

ST2334 – Chapter 2 – Concepts of Random Variables

2.1 – Introduction

2.2 – Discrete Probability Distributions

2.3 – Continuous Probability Distributions

2.4 – Cumulative Distribution Function

2.5 – Mean and Variance of a Random Variable

2.6 – Chebychev's Inequality

2.1 - Introduction

- Some function of the outcome > actual outcome

2.1.1 - Random Variable

- **Random variable**: a function X which assigns a number to every event in S element $s \in S$.

$$R_X = \{x \mid x = X(s), s \in S\}$$

2.1.2 - Equivalent Events

- Suppose $A = \{s \in S \mid X(s) \in B\}$

Then A and B are **equivalent events** and $\Pr(B) = \Pr(A)$.

A consists of all sample points, s , in S for which $X(s) \in B$.

2.2 - Discrete Probability Distributions

2.2.1 - Discrete Random Variable $\rightarrow R_X$, the range space

- Definition 2.3 : Let X be a random variable

Number of possible values of X is finite / countable infinite

$\rightarrow X$ is a discrete random variable

2.2.2 - Probability Function \rightarrow a.k.a probability mass function, p.m.f.

- Each value of X has a probability function $f(x)$
- Probability distribution of X : collection of pairs $(x_i, f(x_i))$

- Must satisfy :

$$\textcircled{1} \quad f(x_i) \geq 0 \quad \forall x_i$$

$$\textcircled{2} \quad \sum_{i=1}^{\infty} f(x_i) = 1$$

2.2.3 - Another View of Probability Function

- Specifying a mathematical model for a finite population

2.3 - Continuous Probability Distributions

2.3.1 - Continuous Random Variables

- Definition 2.4 : Suppose R_x , the range space of a random variable X , is an interval or a collection of intervals.

Then X is a continuous random variable.

2.3.2 - Probability Density Function

- Definition 2.5 : Let X be a continuous random variable.

The probability density function $f(x)$ satisfies :

$$\textcircled{1} \quad f(x) \geq 0 \quad \forall x \in R_x$$

$$\textcircled{2} \quad \int_{R_x} f(x) dx = 1 \quad \text{or} \quad \int_{-\infty}^{\infty} f(x) dx = 1 \quad \text{since } f(x) = 0 \text{ for } x \notin R_x.$$

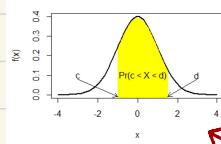
$$\textcircled{3} \quad \forall c, d \text{ st. } c < d, \quad \Pr(c \leq X \leq d) = \int_c^d f(x) dx$$

- In a probability statement, \leq and $<$ can be used interchangeably

- $\Pr(A) = 0 \Rightarrow A = \emptyset$

- X assumes values only in some interval $[a, b]$

→ $f(x) = 0 \quad \forall X \text{ outside } [a, b]$



2.4 - Cumulative Distribution Function

- Definition 2.6: Let X be a random variable, discrete or continuous.

Cumulative distribution function of X : $F(x) = \Pr(X \leq x)$

2.4.1 - CDF for Discrete Random Variables

- X is a ^{step function} discrete random variable

$$\rightarrow F(x) = \sum_{t \leq x} f(t) = \sum_{t \leq x} \Pr(X=t)$$

- For any a, b with $a \leq b$,

$$\begin{aligned} \Pr(a \leq X \leq b) &= \Pr(X \leq b) - \Pr(X < a) \\ &= F(b) - F(a^-) \end{aligned}$$

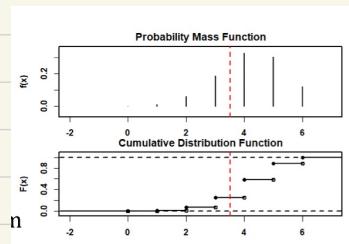
↗ largest possible $X < a$

- The only possible values are integers and a, b are integers

$$\rightarrow \Pr(a \leq X \leq b) = \Pr(X=a \text{ or } a+1 \text{ or } \dots \text{ or } b)$$

$$\hookrightarrow \Pr(a \leq X \leq b) = F(b) - F(a-1)$$

$$- a = b \rightarrow \Pr(X=a) = F(a) - F(a-1)$$



2.4.2 - CDF for Continuous Random Variables

- X is a continuous random variable $\rightarrow F(x) = \int_{-\infty}^x f(t) dt$

