(https://stanford.edu/~shervine/teaching/cs-229/refresher-probabilities-statistics#cheatsheet)Probabilities and Statistics refresher

By Afshine Amidi (https://twitter.com/afshinea) and Shervine Amidi (https://twitter.com/shervinea)

(https://stanford.edu/~shervine/teaching/cs-229/refresher-probabilities-statistics#introduction) Introduction to Probability and Combinatorics

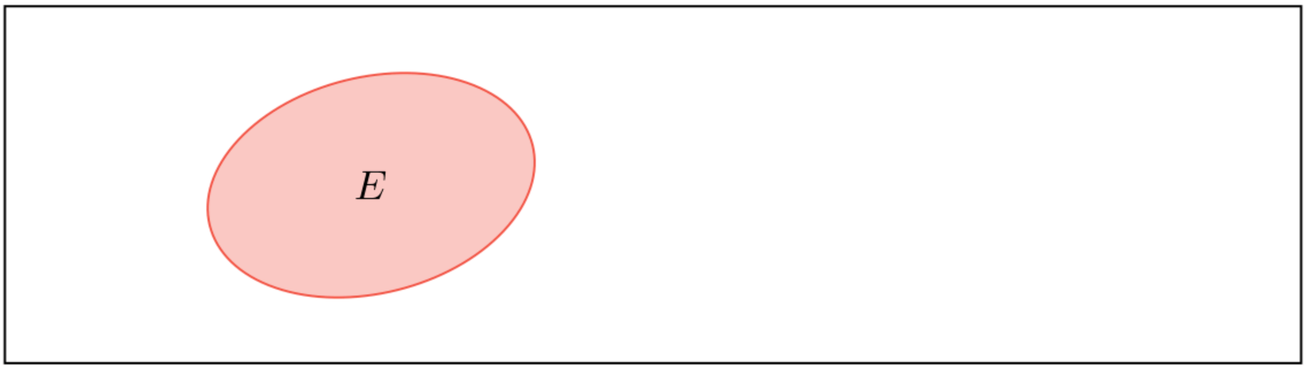
❐ **Sample space** — The set of all possible outcomes of an experiment is known as the sample space of the experiment and is denoted by S.

❐ **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.

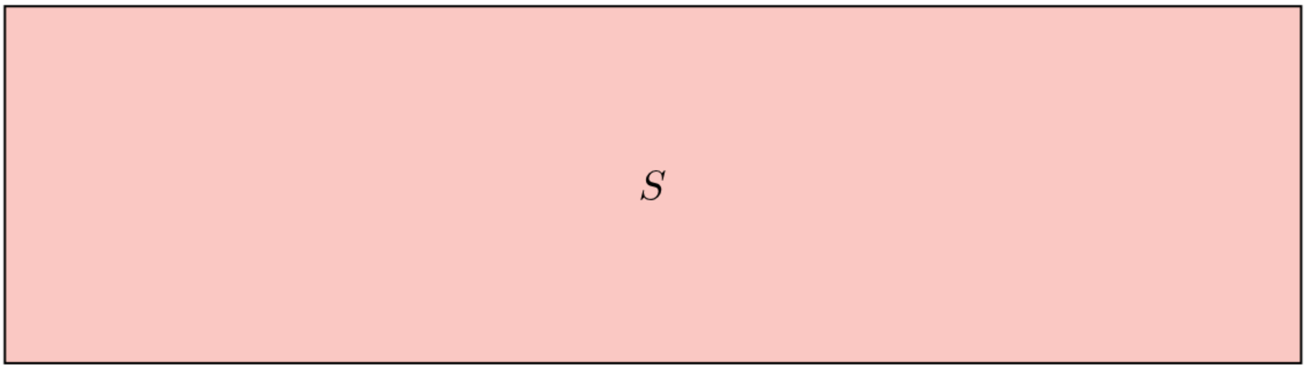
Axiom 1 — Every probability is between 0 and 1 included, i.e:

