3: Sampling from Poisson Process Based on Data

Introduction

In any real application of simulation; we are not going to be given, a priori, any information about the underlying process. We will have to make the following modeling decisions:

1. Does our data come from a homogenous or nonhomogeneous poisson process? (There are, of course, other processes for discrete-time events, but they are beyond the scope of this course)
2. If our data is non-homogeneous, what is its period? (i.e. does our intensity function $\lambda(t)$ repeat on a daily basis, weekly basis, etc)

Once we have made the modeling decision; we must then use data to actually estimate the parameter(s) of the process?

Homogeneous Poisson Process

If we believe that our data comes from a homogeneous Poisson Process (or something very close to it), we need only estimate a single parameter: the intensity λ. There are going to be two cases in which our data is collected, which will necessitate slightly different methods of estimation:

1. We data was collected over a fixed time period $[0, T]$ (i.e. a day, a week, a month, etc)
2. The data was collected until we observed a certain number of arrivals (10 arrivals, 100 arrivals)

Case 1: Data collected over a Fixed Time Period

Suppose we observe $N(T)$ arrivals, then intuitively, our estimate should be:

$$\hat{\lambda} = \frac{N(T)}{T}$$

Owing to the fact that the expected number of arrivals by time T is λT.

Case 2: Data collected until we observe n arrivals

In this case, the end time is random, which can complicate the modeling a bit. Instead we can focus on the fact that the inter-arrival times all following an $\exp(\lambda)$ distribution. If we observe that $X_1 = x_1, \ldots, X_n = x_n$ (X_i being the i-th interarrival time), then this is precisely the problem of maximum likelihood estimation for an exponential distribution:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i} = (\bar{X})^{-1}$$

Nonhomogeneous Poisson Process (One Cycle)

As we have mentioned in the past, the homogeneous Poisson Process is a highly restrictive assumption, and unlikely to be well-reflected in the data for (many, if not most) applications.

The issue, though, with using a nonhomogeneous Poisson process is that we have to estimate the intensity function $\lambda(t)$ or the mean function $\Lambda(t)$ from the data; this is not simple!

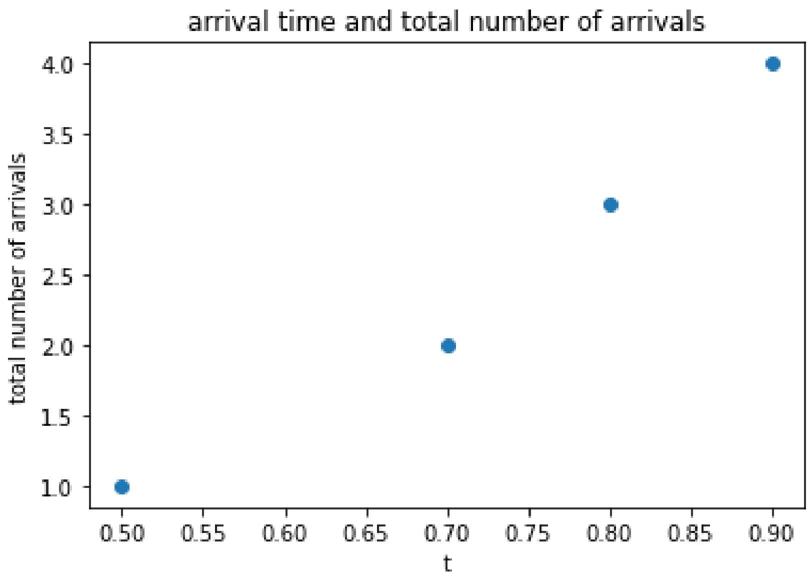
In theory, we could assume a parametric form for $\lambda(t)$ or $\Lambda(t)$ and try and use maximum likelihood estimation for it. This method, though, involves making strong assumptions about the shape of $\lambda(t)$ or $\Lambda(t)$, and can be exceptionally laborious to come up with a good specification. As such, this method is generally not advised.

Instead, we recommend using a non-parametric approach. The non-parametric approach essentially makes minimal assumptions about the form of function $\Lambda(t)$; and adapts as we collect more data.

