Poisson process

Math 478

April 18, 2019

1 General characterization of a stochastic process

Recall that the most general characterization of a stochastic process is a family of RVs X_t where t is in a time set $\mathcal{T}$. The time set $\mathcal{T}$ can be discrete or continuous, finite or infinite. For example, it can be $\mathbb{Z}, \mathbb{N}, [0,1], [0,\infty)$ etc. A general way to understand a stochastic process X_t is to know its finite dimensional distribution:

$$P(X_{t_1} \in E_1, X_{t_2} \in E_2, \cdots, X_{t_n} \in E_n)$$

for all $t_1, t_2, \cdots, t_n \in \mathcal{T}$ and all $E_1, E_2, \cdots, E_n$ subsets of the state space. On the other hand, depending on the situation, the finite dimensional distribution characterization may not be the best way to characterize the stochastic process. For example, for Markov chain, we use the transitional probabilities approach, which is a variation of the finite dimensional distribution (because the process is Markov and the time set is discrete, specifying the transitional probabilities is the same as specifying the finite dimensional distribution - verify this for yourself). An interesting question is how do we extend the transitional probabilities approach to a continuous time process ? (see the answer below)

A different point of view is to describe the dynamics of the process, i.e. how it moves in time. Roughly speaking, we describe the change of the process rather than the process itself. For example, in discrete time, we specify $\Delta X_k := X_{k+1} - X_k$ for all k. In continuous time, it is more delicate. Roughly here we want to specify $dX_t := X_{t+dt} - X_t$ for arbitrarily small dt. The term dX_t is referred to as the differential of X_t.

Generally speaking, the transitional probabilities and dynamics approach are equivalent. We can transform the characterization using one approach to the other (demonstrate this with random walk in discrete time). Pay attention to how both approaches (finite dimensional / dynamics) are used in describing the Poisson process in this chapter.

2 Overview of Poisson process

A Poisson process has discrete state space $\mathbb{N}$ and continuous time set $[0, \infty)$. We can characterize a Poisson process in two equivalent ways.

Definition 2.1. *$N_t, t \geq 0$ is a Poisson process with rate $\lambda > 0$ if*

- $N_0 = 0$.
- *N_t has independent and stationary increments. That is for $r < s < t$ $N_t - N_s$ is independent of $N_s - N_r$. Moreover, the distribution of $N_t - N_s$ only depends on $t - s$.*
- *$N_t - N_s$ has Poisson($\lambda(t-s)$) distribution.*

Definition 2.2. *$N_t, t \geq 0$ is a Poisson process with rate $\lambda > 0$ if*

- $N_0 = 0$.
- *N_t has independent and stationary increments.*
- $P(N(h) = 1) = \lambda h + o(h)$.
- $P(N(h) \geq 2) = o(h)$.

Observe that the independent increments property makes a process Markov (prove this). The stationary increments property implies that the process dynamics or transitional probabilities is not a function of time (this is reflected in particular in the Poisson process by λ as a constant not depending on t). Finally, We see that the first definition has the flavor of the transitional probabilities approach while the second definition has the flavor of the dynamics approach. Observe that the second definition gives the dynamical description of the increments of the Poisson process at any time t and not only at time $t = 0$. Indeed, because of independent and stationary increments

$$\begin{aligned} P(N_{t+h} = j | N_t = i) &= P(N_{t+h} - N_t = j - i | N_t = i) \\ &= P(N_{t+h} - N_t = j - i | N_t - N_0 = i) \\ &= P(N_{t+h} - N_t = j - i) = P(N_h = j - i), j - i \geq 0, \end{aligned}$$

for the second definition.

