For ***any*** function $g(x,y)$ of two random variables *x* and *y* whose (joint) P.D.F. is $f(x,y)$ the ***expectation value/true mean*** of $g(x,y)$ is:

$$E[g(x,y)] = \int\int g(x,y)\, f(x,y)\, dxdy = \int dx \int dy\, g(x,y)\, f(x,y) = \int dy \int dx\, g(x,y)\, f(x,y)$$

$$E[x] \equiv \hat{x} = \int x \underbrace{\left\{\int f(x,y)\, dy\right\}}_{\text{marginal distribution of } x,\ g(x)} dx = \int x g(x)\, dx$$

$$E[y] \equiv \hat{y} = \int y \underbrace{\left\{\int f(x,y)\, dx\right\}}_{\text{marginal distribution of } y,\ h(y)} dy = \int y h(y)\, dy$$

The individual ***variances*** of the two random variables *x* and *y* are:

$$\mathrm{var}(x) \equiv \sigma_x^2 \equiv E[(x-\hat{x})^2] = E[(x-\hat{x})(x-\hat{x})]$$

$$\mathrm{var}(y) \equiv \sigma_y^2 \equiv E[(y-\hat{y})^2] = E[(y-\hat{y})(y-\hat{y})]$$

We also define a new quantity, known as the ***covariance*** of the two random variables *x* and *y*:

$$\begin{aligned}
\mathrm{cov}(x,y) \equiv \sigma_{xy}^2 &\equiv E[(x-\hat{x})(y-\hat{y})] \\
&= E[xy - \hat{x}y - x\hat{y} + \hat{x}\hat{y}] \\
&= E[xy] - \hat{x}E[y] - E[x]\hat{y} + \hat{x}\hat{y} \\
&= E[xy] - \hat{x}\hat{y} - \hat{x}\hat{y} + \hat{x}\hat{y} = E[xy] - 2\hat{x}\hat{y} + \hat{x}\hat{y} \\
&= E[xy] - \hat{x}\hat{y} \\
&= E[xy] - E[x]\,E[y]
\end{aligned}$$

Often, we also use the ***coefficient of correlation*** (*aka* the ***correlation coefficient***), $\rho(x,y)$:

$$\rho(x,y) \equiv \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y}$$

We show that ***magnitude*** of the coefficient of correlation $\rho(x,y) \equiv \mathrm{cov}(x,y)/\sigma_x\sigma_y$ ***cannot*** exceed unity (1). For any two numbers α, β (constants):

$$\begin{aligned}
\mathrm{var}(\alpha x+\beta y) &= E\left[(\alpha x+\beta y - E[\alpha x+\beta y])^2\right] \\
&= E\left[(\alpha x+\beta y-\alpha\hat{x}-\beta\hat{y})^2\right] \\
&= E\left[\{\alpha(x-\hat{x})+\beta(y-\hat{y})\}^2\right] \\
&= E\left[\alpha^2(x-\hat{x})^2+\beta^2(y-\hat{y})^2+2\alpha\beta(x-\hat{x})(y-\hat{y})\right] \\
&= \alpha^2 \underbrace{E[(x-\hat{x})^2]}_{\equiv \mathrm{var}(x)=\sigma_x^2} + \beta^2 \underbrace{E[(y-\hat{y})^2]}_{\equiv \mathrm{var}(y)=\sigma_y^2} + 2\alpha\beta \underbrace{E[(x-\hat{x})(y-\hat{y})]}_{\equiv \mathrm{cov}(x,y)} \\
&= \alpha^2\sigma_x^2+\beta^2\sigma_y^2+2\alpha\beta\,\mathrm{cov}(x,y) \\
&\geq 0
\end{aligned}$$

Since $\mathrm{var}(anything) = \int (anything)^2 f(x,y)\,dx\,dy$ and (by definition) the P.D.F. $f(x,y) \geq 0$.

