Normal Approximation to the Binomial

Donna Dietz

American University

dietz@american.edu

STAT 202 - Spring 2020

Back to Normal

Although we probably do use the Normal distribution more than we should, it has earned its status fair and square. Mathematics and history both lead us to this conclusion. History, because there was a time when computers were not readily available. Mathematics, because of the beautiful properties of the Normal distribution which have permitted it to stand in for many other distributions over the years!

Back to Normal

However, when the Normal distribution stands in for other distributions, there are dead giveaways, like rules of thumb and adjustments you have to do prior to making your calculations.

Two problems

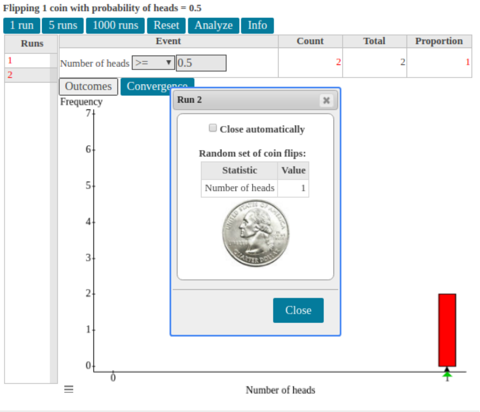
As often happens in math and science, we have a forward problem and an inverse problem. The forward problem is this: If we know what our distribution looks like, then what will the samples look like?

Our inverse problem is: If we have a sample, what can reasonably be said about the population it came from?

These slides only focus on the forward problem, but later we will talk about the inverse problem!

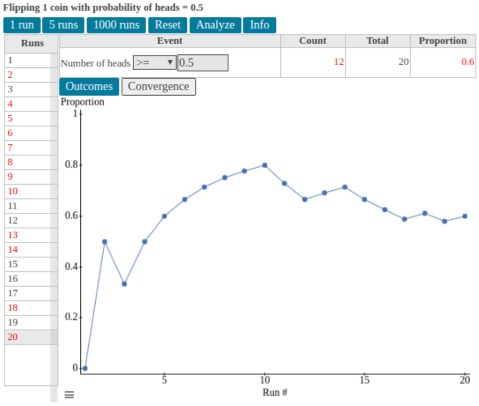
Coins

We can simulate coin flipping in StatCrunch. This is a Binomial distribution, but we'll show how to use the Normal distribution to estimate it rather well!



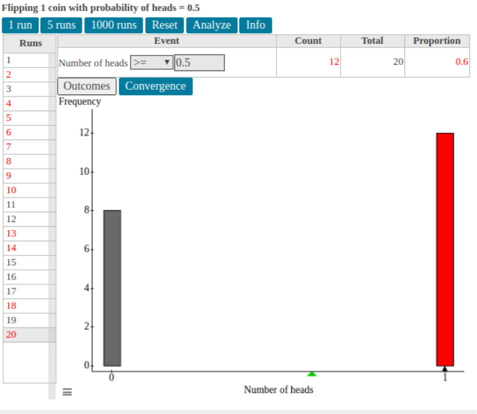
One at a time coin flipping will not get us anywhere fast.

Convergence



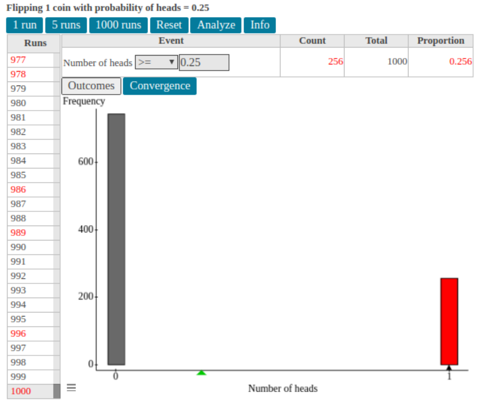
Notice there is a convergence window. This shows you the running average as it is computed.

20 flips



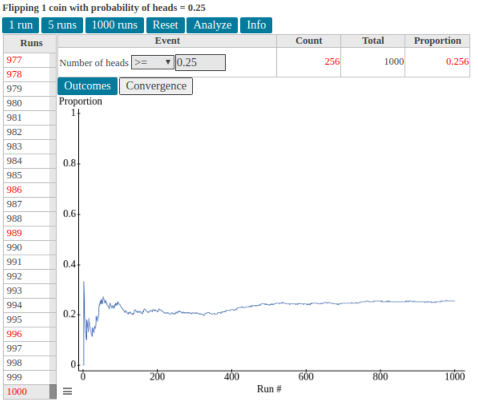
This is what it could look like after 20 coin flips.