Intuitively, a Poisson process N_t models the number of occurences of some type of events (customters arriving at the store, earthquakes) over some time interval (s, t), which is captured by $N_t - N_s$. The iid increments property simply says occurences over non-overlapping time interals are independent and the distribution of the number of occurences is homogeneous in time. The first definition just directly says that this distribution is Poisson $(\lambda(t-s))$. One of the implications here is that longer time interval has higher expected number of occurences which makes intuitive sense. The second definition says that for small time interval of length h (so that $\lambda h < 1$) the distribution of the number of occurences is roughly a Bernoulli with success probability λh. This implies that over a longer time interval (s, t) (whose length $t-s$ is multiple of h) the distribution of the number of occurences is a Binomial success probability λh. More importantly, the expectation of this Binomial random variable is $\lambda(t-s)$, *no matter how small h is.* Thus as we keep s, t constant and pushes $h \to 0$ (so that the success rate converges to 0) by the Poisson approximation for the Binomial we see that the distribution of the number of occurences converges to a Poisson RV rate $\lambda(t-s)$ which is exactly the first definition. Note that by the second definition, the number of occurences of a Poisson process increases by 1 (or not at all) over small time intervals. Thus $N(t)$ can be viewed as a process that keeps the total number of occurences up to time t with single increment. For this reason a Poisson process is also referred to as a *counting process*. Finally in an interal of length h there is an average of λh numbers of events. Thus in 1 unit of time there is an average of λ events happening (e.g. there is λ arrivals per unit of time). This explains why we say λ is the rate of the Poisson process.

We refer the readers to the textbook for a rigorous derivation of the equivalence of the two definitions.

3 Interarrival and waiting time of a Poisson process

The reason why a stochastic process is more than just a family of RVs is because of its time aspect. In other words, we can ask probabalistic questions about the temporal aspects of a stochastic process. We have seen this in the average time spent in recurrent states and other variations with a Markov chain. With a Poisson process, because it has increments by 1 and in general discrete increments, we can ask about the time it takes for a Poisson process to increase by 1. This is referred to as the interarrival time. We can also ask the time it takes for a Poisson process to increase by n for a generic n. This is referred to as the waiting time.

To be more precise, let N_t be a Poisson λ process and $S_n = \inf\{t \geq 0 : N_t = n\}, n \geq 1$. Let $T_n = S_n - S_{n-1}, n \geq 1$, where we take $S_0 := 0$. The first important result is that all the T_n's are iid Exponential(λ) RVs. The iid part intuitively comes from the independent and stationary increments of the Poisson process. The exponential distribution part needs some explanation. We first comment on the rate λ of the exponential distribution. First this means the expected interarrival time is $\frac{1}{\lambda}$, which is the average time it takes for one event to occur after the previous one. Second, this is consistent with the definition of a Poisson process because we know that in an interal of length h there is an average of λh numbers of events. This means the average time for 1 event to occur is $\frac{h}{\lambda h} = \frac{1}{\lambda}$.

Accepting the iid property of the T_n, we only need to show T_1 has Exponential(λ) distribution. But then for any $t \geq 0$

$$\{T_1 > t\} = \{N_t = 0\}.$$

Thus $P(T_1 > t) = P(N_t = 0) = e^{-\lambda t}$ or $P(T_1 \leq t) = 1 - e^{-\lambda t}$ which is the cdf of an Exponetial (λ) distribution.

Finally, since $S_n = \sum_{i=1}^{n} T_i, S_n$ has as Gamma(n, λ) distribution. That is

$$f_{S_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}, t \geq 0.$$

Example 3.1. Suppose that people immigrate into a territory at a Poisson rate $\lambda = 1$ per day.

a. What is the expected time until the tenth arrival?

b. What is the probability that the elapsed time between the 10th and 11th arrival exceeds two days?

Ans :

a. The interarrival time is Exp(λ) with expectation $1/\lambda$. Thus the expected time for the nth arrival is n/λ where $n = 10$ and $\lambda = 1$ here.

b. We have $P(X > 2) = e^{-2\lambda} = e^{-2}$ where $X \sim$ Exp(λ).

