Exponential Distribution

→ A continuous distribution used to model the waiting time.
→ "Continuous version of the geometric distribution".

Definition: A random variable T is said to have an exponential distribution with rate λ, if its cumulative distribution function is $1-e^{-\lambda t}$.

Properties: Let $T \sim Exp(\lambda)$ be an exponential r.v. with rate λ:

① $\Pr[T \leq t] = 1 - e^{-\lambda t} \quad \forall\, t \geq 0$

② $f_T(t) = \lambda e^{-\lambda t} \quad \forall\, t \geq 0$

③ $E[T] = \frac{1}{\lambda}$, ④ $Var[T] = \frac{1}{\lambda^2}$

→ You can think of $Exp(\lambda)$ as the time one must wait before a rare event that happens λ times every second, occurs. The rare event is such that at any given moment, the probability of it happening shortly after is the same.
→ The smaller λ is, the less likely will the event happen, and the more we should wait for its occurence.

Memorylessness The exponential distribution is the only continuous, memoryless random variable. A random variable T is memoryless if

$$\forall\, s,t \geq 0 \quad \Pr[T \geq s+t \mid T \geq s] = \Pr[T \geq t].$$

→ The propability that we have to wait at least t seconds is the same regardless of how much we have already waited.

Exponential Distribution (Cont.)

Exponential Races: Let $T_1, \dots, T_n$ be independent, $T_i \sim Exp(\lambda_i)$ and let $S = \min(T_1, \dots, T_n)$. Then $S \sim Exp(\lambda_1 + \dots + \lambda_n)$. Furthermore, the index of the minimum is distributed proportional to the rates. That is,

$$\Pr[T_i = \min(T_1, \dots, T_n)] = \frac{\lambda_i}{\lambda_1 + \dots + \lambda_n}.$$

→ Here's one way to think about it. There is a memoryless clock that rings λ times every minitue (on average). Now, T is the time we must wait for this clock to ring.

Now if we have n independent clocks, we can think of them as one big clock that rings $\lambda_1 + \dots + \lambda_n$ times every minitue. Thus, the exponentiality of their minimum.

The Poisson Clock!

Sum: Let $T_1, \dots, T_n$ be i.i.d. $Exp(\lambda)$ variables. Their sum, $\tau_n = \sum_{i=1}^{n} T_i$ follows Gamma (n, λ) distribution:

$$f_\tau(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}$$

→ Think of τ_n as the time it takes for the Poisson clock to ring n times.

Poisson Distribution

→ A discrete distribution used to model the number of events occurring in a fixed interval, if the events occur with a constant mean rate and independently of the time since the last event.

Definition: A random variable N is said to have a poisson distribution with parameter $\lambda > 0$, written as $N \sim \text{Poisson}(\lambda)$, if:

$$\mathbb{P}r[N=n] = \frac{e^{-\lambda}\lambda^n}{n!} \quad \text{for all } n \in \{0,1,2,\dots\}$$

Properties:

① $\mathbb{P}r[N=n] = \frac{e^{-\lambda}\lambda^n}{n!}$

② $\mathbb{E}[N] = \text{Var}[N] = \lambda$

→ If a Poisson clock rings λ times every minitue, on average, then the number of times that it rings in a minitue follows $\text{Poisson}(\lambda)$.

$\boxed{\text{Why? Later!}}$

→ Another intuitive note: Consider the Poisson clock that rings λ times every minitue. The number of times it rings in time t follows $\text{Poisson}(\lambda t)$.
on average

Poisson Process: I (Constructive View)

Let $T_1, T_2, \dots$ be i.i.d. rv's with distribution $T_i \sim Exp(\lambda)$.
Define $\tau_n = \sum_{i=1}^{n} T_i$. We can think of τ_n as the time that a rate λ poisson clock rings for the n-th time.

→ As was mentioned earlier (P. 2), $\tau_n \sim \Gamma(n, \lambda)$.

