Chapter 0

Markov Processes and Simple Models of Complex Systems

Goal of the tutorial is for students to get familiar with Poisson processes, and more generally, with Markovian processes. After this tutorial, students should:

- know what a homogeneous Poisson process is, know of examples and where they can be used.
- know that the distribution of number events of a Poisson process in a given time interval is the Poisson distribution.
- know that the distribution of waiting time in a Poisson process is the exponential distribution, and how to recover such distribution.
- know how to generate events following a Poisson process with rate λ.
- know how to check if a distribution is normalized.
- know how to compute averages for discrete probability distributions and continuous probability distributions.
- write down the equation of evolution for quantities following a continuous time Markov process.
- understand what *stationary* means for a stochastic process.

0.1 Homogeneous Poisson process

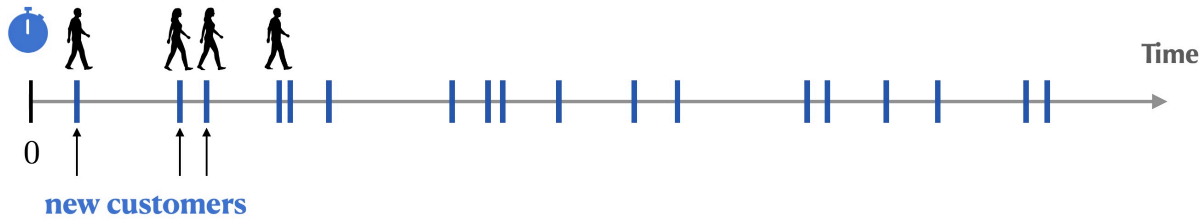
The **Poisson process** is one of the most widely-used counting processes. It is a model for a series of discrete events where the average time between events is known, but the exact time at which events occur is random. In particular, the occurrence of a new event is independent of the previous events: we say that the process is memoryless, or *Markovian*.

(Homogeneous) Poisson process: Collection of **random events** that:

- are **independent from each other**;
- occur with a **constant rate**.

For example, consider a supermarket in which a new customer arrives on average every 2 minutes. The arrival time of a new customer is random, independent of the previous customers, and will not influence the arrival time of the next customers. The Poisson process is a good model for this problem. In practice, Poisson processes have been used to model many other phenomena, such as timing of calls at a help center, of visitors to a website, of patients arriving in emergency room, of meteors hitting Earth, radioactive decay in atoms, occurrences of earthquakes, spiking time of neurons, or movements in a stock price.

> Poisson processes are generally associated with time, but they can also be associated with space. For instance, consider trees randomly placed in a forest and the problem of counting the number of trees in an acre (events per area).



0.1.1 Simple considerations

Let us take the example of modeling the times at which customers arrive at a supermarket. At time $t = 0$, we start counting the number of customers that arrive. We denote by $N(t)$ the total number of customers that have entered the shop until a time t , with $N(0) = 0$.

Q1. We know that on average there is one customer entering the shop every 2 minutes. What is the rate λ at which customers enter the shop? What is the average number of customers $\overline{N(t)}$ that one can expect to see enter the shop within $t = 10$ minutes?

0.1.2 Definitions of a Poisson process and Poisson distribution.

> **Definition 1.** The counting process $N(t)$ is a **Poisson process** with rate $\lambda > 0$, if:
>
> a. $N(0) = 0$;
>
> b. $N(t)$ has independent increments;
>
> c. the number of arrivals in any interval of length $t > 0$ follows a **Poisson distribution** with parameter $\mu = \lambda\, t$, i.e. the probability to observe k arrivals during the time t is given by:
>
> $$\mathbb{P}[N(t) = k] = \frac{\mu^k\, \exp(-\mu)}{k!}\,. \tag{1}$$

Q2. Can you check that the probability distribution defined in Eq. (1) is well normalized, and show that the mean of the distribution is μ? Compare with the expression for the average number of customers $\overline{N(t)}$ that you used in question Q1.

Q3. For the supermarket problem described above, assuming that the number of customers arriving at the shop follows a Poisson process, can you compute the probability that only 2 customers have arrived within a 10 minutes time window? What about 5 customers? 15 customers? Can you plot the distribution of $N(t)$ for $t = 10$ minutes, for $N(t) = 0$ to $N(t) = 20$ (included)?

