Statistics and Error propagation

Christian Bohm - Stockholm University

Measurement statistics

To use a measurement result one must know about its reliability and precision

Most measurements are affected by many random processes and are only fully characterized by their probability distribution

In statistical terms this is a **stochastic variable**

The probability distribution function can be determined from knowledge of the random processes involved or determined experimentally by performing a large number of measurements

A stochastic variable x can assume different values with the probability density function $f(x)$ and x is therefore completely defined by f

$y = 2x$ has a probability density as well and is thus also a stochastic variable, now with the probability distribution $f(x/2)$

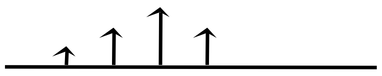
Distribution functions

Probability distribution function (PDF) -> **Complete information** about all statistic properties of the random variable

Main classification **discrete - continuous** distributions

which distribution function - depend on the measuring process

Other names: density function or frequency function

$f(x)$  

$f(x) \geq 0$ — **Probabilities are always positive**

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

The probability for any value is 1

The measurement result is completely characterized by its PDF

If it is not possible to **identify the pdf** of the result - one should **characterize** it as well as possible

The most important parameter is **position**, then **width**, **skewness**, etc.
(these parameters can be determined with good precision from a smaller amount of data)

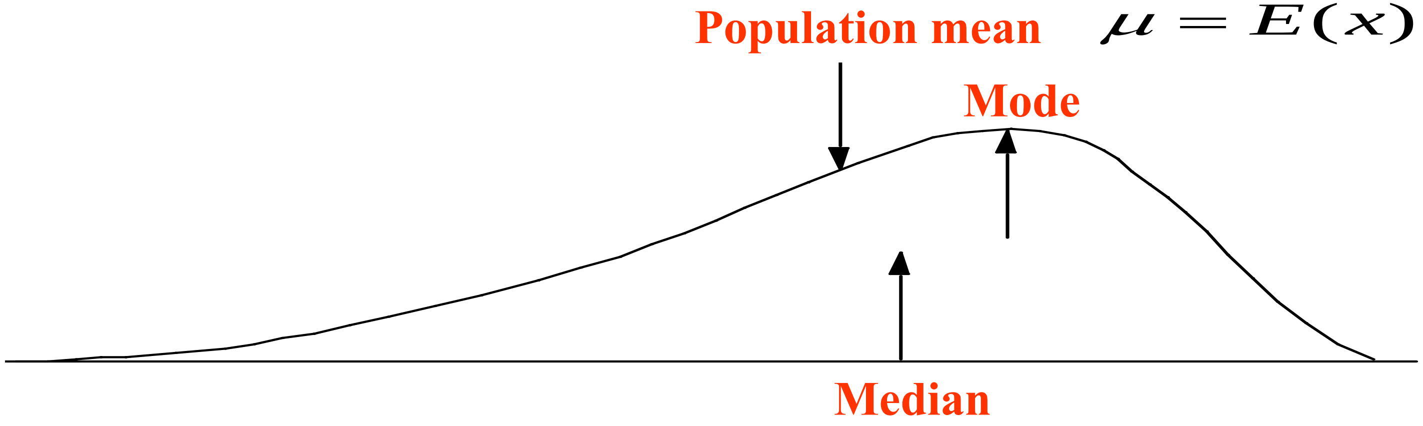
Position measures

The **expectation value** of x

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

f's 1:st moment (center of gravity)