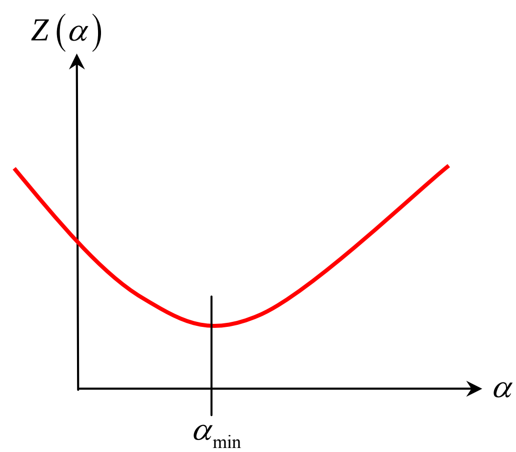
This ***must*** hold for ***any*** arbitrary choice of α and β, so *e.g.* pick $\beta = 1$ and look at the quantity:

$$Z(\alpha) = \alpha^2\sigma_x^2 + \sigma_y^2 + 2\alpha\,\mathrm{cov}(x,y) \geq 0$$

Now $Z(\alpha) = \sigma_x^2\alpha^2 + 2\,\mathrm{cov}(x,y)\,\alpha + \sigma_y^2$ is the equation of a parabola $\left(y(x) = ax^2+bx+c\right)$ whose minimum occurs at a value of α given by the solution of

$$\frac{dZ(\alpha)}{d\alpha} = 0 = 2\sigma_x^2\alpha + 2\,\mathrm{cov}(x,y),$$

which yields $\alpha_{\min} = -\mathrm{cov}(x,y)/\sigma_x^2$.



The minimum value of $Z\left(\alpha_{\min} = -\mathrm{cov}(x,y)/\sigma_x^2\right)$ is:

$$Z(\alpha_{\min}) = \sigma_x^2\left(\frac{\mathrm{cov}(x,y)}{\sigma_x^2}\right)^2 - 2\frac{\mathrm{cov}(x,y)^2}{\sigma_x^2} + \sigma_y^2 = \frac{\mathrm{cov}(x,y)^2}{\sigma_x^2} - 2\frac{\mathrm{cov}(x,y)^2}{\sigma_x^2} + \sigma_y^2 = -\frac{\mathrm{cov}(x,y)^2}{\sigma_x^2} + \sigma_y^2 \geq 0$$

Thus: $\sigma_y^2 - \dfrac{\mathrm{cov}(x,y)^2}{\sigma_x^2} \geq 0$ or: $|\mathrm{cov}(x,y)| \leq \sigma_x\sigma_y$ or: $\dfrac{|\mathrm{cov}(x,y)|}{\sigma_x\sigma_y} \leq 1$.

In terms of $\rho(x,y) \equiv \dfrac{\mathrm{cov}(x,y)}{\sigma_x\sigma_y}$, this becomes: $-1 \leq \rho(x,y) \equiv \dfrac{\mathrm{cov}(x,y)}{\sigma_x\sigma_y} \leq +1$ {Q.E.D.}.

Now suppose that the two random variables *x* and *y* are ***independent*** random variables**.**

In this case, the joint P.D.F. satisfies: $f(x,y) = f_x(x)\cdot f_y(y)$.

We show that $\text{cov}(x,y) = 0$ for ***independent*** random variables.

Recall that: $\text{cov}(x,y) = E[xy] - E[x]\,E[y]$.

$$E[xy] = \int dx \int dy\; x\, y\; f(x,y) = \int dx \int dy\; x\, y\; f_x(x)\cdot f_y(y)$$
$$= \int x\; f_x(x)\,dx \cdot \int y\; f_y(y)\,dy$$

$$E[x] = \int dx \int dy\; x\; f(x,y) = \int dx \int dy\; x\; f_x(x)\cdot f_y(y)$$
$$= \int x\; f_x(x)\,dx \cdot \int f_y(y)\,dy \quad \text{but:} \quad \int f_y(y)\,dy = 1 \quad \text{by normalization}$$

$$\therefore\; E[x] = \int x\; f_x(x)\,dx \quad \text{and similarly:} \quad E[y] = \int y\; f_y(y)\,dy$$

Thus: $$E[xy] = \int x\; f_x(x)\,dx \cdot \int y\; f_y(y)\,dy = E[x]\,E[y]$$

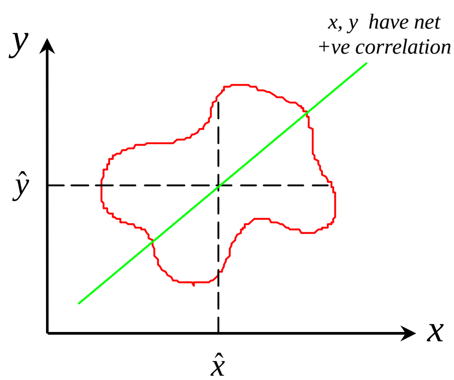
and thus: $\text{cov}(x,y) = E[xy] - E[x]\,E[y] = E[x]\,E[y] - E[x]\,E[y] = 0$ {Q.E.D.}.

