Probability

We live by chance

Probability

- Probability rules our lives
- It is everywhere!
- What are the chances it rains tomorrow?
- What are the chances you win the lottery?

Probability

- Humans are really bad at interpreting probabilities
- Even worse at calculating (estimating) probabilities

The Media Has A Probability Problem

The media's demand for certainty — and its lack of statistical rigor — is a bad match for our complex world.

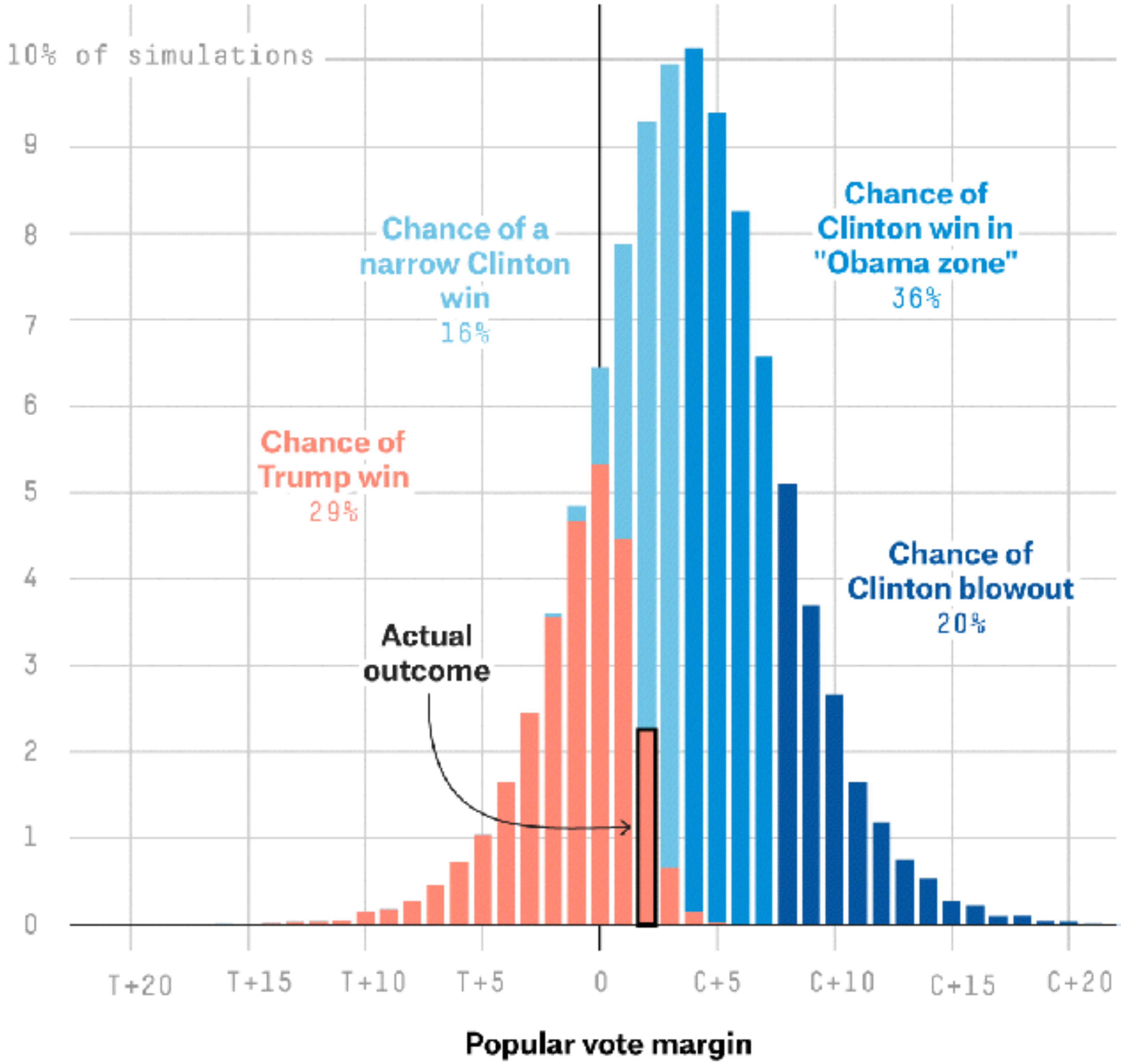
By Nate Silver

Filed under The Real Story Of 2016

Published Sep. 21, 2017

FiveThirtyEight's final forecast for 2016

Likelihood of popular vote outcomes according to FiveThirtyEight's polls-only model at 9:35 a.m. on Election Day 2016. Based on 20,000 simulations.



FiveThirtyEight

DAY		DESCRIPTION	HIGH / LOW	PRECIP	WIND	HUMIDITY
TODAY OCT 25		Sunny	78°/53°	0%	WNW 7 mph	41%
THU OCT 26		Sunny	83°/57°	0%	SSW 17 mph	41%
FRI OCT 27		Partly Cloudy/Wind	63°/38°	10%	N 21 mph	50%
SAT OCT 28		Sunny	62°/38°	0%	NNW 13 mph	34%
SUN OCT 29		Sunny	70°/49°	0%	SW 8 mph	30%
MON OCT 30		Sunny	80°/57°	0%	SSW 12 mph	45%
TUE OCT 31		PM Showers	77°/62°	40%	S 10 mph	72%
WED NOV 1		Scattered Thunderstorms	73°/64°	60%	SSW 9 mph	78%
