The Expectation Value of a Random Variable:

The ***expectation value*** $E[x]$ of a ***random variable*** x is the ***mean*** value of x, *i.e.* $\hat{x}$ (*aka* μ).

For ***discrete*** x_i, $E[x]$ is the ***sum*** of all x_i, weighted by their associated probabilities $P(x_i)$:

Discrete x_i: $$E[x] \equiv \hat{x} \equiv \mu \equiv \sum_{i=1}^{N} x_i P(x_i)$$

For ***continuous*** x, $E[x]$ is the ***integral*** over all x, weighted by the ***probability density function*** of x, $f(x)$:

Continuous x: $$E[x] \equiv \hat{x} \equiv \mu \equiv \int x\, f(x)\, dx$$

Note: $\hat{x}$ is ***not*** a ***random variable*** – it is a single, well-defined ***number*** that characterizes the ***true mean*** of the particular/specific distribution of a ***random variable***.

It is clear that for ***any*** constant, a:

Discrete: $$E[x+a] = \sum_{i=1}^{N}(x_i+a)P(x_i) = \sum_{i=1}^{N} x_i P(x_i) + a\underbrace{\sum_{i=1}^{N} P(x_i)}_{=1} = \hat{x}+a$$

Continuous: $$E[x+a] = \int (x+a)\, f(x)\, dx = \int x\, f(x)\, dx + a\underbrace{\int f(x)\, dx}_{=1} = \hat{x}+a$$

In fact, ***any*** function of x, $H(x)$ has an expectation value:

Discrete: $$E[H(x)] \equiv \hat{H}(x) \equiv \sum_{i=1}^{N} H(x_i)P(x_i)$$

Continuous: $$E[H(x)] \equiv \hat{H}(x) \equiv \int H(x)\, f(x)\, dx$$

The ***true mean***/expectation value $E[x] \equiv \hat{x} \equiv \mu$ is actually the ***first moment*** of the random variable x's probability distribution, ***relative to/taken about the origin*** $x = 0$.

The L^{th} ***moment*** of the ***random variable*** x taken about an ***arbitrary*** point $x = c$ is defined as:

$$\mu_L \equiv E[(x-c)^L]$$

n.b. Moments taken about $x = 0$ are known as ***algebraic*** moments.
Moments taken about the ***true mean*** $x = \mu$ are known as ***central*** moments.

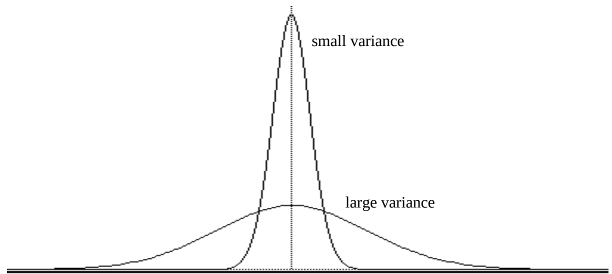
In particular, the 2^{nd} moment of x taken about/with respect to the ***true mean*** $\hat{x}$ is:

$$\mu_2 \equiv E[(x-\hat{x})^2] \equiv \mathrm{var}(x) \equiv \sigma_x^2$$ ≡ the ***variance*** of x.

The quantity $\sigma_x \equiv \sqrt{\mu_2} \equiv \sqrt{\mathrm{var}(x)} = \sqrt{\sigma_x^2}$ is known as the ***dispersion***, or ***standard deviation*** of x.

The ***true mean*** $\hat{x}$ tells us where (in x) the distribution is "important" – *i.e.* $\hat{x}$ tells us where the random variable x is most likely to be found).

The ***dispersion/standard deviation*** σ_x is a measure of the spread (or width) of the distribution over the space of x.

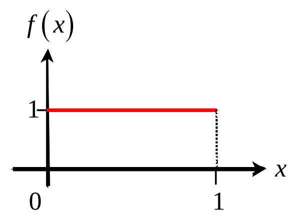


The two distributions in the above figure have the same ***expectation value***/***true mean*** $E[x]=\hat{x}=\mu$, but obviously have very different widths – and hence have different ***variances***.

