1. **A brief reminder on conditional distributions** [4 Points]

In this problem we recall the notion of **conditional distributions** *for two random variables X and Y. This is in particular relevant for the notion of sufficient statistics.*

- *If X and Y are discrete random variables with values in Ω_X and Ω_Y respectively, then for any $y \in \Omega_Y$ with $\mathbf{P}[Y = y] > 0$, the* conditional distribution of X given $Y = y$ *is characterized by the conditional probability mass function*

$$p_{X|Y=y}(x) = \frac{\mathbf{P}[X = x, Y = y]}{\mathbf{P}[Y = y]}, \qquad x \in \Omega_X.$$

This is the distribution of X under the probability measure $\mathbf{P}[\,\cdot\,|Y = y]$.

- *If X and Y are continuous real random variables, then for any $y \in \mathbb{R}$ with $f_Y(y) > 0$, the* conditional distribution of X given $Y = y$ *is characterized by the conditional probability density function*

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \qquad x \in \mathbb{R},$$

where $f_{X,Y}$ is the joint probability density function of (X, Y) and f_Y is the probability density function of Y.

(a) A fair coin is tossed 4 times. Let X denote the number of times that heads comes up and $Y = 1$ if heads comes up on the first toss and $Y = 0$ otherwise. Determine the conditional distribution of X given $Y = 1$ and the conditional distribution of Y given $X = 3$.

(a) $\mathbb{P}(X = i \mid Y = 1) = \binom{3}{i-1}\left(\frac{1}{2}\right)^3 \quad i \geq 1 \qquad \mathbb{P}(X = 0 \mid Y = 1) = 0$

$\mathbb{P}(Y = 1 \mid X = 3) = \frac{3}{4}$

$\mathbb{P}(Y = 0 \mid X = 3) = \frac{1}{4}$

(b) Consider jointly continuous random variables X, Y with density

$$f_{X,Y}(x, y) = 4ye^{-2y(x+1)}\mathbb{1}_{\{x,y>0\}}.$$

Determine the conditional probability density function $f_{X|Y=y}$. What is the conditional distribution of X given $Y = y$?

$$f_{X|Y=y} = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

$$f_Y(y) = \int_0^{+\infty} f_{X,Y}(x,y)\,dx = \int_0^{+\infty} 4y e^{-2y(x+1)}\,dx = -2\int_0^{+\infty} -2y\, e^{-2y(x+1)}\,dx$$

$$= -2 \; e^{-2y(x+1)}\Big|_0^{+\infty}$$

$$= -2(0 - e^{-2y})$$

$$= 2e^{-2y} \cdot \mathbb{1}_{\{y>0\}}$$

$$\Rightarrow f_{X|Y=y} = \frac{4ye^{-2y(x+1)}}{2e^{-2y}} \mathbb{1}_{\{x,y>0\}}.$$

$$= 2y e^{-2yx}$$

$\Rightarrow$ Exponential Distribution.

2. **Sufficient statistics** [4 Points]

(a) Let $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} Pois(\lambda)$ with unknown parameter $\lambda > 0$. Show that $T(\mathbf{X}) = \sum_{j=1}^{n} X_j$ is a sufficient statistic for λ

(i) ...directly, using the definition of sufficiency.

(ii) ...using the Neyman-characterization of sufficiency.

Hint: For (i), use the result of Problem 2 (a) on Problem set 1. You need to calculate $\mathbf{P}[X_1 = x_1, ..., X_n = x_n | T(\mathbf{X}) = t]$ for all possible values of $x_1, ..., x_n, t \in \mathbb{N}_0$.

