20200212【Data Analyst Nanodegree】P04M01L09

> Part 04 : Practical Statistics
> Learn how to apply inferential statistics and probability to important, real-world scenarios, such as analyzing A/B tests and building supervised learning models.

Module 01: Practical Stats

Lesson 09: Normal Distribution Theory

> Learn the mathematics behind moving from a coin flip to a normal distribution.

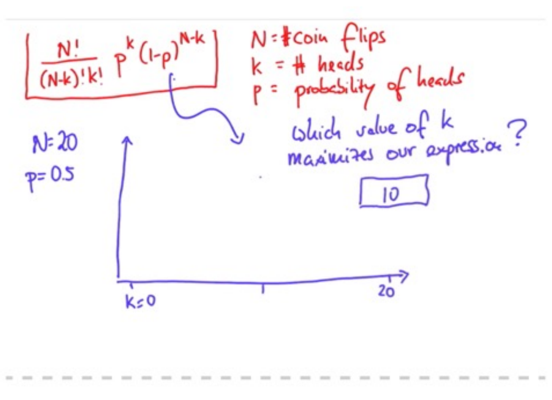
01. Maximum Probability

We start out with the **binomial distribution** that you're familiar with from our last unit and then we move into the **Central Limit Theorem(中心极限理论)** which basically means we take a number of coin flips to infinity. From that, we arrive at the **normal distribution** which is basis to so much in statistics all of **testing** and **confidence intervals** are defined though the normal distribution.

And the reason why this matters is much of what we've done in coin flips had one or two coin flips but in statistics experiments, you often have 1000 of patients or 1000 of data points and then **starting the normal distribution as an approximation to the binomial distribution** is much more practical.

Solution

The reason why the answer is 10 is because **the number of combinations to place 10 positives and 10 negatives into our list of 20 is larger than any other number**. This term over here is maximized when k is exactly half of N--so 10.



02. Shape

This curves is often called a **Bell Curve(贝尔/钟形曲线)** because it's quite feasible to think of it as a **church bell**-- that's move left and right and rings the bells.

Flip a coin a 1000 times and from that, I looked at the **empirical frequency** which is the same as that count of heads divided by a 1000, but this one scales between 0 and 1. I called this thing an **Experiment(实验)**

03. Better Formula

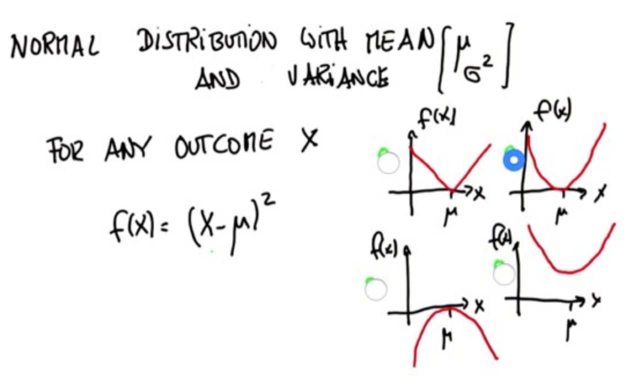
We can find a better formula for this vell-shaped curve, which applies to almost any distribution that is sampled many times and is a very deep statistical result.

04. Quadratics

I will define for you a normal distribution with a specific mean that's often called μ, Greek letter μ, and a variance that's often called σ^2.

We already know that variance is a **Quadratic Expression(二次表达式)**. In normal land we often use μ and σ^2.

The very first element is that for any outcome x we write the quadratic difference between this outcome x and μ. This is indeed a function in x. The horizontal axis is x, and we're graphing f(x). The first I'll give you is a **Triangular Function(三角函数)**. The second is a **Quadratic Function(二次函数)**. The third one is a negative quadratic function. And the fourth one is a quadratic function that doesn't quite touch μ.



