Sample Space and Events

The **sample point** of a random experiement is a outcomes that cannot be decomposed into other results. The **sample space** is defiend as all the possible outcomes. IF a sample space is countable, ie the outcomescan be put into a one-to-one correspondence with the positive integers, then it is a **discrete sample space** if it cannot, then it is a **continuous sample space**. An **event** is a subset of the sample space. If the event is the entire sample space then it is a **certain event** if it contains no sample points than it is **impossible** or **null event**. If it contains only one outcome then it is an **elementary event**.

Set Operations

We have that the **union** of two events A and B, $A \cup B$ is the event that either A or B or both occur. The **intersection** of two events A and B, $A \cap B$ is the event that both A and B occur. The **complement** of an event A, A^c is the event that A does not occur. If A is a subset of B then we can say that A **implies** B. And we can say that two events are **equal** if they contain the same outcomes. Some following set properties are usefull

$\boxed{A \cup B = B \cup A}$ $\boxed{A \cap B = B \cap A}$ $\boxed{A \cup (B \cup C) = (A \cup B) \cup C}$ $\boxed{A \cap (B \cap C) = (A \cap B) \cap C}$ $\boxed{(A \cap B)^c = A^c \cup B^c}$
$\boxed{(A \cup B)^c = A^c \cap B^c}$ $\boxed{A \cup (B \cap C) = (A \cup B) \cap (A \cup C)}$ $\boxed{A \cap (B \cup C) = (A \cap B) \cup (A \cap C)}$

Axioms of Probability

Let a random experiment have a sample space S, A probability law for the experiment is a rule that assigns for each event A a number P[A] called the **probability** of A. **Axiom 1:** $P[A] \geq 0$ for all events A. **Axiom 2:** $P[S] = 1$. **Axiom 3:** If $A_1, A_2, \ldots$ are disjoint events, ie $A_1 \cap A_2 = \emptyset$ then $P[A_1 \cup A_2] = P[A_1] + P[A_2]$. A generalization of this is that given a family of disjoint events $A_1, A_2, \ldots$ then $P[\cup_{i=1}^n A_i] = \sum_{i=1}^n P[A_i]$.

Combinatorics

The **binomial coeeficent** is the number of ways to choose an unordered set of k elements from a set of n elements without replacement. We have that $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. If we do have **replacement** the number of ways for choosing an unordered set of k elements from a set of n elements is $\binom{n-k+1}{k}$. The binomial coefficent is a special case of the **multinomial coefficent** Let us suppose we are partitioning a set of n distinct objects into $\mathcal{F}$ subsets, $B_1, B_2, \ldots, B_{\mathcal{F}}$, where B_i has k_i elements and $\sum_{i=1}^{\mathcal{F}} k_i = n$. Then the number of ways to do this is this is $\binom{n}{k_1, k_2, \ldots, k_{\mathcal{F}}} = \frac{n!}{k_1! k_2! \cdots k_{\mathcal{F}}!}$.

Conditional Probability and its applications

The **conditional probability** of an event A given B is defined as $P[A|B] = \frac{P[A \cap B]}{P[B]}$. From this we can get the **chain rule** $P[A_1 \cap A_2 \cap \cdots \cap A_n] = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$. The **total probability rule** is $P[A] = \sum_{i=1}^n P[A|B_i]P[B_i]$ for some $B_1, ...B_n$ that form a partition of the sample space (ie $B_1 \cap B_2 = \emptyset$ and $\cup_{i=1}^n B_i = S$). Therefore from the definition of conditional probability we can derive the bayes rule: $P[B_i|A] = \frac{P[A|B_i]P[B_i]}{P(A)}$

Independence

We say that two events A and B are **independent** if $P[A|B] = P[A]$. From the definition of conditional probability we can derive that independence can be also expressed as $P[A \cap B] = P[A]P[B]$. Furthermore let two events A and B be **conditionally independent** given C if $P[A \cap B|C] = P[A|C]P[B|C]$.