Definition: For every $t \geqslant 0$, let N_t be the number of times that the Poisson clock has rung, until time t:

$$\forall t \geqslant 0 \quad N_t := |\{n : \tau_n \leqslant t\}|$$

Definition: $(N_t)_{t \geqslant 0}$ is called a Poisson Process with rate λ.

Distribution of N_t: First, note that $supp(N_t) = \{0, 1, 2, \dots\}$.

N_t is the number of times a Poisson clock rings in time t. Intuitively it must ring λt times in this period, on average. Indeed,

$$\boxed{N_t \sim Poisson(\lambda t)}$$

Proof: $\Pr[N_t = n] = \Pr[\tau_n \leqslant t, \tau_{n+1} > t] = \Pr[\tau_n \leqslant t, T_{n+1} > t - \tau_n]$

$$= \int_{s=0}^{t} \underbrace{f_{\tau_n}(s)}_{\text{Pdf of } \Gamma(n,\lambda)} \cdot \underbrace{\Pr[T_{n+1} > t-s]}_{1 - \text{CDF of } Exp(\lambda)} \, ds = \dots = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \sim Poisson(\lambda t)$$

$\blacksquare$

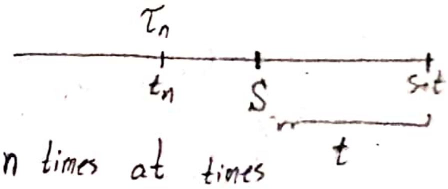
Poisson Process : I (Cont.)

Consider N_t's that were defined. Intuitively, $N_t - N_s$ is the number of times that the clock rings from time s until t. It should be that this quantity follows Poisson$(\lambda(t-s))$ distribution and is independent of how many times the clock has rung before time s.

<u>Theorem:</u> [Distribution of $N_{t+s} - N_s$, independence from N_r, $r \leq s$]

For all $t \geq 0$, $N_{t+s} - N_s \sim$ Poisson(λt) and is independent of N_r, $0 \leq r \leq s$.

<u>Proof:</u>



Suppose that by time s, the clock has rung n times at times

$\tau_1 = t_1, \tau_2 = t_2, \dots, \tau_n = t_n$.

The waiting time for the $n+1$th ring must be after $s - t_n$, but by the memorylessness of the exponential distribution,

$$\forall t \geq 0: \Pr[\overbrace{T_{n+1} > s - t_n + t}^{n+1\text{th ring after } s+t} \mid \overbrace{T_{n+1} > s - t_n}^{n+1\text{th ring after } s}] = \Pr[T_{n+1} > t] = e^{-\lambda t}.$$

So, the distribution of the time of the first ringing <u>after</u> s is Exp(λ) and independent of $\tau_1, \dots, \tau_n$. It is clear that $T_{n+2}, T_{n+3}, \dots$ are independent of $\tau_1, \dots, \tau_n, T_{n+1}$. Thus the ringing times after $\underline{s}$ are all i.i.d. Exp(λ).

Therefore $N_{t+s} - N_s$ is a Poisson process independent of what has happened before.

We showed that: given that the nth ring was before $\underline{s}$, the time we have to wait for the $n+1$th ring is Exp(λ). But, if we are told that the $n+1$th ring doesn't happen before $\underline{s}$, then the waiting time <u>from s</u> is also Exp(λ).

Poisson Process : II

We saw one way of defining a Poisson process is by using a Poisson clock that rings with rate λ, giving us a sequence of ring-times $\tau_1, \tau_2, \dots$.

We also showed that in a Poisson process, $N_{s+t} - N_s \sim \text{Poisson}(\lambda t)$ and that the number of rings in separate intervals are independent.

We will now show the converse: Any process $(N_t)_{t \in \mathbb{R}^+}$ that satisfies the above conditions is indeed a Poisson process, that is, the time between its "jumps" are iid $\text{Exp}(\lambda)$ variables.

Theorem: [Alternative definition of the Poisson process]

A process $(N_t)_{t \in \mathbb{R}^+}$ is a Poisson process with rate λ iff:

i) $N_0 = 0$

ii) $N_{s+t} - N_s \sim \text{Poisson}(t\lambda)$

iii) # of jumps in N_t in non-overlapping intervals are independent.

