Chapter 8

Stochasticity and Probability Distribution

What is stochastic?

In the preceding chapters, we used ordinary differential equations to understand the dynamics of a system. The key idea behind those analyses is that if we know the rules of the time evolution of a system and the initial conditions, we will predict the exact state of the system in the future. This is determinism.

That is why our ODE-based models are called deterministic models. In our models, we define the rules of time evolution using the ODEs. However, rules need not be in that form. You can use some other logical representation of the rules. However, you must know the rules and also the initial state of the system.

It is not difficult to imagine that these two requirements are hard to meet for most real-life problems. Take the example of road accidents. It is extremely difficult to know all the traffic movement rules and the vehicles' initial positions and velocities.

There are systems/processes for which we are ignorant of all the rules of the time evolution of the system or the system's initial state or both. For these systems, we cannot make an exact prediction about their future states. Such systems/processes are called stochastic.

As we can not make an exact prediction, stochastic processes have uncertainties and surprise us. However, is it an issue of mere ignorance, or there could be some inherently uncertain systems? This question has bothered us for a long. It is mostly accepted that certain systems are inherently stochastic, as the radioactive decay of an isotope.

However, we will not deal with such inherently stochastic systems but focus only on uncertainties arising out of our ignorance.

Take the example of a coin toss – a staple example of a stochastic process. If we know the orientation of the coin and all the forces applied, we will be largely able to predict the outcome of a coin toss. You can make a robotic coin flipper that would toss a coin with quite a predictable outcome. I am not assuring an exact outcome as even for a robot, there will be some minor variations beyond our control, and those variations would affect the coin flip.

Therefore, I will consider coin toss as an example of stochasticity out of our ignorance. Take another textbook example to bring home the message – Brownian motion of pollen. The pollen moves on the surface of the water as water molecules collide with it repeatedly. The travel of the pollen is unpredictable to our eye. However, what if we know the exact position and momentum of all water and pollen molecules? Won't we be able to predict the future movement of the pollen just by using the equation of classical mechanics?

Possibly yes. However, we will prefer to avoid such a herculean effort to know the initial states of all molecules and to do innumerable calculations. Instead, it is smarter to do what Einstein and Smoluchowski did. Consider the Brownian motion as a stochastic process and use probabilistic reasoning to relate diffusion with the mean square displacement of the pollen.

That brings us to another difference between a deterministic process and a stochastic process. For stochastic processes, we need to change the questions that we ask. Take the example of a coin toss. I have a fair coin. If I toss it, will I get a *head*? One can not answer this question. However, what if I change the question. What is the probability of getting a *head* in a coin toss? I am sure you will be able to answer this revised question.

For a stochastic process, we determine the probability of the system being in a particular state and predict how this probability changes with time. Such calculations are often difficult, and we focus on specific characteristics of the underlying probability distribution, like mean and variance.

A stochastic process is also called a random process. The variables in a stochastic process are random variables. For example, a coin toss is a stochastic process, and the number of *heads* obtained in ten sequential coin toss is a random variable.

However, in colloquial English, random has multiple meanings. Sometimes it means something bizarre or arbitrary. Other times we call something random if it had happened without a cause or plan. We may have uncertainties and ignorance of the underlying mechanisms, but there is nothing bizarre about a stochastic process. Further, a stochastic process does not violate the principle of causality, even though we may not know all the causes behind an event.

Sometimes it is assumed that in a random process, everything is equally likely. That is not true. We can have a coin that is biased to head. Still, flipping the coin will be a random process, and we will have uncertainty about the outcome.

To avoid these misconceptions, we will instead stick to the "stochastic process" in place of "random process."

Uncertainty is inherent to stochastic processes. However, the uncertainty increases when the system size is small, or the probability of an event is low. Take the example of road accidents. If an accident's probability is very high, then almost every other moment there will be an accident, and the uncertainty will drop.

On the other hand, let the probability of an accident be low, but you count all the road accidents happening on earth. The uncertainty will drop, and you will be sure that there will be one accident somewhere on earth in the next moment.

We have earlier discussed the effect of system size on the uncertainty in a biochemical reaction. Biochemical reactions in a cell are stochastic. We never have the complete information to predict the exact behavior of a biochemical reaction inside a cell. We have uncertainty about the timing of individual reaction events. This uncertainty increases when the number of molecules is very low.

In biology, a process is affected by two types of stochasticity – intrinsic and extrinsic. The stochasticity in the dynamics of the process itself is called the intrinsic stochasticity. In contrast, extrinsic stochasticity originates from the stochasticity of other associated processes inside the cell or in the environment.

For example, suppose we induced a gene expression in *E. coli* by adding IPTG in the media. The transcription of the target gene in an *E. coli* cell has intrinsic stochasticity. Additionally, uptake of IPTG by a cell is another stochastic process. In a population of cells, all cells will not get the same amount of IPTG. Variation in the uptake of IPTG will also affect transcription. That is extrinsic stochasticity.

Suppose we count the number of copies of a particular mRNA produced in an *E. coli* after the induction with IPTG. Just now, one copy of mRNA is produced. As transcription is stochastic, we have uncertainty when the next copy of the mRNA is produced. The time interval between the production of two mRNAs is a random variable.

Therefore, the data collected from a single cell will not be smooth but would look like a step plot. The blue line in Figure 1 shows the data for a particular cell. The other lines, pink and red, are for two different cells. Each cell is independent, and the dynamics of transcription in these cells are also distinctly different.

