INTRODUCTION

- Language conventions are sometimes different between physicists and statisticians.

Example 1:

Physicist say	Statistician say
determine	estimate
estimate	guess

Example 2:

Demographic language	Physics language
Sample	Data(set)
Draw a sample	Observe, measure
Sample of size N	N observations
Population	Observable space

→ Think about a census, or an election poll

→ We need to distinguish between the properties of the sample an and those of the underlying population.

↳ We will specify: "parent mean = population mean = mean of the underlying distribution"

or: "sample mean".

- Avoid misleading terms:

error → variance, confidence interval, interval estimate, credible interval

propagation of error → change of variable

- Two philosophies: Bayesian and frequentist (or classical)

→ Bayesian approach measures the "degree of belief" that a statement is true.

→ Frequentist approach measures the relative frequency of something happening.

~~Bayesian approach~~ ~~Frequentist approach~~

~~Statement depends on observer~~ ~~Statement independent of the observer~~

~~Applies to repeatable cases~~

	Bayesian	Frequentist
Statement depends on observer	×	
Applies to repeatable cases	×	×
Applies to future unknown facts	×	
Applies to past unknown facts	×	
Applies to possible outcome of an experiment	×	×
Applies to the true value of a parameter	×	
Allows goodness-of-fit (single hypothesis testing)		×
Allows decision theory	×	

- Notation
  - → Greek letters: parameters of the theory: $\theta, \mu, \sigma, \dots$
  - → Roman letters: random variables corresponding to physical observables: $x, E, \dots$
  - → Capitalized P, F : ~~Cum~~ probability distribution or cumulative distribution
  - → Lowercase p, f : probability density function
  - → Bar : average value : $\bar{x}$
  - → Hat : estimate of parameter (often mode of the parameter) : $\hat{\theta}, \hat{\mu}$

PROBABILITY THEORY

- History: → mathematically formalized by Kolmogorov in 1933 with "Foundations of the Theory of Probability"
  → many concepts and ~~[illegible]~~ theorems known already before, even if they hadn't been fully integrated in a coherent theory yet. E.g., the Bayes theorem dates ~~[illegible]~~ back to the 18^{th} century.

- Definitions of probability:

→ <u>Mathematical probability</u>

Let Ω be the set of all elementary events x_i, which are mutually exclusive.

We define the probability of the occurrence of event x_i to obey the Kolmogorov axioms:

$$\boxed{\begin{aligned} &P(x_i) \geq 0 \quad \forall i \\ &P(x_i \vee x_j) = P(x_i) + P(x_j) \\ &\sum_i P(x_i) = 1 \end{aligned}}$$

→ abstract definition

→ holds for any quantity that satisfies the axioms

→ <u>Frequentist probability</u>

Consider an experiment in which a series of N events is observed.

Suppose k events are of type X.

The frequentist probability for any <u>single</u> event to be of type X is the empirical limit of the ratio:

$$\boxed{P(X) = \lim_{N\to\infty} \frac{k}{N}}$$

In other words: $P(X) = \lim_{N\to\infty} \frac{\text{\# of favorable cases}}{\text{\# of possible cases}}$

→ In principle, $P(X)$ can only be known for $N = \infty$. But often it can be computed analytically or numerically to great precision.

→ Can only be applied to repeatable experiments. Cannot predict if Italy will win the next world cup.

$\boxed{\text{Prob 1}}$

→ Repeatability is in principle impossible under the same exact conditions. However, it is the job of the physicist to ensure that all relevant conditions are repeatable, or to make corrections if need be.

→ Bayesian probability

~~Degree of belief: amount that one person is willing to bet that x will occur, knowing that if he/she wins he/she will get a fixed amount. (This is called "coherent bet")~~

~~→ $P(x) = \emptyset$ if we are sure x will not happen~~

~~→ $P(x) = 1$ if we are sure x will happen~~

Degree of belief: amount $E(x)$ that we are willing to bet that x will occur, knowing that if we win we get a fixed amount K. (This is called coherent bet).

→ $P(x) = \dfrac{F(x)}{K}$ → $P(x) = 0$ if we are sure x will not happen

→ $P(x) = 1$ if we are sure x will happen

→ $0 < P(x) < 1$ otherwise

→ $\sum_i P(x_i) = 1$

→ It is a property of the system to be observed, as well as of the observer

→ It depends on the knowledge of the observer, and will change if the knowledge of the observer ~~changes~~ improves.

→ Can be applied to non-repeatable phenomena, e.g. Italy winning the next world cup, me getting bald, or the aliens to have colonized the Earth in prehistoric times.

→ Can be applied to the true value of a Physics Theory!

- Applicability of frequentist and Bayesian probability

Based on these definitions, we can immediately make a decision on which type of probability to use depending on the situation, i.e. depending on the question that we want to address. ~~For example~~ This is particularly important if we consider the fact that the physical parameters of a theory are fixed by Nature, but unknown to us.
For example, we can ask the following types of questions:

- Based on the result of a measurement, what is the true value of a parameter of the theory?
- Based on the result of a measurement, what is the interval that contains the true (unknown) value of a given parameter with a given amount of probability?
- Is my parameterization of the measured data good enough? Or does it indicate the presence of some "new physics"?
- Suppose I want to compare two alternative models based on some experimental data. Which of the models describes better the data?
- Based on the results of previous experiments, what is the expected outcome of a future experiment measuring the same quantity, or a quantity connected to it?

$\boxed{\text{Prob 3}}$

- Properties of probability

→ apply to any probability that satisfies the Kolmogorov axioms

→ A set of A of elementary events X_i can be treated as an event.

The occurrence of A is defined as the occurrence of any event $X_i \in A$.

$$P(A) = P(\text{any } X_i \in A \text{ occurs})$$

→ Addition law

Let A and B be non-exclusive sets of events X_i.

