Random Variables

Random Variables

A random variable is a variable whose possible values are numerical outcomes of a random phenomenon

Examples of Random Variables

Let C = outcome when we toss a coin

C can take take two values, H and T.

Let X = number of heads when we toss a coin

X can take two values, 0 and 1.

Let Y = number that comes up when we roll a die

Y can take values 1, 2, 3, 4, 5 or 6.

Let Z = weight of a randomly selected student in this class

Z can take any value between 0 and ∞.

Notation

It is conventional to denote the **random variable by a capital letter** and the **possible values it can take by a small letter**.

Examples:

Let $\mathbf{X}$ = number of heads when we toss a coin, then $\mathbf{x} \in \{0, 1\}$

Let $\mathbf{Z}$ = weight of a randomly selected student in this class, then $\mathbf{z} \in (0, \infty)$

Random Variable : Another Perspective

A **sample space** is the **set** of all **possible outcomes** of an **experiment**.

A random variable is a rule for associating a number with each element in a sample space.

So, if **w** is an element of the sample space **S** (i.e, w is one of the possible outcomes of the experiment concerned) and the number **x** is associated with this outcome, then **X(w) = x**.

Example:

Experiment: **Tossing a coin**
Sample space: **S = {H, T}**.
X is the number of heads when we toss a coin.
Then, **X(H) = 1** and **X(T) = 0**.

Example

Suppose there are 8 balls in a bag.
The random variable X is the weight, in kg, of a ball selected at random.
Balls 1, 2 and 3 weigh 0.1kg, balls 4 and 5 weigh 0.15kg and balls 6, 7 and 8 weigh 0.2kg.

Experiment: Selecting a ball at random
Sample Space: $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$
X is weight of a ball selected at random

Then,
$X(1) = 0.1, X(2) = 0.1, X(3) = 0.1,$
$X(4) = 0.15, X(5) = 0.15,$
$X(6) = 0.2, X(7) = 0.2, X(8) = 0.2$

Types of Random Variables

Discrete Random Variables

- Countable
- Discrete Points
- E.g. X = number that comes when you roll a dice, i.e, $X \in \{1,2,3,4,5,6\}$

Continuous Random Variables

- Uncountable
- Continuous Intervals
- E.g. X = height of a randomly chosen student in this class

Discrete Random Variables

Probabilities

Probabilities are defined on events (subsets of S).

So what is meant by "**P(X = x)**"?

Experiment: **Selecting a ball at random**

Sample Space: **S = {1, 2, 3, 4, 5, 6, 7, 8}**

X is weight of a ball selected at random

What does P(X=0.1) mean?

P(X = 0.1) means the probability that either ball 1, ball 2 or ball 3 is selected.

If E = event that either ball 1, ball 2 or ball 3 is selected, then E = {1, 2, 3}

P(X=0.1) = P(E).

Probability Distribution Function

The function $\mathbf{f_x(x)}$ = P(X=x) for each x in the range of X is the **probability function** (PF) of X.
It specifies how the total probability of 1 is divided up amongst the possible values of X and so gives the **probability distribution** of X.
Probability functions are also known as "**probability distribution functions**".

Properties: (i) $\mathbf{f_x(x) > 0}$ **for all x in range of X** (ii) $\mathbf{\Sigma f_x(x) = 1}$

X is weight of a ball selected at random.
$\mathbf{f_x(0.1) = P(X = 0.1) = 3/8}$
$\mathbf{f_x(0.15) = P(X = 0.15) = 2/8 = 1/4}$
$\mathbf{f_x(0.2) = P(X = 0.20) = 3/8}$

Cumulative Distribution Function

The **cumulative distribution function** (CDF) of X is $\mathbf{F_x(x)}$ = P(X ≤ x).
It gives the probability that X assumes a value that does not exceed x.
CDFs are also known as “**Distribution functions (DF)**”.

