STATISTICAL MEMOIRS
Institute of Mathematics
University of Oslo

No 1
January 1972

SUPPLEMENTARY NOTES ON LINEAR MODELS

by

Erik N. Torgersen

Content.

MATEMATISK INSTITUTT

Blindern, 22/10-71.

Varians analyse
Høsten 1971
Torgersen

I. The Lehmann - Scheffé theorem on minimum variance unbiased estimation.

In their joint 1950 paper (Sankhya Ser A Vol 10 p.324) Lehmann and Scheffé gave a criterion for "uniformly minimum variance unbiasedness" of estimators. The criterion has many simple generalizations, most of which may be found in Rao's book (Linear Statistical...) or in Zack's book (The theory of Statistical...) We shall here consider the obvious generalization to the situation where we choose to restrict ourselves to estimators belonging to some linear space L of everywhere square integrable random variables.

Consider an experiment $((\chi, \mathcal{A}), P_\theta : \theta \in \Theta)$ a real valued function g on Θ and a linear space of everywhere square integrable random variables. The property of beeing minimum variance unbiased within L will be shortened* UMVU(L). Formally a δ in L will be called UMVU(L) if and only if it is unbiased and $\text{var}_\theta \delta \leq \text{var}_\theta \gamma$; $\theta \in \Theta$ for any other unbiased estimor γ in L.

The Lehmann Scheffé theorem extends immediately to this situation.

<u>Theorem.</u> Let $\delta \in L$ be an unbiased estimator of g. Then δ is UMVU(L) if and only if δ is uncorrelated with every unbiased estimator of zero which is in L. A UMVU(L) estimator for g is unique up to equivalence.

* If L is the space of all everywhere square integrable random variables then we write UMVU instead of UMVU(L).

Remark: The Lehmann Scheffé theorem is the particular case where L consists of all square integrable random variables.

Proof of the theorem: The only difference between this proof and that of Lehmann Scheffé is that the restriction "$\in L$" is inserted in the natural places.

1^o Suppose δ is UMVU(L) for g and let $\phi \in L$ be an unbiased estimator of 0. Then $\delta + \lambda\phi$ is - for each $\lambda \in]-\infty,\infty[$ - an unbiased estimator of g, and it belongs to L. It follows that-for each θ- the polynomial $\text{var}_\theta (\delta+\lambda\phi) = (\text{var}_\theta \phi)\lambda^2 + 2 \text{cov}_\theta(\delta,\phi)\lambda + \text{var}\,\delta$ has a minimum for $\lambda=0$, and this implies $\text{cov}_\theta(\delta,\phi) = 0$.

2^o Suppose δ is uncorrelated with any unbiased estimator of zero which is in L. Let $\gamma \in L$ be another unbiased estimator of g. Then -by assumption $\text{cov}_\theta (\delta,\gamma-\delta) \underset{\theta}{\equiv} 0$ so that $\text{var}_\theta \gamma = \text{var}\{\delta + (\gamma - \delta)\} = \text{var}_\theta \delta + \text{var}_\theta (\gamma - \delta) \geqq \text{var}_\theta \delta$; $\theta \in \Theta$

3^o Finally let δ γ both be UMVU(L) estimators of g.

Then: $E_\theta(\delta(\delta-\gamma) \underset{\theta}{\equiv} 0$ and $E_\theta\gamma(\delta-\gamma) \underset{\theta}{\equiv} 0$

Hence $E_\theta\delta^2 \underset{\theta}{\equiv} E_\theta\delta\gamma \underset{\theta}{\equiv} E_\theta\gamma^2$

So that
$$E_\theta(\delta-\gamma)^2 \underset{\theta}{\equiv} E_\theta\delta^2 - 2E_\theta\delta\gamma + E_\theta\gamma^2 \underset{\theta}{\equiv} 0$$

Hence $P_\theta(\delta=\gamma) = 1$; $\theta \in \Theta$ □

Corollary:
Let $g_1, g_2, \ldots g_r$ be r real valued functions on Θ and let $c_1, c_2, \ldots c_r$ be r constants. Suppose δ_i is a UMVU(L) estimator of g_i ; $i=1,2,\ldots r$. Then $\Sigma c_i\delta_i$ is a UMVU(L) estimator of $\Sigma c_i g_i$.

Proof: Let $\phi \in L$ be an unbiased estimator of 0 with everywhere finite variance.

Then: $\mathrm{cov}_\theta(\Sigma c_i \delta_i, \phi) = \Sigma c_i \mathrm{cov}_\theta(\delta_i, \theta) = 0; \ \theta \in \Theta$. □

Finally let us consider the problem of estimating vector valued functions. Let $g = (g_1, g_2, \ldots g_r)$ be a function from Θ to R^r and let $\delta = (\delta_1, \delta_2, \ldots, \delta_r) \in L^r$ be an unbiased estimator of g. Then δ will be called a UMVU(L) estimator of g if and only if $\ddagger_\theta \delta \leqq \ddagger_\theta \tilde{\delta}$; $\theta \in \Theta$ for any other unbiased estimator $\tilde{\delta}$ in L^r. The case of vector valued estimands may be reduced to the case of real valued estimands by

<u>Proposition</u>. Let $\delta = (\delta, \ldots \delta_r) \in L^r$ be an unbiased estimator of g where $g = (g_1, g_2, \ldots g_r)$ is a function from Θ to R^r. Then δ is UMVU(L) for δ if and only if δ_i is a UMVU(L) estimator of g_i; $i=1,2 \ldots r$.

Proof: "if": Suppose the condition is satisfied and let $\tilde{\delta} \in L^r$ be another unbiased estimator of g. Let c be a given $r \times 1$ matrix. By assumption: $c' \ddagger_\theta \delta c = \mathrm{var}_\theta \ \Sigma c_i \delta_i \leqq$ (by the corollary) $\mathrm{var}_\theta \Sigma \ c_i \tilde{\delta}_i = c' \ddagger_\theta \tilde{\delta} c$.

"only if": Suppose δ is UMVU(L). Let $\tilde{\delta}_i \in L$ be an unbiased estimator of g_i. Extend $\tilde{\delta}_i$ to an unbiased estimator $\tilde{\delta} \in L^r$ of g. The inequality $\ddagger_\theta \tilde{\delta} \geqq \ddagger_\theta \delta$ implies $\mathrm{var}_\theta \tilde{\delta} \geqq \mathrm{var}_\theta \delta$. □

A function g with domain Θ (and any range space) will be called identifiable if g is a function of θ via P_θ.

