Multivariate Analysis and Its Applications
IMS Lecture Notes – Monograph Series (1994) Volume 24

INTERIM ANALYSIS FOR NORMALLY DISTRIBUTED OBSERVABLES

BY SEYMOUR GEISSER AND WESLEY JOHNSON

University of Minnesota and University of California at Davis

We address the problem of whether an experiment should be continued or aborted when N observations are in hand and a total of $K > N$ have been scheduled for a decision. A Bayesian predictive approach is used to determine the probability that if one continued the trial with a further sample of size M where $N + M \geq K$ one would come to a particular decision regarding some set of parameters. In particular, sampling from a multivariate normal distribution will be discussed.

1. Introduction.

Often experiments will consist of a series of independent observations with some minimum sample size required, say K, before a conclusion is reached concerning the efficacy of a new treatment. Many such trials are costly and time consuming. Frequently an investigator would like to know at some interim point whether the continuation of the trial is worthwhile. With regard to a new treatment or a therapy, the issue is invariably whether continuation will lead to a conclusion that the treatment is at least as effective as some standard. There are frequentist methods which control type I and type II errors if interim analyses are made at preset sample sizes in a sequential trial. Depending on the number of such interim analyses, the required sample difference can be much larger than in a trial where no interim analyses are made. Also it is not always convenient to conduct such analyses at preset sample sizes in a trial. Other methods that allow for analyses at arbitrary sample sizes involve highly conservative tests which render even more difficult the detection of differences.

Although Bayesian statisticians ordinarily do not suffer from such restrictions they also may be subject to an important trial which requires at least some fixed number of observations before a conclusion is reached. This is

---

Research supported in part by NIGMS Grant 25271.

AMS 1980 Subject Classifications: 62F15, 62H15, 62L99.

Key words: Bayesian approach, Hotelling T^2, interim analysis, multivariate distance, non-central chi-squared distribution, normal distribution, predictive distributions.

inarguably true if the conclusion is to be convincing to a wider public or in particular to a regulatory agency which licenses new therapies. Hence Bayesians also need to consider interim analyses in order to decide whether to abandon a trial or to continue a trial to its specified term.

2. A Mixed Metaphor. In the last few years some Bayesian procedures have been suggested for those who prefer a frequentist analysis, Choi and Pepple (1989), Choi et al. (1985) and Spiegelhalter et al. (1986). First we shall illustrate the procedures suggested in a very simple case and indicate certain difficulties that arise if they are used.

Suppose $X_1, \ldots, X_{N+M}$ are i.i.d. $N(\mu, 1)$ and a test of the following hypotheses is required, $H_0 : \mu \le \mu_0$ vs. $H_1 : \mu > \mu_0$. The standard test for testing H_0 vs. H_1 at level α is to reject H_0 if

$$\sqrt{N+M}(\bar{x} - \mu_0) > z_\alpha, \tag{2.1}$$

where $\alpha = 1 - \Phi(z_\alpha)$, and $\Phi(\cdot)$ is the standard normal distribution function.

To conduct an interim analysis at N observations, it is suggested that the probability of achieving the above event (2.1) be calculated. A syncretic approach has been proposed and developed in the previously mentioned papers which apply Bayesian predictive ideas towards the solution of this problem. It is assumed that the prior for μ is constant to conform as closely as possible to a frequentist analysis. After N observations are in hand, this results in a posterior distribution for μ as $N(\bar{x}_N, 1/N)$. Now we compute the probability of the rejection set in (2.1)

$$\Pr\left[(N+M)^{1/2}\left[\frac{N\bar{x}_N + M\overline{X}_M}{N+M} - \mu_0\right] > z_\alpha\right] = P_\alpha \tag{2.2}$$

where $x_1 + x_2 + \cdots + x_N = N\bar{x}_N$ and $X_{N+1} + \cdots + X_{N+M} = M\overline{X}_M$, noting that now $\bar{x}_N$ is fixed but the as yet unobserved $\overline{X}_M$ is random. The predictive distribution of $\overline{X}_M$ is easily obtained to be $N(\bar{x}_N, 1/N + 1/M)$. Regrouping terms in (2.2) and letting

$$Z = \frac{(\bar{X}_M - \bar{x}_N)}{\left(\frac{1}{N} + \frac{1}{M}\right)^{1/2}}$$

we obtain

$$P_\alpha = \Pr\left[Z \ge \frac{z_\alpha\sqrt{N+M} - (M+N)(\bar{x}_N - \mu_0)}{M\sqrt{\frac{M+N}{MN}}}\right]$$

where Z is $N(0,1)$. Finally, this yields

$$P_\alpha = 1 - \Phi\left(\left(\frac{N}{M}\right)^{1/2}\left[z_\alpha - (M+N)^{1/2}(\bar{x}_N - \mu_0)\right]\right), \tag{2.3}$$

the probability that if the trial were continued for an additional M observations, H_0 would be rejected at level α. Small values of P_α would discourage while large values would encourage the continuation of the trial . It follows from (2.3) that

$$\lim_{M\to\infty} P_\alpha = 1 - \Phi(-\sqrt{N}(\bar{x}_N - \mu_0)) = 1 - P, \tag{2.4}$$

where

$$P = \Pr[Z \geq \sqrt{N}(\bar{x}_N - \mu_0)],$$

the P-value at N observations for testing H_0 which is independent of α. The limiting result of (2.4) turns out to be the posterior probability of the alternative as is indicated subsequently in (3.2).

