Basic Concepts of Probability and Probability Distributions

Arun Nayak [1]

ML4HEP-2025, IISER, Kolkata

14 - 15th May 2025

[1] Institute of Physics, Bhubaneswar

Reference Books

1. G. Cowan, Statistical Data Analysis, Oxford.
2. L. Lista, Statistical Methods for Data Analysis in Particle Physics, Springer.
3. L. Lyons, Statistics for Nuclear and Particle Physics.
4. P. R. Bevington and D. K. Robinson, Data Reduction and Error Analysis.

(Some Figures in these slides are taken from refs. 1 and 2)

Measurements

Purpose of doing experiments:

- Parameter Determination

  e.g. Measuring the value of gravitational constant, G
- Hypothesis Testing

  e.g. Checking whether the gravitational constant varies over time.

In reallity there may be some degree of overlap

Measurements and Uncertainties

Why estimate Uncertainties/Errors?

- It gives a degree of accuracy of the measurement – the estimation of how far the measured value is from the true value of the parameter
- It can also help to test a theory or compare it to the results of other experiments

For example: suppose the last measured value of $G = 6.6743 \times 10^{-11}\ m^3 \cdot kg^{-1} \cdot s^{-2}$, and our measured value is $6.675 \times 10^{-11}\ m^3 \cdot kg^{-1} \cdot s^{-2}$.

1. If $\epsilon\ =\ \pm 0.002 \times 10^{-11} \implies$ Consistent with old one
2. If $\epsilon\ =\ \pm 0.0002 \times 10^{-11} \implies$ Inconsistent with old one, possibility of variation over time
3. If $\epsilon\ =\ \pm 0.2 \times 10^{-11} \implies$ The accuracy is too low, need to perform better experiment.

Uncertainties

In particle physics there are various sources of uncertainties

- Random Errors
  - Inherent quantum mechanical fluctuations
  - Random measurement errors even without quantum effects
- Systematic Errors
  - Errors in the measuring device, e.g. mistakes in calibrations
  - Errors in theoretical predictions, e.g. cross sections of simulated events
  - ...

The factors behind the uncertainties in principle could be known, but practically not.

Uncertainty can be quantified using the concept of Probability

Probability

Can be defined using set theory

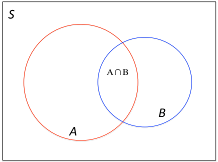
Consider a set S (sample space), with subsets A, B

Kolmogorov Axioms:

For all $A \subset S$, $P(A) \geq 0$

$P(S) = 1$

If $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$



From these axioms further properties can be derived, e.g.

$0 \leq P(A) \leq 1$

$P(\overline{A}) = 1 - P(A)$, $\overline{A}$ is complement of A

$P(A \cup \overline{A}) = 1$

$P(\emptyset) = 0$

if $A \subset B$, $P(A) \leq P(B)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Conditional probability

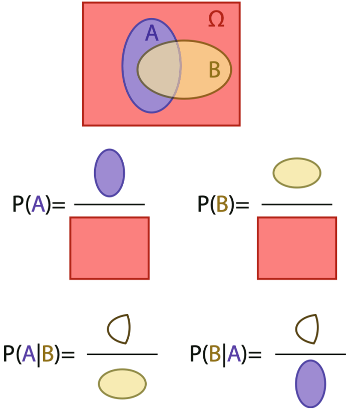
Define conditional probability of A given B (provided $P(B) \neq 0$)

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

A and B are said to be independent, if, $P(A \cap B) = P(A)P(B)$

If A, B are independent, $P(A|B) = P(A)$.

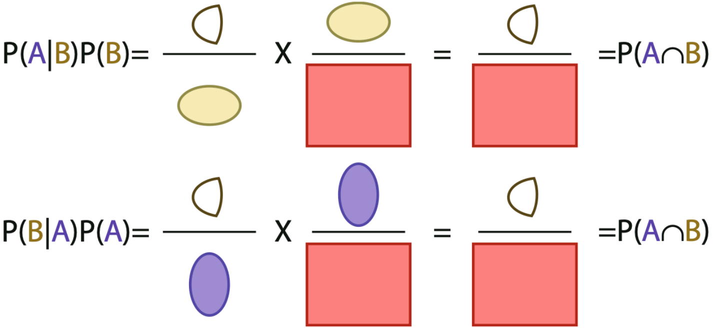
N.B. Don't confuse independent subsets as disjoint subsets (i.e. $P(A \cap B) = \emptyset$)



Conditional probability – contd.

From the definition of conditional probability,

$$P(A|B)P(B) \ = P(B|A)P(A), \qquad \text{since } P(A \cap B) \ = \ P(B \cap A)$$



Hence,

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

Bayes' theorem

Bayes' theorem

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

How do we interpret this equation?

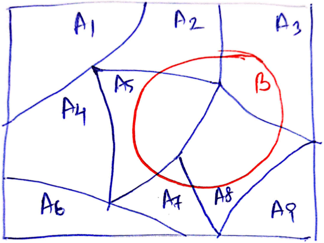
- Before knowing B is true, our *degree of belief* in the event A $=$ prior probability $P(A)$
- After knowing B is true, our *degree of belief* in the event A changes, and becomes equal to the posterior probability $P(A|B)$

That is the *degree of belief* in event A is updated based on the state of our knowledge that B is true.

Law of total probability

Consider a sample space S divided into disjoint subsets A_i,

i.e. $S = \cup_i A_i$, and $A_i \cap A_j = \emptyset$ for $i \neq j$.



Consider a subset $B \subset S$, it can be expressed as

$$B = B \cap S = B \cap (\cup_i A_i) = \cup_i (B \cap A_i)$$

$$\implies P(B) = P(\cup_i (B \cap A_i)) = \sum_i P(B \cap A_i)$$

$$\implies P(B) = \sum_i P(B|A_i) P(A_i)$$ law of total probability

Thus, Bayes' theorem becomes

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

Example of using Bayes' theorem

Consider a disease D carried by 0.1% of people

i.e., the prior probabilities are

$P(D) = 0.001,$
$P(no\ D) = 0.999$

Consider a test that identifies the disease, the result is $+$ve or -ve

Suppose the probabilities to (in)correctly identify a person with the disease are,

$P(+|D) = 0.98,$
$P(-|D) = 0.02$

Similarly, suppose the probabilities to (in)correctly identify a healthy person

$P(+|no\ D) = 0.01,$
$P(-|no\ D) = 0.99$

What is the probability to have the disease if someone is tested $+$ve?