The probability of an event belonging to either A or B is:

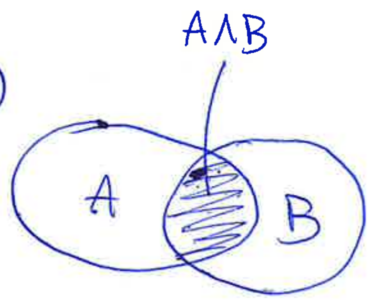
$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

→ Conditional probability and independence

The probability that an elementary event X_i, known to belong to the set B, is also a member of the set A is given by:

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

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



The sets A and B are independent if:

$$P(A|B) = P(A)$$

which means that the previous occurrence of B is irrelevant to the occurrence of A.

If A is independent of B, the probability of the simultaneous occurrence of A and B is the product of their probabilities:

$$P(A \wedge B) = P(A)\cdot P(B)$$

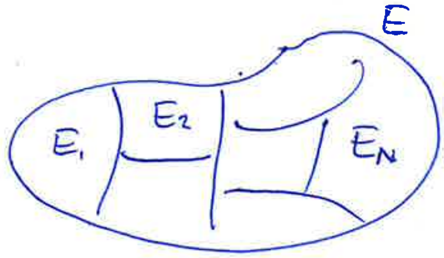
- **Law of Total probability**

Consider N events corresponding to the sets $E_1, \dots, E_N$, which are subsets of another set E included in the sample space Ω.

Assume that the set of E_i is a partition of E, i.e.:

$$E_i \cap E_j = \phi \quad \forall\, i,j$$

$$\bigcup_{i=1}^{N} E_i = E$$



The probability of the set E is:

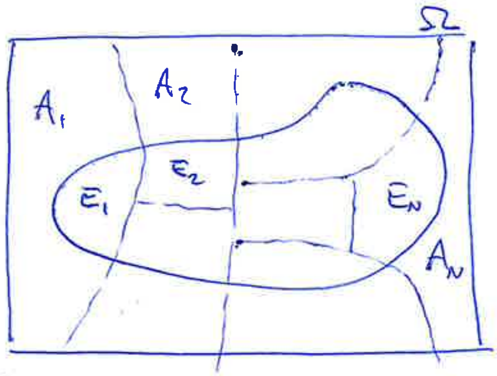
$$P(E) = \sum_{i=1}^{N} P(E_i)$$

Now let's choose a disjoint partition $A_1, \dots, A_N$ of the sample space Ω, such that:

$$\bigcup_i A_i = \Omega$$

$$A_i \cap A_j = \phi$$

$$\sum_i P(A_i) = 1$$



And such that: $E_i = E \cap A_i$

We then have: $P(E_i) = P(E \cap A_i) = P(E|A_i)\,P(A_i)$

Therefore: $$P(E) = \sum_{i=1}^{N} P(E|A_i)\,P(A_i)$$

This is called "law of total probability".

It can be interpreted as the weighted average of the probabilities $P(A_i)$ with weights $w_i = P(E|A_i)$

Bayes Theorem for discrete events

Recall the law of conditional probability (with two sets A and B)

$$P(A \wedge B) = P(B|A)\,P(A) = P(A|B)\,P(B).$$

Therefore:

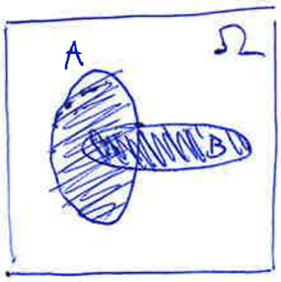
$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

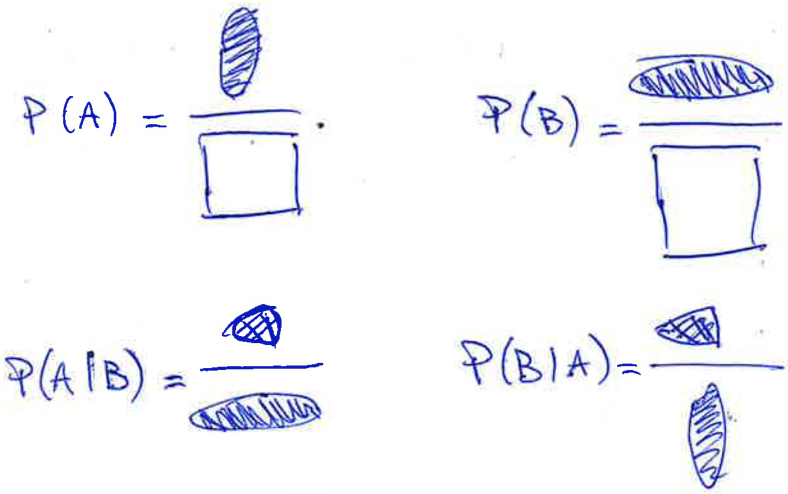
More generally, if $A_i = A_1, \dots, A_n$ are ~~a set of~~ exclusive and exhaustive sets; and if B is any event:

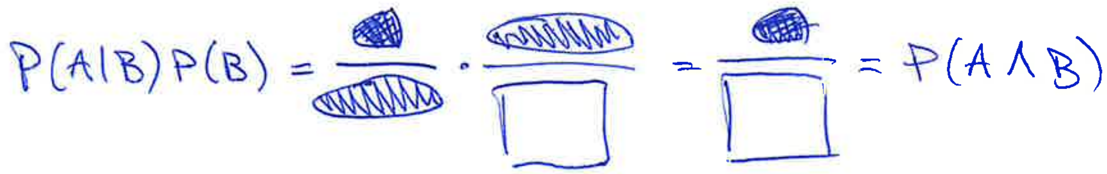
$$\boxed{P(A_i|B) = \frac{P(B|A_i)\,P(A_i)}{\sum_i P(B|A_i)\,P(A_i)}}$$

