TRAINING COURSE ON RADIATION DOSIMETRY:

Statistical analysis and data handling

Marco CARESANA, *POLIMI*

Thu. 22/11/2012, 16:30 – 18:30 pm

What are statistics

Statistics are like a drunk with a lamppost: used more for support than illumination.

Winston Churchill *British politician*

Statistics are like bikinis. What they reveal is suggestive, but what they conceal is vital.

Aaron Levenstein *Professor emeritus at Baruch College*

Characterization of data

Let us consider a series of independent measurements

$(x_{1,}\ x_{2,}\ x_{3\ldots\ldots\ldots}\ x_N)$

Two elementary properties are:

Sum $\quad S = \sum_{k=0}^{N} x_k$

Experimental mean $\quad \bar{x}_e = \frac{S}{N}$

Characterization of data

A convenient representation is in terms of
frequency distribution function F(x)

$$F(x) = \frac{number\ of\ occurences\ of\ the\ value\ x}{number\ of\ measurement\ N}$$

The distribution is automatically normalized

$$\sum_{X=0}^{\infty} F(x) = 1$$

Characterization of data

It may be that x_i are all different. In this case

$$F(x) = \frac{number\ of\ the\ values\ x\ within\ a\ bin\ \Delta x}{number\ of\ measurement\ N}$$

The distribution is automatically normalized

$$\sum_{X=0}^{n\ of\ bins} F(x) = 1$$

Characterization of data

The frequency distribution function allows the calculation of the mean value as follows

$$\bar{x}_e = \sum_{X=0}^{\infty} x \cdot F(x)$$

It remains to evaluate the spread of the experimental data. This is possible by introducing the sample variance.
As a first step let us define the residual of any data point:

$$d_i = x_i - \bar{x}_e \quad \text{and } \epsilon_i = x_i - \bar{x}$$

Characterization of data

Because d_i and ϵ_i can assume positive and negative values it is easy to understand that

$$\bar{d} \quad = \sum_{X=0}^{\infty} d_i = \bar{\epsilon} \quad = \sum_{X=0}^{\infty} \epsilon_i = 0$$

It is better to use the square of the residual

$$d_i^2 = (x_i - \overline{x_e})^2 \qquad \epsilon_i^2 = (x_i - \bar{x})^2$$

The variance is the mean value of ϵ_i^2

$$\overline{\epsilon^2} = \frac{1}{N} \sum_{X=0}^{\infty} \epsilon_i^2$$

Characterization of data

$$\overline{\epsilon^2} = \frac{1}{N}\sum_{X=0}^{\infty}(x_i - \bar{x})^2$$

This definition of variance involves the mean true value $\bar{x}$ that, in practical cases, is unknown.

The best estimate s^2 of $\overline{\epsilon^2}$ can be obtained replacing $\bar{x}$ with $\overline{x_e}$.

$$s^2 = \frac{1}{N-1}\sum_{X=0}^{\infty}(x_i - \overline{x_e})^2$$

The division by *N-1* accounts for the dependence of $\overline{x_e}$ in the experimental data set.

Characterization of data

Considering the frequency distribution function it can be written:

$$s^2 = \sum_{X=0}^{\infty} (x_i - \bar{x})^2 \cdot F(x)$$

The variance is a useful indicator of the degree of internal scattering of experimental data.

Statistical model

The frequency distribution function is an «a posteriori» distribution assessed experimentally.
A model of distribution can be derived from "a priori" information about the statistical quantity.
Let us consider a binary process in that only two results are possible, success or failure.

For instance
Toss a coin (success=head, p=1/2)
Roll a die (success=a six, p=1/6)
Observe a radioactive nucleus for a time t (success=decays, p=$1 - e^{-\lambda \cdot t} \approx \lambda \cdot t$)

Binomial distribution

The question to address is:
Let us consider an honest die and define: success=a six.
What is the probability to obtain *x* successes after *n* trials (i.e. *n* rolls)

$$P(x) = \underbrace{p \cdot p \cdot p \ldots \ldots . . \cdot p}_{x} \cdot \underbrace{(1-p) \cdot (1-p) \ldots \ldots (1-p)}_{n-x} = p^x \cdot (1-p)^{n-x}$$

This is the probability of *x* consecutive successes and *n*-*x* consecutive failures

$$P(x) = \frac{n!}{(n-x)! \cdot x!} \cdot p^x \cdot (1-p)^{n-x}$$

Binomial distribution

Let us calculate mean value and variance for the binomial distribution

$$\bar{x} = \sum_{k=0}^{n} x \cdot P(x) = n \cdot p$$

$$\sigma^2 = \sum_{k=0}^{n} (x - \bar{x})^2 \cdot P(x) = n \cdot p \cdot (1 - p)$$

Binomial distribution

Toss a coin $\bar{x} = n \cdot p = n \cdot 1/2$ $\sigma^2 = n \cdot p \cdot (1 - p) = n \cdot 1/4$

Roll a die $\bar{x} = n \cdot p = n \cdot 1/6$ $\sigma^2 = n \cdot p \cdot (1 - p) = n \cdot 5/36$

Observe a radioactive nucleus for a time t (and assuming n constant)
(success=decays, p=$1 - e^{-\lambda \cdot t} \approx \lambda \cdot t$)

$\bar{x} = \sum_{k=0}^{n} x \cdot P(x) = n \cdot p = n \cdot \lambda \cdot t$ ⇒ $A = \frac{\bar{x}}{t} = n \cdot \lambda$

$$\sigma^2 = \sum_{k=0}^{n} (x - \bar{x})^2 \cdot P(x) = n \cdot p \cdot (1 - p) = n \cdot \lambda \cdot t \cdot (1 - \lambda \cdot t)$$

Poisson distribution

$$P(x) = \frac{n!}{(n-x)! \cdot x!} \cdot p^x \cdot (1-p)^{n-x}$$

p<<1 $\lambda \cdot t \ll 1$

The observation time much lower than the decay time

$$P(x) = \frac{(p \cdot n)^x \cdot e^{-p \cdot n}}{x!}$$

Poisson distribution

$$P(x) = \frac{n!}{(n-x)! \cdot x!} \cdot p^x \cdot (1-p)^{n-x}$$

p<<1 $\lambda \cdot t \ll 1$

The observation time much lower than the decay time

$$P(x) = \frac{(p \cdot n)^x \cdot e^{-p \cdot n}}{x!} = \frac{(\bar{x})^x \cdot e^{-\bar{x}}}{x!}$$

Poisson distribution

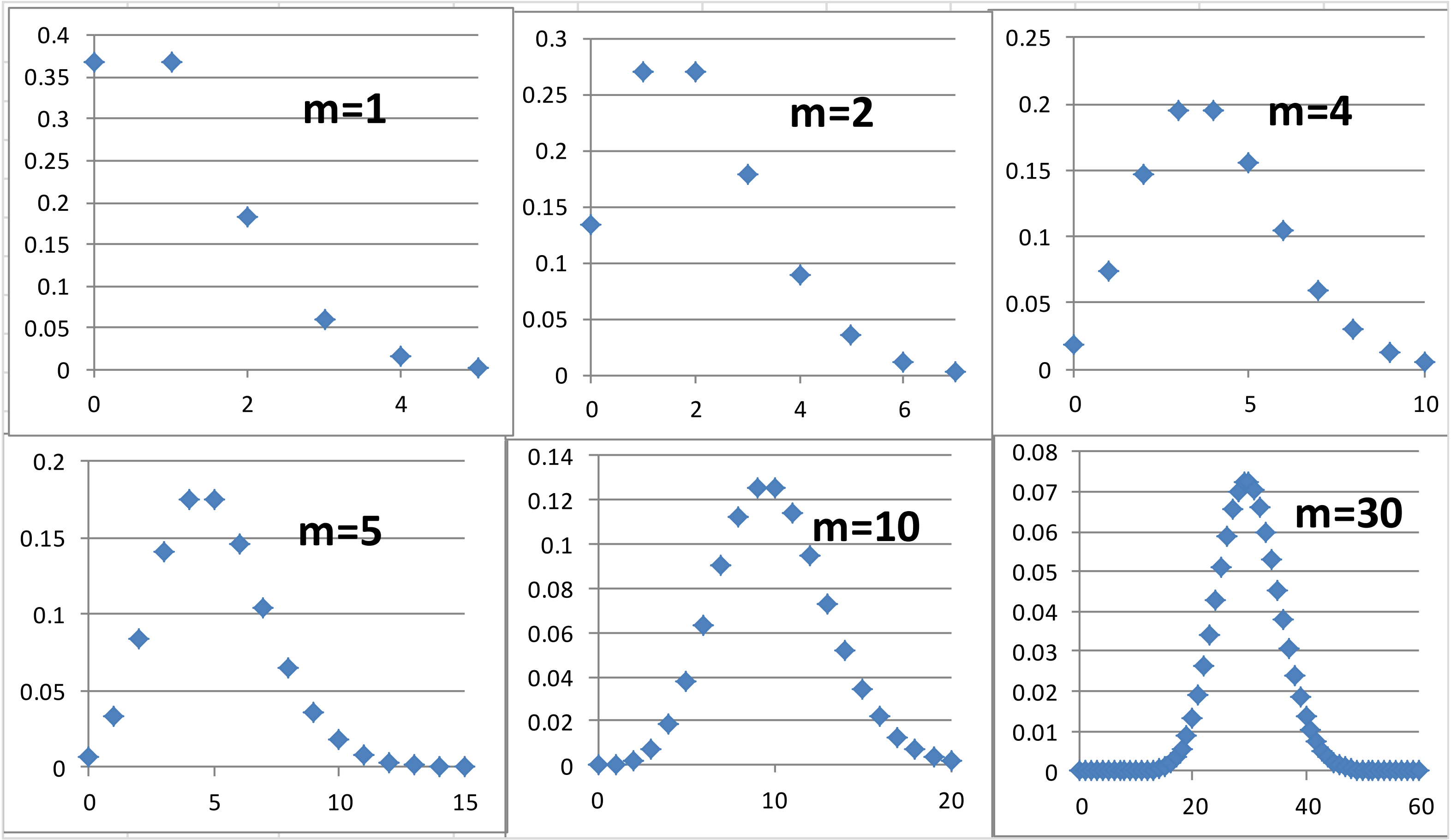
$$P(x) = \frac{(p \cdot n)^x \cdot e^{-p \cdot n}}{x!} = \frac{(\bar{x})^x \cdot e^{-\bar{x}}}{x!}$$

$$\bar{x} = \sum_{k=0}^{n} x \cdot P(x) = n \cdot p$$

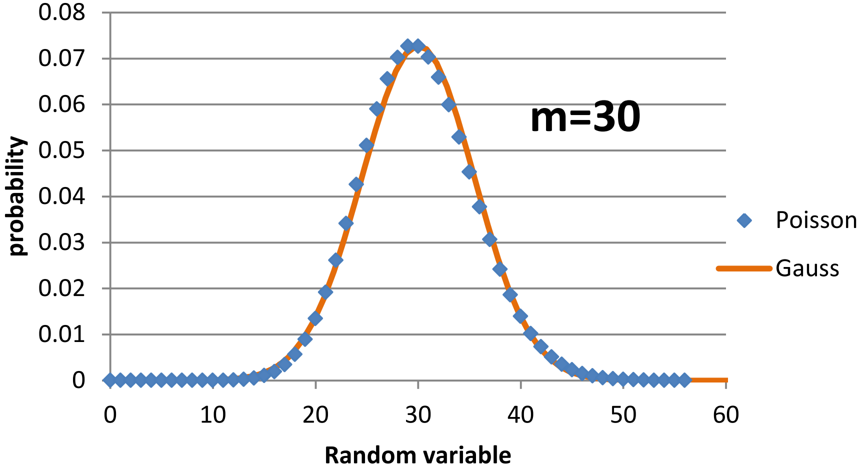
$$\sigma^2 = \sum_{k=0}^{n} (x - \bar{x})^2 \cdot P(x) = n \cdot p$$

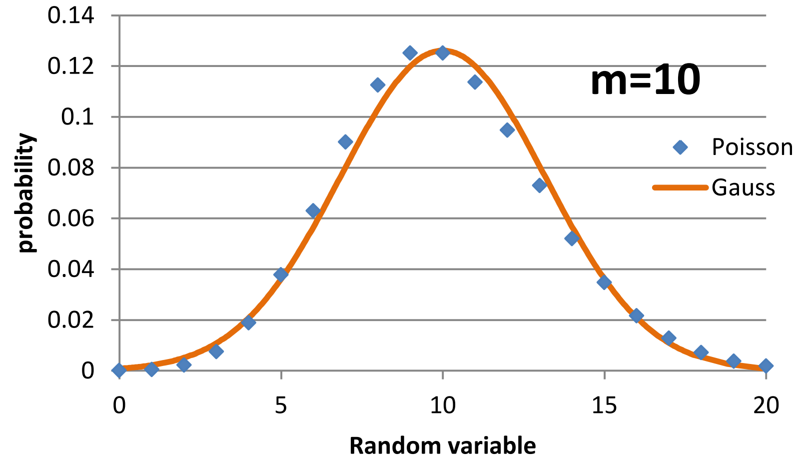
$$\sigma^2 = \bar{x}$$

Poisson distribution



Gauss distribution





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

$\sigma^2 = \bar{x}$

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

Discrete/continuous distributions

Discrete (Poisson)

$$\sum_{x=x1}^{x2} P(x) =$$ Probability of observing a value of x in the range $x_1 - x_2$