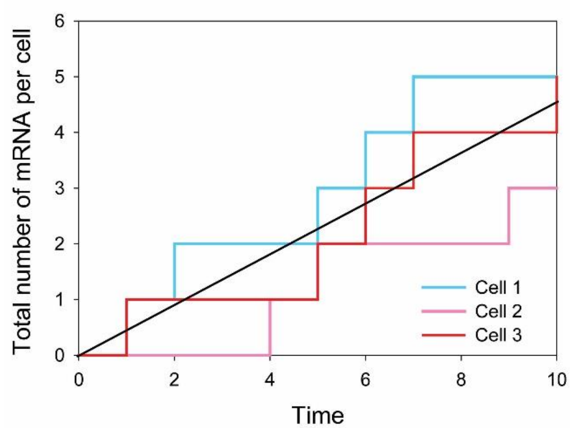

Figure 1: Stochastic fluctuation in the expression of a gene. Blue, pink and red lines are for data collected from three cells. The black line is the average of the data collected from thousands of cells.

However, if we collect data from many cells and average those, we will get a smooth curve (black line in Figure 1). Averaging over many observations reduces the fluctuations in the data and makes the system equivalent to a deterministic system.

The fluctuations observed in the data of a stochastic process is known as the noise. For this reason, a stochastic process is called a noisy process. Often, 'noise' is used as an empirical measure of stochasticity and typically defined as the ratio of the standard deviation of a random variable to its mean.

Random variables often follow well-characterized probability distributions. In this chapter, we will learn some frequently used probability distributions and their properties.

Flip a coin

Take a fair coin and toss it four times. How many times will you get a *head*? You may tempt to say – two. However, we will not get two *heads* every time we toss the coin four times. Sometimes we may get four *heads*, sometimes may get just one, and so on.

The number of *heads* in four successive coin tosses is a random variable. So, we need to change the question. It would be better to ask, "What is the probability of getting two *heads* in four coin tosses?"

Let P_H and P_T are the probabilities of getting *head* and *tail*, respectively. For a fair coin $P_H = P_T = 0.5$. We can get two *heads* in four tosses in ${}^4C_2 = 6$ ways: HHTT, HTHT, HTTH, THTH, TTHH, THHT. As successive coin tosses are independent, the probability of getting HHTT is $P_H P_H P_T P_T = P_H{}^2 P_T{}^{(4-2)} = P_H{}^2 P_T{}^2$. The rest of the combinations have the same probability.

As the six sets of coin tosses are mutually exclusive, the overall probability of getting two *heads* in four tosses is,

$$
\begin{aligned}
P(H=2) &= P_H{}^2 P_T{}^2 + P_H{}^2 P_T{}^2 + P_H{}^2 P_T{}^2 + P_H{}^2 P_T{}^2 + P_H{}^2 P_T{}^2 + P_H{}^2 P_T{}^2 \\
&= {}^4C_2 P_H{}^2 P_T{}^2 \\
&= 6\left(\tfrac{1}{2}\right)^2 \left(\tfrac{1}{2}\right)^2 = 0.375
\end{aligned}
$$

We can generalize this example. For that, change the terminology. Getting a *head* in a coin toss is an *event* and tossing the coin is a trial. Let X is the total number of events in N trials. So the probability of k events happening in N trials is,

$$
P(X=k) = {}^NC_k\, p^k (1-p)^{(N-k)} \qquad (1)
$$

Here, $0 \le X \le N$ and $0 \le k \le N$. p is the probability of the event and $(1-p)$ is the probability that the event will not happen.

The function $P(X=k)$ is called the Probability Mass Function (PMF). One can define a PMF as a function that gives the probability that a discrete random variable is exactly equal to a specific value.

The outcome of a coin toss is random. It is discrete as we either get a *head* or a *tail*. We can not have 1.23 *head*. Therefore, the number of *heads* obtained in N number of coin tosses is a discrete random variable.

Let $a \le X \le b$, where X is a discrete random variable and $P(X = k)$ is the PMF of X. Following the theorems of probability, one can write,

a) $0 \le P\left(X = k\right) \le 1$, where $a \le k \le b$

b) $\sum_{k=a}^{b} P\left(X = k\right) = 1$

Therefore, for the PMF in Equation 1,

a) $0 \le {}^{N}C_{k}\, p^{k} \left(1 - p\right)^{(N-k)} \le 1$, where $0 \le k \le N$

b) $\sum_{k=0}^{N} {}^{N}C_{k}\, p^{k} \left(1 - p\right)^{(N-k)} = 1$

The PMF in Equation 1 is called the PMF of Binomial Distribution. The binomial distribution is a discrete probability distribution of the number of success in a sequence of N independent trials, where only two outcomes are possible in a trial. The two outcomes are mutually exclusive – head or tail, yes or no, success or failure, an event or no event. Note that this distribution is discrete as we have a discrete number of trials with a discrete number of successes.

Figure a shows the PMF of a binomial distribution with probability of success p = 0.3 and the total number of trials N = 10.

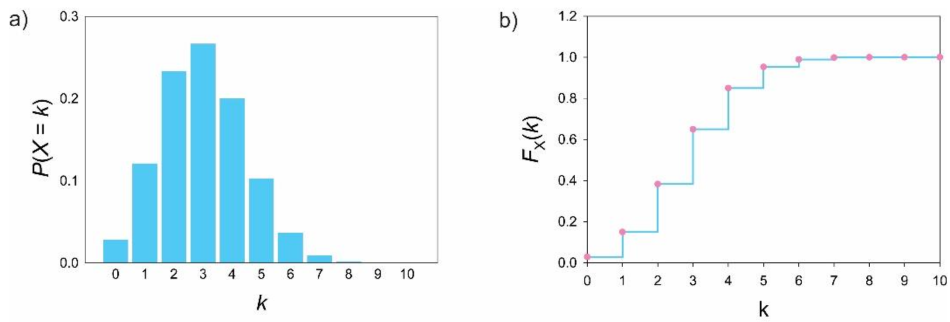

Figure 2: a) Probability Mass Function (PMF) and b) Cumulative Density Function of a binomial distribution with probability of success $p = 0.3$ and the total number of trails $N = 10$.

