Mathematics for Analytics and Finance

Sami Najafi

MSIS2402/ 2405

Module 6

Random Variables

Random Variables

Random Variable: a function that assigns a real value to each possible outcome in the sample space Ω.

- We often find ourselves more interested in the consequences of a random experiment's outcome than the experiment itself. For example, gamblers tend to be more concerned with the amount of their losses than the specific games that led to them.
- When these consequences can be expressed as real numbers, we can conceptualize them as functions that map the outcomes in the sample space Ω to the real line $\mathbb{R}$. These functions are known as random variables.

Example: A fair coin is tossed twice, and the sample space is $\Omega = \{HH, HT, TH, TT\}$. For each outcome $\omega \in \Omega$, let $X(\omega)$ be the number of heads H. Characterize the probability distribution based on the values of X.

Solution

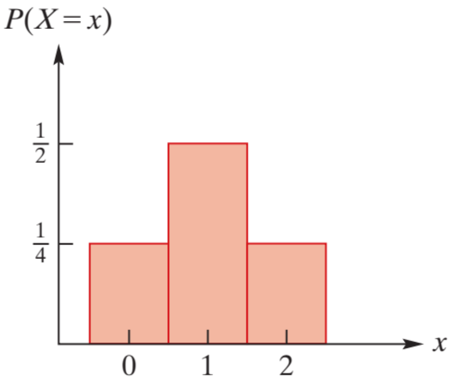
A fair coin is tossed twice: $\Omega = \{HH, HT, TH, TT\}$. For $\omega \in \Omega$, let $X(\omega)$ be the number of heads, so that:

$$X(HH) = 2, \qquad X(HT) = 1, \qquad X(TH) = 1, \qquad X(TT) = 0$$

$\mathrm{P}(X = 0) = \mathrm{P}(\{\omega \in \Omega : X(\omega) = 0\}) = \mathrm{P}(\{TT\}) = \frac{1}{4}.$

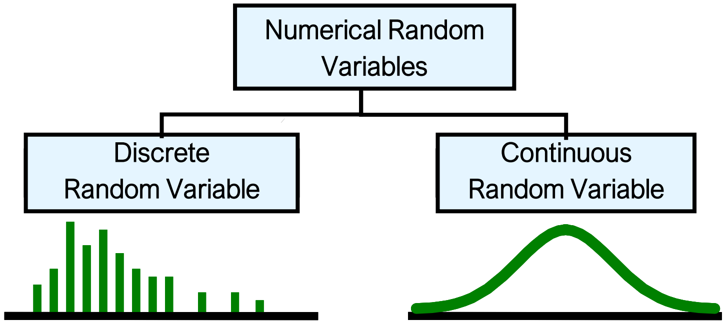
$\mathrm{P}(X = 1) = \mathrm{P}(\{\omega \in \Omega : X(\omega) = 1\}) = \mathrm{P}(\{HT, TH\}) = \frac{1}{2}.$

$\mathrm{P}(X = 2) = \mathrm{P}(\{\omega \in \Omega : X(\omega) = 2\}) = \mathrm{P}(\{HH\}) = \frac{1}{4}.$



Types of Random Variables

- Categorical: Values refer to defined groups or categories.
- Numerical: Values are continuous or discrete numbers:
  - Discrete Random Variables:
    - Produce countable discrete (distinct) values.
    - Produce individual values, with gaps between them
    - Example: Number of classes a student takes.
  - Continuous Random Variables:
    - Produce continuous values.
    - Can take on any value within a specific range or interval.
    - Examples: Salary, weight, etc.