Thus (here), the correlation coefficient $\rho(x,y) \equiv \dfrac{\text{cov}(x,y)}{\sigma_x\sigma_y} = \dfrac{E[xy] - E[x]\,E[y]}{\sigma_x\sigma_y} = 0$

Conclusion: *independent* random variables are ***uncorrelated*** !

Although this is obvious, the ***reverse*** is ***not*** necessarily true, *i.e.* it ***is*** possible for the correlation coefficient $\rho(x,y)$ to be ***zero*** even if *x* and *y* are ***not*** independent !

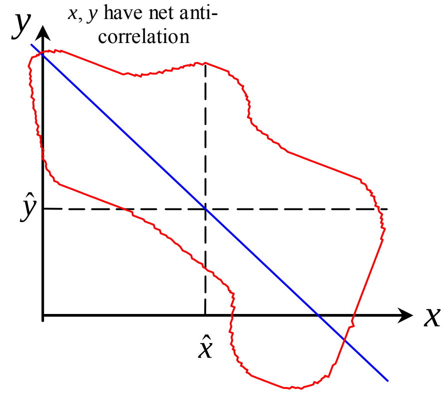
From the definition of covariance: $\text{cov}(x,y) \equiv E[(x-\hat{x})(y-\hat{y})] = E[xy] - E[x]\,E[y]$, we see that the $\text{cov}(x,y)$ will be ***positive*** if the values of ***both*** *x* and *y* tend to be larger (or smaller) than their expectation values $\hat{x},\ \hat{y}$ – here, the random variables *x* and *y* ***tend*** to be ***correlated*** with each other:



In this case:
Here we imagine the random variables *x* and *y* to be distributed ***uniformly*** within the **red** boundary.

There is more area in the 1st & 3rd "quadrants" ($x > \hat{x}$ and $y > \hat{y}$) .and. ($x < \hat{x}$ and $y < \hat{y}$), and thus as a result, $\text{cov}(x,y) > 0$.

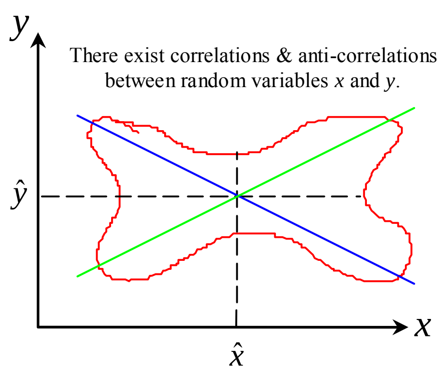
On the other hand, the following sketch shows a situation where there is more area in the 2nd & 4th "quadrants" ($x > \hat{x}$ and $y < \hat{y}$) .and. ($x < \hat{x}$ and $y > \hat{y}$), and thus in this case $\text{cov}(x, y) < 0$, *i.e.* the random variables *x* and *y* ***tend*** to be ***anti-correlated*** with each other:



In this case:
$$\text{cov}(x,y) \equiv E[(x-\hat{x})(y-\hat{y})]$$
$$= E[xy] - E[x]\,E[y]$$
will be ***negative*** .

For the case of where there is ~ equal area in each of the four "quadrants" ($x > \hat{x}$ and $y > \hat{y}$) .and. ($x < \hat{x}$ and $y < \hat{y}$) .and. ($x > \hat{x}$ and $y < \hat{y}$) .and. ($x < \hat{x}$ and $y > \hat{y}$), as shown in the following sketch, this is a situation where $\text{cov}(x, y) \sim 0$, *i.e.* the random variables *x* and *y* tend to have no ***net*** correlations with each other:

In this case:
$$\text{cov}(x,y) \equiv E[(x-\hat{x})(y-\hat{y})]$$
$$= E[xy] - E[x]\,E[y]$$
$$\sim 0$$
although *x* and *y* are obviously ***not*** independent.