Now we ask another question. What is the probability that the number of *heads* obtained in four coin tosses is less than equal to three?

This probability is,

$$\begin{aligned} P(H \leq 3) &= P(H=0)+P(H=1)+P(H=2)+P(H=3) \\ &= {}^4C_0 P_H{}^0 P_T{}^4 + {}^4C_1 P_H{}^1 P_T{}^3 + {}^4C_2 P_H{}^2 P_T{}^2 + {}^4C_3 P_H{}^3 P_T{}^1 \end{aligned}$$

Considering $P_H = P_T = 0.5$,

$$P(H \leq 3) = 0.9375$$

Taking the cue from the coin toss, we can again generalize. Let X is a discrete random variable and $a \leq X \leq b$. Then the probability that X is less than equal to k, where $a \leq k \leq b$, is

$$F_X(k) = P(X \leq k) = \sum_{i=a}^{k} P(X=i) \tag{2}$$

$F_X(k)$ in Equation 2 is called the Cumulative Distribution Function (CDF).

If X follows a binomial distribution, we can use the PMF of the binomial distribution to calculate its CDF,

$$F_X(k) = P(X \le k) = \sum_{i=0}^{k} {}^N C_i p^i (1-p)^{N-i} \tag{3}$$

Figure b shows the binomial distribution's CDF with the probability of success p = 0.3 and the total number of trials N = 10. As X is a discrete variable, the CDF is a step plot.

Coin-toss is easy to understand, and calculation using binomial distribution is easy. We will return to binomial distribution time and again. With a few clever assumptions, we can convert a real-life problem into a coin-toss problem and use binomial distribution for its analysis. Take the example of the transcription of a gene. Assume time as discrete, t = 1, 2, 3, ⋯, N. Further assume that either an mRNA is produced or nothing happens at a one-time point.

With these two assumptions, each time point is equivalent to a coin toss and having one mRNA is equivalent to having a *head* in a toss. Therefore, the number of mRNAs produced in $t = N$ is a discrete random number that follows the binomial distribution.

With this formulation, the probability that there will be k number of mRNA at $t = N$ is,

$$P(X = k) = {}^N C_k p^k (1-p)^{(N-k)}$$

Here, p is the probability of the production of one mRNA at a time point.

Expected value, variance, and moments

We have a machine that pops discrete random numbers from one to four following a PMF $P(X = k)$. Suppose it has generated the following numbers: 1, 2, 3, 1, 4, 4, 1.

The arithmetic mean of these numbers is,

$$\begin{aligned}\mu &= \frac{1+2+3+1+4+4+1}{7} \\ &= 1\times\frac{3}{7}+2\times\frac{1}{7}+3\times\frac{1}{7}+4\times\frac{2}{7}\end{aligned}$$

You must have noticed that 3/7, 1/7, 1/7, and 2/7 are the frequencies of 1, 2, 3, and 4, respectively, in the data. So we can say,

$$\mu = \sum_{i=1}^{4} i \times f_i$$

Let's generalize this definition of mean. Suppose the machine can generate random number $X = k$, such that $a \le k \le b$ and the frequency of appearance of a particular random number is f_k. Then the mean of X is,

$$\mu = \sum_{k=a}^{b} k \times f_k$$

Consider frequency is equivalent to probability and make further generalization of this definition of mean. Suppose X is a discrete random variable with PMF $P(X = k)$ and $a \le X \le b$. Then the mean or expected value of X is,

$$E(X) = \sum_{k=a}^{b} kP(X = k) \tag{4}$$

Let us get back to the random numbers generated by our machine. Now we want to calculate the mean of the square of those numbers. The mean of squares of those numbers is,

$$\begin{aligned}\mu &= \frac{1^2 + 2^2 + 3^2 + 1^2 + 4^2 + 4^2 + 1^2}{7} \\ &= 1^2 \times \frac{3}{7} + 2^2 \times \frac{1}{7} + 3^2 \times \frac{1}{7} + 4^2 \times \frac{2}{7} \\ &= \sum_{i=1}^{4} i^2 \times f_i\end{aligned}$$

By now, you must have got the clue for generalization. If X is a discrete random variable with PMF $P(X = k)$, then the mean or expected value of X^2 is,

$$E(X^2) = \sum_{k=a}^{b} k^2 P(X = k) \tag{5}$$

where, $a \le X \le b$.

Following Equation 4 and 5, the m^{th} Moment of a discrete random variable X, with PMF P(X = k) is defined as

$$\mu_m = E\left(X^m\right) = \sum_{k=a}^{b} k^m P\left(X = k\right) \tag{6}$$

where, $a \leq X \leq b$.

Therefore, the mean or expectation of a random variable is the first moment.

Similarly, the central moment is defined as,

$$E\left(\left[X - E\left(X\right)\right]^m\right) = \sum_{k=a}^{b} \left(k - E\left(X\right)\right)^m P\left(X = k\right) \tag{7}$$

The 2nd central moment of a random variable is called the variance,

$$\text{var}\left(X\right) = E\left(\left[X - E\left(X\right)\right]^2\right) = \sum_{k=a}^{b} \left(k - E\left(X\right)\right)^2 P\left(X = k\right) \tag{8}$$

There are several very useful lemmas for expected value and variance. Those are listed here.