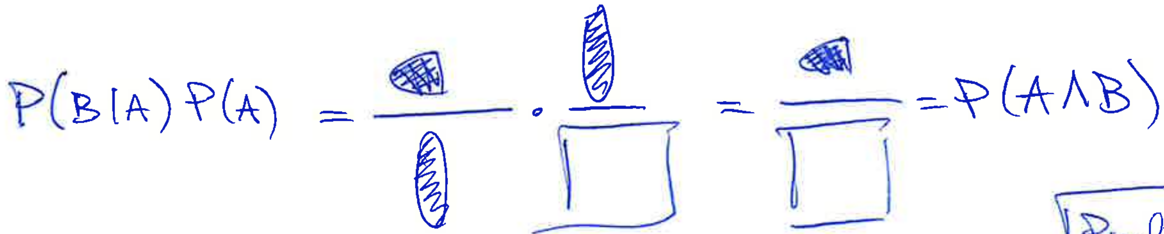
→ $P(A_i)$ is the "prior probability" of A_i, i.e. the probability of set A_i before the knowledge that event B has occurred.

→ $P(A_i|B)$ is the "posterior probability" of A_i, i.e. the probability of set A_i after having collected the information that event B has occurred









Prof. 6

→ Example: measuring protons with particle detector

$P(B)$ = Probability of any particle giving a triggered event

$P(A)$ = Probability of a proton hitting the detector

$P(B|A)$ = Probability of a proton giving a triggered event

$P(A|B)$ = Probability of a triggered event to be induced by a proton.

- Bayes Theorem for hypotheses testing and parameter estimation.

Assume H_0, H_1 are a complete set of hypotheses, i.e. a complete set of physics models describing a given physical phenomenon.

For example: H_0 = background-only hypothesis = The known physics processes are enough to explain the data

H_1 = signal + background hypothesis = There is an additional component due to new physics and that we know how to model.

H_0 and H_1 will depend on some parameters, which can differ between the two hypotheses, but that we will indicate generally as $\vec{\vartheta} \in \Omega$.
Let's indicate the data with $\vec{x}$.

→ Parameter estimation

$$P(\vec{\vartheta}|\vec{x}) = \frac{P(\vec{x}|\vec{\vartheta})\,\pi(\vec{\vartheta})}{\int_\Omega P(\vec{x}|\vec{\vartheta})\,\pi(\vec{\vartheta})\,d\Omega} = \frac{P(\vec{x}|\vec{\vartheta})\,\pi(\vec{\vartheta})}{P(\vec{x})}$$

$P(\vec{\vartheta}|\vec{x})$ = Posterior probability for parameters $\vec{\vartheta}$ given the data $\vec{x}$ and the model H_0 or H_1. This is valid both for H_0 and H_1

$P(\vec{x}|\vec{\vartheta})$ = Probability of obtaining exactly the data $\vec{x}$ given ~~all possible values of~~ the parameters $\vec{\vartheta}$

$\pi(\vec{\vartheta})$ = Prior probability of parameters $\vec{\vartheta}$ under the assumption of H_0 or H_1

$P(\vec{x})$ = Probability of getting data $\vec{x}$ given any possible value of $\vec{\theta}$ assuming model H_0 or H_1. In the end, it's a normalization factor.

Prob. 7

Considering both hypotheses H_0 and H_1: → prior on hypothesis H_0

$$P(\vec{\vartheta}, H_0 | \vec{x}) = \frac{P(\vec{x} | H_0(\vec{\vartheta}))\, \pi(\vec{\vartheta}, H_0)\, \cancel{\pi(H_0)}}{\underbrace{\sum_i \int_\Omega P(\vec{x} | \vec{\vartheta}, H_i)\, \pi(\vec{\vartheta}, H_i)\, \pi(H_i)\, d\vec{\vartheta}}_{\to \text{ constant}}}$$

~~→ Hypothesis Testing:~~

$$\cancel{P(H_0 | \vec{x}) = \frac{\int_\Omega P(\vec{\vartheta}, H_0 | \vec{x})\, d\vec{\vartheta} \cdot \pi(H_0)}{\sum_i P(\vec{x} | H_i)}}$$

~~where:~~

~~$P(H_0 | \vec{x})$ = Posterior probability for hypothesis H_0 given data $\vec{x}$~~

~~$\int P(\vec{\vartheta}, H_0 | \vec{x})\, d\vec{\vartheta}$ = overall~~

→ Hypothesis Testing:

$$P(H_i | \vec{x}) = \frac{P(\vec{x} | H_i)\, \pi(H_i)}{\sum_i P(\vec{x} | H_i)\, \pi(H_i)}$$

where: $P(H_i | \vec{x})$ = posterior probability for hypothesis H_i after measuring the data $\vec{x}$

$\pi(H_i)$ = prior probability for hypothesis H_i.
This is the <u>subjective</u> part of the method.

$P(\vec{x} | H_i) = \int_\Omega P(\vec{x} | \vec{\vartheta}, H_i)\, \pi(\vec{\vartheta})\, d\vec{\vartheta}$ = Probability of obtaining data $\vec{x}$ given all possible values of the parameters $\vec{\vartheta}$ assuming model H_i.

Denominator = normalization factor.

→ <u>Example</u>: Religious belief

$\boxed{\text{Prob. 8}}$

- Random variables and probability distributions

A random event is an event that has more than one possible outcome, to which a probability may be associated.

The outcome of a random event is not known, only the probabilities of the possible outcomes are known.

We can associate a random variable x to a random event X.

The possible numerical values $x_1, x_2, \dots$ ~~of the~~ corresponding to the possible outcomes are the probabilities $P(x_1), P(x_2), \dots$, which form a "probability distribution", obeying to the normalization condition:

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

→ A set of N observations of the random variable x can be considered as a single observation of a vector $\vec{x} = (x_1, \dots, x_N)$

→ This is the definition of a histogram.

