Introduction to Stochastic Processes

Stephanie Palmer
sepalmer@uchicago.edu

Graham Smith
grahams@uchicago.edu

BSD qBio^3 boot camp @ MBL

Goal: This tutorial will cover basic probability theory, to give you a set of tools to model variability in biological data. You will also be able to understand and interpret common comparisons made between biological data and the behavior of standard stochastic processes, such as a Poisson process. We will derive equations for the mean and variance of a Poisson process, to build your intuition for these results instead of simply presenting these equations as facts to memorize. We will explore the limiting case where a Poisson distribution approaches a Gaussian distribution, another useful result of probability theory often used in biology.

1 Randomness in biology

This tutorial highlights an important and pervasive aspect of biological systems: stochasticity. (NB: 'stochasticity', 'variability', 'uncertainty', 'noise', and 'fluctuations' will all be used interchangeably here, though most of these terms have more technical and specific uses in other contexts.) Many of the variables that we observe in biological recordings fluctuate, sometimes because we cannot control all the states of the external and internal world of the organism, other times because thermal noise and other microscopic factors make the state of the biological system we interrogate inherently noisy. It is useful to model not only a median value for a fluctuating variable, but the full shape of its distribution of values.

For example, if we observe the firing of neurons in the brain to repeats of the same external stimulus, the precise times of spikes will vary between repeats. By fitting the statistics of this noise to models, we deepen our understanding of the neural response.

To build your intuition about quantifying uncertainty, let's start with a toy problem I first encountered in David MacKay's lectures on information theory and inference.

2 Testing your intuition: the bent coin lottery

A biased coin is used to generate sequences of digits, 1 for heads, 0 for tails, in a lottery. The coin is tossed 25 times to determine the winning sequence. The probability of heads is 0.1. Tickets for the lottery cost $\$1$ and the prize is $\$10,000,000$.

Exercise 2.1

1. You are only allowed to purchase one ticket. Which ticket would you buy?

Solution. The most likely ticket is the all-zeros ticket

$$00000\,00000\,00000\,00000\,00000,$$

with a probability of $P(000\ldots0) = (0.9)^{25} = 0.7178$.

To compute the probability of any given sequence, you multiply together the probability of every coin flip, so

$$P(000\ldots0) = P(0)P(0)P(0)\ldots P(0).$$

How do we know this ticket has the highest probability? For any 0 that we replace by a 1, the probability correspondingly changes one of the 25 $P(0)$'s to $P(1)$, since $P(1)$ is 0.1 (which is less than $P(0) = 0.9$), then the resulting product must be less than the probability of all-zeros.

2. How many tickets to would you have to buy to cover every possible outcome?

   Solution. The total number of tickets is $2^{25} = 33,554,432$.

3. Is this lottery worth playing?

   Solution. At first glance, the prize money ($10,000,000) is less than the cost of buying all the tickets ($33,554,432), so we might think the lottery is not worth playing. However, not all tickets are equally likely, so we don't want to buy all the tickets. We only want to buy the most likely tickets. This doesn't guarantee that there are enough likely tickets to make lottery worthwhile, but we will show later that there are.

3 Binomial distribution

Each flip of a coin like this with probability, p, of heads is an example of a Bernoulli trial, the general term for an experiment with only two output states, success or failure. The number of heads in the sequence of independent coin flips generated by our lottery will follow a binomial distribution.

Exercise 3.1

1. Write down the probability of observing k heads in n coin flips, if the probability of heads is p.

   Solution. The probability of observing k heads in n coin flips is the same as the probability of observing **any** sequence of length n with k heads. We can find this probability by adding together the probabilities of every **particular** sequence of length n with k heads. That's a lot of terms to sum up. Luckily, the probability of a sequence of length n with k heads is $p^k(1-p)^{n-k}$, regardless of how the heads and tails are ordered. Why? As in Exercise 1.1, we find the probability of a sequence by multiplying the probabilities of each coin flip. Since the order of multiplication doesn't matter, the product is always $p^k(1-p)^{n-k}$.

