SOME PROPERTIES OF CONDITIONAL DISTRIBUTIONS OF A SPECIAL TYPE

by

Jacob Van Bowen, Jr.

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

MASTER OF SCIENCE

in

Statistics

Approved:

Chairman, Dr. D.R. Jensen

Dr. B. Harshbarger   Dr. R.H. Myers

March 3, 1966

Blacksburg, Virginia

TABLE OF CONTENTS

INTRODUCTION

The subject to be treated in this thesis is the conditional distribution of a random variable given that the outcome of an associated random variable lies within a specified interval. This may be considered to be an extension of the classical case in which the outcome of the associated random variable is known to assume a specific numerical value. A brief resume of the classical theory will lay a foundation for the development of the more general case and will help to establish the notation which is to be used in the sequel.

Although the normal distribution is not the only useful distribution, it is perhaps the most well known. Since the properties of the multivariate normal distribution and related distributions have been characterized in detail it will serve to introduce our notation.

Let $\underline{X}'$ be a random vector $[X_1, X_2, \ldots X_p]$ such that the joint density of $X_1, X_2, \ldots X_p$ is the p-variate normal. Let $\underline{X}$ be partitioned as follows: $\underline{X} = \left(\frac{\underline{X}_1}{\underline{X}_2}\right)$.

Let $E(\underline{X}) = \underline{\mu} = E\left(\frac{\underline{X}_1}{\underline{X}_2}\right) = \left(\frac{\underline{\mu}_1}{\underline{\mu}_2}\right)$ be the mean vector.

Let $V(\underline{X}) = \sum = E(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})'$ be the variance - covariance matrix.

Let $E(\underline{X}_1-\mu_1)(\underline{X}_1-\mu_1)' = \Sigma_{11}$, $E(\underline{X}_2-\mu_2)(\underline{X}-\mu_2)' = \Sigma_{22}$ and $E(\underline{X}_1-\mu_1)(\underline{X}_2-\mu_2)' = \Sigma_{12} = \Sigma_{21}'$. It follows that

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} .$$

See Anderson {1} for a more extensive discussion in sec. 2.3.

It is well known that $E(\underline{X}_1|\underline{X}_2=\underline{x}_2) = \mu_1+\Sigma_{12}\,\Sigma_{22}^{-1}(\underline{x}_2-\mu_2)$ and that $V(\underline{X}_1|\underline{X}_2=\underline{x}_2) = \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \equiv \Sigma_{11\cdot 2}$.

Usually $\Sigma_{12}\,\Sigma_{22}^{-1}$ is said to be the matrix of coefficients of the regression of $\underline{X}_1$ on $\underline{x}_2$. The elements of $\Sigma_{12}\,\Sigma_{22}^{-1}$ will be denoted by β_{ij}.

In practice, when we assume $E(\underline{Y}|\underline{X})=\underline{\alpha}+\beta\underline{X}$, we estimate the β_{ij} and α_i using the method of least squares or maximum likelihood. These two methods turn out to give the same results under our assumption of normality.

After this is done we have a relation with which we may predict $\underline{Y}$ given that $\underline{X}=\underline{x}$. This type of conditional prediction model is adequate if the precise numerical values of X_i are to be observed, but suppose that because of cost or because of poor precision in measurement it is found impractical to observe the X_i precisely. Is there an extension of this model which will enable us to estimate the expected value of $\underline{Y}$ given only that $\underline{X}$ is known to be in some region?

The purpose of this thesis is to explore the properties of a system in which the conditioning statement specifies something other than that the conditioning variables have assumed fixed values. The more general problem, i.e. some of the variates are known to lie in some arbitrary measurable subset of their Cartesian product space, seems rather impractical. We shall consider only the case where it is known that each component of our conditioning vector $\underline{X}$ lies in some interval in Euclidean space. The theory will be developed considering arbitrary closed intervals, but this can be extended to apply to open or half open intervals very simply. Although a more general approach might warrant some consideration, it appears to be sufficient for the purposes of the present study to consider only conditioning statements of the form: $X_i \epsilon [a_i,b_i]$, where X_i is the i^{th} component of $\underline{X}' = (X_1,X_2,\ldots X_p)$ and $a_i \leq b_i$ for each $i=1,2,\ldots p$.

One might ask when this type of prediction could be used to benefit the experimenter. There are many practical examples that illustrate the need and use of such an extension to the classical prediction model. The most succinct way to demonstrate this need is to consider any problem in which we feel justified in assuming a linear model to predict the outcome of some event. We shall mean by a linear model that:

1) $E(\underline{Y}|\underline{X}) = \underline{\alpha} + \beta\underline{X}$

2) The variance-covariance matrix of $\underline{Y}$ given that $\underline{X}=\underline{x}$ does not depend on $\underline{x}$.

Let us assume that one of our conditioning variates, say X_1, may be measured in two different ways. At great expense we may observe X_1 with almost no error; or at little expense we may observe that $X_1 \in (a,b]$, where a and b can be assumed not to be random variables. In order to make the example more concrete, assume that we have been estimating $E(Y|\underline{X}=\underline{x})$ using the classical linear model for some time and therefore we have estimates of α and $(\beta_1,\beta_2,\ldots\beta_p)$. Let us assume that the device for measuring X_1 very precisely breaks and that we are faced with the problem of replacing that device with another expensive one or replacing it with a much cheaper one which will be far less precise. The more expensive device might register x_1 to three decimal places and the less expensive device may only register integral values, i.e. we shall be able to determine only whether x_1 is in some set of intervals $(I-.5, I+.5]$ where I is an integer.

We shall attempt to answer some of the following questions: What effect will replacing the more expensive device with the less expensive device have on the mean and variance of what we are trying to predict? What can be said about the effect that the length of the interval in question will have on the moments of Y given $X_j = x_j$ in contrast to the conditioning statement $X_j \in (a_j,b_j]$; $j = 1,2,\ldots p$?

We shall refer to the classical conditional distributions as point conditional distributions. If we are given that some of the random variables $\{X_i\}_{i=1}^p$ are in $\{[a_i,b_i]\}_{i=1}^p$, we shall refer to the distribution of $\underline{Y}$ given this information as the interval conditioned distribution.

It will not be attempted to codify estimation procedure, although in many cases this will be obvious. The primary goal is to explore and characterize the system formed when interval conditioning is imposed on some set of dependent random variables.

We shall discover that if in the interval conditioned distribution we allow each interval $[a_i,b_i]$ to decrease in length to a point c_i, then we obtain, in the limit, precisely the point conditioned distribution. If we allow one of the intervals to increase in length without bound, we find that in the limit we are considering a marginal conditional prediction model, namely with that variable omitted in the model. Therefore the properties of the interval conditioned distributions include as special cases the classical point conditioned distributions and the marginal or unconditioned distributions.

I THE BIVARIATE CASE

It will be assumed that X and Y are dependent random variables. Let $F(Y,X)$ be the joint probability distribution function of X and Y, and let $F(X)$ and $F(Y)$ be the marginal probability distribution functions of X and Y, respectively. Let $f(Y,X)$ be the joint frequency function of X and Y, and let $f(X)$ and $f(Y)$ be the marginal frequency functions of X and Y, respectively. We shall assume that $f(X)$ has bounded moments of all orders.

The model which will be considered here is:

1) $E(Y|X=x) = \alpha + \beta x$, where α and β are functions only of some of the parameters of $f(Y,X)$,

2) $V(Y|X=x) = cV(Y)$, where c is a function only of some of the parameters of $f(Y,X)$ and $V(Y) = E(Y-E(Y))^2$ and $V(Y|X=x) = E\{(Y-E(Y|X=x))^2|X=x\}$.

In principle we may find $f(Y|X\epsilon[a,b])$, where a and b are pre-assigned constants with $a \leq b$; however, in most practical cases we shall be content to find the moments of this distribution. We shall assume that in no case is $Pr(X\epsilon[a,b]) = 0$.

Assuming that $f(Y,X)$ is a density function, we now have the relation:

$$f(Y|X\epsilon[a,b]) = \frac{\int_a^b f(Y,X)dX}{\int_a^b f(X)dX}\text{ , when } X\epsilon[a,b] \text{ and}$$

$f(Y|X\epsilon[a,b]) = 0$ otherwise.

There are a few theorems which will be of use throughout the sequel. These will be presented in the next section.

1.1 PRELIMINARY RESULTS

<u>Theorem 1</u>. The expectation of a function of Y and X, $h(Y,X)$, given that $X\epsilon[a,b]$, is the expectation of $E(h(Y,X)|X)$, where the expectation is taken with respect to the truncated distribution of X, i.e. truncated to $[a,b]$.

Proof: $\Pr(Y \leq y | X\epsilon[a,b]) = \dfrac{\Pr(Y \leq y,\ a \leq X \leq b)}{\Pr(a \leq X \leq b)}$,

provided $\Pr(X\epsilon[a,b])$ is not zero. This is seen to be the truncated distribution of Y and X, where X can occur only in $[a,b]$. Letting $\int_a^b f(X)dX = \varphi(a,b)$ we can write

$$E(h(Y,X)|X\epsilon[a,b]) = \frac{\int_{-\infty}^{\infty}\int_a^b h(Y,X)f(Y,X)dXdY}{\varphi(a,b)} .$$

But $f(Y,X) = f(X)f(Y|X)$, therefore

$$E(h(Y,X)|X\epsilon[a,b]) = \frac{\int_a^b f(X) \int_{-\infty}^{\infty} h(Y,X)f(Y|X)dYdX}{\varphi(a,b)}$$

$$= \frac{\int_a^b f(X)\ E\{h(Y,X)|X\}dX}{\varphi(a,b)} .$$

We see that $\frac{f(X)}{\varphi(a,b)}$ is the form of the density of X truncated to the interval [a,b] . Therefore

$$\frac{\int_a^b f(X)E(h(Y,X)|X)dX}{\varphi(a,b)}$$ is simply the expectation of

$E(h(Y,X)|X)$ in the appropriate truncated distribution of X.

We should note here that if a = b then

$$E(h(Y,X)|X=b=a) = \frac{\int_{-\infty}^{\infty} f(Y,b)h(Y,b)dY}{f_x(b)} ,$$

provided $f(X)|_b = f_x(b)$ is not zero. This is the classical point conditioning case.

We may use this theorem to obtain the expectation of Y given that $X\epsilon[a,b]$.

<u>Theorem 2</u>. If $Pr(a \leq X \leq b)$ is not zero, and if we assume the linear model, then $E(Y|X\epsilon[a,b]) = \alpha + \beta M_T(X)$, where $M_T(X)$ is the mean of the appropriate truncated distribution of X .

Proof: If we apply Theorem 1, where $h(Y,X) = Y$, and

$$\int_a^b f(X)dX = \varphi(a,b) ,$$

it follows that:

$$E(Y|X\epsilon[a,b]) = \frac{\int_a^b f(X)\{\alpha+\beta X\}dX}{\varphi(a,b)} .$$

Therefore $E(Y|X\epsilon[a,b]) = \alpha+\beta M_T(X)$, where $M_T(X) = \dfrac{\int_a^b f(X)XdX}{\varphi(a,b)}$.

It is worth noting here, that

$$E(XY|X\epsilon[a,b]) = \frac{\int_a^b X \int_{-\infty}^{\infty} Yf(Y,X)dYdX}{\varphi(a,b)}$$

$$= \frac{\int_a^b Xf(X) \int_{-\infty}^{\infty} Yf(Y|X)dYdX}{\varphi(a,b)} .$$

$\therefore$ $E(XY|X\epsilon[a,b]) = \dfrac{\int_a^b E(Y|X)Xf(X)dX}{\varphi(a,b)}$. Assuming a linear model we obtain:

$$E(XY|X\epsilon[a,b]) = \alpha M_T(X) + \beta \frac{\int_a^b X^2f(X)dX}{\varphi(a,b)} .$$

It will also be very convenient to use Theorem 1 in obtaining $V(Y|X\epsilon[a,b])$, which is accomplished in the next theorem.

Before we state the next theorem we need to introduce the notation which will be used. Let $M_{Y|X} = E(Y|X)$ and

$M_{Y|X\epsilon[a,b]} = E(Y|X\epsilon[a,b])$.

