PROBABILITY AND DISCRETE DISTRIBUTIONS

The set of all **outcomes** for an **experiment** is called the **sample space** (usually S, but some authors use Ω). An **event** is a set of outcomes. The assignment of probabilities to events must obey the following rules:

Axioms of Probability. *Axioms for probability:*

1. $P(E) \geq 0$ *for any event* E
2. $P(S) = 1$
3. *If* $E_1, E_2, E_3, \ldots$ *are disjoint events, then* $P(E_1 \cup E_2 \cup E_3 \cup \ldots) = \sum_{i=1}^{\infty} P(E_i)$

Theorem (Basic theorems of probability)**.** *Let* A *and* B *be events.*

1. $P(\emptyset) = 0$
2. *If* $A \subseteq B$, *then* $P(A) \leq P(B)$
3. $P(A) = 1 - P(A^C)$
4. $P(A) = P(A \cap B) + P(A \cap B^C)$
5. $\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$

Note that part 5 can be rearranged: $P(A \cap B) = P(A) + P(B) - P(A \cup B)$.

Theorem. *If an experiment has* N *equally likely outcomes, then* $P(E) = \dfrac{\textit{number of outcomes in } E}{N}$.

Method. (Multiplication rule for counting) If a process occurs in two steps and there are m options for the first step and n options for the second, then there are mn total possibilities.

Method. The number of ways to select k elements from an n-element set is...

	Order matters	Order doesn't matter
With replacement	n^k	$\binom{n+k-1}{k}$
Without replacement	$\dfrac{n!}{(n-k)!}$	$\binom{n}{k} = \dfrac{n!}{(n-k)!k!}$

Definition. Let A and B be events with $P(A) \neq 0$. The **conditional probability of** B **given** A is

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

.

Definition. Events A and B are **independent** if and only if $P(A \cap B) = P(A)P(B)$.

Note that it follows that A and B are independent if $P(B|A) = P(B)$, that is, knowing that event A has occurred does not affect our calculation of $P(B)$.

Theorem (Multiplication rule for probabilities)**.** *Let* A *and* B *be events with* $P(B) \neq 0$. *Then*

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

Theorem (The Law of Total Probability)**.** *If event* A *has probability strictly between* 0 *and* 1, *then for any event* B, $\boxed{P(B) = P(B|A)P(A) + P(B|A^C)P(A^C)}$.

Theorem (Bayes' Law)**.** *If* A *and* B *are events with positive probability, then*

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

Note: often the law of total probability is used to calculate $P(B)$.

Date: February 7, 2023.

Theorem (Extended multiplication rule for probabilities)**.** *Let $E_1, E_2, \ldots, E_n$ be events with $P(E_i) \neq 0$. Then*

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1)\, P(E_2|E_1)\, P(E_3|E_1 \cap E_2) \ldots P(E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1})$$

Definition. Events $E_1, E_2, \ldots, E_2$ form a **partition** of S if

1. $E_1 \cup E_2 \cup \cdots \cup E_2 = S$ and
2. $E_i \cap E_j = \emptyset$ if $i \neq j$ (the events are pairwise disjoint).

Theorem (The Law of Total Probability Extended)**.** *If events $E_1, E_2, \ldots, E_n$ each have probability strictly between 0 and 1 and form a partition of S, then for any event A,*

$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_n)P(E_n)$$

Definition. A **random variable** X assigns a number to each outcome in the sample space S.

1. All random variables have a **cumulative distribution function (CDF)**: $F(x) = P(X \leq x)$.
2. A discrete random variable has a **probability mass function (PMF)**: $p(x) = P(X = x)$.

Theorem. *The PMF of any (discrete) random variable has the following properties:*

1. $0 \leq p(x) \leq 1$ *for all* x
2. $\sum_x p(x) = 1$ *(where the sum is over all possible values of X)*

Theorem. *The CDF of any random variable has the following properties:*

1. *it is non-decreasing: if $a \leq b$, then $F(a) \leq F(b)$*
2. $\lim_{x\to-\infty} F(x) = 0$ *and* $\lim_{x\to\infty} F(x) = 1$
3. *If $a < b$, then $P(a < X \leq b) = F(b) - F(a)$*

Definition. The **expected value** (or **mean**) of a random variable is a weighted average and is denoted $E(X)$ or μ_X. If X is a discrete RV with PMF $p(x)$, then $\boxed{E(X) = \sum_x xp(x)}$ (where the sum is taken over all possible values for X).

