Análisis de Datos Categóricos

Variables Aleatorias y Distribuciones Discretas

Mauricio Bucca
github.com/mebucca
mebucca@uc.cl

18 August, 2024

Variables Aleatorias

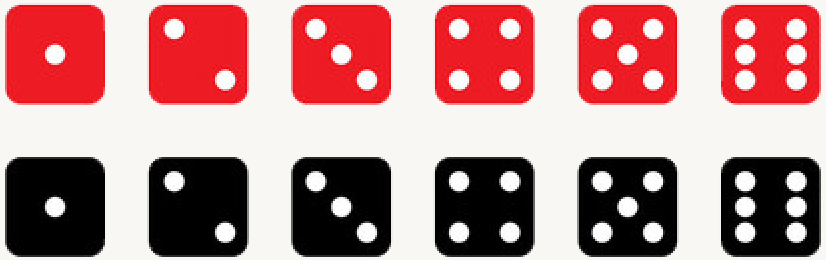
Intuición

Variables Aleatorias

Una variable aleatoria es una variable cuyos valores son el resultado de un experimento aleatorio.

Si Ω es el espacio muestral de un experimento, una variable aleatoria es una función que *mapea* el espacio muestral a los números reales: $\Omega \to \mathbb{R}$.

- Experimento: tirar 2 dados "justos" simultáneamente



- Espacio muestral Ω: $\{(1;1),(1;2),\ldots,(5;6),(6;6)\}$

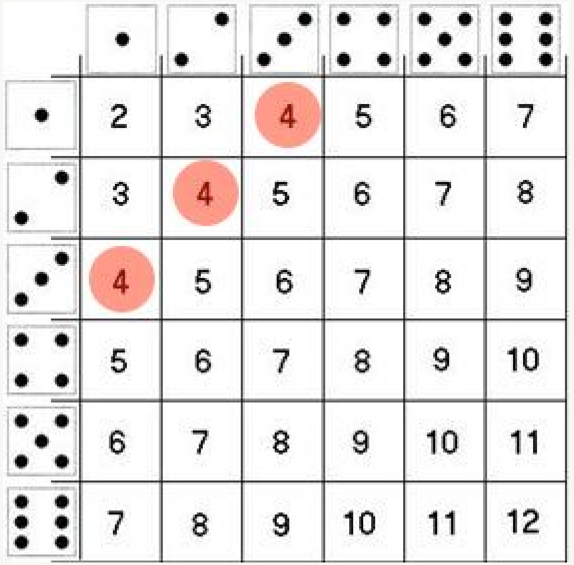
X es la variable aleatoria que resulta de sumar el resultado de ambos dados,

$$X : \{2,3,4,5,6,7,8,9,10,11,12\}$$

Variables Aleatorias

Cada valor posible de una variable aleatoria tiene una probabilidad conocida de ocurrencia, denotada como $\mathbb{P}(X = x)$.

Ejercicio rápido: Si X es la variable que resulta de sumar los dos dados (justos) obtenidos ...



¿Cuál es la probabilidad de que la variable X tome valor 4?

$$\mathbb{P}(X = 4) = \frac{3}{36} = \frac{1}{12}$$

Distribución de una variable aleatoria

El conjunto de las probabilidades asociadas a cada posible resultado de una variable aleatoria se denomina la **distribución** de la variable.

Continuando con nuestro ejemplo, podemos caracterizar la distribución de X con una función:

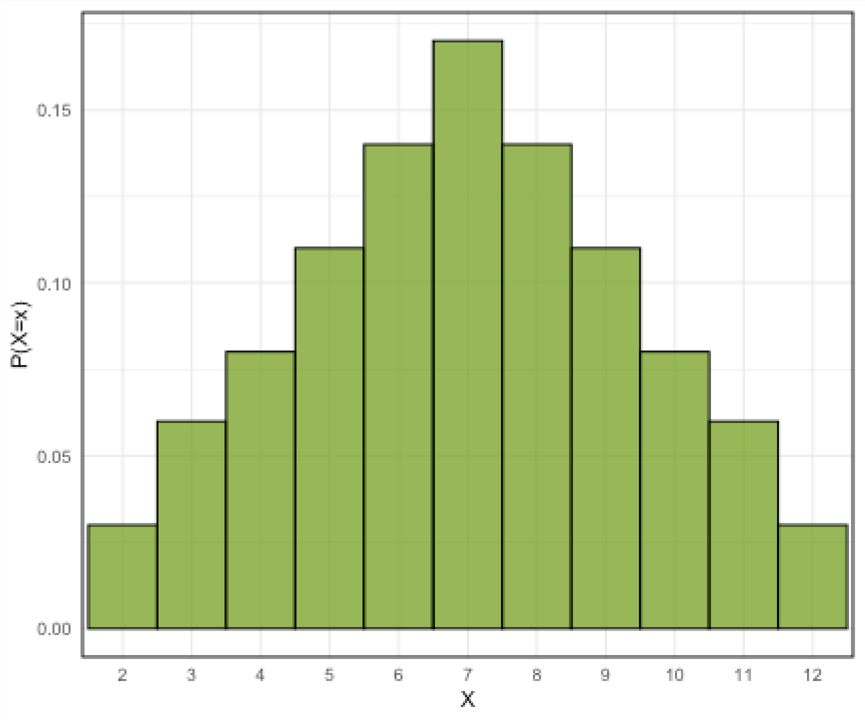
$$f(x) = \begin{cases} \frac{1}{36} & \text{si } x = 2 \text{ o } x = 12 \\ \frac{2}{36} & \text{si } x = 3 \text{ o } x = 11 \\ \frac{3}{36} & \text{si } x = 4 \text{ o } x = 10 \\ \frac{4}{36} & \text{si } x = 5 \text{ o } x = 9 \\ \frac{5}{36} & \text{si } x = 6 \text{ o } x = 8 \\ \frac{6}{36} & \text{si } x = 7 \\ 0 & \text{otherwise} \end{cases}$$

y	P(X=x)
2	0.03
3	0.06
4	0.08
5	0.11
6	0.14
7	0.17
8	0.14
9	0.11
10	0.08
11	0.06
12	0.03

Distribución de una variable aleatoria

Podemos caracterizar la distribución de X con una función:

$$f(x) = \begin{cases} \frac{1}{36} & \text{si } x = 2 \text{ o } x = 12 \\ \frac{2}{36} & \text{si } x = 3 \text{ o } x = 11 \\ \frac{3}{36} & \text{si } x = 4 \text{ o } x = 10 \\ \frac{4}{36} & \text{si } x = 5 \text{ o } x = 9 \\ \frac{5}{36} & \text{si } x = 6 \text{ o } x = 8 \\ \frac{6}{36} & \text{si } x = 7 \\ 0 & \text{otherwise} \end{cases}$$



Importante: $\sum_{i=2}^{12} \mathbb{P}(X = i) = 1$

Distribuciones de Probabilidad

Tipos de variables aleatorias

- **Variables discretas:** variables que solo pueden tomar un número contable de valores distintos y separados, sin valores intermedios posibles entre ellos. Usualmente medidas con números enteros.

Ejemplo: cara/sello, número de accidentes de tránsito, etc.

- **Variables continuas:** variables que pueden tomar cualquier valor dentro de un rango especificado. Estas variables tienen un número infinito de posibles resultados y no están limitadas a valores aislados.

Ejemplo: Los ingresos de una persona pueden tomar cualquier valor entre 0 y ∞+.

Distribución de una variable aleatoria

El conjunto de las probabilidades asociadas a cada posible resultado de una variable aleatoria se denomina la **distribución** de la variable.

La distribución de una variable se puede caracterizar de manera **única** de (al menos) dos maneras:

- **Función de masa/densidad de probabilidad** (PMF/PDF): $f(x)$
  - En el caso de variables discretas $f(x) = \mathbb{P}(X = x)$.
  - Ejemplo, evaluando $f(5)$ obtenemos la probabilidad de que X tome valor 5.

- **Función de distribución acumulada** (CDF): $F(x) = \mathbb{P}(X \leq x)$
  - Ejemplo, evaluando $F(5)$ obtenemos la probabilidad de que X tome un valor igual o menor a 5.

Distribuciones Discretas

Intuición

Tiro de penal

Final de la Copa Mundial de la FIFA 1994. Roberto Baggio va a tirar su penal ...