Random Variables

A random variable is a mapping from $(\Omega, \mathcal{F}, P)$ to a measurable space $(\Omega', \mathcal{F}')$, where Σ is the sample space, $\mathcal{F} = 2^{\Omega}$ is the set of events and P is the probability measure ie $P : \mathcal{F} \to [0, 1]$. More formally a random variable $X : \Omega \to \Omega'$ is a mapping that satisfies: $X^{-1}(B) = \{\omega' \in \Omega : X(\omega') = B\} \in \mathcal{F}$ for all $B \in \mathcal{F}'$.

Discrete Random Variable

In particular, a discrete random variable is a real-valued function of the outcome of the experiment that can take a finite or countably infinite number of values. It is described by its **PMF** which gives the probability of each numerical value that the random variable can take. Furthermore the function of a discrete random variable defines another discrete random variable.

Expectation and Variance

The **expectated value** of a random variable X is defined as $E[X] = \sum_{x \in \Omega'} xP[X = x]$. More generally for a function of a random variable $g(X)$ we have that $E[g(X)] = \sum_{x \in \Omega'} g(x)P[X = x]$. The **variance** of a random variable X is defined as $Var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$. If $Y = aX + b$ we have that $E[Y] = aE[X] + b$ and $Var(Y) = a^2 Var(X)$. For two random variables A and B we have $E[A + B] = E[A] + E[B]$ and $Var(A + B) = Var(A) + Var(B) + 2Cov(A, B)$, therefore if A and B are independent $Var(A + B) = Var(A) + Var(B)$ and $E[AB] = E[A]E[B]$.

Common Discrete Random Variables

Bernoulli	Binomial	Geometric	Negative Binomial
$X \sim Ber(p)$	$X \sim Bin(n, p)$	$X \sim Geo(p)$	$X \sim NegBin(r, p)$
$X \in \{0, 1\}$	$X \in \{0, 1, \ldots, n\}$	$X \in \{0, 1, \ldots\}$	$X \in \{0, 1, \ldots\}$
$P[X = 1] = p$	$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$	$P[X = k] = p(1-p)^k$	$P[X = k] = \binom{k+r-1}{k} p^k (1-p)^r$
$E[X] = p$	$E[X] = np$	$E[X] = \frac{1}{p}$	$E[X] = \frac{rp}{1-p}$
$Var[X] = p(1-p)$	$Var[X] = np(1-p)$	$Var[X] = \frac{1-p}{p^2}$	$Var[X] = \frac{rp}{(1-p)^2}$

Poisson	Hypergeometric	Discrete Uniform	zipf
$X \sim Pois(\lambda)$	$X \sim Hype(n, N, K)$	$X \sim DisUnif(a, b)$	$X \sim Zipf(s, N)$
$X \in \{0, 1, \ldots\}$	$X \in \{0, 1, \ldots, K\}$	$X \in \{a, a+1, \ldots, b\}$	$X \in \{1, 2, \ldots N\}$
$P[X = k] = \frac{\lambda^k e^{-\lambda}}{k!}$	$P[X = k] = \frac{\binom{K}{k}\binom{n-K}{n-k}}{\binom{n}{k}}$	$P[X = a] = \frac{1}{b-a+1}$	$P[X = k] = \frac{1}{H_{N,s}} \frac{1}{k^s}$
$E[X] = \lambda$	$E[X] = n\frac{K}{N}$	$E[X] = \frac{a+b}{2}$	$E[X] = \frac{H_{N,s-1}}{H_{N,s}}$
$Var[X] = \lambda$	$Var[X] = \frac{nK(N-K)(N-n)}{N^2(N-1)}$	$Var[X] = \frac{(b-a+1)^2-1}{12}$	$Var[X] = \frac{H_{N,s-2}}{H_{N,s}} - \frac{H_{N,s-1}^2}{H_{N,s}^2}$

Where $H_{N,s} = \sum_{k=1}^N \frac{1}{k^s}$

Countinous Random Variable

A continous random variable is a random variable that can take any value in a range of values. We can define a continous random variable by its probability density function $f_X(x)$, some important properites fo the pdf is $f_X(x) \geq 0$ and $\int_{\Omega'} f_X(x)dx = 1$. We also have that $P([x, x+\delta]) = f_X(x)\delta$ as $\delta \to 0$. and $P(X \in A) = \int_A f_X(x)dx$.

