- **Random variables and probabilities**

- A variable is said to be random if its value is not known (or cannot be) with complete certainty

- Two different sources of randomness
  a) Imperfect measurement
  b) Intrinsic randomness (quantum systems)

- Probabilities are the way we know how to deal with randomness

e.g. a random variable being the color of the ball you pick from a black box

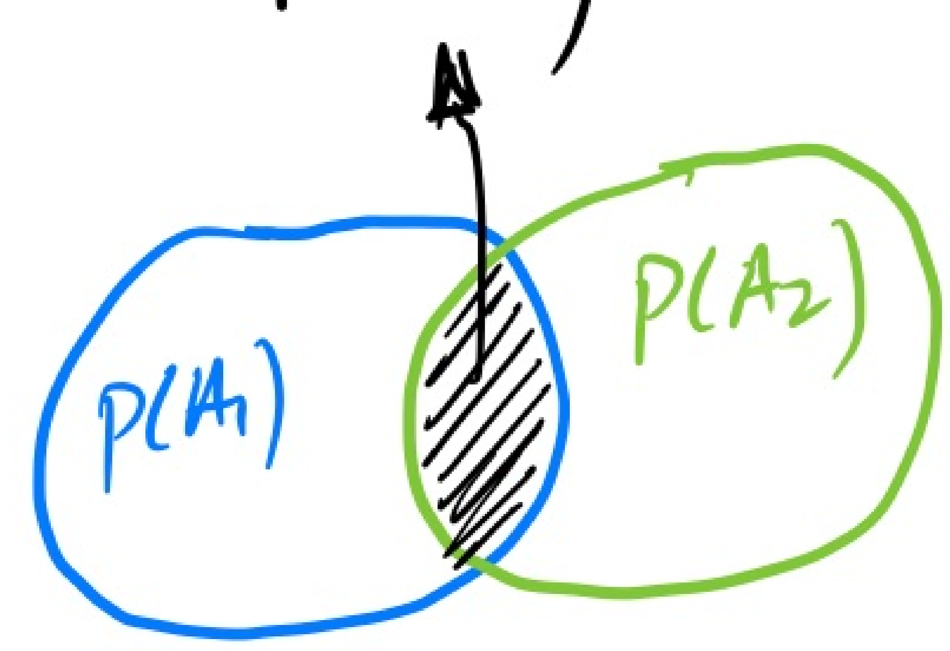
$$C = \{R, G, B\}$$

- The probability is defined through a # of axioms:

i) $p(A_i) \geq 0 \qquad i = 1, 2, \ldots, n$ events

ii) $p(A_1 \cup A_2) = p(A_1) + p(A_2) - p(A_1 \cap A_2)$

iii) $\sum_i p(A_i) = 1$



- Other useful properties:

iv) $p(A|B) = \dfrac{p(A \cap B)}{p(B)} \equiv \dfrac{p(A, B)}{p(B)}$ } a.k.a. the joint probability

v) $P(A,B) = P(A|B)\,P(B) = P(B|A)\,P(A)$

product rule

from which one has the Bayes theorem

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}$$ more on this later!

vi) $P(A) = \sum_B P(A,B)$ sum rule

"marginalisation" over B

- Interpretation of probability

1) Probability as a relative frequency

("frequentist approach")

$$P(A) = \lim_{n\to\infty} \frac{\#\text{ of occurrences of } A \text{ in } n \text{ measurements}}{n}$$

this interpretation makes sense in the typical setup of physics experiments (e.g. collider) where a priori one has n measurements with -ideally- identical initial conditions

It makes less sense if we only have at hand one experiment (say, the observable universe)

2) Subjective probability (Bayesian)

Here we talk about the probabilities of hypothesis:

$P(A)$: degree of believe that hypothesis A is true

Hypothesis A is random because one doesn't know with certainty whether it is true or false

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

in the context of Bayesian inference, typically

A: hypothesis that an experiment will yield a particular result

B: hyp. that a certain theory is true

(more concrete examples later)

So depending on the question we pose, we will need one or the other approach:

- what is the prob. that an experiment will yield a particular result? $\Rightarrow$ Frequentist
- what is the prob. that the mass of the Higgs is contained in some interval? $\Rightarrow$ Bayesian

