Variables_aleatorias_discretas_

15 de junio de 2024

1. Variables aleatorias discretas

[1]:
```
import numpy as np
import numpy.random as npr
import scipy.stats as sps
import matplotlib.pyplot as plt
import random

import warnings
warnings.filterwarnings("ignore")
```

2. Variable aleatoria uniforme discreta

Una variable aleatoria X tiene distribución uniforme discreta en el conjunto $\{x_1, \ldots, x_n\}$ si su función de masa de probabilidades está dada por

$$f_X(k) = \mathbb{P}(X = k) = \begin{cases} \frac{1}{n} & \text{si } k \in \{x_1, \ldots, x_n\} \\ 0 & \text{en otro caso} \end{cases}$$

Notación: $X \sim Unif(x_1, \ldots, x_n)$.

Se llama uniforme porque cada uno de los posibles valores de X tienen la misma probabilidad.

Si $X \sim Unif(x_1, \ldots, x_n)$, en este caso, se tiene:

$$\mathbb{E}[X] = \sum_k k f_X(k) = \frac{x_n + x_1}{2}.$$

y

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_k (k - \mathbb{E}[X])^2 f_X(k) = \frac{(x_n - x_1 + 1)^2 - 1}{12}$$

Ejercicio 1. Verifica que, en efecto, la esperanza y varianza de una variable aleatoria geométrica está dada como antes.

Ejemplo Se lleva a cabo una rifa donde los boletos están enumerados del 00 al 99. Si Y es la variable aleatoria definida como el número del boleto ganador, entonces:

$$\mathbb{P}(Y=k)=\begin{cases}\frac{1}{100} & \text{si } k=00,01,\ldots,99\\ 0 & \text{en otro caso}\end{cases}$$

Consideremos que el premio de la rifa se determina a partir del número premiado de la siguiente forma: $X=Y+1$, donde X es el monto del premio en pesos y Y es el número premiado, entonces X es una variable aleatoria, pues es una función de Y, y además se tiene

$$f_X(k)=\begin{cases}\frac{1}{100} & \text{si } x=1,2,\ldots,100\\ 0 & \text{en otro caso}\end{cases}$$

Entonces, la esperanza y varianza de X están dados por:

$$\mathbb{E}[X]=\sum_{k=1}^{100}k\left(\frac{1}{100}\right)=\frac{1}{100}\sum_{k=1}^{100}k=\frac{1}{100}\left[\frac{100(100+1)}{2}\right]=50{,}5$$

y

$$\operatorname{Var}(X)=\mathbb{E}[X^2]-(\mathbb{E}[X])^2=\sum_{k=1}^{100}k^2\left(\frac{1}{100}\right)-(50{,}5)^2=833{,}25$$

Supongamos que nos interesa calcular la probabilidad de que el premio sea mayor a \$80, entonces

$$\mathbb{P}(X>80)=\sum_{k=81}^{100}\frac{1}{100}=\frac{20}{100}=0{,}2$$

Su función de distribución esta dada por:

$$F_X(x)=\mathbb{P}(X\leq x)=\begin{cases}0 & \text{si } x<1\\ \frac{x}{100} & \text{si } x=1,2,\ldots,100\\ 1 & \text{si } x>100\end{cases}$$

Ejercicio 2. Dibuja a la función de distribución acumulada de la variable aleatoria:

$$S:\{1,2,3,4,5,6\}^2\to\{2,3,4,5,6,7,8,9,10,11,12\}$$

dada por $S(i,j)=i+j$.

2.1. Variable aleatoria Bernoulli con parámetro $p\in(0,1)$.

El modelo probabilístico Bernoulli se aplica a un experimento cuyo espacio muestral está constituido sólo por dos resultados posibles, éxito y fracaso:

Se considerará una v.a X sobre el espacio muestral

$$\Omega=\{\text{éxito, fracaso}\}$$

de tal forma que * $X(\{\text{éxito}\})=1$ * $X(\{\text{fracaso}\})=0$

Las probabilidades asociadas a este modelo son:

- $\mathbb{P}(\{\text{éxito}\}) = p$
- $\mathbb{P}(\{\text{fracaso}\}) = 1 - p$

donde $0 < p < 1$.

Entonces la función de densidad de masa de probabilidades de X está dada por:

$$f_X(k) = \begin{cases} p & \text{si } k = 1 \\ 1 - p & \text{si } k = 0 \end{cases}$$

Notación: $X \sim \text{Ber}(p)$.

Esta variable aleatoria es muy importante. En la practica es usada para modelar situaciones en las cuales hay dos posibles resultados como: * el estado de un teléfono en un momento dado: libre u ocupado. * una persona pueda estar enferma o sana de una determinada enfermedad. * la preferencia de una persona, la cual puede estar a favor o en contra de un candidato determinado.

Mediante la combinación de variables aleatorias Bernoulli es posible construir otras variables aleatorias.

