Steve Sawin
Probability 3352

Problem Set V (Solutions)

Due April 9

All write-ups should be individually done, complete, neat and precise. Please tell me anyone you worked with or got help from and AI tools, websites or books you used and exactly how.

1. [1pt] Question 10.13, p. 326: Show if T is an unbiased estimator of θ and if $\mathrm{Var}(T) \neq 0$ then T^2 is a *biased* estimator of θ^2.

$$\begin{aligned} \mathrm{Var}(T) &= E(T^2) - E(T)^2 \neq 0 \\ E(T^2) &\neq E(T)^2 = \theta^2. \end{aligned}$$

2. [3pt] Question 10.15, p. 326: Show the mean of a random sample of size n is a minimum variance unbiased estimator of the parameter λ for a Poisson distribution. Unbiased:

$$E(\overline{X}) = \mu_X = \lambda$$

for the Poisson distribution.

Minimum variance:

$$\begin{aligned} p(x;\lambda) &= \frac{\lambda^x e^{-\lambda}}{x!} \\ \ln(p(x;\lambda)) &= x\ln(\lambda) - \lambda - \ln(x!) \\ \frac{\partial}{\partial\lambda}\ln(p(x;\lambda)) &= \frac{x}{\lambda} - 1 = \frac{1}{\lambda}(x-\lambda) \\ E\left(\left[\frac{\partial}{\partial\lambda}\ln(p(x;\lambda))\right]^2\right) &= E\left(\frac{1}{\lambda^2}(x-\lambda)^2\right) \\ &= \frac{1}{\lambda^2}E\left((x-\lambda)^2\right) \\ &= \frac{1}{\lambda^2}\lambda = \frac{1}{\lambda} \end{aligned}$$

because $E\left((x-\lambda)^2\right) = E\left((x-\mu)^2\right) = \mathrm{Var}(x) = \lambda$ by definition. So the smallest variance an unbiased estimator of λ can hope to have is

$$\frac{1}{n(1/\lambda)} = \frac{\lambda}{n}.$$

On the other hand $\mathrm{Var}(\overline{X}) = \mathrm{Var}(X)/n = \lambda/n$, so $\overline{X}$ achieves this minimum variance.

3. [3pt] Question 10.23, p. 327: Suppose that $\overline{X}_1$ and $\overline{X}_2$ are independent samples of size n_1 and n_2 respectively from the same normal population with mean μ and standard deviation σ. show that for each real number ω the following is an unbiased estimator of μ

$$\omega\overline{X}_1 + (1-\omega)\overline{X}_2.$$

Show that *among these* the one with the minimum variance is the one with

$$\omega = \frac{n_1}{n_1+n_2}.$$

Unbiased:

$$E\left(\omega\overline{X}_1 + (1-\omega)\overline{X}_2\right) = \omega E\left(\overline{X}_1\right) + (1-\omega)E\left(\overline{X}_2\right) = \omega\mu + (1-\omega)\mu = \mu.$$

minimum variance:

$$\begin{aligned}\operatorname{Var}\left(\omega\overline{X}_1 + (1-\omega)\overline{X}_2\right) &= \omega^2 \operatorname{Var}(\left(\overline{X}_1\right) + (1-\omega)^2 \operatorname{Var}\left(\overline{X}_2\right) \\ &= \omega^2\frac{\sigma^2}{n_1} + (1-\omega)^2\frac{\sigma^2}{n_2} \\ &= \sigma^2\left(\frac{\omega^2}{n_1} + \frac{(1-\omega)^2}{n_2}\right)\end{aligned}$$

now we minimize by setting the derivative equal to 0

$$\begin{aligned}0 &= \sigma^2\left(\frac{2\omega}{n_1} - \frac{2(1-\omega)}{n_2}\right) \\ &= \frac{\omega}{n_1} - \frac{1}{n_2} + \frac{\omega}{n_2} \\ \frac{1}{n_2} &= \omega\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \\ \omega &= \frac{1/n_2}{1/n_1 + 1/n_2} \\ \omega &= \frac{n_1}{n_1+n_2}.\end{aligned}$$

4. [1pt] Question 10.51, p. 342: Given a random sample of size n from a Poisson distribution, use the method of moments to find an estimator for the parameter λ. There is only one parameter so we express it in terms of the first moment, which is μ. In fact

$$\lambda = \mu,$$

so we could not express it any more simply than that.