Then $V(Y|X\epsilon[a,b]) = E\{(Y-M_{Y|X\epsilon[a,b]})^2|X\epsilon[a,b]\}$. We shall let

$$\sigma_T^2(X) = \frac{\int_a^b X^2 f(X)dX}{\varphi(a,b)} - \{M_T(X)\}^2 \text{ , where } \varphi(a,b) = \int_a^b f(X)dX.$$

<u>Theorem 3</u>. The variance of Y given that $X\epsilon[a,b]$ is $V(Y|X=x) + \beta^2\sigma_T^2(X)$, where we have assumed the linear model.

Proof: Consider the expression

$$E\left[(Y-M_{Y|X}+M_{Y|X}-M_{Y|X\epsilon[a,b]})^2|X\epsilon[a,b]\right] .$$

This is simply $V(Y|X\epsilon[a,b])$ and is equal to

$$E\left[(Y-M_{Y|X})^2 + 2(Y-M_{Y|X})(M_{Y|X}-M_{Y|X\epsilon[a,b]}) + (M_{Y|X}-M_{Y|X\epsilon[a,b]})^2|X\epsilon[a,b]\right] .$$

We apply Theorem 1 to this expression and obtain:

$$V(Y|X\epsilon[a,b]) = \frac{\int_a^b f(X)E\{(Y-M_{Y|X})^2|X\}dX}{\varphi(a,b)} + \frac{\int_a^b f(X)\{\beta X-\beta M_T(X)\}^2 dX}{\varphi(a,b)} + 2E\left[(Y-M_{Y|X})(M_{Y|X}-M_{Y|X\epsilon[a,b]})|X\epsilon[a,b]\right].$$

The first two terms are seen to be

$$V(Y|X=x) + \beta^2\sigma_T^2(X)$$

and it remains to be proved that

$$2E\left[(Y-M_{Y|X})(M_{Y|X}-M_{Y|X\epsilon[a,b]})|X\epsilon[a,b]\right] \text{ is zero.}$$

We again invoke Theorem 1 and obtain

$$2E\left[E\{(Y-M_{Y|X})(M_{Y|X}-M_{Y|X\epsilon[a,b]})|X\}|X\epsilon[a,b]\right];$$ but given X,

$M_{Y|X}-M_{Y|X\epsilon[a,b]}$ is fixed and therefore

$$E\{(Y-M_{Y|X})(M_{Y|X}-M_{Y|X\epsilon[a,b]})|X\} =$$

$$(M_{Y|X}-M_{Y|X\epsilon[a,b]})E\{(Y-M_{Y|X})|X\} = 0.$$

Therefore we see that

$$V(Y|X\epsilon[a,b]) = V(Y|X=x) + \sigma_T^2(X) .$$

Since we assumed $V(Y|X=x) = cV(Y)$ we may write this expression as follows:

$$V(Y|X\epsilon[a,b]) = cV(Y) + \sigma_T^2(X) .$$

At this point we can see that it will be very useful to know the behavior of $\sigma_T^2(X)$ as we alter the interval $[a,b]$.

Let us digress for a few paragraphs in order to present a general method for obtaining the variance of a function of X and Y given that $X\epsilon[a,b]$.

It is a rather well known fact that if X and Y have a joint distribution, then we may obtain $V\{Y\}$ by the following relation: $V\{Y\} = E\{V(Y|X)\} + V\{E(Y|X)\}$. See section 2.2 of Parzen {12}.

A corresponding relationship may be derived which is useful in the case of interval conditioning. This will be developed below.

<u>Theorem 4.</u>

(1) $$V\{h(Y,X)|X\epsilon[a,b]\} = E\Big[V\{h(Y,X)|X\}|X\epsilon[a,b]\Big] + V\Big[E\{h(Y,X)|X\}|X\epsilon[a,b]\Big] .$$

Proof: Let I designate [a,b]. We shall denote $E\{h(Y,X)|X\}$ by $M_{h|X}$ and $E\{h(Y,X)|X\epsilon[a,b]\}$ by $M_{h,I}$.

We may write

$$V\{h(Y,X)|X\epsilon I\} = E\Big[\{h(Y,X)-M_{h|X}+M_{h|X}-M_{h,I}\}^2|X\epsilon I\Big] .$$

We apply Theorem 1 to this expression to obtain:

$$V\{h(Y,X)|X\epsilon I\} = E\Big[V\{h(Y,X)|X\}|X\epsilon I\Big] + V\Big[E\{h(Y,X)|X\}|X\epsilon I\Big] + 2E\Big[(h(Y,X)-M_{h|X})(M_{h|X}-M_{h,I})|X\epsilon I\Big] .$$

Now all that is left to be shown is that the last term is zero.

We invoke Theorem 1 once again. Letting

$$\int_a^b f(X)dX = \varphi(a,b),$$

the last term becomes

$$\left[\frac{2}{\varphi(a,b)}\right] \int_a^b f(X)E\Big[(h(Y,X)-M_{h|X})(M_{h|X}-M_{h,I})|X\Big]dX,$$

but obviously $E\left[(h(Y,X)-M_{h|X})(M_{h|X}-M_{h,I})|X\right]$ is zero,

since $M_{h|X}$ is fixed if we are given X.

This proves the theorem and it also shows that

$$E\{(h(Y,X)(M_{h|X}))|X\epsilon I\} = E\{h(Y,X)|X\epsilon I\}E\{M_{h|X}|X\epsilon I\} .$$

If we let [a,b] consist of the domain of X in equation (1) we obtain Parzen's relation.

We should note here that the relation

$$V(Y) = E\{V(Y|X)\} + V\{E(Y|X)\}$$

could be applied to the newly formed random variable $Y' = Y|X\epsilon[a,b]$, where

$$g(Y',X) = f(Y,X|X\epsilon[a,b]) = \frac{f(Y,X)}{\varphi(a,b)} ,$$

$X\epsilon[a,b]$ and $g(Y',X) = 0$, otherwise.

We apply the relation to Y' and obtain:
$V(Y') = E\{V(Y'|X)\} + V\{E(Y'|X)\}$, $X\epsilon[a,b]$, which is equivalent to the results we obtained above.

We have found that if we assume a linear model, $E(Y|X) = \alpha + \beta X$ and $V(Y|X)$ independent of X, then

$$E(Y|X\epsilon[a,b]) = \alpha + \beta M_T(X) \text{ and}$$

$$V(Y|X\epsilon[a,b]) = V(Y|X) + \beta^2\sigma_T{}^2(X).$$

We may deduce from these results that

$$V(Y|X\epsilon[a,b]) \geq V(Y|X), \text{ for } \beta^2\sigma_T{}^2(X) \geq 0.$$

The equality holds if and only if Y and X are independent, where $\sigma_T^{2}(X) \neq 0$.

It may also be shown that if $V(Y|X)$ is independent of X then $V(Y) \geq V(Y|X)$, for as we have seen

$$V(Y) = V\{E(Y|X)\} + E\{V(Y|X)\}.$$

Since $V(Y|X)$ is independent of X, then $E\{V(Y|X)\} = V(Y|X)$ and therefore

$$V(Y) = V(Y|X) + V\{E(Y|X)\}.$$

Obviously $V\{E(Y|X)\} \geq 0$. Therefore $V(Y) \geq V(Y|X)$.

In the next section we shall ponder the proposition:

$$V(Y) \geq V(Y|X\epsilon I) \geq V(Y|X),$$

where $V(Y|X)$ is independent of X. We shall find that if $I \to (-\infty, \infty)$ then the first equality holds and that if $I \to x_0$ then the second equality holds. Obviously both equalities hold if Y and X are independent. In the appendix we shall see that unless there is some kind of monotonicity of $V(Y|X\epsilon I)$ with respect to I, then the statement $V(Y) \geq V(Y|X\epsilon I)$ may not hold.

1.2 LIMITING PROPERTIES

At this point it is quite natural to ask what happens to the mean and variance of our conditioned random variable $(Y|X\epsilon[a,b])$ as $a \to \alpha_0^-$ and $b \to \alpha_0^+$. We might rephrase this and ask what happens to the mean and variance as $[a,b]$ is allowed to approach a point $\alpha_0 \epsilon [a,b]$.

It will be tacitly assumed that $b \geq a$ in all cases and that the density function $f(X)$ is positive and continuous at each a_o which will be considered.

Let $\Psi(u,v) = \int_u^v Xf(X)dX$ and $\varphi(u,v) = \int_u^v f(X)dX$.

It was shown that $E(Y|X\epsilon[a,b]) = \alpha + \beta \frac{\Psi(a,b)}{\varphi(a,b)}$.

Consider $\lim_{a \to a_o} E(Y|X\epsilon[a,b])$. Since we have assumed that for each $\epsilon > 0$ $\varphi(a_o,a_o+\epsilon) > 0$, by continuity, we may write

$$\lim_{a \to a_o} E(Y|X\epsilon[a,b]) = \alpha + \beta \frac{\Psi(a_o,b)}{\varphi(a_o,b)} .$$

We apply L'Hospital's rule to $\frac{\Psi(a_o,b)}{\varphi(a_o,b)}$ to obtain the limit as $b \to a_o$.

$$\lim_{t \to a_o} \left[\frac{d}{dt} \Psi(a_o,t)\right] = \lim_{t \to a_o} t\ f(t)$$

$$\lim_{t \to a_o} \left[\frac{d}{dt} \varphi(a_o,t)\right] = \lim_{t \to a_o} f(t) .$$

Therefore $\lim_{b \to a_o} E(Y|X\epsilon[a_o,b]) = \alpha + \beta \lim_{t \to a_o} \frac{t\ f(t)}{f(t)}$.

We assumed $\lim_{t \to a_o} f(t) > 0$, therefore

$$\lim_{b \to a_o} E(Y|X\epsilon[a_o,b]) = \alpha + \beta a_o .$$

We might abuse our notation and write:

$$\lim_{[a,b] \to X} E(Y|X\epsilon[a,b]) = \alpha + \beta X, \text{ where it exists.}$$

This is immediately seen to be $E(Y|X)$, the classical point conditioning problem.

We next examine what happens to $V(Y|X\epsilon[a,b])$ as $a \to a_0^-$ and $b \to a_0^+$. As before, we find that

$$\lim_{a \to a_0} V(Y|X\epsilon[a,b]) = V(Y|X\epsilon[a_0,b]).$$

If the probability that $a_0 \leq X \leq t$ is not zero for each $t > a_0$, then we seek the limit of $V(Y|X\epsilon[a_0,b])$ as $b \to a_0$.

We shall have need of L'Hospital's rule as we did in the last argument. It has been shown that

$$V(Y|X\epsilon[a,b]) = V(Y|X) + \beta^2 V(X|X\epsilon[a,b]).$$

Applying L'Hospital's rule to

$$E(X^2|X\epsilon[a,b]) - \left[E(X|X\epsilon[a,b])\right]^2,$$

applying it twice to the second term, we obtain:

$$\lim_{t \to a_0} V(Y|X\epsilon[a_0,t]) = V(Y|X) + \beta^2 \left[\lim_{t \to a_0} \left(\frac{t^2 f(t)}{f(t)} - \frac{\{t\ f(t)\}^2}{\{f(t)\}^2}\right)\right]$$

$$= V(Y|X) + \beta^2(0) = V(Y|X),$$

which was assumed to be $cV(Y)$.

This is precisely the result obtained in the classical case of point conditioning.

Thus we have found that as our interval "shrinks" to a point we obtain point conditional results as a special case of our extension to interval conditioning. This is not an unexpected result if one recalls the definition of a conditional distribution.

Now we examine the behavior of the mean and variance of Y, given that $X \epsilon [a,b]$, as $a \to -\infty$ and $b \to +\infty$. It is easily seen that $\lim_{\Delta \to \infty} E(Y|X \epsilon [a-\Delta, b+\Delta]) = \alpha + \beta\, E(X)$, for $M_T(X) \to E(X)$.