Q4. Consider a very short time interval $dt \ll 1$. Using a Taylor expansion for small dt, can you show that the probability that no customer arrives during dt is close to 1 and takes the form:

$$\mathbb{P}[N(dt) = 0] = 1 - \lambda\, dt + o(dt), \tag{2}$$

where $o(dt)$ indicate terms that are negligable (very small) compared to dt. Similarly, can you show that the probability that the probability that 1 customer (respectively or 2 or more customers) arrives during dt takes the form:

$$\mathbb{P}[N(dt) = 1] = \lambda\, dt + o(dt) \tag{3}$$

$$\mathbb{P}[N(dt) \geq 2] = o(dt) \tag{4}$$

Definition 2. The counting process $N(t)$ is a **Poisson process** with rate $\lambda > 0$, if:

a. $N(0) = 0$;

b. $N(t)$ has independent and stationary increments;

c. during a very short time interval dt:
   - the probability that 1 customer enter the shop is: $\lambda\, dt + o(dt)$
   - the probability that 2 or more customers enter the shop is: $o(dt)$
   - and, therefore, the probability that no customer enter the shop is: $1 - \lambda dt + o(dt)$

   which can be written formally as:

$$\mathbb{P}[N(dt) = 0] = 1 - \lambda\, dt + o(dt) \tag{5}$$

$$\mathbb{P}[N(dt) = 1] = \lambda\, dt + o(dt) \tag{6}$$

$$\mathbb{P}[N(dt) \geq 2] = o(dt) \tag{7}$$

Q5. How can you simulate events that follow such rules? Write a small program that simulate the supermarket problem described above, by generating random times (in minutes) at which customers arrive at the supermarket. Simulate data for three days at the supermarket: a day starts at 8 am and finishes at 10 pm. What is the order of magnitude of the number of time steps that the simulation will do if you take $dt = 0.1$ for instance? Discuss the choice of dt.

0.1.3 Waiting time (or interarrival times)

The general question of this section is: How long do we have to wait until the next event occurs? (i.e. until the next customer arrives)

Let's assume that $N(t)$ follows a Poisson process. We would like to study the distribution of waiting times τ between two customers. Let us denote $P(\tau)$ this distribution.

We recall that by definition of a probability distribution $P(\tau)$, one has the following:
Assume that a customer just entered, the probability that the next customer enters within the time interval $[\tau;\ \tau + d\tau)$ later is given by $P(\tau)\, d\tau$, where $d\tau$ is a very short time interval ("infinitesimal time interval") chosen such that at maximum only 1 customer arrives during $d\tau$.

Q6. Let's introduce the **cumulative function** $U(T)$ as the probability that the next customer arrives at any time τ between 0 and T . Can you express $U(T)$ as a function of $P(\tau)$? What is the value of $U(T)$ for $T = 0$? What is the limit value of $U(T)$ for infinitely large time T?

Q7. Consider an infinitesimal time interval dT. The quantity $U(T+dT)$ is then the probability that the next customer arrives before the time $T + dT$. Assuming that $N(t)$ is a Poisson process, can you express $U(T + dT)$ as a function of $U(T)$, dT and λ? Deduce that $U(T)$ verifies the following equation of evolution:

$$\frac{d\,U(T)}{dT} = \lambda\,(1 - U(T))\,. \tag{8}$$

Q8. Solve this equation to find $U(T)$. Which probability distribution do you obtain for $P(\tau)$? Check that you recover an average waiting time of $1/\lambda$, as original defined in the exercise.

Poisson process: Distribution of waiting times.

In this question, you have recovered that the **distribution of the waiting times** τ between two consecutive events of a Poisson process with rate λ is the **exponential distribution**:

$$P(\tau) = \lambda\, \exp(-\lambda\, \tau)\,. \tag{9}$$

Q9. In the data generated in question Q2, can you check that the distribution of time intervals between two successive customers corresponds to the distribution $P(\tau)$ that you just found? You can also test this on the data provided with this tutorial. What is the average waiting time in the data? Compare with $1/\lambda$.

Q10. If we arrive at a random time in the shop, how long can we expect to wait to see the next customer arriving? What is the probability that the next customers arrives in less than 1 minute? What is the probability that the next customer arrives in more than 5 minutes?

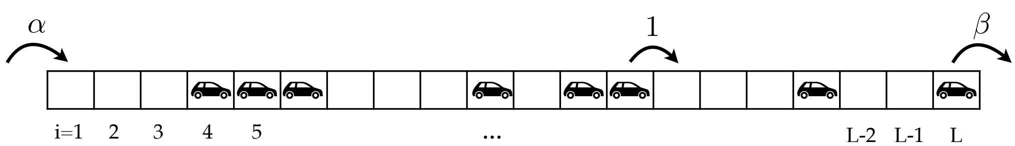
Q11. Sampling from an exponential distribution. Using the expression of the cumulative distribution $U(T)$ found in question Q5, how can you directly sample the next time at which a customer will arrive? Can you use this result to generate random data for the supermarket problem more efficiently?

0.1.4 Summary

Can you summarize what you have seen in this section?

