This research was supported by the U. S. Army Research Office (Durham) Grant No. DA-ARO(D)-31-124-G-746.

SOME CONTRIBUTIONS TO ORDER STATISTICS

by

Prakash Chandra Joshi

University of North Carolina

Institute of Statistics Mimeo Series No. 623

May 1969

SOME CONTRIBUTIONS TO ORDER STATISTICS

by

Prakash Chandra Joshi

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics

Chapel Hill

1969

Approved by:

_____

Adviser

PRAKASH CHANDRA JOSHI. Some Contributions to Order Statistics. (Under the direction of HERBERT A. DAVID).

Three different problems in order statistics are considered in this dissertation.

The first problem deals with the recurrence relations between moments and other functions of order statistics. It is shown that recurrence relations valid for independent and identically distributed random variables continue to hold for exchangeable variables.

In the second problem a method, based upon orthogonal polynomials, for obtaining bounds and approximations for the moments of order statistics is given. These bounds and approximations depend on the distribution function only through certain moments of order statistics in small samples. It is shown that for the Cauchy distribution bounds and approximations of all finite moments can be obtained.

Finally, the problem of detecting a single outlier in a fixed effects linear regression model is considered in some detail. The various cases considered are: (i) known variance, (ii) external studentization and (iii) pooled studentization. In each case, one- and two-sided test statistics for detecting a single outlier are proposed. These statistics are maxima of suitably standardized or studentized weighted residuals. With the help of Bonferroni and other inequalities upper and lower limits for the true upper percentage points of the proposed statistics are developed and some tables are provided. Some measures of performance, appropriate for our purposes are also introduced and studied. Finally, a comparison between external and pooled studentization is made.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF TABLES

ACKNOWLEDGEMENTS

I acknowledge my sincere gratitude to Professor H. A. David, not only for proposing the problems discussed in this dissertation, but also for his guidance and encouragement throughout my stay in Chapel Hill. It has been a pleasure to work under him as a student and also as a research assistant.

I also wish to thank the other members of my examination committee: Professors I. M. Chakravarti, N. L. Johnson, P. K. Sen and N. M. Wigley. Thanks are also extended to other members of the faculty of the Departments of Statistics, Biostatistics and Mathematics who have contributed towards my graduate training. I am particularly indebted to Professor T. G. Donnelly for a stimulating course in computer programming and allowing me to use some of his CALL A COMPUTER programs in this dissertation.

For financial support, I sincerely thank the Department of Biostatistics and the U. S. Army Research Office, Durham. I also wish to thank the State Government of Uttar Pradesh for providing a partial travel grant for this purpose.

Finally, I wish to thank Mrs. Delores Gold for her excellent typing of the manuscript.

ABSTRACT

Three different problems in order statistics are considered in this dissertation.

The first problem deals with the recurrence relations between moments and other functions of order statistics. It is shown that recurrence relations valid for independent and identically distributed random variables continue to hold for exchangeable variables.

In the second problem a method, based upon orthogonal polynomials, for obtaining bounds and approximations for the moments of order statistics is given. These bounds and approximations depend on the distribution function only through certain moments of order statistics in small samples. It is shown that for the Cauchy distribution bounds and approximations of all finite moments can be obtained.

Finally, the problem of detecting a single outlier in a fixed effects linear regression model is considered in some detail. The various cases considered are: (i) known variance, (ii) external studentization and (iii) pooled studentization. In each case, one- and two-sided test statistics for detecting a single outlier are proposed. These statistics are maxima of suitably standardized or studentized weighted residuals. With the help of Bonferroni and other inequalities upper and lower limits for the true upper percentage points of the proposed statistics are developed and some tables are provided. Some measures of performance, appropriate for our purposes are also introduced and studied. Finally, a comparison between external and pooled studentization is made.

CHAPTER I

INTRODUCTION AND SUMMARY

1.1. Scope

In this dissertation some well known problems in the theory of order statistics are considered. Throughout this work, we shall assume that the order statistics are obtained by re-arranging in non-decreasing order of magnitude variates having common marginal c.d.f.

In all, three different problems, along with some applications, are treated. These are:

1. order statistics for exchangeable variates,
2. bounds and approximations for the moments of order statistics and
3. detection of outliers in a linear regression model.

In each case, some existing results are generalized and, where necessary, new concepts are introduced. Out of these three problems, the third has been considered in greatest detail and constitutes the bulk of the dissertation.

1.2. Recurrence relations for order statistics

Recurrence relations between the moments and other functions of order statistics have been derived by many authors, usually on the assumption that the random variables $X_1, X_2, \ldots, X_n$ are independent continuous variates with common marginal c.d.f. $P(x)$. The simplest result is for $r=1,2,\ldots,n-1$,

$$(1.2.1) \qquad n\,\mu_{r:n-1} = r\,\mu_{r+1:n} + (n-r)\mu_{r:n},$$

where $\mu_{r:n}$ is the expected value of r^{th} order statistic in a sample of size n. In a recent paper, Young [43] shows that (1.2.1) and other results deducible from (1.2.1) continue to hold when the X_i are exchangeable variates, i.e., when $Pr(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$ is symmetric in $x_1, x_2, \ldots, x_n$.

In Chapter II, we give a simplified version of the proof given by Young and another simple probabilistic argument which establishes (1.2.1) and multivariate generalizations for exchangeable variates. Some applications of the results are also included.

1.3. Bounds and approximations for the moments of order statistics

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a continuous distribution with c.d.f. P(x). Let

$$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$$

be the corresponding order statistics.

The problem of finding bounds and approximations for the moments of order statistics has drawn considerable attention in the literature. One of the early papers is due to Plackett [31], who showed that there is a universal upper bound for $E(X_{n:n}-X_{1:n})/\sigma$, where σ is the standard deviation of X. Later work on these lines is by Moriguti [28], Gumbel [18], Hartley and David [21] and finally by Sugiura [42]. All these authors with the exception of Sugiura only give the bounds, while Sugiura gives both bounds and approximations.

In Chapter III, we generalize Sugiura's method and show that even with less stringent conditions, one or more different sequences of bounds and approximations for all finite moments can be obtained. These bounds and approximations depend on the distribution function only through certain moments of order statistics in small samples. Further, it has been shown that for the Cauchy distribution bounds and approximations of all finite moments can be obtained. Some numerical calculations for normal and Cauchy distributions are also given.

1.4. Outliers in regression models

1.4.1. The problem of outlier detection

The remainder of this dissertation has been devoted to the problem of detecting outliers in fixed effects additive regression models. The problem can be formulated as follows:

A sample of size n has been observed on a variable Y which has a linear regression on a known set of m variables $X_1, X_2, \ldots, X_m$. The usual normal theory model can be described by

$$\underset{\sim}{y} \overset{d}{=} N(X'\underset{\sim}{\beta}, \sigma^2 I), \tag{1.4.1}$$

where $\underset{1\times n}{y'} = (y_1, y_2, \ldots, y_n)$ is the observation vector, $\underset{m\times n}{X}$ is a known matrix of rank $m(<n)$, $\beta_1, \beta_2, \ldots, \beta_m$, σ^2 are unknown parameters and I is the identity matrix of order n. Also the symbol "$\overset{d}{=}$" stands for "is distributed according to".

For simplicity, we shall neglect the possibility of having any "errors" arising from the entries of the design matrix X. Among the n observations, one or more may show some sign of deviation from the assumed regression or may have a different variability or both. Such

observations are generally termed outliers and can arise for various reasons (see e.g. Kruskal [26]).

It is clear that the inclusion of such observations in any analysis may yield quite erroneous conclusions. Moreover, at times, these observations may themselves be of interest. Therefore, the main problem is to develop some suitable test procedures which can be used to isolate such observations. We shall restrict our attention to this aspect of the outlier problem.

In practice, we do not usually know the total number of outliers in any given data set. This makes the problem of detecting outliers much more difficult. In the special case, when we have a sample from a $N(\mu,\sigma^2)$ parent, the problem has been studied by many authors (see e.g. Chew [6], who summarizes these and other results). The outlying observations are usually assumed to differ in mean or in variance from the other observations. Numerous test procedures, mostly depending on a preassigned number of outliers, have been studied in this case. The general problem of a single outlier in regression models has been considered by Srikantan [40].

In the present study, we shall assume that at most one observation -- we do not know which one -- is an outlier. Moreover, this observation is assumed to differ from the assumed regression only in mean. In this case, certain test statistics have been proposed by Srikantan. For these test statistics, he has obtained the nominal percentage points, which control the error of the first kind at a level not exceeding the specified one. He has also shown that under certain conditions, which in general hold in small samples, the nominal

percentage points coincide with the true percentage points. Here, we consider some additional test procedures, depending on the knowledge available about σ^2 and study their performance. This is summarized in the following section.

1.4.2. Notations and summary

Let

$$\underset{n\times n}{\Lambda} = I - X'(XX')^{-1}X = ((\lambda_{ij})),$$

$$\underline{e} = \Lambda \underline{y} = \text{residual vector},$$

$$S^2 = \underline{e}'\underline{e} = \text{error sum of squares},$$

$$\rho_{ij} = \frac{\lambda_{ij}}{(\lambda_{ii}\lambda_{jj})^{\frac{1}{2}}} = \text{correlation between } e_i \text{ and } e_j,$$

$$\rho_{min} = \underset{i\neq j}{\text{Min}} \ \rho_{ij}$$

and

$$\rho_{max} = \underset{i\neq j}{\text{Max}} \ \rho_{ij}.$$

Throughout this study we shall assume that

$$-1 < \rho_{min} \leq \rho_{max} < 1.$$

In Chapter IV, we propose one- and two-sided test statistics for detecting a single outlier in situations where 1) σ^2 is known and 2) σ^2 is unknown, but an independent root mean square estimator, s_ν, of σ based on ν degrees of freedom is available. The latter case has been further divided into two categories which are termed external studentization and pooled studentization. The one-sided test statistics for σ^2 known, external studentization and pooled studentization are U_1, U_2 and U_3

respectively, where

$$U_1 = \max_i b_i, \quad b_i = \frac{e_i}{\lambda_{ii}^{\frac{1}{2}}} / \sigma,$$

$$U_2 = \max_i t_i, \quad t_i = \frac{e_i}{\lambda_{ii}^{\frac{1}{2}}} / s_\nu,$$

$$U_3 = \max_i w_i, \quad w_i = \frac{e_i}{\lambda_{ii}^{\frac{1}{2}}} / (S^2+\nu s_\nu^2)^{\frac{1}{2}} \tag{1.4.2}$$

and the maximum is taken over the set $\{1,2,\ldots,n\}$. The corresponding two-sided statistics are denoted by V_1, V_2 and V_3 respectively, where $V_1 = \max_i |b_i|$ etc.

The first two Bonferroni inequalities are suggested to get upper and lower limits for the true upper $100\alpha\%$ points of these statistics. These are improved in later chapters for some special cases.

To study the performance of the proposed statistics, we assume the null hypothesis, H_o, that there are no outliers and that under H_o our model is given by equation (1.4.1). The alternative hypothesis H_a is the union of n mutually exclusive hypotheses $H_1, H_2, \ldots, H_n$ and under H_k

$$\underline{y} \overset{d}{=} N(\cdot, \sigma^2 I),$$

where

$$E(\underline{Y}|H_k) = X'\underline{\beta} + \underline{\varepsilon}_k\theta,$$

where $\underline{\varepsilon}_k'$ is an n-vector with k^{th} component 1 and all other components 0.

Now under H_o, all the proposed test statistics are of the form $Z = \max\{z_i: \ i=1,2,\ldots,n\}$, where $z_1, z_2, \ldots, z_n$ are identically distributed random variables. Let z_α be the true upper $100\alpha\%$ point of Z and for $k=1,2,\ldots,n$ let

P_k = Probability that z_k is significantly large when H_k is true

$= \Pr(z_k > z_\alpha | H_k)$,

Q_k = Probability of rejecting H_o when H_k is true

$= \Pr(Z > z_\alpha | H_k)$.

The measures of performance introduced are:

$$P_a = \min_k P_k , \qquad P_b = \frac{1}{n} \sum_{k=1}^{n} P_k,$$

$$Q_a = \min_k Q_k , \qquad Q_b = \frac{1}{n} \sum_{k=1}^{n} Q_k.$$

In some special cases, these reduce to the measures considered by David and Paulson [10].

Finally, three examples of regression models, viz., 1) sample from a $N(\mu,\sigma^2)$ parent, 2) one-way layout and 3) two-way layout are briefly mentioned. These are considered in later chapters as an application of the theory developed.

In Chapter V, the statistics U_1 and V_1 are treated in detail. It is shown that one can always obtain better upper limits for the true upper $100\alpha\%$ points of V_1 than those obtained by the first Bonferroni inequality. Some interesting results for U_1 in the case $\rho_{max} \leq 0$ are also obtained. Among the measures of performance, it was observed that for both U_1 and V_1, P_a depends on $\lambda_{(1)}$, where $\lambda_{(1)} = \min_k \lambda_{kk}$. Moreover, P_a is a poor measure if any one of the λ_{kk} is much smaller than the others and in such cases P_b might be worth computing. Upper and lower limits for the true upper percentage points of U_1 and V_1 for the two-way layout are tabulated and the measure P_a has been studied for this case.

The statistics U_2 and V_2 are investigated in Chapter VI. A number of results in this case are analogous to that of Chapter V. For the measures of performance we have mainly considered P_a and provided a table for computing P_a for the two-sided test statistic.

In Chapter VII, a number of distribution theory results related to U_3 and V_3 are obtained. Findings of this chapter generalize the results due to Doornbos et al. (see e.g. Doornbos [11]) and Srikantan [40]. The joint distribution of w_1 and w_2, where w_i is given at (1.4.2) is derived and an expression useful for evaluating $\Pr(w_1>c_1,\ w_2>c_2)$ is obtained. It is shown that if $\rho_{12} \leq 0$, then

$$\Pr(w_1 \leq c_1,\ w_2 \leq c_2) \geq \prod_{i=1}^{2} \Pr(w_i \leq c_i),$$

provided both c_1 and c_2 are of the same sign.

Following a geometrical argument, it is shown that for some small values of n and ν, the upper limits for the true upper percentage points for both U_3 and V_3 coincide with the true upper percentage points. A comparison between U_2 and U_3 using P_k is also included. It was found that U_3 has a definite edge over U_2, when we use the upper limits in the expressions for P_k.

CHAPTER II

RECURRENCE RELATIONS BETWEEN MOMENTS OF ORDER STATISTICS FOR EXCHANGEABLE VARIATES

2.1. Introduction

Let $X_{i:n}$ $(i=1,2,\ldots,n)$ be the order statistics obtained by rearranging in non-decreasing order of magnitude the variates X_i having common marginal c.d.f. $P(x)$. Denote by $F_{i:n}(x)$ and $\mu_{i:n}$ the c.d.f. and expected value of $X_{i:n}$ respectively. Recurrence relations for moments and other functions of the $X_{i:n}$ have been derived by many authors, usually on the assumption that the X_i are independent continuous variates. The most basic of these relations states that for $r=1,2,\ldots,n-1$,

$$n\,\mu_{r:n-1} = r\,\mu_{r+1:n} + (n-r)\mu_{r:n}. \tag{2.1.1}$$

In a recent paper, Young [43] shows (in effect) that (2.1.1) and hence results deducible from (2.1.1) continue to hold if the X_i are exchangeable, continuous or discrete variates, i.e., if $\Pr(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$ is symmetric in $x_1, x_2, \ldots, x_n$. In Section 2.2, we give a simplified version of the proof given by Young and in Section 2.3, another simple probabilistic argument which establishes (2.1.1) and multivariate generalizations for exchangeable variates is given. In Section 2.4, some applications of the results are mentioned.

2.2. Recurrence relations

Let A_i $(i=1,2,\ldots,n)$ be the event $X_i \leq x$. Define

$$S_m = \sum \Pr(A_{i_1}, A_{i_2}, \ldots, A_{i_m}), \tag{2.2.1}$$

the summation extending over all $\binom{n}{m}$ sets of integers $1 \leq i_1 < i_2 < \ldots < i_m \leq n$. Then for $r=1,2,\ldots,n$,

$$\begin{aligned} F_{r:n}(x) &= \Pr(X_{r:n} \leq x) \\ &= \Pr(\text{at least } r \text{ of the } X_i\text{'s are} \leq x) \\ &= \sum_{m=r}^{n} (-1)^{m-r} \binom{m-1}{r-1} S_m. \end{aligned} \tag{2.2.2}$$

The last equality follows by using a well known theorem for the realization of at least r out of n events (see e.g. Feller [14]).

Equation (2.2.2) is valid for any set of n random variables. In particular, if the X_i are exchangeable variates, then from (2.2.1)

$$\begin{aligned} S_m &= \binom{n}{m} \Pr(X_1 \leq x, X_2 \leq x, \ldots, X_m \leq x) \\ &= \binom{n}{m} \Pr(X_{m:m} \leq x) \\ &= \binom{n}{m} F_{m:m}(x). \end{aligned}$$

Substituting in (2.2.2), we get

$$F_{r:n}(x) = \sum_{m=r}^{n} (-1)^{m-r} \binom{m-1}{r-1} \binom{n}{m} F_{m:m}(x), \tag{2.2.3}$$

a relation linking the c.d.f. of $X_{r:n}$ with the c.d.f.'s of the maximum in samples of $r, r+1, \ldots, n$. Differentiating or differencing, multiplying

by e^{itx} and integrating or summing, we obtain the same relation between p.d.f.'s, characteristic functions, and hence raw moments. Equation (2.2.3) has been proved by Young [43] by a similar but rather complicated method. (See equation (11) of Young's paper).

With the help of (2.2.3) we can derive what may be called the basic recurrence relation for order statistics:

$$(2.2.4) \qquad n\, F_{r:n-1}(x) = r\, F_{r+1:n}(x) + (n-r)F_{r:n}(x).$$

This formula, usually stated in terms of moments, follows on applying (2.2.3) to each term of (2.2.4).

We have here reversed the usual sequence whereby (2.2.3) and indeed some more general related results are deduced by repeated application of (2.2.4). It may also be noted that for identically distributed standardized multi-normal variates with equal correlation ρ, (2.2.4) for arbitrary ρ follows readily from (2.2.4) for $\rho=0$ in view of the representations of X_i by (Owen and Steck [30])

$$X_i = \rho^{\frac{1}{2}}Y_0 + (1-\rho)^{\frac{1}{2}}Y_i \qquad (\rho \geq 0;\ i=1,2,\ldots,n),$$

where $Y_0, Y_1, \ldots, Y_n$ are independent N(0,1) variates and

$$X_i = (-\rho)^{\frac{1}{2}}Z_0 + (1-\rho)^{\frac{1}{2}}Z_i \qquad (-1/(n-1) \leq \rho < 0;\ i=1,2,\ldots,n),$$

where $Z_1, Z_2, \ldots, Z_n$ are independent N(0,1) variates, Z_0 is N(0,1) and

$$E(Z_0 Z_i) = -(-\rho)^{\frac{1}{2}}/(1-\rho)^{\frac{1}{2}}.$$

If in addition to exchangeability multivariate symmetry (say about zero) also holds, i.e.

$$\Pr(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) = \Pr(X_1 \geq -x_1, X_2 \geq -x_2, \ldots, X_n \geq -x_n),$$

then as for independent symmetric variates

$$\Pr(X_{r:n} \leq x) = \Pr(X_{n-r+1:n} \geq -x). \tag{2.2.5}$$

To prove this, we only need to note that in view of multivariate symmetry, the joint distributions of

$$(X_{1:n}, X_{2:n}, \ldots, X_{n:n})$$

and

$$(-X_{n:n}, -X_{n-1:n}, \ldots, -X_{1:n})$$

coincide. Hence for any x

$$\Pr(X_{r:n} \leq x) = \Pr(-X_{n-r+1:n} \leq x)$$

and (2.2.5) follows.

For continuous variates we have, in particular,

$$F_{n:n}(x) = 1 - F_{1:n}(-x)$$

$$= 1 - \sum_{m=1}^{n} (-1)^{m-1} \binom{n}{m} F_{m:m}(-x) \qquad \text{(by (2.2.3))},$$

a result obtained by Steck [41] for the equi-correlated multinormal variates.

2.3. Direct proof and generalizations

Of the n variates X_i drop one at random and let $Y_{i:n-1}$ (i=1,2,...,n-1) denote the i^{th} order statistic in the reduced set of n-1 exchangeable variates. If $X_{i:n}$ is dropped (i=1,2,...,r) the r^{th}

largest variate in the set of n-1 was the $(r+1)^{th}$ largest out of n, i.e.

(A) $$Y_{r:n-1} = X_{r+1:n}.$$

Likewise, if $X_{i:n}$ is dropped (i=r+1,r+2,...,n), then

(B) $$Y_{r:n-1} = X_{r:n}.$$

Since (A) and (B) have respective probabilities r/n and (n-r)/n, it follows that for any x

(2.3.1) $$\Pr(Y_{r:n-1} \leq x) = \frac{r}{n} \Pr(X_{r+1:n} \leq x) + \frac{n-r}{n} \Pr(X_{r:n} \leq x),$$

that is

$$n\, F_{r:n-1}(x) = r\, F_{r+1:n}(x) + (n-r) F_{r:n}(x),$$

which is equation (2.2.4).