Bayes Rule

For two discrete random variables X and Y, the Bayes rule is given by $P(Y=y|X=x) = \frac{P(X=x|Y=y)P(Y=y)}{P(X=x)}$. For two continous random variables X and Y, the Bayes rule is given by $f(Y=y|X=x) = \frac{f(X=x|Y=y)f(Y=y)}{f(X=x)}$. And if X is a continous random variable and Y is a discrete random variable, the Bayes rule is given by $P(Y=y|X=x) = \frac{f(X=x|Y=y)P(Y=y)}{f(X=x)}$ and $f(X=x|Y=y) = \frac{P(Y=y|X=x)f(X=x)}{P(Y=y)}$.

Cumualtive Distribution Functions

For a discrete random variables the CDF is defined as $F_X(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i)$. For a continous random variable the CDF is defined as $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(x)dx$. The CDF has the following properites, it is monotonically non decreasing, as $x \to -\infty$ $F_X(x) \to 0$ and as $x \to \infty$ $F_X(x) \to 1$. We can obtain the PDF from the CDF by taking the derivative of the CDF, ie: $f_X(x) = \frac{d}{dx}F_X(x)$. Likewise for a discrete random variable we can obtain the PMF from the cdf with $p_X(x) = F_X(x) - F_X(x-1)$.

Expectation and Covariance(cont')

For a continous random variable we can define the expectation as: $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx$, and more generally $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$. We have the following properties of the expectation: $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$. And a similar formula for variance $Var(X) = \mathbb{E}[Var(X|Y)] + var(E[X|Y])$. An interesting result of this is that given $X_1, X_2, ..$ that are iid, and another random variable that takes nonnegative integer values N independent of $X_1, X_2, ..$ we have that $Var(\sum_{i=1}^{N} X_i) = E[N]Var(X) + E^2[X]Var(N)$ and $E[\sum_{i=1}^{N} X_i] = E[N]E[X]$. We can also define the covariance of two random variables X and Y as $Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. We have that $Var(\sum_{i=1}^{n} iX_i) = \sum_{i=1}^{n} Var(X_i) + \sum_{i=1}^{n}\sum_{j=1, i\neq j}^{n} Cov(X_i, X_j)$. If $Cov(X, Y) = 0$ then we can say that X and Y are uncorrelated, all independent random variables are uncorrelated, but not all uncorrelated random variables are independent, one example of this is $X \sim U(-1, 1)$ and $Y = X^2$. For n random variables $X_1, X_2, .., X_n$ we can define a the covariance matrix Σ as a n by n matrix where $\Sigma_{ij} = Cov(X_i, X_j)$.

Normal Distribution

The normal distribution is a continous random variable with a bell shaped pdf, it is defined by the following pdf: $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. where μ is the mean and σ is the standard deviation. Some important properties of the normal distribution are that it is symmetric about the mean, and that linear transformations of a normal random variable are also normal specifically $A\mathcal{N}(\mu, \sigma^2) + b \sim \mathcal{N}(A\mu + b, A^2\sigma^2)$. Thus we can say that for a random variable N with mean μ and variance σ^2 we have that $Z = \frac{N-\mu}{\sigma} \sim \mathcal{N}(0, 1)$ and $N = \sigma Z + \mu \sim \mathcal{N}(\mu, \sigma^2)$. We can extend this definition to a multivariate normal distribution by defining a covariance matrix Σ as a n by n matrix where $\Sigma_{ij} = Cov(X_i, X_j)$, and a mean vector μ as a n by 1 vector where $\mu_i = \mathbb{E}[X_i]$. The multivariate normal distribution is defined by the following pdf: $f_X(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$. Linear transformations on this distribution are also multivariate normal, specifically given an $m \times n$ matrix A and a $m \times 1$ vector b we have that $A\mathcal{N}(\mu, \Sigma) + b \sim \mathcal{N}(A\mu + b, A\Sigma A^T)$.