   Now we just need to count how many sequences there are of length n with k heads, since the sum of their probabilities will equal the product of their count with $p^k(1-p)^{n-k}$. It turns out that counting sequences like this comes up a lot, so there is a notation for their count, $\binom{n}{k}$, pronounced "n choose k," and known as a binomial coefficient. This binomial coefficient counts the number of ways you can choose k out of n objects. (Why is that equivalent to the number of sequences of length n with k heads?) It is defined as

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

Then we have the probability of k heads in a sequence of n coin flips is

$$\binom{n}{k} p^k (1-p)^{n-k}$$

This is the binomial distribution. Rather than memorize this particular form, remember how to write it down as the product of intuitive terms. If we consider the limit of a very small p, we can relate the binomial distribution to the Poisson distribution.

4 Poisson distribution

The Poisson distribution describes the probability of finding k events in a fixed interval if we know the rate of occurrence of these events, λ, in that interval. In terms of the variables we have been working with for the bent coin lottery,

$$\lambda = p * n. \tag{4.1}$$

We are going to take the limit where p is very small and n is very large, but their product remains fixed.

Exercise 4.1

1. Derive an expression for the probability of observing k heads in n tosses in the limit of small p and large n.

   Solution. We just showed that the binomial distribution is given by

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

   Equation 4.1 tells us $p = \lambda/n$, so

   $$P(k) = \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

   To make things easier, let's split that expression up and look at what happens to each part when we make n large. First, we'll take the first two terms:

   $$\frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k = \frac{n \cdot (n-1) \cdot (n-2) \dots (n-k+1) \cdot (n-k) \cdot (n-k-1) \dots 1}{k! \cdot (n-k) \cdot (n-k-1) \dots 1} \left(\frac{\lambda}{n}\right)^k$$

This expansion makes it more apparent that the last $n-k$ terms of $n!$ are the same as $(n-k)!$, so we can simplify the above equation to

$$\begin{aligned}
\frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k &= \frac{n\cdot(n-1)\cdot(n-2)\dots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k \\
&= \frac{n\cdot(n-1)\cdot(n-2)\dots(n-k+1)}{k!\cdot\underbrace{n\cdot n\dots n}_{\text{k times}}}\lambda^k \\
&= \frac{n}{n}\cdot\frac{n-1}{n}\dots\frac{n-k+1}{n}\cdot\frac{\lambda^k}{k!}.
\end{aligned}$$

As $n\to\infty$, all but the last fraction goes to one, so that the product is

$$\frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k \xrightarrow[n\to\infty]{} \frac{\lambda^k}{k!}$$

Now let's look at the other half of the original equation,

$$\left(1-\frac{\lambda}{n}\right)^{n-k} = \left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-k}.$$

The part with $-k$ in the exponent goes to one as n gets large. For the other part, we need to remember from calculus that $e^x = \lim_{n\to\infty}(1+\frac{x}{n})^n$, so

$$\left(1-\frac{\lambda}{n}\right)^{n-k} \xrightarrow[n\to\infty]{} e^{-\lambda}.$$

Combining both parts, we have

$$P(k) \xrightarrow[n\to\infty]{} \frac{\lambda^k}{k!}e^{-\lambda}$$

We have just written down the Poisson distribution. You will see this used as a model for biological variability again and again, either explicitly or implicitly. It is important to think about whether or not it is a good model for the system under study each time you come across it or are deciding to use it for your own research.

4.1 Intervals between events

We can also write down the distribution of intervals between events in a Poisson process. This distribution has an exponential form.

Exercise 4.2

1. Derive an expression for the distribution of an interval, τ, between two events in a Poisson process with rate, λ, in this interval.

   Solution. Let's start by translating this problem back to the language of binomial distributions of sequences. Imagine we divide time into small, discrete bins of length Δt, such that the time interval of length τ is divided into n bins. We'll say the bins are so small that we can ignore the possibility of more than one event occurring in a single bin, so that the bins are essentially Bernoulli random variables. In this scheme, the probability that we observe an interval τ between two events is the same as the probability that we observe n bins with no event followed by a bin with an event.

   Let r denote the rate per unit time of this process, so that $\lambda = r\tau$, and the probability for the Bernoulli random variable describing each bin is $r\Delta t$.

