Permutations

Given a set of size n and a sample of size k, there are...

- *with replacement*: n^k different ordered samples
- *without replacement*:

$$P_k^n = \frac{n!}{(n-k)!} = n(n-1)\ldots(n-k+1)$$

different ordered samples

Corollary: the number of orderings of n elements is

$$n! = n(n-1)(n-2)\ldots 1$$

Combinations

Combinations:enumerates the number of possible *combinations* of k out of n items

$$C_k^n = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

This implies that *order does not matter*

Application: these binomial coefficients occur in

$$(a+b)^n = \sum_{k=0}^{n} C_k^n a^k b^{n-k}$$

Conditional

Definition: the **conditional probability** of A given B is

$$P(A \mid B) = \begin{cases} \frac{P(A\cap B)}{P(B)}, & \text{if } P(B) > 0 \\ 0, & \text{otherwise} \end{cases}$$

The multiplication rule: for any events A and B, $P(A \cap B) = P(A|B)P(B)$

Law of Total Probability

The law of total probability:
Suppose $B_1, B_2, \ldots, B_m$ are disjoint events such that

$$\cup_{i=1}^{m} B_i = \Omega$$

The probability of an arbitrary event A can be expressed as:

$$P(A) = \sum_{i=1}^{m} P(A|B_i)P(B_i)$$

Bayes Rule

Bayes' rule:
Suppose the events $B_1, B_2, \ldots, B_m$ are disjoint and $\cup_{i=1}^{m} B_i = \Omega$. The conditional probability of B_i, given an arbitrary event A, is:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{m} P(A|B_j)P(B_j)}$$

It follows from $P(B_i|A)P(A) = P(A|B_i)P(B_i)$ in combination with the law of total probability applied to $P(A)$

Multiple Independence

Events $A_1, A_2, \ldots, A_m$ are called independent if

$$P(\cap_{i=1}^{m} A_i) = \prod_{i=1}^{m} P(A_i)$$

This holds if any subset is replaced by complements

Probability Mass Function pmf

The **probability mass function** p of a discrete random variable X is the function

$$p : \mathbb{R} \to [0, 1]$$

defined by

$$p(a) = P(X = a)$$

for $-\infty < a < \infty$. If X is a discrete random variable that takes on the values $a_1, a_2, \ldots$, then

$$\begin{aligned} p(a_i) &> 0 \\ \sum_i p(a_i) &= 1 \end{aligned}$$

and $p(a) = 0$ for all other a.

Cumulative Distribution Function cdf

Definition: The distribution function F of a random variable X is the function

$$F : \mathbb{R} \to [0, 1]$$

defined by

$$F(a) = P(X \leq a)$$

for $-\infty < a < \infty$.

- Both the probability mass function and the distribution function of a discrete random variable X contain all the probabilistic information of X
- The probability distribution of X is determined by either of them

Properties of CDF

Properties of the distribution function F of a random variable X:

1. For $a \leq b$ one has that $F(a) \leq F(b)$
2. Since $F(a)$ is a probability, $0 \leq F(a) \leq 1$, and
$$\begin{aligned} \lim_{a \to +\infty} F(a) &= 1 \\ \lim_{a \to -\infty} F(a) &= 0 \end{aligned}$$
3. F is right-continuous, i.e., one has
$$\lim_{\epsilon \downarrow 0} F(a + \epsilon) = F(a)$$

- NB: $a \leq b$ implies that the event $\{X \leq a\}$ is contained in the event $\{X \leq b\}$
- Conversely, any function F satisfying 1, 2, and 3 is the distribution function of some random variable

Probability Density Function pdf

The probability density function (pdf) $f(x)$ of X is an integrable function such that

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Conditions on f:

- $f(x) \geq 0 \ \forall x \in \Omega$
- $\int_{-\infty}^{\infty} f(x)dx = 1$

The cdf of a continuous r.v. X is defined as

$$F(x) = \int_{-\infty}^{x} f(u)du = P(X \leq x)$$

Expectation of a Discrete RV

Definition: The **expected value** of a discrete random variable X is defined as

$$E(X) = \sum_{x_i \in \Omega} x_i p(x_i)$$

Expectation of a Continuous RV

Definition: The **expected value** of a continuous random variable X is defined as

$$E(X) = \int_{\Omega} x f(x) dx$$

Variance of any RV

Definition: The **variance** of a random variable X is defined as

$$\begin{aligned} Var(X) &= E\left((X - E(X))^2\right) \\ &= E(X^2) - E(X)^2 \end{aligned}$$

