

**BST 140.651**  
**Problem Set 3**

Problem 1. This problem is lots of fun, so enjoy it. Imagine that a person, say his name is Flip, has an oddly deformed coin and tries the following experiment. Flip flips his coin 10 times, 7 of which are heads. You think maybe Flip's coin is biased towards having a greater probability of yielding a head.

- What is the maximum likelihood estimate of  $p$ , the true probability of heads associated with this coin?
- Plot the likelihood associated with this experiment. Renormalize the likelihood so that its maximum is one. Does the likelihood suggest that the coin is fair?
- Whats the probability of seeing 7 or more heads out of ten coin flips if the coin was fair? Does this probability suggest that the coin is fair? Note this number is called a P-value.
- Suppose that Flip told you that he did not fix the number of trials at 10. Instead, he told you that he had flipped the coin until he obtained 3 tails and it happened to take 10 trials to do so. Therefore, the number 10 was random while the number three 3 fixed. The probability mass function for the number of trials, say  $y$ , to obtain 3 tails (called the negative binomial distribution) is

$$\binom{y-1}{2} (1-p)^3 p^{y-3}$$

for  $y = 3, 4, 5, 6, \dots$ . What is the maximum likelihood estimate of  $p$  now that we've changed the underlying mass function?

- Plot the likelihood under this new mass function. Renormalize the likelihood so that its maximum is one. Does the likelihood suggest that the coin is fair?
- Calculate the probability of requiring 10 or more flips to obtain 3 tails if the coin was fair. (Notice that this is the same as the probability of obtaining 7 or more heads to obtain 3 tails.) This is the Pvalue under the new mass function.

(Aside) This problem highlights a distinction between the likelihood and the P-value. The likelihood and the MLE are the same regardless of the experiment. That is to say, the likelihood only seems to care that you saw 10 coin flips, 7 of which were heads. Flip's intention about when he stopped flipping the coin, either at 10 fixed trials or until he obtained 3 tails, are irrelevant as far as the likelihood is concerned. The P-value, in comparison, does depend on Flip's intentions.

Problem 2. (See Rosner page 135) Suppose that the diasolic blood pressures of 35 – 44 year old men are normally distributed with mean 80 (*mm Hg*) and variance 144. For the same population, the systolic blood pressures are also normally distributed and have a mean of 120 and variance 121.

- What is the probability that a randomly selected person from this population has a DBP less than 90?
- What DBP represents the 90<sup>th</sup>, 95<sup>th</sup> and 97.5<sup>th</sup> percentiles of this distribution?
- What's the probability of a random person from this population having a SBP 1, 2 or 3 standard deviations above 120? What's the corresponding probabilities for having DBPs 1, 2 or 3 standard deviations above 80?
- Suppose that 10 people are sampled from this population. What's the probability that 5 of them have a SBP larger than 140.
- If a person's SBP and DBP are independent, what's the probability that a person has a SBP larger than 140 and a DBP greater than 90? Is independence a good assumption?

Problem 3. Recall that R's function `runif` generates (by default) random uniform variables that have means  $1/2$  and variance  $1/12$ .

- Sample 1,000 observations from this distribution. Take the sample mean and sample variance. What numbers should these estimate and why?
- Retain the same 1,000 observations from part a. Plot the sequential sample means by observation number. Hint. If  $x$  is a vector containing the simulated uniforms, then the code `y <- cumsum(x) / (1 : length(x))` will create a vector of the sequential sample means. Explain the resulting plot.
- Plot a histogram of the 1,000 numbers. Does it look like a uniform density?
- Now sample 1,000 *sample means* from this distribution, each comprised of 100 observations. What numbers should the average and variance of these 1,000 numbers be equal to and why? Hint. The command

```
x <- matrix(runif(1000 * 100), nrow = 1000)
```

creates a matrix of size  $1,000 \times 100$  filled with random uniforms. The command `y<-apply(x,1,mean)` takes the sample mean of each row.

- Plot a histogram of the 1,000 sample means appropriately normalized. What does it look like and why?
- Now sample 1,000 *sample variances* from this distribution, each comprised of 100 observations. Take the average of these 1,000 variances. What property does this illustrate and why?