(a)

(i) Using the definition of sufficiency.

$$\mathbb{P}_\lambda (X_1 = x_1, \cdots\cdots, X_n = x_n \mid T(x) = t)$$

$$= \begin{cases} 0 & \text{if } \sum x_i \neq t \\ \dfrac{\mathbb{P}[X_1 = x_1, \cdots, X_n = x_n, T = t]}{\mathbb{P}(T=t)} \end{cases}$$

$$= \frac{\prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}}{\frac{(n\lambda)^t e^{-n\lambda}}{t!}}$$

$$= \frac{t!}{n^t \prod_{i=1}^{n} x_j!}$$

Here we can see $\mathbb{P}[X_1 = x_1, \cdots, X_n = x_n | T = t]$ is independent of λ

which implies T is a sufficient statistic for X.

2° $\mathbb{P}(X_1=x_1, \cdots X_n=x_n) = \dfrac{\lambda^{\sum x_i} e^{-n\lambda}}{x_1! \cdots x_n!}$

$\mathbb{P}(X_1=x_1, \cdots, X_n=x_n) = \lambda^{\sum x_i} e^{-n\lambda} \cdot \dfrac{1}{x_1! \cdots x_n!}$

Let $g_\lambda(x) = \lambda^x \cdot e^{-n\lambda}$

$h(x_1, \cdots, x_n) = \dfrac{1}{x_1! \cdots\cdots x_n!}$

Therefore, $\mathbb{P}(X_1=x_1, \cdots\cdots X_n=x_n) = g_\lambda(T(x_1, \cdots, x_n)) \cdot h(x_1, \cdots, x_n)$

According to Neyman characterization Thm.

$T(\underline{X})$ is a sufficient stats for X.

(b) Let $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} Pareto(\lambda, a)$ with *known* $a > 0$ and unknown $\lambda > 0$, where we say that $X \sim Pareto(\lambda, a)$ if

$$f_X(x) = \frac{\lambda a^\lambda}{x^{\lambda+1}} \mathbb{1}_{(a,\infty)}(x).$$

Find a sufficient statistic $T(\mathbf{X}) \in \mathbb{R}$ for λ, using the Neyman-characterization.

(b) $$f_{X_1,\cdots,X_n}(x_1,\cdots x_n) = \frac{(\lambda a^\lambda)^n}{(\prod_{i=1}^n x_i)^{\lambda+1}} \prod \mathbb{1}_{(a,+\infty)}(x_i)$$

We let $T(x_1,\cdots x_n) = \prod_{i=1}^n x_i$

Let $g_\lambda(t) \frac{\lambda a^{\lambda n}}{t^{\lambda+1}}$

$h(x_1,\cdots x_n) = \prod_{i=1}^n \mathbb{1}_{(a,+\infty)}(x_i)$

and $f_{X_1,\cdots X_n}(x_1,\cdots,x_n) = g_\lambda(T(x_1,\cdots,x_n)) \cdot h(x_1,\cdots x_n)$

According to Neyman's Thm.

$T(x_1,\cdots x_n) = \prod_{i=1}^n x_i$ is a sufficient statistic.

3. **Method of moments** [4 Points]

(a) Consider $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{U}([0,\theta])$ with unknown $\theta > 0$. Calculate an estimator $\widehat{\theta}_n$ for θ based on the method of moments. Check this estimator for consistency and unbiasedness.

(a) $f = \frac{1}{\theta} \mathbb{1}_{[0,\theta]}$

$$\mathbb{E}[X] = \int_0^\theta x f(x)\, dx = \frac{1}{\theta} \cdot \frac{1}{2}\theta^2 = \frac{\theta}{2} \quad \Rightarrow \theta = 2\mathbb{E}[X]$$

$$\Rightarrow \widehat{\theta}_n = 2 \cdot \frac{\sum_{i=1}^n X_i}{n}$$

Consistency: By WLLN, $\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{\mathbb{P}} \mathbb{E}[X_1]$

$$\Rightarrow \widehat{\theta}_n \xrightarrow{\mathbb{P}} 2 \cdot \mathbb{E}[X_1] = \theta \quad \Rightarrow \text{Consistency.}$$

Unbiasedness $\mathbb{E}[\widehat{\theta}_n] = 2\frac{1}{n} \cdot \sum_{i=1}^n \mathbb{E}[X_1] = 2 \cdot \frac{1}{n} \cdot n \frac{\theta}{2} = \theta$

$\Rightarrow$ Unbiasedness

(b) Consider $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} Bin(N,p)$, where $\theta = (N,p) \in \mathbb{N} \times (0,1)$ is unknown (meaning both N and p are unknown). Determine an estimator $(\widehat{N}_n, \widehat{p}_n)$ for (N,p) based on the method of moments.

