FSML Exam

May 30, 2021

Author: Daisuke Kuwabara

0.1 Exercise A

0.1.1 1. Give an estimator of μ and its expectation and variance

As random variable X have same expectation and variance, The Law of Large Number applies
$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow_{p;as} E[X_1]$
If we have $E[X^k] = g(\theta)$ then an estimator $\hat{\theta}$ of θ is solution of:
$g(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} X_i^k = \bar{X}_n$

As we have unknown mean μ and the expectation is the moment of order 1
$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i^k = \bar{X}_n$ This is an estimator of μ

Let $\hat{\theta}$ be an estimator of θ We say that $\hat{\theta}$ is unbiased of $\forall n \in W$
$E[\hat{\theta}] = \theta$
We say that $\hat{\theta}$ is asymptotically unbiased if
$E[\hat{\theta}] \rightarrow_{n\rightarrow+\infty} \theta$

As $E[\hat{\mu}] = E[\bar{X}_n] = E[\frac{1}{n}\sum_{i=1}^{n} X_i^k] = \frac{1}{n}\sum_{i=1}^{n} E[X_i]$
$E[\hat{\mu}] = \frac{n}{n}E[X_i] = E[X_i] = \mu$ It's expectation is X

As $V[\hat{\mu}] = V[\bar{X}_n] = V[\frac{1}{n}\sum_{i=1}^{n} X_i^k]$
$\frac{1}{n^2}V[\sum_{i=1}^{n} X_i]$ because of the independence of X_i $E[\hat{\mu}] = \frac{n}{n^2}V[X_1]$
$= \frac{V[X_1]}{n}$

0.1.2 2. Provide an unbiased estimator of σ^2

The centered moment of order k is
$E[(X - E[X])^k]$

As the variance is the centered moment of order 2,
$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$
To find the unbiased estimator, find the expectation of the estimator

$E[\hat{\sigma}^2] = E[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2]$
$= \frac{1}{n}E[\sum_{i=1}^{n}(X_i - \bar{X}_n)^2]$
$= \frac{1}{n}E[\sum X_i{}^2 - 2n\bar{X}_n\bar{X}_n + n\bar{X}_n{}^2]$
$= \frac{1}{n}E[\sum X_i^2 - n\bar{X}_n{}^2]$
$= \frac{1}{n}\sum_{i=1}^{n} E[X_i^2] - E[\bar{X}_n{}^2]$

$= (V[X_i] + (E[X_i])^2) - (V[\bar{X_n}] + (E[\bar{[X_n]}])^2)$
$= \theta - \frac{\theta}{n} + (E[X_i])^2 - (E[X_i])^2 = \frac{n-1}{n}\theta$

As $\frac{n-1}{n}\theta = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X_n})^2$
For unbiased estimator of $\sigma^2 = V(X)$
$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ This is the unbiased estimator of σ^2

0.2 Exercise B

0.2.1 1. Find the value of h to have a density function

$\int_0^3 hxdx + \int_3^5 3hdx + \int_5^8 h(8-x)dx = 1$
$= \frac{9h}{2} + 6h + \frac{9}{2}h = 1$
$= \frac{1}{15}$

0.2.2 2. Compute its expectation and its variance

$E(x) = \int xf(x)dx$
$= \int_0^3 x\frac{x}{15}dx + \int_3^5 x\frac{x}{5}dx + \int_5^8 x\frac{8-x}{15}dx$
$= \frac{134}{15}$
$Var(X) = E(X-\mu)^2$
As $Var(X) = E(X^2) - E(X)^2$ $E(X^2) = \int x^2 f(x)dx = \int_0^3 x^2\frac{x}{15}dx + \int_3^5 x^2\frac{x}{5}dx + \int_5^8 x^2\frac{8-x}{15}dx$
$\frac{79}{2} = 39.5$
$Var(X) = 39.5 - (\frac{134}{15})^2$

0.2.3 3. Write a code to simulate 1000 observations of the random variable associated to f.

