**MATH/STAT395: Probability II** Spring 2020

# Homework 1

Due **April 17th, 2020** by 11:59pm

*Instructor: Vincent Roulet* *Teaching Assistant: Zhenman Yuen*

Upload your answers to the questions below to Canvas in a PDF file. All answers require a **clear and complete** mathematical explanation unless specified differently. An answer without explanation/derivation/proof will not be given credits. One exercise chosen at random will be graded and the rest will be given points for completion.

**Exercise 1** Expectation and variance computations of classic random variables

See the lecture note for the expressions of the expectation and the variance

1. Derive the proof of the expressions of the expectation and variance of $X \sim \text{Geom}(p)$

   *Hint for expectation:* Denote $g(x) = \frac{1}{1-x}$, then for $0 < x < 1$, $g(x) = \sum_{k=0}^{+\infty} x^k$ and $g'(x) = \sum_{k=0}^{+\infty} kx^{k-1}$

   *Hint for variance:* Decompose $\mathbb{E}[X^2] = \mathbb{E}[X] + \mathbb{E}[X(X-1)]$ and use that for $0 < x < 1$, $g''(x) = \sum_{k=0}^{+\infty} k(k-1)x^{k-2}$

2. Derive the proof of the expressions of the expectation and the variance of $X \sim \text{Poisson}(\lambda)$

   *Hint for variance:* Decompose $\mathbb{E}[X^2] = \mathbb{E}[X] + \mathbb{E}[X(X-1)]$

3. Derive the proof of the expression of the expectation and the variance of $X \sim \text{Unif}([a, b])$

4. Derive the proof of the expressions of the expectations and the variance of $X \sim \mathcal{N}(\mu, \sigma^2)$ from the expression of the p.d.f.

5. Derive the proof of the expression of the expectation and variance of $X \sim \text{Exp}(\lambda)$

6. *(Optional)* Derive the proof of the expressions of the expectation and the variance of $X \sim \text{Gamma}(r, \lambda)$

   *Hint:* Prove that for any $\alpha > 0$, $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$

7. *(Optional)* Derive the proof of the expression of the expectation of a hypergeometric random variable *Hint:* A hypergeometric r.v. $X$ can be written $X = I_1 + \ldots + I_n$ where $I_i$ is the indicator random variable that the i$^{\text{th}}$ pick from the set belongs to the set of items $A$. As shown later in the course, the random variables $I_i$ are exchangeable such that they all have the same distribution. Use this decomposition to compute the expectation.

**Exercise 2** Modelization

1. On average people have 6 partners in their life. What is the probability that you will have exactly one partner in your life?

   *Hint:* Model it as a Poisson variable

2. It's been 30 min since my date hasn't answered my text message. On average she/he answers after 2min. If she/he does not answer now in the next 30 minutes, knowing that I already waited 30 minutes, I'm going to break up. What is the probability that I break up?

   *Hint: Model it as an exponential variable*

3. *(Optional)* Three of my friends schedule to play a game at 9pm remotely. They are never on time and I always have to wait. The average number of my friends coming in a time interval $[a, b]$ is proportional to the size of this time interval, namely the number of friends coming during $[a, b]$ can be modeled as $X \sim \text{Poisson}(\lambda(b-a))$ with $\lambda = 1$. The number of friends coming in two disjoint time intervals $[a, b]$ $[c, d]$, with $[a, b] \cap [c, d] = \emptyset$, are independent.
   (a) What is the distribution of the time before one of my friend comes?
   (b) What is the distribution of the time before two of my friend comes?
   (c) What is the probability that I wait at most 15min before all my friends are there?
   (d) Which distribution do you recognize? How can it be generalized to $n$ friends?

**Exercise 3** Medians

1. Show that the median of continuous random variable with positive p.d.f. is uniquely defined
2. Exhibit an example of a continuous random variable for which the median is not uniquely defined
3. I have a date at a restaurant. On average people are on time for dates with a variance of 5min. How much should I arrive earlier to be sure at 95% that I am there before my date?