$$\mu = E(x)$$



$$\int_{-\infty}^{m} f(x) dx = 0.5$$

Choice of parameter depend on the type of measurement

Mean most common

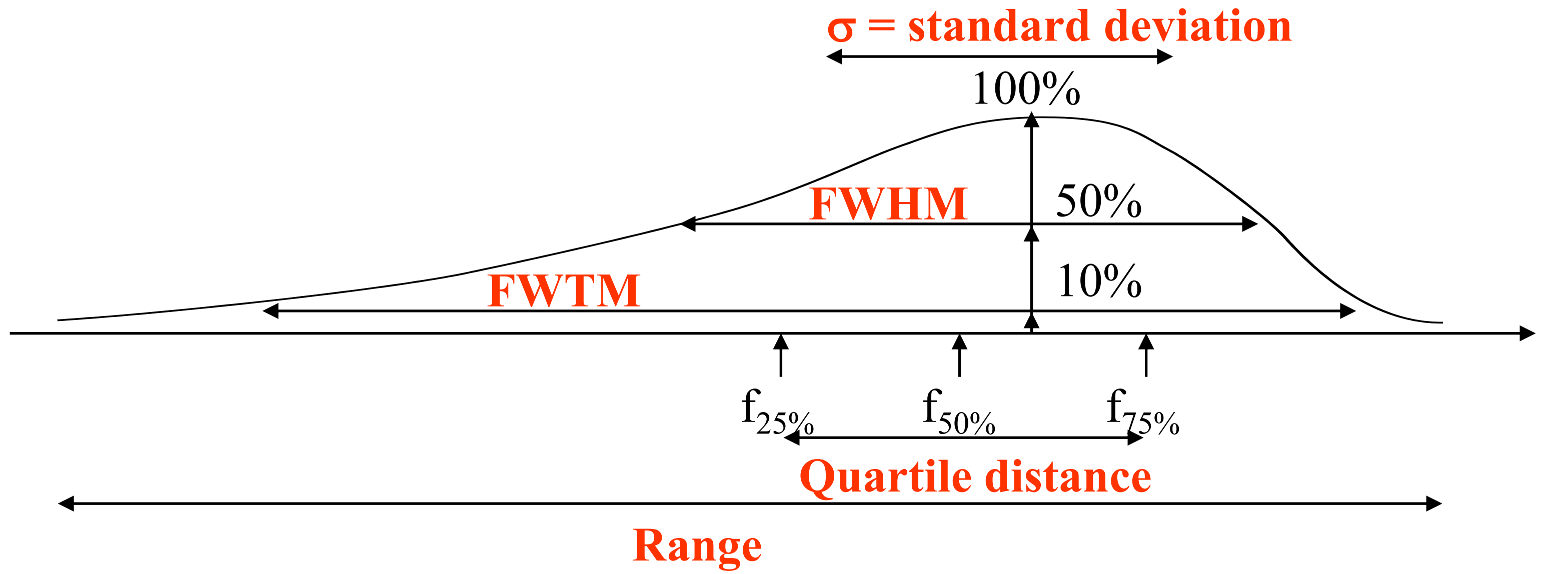
Width measures

Population variance

$$Var(x) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = E((x-\mu)^2) = E(x^2) - 2\mu E(x) + \mu^2 =$$
$$= E(x^2) - \mu^2$$

f's 2:nd central moment

f's 2:nd moment

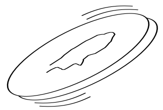


Choice of parameter depend on the type of measurement

Standard deviation and **F**ull **W**idth **H**alf **M**aximum (**FWHM**) most common

For a normal distribution FWHM=2.355σ

Discrete Distributions
Binomial distribution

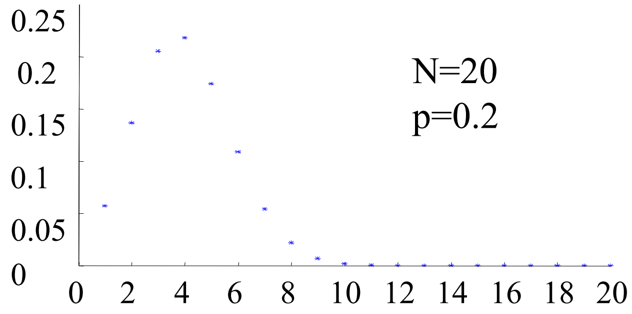


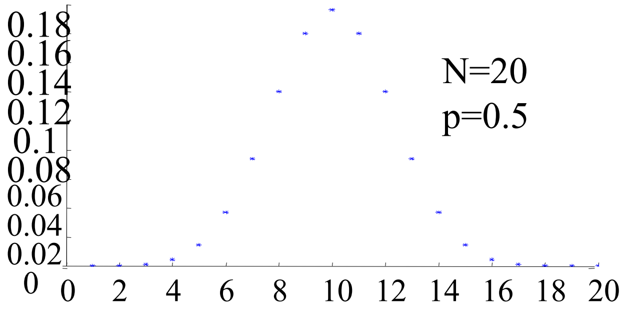
Repeating independent elementary binary events (succeed – fail)
each with the probability p

E.g.

Tossing coins	elementary event – coin toss
Drawing tickets with replacement	elementary event – draw
Radioactive decay	elementary event – decay of a nucleus
Monte Carlo simulations	elementary event – one case

Parameters	$0 \le p \le 1$ probability N>0 number of trails
Variable	r
Probability distribution	$p(r) = \binom{N}{r} p^r (1-p)^{N-r}$
Mean	$E(r) = Np$
Variance	$V(r) = Np(1-p)$

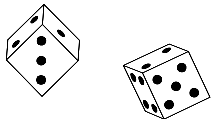




The multinomial distribution

Repeated independent elementary events with many (k) outcomes
each with the probability p_i where $1 \leq i \leq k$

e.g.
Throwing dices
Monte Carlo simulations with several outcomes
Histograms



Parameters $0 \leq p_i \leq 1$ probability
k, the number of outcomes
N number of trails

Variable r_i

Probability distribution $p(r_1, r_2,, r_k) = \frac{N!}{r_1! r_2! \cdot\cdot r_k!} p^{r_1} p^{r_2} \cdot\cdot p^{r_k}$

Mean $E(r_i) = Np_i$

Variance $V(r_i) = Np_i(1 - p_i)$

Covariance $Cov(r_i, r_j)) = -Np_i p_j$

2 dices

Probability for one 5 and one 2

$$p(2,5) = \frac{2!}{0! \cdot 1! \cdot 0! \cdot 0! \cdot 1! \cdot 0!} \cdot \left(\frac{1}{6}\right)^0 \cdot \left(\frac{1}{6}\right)^1 \cdot \left(\frac{1}{6}\right)^0 \cdot \left(\frac{1}{6}\right)^0 \cdot \left(\frac{1}{6}\right)^1 \cdot \left(\frac{1}{6}\right)^0 = \frac{2}{36}$$