Example of the Uniform Distribution *U*(0,1) in the Range $x = 0$ to $x = 1$:

The Uniform Probability Density Function (P.D.F.) $U(0,1)$ is:

$$f(x)=\begin{cases}1 & \text{if} \quad 0\le x<1\\ 0 & \text{otherwise}\end{cases}$$



This is correctly "normalized", *i.e.* $\int_0^1 f(x)\,dx=\int_0^1 1dx=\int_0^1 dx=1$.

(*i.e.* the area under the curve of $f(x)$ *vs.* x is = 1.)

The expectation value/***true mean*** of $U(0,1)$ is: $\hat{x}=\int_0^1 x\; f(x)\,dx=\int_0^1 x\,dx=\tfrac{1}{2}$

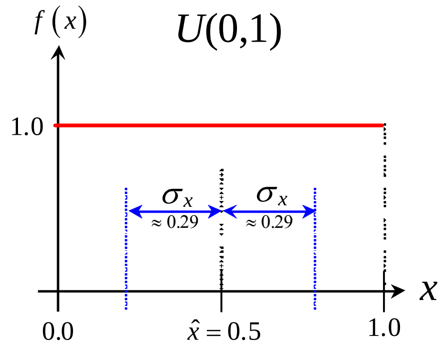
The ***variance*** associated with $U(0,1)$ is: $\operatorname{var}(x)=\sigma_x^2=\int_0^1(x-\hat{x})^2\; f(x)\,dx=\int_0^1\left(x-\tfrac{1}{2}\right)^2\; f(x)\,dx$

We can certainly calculate the above ***variance*** σ^2 for U(0,1) directly, however, in general it is frequently often easier to calculate, if the ***variance*** is first transformed in the following way:

$$\begin{aligned}\operatorname{var}(x)=\sigma_x^2&=\int(x-\hat{x})^2\; f(x)\,dx\\&=\int\left(x^2-2x\hat{x}+\hat{x}^2\right) f(x)\,dx\\&=\int x^2 f(x)\,dx-\int 2x\hat{x}\, f(x)\,dx+\int \hat{x}^2\; f(x)\,dx\\&=\int x^2 f(x)\,dx-2\hat{x}\int xf(x)\,dx+\hat{x}^2\int f(x)\,dx\\&=E[x^2]-2\hat{x}E[x]+\hat{x}^2=E[x^2]-2\hat{x}^2+\hat{x}^2=E[x^2]-\hat{x}^2\end{aligned}$$

$\therefore$ for the $U(0,1)$ distribution: $\boxed{\operatorname{var}(x)=\sigma_x^2=E[x^2]-E[x]^2=\overline{x^2}-\overline{x}^2}$

So for the $U(0,1)$ distribution:



$$\text{var}(x) = \sigma_x^2 = \int_0^1 x^2 dx - \left(\tfrac{1}{2}\right)^2 = \tfrac{1}{3} - \tfrac{1}{4} = \tfrac{1}{12}$$

$$\Rightarrow \; \sigma_x = \frac{1}{\sqrt{12}} \cong 0.29$$

This means that if we perform a measurement of a quantity x that results in the knowledge that it can be ***anywhere*** (with equal likelihood/equal/flat probability) within the range (0,1), our best estimate of its location is that is at the true mean, $x = \hat{x} = 0.5$ with a standard deviation of $\sigma_x = 1/\sqrt{12} \cong 0.29$.