0.2 TASEP: a simple Traffic Model

> We consider a one-dimensional lattice with L sites opened on both sides (which represents a portion of road). Cars randomly hop into the lattice from the left with rate α, jump along the lattice to the right with rate 1, and exit the lattice from the right with rate β (see figure). Each site can only be occupied by one car at a time, and a car can hop to the next site only if it is empty (exclusion interaction). This model for instance the fact that, on a one-lane road, a car moves forward provided that there is no vehicle in front of it.
>
> This simple traffic model is called Totally Asymmetric Simple Exclusion Process (TASEP): "exclusion" because there can only be one car per site, and "Totally Asymmetric" because cars can only move towards the right. It was originally introduced to model the translation of mRNA in protein synthesis, and is useful in modeling a wide range of transport phenomena, including road traffic. The model is known for exhibiting phase transitions even in one dimension, between low-density, high-density and maximum-current phases.



Q1. Consider a very short time interval dt, what are the possible events that can happen during dt and what are the probability that each of these events happen?

Q2. How can you simulate the dynamics of this model? Give the main lines of your algorithm.

Q3. We denote by $n_i(t)$ the number of cars at the site i at time t, which can only take two values: 0 if the site i is empty, and 1 if the site i is occupied. We denote by $\rho_i(t)$ the probability that there is a car at the site i at time t: $\rho_i(t) = P[n_i(t) = 1]$. The probability that there is no car at site i is then: $1 - \rho_i$ (as a site can either be occupied or not).
Can you show that the average number of cars in i at time t is $\overline{n_i(t)} = \rho_i(t)$?

Q4. We define the current $J_i(t)$ of cars that exit site i to the right at time t by:

$$J_i(t) = P[n_i(t) = 1; n_{i+1}(t) = 0]\,, \tag{10}$$

which corresponds to the joined probability that site i is occupied and that site $i+1$ is empty at time t. What is the probability that a car moves from site i to site $i+1$ during the very short time interval $[t, t+dt)$?

Q5. Show that the local density of cars follows the equation of evolution:

$$\frac{d\rho_i}{dt}(t) = J_{i-1}(t) - J_i(t)\,, \tag{11}$$

for all the sites i between $i = 2$ and $i = L - 1$.

Q6. What are the equations of evolution for ρ_1 and ρ_L? Introduce a current $J_0(t)$ that enters the site 1 from the left, and a current $J_L(t)$ that leaves site L to the right.

Q7. In the stationary state, the local densities and currents become time-independent. Deduce from the previous equations that the current is uniform in the stationary state, i.e. for all i, $J_i = J$ is a constant. Show that we have the boundary equations:

$$\rho_1 = 1 - \frac{J}{\alpha} \qquad \text{and} \qquad \rho_L = \frac{J}{\beta} \,. \tag{12}$$

Q8. Play with the simulation provided in the mathematica notebook. The graph in the bottom left display the evolution of the current that goes through the last site $i = L$ as a function of time. Observe that the average current initially increases to then reach a stationary value (the value of J fluctuates around a fixed value). Play with the parameters α and β and observe the different stationary states of the systems. Can you comment on the behavior of the system for $(\alpha,\ \beta) = (1.0, 0.1)$, $(0.1, 1.0)$, $(0.2, 0.2)$ and $(0.6, 0.6)$?

For Thursday: Using the simulation provided in the mathematica notebook, fill in 4 or 5 values of J in the table here. You can accelerate the simulation by selecting '20' for the simulation speed (that will change the refreshing rate of the graphs). Fixing the values of the parameters, the system will slowly evolve towards stationarity: the value of J will fluctuate around a fixed value. Report an estimate of this value in the table of the google doc.

Bonus question: If you have some spare time, feel free to try to implement yourself a numerical simulation of the TASEP model!

0.1 Solutions

0.1.1 Homogeneous Poisson process

Useful definitions

- **Discrete probability distribution:** The Poisson distribution in Eq. (1) is a discrete probability distribution, for which $P(k)$ is the probability that k event happens.
- **Continuous probability distribution**, also called **probability density function (PDF):** The Exponential distribution $P(\tau)$ in Eq. (9) is a probability density function (i.e. $P(\tau)$ is not a probability per se, but a probability density): $P(\tau)\,d\tau$ is the probability that the next event happens within the small time interval $[\tau, \tau + d\tau]$.
- **Normalization and Ensemble Average:**

	Discrete distribution	Continuous distribution
Normalization	$\sum_k P(k) = 1$	$\int \mathrm{d}x\, P(x) = 1$
Average	$\langle A \rangle = \sum_k A(k)\, P(k)$	$\langle A \rangle = \int A(x)\, P(x)\, dx$

  where the summation (resp. the integration) is over the entire domain where the discrete distribution $P(k)$ (resp. the continuous distribution $P(x)$) is defined.

- **Cumulative Distribution Function (CDF):**

$$U(T) = \int_0^T P(\tau) \mathrm{d}\,\tau .$$

- Series expansion of the exponential function:

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \dots$$

  Note: you don't need to remember this expansion for the exam. If it is needed for the exam, the expansion will be written in the first page of the exam sheet.