Central Limit Theorem

The central limit theorem states that the sum of n independent random variables with mean μ and variance σ^2 converges in distribution to a normal distribution with mean $n\mu$ and variance $n\sigma^2$ as $n \to \infty$.

Functions of Random Variables

The CDF for the function of a random variable $Y = g(X)$ is: $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = \int_{\{x|g(x)\leq y\}} f_X(x)dx$. therefore we get that the pdf for the function of a random variable is: $f_Y(y) = \frac{d}{dy}F_Y(y)$. This can also be applied to the multivariate case. One important case of such is the sum of two independent random variables $Z = X + Y$, then we will have that the resulting PMF/PDF is the discrete/continous convolution of the PMF/PDF of X and Y.

Characteristic Function

The characteristic function of a random variable X is defined as the following function: $\phi_X(t) = \mathbb{E}[e^{jtX}]$. This is also the Fourier transform of the pdf of X. From this we can find the kth moment of X as $\mathbb{E}[X^k] = j^{-k}\phi_X^{(k)}(0)$.

Probability Generating Function

The probability generating function of a random variable X is defined as the following function: $\psi_X(t) = \mathbb{E}[z^X] = \sum_{x=0}^{\infty} z^x f_X(x)$. This is also the Taylor series expansion of the CDF of X This is also the z-transform of the PMF of X. From this we can recover the probability mass function of X as $P(X = x) = \frac{\psi_X^{(x)}(0)}{x!}$, where $\psi_X^{(x)}(0)$ is the xth derivative of $\psi_X(t)$ evaluated at $t = 0$.

Some Common Continous Random Variables

Normal	Exponential	Uniform	Erlang	Chi Squared
$X \sim N(\mu, \sigma^2)$	$X \sim Exp(\lambda)$	$X \sim U(a, b)$	$X \sim Erlang(\lambda, k)$	$X \sim \chi^2(k)$
$x \in [-\infty, \infty]$	$x \in [0, \infty]$	$x \in [a, b]$	$x \in [0, \infty]$	$x \in [0, \infty]$
$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$f(x) = \lambda e^{-\lambda x}$	$f(x) = \frac{1}{b-a}$	$f(x) = \frac{\lambda^k}{(k-1)!}x^{k-1}e^{-\lambda x}$	$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}e^{-x/2}$
$E(X) = \mu$	$E(X) = \frac{1}{\lambda}$	$E(X) = \frac{a+b}{2}$	$E(X) = \frac{k}{\lambda}$	$E(X) = k$
$Var(X) = \sigma^2$	$Var(X) = \frac{1}{\lambda^2}$	$Var(X) = \frac{(b-a)^2}{12}$	$Var(X) = \frac{k}{\lambda^2}$	$Var(X) = 2k$
$\phi_X(\omega) = e^{j\mu\omega - \frac{\sigma^2\omega^2}{2}}$	$\phi_X(\omega) = \frac{\lambda}{\lambda - j\omega}$	$\phi_X(\omega) = \frac{e^{j\omega b} - e^{j\omega a}}{j\omega(b-a)}$	$\phi_X(\omega) = \left(\frac{\lambda}{\lambda - j\omega}\right)^k$	$\phi_X(\omega) = \left(\frac{1}{1-j2\omega}\right)^{k/2}$

The Erlang distribution is the sum of k independent exponential random variables with rate λ. The Chi Squared distribution is the sum of the squares of k independent standard normal random variables.