4 Compound Poisson process

$Q(t)$ is called a compound Poisson process if

$$Q(t) = \sum_{i=1}^{N_t} Y_i,$$

where N_t is a Poisson(λ) process and Y_i are iid RVs, N_t is independent of Y_i's. We have seen a similar idea before in a compound RV and indeed the value of a compound Poisson process for a fixed t is a compound RV. Therefore, all results we derived for the compound RVs apply here for a fixed t. In particular

$$\begin{aligned} E(Q_t) &= E(Y_1)E(N_t) = E(Y_1)\lambda t \\ Var(Q_t) &= Var(Y_1)E(N_t) + E^2(Y_1)Var(N_t) = \lambda t E(Y_1^2). \end{aligned}$$

One of the interesting choice for the distribution of Y_1 is where Y_1 can take value $m_k, k = 1, \cdots, n$ with corresponding probabilities $p_k, \sum_{k=1}^n p_k = 1$. In this case, we have the decomposition of $Q(t)$ as followed:

$$Q(t) = \sum_{k=1}^{n} m_k N_t^k,$$

where N_t^k are independent Poisson process with rate λp_k. Conversely, if $N_k(t)$ is a Poisson process with rates $\lambda_k, k = 1, \cdots, n$ and they are independent then $Q(t) = \sum_{i=1}^n m_k N_t^k$ has a compound Poissin distribution:

$$Q(t) = \sum_{i=1}^{N_t} Y_i,$$

where N_t is a Poisson (λ) process where $\lambda = \sum_{k=1}^n \lambda_k$ and Y_1 takes on values $m_k, k = 1, \cdots, n$ with probabilities $p_k = \frac{\lambda_k}{\lambda}$. This result is very similar to the problems of the people arriving at a store according to a Poisson distribution where they can be male or female with probability p and $1-p$ respectively. We leave the details of the verification to the readers.

Example 4.1. *Coupon collecting problem*

There are m different types of coupons. A type i coupon appears with probability $p_i, i = 1, \cdots, m$. Suppose all coupons are independent of one another. Find $E(N)$, the number of coupons needed to have a complete collection.

Ans:

If we let X_i be the number of coupons needed to obtain the first ith coupon then $X = \max_i X_i$. On the other hand, the X_i's are not independent (if the ith coupon appears then the jth coupon cannot appear etc.). Therefore it is not straightforward to calculate $E(X)$ this way.

The key to the problem is to embed the coupon collecting process into a continuous time dimension. Specifically, let the coupon collecting process be modelled by a Poisson(λ) process $N(t)$. $N(t)$ keeps track of how many coupons we have collected so far by time t. Let $N_i(t)$ be the number of coupons of type i we have collected by time t. Then $N_i(t)$ are independent Poisson processes with rate $\lambda_i = \lambda p_i, i = 1, \cdots, m$. Let τ_i be the first time that $N_i(t) = 1$. Then the time it takes to collect all coupon types is $\tau = \max_i \tau_i$. The distribution of τ is straightforward:

$$P(\tau \leq t) = \prod_i P(\tau_i \leq t) = \prod_i (1 - e^{-\lambda_i t}) = \prod_i (1 - e^{-\lambda p_i t}).$$

The final observation is that if we let T_i be the interarrival time of the process $N(t)$ then

$$\tau = \sum_{i=1}^{X} T_i.$$

Since X and T_i are independent, we have

$$E(\tau) = E(X)E(T_i) = \frac{E(X)}{\lambda}.$$

Thus

$$E(X) = \lambda E(\tau) = \lambda \int_0^\infty [1 - \prod_i (1 - e^{-\lambda p_i t})] dt,$$

where we have used without proof (left to the reader) the fact that

$$E(\tau) = \int_0^\infty P(\tau > t) dt.$$

Finally by the substitution $u = \lambda t$ we have

$$E(X) = \int_0^\infty [1 - \prod_i (1 - e^{-p_i t})] dt,$$

independent of λ (as it should be).

Finally, when $p_1 = p_2 = \cdots p_n = \frac{1}{n}$ then

$$\begin{aligned} E(X) &= \int_0^\infty [1 - (1 - e^{-\frac{t}{n}})^n] dt \\ &= \int_0^\infty \sum_{i=1}^{n} \binom{n}{i} (-1)^{i+1} e^{-i\frac{t}{n}} dt \\ &= \sum_{i=1}^{n} \binom{n}{i} (-1)^{i+1} \frac{n}{i}. \end{aligned}$$

This apprently is equal to $\sum_{i=1}^n \sum_{i=1}^n \frac{n}{i}$ but I could not find a (quick?) proof for it. I have verified it with Matlab for several values of n and they all hold !