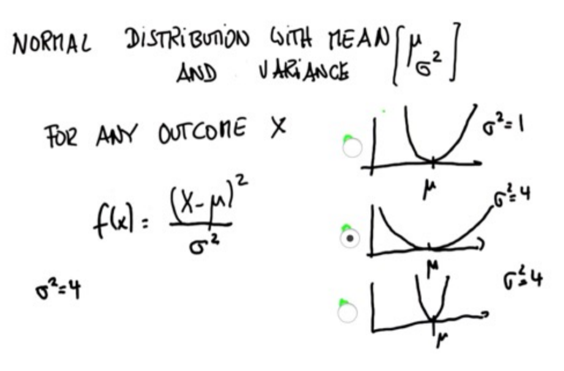
05. Quadratics 2

The next thing I'll do is I'll divide it by σ^2.

Suppose $\sigma^2 = 4$. That means we have a variance of 4 and a standard deviation of 2.

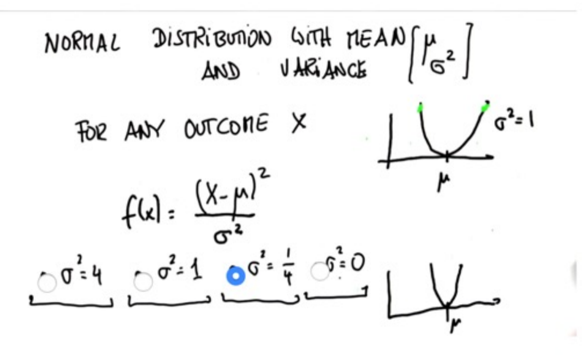
I've given you already the quadratic function when it isn't divided by σ^2. It's the same as saying $\sigma^2 = 1$.

What I'd like to know is whether our new version where σ² = 4 makes this quadratic wider or whether it makes it narrower, assuming that this is our new function f(x).



06. Quadratics 3

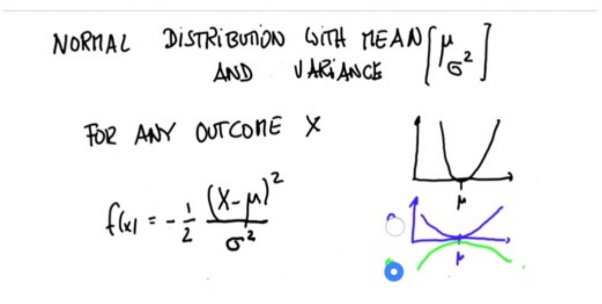
In particular, if we now look at the quadratic over here, which is much tighter, which of the following potential σ^2 would you think is best representative of this narrow function over here, provided that this is the quadratic that corresponds to $\sigma^2 = 1$.



07. Quadratics 4

Let's go further, and let's now take this function and multiply it by $\frac{1}{2}$. Again, I ask you what the affect it. If this is your original quadratic, then what do we get?

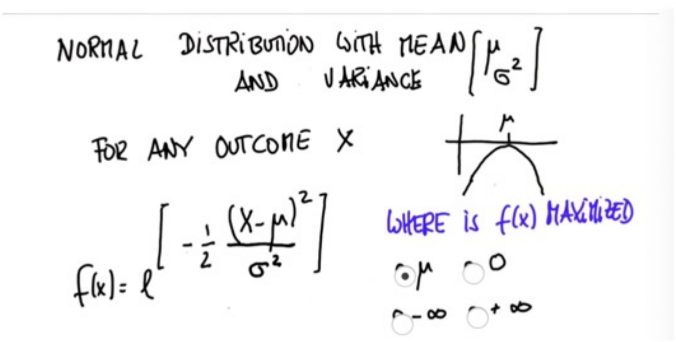
We already know that it's going to flatten it, because you are dividing the f value by half, but are we going to get something like this or perhaps something like this?