   *Hint:* Model it as a Gaussian, use calculators of the error function that you can find on the web to compute the appropriate quantile after an adequate change of variables.

**Exercise 4** Useful Lemma

1. Prove the following lemma

   **Lemma 1.** *Let $X$ be a non-negative r.v. on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that its expectation is defined and denote $F$ its c.d.f.*

   *(a) If $X$ is a discrete integer valued random variable then $\mathbb{E}[X] = \sum_{t=0}^{+\infty}(1 - F(t))$*
   *(b) If $X$ is a continuous random variable then $\mathbb{E}[X] = \int_0^{+\infty}(1 - F(t))dt$*

**Exercise 5** Cauchy Distribution

Choose a point uniformly at random from $\{(x, y) : x > 0, x^2 + y^2 < 1\}$. Let $S$ be the slope of the line through the chosen point and the origin.

1. Find the c.d.f. of $S$

   *Hint:* $\frac{Area\{[Y \leq sX] \cap [X>0, X^2+Y^2<1]\}}{Area\{[X>0, X^2+Y^2<1]\}} = \frac{\arctan(s)+\frac{\pi}{2}}{\pi}$
2. Find the p.d.f. of $S$

**Exercise 6** Not finite moments

1. Give an example of a random variable whose expectation is not finite (check first that it is indeed a random variable)
2. Let $k \in \mathbb{N}$. Give an example of a a random variable such that its $k^{\text{th}}$ moment is finite but not its $k+1$ moment.

| MATH/STAT395: Probability II | Spring 2020 |
|---|---|
| **Homework 2** | |
| Due **April 22th, 2020** by 11:59pm | |
| *Instructor: Vincent Roulet* | *Teaching Assistant: Zhenman Yuen* |

Upload your answers to the questions below to Canvas in a PDF file. All answers require a **clear and complete** mathematical explanation unless specified differently. An answer without explanation/derivation/proof will not be given credits. One exercise chosen at random will be graded and the rest will be given points for completion.

**Exercise 1** Joint distribution for Discrete Random Variables

1. An unfair coin has $\frac{2}{3}$ to show head up and $\frac{1}{3}$ to show tail up. We flip this coin three times. Let $X$ be the number of heads among the first two flips, and $Y$ the number of heads in the last two coin flips.
   (a) Find the joint probability mass function (p.m.f) of $(X, Y)$. (You can either try to find a general formula, or display the function in a table.)
   (b) Find the probability mass function of $XY$.
2. Let $X$ and $Y$ be independent Geom$(p)$ random variables. Let
$$W = \begin{cases} 0, \text{ if } X < Y \\ 1, \text{ if } X = Y \\ 2, \text{ if } X > Y \end{cases} \tag{1}$$
   Find the joint probability mass function of $W$.

**Exercise 2** Joint distribution for Continuous Random Variables

1. A random point $(X, Y)$ is distributed uniformly on the square with vertices $(1, 1), (1, -1), (-1, 1)$, and $(-1, -1)$. That is, the joint pdf is $f(x, y) = \frac{1}{4}$ on the square.
   Determine the probabilities of the following events.
   (a) $X^2 + Y^2 < 1$
   (b) $2X - Y > 0$
   (c) $|X + Y| < 2$
2. A pdf is defined by
$$f(x, y) = \begin{cases} C(x + 3y) & \text{if } 0 < y < 1 \text{ and } 0 < x < 2 \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$
   (a) Find the value of $C$.
   (b) Find the marginal distribution of $X$.
   (c) Find the joint cdf of $X$ and $Y$.
   (d) *(Optional)* Find the pdf of the random variable $Z = \frac{9}{(X+1)^2}$.
3. *(Optional)* A stick is of length 1. You choose two break points $U$ and $V$ independently and $U \sim \text{Unif}([0, 1]), V \sim \text{Unif}([0, 1])$. Now you break this stick into 3 pieces. What's the probability that these 3 pieces could form a triangle?

