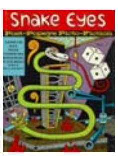
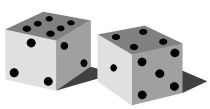
The Normal Approximation to Probability Histograms

Where are we going?

- Probability histograms
- The normal approximation to binomial histograms
- The normal approximation to probability histograms of sums

Dice: Throw a single die twice



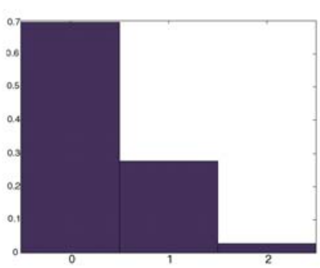


What's the chance of rolling *snake eyes* (two single dots)?

No single dots?

Exactly one single dot?

The Probability Histogram: Area = Probability



Probability histogram for the number of "ones" in two rolls

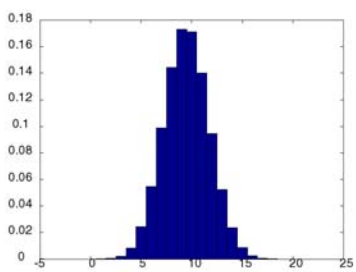
Another Example: Bet on Black in Roulette

What's the chance of winning?

How can we find the chance of winning 10 times in 20 tries? 5 times in 20 tries?

Binomial with 20 trials and chance of success 18/38 on each trial. We can calculate the chance of winning 10 times or 5 times from the binomial formula

Betting on Black in Roulette: probability histogram of the Number of Blacks in 20 tries



Equivalent Box Model

This situation is equivalent to drawing 20 times from a box with 18 ones and 20 zeros and adding the values of the draws. Box average =18/38.

Expected number of blacks =

=

Standard error of number of blacks

=

=

Shortcut formula for SD of a box with two values

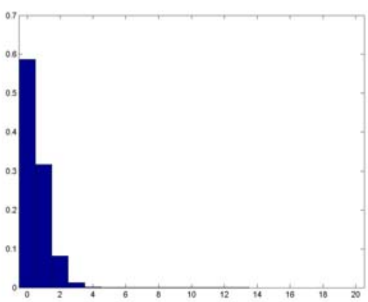
Roulette: Bet on a Number

What’s the chance you win?

What’s the chance you win twice in 20 tries?

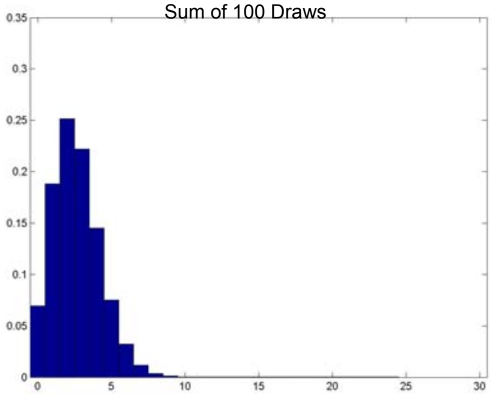
Use binomial with 20 tries and chance of success = 1/38

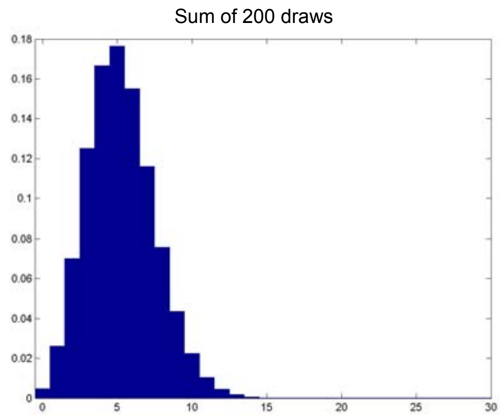
Probability histogram for binomial n=20 & p=1/38

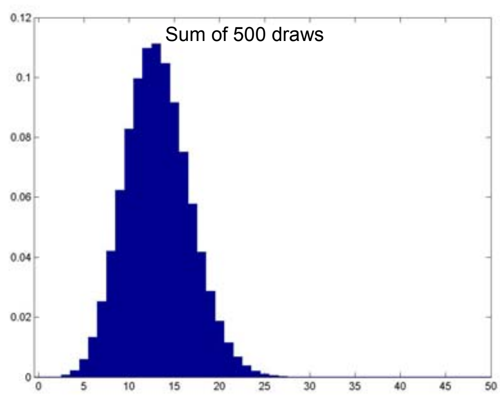


The equivalent box model: sum of draws from a box with zeros and one.