2.1.1. Esperanza y Varianza

Si $X \sim \text{Ber}(p)$, entonces su esperanza es:

$$\mathbb{E}[X] = \sum_{k=0}^{1} k f_X(k) = 1 \cdot p + 0 \cdot (1 - p) = p.$$

Y su varianza se calcula usando,

$$\mathbb{E}[X^2] = \sum_{k=0}^{1} k^2 f_X(k) = 1^2 \cdot p + 0^2 \cdot (1 - p) = p.$$

Por lo tanto,

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1 - p)$$

Módulo Numpy: https://numpy.org Módulo Numpy.random: https://numpy.org/doc/stable/reference/random/gen

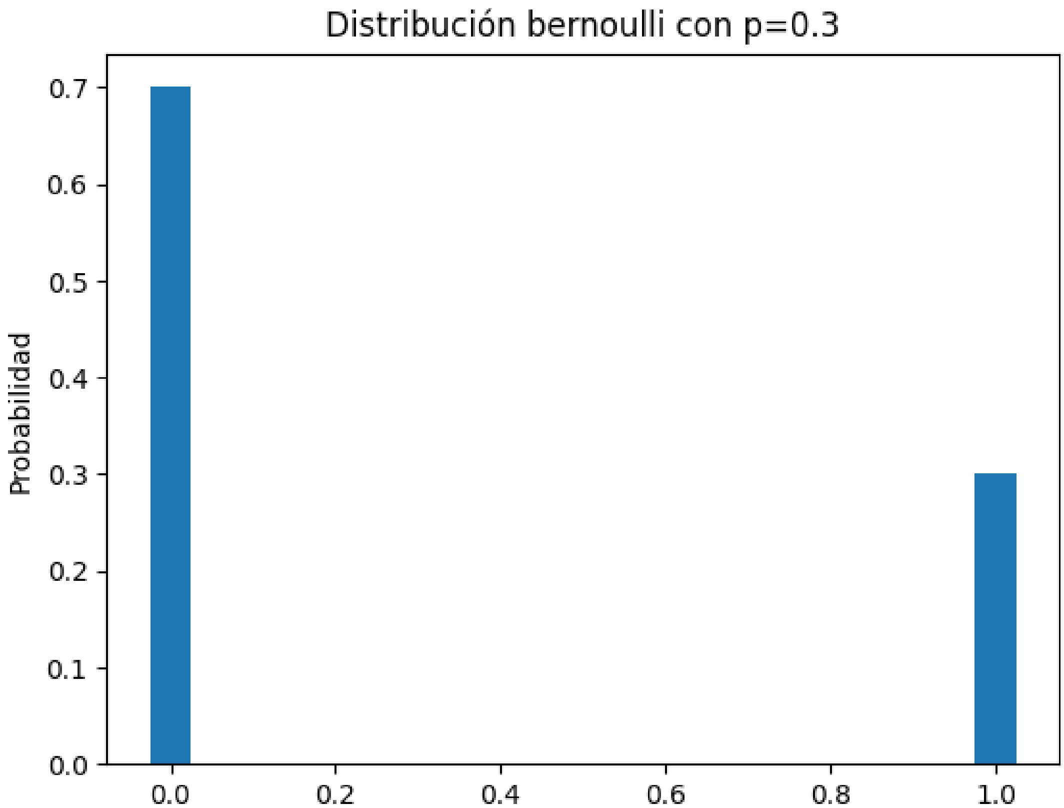
[2]:
```
import numpy as np

n, p = 1, .7  # numbero de lanzamientos, probabilidad de cada ensayo
s = np.random.binomial(n, p, 10)
# muestra de 10 lanzamientos
s
```

[2]:
```
array([1, 1, 1, 1, 1, 0, 1, 1, 1, 1])
```

Lo que usamos es el módulo Scipy.stats: https://docs.scipy.org/doc/scipy/reference/stats.html

[3]:
```python
from scipy.stats import bernoulli

#Definimos la probabilidad el éxito
p=0.3

#Calculo de la función de densidad
x=[0,1]
print("Función de masa = ",bernoulli.pmf(x,p))

#Grafica de la función de densidad
plt.bar(x,bernoulli.pmf(x,p),width=0.05) # color=["r","b"]
plt.xlabel("")
plt.ylabel("Probabilidad")
plt.title("Distribución bernoulli con p=0.3")

# Calculo de la espranza y varianza
media = bernoulli.mean(p)
var = bernoulli.var(p)
print("Esperanza =",media)
print("Varianza =",var)
```

```
Función de masa =  [0.7 0.3]
Esperanza = 0.3
Varianza = 0.21
```



2.2. Variable aleatoria Binomial con parámetros n y $p \in (0, 1)$.

Decimos que una v.a X tiene distribución binomial con párametros n y p, si su función de densidad de masa de probabilidades está dada por:

$$f_X(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{si } k = 0, 1, \ldots, n \\ 0 & \text{en otro caso} \end{cases}$$

donde n es un entero positivo y $0 < p < 1$.

