… finishing Bayes' (from last lecture)

1. Bayes' Theorem in terms of odds

1.1. Definition: Bayes' factor is $\frac{P(B|A)}{P(B|A^c)}$

It quantifies the empirical evidence provided by the data (B) in favor of A. If it is > 1, it increases the odds of A. If < 1, it decreases the odds of A. The posterior odds are the prior odds multiplied by this Bayes' factor.

1.2. Definition: odds$(A) = \frac{P(A)}{P(A^c)}$

Using these two we can reexpress Bayes' Theorem as:

$$\frac{P(A|B)}{P(A^c|B)} = \frac{P(A)}{P(A^c)} \cdot \frac{P(B|A)}{P(B|A^c)}$$

I.e. we solve for LHS by multiplying the odds by Bayes' factor.

Independence

2. Definition: A and B are independent if any of the following is true:

1. $P(A|B) = P(A)$
2. $P(A|B) = P(A|B^c)$
3. $P(B|A) = P(B)$
4. $P(A \cap B) = P(A) \cdot P(B)$
   If any of those are true, they are all true.

2.3. Remark: If A and B are independent, $P(A \cap B) = P(A) \times P(B)$.

This is an extension of the multiplication law and we can generalise to an arbitrary number of events.

2.4. Remark: If $P(A) > 0$ and $P(B) > 0$, independent events cannot be disjoint and disjoint events cannot be independent.

3. Counter-argument to Bayes'

Suppose there have been 1000000 instances in which a miracle could have occurred and no miracle occurred. Consider the sun rising tomorrow and a baby with uniform prior on p.

By Bayes', the chance of $P\left(p > \frac{1}{1600000} | X = 0\right) \approx 0.535$. But if $p > \frac{1}{1600000}$ the probability that there is a real miracle in the next 1000000 trials is, by multiplication rule, greater than $1 - \left(1 - \frac{1}{1600000}\right)^1 000000 \approx 0.465$—clearly absurd.

Random Variables

So far we have considered the probabilities for events, subsets of a sample space. But sample spaces are often very complicated, e.g. HHTTHTHTTHHTTTT, so this is unwieldy. We usually care more about specific numerical properties associated with an outcome, e.g. # of tosses to get first heads. We call this a random variable.

Formally speaking, a random variable is a real-valued function on the sample space Ω mapping elements of Ω, ω, to real numbers, i.e. $\Omega \to \mathbb{R}$ as $\omega \to x = X(\omega)$.

We have two types of random variables: discrete and continuous.

4. Probability Mass Function

The PMF of a random variable X is a function $p(x)$ that maps each possible value x_i to the corresponding probability $P(X = x_i)$. In particular, A PMF $p(x)$ must satisfy $0 \leq p(x) \leq 1$ and $\sum_x p(x) = 1$.

4.5. Bernoulli Distribution

A random variable that can only take two values, 0 and 1, with probabilities $1 - p$ and p, respectively, is called a Bernoulli random variable. Its PMF is thus $p(1) = p$, $p(0) = 1 - p$, and $p(x) = 0$, if $x \neq 0$ or 1. Such a distribution is called Bernoulli distribution with parameter p.

We use this for random trials having only two possible outcomes, e.g. coin flips, whether a drug works, whether a subject answers yes or no.

4.6. Binomial Distribution

Suppose n independent Bernoulli trials are to be performed, each of which results in a success with probability p and a failure with probability $1 - p$. We define X as the number of successes obtained in the n trials. We say X has a binomial distribution with parameters $X \sim Bin(n,p)$ with the PMF $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$.

4.7. Remark: the sum of i.i.d Bernoulli Random Variables is Binomial

4.8. Geometric Distribution

Suppose that a sequence of independent Bernoulli trials are performed, each with probability of success p. Let X be the number of trials required to obtain the first success. The PMF of x is $p(k)=P(X=k)=(1-p)^{k-1}p$ for $k \in \{1,2,3,\ldots\}$. If $k \geq 1$ and $P(X=0)=0$, we denote as $X \sim \text{eometric}(p)$. We say that X has a geometric distribution, since the PMF is a geometric sequence.

4.9. Negative Binomial Distribution

Suppose that a sequence of independent Bernoulli trials are performed, each with probability of success p. Let X be the number of trials required to obtain the th success. For the event $\{X=k\}$ to occur, the kth trial must be a success and the first $k-1$ trials can be -1 successes and $k-$ failures in any order.

Thus, the negative binomial PMF is $P(X=k)=\binom{k-1}{-1}p(1-p)^{k-}$, denoted as $X \sim NB(,p)$.

4.10. Relationship between Negative Binomial & Geometric

If $X_1, X_2, \ldots, X$ are i.i.d. Geometric(p) random variables, then $X_1+X_2+\ +X \sim \text{B}(,p)$. Conversely, let X_1 be the number of trials needed to get the first success, X_2 be the number of additional trials needed to get the second success after the first success, . . ., and X be the number of additional trials needed to get the th success after the (-1)st success. Then $X_1, X_2, \ldots, X$ are independent Geometric(p) random variables.

(This follows quite easily.)

4.11. Poisson Distribution

A random variable X has a Poisson distribution with parameter $\lambda > 0$ if its PMF is $P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$, which we denote as $X \sim \text{Poisson}(\lambda)$. We can prove the Poisson PMF sums to 1 using the Taylor series of $e^u=\sum_{k=0}^{\infty}\frac{u^k}{k!}$ with $u=\lambda$.

4.12. Law of Rare Events/Poisson Approximation

For a binomial distribution with huge n and tiny p such that np is moderate, the Binomial(n, p) is approximately the Poisson$(\lambda = np)$.