- Probability functions

- For a continuous random variable x, the prob. of observing a value in the infinitesimal interval $[x; x+dx]$ is given by the PDF $f(x)$:

$$P_{[x;x+dx]} = f(x)\,dx$$

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

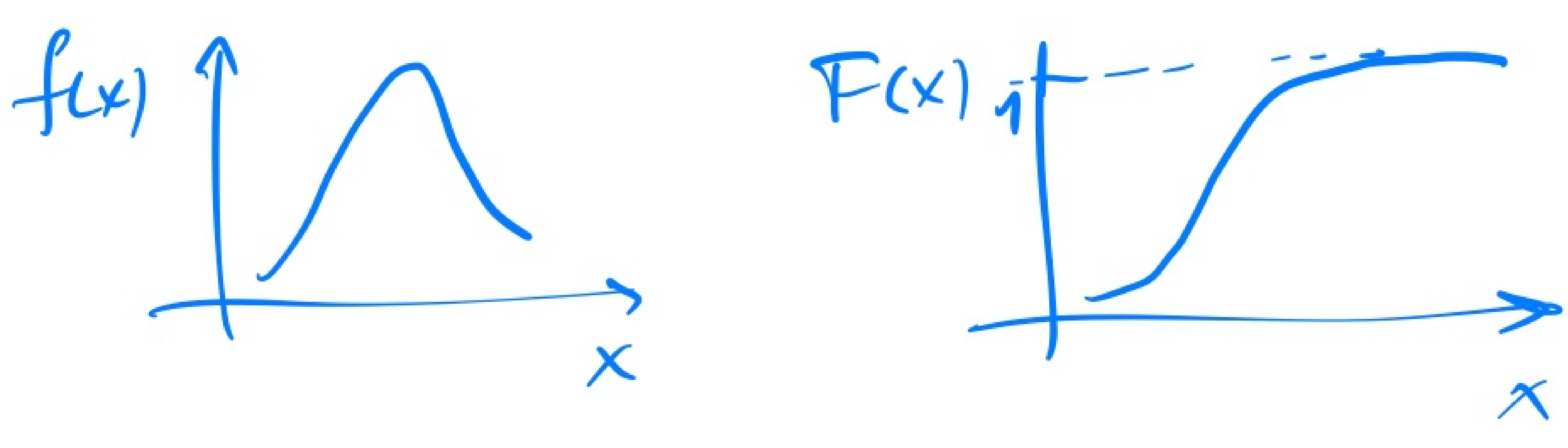
- For a discrete variable instead (e.g. $x \in \mathbb{N}$) we talk about "Prob. Mass Function" and

$$\sum_i f(x_i) = 1$$

- Cumulative distrib. function $F(x)$:

$$F(x) = \int_{-\infty}^{x} f(x')\,dx'$$

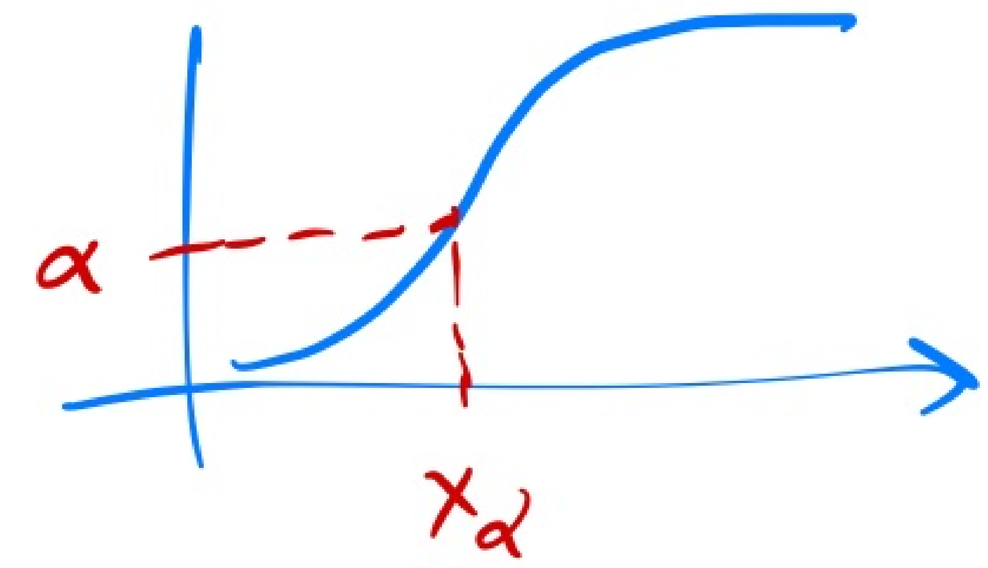
is the prob. for the variable to take a value $\leq x$