Let us investigate an algebraic example:

Suppose *x* is a random variable that is distributed ***symmetrically*** about *x* = 0.

Then: $E[x] = \hat{x} = \int x\, f(x)\, dx = 0$.

Suppose: $y = x^2$. Then: $E[y] = E[x^2] = E[x^2] - E[x]^2 = \text{var}(x) = \sigma_x^2$.

Since $f(x)$ is an ***even*** function of *x* (*i.e.* it is distributed ***symmetrically*** about *x* = 0):

$$\text{cov}(x,y) \equiv E[(x-\hat{x})(y-\hat{y})] = E[xy] - \underbrace{E[x]}_{=0}\,E[y] = E[x^3] = \int x^3 f(x)\, dx = 0$$

Thus *x* and *y* are ***dependent*** random variables, but have no ***net*** correlation. Within a given "quadrant", *x* and *y* will have a net correlation (1st & 3rd) or a net anti-correlation (2nd & 4th).

Consequences of Correlations:

Suppose two random variables x and y are known to have expectation values $\hat{x}$ and $\hat{y}$ and have standard deviations σ_x and σ_y, respectively, where:

$$E[x]=\hat{x},\quad E[y]=\hat{y},\quad E[(x-\hat{x})^2]=\text{var}(x)=\sigma_x^2,\quad E[(y-\hat{y})^2]=\text{var}(y)=\sigma_y^2$$

Let: $A=ax+by$ where a and b are constants. Then: $E[A]=\hat{A}=a\hat{x}+b\hat{y}$, and the variance of A:

$$\begin{aligned}\text{var}(A)=\sigma_A^2&=E[(A-\hat{A})^2]=E[A^2]-E[A]^2\\&=E[a^2x^2+2ab\,xy+b^2y^2]-\left(a^2\hat{x}^2+2ab\,\hat{x}\hat{y}+b^2\hat{y}^2\right)\\&=a^2\underbrace{\left(E[x^2]-\hat{x}^2\right)}_{=\text{var}(x)=\sigma_x^2}+2ab\underbrace{\left(E[xy]-\hat{x}\hat{y}\right)}_{=\text{cov}(x,y)}+b^2\underbrace{\left(E[y^2]-\hat{y}^2\right)}_{=\text{var}(y)=\sigma_y^2}\end{aligned}$$

Thus: $\text{var}(A)=\sigma_A^2=a^2\sigma_x^2+2ab\,\text{cov}(x,y)+b^2\sigma_y^2$

We can alternatively express this result in terms of the correlation coefficient:

$$\rho(x,y)\equiv\frac{\text{cov}(x,y)}{\sigma_x\sigma_y}$$

$$\text{var}(A)=\sigma_A^2=a^2\sigma_x^2+2ab\sigma_x\sigma_y\rho(x,y)+b^2\sigma_y^2$$

Since $-1\le\rho(x,y)\le+1$, we have (for $a>0$ and $b>0$):

$$\underbrace{\left(a^2\sigma_x^2-2ab\sigma_x\sigma_y+b^2\sigma_y^2\right)}_{\rho(x,y)=-1}\le\sigma_A^2\le\underbrace{\left(a^2\sigma_x^2+2ab\sigma_x\sigma_y+b^2\sigma_y^2\right)}_{\rho(x,y)=+1}$$

We can rewrite this relation as: $\underbrace{\left(a\sigma_x-b\sigma_y\right)^2}_{\rho(x,y)=-1}\le\sigma_A^2\le\underbrace{\left(a\sigma_x+b\sigma_y\right)^2}_{\rho(x,y)=+1}$

This holds for all (positive) a and b.

To help understand the consequences, suppose a and b are both = 1, and that $\sigma_x=\sigma_y=\sigma$.