Discrete Random Variables

Probability Mass Function (PMF)

Probability Mass Function: A listing of all possible outcomes for X, along with the probability of each occurrence:

$$p(x) = \mathrm{P}(X = x) = \mathrm{P}(\{\omega \in \Omega : X(\omega) = x\}), \qquad \sum_x p(x) = 1$$

Example: X = the number of classes taken by a student

x	$p(x) = \mathrm{P}(X = x)$
2	0.2
3	0.4
4	0.24
5	0.16

Cumulative Distribution Function (CDF): The cumulative distribution function (CDF) is defined as:

$$F_X(a) = \mathrm{P}(X \leq a) = \sum_{x \leq a} p(x)$$

Example

In the following discrete probability distribution, find $\mathrm{P}(2 \leq X \leq 4)$

x	$p(x)$
1	a
2	$2a$
3	$3a$
4	$4a$

Solution

x	$p(x) = \mathrm{P}(X = x)$
1	a
2	$2a$
3	$3a$
4	$4a$

Step I: Find the parameter a:

$$\sum_x p(x) = 1 \Rightarrow a + 2a + 3a + 4a = 1 \Rightarrow a = 0.1$$

Step II: Find the probability:

$$\mathrm{P}(2 \leq X \leq 4) = 2a + 3a + 4a = 9a = 0.9$$

Expectation

Definition: The mean value, expectation, or expected value of the random vector X with a probability mass function $p(x)$ is defined to be

$$\mu = \mathrm{E}(X) = \sum_x xp(x)$$

Whenever this sum is convergent (well-defined).

Example: Toss 2 coins and take X to be the number of heads. Compute expected value of X :

x	$p(x)$
0	0.25
1	0.5
2	0.25

$$\mu = \mathrm{E}(X) = \sum_x xp(x) = 0 \times 0.25 + 1 \times 0.5 + 2 \times 0.25 = 1$$

Expectation of Functions

Change of Variable Formula: If X has a probability mass function $p(x)$ and $g: \mathbb{R} \to \mathbb{R}$ then

$$\mathrm{E}(g(X)) = \sum_x g(x)p(x)$$

Whenever this sum is convergent (well-defined).

Example: Compute expected value of X^2 :

x	$p(x)$
-1	0.1
1	0.4
2	0.5

$$\mathrm{E}(X^2) = \sum_x x^2 p(x) = 1 \times 0.1 + 1 \times 0.4 + 4 \times 0.5 = 2.5$$

Expectation of Functions

Solving without the use of this proposition:

x	$p(x)$
-1	0.1
1	0.4
2	0.5

$g(x) = x^2$	$f(g)$
1	$0.1 + 0.4 = 0.5$
4	0.5

$$f(g) = \sum_{x:\, x^2 = g} p(x)$$

$$\mathrm{E}(g(X)) = \sum_g g f(g) = 1 \times 0.5 + 4 \times 0.5 = 2.5$$

$$\mathrm{E}(g(X)) = \sum_g g f(g) = \sum_g g \sum_{x:\, x^2 = g} p(x) = \sum_g \sum_{x:\, x^2 = g} g p(x) = \sum_x x^2 p(x)$$

Expectation of a Linear Combination

Change of Variable Formula Result: If $g(X) = aX + b$ where a and b are two constants, then

$$\mathrm{E}\big(g(X)\big) = \mathrm{E}(aX + b) = a\mathrm{E}(X) + b$$

$$\mathrm{E}(aX + b) = \sum_x (ax + b)p(x) = a\sum_x xp(x) + b\sum_x p(x) = a\mathrm{E}(X) + b$$

Variance

Definition: The variance of the random vector X with a probability mass function $p(x)$ is defined to be

$$\sigma^2 = \mathrm{var}(X) = \mathrm{E}\left[(X-\mu)^2\right] = \sum_x (x-\mu)^2 p(x)$$

whenever this sum is convergent (well-defined). The value $\sigma > 0$ is called the standard variation.

Example: Toss 2 coins and take X to be the number of heads. Compute the standard deviation of X:

x	$p(x)$
0	0.25
1	0.5
2	0.25

$$\mu = \sum_x xp(x) = 0 \times 0.25 + 1 \times 0.5 + 2 \times 0.25 = 1$$

$$\sigma = \sqrt{\sum_x (x-\mu)^2 p(x)} = \sqrt{(0-1)^2(0.25) + (1-1)^2(0.50) + (2-1)^2(0.25)} = \sqrt{0.5} = 0.707$$