We may look at this in a different light, as follows:

$$\lim_{\Delta \to \infty} E(Y|X \epsilon [a-\Delta, b+\Delta]) = \frac{\int_{-\infty}^{\infty} Y \int_{-\infty}^{\infty} f(Y,X)\,dX\,dY}{\varphi(-\infty,\infty)} = E(Y) .$$

These two forms are immediately seen to be equivalent, for

$$E(Y) = \underset{X}{E}\{E(Y|X)\} = \alpha + \beta E(X) .$$

See Graybill {7} page 199.

We see that the assumption of the linear model imposes the restriction that $\alpha + \beta E(X) = E(Y)$.

If we write $E(Y|X) = \alpha' + \beta\{X - E(X)\}$, then $\alpha' = E(Y)$, which is well known.

Finally let us look at $V(Y|X \epsilon [a,b])$ as $a \to -\infty$ and $b \to +\infty$.

It immediately follows that

$$\lim_{\Delta \to \infty} V(Y|X \epsilon [a-\Delta, b+\Delta]) = V(Y|X) + \beta^2\, V(X), \text{ for } \sigma_T^{\,2}(X) \to V(X).$$

We should note that an equivalent form may be obtained as follows:

$$V(Y|X \epsilon [a,b]) = E(Y^2|X \epsilon [a,b]) - \{E(Y|X \epsilon [a,b])\}^2$$

and therefore

$$\lim_{\substack{a \to -\infty \\ b \to +\infty}} V(Y|X \epsilon [a,b]) = \lim_{\substack{a \to -\infty \\ b \to +\infty}} E(Y^2|X \epsilon [a,b]) - \{E(Y)\}^2,$$

for we have seen that $\lim\limits_{\substack{a\to-\infty \\ b\to+\infty}} E(Y|X\epsilon[a,b]) = E(Y)$. Also

$$\frac{\int_{-\infty}^{\infty} Y^2 \int_{-\infty}^{\infty} f(Y,X)dXdY}{\varphi(-\infty,\infty)} = E(Y^2) \ .$$

Therefore

$$\lim_{\substack{a\to-\infty \\ b\to+\infty}} V(Y|X\epsilon[a,b]) = E(Y^2) - \{E(Y)\}^2 = V(Y) \ .$$

Again $V(Y)$ and $V(Y|X)+\beta^2V(X)$ are seen to be equivalent. Recall that

$$V(Y) = E\{V(Y|X)\} + V\{E(Y|X)\} \ .$$

We assumed $V(Y|X)$ was not a function of X, and therefore

$$E\{V(Y|X)\} \equiv V(Y|X) \ .$$

Therefore

$$V(Y) = V(Y|X) + V(\alpha+\beta X) = V(Y|X) + \beta^2 V(X) \ .$$

We have shown that in the special case when $I \to x_o$ we have

$$V(Y) \geq V(Y|X\epsilon I) = V(Y|X=x_o),$$

and that when $I \to (-\infty,\infty)$ we have

$$V(Y) = V(Y|X\epsilon I) \geq V(Y|X).$$

Let us recall that

$$V(Y|X\epsilon I) = E\{V(Y|X)|X\epsilon I\} + V\{E(Y|X)|X\epsilon I\}$$

and that

$$V(Y) = V(Y|X) + \beta^2 V(X) \ .$$

Assuming that $V(Y|X)$ is independent of X, we have

$$V(Y|X\epsilon I) = V(Y|X) + V(\beta X|X\epsilon I) = V(Y|X) + \beta^2 V(X|X\epsilon I).$$

Therefore

$$V(Y) - V(Y|X\epsilon I) = \beta^2\left[V(X) - V(X\epsilon I)\right] .$$

Therefore, in order to assert that

$$V(Y) \geq V(Y|X\epsilon I) \geq V(Y|X) \text{ for all } I,$$

where $V(Y|X)$ is independent of X, we must show that

$$V(X) \geq V(X\epsilon I)$$

for all I. This is not the case in general. Sufficient conditions for this property may be found in the appendix. We shall discuss some special cases in the next section which may tend to clarify this question.

1.3 MONOTONICITY OF VARIANCE

Now that we have seen what happens to the variance in the limiting cases, we should like to know what is the behavior in between these two extremes. It would be useful, for purposes of applications, to be able to say that the variance is monotone decreasing as we decrease the interval in some nested manner. This is not true in general. However, we can make some general statements about what happens if we restrict each successive choice of intervals to leave the previous mean unchanged.

Lemma 1. If we consider all frequency functions defined to be zero on the complement of the interval $[a,b]$, the maximum variance is $\frac{(b-a)^2}{4}$.

Proof: Consider $\{f_i\}_{i\epsilon\{I\}}$, the set of all frequency functions defined to be zero on the complement of $[a,b]$. We assume that the family of all distributions $\{P_j(X) = 0 \text{ except at } a \text{ and } b\}_{j\epsilon[0,1]}$, where

$$P_j(X) = j, \text{ if } X=a \;; \quad = 1-j, \text{ if } X=b \,,$$

will contain that distribution with maximal variance.

Let $Pr(X=a) = p$ and $Pr(X=b) = q$, $p + q = 1$. If we compute the variance and then differentiate with respect to p , et cetera, we find that $p = q = 1/2$ maximizes the expression.

Therefore we conclude that the maximal variance,

$$\sum_{X=a,b} (X-E(X))^2 P(X) \,,$$

is $\quad (1/2)\{(a-b)/2\}^2 + (1/2)\{(b-a)/2\}^2 = \{(b-a)/2\}^2 = \dfrac{(b-a)^2}{4}$.

Since $\dfrac{(b-a)^2}{4} \leq \max\{(b-\mu)^2, (\mu-a)^2\}$, we may conclude that the maximum variance can be no larger than $\max\{(b-\mu)^2, (\mu-a)^2\}$, where $\mu = EX$.

With similar assumptions it is easy to show that if we require the mean to be zero, (note $a < 0$ and $b > 0$), then the maximum variance is

(2) $\qquad$ $\text{Max. Var} = (-ab)$.

Let us consider a density function $f(X)$ with bounded mean and variance. Let $f(X)$ be truncated to some interval I_o with mean μ_o. We shall agree to call

$I_n \subset I_{n-1} \subset \ldots \subset I_o$ a nested decreasing truncation about μ_o and

$I_o = J_o \subset J_1 \subset \ldots \subset J_m$ a nested increasing truncation about μ_o,

where $J_i = J_j$ and $I_i = I_j$ if and only if $i=j$ and the mean after each truncation is μ_o. We shall let $VAR(I_j)$ denote the variance of the truncated variate associated with the interval I_j, and require that each truncation not be trivial, i.e. "some probability is excluded (or included) with each successive truncation." Subject to these restrictions and definitions we now prove a theorem.

<u>Theorem 5</u>. A nested increasing truncation implies that the sequence $\{VAR(J_i)\}_{i=0}^{m}$ is a monotone increasing sequence.

Proof: Without loss of generality, we let the constant mean $\mu = \mu_o = 0$. We define $\sigma^2(t)$ to be

$$\int_{u(t)}^{t} X^2 dF(X) \Big/ \int_{u(t)}^{t} dF(X) ,$$

and we let

$$\int_{u(t)}^{t} dF(X) = \varphi\{u(t),t\} .$$

Since

$$\mu = 0 = \int_{u(t)}^{t} XdF(X) \text{ for any } t \geq 0,$$

$$\frac{d\mu}{dt} = tf(t) - u(t)f(u(t))\left[\frac{du(t)}{dt}\right] = 0 .$$

Therefore we have the relation $\frac{du(t)}{dt} = \frac{tf(t)}{u(t)f(u(t))}$,

assuming that $f(u(t))$ is not equal to zero. Therefore we can write:

$$\frac{d\sigma^2(t)}{dt} =$$

$$\left[\frac{1}{\varphi\{u(t),t\}}\right]\left[t^2 f(t) - u(t)tf(t) - \sigma^2(t)\{f(t) - \frac{tf(t)}{u(t)}\}\right] =$$

$$\left[\frac{t(t-u(t))f(t)}{\varphi\{u(t),t\}}\right]\left[1 + \frac{\sigma^2(t)}{tu(t)}\right] .$$

Therefore $\frac{d\sigma^2(t)}{dt}$ is positive provided $1 + \sigma^2(t)/tu(t)$ is positive. We ask now, is $\sigma^2(t) \leq -tu(t)$? We showed that $\sigma^2(t) \leq \{-tu(t)\}$, see equation (2) following Lemma 1.

We have seen that if we increase the length of our interval, i.e. a non-trivial increase, we necessarily increase the variance if the mean remains unaltered. This also implies that a decreasing non-trivial truncation will have associated with it a decreasing sequence of variances.

If we have a discrete variate the same results hold. See Widder's Advanced Calculus page 167 {13}.

The practical aspects of this procedure will usually be found to be of no use whatever, for even if we could choose our intervals as we pleased, the proper way to alter a particular interval, under the conditions of the previous theorem, would depend heavily on our distribution assumptions and would depend on the estimate of our original truncated mean. This would imply that our newly formed interval would have random variables for end points. We would have to assert that the new mean was equal to the original truncated mean in order to be sure that the variance would be larger.

We would prefer to be able to extend the end points and not have to worry about the alteration of the mean. Some characterizations under other hypotheses may be found in the appendix, as well as necessary and sufficient conditions for monotone variance under arbitrary extensions of an interval.

1.4 HIGHER MOMENTS

There may arise a need for higher moments, therefore we shall derive expressions yielding higher moments in general and then, as a special case, obtain them for the case of a linear model.

In order to conserve space we shall use some rather uncommon notation. Let

$$E\{(Y - M_{Y|X})^n|X\} = \sigma^n_{Y|X} \text{ and } \int_I f(X)dX = \varphi\{I\} \ .$$

We are now equipped to apply Theorem 1 to our relation:

$$E\{(Y - M_{Y|X\epsilon I})^n | X\epsilon I\} = E\{(Y - M_{Y|X} + M_{Y|X} - M_{Y|X\epsilon I})^n | X\epsilon I\} .$$

Let us recall that $M_{Y|X}$ is fixed if we are given X.

We obtain

$E\{(Y - M_{Y|X\epsilon I})^n | X\epsilon I\} = \left[\frac{1}{\varphi\{I\}}\right] \int_I f(X)\Psi(X)dX$, where $\Psi(X)$ is

$E\{(Y - M_{Y|X} + M_{Y|X} - M_{Y|X\epsilon I})^n | X\}$, which is equal to

$$E\left[\sum_{m=0}^{n} \binom{n}{m} (Y - M_{Y|X})^m (M_{Y|X} - M_{Y|X\epsilon I})^{n-m} | X\right] .$$

Using the fact that $M_{Y|X}$ is fixed, if we are given X, we obtain

$$\sum_{m=0}^{n} \binom{n}{m} \sigma^m_{Y|X} (M_{Y|X} - M_{Y|X\epsilon I})^{n-m} .$$

Therefore

$$E\{(Y - M_{Y|X\epsilon I})^n | X\epsilon I\} =$$

$$\left[\frac{1}{\varphi\{I\}}\right] \int_I f(X) \sum_{m=0}^{n} \binom{n}{m} \sigma^m_{Y|X} (M_{Y|X} - M_{Y|X\epsilon I})^{n-m} dX .$$

If we assume a linear model, this expression becomes:

$$\left[\frac{1}{\varphi\{I\}}\right] \int_I f(X) \sum_{m=0}^{n} \binom{n}{m} \beta^{n-m} \sigma^m_{Y|X} (X - M_T(X))^{n-m} dX ,$$

where $M_T(X) = \dfrac{\int_I Xf(X)dX}{\varphi\{I\}}$.

Since in most applications higher conditional moments of Y depend on the first two conditional moments of Y, one of which depends on X in the linear model, we see that the previous expression becomes rather messy, but if they were needed, theoretically they could be computed from this expression.

II THE MULTIVARIATE CASE

We shall find the counterpart of Theorem 1 in the multivariate case to be as useful as was Theorem 1 in the bivariate case.

