

Homework

October 16, 2020

0.1 Homework:Roberto Valdivia García

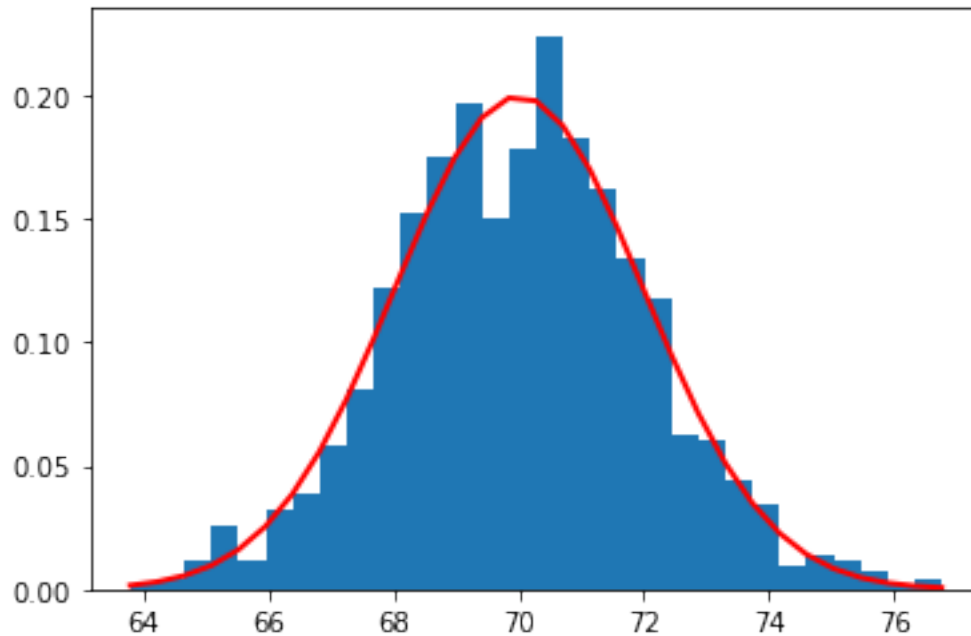
0.1.1 · Find the cumulative distribution function for the Gaussian Distributin, and reproduce the plots. Choose a random number between 0 and 1, and use the CDF to assign the corresponding value of H_0 . Generate as many as you want, and make the histogram of H_0 to verify you did right. Use a $\mu = 70$ and $\sigma = 2$.

-En la primera parte pesento cómo generar una distruibución aleatoria de número, la cual siga una distribución Normal, y la comparo con su función teorica.

```
[2]: import numpy as np
```

```
[3]: mu, sigma = 70, 2  
s = np.random.normal(mu, sigma, 1000)
```

```
[4]: import matplotlib.pyplot as plt  
  
count, bins, ignored = plt.hist(s, 30, density=True)  
plt.plot(bins,1/(sigma*np.sqrt(2*np.pi))*np.exp(-(bins-mu)**2/  
↪(2*sigma**2)),linewidth=2,color='r')  
plt.show()
```



-En esta segunda parte realizo la función de distribución acumulada(CDF), tanto la teórica como la experimental, basada en la distribución anterior. De manera resumida la información que nos da la CDF es la probabilidad de que al seleccionar cierto número de la distribución, éste se encuentre entre un valor $\leq x$

