8106034

NAGARAJA, HAIKADY NAVADA

CONTRIBUTIONS TO THE THEORY OF THE SELECTION DIFFERENTIAL AND TO ORDER STATISTICS

Iowa State University PH.D. 1980

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

Contributions to the theory of the selection differential and to order statistics

by

Haikady Navada Nagaraja

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1980

TABLE OF CONTENTS

I. INTRODUCTION AND SUMMARY

Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ denote the order statistics of a random sample of size n from a distribution with distribution function F, mean μ and variance σ^2. Suppose we select the the top k values in our sample. Sometimes this is referred to as directional selection. The difference between the average of the selected group and the population mean μ expressed in standard deviation units represents a standardized measure of the differences between the selected group and the entire population. This quantity is called the selection differential and may be written as

$$D_{k,n} = \frac{1}{k} \sum_{i=n-k+1}^{n} \frac{(X_{i:n}-\mu)}{\sigma} .$$

"Selection differential" has long been a familiar term to geneticists and breeders who often refer to it as "intensity of selection" (Falconer, 1960). It represents a measure of improvement in the X-trait due to selection. Hence, it is useful in the construction of suitable breeding plans and in the comparison of different plans in plant as well as animal breeding. However, no systematic study of the general theory of the selection differential appears in the literature. Most of the results, developed with genetic applications in mind, concentrate on normal parent populations. Recently, Burrows (1972, 1975) has discussed some asymptotic results for the mean and variance of $D_{k,n}$ restricting consideration essentially to normal and exponential populations. $D_{k,n}$ also serves as a good test statistic in testing for outliers from normal populations. Our primary concern in this work is the study of distributional properties of $D_{k,n}$ both in finite samples and in asymptotic cases.

Sometimes the selection is based on an auxiliary variable, and is then often called indirect selection. Suppose two characters X and Y are associated and selection on the X character is easier to practice than selection on Y. Hence, in order to improve the Y character one may have to choose those with high X values. This is essentially what is done by plant breeders. Animal breeders perform selection on the parent population with the aim of improving a particular trait for the offspring population. In this case also, the selection is based on a concomitant variable. This leads to the definition of the "induced selection differential". Let (X_i,Y_i), i = 1 to n be a random sample from a bivariate population. Let $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ be the order statistics for the X-values and let $Y_{[i:n]}$ be the Y-value associated with $X_{i:n}$. Then $Y_{[i:n]}$ is termed the concomitant of $X_{i:n}$. If we select the top k X-values, then $k^{-1} \sum_{i=n-k+1}^{n} (Y_{[i:n]}-\mu_Y)/\sigma_Y$ represents the difference between the average of the Y-values for the selected group and the mean of the Y-population (μ_Y) expressed in the standard deviation units of the Y-population (σ_Y). This quantity is denoted by $D_{[k,n]}$ and is called the induced selection differential. There is hardly any work in the literature on general distribution theory for $D_{[k,n]}$. Discussion of $D_{[k,n]}$ is also included in our study.

This investigation is made up of five chapters apart from this introduction. Chapters II through IV deal with $D_{k,n}$ providing several small-sample and asymptotic results. In Chapter V we discuss $D_{[k,n]}$. The last chapter is devoted to a few miscellaneous results. Even though there are not many papers dealing with $D_{k,n}$ directly, several results, especially of an asymptotic nature, are available for linear functions of order statistics. $D_{k,n}$ being one such function, we make considerable use of such results. These are brought in and discussed at convenient places and will not be elaborated on here.

In Chapter II we assume that F is continuous and give an expression for the distribution function of $D_{k,n}$. Several bounds using the Cauchy-Schwarz technique and van Zwet's (1964) technique of convex transformation are given for $\mathcal{E}D_{k,n}$. These depend on the degree of restriction on F. Numerical comparison of these bounds are made for the standard normal population when the sample size is 10.

The last section considers the dependent sample case and develops bounds for $\mathcal{E}D_{k,n}$. It is shown there that $\mathcal{E}D_{k,n}$ can never exceed $\sqrt{(n-k)/k}$. This indicates that the breeder can not expect to do any better than this quantity by selection alone.

Chapter III deals with the basic asymptotic theory for $D_{k,n}$. Here the following three cases have to be distinguished: (i) the extreme case where k, the number selected, is held fixed and n, the sample size, becomes infinitely large; (ii) the quantile case where $k = [np]$, $0<p<1$; ([x] stands for the greatest integer not exceeding x); (iii) the asymptotically extreme case where $k\to\infty$, $n\to\infty$ but $k/n\to 0$. In Section 3.1 the limiting distribution of $D_{k,n}$ is obtained in all three cases for the exponential population. For a general parent, the discussion is limited to the first two cases. In the extreme case, by use of the results of Lamperti (1964) and Hall (1978), possible nondegenerate limit laws for $(D_{k,n}-a_n)/b_n$ are given under the assumption that $(X_{n:n}-a_n)/b_n$ has a nondegenerate limit law. The asymptotic distribution of $D_{k,n}$ in the quantile case can be obtained through several different approaches. Apart from the direct approach, one can use the results of Stigler (1974) and Boos (1979), since $D_{k,n}$ is a linear function of order statistics. It turns out that the asymptotic distribution of $D_{k,n}$ properly normalized is normal if and only if the $(1-p)^{th}$ quantile of the parent population is unique.

Some degenerate limit laws for $D_{k,n}$ are considered in Section 3.5. We establish some necessary and sufficient conditions for the existence of sequences of constants c_n and d_n such that $D_{k,n}-c_n \xrightarrow{p} 0$ and $D_{k,n}/d_n \xrightarrow{p} 1$. The discussion owes much to de Haan (1970). An almost sure result for $D_{k,n}$ is also given which requires F to be continuous. The last section investigates how the above asymptotic results apply when the parent distribution is normal, a situation of great practical importance.

In Chapter IV we extend the results on nondegenerate limit laws for $D_{k,n}$ obtained in Chapter III to the situations when some of our basic assumptions are violated. Also, an application of asymptotic theory to testing for outliers is discussed. When μ and σ are unknown and are estimated by the sample mean $\bar{X}$ and the sample standard deviation S, the asymptotic distribution of $\hat{D}_{k,n} = k^{-1} \sum_{i=n-k+1}^{n} (X_{i:n}-\bar{X})/S$, the sample selection differential is obtained. A similar extension is made to the case where the X_i's are independent, have the same first two moments but are not identically distributed. Section 4.5 considers two examples to show that these results may or may not hold for dependent samples. In the last section the problem of outliers is discussed and the use of the asymptotic theory for constructing approximate percentage points for $D_{k,n}$ when sampling from normal population is illustrated. Some comparisons of different asymptotic approaches are presented in the light of empirical percentage points obtained by Barnett and Lewis (1978).

We turn to the induced selection differential ($D_{[k,n]}$) in Chapter V and develop both finite-sample and asymptotic theory. Nondegenerate limit distributions of $D_{[k,n]}$ are obtained in both the extreme and quantile cases. Using a result due to Bhattacharya (1976) we derive the asymptotic joint distribution of $D_{[k,n]}$ and $D_{k,n}$ for the quantile case. The last section is devoted to the study of $D_{[k,n]}$ in the simple linear regression model. This model is often used in biological selection problems and $D_{[k,n]}$ is referred to as the "response to selection" in these applications.

The last chapter deals with two miscellaneous problems. First, we show that the asymptotic distribution of $(X_{n:n}-a_n)/b_n,\ldots,(X_{n-k+1:n}-a_n)/b_n$ in the extreme case is the same as the distribution of the first k lower record values from one of the three extreme value distributions. This observation produces a new canonical representation for the limiting random variables and can be used to give new proofs of some asymptotic results due to Hall (1978). We then prove a bivariate extension of Stigler's (1974) result for linear functions of order statistics. This is applied to obtain the asymptotic distribution of Hogg's (1974) Q statistic, a measure of tail length. As another application, the asymptotic distribution of a quick estimator of the regression coefficient in a simple linear regression model is obtained.

Some well-known results repeatedly referred to in the text are collected in the Appendix for convenience and quick reference. Lemma Ai stands for the i^{th} lemma in the Appendix.

II. SELECTION DIFFERENTIAL - FINITE SAMPLE CASE

2.1. Basic Set-up

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a continuous distribution with mean μ variance σ^2 and distribution function (df) F. Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ denote the order statistics of this sample. Suppose we select the top k X-values. Then $k^{-1} \sum_{i=n-k+1}^{n} (X_{i:n}-\mu)$ represents the average difference between the selected group and the population mean. This quantity expressed in standard deviation units is called the selection differential and may be written as

$$D_X(k,n) = \frac{1}{k} \sum_{i=n-k+1}^{n} (X_{i:n}-\mu)/\sigma. \qquad (2.1.1)$$

In a genetic context $D_X(k,n)$ is often termed "intensity of selection" (Falconer, 1960). For simplicity, $D_X(k,n)$ will be denoted by $D_{k,n}$ from here on. We usually assume that μ and σ are known and without loss of generality (WLOG) take $\mu = 0$, $\sigma = 1$. When μ and/or σ are replaced by $\overline{X}$ and/or S, the sample mean and the sample standard deviation, the resulting quantity will be called the sample selection differential. It will be denoted by $\hat{D}_{k,n}$ if both μ and σ are estimated and by $\hat{D}_{k,n}(\sigma)$ if only μ is estimated.

2.2. Distribution Function of $D_{k,n}$

$$P(D_{k,n} \le x) = P(X_{n-k+1:n} + \ldots + X_{n:n} \le kx)$$

$$= \int_{-\infty}^{\infty} P(X_{n-k+1:n} + \ldots + X_{n:n} \le kx \mid X_{n-k:n} = u)\, dF_{X_{n-k:n}}(u)$$

where $F_{X_{n-k:n}}$ is the df of $X_{n-k:n}$. From Lemma A1 it follows that given $X_{n-k:n}=u$, $X_{n-k+1:n}, \ldots, X_{n:n}$ form the order statistics from a random sample of size k from the df G_u given by

$$G_u(t) = \begin{cases} 0, & t<u \\ \dfrac{F(t)-F(u)}{1-F(u)}, & t \ge u. \end{cases}$$

Hence,

$$P(D_{k,n} \le x) = \int_{-\infty}^{x} G_u^{(k)}(kx)\, dF_{X_{n-k:n}}(u), \qquad (2.2.1)$$

where $G_u^{(k)}$ is the k-fold convolution of G_u; that is, the df of the sum of k independent identically distributed (iid) random variables (rvs) each with df G_u. As is evident from (2.2.1), there is no closed form expression for the df of $D_{k,n}$, in general. However, in the case of the exponential distribution, an expression for the probability density function (pdf) of $D_{k,n}$ can be given as discussed below.

Example 2.2.1:

Let

$$F(x) = \begin{cases} 1-e^{-x}, & x>0 \\ 0, & x\leq 0. \end{cases}$$

This is the df of an exponential rv with mean unity and hence the rv involved will be called Exp(1). From Lemma A2, one obtains,

$$M_{k,n} = \frac{1}{k}\sum_{i=n-k+1}^{n} X_{i:n} \stackrel{d}{=} \frac{Z_1}{n} + \frac{Z_2}{n-1} + \ldots + \frac{Z_{n-k}}{k+1} + \frac{Z_{n-k+1}}{k} + \ldots + \frac{Z_n}{k} \tag{2.2.2}$$

where Z_i's are iid Exp(1) rvs. Hence,

$$M_{k,n} \stackrel{d}{=} Z_1^* + \ldots + Z_{n-k}^* + Z^*/k$$

where $Z_i^* \sim \text{Exp}(\lambda_i^{-1})$, $\lambda_i = (n-i+1)$ and Z^*, the sum of k iid Exp(1) rvs, is Gamma (k,1), and are mutually independent. Consequently,

$$f_{M_{k,n}}(u) = \int_0^\infty f_{Z_1^*+\ldots+Z_{n-k}^*}(u-x)\, f_{Z^*/k}(x)\, dx.$$

From Feller (1966), p. 40, problem 12, it follows that

$$f_{Z_1^*+\ldots+Z_{n-k}^*}(u-x) = \lambda_1\lambda_2\ldots\lambda_{n-k}\Big[\sum_{i=1}^{n-k} \Psi_{i,n-k}\, e^{-\lambda_i(u-x)}\Big],$$

$$u-x>0$$

where

$$\Psi^{-1}_{i,n-k} = (\lambda_1-\lambda_i)\ \ldots\ (\lambda_{i-1}-\lambda_i)(\lambda_{i+1}-\lambda_i)\ldots(\lambda_{n-k}-\lambda_i).$$

Therefore,

$$f_{M_{k,n}}(u) = \lambda_1\lambda_2\ldots\lambda_{n-k} \sum_{i=1}^{n-k} \Psi_{i,n-k}\int_0^u e^{-\lambda_i(u-x)} \cdot \frac{k^k}{(k-1)!}\, e^{-kx}x^{k-1}dx$$

$$= \frac{k^k}{(k-1)!}\lambda_1\lambda_2\ldots\lambda_{n-k} \sum_{i=1}^{n-k} \Psi_{i,n-k}e^{-\lambda_i u} \cdot \int_0^u e^{x(\lambda_i-k)}x^{k-1}dx,\ u>0.$$

For a given k the integral can be evaluated explicitly and hence an explicit expression for the pdf of $D_{k,n}$ is available, since $D_{k,n} = (M_{k,n}-1)$.

2.3. Bounds on the Sample Selection Differential

Let $x_{1:n}\le x_{2:n}\le\ldots\le x_{n:n}$ be the order statistics from an observed sample $x_1,x_2,\ldots,x_n$. Mallows and Richter (1969) have established sharp bounds for $v_k = k^{-1}\sum_{i=n-k+1}^{n} x_{i:n}$, which is the sample selection differential except for a change of location and scale. Their Corollary 6.1 (p. 1931) states that

$$\bar{x} + \frac{(n-k)}{t}\frac{s}{\sqrt{n-1}} \le v_k \le \bar{x} + \sqrt{\frac{n-k}{k}}\, s \qquad (2.3.1)$$

where $t = \max(k,n-k)$ and $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ so that $S^2 = \frac{n}{n-1}s^2$. Assuming that $S\neq 0$ (i.e., x_i's are not all equal), we obtain

$$\frac{n-k}{t}\frac{1}{\sqrt{n}} \le \frac{v_k-\bar{x}}{S} \equiv \hat{D}_{k,n} \le \sqrt{\frac{n-k}{k}}\sqrt{\frac{n-1}{n}}.$$

These bounds are sharp.

2.4. Bounds on $\mathcal{E}D_{k,n}$ - Cauchy-Schwarz Technique

Since $\mu = 0, \sigma = 1$, $\int_0^1 F^{-1}(u)du = 0$ and $\int_0^1 [F^{-1}(u)]^2 du = 1$.

$$\mathcal{E}D_{k,n} = \frac{1}{k}\sum_{i=n-k+1}^{n}\mathcal{E}X_{i:n} = \int_0^1 [\sum_{i=n-k+1}^{n}\frac{1}{k}\frac{n!}{(i-1)!(n-i)!}$$

$$\cdot\, u^{i-1}(1-u)^{n-i} - 1]F^{-1}(u)du$$

$$\le \{\int_0^1 [\sum_{i=n-k+1}^{n}\frac{n}{k}\binom{n-1}{i-1}u^{i-1}(1-u)^{n-i} - 1]^2 du\}^{1/2}$$

$$\{\int_0^1 [F^{-1}(u)]^2 du\}^{1/2},$$

by the Cauchy-Schwarz inequality. Hence,

$$\mathcal{E}D_{k,n} \leq \{(\frac{n}{k})^2 \frac{1}{2n-1} \sum_{i,j=n-k+1}^{n} \frac{\binom{n-1}{i-1}\binom{n-1}{j-1}}{\binom{2n-2}{i+j-2}} - 1\}^{1/2}. \qquad (2.4.1)$$

Of course, $\mathcal{E}D_{k,n} \geq 0$. Equality in (2.4.1) is attained if and only if (iff), for some constant c,

$$F^{-1}(u) = c[\frac{n}{k} \sum_{i=n-k+1}^{n} \binom{n-1}{i-1} u^{i-1}(1-u)^{n-i} - 1]. \qquad (2.4.2)$$

First we note that for k<n, $\sum_{i=n-k+1}^{n} \binom{n-1}{i-1} u^{i-1}(1-u)^{n-i}$ represents the df of $(n-k)^{th}$ order statistic from a random sample of size (n-1) from $\mathcal{U}(0,1)$ distribution, that is, uniform distribution over (0,1). Hence, the right hand side (RHS) in (2.4.2) is increasing if c>0 and consequently there exists an F satisfying (2.4.2). For this F, $\mathcal{E}D_{k,n}$ is the bound given in (2.4.1). However, a closed form expression for such an F is not possible. But, since $\int_0^1 [F^{-1}(u)]^2 du = 1$,

$$c = \{(\frac{n}{k})^2 \frac{1}{2n-1} \sum_{i,j=n-k+1}^{n} \frac{\binom{n-1}{i-1}\binom{n-1}{j-1}}{\binom{2n-2}{i+j-2}} - 1\}^{-1/2}.$$

Also, $F^{-1}(0) = -c$ and $F^{-1}(1) = c.(n-k)/k$. Hence, this extremal F has bounded support, and is nonsymmetric.

Remarks:

1. The above technique has been employed for finding bounds for $\mathcal{E}X_{j:n}$, $1 \leq j \leq n$ in David (1970, p. 51) where it is noted that the bounds are attained only when $j=n$. But, in the case of the selection differential, or equivalently in the case of the average of $X_{n-k+1:n}, \ldots, X_{n:n}$ the bound is attainable for all k.

2. Let h(X) and g(X) be two functions of a rv X where $\mathcal{E}[h(X)]^2$ and $\mathcal{E}[g(X)]^2$ are finite. Let $\mathcal{E}h(X) = 0$. Then sharper bounds can be obtained for $\mathcal{E}h(X)g(X)$ by using the Cauchy-Schwarz inequality for $\mathcal{E}(h(X)-\mathcal{E}h(X))(g(X)-\mathcal{E}g(X))$ instead of the given expectation even though the two integrals are essentially the same. This procedure would yield a tighter bound than the one obtained by direct application of the Cauchy-Schwarz inequality.

3. We can obtain sharper upper bounds for $\mathcal{E}D_{k,n}$ assuming a symmetric parent distribution and using similar techniques. The Cauchy-Schwarz inequality applied to some orthonormal systems can be used to obtain tighter bounds and approximations for $\mathcal{E}D_{k,n}$. These would closely follow Section 4.3 of David (1970, pp. 54-57) and are omitted. But, some nontrivial extensions of his Section 4.4 are possible and we pursue this in the next section.

2.5. c-Comparison and s-Comparison

Let $\mathcal{F}$ be the class of all dfs which have positive continuous derivatives on their supports. If F and F* are in $\mathcal{F}$ then we say that $F \underset{c}{\leq} F^*$ iff $F^{*-1}F$ is convex on I, the support of F, and in such a case F is said to c-precede F*. Van Zwet (1964) has shown that if $F \underset{c}{\leq} F^*$, then

$$F(\mathcal{E}X_{r:n}) \leq F^*(\mathcal{E}X^*_{r:n}) \tag{2.5.1}$$

for all r = 1,2,...,n, and for all n for which $\mathcal{E}X_{r:n}$ and $\mathcal{E}X^*_{r:n}$ exist (see David, 1970, p. 60). We assume that both F and F* have finite variances. Since c-ordering is independent of location and scale, WLOG we take both F and F* to be standardized dfs.

From (2.5.1) we have

$$g(\mathcal{E}X_{r:n}) \leq \mathcal{E}X^*_{r:n}, \quad r = 1,2,\ldots,n$$

where $g = F^{*-1}F$ is a convex function on I. Hence,

$$\frac{1}{k}\sum_{i=n-k+1}^{n} g(\mathcal{E}X_{i:n}) \leq \frac{1}{k}\sum_{i=n-k+1}^{n} \mathcal{E}X^*_{i:n}. \tag{2.5.2}$$

Let Y be a rv which takes values $\mathcal{E}X_{n-k+1:n},\ldots,\mathcal{E}X_{n:n}$ with probability 1/k each. Since g is a convex function on I and these expectations belong to I, we have, by Jensen's inequality

$$g\left(\frac{1}{k}\sum_{i=n-k+1}^{n} \mathcal{E}X_{i:n}\right) = g(\mathcal{E}Y) \leq \mathcal{E}g(Y)$$

$$= \frac{1}{k}\sum_{i=n-k+1}^{n} g(\mathcal{E}X_{i:n}).$$

Hence, we have

$$g\left(\mathcal{E}\left(\frac{1}{k}\sum_{i=n-k+1}^{n} X_{i:n}\right)\right) \leq \frac{1}{k}\sum_{i=n-k+1}^{n} g(\mathcal{E}X_{i:n})$$

$$\leq \mathcal{E}\left(\frac{1}{k}\sum_{i=n-k+1}^{n} X^{*}_{i:n}\right). \tag{2.5.3}$$

Recalling that F and F* are standardized dfs it follows that

$$g(\mathcal{E}D_{k,n}) \leq \frac{1}{k}\sum_{i=n-k+1}^{n} g(\mathcal{E}X_{i:n}) \leq \mathcal{E}(D^{*}_{k,n}).$$

That is,

$$F(\mathcal{E}D_{k,n}) \leq F^{*}\left(\frac{1}{k}\sum_{i=n-k+1}^{n} g(\mathcal{E}X_{i:n})\right) \leq F^{*}(\mathcal{E}D^{*}_{k,n}). \tag{2.5.4}$$

Again, from (2.5.1), we have

$$\mathcal{E}X_{r:n} \leq g^{-1}(\mathcal{E}X^{*}_{r:n}).$$

Hence, proceeding on similar lines as above, and using the fact that g^{-1} is concave, one obtains,

$$F(\mathcal{E}D_{k,n}) \leq F\left(\frac{1}{k}\sum_{i=n-k+1}^{n} g^{-1}(\mathcal{E}X^{*}_{i:n})\right) \leq F^{*}(\mathcal{E}D^{*}_{k,n}). \tag{2.5.5}$$

(2.5.4) can be used to give lower bounds for $\mathcal{E}D^{*}_{k,n}$ whereas (2.5.5) is handy if we are interested in an upper bound for

$\mathcal{E}D_{k,n}$. However, note that the intermediate bounds are not easy to compute. If any of F and F* is not standardized, the corresponding selection differential has to be replaced by the average of the top k order statistics. In that case, one does not even need the finiteness of the mean, just the existence of expectations appearing in (2.5.4) or (2.5.5).

Applications: (i) c-Comparison with the $\mathcal{U}(0,1)$ df gives, for any (standardized) convex F,

$$F(\mathcal{E}D_{k,n}) \leq F\left(\frac{1}{k}\sum_{i=n-k+1}^{n}(i/(n+1))\right) \leq (2n-k+1)/2(n+1);$$

for any concave F, the inequalities are reversed.

(ii) For a standardized df F having increasing failure rate, that is for which $F'(x)/(1-F(x))$ is nondecreasing, we get

$$F(\mathcal{E}D_{k,n}) \leq 1 - \exp(-\mathcal{E}M_{k,n})$$

on c-comparison with Exp(1) distribution. Here $M_{k,n}$ is as given by (2.2.2) and hence

$$\mathcal{E}M_{k,n} = \sum_{i=k+1}^{n}\frac{1}{i} + 1 \leq \int_{k+1/2}^{n+1/2} x^{-1}dx + 1$$

$$= \log\frac{2n+1}{2k+1} + 1.$$

Consequently, $F(\mathcal{E}D_{k,n}) \leq 1 - (2k+1)/e(2n+1)$.

(iii) For the standard normal parent with df $\Phi(x)$, $1/\Phi(x)$ is convex. Hence, with $F(x) = -1/x$, $x<-1$ and $F^*(x) = \Phi(x)$, $F^{-1}F^*$ is concave. Consequently, $g = F^{*-1}F$ is convex since g is increasing and its inverse function is concave. Also, note that F does not have a mean but $\mathcal{E}D_{k,n}$ exists for $k<n$. $\mathcal{E}X_{r:n} = -n/(r-1)$, $r>1$ (David, 1970, p. 61) and hence from (2.5.4) we have

$$\Phi(\mathcal{E}D^*_{k,n}) \geq \Phi\left(\frac{1}{k}\sum_{i=n-k+1}^{n} \Phi^{-1}((i-1)/n)\right)$$

$$\geq F\left(-\frac{n}{k}\sum_{i=n-k}^{n-1} \frac{1}{i}\right).$$

That is,

$$\mathcal{E}D^*_{k,n} \geq \frac{1}{k}\sum_{i=n-k+1}^{n} \Phi^{-1}\left(\frac{i-1}{n}\right) \geq \Phi^{-1}\left(\frac{k}{n\sum_{i=n-k}^{n-1} i^{-1}}\right), \; k<n. \quad (2.5.6)$$

<u>s-Comparison</u>:

Now, we consider a subclass $\mathcal{S}$ of symmetric distributions in $\mathcal{F}$. Let $F(x_0-x) + F(x_0+x) = 1$ for some x_0 and all x if $F\varepsilon\mathcal{S}$. If F and F* are in $\mathcal{S}$, then $F \underset{s}{<} F^*$ iff $g = F^{*-1}F$ is convex for $x>x_0$, $x\varepsilon I$, the support of F. From van Zwet (1964), we have, whenever $F \underset{s}{<} F^*$,

$$g(\mathcal{E}X_{r:n}) \leq \mathcal{E}X^*_{r:n} \quad (2.5.7)$$

for all $(n+1)/2 \leq r \leq n$ and all n for which $\mathcal{E}X^*_{r:n}$ exists (see David, 1970, p. 63). We assume that both F and F* are standardized. Consequently, $x_0=0$ and $g(0) = 0$. Now, noting that $\mathcal{E}X_{r:n} > 0$ for $r > (n+1)/2$ and that g is convex for $x > 0$, we get, on using arguments similar to those leading to (2.5.3),

$$g(\mathcal{E}D_{k,n}) \leq \frac{1}{k} \sum_{i=n-k+1}^{n} g(\mathcal{E}X_{i:n}) \leq \mathcal{E}D^*_{k,n}, \quad k \leq (n+1)/2. \tag{2.5.8}$$

We now show that (2.5.8) is true even when $k > (n+1)/2$. Since F is symmetric about zero, for $k > (n+1)/2$,

$$\mathcal{E}D_{k,n} = \frac{1}{k}\mathcal{E}\left(\sum_{i=n-k+1}^{n} X_{i:n}\right) = \frac{2k-n}{k} \cdot 0 + \frac{1}{k} \sum_{i=k+1}^{n} \mathcal{E}X_{i:n}. \tag{2.5.9}$$

From (2.5.7), since $k > (n+1)/2$,

$$\frac{1}{k} \sum_{i=k+1}^{n} g(\mathcal{E}X_{i:n}) \leq \frac{1}{k} \sum_{i=k+1}^{n} \mathcal{E}X^*_{i:n}. \tag{2.5.10}$$

Define a rv Z which takes values $0, \mathcal{E}X_{k+1:n}, \ldots, \mathcal{E}X_{n:n}$ with probabilities $(2k-n)/k, 1/k, \ldots, 1/k$, respectively. Since g is convex on the support of Z, by Jensen's inequality, it follows that

$$g\left(\frac{2k-n}{k} \cdot 0 + \frac{1}{k} \sum_{i=k+1}^{n} \mathcal{E}X_{i:n}\right)$$

$$\leq \mathcal{E}g(Z)$$

$$= g(0)\, \frac{2k-n}{k} + \frac{1}{k} \sum_{i=k+1}^{n} g(\mathcal{E}X_{i:n})$$

$$= \frac{1}{k} \sum_{i=k+1}^{n} g(\mathcal{E}X_{i:n})$$

$$= \frac{1}{k} \sum_{i=n-k+1}^{n} g(\mathcal{E}X_{i:n})$$

since g is antisymmetric about 0. Now recalling (2.5.9) and (2.5.10) we conclude that (2.5.8) is true for $k > (n+1)/2$ also. This is recorded as a theorem below:

Theorem 2.5.1:

If F and F* are standardized dfs in $\mathcal{S}$, and $F \underset{s}{\leq} F^*$, then $F(\mathcal{E}D_{k,n}) \leq F^*(\frac{1}{k} \sum_{i=t}^{n} g(\mathcal{E}X_{i:n})) \leq F^*(\mathcal{E}D^*_{k,n})$, where $t = \max(k+1, n-k+1)$.

One can also show that

$$F(\mathcal{E}D_{k,n}) \leq F(\frac{1}{k} \sum_{i=t}^{n} g^{-1}(\mathcal{E}X^*_{i:n})) \leq F^*(\mathcal{E}D^*_{k,n}). \qquad (2.5.11)$$

For nonstandardized dfs, the selection differential has to be replaced by $M_{k,n}$, the average of the top k order statistics.

s-Comparison of the standard normal df(F) with the logistic distribution (F*), where $F^*(x) = (1 + \exp(-x))^{-1}$, $-\infty<x<\infty$ shows that $F \underset{s}{\leq} F^*$ (see David, 1970, p. 63) and hence from (2.5.11) we have

$$\mathcal{E}D_{k,n} \leq \frac{1}{k} \sum_{i=t}^{n} \Phi^{-1}(F^*(\mathcal{E}X^*_{i:n})) \leq \Phi^{-1}(F^*(\mathcal{E}M^*_{k,n})). \qquad (2.5.12)$$

It is known from David (1970, p. 64),

$\mathcal{E}X^*_{r:n} = \sum_{i=n-r+1}^{r-1} i^{-1}$ for $r \geq (n+1)/2$ and hence

$$\mathcal{E}M^*_{k,n} = \frac{1}{k}\sum_{i=t}^{n} \mathcal{E}X^*_{i:n}$$

$$= \begin{cases} \sum_{i=1}^{n-1} \frac{1}{i}, & k = 1 \\ \frac{n}{k}\sum_{i=n-k+1}^{n-1} \frac{1}{i} + \sum_{i=1}^{n-k} \frac{1}{i}, & 1 < k \leq \frac{n}{2} \\ \frac{n}{k}\sum_{i=n-k}^{n-1} \frac{1}{i} - \sum_{i=n-k}^{k} \frac{1}{i}, & \frac{n}{2} \leq k < n \end{cases} \quad (2.5.13)$$

on simplification.

