Probability cheatsheet

Introduction to Probability and Combinatorics

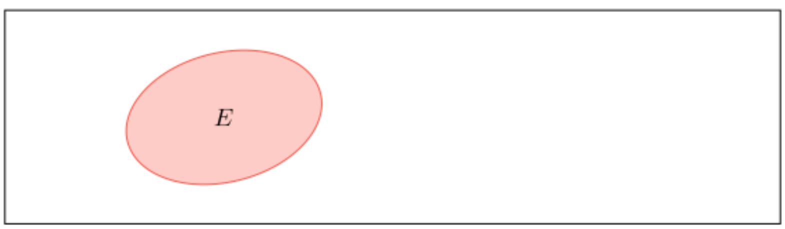
Sample space — The set of all possible outcomes of an experiment is known as the sample space of the experiment and is denoted by S.

Event — Any subset E of the sample space is known as an event. That is, an event is a set consisting of possible outcomes of the experiment. If the outcome of the experiment is contained in E, then we say that E has occurred.

Axioms of probability For each event E, we denote $P(E)$ as the probability of event E occurring.

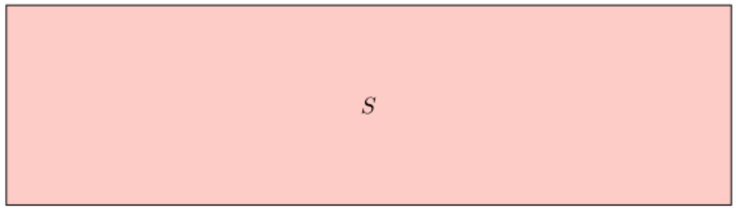
Axiom 1 — Every probability is between 0 and 1 included, i.e:

$$\boxed{0 \leqslant P(E) \leqslant 1}$$



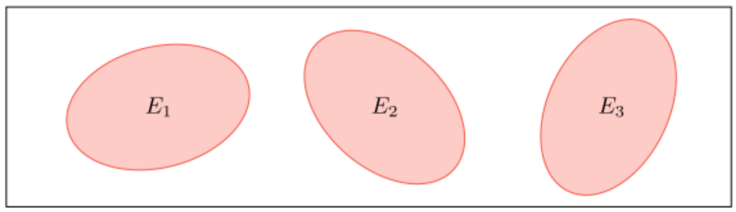
Axiom 2 — The probability that at least one of the elementary events in the entire sample space will occur is 1, i.e:

$$\boxed{P(S) = 1}$$



Axiom 3 — For any sequence of mutually exclusive events $E_1, \ldots, E_n$, we have:

$$\boxed{P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)}$$



Permutation — A permutation is an arrangement of r objects from a pool of n objects, in a given order. The number of such arrangements is given by $P(n, r)$, defined as:

$$\boxed{P(n, r) = \frac{n!}{(n-r)!}}$$

Combination — A combination is an arrangement of r objects from a pool of n objects, where the order does not matter. The number of such arrangements is given by $C(n, r)$, defined as:

$$\boxed{C(n, r) = \frac{P(n, r)}{r!} = \frac{n!}{r!(n-r)!}}$$

Remark: we note that for $0 \leqslant r \leqslant n$, we have $P(n,r) \geqslant C(n,r)$.

Conditional Probability

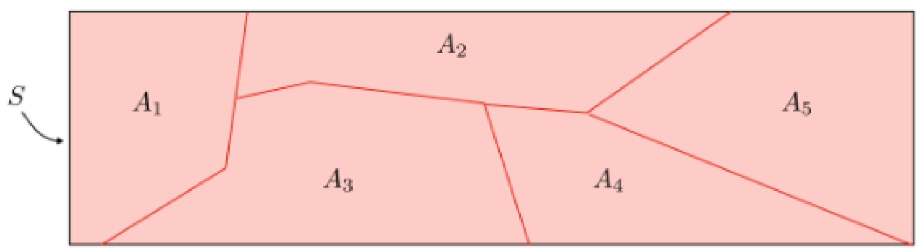
Bayes' rule — For events A and B such that $P(B) > 0$, we have:

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

Remark: we have $P(A \cap B) = P(A)P(B|A) = P(A|B)P(B)$.

