June 26, 2021

Reading Notes: Poisson Distribution

Written by Larry Cui

1 Definition

In the binomial distribution where n is quite large, it's usually a tedious job to calculate $k!$ when computer was not available back in the 18th to early 20th century. So a French mathematician Simeon Denis Poisson came up with a approximation, which is using continuous function to approximate the discrete binomial distribution, and proves to be quite well with a small p.

Poisson Limit:
Suppose X is a binomial random variable, where

$$P_x(k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, 2 \ldots, n$$

If $n \to \infty$ and $p \to 0$, then (let $\lambda = np$)

$$P_x(k) = \frac{\mathsf{e}^{-\lambda} \lambda^k}{k!}$$

2 How to get the Poisson Limit (proof)?

Use $\dfrac{\lambda}{n} = p$ to rewrite the binomial equation, and when $n \to \infty$:

$$\begin{aligned}
P_x(k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n!}{k!(n-k)!} \cdot \frac{\lambda^k}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \cdot \left(1 - \frac{\lambda}{n}\right)^n \\
&= \frac{n!}{k!(n-k)!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{n^k}{(n-\lambda)^k} \cdot \left(1 - \frac{\lambda}{n}\right)^n \\
&= \frac{\lambda^k}{k!} \cdot \frac{n!}{(n-k)!(n-\lambda)^k} \cdot \left(1 - \frac{\lambda}{n}\right)^n
\end{aligned}$$

We know that $\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x = \mathsf{e}$. If we let $-\frac{\lambda}{n} = \frac{1}{x}$, then $n = -\lambda x$, and we get the equation:

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = \lim_{x\to\infty}\left(1+\frac{1}{x}\right)^{-\lambda x} = \mathsf{e}^{-\lambda}$$

Furthermore,

$$\frac{n!}{(n-k)!(n-\lambda)^k} = \frac{n(n-1)\cdots(n-k+1)}{(n-\lambda)^k}$$

Because λ is constant (this is the approximation part of Poisson Limit!), as $n \to \infty$, the above equation tends to be 1 (proved).

Notice: Poisson approximation is turning a discrete distribution to a continuous one.

3 $E(X)$ and $Var(X)$

Lemma 1: E(X) of a binomial distribution

$$E(X) = np = \lambda$$

Proof:

$$E(X) = k \cdot P_x(k) = k \cdot \frac{n!}{k!(n-k)!} \cdot p^k(1-p)^{n-k} \tag{1}$$

$$= np \cdot \frac{(n-1)!}{(k-1)!(n-k)!} \cdot p^{k-1}(1-p)^{n-k} \tag{2}$$

Because the factorial of a negative integer is not defined, so what's the result of $(k-1)!$? Fortunately, if we look at eq.(1) closely, we can see that when $k = 0$, the first term becomes zero and we can proceed with $k = 1$ and other larger integers directly.

Let $j = k-1, m = n-1$, then eq.(2) becomes

$$\begin{aligned} E(X) &= np \cdot \frac{m!}{j!(m-j)!} \cdot p^j(1-p)^{m-j} \\ &= np \cdot \binom{m}{j} p^j(1-p)^{m-j} \\ &= np \end{aligned}$$

Lemma 2: Var(X) of a binomial distribution

$$Var(X) = np(1-p) = \lambda(1-p)$$

Proof:

The whole point here is to find $E(X^2)$, then use the equation $Var(X) = E(X^2) - \mu^2$ (μ is another name for $E(X)$) to find $Var(X)$.

$$\begin{aligned}
E(X^2) &= k^2 \cdot \frac{n!}{k!(n-k)!} \cdot p^k(1-p)^{n-k} \\
&= np \cdot k \cdot \frac{(n-1)!}{(k-1)!(n-k)!} \cdot p^{k-1}(1-p)^{n-k} \\
&= np \cdot (1 + (k-1)) \cdot \frac{(n-1)!}{(k-1)!(n-k)!} \cdot p^{k-1}(1-p)^{n-k} \\
&= eq.(2) + np \cdot (k-1) \cdot \frac{(n-1)!}{(k-1)!(n-k)!} \cdot p^{k-1}(1-p)^{n-k} \\
&= eq.(2) + n(n-1)p^2 \cdot \frac{(n-2)!}{(k-2)!(n-k)!} \cdot p^{k-2}(1-p)^{n-k} \\
&= np + (np)^2 - np^2
\end{aligned}$$

$$\begin{aligned}
Var(X) &= np + (np)^2 - np^2 - (np)^2 \\
&= np(1-p)
\end{aligned}$$

Now let's see how close are the expected value and variance of a Poisson distribution be to the binomial distribution:

Theorem: E(X) and Var(X) of Poisson Distribution

For a Poisson Distribution $P_x(k) = \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}, k = 0, 1, 2, \cdots$

$$E(X) = \lambda, Var(X) = \lambda$$

Proof:

$$
\begin{aligned}
E(X) &= k \cdot P_x(k) \\
&= k \cdot \sum_{k=0}^{\infty} \frac{\mathsf{e}^{-\lambda}\lambda^k}{k!} \\
&= \lambda \mathsf{e}^{-\lambda} \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda \mathsf{e}^{-\lambda} \mathsf{e}^{\lambda} \quad \text{(Taylor Series)} \\
&= \lambda
\end{aligned}
$$

$$
\begin{aligned}
Var(X) &= E(X^2) - \mu^2 \\
&= k^2 \cdot \sum_{k=0}^{\infty} \frac{\mathsf{e}^{-\lambda}\lambda^k}{k!} - \lambda^2 \\
&= \lambda \mathsf{e}^{-\lambda} \cdot \sum_{k=1}^{\infty} \frac{k\lambda^{k-1}}{(k-1)!} - \lambda^2 \\
&= \lambda \mathsf{e}^{-\lambda} \cdot \sum_{k=1}^{\infty} \frac{((k-1)+1)\lambda^{k-1}}{(k-1)!} - \lambda^2 \\
&= \lambda \mathsf{e}^{-\lambda} \cdot \sum_{k=1}^{\infty} \left(\frac{(k-1)\lambda^{k-1}}{(k-1)!} + \frac{\lambda^{k-1}}{(k-1)!} \right) - \lambda^2 \\
&= \lambda \mathsf{e}^{-\lambda} \cdot \left(\lambda \cdot \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right) - \lambda^2 \\
&= \lambda \mathsf{e}^{-\lambda} (\lambda \mathsf{e}^{\lambda} + \mathsf{e}^{\lambda}) - \lambda^2 \\
&= \lambda^2 + \lambda - \lambda^2 \\
&= \lambda
\end{aligned}
$$

Another proof way is by *mgf*:

First of all, *mgf* of the Poisson distribution

$$
\begin{aligned}
M_k(t) = E(\mathsf{e}^{tk}) &= \mathsf{e}^{tk} \cdot \sum_{k=0}^{\infty} \frac{\mathsf{e}^{-\lambda}\lambda^k}{k!} \\
&= \mathsf{e}^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\mathsf{e}^{tk}\lambda^k}{k!} \\
&= \exp[-\lambda + \lambda \mathsf{e}^t]
\end{aligned}
$$

(1) First moment of *mgf*: $M_k^{(1)}(t) = \exp[-\lambda + \lambda \mathsf{e}^t] \cdot \lambda \mathsf{e}^t$;

(2) Second moment of *mgf*: $M_k^{(1)}(t) = \exp[-\lambda + \lambda \mathsf{e}^t] \cdot (\lambda^2 \mathsf{e}^{2t} + \lambda \mathsf{e}^t)$.

Let $t = 0$, we have $E(X) = \lambda$, $E(X^2) = (\lambda^2 + \lambda)$ and $Var(X) = \lambda$.

From above analysis, we can tell that the E(X) is same for both binomial and Poisson distributions, with the only difference at its variance by np^2, which also means that the smaller the p, the more precise the Poisson approximation.

4 Waiting period probability

An interesting application of Poisson distribution is to find the waiting period probability, in another word, to determine how long before the next occurrence of events takes place.

Binomial distribution has a unique feature that, once we know the probability p, then the expected value is simply p times trial number n. On a time axis, we can set the probability within a unit time as λ, the elapsing of time is like traversing multiple time units y, and the total E(X) is just λy. Staying with this notation, the Poisson probability for studying waiting time becomes:

$$P_x(k) = \sum_{k=0}^{\infty} \frac{\mathsf{e}^{-\lambda y}(\lambda y)^k}{k!}$$

Caution: No $P_x(0)$ directly!

It's quite straight-forward to find the probability that during time y, there's no event happening:

$$P_x(0) = \frac{\mathsf{e}^{-\lambda y}(\lambda y)^0}{0!} = \mathsf{e}^{-\lambda y}$$

Wait a minute, what does this mean? Without referring to further reasoning, we know that at time 0, $P_x(0) = 1$, but as y increases (as time goes by), this function will yield less and less value until to 0 in the end. In another word, this function is decreasing, and a decreasing function is not a good *cdf*, if it is still regarded as a *cdf*, very difficult to deal with when we want some pdf with respect to time variable y.

Tips: Happen is Happen, no matter one or more!

Now let's take a look at "$1 - P_x(0)$". It stands for the probability that within time y, some events happen. It's not $P_x(1)$, not $P_x(2)$ or any other numbers that x may take, but the collection of all probabilities for $x = 1, 2, \ldots, \infty$.

Also, you don't need further calculation to find out that at time 0, no event would happen, so the probability equals 0, but as y elapses, the probability grows and eventually reaches 1, which means that something about to happen will happen in the end.

So the *cdf*, which means the probability of something happen during the period of y, is

$$F(Y > y) = 1 - P_x(0) = 1 - \mathsf{e}^{-\lambda y}$$

and, *pdf* with respect to the y that something happen at that moment is

$$\frac{d}{dy}(1 - \mathsf{e}^{-\lambda y}) = \lambda \mathsf{e}^{-\lambda y}$$