Moments of Probability Distributions

Moments are used to characterize a set of similar (comparable) numerical data $\{d_i \mid i = 1,2,....N\}$. For example, if the set represents different measurements of the same quantity, one is interested in the ***average value***, or ***first moment***, of the data set,

$$\langle d \rangle = \frac{1}{N} \sum_{i=1}^{N} d_i \tag{1}$$

This is also known as ***arithmetical average***. The accuracy ***of each individual*** measurement is illustrated by the average scatter of the actual data points d_i about the arithmetic average $\langle d \rangle$. Since by definition, the sum $\sum \left(d_i - \langle d \rangle \right) = 0$, the average deviation defined in terms of this sum is always zero and carries no information about the average accuracy of the measurements. This is the reason why instead the ***variance***, or ***second moment*** of the distribution $\{d_i\}$, is taken as a measure of this accuracy,

$$\sigma_d^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(d_i - \langle d \rangle \right)^2 \tag{2}$$

The sum is over the quadratic deviations of the N individual points d_i from the average. The normalization is to the number of independent measurements of a quadratic deviation. This number is equal to (N -1) rather than equal to N, because the definition of the average $\langle d \rangle$, against which each point d_i is measured, reduces the number of remaining degrees of freedom in the sample from N to N-1. Consider, for example, with two points, d_1 and d_2, the average is defined as $\langle d \rangle = (d_1 + d_2)/2$. Then, there are only N -1=1 independent differences that can be formed from this sample, namely

$\Delta d = (d_1 - d_2)$. Any other difference can be expressed as a combination of $\langle d \rangle$ and Δd.

Experimental errors of the ***individual*** measurements d_i are typically given as ***standard deviation*** (error = $\pm\sigma_d$)

$$\sigma_d = \sqrt{\sigma_d^2} \tag{3}$$

This quantity represents a measure of the scatter of the individual data points d_i. The accuracy of the average value determined by the entire set $\{d_i\}$ is related to this standard deviation but not equal to it. By enlarging the data set by performing additional measurements may not decrease the average scatter of the individual points much, while the average value of all data points may be much better defined. The uncertainty (average error) of the arithmetic average of all individual values d_i is given in terms of the variance

$$\sigma_{\langle d \rangle}^2 = \frac{1}{N(N-1)} \sum_{i=1}^{N} \left(d_i - \langle d \rangle\right)^2 \tag{4}$$

As pointed out above, this variance has to be distinguished from the individual variance, since typically (large N) $\sigma_d^2 >> \sigma_{\langle d \rangle}^2$. Results of a set of individual experimental measurements of an observable are typically given in terms of the average plus/minus the standard deviation: $\langle d \rangle \pm \sigma_{\langle d \rangle}$.

In the following example, a set of 5 measurements $\{d_i | \ i = 1,..,5\}$ of the same observable d is considered, as listed in the table (*Excel* worksheet) below. The individual columns give i, d_i, $(d_i - \langle d \rangle)$, $(d_i - \langle d \rangle)^2$. The results of the calculation of average, variance, and standard deviation are given at the bottom of the Table. The

No. i	d_i	d_i-daver	$(d_i$-daver)^2		
1	17.3	0.06	0.0036		
2	17	-0.24	0.0576		
3	17.3	0.06	0.0036		
4	17.4	0.16	0.0256		
5	17.2	-0.04	0.0016		
	17.24		0.023	0.151658	

"statistical uncertainty" (error) of the arithmetic average of $\langle d \rangle = 17.24$ can be calculated with Equ. 4 as $\sigma_{\langle d \rangle} = 0.152 / \sqrt{5} = 0.068$.

One can use and modify the worksheet, e.g., use it for another data set, by either clicking on the Table or on the link "example ", if the program application *MS Excel* is available on the computer.

