Distribuciones notables II

Distribución Poisson

- Diremos que una v.a. discreta X_t con $X(\Omega) = \mathbb{N}$ tiene distribución de Poisson con parámetro $\lambda > 0$, y lo denotaremos por $Po(\lambda)$ si su función de probabilidad es:

$$P_{X_t}(x) = P(X_t = x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & \text{si } x = 0, 1, \ldots \\ 0 & \text{en otro caso} \end{cases}.$$

- Como el desarrollo en serie Taylor de la exponencial es

$$e^{\lambda} = \sum_{x=0}^{+\infty} \frac{\lambda^x}{x!}.$$

- Es fácil comprobar que todos los valores de la función de probabilidad suman 1 (ejercicio).

La distribución Poisson como "límite" de una binomial.

- La distribución Poisson aparece en el conteo de determinados eventos que se producen en un intervalo de tiempo o en el espacio.
- Supongamos que nuestra variable de interés es

$$X = \text{número de eventos en el intervalo de tiempo}(0, t].$$

- Por ejemplo el número de llamadas a un *call center* del que sabemos que se cumplen las condiciones siguientes:

Condiciones para la distribución Poisson.

a) El número promedio de eventos en el intervalo $(0, t]$ es $\lambda > 0$.
b) Es posible dividir el intervalo de tiempo en un gran número de subintervalos (denotemos por n al número de intervalos) de forma que:
   1) La probabilidad de que se produzcan dos o más eventos en un subintervalo es despreciable.
   2) El número de ocurrencias de eventos en un intervalo es independiente del número de ocurrencias en otro intervalo.
   3) La probabilidad de que un evento ocurra en un subintervalo es $p = \frac{\lambda}{n}$.

La distribución Poisson como límite de una distribución binomial (OPCIONAL)

- Bajo estas condiciones podemos considerar que el número de eventos en el intervalo $(0, t]$ será el número de "éxitos" en n repeticiones independientes de un proceso Bernoulli de parámetro p
- Entonces si $n \to \infty$ y $p \cdot n$ se mantiene igual a λ resulta que la función de probabilidad de X se puede poner como

$$f_X(k) = \lim_{n\to\infty} \left(\begin{array}{c} n \\ k \end{array} \right) p^k q^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

Aproximación de la distribución binomial por la Poisson:

Bajo el punto de vista anterior y si p es pequeño y n suficientemente grande (existen distintos criterios por ejemplo $n > 20$ ó 30 y $p \geqslant 0.1$, $1 - p \geqslant 0.1$) podemos aproximar una $B(n, p)$ por una $Po(n \cdot p)$

Los procesos Poisson.

Propiedades

Si tenemos un experimento Poisson con λ igual al promedio de eventos en una unidad de tiempo (u.t.) entonces si

t es una cantidad de tiempo en u.t., la v.a

X_t = numero de eventos en el intervalo$(0, t]$es una $Po(\lambda \cdot t)$.

A la familia de variables X_t se la denomina proceso de Poisson.

Resumen v.a con distribución Poisson $Po(\lambda)$

X Poisson λ.
$D_X = \{0, 1, \ldots n\}$
$P_X(x) = P(X = x) = \begin{cases} \frac{\lambda^x}{x!} \exp -\lambda & \text{si } x = 0, 1, \ldots, n \\ 0 & \text{en otro caso.} \end{cases}$
$F_X(x) = P(X \leqslant X) =$ Función de R o tabulada
$E(X) = \lambda$ $Var(X) = \lambda$

Ejemplos procesos de Poisson

Podrían, aunque no siempre, seguir una distribución Poisson las siguientes situaciones

- Número de conexiones a un *web service*.
- Numero de erratas por página en un libro.
- Número de clientes en la cola de un cajero de un supermercado.
- Número de insectos capturados en una trampa en un cierto intervalo de tiempo.
- Número de errores en la clasificación de miles de fotografías de una red neuronal....

Ejercicio

$P(X=x)=\frac{\lambda^x}{x!}\cdot e^{-\lambda}$
Supongamos que disponemos de un servicio de correo electrónico y que estamos interesados en contar el número de correos no deseado (SPAM) que recibimos. Supongamos que recibimos 2 correos no deseados por minuto.
¿Cuál es la probabilidad de que en 4 minutos recibamos exactamente 3 correos no deseados?
X_t=número de correos no deseados en t minutos.
Supongamos que es un proceso de Poisson X_t con parámetro $\lambda_t=\lambda\cdot t=2\cdot t$.
X_4 = número de correos no deseados en 4 minutos es una $Po(2\cdot 4)$.
$P(X_4=3)=\frac{8^3}{3!}\cdot e^{-8}=0.0286.$

Ejercicio: Alquiler vacacional

La empresa *Alquilo Beds* (AB) es un servicio de alquiler de apartamentos vacacionales por Internet.
El número de alquileres en el pueblo de la Colonia de Sant Antoni por día a través de AB es de 1.5.