→ What if a random variable covers a continuous interval?

- Probability density functions

Consider a sample space $\Omega \in \mathbb{R}^n$.

A random extraction (experiment) will lead to an outcome (measurement) corresponding to one point $\vec{x} \in \Omega$.

We can associate a "probability density" $f(\vec{x})$ to any point $\vec{x} \in \Omega$, with $f(\vec{x}) \geq 0$.

The probability of an event A is: $P(A) = \int_A f(\vec{x})\, d\vec{x}$

$f(\vec{x})$ is the differential probability, i.e. the infinitesimal probability dP corresponding to the infinitesimal hypervolume $d\vec{x}$.

$$f(\vec{x}) = \frac{dP(\vec{x})}{d\vec{x}}$$

$\boxed{\text{Prob. 9}}$

$f(x)$ obeys the normalization condition: $\int_{\Omega} f(\vec{x})\, d\vec{x} = 1$

→ If x is a 1-dim discrete variable: ~~$f(x) = \delta(x - x_i)$~~

$$f(x) = \sum_{i=1}^{N} P_i \, \delta(x - x_i)$$

In fact: $\int_{-\infty}^{\infty} f(x)\, dx = \sum_{i=1}^{N} P_i \int_{-\infty}^{\infty} \delta(x - x_i)\, dx = \sum_i P_i = 1$

→ If a variable is continuous, but we measure it with a discretized device, we can instead do: $P_i = \int_{x_i}^{x_i + \delta x} f(x)\, dx$

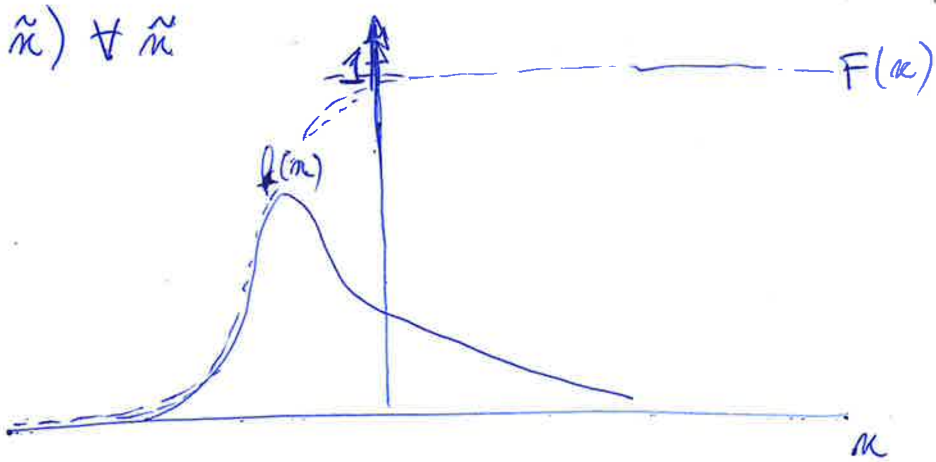
↳ This is the case of a measurement done with a MCA.

↳ This is the case of a measurement of a particle track with a pixelated detector with pixel size δx.

• Cumulative distribution

$$F(x) = \int_{-\infty}^{x} f(y)\, dy \quad \Rightarrow \text{monotonous increasing function from 0 to 1}$$

↳ $F(\tilde{x}) = P(x \le \tilde{x}) \ \forall \ \tilde{x}$



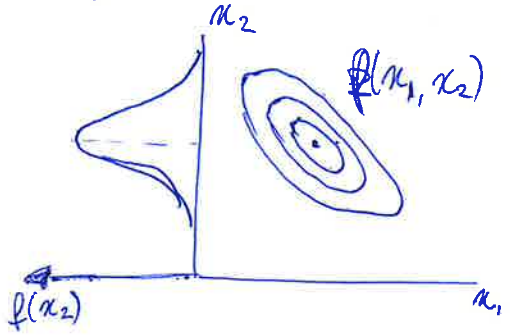
- Probability ~~distr.~~ density function in dim > 1

$$\frac{dP}{d\vec{x}} = f(\vec{x})$$

→ probability density per unit area, volume, ...

→ also called "joint probability distribution"

- Marginal distribution



Take $\vec{x} = x_1, \ldots, x_n$ with PDF $f(\vec{x})$.

The marginal distribution for x_i is:

$$f(x_i) = \int f(\vec{x})\, dx_1 \cdots dx_{i-1} \cdot dx_{i+1} \cdots dx_n \quad \Rightarrow \text{dim } 1$$

We can also define the marginal distribution for a subset of variables:

$$f(x_{i<k}) = \int f(\vec{x})\, dx_{k+1} \cdots dx_n$$

- Independent variables

Recall that two events ~~A~~ A and B are independent if

$$P(A|B) = P(A)$$

or, in other words, if $P(A \wedge B) = P(A) \cdot P(B)$.

Correspondingly, two variables x and y are independent if

$$f(x, y) = f_x(x) \cdot f_y(y)$$

so, if they can be factorized in terms that depend exclusively in x or y.

- Conditional distributions

Take a 2-dim PDF $f(x, y)$ and a fixed value x_0 of variable x.