Variance

Remark: $\sigma^2 = \mathrm{E}(X^2) - \mu^2$

$$\sigma^2 = \mathrm{E}\left[(X-\mu)^2\right] = \mathrm{E}\left(X^2 - 2X\mu + \mu^2\right) = \mathrm{E}\left(X^2\right) - 2\mu\mathrm{E}(X) + \mu^2 = \mathrm{E}\left(X^2\right) - \mu^2 \geq 0$$

x	$p(x)$
0	0.25
1	0.5
2	0.25

$$\mathrm{E}\left(X^2\right) = \sum_x x^2 p(x) = 1 \times 0.5 + 4 \times 0.25 = 1.5$$

$$\sigma^2 = \mathrm{E}\left(X^2\right) - \mu^2 = 1.5 - 1 = 0.5$$

Example

The number of no-shows each day for dinner reservations at the Cottonwood Grille is a discrete random variable with the following probability distribution:

No-shows	Probability
0	0.3
1	0.2
2	0.2
3	0.15
4	0.15

Based on this information, find the standard deviation for the number of no-shows?

Solution

No-shows	Probability
0	0.3
1	0.2
2	0.2
3	0.15
4	0.15

$$\mathrm{E}(X) = \mu = (1)(0.2) + \cdots + (4)(0.15) = 1.65$$

$$\mathrm{E}\left(X^2\right) = \left(1^2\right)(0.2) + \cdots + \left(4^2\right)(0.15) = 4.75$$

$$\sigma^2 = \mathrm{E}\left(X^2\right) - \mu^2 = 4.75 - 1.65^2 = 2.0275 \Rightarrow \sigma = 1.42$$

The other way to find the standard deviation is by using the original formula:

$$\sigma = \sqrt{(0 - 1.65)^2(0.3) + (1 - 1.65)^2(0.2) + \cdots + (4 - 1.65)^2(0.15)} = 1.42$$

Joint Random Variables

Definition: The joint cumulative probability for two random variables X and Y is

$$F(a,b) = \mathrm{P}(X \leq a, Y \leq b) = \sum_{x \leq a} \sum_{y \leq b} p(x,y),$$

Where $p(x,y)$ is the joint probability distribution (like a contingency table we saw before).

X

Y

	1	2	3	$p_Y(y)$
1	0.2	0.25	0.15	0.6
2	0.15	0.1	0.15	0.4
$p_X(x)$	0.35	0.35	0.3	1

$$F(2,2) = \mathrm{P}(X \leq 2, Y \leq 2) = 0.2 + 0.25 + 0.15 + 0.1 = 0.7$$

Joint Random Variables

Properties

- $F_{\mathbf{X}}(-\infty, -\infty) = 0$
- $F_{\mathbf{X}}(+\infty, +\infty) = 1$
- $p_X(a) = \sum_y p(a, y)$ (Law of Total Probability)
- $p_Y(b) = \sum_x p(x, b)$ (Law of Total Probability)
- $F_X(a) = \mathrm{P}(X \leq a) = \mathrm{P}(X \leq a, Y \leq +\infty) = F_{\mathbf{X}}(a, +\infty)$
- $F_Y(b) = \mathrm{P}(Y \leq b) = \mathrm{P}(X \leq +\infty, Y \leq b) = F_{\mathbf{X}}(+\infty, y)$

X

Y

	1	2	3	$p_Y(y)$
1	0.2	0.25	0.15	0.6
2	0.15	0.1	0.15	0.4
$p_X(x)$	0.35	0.35	0.3	1

Expectation of Joint Random Vectors

Change of Variable Formula: If the random vector (X, Y) has the PMF $p(X, Y)$ and $g: \mathbb{R}^2 \to \mathbb{R}$ then

$$\mathrm{E}[g(X,Y)] = \sum_{(x,y)} g(x,y)p(x,y) = \sum_x \sum_y g(x,y)p(x,y)$$

Change of Variable Formula Result: If $g(X, Y) = aX + bY$ then

$$\mathrm{E}[aX + bY] = a\mathrm{E}(X) + b\mathrm{E}(Y)$$

$$\mathrm{E}[aX + bY] = \sum_x \sum_y (ax + by)p(x,y) = a\sum_x x \sum_y p(x,y) + b\sum_y y \sum_x p(x,y)$$

$$= a\sum_x x p_X(x) + b\sum_y y p_Y(y) = a\mathrm{E}(X) + b\mathrm{E}(Y)$$

Covariance and Correlation Coefficient

Definition: The covariance of any two random variables X and Y with the joint probability mass function $p(x, y)$ is

$$\mathrm{cov}(X, Y) = \sigma_{XY} = \mathrm{E}\big[(x - \mathrm{E}(X))(y - \mathrm{E}(Y))\big] = \sum_{(x,y)} (x - \mathrm{E}(X))(y - \mathrm{E}(Y))p(x, y)$$