**Exercise 3** Multinomial Distribution

1. An urn contains 1 green ball, 1 red ball, 1 yellow ball and 1 white ball. I draw 3 balls with replacement. What is the probability that exactly two balls are of the same color?

2. Human being own 4 types of basic blood type: AB, A, B and O. In country C, 40% of people are O type, 30% of people are A type, 20% of people are B type and 10% of people are AB type. Now you are a doctor and have two patients waiting for blood transfusion. One of them own blood type O, can only accept blood type O transfusion; the other own blood type A, can receive either type O or A transfusion. Assume every donor's contribution can only support one patient(that is, one donor cannot donate blood to both patients). If you choose 5 people randomly from the population, what's the probability that both of the patients are able to get blood transfusion? What's the probability that exactly one of them is able to get blood transfusion?

**Exercise 4** Marginal and Independence Discrete Case

1. The random pair $(X, Y)$ has the distribution

| | | | X | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| | 2 | 1/12 | 1/6 | 1/12 |
| Y | 3 | 1/6 | 0 | 1/6 |
| | 4 | 0 | 1/3 | 0 |

   (a) Show that $X$ and $Y$ are dependent.

   (b) Give a probability table for random variables $U$ and $V$ that have the same marginals as $X$ and $Y$ but are independent.

**Exercise 5** Functions of Random Variables

1. Let $Z \sim \mathcal{N}(0, 1)$, find the p.d.f. of $|Z|$

2. *(Optional)* Let $r, s > 0$, a random variables has the beta distribution with parameters $(s, r)$ if $0 < X < 1$ and has a p.d.f.

$$f(x) = \begin{cases} \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} x^{r-1}(1-x)^{s-1} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

   denoted $X \sim \text{Beta}(r, s)$.

   (a) Let $X, Y$ independent with $X \sim \Gamma(r, \lambda)$, $Y \sim \Gamma(s, \lambda)$. Find the joint distribution of $(B, G) = \left(\frac{X}{X+Y}, X + Y\right)$.

   (b) Let $(B, G)$ be independent with $B \sim \text{Beta}(r, s)$ and $G \sim \text{Gamma}(r = s, \lambda)$. Find the joint distribution of $(X, Y) = (BG, (1 - B)G)$

**Exercise 6** Minimum, Maximum, Sums of Independent Random Variable

1. Let $X_1, X_2, ..., X_n$ be independent exponential random variables, with parameter $\lambda_i$ for $X_i$. Let $Y$ be the minimum of these random variables. What's the distribution of Y?

2. Let $X_1, X_2, ..., X_n$ be i.i.d Unif[0,1] random variables. Let $Y^1 := \min\{X_1, ..., X_n\}$, $Y^n := \max\{X_1, ..., X_n\}$. What's the probability distribution function (p.d.f) of $Y^1$ and $Y^n$? [Hint: this is a well-known distribution, which closely related to Gamma distribution] And what's the expectation of $E[Y^1]$, $E[Y^n]$? Describe what you find.

3. Let $X, Y$ be independent with $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Ber}(p)$. Find the p.m.f. of $X + Y$.

| **MATH/STAT395: Probability II** | Spring 2020 |
|---|---|

# Homework 3

Due **May 6th, 2020** by 11:59pm

*Instructor: Vincent Roulet* — *Teaching Assistant: Zhenman Yuan*

Upload your answers to the questions below to Canvas in a PDF file. All answers require a **clear and complete** mathematical explanation unless specified differently. An answer without explanation/derivation/proof will not be given credits. One exercise chosen at random will be graded and the rest will be given points for completion, i.e., if the final answer provided is correct.

**Training exercises.**
Do the examples provided in the lectures (the ones done in class and the additonal ones) without looking at the solution.