Equivalently; g is identifiable if and only if $g(\theta_1) = g(\theta_2)$ when $P_{\theta_1} = P_{\theta_2}$ Let g be R^r valued. Trivially any g having an unbiased estimator is identifiable. On the other hand, any function of θ is identifiable when $\theta \mapsto P_\theta$ is 1-1, and this does positively not exclude the possibility that only the constants have UMVU estimators.

MATEMATISK INSTITUTT

Blindern,

Varians analyse
Høsten 1971
Torgersen

Linear models

We shall in this section consider the situation where our vector* $Y' = (Y_1,\dots,Y_n)'$ of real valued random variables satisfies the following requirements

C_{lin}: (The linearily condition) There is a sub space V of R^n so that EY à priori is, and may be any vector in V.

C_{err}: (This is a condition on the "error" Y - EY) $\mathcal{D}Y = \sigma^2 I$ -where I is the nxn identity matrix and $\sigma > 0$ is more or less unknown.

Let $[a^{(1)}, a^{(2)},\dots, a^{(p)}]$ be a basis for V i.e. the vectors $a^{(1)}, a^{(2)},\dots a^{(p)}$ spans V. Denote by A' the nxp matrix whose i - th column is $a^{(i)}$ ie $A' = (a^{(1)}, a^{(2)},\dots a^{(p)})$.

We will express the projection π on V by A. Note first - since $\pi(y)\perp a^{(i)}, i = 1,\dots,p$ - that $A(\pi(y)-y) = 0$, ie $A\pi(y) = Ay$. Hence $\pi(y)$ may be written $\pi(y) = A'b(y)$ where the px1 matrix b(y) satisfy:

$$AA'b(y) = Ay$$

* We use, throughout, the convention that a vector in some space R^m is - when it is considered as a matrix - a column matrix.

These equations are called the normal equations.

Conversely - let $b(y)$ be a solution of the normal equations. Decomposing $y = A'b(y) + (y-A'b(y))$ we see that $A'b(y) \in V$ while $y-A'b(y) \perp V$. Hence $\pi(y) = A'b(y)$. We have proved:

Proposition.

The projection $\pi(y)$ of a vector $y \in R^n$ on V may be written $A'b(y)$ where the $p \times 1$ matrix $b(y)$ is and may be any solution of the normal equations $AA'b(y) = Ay$.

Let L be the space of linear functions of $Y_1, Y_2, \ldots Y_n$ i e $\delta \in L$ if and only if there are real constants $d_1, d_2, \ldots d_n$ so that $\delta = \sum_{i=1}^{n} d_i Y_i$. L is - for any fixed distribution of $Y - EY$ satisfying C_{err} - an inner product space if we consider $Cov(\Sigma b_i Y_i, \Sigma d_i Y_i) = (\Sigma b_i d_i)\sigma^2$ as the inner product of $\Sigma b_i Y_i$ and $\Sigma d_i Y_i$. The sub space L_V of L consisting of all $\Sigma d_i Y_i$ where $d' = (d_1 \ldots, d_n)' \in V$ will be called the estimation space, and the sub space $L_{V\perp}$ of all $\sum_i d_i Y$ where $d \perp V$ will be called the error space.

Clearly:

$$L = L_V \oplus L_{V\perp}$$

The projection on L_V maps $d'Y$ onto $\pi(d)'Y$.

The justifications for the terms estimation space and error space are:

Proposition

L_V is the space of UMVU(L) estimators and $L_{V\perp}$ is the space of unbiased estimators of 0 which belong to L.

Proof: 1^0 $E < d, Y > \equiv 0 \iff d \in V^\perp$

2° $\langle d,Y\rangle$ is a UMVU(L) estimator (of its expectation)
$\Leftrightarrow \operatorname{Cov}(\langle d,Y\rangle,\langle b,Y\rangle) = \langle b,d\rangle \sigma^2 = 0$ when $b \in V^{\perp}$
$\Leftrightarrow d \in (V^{\perp})^{\perp} = V.$

Here is the basic result on UMVU(L) estimation under C_{lin} and C_{err}:

<u>Theorem</u>[x] Consider the representation $EY = A'\beta$ of EY and let $c \in R^p$. The following conditions on the estimand $\psi: \beta \rightsquigarrow c'\beta$ are equivalent.

(i) ψ is identifiable

(ii) ψ has an unbiased estimator in L.

(iii) ψ has a UMVU(L) estimator.

(iv) $c \in [A']_{row}$

If $\hat{\psi}$ is a UMVU(L) estimator of ψ, then it is unique and:

1) $\hat{\psi}(Y) = \pi(d)'Y$ for any unbiased estimator $d'Y$ of ψ

2) $\hat{\psi}(Y) = c'b(Y)$ where $b(Y)$ is any solution of the normal equations $AA'b(Y) = AY$.

Proof:

(i) $\Leftrightarrow$ (iv): ψ is identifiable $\Leftrightarrow \psi(\beta_1) = \psi(\beta_2)$ when $A'\beta_1 = A'\beta_2 \Leftrightarrow \psi(\beta) = 0$ when $A'\beta = 0 \Leftrightarrow \beta \perp c$ when $\beta \perp [A']_{row} \Leftrightarrow c \in A'_{row}$.

(ii) $\Leftrightarrow$ (iv): $\exists d$ so that $Ed'Y \equiv c'\beta \Leftrightarrow \exists d$ so that $d'A'\beta \equiv c'\beta \Leftrightarrow \exists d$ so that $d'A' = c' \Leftrightarrow c \in [A']_{row}$.

x: If M is any matrix, then M_{row} and M_{col} denotes, respectively the space spanned by the row vectors of M and the space spanned by the column vectors of M.

(ii) $\Leftrightarrow$ (iii): $\exists d$ so that $Ed'y \equiv \psi(\beta)$ $\Leftrightarrow$ (by the previous proposition) $\exists d \in V$ so that $Ed'y \equiv \psi(\beta)$ $\Leftrightarrow$ (by the previous proposition) (iii).