Notations. In the literature (and in this course), you can find the (ensemble) average denoted with different notations, the most common being:

$$\langle A \rangle \qquad \text{or} \qquad \overline{A} \qquad \text{or} \qquad E[A] \tag{13}$$

A1.

$$\lambda = \frac{1}{2\ [\text{min}]} = \frac{1}{2 \times 60\ [\text{s}]} = 8.3 \cdot 10^{-3}\ [\text{s}^{-1}] = 0.5\ [\text{min}^{-1}] \tag{14}$$

$$\overline{N(t = 10\,\text{min})} = \lambda \times t = 5 \text{ customers} \tag{15}$$

A2. To check that the probability distribution is normalized, we must check that $\sum_{k\geq 0} \mathbb{P}[k] = 1$.

$$\sum_{k\geq 0} \mathbb{P}[k] = \sum_{k=0}^{\infty} \frac{\mu^k \exp(-\mu)}{k!} = \exp(-\mu) \sum_{k=0}^{\infty} \frac{\mu^k}{k!} = \exp(-\mu)\ \exp(\mu) = 1\,, \tag{16}$$

where we recognized the power series of the exponential function: $\exp(\mu) = \sum_{k=0}^{\infty} \frac{\mu^k}{k!}$.

The mean $\overline{N(t)}$ of the distribution at a given time t is defined as $\overline{N(t)} = \sum_{k\geq 0} k\,\mathbb{P}[k]$, in which $k = N(t)$.

$$\overline{N(t)} = \sum_{k\geq 0} k\,\mathbb{P}[k], \qquad \text{in this sum, the term for k=0 is null} \tag{17}$$

$$= \exp(-\mu) \sum_{k=1}^{\infty} k\,\frac{\mu^k}{k!} = \exp(-\mu) \sum_{k=1}^{\infty} \frac{\mu^{k-1}\,\mu}{(k-1)!}, \tag{18}$$

$$= \mu\,\exp(-\mu) \sum_{k'=0}^{\infty} \frac{\mu^{k'}}{k'!}, \qquad \text{where we took } k' = k-1 \tag{19}$$

$$= \mu\,\exp(-\mu)\,\exp(\mu) = \mu. \tag{20}$$

A3. We want to compute the probability that $N(t = 10\,\text{min}) = 2$. Assuming that $N(t)$ follows a Poisson process, we have that the probability that $N(t) = k$ is:

$$\mathbb{P}[N(t) = k] = \frac{\mu^k\,\exp(-\mu)}{k!}\,, \tag{21}$$

with $\mu = \lambda\,t$, where $\lambda = 1/2\,[\text{min}^{-1}]$ and $t = 10\,\text{min}$, i.e. $\mu = 10/2 = 5$. Therefore:

$$\mathbb{P}[N(t) = 2] = \frac{5^2\,\exp(-5)}{2!} \simeq 0.08$$
$$\mathbb{P}[N(t) = 5] = \frac{5^5\,\exp(-5)}{5!} \simeq 0.18$$
$$\mathbb{P}[N(t) = 15] = \frac{5^{15}\,\exp(-5)}{15!} \simeq 1.6 \cdot 10^{-4}$$

For other numerical values, see Fig. 1 Left a few pages below.

A4. The probability that no customer arrives during dt is given by the Poisson distribution at $k = 0$ with $\mu = \lambda\,dt$:

$$\mathbb{P}[N(dt) = 0] = \exp(-\lambda\,dt) \tag{22}$$

$$= 1 - \lambda\,dt + \frac{(\lambda\,dt)^2}{2} - \cdots \qquad \text{Taylor series} \tag{23}$$

If dt is small, the terms that include second or higher powers of dt are negligible compared to dt (i.e. when dt goes to 0, these terms are going much faster to 0 than dt itself). We write this as

$$\mathbb{P}[N(dt) = 0] = 1 - \lambda\,dt + o(dt)\,. \tag{24}$$

Little-"o" notation:

If dt is small, the terms that include second or higher powers of dt are negligible compared to dt (i.e. when dt goes to 0, these terms are going much faster to 0 than dt itself). In the equations, we denote all terms that are negligible compared to dt (i.e. the terms that are of order dt^2, dt^3, or higher) by $o(dt)$. For instance, Eq. (23) becomes:

$$\mathbb{P}[N(dt) = 0] = 1 - \lambda\,dt + o(dt)\,. \tag{25}$$

The probability that one customer arrives during dt is given by:

$$\mathbb{P}[N(dt) = 1] = \lambda\,dt\,\exp(-\lambda\,dt) \tag{26}$$

$$= \lambda\,dt - (\lambda\,dt)^2 + \frac{(\lambda\,dt)^3}{2} - \cdots \qquad \text{Taylor series} \tag{27}$$

$$= \lambda\,dt + o(dt)\,. \tag{28}$$

The probability that two or more customers arrives during dt can be obtained using the normalisation:

$$\mathbb{P}[N(dt) = 0] + \mathbb{P}[N(dt) = 1] + \mathbb{P}[N(dt) \geq 2] = 1\,, \tag{29}$$

which leads to:

$$\begin{aligned} \mathbb{P}[N(dt) \geq 2] &= 1 - \mathbb{P}[N(dt) = 0] + \mathbb{P}[N(dt) = 1]\,, && (30)\\ &= o(dt)\,. && (31) \end{aligned}$$

