

# STAT 111

## Recitation 2

Mo Huang

Email: [mohuang@wharton.upenn.edu](mailto:mohuang@wharton.upenn.edu)

Office Hours: Wednesdays 3:00 - 4:00 pm, JMHH F96

Slides (adapted from Gemma Moran): [github.com/mohuangx/STAT111-Fall2018](https://github.com/mohuangx/STAT111-Fall2018)

September 14, 2018

# Parameters ( $\theta$ ), Random Variables ( $X$ ), and Data ( $x$ )

- ▶ A **parameter** represents some underlying numerical constant of a phenomenon. Represented by Greek letters.
- ▶ A **random variable** is a numerical outcome of interest in a future experiment. Can be modeled by a probability **distribution** which depends on the parameter. Represented by capital letters.
- ▶ **Data** is the realization or observed value of a random variable after performing the experiment. Represented by lower-case letters.

Fair coin-flipping example:

- ▶ A **parameter**  $\theta$  is the underlying probability of obtaining a head.  $\theta = 0.5$ .
- ▶ A **random variable**  $X$  is the number of heads obtained in 10 tosses. Can be modeled by a binomial distribution dependent on  $\theta$ .
- ▶ **Data**  $x = 6$  is observing 6 heads in 10 tosses, a realization of  $X$ .

# Random Variables

- ▶ Two types of random variables: **discrete** and continuous.
- ▶ A **discrete random variable** is a random variable that can only take on a countable set of numbers.
- ▶ The **probability distribution** of a discrete random variable is the range of values it can take *and* the probabilities of these values.
- ▶ For example:
  - ▶ Let  $X$  be the number of heads I toss if I toss a fair coin 3 times.

$x$	0	1	2	3
$P(X = x)$	0.125	0.375	0.375	0.125

**Table:** Probability distribution of  $X$  using the tableau method.

# The Binomial Distribution

- ▶ The binomial distribution arises if:
  1. We plan to conduct a fixed number of experiments. We denote the number of experiments as  $n$ .
  2. In each experiment, there are two outcomes: “success” or “failure”.
  3. The experiments are independent.
  4. The probability of a success is the same for each experiment.
  
- ▶ For example:
  1. I plan to toss a coin  $n$  times.
  2. I can toss either a head or a tail.
  3. Each coin toss is independent of the next - it doesn't matter whether I get a head or a tail on the previous toss.
  4. The probability of getting a head is the same for each toss.

# The Binomial Distribution

- ▶ Let  $X$  be a binomial random variable where  $\theta = P(\text{success})$  and there are  $n$  experiments. Then the probability distribution of  $X$  is given by:

$$P(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad \text{for } x = 0, 1, 2, \dots, n.$$

- ▶  $\theta$  is called a **parameter**: a constant whose value may be known or unknown.
- ▶  $\binom{n}{x} = \frac{n!}{x!(n-x)!}$  is said as “ $n$  choose  $x$ ”: it is the number of ways  $x$  successes can occur in  $n$  experiments.

**Note:**  $X \sim \mathcal{B}(n, \theta)$  means “ $X$  is a binomial random variable with  $n$  experiments and probability of success  $\theta$ .”

# The Binomial Distribution: Questions

**Q1:** Let  $X$  be the number of heads if I toss an unbiased coin 3 times ( $n = 3, \theta = 0.5$ ). Find the probability distribution of  $X$  in “tableau” form.

**A1:**

$$\begin{aligned}P(X = 0) &= \binom{3}{0} (0.5)^0 (0.5)^3 \\ &= 0.125\end{aligned}$$

$$\begin{aligned}P(X = 2) &= \binom{3}{2} (0.5)^2 (0.5)^1 \\ &= 0.375\end{aligned}$$

$$\begin{aligned}P(X = 1) &= \binom{3}{1} (0.5)^1 (0.5)^2 \\ &= 0.375\end{aligned}$$

$$\begin{aligned}P(X = 3) &= \binom{3}{3} (0.5)^3 (0.5)^0 \\ &= 0.125\end{aligned}$$

$x$	0	1	2	3
$P(X = x)$	0.125	0.375	0.375	0.125

**Table:** Probability distribution of  $X$  using the tableau method.

## The Binomial Distribution: Tables

Q2:  $X \sim \mathcal{B}(18, 0.15)$ . Find  $P(X = 4)$ .

A2:  $P(X = 4) = 0.1592$

		$\theta$				
$n$	$i$	0.05	0.10	0.15	0.20	0.25
18	0	0.3972	0.1501	0.0536	0.0180	0.0056
	1	0.3763	0.3002	0.1704	0.0811	0.0338
	2	0.1683	0.2835	0.2556	0.1723	0.0958
	3	0.0473	0.1680	0.2406	0.2297	0.1704
	4	0.0093	0.0700	0.1592	0.2153	0.2130

Q3: Find  $P(X \leq 4)$ .

