MATH 563 Statistics

Fred J. Hickernell **Spring 2011**
Test 1 **Thursday, February 24**

Instructions:

i. *This test consists of* THREE *questions. Answer all of them.*

ii. *The time allowed for this test is 75 minutes.*

iii. *This test is closed book, but you may use 2 double-sided letter-size sheets of notes.*

iv. *Calculators, even of the programmable variety, are allowed. Computers using JMP or MATLAB, are also allowed. No internet access, web browsing, email, chat, etc. is allowed.*

v. *Show all your work to justify your answers. Answers without adequate justification will not receive credit.*

1. (30 marks)
For each statement below, determine whether it is true or not and give a brief explanation why.

a) Given i.i.d. data $X_1, \ldots, X_n$, with n large, one can use the Central Limit Theorem to construct a 95% confidence interval for the sample mean, $\bar{X} = (X_1 + \cdots + X_n)/n$.

Answer: False. One only constructs confidence intervals for population quantities.

b) Given only *one* exponentially distributed datum, X, with unknown mean, μ it is possible to construct an unbiased estimate of the standard deviation.

Answer: True. The standard deviation of X is μ, so X is an unbiased estimator of the standard deviation.

c) To decrease the half-width of a confidence interval for the mean by a factor of one tenth, one needs ten times as many data.

Answer: False. The half-width of a confidence interval for the mean is proportional to $1/\sqrt{n}$, where n is the sample size. Therefore, to decrease the half-width by a factor of one tenth, one needs one hundred *times as many data.*

2. (40 marks)
The number of failed attempts that Claire makes before successfully completing a half court basketball shot is a random variable X with a geometric distribution. For this distribution the mean is $\mu = E(X)$, the variance is $\text{var}(X) = \mu(1+\mu)$, and the probability mass function is

$$f(x) = \left(\frac{\mu}{1+\mu}\right)^x \frac{1}{1+\mu}, \quad x = 0, 1, 2, \ldots .$$

a) Compute the moment generating function of X, i.e., $M(t) = E(e^{tX})$.

Answer:

$$\begin{aligned} M(t) = E(e^{tX}) &= \sum_{x=0}^{\infty} e^{tx} f(x) = \sum_{x=0}^{\infty} e^{tx} \left(\frac{\mu}{1+\mu}\right)^x \frac{1}{1+\mu} \\ &= \sum_{x=0}^{\infty} \left(\frac{\mu e^t}{1+\mu}\right)^x \frac{1}{1+\mu} = \frac{1}{(1+\mu)(1-\frac{\mu e^t}{1+\mu})} = \frac{1}{(1+\mu-\mu e^t)}. \end{aligned}$$

b) Starting from the probability mass function or the moment generating function, verify that the parameter μ is truly the mean of this distribution.

Answer:

$$E(X) = M'(0) = \frac{\mathrm{d}}{\mathrm{d}t} \frac{1}{(1+\mu-\mu e^t)}\bigg|_{t=0} = \frac{\mu}{(1+\mu-\mu e^t)^2}\bigg|_{t=0} = \mu.$$

c) Let $X_1, X_2, \ldots$ be i.i.d. instances of this geometric random variable, i.e., X_i is the number of failed attempts before Claire makes her i^{th} half court shot. Let $\bar{X} = (X_1 + \cdots + X_n)/n$ denote the sample mean. Compute $\text{var}(\bar{X})$ in terms of μ and n.

Answer: Note that

$$\text{var}(\bar{X}) = \text{var}\left(\frac{1}{n}(X_1 + \cdots + X_n)\right) = \frac{1}{n^2}\left(\mu(1+\mu) + \cdots + \mu(1+\mu)\right) = \frac{\mu(1+\mu)}{n}.$$

d) As an avid basketball player, Claire shoots half court shots until she makes 100 of them. Her average number of failed attempts before making a half court shot is $\bar{x} = 4.53$. Construct an approximate 95% confidence interval for μ assuming that we are in the large sample size regime.

Answer: Using the Central Limit Theorem we know that

$$\begin{aligned} 95\% &= \text{Prob}\left(\bar{X} - \frac{1.96\sigma}{\sqrt{n}} \le \mu \le \bar{X} + \frac{1.96\sigma}{\sqrt{n}}\right) \quad (\sigma \textit{ is the standard deviation}) \\ &= \text{Prob}\left(\bar{X} - 1.96\sqrt{\frac{\mu(1+\mu)}{n}} \le \mu \le \bar{X} + 1.96\sqrt{\frac{\mu(1+\mu)}{n}}\right) \\ &\approx \text{Prob}\left(\bar{X} - 1.96\sqrt{\frac{\bar{X}(1+\bar{X})}{n}} \le \mu \le \bar{X} + 1.96\sqrt{\frac{\bar{X}(1+\bar{X})}{n}}\right) \end{aligned}$$

Thus, the approximate 95% *confidence interval for* μ *is*

$$\bar{X} \pm 1.96\sqrt{\frac{\bar{X}(1+\bar{X})}{n}}$$

which for this case is $4.53 \pm 0.98 = [3.55, 5.51]$.

3. (30 marks)
In the mayoral election last Tuesday it was found that 357 out of 943 union members surveyed voted for Rahm Emmanuel, while 551 out of 932 non-union members surveyed voted for Rahm Emmanuel. Construct a 95% confidence interval for the the difference in the proportion of all union members in Chicago who voted for Emmanuel and the proportion of all non-union members in Chicago who voted for Emmanuel. Can you confidently say that there is a difference in the proportions of pro-Emmanuel voters for these two different populations?

Answer: From the JMP output, the confidence interval is $[-25.6\%, -16.8\%]$. *Since this confidence interval does not straddle* 0, *there is a definite difference in the proportion of pro-Emmanuel voters in the union and non-union populations.*