STATISTICAL MECHANICS IN CHEMISTRY

Probability, probability distribution, averages, moments

March 1, 2023

Probability (Laplace's definition)

$$P(A) = \lim_{n\to\infty} \frac{N(A)}{N}$$
$$0 \leq P(A) \leq 1$$
$$\sum_A P(A) = 1$$

A – an *event* (e.g., getting a face of a dice with 6 points in a single throw, getting a head in a single coin throw, etc.),
$N(A)$ – the number of trials in which event A occurred,
N – total number of trials.

Opposite events, the space of all events

The opposite event (not A) is usually denoted as $\overline{A}$ (a complement to A). We have

$$P(A) + P(\overline{A}) = 1$$

or

$$P(\overline{A}) = 1 - P(A)$$

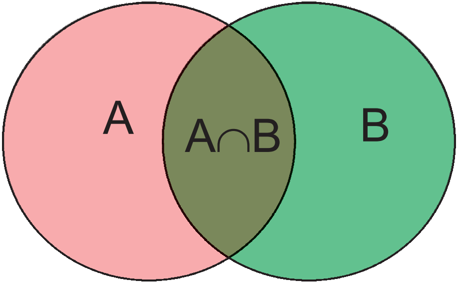
By E we denote the space of all events. We have

$$P(E) = 1$$

Conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

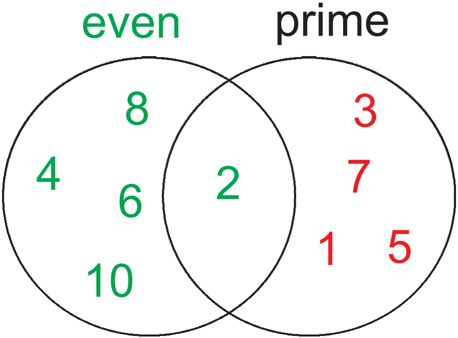
$$P(A \cap B) = P(A|B)P(B)$$



The Venn diagram

Example

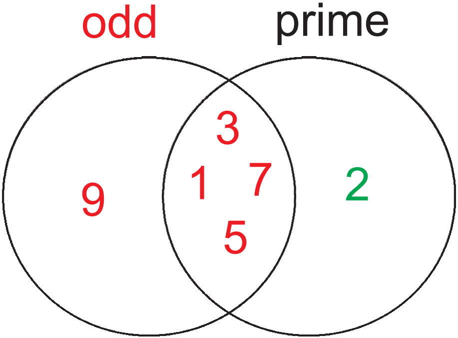
Probability of finding a prime number in the set of integers from 1 to 10. The prime numbers are 1, 2, 3, 5, and 7. There are 5 odd and 5 even numbers in the set.



$$P(prime|even) = \frac{P(prime \cap even)}{P(even)} = \frac{0.1}{0.5} = 0.2$$

Example

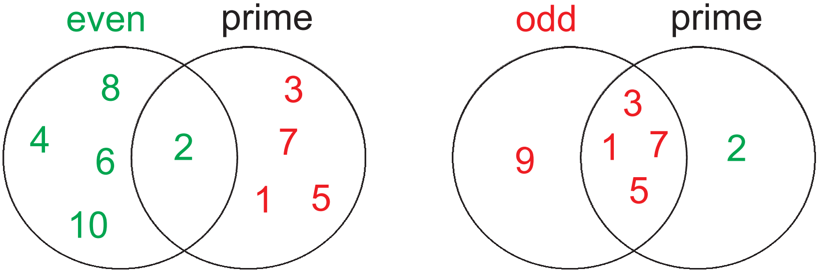
Probability of finding a prime number in the set of integers from 1 to 10. The prime numbers are 1, 2, 3, 5, and 7. There are 5 odd and 5 even numbers in the set.



$$P(prime|odd) = \frac{P(prime \cap odd)}{P(odd)} = \frac{0.4}{0.5} = 0.8$$

Example

Probability of finding a prime number in the set of integers from 1 to 10. The prime numbers are 1, 2, 3, 5, and 7. There are 5 odd and 5 even numbers in the set.



$$P(prime) = P(prime \cap even) + P(prime \cap odd) =$$
$$P(prime|even)P(even) + P(prime|odd)P(odd) =$$
$$0.2 \times 0.5 + 0.8 \times 0.5 = 0.5$$

Total probability of event B to happen

$$P(B) = \sum_i P(B|A_i)P(A_i)$$

$$E = \sum_i A_i$$

The events A_1, $\mathrm{A}_2 \ldots$, form a *complete* space of events.

Real life: I lost my house keys while walking my dog. How do I proceed to find them?