The conditional distribution of y given x_0 is:

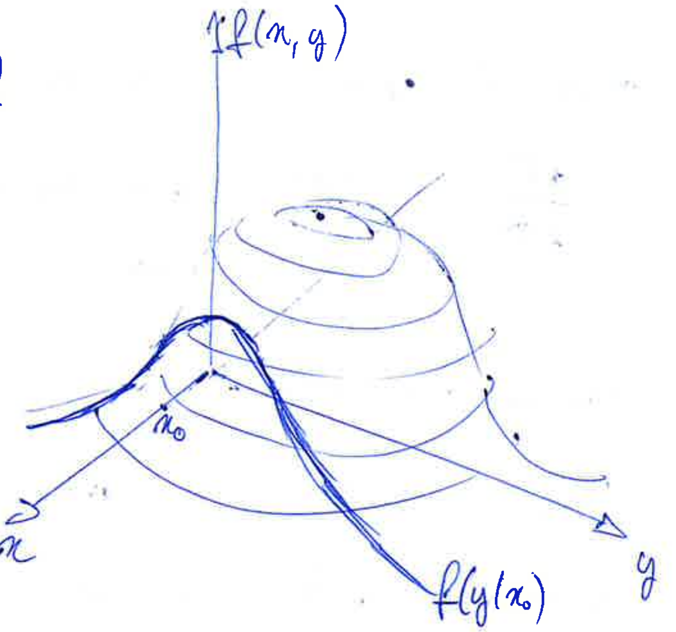
$$f(y|x_0) = \frac{f(x_0, y)}{\underbrace{\int f(x_0, y')\, dy'}_{\text{normalization term}}}$$

This is consistent with $P(B|A) = \dfrac{P(A \wedge B)}{P(A)}$

if we take:

$A = x_0 < \hat{x} < x_0 + \delta x$

$B = y < \hat{y} < y + \delta y$



- Change of variables

→ Discrete case

Take a random variable x, and a second variable $y = Y(x)$.
Assume x can take the values $x_1, \dots, x_n$.
Then y can take the values $y_1 = Y(x_1), \dots, y_n = Y(x_n)$.
The probability of y_i is the sum of the probabilities of all x_j that map into y_i:

$$P(y_i) = \sum_{j:\, Y(x_j) = y_i} P(x_j)$$

→ Continuous case

Take a variable x with PDF $f(x)$, and a second variable $y = Y(x)$.

The PDF of y is: $f(y) = \int \delta(y - Y(x)) f(x)\, dx$

With multiple variables: $f(x', y') = \int \delta(x' - X(x, y))\, \delta(y' - Y(x, y))\, f(x, y)\, dx\, dy$

If the transformation is invertible, the PDF transforms according to the determinant of the Jacobean:

$$f(x_1, \dots, x_n) = \frac{d^n P'}{d^n x} = \frac{d^n P'}{d^n x'} \left| \det\left(\frac{\partial x'_j}{\partial x_i}\right) \right| = f'(x'_1, \dots, x'_n) \left| \det\left(\frac{\partial x'_i}{\partial x_j}\right) \right|$$

In one dimension: $f(x) = f'(x') \left| \dfrac{dx'}{dx} \right| = f(y) \left| \dfrac{dy}{dx} \right|$

EXAMPLE CHANGE OF VARIABLES!

$\boxed{\text{Prob. 12}}$

- ~~Average~~ Expectation operator

Let $g(x)$ be some function of a random variable x with density $f(x)$.

The expectation of $g(x)$ is the number:

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

$\rightarrow$ The expectation is a linear operator:

$$E[a g(x) + b h(x)] = a E(g) + b E(h)$$

- Mean or average

The average is the expectation of the variable itself:

$$E(x) = \langle x \rangle = \bar{x} = \int x f(x)\, dx$$

In the discrete case: $\langle x \rangle = \bar{x} = \sum_{i=1}^{n} x_i P(x_i)$

- Variance

$$V = V(x) = \sigma^2 = E[(x-\bar{x})^2] = \int (x-\bar{x})^2 f(x)\, dx = \langle x^2 \rangle - \langle x \rangle^2$$

In the discrete case: $V = \sum_{i=1}^{n} (x_i - \bar{x})^2 P(x_i)$

- Standard deviation

$$\sigma = \sqrt{V}$$

- Root mean square

$$x_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} x_i^2 P(x_i)} = \sqrt{\langle x^2 \rangle}$$

$\hookrightarrow$ sometimes people denote σ as rms.

$\boxed{\text{Prob. 13}}$

- Covariance

Take two variables x, y with PDF $f(x,y)$.

The covariance is defined as:

$$\boxed{\begin{aligned} \mathrm{cov}(x,y) &= E\left[(x-\bar{x})(y-\bar{y})\right] \\ &\overset{!}{=} E(xy) - E(x)E(y) \end{aligned}}$$

where: $E(xy) = \int xy f(x,y)\,dx\,dy$

$E(x) = \int x f(x,y)\,dx\,dy$

- Correlation

$$\boxed{\mathrm{corr}(x,y) = \rho_{xy} = \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y}}$$

→ always between -1 and 1

where $\sigma_x^2 = E\left[(x-\bar{x})^2\right]$

→ If x and y are independent, we have:

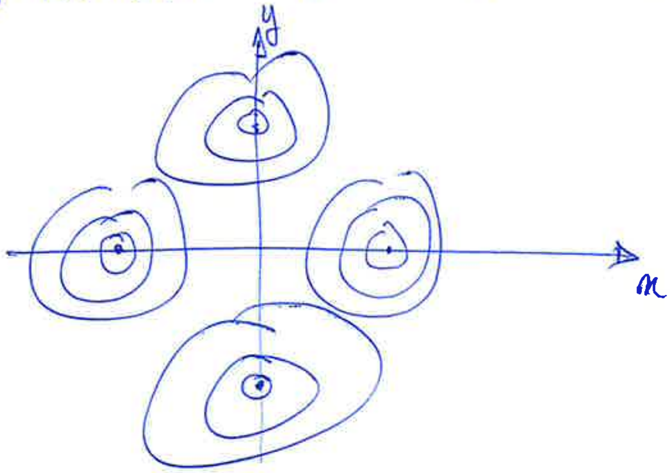
$$E(x,y) = \int xy\, f_x(x) f_y(y)\,dx\,dy = \int x f_x(x)\,dx \int y f_y(y)\,dy = E(x)\,E(y)$$

Therefore, <u>independent variables have null covariance and correlations.</u>

Notice, however, that uncorrelated variables are not necessarily independent!

This is for example the case of:

$$f(x,y) = \frac{1}{4}\Big[g(x;\mu,\sigma)g(y;0,\sigma) + g(x;-\mu,\sigma)g(y;0,\sigma) + g(x;0,\sigma)g(y;\mu,\sigma) + g(x;0,\sigma)\,g(y;-\mu,\sigma)\Big]$$



Prob. 14

- Centroids of a distribution

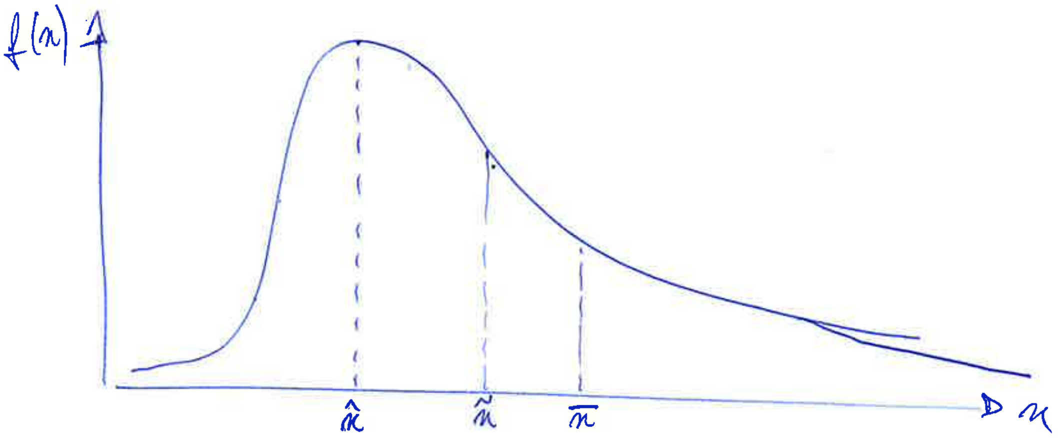
→ Mean: $\bar{x} = \int x f(x)\, dx$

→ Mode: $\hat{x} = \max_x \left(f(x) \right)$

→ Median: $\tilde{x} = x : P(x < \tilde{x}) = P(x > \tilde{x})$

$$= x : \int_{-\infty}^{\tilde{x}} f(x)\, dx = \int_{\tilde{x}}^{+\infty} f(x)\, dx = 0.5$$

Notice that mean, mode and median coincide for a symmetric distributions, but not for an ~~asy~~ asymmetric one!



- Quantiles

The quantity q_α such that:
$$\int_{-\infty}^{q_\alpha} f(x)dx = \alpha = 1 - \int_{q_\alpha}^{+\infty} f(x)\, dx$$

is called α-quantile.

An alternative definition can be given with the cumulative distribution:

$$q_\alpha = x : F(x) = \alpha$$

$\boxed{\text{Prob. 15}}$

• Bernoulli's distribution

The Bernoulli distribution is the discrete distribution of a random variable which takes the value 1 with probability p and the value ∅ with probability 1-p.

It's the probability of extracting a ~~wha~~ red ball from a bag with N balls, of color red or white, and r red balls.

$$p = \frac{r}{N}$$

~~The~~ The distro is expressed by: $f(k;p) = \begin{cases} p & \text{if } k = 1 \\ 1-p & \text{if } k = 0 \end{cases}$

$$= p^k (1-p)^{1-k} \quad \text{for } k \in \{0,1\}$$

~~Mean: $\bar{x} = \sum x f$~~

Mean: $\bar{x} = \sum_{x=0}^{1} x\, f(x;p) = 1 \cdot f(1;p) = p$

~~RM~~ $\langle x^2 \rangle = p$

Variance: $V = \langle x^2 \rangle - \bar{x}^2 = p(1-p)$

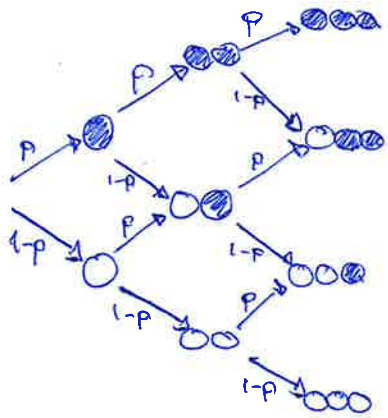
• Binomial distribution

The binomial distribution is the discrete distribution of K independent Bernoulli extractions, each with probability p.

It can be implementing by extracting a ball from the bag k times, and putting it back afterwards.

$$\boxed{P(k;n,p) = \frac{n!}{\underbrace{k!(n-k)!}} p^k (1-p)^{n-k}}$$

→ number of possible paths leading to k successes



...

Mean: $\bar{K} = np$

Variance: $V = np(1-p)$

→ Example Binomial!

- **Multinomial distribution**

Suppose you have a bag with n balls of m colors, and that the probability to extract each color is p_i.

The joint distribution of $k_1, k_2, \dots, k_m$ is:

$$P(k_1, k_2, \dots, k_m; n, p_1, \dots, p_m) = \frac{n!}{k_1! \cdots k_m!} p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m}$$

Average: $\bar{k}_i = N p_i$

Variance: $V[k_i] = N p_i (1 - p_i)$

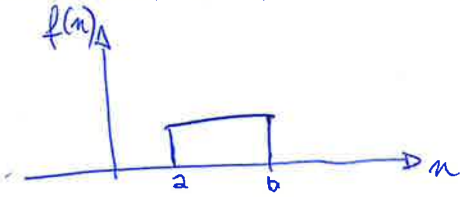
Covariance: $Cov(k_i, k_j) = -N p_i p_j \quad \forall i \neq j$

→ An example of multinomial distribution is a histogram containing n entries distributed in m bins.