If X and Y are random variables such that $Y = aX + b$, where a and b are constants, then,

$$E\left(Y\right) = E\left(aX + b\right) = aE\left(X\right) + b$$

$$\text{var}\left(Y\right) = \text{var}\left(aX + b\right) = a^2 \,\text{var}\left(X\right)$$

X and Y are two independent random variables. If Z is a random variable such that $Z = XY$, then

$$E\left(Z\right) = E\left(XY\right) = E\left(X\right)E\left(Y\right)$$

$$\text{var}\left(Z\right) = \text{var}\left(XY\right) = E\left(X^2\right)E\left(Y^2\right) - E\left(X\right)^2 E\left(Y\right)^2$$

The covariance of two random numbers X and Y is defined as,

$$\text{cov}\left(X, Y\right) = E\left(\left[X - E\left(X\right)\right]\right)E\left(\left[Y - E\left(Y\right)\right]\right) = E(XY) - E(X)E(Y)$$

Therefore, when X and Y are independent, their covariance is zero.

For two random variables X and Y,

$$E(aX + bY) = aE(X) + bE(Y)$$

$$\operatorname{var}(aX + bY) = a^2 \operatorname{var}(X) + b^2 \operatorname{var}(Y) + 2ab\operatorname{cov}(X, Y)$$

$$E(aX - bY) = aE(X) - bE(Y)$$

$$\operatorname{var}(aX - bY) = a^2 \operatorname{var}(X) + b^2 \operatorname{var}(Y) - 2ab\operatorname{cov}(X, Y)$$

Here, a and b are constants.

The expected value and variance help to understand the behavior of a probability distribution and are used to compare different probability distributions. Suppose we know the PMF of a distribution. In that case, we can calculate the expected value and the variance using Equation 4 and 8, respectively.

Here we calculate the expected value of the binomial distribution.

$$\begin{aligned} E(X) &= \sum_{k=0}^{N} k.P(X = k) \\ &= \sum_{k=0}^{N} k\,{}^{N}C_k\, p^k (1-p)^{N-k} \end{aligned} \tag{9}$$

For k = 0, $k\,{}^{N}C_k\, p^k (1-p)^{N-k} = 0$.

So, we can write Equation 9 as,

$$
\begin{aligned}
E(X) &= \sum_{k=1}^{N} k\,{}^{N}C_k\,p^k(1-p)^{N-k} \\
&= \sum_{k=1}^{N} k\frac{N!}{k!(N-k)!}p^k(1-p)^{N-k} \\
&= \sum_{k=1}^{N} k\frac{N(N-1)!}{k(k-1)!(N-k)!}p^k(1-p)^{N-k} \\
&= \sum_{k=1}^{N} \frac{N(N-1)!}{(k-1)!(N-k)!}pp^{k-1}(1-p)^{N-k} \\
&= Np\sum_{k=1}^{N} \frac{(N-1)!}{(k-1)!(N-k)!}p^{k-1}(1-p)^{N-k} \\
&= Np\sum_{k=1}^{N} \frac{(N-1)!}{(k-1)!((N-1)-(k-1))!}p^{k-1}(1-p)^{(N-1)-(k-1)}
\end{aligned} \tag{10}
$$

Let $(N\text{-}1) = b$ and $(k\text{-}1) = a$. Then Equation 10 can be written as,

$$
\begin{aligned}
E(X) &= Np\sum_{a=0}^{b} \frac{b!}{a!(b-a)!}p^a(1-p)^{b-a} \\
&= Np\sum_{a=0}^{b} {}^{b}C_a\,p^a(1-p)^{b-a}
\end{aligned} \tag{11}
$$

${}^{b}C_a\,p^a(1-p)^{b-a}$ is the PMF of a binomial distribution and $\sum_{a=0}^{b} {}^{b}C_a\,p^a(1-p)^{b-a} = 1$. So, from Equation 11, we get the mean or expected value of X as,

$$
E(X) = Np \tag{12}
$$

Similarly, it can be shown that the variance of a binomial distribution is,

$$
\operatorname{var}(X) = Np(1-p) \tag{13}
$$

Multinomial distribution

Replace the coin with a six-sided die. Suppose, all side has an equal probability $p = 1/6$. Throw the die ten times. What is the probability of getting 1, 2, 3, and 4 two times each, and 5 and 6 one time each? To answer this question, we have to use the multinomial distribution.

The multinomial distribution is also discrete. Unlike the binomial, a multinomial problem has more than two outcomes. For the dice throw, there are six possible outcomes.

Suppose there are k mutually exclusive outcomes. x_i is the number of i^{th} outcome in N trails, such that $\sum_{i=1}^{k} x_i = N$ and the propability of i^{th} outcome is p_i, such that $\sum_{i=1}^{k} p_i = 1$. Then the probability of having the $(x_1, x_2, x_3, \cdots, x_k)$ is given by the PMF of the multinomial distribution,

$$P(x_1, x_2, x_3, \cdots, x_n) = \frac{N!}{x_1!x_2!x_3!\cdots x_n!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \cdots p_n^{x_n} \tag{14}$$

For our dice problem, $k = 6$, $N = 10$, $x_1 = x_2 = x_3 = x_4 = 2$, $x_5 = x_6 = 1$. $p = 1/6$ all the outcomes. So, the probability,

$$\begin{aligned} &P(x_1 = 2, x_2 = 2, x_3 = 2, x_4 = 2, x_5 = 1, x_6 = 1) \\ &\quad = \frac{N!}{x_1!x_2!x_3!\cdots x_n!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \cdots p_n^{x_n} \\ &\quad = \frac{10!}{2!2!2!2!1!1!} \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^1 \\ &\quad = 0.00375 \end{aligned}$$

The mean and variance of the i^{th} outcome in multinomial distribution are $E(x_i) = Np_i$ and $\mathrm{var}(x_i) = np_i(1 - p_i)$.

Geometric Distribution

We have used a coin toss to understand the binomial distribution. Each coin toss is called an independent Bernoulli trial. A Bernoulli trial is an experiment whose outcome is random. It can be either of two possible outcomes, like – Yes or No, Success or Failure.