1000 flips



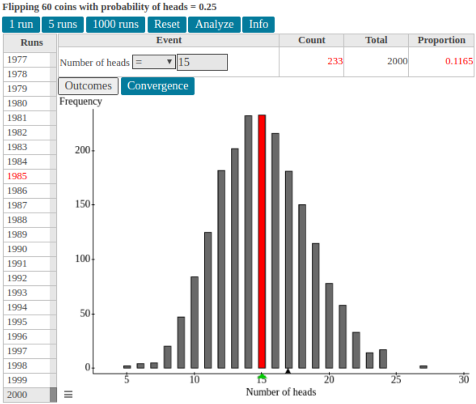
This is what it could look like after 1000 coin flips with $p = 0.25$.

1000 flips



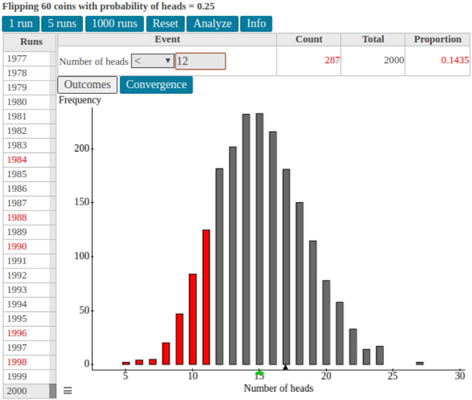
Convergence for 1000 flips with $p = 0.25$.

Exactly 15, the expected outcome



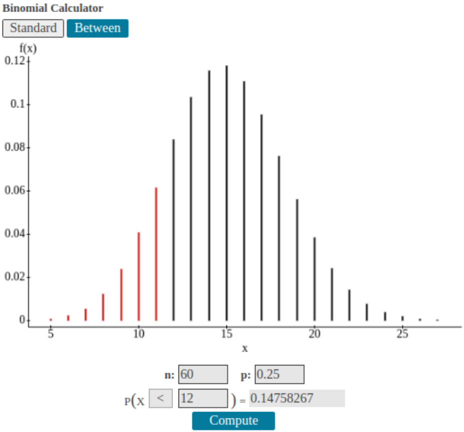
Expected value is 15. How often did we get exactly 15?

cutting off the bottom



$x = 12$ seems to cut off the bottom 14%, very much like $z = -1$ would cut off 16%.

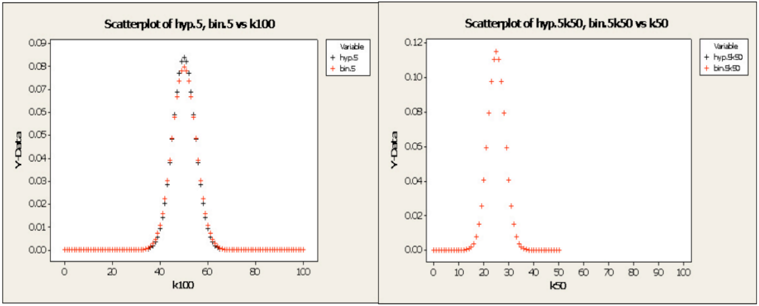
Binomial Calculator



The binomial calculator gives us greater accuracy than a simulation, but it shows our simulation was pretty good!

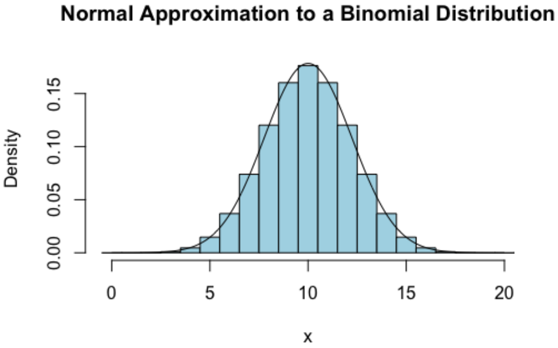
Using the Normal Curve for Binomial Calculations

- Is your expected number of successes 10 or more?
- Is your expected number of failures 10 or more?
- Use $\mu = np$
- Use $\sigma = \sqrt{np(1-p)}$
- Add 0.5 to the upper limit
- Subtract 0.5 from the lower limit
- Don't sample more than 10% without replacement
- Recheck everything for errors!



