Stat 513
Emerson, Winter 04
Homework #7 Key
March 12, 2004

Supplemental written problems due Monday, March 8, 2004 at the beginning of class.

Let $X_i, i = 1, \ldots, n$ be i.i.d. random variables with mean μ and variance $\sigma^2 < \infty$, and let $Y_i, i = 1, \ldots, m$ be i.i.d. random variables with mean ν and variance $\tau^2 < \infty$. We shall assume that as $n \to \infty$, the ratio $n/(m+n) \to \lambda$.

1. Suppose the X_i's are normally distributed. Use Basu's theorem to prove that the sample mean $\overline{X}$ is independent of the sample variance s_X^2. (Hint: Consider first the case where σ^2 is known. Examine the complete sufficient statistic for μ and the distribution of s_X^2. Then argue that the independence of the sample mean and the sample variance will not be altered by the lack of knowledge of σ^2.)

Ans: For σ^2 known, the density of $\vec{X}$ is easily shown to be 1 parameter exponential family

$$\begin{aligned} f_{\vec{X}}(\vec{x}) &= \exp\left(-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - \mu)^2\right) \\ &= \exp\left(-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{\sum_{i=1}^n X_i^2}{2\sigma^2} + \frac{\mu\sum_{i=1}^n X_i}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}\right) \\ &= \exp\left(c_0(\mu) + T_0(\vec{X}) + c_1(\mu)T_1(\vec{X})\right), \end{aligned}$$

with $c_0(\mu) = -(n/2)\log(2\pi\sigma^2) - n\mu/(2\sigma^2)$, $T_0(\vec{X}) = \sum_{i=1}^n X_i^2/(2\sigma^2)$, $c_1(\mu) = n\mu/\sigma^2$, and $T_1(\vec{X}) = \overline{X}_n$. Since the parameter space for μ contains a 1 dimensional open interval, we know that $T_1(\vec{X})$ is complete sufficient for μ.

Because the X_i's are independent identically distributed random variables, we know that $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$, a distribution that does not depend on μ. Hence, s^2 is an ancillary statistic. By Basu's theorem we know that complete sufficient statistics are independent of ancillary statistics, so s^2 is independent of $\overline{X}_n$ in this setting of normal data. Now, if σ^2 is unknown, that will not change the computation or distribution of either s^2 or $\overline{X}$, so the independence of these two statistics in the setting of normally distributed data holds in general.

2. Now suppose that the X_i's have the Bernoulli distribution. Show that the sample mean and sample variance are not independent in this problem.

Ans: Because $X_i = X_i^2$ with binary data, we find the formula for the sample variance

$$\begin{aligned} s^2 &= \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2\right) \\ &= \frac{n}{n-1}\left(\overline{X}_n - \overline{X}_n^2\right) \\ &= \frac{n}{n-1}\overline{X}_n\left(1 - \overline{X}_n\right). \end{aligned}$$

If $\overline{X}_n$ and s^2 are independent, we should find that $E[s^2|\overline{X}_n = x] = E[s^2]$ for all values of x. Now, s^2 is the nonparametric estimator of the variance, so $E[s^2] = p(1-p)$. However, when $X_n = 0$, we also know that $s^2 = 0$, so $E[s^2|\overline{X}_n = 0] = 0 \not\equiv p(1-p)$. Hence, the sample mean and sample variance are not independent for $p \in (0,1)$.

3. Suppose the X_i's and Y_i's are normally distributed.

a. Find the likelihood ratio test for $H_0 : \mu = \nu$ versus $H_1 : \mu \neq \nu$ when $\sigma^2 = \tau^2$.

Ans: Let $\vec{\theta} = (\mu, \nu, \sigma^2)$. We first derive the necessary quantities for likelihood based inference. In the general case (where $\mu, \nu \in \mathcal{R}^1$ and $\sigma^2 > 0$), we find the likelihood and log likelihood functions as

$$L(\vec{\theta}\,|\,\vec{X}, \vec{Y}) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n+m} \exp\left(-\frac{\sum_{i=1}^n (X_i - \mu)^2 + \sum_{i=1}^m (Y_i - \nu)^2}{2\sigma^2}\right)$$
$$\mathcal{L}(\vec{\theta}) = -\frac{n+m}{2}\log(2\pi) - \frac{n+m}{2}\log(\sigma^2) - \frac{\sum_{i=1}^n (X_i - \mu)^2 + \sum_{i=1}^m (Y_i - \nu)^2}{2\sigma^2}$$

The elements of the score vector are thus

$$\frac{\partial}{\partial\mu}\mathcal{L}(\vec{\theta}) = \sum_{i=1}^n \frac{(X_i - \mu)}{\sigma^2}$$
$$\frac{\partial}{\partial\nu}\mathcal{L}(\vec{\theta}) = \sum_{i=1}^m \frac{(Y_i - \nu)}{\sigma^2}$$
$$\frac{\partial}{\partial\sigma^2}\mathcal{L}(\vec{\theta}) = -\frac{n+m}{2\sigma^2} + \frac{\sum_{i=1}^n (X_i - \mu)^2 + \sum_{i=1}^m (Y_i - \nu)^2}{2\sigma^4}$$

The second partials of the log likelihood function are

$$\frac{\partial^2}{\partial\mu^2}\mathcal{L}(\vec{\theta}) = -\frac{n}{\sigma^2}$$
$$\frac{\partial^2}{\partial\mu\partial\nu}\mathcal{L}(\vec{\theta}) = 0$$
$$\frac{\partial^2}{\partial\mu\partial\sigma^2}\mathcal{L}(\vec{\theta}) = -\sum_{i=1}^n \frac{(X_i - \mu)}{\sigma^4}$$
$$\frac{\partial^2}{\partial\nu^2}\mathcal{L}(\vec{\theta}) = -\frac{m}{\sigma^2}$$
$$\frac{\partial^2}{\partial\nu\partial\sigma^2}\mathcal{L}(\vec{\theta}) = -\sum_{i=1}^m \frac{(Y_i - \nu)}{\sigma^4}$$
$$\frac{\partial^2}{\partial(\sigma^2)^2}\mathcal{L}(\vec{\theta}) = \frac{n+m}{2\sigma^4} + \frac{\sum_{i=1}^n (X_i - \mu)^2 - \sum_{i=1}^m (Y_i - \nu)^2}{\sigma^6}$$

Taking the negative expectation of those second partial derivatives results in the information matrix

$$I(\vec{\theta}) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 & 0 \\ 0 & \frac{m}{\sigma^2} & 0 \\ 0 & 0 & \frac{n+m}{2\sigma^4} \end{pmatrix}$$

Now we maximize the log likelihood in the unconstrained problem by setting $\mathcal{U}(\hat{\vec{\theta}}) = 0$. We thus find maximum likelihood estimates

$$\hat{\mu} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$$
$$\hat{\nu} = \overline{Y}_n = \frac{1}{m}\sum_{i=1}^n Y_i$$
$$\hat{\sigma}^2 = \frac{n\hat{\sigma}_X^2 + m\hat{\sigma}_Y^2}{n+m} = \frac{\sum_{i=1}^n (X_i - \hat{\mu})^2 + \sum_{i=1}^m (Y_i - \hat{\nu})^2}{n+m}$$

When we maximize the loglikelihood under the constraint that $\mu = \nu = \omega$, we use log likelihood function

$$\mathcal{L}(\vec{\theta}) = -\frac{n+m}{2}\log(2\pi) - \frac{n+m}{2}\log(\sigma^2) - \frac{\sum_{i=1}^n (X_i - \omega)^2 + \sum_{i=1}^m (Y_i - \omega)^2}{2\sigma^2}$$

yielding score vector

$$\begin{aligned} \frac{\partial}{\partial \omega}\mathcal{L}(\vec{\theta}) &= \frac{\sum_{i=1}^n (X_i - \omega) + \sum_{i=1}^m (Y_i - \omega)}{\sigma^2} \\ \frac{\partial}{\partial \sigma^2}\mathcal{L}(\vec{\theta}) &= -\frac{n+m}{2\sigma^2} + \frac{\sum_{i=1}^n (X_i - \omega)^2 + \sum_{i=1}^m (Y_i - \omega)^2}{2\sigma^4} \end{aligned}$$

In this constrained setting (and using the notation of $\overline{X}_n$, $\overline{Y}_m$, $\hat{\sigma}_X^2$, and $\hat{\sigma}_Y^2$ as implicitly defined above), we find maximum likelihood estimates

$$\begin{aligned} \hat{\mu}_0 &= \hat{\omega} = \frac{n\overline{X}_n + m\overline{Y}_m}{n+m} \\ \hat{\nu}_0 &= \hat{\omega} = \frac{n\overline{X}_n + m\overline{Y}_m}{n+m} \\ \hat{\sigma}_0^2 &= \frac{\sum_{i=1}^n (X_i - \hat{\mu}_0)^2 + \sum_{i=1}^m (Y_i - \hat{\nu}_0)^2}{n+m} \\ &= \frac{n\hat{\sigma}_X^2 + m\hat{\sigma}_Y^2 + mn(\overline{X}_n - \overline{Y}_m)^2/(m+n)}{m+n} \\ &= \hat{\sigma}^2 + \frac{mn}{(m+n)^2}(\overline{X}_n - \overline{Y}_m)^2 \end{aligned}$$

Now the log likelihood evaluated at the unconstrained MLE is

$$\begin{aligned} \mathcal{L}(\hat{\vec{\theta}}) &= -\frac{n+m}{2}\log(2\pi) - \frac{n+m}{2}\log(\hat{\sigma}^2) - \frac{\sum_{i=1}^n (X_i - \hat{\mu})^2 + \sum_{i=1}^m (Y_i - \hat{\nu})^2}{2\hat{\sigma}^2} \\ &= -\frac{n+m}{2}\left(\log(\hat{\sigma}^2) + 1\right) \end{aligned}$$

and the log likelihood evaluated at the constrained MLE is

$$\begin{aligned} \mathcal{L}(\hat{\vec{\theta}}_0) &= -\frac{n+m}{2}\log(2\pi) - \frac{n+m}{2}\log(\hat{\sigma}_0^2) - \frac{\sum_{i=1}^n (X_i - \hat{\mu}_0)^2 + \sum_{i=1}^m (Y_i - \hat{\nu}_0)^2}{2\hat{\sigma}_0^2} \\ &= -\frac{n+m}{2}\left(\log(\hat{\sigma}_0^2) + 1\right) \end{aligned}$$

The likelihood ratio test rejects H_0 when the likelihood ratio is large, or for some appropriate k

$$\mathcal{L}(\hat{\vec{\theta}}) - \mathcal{L}(\hat{\vec{\theta}}_0) = -\frac{m+n}{2}\log\left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right) > k$$

or equivalently $$\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} > k_1 = \exp\left(\frac{2k}{n+m}\right)$$

or equivalently $$1 + \frac{mn}{(m+n)^2}\frac{(\overline{X}_n - \overline{Y}_m)^2}{\hat{\sigma}^2} > k_1$$

or equivalently $$\frac{mn}{(m+n)}\frac{(\overline{X}_n - \overline{Y}_m)^2}{\hat{\sigma}^2} > k_2 = (m+n)(k_1 - 1)$$

or equivalently $$T^2 = \frac{mn}{(m+n)}\frac{(\overline{X}_n - \overline{Y}_m)^2}{s_p^2} > k_3 = \frac{m+n-2}{m+n}k_2$$

where the pooled sample variance $s_p^2 = (m+n)\hat{\sigma}^2/(m+n-2)$. Under the null hypothesis, the test statistic T^2 has an F distribution with 1 and $m+n-2$ degrees of freedom, so the value of k_3 would be the $(1-\alpha)$th quantile of that distribution.

b. Show that the Wald and score tests are equivalent to the LR test in this problem.

Ans: The Wald statistic would be based on the contrast $\vec{a}^T\hat{\vec{\theta}}$, where $\vec{a}^T = (1, -1, 0)$. Now $\vec{a}^T\hat{\vec{\theta}} = \overline{X}_n - \overline{Y}_m$, which is easily shown to be normally distributed with distribution

$$\overline{X}_n - \overline{Y}_m \sim \mathcal{N}\left(\mu - \nu, \sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)\right).$$

We thus compute the quadratic form to test $H_0 : \mu - \nu = 0$ as

$$Z^2 = \frac{mn}{(m+n)}\frac{(\overline{X}_n - \overline{Y}_m)^2}{\sigma^2} \sim \chi^2_1$$

Because σ^2 is unknown, we note the following facts

- $(n+m-2)s_p^2/\sigma^2 \sim \chi^2_{n-m-2}$,
- s_p^2 is independent of $\overline{X}_n$ and $\overline{Y}_m$ in this normal model, and
- if $V \sim \chi^2_k$ and $W \sim \chi^2_\ell$ are independent random variables, then

$$F = \frac{V/k}{W/\ell} \sim F_{k,\ell},$$

  the F distribution with k and ℓ degrees of freedom (this is the definition of the F distribution, which was invented exactly for this setting).

We thus can base inference on

$$T^2 = \frac{\sigma^2}{s_p^2}Z^2 = \frac{mn}{(m+n)}\frac{(\overline{X}_n - \overline{Y}_m)^2}{s_p^2},$$

the exact same statistic used for the likelihood ratio tests.

The score statistic is computed as

$$\mathcal{U}^T(\hat{\vec{\theta}}_0)I^{-1}(\hat{\vec{\theta}}_0)\mathcal{U}(\hat{\vec{\theta}}_0)$$

Now

$$\mathcal{U}(\hat{\vec{\theta}}_0) = \begin{pmatrix} \frac{n\overline{X}_n - n\hat{\omega}}{\hat{\sigma}_0^2} \\ \frac{m\overline{Y}_m - m\hat{\omega}}{\hat{\sigma}_0^2} \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{nm(\overline{X}_n - \overline{Y}_m)}{\hat{\sigma}_0^2} \\ -\frac{nm(\overline{X}_n - \overline{Y}_m)}{\hat{\sigma}_0^2} \\ 0 \end{pmatrix}$$

and the inverse of the information matrix evaluated at $\hat{\vec{\theta}}_0$ is

$$I^{-1}(\hat{\vec{\theta}}_0) = \begin{pmatrix} \frac{\hat{\sigma}_0^2}{n} & 0 & 0 \\ 0 & \frac{\hat{\sigma}_0^2}{m} & 0 \\ 0 & 0 & \frac{2\hat{\sigma}_0^4}{n+m} \end{pmatrix}$$

So the score test would reject H_0 when

$$\frac{n^2m^2(\overline{X}_n - \overline{Y}_m)^2}{\hat{\sigma}_0^4}\frac{\hat{\sigma}_0^2}{n} + \frac{n^2m^2(\overline{Y}_m - \overline{X}_n)^2}{\hat{\sigma}_0^4}\frac{\hat{\sigma}_0^2}{m} > k$$

or equivalently $$\frac{nm(n+m)(\overline{X}_n - \overline{Y}_m)^2}{\hat{\sigma}_0^2} > k$$

or equivalently $$\frac{\hat{\sigma}_0^2}{(\overline{X}_n - \overline{Y}_m)^2} < k_1 = \frac{nm(m+n)}{k}$$

or equivalently $$\frac{\hat{\sigma}^2 + mn(\overline{X}_n - \overline{Y}_m)^2/(m+n)}{(\overline{X}_n - \overline{Y}_m)^2} < k_1$$

or equivalently $$\frac{\hat{\sigma}^2}{(\overline{X}_n - \overline{Y}_m)^2} < k_3 = k_2 - \frac{mn}{m+n}$$

or equivalently $$\frac{(\overline{X}_n - \overline{Y}_m)^2}{\hat{\sigma}^2} > k_4 = \frac{1}{k_3}$$

or equivalently $$T^2 = \frac{mn}{(m+n)}\frac{(\overline{X}_n - \overline{Y}_m)^2}{s_p^2} > k_5 = \frac{mn(m+n-2)k_4}{(m+n)^2},$$

which is again the same statistic used in the likelihood ratio and Wald tests.

c. What is the small sample distribution for the test in part a? What is the critical value of a level α test?

Ans: As noted above, the small sample distribution under the null hypothesis is the F distribution with 1 and $m+n-2$ degrees of freedom, and the critical value is the $(1-\alpha)$th quantile of that distribution. (The statistic has a noncentral F distribution under the alternative.)

4. Now consider the general nonparametric problem (i.e., we only know the means and variances, not the parametric distribution).

a. What is the asymptotic distribution of the test statistic you derived in problem 3? Do not presume that the variances are equal for this problem.

Ans: We first consider the distribution of $\overline{X}_n - \overline{Y}_m$. If $n = m$, we could simply define a new variable $W_i = X_i - Y_i$ which would be distributed with mean $\mu - \nu$ and (by independence) variance $\sigma^2 + \tau^2$. The Levy CLT would then tell us that $\overline{W}_n = \overline{X}_n - \overline{Y}_n$ was asymptotically normally distributed with mean $\mu - \nu$ and variance $(\sigma^2 + \tau^2)/n$.

With unequal sample sizes having some integral ratio, say $r : 1$, we could similarly define variables

$$W_i = \sum_{j=(i-1)r+1}^{ri} X_j/r + Y_i \sim \left(\mu = \nu, \frac{\sigma^2}{r} + \tau^2\right)$$

and again proceed with the Levy CLT.

A more general approach can be based on considering for each sample size a random vector $\vec{W} = (X_1/n, X_2/n, \ldots, X_n/n, -Y_1/m, -Y_2/m, \ldots, -Y_m/m)$. Then

$$S = \sum_{i=1}^{n+m} W_i = \overline{X}_n - \overline{Y}_m.$$

We know then that

$$E[S] = \mu - \nu \qquad \text{and} \qquad Var(S) = \frac{\sigma^2}{n} + \frac{\tau^2}{m}$$

and if the Lindeberg condition holds, the Lindeberg-Feller CLT provides that

$$\frac{(\overline{X}_n - \overline{Y}_m) - (\mu - \nu)}{\sqrt{\sigma^2/n + \tau^2/m}} \rightarrow_d \mathcal{N}(0,1)$$

I note that the Lindeberg condition will hold so long as the minimum of m and n approaches ∞ (shown in Stat 512 notes), a condition that certainly holds in this setting where $n/(m+n) \to \lambda$.

Now we can rewrite the above normalized form as

$$\frac{(\overline{X}_n - \overline{Y}_m) - (\mu - \nu)}{\sqrt{\sigma^2/n + \tau^2/m}} = \sqrt{\frac{mn}{(m+n)}} \frac{(\overline{X}_n - \overline{Y}_m) - (\mu - \nu)}{s_p} \frac{s_p}{\sqrt{(m\sigma^2 + n\tau^2)/(m+n)}}$$

Because $s_X^2 \to_p \sigma^2$, $s_Y^2 \to_p \tau^2$, $(n-1)/(m+n-2) \to \lambda$, and $(m-1)/(m+n-2) \to 1-\lambda$, we can easily see that

$$s_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{m+n-2} \to_p \lambda\sigma^2 + (1-\lambda)\tau^2.$$

Similarly, $(m\sigma^2 + n\tau^2)/(m+n) \to_p (1-\lambda)\sigma^2 + \lambda\tau^2$. Thus by Mann-Wald and other properties of convergence in probability

$$\frac{\sqrt{(m\sigma^2 + n\tau^2)/(m+n)}}{s_p} \to_p \sqrt{\frac{(1-\lambda)\sigma^2 + \lambda\tau^2}{\lambda\sigma^2 + (1-\lambda)\tau^2}},$$

and we find the asymptotic distribution of

$$\sqrt{\frac{mn}{(m+n)}} \frac{(\overline{X}_n - \overline{Y}_m) - (\mu - \nu)}{s_p} \to_d \mathcal{N}\left(0, \frac{(1-\lambda)\sigma^2 + \lambda\tau^2}{\lambda\sigma^2 + (1-\lambda)\tau^2}\right).$$

Under the null hypothesis $H_0 : \mu - \nu = 0$,

$$T = \sqrt{\frac{mn}{(m+n)}} \frac{(\overline{X}_n - \overline{Y}_m)}{s_p} \to_d \mathcal{N}\left(0, \frac{(1-\lambda)\sigma^2 + \lambda\tau^2}{\lambda\sigma^2 + (1-\lambda)\tau^2}\right).$$

Under an alternative hypothesis,

$$T \dot\sim \mathcal{N}\left(\sqrt{\frac{mn}{(m+n)}} \frac{(\mu - \nu)}{\sqrt{\lambda\sigma^2 + (1-\lambda)\tau^2}}, \frac{(1-\lambda)\sigma^2 + \lambda\tau^2}{\lambda\sigma^2 + (1-\lambda)\tau^2}\right).$$

b. Show that the test you derived in problem 3 is not necessarily level α when testing $H_0 : \mu = \nu$ versus $H_1 : \mu \neq \nu$. Show that it is level α when testing $H_0^* : \mu = \nu$ AND $\sigma^2 = \tau^2$ versus $H_1^* : \mu \neq \nu$ OR $\sigma^2 \neq \tau^2$.

Ans: In problem 3, the statistic T^2 was compared to the $(1-\alpha)$th quantile of an F distribution with 1 and $n+m-2$ degrees of freedom. This is equivalent to comparing the absolute value of the statistic T to the $(1-\alpha/2)$th quantile of a t distribution with $n+m-2$ degrees of freedom. As $n+m-2$ gets large, this is equivalent to comparing the absolute value of T to $z_{1-\alpha/2}$. The power function is thus

$$Pwr(\mu - \nu) = Pr[T < z_{\alpha/2} \,|\, \mu - \nu] + Pr[T > z_{1-\alpha/2} \,|\, \mu - \nu].$$

By using the asymptotic normal distribution,

$$Pr[T < z_{\alpha/2} \,|\, \mu - \nu] \dot= \Phi\left(\left(z_{\alpha/2} - \sqrt{\frac{mn}{(m+n)}} \frac{(\mu - \nu)}{\sqrt{\lambda\sigma^2 + (1-\lambda)\tau^2}}\right) \sqrt{\frac{\lambda\sigma^2 + (1-\lambda)\tau^2}{(1-\lambda)\sigma^2 + \lambda\tau^2}}\right)$$

$$Pr[T > z_{1-\alpha/2} \,|\, \mu - \nu] \dot= 1 - \Phi\left(\left(z_{1-\alpha/2} - \sqrt{\frac{mn}{(m+n)}} \frac{(\mu - \nu)}{\sqrt{\lambda\sigma^2 + (1-\lambda)\tau^2}}\right) \sqrt{\frac{\lambda\sigma^2 + (1-\lambda)\tau^2}{(1-\lambda)\sigma^2 + \lambda\tau^2}}\right)$$