- ~~Poisson distribution~~
- **Uniform distribution**

A variable x is uniformly distributed in the interval $[a, b[$ if its PDF is constant in such range, i.e.:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x < b \\ \phi & \text{otherwise} \end{cases}$$



Average: $\bar{x} = \frac{a+b}{2}$ Standard deviation: $\sigma = \frac{b-a}{\sqrt{12}}$

~~Standard deviation~~

$$\langle x^2 \rangle = \int_a^b \frac{x^2}{b-a} dx = \frac{x^3}{3(b-a)}\Big|_a^b = \frac{b^2+ab+a^2}{3(b-a)}$$

Standard deviation: $\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 = \frac{b^2+ab+a^2}{3(b-a)} - \frac{(a+b)^2}{4} \cdot \frac{(b-a)}{3(b-a)}$

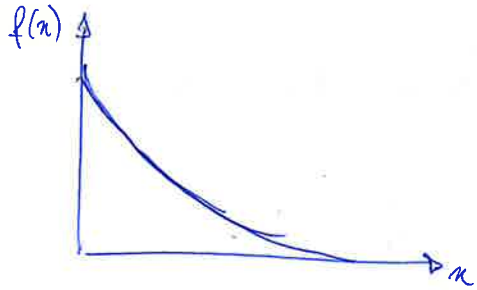
$$= \frac{4b^2+4ab+4a^2}{12(b-a)}$$

- Exponential distribution

Take a variable $x \geq 0$, and a constant $\lambda > 0$.

An exponential distro has the form: $f(x;\lambda) = \lambda e^{-\lambda x}$

Average: $\bar{x} = \dfrac{1}{\lambda}$



Standard deviation: $\sigma = \dfrac{1}{\lambda}$

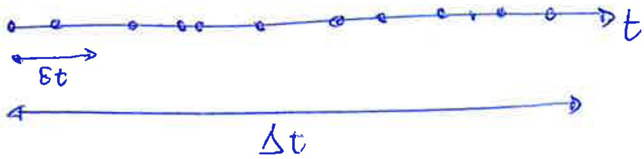
~~Exercise: exponential from uniform!~~

- Poisson distribution

Consider a uniformly distributed variable t over an interval $[0, \Delta t]$.
t could be a space or a time variable, like the coordinate of some particle hit on a pixelated detector, or the time of arrival of some particle.

Assume t is extracted n times in Δt.

The rate of extractions is $r = \dfrac{n}{\Delta t}$.



Let's consider only the extractions k in a shorter interval δt, which are clearly binomial distributed.

Assume n and Δt are constant, and take the limits $n \to \infty$ and $\Delta t \to \infty$, keeping their ratio r fixed.

The expected value ν of the number of extractions in δt is:

$$\nu = \langle k \rangle = \frac{N}{\Delta t} t = rt$$

And k follows a Binomial:

$$P(k;n,\nu) = \frac{n!}{k!(n-k)!} \left(\frac{\nu}{n}\right)^k \left(1-\frac{\nu}{n}\right)^{n-k}$$

$$= \frac{\nu^k}{k!} \cdot \underbrace{\frac{n(n-1)\cdots(n-k+1)}{n^k}}_{\to 1} \cdot \underbrace{\left(1-\frac{\nu}{n}\right)^n}_{e^{-\nu}} \cdot \underbrace{\left(1-\frac{\nu}{n}\right)^{-k}}_{1}$$

$\lim_{k\to\infty}$:

We have then the Poisson distro:

$$P(k;\nu) = \frac{\nu^k e^{-\nu}}{k!}$$

$\boxed{\text{Prob. 18}}$

Average : $\bar{k} = \nu$

Standard deviation : $\sigma = \sqrt{\nu}$

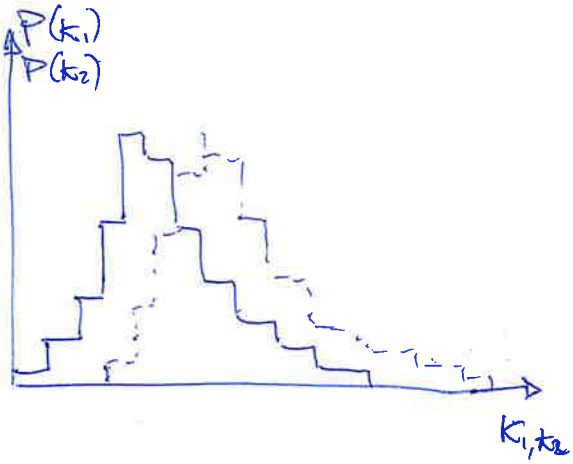
<u>Exercise :</u> Poisson

Properties of Poisson distribution :

→ For large ν, a Poisson distribution can be approximated with a Gaussian with $\mu = \nu$ and $\sigma = \sqrt{\nu}$

→ A binomial distribution with $p \ll 1$ can be approximated with a Poisson distribution with $\nu = pn$

→ If two variables k_1 and k_2 are Poisson distributed with expectation values ν_1 and ν_2, their sum $k = k_1 + k_2$ is Poisson-distributed with expectation value $\nu = \nu_1 + \nu_2$.



$$P(k;\nu_1,\nu_2) = \sum_{j=0}^{k} \frac{e^{-\nu_1}\nu_1^{j}}{j!} \cdot \frac{e^{-\nu_2}\nu_2^{k-j}}{(k-j)!}$$

$$= \underbrace{\sum_{j=0}^{k} \frac{k!}{j!(k-j)!} \cdot \nu_1^{j}\,\nu_2^{k-j}}_{\text{Binomial expansion}} \cdot \frac{e^{-\nu_1}e^{-\nu_2}}{k!}$$

$$= (\nu_1+\nu_2)^k \cdot \frac{e^{-\nu_1}e^{-\nu_2}}{k!}$$

$$= \frac{e^{-\nu}\nu^{k}}{k!}$$

This can be expanded to any number of Poisson variables.