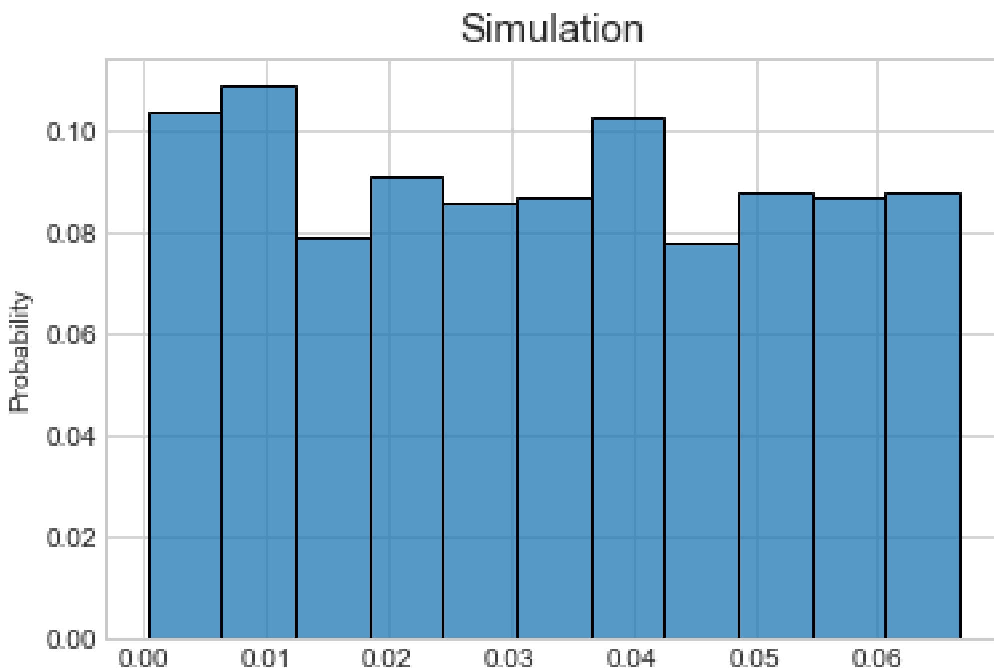
[2]:
```
def pdf(x):
    if 0<=x<=3:
        density= x/15
    elif 3<=x<=5:
        density=x/5
    elif 5<=x<=8:
        density=(8-x)/15
    return density
```

[3]:
```
import numpy as np
import pandas as pd
import seaborn as sns

import matplotlib.pyplot as plt
plt.style.use("seaborn-whitegrid")
np.random.seed(42)

simulation = [pdf(x) for x in np.random.rand(1000)]
#print(f"Simulated values: \n {np.array(simulation)}")
sns.histplot(simulation, stat="probability")
plt.title("Simulation", size=15)
```

[3]: Text(0.5, 1.0, 'Simulation')



0.2.4 4. Make a test to check that the produced data is distributed according to the give law

Kolmogorov-Smirnov test:

[4]:
```
from scipy.stats import chisquare
print(f"P-value: {chisquare(simulation)[1]}")
print("Null hypothesis: The distribution is unifrom") if␣
↪chisquare(simulation)[1]>0.05 else print("Reject Null hypothesis")
```

```
P-value: 1.0
Null hypothesis: The distribution is unifrom
```

Chi-square test:

0.3 Exercise C

A covid screening center can process 500 tests a week. Based on qualitative observation of previous week, they decide to allow 550 people to book for the coming week. The probability of a person to show up after booking is denoted p. Let X be the random variable equal to the number of persons that show up among the 550 possible

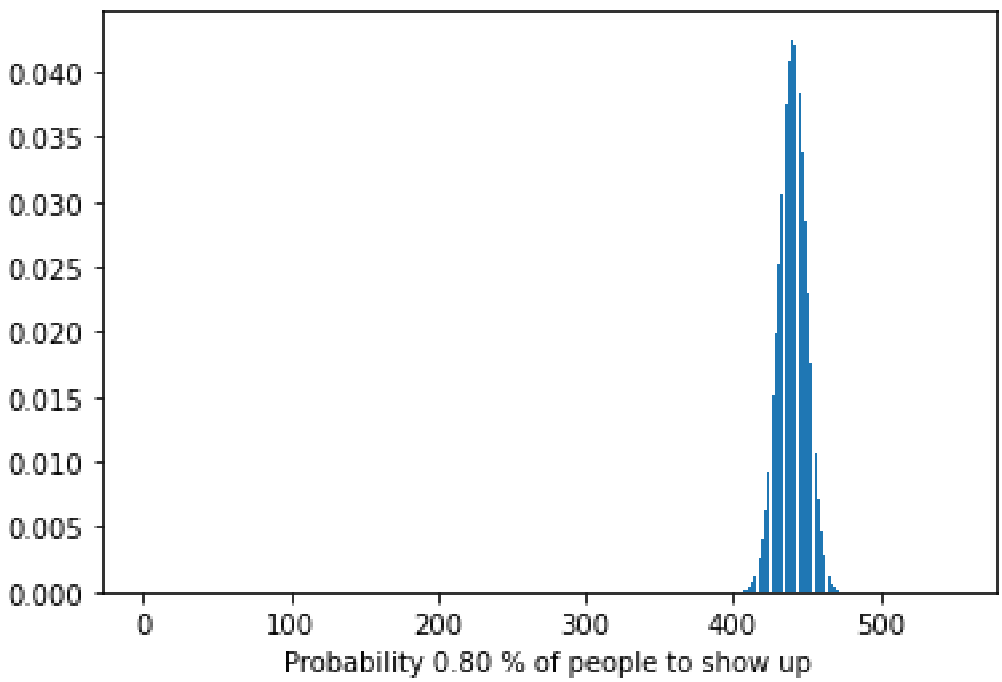
0.3.1 1. What is the distribution of X?