X is weight of a ball selected at random.

$$\mathbf{F_x(0) = P(X \leq 0) = 0}$$
$$\mathbf{F_x(0.1) = P(X \leq 0.1) = P(X = 0.1) = \tfrac{3}{8}}$$
$$\mathbf{F_x(0.15) = P(X \leq 0.15) = P(X = 0.1) + P(X = 0.15) = \tfrac{3}{8} + 2/8 = \tfrac{5}{8}}$$
$$\mathbf{F_x(0.2) = P(X \leq 0.2) = P(X = 0.1) + P(X = 0.15) + P(X = 0.2) = \tfrac{3}{8} + 2/8 + \tfrac{3}{8} = 1}$$

Continuous Random Variables

Probability Density Function

What does **P(X=x)** mean for a continuous random variable?

It's always **zero**.
In case of continuous variables we always take **intervals** into account.

The probability associated with an interval of values, **(a, b)** say, is represented as **P(a < X < b)** – and is the area under the curve of the probability density function (PDF) from a to b. So probabilities can be evaluated by integrating the PDF $f_X(x)$. This relationship defines the PDF.

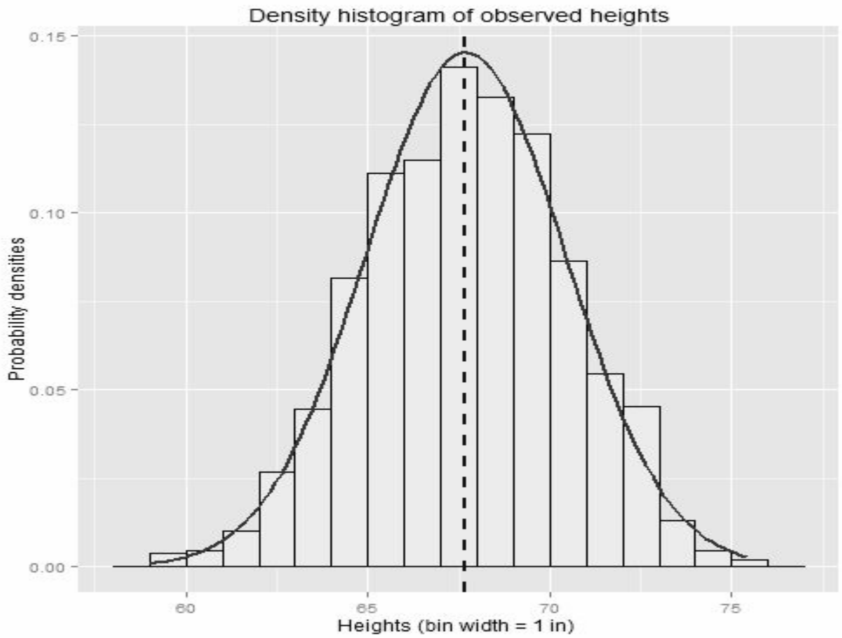
$$P(a < X < b) = \int_a^b f_X(x)\, dx$$

Probability Density Function

- How **dense** are the values in the **vicinity** of x?
- The conditions for a function to serve as a PDF are as follows:

$$f_X(x) \geq 0 \qquad -\infty \leq x \leq \infty$$

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$$



Probability Density Function

Let take an example:

$$f_W(w) = 12w^2(1-w) \qquad 0 < w < 1$$

Check if these conditions hold true:

$$f_X(x) \geq 0 \qquad -\infty \leq x \leq \infty$$

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

Cumulative Distribution Function

The cumulative distribution function (CDF) is defined to be the function:
$F_X(x) = P(X \leq x)$

For a continuous random variable, $F_X(x)$ is a **continuous**, **non-decreasing** function, defined for all real values of x.