Then: $A=x+y$ and $\left(\sigma_x-\sigma_y\right)^2\le\sigma_A^2\le\left(\sigma_x+\sigma_y\right)^2\;\Rightarrow\;0\le\sigma_A^2\le4\sigma^2$ or: $\underbrace{0}_{\rho(x,y)=-1}\le\sigma_A\le\underbrace{2\sigma}_{\rho(x,y)=+1}$

But since: $\sigma_A^2=a^2\sigma_x^2+2ab\sigma_x\sigma_y\rho(x,y)+b^2\sigma_y^2$, we see that if the correlation coefficient $\rho(x,y)=0$ and $a=b=1$ (in which case $A=x+y$), then $\sigma_A^2=\sigma_x^2+\sigma_y^2$, as we all learned a long time ago. We have now learned that that is true ***only*** if the x, y measurements are ***uncorrelated***!

In fact, σ_A^2 can lie anywhere between $\left(\sigma_x-\sigma_y\right)^2$ and $\left(\sigma_x+\sigma_y\right)^2$.

It is easy to imagine cases/situations where the two measurements of the random variables *x* and *y* are completely uncorrelated, so let us construct a situation where they ***are*** correlated:

Suppose that: $y = a + bx$, where *a* and *b* are just numbers (*i.e.* constants). Then:

$$
\begin{aligned}
\operatorname{cov}(x,y) &= E[xy] - E[x]E[y] = E[x(a+bx)] - \hat{x}(a+b\hat{x}) \\
&= E[ax + bx^2] - a\hat{x} - b\hat{x}^2 \\
&= \cancel{a\hat{x}} + b - \cancel{a\hat{x}} - b\hat{x}^2 \\
&= b\underbrace{\left(E[x^2] - \hat{x}^2\right)}_{=\sigma_x^2} = b\sigma_x^2 = b\operatorname{var}(x)
\end{aligned}
$$

Whereas:

$$
\begin{aligned}
\operatorname{var}(y) = \sigma_y^2 &= E[(y-\hat{y})^2] = E[y^2] - \hat{y}^2 \\
&= E[a^2 + 2ab\,x + b^2x^2] - \left(a^2 + 2ab\,\hat{x} + b^2\hat{x}^2\right) \\
&= \cancel{a^2} + \cancel{2ab\hat{x}} + b^2E[x^2] - \cancel{a^2} - \cancel{2ab\hat{x}} - b^2\hat{x}^2 \\
&= b^2\underbrace{\left(E[x^2] - \hat{x}^2\right)}_{=\sigma_x^2} = b^2\sigma_x^2 = b^2\operatorname{var}(x)
\end{aligned}
$$

Therefore, we see (here) that $\sigma_y = |b|\sigma_x$ (since by definition, both $\sigma_x, \sigma_y > 0$).

Thus, (here) the correlation coefficient $\rho(x,y) = \dfrac{\operatorname{cov}(x,y)}{\sigma_x\sigma_y} = \dfrac{b\sigma_x{}^2}{\sigma_x\,|b|\,\sigma_x} = \dfrac{b}{|b|} = \begin{cases} +1 & \text{if } b > 0 \\ -1 & \text{if } b < 0 \end{cases}$

We can now generalize the above to the case of a ***linear*** function of N random variables,

$x_1, x_2, \ldots, x_N$ with P.D.F. $f(x_1, x_2, \ldots, x_N)$.

Let: $$L(x_1, x_2, \ldots, x_N) = \sum_{i=1}^{N} a_i x_i = a_1x_1 + a_2x_2 + \ldots + a_Nx_N\,.$$

Then the ***expectation value/true mean*** of *L* is: $E[L] = \hat{L} = E\left[\sum_{i=1}^{N} a_i x_i\right] = \sum_{i=1}^{N} a_i E[x_i] = \sum_{i=1}^{N} a_i \hat{x}_i$.

