

Topic 3: Poisson Processes

Learning Objectives

1. Review the Exponential, Erlang, and Poisson distributions.
 - a. Convolution of two exponentials is an exponential.
2. Define a Poisson process and describe:
 - a. The time to the next event
 - b. The time to the n^{th} event
 - c. The total number of events in time T
 - d. The superimposition of two Poisson processes
3. Propose a Poisson process describing a natural system and use it to draw biological conclusions
4. Define a non-homogenous Poisson process and calculate its moments
5. Define a compound Poisson process and derive its first two moments.
6. Simulate a Poisson, non-homogeneous Poisson, and compound Poisson process.
7. Use a Gillespie simulation to simulate a biological process

Lecture 3.1 Poisson Processes

Definitions

Let $\{t_1, t_2, \dots\} \stackrel{\text{iid}}{\sim} \text{Exp}(\lambda)$ be waiting times between events such that $T_n = \sum_{j=1}^n t_j$ is the waiting time until the n^{th} event and define $N(\tau) = \max\{n : T_n < \tau\}$ then $N(\tau)$ is a **Poisson process**.

This is a bit of an awkward definition. Instead, we can practically re-word this by considering a process where events occur at a constant rate through time. Each time an event occurs we add 1 to our counter N . The resulting dynamics of N is a Poisson process.

$N(\tau)$ is called a Poisson process because:

$$N(\tau) \sim \text{Poi}(\tau\lambda)$$

Recall that the Poisson distribution describes the number of events that occur in 1 unit of time given that they occur at rate λ , so we can describe the number of events that occur in τ units of time by multiplying the rate by τ (in other words one τ unit of time).

Note that time here is assumed to be "continuous" which in this context means that two events can not occur simultaneously.

Initial condition

Note that the initial condition of a Poisson process, $N(0) = 0$. as no events have occurred yet.

Example:

Seed banks are pools of seeds present in the soil that have not yet germinated. The existence and turnover of seed banks have important implications for genetic diversity and conservation. Suppose that seed from the seed bank germinates at a constant rate of $\lambda = 500$ seeds/year

1. How long does it take for the first seed to germinate?

The waiting time until the first event is exponentially distributed with rate λ . The expected time is $\frac{1}{\lambda} = 0.002$.

2. How long does it take for 50 seeds to germinate?

The waiting time until the $k = 50$ event is Erlang distributed. The mean of this distribution is $\frac{k}{\lambda} = 0.1$.

3. How many seeds have germinated by the end of the second year?

The number of events in a fixed amount of time is Poisson distributed with rate $\lambda * T$ where T is the total time elapsed. The mean here then is $500 * 2 = 1000$.

4. Suppose that a researcher is monitoring the germination of a seed bank. How many seeds have germinated at the beginning of the experiment?

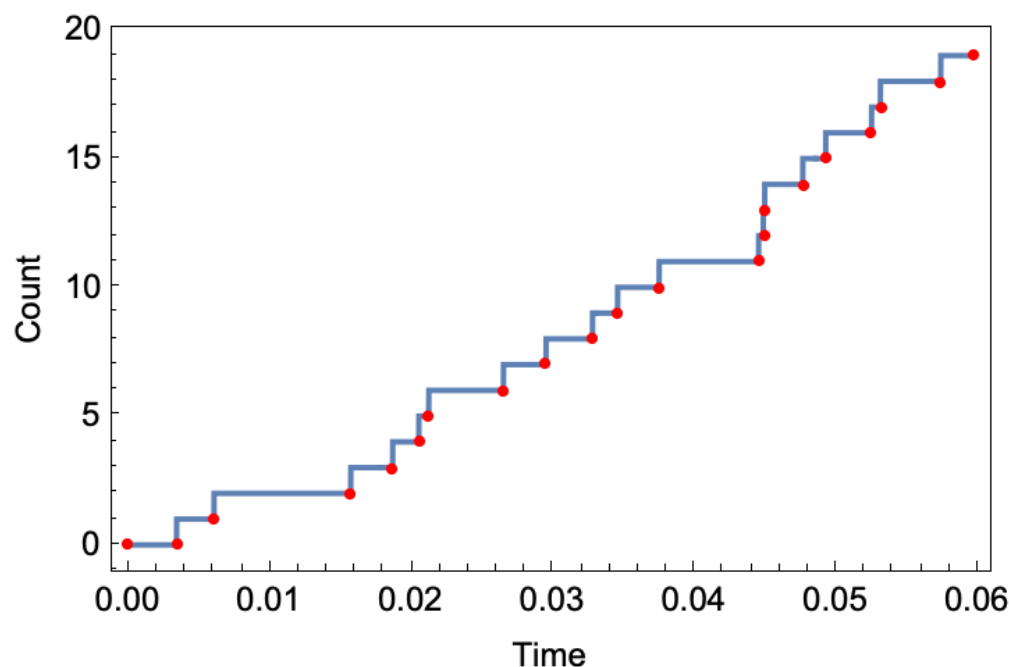
A seed bank refers to all ungerminated seeds at some time. By definition then, no seeds have germinated at the start of the experiment.