The above argument is readily generalized to the joint c.d.f. of two or more order statistics. Let $F_{r,s:n}(x,y)$ denote the joint c.d.f. of $X_{r:n}$ and $X_{s:n}$ $(1 \leq r < s \leq n;\ x \leq y)$ and let $\mu_{r,s:n} = E(X_{r:n} X_{s:n})$. Then corresponding to (2.3.1), we now have for any x,y $(x \leq y)$

$$\begin{aligned}
&\Pr(Y_{r:n-1} \leq x,\ Y_{s:n-1} \leq y) \\
&\quad = \frac{r}{n} \Pr(X_{r+1:n} \leq x,\ X_{s+1:n} \leq y) \\
&\quad + \frac{s-r}{n} \Pr(X_{r:n} \leq x,\ X_{s+1:n} \leq y) \\
&\quad + \frac{n-s}{n} \Pr(X_{r:n} \leq x,\ X_{s:n} \leq y).
\end{aligned}$$

This can be rewritten as

$$(2.3.2)\quad n\,F_{r,s:n-1}(x,y) = r\,F_{r+1,s+1:n}(x,y)+(s-r)F_{r,s+1:n}(x,y) +(n-s)F_{r,s:n}(x,y).$$

As in Section 2.2, this result can be converted into one linking the corresponding product-moments of any order, to give in particular,

$$(2.3.3)\quad n\,\mu_{r,s:n-1} = r\,\mu_{r+1,s+1:n}+(s-r)\mu_{r,s+1:n} + (n-s)\mu_{r,s:n}.$$

(2.3.3) has been established by Govindarajulu [16] for independent identically distributed continuous variates. For the equi-correlated multinormal case (with common marginal c.d.f.) (2.3.3) may also be proved with the help of expressions for the moments of order statistics given by Owen and Steck [30].

2.4. Some applications

For numerical work formula (2.2.3) requires tables of the c.d.f. of largest order statistic. If X_i (i=1,2,...,n) are multivariate normal with means zero, variances unity, and common correlation coefficient ρ, then Gupta [19] tabulates $F_{n:n}(x)$ for n=1(1)12 and several values of ρ.

An important special case of (2.2.3) gives the c.d.f. of $X_{n-1:n}$ as

$$(2.4.1)\quad F_{n-1:n}(x) = n\,F_{n-1:n-1}(x)-(n-1)F_{n:n}(x)$$

As pointed out by Fisher [15] in connection with harmonic analysis a test of the second largest variate $X_{n-1:n}$ becomes of special interest when the test based on $X_{n:n}$ is inconclusive, that is, close to the chosen level of significance.

As an application of (2.4.1) consider the problem of testing n "treatment" means against a "control" mean (Dunnett [12]). Let Z_{ij} and Z_{oh} $(i=1,2,\ldots,n;\ j=1,2,\ldots,k;\ h=1,2,\ldots,\ell)$ be mutually independent normal variates, Z_{ij} and Z_{oh} being respectively $N(\mu_i,\sigma^2)$ and $N(\mu_o,\sigma^2)$, with σ^2 assumed known. In order to test simultaneously whether any of the treatment means $\bar{Z}_i$ differ from the control mean $\bar{Z}_o$ we may use the statistic

$$X_{n:n} = \operatorname{Max}_i X_i,$$

where

$$X_i = \frac{\bar{Z}_i - \bar{Z}_o}{\sigma\left(\frac{1}{k}+\frac{1}{\ell}\right)^{\frac{1}{2}}}, \qquad i=1,2,\ldots,n,$$

$$\bar{Z}_i = \frac{1}{k}\sum_{j=1}^{k} Z_{ij}, \qquad i=1,2,\ldots,n$$

and

$$\bar{Z}_o = \frac{1}{\ell}\sum_{h=1}^{\ell} Z_{oh}.$$

It is easy to show that the X_i are equi-correlated standard normal variates with correlation coefficient $\rho = k/(k+\ell)$ so that Gupta's tables may be used to obtain $F_{n-1:n}(x)$ and hence percentage points of $X_{n-1:n}$. For the case $k=\ell$, i.e. $\rho=\frac{1}{2}$, upper 5 and 1% points are given in Table 2.4.1.

Table 2.4.1. Upper 5 and 1% points of $X_{n-1:n}$, the second largest among n equi-correlated standard normal variates with correlation coefficient $\rho=.5$

n	5%	1%
2	1.100	1.713
3	1.400	1.981
4	1.569	2.134
5	1.685	2.242
6	1.773	2.324
7	1.843	2.390
8	1.901	2.443
9	1.950	2.490
10	1.993	2.532
11	2.031	2.569
12	2.065	2.597

CHAPTER III

BOUNDS AND APPROXIMATIONS FOR THE MOMENTS OF ORDER STATISTICS

3.1. Introduction

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a continuous distribution with c.d.f. P(x) and p.d.f. p(x). Let

$$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$$

be the corresponding order statistics. Denote $E(X_{r:n})$ by $\mu_{r:n}$.

Several authors (see e.g. Blom [5], David and Johnson [8], Sugiura [42]) have given methods of finding approximations for the moments of order statistics. The method due to Sugiura also gives the bounds and requires that the r.v. X has a finite variance. By a generalization of his method, it will be shown here that, even with less stringent conditions, one or more different sequences of bounds for all finite moments can be obtained. These bounds and approximations depend on the distribution function only through certain moments of order statistics in small samples. All such bounds, e.g. for $\mu_{r:n}$, are of the form

$$\mu_{r,n,t} \pm c_{r,n,t},$$

where $\mu_{r,n,t}$ is a (t+1)-term approximation to $\mu_{r:n}$ and $\mu_{r,n,t} \to \mu_{r:n}$, $c_{r,n,t} \to 0$ as $t \to \infty$.

3.2. Notations and some preliminary results concerning orthonormal functions

Let $\{\phi_k(u)\}$; $k=0,1,2,\ldots,\phi_o(u)=1$, be an orthonormal system over the closed interval $[0,1]$, i.e.

$$\int_0^1 \phi_m(u)\phi_n(u)du = \begin{cases} 1 & \text{if } m=n \\ 0 & \text{otherwise.} \end{cases}$$

Let $f(u)$ and $g(u)$ be square-integrable functions over $[0,1]$. Put

$$a_k = \int_0^1 f(u)\phi_k(u)du, \tag{3.2.1}$$

$$b_k = \int_0^1 g(u)\phi_k(u)du. \tag{3.2.2}$$

Thus a_k and b_k are the Fourier coefficients of $f(u)$ and $g(u)$ relative to $\phi_k(u)$ respectively. We then have the following lemma of Sugiura [42].

Lemma 3.2.1. For any integral $t\geq 0$

$$\left|\int_0^1 f(u)g(u)du - \sum_{k=0}^{t} a_k b_k\right| \leq \left[\int_0^1 f^2(u)du - \sum_{k=0}^{t} a_k^2\right]^{\frac{1}{2}} \cdot \left[\int_0^1 g^2(u)du - \sum_{k=0}^{t} b_k^2\right]^{\frac{1}{2}}, \tag{3.2.3}$$

where the equality holds if and only if

$$f(u) - \sum_{k=0}^{t} a_k\phi_k(u) \propto g(u) - \sum_{k=0}^{t} b_k\phi_k(u).$$

Proof: Applying Schwarz inequality to

$$f(u) - \sum_{k=0}^{t} a_k \phi_k(u) \quad \text{and} \quad g(u) - \sum_{k=0}^{t} b_k \phi_k(u)$$

we have

$$\left| \int_0^1 (f(u) - \sum_{k=0}^{t} a_k \phi_k(u))(g(u) - \sum_{k=0}^{t} b_k \phi_k(u)) du \right|$$

$$\leq \left[\int_0^1 (f(u) - \sum_{k=0}^{t} a_k \phi_k(u))^2 du \right]^{\frac{1}{2}} \cdot \left[\int_0^1 (g(u) - \sum_{k=0}^{t} b_k \phi_k(u))^2 du \right]^{\frac{1}{2}}.$$

Multiplying and using (3.2.1), (3.2.2) and the orthonormality of functions $\{\phi_k(u)\}$, the lemma follows.

Remark 1. In Lemma 3.2.1 the members $\phi_k(u)$ $(k=0,1,\ldots,t)$ of the orthonormal system have been used to give (3.2.3). However, the same proof applies for any subset of $\{\phi_k(u)\}$, thus allowing us to select subsets which result in useful bounds.

Remark 2. $\sum_{k=0}^{t} a_k b_k$ provides an approximation to $\int_0^1 fg \, du$, an upper bound to the error involved being given by the R.H.S. of (3.2.3). If, further, $\{\phi_k\}$ constitute a complete orthonormal system in [0,1] (see e.g. [35]), then the R.H.S. of (3.2.3) tends to 0 as $t \to \infty$, and hence the approximation can be made as accurate as we please by choosing t large enough.

As an example of a complete orthonormal system in [0,1], we have the Legendre polynomials in [0,1] given by

$$L_k(u) = \frac{\sqrt{2k+1}}{k!} \frac{d^k}{du^k} u^k (u-1)^k; \quad k=0,1,2,\ldots. \tag{3.2.4}$$

3.3. Bounds and approximations

Let the r.v. X have a continuous c.d.f. P(x) and let $\mu_{r:n}$ be finite. Then by using the probability integral transformation $u = P(x)$ we can write

$$\mu_{r:n} = \frac{1}{B(r,n-r+1)} \int_0^1 x(u) u^{r-1} (1-u)^{n-r} du, \tag{3.3.1}$$

where x(u) denotes that x has been expressed as a function of u.

Theorem 3.3.1. Let $\phi_o=1, \phi_1, \phi_2, \ldots$ be any orthonormal system in [0,1] and let $E(X^2_{2p+1:2p+2q+1})$ be finite for some integral $p, q \geq 0$. Then for $r=1,2,\ldots,n$ and any integral $t \geq 0$

$$\left| \frac{B(p+r,q+n-r+1)}{B(r,n-r+1)} E(X_{p+r:p+q+n}) - \sum_{k=0}^{t} a_k b_k \right| \leq$$

$$[B(2p+1,2q+1)E(X^2_{2p+1:2p+2q+1}) - \sum_{k=0}^{t} a_k^2]^{\frac{1}{2}} \tag{3.3.2}$$

$$\cdot \left[\frac{B(2r-1,2n-2r+1)}{[B(r,n-r+1)]^2} - \sum_{k=0}^{t} b_k^2\right]^{\frac{1}{2}},$$

where a_k and b_k are given by (3.2.1) and (3.2.2) with

$$f(u) = x(u)\, u^p (1-u)^q, \tag{3.3.3}$$

$$g(u) = \frac{1}{B(r,n-r+1)} u^{r-1} (1-u)^{n-r}. \tag{3.3.4}$$

Proof: We have

$$\int_0^1 f^2(u) du = B(2p+1,2q+1)E(X^2_{2p+1:2p+2q+1}),$$

$$\int_0^1 g^2(u)du = \frac{B(2r-1,2n-2r+1)}{[B(r,n-r+1)]^2}$$

and

$$\int_0^1 f(u)g(u)du = \frac{B(p+r,q+n-r+1)}{B(r,n-r+1)} E(X_{p+r:p+q+n}).$$

Applying Lemma 3.2.1, the result follows.

It should be noted that for a given orthonormal system, a_k depends on the c.d.f. P(x) of the r.v. X, but not on r and n; while b_k does not depend on the P(x), but only on r and n. Theorem 3.3.1 then shows that the approximations and bounds for $\mu_{p+r:p+q+n}$ (r=1,2,...,n), which depend on P(x) can be obtained, provided $E(X^2_{2p+1:2p+2q+1})$ is finite. In terms of P(x), a sufficient condition for this is

$$\int_{-\infty}^{\infty} |x|^{\frac{2}{2p+1}} \cdot dP(x) < \infty$$

for $q \geq p \geq 0$ (Sen [37]).

It is clear that the same technique can be used to find the bounds and approximations for moments of any order. Thus, to get similar results for $E(X^s_{p+r:p+q+n})$, where s>0, we only require the square integrability of the function $[x(u)]^s u^p(1-u)^q$. This is equivalent to saying that $E(X^{2s}_{2p+1:2p+2q+1})$ must be finite.

We shall now turn our attention to the case where the distribution is known to be symmetric, say about x=0. This additional information leads to sharper bounds.

Theorem 3.3.2. Let the distribution of X be continuous and symmetric about x=0. Let $\phi_o=1$, $\phi_1,\phi_2,\ldots$ be any complete orthonormal system in [0,1] satisfying

$$\phi_k(u) = (-1)^k \phi_k(1-u), \quad k=1,2,\ldots. \tag{3.3.5}$$

If $E(X^2_{2m+1:4m+1})$ is finite for some integral $m\geq 0$, then for $r=1,2,\ldots,n$ and any integral $t\geq 0$

$$\begin{aligned} &\left|\frac{B(m+r,m+n-r+1)}{B(r,n-r+1)} E(X_{m+r:2m+n}) - \sum_{k=0}^{t} a_{2k+1}b_{2k+1}\right| \\ &\leq [B(2m+1,2m+1)E(X^2_{2m+1:4m+1}) - \sum_{k=0}^{t} a^2_{2k+1}]^{\frac{1}{2}} \\ &\quad \cdot [\frac{B(2r-1,2n-2r+1)-B(n,n)}{2[B(r,n-r+1)]^2} - \sum_{k=0}^{t} b^2_{2k+1}]^{\frac{1}{2}}, \end{aligned} \tag{3.3.6}$$

where a_k and b_k are as in Theorem 3.3.1, with p and q equal to m. If, further, $E(X^4_{2m+1:4m+1})$ is finite for some integral $m\geq 0$, then for $r=1,2,\ldots,n$ and any integral $t\geq 0$

$$\begin{aligned} &\left|\frac{B(m+r,m+n-r+1)}{B(r,n-r+1)} E(X^2_{m+r:2m+n}) - \sum_{k=0}^{t} a'_{2k}b_{2k}\right| \\ &\leq [B(2m+1,2m+1)E(X^4_{2m+1:4m+1}) - \sum_{k=0}^{t} a'^2_{2k}]^{\frac{1}{2}} \\ &\quad \cdot [\frac{B(2r-1,2n-2r+1)+B(n,n)}{2[B(r,n-r+1)]^2} - \sum_{k=0}^{t} b^2_{2k}]^{\frac{1}{2}}, \end{aligned} \tag{3.3.7}$$

where

$$a'_k = \int_0^1 x^2(u)\, u^m(1-u)^m \phi_k(u)du.$$

Proof: Take p=q=m in (3.3.3) and apply Lemma 3.2.1 with k=1,3,5,..., 2t+1,0,2,4,...., thus giving

$$\left|\frac{B(m+r,m+n-r+1)}{B(r,n-r+1)}\,E(X_{m+r:2m+n}) - \sum_{k=0}^{t} a_{2k+1}b_{2k+1} - \sum_{k=0}^{\infty} a_{2k}b_{2k}\right|$$

$$\leq \left[B(2m+1,2m+1)E(X^2_{2m+1:4m+1}) - \sum_{k=0}^{t} a^2_{2k+1} - \sum_{k=0}^{\infty} a^2_{2k}\right]^{\frac{1}{2}}$$

$$(3.3.8)\qquad \cdot\left[\frac{B(2r-1,2n-2r+1)}{[B(r,n-r+1)]^2} - \sum_{k=0}^{t} b^2_{2k+1} - \sum_{k=0}^{\infty} b^2_{2k}\right]^{\frac{1}{2}}.$$

Since the distribution of X is symmetric about x=0, the inverse function x(u) is odd and hence on using (3.3.5)

$$a_{2k} = \int_0^1 x(u)u^m(1-u)^m\phi_{2k}(u)du$$

$$= \int_0^1 x(1-v)v^m(1-v)^m\phi_{2k}(v)dv;\ v=1-u$$

$$= -\int_0^1 x(v)v^m(1-v)^m\phi_{2k}(v)dv = -a_{2k}.$$

So that

$$(3.3.9)\qquad a_{2k} = 0.$$

We now show that

$$(3.3.10)\qquad \sum_{k=0}^{\infty} b^2_{2k} = \frac{B(2r-1,2n-2r+1)+B(n,n)}{2[B(r,n-r+1)]^2}\ .$$

This result has been proved by Sugiura [42] for Legendre polynomials (3.2.4), for which (3.3.5) is satisfied. The proof given here follows on parallel lines. Define

$$(3.3.11)\qquad g^*(u) = \frac{1}{B(r,n-r+1)}\,u^{n-r}(1-u)^{r-1},\quad 0\leq u\leq 1.$$

Then on using (3.3.5)

$$b_{2k} = \int_0^1 \tfrac{1}{2}(g(u)+g^*(u))\phi_{2k}(u)du$$

$$= \int_0^1 h(u)\phi_{2k}(u)du, \tag{3.3.12}$$

where

$$h(u) = \tfrac{1}{2}(g(u)+g^*(u)). \tag{3.3.13}$$

Since $\{\phi_k\}$ is a complete orthonormal system in [0,1], hence

$$\int_0^1 h^2(u)du = \sum_{k=0}^{\infty} c_k^2 , \tag{3.3.14}$$

where c_k is the Fourier coefficient of h(u) relative to $\phi_k(u)$, i.e.

$$c_k = \int_0^1 h(u)\phi_k(u)du.$$

Now by (3.3.12) and (3.3.5) we see that

$$c_{2k} = b_{2k} \quad \text{and} \quad c_{2k+1} = 0.$$

Equation (3.3.14) then reduces to

$$\int_0^1 h^2(u)du = \sum_{k=0}^{\infty} b_{2k}^2 .$$

Therefore

$$\sum_{k=0}^{\infty} b_{2k}^2 = \tfrac{1}{4}\int_0^1 (g(u)+g^*(u))^2du$$

and (3.3.10) follows.

Use of equations (3.3.9) and (3.3.10) in (3.3.8) completes the proof of first part of the theorem.

To prove the second part, note that

$$a'_{2k+1} = 0 \tag{3.3.15}$$

and

$$\sum_{k=0}^{\infty} b^2_{2k+1} = \frac{B(2r-1,2n-2r+1)-B(n,n)}{2[B(r,n-r+1)]^2} .$$

An application of Lemma 3.2.1 with k=0,2,4,...,2t,1,3,5,.... then gives (3.3.7).

It is clear that similar results for any odd and even order moments can be obtained by applying the same techniques as in the proofs of (3.3.6) and (3.3.7) respectively.

3.4. Some applications

As an application of (3.3.6), let X have a Cauchy distribution,

$$p(x) = \frac{1}{\pi(1+x^2)} , \quad -\infty<x<\infty.$$

In this case Barnett [4] has shown that $E(X^s_{r:n})$ is finite for all $s<r\leq n-s$. In particular, this means that $E(X^2_{3:5})$ is finite and hence (3.3.6) is applicable with m=1. We thus get bounds for $E(X_{r+1:n+2})$, r=1,2,...,n, i.e., for all finite moments. However, one can take any integral $m\geq 1$. Thus for m=2 we get another sequence of bounds for $E(X_{r+2:n+4})$, r=1,2,...,n, i.e., for all finite moments except for the second smallest and second largest order statistics.

The case m=0 has been treated by Sugiura [42] who has also given bounds and approximations for a normal distribution for n=10, 20. Here we consider the case m=1 and use the Legendre polynomials (3.2.4) for which (3.3.5) is satisfied. In general, we can write

$$L_k(u) = \sum_{i=0}^{k} c_{k,i} u^i, \tag{3.4.1}$$

where $c_{k,i}$ are constants. For k=0,1,2 and 3 these are given below:

$$L_0(u) = 1,$$
$$L_1(u) = \sqrt{3}\ (2u-1),$$
$$L_2(u) = \sqrt{5}\ (6u^2-6u+1),$$
$$L_3(u) = \sqrt{7}\ (20u^3-30u^2+12u-1).$$

With m=1 and L_k in place of ϕ_k we have

$$a_k = \sum_{i=0}^{k} \frac{c_{k,i}}{(i+2)(i+3)} E(X_{i+2:i+3}), \tag{3.4.2}$$

$$b_k = \sum_{i=0}^{k} c_{k,i} \frac{r(r+1)....(r+i-1)}{(n+1)(n+2)...(n+i)} \tag{3.4.3}$$

and (3.3.6) reduces to

$$\left|\frac{r(n-r+1)}{(n+1)(n+2)} E(X_{r+1:n+2}) - \sum_{k=0}^{t} a_{2k+1} b_{2k+1}\right| \leq$$

$$\left[\frac{1}{30} E(X^2_{3:5}) - \sum_{k=0}^{t} a^2_{2k+1}\right]^{\frac{1}{2}} \tag{3.4.4}$$

$$\cdot \left[\frac{B(2r-1,2n-2r+1)-B(n,n)}{2[B(r,n-r+1)]^2} - \sum_{k=0}^{t} b^2_{2k+1}\right]^{\frac{1}{2}}.$$

Note that a_k is a linear function of the moments $\mu_{i+2:i+3}$ in small samples and can be evaluated by using tables of moments of order statistics for the r.v. X. In cases where no such tables are available, the moments could be evaluated by numerical integration.

The necessary moments for computation are extensively tabulated for the normal distribution in [36] and for the Cauchy distribution in [4] and [34]. Bounds and approximate values for n=8 and n=18, using (3.4.4), are given in Tables 3.4.1 and 3.4.2. t=0 in (3.4.4) gives the first bound and t=1 gives the second bound. These tables show that the approximations and bounds are remarkably good for Cauchy distribution for all r. For the normal distribution, the first bound for $\mu_{n+1:n+2}$ is rather bad. However, a comparison with the corresponding results due to Sugiura [42] (with n=10, 20 and m=0 in (3.3.6)) shows that our first bound is approximately of the same order as his first bound, except for $\mu_{n+1:n+2}$, for which our bound is inferior. But our second bound is far superior to his second bound. In fact, in many cases our second bound is even superior to his third bound, which is obtained by taking t=2 in (3.3.6). This, for example, for $\mu_{11:20}$ his third bound is 0.066 ± 0.008, while our second bound is 0.0644 ± 0.0037.