$P(x) =$ probability

Continuous (Gauss)

$$\int_{x1}^{x2} P(x)dx =$$ Probability of observing a value of x in the range $x_1 - x_2$

$P(x) =$ probability density

Gaussian confidence intervals

$$Prob(x_- \leq x \leq x_+) = \int_{x_-}^{x_+} p(x)\,dx = C$$

We say:

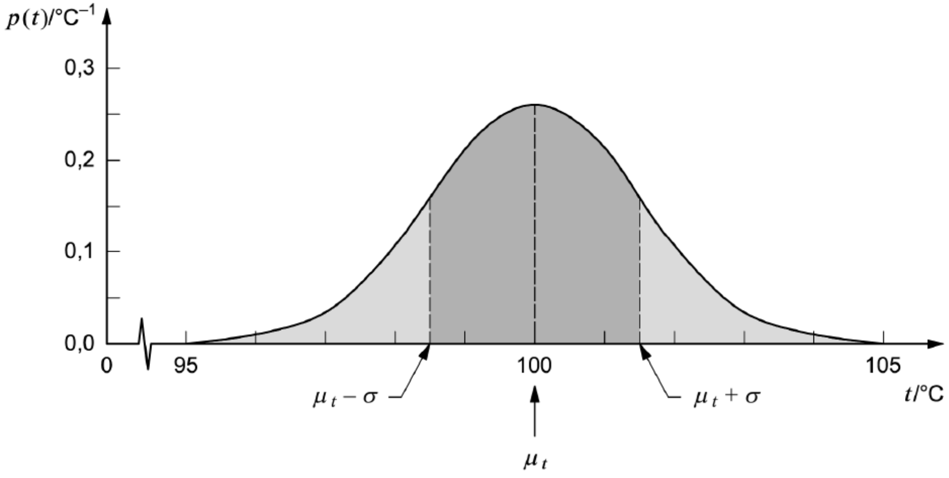
x lies in the interval $[x_- , x_+]$ with **confidence** C

Gaussian confidence intervals

- P(x) = Gaussian distribution with mean μ and variance σ^2 (σ is the standard deviation):
- some examples of confidence intervals:

- $x\pm=\mu\pm k\sigma$ k=1 C = 68%
- $x\pm=\mu\pm k\sigma$ k=2 C = 95.4%
- $x\pm=\mu\pm k\sigma$ k=1.64 C = 90%
- $x\pm=\mu\pm k\sigma$ k=1.96 C = 95%

k is the coverage factor



Poisson distribution

$$\sigma^2 = \bar{x}$$

$If\ \bar{x} > 30 \quad (10)$

$If\ \bar{x} < 30 \quad (10)$

measurement of counts= $\bar{x} \pm \sqrt{\bar{x}}$

~~measurement of counts= $\bar{x} \pm \sqrt{\bar{x}}$~~

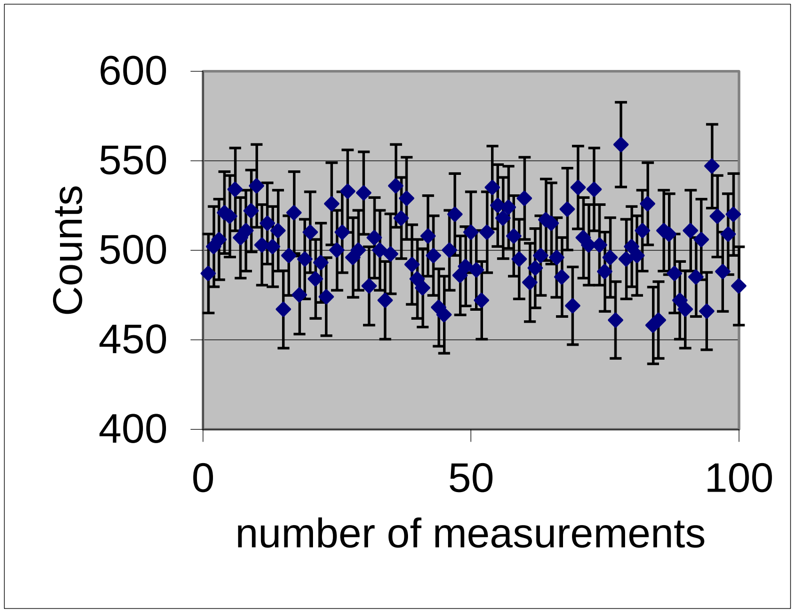
$\chi2$ test

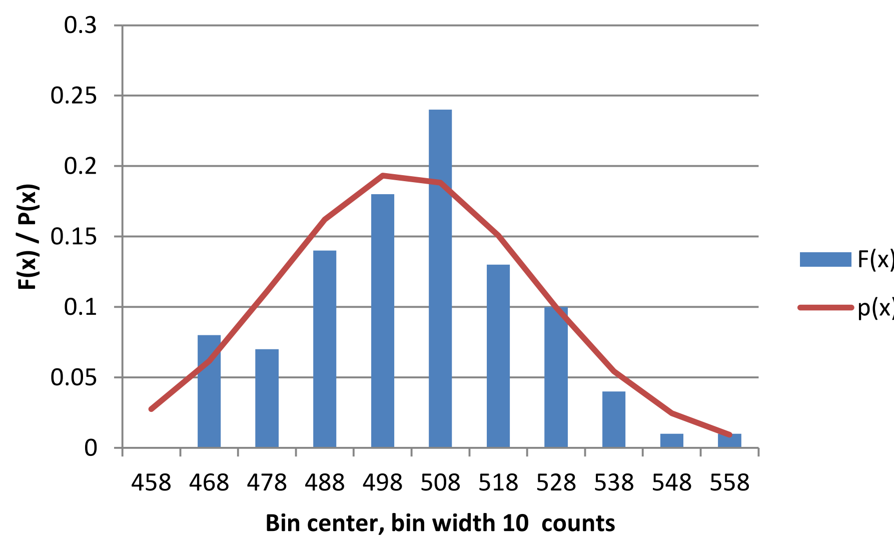
- This test is used to compare an experimental distribution to a theoretical distribution

$F(x)$ frequency distribution ⟺ $P(x)$ probability distribution

χ^2 test

$F(x)$ frequency distribution ⟺ $P(x)$ probability distribution





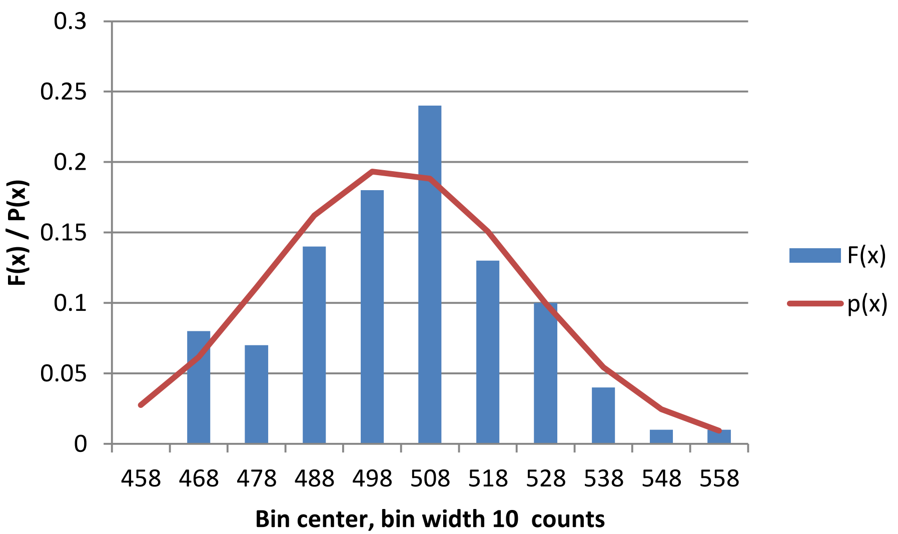
χ2 test

$F(x)$ frequency distribution ⟺ $P(x)$ probability distribution

$$\chi^2 = \sum_{i=1}^{N} \frac{[(nF(x_i) - nP(x_i)]^2}{(\sigma[nF(x_i)])^2}$$

Assuming Poisson

$$(\sigma[nF(x_i)])^2 = nP(x_i)$$



χ2 test

$$\chi^2 = \sum_{i=1}^{N} \frac{[(nF(x_i) - nP(x_i)]^2}{(\sigma[nF(x_i)])^2} \approx \sum_{i=1}^{N} \frac{(\sigma[nF(x_i)])^2}{(\sigma[nF(x_i)])^2} \approx N$$

$$\chi^2 = < \nu > = < N - c >$$

ν = degree of freedom   c=constraint

c=2 for a Poisson distribution

c=3 for a Gauss distribution

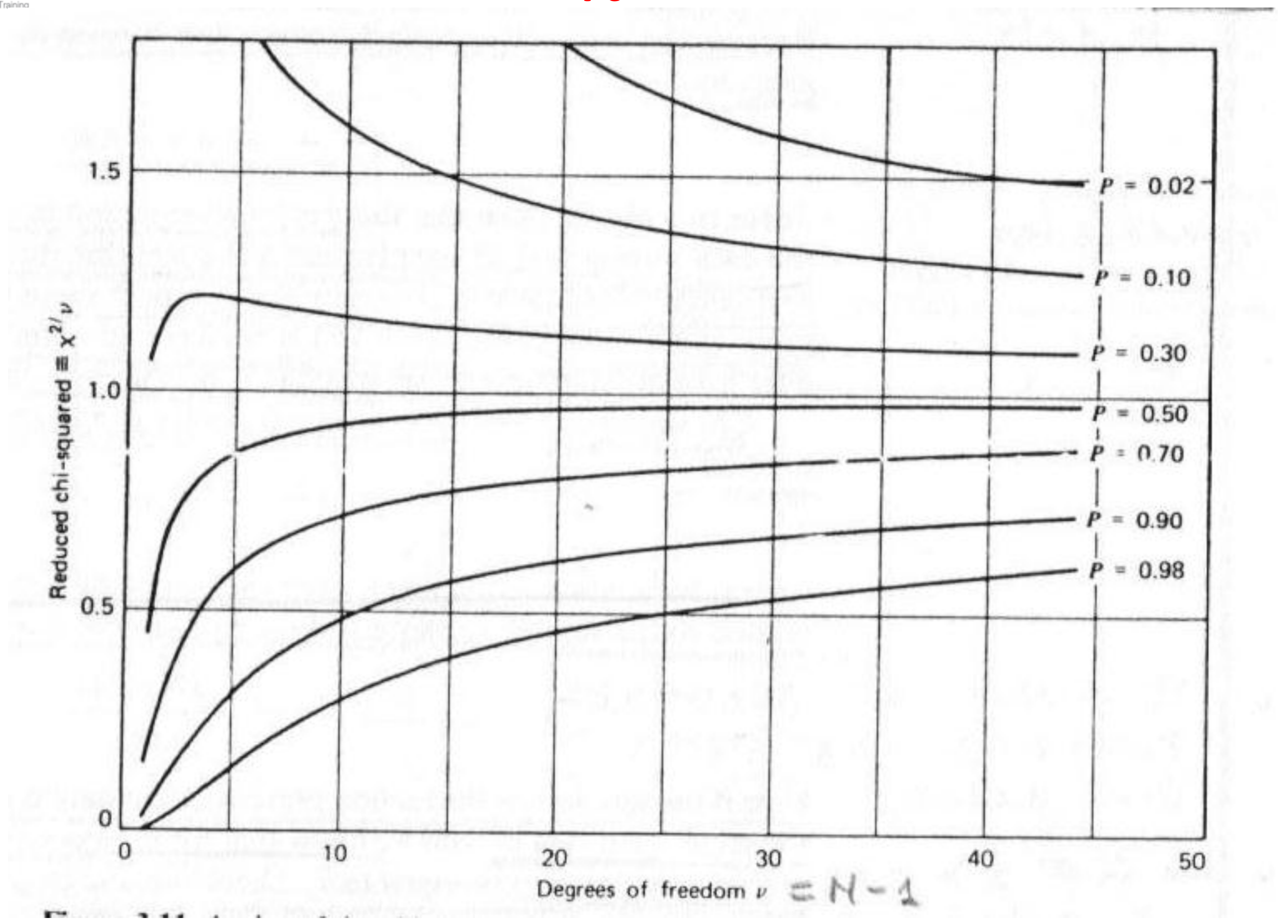
Reduced χ2

$$\chi_r{}^2 = \frac{\chi^2}{\nu} =< 1 >$$

If the $\chi_r{}^2$ <<1 the experimental distribution is «too close» to the target distribution

If the $\chi_r{}^2$ >>1 the experimental distribution is «too far» from the target distribution

χ2 test



P=0.5 is the optimum agreement

χ2 test

The χ^2 can be evaluated without the F(x) distribution

Let us consider a series of n measurements x_i (counts taken in 1 minute) with a mean value X and an experimental variance $\sigma^2(X)$ and let us suppose a Poisson distribution

$$X = \sigma^2(X) \qquad \chi^2 = \sum_{i=1}^{n} \frac{(x_i - X)^2}{X} = \frac{(n-1)s^2}{X} = <(n-1)>$$

s^2 best estimate of the variance

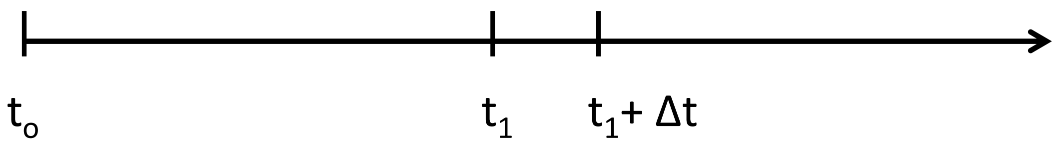
$\chi 2$ N.B.