Suppose we are doing a Bernoulli trial experiment, with N = number of trials and p = probability of success. Then the probability that first success happens on k^{th} trial is,

$$P(X=k)=(1-p)^{k-1}p \tag{15}$$

Here, $1 \le k \le N$. $P(X=k)$ is the PMF of Geometric distribution. The geometric distribution is another discrete probability distribution.

Take the example of the transcription of a gene again. Assume that at one discrete-time point an mRNA is produced with a probability p. Then the probability that the first mRNA will be produced at the T^{th} time point is given by the PMF of the geometric distribution,

$$P(t=T)=(1-p)^{T-1}p$$

The CDF of geometric distribution is,

$$\begin{aligned} Fx(k) &= P(X \le k) \\ &= \sum_{i=1}^{k}(1-p)^{i-1}p \\ &= p\sum_{i=1}^{k}(1-p)^{i-1} \end{aligned} \tag{16}$$

For $r \ne 1$, the sum of first n terms of a geometric series is

$$a+ar+ar^2+\cdots+ar^{n-1}=a\frac{1-r^n}{1-r} \tag{17}$$

Considering a = 1 and r = 1 – p, we use Equation 17 to calculate the sum in Equation 16,

$$Fx(k) = p\sum_{i=1}^{k}(1-p)^{i-1}$$
$$= p\frac{1-(1-p)^k}{1-(1-p)}$$
$$= 1-(1-p)^k \qquad (18)$$

The mean and variance of geometric distribution are $E(X) = \frac{1}{p}$ and $\text{var}(X) = \frac{1-p}{p^2}$.

So, when p is small, the expected number of trials required to observe the first success is high. Take the example of transcription. If the transcription rate is very low, the probability of getting an mRNA at a time point will be very small. Then the expected time that we have to wait to see the first mRNA will be very long.

Poisson distribution

Let us continue with the problem of transcription of a gene. Suppose the average rate of transcription is one mRNA per second. What is the probability that eight mRNAs will be produced in ten seconds? Assume that two mRNAs can not be produced simultaneously.

The mean number of mRNAs that will be produced in ten seconds is,

$$\mu = 1\times 10 = 10$$

Now divide the 10 seconds interval into small discreet intervals, say, in the time scale of a second. Therefore, we will have ten intervals. We assume that in each interval, either one mRNA is produced with probability p or nothing happens with probability $(1 - p)$.

With these assumptions, the problem of transcription is a coin toss problem. Here, the total number of trails N = 10, number of success k = 8, and p is the probability of success.

In binomial distribution, the mean is $\mu = Np$. So for our gene expression problem, $p = \frac{\mu}{N} = \frac{10}{10} = 1.$

Using the PMF of binomial distribution, we get,

$$P(k=8) = {}^{N}C_k p^k (1-p)^{N-k}$$
$$= {}^{10}C_8 1^8 (1-1)^2 = 0 \tag{19}$$

That is quite an unexpected answer and unrealistic.

Now use a smaller time scale – milliseconds. In the time scale of milliseconds, we can divide 10 seconds into N = 10000 intervals. Therefore, the probability of having one mRNA in an interval of a millisecond is $p = \frac{\mu}{N} = \frac{10}{10000} = 0.001$.

Again use the PMF of the binomial distribution to calculate the probability of having eight mRNAs in 10 seconds,

$$P(k=8) = {}^{N}C_k p^k (1-p)^{N-k}$$
$$= {}^{10000}C_8 0.001^8 (1-0.001)^{10000-8} = 0.1126 \tag{20}$$

This probability makes some sense. However, how do we know that this result is correct? Result of our calculation changes with the timescale used. Therefore, for another timescale, say nanosecond, we may get a different result. That is unacceptable.

The problem in our approach is elementary. Time is continuous, not discrete. However, we are dividing time into discrete intervals. This discretization is leading to an error in our estimates. If we use an extremely short time interval, the error will be smaller. But how short is short enough?

Suppose we are using a time interval $\Delta t \to 0$ such that the total number of intervals is $N \to \infty$. For an infinitesimally small time-interval, the discrete-time is equivalent to continuous time. Evaluate the general form of the binomial PMF for $N \to \infty$,

$$
\begin{aligned}
P(X=k) &= \lim_{N\to\infty} {}^{N}C_k\, p^k (1-p)^{N-k} \\
&= \lim_{N\to\infty} {}^{N}C_k \left(\frac{\mu}{N}\right)^k \left(1-\frac{\mu}{N}\right)^{N-k} \\
&= \lim_{N\to\infty} \frac{N!}{k!(N-k)!}\left(\frac{\mu}{N}\right)^k \left(1-\frac{\mu}{N}\right)^{N} \left(1-\frac{\mu}{N}\right)^{-k} \\
&= \lim_{N\to\infty} \frac{N(N-1)\cdots(N-k+1)}{k!}\left(\frac{\mu}{N}\right)^k \left(1-\frac{\mu}{N}\right)^{N} \left(1-\frac{\mu}{N}\right)^{-k} \\
&= \frac{\mu^k}{k!} \lim_{N\to\infty} \frac{N(N-1)\cdots(N-k+1)}{N^k} \lim_{N\to\infty}\left(1-\frac{\mu}{N}\right)^{N} \lim_{N\to\infty}\left(1-\frac{\mu}{N}\right)^{-k} \qquad (21)
\end{aligned}
$$

As $\lim_{N\to\infty} N(N-1)\cdots(N-k+1) = N^k$, Equation 21 can be written as

$$P(X=k) = \frac{\mu^k}{k!} \lim_{N\to\infty}\left(1-\frac{\mu}{N}\right)^{N} \lim_{N\to\infty}\left(1-\frac{\mu}{N}\right)^{-k} \qquad (22)$$

$\lim_{N\to\infty}\left(1-\frac{\mu}{N}\right)^{-k} = 1$ and by the definition of the exponential function $\lim_{N\to\infty}\left(1-\frac{\mu}{N}\right)^{N} = e^{-\mu}$. Therefore we get a reduced form of Equation 22,

$$P(X=k) = \frac{\mu^k}{k!} e^{-\mu} \qquad (23)$$

Equation 23 is the PMF of Poisson distribution.

