Bayes' Theorem, Random Variables

1 Motivation for Bayes' Theorem

- Want to compute the probability that an event occurs, on the basis of partial evidence
- Involved in probability based decision making
- Example of application
  - Assume there is an accurate test for some disease
  - Compute the probability that a positive test case implies disease
  - Can use additional info like how many positive tests were wrong

2 Illustrative Example

We have two boxes. The first box contains **2 green balls** and **7 red balls,** and the second one contains **4 green balls** and **3 red balls**. Bob selects a ball first by choosing one of the two boxes at random, and then randomly selecting a ball from the chosen box. If Bob selected a red ball, what is the probability that he selected the first box?

- Important: if we don't know the colour of the ball, the probability is just 1/2
- We will now see that additional knowledge changes the probability
- Let E be the event "red ball" and F be the event "first box"
- Need to find $p(F|E)$. Know $p(F|E) = p(F \cap E)/p(E)$. Try to find these.
- Know: $p(E|F) = 7/9$ and $p(E|\overline{F}) = 3/7$. Also, $p(F) = p(\overline{F}) = 1/2$
- Have $p(F \cap E) = p(E \cap F) = p(E|F) \cdot p(F) = (7/9) = 7/18$
- Express E as $E = (E \cap F) \cup (E \cap \overline{F})$
- Note: $(E \cap F)$ and $(E \cap \overline{F})$ are disjoint, so $p(E) = p(E \cap F) + p(E \cap \overline{F})$
- Have $p(E \cap \overline{F}) = P(E|\overline{F}) \cdot p(\overline{F}) = (3/7) \cdot (1/2) = 3/14$
- Putting all together, $p(E) = (7/18) + (3/14) = 76/264 = 38/63$, and $p(F|E) = p(F \cap E)/p(E) = (7/18)/(38/63) = 49/76 \approx 0.645$

3 Bayes' Theorem

3.1 Theorem

Let E and F be events in sample space such that $p(E) \neq 0$ and $p(F) \neq 0$ Then

$$p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\overline{F})p(\overline{F})}$$

3.2 Example

Suppose 1 person in 100,000 has a particular rare disease, for which there is a fairly accurate diagnostic test. The test is correct 99% of the time when given to someone with the disease and 99.5% of the time when given to someone without the disease. Given this information, find

- The probability that someone who tests positive has the disease
- The probability that someone who tests negative does not have the disease
- Let E be "tests positive" and F be "has disease"
- We want to compute $p(F|E)$ and $p(F|E)$ respectively

- In the first case, we need to know $p(E|F), p(E|\overline{F}), p(F),$ and $p(\overline{F})$
- We know $p(F) = 1/100,000 = 0.00001,$ so $p(\overline{F}) = 0.99999$
- We also know $p(E|F) = 0.99$ and $p(\overline{E}|\overline{F}) = 0.995$
- Since $p(\overline{E}|\overline{F}) = 0.995,$ have $p(E|\overline{F}) = 0.005$
- All in all

$$p(F|E) = \frac{0.99 \cdot 0.0001}{0.99 \cdot 0.00001 + 0.005 \cdot 0.99999} \approx 0.002$$

4 Bayesian spam filters

- Some of the first tools to detect span were based on Bayes' theorem
- A Bayesian spam filter uses info about previously seen emails to guess whether an incoming email is spam on the basis of occurrences of certain words
- When spam filter fails to identify spam email as spam, this is a false negative.
- When non-spam is identified as spam, this is a false positive, should be avoided as much as possible.
- Assume we have a set B of messages known to be spam and a set G of non-spam messages.
- For a word w, let $n_B(w)$ and $n_G(w)$ denote the number of messages in B and G, respectively, that contain w
- Then the empirical probability that spam message contains w is $p(w) = n_B(w)/|B|$
- Similarly, the empirical probability that a good message contains w is $q(w) = n_G(w)/|G|$
- Then the empirical probability that spam message contains w is $p(w) = n_B(w)/|B|$
- Similarly, the empirical probability that a good message contains w is $q(w) = n_G(w)/|G|$
- If a message arrives, what is the probability that it is spam?
- Let F be the event "spam" and E the event "contains w".
- Apply Bayes' theorem to find

$$p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\overline{F})p(\overline{F})}$$

- Can assume that $p(E|F) = p(w)$ and $p(E|\overline{F}) = q(w)$
- Estimate somehow $p(F)$. let's assume for the moment for simplicity that $p(F) = p(\overline{F}) = 1/2$, that is, spam and non-spam are equally likely
- Then

$$p(F|E) = \frac{p(w)}{p(w) + q(w)}$$

5 Random variables

Many problems concern with a numerical value associated with the outcome of an experiment. Examples:

- The number of 1s in a randomly generated 10 bit string
- The number of steps a sorting algorithm makes to sort n random numbers

5.1 Definition

A **random variable** is a function from the sample space of an experiment to the real numbers. In other words, a random variable assigns a number to each possible outcome

5.2 Example

A coin is flipped three times. Let X(t) be the random variable that equals the number of heads in the outcome t. Then we have

$$\begin{aligned}
&X(HHH) = 3\\
&X(HHT) = X(HTH) = X(THH) = 2\\
&X(TTH) = X(THT) = X(HTT) = 1\\
&X(TTT) = 0
\end{aligned}$$