A. Binomial Distribution

[1]:
```python
from scipy.stats import binom
import numpy as np
import matplotlib.pyplot as plt
# setting the values
# of n and p
n = 550
p = np.random.uniform() # pseudo-random value ranged from 0 to 1

# defining list of r values
r_values = list(range(n + 1))
# list of pmf values
dist = [binom.pmf(r, n, p) for r in r_values]
# plotting the graph
plt.bar(r_values, dist)
plt.xlabel(f"Probability {p:.2f} % of people to show up")
```

[1]:
```
Text(0.5, 0, 'Probability 0.80 % of people to show up')
```



0.3.2 2. What is the maximum value of p such that $(X > 500) \leq 0.05$?

According to the setting, hypothesis testing can be formulated in the following way.
$H_0 : X \leq 500$
$H_1 : X > 500$

As the expectation of binomial distribution can be formulated as

$E[X_1] = P$
As $E[X_1] = P$ is the classical estimator for p
$\hat{P_n} = np$ can be obtained
To obtain the confidence interval, we use the **Central Limit Theorem**

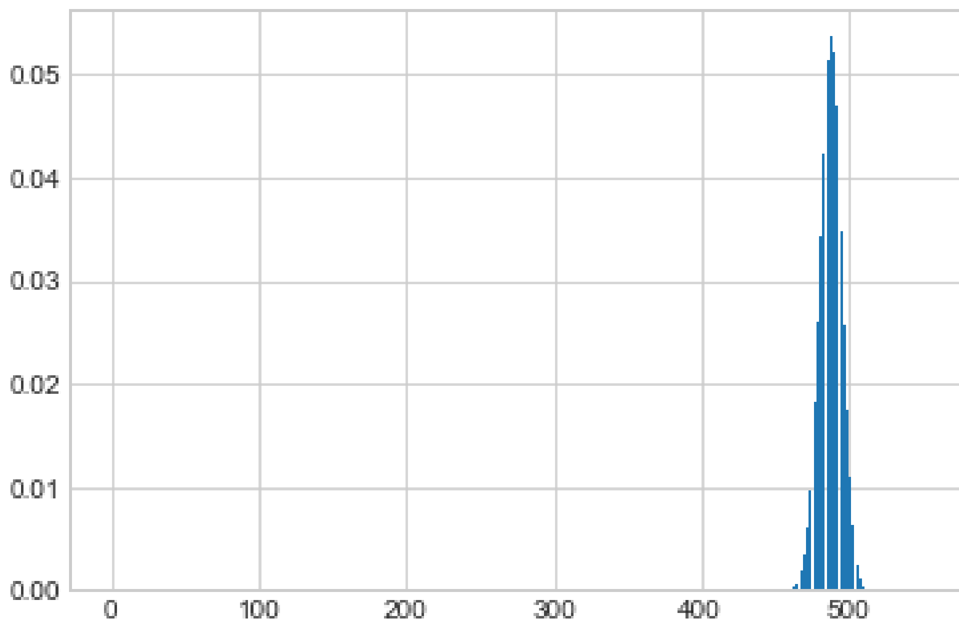
$\sqrt{n}(\frac{\bar{X_n}-\mu}{\sigma}) \to^d \mathcal{N}(0,1)$

$P(X > 500) \leq 0.05 = P(X \leq 500) \geq 0.95$
$\frac{X-np}{\sqrt{np(1-p)}} \leq \frac{500-np}{\sqrt{np(1-p)}}$

