Why Poisson? Why Exponential?

Study Notes by Liutong Zhou

November 12, 2016

Many experiments consist of observing the occurrence times of random arrivals. Examples include arrivals of customers for service, arrivals of calls at a switch-board, occurrences of floods and other natural and man-made disasters in a given period, and so forth. Though the family of Poisson distributions has been commonly used to model the number of such rare events that occur in a fixed time period, few people understand why it is chosen. I prepare this note to explain the reason behind.

1 Poisson Distribution

Definition 1 (Poisson Distribution). *A random variable X has the Poisson Distribution with mean λ if the p.d.f. of X is*

$$f(x|\lambda) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & for\ x \in N \\ 0 & otherwise \end{cases} \tag{1}$$

Example 1 (Customer Arrivals). A store owner believes that customers arrive at his store at a rate of 4.5 customers per hour. He wants to find out the distribution of X, the number of customers who will arrive during a particular one-hour period later in the day.

He models customer arrivals in different time periods as *independent* of each other. As a first approximation, he divides the one-hour period into 3600 seconds and thinks of the arrival rate as being $4.5/3600 = 0.00125$ per second. He then says that during each second, either 0 or 1 customers will arrive, and the probability of an arrival during any single second is 0.00125. He then tries to use the *Binomial Distribution* with parameters n = 3600 and p = 0.00125 to describe X.

He starts calculating f , the p.m.f. of this binomial distribution, and quickly discovers how cumbersome the calculations are. However, he realizes that the successive values of $f(x)$ are closely related to each other because $f(x)$ changes in a systematic way as x increases. So he computes

$$\frac{f(x+1)}{f(x)} = \frac{\binom{n}{x+1}p^{x+1}(1-p)^{n-x-1}}{\binom{n}{x}p^x(1-p)^{n-x}} = \frac{(n-x)p}{(x+1)(1-p)} \approx \frac{np}{x+1},$$

The last approximation follows the fact that n=3600 is large while p=0.00125 is small. When x is mall (i.e. assume for now we are only interested in x=0,...,30), $(n-x) \to n$ and $(1-p) \to 1$.
This approximation suggests defining $\lambda = np$ and approximating $f(x+1) \approx f(x)\lambda/(x+1)$ for all the values of x that matter.

$$f(1) = f(0)\lambda,$$
$$f(2) = f(1)\frac{\lambda}{2} = f(0)\frac{\lambda^2}{2},$$
$$f(3) = f(2)\frac{\lambda}{3} = f(0)\frac{\lambda^3}{6},$$
$$\vdots$$

Continue the above pattern yields $f(x) = f(0)\frac{\lambda^x}{x!}$. But to obtain a valid p.m.f for X, he needs to make sure that

$$\sum_{x=0}^{\infty} f(x) = 1 \Rightarrow \sum_{x=0}^{\infty} f(0)\frac{\lambda^x}{x!} = 1 \Rightarrow f(0) = \frac{1}{\sum_{x=0}^{x=\infty} \lambda^x/x!} = e^{-\lambda} \Rightarrow f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

This is exactly p.m.f of Poisson Distribution 1

2 From Binomial to Poisson

Theorem 1 (Poisson Approximation of Binomial). *For $\forall n \in N$ and each $0 < p < 1$, let $f(x|n,p)$ denote the p.m.f. of the binomial distribution $Binomial(n,p)$. Let $f(x|\lambda)$ denote the p.m.f. of the Poisson distribution with mean λ. Let $\{p_n\}_{n=1}^{\infty}$ be a sequence of numbers between 0 and 1 such that $\lim_{n\to\infty} np_n = \lambda$. Then*

$$\lim_{n\to\infty} f(x|n,p) = f(x|\lambda)$$

which says when n goes to infinity, binomial distribution has no difference from Poisson distribution.

Proof.

$$
\begin{aligned}
X \sim Binomial(n,p) \Rightarrow \lim_{n\to\infty} P(X=k) &= \lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} \\
&= \lim_{n\to\infty} \frac{n(n-1)...(n-(k-1))}{k!} \frac{\lambda^k}{n^k} \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k} \\
&= \lim_{n\to\infty} \frac{n(n-1)...(n-(k-1))}{n^k} \frac{\lambda^k}{k!} \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k} \\
&= 1 \cdot \frac{\lambda^k}{k!} \frac{e^{-\lambda}}{1}
\end{aligned}
$$

$\square$

3 Why Poisson

In example 1, the division of the one-hour period into 3600 seconds was somewhat arbitrary. The owner could have divided the hour into 7200 half-seconds or 14400 quarter-seconds, etc.

