Hedmark distriktshøgskole
Postboks 104
2451 Rena
Telf: 62 44 08 00

On the distribution of t- type statistics involving a weighted sum of χ^2 and some related tests

Erik Mønness

Arbeidsnotat 1993:1
Hedmark distriktshøgskole
Rena

On the distribution of t-type statistics involving a weighted sum of Chi-squares, and some related tests.

Abstract.

The distribution of $U/\sqrt{(V+W\cdot f)}$, where U, V and W are independent, U is normal, V and W are Chi-squares, $f\geq 1$, is given as an infinite weighted sum of t-variables. The weights are given as the probabilities of the negative binomial distribution. Thus the distribution is practically exact, and it can be calculated using commercially available distribution software. Some tests in the Behrens-Fisher situation are of this kind. An alternative test of the mean in a variance component situation, based on this kind of statistic, is shown to be superior in power on certain alternatives as compared to the uniformly most powerful unbiased test. This alternative test is a practical alternative on small samples.

Key words: Behrens-Fisher, Variance components.

1. Introduction.

Let U be a normal random variable, $U \sim N(\mu,1)$. Let V and W be Chi-square random variables with v and w degrees of freedom, respectively. Let U, V, W be stochastically independent. The purpose of this paper is to study the distribution of the statistic $T_f = U/\sqrt{(V+W \cdot f)}$, where $f \geq 1$. When f=1 we have the well known result that $t_1\sqrt{(v+w)}$ has a t distribution.

This statistic emerges in certain normal linear models, e.g. in Behrens-Fisher problem, and in variance component analysis. The distribution of T_f, is derived in Ray & Pitman (1961), in the case when the degrees of freedom involved are even. Lee & Gurland (1975) derived the distribution (without any restrictions) using numerical integration . The present paper combines a result on the distribution of $V+W \cdot f$ given by Robbins & Pitman (1949) with a short distribution calculation to give a formula that can be calculated using only standard statistical computer software.

2. The Behrens-Fisher problem:

Given the model

$$Y_{ij} = \mu_i + E_{ij}, \qquad i=1,2$$

$$E_{ij} \sim NID(0,\sigma_i), \qquad j=1,\ldots,J_i.$$

Let

$$\overline{Y_{i.}}=\frac{1}{J_i}\sum_{j=1}^{J_i} Y_{ij} \qquad \text{and} \qquad SS_i^2=\sum_{j=1}^{J_i}\left(Y_{ij}-\overline{Y_{i.}}\right)^2 \qquad i=1,2$$

The Behrens-Fisher problem is to make inference on $\delta=\mu_2-\mu_1$. Various tests for this purpose have been developed by several authors. Lee and Gurland (1975) give an excellent overview and they also give the power of all these tests. Let $\Delta^2=(\sigma_2/\sigma_1)^2$.
The test statistics are mainly based upon

$$T'=\frac{\left(\overline{Y_{2.}}-\overline{Y_{1.}}\right)}{\sqrt{\dfrac{SS_1^2}{J_1(J_1-1)}+\dfrac{SS_2^2}{J_2(J_2-1)}}}=\frac{\dfrac{\left(\overline{Y_{2.}}-\overline{Y_{1.}}\right)}{\sigma_1\sqrt{\dfrac{\Delta^2}{J_2}+\dfrac{1}{J_1}}}}{\sqrt{\dfrac{SS_1^2}{\sigma_1^2}+\left(\Delta^2\dfrac{J_1(J_1-1)}{J_2(J_2-1)}\right)\dfrac{SS_2^2}{\sigma_2^2}}}\cdot\sqrt{\left(\frac{\Delta^2}{J_2}+\frac{1}{J_1}\right)J_1(J_1-1)}$$

The statistics are of the general form T_f times a constant (the square root) with $f=\Delta^2*(J_1(J_1-1))/(J_2(J_2-1))$. However, most proposed tests use an estimated degree of freedom, thus they are of the type

$$Pr\left(\frac{U}{\sqrt{V+Wf}} > k(V,W)\right)$$

(Lee & Gurland, 1975). The work presented here does not allow the critical value to be random. However, the distribution of T_f is at hand using only ordinary and commercially available distributions. If one make use of some non-random choice of critical value, or due to enough observations believe the statistic to be normally distributed, the true power is easy calculated.

If one incorrectly assumes $\Delta^2 = a^2$ (a known number, usually taken to be 1) we have the usual pooled statistic:

$$T=\frac{\left(\overline{Y_{2.}}-\overline{Y_{1.}}\right)}{\sqrt{\left(\frac{1}{J_1}+\frac{a^2}{J_2}\right)\frac{SS_1^2+\frac{SS_2^2}{a^2}}{J_1+J_2-2}}}=\frac{\frac{\left(\overline{Y_{2.}}-\overline{Y_{1.}}\right)}{\sigma_1\sqrt{\frac{\Delta^2}{J_2}+\frac{1}{J_1}}}}{\sqrt{\frac{SS_1^2}{\sigma_1^2}+\frac{\Delta^2}{a^2}\frac{SS_2^2}{\sigma_2^2}}}\cdot\sqrt{J_1+J_2-2}\cdot\sqrt{\frac{\frac{\Delta^2}{J_2}+\frac{1}{J_1}}{\frac{a^2}{J_2}+\frac{1}{J_1}}}$$

In this case $f= \Delta^2/a^2$.