More generally, for the flat/uniform distribution $U(0,L) = 1/L$:

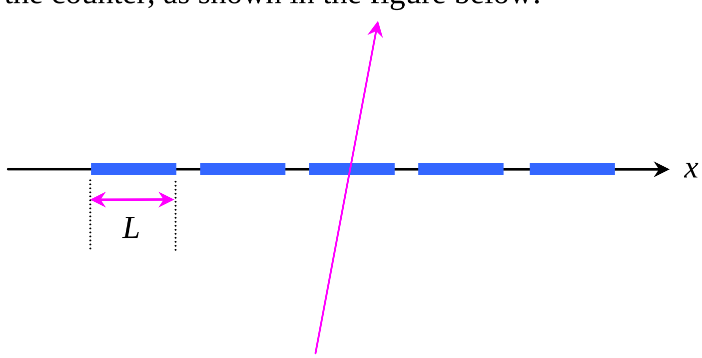
$$\text{mean: } \hat{x} = \frac{L}{2} \qquad \text{variance: } \sigma_x^2 = \frac{L^2}{12} \qquad \Rightarrow \quad \text{standard deviation: } \sigma_x = \frac{L}{\sqrt{12}}$$

If the allowed range of x is $a \le x \le b$, it is also easy to show that $U(a,b) = 1/(b-a)$, and that:

$$\text{mean: } \hat{x} = \frac{b+a}{2} \qquad \text{variance: } \sigma_x^2 = \frac{(b-a)^2}{12} \quad \Rightarrow \quad \text{standard deviation: } \sigma_x = \frac{b-a}{\sqrt{12}}$$

<u>Example</u> – <u>The HEP Hodoscope</u>:

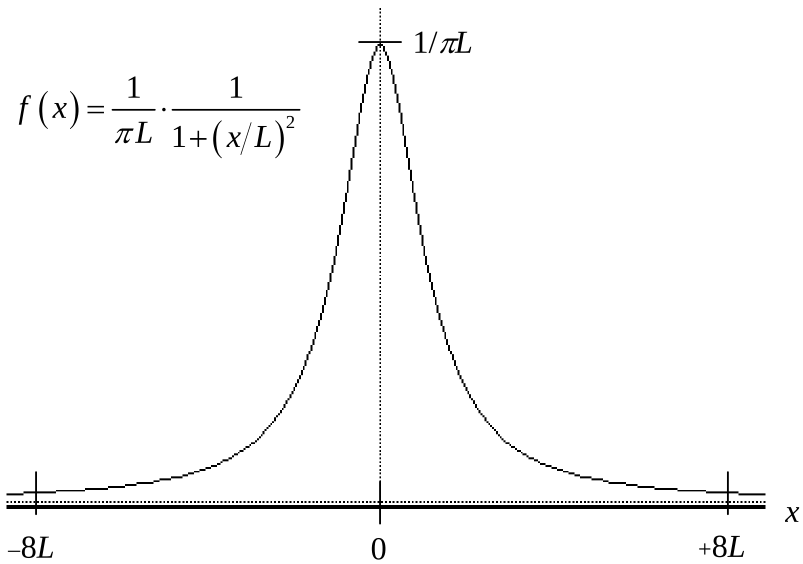
In particle physics experiments, a planar array of scintillation counters ("hodoscope") is often used to detect the passage of ionizing (charged) particles. If a particular counter is hit (and that is all that on one knows), then this information can be considered a measurement of (say) the x coordinate of the trajectory at the plane of the hodoscope. The result of the measurement is that $\hat{x}$ is the x-coordinate of the *center* of the particular counter that was hit by the particle. The standard deviation of the measurement is $\sigma = L/\sqrt{12} \approx 0.29L$, where L is the lateral width (in the x-direction) of the counter, as shown in the figure below:



This example is also relevant for determining the uncertainty in ADC count data in the regime of noise-free data – *e.g*. a 12-bit ADC measures quasi-DC voltages over a ±5.0 volt range. 2^{12} bits = 4096. Note that: 10 Volts/4096 bits = 2.44 *mV*/ADC count.