THU NOV 2		Partly Cloudy	75°/56°	10%	SSE 10 mph	63%
FRI NOV 3		Partly Cloudy	74°/59°	20%	SSE 8 mph	63%
SAT NOV 4		Partly Cloudy	78°/61°	20%	SSE 9 mph	65%
SUN NOV 5		PM Thunderstorms	77°/59°	50%	SSE 8 mph	68%
MON NOV 6		Scattered Thunderstorms	74°/61°	50%	SSE 7 mph	82%
TUE NOV 7		Scattered Thunderstorms	74°/58°	50%	S 7 mph	82%

Some Rules of Probability

- Sample space: set of all possible outcome
    - examples: dice, deck of cards?
- Can think of an events probability as the *frequency* with which it happens if the process is repeated over and over again independently and under same conditions

Some Rules of Probability

- Probabilities are always between 0 and 1 (or 0 and 100%)
- Probability of even A is equal to 1 - probability of (not A)
- Total (sum of) probabilities for all possible outcomes must be 1 (or 100%
  - Sometimes it is easier to find the probability of **(not A) instead of A**, meaning calculate P(A) = 1 - P(not A)

- Probability of event A is equal to the number of outcomes that make it true divided by the total number of possible outcomes (sample space)

- In essence, we look at the share of outcomes in the sample space that would make the event happen

