1 Mathematical Expectation

1.1 Mean of a Random Variable

- Suppose that a probability distribution of a random variable X is specified.
- For a measure of central tendency of the random variable X we use the terms **expectation**, **expected value**, and **average value** for the same concept.
- Intuitively, the expected value of X is the average value that the random variable takes on.
- However, some of the values of the random variable X could be more (or less) probable than the other in the distribution unless the random variable is distributed uniformly.
- Hence, in order to consider an **average** value of X we need to take its probability into account.
- If I repeat the experiment many times, what would be the average number of an outcome of a random variable?
- **Definition 4.1**:

  Let X be a random variable with probability distribution $f(x)$. The **mean** or **expected values** of X is

$$\left\{ \begin{array}{l} \mu = E(X) = \sum_x xf(x) \; if \; X \; is \; discrete \\ \mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx \; if \; X \; is \; continuous \end{array} \right\}$$

- The expected value is used as a measure of centering or location of the distribution of a random variable X.
- By the uniform distribution assumption, i.e. all values of X are equally likely to occur in population with size N, $f(x) = \frac{1}{N}$ for all x,

$$E(X) = \sum_x xf(x) = \sum_x x(\frac{1}{N}) = (\frac{1}{N})\sum_i x_i = \mu = \bar{x}$$

- **Example**: If two coins are tossed 16 times and X is the number of heads that occur per toss, then the value of X can be 0, 1, 2.

- The experiment yields no heads, one head, and two heads a total of 4, 7, and 5 times, respectively.
- The average number of heads per toss is then

$$0 * \frac{4}{16} + 1 * \frac{7}{16} + 2 * \frac{5}{16}$$

where $\frac{4}{16}, \frac{7}{16}, \frac{5}{16}$ are relative frequencies

x	0	1	2
$f(x)$	4/16	7/16	5/16

$$0 * \frac{4}{16} + 1 * \frac{7}{16} + 2 * \frac{5}{16} = \frac{17}{16} = 1.0625$$

- **Example 4.1**: A lot contain 4 good components and 3 defective components.
  - A sample of 3 is taken by a quality inspector.
  - Find the expected value of the number of good components in this sample.
- Solution: X represents the number of good components

$$f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, x = 0, 1, 2, 3$$

$$\mu = E(X) = 0 * f(0) + 1 * f(1) + 2 * f(2) + 3 * f(3) = \frac{12}{7}$$

- **Example 4.3**: Let X be the random variable that denotes the life in hours of a certain electronic device. The probability density function is as the following.

$$f(x) = \left\{ \begin{array}{ll} \frac{20000}{x^3}, & x > 100 \\ 0, & elsewhere \end{array} \right\}$$

  Find the expected life of this type of device.

- Solution:

$$\mu = E(X) = \int_{100}^{\infty} x \frac{20000}{x^3} dx = -\frac{20000}{x} j_{100}^{\infty} = 200$$

- **Mean of** $g(X)$ (any real-valued function): If X is a discrete random variable with $f(x)$, for $x = -1, 0, 1, 2$, and $g(X) = X^2$ then

$$\begin{aligned} &P[g(X) = 0] = P(X = 0) = f(0), \\ &P[g(X) = 1] = P(X = -1) + P(X = 1) = f(-1) + f(1), \\ &P[g(X) = 4] = P(X = 2) = f(2), \end{aligned}$$

- The probability distribution of $g(X)$ can be written

$g(x)$	0	1	4
$P[g(X) = 4]$	f(0)	f(-1)+f(1)	f(2)

-

$$\begin{aligned} E(g(X)) &= 0 * f(0) + 1 * [f(-1) + f(1)] + 4 * f(2) \\ &= (-1)^2 * f(-1) + (0)^2 * f(0) + (1)^2 * f(1) + (2)^2 * f(2) \\ &= \textstyle\sum_x g(x) * f(x) \end{aligned}$$

- **Theorem 4.1:**:

Let X be a random variable with probability distribution $f(x)$. The mean of the random variable $g(X)$ is

$$\left\{ \begin{array}{l} \mu_{g(X)} = E[g(X)] = \sum_x g(x)f(x) \ if \ X \ is \ discrete \\ \mu_{g(X)} = E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \ if \ X \ is \ continuous \end{array} \right\}$$

