Lecture 3B: The Normal Distribution

"the Bell Curve"

→ consider the coin flip probabilities from the previous lecture

→ if we flip 4 times

heads	prob
0	6,25%
1	25,–%
2	37,50%
3	25,–%
4	6,25%

if we flip 10 times

heads	prob
0	0,1%
1	1,0%
2	4,4%
3	11,7%
4	20,5%
5	24,6%
6	20,5%
7	11,7%
8	4,4%
9	1,0%
10	0,1%

→ Notice: buckets get "thinner" as we increase the number of flips

→ probability of a given outcome becomes smaller as we increase number of flips

→ probability of tail event (all heads or all tails) becomes increasingly remote what do those observations mean?

- as the number of possible "heads outcomes" gets larger the probability of any given one gets smaller + smaller
- if we flipped the coin 1,000,000 times, the probability of getting exactly 500,000 heads is almost zero
- and the probability of getting exactly zero heads is even smaller
- also notice that the table traces out a normal distribution

→ so we know that in practice it would be very rare to get exactly 500 heads in 1000 flips (probability is only 2,5%)

→ but there's a good chance that we'll get **close to 50%**

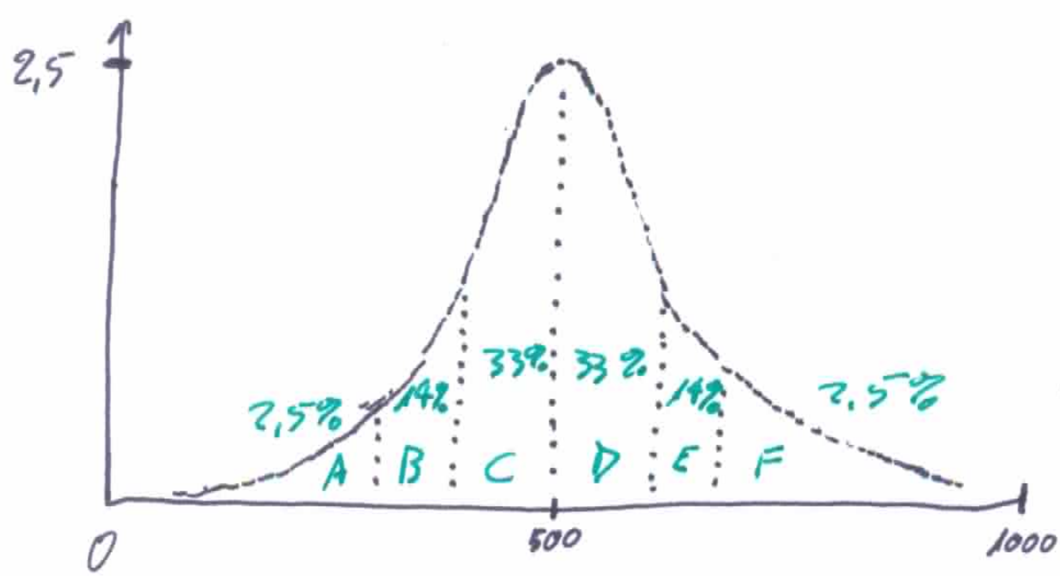
→ how close?

→ prob($485 \leq$ # heads ≤ 515) $\approx$ 67%

→ prob($470 \leq$ # heads ≤ 530) $\approx$ 95%

→ and extremes are very rare

prob(# heads < 470) = 2,5% = prob(# heads > 530)



A: 2,5%
B: 14,-%
C: 33,-%
D: 33,-%
E: 14,-%
F: 2,5%

→ Your textbook approaches the topic by telling you to imagine the measurement of the height of every man in America

→ The idea is the same: going from a small number of dice rolls to an infinite number

going from the measurement of a few dozen men to an infinite number

→ the probability of observing a specific value will be ZERO

BUT the probability of observing values in a given range is positive

→ the probability of observing the entire range of values is 100%

→ so the area under the "bell curve" is 100% (i.e. one)

$$\phi(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$e = 2.71828 \qquad \pi = 3.1415926$

↑ Normal probability density function

→ symmetrical & centered on μ

→ points of inflection at $\mu \pm \sigma$

→ the tails approach, but never touch zero

μ & σ

→ a higher μ shifts the curve right
→ a higher σ makes curve wider

→ many relative frequency distributions that we observe in practice can be treated as Normal

→ some tests designed so that scores are normally distributed

→ error distributions are often assumed to be Normal (because differences caused by multiple unrelated & uncontrollable factors)

→ so it may make sense to treat distribution of men's heights as normal

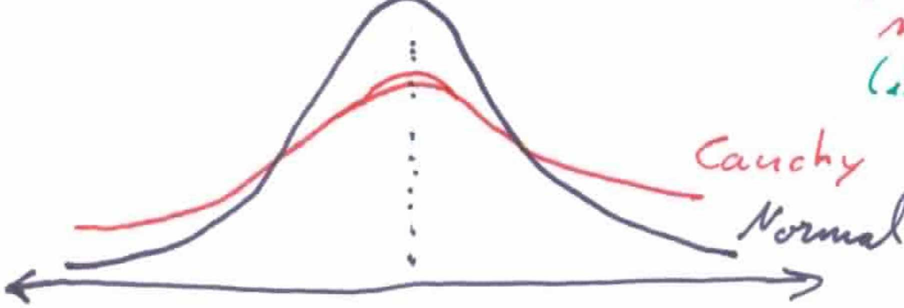
- genetics, nutrition, etc. may affect men's height
- but these factors may be unrelated to each other (and not necessarily known to researcher)

→ what distributions are not Normal?

- income (bounded below at zero) tends to be skewed (not symmetrical)

- <u>stock</u> <u>returns</u>! they're ~~sym~~ symmetrical but have ~~sym~~ fatter tails than Normal

Actually, Cauchy is a little more "pinked" (i.e. more dramatic change in slope)



→ so let's talk about investing

→ suppose your model predicts that on average you'll get a handsome 4% per year (after cost of funds, etc.)

→ but you know that it could be higher or lower than that in a given year

→ so you want to know where will I be 95% of the time?

→ You assume (wrongly) that returns are normally distributed

$$\left.\begin{matrix}\mu = 4 \\ \sigma = 1\end{matrix}\right\} \; \begin{aligned} z &= \frac{x-\mu}{\sigma} \\ &= \frac{x-4}{1}\end{aligned} \; \left.\right\} \; \begin{aligned} P[-2<z<2] &= 95\% \\ P[-1<z<1] &= 68\%\end{aligned}$$

→ so you figure that you'll be **between** 3 + 5 percent 68% of the time AND between 2 + 6 percent 95% of the time

→ but if returns follow a Cauchy distribution, then you ~~could~~ will be in the tails much more often than you anticipated!

→ this is why some traders go from rags to riches & back to rags

→ they assumed normal

→ So why don't they use Cauchy?
(Are they stupid?)

→ The Normal distribution has some very convenient mathematical properties.

→ Suppose a number (e.g. stock price) were to be drawn, what would you expect that number to be?

→ If the distribution is symmetrical, then you would expect the mean = median = mode, but how do we know that?

→ Consider the Likelihood Function

$$\mathcal{L}(x|\mu,\sigma) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

→ Maximizing it (w/ respect to μ & σ) will yield the parameter estimates that are most likely ~~to~~ to have generated the observed data

~~[illegible]~~

→ Take log

$$\ln \mathcal{L}(\cdot) = -N \ln \sigma - \frac{N}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

→ Maximize

$$\frac{\partial \ln \mathcal{L}(\cdot)}{\partial \mu} = \sum_{i=1}^{N} \frac{x-\mu}{\sigma^2} = 0 \Rightarrow \boxed{\frac{1}{N}\sum_{i=1}^{N} x_i = \mu}$$ MEAN

$$\frac{\partial \ln \mathcal{L}(\cdot)}{\partial \sigma} = \frac{-N}{\sigma} + \sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)\cdot\frac{x_i-\mu}{\sigma^2} = 0$$

$$\Rightarrow \boxed{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2 = \sigma^2}$$ VARIANCE

→ The Cauchy Distribution isn't nearly so nice

$$\mathcal{L}(x|\mu,\sigma) = \prod_{i=1}^{N} \frac{1}{\pi}\left(\frac{\gamma}{(x-x_0)^2+\gamma^2}\right)$$

$$\ln \mathcal{L}(\cdot) = -N \ln \pi + N\gamma - \sum_{i=1}^{N} \ln\left((x-x_0)^2+\gamma^2\right)$$

$$\frac{\partial \ln \mathcal{L}(\cdot)}{\partial x_0} = \sum_{i=1}^{N} \frac{2(x-x_0)}{(x-x_0)^2+\gamma^2} = 0 \Rightarrow \boxed{\frac{1}{N}\sum_{i=1}^{N} x_i = x_0}$$

MEDIAN

For technical reasons Cauchy does not have a mean

$$\frac{\partial \ln \mathcal{L}(\cdot)}{\partial \gamma} = N - \sum_{i=1}^{N} \frac{2\gamma}{(x-x_0)^2+\gamma^2} = 0$$

$$\boxed{\frac{1}{N}\sum_{i=1}^{N} \frac{2\gamma}{(x-x_0)^2+\gamma^2} = 1}$$

Now solve for γ

GOOD LUCK!