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



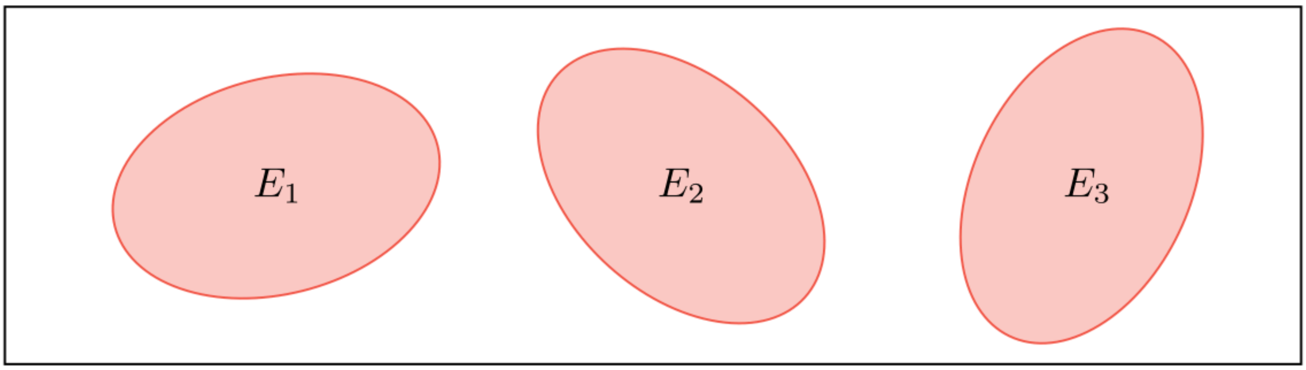
Axiom 2 — The probability that at least one of the elementary events in the entire sample space will occur is 1, i.e:

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



Axiom 3 — For any sequence of mutually exclusive events $E_1, ..., E_n$, 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 $P(n,r)\geqslant C(n,r)$.

(https://stanford.edu/~shervine/teaching/cs-229/refresher-probabilities-statistics#conditional-probability) Conditional Probability

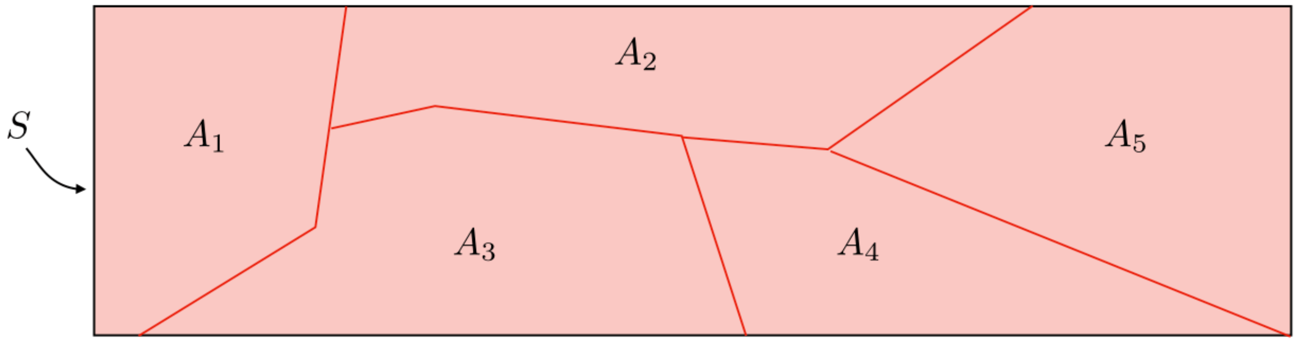
❒ **Bayes' rule** — For events A and B such that $P(B)>0$, we have:

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

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:

