دانشکده‌ی علوم ریاضی

آمار و کاربرد ها	مدرس: دکتر محمد هادی مستفید
تمرین سری دو	
نام و نام‌خانوادگی : کژال باغستانی	*شماره دانشجویی:* ۴۰۱۱۰۰۰۷۱

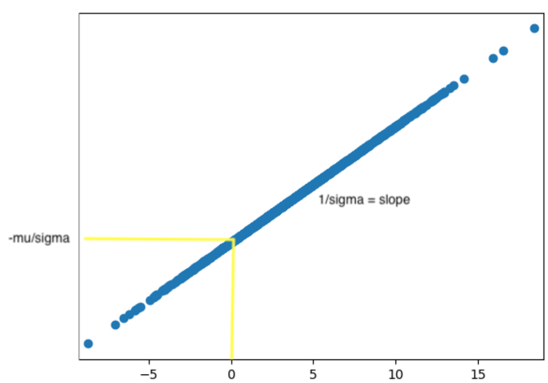
پرسش ۱

آ) فرض کنیم نقاط $x_1, x_2, ..., x_n$ را داریم به صورتی که x_i ها نقاط $quantile$ ها هستند. داریم

$$\frac{1}{n} = \int_a^b f(x)dx = P(a < X < b) = P(a < \sigma Y + \mu < b) = P(\frac{a-\mu}{\sigma} < Y < \frac{b-\mu}{\sigma}) = \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} f(y)dy$$

که $a = x_i$ و $b = x_{i+1}$.

پس نقطه متناظر x_i در نمودار توزیعX برابر است با $\frac{x_i-\mu}{\sigma}$ که این یعنی با رسم نمودار qqplot برای این دو متغیر نقاط روی یه خط صاف با شیب $\frac{1}{\sigma}$ و عرض از مبدا $-\frac{\mu}{\sigma}$ داریم.

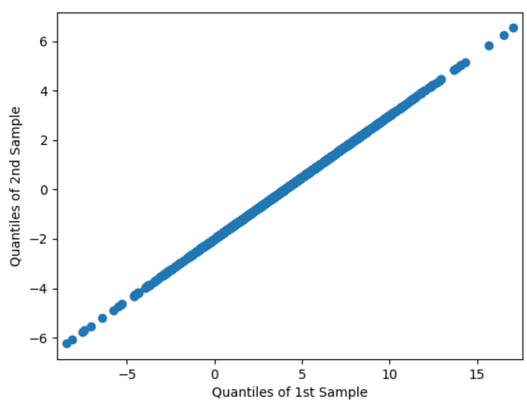


ب) کد پایتون:

src > hw2 > qqplot.py > ...

```python
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm

mu = 4
sigma = 2
X = np.random.normal(mu,sigma**2,1000)
Y = (X-mu)/sigma


fig, ax = plt.subplots()
sm.qqplot_2samples(X, Y,ax = ax)

plt.show()
```

خروجی:



پرسش ۲

آ) قانون قوی اعداد بزرگ میگوید:

$$\lim_{n\to\infty} P(\frac{X1+X2+X3+...+Xn}{n} = \mu) = 1$$

یا به بیان دیگر

$$\bar{X}_n \to \mu \quad a.s. \quad as \quad n \to \infty$$

و یکی از مثال های جالب آن این است که در یک بازی که احتمال برد قمار باز p است و احتمال برد کازینو $1-p$ که $p < \frac{1}{2}$ پس از بازی های زیاد هرچند شاید دفعه اول قمار باز برنده باشد ولی در نهایت کازینو برنده است زیرا

تعداد برد های کازینو نسبت به تعداد کل بازی ها به سمت $1-p$ می رود که بیشتر از $\frac{1}{2}$ است.

ب) تفاوت صورت قوی و ضعیف قانون اعداد بزرگ این است که در صورت قوی آن مجموعه پشامد هایی که در قانون صدق نمیکنند نسبت به مجموعه کل پشامد ها اندازه صفر است و این پیشامد ها ثابت و مشخص هستند در صورتی که در صورت ضیف آن مجموعه این پیشامد ها ثابت و مشخص نیستند و این قانون میگوید صرفا با زیاد کردن تعداد نمونه ها نسبت این پیشامد ها به مجموعه کل پیشامد ها کم و کم تر میشود اما درباره اینکه این پیشامد ها ثابت هستند تضمینی نمیدهد.