Independent events

The A and B events are independent of each other if

$$P(A|B) = P(A) \qquad P(B|A) = P(B)$$
$$P(A \cap B) = P(A)P(B)$$

For independent events A, B, $C \ldots$, we have

$$P(A \cup B \cup C \ldots) = P(A) + P(B) + P(C) + \ldots$$

Random variables and their distributions

Random variable: a number that is assigned to a given event. Any function of a random variable is a random variable.
Examples:

- The number of points on the face of a dice that turns up.
- A temperature read.
- The mass of a bean picked at random.

Important: that a variable is random does not mean that it will take any value with equal probability. It obeys a certain *distribution*.

Cumulative distribution function and distribution function

Cumulative distribution function

$$F(x) = P(y \leq x)$$

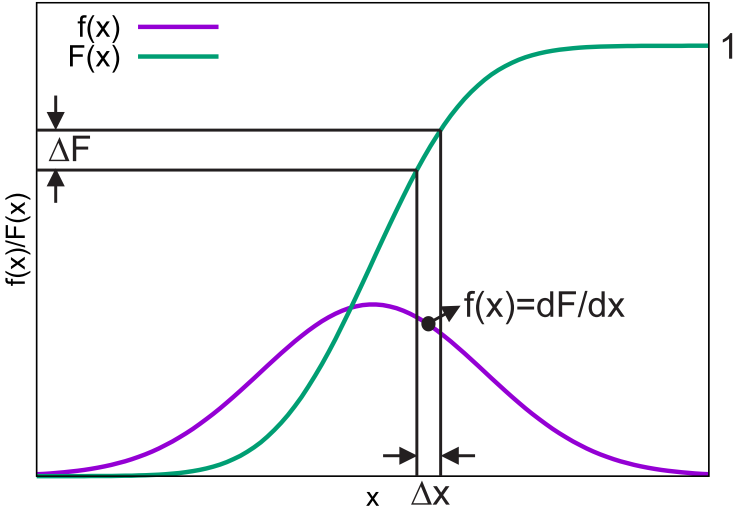
Distribution function

$$f(x) = \frac{dF(x)}{dx} = \lim_{\Delta x \to 0} \frac{P(x < y \leq x + \Delta x)}{\Delta x}$$

Discrete distribution function

$$f(n) = P(n)$$

Cumulative distribution function and distribution function



Discrete distributions

Binomial distrubution

$$W(n,k) = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

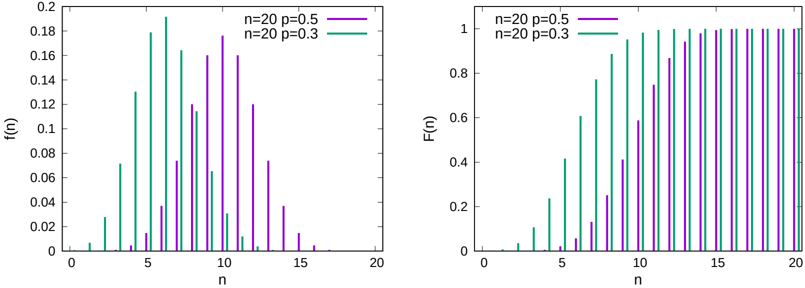
The number of possibilities of a successful draw in n trials with same success and failure probability in a single trial.

$$P(n,k,p) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

The probability of k successful draws in n trials with success probability in a single draw being p.

Discrete distributions

Sample binomial distributions



Discrete distributions

Polynomial distribution

$$W(k_1, k_2, \dots k_m) = \frac{n!}{k_1! k_2! \dots k_m!}$$

$$k_1 + k_2 + \dots k_m = n$$

The number of possibilities of selecting object 1 k_1 times, etc., in n trials given the same chance of selecting an object in a single trial.

$$P(k_1, k_2, \dots k_m, p_1, p_2, \dots p_m) = \frac{n!}{k_1! k_2! \dots k_m!} p_1^{k_1} p_2^{k_2} \dots p_m^{k_m}$$

$$p_1 + p_2 + \dots + p_m = 1$$

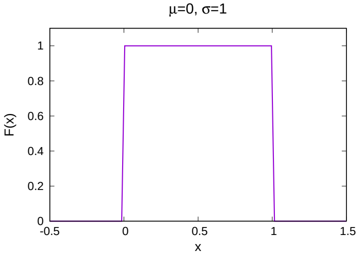
The probability of a successful selection of object 1 k_1 times, etc., in n trials.