The Poisson distribution

The **probability for a certain number of events during a time period** if the **probability per time unit** for such a event is **constant** (l) and **independent of what happened before**. One can say that the process have **no memory**

E.g. Telephone switchboard load

Parameter $0 < \lambda$, events/time unit

Variabel $r \geq 0$, the number of events

Probability distribution $$p(r) = \frac{\lambda^r e^{-\lambda}}{r!}$$

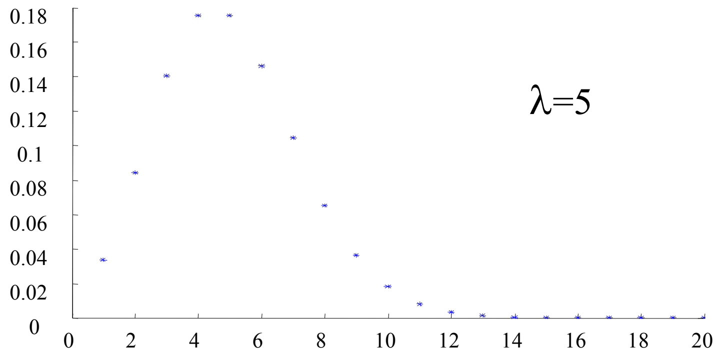
Mean $$E(r) = \lambda$$

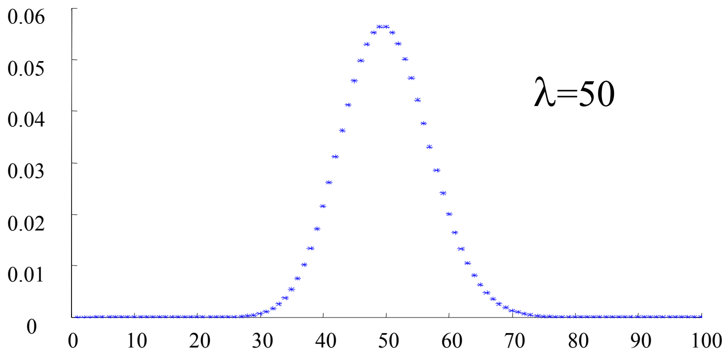
Variance $$V(r) = \lambda$$

Binomial distribution --> Poisson distribution with $N \to \infty$ $p \to 0$ Np=const $$\lambda = Np$$

Radioactive decays (approx. Poisson)

Histograms with many events (approx Poisson)



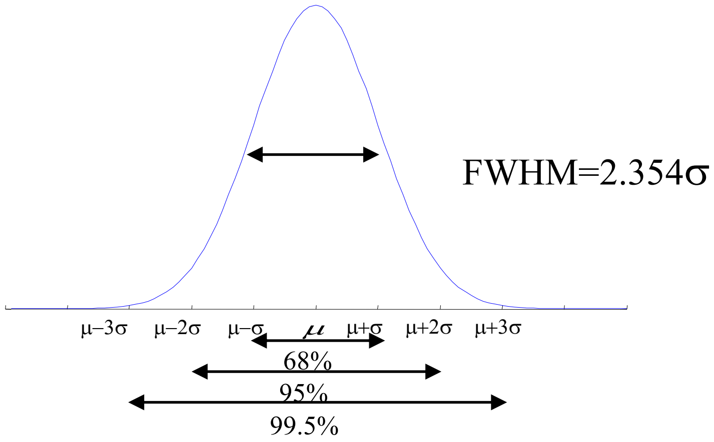


Poisson distribution with n>50 looks like a normal distribution

Normal distribution

Variable	x, real number
Parameter	σ,μ
Probability distribution	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$
Mean	μ
Standard deviation	σ

N(μ,σ^2) denotes a normal distributed parameter with mean μ and standard deviation σ



Also called **Gauss** distribution

The law of the large numbers

According to the law of large numbers

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} X_n = \lim_{n \to \infty} \overline{X} = \mu$$

The sample mean will approach the true value as the size of the sample increases:

More generally, one can say:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} g(X_n) = \int_{-\infty}^{\infty} g(x) \cdot f(x) \cdot dx = g(\mu)$$

When applied to the variance this implies:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} (X_n - \mu)^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) \cdot dx = \sigma^2$$

