EE 3070 Statistics

Practice #2

No need to turn it in. Strongly recommend to practice these questions.
You are allowed to use computers, software packages, and online tools.

1. Let X and Y have a bivariate normal distribution with parameters $\mu_1 = 3, \mu_2 = 1, \sigma_1^2 = 16, \sigma_2^2 = 25$, and $\rho = \frac{3}{5}$. Determine the following probabilities:
(a) $P(3 < Y < 8)$
(b) $P(3 < Y < 8 \mid X = 7)$
(c) $P(-3 < X < 3)$
(d) $P(-3 < X < 3 \mid Y = -4)$

Solution:
(a) To find $P(3 < Y < 8)$, we need to standardize the variables using the means and variances provided.

$$\begin{aligned} P(3 < Y < 8) &= P(\frac{3-1}{5} < \frac{Y-1}{5} < \frac{8-1}{5}) \\ &= P(\frac{2}{5} < z < \frac{7}{5}) = \phi(\frac{2}{5}) - \phi(\frac{7}{5}) \\ &= 0.9192 - 0.6554 \\ &= 0.2638 \end{aligned}$$

(b) To find $P(3 < Y < 8 \mid X = 7)$, a conditional distribution of Y given $X = 7$, which is still a normal distribution, we need to compute the mean and variance.

$$E(Y \big| X = x) = \mu_Y + \rho \cdot \frac{\sigma_Y}{\sigma_X}(x - \mu_X) = \mu_2 + \rho \cdot \frac{\sigma_2}{\sigma_1}(7 - \mu_1) = 1 + \frac{3}{5} \cdot \frac{5}{4} \cdot (7-3) = 1 + 3 = 4$$

$$\sigma_{Y \big| X=x} = \sigma_Y \cdot \sqrt{1-\rho^2} = \sigma_2 \cdot \sqrt{1 - (\frac{3}{5})^2} = 5 \cdot \sqrt{\frac{16}{25}} = 5 \cdot \frac{4}{5} = 4$$

Recall that

$$P(a < Y < b \mid X = c) = P(\frac{y - \mu_{Y|X=c}}{\sqrt{\sigma^2_{Y|X=c}}}) < P(\frac{b - \mu_{Y|X=c}}{\sqrt{\sigma^2_{Y|X=c}}}) - P(\frac{y - \mu_{Y|X=c}}{\sqrt{\sigma^2_{Y|X=c}}}) < P(\frac{a - \mu_{Y|X=c}}{\sqrt{\sigma^2_{Y|X=c}}})$$

Therefore

$$\begin{aligned} P(3 < Y < 8 \mid X = 7) &= P(z < \frac{8-4}{4}) < P(z - \frac{3-4}{4}) \\ &= P(z < 1) - P(z < -0.25) \\ &= 0.8413 - 0.4013 \\ &= 0.4400 \end{aligned}$$

(c) To find $P(-3 < X < 3)$, again, we need to standardize the variables using the means and variances provided.

$$\begin{aligned}
P(-3 < X < 3) &= P(\frac{-3-3}{4} < \frac{X-3}{4} < \frac{3-3}{4}) \\
&= P(\frac{-6}{4} < z < 0) = \phi(0) - \phi(-\frac{3}{2}) = 0.5 - 0.0668 \\
&= 0.4332
\end{aligned}$$

(d) To find $P(-3 < X < 3 \mid Y = -4)$, a conditional distribution of X given $Y = -4$, which is still a normal distribution, and again, we need to compute the mean and variance.

$$\begin{aligned}
E(X\big|Y = y) &= \mu_X + \rho \cdot \frac{\sigma_X}{\sigma_Y}(y - \mu_Y) = \mu_1 + \rho \cdot \frac{\sigma_1}{\sigma_2}(-4 - \mu_2) \\
&= 3 + \frac{3}{5} \cdot \frac{4}{5} \cdot (-4 - 1) = 3 - \frac{12}{5} = \frac{3}{5} = 0.6 \\
\sigma_{X\big|Y=y} &= \sigma_X \cdot \sqrt{1-\rho^2} = \sigma_1 \cdot \sqrt{1 - (\frac{3}{5})^2} = 4 \cdot \frac{4}{5} = \frac{16}{5} = 3.2
\end{aligned}$$