→ Randomly picking with probability ε from a Poisson process gives again a Poisson process.

Take a Poisson ~~distribution~~ variable n_0 with expectation value ν_0.

Then a Binomial variable k with probability ε and sample size n_0 is distributed as a Poisson with average $\nu = \varepsilon \nu_0$.

$$P(k;\nu_0,\varepsilon) = \sum_{n=0}^{\infty} \frac{e^{-\nu_0}\nu_0^{n}}{n!} \cdot \frac{n!}{k!(n-k)!} \varepsilon^{k}(1-\varepsilon)^{n-k}$$

$$= \sum_{n=0}^{\infty} \frac{\varepsilon^k \nu_0^k}{k!} \cdot \frac{e^{-\nu_0}}{(n-k)!} \nu_0^{n-k} (1-\varepsilon)^{n-k}$$

$$= \frac{\nu^k}{k!} e^{-\nu_0} \sum_{n=0}^{\infty} \frac{(\nu_0(1-\varepsilon))^{n-k}}{(n-k)!} = \frac{\nu^k e^{-\nu_0} e^{(1-\varepsilon)\nu_0}}{k!}$$

$$= \frac{\nu^k e^{-\nu}}{k!}$$

This is the case of a detector with efficiency ε, which measures a Poisson process with expectation value ν_0!

Examples:

Suppose you have n unstable isotopes, with decay rate λ (number of ~~emitted~~ ~~The distribution Suppose you measure~~ decays per unit time).

In a time t, ~~we~~ the probability of a decay is ~~then~~ $p = \lambda t$

The distribution of the number of decayed nuclei in a time t is:

a) Binomial, if $p \gtrsim 0.1$

b) Poisson, if $p \lesssim 0.1$

Prob. 20

- Normal (Gaussian) distribution

$$\boxed{g(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$ with x continuous

For $\mu = 0$ and $\sigma = 1$, we have a standard-normal distribution:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

~~The cumulative~~

Mean : μ

Standard deviation : σ

Cumulative of standard normal: $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx = \frac{1}{2}\left[\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) + 1\right]$

Cumulative of normal: $G(x;\mu,\sigma) = \frac{1}{2}\left[\operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\,\sigma}\right) + 1\right]$

Properties:

→ If x_1 and x_2 are Gaussian distributed with means μ_1, μ_2 and sigmas σ_1, σ_2,
their ~~sum~~ combination $x = a x_1 + b x_2$ is Gaussian distributed
with mean $\mu = a\mu_1 + b\mu_2$
and STD $\sigma = \sqrt{a^2\sigma_1^2 + b^2\sigma_2^2}$

→ The sum of two Gaussians is not a Gaussian!

- Multi-variate distribution

A multi-variate distribution is a Gaussian in ~~more~~ dim > 1.

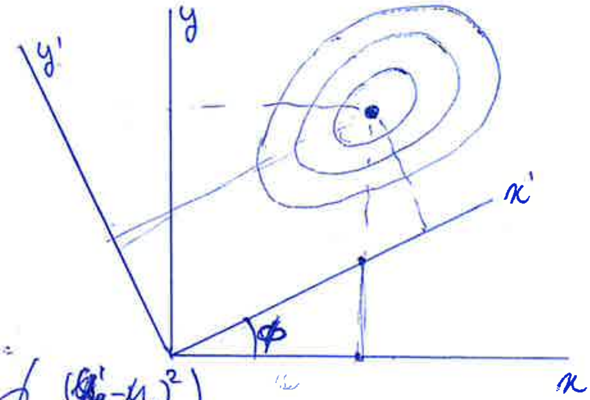
~~$g(x,y)$~~

Dim = 2 : $g(x,y) = \frac{1}{2\pi\,|C|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x,y)\,C^{-1}\begin{pmatrix} x \\ y \end{pmatrix}\right]$

where C is the covariance matrix : $C = \begin{pmatrix} \sigma_x^2 & \rho_{xy}\sigma_x\sigma_y \\ \rho_{xy}\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$

Prob. 21

One can also forget about the covariance matrix and define a rotation of the ~~plane~~ system of reference:



$$\begin{cases} x' = \cos\phi \cdot x + \sin\phi \cdot y \\ y' = -\sin\phi \, x + \cos\phi \cdot y \end{cases}$$

$$\cancel{G(x', y'; \mu_1, \mu_2, \sigma_1, \sigma_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(x'_\phi - \mu_1)^2}{2\sigma_1^2}\right) \exp\left(-\frac{(y'_\phi - \mu_2)^2}{\sigma}\right)}$$

$$G(x', y') = \frac{1}{2\pi\,\sigma_1\sigma_2} \exp\left(-\frac{(x' - \mu_1)^2}{2\sigma_1^2}\right) \exp\left(-\frac{(y' - \mu_2)^2}{2\sigma_2^2}\right)$$

↳ By inverting the rotation, one can get rid of the correlation.

General formula for dim = n:

$$g(\vec{x}) = \frac{1}{(2\pi)^{n/2} \, |C|^{1/2}} \exp\left[-\frac{1}{2}(\vec{x} - \vec{\mu})^T C^{-1} (\vec{x} - \vec{\mu})\right]$$

- Chi-square distribution

A χ^2 random variable with n degrees of freedom is the sum of n standard normal variables.

$$\boxed{f(\chi^2; n) = \frac{2^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \chi^{n-2} e^{-\frac{\chi^2}{2}}}$$

where Γ is the gamma-function, which is the extension of the factorial.
For integers, $\Gamma(n) = (n-1)!$

Expectation value: $\langle \chi^2 \rangle = n$
Standard deviation: $\sigma = \sqrt{2n}$

<u>Exercise: χ^2 distro</u>

Prob 22