For noise-free {or nearly noise-free} ADC data, in the absence of using so-called ADC ***dithering techniques*** (adding and systematically {or randomly} varying a small voltage offset), then $\sigma_{ADC} = 1/\sqrt{12}$ ADC counts, corresponding to $\sigma_V = \left(1/\sqrt{12}\right)\cdot 2.44mV = 0.70mV$.

<u>Example</u> – <u>The Cauchy Distribution</u>: $f(x) = \dfrac{1}{\pi L}\cdot\dfrac{1}{1+(x/L)^2}$



A careful check shows that: $\int_{-\infty}^{+\infty} f(x)\,dx = 1 \quad \Leftarrow$ **OK**

but: $\hat{x} = \int_{-\infty}^{+\infty} x\, f(x)\,dx = \frac{L}{2\pi}\ln(x^2+L^2)\Big|_{-\infty}^{+\infty} \quad \Leftarrow$ **Undefined !!!**

Obviously, the expectation value of *x* is/should be $\hat{x}$ = 0. Hmmm…
There are mathematical complications with this P.D.F. !!!

Using $\hat{x}$ = 0, we can calculate the ***variance*** of the Cauchy Distribution (using a table of integrals):

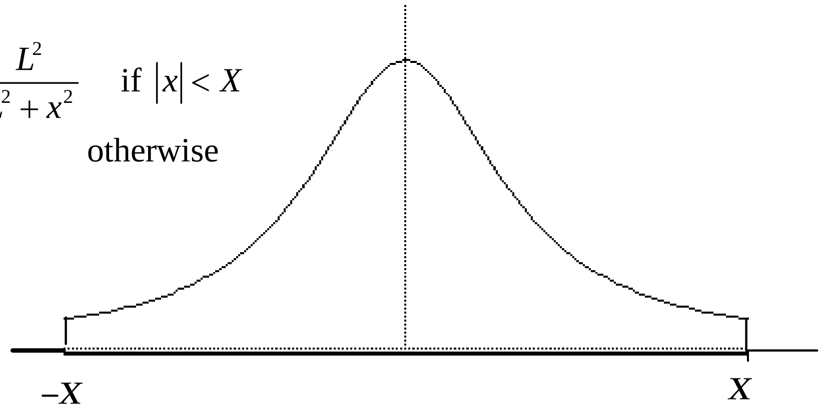
$$\text{var}(x) = \sigma^2 = \int_{-\infty}^{+\infty} x^2 f(x)\,dx = \frac{L}{\pi}\left\{ x\Big|_{-\infty}^{+\infty} - L^2\int_{-\infty}^{+\infty}\frac{dx}{L^2+x^2}\right\} \quad \Leftarrow \quad \textbf{Infinite !!!}$$

⇑ **infinite** ⇑ **finite**

More mathematical problems…!

In order to avoid these difficulties, in practice a ***truncated*** Cauchy distribution is used:

$$f(x) = \begin{cases} c\dfrac{L^2}{L^2+x^2} & \text{if } |x| < X \\ 0 & \text{otherwise} \end{cases}$$



The requirement that: $\int_{-\infty}^{+\infty} f(x)\,dx = 1$ requires: $c = \dfrac{1}{2L\tan^{-1}(X/L)}$.

(Note that for $X \to \infty$, $c \to 1/L\pi$ as we expect.)

Now: $\hat{x} = \int_{-X}^{+X} x\, f(x)\,dx = \dfrac{L}{2\pi}\ln\left(x^2+L^2\right)\Big|_{-X}^{+X} = 0$ and $\operatorname{var}(x) = \sigma^2$ is finite as well.

More About the Moments of a Probability Distribution Function / P.D.F.

The 1st ***central*** moment (*i.e.* taken about the ***true mean*** $\hat{x} = \mu$) is ***not*** interesting, since:

$$\mu_1 \equiv E[x-\hat{x}] = E[x] - E[\hat{x}] = \hat{x} - \hat{x} = 0$$

Higher moments of a probability distribution are ***sometimes*** useful.