Proof:

($\rightarrow$) The theorem in page 6 shows that if N_t is a Poisson process, it satisfies (ii) and (iii). Condition (i) is trivially true.

($\leftarrow$) Consider the sequence of points at which N_t "jumps" (the clock rings) : $\tau_1, \tau_2, \dots$. Let $T_i = \tau_i - \tau_{i-1}$ with $T_1 = \tau_1$.

We shall prove that T_i's are iid $\text{Exp}(\lambda)$ variables

Poisson Process: II (Cont.)

$(\Leftarrow)$ $\Pr[T_1 = \tau_1 > t] = \Pr[N_t = 0] = \Pr[N_t - 0 \sim \text{Poisson}(\lambda t) = 0] = e^{-\lambda t}$

So $T_1 \sim Exp(\lambda)$.

$\Pr[T_2 > t \mid T_1 = s] = \Pr[N_{t+s} - N_s = 0 \mid T_1 = s] = \Pr[\text{Poisson}(\lambda t) = 0] = e^{-\lambda t}$

So $T_2 \sim Exp(\lambda)$ and T_1, T_2 are independent.

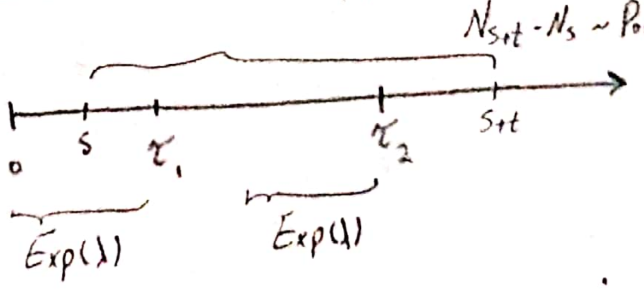
Applying the same argument will show that $T_i \perp\!\!\!\perp T_1, \dots, T_{i-1}$ and that $T_i \sim Exp(\lambda)$.

Putting it all together:

① A Poisson process with rate λ is given by $(N_t)_{t \in \mathbb{R}^+}$ where N_t's take values in $\{0, 1, 2, \dots\}$.

② The time we have to wait for N_t to "jump" (i.e., increase by 1) follows an $Exp(\lambda)$ distribution (a Poisson clock) and is independent of the past.

③ The number of jumps in a length t interval follows a $\text{Poisson}(\lambda t)$ distribution and is independent of what happens in other intervals.



Transformations

SuperPosition: Adding a number of independent processes is called superposition.

Theorem: [Sum of independent Poisson processes is itself a Poisson process]

Suppose $N_1(t), \dots, N_k(t)$ are independent Poisson processes with rates $\lambda_1, \dots, \lambda_k$. $N_1(t) + \dots + N_k(t)$ is a Poisson process with rate $\lambda_1 + \dots + \lambda_k$.

Proof: Waiting time is now $T = \min(T_1, \dots, T_k) \sim Exp(\lambda_1 + \dots + \lambda_k)$.

Because waiting times are independent (memorylessness) and i.i.d, the theorem follows. $\boxtimes$

Thinning: We can also break down a Poisson process into independent ones.

Theorem: If we label each point (jump) of a Poisson process with rate λ independently with $1, \dots, n$ with probabilities $p_1, \dots, p_n$, the points with similar labels constitute independent Poisson processes with rates $p_1\lambda, \dots, p_n\lambda$.

Proof: Use the N_t definition of Poisson process. For $n=2$,

$$\Pr[N_t^1 = n_1, N_t^2 = n_2] = \Pr[N_t = n_1+n_2] \binom{n_1+n_2}{n_1} p_1^{n_1} p_2^{n_2}$$

$$= \dots = \frac{e^{-\lambda p_1 t} (p_1\lambda t)^{n_1}}{n_1!} \times \frac{e^{-\lambda p_2 t} (p_2\lambda t)^{n_2}}{n_2!}$$

Transformations (Cont.)

