Homework 2 - Probability review

---

1) Obtaining the pmf from the cdf of a discrete random variable. Consider a discrete random variable (RV) X with cumulative distribution function (cdf) given by

$$F(x) = \mathrm{P}[X \leq x] = \begin{cases} 0, & x < 0 \\ \frac{1}{2}, & 0 \leq x < 1 \\ 1, & 1 \leq x < \infty \end{cases}$$

Use the identity $p(x) = F(x) - \lim_{y \to x^-} F(y)$ to obtain the probability mass function (pmf) $p(x) = \mathrm{P}[X = x]$ of X. Plot both $F(x)$ and $p(x)$.

2) Continuous random variables. Let X be a continuous RV with probability density function (pdf)

$$f(x) = \begin{cases} c(4x - 2x^2), & 0 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

Determine the value of c, $\mathrm{P}[1/2 < X < 3/2]$, $\mathbb{E}[X]$, $\mathrm{var}[X]$, and $\mathrm{var}[5X + 2]$.

3) Expectation of nonnegative integer-valued random variables. If X is a nonnegative integer-valued RV, show that

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} \mathrm{P}[X \geq n] = \sum_{n=0}^{\infty} \mathrm{P}[X > n].$$

(Hint: Define the indicator RVs $I_n = \mathbb{I}\{X \geq n\}$, $n \geq 1$, and express X in terms of the I_n.)

4) Joint pmf, marginal pmfs, and independence. Suppose X and Y have the following joint pmf

$$\mathrm{P}[X = -1, Y = 2] = 1/4, \quad \mathrm{P}[X = -1, Y = 4] = 1/6$$
$$\mathrm{P}[X = 0, Y = 2] = 1/3, \quad \mathrm{P}[X = 0, Y = 4] = 1/4.$$

Derive the marginal pmfs of X and Y. Are X and Y independent? Compute $\mathrm{Cov}[X, Y]$.

5) Conditional probability satisfies the axioms. Consider a probability space $(S, \mathcal{F}, \mathrm{P}[\cdot])$ and let $F \in \mathcal{F}$ be a fixed event with $\mathrm{P}[F] > 0$. For each $E \in \mathcal{F}$ define $\mu[E] = \mathrm{P}\left[E \,\middle|\, F\right]$. Show that $\mu[\cdot]$ satisfies the axioms of probability.

6) A variance paradox? For independent identically distributed (i.i.d.) RVs $X_1, \ldots, X_n$, each with distribution F and variance σ^2 we know that $\mathrm{var}[X_1 + \ldots + X_n] = n\sigma^2$. On the other hand, if $X \sim F$, then $\mathrm{var}[X + X] = \mathrm{var}[2X] = 4\sigma^2$. Is there a contradiction here? Explain.

7) Bernoulli, binomial, Poisson and normal distributions. In this exercise we deal with Bernoulli, binomial, Poisson and normal RVs. A Bernoulli RV X models experiments, such as a coin toss, where success happens with probability p and failure with probability $1 - p$. Success is indicated by $X = 1$ and failure by $X = 0$. Therefore, the pmf of X is

$$\mathrm{P}[X = 0] = 1 - p, \quad \mathrm{P}[X = 1] = p. \tag{1}$$

A binomial RV with parameters (n, p) counts the number of successes in n independent Bernoulli trials that succeed with probability p. Thus, we can write a binomial RV Y as

$$Y = \sum_{i=1}^{n} X_i \tag{2}$$

where the X_i are Bernoulli RVs with pmfs as in (1). The pmf of a binomial RV is easily derived by noting that we have $X = x$ for some integer x between 0 and n if and only there are x successful Bernoulli trials – something that happens with probability p^x – and $n - x$ failed experiments – which happens with probability $(1-p)^{n-x}$ – and that there are $\binom{n}{x}$ different ways in which this could happen. Thus

$$p(x) := \mathrm{P}\left[X = x\right] = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n. \tag{3}$$