- This means we can count possible outcomes and figure out probabilities

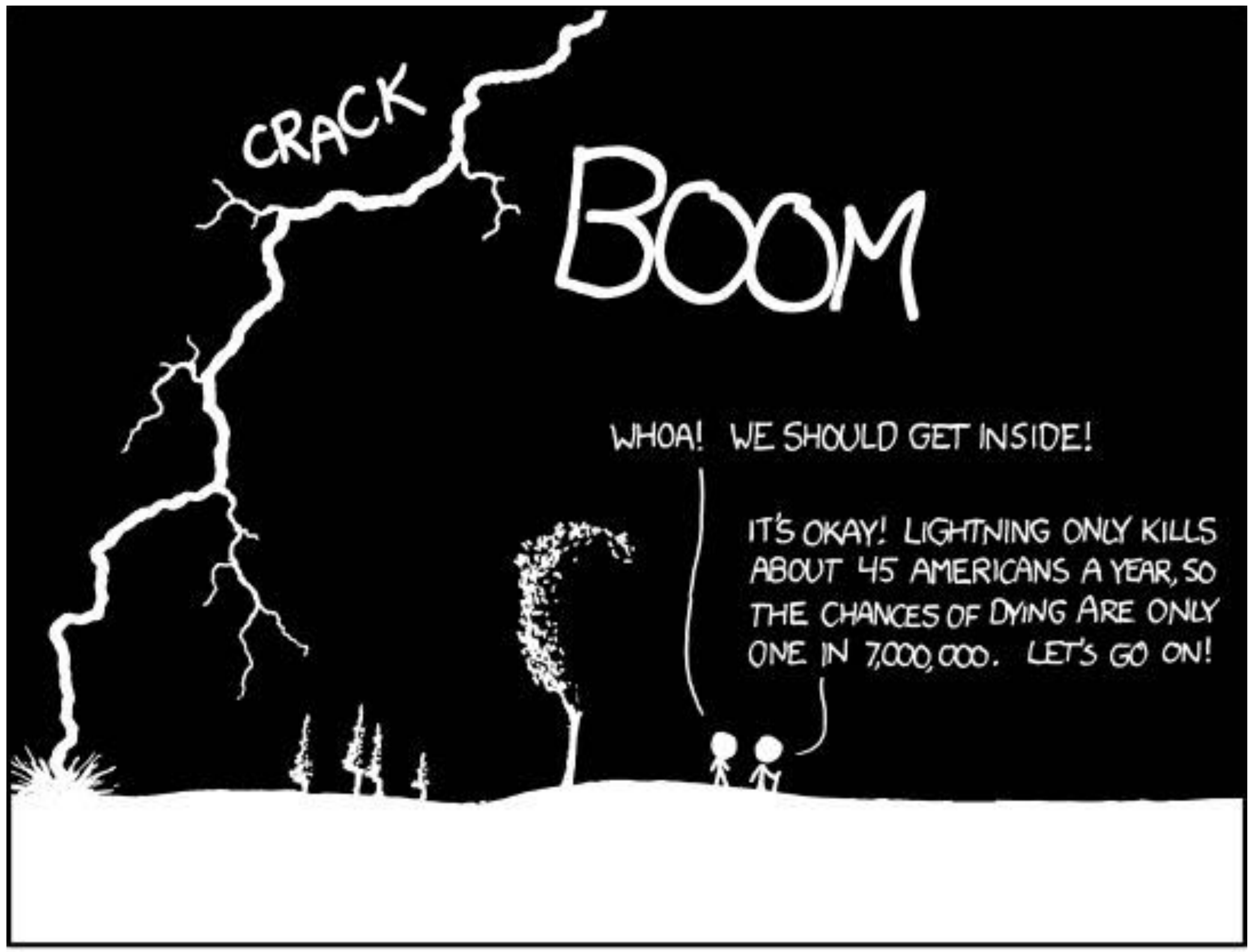
Probabilities

- Let’s think about the probabilities associated with rolling a single die
- What is the sample space (i.e. all possible outcomes)?
- What is the share of the sample space for which our condition is true?
- What is the probability of rolling a 6?
- What is the probability of rolling a 1 or a 6?

- A deck of cards
- What is the sample space?
- What is the share of outcomes for which our condition is true?
- Drawing a queen of hearts?
- Drawing a jack?
- What is the probability of drawing any card between 2 and 10, or jack, queen, king in any color?

Probabilities

- Let’s think about the probabilities associated with rolling a single die
- What is the sample space?
- What is the probability of rolling a 6?
- What is the probability of rolling a 1 or a 6?



THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT STATISTIC IS ONE IN SIX.

Conditional Probabilities

- Probabilities change with more information
- What is the probability of rolling a 5 and then a 6?
- What is the probability of rolling a 5 and then a 6 conditional on having first rolled a 5?
- P(rolled 5 then 6)
- P(rolled 5 then 6 | rolled 5)

Multiplying Probabilities

- When trying to figure out the probability of multiple events happening we often (most) multiply their probabilities

- For example, what is the probability of rolling a 6 and then another 6?