برهان:
می خواهیم نشان دهیم :

$$\forall \epsilon > 0 \qquad \lim_{n\to\infty} P(|\bar{X}-\mu| > \epsilon) = 0$$

با توجه به نامساوی مارکوف داریم

$$P(X > a) \leq \frac{E[X]}{a}$$

پس داریم

$$P((\bar{X}-\mu)^2 > a^2) \leq \frac{E[(\bar{X}-\mu)^2]}{a^2}$$

که این یعنی

$$P((\bar{X}-\mu)^2 > a^2) \leq \frac{var(\bar{X})}{a^2}$$

و با توجه به $i.i.d$ بودن Xi ها

$$var(\bar{X}) = \frac{\sigma^2}{n}$$

$$P(|\bar{X}-\mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon} \quad for \quad fixed \quad \epsilon \quad and \quad \sigma < \infty$$

پس:

$$0 \leq \lim_{n\to\infty} P(|\bar{X}-\mu| > \epsilon) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon} = 0$$

که ثابت میشود:

$$\forall \epsilon > 0 \qquad \lim_{n\to\infty} P(|\bar{X}-\mu| > \epsilon) = 0$$

پ) کد پایتون:

```
src > hw2 > LLN.py > ...
import numpy as np

X1 = np.random.uniform(1,100,100)
X2= np.random.uniform(1,100,1000)
X3 = np.random.uniform(1,100,10000)
X4 = np.random.uniform(1,100,100000)
X5 = np.random.uniform(1,100,1000000)
X6 = np.random.uniform(1,100,10000000)


mean1 = np.mean(X1)
mean2 = np.mean(X2)
mean3 = np.mean(X3)
mean4 = np.mean(X4)
mean5 = np.mean(X5)
mean6 = np.mean(X6)


print(mean1,"-> 100")
print(mean2,"-> 1000")
print(mean3,"-> 10000")
print(mean4,"-> 100000")
print(mean5,"-> 1000000")
print(mean6,"-> 10000000")
```

خروجی کد:

```
54.29280128531462 -> 100
49.519823405486044 -> 1000
50.689668124716846 -> 10000
50.48996036208122 -> 100000
50.48932856481535 -> 1000000
50.49718798076861 -> 10000000
```

پرسش ۳

آ) قضیه حد مرکزی: فرض کنیم دنباله‌ای از متغیرهای تصادفی مستقل و هم توزیع (i.i.d) را داشته باشیم، که به صورت $X_1, X_2, \dots, X_n$ نشان داده می‌شوند. قضیه حد مرکزی بیان می‌کند که توزیع متغیر تصادفی

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

به توزیع متغیر تصادفی نرمال استاندارد نزدیک و نزدیک‌تر می‌شود وقتی که مقدار n به خوبی زیاد شود. به صورت رسمی، این قضیه به صورت زیر بیان می‌شود:

$$lim_{n\to\infty} P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq x\right) = \Phi(x)$$

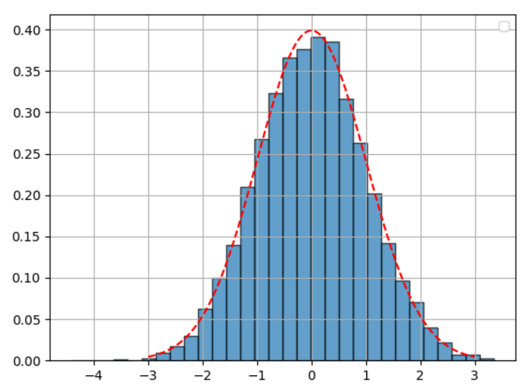
که در این‌جا $\Phi(x)$ تابع توزیع متغیر تصادفی نرمال استاندارد است.

