Homework 1: random variables

EE 325: Probability and Random Processes, Autumn 2019

Instructor: Animesh Kumar, EE, IIT Bombay

Instructions: Some of these questions will be asked in a quiz in the class on 19 Aug 2019 (Monday). *If you have queries, then meet the instructor or the TA during office hours.*

1. Let $X \sim \text{Uniform}[-2, 2]$ random variable and Y be obtained by clipping X. That is,

$$\begin{aligned} Y &= X, \text{ if } |X| \leq 1 \\ &= 1, \text{ if } X > 1 \\ &= -1, \text{ if } X < -1. \end{aligned}$$

What are the values of $\mathbb{P}(Y = 1)$, $\mathbb{P}(Y = -1)$, and $\mathbb{P}(Y = 0)$? Is Y continuous or discrete? Give reasons for your answer.

Solution: The rv $Y = 1$ if and only if $X \geq 1$. Therefore,

$$\begin{aligned} \mathbb{P}(Y = 1) = \mathbb{P}(X \geq 1) &= 1 - \mathbb{P}(X < 1), \\ &= 1 - \mathbb{P}(X \leq 1), \end{aligned}$$

since $\mathbb{P}(X = 1) = 0$ as X is a continuous rv. So,

$$\mathbb{P}(Y = 1) = 1 - F_X(1) = 1 - \frac{3}{4} = \frac{1}{4}.$$

Similarly, random variable Y takes value -1 if and only if $X \leq -1$. Therefore.

$$\begin{aligned} \mathbb{P}(Y = -1) &= \mathbb{P}(X \leq -1) \\ &= F_X(-1) = \frac{1}{4}. \end{aligned}$$

The random variable $Y = 0$ if and only if $X = 0$. So, $\mathbb{P}(Y = 0) = \mathbb{P}(X = 0) = 0$. The cdf of Y is plotted in Figure 1. From the cdf of Y, it can be concluded that Y is a mixed random variable (i.e., neither continuous nor discrete).

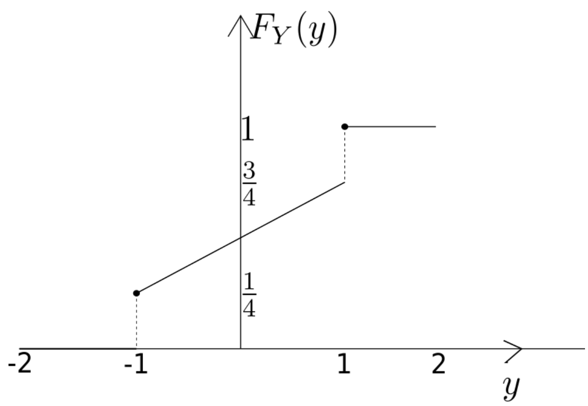


Figure 1: The cdf of Y is illustrated in this figure. Observe that the cdf has a non-zero derivatives in the interval $[-1, 1]$ and has discontinuities at $y = \pm 1$. So, Y is a mixed rv.

2. Using the cdf $F_X(x)$ of a random variable X, and the definition of a random variable, how will compute $\mathbb{P}(1 \le X \le 2)$, $\mathbb{P}(3 \le X < 4)$, and $\mathbb{P}(\{1 \le X \le 2\} \cup \{3 \le X \le 4\})$? Your answers should be explicit formulas, with reasoning, in terms of $F_X(x)$.

   Solution: The key ideas in this problem include the right continuous nature of cdf, and the addition of probabilities of disjoint events. For any $a, b \in \mathbb{R}$ such that $a < b$, we can write the interval $(-\infty, b]$ as the union of disjoint intervals $(-\infty, a)$ and $[a, b]$ i.e. $(-\infty, b] = (-\infty, a) \cup [a, b]$. Therefore,

$$\begin{aligned} &\mathbb{P}(-\infty < X \le b) = \mathbb{P}(-\infty < X < a) + \mathbb{P}(a \le X \le b), \\ &\text{or } \mathbb{P}(a \le X \le b) = \mathbb{P}(-\infty < X \le b) - \mathbb{P}(-\infty < X < a). \end{aligned} \tag{1}$$

   From (1), with $a = 1$ and $b = 2$, we obtain

$$\begin{aligned} \mathbb{P}(1 \le X \le 2) &= F_X(2) - F_X(1) + \mathbb{P}(X = 1), \\ &= F_X(2) - F_X(1^-). \end{aligned}$$