Let $f(Y,X_1,\ldots X_p)$ be the joint density of $Y,X_1,\ldots X_p$, and let $f(Y)$ be the marginal density of Y. Let $f(\underline{X}_p)$ be the joint marginal density of $X_1,\ldots X_p$. Let it be given as a condition that $X_1 \epsilon I_1,\ldots X_p \epsilon I_p$. We shall write

$$C = \int_{I_1} \cdots \int_{I_p} f(\underline{X}_p) dX_1 \cdots dX_p$$

if the expression is not zero. For the sake of brevity, let R_x denote the Cartesian product space of $\{I_i\}_{i=1}^{p}$ and $dX_1 . dX_2 \ldots dX_p = D(\underline{X}_p)$. Let $h(Y,X_1,\ldots X_p)$, any scalar valued function of $Y,X_1,\ldots X_p$, be denoted by h.

Under these assumptions we shall now prove some theorems which will be useful in later sections.

2.1 PRELIMINARY RESULTS

<u>Theorem 1A</u>. $E\{h|\underline{X}_p \epsilon R_x\} = (1/C) \int_{R_x} f(\underline{X}_p) E\{h|\underline{X}_p\} D(\underline{X}_p)$.

Proof:

$$E\{h|\underline{X}_p \epsilon R_x\} = \int_{-\infty}^{\infty} hf(Y|\underline{X}_p \epsilon R_x)dY$$

$$= \{1/C\} \int_{-\infty}^{\infty} \int_{R_x} hf(Y,X_1,\ldots X_p)D(\underline{X}_p)dY \ .$$

But $\quad f(\underline{X}_p)f(Y,X_1,\ldots X_p)/f(\underline{X}_p) = f(\underline{X}_p)f(Y|\underline{X}_p)$.

Therefore we may write:

$$E\{h|\underline{X}_p \epsilon R_x\} = \{1/C\} \int_{-\infty}^{\infty} \int_{R_x} hf(\underline{X}_p)f(Y|\underline{X}_p)D(\underline{X}_p)dY$$

$$= \{1/C\} \int_{R_x} f(\underline{X}_p) \int_{-\infty}^{\infty} hf(Y|\underline{X}_p)dYD(\underline{X}_p) \ .$$

Therefore we have $\quad E\{h|\underline{X}_p \epsilon R_x\} = \{1/C\} \int_{R_x} E\{h|\underline{X}_p\}f(\underline{X}_p)D(\underline{X}_p)$.

Assuming the linear model $E(Y) = \alpha + \underline{\beta}' \ \underline{X}_p$ and $V(Y|\underline{X}_p)$ is independent of $\underline{X}_p$, we now shall seek the conditional mean and variance for the multivariate case corresponding to Theorems 2 and 3 in section 1.1 .

<u>Theorem 2A</u>. If we assume that $E\{Y|\underline{X}_p\} = \alpha + \underline{\beta}' \ \underline{X}_p$ and that $V(Y|\underline{X}_p)$ is independent of $\underline{X}_p$, then if C is not zero we may write $E(Y|\underline{X}_p \epsilon R_x) = \alpha + \underline{\beta}' \ \underline{M}_T$, where $\underline{M}_T$ is the column vector with $\left[\{1/C\} \int_{R_x} f(\underline{X}_p)X_i D(\underline{X}_p)\right]$ as its i^{th} entry.

Proof: We apply Theorem 1A and obtain:

$$E(Y|\underline{X}_p \epsilon R_x) = \{1/C\} \int_{R_x} f(\underline{X}_p) E\{Y|\underline{X}_p\} D(\underline{X}_p)$$

$$= \{1/C\} \int_{R_x} f(\underline{X}_p)\left[\alpha + \underline{\beta}'\ \underline{X}_p\right] D(\underline{X}_p)$$

$$= \alpha + \underline{\beta}'\underline{M}_T \ .$$

Theorem 3A. If we assume $E(Y|\underline{X}_p) = \alpha + \underline{\beta}'\underline{X}_p$ and that $V(Y|\underline{X}_p)$ is independent of $\underline{X}_p$, then

$$V\{Y|\underline{X}_p \epsilon R_x\} = V\{Y|\underline{X}_p\} + \underline{\beta}'\Sigma\ \underline{\beta}\ , \text{ where}$$

$$\Sigma = (\sigma_{ij}) \text{ and } \sigma_{ij} = \{1/C\} \int_{R_x} (X_i - M_i)(X_j - M_j) f(\underline{X}_p)\ D(\underline{X}_p)\ ;$$

$i,j = 1,2,\ldots p$, where M_i is the i^{th} element of $\underline{M}_T$.

Proof: We apply Theorem 1A to the following expression:

$$V(Y|\underline{X}_p \epsilon R_x) = E\{(Y - M_{Y|\underline{X}_p} + M_{Y|\underline{X}_p} - M_{Y|\underline{X}_p \epsilon R_x})^2 | \underline{X}_p \epsilon R_x\}\ .$$

As before, the cross product terms are zero for the same reason as they were previously in the bivariate case. That which remains is:

$$V(Y|\underline{X}_p \epsilon R_x) =$$

$$V(Y|\underline{X}_p) + \{1/C\} \int_{R_x} f(\underline{X}_p) E\{[(\alpha + \underline{\beta}'\underline{X}_p) - (\alpha + \underline{\beta}'\underline{M}_T)]^2 | \underline{X}_p\} D(\underline{X}_p)\ .$$

Therefore we have:

$$V(Y|\underline{X}_p \epsilon R_x) =$$

$$V(Y|\underline{X}_p) + \{1/C\} \int_{R_x} f(\underline{X}_p) E\{[\sum_{i=1}^{p} \beta_i \{X_i - M_T(X_i)\}]^2 | \underline{X}_p\} D(\underline{X}_p) =$$

$$V(Y|\underline{X}_p) + \underline{\beta}' \ddagger \underline{\beta} .$$

We might refer to $\ddagger$ as the "local" variance-covariance matrix of the conditioned variables.

The multivariate version of Theorem 4 holds also.

Theorem 4A. Given that C is not zero, then

$$V(h|\underline{X}_p \epsilon R_x) = E\{V(Y|\underline{X}_p)|\underline{X}_p \epsilon R_x\} + V\{E(Y|\underline{X}_p)|\underline{X}_p \epsilon R_x\} .$$

Proof: Apply Theorem 1A to the following expression:

$$V(h|\underline{X}_p \epsilon R_x) = E\{(h - M_{h|\underline{X}_p} + M_{h|\underline{X}_p} - M_{h|\underline{X}_p \epsilon R_x})^2 | \underline{X}_p \epsilon R_x\} ,$$

where $E(h|\underline{X}_p) = M_{h|\underline{X}_p}$ and $E(h|\underline{X}_p \epsilon R_x) = M_{h|\underline{X}_p \epsilon R_x}$.

Again it is obvious that the cross products are zero, and we obtain:

$$V(h|\underline{X}_p \epsilon R_x) = E\{V(h|\underline{X}_p)|\underline{X}_p \epsilon R_x\} + V\{(M_{h|\underline{X}_p})|\underline{X}_p \epsilon R_x\} .$$

The last term is equal to $V\{[E(h|\underline{X}_p)]|\underline{X}_p \epsilon R_x\}$.

2.2 THE MIXED CASE

In the multivariate case we have a more versatile situation than in the bivariate case. We may consider problems in which it is known that $X_1, X_2, \ldots X_p$ take on fixed numerical values and the remaining variates are in some rectangular region R. Before we begin the next theorem we need to introduce some notation.

Let $h(Y, X_1, X_2, \ldots X_p, Z_1, Z_2, \ldots Z_n)$ be a scalar valued function of $Y, X_1, X_2, \ldots X_p, Z_1, Z_2, \ldots Z_n$. Let the column vector $(X_1, X_2, \ldots X_p, Z_1, Z_2, \ldots Z_n)' = (\underline{W}) = \left(\frac{\underline{W}_x}{\underline{W}_z}\right) =$ $(w_1, w_2, \ldots w_{p+n})'$, where

$(\underline{W}_x)' = (X_1, X_2, \ldots X_p)$ and $(\underline{W}_z)' = (Z_1, Z_2, \ldots Z_n)$.

We shall denote $f(X_1, X_2, \ldots X_p, Z_1, Z_2, \ldots Z_n)\big|_{\underline{W}_z = \underline{a}_n}$

by $f(X_1, X_2, \ldots X_p, a_1, a_2, \ldots a_n)$ or just $f(\underline{W}_x, \underline{a}_n)$.

Let $E\{h(Y, \underline{W}) \mid \underline{W}_x \epsilon R_x, \underline{W}_z = \underline{a}_n\} = M_{h\sim}$ and

$$E\{h(Y, \underline{W}) \mid \underline{W}\} = M_{h|\underline{W}}.$$

<u>Theorem 6</u>. If we assume $E(Y|\underline{W}) = \alpha + \underline{\beta}'(\underline{W})$, where $\underline{\beta}' = (\beta_1, \beta_2, \ldots \beta_p, \beta_{p+1}, \ldots \beta_{p+n})$, then

$$E\{h(Y, \underline{W}) \mid \underline{W}_x \epsilon R_x,\ \underline{W}_z = \underline{a}_n\} = M_{h\sim} = \frac{\int_{R_x} M_{h|\underline{W}}\, f(\underline{W}_x \mid \underline{W}_z = \underline{a}_n) D(\underline{X}_p)}{\int_{R_x} f(\underline{W}_x \mid \underline{W}_z = \underline{a}_n) D(\underline{X}_p)},$$

if it exists.

Proof: In Theorem 1A we showed that

$$E\{h(Y,\underline{W})|\underline{W}\epsilon R_{\underline{W}}\} = E\left[E\{h(Y,\underline{W})|\underline{W}\}|\underline{W}\epsilon R_{\underline{W}}\right] .$$

Therefore $M_{h\sim} = E\left[\{\int_{-\infty}^{\infty} h(Y,\underline{W})f(Y|\underline{W})dY\}|\underline{W}\epsilon R_{\underline{W}}\right]$

$$= \frac{\int_{R_{\underline{W}}}\left[\int_{-\infty}^{\infty} h(Y,\underline{W})f(Y|\underline{W})dY\right] f(\underline{w}_x,\underline{w}_z)D(\underline{W})}{\left[\int_{R_{\underline{W}}} f(\underline{w}_x,\underline{w}_z)D(\underline{W})\right]'} ,$$

provided the denominator is not zero, which it would be without the prime.

We mean by this that:

$$\left[\int_{R_{\underline{W}}} f(\underline{w}_x,\underline{w}_z)D(\underline{W})\right]' = \int_{R_{\underline{w}_x}} f(\underline{w}_z)f(\underline{w}_x|\underline{w}_z)D(\underline{w}_x)$$

$$= f(\underline{w}_z)\int_{R_{\underline{w}_x}} f(\underline{w}_x|\underline{w}_z)D(\underline{w}_x) ,$$

where $\underline{w}_z = \underline{a}_n$.

If this change, i.e. $f(\underline{w}_x,\underline{w}_z) = f(\underline{w}_z)f(\underline{w}_x|\underline{w}_z)$, is made in the numerator and we cancel $f(\underline{w}_z)$, if it is not zero, we obtain $M_{h\sim} = \dfrac{\int_{R_{\underline{w}_x}} \{M_{h|\underline{W}}\}f(\underline{w}_x|\underline{w}_z)D(\underline{w}_x)}{\int_{R_{\underline{w}_x}} f(\underline{w}_x|\underline{w}_z)D(\underline{w}_x)}$, where we have assumed that the conditioning, $\underline{w}_z = \underline{a}_n$ was a non-trivial condition. Now we may write:

$$M_{h\sim} = \frac{\int_{R_x} M_{h|\underline{W}}\, f(\underline{w}_x|\underline{w}_z=\underline{a}_n)D(\underline{w}_x)}{\int_{R_x} f(\underline{w}_x|\underline{w}_z=\underline{a}_n)D(\underline{w}_x)} \quad .$$

This is very much like Theorem 1A in that we integrate the point conditioned mean over the region of truncation. In this case the only difference we encounter is that we integrate that same mean in a conditional distribution over a marginal truncation region.

It is easily verified that this is the same result that one gets if he considers the expectation given that $\underline{X}_p \epsilon R_x$ and $\underline{Z}_n \epsilon R_z$ and then lets $R_z \rightarrow (a_1, a_2, \ldots a_n)$. This follows from the definition of a multivariate conditional probability density.