ب) کد پایتون:

src > hw2 > CLT.py > ...

```python
import numpy as np
import matplotlib.pyplot as plt

n = 30# Sample size
num_samples = 10000
mu = 50
lb = 0
ub = 100

random_vars = np.random.uniform(lb, ub, (num_samples, n))
sample_means = np.mean(random_vars, axis=1)
# variance of uniform distribution 0 to 100 is 2500/3
variance = 2500/3
Z = (sample_means - mu) / (np.sqrt(variance) / np.sqrt(n))

plt.hist(Z, bins=30, density=True, edgecolor='black', alpha=0.7)

x = np.linspace(-3, 3, 1000)
standard_normal = (1 / np.sqrt(2 * np.pi)) * np.exp(-0.5 * x**2)
plt.plot(x, standard_normal, color='red', linestyle='dashed')

plt.legend()
plt.grid(True)
plt.show()
```

خروجی:



پرسش ۴

آ) تعریف استقلال برای دو متغیر تصادفی به صورت زیر است:

$$E[XY] = E[X]E[Y]$$

که برای این دومتغیر تصادفی اگر تعریف کنیم $Xi = flip(i) = H- > 1 : 0$ داریم:

$$E[X] = E[X1+X2] = E[X1]+E[X2] = \frac{1}{2}+\frac{1}{2} = E[X2]+E[X3] = E[X2+X3] = E[Y] = 1$$

حالا کافیست $E[XY]$ را محاسبه کنیم که داریم:

p	event	XY	Y	X
$\frac{1}{8}$	TTT	0	0	0
$\frac{1}{8}$	TTH	0	1	0
0	impossible	0	2	0
$\frac{1}{8}$	HTT	0	0	1
$\frac{2}{8}$	THT or HTH	1	1	1
$\frac{1}{8}$	THH	2	2	1
0	impossible	0	0	2
$\frac{1}{8}$	HHT	2	1	2
$\frac{1}{8}$	HHH	4	2	2

در نتیجه داریم:

$$E[XY] = \frac{2}{8} \times 1 + \frac{1}{8} \times 2 + \frac{1}{8} \times 2 + \frac{1}{8} \times 4 = 1.25$$

پس مستقل نیستند و کوواریانس آنها برابر است با:

$$cov(X, Y) = E[XY] - E[X]E[Y] = 1.25 - 1 = 0.25$$

ب) چون پرتاب های سکه مستقل هستند توزیع و میانگین و واریانس برای Y برابر با X است. پس به محاسبه این پارامتر ها برای X و Z می پردازم.

$$P(X = 0) = \frac{1}{4} \quad P(X = 1) = \frac{1}{2} \quad P(X = 2) = \frac{1}{4}$$

که عبارت بالا توزیع X و Y است.

$$E[X] = E[Y] = 1$$

$$var(X) = var(Y) = E[X^2] - E[X]^2 = 1.5 - 1 = 0.5$$

و برای Z داریم :

$$P(Z = 0) = \frac{1}{8} \quad P(Z = 1) = \frac{3}{8} \quad P(Z = 2) = \frac{3}{8} \quad P(Z = 3) = \frac{1}{8}$$

که عبارت بالا توزیع Z است.

$$E[Z] = 1.5$$

$$var(Z) = E[Z^2] - E[Z]^2 = 3 - 2.25 = 0.75$$

که به صورت کلی متغیر های X و Y توزیع دو جمله دارند با پارامتر های $(2, \frac{1}{2})$ و Z توزیع دو جمله ای با پارامتر $(3, \frac{1}{2})$ دارد.

پرسش ۵

فرض می‌کنیم X متغیر تصادفی هندسی است که نشان دهنده اولین دفعه ای است که سکه را پرتاب میکنیم و شیر می آید و دوبار بعد از آن نیز شیر می آید. به این ترتیب داریم:

$$P(X = i) = (1 - x)^{i-1}x$$

که در آن x احتمال رخ دادن سه شیر پشت سر هم است که برابر است با p^3 و میدانیم تعداد دفعات پرتاب برای رسیدن به اولین بار سه شیر پشت سر هم برابر است با :

$$X + 2$$

حالا میخواهیم متوسط تعداد دفعات پرتاب را محاسبه کنیم که یعنی $E[X + 2]$ که داریم :

$$E[X + 2] = 2 + E[X] = 2 + \frac{1}{p^3}$$

علت اینکه $E[X]$ برابر با $\frac{1}{p^3}$ شد این است که امید ریاضی متغیر تصادفی هندسی با پارامتر p برابر با $\frac{1}{p}$ است. در نتیجه متوسط تعداد دفعات پرتاب برای رسیدن به این منظور برابر با $\frac{1}{p^3} + 2$ بار است.