CMPT 210: Probability & Computing
Assignment 4

Total marks: 200

Due: Via Coursys at 11.59 pm, Tuesday, April, 9 April
Late Submission: 11.59 pm, Thursday, 11 April

Instructions on how to solve, write, and submit your assignment

Solutions: Solutions to assignments must be your own. Use sample exercises from the lectures to learn methods and approaches you can use. In some cases expect that you will need to make a substantial effort to solve a problem. Discussing and collaborating with other people is okay, as long as you produce your own solution. For instance, even when two people try to solve the problem together and extensively discuss possible solutions, if they write down that solution independently, it is acceptable.

We treat any kind of academic dishonesty very seriously. For the SFU policy on academic dishonesty see the part of University Policy S 10.01 relevant for us:

> e. Cheating in assignments, projects, examinations or other forms of evaluation by:
> i. using, or attempting to use, another student's answers;
> ii. providing answers to other students;
> iii. failing to take reasonable measures to protect answers from use by other students; or
> iv. in the case of students who study together, submitting identical or virtually identical assignments for evaluation unless permitted by the course Instructor or supervisor.
>
> University Policy S 10.01
> Code of Academic Integrity and Good Conduct
> 4.1.2 Forms of Academic Dishonesty

Note that this policy treats copying and allowing to copy equally. Should anyone be caught submitting a work too similar to someone else's work, or a source found on the web, a record of the violation will permanently stay in their student file.

Writing the solution: There is no strict prescribed way to present your solution. However, make sure that your solution can be understood by another person without any help on from you. The onus to present your solution in a clear understandable way is on you. Any unclear steps or arguments will be considered incorrect. If you use some method, or result presented in this course you do not need to explain it. However, if you would like to use a result or approach from elsewhere, please, give a reference and explain what you are doing in more details.

Submission: The assignment needs to be submitted via Coursys. For some flexibility, each student is allowed 1 late-submission.

(1) **[45 marks]** Let X be the number of 1's and Y the number of 2's that occur in n rolls of a standard dice.

- Compute $\mathbb{E}[X]$ and $\mathrm{Var}[X]$. [5 marks]
- Are the random variables X and Y independent? Prove or give a counter-example. [5 marks]
- If $Z = X + Y$, compute $\mathbb{E}[Z]$ and $\mathrm{Var}[Z]$. [15 marks]
- Use the above results to compute $\mathbb{E}[(X+Y)^2]$ and $\mathbb{E}[X\,Y]$. [10 marks]
- Use the above results to compute $\mathrm{Cov}[X, Y]$. [5 marks]
- Use the above results to compute $\mathrm{Corr}[X, Y]$. Are X and Y positively or negatively correlated? [5 marks]

(2) **[25 marks]** **Game 1**: You are playing a game where you get n turns. Each of your turns involves flipping a coin a number of times. On the first turn, you have 1 flip, on the second turn you have two flips, and so on until your n-th turn when you flip the coin n times. All the flips are mutually independent. The coin you are using is biased to flip Heads with probability p. You win a turn if you flip ***all*** Heads in that turn. Let X be the r.v. equal to the number of winning turns.

- Calculate $\mathbb{E}[X]$ and $\mathrm{Var}[X]$. [10 marks].

Game 2: You are playing a game where you get n turns. Each of your turns involves flipping a coin a number of times. On the first turn, you have 1 flip, on the second turn you have two flips, and so on until your nth turn when you flip the coin n times. All the flips are mutually independent. The coin you are using is biased to flip Heads with probability p. You win a turn if you flip ***at least one*** Heads in that turn. Let Y be the r.v. equal to the number of winning turns.

- Calculate $\mathbb{E}[Y]$ and $\mathrm{Var}[Y]$. [15 marks].

(3) **[20 marks]** In class, we have seen the two-sided Chebyshev's theorem,

$$\Pr[|R - \mathbb{E}[R]| \geq x] \leq \frac{\mathrm{Var}[R]}{x^2}.$$

The above result bounds the probability that R can take values smaller or larger than $\mathbb{E}[R]$ by an amount x. If we are only interested in bounding the probability that R takes values larger than $\mathbb{E}[R]$, we can use the one-sided Chebyshev's Theorem to obtain a tighter bound.

$$\Pr[R - \mathbb{E}[R] \geq x] \leq \frac{\mathrm{Var}[R]}{x^2 + \mathrm{Var}[R]}$$

- In order to prove the above statement, define $Y := (R - \mathbb{E}[R] + a)^2$ for $a \geq 0$. This implies that if $R - \mathbb{E}[R] \geq x$, then, $Y \geq (x+a)^2$. Using this reasoning, apply the Markov bound to the r.v Y, and prove the following statement:

$$\Pr[R - \mathbb{E}[R] \geq x] \leq \frac{a^2 + \mathrm{Var}[R]}{(a+x)^2} \text{ [15 marks]}$$

- Prove the one-sided Chebyshev's Theorem by finding the best value of a. To obtain the tightest bound, minimize the RHS w.r.t a. [5 marks]

(4) **[50 marks]** From past experience a professor knows that the number of marks that a student gets on their final examination is a r.v. with mean 75 (out of 100).