    - Chance of two things both happening is the chance that the first will happen, times the chance the second will happen **conditional** that the first has happened

When two events are independent

- Their probabilities are are independent of each other
- This means $P(A) = P(A \mid B)$
- The probability that A happens does not depend on whether B happens or not

Independence

- When I roll a die, the probability of rolling a 6 is independent of the last roll

- The probability of me sleeping through the night is **not** independent of the probability that my daughters sleep through the night

Recall multiplication

- Multiplication of probabilities always works when events are independent
- Otherwise be careful to use conditional probabilities!

Addition of probabilities

- If two events are mutually exclusive (i.e. only one of them can happen), then the probability of at least one of them happening is the addition of their individual probabilities

- Example: rolling a 1 or a 6

- **Don’t do this for events that are not mutually exclusive!**

Lottery example

- Every week you buy a lottery ticket that gives you a 1 in 1million chance of winning. Assume you buy a ticket once a week for 10 years, what is the probability you never win?

Lottery example

- Every week you buy a lottery ticket that gives you a 1 in 1million chance of winning. Assume you buy a ticket once a week for 10 years, what is the probability you never win?
    - Chance of not winning any given week 999,999/1,000,000
    - These are independent events
    - $(999{,}999/1{,}000{,}000)^{520} = 0.9995$

Remember

- **Writing Assignment 2 first draft is due on Monday**
  - **Submit via eCampus**
  - **Peer review also done via eCampus!**

Law of Averages

- What does it say?
- In the long run….

Law of Averages

- What does it say?
- In the long run…. Or with a large number of trials the frequency of events will approximate their probabilities & the larger the number of trials the larger will be the **absolute error** but the smaller will be the **error in percentage**

Law of Averages

- What does it say?
- e.g. for coin flip:
  number of heads = half number of tosses + chance error
- Chance error gets larger with larger number of trials but smaller as a percentage

Sum of Draws

- We use the sum of draws from a box of tickets (or sum of rolls from a dice) as an analogy to real world probabilities
- Think about rolling a dice 6 times and recording each roll, then summing the results.
- If we do that a number of times, the resulting sums will be different each time
- How much different?

Box models

- We can translate many problems in probability into box draws, i.e. a box with a number of different tickets with different prices
- Example: roulette red wins $1, black loses $1 and green loses $1, 18 winning, 20 losing spots
    - Same as box with 18, $1 Dollar tickets and 20, -$1 tickets

Exercises

- Two tickets are drawn at random without replacement from the box 1, 2, 3, 4

    A. What is the probability that the second ticket is a 4?

    B. What is the probability that the second ticket is a 4, given the first is a 2?

Exercises

- Five cards are dealt off the top of a well-shuffled deck

  A. What is the probability that the 5th card is a queen?

  B. What is the probability that the 5th card is the queen of spades, given the first for cards are hearts?

Exercise

- A deck is shuffled and two cards are dealt

  A. What is the probability that the second card is a heart, given the first is a heart?

  B. What is the probability that the first is a heart and the second is a heart?

Exercises

- A coin is tosses 3 times

  A. What is the chance of 3 heads?

  B. What is the chance of not flipping 3 heads?

  C. What is the chance of getting at least 1 tail?

Exercises

- Two draws are made at random without replacement from the box 1, 2, 3 ,4. The first ticket is lost and nobody knows what was written on it. True or False, the two draws are independent.

In a psychology experiment each subject is presented with three ordinary playing cards, face down. The subject also takes one cards at random from a separate, full deck of playing cards. If the two cards are from the same suit, the subject wins a prize. What is the probability of winning?

Exercise

- A pair of dice is thrown a 1000 times. What total should appear most often? What total should appear least often?