**Exercise 1** Exchangeability

1. We have an urn with 20 red, 10 black and 15 green balls. We take a sample of 30 balls without replacement, with order. Find the probability that the 3rd, 10th and 23rd picks are of different colors.
2. Let $X_1, X_2, X_3$ be i.i.d. $\text{Exp}(\lambda)$ random variables. Find $\mathbb{P}(X_1 < X_2 < X_3)$.
3. In the lectures, we showed that: i.i.d $\Rightarrow$ exchangeability. The following exercise explores the reverse claim and the necessary conditions of the claim namely having both independent **and** identically distributed random variables.
   (a) Give an example of three random variables $(X, Y, Z)$ that are exchangeable, but not i.i.d.
   (b) Give an example of three random variables $(X, Y, Z)$ that are independent, but not exchangeable.
   (c) (Optional) Give an example of three random variables $(X, Y, Z)$ that are identically distributed, but not exchangeable.

**Exercise 2** Expectation and Variance

1. Let $X$ and $Y$ be two i.i.d. r.v. with mean $\mu$ and variance $\sigma^2$, compute $\mathbb{E}[(X-Y)^2]$
2. Let $(X, Y)$ be a point chosen at random on the unit square $[0,1] \times [0,1]$. Find $\text{Var}[XY]$.
   *Hint:* Are $X, Y$ independent? What are their marginals? Finally use the definition of the variance for $Z = XY$.

**Exercise 3** Modelization about Expectation and Variance
In the following exercises, the idea is to decompose a random variable as a sum of simple independent random variables to compute its expectation and variance. In particular, it is sometimes useful to decompose a random variable as the sum of indicator random variables.
Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the indicator random variable of an event $A \subset \Omega$ is defined as

$$\mathrm{I}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

This is naturally a discrete random variable as a function from $\Omega$ to $\{0, 1\}$.
The first exercise illustrates the method on a simple problem.

1. Adam has 3 siblings, Ben, Chris and Daisy. On a random day Ben calls Adam with probability 0.3, Chris calls Adam with probability 0.4 and Daisy calls Adam with probability 0.7. Let $X$ be the number of siblings that call Adam tomorrow.

   (a) What is the expectation of $X$?

   (b) Assuming that all siblings call Adam in an independent manner, what is the variance of $X$?

   *Method:* Denote $A_1$ the event that "Ben calls Adam". Similarly denote $A_2$ and $A_3$ the events that Chris calls Adam and Daisy calls Adam respectively.

   Using the indicator random variables of $A_1$, $A_2$, $A_3$, we have

$$X = \mathrm{I}_{A_1} + \mathrm{I}_{A_2} + \mathrm{I}_{A_3}$$

   The expectation and variances can then easily be computed by observing that these random variables are Bernoulli random variables with parameters given in the statement of the exercise.

2. We choose a number from the set $\{0, ..., 9999\}$ randomly, and denote by $X$ the sum of its digits. Find the expected value of $X$.

3. I roll a fair die 4 times. Let $X$ be the number of different outcomes that I see. (For example, if the die rolls are 5,3,6,6 then X = 3 because the different outcomes are 3, 5 and 6.)

   (a) Find E[X].

   (b) Find Var[X].

   *Hint:* Again you need to define $X$ as a sum of indicator random variables that depend on the rolls of the die.

4. Let $(X_1, ..., X_n)$ and $(Y_1, ..., Y_n)$ be two i.i.d random vectors with each vector being a random ordering of $k$ ones and $n-k$ zeros. That is, the joint probability mass function of $(X_1, \ldots X_n)$ (same for $(Y_1, \ldots, Y_n)$) is

$$\mathbb{P}(X_1 = i_1, ..., X_n = i_n) \begin{cases} 1/\binom{n}{k} & \text{if } \sum_{j=1}^{n} i_j = k \text{ and } i_j \in \{0,1\} \text{ for all } j \in \{1, \ldots n\} \\ 0 & \text{otherwise} \end{cases}$$

   Let $N = \sum_{i=1}^{n} |X_i - Y_i|$ denote the number of coordinates at which the two vectors have different values.