Statistics

A **statistic** is a function of stochastic variables

$T_N = f(X_1, X_2, , X_N)$ is a **statistic**

The calculation $\{X\}$ --> T_N implies a **data reduction**

Estimators

Let us use the statistics T_N to **estimate** the physical parameter θ

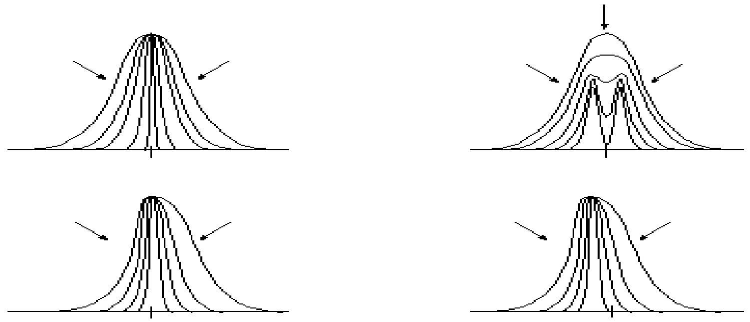
T_N is the called an **estimator**

An infinitely large sample should give the true value

If $\lim_{N \to \infty} T_N = \theta$ then T_N is **consistent**

The mean of a large number of small sample estimators should give the true value

If $E(T_N) = \theta$ for all N then T_N **unbiased**



Unbiased

biased

consistent **inconsistent**

If T_N uses the information well it is **effective**

If T_N is not sensitive to small variations in the distribution then T_N is **robust**

One can say that lack of consistency correspond to systematical errors

And lack of efficiency correspond to statistical errors

Samples

If you have a sample with N measured values x_i then The sample mean is

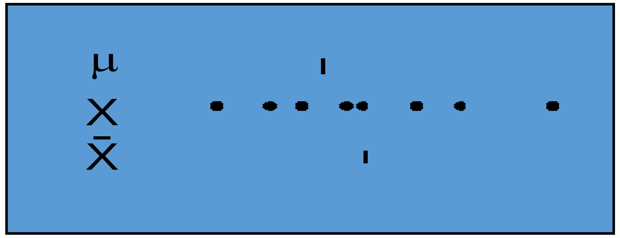
$$\bar{x} = \frac{1}{N}\sum_i x_i$$

It is a consistent estimator of the population mean μ (the law of large numbers)

One also easily show that it is unbiased, since the mean of many small samples is the same as the mean of one large sample

$$s^2 = \frac{1}{N}\sum_i (x_i - \bar{x})^2 \quad \text{and} \quad s^2 = \frac{1}{N-1}\sum_i (x_i - \bar{x})^2$$

are both consistent estimators of σ^2 but only the right one is unbiased, but why N-1?



$\bar{x}$ is more central in the sample than μ thus

$$\sum_i (x_i - \bar{x})^2 \le \sum_i (x_i - \mu)^2$$

N-1 compensates for the under estimation

Estimator examples

If we know that **r** is binomially distributed then **r/N** is a consistent estimator of **μ** or **p** (according to the law of large numbers):

$$\hat{p} = r / N$$

If we know that **r** is Poisson distributed then **n** is a consistent estimator of **λ** (according to the law of large numbers):

$$\hat{n} = \lambda$$

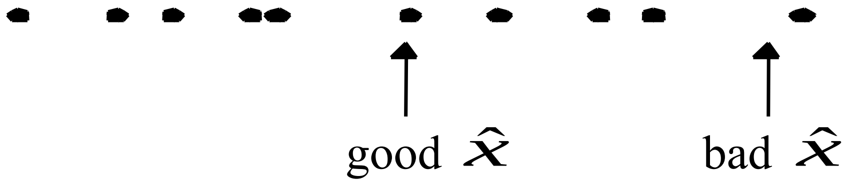
Since small samples also have the mean **λ** it is also unbiased

The Poisson distribution also implies that variance can be estimated by $\hat{n}$ and

$$\sigma = \sqrt{\lambda} \approx \sqrt{n}$$

Simple estimators

Find a representative value (estimator) for a physical parameter which corresponds to X



In order to find which estimator gives the most representative value you need a figure of merit to minimize

E.g. you can minimize $$\sum_i (X_i - \hat{x})^2$$

giving $$\hat{x} = \frac{1}{N}\sum_i X_i$$

If X_i has different variances σ_i^2 you can instead use $$\sum_i \frac{(X_i - \hat{x})^2}{\sigma_i^2}$$

Minimizing --> $$\hat{x} = \frac{\sum x_i / \sigma_i^2}{\sum 1/\sigma_i^2}$$

The Likelihood function

$L(X|\theta)=P(X|\theta)$ is called the likelihood function

which expresses the probability to get the result X if the parameter is θ

$L(X_1X_2X_3|\theta) = L(X_1|\theta)L(X_2|\theta)L(X_3|\theta)$ if X_1, X_2 and X_3 are independent

In the **maximum likelihood (ML) method** you choose the θ that gives maximum L

or, which is the same, maximum lnL.

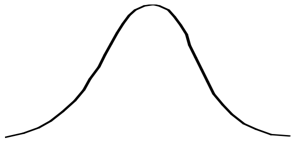
If the X are normally distributed ML is identical to LSM (the least square method)

Information

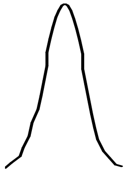
The precision in the ML-determination is better with a more narrow maximum .