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

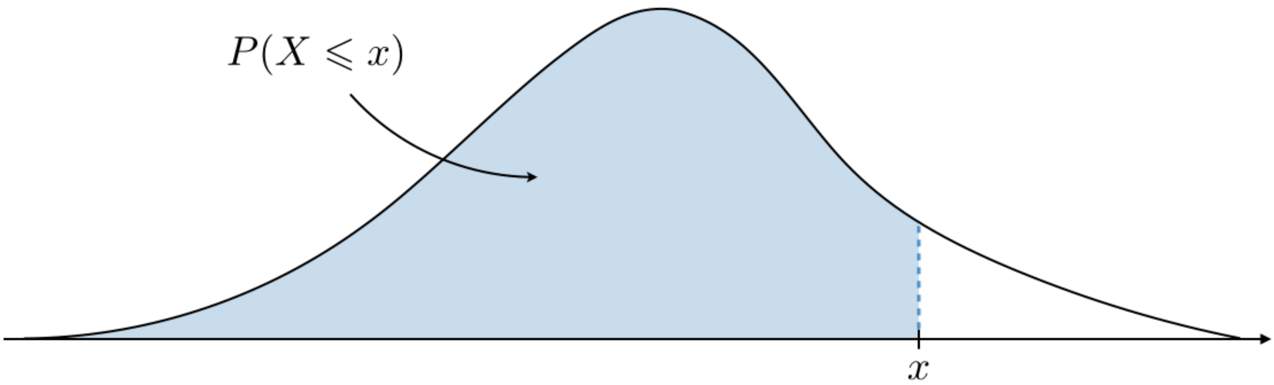
(https://stanford.edu/~shervine/teaching/cs-229/refresher-probabilities-statistics#random-variables) Random Variables

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 $\displaystyle\lim_{x\rightarrow-\infty}F(x)=0$ and $\displaystyle\lim_{x\rightarrow+\infty}F(x)=1$, 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** — Here are the important properties to know in the discrete (D) and the continuous (C) cases.

Case	**CDF** F	**PDF** f	**Properties of PDF**
(D)	$F(x) = \displaystyle\sum_{x_i \leqslant x} P(X = x_i)$	$f(x_j) = P(X = x_j)$	$0 \leqslant f(x_j) \leqslant 1$ and $\displaystyle\sum_j f(x_j) = 1$
(C)	$F(x) = \displaystyle\int_{-\infty}^{x} f(y)dy$	$f(x) = \dfrac{dF}{dx}$	$f(x) \geqslant 0$ and $\displaystyle\int_{-\infty}^{+\infty} f(x)dx = 1$

❒ **Expectation and Moments of the Distribution** — Here are the expressions of the expected value $E[X]$, generalized expected value $E[g(X)]$, k^{th} moment $E[X^k]$ and characteristic function $\psi(\omega)$ for the discrete and continuous cases:

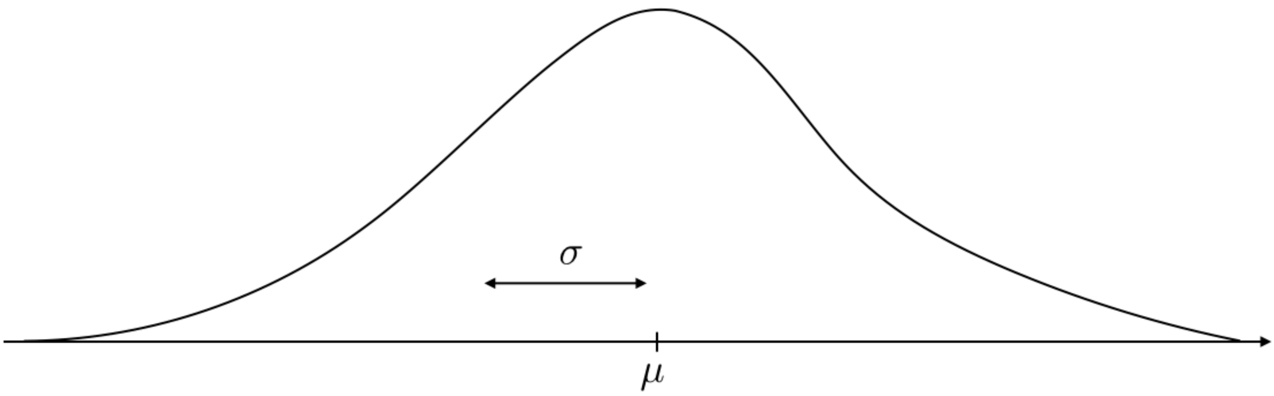
Case	$E[X]$	$E[g(X)]$	$E[X^k]$	$\psi(\omega)$
(D)	$\displaystyle\sum_{i=1}^{n} x_i f(x_i)$	$\displaystyle\sum_{i=1}^{n} g(x_i) f(x_i)$	$\displaystyle\sum_{i=1}^{n} x_i^k f(x_i)$	$\displaystyle\sum_{i=1}^{n} f(x_i) e^{i\omega x_i}$
(C)	$\displaystyle\int_{-\infty}^{+\infty} x f(x) dx$	$\displaystyle\int_{-\infty}^{+\infty} g(x) f(x) dx$	$\displaystyle\int_{-\infty}^{+\infty} x^k f(x) dx$	$\displaystyle\int_{-\infty}^{+\infty} f(x) e^{i\omega x} dx$