For example – the ***coefficients of skewness*** and ***kurtosis***, (Gr. "***kurtos***" = "bulging/swelling") respectively are the 3rd and 4th standardized ***central*** moments of a distribution (*i.e.* also taken about the ***true mean*** $\hat{x} = \mu$):

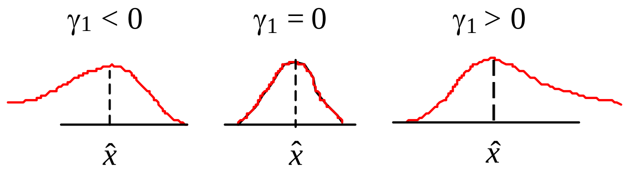
Skewness: $$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{E[(x-\hat{x})^3]}{\{\underbrace{E[(x-\hat{x})^2]}_{=\sigma^2}\}^{3/2}} = \frac{E[(x-\hat{x})^3]}{\sigma^3}$$

Excess Kurtosis: $$\gamma_2 = \frac{\mu_4}{\mu_2^{4/2}} - 3 = \frac{E[(x-\hat{x})^4]}{\{\underbrace{E[(x-\hat{x})^2]}_{=\sigma^2}\}^{4/2}} - 3 = \frac{E[(x-\hat{x})^4]}{\sigma^4} - 3$$

The –3 in the ***excess*** kurtosis definition is added so that $\gamma_2 = 0$ for a Gaussian (*aka* normal) probability distribution. Without the –3, "***regular/normal***" kurtosis is defined as $\mu_4 / \mu_2^{4/2}$.

Skewness γ_1 tells us whether the P.D.F. $f(x)$ is:

$\gamma_1 < 0$: skewed toward $x < \hat{x}$
$\gamma_1 = 0$: symmetric about $\hat{x}$
$\gamma_1 > 0$: skewed toward $x > \hat{x}$

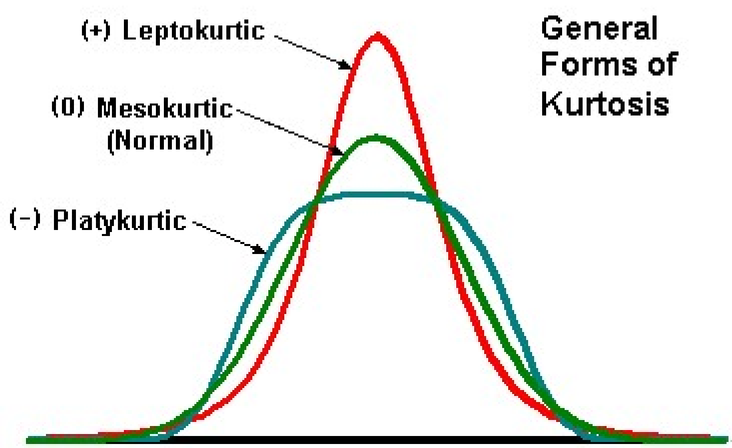


Obviously, the "long tail" dominates in $E[(x-\hat{x})^3]$.

A probability distribution (*e*.**g**. the Gaussian/normal distribution – see below…) that has ***zero*** kurtosis $(\gamma_2 = 0)$ is called ***mesokurtic***, or mesokurotic (Gr. "***mesos***" = "middle").

Positive kurtosis ($\gamma_2 > 0$) indicates a peaked/slender/narrow probability distribution near the mean with abnormally ***long*** tails, and is called ***leptokurtic***, or ***leptokurotic*** (Gr. "***leptos***" = "slender/thin").

Negative kurtosis ($\gamma_2 < 0$) indicates a flattened/wide/broad probability distribution near the mean with abnormally ***shortened*** tails, and is called ***platykurtic***, or ***platykurotic*** (Gr. "***platys***" = "broad/flat").