   In this case $n \to \infty$ provides the best approximation (since we don't really want to discretize time), so the probability that we observe n bins with no event is given by Poisson $P(0) = e^{-\lambda}$. The probability of one event in the last time bin is simply a Bernoulli random variable with probability $r\Delta t$. Therefore we can write the density function of the interval, τ, as

$$\frac{P(\tau)}{\Delta t} = re^{-r\tau}.$$

5 Gaussian distribution

A Gaussian or 'normal' distribution (also called a bell-curve) of a variable x with mean μ and variance σ takes the form

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5.1}$$

Exercise 5.1

As λ gets very large, show that the Poisson distribution approaches a Gaussian distribution with mean λ and variance λ.

Solution. To do this, we need to use moment generating functions (MGFs), which are an alternative (often more useful) way of describing a probability distribution. An MGF is the expectation of e^{tX}, where X is the random variable. We will show that the standardized Poisson variable approaches the standardized Gaussian variable as $\lambda \to \infty$. So, letting X be our Poisson random variable with mean and variance λ,

$$
\begin{aligned}
E\left[e^{t\frac{X-\lambda}{\sqrt{\lambda}}}\right] &= \exp\left(-t\sqrt{\lambda}\right) E\left[\exp\left(\frac{tX}{\sqrt{\lambda}}\right)\right] \\
&= \exp\left(-t\sqrt{\lambda}\right) \sum_{k=0}^{\infty} P(k) \exp\left(\frac{tk}{\sqrt{\lambda}}\right) \\
&= \exp\left(-t\sqrt{\lambda}\right) \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} \exp\left(\frac{tk}{\sqrt{\lambda}}\right) \\
&= \exp\left(-t\sqrt{\lambda}-\lambda\right) \sum_{k=0}^{\infty} \frac{\left(\lambda \exp\left(\frac{t}{\sqrt{\lambda}}\right)\right)^k}{k!} \\
&= \exp\left(-t\sqrt{\lambda}-\lambda\right) \exp\left(\lambda \exp\left(\frac{t}{\sqrt{\lambda}}\right)\right) \\
&= \exp\left(\lambda \exp\left(\frac{t}{\sqrt{\lambda}}\right) - t\sqrt{\lambda} - \lambda\right) \\
&= \exp\left(-t\sqrt{\lambda} - \lambda + \lambda\left(1 + t\lambda^{-1/2} + \frac{t^2\lambda^{-1}}{2!} + \dots\right)\right) \\
&= \exp\left(-t\sqrt{\lambda} - \lambda + \lambda + t\lambda^{1/2} + \frac{t^2}{2!} + \frac{t^3\lambda^{-1/2}}{3!} \dots\right) \\
&= \exp\left(\frac{t^2}{2!} + \frac{t^3\lambda^{-1/2}}{3!} \dots\right) \\
&\xrightarrow[\lambda\to\infty]{} \exp\left(\frac{t^2}{2!}\right).
\end{aligned}
$$

This limit is the MGF of a standard Gaussian distribution! Therefore the Poisson distribution approaches the Gaussian distribution with the same mean and variance λ, for large λ.

Exercise 5.2

1. Derive the mean of a Poisson distribution with rate λ:

   Solution. If K is our random variable with a Poisson distribution,

$$
\begin{aligned}
E[K] &= \sum_{k=0}^{\infty} kP(k) \\
&= \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} \\
&= 0 + e^{-\lambda}\sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \\
&= 0 + \lambda e^{-\lambda}\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}
\end{aligned}
$$

We now substitute $j = k - 1$ to shift the sum,

$$\begin{aligned}
&= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda
\end{aligned}$$

just as we wanted.

