# 3.1 Random Variables: Definition and CDF.

**Remark.** Intuitively, one can think random variables as random quantity.

For example, number of heads when flipping coin 5 times.

Or, a lifespan of a lightbulb until it dies.

**Def.** Given $(\Omega, \mathcal{A}, P)$, a random variable is a function

$X : \Omega \to \mathbb{R}$ such that $\underbrace{\{\omega \in \Omega \mid X(\omega) \le x\} \in \mathcal{A} \quad \forall x \in \mathbb{R}}$.

**Remark.** i) $X$ is a "measurable function" if it satisfies —.

ii) Usually, we use uppercase letters, e.g. $X, Y, \cdots$, for random variables, and lower case letters, e.g. $x, y, \cdots$, for the corresponding values.

iii) Usually, we abbreviate $\{\omega \in \Omega \mid X(\omega) \le x\}$ as $\{X \le x\}$.

It applies for other predicates too.

iv) Usually, we abbreviate $P(\{X \le x\})$ as $P(X \le x)$

**Def.** The CDF (Cumulative Distribution Function) of a r.v. (random variable) is the function $F : \mathbb{R} \to [0,1]$ such that $F(x) = P(X \le x)$.

pf) i) Monotonicity. (If $x \le y$, then $F(x) \le F(y)$)

Suppose that $x \le y$. Then, $\{X \le x\} \subset \{X \le y\}$.

Therefore $P(X \le x) = F(x) \le P(X \le y) = F(y)$

ii) Limiting values. $\left(\lim_{x \to -\infty} F(x) = 0\right)$.

Since $F(x)$ is monotonic and bounded below by zero, it converges as $x \to -\infty$,

So the limit exists for every sequence $\{x_n\}$ converging to $-\infty$.

Let $x_n = -n$. $\lim_{x \to -\infty} F(x) \overset{i)}{=} \lim_{n \to \infty} F(x_n) = \lim_{n \to \infty} F(-n) = \lim_{n \to \infty} P(X \le -n) \overset{ii)}{=} P(\phi) = 0$.

Note. i) $\{X \le x\} = \bigcap_{n=1}^{\infty} \{X \le x_n\} = \lim_{n\to\infty} \{X \le x_n\}$ by continuity from above.

ii) $\{X \le -\infty\} = \{w \in \Omega \mid X(w) \le -\infty\}$.

because if we assume there is some $x \in \mathbb{R}$ such that $x \le -\infty$, it violates the definition of $\infty$. Therefore $\{X \le -\infty\} = \phi$.

iii) Limiting values $\left( \lim_{x\to\infty} F(x) = 1 \right)$

Since $F(x)$ is monotonic and bounded above by 1, it converges as $x \to \infty$.

So the limit exists for every sequence $\{x_n\}$ converging to $\infty$.

Let $x_n = n$. $\lim_{x\to\infty} F(x) = \lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} F(n) = \lim_{n\to\infty} P(X \ge n) = P(\Omega) = 1.$

iv) Right continuity. $\left( \lim_{x\to a^+} F(x) = F(a) \right)$

Consider a decreasing sequence $\{x_n\}$ that converges to $a$.

The sequence of events $\{X \le x_n\}$ is decreasing,

so $\bigcap_{n=1}^{\infty} \{X \le x_n\} = \lim_{n\to\infty} \{X \le x_n\} = \{X \le x\}$ by continuity from above.

$\lim_{x\to a^+} F(x) = \lim_{x\to a^+} P(X \le x) = \lim_{n\to\infty} P(\{X \le x_n\}) = P(X \le a) = F(a).$

This applies for every such sequence $\{x_n\}$, QED.

Def. The distribution of a r.v. $X$, also called the probability law of $X$, is a probability measure $P^X$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $P^X(A) = P(X \in A)$ $\forall A \in \mathcal{B}(\mathbb{R})$.

Note. Original measure $P$ is a measure on $(\Omega, \mathcal{A})$.