The correlation (or correlation coefficient) of X and Y is

$$\rho(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}$$

The covariance measures the strength of the linear relationship between two random variables X and Y.

- A positive covariance indicates a positive relationship.
- A negative covariance indicates a negative relationship.

Example

Consider the return per $\$1000$ for two types of investments as shown in the below table.

Prob.	Econ. Condition	Investment	
		Passive Fund (X)	Aggressive Fund (Y)
0.2	Recession	$-\$25$	$-\$200$
0.5	Stable	$+\$50$	$+\$60$
0.3	Expanding	$+\$100$	$+\$350$

(a) Which one has a higher mean rate of return?

Example

Consider the return per $\$1000$ for two types of investments as shown in the below table.

Prob.	Econ. Condition	Investment	
		Passive Fund (X)	Aggressive Fund (Y)
0.2	Recession	$-\$25$	$-\$200$
0.5	Stable	$+\$50$	$+\$60$
0.3	Expanding	$+\$100$	$+\$350$

(b) Which one has a higher investment risk?

Example

Consider the return per $\$1000$ for two types of investments as shown in the below table.

Prob.	Econ. Condition	Investment	
		Passive Fund (X)	Aggressive Fund (Y)
0.2	Recession	$-\$25$	$-\$200$
0.5	Stable	$+\$50$	$+\$60$
0.3	Expanding	$+\$100$	$+\$350$

(c) Is there any relationship between the two investments?

Solution

a)

$$\mathrm{E}(X) = \mu_X = (-25)(0.2) + (50)(0.5) + (100)(0.3) = 50$$

$$\mathrm{E}(Y) = \mu_Y = (-200)(0.2) + (60)(0.5) + (350)(0.3) = 95$$

Aggressive fund has a higher mean return per each $1000 invested.

b)

$$\sigma_X = \sqrt{(-25-50)^2(0.2) + (50-50)^2(0.5) + (100-50)^2(0.3)} = 43.3$$

$$\sigma_Y = \sqrt{(-200-95)^2(0.2) + (60-95)^2(0.5) + (350-95)^2(0.3)} = 193.71$$

Even though fund Y has a higher average return, it is subject to much more variability and the probability of loss is higher.

c)

$$\mathrm{cov}(X,Y) = (-25-50)(-200-95)(0.2) + (50-50)(60-95)(0.5) + (100-50)(350-95)(0.3) = 8250$$

$$\rho = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{8250}{(43.3)(193.7)} = 0.98$$

Since the correlation coefficient is large and positive, there is a positive relationship between the two investment funds, meaning that they will likely rise and fall together.

Example

Calculate the covariance and the correlation coefficient between the two random variables X and Y from the below probability distribution.

