Mathematics 376 – Probability and Statistics 2
Solutions for Midterm Examination 2
April 22, 2010

I. Let $Y_1, \ldots, Y_n$ be independent samples from the distribution with pdf containing the unknown parameter $\theta > 0$:

$$f(y|\theta) = \begin{cases} \frac{1}{\theta} y^{1/\theta - 1} & \text{if } 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

A) (10) Determine the method of moments estimator for θ using the Y_i.

Solution: For a random variable with this density,

$$\begin{aligned} E(Y) &= \int_0^1 y \cdot \frac{1}{\theta} y^{1/\theta - 1}\, dy \\ &= \frac{1}{\theta} \int_0^1 y^{\frac{1}{\theta}}\, dy \\ &= \frac{\frac{1}{\theta}}{\frac{1}{\theta} + 1} y^{\frac{1}{\theta}+1} \Big|_0^1 \\ &= \frac{1}{\theta + 1}. \end{aligned}$$

Then by the method of moments, we set

$$\overline{Y} = \frac{1}{\theta + 1}$$

and solve for theta to get the estimator:

$$\theta\overline{Y} + \overline{Y} = 1 \Rightarrow \widehat{\theta}_{MM} = \frac{1 - \overline{Y}}{\overline{Y}}\ .$$

B) (15) Determine the maximum likelihood estimator for θ using the Y_i.

Solution: We have

$$L(y_1, \ldots, y_n|\theta) = \frac{1}{\theta^n}(y_1 \cdots y_n)^{1/\theta - 1},$$

so

$$\ln(L) = -n\ln(\theta) + \left(\frac{1}{\theta} - 1\right)\ln(y_1 \cdots y_n).$$

This is a case where $\ln(L)$ *does have* a critical point (exactly one):

$$\frac{d}{d\theta}\ln(L) = \frac{-n}{\theta} - \frac{1}{\theta^2}\ln(y_1 \cdots y_n) = 0$$

when θ equals:

$$\widehat{\theta}_{ML} = \frac{-\ln(Y_1 \cdots Y_n)}{n}.$$

It can be checked without too much difficulty that this is a local maximum. For instance, by the Second Derivative Test,

$$\frac{d^2}{d\theta^2}\ln(L) = \frac{n}{\theta^2} + \frac{2}{\theta^3}\ln(y_1 \cdots y_n)$$

Substituting $\theta = \frac{-\ln(y_1 \cdots y_n)}{n}$ gives the second derivative value:

$$\frac{n^3}{\ln(y_1 \cdots y_n)^2} - \frac{2n^3}{\ln(y_1 \cdots y_n)^2} < 0.$$

Therefore, this is a local maximum.

II. A shop manufactures O-rings for the Space Shuttle booster rockets for NASA. Let d be the proportion of defectives in the shop's output. A random sample of size $n = 70$ O-rings produced 6 defectives.

A) (15) Test the hypothesis $H_0 : d = .1$ versus $H_a : d \neq .1$ using this data. Take $\alpha = .02$ (probability of Type I error). Also give the p-value of your test.

Solution: Since $n = 70 > 30$, a large-sample test is appropriate. The test statistic is

$$z = \frac{6/70 - .1}{\sqrt{\frac{(.1)(.9)}{70}}} = -.3984.$$

For a two-tail test at level $\alpha = .02$, we would have

$$RR = \{z \mid |z| > z_{.01}\} = \{z \mid |z| > 2.33\}.$$

Since z is not in the rejection region, we cannot reject H_0. The p-value of the test would be about $2 \times .3446 = .6892$ from the standard normal table. This is much too large to reject H_0.

B) (15) For the test in part A, what is β (probability of Type II error) if the true value of d is .03?

Solution: The Type II error probability is the probability of *not rejecting* H_0 when it is not true. This corresponds to having $\frac{Y}{70}$ in the complement of the rejection region so:

$$.1 - 2.33\sqrt{\frac{(.1)(.9)}{70}} \leq \frac{Y}{70} \leq .1 + 2.33\sqrt{\frac{(.1)(.9)}{70}}. \tag{1}$$

If the true value of d is .03, then this probability can be estimated using the fact that

$$\frac{\frac{Y}{70} - .03}{\sqrt{\frac{(.03)(.97)}{70}}}$$

is approximately standard normal, so the probability that $\frac{Y}{70}$ is in the range from (1) is

$$P\left(\frac{.07 - 2.33\sqrt{\frac{(.1)(.9)}{70}}}{\sqrt{\frac{(.03)(.97)}{70}}} < Z < \frac{.07 + 2.33\sqrt{\frac{(.1)(.9)}{70}}}{\sqrt{\frac{(.03)(.97)}{70}}}\right)$$

This is

$$P(-.66 < Z < 7.53) \doteq P(Z \le +.66) = 1 - .2546 = .7454.$$

This value of β would be clearly unacceptable in general for a reliable test. A larger sample size is indicated!

