probability

Thursday, May 18, 2017 5:02 PM

frequentist vs bayesian

A Bayesian defines a "probability" in exactly the same way that most non-statisticians do - namely an indication of the plausibility of a proposition or a situation. If you ask him a question, he will give you a direct answer assigning probabilities describing the plausibilities of the possible outcomes for the particular situation (and state his prior assumptions).

A Frequentist is someone that believes probabilities represent long run frequencies with which events occur; if needs be, he will invent a fictitious population from which your particular situation could be considered a random sample so that he can meaningfully talk about long run frequencies. If you ask him a question about a particular situation, he will not give a direct answer, but instead make a statement about this (possibly imaginary) population. Many non-frequentist statisticians will be easily confused by the answer and interpret it as Bayesian probability about the particular situation.

However, it is important to note that most Frequentist methods have a Bayesian equivalent that in most circumstances will give essentially the same result, the difference is largely a matter of philosophy, and in practice it is a matter of "horses for courses".

From <https://stats.stackexchange.com/questions/22/bayesian-and-frequentist-reasoning-in-plain-english>

random variable

discrete

The expression $p(A)$ denotes the probability that the event A is true. For example, A might be the logical expression "it will rain tomorrow". We require that $0 \leq p(A) \leq 1$, where $p(A) = 0$ means the event definitely will not happen, and $p(A) = 1$ means the event definitely will happen. We write $p(\overline{A})$ to denote the probability of the event not A; this is defined to $p(\overline{A}) = 1 - p(A)$. We will often write $A = 1$ to mean the event A is true, and $A = 0$ to mean the event A is false.

We can extend the notion of binary events by defining a **discrete random variable** X, which can take on any value from a finite or countably infinite set $\mathcal{X}$. We denote the probability of the event that $X = x$ by $p(X = x)$, or just $p(x)$ for short. Here $p()$ is called a **probability mass function** or **pmf**. This satisfies the properties $0 \leq p(x) \leq 1$ and $\sum_{x \in \mathcal{X}} p(x) = 1$.

- event probability can be thought of as a binary discrete random variable

continuous

for continuous we have **probability density function**

distribution

formally, p(X) denotes whole distribution, while p(X=x) denotes concrete value

combining variables

Probability of a union of two events

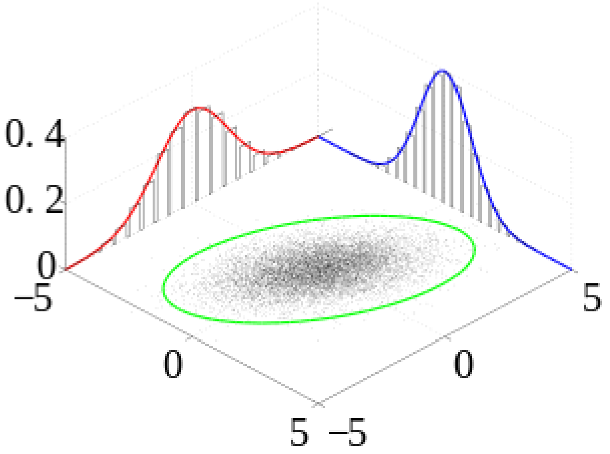
Given two events, A and B, we define the probability of A or B as follows:

$$\begin{aligned} p(A \vee B) &= p(A) + p(B) - p(A \wedge B) \\ &= p(A) + p(B) \text{ if } A \text{ and } B \text{ are mutually exclusive} \end{aligned}$$

Joint probabilities

We define the probability of the joint event A and B as follows:

$p(A, B) = p(A \wedge B) = p(A|B)p(B)$



3 different probabiities are involved in this formula:

- marginal probability - P(B)
- conditional probability - P(A|B)
- joint probability - P(A,B)

Joint vs conditional

we calculate B assuming that it happened - but if we want probability of both we take into account the fact that B also has certain probability of happening, so **P(A|B) >= P(A,B)**

high-school formulae for conditional probability

P(A|B) = P(A,B)/P(B)

we assume that B has 100% chance of happening (P(B) <= 1f, division increases value)

correclated/uncorrelated

if variables are not correlated, P(A|B) = P(A) (should be larger if correlated)

then P(A,B)=P(A)P(B)

after inverting

P(A|B) is usually different than P(B|A)

but P(A,B) should be the same as P(B,A)

P(A|B)P(B) = P(B|A)P(A)

trival if uncorrelated

generalization for more distribution

$$p(X_{1:D}) = p(X_1)p(X_2|X_1)p(X_3|X_2, X_1)p(X_4|X_1, X_2, X_3) \ldots p(X_D|X_{1:D-1}) \tag{2.5}$$

(chain rule)

marginal distribution

This is sometimes called the **product rule**. Given a **joint distribution** on two events $p(A, B)$, we define the **marginal distribution** as follows:

$$p(A) = \sum_b p(A, B) = \sum_b p(A|B = b)p(B = b) \tag{2.4}$$

bayes theorem

Combining the definition of conditional probability with the product and sum rules yields **Bayes rule**, also called **Bayes Theorem**[2]:

$$p(X = x|Y = y) = \frac{p(X = x, Y = y)}{p(Y = y)} = \frac{p(X = x)p(Y = y|X = x)}{\sum_{x'} p(X = x')p(Y = y|X = x')} \tag{2.7}$$

in numerator we use P(A,B) = P(B,A)
in denominator marginal distribution

base rate fallacy

As an example of how to use this rule, consider the following medical diagonsis problem. Suppose you are a woman in your 40s, and you decide to have a medical test for breast cancer called a **mammogram**. If the test is positive, what is the probability you have cancer? That obviously depends on how reliable the test is. Suppose you are told the test has a **sensitivity**

of 80%, which means, if you have cancer, the test will be positive with probability 0.8. In other words,

$$p(x = 1|y = 1) = 0.8 \tag{2.8}$$

where $x = 1$ is the event the mammogram is positive, and $y = 1$ is the event you have breast cancer. Many people conclude they are therefore 80% likely to have cancer. But this is false! It ignores the prior probability of having breast cancer, which fortunately is quite low:

$$p(y = 1) = 0.004 \tag{2.9}$$

Ignoring this prior is called the **base rate fallacy**. We also need to take into account the fact that the test may be a **false positive** or **false alarm**. Unfortunately, such false positives are quite likely (with current screening technology):

$$p(x = 1|y = 0) = 0.1 \tag{2.10}$$

Combining these three terms using Bayes rule, we can compute the correct answer as follows:

$$\begin{aligned} p(y = 1|x = 1) &= \frac{p(x = 1|y = 1)p(y = 1)}{p(x = 1|y = 1)p(y = 1) + p(x = 1|y = 0)p(y = 0)} && (2.11) \\ &= \frac{0.8 \times 0.004}{0.8 \times 0.004 + 0.1 \times 0.996} = 0.031 && (2.12) \end{aligned}$$

where $p(y = 0) = 1 - p(y = 1) = 0.996$. In other words, if you test positive, you only have about a 3% chance of actually having breast cancer![3]

todo

negation
inverse in conditional and joint
marginal distribution