It is worthwhile noting that for symmetric parent distributions, equations (3.3.9) and (3.3.15) can also yield some simple recurrence relations of the form mentioned in Chapter II. Thus, for example, on taking k=1 and L_k in place of ϕ_k in (3.3.9), we get

$$\int_0^1 x(u)u^m(1-u)^m L_2(u)du = 0,$$

that is,

$$\int_0^1 x(u)u^m(1-u)^m(6u^2-6u+1)du = 0$$

and this simplifies to

$$(m+2)\mu_{m+3:2m+3} = (2m+3)\mu_{m+2:2m+2}.$$

Table 3.4.1. Approximate values and bounds for $\mu_{r:10}$

(a) Normal distribution

r	First bound	Second bound	Tabled value
9	1.299 ± 0.298	1.0041 ± 0.0028	1.00136
8	0.530 ± 0.133	0.6509 ± 0.0053	0.65606
7	0.248 ± 0.136	0.3788 ± 0.0031	0.37576
6	0.074 ± 0.057	0.1249 ± 0.0025	0.12267

(b) Cauchy distribution

r	First bound	Second bound	Tabled value
9	3.053 ± 0.073	2.9822 ± 0.0015	2.9814
8	1.246 ± 0.032	1.2749 ± 0.0028	1.2755
7	0.582 ± 0.033	0.6129 ± 0.0016	0.6132
6	0.174 ± 0.014	0.1866 ± 0.0013	0.1866

Table 3.4.2. Approximate values and bounds for $\mu_{r:20}$

(a) Normal distribution

r	First bound	Second bound	Tabled value
19	2.805 ± 1.481	1.4693 ± 0.0619	1.40760
18	1.310 ± 0.377	1.1023 ± 0.0310	1.13095
17	0.804 ± 0.247	0.9002 ± 0.0225	0.92098
16	0.545 ± 0.228	0.7390 ± 0.0116	0.74538
15	0.382 ± 0.222	0.5940 ± 0.0063	0.59030
14	0.267 ± 0.213	0.4570 ± 0.0094	0.44833
13	0.177 ± 0.188	0.3241 ± 0.0115	0.31493
12	0.101 ± 0.134	0.1937 ± 0.0095	0.18696
11	0.033 ± 0.049	0.0644 ± 0.0037	0.06200

(b) Cauchy distribution

r	First bound	Second bound	Tabled value
19	6.590 ± 0.361	6.2705 ± 0.0324	6.2648
18	3.079 ± 0.092	3.0287 ± 0.0162	3.0293
17	1.890 ± 0.060	1.9128 ± 0.0118	1.9140
16	1.279 ± 0.056	1.3259 ± 0.0060	1.3268
15	0.897 ± 0.054	0.9480 ± 0.0033	0.9484
14	0.626 ± 0.052	0.6718 ± 0.0049	0.6720
13	0.415 ± 0.046	0.4506 ± 0.0060	0.4506
12	0.238 ± 0.033	0.2600 ± 0.0050	0.2599
11	0.078 ± 0.012	0.0851 ± 0.0019	0.0850

3.5. Concluding remarks and comments

We conclude this chapter by briefly mentioning a few points of interest. Work on some of these, and related problems, could be further pursued, although a good deal of computation might be needed.

1. First, we consider the effect on the approximation as n increases. To this end note that g(u) is a polynomial of degree n-1 and hence can be written as a linear combination of Legendre polynomials. Thus

$$g(u) = \sum_{k=0}^{n-1} b_k L_k(u) \tag{3.5.1}$$

and

$$\int_0^1 g^2(u)du = \sum_{k=0}^{n-1} b_k^2,$$

where b_k is the Fourier coefficient of g(u) relative to L_k. Now using Theorem 3.3.1, say with p=q=0 and t=n-1, we have

$$\mu_{r:n} = \sum_{k=0}^{n-1} a_k b_k, \tag{3.5.2}$$

where

$$a_k = \int_0^1 x(u)L_k(u)du$$

$$= \sum_{i=0}^{k} \frac{c_{k,i}}{i+1} \cdot \mu_{i+1:i+1}. \tag{3.5.3}$$

Equations (3.5.2) and (3.5.3) also show that $\mu_{r:n}$ is a linear combination of $\mu_{i:i}$, i=1,2,...,n (see equation (2.2.3)). Further, the approximation of $\mu_{r:n}$ is obtained by taking the first (t+1) terms of

the R.H.S. of (3.5.2). From this, it may be conjectured that if n is increased, then more terms will be needed to get the same degree of approximation. Some numerical computations for the normal distribution appear to confirm this.

2. It is clear that the results of Theorems 3.3.1 and 3.3.2 apply to discrete random variables provided only that the necessary moments are finite. From equation (3.5.2) it can be shown that

$$n\,\mu_{r:n-1} = r\,\mu_{r+1:n} + (n-r)\mu_{r:n}.$$

Now equation (2.2.4) shows that this result holds for exchangeable r.v.'s also. Consequently, the approximation for $\mu_{r:n}$ is valid for exchangeable r.v.'s as well.

3. As pointed out in Section 3.4, we can get several different sequences of bounds converging to $\mu_{r+m:n+2m}$ by taking different values of m and n in (3.3.6). However, there does not seem to be any theoretical way of finding the value of m, for which the results are best possible (in the sense that for fixed t, the error bound term is a minimum). Some numerical computations performed for the normal distribution with n+2m=20 and 50 show that, in general, m=1 gives better results than m=0 or m=2 for the second approximation (t=1).

4. The results of Section 3.3 hold for any orthonormal system which satisfies the given conditions. Thus, instead of using the Legendre polynomials, we can also use the trigonometric functions defined by

$$\phi_0(u) = 1,$$

$$\phi_{2k-1}(u) = \frac{1}{\sqrt{2}} \sin 2k\pi u,\ \phi_{2k}(u) = \frac{1}{\sqrt{2}} \cos 2k\pi u,\ k \geq 1.$$

These are known to form a complete orthonormal system in [0,1] and also satisfy (3.3.5). Obviously, the Fourier coefficient a_k is no longer a linear combination of moments of order statistics. However, a_k and b_k could be evaluated by using numerical integration.

This observation raises an interesting question: Does there exist an orthonormal system for a given distribution function, for which the results are best possible, uniformly in n? We leave this question open for further investigation.

CHAPTER IV

SINGLE OUTLIER IN A REGRESSION MODEL

4.1. Formulation of the problem and the test procedures

Let $y_1, y_2, \ldots, y_n$ be n independently and normally distributed observations such that for $j=1,2,\ldots,n$

$$(4.1.1) \qquad E(Y_j) = \mu_j = \sum_{i=1}^{m} x_{ij}\beta_i, \quad Var(Y_j) = \sigma^2,$$

where σ^2, $\beta_1, \beta_2, \ldots, \beta_m$ $(m<n)$ are unknown parameters and x_{ij}'s are known real coefficients. This is the model when there are no outliers present in the data. Letting

$$\underset{n\times 1}{\underline{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \underset{m\times 1}{\underline{\beta}} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}, \quad \underset{m\times n}{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix},$$

the model can be written as

$$\underline{y} \overset{d}{=} N(X'\underline{\beta}, \sigma^2 I),$$

where I is the identity matrix of order n. Without any loss of generality, let X be of rank m. Then the least-square estimate of $\underline{\beta}$ is given by

$$\hat{\underline{\beta}} = (XX')^{-1} X \underline{y} .$$

The residual vector, $\underline{e}$, is

(4.1.2) $$\underset{n\times 1}{\underline{e}} = \Lambda\, \underline{y}\ ,$$

where

(4.1.3) $$\underset{n\times n}{\Lambda} = I - X'(XX')^{-1}X = ((\lambda_{ij}))$$

is an idempotent matrix of rank n-m. Further

(4.1.4) $$\underline{e} \overset{d}{=} N(\underline{0},\ \Lambda\,\sigma^2).$$

The error sum of squares is

(4.1.5) $$S^2 = \underline{e}'\underline{e} = \underline{y}'\,\Lambda\,\underline{y}\ ,$$

which is distributed as $\sigma^2\chi^2$ with n-m degrees of freedom.

The method for detecting a single outlier will be based on these residuals, standardized in some way. We now assume that $\lambda_{ii} > 0$ (i=1,2,...,n). A necessary and sufficient condition for this is given at the end of Theorem 4.1.1. We need the following lemma (see e.g. Rao [33], p. 29).

Lemma 4.1.1. Let A and D be non-singular square matrices of orders m and n respectively and B be a m×n matrix. If $B'A^{-1}B + D^{-1}$ is invertible, then A + BDB' is also invertible and

$$(A+BDB')^{-1} = A^{-1} - A^{-1}B(B'A^{-1}B+D^{-1})^{-1}B'A^{-1}.$$

Theorem 4.1.1. With the notations used as above, let X and Λ be partitioned according to

$$X = [X_1\ X_2],\quad \Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix},$$

where X_1 is of order m×k and Λ_{11} is of order k×k. A necessary and sufficient condition for Λ_{11} to be positive definite is that X_2 is of rank m.

Proof: By definition

$$\Lambda_{11} = I - X_1'(XX')^{-1}X_1.$$

Let $\underline{e}_1$ be the first k components of the residual vector $\underline{e}$. Then by (4.1.4)

$$\underline{e}_1 \overset{d}{=} N(\underline{0}, \Lambda_{11}\sigma^2),$$

which shows that Λ_{11} is at least positive semidefinite. If X_2 is of rank m, then X_2X_2' is also of rank m and by Lemma 4.1.1.

$$\Lambda_{11}^{-1} = I + X_1'(X_2X_2')^{-1}X_1,$$

which proves that Λ_{11} is positive definite.

Conversely, if Λ_{11} is positive definite and X is of rank m, then on applying the lemma to $XX'-X_1X_1'$ we see that $XX'-X_1X_1'$ is invertible. But $X_2X_2' \equiv XX'-X_1X_1'$. Hence X_2X_2' is of rank m and so is X_2.

Applying Theorem 4.1.1 with k=1, we see that $\lambda_{ii} > 0$ provided that the rank of the matrix obtained by deleting the i^{th} column of X is m.

Denote the correlation coefficient between e_i and e_j by ρ_{ij}. Then

$$\rho_{ij} = \lambda_{ij}/(\lambda_{ii}\lambda_{jj})^{\frac{1}{2}}.$$

Note that ρ_{ij} is well defined, because by our assumption $\lambda_{ii} > 0$ for all i. Let

$$\rho_{min} = \underset{i\neq j}{\text{Min}}\ \rho_{ij},\ \rho_{max} = \underset{i\neq j}{\text{Max}}\ \rho_{ij}. \tag{4.1.6}$$

We will test for a single outlier under the following assumption:

$$-1 < \rho_{min} \leq \rho_{max} < 1. \tag{4.1.7}$$

This condition holds for a wide class of regression models and insures that the joint distribution of e_i and e_j is non-singular. Now applying Theorem 4.1.1 with k=2, it follows that assumption (4.1.7) holds, provided that the rank of the matrix obtained by deleting the i^{th} and j^{th} columns of X is m. Another sufficient condition for the validity of (4.1.7) is given later in Corollary 4.2.1.

A test procedure for detecting a single specified outlier with a shift in location has been given by Srikantan [40]. He also gives the test statistics when the outlier is not specified. For this case, he only obtains the nominal percentage points which control the error of the first kind at a level not exceeding the specified one. These points essentially give an upper limit for the true percentage points. In this dissertation, we shall consider some additional test statistics depending on the knowledge available about σ^2. One-sided statistics, appropriate for testing an outlier on the right are denoted by U and the corresponding two-sided statistics are denoted by V. All these test statistics are maxima of suitably standardized or studentized residuals and in each case, the maximum is taken over the set $\{1,2,\ldots,n\}$.

Case 1: σ^2 known

$$(4.1.8) \qquad U_1 = \underset{i}{\text{Max}} \frac{e_i}{\sigma \lambda_{ii}^{\frac{1}{2}}}, \quad V_1 = \underset{i}{\text{Max}} \left|\frac{e_i}{\sigma \lambda_{ii}^{\frac{1}{2}}}\right| .$$

Case 2: σ^2 unknown, but a root mean square estimator s_ν of σ based on ν degrees of freedom and independent of $y_1, y_2, \ldots, y_n$ is available. In this case, we can either use external studentization or pooled studentization. In the former method, any knowledge about σ^2 from the sample is totally ignored, while in the latter, it is pooled with s_ν. The test statistics considered for external studentization are

$$(4.1.9) \qquad U_2 = \underset{i}{\text{Max}} \frac{e_i}{s_\nu \lambda_{ii}^{\frac{1}{2}}}, \quad V_2 = \underset{i}{\text{Max}} \left|\frac{e_i}{s_\nu \lambda_{ii}^{\frac{1}{2}}}\right|$$

and for pooled studentization are

$$(4.1.10) \qquad U_3 = \underset{i}{\text{Max}} \frac{e_i}{S_p \lambda_{ii}^{\frac{1}{2}}}, \quad V_3 = \underset{i}{\text{Max}} \left|\frac{e_i}{S_p \lambda_{ii}^{\frac{1}{2}}}\right|,$$

where

$$S_p^2 = S^2 + \nu s_\nu^2 .$$

When no such s_ν is available, we will use U_3 and V_3 with $\nu=0$. (The case considered by Srikantan).

In view of continuity of $e_1, e_2, \ldots, e_n$ and the assumption (4.1.7), it follows that the maximum in all of these test statistics will occur for a single i, say $i=i_o$. Large values of U or V then indicate that y_{i_o} is an outlier. To find the percentage points of these test

statistics, we need their distribution in the null case (when there is no outlier). These, in general, involve the elements of Λ. However, upper and lower limits for the true percentage points can be obtained by using the Bonferroni inequalities. Some general results for this are given in Section 4.3.

4.2. Bounds for correlation coefficients

It is of some theoretical interest to get bounds for the magnitude of the correlation coefficients $\{\rho_{ij}\}$, where ρ_{ij} has been defined in Section 4.1. In this section, we will obtain some simple upper bounds for $|\rho_{ij}|$ $(i\neq j)$. Among other things, this will also provide sufficient conditions for the validity of assumption (4.1.7).

Theorem 4.2.1. Let Λ be given by equation (4.1.3). Then for $i\neq j$

$$\rho_{ij}^2 \leq \text{Min}\left(\frac{1-\lambda_{ii}}{\lambda_{jj}}, \frac{1-\lambda_{jj}}{\lambda_{ii}}\right). \tag{4.2.1}$$

Proof: By considering the equation $\Lambda^2=\Lambda$ and equating the $(i,i)^{th}$ element on both sides we get

$$\lambda_{i1}^2 + \ldots + \lambda_{ii}^2 + \ldots + \lambda_{ij}^2 + \ldots + \lambda_{in}^2 = \lambda_{ii}.$$

Hence for $i\neq j$

$$\lambda_{ii}^2 + \lambda_{ij}^2 \leq \lambda_{ii},$$

that is,

$$\lambda_{jj}\rho_{ij}^2 \leq 1-\lambda_{ii}. \tag{4.2.2}$$

Similarly

$$\lambda_{ii}\rho_{ij}^2 \leq 1-\lambda_{jj}. \tag{4.2.3}$$

Combining (4.2.2) and (4.2.3) we have

$$\rho_{ij}^2 \leq \text{Min}\left(\frac{1-\lambda_{ii}}{\lambda_{jj}}, \frac{1-\lambda_{jj}}{\lambda_{ii}}\right).$$

Using equation (4.2.1), a sufficient condition for assumption (4.1.7) can be easily obtained. However, the following corollary gives a rather compact result.

Corollary 4.2.1. A sufficient condition for the validity of assumption (4.1.7) is that

$$\lambda_{(1)} + \lambda_{(2)} > 1,$$

where $\lambda_{(1)}$ is the smallest of λ_{kk} (k=1,2,...,n) and $\lambda_{(2)}$ is the second smallest.

Proof: Adding equations (4.2.2) and (4.2.3) we see that for $i \neq j$

$$\rho_{ij}^2 \leq \frac{2}{\lambda_{ii}+\lambda_{jj}} - 1. \tag{4.2.4}$$

Now if

$$\lambda_{(1)} + \lambda_{(2)} > 1,$$

then

$$\lambda_{ii} + \lambda_{jj} > 1 \quad \text{for all } i \neq j.$$

Hence from equation (4.2.4)

$$\rho_{ij}^2 < 1 \qquad \text{for all } i \neq j$$

and the result follows.

Remark 1. We shall show later that for $\lambda_{ii} + \lambda_{jj} > 1$, equation (4.2.4) can be replaced by a stronger inequality

$$|\rho_{ij}| \leq \frac{2}{\lambda_{ii}+\lambda_{jj}} - 1. \tag{4.2.5}$$

The proof of this involves some geometric considerations and is deferred till Section 7.2. Note that

$$\text{Min}\left(\frac{1-\lambda_{ii}}{\lambda_{jj}}, \frac{1-\lambda_{jj}}{\lambda_{ii}}\right) \leq \frac{2}{\lambda_{ii}+\lambda_{jj}} - 1.$$

Consequently, equation (4.2.1) gives better bounds than (4.2.4). However, a comparison between (4.2.1) and (4.2.5) is difficult.

So far we have not made any assumption about the homoscedasticity of the residuals, which we do now. Anscombe [3] lists a number of examples where the residuals have equal variance, viz. $(n-m)\sigma^2/n$. He has also shown that a lower bound for the magnitude of the largest correlation coefficient is $[m/(n-m)(n-1)]^{\frac{1}{2}}$. From equation (4.2.5) we see that in this case

$$|\rho_{ij}| \leq \frac{n}{n-m} - 1 = \frac{m}{n-m} \tag{4.2.6}$$

and a sufficient condition for the validity of (4.1.7) is simply $n>2m$.

In practice, the sample size n is usually much larger than the number of parameters m and the condition $n>2m$ is likely to be satisfied. Moreover, if m is small and n is large, then equation (4.2.6) shows that the correlations will be close to 0 and can often be ignored.

4.3. Upper percentage points of statistics expressible as maxima

Let $z_1, z_2, \ldots, z_n$ be n identically distributed r.v.'s and let $Z = \text{Max}\{z_i : i=1,2,\ldots,n\}$. Then clearly

$$\Pr(Z>z) = \Pr\left(\bigcup_{i=1}^{n} (z_i>z)\right). \tag{4.3.1}$$

Applying the first two Bonferroni inequalities to (4.3.1), we get

$$n \Pr(z_1>z) - \sum_{i<j} \Pr(z_i>z, z_j>z) \leq \Pr(Z>z) \leq n \Pr(z_1>z). \tag{4.3.2}$$

Let z_α be the true upper $100\alpha\%$ point of Z. Then an upper limit $\bar{z}_\alpha$ for z_α is obtained by solving for z in

$$n \Pr(z_1>z) = \alpha. \tag{4.3.3}$$

Similarly, a lower limit $\underline{z}_\alpha$ for z_α may be obtained by solving for z in

$$n \Pr(z_1>z) - \sum_{i<j} \Pr(z_i>z, z_j>z) = \alpha. \tag{4.3.4}$$

We here point out that it is possible for (4.3.4) not to have a solution for all values of α, especially for large α.

Now if $z_1, z_2, \ldots, z_n$ are such that

$$\Pr(z_i>\bar{z}_\alpha, z_j>\bar{z}_\alpha) \leq [\Pr(z_1>\bar{z}_\alpha)]^2 \quad \text{for all } i<j, \tag{4.3.5}$$

then (4.3.2) gives

$$\Pr(Z>\bar{z}_\alpha) \geq \alpha - \frac{(n-1)\alpha^2}{2n} \geq \alpha - \tfrac{1}{2}\alpha^2. \tag{4.3.6}$$

For α small, the lower bound given at (4.3.6) is close to α and hence $\bar{z}_\alpha$ will be approximately equal to z_α. Further, if

(4.3.7) $\Pr(z_i > \bar{z}_\alpha, z_j > \bar{z}_\alpha) = 0$ for all $i<j$

then equation (4.3.2) shows that $\bar{z}_\alpha$ coincides with z_α.

Apart from these simplifications, it is possible to sharpen the upper limits in some special cases by using the properties of the joint distribution of $z_1, z_2, \ldots, z_n$. This will be illustrated in later chapters.

4.4. Measures of performance

We now study the performance of the various test statistics proposed in Section 4.1 in the non-null situation when outliers are present. It is assumed that just one of the observations -- we do not know which one -- is an outlier. The null hypothesis, H_o, specifies that there is no outlier and under H_o our model is given by equation (4.1.1). The alternative hypothesis is the union of n mutually exclusive hypotheses $H_1, H_2, \ldots, H_n$ and under H_k

$$\underline{y} \overset{d}{=} N(\cdot, \sigma^2 I),$$

where for $j=1,2,\ldots,n$

(4.4.1) $$E(Y_j) = \begin{cases} \mu_j & \text{if } j \neq k \\ \mu_k + \theta & \text{if } j = k \end{cases}$$

and μ_j is defined in (4.1.1). For outliers on the right $\theta>0$ and for outliers in either direction $\theta \neq 0$.

Under H_o, the distribution of the residual vector $\underline{e}$ is given by (4.1.4). From equation (4.1.2) it is clear that under H_k, $\underline{e}$ is normally distributed with variance-covariance matrix $\Lambda\, \sigma^2$ and mean

$$E(\underline{e}|H_k) = \Lambda\ E(\underline{Y}|H_k)$$
$$= \Lambda[E(\underline{Y}|H_o) + \underline{\varepsilon}_k\theta],$$

where $\underline{\varepsilon}_k'$ is an n-vector with k^{th} component 1 and all other components 0. Thus

$$E(\underline{e}|H_k) = \Lambda X'\underline{\beta} + \Lambda\underline{\varepsilon}_k\theta$$
$$= \Lambda\underline{\varepsilon}_k\theta\ , \tag{4.4.2}$$

because

$$\Lambda X' = [I-X'(XX')^{-1}X]X'$$
$$= X'-X' = 0.$$

Further the error sum of squares, S^2, has a noncentral $\sigma^2\chi^2$ distribution with n-m degrees of freedom and noncentrality parameter Δ_{kk}^2, where

$$\sigma^2\Delta_{kk}^2 = E(\underline{Y}'|H_k)\ \Lambda\ E(\underline{Y}|H_k)$$
$$= (\underline{\beta}'X + \underline{\varepsilon}_k'\theta)\Lambda(X'\underline{\beta} + \underline{\varepsilon}_k\theta)$$
$$= \underline{\varepsilon}_k'\ \Lambda\ \underline{\varepsilon}_k\theta^2$$
$$= \lambda_{kk}\theta^2\ . \tag{4.4.3}$$

For a random sample of size n from a $N(\mu,\sigma^2)$ parent, David and Paulson [10] have proposed a number of measures of performance. However, their measures are not suitable in the present case because the distribution of $\underline{e}$ under H_k depends on Λ. Thus, we are led to consider some other measures, which in special cases reduce to the measures given by David and Paulson.