Let $\langle d,y \rangle$ and $\langle b,y \rangle$ both be UMVU(L) estimators of ψ then $\langle d-b,y \rangle$ is a UMVU(L) estimator of 0 ie: $d-b \in V \cap V^{\perp} = 0$. Suppose ψ is identifiable and let ϕ be the UMVU(L) estimator. Let $\langle d,y \rangle$ be any unbiased estimator of ψ. It follows directly from the previous proposition that $\pi(d)'Y$ is the UMVU(L) estimator of ψ so that: $\phi(y) = \pi(d)'Y$. Let $y \in R^n$ and suppose $AA'b(y) = Ay$. Then - as we have seen - $\pi(y) = A'b(y)$ and $\phi(y) = \pi(d)'[\pi(y) + y - \pi(y)] = \pi(d)'\pi(y) = \pi(d)'A'b(y) = c'b(y)$ since - by unbiasedness - $\pi(d)'A' = c'$.

The matrix A is -in general- not of maximal rank. If rank $A < p$ then the solutions β of the equation $EY = A'\beta$ - for given EY - fill up a infinite affine space. Uniqueness can only be obtained by imposing conditions on the solutions. A result in this direction is:

<u>Theorem.</u>

Let W be a sub space of R^p, having the property that to each $\beta \in R^p$ there is at least one $\tilde{\beta} \in W$ so that $A'\beta = A'\tilde{\beta}$.

Then the normal equations

$$AA'b = Ay$$

has - for given $y \in R^n$ - at least one solution $b \in W$.

If - in addition - the correspondence $\beta \to \tilde{\beta}$ is single valued - then $\tilde{\beta}$ is a linear identifiable (and consequently has a unique UMVU(L) estimator) function of β.

Proof:

1° if $AA'b = Ay$ then - since $A'b = A'\tilde{b}$ - $AA'\tilde{b} = Ay$.

2° Suppose $\beta \mapsto \tilde{\beta}$ is single valued. Let $\beta^{(1)}$,

$\beta^{(2)} \in R^p$ be such that $A'\beta^{(1)} = A'\beta^{(2)}$. Then $A'\tilde{\beta}^{(1)} = A'\tilde{\beta}^{(2)}$. By uniqueness: $\tilde{\beta}^{(2)} = \tilde{\beta}^{(1)}$. It follows that $\tilde{\beta}$ is identifiable. Let $k_1, k_2 \in R$ and let $\beta^{(1)}, \beta^{(2)} \in R^p$. Then $A'(k_1\beta^{(1)} + k_2\beta^{(2)}) = k_1A'\beta^{(1)} + k_2A'\beta^{(2)} = k_1A'\tilde{\beta}^{(1)} + k_2A'\tilde{\beta}^{(2)} = A'(k_1\tilde{\beta}^{(1)} + k_2\tilde{\beta}^{(2)})$. By uniqueness again: $\widetilde{k_1\beta^{(1)} + k_2\beta^{(2)}} = k_1\tilde{\beta}^{(1)} + k_2\tilde{\beta}^{(2)}$ i e the map $\beta \mapsto \tilde{\beta}$ is linear. The remark in the parantheses follows from the previous theorem.

MATEMATISK INSTITUTT

Blindern, 24.11.1971.

Variansanalyse
Høsten 1971
Torgersen

Linear models with normally distributed variables. Sufficiency, minimal sufficiency and completeness.

In addition to our assumptions C_{lin} and C_{err}, we will in this section assume:

C_{norm}: Y is multinormally distributed.

An immediate and fundamental consequence of C_{norm} is:

Theorem. $Y_1, \ldots, Y_n$ are independently and normally distributed. Any finite dimensional random vector with coordinates in L is multinormally distributed.

This theorem and the fact that uncorrelated coordinates of multinormally distributed random variables are stochastically independent yield:

Theorem. The estimation space and the error space are stochastically independent. In particular $\pi(Y)$ and $Y-\pi(Y)$ are independent random vectors.

The main concern, in this section, will be with sufficiency, minimal sufficiency and completeness. In this section - as in the previous section - the condition C_{lin} may, occasionally, be weakened. Some weakenings of C_{lin} will here be explicitly stated. The reader should, however, go over the results in the previous section with this point in view.

It will be necessary to distinguish two dimension concepts. Let A be any non empty sub set of a finite dimensional vector space. The largest number k with the property that there are k linearily independent vectors in A will be denoted by dim A. This is, of course, nothing but the dimension of the vector space spanned by A. The affine (or geometrical) dimension of A - written dimg A - is the dimension of the affine space generated by A. For any $a \in A$, the set $A-a$ generates the same vector space as as $A-A$. This vector space is the vector space part of the affine space generated by A. dimg A is the dimension of this vector space. dimg A = dim A or dim $A-1$ as $0 \in$ the affine space generated by A or not.

The basic result on sufficiency is:

<u>Theorem:</u> $(\pi(Y), \|Y-\pi(Y)\|)$ is sufficient. It is minimal sufficient if and only if dimg $\{(EY/\sigma^2, 1/\sigma^2)\}$ = dim $V+1$.

Let σ^2 be fixed. Then $\pi(Y)$ is sufficient. It is minimal sufficient if and only if the affine dimension of the set of vectors EY which are à priori compatible with the chosen σ^2, is dim V.

If EY is fixed, then $\|Y-EY\|$ is sufficient and it is minimal sufficient if and only if there is at least two possible values of σ^2 for the chosen EY.

Proof: The joint density of $Y_1, \ldots, Y_n$ may be written $(2\pi\sigma^2)^{-n/2}$ $\exp(-\|y-EY\|^2/2\sigma^2) = (2\pi\sigma^2)^{-n/2}$ $[\exp(-\|\pi(y) - EY\|^2/2\sigma^2)]$ $[\exp(-\|y - \pi(y)\|^2/2\sigma^2)$. The statements on sufficiency follows directly from the factorization criterion for sufficiency.

Denote EY by η and let P_{η,σ^2} be the probability distribution of Y. We use the sign :: for "induce the same set of events as". Let (η^o,σ_o^2) be a possible pair. Then $\{dP_{\eta,\sigma^2}/dP_{\eta^o,\sigma_o^2}\} :: \{y \longrightarrow \langle \eta/\sigma^2-\eta^o/\sigma_o^2,\pi(y)\rangle+(\sigma^{-2}-\sigma_o^{-2})\|y\|^2/2\}$

At this point we may apply a theorem of Halmos and Savage which states that pairwise sufficiency and sufficiency is the same for dominated experiments. Moreover - since $\{P_{\eta,\sigma^2}\}$ is homogeneous- we may in this case restrict our attention to the set of all possible pairs $(P_{\eta,\sigma^2},P_{\eta^o,\sigma_o^2})$.