Deck of Cards is shuffled. TRUE or FALSE?

A. Chance that top card Is jack of spades is 1/52

B. Chance that bottom card is jack of diamonds is 1/52

C. Chance that top cards is jack of clubs or the bottom card is the jack of diamonds is 2/52

D. Chance the top card is jack of clubs or the bottom card is the jack of clubs is 2/52

E. Chance that the top card is jack of clubs and the bottom card is the jack of diamonds is 1/52 * 1/52

F. Chance that the top card is jack of clubs and the bottom card is jack of clubs is 1/52*1/52

Expected Value

- The value that we should expect, or is most likely
- For a sum of draws with replacement, the expected value is the **number of draws times the average in the box**
- why?

Expected Value

- Example:
  - Box with 2, -1, -1, -1,
  - If I play 25 games, how much am I **expected to win**?

Expected Value

- Example:
  - Box with 2, -1, -1, -1,
  - If I play 100 games, how much am I **expected to win**?
  - (2 - 1 - 1 -1)/4 = -0.25
  - -0.25*25 = -6.25

Standard Error

- The standard error tells us about the uncertainty around the expected value
- The standard error is an approximation of the chance error
- Remember in repeated draws:
    - Sum = expected value + chance error/standard error

Standard Error

- Standard Error: squareroot(number of draws)* SD of box
- For previous example of box 2, -1, -1, -1
    - SD = 1.1, Standard error 5*1.1
    - So we would expect to lose 6.25 +/- 5.5
    - Rarely are observed sums more than 2 SEs away from the expected value

In fact!

- We can use the expected value and standard error on the normal curve to better understand our expectation about the sum of draws
- With the previous example, we expected to lose $-6.25 +/- 5
- What is the chance we actually win more than 3.75?
- Would be 2 SU away from expected value

Roulette Example

- 10,000 independent plays on a roulette table, suppose gambler only puts $1 on red each play. Estimate chance house will win more than $250 from the 10.000 plays

- What is the expected value?

Roulette Example

- 10,000 independent plays on a roulette table, suppose gambler only puts $1 on red each play. Estimate chance house will win more than $250 from the 10.000 plays
- What is the expected value?
- Box of 18 tickets worth -$1 and 20 tickets worth $1 (casino perspective)
- Expected value 10,000*(2/38) = ca. 500

Roulette Example

- SE: 100*SD = 100*approx. 1 = 100
- Expected value of 500, SE of 100
- What is the chance house wins $250 or more?

Short-cut to SD

- If box only has two numbers:
- SD = (big number - small number) *sqrt(fraction w. big number *fraction w. small number)
- E.g. box with 2, -1, -1, -1:
  - (2 - (-1)) *sqrt(1/4 * 3/4)

Short-cut to SD

- Roulette Example:
  - (1 - (-1) * sqrt(20/38 * (18/38)) = 2* 0.49

Using the box method for counting

- What is the expected number of sixes in 60 rolls of a die?
  - Realize this is like a box with five 0s and one 1!
    - 0, 0, 0, 0, 0, 1
    - Only if we roll a six is our condition satisfied

Using the box method for counting

- Realize this is like a box with five 0s and one 1!
  - 0, 0, 0, 0, 0, 1
  - Only if we roll a six is our condition satisfied
  - Expected value 60*1/6 = 10
  - SD = 1*sqrt(1/6*5/6) = 0.37
  - SE = sqrt(60)*0.37 = approx 3

Box method on coin tosses

- A coin is tossed 100 times. What is the expected value and SE for number of tails? What is the chance of getting between 40 and 60 tails?

Box method on coin tosses

- A coin is tossed 100 times. What is the expected value and SE for number of tails? What is the chance of getting between 40 and 60 tails?