6 Distribution of a random variable

Let X(S) denote the set of all values taken by X

6.1 Definition

The distribution of a random variable X on a sample space S is the set of pairs $(r, p(X = r))$ for all values $r \in X(S)$, where $p(X = r)$ is the probability that X takes value r

6.2 Example

A distribution is usually described by specifying $p(X = r)$ for each value r
Example: recall the random variable X from the previous example

$$\begin{aligned}
&X(HHH) = 3\\
&X(HHT) = X(HTH) = X(THH) = 2\\
&X(TTH) = X(THT) = X(HTT) = 1\\
&X(TTT) = 0
\end{aligned}$$

Then the distribution of X is given by

$$p(X = 3) = 1/8, p(X = 2) = 3/8, p(X = 1) = 3/8, p(X = 0) = 1/8$$

7 Expected value

- The expected value of a random variable is a weighted average of its values
- It can be used, for example, to determine who has an advantage in a gambling game, or to compute the average-case complexity of an algorithm

7.1 Definition

The expected value, or expectation of a random variable X on a sample space S with possible outcomes $s_1, ..., s_n$ is equal to

$$E(X) = \sum_{i=1}^{n} p(s_i)\, X(s_i)$$

7.2 Example

Recall the random variable X from the two previous examples
Note that the probability of each outcome is 1/8.

$$E(X) = (1/8) \cdot (3 + 2 + 2 + 2 + 1 + 1 + 1 + 0) = 12/8 = 3/2$$

7.3 Theorem

if X is a random variable and $p(X = r)$ is the probability that $p(X) = r$ so that, $p(X = r) = \sum_{s \in S, X(s)=r} p(s)$, then

$$E(X) = \sum_{r \in X(S)} p(X = r) \cdot r$$

7.4 Proof

Follows directly from definition, just group terms in the sum by $r = X(s_i)$
Example: use the same old random variable X.
It takes values 0,1,2,3 and we know that

$$p(X=3)=1/8, p(X=2)=3/8, p(X=1)=3/8, p(X=0)=1/8$$

Hence, $E(X) = (1/8)\cdot 3 + (3/8)\cdot 2 + (3/8)\cdot 1 + (1/8)\cdot 0 = 12/8 = 3/2$

8 Linearity of expectation

8.1 Theorem

If $X_i, i = 1, ..., n$ are random variables on a sample space S, and a and b are real numbers

- $E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$
- $E(aX + b) = aE(X) + b$

8.2 Proof

What is the expected number of successes in n independent Bernoulli trials with a probability of success p?

- Let S be the sample space of all n-tuples $t = (t_1, ..., t_n)$ where each t_i is either success or failure
- Let X_i be the random variable on S such that $X_i(t) = 1$ if t_i is a success and $X_i(t) = 0$ if t_i is a failure. We have $E(X_i) = 1\cdot p + 0\cdot(1-p) = p$
- We want to compute $E(X)$ where $X = X_1 + ... + X_n$. By linearity of expectation, $E(X) = E(X_1) + ... + E(X_n) = np$

9 Variance of standard deviation

Let X be a random variable on a sample space S. The variance of X is given by

$$V(X) = \sum_{i=1}^{n} (X(s_i) - E(X))^2 \cdot p(s_i)$$

The standard deviation of X, denoted $\sigma(X)$, is defined as $\sqrt{V(X)}$

9.1 Theorem

If X is a random variable then $V(X) = E(X^2) - E(X)^2$

10 Examples

What is the variance in a single Bernoulli trial with a probability?

- Let X be the random variable such that X(t)=1 if t is a success and X(t)=0 if t is a failure
- Since X takes only 0 and 1 values, we have $X^2 = X$
- Hence, $V(X) = E\left(X^2\right) - E(X)^2 = E(X) - E(X)^2 = p - p^2 = p(1-p) = pq$

What is the variance of the random variable equal to the number that comes up when a fair die is rolled?

- We have
$$E(X) = \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 2 + \frac{1}{6}\cdot 3 + \frac{1}{6}\cdot 4 + \frac{1}{6}\cdot 5 + \frac{1}{6}\cdot 6 = 7/2$$
- By definition of expectation
$$E\left(X^2\right) = \frac{1}{6}\cdot 1^2 + \frac{1}{6}\cdot 2^2 + \frac{1}{6}\cdot 3^2 + \frac{1}{6}\cdot 4^2 + \frac{1}{6}\cdot 5^2 + \frac{1}{6}\cdot 6^2 = 91/6$$
- Hence $V(X) = 91/6 - (7/2)^2 = 35/12$

11 Two important inequalities

11.1 Chebysev's inequality

Let X be a random variable on a sample space S with a probability distribution p. If $r > 0$ is a real number then

$$p(|X(s) - E(X)| \geq r) \leq V(X)/r^2$$

11.2 Markov's inequality

Let X be a random variable on a sample S with $X(s) \geqslant 0$ for all s. Then, for any real number $a > 0$

$$p(X(s) \geq a) \leq E(x)/a$$