Hint: You need both $\mathbf{E}_{(N,p)}[X_1]$ and $\mathbf{E}_{(N,p)}[X_1^2]$.

(b)

$$\mathbb{E}_{(N,p)}[X_1] = Np$$

$$\mathbb{E}_{(N,p)}[X_1^2] = Var[X_1] + \mathbb{E}^2[X_1] = Np(1-p) + N^2p^2$$

$$\Rightarrow p = 1 - \frac{\mathbb{E}_{(N,p)}[X_1^2] - \mathbb{E}^2_{(N,p)}[X_1]}{\mathbb{E}_{(N,p)}[X_1]}$$

$$N = \frac{\mathbb{E}_{(N,p)}[X_1]}{1 - \frac{\mathbb{E}_{(N,p)}[X_1^2] - \mathbb{E}^2_{(N,p)}[X_1]}{\mathbb{E}_{(N,p)}[X_1]}}$$

Therefore.

$$\begin{pmatrix} \widehat{N}_n \\ \widehat{p}_n \end{pmatrix} = \begin{bmatrix} \dfrac{\frac{\sum_{i=1}^n X_i}{n}}{1 - \dfrac{\frac{\sum_{i=1}^n X_i^2}{n} - \frac{(\sum_{i=1}^n X_i)^2}{n^2}}{\frac{\sum_{i=1}^n X_i}{n}}} \\[2em] 1 - \dfrac{\frac{\sum_{i=1}^n X_i^2}{n} - \frac{(\sum_{i=1}^n X_i)^2}{n^2}}{\frac{\sum_{i=1}^n X_i}{n}} \end{bmatrix}$$

4. **(R exercise) The empirical distribution** [4 Points]

Note: Please provide your source code and images obtained with your solution.

Suppose that $X_1, \ldots, X_n$ are i.i.d. real random variables such that X_1 has cumulative distribution function F. Given a realization of these random variables, we can consider the *empirical cumulative distribution function*

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}.$$

(a) Explain why $\widehat{F}_n(x)$ fulfills the following:

(i) For every $x \in \mathbb{R}$, one has $\mathbf{E}[\widehat{F}_n(x)] = F(x)$ and $\mathrm{Var}[\widehat{F}_n(x)] = \frac{1}{n}F(x)(1-F(x))$.

(ii) For every $x \in \mathbb{R}$, one has $\widehat{F}_n(x) \xrightarrow[n\to\infty]{\mathbf{P}} F(x)$.

Hint: For any event A in some probability space, one has $\mathbf{E}[\mathbb{1}_A] = \mathbf{P}[A]$.

(a)

(i) $\mathbb{E}[\widehat{F}_n(x)] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{x_i \leq x\}\right]$

$= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{X_i \leq x\}\right] = \frac{1}{n}\cdot n\,\mathbb{E}[\mathbb{1}\{X_1 \leq x\}]$

$= \mathbb{P}[X \leq x] = F(x)$

$\mathrm{Var}[\widehat{F}_n(x)] = \mathbb{E}\left[\frac{1}{n^2}\left(\sum_{i=1}^{n}\mathbb{1}\{x_i \leq x\}\right)^2\right] - F^2(x)$

$= \frac{1}{n^2}\,\mathbb{E}\left[\sum_{i=1}^{n}\mathbb{1}\{x_i \leq x\}^2 + 2\sum_{j,k}\mathbb{1}\{x_k \leq x\}\mathbb{1}\{x_j \leq x\}\right] - F^2(x)$

