[AI2613 Lecture 9] Poisson Distribution, Poisson Process

May 4, 2023

1 Poisson Distribution

Example 1 *Let's consider a scenario where there is a restaurant that has had 100, 120, 80, 75, and 110 customers in the past five days. To ensure that they have the right amount of ingredients for tomorrow, it's important to estimate the number of customers that they will have based on the information from the previous days. Although the natural approach would be to calculate the average number of customers (which in this case would be 97), it's worth noting that this method could lead to a shortage of food on three out of the first five days if it were implemented in practice.*

In order to examine the distribution of the number of customers coming to the restaurant, we need to make some assumptions. Let's assume that there are a total of n equally-sized time slots throughout the day, with each slot being small enough so that no more than one customer can enter the restaurant during a given slot. We'll also assume that the probability of a customer entering the restaurant during a slot is denoted by p, and that the occurrence of a customer entering during one slot is independent of any other slot.

Formally, let $X_i \triangleq \mathbf{1}[\text{there is a customer coming in the } i\text{-th slot}]$ for $i \in [n]$. Then we know $X_i \sim \mathrm{Ber}(p)$ and X_i's are mutually independent. Let $Z_n = \sum_{i=1}^n X_i$ and $\lambda = \mathbf{E}[Z_n] = pn$. Now let's compute the distribution of the number of customers Z_n. For any constant $k \in \mathbb{N}$,

$$\begin{aligned}\mathbf{Pr}[Z_n = k] &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{k!} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \frac{\lambda^k}{k!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k}. \end{aligned} \tag{1}$$

Note that λ and k are constants. Thus, when $n \to \infty$, Equation (1) equals to $\frac{\lambda^k}{k!} e^{-\lambda}$ and Z_n follows Poisson distribution with mean λ.

Definition 1 (Poisson Distribution) *We say a random variable X follows Poisson distribution with mean λ, written as $X \sim \mathrm{Pois}(\lambda)$, if for any $k \in \mathbb{Z}$,*

$$\mathbf{Pr}[X = k] = \begin{cases} \frac{\lambda^k}{k!} e^{-\lambda} & \text{if } k \geq 0, \\ 0 & \text{if } k < 0. \end{cases}$$

Since we get the distribution of Z_n by taking the limit, we need to verify that it is a distribution and it's mean is indeed λ:

- We have $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1$. Thus it is indeed a distribution.
- Since

$$\mathbf{E}\left[Z_n\right] = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{(k)!} e^{-\lambda} = \lambda,$$

the expectation of Z_n indeed equals to λ.

What is the distribution of the number of customers in two days? It follows from the following property of Poisson distributions.

Proposition 2 *Suppose* $X_1 \sim \mathrm{Pois}(\lambda_1)$ *and* $X_2 \sim \mathrm{Pois}(\lambda_2)$ *are two independent random variables. Then*

$$X_1 + X_2 \sim \mathrm{Pois}(\lambda_1 + \lambda_2).$$

Proof. For $n \geq 0$,

$$\begin{aligned}
\mathbf{Pr}\left[X_1 + X_2 = n\right] &= \sum_{m=0}^{n} \mathbf{Pr}\left[X_1 = m\right] \cdot \mathbf{Pr}\left[X_2 = n - m\right] \\
&= \sum_{m=0}^{n} \frac{\lambda_1^m}{m!} e^{-\lambda_1} \cdot \frac{\lambda_2^{n-m}}{(n-m)!} e^{-\lambda_2} \\
&= e^{-(\lambda_1+\lambda_2)} \cdot \sum_{m=0}^{n} \binom{n}{m} \frac{\lambda_1^m \lambda_2^{n-m}}{n!} \\
&= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1 + \lambda_2)^n}{n!}
\end{aligned}$$

□

It is easy to extend the proposition to more independent Poisson random variables.

Corollary 3 *Suppose that* $X_1, X_2, \ldots, X_n$ *are* n *mutually independent random variables where* $X_i \sim \mathrm{Pois}(\lambda_i)$. *Then*

$$\sum_{i=1}^{n} X_i \sim \mathrm{Pois}\left(\sum_{i=1}^{n} \lambda_i\right).$$

2 *Poisson Processes*

2.1 *Definition of a Poisson Process*

If we count the number of customers during a period of time rather than a single day, e.g., from day t_1 to day t_2, it should follow $\mathrm{Pois}\left((t_2 - t_2)\lambda\right)$. A Poisson process summarizes all the relevant information.

Definition 4 *A Poisson process* $\{N(s) : s \geq 0\}$ *with rate* λ *is a stochastic process that*

1. *$N(0) = 0$;*

2. *$\forall t, s \geq 0$, $N(t+s) - N(s) \sim \mathrm{Pois}(\lambda t)$;*

3. *$\forall t_0 \leq t_1 \leq \cdots \leq t_n$, $N(t_1) - N(t_0), N(t_2) - N(t_1), \cdots, N(t_n) - N(t_{n-1})$ are mutually independent.*

In fact, we can view the Poisson process in another way by considering the time gaps between arrivals. To see this, we first recall the exponential distribution.

Definition 5 *The probability density function of the exponential distribution with rate $\lambda > 0$ is*

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The corresponding cumulative distribution function is

$$F(t) = \int_{-\infty}^{t} f(x)\, dx = \int_{0}^{t} \lambda e^{-\lambda x}\, dx = 1 - e^{-\lambda t}.$$

Then the following proposition gives another characterization of the Poisson process.

In Proposition 6, we can regard τ_i as the time gap between the arrival of the $i-1$-th and the i-th customer. The parameter λ can be understood as the coming rate. Then the CDF of τ_i $F(t) = 1 - e^{-\lambda t}$ is the probability that the i-th customer comes within time t after the arrival of the $i-1$-th person.

Proposition 6 *Suppose that $\tau_1, \tau_2, \ldots, \tau_n$ is a sequence of independent random variables that $\tau_i \sim \mathrm{Exp}(\lambda)$. Let $T_n = \sum_{i=1}^{n} \tau_i$. For $s \geq 0$, let $N(s) = \max\{n \mid T_n \leq s\}$. Then $N(s)$ is a Poisson process with rate λ.*

Before proving this proposition, we discuss some properties of the exponential distribution.

2.2 Properties of Exponential Distribution

Proposition 7 *Let $X \sim \mathrm{Exp}(\lambda)$. Then $\mathbf{E}[X] = \frac{1}{\lambda}$.*

Since λ is the arriving rate, we can imagine that the average time between arrivals $\mathbf{E}[\tau_i]$ is the reciprocal of λ. This gives an intuition of Proposition 7.

Proof.

$$\begin{aligned}\mathbf{E}[X] &= \int_0^\infty t \cdot \lambda e^{-\lambda t} dt = \left(-te^{-\lambda t}\right)\Big|_0^\infty + \int_0^\infty e^{-\lambda t} dt \\ &= -\frac{1}{\lambda} e^{-\lambda t}\Big|_0^\infty = \frac{1}{\lambda}.\end{aligned}$$

□

Proposition 8 *Let $X \sim \mathrm{Exp}(\lambda)$. Then $\mathbf{Var}[X] = \frac{1}{\lambda^2}$.*

Proof. Note that

$$\mathbf{Var}[X] = \mathbf{E}\left[X^2\right] - \mathbf{E}[X]^2 = \mathbf{E}\left[X^2\right] - \frac{1}{\lambda^2}.$$

And

$$\mathbf{E}\left[X^2\right] = \int_0^\infty t^2 \cdot \lambda e^{-\lambda t} dt = \left(-t^2 e^{-\lambda t}\right)\Big|_0^\infty + \int_0^\infty 2te^{-\lambda t} dt^2$$
$$= 2\int_0^\infty t \cdot e^{-\lambda t} dt = \mathbf{E}\left[X\right] \cdot \frac{2}{\lambda} = \frac{2}{\lambda^2}.$$

Thus we have $\mathbf{Var}\left[X\right] = \frac{1}{\lambda^2}$. □

Proposition 9 (Lack of Memory) *Let $X \sim \text{Exp}(\lambda)$. Then for any $t, s > 0$,*

$$\mathbf{Pr}\left[X > t + s \mid X > s\right] = \mathbf{Pr}\left[X > t\right].$$

Proof.

$$\mathbf{Pr}\left[X > t + s \mid X > s\right] = \frac{\mathbf{Pr}\left[X > t + s \wedge X > s\right]}{\mathbf{Pr}\left[X > s\right]} = \frac{\mathbf{Pr}\left[X > t + s\right]}{\mathbf{Pr}\left[X > s\right]}$$
$$= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}.$$

□

Proposition 10 (Exponential Races) *Let $X_1 \sim \text{Exp}(\lambda_1)$ and $X_2 \sim \text{Exp}(\lambda_2)$ be two independent random variables. Then $Y \triangleq \min\{X_1, X_2\} \sim \text{Exp}(\lambda_1 + \lambda_2)$.*

Proof. By the independence, we have

$$\mathbf{Pr}\left[Y > t\right] = \mathbf{Pr}\left[X_1 > t \wedge X_2 > t\right] = \mathbf{Pr}\left[X_1 > t\right] \cdot \mathbf{Pr}\left[X_2 > t\right] = e^{-(\lambda_1+\lambda_2)t}.$$

□

Proposition 10 describes the distribution of the earliest customer of two restaurants. And we can easily generalize this to the case of more restaurants.

Corollary 11 *Let $X_1, X_2 \ldots, X_n$ be n mutually independent random variables where $X_i \sim \text{Exp}(\lambda_i)$. Then $Y \triangleq \min\{X_1, X_2, \ldots, X_n\}$ has an exponential distribution with rate $\sum_{i=1}^n \lambda_i$.*

Now we consider the problem “who wins the race?”. That is, the restaurants are racing to see who will first have a customer. We first assume that there are only two random variables. Let f_λ be the probability density function of exponential distribution with rate λ. Using the law of total probability, we can compute the probability that X_1 wins the race as follows:

$$\mathbf{Pr}\left[X_1 < X_2\right] = \int_0^\infty f_{\lambda_1}(t)\mathbf{Pr}\left[X_2 \ge t\right] dt$$
$$= \int_0^\infty \lambda_1 e^{-\lambda_1 t} e^{-\lambda_2 t} dt$$
$$= \lambda_1 \int_0^\infty e^{-(\lambda_1+\lambda_2)t} dt = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

Thus, clearly, the probability that X_i wins the race among n random variables is $\frac{\lambda_i}{\sum_{j=1}^n \lambda_j}$.

2.3 *Proof of Proposition 6*

Proposition 12 (Proposition 6 restated) *Suppose that $\tau_1, \tau_2, \ldots, \tau_n$ is a sequence of independent random variables that $\tau_i \sim \mathrm{Exp}(\lambda)$. Let $T_n = \sum_{i=1}^{n} \tau_i$. For $s \geq 0$, let $N(s) = \max\{n \mid T_n \leq s\}$. Then $N(s)$ is a Poisson process with rate λ.*

Proof. Note that $T_n = \sum_{i=1}^{n} \tau_i$ is the arrival time of the n-th customer. Let g_n be the probability density function of T_n. First we prove that the distribution of T_n follows the *Gamma distribution* $\Gamma(n, \lambda)$:

$$g_n(t) = \begin{cases} \lambda e^{-\lambda t} \cdot \frac{(\lambda t)^{n-1}}{(n-1)!} & t \geq 0, \\ 0 & t < 0. \end{cases}$$

We prove this by induction. Note that when $n = 1$, $T_1 = \tau_1 \sim \mathrm{Exp}(\lambda) = \Gamma(1, \lambda)$. Suppose that $T_n \sim \Gamma(n, \lambda)$ for some $n \geq 1$. By the independence of T_n and τ_{n+1}, for $t \geq 0$ we have

$$\begin{aligned} g_{n+1}(t) &= \int_0^t g_n(s) \cdot f_\lambda(t-s) ds \\ &= \int_0^t \lambda e^{-\lambda s} \cdot \frac{(\lambda s)^{n-1}}{(n-1)!} \cdot \lambda e^{-\lambda(t-s)} ds \\ &= \lambda e^{-\lambda t} \frac{\lambda^n}{(n-1)!} \int_0^t s^{n-1} ds \\ &= \lambda e^{-\lambda t} \frac{\lambda^n}{(n-1)!} \cdot \frac{t^n}{n} = \lambda e^{-\lambda t} \cdot \frac{(\lambda t)^n}{n!}. \end{aligned}$$

Then we compute the distribution of $N(t)$.

$$\begin{aligned} \mathbf{Pr}\left[N(t) = n\right] &= \mathbf{Pr}\left[T_n \geq t \wedge T_{n+1} > t\right] \\ &= \int_0^t g_n(s) \cdot \mathbf{Pr}\left[\tau_{n+1} > t - s\right] ds \\ &= \int_0^t \lambda e^{-\lambda s} \cdot \frac{(\lambda s)^{n-1}}{(n-1)!} \cdot e^{-\lambda(t-s)} ds \\ &= \lambda^n e^{-\lambda t} \frac{t^n}{n!}. \end{aligned}$$

Thus, $N(t) \sim \mathrm{Pois}(\lambda t)$. Then we verify that $\{N(t) : t \geq 0\}$ satisfies the three conditions in Definition 4.

First it is clear that $N(t) = 0$ when $t = 0$. By the lack of memory property, we know that $N(s+t) - N(s)$ follows the same distribution as $N(t) - N(0)$, which $\mathrm{Pois}(\lambda t)$. Furthermore, it is easy to see that $N(s+t) - N(s)$ is independent of $N(r)$ for all $r \leq s$ again by the lack of memory property. It implies that $N(t)$ has independent increments, and hence completes our proof of Proposition 6. □

Imagine the difference between $N(s+t) - N(s)$ and $N(t) - N(0)$. In the $N(t) - N(0)$, we start to wait for the first customer at time 0, while in $N(s+t) - N(s)$, at time s, we might have waited for the first customer in the period for sometime. However, due to the lack of memory property of the waiting time, this is equivalent to start to wait at time s.

2.4 *Thinning*

In the example of customers coming into the restaurant, sometimes we have a more detailed characterization of customers, such as the gender.

We associate an i.i.d. random variable Y_i with i-th arrival, and then use the value of Y_i to label the arrival and separate the Poisson process into several. Suppose that $Y_i \in \mathbb{N}$ and let $p_j = \mathbf{Pr}[Y_i = j]$. For all $j \in \text{Range}(Y_i)$, let $N_j(t)$ denote the number of arrivals with label j that have arrived by time t. Then $\{N_j(t)\}$ is called a thinning of a Poisson process. We have the following useful and surprising proposition.

Proposition 13 *For each j, $\{N_j(t) : t \geq 0\}$ is a Poisson process with rate $p_j\lambda$. Moreover, the collections of processes $\{\{N_j(t) : t \geq 0\} : j \in \mathsf{Range}(Y)\}$ are mutually independent.*

Here is an example explains why this proposition is surprising. Assume that the customers coming into a restaurant is a Poisson process, and each customer is male or female independently with probability 1/2 and 1/2 respectively. In fact we can assume that we flip a coin to determine whether the arriving customer is male or female. So intuitively, one might think that a large number of men (such as 50) arriving in one hour would indicates a large volume of business and hence a larger than normal number of women arriving. However this proposition tells us that the number of men arriving and the number of women arriving are independent.

Proof. For convenience we assume that $Y_i \in \{0, 1\}$. Then the following calculation concludes the independence and the distribution of $N_j(t)$ at the same time.

$$\begin{aligned}
\mathbf{Pr}[N_0(t) = j \wedge N_1(t) = k] &= \mathbf{Pr}[N_0(t) = j \wedge N(t) = k + j] \\
&= \mathbf{Pr}[N(t) = k + j] \cdot \mathbf{Pr}[N_0(t) = j \mid N(t) = k + j] \\
&= e^{-\lambda t}\frac{(\lambda t)^{j+k}}{(j+k)!} \cdot \binom{j+k}{j} p_0^j p_1^k \\
&= e^{-p_0\lambda t}\frac{(p_0\lambda t)^j}{j!} \cdot e^{-p_1\lambda t}\frac{(p_1\lambda t)^k}{k!}.
\end{aligned}$$

Thus, when there are n labels, it easy to verify that $N_j(t) \sim \text{Pois}(p_j\lambda)$ and they are mutually independent. □

Let's see an application of Poisson process.

Example 2 (Maximum Likelihood of Poisson Process) *Suppose there are two editors reading a book of* 300 *pages. Editor A finds* 100 *typos in the book, and editor B finds* 120 *typos,* 80 *of which are in common.*

Suppose that the author's typos follow a Poisson process with some unknown rate λ per page. The two editors catch typos with unknown probabilities of success p_A and p_B respectively. We want to know how many typos there actually are. We can estimate this by determining λ, p_A and p_B. Clearly, there are four types of typos:

Type 1 *The typo is found by neither of the editors. This happens w.p.* $q_1 = (1 - p_A)(1 - p_B)$.

Type 2 *The typo is found only by editor A. This happens w.p.* $q_2 = (1 - p_A)p_B$.

Type 3 *The typo is found only by editor B. This happens w.p.* $q_3 = (1 - p_B)p_A$.

Type 4 *The typo is found by both editors. This happens w.p.* $q_4 = p_A p_B$.

So the occurrence of type i typos follows an independent Poisson process with rate $q_i\lambda$. That is, letting X_1, X_2, X_3 and X_4 be the occurrence time of the corresponding type of typos in this book, then $X_i \sim \text{Pois}(300q_i\lambda)$. Note that *there are* 20 *typos of type* 2, 40 *typos of type* 3 *and* 80 *typos of type* 4. *We claim that the most likely values of the rates are*

$$\begin{cases} 300(1-p_A)p_B\lambda = 20, \\ 300(1-p_B)p_A\lambda = 40, \\ 300p_Ap_B\lambda = 80. \end{cases}$$

This yields that $p_A = \frac{2}{3}$, $p_B = \frac{4}{5}$ *and* $\lambda = \frac{1}{2}$.

It remains to prove that claim. Suppose $X \sim \mathrm{Pois}(\theta)$ *with some unknown* θ. *Then given* z, *our goal is to find* $\arg\max_\theta \mathbf{Pr}\left[X = z \mid X \sim \mathrm{Pois}(\theta)\right]$. *Note that* $\mathbf{Pr}\left[X = z \mid X \sim \mathrm{Pois}(\theta)\right] = e^{-\theta}\frac{\theta^z}{z!}$ *and* $\log e^{-\theta}\frac{\theta^z}{z!} = -\theta + z \cdot \log\theta$. *So it is equivalent to find*

$$\arg\max_\theta -\theta + z \cdot \log\theta \tag{2}$$

Let the derivation of Equation (2) *equals to* 0. *We have* $\theta = z$, *that is,* $\arg\max_\theta e^{-\theta}\frac{\theta^z}{z!} = z$.