Now we compare some of the bounds discussed so far when the parent distribution is standard normal and the sample size is 10. For this define the following:

UB1 = $\frac{1}{k}\sum_{i=t}^{n} \Phi^{-1}(F^*(\mathcal{E}X^*_{i:n}))$ of (2.5.12)

UB2 = $\Phi^{-1}(F^*(\mathcal{E}M^*_{k,n}))$ of (2.5.12) where $\mathcal{E}M^*_{k,n}$ is given by (2.5.13)

UB3 = Bound given by (2.4.1) using the Cauchy-Schwarz technique.

$LB = \frac{1}{k} \sum_{i=t}^{n} \phi^{-1}(\frac{i-1}{n})$, an improved version of the intermediate bound of (2.5.6) which exploits the symmetry of the normal distribution.

$\mathcal{E}D_{k,n}$ was computed using the table of expected values given by Teichroew (1956). Table 4.4 of David (1970) was used to compute UB1. All these bounds and $\mathcal{E}D_{k,n}$ are given for $k = 1(1)9$, $n = 10$ in the following table.

Table 2.5.1. Bounds for $\mathcal{E}D_{k,n}$ for n = 10

k	$\mathcal{E}D_{k,n}$	UB1	UB2	UB3	LB
1	1.539	1.591	1.591	2.065	1.282
2	1.270	1.309	1.321	1.526	1.062
3	1.065	1.096	1.115	1.211	0.883
4	0.893	0.918	0.942	0.987	0.725
5	0.739	0.760	0.787	0.810	0.580
6	0.595	0.612	0.641	0.658	0.483
7	0.457	0.470	0.499	0.519	0.378
8	0.318	0.328	0.354	0.381	0.265
9	0.171	0.177	0.196	0.229	0.142

Of the upper bounds, the ones obtained using s-comparison perform well in comparison with the one which uses the Cauchy-Schwarz technique. The lower bound is too low to be useful.

2.6. Dependent Sample Case

In this section we first consider bounds on the expectation of any linear function of order statistics when the variables are dependent and possibly nonidentically distributed. While doing so, we improve a result due to Arnold and Groeneveld (1979). Then, we discuss the case of the selection differential.

Suppose $X_1, X_2, \ldots, X_n$ are possibly dependent rvs with $\mathcal{E}X_i = \mu_i$ and $\mathrm{Var}(X_i) = \sigma_i^2$. Let $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ be the order statistics with $\mu_{i:n} = \mathcal{E}X_{i:n}$. Let $\bar{X}$ be the sample mean and $s^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$.

$$\mathcal{E}s^2 = \frac{1}{n} \sum_{i=1}^{n} \mathcal{E}X_i^2 - \mathcal{E}\bar{X}^2 \leq \frac{1}{n} \sum_{i=1}^{n} \mathcal{E}X_i^2 - (\mathcal{E}\bar{X})^2,$$

since $\mathrm{Var}(\bar{X}) \geq 0$

$$= \frac{1}{n} \sum_{i=1}^{n} (\mu_i^2 + \sigma_i^2) - \bar{\mu}^2, \text{ where } \bar{\mu} = \frac{1}{n} \Sigma\mu_i = \frac{1}{n} \Sigma\mu_{i:n}$$

$$= \frac{1}{n} \sum_{i=1}^{n} [\sigma_i^2 + (\mu_i - \bar{\mu})^2]$$

and the equality holds iff $\bar{X}$ = constant almost surely (a.s.). Also,

$$(\mu_{i:n} - \bar{\mu})^2 = [\mathcal{E}(X_{i:n} - \bar{X})]^2 \leq \mathcal{E}(X_{i:n} - \bar{X})^2$$

and hence

$$\sum_i (\mu_{i:n}-\bar{\mu})^2 \le \sum_i \mathcal{E}(X_{i:n}-\bar{X})^2 = \mathcal{E}(\sum_i (X_i-\bar{X})^2) = n\mathcal{E}s^2$$

where the equality holds iff $X_{i:n}-\bar{X} = c_i$ a.s. with $\Sigma c_i = 0$. Hence, we have the following:

$$\sum_{i=1}^{n} (\mu_{i:n}-\bar{\mu})^2 \le n\mathcal{E}s^2 \le \sum_{i=1}^{n} [\sigma_i^2 + (\mu_i-\bar{\mu})^2]. \qquad (2.6.1)$$

Arnold and Groeneveld (1979, pp. 220-221) have shown that:

$$\sum_{i=1}^{n} (\mu_{i:n}-\bar{\mu})^2 \le \sum_{i=1}^{n} [\sigma_i^2 + (\mu_i-\bar{\mu})^2]$$

and hence, for constants λ_i, $1 \le i \le n$, that

$$|\Sigma\lambda_i(\mu_{i:n}-\bar{\mu})| \le [\Sigma(\lambda_i-\bar{\lambda})^2]^{1/2}[\Sigma(\mu_{i:n}-\bar{\mu})^2]^{1/2}$$

$$\le [\Sigma(\lambda_i-\bar{\lambda})^2]^{1/2}[\Sigma(\sigma_i^2 + (\mu_i-\bar{\mu})^2)]^{1/2}. \qquad (2.6.2)$$

However, using the first inequality in (2.6.1), we obtain

$$|\Sigma\lambda_i(\mu_{i:n}-\bar{\mu})| \le \sqrt{n}[\Sigma(\lambda_i-\bar{\lambda})^2]^{1/2}(\mathcal{E}s^2)^{1/2} \qquad (2.6.3)$$

which is strictly better than (2.6.2) unless the sample mean is a constant a.s. Also, if we start with $\Sigma\lambda_i(X_{i:n}-\bar{X})$ instead of $\Sigma\lambda_i(\mu_{i:n}-\bar{\mu})$, use the Cauchy-Schwarz inequality, and take expectations at the end, we end up with still better bounds. To be precise, consider

$$|\Sigma\lambda_i(X_{i:n}-\bar{X})| = |\Sigma(\lambda_i-\bar{\lambda})(X_{i:n}-\bar{X})|$$

$$\leq [\Sigma(\lambda_i-\bar{\lambda})^2]^{1/2}[\Sigma(X_{i:n}-\bar{X})^2]^{1/2}$$

$$= \sqrt{n}\ [\Sigma(\lambda_i-\bar{\lambda})^2]^{1/2}s.$$

Therefore,

$$|\Sigma\lambda_i(\mu_{i:n}-\bar{\mu})| = |\mathcal{E}\Sigma\lambda_i(X_{i:n}-\bar{X})| \leq \mathcal{E}|\Sigma\lambda_i(X_{i:n}-\bar{X})|$$

$$\leq \sqrt{n}[\Sigma(\lambda_i-\bar{\lambda})^2]^{1/2}\mathcal{E}s.$$

That is,

$$|\Sigma\lambda_i(\mu_{i:n}-\bar{\mu})| \leq \sqrt{n}[\Sigma(\lambda_i-\bar{\lambda})^2]^{1/2}\mathcal{E}s. \qquad (2.6.4)$$

Noting that $\mathcal{E}s^2 \geq [\mathcal{E}(s)]^2$ we see that (2.6.4) gives a sharper bound than (2.6.3), with equality of bounds occurring only when s^2 is a constant a.s. The only shortcoming of (2.6.3) or (2.6.4) is that we need to know $\mathcal{E}s^2$ or $\mathcal{E}s$ in order to compute the bound. But, at the same time, one can dispense with the knowledge of σ_i^2's which are needed in (2.6.2).

Finally, we consider a special case of dependence where X_i's are uncorrelated. Then, it can be shown that

$$n\mathcal{E}(s^2) = \sum_i[(\mu_i-\bar{\mu})^2 + \sigma_i^2(\tfrac{n-1}{n})]$$

and hence (2.6.3) reduces to

$$|\Sigma\lambda_i(\mu_{i:n}-\bar{\mu})| \leq [\Sigma(\lambda_i-\bar{\lambda})^2]^{1/2}[\Sigma((\mu_i-\bar{\mu})^2+(1-\frac{1}{n})\sigma_i^{\,2})]^{1/2}$$

indicating clearly the improvement over (2.6.2). The above inequality is dealt with in Exercise 4.5.1 of David (1980).

Now we can assume that X_i's have the same mean μ and the same variance σ^2 and turn our attention to the selection differential. Here, sharp bounds can be obtained by dealing with (2.3.1), rather than appealing to any of the inequalities derived above. Taking expectations in (2.3.1), we get

$$\mu + \frac{n-k}{t} \cdot \frac{\mathcal{E}s}{\sqrt{n-1}} \leq \mathcal{E}\left(\frac{1}{k} \sum_{i=n-k+1}^{n} X_{i:n}\right) \leq \mu + \sqrt{\frac{n-k}{k}}\,\mathcal{E}s$$

where $t = \max(k, n-k)$. Therefore,

$$\frac{n-k}{\max(k,n-k)} \; \frac{1}{\sqrt{n-1}} \frac{\mathcal{E}s}{\sigma} \leq \mathcal{E}D_{k,n} \leq \sqrt{\frac{n-k}{k}} \frac{\mathcal{E}s}{\sigma} . \qquad (2.6.5)$$

Since the bounds in (2.3.1) are sharp, these bounds are also sharp. (A necessary condition is that s is constant a.s.). If $\mathcal{E}s$ is unknown, the fact that $\mathcal{E}s \leq \sigma$ can be used to replace the upper bound in (2.6.5) by $\sqrt{(n-k)/k}$. In addition, if X_i's are uncorrelated,

$$\mathcal{E}s \leq \sqrt{\mathcal{E}s^2} = \sigma\sqrt{(n-1)/n}$$

gives a slightly better upper bound, namely $\sqrt{(n-k)(n-1)/kn}$. But, a good lower bound for $\mathcal{E}s/\sigma$ is not possible without additional conditions on the parent distribution.

III. SELECTION DIFFERENTIAL - BASIC ASYMPTOTIC THEORY

In this chapter we investigate the asymptotic properties of $D_{k,n}$. We derive nondegenerate limit laws for $D_{k,n}$ as well as degenerate limit laws when k is a fixed integer and when k is a fixed proportion of n. Most of these results do not require the basic assumption of continuity made in Chapter II. More general results in this direction such as when the iid assumption is violated or when μ and σ are unknown and are estimated by $\overline{X}$ and S, are reported in Chapter IV.

3.1. Nondegenerate Limit Laws - Exponential Case

As in Example 2.2.1, let the X_i's be iid Exp(1) rvs. Define

$$\begin{aligned} S_{k,n} &= X_{n-k+1:n} + \ldots + X_{n:n} \\ &= kX_{n-k+1:n} + (k-1)(X_{n-k+2:n} - X_{n-k+1:n}) \\ &\quad + \ldots + 1(X_{n:n} - X_{n-1:n}) \\ &= kX_{n-k+1:n} + Z_1 + Z_2 + \ldots + Z_{k-1}, \text{ say.} \end{aligned}$$

It is known that the Z_i's and $X_{n-k+1:n}$ are mutually independent and $Z_i \sim \text{Exp}(1)$. Since $\mu = 1, \sigma = 1$ for the parent distribution,

$$D_{k,n} = \left(\frac{S_{k,n}}{k} - 1\right) = X_{n-k+1:n} + \left(\frac{Z_1+Z_2+\ldots+Z_{k-1}}{k} - 1\right). \tag{3.1.1}$$

We obtain the asymptotic distribution of $D_{k,n}$ in the following cases:

(i) k is a fixed integer and $n\to\infty$ (extreme case)

(ii) $k = [np]$, $0<p<1$, and $n\to\infty$ (quantile case)

(iii) $k\to\infty$ and $k = o(n)$ (asymptotically extreme case).

Case (i):

It is well-known (see e.g., Galambos, 1978, p. 102) that $(X_{n-k+1:n}-\log n)$ converges in law to a rv A whose df is given by

$$F_A(x) = \exp(-e^{-x}) \sum_{i=0}^{k-1} e^{-ix}/i!$$

(We will elaborate on this and related results later in Section 3.2).

Also, $Z_1+Z_2 +\ldots+ Z_{k-1} = B \sim \text{Gamma}\ (1,k-1)$ and hence

$$D_{k,n} - \log n \xrightarrow{\mathcal{L}} A + (B/k - 1)$$

where A and B are independent. The df of $A + (B/k-1)$ can be written explicitly and is dealt with in Theorem 3.2.2.

Case (ii):

In this case $(n-k+1)/n \to 1-p = q$, say. Let ξ_q be the qth quantile, that is $F(\xi_q) = q$. Here $\xi_q = -\log p$. Also, if $X \sim \text{Exp}(1)$, the conditional distribution of $(X-a)$ given

$X>a$ is also Exp(1) for any $a>0$. Consequently, $\mathcal{E}(X-\xi_q|X>\xi_q) = 1$ and $\text{Var}(X-\xi_q|X>\xi_q) = 1$.

Let $\mu_p = \mathcal{E}(X|X>\xi_q)$ and $\sigma_p^2 = \text{Var}(X|X>\xi_q)$. Then

$$\mu_p = \mathcal{E}(X-\xi_q|X>\xi_q) + \xi_q = 1 + \xi_q$$

and

$$\sigma_p^2 = \text{Var}(X-\xi_q|X>\xi_q) = 1.$$

Recalling (3.1.1) we have

$$(D_{k,n}-\mu_p) = (X_{n-k+1:n}-\xi_q) + (\frac{k-1}{k}\bar{Z}_{k-1}-1)$$

where $\bar{Z}_{k-1}$ is the mean of the Z_i's. Therefore,

$$\sqrt{k}(D_{k,n}-\mu_p) = \sqrt{k}(X_{n-k+1:n}-\xi_q) + \sqrt{k}(\bar{Z}_{k-1}-1) - \frac{\bar{Z}_{k-1}}{\sqrt{k}}$$

$$= A_n + B_n + C_n, \text{ say.}$$

By the Central Limit Theorem (CLT) for iid rvs, we have

$$\sqrt{k-1}(\bar{Z}_{k-1}-1) \stackrel{\mathcal{L}}{\rightarrow} N(0,1) \quad \text{as } n\rightarrow\infty$$

and hence

$$B_n = \sqrt{\frac{k}{k-1}} \cdot \sqrt{k-1}(\bar{Z}_{k-1}-1) \stackrel{\mathcal{L}}{\rightarrow} N(0,1).$$

Also,

$$C_n \stackrel{P}{\rightarrow} 0 \text{ since } \bar{Z}_{k-1} \stackrel{P}{\rightarrow} 1.$$

From Lemma A3 it follows that

$$p\sqrt{\frac{n}{pq}}\,(X_{n-k+1:n}-\xi_q) \stackrel{\mathcal{L}}{\rightarrow} N(0,1)$$

since $f(\xi_q) = p$ in the Exp(1) case.

Hence

$$A_n = \sqrt{k}(X_{n-k+1:n}-\xi_q) = \sqrt{\frac{k}{n}} \cdot \sqrt{n}(X_{n-k+1:n}-\xi_q)$$

$$\overset{\mathcal{L}}{\to} N(0,p\cdot \frac{q}{p}) = N(0,q), \text{ as } n \to \infty.$$

Since A_n and B_n are independent for every n,

$$A_n + B_n \overset{\mathcal{L}}{\to} N(0,1+q)$$

and hence

$$\sqrt{k}(D_{k,n}-\mu_p) = A_n + B_n + C_n$$
$$\overset{\mathcal{L}}{\to} N(0,1+q).$$

Case (iii):

$$\frac{S_{k,n}}{k} = X_{n-k+1:n} + \frac{Z_1+\ldots+Z_{k-1}}{k}.$$

Also, from Lemma A2,

$$X_{n-k+1:n} = \frac{E_1}{n} + \frac{E_2}{n-1} +\ldots+ \frac{E_{n-k+1}}{k} = S_n, \text{ say}$$

where $E_i \sim$ Exp(1) rvs and are independent. Hence S_n is the sum of independent rvs. Now,

$$\text{Var}(S_n) = \sigma^2(S_n) = \sum_{i=k}^{n} 1/i^2 \doteq (\frac{1}{k} - \frac{1}{n})$$

so that $\sqrt{k}\sigma(S_n) \to 1$. Therefore,

$$\frac{1}{\sigma(S_n)^{2+\delta}} \sum_{i=k}^{n} \mathcal{E}|(E_{n-i+1}-1)/i|^{2+\delta}$$

$$= \mathcal{E}|E_1-1|^{2+\delta}[\sum_{i=k}^{n}(\frac{1}{i})^{2+\delta}] \cdot [\sum_{i=k}^{n}\frac{1}{i^2}]^{-1-\delta/2}$$

$$\doteq c(\frac{1}{k^{1+\delta}}-\frac{1}{n^{1+\delta}})(\frac{1}{k}-\frac{1}{n})^{-1-\delta/2}$$

$$= \frac{c}{k^{\delta/2}}(1-(\frac{k}{n})^{1+\delta})(1-\frac{k}{n})^{-1-\delta/2}$$

$$\simeq c/k^{\delta/2}$$

$$\to 0, \text{ since } k\to\infty \text{ and } k/n \to 0.$$

Therefore, CLT holds for S_n (see Lemma A4) and hence

$$\frac{S_n-\mathcal{E}S_n}{\sigma(S_n)} \overset{\mathcal{L}}{\to} N(0,1).$$

Since $\sqrt{k}\sigma(S_n) \to 1$ and

$$0 \le \mathcal{E}S_n - \log\frac{n}{k} = \sum_{i=k}^{n}\frac{1}{i} - \int_k^n \frac{dx}{x} \le \sum_{i=k}^{n}\frac{1}{i} - \sum_{i=k+1}^{n}\frac{1}{i} = \frac{1}{k}$$

$\mathcal{E}S_n$ and $\sigma(S_n)$ can be replaced by $\log(n/k)$ and $k^{-1/2}$, respectively in CLT. That is,

$$\sqrt{k}(S_n-\log(n/k)) \overset{\mathcal{L}}{\to} N(0,1).$$

Also,

$$\sqrt{k}(\frac{Z_1+\ldots+Z_{k-1}}{k} - 1) \overset{\mathcal{L}}{\to} N(0,1)$$

as in case (ii). By the independence of S_n and Z_i's for all n, we obtain

$$\sqrt{k}(D_{k,n} - \log(n/k)) \xrightarrow{\mathcal{L}} N(0,2).$$

Section 3.2 is concerned with the extreme case, that is case (i) where k is a fixed integer and the sample size n approaches infinity, for a general distribution. Sections 3.3 and 3.4 deal with the quantile case, that is case (ii), in general. As we shall see later, the absence of the special properties enjoyed by the exponential distribution makes our proofs longer and more involved. The asymptotically extreme case, where $k\to\infty$ with $k/n\to 0$, for an arbitrary distribution is not pursued in this work.

3.2. Nondegenerate Limit Laws - Extreme Case

Suppose that there exist constants a_n, and $b_n>0$ such that for a df F,

$$P((X_{n:n}-a_n)/b_n \le x) = F^n(a_n+b_n x) \to G(x) \tag{3.2.1}$$

as $n\to\infty$, where G is a nondegenerate df. In such a case, we say that F is in the domain of attraction of G and we write $F\in D(G)$. Gnedenko (1943) has shown that G can be one of the three types of distributions Φ_α, Ψ_α and Λ, and has derived necessary and sufficient conditions for F to be in

D(G) in each of the three cases. F need not be continuous for (3.2.1) to hold. Appropriate sequences a_n and b_n which would facilitate convergence, are also known. Of course, the maximum from a given df F, need not have a nondegenerate limit distribution, whatever the normalization.

Lamperti (1964) has shown that if (3.2.1) holds, then for each $k \geq 1$, the vector $((X_{n:n}-a_n)/b_n, (X_{n-1:n}-a_n)/b_n, \ldots, (X_{n-k+1:n}-a_n)/b_n)$ has a limiting joint distribution of $(T_1, T_2, \ldots, T_k)$ which again can be only one of three types. In this situation T_k has one of the following distributions:

$$\Phi_\alpha(x;k) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}) \sum_{i=0}^{k-1} x^{-\alpha i}/i!, & x>0,\ \alpha>0 \end{cases} \tag{3.2.2a}$$

$$\Psi_\alpha(x;k) = \begin{cases} \exp(-|x|^{\alpha}) \sum_{i=0}^{k-1} |x|^{\alpha i}/i!, & x<0 \\ 1, & x \geq 0,\ \alpha>0 \end{cases} \tag{3.2.2b}$$

$$\Lambda(x;k) = \exp(-e^{-x}) \sum_{i=0}^{k-1} e^{-ix}/i!, \quad -\infty<x<\infty \tag{3.2.2c}$$

Dwass (1966) gave the joint pdf of $(T_1, T_2, \ldots, T_k)$. Hall (1978) has provided a canonical representation of the stochastic process $\{T_k, k \geq 1\}$ in terms of exponential rvs. First we use this representation to obtain the possible limiting distributions for $D_{k,n}$. Later we sketch a direct proof without using his representation. In Section 6.1, we take a closer look at T_i's to discover that they are in fact, lower record values from Φ_α, Ψ_α or Λ.

Lemma 3.2.1: (Hall, 1978)

If $F \varepsilon D(\Phi_\alpha)$, then

$$T_n \equiv T_n^{(1)} \overset{d}{=} \exp\{\frac{1}{\alpha}[\sum_{j=n}^{\infty} \frac{(Z_j-1)}{j} + \gamma - \sum_{j=1}^{n-1} \frac{1}{j}]\},\ n \geq 1 \qquad (3.2.3a)$$

If $F \varepsilon D(\Psi_\alpha)$, then

$$T_n \equiv T_n^{(2)} \overset{d}{=} -\exp\{-\frac{1}{\alpha}[\sum_{j=n}^{\infty} \frac{Z_j-1}{j} + \gamma - \sum_{j=1}^{n-1} \frac{1}{j}]\},\ n \geq 1 \qquad (3.2.3b)$$

If $F \varepsilon D(\Lambda)$, then

$$T_n \equiv T_n^{(3)} \overset{d}{=} \sum_{j=n}^{\infty} \frac{Z_j-1}{j} + \gamma - \sum_{j=1}^{n-1} \frac{1}{j},\ n \geq 1 \qquad (3.2.3c)$$

where Z_j's are iid Exp(1) rvs and $\sum_{j=1}^{0} 1/j$ is interpreted as zero.

As usual, we take $\mu = 0$, $\sigma = 1$. Suppose (3.2.1) holds. Then we know that

$$\left(\frac{X_{n:n}-a_n}{b_n}, \ldots, \frac{X_{n-k+1:n}-a_n}{b_n}\right) \overset{\mathcal{L}}{\rightarrow} (T_1, \ldots, T_k).$$

Since $(D_{k,n}-a_n)/b_n = \frac{1}{k} \sum_{i=1}^{k} (X_{n-i+1:n}-a_n)/b_n$ is a continuous function of the above components, it is immediate that

$$(D_{k,n}-a_n)/b_n \overset{\mathcal{L}}{\rightarrow} (T_1+\ldots+T_k)/k = D_k, \text{ say}$$

Hence we will try to find the distribution of D_k using (3.2.3). As we shall see later, only in the Λ-case can the df and pdf of D_k be found explicitly.

If $F\varepsilon D(\Phi_\alpha)$, using (3.2.3a) we have

$$T_i = T_{i+1} \exp(Z_i/i\alpha)$$

and hence

$$T_1 + \ldots + T_k \overset{d}{=} T_k^{(1)} (\exp(\frac{1}{\alpha} \sum_{j=1}^{k-1} \frac{Z_j}{j}) + \exp(\frac{1}{\alpha} \sum_{j=2}^{k-1} \frac{Z_j}{j})$$

$$+ \ldots + \exp(\frac{1}{\alpha} \frac{Z_{k-1}}{k-1}) + 1)$$

$$= T_k^{(1)} (1 + Y_1 + Y_1 Y_2 + \ldots + Y_1 Y_2 \ldots Y_{k-1})$$

where $Y_j = \exp(Z_{k-j}/\alpha(k-j))$ is a Pareto rv with parameter $\alpha(k-j)$. That is,

$$P(Y_j \leq u) = 1 - u^{-\alpha(k-j)}, \; u \geq 1.$$

Also, note that $T_k^{(1)}, Y_1, \ldots, Y_{k-1}$ are mutually independent. One can obtain the df of $Y_1 Y_2 \ldots Y_j$ either by induction or from Feller (1966, p. 40, Problem 12), recalling the relation between Pareto and exponential rvs. It turns out that

$$P(Y_1 Y_2 \ldots Y_j > u) = j\binom{k-1}{j} \sum_{i=1}^{j} \frac{(-1)^{i-1}}{k-i} \binom{j-1}{i-1} u^{-\alpha(k-i)},$$

$$u > 1, \; j \leq k-1.$$

However, this is of no help in the evaluation of the df of D_k.

If $F\varepsilon D(\Psi_\alpha)$, the situation is essentially the same as above, and $T_1 + \ldots + T_k \overset{d}{=} T_k^{(2)}(1 + Y_1^* + Y_1^*Y_2^* + \ldots + Y_1^*Y_2^* \ldots Y_{k-1}^*)$, where $Y_j^* = Y_j^{-1}$.

If $F \varepsilon D(\Lambda)$, it follows that

$$T_i = \frac{Z_i}{i} + T_{i+1} = \ldots = \frac{Z_i}{i} + \frac{Z_{i+1}}{i+1} + \ldots + \frac{Z_{k-1}}{k-1} + T_k$$

and hence that

$$T_1 + T_2 + \ldots + T_k = Z_1 + Z_2 + \ldots + Z_{k-1} + kT_k^{(3)}$$

Hence $D_k \overset{d}{=} A + B/k$ where $B \sim$ Gamma $(1,k-1)$ and A has the df $\Lambda(x;k)$ and A and B are independent.

The above discussion leads to the following theorem.

<u>Theorem 3.2.1</u>: If $(X_{n:n} - a_n)/b_n$ has a nondegenerate limiting distribution, then $(D_{k,n} - a_n)/b_n$ converges in distribution to a rv D_k where

(i) $D_k \overset{d}{=} T_k^{(1)}(1 + Y_1 + Y_1Y_2 + \ldots + Y_1 \ldots Y_{k-1})/k$

if $F \varepsilon D(\Phi_\alpha)$ (3.2.4a)

(ii) $D_k \overset{d}{=} T_k^{(2)}(1 + Y_1^* + Y_1^*Y_2^* + \ldots + Y_1^* \ldots Y_{k-1}^*)/k$

if $F \varepsilon D(\Psi_\alpha)$ (3.2.4b)

(iii) $D_k \overset{d}{=} B/k + T_k^{(3)}$ if $F \varepsilon D(\Lambda)$ (3.2.4c)

where $Y_i \sim$ Pareto $(\alpha(k-i))$, $Y_i^* = 1/Y_i$, $B \sim$ Gamma $(1,k-1)$, and $T_k^{(i)}$, $i = 1,2,3$ have the dfs given by (3.2.2a), (3.2.2b) and (3.2.2c) respectively. Furthermore, the rvs on the RHS in each of the three cases are mutually independent.

The next result provides the df, pdf and the characteristic function of D_k when $F\varepsilon D(\Lambda)$.

Theorem 3.2.2: If $F\varepsilon D(\Lambda)$, then $(D_{k,n}-a_n)/b_n$ converges in law to a rv D_k with the df, pdf and characteristic function given by (3.2.5), (3.2.6) and (3.2.7), respectively:

$$F_k(x) = \frac{k^{k-1}}{(k-2)!} \sum_{j=0}^{k-1} \frac{e^{-xj}}{j!} \int_0^\infty \exp(-e^{u-x}) e^{-u(k-j)} u^{k-2} du, \tag{3.2.5}$$

$$k \geq 2, \quad -\infty < x < \infty$$

$$f_k(x) = \frac{k^{k-1}}{(k-1)!(k-2)!} e^{-kx} \int_0^\infty \exp(-e^{u-x}) u^{k-2} du, \tag{3.2.6}$$

$$k \geq 2, \quad -\infty < x < \infty$$

$$\phi_{D_k}(t) = \mathcal{E} e^{itD_k} = \frac{\Gamma(k-it)}{\Gamma(k)} \left(\frac{k}{k-it}\right)^{k-1}. \tag{3.2.7}$$

Proof:

The pdf of B/k in (3.2.4c) is

$$f(x) = \frac{k^{k-1}}{(k-2)!} e^{-kx} x^{k-2}, \quad 0<x<\infty$$

since $B \sim$ Gamma $(1,k-1)$. For $k \geq 2$,

$$\begin{aligned} P(D_k \leq x) &= P(T_k^{(3)} + B/k \leq x) \\ &= \int_0^\infty P(T_k^{(3)} \leq x-u \mid B/k = u) f(u) du \\ &= \int_0^\infty P(T_k^{(3)} \leq x-u) f(u) du \end{aligned}$$

since $T_k^{(3)}$ and B/k are independent. Now, recalling (3.2.2c) and substituting the pdf of B/k, we obtain

$$P(D_k \le x) = \frac{k^{k-1}}{(k-2)!} \int_0^\infty \exp(-\exp(u-x)) \sum_{j=0}^{k-1} \frac{e^{-j(x-u)}}{j!} e^{-ku} u^{k-2} du$$

$$= \frac{k^{k-1}}{(k-2)!} \sum_{j=0}^{k-1} \frac{e^{-xj}}{j!} \int_0^\infty \exp(-\exp(u-x)) e^{-u(k-j)} u^{k-2} du,$$

$$-\infty < x < \infty.$$

This establishes (3.2.5). Differentiating $F_k(x)$, after several cancellations, one obtains $f_k(x)$ as given by (3.2.6). Direct derivation of the pdf using a transformation is also easy.

$$\phi_{D_k}(t) = \mathcal{E} e^{itD_k} = \mathcal{E} e^{itT_k^{(3)}} \cdot \mathcal{E} e^{i(t/k)B}$$

$$= \frac{\Gamma(k-it)}{\Gamma(k)} \left(1 - \frac{it}{k}\right)^{-(k-1)}$$

since

$$\mathcal{E} e^{itT_k^{(3)}} = \frac{1}{(k-1)!} \int_{-\infty}^\infty e^{itx} e^{-kx} \exp(-\exp(-x)) dx$$

(pdf comes from (3.2.2c))

$$= \frac{1}{(k-1)!} \int_0^\infty e^{-u} u^{k-it-1} du$$

$$= \Gamma(k-it)/\Gamma(k).$$

Therefore, ϕ_{D_k} is given by (3.2.7), and hence the proof of the theorem.