La esperanza de X es:

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{j=0}^{m} \binom{m}{j} p^j (1-p)^{m-j} \\
&= np
\end{aligned}
$$

Ahora,

$$
\begin{aligned}
\mathbb{E}[X^2] &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} kn \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} k \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{j=0}^{m} (j+1) \binom{m}{j} p^j (1-p)^{m-j} \\
&= np \left(\sum_{j=0}^{m} j \binom{m}{j} p^j (1-p)^{m-j} + \sum_{j=0}^{m} \binom{m}{j} p^j (1-p)^{m-j} \right) \\
&= np \left(\sum_{j=0}^{m} m \binom{m-1}{j-1} p^j (1-p)^{m-j} + \sum_{j=0}^{m} \binom{m}{j} p^j (1-p)^{m-j} \right) \\
&= np \left((n-1)p \sum_{j=1}^{m} \binom{m-1}{j-1} p^{j-1} (1-p)^{(m-1)-(j-1)} + \sum_{j=0}^{m} \binom{m}{j} p^j (1-p)^{m-j} \right) \\
&= np \left((n-1)p(p+(1-p))^{m-1} + (p+(1-p))^m \right) \\
&= np((n-1)p+1) \\
&= (np)^2 + np(1-p),
\end{aligned}
$$

por lo que la varianza de X es:

$$\mathrm{Var}(X) = np(1-p).$$

Ejemplo Una moneda no justa se tira 6 veces, donde la probabilidad de obtener sol es de 0,3. Sea X el número de veces que cae sol, entonces X tiene una distribución binomial con parámetros $n = 6$ y $p = 0{,}3$. Por lo que

$$
\begin{aligned}
\mathbb{P}(X = 2) &= \binom{6}{2}(0{,}3)^2(0{,}7)^4 = 0{,}3241 \\
\mathbb{P}(X = 3) &= \binom{6}{3}(0{,}3)^3(0{,}7)^3 = 0{,}1852 \\
\mathbb{P}(1 < X \leq 5) &= \mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \mathbb{P}(X = 4) + \mathbb{P}(X = 5) \\
&= 0{,}3241 + 0{,}1852 + \binom{6}{4}(0{,}3)^4(0{,}7)^2 + \binom{6}{5}(0{,}3)^5(0{,}7)^1 \\
&= 0{,}5093 + 0{,}0595 + 0{,}0102 = 0{,}579
\end{aligned}
$$

Note que la probabilidad

$$\mathbb{P}(1 < X \leq 5) = F_X(5) - F_X(1).$$

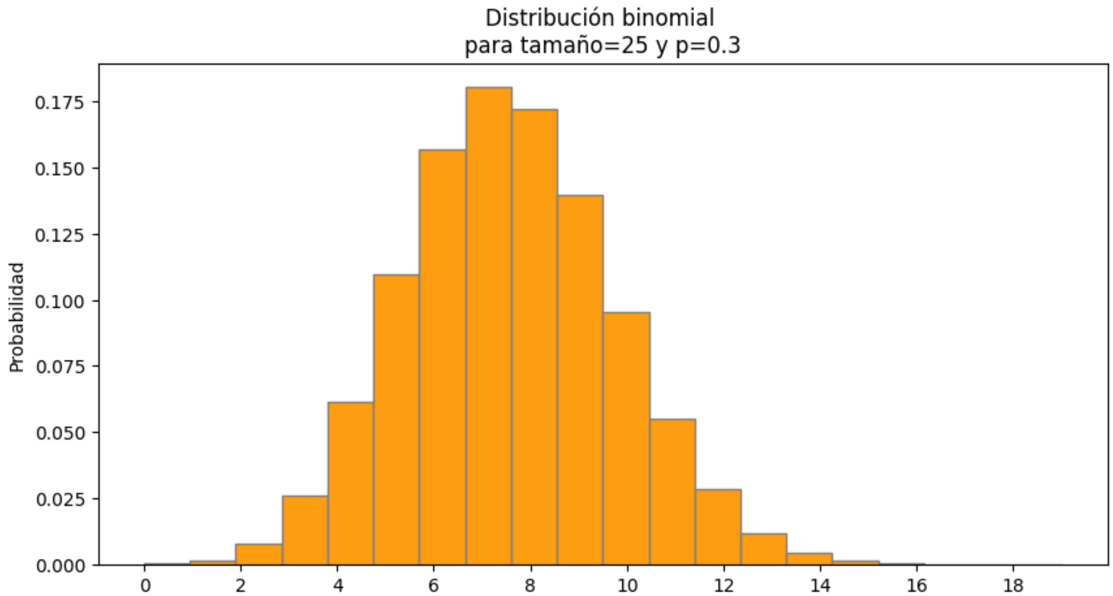
Usando el atributo `.cdf`, podemos calcular la probabilidad anterior:

[4]:
```
from scipy.stats import binom
sps.binom.cdf(5, 6, 0.3)-binom.cdf(1, 6, 0.3)
```