Stock market investors are ***very*** interested in these particular higher moments (skewness & kurtosis) of various probability distribution functions…

<u>Important Note:</u>

<u>Each/every</u> measurement of a random variable x contains ***<u>information</u>*** about the PDF $f(x)$ from which it originates – and its expectation value/true mean $\hat{x}$, variance $\text{var}(x) = \sigma_x^2$, and the PDF's higher order moments…

In the limit of an ***infinite*** # of measurements $N \to \infty$ of the random variable *x*, the PDF $f(x)$ and all of its associated moments are ***precisely*** known…

Any random variable x can be "normalized" or "standardized" and can thus be reduced to a ***dimensionless*** quantity.

Suppose that the ***random variable*** x has P.D.F. $f(x)$ and has a well-defined ***expectation value / true mean*** $\hat{x}$ and ***variance*** σ_x^2.

Consider a ***new random variable*** defined as: $u(x) \equiv \dfrac{(x-\hat{x})}{\sigma_x} \Leftarrow$ *n.b.* dimensionless quantity

Aside: in general, if c is a ***constant***, then: $E[cg(x)] = cE[g(x)]$.

Thus: $E[u(x)] = \hat{u} = \int u(x)\, f(x)\,dx = \int \dfrac{x-\hat{x}}{\sigma_x} f(x)\,dx = \dfrac{1}{\sigma_x}\int (x-\hat{x}) f(x)\,dx = 0$, *i.e.* $\hat{u} = 0$

and:

$$\mathrm{var}(u) = \int (u-\hat{u})^2\, f(x)\,dx = \int u^2(x)\, f(x)\,dx - 2\hat{u}\underbrace{\int u(x)\, f(x)\,dx}_{=\hat{u}} + \hat{u}^2\int f(x)\,dx = \int u^2 f(x)\,dx$$

$$= \frac{1}{\sigma_x^2}\underbrace{\int (x-\hat{x})^2\, f(x)\,dx}_{\equiv \mathrm{var}(x)=\sigma_x^2} = \frac{1}{\sigma_x^2}\sigma_x^2 = 1$$

Thus, the ***random variable*** $u(x) \equiv (x-\hat{x})/\sigma_x$ has:

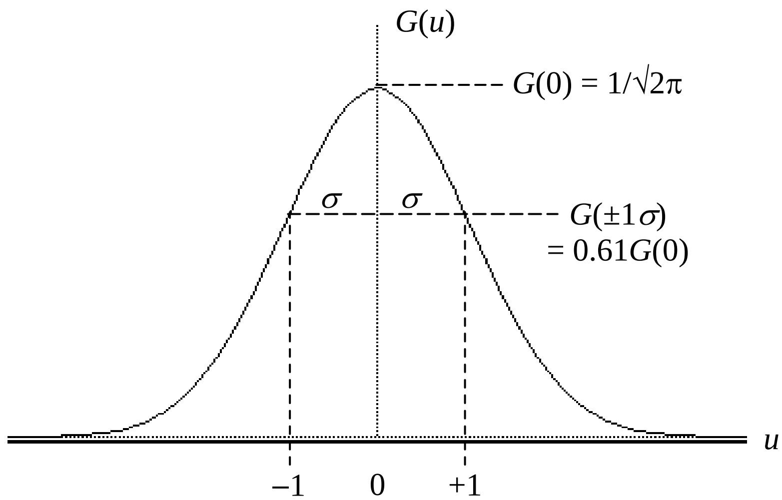
Expectation value/true mean $E[u(x)] = \hat{u} = 0$

Variance $\mathrm{var}(u) = \sigma_u^2 = 1$, hence standard deviation $\sigma_u = 1$

<u>The 1-D Gaussian (*aka* Normal) Distribution</u>:

One of the most common P.D.F.'s that is encountered in the everyday world of doing experimental physics is the Gaussian (*aka* normal) probability distribution function (P.D.F.):

$$\boxed{G(u) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-u^2/2\sigma^2}} \quad \text{with: } \boxed{\sigma = 1}$$



It is easily verified that if u has the above P.D.F., then the ***expectation value***/***true mean*** of the Gaussian/normal distribution is $E[u] = 0$ (by symmetry) and its standard deviation, $\sigma_u = 1$.

