James Green-Armytage, econ 229, fall 2015

Bayes's Rule

1. Foundations

Suppose that A and B are events that may occur. Define $\mathrm{P}(A)$ as the probability that A will occur, and $\mathrm{P}(B)$ as the probability that B will occur. Define $\mathrm{P}(A \cup B)$ as the probability that A, B, or both A and B will occur. Define $\mathrm{P}(A \cap B)$ as the probability that both A and B will occur. In this section we state and discuss formulae for $\mathrm{P}(A \cup B)$ and $\mathrm{P}(A \cap B)$, in terms of $\mathrm{P}(A)$ and $\mathrm{P}(B)$.

First, we state an identity for $\mathrm{P}(A \cup B)$:

$$\boxed{\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)}$$

If A and B are disjoint, i.e. $\mathrm{P}(A \cap B) = 0$, this simplifies to $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$.

Example 1: When drawing a card from a standard 52-card deck, what is the probability of drawing a king or a spade? These events are not disjoint, because it is possible to draw a king of spades. So, the probability of drawing a king is $\mathrm{P}(king) = \frac{4}{52} = \frac{1}{13}$, the probability of drawing a spade is $\mathrm{P}(spade) = \frac{13}{52} = \frac{1}{4}$, and the probability of drawing a king or a spade is $\mathrm{P}(king) + \mathrm{P}(spade) - \mathrm{P}(king \cap spade) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{17}{52}$.

Example 2: When drawing a card from a standard 52-card deck, what is the probability of drawing a face card or an ace? These events are disjoint, because no card is both an ace and a face card. So the probability of drawing a face card is $\mathrm{P}(face) = \frac{12}{52} = \frac{3}{13}$, the probability of drawing an ace is $\mathrm{P}(ace) = \frac{4}{52} = \frac{1}{13}$, the probability of drawing a face card or an ace is $\mathrm{P}(face) + \mathrm{P}(ace) = \frac{3}{13} + \frac{1}{13} = \frac{4}{13}$.

Second, we state an identity for $\mathrm{P}(A \cap B)$:

$$\boxed{\mathrm{P}(A \cap B) = \mathrm{P}(A|B) \cdot \mathrm{P}(B) = \mathrm{P}(B|A) \cdot \mathrm{P}(A)}$$

If A and B are independent, i.e. if $\mathrm{P}(A|B) = P(A)$ and $P(B|A) = P(B)$, then $\mathrm{P}(A \cap B) = \mathrm{P}(A) \cdot \mathrm{P}(B)$.

Example 3: When drawing a card from a standard 52-card deck, what is the probability of drawing the king of spades? One can come immediately to the answer $\frac{1}{52}$, but it is useful to verify that the formula gives the same result. In this case, the events are independent, because the card being a king doesn't change the odds that the card is a spade, or vice versa. So we can calculate $\mathrm{P}(king \cap spade) = \mathrm{P}(king) \cdot \mathrm{P}(spade) = \frac{1}{13} \cdot \frac{13}{52} = \frac{1}{52}$.

Example 4: When drawing a card from a deck that is missing the king of clubs, what is the probability of drawing the king of spades? Again, one can come immediately to the answer $\frac{1}{51}$, but it is useful to verify that the formula gives the same result. In this case the events are not independent, because if the card is a king, we know it is more likely than otherwise to be a spade, because it can't be a club. Likewise, we know that if the card is a spade, it is more likely than otherwise to be a king. So, we can calculate $\mathrm{P}(king \cap spade) = \mathrm{P}(king|spade) \cdot \mathrm{P}(spade) = \frac{1}{13} \cdot \frac{13}{51} = \frac{1}{51}$. Or, we can calculate $\mathrm{P}(king \cap spade) = \mathrm{P}(spade|king) \cdot \mathrm{P}(king) = \frac{1}{3} \cdot \frac{3}{51} = \frac{1}{51}$.

2. The formula in two equalities

Let A_* be a possible state of a random variable A. Suppose that we aren't able to observe directly whether A_* has occurred, but we are interested in estimating the probability with which it has occurred. Further, suppose that the random variable B is influenced by the random variable A, suppose that we've observed that B has taken on state B_o, and suppose that we know the conditional probabilities $B_o|A_i$, where the n distinct A_is represent all of the states that the variable A can take on.

Bayes's rule states that

$$\boxed{\mathrm{P}(A_*|B_o) = \frac{\mathrm{P}(A_* \cap B_o)}{\mathrm{P}(B_o)} = \frac{\mathrm{P}(B_o|A_*)\mathrm{P}(A_*)}{\sum_{i=1}^{n} \mathrm{P}(B_o|A_i) \cdot \mathrm{P}(A_i)}}$$

That is, we are interested in the probability that A has taken on state A_*, conditional on the fact that B has taken on the observed state B_o, but we don't have direct access to the conditional probability $A_*|B_o$. However, if we know the prior probabilities $\mathrm{P}(A_i)$, and the conditional probabilities $\mathrm{P}(B_o|A_i)$, Bayes's rule allows us to work backwards from these to assess the conditional probability we are interested in.

The rule as stated above includes two equalities, which we will derive in turn in the next two sections.

3. The first equality

Here we derive the first equality, $\boxed{\mathrm{P}(A_*|B_o) = \frac{\mathrm{P}(A_* \cap B_o)}{\mathrm{P}(B_o)}}$.

By the formula for $\mathrm{P}(A \cap B)$ above, we have $\mathrm{P}(A_* \cap B_o) = \mathrm{P}(A_*|B_o) \cdot \mathrm{P}(B_o)$.

We can divide both sides by $\mathrm{P}(B_o)$ to obtain $\mathrm{P}(A_*|B_o) = \frac{\mathrm{P}(A_* \cap B_o)}{\mathrm{P}(B_o)}$.

This, perhaps, is the most intuitive version of Bayes's rule. In simple examples, it is possible to just calculate $\mathrm{P}(A_* \cap B_o)$ and $\mathrm{P}(B_o)$ directly, which means that one does not necessarily have to bother with anything beyond this in order to solve them. Nonetheless, we derive the second equality below, and thus the more explicit version of the rule, in the interest of completeness.

4. The second equality

Here we derive the second equality, $\boxed{\frac{\mathrm{P}(A_* \cap B_o)}{\mathrm{P}(B_o)} = \frac{\mathrm{P}(B_o|A_*)\mathrm{P}(A_*)}{\sum_{i=1}^{n} \mathrm{P}(B_o|A_i) \cdot \mathrm{P}(A_i)}}$.

By the axiom $\mathrm{P}(A \cap B) = \mathrm{P}(B|A) \cdot \mathrm{P}(A)$, it follows immediately that $\mathrm{P}(A_* \cap B_o) = \mathrm{P}(B_o|A_*)\mathrm{P}(A_*)$. So the numerators are the same.

Further, we can obtain the identity $\mathrm{P}(B_o) = \sum_{i=1}^{n} \mathrm{P}(A_i \cap B_o)$ via the understanding that the probability of B taking on state B_o is equal to the sum of all the mutually disjoint probabilities $\mathrm{P}(A_i \cap B_o)$, i.e. the probabilities with which B_o can coincide with each of the disjoint states of A. Combining this with the identity $\mathrm{P}(A_i \cap B_o) = \mathrm{P}(B_o|A_i) \cdot \mathrm{P}(A_i)$, we obtain $\mathrm{P}(B_o) = \sum_{i=1}^{n} \mathrm{P}(B_o|A_i) \cdot \mathrm{P}(A_i)$. So the denominators are the same, and the equality holds.