Next we replace each instance of a population moment with the corresponding sample moment, to get an expression for the estimator. If we call L the estimator for λ

$$L = \overline{X}.$$

5. [2pt] Question 10.59, p. 342: Repeat the previous question using the method of maximum likelihood. Use the trick of minimizing the log of the likelihood. the likelihood function of λ is

$$L(\lambda) = f(x_1, \ldots, x_n; \lambda) = e^{-n\lambda} \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \lambda^{\sum x_i} \prod_{i=1}^{n} \frac{1}{x_i!}$$

so we will maximize $\ln(L)$

$$\begin{aligned}
0 &= \frac{\partial}{\partial\lambda}(\ln(L)) = \frac{\partial}{\partial\lambda}\left(\ln\left[e^{-n\lambda}\lambda^{\sum x_i}\prod_{i=1}^{n}\frac{1}{x_i!}\right]\right) \\
&= \frac{\partial}{\partial\lambda}\left(-n\lambda + \ln(\lambda)\sum_i x_i - \sum_i \ln(x_i!)\right) \\
&= -n + \frac{\sum_i x_i}{\lambda} \\
n\lambda &= \sum_i x_i \\
\lambda &= \frac{1}{n}\sum_i x_i = \overline{x}
\end{aligned}$$

so $\overline{X}$ is the maximum likelihood estimate for λ.

6. [3pt] Suppose again that we are estimating μ from a normal population with σ known, using a simple random sample with sample mean $\overline{x}$. Suppose you want to pick your sample size n so that the margin of error (which is half the total width of the confidence interval) is equal to e. Give a formula for the minimum n that will accomplish this as a function of σ, e and z^*_α. Remember $z^*_\alpha = z_{(1-\alpha)/2}$ is the z-score such that α is the chance a standard normal random variable will fall between $-z^*_\alpha$ and z^*_α. Since the α confidence interval has a margin of error

$$z^*_\alpha \sigma/\sqrt{n},$$

we have a probability α that

$$|\overline{x} - \mu| < z^*_\alpha \sigma/\sqrt{n}$$

and therefore if we need α confidence that

$$|\overline{x} - \mu| < e$$

it is sufficient to make sure

$$e > z^*_\alpha \sigma/\sqrt{n}.$$

Solving for n in this

$$\begin{aligned}
e &> z^*_\alpha \sigma/\sqrt{n} \\
\sqrt{n} &> z^*_\alpha \sigma/e \\
n &> (z^*_\alpha)^2 \frac{\sigma^2}{e^2}.
\end{aligned}$$

7. [2pt] Question 11.24, p. 370: A study of the annual growth of certain cacti showed that 64 of them, selected at random in a desert region, grew an average of 52.80 mm with a standard deviation of 4.5 mm. Construct a 99% confidence interval for the true average annual growth of the given kind of cactus. Note we are given the *sample* standard deviation. so we have to use the t-procedure, not the z procedure appropriate when you know the *population* s.d. For 99% confidence with 63 degrees of freedom we get a critical t-value of 2.65614 (notice because n is so large this is nearly the same as the critical z value, which would have been 2.575, so those who used the σ know procedure got pretty close). The confidence interval is

$$\overline{x} \pm t^*_{.99,63}\frac{s}{\sqrt{n}} = 52.80 \pm 1.49$$

or within the interval $[51.31, 54.29]$.

8. [2pt] Question 11.38, p. 372: A sample survey at a supermarket showed that 204 of 300 shoppers regularly used cents-off coupons. Give a 95% confidence interval for the true proportion.

$$\frac{204}{300} \pm z^*_{.95}\sqrt{\frac{204/300(1-204/300)}{300}} = 68.00\% \pm 5.28\%.$$

https://www.faculty.fairfield.edu/ssawin/stats/data/High_Jump.XLS

9. [3pt] Get the file at the link immediately above.

The first two columns give the heights jumped by a sample of men and another sample of women athletes. Use the two sample mean confidence interval to give a 90% confidence interval for the amount by which the average height jumped by men exceeds the average height jumped by women. Tell me what you would need to know about the population sizes, variables and sampling process to know that the assumptions for using this procedure were met. Again we use the two sample t procedure for the difference of means. Pasting the two columns in and setting the conf. level to 90% gives us that men jumped an averaqe of $0.25 \pm .09$ meters higher. Again we would need to know that the sample was a random sample of all athletes (I'm skeptical, but who knows), that the population had at least 500 men and 360 women (certainly if it is really all athletes, almost certainly even if it is much smaller population, say all American college high jumpers). Finally, since the samples are between 15 and 40 and highly skewed, it does not meet this assumption unless you knew that the population distributions were normal, which is pretty implausible.

Out of 20points.