Número finito de resultados posibles:

- Gol
- Fallo

Evento aleatorio:

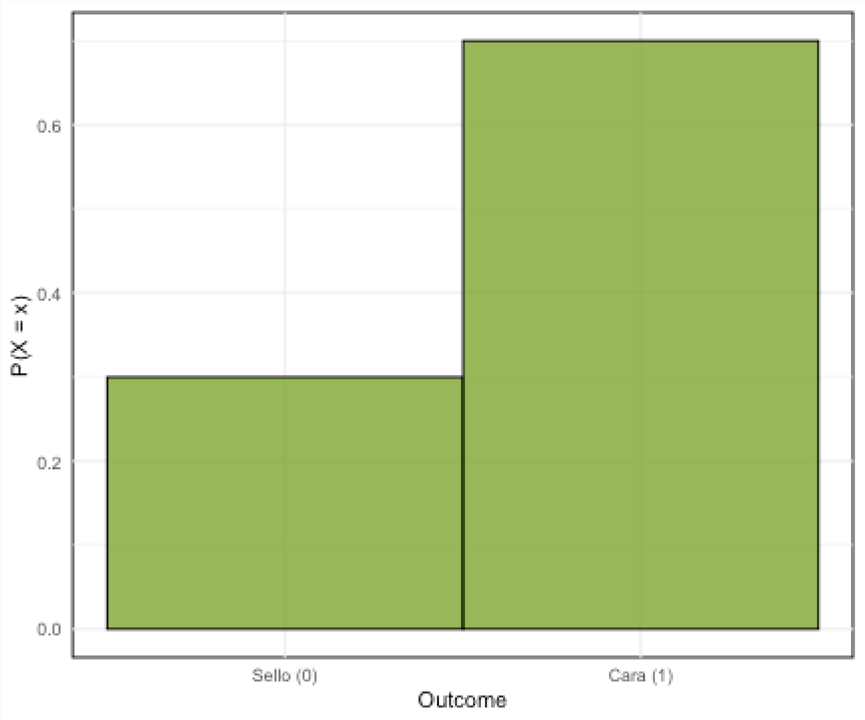
- Existe una probabilidad asociada a cada resultado

Distribuciones Discretas

Distribución Bernoulli

Distribución Bernoulli

- Experimento: tiramos una moneda al aire.
- X es una variable aleatoria tal que $X = 1$ (Cara) o $X = 0$ (Sello).
- La probabilidad de obtener Cara es $p = 0.7$ y la de Sello es $1 - p = 0.3$



Función de masa probabilistica (PMF):

X es una variable aleatoria Bernoulli, es decir

$$f(x) = \begin{cases} p & \text{si } x = 1 \\ 1 - p & \text{si } x = 0 \\ 0 & \text{otro} \end{cases}$$

En modo más sintético:

$$f(x) = p^x (1-p)^{1-x} \quad \text{si } x = 1 \text{ o } x = 0$$

Distribución Bernoulli

Ilustración via simulación en R

Tiremos una moneda con probabilidad de obtener "Cara" (1) de 70% ($p = 0.7$)

```
#set.seed(12347)
moneda <- rbinom(n=1, size=1, p=0.7)
print(moneda)
```

```
## [1] 0
```

Repitamos el proceso 1000 veces ...

```
#set.seed(12347)
monedas <- rbinom(n=1000, size=1, p=0.7)
```

```
##  [1] 1 1 1 1 1 1 0 1 0 1 1 1 1 0 1 0 1 1 0 1
```

```
## [1] "P(Cara) =  0.744"
```

Distribuciones Discretas

Distribución Binomial

Distribución Binomial

La distribución binomial es la distribución de la suma de variables Bernoulli *independientes y con distribución idéntica* (**iid**).

Ejemplo,

- Supongamos que X es una variable de Bernoulli que toma el valor 1 cuando se obtiene "Cara" al lanzar una moneda
- $\mathbb{P}(X = 1) = p$

- Ahora, supongamos que lanzamos la misma moneda 3 veces. Llamamos a estas variables X_1, X_2, X_3
- Definamos $Y = X_1 + X_2 + X_3$
- Y sigue una distribución Binomial, o $Y \sim \text{Binomial}$

Distribución Binomial

Ejercicio rápido:

Pregunta 1: ¿Cuál es la probabilidad de obtener tres "Caras"? Es decir, ¿Cuál es la probabilidad de que $Y = 3$?

