Variable aleatoria Poisson

La v.a. Poisson tiene muchas aplicaciones. Al igual que la binomial, la v.a. Poisson modela conteos. Vamos a deducir la distribución de la v.a. Poisson como límite de la distribución de la v.a. binomial

Sabemos que $X \sim Bin(n,p)$

$\Rightarrow$ X cuenta el número de éxitos en $\underline{n}$ experimentos Bernoulli(p) independientes, pero que pasa si $\underline{n}$ es muy grande? en particular si $n \to \infty$

Si $X \sim Bin(n,p)$

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Si hacemos $\lambda = np \Rightarrow p = \frac{\lambda}{n}$

$$f(x) = \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{n!}{(n-x)!\,x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{n(n-1)\cdots(n-x+1)\cancel{(n-x)!}}{\cancel{(n-x)!}\,x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

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

$$n(n-1)\cdots(n-x+1) = n\left[n\left(1-\tfrac{1}{n}\right) n\left(1-\tfrac{2}{n}\right)\cdots n\left(1-\tfrac{x-1}{n}\right)\right]$$

$$= n^x\left[\left(1-\tfrac{1}{n}\right)\left(1-\tfrac{2}{n}\right)\cdots\left(1-\tfrac{x-1}{n}\right)\right]$$

$$= \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)}{x!}\,\lambda^x\,\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^x}$$

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

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

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

Diremos que $X \sim \text{Poisson}(\lambda)$ si

$$f(x) = \frac{\lambda^x}{x!}e^{-\lambda}, \quad x = 0, 1, 2, 3, \ldots$$

¿Como interpretamos esta v.a.?

$X \sim \text{Bin}(n, p)$ y tomando

$\lambda = np \Rightarrow p = \lambda/n$

y $n \to \infty$

$\Rightarrow X \sim \text{Poisson}(\lambda)$

X cuenta el número de éxitos en $\underline{n}$ experimentos Bernoulli(p), pero a medida que $\underline{n}$ crece

$\rightarrow$ p se vuelve cada vez más pequeña pero $np = \lambda$, el numero esperado de éxitos, es constante!

$\Rightarrow$ $X \sim \text{Poisson}(\lambda)$ cuenta el numero de éxitos en $\underline{n}$ experimentos Bernoulli(p) independientes cuando la probabilidad de éxito es muy pequeña pero hacemos una cantidad de experimentos $\underline{n}$ muy grande.

- Por eso se dice que cuenta eventos probabilidades pequeñas (eventos raros) pero que ocurren con una frecuencia media constante $\lambda = np$.

Diremos que $X \sim \text{Poisson}(\lambda)$ si

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, 2, 3, \ldots$$

¿Es una función de distribución?

$$\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \underbrace{\sum_{x=0}^{\infty} \frac{\lambda^x}{x!}}_{e^{\lambda}} = e^{-\lambda} e^{\lambda} = 1$$

Función generadora de momentos

$$M_X(t) = \mathbb{E}(e^{Xt})$$

$$e^{xt} = \sum_{k=0}^{\infty} \frac{(xt)^k}{k!} = 1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \dots$$

$$\mathbb{E}(e^{Xt}) = 1 + t\,\mathbb{E}(X) + \frac{t^2}{2!}\mathbb{E}(X^2) + \frac{t^3}{3!}\mathbb{E}(X^3) + \dots$$

$$\frac{d}{dt} M_X(t) = \frac{d}{dt} \mathbb{E}(e^{Xt}) \Big|_{t=0}$$

$$= \mathbb{E}(X) + \frac{2t}{2!}\mathbb{E}(X^2) + \frac{3t^2}{3!}\mathbb{E}(X^3) + \dots \Big|_{t=0}$$

