# M3S1 - Binomial Distribution

Professor Jarad Niemi

STAT 226 - Iowa State University

September 28, 2018

# Outline

- Random variables
  - Probability distribution function
  - Expectation (mean)
  - Variance
- Discrete random variables
  - Bernoulli
  - Binomial

# Probability

### Definition

A probability is a mathematical function, $P(E)$, that describes how likely an event $E$ is to occur. This function adheres to two basic rules:

1. $0 \le P(E) \le 1$
2. For mutually exclusive events $E_1, \ldots, E_K$,

$$P(E_1 \text{ or } E_2 \text{ or } \cdots \text{ or } E_K) = P(E_1) + P(E_2) + \cdots + P(E_K).$$

# Flipping a coin

Suppose we are flipping an unbiased coin that has two sides: heads ($H$) and tails ($T$). Then

$$P(H) = 0.5 \qquad P(T) = 0.5.$$

which adheres to rule 1) and

$$P(H \text{ or } T) = P(H) + P(T) = 0.5 + 0.5 = 1$$

which adheres to rule 2). So this is a valid probability.

# Rolling a 6-sided die

Suppose we are rolling an unbiased 6-sided die. If we count the number of pips on the upturned face, then the possible events are 1, 2, 3, 4, 5, and 6. Then

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$$

which adheres to 1). What is

$$P(1 \text{ or } 2 \text{ or } 3 \text{ or } 4 \text{ or } 5 \text{ or } 6) = 1.$$

To verify 2), we would need to calculate the probability of the $2^6$ possible colections of mutually exclusive events and find that their probability is the sum of the individual probabilities.

# Random variable

**Definition**

A random variable is the uncertain, numeric outcome of a random process. A discrete random variable takes on one of a list of possible values. A continuous random variable takes on any value in an interval.

A random variable is denoted by a capital letter, e.g. X or Y.

Discrete random variables:

- result of a coin flip
- the number of pips on the upturned face of a 6-sided die roll
- whether or not a company beats its earnings forecast
- the number of HR incidents next month

Continuous random variables:

- my height
- how far away a 6-sided die lands
- a company's next quarterly earnings
- a company's closing stock price tomorrow

# Probability distribution function

## Definition

A probability distribution function describes all possible outcomes for a random variable and the probability of those outcomes.

For example,

- Coin flipping:

$$P(H) = P(T) = 1.$$

- Unbiased 6-sided die rolling

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6.$$

- Company earnings compared to forecasts

$$\begin{aligned} P(\text{Earnings within 5\% of forecast}) &= 0.6 \\ P(\text{Earnings less than 5\% of forecast}) &= 0.1 \\ P(\text{Earnings greater than 5\% of forecast}) &= 0.3 \end{aligned}$$

# Events

**Definition**

An event is a set of possible outcomes of a random variable.

Discrete random variables:

- a coin flipping heads is heads
- the number of pips on the upturned face of a 6-sided die roll is less than 3
- a company beats its earnings forecast
- the number of HR incidents next month is less between 5 and 10

Continuous random variables:

- my height is greater than 6 feet
- how far away a 6-sided die lands is less than 3 feet
- a company's next quarterly earnings is within 5% of forecast
- a company's closing stock price tomorrow is less than today's

# Die rolling

Suppose we roll an unbiased 6-sided die. Determine the probabilities of the following events. The number of pips is

- exactly 3
- less than 3
- is greater than or equal to 3
- is odd
- is even and less than 5

# Bernoulli random variable

**Definition**

A Bernoulli random variable has two possible outcomes:

- 1 (success)
- 0 (failure)

A Bernoulli random variable is completey characterized by a single probability $p$, the probability of success (1). We write $X \sim Ber(p)$ to indicate that $X$ is a random variable that has a Bernoulli distribution with probability of success $p$. If $X \sim Ber(p)$, then we know $P(X = 1) = p$ and $P(X = 0) = 1 - p$.