$= \frac{1}{n^2}\cdot[n\cdot F(x)] + \frac{1}{n^2}2\frac{n(n-1)}{2}F^2(x) - F^2(x)$

$= \frac{1}{n}F(x) + \frac{n-1}{n}F^2(x) - F^2(x) = \frac{1}{n}F(x) - \frac{1}{n}F^2(x) = \frac{1}{n}F(x)(1-F(x))$

(ii)

$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{x_i \leq x\}$

By law of large numbers. $\widehat{F}_n(x) \xrightarrow{\mathbb{P}} \mathbb{E}[\mathbb{1}\{x_i \leq x\}]$

$= \mathbb{P}(X_i \leq x)$

$= F(x)$

(b) Use R to generate $n \in \{5, 20, 500\}$ samples of i.i.d. random variables $X_1, ..., X_n$ following a $\mathscr{U}([0,1])$-distribution or a $\mathscr{E}(3)$-distribution. Plot the empiricial cumulative distribution function together with the graph of the cumulative distribution function F_{X_1}. What do you observe?

Hint: The R-command `ecdf()` calculates the empirical distribution function of a vector and `plot(ecdf())` plots the respective graph.

See code file.

(c) Data on the magnitudes of earthquakes near Fiji are available from R, using the command `quakes`.[1] For help on this dataset type `?quakes`. Plot a histogram and the empirical cumulative distribution function for the *magnitudes*. Calculate the average $\overline{X}_n$ and sample variance S_n^2 for the magnitude.

Hint: The data set `quakes` is a data frame containing information on 5 observations (i.e. a table with 5 columns). To obtain a vector *only* containing the data in column 1, use `quakes[,4]`.

See Code file.

(d) Suppose it is suggested that the data for the magnitudes $X_1, \ldots, X_n$ can be modelled by a $\Gamma(\alpha, \beta)$ distribution. Find consistent estimators $\widehat{\alpha}_n$ and $\widehat{\beta}_n$ for α and β, and calculate the estimates using the data from `quakes`. Plot the cumulative distribution function of $\Gamma(\widehat{\alpha}_n, \widehat{\beta}_n)$ together with the empirical distribution function of the data. What do you observe?

Hint: Recall that for $X \sim \Gamma(\alpha, \beta)$, we have $\mathbf{E}[X] = \frac{\alpha}{\beta}$ and $\mathrm{Var}[X] = \frac{\alpha}{\beta^2}$.

$$\bar{X}_n \xrightarrow[\mathbb{P}]{n\to\infty} \mathbb{E}[X] = \frac{\alpha}{\beta}$$

$$S_n^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}\left((X_i^2) - \bar{X}_n^2\right)$$

By law of large numbers.

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i)^2 \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}(X_i^2) = \mu^2 + \sigma^2$$

By continuous mapping Thm.

$$\frac{1}{n-1}\sum_{i=1}^{n}(\bar{X}_n)^2 \xrightarrow[n\to\infty]{\mathbb{P}} \mu^2$$

$$\Rightarrow S_n^2 \xrightarrow[n\to\infty]{\mathbb{P}} \sigma^2 = Var[X] = \frac{\alpha}{\beta^2}$$

define.

$$\hat{\alpha}_n = \frac{(\bar{X}_n)^2}{S_n^2}$$

$$\hat{\beta}_n = \frac{(\bar{X}_n)}{S_n^2}$$

By properties of Convergence in $\mathbb{P}$.

$$\hat{\alpha}_n \xrightarrow{\mathbb{P}} \frac{\frac{\alpha^2}{\beta^2}}{\frac{\alpha}{\beta^2}} = \alpha$$

$$\hat{\beta}_n \xrightarrow{\mathbb{P}} \frac{\frac{\alpha}{\beta}}{\frac{\alpha}{\beta^2}} = \beta.$$