Example contd..

We can calculate it using the Bayes' theorem

i.e., the probability to have the disease given a +ve test result is

$$\begin{aligned} P(D|+) \;&=\; \frac{P(+|D)P(D)}{P(+|D)P(D) \;+\; P(+|no\ D)P(no\ D)} \\ &= \frac{0.98 \times 0.001}{0.98 \times 0.001 \;+\; 0.01 \times 0.999} \\ &= \; 0.089 \quad \text{(posterior probability)} \end{aligned}$$

What does it mean?

Patient's view: Probabilty for him to have the disease is 8.9%.

Doctor's view: 8.9% of people like this have the disease.

Interpretation of probability

Two definitions of probabilities: $frequentist\ probability$ and $Bayesian\ probability$

- frequentist: Relative frequency

  Suppose X, Y, ... are outcomes of a repeatable experiment

  $$P(X) \ = \ lim_{N\rightarrow\infty} \frac{\text{number of times the outcome is X}}{N}$$

  e.g., Probability of monsoon reaching India in 1st weak of June.

  Probability for producing a certain particle in pp collisions at a particular centre of mass energy.

  Note: $P(X) \ \geq \ 0$, $\quad P(S) \ = \ 1$, i.e. consistent with axioms of probability

Interpretation of probability - II

- Bayesian: Subjective probability

  Degree of belief that something will happen.
  In this case X, Y, ... are hypotheses (statements that are true or false)

$$P(X) = \text{degree of belief that X is true}$$

e.g.

It will rain tomorrow.

The mass of SUSY candidate particle is between 2 and 3 TeV

The sample space S (also called as "hypothesis space") consists of elementary hypotheses that are mutually exclusive. One of the hypotheses must necessarily be true.

So, $P(X) \geq 0$, and $P(S) = 1$

Consistent with Kolmogorov axioms

Bayesian Statistics

Subjective probability is the basis of Bayesian statistics e.g., Probability for a hypothesis to be true given observations (data) from an experiment.

$$P(H|data) = \frac{P(data|H)P(H)}{\sum P(data|H_i)P(H_i)}$$

$P(H)$ is the **prior probability**, i.e. knowledge or degree-of-belief in hypothesis H before seeing the data.
$P(data|H)$ is the probability of observing this data assuming the hypothesis **(likelihood)**.
$P(H|data)$ is the **posterior probability**, i.e., after seeing the result/data of the experiment.

There is no general rule for assigning the priors (subjective!). However, given a prior it says how the probabilities change in the light of experimental data

Probability Distributions

A variable that takes on a specific value for each element of the sample space is called a **random variable**.
The outcome of a random event is not predictable, only the probabilities of the possible outcomes are known.

Random variable can be discrete or continuous. Suppose the outcome of an experiment is a continuous variable x, probability for x to lie between x and $x + dx$ is

$$P(x\ in\ [x, x + dx]) \ = \ f(x)dx$$

$f(x)$ is probability density function (PDF)

Normalization condition

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

Probability Distributions – contd.

For discrete variables x_i, $i = 1, 2, ...$,

$$P(x_i) \;=\; f_i$$

Normalization condition

$$\sum_i f_i \;=\; 1$$

in continous PDF formalism

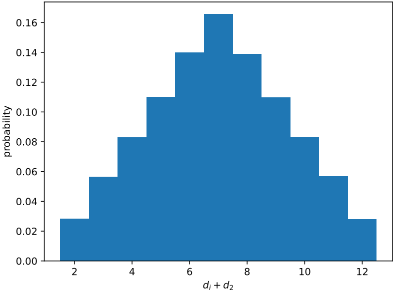
$$f(x) = \sum_i f_i \delta(x - x_i)$$

The normalization then

$$\int_{-\infty}^{+\infty} f(x)dx \;=\; \sum_i f_i \int_{-\infty}^{+\infty} \delta(x - x_i)dx = \sum_i f_i \;=\; 1$$

Probability Distributions – Example

Consider
tossing of two dice:



What
are the outcomes and their corresponding probabilities?

Possible outcomes are
$2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$

The corresponding probabilities are:
$1/36, 2/36, 3/36, 4/36, 5/36,$
$6/36, 5/36, 4/36, 3/36, 2/36, 1/36$

Statistical Indicators

Consider random variable x with pdf $f(x)$.
Expectation value (also called **population mean**) is defined as
$< x > \ = \ \int x f(x) dx \ = \ \mu$

Variance:
$$V[x] = \sigma^2(x) = \int (x \ - \ < x >)^2 f(x) dx =< x^2 > - < x >^2$$
$$=< x^2 > \ - \ \mu^2$$

σ is called the **standard deviation**.
$\sigma \ \sim$ width of pdf (same units as x).

For discrete variables
The *mean* or *arithmetic mean* is defined as
$\bar{x} \ = \ \frac{1}{N} \sum_{i=1}^{N} x_i$
and variance $Var[x] \ = \ \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$

Statistical Indicators – contd.

The kth moment of x is defined as

$$m'_k(x) \;=\; \int x^k f(x)dx$$

and the kth central moment of x is defined as

$$m_k(x) \;=\; <(x-<x>)^k> \;=\; \int (x-<x>)^k f(x)dx$$

The mode of a PDF f is the value $\hat{x}$ corresponding to the maximum of $f(x)$

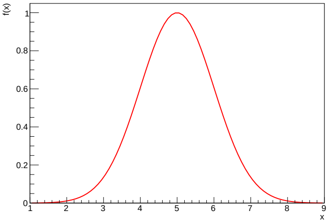
$$f(\hat{x}) \;=\; \max f(x) \text{ or } \hat{x} \;=\; \arg\max f(x)$$

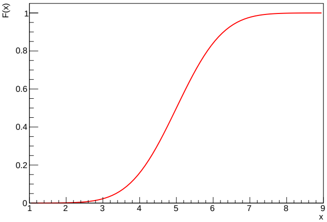
Cumulative distribution function

Probability to obtain an outcome less than or equal to x:

$$F(x) \;=\; \int_{-\infty}^{x} f(x')dx' \qquad\qquad \text{(cdf)}$$

$$\text{Or define pdf as,} \qquad f(x) \;=\; \frac{\partial F(x)}{\partial x}$$





A relative quantity is **quantile of order** α or α-**point**:

