Axioms of Probability

- The subject of probability is concerned with *random process*, a process that can have multiple outcomes.
- A *probability space* consists of three components.
  1. A *sample space* Ω of all outcomes of a random process. An element of the sample space is called a *simple element*.
  2. A family of sets $\mathcal{F}$ of *allowable events*. Each element of $\mathcal{F}$, called an *event*, is a subset of Ω.
  3. A *probability function* $\Pr : \mathcal{F} \to \mathbb{R}$, satisfying the following properties:
     (a) For any $E \in \mathcal{F}$, we have $0 \leq \Pr(E) \leq 1$.
     (b) $\Pr(\Omega) = 1$.
     (c) For any finite or countably infinite sequence of pairwise mutually disjoint events $E_1, E_2, E_3, \ldots$,

$$\Pr\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \Pr(E_i).$$

- *Example:* Consider the process of rolling two fair die. We can model the sample space as the set $\{(a,b) : 1 \leq a, b \leq 6\}$. We also have $\Pr(\{(a,b)\}) = \frac{1}{36}$ for all (a,b) pairs. Now, let's look at some more interesting events:
  - $\Pr(\text{sum of rolls are even}) = \frac{18}{36} = \frac{1}{2}$.
  - $\Pr(\text{the first roll is equal to the second}) = \frac{6}{36} = \frac{1}{6}$.
  - $\Pr(\text{the first roll is larger than the second}) = \frac{15}{36} = \frac{5}{12}$.
  - $\Pr(\text{at least one roll equals 4}) = \frac{11}{36}$.
- We will be only interested in *discrete probability space*, the probability space where Ω is finite or countably infinite. In this case, if event $E = \{s_1, s_2, s_3, \ldots\}$, we have that

$$\Pr(E) = \Pr(\{s_1\}) + \Pr(\{s_2\}) + \Pr(\{s_3\}) + \cdots.$$

  A common special case is when Ω is finite, and every simple event has equal probability. In this case, we have that

$$\Pr(E) = \frac{\#(E)}{\#(\Omega)}.$$

- Some properties of the probability functions:
  - If $A \subseteq B$, then $\Pr(A) \leq \Pr(B)$.
  - **Union bound:**

$$\Pr\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} \Pr(E_i).$$

  - **Inclusion-exclusion principle:**

$$\Pr\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \left(\sum_{\substack{S \subseteq [n] \\ \#(S)=k}} \Pr\left(\bigcap_{i \in S} E_i\right)\right).$$

Conditional Probability

- The *conditional probability* that event A occurs given that event F occurs is

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

  The conditional probability is defined only when $P(B) \neq 0$.

- Conditional probability satisfies all the probability axioms.
  - The probability space Ω maps to B.
  - The set of allowable events $\mathcal{F}$ maps to $\{A \cap B : A \in \mathcal{F}\}$.
  - The probability function $\Pr(A)$ maps to $\Pr(A|B)$.

- *Example:* You are dealt two cards face down from a shuffled deck of 8 cards consisting of the four queens and four kings from a standard bridge deck.
  - The dealer looks at both of your two cards (without showing them to you) and tells you (truthfully) that at least one card is a queen. What is the probability that you have been given two queens?

    Answer: Let A be the event that I get two queens, and let B be the event that one card is a queen. We have that $\#(A \cap B) = \#(A) = 6$, and $\#(B) = 4 \times 4 + 6 = 22$, so $\Pr(A|B) = 6/22 = 3/11$.
  - What is this probability if the dealer tells you instead that at least one card is a red queen?

    Answer: Let C be the event that at least one card is a red queen. We have that $\#(A \cap C) = 1 + 2 \times 2 = 5$, and $\#(C) = 12 + 1 = 13$. So, $\Pr(A|C) = 5/13$.
  - What is this probability if the dealer tells you instaed that one card is the queen of hearts?

    Answer: Let D be the event that one card is the queen of hearts. We have that $\#(A \cap D) = 3$, and $\#(D) = 7$. So $\Pr(A|D) = 3/7$.

- **Multiplication Rule:** Assuming all conditioning events have positive probability, we have

$$\Pr(\cap_{i=1}^{n} A_i) = \Pr(A_1)\Pr(A_2|A_1)\Pr(A_3|A_1 \cap A_2)\cdots\Pr(A_n| \cap_{i=1}^{n-1} A_i).$$

- *Example:* A class consisting of 4 graduate and 12 undergraduate students is randomly divided into 4 groups of 4. What is the probability that each group includes a graduate students?