It follows that

$\{dP_{\eta,\sigma^2}/dP_{\eta^o,\sigma_o^2}\} :: \{y \longrightarrow \pi(y), y \longrightarrow \|y\|\} :: \{y \longrightarrow \pi(y), y \longrightarrow \|y-\pi(y)\|\}$ when $\operatorname{dimg}\{(\eta/\sigma^2,1/\sigma^2)\} = \dim V+1$. If $\operatorname{dimg}\{(\eta/\sigma^2,1/\sigma^2)\} \leq \dim V$, then - by the same result - $\pi(Y)$ and $\|Y-\pi(Y)\|$ is no longer minimal sufficient.

The proofs of the other statements on minimal sufficiency are similar but simpler - and is therefore omitted. □

Finally we have the following result on completeness.

Theorem. $(\pi(Y),\|Y-\pi(Y)\|)$ is complete provided the interior of $\{(EY,\sigma^2)\}$ relative to $V\times R$ is non empty.

Let σ^2 be fixed. Then $\pi(Y)$ is complete provided the interior relative to V of the set of vectors which are compatible with the chosen σ^2 is non empty.

If EY is fixed then $\|Y-EY\|$ is complete provided the set of numbers σ^2 which are compatible with the chosen EY contains a right sided accumulation point.

Remark. If D is a set of real numbers then a right sided accumulation point x for D is a real real number x such that

$]x,x+\varepsilon[\cap D \neq \emptyset$ when $\varepsilon>0$.

Proof: Write $\eta = EY$. The joint density of Y may be written:

$(\ )[\exp(\|y\|^2/\sigma^2 - 2<\eta/\sigma^2, y>](\ )$ where the first $(\ )$ is a function of (η,σ^2) and the last $(\ )$ is a function of y only. If $(\rho_1,\rho_2,\ldots\rho_r)$ is a linearily independent basis for V then $<\eta/\sigma^2, y> = \sum_{i=1}^{r} \zeta_i<\rho_i/\sigma^2, y>$ where $\zeta_1\ldots,\zeta_r$ are the $(\rho_1,\ldots,\rho_r)$ coordinates of η. It is easily seen that $\{(\zeta_1/\sigma^2,\ldots,\zeta_r/\sigma^2,1/\sigma^2)\}$ has interior points - relative to R^{r+1} - if and only if $\{(\eta,\sigma^2)\}$ has interior points relative to V×R. Completeness follows now from the basic completeness theorem for exponential experiments.

The proof of the second statement is similar but simpler and is therefore omitted.

Write $Z = \|Y-EY\|^2$. Then Z/σ^2 has a χ^2 distribution with n degrees of freedom. It follows that the density of Z may written

(∗) $\quad$ constant $(\sigma^2)^{-n/2} z^{n/2-1} \exp(-z/2\sigma^2)$

Suppose $E_{\sigma^2}h(Z) = 0$ for all à priori possible values on σ^2. h is integrable w.r.t. the density (∗) - for any $\sigma^2>0$- if and only if $I(\sigma^2) = \int_0^\infty |h(z)|z^{n/2-1} \exp(-z/2\sigma^2)\,dz < \infty$.

Hence - since $I(\sigma^2)$ is a non decreasing function of σ^2- the set of numbers $\sigma^2>0$ so that $I(\sigma^2)<\infty$ is an interval of the form $]0,t)$ Moreover the map $\sigma^2 \dashrightarrow E_{\sigma^2}h(Z)$ from $]0,t[$ to R is analytic. It follows from indentity theorem for analytic functions that $I(\sigma^2) = 0$ for all $\sigma^2 \in]0,t[$. Hence - by the basic completeness theorem for exponential experiments - $h = 0$ a.e. □

MATEMATISK INSTITUTT

Blindern, 14.12.71.

Varians analyse,
Høsten 1971,
Torgersen.

Stochastic ordering of probability distribution and monotonically increasing likelihood ratios.

A probability distribution F on R is called stochastically larger then the probability distribution G on R if and only if $F([x,\infty[) \geq G([x,\infty[)$ for all $x \in R$.

Example.

Let the real random variables X and Y be, respectively, normal $(\xi,1)$ and normal $(\eta,1)$. Then the distribution of X^2 is stochastically larger then the distribution of Y^2 provided $\xi^2 \geq \eta^2$. Demonstration: Let $a \geq 0$ We must show that $P(X^2 \geq a^2) \geq P(Y^2 \geq a^2)$ or - equivalently - that $P(|X| \leq a) \leq P(|Y| \leq a)$ If Φ denotes the normal $(0,1)$ distribution then this inequality may be written:

$$\Phi[-\xi-a,-\xi+a] \leq \Phi[-\eta-a,-\eta+a]$$

or - using the symmetry of Φ:

$$\Phi[|\xi|-a,|\xi|+a] \leq \Phi[|\eta|-a,|\eta|+a]$$

It suffices therefore to show that $\Phi[x-a,x+a]$ is a decreasing function of $x \in [0,\infty[$.

The derivative of this function, however, is $\Phi'(x+a)-\Phi'(x-a) = \Phi'(x+a) - \Phi'(|x-a|) \leq 0$ since $x+a \geq |x-a|$.

The only properties of the normal (0,1) distribution we used was unimodality and symmetry Let G be any symmetrid unimodal -ie $x \mapsto G\,]-\infty, x[$ is convex on $]-\infty, 0[$ and concave on $]0,\infty[$ - distribution. Suppose U is a real random variable having the distribution G. Then the distribution of $(\xi + V)^2$ is stochastically larger than the distribution of $(\eta + V)^2$ provided $\xi^2 \geqq \eta^2$.

Let F be any probability distribution on R. Then the (lower) fractile function of F - notation F^{-1} - is the function from $]0,1[$ to R which maps p into $F^{-1}(p) = \inf \{x: F]-\infty, x[\geqq p\}$

The following result is often useful:

Proposition

If U is uniformly distributed on $]0,1[$ then $F^{-1}(U)$ has distribution F.