Note that, under H_o, the U and V test statistics of Section 4.1 are of the form $Z = \text{Max}\{z_i: i=1,2,\ldots,n\}$, where $z_1,z_2,\ldots,z_n$

are identically distributed random variables. With the notation used in Section 4.3, let

P_k = Probability that z_k is significantly large when H_k is true

(4.4.4) $= \Pr(z_k > z_\alpha | H_k)$,

Q_k = Probability of rejecting H_o when H_k is true

(4.4.5) $= \Pr(Z > z_\alpha | H_k)$.

In general, both P_k and Q_k depend on k. Also, any measure of performance must be symmetric with respect to hypotheses $H_1, H_2, \ldots, H_n$. If we now assume that a priori each observation has an equal chance of being declared an outlier, then the following measures seem reasonable:

$$(4.4.6) \qquad P_a = \min_k P_k, \quad P_b = \frac{1}{n} \sum_{k=1}^{n} P_k,$$

$$(4.4.7) \qquad Q_a = \min_k Q_k, \quad Q_b = \frac{1}{n} \sum_{k=1}^{n} Q_k,$$

where the minimum is taken over the set $\{1,2,\ldots,n\}$. In general $P_a \leq P_b$ and $Q_a \leq Q_b$. But, in the special case when P_k and Q_k do not depend on k, then

$P_a = P_b = P_1$ = Probability that z_1 is significantly large when the alternative hypothesis is true

and

$Q_a = Q_b = Q_1$ = Power function.

These are the two measures considered in detail by David and Paulson [10].

It is easy to show that

(4.4.8) $$Q_k \geq P_k, \quad k=1,2,\ldots,n.$$

Hence

$$Q_a \geq P_a \quad \text{and} \quad Q_b \geq P_b.$$

Evaluation of Q_k is quite laborious, but again Bonferroni inequalities can be used to get upper and lower bounds for Q_k. Thus

(4.4.9) $$\text{Max}(P_k, \underline{Q}_k) \leq Q_k \leq \bar{Q}_k,$$

where

$$\bar{Q}_k = \sum_{i=1}^{n} \Pr(z_i > z_\alpha | H_k)$$

(4.4.10) $$= P_k + \sum_{\substack{i=1 \\ i\neq k}}^{n} \Pr(z_i > z_\alpha | H_k)$$

and

(4.4.11) $$\underline{Q}_k = \bar{Q}_k - \sum_{i<j} \Pr(z_i > z_\alpha, z_j > z_\alpha | H_k).$$

Using these results, lower and upper bounds for Q_a and Q_b can be obtained. In practice, z_α is available for very few regression models and as an approximation we will use the upper limit $\bar{z}_\alpha$ in (4.4.4) and (4.4.5). This will give a lower bound for P_k and Q_k.

4.5. Examples

We shall consider the following examples of the fixed effects models, which are of interest in practice:

Example 4.5.1. Sample from a $N(\mu,\sigma^2)$ parent. In this case, statistics related to U and V have been studied by several authors (see e.g. Chew [6] for a good review of the literature). We have included it here for the sake of completeness and performance studies. The variance of any residual is $(n-1)\sigma^2/n$ and the rank of matrix Λ is n-1. The correlation between any two residuals is $-1/(n-1)$ and hence the condition (4.1.7) holds for $n \geq 3$.

Example 4.5.2. One-way layout. Let y_{ij} be the j^{th} observation from the i^{th} class ($j=1,2,\ldots,n_i$; $i=1,2,\ldots,m$; $m>1$) and let e_{ij} be the corresponding residual, i.e.

$$e_{ij} = y_{ij}-y_{i\cdot} ,$$

where

$$y_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}.$$

It is easy to show that the variance of e_{ij} is $(n_i-1)\sigma^2/n_i$ and the rank of matrix Λ is n-m, where $n = \sum_{i=1}^{m} n_i$. Further

$$\rho(e_{ij},e_{i'j'}) = \begin{cases} -1/(n_i-1) & \text{if } i=i', j\neq j' \\ 0 & \text{otherwise.} \end{cases}$$

This shows that $\rho_{max} = 0$ and the condition (4.1.7) holds provided that $n_i \geq 3$ for all i.

Example 4.5.3. Two-way layout. Let y_{ij} be the observation in the i^{th} row and j^{th} column of a two-way layout ($i=1,2,\ldots,r$; $j=1,2,\ldots,c$) and let e_{ij} be the corresponding residual, i.e.

$$e_{ij} = y_{ij}-y_{i\cdot}-y_{\cdot j}+y_{\cdot\cdot} ,$$

where

$$y_{i\cdot} = \frac{1}{c}\sum_{j=1}^{c} y_{ij},\quad y_{\cdot j} = \frac{1}{r}\sum_{i=1}^{r} y_{ij},\quad y_{\cdot\cdot} = \frac{1}{rc}\sum_{i}\sum_{j} y_{ij}.$$

Here

$$n = rc \quad \text{and} \quad m = r+c-1.$$

The variance of e_{ij} is $(n-m)\sigma^2/n$ and the rank of matrix Λ is $n-m$. Further

$$\rho(e_{ij},e_{i'j'}) = \begin{cases} -1/(c-1) & \text{if } i=i',\ j\neq j' \\ -1/(r-1) & \text{if } i\neq i',\ j=j' \\ 1/((r-1)(c-1)) & \text{if } i\neq i',\ j\neq j'. \end{cases}$$

From this, it follows that the correlations are both positive and negative. But these can assume at most three different values. Also, the condition (4.1.7) holds provided both r and c are greater than or equal to 3.

CHAPTER V

DISTRIBUTION THEORY WHEN VARIANCE IS KNOWN

5.1. Introduction

In this chapter, we will study the statistics U_1 and V_1 of Section 4.1. For convenience, we shall drop the suffixes and write the test statistics as U and V. Letting

$$b_i = \frac{e_i}{\sigma \lambda_{ii}^{\frac{1}{2}}}, \quad i=1,2,\ldots,n, \tag{5.1.1}$$

we have from (4.1.8)

$$U = \underset{i}{\text{Max}}\ b_i, \quad V = \underset{i}{\text{Max}}|b_i|. \tag{5.1.2}$$

Further from equation (4.1.4), under H_o, $\underline{b}' = (b_1,b_2,\ldots,b_n)$ has a singular normal distribution with means zero, unit variances and correlations $\{\rho_{ij}\}$. Hence the joint distribution of b_i, b_j is $N(0,0,1,1,\rho_{ij})$, which in view of (4.1.7) is non-singular. Moreover, marginally, each b_i has a unit normal distribution.

In Section 5.2, we give upper and lower limits for the true percentage points of U and V. Some results regarding the performance of these statistics are given in Section 5.3. Finally some applications are considered in Section 5.4.

5.2. Percentage points

5.2.1. Upper limits

Following the notations introduced in Section 4.3, we shall denote the true upper 100α% points of U and V by u_α and v_α respectively. Then the upper limits for u_α and v_α, denoted by $\bar{u}_\alpha$ and $\bar{v}_\alpha$ respectively, are obtained by solving

$$n \Pr(b_1 > \bar{u}_\alpha) = \alpha, \quad n \Pr(|b_1| > \bar{v}_\alpha) = \alpha, \tag{5.2.1}$$

that is,

$$\Phi(\bar{u}_\alpha) = 1 - \frac{\alpha}{n}, \quad \Phi(\bar{v}_\alpha) = 1 - \frac{\alpha}{2n},$$

where

$$\Phi(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}x^2)dx.$$

5.2.2. Improved upper limits

We now show that it is always possible to get better upper limits for v_α than $\bar{v}_\alpha$ of equation (5.2.1). We first state some inequalities involving multivariate normal distributions, which are of value in the present context (see e.g. Šidák [38]).

Theorem 5.2.1. Let $(X_1,\ldots,X_n)$ be a random vector having a n-variate normal distribution with mean values 0, variances 1, and having, under the probability law P_K, the correlation matrix $K = ((k_{ij}))$, and, under the probability law P_R, the correlation matrix $R = ((r_{ij}))$. If

$$k_{ij} \leq r_{ij} \quad \text{for all } i,j,$$

then

$$(5.2.2) \qquad P_K(X_1 \leq c_1, \ldots, X_n \leq c_n) \leq P_R(X_1 \leq c_1, \ldots, X_n \leq c_n)$$

for any numbers $c_1, c_2, \ldots, c_n$.

Corollary 5.2.1. If $k_{ij} \leq 0$ for $i \neq j = 1,2,\ldots,n$, then

$$(5.2.3) \qquad P_K(X_1 \leq c_1, \ldots, X_n \leq c_n) \leq \prod_{i=1}^{n} \Pr(X_i \leq c_i).$$

Proof: Take $R = I_n$ in (5.2.2).

Corollary 5.2.2. If $r_{ij} \geq 0$ for $i \neq j = 1,2,\ldots,n$, then

$$(5.2.4) \qquad P_R(X_1 \leq c_1, \ldots, X_n \leq c_n) \geq \prod_{i=1}^{n} \Pr(X_i \leq c_i).$$

Proof: Take $K = I_n$ in (5.2.2).

Theorem 5.2.2. Let $(X_1, \ldots, X_n)$ have a multivariate normal distribution with mean values 0 and with an arbitrary correlation matrix, then

$$(5.2.5) \qquad \Pr(|X_1| \leq c_1, \ldots, |X_n| \leq c_n) \geq \prod_{i=1}^{n} \Pr(|X_i| \leq c_i)$$

for any non-negative numbers $c_1, c_2, \ldots, c_n$.

We now apply Theorem 5.2.2 to get improved upper limits for v_α.

Theorem 5.2.3. $\bar{v}'_\alpha$ as obtained by solving

$$(5.2.6) \qquad \Pr(|b_1| > \bar{v}'_\alpha) = 1 - (1-\alpha)^{\frac{1}{n}}$$

gives an upper limit for v_α which is lower (and hence better) than $\bar{v}_\alpha$ of (5.2.1).

Proof: By definition

$$\Pr(V>\bar{v}'_\alpha) = \Pr(\max_i |b_i| > \bar{v}'_\alpha)$$

$$= 1-\Pr(\bigcap_{i=1}^{n} (|b_i| \leq \bar{v}'_\alpha))$$

$$\leq 1 - \prod_{i=1}^{n} \Pr(|b_i| \leq \bar{v}'_\alpha) \quad \text{by (5.2.5)}$$

$$= 1 - \prod_{i=1}^{n} (1-\alpha)^{\frac{1}{n}} \quad \text{by (5.2.6)}$$

$$= \alpha.$$

Hence $\bar{v}'_\alpha$ is a nominal upper $100\alpha\%$ point of V. Further it is easy to show that

$$\frac{\alpha}{n} \leq 1-(1-\alpha)^{\frac{1}{n}},$$

that is,

$$\Pr(|b_1| > \bar{v}_\alpha) \leq \Pr(|b_1| > \bar{v}'_\alpha).$$

This implies that

$$\bar{v}_\alpha \geq \bar{v}'_\alpha ,$$

which shows that $\bar{v}'_\alpha$ is closer to the true percentage point than $\bar{v}_\alpha$; i.e., $\bar{v}'_\alpha$ is an improved upper limit for v_α. Note that

$$1-(1-\alpha)^{\frac{1}{n}} \doteq \frac{\alpha}{n} + \frac{\alpha^2}{2n}$$

and the improvement is slight if α is small and n is large.

We now consider the statistic U. If $\rho_{min} \geq 0$ then an application of (5.2.4) shows that an improved upper limit for u_α is $\bar{u}'_\alpha$, where

$$\Pr(b_1 > \bar{u}'_\alpha) = 1-(1-\alpha)^{\frac{1}{n}}.$$

Remark 1. Regression models satisfying $\rho_{min} \geq 0$ are rare. One such example is given below. However, for this example the condition (4.1.7) does not hold. We are not aware of any example where both these conditions hold simultaneously.

Example 5.2.1. Let the design matrix X of the regression model be

$$X = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

Then

$$XX' = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

and

$$\Lambda = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

so that ρ_{ij} $(i \neq j)$ is either 0 or 1.

5.2.3. Lower bound for the significance level attained

It is clear that by using the upper limits instead of the true percentage points, we are actually working at a reduced significance level. But how big is the reduction? Obviously, this will depend on the particular regression model under consideration. In an important case when $\rho_{max} \leq 0$, the following theorem shows that the significance level for the one-sided test statistic cannot drop below $1-e^{-\alpha}$ by using $\bar{u}_\alpha$.

Theorem 5.2.4. If $\rho_{max} \leq 0$, then

(5.2.7) $$\Pr(U>\bar{u}_\alpha) \geq 1-(1-\frac{\alpha}{n})^n \geq 1-e^{-\alpha} ,$$

where $\bar{u}_\alpha$ is given by (5.2.1).

Proof: For any c,

$$\Pr(U>c) = \Pr(\bigcup_{i=1}^{n}(b_i>c))$$

$$= 1-\Pr(\bigcap_{i=1}^{n}(b_i\leq c))$$

$$\geq 1 - \prod_{i=1}^{n}\Pr(b_i\leq c) \quad \text{by (5.2.3)}$$

(5.2.8) $$= 1-[\Phi(c)]^n.$$

The first inequality of (5.2.7) follows if we set $c = \bar{u}_\alpha$ and use equation (5.2.1). To prove the second inequality, note that the function

$$h_\alpha(t) = (1-\frac{\alpha}{t})^t \text{ for } t\geq 1,\ 0\leq\alpha\leq 1$$

is a monotonic non-decreasing function of t and tends to $e^{-\alpha}$ as t tends to ∞.

Note that the same approach applied to the second Bonferroni bound (equation (4.3.6)) gives

$$\Pr(U>\bar{u}_\alpha) \geq \alpha - \frac{(n-1)\alpha^2}{2n} .$$

This is slightly less than $1-(1-\frac{\alpha}{n})^n$.

5.2.4. Lower limits

For the lower limits $\underline{u}_\alpha$ and $\underline{v}_\alpha$, we solve for u and v in

$$(5.2.9) \qquad n\ \Pr(b_1>u) - \sum_{i<j} \Pr(b_i>u,\ b_j>u) = \alpha$$

and

$$(5.2.10) \qquad n\ \Pr(|b_1|>v) - \sum_{i<j} \Pr(|b_i|>v,\ |b_j|>v) = \alpha$$

respectively. Let

$$(5.2.11) \quad L(h,k,\rho) = \int_h^\infty \int_k^\infty \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)\right] dydx.$$

Then

$$\Pr(b_i>u,\ b_j>u) = L(u,u,\rho_{ij})$$

and

$$\Pr(|b_i|>v,\ |b_j|>v) = 2L(v,v,\rho_{ij}) + 2L(v,v,-\rho_{ij}).$$

The function $L(h,k,\rho)$ has been tabulated by the National Bureau of Standards [29]. But, here we are interested in the "tail" probabilities and these tables are not of much use. It is well known that (see e.g. [29]) for positive values of h and k

$$(5.2.12) \quad L(h,k,\rho) = \frac{1}{2\pi} \int_{\arccos\rho}^{\pi} \exp\left[-\tfrac{1}{2}(h^2+k^2-2hk\cos\omega)\operatorname{cosec}^2\omega\right] d\omega.$$

Therefore

$$L(h,h,\rho) = \frac{1}{2\pi} \int_{\arccos\rho}^{\pi} \exp\left[\frac{-h^2}{1+\cos\omega}\right] d\omega\ .$$

This integral can be evaluated by using the quadrature method given in the Appendix. It is clear that unless ρ_{ij} can assume very few values, the evaluation of $\underline{u}_\alpha$ and $\underline{v}_\alpha$ will require a good deal of computation.

In the special case when $\rho_{max} \leq 0$, the solution of equation

$$\Phi(u) = (1-\alpha)^{\frac{1}{n}} \tag{5.2.13}$$

for u, gives another lower limit, $\underline{u}'_\alpha$, for u_α. This follows immediately by setting c=u in equation (5.2.8). In general, $\underline{u}'_\alpha$ is expected to be slightly inferior to $\underline{u}_\alpha$ of (5.2.9). However, if ρ_{ij}'s are close to 0 then it may be even superior to $\underline{u}_\alpha$. Since

$$(1-\alpha)^{\frac{1}{n}} \doteq 1 - \frac{\alpha}{n} - \frac{\alpha^2}{2n} ,$$

it follows, from (5.2.1) and (5.2.13), that the difference between the upper limit $\bar{u}_\alpha$ and lower limit $\underline{u}'_\alpha$ will be small. Thus, for example, for α=.05 and n=10 we have $\bar{u}_\alpha$ = 2.576 and $\underline{u}'_\alpha$=2.568.

5.3. Performance of test statistics

In this section, we shall continue to use the notation introduced in Section 4.4. From equations (4.4.2) and (5.1.1), it follows that under H_k the joint distribution of $b_1, b_2, \ldots, b_n$ is multivariate normal with unit variances, correlation matrix $((\rho_{ij}))$ and means given by

$$E(b_i | H_k) = \delta_{ik} \cdot \frac{\theta}{\sigma} , \quad i=1,2,\ldots,n, \tag{5.3.1}$$

where

$$\delta_{ik} = \frac{\lambda_{ik}}{\lambda_{ii}^{\frac{1}{2}}} . \tag{5.3.2}$$

Since all the measures of performance depend on the ratio θ/σ, we may without loss of generality take σ=1.

One-sided test statistic U. From equation (4.4.4)

$$P_k = \Pr(b_k > u_\alpha | H_k), \quad k=1,2,\ldots,n$$

$$= \Phi(-u_\alpha + \lambda_{kk}^{\frac{1}{2}}\theta). \tag{5.3.3}$$

For fixed $\theta>0$, P_k will attain a minimum when λ_{kk} is a minimum. Hence

$$P_a = \min_k P_k = \Phi(-u_\alpha + \lambda_{(1)}^{\frac{1}{2}}\theta),$$

where $\lambda_{(1)} = \min_k \lambda_{kk}$.

The computation of P_b requires the evaluation of at most n different normal probabilities given at (5.3.3). If all λ_{kk} are equal, then $P_a = P_b = P_1$. But if any one of the λ_{kk} is much smaller than the others, then P_b will be a better measure of performance and will be worth computing.

Similarly, equation (4.4.5) gives

$$Q_k = \Pr(U > u_\alpha | H_k), \quad k=1,2,\ldots,n$$

$$= \Pr(\bigcup_{i=1}^{n} (b_i > u_\alpha | H_k)). \tag{5.3.4}$$

Equations (4.4.10) and (4.4.11) then give

$$\bar{Q}_k = P_k + \sum_{\substack{i=1 \\ i\neq k}}^{n} \Phi(-u_\alpha + \delta_{ik}\theta) \tag{5.3.5}$$

and

$$\underline{Q}_k = \bar{Q}_k - \sum_{i<j} \Pr(b_i>u_\alpha, b_j>u_\alpha | H_k)$$

$$= \bar{Q}_k - \sum_{i<j} L(u_\alpha-\delta_{ik}\theta, u_\alpha-\delta_{jk}\theta, \rho_{ij}), \tag{5.3.6}$$

where the function L has been defined at (5.2.11). Since the bivariate probabilities in (5.3.6) are difficult to obtain, it is useful to note that in the important case when $\rho_{max} \leq 0$

$$Q_k \geq \underline{Q}'_k ,$$

where

$$\underline{Q}'_k = 1 - \prod_{i=1}^{n} \Phi(u_\alpha - \delta_{ik}\theta). \tag{5.3.7}$$

Proof: From equations (5.3.1) and (5.3.4)

$$Q_k = \Pr\left(\bigcup_{i=1}^{n} (w_i > u_\alpha - \delta_{ik}\theta)\right),$$

where $w_1, w_2, \ldots, w_n$ have a multinormal distribution with mean values 0, variances 1 and correlation matrix $((\rho_{ij}))$. Now on applying the same technique as in the proof of Theorem 5.2.4, we get

$$Q_k \geq 1 - \prod_{i=1}^{n} \Pr(w_i \leq u_\alpha - \delta_{ik}\theta),$$

that is, $Q_k \geq \underline{Q}'_k$.

A direct comparison between (5.3.6) and (5.3.7) is difficult, but (5.3.6) is expected to give better results, unless the ρ_{ij}'s are close to 0. Some comparative results for Example 4.5.1 are given in the next section. However (5.3.7) is better than (4.4.8), because from (5.3.7)

$$\underline{Q}'_k = 1-\Phi(u_\alpha - \lambda_{kk}^{\frac{1}{2}}\theta) \cdot \prod_{\substack{i=1 \\ i\neq k}}^{n} \Phi(u_\alpha - \delta_{ik}\theta)$$

$$\geq 1-\Phi(u_\alpha - \lambda_{kk}^{\frac{1}{2}}\theta) = P_k.$$

Two-sided test statistic V. Now

$$P_k = \Pr(|b_k| > v_\alpha | H_k), \quad k=1,2,\ldots,n$$

$$= \Phi(-v_\alpha + \lambda_{kk}^{\frac{1}{2}}\theta) + \Phi(-v_\alpha - \lambda_{kk}^{\frac{1}{2}}\theta). \tag{5.3.8}$$

To find P_a, we need the following easily proved lemma.

Lemma 5.3.1. For $\lambda \geq 0$, $a \geq 0$ and fixed θ

$$h(\lambda) = \Phi(-a+\lambda\theta) + \Phi(-a-\lambda\theta)$$

is a non-decreasing function of λ.

Applying the lemma, we see that

$$P_a = \min_k P_k = \Phi(-v_\alpha + \lambda_{(1)}^{\frac{1}{2}}\theta) + \Phi(-v_\alpha - \lambda_{(1)}^{\frac{1}{2}}\theta), \tag{5.3.9}$$

where $\lambda_{(1)} = \min_k \lambda_{kk}$.