Partition — Let $\{A_i, i \in [\![1,n]\!]\}$ be such that for all i, $A_i \neq \varnothing$. We say that $\{A_i\}$ is a partition if we have:

$$\boxed{\forall i \neq j, A_i \cap A_j = \emptyset \quad \textrm{and} \quad \bigcup_{i=1}^{n} A_i = S}$$



Remark: for any event B in the sample space, we have $P(B) = \displaystyle\sum_{i=1}^{n} P(B|A_i)P(A_i)$.

Extended form of Bayes' rule — Let $\{A_i, i \in [\![1,n]\!]\}$ be a partition of the sample space. We have:

$$\boxed{P(A_k|B)=\frac{P(B|A_k)P(A_k)}{\displaystyle\sum_{i=1}^{n}P(B|A_i)P(A_i)}}$$

Independence — Two events A and B are independent if and only if we have:

$$\boxed{P(A\cap B)=P(A)P(B)}$$

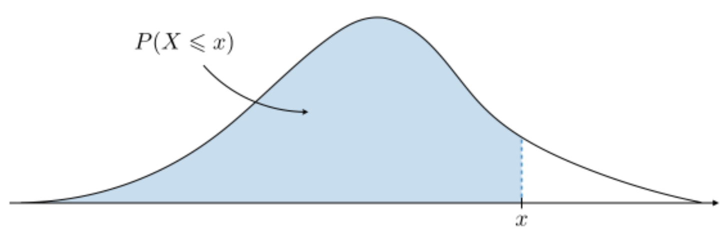
Random Variables

Definitions

Random variable — A random variable, often noted X, is a function that maps every element in a sample space to a real line.

Cumulative distribution function (CDF) — The cumulative distribution function F, which is monotonically non-decreasing and is such that $\underset{x\rightarrow-\infty}{\lim}F(x)=0$ and $\underset{x\rightarrow+\infty}{\lim}F(x)=1$, is defined as:

$$\boxed{F(x)=P(X\leqslant x)}$$



Remark: we have $P(a<X\leqslant B)=F(b)-F(a)$.

Probability density function (PDF) — The probability density function f is the probability that X takes on values between two adjacent realizations of the random variable.

Relationships involving the PDF and CDF

Discrete case — Here, X takes discrete values, such as outcomes of coin flips. By noting f and F the PDF and CDF respectively, we have the following relations:

$$\boxed{F(x)=\sum_{x_i\leqslant x}P(X=x_i)}\quad\textrm{and}\quad\boxed{f(x_j)=P(X=x_j)}$$

On top of that, the PDF is such that:

$$\boxed{0\leqslant f(x_j)\leqslant1}\quad\textrm{and}\quad\boxed{\sum_{j}f(x_j)=1}$$

Continuous case — Here, X takes continuous values, such as the temperature in the room. By noting f and F the PDF and CDF respectively, we have the following relations:

$$\boxed{F(x)=\int_{-\infty}^{x}f(y)dy}\quad\textrm{and}\quad\boxed{f(x)=\frac{dF}{dx}}$$

On top of that, the PDF is such that:

$$\boxed{f(x)\geqslant0}\quad\textrm{and}\quad\boxed{\int_{-\infty}^{+\infty}f(x)dx=1}$$

Expectation and Moments of the Distribution

In the following sections, we are going to keep the same notations as before and the formulas will be explicitly detailed for the discrete **(D)** and continuous **(C)** cases.

Expected value — The expected value of a random variable, also known as the mean value or the first moment, is often noted $E[X]$ or μ and is the value that we would obtain by averaging the results of the experiment infinitely many times. It is computed as follows:

$$\textrm{(D)}\quad\boxed{E[X]=\sum_{i=1}^nx_if(x_i)}\quad\quad\textrm{and}\quad\textrm{(C)}\quad\boxed{E[X]=\int_{-\infty}^{+\infty}xf(x)dx}$$

Generalization of the expected value — The expected value of a function of a random variable $g(X)$ is computed as follows:

$$\textrm{(D)}\quad\boxed{E[g(X)]=\sum_{i=1}^ng(x_i)f(x_i)}\quad\quad\textrm{and}\quad\textrm{(C)}\quad\boxed{E[g(X)]=\int_{-\infty}^{+\infty}g(x)f(x)dx}$$

k^{th} **moment** — The k^{th} moment, noted $E[X^k]$, is the value of X^k that we expect to observe on average on infinitely many trials. It is computed as follows:

$$\textrm{(D)}\quad\boxed{E[X^k]=\sum_{i=1}^nx_i^kf(x_i)}\quad\quad\textrm{and}\quad\textrm{(C)}\quad\boxed{E[X^k]=\int_{-\infty}^{+\infty}x^kf(x)dx}$$

