PROBABILITY AND DISCRETE DISTRIBUTIONS

Definition. 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.

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

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. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
6. $P(A \cap B) = P(B|A)P(A)$

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

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	$\dbinom{n+k-1}{k}$
Without replacement	$\dfrac{n!}{(n-k)!}$	$\dbinom{n}{k} = \dfrac{n!}{(n-k)!k!}$

Note. The number of **combinations** (bottom right corner) $\binom{n}{k} = \frac{n!}{(n-k)!k!}$ is implemented in **R** as `choose(n,k)`. It is implemented in most calculators as `nCk` or `C(n,k)`. The number of **permutations** (bottom left corner) $\frac{n!}{(n-k)!}$ is implemented in most calculators as `nPk` or `P(n,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 if $P(B|A) = P(B)$, then A and B are independent. In other words, if knowing that event A has occurred does not affect our calculation of $P(B)$, then A and B are independent.

Theorem (The Law of Total Probability)**.** If event A has probability strictly between 0 and 1, then for any event B,

$$P(B) = P(B|A)P(A) + P(B|A^C)P(A^C)$$

Date: February 12, 2024.

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

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

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

Example. Suppose 0.01% of a population carry a genetic disease (and have no obvious symptoms). Doctors screen for the disease using a test that correctly identifies carriers 99.9% of the time. The test also gives false positives 0.2% of the time. What is the probability that a random person who tests positive actually carries the disease?

Example. There are two taxi companies in Extremistan: Crimson Cab (which has 80% of the cars) and Tangerine Taxi (which owns the remaining 20%). A taxi from one of these companies sideswiped some parked cars. An eyewitness claims to have seen a Tangerine Taxi, but given the conditions (dusk) and distance (1 block) the eyewitness is considered to be only 75% reliable. Should Tangerine Taxi be found liable?

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)** defined for all real numbers: $F(x) = P(X \leq x)$.
2. If X is a discrete random variable with possible values $x_1, x_2, x_3, \ldots$, then X has a **probability mass function (PMF)**: $p(x_i) = P(X = x_i)$.

Theorem. The PMF of any (discrete) random variable with possible values $x_1, x_2, x_3, \ldots$ has the following properties:

1. $0 \leq p(x_i) \leq 1$
2. $\displaystyle\sum_i p(x_i) = 1$

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

1. 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 possible values $x_1, x_2, x_3, \ldots$, then

$$E(X) = \sum_i x_i P(X = x_i)$$

Definition. The **variance** of a random variable is denoted $Var(X)$ or σ_X^2. If X is a discrete RV with possible values $x_1, x_2, x_3, \ldots$, then

$$Var(X) = \sum_i (x_i - \mu_X)^2 \, P(X = x_i)$$

Definition. The **standard deviation** of a random variable is $\sigma = \sqrt{Var(X)}$.

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.

Proposition. The mean and variance of a Bernoulli random variable are $\mu = p$ and $\sigma^2 = p(1 - p)$.

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

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

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

Theorem. The mean and variance of a binomial distribution are $\mu = np$ and $\sigma^2 = np(1 - p)$.

Definition. If independent, identically distributed Bernoulli trials with parameter p are repeated until the first success, then the total number of trials (counting the success) has a **geometric** distribution. The PMF of a random variable X having a geometric distribution with parameter p is

$$P(X = x) = (1 - p)^{x-1} p \text{ for } x = 1, 2, 3, \ldots$$

and the CDF is

$$P(X \leq x) = 1 - (1 - p)^x \text{ for } x = 1, 2, 3, \ldots$$

R syntax. If $X \sim \text{geom}(p)$, then the PMF is `dgeom(x, p)` and the CDF is `pgeom(x, p)`.

Theorem. The mean and variance of a geometric distribution are $\mu = \frac{1}{p}$ and $\sigma^2 = \frac{1-p}{p^2}$.

Definition. Suppose n elements are to be selected without replacement from a population of size $M_1 + M_2$ consisting of M_1 successes and M_2 failures and the order of selection doesn't matter. If $n \leq M_1$, then the number of successes selected is a **hypergeometric** random variable and its PMF is

$$P(X = x) = \frac{\binom{M_1}{x}\binom{M_2}{n-x}}{\binom{M_1+M_2}{n}} \text{ for } x = 0, 1, 2, \ldots, n$$

Proposition. The mean and variance of a hypergeometric distribution are $\mu = n\left(\frac{M_1}{M_1 + M_2}\right)$ and $\sigma^2 = n\left(\frac{M_1}{M_1 + M_2}\right)\left(\frac{M_2}{M_1 + M_2}\right)\left(\frac{M_1 + M_2 - n}{M_1 + M_2 - 1}\right)$.

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

Definition. A **Poisson** random variable with parameter $\lambda > 0$ has the PMF

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \text{ for } x = 0, 1, 2, \ldots$$

Proposition. The mean and variance of a Poisson distribution are $\mu = \lambda$ and $\sigma^2 = \lambda$.

R syntax. If $X \sim \text{pois}(\lambda)$, then the PMF is `dpois(x, λ)` and the CDF is `ppois(x, λ)`.