- **Example 4.5**: Let X be a random variable with density function

$$f(x) = \left\{ \begin{array}{ll} \frac{x^2}{3}, & -1 < x < 2 \\ 0, & elsewhere \end{array} \right\}$$

- Find the expected value of $g(X) = 4X + 3$.
- Solution:

$$E[g(X)] = E(4X + 3) = \frac{1}{3} \int_{-1}^{2} (4x^3 + 3x^2) dx = 8$$

- **Theorem 4.2:**:

Let X and Y be random variables with joint probability function $f(x), y$. The mean of the random variable $g(X,Y)$ is

$$\left\{ \begin{array}{l} \mu_{g(X,Y)} = E[g(X,Y)] = \sum_x \sum_y g(x,y)f(x,y) \\ if\ X\ and\ Y\ are\ discrete \\ \mu_{g(X,Y)} = E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy \\ \ if\ X\ and\ Y\ are\ continuous \end{array} \right\}$$

- **Example 4.7**: Find $E(Y/X)$ for the density function

$$f(x,y) = \left\{ \begin{array}{ll} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ ,0 < y < 1 \\ 0, & elsewhere \end{array} \right\}$$

- Solution:

$$E(\frac{Y}{X}) = \int_0^1 \int_0^2 \frac{y}{x} \frac{x(1+3y^2)}{4} dxdy = \frac{5}{8}$$

- If $g(X,Y) = X$ is

$$E(X) = \left\{ \begin{array}{l} \sum_x \sum_y xf(x,y) = \sum_x xg(x) \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x,y)dxdy = \int_{-\infty}^{\infty} xg(x)dx \end{array} \right\}$$

where $g(x)$ is the marginal distribution of X

- If $g(X,Y) = Y$ is

$$E(Y) = \left\{ \begin{array}{l} \sum_x \sum_y yf(x,y) = \sum_y yh(y) \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x,y)dxdy = \int_{-\infty}^{\infty} yh(y)dy \end{array} \right\}$$

where $h(y)$ is the marginal distribution of Y

1.2 Variance and Covariance

- A mean does not give adequate description of the shape of a random variable (probability distribution).
- We need to characterize the variability (or dispersion) of the random variable X in the distribution.

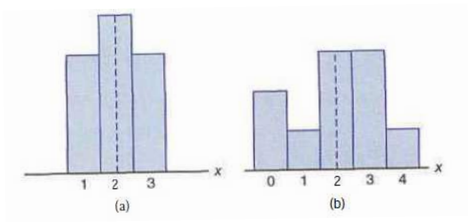

Figure 1: Distributions with equal means and unequal dispersions.

- **Definition 4.3**:

  Let X be a random variable with probability distribution $f(x)$ and mean μ. The **variance** of X is

  $$\left\{ \begin{array}{l} \sigma^2 = E\left[(X-\mu)^2\right] = \sum_x (x-\mu)^2 f(x), \; if \; X \; is \; discrete \\ \sigma^2 = E\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx, \; if \; X \; is \; continuous \end{array} \right\}$$

  σ is called the **standard deviation** of X.

- **Example 4.8**:Let the random variable X represent the number of automobiles that are used for official business purposes on any given workday.

- The probability distribution for company A and B is as follows.

x	1	2	3
$f(x)$	0.3	0.4	0.3

x	0	1	2	3	4
$f(x)$	0.2	0.1	0.3	0.3	0.1

- Show that the variance of the probability distribution for company B is greater than that of company A.

- Solution:

$$\mu_A = E(X) = 1 * 0.3 + 2 * 0.4 + 3 * 0.3 = 2.0$$

$$\sigma_A^2 = \sum_{x=1}^{3} (x-2.0)^2 f(x) = (1-2)^2 * 0.3 + (2-2)^2 * 0.4 + (3-2)^2 * 0.3 = 0.6$$

$$\mu_B = 2.0 \;\&\; \sigma_B^2 = 1.6$$

- **Theorem 4.2**:

  The **variance** of a random variable X is
  $$\sigma^2 = E(X^2) - \mu^2$$

- **Example 4.9**: Let the random variable X represent the number of defective parts for a machine when 3 parts are sampled from a production line and tested.

