Mathematics 376 – Mathematical Statistics
Solutions for Midterm Exam I
March 15, 2012

I. Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be two independent random samples of equal sizes from a population with a normal distribution having a *known* mean μ, but *unknown* variance σ^2. Let Σ_X^2 and Σ_Y^2 be the statistics:

$$\Sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \qquad \Sigma_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu)^2.$$

Answer the following questions, giving complete explanations.

A) (5) What is the distribution of $\frac{X_i - \mu}{\sigma}$?

Solution: This is the usual standardization of the normal random variable X_i, so it has a standard normal distibution.

B) (5) What is the distribution of

$$W = \frac{n\Sigma_X^2}{\sigma^2}?$$

Solution: We see that

$$n\Sigma_X^2/\sigma^2 = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2$$

is the sum of squares of n independent standard normals. Hence W has a χ^2 distribution with n degrees of freedom. (It's *not* $n-1$ degrees of freedom here because we know the exact population μ and we use that rather than the sample mean $\overline{X}$.)

C) (10) What is the distribution of

$$V = \frac{\Sigma_X^2}{\Sigma_Y^2}?$$

Solution: Rewriting the quotient as

$$V = \Sigma_X^2/\Sigma_Y^2 = \frac{\frac{n\Sigma_X^2}{\sigma^2}/n}{\frac{n\Sigma_Y^2}{\sigma^2}/n}$$

using part A we see that V has an F distribution with n degrees of freedom in the numerator and n degrees of freedom in the denominator.

D) (10) What is the distribution of

$$U = \frac{Y_1 - \mu}{\Sigma_X}?$$

Solution: To see what is going on here, we need to note that Σ_X appears here rather than Σ_X^2. So a square root has been extracted in the denominator. Rewriting we have

$$\frac{(Y_1 - \mu)/\sigma}{\sqrt{(n\Sigma_X^2/\sigma^2)/n}}$$

Therefore we have a standard normal, divided by the square root of a $\chi^2(n)$ divided by its number of degrees of freedom. The standard normal and the χ^2 are independent by assumption. Hence U has a t distribution with n degrees of freedom.

II. Assume that the calorie content of Nature Valley granola bars is normally distributed. A random sample of 10 granola bars yields calorie content values of:

$$210, 220, 234, 227, 229, 240, 221, 228, 225, 219$$

A) (15) Find a 90% confidence interval for the population mean calorie content, μ. Explain how you know which method to use.

Solution: We estimate μ with the sample mean $\overline{Y} \doteq 225.3$. The sample variance of the calorie content values above is $S^2 \doteq 70.68$, so $S \doteq 8.407$ cal. Since $n = 10$, we use the small sample formula (using percentage points of the t-distribution with $n - 1 = 9$ degrees of freedom). Since we want a 90% confidence interval we want $\alpha = (1 - .9)/2 = .05$.

$$\begin{aligned}\mu &= 225.3 \pm t_{.05}\frac{8.407}{\sqrt{10}}\\ &= 225.3 \pm (1.833)\frac{8.407}{\sqrt{10}}\\ &= 225.3 \pm 4.87\end{aligned}$$

B) (10) Find a 95% confidence interval for the calorie content variance σ^2.

Solution: The 95% confidence interval is found using the percentage points of the χ^2 distribution with $n - 1 = 9$ degrees of freedom:

$$\begin{aligned}\sigma^2 &\in \left[\frac{9(70.68)}{\chi^2_{.025}}, \frac{9(70.68)}{\chi^2_{.975}}\right]\\ &= \left[\frac{9(70.68)}{19.0228}, \frac{9(70.68)}{2.70039}\right]\\ &= [33.44, 235.57]\end{aligned}$$

(Note: The fact that this interval is so large comes from the fact that the number of samples is quite small. It is difficult to pin down the population variance from this little data.)

III. A researcher found that 107 out of 200 apple seeds germinated when they were planted in soil maintained at $5°C$, while 123 out of 200 seeds germinated when they were planted in soil maintained at $15°C$.

A) (20) Construct a 95% confidence interval for the difference in the proportions of seeds that germinate at the two temperatures.

