# Lecture 17:
# Expectation and Variance - Part IV
# &
# Continuous Random Variables - Part I

Satyajit Thakor
IIT Mandi

## Example

- There are $k$ distinguishable balls and $n$ distinguishable boxes. The balls are randomly placed in the boxes, with all $n^k$ possibilities equally likely. Problems in this setting are called occupancy problems, and are used in computer science (for example, for randomized algorithms).
- Find the expected number of empty boxes in terms of the parameters $n, k$.

- Let $X_i$ be the r.v. such that $\{X_i = 1\}$ corresponds to $i$th box being empty and $\{X_i = 0\}$ corresponds to $i$th box not empty.
- Then, the no. of empty boxes is the r.v. $X = \sum_{i=1}^{n} X_i$.
- Expected no. of empty boxes is

# Example

$$E(X) = E\sum_{i=1}^{n} X_i = \sum_{j=1}^{n} E(X_i) \quad (\because \text{ linearity of } E(\cdot))$$

- Now, $E(X_i) = P(X_i = 1)\cdot 1$

$$= P(\{\text{there are no balls in } i^{th} \text{ box}\})$$

$$= P\Big(\bigcap_{j=1}^{K} \underbrace{\{j^{th} \text{ ball is not placed in } i^{th} \text{ box}\}}\Big)$$

$\Downarrow$ $\because$ independent events

$$= \prod_{j=1}^{K} P(\{j^{th} \text{ ball is not placed in } i^{th} \text{ box}\})$$

$$= \prod_{j=1}^{K} \Big[1 - \underbrace{P(\{j^{th} \text{ ball is placed in } i^{th} \text{ box}\})}\Big]$$

$\because$ ball is placed in any of n boxes randomly.

$$= \prod_{j=1}^{K} \Big[1 - \frac{1}{n}\Big] = \Big[1 - \frac{1}{n}\Big]^K \Rightarrow E(X) = n\Big[1 - \frac{1}{n}\Big]^K.$$

# Example

- Suppose that $X$ and $Y$ are independent r.v.s for which $\text{Var}(X) = \text{Var}(Y) = 3$. Find $\text{Var}(X - Y)$.

- $X$ and $Y$ are independent $\Rightarrow$ $X$ and $-Y$ are independent.

($\because$ functions of independent r.v.s are independent)

- Now,

$$
\begin{aligned}
\text{Var}(X-Y) &= \text{Var}(X+(-Y)) \\
&= \text{Var}(X) + \text{Var}(-Y) \quad \left(\because Z \text{ and } W \text{ are independent} \Rightarrow \text{Var}(Z+W) = \text{Var}(Z) + \text{Var}(W)\right) \\
&= \text{Var}(X) + (-1)^2\,\text{Var}(Y) \quad \left(\because \text{Var}(cX) = c^2\,\text{Var}(X)\right) \\
&= \text{Var}(X) + \text{Var}(Y) \\
&= 3 + 3 = 6.
\end{aligned}
$$

# Example

- A person wishes to insure his car for 200,000 rupees. The insurance company estimates that a total loss[1] will occur with probability 0.002, a 50% loss with probability 0.01, and a 25% loss with probability 0.1. Ignoring all other partial losses, what premium should the insurance company charge each year to realize an average profit of 500 rupees?

– Let $X$ = Claim amount. Then,

| $x$: | 200,000 | 100,000 | 50,000 | 0 |
|---|---|---|---|---|
| $p_X(x)$: | 0.002 | 0.01 | 0.1 | 0.888 |

$$\Rightarrow \text{Expected claim} = E(X) = 200000 \cdot 0.002 + 100000 \cdot 0.01 + 50000 \cdot 0.1$$

$$= 6400/-$$

Hence, the company should charge the premium : $6400 + \underbrace{500}_{\text{required average profit}}$

$= 6900/-$

[1] Here, loss of the insurance company means the claim of the person after accident of the car. This amount the company will pay to the person.

# Example

- A coin is biased such that a head is three times as likely to occur as a tail. Find the expected number of tails when this coin is tossed twice.