Proof:

$$F\,]-\infty,x[\;=\; P(U \leqq F]-\infty,x[) \begin{cases} \leqq P\,(F^{-1}(U) \leqq x) \\ \geqq P\,(F^{-1}(U) < x) \end{cases}$$

Stochastic ordering of probability distributions is related to the ordering of random variables through:

Proposition.

If X and Y are real random variables such that $X \geqq Y$ a.s. then the distribution of X is stochastically larger than the distribution of Y.

Conversely; let the distribution F be stochastically larger than the distribution G. Then there are real random variables X and Y on some probability space having, respectively, distributions F and G such that $X \geqq Y$.

Proof:

Only the last statement needs a proof. Suppose F is stochastically larger than G. Then $F^{-1} \geq G^{-1}$ Let U be uniformly distributed on $]0,1[$. Then $X = F^{-1}(U)$ and $Y = G^{-1}(U)$ have the desired properties. $\square$

Many properties of the stochastic ordering of distribution functions are simple consequences of the last proposition. We restrict ourselves to the following two propositions.

Proposition.

Let F be stochastically larger then F. Then $\int h dF \geq \int h dF$ for any monotonically increasing function h such that both integrals exists.

Remark.

A converse may be obtained by considering the indicator functions h of intervals of the form $[x,\infty[$.

Proposition.

If F_i is stochastically larger than G_i, $i = 1,2,...n$, then $F_1 * ... * F_n$ is stochastically larger than $G_1 * ... * G_n$.

A much stronger property that "beeing stochastically larger" is the property of having monotonically increasing likelihood ratio. Suppose F and G are given by densities f and g respectively w.r.t. some measure μ on R (ie on the class of Borel sub sets of R) R). Then we shall say that G has monotonically increasing likelihood ratio w.r.t. F if it is possible to specify μ, f and g so that g/f is monotonically increasing (on its domain which is $[f > 0] \cup [g > 0]$) . It is easily seen that if G has monotonically increasing likelihood ratio w.r.t. F, and ν is any measure such that F and G have densities f and g, respectively, w.r.t. ν, then f and g may be specified so that g/f is monotonically increasing. In particular we may use $\nu = F+G$ and choose f and g so that, $f \geqq 0$, $g \geqq o$, $f + g = 1$ and g is monotonically increasing.

Proposition

Let G have monotonically increasing likelihood ratio w.r.t. F. Then any test δ of the form

$$\delta = \gamma I_{\{c\}} + I_{]c,\infty[}$$

is a most powerful level $\alpha = \int \delta dF$ test for testing F against G provided $\alpha > 0$.

Proof:

Let $f = dF/d(F+G)$, $g = d\sigma/d(F+G) = 1-f$ We may choose g so that it is monotonically increasing and $0 \leqq g \leqq 1$. Then $f(c)>0$ (if otherwise, then $\alpha = 0$). Let ϕ be another level α test. Then: $\int \delta dG - \int \phi dG = \int (\delta-\phi)\ (g-(g(c)/f(c))f)d(F+G)+(g(c)/f(c))\int(\delta-\phi)fd(F+G)$.

$\therefore$ The first term on the right is $\geqq 0$ since the integrand is $\geqq 0$, and the last term is $\geqq 0$ since $\int \phi dF \leqq \alpha$. Hence $\int \delta dG \geqq \int \phi dG$.

Problem

Does this proposition have a converse?

Proposition.

Let G have monotonically increasing likelihood ration w.r.t. F. Then G is stochastically larger than F.

Proof:

We must show that $F[c,\infty[\leqq G[c,\infty]$ for any real number c. It suffices to consider the case where $\alpha = F[c,\infty[> 0$. Then $I_{[c,\infty[}$ is a most powerful α test for testing F against G. Hence - since a most powerful test is unbiased - $\alpha = F[c,\infty[\leqq$

$\int I_{[c,\infty[} dG \leq G[c,\infty[$

Example

Let F and G have, respectively, densities $x \rightsquigarrow [\pi(1+x^2)]^{-1}$ and $x \rightsquigarrow [\pi(1+(x-1))^2]^{-1}$ w.r.t. Lebesque measure. Then G is stochastically larger than F. G does not, however, have monotonically increasing likelihood ratio w.r.t. F.

If F and G have densities f and g w.r.t. some measure μ such that g/f is monotonically decreasing in x then we shall say that G has monotonically decreasing likelihood ratio w.r.t. F. If so, then $dF/d\nu$ and $dG/d\nu$ may - for any σ finite measure ν dominating F and G - be specified so that $(dG/d\nu)/(dF/d\nu)$ is monotonically decreasing. In particular we may specify $dG/d(F+G)$ so that it is monotonically decreasing.

To any measure F on $]-\infty,\infty[$ we may construct another $\hat{F}$ by putting $\hat{F}(B)=F(-B)$ for each Borel set B. It is easily seen that G has monotonically decreasing likelihood ratio w.r.t. F if and only if $\hat{G}$ has monotonically increasing likelihood ratio w.r.t. $\hat{F}$ and this holds if and only if F has monotonically increasing likelihood ratio w.r.t. G. The last two propositions yields therefore:

Corollary

Let G have monotonically decreasing likelihood ratio w.r.t. F. Then any test of the form

$$\delta = \gamma I_{\{c\}} + I_{]-\infty,c[}$$

is a most powerful level $\alpha = \int \delta dF$ test for testing F against G provided $\alpha > 0$.

Corollary

Let G have monotonically decreasing likelihood ratio w.r.t. F. Then G is stochastically smaller that F.

References on monoton likelihood ratio are:

Rubin, Karlin. 1956	The theory of decision procedures for distributions with monotone likelihood ratio, A.M.S, 27:272-300
Karlin, 1956	Decision theory for Pólya type distributions Case of two actions I. Proc. Third Berkeley Symp. Math. Statist. Prod.,1:115-128
Karlin, 1957	Pólya type distributions II A.M.S. 28:281.
Karlin, 1958	Pólya type distributions III A.M.S. 29:406.

MATEMATISK INSTITUTT

Blindern,

Varians analyse
Høsten 1971
Torgersen.

A few univariate distributions which have applications in the analysis of varians.

Definition Let $\psi \geq 0$ and $r = 1,2,\ldots$. The non central χ^2 distribution with r degrees of freedom and non centrality parameter ψ^2 is the distribution of $(U_1+\psi)^2 + U_2^2 + U_3^2 + \ldots + U_r^2$ where $U_1,\ldots,U_r$ are independently and normally $(o,1)$ distributed. We shall denote this distribution by K_{r,ψ^2} and we will write K_r instead of $K_{r,0}$.