Corollary 1. If we assume $E(Y|\underline{W}) = \alpha + \underline{\beta}'\underline{W}$, and if we let $\underline{W}\epsilon R_{\underline{W}}$ mean $\underline{X}_p \epsilon R_x$ and $\underline{Z}_n = \underline{a}_n$, then

$$M_{Y|\underline{W}\epsilon R_{\underline{W}}} = \alpha + \underline{\beta}' \; (m_1, m_2, \ldots m_p, a_1, a_2, \ldots a_n)' \; ,$$

where $m_i = E\{X_i|\underline{W}\epsilon R_{\underline{W}}\}$. This is a simple consequence of Theorem 6. We might note here that

$(m_1, m_2, \ldots m_p, a_1, a_2, \ldots a_n)$ could be written

$$(m_1, m_2, \ldots m_p, m_{p+1}, m_{p+2}, \ldots m_{p+n}) \text{ , for}$$

$$E\{Z_i | \underline{W} \epsilon R_{\underline{W}}\} = E\{w_{i+p} | \underline{W} \epsilon R_{\underline{W}}\} = a_i \text{ .}$$

Now we should like to find the variance of Y given $\underline{X}_p \epsilon R_x$ and $\underline{Z}_n = \underline{a}_n$. We shall retain the same notation that we used in Theorem 6.

Theorem 7. If we assume that $E(Y|\underline{W}) = \alpha + \underline{\beta}'(\underline{W})$ and that $V(Y|\underline{W})$ is independent of $\underline{W}$, then

$$V(Y|\underline{w}_x \epsilon R_x, \underline{w}_z = \underline{a}_n) = V(Y|\underline{W}) + \underline{\beta}' \, T \, \underline{\beta} \text{ ,}$$

where T is a (p+n)x(p+n) matrix such that row i is $(t_{i1}, t_{i2}, \ldots t_{ip}, o, o, \ldots o)$ if $i \leq p$ and $(\underline{o})'$ if $i > p$, where $t_{ij} = E\{(x_i - m_i)(x_j - m_j) | \underline{W} \epsilon R_{\underline{W}}\}$; $i, j \leq p$ and the $m_i = E\{x_i | \underline{W} \epsilon R_{\underline{W}}\}$.

Proof: We shall apply Theorem 6 to $V(Y|\underline{w}_x \epsilon R_x, \underline{w}_z = \underline{a}_n)$, which may be written $E\{(Y - M_{Y|\underline{W}} + M_{Y|\underline{W}} - M_{Y|\underline{W} \epsilon R_{\underline{W}}})^2 | \underline{W} \epsilon R_{\underline{W}}\}$. We defined $M_{Y|\underline{W}}$ in Theorem 6. Let $M_{Y|\underline{W} \epsilon R_{\underline{W}}} = E\{Y | \underline{W} \epsilon R_{\underline{W}}\}$.

If we expand this expression and then apply Theorem 6 we obtain:

$$V(Y|\underline{W} \epsilon R_{\underline{W}}) =$$

$$\frac{\int_{R_x}\left[V(Y|\underline{W}) + 2E\{C.P.|\underline{W}\} + E\{(M_{Y|\underline{W}}-M_{Y|\underline{W}\epsilon R_{\underline{W}}})^2|\underline{X}_p\}\right]f(\underline{w}_x|\underline{w}_z)D(\underline{X}_p)}{\int_{R_x} f(\underline{w}_x|\underline{w}_z)D(\underline{X}_p)} ,$$

where $C.P. = (Y - M_{Y|\underline{W}})(M_{Y|\underline{W}} - M_{Y|\underline{W}\epsilon R_{\underline{W}}})$.

Since $V(Y|\underline{W})$ and $M_{Y|\underline{W}}$, given $\underline{W}$, are fixed we obtain

$V(Y|\underline{W}\epsilon R_{\underline{W}}) =$

$$V(Y|\underline{W})+o+E\left[\{\alpha+(\underline{\beta})'\underline{W}-[\alpha+(\underline{\beta})'(m_1,m_2,\ldots m_p,\underline{w}_z)']\}^2|\underline{W}\epsilon R_{\underline{W}}\right] .$$

The last term is

$$E\left[\{\sum_{i=1}^{p} \beta_i(x_i - m_i)\}^2|\underline{W}\epsilon R_{\underline{W}}\right] = \underline{\beta}'T\,\underline{\beta} .$$

If we partition $\underline{\beta}$ such that

$$\underline{\beta}' = (\beta_1,\beta_2,\ldots\beta_p,\beta_{p+1},\beta_{p+2},\ldots\beta_{p+n})$$

$$= (\underline{\beta}_p',\underline{\beta}_n') ,$$

we can write:

$V(Y|\underline{W}\epsilon R_{\underline{W}}) = V(Y|\underline{W}) + \underline{\beta}_p'\,(T_c)\,\underline{\beta}_p$, where T_c is the variance-covariance matrix of $X_1,X_2,\ldots X_p$ in R_x given that $\underline{w}_z = \underline{a}_n$.

If we are concerned with predicting a vector $\underline{Y}_m' = (Y_1,Y_2,\ldots Y_m)$, the above theorems will be valid for most of the marginal properties of $\underline{Y}_m$. The more general multivariate situations will be considered in the following section.

2.3 EXTENSIONS TO PREDICTING A VECTOR

Consider the model:

$$E(\underline{Y}_m | \underline{X}_p \text{ and } \underline{Z}_n) = \underline{d}_m + (\beta)(\underline{X}'_p, \underline{Z}'_n)' ,$$

where $\underline{d}'_m = (d_1, d_2, \ldots d_m)$ is a vector of constants and β is an $(m)x(p+n)$ matrix of constants. Again we shall let

$$(\underline{X}'_p, \underline{Z}'_n) = (\underline{w}'_x, \underline{w}'_z) = (w_1, w_2, \ldots w_{n+p}) = (\underline{W})' \text{ and}$$

$$\underline{M}_{Y|\underline{W}} = E(\underline{Y}_m | \underline{W}) \text{ and } \underline{M}_{Y|\underline{W} \epsilon R_{\underline{W}}} = E(\underline{Y}_m | \underline{W} \epsilon R_{\underline{W}}) .$$

We shall insist that throughout this section we denote the variables which are point conditioned only by $\underline{w}_z$ and that we shall denote only the interval conditioned variables by $\underline{w}_x$, which have been defined above. We shall, without loss of generality, use $\underline{W}$ to denote whatever combination is appropriate. Note that $\underline{w}_z$ or $\underline{w}_x$ may be a void vector.

<u>Theorem 8</u>. If $E(\underline{Y}_m | \underline{W}) = \underline{d}_m + (\beta)\underline{W}$,

then $$E(\underline{Y}_m | \underline{W} \epsilon R_{\underline{W}}) = \underline{d}_m + (\beta)(\underline{M}'_{\underline{w}_x \epsilon R_x | \underline{w}_z}, \underline{w}'_z)' ,$$

where $$\underline{M}_{\underline{w}_x \epsilon R_x | \underline{w}_z} = E\{\underline{w}_x | \underline{w}_x \epsilon R_x, \underline{w}_z\} .$$

Proof: By $E(\underline{Y}_m | \underline{W} \epsilon R_{\underline{W}})$ we mean that column vector $\underline{M}_{Y|\underline{W}\epsilon R_{\underline{W}}} = (E\{Y_1|\underline{W}\epsilon R_{\underline{W}}\},\ldots E\{Y_m|\underline{W}\epsilon R_{\underline{W}}\})'$.

We may apply Theorem 6, or Corollary 1, to each of the marginal entries. We obtain

$$\underline{M}_{Y|\underline{W}\epsilon R_{\underline{W}}} = \underline{d}_m + (\beta)(m_1,m_2,\ldots m_p,a_1,a_2,\ldots a_n)' ,$$

where $m_i = E\{X_i|\underline{W}\epsilon R_{\underline{W}}\}$ and $\underline{w}_z' = (a_1,a_2,\ldots a_n)$.

Therefore: $\underline{M}_{Y|\underline{W}\epsilon R_{\underline{W}}} = \underline{d}_m + (\beta)(\underline{M}'_{\underline{w}_x\epsilon R_x|\underline{w}_z}, \underline{w}_z')'$.

We shall need a few definitions before proving the next theorem. Let it be given that $\underline{X}_p\epsilon R_x$ and $\underline{Z}_n = \underline{a}_n$, then we shall denote this by $\underline{W}\epsilon R_{\underline{W}}$. We shall denote the condition, given $\underline{X}_p$ and $\underline{Z}_n = \underline{a}_n$, by given $\underline{W}$. We shall let $V(\underline{Y}_m|\underline{W}\epsilon R_{\underline{W}}) = U$, where $U = (u_{ij})$ is an $(m)x(m)$ matrix. We shall let $V(\underline{Y}_m|\underline{W}) = V$, where $V = (v_{ij})$ is an $(m)x(m)$ matrix independent of $\underline{W}$. We shall denote $E\{w_i|\underline{W}\epsilon R_{\underline{W}}\}$ by m_i , and we shall denote the $(p+n)x(p+n)$ matrix with the ij^{th} element, $E\{(w_i - m_i)(w_j - m_j)|\underline{W}\epsilon R_{\underline{W}}\}$, by $\Sigma = (\sigma_{ij})$. We shall also let

$$E\{Y_i|\underline{W}\epsilon R_{\underline{W}}\} = M_{Y_i|R} \quad \text{and}$$

$$E\{Y_i|\underline{W}\} = M_{Y_i|\underline{W}} \qquad \text{and} \qquad E\{\underline{W}|\underline{W}\epsilon R_{\underline{W}}\} = \underline{M}_T = (\underline{m}_i) .$$

<u>Theorem 9</u>. If $E(\underline{Y}_m|\underline{W}) = \underline{d}_m + (\beta)\underline{W}$, where $\underline{W}' = (\underline{X}'_p, \underline{Z}'_n)$ and we are given that $\underline{W} \epsilon R_{\underline{W}}$ and V is independent of $\underline{W}$, then

$$U = V + (\beta) \ddagger (\beta)' .$$

Proof: Consider $E\{(Y_i - M_{Y_i|R})(Y_j - M_{Y_j|R})|\underline{W}\epsilon R_{\underline{W}}\}$. This is the ijth entry in U, i.e. U_{ij} . We now expand the term and get:

$$E\{Y_i Y_j|\underline{W}\epsilon R_{\underline{W}}\} - E\{M_{Y_i|R}\, M_{Y_j|R}|\underline{W}\epsilon R_{\underline{W}}\} .$$

We now add and subtract from this:

$E\{M_{Y_i|\underline{W}}\, M_{Y_j|\underline{W}}|\underline{W}\epsilon R_{\underline{W}}\}$. We now have:

$$E\{(Y_i Y_j - M_{Y_i|\underline{W}}\, M_{Y_j|\underline{W}})|\underline{W}\epsilon R_{\underline{W}}\} +$$

$$E\{(M_{Y_i|\underline{W}}\, M_{Y_j|\underline{W}} - M_{Y_i|R}\, M_{Y_j|R})|\underline{W}\epsilon R_{\underline{W}}\} .$$

We apply Theorem 6 to the first term and obtain v_{ij} . Now we have:

$$u_{ij} = v_{ij} + E\{(M_{Y_i|\underline{W}}\, M_{Y_j|\underline{W}} - M_{Y_i|R}\, M_{Y_j|R})|\underline{W}\epsilon R_{\underline{W}}\} .$$

We now replace the terms in the expectation by equivalent terms obtained via Corollary 1 of Theorem 6 and our model. Rewriting this we have:

$$u_{ij} = v_{ij} + E\{(d_i+\underline{\beta}_i'\underline{W})(d_j+\underline{\beta}_j'\underline{W}) - (d_i+\underline{\beta}_i'\underline{M}_T)(d_j+\underline{\beta}_j'\underline{M}_T)|\underline{W}\epsilon R_{\underline{W}}\} ,$$

where $\underline{\beta}_k'$ is the k^{th} row of (β) .