   Similarly, the interval $(-\infty, 4)$ is the union of disjoint intervals $(-\infty, 3)$ and $[3, 4)$. So,

$$\begin{aligned} \mathbb{P}(3 \le X < 4) &= \mathbb{P}(-\infty < X < 4) - \mathbb{P}(-\infty < X < 3), \\ &= F_X(4^-) - F_X(3^-). \end{aligned}$$

   Since, $[1, 2]$ and $[3, 4]$ are disjoint intervals, using the properties of rv we have

$$\begin{aligned} \mathbb{P}(\{1 \le X \le 2\} \cup \{3 \le X \le 4\}) &= \mathbb{P}(\{1 \le X \le 2\} + \mathbb{P}(\{3 \le X \le 4\} \\ &= F_X(2) - F_X(1) + \mathbb{P}(X = 1) + F_X(4) - F_X(3) + \mathbb{P}(X = 3), \\ &= F_X(2) - F_X(1^-) + F_X(4) - F_X(3^-). \end{aligned}$$

   The answer is now complete.

3. Let $F(x, y)$ be the joint cdf of two random variables (X, Y). Show that

$$F(2, 2) + F(1, 1) \ge F(2, 1) + F(1, 2).$$

   How can this inequality be generalized?

   Solution: Consider the figure shown below. The regions are defined as

$$\begin{aligned} A_1 &= \{(x, y) : x \le 1, y \le 1\} \\ A_2 &= \{(x, y) : x \le 1, 1 < y \le 2\} \\ A_3 &= \{(x, y) : 1 < x \le 2, y \le 1\} \\ A_4 &= \{(x, y) : 1 < x \le 2, 1 < y \le 2\}. \end{aligned}$$

   Further define $B_2 = A_1 \cup A_2$, $B_3 = A_1 \cup A_3$, and $B_4 = A_1 \cup A_2 \cup A_3 \cup A_4$. Observe that A_1, A_2, A_3, A_4 are disjoint. By the definition of random vectors, we get

$$\begin{aligned} \mathbb{P}((X, Y) \in B_2) &= \mathbb{P}((X, Y) \in A_1) + \mathbb{P}((X, Y) \in A_2) \\ \mathbb{P}((X, Y) \in B_3) &= \mathbb{P}((X, Y) \in A_1) + \mathbb{P}((X, Y) \in A_3) \\ \mathbb{P}((X, Y) \in B_4) &= \mathbb{P}((X, Y) \in A_1) + \mathbb{P}((X, Y) \in A_2) + \mathbb{P}((X, Y) \in A_3) + \mathbb{P}((X, Y) \in A_4). \end{aligned}$$

   Then,

$$\begin{aligned} \mathbb{P}((X, Y) \in A_4) &= \mathbb{P}((X, Y) \in B_4) + \mathbb{P}((X, Y) \in A_1) - \mathbb{P}((X, Y) \in B_2) - \mathbb{P}((X, Y) \in B_3) \\ & F(2, 2) + F(1, 1) - F(1, 2) - F(2, 1) \end{aligned}$$

   where the last step follows by the definition of the cdf. Since probability is positive, i.e., $\mathbb{P}((X, Y) \in A_4) \ge 0$, therefore the inequality follows. A general version of this inequality is that for any pair of points (x_1, y_1) and (x_2, y_2) such that $x_1 < x_2, y_1 < y_2$, it can be stated that

$$F(x_1, y_1) + F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) \ge 0.$$

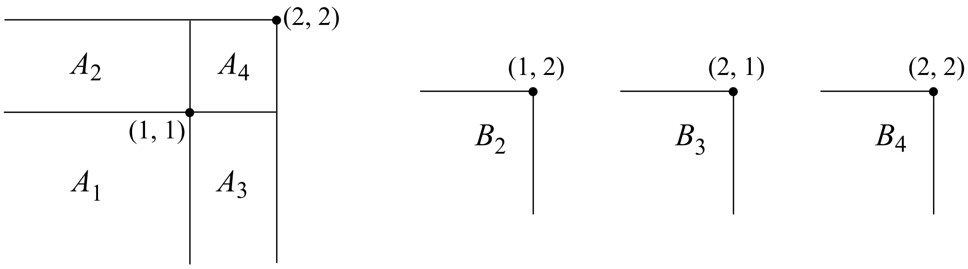


Figure 2: Qn 3.