[4]: 0.5790960000000003

[5]:
```
n = 25  # numbero de ensayos Bernoulli
p = 0.3  # probabilidad de éxito
s = 100000  # tamaño de muestras

random.seed(3)
binom_numberos = sps.binom.rvs(n, p, size=s) # genera variables aleatorias␣
↪binomiales

plt.figure(figsize=(10, 5))
plt.hist(
    binom_numberos,
    density=True,
    bins=len(np.unique(binom_numberos)),
    color="#fc9d12",
    edgecolor="grey",
)
plt.xlabel(" ")
plt.ylabel("Probabilidad")
plt.title("Distribución binomial \npara tamaño=25 y p=0.3")
plt.xticks(
    np.arange(min(binom_numberos), max(binom_numberos) + 1, 2.0)
)
```

```
plt.show()
```



3. Variable aleatoria Poisson

Es una distribución de probabilidad discreta que sirve para calcular la probabilidad de que ocurra un determinado número de eventos raros durante un intervalo dado (puede ser de tiempo, longitud, área, etc.

Esta variable aleatoria toma los valores sobre el conjunto $\{0, 1, 2, \dots\}$ y tiene un parámetro $\lambda > 0$, el cual representa el número de veces que se espera que ocurra un evento durante un intervalo dado.

La función de masa de probabilidades de una variable aleatoria Poisson, X, está dada por:

$$f_X(k) = \begin{cases} e^{-\lambda}\frac{\lambda^k}{k!} & \text{si } k = 0, 1, 2, \dots \\ 0 & \text{en otro caso} \end{cases}$$

En efecto, es una función de masa de probabilidades debido a que

$$\begin{aligned} \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} &= e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} \\ &= e^{-\lambda}\underbrace{\left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots\right)}_{\text{desarrollo de Taylor de } e^{\lambda}} \\ &= e^{-\lambda}e^{\lambda} = 1. \end{aligned}$$

Aplicaciones

Ejemplo: Supongamos que el número de accidentes que ocurre en un punto tiene una distribución Poisson con parámetro $\lambda = 2$. * ¿Cuál es la probabilidad de que en un día ocurran más de dos accidentes?

$$\begin{aligned}
\mathbb{P}(X > 2) &= 1 - \mathbb{P}(X \leq 2) \\
&= 1 - [\mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2)] \\
&= 1 - \left[\frac{e^{-2}2^0}{0!} + \frac{e^{-2}2^1}{1!} + \frac{e^{-2}2^2}{2!}\right] \\
&= 1 - e^{-2}[1 + 2 + 2] = 1 - 5e^{-2} = 0{,}3233
\end{aligned}$$

- ¿Cuál es la probabilidad de que en un día ocurran más de dos accidentes, sabiendo que ocurre por lo menos uno?

$$\mathbb{P}(X > 2 \mid X \geq 1) = \frac{\mathbb{P}(X > 2 \mid X \geq 1)}{\mathbb{P}(X \geq 1)} = \frac{\mathbb{P}(X > 2)}{\mathbb{P}(X \geq 1)} = \frac{1 - 5e^{-2}}{1 - e^{-2}} = \frac{0{,}3233}{0{,}8646} = 0{,}3739$$

  ya que $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X < 1) = 1 - \mathbb{P}(X = 0) = 1 - e^{-2}$

Ejercicio 3. Usando el atributo `.cdf` Poisson calcula las probabilidades anteriores.

3.1. Aproximación de Poisson a la Binomial

La distribución de Poisson es una forma límite de la distribución binomial, es decir, es una buena aproximación cuando n es suficientemente grande y p suficientemente pequeña.

Teorema (Poisson).- Sean $S_n \sim Bin(n, p_n)$ bajo el regimen

$$\lim_{n\to\infty} np_n = \lambda > 0.$$

Consideremos la siguiente sucesión de números reales:

$$a_j(n, p_n) = \begin{cases} \binom{n}{j}(p_n)^j(1 - p_n)^{n-j} & j \leq n \\ 0 & j \geq n + 1 \end{cases}$$

Entonces,

$$\lim_{n\to\infty} a_j(n, p_n) = a_j = e^{-\lambda}\frac{\lambda^j}{j!} \quad \forall j \in \mathbb{N}.$$

El teorema anterior implica que la distribución de Poisson ofrece un modelo probabilístico adecuado para todos aquellos experimentos aleatorios en los que las repeticiones son independientes unas de otras y en los que sólo hay dos posibles resultados: éxito o fracaso, con probabilidad de éxito pequeña, y en los que el interés se centra en conocer el número de éxitos obtenidos al realizar el experimento un número suficientemente grande de veces.

Empíricamente se ha establecido, que la aproximación se puede aplicar con seguridad si $n \geq 100$, $p \leq 0{,}01$ y $np \leq 20$.