We see that cross product terms cancel when the expectation is taken, therefore we write:

$$\begin{aligned} u_{ij} &= v_{ij} + E\{(\underline{\beta}_i' \underline{W})(\underline{\beta}_j' \underline{W}) - (\underline{\beta}_i' \underline{M}_T)(\underline{\beta}_j' \underline{M}_T)|\underline{W}\epsilon R_{\underline{W}}\} \\ &= v_{ij} + E\{\underline{\beta}_i'(\underline{W} - \underline{M}_T)\underline{\beta}_j'(\underline{W} - \underline{M}_T)|\underline{W}\epsilon R_{\underline{W}}\} \\ &= v_{ij} + ((\beta) \Sigma (\beta)')_{ij} . \end{aligned}$$

We see that σ_{ij} is zero if i or j is greater than p.

$$\therefore U = V + (\beta) \Sigma (\beta)' .$$

This theorem raises an interesting question. Let us assume that we know the structure of the (β) matrix and that we could find a definitive mathematical relationship between the σ_{ij} involving the lengths of each of the intervals $[a_i,b_i]$. Let us also assume that if we decrease the length of some interval, say $[a_i,b_i]_o \rightarrow [a_i,b_i]_1$, then $(\sigma_{ij})_o - (\sigma_{ij})_1$ is positive definite, where $(\sigma_{ij})_o$ is the variance-covariance matrix associated with $[a_i,b_i]_o$ and $(\sigma_{ij})_1$ is the variance-covariance matrix associated with $[a_i,b_i]_1$. Can we minimize the variance of Y_i with some set of constraints on our set of intervals, namely cost constraints? We saw that if we point condition $\underline{Z}_n$ then we obtain a minimal variance situation if we point condition the other p variates. If we consider p variates interval conditioned, then the point conditioning of n others caused the upper left (p)x(p) matrix of Σ to be not the local variance-covariance matrix, but the local variance-covariance matrix given $\underline{Z}_n = \underline{a}_n$, which we assumed to be "smaller" than if the n Z variates were interval conditioned. Just how could we attack this problem with some cost levels attached to the intervals? We shall not attempt to answer this question here. We only point out that this seems to be a very interesting problem for future study.

2.4 LIMITING PROPERTIES

We shall consider the limiting properties for the linear model: $E(\underline{Y}_m|\underline{W}) = \underline{d}_m + (\beta)\underline{W}$, where $V(\underline{Y}_m|\underline{W})$ is independent of $\underline{W}$. This will provide us with the results we might want for the various cases; $\underline{Y}_m = Y_1$, $\underline{W} = X_1$, and the mixed cases. We shall continue to use the notation of Theorems 8 and 9, except we shall not insist that $\underline{w}_z$ be a given vector of constants as we did in the previous section.

We found in Theorem 8 that:

$$(3) \qquad E(\underline{Y}_m | \underline{W} \epsilon R_{\underline{W}}) = \underline{d}_m + (\beta)\, \underline{M}'_T ,$$

where $\quad \underline{M}'_T = (m_1, m_2, \ldots m_{n+p})$ with $m_i = E\{w_i | \underline{W} \epsilon R_{\underline{W}}\}$.

Here we are using $m_i = w_i$ if w_i is given. This is consistent since if $w_{p+i} = a_i$, then $E(w_{p+i} | w_{p+i} = a_i) = a_i$.

Now we ask what happens to $E(\underline{Y}_m | \underline{W} \epsilon R_{\underline{W}})$ as $R_x \to E_p$, i.e. Euclidean p-space, and R_z or $R_{\underline{w}_z} \to (a_1, a_2, \ldots a_n)$.

All we need to do to answer this question is to consider one of the m_i , since equation (3) holds for arbitrary $R_{\underline{W}}$. But an immediate consequence of Theorem 6 is that

$$m_i = E(w_i | \underline{w}_z \epsilon R_z) .$$

Applying the definition of multivariate conditional distributions to this expression we obtain:

$$m_i = E(w_i | \underline{w}_z = \underline{a}_n) .$$

This was not unexpected, for we were essentially given nothing concerning the $\underline{w}_x$ and therefore obtained the marginal expectation given only that $\underline{w}_z = \underline{a}_n$. The intermediate step, $m_i = E(w_i | \underline{w}_z \epsilon R_z)$, is the case where we have some variates interval conditioned, in R_z, and we then let the intervals of R_x increase without bound.

Retaining the same model, we now wish to know what happens to $V(\underline{Y}_m|\underline{W}\epsilon R_{\underline{W}})$ as we let $R_x \to E_p$. In this discussion we shall not require Σ, of Theorem 9, to have zero entries in the (n)x(n) lower right hand submatrix, i.e. we allow R_z to be an arbitrary rectangular region. Recall that Theorem 9 stated that

$$U = V + (\beta)\ \Sigma\ (\beta)' ,$$

where V is fixed, for any $R_{\underline{W}}$. Therefore all we need consider is the behavior of Σ as $R_x \to E_p$. We defined σ_{ij} to be $E\{(w_i - m_i)(w_j - m_j)|\underline{W}\epsilon R_{\underline{W}}\}$, for any rectangular region $R_{\underline{W}}$. Applying Theorem 6 again we see that as $R_x \to E_p$ we obtain a new

$$\sigma_{ij} = E\{(w_i - m_i)(w_j - m_j)|\underline{w}_z \epsilon R_z\} ,$$

where now $m_k = E\{w_k|\underline{w}_z \epsilon R_z\}$.

We next consider the case where we are given $\underline{W}\epsilon R_{\underline{W}}$ and R_z is a rectangular region with all intervals of positive length. We wish to examine the behavior of Σ as $R_z \to (a_1, a_2, \ldots a_n)$.

Here we first consider $f(\underline{w}_x', \underline{w}_z')$ as $R_z \to (a_1, a_2, \ldots a_n)$.

By the definition of conditional distributions we obtain $f(\underline{w}'_x | \underline{w}_z = \underline{a}_n)$, if it exists. We now apply Theorem 6 in the same way as we did in Theorem 9 and we obtain a new

$$\sigma_{ij} = E\{(w_i - m_i)(w_j - m_j) | \underline{W} \epsilon R'_{\underline{W}}\} ,$$

where $R'_{\underline{W}}$ is $\{(\underline{w}'_x, a_1, a_2, \ldots a_n)\}$ such that $\underline{w}'_x \epsilon R_x$ and

$$m_i = E\{w_i | \underline{w}_x \epsilon R_x \text{ and } \underline{w}_z = \underline{a}_n\} .$$

We see that we get the same expression that we had in Theorem 9;

$$U = V + \beta \textstyle\sum \beta' .$$ [1]

---

[1] It may seem that the last few discussions set forth are incomplete. This is why these discussions are not labelled theorems. They are meant only to be an outline for the application of previous theorems.

III SUMMARY

We have seen that there is essentially no difference in $f(Y,X|X\epsilon I)$ and the joint frequency function of Y and X where we have truncated the distribution of X in I. Even though there is no difference between the frequency functions there is a very important difference in their origins. The classical meaning of truncation is that we exclude some of the outcomes which might occur and therefore delete our sample space. Consequently, we map the elements of this new sample space with a new random variable which obviously has a different frequency function. In our interval conditioning problem we allow the random variable to take on any value, unknown to the experimenter, and then we observe only in which interval that value falls.

Although we have developed the properties with the simple statement, "given $\underline{X}\epsilon R_x$", it does not seem feasible to apply this to any experiment unless an exhaustive disjoint family of intervals may be chosen for each X_i prior to running the experiment. Recall that we assumed that each interval had end points which were not random variables. There may be some cases in which we may feel justified in approximating some intervals, but these cases are subjective and only one of them will be discussed below.

Assuming $E(Y|X) = \alpha + \beta X$, we also considered what happened to the variance of Y given $X\epsilon I$ as I was altered. We found that if $E(X|X\epsilon I) = \mu_o$ then the variance of Y given $X\epsilon I'$, $I \subset I'$, would be at least no smaller than the variance of Y given that $X\epsilon I$ if $E(X|X\epsilon I') = \mu_o$. We mentioned why this was not very useful, and referred the reader to the appendix.

One might feel that this monotone property is obvious for all practical purposes, but this is not the case. We now offer a very simple example to demonstrate that this is not true.

We shall show that if

$$V(Y|X\epsilon\{-1,k\}) = C + \beta^2 V(X|X\epsilon\{-1,k\}) ,$$

then it will depend on the frequency function of X whether

$$V(X|X\epsilon\{o,k\}) < V(X|X\epsilon\{-1,k\}) \text{ or not.}$$

In order to conserve space let us assume that the frequency function of X is approximated very closely by

$$Pr\{X=-1\}=1/3,\ Pr\{X=o\}=1/3 \text{ and } Pr\{X=k\}=1/3 .$$

Then $E(X) = (1/3)(k-1)$ and $V(X) = (2/9)(k^2+k+1)$. We now exclude $X=-1$ and have $Pr\{X'=o\}=1/2$ and $Pr\{X'=k\}=1/2$.

Therefore $E(X')=k/2$ and $V(X') = (1/4)k^2$.

We see immediately that $(1/4)k^2 > 2/9(k^2+k+1)$ for $k \geq 10$, and we see that the variance of X' may be made as much larger than the variance of X as we like.

We see that we do need to examine the exact behavior of the "local variance" before we make any decisions about selecting a family of intervals.

Since the normal distribution is usually assumed in conjunction with the linear model we should like to characterize the local variance properties for this distribution. We could not prove or disprove that in all cases the normal distribution had the monotone variance properties we had anticipated. Some of the particular intervals for which a normal random variable has the monotone variance property will be discussed in the appendix. Table 1, presented by Clark {3}, seems to be a rather good indication that the variance will increase if we extend any interval under consideration in the case of a normal variate with mean zero and variance one.

If we move up or to the left in the table from any point, the entry is found to be larger than the original entry. We do, however, recognize that this is only an indication and not conclusive.

If we are willing to accept this as sufficient evidence for the said property, then it follows that any normally distributed random variable with bounded mean and variance also has the property of monotone variance in a nested truncation.

Much more work can be done on this particular question. Some numerical work on this aspect would appear to be particularly helpful in characterizing some properties of

TABLE 1

VALUES OF THE STANDARD DEVIATION $\sigma_{a,b}$ OF THE STANDARD NORMAL POPULATION TRUNCATED AT a AND b($a \leq b$).

NOTE THAT $\sigma_{-b,-a} = \sigma_{a,b}$.

a: -3.00	-2.75	-2.50	-2.25	-2.00	-1.75	-1.50	-1.25	-1.00	-.75	-.50	-.25	0	.25	.50	a/b
.9866	.9803	.9707	.9557	.9344	.9063	.8713	.8305	.7849	.7367	.6869	.6375	.5889	.5432	.4989	3.00
	.9745	.9657	.9495	.9281	.8996	.8646	.8234	.7774	.7298	.6803	.6276	.5784	.5303	.4838	2.75
		.9546	.9394	.9176	.8891	.8535	.8118	.7605	.7155	.6639	.6120	.5610	.5106	.4610	2.50
			.9239	.9020	.8757	.8372	.7944	.7468	.6959	.6428	.5891	.5357	.4823	.4290	2.25
				.8796	.8500	.8131	.7697	.7209	.6686	.6136	.5577	.5016	.4449	.3878	2.00
					.8196	.7817	.7371	.6868	.6327	.5756	.5171	.4579	.3975	.3359	1.75
						.7425	.6966	.6446	.5885	.5291	.4679	.4055	.3414	.2792	1.50
							.6490	.5952	.5371	.4755	.4118	.3469	.2802	.2118	1.25
								.5394	.4793	.4154	.3494	.2822	.2129	.1410	1.00
									.4173	.3514	.2835	.2148	.1447	.0734	.75
										.2835	.2139	.1439	.0734	0	.50
											.1423	.0706	0		.25
												0			0

Table presented by Clark {3} (abbreviated).

the system studied here.

Most of the work done in section I, The Bivariate Case, carried over to the multivariate case as expected, with a few more variations in the limiting cases.

Since only the fundamental properties of this system could be explored in this thesis, many of the problems and applications can only be conjectured at this time.

Some of the problems to be solved may be presented best in the form of questions. How can we estimate parameters if we do not observe the variate "precisely"? How could one attack the problems concerning robustness? How would one decide which choice of intervals to use for a minimal cost estimation with a pre-assigned level of variance? These and many other problems must be left for future investigation.