In terms of probability, the probability that u will be found in an interval du is:

$$P = \begin{cases} \dfrac{1}{\sqrt{2\pi}} du & \text{at } u = 0 \\ \dfrac{1}{\sqrt{2\pi}} e^{-1/2} du & \text{at } u = \pm\sigma \end{cases}$$

The ratio $G(\pm\sigma)/G(0)$ is $e^{-1/2} = 0.6065... \approx 0.61$

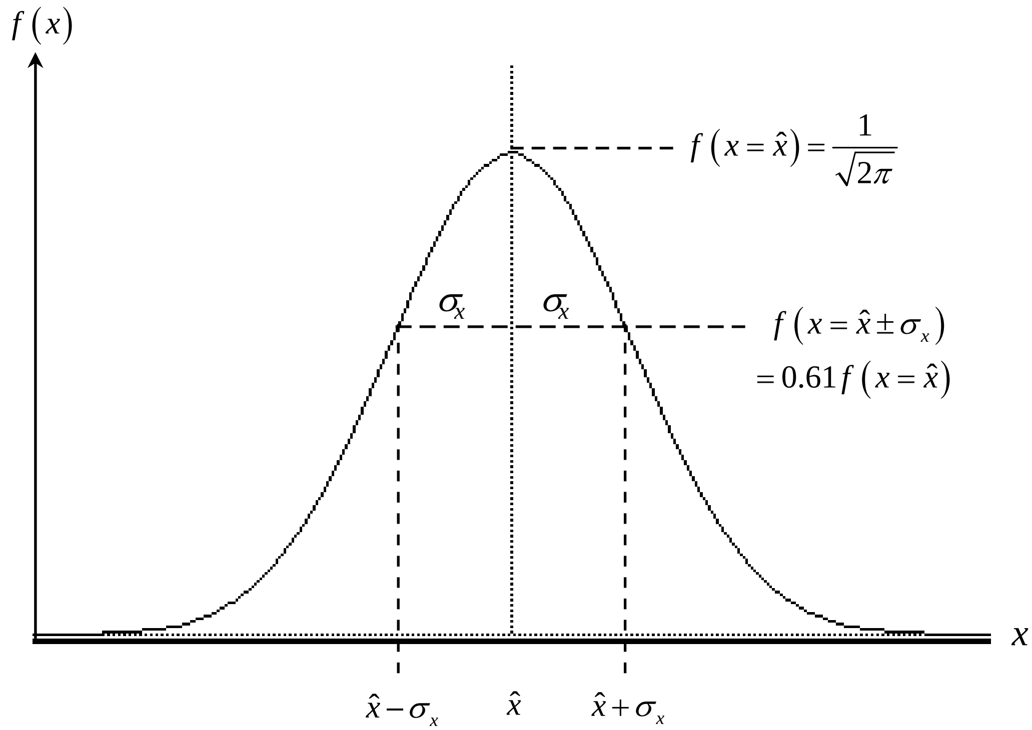
Later we shall show ***why*** the normal distribution is so universal in experimental physics…

In many situations in experimental physics, we need to describe a normally distributed ***random variable*** x with ***expectation value*** / ***true mean*** $\hat{x}$ and standard deviation σ_x.

We can also ***invert*** the transformation $u = \dfrac{x - \hat{x}}{\sigma_x}$:

The P.D.F. of x satisfies: $f(x)\,dx = G(u)\,du = \dfrac{1}{\sqrt{2\pi}} e^{-(x-\hat{x})^2/2\sigma_x^2} \dfrac{dx}{\sigma_x}$

i.e. $\boxed{f(x) = \dfrac{1}{\sqrt{2\pi}\,\sigma_x} e^{-(x-\hat{x})^2/2\sigma_x^2}}$ ⇐ Shifted Gaussian distribution, centered at $x = \hat{x}$ with standard deviation σ_x.