- Dado que los 3 ensayos son independientes podemos expresar esta probabilidad como:

$$\mathbb{P}(Y = 3) = \mathbb{P}(X_1 = 1, X_2 = 1, X_3 = 1) = \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1)\mathbb{P}(X_3 = 1)$$

- Y dado que las tres variables distribuyen Bernoulli con la misma probabilidad p, obtenemos:

$$\mathbb{P}(Y = 3) = p \times p \times p = p^3$$

Distribución Binomial

Pregunta 2: ¿Cuál es la probabilidad de obtener 2 "Caras" con 3 tiros? Es decir, ¿Cuál es la probabilidad de que $Y = 2$?

- Por simpleza, consideremos la siguiente secuencia: $\{X_1 = 1, X_2 = 1, X_3 = 0\}$, que satisface $Y = 2$
- La probabilidad de obtener esta secuencia es:

$$\begin{aligned} \mathbb{P}(X_1 = 1, X_2 = 1, X_3 = 0) &= \mathbb{P}(X_1 = 1) \times \mathbb{P}(X_2 = 1) \times \mathbb{P}(X_3 = 0) \\ &= p \times p \times (1 - p) = p^2(1 - p) \end{aligned}$$

- Sin embargo, hay 3 secuencias que satisfacen $Y = 2$. También $\{X_1 = 1, X_2 = 0, X_3 = 1\}$ y $\{X_1 = 0, X_2 = 1, X_3 = 1\}$, cada una con probabilidad de ocurrencia $p^2(1 - p)^1$. Por tanto:

Respuesta: la probabilidad de conseguir 2 "Caras" con 3 tiros es:

$$\mathbb{P}(Y = 2) = 3 \times p^2(1 - p)^1$$

Distribución Binomial

Generalización: lanzamos la misma moneda n veces y la variable Y cuantifica el número de "Caras" (1) obtenidas.

$$Y = \sum_{i=1}^{n} X_i$$

Pregunta: ¿Cuál es la probabilidad de conseguir y "Caras" con n tiros?

- La probabilidad de obtener una secuencia particular con y "Caras" a partir de n lanzamientos es $p^y(1-p)^{n-y}$
- Existen $\binom{n}{y} = \frac{n!}{y!(n-y)!}$ secuencias de este tipo... Por tanto,

$$\mathbb{P}(Y = y) = f(y) = \frac{n!}{y!(n-y)!} \times p^y(1-p)^{n-y}$$

En otras palabras, Y distribuye binomial con **parámetros** n y p: $Y \sim \text{Binomial}(n, p)$

Distribución Binomial

En práctica ...

- **Contexto**: Tenemos una moneda que, cuando se lanza, tiene una probabilidad de $p = 0.6$ de caer en "Cara" y una probabilidad de $1 - p = 0.4$ de caer en "Sello".
- **Problema**:¿Cuál es la probabilidad de obtener 3 "Caras" con 10 lanzamientos?
- **Solución**: $X \sim \text{Binomial}(n = 10, p = 0.6)$