Ejemplo Supongamos que la probabilidad de que un producto producido por cierta máquina es defectuoso es de 0,1. ¿Cuál es la probabilidad de que un lote de 10 productos contenga a lo más un producto defectuoso?

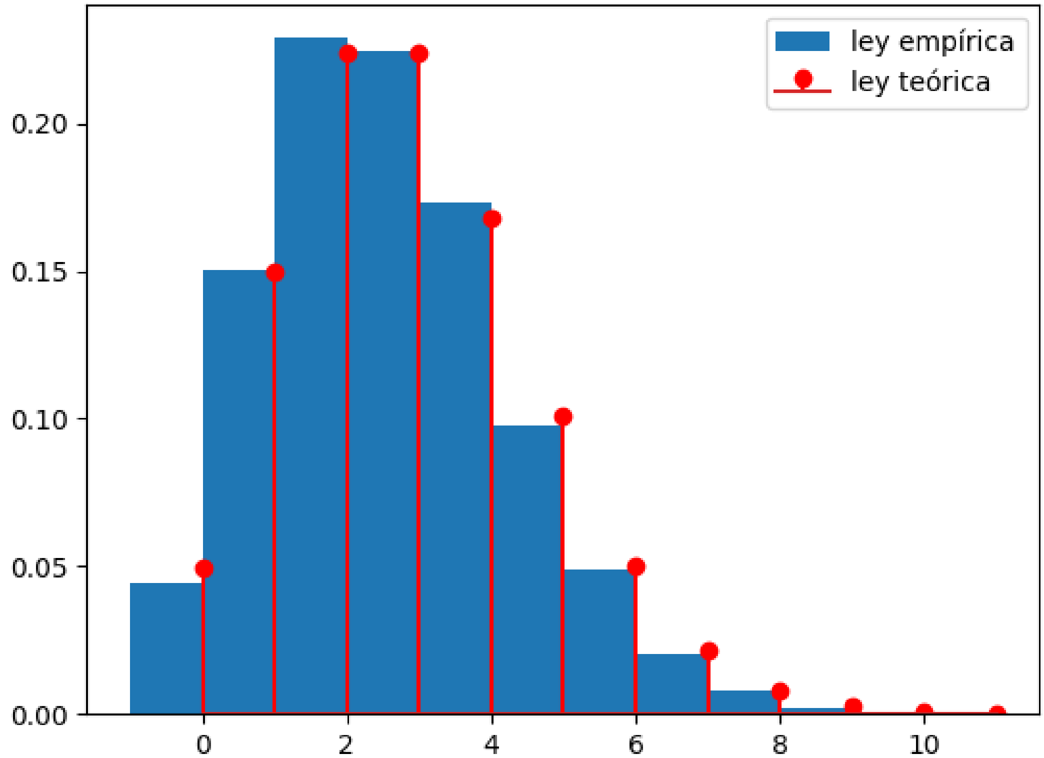
Sea X el número de productos defectuosos, y sabemos que X tiene una distribución binomial con parámetros $n = 10$ y $p = 0{,}1$, entonces

$$\begin{aligned}\mathbb{P}(X \leq 1) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) \\ &= \binom{10}{0}(0{,}1)^0(0{,}9)^{10-0} + \binom{10}{1}(0{,}1)^1(0{,}9)^{10-1} \\ &= 0{,}7361\end{aligned}$$

Ahora, con la distribución Poisson, tenemos que $\lambda = 10(0{,}1) = 1$, por lo que

$$\mathbb{P}(X \leq 1) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = \frac{e^{-1}1^0}{0!} + \frac{e^{-1}1^1}{1!} = e^{-1} + e^{-1} = 0{,}7358$$

[6]:
```python
# Simulación de la aproximación
param=3
n=1000
N=5000

X=npr.binomial(n,param/n,N)

counts = np.bincount(X) / float(N)
x = np.arange(len(counts))
f_x = sps.poisson.pmf(x, param)
plt.close()
plt.bar(x - 0.5, counts, width=1., label="ley empírica")
p2 = plt.stem(x, f_x, "r", label="ley teórica")
plt.legend()
plt.show()
```



3.2. Variable aleatoria Geométrica con parámetro $p \in (0, 1)$.

Esta variable aleatoria cuenta el número de fracasos antes del primer éxito en ensayos bernoulli independientes con parámetro $0 < p < 1$, y su función de masa de probabilidades está dada por:

$$f_X(k) = \begin{cases} p(1-p)^{k-1} & \text{si } k = 1, 2, \ldots \\ 0 & \text{en otro caso} \end{cases}$$

Notación. $X \sim Geo(p)$

La cual es una función de densidad ya que: * $0 \leq f_X(k) \leq 1$ para toda x,

$$\begin{aligned} \sum_{k \in R_X} f_X(k) &= \sum_{k=1}^{\infty} (1-p)^{k-1} p \\ &= p \sum_{y=0}^{\infty} (1-p)^y \\ &= p \left(\frac{1}{1-(1-p)} \right) = 1 \end{aligned}$$