- ¿Cuál es la probabilidad de que en una semana no se alquile ningún apartamento a través de AB?
- ¿Cuál es la probabilidad de que en 15 días se alquilen como mínimo 2 apartamentos mediante AB?

Ejercicio: Alquiler vacacional

La empresa *Alquilo Beds* (AB) es un servico de alquiler de apartamentos vacacionales por internet.
El número de alquileres en el pueblo de la Colonia de Sant Antoni por día a través de AB es de 1.5.
X_t = número de apartamentos alquilados en t días, supongamos que es $Po(1.5\cdot t)$.

- ¿Cuál es la probabilidad de que en una semana no se alquile ningún apartamento a través de AB?
  X_7 es $Po(10.5)$.

$$P(X_7=0)=\frac{10.5^0}{0!}\cdot e^{-10.5}=e^{-10.5}=2.75\times 10^{-5}$$

Ejercicio: Alquiler vacacional

La empresa *Alquilo Beds* (AB) es un servico de alquiler de apartamentos vacacionales por internet.
El número de alquileres en el pueblo de la Colonia de Sant Antoni por día a través de AB es de 1.5.

- ¿Cuál es la probabilidad de que en 15 días se alquilen como mínimo 2 apartamentos mediante AB?
  X_{14} es $Po(21)$:

$$\begin{aligned} P(X_{14} \geqslant 2) &= 1 - P(X_{14} \leqslant 1) \\ &= 1 - (P(X_{14} = 0) + P(X_{14} = 1)) \\ &= 1 - e^{-21} \cdot \frac{21^0}{0!} - e^{-21} \cdot \frac{21^1}{1!} \\ &\approx 1 \end{aligned}$$

Distribución hipergeométrica

- Consideremos el experimento en el que "extraemos de golpe" (o una detrás de otra, sin devolverlas) n objetos de una "urna" en la que hay N de tipo A y M de tipo B (así que en total hay $N + M$ objetos).
- Sea X la v.a qque a cada suceso elemental le asigna el número de objetos de tipo A.
- Diremos que X es una variables hipergeométrica (o que tiene distribución hipergeométrica) de parámetros N, M, n y lo denotaremos por $H(N, M, n)$.

Variable con distribución hipergeométrica $H(N, M, n)$.

X hipergeométrica $H(N, M, n)$.
$D_X = \{x \in \mathbb{N}$ que cumplan que $\min\{N, n\} \leqslant x \leqslant \max\{n, N\}\}$
$P_X(x) = P(X = x) = \begin{cases} \dfrac{\binom{N}{x} \cdot \binom{M}{n-x}}{\binom{N+M}{n}} & \text{si } x \in D_X \\ 0 & \text{en otro caso.} \end{cases}$
$F_X(x) = P(X \leqslant X) =$ Función de R o tabulada
$E(X) = \frac{n \cdot N}{N+M}$. $Var(X) = \frac{n \cdot N \cdot M}{(N+M)^2} \cdot \frac{N+M-n}{N+M-1}$.

Ejemplo: Distribución hipergeométrica

$f(k) = \binom{N}{k} \cdot \binom{M}{n-k} / \binom{N+M}{n}$

En un lago de un parque natural hay 500 peces, los guardias del parque han anillado 20 ejemplares. Si capturamos 15 ¿cuál es la probabilidad de que capturemos al menos uno marcado?

$X=$ número de peces marcados capturados

Es $H(20, 480, 15)$

$$\begin{aligned} P(X \geqslant 1) &= 1 - P(X = 0) = 1 - f(0) \\ &= 1 - \frac{\binom{20}{0} \cdot \binom{480}{15}}{\binom{500}{15}} = 0.4627 \end{aligned}$$

Distribució hipergeomètrica

$E(X) = \frac{n \cdot N}{N+M}$

En un lago de un parque natural hay 500 peces, los guardia del parque han anillado 20 ejemplares. Si capturamos 15 ¿cuál es el número esperado de peces marcados que hemos pescado?

$X=$ número de peces marcados capturados
Es $H(20, 480, 15)$

$$E(X) = \frac{15 \cdot 20}{500} = 0.6$$