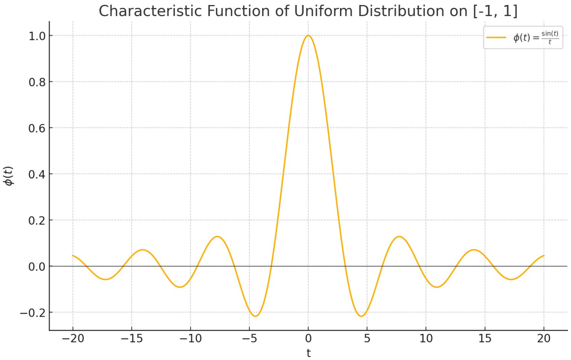
$$x_\alpha \;=\; F^{-1}(\alpha) \qquad\qquad x_{1/2} \text{ is } \mathbf{median}$$

Characteristic function

The characteristic function (CF) of a r.v. x, with PDF $f(x)$ and CDF $F(x)$, is defined as

$$\varphi_x(t) = < e^{itx} > = \int_{\mathbb{R}} e^{itx} f(x) dx = \int_{\mathbb{R}} e^{itx} dF(x)$$

i.e. It is the *Fourier transform* of PDF with sign reversal.



Characteristic function – contd.

- CFs can be used to find **moments** of a r.v., i.e. for k^{th} moment,

$$< x^k >= i^{-k} \left[\frac{d^k}{dt^k} \varphi_x(t) \right]_{t=0} = i^{-k} \varphi_x^{(k)}(0)$$

  For example, for Gaussian PDF, $\varphi_x(t) = exp(i\mu t - \frac{1}{2}\sigma^2 t^2)$, Thus, the mean

$$< x >= i^{-1} \left[\frac{d}{dt} \varphi_x(t) \right]_{t=0} = i^{-1} \left[\left(i\mu - \sigma^2 t \right) \varphi_x(t) \right]_{t=0} = \mu$$

  Similarly $< x^2 >= \mu^2 + \sigma^2$.
- If $x_1, .., x_n$ are independent r.v. and $w_1, .., w_n$ are some constants, the the CF of linear combination is

$$\varphi_{w_1 x_1 + \cdots + w_n x_n}(t) = \varphi_{x_1}(w_1 t) \cdots \varphi_{x_n}(w_n t)$$

  e.g. for sum two variables,

$$\varphi_{x_1 + x_2}(t) = \varphi_{x_1}(t) \cdot \varphi_{x_2}(t)$$

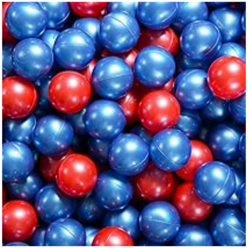
Common Discrete Probability Functions

These are some of the common discrete probability distribution functions:

Bernoulli	$f(1) = p,\ f(0) = 1 - p$	
Binomial	$f(n; N, p)$	${}^N C_n p^n (1-p)^{N-n}$
Mutinomial	$f(n_1, .., n_m; N, p_1, .., p_m)$	$\frac{N!}{n_1!...n_m!} p_1^{n_1} ... p_m^{n_m}$
Poisson	$f(n; \nu)$	$\frac{\nu^n}{n!} e^{-\nu}$

Bernoulli Distribution

Consider a bag containing a number of balls each having one of the two possible colours.
Assume number of RED balls = R
and number of BLUE balls = B



Probability
to randomly pick a RED ball, $p = \frac{R}{R+B}$

A variable x = the outcomes
of a trial, is called Bernoulli variable,
i.e. $x = 0(\text{failure})$ or $1(\text{success})$

The probability distribution of x is simply

$$P(1) = p,$$
$$P(0) = 1 - p$$

Bernoulli Distribution – contd.

The average of Bernoulli variable

$$< x >= 0 \times P(0) + 1 \times P(1) = P(1) = p$$

Variance

$$V[x] =< x^2 > - < x >^2 = 0^2 \times P(0) + 1^2 \times P(1) - p^2 = p - p^2 = p(1-p)$$

The ratio

$$O = \frac{p}{1-p}$$

is called *odds*

Binomial Distribution

Consider N independent trials (or observations), each with two possible outcomes: *success* or *failure*
Probability of *success* $= p$

The total probability to have "n" suceesses in N trials is

$$f(n; N, p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

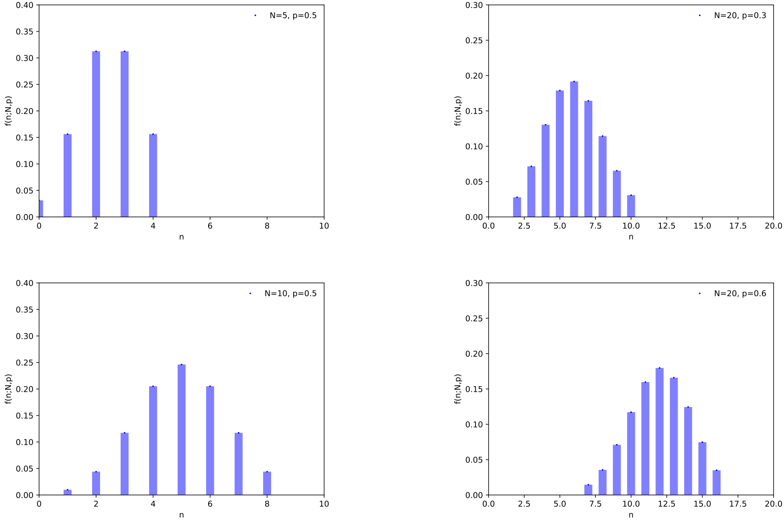
where $n = 0, 1, .., N$.
This is known as the **binomial distribution**

And
$\frac{N!}{n!(N-n)!}$ is called *binomial coefficient*

Exercise: Show that $< n > \; = \; Np$ and $V[n] \; = \; Np(1-p)$

Binomial Distribution – contd.

Binomial distributions for different values of N and p



Poisson Distribution

When N is very large and p is very small, such that $\nu = Np$ (rate parameter) is finite, the binomial distribution becomes *Poisson distribution*.