In the above example, all measurements are considered equally accurate or representative of the actual value of the observable d. However, it is often the case that measurements are compared that have been performed with different accuracy, e.g., with different instruments, different attention to detail, at different times and in different laboratories. Then, one assigns different ***weights*** $\boldsymbol{w_i}$ to the different measurements and forms a ***weighted average***,

$$\langle d \rangle = \frac{\sum_{i=1}^{N} w_i d_i}{\sum_{i=1}^{N} w_i} \qquad (5)$$

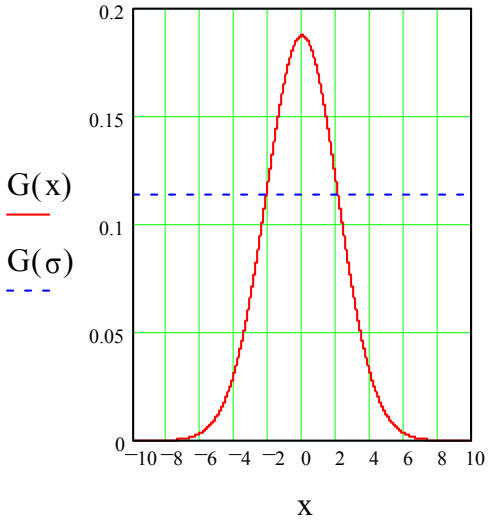
If all weights are equal, $\boldsymbol{w_i = 1}$, one recovers formula (1) for the unweighted, arithmetical average. Obviously, if one normalizes the set of weights $\tilde{w}_i$, such that $\sum \tilde{w}_i = 1$, one obtains the simplified form of the weighted average

$$\langle d \rangle = \sum_{i=1}^{N} \tilde{w}_i d_i \tag{6}$$

The variance (statistical uncertainty) for the weighted average set of data with different weights w_i is calculated as

$$\sigma^2_{\langle d \rangle} = \frac{\sum_{i=1}^{N} w_i (d_i - \langle d \rangle)^2}{(N-1)\sum_{i=1}^{N} w_i} \tag{7}$$

The rules for calculating averages and statistical uncertainty of data sets are derived in probability theory, assuming that the individual elements x_i of an actual data set $\{x_i\}$ are distributed according to a statistical probability distribution $P(x)$. The probability distribution determines average values and the spreads in values of such data sets.



The most important probability distributions are the ***Gaussian***, or ***normal***, distribution, the ***Poisson*** distribution, and the ***binomial*** distribution. In experimental applications, often a ***normally distributed data set*** is postulated or argued. Here, the probability to find a certain value *x* is given by a continuous Gaussian function

$$P(x)=\frac{1}{\sqrt{2\pi\sigma_X^2}}\cdot\exp\left\{-\frac{\left(x-\langle x\rangle\right)^2}{2\sigma_X^2}\right\} \qquad (8)$$

as illustrated in the figure. ***Average value $\langle x\rangle$ and variance*** σ_x^2, i.e., the ***first and second moments of P(x)***, ***define a Gaussian probability distribution function*** for the observable *x* completely. The probability function in Equ. 8 is normalized to unity, i.e., the integral $\int dx P(x) = 1$, and

$$\langle x\rangle = \int_{-\infty}^{+\infty} dx'\, x'\, P(x')$$

$$\sigma_x^2 = \langle x^2\rangle - \langle x\rangle^2 = \int_{-\infty}^{\infty} dx' \left(x'-\langle x\rangle\right)^2 P(x') \qquad (9)$$

Since the variance is calculated with respect to the average $\langle x\rangle$, the ***central value*** of *P*, the variance is the second ***central moment of P(x)***. Higher central moments, like the skewness of *P(x)*, are defined in an analog fashion. For a Gaussian, all odd moments are zero. It is completely determined by only the two first moments.