In binomial distribution, for a given μ, when total number of trials $N \to \infty$, the probability of success $p \to 0$. Therefore, Poisson distribution can be considered as a special case of Binomial distribution with very low probability of successes and very large number of trails.

We will use the PMF of Poisson distribution to answer our problem of transcription. Notice that Equation 23 does not have any probability term *p* on the right-hand side. So, we do not need to divide time into short intervals and calculate *p*.

For the transcription problem, μ = 10 and *k* = 8. So the probability,

$$P(k=8)=\frac{\mu^k}{k!}e^{-\mu}=\frac{10^8}{8!}e^{-10}=0.1126$$

You must have noticed that this result is the same as the one we got using binomial distribution and millisecond timescale. However, the result obtained using Poisson distribution is exact and is independent of any time scale assumption.

In general, a discrete random variable X is said to have a Poisson distribution with parameter $\mu > 0$, if for $k = 0, 1, 2, ...,$ the PMF of X is given by,

$$P(X=k)=\frac{\mu^k}{k!}e^{-\mu} \tag{24}$$

Figure 3 shows the PMF of Poisson distribution with $\mu = 5$. For Poisson distribution, the expected value $E(X) = \mu$ and the variance $\mathrm{var}(X) = \mu$.

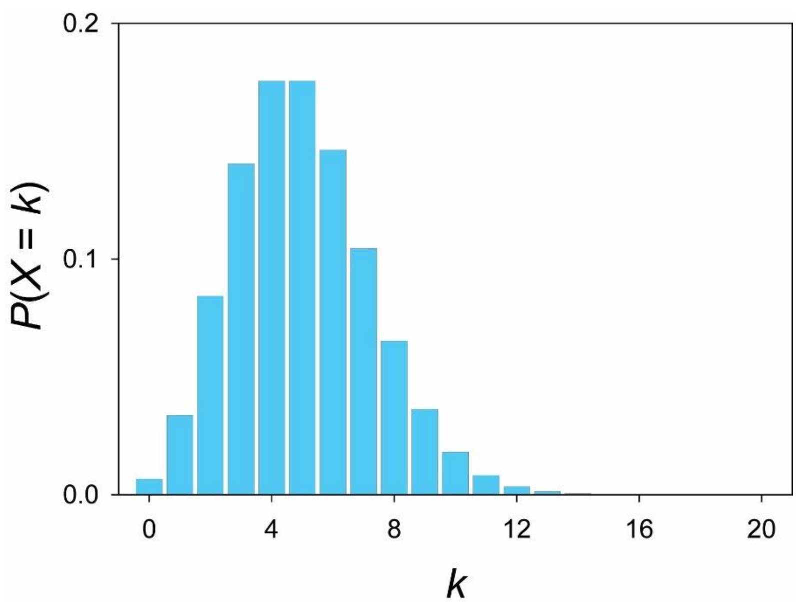

Figure 3: Probability mass function of Poisson distribution with parameter $\mu = 5$.

Continuous Distribution

Some random variables are continuous. A continuous random variable can not have a PMF. Suppose X is a continuous random variable and $m \le X \le n$. $P(X = x)$ is the probability that X is equal to a particular value x. Considering all the possible values of X, we get $\sum_{\text{all } x} P(X = x) = 1$.

However, there are infinite numbers between m and n. Therefore, the probability that X takes a particular value in this range is infinitesimally small; in other words, zero. That is why we do not define PMF for a continuous random variable. Instead, we calculate the probability of X being in a particular interval, say between a and b.

Let there be a function $f_X(x)$ called Probability Density Function (PDF), such that the probability that X will be in the interval $[a, b]$ is,

$$P(a \le X \le b) = \int_a^b f_X(x)\,dx \tag{25}$$

Here, $m \le a < b \le n$.

The PDF $f_X(x)$ must satisfy the following properties,

a) $f_X(x) \ge 0$

b) $P(m \le X \le n) = \int_m^n f_X(x)\,dx = 1$

Note the word 'Density' in PDF. The PDF $f_X(x)$ is not a probability. It is a density – probability per interval. Thefore, unlike a PMF, it can be greater than 1.

The Cumulative Distribution Function (CDF) for the same continuous random variable is defined as,

$$F_X(z) = P(X \le z) = \int_m^z f_X(x)\,dx \tag{26}$$

Like a discrete random variable, a continuous random variable will also have moments. The mth moment's definition remains the same except that in place of discrete summation, we do an integration over a range. So the μ[th] moment of a continuous random variable X, $m \le X \le n$, is

$$E\left(X^{\mu}\right) = \int_{m}^{n} x^{\mu} f_X\left(x\right) dx \tag{27}$$

Therefore, the expectation or mean of this continuous random variable is,

$$E\left(X\right) = \int_{m}^{n} x \, f_X\left(x\right) dx \tag{28}$$

The variance of this random variable is,

$$\text{var}(X) = E\left(\left[X - E\left(X\right)\right]^2\right) = \int_{m}^{n} \left(X - E\left(X\right)\right)^2 f_X\left(x\right) dx \tag{29}$$

The theorems for expected value and variance discussed earlier for discrete random variables are equally applicable for continuous random variables.

There are many well known continuous random distributions. In this chapter, we will discuss a few of those relevant ones.