Some percentage points of F_k for $k \le 5$ are given in Table 3.2.1 below. $F_1(x) = \exp(-\exp(-x))$ yields these points directly for k=1. For $k \geq 2$, the integral in (3.2.5) was evaluated using the IMSL DECADRE subroutine and increasing the upper limit of integration until the increase in the calculated value was insignificant. Dr. W. Q. Meeker provided an efficient iterative algorithm to obtain the solution of $F_k(x) = p$.

Table 3.2.1. Values of $\xi_{k,p} = F_k^{-1}(p)$ for some selected p

k \ p	0.50	0.95	0.99
1	0.366513	2.970195	4.600149
2	-0.037107	1.799911	2.812969
3	-0.334556	1.154068	1.932540
4	-0.565820	0.714566	1.363627
5	-0.754310	0.384305	0.949440

Table 3.2.2a exhibits $f_k(x)$, for x = -2.50(0.25)3.50 and Table 3.2.2b lists the modal points and corresponding f_k values for k = 2,3,4,5.

Table 3.2.2a. Values of $f_k(x)$ for some selected x

x \ k	2	3	4	5
-2.50	.00012	.00023	.00044	.00084
-2.25	.00131	.00248	.00463	.00835
-2.00	.00814	.01484	.02626	.04454
-1.75	.03163	.05506	.09180	.14518
-1.50	.08513	.14050	.21885	.31983
-1.25	.17187	.26675	.38481	.51509
-1.00	.27683	.40065	.53055	.64470
-0.75	.37319	.49933	.60163	.65794
-0.50	.43708	.53594	.58248	.56847
-0.25	.45784	.51003	.49582	.42840
0.00	.43877	.44029	.37977	.28830
0.25	.39157	.35103	.26659	.17656
0.50	.32996	.26216	.17404	.09988
0.75	.26541	.18549	.10692	.05285
1.00	.20555	.12546	.06239	.02641
1.25	.15432	.08171	.03485	.01256
1.50	.11292	.05155	.01875	.00573
1.75	.08089	.03164	.00977	.00252
2.00	.05691	.01898	.00495	.00107
2.25	.03945	.01116	.00245	.00044
2.50	.02700	.00645	.00119	.00018
2.75	.01827	.00367	.00056	.00007
3.00	.01225	.00206	.00026	.00003
3.25	.00815	.00114	.00012	.00001
3.50	.00539	.00063	.00005	.00000

Table 3.2.2b. Modal points of the distribution of D_k when $F\varepsilon D(\Lambda)$

k	max $f_k(x)$	mode
2	0.45784	-0.25
3	0.53603	-0.49
4	0.60494	-0.68
5	0.66705	-0.85

Figure 3.2.1 describes $f_k(x)$ for $k = 2,\ldots,5$. All these four distributions are positively skewed and as k increases the pdf becomes more peaked.

Now we sketch briefly a direct but long approach which also proves Theorem 3.2.1. To fix the ideas we assume $F\varepsilon D(\Lambda)$ since the remaining cases can be handled by means of a transformation.

$$P\left(\sum_{j=n-k+1}^{n} \frac{X_{j:n}-a_n}{b_n} \le x\right)$$

$$= \int P\left(\sum_{j=n-k+1}^{n} \frac{X_{j:n}-a_n}{b_n} \le x \,\middle|\, \frac{X_{n-k+1:n}-a_n}{b_n} = u\right)$$

$$\cdot \, dF_{(X_{n-k+1:n}-a_n)/b_n}(u)$$

which can be written, following the approach leading to (2.2.1) as

$$\int G_{n,u}^{(k-1)}(x-u)\, dF_{(X_{n-k+1:n}-a_n)/b_n}(u)$$

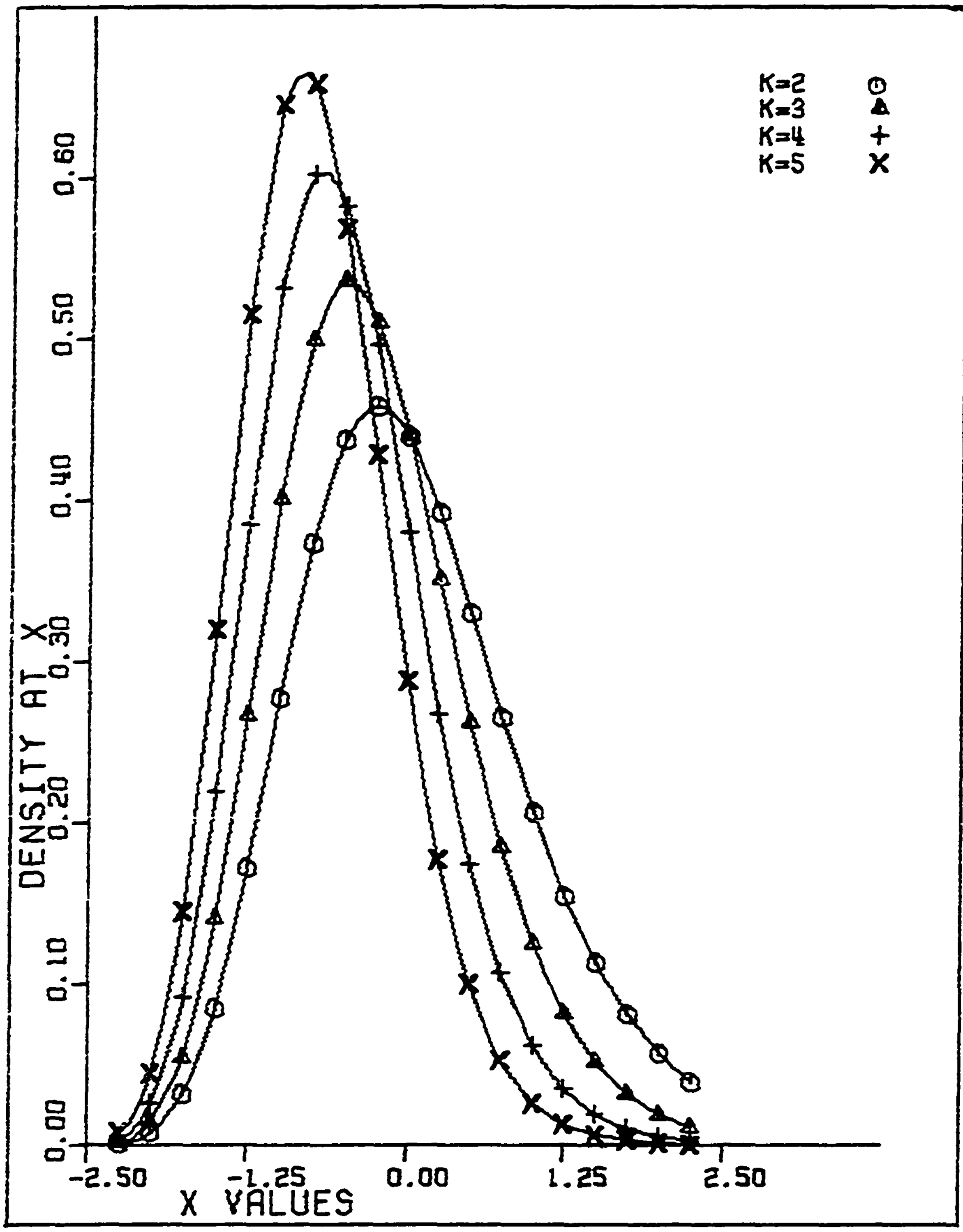

Figure 3.2.1. Probability density function of D_k for k = 2(1)5, when $F\epsilon D(\Lambda)$

where $G_{n,u}^{(k-1)}$ stands for the (k-1)-fold convolution df of

$$G_{n,u}(x)=\begin{cases}0, & x<u \\ \dfrac{F(a_n+b_nx)-F(a_n+b_nu)}{1-F(a_n+b_nu)}, & x\geq u.\end{cases}$$

Since $F^n(a_n+b_nx) \to \exp(-\exp(-x))$, $n[1-F(a_n+b_nx)] \to \exp(-x)$ for all x. Consequently,

$$G_{n,u}(x-u) = \begin{cases}0, & u>x/2 \\ \dfrac{F(a_n+b_n(x-u))-F(a_n+b_nu)}{1-F(a_n+b_nu)}, & u\leq x/2\end{cases}$$

$$\to G_u(x-u) = \begin{cases}0, & u>x/2 \\ \dfrac{e^{-u}-e^{-(x-u)}}{e^{-u}}, & u\leq x/2\end{cases} \quad \text{as } n\to\infty.$$

That is

$$G_u(x-u) = \begin{cases}0, & u>x/2 \\ 1-e^{-(x-2u)}, & u\leq x/2.\end{cases}$$

For a fixed x, $1-G_u(x-u)$ and $1-G_{n,u}(x-u)$ both behave as continuous dfs as functions of u, the former being the limit of the latter as $n\to\infty$. Since $1-G_u(x-u)$ is continuous, the convergence is uniform (Lemma A5). Hence, from Lemma A6, it follows that

$$\int [1-G_{n,u}(x-u)]\,dF_{(X_{n-k+1:n}-a_n)/b_n}(u)$$

$$\to \int [1-G_u(x-u)]\,dF_{T_k^{(3)}}(u).$$

Hence, we have shown that

$$P\left(\sum_{j=n-k+1}^{n}(X_{j:n}-a_n)/b_n\le x\right)$$

$$= \int G_{n,u}^{(k-1)}(x-u)\,dF_{(X_{n-k+1:n}-a_n)/b_n}(u)$$

$$\to \int G_u^{(k-1)}(x-u)\,dF_{T_k^{(3)}}(u) \qquad (3.2.8)$$

when $k=2$.

Also, for a fixed u, $G_{n,u}(x-u)$ behaves as a df converging to a continuous df $G_u(x-u)$ as a function of x. Hence, the convergence is also uniform in x. We had earlier shown that convergence is uniform in u for a given x. Using these facts inductively, one can show that for $j\ge 2$ $G_{n,u}^{(j)}(x-u) \to G_u^{(j)}(x-u)$ as $n\to\infty$ with uniform convergence in u for a given x (and the same in x for a given u). An appeal to Lemma A6 would then complete the proof of (3.2.8) for $k>2$.

Now, note that if Y has the df G_u, $Y = u+Z$, where $Z \sim \text{Exp}(1)$. Hence,

$$G_u^{(k-1)}(x-u) = P(Z_1+Z_2+\ldots+Z_{k-1} \le x-ku)$$

with independent Z_i's and consequently

$$\lim_{n\to\infty} P(k(D_{k,n}-a_n)/b_n\le x)$$

$$= \int P(Z_1+Z_2+\ldots+Z_{k-1} \le x-ku)\,dF_{T_k^{(3)}}(u)$$

$$= P(Z_1+Z_2+\ldots+Z_{k-1}+kT_k^{(3)} \le x).$$

Hence, $(D_{k,n}-a_n)/b_n \overset{\mathcal{L}}{\to} (Z_1+\ldots+Z_{k-1})/k + T_k^{(3)}$

which is (3.2.4c).

Remarks:

Had we conditioned on $X_{n-k:n}$ instead of $X_{n-k+1:n}$ in the above discussion we would have ended up showing that

$$(D_{k,n}-a_n)/b_n \xrightarrow{\mathcal{L}} (Z_1+Z_2+\ldots+Z_k)/k + T_{k+1}^{(3)}$$

which in view of Hall's representation (3.2.3c) is equivalent to (3.2.4c).

3.3. Nondegenerate Limit Laws - Quantile Case

Here we assume that $k = [np]$, $0<p<1$ where [ ] is the greatest integer function and derive the asymptotic distribution of $D_{k,n}$ appropriately normalized, as $n\to\infty$. In Section 3.1 it was shown that for the exponential parent distribution, the limiting distribution is normal. Now we will show this indeed is the case in a fairly general set-up. In the next section, using different approaches we derive all the possible limiting distributions.

Let F be absolutely continuous with pdf f and let ξ_q be the q^{th} quantile with $f(\xi_q) \neq 0$. Also assume that the $(2+\delta)^{th}$ moment exists for F. Let μ_p and σ_p be the mean and the standard deviation when F is truncated below at ξ_q. Then, as we shall see in the following steps, it follows that

$$\sqrt{k}(D_{k,n}-\mu_p) \xrightarrow{\mathcal{L}} N(0,\ \sigma_p^{\ 2} + q(\mu_p-\xi_q)^2).$$

Step 1: From Lemma A3 (due to Ghosh, 1971), we have, as $n\to\infty$,

$$Y_{k,n} = (X_{n-k:n}-\xi_q)\sqrt{\frac{n}{pq}}\ f(\xi_q) \xrightarrow{\mathcal{L}} N(0,1). \tag{3.3.1}$$

Let $\epsilon>0$ be given. Then there exists a constant c such that

$$\int_{|u|>c} d\Phi(u) < \epsilon/8 \tag{3.3.2}$$

where Φ is the standard normal df. Since $Y_{k,n} \xrightarrow{\mathcal{L}} N(0,1)$, there exists a positive integer $N_1(\epsilon)$ such that for all $n>N_1(\epsilon)$

$$\int_{|u|>c} dF_{Y_{k,n}}(u) < \epsilon/4. \tag{3.3.3}$$

Step 2:

Fix x and consider

$$P\left(\frac{\sqrt{k}(D_{k,n}-\mu_p)}{\sigma_p} \le x\right) = \int P\left(\frac{S_{k,n}-k\mu_p}{\sqrt{k}\ \sigma_p} \le x \mid Y_{k,n}=u\right) dF_{Y_{k,n}}(u). \tag{3.3.4}$$

From Lemma A1, given $Y_{k,n} = u$, $S_{k,n}$ is distributed as the sum of k iid rvs, say $A_{j,n}$ having mean $\mu_{\overline{F}(u_n)}$, standard deviation $\sigma_{\overline{F}(u_n)}$, where $\overline{F} = 1-F$, $u_n = \xi_q + \sqrt{pq/n}\cdot u/f(\xi_q)$, and df

$$H_{n,u}(x) = \begin{cases} 0, & x<u_n \\ \dfrac{F(x)-F(u_n)}{1-F(u_n)}, & x\geq u_n. \end{cases}$$

Define $Z_{j,n} = (A_{j,n}-\mu_{\overline{F}(u_n)})/\sqrt{k}\,\sigma_{\overline{F}(u_n)}$, $j = 1,\ldots,k$.

$\{Z_{j,n},\ j = 1 \text{ to } k,\ n = 1,2,\ldots\}$ is a double sequence of independent rvs, $Z_{j,n}$'s being iid for a given n. $\mathcal{E}Z_{j,n} = 0$ and $\sum_{j=1}^{k} \sigma^2(Z_{j,n}) = 1$.

$$\mathcal{E}|Z_{j,n}|^{2+\delta} = \frac{\mathcal{E}|A_{j,n}-\mu_{\overline{F}(u_n)}|^{2+\delta}}{k^{1+\delta/2}\sigma^{2+\delta}_{\overline{F}(u_n)}} < \infty \text{ since } (2+\delta)^{th} \text{ moment}$$

exists for the parent distribution.

$$\frac{\Sigma\mathcal{E}|Z_{j,n}|^{2+\delta}}{(\Sigma\sigma^2(Z_{j,n}))^{1+\delta/2}} = \frac{k\mathcal{E}|A_{j,n}-\mu_{\overline{F}(u_n)}|^{2+\delta}}{k^{1+\delta/2}\sigma^{2+\delta}_{\overline{F}(u_n)}} \sim \frac{1}{k^{\delta/2}} \to 0 \text{ as}$$

$n\to\infty$. Hence, from the CLT (Lemma A4), it follows that, given $Y_{k,n} = u$,

$$\frac{S_{k,n}-k\mu_{\overline{F}(u_n)}}{\sqrt{k}\,\sigma_{\overline{F}(u_n)}} \xrightarrow{\mathcal{L}} N(0,1). \tag{3.3.5}$$

<u>Step 3</u>:

$$\mu_{\overline{F}(u)} = \frac{1}{\overline{F}(u)}\int_u^\infty w\,dF(w) = u + \frac{1}{\overline{F}(u)}\int_u^\infty \overline{F}(w)\,dw.$$

Therefore,

$$\mu_{\bar{F}(u_n)} - \mu_{\bar{F}(\xi_q)} = (u_n - \xi_q) + \frac{1}{\bar{F}(u_n)} \int_{u_n}^{\infty} \bar{F}(w)\,dw$$

$$- \frac{1}{\bar{F}(\xi_q)} \int_{\xi_q}^{\infty} \bar{F}(w)\,dw$$

$$= (u_n - \xi_q) + \frac{\bar{F}(\xi_q) - \bar{F}(u_n)}{\bar{F}(u_n)\bar{F}(\xi_q)} \int_{\xi_q}^{\infty} \bar{F}(w)\,dw$$

$$- \frac{1}{\bar{F}(u_n)} \int_{\xi_q}^{u_n} \bar{F}(w)\,dw \, .$$

From the Mean Value Theorem of Integral Calculus, there exists a v_n between ξ_q and u_n such that

$$\int_{\xi_q}^{u_n} \bar{F}(w)\,dw = (u_n - \xi_q)\bar{F}(v_n) \, .$$

Also

$$\int_{\xi_q}^{\infty} \bar{F}(w)\,dw = p(\mu_p - \xi_q) .$$

Hence,

$$\mu_{\bar{F}(u_n)} - \mu_{\bar{F}(\xi_q)} = (u_n - \xi_q)\{1 + \frac{(\mu_p - \xi_q)}{\bar{F}(u_n)}$$

$$\cdot \frac{\bar{F}(\xi_q) - \bar{F}(u_n)}{u_n - \xi_q} - \frac{\bar{F}(v_n)}{\bar{F}(u_n)}\} .$$

That is,

$$\sqrt{k}(\mu_{\bar{F}(u_n)}-\mu_{\bar{F}(\xi_q)}) = \sqrt{\frac{k}{n}}\sqrt{pq}\,\frac{u}{f(\xi_q)}\,\{1+\frac{(\mu_p-\xi_q)}{\bar{F}(u_n)}$$

$$\cdot\frac{F(u_n)-F(\xi_q)}{u_n-\xi_q} - \frac{\bar{F}(v_n)}{\bar{F}(u_n)}\}$$

$$\rightarrow p\sqrt{q}\,\frac{u}{f(\xi_q)}\,\{1+\frac{(\mu_p-\xi_q)}{p}\cdot f(\xi_q)-1\} \quad \text{as } n\rightarrow\infty.$$

Therefore,

$$\sqrt{k}(\mu_{\bar{F}(u_n)}-\mu_p) \rightarrow \sqrt{q}\,u(\mu_p-\xi_q) \quad \text{as } n\rightarrow\infty. \tag{3.3.6}$$

Step 4:

$$\frac{S_{k,n}-k\mu_p}{\sqrt{k}\,\sigma_p} = \frac{S_{k,n}-k\mu_{\bar{F}(u_n)}}{\sqrt{k}\,\sigma_{\bar{F}(u_n)}}\;\frac{\sigma_{\bar{F}(u_n)}}{\sigma_p} + \frac{\sqrt{k}(\mu_{\bar{F}(u_n)}-\mu_p)}{\sigma_p}\,.$$

From (3.3.6) and the fact that $\sigma_{\bar{F}(u_n)} \rightarrow \sigma_p$ as $n\rightarrow\infty$ it follows from (3.3.5) that given $Y_{k,n} = u$,

$$\frac{S_{k,n}-k\mu_p}{\sqrt{k}\,\sigma_p} \overset{\mathcal{L}}{\rightarrow} N(\frac{\sqrt{q}\,u(\mu_p-\xi_q)}{\sigma_p},\,1).$$

That is, for a fixed x,

$$H_n(u) \equiv P(\frac{S_{k,n}-k\mu_p}{\sqrt{k}\,\sigma_p} \le x \mid Y_{k,n} = u) \rightarrow \Phi(x-\frac{\sqrt{q}\cdot u(\mu_p-\xi_q)}{\sigma_p})$$

$$\text{as } n\rightarrow\infty.$$

Now we will show that the convergence is in fact uniform in u. For this, first note that $P(S_{k,n} \le x \mid X_{n-k:n} = u)$ is a continuous decreasing function of u for every n, for a fixed

x. Hence $P(S_{k,n} \leq k\mu_p + x\sqrt{k}\,\sigma_p | X_{n-k:n} = u)$ is a decreasing continuous function of u. Also, u_n is an increasing continuous function of u and consequently, $H_n(u)$ is a decreasing continuous function of u. The limit function $\Phi^*(u) = \Phi(x-\sqrt{q}\,u(\mu_p-\xi_q)/\sigma_p)$ is also a decreasing continuous function of u. Hence, we have a sequence of uniformly bounded decreasing continuous functions $H_n(u)$ converging to a decreasing continuous function $\Phi^*(u)$ on $[-c,c]$. Then it can be shown on lines similar to the proof of Lemma A5, that the convergence is uniform in u.

Step 5:

$H_n(u) \to \Phi^*(u)$ uniformly in $u\varepsilon[-c,c]$, from Step 4. Also, $F_{Y_{k,n}} \to \Phi$. Then, from Lemma A6 we conclude that

$$\int_{-c}^{c} H_n(u)\,dF_{Y_{k,n}}(u) \to \int_{-c}^{c} \Phi^*(u)\,d\Phi(u).$$

Hence, there exists an integer $N_3(\epsilon)$ such that for $n>N_3(\epsilon)$

$$\left| \int_{-c}^{c} H_n(u)\,dF_{Y_{k,n}}(u) - \int_{-c}^{c} \Phi^*(u)\,d\Phi(u) \right| < \epsilon/2. \qquad (3.3.7)$$

Step 6:

For $n>N(\epsilon) = \max(N_1,N_2,N_3)$,

$$\left|\int_{-\infty}^{\infty} H_n(u)\,dF_{Y_{k,n}}(u) - \int_{-\infty}^{\infty} \phi^*(u)\,d\Phi(u)\right|$$

$$\leq \left|\int_{-c}^{c} H_n(u)\,dF_{Y_{k,n}}(u) - \int_{-c}^{c} \phi^*(u)\,d\Phi(u)\right|$$

$$+ \int_{|u|>c} dF_{Y_{k,n}}(u) + \int_{|u|>c} d\,\Phi(u)$$

$$< \epsilon/2 + \epsilon/4 + \epsilon/8$$

from (3.3.7, 3.3.4, and 3.3.2).

$$< \epsilon.$$

Hence,

$$\lim_{n\to\infty} P\left(\frac{S_{k,n}-k\mu_p}{\sqrt{k}\,\sigma_p} \leq x\right) = \int_{-\infty}^{\infty} \Phi\left(x - \sqrt{q}\,\frac{u(\mu_p-\xi_q)}{\sigma_p}\right) d\Phi(u)$$

for all x.

The RHS is the df of $W_1 + W_2\sqrt{q}(\mu_p-\xi_q)/\sigma_p$ where W_1 and W_2 are N(0,1) rvs. Hence, we have shown that

$$\frac{\sqrt{k}(D_{k,n}-\mu_p)}{\sigma_p} = \frac{S_{k,n}-k\mu_p}{\sqrt{k}\,\sigma_p} \xrightarrow{\mathfrak{L}} N\left(0,\ 1 + \frac{q(\mu_p-\xi_q)^2}{\sigma_p^2}\right).$$

This will be stated as:

<u>Theorem 3.3.1</u>:

Let the parent df F be absolutely continuous and have finite $(2+\delta)^{th}$ moment. Let its pdf be positive at ξ_q, the q^{th} quantile. Then for $k = [np]$, $0<p<1$,

$$\sqrt{k}(D_{k,n}-\mu_p) \xrightarrow{\mathfrak{L}} N(0,\ \sigma_p^2 + q(\mu_p-\xi_q)^2) \tag{3.3.8}$$

where μ_p and σ_p^2 are the mean and variance of the distribution obtained by truncating F below at ξ_q.

3.4. Alternative Approaches in the Quantile Case

Several approaches are available for finding the limiting distribution of $D_{k,n}$ in the quantile case, since it is a linear combination of order statistics with a smooth weight function. These approaches besides being more general, have fewer conditions than demanded by Theorem 3.3.1. However, the proofs involved are more complicated appealing to deeper results in the literature. We examine two of them, and make some comparisons among all three approaches. Finally, we make use of a result on the trimmed mean, due to Stigler (1973), to give the most general version of Theorem 3.3.1.

Boos' Approach (1979):

Let us introduce a weight function J on (0,1) and define

$$T_n = \sum_{i=1}^{n} \left(\int_{(i-1)/n}^{i/n} J(u)\,du \right) X_{i:n} = \int F_n^{-1}(t) J(t)\,dt,$$

where F_n is the empirical df. Let $\mu(J,F) = \int F^{-1}(t)J(t)\,dt$ and

$$q(t) = [t(1-t)]^{1/2-\delta}, \quad 0<\delta<1/2. \tag{3.4.1}$$

Lemma 3.4.1: (Boos, 1979, p. 958)

Let J be bounded and continuous a.e. Lebesgue and a.e. F^{-1}, and let $\int q(F(x))dx < \infty$. Define

$$\sigma^2(J,F) = \iint J(F(u))J(F(v))[F(\min(u,v)) - F(u)F(v)]dudv \tag{3.4.2}$$

and assume that $0 < \sigma^2(J,F) < \infty$. If the parent df is F, then

$$\sqrt{n}(T_n - \mu(J,F)) \xrightarrow{\mathcal{L}} N(0, \sigma^2(J,F)) \tag{3.4.3}$$

and

$$\limsup_{n\to\infty} \frac{\sqrt{n}(T_n-\mu(J,F))}{\sqrt{2\sigma^2(J,F)\log\log n}} = 1 \text{ with probability 1.} \tag{3.4.4}$$

The J function for the selection differential is

$$J(u) = \begin{cases} 1, & u \geq q \\ 0, & u < q. \end{cases} \tag{3.4.5}$$

This J is bounded and is also continuous a.e. F^{-1} if ξ_q, the q^{th} quantile of F, is unique. We assume this from here onwards.

$$T_n = \sum_{i=1}^{n} [\int_{(i-1)/n}^{i/n} J(u)du]X_{i:n}$$

$$= S_{k,n}/n \quad \text{if np is an integer}$$

$$= S_{k,n}/n + (np-[np])X_{n-k:n}/n \text{ if np is not an integer.}$$

Hence,

$$|T_n - \frac{S_{k,n}}{n}| \leq \frac{|X_{n-k:n}|}{n}$$

and consequently $\sqrt{n}(T_n-S_{k,n}/n) \xrightarrow{P} 0$ since, ξ_q being unique,

$X_{n-k:n} \overset{P}{\rightarrow} \xi_q$ (Smirnov, 1952, p. 12). Therefore,

$$\sqrt{n}(S_{k,n}/n-\mu(J,F)) = \sqrt{n}(T_n-\mu(J,F))-\sqrt{n}(T_n-S_{k,n}/n)$$

would have the same asymptotic distribution as $\sqrt{n}(T_n-\mu(J,F))$. That is

$$\sqrt{n}(\frac{S_{k,n}}{n} - \mu(J,F)) \overset{\mathcal{L}}{\rightarrow} N(0, \sigma^2(J,F)) \tag{3.4.6}$$

using (3.4.3). We now prove the following lemma.

