Homework 7

by Mher Movsisyan

Problem 1 (10 points)

Let X and Y be continuous random variables with joint density function with a constant c:

$f(x, y) = x + cy^2$ when $0 < x < 1$ and $0 < y < 1$

a) Find the value of c (2 points)

$$\iint_0^1 (x + cy^2)dxdy = 1$$

$$\int_0^1 \frac{2cy^2 + 1}{2} dy = 1$$

$$\frac{2c + 3}{6} = 1$$

$$c = 1.5$$

b) Calculate $f_{X|Y}(x, y)$ (4 points)

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

$$f_y(y) = \int_0^1 (x + 1.5y^2)dx = \frac{3y^2 + 1}{2}$$

$$f_{X|Y}(x, y) = \frac{2x + 3y^2}{3y^2 + 1}$$

c) Calculate $f_{X|Y=0.5}(x)$ (4 points)

$$f_{X|Y=0.5}(x, y) = \frac{2x + 3 * 0.25}{3 * 0.25 + 1} = \frac{2x + 0.75}{1.75}$$

Problem 2 (10 points)

Consider two random variables X and Y with joint PMF given in below:

X\Y	0	2	3
1	0.1	0.2	0.1
2	0.1	0.1	0.2
3	0.05	0.05	0.1

a) Calculate Cov(x, y) (5 points)

$$E(x) = 0.4 * 1 + 0.4 * 2 + 0.2 * 3 = 1.8$$

$$E(y) = 2 * 0.35 + 3 * 0.4 = 1.9$$

$$E(xy) = 2 * 1 * 0.2 + 2 * 2 * 0.1 + 2 * 3 * 0.05 + 3 * 1 * 0.1 + 3 * 2 * 0.2 + 3 * 3 * 0.1$$
$$= 0.4 + 0.4 + 0.3 + 0.3 + 1.2 + 0.9 = 3.5$$
$$Cov(x, y) = E(xy) - E(x)E(y) = 3.5 - 1.9 * 1.8 = 0.08$$

b) Calculate p(x, y) (Correlation) (5 points)

$$Var(x) = E(x^2) - E(x)^2 = (0.4 * 1 + 0.4 * 4 + 0.2 * 9) - 1.8^2 = 0.56$$
$$Var(x) = E(x^2) - E(x)^2 = (4 * 0.35 + 9 * 0.4) - 1.9^2 = 1.39$$
$$p(x, y) = \frac{Cov(x, y)}{\sqrt{Var(x)Var(y)}} = \frac{0.08}{\sqrt{1.39 * 0.56}} = 0.1028$$

Problem 3 (25 points)

In Armenia average height of Male is 171.5 cm and standard deviation is 8 cm. For females it is 159.2 cm and standard deviation is 8. We are taking female group of size 50 and the male group of size 70.

a) Approximate the probability that the average height in the group of females exceed 160. (5 points)

Using the central limit theorem:

$$P(\frac{X_1 + \ldots + X_{50} - 50 * 159.2}{8\sqrt{50}} \leqslant a) \approx \phi(a)$$
$$P(\frac{X_1 + \ldots + X_{50} - 50 * 159.2}{8\sqrt{50}} \geqslant a) \approx 1 - \phi(a)$$
$$P(\frac{X_1 + \ldots + X_{50}}{50} \geqslant \frac{8 * a}{\sqrt{50}} + 159.2) \approx 1 - \phi(a)$$
$$\frac{8 * a}{\sqrt{50}} + 159.2 = 160$$

$$a = 0.707107$$
$$P(\frac{X_1 + \ldots + X_{50}}{50} \geqslant 160) \approx 1 - \phi(0.707107) \approx 1 - 0.76 = 0.24$$

b) Approximate the probability that the average height in the men group exceeds that of the other group by over 15 cm. (10 points)

$$P(\frac{X_1 + \ldots + X_{70} - 70 * 171.5}{8\sqrt{70}} \geqslant 159.2 + 15)$$
$$P(\frac{X_1 + \ldots + X_{70}}{70} \geqslant \frac{a * 8}{\sqrt{70}} + 171.5) = 1 - \phi(a)$$
$$\frac{a * 8}{\sqrt{70}} + 171.5 = 159.2 + 15$$
$$a = 2.82$$
$$P(\frac{X_1 + \ldots + X_{70}}{70} \geqslant 159.2 + 15) = 1 - \phi(2.82) = 1 - 0.9976 = 0.0024$$

c) Approximate the probability that the difference between average height of female group and the average height of group of males is less than 10. (10 points)

$$Let : \mu_x = \frac{X_1 + \ldots + X_{70}}{70}$$
$$Let : \mu_y = \frac{Y_1 + \ldots + Y_{50}}{50}$$
$$P(|\mu_x - \mu_y| \leqslant 10) = P(-10 \leqslant \mu_x - \mu_y \leqslant 10)$$
$$P(-10 \leqslant \mu_x - \mu_y \leqslant 10) = P(\mu_x - \mu_y \leqslant 10) - P(-10 \leqslant \mu_x - \mu_y)$$

Let's try computing it

In [54]:
```
# Import necessary library, set a seed for the random generator
import numpy as np
import  seaborn as sns
from scipy.stats import norm
import matplotlib.pyplot as plt
from matplotlib.lines import Line2D
%matplotlib inline
np.random.seed(173)

# Create a place to store the difference of the means
mean_diffs = np.zeros(10000)

for i in range(10000):
    # Take two random samples, doesn't matter much which distribution we draw
from
    X = np.random.normal(171.5, 8, 70)
    Y = np.random.normal(159.2, 8, 50)

    # Compute and store the difference of means
    mean_diffs[i] = np.mean(X) - np.mean(Y)

mean_diffs.sort()
print("The mean of M is", np.mean(mean_diffs), "and the standard deviation i
s", np.std(mean_diffs))
```