The graph above shows a specific example of such a Gaussian function for the variable *x*. It is positioned at *x=0*, i.e., its center and average is equal to $\langle x \rangle = 0$, its variance equals $\sigma_x^2 = 4.52$, corresponding to a ***full width at half-maximum*** of $\Gamma_{FWHM} = 2.35 \cdot \sigma_x = 5.0$ units in *x*. The value of the Gaussian one ***standard deviation*** above and below the average, i.e., at $x = \langle x \rangle \pm \sigma_x$, is indicated in the figure by the intersection of the dashed horizontal line with the Gaussian.

The Gaussian or normal distribution is the limit of the ***binomial distribution***,

$$P_{binomial}(m) = \binom{N}{m} p^m (1-p)^{N-m} \qquad (10)$$

defined for the integer variable *m*. The binomial distribution describes, for example, the probability for *m* successes out of a total of *N* attempts, when the probability for a success in any one attempt is given by *p*. The quantity p^m gives the probability for having *m* attempts succeed, regardless of what happens to the other *N-m* attempts, while the quantity $(1-p)^{N-m}$ gives the probability that these *N-m* attempts do not succeed. The product of these two probabilities is, hence, equal to the probability that a given set of *m* attempts, chosen out of *N* total attempts, succeed. The combinatorial ***binomial coefficient***

$$\binom{N}{m} = \frac{N!}{m!(N-m)!} = \frac{(N-m+1)\cdots N}{1\cdots m} \qquad (11)$$

gives the number of ways that a group of *m* attempts can be chosen out of *N* total attempts. It multiplies the product of the probabilities, since one is only interested in the number *m* of successful at-

tempts, for example $m = 3$ out of $N = 56$, but not in which of the attempts succeed, e.g., $i = 15, 28,$ and 44.

The binomial distribution is normalized to unity:

$$1 = \sum_{m=0}^{N} P_{binomial}(m) = \sum_{m=0}^{N} \binom{N}{m} p^m (1-p)^{N-m} \qquad (12)$$

Average value and variance for the binomial distribution are given by

$$\overline{m} = N \cdot p \qquad and \qquad \sigma_m^2 = N \cdot p(1-p) \qquad (13)$$

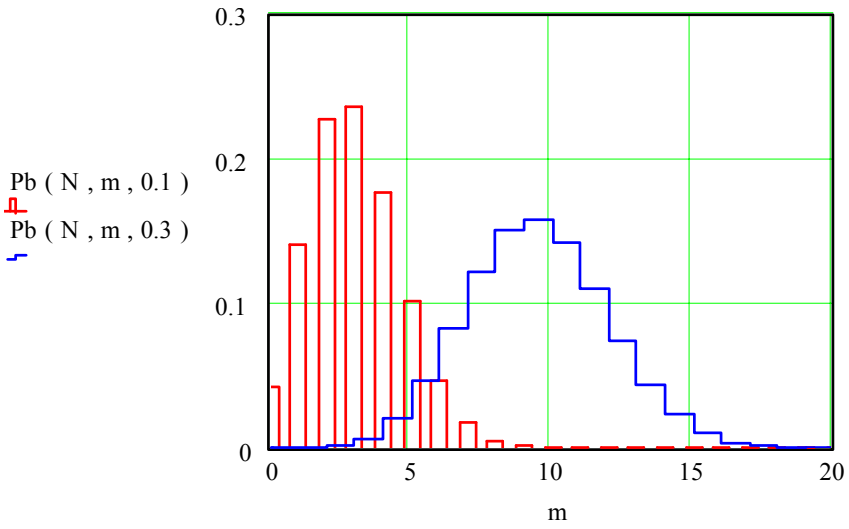
This is proven by considering the definition of the average

$$\overline{m} = \sum_{m=0}^{N} m P_{binomial}(m) = \sum_{m=0}^{N} m \cdot \binom{N}{m} p^m (1-p)^{N-m} \qquad (14)$$

m=k+1 N=M+1

$$\overline{m} = \sum_{m=0}^{N} m \cdot \frac{N!}{m!(N-m)!} p^m (1-p)^{N-m} \qquad (15)$$

As examples. The figure shows two binomial distributions, each for a total of $N = 30$ trials. The distribution for $p = 0.1$ is concentrated at $\langle m \rangle \approx 3$ and somewhat asymmetric. For the larger elementary probability of $p = 0.3$, the distribution moves to larger values of m. Simultaneously, it becomes broader and more symmetric. The distribution for $p = 0.3$ almost looks like a Gaussian. In fact, the Gaussian is a good approximation to the binomial distribution even for a relatively small number of trials $N \approx 10$.



