20200210【Data Analyst Nanodegree】P04M01L06

> 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

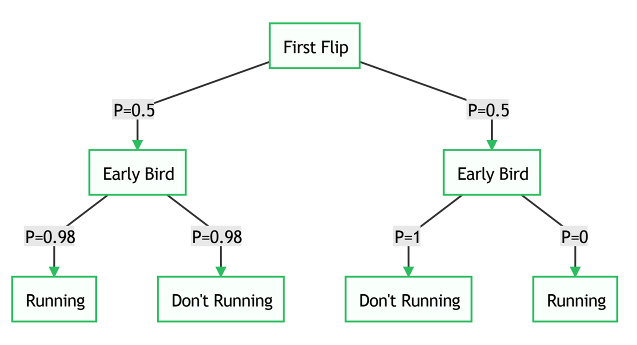
Lesson 06: Conditional Probability

> Not all events are independent. Learn the probability rules for dependent events.

01. Introduction to Conditional Probability

Case

In real life, events often depend on each other. Say you can be an early bird or night owl. And for the sake of simplicity, let's assume there's a 50 percent chance for each one. Now whether you decide to go for a run at 5:00 a.m. tomorrow is not entirely independent. If you're an early bird, the probability might be a two percent chance, while if you're a night owl, the probability might be zero.



02. Medical Example 1

Medical Example

Supposed there's a patient in the hospital who might suffer from a medical condition like cancer. Let's say the probability of having this cancer is 0.1. That means you can tell me what's the probability of being cancer free.

$P(CANCER) = 0.1$
$P(\neg CANCER) = 0.9$

03. Medical Example 2

$P(CANCER) = 0.1$
$P(\neg CANCER) = 0.9$

In reality, we don't know whether a person suffers cancer, but we can run a test like **a blood test**. The outcome of it blood test may be positive or negative, but like any good test, it tells me something about the thing I really care about--whether the person has cancer or not.

Let's say, if the person has the cancer, the test comes up **positive** with the probability of 0.9, and that implies if the person has cancer, the **negative** outcome will have 0.1 probability and that's because these two things have to add to 1.

We call this thing over here a **Conditional Probability(条件概率)**, and the way to understand this is a very funny notation. There's a **bar** in the middle, and the bar says what's the probability of the stuff on the left given that we assume the stuff on the right is actually the case.

$P(Positive|CANCER) = 0.9$
$P(Negative|CANCER) = 0.1$
$P(Positive|\neg CANCER) = 0.2$
$P(Negative|\neg CANCER) = 0.8$

04. Medical Example 3

This is very nontrivial but armed with this, we can now build up the **truth table** for all the cases of the two different variables, cancer and non-cancer and positive and negative tests outcome.

So, please give me the **probability of the combination** of those for the very first one, and as a hint, it's kind of the same as before where we multiply two things, but you have to find the right things to multiple in this table over here.

$P(CANCER) = 0.1$
$P(\neg CANCER) = 0.9$
$P(Positive|CANCER) = 0.9$
$P(Negative|CANCER) = 0.1$
$P(Positive|\neg CANCER) = 0.2$
$P(Negative|\neg CANCER) = 0.8$

True Table

Cancer	Test	P
Y	P	0.09
Y	N	
N	P	
N	N	

05. Medical Example 4

$P(CANCER) = 0.1$
$P(\neg CANCER) = 0.9$
$P(Positive|CANCER) = 0.9$
$P(Negative|CANCER) = 0.1$
$P(Positive|\neg CANCER) = 0.2$
$P(Negative|\neg CANCER) = 0.8$

True Table

Cancer	Test	P
Y	P	0.09
Y	N	0.01
N	P	
N	N	

06. Medical Example 5

$P(CANCER) = 0.1$
$P(\neg CANCER) = 0.9$
$P(Positive|CANCER) = 0.9$
$P(Negative|CANCER) = 0.1$
$P(Positive|\neg CANCER) = 0.2$
$P(Negative|\neg CANCER) = 0.8$

True Table

Cancer	Test	P
Y	P	0.09
Y	N	0.01
N	P	0.18
N	N	

07. Medical Example 6

$P(CANCER) = 0.1$
$P(\neg CANCER) = 0.9$
$P(Positive|CANCER) = 0.9$
$P(Negative|CANCER) = 0.1$
$P(Positive|\neg CANCER) = 0.2$
$P(Negative|\neg CANCER) = 0.8$

True Table

Cancer	Test	P(AB)
Y	P	0.09
Y	N	0.01
N	P	0.18
N	N	0.72

08. Medical Example 7

$P(CANCER) = 0.1$
$P(\neg CANCER) = 0.9$
$P(Positive|CANCER) = 0.9$
$P(Negative|CANCER) = 0.1$
$P(Positive|\neg CANCER) = 0.2$
$P(Negative|\neg CANCER) = 0.8$