The χ^2 test holds for raw data only!

	Count (60s)		CPS
	31		0.52
	30		0.50
	36		0.60
	25		0.42
	24		0.40
	33		0.55
	38		0.63
	27		0.45
	22		0.37
	35		0.58
reduced chi^2	0.99		0.02

Poisson distribution in a time domain

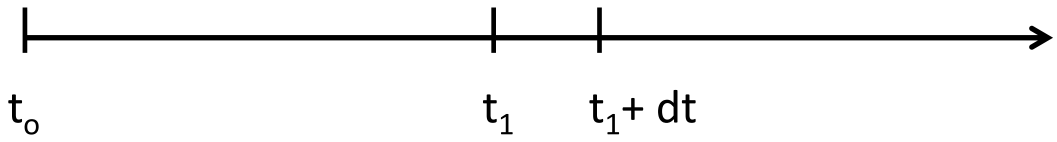
If a Poisson event happens at the time $t_{0,}$ what is the probability P(t) to obtain another Poisson event at the time t_1+Δt.



P(t)dt=(prob. of no event in the interval t_0 - t_1) x (probability of an event in the time interval Δt

Let us call *r* the number of events per second (i.e. the countrate of a deterctor)

Poisson distribution in a time domain



P(t)dt=P(0) x rdt

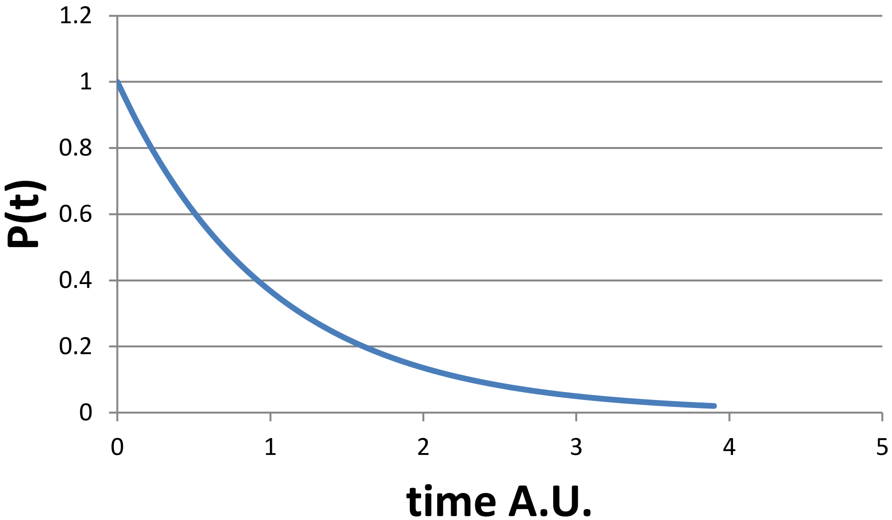
$$P(x) = \frac{(p{\cdot}n)^{x}{\cdot}e^{-p{\cdot}n}}{x!} = \frac{(\bar{x})^{x}{\cdot}e^{-\bar{x}}}{x!}$$

$$P(0) = \frac{(rt)^{0}{\cdot}e^{-rt}}{0!} = e^{-rt}$$

$$P(t)dt = re^{-rt}dt$$

Poisson distribution in a time domain

$$P(t)dt = re^{-rt}dt$$



$$\bar{t} = \frac{\int_0^\infty tP(t)dt}{\int_0^\infty P(t)dt} = \frac{\int_0^\infty tre^{-rt}dt}{\int_0^\infty re^{-rt}dt} = \frac{1}{r}$$

Uncertainty assessement

Uncertainty: parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand

Model

$$Y = (X_1 - X_2) \cdot \frac{X_5 \cdot X_7 \cdot \ldots\ldots X_{N-1}}{X_6 \cdot X_8 \cdot \ldots\ldots X_N} = (X_1 - X_2) \cdot W$$

X_1 gross signal
X_2 background signal

X_5 to X_N correction factors (calibration, environmental parameters etc.)

Uncertainty assessement

$$Y = (X_1 - X_2) \cdot \frac{X_5 \cdot X_7 \cdot ..... X_{N-1}}{X_6 \cdot X_8 \cdot ..... X_N} = (X_1 - X_2) \cdot W$$

We have:

- to assess the uncertainty of every single input variable and the associated probability distribution.
- To compose all the uncertainties obtain the uncertainty associated with *y* (combined uncertainty $u_c(y)$)
- To evaluate the probability distribution associated with *y*

MIND to check the correlation of the input variables

Uncertainty assessement

ISO/IEC GUIDE 98-3:2008 Guide to the expression of uncertainty in measurement

Type A evaluation of standard uncertainty: are founded on frequency distributions.

Type B evaluation of standard uncertainty: are founded on *a priori* distributions.

The standard uncertainty is indicated with the letter *u* (low case)

Uncertainty assessement

Type A evaluation of standard uncertainty

$$\bar{q} = \frac{1}{n}\sum_{k=1}^{n} q_k \qquad s^2(q_k) = \frac{1}{n-1}\sum_{j=1}^{n}\left(q_j - \bar{q}\right)^2 \qquad s^2(\bar{q}) = \frac{s^2(q_k)}{n}$$

MIND that you can use this kind of assessment if you are reasonably sure that the random variable has a Gaussian distribution.

Uncertainty assessement

Type B evaluation of standard uncertainty

— previous measurement data;

— experience with or general knowledge of the behaviour and properties of relevant materials and instruments;

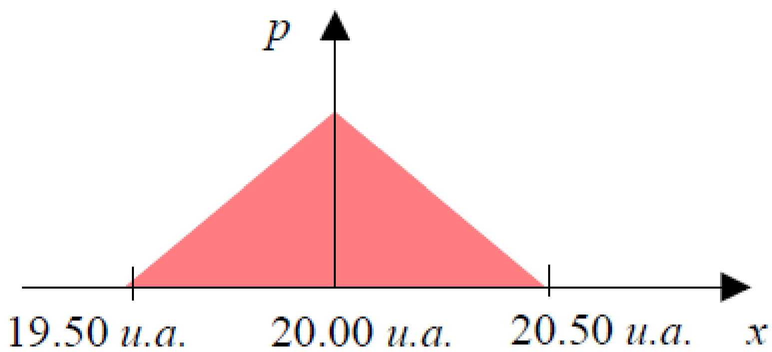
— manufacturer's specifications;

— data provided in calibration and other certificates;

— uncertainties assigned to reference data taken from handbooks.

Examples of Type B evaluation

If I have only 3 repeated measurements of the same quantity the experimental standard deviation is meaningless. One technique is to define «a priori» a reasonable probability distribution

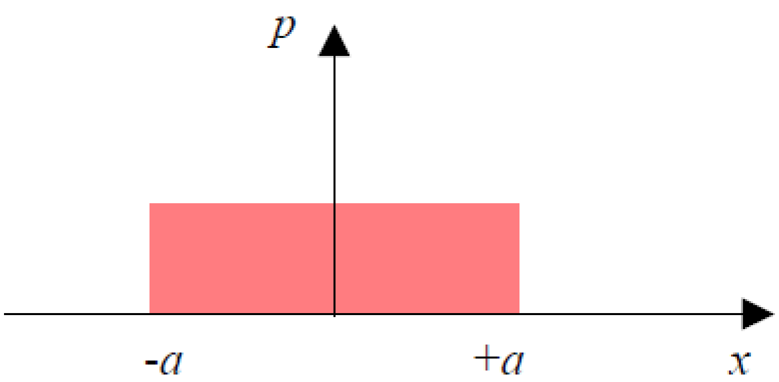


$$\sigma^2 = \int_{-\infty}^{+\infty} p(x)(x - \bar{x})^2 \, dx$$

$$u(x_i) = \frac{0.50}{\sqrt{6}} = 0.20 \; u.a.$$

Examples of Type B evaluation

I have only 2 repeated measurements of the same quantity

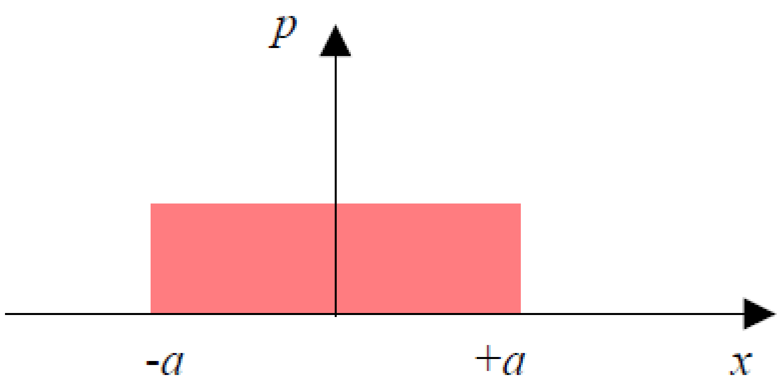


$$\sigma^2 = \int_{-\infty}^{+\infty} p(x)(x-\bar{x})^2\,dx$$

$$u(x_i) = \frac{a}{\sqrt{3}}$$

Examples of Type B evaluation

Reading of a digital instrument, for instance a voltmeter

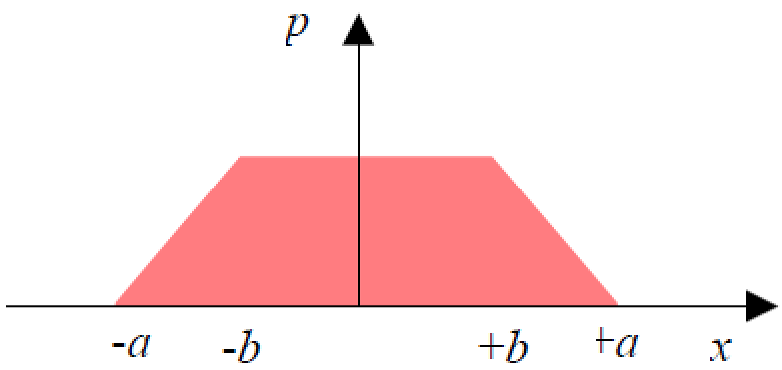


V=1.5 resolution 0.1 V

It can be assumed V in the range 1.45 – 1.55

Examples of Type B evaluation

Another possibility is a trapezoidal distribution



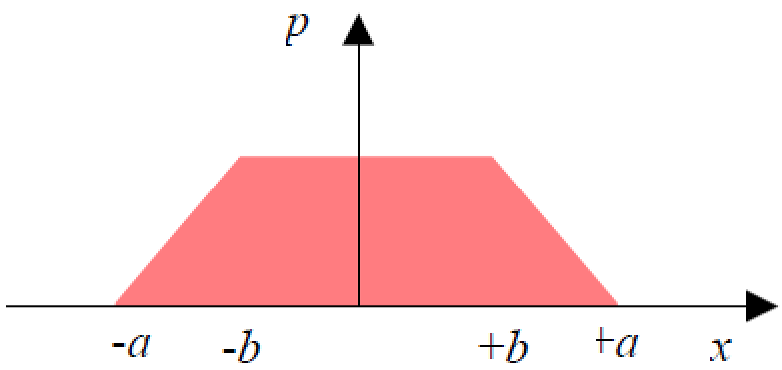
$$\sigma^2 = \int_{-\infty}^{+\infty} p(x)(x-\bar{x})^2\, dx$$

$$2b = 2a\beta$$

$$u(x_i) = \frac{a\sqrt{1+\beta^2}}{\sqrt{6}}$$

$$0 \leq \beta \leq 1$$

Examples of Type B evaluation



The digital indication oscillates between

V=1.5 and V=1.7

1.45-1.55 1.65-1.75 Resol. 0.1V

It can be assumed

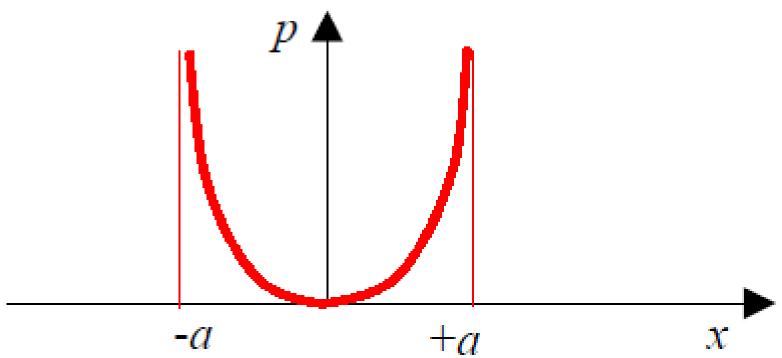
Vmean =1.6 V

b=0.1 V (oscillation)

a=0.05V (resolution)