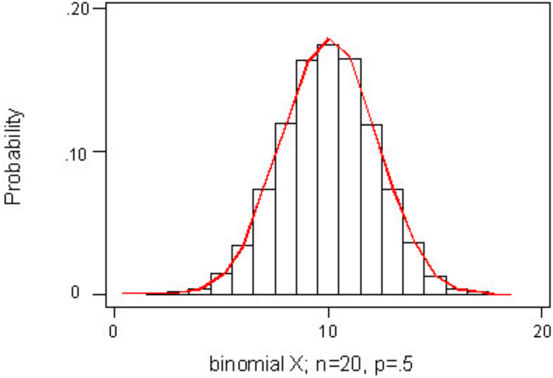
If you are interested in a full explanation of this, please visit:
https://web.ma.utexas.edu/users/mks/M358KInstr/TenPctCond.pdf

Some examples from the Internet!



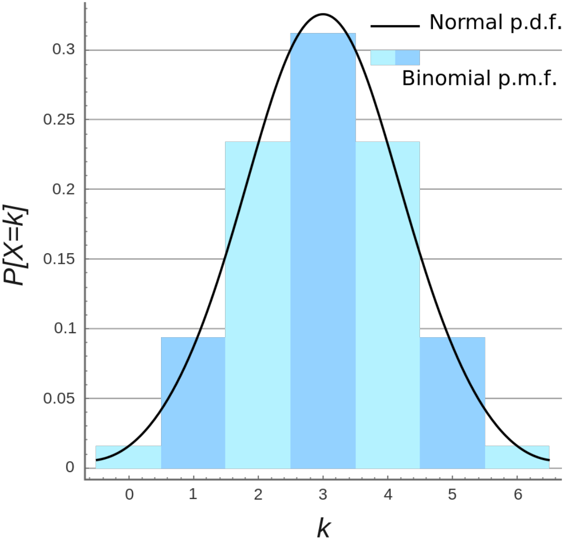
From Emory

Some examples from the Internet!



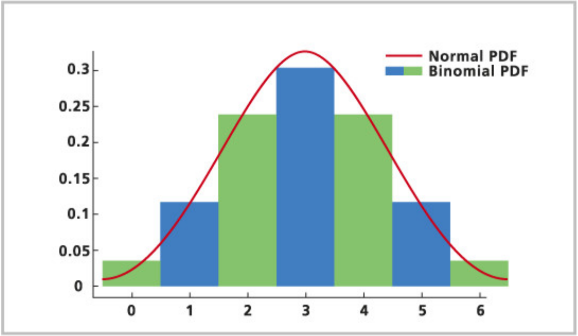
From U. Florida

Some examples from the Internet!



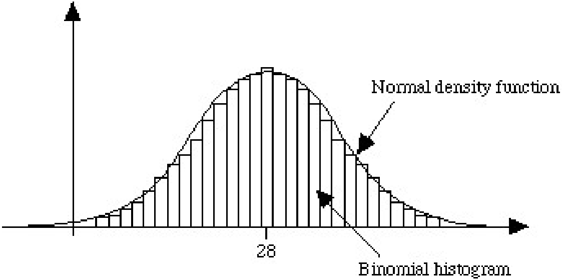
From ChampionBets

Some examples from the Internet!



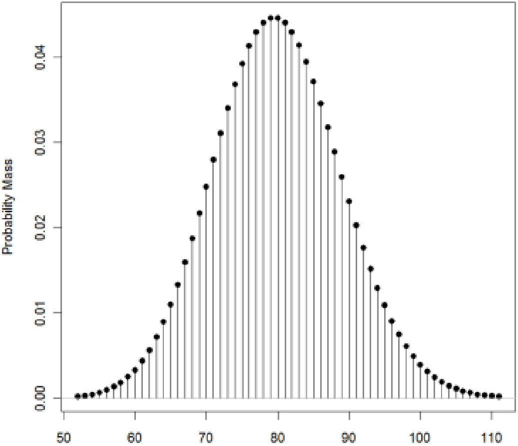
From TinyCards

Some examples from the Internet!



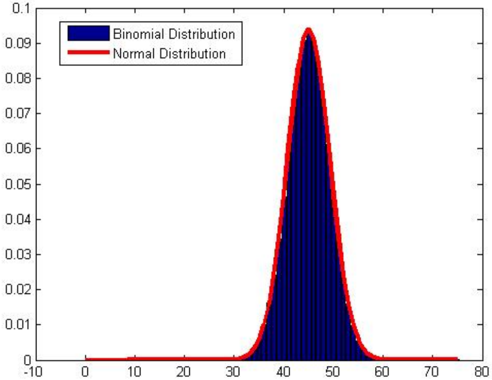
From J. Sevy (jsevy.com)

Some examples from the Internet!