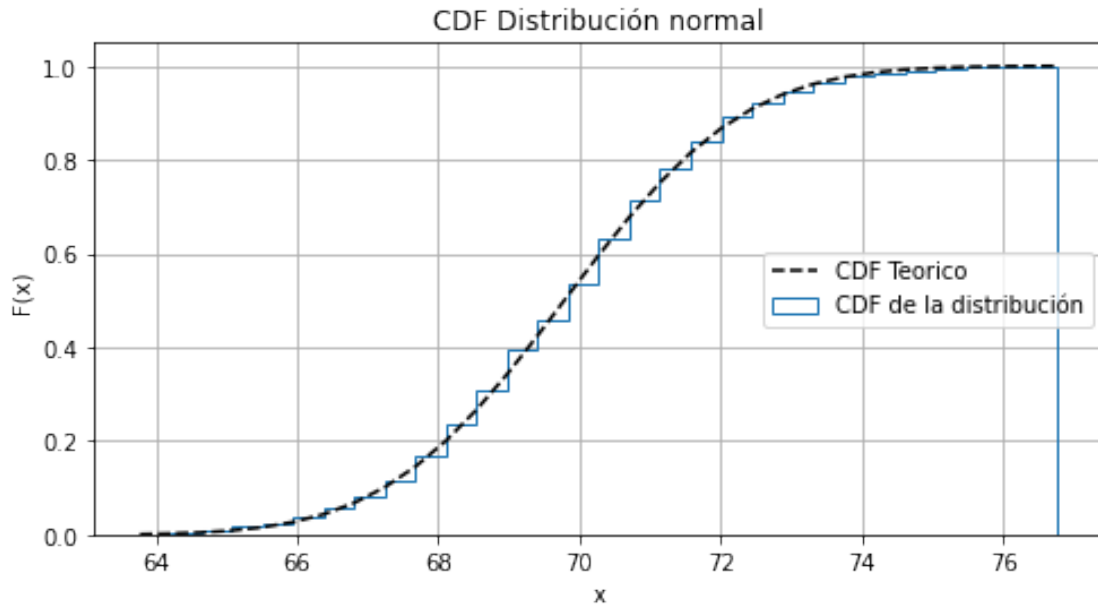
```
[5]: fig, ax = plt.subplots(figsize=(8, 4))
n, bins, patches = ax.hist(s, bins, density=True,
    ↪ histtype='step', cumulative=True, label='CDF de la distribución')

y = ((1 / (np.sqrt(2 * np.pi) * sigma)) *
      np.exp(-0.5 * (1 / sigma * (bins - mu)**2))
y = y.cumsum()
y /= y[-1]

ax.plot(bins, y, 'k--', linewidth=1.5, label='CDF Teorico')

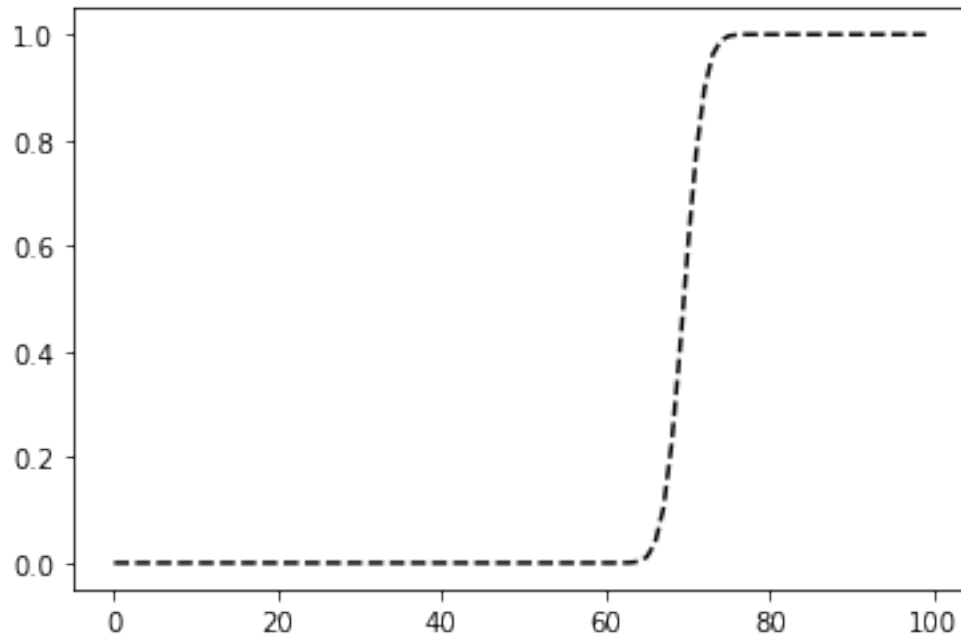
ax.grid(True)
ax.legend(loc='right')
ax.set_title('CDF Distribución normal')
ax.set_xlabel('x')
ax.set_ylabel('F(x)')

plt.show()
```