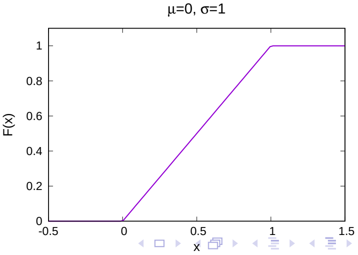
Continuous distribution functions

The uniform distribution

$$f(x;a,b) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \le x \le b \\ 0 & x > b \end{cases}$$

$$F(x;a,b) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$$



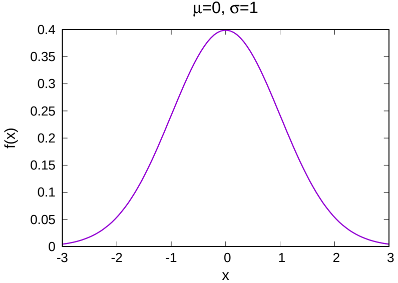


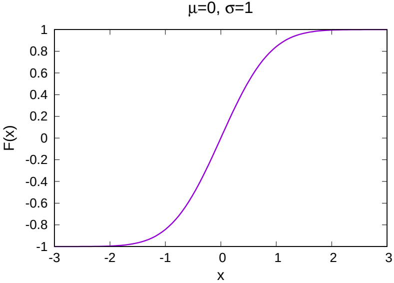
Continuous distribution functions

The normal distribution

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

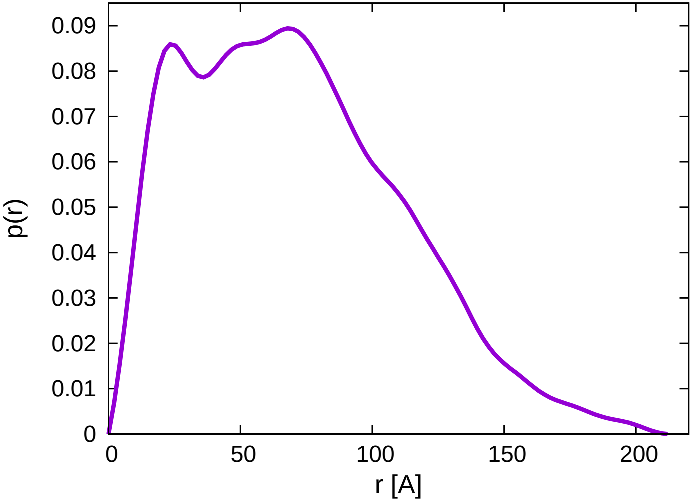
$$F(x;\mu,\sigma) = \mathrm{erf}\left(\frac{x-\mu}{\sigma}\right)$$





Continuous distribution functions

Sample multimodal distribution (distance distribution from SAXS measurements)



Moments of probability distribution

Average

$$\overline{x} = \frac{x_1 + x_2 + \dots x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

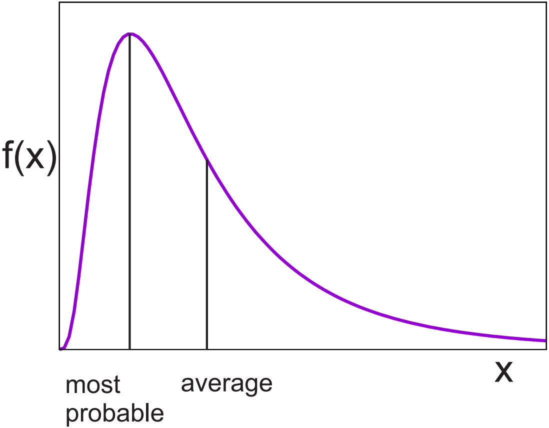
If the values of $x_1, x_2, \dots x_n$ occurred n_1, $n_2, \dots, n_m$ times, respectively,

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i = \sum_{i=1}^{m} P_i x_i$$

$$n = \sum_{i=1}^{m} n_i, \quad P_i = \frac{n_i}{n}$$

Moments of probability distribution

The **most probable** value of a random variable is the value for which the distribution has the global maximum. It is equal to the average value for unimodal symmetric distributions.



Moments of probability distribution

Variance

$$\sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}x_i^2 - \frac{1}{n(n-1)}\left(\sum_{i=1}^{n}x_i\right)^2 \approx$$
$$\overline{(x-\overline{x})^2} = \overline{x^2} - (\overline{x})^2$$

When the same results are grouped

$$\sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{m}n_i x_i^2 - \frac{1}{n(n-1)}\left(\sum_{i=1}^{m}n_i x_i\right)^2$$
$$\approx \sum_{i=1}^{m}P_i x_i^2 - \left(\sum_{i=1}^{m}P_i x_i\right)^2$$

Moments of probability distribution

Average for discrete and continuous distribution of random variable x.

$$\lambda_1(\{x\}) = \sum_{i=1}^{n} P_i x_i \qquad \lambda_1(\{x\}) = \int_{-\infty}^{\infty} x P(x) dx$$