A3:

$$\begin{aligned}P(X \leq 4) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) \\&= 0.0536 + 0.1704 + 0.2556 + 0.2406 + 0.1592 \\&= 0.8794\end{aligned}$$

# The Binomial Distribution: Tables

Q2:  $X \sim \mathcal{B}(12, 0.8)$ . Find  $P(X = 3)$ .

A2: Finding 3 successes with  $\theta = 0.8$  is the same as finding 9 failures with  $\theta$  of failure 0.2. Hence,  $P(X = 3) = 0.0001$ .

$\theta$

$n$	$i$	0.05	0.10	0.15	0.20	0.25	0.30	0.35
12	0	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057
	1	0.3413	0.3766	0.3012	0.2062	0.1267	0.0712	0.0368
	2	0.0988	0.2301	0.2924	0.2835	0.2323	0.1678	0.1088
	3	0.0173	0.0852	0.1720	0.2362	0.2581	0.2397	0.1954
	4	0.0021	0.0213	0.0683	0.1329	0.1936	0.2311	0.2367
	5	0.0002	0.0038	0.0193	0.0532	0.1032	0.1585	0.2039
	6	0.0000	0.0005	0.0040	0.0155	0.0401	0.0792	0.1281
	7	0.0000	0.0000	0.0006	0.0033	0.0115	0.0291	0.0591
	8	0.0000	0.0000	0.0001	0.0005	0.0024	0.0078	0.0199
	9	0.0000	0.0000	0.0000	0.0001	0.0004	0.0015	0.0048
	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0008
	11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
	12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000



# Random Variables: Mean

- ▶ The **mean of a random variable** (expected value) is the long-run average of realizations of the random variable over repeated experiments.
- ▶ This mean of random variable  $X$  is different from the sample mean, which is the average of a *finite* number of observations ( $x$ ).
- ▶ Let  $X$  be a discrete random variable that can take values  $\{v_1, v_2, \dots, v_k\}$ . Then the mean of  $X$  is given by:

$$\begin{aligned}\mu &= \sum_{i=1}^k v_i P(X = v_i) \\ &= v_1 P(X = v_1) + v_2 P(X = v_2) + \dots + v_k P(X = v_k).\end{aligned}$$

**Note:** We use  $\mu$  to denote the mean of a random variable.

**Note:** Can think of it as a weighted average, weighted by probability.

## Random Variables: Questions

Q2: Let  $X$  be the outcome of one roll of a biased dice where the probability of the number  $j$  turning up is  $j/21$ . Find the mean of  $X$ .

A2:

$$\begin{aligned}\mu &= 1 \times 1/21 + 2 \times 2/21 + 3 \times 3/21 + 4 \times 4/21 + 5 \times 5/21 + 6 \times 6/21 \\ &= 91/21\end{aligned}$$

Q3: Let  $X$  be a binomial random variable with  $n = 2$  and  $\theta = 0.8$ . Find the mean of  $X$  using the binomial table.

A3:

$$\begin{aligned}\mu &= 0 \times 0.04 + 1 \times 0.32 + 2 \times 0.64 \\ &= 1.6\end{aligned}$$

# The Binomial Distribution: Mean

- ▶ For the binomial distribution, we have a simpler formula for the mean: if  $X \sim \mathcal{B}(n, \theta)$ , then the mean of  $X$  is

$$\mu = n\theta.$$

- ▶ For example:
  - ▶ On the previous slide, we calculated the mean of a binomial random variable with  $n = 2$  and  $\theta = 0.8$  to be 1.6. This is exactly  $n\theta$ .

**Note:** This formula is *only* for the binomial distribution.

## Random Variables: Variance

- ▶ The **variance of a random variable** is a measure of the *spread* of a distribution - that is, how far away values are from the mean.
- ▶ Let  $X$  be a discrete random variable that can take values  $\{v_1, v_2, \dots, v_k\}$ . Then the variance of  $X$  is given by:

$$\sigma^2 = (v_1 - \mu)^2 P(X = v_1) + \dots + (v_k - \mu)^2 P(X = v_k) \quad (1)$$

or

$$\sigma^2 = v_1^2 P(X = v_1) + \dots + v_k^2 P(X = v_k) - \mu^2 \quad (2)$$

- ▶ The **standard deviation of a random variable** is the square root of the variance and is denoted by  $\sigma$ .
- ▶ For a **binomial** random variable  $X \sim \mathcal{B}(n, \theta)$ :

$$\sigma^2 = n\theta(1 - \theta)$$

**Note:** The variance is always positive - you can't have a negative spread of a distribution.