Notation If $\lambda \geq 0$ and $n=0,1,\ldots$ then we will write $b_\lambda(n) = (\lambda^n/n!)\exp-\lambda$; ie $b_\lambda(n)$ is the probability that $X = n$ when X has the Poisson distribution with parameter λ. Note that the possibility $\lambda = 0$ is not excluded.

Proposition The following random variables have the distribution K_{r,ψ^2}.

(i) $X_1^2 + \ldots + X_r^2$ where $X_1,\ldots,X_r$ are independently and normally distributed such that $\psi^2 = (EX_1)^2 + \ldots + (EX_r)^2$, and $\operatorname{var} X_i = 1$; $i = 1,\ldots,r$.

(ii) $X_1^2 + \ldots + X_{2N+r}^2$ where $X_1, X_2, \ldots$ and N are independent random variables such that X_i is normal $(0,1)$; $i = 1,\ldots$ and N has the Poisson distribution with expectation $\psi^2/2$

(iii) Z where the conditional distribution of Z given N is K_{2N+r} and N has the Poisson distribution with expectation $\psi^2/2$

We will prove this proposition by first proving part (i) and - using that result - prove:

Proposition

$$K_{r,\psi^2} * K_{s,\phi^2} = K_{r+s,\psi^2+\phi^2}$$

Finally this proposition will be used to prove parts (ii) and (iii) of the first proposition.

Proof of part (i) of the first proposition: Let C be a $r \times r$ orthonormal matrix such that:

$$(\sqrt{(EX_1)^2+\ldots+(EX_r)^2},\ 0,\ldots,0)' = C(EX_1,\ldots,EX_r)'$$

Introduce new random variables $Y_1,\ldots,Y_r$ by: $(Y_1,\ldots,Y_r)' = C(X_1\ldots X_r)'$. Then $Y_1,\ldots,Y_r$ are independently and normally distributed, each Y_i having variance 1 and $EY_1 = \sqrt{(EX_1)^2+\ldots+(EX_r)^2}$, $EY_2 = \ldots = EY_r = 0$. It follows that the distribution of $X_1^2+\ldots+X_r^2 = Y_1^2+\ldots+Y_r^2$ is K_{r,ψ^2}. □

Proof of the second proposition: Let $U_1,\ldots,U_r$, $V_1,\ldots,V_s$ be independently and normally (0,1) distributed. By the definition of the non central χ^2 distribution $(U_1+\psi)^2+U_2^2+\ldots+U_r^2+(V_1+\phi)^2+V_2^2+\ldots+V_r^2$ has the distribution $K_{r,\psi^2} * K_{s,\phi^2}$. Part (i) of the first proposition, however, imply that this distribution is $K_{r+s,\psi^2+\phi^2}$.

Proof of parts (ii) and (iii) of the first proposition. Simple calculations show that

$$K'_{1,\psi^2}(x) = \text{constant}\ \ x^{-\frac{1}{2}}\ (\exp - x/2)\ (\exp - \psi^2/2\ \ \cosh \sqrt{x}\psi\ ;\ x>0$$

In particular

$$K'_{1,0}(x) = \text{constant}\ \ x^{-\frac{1}{2}}\ (\exp - x/2)\ ;\ x>0$$

By induction - using the convolution formula -

$K'_{r,0}(x) \quad = \text{constant} \quad x^{r/2-1} - \exp(-x/2), \; x > 0.$

Using the power series expansion of cosh we get:

$$K'_{1,\psi^2}(x) = \sum_{n=0}^{\infty} \psi^{2n} (\exp - \psi^2/2) \text{ constant}_n \; x^{n-\frac{1}{2}} (\exp - x/2)$$

$$= \sum_{n=0}^{\infty} \text{constant}_n \; b_{\psi^2/2}(n) \; K'_{2n+1}(x), \text{ so that}$$

$$K_{1,\psi^2} = \sum_{n=0}^{\infty} \text{constant}_n \; b_{\psi^2/2}(n) \; K_{2n+1}$$

Hence

$$1 = \sum_{n=0}^{\infty} \text{constant}_n \; b_{\psi^2/2}(n)$$

It follows - since constant_n does not depend on ψ -by the completeness of the family of Poisson distributions that

$$\text{constant}_n = 1, n=1,2,\ldots$$

Hence $K_{r,\psi^2} = \sum_{n=0}^{\infty} b_{\psi^2/2}(n) \; K_{2n+r}$ when $r = 1$

If $r \geq 2$ then the convolution formula yields:

$$K_{r,\psi^2} = K_{1,\psi^2} * K_{r-1} = \sum_{n=0}^{\infty} b_{\psi^2/2}(n) \; K_{2n+1} * K_{r-1} = \sum_{n=0}^{\infty} b_{\psi^2/2}(n) K_{2n+r}$$

This proves (iii) and (ii) is just a particular form of (iii) □

<u>Problem</u>

(i) Work out the constant factor in K'_2

(ii) Use (i) to find expectation and variance in K_2

(iii) Use (ii) and the convolution formula $K_1 * K_1 = K_2$ to find expectation and variance in K_1

(iv) Use (iii) and the convolution formula to find expectation and variance in K_r

(v) Use (iv), part (iii) of the first proposition and the formulas:

$EZ = E(EZ1N)$

and

var Z = E(varZ|N) + var(EZ|N)

to show that expectation and variance in K_{r,ψ^2} are, respectively, $r+\psi^2$ and $2r + 4\psi^2$

Problem

(i) Work out the constant factor in K_r'.

(ii) Using (i) show that

$$\sum_{n=0}^{\infty} x^n/(n!)^2 = 2e^{x+1} K'_{2,2}(2x)$$

(iii) Define the non central χ^2 distribution with 0 degrees of freedom and non sentrality parameter ψ^2 by:

$$K_{0,\psi^2} \stackrel{def}{=} \sum_{n=0}^{\infty} b_{\psi^2/2}(n)\ K_{2n}$$

where K_0 is the one point distribution in 0.
Extend some of the previous results so that the extensions permits 0 degrees of freedom. In particular show that $K_{r,\psi^2} = K_{r,o} * K_{0,\psi^2}$ andfind - using (ii) - a "closed" form for the density of the absolute continuous part $\sum_{n=1}^{\infty} b_{\psi^2/2}(n)\ K_{2n}$ of K_{0,ψ^2}

(iii) of the last problem tells us that a non central χ^2 distribution with r degrees of freedom is a random translation of the central χ^2 distribution with r degrees of freedom. If Ψ^2 is the non sentrality parameter, then the random translation has the distribution K_{0,ψ^2} which has expectation ψ^2 and variance $4\psi^2$.