Narrow maximum (small variance) --> more information about θ

More observations (smaller variance) --> narrower maximum



approximate information about position



better information information about position

You can define **information** (according to Fischer) as

$$I = -E\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right)$$ evaluated where L is maximal

The information is then additive

$I(X_1X_2X_3)=I(X_1)+I(X_2)+I(X_3)$

if X_1,X_2 and X_3 are independent

Covariances and correlations

If we have two random variables then as x varies around μ_x, y will vary around μ_y
The covariance will tell us if these variations are connected:

$$\sigma_{XY} = \mathrm{cov}(X,Y) = \sum_i (X_i - \mu_X)(Y_i - \mu_Y) f(X,Y)$$

or for continuous variables:

$$\mathrm{cov}(X,Y) = E((X-\mu_X)(Y-\mu_Y)) = \int_{-\infty}^{\infty} (X-\mu_X)(Y-\mu_Y) f(X,Y)\,dX\,dY$$

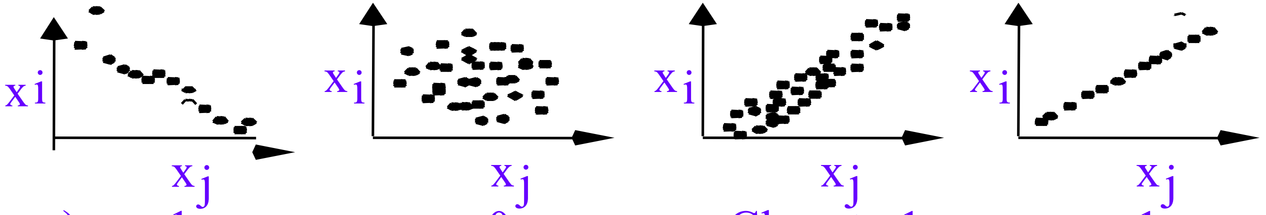
The covariance matrix is defined as:: $V = \begin{pmatrix} \mathrm{var}(x) & \mathrm{cov}(x,y) \\ \mathrm{cov}(x,y) & \mathrm{var}(y) \end{pmatrix}$

cov(x,y) = cov(y,x) → V is symmetric

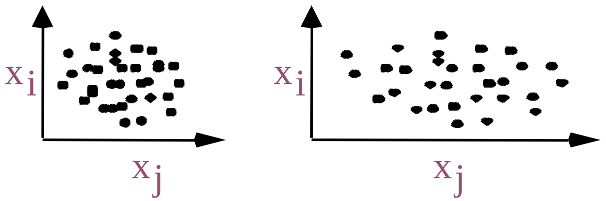
The magnitude of the normalized correlation coefficient is defined as:

$$corr(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$

is always less or equal to 1:



$corr(x_i,x_j)$	-1	0	Close to 1	1
V	Non-diagonal	Diagonal	Non-diagonal	Non-diagonal



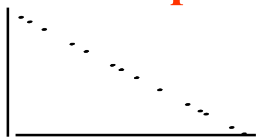
corr(x_i,x_j): 0 0

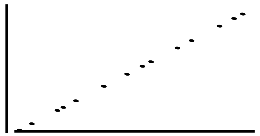
Whether the pattern is "circular" or "elliptic" along the coordinate axes, does not affect the correlation

Since the "ellipticity" can be removed by re-scaling

But it is important to realize that:

Independent $\underset{\not\leftarrow}{\rightarrow}$ **Uncorrelated**





corr(x,y) = -1 for sample 1 corr(x,y) = +1 for sample 2

thus:

Here corr (x,y)= 0 for the combined sample 1 + 2 but x and y are definitely not independent

Addition of two stochastic variables

$$Var(x+y)=\sigma(x+y)^2=\int_{-\infty}^{\infty}(x+y-\mu_x-\mu_x)^2 f(x)g(y)dxdy=$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}((x-\mu_x)^2+(y-\mu_y)^2+2(x-\mu_x)(y-\mu_y))f(x)g(y)dxdy=$$

$$=\int_{-\infty}^{\infty}(x-\mu_x)^2 f(x)dx+\int_{-\infty}^{\infty}(y-\mu_y)^2 g(y)dy-2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}((x-\mu_x)(y-\mu_y))f(x)g(y)dxdy=$$

$$=Var(x)+Var(y)-2\operatorname{cov}(x,y)$$

If you combine two measurements negatively covariance helps

If x and y are uncorrelated → $\sigma(x+y)=\sqrt{\sigma^2{}_x+\sigma^2{}_y}=\sqrt{Var(x)+Var(y)}$

If you subtract two random variables you get the same formula. X can be signal and y background

If the signal plus background is 25 and the background 16 the error in N-B=9 is about 6

One can easily show that: $\sigma(ax)=a\sqrt{\sigma^2{}_x}=a\sqrt{Var(x)}$

and more general after linearizing: $\sigma(f(\mathbf{x}))=a\sqrt{\left(\frac{\partial f}{x_1}\right)^2\sigma^2{}_{x_1}+\sqrt{\left(\frac{\partial f}{x_2}\right)^2\sigma^2{}_{x_2}+\ldots+\sqrt{\left(\frac{\partial f}{x_n}\right)^2\sigma^2{}_{x_n}}}}$

This is called the error propagation formula

Negative correlation

Estimating the DC bias of an AC signal by random sampling require many samples to get a precise result using averaging.