It is not certain that $f\geq 1$ in these cases, but if $f<1$ then one can use $1/f$ as multiplier to the opposite χ^2 statistic). If $a^2=1$ and $J_1=J_2$ the two statistics are identical.

3. Inference about the mean in a variance component situation.

Let

$$Y_{ij} = \mu + D_i + E_{ij} \qquad i=1,\ldots,I \text{ and } j=1,\ldots,J$$

where

$$D_i \sim NID(0,\sigma_d) \quad \text{and} \quad E_{ij} \sim NID(0,\sigma_e)$$

We have

$$\overline{Y_{..}} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{ij} \qquad \overline{Y_{..}} \sim N\left(\mu, \frac{\sqrt{J\sigma_d^2+\sigma_e^2}}{\sqrt{IJ}}\right)$$

and

$$SS_{de}=\sum_{i=1}^{I} J\left(\overline{Y_{i}.}-\overline{Y..}\right)^2 \qquad \frac{SS_{de}}{J\sigma_d^2+\sigma_e^2} \sim \chi^2_{I-1}$$

and

$$SS_e=\sum_{i=1}^{I}\sum_{j=1}^{J}\left(Y_{ij}-\overline{Y_{i}.}\right)^2 \qquad \frac{SS_e}{\sigma_e^2} \sim \chi^2_{I(J-1)}$$

We want to test the hypothesis $\mu=0$ against the alternative $\mu>0$. Let $k_{i,\alpha}$ be the $1-\alpha$ quantile of the central t distribution with i degrees of freedom.

The uniformly most powerful unbiased test is to reject when

$$\frac{\overline{Y..}\sqrt{I-1}\sqrt{IJ}}{\sqrt{SS_{de}}} > k_{I-1,\alpha}$$

However, since SS_e usually has many more degrees of freedom, and since it also carries information about the variability, it should be possible to use it to gain testing power on certain alternatives.

Let

$$\Delta^2=\frac{\sigma_e^2}{J\sigma_d^2+\sigma_e^2}$$

Thus $\Delta^2 \in (0,1)$.

If Δ^2 had been known to have a value (a^2, say), then the uniformly most powerful unbiased test is to reject when

$$\frac{\overline{Y..}\sqrt{I-1\ +\ I(J-1)}\sqrt{IJ}}{\sqrt{SS_{de}+\frac{SS_e}{a^2}}} > k_{I-1+I(J-1),\alpha}$$

Now assume that Δ^2 has a lower bound which is strictly greater than zero:

$$\Delta^2 \in [a^2,1), \quad \text{where } a>0 \text{ is known.}$$

Then we have the following:

$$Pr\left(\frac{\overline{Y..}}{\sqrt{SS_{de}+\frac{SS_e}{a^2}}}>k\right) = Pr\left(\frac{\frac{\overline{Y..}}{J\sigma_d^2+\sigma_e^2}}{\sqrt{\frac{SS_{de}}{J\sigma_d^2+\sigma_e^2}+\frac{SS_e}{(J\sigma_d^2+\sigma_e^2)\Delta^2}\frac{\Delta^2}{a^2}}}>k\right) \le$$

$$\le Pr\left(\frac{\frac{\overline{Y..}}{J\sigma_d^2+\sigma_e^2}}{\sqrt{\frac{SS_{de}}{J\sigma_d^2+\sigma_e^2}+\frac{SS_e}{\sigma_e^2}}}>k\right)$$

The test that is optimal when $\Delta^2=a^2$ thus also hold its size when $\Delta^2 \in [a^2,1)$. Due to continuity, this test has greater power than the uniformly most powerful unbiased test also in a certain area of the alternative $\mu>0$, typically when Δ^2 is close to a^2. This test is of the general form T_f, where $f= \Delta^2/a^2 \ge 1$. To a given μ, the power $Pr(T_f>k)$ will decrease monotonically from its maximum at f=1 to zero for large f values. For given degrees of freedom v and w, it is sufficient to calculate a power table $Pr(T_f\sqrt{(v+w)}>t_{v+w,\alpha})$ for different μ and different f values.

In a practical situation a^2 has to be fixed to a specific positive value. Then substitute f with Δ^2/a^2 in the power table. Since $\Delta^2 \in [a^2,1]$ the actual part of the power table involved is where $f\in [1,1/a^2]$. Thus if a^2 is close to 1, only a small set of f values near 1 is involved; this is where the power is high. If a^2 is small, also higher f values with low test power is involved. Conclusion: A big a^2 favors this alternative test. In the original parameter space, to fix a^2 is equivalent to fix a value c such that $\sigma_e \geq c*\sigma_d$, giving $a^2=c^2/(J+c^2)$. A large "c" gives a large "a". A large "c" means that the variability in the data is mainly in the error E_{ij}, giving approximately IJ independent observations. A small "c" means that the variability can be mainly in the D_i's. If so, it is basically only I independent observations. Increasing J gives more degrees of freedom (=IJ-1), but it lowers the a^2 value.