Remark: One could also see that the number of coupons needed for a complete collection is $N(\tau)$. Thus we could have tried to compute $E(N(\tau))$ which is intuitively $\lambda E(\tau)$. However, this computation requires that we establish the independence between $N(t)$ and τ, which may not be as clear as the independence between X and T_i in our actual computation.

5 The exponential distribution

We have seen how the Poisson process is closely connected to the exponential distribution. In fact, the time dimension of the Poisson process is the sum of independent exponential RVs. The exponential distribution is a popular model for time spent doing a certain activity. If we also known that the phenomenon being modelled has the so called "memoryless property" then the exponential distribution is the unique choice for the model. We explore this and further properties of the exponential distribution in this section.

5.1 Memoryless property

A RV X is said to be *memoryless* if for all $s, t \geq 0$

$$P(X > s + t | X > t) = P(X > s).$$

Thinking of X as the "survival" time, the interpretation is that knowing the phenomenon has last beyond time t, asking the probability that it would last an extra s units of time is the same as asking this question at time 0. That is the phenomenon does not “remember” that it has survived the first t units of time. We can also write the above statement as

$$X - t | X > t \sim Exp(\lambda)$$

if $X \sim Exp(\lambda)$. A consequence of this property is the following : let X be an Exponential RV. Then

$$\begin{aligned} E(X - t | X > t) &= E(X) \\ Var(X - t | X > t) &= Var(X). \end{aligned}$$

The proof is as followed. Let $Y = X - t|X > t$. Then we see that Y has the same distribution as X from the memoryless property. Thus $E(Y) = E(X)$ and $Var(Y) = Var(X)$. Observe that

$$\int_t^\infty (X - t) f_X(x) dx \neq E(X - t | X > t).$$

In fact,

$$\int_t^\infty (X - t) f_X(x) dx = E((X - t)\mathbf{1}_{X>t}).$$

Remark: The Geometric distribution also has memoryless property. Indeed, if $X \sim Geometric(p)$ then $P(X > n) = \sum_{i=n+1}^\infty (1-p)^{i-1} p = (1-p)^n$. Therefore $P(X > n|X > m) = P(X > n - m)$. This does not contradict what we said above about the Exponential distribution being the unique one with memoryless property. That statement obviously was meant for continuous RVs. Indeed, we see that the Geometric is the discrete analog of the Exponential(λ) in this sense where $\lambda = -\log(1-p)$. What is more curious is this opens up a possibility to define an analog of the Poisson process in descrete time : one would wait a Geometric(p) time and then jump with size 1 and then repeat (what is the finite dimensional distribution of such a process) ? Another curious feature is in the above coupon problem, the precise issue was because if we view the process from the Geometric clock angle then they are not independent. On the other hand, transforming the process to continuous time and viewing the process from the Exponential clock angle then they are independent, allowing us to compute the distribution of the max !

Example 5.1. Consider a post office that is run by two clerks. Suppose that when Mr. X enters the post office he discovers that Mr. Y and Mr. Z are being served by the two clerks. Mr. X will be served immediately after Mr. Y or Mr. Z is finished. Suppose that the service time for all of these gentlemen are independent Exponential(λ). What is the probability that Mr. X will be the last one to leave the office?

Ans : We analyze the problem the moment Mr. X begins service. At that moment either Mr. Y or Mr. Z is still being served (the probability that both gentlemen finishing at the same time is 0 !). Without loss suppose it's Mr. Z. Noticing that the remaining service time of Mr. Z is still an Exponential (λ) (due to the memoryless property), the probability that Mr. X leaves before Mr. Z is 1/2.

Further question : Answer the same question with the following change :

a) The three gentlemen's service times have different rates.

b) Changing Mr. Y or Mr. Z for Mr. X.