Uniform distribution

Suppose X is a continuous random variable in the interval [m, n], such that the probabilities of it being in any intervals of equal length within m and n are equal. In other words,

$$P\left(a \le X \le b\right) = P\left(p \le X \le q\right)$$

where $\left(b - a\right) = \left(q - p\right)$, $m \le a < b \le n$ and $m \le p < q \le n$.

The distribution of this random variable is called the Unifrom distribution. If X is a uniformly distributed random variable in the interval [m, n], it is written as $X \sim U(m, n)$.

The PDF of the uniform distribution is,

$$f_X(x) = \begin{cases} \frac{1}{n-m} & \text{for } m \le x \le n \\ 0 & \text{otherwise} \end{cases}$$

The CDF of the uniform distribution is,

$$F_X(x) = \begin{cases} 0 & \text{for } x < m \\ \dfrac{x-m}{n-m} & \text{for } m \le x \le n \\ 1 & \text{for } x > n \end{cases}$$

The PDF and the CDF of this random variable are shown in Figure 4. The mean and variance of uniform distribution are $E(X) = \dfrac{m+n}{2}$ and $\text{var}(X) = \dfrac{(n-m)^2}{12}$, respectively.

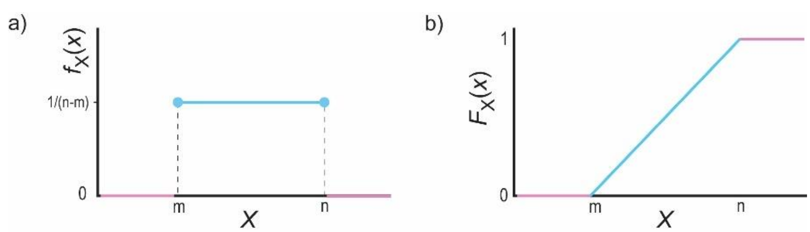

Figure 4: (a) Probability Density Function $f_X(x)$ and (b) Cumulative Distribution Function $F_X(x)$ of an Uniform distribution $U(m,n)$.

Exponential Distribution

The exponential distribution is another continuous distribution. The PDF of an exponential distribution is,

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{when } x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

This PDF has an exponential decay, and the decay rate changes with the parameter λ (Figure 5a). λ is called the rate parameter. A random variable X following exponential distribution is represented as $X \sim \text{Exp}(\lambda)$.

The CDF of an exponential distribution is,

$$F_X(x) = 1 - e^{-\lambda x} \quad \text{when } x \geq 0$$

The CDFs of exponential distributions with different λ are shown in Figure 5b.

For exponential distribution, mean and the variance are $E(X) = \dfrac{1}{\lambda}$ and $\text{var}(X) = \dfrac{1}{\lambda^2}$, respectively.

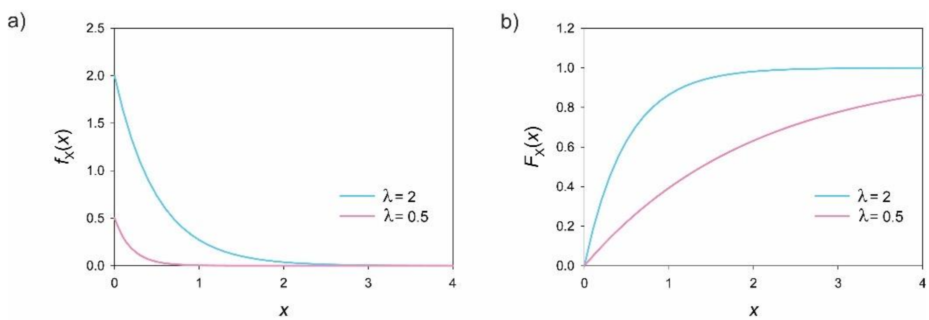

Figure 5: (a) Probability Density Function $f_X(x)$ and (b) Cumulative Distribution Function $F_X(x)$ of an exponential distribution for two different values of the parameter, λ = 2 and 0.5.

Suppose, $X \sim \text{Exp}(\lambda)$ with λ = 0.2. We want to calculate the probability that X is in the interval [m = 1, n = 1.5]. Following Equation 25,

$$P(m \leq X \leq n) = \int_m^n \lambda e^{-\lambda x} dx$$

$$= \lambda \left[-\frac{e^{-\lambda x}}{\lambda} \right]_m^n$$

$$= e^{-m\lambda} - e^{-n\lambda}$$

Using the numerical values for m, n, and λ, $P(1 \leq X \leq 1.5) = e^{-1\times0.2} - e^{-1.5\times0.2} = 0.0779$. This probability is equal to the pink colored area in Figure 6.

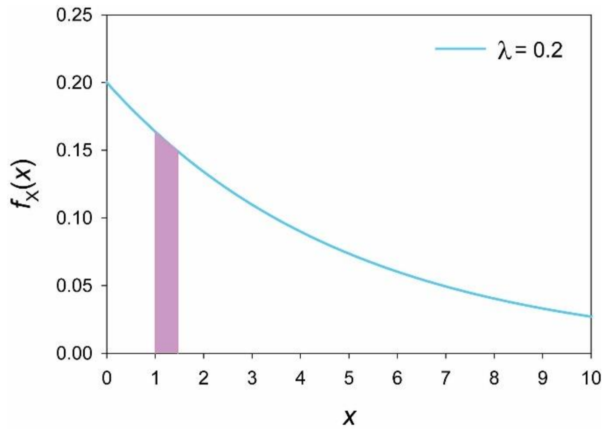

Figure 6: Graphical representation of $P(1 \le X \le 1.5)$ for the exponential distribution $X \sim \text{Exp}(\lambda = 0.1)$. The blue line is the PDF. The pink area is the probability that X is in the interval [1, 1.5].