Examples of Type B evaluation

A U shaped distribution is the typical distribution of the temperature in an air-conditioned lab. Usually the chiller starts and stops according to a temperature sensor. This causes a sinusoidal behavior of the temperature



$$\sigma^2 = \int_{-\infty}^{+\infty} p(x)(x-\bar{x})^2\,dx$$

$$p(x) = \frac{1}{\pi}\frac{1}{\sqrt{a^2 - x^2}}$$

$$u(x_i) = \frac{a}{\sqrt{2}}$$

Examples of Type B evaluation

Uncertainty associated to the calibration factor

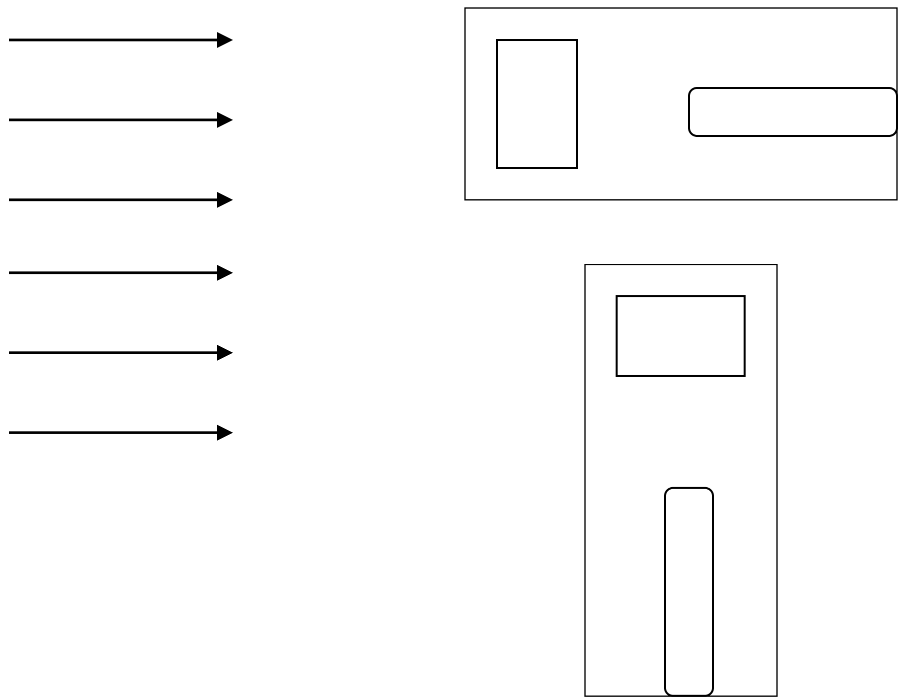
Usually the calibration factor is given with an associated uncertainty (if the calibration lab is honest).

But sometimes the calibration factor and the uncertainty cannot be used "as they are" because it's impossible to reproduce the same experimental conditions of the calibration lab.

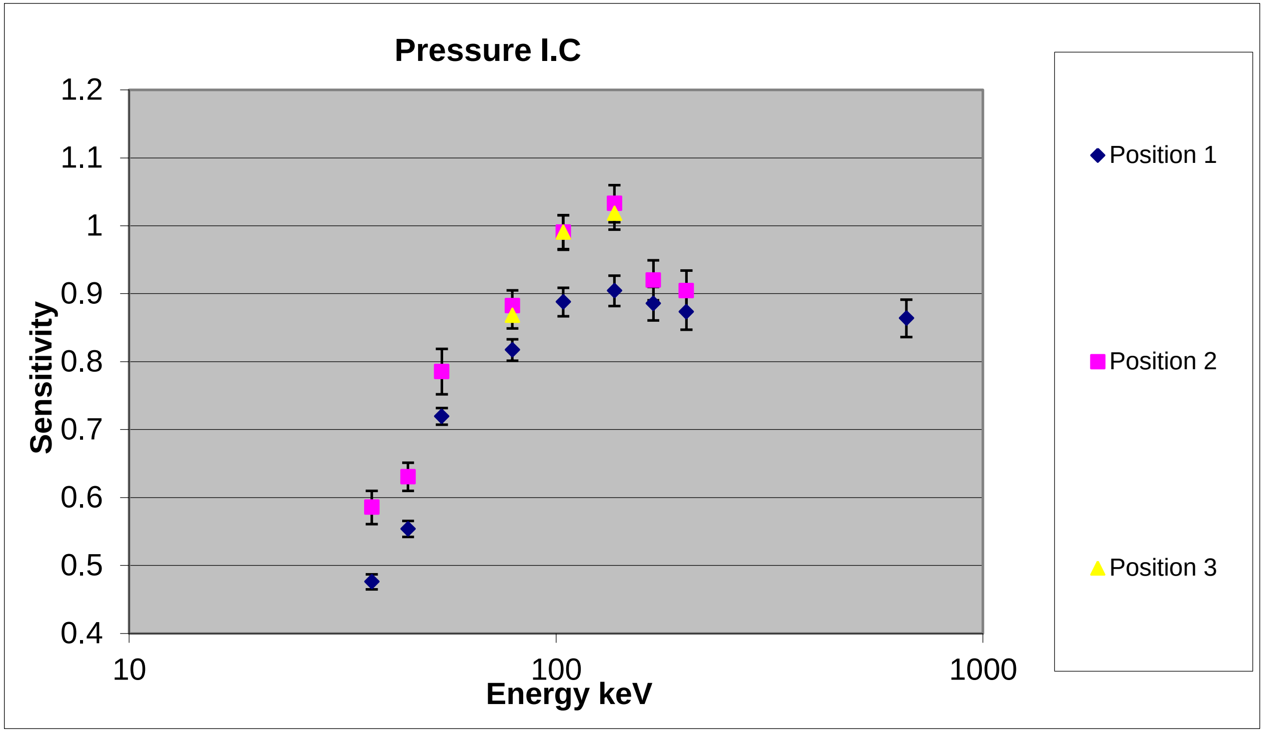
Let's consider the following problem: I have to measure the air kerma in a photon field with a survey meter. I have the instrument calibration factor for different photon energies, but I don't know exactly the energy distribution of the photon field I'm going to measure.

Uncertainty associated to the calibration factor

Beam direction



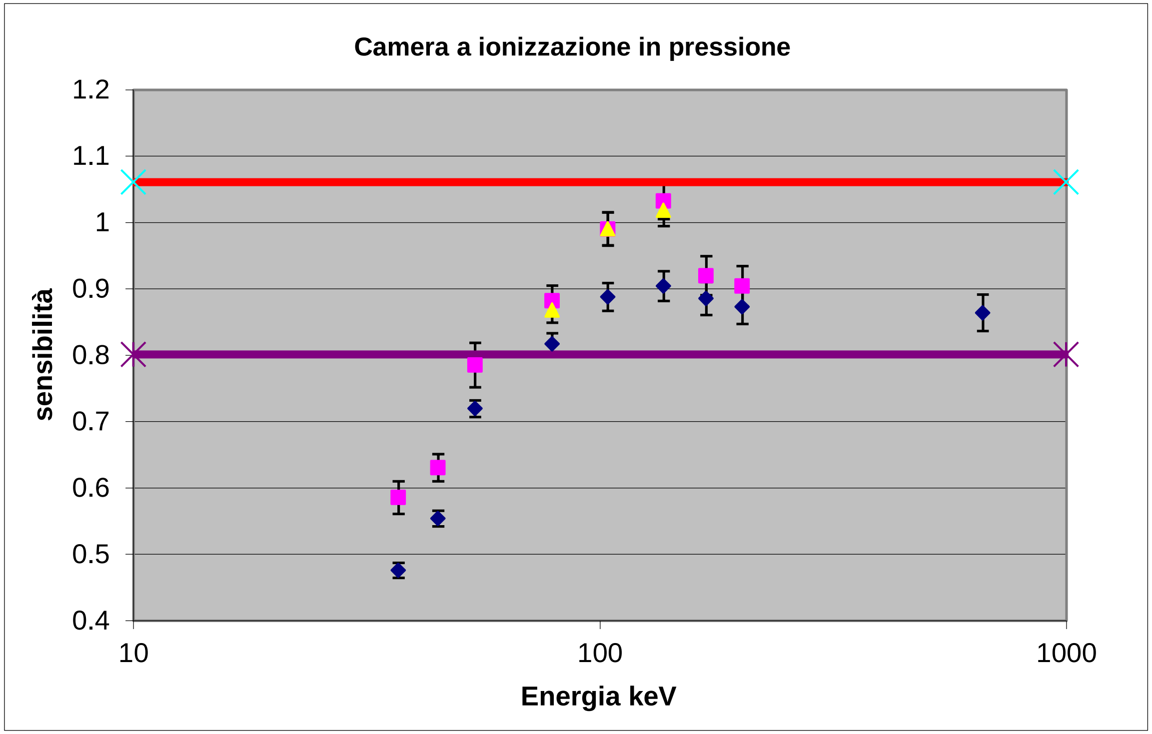
Uncertainty associated to the calibration factor



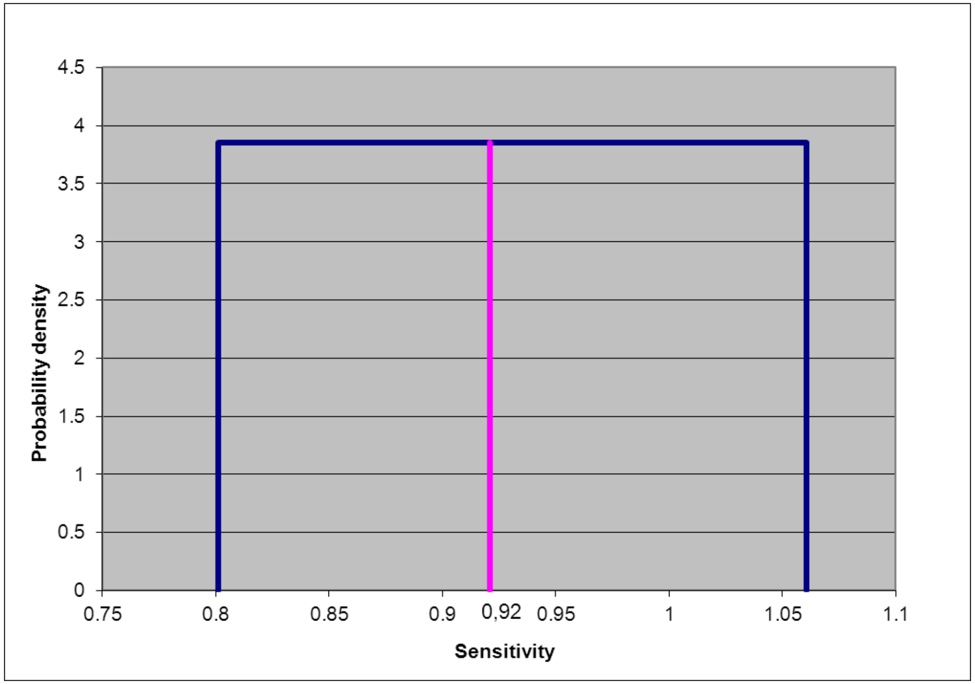
Sensitivity in term of air kerma

Uncertainty associated to the calibration factor

I can suppose that the photon energy is in the range 80 keV – 200 keV



Uncertainty associated to the calibration factor



$$0{,}92 = \bar{S} = \frac{1}{n}\sum_{i=1}^{n} S_i$$

$$0{,}93 = \bar{\bar{S}} = \frac{\max - \min}{2}$$

$$u(\bar{S}) = \frac{0{,}14}{\sqrt{3}} = 0{,}0807$$

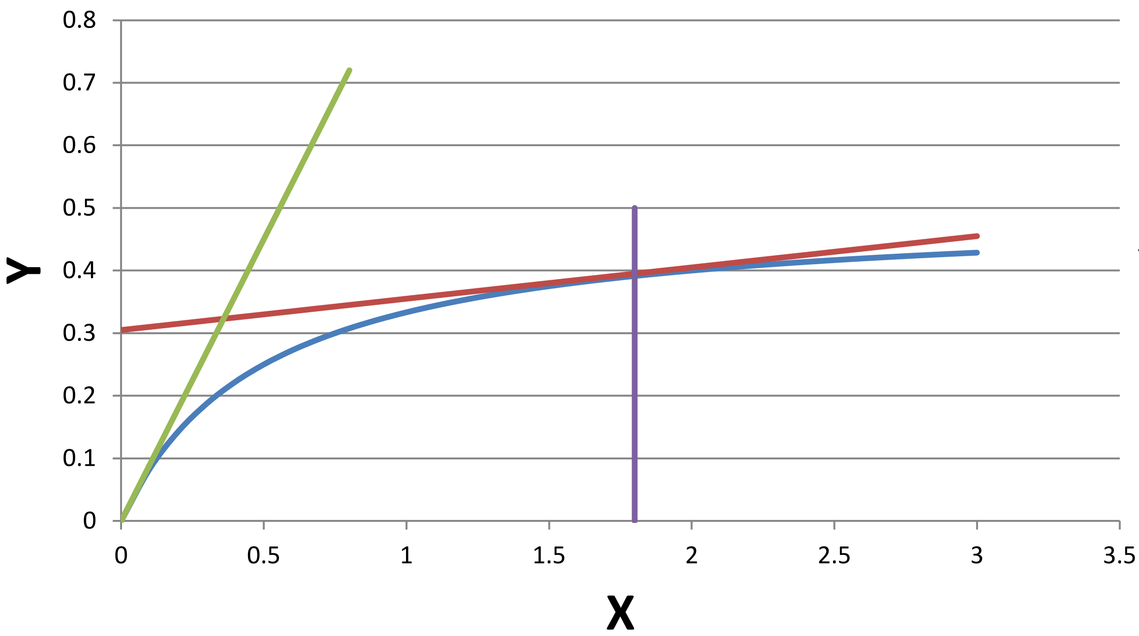
$$u_{\%}(N) = u_{\%}(\bar{S}) = \frac{u(\bar{S})}{S} = \frac{0{,}0807}{0{,}92} = 8{,}8\%$$