5. Simulate the first 20 events in the Poisson Process described above.

Step 1: Draw random waiting times $t_i \sim \text{Exp}(\lambda = 500) \quad i \in \{1, 2, \dots, 20\}$.

Step 2: Accumulate waiting times: $T_i = \sum_{j=1}^i t_j$

Step 3: Assign counter $N(x) = i - 1 \quad T_{i-1} \leq x < T_i$ where $T_0 = 0$



When studying Poisson processes, we are interested then in both the *number of events* and the *distribution of waiting times* between events:

While one way to think of a Poisson process is as events occurring in time, another way to think about them is occurring in space.

Example: tree density

Suppose that along a linear transect, we observe 400 trees/km.

1. What is the average spacing between trees?

This waiting space is exponentially distributed $1/400 = 0.0025 \frac{\text{km}}{\text{tree}} = 2.5m$

2. Suppose that the optimal spacing between trees is 5m, what proportion of trees are closer than this optimal spacing?

$$CDF_{exp}(X \leq 0.005 | \lambda = 400) = 86.5\%$$

3. Given a researcher is surveying 5km long elevational transects how many trees should be observed along each?

Markov Property

Poisson processes are **markovian**. In continuous time the Markov Property can be expressed as:

$$\Pr(N(s+t) = n | N(s) \forall s < t) = \Pr(N(s+t) = n | N(s))$$

This means that the Poisson process $N(t)$ with rate λ has **independent increments** such that for time. $t_0 < t_1 < \dots < t_n$

$$N(t_n) - N(t_{n-1}), N(t_{n-1}) - N(t_{n-2}), \dots, N(t_2) - N(t_1), \text{ and } N(t_1) - N(t_0)$$

are independent and $M(t) = N(t+s) - N(s)$ is Poisson distributed with rate λt .

If we think of a Poisson process as events occurring along a timeline, then any subsection of the number of events that occurs within non-overlapping subsections of that timeline is also Poisson distributed.

Compound Poisson Process

Definition

Let $N(\tau)$ be a Poisson process; in other words $N(\tau)$ is the number of events in τ time given that events occur at rate λ . A **Compound Poisson Process** describes the random arrival of events with random magnitude $X \sim \mathcal{D}ist$. Such that the poisson counter is given by $Y(\tau) = \sum_{i=1}^{N(\tau)} X_i$.

Moments

Mean:

$$E[Y(\tau)] = E[D_1 + D_2 + \dots + D_{N(\tau)}] = \lambda \tau E[D]$$

Variance:

Recall the law of total variance

$$\text{Var}(Y) = \underbrace{E_N[\text{Var}(Y|N)]}_{\text{Variance within}} + \underbrace{\text{Var}_N[E[Y|N]]}_{\text{Variance Between}}$$

Example: Synchronous Life Histories

In ecological studies, the compound Poisson process is frequently used to model the dynamics of a population, where events of interest occur in clusters rather than independently. Consider a scenario where the reproduction events of a particular species are modelled as a compound Poisson process. The number of offspring produced in each reproductive event follows a binomial distribution, and these events occur according to an underlying Poisson process.

1. Define a compound Poisson process describing the number of offspring produced by time t .

Let $N(t)$ be the number of reproductive events by time t , where events are assumed to occur randomly at rate λ . Let $X \sim \mathcal{B}(n, p)$ be a random variable giving the number of offspring produced per reproductive event. Then the number of offspring produced by time t is given by:

$$Y(t) = \sum_{i=0}^{N(t)} X_i$$

2. What is the expected number of offspring after time $t = 1$ year? How certain is this projection?

The expected number of offspring is given by $\lambda \times t \times E[X] = \lambda np$.