$P(\frac{500-np}{\sqrt{np(1-p)}} < \frac{X-np}{\sqrt{np(1-p)}}) \leq 0.05$
As $z_{0.05} = 1.645$
$P(\frac{500-np}{\sqrt{np(1-p)}}) = 1.645$
$500 - 550p = 1.645(\sqrt{550p - 550p^2})$
$500 = 550p + 1.645(\sqrt{550p - 550p^2})$
$p = 0.886873$
This is the maximum value of p which satisfies the condition

#Personal memo
unbiased estimator cannot give the precise estimation of p

[6]:
```
from scipy.stats import binom
import matplotlib.pyplot as plt
# setting the values
# of n and p
n = 550
p = 0.886873
# defining list of r values
r_values = list(range(n + 1))
# list of pmf values
dist = [binom.pmf(r, n, p) for r in r_values]
# plotting the graph
plt.bar(r_values, dist)
plt.show()
```

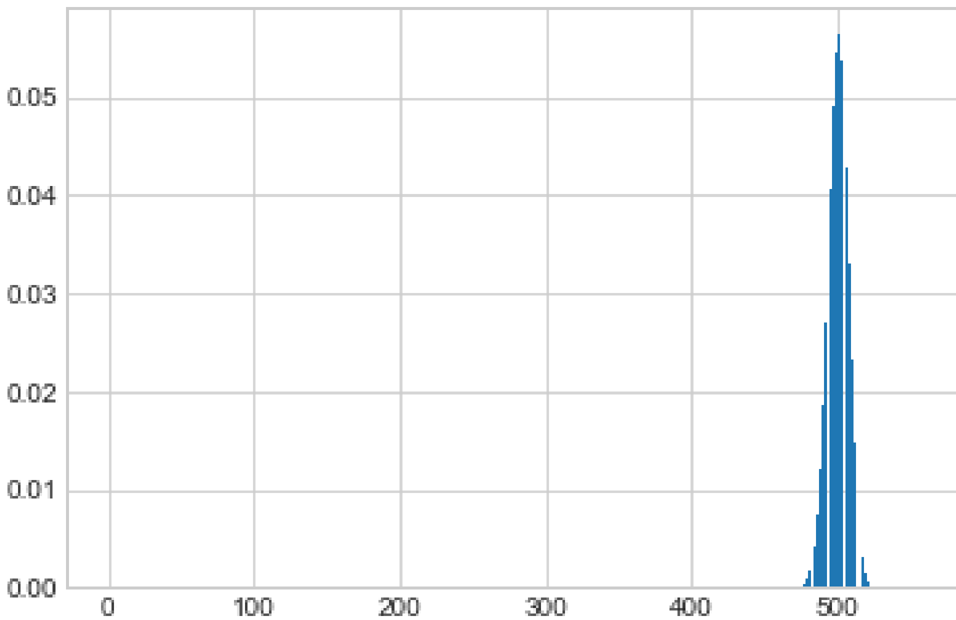


0.3.3 3. After a few weeks, they make an estimation of $p = 0.9$. How many people would you recommend them to allow for booking?

As $p = 0.9$ Set the mean of people who show up to the 500 so that we can maximize the capacity of the medicine

As binomial distribution has its property, $E[X] = np$
np = 500
0.9n = 500
$n = 556$
I reccomend them to allow around 556 people

[7]:
```
from scipy.stats import binom
import matplotlib.pyplot as plt
# setting the values
# of n and p
n = 556
p = 0.9
# defining list of r values
r_values = list(range(n + 1))
# list of pmf values
dist = [binom.pmf(r, n, p) for r in r_values]
# plotting the graph
plt.bar(r_values, dist)
plt.show()
```



0.4 Exercise D

With $x_0 > 0$, let us consider the function d given by:
$

$$d(t) = \begin{cases} a\frac{x_0^x}{t^{a+1}}, & \text{if } t \geq x_0. \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

$

0.4.1 1. Verify that d is a density function

To verify the probabilty density function, two important properties should be verified

1.$f(x) \geq 0$ for all x
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$

$\int_{-\infty}^{+\infty} f(x)dx$
$= \int_{x_0}^{+\infty} f(t)dt$
$= \int_{x_0}^{+\infty} a\frac{x_0^a}{t^{a+1}}dt$
$= ax_0^a[\frac{1}{-at^a}]_{x_0}^{+\infty}$
$= -\frac{a}{a}[(\frac{x_0}{t})^a]_{x_0}^{+\infty}$
As t is always grater than x_0
$= -1[0 - (\frac{x_0}{x_0})^a] = -1[0 - 1] = 1$