		X		
		1	2	Total
Y	1	0.25	0.3	0.55
	2	0.2	0.25	0.45
	Total	0.45	0.55	1

Solution

$\mathrm{E}(X) = 0.45 + 0.55 \times 2 = 1.55$

$\mathrm{E}(Y) = 0.55 + 0.45 \times 2 = 1.45$

$\mathrm{E}(X^2) = 0.45 + 0.55 \times 4 = 2.65$

$\mathrm{E}(Y^2) = 0.55 + 0.45 \times 4 = 2.35$

$\sigma_X^2 = 2.65 - 1.55^2 = 0.2475 \Rightarrow \sigma_X = 0.4975$

$\sigma_Y^2 = 2.35 - 1.45^2 = 0.2475 \Rightarrow \sigma_Y = 0.4975$

$\mathrm{cov}(X, Y) = (1 - 1.55)(1 - 1.45) \times 0.25$

$+(1 - 1.55)(2 - 1.45) \times 0.2$

$+(2 - 1.55)(1 - 1.45) \times 0.3$

$+(2 - 1.55)(2 - 1.45) \times 0.25$

$= 0.0025$

		X: 1	X: 2	Total
Y	1	0.25	0.3	0.55
	2	0.2	0.25	0.45
	Total	0.45	0.55	1

$$\rho = \frac{0.0025}{0.4975 \times 0.4975} = 0.01$$

Covariance Properties

- $\mathrm{cov}(X, X) = \mathrm{var}(X)$
- $\mathrm{cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$
- $\mathrm{cov}(X, a) = 0$ (a: a fixed value)
- $\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X)$
- $\mathrm{cov}(X, Y + Z) = \mathrm{cov}(X, Y) + \mathrm{cov}(X, Z)$
- $\mathrm{cov}(aX, bY) = ab\mathrm{cov}(X, Y)$ (a, b: fixed values)
- $\mathrm{var}(aX + bY) = \mathrm{cov}(aX + bY, aX + bY) = a^2\mathrm{var}(X) + b^2\mathrm{var}(Y) + 2ab\mathrm{cov}(X, Y)$ (a, b: fixed values)

Covariance Properties

Proofs (for interested students):

- $\mathrm{cov}(X,X) = \mathrm{E}[(X-\mathrm{E}(X))(X-\mathrm{E}(X))] = \mathrm{E}\left[\left(X-\mathrm{E}(X)\right)^2\right] = \mathrm{var}(X)$
- $\mathrm{cov}(X,Y) = \mathrm{E}[(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))] = \mathrm{E}\big(XY-\mathrm{E}(X)Y - X\mathrm{E}(Y) + \mathrm{E}(X)\mathrm{E}(Y)\big) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) - \mathrm{E}(X)\mathrm{E}(Y) + \mathrm{E}(X)\mathrm{E}(Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$
- $\mathrm{cov}(X,a) = \mathrm{E}[(X-\mathrm{E}(X))(a-\mathrm{E}(a))] = \mathrm{E}[(X-\mathrm{E}(X))(a-a)] = 0$
- $\mathrm{cov}(X,Y) = \mathrm{E}[(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))] = \mathrm{E}[(Y-\mathrm{E}(Y))(X-\mathrm{E}(X))] = \mathrm{cov}(Y,X)$
- $\mathrm{cov}(X,Y+Z) = \mathrm{E}[(X-\mathrm{E}(X))(Y+Z-\mathrm{E}(Y+Z))] = \mathrm{E}\big[\big(X-\mathrm{E}(X)\big)\big(Y-\mathrm{E}(Y)\big) + \big(X-\mathrm{E}(X)\big)\big(Z-\mathrm{E}(Z)\big)\big] = \mathrm{cov}(X,Y) + \mathrm{cov}(X,Z)$
- $\mathrm{cov}(aX,bY) = ab\mathrm{cov}(X,Y) = \mathrm{E}[(aX-\mathrm{E}(aX))(bY-\mathrm{E}(bY))] = ab\mathrm{E}[(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))] = ab\mathrm{cov}(X,Y),$
- $\mathrm{var}(aX+bY) = \mathrm{cov}(aX+bY, aX+bY) = a^2\mathrm{cov}(X,X) + b^2\mathrm{cov}(Y,Y) + ab\mathrm{cov}(X,Y) + ba\mathrm{cov}(Y,X) = a^2\mathrm{var}(X) + b^2\mathrm{var}(Y) + 2ab\mathrm{cov}(X,Y)$

Independent Random Variables

Two random variables X and Y are independent if and only if:

$$p(a, b) = \mathrm{P}(X = a, Y = b) = p_X(a)p_Y(b) \text{ for all } a, b \in \mathbb{R}, \quad \text{or equivalently:}$$

$$F(a, b) = \mathrm{P}(X \leq a, Y \leq b) = \mathrm{P}(X \leq a)\mathrm{P}(Y \leq b) \text{ for all } a, b \in \mathbb{R}$$