This implies that if one continued the trial indefinitely, the predictive probability of rejecting H_0 approaches $1 - P$ irrespective of α. This is a "Bayesian" interpretation of $1-P$ that naive students and some investigators often make with regard to significance tests. Further, teachers of frequentist statistics often strive mightily to disabuse students of this flawed interpretation. The result does not have an acceptable frequentist interpretation and furthermore, this is not the kind of test a Bayesian would apply. Hence one needs to be rather careful in mixing metaphors.

3. The Bayes Approach. A Bayesian approach in this situation would reject H_0, say, if the posterior probability, for a specified p, is

$$\Pr\left[\mu > \mu_0 | x_1, \ldots, x_{N+M}\right] \geq p$$

assuming a prior $\pi(\mu)$ for μ. Hence, after N observations one would calculate the predictive probability of the above event assuming $x_1, \ldots, x_N$ have been observed and future observables $X_{N+1}, \ldots, X_{N+M}$ are random. In this example if the previous prior for μ is used, then $\mu \sim N(\bar{x}_{N+M}, 1/(N+M))$. Hence H_0 is rejected if

$$\Pr[\sqrt{N+M}(\mu - \bar{x}_{N+M}) > \sqrt{N+M}(\mu_0 - \bar{x}_{N+M})] \geq p$$

or

$$1 - \Phi(\sqrt{N+M}(\mu_0 - \bar{x}_{N+M})) \geq p,$$

where $(N+M)\bar{x}_{N+M} = (x_1 + \cdots + x_{N+M})$. Now stopping at N, we need to find the predictive probability of the above event i.e.

$$P_p = \Pr\left[1 - \Phi\left(\sqrt{N+M}\left(\mu_0 - \frac{N\bar{x}_N + M\overline{X}_M}{N+M}\right)\right) \geq p\right].$$

After some algebra, and denoting $\Phi^{-1}(p)$ as the inverse distribution function,

$$P_p = 1 - \Phi\left(\left(\frac{N}{M}\right)^{1/2}\left[(\mu_0 - \bar{x}_N)(N+M)^{1/2} - \Phi^{-1}(1-p)\right]\right) \tag{3.1}$$

is the chance of rejecting H_0 if the trial were continued.

Now if the trial were contemplated to be continued indefinitely,

$$\lim_{M\to\infty} P_p = 1 - \Phi((\mu_0 - \bar{x}_N)\sqrt{N}) = \Pr\left[\mu > \mu_0 | x_1, \dots, x_N\right] \tag{3.2}$$

which does not depend on p and is obviously the posterior probability given N observations. This is perfectly sensible as the best prediction of what would occur if one were to continue sampling indefinitely.

4. Normal Sampling with Mean and Variance Unknown. Let $X_i, i = 1, \dots, N + M$ be i.i.d. $N(\mu, \sigma^2)$ and $\pi(\mu, \sigma^2) \propto 1/\sigma^2$. Hence it is well known that

$$\frac{(\mu - \bar{x}_{N+M})\sqrt{N+M}}{S_{N+M}} \sim t_\nu,$$

where t_ν is a student random variable with $\nu = N + M - 1$ degrees of freedom. To test

$$H_0 : \mu \le \mu_0 \quad \text{vs.} \quad \mu > \mu_0,$$

we will decide for H_0 if the posterior probability

$$\Pr\left[\mu \le \mu_0 | x^{(N+M)}\right] = S_\nu \left[\frac{(\mu_0 - \bar{x}_{N+M})\sqrt{N+M}}{s_{N+M}}\right] \ge p,$$

where

$$\nu s_{N+M}^2 = \sum_{i=1}^{N+M} (x_i - \bar{x}_{N+M})^2$$

and $S_\nu(\cdot)$ is the student distribution function with ν degrees of freedom. After observing $x^{(N)} = (x_1, \cdots, x_N)$ and some algebraic manipulation we find that we need to calculate

$$P_p = \Pr\left(\frac{\left[\mu_0 - \frac{N\bar{x}_N}{N+M} - \frac{MX}{N+M}\right](N+M)^{1/2}(N+M-1)^{1/2}}{\left(z + Y + \frac{NM}{N+M}(X - \bar{x}_N)^2\right)^{1/2}} \ge S_\nu^{-1}(p)\right) \tag{4.1}$$

for

$$X = M^{-1} \sum_{i=N+1}^{N+M} X_i,$$

$$z = (N-1)s_N^2 = \sum_{i=1}^{N} (x_i - \bar{x}_N)^2,$$

$$Y = (M-1)s_M^2 = \sum_{i=N+1}^{N+M} (X_i - X)^2,$$

and $S_\nu^{-1}(p)$ is the inverse student distribution function or the quantile function, where X and Y are the as yet unobserved random quantities. This requires the calculation of the joint predictive distribution of X and Y. This can easily be obtained, Geisser (1992), as

$$f(x,y|x^{(N)}) = \frac{\sqrt{MN}\Gamma\left(\frac{M+N-1}{2}\right) z^{\frac{N-1}{2}} y^{\frac{M-3}{2}}}{\sqrt{M+N}\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{N-1}{2}\right)\Gamma\left(\frac{M-1}{2}\right)} \cdot \left[z + y + \frac{NM}{N+M}(x-\bar{x}_N)^2\right]^{-\frac{M+N-1}{2}}.$$