Example 5.2. The dollar amount of damage involved in an automobile accident is an exponential random variable with mean 1000. Of this, the insurance company only pays for the amount exceeding 400. Find the expectation and variance of the amount paid per accident.

Ans: Let X be an Exponential(λ) RV. The amount the insurance company pays is $\max(X-400,0)=(X-400)^+$. Normally we would proceed with

$$E((X-400)^+) = E((X-400)^+|X) = \int_{400}^{\infty}(x-400)e^{-\frac{x}{1000}}dx.$$

On the other hand, because X is an Exponential RV we can utilize the memoryless property as followed. Let $Y=\mathbf{1}_{X>400}$. Then

$$\begin{aligned}(X-400)^+ \mid (Y=1) &= X-400 \mid Y=1 \sim Exp(\frac{1}{1000})\\ (X-400)^+ \mid (Y=0) &= 0.\end{aligned}$$

Thus we see that $(X-400^+|Y$ is a RV whose distribution is an Exp(1/1000) when $Y=1$ (with probability $e^{-\frac{4}{10}}$) and is 0 when $Y=0$ (with probability $1-e^{-\frac{4}{10}}$.) Therefore

$$E((X-400)^+|Y) = 1000\ \mathbf{1}_{Y=1},$$

and

$$\begin{aligned}E((X-400)^+) &= E(E((X-400)^+|Y))\\ &= 1000P(Y=1) = 1000e^{-\frac{4}{10}},\end{aligned}$$

Similarly,

$$Var((X-400)^+) \quad = \quad E\big[Var((X-400)^+|Y)\big] + Var\big[E((X-400)^+|Y)\big].$$

Now

$$Var((X-400)^+|Y) = 1000^2\ \mathbf{1}_{Y=1},$$

since $(X-400)+$ is a constant when $Y=0$. Thus

$$\begin{aligned} Var((X-400)^+) &= 1000^2 P(Y=1) + 1000^2 P(Y=1)(1-P(Y=1)) \\ &= 1000^2 e^{\frac{-4}{10}}(2-e^{\frac{-4}{10}}). \end{aligned}$$

5.2 Hazard rate

Let X be a positive continuous RV. The hazard (failure) rate of X, denoted as $r(t)$, is defined as

$$r(t) = \frac{f(t)}{1-F(t)}.$$

The hazard rate has the following interpretation:

$$\begin{aligned} r(t)dt &= \frac{f(t)dt}{1-F(t)} \approx \frac{P(X \in (t, t+dt)}{P(X > t)} \\ &= P(X \in (t, t+dt)|X > t) = P(X < t + dt|X > t). \end{aligned}$$

Thus the quantity $r(t)dt$ is approximately the conditional probability that the RV survives only (fails after) an extra dt units of time given that it has survived t units of time. Thus the higher $r(t)$, the more likely that X breaks down after having survived until time t. Moreover, we also have

$$F(t) = 1 - \exp(-\int_0^t r(u)du).$$

(See proof in Ross). Thus a distribution is uniquely determined by its hazard rate and vice versa. From thie equation, it's also clear that an exponential random variable with mean $\frac{1}{\lambda}$ as a constant hazard rate λ. It is sometimes referred to as just an exponential RV with rate λ. Because the hazard rate is constant, we can also intuitively see the reason for its memoryless propery (it is just as likely to fail after having survived t units of time as at the beginning).

Example 5.3. *The failure rate of a hyper-exponential RV*

A bin contains n different types of batteris, with battery of type i lasting an exponential distributed time with rate $\lambda_i, i = 1, \cdots, n$. Also let p_i be the proportion of batteries in the bin of type i. Let Y be the lifetime of a battery randomly chosen. Find the failure rate of Y.

Ans: Let N be the type of the battery selected. That is $N = i$ with probability p_i. Also let the life time of the ith type battery be $X_i, i = 1 \cdots, n$. Obseve that N is independent of X_i's. We have $Y = X_N$ and we refer to Y as a hyper-exponential distribution.

$$\begin{aligned} F_{X_N}(t) = P(X_N < t) &= \sum_i P(X_N < x | N = i) p_i \\ &= \sum_i P(X_i < t) p_i = \sum_i p_i (1 - e^{-\lambda_i t}). \end{aligned}$$

Thus

$$f_{X_N}(t) = \sum_i p_i \lambda_i e^{-\lambda_i t}.$$

Finally,

$$r_{X_N}(t) = \frac{\sum_i p_i \lambda_i e^{-\lambda_i t}}{\sum_i p_i e^{-\lambda_i t}}.$$

It is easy to see that

$$\lim_{t \to \infty} r_{X_N}(t) = \min_i \lambda_i.$$

Thus the hazard rate of a randomly chosen battery converges to the hazard rate of the "best" battery. This makes sense because the longer the battery survives, the more likely that it is the "best" type.

5.3 Minimum of exponentials and some simple queueing models

Since Exponential RVs are the default model for waiting time, the minimum of independent exponentials serves as the model of first arrival time of a series of events. Let $X_1, X_2, \cdots, X_n$ be iid Exponentials(λ_i), $i = 1, \cdots, n$. We have

$$\begin{aligned} P(\min_i X_i \geq x) &= P(X_i \geq x, \forall i) = \prod_i P(X_i \geq x) \\ &= \prod_i e^{-\lambda_i x} = e^{-x \sum_i \lambda_i}. \end{aligned}$$

Thus the minimum of independent exponentials is also an exponential with rate being the sum of the components. Moreover, for any particular X_i,

$$\begin{aligned}
P(X_i = \min_{j\neq i} X_j) &= P(X_i \leq X_j, \forall j \neq i) \\
&= \int_0^\infty \lambda_i P(x \leq X_j, \forall j \neq i | X_i = x) e^{-\lambda_i x} dx \\
&= \int_0^\infty \lambda_i P(x \leq X_j, \forall j \neq i) e^{-\lambda_i x} dx \\
&= \int_0^\infty \lambda_i \prod_{j\neq i} P(x \leq X_j) e^{-\lambda_i x} dx \\
&= \int_0^\infty \lambda_i \prod_{j\neq i} e^{-\lambda_j x} e^{-\lambda_i x} dx \\
&= \int_0^\infty \lambda_i e^{-\sum_j \lambda_j x} dx \\
&= \frac{\lambda_i}{\sum_j \lambda_j}.
\end{aligned}$$

This result makes intuitive sense because the larger the rate of the exponential (the smaller its mean), the more likely it is the minimum of the exponential collection.

Example 5.4. *A queueing example with two servers and no waiting line*

Suppose you arrive at a post office having two clerks who are both busy but there is no one else waiting in line. You will enter service when either clerk becomes free. Let service time for clerk i be exponential with rate $\lambda_i, i = 1, 2$. Find $E(T)$, the expected amount of time that you spend in the post office.

Ans: Let X_i be clerk i's service time for the current customer and Y_i be the clerk i's service time for "you," conditional on the fact that "you" are served by clerk i. Then

$$T = \min_i X_i + Y_1 \mathbf{1}_{X_1 \leq X_2} + Y_2 \mathbf{1}_{X_2 < X_1}.$$

Now $E(\min_i X_i) = \frac{1}{\lambda_1+\lambda_2}$. And

$$\begin{aligned}
E(Y_1 \mathbf{1}_{X_1 < X_2}) &= E(Y_1 | X_1 \leq X_2) P(X_1 \leq X_2) \\
&= \frac{1}{\lambda_1} \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{1}{\lambda_1 + \lambda_2}.
\end{aligned}$$

Similarly,

$$E(Y_2 \mathbf{1}_{X_2 < X_1}) = \frac{1}{\lambda_1 + \lambda_2}.$$

Thus $E(T) = \frac{3}{\lambda_1+\lambda_2}$.

