

0.1 Poisson Processes

We are going to derive the Poisson and Gamma distributions for waiting time processes. You may have done this already in the past, but it is a good exercise to understand the mechanism behind waiting time/queuing problems. The Poisson distribution is a solution to a system of first order linear ordinary differential equations, which you will derive below and thereafter solve.

Suppose you throw an epic party where, of course, you get visitors. Suppose that visitors come in randomly. Let at any moment in time p be the probability per minute that someone arrives. And for the moment we suppose that arrivals come in independent of the time of day, how many people are already there, or any other factor, so p is a constant.

Now if Δt is a small time interval (say 1 second) then $p\Delta t$ is (approximately) the probability that someone will arrive within this time interval¹.

- a. If $p\Delta t$ is the probability that someone arrives within the small time interval Δt , then what is the probability that no one arrives within this small time interval?

Suppose you start the party at time t_0 . Denote by $P(n, t, t_0)$ the probability that by time t there have been n visitors coming to your party.

If by time $t + \Delta t$ you have n visitors at your party, then this could have come from two possibilities a moment earlier at time t . It could have been that you had $n - 1$ visitors at time t and one extra came in in the small time interval Δt , or it could have been that you had already n visitors at time t and none came in in the small time interval.

- b. Express the probability $P(n, t + \Delta t, t_0)$ in terms of probabilities $P(n - 1, t, t_0)$ and $P(n, t, t_0)$.
- c. Derive from this expression the differential equation

$$\frac{dP(n; t, t_0)}{dt} = -pP(n; t, t_0) + pP(n - 1; t, t_0) \quad (1)$$

- d. Now also express the probability $P(0, t + \Delta t, t_0)$ of having no visitors in terms of probabilities at time t , and derive from this the differential equation

$$\frac{dP(0; t, t_0)}{dt} = -pP(0; t, t_0) \quad (2)$$

The above form a system of coupled differential equations, one for each n . They are coupled because the probability of each number n depends on the probability of $n - 1$. These can be solved exactly.

Since at time t_0 you start the party, you are certain that you have no visitors, so the initial conditions to the differential equations are $P(0, t_0, t_0) = 1$ and $P(n, t_0, t_0) = 0$ for $n > 0$.

- e. Use the ansatz $P(n, t, t_0) = A_n(t)e^{-p(t-t_0)}$ to derive that

$$P(n; t, t_0) = \frac{(p(t - t_0))^n e^{-p(t-t_0)}}{n!} \quad (3)$$

¹This is an approximation which is exact when taken in the limit $\Delta t \rightarrow 0$.

are the solutions for the differential equations 1 and 2 that agree with the initial conditions.

Now with equation 3 you have for each n a probability that this many visitors will come to your party between time t_0 and t , which is a probability distribution of n . This probability distribution is a Poisson distribution.

- f. Derive from equation 3 the average/expected number of visitors coming to your party between time t_0 and t .
- g. Derive from equation 3 the most likely number of visitors coming to your party, i.e. the number n where $P(n; t, t_0) > P(n-1; t, t_0)$ and $P(n; t, t_0) > P(n+1; t, t_0)$.

You are going to write a simulation program where $p = 0.5$ per minute.

- h. Divide a period of 60 minutes into steps of $\Delta t = 1$ minute. For each step determine randomly if someone arrives according to probability $p\Delta t$ and count the number of visitors after 60 minutes. Repeat this procedure 10000 times and plot the distribution of visitors after 60 minutes in a histogram.
- i. Now divide the period of 60 minutes into steps of $\Delta t = 1$ second and plot the distribution again. Also draw the Poisson distribution that you expect. Why are the two simulated distributions not the same? Which one is more accurate? And why?

We can also determine at what time we expect the n^{th} visitor to arrive. For the n^{th} visitor to arrive at time t we need $n-1$ visitors to have arrived already AND another person arriving at time t .

- j. Show that the probability for getting the n^{th} visitor at time t is

$$dtP(t; n, t_0) = dt \frac{p(p(t-t_0))^{n-1} e^{-p(t-t_0)}}{(n-1)!} \quad (4)$$

Here dt is the expression for Δt as it becomes infinitesimally small.

