

# S1211Q Introduction to Statistics

## Lecture 7

Wei Wang

July 11, 2012

# Understanding Random Variables

- ▶ We typically write a RV in capital letter, such as  $X$ , and the value it takes in lower case letter, such as  $x$ . So a typical probability statement looks like  $P(X = x)$ .

# Understanding Random Variables

- ▶ We typically write a RV in capital letter, such as  $X$ , and the value it takes in lower case letter, such as  $x$ . So a typical probability statement looks like  $P(X = x)$ .
- ▶ Also, since probability is always associated with some event, what do we mean when we write  $P(X = x)$ ?

# Understanding Random Variables

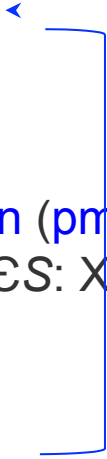
- ▶ We typically write a RV in capital letter, such as  $X$ , and the value it takes in lower case letter, such as  $x$ . So a typical probability statement looks like  $P(X = x)$ .
- ▶ Also, since probability is always associated with some event, what do we mean when we write  $P(X = x)$ ?
- ▶ By the definition of RV,  $X = x$  corresponds to an event that consists of outcomes which is associated with the value  $x$  by random variable  $X$ .

$$P(X = x) = P(\{s \in S : X(s) = x\})$$

# PMF

- The probability model for a discrete random variable  $X$ , lists its possible values and their probabilities.

Value of $X$	$x_1$	$x_2$	.....	$x_k$
Probability	$p_1$	$p_2$	.....	$p_k$

- Every probability,  $p_i$ , is a number between 0 and 1.
  - $p_1 + p_2 + \dots + p_k = 1$
  - The probability distribution or probability mass function (pmf) of a discrete rv is defined for every number  $x$  by  $p(x) = P(X=x) = P(\text{all } s \in S: X(s)=x)$ .
  - How to check if some function  $p(x)$  is a proper PMF?
- 

# Bernoulli RV

- ▶ The arguably simplest probability model is Bernoulli. Any random variable whose possible values are only 0 and 1 is called a Bernoulli random variable.
- ▶ Ex. Flip a coin.  $S = \{H, T\}$ . We can define a Bernoulli random variable,  $X(H) = 1, X(T) = 0$ . Then the distribution of  $X$  is

$$P(X = 1) = .5, P(X = 0) = .5$$

- ▶ Ex. Roll a die.  $S = \{1, 2, 3, 4, 5, 6\}$ . We can define a bernoulli random variable,  $X(1) = X(2) = \mathbf{1}, X(3) = X(4) = X(5) = X(6) = \mathbf{0}$ . Then the distribution is

$$P(X = \mathbf{1}) = 1/3, P(X = \mathbf{0}) = 2/3$$

# Example

Ex. Flip three fair coins. (*Binomial*)

$S = \{\text{HHH, HHT, HTH, HTT, THT, THH, TTH, TTT}\}$ . Let's define random variable  $X$  to be the number of heads in the experiment, i.e.,  $X(\text{HHH})=3$ ,  $X(\text{THT})=1$ , etc.

$X$

0 TTT

1 TTH THT HTT

2 THH HTH HHT

3 HHH

Value of $X$	0	1	2	3
Probability	0.125	0.375	0.375	0.125

One can calculate the probability of an event by adding the probabilities  $p_i$  of the particular values of  $x_i$  that make up the event. For example, if we want to know the probability of getting less than 2 heads, we can use

$$P(X < 2) = P(X=0) + P(X=1) = 0.125 + 0.375 = 0.5$$

$$\text{Note: } P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = 0.875$$

# CDF

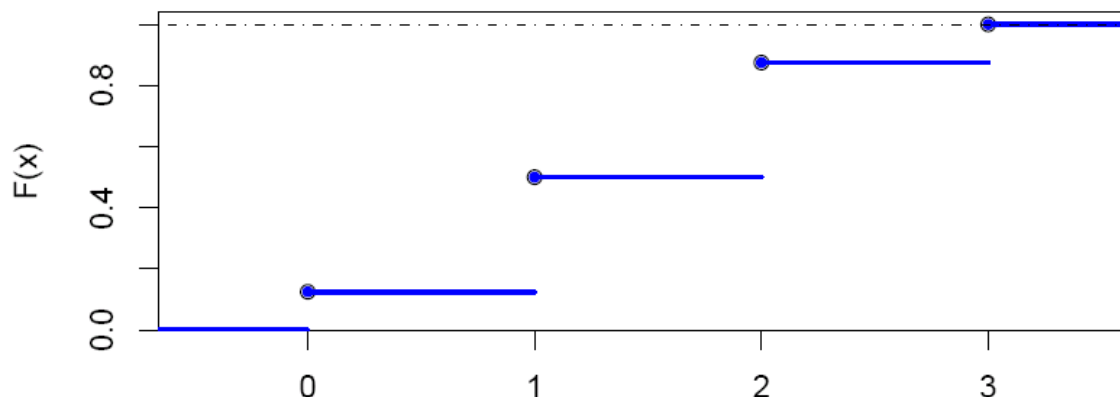
- The **cumulative distribution function** (cdf)  $F(x)$  of a discrete rv variable  $X$  with pmf  $p(x)$  is defined for every number  $x$  by