4. Sketch the cdf of the following random variables:

   (a) A Poisson random variable with the parameter $\lambda = 2$.

   (b) A Cauchy random variable with the pdf as follows:

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \quad x \in \mathbb{R}.$$

Solution:

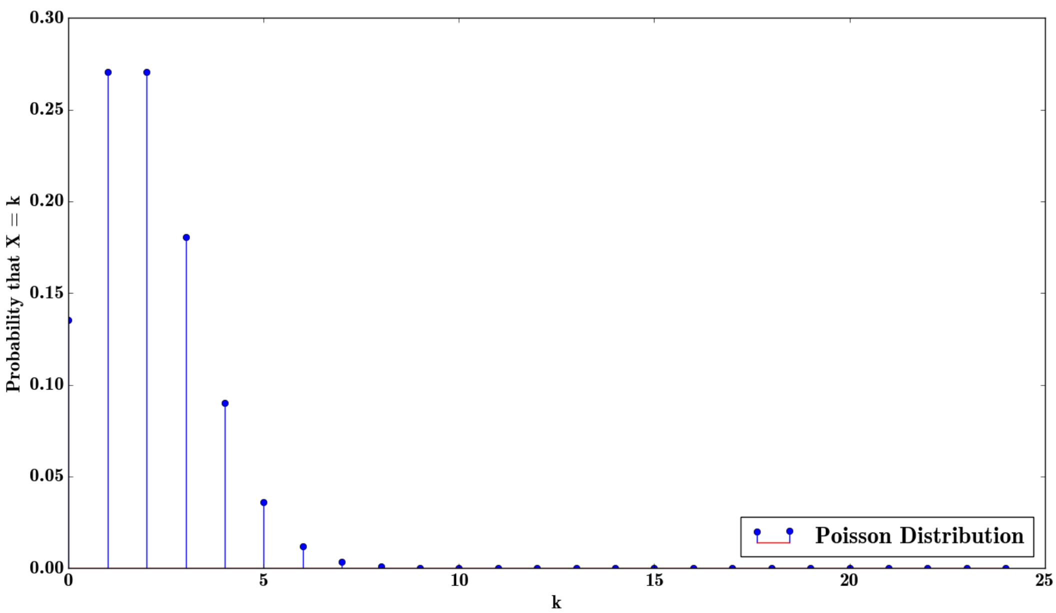
(a) Poisson distribution is defined over whole numbers ($k \geq 0$) and is given by:

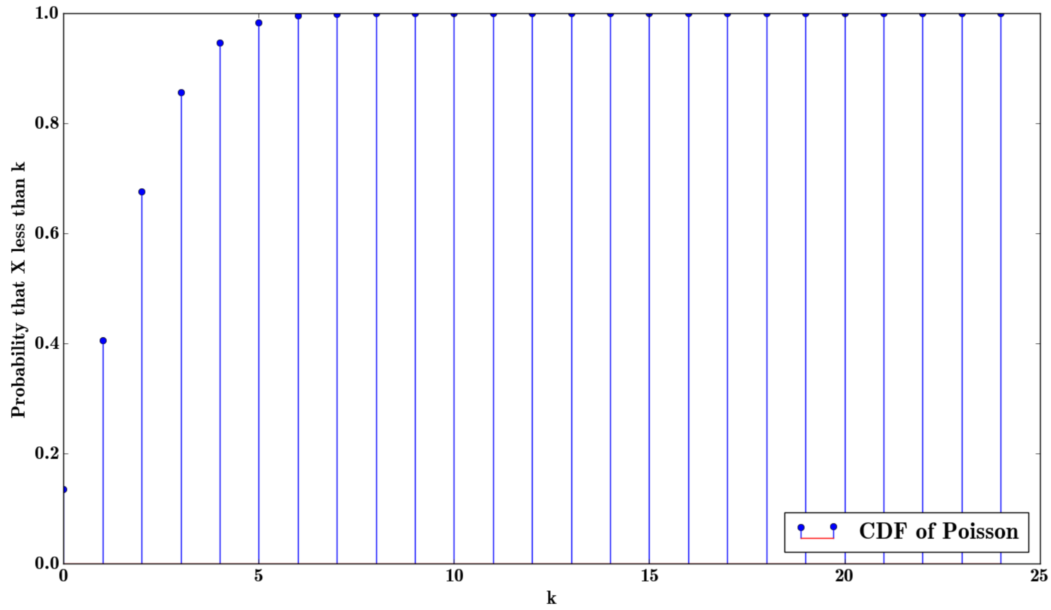
$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