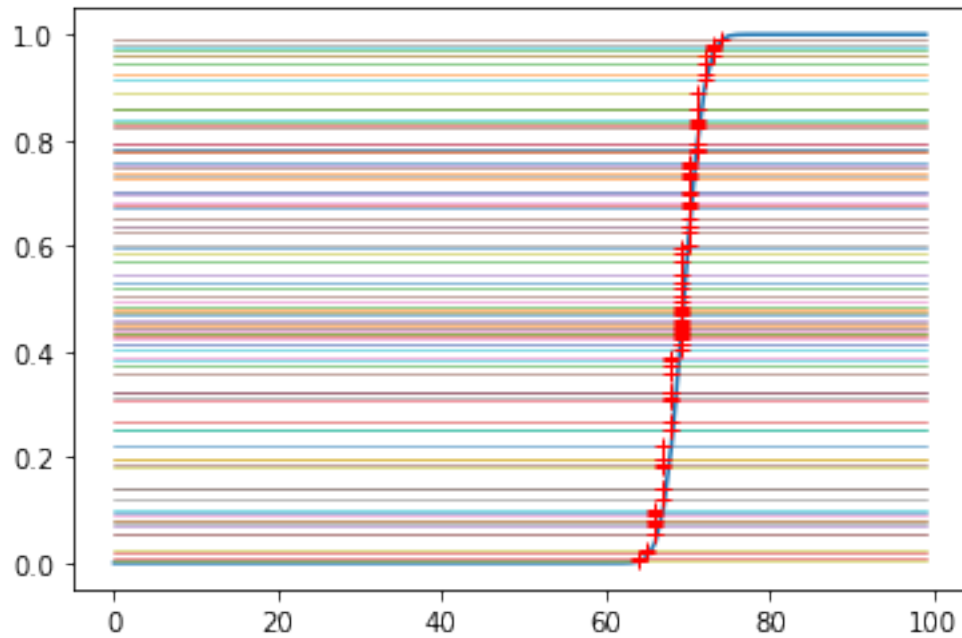


0.1.2 Para la parte sobre H0:

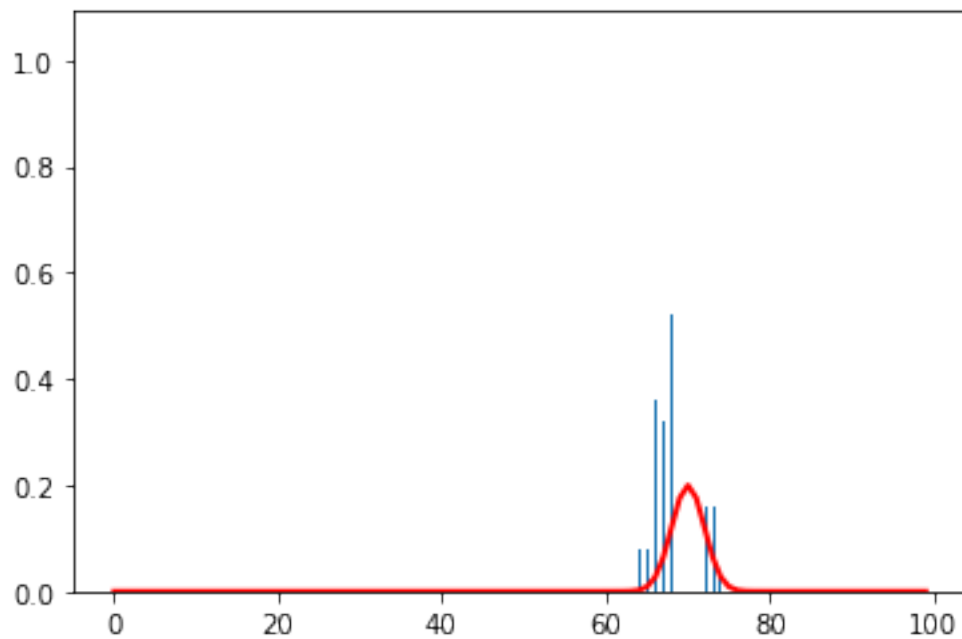
```
[6]: import numpy as np
import matplotlib.pyplot as plt
import random
mu=70
sigma=2
bins=40
H0=np.zeros(100, dtype=float)
x = np.arange(0, 100)
#definimos la CDF
y = ((1 / (np.sqrt(2 * np.pi) * sigma))*np.exp(-0.5 * (1 / sigma * (x -
→mu))**2))
y=y.cumsum()
y /= y[-1]
plt.plot(x, y, 'k--', linewidth=1.5, label='CDF Teorico')
plt.show()
```



```
[7]: plt.plot(x, y, '-')
for j in range(len(H0)):
    #encontramos un número aleatorio
    r=random.uniform(0, 1)
    f = np.array([r for i in range(len(x))])
    plt.plot(x, f, '-',linewidth=0.5)
    #encontramos el valor de H0 donde cruza el número aleatorio
    idx = np.argwhere(np.diff(np.sign(f - y))).flatten()
    plt.plot(x[idx], f[idx], 'r+',linewidth=0.5)
    H0[j]=idx[0]
plt.show()
```



```
[8]: plt.hist(H0, bins=40, density=True)
plt.plot(x, 1/(sigma*np.sqrt(2*np.pi))*np.exp(-(x-mu)**2/
↪ (2*sigma**2)),linewidth=2,color='r')
plt.show()
```