$$F(x) = P(X \leq x) = \sum_{y: y \leq x} p(y).$$

For any number  $x$ ,  $F(x)$  is the probability that the observed value of  $X$  will be at most  $x$ .

- For  $X$  a discrete rv, the graph of  $F(x)$  will have a jump at every possible value of  $X$  and will be flat between possible values. Such a graph is called a **step function**.

**The three coin flips example**





# Parameter and Family

- Suppose  $p(x)$  depends on a quantity that can be assigned any one of a number of possible values, with each different value determining a different probability distribution. Such a quantity is called a **parameter** of the distribution. The collection of all probability distributions for different values of the parameter is called a **family** of probability distributions.

Ex. For Bernoulli rv's, the parameter is the probability of being 1 (or 0), that is,

$$p = P(X=1)$$

# Expectation and Variance

- Random variables have distributions, so they have centers and spreads.
- The **expected value** (**mean value** or **expectation**) of a random variable describes its **theoretical long-run average value**.
- We typically use  $\mu$  or  $E(X)$  to denote the mean,  $\text{Var}(X)$  to denote the variance and  $\sigma$  or  $\text{SD}(X)$  to denote the standard deviation of a rv  $X$ .

# Motivating examples

Ex. How many heads would you expect if you flipped a fair coin twice?

$S = \{\text{HH}, \text{HT}, \text{TH}, \text{TT}\}.$

$X =$  number of heads.

0    TT

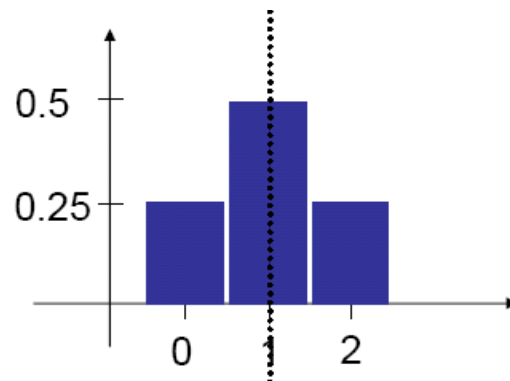
1    HT TH

2    HH

$p(X=0) = 0.25; p(X=1) = 0.5; p(X=2) = 0.25.$

Each outcome is weighted by its probability.

$$\mu = 0 \times 0.25 + 1 \times 0.5 + 2 \times 0.25 = 1$$



# Example

Ex. How many heads would you expect if you flipped a coin three times?

$$\mu = 0 \times 0.125 + 1 \times 0.375 + 2 \times 0.375 + 3 \times 0.125 = 1.5$$

This can never occur in a single trial of 3 flips. However, **on average** we would expect to get 1.5 heads if we repeated the experiment many times.

# Definition

- Suppose  $X$  is a discrete random variable whose probability model is given by

Value of $X$	$x_1$	$x_2$	.....	$x_k$
Probability	$p_1$	$p_2$	.....	$p_k$

The expected value of  $X$  is given by

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x) = x_1 p_1 + x_2 p_2 + \cdots x_k p_k$$

# Example

Ex. Expectation of a Bernoulli rv.

$$p(x) = \begin{cases} 1 - p & x = 0 \\ p & x = 1 \\ 0 & x \neq 0, 1 \end{cases}$$

$$\mu = 0 \times (1-p) + 1 \times p = p.$$

# The Expected Value of a Function

- ▶ A bookstore purchases ten copies of a books at \$60 each to sell at \$120, and any unsold copies after three months can be redeemed for \$20. If the number of copies sold is  $X$ , what is the profit of the bookstore?

# The Expected Value of a Function

- ▶ A bookstore purchases ten copies of a books at \$60 each to sell at \$120, and any unsold copies after three months can be redeemed for \$20. If the number of copies sold is  $X$ , what is the profit of the bookstore?
- ▶ The profit is  $h(X) = 100X - 400$ . Is  $h(X)$  a Random Variable? Then what is the expectation of  $h(X)$ ?