4. The distribution of the general statistic T_f.

The calculations depends on the following result on the distribution of sums of χ^2 variables (Robbins & Pitman, (1949), derived a more general result with more than two χ^2 variables in the sum. This is applied and developed here on the specific V+Wf situation.) Let $V \sim \chi^2$ with v degrees of freedom be independent of $W \sim \chi^2$ with w degrees of freedom.

The cumulative distribution of V+Wf, f≥1 is given by

$$Pr(\,(V+Wf) \leq k) \quad = \quad \sum_{j=0}^{\infty} c_j(f)\chi^2_{v+w+2j}(k)$$

where $\chi^2(.)$ is the cumulative Chi-square distribution with the given degrees of freedom, and $c_j(f)$ is given by the "probabilities" in the negative binomial distribution:

$$c_j(f) = \binom{\frac{w}{2}+j-1}{\frac{w}{2}-1}\left(\frac{1}{f}\right)^{\frac{w}{2}}\left(1-\frac{1}{f}\right)^j \qquad j\geq 0 \qquad w\geq 1$$

Now let us return to the T_f statistic. Let k > 0. Then

$$Pr\left(\frac{U}{\sqrt{V+Wf}}>k\right) = \int Pr\left(\frac{t}{\sqrt{V+Wf}}>k \;\middle|\; U=t\right)dPU^{-1}(t) =$$

$$= \int_0^\infty Pr\left(\frac{t}{\sqrt{V+Wf}}>k\right)dPU^{-1}(t) = \int_0^\infty Pr\left(V+Wf<\frac{t^2}{k^2}\right)dPU^{-1}(t) =$$

$$= \int_0^\infty \sum_{j=0}^\infty c_j(f)\chi^2_{v+w+2j}\left(\frac{t^2}{k^2}\right)dPU^{-1}(t) = \sum_{j=0}^\infty c_j(f)\int_0^\infty Pr\left(V_j \leq \frac{t^2}{k^2}\right)dPU^{-1}(t) =$$

$$= \sum_{j=0}^\infty c_j(f)Pr\left(\frac{U}{\sqrt{V_j}}\sqrt{v+w+2j}>k\sqrt{v+w+2j}\right), \qquad \text{where} \quad V_j \sim\chi^2_{v+w+2j}$$

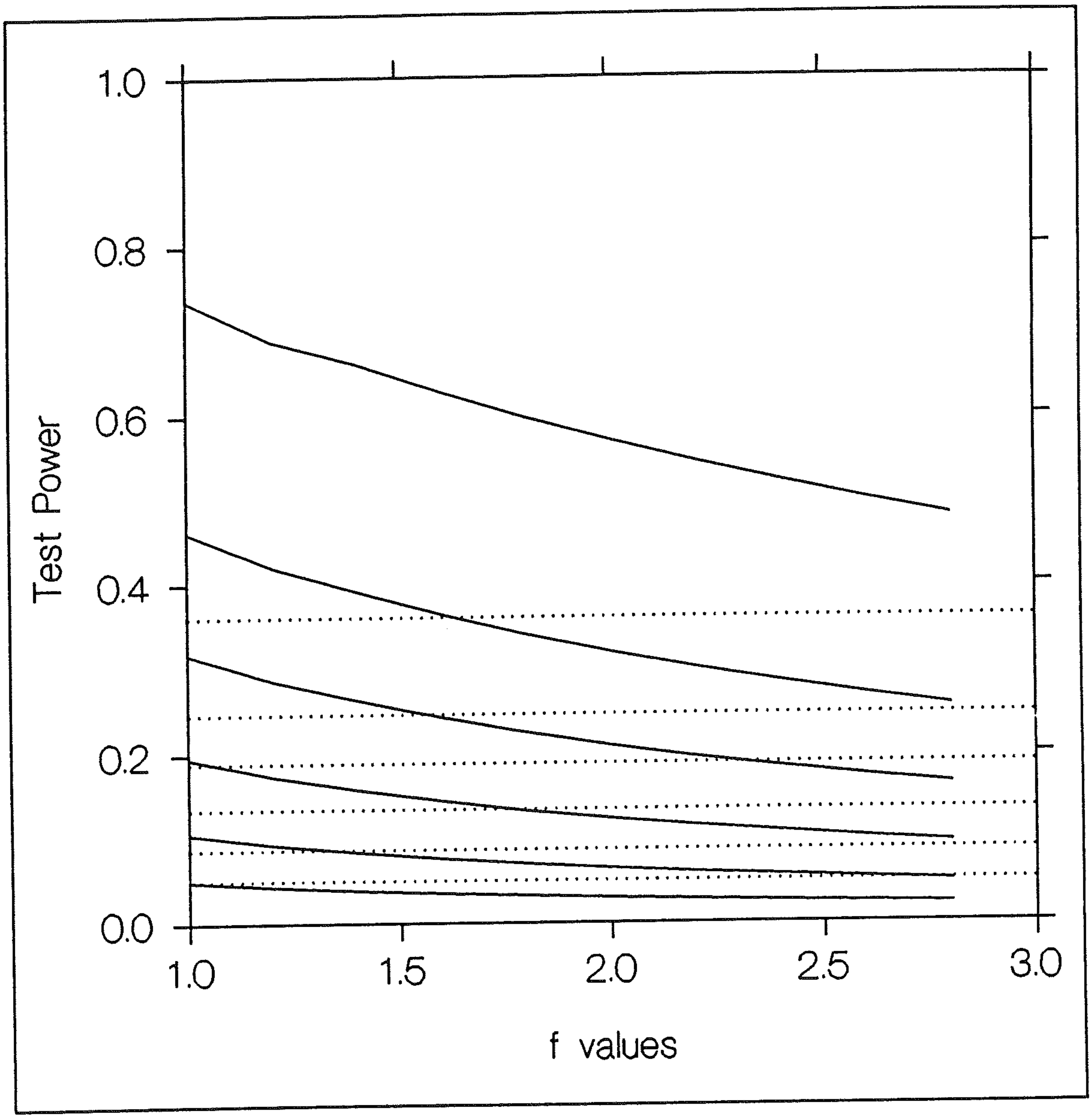
$$= \sum_{j=0}^\infty c_j(f)\left(1-T_{v+w+2j}(k\sqrt{v+w+2j})\right)$$