- Expected value: 1/2* 100 = 50

- SE = 10* (sqrt((1/2)*1/2)) = 10* 1/2 = 5

Box method on coin tosses

- What is the chance of getting between 40 and 60 tails?
- Expected value: 1/2* 100 = 50
- SE = 10* (sqrt((1/2)*1/2)) = 10* 1/2 = 5
- 2 SU below and above 50, 95%

Exercises

- A coin is tossed 100 times, landing heads 53 times. However, the last seven tosses are all heads. True or false: the chance that the next toss will be heads is somewhat less than 50%. Explain

Exercises

- With a Nevada roulette wheel, there are 18 chances in 38 that the ball will land in a red pocket. A wheel is going to be spun many times. There are two choices:

  A. 38 spins, and you win a dollar if the ball lands in a red pocket 20 or more times.

  B. 76 spins, and you win a dollar if the ball lands in a red pocket 40 or more times.

- Which is better for you? Or are they the same? Explain

Exercises

One hundred draws are made at random with replacement from the box with tickets; 1, 2

A. How small can the sum be? How large?

B. How many times should we expect ticket 1?

C. How much do you expect the sum to be?

Exercises

You gamble four times at a casino. You win $4 on the first play, lose $2 on the second, win $5 on the third, lose $3 on the fourth. Which of the following calculations tells how much you have come out ahead? (more than 1 might be correct)

A. $4 + $5 - ($2 + $3)

B. $4 + (-$2) + $5 + (-$3)

C. $4 +$2 + $5 - $3

D. -$4 + $2 + $5 + $3

Exercises

In one version of chuck-a-luck, 3 dice are rolled out of a cage. You can bet that all 3 show 6. The house pays 36 to 1, and the bettor has 1 chance in 216 to win. Suppose you make the bet 10 times, staking $1 each time. Your net gain is like the sum of ____ draws made at random with replacement from the box _____. Fill in the blanks.

Exercises

In one version of chuck-a-luck, 3 dice are rolled out of a cage. You can bet that all 3 show 6. The house pays 36 to 1, and the bettor has 1 chance in 216 to win. Suppose you make the bet 10 times, staking $1 each time. Your net gain is like the sum of **10** draws made at random with replacement from the box **1 * $36 and 215 * -$1**. Fill in the blanks.

What is the expected value?

Standard Error?

Exercise

Find the expected value for the sum of 100 draws at random with replacement from the following boxes:

A. 0, 1, 1, 6

B. -2, -1, 0, 2

C. -2, -1, 3

D. 0, 1, 1

Exercise

You play roulette 100 times, staking $1 on red or black (odds are 18/38). Find the expected value for your net gain.

Exercise

There are two options for draws with replacement from the following boxes:

A. 100 draws: 1, 1, 5, 7, 8, 8

B. 25 draws from: 14, 17, 21, 23, 25

For each of the three payoffs, which box is better?

1. $1 when the sum is 550 or more
2. $1 when the sum is $450 or less
3. $1 when sum is between 450 and 550

Exercise

At Nevada roulette tables, the “house special” is a bet on the numbers, 0, 00, 1, 2, 3. The bet pays 6 to 1, and there are 5 chances in 38 to win.

A. For all other bets at Nevada roulette tables, the house expects to make about 5 cents out of every dollar put on the table. How much should it expect for the “house special”?

B. Someone plays the “house special” a 100 times. Estimate the chance the person comes out ahead.

Normal Approximation for Probability Histograms

- Already previewed in last lecture when we used EV and SE to find out chance/probability between certain values
- Probability histograms represent chances, not distributions

Probability Histograms

- Probability histograms represent the chance by area
- Area above values represents their probability
- Empirical histograms of the values drawn converge (become more similar) to the probability histogram as number of repetitions increase

