4.1 Bayes' Rule

Zongyi Liu

2023-06-08

4.1 Bayes' Rule

Bayes' Rule was invented by Thomas Bayes, a mathematician from England. The simple rule that relates conditional probabilities has vast ramifications for statistical inference, and therefore as long as his name is attached to the rule, we'll continue to see his name in textbooks.

There is another branch of statistics, called null hypothesis significance testing (NHST), which relies on the probability of data given the model and does not use Bayes' rule.

4.1.1 Derived From Definitions of Conditional Probability

Based from the definition of conditional probability, which is, $p(y|x)p(x) =$, we can get the new expression:

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}.$$

Since we know the relationship between conditional probability and sum, we can get another expression

$$p(y|x) = \frac{p(x|y)p(y)}{\sum_y p(x|y)p(y)}.$$

4.1.2 Intuited From a Two-way Discrete Table

It is known to us that

# Switches	**# Heads** 0	**1**	**2**	**3**	**Marginal (# Switches)**
0	1/8	0	0	1/8	2/8
1	0	2/8	2/8	0	4/8
2	0	1/8	1/8	0	2/8
Marginal (# Heads):	1/8	3/8	3/8	1/8	

Thus we can arrive at a new table based on Bayes' rule not merely special but also spatial

Row	Column			Marginal
	$\cdots$	**j**	$\cdots$	
$\vdots$		$\vdots$		
i	$\cdots$	$p(R_i, C_j)$ $= p(R_i \mid C_j) p(C_j)$ $= p(C_j \mid R_i) p(R_i)$	$\cdots$	$p(R_i)$
$\vdots$		$\vdots$		
Marginal:		$p(C_j)$		

Suppose we know that event R_i has happened, but we don't know the column value. In this case, the remaining possibilities are the cells in row R_i, and therefore we can restrict our attention to only the i-th row of Table 4.1. Because we know that R_i is true, our universe of remaining possibilities has collapsed to that row, and therefore we know that the sum of the probabilities in the row must be 1, instead of $p(R_i)$.

4.1.3 The Denominator as Integral over Continuous Values

For continuous variables, we can get the Bayes' rule as

$$p(y|x) = \frac{p(x|y)p(y)}{\int dy\, p(x|y)p(y)}.$$

In this equation, the y in the numerator is a specific fixed value, whereas the y in the denominator is a variable that takes on all possible values of y over the integral.