Uncertainty associated to the calibration factor

This last example is important because it shows that the important uncertainties arise from an incomplete definition of the quantity under measurement (energy distribution).
For "on field" measurements it is important, and difficult, to assess these kinds of uncertainties. The researcher experience plays a key role.
The main issues are:

- Find out all the uncertainty sources
- Define the most reasonable probability distributions

Uncertainty propagation



$$y = f(x)$$

$$u^2(y) = \left(\frac{\partial(f(x))}{\partial x}\right)^2 u^2(x)$$

Uncertainty propagation

$$y = f(x)$$

$$countrate = \frac{counts}{t}$$

$$u^2(y) = \left(\frac{\partial(f(x))}{\partial x}\right)^2 u^2(x)$$

$$u^2(counrate) = \frac{u^2(counts)}{t^2}$$

Uncertainty propagation

In case of uncorrelated input variables

$$y = f(x_1, x_2, \dots, x_N)$$

$$u_c^2(y) = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i)$$

Uncertainty propagation

What is the probability distribution of *y*

$$y = f(x_1, x_2, \ldots, x_N)$$

$$u_c^2(y) = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i)$$

Central limit theorem (**CLT**) states that, given certain conditions, the mean of a sufficiently large number of independent random variables, each with finite mean and variance, will be approximately normally distributed

Uncertainty propagation

The Central Limit Theorem is significant because it shows the very important role played by the variances of the probability distributions of the input quantities. It implies that the convolved distribution converges towards the normal distribution as the number of input quantities contributing to the variance of Y increases and that the convergence will be more rapid the closer the values of $\left(\frac{\partial(f(x))}{\partial x}\right)^2 \sigma^2(x)$ are to each other (equivalent in practice to each input estimate x_i contributing a comparable uncertainty to the uncertainty of the estimate y of the measurand Y)

Uncertainty propagation

Data are usually expressed in term of expanded uncertainty *U* (upper case).

The expanded uncertainty *U* is obtained by multiplying the combined standard uncertainty $u_c(y)$ by a coverage factor *k*: $U = ku_c(y)$

The value of the coverage factor *k* is chosen on the basis of the level of confidence required of the interval $y - U$ *to* $y + U$.

The standard choice is a 95% level of confidence. If a normal distribution is assumed, this means *K=2*

Uncertainty propagation

Let's get back to the air kerma measurement in a photon field with a survey meter.

Measuring model $$R = M \cdot N = \frac{M}{S}$$

R measurement result
M instrument reading
N calibration factor
S sensitivity

M=1.00 mGy

$$u_{\%}(R) = \sqrt{u_{\%}^{2}(M) + u_{\%}^{2}(S)} \qquad u_{\%}(N) = u_{\%}(S) = 8{,}8\% \qquad u_{\%}^{2}(M) << u_{\%}^{2}(S)$$

Uniform probability distribution

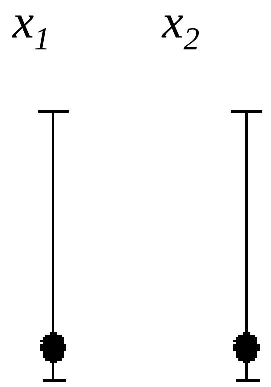
C.L. about 70% $$u_{\%}(R) = 8{,}8\%$$

C.L. about 95%
K=1,65 $$U_{\%}(R) = 14{,}5\%$$

Correlation among input variables

$$y = f(x_1, x_2)$$

Uncorrelated variables

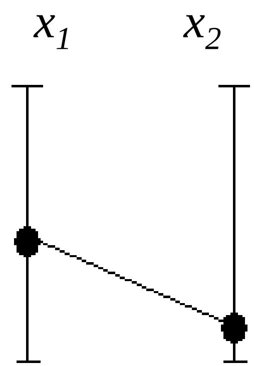


$$u^2(y) = \left(\frac{\partial f}{\partial x_1}\right)^2 u^2(x_1) + \left(\frac{\partial f}{\partial x_2}\right)^2 u^2(x_2)$$

Correlation among input variables

$$y = f(x_1, x_2)$$

Correlated variables



Term of covariance

$$u^2(y) = \left(\frac{\partial f}{\partial x_1}\right)^2 \cdot u^2(x_1) + \left(\frac{\partial f}{\partial x_2}\right)^2 \cdot u^2(x_2) + 2 \cdot \frac{\partial f}{\partial x_1} \cdot \frac{\partial f}{\partial x_2} \cdot u(x_1, x_2)$$

Covariance

Correlated input variables

$$u^2(x) = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$u(x_1, x_2) = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_{1i} - \bar{x}_1) \cdot (x_{2i} - \bar{x}_2)$$

$$u^2(\bar{x}) = \frac{1}{n \cdot (n-1)} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$u(\bar{x}_1, \bar{x}_2) = \frac{1}{n \cdot (n-1)} \cdot \sum_{i=1}^{n} (x_{1i} - \bar{x}_1) \cdot (x_{2i} - \bar{x}_2)$$

Correlated input variables

Correlation coefficient

$$r(x_1, x_2) = \frac{u(x_1, x_2)}{u(x_1) \cdot u(x_2)}$$

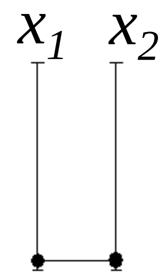
Term of covariance

$$2 \cdot \frac{\partial f}{\partial x_1} \cdot \frac{\partial f}{\partial x_2} \cdot u(x_1) \cdot u(x_2) \cdot r(x_1, x_2)$$

$$-1 < r(x_1, x_2) < 1$$

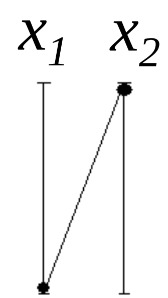
Correlated input variables

$$r(x_1, x_2) = 1$$ Positive correlation



$$r(x_1, x_2) = 0$$ Uncorrelated variables

$$r(x_1, x_2) = -1$$ Negative correlation



$$u(x_1, x_2) = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_{1_i} - \bar{x}_1) \cdot (x_{2_i} - \bar{x}_2)$$

Uncertainty propagation

Experimental evaluation of the covariance
(type A evaluation)

$$s(\bar{q},\bar{r}) = \frac{1}{n(n-1)}\sum_{k=1}^{n}(q_k-\bar{q})(r_k-\bar{r})$$

s is an estimator of the covariance

Correlation of input variables

A type B evaluation of the correlation can be done by observing the a variation δ_1 in x_1 produces a variation δ_2 in x_2. The correlation coefficient can be evaluated as follows:

$$r(x_1, x_2) = \frac{u(x_1) \cdot \delta_2}{u(x_2) \cdot \delta_1}$$

Correlation of input variables

Sometimes it is possible to remove the correlation modifying the measuring model.

$$y = f\left(x_1(t), x_2(t)\right) \quad \Rightarrow \quad y = g\left(x_1, x_2, t\right)$$

Uncertainty propagation

In case of correlated input variables

$$y = f(x_1, x_2, \ldots, x_N)$$

$$u_c^2(y) = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}u(x_i, x_j) = \sum_{i=1}^{N}\left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}u(x_i, x_j)$$

Example

ISO/IEC GUIDE 98-3:2008H.4 Measurement of activity

Problem: the unknown radon activity concentration in a water sample is determined by liquid-scintillation counting against a radon-in-water standard sample

C_S, C_B, C_x are the number of counts recorded in the dead-time-corrected counting intervals T_0 = 60 min for the standard, blank, and sample vials, respectively.

t_S, t_B, t_x are the times from the reference time t = 0 to the midpoint of the dead-time-corrected counting intervals T_0 = 60 min for the standard, blank, and sample vials, respectively

Example

ISO/IEC GUIDE 98-3:2008H.4 Measurement of activity

The observed counts may be expressed as

$$C_S = C_B + \varepsilon A_S T_0 m_S e^{-\lambda t_S}$$

$$C_x = C_B + \varepsilon A_x T_0 m_x e^{-\lambda t_x}$$

ε	is the liquid scintillation detection efficiency for ^{222}Rn for a given source composition, assumed to be independent of the activity level;
A_S	is the activity concentration of the standard at the reference time $t = 0$;
A_x	is the *measurand* and is defined as the unknown activity concentration of the sample at the reference time $t = 0$;
m_S	is the mass of the standard solution;
m_x	is the mass of the sample aliquot;
λ	is the decay constant for ^{222}Rn: $\lambda = (\ln 2)/T_{1/2} = 1{,}258\ 94 \times 10^{-4}\ \text{min}^{-1}$ ($T_{1/2} = 5\ 505{,}8$ min).

Example

ISO/IEC GUIDE 98-3:2008H.4 Measurement of activity

Table H.7 — Counting data for determining the activity concentration of an unknown sample

Cycle	Standard		Blank		Sample	
k	t_S (min)	C_S (counts)	t_B (min)	C_B (counts)	t_x (min)	C_x (counts)
1	243,74	15 380	305,56	4 054	367,37	41 432
2	984,53	14 978	1 046,10	3 922	1 107,66	38 706
3	1 723,87	14 394	1 785,43	4 200	1 846,99	35 860
4	2 463,17	13 254	2 524,73	3 830	2 586,28	32 238
5	3 217,56	12 516	3 279,12	3 956	3 340,68	29 640
6	3 956,83	11 058	4 018,38	3 980	4 079,94	26 356

Example

ISO/IEC GUIDE 98-3:2008H.4 Measurement of activity

Measuring model

$$A_x = f\left(A_S, m_S, m_x, C_S, C_x, C_B, t_S, t_x, \lambda\right)$$

$$= A_S \frac{m_S}{m_x} \frac{\left(C_x - C_B\right) e^{\lambda t_x}}{\left(C_S - C_B\right) e^{\lambda t_S}}$$

$$= A_S \frac{m_S}{m_x} \frac{C_x - C_B}{C_S - C_B} e^{\lambda (t_x - t_S)}$$

$$A_x = f\left(A_S, m_S, m_x, R_S, R_x\right) = A_S \frac{m_S}{m_x} \frac{R_x}{R_S}$$

$$R_x = \left[\left(C_x - C_B\right)/T_0\right] e^{\lambda t_x}$$

$$R_S = \left[\left(C_S - C_B\right)/T_0\right] e^{\lambda t_S}$$

Example

ISO/IEC GUIDE 98-3:2008H.4 Measurement of activity

The arithmetic means $\overline{R_S}$, $\overline{R_x}$ and $\overline{R}$, and their experimental standard deviations s($\overline{R_S}$), s($\overline{R_x}$), and s($\overline{R}$), are calculated in the usual way:

$$\bar{q} = \frac{1}{n}\sum_{k=1}^{n} q_k$$

$$s^2(q_k) = \frac{1}{n-1}\sum_{j=1}^{n}\left(q_j - \bar{q}\right)^2$$

$$R = R_x/R_S = \left[\left(C_x - C_B\right)/\left(C_S - C_B\right)\right]e^{\lambda(t_x - t_S)}$$

$$R_x = \left[\left(C_x - C_B\right)/T_0\right]e^{\lambda t_x}$$

$$R_S = \left[\left(C_S - C_B\right)/T_0\right]e^{\lambda t_S}$$

Example

ISO/IEC GUIDE 98-3:2008H.4 Measurement of activity

The correlation coefficient $r(\overline{R_S}, \overline{R_x})$ is assessed with a type A calculation

$$s(\overline{q}, \overline{r}) = \frac{1}{n(n-1)} \sum_{k=1}^{n} (q_k - \overline{q})(r_k - \overline{r})$$

$$r(x_i, x_j) = \frac{u(x_i, x_j)}{u(x_i)u(x_j)}$$

$$R_x = [(C_x - C_B)/T_0]e^{\lambda t_x}$$

$$R_S = [(C_S - C_B)/T_0]e^{\lambda t_S}$$

Example

ISO/IEC GUIDE 98-3:2008H.4 Measurement of activity

There are two ways to face the problem:

With correlation

$$A_x = A_s \frac{m_S \overline{R_x}}{m_x \overline{R_S}}$$

without correlation

$$A_x = A_s \frac{m_S}{m_x} \bar{R}$$

Example

ISO/IEC GUIDE 98-3:2008H.4 Measurement of activity

Table H.8 — Calculation of decay-corrected and background-corrected counting rates