Si por el contrario queremos contar el número de éxitos antes del primer fracaso, tenemos que la función de está dada por:

$$f_X(k) = \begin{cases} p^k(1-p) & \text{si } k = 0, 1, 2, \ldots \\ 0 & \text{en otro caso} \end{cases}$$

Para calcular la Esperanza y Varianza de X, necesitamos del siguiente lema:

Lema 1 Sea x un número real tal que $|x| < 1$. Entonces,

$$\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}.$$

Demostración. Se tiene que

$$\frac{d}{dx}\sum_{k=1}^{\infty} x^k = \sum_{k=1}^{\infty} kx^{k-1}.$$

Además,

$$\sum_{k=1}^{\infty} x^k = \frac{x}{1-x}.$$

Por lo tanto,

$$\frac{d}{dx}\frac{x}{1-x} = \frac{1}{(1-x)^2}.$$

Como corolario,

$$\sum_{k=1}^{\infty} k(k+1)x^{k-1} = \frac{2}{(1-x)^2}.$$

En efecto,

$$\frac{d}{dx}\sum_{k=1}^{\infty}(k+1)x^k = \sum_{k=1}^{\infty} k(k+1)x^{k-1}.$$

Ahora,

$$\sum_{k=1}^{\infty}(k+1)x^k = \sum_{m=2}^{\infty} mx^{m-1} = \frac{1}{(1-x)^2} - 1.$$

Usando estos resultados, la esperanza de X es:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x \geq 1} kp(1-p)^{k-1} \\
&= p\frac{1}{(1-(1-p))^2} \\
&= \frac{p}{p^2} \\
&= \frac{1}{p}
\end{aligned}$$

Calculamos,

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_{k=1}^{\infty} x^2 p(1-p)^{k-1} \\ &= \sum_{k=1}^{\infty} k(k+1)(1-p)^{k-1}p - \sum_{k=1}^{\infty} p(1-p)^{k-1}\end{aligned}$$

Por el Lema 1,

$$\mathbb{E}[X^2] = p\frac{2}{p^3} - \frac{1}{p} = \frac{2}{p^2} - \frac{1}{p}.$$

Por lo tanto,

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

Ejemplo. Cada una de 12 personas lanza una moneda indefinidamente, deteniendo sus lanzamientos en el momento en que se obtiene por primera vez águila. ¿Cuál es la probabilidad de que por lo menos 10 de ellas realicen menos de 4 lanzamientos?

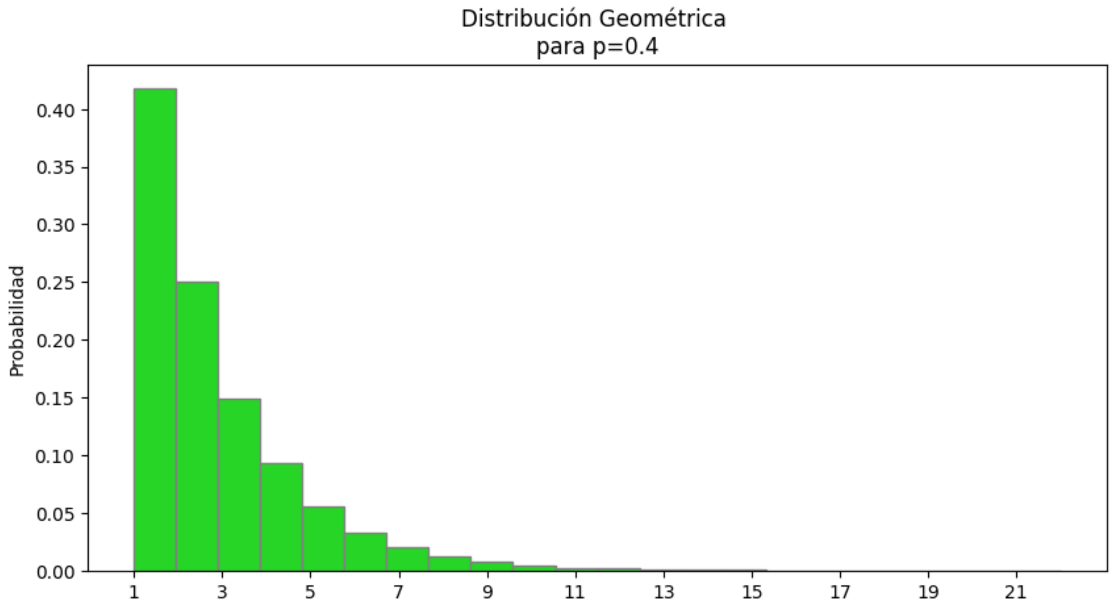
Sea X el número de lanzamientos que se realizan antes de obtener por primera vez águila al lanzar una moneda indefinidamente y sea Y el número de personas que realizan menos de 4 lanzamientos. Así, $Y \sim \text{Binom}\,(12, p = \mathbb{P}(X < 3))$, en donde $X \sim \text{Geo}(1/2)$. Así, - $p = \frac{7}{8}$. - $\mathbb{P}(Y \geq 10) = 0{,}818$