Sum of 100 Draws







Normal curve approximation

We will approximate probability histograms by normal curves.

When we approximated areas of histograms of data by the normal curve we converted to standard units using the average and the SD of the data.

To approximate areas of probability histograms of sums, we will convert to standard units using the EV of the sum and the SE of the sum.

Normal Curve Approximation to Binomial Probability Histograms

To approximate areas of probability histograms of sums, we will convert to standard units using the EV of the sum and the SE of the sum.

For the binomial:

$$EV = n \times p$$

$$SE = \sqrt{n} \times \sqrt{p \times (1-p)}$$

Normal curve approximation to binomial:a rule of thumb

n x p > 5 and n x (1-p) > 5

Examples

20 x (18/38) = 9.5 20 x (20/38) = 11.5

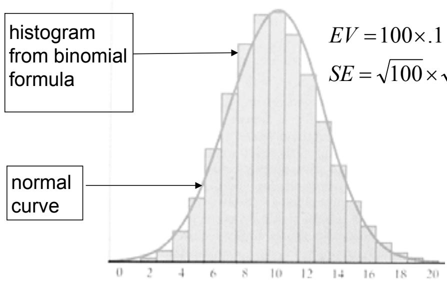
20 x (1/38) = .5

200 x (1/38) = 5.2

500 x (1/38) = 13

More binomial histograms

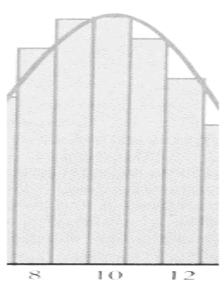
A closer look: Sum of 100 throws with probability p= .1 of success and a normal curve



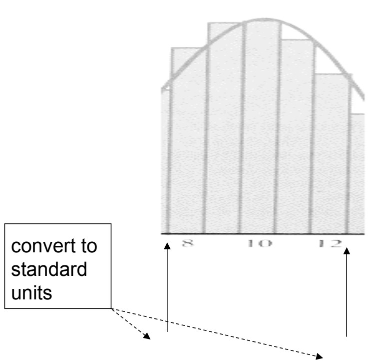
$$EV = 100 \times .1 = 10$$

$$SE = \sqrt{100} \times \sqrt{.1 \times .9} = 3$$

Normal curve approximation of the chance of between 8 and 12 (inclusive)



Between what two points should we find the area under the normal curve?



The normal curve approximation

The area under the normal curve between -.83 and .83 standard units is about .60, from the table.

The exact answer (using the binomial formula) is .60 (to two decimal places).

Normal curve approximation without keeping track of endpoints (the “continuity correction”)

Expected Number of Succeses = 10

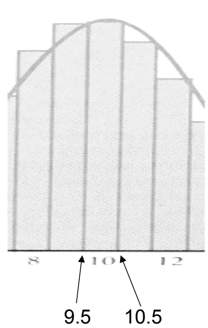
SE = 3

Want chance of between 8 and 12.

Area under normal curve within .66 standard units is 49%. Not as accurate.

Last time we got 60%. Where did the 60%-49% = 11% go?

Normal Curve Approximation to Chance of 10



Expected Number of Successes = 10

Standard Error = 3

From the table, the chance is about 11.9%. That's using the normal curve approximation.

The exact chance is 13%, found from the formula for the binomial probabilities.

When to use the continuity correction?

Rule of thumb: you don't need to use the continuity correction if the SE is greater than 10 or so and you are finding an area over a region that is at least a sizeable fraction of an SE wide.

The Central Limit Theorem

The probability histogram of the sum of a large number of independent draws from a box can be approximated by a normal curve.

To find the approximating normal curve for the sum of the draws you only need to know:

The box average ⟹ expected value of sum

The box SD ⟹ standard error of sum

When the number of draws is large, other facts about the box don't matter.

Using the Normal Curve Approximation

For a large number of draws:

- The chance of getting the EV plus or minus one SE is about 68%
- The chance of getting the EV plus or minus two SE is about 95%

To approximate the chance of being in a given range, express the numbers in standard units and use the normal curve.

How Many is a Large Number?

There is no simple, universal answer. If the histogram of the values of the tickets is not too far from normal, the number can be small (say 30).