Cycle k	R_x (min^{-1})	R_S (min^{-1})	$t_x - t_S$ (min)	$R = R_x / R_S$
1	652,46	194,65	123,63	3,352 0
2	666,48	208,58	123,13	3,195 3
3	665,80	211,08	123,12	3,154 3
4	655,68	214,17	123,11	3,061 5
5	651,87	213,92	123,12	3,047 3
6	623,31	194,13	123,11	3,210 7
	$\bar{R}_x = 652{,}60$ $s(\bar{R}_x) = 6{,}42$ $s(\bar{R}_x)/\bar{R}_x = 0{,}98 \times 10^{-2}$	$\bar{R}_S = 206{,}09$ $s(\bar{R}_S) = 3{,}79$ $s(\bar{R}_S)/\bar{R}_S = 1{,}84 \times 10^{-2}$		$\bar{R} = 3{,}170$ $s(\bar{R}) = 0{,}046$ $s(\bar{R})/\bar{R} = 1{,}44 \times 10^{-2}$
	$\bar{R}_x/\bar{R}_S = 3{,}167$ $u(\bar{R}_x/\bar{R}_S) = 0{,}045$ $u(\bar{R}_x/\bar{R}_S)/(\bar{R}_x/\bar{R}_S) = 1{,}42 \times 10^{-2}$			
Correlation coefficient				
$r(\bar{R}_x, \bar{R}_S) = 0{,}646$				

Example

ISO/IEC GUIDE 98-3:2008H.4 Measurement of activity

Result using the approach with correlation

$$A_x = A_S \frac{m_S}{m_x} \frac{\overline{R}_x}{\overline{R}_S} = 0{,}430\,0 \text{ Bq/g}$$

$$\frac{u_c^2(A_x)}{A_x^2} = \frac{u^2(A_S)}{A_S^2} + \frac{u^2(m_S)}{m_S^2} + \frac{u^2(m_x)}{m_x^2} + \frac{u^2(\overline{R}_x)}{\overline{R}_x^2} + \frac{u^2(\overline{R}_S)}{\overline{R}_S^2} - 2r(\overline{R}_x, \overline{R}_S) \frac{u(\overline{R}_x) u(\overline{R}_S)}{\overline{R}_x \overline{R}_S}$$

$$u_c(A_x) = 0{,}008\,3 \text{ Bq/g}$$

Example

ISO/IEC GUIDE 98-3:2008H.4 Measurement of activity

Result using the approach without correlation

$$A_x = A_S \frac{m_S}{m_x} \overline{R} = 0{,}430\,4\ \mathrm{Bq/g}$$

$$\frac{u_c^2(A_x)}{A_x^2} = \frac{u^2(A_S)}{A_S^2} + \frac{u^2(m_S)}{m_S^2} + \frac{u^2(m_x)}{m_x^2} + \frac{u^2(\overline{R})}{\overline{R}^2}$$

$$u_c(A_x) = 0{,}008\,4\ \mathrm{Bq/g}$$

Example

ISO/IEC GUIDE 98-3:2008H.4 Measurement of activity

Comparison of the two approaches

With correlation

$$A_x = A_S \frac{m_S}{m_x} \frac{\overline{R}_x}{\overline{R}_S} = 0{,}430\,0 \text{ Bq/g}$$

$$u_c(A_x) = 0{,}008\,3 \text{ Bq/g}$$

without correlation

$$A_x = A_S \frac{m_S}{m_x} \overline{R} = 0{,}430\,4 \text{ Bq/g}$$

$$u_c(A_x) = 0{,}008\,4 \text{ Bq/g}$$

Characteristics limits
decision threshold and detection limit

Suppose we measure the activity in an unknown sample:

- the "decision threshold" gives a decision on whether or not the physical effect quantified by the measurand is present;

- the "detection limit" indicates the smallest true value of the measurand which can still be detected; this gives a decision on whether or not the measurement procedure satisfies the requirements and is therefore suitable for the intended measurement purpose

decision threshold

Let us suppose to know, « a priori» that in the sample there is no activity. The problem is: define a threshold "decision threshold" that permits to define a probability of false positive.

decision threshold

Measuring model

$$y = (x_1 - x_2) \cdot w = \left(\frac{N_S}{T} - \frac{N_B}{T} \right) \cdot w$$

$$u(y) = \sqrt{w^2 \cdot \left(u^2(x_1) + u^2(x_2)\right) + y^2 \cdot u_{rel}^2(w)}$$

$$u^2(x_1) = \frac{N_S}{T^2}$$

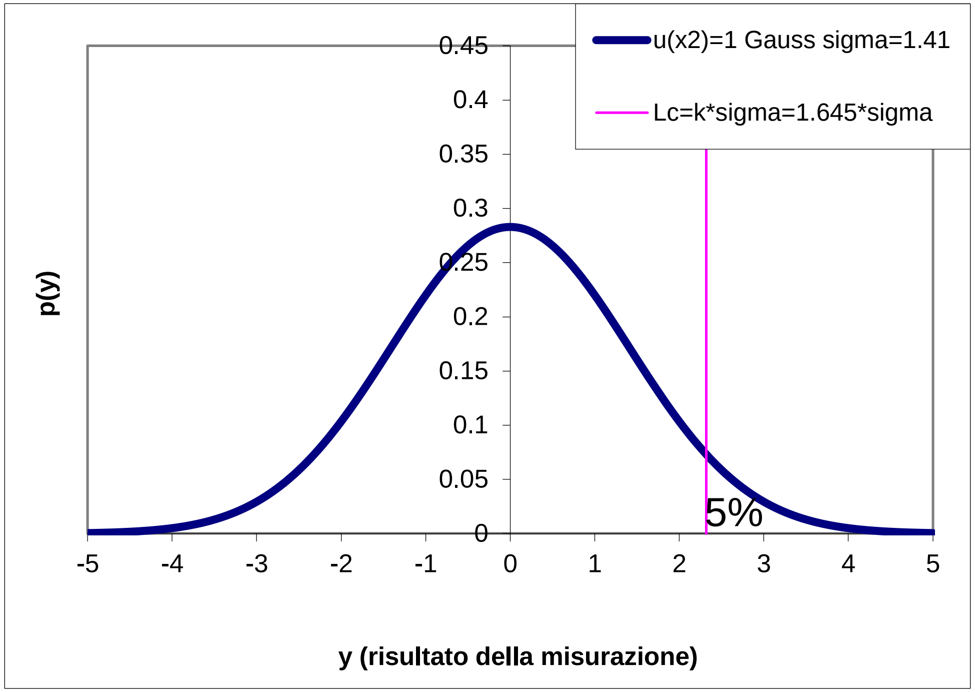
$$u^2(x_2) = \frac{N_B}{T^2}$$

No activity in the sample $y=0$; $x_1=x_2=$background

$$u(0) = w \cdot \sqrt{\left(u^2(x_1) + u^2(x_2)\right)} = w \cdot \sqrt{2 \cdot u^2(x_2)}$$

decision threshold

Critical level



Si presuppone y=0

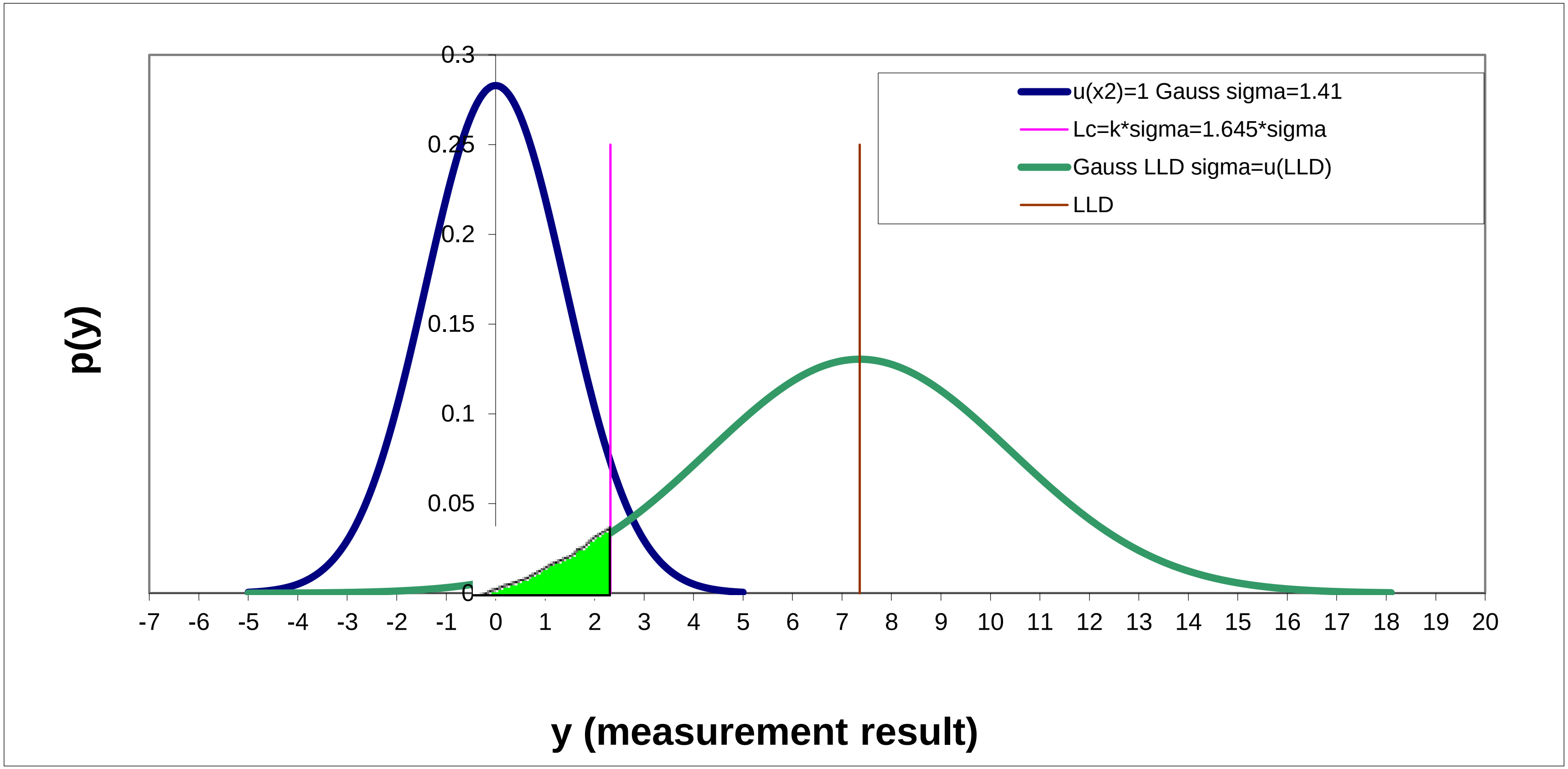
$$u(0) = w \cdot \sqrt{2 \cdot u^2(x_2)}$$

*Lc=1.645*u(0)*

Lc Critical level or decision threshold defines the percentage of false positive

It depends on the uncertainty $u(x_2)$ of the background measurement

detection limit
LLD (Lower limit of detection)



detection limit
LLD (Lower limit of detection)

Let's express the uncertainty as a function of the measurand

$$y = (x_1 - x_2)\cdot w = \left(\frac{N_S}{T} - \frac{N_B}{T}\right)\cdot w \qquad u(y) = \sqrt{w^2\cdot\left(u^2(x_1) + u^2(x_2)\right) + y^2\cdot u_{rel}^2(w)}$$

$$x_1 = g(y) \qquad u(y) = h(y)$$

$$x_1 = \frac{y}{w} + x_2$$

e

$$u(y) = \sqrt{w^2\cdot\left(\frac{y}{w\cdot T} + 2\cdot u^2(x_2)\right) + y^2\cdot u_{rel}^2(w)}$$

$$u^2(x_1) = \frac{N_S}{T^2} \qquad u^2(x_2) = \frac{N_B}{T^2}$$

detection limit
LLD (Lower limit of detection)

LLD ($y^{\#}$) Can be calculated as follows:

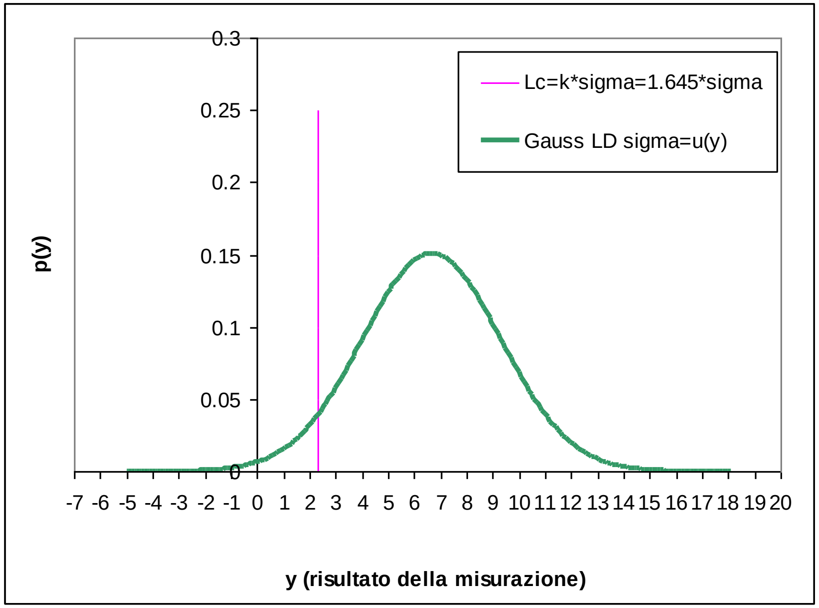
$$y^{\#} = Lc + k \cdot u\left(y^{\#}\right) = Lc + k \cdot \sqrt{w^{2} \cdot \left(\frac{y^{\#}}{w \cdot T} + 2 \cdot u^{2}(x_{2}) \right) + y^{\#2} \cdot u_{rel}^{2}(w)}$$

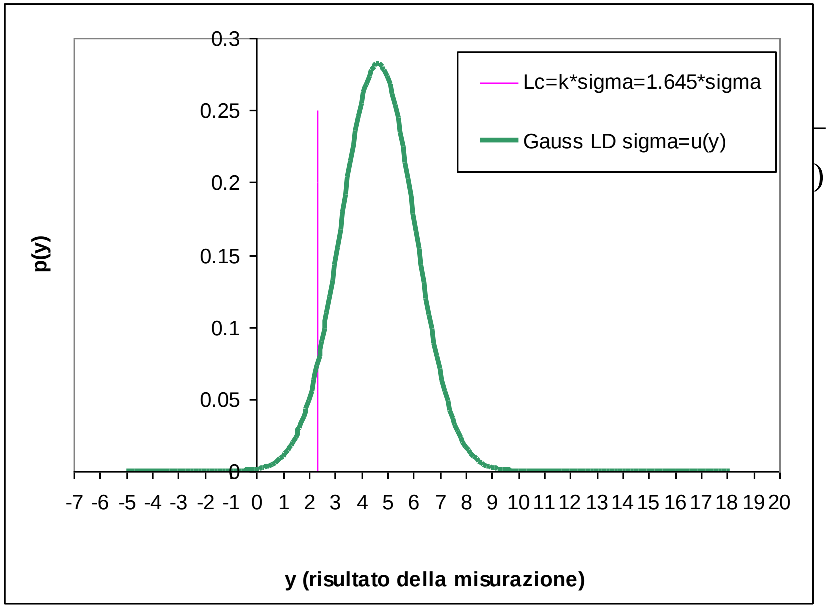
Where k in the coverage factor coresponding to a given probability of false negative. K=1.645 -> p=5%

The equation can be solved in an iterative way.