[7]:
```
p = sps.geom.cdf(3, 0.5)
print(p)
#La probabilidad buscada es:
P = 1-sps.binom.cdf(9, 12, p)
print(P)
```

```
0.875
0.8180006196344038
```

[8]:
```
p = 0.4  # probabilidad de éxito
s = 100000  # tamaño de muestras



random.seed(3)
geom_numberos = sps.geom.rvs(p, size=s) # genera variables aleatorias geométricas

plt.figure(figsize=(10, 5))
plt.hist(
    geom_numberos,
    density=True,
    bins=len(np.unique(geom_numberos)),
    color="#26d526",
    edgecolor="grey",
)
plt.xlabel(" ")
plt.ylabel("Probabilidad")
```

```python
plt.title("Distribución Geométrica \npara p=0.4")
plt.xticks(
    np.arange(min(geom_numberos), max(geom_numberos) + 1, 2.0)
)


plt.show()
```



3.3. Variable aleatoria Binomial Negativa con parámetros $r \geq 1$ y $p \in (0, 1)$.

Supongamos que se realizan ensayos independientes, cada uno con probabilidad $0 < p < 1$ de ser un éxito, hasta obtener un total de r éxitos acumulados. Sea X el número de ensayos que se requieren, entonces su función de masa de probabilidades está dada por:

$$f_X(k) = \begin{cases} \binom{n-1}{r-1} p^r (1-p)^{n-r} & \text{si } n = r, r+1, \ldots \\ 0 & \text{en otro caso} \end{cases}$$

Notación. $X \sim \mathrm{BN}(r, p)$.

Se tiene que

$$\mathbb{E}[X] = \frac{r}{p}$$

y

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

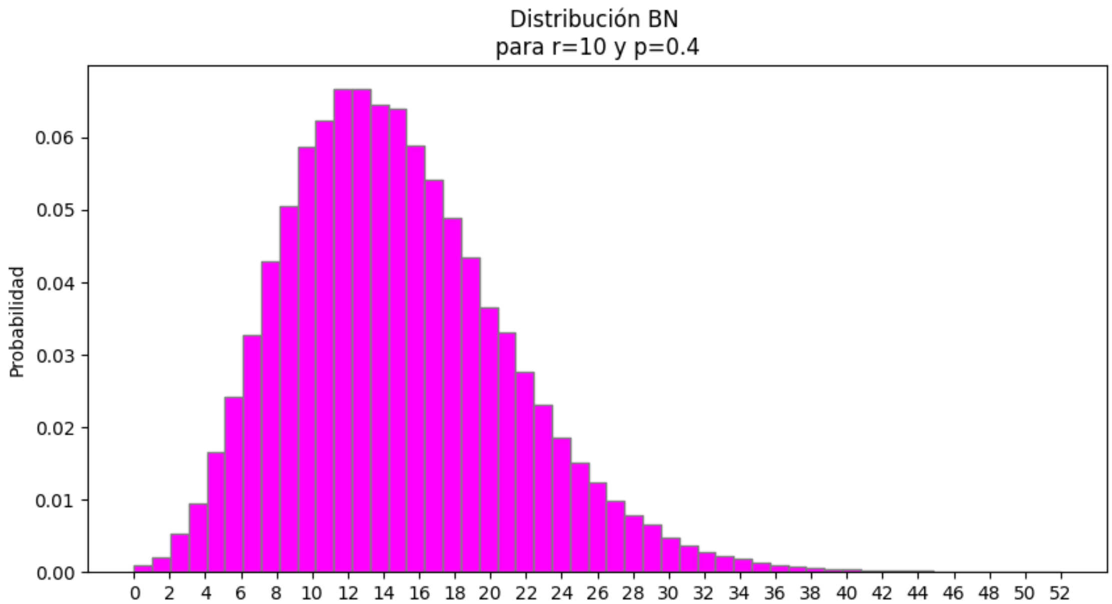
Ejercicio 4. Un examen de Estadística consta de 20 preguntas tipo test y se conoce de experiencias anteriores que un alumno tiene probabilidad 0.7 de contestar bien cada pregunta. Obtener:

a) La probabilidad de que la primera pregunta que contesta bien sea la cuarta.

b) Sabiendo que para aprobar el examen es necesario contestar bien a 10 preguntas, ¿cuál es la probabilidad de que apruebe al contestar la pregunta duodécima?

Ejercicio 5. Investigue sobre el problema de la caja de cerillos de Banach y explique su solución.