Perhaps the store owner would do better simply modeling the number X of arrivals as a *Poisson* random variable with mean 4.5, rather than choosing an arbitrarily sized time interval to accommodate a tedious binomial calculation.

What if the owner is interested in a half-hour period or a 4-hour and 15-minute period? Is it safe to assume that the number of customers that arrive in a half-hour period has the Poisson distribution with mean 2.25?

In order to be sure that all of the distributions for the various numbers of arrivals in Example 1 are consistent with each other, the store owner needs to think about **the overall process of customer arrivals, not just a few isolated time periods**.

To answer all the questions and explain where Poisson distribution comes from (or why Poisson), we need to introduce Poisson Process.

3.1 Poisson Process

Definition 2 (Poisson Process). *A Poisson process with rate λ ($\lambda > 0$) per unit time is a process that satisfies the following two properties:*

- *The number of arrivals in every fixed interval of time of length t has the Poisson distribution with mean λt.*

- *The numbers of arrivals in every collection of* disjoint *time intervals are independent.*

Now under definition 2, we are sure that it is safe to assume that the number of customers that arrive in a half-hour period has the Poisson distribution with mean 2.25.

Generality of Poisson Processes Not only in terms of counts of arrivals during time intervals, Poisson processes are actually more general. For example, a Poisson process can be used to model occurrences in space as well as time. A Poisson process could be used to model telephone calls arriving at a switchboard, atomic particles emitted from a radioactive source, diseased trees in a forest, or defects on the surface of a manufactured product. The reason for the popularity of the Poisson process model is twofold.

- First, the model is computationally convenient.

- Second, there is a mathematical justification for the model if one makes **three plausible assumptions** about how the phenomena occur.

3.2 Assumptions Underlying the Poisson Process Model

Indeed, a Poisson process can be used to model occurrences in **any region that can be subdivided into arbitrarily small pieces**. There are three assumptions that lead to the Poisson process model.

1. The numbers of occurrences in any collection of disjoint intervals of time must be mutually independent.

2. The probability of an occurrence during each very short interval of time must be approximately proportional to the length of that interval.

3. For each very short interval of time, the probability that there will be two or more occurrences in that interval must have a smaller order of magnitude than the probability that there will be just one occurrence.

To express the assumptions in mathematical language, let $o(t)$ denote any function of t having the property that

$$\lim_{t \to 0} \frac{o(t)}{t} = 0 \tag{2}$$

o(t) is a function that approaches 0 as $t \to 0$, and, furthermore, this function must approach 0 at a rate faster than t itself. An example of such a function is $o(t) = t\alpha$, where $\alpha > 1$.

Having 2, the second and third assumptions can then be expressed as: $\exists \lambda > 0$ s.t. for every time interval of length t, the probability of at least one occurrence during that interval has the form

$$\lambda t + o(t) \tag{3}$$

Thus, for every very small value of t, the probability of at least one occurrence during an interval of length t is equal to λt plus a quantity having a **smaller order of magnitude**.

One not so good consequence of the second assumption is that the process being observed must be stationary over the entire period of observation, that is there can be neither busy intervals nor quiet intervals. But Assumption 2 can be relaxed at the cost of more complicated mathematics.

In symbols,

$$P("there\ are\ two\ or\ more\ occurrences\ in\ a\ time\ interval\ of\ length\ t") = o(t) \tag{4}$$

Thus, **the probability of two or more occurrences in a small interval must be negligible in comparison with the probability of one occurrence in that interval**.

Under the preceding three assumptions, it can be shown that the process will satisfy the definition of a Poisson process with rate λ.

3.3 Why Poisson Distribution

We will prove that the three assumptions underlying the Poisson process model indeed implies that the number of occurrences has Poisson distribution.

What we need to show is that, for each t, the number of occurrences during a time interval of length t has the Poisson distribution with mean λt. Let X stand for the number of occurrences during a particular time interval of length t. We will use the following fact: $\forall a \in R$,

$$\lim_{u \to 0} (1 + au + o(u))^{1/u} = e^a$$

Follow the following steps to prove.