However, the distribution of the function of X and Y within the parentheses on the left-hand-side of the greater than or equal sign in (4.1) is fairly complex and is not readily susceptible to being tabled. Hence as a reasonable approximation for P_p for sufficiently large N, we shall approximate s_{N+M}^2 by known s_N^2 so that

$$S_\nu\left(\frac{(\mu_0-\bar{x}_{N+M})\sqrt{N+M}}{s_N}\right) \doteq S_\nu\left(\frac{(\mu_0-\bar{x}_{N+M})\sqrt{N+M}}{s_{N+M}}\right).$$

Then calculate

$$\begin{aligned} P_p &\doteq \Pr\left(\frac{(\mu_0-\bar{x}_{N+M})\sqrt{N+M}}{s_N} \geq S_\nu^{-1}(p)\right) \\ &\doteq \Pr\left(\frac{(\mu_0-\bar{x}_N)\sqrt{N+M} + (\bar{x}_N - X)\frac{M}{\sqrt{N+M}}}{s_N} \geq S_\nu^{-1}(p)\right) \\ &\doteq 1 - S_{N-1}\left(\left(\frac{N}{M}\right)^{1/2}\left(S_\nu^{-1}(p) + \frac{(N+M)^{1/2}(\bar{x}_N-\mu_0)}{s_N}\right)\right). \end{aligned}$$

This should serve as an adequate approximation for $N \geq 25$, until computing algorithms of the distribution function involved in (4.1) can be easily managed.

5. Multivariate Normal Observables. Let $X_i, i = 1, 2, \ldots, n$ be d-dimensional and i.i.d. $N(\mu, \Sigma)$. Suppose for some d-dimensional region R_0^d, we are testing

$$H_0: \mu \in R_0^d \quad \text{vs.} \quad H_1: \mu \notin R_0^d,$$

and we reject H_0 if at $n = N + M$ observations,

$$\Pr[\mu \notin R_0^d | x^{(n)}] \geq p, \tag{5.1}$$

calculated from the posterior distribution of μ, where $x^{(n)} = (x_1, \ldots, x_n)$. Assuming $p(\mu, \Sigma^{-1}) \propto |\Sigma|^{(d+1)/2}$ and $(n-1)S_n = \sum_{i=1}^n (x_i - \bar{x}_n)(x_i - \bar{x}_n)'$, we obtain that the posterior distribution of

$$\mu \sim S\left(n-d, \bar{x}_n, \frac{(n-1)}{n}S_n\right),$$

i.e. a d-variate student distribution whose density is defined as

$$f(x) \propto (1 + (x-\theta)'\Lambda^{-1}(x-\theta))^{-\frac{\alpha+d}{2}}$$

so that $X \sim S(\alpha, \theta, \Lambda)$, Geisser and Cornfield (1963). Stopping at N we need to compute

$$\Pr\left[\Pr\left[(\mu \notin R_0^d \mid x^{(N)}, X_{(M)}\right] \geq p\right]$$

where $X_{(M)} = (X_{N+1}, \ldots, X_{N+M})$ is now random. Now it is clear

$$\Pr\left[\mu \notin R_0^d \mid x^{(N)}, X_{(M)}\right] = U(X_{(M)} \mid x^{(N)}),$$

say, is a scalar random variable so that we are required to find

$$\Pr[U \geq p] = P_p.$$

An important application is when R_0^d is a hyperrectangle or semi-infinite hyperrectangle. Here simulation appears to be the simplest method of calculating P_p if the dimension d is not large.

Another particular application which may be of some interest is the "distance" between two populations or the "distance" of a population from some specified d-dimensional vector, say μ_0. Let γ be the distance of the population from μ_0 so that

$$\gamma = (\mu - \mu_0)'\Sigma^{-1}(\mu - \mu_0).$$

Interest can focus on whether this normed difference of μ from some μ_0 is less than some given distance. A similar situation can be defined for two populations. Further suppose we are interested in testing $H_0 : \gamma \leq \gamma_0$ vs. $H_1 : \gamma > \gamma_0$. Now for $n = N+M$ observations and Σ known, $n\gamma \sim \chi_d^2(\lambda)$ where $\lambda = n(\bar{x}_n - \mu_0)'\Sigma^{-1}(\bar{x}_n - \mu_0)$. Further $\nu\lambda/T^2 \sim \chi_\nu^2$ for $\Sigma^{-1} \sim W(\nu, \nu^{-1}S_n^{-1})$ i.e. Wishart distributed, for

$$\nu = n - 1, \qquad \nu S_n = \sum_{i=1}^{n}(x_i - \bar{x}_n)(x_i - \bar{x}_n)',$$

and

$$T^2 = n(\bar{x}_n - \mu_0)'S_n^{-1}(\bar{x}_n - \mu_0).$$

Now we can find the posterior density,

$$p(\gamma \mid T^2) = \int p(\gamma \mid \lambda, T^2)p(\lambda \mid T^2)\, d\lambda$$

and it is possible to show that

$$p_{T^2}(n\gamma | x^{(n)}) = \sum_{j=0}^{\infty} w_j f_{d+2j}(n\gamma),$$

where $f_{d+2j}(n\gamma)$ is the density of a chi-squared random variable with $d+2j$ degrees of freedom and

$$w_j = \binom{\frac{\nu}{2}+j-1}{j}\left(\frac{\nu}{T^2+\nu}\right)^{\frac{\nu}{2}}\left(\frac{T^2}{T^2+\nu}\right)^j,$$

Geisser (1967).

Note that as n grows,

$$w_j \to \left(\frac{T^2}{2}\right)^j \frac{e^{-\frac{T^2}{2}}}{j!}.$$

Hence $n\gamma$ will tend to a non-central chi-squared variate with d degrees of freedom and non-centrality parameter T^2.