Remark: If X and Y are independent, then for any functions h and g

a) $\mathrm{E}[g(X)h(Y)] = \mathrm{E}[g(X)]\mathrm{E}[h(Y)]$

b) $\mathrm{cov}(X, Y) = 0$

c) $\mathrm{var}(a_1X_1 + a_2X_2) = a_1^2\mathrm{var}(X_1) + a_2^2\mathrm{var}(X_2)$

Remark: $\mathrm{cov}(X, Y) = 0$ does *not* mean independence

Example

Investment Portfolio

Two investments: passive are aggressive considered. The returns on investing each $1000 in the passive and aggressive investments are the random variables X and Y respectively. The detailed information of each investment is provided in the below table. Suppose 40% of the portfolio is in the passive investment and 60% is in the aggressive investment. Find the portfolio's expected return and risk (i.e., standard deviation).

Return and Risk on passive Investment (X)	$\mu_X = 50, \sigma_X = 43.3$
Return and Risk on aggressive Investment (Y)	$\mu_Y = 95, \sigma_Y = 193.21$
Covariance	$\text{cov}(X, Y) = 8250$

Solution

Return and Risk on passive Investment (X)	$\mu_X = 50, \sigma_X = 43.3$
Return and Risk on aggressive Investment (Y)	$\mu_Y = 95, \sigma_Y = 193.21$
Covariance	$\mathrm{cov}(X, Y) = 8250$

The random variable for the return of the investment portfolio is

$$P = 0.4X + 0.6Y.$$

The expectation and standard deviation of P is:

$$\mathrm{E}(P) = 0.4(50) + 0.6(95) = 77$$

$$\sigma_P = \sqrt{(0.4)^2(43.3)^2 + (0.6)^2(193.71)^2 + 2(0.4)(0.6)(8250)} = 133.30$$

The portfolio mean return and portfolio variability are between the values for the passive and aggressive investments considered individually.

Discrete Probability Distributions

Binomial Distribution

Consider n independent trials, each resulting in a "success" with probability p and in a "failure" with probability $1-p$. Let X be the number of successes in these n trials. Then X is called a binomial random variable with parameters (n, p).

The probability mass function (PMF) of this binomial random variable X is:

$$p(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \qquad x = 0,1,2, \dots, n,$$

Example

Suppose the probability of purchasing a defective computer is 0.2. What is the probability of purchasing 2 defective computers in a group of 4?

Solution

$$p(2) = \frac{4!}{2!(4-2)!}(0.2)^2(1-0.2)^{4-2} = 0.1536$$

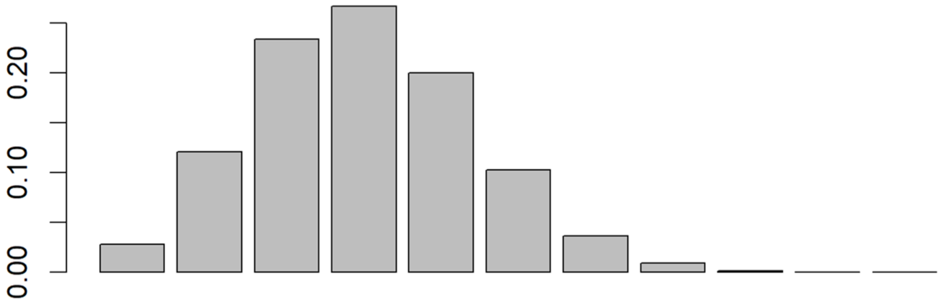
Binomial Distribution in R

$$\mathrm{P}(x) = \mathrm{P}(X = x) = \mathrm{dbinom(x, n, p)}, \qquad x = 0,1,2, \ldots, n$$

$$\mathrm{P}(X \leq x) = \sum_{j=0}^{x} p(j) = \mathrm{pbinom(x, n, p)} = \mathrm{sum(dbinom(c(0{:}\,x), n, p))}, \qquad x = 0,1,2, \ldots, n$$

n	10
p	0.3

x	P(x successes)
0	0.03
1	0.12
2	0.23
3	0.27
4	0.20
5	0.10
6	0.04
7	0.01
8	0.00
9	0.00
10	0.00




```
barplot(dbinom(c(0:10),10,0.3),col='grey')
```

Binomial Mean and Variance

If X is a binomial random variable with parameters (n, p) then:

$$\mu = \mathrm{E}(X) = np$$

$$\sigma^2 = \mathrm{var}(X) = np(1-p)$$

Example

The Nationwide Motel Company has conducted a study, revealing that 70 percent of all calls for motel reservations specifically request nonsmoking rooms. To gain further insights, the customer service manager randomly selected 25 calls for analysis. Calculate both the mean and standard deviation of requests for nonsmoking rooms.