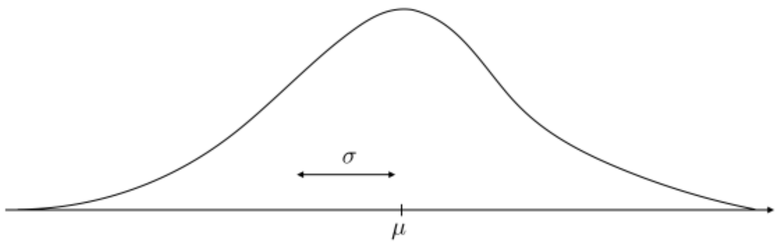
Remark: the k^{th} moment is a particular case of the previous definition with $g:X\mapsto X^k$.

Variance — The variance of a random variable, often noted Var(X) or σ^2, is a measure of the spread of its distribution function. It is determined as follows:

$$\boxed{\textrm{Var}(X)=E[(X-E[X])^2]=E[X^2]-E[X]^2}$$

Standard deviation — The standard deviation of a random variable, often noted σ, is a measure of the spread of its distribution function which is compatible with the units of the actual random variable. It is determined as follows:

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

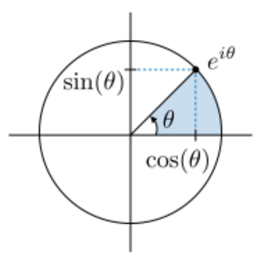


Characteristic function — A characteristic function $\psi(\omega)$ is derived from a probability density function $f(x)$ and is defined as:

$$\textrm{(D)}\quad\boxed{\psi(\omega)=\sum_{i=1}^nf(x_i)e^{i\omega x_i}}\quad\quad\textrm{and}\quad\textrm{(C)}\quad\boxed{\psi(\omega)=\int_{-\infty}^{+\infty}f(x)e^{i\omega x}dx}$$

Euler's formula - For $\theta\in\mathbb{R}$, the Euler formula is the name given to the identity:

$$\boxed{e^{i\theta}=\cos(\theta)+i\sin(\theta)}$$



Revisiting the k^{th} moment — The k^{th} moment can also be computed with the characteristic function as follows:

$$\boxed{E[X^k]=\frac{1}{i^k}\left[\frac{\partial^k\psi}{\partial\omega^k}\right]_{\omega=0}}$$

Transformation of random variables — Let the variables X and Y be linked by some function. By noting f_X and f_Y the distribution function of X and Y respectively, we have:

$$\boxed{f_Y(y)=f_X(x)\left|\frac{dx}{dy}\right|}$$

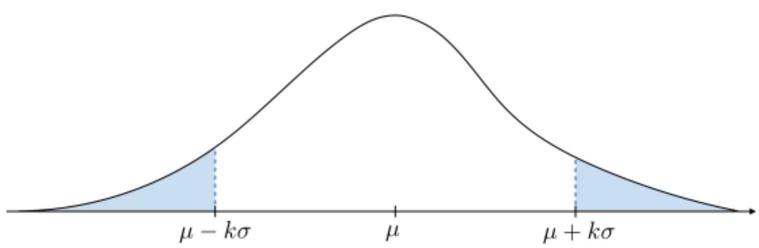
Leibniz integral rule — Let g be a function of x and potentially c, and a,b boundaries that may depend on c. We have:

$$\boxed{\frac{\partial}{\partial c}\left(\int_a^b g(x)dx\right)=\frac{\partial b}{\partial c}\cdot g(b)-\frac{\partial a}{\partial c}\cdot g(a)+\int_a^b\frac{\partial g}{\partial c}(x)dx}$$

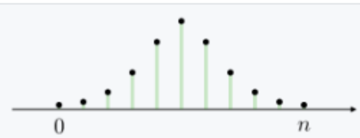
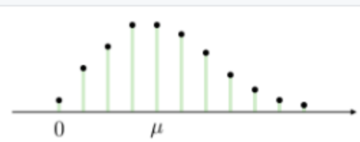
Probability Distributions

Chebyshev's inequality — Let X be a random variable with expected value μ. For $k,\sigma>0$, we have the following inequality:

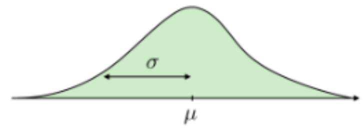
$$\boxed{P(|X-\mu|\geqslant k\sigma)\leqslant\frac{1}{k^2}}$$