- Give an upper bound for the probability that a student's score will exceed 85. [5 marks]
- If we know that no student gets under 40 marks, improve the above upper-bound on the probability that a student's score will exceed 85. [5 marks]
- If we further know the variance in the student scores to be 25, use Chebyshev's inequality to improve the above upper-bound on the probability that a student's score will exceed 85. [5 marks]
- Using the one-sided Chebyshev inequality from Question (3), further improve the above upper-bound on the probability that a student's score will exceed 85. [5 marks]
- Under the same assumptions that the student scores have mean 75, variance 25 and are always greater than 40, define S_i as the score of student i in a class of n students. Assume that the scores of all students are mutually independent. If $Z_i := \frac{S_i - 40}{60}$, prove that $Z_i \in [0, 1]$, $\mathbb{E}[Z_i] = \frac{7}{12}$ and $\text{Var}[Z_i] = \frac{1}{144}$. [10 marks]
- Using the one-sided Chebyshev's inequality to bound the deviation of $Z := \sum_{i=1}^{n} Z_i$ from its mean, calculate how many students would need to take the examination to ensure, with probability of at least 0.999, that the class average would be at most 85? [10 marks]
- Using the Chernoff bound to bound the deviation of $Z := \sum_{i=1}^{n} Z_i$ from its mean,, calculate how many students would need to take the examination to ensure, with probability of at least 0.999, that the class average would be at most 85? (NOTE: $\exp(x) := e^x$ in the Chernoff bound) [10 marks]

(5) **[30 marks]** 1000 files $F_1, F_2, \ldots, F_{1000}$ have reached a disk manager for writing onto disk. Each file's size is between 0 MB and 1 MB. The sum of all files' sizes is 400 MB. The disk manager has 4 disks under its control. For each file F_i, the disk manager chooses a disk uniformly at random from amongst the 4 disks, and F_i is written to that disk. The choices of disk for the different files are mutually independent.

- If D_1 is the r.v. equal to the MB written onto disk 1, calculate $\mathbb{E}[D_1]$. [10 marks]
- Using Markov's inequality, find the upper bound on the probability that 200 MB or more are written on the first disk? [5 marks]
- Using the Chernoff bound, find the upper bound on the probability that 200 MB or more are written on the first disk? [10 marks]
- Using the above result for the Chernoff bound, find an upper-bound on the probability that there is some disk with 200 MB or more written on it? [5 marks]

(6) **[30 marks]** The Poisson distribution is important because it can approximate the binomial distributions with large n and relatively small values of np (rule of thumb is when $n \geq 20$ and $np \leq 10$). If a r.v. X is distributed according to the Poisson distribution, its range is $\{0, 1, 2, \ldots, \infty\}$ and its PDF is given as:

$$\text{PDF}_X[k] = \Pr[X = k] = \frac{\exp(-\lambda)\,\lambda^k}{k!}$$

- Verify that the above PDF indeed corresponds to a valid distribution (use the Taylor series expansion of $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$). [5 marks]
- Prove that $\mathbb{E}[X] = \lambda$ and $\mathrm{Var}[X] = \lambda$ [10 marks]

If X is a random variable from a certain distribution, then the function

$$\phi(t) := \mathbb{E}[\exp^{tX}] := \sum_{x \in \mathrm{Range}(X)} \exp(t\,x) \Pr[X = x]$$

is referred to as the *moment generating function* of the specific distribution.

The name *moment generating function* comes from the fact that derivatives of $\phi(t)$ can be used to generate the different moments of the distribution. For example,

$$\begin{aligned} \mathbb{E}[X] &= \phi'(0) := \frac{d}{dt}[\phi(t)]\Big|_{t=0} \\ \mathbb{E}[X^2] &= \phi''(0) = \frac{d}{dt}[\phi'(t)]\Big|_{t=0} \end{aligned}$$

and so on. To see this, note that,

$$\begin{aligned} \phi'(t) &= \frac{d}{dt}\left[\sum_{x \in \mathrm{Range}(X)} \exp(t\,x) \Pr[X = x]\right] = \left[\sum_{x \in \mathrm{Range}(X)} x \exp(t\,x) \Pr[X = x]\right] \\ \phi'(0) &= \left[\sum_{x \in \mathrm{Range}(X)} x \exp(0\,x) \Pr[X = x]\right] = \left[\sum_{x \in \mathrm{Range}(X)} x \Pr[X = x]\right] = \mathbb{E}[X] \end{aligned}$$

Similarly, we can show that for $q > 1$, $\mathbb{E}[X^q] = \frac{d^q}{dt^q}\phi(t)\Big|_{t=0}$.

Using the above definition,

- Prove that if $X \sim \mathrm{Poisson}(\lambda)$, then $\phi(t) = \exp\Big(\lambda\,(\exp(t) - 1)\Big)$. [8 marks]
- Using this moment generating function, prove that if $X \sim \mathrm{Poisson}(\lambda)$, then $\mathbb{E}[X] = \lambda$ and $\mathrm{Var}[X] = \lambda$. [7 marks]