To reject H_0 we require $\Pr[n\gamma > n\gamma_0 \mid x^{(N+M)}] \geq p$ or $1-F(n\gamma_0 \mid T^2) \geq p$. Now if we stop at N we need to calculate $P_p = \Pr[1 - F(n\gamma_0|T^2) \geq p]$ for $n = N+M$. Because we can show that $1 - F((N+M)\gamma_0|T^2)$ is increasing in T^2, we need to find the minimum T^2, say t_0^2 , such that

$$1 - F((N+M)\gamma_0 \mid t_0^2) \geq p$$

and then

$$P_p = \Pr(T^2 > t_0^2).$$

To demonstrate the monotonicity property we note that

$$1 - F(y|T^2) = \Pr(n\gamma > y) = \sum_{j=0}^{\infty} w_j G_{d+2j}(y)$$

where G_{d+2j} is the distribution function corresponding to $f_{d+2j}(n\gamma)$. Define

$$c_j = \binom{\frac{\nu}{2}+j-1}{j}, \qquad \eta = \left(\frac{T^2}{T^2+\nu}\right).$$

Then it suffices to establish monotonicity in η. Now it is easy to show that

$$\frac{dw_j}{d\eta} = \left(\frac{\nu}{2}+j-1\right)w_{j-1} - \frac{\nu}{2}(1-\eta)^{-1}w_j$$

so that after some algebraic manipulation

$$\frac{d}{d\eta}[1 - F(y|T^2)] \propto \sum_{j=0}^{\infty} w_j(G_{d+2+2j}(y) - G_{d+2j}(y)) + \sum_{j=0}^{\infty} w_j \frac{(2j-T^2)}{\nu+T^2} G_{d+2+2j}(y). \tag{5.2}$$

Now we will show that for any integer k

$$G_{k+2}(y) > G_k(y) \tag{5.3}$$

which will imply that the first term in (5.2) is non-negative. Let

$$G_{k+2}(y) - G_k(y) = h(y) = \int_y^\infty \left(\frac{x}{k} - 1\right) f_k(x)\,dx.$$

Clearly $h(y) > 0$ if $y > k$. Furthermore,

$$\frac{dh(y)}{dy} = -\left(\frac{y}{k} - 1\right) f_k(y) > 0$$

if and only if $y < k$. Thus $h(y)$ is monotonically increasing for $y \le k$ and then monotonically decreasing. Because $h(0) = 0$ it then follows that $h(y) > 0$ for all $0 < y < k$ and (5.3) is established. Hence the first term in (5.2) is positive for $y > 0$.

We now consider the second term in (5.2). Define $g(\cdot)$ to be a differentiable increasing version of $G_{d+2j}(y)$ so that $g(c)$ is differentiable, increasing and $g(c) = G_{d+2j}(y)$ provided that $c = j$. Then the second term of (5.2) is proportional to

$$\sum_{j=0}^{\infty} w_j \left(j - \frac{T^2}{2}\right) g(j). \tag{5.4}$$

A first order Taylor expansion about $T^2/2$ yields

$$g(j) = g\left(\frac{T^2}{2}\right) + g'(j^*)\left(j - \frac{T^2}{2}\right)$$

for $j^* \in (j, T^2/2)$. Expression (5.4) then becomes

$$g\left(\frac{T^2}{2}\right)\sum_{j=0}^{\infty} w_j \left(j - \frac{T^2}{2}\right) + \sum_{j=0}^{\infty} w_j g'(j^*)\left(j - \frac{T^2}{2}\right)^2.$$

But $g'(j^*) > 0$ and

$$\sum_{j=0}^{\infty} j\, w_j = \frac{T^2}{2}.$$

Hence the second term in (5.2) is also positive and the monotonicity property is established.

6. An Approximation to the Predictive Distribution of T^2. Now T^2 depends on the random variables $\bar{X}_M$ and S_M since

$$T^2 = (N+M)\left(\frac{N\bar{x}_N + M\bar{X}_M}{N+M} - \mu_0\right)' S_{N+M}^{-1}\left(\frac{N\bar{x}_N + M\bar{X}_M}{N+M} - \mu_0\right),$$

where

$$(N+M-1)S_{N+M} = (N-1)S_N+(M-1)S_M+\frac{NM}{M+N}(\overline{X}_M-\bar{x}_N)(\overline{X}_M-\bar{x}_N)'.$$

The joint predictive density of $\bar{X}_M = X$ and $(M-1)S_M = Y$ is easily found to be

$$f(x,y\,|\,x^{(N)}) \propto |y|^{\frac{M-d-2}{2}} \left|\frac{NM}{N+M}(x-\bar{x}_N)(x-\bar{x}_N)' + y + z\right|^{-\frac{N+M-1}{2}},$$

for $z = (N-1)S_N$. However, the calculation of the exact density of T^2 is even less tractable from the above than in the univariate case i.e. $d = 1$, discussed in Section 4.