Lemma 3.4.2:

For the J function given by (3.4.5), when ξ_q is unique,

$$\begin{aligned} \mu(J,F) &= p\mu_p \text{ and} \\ \sigma^2(J,F) &= p\sigma_p{}^2 + pq(\mu_p-\xi_q)^2 \end{aligned} \tag{3.4.7}$$

where μ_p and $\sigma_p{}^2$ are the mean and variances of the df G given by

$$G(x) = \begin{cases} (F(x)-q)/p, & x \geq \xi_q \\ 0 & , \quad x < \xi_q. \end{cases}$$

Proof:

$$\begin{aligned} \mu(J,F) &= \int F^{-1}(t)J(t)dt = \int_q^1 F^{-1}(t)dt \\ &= \int_{\xi_q}^{\infty} udF(u) \\ &= p\cdot\mu_p. \end{aligned}$$

From (3.4.2) we have

$$\sigma^2(J,F) = \int_{\xi_q}^{\infty}\int_{\xi_q}^{\infty} [F(\min(x,y)) - F(x)F(y)]dxdy.$$

$$F(\min(x,y)) - F(x)F(y) = (F(\min(x,y))-q) - (F(x)-q)(F(y)-q)$$

$$+ q(p-(F(x)-q)-(F(y)-q))$$

$$= pG(\min(x,y))-p^2G(x)G(y)+pq(1-G(x)-G(y))$$

over the region of integration.

$$= p[G(\min(x,y))-G(x)G(y)]+pq(1-G(x))(1-G(y)).$$

Hence,

$$\sigma^2(J,F) = p\iint [G(\min(x,y))-G(x)G(y)]dxdy$$

$$+ pq\int_{\xi_q}^{\infty} [1-G(x)]dx \cdot \int_{\xi_q}^{\infty} [1-G(y)]dy.$$

The first term is $p\sigma_p^{\,2}$ from a well-known representation for the variance due to Hoeffding (1948). Also,

$$\mu_p = \int_{\xi_q}^{\infty} xdG(x) = -\int_{\xi_q}^{\infty} xd[1-G(x)] = \xi_q + \int_{\xi_q}^{\infty} (1-G(x))dx$$

and hence

$$\int_{\xi_q}^{\infty} (1-G(x))dx = (\mu_p-\xi_q).$$

Therefore, we obtain

$$\sigma^2(J,F) = p\sigma_p^{\,2} + pq(\mu_p-\xi_q)^2$$

which completes the proof.

From (3.4.6) and (3.4.7) it follows that

$$\sqrt{np}\left(\frac{S_{k,n}}{np} - \mu_p\right) \xrightarrow{\mathcal{L}} N(0,\ \sigma^2(J,F)),\ \text{or}$$

$$\sqrt{np}\left(\frac{S_{k,n}}{np} - \mu_p\right) \xrightarrow{\mathcal{L}} N(0,\ \sigma_p{}^2 + q(\mu_p-\xi_q)^2).$$

This implies that $S_{k,n}/np \xrightarrow{P} \mu_p$ and hence

$$\sqrt{np}\left|\frac{S_{k,n}}{k} - \frac{S_{k,n}}{np}\right| \leq \frac{\sqrt{np}}{[np]}\left|\frac{S_{k,n}}{np}\right| \xrightarrow{P} 0\cdot|\mu_p| \quad = 0 \quad \text{as } n\to\infty.$$

Also, $\sqrt{np}/\sqrt{k} \to 1$ and hence combining all these we conclude that

$$\sqrt{k}\left(\frac{S_{k,n}}{k} - \mu_p\right) = \sqrt{k}(D_{k,n}-\mu_p) \xrightarrow{\mathcal{L}} N(0,\ \sigma_p{}^2 + q(\mu_p-\xi_q)^2).$$

Hence, we have proved Theorem 3.3.1 under fewer assumptions, namely we have now assumed that ξ_q is unique instead of the much stronger assumption of absolute continuity and non-zero pdf at ξ_q. However, as the following lemma shows, the existence of the $(2+\delta)^{th}$ moment for some $\delta>0$ and the existence of $\int q(F(x))dx$ for some $0<\delta<1/2$ are equivalent.

<u>Lemma 3.4.3</u>: The following statements are equivalent:

A. $\int |x|^{2+\delta}dF(x) < \infty$ for some $\delta>0$

B. $\int (F(x)(1-F(x)))^{1/2-\delta'}dx < \infty$ for some $0<\delta'<1/2$.

Proof: (A) $\Rightarrow \int_0^\infty x^{2+\delta} dF(x) < \infty$

$$\Rightarrow x^{2+\delta}(1-F(x)) \to 0 \quad \text{as } x\to\infty .$$

Therefore,

$$x^{(2+\delta)(1/2-\delta')}(1-F(x))^{1/2-\delta'} \to 0 \quad \text{if} \quad \delta' < 1/2.$$

That is

$$x^{1+\delta^*}(1-F(x))^{1/2-\delta'} \to 0 \quad \text{where } \delta^* = \delta/2 - \delta\delta' - 2\delta' > 0$$

$$\text{if } \delta' < \delta/(2(\delta+2)).$$

Hence, if $\delta' < \delta/(2(\delta+2))$, then $(1-F(x))^{1/2-\delta'} \underset{x\to\infty}{=} o(\frac{1}{x^{1+\delta^*}})$,

In other words, $q(F(x)) = o(\frac{1}{x^{1+\delta^*}})$, $x\to\infty$.

Since $\int_1^\infty \frac{dx}{x^{1+\delta^*}} < \infty$, we have $\int_1^\infty q(F(x))dx < \infty$. Similarly, by looking at the negative real axis, we obtain $\int_{-\infty}^{-1} q(F(x))dx < \infty$. Also, $\int_{-1}^{1} q(F(x))dx$, being a definite integral, is finite. This concludes the proof of the fact that A$\Rightarrow$B.

Now B implies

(i) $\int_0^\infty [1-F(x)]^{1/2-\delta'}dx < \infty$,

and

(ii) $\int_{-\infty}^0 [F(x)]^{1/2-\delta'}dx < \infty$.

Let $1-G(x) = [1-F(x)]^{1/2-\delta'}$. G(x) is a df and from (i)

$\int_0^\infty [1-G(x)]dx < \infty$.

Also

$$1-G(x) \geq 1-F(x)$$

since

$$\frac{1}{2} - \delta' < \frac{1}{2} < 1 \text{ and } 0<1-F(x)<1.$$

Hence, $G(x) \leq F(x) \leq (F(x))^{1/2-\delta'}$, which in light of (ii) implies that $\int_{-\infty}^0 G(x)dx < \infty$.

It is known that for a df H, $\int |x| dH(x)$ is finite iff both $\int_{-\infty}^0 H(x)dx$ and $\int_0^\infty [1-H(x)]dx$ are finite (see problem 18, p. 49, Chung, 1974).

Therefore, we conclude $\int |x| dG(x) < \infty$ and hence $x(1-G(x)) \to 0$ as $x\to\infty$. That is

$$x(1-F(x))^{1/2-\delta'} \to 0$$

or in other words $x^{(1/2-\delta')^{-1}}(1-F(x)) \to 0$ as $x\to\infty$. (3.4.8)

Now, $(\frac{1}{2} - \delta')^{-1} = \frac{2}{1-2\delta'} = \frac{2(1+2\delta')}{1-4\delta'^2} > 2(1+2\delta')$ since $0<2\delta'<1$.

Therefore,

$x^{(1/2-\delta')^{-1}} > x^{2+4\delta'}$, for $x>1$ and consequently from (3.4.8) we have

$$x^{2+4\delta'}(1-F(x)) \to 0.$$

Hence if $\delta<4\delta'$,

$$x^{2+\delta}(1-F(x)) \to 0. \quad (3.4.9a)$$

Also, since

$$x^{1+\delta}(1-F(x)) = o(\frac{1}{x^{(4\delta'-\delta)+1}})$$

and

$$\int_1^\infty \frac{dx}{x^{1+\delta^*}} < \infty \quad \text{for } \delta^*>0, \text{ it follows that}$$

$$\int_1^\infty x^{1+\delta}(1-F(x))dx < \infty \text{ whenever } \delta<4\delta'. \tag{3.4.9b}$$

(3.4.9a) and (3.4.9b) together imply that $\mathcal{E}(X^+)^{2+\delta} < \infty$ for $\delta<4\delta'$. Now taking $G(x) = [F(x)]^{1/2-\delta'}$, and proceeding on lines similar to the above discussion, we can show that $\mathcal{E}(X^-)^{2+\delta}<\infty$. Hence $\mathcal{E}|X|^{2+\delta} < \infty$.

Stigler's approaches:

To begin with, we state an important result due to Stigler (1974) for which the parent df F need not be continuous.

Theorem 3.4.1: (Stigler)

Let $S_n = \frac{1}{n}\sum_{i=1}^{n} J(\frac{i}{n+1})X_{i:n}$

where the weight function J is bounded and continuous a.e. F^{-1}. If the population variance is finite,

(i) $\lim_{n\to\infty} n\sigma^2(S_n) = \sigma^2(J,F)$, where $\sigma^2(S_n)$ is the variance of S_n and $\sigma^2(J,F)$ is given by (3.4.2)

(ii) If $\sigma^2(J,F) > 0$, $\frac{S_n - \mathcal{E}S_n}{\sigma(S_n)} \overset{\mathcal{L}}{\to} N(0,1)$

(iii) Suppose further that $\int [F(x)(1-F(x))]^{1/2} dx < \infty$ and that $J(u)$ satisfies a Lipschitz condition with index $\alpha > 1/2$ except possibly at a finite number of points of F^{-1} measure zero. Then

$$\sqrt{n}(\mathcal{E}S_n - \mu(J,F)) \to 0$$

where $\mu(J,F) = \int F^{-1}(t)J(t)dt$. Consequently

$$\sqrt{n}(S_n - \mu(J,F)) \overset{\mathcal{L}}{\to} N(0, \sigma^2(J,F)).$$

The above three parts appear as Theorems 1, 2 and 4, respectively in Stigler (1974). However, his proof of Theorem 4 was later discovered to be incomplete. But recently Mason (1979) has been able to prove it without any additional conditions by connecting S_n to the statistic T_n introduced earlier.

For the selection differential with $\mu = 0$, $\sigma = 1$, the J function, given by (3.4.5) satisfies the Lipschitz condition also along with other conditions if ξ_q is unique. Consequently, if $\int [F(x)(1-F(x))]^{1/2} dx < \infty$ and ξ_q is unique, then (3.3.8) holds. Of course, as was done in the Boos approach, one has to show that $\sqrt{n}(S_{k,n}/n - S_n) \overset{P}{\to} 0$, which is not difficult. Also, recall that $\sigma^2(J,F)$ was computed in Lemma 3.4.2. Hence, we can replace the assumption of the finiteness of the $(2+\delta)^{th}$ moment by a milder assumption that

$\int [F(x)\ (1-F(x))]^{1/2}dx$ is finite. But, this is stronger than just assuming finite variance. Now, Theorem 3.3.1, with these relaxed conditions can be restated as -

Theorem 3.4.2:

Let $\int [F(x)(1-F(x))]^{1/2}dx$ be finite and ξ_q be unique. Even when F is not continuous, (3.3.8) holds.

Finally, we use an asymptotic result by Stigler (1973) for trimmed means, since $D_{k,n}$ is essentially a trimmed mean, all the trimming being done on the left side. On examining the bivariate rv with $D_{k,n}$ and the number of sample points less than the (nonunique) q^{th} quantile as its components, and proceeding exactly as in his paper we obtain the following result. In fact, our case is simpler than his, because there is only one-sided trimming here.

Theorem 3.4.3:

Let $a = \sup\{x: F(x) \leq q\}$ and $A = a - \inf\{x:F(x) \geq q\}$. Then as $n\to\infty$,

$$\sqrt{k}(D_{k,n}-\mu_p) \overset{\mathcal{L}}{\to} Y_1 + (a-\mu_p)Y_2 - A\max(0,Y_2) \qquad (3.4.10)$$

where $Y_1 \sim N(0, \sigma_p^2)$, $Y_2 \sim N(0,q)$ and Y_1, Y_2 are independent.

Remarks:

Stigler's (1973) approach would require finite σ_p^2, but our basic assumption $\sigma = 1$ ensures this. Hence, we have imposed absolutely no more conditions than our basic assumptions. In fact, the results in this section are not limited to continuous distributions.

(ii) If A = 0, a = ξ_q, the unique quantile, and (3.4.10) reduces to (3.3.8).

(ii) When ξ_q is not unique the asymptotic distribution of $D_{k,n}$ is not a normal distribution.

Before closing this section we investigate the case when k is not exactly [np] but is fairly close. To be precise, when $\sqrt{n}(p-k/n) \to c$, a constant, we find the asymptotic distribution of $D_{k,n}$.

Theorem 3.4.4:

If $\sqrt{n}(p-k/n) \to c$, when c is a finite constant, then

$$\sqrt{k}(D_{k,n}-\mu_p) \overset{\mathcal{L}}{\to} N(\frac{c}{\sqrt{p}}(\mu_p-\xi_q),\ \sigma_p{}^2+q(\mu_p-\xi_q)^2)$$

if ξ_q is unique.

Proof:

WLOG we take $k \neq [np]$ always in the proof.

$$S_{[np],n}-S_{k,n} = \sum_{j=n-[np]+1}^{n-k} X_{j:n} \quad \text{if} \quad k<[np].$$

Hence, $([np]-k)X_{n-[np]+1:n} \leq S_{[np],n} - S_{k,n}$

$$\leq ([np]-k)X_{n-k:n}.$$

Therefore,

$$X_{n-[np]:n} \le X_{n-[np]+1:n} \le \frac{S_{[np],n}-S_{k,n}}{[np]-k} \le X_{n-k:n},$$

$$\text{if} \quad k<[np].$$

Similarly,

$$X_{n-k:n} \le \frac{S_{k,n}-S_{[np],n}}{k-[np]} \le X_{n-[np]:n} \quad \text{if} \quad k>[np].$$

Hence

$$\min(X_{n-k:n}, X_{n-[np]:n}) \le \frac{S_{k,n}-S_{[np],n}}{k-[np]}$$

$$\le \max(X_{n-k:n}, X_{n-[np]:n}) .$$

If ξ_q is unique, $X_{n-k:n} \xrightarrow{P} \xi_q$ and $X_{n-[np]:n} \xrightarrow{P} \xi_q$ (Smirnov, 1952, p. 9).

Therefore,

$$\frac{S_{k,n}-S_{[np],n}}{k-[np]} \xrightarrow{P} \xi_q$$

and consequently as $n\to\infty$,

$$\frac{S_{k,n}-S_{[np],n}}{\sqrt{k}} = \frac{S_{k,n}-S_{[np],n}}{k-[np]} \cdot \frac{k-[np]}{\sqrt{n}} \sqrt{\frac{n}{k}} \xrightarrow{P} \xi_q(-c)/\sqrt{p}. \tag{3.4.11}$$

Now

$$\sqrt{k}(D_{k,n}-\mu_p) = (S_{k,n}-k\mu_p)/\sqrt{k}$$
$$= \frac{S_{k,n}-S_{[np],n}}{\sqrt{k}} + \frac{[np]}{\sqrt{k}}\left(\frac{S_{[np],n}}{[np]} - \mu_p\right)$$
$$+ \frac{1}{\sqrt{k}}([np]-k)\mu_p$$

where the first term converges to $-\xi_q c/\sqrt{p}$ in probability from (3.4.11), the second term converges in law to $N(0, \sigma_p^2 + q(\mu_p-\xi_q)^2)$ and the last term tends to $c\mu_p/\sqrt{p}$ as $n\to\infty$. Hence,

$$\sqrt{k}(D_{k,n}-\mu_p) \xrightarrow{\mathcal{L}} N(c(\mu_p-\xi_q)/\sqrt{p},\ \sigma_p^2 + q(\mu_p-\xi_q)^2).$$

Note: This does not hold when ξ_q is not unique except when $c = 0$ in which case one obtains (3.4.10). This is because even though $X_{n-k:n}$ does not converge to any value in probability it would be bounded in probability. Then $c = 0$ ensures that $(S_{k,n}-S_{[np],n})/\sqrt{k}$ converges to zero in probability.

3.5. Degenerate Limit Laws

Weak laws - extreme case:

Following Galambos (1978, p. 206) we start with two definitions:

Definition 3.5.1:

A sequence of rvs $\{Y_n\}$ is said to satisfy an additive weak law (AWL) if there is a sequence of constants $\{a_n\}$ such that $Y_n-a_n \xrightarrow{P} 0$, as $n\to\infty$. We write (Y_n, a_n) obeys AWL.

Definition 3.5.2:

A sequence of rvs $\{Y_n\}$ is said to satisfy a multiplicative weak law (MWL) if there exists a sequence of nonzero constants b_n such that $Y_n/b_n \overset{P}{\to} 1$. We say (Y_n, b_n) obeys MWL.

We will examine conditions under which $D_{k,n}$ obeys AWL or MWL. We do not need the continuity of F in this section also. If $x_o = F^{-1}(1)$ is finite, then $D_{k,n} \overset{P}{\to} x_o$ (in fact a.s.ly) and hence $(D_{k,n}, x_o)$ obeys both laws except that when $x_o = 0$, MWL does not hold for $(D_{k,n}, 0)$. But MWL for $D_{k,n}$ in this case will be the same as AWL for $-\log(-D_{k,n})$ which has upper bound $+\infty$. Hence we take $x_o = +\infty$ and obtain some necessary and sufficient conditions, and some sufficient conditions for AWL and MWL to hold. First, we state an interesting lemma.

Lemma 3.5.1:

Let k be a fixed nonnegative integer, $\tau \geq 0$ and $\{p_n\}$ a sequence of real numbers with $0<p_n<1$. Then

$$\sum_{j=0}^{k} \binom{n}{j} p_n^{\,j} (1-p_n)^{n-j} \to e^{-\tau} \sum_{j=0}^{k} \frac{\tau^j}{j!}$$

iff $np_n \to \tau$, finite or infinite.

The proof can be found in Leadbetter (1978, p. 55), where only $\tau>0$ is considered. The same arguments hold when $\tau=0$ and $+\infty$ and when $\tau=0$ or $+\infty$ the RHS is interpreted as 1 or

0, respectively.

Theorem 3.5.1:

Let $x_o = +\infty$. Then for k, any fixed positive integer, $(D_{k,n}, a_n)$ obeys AWL iff $(X_{n:n}, a_n)$ obeys AWL.

Proof:

Suppose first that $(D_{k,n}, a_n)$ obeys AWL. Then

$$P(D_{k,n} \leq a_n + x) \to \begin{cases} 0, & x<0 \\ 1, & x>0. \end{cases} \tag{3.5.1}$$

$X_{n-k+1:n} \leq D_{k,n} \leq X_{n:n} \Rightarrow$

$$P(X_{n:n} \leq a_n + x) = F^n(a_n+x) \to 0, \quad x<0 \tag{3.5.2}$$

and

$$P(X_{n-k+1:n} \leq a_n + x) = \sum_{j=0}^{k-1} \binom{n}{j} [1-F(a_n+x)]^j [F(a_n+x)]^{n-j} \to 1, \quad x<0. \tag{3.5.3}$$

Now fix x>0 and let $p_n = 1-F(a_n+x)$. Then (3.5.2) implies that

$$\sum_{j=0}^{k-1} \binom{n}{j} p_n{}^j (1-p_n)^{n-j} \to 1 = e^0 \sum_{j=0}^{k-1} \frac{0^j}{j!}, \text{ as } n\to\infty.$$

Hence, from Lemma 3.5.1, we have $np_n = n[1-F(a_n+x)] \to 0$ as $n\to\infty$. That is, $F(a_n+x) = 1 + o(1/n)$ so that

$$F^n(a_n+x) = \left(1 + \frac{no(1/n)}{n}\right)^n \to e^0 = 1, \quad \text{for } x>0.$$

Hence, we have

$$F^n(a_n+x) \to \begin{cases} 0, & x<0 \quad \text{from (3.5.2)} \\ 1, & x>0 \quad \text{from the above line.} \end{cases} \tag{3.5.4}$$

Therefore, $(X_{n:n}, a_n)$ obeys AWL.

Now, to prove the converse, suppose that (3.5.4) holds. Again, observing that $D_{k,n} \leq X_{n:n}$ we have $P(D_{k,n} \leq a_n+x) \to 1$, $x>0$. For $x<0$, letting $p_n = 1-F(a_n+x)$, we see that $(1-np_n/n)^n \to 0$. Therefore, $np_n \to +\infty$. By Lemma 3.5.1, we would then have

$$P(X_{n-k+1:n} \leq a_n+x) = \sum_{j=0}^{k-1} \binom{n}{j} p_n{}^j (k-p_n)^{n-j} \to 0, \text{ for } x<0.$$

$X_{n-k+1:n} \leq D_{k,n}$ would then imply that

$$P(D_{k,n} \leq a_n+x) \to 0, \; x<0.$$

This proves that (3.5.1) holds. That is, $(D_{k,n}, a_n)$ obeys AWL.

de Haan (1970) has obtained several necessary and sufficient conditions for $(X_{n:n}, a_n)$ to obey AWL (see pp. 119-120). He has shown that if $(X_{n:n}, a_n)$ obeys AWL, a_n can be taken to be $\bar{a}_n = \inf\{x \mid 1-F(x) \leq 1/n\}$. As a consequence of his results and in view of Theorem 3.5.1, we have the following result.

<u>Theorem 3.5.2:</u>

Let $x_o = +\infty$. Then the following are equivalent:

a. There exists a sequence of constants a_n such that $(D_{k,n}, a_n)$ obeys AWL.

b. $(D_{k,n}, \bar{a}_n)$ obeys AWL.

c. $\lim_{t\to\infty} \frac{1-F(t+x)}{1-F(t)} = 0$ for all $x>0$.

d. $\int_0^\infty [1-F(t)]dt < +\infty$ and $\lim_{x\to\infty} \mathcal{E}(X-x|X>x) = 0$.

Theorem 2.9.4 of de Haan (1970) gives a sufficient condition for $(X_{n:n}, a_n)$ to obey AWL which also holds when $X_{n:n}$ is replaced by $D_{k,n}$. We do not state it here formally except to mention that the sufficient condition is that F' exists for large x and $F'(x)/(1-F(x)) \to +\infty$ as $x\to\infty$.

For the MWL for $D_{k,n}$, the following fact (see de Haan, 1970, p. 120) establishes an important relationship between distributions obeying AWL and MWL and consequently transforms every result on AWL into a corresponding result on MWL:

$X_{n:n}$ from df F has AWL iff $X^*_{n:n}$ from df F^* obeys MWL, where

$$F^*(x) = \begin{cases} 0, & x\leq 0 \\ F(\log x), & x>0. \end{cases} \tag{3.5.5}$$

However, to exploit the above relation and relevant results of de Haan (1970), we need the equivalent of Theorem 3.5.1:

<u>Theorem 3.5.3</u>:

$(D_{k,n}, b_n)$ obeys MWL iff $(X_{n:n}, b_n)$ obeys MWL.

The proof, being similar to that of Theorem 3.5.1, is omitted.

Now, if F and F^* are related as given by (3.5.5), from Theorems 3.5.1 and 3.5.3 and the statement preceding (3.5.5), we have

$$(D_{k,n}, a_n) \text{ has AWL} \Leftrightarrow (X_{n:n}, a_n) \text{ has AWL}$$
$$\Leftrightarrow (X^*_{n:n}, a^*_n) \text{ has MWL} \Leftrightarrow (D^*_{k,n}, a^*_n) \text{ has MWL.}$$

This relation leads to the following result (cf. de Haan, 1970, p. 116):

Theorem 3.5.4:

Let $x_o = +\infty$. The following are equivalent:

a. There exists a sequence of constants a_n such that $(D_{k,n}, a_n)$ obeys MWL.

b. $(D_{k,n}, \bar{a}_n)$ has MWL where $\bar{a}_n = \inf\{x | 1-F(x) \leq 1/n\}$.

c. $\lim\limits_{t\to\infty} \frac{1-F(tx)}{1-F(x)} = 0$ for all $x>0$.

d. $\int_0^\infty [1-F(t)]dt < \infty$ and $\lim\limits_{x\to\infty} \frac{\mathcal{E}(X|X>x)}{x} = 1$.

de Haan (1970) has given two sufficient conditions for $X_{n:n}$ to obey MWL, one obtained as a parallel to the sufficient condition for AWL mentioned earlier and the other in terms of the domain of attraction to Λ (see p. 117). The latter, in view of Theorem 3.5.3, leads to our next result.

Theorem 3.5.5:

If $F\varepsilon D(\Lambda)$ and $x_o = +\infty$, then $(D_{k,n}, \bar{a}_n)$ obeys MWL.

An almost sure result - extreme case:

Unlike the preceding results in this chapter, the following theorem is applicable only to a subclass of continuous distributions.

Theorem 3.5.6:

Let F be a continuous df with

$$\lim_{t\to\infty} (1-F(t))^{1/t} = d, \quad \text{finite or infinite}$$

Then

$$\frac{D_{k,n}}{\log n} \xrightarrow{a.s.} -\frac{1}{\log d}, \quad \text{where } k \text{ is any fixed integer.}$$

Proof:

$(1-F(t))^{1/t} \to d$ iff $t/[-\log(1-F(t))] \to 1/(-\log d) = c$, say. From Nagaraja (1978), it then follows that

$$\frac{X_{n-j+1:n}}{\log n} \xrightarrow{a.s.} c, \quad j = 1,2,\ldots,k.$$

This implies that $D_{k,n}/\log n \xrightarrow{a.s.} c$, completing the proof of the theorem.

Quantile case:

Equation (3.4.4) implies that when $k = [np]$, and ξ_q is unique, $D_{k,n} \xrightarrow{a.s.} \mu_p$ and hence also in probability if the $(2+\delta)^{th}$ moment is finite. In fact, it provides a much stronger result of iterated logarithm for $D_{k,n}$. In view of Theorem 3.4.3, $D_{k,n} \xrightarrow{p} \mu_p$ even when ξ_q is not unique.

3.6. Concluding Remarks

This chapter deals with asymptotic results for $D_{k,n}$ in the iid situation when $\mu = 0$, $\sigma = 1$ (i.e., both parameters known). An important point is that no major result here except for Theorem 3.5.6 required continuity of F. Hence, even if our assumption of continuity for F, as mentioned in Section 2.1 does not hold, these results are still applicable. The next chapter deals with more general situations when μ and/or σ are estimated, and with certain non-iid cases.

Before closing, we will examine the implication of the above results when the parent population is standard normal, to illustrate their applicability.

When k is fixed, since $\Phi \varepsilon D(\Lambda)$, $(D_{k,n}-a_n)/b_n \xrightarrow{\mathcal{L}} B/k + T_k^{(3)}$ of (3.2.4c) and hence its asymptotic df is given by (3.2.5). It is also known that one choice of a_n and b_n is (see Galambos, 1978, p. 65)

$$a_n = \sqrt{2 \log n} - (\log \log n + \log 4\pi)/2\sqrt{2 \log n}$$

and

$$b_n = 1/\sqrt{2 \log n}. \tag{3.6.1}$$

A more detailed study of (a) the choice of a_n and b_n and (b) approximate percentage points for $D_{k,n}$ is postponed to the next chapter.

When $k = [np]$, from Theorem 3.3.1, we have

$$\sqrt{k}(D_{k,n}-\mu_p) \xrightarrow{\mathcal{L}} N(0,\ \sigma_p^{\ 2} + q(\mu_p-\xi_q)^2)$$

where μ_p and $\sigma_p^{\ 2}$ are given by

$$\mu_p = \phi(\xi_p)/p \quad \text{and} \quad \sigma_p^{\ 2} = 1-\mu_p(\mu_p-\xi_q). \qquad (3.6.2)$$

Here ϕ represents the standard normal density.
Burrows (1972, 1975) has tabulated μ_p and $\sigma_p^{\ 2} + q(\mu_p-\xi_q)^2$ for several values of p.

As far as the degenerate limit laws are concerned, since $(X_{n:n}, \sqrt{2 \log n})$ obeys AWL (see David, 1980, p. 321), so does $(D_{k,n}, \sqrt{2 \log n})$. The pair also obeys MWL.

From the well-known fact that

$$1 - \Phi(x) = \frac{1}{x}\,\phi(x)[1 + O(\frac{1}{x^2})],\ x\to\infty,$$

it follows that

$$\frac{-\log(1-\Phi(x))}{x} = \frac{\log x}{x} + \frac{\log \sqrt{2\pi}}{x} - \frac{\log(1+O(1/x^2))}{x} + \frac{x}{2}$$

$$\to \infty \text{ as } x \to \infty.$$

Hence c = 0 in Theorem 3.5.6 and therefore $D_{k,n}/\log n \xrightarrow{a.s.} 0$.

IV. ASYMPTOTIC THEORY - EXTENSIONS AND APPLICATIONS

In Chapter III we assumed that our sample is a random sample from a distribution with known first two moments. Now, we relax some of these assumptions and examine possible limit laws for $D_{k,n}$, or its estimate both in the extreme and in the quantile case. In the extreme case, we give sufficient conditions which ensure the validity of Theorem 3.2.1 for $\hat{D}_{k,n}$, obtained by replacing μ and σ by their best sample estimates $\overline{X}$ and S. In the quantile case, our approach allows us to obtain the limiting distribution of $\hat{D}_{k,n}(\sigma)$, where μ is estimated by $\overline{X}$ and σ is assumed to be known. The independent nonidentically distributed situation is also dealt with in both the cases. Limit laws for $D_{k,n}$ in some special dependent situations are also discussed. The last section deals with the application of the asymptotic theory in the construction of percentage points for $D_{k,n}$, which is of use in testing for outliers.

4.1. Asymptotic Distribution of $\hat{D}_{k,n}$ in the Extreme Case

We now suppose that μ and σ are estimated by $\overline{X}$ and S, and find the asymptotic distribution of $\hat{D}_{k,n} = k^{-1}\sum_{i=n-k+1}^{n}(X_{i:n}-\overline{X})/S$. Since the distribution of $\hat{D}_{k,n}$ does not depend on μ or σ, we take $\mu = 0$, $\sigma = 1$ WLOG. We assume that there exist constants a_n, and $b_n > 0$ such that $(X_{n:n}-a_n)/b_n$ has a nondegenerate limit law as $n\to\infty$. In a series of lemmas we show that if $a_n/\sqrt{n}\, b_n \to 0$, then $\hat{D}_{k,n}$ also has one of the nondegenerate limit laws established in Theorem 3.2.1. In our discussion S^2 is the unbiased estimator of σ^2 given by $(n-1)^{-1} \sum_{i=1}^{n} (X_i-\bar{X})^2$.