LLD depends on x_2 and W

detection limit
LLD (Lower limit of detection)





High $u_{rel}^2(w)$ Low $u_{rel}^2(w)$

$$y^{\#} = y^{*} + k \cdot \sqrt{w^2 \cdot \left(\frac{y^{\#}}{w \cdot T} + 2 \cdot u^2(x_2) \right) + y^{\#2} \cdot u_{rel}^2(w)}$$

Benford's law
first digit distribution

We have the following problem: we need to calibrate a survey meter for X and gamma radiation in a calibration lab. In order to fit our budget we can get 1 point for every full scale.

e.g.

One point in the range 0-10 µSv (10 µSv full scale)

One point in the range 0-100 µSv (100 µSv full scale)

Etc.

Benford's law first digit distribution

How can we choose the calibration point?

e.g. in the range 0-10μSv which is the better choice?

1μSv or 2μSv….or 9μSv. In other words the calibration point must start with the digit 1 or 2 …..or 9.

We can give an answer by addressing another question:

During the routine on field measurements which is the first digit more probable to obtain?

One could say: every digit has the same probability, but……

Benford's law
first digit distribution

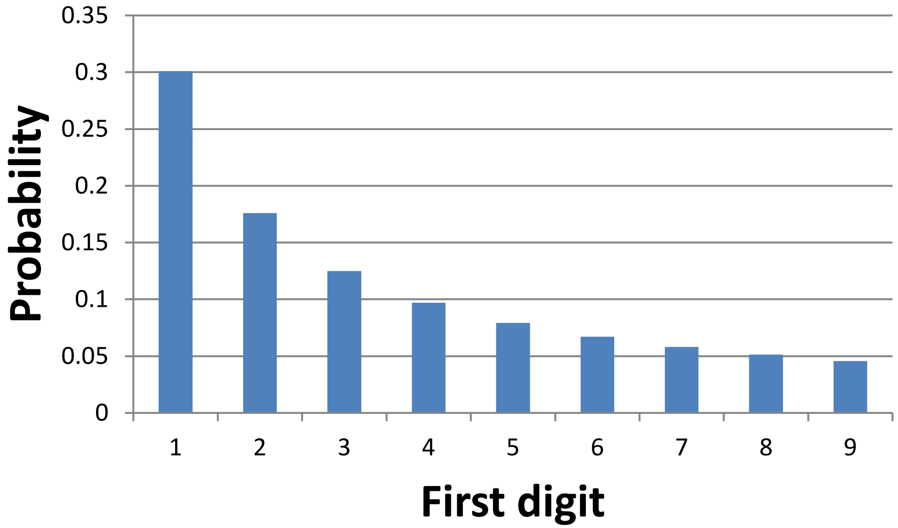
....this is not true!!!

Benford's Law (which was first mentioned in 1881 by the astronomer Simon Newcomb) states that if we randomly select a number from a table of physical constants or statistical data, the probability of occurrence of the first digit is distributed as follows:

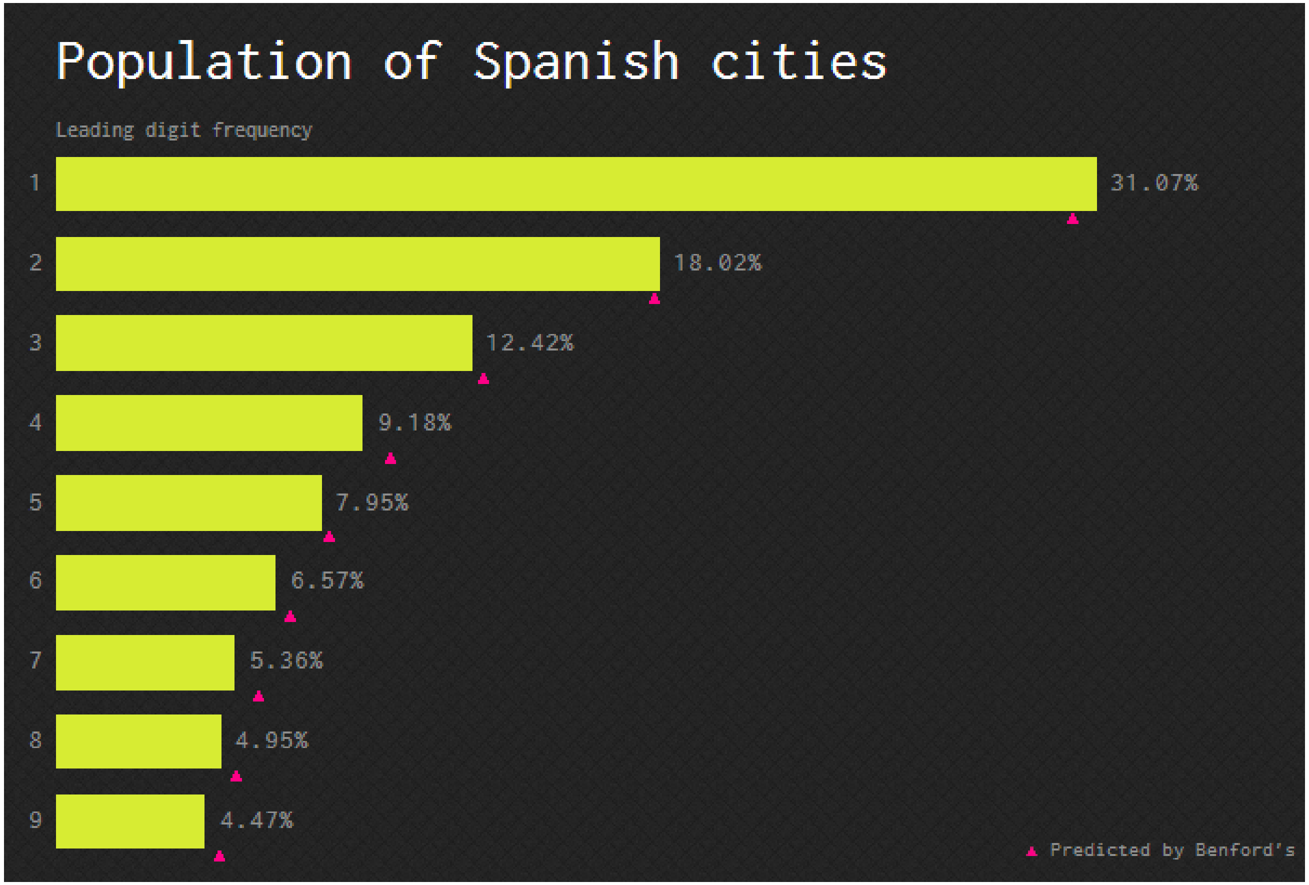
$$P(d) = \frac{Ln(1+\frac{1}{d})}{Ln(10)}$$

Benford's law
first digit distribution

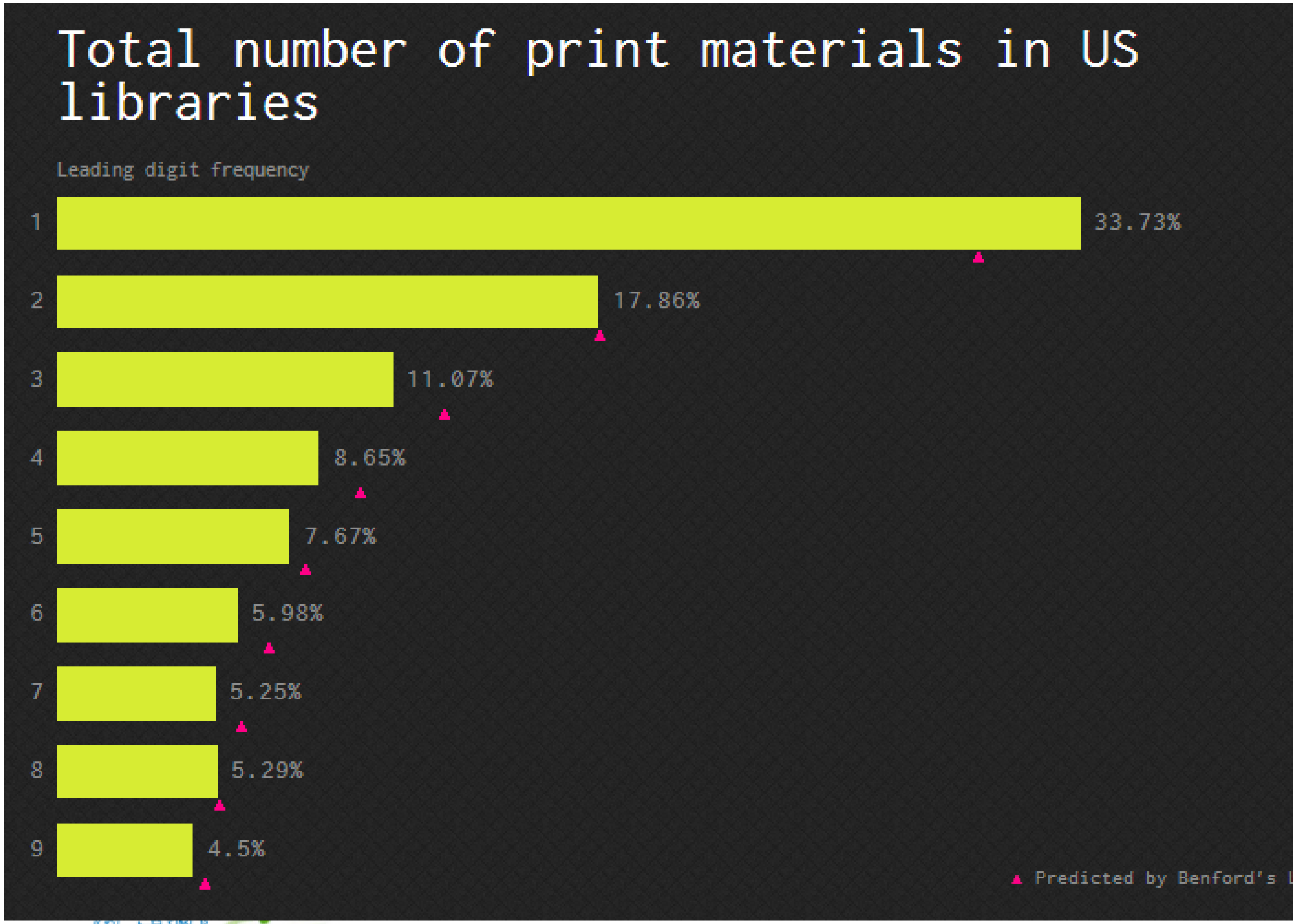
$$P(d) = \frac{Ln(1+\frac{1}{d})}{Ln(10)}$$