08. Maximum

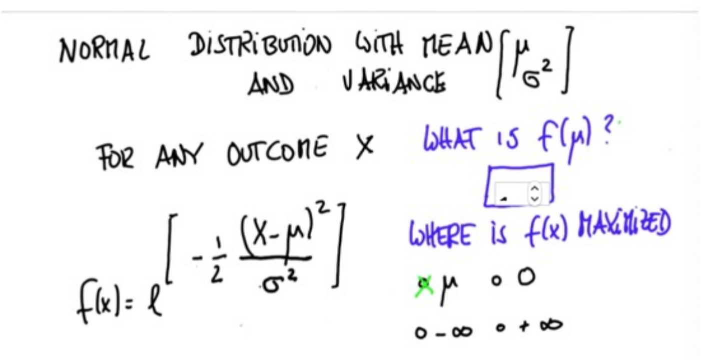
I'm going to make this the exponent of the e function. Remember, the inner argument is a quadratic that points down. This a bit does depend on σ. This mean is μ so I call this `f(x)` where `f(x)` maximize. Where will this thing be the largest?

all the arguments of the exponential are at best 0 and otherwise are negative, because the exponential is **Monotonic(单调的)** that is the larger its argument, the larger its **exponential value**.



09. Maximum Value

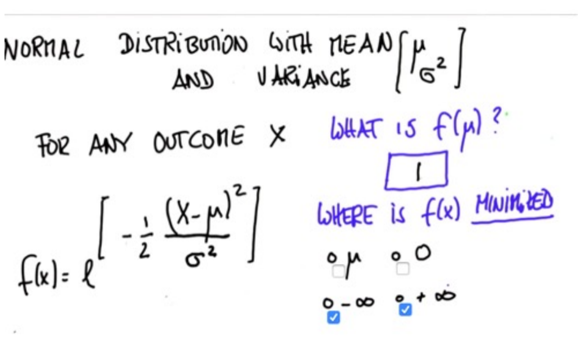
Let me ask you another question. What is the value of this function if we go to the point where it's maximum, which is $x = \mu$? That's the way to write this. Compute for me in your head this what this thing will be when $x = \mu$.



10. Minimum

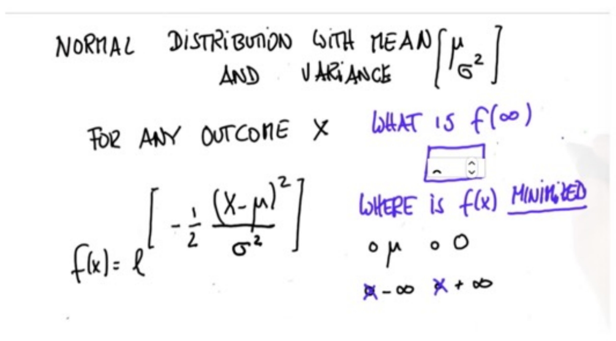
Next I'd like to know where is f(x) minimized?

For what value of x would we get the possible smallest value of this entire expression over here?



11. Minimum Value

In fact, what do you think is the value of this where $x = \infty$?



12. Normalizer

This is what I would consider a relatively simple formula describes the limit of making infinitely many coin flips.

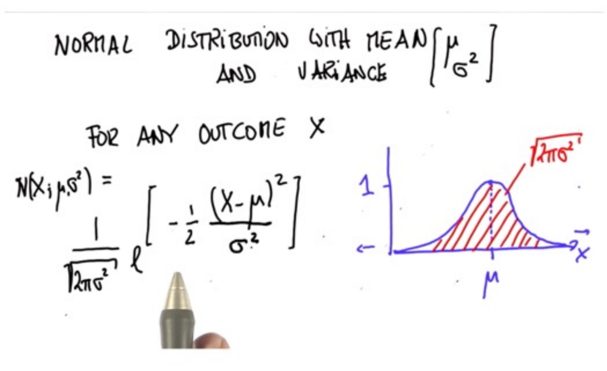
In fact, it describes the limit of computing a mean over any set of experiments.

No matter what you do when you drive n to very large numbers you get a bell curve like this.

There is one flaw here, which is the area underneath this curve doesn't always add up to 1. In fact, without proof, it adds up to $\sqrt{2\pi\sigma^2}$.