Conditioning: Given the number of times a clock has rung in an interval, its ringing times are distributed uniformly.

Theorem: Let $U_1, \dots, U_n$ be i.i.d Uniform $[0, t]$ variables. and $V_1, \dots, V_n$ be their sorted permutation.

$$[(\tau_1, \dots, \tau_n) \mid N_t = n] \sim (V_1, \dots, V_n)$$

Continuous-Time Markov Chains

Markov Property:

A continuous-time stochastic process is a sequence of r.v.'s $(X_t)_{t\geq 0}$ which has the Markov property, i.e., future is independent of past, given the present.

Definition: [Markov property]

$(X_t)_{t\geq 0}$ satisfies the Markov property if for all $s,t \geq 0$,

$$\boxed{\forall y \in S \quad \Pr[X_{s+t} = y \mid X_r \quad 0 \leq r \leq s] = \Pr[X_{s+t} = y \mid X_s]}$$

→ Another way of saying this is to say that for every $n \in \mathbb{N}$, past times $0 \leq r_1 < \dots < r_n < s$ and states $x_1, \dots, x_n \in S$,

$$\Pr[X_{s+t} = y \mid X_{r_1} = x_1, \dots, X_{r_n} = x_n, X_s = x] = \Pr[X_{s+t} = y \mid X_s = x]$$

Homogeneity:

A Markov chain is homogeneous if its transitions are independent of starting time:

$$\forall s,t \geq 0 \quad \Pr[X_{s+t} = y \mid X_s = x] = \Pr[X_t = y \mid X_0 = x]$$
$$= P^{(t)}(x,y)$$

Continuous Time Markov Chains (Cont.)

Based on what has been said, a homogeneous Markov chain is determined by its transition matrices $P^{(t)}$ that specify the transition probabilities after time t:

$$P^{(t)}(x,y) = \Pr[X_t = y \mid X_0 = x]$$

Linearity:

Although it may seem that infinitly many matrices are required to specify a continuous time chain, Chapman-Kolmogorov equation suggest that they all are related and share a common structure:

Assume $|S| < \infty$. Let $P^{(t)}$ be the transition matrix for intervals of length t.

$$\Pr[X_{s+t} = y \mid X_0 = x] = \sum_{z \in S} \Pr[X_t = y \mid X_0 = z]\, \Pr[X_s = z \mid X_0 = x]$$

$$\Rightarrow P^{(t+s)} = P^{(t)} P^{(s)}$$

$\rightarrow$ If we fix s, e.g., be setting $s=0$, then $P^{(t+s)}$ is a linear function of t. This means that the time-derivative of $P^{(t)}$ will completly determine $P^{(t)}$ for all t.

Continuous Time Markov Chains : I

One way of looking at these processes is by considering a regular discrete time process and a Poisson clock at every state. We only move in the discrete chain when the Poisson clock at the current state rings.

Theorem: [Transitions happen when the Poisson clock rings].

Assume that $X_0 = x$. Let τ_1 be the time of the first transition (change of state). The Markov property implies that $\tau_1 \sim Exp(\lambda_x)$

proof:

$$\mathbb{P}r[\tau_1 > t+s \mid \tau_1 > s] = \mathbb{P}r[X_r = x \quad 0 \leq r \leq s+t \mid X_r = x \quad 0 \leq r \leq s]$$

$$\underset{(MP)}{=} \mathbb{P}r[X_r = x \quad s \leq r \leq s+t \mid X_s = x] \overset{\text{homog}}{=} \mathbb{P}r[X_r = x \quad 0 \leq r \leq t \mid X_0 = x]$$

$$= \mathbb{P}r[\tau_1 > t]$$

Therefore, τ_1 is memoryless and it has a $Exp(\lambda_x)$ distribution.

$\boxtimes$

→ Once the Poisson clock rings, we transition according to the *routing matrix* r. This matrix is defined as:

$$r(x,y) := \mathbb{P}r[X_{\tau_1} = y \mid X_0 = x].$$

rate of change from x to y.