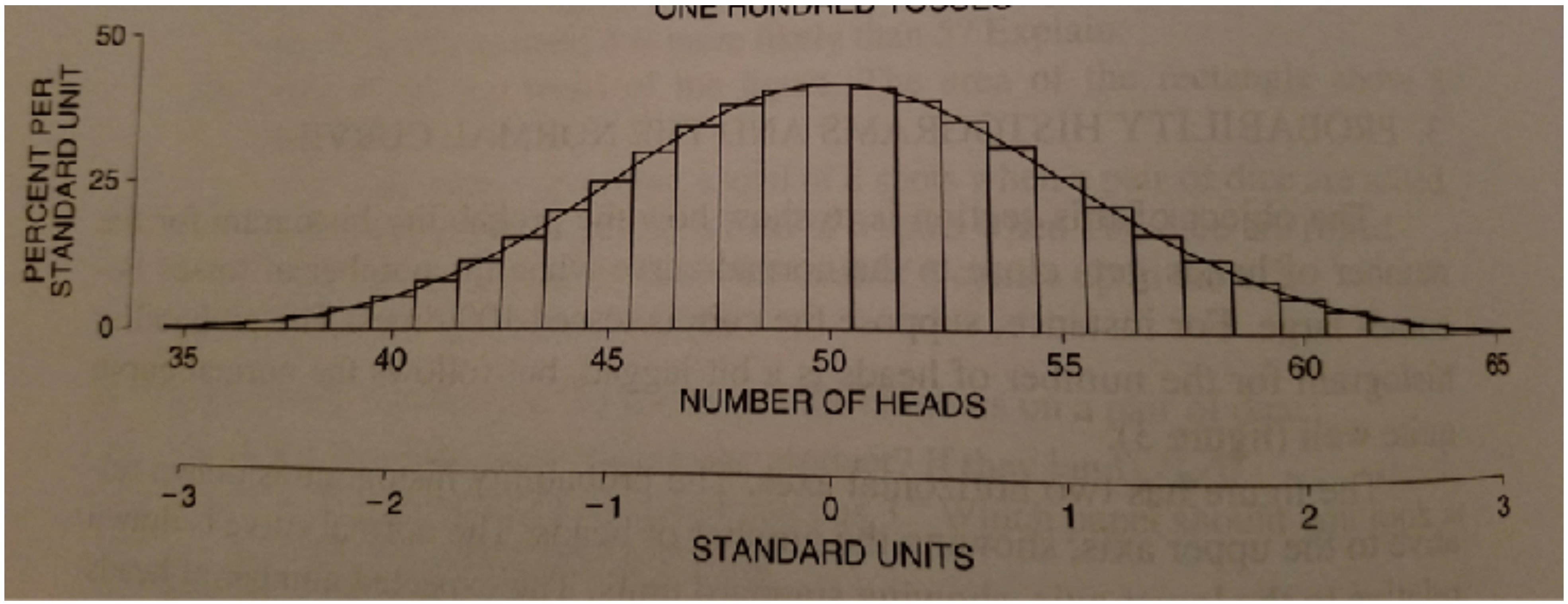
And the normal curve?

- Probability histograms often follow the normal curve
- Especially true for things we can represent as sums of draws
- As the number of draws increases, the normal approximation becomes more accurate

Normal Approximation

- We can also be more precise with our normal approximation and include/exclude endpoints
- Need to check where rectangles end and base approximation on those last included endpoints
- Then just convert to standard units and proceed as usual

EV = 50, SE = 5



What is the probability of getting exactly 50?

What is the probability of getting exactly between 45 and 55 inclusive?

What is the probability of getting exactly between 45 and 55 exclusive?

Normal Approximation

- The more the histogram of numbers in the box differs from the normal curve, the more draws we need to approximate the normal curve
- Importantly, the normal curve is tied to sums — **It does not work for products of draws, etc**

Central Limit Theorem

When drawing at random with replacement from a box, the probability histogram for the sum will follow the normal curve, even if the contents of the box do not. The histogram must be put into standard units and the number of draws must be reasonably large.

Central Limit Theorem

How many draws are enough? It depends on the contents of the box

Generally around a 100 draws should be enough

Average and SD of Box and number of draws is all we need to know!