Similarly, from equations (4.4.10) and (4.4.11) we get

$$\bar{Q}_k = P_k + \sum_{\substack{i=1 \\ i\neq k}}^{n} \Pr(|b_i| > v_\alpha | H_k)$$

$$= P_k + \sum_{\substack{i=1 \\ i\neq k}}^{n} [\Phi(-v_\alpha + \delta_{ik}\theta) + \Phi(-v_\alpha - \delta_{ik}\theta)] \tag{5.3.10}$$

and

$$\underline{Q}_k = \bar{Q}_k - \sum_{i<j} \Pr(|b_i| > v_\alpha, \ |b_j| > v_\alpha | H_k). \tag{5.3.11}$$

The bivariate probability at (5.3.11) is a linear function of at most 4 "L" functions defined in (5.2.11). For computational purposes,

we give the closed form for

$$p_{12}^{(k)} = \Pr(|b_1|>v_\alpha,\ |b_2|>v_\alpha|H_k).$$

Let

$$h_i = v_\alpha-\delta_{ik}\theta,\quad g_i = v_\alpha+\delta_{ik}\theta,\quad i=1,2$$

and

$$\rho = \rho_{12}.$$

Then

$$p_{12}^{(k)} = L(h_1,h_2,\rho) + L(g_1,g_2,\rho)$$

$$+ L(h_1,g_2,-\rho) + L(g_1,h_2,-\rho). \tag{5.3.12}$$

Note that P_k also gives a lower limit for Q_k and may be better than $\underline{Q}_k$ in some cases. Thus

$$Q_k \geq \text{Max}(P_k,\ \underline{Q}_k). \tag{5.3.13}$$

5.4. Applications

We first consider Example 4.5.1 with $\sigma=1$. In this case, the true percentage points of the statistic $A_1 = \underset{i}{\text{Max}}(y_i-\bar{y})$ have been tabulated by Grubbs [17] for n=3(1)25 and several values of α. Note that

$$U = \left(\frac{n}{n-1}\right)^{\frac{1}{2}} A_1.$$

From equation (5.3.4), it is easy to see that in this case $Q_a=Q_b=Q_1$. The lower bounds for Q_1 given at (5.3.6) and (5.3.7) simplify to

(5.4.1) $$\underline{Q}_1 = \Phi(-h_1) + (n-1)\Phi(-h_2) - (n-1)L(h_1,h_2,\rho) - \binom{n-1}{2}L(h_2,h_2,\rho)$$

and

(5.4.2) $$\underline{Q}_1' = 1 - \Phi(h_1) \cdot [\Phi(h_2)]^{n-1},$$

where

$$h_1 = u_\alpha - \left(\frac{n-1}{n}\right)^{\frac{1}{2}}\theta, \quad h_2 = u_\alpha + \frac{\theta}{[n(n-1)]^{\frac{1}{2}}}$$

and

$$\rho = -\frac{1}{n-1}.$$

$\underline{Q}_1$ has been tabulated by David and Paulson [10] for α=.01 and .05 and several values of n. Table 5.4.1 compares $\underline{Q}_1$ and $\underline{Q}_1'$ for α=.05 and n=5(5)20. For $h_1 \geq 0$ and $h_2 \geq 0$, the bivariate probabilities at (5.4.1) were obtained by using equation (5.2.12) and the quadrature method described in the Appendix. For other values of h_1, h_2, these were obtained by using the expressions given in [29]. In general, $\underline{Q}_1'$ is only slightly inferior to $\underline{Q}_1$; the maximum difference between $\underline{Q}_1$ and $\underline{Q}_1'$ is less than .003. Results for α=.01 are not tabled here, but the differences are even less in this case.

Both upper and lower Bonferroni type limits of the statistic $\underset{i}{\text{Max}}|y_i-\bar{y}|$ are tabulated by Halperin et al. [20] (corresponding to their case m=∞) for α=.05 and .01 and various sample sizes. However, their upper limits can be slightly decreased by using Theorem 5.2.3.

We now study the performance of

$$V_1 = (n/(n-1))^{\frac{1}{2}} \underset{i}{\text{Max}}|y_i-\bar{y}|$$

by using the improved upper limits $\bar{v}'_\alpha$ given at (5.2.6). Since all Q_k's are equal, the power function is $Q_a=Q_b=Q_1$ with a lower bound given by (5.3.13). Using $\bar{v}'_\alpha$ in place of v_α in equations (5.3.8) and (5.3.11) we have

$$Q_1 \geq \text{Max}(P_1, \underline{Q}_1), \tag{5.4.3}$$

where

$$P_1 = \Phi(-h_1) + \Phi(-h_2)$$

and

$$\begin{aligned}\underline{Q}_1 = P_1 &+ (n-1)[\Phi(-h_3) + \Phi(-h_4)] \\ &-(n-1)[L(h_1,h_3,\rho) + L(h_2,h_4,\rho) + L(h_1,h_4,-\rho) + L(h_2,h_3,-\rho)] \\ &-\binom{n-1}{2}[L(h_3,h_3,\rho) + L(h_4,h_4,\rho) + 2L(h_3,h_4,-\rho)],\end{aligned}$$

where

$$h_1 = \bar{v}'_\alpha - \left(\frac{n-1}{n}\right)^{\frac{1}{2}}\theta, \quad h_2 = \bar{v}'_\alpha + \left(\frac{n-1}{n}\right)^{\frac{1}{2}}\theta,$$

$$h_3 = \bar{v}'_\alpha + \frac{\theta}{[n(n-1)]^{\frac{1}{2}}}, \quad h_4 = \bar{v}'_\alpha - \frac{\theta}{[n(n-1)]^{\frac{1}{2}}},$$

and $\rho = -1/(n-1)$.

Since P_1 and $\underline{Q}_1$ are symmetric functions of θ, we only need the values for $\theta \geq 0$. For $\alpha=.05$, the lower bound (5.4.3) is tabled in Table 5.4.2 for n=3(1)10(5)30. It was observed that, in general, P_1 was greater than $\underline{Q}_1$ for $\theta \geq 5$ and in this case the difference $(P_1-\underline{Q}_1)$ was largest for small values of n. Note that for fixed θ, the tabulated values first increase and then decrease as the sample size increases.

For a one-way layout (Example 4.5.2), $\rho_{max} \leq 0$. Hence the lower bound (5.2.7) holds and equation (5.2.13) can be used to get the lower limits for u_α. Other results are also expected to be similar to that of previous example.

We next consider the statistic U for a two-way layout with r rows and c columns (Example 4.5.3). The upper limit, $\bar{u}_\alpha$, is immediately given by (5.2.1) with n=rc. The lower limit, $\underline{u}_\alpha$, is obtained by solving equation (5.2.9), which in the present case simplifies to

$$rc\ \Phi(-u) - [c\binom{r}{2}L(u,u,\rho_1) + r\binom{c}{2}L(u,u,\rho_2) + \{\binom{rc}{2}-c\binom{r}{2}-r\binom{c}{2}\}\ L(u,u,\rho_3)] = \alpha,$$

where

$$\rho_1 = -1/(r-1), \quad \rho_2 = -1/(c-1), \quad \rho_3 = 1/[(r-1)(c-1)].$$

Define a function h(u) by

$$h(u) = \Phi(u) + \tfrac{1}{2}[(r-1)L(u,u,\rho_1) + (c-1)L(u,u,\rho_2) + (r-1)(c-1)L(u,u,\rho_3)].$$

Then $\underline{u}_\alpha$ is the solution of the equation

$$h(u) = 1 - \frac{\alpha}{rc} \quad . \tag{5.4.4}$$

To solve this equation, note that

$$h(\bar{u}_\alpha) > 1 - \frac{\alpha}{rc} \ .$$

We also know that the lower limit $\underline{u}_\alpha$ lies between 0 and $\bar{u}_\alpha$. We then find a value d (if possible) such that

(5.4.5) $$h(\bar{u}_\alpha - d) \leq 1 - \frac{\alpha}{rc} .$$

This shows that equation (5.4.4) has a root between $\bar{u}_\alpha - d$ and $\bar{u}_\alpha$. Here, we increased d in steps of 0.1 till the inequality (5.4.5) was satisfied. This method then gives a value of d for which

$$\bar{u}_\alpha - d \leq \underline{u}_\alpha \leq \bar{u}_\alpha - d + 0.1.$$

The range of $\underline{u}_\alpha$ was then narrowed down to get 3-decimal place accuracy. Values of $\underline{u}_\alpha$ are given in Table 5.4.3 for $\alpha=.05$ and r=3(1)12, c=3(1)7. The table is symmetric in r and c and the values above the diagonal are not tabulated. For comparison purposes, the upper limit $\bar{u}_\alpha$ is also included. This table shows that the difference between upper and lower limits is small; the largest difference is .023 for r=c=3.

Using the upper limit $\bar{u}_\alpha$ in (5.3.3) we have

$$P_1 = P_a = P_b = \Phi(-\bar{u}_\alpha + (\frac{(r-1)(c-1)}{rc})^{\frac{1}{2}}\theta).$$

This is tabulated in Table 5.4.4 for $\alpha=.05$ and r=c=3(1)12. It appears that the values of P_1 do not depend greatly on the total number of observations rc, especially for $\theta \geq 3$.

Corresponding results for the two-sided test statistic V are given in Tables 5.4.5 and 5.4.6. The upper and lower limits were obtained from equations (5.2.6) and (5.2.10). The maximum difference between the upper and lower limits is .075, when r=c=3. The difference decreases rapidly as r and c increase. Thus for r=5 and c=6, it is only .010.

Table 5.4.1. Comparison between the lower bounds (5.4.1) (given in top row) and (5.4.2) (given in bottom row) for α=.05

n \ θ	1	2	3	4	5	6
5	.0975	.3059	.6435	.8956	.9842	.9988
	.0960	.3036	.6417	.8950	.9841	.9988
10	.0845	.2700	.6168	.8915	.9853	.9991
	.0832	.2676	.6143	.8905	.9851	.9991
15	.0777	.2445	.5878	.8795	.9835	.9990
	.0768	.2425	.5855	.8785	.9833	.9990
20	.0733	.2258	.5633	.8677	.9813	.9988
	.0726	.2242	.5613	.8668	.9812	.9989

Table 5.4.2. Lower bound (5.4.3) for Q_1 for α=.05

n \ θ	0	1	2	3	4	5	6
3	.045	.094	.271	.562	.816	.955	.994
4	.047	.092	.267	.574	.842	.967	.997
5	.048	.088	.259	.573	.851	.972	.997
6	.049	.085	.251	.569	.853	.973	.998
7	.049	.083	.243	.563	.853	.974	.998
8	.049	.080	.236	.556	.852	.974	.998
9	.049	.078	.230	.550	.850	.974	.998
10	.049	.077	.224	.543	.847	.974	.998
15	.050	.071	.201	.513	.833	.971	.998
20	.050	.068	.185	.490	.820	.968	.998
25	.050	.065	.173	.470	.807	.965	.997
30	.050	.063	.164	.454	.797	.963	.997

Table 5.4.3. Upper and lower limits of the statistic U_1 for a two-way layout with r rows and c columns and $\alpha=.05$

r \ c	3	4	5	6	7
3	2.539 2.516				
4	2.638 2.622	2.734 2.722			
5	2.713 2.700	2.807 2.796	2.878 2.868		
6	2.773 2.761	2.865 2.855	2.935 2.926	2.991 2.983	
7	2.823 2.812	2.914 2.904	2.983 2.974	3.038 3.030	3.084 3.076
8	2.865 2.855	2.955 2.946	3.023 3.015	3.078 3.070	3.124 3.116
9	2.902 2.893	2.991 2.983	3.059 3.051	3.113 3.105	3.158 3.151
10	2.935 2.926	3.023 3.015	3.090 3.082	3.144 3.136	3.189 3.181
11	2.965 2.956	3.052 3.044	3.118 3.111	3.172 3.164	3.216 3.209
12	2.991 2.983	3.078 3.070	3.144 3.136	3.197 3.189	3.241 3.234

Table 5.4.4. Performance P_1 of the test statistic U_1 for a two-way layout with r rows and c columns and $\alpha=.05$

r=c \ θ	1	2	3	4	5	6	7
3	.031	.114	.295	.551	.786	.928	.983
4	.024	.109	.314	.605	.845	.961	.994
5	.019	.101	.316	.626	.869	.973	.997
6	.015	.093	.312	.634	.880	.978	.998
7	.013	.085	.304	.635	.885	.980	.998
8	.011	.079	.295	.632	.887	.982	.998
9	.010	.073	.286	.627	.888	.982	.999
10	.008	.068	.277	.622	.887	.983	.999
11	.007	.064	.269	.615	.885	.983	.999
12	.007	.060	.261	.608	.883	.982	.999

Table 5.4.5. Upper and lower limits of the statistic V_1 for a two-way layout with r rows and c columns and α=.05

r \ c	3	4	5	6	7
3	2.766 2.691				
4	2.858 2.812	2.948 2.922			
5	2.928 2.893	3.016 2.998	3.083 3.071		
6	2.984 2.955	3.071 3.057	3.137 3.127	3.190 3.183	
7	3.031 3.006	3.116 3.105	3.182 3.174	3.234 3.228	3.278 3.273
8	3.071 3.048	3.156 3.145	3.220 3.214	3.272 3.267	3.315 3.312
9	3.106 3.085	3.190 3.180	3.254 3.248	3.305 3.301	3.348 3.345
10	3.137 3.118	3.220 3.212	3.283 3.278	3.335 3.331	3.377 3.374
11	3.165 3.147	3.247 3.240	3.310 3.306	3.361 3.358	3.403 3.401
12	3.190 3.173	3.272 3.265	3.335 3.330	3.385 3.382	3.427 3.425

Table 5.4.6. Performance P_1 of the test statistic V_1 for a two-way layout with r rows and c columns and α=.05

r=c \ θ	1	2	3	4	5	6	7
3	.018	.076	.222	.461	.715	.891	.971
4	.014	.074	.243	.521	.789	.940	.989
5	.011	.069	.247	.547	.820	.957	.994
6	.009	.064	.245	.557	.836	.965	.996
7	.008	.059	.240	.560	.843	.969	.997
8	.007	.055	.233	.559	.847	.971	.997
9	.006	.051	.227	.555	.848	.972	.997
10	.005	.047	.219	.550	.848	.973	.998
11	.004	.044	.213	.544	.846	.973	.998
12	.004	.041	.206	.538	.844	.973	.998

CHAPTER VI

DISTRIBUTION THEORY WHEN VARIANCE IS UNKNOWN — EXTERNAL STUDENTIZATION

6.1. Introduction

Let b_i be as defined in (5.1.1) and let

$$(6.1.1) \qquad t_i = \frac{e_i}{s_\nu \lambda_{ii}^{\frac{1}{2}}} = \frac{b_i}{s_\nu/\sigma}, \quad i=1,2,\ldots,n,$$

where s_ν is a root mean square estimator of σ based on ν degrees of freedom and independent of $y_1, y_2, \ldots, y_n$. We now take $\sigma=1$. The test statistics for locating a single outlier in this case are U_2 and V_2 defined at (4.1.9). Dropping the suffixes from U_2 and V_2, we have

$$(6.1.2) \qquad U = \underset{i}{\text{Max}}\ t_i, \quad V = \underset{i}{\text{Max}} |t_i|.$$

From (6.1.1) and (4.1.4) it follows that, under H_o, each t_i has a Student t distribution with ν degrees of freedom and the joint distribution of, say, t_1 and t_2 is bivariate t with ν degrees of freedom (see e.g. Dunnett and Sobel [13]) with density given by

$$(6.1.3) \qquad g(t_1,t_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{t_1^2 - 2\rho t_1 t_2 + t_2^2}{\nu(1-\rho^2)} \right)^{-\frac{1}{2}(\nu+2)},$$

where $\rho = \rho_{12}$.

In this chapter, our approach is similar to that of Chapter V. The percentage points of these statistics are given in Section 6.2.

In Section 6.3, we study their performance. Some applications are considered in Section 6.4.

6.2. Percentage points

6.2.1. Upper limits

The solution of equations

$$(6.2.1) \qquad n \Pr(t_1 > \bar{u}_\alpha) = \alpha, \quad n \Pr(|t_1| > \bar{v}_\alpha) = \alpha$$

give an upper limit for u_α and v_α respectively, where u_α and v_α are the true upper $100\alpha\%$ points of U and V. The c.d.f. of the t distribution has been tabulated by Hartley and Pearson [22]. Using their tables $\bar{u}_\alpha$ and $\bar{v}_\alpha$ can be readily obtained. Alternatively, tables of the incomplete beta function can be used for this purpose, since for $a \geq 0$

$$\Pr(t_1 > a) = \tfrac{1}{2} I_{\frac{\nu}{\nu+a^2}}(\tfrac{1}{2}\nu, \tfrac{1}{2}),$$

where

$$(6.2.2) \qquad I_x(p,q) = \int_0^x \frac{1}{B(p,q)} u^{p-1}(1-u)^{q-1} du$$

is the incomplete beta function.

6.2.2. Improved upper limits

Similar to the case of known variance, here also it is possible to improve the upper limits in some cases. Note that Corollary 5.2.1 has no analogue in the present case (see Hume [23]). However, the following theorem gives a partial analogue of Corollary 5.2.2.

Theorem 6.2.1. If $\rho_{ij} \geq 0$ for all $i \neq j$, then

$$(6.2.3) \qquad \Pr(t_1 \leq c_1, \ldots, t_n \leq c_n) \geq \prod_{i=1}^{n} \Pr(t_i \leq c_i),$$

where t_i is given in (6.1.1) (with $\sigma=1$) and $c_1, c_2, \ldots, c_n$ are any non-negative numbers.

Proof:

$$(6.2.4) \qquad \Pr(t_1 \leq c_1, \ldots, t_n \leq c_n) = \int_0^{\infty} \Pr(b_1 \leq c_1 s_\nu, \ldots, b_n \leq c_n s_\nu) \cdot f(s_\nu) ds_\nu,$$

where $f(s_\nu)$ is the p.d.f. of s_ν. Now applying Corollary 5.2.2 to the first term of the integrand on the R.H.S. of (6.2.4), we get

$$\begin{aligned} \Pr(t_1 \leq c_1, \ldots, t_n \leq c_n) &\geq \int_0^{\infty} \prod_{i=1}^{n} \Pr(b_i \leq c_i s_\nu) f(s_\nu) ds_\nu \\ &= E\{ \prod_{i=1}^{n} \Pr(b_i \leq c_i s_\nu) \} \\ &\geq \prod_{i=1}^{n} E(\Pr(b_i \leq c_i s_\nu)) \qquad (6.2.5) \\ &= \prod_{i=1}^{n} \Pr(t_i \leq c_i). \end{aligned}$$

The inequality at (6.2.5) follows on using a result due to Kimball [25].

The analogue of Theorem 5.2.2 is given in Theorem 6.2.2 and is due to Šidák (see e.g. [38]).

Theorem 6.2.2. With the notations of Theorem 6.1.1,

$$\Pr(|t_1| \leq c_1, \ldots, |t_n| \leq c_n) \geq \prod_{i=1}^{n} \Pr(|t_i| \leq c_i).$$

Corresponding to Theorem 5.2.3, we now have

Theorem 6.2.3. The solution of

$$\Pr(|t_1|>\bar{v}'_\alpha) = 1-(1-\alpha)^{\frac{1}{n}} \tag{6.2.6}$$

gives an improved upper limit for v_α.

Similarly, if $\rho_{min} \geq 0$ then an improved upper limit for u_α is $\bar{u}'_\alpha$, where

$$\Pr(t_1>\bar{u}'_\alpha) = 1-(1-\alpha)^{\frac{1}{n}} .$$

6.2.3. Lower limits

The evaluation of the lower limits $\underline{u}_\alpha$ and $\underline{v}_\alpha$ is similar to the case of known variance with the exception that now we have to deal with the bivariate t distribution.

Denote $\Pr(t_1>h,\ t_2>k)$ by $L(h,k,\rho,\nu)$, where the joint distribution of t_1 and t_2 is given at (6.1.3). Then

$$\Pr(|t_1|>h,|t_2|>k] = 2L(h,k,\rho,\nu) + 2L(h,k,-\rho,\nu).$$

Tables prepared by Dunnett and Sobel [13] and Siotani [39] could be used to evaluate such probabilities. However, for computational purposes, it is convenient to express the double integral

$$L(h,k,\rho,\nu) = \int_k^\infty \int_h^\infty g(t_1,t_2)dt_1dt_2$$

as a single integral and to use numerical integration. To this end note that for h,k positive

$$L(h,k,\rho,\nu) = \int_0^\infty \Pr(b_1 > hs_\nu,\ b_2 > ks_\nu) \cdot f(s_\nu)ds_\nu,$$

where $f(s_\nu)$ is the p.d.f. of s_ν. Now using (5.2.12) and interchanging the order of integration, we get

$$L(h,k,\rho,\nu) = \frac{1}{2\pi} \int_{\arccos\rho}^{\pi} \int_0^\infty \exp[-\tfrac{1}{2}s^2(h^2+k^2-2hk\cos w)\operatorname{cosec}^2 w]$$

$$\cdot \frac{2\cdot(\frac{1}{2}\nu)^{\frac{1}{2}\nu}}{\Gamma(\frac{1}{2}\nu)} \cdot s^{\nu-1} \exp[-\tfrac{1}{2}\nu s^2]\, dsdw.$$

$$= \frac{1}{2\pi} \int_{\arccos\rho}^{\pi} \left[1 + \frac{h^2+k^2-2hk\cos w}{\nu \sin^2 w}\right]^{-\frac{\nu}{2}} dw.$$

For h=k, this reduces to

$$L(h,h,\rho,\nu) = \frac{1}{2\pi} \int_{\arccos\rho}^{\pi} \left[1 + \frac{2h^2}{\nu(1+\cos w)}\right]^{-\frac{\nu}{2}} dw.$$

6.3. Performance of test statistics

To study the performance of U and V, we shall proceed similarly to the case of known variance. From equations (5.3.1) and (6.1.1), it is clear that under H_k, we can write

$$t_i = \frac{b_i}{s_\nu} = \frac{z+\delta_{ik}\theta}{s_\nu},$$

where δ_{ik} is given by equation (5.3.2) and z has a standard unit normal distribution independent of s_ν. Now letting

$$\Delta_{ik} = \delta_{ik}\theta, \tag{6.3.1}$$

it follows that under H_k, t_i has a noncentral t distribution with ν

degrees of freedom and noncentrality parameter Δ_{ik}. Note that the notation Δ_{ik} is in agreement with equation (4.4.3) for $\sigma=1$.