Then this satisfy the condition of density function

0.4.2 2. Let X be a random variable with density function d. Compute its expectation and variance when they exist

$E(x) = \int xf(x)dx$

$= \int_x^\infty ta\frac{x^a}{t^{a+1}}dt$
$ax_0^a[\frac{1}{(1-a)t^{a-1}}]$
$= \frac{a}{1-a}x_0[(\frac{x_0}{t})^{a-1}]_{x_0}^{+\infty}$
$= x_0\frac{a}{a-1}$

This is the case only when $a > 1$

As $Var(X) = E(X^2) - E(X)^2$

$E(X^2)$
$= \int_x^\infty t^2a\frac{x^a}{t^{a+1}}dt$
$= \int_x^\infty \frac{ax_0^a}{t^{a-1}}dt$
$= \frac{at^{2-a}x^a}{2-a}$
$= [\frac{a}{2-a}x_0^a\frac{1}{t^{a-2}}]_{x_0}^{+\infty}$
$= \frac{a}{a-2}x_0^2$

$Var(X)$
$= \frac{a}{a-2}x_0^2 - (\frac{ax_0}{a-1})^2$
$= \frac{ax_0^2}{(a-2)(a-1)^2}$

0.4.3 3. Compute the distribution function of X

The distribution function F for a continuous random variable X, with density function f is defined by
$\forall x \in \mathbb{R}, F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt$

As distribution function verifies - increasing function - continuous function - $\forall x \in \mathbb{R}, 0 \leq F(x) \leq 1$ - $lim_{x\to-\infty}F(x) = 0$ $lim_{x\to+\infty}F(x) = 1$ - $\forall a \leq b, P(X \in [a,b]) = F(b) - F(a)$ - $f = F'$ for all point x where F is differentiable

Let $x \in \mathbb{R}$, by definition $F(x) = \int_{-\infty}^x f(t)dt$

if $x \geq x_0$, we have

$F(x) = \int_{-\infty}^x f(t)dt$
$= \int_{-\infty}^{x_0} 0dt + \int_{x_0}^x a\frac{x_0^a}{t^{a+1}}dt$
$= \frac{ax_0^a}{-a}[\frac{1}{t^a}]_{x_0}^x$
$= -x_0^a[\frac{1}{x^a} - \frac{1}{x_0^a}] = 1 - (\frac{x_0}{x})^a$

$ =

$$F(x) = \begin{cases} 1 - (\frac{x_0}{x})^a, & \text{if } x \geq x_0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

$

0.4.4 4. Let denote by $Z_n = \mathbf{min}(X_1, ..., X_n)$. Determine its expectation.

$P(Z \geq x)$
$= P(min(X_1, ..., X_n) \geq x) = P(X_1 \geq x \cap X_2 \geq x \cap, ..., \cap X_n \geq x)$ As X is independently identically distributed, $\pi_{i=1}^{n} P(X_i \geq x)$
$\pi_{i=1}^{n}(1 - F(x))$
$(1 - F(x))^n$
$F_{Z_n}(x) = 1 - [1 - F(x)]^n$
$1 - [1 - (1 - (\frac{x_0}{x})^a)]^n$
$= 1 - (\frac{x_0}{x})^{an}$

PDF of Z_n can be defined,
$f_{Z_n}(x) = F'_{Zn}(x) = \frac{anx_0^{an}}{x^{an+1}}$
$E[Z_n]$
$= \int_{x_0}^{+\infty} anx_n^{an} \frac{x}{x^{an+1}} dx$
$E[\frac{an-1}{an} Z_n]$
$= X_0 \; \hat{x_0} = \frac{an-1}{an} Z_n$
$= \hat{x_0} = \frac{an-1}{an} Z_n$ is the uniased estimator of its expectation

0.4.5 5. Assuming that $a > 2$ is known, propose an estimator for x_0

\$

$$d(t) = \begin{cases} a\frac{x_0^x}{t^{a+1}}, & \text{if } t \geq x_0. \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

\$

According to the method of moments,
$\mu = E[X] = \frac{a}{a-1} x_0$
$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}_n$

With these properties, we can find the estimator of x_0 which is
$\hat{x}_0 = \frac{a-1}{a} \bar{X}_n$

As a is strictly greater than 2 and $lim_{a \to \pm\infty} \frac{-1+a}{a} = 1$
I conclude that an estimator for x_0 is
$\frac{1}{2} \bar{X}_n < x_0 \leq \bar{X}_n$