Examples:

- a coin flip landing heads
- a 6-sided die landing on 1
- a 6-sided die landing on 1 or 2
- a company beating its earnings forecast
- a company's stock price closing higher tomorrow

# Coin flipping

Suppose we are flipping an unbiased coin and we let

$$X = \begin{cases} 0 & \text{if coin flip lands on tails} \\ 1 & \text{if coin flip lands on heads} \end{cases}$$

Then $X \sim Ber(0.5)$ which means $p = 0.5$ is the probability of success (heads) and $P(X = 1) = 0.5$ and $P(X = 0) = 0.5$.

# Die rolling

Suppose we are rolling an unbiased 6-sided die and we let

$$X = \begin{cases} 0 & \text{if die lands on 3, 4, 5, or 6} \\ 1 & \text{if die lands on 1 or 2} \end{cases}$$

Then $X \sim Ber(1/3)$ which means $p = 1/3$ is the probability of success (a 1 or 2) and $P(X = 1) = 1/3$ and $P(X = 0) = 2/3$.

# Mean of a random variable

**Definition**

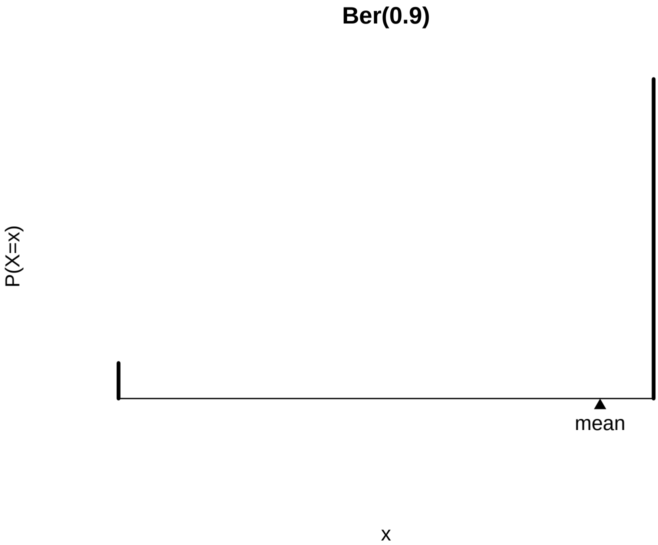
The mean of a random variable is a probability weighted average of the outcomes of that random variable. This mean is also called the expectation of the random variable and for a random variable $X$ is denoted $E[X]$ (or $E(X)$).

For a Bernoulli random variable $X \sim Ber(p)$, we have

$$E[X] = (1-p) \times 0 + p \times 1 = p.$$

The mean of a random variable is analogous to the physics concept of center of mass.

# Expectation is the "center of mass"



# Variance of a random variable

## Definition

The variance of a random variable is the probability-weighted average of the squared difference from the mean. The variance of a random variable $X$ is denoted $Var[X]$ (or $Var(X)$) and $Var[X] = E[(X-\mu)^2]$ where $\mu = E[X]$ is the mean. The standard deviation of a random variable is the square root of the variance of the random variable, i.e. $SE[X] = \sqrt{Var[X]}$.

For a Bernoulli random variable $X \sim Ber(p)$, we have

$$\begin{aligned} Var[X] &= (1-p)\times(0-p)^2 + p\times(1-p)^2 \\ &= (1-p)\times p^2 + p\times(1-2p+p^2) \\ &= p^2 - p^3 + p - 2p^2 + p^3) \\ &= p - p^2 \\ &= p(1-p). \end{aligned}$$

Variance is analogous to the physics concept of moment of inertia.

# Coin flipping

If $X \sim Ber(0.5)$, then

- $E[X] = 1/2$
- $Var[X] = 1/2 \times (1 - 1/2) = 1/2 \times 1/2 = 1/4.$

If $X \sim Ber(1/3)$, then

- $E[X] = 1/3$
- $Var[X] = 1/3 \times (1 - 1/3) = 1/3 \times 2/3 = 2/9.$

If $X \sim Ber(2/9)$, then

- $E[X] = 2/9$
- $Var[X] = 2/9 \times (1 - 2/9) = 2/9 \times 7/9 = 14/81.$

# Die rolling

Let $X$ be the number of pips on the upturned face of an unbiased 6-sided die. Find the probability distribution function, the expected value (mean), and the variance.

