Statistical Foundation of Data Science (MA 589)

Lecture Slides

Topic 02: Random Variable

Random Variables

Definition 2.1: A function $X : \mathcal{S} \to \mathbb{R}$ is called a random variable (RV).

Remark 2.1: The above definition is not complete. However, in this course we will use it as definition of RV.

Example 2.1: Tossing a fair coin n times. Assume that the tosses are independent. Let $X : \mathcal{S} \to \mathbb{R}$ be defined by the number of heads. Take $n = 2$. $P(X = 0) = P(X = 2) = \frac{1}{4}$, $P(X = 1) = \frac{1}{2}$.

Example 2.2: Throwing a fair die twice. Assume the throws are independent. Let $X : \mathcal{S} \to \mathbb{R}$ be defined by the sum of the outcomes. $P(X = 2) = 1/36, P(X = 3) = 2/36, P(X = 4) = 3/36, P(X = 5) = 4/36, P(X = 6) = 5/36, P(X = 7) = 6/36,$
$P(X = 8) = 5/36, P(X = 9) = 4/36, P(X = 10) = 3/36,$
$P(X = 11) = 2/36, P(X = 12) = 1/36.$

Cumulative Distribution Function

Definition 2.2: The cumulative distribution function (CDF) of a random variable X is a function $F_X : \mathbb{R} \to [0, \infty)$ defined by

$$F_X(x) = P(X \leq x)\,.$$

Example 2.3: Toss a coin twice. Let $X =$ number of heads. Then

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < 0, \\ 1/4 & \text{if} \quad 0 \leq x < 1, \\ 3/4 & \text{if} \quad 1 \leq x < 2, \\ 1 & \text{if} \quad x \geq 2\,. \end{cases}$$

Example 2.4: Throwing a fair die twice. Assume the throws are independent. Let $X =$ the sum of the outcomes. Then

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < 2, \\ 1/36 & \text{if} \quad 2 \le x < 3, \\ 3/36 & \text{if} \quad 3 \le x < 4, \\ 6/36 & \text{if} \quad 4 \le x < 5, \\ 10/36 & \text{if} \quad 5 \le x < 6, \\ 15/36 & \text{if} \quad 6 \le x < 7, \\ 21/36 & \text{if} \quad 7 \le x < 8, \\ 26/36 & \text{if} \quad 8 \le x < 9, \\ 30/36 & \text{if} \quad 9 \le x < 10, \\ 33/36 & \text{if} \quad 10 \le x < 11, \\ 35/36 & \text{if} \quad 11 \le x < 12, \\ 1 & \text{if} \quad x \ge 12\,. \end{cases}$$

Proposition: The CDF of a random variable has the following properties:

(1) $F_X(\cdot)$ is non-decreasing.

(2) $\lim_{x \uparrow \infty} F_X(x) = 1$, $\lim_{x \downarrow -\infty} F_X(x) = 0$.

(3) $\lim_{h \downarrow 0} F_X(x+h) = F_X(x), \forall x \in \mathbb{R}$, thus CDF is right continuous.

(4) $\lim_{h \downarrow 0} F_X(x-h) = F_X(x) - P(X = x), \forall x \in \mathbb{R}$.

Theorem 2.1: Let F be a function satisfying properties (1)-(3). Then F is a CDF.

Remarks

- Random variable is just a function and does not depend on the probability. But the distribution of the random variable depends on the probability. So keeping the function same if we change the probability then the random variable will remain the same but its distribution will change. Consider Example 1, but with the probabilities
  $P(HH) = 9/16, P(TT) = 1/16, P(HT) = P(TH) = 3/16.$
  What will be the distribution function in this case?
- If $x \in \mathbb{R}$ is such that $P(X = x) > 0$, then x is said to be an atom of the distribution function of X. Thus if the distribution function of a random variable has no atoms then it is continuous.
- $P(a < X \leq b) = F_X(b) - F_X(a)$.
- $P(a \leq X \leq b) = F_X(b) - F_X(a-)$.
- $P(a < X < b) = F_X(b-) - F_X(a)$.
- $P(a \leq X < b) = F_X(b-) - F_X(a-)$.

Discrete Random Variable

Definition 2.3: [Discrete RV] A random variable is said to have discrete distribution if there exists an at most countable set $S_X \subset \mathbb{R}$ such that $P(X = x) > 0$ for all $x \in S_X$ and $\sum_{x \in S_X} P(X = x) = 1$. S_X is called the support of X.