The ***variance*** of *L* is:

$$
\begin{aligned}
\operatorname{var}(L) = \sigma_L^2 = E[(L-\hat{L})^2] &= E\left[\left\{\sum_{i=1}^{N} a_i(x_i - \hat{x}_i)\right\}^2\right] \\
&= E\left[\left\{\sum_{i=1}^{N} a_i(x_i - \hat{x}_i)\right\}\left\{\sum_{j=1}^{N} a_j(x_j - \hat{x}_j)\right\}\right] \\
&= E\left[\sum_{i,j=1}^{N} a_i a_j (x_i - \hat{x}_i)(x_j - \hat{x}_j)\right]
\end{aligned}
$$

$$\begin{aligned}\operatorname{var}(L)=\sigma_L^2 &= E\left[\sum_{i=1}^{N} a_i^2 (x_i-\hat{x}_i)^2 + 2\sum_{i=1}^{N} a_i \sum_{j=i+1}^{N} a_j (x_i-\hat{x}_i)(x_j-\hat{x}_j)\right]\\ &= \sum_{i=1}^{N} a_i^2 E[(x_i-\hat{x}_i)^2] + 2\sum_{i=1}^{N} a_i \sum_{j=i+1}^{N} a_j \underbrace{E[(x_i-\hat{x}_i)(x_j-\hat{x}_j)]}_{=\operatorname{cov}(x_i,x_j)}\\ &= \sum_{i=1}^{N} a_i^2 \sigma_{x_i}^2 + 2\sum_{i=1}^{N} a_i \sum_{j=i+1}^{N} a_j \operatorname{cov}(x_i,x_j)\end{aligned}$$

For ***uncorrelated*** variables (*i.e.* when $\operatorname{cov}(x_i,x_j)=0$), this reduces to:

$$\operatorname{var}(L)=\sigma_L^2=\sigma_{\Sigma a_i x_i}^2=\sum_{i=1}^{N} a_i^2\sigma_{x_i}^2$$

A very common application of this is the case where the x_i result from N repetitions of the same experiment in which a single random variable has been measured, *e.g.* N independent measurements of the length of a rod.

The linear function commonly defined is the "***average***, ***or mean value of x***" (*aka* the ***sample mean***) associated with the N independent measurements is:

$$\boxed{\bar{x} \equiv \frac{1}{N}\sum_{i=1}^{N} x_i}$$

Let us assume that the ***true mean*** $\hat{x}$ is the same for all of the N individual experiments (*i.e.* the rod does not contract or expand during the course of the experiments.)

The ***true mean*** $\hat{x}$ is the most that we can hope to learn about the ***true*** length of the rod.

The result of a single measurement, x_i is not likely to be a reliable estimate of $\hat{x}$, because it will be distributed according to the P.D.F. $f(x)$ about $\hat{x}$ with standard deviations σ_{x_i}.

We have previously learned that to "determine the ***true*** length" we are supposed to use the average of ***many*** measurements (*i.e.* $N\to\infty$). In fact:

$$E[\bar{x}]=E\left[\frac{1}{N}\sum_{i=1}^{N} x_i\right]=\frac{1}{N}\sum_{i=1}^{N} E[x_i]=\frac{1}{\cancel{N}}\cancel{N}\,\hat{x}=\hat{x}$$

Thus, the average of N independent measurements has as its expectation value $E[\bar{x}]$, which is equal to the expectation value associated with a single measurement $E[x_i]=\hat{x}$.

In order to determine "how close" the ***average/sample mean*** $\bar{x}$ will lie to the ***expectation value/true mean*** $\hat{x}$, we must look at the ***<u>variance</u>*** of the ***sample mean*** $\bar{x}$:

$$\operatorname{var}(\bar{x}) = \sigma_{\bar{x}}^2 = \operatorname{var}\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right) = E[(\bar{x} - E[\bar{x}])^2]$$

$$= \sum_{i=1}^{N}\left(\frac{1}{N}\right)^2 E\left[\left(x_i - \hat{x}_i\right)^2\right] + 2\sum_{i=1}^{N}\frac{1}{N}\sum_{j=i+1}^{N}\frac{1}{N}\underbrace{E[(x_i - \hat{x}_i)(x_j - \hat{x}_j)]}_{=\operatorname{cov}(x_i,x_j)}$$

or:

$$\boxed{\operatorname{var}(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma_x{}^2}{N} + \frac{2}{N^2}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\operatorname{cov}(x_i, x_j)}$$