Writing the binomial distribution in terms of ν

$$
\begin{aligned}
f(n; N, \nu) &= \frac{N!}{n!(N-n)!}\left(\frac{\nu}{N}\right)^n\left(1-\frac{\nu}{N}\right)^{N-n} \\
&= \left(\frac{\nu^n}{n!}\right)\frac{N(N-1)...(N-n+1)}{N^n}\left(1-\frac{\nu}{N}\right)^N\left(1-\frac{\nu}{N}\right)^{-n}
\end{aligned}
$$

Poisson Distribution – contd.

In the limit $N \to \infty$

- $\lim_{N\to\infty} \frac{N(N-1)...(N-n+1)}{N^n} \ = \ 1$
- $\lim_{N\to\infty} \left(1 - \frac{\nu}{N}\right)^N \ = \ \lim_{N\to\infty} exp\left(Nln\left(1 - \frac{\nu}{N}\right)\right) \ = \ e^{-\nu}$
- $\lim_{N\to\infty} \left(1 - \frac{\nu}{N}\right)^{-n} \ = \ 1$, since $\nu/N \to 0$

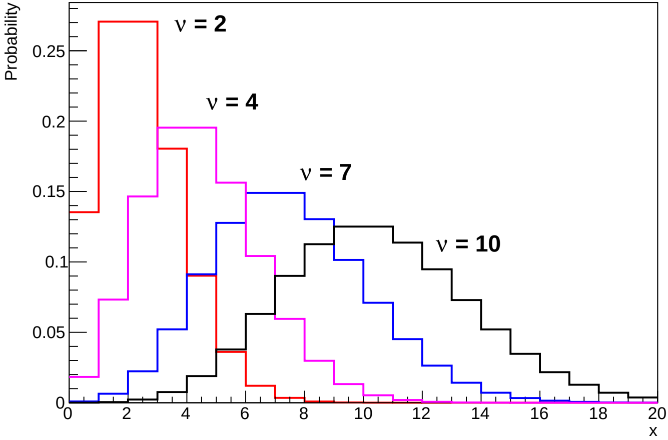
Thus, the distribution becomes

$$f(n;\nu) = \frac{\nu^n}{n!}e^{-\nu}$$

Exercise Show that $< n >= \nu$ and $V[n] = \nu$

Poisson Distribution – contd.

Poisson distribution for different mean value



Common Probability Density Functions for continous variables

Uniform	$f(x; a, b)$	$1/(b-a)$	$x \in [a, b]$
Exponential	$f(x; a)$	$\frac{1}{a} e^{x/a}$	$x \in [0, \infty)$
Gaussian	$f(x; \mu, \sigma)$	$\frac{1}{\sqrt{2\pi\sigma^2}} exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$	$x \in (-\infty, \infty)$
Log-normal	$f(x; \mu, \sigma)$	$\frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x} exp\left(\frac{-(logx-\mu)^2}{2\sigma^2}\right)$	$x \in [0, \infty)$
Chi-square	$f(x; n)$	$\frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}$	$x \in [0, \infty)$
Breit-Wigner	$f(x; \Gamma, M)$	$\frac{1}{\pi} \frac{\Gamma/2}{\Gamma^2/4+(x-M)^2}$	$x \in (-\infty, \infty)$
Gamma	$f(x; a, b)$	$\frac{1}{\Gamma(a)b^a} x^{a-1} e^{-x/b}$	$x \in [0, \infty)$

Uniform Distribution

If the variable is uniformly distributed in the range [a, b), the PDF is constant in the range.

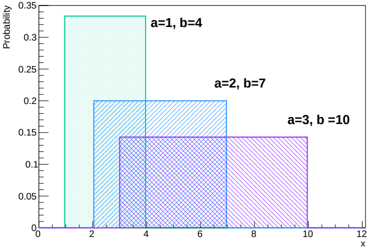
$$p(x) = \begin{cases} \frac{1}{b-a} & \text{if a} \leq \text{x} < \text{b} \\ 0 & \text{otherwise} \end{cases}$$

Mean

$$< x >= \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}$$

and variance

$$V[x] =< (x- < x >)^2 >$$
$$= \frac{1}{12}(b-a)^2$$



Exponential Distribution

The exponential PDF for continous variable $x(0 \leq x < \infty)$

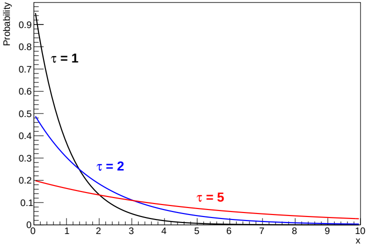
$$f(x;\tau) = \frac{1}{\tau} e^{-x/\tau}$$

The mean

$$< x >= \frac{1}{\tau} \int_0^\infty x e^{-x/\tau} dx = \tau$$

and variance

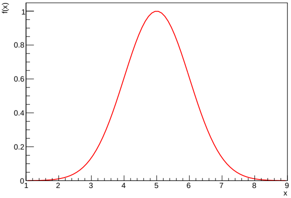
$$V[x] = \frac{1}{\tau} \int_0^\infty (x-\tau)^2 e^{-x/\tau} dx$$
$$= \tau^2$$



Example: The distribution of decay time of an unstable particle, where τ is the mean lifetime of the particle.

Gaussian (Normal) Distribution

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



The mean and variance are:
$< x > \;=\; \mu,\; V[x] \;=\; \sigma^2$

Special case: $\mu = 0,\ \sigma = 1$ (standard Gaussian)

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

And the cumulative distribution:

$$\Phi(z) \;=\; \int_{-\infty}^{z} \phi(z')dz' = \frac{1}{2}\left[erf\left(\frac{z}{\sqrt{2}}\right) + 1\right]$$

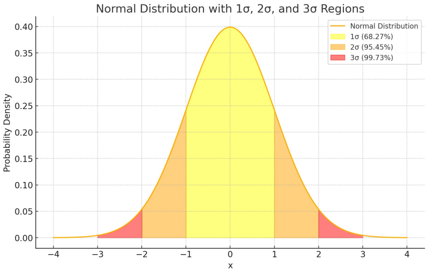
If x follows Gaussian with μ, σ, then $z = (x-\mu)/\sigma$ follows $\phi(y)$.

Normal Distribution – contd.