Lemma 4.1.1:

If $a_n/\sqrt{n}\, b_n \to 0$ then $a_n(1-S)/b_n \overset{P}{\to} 0$ as $n\to\infty$.

Proof:

$$P\left(\left|\frac{a_n}{b_n}(1-S)\right|>\epsilon\right) = P\left(|1-S| > \left|\frac{b_n}{a_n}\right|\epsilon\right)$$

$$\leq \frac{\mathcal{E}(1-S)^2}{\left(\left|\frac{b_n}{a_n}\right|\epsilon\right)^2} \qquad (4.1.1)$$

by Chebychev's inequality. Now, $\mathcal{E}S^2 = \sigma^2 = 1$ and

$\mathcal{E}S = \sqrt{\frac{n}{n-1}}\,[\sigma+O(\frac{1}{n})]$ (Cramér, 1946, p. 353)

It follows that

$$\mathcal{E}(1-S)^2 = 1-2\mathcal{E}S + \mathcal{E}S^2$$
$$= O(1/n).$$

Therefore, from (4.1.1) we have

$$\limsup_{n\to\infty} P\left(\left|\frac{a_n}{b_n}(1-S)\right| >\epsilon\right) \leq \lim_{n\to\infty}\left|\frac{a_n}{\sqrt{n}\, b_n}\right|^2 \cdot \frac{n\,\mathcal{E}(1-S)^2}{\epsilon^2}$$

$$= 0 \text{ for all } \epsilon>0.$$

That is, $a_n(1-S)/b_n \overset{P}{\to} 0$.

Lemma 4.1.2:

If $(X_{n:n}-a_n)/b_n$ has a nondegenerate limiting distribution, then $a_n/\sqrt{n}\, b_n \to 0$ iff $X_{n:n}/\sqrt{n}\, b_n \overset{P}{\to} 0$.

Proof:

We know that (see Section 3.2), $(X_{n:n}-a_n)/b_n \overset{\mathcal{L}}{\to} T_1$, a real rv and hence $(X_{n:n}-a_n)/\sqrt{n}\, b_n \overset{P}{\to} 0$; that is, $(X_{n:n}/\sqrt{n}\, b_n) - (a_n/\sqrt{n}\, b_n) \overset{P}{\to} 0$.

Therefore, $(X_{n:n}/\sqrt{n}\, b_n) \overset{P}{\to} 0$ iff $a_n/\sqrt{n}\, b_n \to 0$.

Lemma 4.1.3: If $(X_{n:n}-a_n)/b_n$ has a nondegenerate limit law and $X_{n:n}/\sqrt{n}\, b_n \overset{P}{\to} 0$ then $1/\sqrt{n}\, b_n \to 0$ if $x_o = F^{-1}(1)$ is nonzero.

Proof:

We consider each of the three possible limiting dfs G, separately.

(i) $G = \Phi_\alpha$: Here $x_o = +\infty$ and one can take $a_n = 0$, $b_n = \xi_{(1-1/n)}$. Hence, $\sqrt{n}\, b_n \to \infty$ and consequently $1/\sqrt{n}\, b_n \to 0$.

(ii) $G = \Psi_\alpha$: One can take $a_n = x_o < +\infty$, $b_n = x_o - \xi(1-1/n)$. Since $X_{n:n}/\sqrt{n}\, b_n \overset{P}{\to} 0$, from Lemma 4.1.2 it follows that $x_o/\sqrt{n}\, b_n \to 0$, where x_o is assumed to be nonzero. Therefore, $1/\sqrt{n}\, b_n \to 0$.

(iii) $G = \Lambda$: $a_n = \xi_{(1-1/n)}$, and $b_n = \mathcal{E}(X_1 - a_n | X_1 > a_n)$. If x_o is finite proceed as in (ii). If $x_o = +\infty$, $X_{n:n}/\sqrt{n}\, b_n \overset{P}{\to} 0$ implies that $F^n(\sqrt{n}\, b_n \epsilon) \to 1$ as $n \to \infty$ for any positive ϵ and hence $\sqrt{n}\, b_n \epsilon \to +\infty$. Therefore, $1/\sqrt{n}\, b_n \to 0$.

Note:

The above proof does not assume anything about μ, even its existence. Lemma 4.1.3 has been proved by Berman (1962) assuming $\mu = 0$ in which case $x_o > 0$. Hence, the above result is, in a sense, more general than Berman's.

Lemma 4.1.4:

Let $F \varepsilon D(G)$ and have zero mean and finite variance. Then $\bar{X}/b_n \xrightarrow{P} 0$ iff $1/\sqrt{n}\, b_n \to 0$, as $n \to \infty$.

Proof:

WLOG, take $\sigma = 1$. Then $\sqrt{n}\, \bar{X} \xrightarrow{\mathcal{L}} N(0,1)$. Hence, if $1/\sqrt{n}\, b_n \to 0$ then $\bar{X}/b_n = \sqrt{n}\, \bar{X}/\sqrt{n}\, b_n \xrightarrow{P} 0$. Conversely, if $\bar{X}/b_n \xrightarrow{P} 0$, for a fixed positive ϵ, $P(\bar{X} < |b_n|\epsilon) \to 1$ and consequently $P(\sqrt{n}\, \bar{X} < \sqrt{n}|b_n|\epsilon) \to 1$. Since $\sqrt{n}\, \bar{X}$ converges in law to an unbounded rv, one can conclude that $\sqrt{n}|b_n|\epsilon \to \infty$ or $1/\sqrt{n}\, b_n \to 0$.

Now we are in a position to answer the main question of interest. We have assumed that $(X_{n:n} - a_n)/b_n \xrightarrow{\mathcal{L}} T_1$. We are interested in knowing whether

$$Y_n = \frac{1}{b_n}\left(\frac{X_{n:n} - \bar{X}}{S} - a_n\right) \xrightarrow{\mathcal{L}} T_1.$$

Since $S \xrightarrow{P} 1$ (recall that $\sigma = 1$), the above convergence is equivalent to

$$Y_n' = \frac{(X_{n:n}-\bar{X}) - a_n S}{b_n} \overset{\mathcal{L}}{\to} T_1 .$$

Theorem 4.1.1:

Let $(X_{n:n}-a_n)/b_n \overset{\mathcal{L}}{\to} T_1$ and let $\sigma = 1$. If $a_n/\sqrt{n}\, b_n \to 0$, then $Y_n \overset{\mathcal{L}}{\to} T_1$.

Proof:

We show that $Y_n' \overset{\mathcal{L}}{\to} T_1$.

$$Y_n' = \frac{X_{n:n}-a_n}{b_n} - \frac{\bar{X}}{b_n} + \frac{a_n(1-S)}{b_n}$$

and hence

$$Y_n' - \frac{X_{n:n}-a_n}{b_n} = \frac{a_n(1-S)}{b_n} - \frac{\bar{X}}{b_n} .$$

Now

$$\frac{a_n}{\sqrt{n}\, b_n} \to 0 \overset{\text{Lemma 4.1.2}}{\iff} \frac{X_{n:n}}{\sqrt{n}\, b_n} \overset{P}{\to} 0$$

$$\overset{\text{Lemma 4.1.3}}{\implies} \frac{1}{\sqrt{n}\, b_n} \to 0$$

$$\overset{\text{Lemma 4.1.4}}{\iff} \frac{\bar{X}}{b_n} \overset{P}{\to} 0.$$

Also, from Lemma 4.1.1 it follows that $a_n(1-S)/b_n \overset{P}{\to} 0$. Hence, $Y_n'-(X_{n:n}-a_n)/b_n \overset{P}{\to} 0$. Using Lemma A7 with k=1, the result now follows.

Remarks:

(i) Berman (1962) has proved this result but our argument is new. Further, Berman's approach tacitly assumes the finiteness of the fourth moment since he uses the fact that $\sqrt{n}(1-S)$ is asymptotically normal to show that $a_n(1-S)/b_n \xrightarrow{P} 0$. Our approach does not require this assumption.

(ii) The proof of Theorem 4.1.1 involved showing that $Y_n' - (X_{n:n}-a_n)/b_n \xrightarrow{P} 0$. Since $Y_n-Y_n' = Y_n'(1-S)/S \xrightarrow{P} 0$, it follows that $Y_n - (X_{n:n}-a_n)/b_n \xrightarrow{P} 0$. Repeating this technique we will show that, if $a_n/\sqrt{n}\, b_n \to 0$, then $Y_j - (X_{j:n}-a_n)/b_n \xrightarrow{P} 0$, where $Y_j = ((X_{j:n}-\bar{X})/S-a_n)/b_n$, $j = n,\ldots,(n-k+1)$.

Define $Y_j' = SY_j$. Then

$$Y_j' - (X_{j:n}-a_n)/b_n = a_n(1-S)/b_n - \bar{X}/b_n \xrightarrow{P} 0 \text{ if } a_n/\sqrt{n}\, b_n \to 0.$$

Therefore, $Y_j' \xrightarrow{\mathcal{L}} T_j$ and hence $(Y_j-Y_j') = Y_j'(1-S)/S \xrightarrow{P} 0$. Consequently,

$$Y_j - (X_{j:n}-a_n)/b_n \xrightarrow{P} 0, \quad j = n,n-1,\ldots,n-k+1. \qquad (4.1.2)$$

This fact is used in establishing the following result.

Theorem 4.1.2:

Let the parent df F be standardized and let $F \in D(G)$. If $a_n/\sqrt{n}\, b_n \to 0$ as $n \to \infty$, then

$$\underline{Y}_n = (Y_n,\ldots,Y_{n-k+1}) \xrightarrow{\mathcal{L}} (T_1,T_2,\ldots,T_k) \equiv \underline{T}$$

where T_j's are as described in Section 3.2, and $Y_j = ((X_{j:n}-\bar{X})/S-a_n)/b_n$.

Proof:

If $(X_{n:n}-a_n)/b_n \xrightarrow{\mathcal{L}} T_1$, then from Lamperti (1964), it follows that

$$\underline{X}_n = \left(\frac{X_{n:n}-a_n}{b_n}, \ldots, \frac{X_{n-k+1:n}-a_n}{b_n}\right) \xrightarrow{\mathcal{L}} \underline{T}. \tag{4.1.3}$$

From (4.1.2) and Lemma A8 we have $\underline{Y}_n - \underline{X}_n \xrightarrow{P} \underline{0}$. It now follows from Lemma A7 and (4.1.3), that $\underline{Y}_n \xrightarrow{\mathcal{L}} \underline{T}$.

Corollary 1:

Under the conditions of Theorem 4.1.2, $k(\hat{D}_{k,n}-a_n)/b_n \xrightarrow{\mathcal{L}} (T_1+T_2+\ldots+T_k)$. Hence, if $a_n/\sqrt{n}\, b_n \to 0$, Theorems 3.2.1 and 3.2.2 continue to hold when $D_{k,n}$ is replaced by $\hat{D}_{k,n}$, where $\hat{D}_{k,n} = k^{-1}\sum_{i=n-k+1}^{n}(X_{i:n}-\bar{X})/S$.

Corollary 2:

If $\sigma(=1)$ is known and $\hat{D}_{k,n}(\sigma)$ is defined to be $k^{-1}\sum_{i=n-k+1}^{n}(X_{i:n}-\bar{X})$, Theorems 3.2.1 and 3.2.2 hold when $D_{k,n}$ is replaced by $\hat{D}_{k,n}(\sigma)$ if $1/\sqrt{n}\, b_n \to 0$.

Lemmas 4.1.2 and 4.1.3 indicate that $a_n/\sqrt{n}\, b_n \to 0$ implies that $1/\sqrt{n}\, b_n \to 0$ for a standardized df. The converse statement does not seem to be true. Hence, it appears that conditions imposed in the theorem when σ is unknown are stronger than those imposed in Corollary 2, where σ is assumed to be known.

4.2. Asymptotic Distribution of $D_{k,n}$ in the Independent, Nonidentically Distributed, Extreme Case

If we relax the assumption of identical distribution of X_i's, but retain the independence assumption, the asymptotic theory developed in Section 3.2 holds with some additional assumptions. This is possible because of the extensions of Lamperti's (1964) results by Weissman (1975) to the case of independent but nonidentically distributed variates. Using Weissman's Theorem 3 we conclude the following:

Let X_i's be a sequence of independent rvs. For $t>0$ define

$$M_n(t) = \operatorname{Max}\{(X_1-a_n)/b_n, (X_2-a_n)/b_n, \ldots, (X_{[nt]}-a_n)/b_n\},$$

if $[nt] \geq 1$ and $M_n(t) = -\infty$ if $[nt] < 1$. Suppose that there exists a family of dfs $\{G_t, t>0\}$, not all identical, such that

$$P(M_n(t) \leq x) \to G_t(x) \quad \text{for all} \quad t>0. \tag{4.2.1}$$

Then

$$\left(\frac{X_{n:n}-a_n}{b_n}, \ldots, \frac{X_{n-k+1:n}-a_n}{b_n}\right) \xrightarrow{\mathcal{L}} (T_1, T_2, \ldots, T_k)$$

as in the iid case. Hence, if we assume that the X_i's have common mean zero and common variance unity and that (4.2.1) holds, then Theorem 3.2.1 holds for these independent, not necessarily identically distributed X_i's.

4.3. Asymptotic Distribution of $\hat{D}_{k,n}(\sigma)$ in the Quantile Case

Assuming that σ is known and $k = [np]$, $0<p<1$, we derive the asymptotic distribution of the sample selection differential $\hat{D}_{k,n}(\sigma)$. Since the discussion here follows closely Stigler's results discussed in Section 3.4, we briefly sketch our steps omitting routine details. The distribution of $\hat{D}_{k,n}(\sigma)$ does not depend on μ and σ, and hence WLOG we take $\mu = 0$, $\sigma = 1$. Then

$$\hat{D}_{k,n}(\sigma) = \frac{1}{k} \sum_{i=n-k+1}^{n} X_{i:n} - \bar{X} \doteq \frac{1}{k} \sum_{i=1}^{n-k} (-p) X_{i:n}$$

$$+ \sum_{i=n-k+1}^{n} qX_{i:n} .$$

Hence, $p\hat{D}_{k,n}(\sigma) \doteq n^{-1} \sum_{i=1}^{n} J(i/n+1) X_{i:n} = S_n$, say, where

$$J(u) = \begin{cases} -p, & u<q \\ q, & u\geq q \end{cases} .$$

We assume that ξ_q, the q^{th} quantile of the parent distribution is unique. Then, one can show that $\sqrt{n}(p\hat{D}_{k,n}(\sigma) - S_n) \overset{P}{\to} 0$. Also, J is continuous a.e. F^{-1} and satisfies a Lipschitz condition with $\alpha>1/2$. Further, we assume that $\int [F(x)(1-F(x))]^{1/2} dx$ is finite. From Theorem 3.4.1, it then follows that, if $0<\sigma^2(J,F)<\infty$,

$$\sqrt{n}(S_n - \mu(J,F)) \overset{\mathcal{L}}{\to} N(0, \sigma^2(J,F))$$

where

$$\mu(J,F) = \int_0^1 J(u)F^{-1}(u)du$$

$$= \int_q^1 F^{-1}(u)du - p\int_0^1 F^{-1}(u)du$$

$$= p\mu_p, \quad \text{since} \quad \mu = 0,$$

and $\sigma^2(J,F)$ is as given in (3.4.2).

Dividing the region of integration into four subregions, viz.,

$$\{x<\xi_q,\ y<\xi_q\},\ \{x<\xi_q, y\geq\xi_q\},\ \{x\geq\xi_q,\ y<\xi_q\},\ \{x\geq\xi_q,\ y\geq\xi_q\}$$

and using the approach employed in Lemma 3.4.2 one obtains

$$\sigma^2(J,F) = pq(p\bar{\sigma}_q^{\ 2} + q\sigma_p^{\ 2} + (p\bar{\mu}_q+q\mu_p-\xi_q)^2)$$

on simplification. Here $\bar{\mu}_q$ and $\bar{\sigma}_q^{\ 2}$ are the mean and the variance of the df G* given by

$$G^*(x) = \begin{cases} F(x)/q, & x\leq\xi_q \\ 1, & x>\xi_q \end{cases}$$

and μ_p and $\sigma_p^{\ 2}$ are as described in Lemma 3.4.2. $\sigma^2(J,F)$ is indeed a finite positive quantity. Now, recalling that $\sqrt{n}(p\hat{D}_{k,n}(\sigma) - S_n) \overset{P}{\to} 0$ and that $k = [np]$, we have proved the following result.

Theorem 4.3.1:

Let ξ_q, the q^{th} quantile of the parent df F, be unique and $\int [F(x)(1-F(x))]^{1/2}dx$ be finite. Then

$$\sqrt{k}(\hat{D}_{k,n}(\sigma)-\mu_p) \xrightarrow{\mathcal{L}} N(0,\ q(p\bar{\sigma}_q^{\ 2} + q\sigma_p^{\ 2} + (p\bar{\sigma}_q+q\mu_p-\xi_q)^2)) \tag{4.3.1}$$

where μ_p, $\sigma_p^{\ 2}$, $\bar{\mu}_q$, $\bar{\sigma}_q^{\ 2}$ are as described above.

The quantities μ_p, $\bar{\mu}_q$, μ and $\sigma_p^{\ 2}$, $\bar{\sigma}_q^{\ 2}$, σ^2 satisfy the following relations:

$$\mu = p\mu_p + q\bar{\mu}_q$$

$$\sigma^2 = p(\sigma_p^{\ 2} + (\mu_p-\mu)^2) + q(\bar{\sigma}_q^{\ 2}+(\bar{\mu}_q-\mu)^2).$$

Since we have $\mu = 0$, $\sigma = 1$, it follows that $\bar{\mu}_q = -p\mu_p/q$, $q\bar{\sigma}_q^{\ 2} = 1-p\sigma_p^{\ 2} - p\mu_p^{\ 2}/q$. Hence, the limiting variance in (4.3.1) can be written as $p + (q-p)\sigma_p^{\ 2} + q(\mu_p-\xi_q)^2 - 2p\mu_p(\mu_p-\xi_q)$; that is, the limiting variance of $\sqrt{k}\ \hat{D}_{k,n}(\sigma)$ can be written as a function of μ_p, $\sigma_p^{\ 2}$ and ξ_q. From (3.3.8), we know that the limiting variance of $\sqrt{k}\ D_{k,n}$ is $\sigma_p^{\ 2} + q(\mu_p-\xi_q)^2$. Hence the limiting variance of $\sqrt{k}\ \hat{D}_{k,n}(\sigma)$ is smaller than that of $\sqrt{k}\ D_{k,n}$ iff

$$p + (q-p)\sigma_p^{\ 2} - 2p\mu_p(\mu_p-\xi_q) < \sigma_p^{\ 2},$$

that is, if

$$\sigma_p^{\ 2} + \mu_p(\mu_p-\xi_q) > 1/2.$$

From (3.6.2), for a standard normal parent, $\sigma_p^2 + \mu_p(\mu_p - \xi_q) = 1$ and hence in this case $\mathrm{Var}(\sqrt{k}\,\hat{D}_{k,n}(\sigma)) < \mathrm{Var}(\sqrt{k}\,D_{k,n})$ asymptotically.

Note:

One can show that, when σ is also replaced by its estimate S,

$$\sqrt{k}(\hat{D}_{k,n} - \mathcal{E}(\hat{D}_{k,n}(\sigma))/S)) \to N(0, q(p\bar{\sigma}_q^{\,2} + q\sigma_p^{\,2} + (p\bar{\mu}_q + q\mu_p - \xi_q)^2)$$

whenever ξ_q is unique. However, our approach does not permit us to replace the stochastic centering quantity $(\mathcal{E}\hat{D}_{k,n}(\sigma))/S$ by a nonrandom quantity.

4.4. Asymptotic Distribution of $D_{k,n}$ in the Independent, Nonidentically Distributed, Quantile Case

We start with Stigler's (1974) Theorem 6, which forms the basis of our discussion of the asymptotic theory for $D_{k,n}$ when the variables involved are independent, not necessarily identically distributed.

For each $n \geq 1$, let $X_{1n}, X_{2n}, \ldots, X_{nn}$ be n independent rvs with (possibly different) dfs $F_{1n}, F_{2n}, \ldots, F_{nn}$ where the F_{in}'s are arbitrary dfs. Let $X_{1:n} \leq \ldots \leq X_{n:n}$ denote the order statistics of this sample and define $S_n = n^{-1} \sum_{i=1}^{n} J(i/(n+1)) X_{i:n}$.

Theorem 4.4.1: (Stigler)

Suppose that there is a df G with associated rv Y such that $\mathcal{E}Y^2$ is finite and whenever $y \le -M$, $F_{jn}(y) \le G(y)$ and whenever $y \ge M$, $F_{jn}(y) \ge G(y)$ where M is some constant. Assume that both

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} F_{jn}(x) = F(x) \tag{4.4.1a}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} [F_{jn}(\min(x,y)) - F_{jn}(x)F_{jn}(y)] = K(x,y) \tag{4.4.1b}$$

exist for a.e. x, y wrt Lebesgue measure. Then, if J(u) is bounded and continuous a.e. F^{-1}, $n\sigma^2(S_n) \to \sigma^2(J,F,K)$, given below, and if $\sigma^2(J,F,K) > 0$, then

$$\sqrt{n}\ (S_n - \mathcal{E}S_n) \xrightarrow{\mathcal{L}} N(0,\sigma^2(J,F,K)) \tag{4.4.2}$$

as $n\to\infty$. Here

$$\sigma^2(J,F,K) = \iint J(F(x))J(F(y))K(x,y)\,dxdy. \tag{4.4.3}$$

If $\sqrt{n}(\mathcal{E}S_n - \mu(J,F)) \to 0$ as $n\to\infty$, $\mathcal{E}S_n$ in (4.4.2) can be replaced by $\mu(J,F) = \int J(u)F^{-1}(u)\,du$.

Stigler (1974) also points out that if $S_n' = n^{-1}\sum_{i=1}^{n} J_n(i/(n+1))X_{i:n}$, where the J_n's are uniformly bounded, and for every continuity point p_o of J there is an open neighborhood of p_o such that $J_n(u) \to J(u)$ uniformly in this neighborhood, then the conclusion of Theorem 4.4.1 is true for S_n' also.

Define

$$M_{k,n} = \frac{1}{k} \sum_{i=n-k+1}^{n} X_{i:n}, \quad \text{where} \quad k = [np]$$

$$= \frac{1}{k} \sum_{i=1}^{n} J_n\left(\frac{i}{n+1}\right) X_{i:n}$$

where

$$J_n(u) = \begin{cases} 1, & u \geq (n-[np]+1)/(n+1) \\ 0, & u < (n-[np]+1)/(n+1). \end{cases}$$

It is easy to see that $J_n(u) \to J(u)$, defined by (3.4.5), namely,

$$J(u) = \begin{cases} 1, & u \geq q \\ 0, & u < q \end{cases}$$

and the convergence is uniform around every continuity point of J. Furthermore, the J_n's are uniformly bounded. Conditions imposed in the theorem also ensure that F is necessarily a df and if F has unique q^{th} quantile ξ_q, J will be continuous a.e. F^{-1}. Hence, from Theorem 4.4.1 and the succeeding observations, we conclude that under the assumptions of that theorem,

$$\sqrt{n}\left(\frac{k}{n} M_{k,n} - \mathcal{E}\left(\frac{k}{n} M_{k,n}\right)\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2(J,F,K))$$

or, in other words

$$\sqrt{k}(M_{k,n} - \mathcal{E}M_{k,n}) \xrightarrow{\mathcal{L}} N(0, \sigma^2(J,F,K)/p)$$

where $\sigma^2(J,F,K)$ is given by (4.4.3) with J as in (3.4.5).

Simplification in the expression for $\sigma^2(J,F,K)$ is possible if we assume that $(\sum_{j=1}^{n} F_{jn}(x)F_{jn}(y))/n \to F(x)F(y)$ as $n\to\infty$ instead of making the weaker assumption (4.4.1b). Then $K(x,y) = F(\min(x,y)) - F(x)F(y)$ and consequently $\sigma^2(J,F,K) = \sigma^2(J,F)$ given by (3.4.7). To be precise, we obtain the following result.

<u>Theorem 4.4.2</u>:

For each $n\geq 1$ let $X_{1n}, X_{2n}, \ldots, X_{nn}$ be n independent rvs with dfs $F_{1n}, F_{2n}, \ldots, F_{nn}$. Define $M_{k,n} = (\sum_{i=n-k+1}^{n} X_{i:n})/k$, $k = [np]$, $0<p<1$. Suppose there exists a df G and an associated rv Y such that $\mathcal{E}Y^2$ is finite and whenever $y \leq -M$, $F_{jn}(y) \leq G(y)$ and whenever $y \geq M$, $F_{jn}(y) \geq G(y)$ where M is some finite constant. Assume also that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} F_{jn}(x) = F(x)$$

and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} F_{jn}(x)F_{jn}(y) = F(x)F(y) \tag{4.4.4}$$

exist for a.e. x,y wrt Lebesgue measure. If F has unique q^{th} quantile ξ_q, then

$$\sqrt{k}\,(M_{k,n}-\mathcal{E}M_{k,n}) \xrightarrow{\mathcal{L}} N(0,\ \sigma_p^2 + q(\mu_p-\xi_q)^2)$$

where μ_p and σ_p^2 are as described in Lemma 3.4.2.

Remarks:

(i) If it can be shown that $\sqrt{k}(\mathcal{E}M_{k,n}-\mathcal{E}M_{k,n}^F) \to 0$, where $M_{k,n}^F$ corresponds to the iid case with the parent df F, then it follows that $\sqrt{k}(\mathcal{E}M_{k,n}-\mu_p) \to 0$ if $\int [F(x)(1-F(x))]^{1/2}dx$ is finite. This is because in that situation it is known from Theorem 3.4.1, that $\sqrt{k}(\mathcal{E}M_{k,n}^F-\mu_p) \to 0$. Hence, with these additional assumptions one can replace $\mathcal{E}M_{k,n}$ by μ_p in the above result.

(ii) If all the F_{jn}'s have the same mean and variance, WLOG one can take the mean to be zero and the variance to be unity. Then $D_{k,n}$, the selection differential, takes the place of $M_{k,n}$ in the above theorem and in Remark (i).

Example 4.4.1:

Let one of $F_{1n}, F_{2n}, \ldots, F_{nn}$ be F* and the rest all be F where F*<F, i.e., one of the populations has slipped to the right. Let $\int (F(x)(1-F(x)))^{1/2}dx$ be finite and let F have the unique quantile ξ_q. Assume that F* has finite variance. Define

$$G(y) = \begin{cases} F(y), & y \leq -M \\ F(-M), & -M<y<M \\ F^*(y), & y \geq M \end{cases}$$

where M is such that $F(-M)<F^*(M)$. This is possible since both F and F* are dfs. Here it is immediate that (4.4.4) is satisfied. Hence, from Theorem 4.4.2, we have

$$\sqrt{k}(M_{k,n}-\mathcal{E}M_{k,n}) \xrightarrow{\mathcal{L}} N(0,\ \sigma_p^{\ 2} + q(\mu_p-\xi_q)^2) \text{ as } n\to\infty.$$

We will now show that $\sqrt{k}(\mathcal{E}M_{k,n}-\mathcal{E}M^F_{k,n}) \to 0$ in the following steps.

The df of $X_{r:n}$ is (David and Shu, 1978),

$$H_{r:n}(x) = F_{r:n-1}(x) + \binom{n-1}{r-1}F^{r-1}(x)(1-F(x))^{n-r}F^*(x),$$

$$r = 1,2,\ldots,(n-1) \qquad (4.4.5)$$

where $F_{s:m}$ is the df of the s^{th} order statistic from a random sample of size m from the df F. Also, the df of $X^F_{r:n}$, viz $F_{r:n}$, satisfies

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

$$r = 1,2,\ldots,(n-1)$$

so that

$$F_{r:n}(x)-H_{r:n}(x) = \binom{n-1}{r-1}F^{r-1}(x)[1-F(x)]^{n-r}[F(x)-F^*(x)].$$

This is true for $r = 1,2,\ldots,n$.

Since

$$\mathcal{E}X_{r:n} = \int_0^\infty [1-H_{r:n}(x)]dx - \int_{-\infty}^0 H_{r:n}(x)dx, \text{ it follows that}$$

$$\mathcal{E}(X_{r:n}-X^F_{r:n}) = \int_{-\infty}^\infty [F_{r:n}(x)-H_{r:n}(x)]dx.$$

Hence

$$\mathcal{E}(M_{k,n}-M^F_{k,n}) = \frac{1}{k}\int_{-\infty}^\infty \{\sum_{j=n-k+1}^{n} \binom{n-1}{j-1}F^{j-1}(x)(1-F(x))^{n-j}\}$$

$$\cdot\ [F(x)-F^*(x)]dx.$$

Since $F^* < F$ and $\sum_{j=n-k+1}^{n} \binom{n-1}{j-1} F^{j-1}(x)(1-F(x))^{n-j} \le 1$
we have

$$0 \le \mathcal{E}(M_{k,n} - M^F_{k,n}) \le \frac{1}{k} \int_{-\infty}^{\infty} [F(x) - F^*(x)]dx$$

$$= \frac{1}{k}(\mu(F) - \mu(F^*)).$$

Therefore

$$\sqrt{k}\ \mathcal{E}(M_{k,n} - M^F_{k,n}) \to 0 \quad \text{as} \quad n \to \infty.$$

Hence, from Remark (i) above, it follows that:

$$\sqrt{k}(M_{k,n} - \mu_p) \xrightarrow{\mathcal{L}} N(0,\ \sigma_p{}^2 + q(\mu_p - \xi_q)^2).$$

Note:

This example can be generalized to handle the case when we have more than one slipped population. Then we will have to write general versions of (4.4.5). Except for messier algebra, we do not expect any other problem here. But, if we have a proportion of the populations slipped, replacing $\mathcal{E}M_{k,n}$ by a fixed centering constant does not appear to be possible even though Theorem 4.4.2 holds in this situation.