Thus CDF ($P(X \leq k)$) is:

$$F(k) = P(X \leq k) = \sum_{i=0}^{k} \frac{e^{-\lambda}\lambda^i}{i!}$$

Probablity Mass Function (PMF) and CDF of Poisson Distribution for $\lambda = 2$:





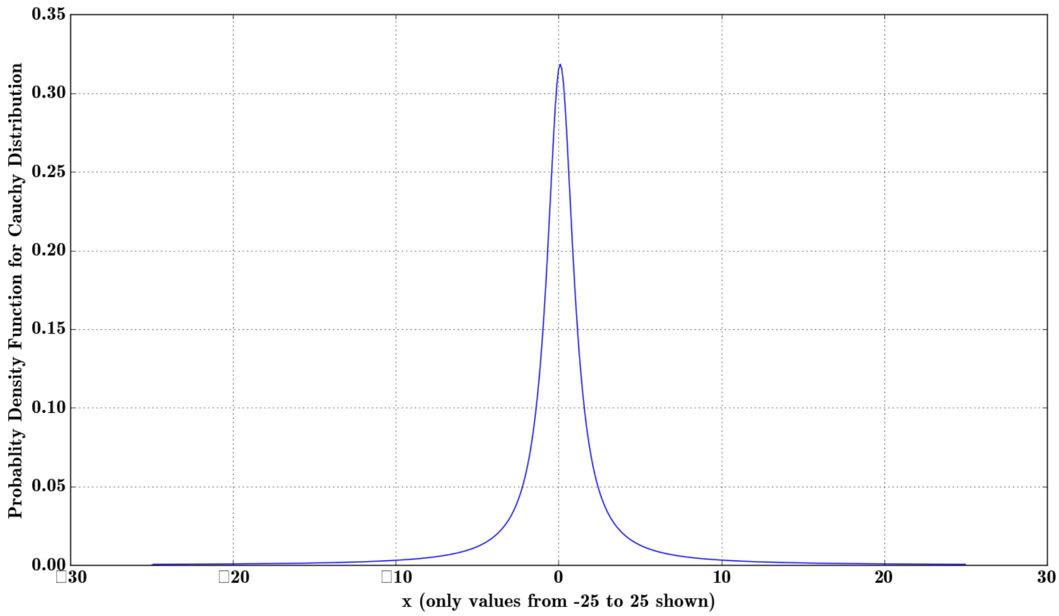
(b) Cauchy Distribution:

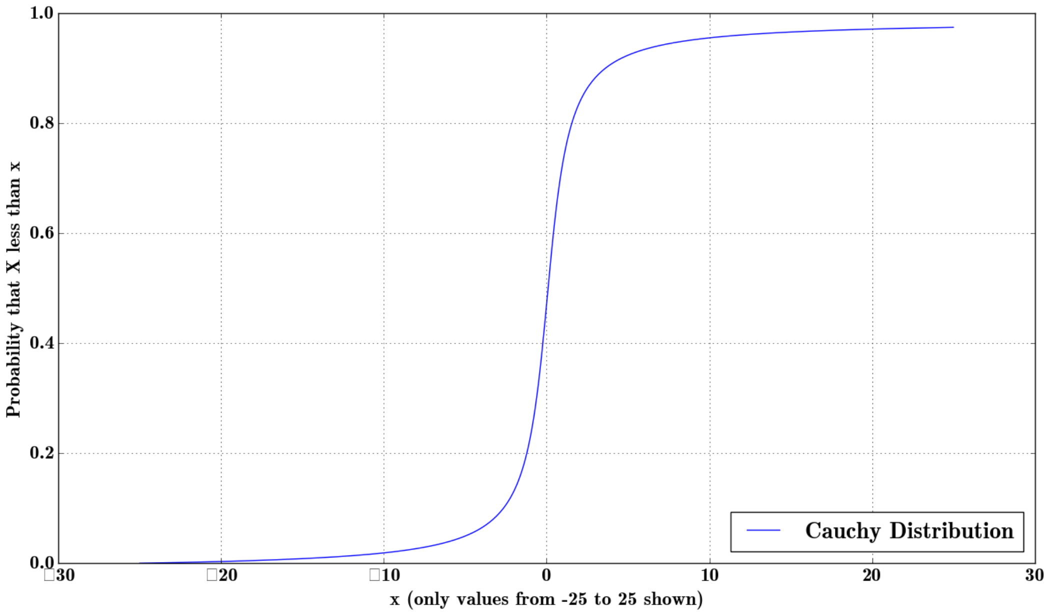
$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad x \in \mathbb{R}.$$

