Variables aleatorias discretas

CC3039

Recordatorio:
Las VA pueden interpretarse como una medición sobre un resultado de un experimento estocástico

Para entender una VA nos interesa saber...

¿Qué miden?

Dato importante número 1

¿Cuáles son sus parámetros?

Dato importante número 2

Variables aleatorias de Bernoulli

Experimentos de un solo intento. La observación es bien un **éxito** o un **fracaso.** La probabilidad de éxito es **p**.

X es una variable aleatoria de Bernoulli con parámetro **p** si su distribución se enuncia así:

$$p_X(x) = \begin{cases} p & if\, x = 1 \\ 1 - p & if\, x = 0 \\ 0 & if\, x \notin R_X \end{cases}$$

Variables aleatorias de Bernoulli

Para X con distribución de Bernoulli, tenemos que:

E[X] = p

Var(X) = p(1 - p)

Variables aleatorias binomiales

Experimentos de intentos independientes.

Cada observación es bien un **éxito** o un **fracaso.**

La probabilidad de éxito es **p**, la cantidad de intentos es **n** y X representa la cantidad de éxitos que ocurren en esos **n** intentos.

X es una variable aleatoria binomial con parámetros **n** y **p**

Variables aleatorias binomiales

La probabilidad de estas variables está dada por:

$$P_i \equiv P\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, 1, \ldots, n$$

$$\binom{n}{i} = \frac{n!}{i!(n-i)!}$$

Variables aleatorias binomiales

Indicadores:

$$E[X] = \sum_{i=1}^{n} E[X_i] = np$$

$$\begin{aligned} \mathrm{Var}(X) &= \sum_{i=1}^{n} \mathrm{Var}(X_i) \quad \text{since the } X_i \text{ are independent} \\ &= np(1-p) \end{aligned}$$

Variables aleatorias binomiales

Para calcular eficientemente una probabilidad binomial, podemos usar una fórmula recursiva:

$$
\begin{aligned}
p_{i+1} &= \frac{n!}{(n-i-1)!(i+1)!} p^{i+1}(1-p)^{n-i-1} \\
&= \frac{n!(n-i)}{(n-i)!i!(i+1)} p^{i}(1-p)^{n-i} \frac{p}{1-p} \\
&= \frac{n-i}{i+1} \frac{p}{1-p} p_i
\end{aligned}
$$

Variables aleatorias de Poisson

Una variable aleatoria de Poisson representa la cantidad de éxitos en un experimento binomial donde tenemos muchos intentos y una probabilidad de éxito muy baja.

Lambda es una medida de *np*

$$P\{X = i\} \approx e^{-\lambda} \frac{\lambda^i}{i!}$$

$$E[X] = \mathrm{Var}(X) = \lambda$$

Variables aleatorias de Poisson

- Modelan el número de éxitos en un intervalo.
- La cantidad de éxitos en el intervalo es independiente de la cantidad de éxitos en otros intervalos.
- Los intervalos no se cruzan
- La probabilidad de éxito es proporcional al intervalo.

Esta distribución modela eventos raros:

Cantidad de quejas cada 100000 clientes

Variables aleatorias de Poisson

Computacionalmente, podemos ahorrar muchos cálculos si usamos un resultado de una probabilidad anterior, para calcular una nueva probabilidad:

$$\frac{p_{i+1}}{p_i} = \frac{\frac{e^{-\lambda}\lambda^{i+1}}{(i+1)!}}{\frac{e^{-\lambda}\lambda^{i}}{i!}} = \frac{\lambda}{i+1}$$

$$p_{i+1} = \frac{\lambda}{i+1} p_i, \quad i \geqslant 0$$

Variables aleatorias de Poisson

Supongan que tenemos un experimento con una variable aleatoria de Poisson **X**, con valor esperado **lambda**.

Supongan que los eventos independientemente de **X** son de **tipo 1** o de **tipo 2**, donde la probabilidad que sea de **tipo 1** es **p** y la probabilidad que sea de **tipo 2** es **1 - p.**

Sea **Ni** la cantidad de eventos que es de **tipo i.**

Resulta que N1 y N2 son independientes, y aún más:

E[N1] = p*lambda y **E[N2] = (1 - p)*lambda**

Variables de Poisson

Sea **X** la variable aleatoria de Poisson (lambda = 3) que representa la cantidad de clientes que llegan a ***Pollo Pinulito*** en la siguiente hora. Se sabe que, a priori, la probabilidad de que un cliente sea mujer es de 0.4.

Sea $\mathbf{X_m}$ la variable aleatoria que indica cuántos clientes mujeres llegarán a Pollo Pinulito en la siguiente hora. Sea $\mathbf{X_h}$ la variable aleatoria que indica cuántos clientes hombres llegarán a Pollo Pinulito en la siguiente hora.

¿Qué sabemos de X_m y X_h?

Variables de Poisson

Sabemos que:

- **Xm** y **Xh** son variables aleatorias de Poisson
- E[**Xm**] = p*lambda = 0.4*3 =1.2
- E[**Xh**] = p*lambda = 0.6*3 =1.8

Este resultado se puede generalizar para más de dos categorías de eventos, donde p1 + p2 + .. + pn = 1.

Variables aleatorias geométricas

Para las mismas condiciones de la variable aleatoria binomial. Si nosotros queremos calcular la probabilidad de que el primer éxito suceda en el n-ésimo intento, tenemos que:

$$P\{X = n\} = p(1-p)^{n-1}, \quad n \geqslant 1$$

Variables aleatorias geométricas

Si una variable aleatoria **X** tiene una distribución de probabilidad como la indicada en el slide anterior, decimos que **X** es geométrica. Los indicadores de **X** son:

$$E[X] = \sum_{n=1}^{\infty} np(1-p)^{n-1} = \frac{1}{p}$$

$$\mathrm{Var}(X) = \frac{1-p}{p^2}$$

Variables aleatorias binomiales inversas (Pascal)

Para las mismas condiciones de la variable aleatoria binomial. Si nosotros queremos calcular la probabilidad de obtener **r éxitos** en **n intentos**

$$P\{X = n\} = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n \geqslant r$$

Variables aleatorias binomiales inversas (Pascal)

Si una variable aleatoria **X** tiene una distribución de probabilidad como la indicada en el slide anterior, decimos que **X** es binomial inversa (o de Pascal). Los indicadores de **X** son:

$$E[X] = \sum_{i=1}^{r} E[X_i] = \frac{r}{p}$$

$$\mathrm{Var}(X) = \sum_{i=1}^{r} \mathrm{Var}(X_i) = \frac{r(1-p)}{p^2}$$

Variables aleatorias Hipergeométricas

Considere un recipiente con N + M ítems. N cumplen con *N(item)* y no cumplen con *M(item)*. M cumplen con *M(item)* y no cumplen con *N(item)*. Si seleccionamos **n** items, y cada subconjunto de n ítems tiene exactamente la misma probabilidad de suceder, sea $\mathbf{X_N}$ la variable aleatoria que indica cuántos ítems en la muestra cumpliran con N.

Variables aleatorias Hipergeométricas

$$P\{X = i\} = \frac{\binom{N}{i}\binom{M}{n-i}}{\binom{N+M}{n}}$$

$$E[X] = \sum_{i=1}^{n} E[X_i] = \frac{nN}{N+M}$$

$$\mathrm{Var}(X) = \frac{nNM}{(N+M)^2}\left(1 - \frac{n-1}{N+M-1}\right)$$