A5. Take a small time interval dt such that the probability that 1 event occurs during dt is reasonable, but the probability that 2 events occur during dt is very small. For instance, if we consider a minute as a unit time, choosing $dt = 0.1$ minute gives that the probability that 1 event occurs during dt is $\lambda\, dt = 0.05$, the probability that two events occur during dt is $(\lambda\, dt)^2 = 0.0025$ (note that the probability that two events or more occur during dt is $(\lambda\, dt)^2 + o(dt^2)$, which is of the order of $(\lambda\, dt)^2 = 0.0025$). Note that we don't want to take dt too small, as otherwise we would need to take too many time steps. With $dt = 0.1$ minute, one needs 10 steps to simulate a minute, 600 steps for an hour, $14 * 600 = 8400$ steps for a whole day, and finally 25200 steps for three days.

A6. By definition:

$$U(T) = \int_0^T P(\tau) d\tau \qquad \text{and} \qquad \lim_{T\to\infty} U(T) = 1\,. \tag{32}$$

The first equation comes from the definition of $U(T)$, the second from the normalization of $P(\tau)$ (as $P(\tau)$ is defined for $\tau \in [0; +\infty)$).

A7.

$$\underbrace{U(T + dT)}_{(1)} = \underbrace{U(T)}_{(2)} + \underbrace{(1 - U(T))}_{(3)} \times \underbrace{\lambda\, dT}_{(4)}\ . \tag{33}$$

where:

(1) is the probability that the next customer enters between time 0 and time $T + dT$;

(2) is the probability that the next customer enters between time 0 and time T;

(3) is the probability that the next customer has not yet entered between time 0 and time T (which is also equal to the probability that the next customer arrives at a time later than T);

(4) is the probability that a customer enters during the short time dT. Note that, as the process is memoryless process, this probability is completely independent of what could have happened before.

This equation can be re-written as:

$$\frac{U(T + dT) - U(T)}{dT} = \lambda\,(1 - U(T))\,. \tag{34}$$

which finally leads to the equation of evolution in the text, in which we use the first order approximation of the derivative:

$$U'(T) = \frac{dU(T)}{dT} = \frac{U(T + dT) - U(T)}{dT}\,. \tag{35}$$

Approximation for the derivative of a function:

We will often use the following approximation for the derivative of a function $f(x)$:

$$f'(x) = \frac{f(x + dx) - f(x)}{dx}\,, \qquad \text{for very small } dx \tag{36}$$

> Note: this is equivalent to cutting the Taylor expansion of $f(x+dx)$ to first order:
>
> $$f(x+dx) = f(x) + dx\, f'(x) + o(dx^2)\,. \tag{37}$$
>
> Note for computer scientists: this is similar to using the finite difference as an approximation of the derivative:
>
> $$f'(T) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \tag{38}$$

A8. We can re-write equation (8) as

$$\frac{d\,V(T)}{dT} = -\lambda\, V(T) \qquad \text{where} \;\; V(T) = 1 - U(T) = \int_T^{+\infty} P(\tau)\, d\tau\,. \tag{39}$$

The solution of this equation is $V(T) = A\;\exp(-\lambda\, T)$, where A is a constant. Using that $V(0) = 1$ (because $P(\tau)$ is normalized), we obtained that $A = 1$. Hence $U(T) = 1 -\; \exp(-\lambda\, T)$. Finally, $P(\tau)$ can be obtained by deriving $U(T)$:

$$P(T) = U'(T) = \lambda \exp(-\lambda T)\,. \tag{40}$$

We just proved that the time intervals between successive events of a Poisson process follow an Exponential distribution (see Fig. 1 Right in the next page).

The average waiting time is calculated using:

$$\begin{aligned}
\langle\tau\rangle &= \int_0^{+\infty} \tau\, P(\tau)\, d\tau = \int_0^{+\infty} \tau\, \lambda\; \exp(-\lambda\tau)\, d\tau \\
&= [-\tau\; \exp(-\lambda\tau)]_0^{+\infty} + \int_0^{+\infty} \exp(-\lambda\tau)\, d\tau \qquad \text{(integration by parts)} \\
&= 0 + [-\frac{1}{\lambda}\; \exp(-\lambda\tau)]_0^{+\infty} \\
\langle\tau\rangle &= \frac{1}{\lambda}
\end{aligned}$$

A9. Note that, if we were to sample many dataset, for each of them, we will find a different value of the average waiting time. However, if one were to run a large number of experiments, one would find that the average waiting time (computed for each experiment) follows a Gaussian distribution (normal distribution) centered around the average value $1/\lambda$, and with a standard deviation that is proportional to $1/\sqrt{N}$, where N is the number of experiments performed.