where $T_{v+w+2j}(k)$ is the (possible non-central) cumulative T-distribution with the given number of degrees of freedom. Thus the statistic T_f is distributed as an infinite weighted sum of ordinary

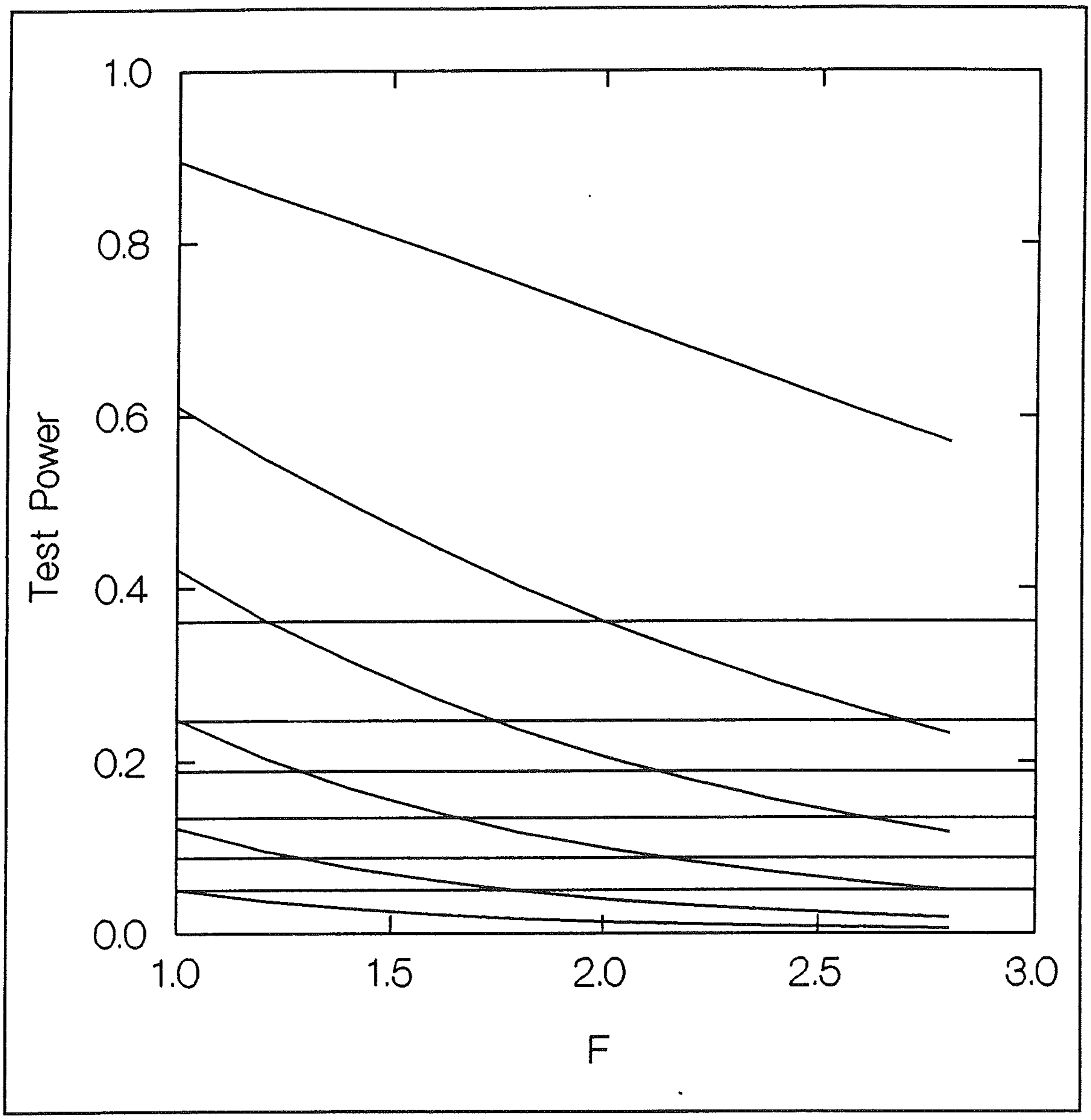
t-variables, the weights being the negative binomial distribution "probabilities". Since all factors of all components in the sum is between 0 and 1, we have full control of the eventually error imposed by only calculating a finite sum.
A similar result also hold when the number of Chi-squares is bigger than two, and also for F-type statistics, see also Robbins & Pitman (1949).