True Table

Cancer	Test	$P(AB) = 1$
Y	P	0.09
Y	N	0.01
N	P	0.18
N	N	0.72

09. Medical Example 8

$P(CANCER) = 0.1$
$P(\neg CANCER) = 0.9$
$P(Positive|CANCER) = 0.9$
$P(Negative|CANCER) = 0.1$
$P(Positive|\neg CANCER) = 0.2$
$P(Negative|\neg CANCER) = 0.8$

True Table

Cancer	Test	$P(AB) = 1$
Y	P	0.09
Y	N	0.01
N	P	0.18
N	N	0.72

$P(PositiveResult) = 0.09 + 0.18 = 0.27$

10. Total Probability

Total Probability

$P(CANCER) \rightarrow P(\neg CANCER)$
$P(Positive|CANCER) \rightarrow P(Negative|CANCER)$
$P(Positive|\neg CANCER) \rightarrow P(Negative|\neg CANCER)$

You complete the probability of a positive test result as the sum of a positive test result given cancer times the probability of cancer, which is our truth table entry for the combination of P and C plus the same given we don't have of cancer.

Law of total probability(全概率公式):

$P(P) = P(P|C) \cdot P(C) + P(P|\neg C) \cdot P(\neg C)$

11. Two Coins 1

This time around, we have a bag, and in the bag are 2 coins,coin 1 and coin 2. And in advance, we know that coin 1 is fair. So P of coin 1 of coming up heads is 0.5 whereas coin 2 is loaded, that is, P of coin 2 coming up heads is 0.9.

$P_1(H) = 0.5$
$P_1(T) = 1 - 0.5 = 0.5$
$P_2(H) = 0.9$
$P_2(T) = 1 - 0.9 = 0.1$

12. Two Coins 2

$P_1(H) = 0.5$
$P_1(T) = 0.5$
$P_2(H) = 0.9$
$P_2(T) = 0.1$

You have a pick event followed by a flip event We can pick coin 1 or coin 2. Then we can flip and get heads or tails for the coin we've chosen.

$P(1) = 0.5$
$P(2) = 0.5$

True Table

PICK	FLIP	P
1	H	0.25
1	T	0.25
2	H	0.45
2	T	0.05

$P(H) = 0.5 \cdot 0.5 + 0.9 \cdot 0.5 = 0.7$

13. Two Coins 3

$P_1(H) = 0.5$
$P_1(T) = 0.5$
$P_2(H) = 0.9$
$P_2(T) = 0.1$

You are picking only one coin from the bag, then flipping that one coin twice and observing heads then tails.

$P(1) = 0.5$
$P(2) = 0.5$
$P(H) = 0.5 \cdot 0.5 + 0.9 \cdot 0.5 = 0.7$
$P(T) = 0.5 \cdot 0.5 + 0.1 \cdot 0.5 = 0.3$

True Table

PICK	FLIP-Once	FLIP-Twice	P
1	H	H	
1	H	T	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$
1	T	H	
1	T	T	
2	H	H	
2	H	T	$0.5 \cdot 0.9 \cdot 0.1 = 0.045$
2	T	H	
2	T	T	

$P(H, T) = 0.125 + 0.045 = 0.17$

14. Two Coins 4

$P(1) = 0.5$
$P(2) = 0.5$

$P(H|1) = 1$
$P(H|2) = 0.6$
$P(T|1) = 0$
$P(T|2) = 0.4$

You choose a coin and then flip it twice. The coin is not put back and chosen again in between flips.

$P(T, T) = 0 + 0.08 = 0.08$

True Table

PICK	FLIP-Once	FLIP-Twice	P
1	H	H	
1	H	T	
1	T	H	
1	T	T	$0.5 \cdot 0 \cdot 0 = 0$
2	H	H	
2	H	T	
2	T	H	
2	T	T	$0.5 \cdot 0.4 \cdot 0.4 = 0.08$

15. Summary

The key thing is we talked about conditional probabilities. We said that the outcome in a variable, like a test is actually not like the random coin flip but it depends on something else, like a disease.

$P(TEST) = P(TEST|DISEASE) \cdot P(DESEASE) + P(TEST|\neg DISEASE) \cdot P(\neg DISEASE)$

16. Text: Summary

Conditional Probability

In this lesson you learned about conditional probability. Often events are not independent like with coin flips and dice rolling. Instead, the outcome of one event depends on an earlier event.

For example, the probability of obtaining a positive test result is dependent on whether or not you have a particular condition. If you have a condition, it is more likely that a test result is positive. We can formulate conditional probabilities for any two events in the following way:

$P(A|B) = \frac{P(A \bigcap B)}{P(B)}$

In this case, we could have this as:

$P(positive|disease) = \frac{P(positive \bigcap disease)}{P(disease)}$

where $|$ represents "given" and $\bigcap$ represents "and".

Looking Ahead

You will get more practice with conditional probability using Bayes rule in the lesson. If you are comfortable with the examples here, the next lesson should be a breeze. And if you are still feeling a bit uncomfortable with these ideas, the next lesson should be good practice to reinforce the topics here with some more examples.

17. Appendix: Glossary

- Conditional Probability(条件概率)
- Law of total probability(全概率公式)