The obvious application, which has been mentioned above, is predicting the outcome of some event by measuring, or observing, some associated random variables with a device which yields measurements which are not sufficiently precise to be considered, or assumed, continuous events.

It seems very feasible that this conditioning may have some practical applications in designs of experiments, particularly in some cases of missing data. If one has some way of knowing that the data lost were within given bounds, he might be able to salvage some of the information for purposes of estimation. We must keep in mind that we would have to assume that these given bounds were not random variables.

The most fruitful application would seem to be the design of a minimal cost experiment by observing more correlated variables in larger intervals. Some of these variables may nullify their usefulness in the conventional regression models because of the cost of precise observation. We might also choose some variables with established estimates for their parameters which might be very inexpensive to observe less precisely.

IV APPENDIX

It was shown, for any probability distribution, that if we considered $V(Y|Y\epsilon I_1)$ where the mean was, say M_1, then non-trivial extensions of the interval leaving the mean unchanged necessarily produced a larger variance. Obviously, this leaves a great deal to be desired. We have mentioned the problems involved in extending an interval in such a way as to leave the mean unchanged. Can we find some distributions which have the property that if we extend some arbitrary interval we may be sure that the variance is larger regardless of the change of the mean?

Let us consider first a simple case in which we assume a linear model $E(Y|X) = \alpha + \beta X$, and where $V(Y|X)$ is independent of X . We found that

$$V(Y|X\epsilon I_1) = \alpha + \beta^2 V_L ,$$

where V_L is the "local" variance of X in I_1. Since we are attempting to lay foundations for estimation and prediction procedure, it becomes very important to know the behavior of V_L as we choose different intervals. If we could master this problem, we could extend our results to the various multivariate cases without great difficulty.

This problem has had no exploration that the author has been able to locate, other than the brief numerical table computed for the normal distribution in Clark's article {3}.

Because regular behavior of V_L in some sense would appear to be essential for certain applications of conditional predictors, we direct our interests toward the search for some characterizations of non-negative functions which will insure the monotonicity of

$$\int_{R_X} \frac{(X-M(X))^2 f(X)dX}{\int_{R_X} f(X)dX}$$

as we extend R_X .

Those discoveries, made thus far by the author, which are pertinent to this problem will be presented in the form of theorems.

We now present a theorem which will give us some idea about what kinds of sets we must exclude in order to discuss the subject of monotonicity of variance. We have in the past reserved this term for discussions about the extensions of intervals. This theorem will help to show why we made this reservation.

Let us consider a random variable X with a frequency function f(X) which is positive on the Lebesgue measurable set R. Let us define a truncation of this variate to a set R_1 where $Pr\{X\epsilon R_1\} > 0$. This defines a new random variable X_1 with a frequency function $f_1(X_1) = c\ f(X)$ in R_1 , where

$\int_{R_1} f_1(X_1)dX_1 = 1$. We define another random variable X_2 similarly in R_2, where $R_2 \cap R_1 = \emptyset$ and $R_2 \cup R_1 = R$, with frequency function $f_2(X_2)$. With these definitions we prove Theorem A.

<u>Theorem A</u>. Let X_1, X_2, R_1 and R_2 be defined as above. Then it is false that $V(X_1) < V(X)$ for each $R_1 \subset R$.

Proof: Let $\int_R f(X)dX = A(R)$, $\int_{R_1} f(X)dX = A(R_1)$ and

$$\int_{R_2} f(X)dX = A(R_2) .$$

We may write:

$$A(R)\ V(X) = A(R_1)\ V(X_1) + A(R_2)\ V(X_2) + A(R_1)\{M_{X_1} - M_X\}^2 + A(R_2)\{M_{X_2} - M_X\}^2 ,$$

where $M_X = E(X)$, $M_{X_1} = E(X_1)$ and $M_{X_2} = E(X_2)$.

Since we need to demonstrate only the existence of some R_1 such that $V(X_1) < V(X)$ we may, without loss of generality, consider R_1 and R_2 such that $M_{X_1} = M_{X_2} = M_X$. Then

$$A(R)\ V(X) = A(R_1)\ V(X_1) + A(R_2)\ V(X_2) .$$

We note that $A(R) = A(R_1) + A(R_2)$. Therefore,

$$A(R_1)\{V(X) - V(X_1)\} = A(R_2)\{V(X_2) - V(X)\} .$$

If $V(X) > V(X_1)$ then the right side must be positive also. We conclude that if $V(X) > V(X_1)$ then $V(X_2) > V(X)$, and vice versa.

Therefore we must preserve some order in the successive choice of R_i if we are to talk, meaningfully, about monotone variance with the R_i . The restriction most suited for our purposes is obviously to require that the sets we shall discuss will be nested intervals.

There is a similar theorem for the multivariate case. Let $\underline{Y}$ be a vector of random variables with the joint frequency function $f(\underline{Y})$ such that $\int_R f(\underline{Y})D(\underline{Y}) = 1$. As in the previous theorem we define $f_1(\underline{Y}_1)$ by truncation in R_1 and $f_2(\underline{Y}_2)$ in R_2 , where $R_1 \cup R_2 = R$ and $R_1 \cap R_2 = \emptyset$.

We define $C = \int_R f(\underline{Y})D(\underline{Y})$, $C_1 = \int_{R_1} f(\underline{Y})\, D(\underline{Y})$ and $C_2 = \int_{R_2} f(\underline{Y})D(\underline{Y})$

and $\qquad E(\underline{Y}) = \underline{M}$, $E(\underline{Y}_1) = \underline{M}_1$ and $E(\underline{Y}_2) = \underline{M}_2$.

Let $\qquad V(\underline{Y}) = \Sigma$, $V(\underline{Y}_1) = \Sigma_1$, and $V(\underline{Y}_2) = \Sigma_2$.

Theorem B. Let $\underline{Y}_1$, $\underline{Y}_2$, Σ_1 , Σ_2 , R_1 , and R_2 be defined as above. Then it is false that $\Sigma_1 < \Sigma$ for each $R_1 \subset R$.

Proof: Consider R_1 and R_2 such that $\underline{M}_1 = \underline{M}_2 = \underline{M}$.

Then $$C\Sigma = C_1\Sigma_1 + C_2\Sigma_2 , \quad C = C_1 + C_2 .$$

$$\therefore C_1\{ \Sigma - \Sigma_1\} = C_2\{ \Sigma_2 - \Sigma\} .$$

If the left side is positive definite, then the right side must be also. Therefore $\Sigma - \Sigma_2$ is negative definite if $\Sigma - \Sigma_1$ is positive definite and vice versa.

In this case the restriction most suited for our purposes is obviously to require that the regions we shall consider will be nested rectangular regions in the space being considered.

Our next theorem is one which can be used, in special cases, with the normal distribution and any distribution when we are considering changing the lengths of intervals where the density function is monotone. We shall define

$$\sigma^2(t) = \left[\frac{1}{\varphi(a,t)}\right] \int_a^t \{X - M_{(t)}\}^2 f(X)dX ,$$

$$M_{(t)} = \int_a^t \frac{X\, f(X)dX}{\varphi(a,t)} \quad \text{and} \quad \varphi(a,t) = \int_a^t f(X)dX .$$

Theorem C. If the density function, $f(X)$, under consideration is a monotone decreasing function of X for each $X \epsilon [a,b]$, then $\frac{d}{dt}\left[\sigma^2(t)\right] \geq 0$ for each $t \epsilon [a,b]$.

Proof: $\frac{d}{dt}\left[\sigma^2(t)\right] = \frac{f(t)}{\varphi(a,t)} \{(t - M_{(t)})^2 - \sigma^2(t)\}$;

see Theorem 5 page 24. We have seen that

$$\sup \sigma^2(t) \leq \max\{(t - M_{(t)})^2, (a - M_{(t)})^2\} ;$$

see Lemma 1 in section 1.3 .

Therefore if f(X) is as we assumed, then

$(t - M_{(t)}) \geq -(a - M_{(t)})$ for each $t \epsilon [a,b]$ and

therefore $(t - M_{(t)})^2 \geq \sigma^2(t)$ for each $t \epsilon [a,b]$ also.

This implies that $\frac{d}{dt}\left[\sigma^2(t)\right] \geq 0$ for each $t \epsilon [a,b]$.

In general, any distribution function which is differentiable in $[a,t]$ and is such that $t-M_{(t)} \geq (M_t-a)$ for each $t \epsilon [a,b]$ will have this property of uniformly monotone increasing variance with t , $t \epsilon [a,b]$. Obviously, if the mean is nearer the right end point of $[t,b]$ for each $t < b$ which is considered, we may extend our interval to the left with assurance that the variance will increase.

Now we shall seek the necessary and sufficient conditions for the density function f(X) to have the monotone increasing variance property under right-hand extensions.

Let us define the following terms:

$$\sigma^2[a,t] = \frac{\int_a^t (X - M_{a,t})^2 dF(X)}{\varphi[a,t]} \quad ,$$

$$M_{[a,t]} = \frac{\int_a^t X dF(X)}{\varphi[a,t]} \quad , \text{ and}$$

$$\varphi[a,t] = \int_a^t dF(X) \quad ,$$

where F(X) is the distribution function of X .

Let $\quad F_T(X) = \frac{F(X) - F(a)}{F(b) - F(a)} \quad , \quad F(b) - F(a) > 0 \quad .$

We know that

$$\int_a^b (X-M_{[a,b]})^2 dF_T(X) + \int_a^b F_T(X) d(X-M_{[a,b]})^2 = (b-M_{[a,b]})^2 \quad .$$

In order for $\sigma^2[a,b]$ to be greater than or equal to $(b-M_{[a,b]})^2$ we must have

$$\int_a^b F_T(X) d(X-M_{[a,b]})^2 \geq 0 \ .$$

This is the necessary and sufficient condition for

$$\frac{d}{dt}\left[\sigma^2[a,t]\right] \geq 0 \text{ at } t = b \ .$$

If this is true for each $t \in [a,b]$ then

$$\frac{d}{dt}\left[\sigma^2[a,t]\right] \geq 0 \text{ for each } t \in [a,b] \ .$$

Let us examine this condition more closely.

If $\int_a^t F_T(X) d(X - M_{[a,t]})^2 \geq 0$ for each $t \epsilon [a,b]$, then

$\int_a^t F_T(X)\ X\ dX \geq (M_{[a,t]}) \int_a^t F_T(X)\ dX$ for each $t \epsilon [a,b]$.

Therefore if

$$(4) \qquad \frac{\int_a^t XF_T(X)\ dX}{\int_a^t F_T(X)\ dX} \geq M_{[a,t]} \text{ for each } t \epsilon [a,b] \text{ , then}$$

$\frac{d}{dt}\left[\sigma^2[a,t]\right] \geq 0$ for each $t \epsilon [a,b]$. How do we interpret this condition? The expression on the left in inequality (4) is simply the mean of $F(X)$ in $[a,t]$, $M_{F(X)[a,t]}$.

Therefore we may write $\frac{d}{dt}\left[\sigma^2[a,t]\right] \geq 0$ for each $t \epsilon [a,b]$ if and only if $M_{F(X)[a,t]} \geq M_{[a,t]}$.

It has been shown that if we choose $u(t)$ such that $M_{[u(t),t]} = M_o$ for each t under consideration, then

$$\frac{d}{dt}\left[\sigma^2[u(t),t]\right] \geq 0 \text{ and } \frac{-d}{du(t)}\left[\sigma^2[u(t),t]\right] \geq 0 \text{ .}$$

Therefore it follows that if $\frac{d}{dt}\left[\sigma^2[a,t]\right] \leq 0$, then

$\frac{-d}{da}\left[\sigma^2[a,t]\right] \geq 0$ if we consider a to be a variable. This particular statement must not be construed to imply that if one is positive the other must be negative!

Therefore we may conclude that:

1) if and only if $M_{F(X)[a,b]} \geq M_{[a,b]}$, then $\frac{d}{db}\left[\sigma^2[a,b]\right] \geq 0$;

2) if and only if $M_{F(X)[a,b]} \leq M_{[a,b]}$, then $\frac{d}{-da}\left[\sigma^2[a,b]\right] \geq 0$.