Define

$$V = \frac{N\bar{x}_N + M\overline{X}_M}{N+M},$$

so that

$$T^2 = (N+M)(V-\mu_0)'S_{N+M}^{-1}(V-\mu_0).$$

As in the univariate case we will approximate S_{N+M} by S_N thus eliminating the random matrix S_M and alter T^2 to

$$\widehat{T}^2 = (N+M)(V-\mu_0)'S_N^{-1}(V-\mu_0)$$

and derive the density of $\widehat{T}^2$. Define $Q = \frac{M}{N(N+M)}S_N, q = N-d$. Then

$$V - \bar{x}_N \sim S(q, 0, (N-1)Q).$$

Now given $x^{(N)}$ consider the random vector $W(q/U)^{1/2}$, where W is $N(0,Q)$ and independent of U which is χ^2_{N-d}. Let $\delta = \bar{x}_N - \mu_0$. Then the vector $V - \mu_0$ is distributed as $W(q/U)^{1/2} + \delta$. Hence $\widehat{T}^2$ is distributed as

$$(N+M)\left(W + \left(\frac{U}{q}\right)^{1/2}\delta\right)' S_N^{-1}\left(W + \left(\frac{U}{q}\right)^{1/2}\delta\right)\left(\frac{q}{U}\right)^{1/2}.$$

Conditional on $U, W + (\frac{U}{q})^{1/2}\delta \sim N((\frac{U}{q})^{1/2}\delta, Q)$ so that

$$\left(W + \left(\frac{U}{q}\right)^{1/2}\delta\right)' Q^{-1}\left(W + \left(\frac{U}{q}\right)^{1/2}\delta\right) \sim \chi^2_d\left(\frac{\delta' Q^{-1}\delta U}{q}\right)$$

i.e. non-central chi-square with d degrees of freedom and non-centrality parameter

$$\delta' Q^{-1}\delta U/q \equiv DU.$$

Thus conditional on U

$$\frac{N}{M}\widehat{T}^2 = A \sim \chi^2_d(DU)/U.$$

The predictive density of A is then

$$\begin{aligned} f(a|x^{(N)}) &= \int_0^\infty f_U(u|x^{(N)}) f(ua|x^{(N)}, u) u\, du \\ &= \sum_{k=0}^\infty \frac{\Gamma\left(2k+\frac{N}{2}\right)}{k!\Gamma\left(k+\frac{d}{2}\right)\Gamma\left(\frac{q}{2}\right)} \left(\frac{D}{1+D}\right)^k \left(\frac{1}{1+D}\right)^{k+\frac{N}{2}} \\ &\quad \cdot a^{k+\frac{d}{2}-1}\left(1+\frac{a}{1+D}\right)^{-(2k+\frac{N}{2})} . \end{aligned} \tag{6.1}$$

From (6.1) it is clear that $A(1+D)^{-1} = B$ has density

$$f(b\,|\,x^{(N)}) = \sum_{k=0}^\infty w_k f\left(b|k+\frac{d}{2}, k+\frac{q}{2}\right),$$

an infinite sum of beta variates, i.e.

$$f\left(b|k+\frac{d}{2}, k+\frac{q}{2}\right) \propto b^{k+\frac{d}{2}-1}(1+b)^{-(2k+\frac{d+q}{2})} \tag{6.2}$$

with negative binomial weights, where

$$w_k = \binom{k+\frac{q}{2}-1}{k}\left(\frac{D}{1+D}\right)^k\left(\frac{1}{1+D}\right)^{q/2}.$$

Hence

$$B = \frac{N\widehat{T}^2}{M(N-d)(1+D)}$$

so that

$$\begin{aligned} P_p &= \Pr(T^2 > t_0^2) \doteq \Pr(\widehat{T}^2 > t_0^2) \\ &= \Pr\left(B > \frac{Nt_0^2}{M(N-d)(1+D)}\right) = \widehat{P}_p \end{aligned}$$

which can be numerically calculated to reasonable accuracy.

7. Monte Carlo comparisons of T^2 and $\widehat{T}^2$ and P_p and $\widehat{P}_p$. We have provided an approximation for the solution of this problem that can be numerically calculated. The question now is how good this approximation is with regard to the exact P_p. To have an idea, we consider a subset of the Iris data of Fisher (1936). The full set consists of 4 variables yielding measurements on sepal and petal widths and lengths in centimeters on 150 plants, 50 each from 3 different species of irises, setosa, versicolor and virginica. For the sake of illustration, we only consider two of the variables, sepal and petal width on each plant, of species Iris versicolor. We first use petal width and sepal width together as bivariate data and then separately as univariate data to check the approximations of $\widehat{T}^2$ to T^2 and $\widehat{P}_p$ to P_p. The latter are calculated using Monte Carlo simulations of 10,000 repetitions for sample sizes $N = 25, 50$ and future samples of size $M = 25$, for varying μ_0 and γ_0 and p. In each figure the graph in the upper left hand corner compares the distribution function of $\widehat{T}^2$ with that of T^2 while the other three compare $\widehat{P}_p$ with P_p. The graphs for the bivariate case are given in Figures 1 through 4, and for the univariate case in Figures 5 through 8.

