4.3 Random Variable

A random variable is a variable taking numerical values determined by the outcome of a random phenomenon. We use capital letter X, Y or Z to denote it.

A random variable is:

- **Discrete** when only certain separate numbers are possible as values.
- **Continuous** when any value on an interval of numbers is possible.

1. Probability distribution of a discrete random variable X

We only consider a discrete random variable with finite number of possible values.

The probability distribution of X lists the values and their probabilities:

Value of X	x_1	x_2	x_3	…	x_k
Probability	p_1	p_2	p_3	…	p_k

The probability p_i must satisfy two requirements:

i) Every probability p_i is a number between 0 and 1.

ii) $p_1 + p_2 + \dots + p_k = 1$.

Note:

- Any event can be expressed in terms of X. Therefore, given probability distribution of X, it is easy to compute probability of any event.
- If probability distribution of X is not given, compute probability distribution first by using the known information.

2. Probability distribution of a continuous random variable X

Density curve: a density curve is a curve that

- Is always on or above the horizontal axis, and
- Has are exactly 1 underneath it.

Note:

- The probability distribution of X is described by a density curve.
- The probability of any event = the area under the curve and above the values of X that make up the event.
- For a continuous random variable X, P(X=x)=0. i.e. probability of every individual outcome is zero. $P(X \geq x) = P(X > x)$, $P(X \leq x)=P(X < x)$.

Examples of continuous distribution: normal distr. $N(\mu, \sigma)$. Uniform distr.

4.4 Means and Variances of Random Variables

Random variable X ⟶ Probability distribution of X
Mean of X: μ or μ_x
Variance of X: σ^2 or σ^2_x

1. The mean of a random variable

a) For a discrete random variable

Value of X	x_1	x_2	x_3	…	x_k
Probability	p_1	p_2	p_3	…	p_k

$$\mu_x = x_1 p_1 + x_2 p_2 + \cdots + x_k p_k$$
$$= \sum x_i p_i$$

b) For a continuous random variable X

μ_x is the point at which the area under the density curve would balance if it were made out of solid material. Therefore if density curve is symmetric, mean lies at its center.

c) The law of large numbers

Suppose we have a population with mean μ, and we draw n independent observations at random from the population:

$$x_1, x_2, x_3, \cdots, x_n$$

$$\overline{X} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

Then

i) $\mu \approx \overline{X}$

ii) If n increases, $\overline{X}$ gets closer to and eventually approaches μ

---Law of Large Numbers.

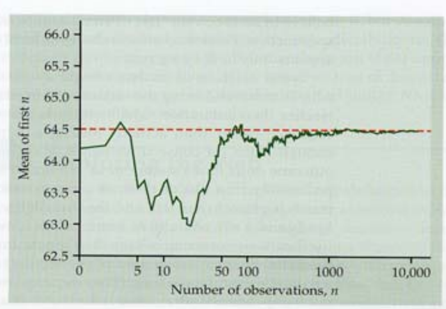

FIGURE 4.14 The law of large numbers in action. As we take more observations, the sample mean $\bar{x}$ always approaches the mean μ of the population.

d) Rules for means

If X and Y are random variables and a, b, c are constants, then a+bX, X+Y and a+bX+cY are also random variables.

$$\left.\begin{aligned} \mu_{a+bX} &= a + b\mu_x \\ \mu_{X+Y} &= \mu_X + \mu_Y \end{aligned}\right\} \Rightarrow \mu_{a+bX+cY} = a + b\mu_x + c\mu_Y$$

Example: Gain Communications sells aircraft communications units to both the military and the civilian markets. Next year's sales depend on market conditions that can not be predicted exactly. Gain follows the modern practice of using probability estimates of sales. The military division estimates its sales as follows:

Units sold	1000	3000	5000	10,000
Probability	0.1	0.3	o.4	o.2

These are personal probabilities that express the informed opinion of Gain's executives. The corresponding sales estimates for the civilian division are

Units sold	300	500	750
Probability	0.4	0.5	0.1

Let X be the no. of military units sold and Y the no. of civilian units. Gain makes a profit of $2,000 on each military unit sold and $3500 on each civilian unit.

What is the estimate of the company's next year's profit?

2. The variance of a random variable

a) For a discrete random variable

Value of X	x_1	x_2	x_3	…	x_k
Probability	p_1	p_2	p_3	…	p_k

The variance of X is:

$$\begin{aligned} \sigma_X^2 &= (x_1 - \mu_X)^2 p_1 + (x_2 - \mu_X)^2 p_2 + \cdots + (x_k - \mu_X)^2 p_k \\ &= \sum (x_i - \mu_X)^2 p_i \end{aligned}$$

b) Rules for variance

Rule 1: $\sigma_{a+bX}^2 = b^2 \sigma_X^2$

Rule 2: addition rule for variances of independent random variables

$$\begin{cases} \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 \\ \sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 \end{cases} \Rightarrow \sigma_{aX+bY}^2 = a^2\sigma_X^2 + b^2\sigma_Y^2$$

Rule 3: General addition rules for variances of random variables. If X and Y have correlation ρ , then:

$$\sigma^2_{X+y} = \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y$$

$$\sigma^2_{X-y} = \sigma^2_X + \sigma^2_Y - 2\rho\sigma_X\sigma_Y$$

Where ρ is defined in a similar way as sample correlation r and is also a number between -1 and 1 that measures the direction and strength of the linear relationship between two random variables.

When two random variables X and Y are independent, ρ=0.