We can integrate from $-\infty$ to x and get the CDF. We get:

$$F_X(x) = \frac{1}{\pi}(\tan^{-1}(x) + \frac{\pi}{2}), \quad x \in \mathbb{R}.$$

PDF and CDF of Cauchy Distribution:





5. Let (X, Y, Z) be independent random variables. Show that any two subset of random variables, for example (X, Y), are also independent. How will your result generalize to more than three random variables?

Solution: Recall that two random variables X, Y are independent if

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

Equivalently, X, Y are independent iff for any x, y, the events $X \leq x, Y \leq y$ are independent events. Extending to a set of n random variables, we have:

$$F_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$

Given that the set $\{X_1, X_2, \dots, X_n\}$ random variables are independent, consider any $\mathcal{I} \subseteq \{1, 2, \dots, n\}$, let $\mathcal{I}^c = \{1, 2, \dots, n\} - \mathcal{I}$. The set of random variables $\{X_j, j \in \mathcal{I}\}$ can easily be proved to be independent using the above definition of independence taking values of $x_i, i \in \mathcal{I}$ to be any set of real numbers but choosing $x_k = \infty$, $\forall k \in \mathcal{I}^c$.

6. Let k and n be non-negative integers, and $0 < p < 1$. A random variable X has a geometric distribution if its pmf is given by $p_X(k) = (1-p)p^k$. Define the residual lifetime distribution function as, $l_X(k, n) := \mathbb{P}(X \geq n + k | X \geq n)$.

   (a) Show that $l_X(k, n) = \mathbb{P}(X \geq k)$ independent of n, i.e., the geometric distribution satisfies the memoryless property.

   (b) Assume that $Y \geq 0$ is any other discrete integer-valued distribution which exhibits memoryless property, i.e., $l_Y(k, n) = \mathbb{P}(Y \geq k)$. Show that $l_Y(k, n)$ has to be of the form α^k for some $0 < \alpha < 1$.

   (c) Using (b), show that if Y satisfies the memoryless property, then it has a geometric distribution.

Solution:

(a) For $X \sim \text{Geometric}(p)$, it is easy to see that $\mathbb{P}(X \geq n) = \sum_{k \geq n}(1-p)p^k = p^n$. By the definition of residual lifetime,

$$\begin{aligned} l_X(k, n) = \mathbb{P}(X \geq n + k | X \geq n) &= \frac{\mathbb{P}((X \geq n + k), (X \geq n))}{\mathbb{P}(X \geq n)} \\ &= \frac{\mathbb{P}(X \geq n + k)}{\mathbb{P}(X \geq n)} \\ &= \frac{p^{n+k}}{p^n} = p^k, \end{aligned}$$

which is independent of n. Thus the Geometric(p) random variable exhibits memoryless property.

(b) By the memoryless property,

$$\mathbb{P}(Y \geq n+k) = \mathbb{P}(Y \geq n)\mathbb{P}(Y \geq k). \tag{2}$$

Put $k = 1$ in the above equation to get $\mathbb{P}(Y \geq n+1) = \mathbb{P}(Y \geq n)\mathbb{P}(Y \geq 1)$. By a recursive argument we can obtain

$$\mathbb{P}(Y \geq n+1) = [\mathbb{P}(Y \geq 1)]^{n+1}, \forall n \geq 0. \tag{3}$$

Denote $\mathbb{P}(Y \geq 1) = \alpha$. Then $l_Y(k,n) = \mathbb{P}(Y \geq k) = \alpha^k$ for some $0 \leq \alpha \leq 1$.

(c) Since Y is a discrete random variable, therefore $\mathbb{P}(Y = n) = \mathbb{P}(Y \geq n) - \mathbb{P}(Y \geq n+1) = \alpha^n(1-\alpha)$. Thus, Y is a geometric random variable. The parameter α can lie in $[0,1]$. The extreme cases i.e., $\alpha = 0$ or $\alpha = 1$ correspond to the trivial cases.