A10. We can expect to wait an average time of $1/\lambda = 2$ minutes. The probability that the next customer arrives in less than 1 minute is given by $U(T = 1\,\text{min}) = 1 - \exp(-\lambda T) = 1 - \exp(-1[\text{min}]/2[\text{min}]) = 0.39$. The probability that the next customer arrives in more than 5 minutes is $1 - U(T = 5\,\text{min}) = \exp(-\lambda T) = \exp(-5[\text{min}]/2[\text{min}]) = 0.08$.

A11. To sample a random time T from an exponential distribution, one can sample uniformly a variable $\epsilon \in [0, 1]$, then find the value of T such that:

$$\int_0^T P(\tau)\, d\tau = \epsilon\,. \tag{41}$$

i.e. such that $U(T) = \epsilon$. As $U(T) = 1 - \exp(-\lambda\, T)$ is a strictly growing function of T, this equation has only a unique solution, which can be easily computed:

$$1 - \exp(-\lambda\, T) = \epsilon \qquad \Longleftrightarrow \qquad T = -\frac{1}{\lambda} \log(1-\epsilon)\,. \tag{42}$$

Note that if ϵ is uniformly distributed over $[0, 1]$, then $\eta = 1 - \epsilon$ is also uniformly distributed over $[0, 1]$. Therefore, to sample a random time T from an exponential distribution with rate λ, one can sample uniformly a variable $\eta \in [0, 1]$ and return $T = -\frac{1}{\lambda} \log(\eta)$.

To generate data for the supermarket problem more efficiently, one can directly sample from the exponential distribution $P(\tau)$ the waiting time until the next customer arrives using the method just described.

Summary 1. Poisson process.

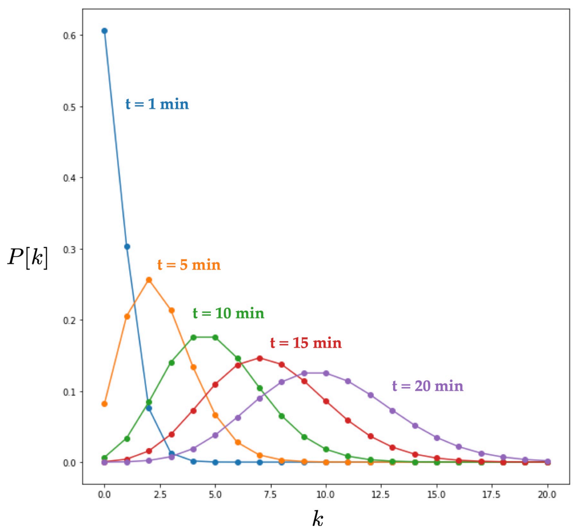
If $N(t)$ follows a Poisson process with parameter $\lambda > 0$, then:

- the **probability to observe** k **events** in a given time interval t is given by the **Poisson distribution**:

$$\mathbb{P}[N(t) = k] = \frac{\mu^k \exp(-\mu)}{k!}, \qquad \text{where} \quad \mu = \lambda\, t\,. \tag{43}$$

- the **distribution of the waiting time** τ between two events is the **exponential distribution**:

$$P(\tau) = \lambda \exp(-\lambda\, \tau)\,. \tag{44}$$



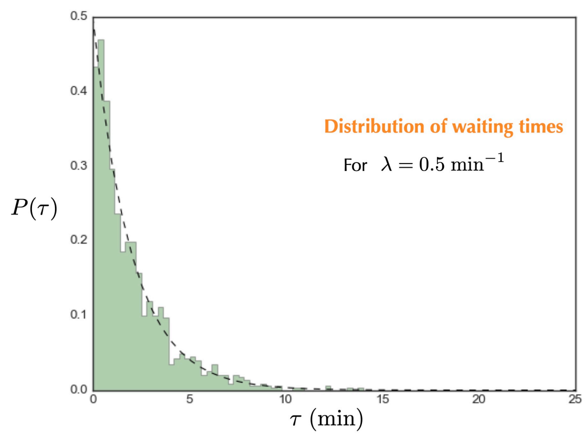


Figure 1: **Left: Poisson distribution.** The probability to observe k events in a given time interval t is given by the Poisson distribution (which is a discrete probability distribution). The figure shows the probability that k customers arrives in a given time interval t for the process in the exercise above with rate $\lambda = 0.5$ customers per minute. **Right: Distribution of waiting times.** The duration of the time-intervals between two consecutive events (also called *waiting times*) in a Poisson process follows an exponential distribution. This is a continuous probability distribution. The figure shows the distribution of waiting times τ for the process in the exercise above with rate $\lambda = 0.5$ customers per minute.

Summary 2. Sampling from an Exponential Distribution.