4.5. Some Special Dependent Cases

Under the assumptions of independence we obtained the asymptotic distribution of $D_{k,n}$ both in the extreme and the quantile cases. We obtained the same results for the independent case as for the iid case, but of course, under some additional assumptions. The following two situations show that these results are not necessarily true for dependent samples.

Equicorrelated Normal Case:

Let X_i, $i = 1,2,\ldots,$ be equicorrelated standard normal rvs with the common correlation coefficient $\rho(>0)$. Then it is well-known that the X_i's can be represented as

$$X_i = \sqrt{\rho}\, U + \sqrt{1-\rho}\, Y_i, \quad i \geq 1$$

where U, $Y_1, Y_2, \ldots$ are all mutually independent standard normal rvs. When k is fixed, since $\Phi \varepsilon D(\Lambda)$,

$$\left(\frac{1}{k} \sum_{i=n-k+1}^{n} Y_{i:n} - a_n\right)/b_n \xrightarrow{\mathcal{L}} B/k + T_k^{(3)}$$

as given by (3.2.4c). Also, a_n and b_n can be chosen to satisfy (3.6.1). $D_X(k,n) = (\sum_{i=n-k+1}^{n} X_{i:n})/k$ here and

$$\frac{D_X(k,n)-a_n^*}{b_n^*} = \frac{\sqrt{\rho}\, U}{b_n^*} + \frac{\sqrt{1-\rho}\, D_Y(k,n)-a_n^*}{b_n^*} .$$

$\sqrt{\rho}\, U/b_n^*$ has a nondegenerate limit law iff b_n^* converges to a nonzero finite number. For $(\sqrt{1-\rho}\, D_Y(k,n)-a_n^*)/b_n^*$ to have a nondegenerate limit law one has to take $b_n^* \sim (\sqrt{2 \log n})^{-1}$ which converges to zero. Hence, if (a_n, b_n) are the appropriate norming constants in the iid normal case, $(\sqrt{1-\rho}\, a_n, b_n)$ would not normalize $D_X(k,n)$ to yield a non-degenerate limit law. However, $D_Y(k,n) - \sqrt{2 \log n} \xrightarrow{P} 0$ (see

Section 3.6), and hence

$$(D_X(k,n) - \sqrt{2(1-\rho)\log n})/\sqrt{\rho} \xrightarrow{\mathcal{L}} N(0,1).$$

Therefore, $(\sqrt{2(1-\rho)\log n},\ \sqrt{\rho})$ can be used as a pair of norming constants and the only possible nondegenerate limit law is normal.

In the quantile case, since $D_Y(k,n) \xrightarrow{P} \mu_p$,

$$D_X(k,n) = \sqrt{\rho}\ U + \sqrt{1-\rho}\ \ D_Y(k,n) \xrightarrow{\mathcal{L}} N(\sqrt{1-\rho}\ \ \mu_p,\ \rho).$$

That is, $(D_X(k,n) - \sqrt{1-\rho}\ \mu_p)/\sqrt{\rho}$ is asymptotically standard normal. As a contrast, if the X_i's were independent also, one would obtain $\sqrt{k}(D_X(k,n)-\mu_p)$ to be asymptotically normal.

<u>Stationary Gaussian Process</u>:

Let $\{X_i,\ i = 0, \pm 1, \pm 2, \ldots\}$ be a stationary Gaussian sequence with $\mathcal{E}X_n = 0$, $\mathcal{E}X_iX_{i+n} = r_n$. If $r_n \log n \to 0$ as $n\to\infty$, Welsch (1973) has shown that, when k is fixed, the asymptotic distribution of $((X_{n:n}-a_n)/b_n, \ldots, (X_{n-k+1:n}-a_n)/b_n)$ is the same as in the iid standard normal case. The same norming constants a_n, b_n work in both cases. (He has shown this for k=2; but the result is true in general.) Hence,

$$(D_{k,n}-a_n)/b_n \xrightarrow{\mathcal{L}} B/k + T_k^{(3)}$$

agreeing with the independent standard normal case.

4.6. Application of the Asymptotic Theory to Testing for Outliers

In this section we obtain approximate percentage points for $D_{k,n}$ for large n when the parent population is normal using the asymptotic theory developed in Chapter 3. This is of considerable interest in the outlier testing problem which is described below.

Let $X_1, X_2, \ldots, X_n$ be independent rvs, $X_i \sim N(\mu_i, \sigma^2)$. Consider the problem of testing the hypothesis

H: $\mu_1 = \mu_2 = \ldots = \mu_n = \mu$
against the alternative

A: k of these μ_i's are equal to $\mu+\delta\ (\delta>0)$ and the remainder are equal to μ.

Then, $(X_{n-k+1:n} + \ldots + X_{n:n} - k\mu)/\sigma = (S_{k,n}-k\mu)/\sigma = kD_{k,n}$ can be used as a test statistic, when μ and σ are known. In fact, when μ and/or σ are estimated by $\overline{X}$ and/or S, Barnett and Lewis (1978) point out that the test which rejects H for large values of $(S_{k,n}-k\overline{X})/S$ is the likelihood ratio test for a location slippage alternative in which k observations arise from a common normal distribution $N(\mu+\delta, \sigma^2)$, $\delta>0$, i.e., the alternative A, above. For this alternative it has the optimal property of being the scale and location invariant test of given size which maximizes the probability of identifying the k contaminants as discordant (pp. 95-96, 112). Early work on this statistic, due to Murphy (1951), was later followed up by McMillan (1971).

Because of the above motivation, considerable attention has been given to the distribution of $D_{k,n}$ and its percentage points. We assume that μ and σ are known and WLOG take $\mu = 0$, $\sigma = 1$. Then, we compare the approximate percentage points for $D_{k,n}$ under H obtained using the asymptotic theory assuming: (a) k fixed, (b) $k = [np]$, $0<p<1$, and (c) Table IXg of Barnett and Lewis (1978), which is based on simulation.

Approach (a):

When k is fixed, since $\Phi \varepsilon D(\Lambda)$, the percentage points of the limiting distribution of $(D_{k,n}-a_n)/b_n$, as $n \to \infty$, are given by Table 3.2.1. But now, the problem is to use "good" choices of a_n and b_n. Often these are given by (3.6.1), namely

$$a_n = \sqrt{2 \log n} - (\log \log n + \log 4\pi)/2\sqrt{2 \log n}$$

and

$$b_n = 1/\sqrt{2 \log n}. \qquad (4.6.1)$$

It is worth recalling that any other sequence a_n' and b_n' such that $b_n/b_n' \to 1$ and $(a_n-a_n')/b_n \to 0$ as $n \to \infty$ would serve asymptotically. Recently, Hall (1979), has shown that the best rate of convergence of

$$\sup_{-\infty<x<\infty} |\Phi^n(a_n+b_nx)-\Lambda(x)|$$

is achieved when a_n and b_n are chosen such that

$$2\pi a_n^2 \exp(a_n^2) = n^2 \quad \text{and} \quad b_n = 1/a_n, \tag{4.6.2}$$

the rate being of the order of 1/log n.

Let a_n^* and b_n^* be the solutions of (4.6.2). The following table illustrates the differences in a_n and b_n as given by (4.6.1), and a_n^*, b_n^*.

Table 4.6.1. Values of the norming constants for selected n

n	a_n	b_n	a_n^*	b_n^*
30	1.8882	.3834	1.9146	.5223
50	2.1009	.3575	2.1118	.4735
100	2.3663	.3295	2.3753	.4210
500	2.9075	.2836	2.9080	.3439
1000	3.1165	.2690	3.1153	.3210

The approximate percentage points of $D_{k,n}$ are then given by $a_n+b_n\xi_{k,p}$ and $a_n^* + b_n^*\xi_{k,p}$ for the two choices of constants, and are labeled Ext(a_n,b_n) and Ext(a_n^*,b_n^*), respectively.

Approach (b):

For given n and k, we can take k/n = p and use the asymptotic theory of the quantile case. Then, from Theorem 3.3.1, it follows that

$$\sqrt{k}(D_{k,n}-\mu_p) \xrightarrow{\mathcal{L}} N(0, \sigma_D^2) \tag{4.6.3}$$

where $\sigma_D^2 = \sigma_p^2 + q(\mu_p-\xi_q)^2$ is tabulated by Burrows (1975) for various values of p. Also, $\mu_p = \phi(\xi_q)/p$ makes it easy to compute μ_p in the standard normal parent case. Burrows (1972) has also obtained a good approximation of $\mathcal{E}D_{k,n}$ which converges to μ_p at the rate of 1/n. Hence, we may also use his approximation, namely

$$\hat{\mu}_p = \mu_p - \frac{n-k}{2k(n+1)} \cdot \frac{1}{\mu_p} \tag{4.6.4}$$

instead of μ_p in (4.6.3). These give another pair of percentage points for $D_{k,n}$, namely $\mu_p + z_\alpha\sigma_D/\sqrt{k}$ and $\hat{\mu}_p + z_\alpha\sigma_D/\sqrt{k}$ where z_α is the upper α percentile point for N(0,1). These are labeled Qnt(μ_p) and Qnt($\hat{\mu}_p$), respectively.

Approach (c):

This is the simulation approach used in the construction of Table IXg of Barnett and Lewis (1978), and the percentage points so obtained are labeled Sim (B&L). We compare these five approximate percentage points for $D_{k,n}$ for k = 2,3,4, n = 20,30,40,50,100 at the 95% and 99% level in Table 4.6.2 below.

Table 4.6.2. Five approximations to the percentage points of $D_{k,n}$ for the normal parent population

n	95% points					99% points				
	Ext (a_n,b_n)	Ext (a_n^*,b_n^*)	Qnt(μ_p)	Qnt$(\hat{\mu}_p)$	Sim(B&L)	Ext (a_n,b_n)	Ext (a_n^*,b_n^*)	Qnt(μ_p)	Qnt$(\hat{\mu}_p)$	Sim(B&L)
k = 2										
20	2.44	2.78	2.46	2.34	2.37	2.86	3.36	2.76	2.64	2.72
30	2.58	2.85	2.61	2.50	2.51	2.97	3.38	2.89	2.78	2.84
40	2.67	2.92	2.70	2.58	2.62	3.05	3.42	2.97	2.85	2.93
50	2.74	2.97	2.78	2.67	2.68	3.10	3.45	3.04	2.93	3.02
100	2.96	3.13	3.00	2.90	2.92	3.29	3.56	3.23	3.13	3.20
k = 3										
20	2.18	2.41	2.17	2.08	2.10	2.50	2.85	2.43	2.34	2.39
30	2.33	2.52	2.33	2.25	2.26	2.63	2.92	2.57	2.49	2.54
40	2.43	2.60	2.44	2.36	2.38	2.72	2.98	2.67	2.59	2.63
50	2.51	2.66	2.52	2.44	2.45	2.79	3.03	2.74	2.66	2.72
100	2.75	2.86	2.76	2.69	2.70	3.00	3.19	2.96	2.89	2.94
k = 4										
20	2.00	2.15	1.96	1.89	1.90	2.26	2.53	2.20	2.13	2.16
30	2.16	2.29	2.14	2.08	2.08	2.41	2.63	2.36	2.30	2.32
40	2.27	2.38	2.26	2.20	2.21	2.51	2.70	2.46	2.40	2.43
50	2.36	2.46	2.34	2.28	2.28	2.59	2.76	2.54	2.48	2.53
100	2.60	2.68	2.60	2.54	2.55	2.82	2.95	2.78	2.72	2.78

<u>Observations and comments:</u>

(i) $\mathrm{Ext}(a_n,b_n)$ does much better than $\mathrm{Ext}(a_n^*,b_n^*)$ for all n, k and the percentages considered, in the sense that it is much closer to Sim(B&L) than the latter. Even though a_n^* and b_n^* are supposed to make the convergence of the df of $X_{n:n}$ faster in the sense of the supremum over the entire real line, $\mathrm{Ext}(a_n^*,b_n^*)$ does not perform well at the 95th and 99th percentile points of $D_{k,n}$.

(ii) At the 95 percent level, $\mathrm{Qnt}(\hat{\mu}_p)$ comes closest to Sim(B&L) being within 0.01 of the latter for $k \geq 3$, $n \geq 30$. However, $\mathrm{Qnt}(\hat{\mu}_p) \leq \mathrm{Sim(B\&L)}$. This suggests that one could use $\mathrm{Qnt}(\hat{\mu}_p)$ to find 95 percent points when $k \geq 3$, $n \geq 30$. It may be noted also that $\mathrm{Ext}(a_n,b_n)$ and $\mathrm{Qnt}(\mu_p)$ approach each other as n increases for $k \geq 3$, even though both are off from Sim(B&L).

(iii) At the 99 percent level $\mathrm{Qnt}(\mu_p)$ does very well indeed, doing better with increased k for a given n.

We now consider some large values of n in an attempt to search for a trend which can be of some help in determining which of these approaches is desirable.

These do not seem to give much insight except to show that for k=4, the $\mathrm{Qnt}(\mu_p)$ and $\mathrm{Ext}(a_n,b_n)$ actually coincide at 95 percent level as is evident from Table 4.6.3 below.

In conclusion, the empirical evidence expressed in Table 4.6.2 seems to suggest that $\mathrm{Qnt}(\hat{\mu}_p)$ provides a close

Table 4.6.3. Approximate 95% points for large n

n	k=2			k=4	
	$Ext(a_n,b_n)$	$Qnt(\mu_p)$	$Qnt(\hat{\mu}_p)$	$Ext(a_n,b_n)$	$Qnt(\mu_p)$
200	3.16	3.30	3.21	2.83	2.83
400	3.36	3.39	3.30	3.04	3.04
500	3.42	3.45	3.37	3.11	3.11
1000	3.47	3.63	3.56	3.31	3.31

approximation at the 95 percent level whereas $Qnt(\mu_p)$ does well at the 99 percent level. For extremely small p (<.005) it might be safer to use $Ext(a_n,b_n)$ rather than the rest.

So far, in our discussion, it was assumed that μ and σ are known. When these are estimated by $\bar{X}$ and S, since $a_n/\sqrt{n}\, b_n \to 0$, from Corollary 1 to Theorem 4.1.2, it follows that the percentage points of the asymptotic distribution of $(\hat{D}_{k,n}-a_n)/b_n$ are the same as those corresponding to $(D_{k,n}-a_n)/b_n$. Hence, our approximations $Ext(a_n,b_n)$, $Ext(a_n^*,b_n^*)$, obtained using the "extreme case" approach remain the same. However, these values fall far away from the simulated percentage points of $\hat{D}_{k,n}$ given by Table IXa of Barnett and Lewis (1978). The quantile case can be used only when σ is known (see Section 4.3). Using Theorem 4.3.1, in this case, one obtains a different set of values for $Qnt(\mu_p)$ as an approximation to the percentage points of $\hat{D}_{k,n}(\sigma)$. The actual computations and comparisons with the simulated percentage points given by Table IXe of Barnett and Lewis (1978) will not be presented.

V. GENERAL DISTRIBUTION THEORY FOR THE INDUCED SELECTION DIFFERENTIAL

In plant and animal breeding, quantity of interest is the "response to selection", i.e., the difference between the mean phenotypic value of the offspring of the selected parents and the mean of the entire population. In breeding problems we select the top p fraction of the parental population and are interested in the performance of their offspring, compared to that of the whole population. A natural measure of performance is provided by the induced selection differential, that is the selection differential based on "concomitants". We study this quantity in the present chapter.

5.1. Finite Sample Theory for the Induced Selection Differential ($D_{[k,n]}$)

Let (X_i, Y_i), i = 1 to n, be iid rvs each having df $F_{X,Y}(x,y)$ where the X_i's are assumed to be continuous with df F_X. Let $X_{r:n}$ be the rth order statistic of the X values and let $Y_{[r:n]}$ be the Y variate paired with $X_{r:n}$. Then $Y_{[r:n]}$ is called the concomitant of $X_{r:n}$. Let μ_Y and σ_Y^2 be the mean and variance of the distribution of Y_i's. Then the induced selection differential, $D_{[k,n]}$ is defined by

$$D_{[k,n]} = \frac{1}{k} \sum_{i=n-k+1}^{n} (Y_{[i:n]} - \mu_Y)/\sigma_Y.$$

If $F_{X,Y}(x,y)$ is an absolutely continuous df, then from Yang (1977, p. 997), we have

$$f_{Y_{[n-k+1:n]},\ldots,Y_{[n:n]}}(y_1,\ldots,y_k)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_2} \prod_{i=1}^{k} f(y_i|x_i) f_{X_{n-k+1:n},\ldots,X_{n:n}}$$

$$(x_1,\ldots,x_k)\,dx_1 dx_2 \ldots dx_k$$

$$= \frac{n!}{(n-k)!} \int_{-\infty}^{\infty} \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_2} \prod_{i=1}^{k} f(y_i|x_i)$$

$$\cdot\ [F_X(x_1)]^{n-k} f_X(x_1) \ldots f_X(x_k)\,dx_1 \ldots dx_k.$$

This can be used to obtain the distribution of $D_{[k,n]}$ even though a closed form expression may not be possible.

Bounds on $\mathcal{E}D_{[k,n]}$:

From (6.1) of Mallows and Richter (1969) we have

$$(M_{[k,n]} - \bar{Y})^2 \le \frac{n-k}{k} s_Y^2,$$

where $M_{[k,n]} = \frac{1}{k} \sum_{i=n-k+1}^{n} Y_{[i:n]} = \mu_Y + \sigma_Y D_{[k,n]}$, $\bar{Y}$ is the mean of the Y values and $s_Y^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$. That is,

$$\overline{Y} - \sqrt{\frac{n-k}{k}}\, s_Y \le M_{[k,n]} \le \overline{Y} + \sqrt{\frac{n-k}{k}}\, s_Y .$$

Hence, taking expectations and noting that $\mathcal{E}s_Y \le (\mathcal{E}s_Y{}^2)^{1/2} \le \sigma_Y\sqrt{\frac{n-1}{n}}$, we obtain

$$\mu_Y - \sqrt{\frac{n-k}{k}}\,\sigma_Y \cdot \sqrt{\frac{n-1}{n}} \le \mathcal{E}M_{[k,n]} \le \mu_Y + \sqrt{\frac{n-k}{k}} \cdot \sigma_Y\sqrt{\frac{n-1}{n}} .$$

Since $(\mathcal{E}M_{[k,n]}-\mu_Y)/\sigma_Y = \mathcal{E}D_{[k,n]}$, it follows that

$$-\sqrt{\frac{(n-k)(n-1)}{nk}} \le \mathcal{E}D_{[k,n]} \le \sqrt{\frac{(n-k)(n-1)}{nk}} . \qquad (5.1.1)$$

If our sample were not random then $(\mathcal{E}s_Y{}^2)^{1/2} \le \sigma_Y$ and hence one obtains

$$-\sqrt{\frac{n-k}{k}} \le D_{[k,n]} \le \sqrt{\frac{n-k}{k}}$$

in the dependent sample case.

So far, we have not made use of the fact that the Y's are the concomitants of the order statistics. To exploit this fact, we further assume that $\mathcal{E}(Y|X = x) = m(x)$ is a monotonic function of x. WLOG take m(x) to be increasing. Let $X^* = m(X)$. Then $\mathcal{E}Y_{k:n} = \mathcal{E}m(X_{k:n}) = \mathcal{E}X^*_{k:n}$. Noting that $X^*_{1:n},\ldots,X^*_{n:n}$ are the order statistics from the distribution of m(X), we can use the bounds for $\mathcal{E}D_{k,n}$, obtained in Section 2.6. Using (2.6.5), we have

$$\mathcal{E}m(X) + \frac{n-k}{t}\,\frac{\mathcal{E}s_{m(X)}}{\sqrt{n-1}} \le \mathcal{E}M_{[k,n]} \le \mathcal{E}m(X) + \sqrt{\frac{n-k}{k}}\,\mathcal{E}s_{m(X)}$$

where $t = \max(k,n-k)$. Note that $\mathcal{E}s_{m(X)} \le \sigma_{m(X)}\sqrt{(n-1)/n}$, $\mathcal{E}m(X) = \mu_Y$. Hence, the above inequality can be rearranged to yield

$$\frac{n-k}{\max(k,n-k)} \; \frac{\mathcal{E}s_{m(X)}}{\sigma_Y\sqrt{n-1}} \le \mathcal{E}D_{[k,n]} \le \sqrt{\frac{(n-k)(n-1)}{nk}} \; \frac{\sigma_{m(X)}}{\sigma_Y} . \tag{5.1.2}$$

Since

$$\begin{aligned} \sigma_Y^2 &= \mathrm{Var}(\mathcal{E}(Y|X)) + \mathcal{E}(\mathrm{Var}(Y|X)) \\ &= \mathrm{Var}(m(X)) + \mathcal{E}(\mathrm{Var}(Y|X)) \\ &\ge \sigma_{m(X)}^2 , \end{aligned}$$

the upper bound in (5.1.2) is better than the one in (5.1.1). The same is the case with lower bounds. Of course we have used the fact that $\mathcal{E}(Y|X)$ is increasing in obtaining (5.1.2). With the same assumptions, one can obtain tighter bounds for $\mathcal{E}D_{[k,n]}$ using the techniques of Section 2.4 and 2.5. The details are omitted.

5.2. Asymptotic Distribution of $D_{[k,n]}$ in the Extreme Case

Let (X_i,Y_i), $i=1$ to n, be iid bivariate absolutely continuous rvs with pdf $f(x,y)$ and df $F(x,y)$. WLOG we take $\mu_Y = 0$ and $\sigma_Y = 1$. We consider the cases where $x_o = F_X^{-1}(1) < \infty$ and where $x_o = +\infty$ separately.

Case A: $x_o < \infty$:

From Yang (1977) we have

$$P(Y_{[n-k+1:n]} \le y_1, \ldots, Y_{[n:n]} \le y_k)$$

$$= \int \cdots \int_{x_1 < \ldots < x_k} \prod_{i=1}^{k} P(Y_i \le y_i \mid X_i = x_i)$$

$$\cdot dF_{X_{n-k+1:n}, \ldots, X_{n:n}}(x_1, \ldots, x_k).$$

$(X_{n-k+1:n}, \ldots, X_{n:n}) \xrightarrow{P} (x_o, \ldots, x_o)$, as $n \to \infty$ since k is a fixed integer. Hence, following Yang's (1977) Theorem 2.1 we have

$$\lim_{n\to\infty} P(Y_{[n-k+1:n]} \le y_1, \ldots, Y_{[n:n]} \le y_k)$$

$$= \prod_{i=1}^{k} P(Y_i \le y_i \mid X_i = x_o).$$

Therefore, $D_{[k,n]} = (\sum_{i=n-k+1}^{n} Y_{[i:n]})/k$ converges in law to the average of k iid rvs each having the pdf $f(y|x_o)$.

Case B: $x_o = +\infty$:

Theorem 5.2.1:

Let (X_i, Y_i), i = 1 to n be a random sample from a bivariate absolutely continuous distribution. Let $x_o = \infty$ and $F_X^n(a_n + b_n x) \to G(x)$, a nondegenerate df; that is, $F_X \varepsilon D(G)$. If

$$P(Y_1 \le A_n + B_n u \mid X_1 = a_n + b_n z) = T_n(u,z) \to T(u,z)$$

uniformly in z, then

$$P(Y_{[n:n]} \leq A_n + B_n u) \to T(u) = \int T(u,z)dG(z).$$

Proof:

First, fix u and note that $T_n(u,z)$ is a sequence of bounded continuous functions converging uniformly in z to $T(u,z)$. Since $F_X{}^n(a_n+b_nx) \to G(x)$, a df, from Lemma A6 we have

$$\begin{aligned} P(Y_{[n:n]} \leq A_n + B_n u) &= \int T_n(u,z)dF_X{}^n(a_n+b_nz) \\ &\to \int T(u,z)dG(z) \quad \text{as } n\to\infty \\ &= T(u). \end{aligned}$$

Note:

(i) Conditions imposed in the above theorem are sufficient to ensure that

$$P(Y_{[n:n]} \leq A_n + B_n u_1,\ Y_{[n-1:n]} \leq A_n + B_n u_2,\ \ldots,\ Y_{[n-k+1:n]} \leq A_n + B_n u_k) \to H_1(u_1)H_2(u_2)\ldots H_k(u_k)$$

where

$H_i(u) = \int T(u,z)dG(x;i)$. Here $G(x;i)$ is one of the distributions represented by (3.2.2a-c). Hence, under the conditions of Theorem 5.2.1, $(D_{[k,n]}-A_n)/B_n$ converges in distribution to that of the mean of k independent rvs, the i^{th} one having the df H_i.

(ii) Galambos (1978) has given the limiting distribution of $(Y_{[n:n]}-A_n)/B_n$ under a different set of conditions when $G = \Phi_\alpha$, (see his Theorem 5.5.1, which we are paralleling). But his proof appears to be incomplete since the use of the dominated convergence theorem is not justified.

Example 5.2.1:

Let $X_i \sim \text{Exp}(1)$ and $Y_i|X_i = x \sim N(x,1)$. Then $F_X \varepsilon D(\Lambda)$ with $a_n = \log n$ and $b_n = 1$. Also, $x_o = +\infty$ and if we take $A_n = \log n$ and $B_n = 1$ we have

$$\begin{aligned} P(Y_1 \leq A_n+B_n u|X_1 = a_n+b_n z) &= \Phi(A_n+B_n u-a_n-b_n z) \\ &= \Phi(u-z) = T(u,z) \end{aligned}$$

and the convergence is uniform in z. Then, from Theorem 5.2.1 we have

$$P(Y_{[n:n]} \leq \log n + u) \to \int \Phi(u-z)d\Lambda(z).$$

5.3. Asymptotic Distribution of $D_{[k,n]}$ in the Quantile Case

We use the results of Bhattacharya (1976) and Yang (1979) to obtain the asymptotic distribution of $D_{[k,n]}$ when $k = [np]$, $0<p<1$. WLOG we take $\mu_Y = 0$, $\sigma_Y = 1$.

Bhattacharya's (1976) approach:

Bhattacharya has essentially obtained the asymptotic distribution of $D_{[k,n]}$ under the following conditions:

B1. F_X is continuous.

B2. $\beta(x) = \mathcal{E}[(Y-m(x))^4 | X=x]$ is bounded. (5.3.1)

B3. $\sigma^2(x) = Var(Y|X=x)$ is of bounded variation.

B4. $h(t) = m(\xi_X(t))$ is a continuous function where $m(x) = \mathcal{E}(Y|X=x)$ and $\xi_X(t) = F_X^{-1}(t)$.

Define $H_n(t) = n^{-1} \sum_{i=1}^{[nt]} Y_{[i:n]}$ and $H(t) = \int_0^t h(s)ds$.

Then, from Bhattacharya (1976, p. 622) it follows that, for $0<a<b<1$,

$$\sqrt{n}(H_n(t)-H(t)) \Rightarrow \zeta(\psi(t)) + \int_0^t \eta(s)dh(s), \text{ on } [a,b] \quad (5.3.2)$$

where

$\psi(t) = \int_{-\infty}^{\xi_X(t)} \sigma^2(x)dF_X(x)$, ζ is a standard Brownian motion and η is a Brownian bridge independent of ζ.

Here $\Rightarrow$ stands for the convergence of a stochastic process.

Therefore, $A_t = \zeta(\psi(t)) \sim N(0, \psi(t))$.

$B_t = \int_0^t \eta(s)dh(s)$ is also normal because η is a normal process and the integral of such a process is again normal (recall that h is continuous). Further $\mathcal{E}B_t = \int_0^t \mathcal{E}(\eta(s))dh(s) = 0$, and

$$\mathrm{Var}(B_t) = 2\int_{u=0}^{t}\int_{v=0}^{u} \mathcal{E}(\eta(u)\eta(v))\,dh(u)\,dh(v)$$

$$= 2\int_{u=0}^{t}\int_{v=0}^{u} v(1-u)\,dh(u)\,dh(v) \quad \text{(from Billingsley, 1968, p. 65)}$$

$$= \int_{u=0}^{t}\int_{v=0}^{t} [\min(u,v)-uv]\,dh(u)\,dh(v).$$

Hence,

$$\sigma_{B_t}^{2} = \int_{-\infty}^{\xi_X(t)}\int_{-\infty}^{\xi_X(t)} [F_X(\min(x,y))-F_X(x)F_X(y)]\,dm(x)\,dm(y)$$

making the transformation $u = F_X(x)$ and $v = F_X(y)$.

Therefore, from (5.3.2) we have

$$\sqrt{n}(H_n(t)-H(t)) \xrightarrow{\mathcal{L}} N(0,\ \psi(t)+\sigma_{B_t}^{2})$$

since A_t and B_t are independent.