   (a) (Optional) Denote $M$ the number of indexes $i$ for which $X_i = 1, Y_i = 0$. Show that $M \sim \text{Hypergeom}(n, k, n-k)$.

   (b) Find the expectation of $N$. (You can admit that $M \sim \text{Hypergeom}(n, k, n-k)$)

5. (Optional) Let $X_1, X_2, ...$ be a sequence of independent and identically distributed continuous random variables. Let $N \geq 2$ be such that

$$X_1 \geq X_2 \geq .... \geq X_{N-1} < X_N \tag{1}$$

   That is, $N$ is the point at which the sequence stops decreasing. Show that $E[N] = e$.

   *Hint:* Write $N$ as a sum of indicator random variables. The probability of each event associated to the indicator random variables can then be computed by using the exchangeability coming from the i.i.d. assumption.

**Exercise 4** (Optional) Mutually Independent and Pairwise Independent

This exercise is outside the scope of the lectures done last week. But this can be seen as an interesting complement of the course and can be done with the tools you have seen before.

Given $n$ random variables $X_1, X_2, ..., X_n$, if $X_i$ are independent of $X_j$ for any pair of them $(X_i, X_j), i \neq j$, then we say $X_1, X_2, ..., X_n$ are *pairwise independent*. Independence (also called mutual independence to stress the difference with pairwise independence) of the random variables $X_1, X_2, ..., X_n$ always implies pairwise independence, while the converse is not true. Here is a simple counter-example.

Suppose $X_1$ and $X_2$ are two independent random variable with $P[X_i = 1] = P[X_i = -1] = \frac{1}{2}$. Let $Z = X_1 X_2$. Then show that:

1. $(X_1, X_2, Z)$ are identically distributed and pairwise independent.

2. $X_1, X_2$ and $Z$ are not mutually independent.

| **MATH/STAT395: Probability II** | Spring 2020 |
|---|---|
| **Homework 4** | |
| Due **May 13th, 2020** by 11:59pm | |
| *Instructor: Vincent Roulet* | *Teaching Assistant: Zhenman Yuan* |

Upload your answers to the questions below to Canvas in a PDF file. All answers require a **clear and complete** mathematical explanation unless specified differently. An answer without explanation/derivation/proof will not be given credits. One exercise chosen at random will be graded and the rest will be given points for completion.

**Training exercises.**
Do the examples provided in the lectures (the ones done in class and the additonal ones) without looking at the solution.

**Exercise 1** Covariance and Correlation

1. The random variables $X$ and $Y$ have a joint density function given by

$$f(x, y) = \begin{cases} \frac{2e^{-2x}}{x}, & 0 \leq x < \infty, 0 \leq y \leq x \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

   Compute Cov(X, Y).

2. If $X_1, X_2, X_3$ and $X_4$ are (pairwise) uncorrelated random variables, each have mean 0 and variance 1, compute the correlations of
   (a) $X_1 + X_2$ and $X_2 + X_3$;
   (b) $X_1 + X_2$ and $X_3 + X_4$.

**Exercise 2** Variance of Sum of Random Variables
Consider a urn that contains 5 red balls and 6 green balls. Draw 3 balls without replacement from this urn and let $X$ be the number of red balls.

1. What is $\mathbb{E}(X)$?
2. What is $\text{Var}(X)$?
3. Denote $Y$ the number of red balls that you get if you drew 11 balls. What is $\text{Var}(Y)$? How could you have found this result without computing the variance?
4. Assume that the number of balls in the urn grows to infinity while the portion of red balls in the urn remains constant and the number of balls you draw remains constant. (In the notations below, $N \to +\infty, K \to +\infty$, such that $N/K = p$ remains constant). What variance do you obtain in this limit? Do you recognize the variance of a classical distribution?

