Probability

Before, we defined probability as the number of 'good' outcomes (size of an event) divided by the total number of outcomes. We do this more formally as: let S be a (countable) sample space. A *probability distribution* is a function $p : S \to \mathbb{R}$ such that:

1. $0 \le p(s) \le 1$ for all $s \in S$ (all probabilities should be nonnegative and less than 1)
2. $\sum_{s\in S} p(s) = 1$ (the sum of all probabilities should be 1)

We call $p(s)$ the probability of an outcome $s \in S$. For an event $E \subset S$, we define the probability as $p(E) = \sum_{s\in E} p(s)$.

Exercises

For the following, let S be a sample space, and p a probability distribution on S.

1. For any event $E \subset S$, show that $p(\overline{E}) = 1 - p(E)$.

$$p(\overline{E}) + p(E) = \sum_{s\in E} p(s) + \sum_{s\in \overline{E}} p(s) = \sum_{s\in S} p(s) = 1$$

   The second equality is because E and $\overline{E}$ are disjoint and their union is the whole sample space. The last equality is from the definition of a probability distribution.

2. For any events $E_1, E_2 \subset S$, show that $p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$.

$$p(E_1)+p(E_2) = \sum_{s\in E_1} p(s)+\sum_{s\in E_2} p(s) = \sum_{s\in E_1\cup E_2} p(s)+\sum_{s\in E_1\cap E_2} p(s) = p(E_1\cup E_2)+p(E_1\cap E_2)$$

   The second equality is because $|E_1 \cup E_2| = |E_1|+|E_2|-|E_1 \cap E_2|$ (inclusion exclusion).

3. For any parwise disjoint events $E_1, E_2, \cdots \subset S$, show that $p(\cup_i E_i) = \sum_i p(E_i)$.

   Since the events E_i are disjoint, we know that $s \in \cup_i E_i \iff s \in E_i$ for exactly one i. Therefore:

$$p(\cup_i E_i) = \sum_{s\in \cup_i E_i} p(s) = \sum_i \sum_{s\in E_i} p(s) = \sum_i p(E_i)$$

4. We say that events $E_1, E_2 \subset S$ are *independent* if $p(E_1 \cap E_2) = p(E_1)p(E_2)$. Under what situation can two disjoint events $E_1, E_2 \subset S$ be independent?

   If $E_1 \cap E_2 = \varnothing$, then $p(E_1 \cap E_2) = p(\varnothing) = 0$. The only way for a product of two real numbers to be 0 is if at least one is zero. Hence at least one of $p(E_1)$ and $p(E_2)$ must be 0. (So usually, disjoint events will not be independent!)

Conditional Probability

Exercises

1. Prove that if $p(E \mid F) = p(E)$, then E and F are independent events.

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)} = p(E) \implies p(E \cap F) = p(E)p(F)$$

2. A *Bernoulli trial* is an experiment with two outcomes, one ("success") with fixed probability p and the other ("failure") with probability $1-p$. Prove that the probability of k successes in n independent Bernoulli trials is $\binom{n}{k}p^k(1-p)^{n-k}$. What is the sum $\sum_{k=0}^{n}\binom{n}{k}(p^k)(1-p)^{n-k}$ in terms of a simple expression?

   Think of n independent (it is important that they are independent!) Bernoulli trials as a string of n letters which are either S (success) or F (failure). The the number of ways to get k successes in n trials is the number of strings with k S's and $n-k$ F's. The probability of getting such a string is $\binom{n}{k}p^k(1-p)^{n-k}$ because we also need to account for the probability of success (p) and the probability of failure ($q = q - p$).

$$\sum_{k=0}^{n}\binom{n}{k}(p^k)(1-p)^{n-k} = 1$$

   Two ways to prove this are: (1) this is the sum of all probabilities in a probability distribution, so must be equal to 1. (2) Using the binomial theorem, this is the expansion of $(p + (1-p))^n = 1^n = 1$.

Bayes' Theorem

Exercises

1. Show that $p(F \mid E)p(E) + p(F \mid \overline{E})p(\overline{E}) = p(F)$. Hence we can also write Bayes' Theorem as $p(E \mid F) = \frac{p(F|E)p(E)}{p(F)}$. Prove this form of Bayes' Theorem using the definition of conditional probability.

$$\begin{aligned} p(F \mid \overline{E})p(\overline{E}) + p(F \mid E)p(E) &= \frac{p(F \cap \overline{E})}{p(\overline{E})}p(\overline{E}) + \frac{p(F \cap E)}{p(E)}p(E) \\ &= p(F \cap \overline{E}) + p(F \cap E) = p(F) \end{aligned}$$

   and this proves the alternative form of Bayes' theorem.

   The last equality is from using problem 3 on the previous side since $F \cap \overline{E}$ and $F \cap E$ are disjoint (since E and $\overline{E}$ are) and $F \cap \overline{E} \cup F \cap E = F$.

   (Often it will be more convenient to use one form over another for computations. For example, some cases may give $p(F)$ directly whereas other cases may cause computing the larger expression $p(F \mid E)p(E) + p(F \mid \overline{E})p(\overline{E})$ to be more intuitive.)