Proof.   1. For each $n \in N^+$, divide the time interval into n disjoint sub-interval of length t/n each. For $i = 1, ..., n$, let $Y_i = 1$ if exactly one arrival occurs in the ith sub-interval and 0 otherwise. Let A_i be the event that there are two or more occurrences during the ith sub-interval. Let $W_n = \sum_{i=1}^{n} Y_i$, $A = \cup_{i=1}^{n} A_i$. Then the event $\{X = k\}$ has the following split[1]:

$$\{X = k\} = (\{X = k\} \cap A) \cup (\{X = k\} \cap A^c) = E_1 \cup E_2 \tag{5}$$

Where $E_1 = (\{X = k\} \cap A)$, $E_2 = (\{X = k\} \cap A^c)$. But $A^c = \cap_{i=1}^{n} A_i^c$, so E_2 is just the event that "there are at most one occurrence in each sub-interval and a total of k occurrences in the whole interval", which is $E_2 = \{W_n = k\}$. And $E_1 \subset A$, $E_2 \subset A^c$. Since E_1 and E_2 are disjoint, we have

$$P(X = k) = P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$
$$= P(E_1) + P(E_2) = P(E_1) + P(W_n = k) \tag{6}$$

   2. Since all sub-intervals are disjoint, the events $A_1, ... A_n$ are mutually independent (follows assumption 1). Since each sub interval has equal length t/n, all the events $A_1, ... A_n$ have same probability to occur (follows assumption 2). So

$$P(A) = 1 - P(A^c) = 1 - P(\cap_{i=1}^{n} A_i^c) = 1 - (P(A_1^c))^n = 1 - (1 - P(A_1))^n$$

where $P(A_1) = o(\frac{\lambda t}{n}) = \lambda\, o(\frac{t}{n})$, which follows equation 4 and $\lim_{n \to \infty} \frac{o(\frac{t}{n})}{\frac{t}{n}} = 0$, which follows equation 2

$$\therefore \lim_{n \to \infty} P(A) = \lim_{n \to \infty} 1 - (1 - \lambda\, o(\frac{t}{n}))^n = 1 - \lim_{n \to \infty} e^{-\lambda t \frac{o(t/n)}{t/n}} = 1 - e^0 = 0$$

$$\therefore E_1 \subset A \Rightarrow P(E_1) <= P(A) \Rightarrow \lim_{n \to \infty} P(E_1) <= \lim_{n \to \infty} P(A) = 0 \Rightarrow \lim_{n \to \infty} P(E_1) = 0$$

$$\therefore Equation\ 6 \Rightarrow \lim_{n \to \infty} P(X = k) = \lim_{n \to \infty} P(W_n = k)$$

---
[1] De Morgan's Law. Note that Sample Space $\Omega = A \cup A^c$.

3. Since $Y_i \overset{i.i.d.}{\sim} Bernoulli(p_n)$, where by equation 3 $p_n = \lambda \frac{t}{n}$. $W_n = \sum_{i=1}^{n} Y_i \sim Binomial(n, p_n)$. But remember, in Theorem 1, we have proved that for binomial distribution

$$\lim_{n \to \infty} P(W_n = k) = \frac{(\lim_{n \to \infty} np_n)^k}{k!} e^{-\lim_{n \to \infty} np_n} = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \tag{7}$$

4. Combine The results from step 1~3, we have

$$\lim_{n \to \infty} P(X = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \tag{8}$$

$\square$

4 From Poisson to Exponential

Having the above result, now we can think of how exponential comes to the stage. Consider an event E="there are X number of occurrences of a rare event during a given time period of length t". Let T denote the waiting time (sometimes called return period) of this rare event. Then $P(T \leq t) = 1 - P(T > t)$, but the event that "waiting time is larger than t" ($\{T > t\}$) is just equivalent to the event "there are zero occurrence of event E during this given time period of length t" ($\{X = 0\}$). So take k=0 in equation 8, we get.

$$P(T > t) = P(X = 0) = e^{-\lambda t} \Rightarrow F_T(t) = P(T \leq t) = 1 - e^{-\lambda t} \tag{9}$$

I think you have noticed the magic here. What we get in equation 9 is just the **CDF of Exponential Distribution**. Simply take derivative w.r.t. t, you will get PDF of Exponential Distribution.

$$f_T(t) = \lambda e^{-\lambda t} \tag{10}$$