362 (conditional probability) Bayes defined conditional probability, using his own notation, as follows:

$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)}$$

This is to be read "the probability that A is true given that B is true is equal to the probability that both are true divided by the probability that B is true". What is wrong with Bayes' notation, and how is conditional probability expressed in our notation?

After trying the question, scroll down to the solution.

§ The notation $P(A)$ meaning "the probability that A is true" is not transparent. For example, let d be the result of throwing a 6-sided die. It is standard to write

$$P(d=3) = 1/6$$

to mean "the probability that the die shows 3 is 1/6 ". When you throw the die, you get a value for d . Suppose $d=2$. Then, according to the transparent law,

$$P(2=3) = 1/6$$

which is nonsense. Suppose $d=3$. Then

$$P(3=3) = 1/6$$

which is also nonsense. As another example, suppose variable q can be 1 , 2 , or 3 , with probabilities (in standard notation)

$$P(q=1) = 1/6$$
$$P(q=2) = 1/3$$
$$P(q=3) = 1/2$$

Now suppose $q=1$. Then $q=2$ and $q=3$ are equal (both *false*). Therefore, by the transparent law, $P(q=2)$ and $P(q=3)$ are equal. But they aren't equal.

A binary expression B may not be a distribution. It becomes a distribution if we divide it by its sum. For example, let x be the only variable. Then $B / (\Sigma x\cdot B)$ is the distribution proportional to B . Define

$$\updownarrow B = B / (\Sigma x\cdot B)$$

We pronounce $\updownarrow B$ as "normalize B ". We can now express Bayes' conditional probability $P(A \mid B)$ as

$$\updownarrow B'. A$$

This describes the situation where we learn that B is true and then ask if A is true. This situation is just one of infinitely many situations for which we may want to calculate a probability. We cannot make a special notation for each one, as Bayes has done for conditional probability. We need to be able to describe the situation using a basic set of connectives, and from that description, calculate probabilities, as we do.

To prove that we have described Bayes' conditional probability, again let x be the only variable, and let n be the size of its domain. Then

	$\updownarrow B'. A$	use definition of $\updownarrow$
$=$	$B' / (\Sigma x'\cdot B').\ A$	use definition of .
$=$	$\Sigma x''\cdot\ B'' / (\Sigma x'\cdot B') \times A''$	rearrange and rename local variables
$=$	$(\Sigma x\cdot A \times B) / (\Sigma x\cdot B)$	divide numerator and denominator each by n
$=$	$\dfrac{(\Sigma x\cdot A \times B) / n}{(\Sigma x\cdot B) / n}$	switch to Bayes' probability notation
$=$	$\dfrac{P(A \wedge B)}{P(B)}$	use Bayes definition of conditional probability
$=$	$P(A \mid B)$	