Lecture 4: Probability

Jacob M. Montgomery

Quantitative Political Methodology

Lecture 4

CLass business

- PROBLEM SET 1 IS DUE RIGHT NOW
- Problem set 2 will be distributed today via the syllabus

Facebook and survey

- Sign up for our Facebook group: `https://www.facebook.com/groups/1071702902960687/`
- Take the class survey! Can't assign teams until you all do.

`https://wustl.az1.qualtrics.com/jfe/form/SV_6rpSYD3xxmbRe5v`

Roadmap

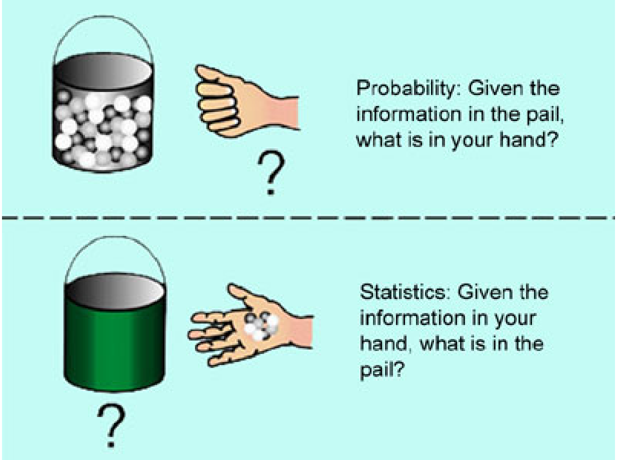
Last time:

- Visualizing data
- Measures of central tendency and spread

This time:

- Understand core concepts of probability
- Understanding concept of a "parameter"
- Introduce some probability distributions

Why are we studying this?



Probability defined

Imagine tossing a coin...

- Can you predict the outcome of a single coin toss?

Probability defined

Imagine tossing a coin...

- Can you predict the outcome of a single coin toss?
- Can you predict the *overall* outcome of 100 coin tosses?

Probability defined

Imagine tossing a coin. . .

- Can you predict the outcome of a single coin toss?
- Can you predict the *overall* outcome of 100 coin tosses?

AF p. 73: "For a particular possible outcome for a random phenomenon, the probability of that outcome is the proportion of times that the outcome would occur in a very long sequence of observations."

Example

Imagine you were rolling two six-sided dice.

1. Write down all possible scores.
2. Calculate the probability of each score
    - What is the probability of rolling a 2?

36 possible outcomes for the two dice:

1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

How many outcomes will generate a total score of 2?

1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

$\text{P(roll=2)} = \frac{1}{36} = 0.028.$

Putting this all togetheer

y_k	$Pr(Y = y_k)$
2	$1/36$
3	$2/36$

Putting this all togetheer

y_k	$Pr(Y = y_k)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

More formal definition

Probability *is the relative frequency of occurrence for some particular outcome if a process is repeated a large number of times under similar conditions*

More formal definition

Probability *is the relative frequency of occurrence for some particular outcome if a process is repeated a large number of times under similar conditions*

- If I flip a coin three times, what is the probability that I will get exactly two heads?
- If I roll two dice, what is the probability of getting a two?
- If I take a random sample of 100 Wash U students, what is the probability that less than 40% of the sample will be male?

Frequency distributions

***Frequency** distribution (discrete probability distribution)}*
is a probability distribution of a discrete variable Y assigns a probability to each possible outcome.

Frequency distributions

Frequency *distribution (discrete probability distribution)}*
is a probability distribution of a discrete variable Y assigns a probability to each possible outcome.

- Let $S = \{y_1, y_2, \ldots, y_k\}$ be the set of all possible outcomes,

Frequency distributions

***Frequency** distribution (discrete probability distribution)}*
is a probability distribution of a discrete variable Y assigns a probability to each possible outcome.

- Let $S = \{y_1, y_2, \ldots, y_k\}$ be the set of all possible outcomes, and Y be the realization of the variable.

Frequency distributions

***Frequency** distribution (discrete probability distribution)}*
is a probability distribution of a discrete variable Y assigns a probability to each possible outcome.

- Let $S = \{y_1, y_2, \ldots, y_k\}$ be the set of all possible outcomes, and Y be the realization of the variable.

- Then, $p(y_k) = Pr(Y = y_k)$, where
- $0 \leq p(y_k) \leq 1 \ \forall \ k$

Frequency distributions

Frequency *distribution (discrete probability distribution)}*
is a probability distribution of a discrete variable Y assigns a probability to each possible outcome.

- Let $S = \{y_1, y_2, \ldots, y_k\}$ be the set of all possible outcomes, and Y be the realization of the variable.

- Then, $p(y_k) = Pr(Y = y_k)$, where
- $0 \leq p(y_k) \leq 1 \ \forall \ k$
- $$\sum_{k=1}^{K} p(y_k) = 1$$

We already made one of these

y_k	$Pr(Y = y_k)$
2	$1/36$
3	$2/36$
4	$3/36$
5	$4/36$
6	$5/36$
7	$6/36$
8	$5/36$
9	$4/36$
10	$3/36$
11	$2/36$
12	$1/36$

We already made one of these

y_k	$Pr(Y = y_k)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

- $p(y_k) = Pr(Y = y_k)$
- $0 \leq p(y_k) \leq 1 \ \forall \ k$
- $\sum_{k=1}^{K} p(y_k) = 1$

Parameters of distributions

- In probability theory we often wish to identify two important characteristics of distribution.

Parameters of distributions

- In probability theory we often wish to identify two important characteristics of distribution. - We wish to know the mean (μ), which is sometimes called the expected value.
  - We wish to know the variance (σ^2)
- NOTE: This is **not** the same as $\bar{x}$ and s^2.