One-sided statistic U. From equation (4.4.4)

$$P_k = \Pr(t_k > u_\alpha | H_k), \quad k=1,2,\ldots,n$$

$$= \Pr(t'_{\nu,\Delta_{kk}} > u_\alpha), \tag{6.3.2}$$

where $t'_{\nu,\Delta_{kk}}$ has the noncentral t distribution with ν d.f. and noncentrality parameter $\Delta_{kk} = \lambda_{kk}^{\frac{1}{2}}\theta$. Tables of the noncentral t distribution are needed to evaluate P_k. However, the existing tables (see e.g. [27]) require a considerable amount of interpolation and one may use the normal approximation to the c.d.f. of noncentral t distribution (see e.g. [24] and [1] p. 949) given by

$$\Pr(t'_{\nu,\Delta} \leq t_o) \doteq \Phi\left(\frac{t_o\left(1 - \frac{1}{4\nu}\right) - \Delta}{\left(1 + \frac{t_o^2}{2\nu}\right)^{\frac{1}{2}}}\right). \tag{6.3.3}$$

This gives a reasonably good approximation. Thus, for example, for $\nu=5$, $\Delta=(\nu+1)^{\frac{1}{2}}$ and $t_o=5$, the approximate value from (6.3.3) is .891 while the exact value is .900. The accuracy of the approximation improves with increasing ν.

To find P_a, note that P_k can be rewritten as

$$P_k = \int_0^\infty \Pr(z > u_\alpha s_\nu - \delta_{kk}\theta) f(s_\nu) ds_\nu$$

$$= \int_0^\infty \Phi(-u_\alpha s_\nu + \lambda_{kk}^{\frac{1}{2}}\theta) f(s_\nu) ds_\nu ,$$

where $f(s_\nu)$ is the p.d.f. of s_ν. For $\theta \geq 0$, the minimum of $\Phi(-u_\alpha s_\nu + \lambda_{kk}^{\frac{1}{2}}\theta)$ occurs when λ_{kk} is a minimum and hence

(6.3.4) $$P_a = \underset{k}{\text{Min}}\, P_k = \Pr(t'_{\nu,\Delta} > u_\alpha),$$

where

(6.3.5) $$\Delta = \lambda_{(1)}^{\frac{1}{2}}\theta, \quad \lambda_{(1)} = \underset{k}{\text{Min}}\, \lambda_{kk}.$$

Similarly from (4.4.10)

$$\bar{Q}_k = P_k + \sum_{\substack{i=1 \\ i\neq k}}^{n} \Pr(t_i > u_\alpha \mid H_k)$$

$$= P_k + \sum_{\substack{i=1 \\ i\neq k}}^{n} \Pr(t'_{\nu,\Delta_{ik}} > u_\alpha).$$

The lower bound $\underline{Q}_k$ given at (4.4.11) involves the bivariate noncentral t distribution and is not considered here.

Two-sided statistic V. Now

$$P_k = \Pr(|t_k| > v_\alpha \mid H_k), \quad k=1,2,\ldots,n$$

$$= \Pr(|t'_{\nu,\Delta_{kk}}| > v_\alpha)$$

$$= \int_0^\infty \Pr(|z+\delta_{kk}\theta| > v_\alpha s_\nu) f(s_\nu)\, ds_\nu$$

$$= \int_0^\infty [\Phi(-v_\alpha s_\nu + \delta_{kk}\theta) + \Phi(-v_\alpha s_\nu - \delta_{kk}\theta)] \cdot f(s_\nu)\, ds_\nu .$$

Now using Lemma 5.3.1, we have

(6.3.6) $$P_a = \underset{k}{\text{Min}}\, P_k = \Pr(|t'_{\nu,\Delta}| > v_\alpha),$$

where Δ is given at (6.3.5).

Expression for the upper bound for Q_k is similar.

6.4. Applications

For Example 4.5.1, the true percentage points of

$$C_1 = \underset{i}{\text{Max}} \left(\frac{y_i - \bar{y}}{s_\nu}\right) = \left(\frac{n-1}{n}\right)^{\frac{1}{2}} U$$

have been tabulated by David [9] for n=3(1)12 and several values of ν and α. Further, the performance P_a has been studied by David and Paulson [10].

The upper and lower limits for the true percentage points of

$$C_2 = \underset{i}{\text{Max}} \left|\frac{y_i - \bar{y}}{s_\nu}\right| = \left(\frac{n-1}{n}\right)^{\frac{1}{2}} V$$

are given by Halperin et al. [20] for α=.05, .01 and several values of n and ν. As for the case of known variance, the upper limits can be slightly improved by using (6.2.6).

We now turn our attention to the general regression model. From equations (6.2.1) and (6.2.6), it follows that the upper limits depend only on α,n and ν. To fix the ideas, we consider the two-sided statistic V and use the improved upper limit $\bar{v}'_\alpha$. Equation (6.3.6) then reduces to

$$P_a = \Pr(|t'_{\nu,\Delta}| > \bar{v}'_\alpha), \tag{6.4.1}$$

where Δ is given at (6.3.5). This shows that for fixed α,n and ν, P_a depends only on Δ and hence is a conservative measure of performance. This is given in Table 6.4.2 for n=4(2)10; ν=5(5)20 and α=.05. For purposes of comparison P_a for σ known ($\nu=\infty$) has also been included. The values of $\bar{v}'_\alpha$ are tabulated in Table 6.4.1 for the above values of n,ν and α.

For the numerical work involving the noncentral t distribution, the series expansion for $\Pr(|t'_{\nu,\Delta}| \le t_o)$, originally due to Craig [7] was found most convenient. However, as pointed out by Amos [2], Craig's formula contains an error. The correct formula is

$$(6.4.2)\quad \Pr(|t'_{\nu,\Delta}| \le t_o) = e^{-\frac{1}{2}\Delta^2} \cdot \sum_{j=0}^{\infty} \frac{(\frac{1}{2}\Delta^2)^j}{j!} I_{\frac{t_o^2}{t_o^2+\nu}}(j+\tfrac{1}{2}, \tfrac{1}{2}\nu),$$

where $I_x(p,q)$ is given at (6.2.2). From (6.4.2) we get

$$(6.4.3)\quad \Pr(|t'_{\nu,\Delta}| > t_o) = e^{-\frac{1}{2}\Delta^2} \cdot \sum_{j=0}^{\infty} (\tfrac{1}{2}\Delta^2)^j \cdot \frac{1}{j!} I_{\frac{\nu}{\nu+t_o^2}}(\tfrac{1}{2}\nu, j+\tfrac{1}{2}).$$

This series was used for the numerical work. The incomplete beta functions were obtained by using the quadrature method given in the Appendix and the series was summed to give at least 3-decimal place accuracy.

The normal approximation (6.3.3) was also studied in this connection. For all the values tabulated, it was found that the approximation is quite good for all $\nu \ge 10$ and in this case the largest difference between the tabled and the approximate values was .003. Even for $\nu=5$, the largest difference was .012. Further, for fixed ν, the approximation appeared to be less accurate for large values of $\bar{v}'_\alpha$.

Note that Table 6.4.2 can be used to get any P_k and hence P_b. However, the normal approximation (6.3.3) is recommended for $\nu \ge 10$.

Table 6.4.1. Improved upper limits $\bar{v}'_\alpha$ from equation (6.2.6) for α=.05

ν \ n	4	6	8	10
5	3.791	4.197	4.501	4.747
10	3.027	3.264	3.434	3.568
15	2.827	3.026	3.166	3.275
20	2.736	2.918	3.045	3.143
∞	2.491	2.631	2.727	2.800

Table 6.4.2. Performance P_a given at equation (6.4.1) for the statistic V_2 for α=.05

n	ν \ Δ	1	2	3	4	5	6
4	5	.040	.147	.355	.610	.819	.936
	10	.051	.215	.519	.809	.954	.994
	15	.056	.244	.579	.861	.975	.998
	20	.059	.260	.608	.883	.982	.999
	∞	.068	.312	.695	.934	.994	1.000
6	5	.028	.108	.278	.513	.737	.887
	10	.037	.169	.444	.748	.928	.988
	15	.041	.196	.510	.816	.961	.996
	20	.043	.212	.544	.845	.973	.998
	∞	.052	.264	.644	.914	.991	1.000
8	5	.022	.086	.231	.447	.672	.842
	10	.029	.141	.393	.700	.905	.981
	15	.033	.167	.462	.779	.948	.993
	20	.035	.182	.498	.815	.963	.996
	∞	.042	.234	.608	.898	.988	.999
10	5	.018	.072	.199	.398	.619	.801
	10	.024	.122	.356	.661	.883	.975
	15	.027	.147	.426	.749	.936	.991
	20	.029	.161	.464	.789	.955	.995
	∞	.036	.212	.579	.885	.986	.999

CHAPTER VII

DISTRIBUTION THEORY WHEN VARIANCE IS UNKNOWN — POOLED STUDENTIZATION

7.1. Introduction

The statistics, which we will consider here are U_3 and V_3 as defined at (4.1.10). Let

$$w_i = \frac{e_i}{\lambda_{ii}^{\frac{1}{2}} S_p}, \quad i=1,2,\ldots,n, \tag{7.1.1}$$

where

$$S_p^2 = S^2 + \nu s_\nu^2 \tag{7.1.2}$$

is the pooled sum of squares based on

$$p = n-m+\nu \tag{7.1.3}$$

degrees of freedom. Note that the assumption (4.1.7) implies that $p \geq 2$. Dropping the suffixes from U_3 and V_3, we have from equation (4.1.10)

$$U = \underset{i}{\text{Max}}\ w_i, \quad V = \underset{i}{\text{Max}}\ |w_i| . \tag{7.1.4}$$

Whereas the results of Chapter VI are similar to that of Chapter V, the situation is quite different in the present case. This is due to the fact that the statistics U and V are the maximum of n bounded random variables. As we shall see in Section 7.5, this

allows us to evaluate the true percentage points in some cases.

The necessary distribution theory results are given in Sections 7.2 and 7.3. In Section 7.4, we give an analogue of Corollary 5.2.1 for a special case. The percentage points of U and V are considered in Section 7.5 and some results regarding the performance are discussed in Section 7.6. A comparison with external studentization is made in Section 7.7.

7.2. Marginal and joint distributions

We first obtain the marginal distribution of w_i. Using the Cochran's theorem, it is easy to show that

$$S_p^2 = \frac{e_i^2}{\lambda_{ii}} + (S^2 + \nu s_\nu^2 - \frac{e_i^2}{\lambda_{ii}}),$$

where

$$\frac{e_i^2}{\lambda_{ii}} \overset{d}{=} \sigma^2 \chi_1^2$$

and

$$S^2 + \nu s_\nu^2 - \frac{e_i^2}{\lambda_{ii}} \overset{d}{=} \sigma^2 \chi_{p-1}^2$$

and the two quadratic forms are independent. Hence

$$w_i^2 = \frac{e_i^2}{\lambda_{ii}} \Big/ S_p^2 \tag{7.2.1}$$

has a beta $(\frac{1}{2}, \frac{1}{2}(p-1))$ distribution. From this we get the marginal distribution of w_i

$$g(w_i) = \frac{1}{B(\frac{1}{2},\frac{1}{2}(p-1))} \cdot (1-w_i^2)^{\frac{1}{2}(p-3)}, \quad -1 \leq w_i \leq 1. \tag{7.2.2}$$

Next, consider the joint distribution of, say, w_1 and w_2. By assumption (4.1.7), we see that the matrix

$$\begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{12} & \lambda_{22} \end{bmatrix}$$

is positive definite. Put

$$Q_1 = (e_1 \quad e_2)\begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{12} & \lambda_{22} \end{bmatrix}^{-1}\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

$$= \frac{\lambda_{22}e_1^2 - 2\lambda_{12}e_1e_2 + \lambda_{11}e_2^2}{\lambda_{11}\lambda_{22} - \lambda_{12}^2}$$

and

$$Q_2 = S_p^2 - Q_1.$$

Case 1. $p \geq 3$

Again by using the Cochran's theorem, we see that Q_1 and Q_2 are independent $\sigma^2\chi^2$ variates with 2 and (p-2) degrees of freedom respectively. Further, Q_1 can be written as

$$Q_1 = z_1^2 + z'^2,$$

where

(7.2.3) $$z_i = e_i/\lambda_{ii}^{\frac{1}{2}}, \quad i=1,2,$$

$$z' = (z_2 - \rho z_1)/(1-\rho^2)^{\frac{1}{2}}$$

and

$$\rho = \rho_{12}.$$

Thus, we have the decomposition

$$S_p^2 = z_1^2 + z'^2 + Q_2,$$

where z_1^2, z'^2 and Q_2 are mutually independent $\sigma^2\chi^2$ variates with 1,1

and (p-2) degrees of freedom respectively. This implies that the joint distribution of z_1, z' and Q_2 is

$$g(z_1,z',Q_2) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(z_1^2 + z'^2)} \cdot \frac{1}{(2\sigma^2)^{\frac{1}{2}(p-2)} \cdot \Gamma(\frac{1}{2}(p-2))} \cdot Q_2^{\frac{1}{2}(p-4)} e^{-\frac{Q_2}{2\sigma^2}}.$$

Making the transformation

$$z_1 = z_1,\ z_2 = \rho z_1 + (1-\rho^2)^{\frac{1}{2}} z',\ Q_2 = Q_2$$

we get

$$g(z_1,z_2,Q_2) = C \cdot Q_2^{\frac{1}{2}(p-4)} \cdot e^{-\frac{1}{2\sigma^2}(Q_2 + \frac{z_1^2 - 2\rho z_1 z_2 + z_2^2}{1-\rho^2})},$$

where

$$C^{-1} = \pi(1-\rho^2)^{\frac{1}{2}}(2\sigma^2)^{\frac{1}{2}p} \cdot \Gamma(\tfrac{1}{2}(p-2)). \tag{7.2.4}$$

To get the joint distribution of w_1 and w_2 we note that

$$w_i = \frac{z_i}{(\frac{z_1^2 - 2\rho z_1 z_2 + z_2^2}{1-\rho^2} + Q_2)^{\frac{1}{2}}}, \quad i=1,2$$

$$= \frac{z_i}{(\underline{z}' P^{-1} \underline{z} + Q_2)^{\frac{1}{2}}},$$

where

$$\underline{z}' = (z_1,z_2), \quad P = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Make a transformation

$$w_i = \frac{z_i}{(\underline{z}'P^{-1}\underline{z}+Q_2)^{\frac{1}{2}}}, \quad i=1,2$$

$$Q_2 = Q_2.$$

The jacobian of the transformation, after some simplifications, is

$$\frac{\partial(z_1,z_2,Q_2)}{\partial(w_1,w_2,Q_2)} = \frac{Q_2}{(1-\underline{w}'P^{-1}\underline{w})^2},$$

where

$$\underline{w}' = (w_1,w_2).$$

Therefore, the joint distribution of w_1, w_2 and Q_2 is

$$g(w_1,w_2,Q_2) = \frac{C \cdot Q_2^{\frac{1}{2}(p-2)} \cdot e^{-\frac{Q_2}{2\sigma^2(1-\underline{w}'P^{-1}\underline{w})}}}{(1-\underline{w}'P^{-1}\underline{w})^2}$$

Now integrating out Q_2 from 0 to ∞, we get

$$g(w_1,w_2) = \frac{p-2}{2\pi(1-\rho^2)^{\frac{1}{2}}} \cdot (1-\underline{w}'P^{-1}\underline{w})^{\frac{1}{2}(p-4)}. \tag{7.2.5}$$

The region of positive density is the interior of the ellipse

$$\underline{w}'P^{-1}\underline{w} = 1,$$

i.e.,

$$w_1^2-2\rho w_1w_2+w_2^2 = 1-\rho^2 . \tag{7.2.6}$$

Case 2. p=2

Using the same notations as in case 1, we now have

$$w_i = \frac{z_i}{(\underline{z}'P^{-1}\underline{z})^{\frac{1}{2}}}, \quad i=1,2.$$

It is easy to show that

$$\underline{w}'P^{-1}\underline{w} \equiv 1$$

and hence the joint density of w_1, w_2 can not exist. However, the marginal distribution of w_i is still given by (7.2.2).

Note that the R.H.S. of (7.2.5) vanishes on putting p=2. Hence we will continue to use (7.2.5), keeping in mind that no joint density exists for p=2.

Remark 1. (7.2.5) generalizes a result obtained by Doornbos et al. (see e.g. Doornbos [11]) for the special case of Example 4.5.1. Their result has been re-derived by Quesenberry and David [32] by a simplified argument.

Remark 2. By using the inverse transformation

$$w_i = \frac{t_i}{(p-2 + \underline{t}'P^{-1}\underline{t})^{\frac{1}{2}}}, \quad i=1,2,$$

where $\underline{t}' = (t_1,t_2)$, it can be shown that the joint distribution of t_1 and t_2 is bivariate t with p-2 degrees of freedom. In view of this result and the corresponding result between univariate beta and univariate t distributions, we may call the joint distribution of w_1^2 and w_2^2 a generalized bivariate beta distribution.

Next consider the ellipse

$$w_1^2 - 2\rho w_1 w_2 + w_2^2 = 1-\rho^2$$

in the (w_1,w_2)-plane. The principal axes of the ellipse are

$$w_1-w_2 = 0, \quad w_1+w_2 = 0$$

and the axis $w_1-w_2 = 0$ intersects the ellipse at the points (c_o,c_o) and $(-c_o,-c_o)$, where $c_o = [\frac{1}{2}(1+\rho)]^{\frac{1}{2}}$ (see Figure 7.2.1). We now have the following lemma (see also Srikantan [40]).

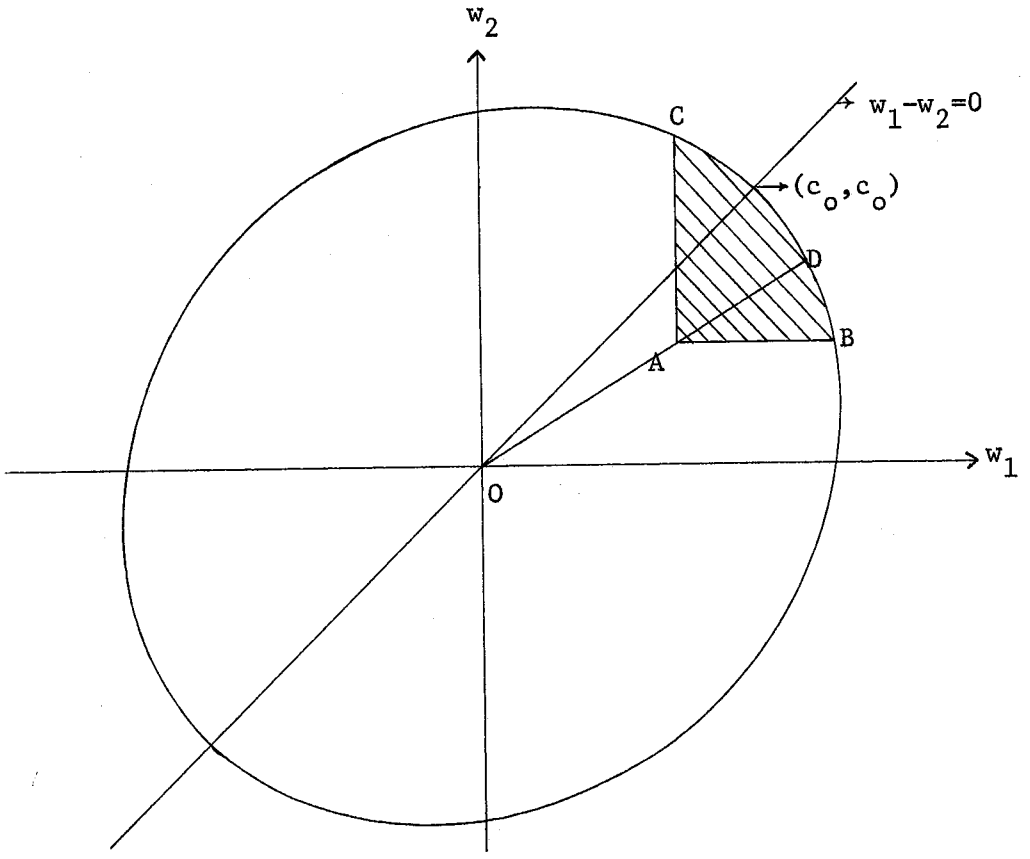


Figure 7.2.1

Lemma 7.2.1. Let the joint distribution of w_1 and w_2 be given by (7.2.5). Then

(a) For any $c \geq [\frac{1}{2}(1+\rho)]^{\frac{1}{2}}$

$$\Pr(w_1>c,\ w_2>c) = 0.$$

(b) For any $c' \geq [\frac{1}{2}(1+|\rho|)]^{\frac{1}{2}}$

$$\Pr(|w_1|>c',\ |w_2|>c') = 0.$$

Proof: Result (a) is obvious. To prove (b), let

$$M(h,k,\rho,p) = \Pr(w_1>h,\ w_2>k). \tag{7.2.7}$$

Then

$$\Pr(|w_1|>h,\ |w_2|>k) = 2\,M(h,k,\rho,p) + 2\,M(h,k,-\rho,p). \tag{7.2.8}$$

Setting h=k=c' in (7.2.8) and applying result (a), we see that both terms appearing on the R.H.S. of (7.2.8) will vanish, provided that

$$c' \geq \mathrm{Max}([\tfrac{1}{2}(1+\rho)]^{\frac{1}{2}},\ [\tfrac{1}{2}(1-\rho)]^{\frac{1}{2}}) = [\tfrac{1}{2}(1+|\rho|)]^{\frac{1}{2}}.$$

We are now in a position to prove the stronger version of equation (4.2.4) as mentioned in Remark 1 of Section 4.2.

