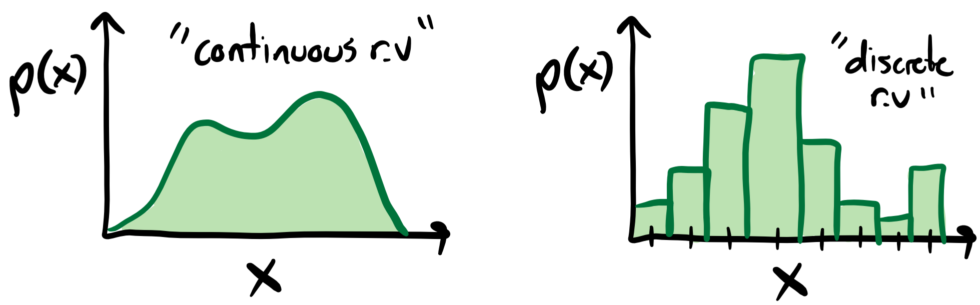
Randomness and Probability

Since many aspects of biology are stochastic, we'll also need to rely on some concepts from ***probability and statistics***. We won't require anything too sophisticated – most biology students will have encountered this material in a statistics course, while physicists will have likely seen it in a previous statmech or quantum class. Both communities tend to use slightly different notation, so we will try to provide a list of common terminology here.

Random variables. We'll assume that you are familiar with the concept of a ***random variable***, $\hat{x}$, which is distributed according to some probability distribution $p(x)$, which could be continuous or discrete:



A continuous random variable might be appropriate for a continuous quantity like position or time, while a discrete random variable might refer to something discrete like the number of ribosomes in a cell or the conformational state of a protein. It's common to refer to the distribution of a random variable using the notation $\hat{x} \sim p(x)$ [pronounced "x is distributed according to the distribution $p(x)$"]. If we're getting sloppy, we might drop the hat.

Normalization. One of the defining properties of a probability distribution is that they are ***normalized*** (i.e. that they sum to one):

$$\int p(x)dx = 1 \quad \text{(continuous r.v.)} \tag{14}$$

$$\sum_i p(x_i) = 1 \quad \text{(discrete r.v.)} \tag{15}$$

This is sometimes known as ***the law of total probability*** – it's simply a statement that *something* has to happen if we know we've enumerated all the possibilities.

Means and averages. The ***average*** / ***mean*** / ***expected value*** of $\hat{x}$ will be denoted by

$$\langle x \rangle \equiv \mathbb{E}[x] \equiv \int x \cdot p(x)\, dx\,, \tag{16}$$

for a continuous random variable, or

$$\langle x \rangle \equiv \mathbb{E}[x] \equiv \sum_i x_i \cdot p(x_i) \tag{17}$$

for a discrete random variable. In many cases, the mean provides a decent estimate of the "typical" value of $\hat{x}$.

Variance and uncertainty The ***variance*** (or ***mean squared deviation***) of $\hat{x}$ is defined by

$$\mathrm{Var}(x) \equiv \sigma_x^2 \equiv \langle x^2 \rangle - \langle x \rangle^2\,. \tag{18}$$

The square root of this quantity, also known as the ***standard deviation***,

$$\mathrm{Std}(x) \equiv \sigma_x \equiv \sqrt{\mathrm{Var}(x)} \tag{19}$$

is often used to quantify the uncertainty in $\hat{x}$ (or the "spread" in its probability distribution $p(x)$). Note that from the definition above, the variance and standard deviation satisfy the scaling property $\mathrm{Var}(c \cdot x) = c^2 \cdot \mathrm{Var}(x)$ and $\mathrm{Std}(c \cdot x) = c \cdot \mathrm{Std}(x)$

Common distributions. We will assume that you are familiar with some common probability distributions. These include discrete distributions like the ***binomial distribution***,

$$n \sim \mathrm{Binomial}(N, p) \implies P(n) = \binom{N}{n} p^n (1-p)^{N-n} \tag{20}$$

which models the number of successes in N independent coin flips with success probability p. Another common distribution we'll encounter is the ***Gaussian*** or ***Normal distribution***,

$$x \sim \mathrm{Gaussian}(\mu, \sigma^2) \implies p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{21}$$

which has mean $\langle x \rangle = \mu$ and variance σ^2. To save space, we will sometimes write this as $x \sim N(\mu, \sigma^2)$.