From MakeMeAnalyst

Some examples from the Internet!



From MathWorks

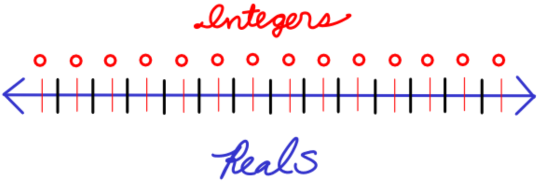
Why all the examples?

I wanted you to see a variety of graphics to get the idea into your mind that we have what is really and truly a set of integers in our physical real distribution. However, we are only using the mechanism of the Normal Distribution for calculations. Thus, each integer bar must occupy ± 0.5 for a width of 1.

When I ask you for something within $\pm n$ of the expected outcome, for example, let's say the expected outcome is 15 and I want to know when we get ± 3. Then, you would want to know whether we got between 12 and 18 of something. If you are using the Binomial calculator, which is discrete, that would be $12 \leq x \leq 18$, however, if you are using the continuous model, that would be $11.5 \leq x \leq 18.5$.

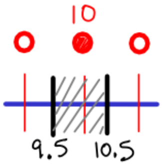
Each integer in our discrete model occupies a width of one in our continuous model.

Numbers: Integers and Real



Think of **Integers** as points, but think of **Real** numbers as having ranges rather than locations.

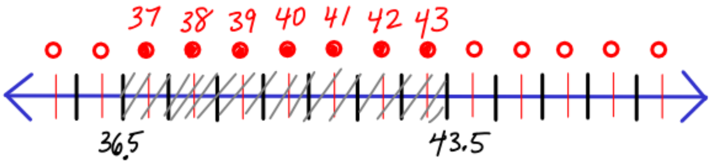
Example: 10



If I want to think of 10 as an integer, it's just "10" because we know what point we mean.

But if we want to represent something meaningful in terms of a range which represents **Real** values that are close to 10, we have to grab the half unit of neighbors on either side of 10.

Numbers: An example



Values from 37 to 43, as integers would be $37 \leq x \leq 43$, but if you are using a continuous model to represent discrete values, you have to stretch that on either side, or $36.5 \leq x \leq 43.5$.

Beware! $<$ vs. $\leq$

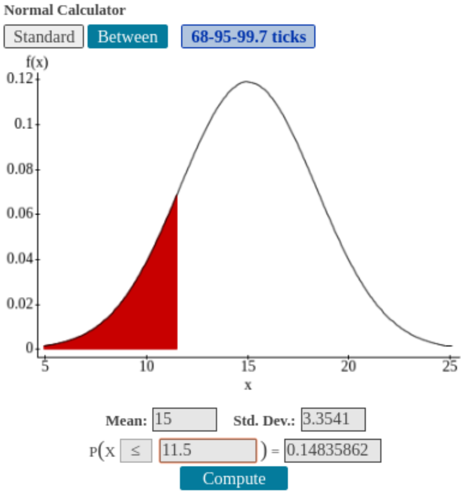
It's easy to get confused as you move between the discrete and continuous worlds! If I ask for the probability of getting fewer than 12 heads, it's really the same as asking for 11 or fewer heads. Which is our x value?

For the discrete value x:
the continuous model will occupy $x - 0.5$ to $x + 0.5$.

Examples

discrete model	think:	continuous model
$x < 7$	include 6	$x < 6.5$
$x > 10$	include 11	$x > 10.5$
$4 < x < 9$	include 5 and 8	$4.5 < x < 8.5$
$x \leq 10$	include 10	$x < 10.5$
$3 < x \leq 11$	include 4 and 11	$3.5 < x < 11.5$

$x < 12$ means $x \leq 11.5$



Using the Normal Table

You can get close to this result using a Normal Table:

$$\mu = np = 60 \cdot 0.25 = 15$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{15 \cdot 0.75} \approx 3.3541$$

$$x = 11.5$$

$$z = \frac{11.5 - 15}{3.3541} = -1.04$$

Table entry for *z* is the area under the standard Normal curve to the left of *z*.