As an example for the application of the binomial probability distribution, one can consider the game of coin tossing. The probability for success, e.g., to obtain "head", in a single trial is equal to $p = 1/2$. The probability for failure in a single trial is given by $(1 - p)$, which happens to be equal to p.

It is easy to show numerically that in the limit of large values of N and a finite probability $p \neq 0$, the binomial distribution becomes Gaussian,

$$\lim_{N\to\infty} P_{binomial}(m) = \frac{1}{\sqrt{2\pi\sigma_m^2}} \cdot \exp\left\{-\frac{\left(x - \langle m\rangle\right)^2}{2\sigma_m^2}\right\} \qquad (13)$$

where average and variance are given by Equ. 12, but with a replacement of the discrete variable m by a continuous variable m. This property is plausible already from the figure above.

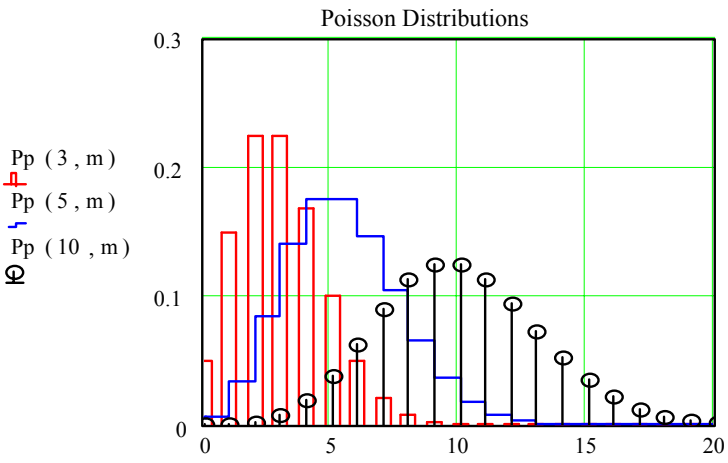
It is interesting to inspect the relation between average value and variance, because that relation is characteristic for statistical probability distributions. Considering the normal (Gaussian) distribution as a reasonably accurate fit to a binomial distribution, one has for the ratio of standard deviation ($\sigma_m = \sqrt{\sigma_m^2}$) to average,

$$\frac{\sigma_m}{\langle m \rangle} = \frac{\sqrt{Np(1-p)}}{Np} = \frac{1}{\sqrt{N}}\sqrt{\frac{1-p}{p}} \propto \frac{1}{\sqrt{N}} \qquad (14)$$

This feature implies that, although the Gaussian becomes broader with increasing N, the ***relative width*** given by Equ. 14 ***decreases with $1/\sqrt{N}$***.

A probability distribution that is important for very unlikely processes with $p \to 0$ is the ***Poisson distribution***. For large numbers N of events (trials), $N \to \infty$, when the product the total number of success remains finite in spite of small elementary probabilities, i.e., when $N \cdot p \neq 0$, the binomial distribution transforms into the Poisson distribution

$$P_{Poisson}(\mu, m) = \frac{\mu^m \cdot e^{-\mu}}{m!} \qquad (15)$$



where $\mu = Np$ is the average value.

Examples of Poisson distributions are given in the figure, for average values of $\mu = 3, 5,$ and 10. One observes again an evolution towards a Gaussian, for increasing values of μ (i.e., N).

Examples of processes described by Poisson-type probability distributions include decay and dissociation processes, tunneling

processes in general, where there a large number of trials (assaults on the potential barrier) but very few successes.