Theorem 7.2.1. Let Λ be given by (4.1.3). Then

$$|\rho_{ij}| \leq \frac{2}{\lambda_{ii}+\lambda_{jj}} - 1, \quad i\neq j. \tag{7.2.9}$$

Proof: From (7.1.1) and (7.1.2) we have

$$\sum_{i=1}^{n} \lambda_{ii} w_i^2 = \frac{\sum_{i=1}^{n} e_i^2}{S^2+\nu s_\nu^2} = \frac{S^2}{S^2+\nu s_\nu^2} .$$

Hence

$$\sum_{i=1}^{n} \lambda_{ii} w_i^2 \leq 1. \tag{7.2.10}$$

Without any loss of generality we now take i=1 and j=2. Then (7.2.10) yields.

$$\lambda_{11} w_1^2 + \lambda_{22} w_2^2 \leq 1. \tag{7.2.11}$$

The equation

$$\lambda_{11} w_1^2 + \lambda_{22} w_2^2 = 1 \tag{7.2.12}$$

describes an ellipse in the (w_1,w_2)-plane (see Figure 7.2.2) and the line $w_1-w_2 = 0$ intersects the ellipse at the points (d,d) and (-d,-d), where

$$d = \frac{1}{(\lambda_{11}+\lambda_{22})^{\frac{1}{2}}} . \tag{7.2.13}$$

Similarly, the line $w_1+w_2 = 0$ intersects the ellipse (7.2.12) at the points (-d,d) and (d,-d).

By (7.2.11), it then follows that

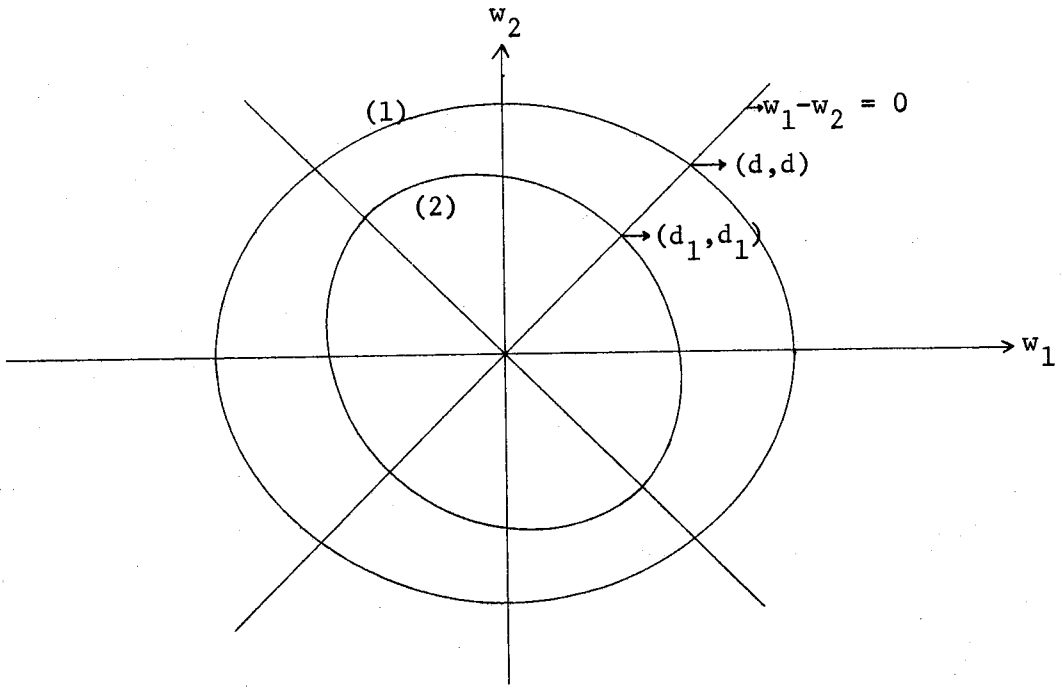
$$\Pr(|w_1|>d, \ |w_2|>d) = 0. \tag{7.2.14}$$

Further from Lemma 7.2.1

$$\Pr(|w_1|>d_1, \ |w_2|>d_1) = 0,$$

where

$$d_1 = [\tfrac{1}{2}(1+|\rho_{12}|)]^{\frac{1}{2}}.$$



Ellipse (1): $\lambda_{11}w_1^2 + \lambda_{22}w_2^2 = 1$

Ellipse (2): $w_1^2-2\rho_{12}w_1w_2 + w_2^2 = 1-\rho_{12}^2$.

Figure 7.2.2

We now claim that $d \geq d_1$; for otherwise equation (7.2.14) will contradict the fact that the joint density of w_1, w_2 is positive inside the ellipse (7.2.6). Hence

$$\frac{1}{(\lambda_{11}+\lambda_{22})^{\frac{1}{2}}} \geq [\tfrac{1}{2}(1+|\rho_{12}|)]^{\frac{1}{2}},$$

i.e.,

$$|\rho_{12}| \leq \frac{2}{\lambda_{11}+\lambda_{22}} - 1.$$

7.3. Evaluation of bivariate probability

Before proceeding any further, we will obtain an expression for the bivariate probability in terms of a single integral. This will be used for the evaluation of lower limits of the true percentage points. From equation (7.2.7)

$$M(h,k,\rho,p) = \Pr(w_1>h,\ w_2>k)$$

$$(7.3.1) \qquad = \frac{p-2}{2\pi(1-\rho^2)^{\frac{1}{2}}} \iint \left(1 - \frac{w_1^2-2\rho w_1 w_2+w_2^2}{1-\rho^2}\right)^{\frac{1}{2}(p-4)} dw_2 dw_1,$$

where the integration is over the region

$$(7.3.2) \qquad w_1>h,\ w_2>k,\ w_1^2-2\rho w_1 w_2+w_2^2 \leq 1-\rho^2 .$$

Define

$$(7.3.3) \qquad Q(a) = \Pr(w_1>a)$$

$$= K_1 \int_a^1 (1-w_1^2)^{\frac{1}{2}(p-3)} dw_1$$

where

$$(7.3.4) \qquad K_1 = \frac{1}{B(\frac{1}{2},\frac{1}{2}(p-1))} .$$

The following properties of the M function are obvious:

$$M(h,k,\rho,p) = M(k,h,\rho,p),$$

$$M(-h,k,\rho,p) + M(h,k,-\rho,p) = Q(k),$$

$$M(-h,-k,\rho,p) = 1-Q(h)-Q(k)+M(h,k,\rho,p).$$

From these relations, it is clear that we only need to consider the case $h,k \geq 0$. Let A be the point (h,k) in the (w_1,w_2)-plane. We

assume that A lies inside the ellipse

$$w_1^2-2\rho w_1 w_2+w_2^2 = 1-\rho^2 \quad , \tag{7.3.5}$$

for otherwise, the desired probability is zero. The region of integration given at (7.3.2) is then the shaded area ABCA (see Figure 7.2.1), where

$$B = (\rho k + (1-\rho^2)^{\frac{1}{2}}(1-k^2)^{\frac{1}{2}},\ k)$$

and

$$C = (h, \rho h + (1-\rho^2)^{\frac{1}{2}}(1-h^2)^{\frac{1}{2}}).$$

The extended line OA intersects the ellipse at the point D, where

$$D = (h(\frac{1-\rho^2}{h^2-2\rho hk+k^2})^{\frac{1}{2}},\quad k(\frac{1-\rho^2}{h^2-2\rho hk+k^2})^{\frac{1}{2}}).$$

If we now define

$$M_1(h,k,\rho,p) = \Pr((w_1,w_2) \in ABDA),$$

then due to symmetry

$$M(h,k,\rho,p) = M_1(h,k,\rho,p) + M_1(k,h,\rho,p). \tag{7.3.6}$$

Hence we only have to find an expression for $M_1(h,k,\rho,p)$. We now assume that $k>0$, for $M_1(h,0,\rho,p) = 0$. Make a transformation

$$z_1 = \frac{w_1-\rho w_2}{(1-\rho^2)^{\frac{1}{2}}}$$

$$z_2 = w_2.$$

Then the joint distribution of z_1, z_2 is

$$g(z_1,z_2) = \frac{p-2}{2\pi}\,(1-z_1^2-z_2^2)^{\frac{1}{2}(p-4)},\quad z_1^2+z_2^2 \leq 1$$

and the region of integration for $M_1(h,k,\rho,p)$ is the shaded region $A_1B_1D_1A_1$ (Figure 7.3.1), where

$$A_1 = \left(\frac{h-\rho k}{(1-\rho^2)^{\frac{1}{2}}},\ k\right),$$

$$B_1 = \left((1-k^2)^{\frac{1}{2}},\ k\right),$$

$$D_1 = \left(\frac{h-\rho k}{(h^2-2\rho hk+k^2)^{\frac{1}{2}}},\ k\left(\frac{1-\rho^2}{h^2-2\rho hk+k^2}\right)^{\frac{1}{2}}\right).$$

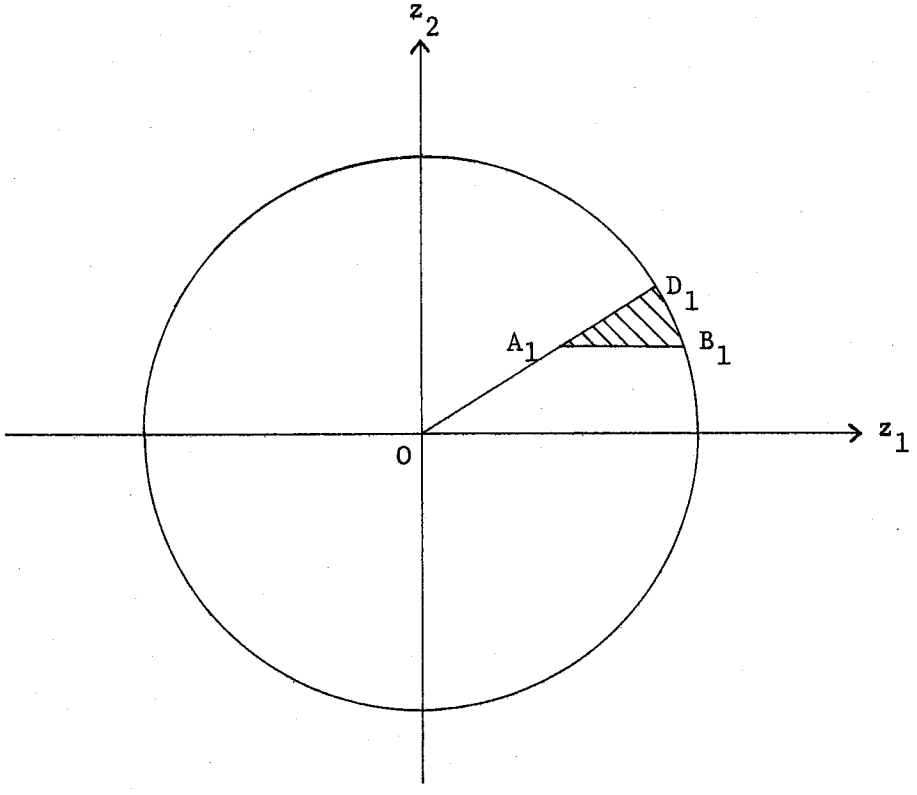


Figure 7.3.1

Put

$$r \cos \theta = z_1$$

and

$$r \sin \theta = z_2.$$

Then, the joint distribution of r and θ is

$$g(r,\theta) = \frac{p-2}{2\pi} \cdot r\,(1-r^2)^{\frac{1}{2}(p-4)}, \quad 0 \leq \theta < 2\pi, \quad 0 \leq r \leq 1.$$

Moreover

$$M_1(h,k,\rho,p) = \int_{\arctan(k/(1-k^2)^{\frac{1}{2}})}^{\arctan \frac{k(1-\rho^2)^{\frac{1}{2}}}{h-\rho k}} \int_{k \operatorname{cosec} \theta}^{1} \frac{p-2}{2\pi} r(1-r^2)^{\frac{1}{2}(p-4)} dr d\theta$$

$$= \frac{1}{2\pi} \int_{\arctan(k/(1-k^2)^{\frac{1}{2}})}^{\arctan \frac{k(1-\rho^2)^{\frac{1}{2}}}{h-\rho k}} [1-k^2 \operatorname{cosec}^2 \theta]^{\frac{1}{2}(p-2)} d\theta.$$

On putting

$$\tan \theta = \frac{k(1-u^2)^{\frac{1}{2}}}{h-uk},$$

this can be rewritten as

$$M_1(h,k,\rho,p) = \frac{1}{2\pi} \int \frac{k(k-uh)}{h^2+k^2-2uhk} \cdot \frac{1}{(1-u^2)^{\frac{1}{2}}} \left(1 - \frac{h^2+k^2-2uhk}{1-u^2}\right)^{\frac{1}{2}(p-2)} du,$$

where the range of integration is

$$hk - (1-h^2)^{\frac{1}{2}}(1-k^2)^{\frac{1}{2}} \leq u \leq \rho \,. \tag{7.3.7}$$

The expression for $M_1(k,h,\rho,p)$ is similar, with h and k interchanged (valid for h>0). Note that the range of integration for $M_1(k,h,\rho,p)$ is same as (7.3.7). Hence on using (7.3.6)

$$M(h,k,\rho,p) = \frac{1}{2\pi}\int \frac{1}{(1-u^2)^{\frac{1}{2}}}\left(1 - \frac{h^2+k^2-2uhk}{1-u^2}\right)^{\frac{1}{2}(p-2)}du,$$

where the range of integration is as in (7.3.7). Putting

$$u = \cos w$$

we get for h>0, k>0

$$(7.3.8)\quad M(h,k,\rho,p) = \frac{1}{2\pi}\int_{\arccos\rho}^{\arccos(hk-(1-h^2)^{\frac{1}{2}}(1-k^2)^{\frac{1}{2}})} \left(1 - \frac{h^2+k^2-2hk\cos w}{\sin^2 w}\right)^{\frac{1}{2}(p-2)}dw.$$

It is easy to see that this expression is also valid when either h or k is equal to 0. In fact, this is valid even for h=k=0. To this end note that

$$\begin{aligned} M(0,0,\rho,p) &= \Pr(w_1>0,\ w_2>0) \\ &= \Pr\left(\frac{e_1}{\lambda_{11}^{\frac{1}{2}}\cdot S_p} > 0,\ \frac{e_2}{\lambda_{22}^{\frac{1}{2}}\cdot S_p} > 0\right) \\ &= \Pr(u_1>0,\ u_2>0), \end{aligned}$$

where u_1 and u_2 have a $N(0,0,1,1,\rho)$ distribution. Hence

$$(7.3.9)\qquad M(0,0,\rho,p) = \tfrac{1}{4} + \frac{1}{2\pi}\arcsin\rho$$

and this is the value we obtain from equation (7.3.8) as well.

7.4. A probability inequality

We now turn to a probability inequality, which gives a restricted analogue to Corollary 5.2.1.

Theorem 7.4.1. Let the joint distribution of w_1 and w_2 be given by (7.2.5). If $\rho \leq 0$, then

$$\Pr(w_1 \leq c_1,\ w_2 \leq c_2) \leq \prod_{i=1}^{2} \Pr(w_i \leq c_i), \tag{7.4.1}$$

provided both c_1 and c_2 are of the same sign.

Remark 1. This theorem has been proved by Doornbos et al. (see Doornbos [11]) for the special case of Example 4.5.1. The proof given here follows on parallel lines.

Proof: First consider the case when both c_1 and c_2 are negative. Let A be the point (c_1, c_2) (see Figure 7.4.1), which lies within the ellipse

$$w_1^2 - 2\rho w_1 w_2 + w_2^2 = 1-\rho^2,\ \rho \leq 0.$$

Then the region $(w_1 \leq c_1,\ w_2 \leq c_2)$ intersecting with the ellipse is ABCA, where

$$B = (\rho c_2 - (1-\rho^2)^{\frac{1}{2}}(1-c_2^2)^{\frac{1}{2}},\ c_2)$$

and

$$C = (c_1,\ \rho c_1 - (1-\rho^2)^{\frac{1}{2}}(1-c_1^2)^{\frac{1}{2}}).$$

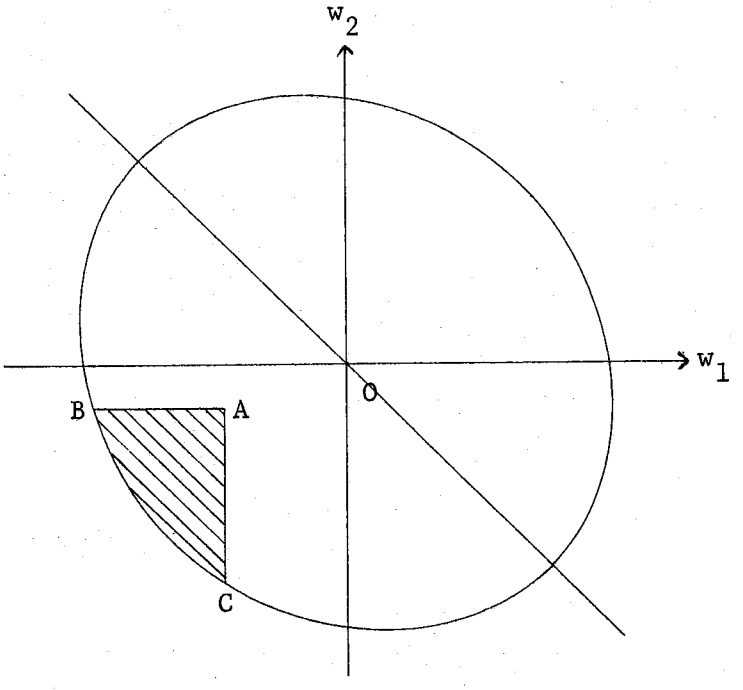


Figure 7.4.1

Now consider the function

$$\phi(c_1,c_2) = \Pr(w_1 \leq c_1)\cdot\Pr(w_2 \leq c_2) - \Pr(w_1 \leq c_1, w_2 \leq c_2). \tag{7.4.2}$$

We shall show that

$$\phi(c_1,c_2) \geq 0,$$

which will prove the result.

From (7.2.2), we see that

$$\Pr(w_i \leq c_i) = K_1 \int_{-1}^{c_i} (1-w_i^2)^{\frac{1}{2}(p-3)} dw_i, \quad i=1,2, \tag{7.4.3}$$

where

$$K_1 = \frac{1}{B(\frac{1}{2},\frac{1}{2}(p-1))} \cdot \tag{7.4.4}$$

Define

$$h(t) = \rho t-(1-\rho^2)^{\frac{1}{2}}(1-t^2)^{\frac{1}{2}}, \quad -1\leq t\leq 0. \tag{7.4.5}$$

Then

$$\Pr(w_1\leq c_1,\ w_2\leq c_2) = \int_{h(c_2)}^{c_1}\int_{h(w_1)}^{c_2} g(w_1,w_2)\,dw_2\,dw_1, \tag{7.4.6}$$

where $g(w_1,w_2)$ is given at (7.2.5). Now making a transformation

$$w_1 = w_1$$

and

$$w_2'(w_1,w_2) = \frac{w_2-\rho w_1}{(1-\rho^2)^{\frac{1}{2}}(1-w_1^2)^{\frac{1}{2}}}, \tag{7.4.7}$$

we get the joint distribution of w_1, w_2' as

$$g_1(w_1,w_2') = \frac{p-2}{2\pi}(1-w_1^2)^{\frac{1}{2}(p-3)}\cdot(1-w_2'^2)^{\frac{1}{2}(p-4)}$$

$$= K_1(1-w_1^2)^{\frac{1}{2}(p-3)}\cdot K_2(1-w_2'^2)^{\frac{1}{2}(p-4)}, \tag{7.4.8}$$

where

$$K_2 = \frac{1}{B(\frac{1}{2},\frac{1}{2}(p-2))}$$

and K_1 is given at (7.4.4). Also (7.4.6) reduces to

$$\Pr(w_1\leq c_1,\ w_2\leq c_2) = \int_{h(c_2)}^{c_1}\int_{-1}^{w_2'(w_1,c_2)} g_1(w_1,w_2')\,dw_2'\,dw_1, \tag{7.4.9}$$

where $w_2'(w_1,c_2)$ is obtained from (7.4.7). Substituting from (7.4.3) and (7.4.9) in (7.4.2), we get

$$\phi(c_1,c_2) = K_1 \int_{-1}^{c_1} (1-w_1^2)^{\frac{1}{2}(p-3)} dw_1 \cdot K_1 \int_{-1}^{c_2} (1-w_2^2)^{\frac{1}{2}(p-3)} dw_2$$

(7.4.10)

$$- K_1 \int_{h(c_2)}^{c_1} [(1-w_1^2)^{\frac{1}{2}(p-3)} \cdot K_2 \int_{-1}^{w_2'(w_1,c_2)} (1-w_2'^2)^{\frac{1}{2}(p-4)} dw_2'] \, dw_1 \, .$$

We now treat it as a function of c_1,c_2 for $c_1 \leq 0$, $c_2 \leq 0$. Differentiating $\phi(c_1,c_2)$ with respect to c_1, we get

$$\frac{\partial\phi(c_1,c_2)}{\partial c_1} = K_1(1-c_1^2)^{\frac{1}{2}(p-3)} \cdot K_1 \int_{-1}^{c_2} (1-w_2^2)^{\frac{1}{2}(p-3)} dw_2$$

(7.4.11)
$$- K_1(1-c_1^2)^{\frac{1}{2}(p-3)} \cdot K_2 \int_{-1}^{w_2'(c_1,c_2)} (1-w_2'^2)^{\frac{1}{2}(p-4)} dw_2',$$

where

(7.4.12)
$$w_2'(c_1,c_2) = \frac{c_2-\rho c_1}{(1-\rho^2)^{\frac{1}{2}}(1-c_1^2)^{\frac{1}{2}}} \, .$$

Writing

(7.4.13)
$$\phi_1(c_1,c_2) = K_1 \int_{-1}^{c_2} (1-w_2^2)^{\frac{1}{2}(p-3)} dw_2$$

$$- K_2 \int_{-1}^{w_2'(c_1,c_2)} (1-w_2'^2)^{\frac{1}{2}(p-4)} dw_2',$$

(7.4.11) can be rewritten as

(7.4.14)
$$\frac{\partial\phi(c_1,c_2)}{\partial c_1} = K_1(1-c_1^2)^{\frac{1}{2}(p-3)} \cdot \phi_1(c_1,c_2).$$

Now differentiating $w_2'(c_1,c_2)$ with respect to c_1, we have

$$\frac{\partial w_2'(c_1,c_2)}{\partial c_1} = \frac{c_1c_2 - \rho}{(1-\rho^2)^{\frac{1}{2}}(1-c_1^2)^{3/2}} \quad . \tag{7.4.15}$$

Since both c_1 and c_2 are negative and $\rho \leq 0$, hence by (7.4.15)

$$\frac{\partial w_2'(c_1,c_2)}{\partial c_1} \geq 0.$$

This implies that $w_2'(c_1,c_2)$ is a nondecreasing function of c_1 for fixed c_2. From (7.4.13), it then follows that $\phi_1(c_1,c_2)$ is a non-increasing function of c_1 for fixed c_2. Also

$$\phi_1(0,0) = K_1 \int_{-1}^{0} (1-w_2^2)^{\frac{1}{2}(p-3)} dw_2 - K_2 \int_{-1}^{0} (1-w_2'^2)^{\frac{1}{2}(p-4)} dw_2'$$

$$= \tfrac{1}{2} - \tfrac{1}{2} = 0.$$

Hence $\phi_1(c_1,0) \geq 0$ on the line PO (Figure 7.4.2). In view of (7.4.14), this implies that $\phi(c_1,0)$ is a nondecreasing function of c_1 on PO. But $\Pr(w_1 \leq c_1, w_2 \leq 0)$ is zero at the point P and hence $\phi(c_1,0) \geq 0$ at the point P. Therefore $\phi(c_1,0) \geq 0$ on entire line segment PO. By symmetry, $\phi(0,c_2) \geq 0$ on the line segment SO.