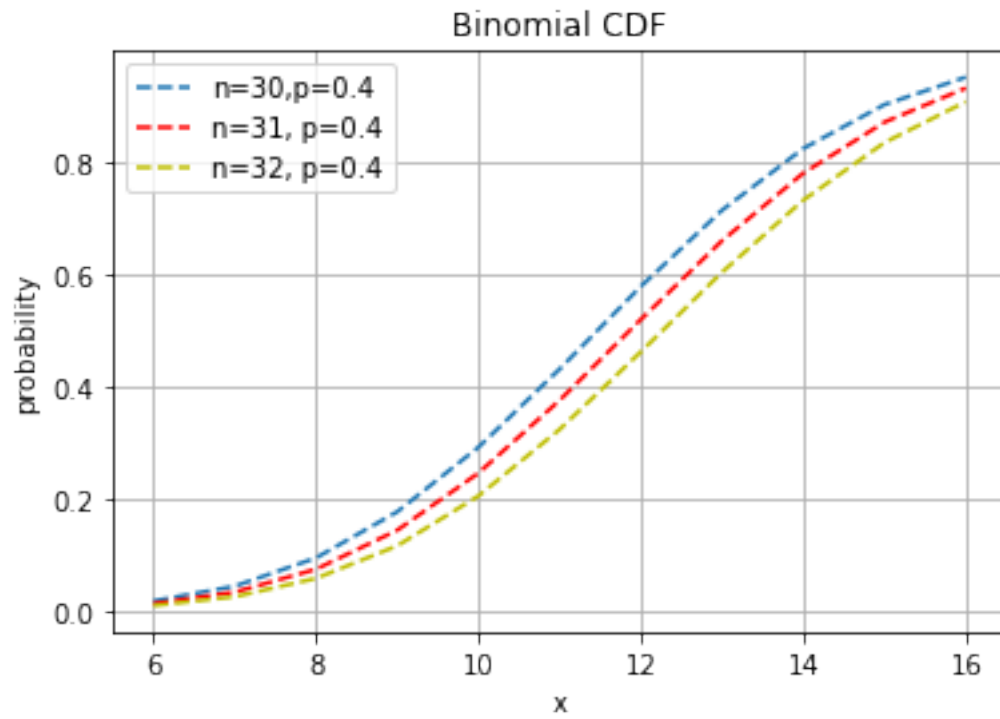


0.1.3 -For the Poisson and Binomial, etc... distribution mentioned, find the CDF mean and standard deviation. Plot both the PDF and CDF, for some different values of mean and sigma.