Now, when $\mu = \nu$, the above power function is asymptotically level α providing $\sigma^2 = \tau^2$ OR $\lambda = 0.5$, because in those cases

$$\sqrt{\frac{\lambda\sigma^2 + (1-\lambda)\tau^2}{(1-\lambda)\sigma^2 + \lambda\tau^2}} = 1,$$

and $Pwr(0) = \alpha/2 + \alpha/2 = \alpha$. But when $\lambda \neq 0.5$ and $\sigma^2 \neq \tau^2$,

$$Pwr(0) = 2\Phi\left(z_{\alpha/2}\sqrt{\frac{\lambda\sigma^2 + (1-\lambda)\tau^2}{(1-\lambda)\sigma^2 + \lambda\tau^2}}\right).$$

Note that $Pwr(0) > \alpha$ when

$$\sqrt{\frac{\lambda\sigma^2 + (1-\lambda)\tau^2}{(1-\lambda)\sigma^2 + \lambda\tau^2}} < 1,$$

which occurs when $(1-2\lambda)\sigma^2 > (1-2\lambda)\tau^2$. Thus, if $\lambda < 0.5$ and $\sigma^2 > \tau^2$ or if $\lambda > 0.5$ and $\sigma^2 < \tau^2$, the test is anti-conservative in that the true type I error is greater than α. This can be said a bit more succinctly as: if the group with the larger variance has the smaller sample size, the test is anti-conservative. If the group with the larger variance has the larger sample size, the test is conservative (the type I error is smaller than α).

It should be obvious that if we are testing H_0^* versus H_1^*, the test is the correct level, because the event that $\sigma^2 \neq \tau^2$ is not possible under the null hypothesis.

c. Show that the test you derived in problem 3 is not consistent when testing $H_0^* : \mu = \nu$ AND $\sigma^2 = \tau^2$ versus $H_1^* : \mu \neq \nu$ OR $\sigma^2 \neq \tau^2$.

Ans: We consider the case where $\mu = \nu$ and $\sigma^2 \neq \tau^2$. This is clearly included in the alternative hypothesis. Now, if $m = n$ ($\lambda = 0.5$), we found in part b that the power of the test was $\alpha < 1$. So the test is clearly not consistent. Note that even if $\lambda \neq 0.5$, the power is still less than 1. This is because when $\mu = \nu$, the power does not depend upon m or n, but only on their ratio.

Moral To This Story

Almost every test comparing a distribution between two groups is based on comparing some functional of the distribution within each group:

- The t test examines the difference in group means.
- A t test performed on log transformed data examines the ratio of group geometric means.
- The parametric accelerated failure time models sometimes used in survival analysis (e.g., the Weibull, log logistic, lognormal, generalized gamma) examine the ratio of group medians.
- The chi squared test can be viewed as looking at either odds ratios or difference in proportions.
- The logrank test (as can be derived as a score test in the proportional hazards model) examines the ratio of hazard functions.
- The Wilcoxon rank sum test examines the probability that a randomly chosen subject from one group might exceed a randomly chosen subject from another group, i.e., whether $Pr(X > Y) = 0.5$.

In the common implementations of all of the above tests (e.g., the t test presuming equal variances), the variance of the test statistic is computed under the assumption that if the functional is at its null value (e.g., no difference in means), the distribution in the two groups is identical. In all of the above tests except the chi squared test, this means that the test might have the wrong type I error when viewed solely as a test of the functional and that the test is inconsistent when viewed as a test of any difference in the distributions.

Had we used the t test which allows for the possibility of unequal variances, the test would be asymptotically of the correct size and consistent to test H_0 versus H_1. Personally, I believe that it is best to use tests based on as few assumptions as possible. I like consistent tests, and I do not like making unnecessary assumptions that are more detailed than the hypotheses I am testing. Hence, I am, for instance, in favor of modifying the computation of the null variance for the Wilcoxon and logrank tests to be similar in approach to the t test which allows for unequal variances.

Lastly, I note that there are some tests based on functionals which do not have this problem. The chi squared test falls in this category, because binary data must be the Bernoulli distribution, and it is entirely specified by the event probability. Other examples would include the Kolmogorov-Smirnov test, which examines the maximum difference between two cdf's. For this test, there is no problem similar to that when testing a more narrowly defined functional like the mean or median.