Chapter 1:

Goals

* Learn how to obtain the probabilities of all possible outcomes from a given model and see how we can compare the theoretical frequencies with those observed in real data.
* Explore a complete example of how to use the Poisson distribution to analyse data on epitope detection.
* See how we can experiment with the most useful generative models for discrete data: Poisson, binomial, multinomial.
* Use the R functions for computing probabilities and counting rare events.
* Generate random numbers from specified distributions.

Poisson distribution is continuous – probabilities of number of mutations in HIV after each replication

A point mutation is a discrete variable – it occurs or it does not (yes or no). It is a categorical variable. Categorical variables can have few or many levels. Tally number of instances. In R, categorical variables are called factors.

Tossing a coin has two possible outcomes. This simple experiment, called a Bernoulli trial, is modeled using a so-called Bernoulli random variable. Understanding this building block will take you surprisingly far. We can use it to build more complex models.

**probability or generative modeling**: all the parameters are known and the mathematical theory allows us to work by **deduction** in a **top-down** fashion.

If instead we are in the more realistic situation of knowing the number of patients and the length of the proteins, but don’t know the distribution of the data, then we have to use statistical modeling. This approach will be developed in Chapter 2. We will see that if we have only the data to start with, we first need to fit a reasonable distribution to describe it

What is the output of the formula for k=3k=3, p=2/3p=2/3, n=4n=4?