# The Expected Value of a Function

- ▶ A bookstore purchases ten copies of a books at \$60 each to sell at \$120, and any unsold copies after three months can be redeemed for \$20. If the number of copies sold is  $X$ , what is the profit of the bookstore?
- ▶ The profit is  $h(X) = 100X - 400$ . Is  $h(X)$  a Random Variable? Then what is the expectation of  $h(X)$ ?
- ▶ If Random Variable  $X$  has range  $D$  and pmf  $p(x)$ , then the expected value of function  $h(X)$  is given by

$$E(h(X)) = \sum_{x \in D} (h(x) \cdot p(x))$$

# The Linear Function Case

- ▶ In the case of linear function, we have a much more convenient formula

$$E(aX + b) = a \cdot E(X) + b$$

# The Linear Function Case

- ▶ In the case of linear function, we have a much more convenient formula

$$E(aX + b) = a \cdot E(X) + b$$

- ▶ To prove,

$$E(aX + b) = \sum_{x \in D} (ax + b) \cdot p(x) = a \sum_{x \in D} x \cdot p(x) + b \sum_{x \in D} p(x) = aE(X) + b$$

# Difference Between Expectation and Measure of Center

- ▶ What's the difference between  $E(X)$  (expectation/population mean) and  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  (sample mean)?

# Difference Between Expectation and Measure of Center

- ▶ What's the difference between  $E(X)$  (expectation/population mean) and  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  (sample mean)?
- ▶ Actually, they live in different worlds. Expected value of a RV  $X$  is the *true* mean in the ideal world, while mean of a sample  $x_1, x_2, x_3, \dots$  is the *observed* mean in the empirical world.

# Difference Between Expectation and Measure of Center

- ▶ What's the difference between  $E(X)$  (expectation/population mean) and  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  (sample mean)?
- ▶ Actually, they live in different worlds. Expected value of a RV  $X$  is the *true* mean in the ideal world, while mean of a sample  $x_1, x_2, x_3, \dots$  is the *observed* mean in the empirical world.
- ▶ In another word, expectation is a concept in probability, and sample mean is a concept in statistics. In inference, we use  $\bar{x}$  to infer  $E(X)$

# Variance of Random Variables

- ▶ We defined the concept of sample variance in the first chapter. Similarly we can define the variance of random variables, which is still a measure of deviation from the center.
- ▶  $X$  has pmf  $p(x)$  and expected value  $\mu$ , then the variance of  $X$ , denoted as  $V(X)$  or  $\sigma^2$  is

$$V(X) = E[(X - \mu)^2]$$

The standard deviation of  $X$  is

$$\sigma = \sqrt{\sigma^2}$$

# Example

- ▶ A reader can check out at most 6 videos from a library at one time. Consider only those who check out videos, let  $X$  denote the number of videos checked out to a randomly selected individual. The pmf is

$x$	1	2	3	4	5	6
$p(x)$	.3	.25	.15	.05	.1	.15

- ▶ How to calculate variance and standard deviation of  $X$ ?



# Shortcut Formula for $\sigma^2$

- ▶ Similar to sample variance, we have a shortcut formula for variance of random variables too

▶

$$V(X) = \sigma^2 = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$$

- ▶ Proof:

$$\sigma^2 = \sum_D x^2 \cdot p(x) - 2\mu \cdot \sum_D x \cdot p(x) + \mu^2 \sum_D p(x) = E(X^2) - \mu^2$$

# Variance of Linear Function of Random Variables

- ▶ How would variance change if we add a constant to the RV?

# Variance of Linear Function of Random Variables

- ▶ How would variance change if we add a constant to the RV?
- ▶ How would variance change if we multiply it by a constant?

# Variance of Linear Function of Random Variables

- ▶ How would variance change if we add a constant to the RV?
- ▶ How would variance change if we multiply it by a constant?

▶

$$V(aX + b) = a^2 \cdot V(X)$$

and

$$\sigma_{aX+b} = |a| \cdot \sigma_X$$

# Discrete Probability Models

- We often conduct trials/experiments repeatedly. Today we will discuss probability models that allow us to answer questions regarding repeated trials. In particular we will be interested in series of  $n$  repeated trials of a random phenomena with two possible outcomes.
- Many popular discrete models are motivated by coin tosses, or more specifically a series of  $n$  Bernoulli trials. A series of  $n$  trials are Bernoulli trials if:
  1. The  $n$  trials are identical.
  2. The trials are independent (the outcome on any particular trial does not influence the outcome on any other trial).
  3. Each trial has two possible outcomes: success or failure.
  4. The probability of success, denoted by  $p$ , is the same for each trial (identical).

# Binomial Experiment

- If in Bernoulli trials, the number of trials  $n$  is **fixed** in advance of the experiment. This experiment is called a binomial experiment.