The distribution of $X$ $P^X$ is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

In many instances, $P$ remains hidden or unused, and one works directly with the more tangible probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P^X)$. ex) Statistical properties of $X$.

pf) i) $P^X(A) = P(X \in A)$ $\forall A \in \mathcal{B}(\mathbb{R})$ is a function $P^X : \mathcal{B}(\mathbb{R}) \to [0,1]$.

ii) $P^X(\phi) = P(X \in \phi) = P(\phi) = 0$.

iii) $P^X(\Omega) = P(X \in \Omega) = P(\Omega) = 1$.

iv) For countable additivity, let $\{B_i\}$ be a countable disjoint subsets of $\mathcal{B}(\mathbb{R})$.

Sets $\{X \in B_i\}$ is also disjoint, as $\{X \in B_i\} = \{\omega \in \Omega \mid X(\omega) \in B_i\}$.

So, $\{X \in \bigcup_{i=1}^{\infty} B_i\} = \bigcup_{i=1}^{\infty} \{X \in B_i\}$.

$P^X(\bigcup_{i=1}^{\infty} B_i) = P(X \in \bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(X \in B_i) = \sum_{i=1}^{\infty} P^X(B_i)$. ▨

Def. A function $F : \mathbb{R} \to [0,1]$ is a distribution function if it satisfies three properties of a CDF, namely, monoticity, limiting values, and right-continuity.

Thm. Let $F$ be a distribution function, and consider a probability space $([0,1], \mathcal{B}([0,1]), P)$ such that $P$ is the lebesgue measure.

There exists a measurable function $X : \Omega \to \mathbb{R}$ whose CDF $F_X$ satisfies $F_X = F$.

Prop. Distribution of r.v. $X$, $P^X$ is the probability measure induced by CDF $F$.

pf) (Outline) $P((-\infty, x]) = F(x) = P(X \leq x) = P(X \in (-\infty, x]) = P^X((-\infty, x])$ ▨

Equivalence between CDF and Borel probability measure

## 3.2 Types of Random Variables.

**Def.** A random variable $X$ is discrete if $\{X(\omega) \mid \omega \in \Omega\}$ is countable, i.e., $\{X(\omega) \mid \omega \in \Omega\} = \{x_1, x_2, \cdots\}$.

**Def.** A random variable $X$ has a density $f$ if the CDF of $X$, $F(x) = \int_{-\infty}^{x} f(u)\,du \;\; \forall x \in \mathbb{R}$ for some integrable $f : \mathbb{R} \to [0, \infty)$

Lebesgue integral.
But if Riemann integral exists, Lebesgue integral always exists, so it's safe to think that it is a Riemann integral for most common cases.

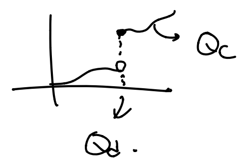
**Motivation** (from a measure theory - perspective) Decompositions of $Q$.

Let $Q = P^X$, $J = \{x \in \mathbb{R} \mid Q(x) > 0\}$.

$Q_d(A) = Q(A \cap J)$, $Q_c(A) = Q(A) - Q(A \cap J)$.

$Q = Q_d + Q_c$. Continuous part of the measure.

Discrete part of the measure




$Q_{ac}((-\infty, x]) = \int_{-\infty}^{\infty} f(u)\,du$.

$Q_c = Q_{ac} + Q_{sc}$ (by Raydon-Nikodym thm.)

Singular Continuous part of the measure

Absolutely Continuous of the measure.

**Claim** $X$ is discrete $\Rightarrow Q = Q_d$.

$X$ has density $\Rightarrow Q = Q_{ac}$

**Remark:** $Q = Q_c \not\Rightarrow X$ has a density. (Because of $Q_{sc}$).

**Example** If one take Cantor function as a CDF, there is no $Q_d$, $Q_{ac}$ part, only $Q_{sc}$.

## 3.3 Discrete Random Variables

$\{X(\omega) \mid \omega \in \Omega\}$ is countable.