HW4_Q4

March 2, 2022

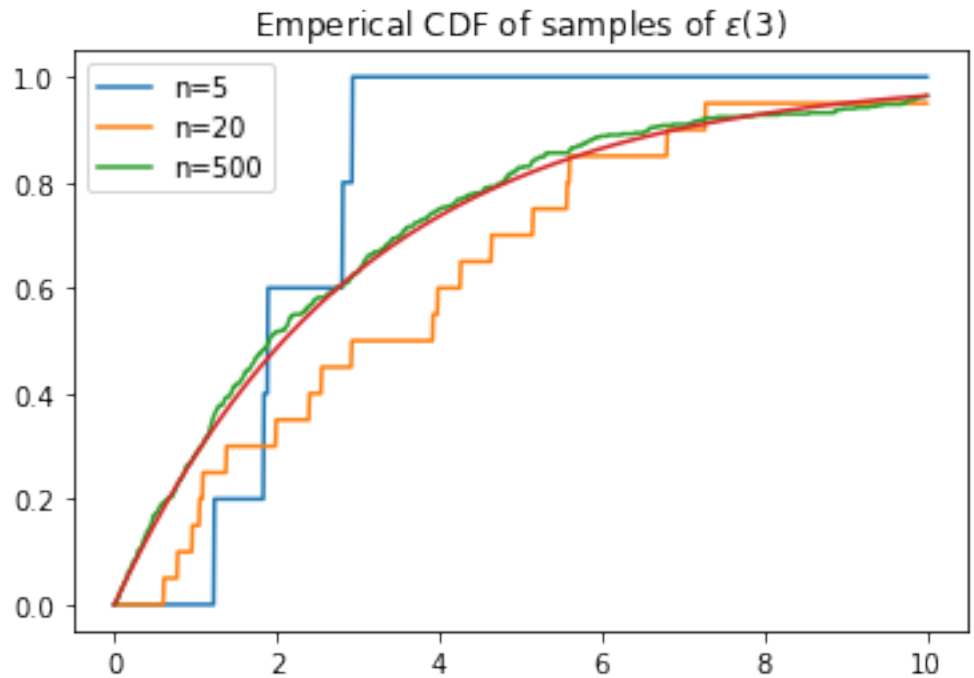
[1]:
```
import numpy as np
from scipy import stats
import matplotlib.pyplot as plt
import pandas as pd
from statsmodels.distributions.empirical_distribution import ECDF
```

0.1 4(b)

[2]:
```
# Set up the data array
expo5 = [np.random.exponential(3) for i in range(5)]
expo20 = [np.random.exponential(3) for i in range(20)]
expo500 = [np.random.exponential(3) for i in range(500)]
```

[3]:
```
# Use ECDF method to generate the emperical cdf
ecdf5 = ECDF(expo5)
ecdf20 = ECDF(expo20)
ecdf500 = ECDF(expo500)
```

[4]:
```
x_space = np.linspace(0, 10, 1000)
plt.plot(x_space, ecdf5(x_space), label="n=5")
plt.plot(x_space, ecdf20(x_space), label="n=20")
plt.plot(x_space, ecdf500(x_space), label="n=500")
plt.title("Emperical CDF of samples of " + r"$\epsilon(3)$")
plt.plot(x_space, stats.expon.cdf(x_space, scale=3))
plt.legend()
plt.show()
```



0.1.1 Obeservations

- The empirical CDF are more or less step functions
- The empirical CDFs simulates the actual CDF
- The bigger times of experiments, n, is, the better the simulation is. In the meantime, the empirical CDF is more smooth as n grows bigger.