III. Consider the following measurements of the heat-producing capacity of the natural gas produced by two fields of gas wells (in calories per cubic meter):

Field 1 :	8.26	8.13	8.35	8.07	8.34	
Field 2 :	7.95	7.89	7.90	8.14	7.92	7.84

Let μ_i $(i = 1, 2)$ be the population mean heat-producing capacity of the natural gas from field i.

A) (10) What assumptions do you need to make in order to use the appropriate test of $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$? How could you determine whether it is reasonable to assume those assumptions are satisfied?

Solution: For the basic small-sample test for equality of means, we would need to assume the usual independence of the two sets of samples, that the two populations have normal distributions, and that the variances of the two normal populations are *equal*. You could do an F-test for equality of the two population variances. (Note that no F-table was provided with the exam(!) If you had one, then you could proceed as follows. The two sample variances are

$$S_1^2 = .01575 \quad \text{and} \quad S_2^2 = .01092.$$

So $F = S_1^2/S_2^2 = 1.4423$. For 4 degrees of freedom in the numerator and 5 degrees of freedom in the denominator, with $\alpha = .05$, $F_{.025} = 7.39$ and $F_{.975} = 1/9.36 \doteq .011$. The rejection region would consist of values $F > 7.39$ and $0 < F < .011$. We do not reject the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$.)

B) (15) Carry out the test from part A with $\alpha = .05$ and state your conclusion clearly and succinctly.

Solution: We use the pooled estimator for the variance:

$$S_p^2 = \frac{4S_1^2 + 5S_2^2}{9} \doteq .0131.$$

Then

$$t = \frac{8.23 - 7.94}{\sqrt{.0131}\sqrt{1/5 + 1/6}} \doteq 4.19$$

For a t-distribution with 9 degrees of freedom $t_{.025} = 2.262$. So there is relatively strong evidence to reject H_0 here. (The p-value is between .01 and .02.)

C) (5) Construct a two-sided 95% confidence interval for the difference of the population mean heat producing capacities $\mu_1 - \mu_2$. How is this related to your answer in B?

Solution: The confidence interval can be computed using all the information developed in the previous part:

$$\mu_1 - \mu_2 = (8.23 - 7.94) \pm (2.262)\sqrt{.0131}\sqrt{1/5 + 1/6}.$$

Under the null hypothesis, $\mu_1 - \mu_2 = 0$, so we "accept" H_0 with sample means $\overline{Y}_1$ and $\overline{Y}_2$ when 0 is in the confidence interval

$$\overline{Y}_1 - \overline{Y}_2 \pm (2.262)\sqrt{.0131}\sqrt{1/5 + 1/6}$$

and reject H_0 otherwise.

IV. (15) Let (x_i, y_i), $i = 1, \ldots, n$ be a collection of data points in the plane. Using the matrix formulation, explain how to derive the normal equations for the least squares estimators for the coefficients $\beta_0, \beta_1, \beta_2, \beta_3$ in the model $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$ fitting the data. (Note: You *do not need to solve the equations.*)

Solution: We have

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{pmatrix}$$

and Y = column vector of with entries $y_1, \ldots, y_n$. Let β be the column vector with entries $\beta_0, \beta_1, \beta_2, \beta_3$. Then the normal equations are

$$X^t X \beta = X^t Y.$$

This can be written in the following form:

$$\begin{pmatrix} n & \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \sum x_i^5 \\ \sum x_i^3 & \sum x_i^4 & \sum x_i^5 & \sum x_i^6 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \\ \sum x_i^3 y_i \end{pmatrix}$$

Extra Credit (10) In the situation of question I, suppose you set up a test of the hypothesis $H_0 : \theta = 1$ versus the alternative $H_a : \theta > 1$ using Y_1, Y_2 (exactly two of the samples). If you reject H_0 when $Y_1 + Y_2 > 1.8$, what is the Type I error probability, α?

Solution: Under H_0, $\theta = 1$, so Y_1, Y_2 have uniform distributions on the interval $[0, 1]$. The samples are assumed independent, so the joint density of Y_1 and Y_2 is 1 on the unit square $[0, 1] \times [0, 1]$ and zero elsewhere. We make a Type I error when we reject H_0, but $\theta = 1$. The probability that $Y_1 + Y_2 > 1.8$ is the area of an isosceles right triangle with legs .2, so $\alpha = (.2)(.2)/2 = .02$.