If you realize that voltages are pairwise negatively correlated if the time interval is close to half the period.

If the interval is exactly half the period the correlation is exactly -1. The variance is then:

$$\sigma^2 + \sigma^2 - 2\sigma^2 = 0$$

Since:

$$corr = \frac{cov}{\sigma\sigma}; cov = corr\,\sigma\sigma = -\sigma^2$$

The average of two sample points with half a periods distance is exactly base line.

Multidimensional probability distributions
The multivariate distribution

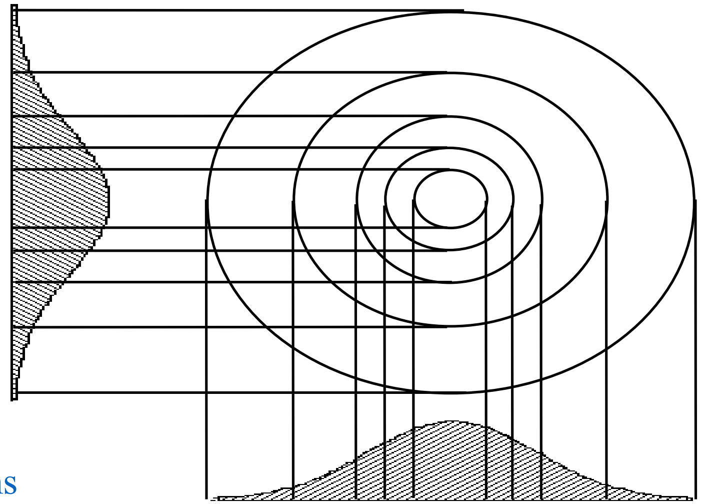
If x_1 and x_2 are independent and normally distributed, the compound 2-d distribution is given by:

$$\frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}\cdot\frac{1}{\sigma_2\sqrt{2\pi}}e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}}=\frac{1}{\sigma_1\sigma_2 2\pi}e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}$$

This expression can be given in matrix form

$$\frac{1}{(2\pi)^{k/2}|\mathbf{V}|^{1/2}}e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{V}^{-1}(\mathbf{x}-\mu)}$$

where the covariance matrix $\mathbf{V}=\begin{pmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{pmatrix}$ is diagonal



Marginal distributions

Normality in several dimensions

When measuring independent normal distributed parameters in connection with events 67% are within one standard deviation from the mean and 95% Within 2 standard deviations.

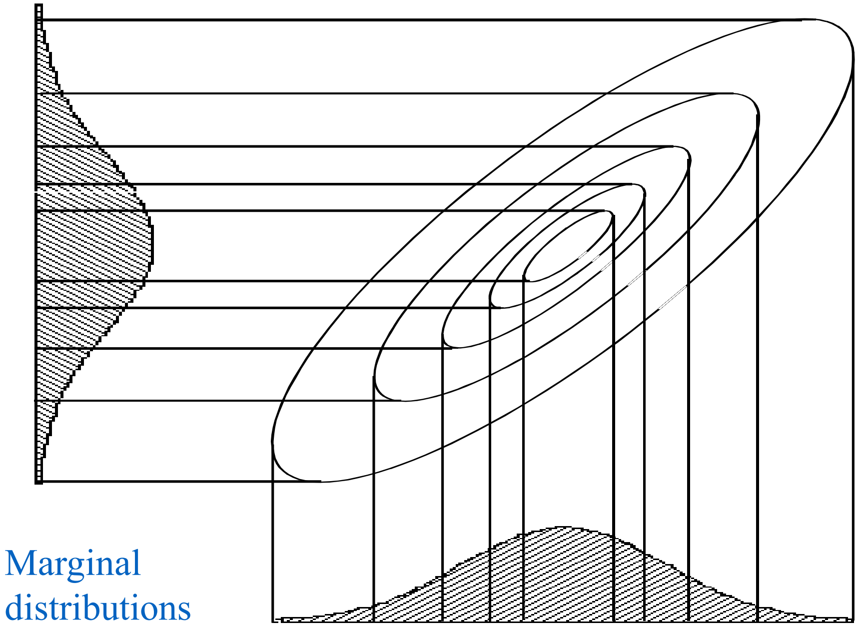
The probability of 10 independent parameters each being within one standard deviation from the mean in is $0.67^{10} = 1.8\%$. The corresponding probability for being within two standard deviations is $0.95^{10} = 60\%$.

Thus when considering several parameters in connection with an event it is probable that some parameters are far from the mean.