$$\mathbb{P}(X = 3) = \binom{10}{3} \times (0.6)^3 \times (0.4)^{10-3}$$

$$\mathbb{P}(X = 3) = \frac{10!}{3! \times 7!} \times (0.6)^3 \times (0.4)^7$$

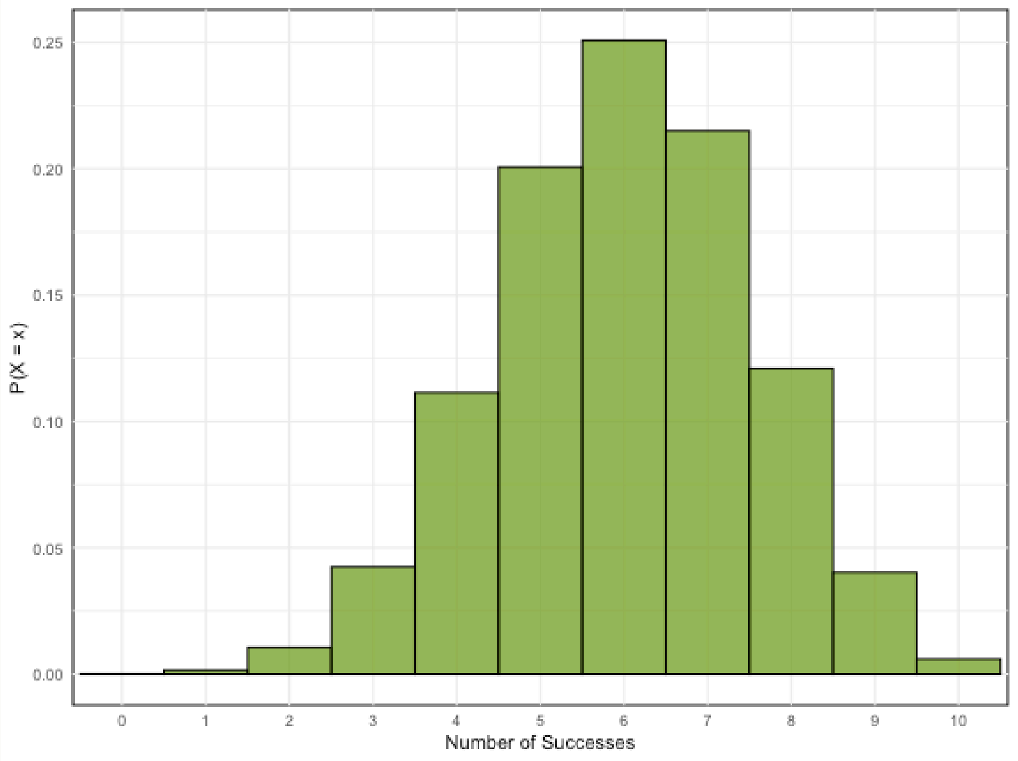
$$\mathbb{P}(X = 3) = \frac{10 \times 9 \times \ldots 1}{(3 \times 2 \times 1) \times (7 \times 6 \times \ldots 1)} \times (0.6)^3 \times (0.4)^7$$

$$\mathbb{P}(X = 3) = 120 \times 0.216 \times 0.0028$$

$$\mathbb{P}(X = 3) \approx 0.0425$$

Distribución Binomial

- Ahora observamos la distribución completa
- La probabilidad de obtener $0, 1, \ldots, 10$ caras es:



Función de masa probabilistica (PMF):

X es una variable aleatoria Binomial, es decir

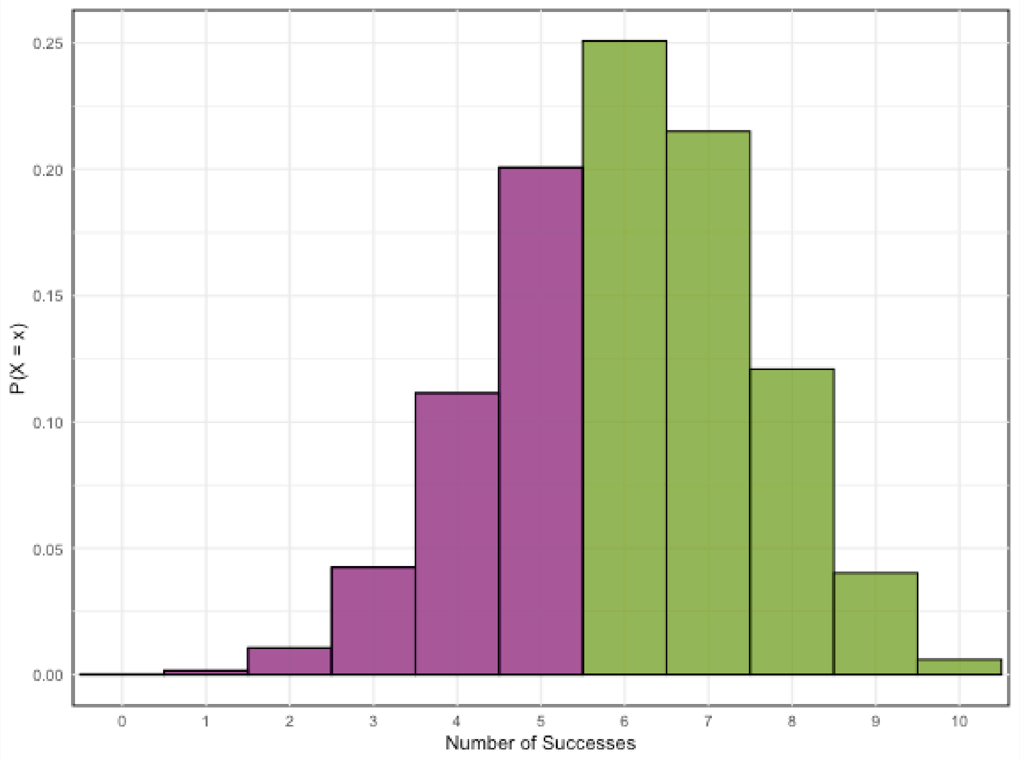
$$\mathbb{P}(X = x) = f(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

En este caso,

$n = 10$ y $p = 0.6$

Distribución Binomial

- Ahora observamos la distribución completa
- ¿Cual es la probabilidad de obtener 5 monedas o menos?