$D_{[k,n]}$ is the average of the concomitants of the top k X-values whereas $H_n(t)$ corresponds to the bottom X-values. Hence, we define

$$H_n^*(t) = \frac{1}{n}\sum_{i=n-[nt]+1}^{n} Y_{[i:n]},$$

$$h^*(t) = m(\xi_X(1-t))$$

and

$$H^*(t) = \int_0^t h^*(s)\,ds = \int_{1-t}^{1} m(\xi_X(s))\,ds.$$

Then, under the conditions B1-B4 of (5.3.1) we have

$$\sqrt{n}(H_n^*(t)-H^*(t)) \xrightarrow{\mathcal{L}} N(0,\ \psi^*(t) + \sigma_{B_t^*}^{2}),\quad 0<t<1$$

where

$$\psi^*(t) = \int_{\xi_X(1-t)}^{\infty} \sigma^2(x)\,dF_X(x)$$

and

$$\sigma^2_{B_t^*} = \int_{\xi_X(1-t)}^{\infty}\int_{\xi_X(1-t)}^{\infty} [F_X(\min(x,y)) - F_X(x)F_X(y)]\,dm(x)\,dm(y). \tag{5.3.3}$$

$D_{[k,n]} = \frac{n}{[np]} H_n^*(p)$ and hence

$$\frac{[np]}{\sqrt{n}}(D_{[k,n]} - \frac{1}{p}H^*(p)) = \sqrt{n}(H_n^*(p) - H^*(p)) + \sqrt{n}\, H^*(p) \cdot \frac{np-[np]}{np} \xrightarrow{\mathcal{L}} N(0,\ \psi^*(p) + \sigma^2_{B_p^*})$$

as $n\to\infty$, since the second term on the right tends to zero.

Therefore,

$$\sqrt{k}(D_{[k,n]} - H^*(p)/p) \xrightarrow{\mathcal{L}} N(0,\ (\psi^*(p) + \sigma^2_{B_p^*})/p).$$

Now

$$\frac{H^*(p)}{p} = \frac{1}{p}\int_{\xi_X(q)}^{\infty} m(x)\,dF_X(x) = \mathcal{E}(m(X)\,|\,X>\xi_X(q))$$

$$= \mu_{m(X)}(p), \quad \text{say where } q = 1-p.$$

Formally, we state this as a theorem.

Theorem 5.3.1:

Let $\mu_Y = 0$, $\sigma_Y = 1$ and $k = [np]$, $0<p<1$. Under the conditions B1-B4 of (5.3.1), as $n\to\infty$,

$$\sqrt{k}(D_{[k,n]}-\mu_{m(X)}(p)) \xrightarrow{\mathcal{L}} N(0, \psi^*(p) + \sigma^2_{B^*_p}) \tag{5.3.4}$$

where ψ^* and $\sigma^2_{B^*_p}$ are defined by (5.3.3).

Remarks:

(i) Bhattacharya's expression for the limiting variance of $\sqrt{n}(H_n(t)-H(t))$ as given on the top of page 623, namely, $D(t) + t(1-t) - 2(1-t)h(t)H(t) - H^2(t)$, is wrong. For the bivariate normal parent case one can show that the above representation does not give the right answer.

(ii) His proof can be used to obtain the joint limit distribution of the selection differential and the induced selection differential. This will be done in the next section.

Yang's (1979) approach:

Recently, Yang, paralleling the work of Stigler (1974), has obtained the asymptotic distribution of linear functions of the concomitants of order statistics. He makes the following assumptions:

Y1. F_X is continuous (same as B1).

Y2. $\mathcal{E}Y^2 < +\infty$ (follows from B3).

Y3. m(x) is a right continuous function of bounded variation in any finite interval (implied by B4).

Y4. J is bounded and continuous a.e. $m(F_X^{-1})$ (implied by B4 for the particular J there).

Let, $S_n = \frac{1}{n}\sum_{i=1}^{n} J(\frac{i}{n+1})Y_{[i:n]}$. Under this set-up, Yang has shown the following:

$$\lim_{n\to\infty} \mathcal{E}S_n = \mu(J,F_X) = \int m(x)J(F_X(x))dF_X(x)$$

$$\begin{aligned}\lim_{n\to\infty} n\,\mathrm{Var}(S_n) &= \sigma^2(J,F_X)\\ &= \int J^2(F_X(x))\sigma^2(x)dF_X(x)\\ &\quad + \iint [F_X(\min(x,y))\\ &\quad - F_X(x)F_X(y)]J(F_X(x))J(F_X(y))dm(x)dm(y).\end{aligned}$$

The first term is comparable with $\psi(t)$ whereas the second corresponds to $\sigma^2_{B_t}$ in Bhattacharya's approach.

If $\sigma^2(J,F_X) > 0$, then $(S_n-\mathcal{E}S_n)/\sqrt{\mathrm{Var}(S_n)} \xrightarrow{\mathcal{L}} N(0,1)$ as $n\to\infty$. Equivalently

$$\sqrt{n}(S_n-\mathcal{E}S_n) \xrightarrow{\mathcal{L}} N(0,\ \sigma^2(J,F_X)). \tag{5.3.5}$$

Remarks:

Bhattacharya (1976) had a particular J function, namely

$$J(u) = \begin{cases} 1, & u\le t\\ 0, & u>t\end{cases}$$

and had more conditions than Yang. But, his result is stronger than Yang's in three respects: (i) He has a fixed centering constant, H(t) whereas Yang's $\mathcal{E}S_n$, depends on n. If $\sqrt{n}(\mathcal{E}S_n - \mu(J,F_X)) \to 0$ then can we replace $\mathcal{E}S_n$ by $\mu(J,F_X)$ in (5.3.5). (ii) Bhattacharya's result deals with the convergence of the process $\sqrt{n}(H_n(t)-H(t))$ and hence gives the asymptotic distribution of any finite dimensional law from this process. (iii) Bhattacharya decomposes the limiting process into two independent normal components which is not presented in Yang's results.

Under some additional assumptions we extend Yang's (1979) result as given by (5.3.5) to include a fixed centering constant. Assume that

Y3': m(x) is a continuous monotonic function of x.

Y5. J satisfies a Lipschitz condition of order $\alpha > 1/2$ except perhaps at a finite number of continuity points of $m(F_X^{-1})$.

Y6. $\int [F_m(x)(1-F_m(x))]^{1/2} dx < \infty$ where F_m is the df of m(X).

WLOG we take m to be monotonically increasing. Then

$$\mathcal{E}S_n = \frac{1}{n} \sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) \mathcal{E} m(X_{i:n})$$

where $m(X_{i:n})$ is the i^{th} order statistic from the distribution of m(X). Conditions Y4, Y5 and Y6, in view of Mason (1979) imply that

$$\sqrt{n}(\mathcal{E}S_n - \mu(J,F_X)) \to 0$$

as $n \to \infty$. Hence, under Y1, Y2, Y3', Y4-Y6,

$$\sqrt{n}(S_n - \mu(J,F_X)) \xrightarrow{\mathcal{L}} N(0, \sigma^2(J,F_X)).$$

For the induced selection differential,

$$J(u) = \begin{cases} 1, & u \geq 1-p \\ 0, & u < 1-p \end{cases}.$$

Hence, Y5 is satisfied and Y3' implies Y4 here. Also, $\sqrt{n}(D_{[k,n]}/p - S_n)$ can be shown to tend to zero in probability as $n \to \infty$. Combining all these we have the following result.

Theorem 5.3.2:

Under the assumptions Y1, Y2, Y3', Y6, the asymptotic distribution of $D_{[k,n]}$ is given by (5.3.4).

5.4. Asymptotic Joint Distribution of $D_{[k,n]}$ and $D_{k,n}$ in the Quantile Case

Using Bhattacharya's (1976) methods we now obtain the limiting distribution of the bivariate random variable with the induced selection differential and the selection differential as its components. To start with, we assume the following in addition to assumptions B1-B4 of the previous section:

B5. $\xi_X(t) = F_X^{-1}(t)$ is continuous and σ_X^2 is finite.

Following Bhattacharya's notation we define

$$U_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} (Y_{[i:n]} - m(X_{i:n})) \text{ and } V_n(t) = \sqrt{n}(G_n(t)-t),$$

where $G_n = F_n(F^{-1})$ with F_n being the empirical df of the X_i's.

It then follows from Bhattacharya (1976, p. 622) that

$$(U_n, V_n) \Rightarrow (U,V) \equiv (\zeta(\psi), \eta) \tag{5.4.1}$$

where $\psi(t) = \int_{-\infty}^{\xi_X(t)} \sigma^2(x)\,dF_X(x)$, ζ is a standard Brownian motion and η is a Brownian bridge independent of ζ.

Let

$$H_n(t) = \frac{1}{n}\sum_{i=1}^{[nt]} Y_{[i:n]}, \quad H(t) = \int_0^t h(s)\,ds,$$

$$K_n(t) = \frac{1}{n}\sum_{i=1}^{[nt]} X_{i:n}, \quad K(t) = \int_0^t \xi_X(s)\,ds$$

and define

$$C_n(t) = \sqrt{n}(H_n(t)-H(t)) \quad \text{and} \quad D_n(t) = \sqrt{n}(K_n(t)-K(t)).$$

It can be shown that (see Bhattacharya, 1976, p. 621),

$$\begin{aligned} C_n(t) &= U_n(t) - \int_0^t V_n(s)\,dh(s) + R_{1n}(t) \\ D_n(t) &= -\int_0^t V_n(s)\,d\xi_X(s) + R_{2n}(t) \end{aligned} \tag{5.4.2}$$

where

$$\sup_{a \le t \le b} |R_{1n}(t)| \xrightarrow{P} 0 \quad \text{and} \quad \sup_{a \le t \le b} |R_{2n}(t)| \xrightarrow{P} 0$$

for $[a,b] \subset (0,1)$.

Theorem 5.4.1:

Under the conditions B1-B5,

$$(C_n(t),\ D_n(t)) \Rightarrow (C(t),\ D(t))$$
$$\equiv (\zeta(\psi(t)) + \int_0^t \eta(s)dh(s),$$
$$\int_0^t \eta(s)d\xi_X(s)),$$

where $t\varepsilon[a,b] \subset (0,1)$, and ζ, η and $\psi(t)$ are as described above.

Proof:

$(u(t) - \int_0^t v(s)dh(s),\ - \int_0^t v(s)d\xi_X(s))$ is a continuous function of u and v. Hence, recalling (5.4.1) and (5.4.2) it follows that

$$(C_n(t),\ D_n(t)) \Rightarrow (\zeta(\psi(t)) - \int_0^t \eta(s)dh(s),$$
$$- \int_0^t \eta(s)d\xi_X(s)).$$

Now note that $-\eta(s) \overset{d}{=} \eta(s)$ to conclude the proof of the theorem.

Theorem 5.4.2:

Whenever Theorem 5.4.1 holds,

$$\mathrm{Cov}(C(t),D(t)) = [\xi_X(t)-t\xi_X(t)+K(t)][th(t)-H(t)]$$
$$- \int_0^t K(u)dh(u).$$

Proof:

Since ζ and η are independent,

$$\begin{aligned}
\mathrm{Cov}(C(t),\, D(t)) &= \mathrm{Cov}\left(\int_0^t \eta(s)\,dh(s),\right. \\
&\qquad \left.\int_0^t \eta(s)\,d\xi_X(s)\right) \\
&= \int_{u=0}^t \int_{v=0}^u v(1-u)\,dh(u)\,d\xi_X(v) \\
&\quad + \int_{u=0}^t \int_{v=u}^t u(1-v)\,dh(u)\,d\xi_X(v) \\
&= \int_{u=0}^t (1-u)\,[u\xi_X(u)-K(u)]\,dh(u) \\
&\quad + \int_{u=0}^t u[(\xi_X(t)-\xi_X(u)) \\
&\quad - (t\xi_X(t)-u\xi_X(u)) + K(t)-K(u)]\,dh(u)
\end{aligned}$$

on integration by parts of one of the integrals. Hence,

$$\begin{aligned}
\mathrm{Cov}(C(t),D(t)) &= \int_0^t \xi_X(u)\,[u-u^2-u+u^2]\,dh(u) \\
&\quad + \int_0^t K(u)\,[-1+u-u]\,dh(u) \\
&\quad + \int_0^t u\,[\xi_X(t)-t\xi_X(t)+K(t)]\,dh(u) \\
&= [\xi_X(t)-t\xi_X(t)+K(t)]\int_0^t u\,dh(u) \\
&\quad - \int_0^t K(u)\,dh(u).
\end{aligned}$$

But $\int_0^t u\,dh(u) = th(t)-H(t)$ and hence the proof is over.

Remarks:

Both C(t) and D(t) are normal. Noting that $a_1C(t) + a_2D(t)$ is univariate normal for all real a_1 and a_2, we conclude that (C(t), D(t)) is a bivariate normal rv. Hence, C(t) and D(t) are independent iff they are uncorrelated. In view of Theorem 5.4.2, this is true iff

$$[\xi_X(t)-t\xi_X(t)+K(t)][th(t)-H(t)] - \int_0^t K(u)\,dh(u) = 0.$$

The natural question is whether this is possible at all. The following example shows that the answer is in the affirmative.

Example 5.4.1:

Let $X \sim \mathcal{U}(0,1)$ so that $\xi_X(u) = u$ and let $Y|X = s \sim N(s-3s^2,1)$. Then, $m(s) = s-3s^2 = m(\xi_X(s)) = h(s)$. Conditions B1-B5 are satisfied. $H(t) = \int_0^t h(s)\,ds = t^2/2 - t^3$, $K(t) = \int_0^t x\,dx = t^2/2$ and $\int_0^t K(u)\,dh(u) = (2t^3-9t^4)/12$. Hence,

$$\begin{aligned}\mathrm{Cov}(C(t),D(t)) &= (t-t^2+t^2/2)(t^2-3t^3-t^2/2+t^3)\\ &\quad - (2t^3-9t^4)/12 = 0\\ &\Rightarrow t^3(6t^2-9t+2) = 0.\end{aligned}$$

$t = (9-\sqrt{33})/12 \doteq 0.2719$ is the only solution of this equation in (0,1). Hence, for this value of t $C_n(t)$ and

$D_n(t)$ are asymptotically independent.

Now we assume that $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$, and find the asymptotic joint distribution of $D_{[k,n]}$ and $D_{k,n}$ when $k = [np]$, $0<p<1$, after appropriate normalization. For this we define the following:

$$H_n^*(t) = \frac{1}{n} \sum_{i=n-[nt]+1}^{n} Y_{[i:n]}, \quad H^*(t) = \int_{1-t}^{1} h(s)ds,$$

$$K_n^*(t) = \frac{1}{n} \sum_{i=n-[nt]+1}^{n} X_{i:n}, \quad K^*(t) = \int_{1-t}^{1} \xi_X(s)ds = t\mu_X(t).$$

Then, under the conditions B1-B5, exactly on the lines of Theorem 5.4.1 one can show that

$$(\sqrt{n}(H_n^*(p)-H^*(p)),\ \sqrt{n}(K_n^*(p)-p\mu_X(p))) \xrightarrow{\mathcal{L}}$$

$$(\zeta(\psi^*(p)) - \int_{1-p}^{1} \eta(s)dh(s),\ \int_{1-p}^{1} \eta(s)d\xi_X(s)).$$

where

$$\psi^*(p) = \int_{\xi_X(1-p)}^{\infty} \sigma^2(x)dF_X(x).$$

Also, one can show that

$$\sqrt{k}(D_{[k,n]}-H^*(p)/p) - \sqrt{n/p}.(H_n^*(p)-H^*(p)) \xrightarrow{P} 0$$

and

$$\sqrt{k}(D_{k,n}-\mu_X(p)) - \sqrt{n/p}\cdot(K_n^*(p) - p\mu_X(p)) \xrightarrow{P} 0.$$

This proves the following result.

Theorem 5.4.3:

Under the assumptions B1-B5, when $k = [np]$, $0<p<1$

$$(\sqrt{k}(D_{[k,n]}-H^*(p)/p),\ \sqrt{k}(D_{k,n}-\mu_X(p)) \stackrel{\mathcal{L}}{\to}$$

$$(\frac{1}{\sqrt{p}}[\zeta(\psi^*(p)) - \int_{1-p}^{1} \eta(s)dh(s)],$$

$$(\frac{1}{\sqrt{p}} \int_{1-p}^{1} \eta(s)d\xi_X(s)).$$

Hence, the asymptotic distribution is bivariate normal. Also, in view of Example 5.4.1, it is possible to have asymptotic independence of $D_{k,n}$ and $D_{[k,n]}$ for some p, even though X_i and Y_i are not independent.

5.5. Linear Regression Model

Suppose $Y = \alpha + \beta X + E$ where X and E are mutually independent rvs with finite variance and $\mu_E = 0$. Let (X_i, Y_i), $i = 1$ to n, be a random sample from this simple linear regression model. Then, it is known that $\alpha + E_{[i]} = Y_{[i:n]} - \beta X_{i:n}$, $i = 1$ to n, are iid rvs independent of $(X_{1:n}, X_{2:n}, \ldots, X_{n:n})$. Therefore,

$$D_{[k,n]} - \beta \frac{\sigma_X}{\sigma_Y} D_{k,n} \stackrel{d}{=} \frac{\bar{E}_k}{\sigma_Y}, \qquad (5.5.1)$$

where $\bar{E}_k$ is the average of k iid rvs each having the same distribution as E and is independent of $(X_{1:n}, \ldots, X_{n:n})$. Note that the coefficient of correlation between X and Y $\equiv \rho = \beta\sigma_X/\sigma_Y$. We will find the limit distribution of $D_{[k,n]}$

under this model in various cases. The quantile case leads to the asymptotic distribution of selection to response, a quantity used in breeding problems.

Quantile case: $k = [np]$, $0<p<1$:

Using (5.5.1) and CLT, we have

$$A_k = \sqrt{k}(D_{[k,n]}-\rho D_{k,n}) = \sqrt{k}\,\bar{E}_k/\sigma_Y \xrightarrow{\mathcal{L}} N(0,\ \sigma_E^2/\sigma_Y^2)$$

as $k\to\infty$. Assuming that $\xi_X(q)$, the q^{th} quantile of the distribution of X, is unique, it follows from Theorem 3.4.3 that

$$B_k = \sqrt{k}(D_{k,n}-(\mu_X(p)-\mu_X)/\sigma_X) \xrightarrow{\mathcal{L}} N(0,\ (\sigma_X^2(p)+q(\mu_X(p) - \xi_X(q))^2)/\sigma_X^2),$$

where $\mu_X(p)$ and $\sigma_X^2(p)$ are the conditional mean and variance of the distribution of X when truncated below at $\xi_X(q)$. Since A_k and B_k are independent (this is because $\bar{E}_k$ and $X_{i:n}$'s are independent), we conclude that

$$\sqrt{k}(D_{[k,n]}-\rho(\mu_X(p)-\mu_X)/\sigma_X) \overset{d}{=} A_k + \rho B_k$$

$$\xrightarrow{\mathcal{L}} N(0,\ [\sigma_E^2 + \beta^2(\sigma_X^2(p) + q(\mu_X(p)- \xi_X(q))^2)]/\sigma_Y^2). \quad (5.5.2)$$

The results of Section 5.3 can also be used to obtain this result after imposing some additional conditions. Even if $\xi_X(q)$ is not unique, the limit distribution of $\sqrt{k}(D_{[k,n]}-\rho(\mu_X(p)-\mu_X)/\sigma_X)$ exists, but will not be normal. This is because the limit distribution of B_k is not normal when $\xi_X(q)$ is not unique (see Theorem 3.4.3).

In a genetic context the term $D_{[k,n]}$ is often called average selection response (Burrows, 1975) or response to selection (Falconer, 1960) where we assume that $\mu_Y = \mu_X$ and $\sigma_Y = 1$. Hence (5.5.2) shows that if the top p fraction of the parents is selected from an infinite population under the commonly used linear regression set-up, the average selection response, appropriately normalized, is normal. This can be used to make inferences about improvement due to selection.

Also,

$$(\sqrt{k}(D_{[k,n]}-\rho(\mu_X(p)-\mu_X)/\sigma_X),\ \sqrt{k}(D_{k,n}-(\mu_X(p)-\mu_X)/\sigma_X))$$

$$\stackrel{d}{=} (A_k+\rho B_k, B_k)$$

$$\stackrel{\mathfrak{L}}{\to} (A+\rho B, B)$$

where A and B are independent and $A \sim N(0,\ \sigma_E^2/\sigma_Y^2)$, and $B \sim N(0,\ [\sigma_X^2(p)+q(\mu_X(p)-\xi_X(q))^2]/\sigma_X^2)$. Hence, the limiting covariance is $\rho \text{Var}(B)$. Consequently, $D_{[k,n]}$ and $D_{k,n}$, appropriately normalized are asymptotically independent iff $\rho = 0$.

<u>Extreme case: k fixed:</u>

If $x_o = F_X^{-1}(1) < \infty$, then $D_{k,n} \stackrel{P}{\to} (x_o-\mu_X)/\sigma_X$ and hence from (5.5.1) it follows that

$$D_{[k,n]} \stackrel{\mathfrak{L}}{\to} [\bar{E}_k + \beta(x_o-\mu_X)]/\sigma_Y.$$

If $x_o = +\infty$ then the situation is more involved and is analyzed further below:

(i) If $D_{k,n}$ has AWL, i.e., there exists a sequence of constants c_n such taht $D_{k,n}-c_n \xrightarrow{P} 0$, then

$$D_{[k,n]}-\rho c_n \overset{d}{=} \rho(D_{k,n}-c_n) + \bar{E}_k/\sigma_Y \xrightarrow{\mathcal{L}} \bar{E}_k/\sigma_Y . \qquad (5.5.3)$$

(ii) If $F_X \varepsilon D(\Phi_\alpha)$ then $F^n(a_n+b_nx) \to \Phi_\alpha(x)$ where a_n and b_n can be taken to be 0 and $\xi_X(1-1/n)$, respectively. Hence, from (5.5.1),

$$D_{[k,n]}/b_n \overset{d}{=} \rho D_{k,n}/b_n + \bar{E}_k/\sigma_Y b_n \xrightarrow{\mathcal{L}} \rho D_k$$

since $b_n \to \infty$. Here D_k has the representation given by (3.2.4a).

(iii) If $F_X \varepsilon D(\Lambda)$ we have to examine further.

a. If $b_n \to 0$, then $D_{k,n}$ has AWL and hence (5.5.3) holds where c_n can be taken to be a_n.

b. If $b_n \to \infty$, then $(D_{[k,n]}-\rho a_n)/b_n \xrightarrow{\mathcal{L}} \rho D_k$.

c. If $b_n \to b \neq 0$, then $(D_{[k,n]}-\rho a_n) \xrightarrow{\mathcal{L}} b(\rho D_k+\bar{E}_k/\sigma_Y)$.

In both b and c, the df of D_k is given by (3.2.5).

VI. MISCELLANEOUS RESULTS

In this final chapter we consider two problems which came up while pursuing the asymptotic theory of the selection differential, but were not connected directly with the selection differential. In Section 3.2, it was seen that if $F \varepsilon D(G)$ where G can be ϕ_α, ψ_α or Λ, then

$$((X_{n:n}-a_n)/b_n, \ldots, (X_{n-k+1:n}-a_n)/b_n) \overset{\mathcal{L}}{\to} (T_1, \ldots, T_k)$$

where the df of T_i was given by (3.2.2a-c). We consider the joint distribution of $T_1, T_2, \ldots, T_k$, called the k-dimensional extremal distribution and connect it to record value theory. This is done in Section 6.1 and can be used to give new proofs of some of the results of Hall (1978). Section 6.2 deals with the bivariate extension of Stigler's (1974) result (Theorem 3.4.1) for linear functions of order statistics. Two applications of this extension in finding the asymptotic distribution of Hogg's Q statistic and the asymptotic distribution of a quick estimator of the regression coefficient in a simple linear regression model are also given.

6.1. Extremal Distributions and Lower Record Values

Dwass (1966) defines a k-dimensional extremal distribution as follows: A random vector $(Y_1,\ldots,Y_k)$ is said to have a k-dimensional extremal distribution with parameter G (a df) if

a. $G^{-1}(0) \le Y_k \le \cdots \le Y_1 \le G^{-1}(1)$ with probability 1 and

b. if $G^{-1}(0) \le v_k < u_k < v_{k-1} < u_{k-1} < \cdots < v_1 < u_1$

$\le G^{-1}(1)$, then

$$P\left(\bigcap_{i=1}^{k} [v_i < Y_i \le u_i]\right) = [G(u_k) - G(v_k)] \prod_{i=1}^{k-1} (-\log(G(v_i)/G(u_i))) \quad (6.1.1)$$

where $\prod_{i=1}^{0} \equiv 1$. Further, it follows from Lamperti (1964) that if $F^n(a_n+b_nx) \to G(x)$, a nondegenerate df, i.e., $F \in D(G)$, then

$$\left(\frac{X_{n:n}-a_n}{b_n}, \ldots, \frac{X_{n-k+1:n}-a_n}{b_n}\right) \xrightarrow{\mathcal{L}} (T_1, T_2, \ldots, T_k)$$

where T_i's replace Y_i's above and G is one of Φ_α, ψ_α or Λ.

Now, suppose G is any absolutely continuous df with pdf g and $(Y_1,\ldots,Y_k)$ satisfies (6.1.1). Then the joint pdf of $(Y_1,Y_2,\ldots,Y_k)$ is given by

$$g(y_1,y_2,\ldots,y_k) = \lim_{\substack{h_i\to 0+\\ i=1 \text{ to } k}} \frac{P(\bigcap_{i=1}^{k} [y_i<Y_i\le y_i+h_i])}{h_i}$$

$$= \lim_{h_k\to 0+} \frac{G(y_k+h_k)-G(y_k)}{h_k} \prod_{i=1}^{k-1} \lim_{h_i\to 0+} \frac{\log G(y_i+h_i)-\log G(y_i)}{h_i}$$

$$= g(y_k) \prod_{i=1}^{k-1} \frac{d[\log G(y_i)]}{dy_i}, \quad y_1<y_2<\ldots<y_k$$

$$= \begin{cases} g(y_k) \prod_{i=1}^{k-1} \dfrac{g(y_i)}{G(y_i)}, & y_1<y_2<\ldots<y_k \\ 0, & \text{otherwise.} \end{cases}$$

But, from record value theory, it is immediate that this is the joint pdf of the first k lower record values from the df G (see, e.g., Chandler, 1952).

We exploit this relationship between extremal distributions and lower record values from one dimensional extremal distributions to:

(i) give a different canonical representation of the T_i's in the three cases and to

(ii) reprove the limit laws of Hall (1978) for T_k, using record value theory.

Let the Y_i's be the upper record values from Exp(1) distribution. Then, it is known (see e.g., Resnick (1973, p. 69) that $Y_i \overset{d}{=} \sum_{j=1}^{i} Z_j$, where Z_j's are iid Exp(1) rvs. Now, let T_i be lower record values from a continuous df G. Then $G(T_i)$ form lower record values from $\mathcal{U}(0,1)$ and consequently, $-\log G(T_i)$ are upper record values from Exp(1) distribution.

That is, $-\log G(T_i) \overset{d}{=} Y_i \overset{d}{=} \sum_{j=1}^{i} Z_j$ and hence,

$$T_i = G^{-1}(\exp(-\sum_{j=1}^{i} Z_j)).$$

Therefore,

$$T_i \overset{d}{=} (\sum_{j=1}^{i} Z_j)^{-1/\alpha} \quad \text{if} \quad G = \Phi_\alpha \tag{6.1.2a}$$

$$\overset{d}{=} -(\sum_{j=1}^{i} Z_j)^{1/\alpha} \quad \text{if} \quad G = \Psi_\alpha \tag{6.1.2b}$$

$$\overset{d}{=} -\log(\sum_{j=1}^{i} Z_j) \quad \text{if} \quad G = \Lambda \tag{6.1.2c}$$

These representations involve only a finite number of exponential rvs whereas, Hall's (1978) representations, given by (3.2.3a-c), consist of an infinite number of exponential rvs. The above representations have also been obtained by Weissman using a Poisson process approach (personal communication).

Using (6.1.2a-c) we study the asymptotic behavior of T_i. It is evident that for this purpose we have to study the behavior of $S_i = \sum_{j=1}^{i} Z_j$. But, from the classical limit theory it is easy to see that, for an Exp(1) parent,

$$\frac{S_k - k}{\sqrt{k}} \xrightarrow{\mathcal{L}} N(0,1) \qquad \text{(CLT)}$$

$$\frac{S_k}{k} \xrightarrow{a.s.} 1 \qquad \text{(SLLN)}$$

and

$$\left.\begin{aligned} \limsup_{k\to\infty} \frac{S_k - k}{\sqrt{2k \log \log k}} &= 1 \text{ a.s.} \\ \liminf_{k\to\infty} \frac{S_k - k}{\sqrt{2k \log \log k}} &= -1 \text{ a.s.} \end{aligned}\right\} \qquad \text{(LIL)}$$

We can now take G to be one of the three extreme value distributions. To fix the ideas, we take $G = \Lambda$. From (6.1.2c) it follows that

$T_k = -\log S_k$ and hence

$$T_k + \log k = -\log(S_k/k) \xrightarrow{a.s.} -\log 1 = 0 \quad \text{as } k\to\infty. \qquad (6.1.3)$$

Now to prove CLT for T_k, we recall the following result (Rao, 1973, p. 385): If

$\sqrt{k}(U_k - \theta) \xrightarrow{\mathcal{L}} N(0, \sigma^2(\theta))$ and g is a differentiable function, then

$$\sqrt{k}(g(U_k) - g(\theta)) \xrightarrow{\mathcal{L}} N(0, \sigma^2(\theta)[g'(\theta)]^2), \text{ as } k\to\infty.$$

Take $U_k = S_k/k, \theta = 1, \sigma^2(\theta) = 1, g(x) = -\log x$ so that

$$\sqrt{k}(T_k + \log k) = \sqrt{k}(-\log(S_k/k) - 0) \xrightarrow{\mathcal{L}} N(0,1). \qquad (6.1.4)$$

We can also prove LIL for T_k using elementary analytical methods by exploring the concept of limit superior and limit inferior and the relation between T_k and S_k. We do not present this long but conceptually simple derivation here. It turns out that

$$\lim_{k\to\infty} \sup \sqrt{\frac{k}{2 \log \log k}}(T_k + \log k) = +1 \text{ a.s.}$$

$$\lim_{k\to\infty} \inf \sqrt{\frac{k}{2 \log \log k}}(T_k + \log k) = -1 \text{ a.s.} \qquad (6.1.5)$$

(6.1.3)-(6.1.5) have been obtained by Hall (1978) using a different canonical representation for T_k's as given by (3.2.3a-c) and some martingale convergence theorems. One can also use the general asymptotic theory for record values to obtain these results.