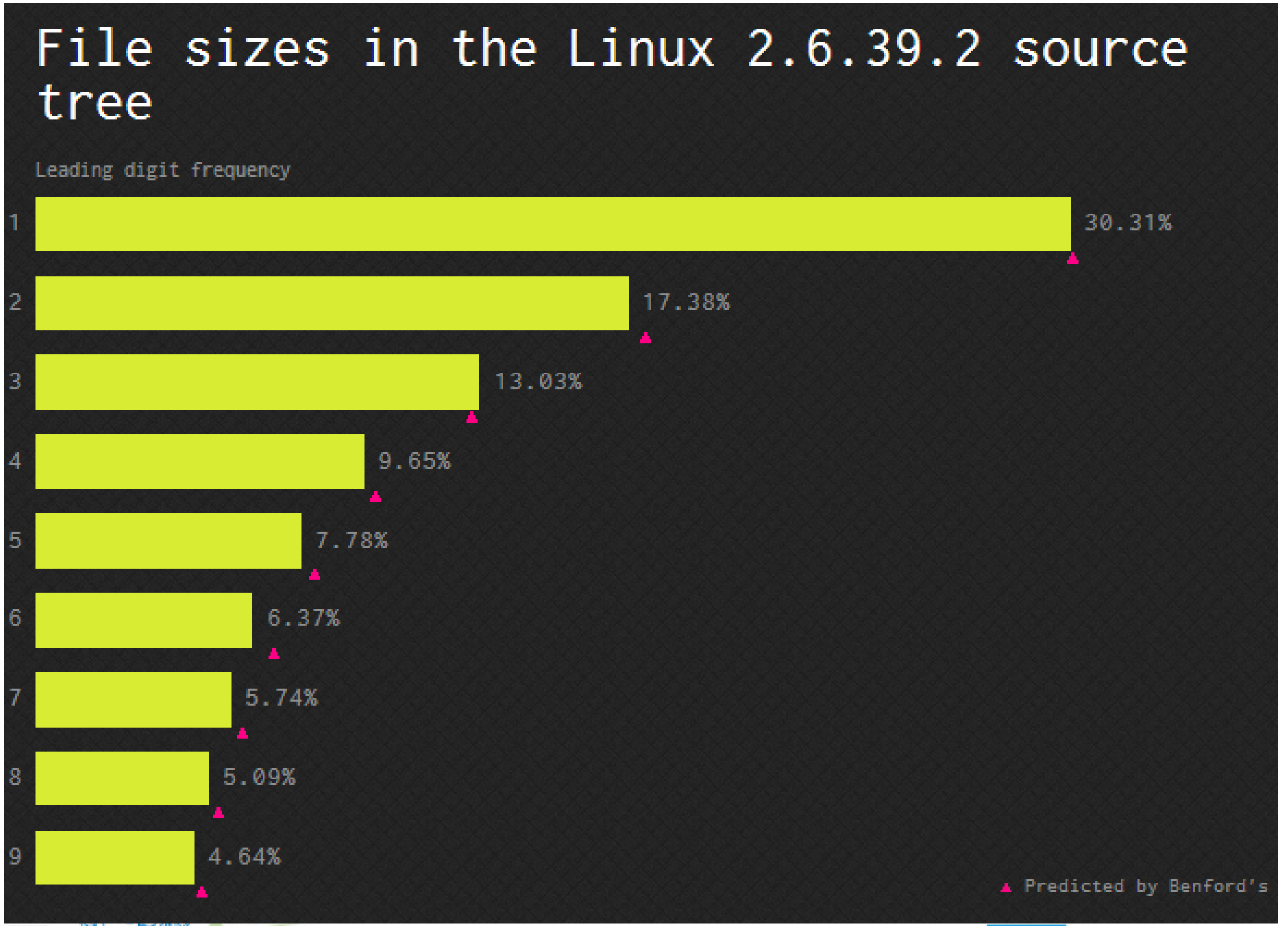
Benford's law first digit distribution



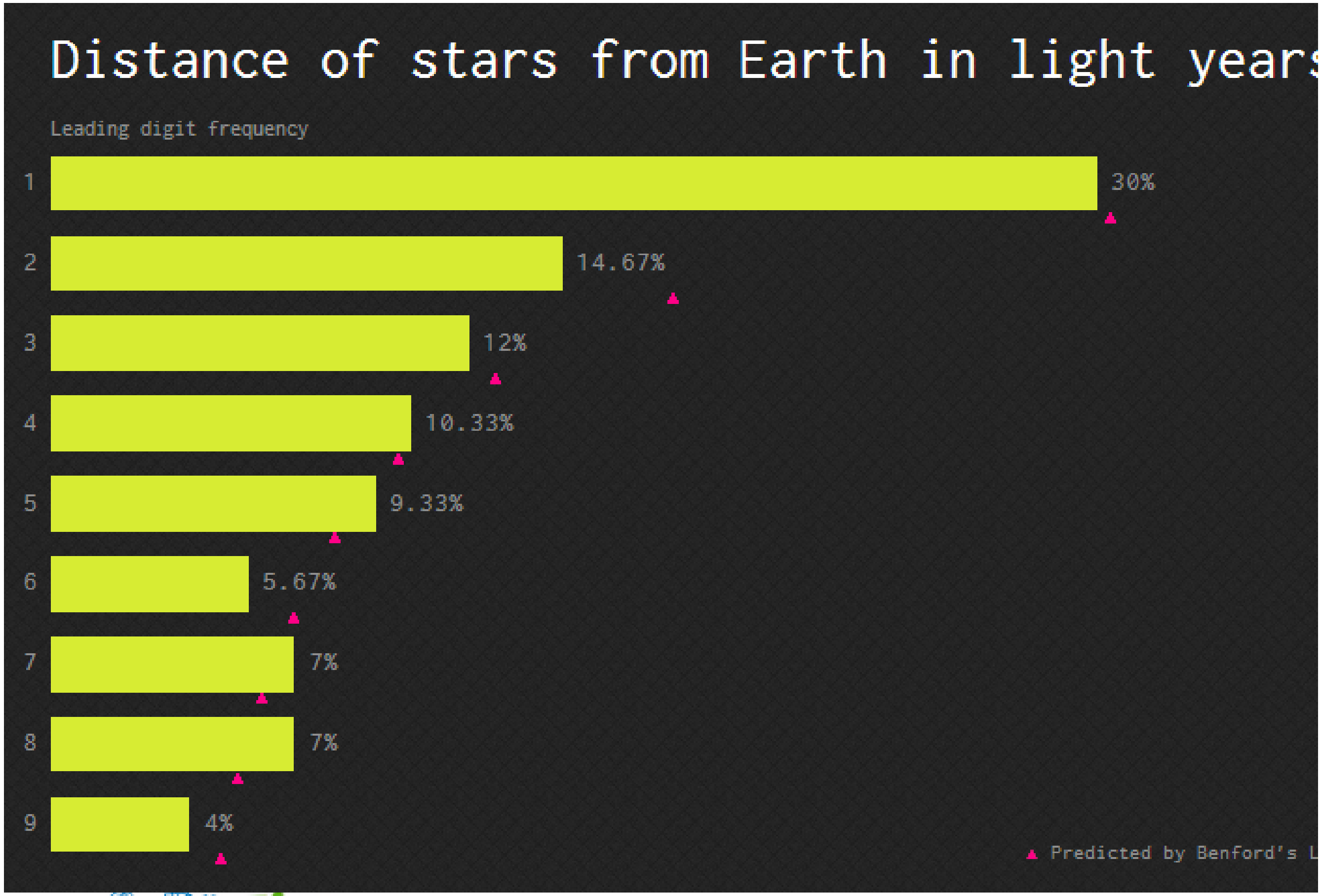
Benford's law
first digit distribution



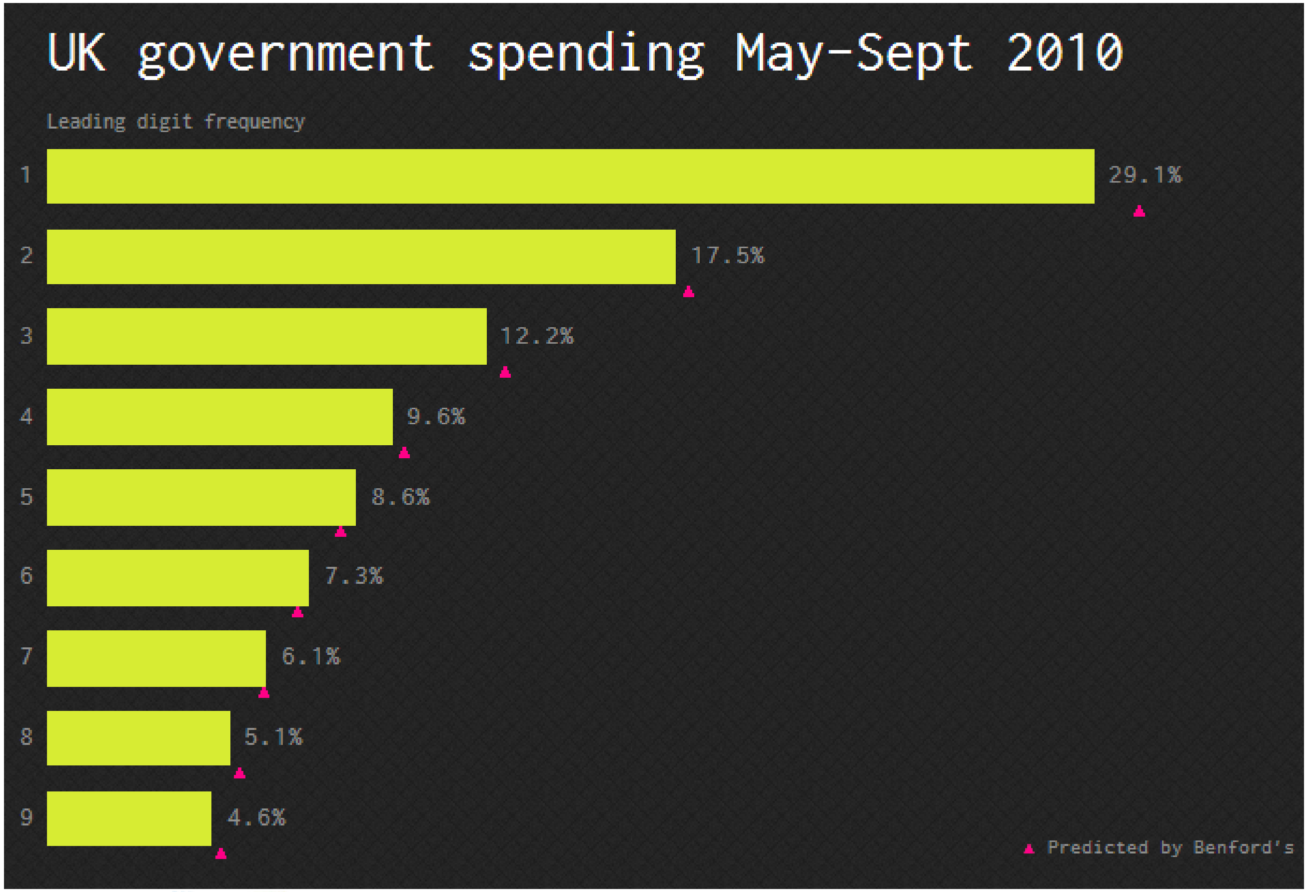
Benford's law
first digit distribution



Benford's law
first digit distribution



Benford's law
first digit distribution



Benford's law
first digit distribution

What does the Benford's law conceal about nature?

1………………..2……..3……….4…..5….6…7..8.9

P(1) = P(2,3) = P(4,5,6,7)

This means that the nature behaves in a logarithmic way

Benford's law
first digit distribution

Getting back to the calibration problems, it is better to choose a calibration point in the range:

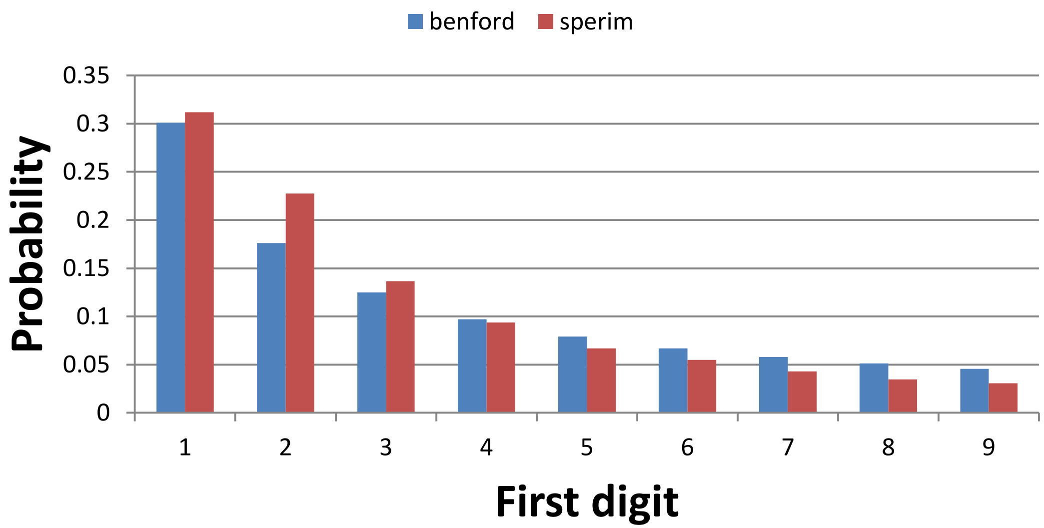
1 – 2μSv for the scale 0-10 μSv

10 – 20 μSv for the scale 0-100μSv

And so on

Benford's law
first digit distribution

Let us suppose to measure at one meter from a radiation source a doserate of 9.9 μSv/h. Let us measure up to 8 meters from the source in steps of 1 cm. According to the $1/r^2$ law the first digit is distributed according to the Benford law



Bibliography

- An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements by John R. Taylor
- Radiation detection and measurements Glenn Knoll
- ISO/IEC GUIDE 98-3:2008(E) Guide to the expression of uncertainty in measurement
- ISO 11929 (2010) Determination of the characteristic limits (decision threshold, detection limit and limits of the confidence interval) for measurements of ionizing radiation — Fundamentals and application