Another way to obtain this result is to look at:

$$\int_a^b (X-M_{[a,b]})^2 d\{F_T(X)-1\} + \int_a^b (F_T(X)-1)d(X-M_{[a,b]})^2 = (a - M_{[a,b]})^2 \ .$$

In order to insure $(a - M_{[a,b]})^2 \geq \sigma^2[a,b]$, we have

$$\int_a^b \{F_T(X)-1\}(X - M_{[a,b]})dX \geq 0$$

for each a under consideration. This implies that

$$\int_a^b X\{F_T(X)-1\}dX \geq M_{[a,b]} \int_a^b (F_T(X)-1)dX \ .$$

Note that $F_T(X)-1$ cannot be positive and therefore:

$$\frac{\int_a^b X(F_T(X)-1)dX}{\int_a^b (F_T(X)-1)dX} \leq M_{[a,b]} \ .$$

The expression on the left is $M_{F(X)[a,b]}$. We have that

$\frac{-d}{dt}\left[\sigma^2[t,b]\right] \geq 0$ if and only if $M_{F(X)_{[t,b]}} \leq M_{[t,b]}$ for each t under consideration. This is condition 2) above.

In more crude terms we may say that if we are considering absolutely continuous distribution functions and decrease our variance with a "slight" increase of the upper end point then we do increase our variance if we decrease the left end point "slightly." This situation is governed more accurately and totally by statements 1) and 2) above.

Let us now consider the changes in variance as we extend both end points. Let $u = \Psi(t)$, where t is the upper end point and u is the lower. Let the original interval be $[a,b]$. We shall consider $\sigma^2[u,t]$, the variance on $[u,t]$, where $[a,b] \subset [u,t]$. If $u = a$ for each $t \geq b$ then $\frac{du}{dt} = 0$, otherwise we shall consider only those cases where

$$\frac{du}{dt} = u' < 0 .$$

We see that $$\frac{d}{dt}\left[\sigma^2[u,t]\right] =$$

$$\frac{1}{A(t)}\left[f(t)\{(t-M)^2 - \sigma^2[u,t]\} - u'f(u)\{(u-M)^2 - \sigma^2[u,t]\}\right] ,$$

where $A(t) = \int_u^t dF(X)$ and $M = M_{[u,t]}$.

Recall: $\int_u^t (X-M)^2 dF_T(X) + \int_u^t F_T(X)d(X-M)^2 = (t-M)^2$ and

$$\int_u^t (X-M)^2 d\{F_T(X)-1\} + \int_u^t \{F_T(X)-1\} d(X-M)^2 = (u-M)^2 ,$$

$$\text{where} \quad F_T(X) = \frac{F(X) - F(a)}{F(t) - F(a)} .$$

Multiply the first equation by $f(t)$ and the second by $-u'f(u)$. Add the results. This yields: $\frac{d}{dt}\left[\sigma^2[u,t]\right] \geq 0$

if and only if

$$f(t) \int_u^t F_T(X) d(X-M)^2 - u'f(u) \int_u^t \{F_T(X)-1\} d(X-M)^2 \geq 0 .$$

In fact, $\frac{d}{dt}\left[\sigma^2[u,t]\right]$ is $\{1/A(t)\}$ times the expression on the left. In another form we have that the derivative is not negative if and only if:

$$\{f(t) - u'f(u)\} \int_u^t F_T(X) d(X-M)^2 \geq u'f(u)\{(t-M)^2 - (u-M)^2\} .$$

The most useful case is probably that in which $u' = -1$, (i.e. an equal numerical change of endpoints).

NOTE: If the distribution is symmetric about M_o for $[a,b]$, then the right side is zero and the left side is positive ($M_F > M$) unless $f(t)$ and $f(u)$ are both zero; in which case we know $\frac{d}{dt}\left[\sigma^2[u,t]\right]$ is zero. This is consistent with our previous results, for $(t-M)^2 = (u-M)^2 \geq (t-u)^2/4$.

We discussed the proposition, $V(Y) \geq V(Y|X\epsilon I) \geq V(Y|X)$, for the case of a linear model where $V(Y|X)$ was independent of X. We proved that both $V(Y)$ and $V(Y|X\epsilon I)$ were greater than $V(Y|X)$. Assuming the linear model we prove two theorems.

<u>Theorem D</u>. If $V(Y|X\epsilon I_j)$ is monotone increasing for each arbitrary nested increasing truncation, then

$$V(Y) \geq V(Y|X\epsilon I_j) \text{ for each } j = 0,1,\ldots \quad .$$

Proof: Let $I = (-\infty,\infty)$. Then $V(Y|X\epsilon I) = V(Y)$ as was seen in section 1.2 . Therefore if $V(Y|X\epsilon I_j)$ is monotone in the sense we have discussed, then

$$V(Y|X\epsilon I) = V(Y) \geq V(Y|X\epsilon I_j) \text{ for each } I_j \subset I \ .$$

<u>Theorem E</u>. If $V(Y|X\epsilon I_j)$ is not monotone increasing for each arbitrary nested increasing truncation, then $V(Y)$ is not necessarily greater than or equal to $V(Y|X\epsilon I_j)$ for each j , where $I_o \subset I_1 \subset I_2 \subset\ldots\subset I_n\ldots\subset I = (-\infty,\infty)$.

Proof: Consider a density function $f(Y,X)$ which is such that: $V(Y) = K$ and

$$k = V(Y|X) < V(Y|X\epsilon I_1) < V(Y|X\epsilon I_2) \ldots$$

$$< V(Y|X\epsilon I_n) > \left[V(Y|X\epsilon I_{n+1})\right] =$$

$$K' < V(Y|X\epsilon I_{n+2}) < \ldots < V(Y) = K \ .$$

If $V(Y) = V(Y|X\epsilon I_{n+1})$ the theorem holds. If not, choose some $f_1(Y',X') = cf(Y,X)$, $X\epsilon I_{n+1}$; $f_1(Y',X') = 0$ otherwise. Then

$$V(Y') = K' > V(Y'|X'\epsilon I_n) \equiv V(Y|X\epsilon I_n) \ .$$

Therefore if we have a density which is such that $V(Y|X\epsilon I_j)$ is not monotone we can find a density which is such that $V(Y') < V(Y'|X'\epsilon I)$ for some I .

We have shown that the proposition: $V(Y) \geq V(Y|X\epsilon I)$ may not hold unless we have the monotonicity of variance property with nested increasing truncations. We recall that $V(Y|X\epsilon I) \geq V(Y|X)$ in all cases under our assumptions.

We might ask what good are the conditions we presented on page 60. They may be used in some cases to give definitive results. We shall present one such example.

Let $f(X) = e^{X-t}$, $X\epsilon(-\infty,t]$. Let $u' = 0$. Then we have that $F(X) = f(X) = e^{X-t}$.

Therefore $M_{F(X)_{(-\infty,t]}} \equiv M_{(-\infty,t]}$. Therefore, $\sigma^2(-\infty,t]$ is constant for each $t < \infty$. This may be verified by noting that

$$\frac{d}{dt} \int_{-\infty}^{t} (X-M)^2 e^{X-t} dX = 0 \ .$$

These conditions may be applied to the negative exponential in the same way.

In the case of a normal variable the necessary and sufficient conditions presented in this appendix are not appropriate. Consider a normal variate with mean zero and variance one. It can be shown, and is shown by Clark {3}, that if one truncates this distribution to the interval $[a,\infty)$, then $\sigma^2[a,\infty) = 1 - \mu_{[a,\infty)}(\mu_{[a,\infty)} - a)$.

This may be obtained from the expression:

$$\sigma^2[a,b] = 1 - \mu^2_{[a,b]} + \frac{af(a) - bf(b)}{F(b) - F(a)} .$$

Since $\mu_{[a,\infty)} - a \geq 0$ and $\mu_{[a,\infty)}$ is monotone increasing with a, we see that $\sigma^2[a,\infty)$ is monotone decreasing with a, for each $a < \infty$. By symmetry, we see that this holds for right extensions of $(-\infty,a]$ also.

The author has not been able to prove that arbitrary extensions of intervals imply an increase in variance in the case of the normal distribution, although the table on page 49 is a good indication that the normal distribution does have this property.

Every attempt by the author to classify, in generality, the set of $f(X)$ for which right extensions imply monotonic variances has failed.

This is a very interesting question, but much more work can be done to improve our knowledge about this property and determine its implications.

V BIBLIOGRAPHY

{1} Anderson, T.W. An Introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons, Inc. 1958.

{2} Birnbaum, Z.W., "Effect of Linear Truncation on a Multi-normal Population", Annals of Mathematical Statistics, Vol. 21 (1950), pp. 272-279.

{3} Clark, F.E., "Truncation To Meet Requirements On Means", Journal of the American Statistical Association, Vol. 52 (December, 1957), pp. 527-536.

{4} Cohen, A.C., Jr., "Estimating the Mean and Variance of Normal Populations From Singly Truncated and Doubly Truncated Samples", Annals of Mathematical Statistics, Vol. 21 (1950), pp. 557-569.

{5} Cohen, A.C., Jr., "Restriction and Selection in Samples From Bivariate Normal Distributions", Journal of the American Statistical Association, Vol. 50 (September, 1955), pp. 884-893.

{6} Feller, W. An Introduction to Probability Theory and Its Applications (second edition). New York: John Wiley and Sons, Inc. 1957, Vol. I.

{7} Graybill, F.A. An Introduction to Linear Statistical Models. New York: McGraw-Hill Book Company, Inc. 1961, Vol. I.

{8} Kendall, M.G., and Stuart, A. The Advanced Theory of Statistics. New York: Hafner Publishing Company, 1963, Vols. I (second edition) and II.

{9} Kolmogorov, A.N. Foundations of the Theory of Probability (second edition). New York: Chelsea Publishing Company, 1956.

{10} Mood, A.M., and Graybill, F.A. Introduction to the Theory of Statistics (second edition). New York: McGraw-Hill Book Company, Inc. 1963.

{11} Parzen, E. Modern Probability Theory and Its Applications. New York: John Wiley and Sons, Inc. 1960.

{12} Parzen, E. Stochastic Processes. San Francisco: Holden-Day, Inc. 1962.

{13} Widder, D.V. Advanced Calculus (second edition). Englewood Cliffs, N.J.: Prentice-Hall, Inc. 1961.

The vita has been removed from the scanned document

ABSTRACT

The subject treated in this thesis is the conditional distribution of a random variable given that the outcome of an associated random variable lies within a specified interval. This may be considered to be an extension of the classical case in which the outcome of the associated random variable is known to assume a specific numerical value.

The primary purpose of the study was to examine the properties of a system formed by interval conditioning under the assumption of a suitable linear model. No attention was given to appropriate estimation procedures.

The principal conclusions of the study follow. Let X and Y be jointly distributed random variables such that $E(Y|X) = \alpha + \beta X$, where α and β are constants, and such that the variance of Y given X is independent of X. Then

$$E(Y|X\epsilon I) = \alpha + \beta\ E(X|X\epsilon I)$$

and the variance of Y given $X\epsilon I$ is equal to the variance of Y given X plus β^2 times the variance of X in its truncated distribution, i.e. truncated in the conditioning interval I.

It was shown that the limiting cases of the system led to the classical conditional results as the conditioning interval degenerates to a point, and to the classical marginal results as the interval expands to encompass the real line. These results were generalized into the case

where a random variable Y is conditioned on a set of associated variables, $\{X_i\}_{i=1}^{p}$, such that $X_i \epsilon I_i$, $i = 1,2,\ldots p$.

Higher conditional moments were found in general. Since third and higher conditional moments are usually functions of the conditioned variables, only an analytic form was given.

Consideration was given to the case in which a vector of random variables is to be predicted given that an associated vector of random variables lies in a specified rectangular region. Two types of conditioning were considered simultaneously at this point, namely, the case in which part of the associated variables are conditioned to points and the remainder to intervals.

In various places in the body of the thesis and in the appendix consideration was given to the conditions under which the variance of a truncated random variable increases monotonically with the interval of truncation. This was found to be a complicated problem, but necessary and sufficient conditions for this property were developed in the appendix.