Again, with

$$P(a < X < b \mid Y = c) = P(\frac{x - \mu_{X|Y=c}}{\sqrt{\sigma^2_{X|Y=c}}}) < P(\frac{b - \mu_{X|Y=c}}{\sqrt{\sigma^2_{X|Y=c}}}) - P(\frac{x - \mu_{X|Y=c}}{\sqrt{\sigma^2_{X|Y=c}}}) < P(\frac{a - \mu_{X|Y=c}}{\sqrt{\sigma^2_{X|Y=c}}})$$

Therefore

$$\begin{aligned}
P(-3 < X < 3 \mid Y = -4) &= P(z < \frac{3-0.6}{3.2}) - P(z < \frac{-3-0.6}{3.2}) \\
&= P(z < 0.75) - P(z < -1.25) = 0.7734 - 0.1303 \\
&= 0.6431
\end{aligned}$$

■

2. Let X and Y have a bivariate normal distribution with parameters $\mu_1 = 5$, $\mu_2 = 10$, $\sigma_1^2 = 1$, $\sigma_2^2 = 25$, and $\rho > 0$. If $P(4 < Y < 16 \mid X = 5) = 0.954$, determine ρ.

Solution:
Since

$$\begin{aligned}
&E(Y\big|X = x) = \mu_Y + \rho \cdot \frac{\sigma_Y}{\sigma_X}(x - \mu_X) = \mu_2 + \rho \cdot \frac{\sigma_2}{\sigma_1}(x - \mu_1) = 10 + \rho \cdot \frac{10}{5} \cdot (5-5) = 10 \\
&\sigma_{Y\big|X=x} = \sigma_Y \cdot \sqrt{1-\rho^2} = \sigma_2 \cdot \sqrt{1-\rho^2} = 5 \cdot \sqrt{1-\rho^2} \\
\Rightarrow &P(4 < Y < 16 \mid X = 5) = P(z < \frac{16-10}{5 \cdot \sqrt{1-\rho^2}}) < P(z - \frac{4-10}{5 \cdot \sqrt{1-\rho^2}}) = 0.954 = \phi(2) - \phi(-2)
\end{aligned}$$

This probability requires that (from Example 3.4.4.)

$$\frac{16-10}{5\sqrt{1-\rho^2}} = 2, \quad \frac{3}{5} = \sqrt{1-\rho^2}, \quad \frac{9}{25} = 1 - \rho^2, \quad \text{and therefore} \quad \rho = \frac{4}{5}.$$

■

3. Let T have a t-distribution with $r > 4$ degrees of freedom. Use expression (3.6.4) to determine the kurtosis of T. See Exercise 1.9.15 for the definition of kurtosis.

Solution:
Equation 3.6.4

$$E(T^k) = E(W^k)\frac{2^{-\frac{k}{2}} \cdot \Gamma(\frac{r}{2} - \frac{k}{2})}{\Gamma(\frac{r}{2}) \cdot r^{-\frac{k}{2}}}, \quad \text{for } k < r.$$

From Exercise 1.9.15, we have the definition of kurtosis

$$\begin{aligned}
K &= \frac{E((T-\mu)^4)}{\sigma^4} = \frac{E(T^4)}{\mathrm{Var}(T)^4} \\
&= E(W^4) \cdot \frac{2^{-\frac{4}{2}} \cdot \Gamma(\frac{r}{2} - \frac{4}{2})}{\Gamma(\frac{r}{2}) \cdot r^{-\frac{4}{2}}} \\
&= \frac{1}{4} \cdot \frac{E(W^4) \cdot \Gamma(\frac{r}{2} - 2)}{\frac{r^2}{(r-2)^2} \cdot \Gamma(\frac{r}{2}) \cdot r^{-2}} \\
&= \frac{1}{4} \cdot \frac{3 \cdot \Gamma(\frac{r}{2} - 2)}{\Gamma(\frac{r}{2})} \cdot (r-2)^2 \\
&= \frac{3 \cdot (r-2)^2}{4} \cdot \frac{1}{(\frac{r}{2} - 1) \cdot (\frac{r}{2} - 2)} \\
&= \frac{3 \cdot (r-2)^2}{4} \cdot \frac{4}{(r-2)(r-4)} \\
&= \frac{3 \cdot (r-2)}{r-4}
\end{aligned}$$

Note: how we know that $E(W^4) = 3$?