2. Derive the variance of a Poisson distribution with rate λ:

Solution.

$$\begin{aligned}
\sigma^2 &= E\left[K^2\right] - E\left[K\right]^2 \\
&= \sum_{k=0}^{\infty} k^2 P(k) - \lambda^2 \\
&= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} - \lambda^2 \\
&= 0 + \lambda e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} - \lambda^2 \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \left((k-1) \frac{\lambda^{k-1}}{(k-1)!} + \frac{\lambda^{k-1}}{(k-1)!} \right) - \lambda^2 \\
&= \lambda e^{-\lambda} \left(\sum_{k=1}^{\infty} (k-1) \frac{\lambda^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right) - \lambda^2 \\
&= \lambda e^{-\lambda} \left(0 + \sum_{k=2}^{\infty} (k-1) \frac{\lambda^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right) - \lambda^2 \\
&= \lambda e^{-\lambda} \left(\lambda \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right) - \lambda^2
\end{aligned}$$

and we again substitute $j = k - 1$ and $l = k - 2$ to shift the sums,

$$\begin{aligned}
&= \lambda e^{-\lambda} \left(\lambda \sum_{l=0}^{\infty} \frac{\lambda^l}{l!} + \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \right) - \lambda^2 \\
&= \lambda e^{-\lambda} \left(\lambda e^{\lambda} + e^{\lambda} \right) - \lambda^2 \\
&= \lambda^2 + \lambda - \lambda^2 \\
&= \lambda
\end{aligned}$$

just as we wanted.

These quantities are often summarized as the ratio of the variance and the mean, or Fano Factor (FF). The FF for a Poisson process is clearly equal to one.

Exercise 5.3

1. Does observing an FF $= 1$ in data mean that the underlying stochastic process is a Poisson process?

   Solution. No! Suppose you have a process such that the waiting time, W, is either 2 or 0 with equal probability. The average waiting time is $E[W] = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 2 = 1$. The variance of the waiting time is also 1 because

$$\begin{aligned} \sigma^2 &= E\left[(W - E[W])^2\right] \\ &= \sum_{w=0,2} p(w)(w-1)^2 \\ &= \frac{1}{2}(0-1)^2 + \frac{1}{2}(2-1)^2 \\ &= 1 \end{aligned}$$

   Since the mean and variance are both 1, then the Fano factor is also 1. However, the process is certainly not Poisson. Therefore we have constructed a counterexample that shows that a Fano factor of 1 does not imply that the underlying process is Poisson.

6 Winning the bent coin lottery

Now that we have all of these distributions at our fingertips, let's return to the bent coin lottery.

Exercise 6.1

1. Derive how many tickets you need to buy to guarantee yourself a 99% chance of winning.

   Solution. Let's use the Gaussian approximation for the probability of getting k heads out of n coin flips. Recall that the probability of one head is $p = 0.1$ and the number of coin flips is $n = 25$. Then we can calculate the mean number of heads is $np = 2.5$, and the variance is $np(1-p) = 2.25$, which gives a standard deviation of $\sigma = \sqrt{2.25} = 1.5$.

   The nature of a Gaussian distribution is that 95% of the distribution is within two standard deviations of the mean. That means that if we get all the lottery tickets less than two standard deviations above the mean, we will have at least a 95% chance of winning. In fact, we will have a higher chance of winning because the left tail of the distribution is heavier than a Gaussian, since the probability of $k = 0$ is so high. That means we want to buy up to 5.5, which doesn't really make sense since you can't have half a head. Instead, rounding up we guess that if you buy all tickets with 6 or fewer heads, we will have $\sim 99\%$ chance of winning.

   Now we can count how many tickets we need to buy according to our approximation:

$$\binom{25}{0} + \binom{25}{1} + \binom{25}{2} + \ldots + \binom{25}{6} = 1 + 25 + 300 + \ldots 177,100$$

   That's way less than $10,000,000$, so at a dollar per ticket the lottery is definitely worth playing!

6.1 Generating samples of a stochastic process

When modeling biological systems, it is often necessary to generate sequences from a Poisson or other stochastic process. We did this to generate our draws from the bent coin lottery.

6.2 Markov processes

One feature of the stochastic processes we have been considering today is that they are independent. A flip of the coin doesn't depend on the flip before, or any of the other previous flips. In biological systems, what came before often influences a fluctuating quantity. For example, having spiked, a neuron is unable to spike for a millisecond or two. Modeling this type of variability falls requires using stochastic processes that have an explicit history dependence. Markov processes depend only on the previous time step, in generating the current state.