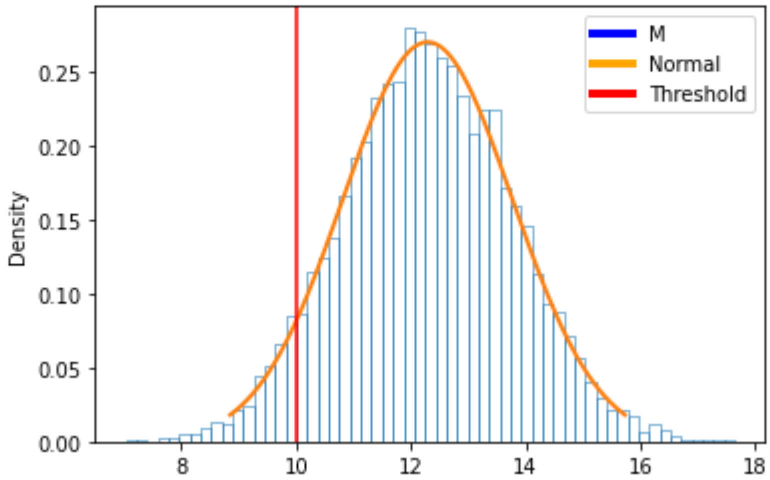
```
The mean of M is 12.292390462594723 and the standard deviation is 1.4774510
383987165
```

From the central limit theorem, we know that the means of X and Y are normally distributed, let's see if the new variable M (mean differences) is normally distributed or not.

In [59]:
```
x = np.arange(
    norm.ppf(0.01, np.mean(mean_diffs), np.std(mean_diffs)),
    norm.ppf(0.99, np.mean(mean_diffs), np.std(mean_diffs)),
    step = 0.001)

fig, ax = plt.subplots()
sns.histplot(x = mean_diffs, ax=ax, fill=False, stat="density")
sns.lineplot(x = x, y = norm.pdf(x, np.mean(mean_diffs), np.std(mean_diffs)),
linewidth=2, ax=ax)
plt.axvline(x=10, color="red")

custom_lines = [Line2D([0], [0], color="Blue", lw=4),
                Line2D([0], [0], color="Orange", lw=4),
                Line2D([0], [0], color="Red", lw=4)]
ax.legend(custom_lines, ["M", "Normal", "Threshold"]);
```



The new variable M is indeed normally distributed, now let's compute the probability of M being less than 10

In [61]:
```
np.sum(mean_diffs < 10) / len(mean_diffs)
```

Out[61]: 0.0624

Problem 4 (15 points)

The number of exercises a student is doing during the week is a random variable with mean 45.

a) What can be said about the probability that this week's student will do at least 115 exercises? (7 points)

Since we only have the mean, we can use Markov's inequality

$$P(X \geqslant 115) \leqslant \frac{45}{115} \approx 0.39$$

b) If the variance of a week's exercises is known to equal 100, can we obtain a better bound for (a)? (8 points)

Now we can use Chebyshev's almighty inequality.

$$P(X \geqslant 115) = 0.5 * P(|X - 45| \geqslant 70) \leqslant \frac{100 * 0.5}{70^2}$$

$$P(X \geqslant 115) \leqslant \frac{1}{98} \approx 0.01$$

Problem 5 (15 points)

Suppose someone gives you a die and claims that this die is biased (that it gave 6 only 15% of the time). You decide to test the die by yourself and draw the die 70 times. Find a probability that there will be less than 9 six in the all 70 trials.

This is a simple binomial problem which can be approximated by a normal variable,

$$p = 0.15, n = 70$$
$$\mu = np = 70 * 0.15 = 10.5, \sigma^2 = 10.5 * (0.85) = 8.925$$
$$P(X < 9) = \phi(\frac{9 - 10.5}{\sqrt{8.925}}) = \phi(-0.5) = 1 - \phi(0.5)$$
$$P(X < 9) \approx 0.31$$

Problem 6 (25 points)

In The university a distribution of the final exam grade of students has a mean of 82 (in a range from 0 to 100) and a variance 144. You randomly select 40 students from the population of students.

a) Find the probability that the sample mean is between 78 and 90. (10 points)

$$Let \mu = \frac{X_1 + \ldots + X_{40}}{40}$$
$$P(78 \leqslant \mu \leqslant 90) = ?$$
$$P(\mu \leqslant 90) = P(\frac{X_1 + \ldots + X_{40} - 40 * 82}{12\sqrt{40}} \leqslant a_1) \approx \phi(a_1)$$
$$P(\frac{X_1 + \ldots + X_{40}}{40} \leqslant \frac{12a_1\sqrt{40} + 40 * 82}{40})$$
$$\frac{12a_1\sqrt{40} + 40 * 82}{40} = 90$$
$$a_1 = \frac{4\sqrt{10}}{3} \approx 4.22$$
$$P(\mu \leqslant 90) \approx 1$$
$$P(\mu \geqslant 78) = 1 - P(\mu \leqslant 78)$$
$$\frac{12a_2\sqrt{40} + 40 * 82}{40} = 78$$
$$a_2 = -\frac{2\sqrt{10}}{3} \approx -2.11$$
$$P(78 \leqslant \mu \leqslant 90) = P(\mu \geqslant 78) = 1 - \phi(-2.11) = 1 - (1 - \phi(2.11)) =$$
$$= \phi(2.11) \approx 0.9826$$

b) Find the value of final exam score that is two standard deviations above the population mean, 82, of the sample mean. (5 points)

$$82 + \frac{2 * 12}{\sqrt{40}} \approx 85.79$$