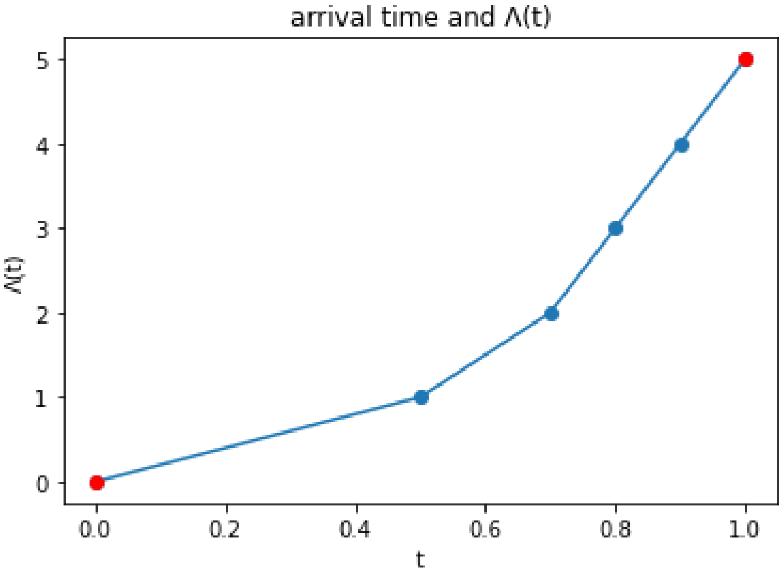
Suppose our nonhomogeneous poisson process is cyclical (i.e. the intensity function repeats at some regular interval c, i.e.: $\lambda(t+c)=\lambda(t)$ for all t. If we're modeling the arrivals of customers to a restaurant, a daily period means we have the same lunch-time rush, afternoon lull, and dinner rush every day. A weekly period might be more realistic, different days have different patterns: Friday and Saturday dinners will be very busy, brunch doesn't exist on the weekdays, etc.

Suppose our data set consists of a single observation of the period (t=0 to T=1). This isn't a great practice; but is good illustrating the basic idea that we'll generalize in the next section.

Let's look at plot of the cumulative arrivals from an NPP:



The simplest continuous $\Lambda(t)$ that is consistent with the data would be one that is equal to the number of cumulative arrivals at the arrival times, and is linearly interpolated in between! That is: $\Lambda(0.5)=1, \Lambda(0.7)=2, \Lambda(0.8)=3, \Lambda(0.9)=4$. Let's also assume that $\Lambda(0)=0$, and for the purposes of having a nice picture $\Lambda(1) = 5$:



The portion of $\Lambda(t)$ between t_i and t_{i+1} is equal to $i + \frac{t-t_i}{t_{i+1}-t_i}$

We can write this out more formally:

$$\Lambda(t) = \begin{cases} \frac{t}{0.5} & , 0 \leq < t < 0.5 \\ 1 + \frac{t-0.5}{0.2} & , 0.5 \leq t < 0.7 \\ 2 + \frac{t-0.7}{0.1} & , 0.7 \leq t < 0.8 \\ 3 + \frac{t-0.8}{0.1} & 0.8 \leq t < 0.9 \\ 4 + \frac{t-0.9}{0.1} & 0.9 \leq t \leq 1 \end{cases}$$

Technically, we could write this a bit more simply since the last few arrivals are evenly spaced, so the slope is the same.

Now that we have written $\Lambda(t)$, we can generate samples from this poisson process using the inversion method:

- Generate the arrival times $a_1, \ldots, a_n$ from a HPP with $\lambda = 1$ from 0 to $\Lambda(T)$
- The arrival times of the NPP are $\Lambda^{-1}(a_1), \ldots, \Lambda^{-1}(a_n)$

To perform linear interpolation in Python, we can use :

- `scipy.interpolate.interp1d(x,y)`
  - `x` and `y` are arrays of values used to approximate some function f: `y = f(x)`. This class returns a function whose call method uses interpolation to find the value of new point.
- We also want use construct the inverse of of the linear interpolation. This is easy, we can simply use `scipy.interpolate.interp1d(y,x)` (the arguments are switched)

Nonhomogeneous Poisson Process (Multiple Cycles)

More commonly, we will observe multiple cycles of our Poisson process (the last section is largely illustrative). Let's suppose the cycle is daily, but you can just as easily substitute another cycle length in here. We get the following observations:

Day 1: $a_1, a_2, \ldots$

Day 2: $a_1, a_2, \ldots$

Day 3: $a_1, a_2, \ldots$

By convention, we will restart at t=0 every day.

etc.

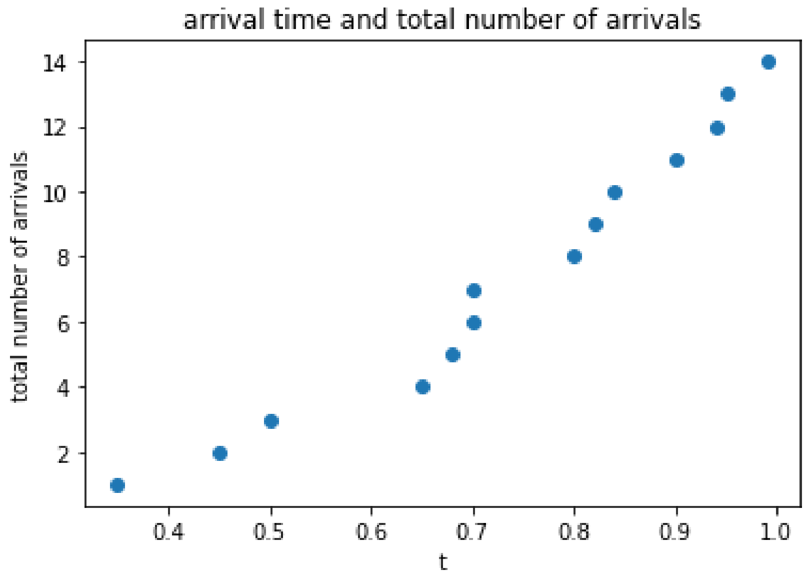
Toy data:

Day1: $t_1 = 0.5$, $t_2 = 0.7$, $t_3 = 0.8$, $t_4 = 0.9$

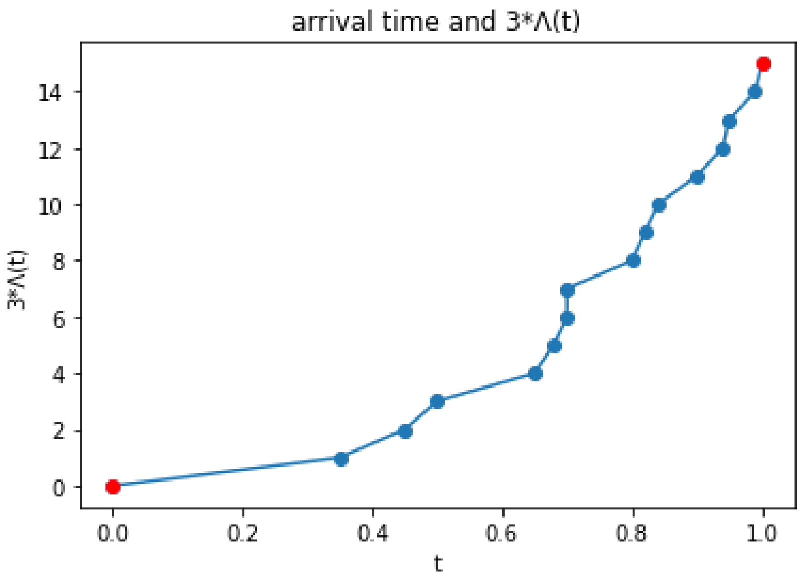
Day2: $t_1 = 0.45$, $t_2 = 0.68$, $t_3 = 0.84$, $t_4 = 0.95$, $t_5 = 0.99$

Day3: $t_1 = 0.35$, $t_2 = 0.65$, $t_3 = 0.7$, $t_4 = 0.82$, $t_5 = 0.94$

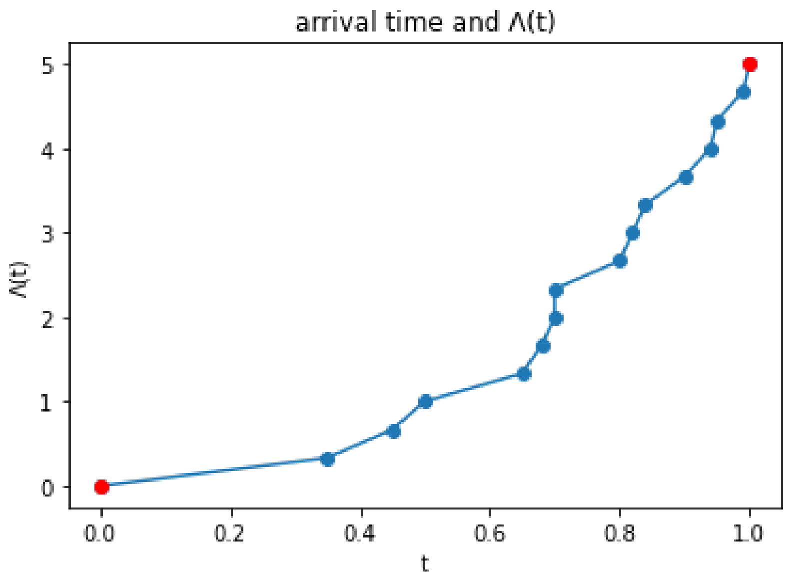
Let's try to combine all the arrivals and plot what we had before:



Again, we can assume that at *t*=1 the last arrival happens. We can then connect the dots:



However, in this case, as we are combining the arrivals from 3 periods instead of just 1 period. Thus, we need to divide the values on y_axis by 3 in order to get $\Lambda(t)$.



If we use k cycles, our estimate of $\Lambda(t)$ between t_i and t_{i+1} is $\frac{i+\frac{t-t_i}{t_{i+1}-t_i}}{k}$

Nonhomogeneous

This code slide does not have a description.

Reading

1

Automatic Zoom