[STAT-E] Additional exercises

1. In a certain town there are 10,000 bicycles, each of which is assigned a license number from 1 to 10,000 (no two bicycles can receive the same number). What is the probability that the number on the first bicycle one encounters will not have any 8's among its digits?

2. This is a classic, the *Bernouilli-Euler problem of the misaddressed letters*. You write n letters and then the corresponding addresses on n envelopes, placing the letters at random in the envelopes. Let q_n denote the probability that no letter is placed in the correct envelope. Give a general formula for q_n and see what happens as n increases.

3. A family has three different children of different ages. We name them, randomly, A, B and C. What is the probability that A is older than B, given that A is older than C?

4. In a town of $n + 1$ inhabitants, a person creates a rumor, and tells it to a second person, who in turn repeats it to a third person, etc. At each step, the recipient of the rumor is chosen at random among the other n persons. Find the probability of the rumor being told exactly r times (including the first person telling it to the second person):

(a) Before returning to the originator.

(b) Without being repeated to any person.

5. A jar contains r red balls and g green balls. Two balls are drawn from the jar randomly.

(a) Which is higher, the probability that the first ball is red, or the probability that the second ball is red?

(b) Suppose that there are 16 balls in total ($r + g = 16$), and that the probability that the two balls are the same color is the the same as the probability that they are different colors. What are r and g?

6. In a bestselling book, the author reports a test on the ability of German physicians to read the numbers. About 95% of those enrolled in a study were wrong in the following one:

The probability that a woman in a certain population has breast cancer is 0.8 percent. If a woman has breast cancer, the probability is 90% that she will have a positive mammogram and, if she does not have breast cancer, the probability is 7% that she still will have a positive mammogram. If a woman gets a positive mammography, what is the probability that she actually has breast cancer?

¶ Source: G Gigerenzer (2002), *Reckoning with Risk*, Penguin.

7. Consider the following 7-door version of the Monty Hall problem. There are seven doors, behind one of which there is a car and behind the rest of which there are gots. You choose a door. Monty then opens three goat doors, and offers you the option of switching to any of the remaining three doors. What is your probability of success if you switch to one of the remaining three doors?

¶ Source: Joe Blitzein, Harvard University.

8. This is the *four liars problem*. A, B, C and D are four friends who tell the truth with probability 1/3 and lie with probability 2/3. Suppose that A makes a statement and then D says that C says that B says that A is telling the truth. What is the probability that A was actually telling the truth?

9. Nine passengers board a train consisting of three cars. Each passenger selects at random which car he/she will seat in. What is the probability that:

(a) There will be three people in the first car?

(b) There will be three people in each car?

10. Calculate the skewness and the kurtosis of the uniform distribution.

11. Take $U_1,\ U_2 \sim \mathcal{U}(0,1)$, independent, and define $X = U_1/U_2$. How is the distribution of this variable? Draw a random sample of size one million and take a look.

12. Calculate the skewness and the kurtosis of the exponential distribution.

13. Prove that the skewness and the kurtosis of a standard normal distribution are null.

14. Assuming normality, simulate the sampling distribution of the sample skewness and kurtosis with size $n = 100$. Look they normal? Do you get the expected standard deviations ($\sqrt{6/n}$ and $\sqrt{24/n}$, respectively)? Does this improve with a bigger sample size? What about the Jarque-Bera statistic?

15. Suppose that θ ($0 \le \theta \le 1$) is an unknown parameter and X is a random variable which can take the values $1, 2, 3, 4, 5$, with probabilities

$$p_1 = \theta^3, \quad p_2 = \theta^2(1-\theta), \quad p_3 = 2\theta(1-\theta), \quad p_4 = \theta(1-\theta)^2, \quad p_5 = (1-\theta)^3.$$

(a) Check that this defines a probability distribution for any θ in the specified range.

(b) Given a constant c, consider an estimator $\hat{\theta} = h(X)$ defined by

$$h(1) = 1, \qquad h(2) = 2 - 2c, \qquad h(3) = c, \qquad h(4) = 1 - 2c, \qquad h(5) = 0.$$

Show that $\hat{\theta}$ is an unbiased estimator of θ for any c.

(c) Find the value $c(\theta)$ for which $\hat{\theta}$ has minimum variance.

16. A survey on customer satisfaction uses a question about general satisfaction with options ranging from 1 (completely unsatisfied) to 5 (completely satisfied). A sample of $n_1 = 125$ customers from segment 1 gives an average response $\bar{x}_1 = 4.27$, with standard deviation $s_1 = 0.81$. For a sample of customers from segment 2, we get $n_2 = 129$, $\bar{x}_2 = 4.16$ and $s_2 = 0.75$.

(a) Calculate a t statistic for testing the mean difference between the two segments.

(b) Can we conclude that there is no difference in average satisfaction between the two segments?

17. Draw 250 independent random samples of size 5 from $\mathcal{N}(0,1)$ and calculate the sample variance for each sample. The same for $\mathcal{N}(1,1)$. Divide the first by the second, getting 250 F statistic values, and plot a histogram. Compare this histogram with Figure 1.

18. Generate 1,000 independent samples of size 10 of (X, Y) as follows. Take X and ϵ independent, with $X \sim \mathcal{N}(2,1)$ and $\epsilon \sim \mathcal{N}(0, 0.04)$. Then define Y as $Y = 3 + X + \epsilon$. For every sample, fit a regression line. Save the coefficients obtained and examine the joint distribution of the slope and the intercept.

Perform the simulation again, with $X \sim \mathcal{N}(0,1)$. Explain the different results obtained.