Definition 2.4: Define a function $f_X : \mathbb{R} \to [0, 1]$ by

$$f_X(x) = \begin{cases} P(X = x) & \text{if } x \in S_X \\ 0 & \text{otherwise}. \end{cases}$$

The function f_X is called the probability mass function of X.

Example 2.5: RV X in Examples 2.1 and 2.2 are discrete.

Remarks

- For a discrete random variable X,

$$F_X(x) = \sum_{\substack{y \in S_X \\ y \leq x}} f_X(y).$$

- For a discrete random variable X,

$$f_X(x) = F_X(x) - F_X(x-).$$

Properties of PMF

1. $f_X(x) \geq 0$ for all $x \in \mathbb{R}$.
2. $\sum_{x \in S_X} f_X(x) = 1$.

Theorem 2.2: Suppose a real valued function $h : \mathbb{R} \to \mathbb{R}$ satisfies the following conditions:

1. $h(x) \geq 0$ for all $x \in \mathbb{R}$. $D = \{x : h(x) > 0\}$ is atmost countable.
2. $\sum_{x \in D} h(x) = 1$.

Then $h(\cdot)$ is a probability mass function of some discrete random variable.

Example

1. (Bernoulli Distribution) $X \sim Bernoulli(p)$: $S_X = \{0, 1\}$, $f_X(0) = 1 - p$, $f_X(1) = p$.
2. (Binomial Distribution) $X \sim Bin(n\,, p)$: $S_X = \{0,\ 1,\ \ldots,\ n\}$, $f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$.
3. (Geometric Distribution) $X \sim Geo(p)$: $S_X = \{0,\ 1,\ \ldots\}$, $f_X(k) = p(1-p)^k$.
4. (Poisson Distribution) $X \sim Poi(\lambda)$ $(\lambda > 0)$: $S_X = \{0,\ 1,\ \ldots\}$, $f_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}$.

Application

Example 2.6: Suppose that an airplane engine will fail, when in flight, with probability $1 - p$ independently from engine to engine. The airplane will make a successful flight if at least 50 percent of its engines remain operating. For what values of p is a four engine plane preferable to a two engine plane? (Ans: $p > 2/3$)

Continuous Random Variable

Definition 2.5: A random variable is said to have a continuous distribution if there exists a non-negative integrable function $f_X : \mathbb{R} \to [0, \infty)$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

for all $x \in \mathbb{R}$. The function f_X is called the probability density function. The set $S_X = \{x \in \mathbb{R} : f_X(x) > 0\}$ is called support of X.

Example 2.7: (Exponential Distribution: $Exp(\lambda)$)

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Example 2.8: (Uniform Distribution: $U(a, b)$)

$$F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x < b \\ 1 & \text{if } x \geq b. \end{cases}$$

Example 2.9: (Normal Distribution: $N(\mu, \sigma^2)$)

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad \text{if } -\infty < x < \infty.$$

Remarks on CRV

- For a continuous random variable X, $P(X = a) = 0$ for all $a \in \mathbb{R}$.
- CDF of a continuous random variable is continuous.
- PDF is not unique.
- Support of a continuous random variable is not unique.
- $P(a \leq X \leq b) = \displaystyle\int_a^b f_X(t)dt$.
- $f_X(x)$ is not $P(X = x)$.

Properties of PDF

1. $f_X(x) \geq 0$ for all $x \in \mathbb{R}$.
2. $\displaystyle\int_{-\infty}^{\infty} f_X(x) = 1.$

Theorem 2.3: Suppose a real valued function $g : \mathbb{R} \to \mathbb{R}$ satisfies the following conditions:

1. $g(x) \geq 0$ for all $x \in \mathbb{R}$.
2. $\displaystyle\int_{-\infty}^{\infty} g(x)dx = 1.$

Then $g(\cdot)$ is a probability density function of some continuous random variable.

RV which is neither discrete nor continuous

Consider the random variable X whose distribution function is given by

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < -1 \\ x+1 & \text{if} \quad -1 \le x < -\frac{1}{2} \\ 1 & \text{if} \quad x \ge -\frac{1}{2}. \end{cases}$$

Expectation of DRV

Definition 2.6: Let X be a discrete RV with PMF $f_X(\cdot)$ and support S_X. The expectation or mean of X is defined by

$$E(X) = \sum_{x \in S_X} x f_X(x) \quad \text{provided} \quad \sum_{x \in S_X} |x| f_X(x) < \infty \,.$$

- E(X) is the weighted average of the values taken by X.
- If $\sum_{x \in S_X} |x| f_X(x) = \infty$ then we say that expectation does not exist.

