Conditional Probability continued

Let S be the sample space.
$\mathcal{A}$ the collection of sets on which a probability P is defined.

Fix B and let's look at the set function $Q(A) = P(A|B), A \in \mathcal{A}$.
This is also a probability function.
Why?

So all probability rules apply to $Q(A)$ as well.
$P(A|B) = 1 - P(A^c|B)$,
$P(A \cup C|B) = P(A|B) + P(C|B) - P(A \cap C|B)$.

Bayes Rule

- There are many times that we want $P(B|A)$.
- However, we might only have information on $P(A|B)$.
- E.g. from medical tests, we often have a lot of knowledge of the probability of a test resulting positive given an patient has a disease, or the probability of a test resulting positive given a patient does not have a disease.
- But when a test is run, we want the probability that a patient has a disease given that a test is positive or negative.

Motivation

Data from OpenIntro p98. Breast cancer for women in Canada.

$$P(\text{positive}|\text{cancer}) = 0.89 \text{ so } P(\text{negative}|\text{cancer}) = 0.11.$$
$$P(\text{positive}|\text{not cancer}) = 0.07 \text{ so } P(\text{negative}|\text{not cancer}) = 0.93.$$
$$P(\text{cancer}) = 0.0035 \text{ so } P(\text{not cancer}) = 0.9965$$

- But we want to know $P(\text{cancer}|\text{positive})$.

Bayes Rule

Recall Multiplication Rule:

$$
\begin{aligned}
P(B|A)P(A) &= P(A \cap B) = P(A|B)P(B) \\
P(B|A)P(A) &= P(A|B)P(B) \\
P(B|A) &= \frac{P(A|B)P(B)}{P(A)}
\end{aligned}
$$

This is known as Bayes rule.

Bayes Rule

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

Law of Total Probability

You almost always get $P(A)$ by using the law of total probablity:

Law of total probability (more general form)

Suppose $B_1, B_2, \ldots, B_K$ is a partition of the sample space S. I.e. $B_1 \cup B_2 \cup \cdots \cup B_K = S$ and $B_i \cap B_j = \emptyset$ for all $i \neq j$, then

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_K) \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_K)P(B_K). \end{aligned}$$

We previously defined this law using $K = 2$:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c).$$

Our Cancer Example

$$P(\text{positive}|\text{cancer}) = 0.89 \text{ so } P(\text{negative}|\text{cancer}) = 0.11.$$
$$P(\text{positive}|\text{not cancer}) = 0.07 \text{ so } P(\text{negative}|\text{not cancer}) = 0.93.$$
$$P(\text{cancer}) = 0.0035 \text{ so } P(\text{not cancer}) = 0.9965$$

$P(\text{cancer}|\text{positive}) = \frac{P(\text{positive}|\text{cancer})P(\text{cancer})}{P(\text{positive})}$.
We need

$$\begin{aligned} P(\text{positive}) &= P(\text{positive}|\text{cancer})P(\text{cancer}) \\ &\quad + P(\text{positive}|\text{not cancer})P(\text{not cancer}) \\ &= 0.89 * 0.0035 + 0.07 * 0.9965 = 0.07287 \end{aligned}$$

Our Cancer Example

$$\begin{aligned} P(\text{positive}|\text{cancer}) &= 0.89 \text{ so } P(\text{negative}|\text{cancer}) = 0.11. \\ P(\text{positive}|\text{not cancer}) &= 0.07 \text{ so } P(\text{negative}|\text{not cancer}) = 0.93. \\ P(\text{cancer}) &= 0.0035 \text{ so } P(\text{not cancer}) = 0.9965 \\ P(\text{positive}) &= 0.07287 \end{aligned}$$

$$\begin{aligned} P(\text{cancer}|\text{positive}) &= \frac{P(\text{positive}|\text{cancer})P(\text{cancer})}{P(\text{positive})} \\ &= \frac{0.89 * 0.0035}{0.07287} \\ &= 0.04275 \end{aligned}$$

Intuition

- So the probability you have cancer given a positive test is only about 4%!
- Even though the test is fairly accurate, because there are so many more people who do not have cancer than who have cancer, they make up a majority of the population who have a positive test result.

Non-uniform probabilities I

Up to now in all our examples elements of $\mathcal{S}$ are equally likely.

If $\#S = N$, $p\{s\} = 1/N$ for any $s \in S$.

Now imagine coloring the elements in S with three colors R, G, B, with n_R-red, n_G-green, n_B-blue.

We draw an element from S at random, but only record the color.

So we have a new sample space $\{\text{red}, \text{green}, \text{blue}\}$
$p_{red} = n_R/N, p_{green} = n_G/N, p_{blue} = n_B/N$.