Discrete distributions — Here are the main discrete distributions to have in mind:

Distribution	$P(X=x)$	$\psi(\omega)$	$E[X]$	$\mathrm{Var}(X)$	**Illustration**
$X\sim\mathcal{B}(n,p)$	$\binom{n}{x}p^xq^{n-x}$	$(pe^{i\omega}+q)^n$	np	npq	
$X\sim\mathrm{Po}(\mu)$	$\frac{\mu^x}{x!}e^{-\mu}$	$e^{\mu(e^{i\omega}-1)}$	μ	μ	

Continuous distributions — Here are the main continuous distributions to have in mind:

Distribution	$f(x)$	$\psi(\omega)$	$E[X]$	$\textrm{Var}(X)$	Illustration
$X\sim\mathcal{U}(a,b)$	$\dfrac{1}{b-a}$	$\dfrac{e^{i\omega b}-e^{i\omega a}}{(b-a)i\omega}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	
$X\sim\mathcal{N}(\mu,\sigma)$	$\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$e^{i\omega\mu-\frac{1}{2}\omega^2\sigma^2}$	μ	σ^2	
$X\sim\textrm{Exp}(\lambda)$	$\lambda e^{-\lambda x}$	$\dfrac{1}{1-\frac{i\omega}{\lambda}}$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$	

Jointly Distributed Random Variables

Joint probability density function — The joint probability density function of two random variables X and Y, that we note f_{XY}, is defined as follows:

$$\textrm{(D)}\quad\boxed{f_{XY}(x_i,y_j)=P(X=x_i\textrm{ and }Y=y_j)}$$

$$\textrm{(C)}\quad\boxed{f_{XY}(x,y)\Delta x\Delta y=P(x\leqslant X\leqslant x+\Delta x\textrm{ and }y\leqslant Y\leqslant y+\Delta y)}$$

Marginal density — We define the marginal density for the variable X as follows:

$$\textrm{(D)}\quad\boxed{f_X(x_i)=\sum_{j}f_{XY}(x_i,y_j)}\quad\textrm{and}\quad\textrm{(C)}\quad\boxed{f_X(x)=\int_{-\infty}^{+\infty}f_{XY}(x,y)dy}$$

Cumulative distribution — We define cumulative distrubution F_{XY} as follows:

$$\textrm{(D)}\quad\boxed{F_{XY}(x,y)=\sum_{x_i\leqslant x}\sum_{y_j\leqslant y}f_{XY}(x_i,y_j)}\quad\textrm{and}\quad\textrm{(C)}\quad\boxed{F_{XY}(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f_{XY}(x',y')dx'dy'}$$

Conditional density — The conditional density of X with respect to Y, often noted $f_{X|Y}$, is defined as follows:

$$\boxed{f_{X|Y}(x)=\frac{f_{XY}(x,y)}{f_Y(y)}}$$

Independence — Two random variables X and Y are said to be independent if we have:

$$\boxed{f_{XY}(x,y)=f_X(x)f_Y(y)}$$

Moments of joint distributions — We define the moments of joint distributions of random variables X and Y as follows:

$$\text{(D)}\quad \boxed{E[X^pY^q]=\sum_i\sum_j x_i^py_j^qf(x_i,y_j)}\qquad\text{and}\quad\text{(C)}\quad\boxed{E[X^pY^q]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x^py^qf(x,y)dydx}$$

Distribution of a sum of independent random variables — Let $Y=X_1+...+X_n$ with $X_1,...,X_n$ independent. We have:

$$\boxed{\psi_Y(\omega)=\prod_{k=1}^n\psi_{X_k}(\omega)}$$

Covariance — We define the covariance of two random variables X and Y, that we note σ_{XY}^2 or more commonly $\mathrm{Cov}(X,Y)$, as follows:

$$\boxed{\mathrm{Cov}(X,Y)\triangleq\sigma_{XY}^2=E[(X-\mu_X)(Y-\mu_Y)]=E[XY]-\mu_X\mu_Y}$$

Correlation — By noting σ_X,σ_Y the standard deviations of X and Y, we define the correlation between the random variables X and Y, noted ρ_{XY}, as follows:

$$\boxed{\rho_{XY}=\frac{\sigma_{XY}^2}{\sigma_X\sigma_Y}}$$

Remark 1: we note that for any random variables X,Y, we have $\rho_{XY}\in[-1,1]$.

Remark 2: If X and Y are independent, then $\rho_{XY}=0$.