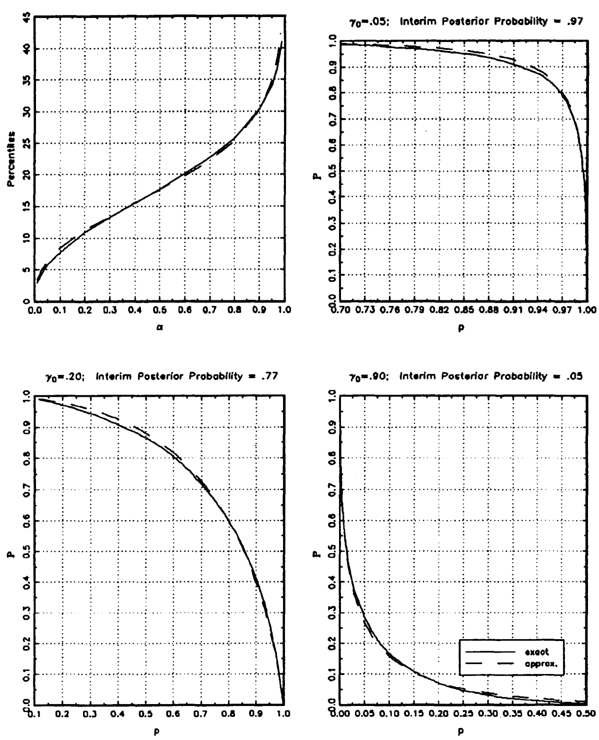

Figure 1 SP-Width Data: N=M=25, μ_0=(2.7,1.4), T_N^2=8.2

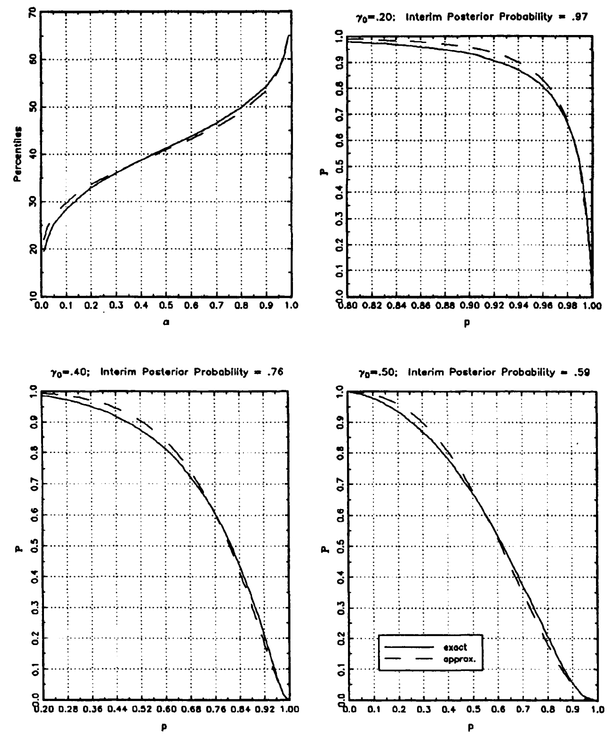

Figure 2 SP Data: N=50, M=25, μ_0=(2.7,1.4), T_N^2=26.9

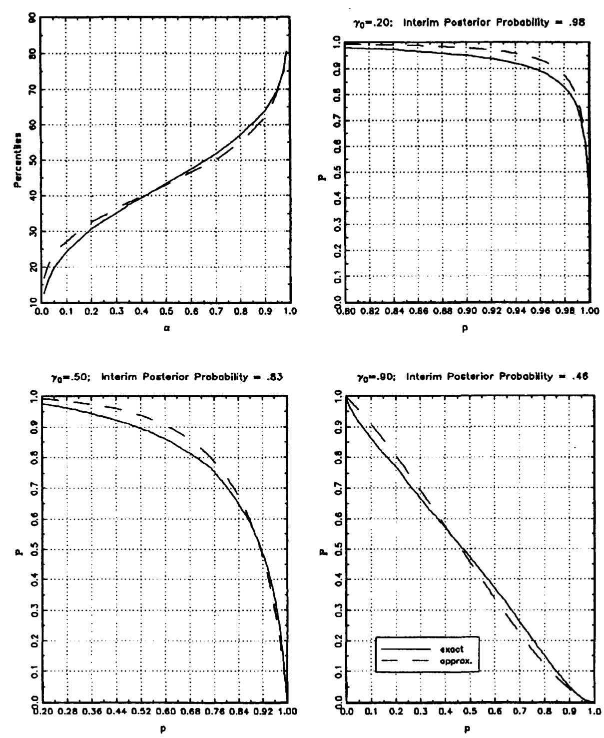

Figure 3 SP-Width Data: N=M=25, μ_0=(2.6,1.4), T_N^2=21.0

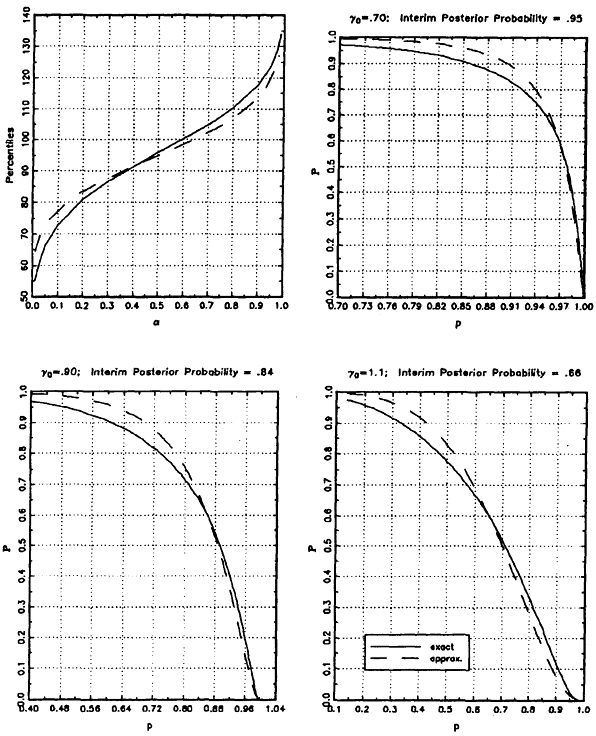

Figure 4 SP Data: N=50, M=25, μ_0=(2.6,1.4), T_N^2=62.9

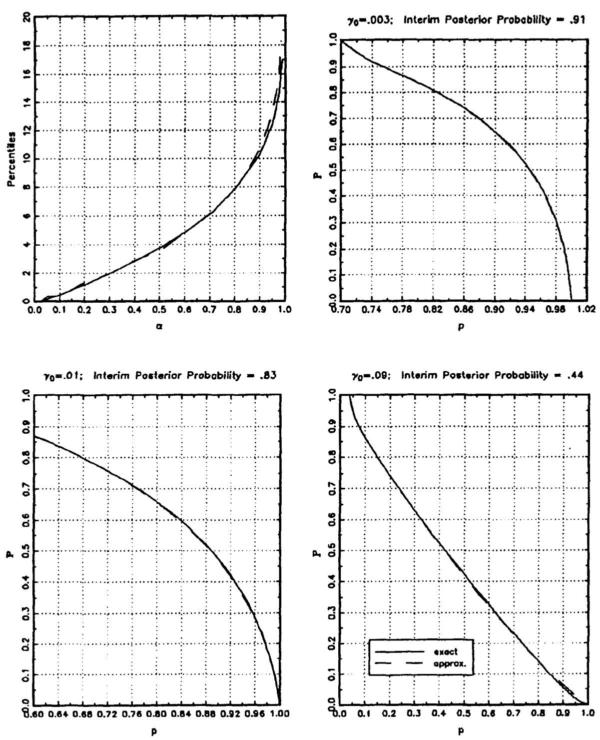

Figure 5 P-Width Data: N=25, M=25, μ_0=(1.4), T_N^2=1.8

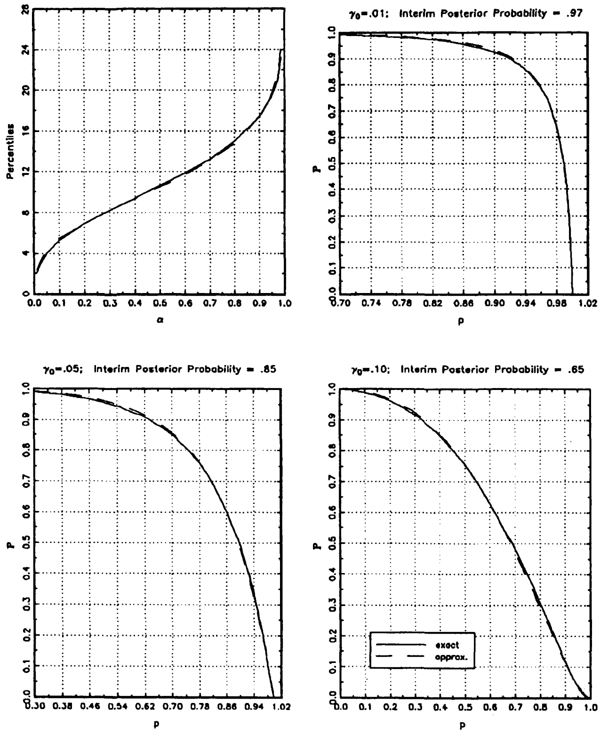

Figure 6 P-Width Data: N=50, M=25, μ_0=(1.4), T_N^2=7.0

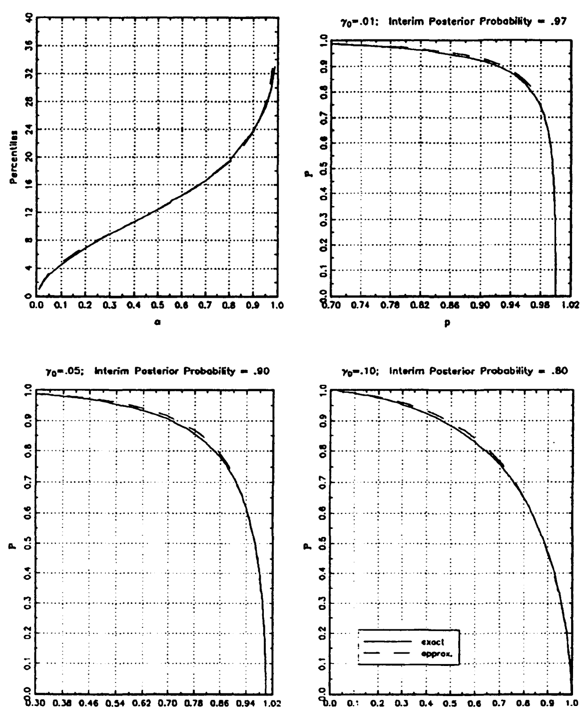

Figure 7 S-Width Data: N=25, M=25, μ_0=(2.6), T_N^2=6.2

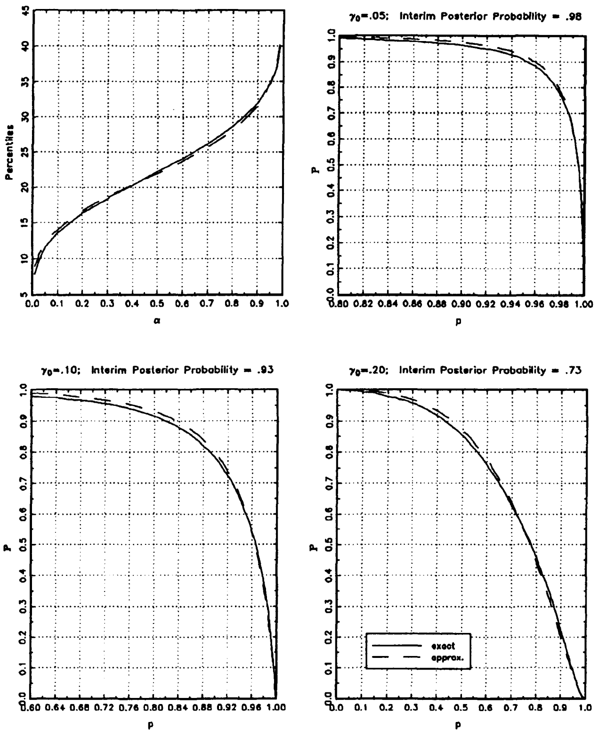

Figure 8 S-Width Data: N=50, M=25, μ_0=(2.6), T_N^2=14.7

The approximations appear to be quite good in all situations for which T_N^2, the interim value of T^2, is small or moderate. This is true for all the univariate cases considered and for the bivariate case illustrated in Figure 1 with $N = M = 25$ and $\mu_0 = (2.7, 1.4)$. We note that

$$\overline{X}_N = \begin{pmatrix} 1.326 \\ 2.770 \end{pmatrix} \qquad S_N = \begin{pmatrix} 0.0985 & 0.0412 \\ 0.0412 & 0.0391 \end{pmatrix}.$$

The cases illustrated in Figures 2–4 indicate that the error in approximation can range up to 10%. For example, in the cases depicted by Figure 4, if γ_0 were set at 0.70 and p were set at 0.91, we obtain $P_p = 0.83$ while $\widehat{P}_p = 0.89$. However, we note that in this case, if we consider the null sampling distribution of T_N^2 (i.e. the usual Hotelling T^2 distribution), it will exceed its observed value of 62.9 with probability 8×10^{-14}. The value $\gamma_0 = 0.70$ was chosen to give an interim posterior probability of 0.95 and is