The certainty of this outcome is determined by the variance:

$$\text{Var}(Y) = E_N[\text{Var}(Y|N)] + \text{Var}_N[E[Y|N]]$$

Breaking this down:

$$\text{Var}(Y|N) = \text{Var} \left(\sum_{i=0}^{N(t)} X_i \mid N(t) = N \right)$$

Given that each X_i is independent the variance in the sum is the sum of the variances:

$$\text{Var}(Y|N) = \sum_{i=0}^N \text{Var}(X_i) = \sum_{i=0}^N np(1-p) = Nnp(1-p)$$

Now we can take the expectation over N :

$$E_N [\text{Var}(Y|N)] = E_N [Nnp(1-p)] = np(1-p)\lambda$$

For the other term:

$$E[Y|N] = E \left[\sum_{i=0}^N X_i \right] = \sum_{i=0}^N E[X_i] = Nnp$$

Hence:

$$\text{Var}_N [E[Y|N]] = \text{Var}_N [Nnp] = np\lambda$$

Recall that the mean and variance of a Poisson distribution are both equal to λ . Combining we have:

$$\text{Var}(Y) = np(1-p)\lambda + np\lambda = np\lambda(2-p)$$

Comparing the mean and the variance:

$$\frac{\text{Var}(Y)}{E[Y]} = \frac{np\lambda(2-p)}{\lambda np} = 2-p$$

Discussion: What does this tell us about the certainty of the outcome?

Heterogenous Poisson Processes

So far we have assumed that the rates at which events occur are independent of time (time-homogenous processes). This is certainly not always the case (e.g., due to seasonality). We can incorporate time-dependence, events occur at rate $\lambda(t)$. If it is currently time t_0 , the probability that the next event occurs at time $t_1 = t_0 + \Delta t$ is given by the **Heterogenous Exponential distribution**

$$\text{Pr}(\Delta t|t_0) = \lambda(t_0 + \Delta t)e^{-\int_{t_0}^{t_0+\Delta t} \lambda(x)dx}$$

Discussion: How does this distribution relate to the exponential distribution?

Lecture 3.2 Simulating Poisson Processes

Thinning and Superposition

Thinning and superposition describe the convenient property of Poisson processes that the sum of multiple Poisson processes is a Poisson process (Superposition) and that a single Poisson process can be decomposed into several processes (thinning).

Theorem: Suppose that $N_1(t), N_2(t), \dots, N_k(t)$ are independent Poisson processes occurring with rates $\lambda_1, \lambda_2, \dots, \lambda_k$ respectively. Then $N_1(t) + N_2(t) + \dots + N_k(t)$ is a Poisson process with rate $\lambda_1 + \lambda_2 + \dots + \lambda_k$.

As above time/space here is assumed to be "**continuous**" which in this context means that two events can not occur simultaneously or in the same place.

We can show this is true by considering the case when $k = 2$ and use the Markov property of Poisson distributions to show that this holds by induction. Let's define $M(t) = N_1(t) + N_2(t)$ then at time $t = 0$:

$$M(0) = N_1(0) + N_2(0) = 0$$

Hence M satisfies the initial conditions of a Poisson process. Let $Y = N_1(t + s) - N_1(s)$ and $Z = N_2(t + s) - N_2(s)$ then:

$$\begin{aligned} \Pr(M(t + s) - M(t)) &= \Pr(Y + Z = n) = \sum_{m=0}^n \Pr(Y = m) \Pr(Z = n - m) \\ &= \sum_{m=0}^n e^{-\lambda_1 t} \frac{(\lambda_1 t)^m}{m!} e^{-\lambda_2 t} \frac{(\lambda_2 t)^{n-m}}{(n-m)!} \\ &= e^{-(\lambda_1 + \lambda_2)t} \frac{(\lambda_1 t + \lambda_2 t)^n}{n!} \sum_{m=0}^n \frac{n!}{m!(n-m)!} \frac{(\lambda_1 t)^m (\lambda_2 t)^{n-m}}{(\lambda_1 t + \lambda_2 t)^{m+n-m}} \\ &= e^{-(\lambda_1 + \lambda_2)t} \frac{(\lambda_1 t + \lambda_2 t)^n}{n!} \sum_{m=0}^n \binom{n}{m} \underbrace{\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^m}_{p^m} \underbrace{\left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-m}}_{(1-p)^{n-m}} \\ &= e^{-(\lambda_1 + \lambda_2)t} \frac{(\lambda_1 t + \lambda_2 t)^n}{n!} \end{aligned}$$

which is a poisson distribution with rate $\lambda_1 + \lambda_2$.

The probability that the next event is of type 1 is: $p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$.

Example: Mutations

Genetic mutations can be considered as occurring as random events over time. Suppose a population experiences two types of mutations: insertion/deletions (indels) which typically occur due to DNA repair and pair mutations resulting from replication errors.