- Calculate σ^2 using the following probability distribution.

x	0	1	2	3
$f(x)$	0.51	0.38	0.10	0.01

- Solution:
$$\mu = E(X) = 0 * 0.51 + \ldots = 0.61$$
$$E(X^2) = \sum_{x=0}^{3} x^2 f(x) = 0^2 * 0.51 + \ldots = 0.87$$
$$\sigma^2 = E(X^2) - \mu^2 = 0.87 - 0.61^2 = 0.4979$$

- **Theorem 4.3**:

  Let X be a random variable with probability distribution $f(x)$. The variance of the random variable $g(X)$ is
  $$\left\{ \begin{array}{l} \sigma^2_{g(X)} = E\left\{[g(X) - \mu_{g(X)}]^2\right\} = \sum_x [g(X) - \mu_{g(X)}]^2, \\ if\ X\ is\ discrete \\ \sigma^2_{g(X)} = E\left\{[g(X) - \mu_{g(X)}]^2\right\} = \int_{-\infty}^{\infty} [g(X) - \mu_{g(X)}]^2 f(x)dx, \\ if\ X\ is\ continuous \end{array} \right\}$$

- **Example 4.11**: Calculate the variance of $g(X) = 2X + 3$, where X is a random variable with probability distribution.

x	0	1	2	3
$f(x)$	1/4	1/8	1/2	1/8

- Solution:
$$\mu_{2X+3} = E(2X + 3) = \sum_{x=0}^{3} (2X + 3) f(x) = 6$$

$$\sigma^2_{2X+3} = E\left\{[2X+3-\mu_{2X+3}]^2\right\} = E\left\{[2X+3-6]^2\right\}$$
$$= E(4X^2-12X+9) = \sum_{x=0}^{3}(4X^2-12X+9)f(x) = 4$$

- **Definition 4.4**:

> Let X and Y be random variables with joint probability distribution $f(x,y)$. The covariance of X and Y is
>
> $$\left\{\begin{array}{l} \sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x \sum_y (x-\mu_X)(y-\mu_Y)f(x,y), \\ \ \ if\ X\ and\ Y\ are\ discrete \\ \sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f(x,y)dxdy,\ if\ X\ andY\ are\ continuous \end{array}\right\}$$

- The covariance between two random variables is a measurement of the nature of the **association** between the two.
- The **sign** of the covariance indicates whether the relationship between two dependent random variables is positive or negative.
- When X and Y are statistically independent, it can be shown that the covariance is zero.
- The converse, however, is not generally true. Two variables may have zero covariance and still not be statistically independent.
- The covariance only describe the linear relationship between two random variables.
- If a covariance between X and Y is zero, X and Y may have a nonlinear relationship, which means that they are not necessarily independent.
- **Theorem 4.4**:

> The covariance of two random variables X and Y with means μ_X and μ_Y respectively, is given by
>
> $$\sigma_{XY} = E(XY) - \mu_X\mu_Y$$

- **Definition 4.5**:

> Let X and Y be random variables with covariance σ_{XY} and standard deviations σ_X and σ_Y. The **correlation coefficient** of X and Y is
>
> $$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y},\ -1 \leq \rho_{XY} \leq 1$$

- Exact linear dependency: $Y = a + bX$

$$\rho_{XY} = 1,\ if\ b > 0\ \ ;\ \ \rho_{XY} = -1,\ if\ b < 0$$

1.3 Means and Variance of Linear Combinations of Random Variables

- Some useful properties that will simplify the calculations of means and variances of random variables.
- These properties will permit us to deal with expectations in terms of other parameters that are either known or are easily computed.
- **Theorem 4.5**:

If a and b are constants, then $$E(aX + b) = aE(X) + b$$

- Corollary 4.1: $E(b) = b$
- Corollary 4.2: $E(aX) = aE(X)$
- **Example 4.16**: Applying Theorem 4.5 to the continuous random variable $g(X) = 4X + 3$, the density function of X is as follows.