  Let us denote the four graduate students by 1, 2, 3, and 4.
  Let $A_1 = \{$students 1 and 2 are in different group$\}$.
  Let $A_2 = \{$students 1, 2, and 3 are in different group$\}$.
  Let $A_3 = \{$students 1, 2, 3, and 4 are in different group$\}$.
  We have that

$$\Pr(A_3) = \Pr(A_1 \cap A_2 \cap A_3) = \Pr(A_1)\Pr(A_2|A_1)\Pr(A_3|A_1 \cap A_2) = \Pr(A_1)\Pr(A_2|A_1)\Pr(A_3|A_2).$$

  Now, $\Pr(A_1) = 12/15$ because there are 12 slots out of 15 slots that student 2 can occupy. Similarly, $\Pr(A_2|A_1) = \frac{8}{14} = \frac{4}{7}$, and $\Pr(A_3|A_2) = \frac{4}{13}$. So,

$$\Pr(A_3) = \frac{12}{15} \cdot \frac{4}{7} \cdot \frac{4}{13}.$$

- **The Law of Total Probability:** Let $A_1, A_2, \dots, A_n$ be a partition of the sample space. Then, for any event B, we have

$$\Pr(B) = \sum_{i=1}^{n} \Pr(A_i)\Pr(B|A_i).$$

The form of this fomula that turns up very often is:

$$\Pr(B) = \Pr(A)\Pr(B|A) + \Pr(A^c)\Pr(B|A^c).$$

- **Baye's Rule:** Let $A_1, A_2, \dots, A_n$ be a partition of the sample space. Then, for any event B such that $\Pr(B) > 0$, we have, for any i,

$$\Pr(A_i|B) = \frac{\Pr(A_i \cap B)}{\Pr(B)} = \frac{\Pr(A_i)\Pr(B|A_i)}{\sum_{j=1}^{n}\Pr(A_j)\Pr(B|A_j)}.$$

Baye's rule is certainly very useful, but we will not use it much here.

Independence

- Event A is said to be *independent* of event B if $\Pr(A \cap B) = \Pr(A)\Pr(B)$. This implies that, if $\Pr(B) \neq 0$, then $\Pr(A|B) = \Pr(A)$.

  Similarly, we say that events $A_1, A_2, \dots, A_n$ are independent if, for every subset S of $\{1, 2, \dots, n\}$,

$$\Pr\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} \Pr(A_i).$$

In other words, for any subset S and for any $j \notin S$,

$$\Pr\left(A_j \,\middle|\, \bigcap_{i \in S} A_i\right) = \Pr(A_j).$$

- *Example:* An unfair coin (probability p of showing heads) is tossed n times. What is the probability that the number of heads will be even?

  It is natural to assume that the coin tosses are independent. Let us solve a similar problem first. Let $A_{n,k}$ be the event that we get exactly k heads from n tosses. With some counting, we know that

$$\Pr(A_{n,k}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Letting E_n be the event that we get an even number of heads from n tosses, we have

$$\begin{aligned}\Pr(E_n) &= \Pr(A_{n,0}) + \Pr(A_{n,2}) + \Pr(A_{n,4}) + \ldots + \Pr(A_{n,2\lfloor n/2 \rfloor}) \\ &= \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{k} p^{2k} (1-p)^{n-2k}.\end{aligned}$$

Evaluating this sum is a pain in the ass, so we should find another method. Say, let O_n be the event that we get an odd number of heads after n tosses. We have

$$\begin{aligned}\Pr(E_n) &= p\Pr(O_{n-1}) + (1-p)\Pr(E_{n-1}) \\ &= p(1 - \Pr(E_{n-1})) + (1-p)\Pr(E_{n-1}) \\ &= p + (1-2p)\Pr(E_{n-1}).\end{aligned}$$

Recall that this is an inhomogeneous linear recurrence relation. The homogeneous part

$$P(E_n) = (1-2p)\Pr(E_{n-1})$$

has solution $\Pr(E_n) = \alpha(1-2p)^n$. Next, we guess the solution of the inhomogeneous relation to be $\alpha(1-2p)^n + \beta n + \gamma$, so

$$\begin{aligned}
1 &= \alpha(1-2p)^0 + 0\cdot\beta + \gamma \\
1-p &= \alpha(1-2p)^1 + 1\cdot\beta + \gamma \\
p^2 + (1-p)^2 &= \alpha(1-2p)^2 + 2\cdot\beta + \gamma
\end{aligned}$$

Solving, we have that $\alpha = \frac{1}{2}, \beta = 0$, and $\gamma = \frac{1}{2}$. So,

$$\Pr(E_n) = \frac{1+(1-2p)^n}{2}.$$

Discrete Random Variables

- A *random variable* X on a sample space Ω is a real-valued function $X : \Omega \to \mathbb{R}$.

  We are only interested in *discrete* random variable, random variable X whose range is the integers.