- For the given biased coin: $P(\{H\}) = 3/4$, $P(\{T\}) = 1/4$.

- Let $X$ be the no. of tails in two tosses. Then

$$P(X=0) = P(\{H,H\}) = 9/16$$

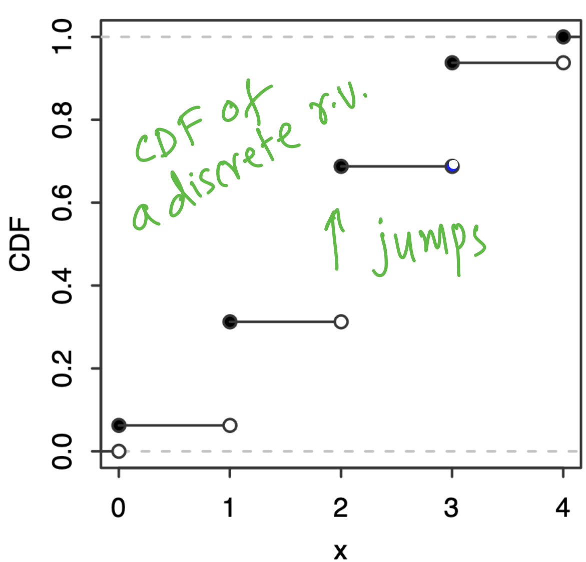
$$P(X=1) = P(\{H,T\}) + P(\{T,H\}) = 2 \cdot 3/4 \cdot 1/4 = 3/8$$

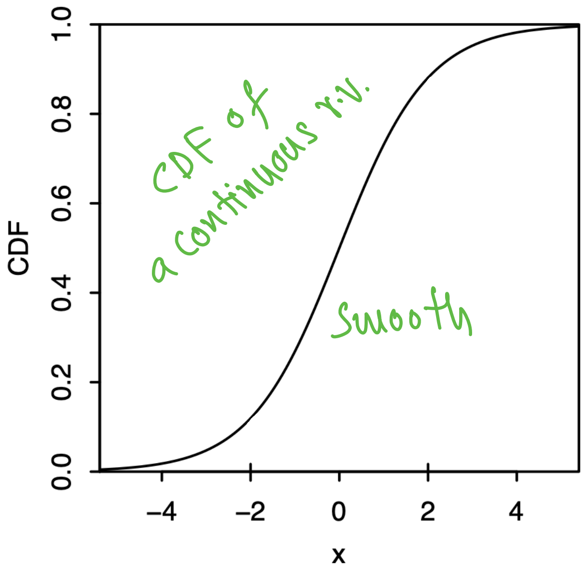
$$P(X=2) = P(\{T,T\}) = 1/4 \cdot 1/4 = 1/16$$

- Then, $E(X) = 0 \cdot 9/16 + 1 \cdot 3/8 + 2 \cdot 1/16 = 4/8 = 1/2$.

# Continuous random variables

- An r.v. has a continuous distribution if its CDF is differentiable. (or if it is at least (1) continuous and (2) not differentiable at finite number of points)
- A continuous r.v. is an r.v. with a continuous distribution.




- CDF of a discrete r.v. has jumps, CDF of a continuous r.v. is smooth.

# Continuous random variables

CDF to PDF

- For a continuous r.v. $X$ with CDF $F_X$, the probability density function (PDF) of $X$ is the derivative $f_X$ of the CDF, given by

$$\boxed{f_X(x) = F'_X(x) = \frac{d}{dx}F_X(x).}$$

- PDF to CDF: Since $F_X$ is an antiderivative of $f_X$, CDF can be obtained by integration of PDF:

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

- Caution:
  1. For a continuous r.v. $X$, $P(X = x) = 0$ for all $x$ because $P(X = x)$ is the height of a jump in the CDF at $x$, but the CDF of $X$ has no jumps.
  2. $f_X(x)$ is not a probability, and in fact it is possible to have $f_X(x) > 1$ for some values of $x$ (example will be discussed in the next lecture).