Solution

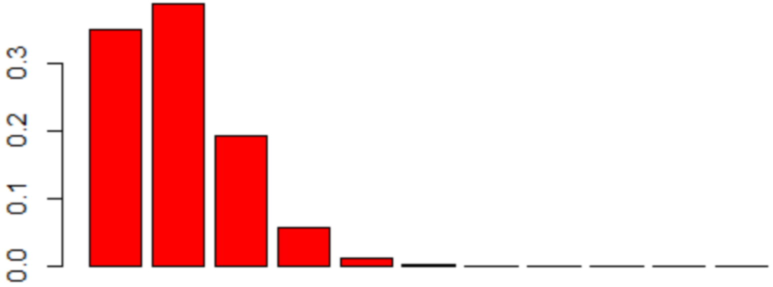
$$\mu = 25 \times 0.7 = 17.5$$

$$\sigma = \sqrt{25 \times 0.7 \times 0.3} = 2.29$$

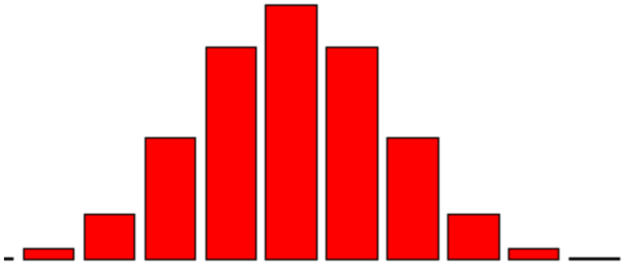
The Binomial Distribution Shape

The shape of the binomial distribution depends on the values of p and n

- $p > 0.5$ and $n < \infty$: the distribution is left-skewed
- $p = 0.5$ and any value of n: the distribution is symmetric
- $p < 0.5$ and $n < \infty$: the distribution is right-skewed
- If $n \to \infty$: the distribution becomes symmetric



$n = 10, p = 0.1$



$n = 10, p = 0.5$

Example

Which of the following statements is true?

A) A binomial distribution with $n = 20$ and $p = 0.05$ will be right-skewed.

B) A binomial distribution with $n = 6$ and $p = 0.5$ will be symmetric.

C) A binomial distribution with $n = 20$ and $p = 0.05$ has an expected value equal to 1.

D) A, B, and C are all true.

Poisson Distribution: Motivation

- What is the Poisson Distribution?
  - The Poisson distribution is used to model the number of times an event happens within a fixed interval of time or space, given that these events occur with a known average rate and independently of the time since the last event.
  - Let us use a scenario of customers arriving at a bank to motivate and build up to the Poisson distribution.
- Scenario Setup:
  - Suppose customers arrive at a bank with rate λ (say, 10 customers per hour).
  - We want to calculate the probability that x customers arrive between time 0 and time t.
- Dividing Time into Small Intervals:
  - To simplify the problem, imagine breaking up the time period t into many small intervals (Δt).
  - In each tiny interval, the probability of a customer arriving is very small:

$$p = \mathrm{P}(X = 1) = \lambda\Delta t + o(\Delta t), \qquad p(X \geq 2) = o(\Delta t)$$

  - As we increase the number of intervals to infinity (making each interval smaller and smaller), i.e., $n \to \infty$, $p \to \lambda\Delta t$, X becomes a random variable from binomial distribution with parameters (n, p):

Poisson Distribution: Motivation

$$p(x) = \lim_{n\to\infty} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$= \lim_{n\to\infty} \frac{n(n-1)\dots(n-1+x)}{x!} (\lambda\Delta t)^x (1-\lambda\Delta t)^{n-x} = \lim_{n\to\infty} \frac{n^x}{x!} \left(\frac{\lambda t}{n}\right)^x (1-\frac{\lambda t}{n})^{n-x} = \frac{(\lambda t)^x}{x!} e^{-\lambda t}$$