Then the probability distribution function is

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

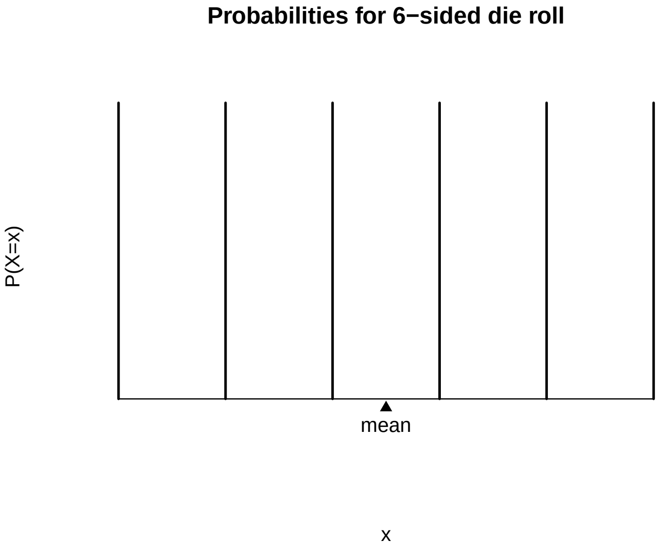
The expected value, $E[X]$, is

$$\begin{aligned} E[X] &= 1/6 \times 1 + 1/6 \times 2 + 1/6 \times 3 + 1/6 \times 4 + 1/6 \times 5 + 1/6 \times 6 \\ &= 3.5. \end{aligned}$$

The variance, $Var[X]$, is

$$\begin{aligned} Var[X] &= 1/6 \times (1 - 3.5)^2 + 1/6 \times (2 - 3.5)^2 + 1/6 \times (3 - 3.5)^2 \\ &\quad +1/6 \times (4 - 3.5)^2 + 1/6 \times (5 - 3.5)^2 + 1/6 \times (6 - 3.5)^2 \\ &= 2.91\overline{6}. \end{aligned}$$

# Expectation is the “center of mass”



# Independence

**Definition**

Two random variables are independent if the outcome of one random variable does not affect the probabilities of the outcomes of the other random variable.

For independent random variables $X$ and $Y$ and constants $a$, $b$, and $c$, we have the following properties

$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

and

$$Var[aX + bY + c] = a^2 Var[X] + b^2 Var[Y].$$

# Sum of independent Bernoulli random variables

Let $X_i$, for $i = 1, \ldots, n$ be independent Bernoulli random variable with a common probability of success $p$. We write

$$X_i \stackrel{ind}{\sim} Ber(p).$$

Then the sum

$$Y = \sum_{i=1}^{n} X_i$$

is a binomial random variable.

# Binomial

### Definition

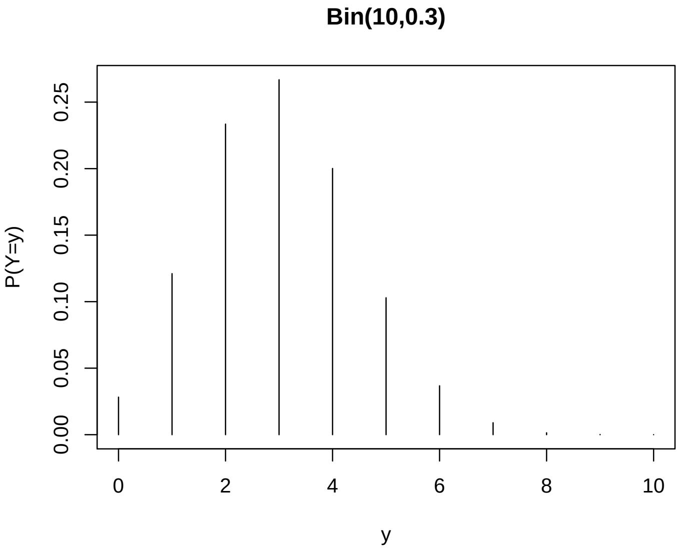
A binomial random variable with $n$ attempts and probability of success $p$ has a probability distribution function

$$P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}$$

for $0 \leq p \leq 1$ and $y = 0, 1, \ldots, n$ where

$$\binom{n}{y} = \frac{n!}{(n-y)!y!}.$$

We write $Y \sim Bin(n, p)$.


