

$$= k \exp- \frac{1}{2}\left[Q_1(U) + U'BC^{-1}B^{1}U\right] \qquad (3.10)$$

$$= k \exp- \frac{1}{2} \sum_{i}^{4} \sum_{=1}^{4} O_{ij}\ U_i U_j \qquad (3.11)$$

The symmetric matrix $D = BC^{-1}B'$ has elements d_{ij} given by

$$d_{11} = d_{33} = 23.076795 + 2.410104p + 37248.059452p^2$$

$$d_{12} = d_{34} = 11.076822 - 19.923340p + 26681.924854p^2$$

$$d_{13} = 15.076795 - 16.978077p + 35587.620459p^2$$

$$d_{14} = d_{23} = 15.076822 - 19.215344p + 27498.425898p^2$$

$$d_{22} = d_{44} = 16.074844 - 21.452619p + 21073.745121p^2$$

$$d_{24} = 15.073771 - 19.832951p + 20666.131167p^2.$$

and the (θ_{ij}) matrix has elements θ_{ij}, $1 \leq i \leq 4$, $1 \leq j \leq 4$:

$$\theta_{11} = \theta_{33} = 55.076795 + 2.410104p + 37319.832562p^2$$

$$\theta_{12} = \theta_{34} = 27.076822 - 48.297655p + 26716.815409p^2$$

$$\theta_{13} = 31.076795 - 42.829036p + 35623.507013p^2$$

$$\theta_{14} = \theta_{23} = 31.076822 - 76.143124p + 27534.312453p^2$$

$$\theta_{22} = \theta_{44} = 40.076844 - 47.303577p + 21112.571900p^2$$

$$\theta_{24} = 31.073771 - 45.683909p + 20702.017721p^2.$$

and $(\theta_{ij})^{-1} = (\theta^{ij})$ has elements:

$$\theta^{22} = \theta^{44} = .075653 - .049438p + 23.215799p^2$$

and $$\theta^{24} = .044310 + .036826p + 6.295394p^2.$$

Thus with $$U_1 = .\sqrt{4n}\left[-\frac{1}{4} + \frac{1}{2\pi}\ \bar{y}_3\right]$$

$$U_2 = .\sqrt{4n}\left[\frac{1}{2} - \frac{1}{2\pi}(\bar{y}_3 - \bar{y}_1)\right]$$

$$U_3 = .\sqrt{4n}\left[-\frac{1}{4} + \frac{1}{2\pi}\ \bar{x}_3\right]$$

$$U_4 = .\sqrt{4n}\left[\frac{1}{2} - \frac{1}{2\pi}(\bar{x}_3 - \bar{x}_1)\right]$$

U_i's follow the multivariate normal distribution with covariance matrix $(\theta_{ij})^{-1} = (\theta^{ij})$ and zero means.

Page 86 of Morrison's book 7 records the fact that if the joint distribution of a set of elements, written as the coordinates of a vector $\underline{x}$ is multivariate normal with mean vector $\underline{\mu}$ and covariance matrix $\sum$ then any linear transformation e.g. $\underline{y} = A\underline{x}$ gives a vector whose components again follow the multivariate normal distribution but with mean vector $A\underline{\mu}$ and covariance matrix $A' \sum A$.

Thus using matrix $A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ one verifies that U_2, U_4 are jointly distributed as a bivariate normal distribution with zero means and covariance matrix $\begin{pmatrix} \theta^{22} & \theta^{24} \\ \theta^{24} & \theta^{44} \end{pmatrix}$

Hence $(\bar{y}_3-\bar{y}_1)$ and $(\bar{x}_3-\bar{x}_1)$ are jointly distributed as the bivariate normal with means $\sqrt{\frac{\pi}{2}}$ and variances $\frac{\pi\sigma^{22}}{2n} = \frac{\pi\sigma^{44}}{2n}$ and covariance $\frac{\theta^{24}}{4n}$.