If η is uniformly distributed over $[0, 1]$, then $T = -\frac{1}{\lambda}\log(\eta)$ follows an exponential distribution with rate λ.

0.1.2 TASEP: a simple Traffic Model

A1. During dt, the following events can happen:

- if the site $i = 1$ is empty, then a car can enter from the left with probability $\alpha\, dt$;
- if the site $i = L$ is occupied, then the car in $i = L$ can exit the lattice from the right with probability $\beta\, dt$;
- each car inside the lattice can hop to the right with probability dt if the site to their right is empty.

A2. A first possible algorithm consists in simulating the evolution of the system by small time steps of length dt. During each time step, one must check if any of the events listed in the answer of question Q1 happens using their respective probability to occur. One must take dt sufficiently small so that the probability that the same car moves twice (or more) during

dt is negligeable, but large enough for the simulation to be executed in a reasonable time.

A second version of the algorithm consists in directly sampling the next time at which one event will happen, and thus to jump forward in time directly to that time. This approach is based on the fact that, for each of the possible action (car enters from the left, car moves forward, car exits from the right), the time interval between two actions are exponentially distributed (with respective rates α, 1, β). This version of the algorithm will be much faster than the first version. In this version, we iterate in the following way: (1) at the current time t, we sample:

- if the first site $i = 1$ is empty, we sample the time T_{in} at which the next car with enter the lattice from the left; this time is sampled from an exponential distribution with rate α;
- if the last site $i = L$ is occupied, we sample the time T_{out} at which the car will leave the lattice from the right; this time is sampled from an exponential distribution with rate β;
- the time T_{hop} at which the next car on the lattice will move to the right; this time is sampled from an exponential distribution with rate $N \times 1$, where N is the number of cars on the lattice that have an empty spot on their right (i.e. that can move to the right).

We then take the shorter of the sampled times ($T = \min(T_{in}, T_{out}, T_{hop})$), move forward in time to that time ($t = t + T$), and perform the corresponding action. Note that if T_{hop} is the shortest time, we then uniformly sample one of the N cars on the lattice to hop to the right.

A3. By definition, the average number of cars in i at time t is given by: $\overline{n_i(t)} = \sum_{n_i} n_i P[n_i(t)]$, where the sum is over the values that can be taken by n_i. As n_i can only take the two values 0 and 1, we get:

$$\overline{n_i(t)} = 1 \times P[n_i(t) = 1] + 0 \times P[n_i(t) = 0] \tag{45}$$

$$= \rho_i(t) \tag{46}$$

A4. The probability that a car hops from node i to node $i + 1$ during the time interval $[t, t + dt)$ is:

$$J_i(t)\, dt = \underbrace{P[n_i(t) = 1; n_{i+1}(t) = 0]}_{(1)} \underbrace{dt}_{(2)} \tag{47}$$

where:

(1) is the probability that there is a car in i and no car in $i + 1$ at time t;

(2) is the probability that the car moves to the right.

Comment: there is a common mistake done by students here, which is to write the probability of $n_i(t) = 1$ and $n_{i+1} = 0$ as a product of the two probabilities, $P[n_i(t) = 1] \times P[n_{i+1}(t) = 0]$, instead of as the joint probability $P[n_i(t) = 1; n_{i+1}(t) = 0]$. This is not correct. $P[n_i(t); n_{i+1}(t)]$ is a joint probability distribution over 4 possibles states ($(n_i(t), n_{i+1}(t)) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$). Writing down that:

$$P[n_i(t); n_{i+1}(t)] = P[n_i(t)] \times P[n_{i+1}(t)] \qquad \text{(not true in general)} \tag{48}$$

assumes that the states taken by the variables n_i are independent of the states taken by n_{i+1} (and reciprocally that the states of n_{i+1} doesn't depends on the state of n_i). For instance, Eq. (48) assumes that the probability that $n_{i+1} = 1$ doesn't depends on the state of n_i. A priori, there is no reason to assume independence between n_i and n_{i+1} without any specific knowledge about the system, so in general Eq. (48) is not valid. Note that in the tutorial on mean-field theory, you will see that this independence is assumed in the mean-field approximation.

A5. To compute the equation of evolution of $\rho_i(t)$, one must consider the evolution of the system during a very short time interval $[t, t + dt)$. For that, the idea is to compute the probability $\rho_i(t + dt)$ that the site i is occupied by a car at time $t + dt$ by relating it to the states of the system at time t. There are two possibilities for the site i to be occupied at time $t + dt$:

1. at time t, the site i was empty and the site $(i-1)$ was occupied, AND the car moved from $(i-1)$ to i during dt; The probability of this event corresponds to item (1) in the equation below.
2. at time t, the site i was occupied by a car that didn't move out to the right during dt. This corresponds to item (2) in the equation below and can be decomposed into two cases:
   - 2a. at time t, the site i and $i+1$ were both occupied (and therefore the car in i can't move to the right during dt). The probability of this event corresponds to item (2a) in the equation below.
   - 2b. at time t, the site i was occupied and the site $i+1$ was empty, but the car didn't move to the right during dt. The probability of this event corresponds to item (2b) in the equation below.