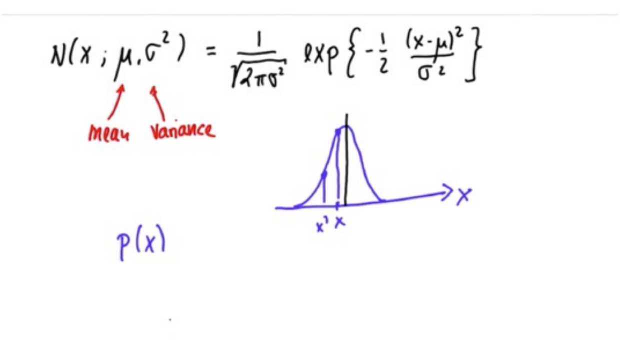
The true normal distribution is normalized by just the inverse of this thing over here $\frac{1}{\sqrt{2\pi\sigma^2}}$.

So, that is the normal distribution of any value x indexed by the parameter μ and σ^2.



13. Formula Summary

What I want to get into your brains is not the complexity of the formula. I want you to really understand how this formula is constructed. I'd want you to understand the quadratic **Penalty Term(惩罚项)** of deviations from the expectation of the mean of this expression. Then the exponential that squeezes it back into the curves.



14. Central Limit Theorem

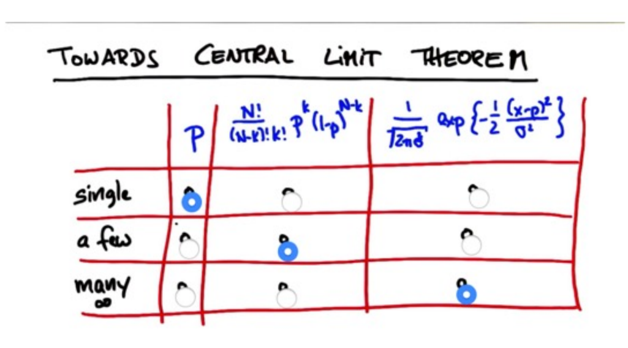
- Single Probability
  P
- **Exponential Families of Distributions(指数族分布)**——Binomial Distribution
  $P(X=x) = \frac{n!}{(k)!(n-k)!} * p^k(1-p)^{(n-k)}$
  $\sigma^2 = \frac{p(1-p)}{N}$
- Exponential Families of Distributions——**Gaussian Exponential(高斯指数)**
  $\frac{1}{\sqrt{2\pi\sigma^2}} exp\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\}$

For 1 coin flip: What is the formula for figuring out the probability of the flip being heads?

For many coin flips: What is the formula for figuring out the probability of a given number of flips being heads?

For infinitely many flips: What is the formula for figuring out the probability of a given proportion of flips being heads?

Note:
If the coin flip is zero to one, then the mean is always in the positive. Gaussian Exponential can assume negative values, but it's an approximation.

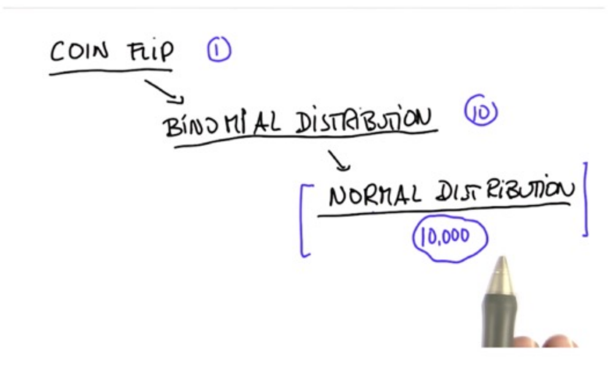


15. Summary

In fact, if you're a medical doctor and you have **one patient**, you might think of it as **a coin flip**.

If you have **10 patients**, you might think of it as **binomial distribution**.

If you do what's normally done, when your test say a new drug and you have, say **10,000 patients** then this thing over here is a beautiful and very compact representation of it.



16. Appendix: Glossary

- Central Limit Theorem(中心极限理论)
- Experiment(实验)
- Triangular Function(三角函数)
- Exponential Families of Distributions(指数族分布)
- Gaussian Exponential(高斯指数)
- Bell Curve(贝尔/钟形曲线)
- Quadratic Expression(二次表达式)
- Quadratic Function(二次函数)
- Monotonic(单调的)
- Penalty Term(惩罚项)