A Poisson RV X takes values in the nonnegative integers. We say that X is Poisson with parameter λ it its pmf is

$$p(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, \ldots \tag{4}$$

Different from the other two, a normal RV X can take any real value (not just 0 or 1 like the Bernoulli or integers between 0 and n for the binomial or nonnegative integers for the Poisson). We say X is a continuous RV. Probabilities are described now using a pdf. For a normal RV with mean μ and variance σ^2, the pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad x \in \mathbb{R}. \tag{5}$$

Another concept of interest here is that of a cdf defined as the probability of X not exceeding x, i.e., $F(x) = \mathrm{P}\left[X \leq x\right]$. For nonnegative discrete RVs (Bernoulli, binomial and Poisson) we can write

$$F(x) = \sum_{u=0}^{x} p(u). \tag{6}$$

For the continuous normal cdf, the sum is replaced by an integral to write

$$F(x) = \int_{u=-\infty}^{x} f(u)\, du. \tag{7}$$

Unlikely as it may seem binomial, Poisson and normal RVs are in fact closely related. You will explore these connections in this exercise.

A) Binomial distribution. Prove that the expected value of a binomial RV X_n with parameters (n, p) is $\mathbb{E}\left[X_n\right] = np$ and that the variance is $\mathbb{E}\left[(X_n - \mathbb{E}\left[X_n\right])^2\right] = np(1-p)$. You are advised to start from (2). Fix the expected value at $\mathbb{E}\left[X_n\right] = np = 5$ and plot the pmf and cdf for $n = 6, 10, 20, 50$. Values of p have to be modified appropriately.

B) Binomial and Poisson distributions. Prove that the expected value of a Poisson RV X_P with parameter λ is $\mathbb{E}\left[X_P\right] = \lambda$. Plot the pmf of a Poisson distribution with parameter $\lambda = 5$. Notice that this pmf is quite similar to the binomial pmfs of Part A when n is large. In fact we can quantify this proximity by evaluating the following mean-squared error (MSE)

$$\Delta(X_n, X_P) = \sum_{x=0}^{\infty} \Big(\mathrm{P}\left[X_n = x\right] - \mathrm{P}\left[X_P = x\right]\Big)^2 \mathrm{P}\left[X_P = x\right]. \tag{8}$$

To evaluate the MSE in (8) numerically the infinite sum needs to be truncated. You can neglect probabilities smaller than 5×10^{-2}. Compute $\Delta(X_n, X_P)$ for $n = 6, 10, 20, 50$.

C) Binomial and Poisson distributions again. Having noticed this interesting fact, consider binomial RVs X_n with parameters n and $p = \lambda/n$. Prove that as $n \to \infty$ the pmf of X_n converges to the pmf of X_P.

D) Binomial and normal distributions. An important result in probability theory is the central limit theorem (CLT). The CLT concerns sum of i.i.d. RVs X_i with mean $\mathbb{E}[X_i] = \mu$ and variance $\text{var}[X_i] = \sigma^2$. Specifically, define

$$Z_n := \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}. \tag{9}$$

The CLT states that the pdf of Z_n is approximately standard normal for sufficiently large n. Formally

$$\lim_{n\to\infty} \mathrm{P}[Z_n \leq z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\, du. \tag{10}$$

Now, since a binomial RV is a sum of i.i.d Bernoulli RVs it is possible to approximate a binomial cdf with a normal cdf. You are asked to approximate a binomial cdf with a normal pdf for $p = 0.5$ and n = $10, 20, 50$. Show the equations you used for the approximations and corresponding plots.

E) Normal and Poisson approximations. In parts B and C you showed that for large n it is possible to approximate a binomial RV with a Poisson RV. In part D you showed that it is possible to approximate a binomial RV with a normal RV. Poisson and normal RVs are clearly different but this results cannot contradict each other because both are true. Please explain why these two approximations do not contradict each other. The answer is *not* that Poisson and normal are similar.