Función de distribución acumulada (CDF):

X es una variable aleatoria Binomial, es decir

$$F(k;n,p) = \mathbb{P}(X \leq k)$$

$$= \sum_{i=0}^{k} \mathbb{P}(X = i)$$

$$= \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}$$

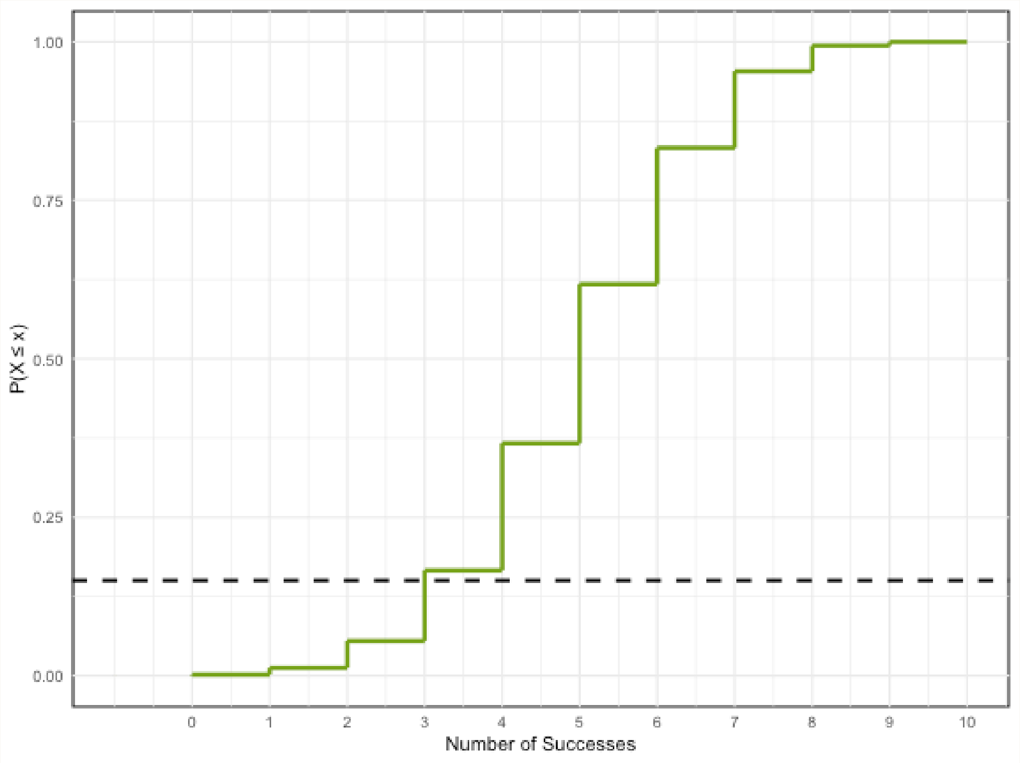
En este caso,

$$F(k=5; n=10, p=0.6) = \mathbb{P}(X=0) + \cdots + \mathbb{P}(X=5)$$

$$\approx 0.366$$

Distribución Binomial

- Ahora observamos la distribución completa
- ¿Cual es la probabilidad de obtener 5 monedas o menos?



Función de distribución acumulada (CDF):

X es una variable aleatoria Binomial, es decir

$$F(k;n,p) = \mathbb{P}(X \leq k)$$

$$= \sum_{i=0}^{k} \mathbb{P}(X = i)$$

$$= \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}$$

En este caso,

$$F(k=5; n=10, p=0.6) = \mathbb{P}(X=0) + \cdots + \mathbb{P}(X=5)$$

$$\approx 0.366$$

Hasta la próxima clase. Gracias!

Mauricio Bucca
https://mebucca.github.io/
github.com/mebucca