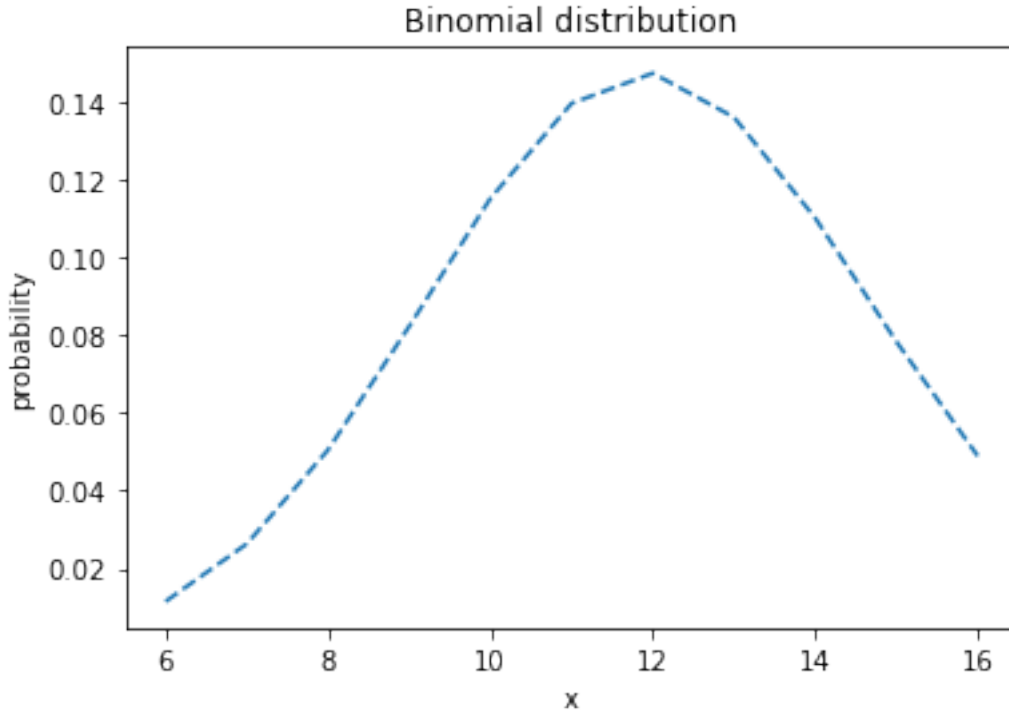
0.1.4 -Investigate about other useful distribution functions.

```
[9]: from scipy.stats import binom
N, p = 30, 0.4
N2, p2 = 31, 0.4
N3, p3 = 32, 0.4
mu = N*p
sigma = np.sqrt(N*p*(1-p))

binomial = binom(N, p)
binomial2 = binom(N2,p2)
binomial3 = binom(N3, p3)
x = np.arange(6,17,1)
CDF = binomial.cdf(x)
CDF2 = binomial2.cdf(x)
CDF3 = binomial3.cdf(x)
plt.plot(x, CDF, '--', label='n=30,p=0.4')
plt.plot(x,CDF2, '--', color='r', label='n=31, p=0.4')
plt.plot(x,CDF3, '--', color='y', label='n=32, p=0.4')
plt.legend()
plt.title('Binomial CDF')
plt.ylabel('probability')
plt.grid()
plt.xlabel('x')
plt.show()
```



```
[10]: fmp = binomial.pmf(x)
plt.plot(x, fmp, '--')
plt.title('Binomial distribution')
plt.ylabel('probability')
plt.xlabel('x')
plt.show()
```



Otras funciones de distribución relevantes, son la binomial, de Poisson, distribución uniforme, distribución exponencial.

0.1.5 -If $y = \phi(x) = \exp(x)$ and $p(x)=1$ for $0 \leq x \leq 1$ (a uniform distribution). What is the resultant distribution for y .