probably larger than one would choose in practice. Values of γ_0 smaller than 0.3 correspond to interim posterior probabilities greater than 0.9994 and result in values of P_p and $\widehat{P}_p$ that are nearly 1 for all choices of p. For example, when $\gamma_0 = 0.3$, and $p = 0.99$, we obtain $\widehat{P}_p = .999$ and $P_p = 0.99$. This is the worst case for $p \le 0.99$. In passing, we point out that the tail probabilities calculated from the null sampling distribution for cases 1–8 are .004, $3 \times 10^{-8}, 1.1 \times 10^{-5}, 8 \times 10^{-14}$, 0.19, 0.02, 0.01 and 0.00035, respectively, indicating that the approximation works well when $\overline{X}_N$ is within a reasonable distance of μ_0 while it fails to be precise when it is not.

The reason that the approximation fails to be precise when T_N^2 is large is due to the fact that the distribution of T^2 has a noticeably fatter left tail than that of $\widehat{T}^2$ in this instance, as can be seen in the figures. The right hand tail can also be fatter as indicated in Figure 4. We point out that the mean of T^2 and that of $\widehat{T}^2$ were consistently very close to one another across all cases considered. In any event, suppose the value t_0^2 corresponding to p, namely such that

$$1 - F_{t_0^2}((N+M)\gamma_0) = p,$$

corresponded to the first percentile of $\widehat{T}^2$, say $\widehat{T}^2_{0.01}$. Then if the left tail is fatter for T^2 than it is for $\widehat{T}^2$, we must have

$$\widehat{P}_p = 0.99 > P_p.$$

Similarly if, for given p, t_0^2 corresponds to $\widehat{T}^2_{0.99}$, we must have

$$\widehat{P}_p = 0.01 < P_p.$$