[9]:
```
n = 10  # numbero de ensayos Bernoulli
p = 0.4  # probabilidad de éxito
s = 100000  # tamaño de muestras


random.seed(3)
geom_numeros = sps.nbinom.rvs(n,p, size=s) # genera variables aleatorias BNs

plt.figure(figsize=(10, 5))
plt.hist(
    geom_numeros,
    density=True,
    bins=len(np.unique(geom_numeros)),
    color="#FF00FF",
    edgecolor="grey",
)
plt.xlabel(" ")
plt.ylabel("Probabilidad")
plt.title("Distribución BN \npara r=10 y p=0.4")
plt.xticks(
    np.arange(min(geom_numeros), max(geom_numeros) + 1, 2.0)
)


plt.show()
```



3.4. Variable aleatoria Hipergeométrica con parámetros n, N, m.

Supongamos que se elige, sin reemplazo, una muestra de tamaño n de una urna que contiene N bolas, de las cuales m son rojas y $N-m$ son verdes. Sea X el número de ebolas rojas seleccionadas, entonces su función de masa de probabilidades está dada por:

$$f_X(k) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \quad \text{si } i = 0, 1, \dots, n$$

Notación. $X \sim \text{Hiper}(n, N, m)$.

Se tiene que

$$\mathbb{E}[X] = \frac{nm}{N}$$

y

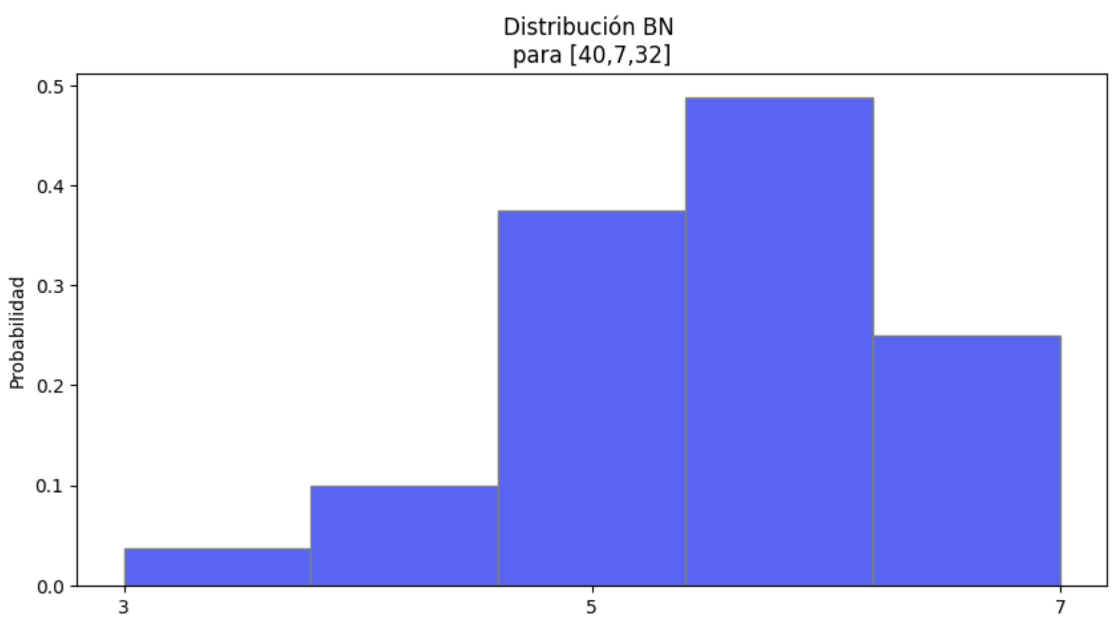
$$\text{Var}[X] = \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\right].$$

Nota. Si $i \leq n$ y $X \sim \text{Hiper}(n, N, m)$, cuando $p = \frac{m}{N}$ y m, N son muy grandes con respecto a n e i:

$$\mathbb{P}(X = i) \approx \binom{n}{i} p^i (1-p)^{n-i}.$$

[10]:
```
n = 40
m = 32
N = 7
```

```
s = 100


random.seed(3)
geom_numeros = sps.hypergeom.rvs(n, m, N, size=s) # genera variables aleatorias␣
↪hgeométricas

plt.figure(figsize=(10, 5))
plt.hist(
    geom_numeros,
    density=True,
    bins=len(np.unique(geom_numeros)),
    color="#5865F2",
    edgecolor="grey",
)
plt.xlabel(" ")
plt.ylabel("Probabilidad")
plt.title("Distribución BN \npara [40,7,32]")
plt.xticks(
    np.arange(min(geom_numeros), max(geom_numeros) + 1, 2.0)
)


plt.show()
```



Ejercicio 6. Una compañía petrolera realiza un estudio geológico que indica que un pozo petrolero

exploratorio debería tener un 20 % de posibilidades de encontrar petróleo.

- ¿Cuál es la probabilidad de que el primer pozo se produzca en el tercer pozo perforado?
- ¿Cuál es la probabilidad de que el tercer pozo se produzca en el séptimo pozo perforado?
- ¿Cuál es la media y la varianza del número de pozos que se deben perforar si la compañía petrolera quiere establecer tres pozos productores?

[10]:

```

```