Assignment 2

Pietro Galliani, Werner Nutt
Andy Rivkin

Special Random Variables

Instructions: You have time until 7 January 2021, 23:55 h, to submit your solution to this assignment on OLE. This means, you have two weeks of regular course time.
Recall that your answers to each question count towards the final mark. Therefore, it is worthwhile to submit the assignment even if you feel not able to answer all questions.

- You can work out your solutions with a word processing system (Word, Latex) or by hand. It may be easier to write up your answers by hand because they are likely to contain symbolic calculations with fractions, powers, and integrals. If you submit handwritten solutions, make an effort to write clearly and to structure and comment your answers so that they are legible. If you write by hand, submit the answer as a scanned PDF document. (Don't submit a photo, but a scan. Photos are usually very hard to read.)
- Also if you write by hand, organize your submission clearly and and write legibly. Otherwise, we cannot accept your work.
- Explain your answers and the approach by which you obtained them. Mention when you apply specific rules (like "the variance of a sum of independent variables is the sum of the variances"). A chain of formulas alone is not enough.
- Each question has a weight, expressed in points. For your answer you get a mark on the scale from 0 to 30. The weight determines how much the mark contributes to the coursework mark and therefore to the final mark. Note that for the final mark, we consider the maximum of the exam mark and the mark for the question. Therefore, not much is lost if you get a low mark for one question: (i) it can be compensated by the exam mark, and (ii) it does not influence the marks for the other questions.
- Your assignment should represent your own effort. However, you are not expected to work alone. It is fine to discuss the exercises and try to find solutions together, but each student shall write down and submit his/her solutions separately.
- If you are a student who took this course already last year, you need not submit the same answers as past year to get points. Only submit answers to a question if they are different from the answers of last year. For other questions just indicate in your submssion that you would like to get the same mark as before. That spares us repeatedly marking the same work.

1 Chicks Hatching

Suppose, we are managing a chicken farm. Experience shows that not every egg hatches a chick, but that there is a probability p, $0 < p < 1$, that this happens. Also, not every hatched chick survives to adulthood, but there is a probability r, $0 < r < 1$, that this happens. Since conditions for eggs and chickens are well-controlled, it is reasonable to assume that hatching of different eggs is independent and that the same holds for survival. We have currently n eggs in the incubator.

1. What is the probability that k chickens hatch out of these n eggs?

2. What is the probability that from the n eggs k chickens survive to adulthood?

For each question, give the probability mass function and give the name of the distribution and its parameters.

(Weight: 8% of this CW)

2 Spaghetti Breaking

Suppose, we repeatedly drop a spaghetto so that it breaks in two pieces. A spaghetto has length l. We assume that the position of the breaking point is uniformly distributed. We are interested in the length $\mathcal{L}$ of the longer piece.

1. What is the cumulative distribution function of $\mathcal{L}$? What is the density?

2. What is the average length of the longer piece?

Hint: Assume first that $l = 1$.

(Weight: 10% of this CW)

3 Alice Waiting

A post office has 2 clerks. Alice enters the post office while 2 other customers, Bob and Claire, are being served by the 2 clerks. Alice is next in line. Assume that the time a clerk spends serving a customer has the *Exp*(λ) distribution.

1. What is the probability that Alice is the last of the 3 customers to finish being served?

2. What is the expected total time that Alice needs to spend at the post office?

Hint: What do you know about the minimum of two exponentially distributed variables?

(Weight: 8% of this CW)

4 How Skewed is the Exponential Distribution?

Mean and variance are two parameters that describe the shape of the distribution of a random variable. The mean is the center of gravity while the standard deviation, the square root of the variance, indicates how far values are spread out from the mean.

The mean of the distribution of $\mathcal{X}$ is the expected value $\mu = E[\mathcal{X}]$ of $\mathcal{X}$. To obtain the variance, the variable is normalized so that its distribution has mean 0, which is achieved by subtracting μ from $\mathcal{X}$. The variance is then the expected value $E[(\mathcal{X} - \mu)^2]$ of the *square* of the normalized version of $\mathcal{X}$.

One idea to measure skewness is the following. We normalize $\mathcal{X}$ to a variable whose distribution has the same proportions as the one of $\mathcal{X}$, but which has mean 0 and variance 1. This is achieved by first subtracting μ and then dividing the difference by the standard deviation σ, that is, the square root of the variance. We then define as the *skewness* of the distribution of $\mathcal{X}$ the expected value of the *cube* of the normalized version of $\mathcal{X}$, that is

$$sk(\mathcal{X}) = E\left[\left(\frac{\mathcal{X} - \mu}{\sigma}\right)^3\right]. \tag{1}$$

The intuition is the following:

- If the distribution is symmetric around the mean, then cubes of values to the left of μ, which are negative, are offset by cubes of the values to the right of μ, so that such a distribution has skewness 0.

- When cubing a number z with $|z| < 1$, then $|z^3| < |z|$, and when cubing a z with $|z| > 1$, then $|z^3| > |z|$. Therefore, when cubing $\frac{\mathcal{X}-\mu}{\sigma}$ and taking the expected value, then the contribution of values of $\mathcal{X}$ within a distance of σ from μ is decreased, and the contribution of values beyond that distance is increased.

In this way, most weight is given to those values of $\mathcal{X}$ far away from μ that are not balanced by values on the other side of μ. If there are more such values to the left of μ, expression (1) is negative, if they balance out it is close to 0, and if there are more to the right it is positive.

Remember that the variance of $\mathcal{X}$ satisfies the equation

$$Var(\mathcal{X}) = E\left[\mathcal{X}^2\right] - \mu^2.$$

1. Find an equation where $sk(\mathcal{X})$ is expressed in terms of $E[\mathcal{X}^3]$ and possibly μ and σ.

Suppose $\mathcal{X}$ is exponentially distributed with rate λ, that is $\mathcal{X} \sim Exp(\lambda)$.

2. Determine $E[\mathcal{X}^3]$.

   Hint: Use LOTUS and the fact that $E[\mathcal{X}] = 1/\lambda$ and $Var(\mathcal{X}) = 1/\lambda^2$. You can obtain the answer with a single application of integration by parts.

3. Using your answer to the first subquestion, determine $sk(\mathcal{X})$. How does $sk(\mathcal{X})$ depend on the rate λ?

(Weight: 10% of this CW)

5 Bob's Emails

Bob gets different numbers of emails on week days (Monday–Friday) and on weekends (Saturday–Sunday). In both cases, the number of incoming mails can be modeled by a Poisson distribution, with a rate of one email every 6 minutes for the weekdays and one email every 30 minutes for the weekends.

1. What is the probability that Bob receives k emails during a given hour on a week day?

2. What is the probability that Bob will get no emails during the four hours from 1pm to 5pm on Sunday?

3. What is distribution of the number of emails Bob receives during a working day of 24 hours? (That is, what is the probability that he receives 0, 1, 2, etc. emails per day?)

4. What is the average number of emails he receives during a day? Explain your answer.

5. Suppose we choose a random day (during the week or on a weekend) and a random interval of length one hour on that day. Bob receives 5 mails during that interval. What is the probability that the chosen day is a week day?

(Weight: 10% of this CW)

6 When Do We See the First Six?

Suppose you repeatedly roll a die until you get for the first time a six.

1. What is the probability that you need exactly k rolls, $1 \leq k < \infty$, until you see a six for the first time?

2. Generalize your result to the case that you are performing k independent repetitions of a Bernoulli experiment with distribution $Bernoulli(p)$ until there is the first success.

3. Let $\mathcal{X}$ stand for the number of repetitions until the first success. What are the expected value $E[\mathcal{X}]$ and the variance of $Var(\mathcal{X})$? What are the expected value and the variance in the case of the dice roll experiment at the beginning?

(Weight: 20% of this CW)

7 How Tall are US Women?

The height of adult women in the United States is distributed with mean 164cm and standard deviation 6cm. We assume that the distribution is normal. Find the probability that a randomly chosen woman is

1. less than 160cm tall;

2. less than 178cm tall;

3. between 160 and 178 cm tall.
4. Alice is 182 cm tall. What percentage of women is shorter than Alice?
5. Find the probability that the average of the heights of two randomly chosen women exceeds 167 cm.
6. Find out the analogous probability for four randomly chosen women.

(Weight: 12% of this CW)

8 Who Wins at Roulette?

A roulette wheel has 37 slots, numbered 0, and 1 through 36. If you bet 1 on a specified number, you either win 35 if the roulette ball lands on that number or lose 1 if it does not. If you continually make such bets, approximate the probability that

1. you are winning after 35 bets;
2. you are winning after 1,000 bets;
3. you are winning after 100,000 bets.

Note that the question here is not to approximate how much you win or lose, but whether at the end you have more money that you had at the start.
Assume that each roll of the roulette ball is equally likely to land on any of the 37 numbers.

(Weight: 12% of this CW)

9 How Wrong are Rounded Numbers?

Fifty numbers are rounded off to the nearest integer and then summed. Suppose the individual roundoff errors are uniformly distributed between $-.5$ and $.5$.

1. What is the approximate probability that the resultant sum differs from the exact sum by more than 3?
2. Generalize the previous result to the case that there are n numbers and we are interested in the error bound ϵ.

(Weight: 10% of this CW)