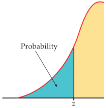


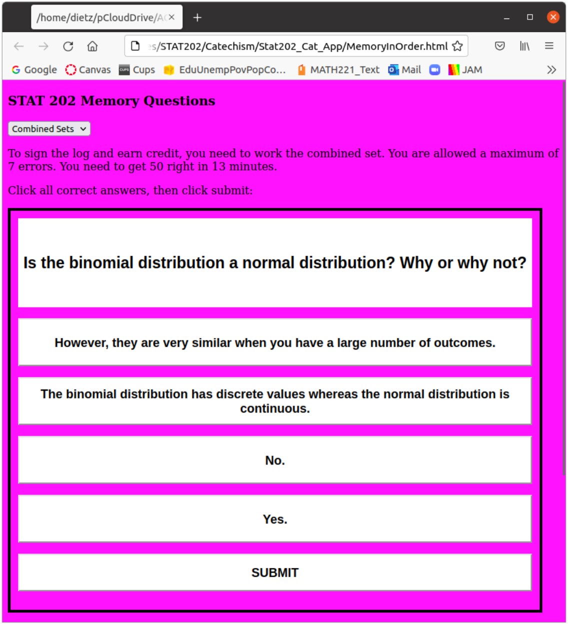
TABLE A

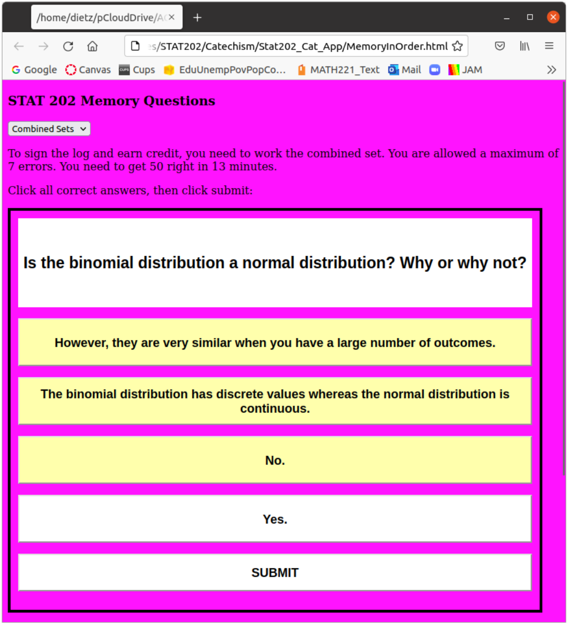
Standard Normal probabilities

z	.00	.01	.02	.03	.04	.05
−3.4	.0003	.0003	.0003	.0003	.0003	.000
−3.3	.0005	.0005	.0005	.0004	.0004	.000
−3.2	.0007	.0007	.0006	.0006	.0006	.000
−3.1	.0010	.0009	.0009	.0009	.0008	.000
−3.0	.0013	.0013	.0013	.0012	.0012	.001
−2.9	.0019	.0018	.0018	.0017	.0016	.001
−2.8	.0026	.0025	.0024	.0023	.0023	.002
−2.7	.0035	.0034	.0033	.0032	.0031	.003
−2.6	.0047	.0045	.0044	.0043	.0041	.004
−2.5	.0062	.0060	.0059	.0057	.0055	.005
−2.4	.0082	.0080	.0078	.0075	.0073	.007
−2.3	.0107	.0104	.0102	.0099	.0096	.009
−2.2	.0139	.0136	.0132	.0129	.0125	.012
−2.1	.0179	.0174	.0170	.0166	.0162	.015
−2.0	.0228	.0222	.0217	.0212	.0207	.020
−1.9	.0287	.0281	.0274	.0268	.0262	.025
−1.8	.0359	.0351	.0344	.0336	.0329	.032
−1.7	.0446	.0436	.0427	.0418	.0409	.040
−1.6	.0548	.0537	.0526	.0516	.0505	.049
−1.5	.0668	.0655	.0643	.0630	.0618	.060
−1.4	.0808	.0793	.0778	.0764	.0749	.073
−1.3	.0968	.0951	.0934	.0918	.0901	.088
−1.2	.1151	.1131	.1112	.1093	.1075	.105
−1.1	.1357	.1335	.1314	.1292	.1271	.125
−1.0	.1587	.1562	.1539	.1515	.1492	.146
−0.9	.1841	.1814	.1788	.1762	.1736	.171

The result using the Normal Table is 0.1492 which is quite close to our result from before.

MEMORY QUESTION


/home/dietz/pCloudDrive/A
s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
Is the binomial distribution a normal distribution? Why or why not?
However, they are very similar when you have a large number of outcomes.
The binomial distribution has discrete values whereas the normal distribution is continuous.
No.
Yes.
SUBMIT


/home/dietz/pCloudDrive/A
s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
Is the binomial distribution a normal distribution? Why or why not?
However, they are very similar when you have a large number of outcomes.
The binomial distribution has discrete values whereas the normal distribution is continuous.
No.
Yes.
SUBMIT