Standard Deviation of any RV

The **standard deviation** of a rv X is

$$\sigma(X) = \sqrt{Var(X)}$$

Expectation Properties

Expectation:

$$\begin{aligned} E(aX) &= aE(X) \quad \forall\ a \text{ constant} \\ E(XY) &= E(X)E(Y) \quad \text{if X and Y are independent} \\ E(a + bX) &= a + bE(X) \quad \text{linearity} \\ E(X + Y) &= E(X) + E(Y) \quad \text{linearity} \\ E[\sum_{i=1}^{n} X_i] &= \sum_{i=1}^{n} E[X_i] \end{aligned}$$

Variance Properties

Variance:

$$\begin{aligned} Var(aX) &= a^2 Var(X) \quad \forall\ a \text{ constant} \\ Var(a + X) &= Var(X) \quad \forall\ a \text{ constant} \end{aligned}$$

Bernoulli Distribution

Definition: A discrete random variable X has a Bernoulli distribution with parameter p, where $0 \leq p \leq 1$, if its probability mass function is given by

$$p_X(1) = P(X = 1) = p$$

and

$$p_X(0) = P(X = 0) = 1 - p$$

Notation: $X \sim Ber(p)$.

Binomial Distribution

Definition: A discrete random variable X has a **Binomial distribution** with parameters n and p, where $n = 1, 2, \ldots$ and $0 \leq p \leq 1$, if its probability mass function is given by

$$p_X(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

for $k = 0, 1, \ldots, n$

- We denote this distribution by $Bin(n, p)$
- The expectation of a Binomial distribution $Bin(n, p)$ is

$$E(X) = np$$

- Its variance is

$$Var(X) = np(1 - p)$$

Hypergeometric Distribution

Hypergeometric Distribution

$$(X = x) = \frac{\binom{s}{x}\binom{N-s}{n-x}}{\binom{N}{n}}$$

$$\mu = \frac{ns}{N} \qquad \sigma^2 = n\left(\frac{s}{N}\right)\left(\frac{N-s}{N}\right)\left(\frac{N-n}{N-1}\right)$$

where,

- N = Total number of elements.
- s = Number of special items in N elements.
- n = Number of elements drawn.
- x = Number of special items in the n elements.

Geometric Distribution

Definition: A discrete random variable X has a geometric distribution with parameter p, where $0 < p \leq 1$, if its probability mass function is given by

$$p_X(k) = P(X = k) = (1-p)^{k-1}p$$

for $k = 1, 2, \ldots$.

- We denote this distribution by $Geo(p)$
- The expectation of a Geometric distribution $Geo(p)$ is

$$E(X) = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = \frac{1}{p}$$

- Its variance is

$$Var(X) = \frac{1-p}{p^2}$$

Geometric Distribution: Memoryless Property

Memoryless property: for $n, k = 0, 1, 2, \ldots$ one has

$$P(X > n + k | X > k) = P(X > n)$$

Poisson Distribution

Definition: A discrete random variable X has a Poisson distribution with parameter $\lambda > 0$ if its probability mass function p is given by

$$p(k) = P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

- We denote this distribution by $Poi(\lambda)$.
- Derivation of the expectation of a Poisson rv X with rate λ:

$$\begin{aligned} E(X) &= \textstyle\sum_{k=0}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} &= \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda}\textstyle\sum_{j=0}^{\infty}\frac{\lambda^j}{j!} &= \lambda \end{aligned}$$

- The variance can be derived in a similar way:

$$Var(X) = \lambda$$

Poisson Distribution: Property

Suppose we sum two Poisson random variables, then the sum is also Poisson.

That is, if

$$X \sim \text{Poisson}(\lambda) \text{ and } Y \sim \text{Poisson}(\mu),$$

then

$$X + Y \sim \text{Poisson}(\lambda + \mu).$$