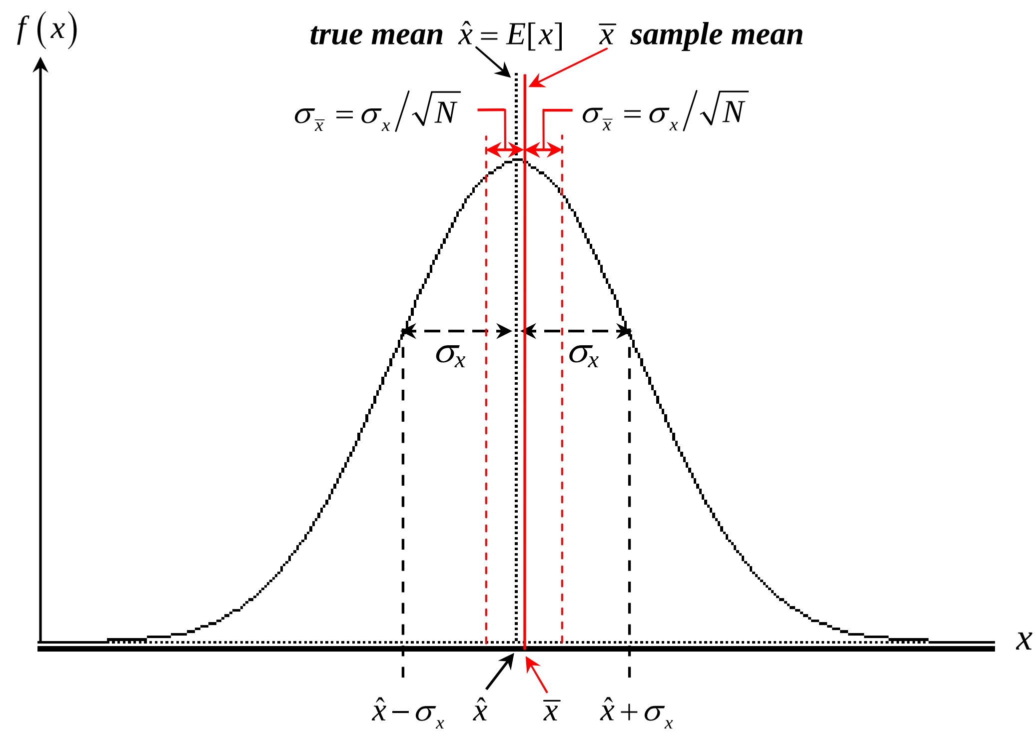
If the N experimental measurements x_i are ***<u>independent</u>*** of each other, then all of the individual covariances $\operatorname{cov}(x_i, x_j)$ are ***<u>zero</u>***, and we obtain the simple result that:

$$\boxed{\operatorname{var}(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma_x{}^2}{N}}$$ = ***variance of the sample mean (N <u>independent</u> samples)***

$$\boxed{\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}}}$$ = ***standard deviation of the sample mean***

Thus, we see that the ***standard deviation of the sample mean*** $\sigma_{\bar{x}} = \sigma_x/\sqrt{N}$ is a factor of $1/\sqrt{N}$ times smaller than the standard deviation associated with an ***<u>individual</u>*** measurement σ_x.

Therefore, we also see that (obviously) the ***sample mean*** $\bar{x} \equiv \frac{1}{N}\sum_{i=1}^{N} x_i$ is a statistically much better/more accurate ***<u>estimator</u>*** of the ***true mean*** $\hat{x}$ than any individual/single measurement x_i.



Comments:

• We have ***not*** specified the ***explicit*** form of the P.D.F. $f(x)$ in ***any*** of the above derivations.

• We ***must*** distinguish between a "finite sample" which allows us to calculate the ***average*** or ***sample mean*** $\overline{x}$ *vs*. an **idealization** of the mean (*i.e*. the **true** mean, $\hat{x}$) where ***all possible outcomes*** of an experiment can be sampled. In that case we ***would*** get the ***true mean*** $\hat{x}$, but of course this would also require us to carry out an ***infinite*** number of experiments.

Thus, if $N \to \infty$, then $\sigma_{\overline{x}} \to 0$, *i.e*. ***sample mean*** $\overline{x}$ → ***true mean*** $\hat{x}$.

• We will use the lower case $\overline{x}$ to represent an ***experimental*** determination of the ***sample mean***.

• We will ***NEVER*** use $\overline{x}$ to denote the ***expectation value/true mean*** $E[x] \equiv \hat{x}$.