The probability for a Gaussian distribution corresponding to the symmetric interval around μ $[\mu - Z\sigma, \mu + Z\sigma]$ is given by

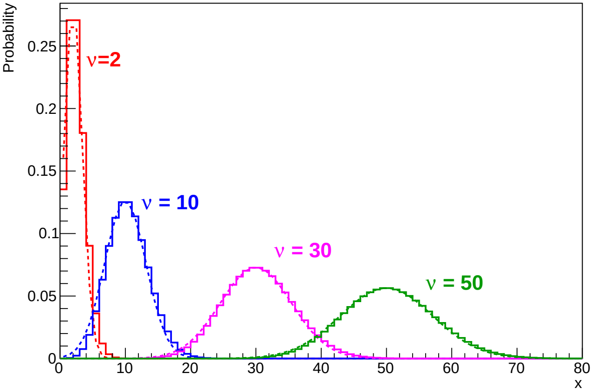
$$P(Z\sigma) = \frac{1}{\sqrt{2\pi}} \int_{-z}^{z} e^{-z'^2/2} dz' = \phi(Z) - \phi(-Z) = erf\left(\frac{Z}{\sqrt{2}}\right)$$

Most frequently used are: 1σ, 2σ, and 3σ $(Z = 1, 2, 3)$, having corresponding probabilities of 68.27%, 95.45% and 99.73%.



Poisson vs Gaussian Distribution

For large ν, Poisson distribution can be approximated with a Gaussian distribution having mean $\mu = \nu$ and standard deviation $\sigma = \sqrt{\nu}$.



Central Limit Theorem and Gaussian pdf

The theorem states that:
If x_i are n independent random variables with variances σ_i^2, then, in the limit $n \to \infty$ the **sum**

$$y = \sum_{i=1}^{n} x_i$$

becomes a **Gaussian random variable** with

$$< x > = \sum_{i=1}^{n} \mu_i \qquad V[y] = \sum_{i=1}^{n} \sigma_i^2$$

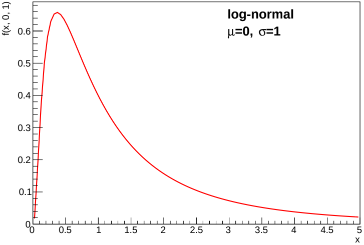
irrespective of the form of the individual pdfs of the x_i.
(The proof can be found in the references).
Measurement errors are often the sum of large number of small contributions, so, these are normally treated as Gaussian random variables.

Log-normal distribution

If y is Gaussian with mean μ and variance σ^2, then $x = e^y$ follows the **log-normal** distribution.

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{x}exp\left(\frac{-(logx-\mu)^2}{2\sigma^2}\right)$$

Recall CLT: If $y = \sum_{i=1}^{n} x_i$, then y is Gaussian.
Similarly, if $y = \prod_{i=1}^{n} x_i$, then y follows log-normal.
Therefore,
random errors that change the result by a multiplicative factor are modeled as log-normal.



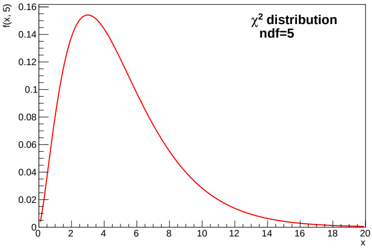
Exercise: show that
$< x >= exp(\mu + \frac{1}{2}\sigma^2)$
$V[x] =$
$exp(2\mu + \sigma^2)[exp(\sigma^2) - 1]$

Chi-square (χ^2) distribution

The χ^2 distribution for the continuous random variable z ($z \geq 0$),

$$f(z;n) = \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2}$$

n = number of degrees of freedom (dof).
$< z > \ = \ n, \qquad v[z] \ = \ 2n$
Given n independent Gaussian variables x_i, with mean μ_i and variances σ_i^2, $z \ = \ \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2}$ follows χ^2 distribution with n dof.



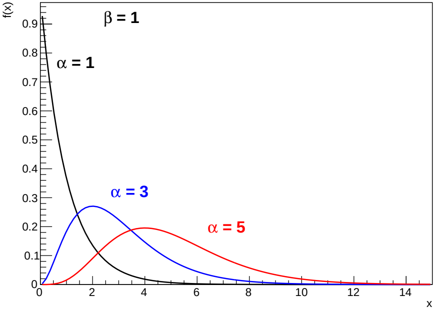
Plays important role in goodness-of-fit test, especially in method of least squares.

Gamma Distribution

The Gamma distribution for x ($x \geq 0$),

$$f(x; \alpha, \beta) = \frac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$$

where $\alpha, \beta > 0$.
α is called shape parameter and β is scale parameter



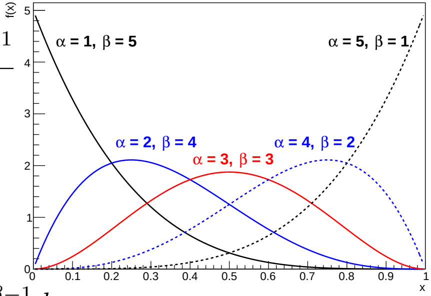
Exercise: Show that mean $= \alpha/\beta$ and variance $= \alpha\beta^2$

Beta Distribution

The Beta distribution
for x $(0 < x < 1)$,

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$$

where $\alpha, \beta > 0$,



$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$$
$$= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

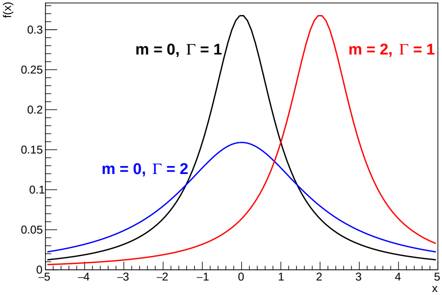
Breit–Wigner Distribution

Breit–Wigner distribution (aka *Lorentz distribution or Cauchy distribution*) is given by

$$f(x;\Gamma,m) = \frac{1}{\pi}\frac{\Gamma}{(x-m)^2+\Gamma^2}$$

where,
m determines the position of the peak, and 2Γ = Full width at half maximum.

Note: The mean and variance are undefined, since both $\int x f(x)dx$ and $\int x^2 f(x)dx$ are divergent.