Ex. The same coin is tossed successively and independently 10 times.

Ex. Suppose there are 50 colored socks in the drawer, of which 16 are red and the other 34 are blue. We are going to randomly draw 10 socks out of the drawer **without replacement**. We label the  $i$ th trial as a success if the  $i$ th sock is blue. (Is this a binomial experiment? What if it's **with replacement**?)

Ex. The previous example, what if we have 500,000 socks, of which 400,000 are blue. A sample of 10 socks are drawn **without replacement**.

$$P(\text{success on 2} \mid \text{success on 1}) = 399,999/499,999 = .80000$$

$$P(\text{success on 10} \mid \text{success on first 9}) = 399,991/499,991 = .799996 \approx .80000$$

# A Rule of Thumb

- For drawing **without** replacement (*hypergeometric*), as the previous example suggests, although the trials are not exactly independent, the conditional probability differ so slightly from one another that for practical purposes the trials can be regarded as independent. Thus, to a very good approximation, the previous experiment is binomial with  $n = 10$  and  $p = .8$ .
- As a **rule of thumb**: consider sampling without replacement from a dichotomous population of size  $N$ . If the sample size (number of trials)  $n$  is at most 5% of the population size, the experiment can be analyzed as though it were exactly a binomial experiment.

# Binomial RV

- The **binomial random variable**  $X$  associated with a binomial experiment consisting of  $n$  trials is defined as

$X$  = the number of successes among the  $n$  trials.

- The pmf of a binomial rv  $X$  depends on the two parameters  $n$  and  $p$ , we denote the pmf by  $b(x; n, p)$ . The cdf will be denoted by

$$P(X \leq x) = B(x; n, p) = \sum_{y=0}^x b(y; n, p).$$

Note that  $x$  can only take values in  $\{0, 1, \dots, n\}$ .



# Example

Ex. Roll a ten-sided die four times. What is the probability of getting exactly one three?

S = rolling a three.

F = rolling something other than a three.

$P(S) = p = 0.1$  and  $P(F) = 1-p = 0.9$

Let  $X$  = the number of threes, then  $X$  is  $\text{Bin}(4, 0.1)$  and we want to calculate  $P(X=1)$ . There are four possible ways of rolling a three: **SFFF**, **FSFF**, **FFSF**, **FFFS**.

$$P(\text{SFFF}) = P(S)P(F)P(F)P(F) = (1-p)^3p = (.9)^3(.1) = 0.0729$$

Similarly,  $P(\text{FSFF}) = P(\text{FFSF}) = P(\text{FFFS}) = 0.0729$ .

$$\begin{aligned} P(X=1) &= P(\text{SFFF}) + P(\text{FSFF}) + P(\text{FFSF}) + P(\text{FFFS}) \\ &= 4(0.0729) = 0.2916 \end{aligned}$$

# Binomial pmf

- From the previous example, we see that

$$P(X=1) = b(1; 4, p) = 4(1-p)^3p$$

$$= \{\text{\# of outcomes with } X=1\} \cdot \{\text{prob. of any particular outcome with } X=1\}$$

- Thus more generally, we have

$$b(x; n, p) = \{\text{\# of outcomes with } X=x\} \cdot \{\text{prob. of any particular outcome with } X=x\}$$

- The pmf of a binomial rv is

$$b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, 3, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

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

# Example

Ex. (Ten-sided die cont.) Use binomial pmf to verify  $P(X=1)$  we have calculated.

$$P(X = 1) = \binom{4}{1} (1/10)^1 (9/10)^3 = \frac{4!}{1!3!} (1/10)^1 (9/10)^3 = 0.2916$$

What is the probability of getting less than two 3's in four rolls?

$$\begin{aligned} P(X < 2) &= P(X = 0) + P(X = 1) \\ &= \binom{4}{0} (1/10)^0 (9/10)^4 + \binom{4}{1} (1/10)^1 (9/10)^3 \\ &= 0.6561 + 0.2916 = 0.9477 \end{aligned}$$

- Try using `dbinom()` ; `pbinom()` to calculate the things above.

# Example

Ex. Suppose we are searching for new apartments in the city, and our goal is to find an apartment among the top 5% (based on some criteria). Our strategy is to randomly sample 20 apartments from the pool, and choose the best out of these 20. What is the probability that we will accomplish our goal?

# Mean and Variance of Binomial

- Proposition:

If  $X \sim \text{Bin}(n, p)$ , then  $E(X) = np$ ,  $\text{Var}(X) = np(1 - p) = npq$ , and  $\sigma_X = \sqrt{npq}$  (where  $q = 1 - p$ ).

We'll show an easy proof in chapter 5.