There are several properties characterising the prob. distributions:

- The "quantile of order α"

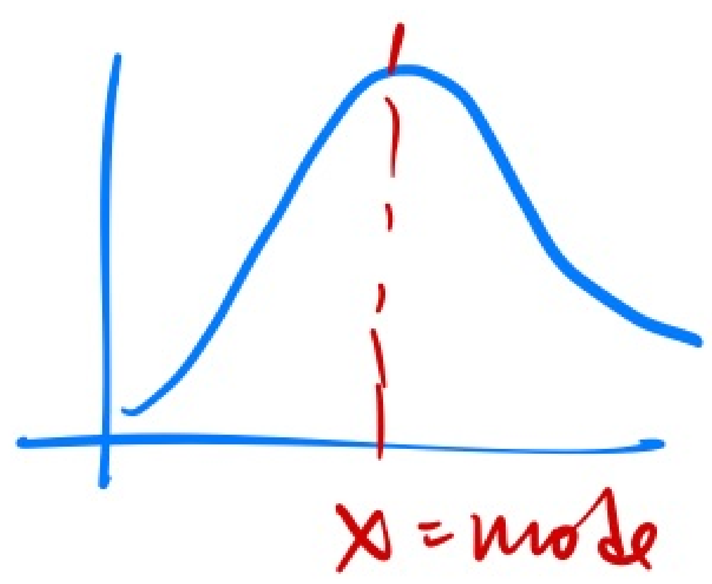
  Is the value $x = x_\alpha$ s.t. $F(x_\alpha) = \alpha$



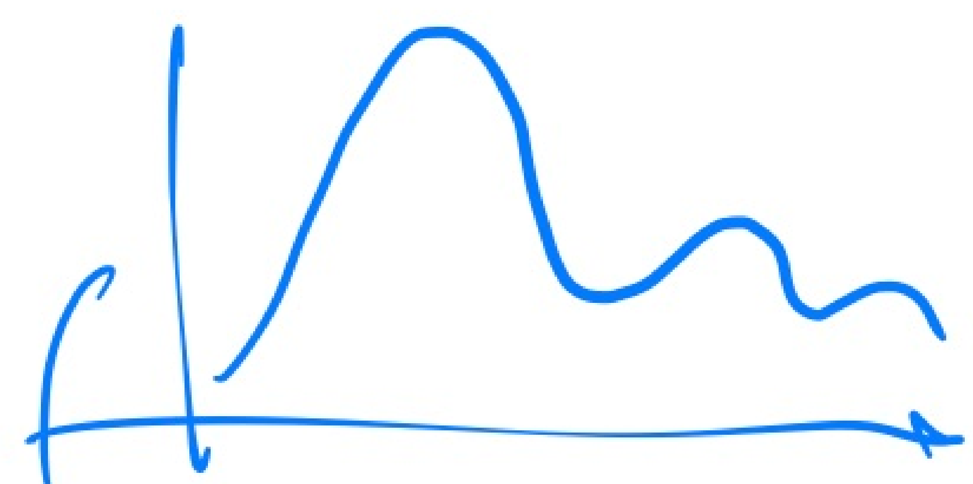
- The median:

  Is the value x_α for $\alpha = 0.5$
  (so equal prob. for $x \leq x_\alpha$ and $x > x_\alpha$)

- The mode



mode = argmax $f(x)$



"multimodal distribution"

- The mean of x ≡ the expectation value of x
  ≡ the 1st moment of $f(x)$

$$\mu = \mathbb{E}[x] = \int_{-\infty}^{\infty} x \, f(x) \, dx$$

arguably the most important use of the PDF

measures where it is likely to observe x

- The variance of x ≡ the 2nd moment of $f(x)$

$$\sigma^2 = \mathrm{Var}[x] = \mathbb{E}[(x-\mu)^2]$$

$$= \int_{-\infty}^{\infty} (x-\mu)^2 f(x) \, dx$$

$$Var[x] = \mathbb{E}[x^2] - 2\mu\mathbb{E}[x] + \mu^2 = \mathbb{E}[x^2] - \mu^2$$

measures how widely x is spread around the mean

- Other higher moments:
  - "Skewness": $\mathbb{E}\left[\left(\frac{x-\mu}{\sigma}\right)^3\right]$ → measures the degree of asymmetry of $f(x)$
  - "Kurtosis": $\mathbb{E}\left[\left(\frac{x-\mu}{\sigma}\right)^4\right]$ → Tailedness, characterising the outliers

- What is the PDF of a function $a(x)$?