- For a continuous random variable X , $f(x) = \frac{d F(x)}{dx}$ \rightarrow if the derivative exists

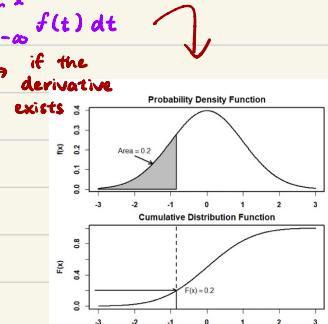
$$\Pr(a \leq X \leq b) = \Pr(a < X \leq b)$$

$$= F(b) - F(a)$$

- $F(x)$ is a non-decreasing function

$$\text{i.e. } x_1 < x_2 \rightarrow F(x_1) \leq F(x_2)$$

$$- 0 \leq F(x) \leq 1$$



2.5 - Mean and Variance of a Random Variable

2.5.1 - Expected Values

- Definition 2.7a: X is a discrete random variable, taking on values x_1, x_2, \dots

with probability function $f_X(x)$

$$\rightarrow \text{mean / expected value of } X : \mu_X = E(X) = \sum_i x_i f_X(x_i) = \sum_x x f_X(x)$$

Not necessarily a possible value of X

- Definition 2.7b: X is a continuous random variable with probability density function $f_X(x)$

$$\rightarrow \text{mean of } X : \mu_X = E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

(weighted) average

- The expected value exists provided the sum / integral in the above definitions exists

- Discrete case: $f_X(x) = \frac{1}{N}$ for each N of x

$$\rightarrow E(X) = \sum_i x_i f(x_i) = \frac{1}{N} \sum x_i$$

2.5.2 - Expectation of a Function of a RV

- Definition 2.8: For any function $g(x)$ of a random variable X with p.f./p.d.f. $f_X(x)$,

① If X is a discrete r.v. providing sum exists: $E[g(x)] = \sum_x g(x) f_X(x)$

② If X is a continuous r.v. providing integral exists: $E[g(x)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$

- Special case 1: \rightarrow definition of variance of a given r.v. X

$g(x) = (x - \mu_X)^2$

- Definition 2.9: Let X be a r.v. with p.f./p.d.f. $f(x)$.

Then variance of X : $\sigma_X^2 = V(X) = E[(X - \mu_X)^2]$

$$= \begin{cases} \sum_x (x - \mu_X)^2 f_X(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

- $V(X) \geq 0$

- $V(X) = E(X^2) - [E(X)]^2$ \rightarrow positive

- Standard deviation of X : $\sigma_X = \sqrt{V(X)}$

- Special case 2: \rightarrow k-th moment of X

$$g(x) = x^k$$

2.5.3 - Properties of Expectation

① $E(aX + b) = aE(X) + b \longrightarrow a, b \text{ are constants}$

② $V(X) = E(X^2) - [E(X)]^2$

③ $V(aX + b) = a^2 V(X)$

2.6 - Chebyshov's Inequality

- We know probability distribution of a r.v. $X \rightarrow$ we can compute $E(X)$ and $V(X)$, but
- We know probability distribution of a r.v. $X \cancel{\rightarrow}$ we can compute $E(X)$ and $V(X)$
- \Rightarrow we cannot compute quantities e.g. $\Pr(|X - E(X)| \leq c)$ positive constant

- Chebyshov's inequality:** let X be a r.v. with $E(X) = \mu$ and $V(X) = \sigma^2$.
Then for any $k > 0$, we have $\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$
or $\Pr(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$

- $k > 0$

- Inequality is true \forall distributions with finite mean and variance

- Theorem gives a lower probability on the probability that $|X - \mu| < k\sigma$

Probability that value of X lies $\geq k$ s.d. from its mean is $\leq \frac{1}{k^2}$