$q(x,y)$

→ We can also imagine a Poisson clock with rate $\boxed{\lambda_x r(x,y)}$ on the edge connecting x and y. The Markov chain moves from x to y when this clock rings. (Using thinning) we could say that in each state, we wait for an adjacent clock to ring and use the edge whose clock rings first.

Continuous Time Markov Chains : II

We will now return to examine $P^{(t)}$, the transition matrix after time t. Remember from p.12 that due to its linearity, it suffices to calculate $\frac{d}{dt}P^{(t)}$ to find $P^{(t)}$ for all values of t.

Using the definition of derivative and Chapman-Kolmogorov equality, we can write

$$\frac{d}{dt}P^{(t)} = \lim_{\varepsilon\to 0}\frac{P^{(t+\varepsilon)} - P^{(t)}}{\varepsilon} = \lim_{\varepsilon\to 0}\frac{P^{(t)}(P^{(\varepsilon)} - I)}{\varepsilon} = P^{(t)}\lim_{\varepsilon\to 0}\frac{P^{(\varepsilon)} - I}{\varepsilon}$$

But since $P^{(0)} = I$, the last limit is just the derivative of $P^{(t)}$ evaluated at $\underline{0}$. So we have:

$$\frac{d}{dt}P^{(t)} = P^{(t)}\frac{d}{dt}P^{(0)}$$

The next step is to find $P^{(\varepsilon)}$ so that we can find $\frac{d}{dt}P^{(0)}$. To do so, notice that if $\varepsilon \ll 1$ is small enough, the Poisson clocks ring at most once in the interval $[0,\varepsilon]$. This will simplify calculating P^{ε} as we have:

$$\boxed{P^{(\varepsilon)}(x,y) = \Pr[X_\varepsilon = y \mid X_0 = x]}$$

If we know that at most <u>one</u> transition occurs in $[0,\varepsilon]$, the above probability can be derived from the routing matrix r.

→ Once we find $P^{(\varepsilon)}$, we can also find $\frac{d}{dt}P^{(0)}$. This will then help us solve the differential equation $\frac{d}{dt}P^{(t)} = P^{(t)}\frac{d}{dt}P^{(0)}$ to find $P^{(t)}$.

Theorem: [Poisson clock rings at most once in short time intervals.]

Let N_ε be the number of times the Poisson clock in one of the states rings in the time interval $[0, \varepsilon]$. For sufficiently small ε, $N_\varepsilon \leq 1$ with high probability.

Proof: N_ε follows the Poisson$(\lambda\varepsilon)$ distribution. Using the Taylor expansion of the exponential function, we have

$$\left.\begin{aligned} \Pr[N_\varepsilon = 0] &= e^{-\lambda\varepsilon} = 1 - \lambda\varepsilon + o(\lambda\varepsilon) \\ \Pr[N_\varepsilon = 1] &= \lambda\varepsilon e^{-\lambda\varepsilon} = \lambda\varepsilon + o(\lambda\varepsilon) \end{aligned}\right\} \Rightarrow \Pr[N_\varepsilon > 1] = o(\lambda\varepsilon)$$

$\boxtimes$

Having proved the above theorem, we can now condition on $N_\varepsilon = 0$ or 1 to find the transition probabilities for the small time interval $[0, \varepsilon]$:

$$P^{(\varepsilon)}(x,y) = \Pr[X_\varepsilon = y \mid X_0 = x] = \begin{cases} 1 - \lambda_x \varepsilon + o(\lambda_x \varepsilon) & x = y \\ \lambda_x r(x,y)\varepsilon + o(\lambda_x r(x,y)\varepsilon) & x \neq y \end{cases}$$

(If $\max(\lambda_x) < \infty$ we can replace $o(\cdot)$s with $o(\varepsilon)$)

Now that we know $P^{(\varepsilon)}$, the stage is set for calculating $\frac{d}{dt} P^{(t)}$

Continuous-Time Markov Chains: II (Cont.)

Transition Rate Matrix (Infinitesimal Generator)

Let $Q = \frac{d}{dt} P^{(0)} = \lim_{\varepsilon \to 0} \frac{P^{(\varepsilon)} - I}{\varepsilon}$. We call Q the transition rate matrix and we have:

$$Q(x,x) = \lim_{\varepsilon \to 0} \frac{1 - \lambda_x \varepsilon + o(\varepsilon) - 1}{\varepsilon} = -\lambda_x .$$

$$Q(x,y) = \lim_{\varepsilon \to 0} \frac{\lambda_x r(x,y) \varepsilon + o(\varepsilon)}{\varepsilon} = \lambda_x r(x,y).$$

So, the transition rate matrix has the following form:

$$Q = \begin{bmatrix} -\lambda_1 & \lambda_1 r(1,2) & \cdots & \lambda_1 r(1,n) \\ \lambda_2 r(2,1) & -\lambda_2 & \cdots & \lambda_2 r(2,n) \\ \vdots & & \ddots & \\ \lambda_n r(n,1) & \cdots & & -\lambda_n \end{bmatrix}$$

where λ_i is the rate at which the Poisson clock at state i rings and $r(i,j)$ is the probability of transitioning from i to j upon the clock ringing.

Having identified $Q = \frac{d}{dt} P^{(0)}$, we can now solve for $P^{(t)}$.

Continuous-Time Markov Chains (Cont.)

Kolmogorov Forward Equation:

Recall from p.14 that we had the following differential equation describing $P^{(t)}$:

$$\frac{d}{dt} P^{(t)} = P^{(t)} Q$$

$$\Rightarrow P^{(t)} = e^{tQ} = \sum_{n=0}^{\infty} \frac{(tQ)^n}{n!}$$

known as the Kolmogorov's forward equation.

Similarly we can get the backward equation:

$$\frac{d}{dt} P^{(t)} = Q P^{(t)}$$

Stationary Distribution

Consider the transition matrix after time t, $P^{(t)}$. If the starting distribution is $X_0 \sim \pi$, then $X_t \sim \pi P^{(t)}$.

For π to be the stationary distribution, we must have

$$\pi P^{(t)} = \pi .$$

Taking the derivative with respect to t, we get

$$\frac{d}{dt} \pi P^{(t)} = 0 \quad \Rightarrow \underbrace{\pi P^{(t)}}\, Q = 0$$
$$\Rightarrow \pi Q = 0$$

So π must be in the (left) kernel of the transition rate matrix, Q.

Irreducibility :

We can extend the notion of irreducibility to continuous-time chains. Remember that in the discrete time setting, irreducibility meant that every state was reachable from every other state. Here we again use the same definition.

Definition: Two states $x, y \in S$ communicate, written as $x \to y$, if there exists a sequence of states $x_0 = x, x_1, \ldots, x_n = y$ such that for all i,

$$q(x_i, x_{i+1}) = \lambda_{x_i} r(x_i, x_{i+1}) > 0 .$$

Definition: A Markov chain is irreducible if $\forall x, y \in S$, $x \to y$.

Convergence Theorem

Theorem: If a continuous time Markov chain is irreducible and has an stationary distribution π, then

$$\lim_{t\to\infty} \mu_t = \pi$$

Theorem: Let $(X_t)_{t\geq 0}$ be a continuous-time Markov chain with an irreducible, positive-recurrent jump chain (specified by the routing matrix r).

Suppose the unique stationary distribution of the jump chain is given by $\tilde{\pi} = [\tilde{\pi}_0, \tilde{\pi}_1, \dots]$. Also assume that $0 < \sum_{k\in S} \frac{\tilde{\pi}_k}{\lambda_k} < \infty$.

Then,

$$\pi_j = \lim_{t\to\infty} \Pr[X_t = j \mid X_0 = i] = \frac{\frac{\tilde{\pi}_j}{\lambda_j}}{\sum_{k\in S} \frac{\tilde{\pi}_k}{\lambda_k}}.$$

For all $i, j \in S$. This means that π is the limiting distribution of the chain.