These considerations can be translated into the following equation:

$$P[n_i(t+dt)=1] = \underbrace{P[n_{i-1}(t)=1; n_i(t)=0] \times \underbrace{dt}_{(\star)}}_{(1)} + \underbrace{\underbrace{P[n_i(t)=1; n_{i+1}(t)=1]}_{(2a)} + \underbrace{P[n_i(t)=1; n_{i+1}(t)=0]\,(1-dt)}_{(2b)}}_{(2)}, \tag{49}$$

where

(1) is the probability that a car moves from $(i-1)$ to i between the times t and $t+dt$;

($\star$) is the probability that a chosen car moves to the right during dt;

(2) is the probability that the site i was already occupied at time t and stays occupied during dt;

Using that $P[n_i(t)=1; n_{i+1}(t)=1] + P[n_i(t)=1; n_{i+1}(t)=0] = P[n_i(t)=1]$, we can re-write the term (2) as:

$$(2) = P[n_i(t)=1] - P[n_i(t)=1; n_{i+1}(t)=0]\,dt\,, \tag{50}$$

which could have also been directly obtained as the probability that i was occupied at time t minus the probability that i becomes empty during dt. Finally, using the definition of $\rho_i(t)$ and $J_i(t)$, we get:

$$\frac{\rho_i(t+dt)-\rho_i(t)}{dt} = J_{i-1}(t) - J_i(t)\,, \qquad \text{for all } i \in \{2, \cdots, (L-1)\}\,, \tag{51}$$

which leads to the equation of evolution for $\rho_i(t)$ given in Eq. (3.9).

A6. Similarly to question Q5, we can write the probability that the site $i=1$ is occupied at time $t+dt$:

$$P[n_1(t+dt)=1] = P[n_1(t)=0]\alpha\,dt + P[n_1(t)=1] - P[n_1(t)=1; n_2(t)=0]\,dt\,, \tag{52}$$

which gives:

$$\frac{d\rho_1}{dt}(t) = J_0(t) - J_1(t)\,, \qquad \text{where} \qquad J_0(t) = \alpha\,(1-\rho_1(t))\,. \tag{53}$$

Similarly, one gets for the current on the right end of the lattice:

$$\frac{d\rho_L}{dt}(t) = J_{L-1}(t) - J_L(t)\,, \qquad \text{where} \qquad J_L(t) = \beta\,\rho_L(t)\,. \tag{54}$$

A7. At stationarity (i.e. in the stationary state), the local densities are independent from time, which means that $\frac{d\rho_i}{dt}(t) = 0$. Replacing this value in Eq. (3.9), we obtain that $J_{i-1} = J_i$ for all i from 2 to $(L-1)$, i.e. the current is constant:

$$J_i = J = \text{constant}\,, \qquad \text{for all } i \in \{1, \cdots, (L-1)\}\,. \tag{55}$$

Similarly, at stationarity, $\frac{d\rho_1}{dt}(t) = 0$ and $\frac{d\rho_L}{dt}(t) = 0$. Replacing this respectively in Eq. (53) and in Eq. (54), we get that:

$$J_0 = J = J_L\,, \tag{56}$$

where $J_0 = \alpha\,(1-\rho_1)$ and $J_L = \beta\,\rho_L$.

A8. The answers to this question are discussed at the beginning of the lecture L1 (and see Fig. 2 below).

1,0		0,09	0,16	0,21	0,24	0,25	0,26	0,27	0,27	0,27	0,26
0,9		0,09	0,16	0,21	0,24	0,25	0,26	0,26	0,27	0,27	0,27
0,8		0,09	0,16	0,21	0,24	0,24	0	0	0,27	0	0,27
0,7		0,09	0,16	0,21	0,24	0	0	0,26	0	0,25	0,27
0,6		0,09	0,17	0,21	0,24	0	0,26	0	0,22	0	0,26
0,5		0,09	0,15	0,21	0,24	0,25	0,25	0,25	0	0,26	0,25
0,4		0,09	0,16	0,2	0,23	0,24	0	0	0	0,23	0,24
0,3		0,09	0,16	0,2	0,2	0,21	0,21	0,21	0,21	0,21	0,21
0,2		0,09	0,15	0,15	0,15	0,16	0,16	0,16	0,16	0,15	0,16
0,1		0,08	0,09	0,09	0,09	0,09	0,09	0,09	0,08	0,09	0,09
beta											
	alpha	**0,1**	**0,2**	**0,3**	**0,4**	**0,5**	**0,6**	**0,7**	**0,8**	**0,9**	**1,0**

Figure 2: **TASEP phase diagram:** Numerical values of the stationary current J for different values of α and β, filled in by the students of year 2023, based on the mathematica program available in canvas.