$$f(x) = \left\{ \begin{array}{l} \frac{x^2}{3}\ for\ -1 < x < 2 \\ 0,\ elsewhere \end{array} \right\}$$

- Solution:

$$E(4X + 3) = 4E(X) + 3 = 4\left(\int_{-1}^{2} x\frac{x^2}{3}dx\right) + 3 = 8$$

- **Theorem 4.6**:

$$E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)]$$

- **Theorem 4.7**:

$$E[g(X, Y) \pm h(X, Y)] = E[g(X, Y)] \pm E[h(X, Y)]$$

- Corollary 4.3: Setting $g(X, Y) = g(X)$ and $h(X, Y) = h(Y)$.

$$E[g(X) \pm h(Y)] = E[g(X)] \pm E[h(Y)]$$

- Corollary 4.4: Setting $g(X,Y) = X$ and $h(X,Y) = Y$.

$$E[X \pm Y] = E(X) \pm E(Y)$$

- **Theorem 4.7**:

  Let X and Y be two independent random variables. Then

  $$E(XY) = E(X)E(Y)$$

- Corollary 4.5: Let X and Y be two independent random variables, Then $\sigma_{XY} = 0$
  - $E(XY) = E(X)E(Y)$ for independent variables
  - $\sigma_{XY} = E(XY) - E(X)E(Y) = 0$

- **Example 4.19**: In producing gallium-arsenide microchips, it is known that the ratio between gallium and arsenide is independent of producing a high percentage of workable wafers.

- Let X denote the ratio of gallium to arsenide and Y denote the percentage of workable wafers retrieved during a 1-hour period.

- X and Y are independent random variables with the joint density being known as

$$f(x) = \left\{ \begin{array}{l} \frac{x(1+3y^2)}{4} \; for \; 0 < x < 2, \quad 0 < y < 1 \\ 0, \; elsewhere \end{array} \right\}$$

  Illustrate that $E(XY) = E(X)E(Y)$.

- Solution:

$$E(XY) = \int_0^1 \int_0^2 xyf(x,y)dxdy = \int_0^1 \int_0^2 xy\frac{x(1+3y^2)}{4}dxdy = \frac{5}{6}$$

$$E(X) = \int_0^1 \int_0^2 xf(x,y)dxdy = \int_0^1 \int_0^2 x\frac{x(1+3y^2)}{4}dxdy = \frac{4}{3}$$

$$E(Y) = \int_0^1 \int_0^2 yf(x,y)dxdy = \int_0^1 \int_0^2 y\frac{x(1+3y^2)}{4}dxdy = \frac{5}{8}$$

$$(E(XY) =)\frac{5}{6} = \frac{4}{3} * \frac{5}{8}(= E(X) * E(Y))$$

- **Theorem 4.9**:

  If a and b are constants, then

  $$\sigma^2_{aX+b} = a^2\sigma^2_X = a^2\sigma^2$$

- Corollary 4.6: $\sigma^2_{X+b} = \sigma^2_X = \sigma^2$
  - The variance is unchanged if a constant is added to or subtracted from a random variable.
  - The addition or subtraction of a constant simply shifts the values of X to the right/left but does not change their variability.

- Corollary 4.7: $\sigma^2_{aX} = a^2\sigma^2_X = a^2\sigma^2$
  - The variance is multiplied or divided by the square of the constant.

- **Theorem 4.10**:

  If X and Y are random variables with joint probability distribution $f(x, y)$, then

  $$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma_{XY}$$

- Corollary 4.8: If X and Y are independent random variables, then

  $$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y$$

- Corollary 4.9: If X and Y are independent random variables, then

  $$\sigma^2_{aX-bY} = a^2\sigma^2_X + b^2\sigma^2_Y$$

- Corollary 4.10: If $X_1, X_2, \ldots X_n$ are independent random variables, then

  $$\sigma^2_{a_1X_1+a_2X_2+\ldots a_nX_n} = a_1^2\sigma^2_{X_1} + a_2^2\sigma^2_{X_2} + \ldots + a_n^2\sigma^2_{X_n}$$