Example 5.5. *A queueing example with one server and a waiting line*

Suppose that customers are in line to receive service that is provided sequentially by a server. Once a service is complete, the next service begins immediately. However, each waiting customer will only wait an exponentially distributed time with rate θ; if its service has not yet begun by this time then it will depart the system. The service time for each customer is exponetial with rate μ. Suppose all times are independent. Suppose that someone is being served and consider the nth person in line.

a. Find P_n, the probability that this customer is eventually served.

b. Find W_n, the conditional expected amount of time this person spends waiting in line given that he / she is eventually served.

Ans :

a. Let's analyze the question from the point of view of the nth person in the queue. Clearly to be served, he must move up in the queue until he reaches position 0 (being the person being served). How does he get there? There is the first moment when either one of the $n-1$ people in the queue or the one being served leaves the system. If the nth person survived (did not leave before such moment) then he will move up one position. The process then repeats itself. If he survives n such moments then he will be served. Incidentally, the process where people in front of the nth person leaving the queue can be viewed as a time non-homogeneous Poisson process.

Thus we let T_n to be the first time a person leaves the system where there are $n-1$ people in the queue and 1 person in the server. Then T_n is the minimum of $n-1$ Exponentials(θ) and one Exponential(μ), all independent. Thus T_n has an Exponential$[(n-1)\theta+\mu]$ distribution. We can continue in the same fashion to define T_{n-1}, the first time afer T_n when a person leaves the system where there are $n-2$ people in the queue and 1 person in the server (if the person left was in the server then one person from the queue will move up immediately to the server, leaving $n-2$ people in the queue). Then T_{n-1} has an Exponential$[(n-2)\theta+\mu]$ distribution. Finally T_1 has an Exponential(μ) distribution, being the situation where there's 1 person in the server and only "you" (or the nth person) waiting in line to be served. Observe that the T_i's are independent.

Let Y_n be the nth person "survival time". Then he will be served if $Y_n > \sum_{k=1}^{n} T_i$. On the other hand, the distribution of $\sum_{i=1}^{n} T_i$ is not elegant to deal with (it is related to the Gamma but not quite, since T_i's don't have the same rate). The technique to circumvent this is clearly recursion by conditioning. That is

$$\begin{aligned}
P_n = P(\text{ nth person served }) &= P(\text{ nth person served } |Y_n \le T_n)P(Y_n \le T_n) \\
&+ P(\text{ nth person served } |Y_n > T_n)P(Y_n > T_n) \\
&= P(\text{ nth person served } |Y_n > T_n)P(Y_n > T_n) \\
&= P_{n-1}\frac{(n-1)\theta + \mu}{n\theta + \mu}.
\end{aligned}$$

The last equality is from the interpretation that if the person has survived until one person in front of him left the queue, then it is the same as if he has to wait with $n-1$ people left in the queue (the memoryless property of the Exponentials play a crucial role here !). Finally the above product is a telescoping product with the last term $P_1 = \frac{\mu}{n\theta+\mu}$. Thus

$$P_n = \frac{\mu}{n\theta + \mu}.$$

Remark : The above analysis suggests the way to do the following computation :

$$P(X > \sum_{i=1}^{n} Y_i),$$

where X is Exp(μ) and Y_i are Exp(λ_i), all independent. In fact, the following is the rigorous justification of the above argument. We have

$$\begin{aligned}
P(X > \sum_{i=1}^{n} Y_i) &= P(X > \sum_{i=1}^{n} Y_i | X > Y_1)P(X > Y_1) \\
&= P(X > \sum_{i=1}^{n} Y_i | X > Y_1 + Y_2)P(X > Y_1 + Y_2 | X > Y_1)P(X > Y_1) \\
&= \cdots
\end{aligned}$$

The only thing left to justify is

$$P(X > Y_1 + Y_2 | X > Y_1) = P(X > Y_2).$$

This obviously reminds us of the memoryless property and can be done by conditioning on $Y_1 = s, Y_2 = t$.

b. If the nth person is actually served, then the wait time is just $\sum_{i=1}^{n} T_i$. Thus the (conditional) expected wait time is

$$\sum_{i=1}^{n} E(T_i) = \sum_{i=1}^{n} \frac{1}{i-1)\theta + \mu}.$$