❒ **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:

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



❒ **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:

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

(https://stanford.edu/~shervine/teaching/cs-229/refresher-probabilities-statistics#probability-distributions) Probability Distributions

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

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

❒ **Main distributions** — Here are the main distributions to have in mind:

Type	**Distribution**	**PDF**	$\psi(\omega)$	$E[X]$	$\mathrm{Var}(X)$	
(D)	$X \sim \mathcal{B}(n,p)$	$\binom{n}{x} p^x q^{n-x}$	$(pe^{i\omega} + q)^n$	np	npq	
(D)	$X \sim \mathrm{Po}(\mu)$	$\frac{\mu^x}{x!} e^{-\mu}$	$e^{\mu(e^{i\omega}-1)}$	μ	μ	
(C)	$X \sim \mathcal{U}(a,b)$	$\frac{1}{b-a}$	$\frac{e^{i\omega b} - e^{i\omega a}}{(b-a)i\omega}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	
(C)	$X \sim \mathcal{N}(\mu,\sigma)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$e^{i\omega\mu - \frac{1}{2}\omega^2\sigma^2}$	μ	σ^2	
(C)	$X \sim \mathrm{Exp}(\lambda)$	$\lambda e^{-\lambda x}$	$\frac{1}{1-\frac{i\omega}{\lambda}}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	

(https://stanford.edu/~shervine/teaching/cs-229/refresher-probabilities-statistics#joint-rv) Jointly Distributed Random Variables

❒ **Marginal density and cumulative distribution** — From the joint density probability function f_{XY} , we have

Case	Marginal density	Cumulative function
(D)	$f_X(x_i) = \sum_j f_{XY}(x_i, y_j)$	$F_{XY}(x, y) = \sum_{x_i \leqslant x} \sum_{y_j \leqslant y} f_{XY}(x_i, y_j)$
(C)	$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y)dy$	$F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(x', y')dx'dy'$

❒ **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:

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

❒ **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 $\rho_{XY} = 0$.

(https://stanford.edu/~shervine/teaching/cs-229/refresher-probabilities-statistics#parameter-estimation) Parameter estimation

Definitions

❒ **Random sample** — A random sample is a collection of n random variables $X_1, ..., X_n$ that are independent and identically distributed with X.

❒ **Estimator** — An estimator is a function of the data that is used to infer the value of an unknown parameter in a statistical model.

❒ **Bias** — The bias of an estimator $\hat{\theta}$ is defined as being the difference between the expected value of the distribution of $\hat{\theta}$ and the true value, i.e.:

$$\boxed{\textrm{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta}$$

Remark: an estimator is said to be unbiased when we have $E[\hat{\theta}] = \theta$.

Estimating the mean

❒ **Sample mean** — The sample mean of a random sample is used to estimate the true mean μ of a distribution, is often noted $\overline{X}$ and is defined as follows:

$$\boxed{\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

Remark: the sample mean is unbiased, i.e $E[\overline{X}] = \mu$.

❒ **Central Limit Theorem** — Let us have a random sample $X_1, ..., X_n$ following a given distribution with mean μ and variance σ^2, then we have:

$$\boxed{\overline{X} \underset{n\to+\infty}{\sim} \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

Estimating the variance

❐ **Sample variance** — The sample variance of a random sample is used to estimate the true variance σ^2 of a distribution, is often noted s^2 or $\hat{\sigma}^2$ and is defined as follows:

$$\boxed{s^2 = \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Remark: the sample variance is unbiased, i.e $E[s^2] = \sigma^2$.

❐ **Chi-Squared relation with sample variance** — Let s^2 be the sample variance of a random sample. We have:

$$\boxed{\frac{s^2(n-1)}{\sigma^2} \sim \chi^2_{n-1}}$$

For a more detailed overview of the concepts above, check out the **Probabilities and Statistics cheatsheets (teaching/cme-106)**!