- **Example 4.20**: X and Y are random variables with variances $\sigma^2_X = 2$, $\sigma^2_Y = 2$, and covariance $\sigma_{XY} = -2$,

- Find the variance of the random variable $Z = 3X - 4Y + 8$

- Solution:

$$\sigma^2_Z = \sigma^2_{3X-4Y+8} = \sigma^2_{3X-4Y} \text{ (by Theorem 4.9)}$$
$$= 9\sigma^2_X + 16\sigma^2_Y - 24\sigma_{XY} \text{ (by Theorem 4.10)}$$
$$= 130$$

- **Example 4.21**: Let X and Y denote the amount of two different types of impurities in a batch of a certain chemical product.
- Suppose that X and Y are independent random variables with variances $\sigma_X^2 = 2$, $\sigma_Y^2 = 3$
- Find the variance of the random variable $Z = 3X - 2Y + 5$
- Solution:

$$\begin{aligned}\sigma_Z^2 = \sigma_{3X-2Y+5}^2 &= \sigma_{3X-2Y}^2 \ (by\ Theorem\ 4.9) \\ &= 9\sigma_X^2 + 4\sigma_Y^2 \ (by\ Corollary\ 4.9) \\ &= 30\end{aligned}$$

1.4 Chebyshev's Theorem

- If a random variable has a small variance or standard deviation, we would expect most of the values to be grouped around the mean
- A large variance indicates a greater variability, so the area of distribution should be spread out more.

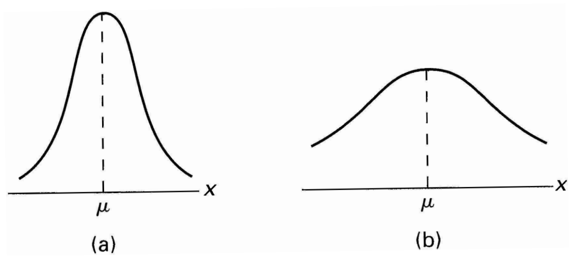

Figure 2: Variability of continuous observations about the mean.

- **Theorem 4.11**:

(**Chebyshev's theorem**) The probability that any random variable X will assume a value within k standard deviation of the mean is at least $1 - 1/k^2$. That is

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

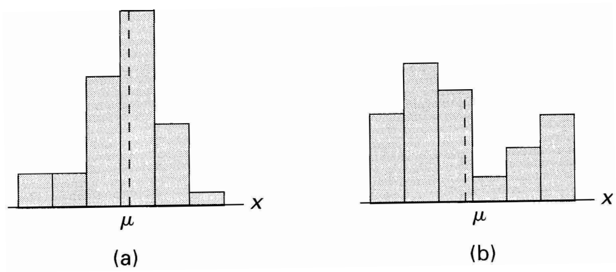

Figure 3: Variability of discrete observations about the: mean.

- **Example 4.22**: A random variable X has a mean $\mu = 8$, a variance $\sigma^2 = 9$, and an unknown probability distribution. Find
- $P(-4 < X < 20)$

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

$$P(-4 < X < 20) = P(8 - 4 * 3 < X < 8 + 4 * 3) \geq 1 - \frac{1}{4^2} = \frac{15}{16}$$

- $P(|X - 8| \geq 6)$

$$P(|X - 8| \geq 6) = 1 - P(|X - 8| < 6) = 1 - P(-6 < X - 8 < 6)$$

$$= 1 - P(8 - 6 < X < 6 + 8) = 1 - P(8 - 2 * 3 < X < 8 + 2 * 3) \leq \frac{1}{2^2} = \frac{1}{4}$$

- The Chebyshev inequality is a useful tool as well as a relation that connects the variance of a distribution with the intuitive notation of dispersion in a distribution.
- For any population or sample, this provides that the minimum probability of the data within $k\sigma$ from the mean μ is $1 - \frac{1}{k^2}$.
- The use of Chebyshev's theorem;
  - holds for any distribution of observations
  - gives a lower bound only
  - is suitable to situations where the form of the distribution is unknown (a distribution-free result)