The result (excluding the infinitesimal dt) $P(t; n, t_0)$ is a probability **density** function (PDF) of the continuous time variable t (where the Poisson distribution $P(n; t, t_0)$ in equation 3 is a probability distribution of discrete variable n). This probability density function (PDF) is a Gamma distribution with shape parameter n and rate parameter p .

The probability for the first visitor to arrive at time t has shape parameter $n = 1$ and reduces to the simple exponential PDF $P(t; 1, t_0) = pe^{-p(t-t_0)}$.

- k. What is the average time it takes for the first visitor to arrive (derive this).

The exponential PDF $pe^{-p(t-t_0)}$ is the probability that it takes $t-t_0$ minutes for the first visitor to arrive. In fact the probability that it takes this amount of time for any next person to arrive is this same PDF. This you can imagine because you can privately reset your count to zero after every visitor. To illustrate that this indeed works

1. Write down the probability that starting from t_0 the first visitor arrives at time t_1 AND **from that moment** t_1 the next visitor arrives at time t_2 . Now allow the time that the first visitor arrives to be anywhere between t_0 and t_2 by integrating the probability over t_1 between t_0 and t_2 , and show that the result is the same as $P(t_2; 2, t_0)$ the probability of the second visitor arriving at time t_2 .

Thus $P(t; 2, t_0)$ is really just the probability of the second visitor arriving at time t and the first one anywhere before that. In a similar fashion you can show that multiplying n exponential PDFs $e^{-p(t_i - t_{i-1})}$ for n consecutive customers and integrating over t_1 to t_{n-1} leads to $P(t_n; n, t_0)$. Thus $P(t; n, t_0)$ is really just the probability of the n^{th} visitor arriving at time t and all the previous ones anywhere before that.

You are going to write a simulation program using the exponential PDF to decide customers coming into your store instead of taking small time steps like you previously used.

- m. Use the exponential PDF pe^{-pt} with $p = 0.5$ to generate random times until their sum is larger than 60 minutes. The number of visitors that arrived in the 60 minutes is one less than generated. Repeat this procedure 10000 times and plot the number of visitors in 60 minutes in a histogram. Compare with the previous simulations and with the expected Poisson distribution.
- n. How many random times did you generate in total? How does this compare to the total number of steps that you previously simulated for $\Delta t = 1$ second?

We obtained equation 4 the probability that the n^{th} visitor arrives at time t from the Poisson probability of having $n - 1$ visitors before that time AND one coming in at that time. You can also get the Poisson probability of n visitors back from the probability of the time that the n^{th} visitor arrives.

- o. Write down the probability that the n^{th} visitor arrives at time t_n AND no more customers come in after that between time t_n and t . Now allow the time t_n where the n^{th} visitor arrives to be anywhere between t_0 and t by integrating the probability over t_n between t_0 and t , and show that the result is the same as $P(n; t, t_0)$ the probability of having n visitors between t_0 and t .

0.2 Time dependent rate

Now suppose that the rate p is not constant but depends on time, i.e. $p = p(t)$. People are more likely to arrive at your party at certain hours.

- a. The previously derived Poisson and Gamma distributions all involved the quantity $p \times (t - t_0)$. This is the integral of constant p over time from t_0 to t . Knowing this, write down the algebraic solutions $P(n; t, t_0)$ to the differential equations when the rate is not constant.
- b. What is the probability of no one coming, ever ('ever' is $t \rightarrow \infty$)? Under what circumstances is this probability not zero? Can this happen with a constant rate?

- c. And write down $P(t; n, t_0)$ the probability of getting the n^{th} visitor at time t in this case.
- d. Starting from t_0 the probability of getting your first visitor at time t is $p(t)e^{-\int_{t_0}^t dsp(s)}$. Write down the probability that starting from t_0 the first visitor arrives at time t_1 AND from that moment t_1 the next visitor arrives at time t_2 . Integrate this probability over t_1 between t_0 and t_2 and show that the result is the same as $P(t_2; 2, t_0)$ the probability of the second visitor arriving at time t_2 .

As with a constant rate you can chain probabilities for individual visitors to arrive. You are going to write a simulation doing this with a time dependent rate. The rates as a function of the hour of the day are given in table 1.

start time	end time	rate (per minute)
18:00	10:00	0.1
19:00	11:00	0.2
20:00	12:00	0.3
21:00	13:00	0.4
22:00	14:00	0.5
23:00	15:00	0.6
00:00	16:00	0.3
01:00	17:00	0.1

Table 1: Rate depending on the hour of the day. The rate is zero at any other hour.