**Def.** The probability mass function (PMF) of a discrete r.v. $X$ is the function $p: \mathbb{R} \to [0,1]$ such that $p(x) = P(X=x)$

**Remark.** $P(X \in A) = \sum_{x \in A \cap S} p(x)$ where $S = \{X(\omega) \mid \omega \in \Omega\}$.

pf) $P(X \in A) = P^X(A) = P^X(A \cap S) + \cancel{P^X(A \cap S^c)}^{\,0} = \sum_{x \in A \cap S} P(X=x) = \sum_{x \in A \cap S} p(x)$.

**Remark.** $P(X \in \mathbb{R}) = \sum_{x \in S} p(x) = 1$.

**Notation.** For a discrete r.v. $X$, $X \sim p$ means $X$ is distributed according to pmf $p$. with an underlying probability measure space $(\Omega, \mathcal{A}, P)$.

Also, $X \sim F$ means $X$ has a cdf $F$.

Also, $X \sim Q$ means $X$ has a distribution $Q$

**Examples** i) $X \sim \text{Bernoulli}(\underset{\text{Parameters}}{\alpha})$, $\alpha \in [0,1]$. ⇝ One Coin flip

$p(1) = \alpha$, $p(0) = 1-\alpha$.

ii) $X \sim \text{Binomial}(n, \alpha)$, $n \in \mathbb{N}$, $\alpha \in [0,1]$. ⇝ $n$ coin flip

$p(k) = \binom{n}{k} \alpha^k (1-\alpha)^{n-k}$ where $k \in \{0, 1, \cdots, n\}$

iii) $X \sim \text{Geometric}(\alpha)$, $\alpha \in [0,1]$. ⇝ Coin flip until heads

$p(k) = (1-\alpha)^{k-1} \alpha$ where $k \in \{0, 1, \cdots\}$

iv) $X \sim \text{Poisson}(\lambda)$, $\lambda \in \mathbb{R}$, $\lambda \geq 0$. ⇝ Number of customers arriving if "rate" is constant. i.e. $\lambda$ persons per hour.

$p(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ where $k \in \{0, 1, \cdots\}$

## 3.4) Random Variables with Densities

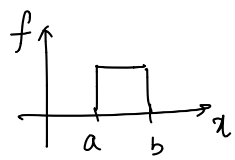
**Notation.** We call $f$ the probability density function (PDF) of a r.v. $X$.
We write it as $X \sim f$.

Usually, $P$ : Probability measures, $p$ : pmfs, $f$ : pdfs.
But it can change over context, so be aware always.

Indicator function of $A$ $I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$

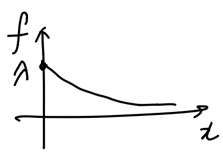
**Examples** i) $X \sim \text{Uniform}(a, b)$, $a < b$, $a, b \in \mathbb{R}$.

$$f(x) = \begin{cases} 1/b-a, & x \in [a, b] \\ 0, & \text{otherwise} \end{cases}$$




ii) $X \sim \text{Exponential}(\lambda)$, $\lambda > 0$, $x \in \mathbb{R}$.

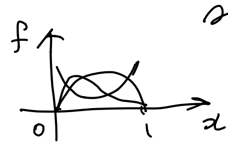
$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$




→ "Memoryless property"
Same distribution whether or not "lightbulb" has been already used.

iii) $X \sim \text{Beta}(\alpha, \beta)$, $\alpha, \beta > 0$, $\alpha, \beta \in \mathbb{R}$.

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, & x \in [0, 1] \\ 0, & \text{otherwise} \end{cases}$$




→ Often used to model the distribution of a r.v. which, itself is a probability.
"r.v. which itself is a probability".

iv) $X \sim \text{Normal}(\mu, \sigma^2)$, $\mu, \sigma^2 \in \mathbb{R}$, $\sigma^2 > 0$.

→ "Gaussian Distribution"

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