Note: *Wikipedia is extremely useful for common probability distributions.*[2] *It lists formulas for the means, variances, and other moments (when they are known), as well as useful identities connecting the different distributions.*

Strongly peaked distributions. Throughout this course, we will often encounter distributions that are ***strongly peaked*** — that is, most of their weight is concentrated within a narrow band of x values where $p(x)$ is highest. An example might be the binomial distribution in Eq. (20) with a very large value of N: if we flip a million fair coins, we are fairly confident that the number of heads

[2] e.g. `https://en.wikipedia.org/wiki/Binomial_distribution`.

will be close to 50%. In this case, it is common to summarize these distributions by Taylor expanding $p(x)$ in the neighborhood of this "most likely" value.

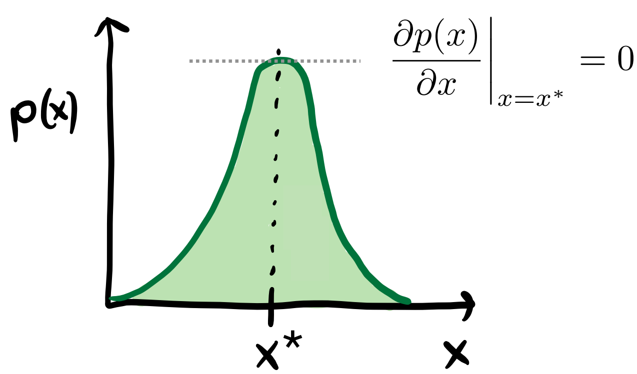
For example, if the point x^* is a maximum of $p(x)$, then we know from calculus that the first derivative must vanish at x^*:

$$\left.\frac{\partial p(x)}{\partial x}\right|_{x=x^*} = 0 \tag{22}$$

In practice, it is more common (and often easier) to work with the logarithm of $p(x)$ instead, which satisfies the same condition:

$$\left.\frac{\partial \log p(x)}{\partial x}\right|_{x=x^*} = 0 \tag{23}$$

We will often use this criterion to identify where the most likely value of x is. In many cases of interest the most likely value x^* will also coincide with the mean $\langle x \rangle$ defined above. This makes it another good summary of the "typical" value of x (which is often easier to calculate, since it doesn't require us to perform an integral).



Joint distributions. We'll also need to use the concept of a ***joint distribution***, which describes how a collection of 2 (or more) random variables are distributed at the same time:

$$p(x, y) \equiv \begin{array}{c}\text{“probability that } \hat{x} = x \text{ and} \\ \hat{y} = y \text{ at the same time”}\end{array} \tag{24}$$

If we know the joint distribution, we can calculate the single-variable distribution for any one of the variables (also known as the ***marginal distribution***) by integrating over the possible values of the others:

$$p(x) \equiv \int p(x, y)\, dy \tag{25}$$

This is sometimes known as the ***law of total probability***. We can also define the ***conditional probability***,

$$p(x|y) \equiv \frac{p(x, y)}{p(y)} \equiv \begin{array}{c}\text{“probability that } \hat{x} = x \\ \text{if we know that } \hat{y} = y\text{”}\end{array} \tag{26}$$

An important concept is ***statistical independence***, which occurs when the joint distribution factorizes:

$$p(x, y) = p(x)p(y) \tag{27}$$

Using the definition of the conditional probability in Eq. (26), we can equivalently write this as

$$p(x|y) = p(x) \iff \text{“x is independent of y”} \tag{28}$$

In other words, two random variables are independent if knowing the value of y provides no extra information about the value of x (and vice versa).

Sums of random variables and the central limit theorem. Throughout the course, we'll often encounter phenomena that depend on sums of independent random variables. E.g. the total force imparted on an E. coli cell by collisions with a large number of solvent molecules. In this case, an important result will be the ***central limit theorem***. This result says that $x_1, x_2, \ldots, x_n$ are independent random variables, then for sufficiently large n, their sum will approach a Gaussian distribution:

$$\sum_{i=1}^{n} x_i \approx \text{Gaussian}\left[\sum_i \langle x_i \rangle, \sum_i \text{Var}(x_i)\right] \tag{29}$$

whose mean and variance are equal to the sums of the means and variances of the x_i. If the x_i all the same mean and variance, we will often write this as

$$\frac{1}{n}\sum_{i=1}^{n} x_i \approx \langle x \rangle \pm \sqrt{\frac{\text{Var}(x)}{n}} \tag{30}$$

which shows that the spread of the mean of a bunch of observations scales like $1/\sqrt{n}$ when n is large. The central limit theorem will become important when we consider diffusion and random walks later in the course.