- e. Make a plot of the function $p(t)e^{-\int_{t_0}^t dsp(s)}$, with t_0 at 18:00, and $p(t)$ given by the hourly rates from table 1. This is the distribution of times where your first visitor arrives. You don't have to work with actual times, it is also fine to work with e.g. the number of minutes since 18:00 to avoid the horrible datetime libraries. This plot will serve as a check for your simulation to come.
- f. Make a plot of the function $p(t) \left(\int_{t_0}^t dsp(s) \right)^9 e^{-\int_{t_0}^t dsp(s)} / 9!$, with t_0 at 18:00, and $p(t)$ given by the hourly rates from table 1. This is the distribution of times where your 10th visitor arrives. This plot will also serve as a check for your simulation to come.

There is no standard probability distribution that you can use to generate times for a time dependent rate, so you will build your own random time generator that generates times randomly according to $p(t)e^{-\int_{t_0}^t dsp(s)}$. The idea is simple, you take a standard random number generator and generate a uniform random number x (between 0 and 1 with flat probability distribution). Then you find the time t for which $\int_{t_0}^t d\tau p(\tau) e^{-\int_{t_0}^{\tau} dsp(s)} = x$. This is a general way to construct random number generators with a simple uniform distribution: generate a simple uniformly distributed random number, then integrate your PDF until it hits the generated number.

- g. Build a random number generator for $p(t)e^{-\int_{t_0}^t dsp(s)}$ as indicated above, using the hourly rates from table 1. Note that t_0 the point from which on you generate your next visitor is not fixed but is the time that your last visitor arrived, so pay attention to that in your generator.
- h. Start from 18:00 and generate visitors arriving by generating random times until you pass 02:00 at night. Repeat this procedure 1000 times. Each time remember the time the first visitor arrived, the time the 10th visitor arrived and the total number of visitors. Make a histogram of the arrival time of the first visitor, and a histogram of the arrival time of the 10th visitor. Compare these to the plots you made before of the distributions for these to ensure you have built your simulation properly. Then also make a histogram of the total number of visitors. What is the average/expected number of visitors to your party from your simulation? What does the probability distribution (Poisson distribution) say the average should be?

0.3 State dependent rate

Now suppose that the rate p not only depends on the hour, but somehow also depends on how many visitors came already, $p = p(n, t)$. Perhaps someone is more or is less likely to come when it's busy. In this case the rate is state dependent, the state being the number of visitors that have come already. Another state dependence can for example be that p depends on the time that the last visitor came.

- a. Starting from any time t_0 the probability that the next visitor arrives at time t is still an exponential, but is in this case $p(n, t)e^{-\int_{t_0}^t dsp(n, s)}$. Explain/interpret this exponential distribution.
- b. Starting from t_0 , write down the probability of getting your first visitor at time t_1 AND from that moment t_1 the second visitor at time t_2 .
- c. Formally when you integrate this probability over t_1 between t_0 and t_2 , you end up with $P(t_2; 2, t_0)$ the probability for having the second visitor arrive at time t_2 (and the first somewhere between t_0 and t_2). However the integral does not work out nicely. Why does it not work out to a similar expression, the Gamma distribution of equation 4, as in the previous cases?

When the rate is constant, or even time dependent but not state dependent, you don't actually need a simulation to calculate how many visitors to expect at your party. The Poisson distribution gives you the probability for each number. All the (infinite) possibilities of any number of people arriving at any times are already accumulated in the Poisson and Gamma distribution for arrival times. But when your rate is state dependent you will generally not find such single nice formulas, and you have to resort to simulation.

- d. Build a random number generator for $p_n(t)e^{-\int_{t_0}^t dsp_n(s)}$. For $p_n(t)$ use the hourly rates from table 1 multiplied by

$$\frac{e^{-0.05(n-100)}}{1 + e^{-0.05(n-100)}}$$

I.e. when you have over 100 visitors then people are less and less likely to come in.

- e. Start from 18:00 and generate visitors arriving by generating random times until you pass 02:00 at night. Repeat this procedure 1000 times. Each time remember the total number of visitors. Then make a histogram of the total number of visitors. What is the average/expected number of visitors to your party from your simulation?