Problem 4. Note that R's function `rexp` generates random exponential variables. The exponential distribution with rate 1 (the default) has a theoretical mean of 1 and variance of 1.

- Sample 1,000 observations from this distribution. Take the sample mean and sample variance. What numbers should these estimate and why?
- Retain the same 1,000 observations from part a. Plot the sequential sample means by observation number. Explain the resulting plot.
- Plot a histogram of the 1,000 numbers. Does it look like a exponential density?

- d. Now sample 1,000 *sample means* from this distribution, each comprised of 100 observations. What numbers should the average and variance of these 1,000 numbers be equal to and why?
- e. Plot a histogram of the 1,000 sample means appropriately normalized. What does it look like and why?
- f. Now sample 1,000 *sample variances* from this distribution, each comprised of 100 observations. Take the average of these 1,000 variances. What property does this illustrate and why?

Problem 5. Note that the code `x <- sample(1 : 6, 1000, replace = TRUE)` simulates 1,000 die rolls in R.

- a. Sample 1,000 observations from this distribution. Take the sample mean and sample variance. What numbers should these estimate and why?
- b. Retain the same 1,000 observations from part a. Plot the sequential sample means by observation number. Explain the resulting plot.
- c. Plot a histogram of the 1,000 numbers. Does it look like equal probability was placed on each number 1 – 6?
- d. Now sample 1,000 *sample means* from this distribution, each comprised of 100 observations. What numbers should the average and variance of these 1,000 numbers be equal to and why?
- e. Plot a histogram of the 1,000 sample means appropriately normalized. What does it look like and why?
- f. Now sample 1,000 *sample variances* from this distribution, each comprised of 100 observations. Take the average of these 1,000 variances. What property does this illustrate and why?

Problem 6. Consider the distribution of a fair coin flip (i.e. a random variable that takes the values 0 and 1 with probability  $1/2$  each.)

- a. Sample 1,000 observations from this distribution. Take the sample mean and sample variance. What numbers should these estimate and why?
- b. Retain the same 1,000 observations from part a. Plot the sequential sample means by observation number. Explain the resulting plot.
- c. Plot a histogram of the 1,000 numbers. Does it look like it places equal probability on 0 and 1?
- d. Now sample 1,000 *sample means* from this distribution, each comprised of 100 observations. What numbers should the average and variance of these 1,000 numbers be equal to and why?
- e. Plot a histogram of the 1,000 sample means appropriately normalized. What does it look like and why?

- f. Now sample 1,000 *sample variances* from this distribution, each comprised of 100 observations. Take the average of these 1,000 variances. What property does this illustrate and why?

Problem 7. A special study is conducted to test the hypothesis that persons with glaucoma have higher blood pressure than average. Two hundred subjects with glaucoma are recruited with a sample mean systolic blood pressure of  $140\text{mm}$  and a sample standard deviation of  $25\text{mm}$ . (Do not use a computer for this problem.)

- What is the estimated standard error of the mean? What is the difference in interpretation between the standard error of the mean and the standard deviation? Explain your answer in words.
- Construct a 95% confidence interval for the mean systolic blood pressure among persons with glaucoma. Do you need to assume normality? Explain.
- If the average systolic blood pressure for persons without glaucoma of comparable age is  $130\text{mm}$ . Is there statistical evidence that the blood pressure is elevated?

Problem 8. Suppose we wish to estimate the concentration  $\mu\text{g}/\text{m}\ell$  of a specific dose of ampicillin in the urine. We recruit 25 volunteers and find that they have sample mean concentration of  $7.0\ \mu\text{g}/\text{m}\ell$  with sample standard deviation  $3.0\ \mu\text{g}/\text{m}\ell$ . Let us assume that the underlying population distribution of concentrations is normally distributed.

- Find a 90% confidence interval for the population mean concentration.
- How large a sample would be needed to insure that the length of the confidence interval is  $0.5\ \mu\text{g}/\text{m}\ell$  if it is assumed that the sample standard deviation remains at  $3.0\ \mu\text{g}/\text{m}\ell$ ?