1. Propose a Poisson process for the occurrence of mutations across a genome from generation to generation.

Discussion: What assumptions do we have to make about how mutations occur to think about them as a Poisson process?

Let μ_1 be the rate at which indels occur per megabase per generation and μ_2 the rate at which point mutations occur.

Let's assume that the genome is large enough that is unlikely for two mutations to occur in the same position in the genome in the same generation.

Then the number of mutations in a genome of X megabases/generation is Poisson distributed with rate $\mu_1 X + \mu_2 X$.

The occurrence of mutations in space (along the genome) and time (generations) can be modelled as a Poisson process with the total rate $\lambda = gX(\mu_1 + \mu_2)$

2. Given that indels occur at a rate of 15 mutations per 100 megabases per generation and point mutations occur at a rate of 1 mutation per megabase per generation. What is the probability that a new mutation is an indel?

Here we only care about the relative mutation rates: $\mu_1 = 15/100 = 0.15$ and $\mu_2 = 1$.

$$\frac{0.15}{0.15 + 1} = 0.13$$

or 13% of mutations

The Gillespie Algorithm

The ability to thin and superimpose Poisson processes forms the basis of the **Gillespie Algorithm** which is a general algorithm for simulating **continuous time stochastic processes**.

The Gillespie algorithm was originally developed (Gillespie 1976) to describe the dynamics of a well-stirred chemical system with stochastic (random) reaction events. But is used widely in population biology as it is particularly useful for studying the dynamics of continuous time discrete-state stochastic processes.

For the Gillespie algorithm we express the state of the system with a vector. For example if we are following the number of prey, X and the number of predators Y we have: $[X, Y]$.

Example: Lotka-Volterra Predator-Prey dynamics

Interactions between a predator Y and its prey X are often modelled with the following system of differential equations known as the Lotka-Volterra model:

$$\begin{aligned}\frac{dX}{dt} &= \alpha X - \beta XY \\ \frac{dY}{dt} &= \beta c XY - \gamma Y\end{aligned}$$

1. What do you think the parameters mean? And what are reasonable values for them?

- α Prey birth rate. We can always set one parameter to 1 (as this defines the time scale).
- β Predation rate. If predators and prey have population sizes of 100 then for this term to be on the same order as birth/death we have to have $\beta = \mathcal{O}(0.01)$. Let's define $\beta = 0.02$
- c A conversion parameter from captured prey to predator offspring. The conversion rate should be less than 1, say $c = 0.25$.
- γ predator death. This has to be on a similar scale as α so suppose $\gamma = 0.2$

The predator in this model is known as a 'keystone' predator in that it controls the prey population size. If there were no predation then the prey population size would grow exponentially.

2. Propose a Poisson process that is 'analogous' to this continuous time model.

Let's propose a continuous-time process with five superimposed processes:

1. Prey birth occurs at rate $\alpha X(t)$ and increases the number of prey by 1.
2. Predation occurs at rate $\beta X(t)Y(t)$ and decreases the number of prey by 1.
3. Predator birth occurs at rate $c\beta X(t)Y(t)$ and increases the number of predators by 1.
4. Predator death occurs at rate $\gamma Y(t)$ and decreases the number of predators by 1.

The state of the system is given by the vector $[X, Y]$

Step 1: List the set of events that can occur, the rates at which they occur and the "Change vector" describing their effect on the state.

Event	Rates	Δ state
Prey Birth	$\alpha X(t)$	$[1, 0]$
Predation	$\beta X(t)Y(t)$	$[-1, 0]$
Pred. Birth	$\beta c X(t)Y(t)$	$[0, 1]$
Pred. Death	$\gamma Y(t)$	$[0, -1]$

Step 2: Set initial conditions

Step 3: given rates $\vec{\lambda}([X, Y])$, draw the time until the next event from the exponential distribution with rate $\sum_i \lambda_i([X, Y]) = \lambda_{tot}([X, Y])$.

Step 4: Choose the next event. Event e occurs with probability

$$\frac{\lambda_e([X, Y])}{\lambda_{tot}([X, Y])}$$

Step 5: Update state vector and rates.

Repeat: Steps 3-5 until time t_{Max} or some other stopping condition is reached.

Example Continued: Lotka-Volterra Predator-Prey dynamics

3. Use a Gillespie Simulation to simulate 50 Trajectories of the Lotka-Volterra Model if there are initially $X(0) = 100$ and $Y(0) = 20$.

4. Compare the stochastic dynamics to the deterministic dynamics for this model.

[Python: Lecture3_2.ipynb](#)