Normal Distribution

Normal or Gaussian or Laplace-Gauss distribution is a continuous distribution. A random variable X in the interval $[-\infty, +\infty]$ follows Normal distribution $X \sim N(\mu, \sigma^2)$ if the PDF is,

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

here, μ and σ^2 are mean and variance of the distribution, respectively.

Shows the PDF and CDF of a Normal distribution.

Standard Normal distribution $N(0, 1)$, frequently used in stochastic modelling and statistics, has mean $\mu = 0$ and variance $\sigma^2 = 1$. The PDF of the standard Normal distribution is,

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

The normal distribution is a symmetric bell-shaped distribution. The binomial distribution is also symmetric and bell-shaped for the probability of success, $p = 0.5$. The normal distribution can be used to approximate binomial distribution when p is close to 0.5, and the total number of trail N is large.

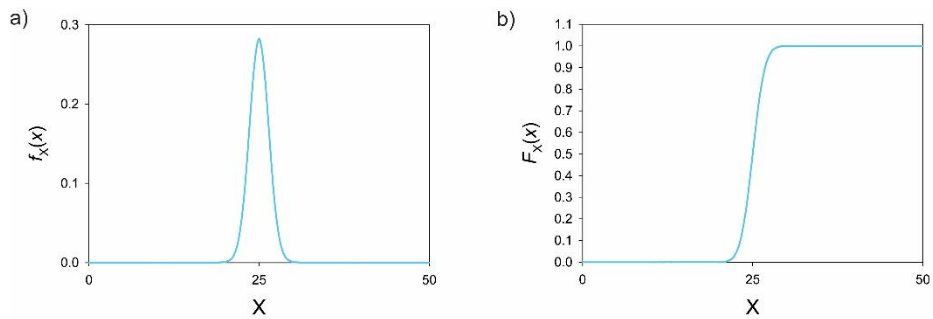

Figure 7: a) PDF and b) CDF of a Normal distribution with $\mu = 25$ and $\sigma^2 = 2$.

The mean and variance of the binomial distribution are Np and $Np(1-p)$. So a binomial distribution can be approximated by a Normal distribution $N(\mu, \sigma^2)$ where $\mu = Np$ and $\sigma^2 = Np(1-p)$. Figure 8 shows the PMF of the binomial distribution for $p = 0.52$ and $N = 40$. The blue line on the figure is the corresponding Normal distribution. The approximation by the Normal distribution is excellent.

Such approximation helps in probability calculations and drawing statistical inferences. We can also approximate a Poisson distribution by Normal distribution. For Poisson distribution, the mean μ is equal to the variance. For large μ, the Normal distribution $X \sim N(\mu, \mu)$ can approximate the Poisson distribution with rate parameter μ.

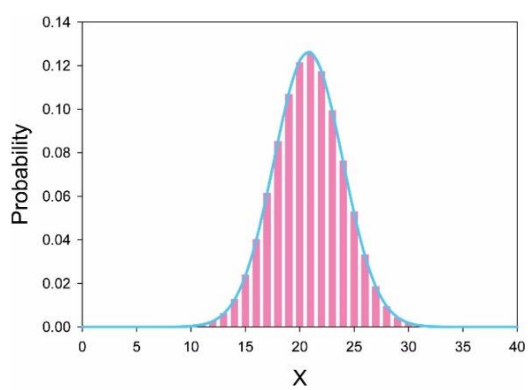

Figure 8: Normal approximation of Binomial distribution. The pink histogram shows the PMF of Binomial distribution for $N = 40$ and $p = 0.52$. The mean $\mu = 20.8$ and variance $\sigma^2 = 9.984$. The blue line is the Normal distribution $X \sim N(\mu = 20.8, \sigma^2 = 9.984)$.

Exercises

1. What is the probability that the number of *heads* obtained in four consecutive coin tosses is greater than one but less-equal to three? Consider $P_H = P_T = 0.5$.

2. X and Y are two random variables. Prove that $\mathrm{cov}(aX, Y) = a.\mathrm{cov}(X, Y)$.

3. X and Y are two random variables. Prove that $E(X+Y) = E(X) + E(Y)$.

4. Draw the diagrams of the PMF of Poisson distribution for $\mu = 1, 5$, and 10.

5. Some cells types in a tissue represent only a tiny portion of the total cell population, such as hematopoietic stem cells in the blood. Imagine that a particular cell type is present at a rate of 1 cell in 100000. If we are given a sample with 50000 cells, what is the probability of finding exactly 5 of these rare cells?

6. The probability of having a mutation in a gene in a cell is 1/1000. If we randomly sequence ten different genes of the same cell, what is the probability that any one of these genes will have a mutation? (Here, we are not bothered about the number of mutations in a single gene or whether the mutation will change amino acid or not.)

7. Calculate the CDF of a continuous distribution with PDF $f_X(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}$. Here, $x \geq 0,\ k > 0,\ \lambda > 0$.

8. M332 is a dye that gives red fluorescence when it is in the cytoplasm of a cell. 10^6 cells are incubated with a solution of this dye for 30 minutes. The number of dye molecules in this staining solution is more than the number of cells. The dye crosses the membrane and enters the cell's cytoplasm at a rate of 1 molecule per 10 sec. Assume that one dye enters at a time.

   After 30 minutes of incubation, cells were washed, and we measured the number of molecules of the dye in each cell. It was observed that the uptake of the dye is not uniform. What will be the mean number of dye molecules per cell? What will be the variance for the number of dye molecules per cell?

9. A malfunctioning clock stops at any time of the day, randomly, without any particular bias. What is the probability that it will stop between 1 PM and 1:15 PM?

10. Let,

$$f_X(x) = \begin{cases} kx & \text{when } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Find k so that $f_X(x)$ is a valid PDF for a continuous random variable X.