Requirement $\quad P_{[x;\,x+dx]} = P_{[a;\,a+da]}$

$$g(x)\,dx = g(a)\,da$$

$$g(a) = f(x(a))\left|\frac{dx}{da}\right|$$

- Joint PDF
  Suppose 2 variables $x, y \Rightarrow f(x,y)$

$$\underbrace{f(x) = \int f(x,y)\,dy}_{\text{marginal}} \qquad \underbrace{h(y|x)}_{\text{conditional}} = \frac{f(x,y)}{\int f(x,y')\,dy'}$$

- Covariance of 2 random variables

$$Cov[x,y] = \mathbb{E}[(x-\mu_x)(y-\mu_y)] = \mathbb{E}[xy] - \mu_x\mu_y$$

$$= \iint xy\,f(x,y)\,dx\,dy - \mu_x\mu_y$$

- Useful probability distributions

— Binary variables, $x = \{0, 1\}$

Let $p(x=1) = \mu$, $p(x=0) = 1-\mu$

then the PMF is the Bernoulli distrib:

$$\text{Bern}(x|\mu) = \mu^x (1-\mu)^{1-x}$$

The mean:

$$\mathbb{E}[x] = \sum_{x=0,1} x \, \text{Bern}(x|\mu) = \mu$$

The variance: $\text{Var}[x] = \mu(1-\mu)$

— Categorical variables $x = \{1, 2, \ldots, K\}$
$= \{A, B, \ldots\}$

A common way to represent this type of variable is the so-called "1-to-K" scheme or "one-hot-encoding"

$$x \to \vec{x} = \{0, 1\}^K$$

e.g. if $K=3$ variables and $\underbrace{x = 2}_{\text{"class 2"}}$

$$\vec{x} = (0, 1, 0)$$

So each class k has an associated prob. μ_k:

$$p(x_k = 1) = \mu_k$$

Then the distrib. is the "generalised Bernoulli" or the "categorical distrib":

$$gBern(\vec{x}|\vec{\mu}) = \prod_{k=1}^{K} \mu_k^{x_k}$$

e.g. if $\vec{x} = (0,1,0)$:

$$gBern(\vec{x}|\vec{\mu}) = \mu_1^0 \mu_2^1 \mu_3^0 = \mu_2$$

$$\mathbb{E}[\vec{x}] = \vec{\mu}$$

- Very related to the Bernoulli distribution: the Binomial distrib:

  It is the prob of having n successes (that a binary variable $x=1$) out of N trials

$$Bin(n|\mu, N) = \binom{N}{n} \mu^n (1-\mu)^{N-n}$$

- Multinomial (analogous to the binomial, but for categorical variables)

$$Mult(\vec{n}|\vec{\mu}, N) = \frac{N!}{n_1! \dots n_K!} \mu_1^{n_1} \mu_2^{n_2} \dots \mu_K^{n_K}$$

- Poisson distrib. $\quad n \in \mathbb{N}$

$$p(n|\lambda) = \frac{\lambda^n e^{-\lambda}}{n!} \qquad \lambda \in \mathbb{R}$$

$$\mathbb{E}[n] = Var[n] = \lambda$$

– Gaussian (or normal)

for $x \in \mathbb{R}$

$$N(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

For the multivariate case; $\vec{x}$: $\dim(\vec{x}) = d$

$$N(\vec{x} \mid \vec{\mu}, \Sigma) = \frac{1}{(2\pi)^{d/2}} \frac{1}{(\det\Sigma)^{1/2}} \times \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \cdot \Sigma^{-1} (\vec{x}-\vec{\mu})\right\}$$

covariance matrix

- **The Central Limit Theorem**

: The sum of n indep. random variables x_i, with means μ_i and variances σ_i^2 (but otherwise unspecified PDF) will have a distrib. that converges, in the limit $n \to \infty$, to a Gaussian with:

$$\mu = \sum_i \mu_i \qquad \sigma^2 = \sum_i \sigma_i^2$$

[proof. Chp. 10 Cowan]

This is the reason why the Gaussian distrib. is so important in statistics

Note! – Variables x_i don't need to be identically distributed

– watch out far away from the center of the Gaussian, if the underlying distrib. of x is quite non-Gaussian