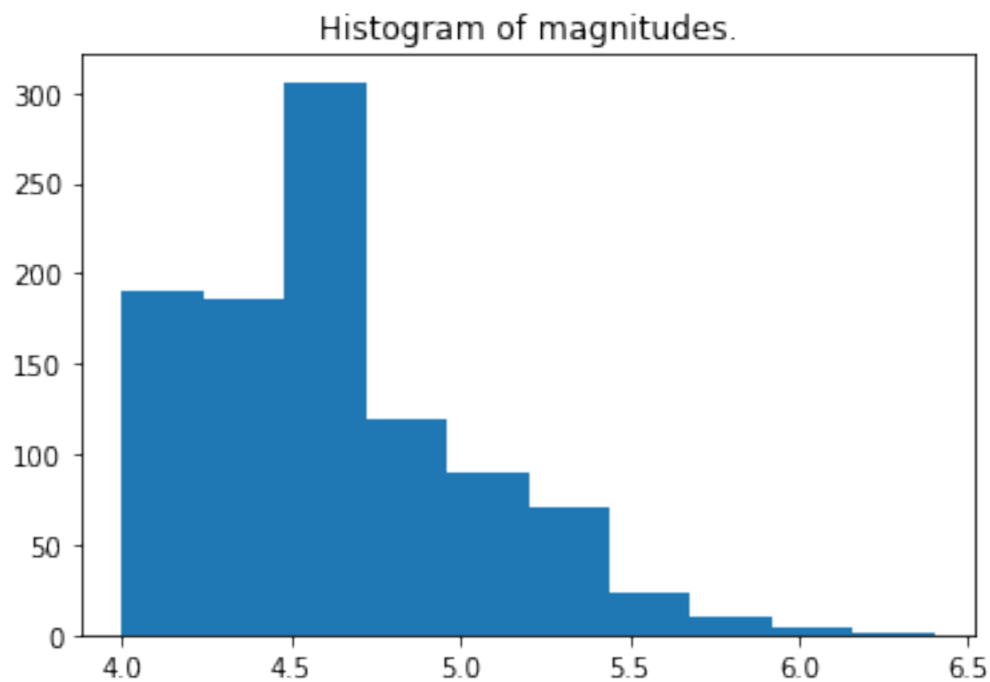
0.2 4(c)

[5]:
```
# use pandas to extract the data
df = pd.read_csv(
    "fijiquakes.dat",
    sep="\s+",
    skiprows=1,
    usecols= [1, 2, 3, 4, 5],
    names= ["lat", "long", "depth", "mag", "stations"]
)

# We are 'using the maginitudes
mag = df["mag"].to_numpy()

mean = mag.mean()
variance = mag.var(ddof=1)
print("Sample mean: {:.4f}".format(mean))
```

```
print("Sample variance: {:.4f}".format(variance))
```

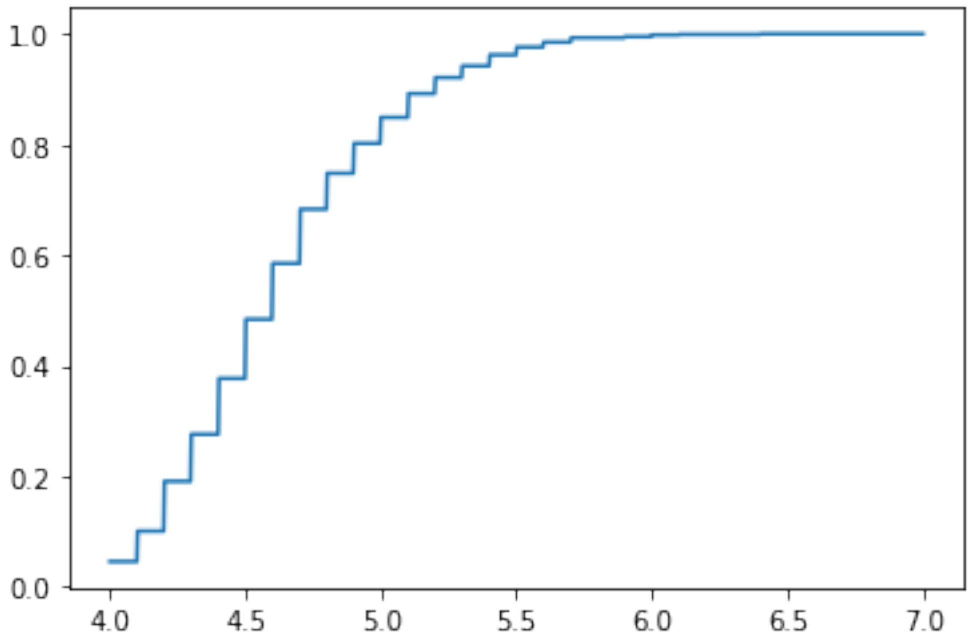
```
Sample mean: 4.6204
Sample variance: 0.1622
```