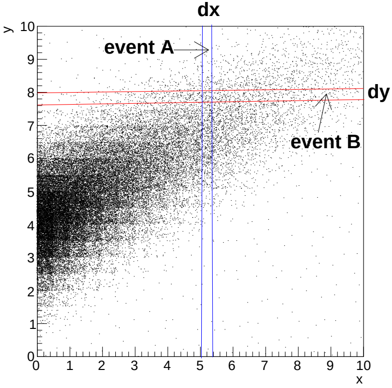
Joint pdf

Suppose the result of
a measurement is characterized
by several quantities,
the joint pdf $f(x, y)$ is given by

$$P(A \cap B) \;=\; f(x, y)dxdy$$

Normalization:

$$\int\int f(x, y)dxdy \;=\; 1$$

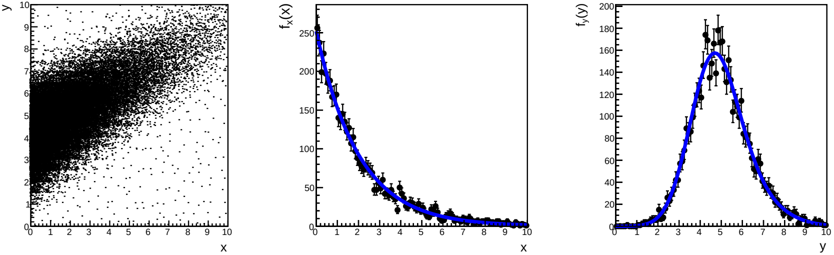


The pdf of x regardless
of y – **marginal pdf**:

$$f_x(x) \;=\; \int f(x, y)dy$$

Marginal pdf

It is the projection of joint pdf onto individual axes.



Conditional pdf

Probability for y to be in $[y, y+dy]$, given x is in $[x, x+dx]$ (i.e. treating x to be fixed)
Definition of conditional probability:

$$P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{f(x,y)dxdy}{f_x(x)dx}$$

$$\text{conditional pdf}: \quad h(y|x) = \frac{f(x,y)}{f_x(x)}, \qquad g(x|y) = \frac{f(x,y)}{f_y(y)}$$

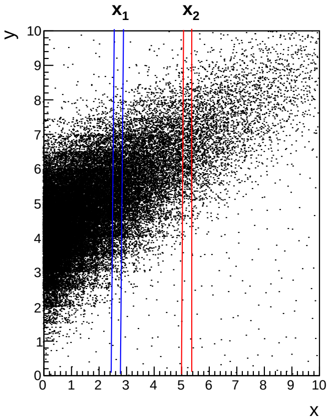
Thus, Bayes' theorem becomes:

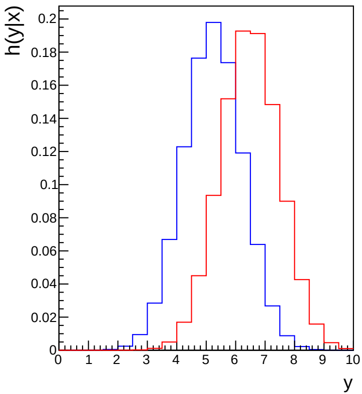
$$g(x|y) = \frac{h(y|x) f_x(x)}{f_y(y)}$$

Since, $A,\ B$ are independent if $P(A\cap B) = P(A)P(B)$
$\implies x,\ y$ are independent if $f(x,y) = f_x(x)f_y(y)$. i.e. the corresponding joint pdf factorizes.

Conditional pdfs

Conditional pdfs $h(y|x_1),\ h(y|x_2)$ obtained from joint pdf $f(x,y)$.





Covariance

The covariance of two random variables is defined as

$$cov[x,y] \text{ (or } V_{xy}) =< (x- < x >)(y- < y >) >$$
$$=< xy > - < x >< y >= \int\int xyf(x,y)dxdy - \mu_x\mu_y$$

Correlation coefficient defined as

$$\rho_{xy} = \frac{cov[x,y]}{\sigma_x\sigma_y} \qquad -1 \le \rho_{xy} \le 1$$

If x and y are independent, i.e. $f(x,y) = f_x(x)f_y(y)$,

$$< xy > = \int\int xyf(x,y)dxdy$$
$$= \int xf_x(x)dx \int yf_y(y)dy = \mu_x\mu_y$$
$$\Longrightarrow \quad cov[x,y] = 0 \qquad (x \text{ and } y \text{ are uncorrelated})$$

Note that, $cov[x,y] = 0$ does not always mean x and y are independent.

Multinomial Distribution

It is the generalization of binomial distribution to the case where there m different possible outcomes.

Suppose the probability of each outcome i, for a particular trial, is p_i.

Then, $\sum_{i=1}^{m} p_i = 1$

The joint probability to observe n_1 outcomes of type 1, n_2 outcomes of type 2 etc. in N total trials is given by

$$f(n_1, ..., n_m; N, p_1, ..., p_m) = \frac{N!}{n_1!...n_m!} p_1^{n_1} ... p_m^{n_m}$$

Exercise Show that $< n_i >= N p_i$ and variance $V[n_i] = N p_i (1 - p_i)$.

Multinomial Distribution – contd.

If we have only three possible outcomes: i, j, and everything else

Then, the joint probability distribution for n_i outcomes of type i, n_j outcomes of type j, and $N - n_i - n_j$ of rest

$$f(n_i, n_j; N, p_i, p_j) = \frac{N!}{n_i! n_j! (N - n_i - n_j)!} p_i^{n_i} p_j^{n_j} (1 - p_i - p_j)^{N - n_i - n_j}$$

Exercise Show that the covariance $V_{ij} = cov[n_i, n_j]$, for $i \neq j$, is

$$V_{ij} = < (n_i - < n_i >)(n_j - < n_j >) > = -N p_i p_j$$

For $i = j$, $V_{ii} = \sigma_i^2 = N p_i (1 - p_i)$

Multivariate Gaussian distribution

Gaussian distribution for n-dimesional vector $\vec{x} = (x_1, ..., x_n)$,

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

$\vec{x}$, $\vec{\mu}$ are column vectors. The expectation values and (co)variances,

$$< x_i >= \mu_i, \quad V[x_i] = \sigma_i^2, \quad cov[x_i, x_j] = V_{ij}$$

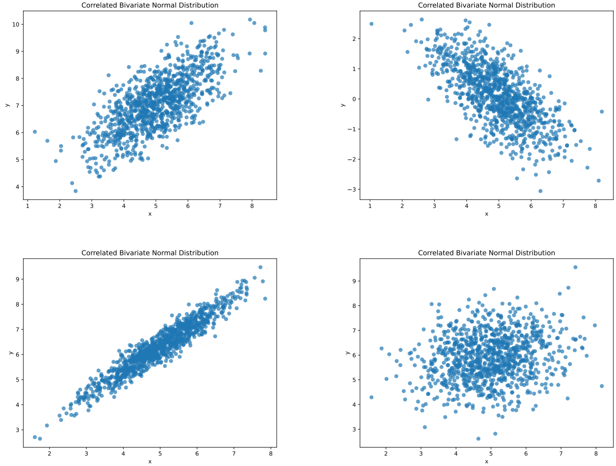
In 2d,

$$f(x_1, x_2; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \;=\; \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

$$\times\, exp\left[-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right]$$

$\rho \;=\; cov[x_1, x_2]/(\sigma_1\sigma_2)$ is the correlation coefficient.