Another consequence is the inequality:

$$\|K_{r,\psi^2} - K_{r,0}\| \leqq 2(1-(\exp-\psi^2/2))$$

where $\|\cdot\|$ denotes total variation.

Demonstration: $\|K_{r,\psi^2} - K_{r,0}\| = \|K_{r,0} * (K_{0,\psi^2} - K_0)\| <$

$\leq \|K_{0,\psi^2} - K_0\| = 2K_{0,\psi^2}(\complement\{0\}) = 2(1-(\exp-\psi^2/2))$.

We shall need the following result:

Proposition

Let $\Sigma a_k Z^k$ and $\Sigma b_k Z^k$ be two powerseries and let Z_0 belong to the interiors of both circles of convergence. Then:

$$(\Sigma a_k Z_0^k)\left[\frac{d}{dZ} \Sigma b_k Z^k\right]_{Z_0} - \Sigma b_k Z_0^k \left[\frac{d}{dZ} \Sigma a_k Z^k\right]_{Z_0}$$

$$= \Sigma\left\{(n-k)(a_k b_n - a_n b_k) Z_0^{k+n-1} : 0 \leq k<n\right\}$$

Proof:

$$(\Sigma a_k Z_0^k)\left[\frac{d}{dZ} \Sigma b_k Z^k\right]_{Z_0} - (\Sigma b_k Z_0^k)\left[\frac{d}{dZ} \Sigma a_k Z^k\right]_{Z_0} =$$

$$\sum_{t=0}^{\infty}\left[\Sigma\left\{(k_1+1) b_{k_1+1} a_{k_2} - b_{k_2} a_{k_1+1} (k_1+1) : k_1+k_2 = t\right\}\right] Z_0^t$$

$$= \sum_{t=0}^{\infty}\left[\sum_{i=0}^{t+1} (t+1-2i)\, a_i b_{t+1-i}\right] Z_0^t$$

$$= \sum_{t=0}^{\infty} \Sigma\left\{(n-(t+1-n))(a_{t+1-n} b_n - a_n b_{t+1-n}) Z_0^{(t+1-n)+n-1} : 0 \leq t+1-n<(t+1)/2\right\}$$

$$= \Sigma\left\{(n-k)(a_k b_n - a_n b_k) Z_0^{k+n-1} : 0 \leq k<n\right\}$$

Corollary

Let $a_0, a_1, \ldots b_0, b_1, \ldots$ be real numbers. Suppose $\Sigma a_k x^k$ and $\Sigma b_k x^k$ both converges in a neighbourhood of $x_0 > 0$ and that:

(i) $a_k \geq 0,\ k = 1,2,\ldots$

with strict inequality for at least one k

(ii) $a_k b_n \geq a_n b_k$ when $0 \leq k < n$

with strict inequality for at least one pair (k,n)

Then $x \curvearrowright (\Sigma b_k x^k)/(\Sigma a_k x^k)$ is strictly increasing in a a neighbourhood of x_0.

Proof: By the previous proposition:

$$\left[\frac{d}{dx}(\Sigma b_k x^k)/(\Sigma a_k x^k)\right]_{x_0} = (\Sigma a_k x_0{}^k)^{-2}\ \Sigma\left\{(n-k):(a_k b_n - a_n b_k)x_0{}^{k+n-1} : 0 \leq k < n\right\} > 0$$

□

Remark

If $a_k > 0;\ k = 0,1,\ldots.$

then (ii) may be written (ii'): the sequence $b_1/a_1, b_2/a_2, \ldots.$

is non constant and monotonically increasing.

Corollary

K_{r,ψ_2^2} has monotonically increasing likelihood ration w.r.t. K_{r,ψ_1^2}, when $\psi_2^2 \geq \psi_1^2$.

Proof: This is an immediate consequence of the correspon-ding fact for pairs of Poisson distributions, the last propo-sition and the formula $K_{r,\psi^2} = \sum_{n=0}^{\infty} b_{\psi^2/2}(n) K_{2n+r}$.

Problem

Let δ be the one point distribution in 0 and let μ be Leberquemeasure on $[0,\infty[$. Define $\varepsilon(x);\ x \geq 0$

by $\varepsilon(x) = 0$ or 1 as $x = 0$ or $x > 0$

(i) $d\delta/d(\mu+\delta) = \varepsilon$, $d\mu/d(\mu+\delta) = 1-\varepsilon$

(ii) Find the density of K_{0,ψ^2} w.r.t. $\mu+\delta$.

(iii) Show that K_{0,ψ_2^2} has monotonically increasing likelihood ratio w.r.t. K_{0,ψ_1^2} when $\psi_2^2 \geq \psi_1^2$.

Definition

Let $\delta \in]-\infty,\infty[$ and $r = 1,2,\ldots.$ The non central t distribution with non centrality parameter δ and r degrees of freedom is the distribution of $(U+\delta)/\sqrt{Z/r}$ where U and Z are independent random variables - distributed as, respectively, normal (0,1) and K_r. This distribution function will be denoted by $T_{r,\delta}$ and we shall write T_r instead of $T_{r,0}$

Proposition

The density - g_δ - of $T/\sqrt{r}$ where T has the non central t-distribution with r degrees of freedom and non centrality parameter δ is:

$$g_\delta(x) = \text{constant}\ g_0(x) \int_0^\infty z^r \left[\exp - \tfrac{1}{2}(z^2-2z\delta x/\sqrt{1+x^2}+\delta^2)\right]dz ; x \in]-\infty,\infty[$$

where

$$g_0(x) = \text{constant}\ (1+x^2)^{-(r+1)/2}$$

and where the constants does not depend on δ.