Example 2.10: X = outcome of a roll of a fair die. What is $E(X)$?
Example 2.11: $X \sim Bin(n, p)$. What is $E(X)$?
Example 2.12: $X \sim Geo(p)$. What is $E(X)$?
Example 2.13: $X \sim Poi(\lambda)$. What is $E(X)$?
Example 2.14:

$$f_X(x) = \begin{cases} \frac{c}{x^2}, & x \in \mathbb{N}, \quad \text{where} \quad c = \left(\sum_{n=1}^{\infty} \frac{1}{n^2}\right)^{-1} \\ 0 & \text{otherwise}. \end{cases}$$

Let X be a DRV having the above PMF, then $E(X)$ does not exist.

Expectation of CRV

Definition 2.7: Let X be a CRV with PDF $f_X(.)$. The expectation of X is defined by

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx \quad \text{provided} \quad \int_{-\infty}^{\infty} |x| f_X(x) dx < \infty \,.$$

Example 2.15: $X \sim U(a, b)$, what is $E(X)$?
Example 2.16: $X \sim Exp(\lambda)$, what is $E(X)$?
Example 2.17: $X \sim N(\mu, \sigma^2)$, what is $E(X)$?
Example 2.18: Let X be a CRV having PDF $f_X(x) = \frac{1}{\pi(1+x^2)}, \forall x \in \mathbb{R}$. What is $E(X)$?

Transformation of RV

1. If X is a random variable then $Y = g(X)$ is a random variable where $g : \mathbb{R} \to \mathbb{R}$.
2. Our aim is to find the distribution (CDF/PMF/PDF) of $Y = g(X)$ for a known distribution of X.
3. There are mainly three techniques.

Technique 1

- Find the CDF of $Y = g(X)$ from the distribution of X.

Example 2.19: Let the random variable X has the following PDF:

$$f(x) = \begin{cases} e^{-x} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Find the distribution of $Y = [X]$.

Example 2.20: Let the random variable X has the following PDF:

$$f(x) = \begin{cases} \frac{|x|}{2} & \text{if } -1 < x < 1 \\ \frac{x}{3} & \text{if } 1 \leq x < 2 \\ 0 & \text{otherwise.} \end{cases}$$

Find the distribution of $Y = X^2$.

Technique 2 for DRV

Example 2.21: Let the random variable X has the following PMF:

$$f(x) = \begin{cases} \frac{1}{7} & \text{if } x = -2,\ -1,\ 0,\ 1 \\ \frac{3}{14} & \text{if } x = 2,\ 3 \\ 0 & \text{otherwise.} \end{cases}$$

Find the distribution of $Y = X^2$.

Technique 2 for DRV (Contd.)

Theorem 2.4: Let X be a DRV with PMF $f_X(\cdot)$ and support S_X. Let $g : \mathbb{R} \to \mathbb{R}$ and $Y = g(X)$. Then Y is a DRV with support $S_Y = \{g(x) : x \in S_X\}$ and PMF

$$f_Y(y) = \begin{cases} \sum_{x \in A_y} f_X(x) & \text{if } y \in S_Y \\ 0 & \text{otherwise,} \end{cases}$$

where $A_y = \{x \in S_X : g(x) = y\}$.

Example 2.22: $X \sim Bin(n, p)$. Find the distribution of $Y = n - X$.

Technique 2 for CRV

Theorem 2.5: Let X be a CRV with PDF $f_X(\cdot)$ and support S_X, which is an interval. Let $g : S_X \to \mathbb{R}$ be a differentiable function and either $g'(x) < 0$ for all $x \in S_X$ or $g'(x) > 0$ for all $x \in S_X$. Then the RV $Y = g(X)$ is a CRV with PDF

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & \text{for } y \in g(S_X) \\ 0 & \text{otherwise.} \end{cases}$$

Example 2.23: Let $X \sim U(0, 1)$, then $Y = -\ln X \sim Exp(1)$.

Example 2.24: Let $X \sim Exp(1)$, then find the distribution of $Y = X^2$.