Bivariate Gaussian distributions



(a) $\rho = 0.70$, (b) $\rho = -0.70$, (c) $\rho = 0.95$, (d) $\rho = 0.20$

Functions of a random variables

Functions of random variables are themselves random variables.
Suppose $f(x)$ is the pdf of x, and $a(x)$ is a continuous function.
What is pdf of a?

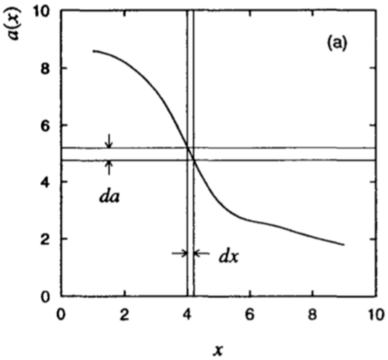
If $g(a)$ is pdf of a, then

$$g(a)da \;=\; \int_{dS} f(x)dx$$

dS is the region x-space for which a lies within $[a, a+da]$.

If $a(x)$ can be inverted to obtain $x(a)$,

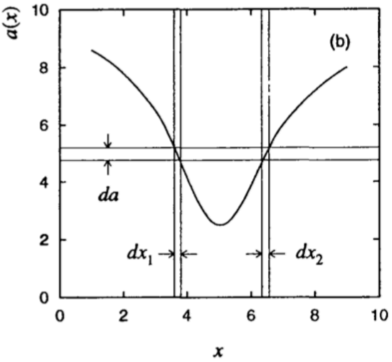
$$g(a)da \;=\; f(x)dx$$
$$\implies \;\; g(a) \;=\; f(x(a))\left|\frac{dx}{da}\right|$$



Functions of a random variables – contd.

If $a(x)$ does not have a unique inverse, include all dx intervals in dS that correspond to da.

Exercise: What is the pdf $g(a)$ for the function $a(x) = x^2/4$?

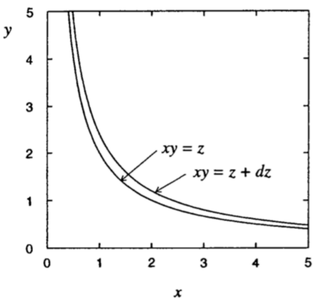


Functions of more than one random variables

If $a(\vec{x})$ is a function of n random variables $\vec{x} = (x_1, x_2, ..., x_n)$

$$g(a')da' = \int ... \int_{dS} f(x_1, x_2, ..., x_n) dx_1 ... dx_n$$

dS is the region in $\vec{x}$-space between two (hyper)surfaces:
$a(\vec{x}) = a'$, $a(\vec{x}) = a' + da'$.



Example:
Suppose x and y are independent with pdfs $g(x)$ and $h(y)$. Consider function $z = xy$. What is $f(z)$?

$$f(z)dz = \int \int_{dS} g(x)h(y)dxdy$$
$$= \int_{-\infty}^{\infty} g(x)dx \int_{z/x}^{(z+dz)/x} h(y)dy$$

$$f(z) = \int_{-\infty}^{\infty} g(x)h(z/x)\frac{dx}{x}$$

or $f(z) = \int_{-\infty}^{\infty} g(z/y)h(y)\frac{dy}{y}$

(Mellin convolution)

Functions of more than one random variables – contd.

Similarly, the pdf $f(z)$ of the sum $z = x + y$

$$\begin{aligned} f(z) &= \int_{-\infty}^{\infty} g(x)h(z-x)dx \\ &= \int_{-\infty}^{\infty} g(z-y)h(y)dy \end{aligned}$$

f is called the (Fourier convolution).

Exercise: Show that if x and y are distributed uniformly between 0 and 1, then PDF of z will be triangular between 0 and 2.

Functions of more than one random variables – contd.

For n random variables $\mathbf{x} = (x_1, ..., x_n)$, to determine the joint pdf of n linearly independent functions $a_i(bfx)$, assuming a_is can be inverted to get $x_i(a_1, ..., a_n)$

$$g(a_1, ..., a_n) = f(x_1, ..., x_n)|J|$$

where $|J|$ is the absolute value of the Jacobian determinant,

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial a_1} & \frac{\partial x_1}{\partial a_2} & \dots & \frac{\partial x_1}{\partial a_n} \\ \frac{\partial x_2}{\partial a_1} & \frac{\partial x_2}{\partial a_2} & \dots & \frac{\partial x_2}{\partial a_n} \\ \vdots & & & \vdots \\ \frac{\partial x_n}{\partial a_1} & & \dots & \frac{\partial x_n}{\partial a_n} \end{vmatrix}$$

To get the marginal pdf for one of the functions the joint pdf has to be integrated over the remaining a_i.

Error Propagation

Suppose we have a set of random variables $\vec{x} = (x_1, ..., x_n)$, distributed according to joint pdf $f(\vec{x})$.
Consider a function $y(\vec{x})$.
How to estimate the mean and variance of $y(\vec{x})$?
Ans: Use joint pdf $f(\vec{x})$ to find the pdf $g(y)$, then using this pdf compute $< y >$ and $V[y]$.

However, in practical cases, $f(\vec{x})$ may not be fully known.
But, the mean values of the x_i, $\vec{\mu} = (\mu_1, ..., \mu_n)$ and the covariance matrix $V_{ij} = cov[x_i, x_j]$ are known.
We can approximate the value of $< y >$ and $V[y]$.