$$\therefore \text{Exp. } (\bar{y}_3-\bar{y}_1)(\bar{x}_3-\bar{x}_1) = \frac{\pi\sigma^{24}}{2n} + \frac{\pi}{2}$$

$$\text{Exp } (y_3-y_1)(\bar{x}_3-x_1)^2 = \text{Exp}\left\{\left(\sqrt{\frac{\pi}{2}} - \sqrt{\frac{\pi}{2n}}\ U_2\right)^2\left(\sqrt{\frac{\pi}{2}} - \sqrt{\frac{\pi}{\sqrt{2n}}}\ U_4\right)^2\right\}$$

$$= \text{Exp}\left\{\frac{\pi^2}{4} - \frac{\pi^2}{2\sqrt{n}}(U_2+U_4) + \frac{\pi^2}{4n}(U_2^2+U_4^2) + \frac{\pi^2}{n}U_2U_4\right.$$

$$- \frac{\pi^2}{2n\sqrt{n}} \quad (U_2U_4^2 + U_4U_2^2) + \frac{\pi^2}{4n^2} U_2^2U_4^2 \Big\}$$

$$= \frac{\pi^2}{4} + \frac{\pi^2}{4n}(\theta^{22} + \theta^{44}) + \frac{\pi^2}{n}\theta^{24} + \frac{\pi^2}{4n^2}\left[(\theta^{22})^2 + (2\theta^{24})^2\right]$$

Whence variance of $(\bar{y}_3-\bar{y}_1)(\bar{x}_3-\bar{x}_1) = \text{Exp.}\Big\{(y_3-y_1)^2(x_3-x_1)$

$$- E(\bar{y}_3-\bar{y}_1)(\bar{x}_3-\bar{x}_1)\Big\}^2$$

$$= \frac{\pi^2}{4n}\left\{2(\theta^{22} + \theta^{24}) + \frac{1}{n}\left[(\theta^{22})^2 + (\theta^{24})^2\right]\right\}$$

and the efficiency of the linear function of $(\bar{y}_3-\bar{y}_1)(\bar{x}_3-\bar{x}_1)$ used to estimate p from a sample of size (4n-2) may be ascertained to be

$$\frac{1}{4n+2} \cdot \frac{1-p^2}{1+p^2} \cdot \left(\frac{.036826}{n}\right) \cdot 4n \cdot \left[2\theta^{22} + 2\theta^{24} + \frac{1}{n}\left[(\theta^{22})^2 + (\theta^{24})^2\right]\right]^{-1}$$

In like manner one may consider the statistic $(\bar{y}_3-\bar{x}_1)(\bar{x}_3-\bar{y}_1)$ and possibly linear combinations of these could yield a more efficient statistic for estimating p.

BIBLIOGRAPHY

[1] S. K. Banerjee (1964) - Contributions to order Statistics - Ph. D. Thesis submitted to the University of Toronto

[2] Allen T. Craig - On the distribution of certain Statistics: American Journal of Mathematics, Vol. 54 (1932) p.p. 353 - 366

[3] R. A. Fisher and L. H. C. Tippett - Limiting forms of the frequency distribution of the largest and smallest member of a sample: Proceedings of Cambridge Philosophical Society, Vol. 24 (1928) p.p. 180 - 190

[4] H. Freeman - Introduction to Statistical Inference (1963) Addison Wesley Publishing Co. Inc. p. 192

[5] Grumbel - "The return of flood flows" - Annals of Mathematical Statistics, Vol. 19 (1941) p.p. 163 - 190

[6] Kendall and Stuart - Advanced Statistics, Vol. 1 New York Hafner Publishing Co. (1967)

[7] Donald F. Morrison (1967) - Multivariate Statistical Methods McGraw - Hill Series in Probobility and Statistics

[8] F. Mosteller - "On some useful 'inefficient' statistics", Annals of Mathematical Statistics, Vol. 17 (1946) p.p. 377-408

[9] National Bureau of Standards - Applied Mathematics Series, NO. 23

[10] National Bureau of Standards - Applied Mathematics Series, NO. 50

[11] K. Pearson - "Note on Frances Galton's problem" Biometrika, Vol. 1 (1902) p.p. 390-399

[12] William R. Thomson - "On confidence ranges for the median and other expectation distributions for populations of unknown distribution function. Annals of Mathematical Statistics, Vol. 7 (1936) p.p. 122 - 128

[13] L. H. C. Tippett - "On the extreme individuals and the range of samples taken from a normal population. Biometrika, Vol. 17 (1925) p.p. 364 - 387

[14] S. S. Wilks - Mathematical Statistics (1962) New York, John Wiley and Sons Inc.

[15] S. S. Wilks - "Order Statistics"
- Bulletin of American Mathematical Society, Vol. 54 (1948) p.p. 6 - 50

[16] S. S. Wilks - "Non-parametric Statistical Inference"
- Probability and Statistics: The Harald Cramer Vab. (ULF Grenander Ed.)
- Almquist and Wiksell, Stockholm p.p. 331-354

[17] S. S. Wilks - Collected Papers: Contributions to Mathematical Statistics
- Edited by Professor T. W. Anderson John Wiley and Sons Inc. (1967)