5. The variance component situation, examples.

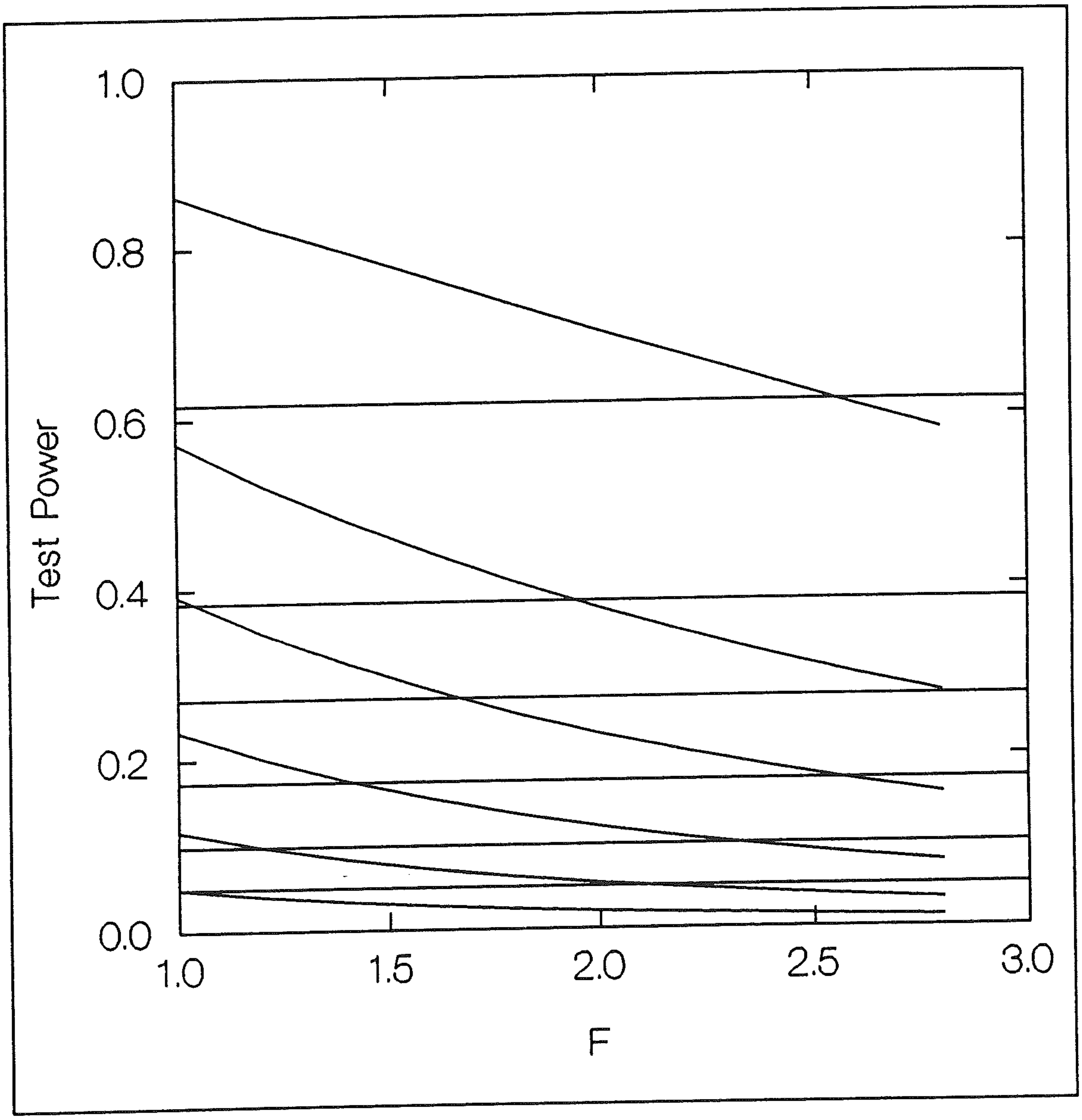
Let I be the number of groups and J be the number of replicates in each group of the variance component layout. Let the test size be 0.05. We are going to evaluate the test power on alternatives with noncentrally parameters 0.0, 0.5, 1.0, 1.5, 2.0, and 3.0, and f values from 1.0 to 2.8.
In an ordinary t-test situation, the value of increasing the sample size to gain power is higher when the sample originally is small. So also in this case. Also this alternative test is biased and has lower power than the most powerful unbiased test (when f is "large"). All this suggest that the possibly practical value is when the number of groups I is small (when then most powerful unbiased test has few degrees of freedom). Calculations show that this test is worth considering when $I \in [2,5]$. There are 6 figures, on I=2 & J=2, I=2 & J=10, I=3 & J=3, I=3 & J=10, I=4 & J=4, I=6 & J=6.
Each figure shows 6 curves and 6 straight lines. These are the power of the alternative test (solid curves) and the most powerful unbiased test (straight lines), for noncentrally parameters 0.0, 0.5, 1.0, 1.5, 2.0, and 3.0, repectively, taken from below.