- Two random variables X and Y are said to be *independent* if and only if, for pairs of integers x, y, it is true that

  $$\Pr(\{X = x\} \cap \{Y = y\}) = \Pr(X = x)\Pr(Y = y).$$

  Similarly, random variables X_1, X_2, $\ldots$, X_n are independent if and only if, for any subset $I \subset \{1, 2, \ldots, n\}$, and for any integer value x_i, $i \in I$,

  $$\Pr\left(\bigcap_{i\in I} X_i = x_i\right) = \prod_{i\in I}\Pr(X_i = x_i).$$

- The *expectation* of a discrete random variable X, denoted by $\mathbf{E}[X]$, is given by

  $$\mathbf{E}[X] = \sum_{i\in\mathbb{Z}} i\Pr(X = i).$$

- **Bernoulli Random Variable:** A biased coin with probability p of turning head is flipped. The Bernoulli random variable X is defined to be

  $$X = \begin{cases} 1, & \text{if head,} \\ 0, & \text{if tail.} \end{cases}$$

  So, $\Pr(X = 1) = p, \Pr(X = 0) = 1-p$, and $\Pr(X = i) = 0$ for other i. Thus, $\mathbf{E}[X] = (1-p)\cdot 0 + p\cdot 1 = p$.

- **Binomial Random Variable:** The above biased coin is flipped (independently) n times, and X is the number of heads. We have that, if X_i is 1 if the ith flip resulted in a head and 0 otherwise, then $X = X_1 + X_2 + \cdots + X_n$. Note that X_i is a Bernoulli random variable for all i.

  From a problem in the last section, we have

  $$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

So,

$$\mathbf{E}[X] = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=1}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=1}^{n} k \cdot \frac{n}{k}\binom{n-1}{k-1}p^k(1-p)^{n-k}$$
$$= np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k} = np\sum_{k=0}^{n-1}\binom{n-1}{k}p^k(1-p)^{n-1-k} = np.$$

- **Linearity of Expectations:** If X and Y are two random variables, then

$$\begin{aligned}\mathbf{E}[X+Y] &= \sum_{i\in\mathbb{Z}}\sum_{j\in\mathbb{Z}}(i+j)\Pr(\{X=i\}\cap\{Y=j\})\\ &= \sum_{i\in\mathbb{Z}}\sum_{j\in\mathbb{Z}} i\Pr(\{X=i\}\cap\{Y=j\}) + \sum_{i\in\mathbb{Z}}\sum_{j\in\mathbb{Z}} j\Pr(\{X=i\}\cap\{Y=j\})\\ &= \sum_{i\in\mathbb{Z}} i\sum_{j\in\mathbb{Z}}\Pr(Y=j)\Pr(X=i \mid Y=j) + \sum_{j\in\mathbb{Z}} j\sum_{i\in\mathbb{Z}}\Pr(X=i)\Pr(Y=j \mid X=i)\\ &= \sum_{i\in\mathbb{Z}} i\Pr(X=i) + \sum_{j\in\mathbb{Z}} j\Pr(Y=j)\\ &= \mathbf{E}[X]+\mathbf{E}[Y].\end{aligned}$$

  This imples that, for any random variables $X_1, X_2, \ldots, X_n$,

$$\mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n}\mathbf{E}[X_i].$$

  Linearity of expectations is very powerful because it holds for *any* random variable.

- *Example:* Note that linearity of expection gives us a very easy way to calculate the expectation of the binomial random variable: because $X = X_1+X_2+\ldots+X_n$, we have $\mathbf{E}[X] = \mathbf{E}[X_1]+\cdots+\mathbf{E}[X_n] = np$.

- *Example:* n people wearing hats enter a shop. They give their hats to the door boy for safe keeping. When they leave, the door boy hands each of them a random hat. How many people get their own hats back on average?

  Let X_i be the *indicator random variable* (random variable that takes either 0 or 1 as its value) such that $X_i = 1$ if person i gets his hat back, and $X_i = 0$ otherwise. We have that $\mathbf{E}[X_i] = 1/n$ because i gets a random hat back. Now, let X be the number of people who get their own hats back. We have that $X = X_1 + X_2 + \ldots + X_n$. So, $\mathbf{E}[X] = n\mathbf{E}[X_1] = 1$.

- More properties of expectations:
  - For any constant $c, \mathbf{E}[cX] = c\mathbf{E}[X]$.
  - For any random variable X taking only positive values, $\mathbf{E}[X] = \sum_{i\geq 1}\Pr(X \geq i)$.

- **Geometric Random Variable:** A biased coin (with probability p of showing head) is flipped until the first head shows up. Let X be the number of flips until it happens.

  We have that $P(X=k) = (1-p)^{k-1}p$ for all $k \geq 1$. Since X takes only positive values:

$$\mathbf{E}[X] = \sum_{i\geq 1}\Pr(X\geq i) = 1 + (1-p) + (1-p)^2 + \cdots = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

- **Markov's Inequality:** For any positive random variable X, for any $a > 0$,

$$\Pr(X \geq a) \leq \frac{\mathbf{E}[X]}{a}.$$