Non-uniform probabilities II

More generally on a sample space $\mathcal{S} = \{1, \ldots, N\}$ we can define a probability P by defining $p_1 = P\{1\}, \ldots, p_N = P\{N\}$, such that:

$p_i \geq 0, i = 1, \ldots, N$

$\sum_{i=1}^{N} p_i = 1.$

And then $P(A) = \sum_{i \in A} p_i$ is a probability.

Random Variables I

Different sample spaces with different event definitions yield the same probabilities:

- Flip a coin 5 times.
- Roll a die 5 times and record if it is odd or even.
- Draw five cards at random from a deck of cards, with replacement and record if they are red or black.

All of these can be described as having a $0/1$ outcome each with probability $1/2$.

- Define X - 0 if heads, 1 if tails.
- Define Y - 0 if odd, 1 if even.
- Define Z - 0 if red, 1 if black.

$P(X = 0) = P(Y = 0) = P(Z = 0) = 1/2$
$P(X = 1) = P(Y = 1) = P(Z = 1) = 1/2$
they all have the same *distribution.*

Random Variables II

Random variable: A function from a sample space S into R, $X : S \to R$.
Discrete random variable: The *range* of X is discrete (either finite or countable).
Let X be discrete and finite $S \to R$. Let the range of X be R_X.
Each value in the range defines an *event*:
$A_x = \{s \in S : X(s) = x\}$.

- $A_x, x \in R_X$ are disjoint.
- $\cup_{x \in R_X} A_x = S$.

Define $f_X(x) = P(A_x) = \ldots$ (shorthand) $P(X = x)$.
$f_X(x), x \in R_X$ is the distribution of X. $\sum_{x \in R_X} f_X(x) = 1$.
Many different random variables with range R_X can have the same *distribution*.

Mean and Variance I

Mean of a random variable: $\mu_X = E(X) = \sum_{x \in R_X} x f_X(x)$

Box example: Box with 10 cards: 4 with value 1, 5 with value 3, 1 with value 5.

X - draw card from box and record its number.
$R_X = \{1, 3, 5\}, \quad f_X(1) = 2/5, f_X(3) = 1/2, f_X(5) = 1/10.$
$\mu_X = E(X) = 1 \cdot 2/5 + 3 \cdot 1/2 + 5 \cdot 1/10 = 2.$
This is also the average of the values in the box:
$(4 * 1 + 5 * 3 + 1 * 5)/10 = 2$
Just like the average of a list is a one number description of the list, or the 'center' of the list, so the mean of a random variable is a one number description of its distribution, or a 'center' of the distribution.

Mean and Variance II

One measure of the spread of a distribution is given by its variance:
$Var(X) = \sum_{x \in R_X} (x - \mu_X)^2 f_X(x)$.
Compute the variance in the example above.
It is the same as the MSD of the box. Why?

Bernoulli and Binomial distributions I

Bernoulli distribution: $R_X = \{0, 1\}, f_X(1) = p, f_X(0) = 1 - p$.
$EX = 1 \cdot p + 0 \cdot (1 - p) = p$.
$Var(X) = (1 - p)^2 p + (0 - p)^2 (1 - p) = p \cdot (1 - p)$.

Bernoulli and Binomial distributions II

Binomial distribution:
$R_X = \{0, 1, \ldots, n\}, f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, \ldots, n$

Let X be the number of heads in n independent tosses of a coin with probability p of heads.
To observe exactly k heads in n tosses we need to observe $n-k$ tails.
The probability of a particular sequence of k heads and $n-k$ tails is always $p^k(1-p)^{n-k}$.
There are $\binom{n}{k}$ different sequences of length n with k heads.
Consequently $f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ and X has the binomial distribution.

Functions of a random variable I

If $X : S \to R$ is a random variable, and $g : R \to R$ is a function then $Y = g(X)$ is also a random variable with range $R_Y = g(R_X)$.

Example: For the box example take $g(x) = \begin{cases} 3 & \text{if } x = 1 \\ 1 & \text{if } x > 1 \end{cases}$.

Then $Y = g(X)$ has range $R_Y = \{1, 3\}$ and distribution
$f_Y(1) = P(Y = 1) = P(X = 3 \text{ or } X = 5) = 6/10$.
$f_Y(3) = P(Y = 3) = P(X = 1) = 4/10$.
So we can compute the mean: $E(Y) = 1 \cdot 6/10 + 3 \cdot 4/10 = 1.8$

Sometimes it is difficult to compute the new distribution of Y.
It's still easy to compute $E(Y)$ in terms of the distribution of X:
$E(Y) = Eg(X) = \sum_{k=1}^{K} g(x_k) f_X(x_k)$.
So in our example we compute

$$\begin{aligned} E(Y) =& g(1) \cdot 2/5 + g(3) \cdot 1/2 + g(5) \cdot 1/10 \\ =& 3 \cdot 2/5 + 1 \cdot 1/2 + 1 \cdot 1/10 = 1.7 \end{aligned}$$