Math 472: Homework 04

Due Monday Feb 16

Problem 1 (Geometric distribution)**.** Consider a sequence of independent coin flips, each of which has probability p of being heads. Define a random variable X as the length of the run (either heads or tails) started by the first trial. (For example, if `TTTH` or `HHHT` is observed).

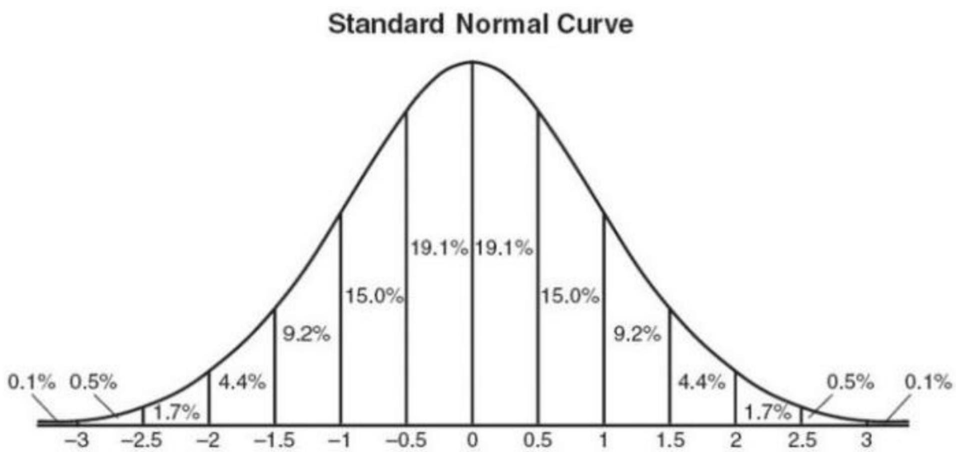
(a) What is the sample space? What are the sample points?

(b) Find the distribution of X (i.e., compute the probability mass function of X).

(c) Find $\mathbb{E}[X]$.

(d) Use `R` to generate 50 samples from the distribution of X (use whatever value of $p \in (0, 1)$ you want). What is the sample mean? What is the sample variance?

For Problems 2 to 6, the following plot may be helpful:



Problem 2. Let S be the number of heads in 100 tosses of a fair coin. Use the Central Limit Theorem to compute $\mathbb{P}[45 \leq S \leq 55]$. *Your answer should be a number.*

Problem 3. Use the Central Limit Theorem to approximate the probability of obtaining more than 65 heads when flipping a fair coin 100 times.

Problem 4. The nation of Oceania has always been at war with the nation of Eurasia. To establish peace, Oceania's leadership has unanimously authorized a preemptive strike to be carried out tomorrow morning against 48 pre-selected military and industrial targets in enemy territory. Analysts at the Ministry of Peace estimate that each target has a probability of 3/4 of being destroyed in this first strike.

The operation will be regarded as a strategic success under Oceania's peace doctrine if at least 30 targets are destroyed. Using the Central Limit Theorem, estimate the probability that the first strike achieves this benchmark.[1]

Problem 5. A mandolin-making machine in Mordecai's mandolin manufactory makes about 5% defective mandolins even when properly functioning. The mandolins are then packed in crates containing 1900 mandolins each. A crate is examined and found to contain 115 defective mandolins. What is the approximate probability of finding at least this many defective mandolins if the machine is properly adjusted? If you were Mordecai, would hire a technician to check out the machine?

Problem 6. A candidate believes that she can win a city election if she can earn at least 55% of the votes in precinct 1. She also believes that about 50% of the city's voters favor her. If $n = 100$ voters show up to vote at precinct 1, what is the probability that she will recieve at least 55% of their votes?

[1] Responses deviating significantly from 100% confidence will be flagged for review by the Ministry of Truth.

Problem 7 (Exercises 8.2, 8.3, and 8.4 in the textbook)**.** Suppose $\widehat{\theta}$ is an estimator for a target parameter θ.

(a) If $\widehat{\theta}$ is unbiased, what is $\text{Bias}(\widehat{\theta})$?

(b) If $\text{Bias}(\widehat{\theta}) = 5$, what is $\mathbb{E}[\widehat{\theta}]$?

(c) Suppose $\mathbb{E}[\widehat{\theta}] = a\theta + b$ for some nonzero constants a and b. In terms of a and b, what is $\text{Bias}(\widehat{\theta})$? Find a function of $\widehat{\theta}$—say, $\widehat{\theta}^*$—that is an unbiased estimator for θ.

(d) If $\widehat{\theta}$ is unbiased, how does $\text{MSE}(\widehat{\theta})$ compare to $\text{Var}(\widehat{\theta})$? What about when $\widehat{\theta}$ is biased?

Problem 8 (Exercise 8.8 in the textbook)**.** Suppose that Y_1, Y_2, Y_3 are a random sample from an exponential distribution with density function

$$f(y) = \frac{1}{\theta} e^{-y/\theta} \mathbf{1}_{[y>0]}$$

Consider the following estimators

$$\hat{\theta}_1 = Y_1, \qquad \widehat{\theta}_2 = \frac{Y_1 + Y_2}{2}, \qquad \widehat{\theta}_3 = \frac{Y_1 + 2Y_2}{3}, \qquad \hat{\theta}_4 = \min(Y_1, Y_2, Y_3), \qquad \hat{\theta}_5 = \overline{Y}$$

(Hint: recall the fact that the minimum of exponential random variables is itself exponentially distributed.).

(a) Which of these estimators is unbiased?

(b) Among the unbiased estimators, which has the smallest variance?

Problem 9 (Exercise 8.12 in the textbook)**.** The reading on a voltage meter connected to a test circuit is uniformly distributed over the interval $(\theta, \theta + 1)$, where θ is the true but unknown voltage of the circuit. Suppose that $Y_1, \ldots, Y_n$ denote a random sample of such readings

(a) Show that $\overline{Y}$ is a biased estimator of θ and compute the bias.

(b) Find a function of $\overline{Y}$ that is an unbiased estimator of θ.

(c) Find $\text{MSE}(\overline{Y})$ when $\overline{Y}$ is used as an estimator of θ.

Problem 10 (Exercise 8.23 in the textbook – similar to textbook Example 8.2)**.** The Environmental Protection Agency and the University of Florida recently cooperated in a large study of the possible effects of trace elements in drinking water on kidney-stone disease. The accompanying table presents data on age, amount of calcium in home drinking water (measured in parts per million), and smoking activity. These data were obtained from individ- uals with recurrent kidney-stone problems, all of whom lived in the Carolinas and the Rocky Mountain states.

	Carolinas	Rockies
Sample size	467	191
Mean age	45.1	46.4
Standard deviation of age	10.2	9.8
Mean calcium component (ppm)	11.3	40.1
Standard deviation of calcium	16.6	28.4
Proportion now smoking	0.78	0.61

(a) Estimate the average calcium concentration in drinking water for kidney-stone patients in the Carolinas. Place a bound on the error of estimation.

(b) Estimate the difference in mean ages for kidney-stone patients in the Carolinas and in the Rockies. Place a bound on the error of estimation.

(c) Estimate and place a 2-standard-deviation bound on the difference in proportions of kidney-stone patients from the Carolinas and Rockies who were smokers at the time of the study.