Noting that:

$$\lim_{n\to\infty} (1-\frac{\lambda t}{n})^{n-x} = \frac{\lim_{n\to\infty} (1-\frac{\lambda t}{n})^n}{\lim_{n\to\infty} (1-\frac{\lambda t}{n})^x} = \lim_{n\to\infty} (1-\frac{\lambda t}{n})^n = e^{-\lambda t}.$$

$$p(x) = \frac{(\lambda t)^x}{x!} e^{-\lambda t}$$

Poisson Distribution

A random variable X is a Poisson random variable with parameter $\mu = \lambda t$, if for some $\mu > 0$:

$$p(x) = \frac{e^{-\mu}\mu^x}{x!}, \qquad x = 0,1,2, \ldots,$$

where:

- μ: the average number of **events** in a **given interval** t
- λ: the average number of **events** in a **unit** of the interval (the rate of occurrence of the events)
- t: the **interval of time or space** over which events occur (e.g., a specific **time duration**, **area**, or **volume**)

Example:

- The number of scratches in a car's paint
- The number of computer crashes in a day
- The number of telephone calls one receives per day

Example

The number of customers who enter a bank is thought to be Poisson distributed with a mean equal to 10 per hour.

What are the chances that no customers will arrive in a 15-minute period?

Solution

$$\left.\begin{array}{l}\lambda = 10/\text{hour} \\ t = \dfrac{15}{60} = \frac{1}{4}\end{array}\right\} \Rightarrow \lambda \times t = \frac{10}{4} = 2.5$$ arrivals in a 15-minute period

$$p(0) = e^{-\lambda t} = \mathrm{e}^{-2.5} = 0.0821$$

Poisson Mean and Variance

If X is a Poisson random variable with parameter $\mu = \lambda t$ then

$$\mathrm{E}(X) = \mu = \lambda t$$

$$\mathrm{var}(X) = \mu = \lambda t$$

Example

If the standard deviation for a Poisson distribution is known to be 3, the expected value of that Poison distribution is:

A) 3.

B) about 1.73.

C) 9.

D) Can't be determined without more information.

Example

Answer: C

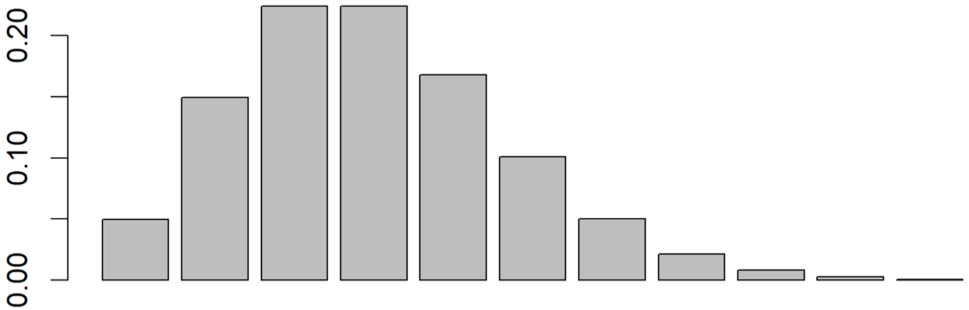
$\mu = \sigma^2 = 3^2 = 9.$

Poisson Distribution in R

$$p(x) = \mathrm{P}(X = x) = \mathrm{dpois}(\mathrm{x}, \mu), \qquad x = 0,1,2,\ldots$$

$$\mathrm{P}(X \leq x) = \sum_{j=0}^{x} p(j) = \mathrm{ppois}(\mathrm{x}, \mu) = \mathrm{sum}\big(\mathrm{dpois}(\mathrm{c}(0\!:\!\mathrm{x}), \mu)\big), \qquad x = 0,1,2,\ldots$$

x	P(x)
0	0.05
1	0.15
2	0.22
3	0.22
4	0.17
5	0.10
6	0.05
7	0.02
8	0.01
9	0.00
10	0.00




```
barplot(dpois(c(0:10),3),col='grey')
```