Next, consider the line QR for which c_2 is constant (Figure 7.4.2). Since $\phi_1(c_1,c_2)$ is a nonincreasing function of c_1 for fixed c_2, hence by (7.4.14), we see that there exists a point T on QR (which may coincide with either Q or R) such that

$$\frac{\partial \phi(c_1,c_2)}{\partial c_1} \geq 0 \text{ on QT}$$

and

$$\frac{\partial\phi(c_1,c_2)}{\partial c_1} \leq 0 \text{ on TR.}$$

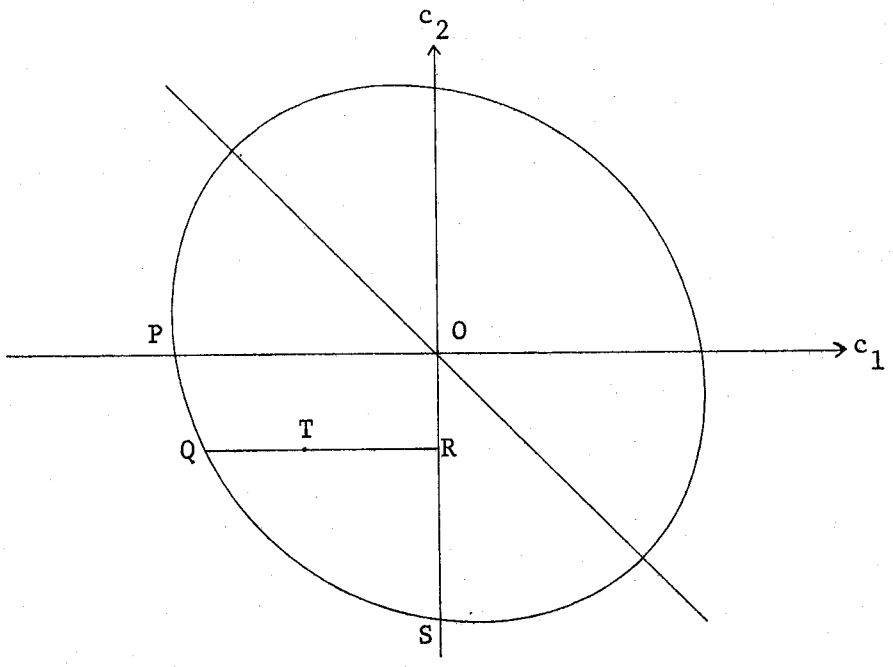


Figure 7.4.2

Therefore $\phi(c_1,c_2)$ increases from Q to T and decreases from T to R. But the function ϕ is non-negative at Q (since $\Pr(w_1 \leq c_1, w_2 \leq c_2)=0$ at Q) and at R (since R lies on the line SO). This shows that $\phi(c_1,c_2) \geq 0$ on QR for any fixed c_2, which completes the proof when both c_1 and c_2 are negative.

To complete the proof when both c_1 and c_2 are positive, note that (7.4.1) is equivalent to

$$\Pr(w_1 \geq c_1,\ w_2 \geq c_2) \leq \prod_{i=1}^{2} \Pr(w_i \geq c_i), \tag{7.4.16}$$

where c_1 and c_2 are negative. Since the probability distribution of (w_1,w_2) is same as that of $(-w_1,-w_2)$, hence (7.4.16) yields

$$\Pr(w_1 \leq d_1,\ w_2 \leq d_2) \leq \prod_{i=1}^{2} \Pr(w_i \leq d_i),$$

where $d_i = -c_i$ $(i=1,2)$ is positive.

Remark 2. In view of Lemma 7.2.1, one suspects that this theorem is valid for $\rho>0$ in some region near the "boundary" of the ellipse. However, this is not true for all points within the ellipse as can be easily seen by using equation (7.3.9).

7.5. Percentage points

7.5.1. Upper and lower limits

The solution of equations

$$n \Pr(w_1 > \bar{u}_\alpha) = \alpha,\ n \Pr(|w_1| > \bar{v}_\alpha) = \alpha \tag{7.5.1}$$

give an upper limit for u_α and v_α, where u_α and v_α denote the true upper $100\alpha\%$ points of U and V respectively. These can be obtained by using the tables of the incomplete beta function. Thus to get $\bar{u}_\alpha$ and $\bar{v}_\alpha$, we solve

$$I_{1-\bar{u}_\alpha^2}\left(\tfrac{1}{2}(p-1),\ \tfrac{1}{2}\right) = \frac{2\alpha}{n}$$

and

$$I_{1-\bar{v}_\alpha^2}\left(\tfrac{1}{2}(p-1),\ \tfrac{1}{2}\right) = \frac{\alpha}{n},$$

where $I_x(p,q)$ is defined at (6.2.2).

The method of finding the lower limits is similar to the case of known variance with the exception that now we use the bivariate probabilities given by (7.3.8) and (7.2.8).

Note that if $\rho_{max} \leq 0$, then on using Theorem 7.4.1

$$\Pr(U>\bar{u}_\alpha) \geq \alpha - \frac{(n-1)\alpha^2}{2n} .$$

7.5.2. True percentage points

For the case $\nu=0$, Srikantan [40] has shown that under certain conditions, which in general hold for small values of n, the upper limits given at (7.5.1) coincide with the true percentage points. This is true in the present case as well

Theorem 7.5.2. (a) $\bar{u}_\alpha$ coincides with u_α, provided that

(7.5.2) $$\bar{u}_\alpha \geq [\tfrac{1}{2}(1+\rho_{max})]^{\frac{1}{2}} .$$

(b) $\bar{v}_\alpha$ coincides with v_α, provided that

(7.5.3) $$\bar{v}_\alpha \geq [\tfrac{1}{2}(1+|\rho_{ij}|)]^{\frac{1}{2}} \quad \text{for all } i\neq j.$$

Proof: If the conditions (7.5.2) and (7.5.3) are satisfied, then by Lemma 7.2.1, we have for all $i\neq j$

$$\Pr(w_i>\bar{u}_\alpha,\ w_j>\bar{u}_\alpha) = 0$$

and

$$\Pr(|w_i|>\bar{v}_\alpha,\ |w_j|>\bar{v}_\alpha) = 0.$$

This means that condition (4.3.7) is satisfied by both $\bar{u}_\alpha$ and $\bar{v}_\alpha$ and hence $\bar{u}_\alpha$ and $\bar{v}_\alpha$ coincide with u_α and v_α respectively.

It should be noted that this theorem is of limited use, as it gives the true percentage points for some small values of n and ν only.

Illustration. For Example 4.5.3 with r=c=4 and α=.05, the upper limits $\bar{u}_\alpha$ and $\bar{v}_\alpha$ give the true percentage points for $\nu \leq 1$ and $\nu=0$ respectively.

7.6. Performance of test statistics

For simplicity, we only consider the measures P_a and P_b. We now need the distribution of w_k under H_k. From equation (7.2.1), we can write

$$w_k^2 = \frac{Q_1}{Q_1+Q_2} , \tag{7.6.1}$$

where

$$Q_1 = \frac{e_k^2}{\lambda_{kk}}$$

and

$$Q_2 = S_p^2 - Q_1 = \nu s_\nu^2 + (S^2-Q_1).$$

Now, by the results established in Section 4.4, under H_k, e_k has a $N(\lambda_{kk}\theta, \lambda_{kk}\sigma^2)$ distribution and S^2 has a noncentral $\sigma^2\chi^2$ distribution with n-m degrees of freedom and noncentrality parameter

$$\Delta_{kk}^2 = \lambda_{kk} \cdot \frac{\theta^2}{\sigma^2} .$$

It is clear that the distribution of w_k^2 depends on the ratio θ/σ and hence we may take σ=1. Considering the decomposition

$$S^2 = Q_1 + (S^2-Q_1),$$

we see that Q_1 has a noncentral χ^2 distribution with 1 d.f. and noncentrality parameter $\Delta_{kk}^2 = \lambda_{kk}\theta^2$, (S^2-Q_1) has a central χ^2 distribution with (n-m-1) d.f. and the two χ^2's are independent. Hence

$$w_k^2 \overset{d}{=} \frac{\chi'^2_{1,\Delta_{kk}^2}}{\chi'^2_{1,\Delta_{kk}^2} + \chi^2_{p-1}} , \tag{7.6.2}$$

where $\chi'^2_{1,\Delta_{kk}^2}$ stands for a noncentral χ^2 variate with 1 d.f. and noncentrality parameter Δ_{kk}^2, χ^2_{p-1} stands for a central χ^2 variate with p-1 d.f. and the two χ^2's are independent. Thus, under H_k, w_k^2 has a noncentral beta $(\frac{1}{2},\frac{1}{2}(p-1),\Delta_{kk}^2)$ distribution. The distribution of w_k is now immediate.

Note that (7.6.2) can be rewritten as

$$w_k^2 \overset{d}{=} \frac{t'^2_{f,\Delta_{kk}}}{t'^2_{f,\Delta_{kk}} + f} , \tag{7.6.3}$$

where $t'_{f,\Delta_{kk}}$ is the noncentral t distribution with f=p-1 d.f. and noncentrality parameter Δ_{kk}.

For the one-sided statistic, we now have

$$P_k = \Pr(w_k > u_\alpha \mid H_k), \quad k=1,2,\ldots,n$$

$$= \Pr\left(\frac{t'_{f,\Delta_{kk}}}{(t'^2_{f,\Delta_{kk}} + f)^{\frac{1}{2}}} > u_\alpha\right)$$

$$= \Pr\left(t'_{f,\Delta_{kk}} > \frac{u_\alpha \cdot f^{\frac{1}{2}}}{(1-u_\alpha^2)^{\frac{1}{2}}}\right) . \tag{7.6.4}$$

This can be compared with P_k of equation (6.3.2). Note that Δ_{kk} is same in both expressions. Now

$$P_a = \Pr(t'_{f,\Delta} > \frac{u_\alpha \cdot f^{\frac{1}{2}}}{(1-u_\alpha^2)^{\frac{1}{2}}}),$$

where Δ is given at (6.3.5).

Similarly, for the two-sided statistic

$$P_k = \Pr(|w_k| > v_\alpha | H_k), \quad k=1,2,\ldots,n$$

$$= \Pr(|t'_{f,\Delta_{kk}}| > \frac{v_\alpha \cdot f^{\frac{1}{2}}}{(1-v_\alpha^2)^{\frac{1}{2}}})$$

and

$$P_a = \Pr(|t'_{f,\Delta}| > \frac{v_\alpha \cdot f^{\frac{1}{2}}}{(1-v_\alpha^2)^{\frac{1}{2}}}).$$

7.7. Comparison between external and pooled studentization

In this section, we deviate from the notations used so far and retain the suffixes in various statistics and related quantities. Thus, for example, the true upper $100\alpha\%$ point of U_2 will now be denoted by $u_{2,\alpha}$ and ν as used in Chapter VI for external studentization will be denoted by ν_2.

The non-availability of true percentage points for most regression models makes the comparison difficult. One way to compare the two cases is to use the upper limits and study the various measures of performance mentioned in Section 4.4.

We now restrict our attention to the one-sided statistics U_2 and U_3. From equations (6.2.1) and (7.5.1) we see that $\bar{u}_{2,\alpha}$ and $\bar{u}_{3,\alpha}$ satisfy

(7.7.1) $$I_{\frac{\nu_2}{\nu_2+\bar{u}^2_{2,\alpha}}}(\tfrac{1}{2}\nu_2,\tfrac{1}{2}) = \frac{2\alpha}{n}$$

and

(7.7.2) $$I_{1-\bar{u}^2_{3,\alpha}}(\tfrac{1}{2}(n-m+\nu_3-1),\tfrac{1}{2}) = \frac{2\alpha}{n}\ .$$

Similarly, using the upper limits in the expressions for $P_{2,k}$ and $P_{3,k}$ given at (6.3.2) and (7.6.4) respectively, we get for k=1,2,...,n

(7.7.3) $$P_{2,k} = \Pr(t'_{\nu_2,\Delta_{kk}} > \bar{u}_{2,\alpha})$$

and

(7.7.4) $$P_{3,k} = \Pr(t'_{f_3,\Delta_{kk}} > \frac{f_3^{\frac{1}{2}}\cdot\bar{u}_{3,\alpha}}{(1-\bar{u}^2_{3,\alpha})^{\frac{1}{2}}}),$$

where

$$\Delta_{kk} = \lambda_{kk}^{\frac{1}{2}}\,\theta$$

and

$$f_3 = n-m+\nu_3-1.$$

From equations (7.7.1)-(7.7.4), it follows that if $f_3=\nu_2$ then

$$\bar{u}_{2,\alpha} = \frac{f_3^{\frac{1}{2}}\cdot\bar{u}_{3,\alpha}}{(1-\bar{u}^2_{3,\alpha})^{\frac{1}{2}}}$$

and

$$P_{2,k} \equiv P_{3,k}.$$

Consequently $P_{2,a}=P_{3,a}$ and $P_{2,b}=P_{3,b}$.

Next consider the case when $\nu_2=\nu_3$. By our assumption (4.1.7), the matrix Λ is at least of rank 2 and hence $n-m-1\geq 1$. This implies that $f_3>\nu_2$ and $P_{3,k}>P_{2,k}$. Hence, if we use the upper limits, then for the measures P_a and P_b, U_3 will have a definite edge over U_2. Obviously, the gain will be large when $(n-m-1)$ is large compared to ν_2.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Abramowitz, M. and Stegun, I. A. (eds.) (1964). Handbook of Mathematical Functions. National Bureau of Standards, Applied Mathematics Series 55, U. S. Government Printing Office, Washington, D. C..

[2] Amos, D. E. (1964). "Representations of the central and non-central t distributions", Biometrika, 51, 451-458.

[3] Anscombe, F. J. (1960). "Rejection of outliers", Technometrics, 2, 123-147.

[4] Barnett, V. D. (1966). "Order statistics estimators of the location of the Cauchy distribution". Journal of the American Statistical Association, 61, 1205-1218; correction 63, 383-385.

[5] Blom, G. (1958). Statistical Estimates and Transformed Beta Variables. Almqvist and Wiksell, Uppsala, Sweden.

[6] Chew, V. (1964). "Tests for the rejection of outlying observations". RCA Systems Analysis Technical Memorandum No. 64-7, Patrick Air Force Base, Florida.

[7] Craig, C. C. (1941). "Note on the distribution of non-central t with an application". Annals of Mathematical Statistics, 12, 224-228.

[8] David, F. N. and Johnson, N. L. (1954). "Statistical treatment of censored data, Part I, Fundamental formulae". Biometrika, 41, 228-240.

[9] David, H. A. (1956). "Revised upper percentage points of the extreme studentized deviate from the sample mean". Biometrika, 43, 449-451.

[10] David, H. A. and Paulson, A. S. (1965). "The performance of several tests for outliers". Biometrika, 52, 429-436.

[11] Doornbos, R. (1966). Slippage Tests. Mathematisch Centrum, Amsterdam.

[12] Dunnett, C. W. (1955). "A multiple comparison procedure for comparing several treatments with a control". Journal of the American Statistical Association, 50, 1096-1121.

[13] Dunnett, C. W. and Sobel, M. (1954). "A bivariate generalization of Student's t-distribution, with tables for certain special cases". Biometrika, 41, 153-169.

[14] Feller, W. (1957). An Introduction to Probability Theory and Its Applications, Volume I. John Wiley and Sons, Inc., New York.

[15] Fisher, R. A. (1940). "On the similarity of the distributions found for the test of significance in harmonic analysis, and in Stevens's problem in geometrical probability". Annals of Eugenics, 10, 14-17.

[16] Govindarajulu, Z. (1963). "On moments of order statistics and quasi-ranges from normal populations". Annals of Mathematical Statistics, 34, 633-651.

[17] Grubbs, F. E. (1950). "Sample criteria for testing outlying observations". Annals of Mathematical Statistics, 21, 27-58.

[18] Gumbel, E. J. (1954). "The maxima of the mean largest value and of the range". Annals of Mathematical Statistics, 25, 76-84.

[19] Gupta, S. S. (1963). "Probability integrals of multivariate normal and multivariate t". Annals of Mathematical Statistics, 34, 792-828.

[20] Halperin, M., Greenhouse, S. W., Cornfield, J. and Zalokar, J. (1955). "Tables of percentage points for the studentized maximum absolute deviate in normal samples". Journal of the American Statistical Association, 50, 185-195.

[21] Hartley, H. O. and David, H. A. (1954). "Universal bounds for mean range and extreme observation". Annals of Mathematical Statistics, 25, 85-99.

[22] Hartley, H. O. and Pearson, E. S. (1950). "Table of the probability integral of the t-distribution". Biometrika, 37, 168-172.

[23] Hume, M. W. (1965). "The distribution of statistics expressible as maxima". The Virginia Journal of Science, 16, New Series No. 2, 120-127.

[24] Johnson, N. L. and Welch, B. L. (1940). "Applications of the noncentral t-distribution". Biometrika, 31, 362-389.

[25] Kimball, A. W. (1951). "On dependent tests of significance in the analysis of variance". Annals of Mathematical Statistics, 22, 600-602.

[26] Kruskal, W. H. (1960). "Some remarks on wild observations". Technometrics, 2, 1-3.

[27] Locks, M. O., Alexander, M. J. and Byars, B. J. (1963). New Tables of the Noncentral t Distribution. Aeronautical Research Laboratories, Wright-Patterson Air Force Base, Ohio.

[28] Moriguti, S. (1951). "Extremal properties of extreme value distributions". Annals of Mathematical Statistics, 22, 523-536.

[29] National Bureau of Standards (1959). Tables of the Bivariate Normal Distribution Function and Related Functions. Applied Mathematics Series 50, U. S. Government Printing Office, Washington, D. C..

[30] Owen, D. B. and Steck, G. P. (1962). "Moments of order statistics from the equicorrelated multivariate normal distribution." Annals of Mathematical Statistics, 33, 1286-1291.

[31] Plackett, R. L. (1947). "Limits of the ratio of mean range to standard deviation". Biometrika, 34, 120-122.

[32] Quesenberry, C. P. and David, H. A. (1961). "Some tests for outliers". Biometrika, 48, 379-390.

[33] Rao, C. R. (1965). Linear Statistical Inference and Its Applications. John Wiley and Sons, Inc., New York.

[34] Rider, P. R. (1960). "Variance of the median of samples from a Cauchy distribution". Journal of the American Statistical Association, 55, 322-323.

[35] Sansone, G. (1959). Orthogonal Functions. Interscience, Inc., New York.

[36] Sarhan, A. E. and Greenberg, B. G. (eds.) (1962). Contributions to Order Statistics. John Wiley and Sons, Inc., New York.

[37] Sen, P. K. (1959). "On the moments of the sample quantiles". Calcutta Statistical Association Bulletin, 9, 1-19.

[38] Šidák, Z. (1968). "On multivariate normal probabilities of rectangles: their dependence on correlations". Annals of Mathematical Statistics, 39, 1425-1434.

[39] Siotani, M. (1964). "Interval Estimation for linear combinations of means". Journal of the American Statistical Association, 59, 1141-1164.

[40] Srikantan, K. S. (1961). "Testing for the single outlier in a regression model". Sankhyā Series A, 23, 251-260.

[41] Steck, G. P. (1962). "Orthant probabilities for the equicorrelated multivariate normal distribution". Biometrika, 49, 433-445.

[42] Sugiura, N. (1962). "On the orthogonal inverse expansion with an application to the moments of order statistics". Osaka Mathematical Journal, 14, 253-263.

[43] Young, D. H. (1967). "Recurrence relations between the P.D.F.'s of order statistics of dependent variables, and some applications". Biometrika, 54, 283-292.

APPENDIX

EVALUATION OF DEFINITE INTEGRALS

For our computational needs, the definite integrals were evaluated by using the Gauss quadrature formula (see e.g. [1], p. 887). The formula can be briefly summarized as below:

Let a and b be finite real numbers. Then for n point formula

$$\int_a^b f(y)dy \doteq \tfrac{1}{2}(b-a) \sum_{i=1}^{n} w_i f(y_i), \tag{A.1}$$

where

$$y_i = \tfrac{1}{2}(b-a)x_i + \tfrac{1}{2}(b+a), \quad i=1,2,\ldots,n$$

and the points x_i and weights w_i are constants extensively tabulated in [1] for various values of n. It should be noted that (A.1) gives an exact result when f(y) is a polynomial of degree (2n-1) or less.

The value n=20 was found to give sufficient accuracy for all of our computations. The points x_i and weights w_i were directly taken from the above tables and the computations were performed on a GE235 CALL A COMPUTER.