Tarea
Curso: Labaratorio avanzado I
Fecha: 16 de octubre de 2020
Alumno: Edgar Iván Preciado Govea
NUA: 341087
Profesor: Dr. Alma Gonzalez

Distribución de Poisson

Sea x una variable aleatoria discreta. Se dice que ésta tiene una distribución de Poisson con parametro $\lambda > 0$, si, para $k = 0, 1, 2, ...$ su función de probabilidad está dada por

$$f(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{1}$$

A su vez, podemos definir la función de distribución acumulativa asociada a (1) como

$$F(x_k) \equiv P(\leq x_k) = \sum_{x_i \leq x_k} f(x_i) \tag{2}$$

con propiedades

$$F(-\infty) = 0, \tag{3}$$

$$F(\infty) = 1. \tag{4}$$

Usando (1) en (2) tenemos que la CDF está dada por

$$F = e^{\lambda} \sum_{k=0} \frac{\lambda^k}{k!}. \tag{5}$$

La distribución (1) cumple con las propiedad de que su media y su varianza son iguales

$$\mathcal{P}(\lambda)\mathbf{E}[\mathbf{X}] = \mathbf{Var}[\mathbf{X}] = \lambda. \tag{6}$$

Calculemos el valor esperado. Se sigue de (1) que

$$E[X] = \sum_{o}^{\infty} x e^{\lambda} \frac{\lambda^x}{x!} = \lambda \sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!}. \tag{7}$$

Efectuando el cambio de variable $y = x - 1$, rescribimos

$$E[X] = \lambda \sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!} = \lambda. \tag{8}$$

Para la ultima relación hemos hecho uso de que

$$\sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!} = 1. \tag{9}$$

Para calcular la varianza observamos que

$$Var[X] = E[X^2] - (E[X])^2 \tag{10}$$

y

$$E[X^2] = E[X(X-1)] + E[X]. \tag{11}$$

Un procedimiento idéntico al seguido en (7), nos permite identiciar $E[X(X-1)] = \lambda^2$. Por lo que $E[X^2] = \lambda^2 + \lambda$. Con lo que finalmente, usando (10), escribimos

$$Var[X] = \lambda. \tag{12}$$

Distribución Binomial

La distribución binomial se usa cuando hay exactamente dos resultados mutuamente excluyentes de un ensayo. Estos resultados se denominan **éxito** y **fracaso**. La distribución binomial se utiliza para obtener la probabilidad de observar X éxitos en N ensayos, con la probabilidad de éxito en un único ensayo denotado por p. La distribución binomial supone que p es fijo para todos los ensayos. La función de distribución de probabilidad binoial está dada por

$$f(x; p, n) = \binom{n}{x} p^x (1-p)^{(n-x)} \tag{13}$$

para $x = 0, 1, 2, ...n$ donde

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}. \tag{14}$$

Con (13) en (2), escribimos la **CDF** como

$$F(x; p, n) = \sum_{i=0}^{x} \binom{n}{i} (p)^i (1-p)^{(n-i)}. \tag{15}$$

La función de probabilidad binomial cumple con la propiedad de que su valor esperado y su varianza están dados por np y $\sqrt{np(1-p)}$ respectivamente.

Transformación de Variables

Sea x la variable de la cual obtenemos mediciones. Nos interesa, sin embargo, conocer $y(x)$. Si conocemos la PDF $p(x)$, podemos escribir

$$p(y) = p_x[\Phi^{-1}(y)] \left| \frac{d\Phi^{-1}(y)}{dy} \right|. \tag{16}$$

Ahora, si consideramos el caso particular en que $p(x) = 1$ para $0 \leq x \leq 1$ y $y = \Phi(x) = exp(x)$. Podemos escribir $\Phi^{-1}(y) = \ln y$, talque

$$p(y) = \frac{d(\ln y)}{dy} = \frac{1}{y}. \tag{17}$$

donde hemos hecho $p_x(\Phi^{-1}(y)) = 1$.

χ^2 y minimos cuadrados.
Sea χ^2 dada por

$$\chi^2 = \frac{\sum (y_i - y(x_i, \theta))^2}{\sigma^2}. \tag{18}$$

Con y_i y x_i los datos y y la función a ajustar con parámetros θ. Para encontrar los parametos libre debemos minimisar χ^2,

$$\frac{\partial \chi^2}{\partial \theta} = 0. \tag{19}$$

Para una linea recta $y = mx + b$, donde $\theta_{1,2} = m, b$. Minimizando con respecto a estos parámetros obtenemos el sistema de ecuaciones dado por

$$\sum (y_i - mx_i - b)x_i = 0 \tag{20a}$$

$$\sum (y_i - ax_i - b) = 0. \tag{20b}$$

De lo anterior podemos escribir el parámetro b como

$$b = \frac{1}{n} \sum (y_i - ax_i). \tag{21}$$

Con esta expresion para b en (20a) obtenemos

$$\sum n y_i x_i - mn \sum x_i^2 - \sum y_i \sum x_i + m \sum x_i \sum x_i = 0. \tag{22}$$

Despejando m obtenemos

$$m = \frac{n \sum y_i x_i - \sum y_i \sum x_i}{n \sum x_i^2 - (\sum x_i)^2} \tag{23}$$

Finalmente usando está expresión de m en b, relación (21) obtenemos

$$b = \frac{\sum y_i \sum x_i^2 - \frac{1}{n} - \sum y_i x_i}{n \sum x_i^2 - (\sum x_i)^2}. \tag{24}$$

Los resultados (23) y (24) son los mismos referentes a el método de mínimos cuadrados.

Poisson likelihood
Al tratar con conteos por unidad de tiempo, es útil usar la estadística de Poissson (Poisson likelihood) para evaluar que tan bien un modelo en particular describe los datos.

En una distribución de Poisson, la probabilidad de observar k numero de cuentas en los datos, dado un valor de λ (6), esta dado por

$$p(k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{25}$$

Ahora, si tenemos N puntos de datos, tal que $k = 1, ..., N$, la probabilidad de observar aquellos datos con las predicciones dl modelo para cada punto, λ_i, es

$$\mathcal{L} = \prod_{i=1}^{N} \frac{\lambda_i^{k_i} e^{-\lambda_i}}{k_i!} \tag{26}$$

Los parámetros óptimos maximizan (26). En términos de logaritmo podemos escribir,

$$-\log \mathcal{L} = \sum_{i=1}^{N} (-k_i \log \lambda_i + \lambda_i), \tag{27}$$

donde hemos descartado el termino $-\log k_i$ ya que éste es independiente del modelo. Debemos derivar para encontrar los parámetros que maximizan (27). En el caso de un sólo parámetro λ, tras derivar encontramos que

$$\frac{1}{\lambda} \sum_{i=1}^{N} x_i - N = 0, \tag{28}$$

lo que implica

$$\lambda = \bar{x}. \tag{29}$$