20200210【Data Analyst Nanodegree】P04M01L05

> 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 05: Binomial Distribution

> Learn about one of the most popular distributions in probability – the Binomial Distribution.

01. Binomial

Binomial Distribution(二项分布)

Suppose we do **2 coin flips**. I would like to know how many outcomes of these 2 coin flips are there and which number of heads equals the number of tails.

Truth Table

Flip–1	Flip–2
H	T
T	H

$C(HEADS = TAILS) = 2$

02. Heads Tails

Binomial Distribution

Suppose we do **4 coin flips**. I would like to know how many outcomes of these 4 coin flips are there and which number of heads equals the number of tails.

Truth Table

Flip–1	Flip–2	Flip–3	Flip–4
H	H	T	T
H	T	H	T
H	T	T	H
T	T	H	H
T	H	T	H
T	H	H	T

$C(HEADS = TAILS) = 6$

03. Heads Tails 2

Binomial Distribution

Suppose we do **5 coin flips**. I would like to know how many outcomes of these 5 coin flips are there and which number of heads equals the number of tails.

$C(HEADS = TAILS) = 0$

04. 5 Flips 1 Head

Binomial Distribution

Suppose we do **5 coin flips**. With 5 coin flips, how many outcomes will have exactly 1 heads, hence 4 tails.

$C(\text{1 HEADS and 4 TAILS}) = 5$

To place heads––could be first, second, third, fourth or fifth. So these are 5 different ways.

05. 5 Flips 2 Heads

Binomial Distribution

Suppose we do **5 coin flips**. I would like to know how many outcomes of these 5 coin flips are there and which number of heads equals 2.

$C(\text{2 HEADS and 3 TAILS}) = (5 * 4)/(2 * 1) = 10$

Prove

You could go and place the first heads anywhere in these five element and there's five different ways to place heads. You can now place the second heads among the remaining four––for example, you could place it. over here and it gives you a factor of four different ways of placing the second heads. But when you do this, you over count––you over count exactly by a factor of 2 in the business.

06. 5 Flips 3 Heads

Binomial Distribution

Suppose we do **5 coin flips**. I would like to know how many outcomes of these 5 coin flips are there and which number of heads equals 3.

$C(\text{3 HEADS and 2 TAILS}) = (5 * 4 * 3)/(3 * 2 * 1) = 10$

Prove 1

One is I can just flip heads and tails. So three heads means two tails. I can do the exact same game as before where I placed tails as opposed to heads and it gives me the same equation as before, but let's do it the new way, three heads.

Prove 2

I can place $5 * 4 * 3$––the first heads, the second and the third. For the first, I have five positions, for the second––four, and for the third––three are left.

How much am I over counting?
Well, suppose I'm committed to put the three heads into the three slots over here and that's not given. For the first one placed in here, there's now three different ways of placing it. For the second one, there's two different ways of placing it. For the third one, it's not deterministic––there's just one slot left.

07. 10 Flips 5 Heads

Binomial Distribution

Suppose we do **10 coin flips**. I would like to know how many outcomes of these 10 coin flips are there and which number of heads equals 5.

$C(\text{5 HEADS and 5 TAILS}) = (10 * 9 * 8 * 7 * 6)/(5 * 4 * 3 * 2 * 1) = 252$

08. Formula

Binomial Distribution

Suppose we do **10 coin flips**. I would like to know how many outcomes of these 10 coin flips are there and which number of heads equals 5.

Factorial(阶乘)

$n! = n(n-1)(n-2)\cdots 1$
n = **10 coin flips**
k = **5 HEADS**

$C(\text{k HEADS}) = \frac{n!}{(k)!(n-k)!}$

09. Arrangements

Binomial Distribution

Suppose we do **125 coin flips**. I would like to know how many outcomes of these 125 coin flips are there and which number of heads equals 3.

Factorial

$n! = n(n-1)(n-2)\cdots 1$
n = **125 coin flips**
k = **3 HEADS**

$C(\text{k HEADS}) = \frac{n!}{(k)!(n-k)!} = \frac{125!}{(125-3)!3!} = 317750$

10. Binomial 1

Probability

we have a fair coin with a probability of heads is 0.5.

$P(\text{HEADS EACH TIME}) = 0.5$

$P(\#\text{HEADS} = 1)$

If I flip a coin **5** times, what's the **probability** the number of heads is exactly 1.

$\text{Size of table} = 2^5 = 32$
$C(\text{1 HEADS}) = \frac{5!}{(1)!(5-1)!} = 5$
$P(\text{1 HEADS}) = \frac{Count(\text{1 HEADS})}{\text{Size of table}} = \frac{5}{32} = 0.15625$

11. Binomial 2

Probability

we have a loaded coin with a probability of heads is 0.8.