Bin(10,0.3)
P(Y=y)
0.00
0.05
0.10
0.15
0.20
0.25
0
2
4
6
8
10
y

# Binomial expected value and variance

The expected value (mean) is

$$\begin{aligned} E[Y] &= E[X_1 + X_2 + \cdots + X_n] \\ &= E[X_1] + E[X_2] + \cdots + E[X_n] \\ &= p + p + \cdots + p \\ &= np. \end{aligned}$$

The variance is

$$\begin{aligned} Var[Y] &= Var[X_1 + X_2 + \cdots + X_n] \\ &= Var[X_1] + Var[X_2] + \cdots + Var[X_n] \\ &= p(1-p) + p(1-p) + \cdots + p(1-p) \\ &= np(1-p). \end{aligned}$$

# Examples

If $Y \sim Bin(10, .3)$, then

$$E[Y] = 10 \times 0.3 = 3$$

and

$$Var[Y] = 10 \times 0.3 \times (1 - 0.3) = 10 \times 0.3 \times 0.7 = 2.1.$$

If $Y \sim Bin(65, 1/4)$, then

$$E[Y] = 65 \times 1/4 = 16.25$$

and

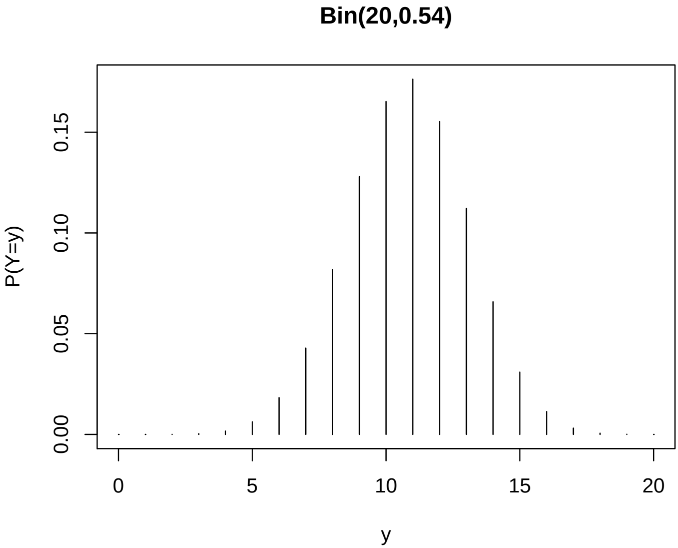
$$Var[Y] = 65 \times 1/4 \times (1 - 1/4) = 65 \times 1/4 \times 3/4 = 12.1875.$$

# AVP Example

In the 2018 AVP Gold Series Championships in Chicago, IL, Alex Klineman and April Ross beat Sara Hughes and Summer Ross in 2 sets with scores 25-23, 21-16. Suppose that these scores actually determine the probability that Klineman/Ross will score a point against Hughes/Ross, i.e. $p = (25 + 21)/(25 + 23 + 21 + 16) = 0.54$ and that each point is independent.

Let $Y$ be the number of points Klineman/Ross will win (against Hughes/Ross) over the next 20 points. Based on our assumptions $Y \sim Bin(20, 0.54)$.

# AVP Example (cont.)



# AVP Example (cont.)

Here are some questions we can answer:

- How many points do we expect Klineman/Ross to score?

$$E[Y] = 20 \times .54 = 10.8 \text{ points}$$

- What is the variance around this number?

$$Var[Y] = 20 \times .54 \times (1 - .54) = 4.966 \text{ points}^2$$

- What is the standard deviation around this number?

$$SD[Y] = \sqrt{Var[Y]} = \sqrt{4.966} = 2.23 \text{ points}$$

- What is the probability that Klineman/Ross will win at least 10 points?

$$P(Y >= 10) = P(Y = 10) + P(Y = 11) + \cdots + P(Y = 20) = 0.72$$

# AVP Example (cont.)