This is illustrated in Figures 2–4. Further study is necessary to see if the distribution of $\widehat{T}^2$ can be adjusted appropriately so that it better mimics the tails of the distribution of T^2 when T_N^2 is large. However, our simulations suggest that when the given value of T_N^2 is such that the chance of it being exceeded is not miniscule using the null sampling distribution, the suggested approximation is quite adequate.

8. Remarks. As noted before, this work can be adapted to the two population problem. However, the case of k populations presents further complications and will be the scope of further work including interim analysis in multivariate normal classification problems.

REFERENCES

CHOI, S. C. and PEPPLE, P. A. (1989). Monitoring clinical trials based on predictive probability of significance. *Biometrics* **45**, 317–323.

CHOI, S. C., SMITH, P.J. and BECKER, D.P. (1985). Early decision in clinical trials when treatment differences are small. *Controlled Clinical Trials* **6**, 280–288.

FISHER, R. A. (1936). The use of multiple measurements in toxonomic problems. *Annals of Eugenics* Vol VII; Pt. II, 179–188.

GEISSER, S. (1967). Estimation associated with linear discriminants. *Annals of Mathematical Statistics* **38** 3, 807–817.

GEISSER, S. (1992). On the curtailment of sampling. *Canadian Journal of Statistics*, **20**, 3, 297–309.

GEISSER, S. and CORNFIELD, J. (1963). Posterior distributions for multivariate normal parameters. *Journal of the Royal Statistical Society B*, **25**, 368–376.

SPIEGELHALTER, D. J., FREEDMAN, D. S., and BLACKBURN, P. R. (1986). Monitoring clinical trials: Conditional or predictive power? *Controlled Clinical Trials*, **7**, 8–17.

SCHOOL OF STATISTICS
UNIVERSITY OF MINNESOTA
270 VINCENT HALL, 206 CHURCH STREET S.E.
MINNEAPOLIS, MN 55455-0488
USA

DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA AT DAVIS
KERR HALL
DAVIS, CA 95616
USA