Example 2.25: Let $X \sim N(0, 1)$, then find the distribution of $Y = X^2$.

Expectation of Function of RV

Example 2.26: Let the random variable X be a DRV with PMF

$$f_X(x) = \begin{cases} \frac{1}{7} & \text{if } x = -2,\ -1,\ 0,\ 1 \\ \frac{3}{14} & \text{if } x = 2,\ 3 \\ 0 & \text{otherwise.} \end{cases}$$

Let $Y = X^2$. Find the expectation of Y.

Expectation of Function of RV

Theorem 2.6: Let X be a DRV with PMF $f_X(\cdot)$ and support S_X. Let $g : \mathbb{R} \to \mathbb{R}$. Then

$$E\left[g(X)\right] = \sum_{x \in S_X} g(x) f_X(x) \quad \text{provided} \quad \sum_{x \in S_X} |g(x)| f_X(x) < \infty.$$

Theorem 2.7: Let X be a CRV with PDF $f_X(\cdot)$. Let $g : \mathbb{R} \to \mathbb{R}$. Then

$$E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \quad \text{provided} \quad \int_{-\infty}^{\infty} |g(x)| f_X(x) dx < \infty.$$

Expectation of Function of RV

Theorem 2.8: Let X be a RV (either DRV or CRV). Then

1. Let $A \subset \mathbb{R}$. Then $E(I_A(X)) = P(X \in A)$.
2. $h_1(x) \leq h_2(x)$, for all $x \in \mathbb{R}$, then $E\left[h_1(X)\right] \leq E\left[h_2(X)\right]$, provided all the expectations exist.
3. $a < b$ are two real numbers such that $S_X \subset [a, b]$, then $a \leq E(X) \leq b$, provided the expectation exists.
4. $E(a + bX) = a + bE(X)$, where a and b are two real numbers.
5. Let $h_1(\cdot), \ldots, h_p(\cdot)$ be real valued function of real numbers such that $E(h_i(X))$ exists for all $i = 1, 2, \ldots, p$, then

$$E\left(\sum_{i=1}^{p} h_i(X)\right) = \sum_{i=1}^{p} E\left(h_i(X)\right).$$

Some Special Expectations

- For $r = 1, 2, \ldots, \mu_r = E(X^r)$ is called rth raw moment of X, if the expectation exists.
- $\mu'_r = E\left[(X - E(X))^r\right]$ is called rth central moment of X, if the expectations exist.
- $\mu'_2 = E\left[(X - E(X))^2\right]$ is called variance of X when it exists and is denoted by $Var(X)$.
- $Var(X) = E(X^2) - (E(X))^2$.

Moment Generating Function

Definition 2.8: The moment generating function of random variable X is defined by

$$M_X(t) = E\left(e^{tX}\right)$$

provided the expectation exists in a neighbourhood of the origin.

Example 2.27: $X \sim Bin(n, p)$, then $M_X(t) = (1 - p + pe^t)^n$ for all $t \in \mathbb{R}$.

Example 2.28: $X \sim Exp(\lambda)$, then $M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}$ for all $t < \lambda$.

Example 2.29: $X \sim N(\mu, \sigma^2)$, then $M_X(t) = e^{\mu t + \frac{t^2\sigma^2}{2}}$ for all $t \in \mathbb{R}$.

Remark

If the MGF $M_X(t)$ exist for $t \in (-a, a)$ for some $a > 0$, the derivatives of all order exist at $t = 0$ and

$$E\left(X^k\right) = \frac{d^k}{dt^k} M_X(t)\bigg|_{t=0}$$

for all positive integer k.

Technique 3

Definition 2.9: X and Y are said to be same in distribution (denoted by $X \stackrel{d}{=} Y$) if $F_X(x) = F_Y(x)$ for all $x \in \mathbb{R}$.

Theorem 2.9: Let X and Y be two random variables having MGFs $M_X(\cdot)$ and $M_Y(\cdot)$, respectively. Suppose that there exists a positive real number a such that $M_X(t) = M_Y(t)$ for all $t \in (-a, a)$. Then X and Y are same in distribution.

Example 2.30: Let $X \sim N(\mu, \sigma^2)$. Find the distribution of $Y = a + bX$.

Example 2.31: Let $X \sim N(0, \sigma^2)$. Find the distribution of $Y = X^2$.