Multivariate distributions

$$\frac{1}{(2\pi)^{k/2}|\mathbf{V}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{V}^{-1}(\mathbf{x}-\mu)}$$

If V is not diagonal then x_1 and x_2 are correlated $\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$



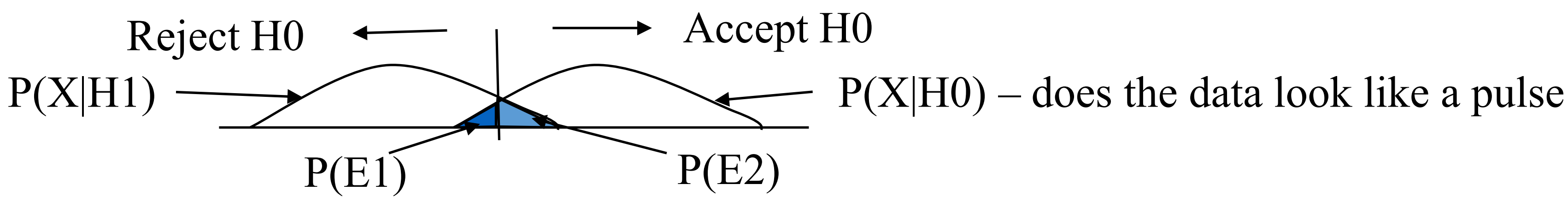
Marginal distributions

The marginal distributions do not tell the whole story

Tests of hypotheses

H0 null-hypothesis- the hypothesis you want to test - e.g. there is a pulse
H1 an alternative hypothesis – there was no pulse

Error of the first kind (E1): Erroneous rejection of the null-hypothesis – the pulse was lost, inefficiency
Error of the second kind (E2): Erroneous rejection of the alternate hypothesis – noise



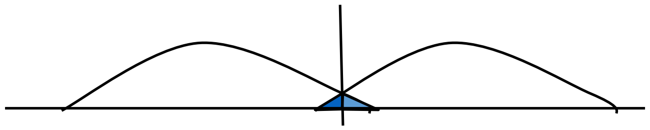
Choose a limit so that P(E2) becomes sufficiently small – below a significance level 5% is common.

In particle physics you demand 5σ for a discovery of a new particle (this corresponds to P(E1) = 0.00003%).

If P(E2) becomes too large improve the data (improve the measurements)

Find a cost function which includes the probabilities and the cost caused by errors

Choose the hypothesis that minimizes the cost function



Why we need to record many events

To determine if our **N** new observed events constitute a discovery we must determine if the same data could be produced by combinations of well-known events. The probability for is the background **B**.

For **N** to be a discovery **N** must be significantly larger than **B**

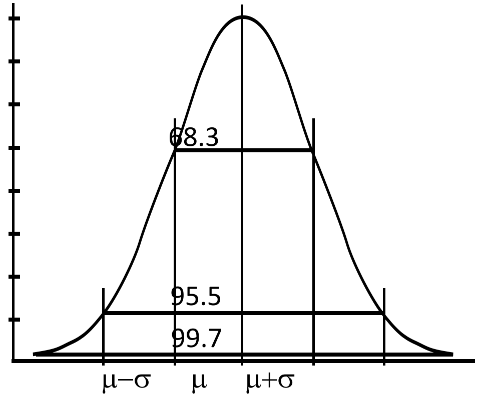
For example if **N** is 80 and **B** is 64 then σ(**B**) is 8 (assume Poisson distribution σ^2=N)

N is 2σ above i.e. 2% probability that N is just random noise

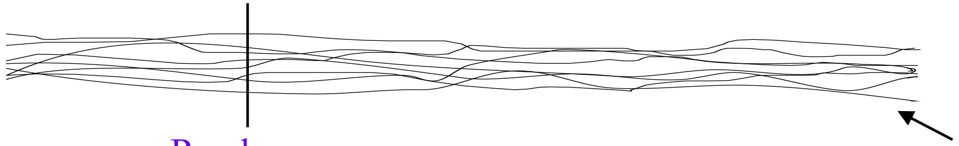
If we measure twice as long **N** will be 320, **B** is 256 and σ(B) is 16 i.e. about 4σ above (0.004% that it is random noise). Much smaller probability that **N** is due to random noise but not enough.

5σ (0.00002% it is random noise) is required for discovery.

Normal distribution
Almost the same
as Poisson if N>50



Stochastic processes



A stochastic process is a family (**ensemble**) of functions

$$\mathbf{x}(t,\varsigma)$$

depends on time t and the outcome of the experiment ζ (family member)

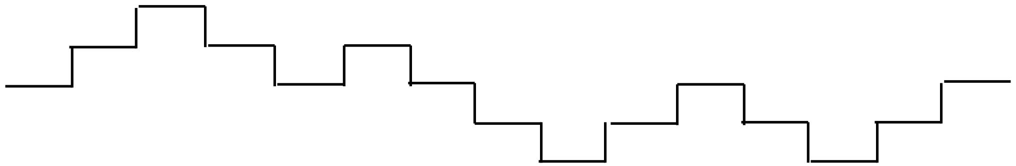
for each t, $\mathbf{x}(t,\zeta)$ is a stochastic variable and
for each z, $\mathbf{x}(t,\zeta)$ is an ordinary time function

It is thus a time dependent stochastic variable whose values are described by a multi-ordered probability distribution function:

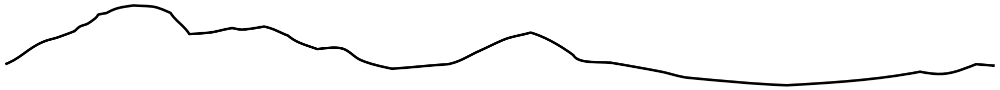
$$f(x_1, t_1, x_2, t_2, \ldots.)$$

Examples of stochastic processes

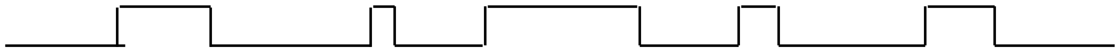
Random walk



Weiner process



Telegraph process



Poisson process

.