The *N*-D Probability Distributions Functions, *N*-D Cumulative Distribution Functions:

In physics experiments, we often measure more than one property of a system (*e.g*. the three spatial coordinates of a particle's position, its 3-momentum (or velocity), the energy and/or lifetime of an excited atomic state, *etc*.) An experiment may then result in k different ***random variables***, which can be conveniently plotted on k mutually orthogonal axes (*i.e*. a mathematical hyper-space consisting of k orthogonal dimensions) and treated as if they were the components of a k-dimensional vector.

First, let us consider only ***two*** such ***random variables*** x and y.

The 2-D ***Cumulative*** Distribution Function: $F(X,Y) \equiv Prob(x < X \text{ .and. } y < Y)$

The 2-D Probability Density Function: $f(x,y) \equiv \frac{\partial}{\partial x}\frac{\partial}{\partial y} F(x,y)$

If the ***random variables*** x and y are ***independent***, then $f(x,y)$ is ***separable*** in x & y, *i.e*. $f(x,y) = f_x(x) \cdot f_y(y) \Leftarrow$ x & y have no ***correlations*** (*i.e*. are ***uncorrelated***).

Then: $\boxed{P(a \le x < b,\ c \le y < d) = \int_a^b dx \int_c^d dy\ f(x,y)}$ with the requirement: $\boxed{\int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy\ f(x,y) = 1}$

where the x, y integrals run over the allowed values of x and y: $-\infty \le a \le x \le b \le +\infty$
and: $-\infty \le c \le y \le d \le +\infty$.

In certain experimental situations, we may not care (or even know) what y is, hence we integrate y over its entire range and get: $\boxed{g(x)=\int_{-\infty}^{+\infty} f(x,y)\,dy}$.

In this situation: $\boxed{P(a \le x < b,\ -\infty \le y < +\infty) = \int_a^b g(x)\,dx}$.

Thus we see that $g(x)$ is ***also*** a P.D.F.; $g(x)$ is known as the ***marginal distribution*** of x.

Change of k Random Variables:

Suppose we have k ***random variables*** $(x_1, x_2, \ldots, x_k)$, with corresponding P.D.F. $f(x_1, x_2, \ldots, x_k)$.

We may wish (or need) to make a change variables from $(x_1, x_2, \ldots, x_k)$ to $(y_1, y_2, \ldots, y_k)$, where ***each*** of the $y_i = y_i(x_1, x_2, \ldots, x_k)$.

For the case of a ***single random variable*** x, for a change of variables $y = H(x)$, from P598AEM Lecture Notes 2, *p*. 8, the relation $g(y)\,dy = f(x)\,dx$ leads to the relation (*here*):

$$\boxed{g(y)=\frac{f(x)}{|H(x)|}} \quad \text{where: } |H(x)| = |dy/dx| = |y\text{-slope}|$$

← ***n.b.*** absolute value

For k ***random variables***, this generalizes to:

$$\boxed{g(y_1, y_2, \cdots, y_k) = \frac{f(x_1, x_2, \ldots, x_k)}{|J|}}$$

← ***n.b.*** absolute value

where J is the ***Jacobian***:

$$J\left(\frac{y_1 y_2 \ldots y_k}{x_1 x_2 \ldots x_k}\right) \equiv \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_k} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_k}{\partial x_1} & \frac{\partial y_k}{\partial x_2} & \cdots & \frac{\partial y_k}{\partial x_k} \end{vmatrix}$$

⇐ ***n.b.*** The Jacobian ***J*** is the ***determinant*** of the $k \times k$ matrix of derivatives $\partial y_i / \partial x_j$

i.e. compute the individual entries of the $k \times k$ matrix of partial derivatives $\partial y_i / \partial x_j$, then compute the determinant of this matrix, then take the absolute value of the determinant.