6.2. Bivariate Extension of Stigler's (1974) Result with Applications

As in Section 3.4, let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the order statistics of a random sample of size n from a distribution with df F with finite variance. Put

$$S_{1n} = \frac{1}{n} \sum_{i=1}^{n} J_1\left(\frac{i}{n+1}\right) X_{i:n}, \quad S_{2n} = \frac{1}{n} \sum_{i=1}^{n} J_2\left(\frac{i}{n+1}\right) X_{i:n}$$

where J_1 and J_2 are bounded and continuous a.e. F^{-1}. Let,

$$\sigma^2(J_i, F) = \iint J_i(F(x)) J_i(F(y)) [F(\min(x,y)) - F(x)F(y)] dxdy, \quad i = 1,2,$$

be positive. Also, let

$$\sigma_{12}(J_1,J_2,F) = \iint J_1(F(x))J_2(F(y))[F(\min(x,y)) - F(x)F(y)]dxdy.$$

Theorem 6.2.1:

Under the above assumptions,

(i) $n\,\mathrm{Var}(S_{in}) \to \sigma^2(J_i,F), \quad i = 1,2$ (6.2.1)

$n\,\mathrm{Cov}(S_{1n},S_{2n}) \to \sigma_{12}(J_1,J_2,F)$ (6.2.2)

(ii) $(\sqrt{n}(S_{1n}-\mathcal{E}S_{1n}),\ \sqrt{n}(S_{2n}-\mathcal{E}S_{2n})) \xrightarrow{\mathcal{L}} (S_1,S_2)$ (6.2.3)

where (S_1,S_2) is a bivariate normal random variable with mean vector (0,0) and covariance matrix

$$\begin{pmatrix} \sigma^2(J_1,F) & \sigma_{12}(J_1,J_2,F) \\ \sigma_{12}(J_1,J_2,F) & \sigma^2(J_2,F) \end{pmatrix}.$$

(iii) Suppose further that $\int(F(x)(1-F(x)))^{1/2}dx$ is finite and that J_1 and J_2 satisfy Lipschitz conditions with indices $\alpha_1 > 1/2$ and $\alpha_2 > 1/2$, respectively, except possibly at a finite number of points of F^{-1} measure zero. Then

$\sqrt{n}(\mathcal{E}S_{in}-\mu(J_i,F)) \to 0, \quad i = 1,2$ (6.2.4)

where $\mu(J_i,F) = \int F^{-1}(t)J_i(t)dt$. Consequently, one can replace $\mathcal{E}S_{in}$ by $\mu(J_i,F)$, $i = 1,2$ in (6.2.3).

Proof:

Since J_1 and J_2 are bounded and continuous a.e. F^{-1}, from part (i) of Theorem 3.4.1, (6.2.1) follows. To show the asymptotic bivariate normality we show that $\sqrt{n}[c_1(S_{1n}-\mathcal{E}S_{1n}) + c_2(S_{2n}-\mathcal{E}S_{2n})]$ converges in law to a univariate normal distribution for all real c_1 and c_2. For this, let $J = c_1J_1 + c_2J_2$ and $S_n = c_1S_{1n} + c_2S_{2n}$. Then, applying Theorem 3.4.1 for S_n we conclude that

$$\sqrt{n}(S_n-\mathcal{E}S_n) \overset{\mathcal{L}}{\to} N(0, \sigma^2(J,F)),$$

where

$$\sigma^2(J,F) = \iint [c_1J_1(F(x))+c_2J_2(F(x))][c_1J_1(F(y)) + c_2J_2(F(y))] \cdot [F(\min(x,y)) - F(x)F(y)]dxdy$$

$$= c_1{}^2\sigma^2(J_1,F)+c_2{}^2\sigma^2(J_2,F)+2c_1c_2\sigma^2(J_1,J_2,F). \tag{6.2.5}$$

Since c_1 and c_2 are arbitrary, (6.2.3) follows. Also, using (6.2.1), (6.2.5) and the fact that $n\ \mathrm{Var}(S_n) \to \sigma^2(J,F)$ we obtain (6.2.2). Applying part (iii) of Theorem 3.4.1, for both S_{1n} and S_{2n}, (6.2.4) follows. Therefore, one can replace $\mathcal{E}S_{in}$ by $\mu(J_i,F)$ in (6.2.3).

Corollary:

Under the above conditions, $\sqrt{n}(S_{1n}-\mu(J_1,F))$ and $\sqrt{n}(S_{2n}-\mu(J_2,F))$ are asymptotically independent iff $\sigma_{12}(J_1,J_2,F) = 0$ or iff $\lim_{n\to\infty} n\ \mathrm{Cov}(S_{1n},S_{2n}) = 0$.

Example 6.2.1:

Following Hogg (1974), define $\bar{U}_n(p) = \frac{1}{[np]} \sum_{i=n-[np]+1}^{n} X_{i:n}$ and $\bar{L}_n(p) = \frac{1}{[np]} \sum_{i=1}^{[np]} X_{i:n}$. Let $0<p<0.5$. Then, using the theorem above, one can show that $\sqrt{[np]}(\bar{U}_n(p)-\mu_p)$ and $\sqrt{[np]}(\bar{L}_n(p)-\bar{\mu}_p)$ are asymptotically bivariate normal when ξ_p and ξ_q, the p^{th} and q^{th} quantile are unique. Here, $\mu_p = \frac{1}{p}\int_{\xi_q}^{\infty} x dF(x)$ and $\bar{\mu}_p = \frac{1}{p}\int_{-\infty}^{\xi_p} x dF(x)$. Further, the limiting covariance is $(\xi_p-\bar{\mu}_p)(\mu_p-\xi_q)$, a positive quantity. Hence, $\bar{L}_n(p)$ and $\bar{U}_n(p)$ are not asymptotically independent. This is in contrast to the independence of $g_1(X_{1:n},\ldots,X_{k:n})$ and $g_2(X_{n-k+1:n},\ldots,X_{n:n})$ when $k/n\to 0$ (Rossberg, 1965, David, 1980, p. 306).

Example 6.2.2:

Let F be symmetric about zero and let the median be unique. Define

$$J_1(u) = 1,\ 0\le u\le 1;\quad J_2(u) = \begin{cases} -1, & u\le p \\ 1, & u>1-p \end{cases}.$$

Then, $S_{1n} = \bar{X}$ and $S_{2n} = \frac{1}{2}\{\bar{U}_n(p) - \bar{L}_n(p)\}$. Because of symmetry, $\mathrm{Cov}(S_{1n},S_{2n}) = 0$ for every n. Hence, $\sqrt{n}\,\bar{X}$ and

$\sqrt{n}(S_{2n}-2\mu_p)$ are both asymptotically normal and asymptotically independent from the theorem and the corollary above.

Asymptotic distribution of Hogg's Q statistic:

Hogg (1974) suggested the following statistic as a good indicator of tail length in symmetric populations:

$$Q_n = \frac{\bar{U}_n(p_1) - \bar{L}_n(p_1)}{\bar{U}_n(p_2) - \bar{L}_n(p_2)} .$$

In fact, he took $p_1 = 0.05,\ 0.2$ and $p_2 = 0.5$ in his study. We use Theorem 6.2.1 to obtain the asymptotic distribution of Q_n. Define

$$J_i(u) = \begin{cases} 1, & u>1-p_i \\ -1, & u<p_i \\ 0, & \text{otherwise} \end{cases} , \quad i = 1,2$$

$$W_n = \frac{1}{n}\Big(\sum_{i=n-k_1+1}^{n} X_{i:n} - \sum_{i=1}^{k_1} X_{i:n}\Big) \Big/ \frac{1}{n}\Big(\sum_{i=n-k_2+1}^{n} X_{i:n} - \sum_{i=1}^{k_2} X_{i:n}\Big)$$

where $k_i = [np_i]$

$$= S_{1n}/S_{2n} \quad \text{where} \quad S_{1n} = \Sigma J_1\Big(\frac{i}{n+1}\Big) X_{i:n},$$

$$S_{2n} = \Sigma J_2\Big(\frac{i}{n+1}\Big) X_{i:n};$$

$$Q_n = [np_2] W_n/[np_1].$$

Under the assumptions of Theorem 6.2.1, it follows that $(\sqrt{n}(S_{1n}-\mu(J_1,F)),\ \sqrt{n}(S_{2n}-\mu(J_2,F)))$ converges in law to a bivariate normal distribution. Hence

$$\sqrt{n}\left(\frac{S_{1n}}{S_{2n}}-\frac{\mu(J_1,F)}{\mu(J_2,F)}\right)=\frac{\sqrt{n}(S_{1n}-\mu(J_1,F))\mu(J_2,F)-\sqrt{n}\mu(J_1,F)(S_{2n}-\mu(J_2,F))}{S_{2n}\mu(J_2,F)}$$

being a continuous function of these components converges in law to $N(0, \sigma^2(W,F))$ where

$$\sigma^2(W,F) = \frac{\mu^2(J_2,F)\sigma^2(J_1,F)+\mu^2(J_1,F)\sigma^2(J_2,F) - 2\mu(J_1,F)\mu(J_2,F)\sigma_{12}(J_1,J_2,F)}{\mu^4(J_2,F)}.$$

Here we have also made use of the fact that $S_{2n} \xrightarrow{P} \mu(J_2,F)$. We do not need symmetry for this result to be true. However, under Hogg's assumption of symmetry some simplifications in the expression for $\sigma^2(W,F)$ is possible . WLOG we assume F is symmetric about zero and $p_1<p_2$. Then one can show, after some algebra, that

$$\mu(J_i,F) = 2p_i\mu_{p_i}, \quad i = 1,2$$

$$\sigma^2(J_i,F) = 2p_i\sigma^2_{p_i} + 2p_i(q_i-p_i)(\mu_{p_i}-\xi_{q_i})^2, \quad i = 1,2$$

and

$$\sigma_{12}(J_1,J_2,F) = \sigma^2(J_1,F) + 2p_1(\mu_{p_1}-\xi_{q_1})[(\xi_{q_1}-\xi_{q_2}) - 2p_2(\mu_{p_2}-\xi_{q_2})].$$

As usual, μ_{p_i} and $\sigma^2_{p_i}$ are the mean and variance of the df obtained by truncating F below at ξ_{q_i}, the q_i^{th} quantile point. Of course, the above have to be substituted in the expression for $\sigma^2(W,F)$. Now

$$\sqrt{n}(Q_n - \frac{\mu_{p_1}}{\mu_{p_2}}) \doteq \frac{p_2}{p_1} \sqrt{n}\ (W_n - \frac{\mu(J_1,F)}{\mu(J_2,F)}) \text{ and hence}$$

$$\xrightarrow{\mathcal{L}} N(0,\ {p_2}^2\sigma^2(W,F)/{p_1}^2). \qquad (6.2.6)$$

The assumptions which ensure (6.2.6) are apart from the symmetry that the quantiles concerned are unique and that $\int (F(x)(1-F(x)))^{1/2}dx$ is finite. De Wet and van Wyk (1979) have also considered the asymptotic distribution of Q_n. It appears that their use of Moore's (1968) result in establishing the asymptotic normality is questionable.

This problem was brought to my attention by Dr. Robert Stephenson and reference to de Wet and van Wyk (1979) was indicated by Dr. Robert Hogg.

Simple linear regression model:

As another application of Theorem 6.2.1 we obtain the asymptotic distribution of a quick estimator of the regression coefficient in a simple linear regression model. The asymptotic distribution of this estimator in the bivariate normal case has been obtained by Barton and Casley (1958). Let (X_i, Y_i), i = 1 to n be a random sample, from the simple linear regression model described in Section 5.5. Let

$\hat{\beta} = (\overline{Y}_k' - \overline{Y}_k)/(\overline{X}_k' - \overline{X}_k)$ where $k = [np]$, $0<p<1/2$ and

$$\overline{X}_k' = \frac{1}{k}\sum_{i=n-k+1}^{n} X_{i:n},\ \overline{X}_k = \frac{1}{k}\sum_{i=1}^{k} X_{i:n},$$

$$\overline{Y}_k' = \frac{1}{k}\sum_{i=n-k+1}^{n} Y_{[i:n]},\ \overline{Y}_k = \frac{1}{k}\sum_{i=1}^{k} Y_{[i:n]}.$$

We obtain the asymptotic distribution of $\hat{\beta}$. To start with, we prove some general results for the linear regression model using Theorem 6.2.1.

Under our model it is known that $\alpha + E_{[i]} = Y_{[i:n]} - \beta X_{i:n}$, $i = 1$ to n are iid rvs and are independent of $(X_{1:n},\ldots,X_{n:n})$.

Define the following linear functions of $X_{i:n}$'s and $Y_{[i:n]}$'s:

$$U_n = \frac{1}{n}\sum_{i=1}^{n} J_1(\tfrac{i}{n+1})Y_{[i:n]} = \beta\frac{1}{n}\sum_{i=1}^{n} J_1(\tfrac{i}{n+1})X_{i:n} + \frac{1}{n}\sum_{i=1}^{n} J_1(\tfrac{i}{n+1})(E_{[i]}+\alpha)$$

$$= \beta S_n + R_n \quad \text{say;}$$

$$T_n = \frac{1}{n}\sum_{i=1}^{n} J_2(\tfrac{i}{n+1})X_{i:n}$$

where J_1 and J_2 are bounded and continuous a.e. F_X^{-1}. Note that S_n and R_n are independent and S_n and T_n play the role of S_{1n} and S_{2n} in Theorem 6.2.1.

<u>Theorem 6.2.2</u>:

In addition to the above assumptions, let J_1 be integrable and $\mathcal{E}|E|^{2+\delta} < \infty$ for some $\delta>0$. Let $\sigma^2(J_i, F_X)$ be positive and finite. Then

$$(\sqrt{n}(U_n - \mathcal{E}U_n), \sqrt{n}(T_n - \mathcal{E}T_n)) \xrightarrow{\mathcal{L}} (R + \beta S, T)$$

where $R \sim N(\alpha, \sigma_E^2 \int_0^1 J_1^2(x)dx)$, (S,T) has the distribution of (S_1, S_2) of Theorem 6.2.1 with $F = F_X$. Further R and (S,T) are independent.

<u>Proof</u>:

$(\sqrt{n}(U_n - \mathcal{E}U_n), \sqrt{n}(T_n - \mathcal{E}T_n)) = (\sqrt{n}(R_n - \mathcal{E}R_n) + \beta\sqrt{n}(S_n - \mathcal{E}S_n), \sqrt{n}(T_n - \mathcal{E}T_n))$. First note that $\sqrt{n}(R_n - \mathcal{E}R_n)$ and $(\sqrt{n}(S_n - \mathcal{E}S_n), \sqrt{n}(T_n - \mathcal{E}T_n))$ are independent. The convergence of the bivariate rv follows from Theorem 6.2.1. R_n is the mean of independent nonidentically distributed rvs. $\mathcal{E}|E|^{2+\delta} < \infty$ implies $\mathcal{E}|J_1(\frac{i}{n+1})E_{[i]}|^{2+\delta} < \infty$. Using CLT (Lemma A4), and the fact that J_1 is integrable, one can show that

$$\sqrt{n}(R_n - \mathcal{E}R_n) \xrightarrow{\mathcal{L}} R \sim N(0, \sigma_E^2 \int_0^1 J_1^2(u)du).$$

This completes the proof.

One can show that if J_1 satisfies a Lipschitz condition with index $\alpha > 1/2$ except at a finite number of points, then

$\sqrt{n}(\mathcal{E}R_n - \alpha\int_0^1 J_1(u)du) \to 0$ as $n\to\infty$. Hence, if in addition to the assumptions of Theorem 6.2.2, J_1, and J_2 satisfy

Lipschitz conditions with indices bigger than 1/2 except possibly at a finite number of points of F_X^{-1} measure zero, and $\int (F_X(x)(1-F_X(x)))^{1/2}dx$ is finite, then one can replace $\mathcal{E}U_n$ and $\mathcal{E}T_n$ by $\beta\mu(J_1,F_X) + \alpha\int_0^1 J_1(u)du$ and $\mu(J_2,F_X)$, respectively.

To obtain the asymptotic distribution of $\hat{\beta}$, take

$$J_1(u) = J_2(u) = \begin{cases} -1, & u<p \\ 1, & u\geq q \\ 0, & \text{otherwise} \end{cases}$$

and assume that $\xi_X(p)$ and $\xi_X(q)$ are unique.

Then

$$S_n = T_n; \quad \hat{\beta} = \frac{U_n}{S_n}.$$

$$\sqrt{n}(U_n-\beta S_n) = \sqrt{n}\ R_n = \sqrt{n}(R_n-\alpha\int_0^1 J_1(u)du)$$

$$\xrightarrow{\mathcal{L}} N(0,\ \sigma_E^2\int_0^1 J_1^2(u)du)$$

$$= N(0,\ 2p\sigma_E^2).$$

Also, since $\sqrt{n}(S_n-\mu(J_1,F_X))$ converges in distribution, $S_n \xrightarrow{P} \mu(J_1,F_X) = p(\mu_p-\bar{\mu}_p)$. Hence

$$\sqrt{n}\left(\frac{U_n-\beta S_n}{S_n}\right) = \sqrt{n}(\hat{\beta}-\beta) \xrightarrow{\mathcal{L}} N\left(0,\frac{2p\sigma_E^2}{p^2(\mu_p-\bar{\mu}_p)^2}\right);$$

that is,

$$\sqrt{k}(\hat{\beta}-\beta) \xrightarrow{\mathcal{L}} N\left(0,\ \frac{2\sigma_E^2}{(\mu_p-\bar{\mu}_p)^2}\right).$$

This proof has not made any explicit use of Theorem 6.2.2. But, one can write a proof using that theorem on the lines similar to those used in obtaining the asymptotic distribution of Hogg's Q statistic.

VII. BIBLIOGRAPHY

Arnold, B. C. and R. A. Groeneveld. 1979. Bounds on expectations of linear systematic statistics based on dependent samples. Ann. Statist. 7:220-223.

Barnett, V. and T. Lewis. 1978. Outliers in statistical data. John Wiley and Sons Ltd., Chichester. 365 pp.

Barton, D. E. and D. J. Casley. 1958. A quick estimate of the regression coefficient. Biometrika 45:431-435.

Berman, S. M. 1962. Limiting distribution of the studentized largest observation. Skand. Akt. 45:154-161.

Bhattacharya, P. K. 1976. An invariance principle in regression analysis. Ann Statist. 4:621-624.

Billingsley, P. 1968. Convergence of probability measures. John Wiley and Sons, Inc., New York, N.Y. 253 pp.

Billingsley, P. 1979. Probability and measure. John Wiley and Sons, Inc., New York, N.Y. 515 pp.

Boos, D. D. 1979. A differential for L-statistics. Ann. Statist. 7:955-959.

Burrows, P. M. 1972. Expected selection differentials for directional selection. Biometrics 28:1091-1100.

Burrows, P. M. 1975. Variances of selection differentials in normal samples. Biometrics 31:125-133.

Chandler, K. N. 1952. The distribution and frequency of record values. J. R. Statist. Soc. B 14:220-228.

Chung, K. L. 1974. A course in probability theory. 2nd ed. Academic Press, New York, N.Y. 365 pp.

Cramér, H. 1946. Mathematical methods of statistics. Princeton University Press, Princeton, N.J. 575 pp.

David, H. A. 1970. Order statistics. John Wiley and Sons, Inc., New York, N.Y. 272 pp.

David, H. A. 1980. Order statistics. 2nd ed. (to appear).

David, H. A. and V. S. Shu. 1978. Robustness of location estimators in the presence of an outlier. Pages 235-250 in H. A. David, ed. Contributions to survey sampling and applied statistics: Papers in honor of H. O. Hartley. Academic Press, New York, N.Y.

de Haan, L. 1970. On regular variation and its application to the weak convergence of sample extremes. Mathematical Centre Tracts 32. Mathematical Centrum, Amsterdam. 124 pp.

de Wet, T. and J. W. J. van Wyk. 1979. Some large sample properties of Hogg's adaptive trimmed means. S. Afric. Statist. J. 13:53-69.

Dwass, M. 1966. Extremal processes II. Ill. J. Math. 10: 381-391.

Falconer, D. S. 1960. Introduction to quantitative genetics. The Ronald Press Company, New York, N.Y. 365 pp.

Feller, W. 1966. An introduction to probability theory and its applications. Vol. II. John Wiley and Sons, Inc., New York, N.Y. 626 pp.

Galambos, J. 1978. The asymptotic theory of extreme order statistics. John Wiley and Sons, Inc., New York, N.Y. 352 pp.

Ghosh, J. K. 1971. A new proof of the Bahadur representation of quantiles and an application. Ann. Math. Statist. 42:1957-1961.

Gnedenko, B. 1943. Sur la distribution limite du terme maximum d'une série aléatoire. Ann. Math. 44:423-453.

Hall, P. 1978. Representations and limit theorems for extreme value distributions. J. Appl. Probab. 15: 639-644.

Hall, P. 1979. On the rate of convergence of normal extremes. J. Appl. Probab. 16:433-439.

Hoeffding, W. 1948. A class of statistics with asymptotically normal distribution. Ann. Math. Statist. 19:293-325.

Hogg, R. V. 1974. Adaptive robust procedures: a partial review and some suggestions for future applications and theory. J. Amer. Statist. Assn. 69:909-923.

Lamperti, J. 1964. On extreme order statistics. Ann. Math. Statist. 35:1726-1737.

Leadbetter, M. R. 1978. Extreme value theory under weak mixing conditions. Pages 46-110 in M. Rosenblatt, ed. Studies in probability theory. The Mathematical Association of America.

McMillan, R. G. 1971. Tests for one or two outliers in normal samples with unknown variance. Technometrics 13: 87-100.

Mallows, C. L. and D. Richter. 1969. Inequalities of Chebychev type involving conditional expectations. Ann. Math. Statist. 40:1922-1932.

Mason, D. M. 1979. On centering linear combinations of order statistics. Department of Statistics, University of Kentucky, Tech. Report No. 148.

Moore, D. S. 1968. An elementary proof of asymptotic normality of linear functions of order statistics. Ann. Math. Statist. 39:263-265.

Murphy, R. B. 1951. On tests for outlying observations. Ph.D. Dissertation. Princeton University. 107 pp. University Microfilms, Ann Arbor, Mich. (Dissertation Abstracts 15, No. 3:425)

Nagaraja, H. N. 1978. A note on almost sure convergence of sums of order statistics. J. Indian Statist. Assn. 16:131-134.

Rao, C. R. 1973. Linear statistical inference and its applications. 2nd ed. John Wiley and Sons, Inc., New York, N.Y. 625 pp.

Resnick, S. I. 1973. Limit laws for record values. Stochastic Processes Appl. 1:67-82.

Rossberg, H. J. 1965. Die asymptotische Unabhängigkeit der kleinsten und größten Werte einer Stichprobe vom Stichprobenmittel. Math. Nachr. 28:305-318.

Smirnov, N. V. 1952. Limit distributions for the terms of a variational series. Amer. Math. Soc. Transl. Ser. 1, No. 67. 64 pp. (Original in Russian published in 1949).

Stigler, S. M. 1973. The asymptotic distribution of trimmed mean. Ann. Statist. 1:472-477.

Stigler, S. M. 1974. Linear functions of order statistics with smooth weight functions. Ann. Statist. 2:676-693. (Correction) Ann. Statist. 7:466.

Teichroew, D. 1956. Tables of expected values of order statistics and products of order statistics for samples of size 20 and less from the normal distribution. Ann. Math. Statist. 27:410-426.

van Zwet, W. R. 1964. Convex transformations of random variables. Mathmeatical Centre Tracts 7. Mathmematisch Centrum, Amsterdam. 116 pp.

Weissman, I. 1975. Multivariate extremal processes generated by independent nonidentically distributed random variables. J. Appl. Probab. 12:477-487.

Welsch, R. E. 1973. A convergence theorem for extreme values from Gaussian sequences. Ann. Probability 1: 398-404.

Yang, S. S. 1977. General distribution theory of the concomitants of order statistics. Ann. Statist. 5: 996-1002.

Yang, S. S. 1979. Linear function of concomitants of order statistics with application to nonparametric estimation of regression function. Preprint. Kansas State University, Manhattan, Kansas.

VIII. ACKNOWLEDGMENTS

I wish to express my sincere appreciation to Professor Herbert A. David for: (i) admitting me to this school, (ii) awarding me an assistantship and, (iii) supervising my dissertation. This work grew out of problems posed by him. I thank him for his very practical advice coupled with profound insight and for the inspiration and support he has given me during the period of my stay here.

Dr. Barry C. Arnold and Dr. Glen D. Meeden have contributed a lot to my life both at professional and personal levels. I thank them for being helpful and encouraging at all times.

I would not have been here except for the support and help of Dr. N. R. Mohan during the period of my stay at Mysore, India. I am greatly indebted to him.

I thank all the members of my family especially my elders for providing exceptional love and support for all these years which have been keys to my progress.

Mrs. Pat Gunnells has done an excellent job of typing the manuscript. Special thanks are due for her remarkable speed.

This research was supported by a grant from the U.S. Army Research Office.

IX. APPENDIX

Lemma A1: For a random sample of n from a continuous parent, the conditional distribution of $X_{s:n}$ given $X_{r:n} = x$ (s>r), is just the distribution of the $(s-r)^{th}$ order statistic in a random sample of (n-r) drawn from the parent distribution truncated on the left at x.

Proof: See David (1980, p. 20).

Lemma A2: Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the order statistics in a random sample of size n from the Exp(1) distribution. Then $Y_r = (n-r+1)(X_{r:n}-X_{r-1:n})$, r = 1,2,...,n with $X_0 \equiv 0$, are iid Exp(1). Consequently, $X_{r:n} \stackrel{d}{=} \sum_{i=1}^{r} Y_i/(n-i+1)$, r = 1,2,...,n, where the Y_i's are iid Exp(1) rvs. Here $\stackrel{d}{=}$ stands for the identical distribution of the rvs on either side of this symbol.

Proof: See David (1980, pp. 20-21).

Lemma A3: Let $0<f(\xi_p)<\infty$, $0<p<1$, where ξ_p is the p^{th} quantile and f is the parent pdf. If $p_n - p = O(1/\sqrt{n})$ then

$$X_{[np_n]:n} = \xi_p + [p-F_n(\xi_p)]/f(\xi_p) + R_n$$

where $F_n(\xi_p)$ is the empirical df of $X_1, X_2, \ldots, X_n$ evaluated at ξ_p and where $\sqrt{n}\, R_n \xrightarrow{p} 0$ as $n \to \infty$. (this is a weaker version of Bahadur's representation).

Consequently, $\sqrt{n}(X_{[np_n]:n} - \xi_p) \xrightarrow{\mathcal{L}} N(0, p(1-p)/f^2(\xi_p))$. This result is due to J. K. Ghosh (1971).

Lemma A4 (CLT): Suppose that for each n, the sequence of rvs $X_{n1},\ldots,X_{nr_n}$ is independent. Let $\mathcal{E}X_{nk} = 0$, $\sigma_{nk}^2 = \mathrm{Var}(X_{nk})$, $s_n^2 = \sum_{k=1}^{r_n} \sigma_{nk}^2$. Suppose that $\mathcal{E}|X_{nk}|^{2+\delta}$ exists for some $\delta>0$ and that Lyapounov's condition

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} \mathcal{E}(|X_{nk}|^{2+\delta}) = 0$$

is satisfied. Then,

$$\sum_{i=1}^{r_n} X_{ni}/s_n \xrightarrow{\mathcal{L}} N(0,1).$$

A proof is given, e.g., in Billingsley (1979, pp. 310-312).

Lemma A5: If $G_n(x)$ is a sequence of dfs converging to a continuous df $G(x)$ then the convergence is uniform in x.

This is known as Polya's lemma. For a proof see e.g., Galambos (1978, p. 111).

Lemma A6: If $G_n(x)$ is a sequence of dfs converging to a df $G(x)$ and if the g_n's are bounded continuous functions converging uniformly to g, then

$$\lim_{n\to\infty} \int g_n dG_n = \int g dG.$$

This appears in Chung (1974, p. 93).

Lemma A7: If $\underline{X}_n - \underline{Y}_n \xrightarrow{P} 0$ and $\underline{Y}_n \xrightarrow{\mathcal{L}} \underline{Y}$ where all the rvs involved are of k dimensions, then $\underline{X}_n \xrightarrow{\mathcal{L}} \underline{Y}$.

A proof for the case k=1 is given in Rao (1973, p. 123). A similar proof can be written for k>1.

Lemma A8: Let $\underline{X}_n = (X_n^{(1)},\ldots,X_n^{(k)})$ and $\underline{X} = (X^{(1)},\ldots,X^{(k)})$ be k-dimensional rvs. Then, with the usual Euclidean distance function in the definition of convergence in probability, $X_n^{(j)} \xrightarrow{P} X^{(j)}$, j = 1 to k, iff $\underline{X}_n \xrightarrow{P} \underline{X}$.

The proof is easy.