Definition. A random variable X has a **Bernoulli distribution** with parameter p (with $0 < p < 1$) if X has possible values 0 and 1 with $P(X = 1) = p$ and $P(X = 0) = 1 - p$. The outcome 1 is often referred to as "success" while 0 is "failure" and the experiment is often called a Bernoulli trial.

Definition. The total number of successes in n independent, identically distributed (iid) Bernoulli trials with parameter p is a random variable with a **binomial distribution**. The PMF of a random variable X having a binomial distribution with parameters n and p is

$$\boxed{b(x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 0, 1, \ldots, n}$$

R Implementation. If $X \sim \text{binom}(n, p)$, then the PMF is `dbinom(x, n, p)` and the CDF is `pbinom(x, n, p)`.

Definition. Suppose n elements are to be selected without replacement from a population of size $N = M_1 + M_2$ where M_1 is the number of successes and M_2 is the number of failures. The number of successes selected is a **hypergeometric** random variable and its PMF is

$$h(x) = \frac{\binom{M_1}{x}\binom{N-M_1}{n-x}}{\binom{N}{n}} = \frac{\binom{M_1}{x}\binom{M_2}{n-x}}{\binom{M_1+M_2}{n}}$$

R Implementation. If $Y \sim \text{hyper}(M_1, M_2, n)$, then the PMF is `dhyper(x, M1, M2, n)` and the CDF is `phyper(x, M1, M2, n)`.

Note. A standard deck of cards has 52 cards in 4 suits: spades $\spadesuit$, hearts $\heartsuit$, clubs $\clubsuit$, and diamonds $\diamondsuit$. Spades and clubs are black suits, while hearts and diamonds are red suits. Within each suit there are 13 ranks: 2, 3, 4, . . . , 10, Jack, Queen, King, and Ace. Together, a rank and a suit uniquely identify the card (cards are 2-dimensional).

1. Suppose 10 cards are dealt from a well-shuffled deck.

a) Find the probability that exactly 5 are hearts.

b) Find the probability that 5 or more are hearts.

2. Suppose now that you shuffle the deck and look a the top card 10 times.

a) Find the probability that exactly 5 of the cards you see are hearts.

b) Find the probability that 5 or more are hearts.

3. For this problem, suppose you have created a mega deck by shuffling together 10 regular decks (for a total of 520 cards).

a) Find the probability that exactly 5 are hearts if you deal 10 cards from the mega deck.

b) Find the probability that exactly 5 cards are hearts when you shuffle and look at the top card 10 times.

c) Repeat both parts a and b for a super-mega deck of 5200 cards.

d) Any observations?

4. Return to a regular deck of 52 cards. The goal this time is to find the PMF of a new distribution in which you count the number of cards you must look at to find a heart. For this problem, take the inefficient approach of shuffling and looking at just the top card. Return the card to the deck an repeat until you see a heart. Let X be the number of times you do this (including the time when you see the heart and stop).

a) What are the possible values for X?

b) Find the PMF for X.

5. This is the more reasonable version of the previous problem: deal cards from a well-shuffled deck until you deal a heart. Let Y be the number of cards you deal (including the heart).

a) What are the possible values for Y?

b) Find the PMF for Y.

6. Suppose you're studying abroad in Florence and, as you leave for school one morning, you find a coin in the pocket of your jacket. You don't look at the coin, but you can tell it's either a €1 or a €2 coin; you figure both are equally likely. On your way out, a friend pays you back for the coffee you bought the other day with a €1 coin that you drop in the same pocket. Then you stop for coffee and randomly select a coin from the pocket: it's a €1 coin. What is the probability that the remaining coin is €2?