$P(\text{HEADS EACH TIME}) = 0.8$

$P(\#\text{HEADS} = 1)$

If I flip a coin **3** times, what's the **probability** the number of heads is exactly 1.

Truth Table

Flip–1	Flip–2	Flip–3	Probability
H	T	T	$0.8 * 0.2 * 0.2$
T	H	T	$0.2 * 0.8 * 0.2$
T	T	H	$0.2 * 0.2 * 0.8$

$P(\text{1 HEADS}) = 0.2 * 0.2 * 0.8 * 3 = 0.096$

13. Binomial 4

Probability

we have a loaded coin with a probability of heads is 0.8.

$P(\text{HEADS EACH TIME}) = 0.8$

$P(\#\text{HEADS} = 4)$

If I flip a coin **5** times, what's the **probability** the number of heads is exactly 4.

$C(\text{4 HEADS}) = \frac{5!}{(4)!(5-4)!} = 5$
$P(\text{4 HEADS}) = 5 * (0.8)^4 * (0.2)^1 = 0.4096$

14. Binomial 5

Probability

we have a loaded coin with a probability of heads is 0.8.

$P(\text{HEADS EACH TIME}) = 0.8$

$P(\#\text{HEADS} = 3)$

If I flip a coin **5** times, what's the **probability** the number of heads is exactly 3.

$C(\text{3 HEADS}) = \frac{5!}{(3)!(5-3)!} = 10$
$P(\text{3 HEADS}) = 10 * (0.8)^3 * (0.2)^2 = 0.2048$

15. Binomial 6

Probability

we have a loaded coin with a probability of heads is 0.8.

$P(\text{HEADS EACH TIME}) = 0.8$

$P(\#\text{HEADS} = 9)$

If I flip a coin **12** times, what's the **probability** the number of heads is exactly 9.

$C(\text{9 HEADS}) = \frac{12!}{(9)!(12-9)!} = 220$
$P(\text{9 HEADS}) = 220 * (0.8)^9 * (0.2)^3 = 0.236$

16. Binomial Conclusion

Binomial Distribution Conclusion

In this lesson, we found the **probability** that A coin would land on heads k times out of n flips. If the **probability of heads** for the coin is p, we get the following **formula** for the probability that the number of heads will be k.

$P(\text{Binomial Distribution}) = \frac{n!}{(k)!(n-k)!} * p^k(1-p)^{(n-k)}$

First part keeps track of the **total number of ways** we can get k heads and n total flips of the coin.

$C(\#\text{K Heads}) = \frac{n!}{(k)!(n-k)!}$

Second Part keeps track of each one of these ways have **probability of even occurring**.

$P(\#\text{K Heads}) = p^k(1-p)^{(n-k)}$

Although the examples in this lesson use **coin flips**, you can really perform these calculations with **any events that have two outcomes**. Does a **customer** buy or not? Is a **transaction** fraud or not? Or if a **coin flip** is heads or tails. Binomial distributions are used to give us insight about all of these.

17. Text: Recap + Next Steps

The Binomial Distribution

The **Binomial Distribution** helps us determine the probability of a string of independent 'coin flip like events'.

The probability mass function(Probability Mass Function(概率质量函数,PMF)) associated with the binomial distribution is of the following form:

$P(X = x) = \frac{n!}{(k)!(n-k)!} * p^k(1-p)^{(n-k)}$

where **n** is the number of events, **x** is the number of "successes", and **p** is the probability of "success".

We can now use this distribution to determine the probability of things like:

- The probability of 3 heads occurring in 10 flips.
- The probability of observing 8 or more heads occurring in 10 flips.
- The probability of not observing any heads in 20 flips.

Looking Ahead

The truth is that in practice, you will commonly be working with data, which might follow a binomial distribution. So it is **less important** to **calculate these probabilities** (though this can be useful in some cases), and it is **more important** that you **understand what the Binomial Distribution is used for**, as it shows up in a lot of modeling techniques in machine learning, and it can sneak up in our datasets with tracking any outcome with two possible events. You will get some practice with this in the Python Probability Practice lesson.

One of the most popular places you see the **Binomial distribution** is in **logistic regression**, which you will learn about in the last lesson of this statistics course.

In the next section, you will begin to work with **events that aren't independent**. The events we have seen so far haven't influenced one another, but it turns out the real world is usually more complicated than this. The next section will introduce the idea of dependence, and you will learn even more with Bayes rule in the following section.

18. Appendix: Gloassary

Binomial Distribution(二项分布)
Factorial(阶乘)
Probability Mass Function(概率质量函数,PMF)