Noise can be expressed as a wiener process

Correlation in stochastic processes

In stochastic processes it is possible to calculate the correlation between the stochastic process at different times. This is called **autocorrelation**.

If the autocorrelation is localized measurement separated with an interval larger than the **width** of the autocorrelation function, these values are **uncorrelated**. If the autocorrelation function is a delta infinitely close data are uncorrelated (clearly unphysical). This is the case of white noise (also unphysical).

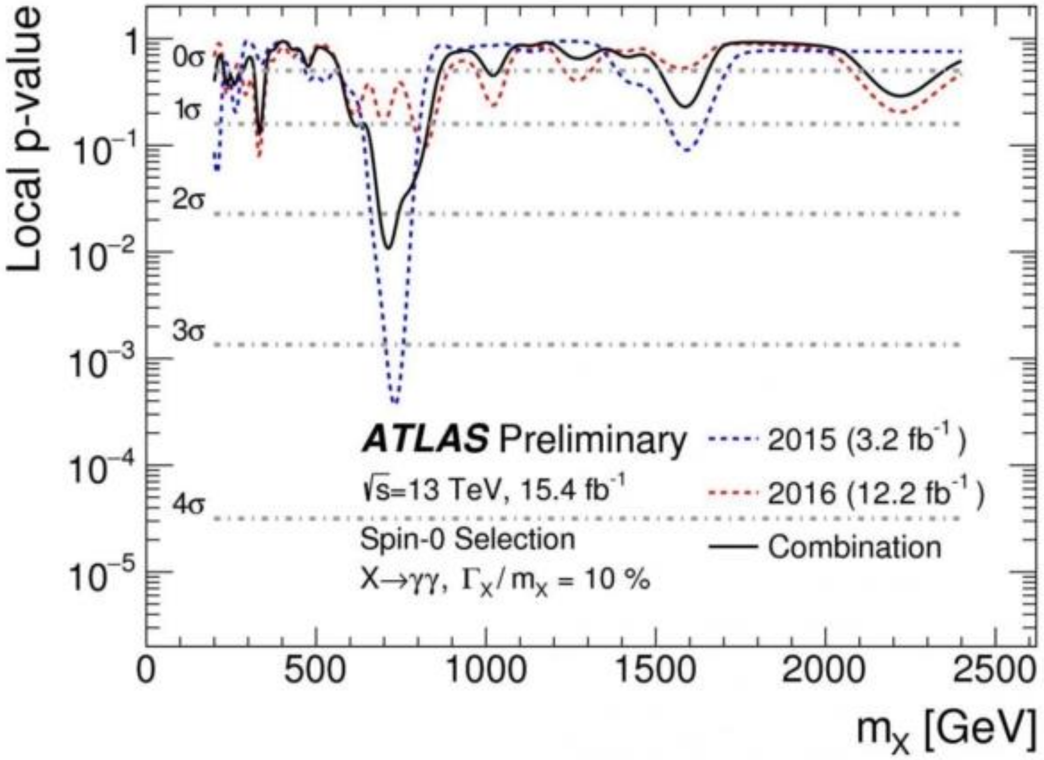
If you sample a stochastic process so that the samples are uncorrelated but normal every third sample is more than one standard deviation away from the mean. 5σ is a good criterion if you look at one measurement. If you have many measurements this reasoning is **not valid anymore**.

If you have a digital transmission you need a **Bit Error Rate** (BER), i.e. the probability that noise would corrupt one bit, of the order of or better that 10^{-16}. With 5σ for each sample you would find 2 pulses/second if you sample with 40 MHz.

This argument can be applied to the situation where you look for a pulse in noise or a peak in a noisy spectrum. This is sometimes called the **“Look elsewhere effect”**
If a peak could happen in any of n bins you need to improve the 5σ margin with the factor n.

Example of interpretation of uncertain experimental results

The **750 GeV diphoton excess** reported by ATLAS and CMS in 2015 disappeared in 2016 data, in the meantime about 500 theoretical studies were made to explain the early results. It never reached the 5σ level but showed promise. There was also a hope to find something new after the Higgs (see Wikipedia for more information).



From https://physicsworld.com > and-so-to-bed-for-the-750-gev-bump

Literature

My favorite statistic book:

Statistical methods in experimental physics
By Frederic James

This book contains everything that is necessary to know in experimental statistic,
But it is rather extensive and takes time to read if you want read it thoroughly.
If you don't intend to spend much time on the project there are many other good books on statistics.