Proof:

Let Y and Z be independent real random variables such that Y is normal $(\delta,1)$ and Z has a χ^2 distribution with r degrees of freedom. Then $T/\sqrt{r}$ has the same distribution as $Y/\sqrt{Z}$. The density of (Y,Z) is: constant $[\exp-(y-\delta)^2/2]Z^{r/2}(\exp-Z/2)$
It follows that the density of (X,Z) where $X = Y/\sqrt{Z}$ is:
constant $[\exp-(x\sqrt{z}-\delta)^2/2]z^{r/2-1}(\exp-z/2)$

Integrating over z we get:

$$g_\delta(x) = \text{constant} \int_0^\infty [\exp-(x\sqrt{z}-\delta)^2/2]z^{r/2-1}(\exp-z/2)\,dz$$

Substituting $y = \sqrt{z(1+x^2)}$ and then replacing y with z we get:

$$g_\delta(x) = \text{constant}\ (1+x^2)^{-(r+1)/2}\int_0^\infty z^r[\exp-\tfrac{1}{2}(z^2-2\delta zx/\sqrt{1+x^2}+\delta^2)]dz$$

In particular:

$$g_0(x) = \text{constant}\ (1+x^2)^{-(r+1)/2}$$

□

Corollary

$T_{r,\delta}$ has monotonically increasing - or decreasing likelihood ratio w.r.t. T_r as $\delta > 0$ or $\delta < 0$.

Proof: By the previous proposition:

$$g_\delta(x)/g_0(x) = \text{constant} \int_0^\infty z^r\left[\exp-\tfrac{1}{2}(z^2-2\delta zx/\sqrt{1+x^2}+\delta^2)\right]dz$$

and the integrand is increasing or decreasing in x as $\delta > 0$ or $\delta < 0$.

□

Definition Let $r = 1,2,\ldots,s = 1,2,\ldots$ and $\psi^2 \geqq 0$. The non central F-distribution with r and s degrees of freedom and non centrality parameter ψ^2 is the distribution of $(V/r)/(W/s)$ where V and W are independent random variables having, respectively, distributions K_{r,ψ^2} and K_s. This distribution will be denoted by F_{r,s,ψ^2} and we will write $F_{r,s}$ instead of $F_{r,s,0}$.

Definition Let $\alpha > 0$, $\beta > 0$ and $\psi^2 \geqq 0$. The non central β-distribution with parameters α and β and non centrality parameter ψ^2 is the distribution whose density is:

$$\sum_{n=0}^{\infty} b_{\psi^2/2}(n)\; x^{n+\alpha-1}(1-x)^{\beta-1}/ \int_0^1 x^{n+\alpha-1}(1-x)^{\beta-1}dx;\; x\varepsilon]-\infty,\infty[$$

This distribution will be denoted by J_{α,β,ψ^2} and we shall write $J_{\alpha,\beta}$ instead of $J_{\alpha,\beta,0}$. It follows directly from the definition that

$$J_{\alpha,\beta,\psi^2} = \sum_{n=0}^{\infty} b_{\psi^2/2}(n) J_{n+\alpha,\beta}$$

Proposition Let the real random variables V,W and N be such that

(i) (V,N) is independent of W

(ii) N has the Poisson distribution with expectation $\psi^2/2$

(iii) W has the distribution K_s

(iv) The conditional distribution of V given N is K_{2N+r}.

Then $(V/r)/(W/r)$ has the distribution F_{r,s,ψ^2}

Proof: The distribution of V is - by a previous proposition - K_{r,ψ^2} and V and W are independent. □

Proposition

If F has the distribution K_{r,s,ψ^2} then $[(r/s)F]/[(r/s)F+1]$ has the distribution $J_{r/2,s/2,\psi^2}$.

Proof: We may assume that $F = (V/r)/(W/r)$ where V, W and N has the properties (i) - (iv) of the previous proposition. It follows that

$$[(r/s)F]/[(r/s)F+1] = V/(V+W).$$

The density of the conditional distribution of (V,W) given N is:

$$\text{constant } v^{N+(r/2)-1} w^{s/2-1} \exp-(v+w)/2 \ ; \ v > 0, w > 0.$$

Put $X = V/(V+W)$. It follows that the density of the conditional distribution of (V,X) given N is:

$$\text{constant}\, x^{-s/2-1}(1-x)^{s/2-1} \int_0^\infty v^{N+(r+s)/2-1} (\exp-v/2x)\, dv$$

$$= \text{constant } x^{N+r/2-1}(1-x)^{s/2-1} = J'_{N+r/2,s/2}(x)$$

so that the distribution of X is:

$$E\, J_{N+r/2,s/2} = J_{r/2,s/2,\psi^2}.$$

Proposition

$J_{\alpha,\beta,\psi_2^2}$ has monotonically increasing likelihood ratio w.r.t. $J_{\alpha,\beta,\psi_1^2}$ when $\psi_2^2 > \psi_1^2$

Proof: By the definition: $J_{\alpha,\beta,\psi_2^2}(x)/J_{\alpha,\beta,\psi_1^2}(x) =$

$$\sum_{n=0}^{\infty} b_{\psi_2^2/2}(n) \text{ constant}_n\, x^n \Big/ \sum_{n=0}^{\infty} b_{\psi_1^2/2}(n) \text{ constant}_n\, x^n$$

and this ratio is - by the argument used to prove the corresponding statement for the non central χ^2 distributions - a strictly increasing function in x. □

Corollary F_{r,s,ψ_2^2} has monotonically increasing likelihood ratio w.r.t. F_{r,s,ψ_1^2} when $\psi_2^2 > \psi_1^2$

Proof:
Follows directly from the last two propositions. □

Let $U_1, U_2, \ldots, V_1, V_2, \ldots$ be independently and normally (o,1) distributed. Then

$$F_{r,s}[0,x] = P((U_1^2 + \ldots + U_r^2)/(V_1^2 + \ldots + V_s^2) \leq (r/s)x)$$

Hence:

Proposition Let $p \in]0,1[$. Then $(r/s)\, F_{r,s}^{-1}(p)$ is strictly increasing in r and strictly decreasing in s.

A few simple - but useful - facts are collected in:

Proposition:

(i) If T has the distribution $T_{r,\delta}$, then the distribution of T^2 is F_{1,r,δ^2}.

(ii) If F has the distribution $F_{r,s}$ then the distribution of F^{-1} is $F_{s,r}$.

(iii) If V and W are independent and has, respectively, distributions K_{r,ψ^2} and $K_{r,0}$ then $V/(V+W)$ has the distribution $J_{r/2, s/2, \psi^2}$.