Parameters of distributions

- In probability theory we often wish to identify two important characteristics of distribution. - We wish to know the mean (μ), which is sometimes called the expected value.
    - We wish to know the variance (σ^2)
- NOTE: This is **not** the same as $\bar{x}$ and s^2. Why are these greek letters?

Expected value

y_k	$Pr(Y = y_k)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Mean of probability distribution (**expected value**)

$$\mu = \sum_{k=1}^{K} y_k Pr(Y = y_k)$$

Expected value

y_k	$Pr(Y = y_k)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Mean of probability distribution (**expected value**)

$$\mu = \sum_{k=1}^{K} y_k Pr(Y = y_k) = 2(1/36) + 3(2/36) + \ldots + 12(1/36)$$

Expected value

y_k	$Pr(Y = y_k)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Mean of probability distribution (**expected value**)

$$\mu = \sum_{k=1}^{K} y_k Pr(Y = y_k) = 2(1/36) + 3(2/36) + \ldots + 12(1/36) = 7$$

Expected value

y_k	$Pr(Y = y_k)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Mean of probability distribution (**expected value**)

$$\mu = \sum_{k=1}^{K} y_k Pr(Y = y_k) = 2(1/36) + 3(2/36) + \ldots + 12(1/36) = 7$$

The variance of a distribution

y_k	$Pr(Y = y_k)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

$$\sigma^2 = E(Y - \mu)^2$$

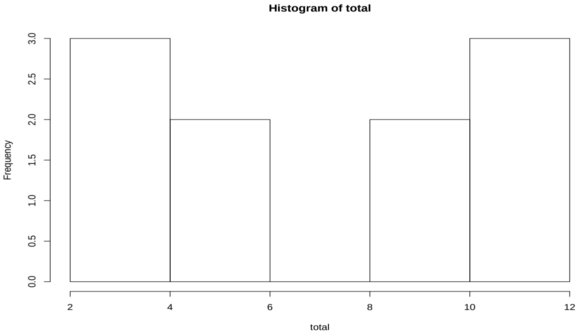
The variance of a distribution

y_k	$Pr(Y = y_k)$
2	$1/36$
3	$2/36$
4	$3/36$
5	$4/36$
6	$5/36$
7	$6/36$
8	$5/36$
9	$4/36$
10	$3/36$
11	$2/36$
12	$1/36$

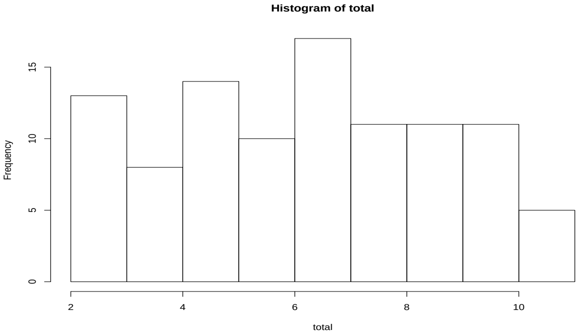
$\sigma^2 = E(Y - \mu)^2$ requires extra calculations

A little simulation

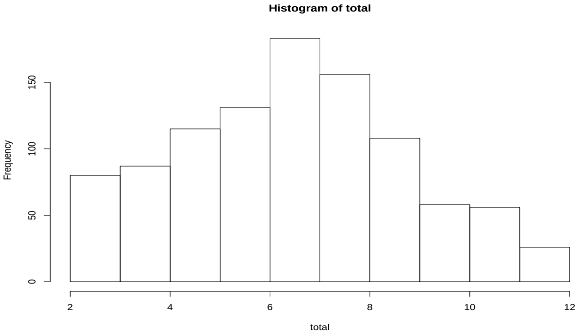
```
posVal<-c(1,2,3,4,5,6)
numRoll<-10
die1<-sample(x = posVal, size=numRoll, replace=TRUE)
die2<-sample(x = posVal, size=numRoll, replace=TRUE)
total<-die1+die2
hist(total)
```



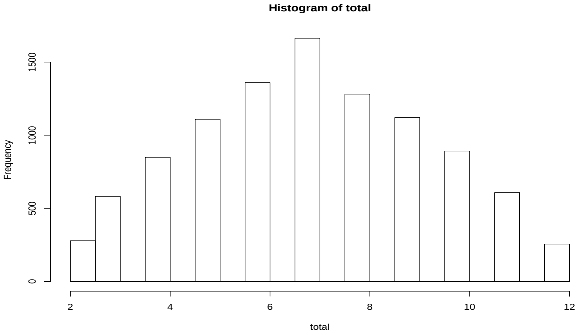
```r
posVal<-c(1,2,3,4,5,6)
numRoll<-100
die1<-sample(x = posVal, size=numRoll, replace=TRUE)
die2<-sample(x = posVal, size=numRoll, replace=TRUE)
total<-die1+die2
hist(total)
```



```r
posVal<-c(1,2,3,4,5,6)
numRoll<-1000
die1<-sample(x = posVal, size=numRoll, replace=TRUE)
die2<-sample(x = posVal, size=numRoll, replace=TRUE)
total<-die1+die2
hist(total)
```



```r
posVal<-c(1,2,3,4,5,6)
numRoll<-10000
die1<-sample(x = posVal, size=numRoll, replace=TRUE)
die2<-sample(x = posVal, size=numRoll, replace=TRUE)
total<-die1+die2
hist(total)
```



End of part 1

- What is a probability?
- What is a frequency distribution?
- What are the two most important parameters for characterizing a distribution?