Central moments of discrete and continuous distributions of x

$$\mu_k(\{x\}) = \sum_{i=1}^{n} (x_i - \overline{x})^k P_i \qquad \mu_k(\{x\}) = \int_{-\infty}^{\infty} (x - \overline{x})^k P(x) dx$$

Useful central moments and derived quantities

Variance (σ^2) and standard deviation (σ).

$$\sigma^2(\{x\}) = \mu_2(\{x\}) = \int_{-\infty}^{\infty} (x - \overline{x})^2 f(x) dx, \quad \sigma(\{x\}) = \sqrt{\sigma^2(\{x\})}$$

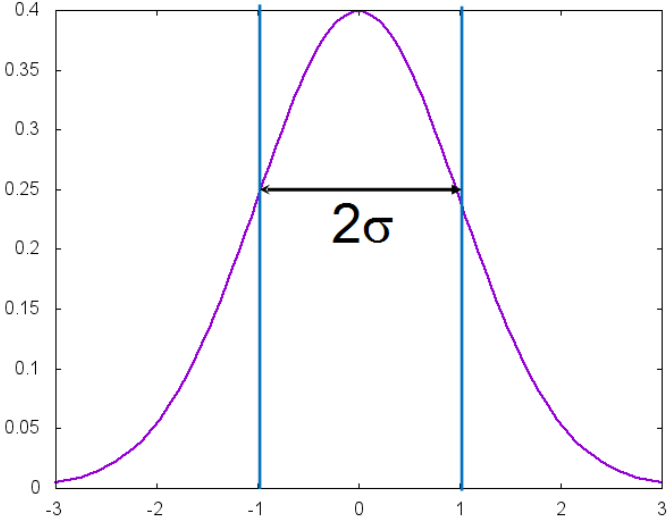
Skewness (γ)

$$\gamma(\{x\}) = \frac{\mu_3(\{x\})}{\mu_2^{\frac{3}{2}}(\{x\})} = \frac{1}{\sigma^3(\{x\})} \int_{-\infty}^{\infty} (x - \overline{x})^3 f(x)$$

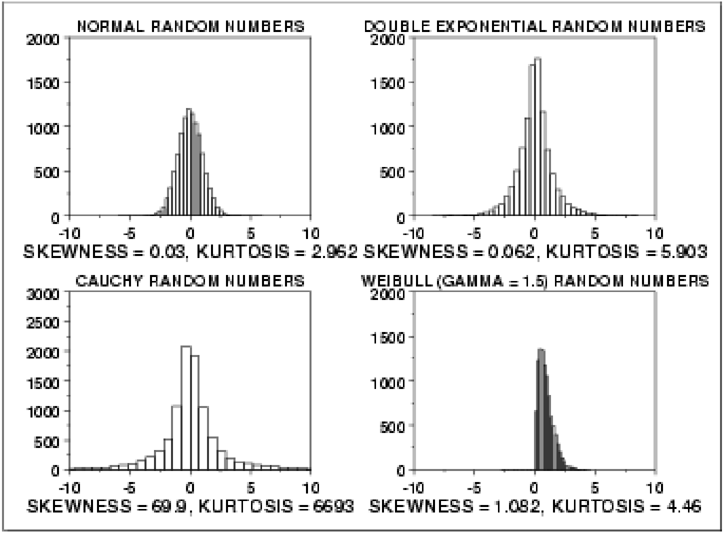
Kurtosis (κ)

$$\kappa(\{x\}) = \frac{\mu_4(\{x\})}{\mu_2(\{x\})^2} = \frac{1}{\sigma^4(\{x\})} \int_{-\infty}^{\infty} (x - \overline{x})^4 f(x)$$

Illustration of the standard deviation



Central moments of some distributions



The Central Limit Theorem

Given the random variable x, distributed with mean a and variance b^2, the variable

$$\xi = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i$$

has a normal distribution with mean a and variance b^2/n.

Distribution of two random variables, $P(x, y)$. Covariance.

$$\lambda_{10} = \overline{x}$$

$$\lambda_{01} = \overline{y}$$

$$\mu_{20} = \sigma^2(\{x\})$$

$$\mu_{02} = \sigma^2(\{y\})$$

$$\mu_{11} = \overline{(x - \overline{x})(y - \overline{y})} = \overline{xy} - \overline{x} \bullet \overline{y} = \mathrm{cov}(\{x\}, \{y\})$$

Correlation coefficient

$$\rho(\{x\}, \{y\}) = \frac{\mathrm{cov}(\{x\}, \{y\})}{\sigma(\{x\})\sigma(\{y\})}$$

Examples of uncorrelated and correlated bivariate distributions