Handout 1: Random variables - Inequalities - Limit theorems

Due date : February 3^{rd}

Brief recap 1.

1. PMFs are usually more intuitive to retrieve for discrete rvs, whereas CDFs are more intuitive for continuous rvs.
2. For a PMF to be valid, it needs to satisfy two requirements, $p(e) \geq 0$ for any event, and $\sum_{i \in E} = 1$, where E is the sample space.
3. For a PDFs f to be valid, it needs to satisfy three requirements, $f \geq 0$ for any event, f is piecewise continuous, and $\int_{-\infty}^{\infty} f(x)dx = 1$, where E is the sample space.
4.

Exercise 1. Consider a sequence of independent Bernoulli trials, each with the same success probability $p \in [0, 1]$, with trials performed until a success occurs. Let X be the number of failures before the first successful trial.

1. Determine $P(X = k)$.
2. Show that it is a valid PMF.
3. Calculate $E[X]$.

This distribution is called the geometric distribution $X \sim \text{Geom}(p)$.

Exercise 2.

1. Show that $p(n) = (1/2)^{n+1}$ for $n = 0, 1, 2, \ldots$ is a valid PMF for a discrete r.v.
2. Find the CDF of a random variable with the PMF in 1.

Exercise 3.

1. Benford's law states that in a very large variety of real-life data sets, the first digit approximately follows a particular distribution with about a 30% chance of a 1, an 18% chance of a 2, and in general

$$P(D = j) = \log_{10}\left(\frac{j+1}{j}\right), \quad \forall j \in \{1, 2, 3, \ldots, 9\}.$$

where D is the first digit of a randomly chosen element. Check that this is a valid PMF.

Exercise 4. Let $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p)$, independent of X.

1. Show that $X + Y$ is Binomial and give its distribution.
2. Is $X - Y$ Binomial? If yes, give its distribution. If not, why?

Exercise 5. There are two coins, one with probability p_1 of Heads and the other with probability p_2 of Heads. One of the coins is randomly chosen (with equal probabilities for the two coins). It is then flipped $n \geq 2$ times. Let X be the number of times it lands Heads.

1. Find the PMF of X.
2. What is the distribution of X if $p_1 = p_2$?

Exercise 6. Student A flips a fair coin n times and student B flips another fair coin $n+1$ times, resulting in independent random variables $X \sim \text{Bin}(n, \frac{1}{2})$ and $Y \sim \text{Bin}(n+1, \frac{1}{2})$.

1. Show that $P(X < Y) = P(n - X < n + 1 - Y)$.
2. Compute $P(X < Y)$.

Exercise 7. Prove that $E[X]^2 \leq E[X^2]$.

Exercise 8. Let X be a non-negative rv and $a > 0$ be some fixed number.

1. Show that a $\mathbb{I}_{|X| \geq a} \leq X$, where $\mathbb{I}$ is the indicator rv.
2. Prove the Markov inequality using the previous result.

Exercise 9. Let X be the number of purchases that Fred will make on the online site for a certain company (in some specified time period). Suppose that the probability mass function (PMF) of X is

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

This distribution is called the Poisson distribution with parameter λ, and it will be studied extensively in later chapters.

1. Find $P(X \geq 1)$ and $P(X \geq 2)$ using 2 different methods.
2. Suppose that the company only knows about people who have made at least one purchase on their site (a user sets up an account to make a purchase, but someone who has never made a purchase there doesn?t appear in the customer database). If the company computes the number of purchases for everyone in their database, then these data are draws from the conditional distribution of the number of purchases, given that at least one purchase is made. Find the conditional PMF of X given $X \geq 1$. (This conditional distribution is called a truncated Poisson distribution.)

Exercise 10. In a national survey, a random sample of people are chosen and asked whether they support a certain policy. Assume that everyone in the population is equally likely to be surveyed at each step, and that the sampling is with replacement. Let n be the sample size, and let $\hat{p}$ and p be the proportion of people who support the policy in the sample and in the entire population, respectively. Show that

$$\forall c > 0, \quad P(|\hat{p} - p| > c) \leq \frac{1}{4nc^2}. \tag{1}$$

Exercise 11. For i.i.d. r.v.s $X_1, \ldots, X_n$ with mean μ and variance σ^2, give a value of n that will ensure that there is at least a 99% chance that the sample mean will be within 2 standard deviations of the true mean μ.

Exercise 12. Let $X_1, X_2, \ldots$ be i.i.d. positive random variables with mean 2. Let $Y_1, Y_2, \ldots$ be i.i.d. positive random variables with mean 3.

1. Show that
$$\frac{X_1 + X_2 + \ldots + X_n}{Y_1 + Y_2 + \ldots + Y_n} \rightarrow \frac{2}{3}$$
with probability 1. Does it matter whether the X_i are independent of the Y_j ?

Exercise 13. Let $Y_1, Y_2, \ldots, U_{100}$ be i.i.d. r.v.s with distribution $\text{Unif}(0, 1)$ and $X = Y_1 + Y_2 + \ldots + U_{100}$.

1. Which important distribution is the distribution of X very close to? Specify what the parameters are, and state which theorem justifies your choice.

2. Give a simple but accurate approximation for $P(X > 17)$. Justify briefly.

References and acknowledgments: Introduction to probability (Blitzstein and Huang) - Stochastic Processes (Gallager)