Partiendo de la PDF tenemos:

$$\mathbf{0.1.6} \quad p(y) = p_x [\Phi^{-1}(y)] \left| \frac{d\Phi^{-1}(y)}{dy} \right|$$

si consideramos que $p(x) = 1$ para $0 \leq x \leq 1$ y $y = \phi(x) = e^x$, podemos escribir entonces su inversa como $\phi^{-1}(y) = \ln y$ tal que

$$\mathbf{0.1.7} \quad p(y) = \frac{d(\ln y)}{dy} = \frac{1}{y}$$

de donde hicimos que $p_x(\phi^{-1}(y)) = 1$.

0.1.8 -Show explicitly that the linear least square method is derived from the minimization of the chi-square when the model is a straight line.

Tenemos Xi-Square:

$$\mathbf{0.1.9} \quad \chi^2 = \frac{\sum (y_i - y(x_i, \theta))^2}{\sigma^2}$$

y para encontrar los los punto singulares tenemos

$$\mathbf{0.1.10} \quad \frac{\partial \chi^2}{\partial \theta} = 0$$

La ecuación de la recta es

$$\mathbf{0.1.11} \quad y = ax + b$$

donde los parámetro libres θ son a y b. Entonces tenemos las derivadas

$$\mathbf{0.1.12} \quad \frac{\partial \chi^2}{\partial a} = \frac{2 \sum (y_i - y(x_i, \theta)) x_i}{\sigma^2} = 0$$

y

$$\mathbf{0.1.13} \quad \frac{\partial \chi^2}{\partial b} = \frac{2 \sum (y_i - y(x_i, \theta))}{\sigma^2} = 0$$

O bien, el sistema siguiente de ecuaciones

$$\mathbf{0.1.14} \quad \sum (y_i - ax_i - b) x_i = 0$$

$$\mathbf{0.1.15} \quad \sum (y_i - ax_i - b) = 0$$

despejando b de la segunda ecuación tenemos

$$\mathbf{0.1.16} \quad b = \frac{1}{n} (\sum y_i - a \sum x_i)$$

y de la otra ecuación tenemos

$$\mathbf{0.1.17} \quad \sum y_i x_i - a \sum x_i^2 - b \sum x_i = 0$$

sustituyendo el valor de b en la ecuación anterior

$$\mathbf{0.1.18} \quad \sum y_i x_i - a \sum x_i^2 - \frac{1}{n} (\sum y_i - a \sum x_i) \sum x_i = 0$$

$$\mathbf{0.1.19} \quad n \sum y_i x_i - an \sum x_i^2 - \sum y_i \sum x_i + a \sum x_i \sum x_i = 0$$

y despejando a:

$$\mathbf{0.1.20} \quad a = \frac{n \sum y_i x_i - \sum y_i \sum x_i}{n \sum x_i^2 - (\sum x_i)^2}$$

ahora esto en la expresión de b

$$\mathbf{0.1.21} \quad b = \frac{1}{n} \left(\sum y_i - \left[\frac{n \sum y_i x_i - \sum y_i \sum x_i}{n \sum x_i^2 - (\sum x_i)^2} \right] \sum x_i \right)$$

podemos reducir b como:

$$\mathbf{0.1.22} \quad b = \frac{\sum y_i \sum x_i^2 - \frac{1}{n} \sum y_i (\sum x_i)^2 - \sum y_i x_i \sum x_i + \frac{1}{n} \sum y_i (\sum x_i)^2}{n \sum x_i^2 - (\sum x_i)^2}$$

$$\mathbf{0.1.23} \quad b = \frac{\sum y_i \sum x_i^2 - \sum y_i x_i \sum x_i}{n \sum x_i^2 - (\sum x_i)^2}$$

por lo que finalmente obtenemos:

$$\mathbf{0.1.24} \quad a = \frac{n \sum y_i x_i - \sum y_i \sum x_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$\mathbf{0.1.25} \quad b = \frac{\sum y_i \sum x_i^2 - \sum y_i x_i}{n \sum x_i^2 - (\sum x_i)^2}$$

[]: