**Autumn Semester** **August 12, 2022**

# DS203:Programming for Data Science

Assignment 2

---

**Exercise 1 (5 points)** *If $X_1$ and $X_2$ are independent binomial random variables with respective parameters $(n_1, p)$ and $(n_2, p)$. Calculate the conditional probability mass function of $X_1$ given that $X_1 + X_2 = m$.*

**Exercise 2 (5 points)** *Give an example of two random variables X and Y that are uncorrelated but not independent.*

**Exercise 3 (5 points)** *Suppose $X$ and $Y$ have joint density function $f_{X,Y}(x, y) = c(1 + xy)$ if $2 \leq x \leq 3$ and $1 \leq y \leq 2$, and $f_{X,Y}(x, y) = 0$ otherwise.*

1. *Find c.*
2. *Find $f_X$ and $f_Y$.*

**Exercise 4 (5 points)** *An insurance company supposes that the number of accidents that each of its policyholders will have in a year is Poisson distributed, with the mean of the Poisson depending on the policyholder. If the Poisson mean of a randomly chosen policyholder has a gamma distribution with density function,*

$$g(\lambda) = \lambda e^{-\lambda}, \qquad \lambda \geq 0$$

*what is the probability that a randomly chosen policyholder has exactly n accidents next year?*

**Exercise 5 (5 points)** *Suppose that the number of people who visit a yoga studio each day is a Poisson random variable with mean λ. Suppose further that each person who visits is, independently, female with probability p or male with probability $1 - p$. Find the joint probability that exactly n women and m men visit the academy today.*

**Exercise 6 (5 points)** *Let $X_1, X_2, X_3$ are RVs and $a, b, c, d$ are constants. Show that*

- $Cov(aX_1 + b, cX_2 + b) = acCov(X_1, X_2)$
- $Cov(X_1 + X_2, X_3) = Cov(X_1, X_3) + Cov(X_2, X_3)$.

**Exercise 7** *Consider $n$ RVs $X_i : i = 1, 2, \ldots, n$ where each $X_i \sim Ber(p)$ for some $p \in (0, 1)$. Assume that $X_i$s are independent. Show that $X = \sum_{i=1}^{n} X_i$ is $Bin(n, p)$.*

**Exercise 8 (5 points)** *You are given $n = 100$ i.i.d. samples generated from a random experiment where the outcome of experiment lies in $[0, 1]$. Let the estimate of mean from these samples is $\hat{\mu} = 0.45$. We know that true mean lies somewhere around $\hat{\mu}$ and we would like to find an interval (around $\hat{\mu}$) such that the true value lies in the interval with probability at least $0.95$.*

- *What would be your (confidence) interval? Specify the method you used to come up with the interval.*
- *If you want the your confidence interval to shrink by half, how many more samples would you need? (the estimate could be different now)*