Solution: Since $n = 200 > 30$ for the two samples, we use the large-sample formula for the difference of proportions (with the percentage points of the standard normal distribution). The point estimators for the two proportions are $\hat{p}_{5^\circ} = 107/200 = .535$ and $\hat{p}_{15^\circ} = 123/200 = .615$. Then

$$p_{15^\circ} - p_{5^\circ} = .615 - .535 \pm (1.96)\sqrt{\frac{(.615)(.385)}{200} + \frac{(.535)(.465)}{200}} \doteq 0.08 \pm .0966$$

B) (5) Do the results of this study suggest that a greater proportion of apple seeds will germinate at the higher temperature in general? Why or why not?

Solution: No – the data does not support that conclusion because the confidence interval contains zero and positive numbers as well as negative numbers. The fact that $\hat{p}_{5^\circ} < \hat{p}_{15^\circ}$ might be a result of random variation.

IV. (15)Let $Y_1, \ldots, Y_n$ denote a random sample from a uniformly distributed population on an interval $[\theta, 2\theta]$ where θ is unknown. Consider the estimator $\widehat{\theta} = \min(Y_1, \ldots, Y_n)$. Is $\widehat{\theta}$ biased or unbiased?

Solution: We need to compute the expected value of $\widehat{\beta} = Y_{(1)}$, so we need the pdf for the sample minimum in this case. The general formula is $f_{(1)}(y) = n(1-F(y))^{n-1}f(y)$, where $F(y)$ is the cdf for the distribution. From the formulas for uniform distributions we have

$$f(y) = \begin{cases} \frac{1}{\theta} & \text{if } \theta \le y \le 2\theta \\ 0 & \text{otherwise} \end{cases}$$

and

$$F(y) = \begin{cases} 0 & \text{if } y < \theta \\ \frac{(y-\theta)}{\theta} & \text{if } \theta \le y \le 2\theta \\ 1 & \text{if } y > 2\theta \end{cases}.$$

Hence

$$f_{(1)}(y) = n\left(1 - \frac{(y-\theta)}{\theta}\right)^{n-1} \frac{1}{\theta} = n\frac{(2\theta - y)^{n-1}}{\theta^n}$$

if $\theta \le y \le 2\theta$ and 0 otherwise. To find the expected value, we can integrate by parts with $u = y$, $dv = n\frac{(2\theta-y)^{n-1}}{\theta^n}\ dy$. Then $du = dy$, $v = \frac{-(2\theta-y)^n}{\theta^n}$ and

$$
\begin{aligned}
E(Y_{(1)}) &= \int_\theta^{2\theta} yn\frac{(2\theta-y)^{n-1}}{\theta^n}dy \\
&= \left.\frac{-y(2\theta-y)^n}{\theta^n}\right|_\theta^{2\theta} + \int_\theta^{2\theta} \frac{(2\theta-y)^n}{\theta^n}\ dy \\
&= \theta + \left.\frac{-(2\theta-y)^{n+1}}{(n+1)\theta^n}\right|_\theta^{2\theta} \\
&= \theta + \frac{\theta}{n+1} \\
&= \frac{(n+2)\theta}{n+1}
\end{aligned}
$$

Since we do not get exactly θ, this is a *biased* estimator.

Extra Credit (10) In honor of π-day, here is the fact behind one of the demos that we were running in the McBrien Student Common Room. Suppose we drop a rod 1 unit long so that its midpoint lands at a point with y-coordinate uniformly distributed on the interval $[0, 1/2]$, and the angle between the horizontal and the rod is also uniformly distributed on the interval $[0, \pi]$. Show that the probability that the rod intersects the x-axis is $\frac{2}{\pi}$.

Solution: Supposing the midpoint lands at a position making an angle θ with the horizontal, the rod will hit the x-axis if $0 \leq y \leq \frac{\sin(\theta)}{2}$ (draw the picture)! Hence the probability we want is

$$\int_0^\pi \int_0^{\sin(\theta)/2} \frac{2}{\pi}\ dy\ d\theta = \frac{2}{\pi} \cdot \int_0^\pi \frac{\sin(\theta)}{2}\ d\theta = \frac{2}{\pi} \cdot (-\cos(\pi) + \cos(0))/2 = \frac{2}{\pi}$$