[6]:
```
# Visualization
plt.hist(mag)
plt.title("Histogram of magnitudes.")
plt.show()
```



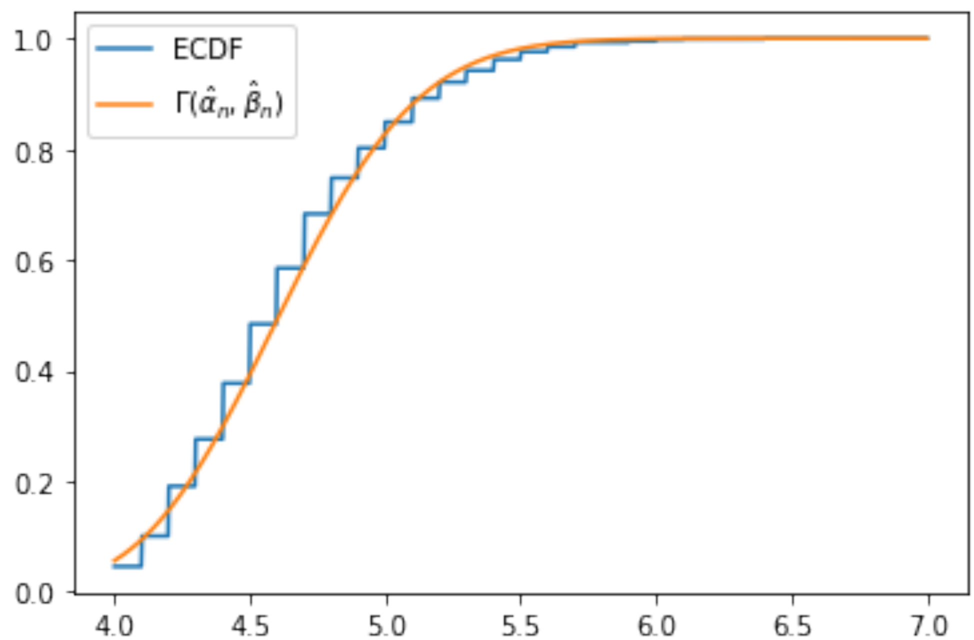
[7]:
```
x_space = np.linspace(4, 7, 1000)
ecdf = ECDF(mag)
plt.plot(x_space, ecdf(x_space))
```

[7]: `[<matplotlib.lines.Line2D at 0x7ff79f06c220>]`



[8]:
```
alpha_hat_n = mean**2 / variance
beta_hat_n = mean / variance
print("alphs_hat_n = ", alpha_hat_n)
print("beta_hat_n = ", beta_hat_n)
plt.plot(x_space, ecdf(x_space), label="ECDF")
plt.plot(x_space, stats.gamma.cdf(x_space, a=alpha_hat_n, scale=1/beta_hat_n),\
    label=r"$\Gamma(\hat{\alpha}_n, \hat{\beta}_n)$")

plt.legend()
plt.show()
```

```
alphs_hat_n =  131.59473491335265
beta_hat_n =  28.48124294722376
```



0.2.1 Observations

- The empirical CDF is more or less a step function
- It, however, simulates the actual CDF very well.