*Hint:* This is an instance of a hyper-geometric random variable. Denote $N = 11$ the total number of balls. $K = 5$ the number of red balls and $n = 3$ the number of balls that are drawn from the urn. (Try to express your result first in terms of $N$, $n$, $p = K/N$ and $q = 1 - p = (N - K)/N$. Then you can plug the values of $N$, $K$, $n$). The p.m.f. of $X$ is known such that one can use the definition of expectation and variance with respect to the p.m.f. Computing the expectation and the variance can be done more easily by decomposing $X$ in a sum of exchangeable indicator random variables.

**Exercise 3** Multivariate Normal Distribution

1. Suppose $(X, Y, Z)$ follow the multivariate normal distribution:

$$(X, Y, Z) \sim \mathcal{N}\left[\begin{pmatrix}1\\3\\4\end{pmatrix}, \begin{pmatrix}1 & 0.4 & 0.2\\0.4 & 1 & 0.3\\0.2 & 0.3 & 1\end{pmatrix}\right] \tag{2}$$

   (a) Find the distribution of $X + Y + Z$;

   (b) Find $\mathrm{Cov}(2X - 3Y + Z, X + Y - Z)$.

   *Hint for 1.:* A linear transform of a univariate/multivariate normal r.v. is a univariate/multivariate normal random variable. A univariate/multivariate normal random variable is entirely characterized by its mean and its variance/covariance matrix.

2. (Optional) Let $\mathbf{X} = (X_1, \ldots, X_n)^\top$ be a random vector and $S$ be its covariance matrix.
   Show that $S$ is a *positive semi-definite matrix*, i.e., for any $\mathbf{u} \in \mathbb{R}^n$, $\mathbf{u}^\top S \mathbf{u} \geq 0$

**Exercise 4** Moment Generating Functions

1. Derive the moment generating function of the following r.v.:

   (a) $X \sim \mathrm{Geom}(p)$ for $p \in (0, 1)$

   (b) $X \sim \mathrm{Unif}([a, b])$ for $a, b \in \mathbb{R}$, $a < b$

   (c) (Optional) $\mathrm{Gamma}(r, \lambda)$, $r \geq 1$, $\lambda \geq 0$

2. Given that $X$ has moment-generating function

$$M(t) = \frac{1}{6}e^{-2t} + \frac{1}{3}e^{-t} + \frac{1}{4}e^{t} + \frac{1}{4}e^{2t}, \tag{3}$$

   find $P(|X| \leq 1)$.

3. Let $\lambda > 0$ be a fixed number. Suppose $X_n \sim \mathrm{Bin}(n, \frac{\lambda}{n})$, for $n \in \mathbb{N}^+$.

   (a) Find the moment-generating function $M_{X_n}(t)$ for $X_n$.
   *Hint*: Use the decomposition of $X_n$ as a sum of simple r.v.

   (b) Let $n \to \infty$, what is the point-wise limit of $M_{X_n}(t)$? What is the distribution of the m.g.f. that correspond to this limit?

4. (a) Let $Z \sim \mathcal{N}(0, 1)$. Show that

$$E[Z^n] = \begin{cases} 0 & \text{if } n \text{ is odd} \\ \frac{(2j)!}{2^j j!} & \text{if } n = 2j \end{cases}$$

   (b) Find $\mathrm{Cov}[Z^2, Z^4]$.

**Exercise 5** Modelization

1. We have an urn with 6 red balls and 4 green balls. We draw balls from the urn one by one without replacement, noting the order of the colors, until the urn is empty. Let $X$ be the number of red balls in the first five draws, and $Y$ the number of red balls in the last five draws. Compute the covariance Cov[X, Y].

2. Let $X$ equal the outcome when a fair four-sided die that has its faces numbered 0, 1, 2, and 3 is rolled. Let $Y$ equal the outcome when a fair four-sided die that has its faces numbered 0, 4, 8, and 12 is rolled.

(a) Define the m.g.f. of $X$.

(b) Define the m.g.f. of $Y$.

(c) Let $W = X + Y$, the sum when the pair of dice is rolled. Find the m.g.f. of $W$.

(d) Give the p.m.f. of $W$; that is, determine $P(W = w), w = 0, 1, ...., 15$, from the m.g.f. of W.