$$= \mathbb{E}(X)$$

de esta podemos ver que

$$\mathbb{E}(X^k) = \frac{d^k}{dt^k} M_X(t) \Big|_{t=0}$$

$$M_X(t) = \mathbb{E}(e^{Xt})$$

$$= \sum_{x=0}^{\infty} e^{xt} \frac{\lambda^x}{x!} e^{-\lambda}$$

$$= \sum_{x=0}^{\infty} \frac{(e^t \lambda)^x}{x!} e^{-\lambda}$$

$$= e^{-\lambda} e^{e^t \lambda} \cancelto{1}{\sum_{x=0}^{\infty} \frac{(e^t \lambda)^x}{x!} e^{-e^t \lambda}}$$

$$= e^{\lambda(e^t - 1)}$$

$$M_X(t) = e^{\lambda(e^t - 1)}$$

$$\mathbb{E}(X) = \frac{d}{dt} M_X(t) \Big|_{t=0}$$

$$= \lambda e^t e^{\lambda(e^t - 1)} \Big|_{t=0}$$

$$= \lambda$$

$$\mathbb{E}(X^2) = \frac{d^2}{dt^2} M_X(t) \Big|_{t=0}$$

$$= \lambda e^t e^{\lambda(e^t-1)} + (\lambda e^t)^2 e^{\lambda(e^t-1)} \Big|_{t=0}$$

$$= \lambda + \lambda^2$$

$$\Rightarrow Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \lambda$$

Una característica interesante de la distribución Poisson es que si

$$X_i \sim Poisson(\lambda_i), \quad i = 1, 2, \ldots, n$$

$$y \quad X_i \perp X_j \quad \forall\, i \neq j$$

$$\Rightarrow Y = \sum_{i=1}^{n} X_i \sim Poisson\left(\sum_{i=1}^{n} \lambda_i\right)$$

Dem

$$M_Y(t) = \mathbb{E}(e^{Yt}) = \mathbb{E}\left(e^{\sum x_i t}\right)$$

$$= \mathbb{E}(e^{X_1 t} e^{X_2 t} e^{X_3 t} \cdots e^{X_n t})$$

$$= \prod_{i=1}^{n} \mathbb{E}(e^{X_i t})$$

$$= \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} e^{\lambda_i(e^t-1)}$$

$$= e^{\sum \lambda_i (e^t - 1)} \qquad \square$$

La distribución Poisson se utiliza para modelar el numero de eventos raros que ocurren en intervalos de tiempo o espacios fijos, e.g.

- Número de clientes que llegan a una tienda, llamadas a un centro de servicio en un intervalo de tiempo determinado

- Número de accidentes de tráfico en un periodo de tiempo especifico

- Número de mutaciones en una secuencia de ADN en cierto periodo de tiempo

- El numero de veces que ocurre una cierta palabra en documentos largos.

Ejemplo:

La probabilidad de encontrar 1 imperfección en 1 metro de cable de fibra óptica es $1/1000$. Y la probabilidad de encontrar más de una imperfección es ≈ 0. Determine la probabilidad de encontrar 5 imperfecciones en un rollo de 3000 Km

$$\left.\begin{array}{l} n = 3{,}000 \\ p = 1/1000 \end{array}\right\} \Rightarrow \lambda = np = \frac{3{,}000}{1{,}000} = 3$$

$$f(5) = \frac{e^{-3}\, 3^{5}}{5!} = 0.101$$

El problema de usar la distribución Poisson en aplicaciones es que

$$\lambda = \mathbb{E}(X) = \operatorname{Var}(X)$$

Analizando datos reales, se tiene que en muchos casos

$\bar{x} < s^2$, en estos casos obviamente el supuesto que implica la Poisson no es válido. Una posible solución es usar la <u>Binomial Negativa</u> o generalizaciones de la Poisson

Una parametrización alternativa de la distribución Poisson es la siguiente

$$f(x) = \frac{\lambda^{x} e^{-\lambda}}{x!} = \frac{(rt)^{x} e^{-rt}}{x!}$$

en este caso se tienen los parámetros $\underline{\underline{rt}} = \lambda$

en donde rt es el número promedio de éxitos en un intervalo de longitud t. Y r es el numero promedio de exitos en un intervalo de longitud 1.