Expand $y(\vec{x})$ to 1st order about $\vec{\mu}$.

$$y(\vec{x}) \approx y(\vec{\mu}) + \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i)$$

Error Propagation - 2

Since $< x_i - \mu_i > \; = \; 0$,

$$< y(\vec{x}) > \approx \; y(\vec{\mu})$$

Variance: $\sigma_y^2 \; = \; < y^2 > - < y >^2$

Where,

$$< y^2(\vec{x}) > \approx \; y^2(\vec{\mu}) + 2y(\vec{\mu}) \cdot \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i} \right]_{\vec{x}=\vec{\mu}} < x_i - \mu_i >$$

$$+ < \left(\sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i} \right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i) \right) \left(\sum_{j=1}^{n} \left[\frac{\partial y}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} (x_j - \mu_j) \right) >$$

$$= \; y^2(\vec{\mu}) \; + \; \sum_{i,j=1}^{n} \left[\frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} V_{ij}$$

$$\Longrightarrow \quad \sigma_y^2 \; \approx \; \sum_{i,j=1}^{n} \left[\frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} V_{ij}$$

Error Propagation - 3

If the x_i are uncorrelated, i.e. $V_{ij} = \delta_{ij}\sigma_i^2$, then,

$$\sigma_y^2 \approx \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]^2_{\vec{x}=\vec{\mu}} V_{ij}$$

Special cases:
If $y = x_1 + x_2$,

$$\sigma_y^2 = \sigma_1^2 + \sigma_2^2 + 2V_{12}$$

Similarly, for $y = x_1 x_2$,

$$\frac{\sigma_y^2}{y^2} = \frac{\sigma_1^2}{x_1^2} + \frac{\sigma_2^2}{x_2^2} + 2\frac{V_{12}}{x_1 x_2}$$

Exercises: Derive the expression of the variance for the following functions: (1) $y = x_1 - x_2$, (2) $y = a\frac{x_1}{x_2}$, (3) $y = ax^b$, (4) $y = ae^{bx}$ (5) $y = aln(bx)$, and (6) $y = acos(bx)$, where a and b are constants.

Error Propagation - 4

Consider a set of m functions $\vec{y}(\vec{x}) = (y_1(\vec{x}), ..., y_m(\vec{x}))$. The covariance matrix will be

$$U_{kl} = cov[y_k, y_l] \approx \sum_{i,j=1}^{n} \left[\frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} V_{ij}$$

In matrix notation, $U = AVA^T$, where

$$A_{ij} = \left[\frac{\partial y_i}{\partial x_j} \right]_{\vec{x}=\vec{\mu}}$$

And, if the x_i are uncorrelated,

$$U_{kl} \approx \sum_{i=1}^{n} \left[\frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_i} \right]_{\vec{x}=\vec{\mu}} \sigma_i^2$$

Orthogonal transformation of random variables

Let $x_1, ..., x_n$ are a set of correlated n r.v. having covariant matrix $v_{ij} = cov[x_i, x_j]$ with non-zero offdiagonal elements.

Useful to define a set of n r.v. $y_1, ..., y_n$ that are uncorrelated. i.e. $U_{ij} = cov[y_i, y_j]$ is diagonal.
It is possible with a linear transformation,

$$y_i = \sum_{j=1}^{n} A_{ij} x_j$$

The cov. matrix

$$U_{ij} = cov[y_i, y_j] = cov\left[\sum_{k=1}^{n} A_{ik} x_k \sum_{l=1}^{n} A_{jl} x_l\right]$$
$$= \sum_{k,l=1}^{n} A_{ik} A_{jl} cov[x_k, x_l] = \sum_{k,l=1}^{n} A_{ik} V_{kl} A_{lj}^T$$

This is simply a special case of error propagation formula $U = AVA^T$.

Orthogonal transformation of r.v. – contd.

Idea is to find a matrix A such that $U = AVA^T$ is diagonal.
$\longrightarrow$ Diagonalization of real, symmetric matrix

Determine the eigenvalues (λ_i) and eigenvectors ($\vec{r^i}$)
$\Longrightarrow$ λ_i are the diagonal elements of U and matrx A is constructed from n eigenvectors $\vec{r^i}$.
i.e. $A_{ij} = r^i_j$ and $A^T_{ij} = r^j_i$.

You can show that $U_{ij} = \sum_{k,l=1}^{n} A_{ik}V_{kl}A^T_{lj} = \lambda_j \delta_{ij}$.

The variances of y_i are λ_is.
Since the eigenvectors are orthonormal, $AA^T = \mathbb{I} \Rightarrow A^T = A^{-1}$.

$\Longrightarrow$ The transformation is orthogonal, i.e. corresponds to rotation of the vector $\vec{x}$ into $\vec{y}$ such that $|\vec{y}|^2 = |\vec{x}|^2$.

Orthogonal transformation of r.v. – contd.

In $2D$, the cov. matrix for the variables (x_1, x_2) can be written as

$$V = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Solving for eigenvalues,

$$\lambda_{\pm} = \frac{1}{2}\left[\sigma_1^2 + \sigma_2^2 \pm \sqrt{(\sigma_1^2 + \sigma_2^2)^2 - 4(1-\rho^2)\sigma_1^2\sigma_2^2}\right]$$

Orthonormal eigenvectors can be parametrized by an angle θ, i.e.,

$$\vec{r}_+ = \begin{pmatrix} cos\theta \\ sin\theta \end{pmatrix} \qquad \vec{r}_- = \begin{pmatrix} -sin\theta \\ cos\theta \end{pmatrix}$$

Substituting back into eigenvalue equation

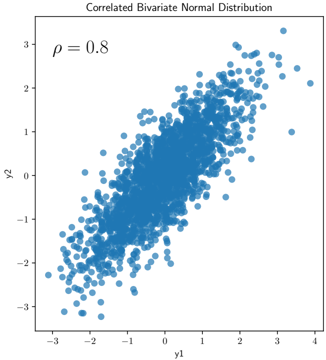
$$\theta = \frac{1}{2} tan^{-1}\left(\frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}\right)$$

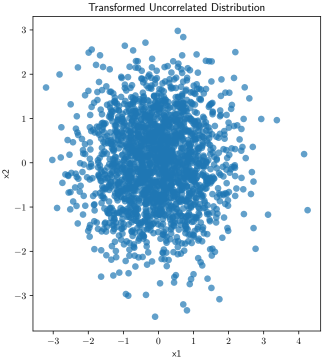
Orthogonal transformation of r.v. – contd.

The desired transformation matrix is

$$A = \begin{pmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{pmatrix}$$

Corresponds to a rotation of the vector (x_1, x_2) by an angle θ.





END