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

Cumulative Distribution Function

Let take an example: $$f_W(w) = 12w^2(1-w) \qquad 0 < w < 1$$

Calculate the CDF: $$F_W(w) = \int_0^w 12t^2(1-t)\,dt = \left[4t^3 - 3t^4\right]_0^w = w^3(4-3w)$$

Mean of Random Variables

Mean

E[X] is a measure of the average/centre/location/level of the distribution of **X**.
It is called the **expected value** of X, or **mean** of X, and is usually denoted **μ**.

$$E[X] = \sum_{x} x f_X(x) \quad \text{for the discrete case}$$

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx \quad \text{for the continuous case}$$

Mean - Discrete Random Variables

X = number that comes when we roll a fair dice; $x \in \{1,2,3,4,5,6\}$

x	1	2	3	4	5	6
P(X=x)	1/6	1/6	1/6	1/6	1/6	1/6

Expected value of X = E[X]

$= \Sigma\, x * P(X=x)$

$= (1 * \frac{1}{6}) + (2 * \frac{1}{6}) + (3 * \frac{1}{6}) + (4 * \frac{1}{6}) + (5 * \frac{1}{6}) + (6 * \frac{1}{6})$

= **3.5**

Mean - Continuous Random Variables

Let's take an example

$$f_W(w) = 12w^2(1-w) \qquad 0 < w < 1$$

Calculate E[W]

$$\int_0^1 w f_W(w)\,dw = \int_0^1 12w^3(1-w)\,dw = \int_0^1 12w^3 - 12w^4\,dw = \left[3w^4 - \frac{12}{5}w^5\right]_0^1 = 0.6$$

Variance of Random Variables

Expected value of g(X)

$$E[g(x)] = \sum_{x} g(x) f_X(x) \quad \text{for the discrete case}$$

$$E[g(x)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx \quad \text{for the continuous case}$$

Variance

The variance **σ^2** is a measure of the spread/dispersion/variability of the distribution. Specifically, it is a measure of the **spread** of the distribution about its mean.

$$\mathrm{var}[X] = E[\{X - E[X]\}^2]$$

$$\mathrm{var}[X] = E[X^2] - \mu^2$$

Variance - Discrete Random Variables

X = number that comes when we roll a fair dice; $x \in \{1,2,3,4,5,6\}$

x	1	2	3	4	5	6
P(X=x)	1/6	1/6	1/6	1/6	1/6	1/6

Expected value of $X^2 = E[X^2]$

$$= \Sigma\, x^2 * P(X=x)$$

$$= (1^2 * \tfrac{1}{6}) + (2^2 * \tfrac{1}{6}) + (3^2 * \tfrac{1}{6}) + (4^2 * \tfrac{1}{6}) + (5^2 * \tfrac{1}{6}) + (6^2 * \tfrac{1}{6})$$

$$= \mathbf{15.167}$$

Variance of X = $15.167 - 3.5^2 =$ **2.917**

Variance - Continuous Random Variables

Let's take an example

$$f_W(w) = 12w^2(1-w) \qquad 0 < w < 1$$

Calculate E[W²]

$$E[W^2] = \int_0^1 12w^4(1-w)\,dw = \int_0^1 12w^4 - 12w^5\,dw = \left[\frac{12}{5}w^5 - 2w^6\right]_0^1 = 0.4$$

Variance, var[W] = E[W²] - (E[W])² = 0.4 - 0.6² = 0.04

Linear Combinations of Random Variables

Mean & Variance of Linear Combinations

Let $Y = aX + b$. Let $E[X] = \mu$.

$$E[Y] = E[aX + b] = a\mu + b$$

So $Y - E[Y] = aX + b - [a\mu + b] = a[X - \mu]$.

Hence $\mathrm{var}[Y] = E\left[\{Y - E[Y]\}^2\right] = a^2 E\left[(X - \mu)^2\right] = a^2\,\mathrm{var}[X]$.

Mean & Variance of Linear Combinations

Y = 2X

Mean X = 10
Variance X = 15

Mean Y = 2 * Mean X = 2 * 10 = 20
Variance Y = **2^2** * Variance X = 4 * 15 = 60