The more the histogram of the tickets in the box differs from a normal curve, the larger the number of draws required in order that the normal curve approximates the probability histogram of the sum.

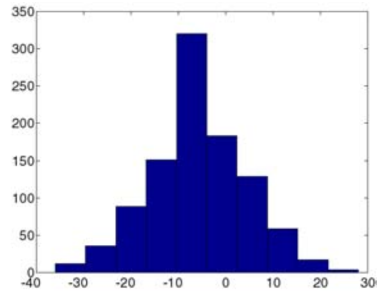
If the box contains rare tickets with very large values, the number will have to be larger.

"Usually," 100 is large enough.

Examples

Roulette: Betting on Black

1000 Simulations of Sums of 100 Draws



Expected Value of Sum = -$5.26 SE of Sum = $9.90

Roulette: Using the Normal Curve

Bet on Black

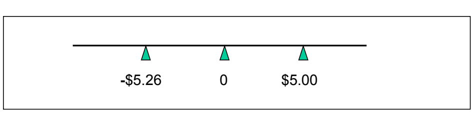
- Box contains 18 tickets = $1 and 20 tickets = -$1
- Box Average = -.0526
- SD of Box = .99

Draw 100 times

- Expected value of sum =
- SE of sum of =

What's the chance of winning $5 or more? Use normal curve to get an approximate answer.

Convert to Standard Units:

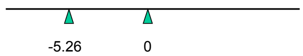


Expected Value = ; SE =

From the normal table, the chances are about 15%

What's the Chance You Come Out Ahead?

Expected Value = -$5.26;   SE = $9.90



From the table of the normal curve, the chance is about 29%.

Is it better to play 100 times for $1or 1 time for $100?

If you play 100 times, your expected winnings are -$5.26. The chance you come out ahead is 29%.

If you play once: the box has 20 tickets with values -$100 and 18 tickets with values $100.

Expected winnings =

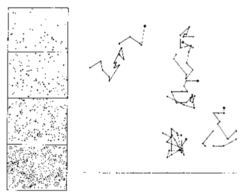
=

If you play 100 times for $1 each time, the chance you come out ahead is about 29%.

However, if you play once for $100, the chance you come out ahead is

How to gamble if you must: if the odds are against you, play a bold strategy.

Brownian Motion



First observed by Brown in 1827.
Theory by Einstein in 1905

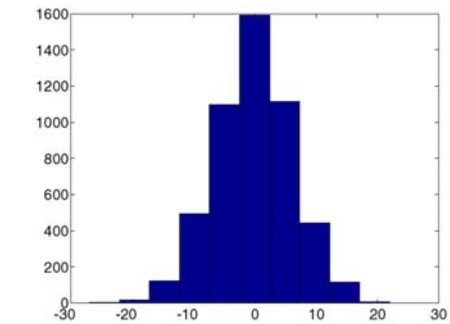
Brownian Motion:
The Movie

A Random Walk

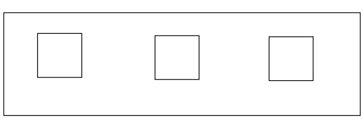
Steps left with probability 1/3, doesn't move with probability 1/3, and steps right with probability 1/3.

Where do you expect him to be after 60 steps? How far away from the starting point will you have to look for him?

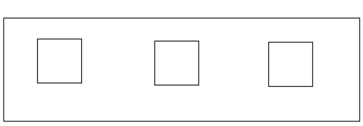
Histogram of positions at the ends of 5000 walks each of 60 steps



Box and Ticket Model



Position after 60 steps is the sum of 60 draws.



Box Average = 0 Box SD = .8

Expected Value of Sum = 0

Standard Error of Sum =

What's the chance that after 60 steps he is more than 20 steps from where he starts? Use normal curve approximation.

From the table, the chance of more than standard units (in either direction) is about

Very small.

(Tricky) Problem: 25 draws are made at random, with replacement, from the following box:

1	1	2	2	2

Find, approximately the chance that 1 shows up more times than 2 does in the 25 draws.

We are interested in the number of times a 1 is drawn in 25 draws. What's the box model for this?

EV of sum of draws =

SE of sum of draws =

So the EV of the sum is 10 and the SE is 2.5 and we want to know the chance that the sum is 13 or greater.

Use normal approximation: use continuity correction and convert to standard units.

How many standard units is 12.5?

Now from the normal table, the chance of being greater than 1.0 is about 16.%