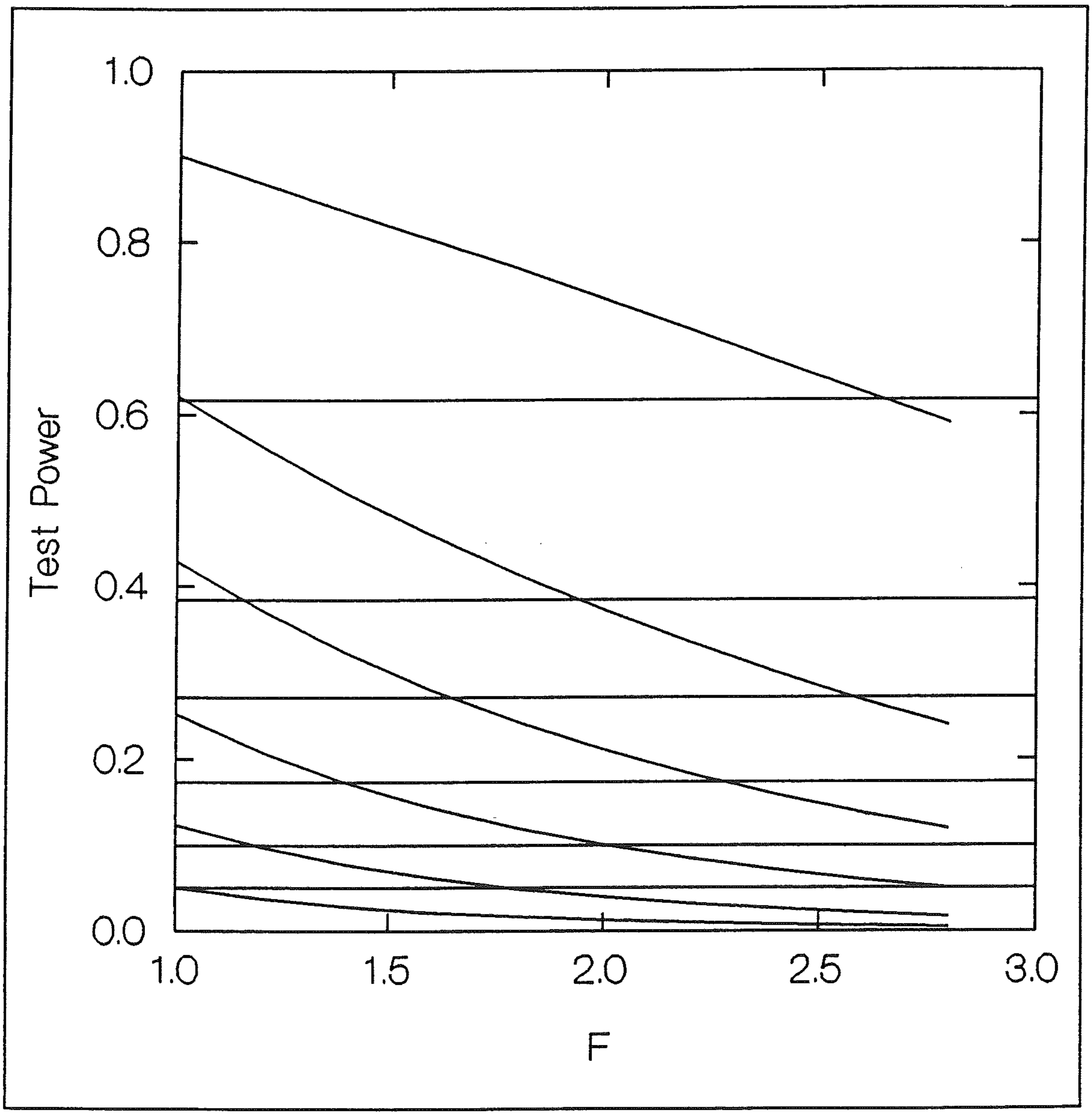
I=2, J=2



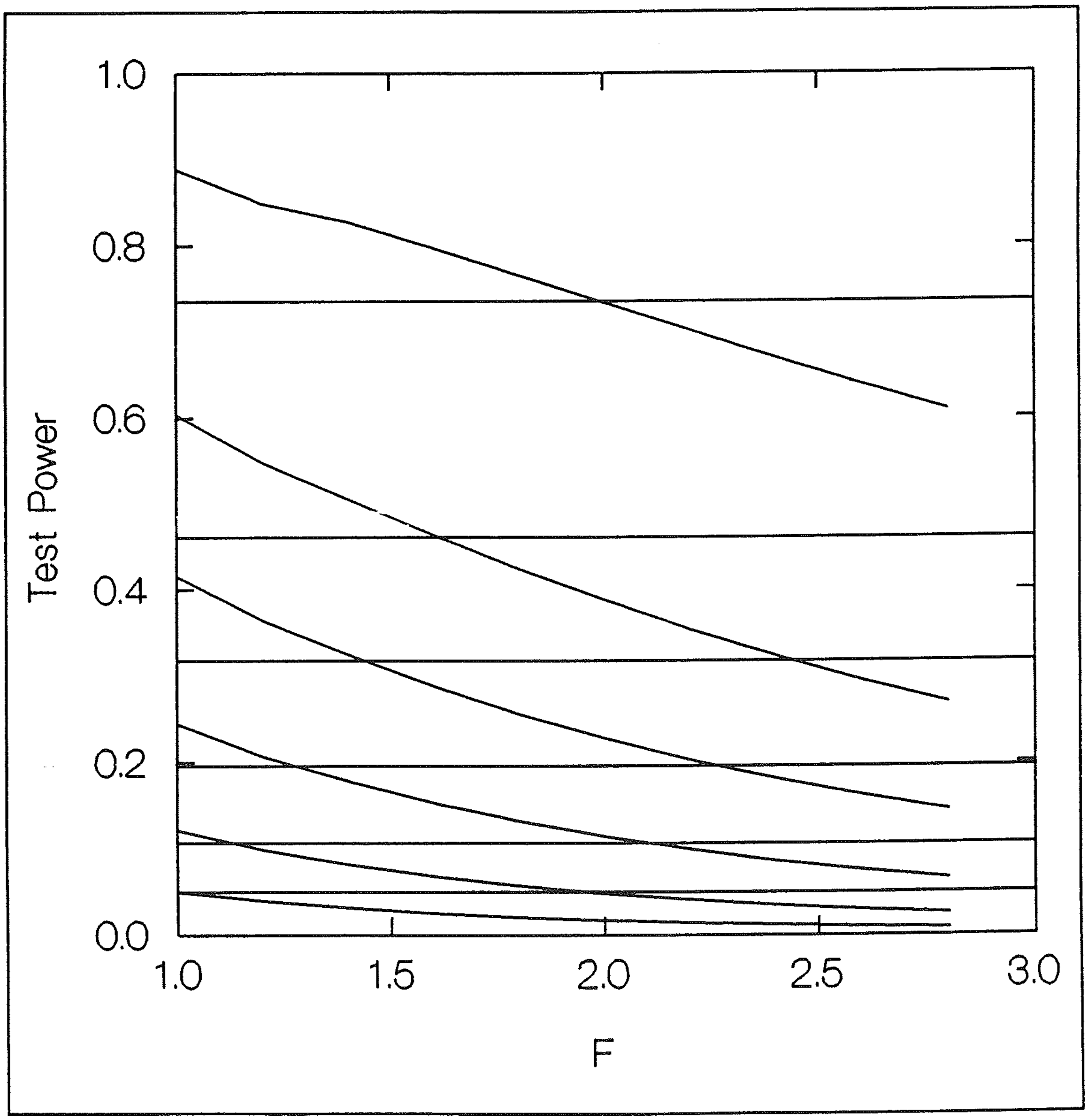
I=2, J=10



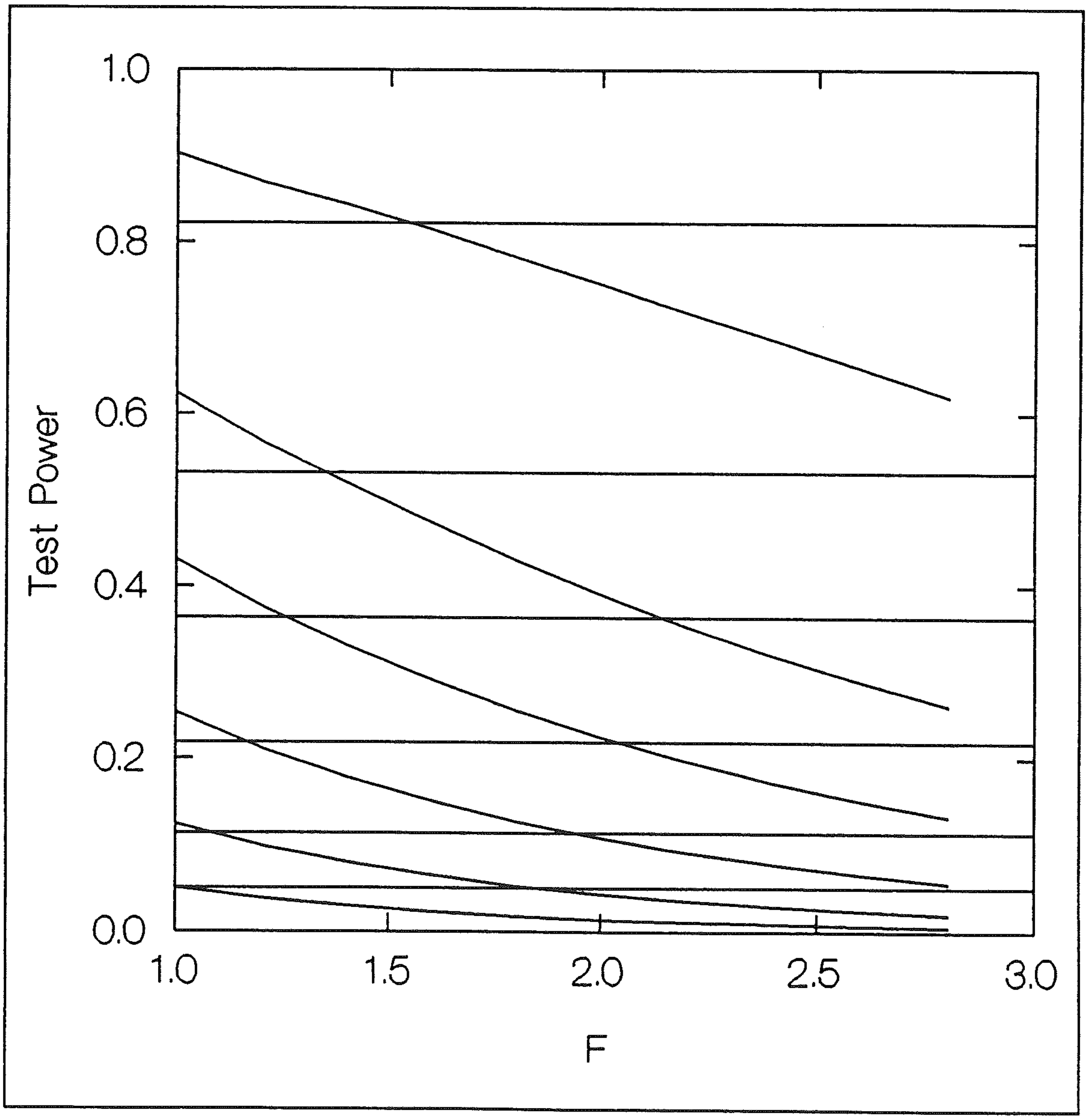
I=3, J=3



I=3, J=10



I=4, J=4



I=6, J=6

Consider the I=3 & J=3 display: As much as 0.2 "power units" can be gained when the noncentrally parameter i greater than 2.0, and the alternative test is superior when f≤2. The I=6 & J=6 display shows that little can be gained for small f values, and much lost for moderate and large f values.

6. Conclusion:

The Power function of the statistic T_f is easily calculated. This may be used in the Behrens-Fisher problem. In a variance component situation, an alternative test is proposed which in certain situation can prove superiority.

Acknowledgement: The main distribution result presented here was derived in my thesis at University of Oslo in 1976. I would like express my thanks to my then adviser, prof. Ragnar Nordberg.

Literature cited:

Lee, A. F. S. & Gurland, J. (1975): Size and power of tests for equality of means of two normal populations with unequal variances. Journal of the American Statistical Association Dec 1975, Vol 70 No 352, p 933-941.

Ray, W. D. & Pitman, A. E. N. T. (1961): An exact Distribution of the Fisher-Behrens-Welch Statistic for Testing the Difference between the Means of Two Normal Populations with Unknown Variates.

Journal of the Royal Statistical Society, 1961, Vol 23, No 2, p 377-384.

Mønness, E. (1976): Hypoteseprøving i et innskrenket parameterrom belyst ut ifra maksiminprinsippet. [Testing hypothesis in a restricted parameter space. A Minimax approach] (In Norwegian). Thesis, Dep. of mathematical statistics, University of Oslo, Norway.

Robbins, H. & Pitman, E. J. G. (1949): Application of the method of mixtures to quadratic forms in normal variates. The Annals of mathematical statistics, 20 p552-560.

SAS,(1985): SAS/IML Guide for Personal Computers, Ver. 6. SAS Inst. Inc. Cary, NC-USA