> $E(W^4) = E(X^2)$, where X is a chi-square and W has a $N(0,1)$ distribution.
> By definition, $\mathrm{Var}(X) = E(X^2) - E(X)^2 \Rightarrow E(X^2) = \mathrm{Var}(X) + E(X)^2$
> $E(X)^2 = E(W^2)^2 = (\mathrm{Var}(W) + E(W)^2)^2 = (\sigma^2 + 0^2)^2 = \sigma^4 = 1^2 = 1$
> (The variance of a chi-squared distribution with n degrees of freedom is $2n\sigma^2$, so with $n = 1$ degree of freedom, we have $\mathrm{Var}(X) = 2(1 \cdot \sigma^2) = 2\sigma^2 = 2 \cdot 1 = 2$)
> $E(W^4) = E(X^2) = 2 + 1 = 3$

Therefore, the kurtosis of T is given by

$$T = \frac{3 \cdot (r-2)}{r-4}.$$

∎

4. Let F have an F-distribution with parameters r_1 and r_2. Show that $1/F$ has a F-distribution with parameters r_2 and r_1.

Solution:
Recall from Equation 3.6.7, we have the definition of F-distribution

$$F = \frac{U/r_1}{V/r_2},$$

where U and V are independent χ-square variables with r_1 and r_2 degrees of freedom, respectively, then

$$\frac{1}{F} = \frac{V/r_2}{U/r_1},$$

which has a F-distribution with r_2 and r_1 degrees of freedom.

$\blacksquare$

5. Consider the mixture distribution, $\frac{9}{10} \cdot N(0,1) + \frac{1}{10} \cdot N(0,9)$. Show that its kurtosis is 8.34.

Solution:

$$f_1(x) \sim N(0,1), \quad f_2(x) \sim N(0,9), \quad f(x) = \frac{9}{10} \cdot f_1(x) + \frac{1}{10} \cdot f_2(x)$$

$$\begin{aligned}
\mathrm{Var}(X) &= \frac{9}{10} \cdot 1 + \frac{1}{10} \cdot 9 = \frac{18}{10} \\
k &= \frac{E(X^4)}{\mathrm{Var}(X)^2} = \frac{\frac{9}{10} \cdot E(X_1^4) + \frac{1}{10} \cdot E(X_2^4)}{(\frac{18}{10})^2} = \frac{9 \cdot (3 \cdot \sigma_1^4) + 1 \cdot (3 \cdot \sigma_2^4)}{18^2/10} = \frac{27 \cdot 1^4 + 3 \cdot 3^4}{18^2/10} \\
&= \frac{270 \cdot 10}{324} = \frac{225}{27} = 8.\bar{3} \approx 8.34
\end{aligned}$$

$\blacksquare$

6. Suppose the number of customers X that enter a store between the hours 9:00 a.m. and 10:00 a.m. follows a Poisson distribution with parameter θ. Suppose a random sample of the number of customers that enter the store between 9:00 a.m. and 10:00 a.m. for 10 days results in the values: 9, 7, 9, 15, 10, 13, 11, 7, 2, 12.
(a) Determine the maximum likelihood estimate of θ. Show that it is an unbiased estimator.
(b) Based on these data, obtain the realization of your estimator in part **(a)**. Explain the meaning of this estimate in terms of the number of customers.

Solution:
(a) To compute $\widehat{\theta}$, we can use the definition of the ML estimator:

$$\begin{aligned}
X \sim P(\theta) \quad P_X(x) &= \frac{e^{-\theta} \cdot \theta^x}{x!}, \quad x = 0, 1, 2, \cdots \\
L(\theta) &= \prod_{i=1}^{n} P(x_i > \theta) = \prod_{i=1}^{n} \frac{e^{-\theta} \cdot \theta^{x_i}}{x_i!} = \frac{e^{-n\theta} \cdot \theta^{x_1+x_2+\cdots+x_n}}{x_1!x_2!\cdots x_n!} \\
\Rightarrow \ln(L(\theta)) &= \ln(e^{-n\theta} \cdot \theta^{x_1+x_2+\cdots+x_n}) - \ln(x_1!x_2!\cdots x_n!) \\
&= \ln(e^{-n\theta}) + \ln(\theta^{x_1+x_2+\cdots+x_n}) + \ln(x_1!x_2!\cdots x_n!) \\
&= -n \cdot \theta + (\sum_{i=1}^{n} x_i) \cdot \ln(\theta) - \ln(\prod_{i=1}^{n} x_i!) \\
\Rightarrow \frac{\partial \ln(L(\theta))}{\partial \theta} &= -n + \frac{1}{\theta} \sum_{i=1}^{n} x_i = 0 \quad \Rightarrow \frac{1}{\theta} \sum_{i=1}^{n} x_i = n \quad \Rightarrow \theta = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x} \\
&\Rightarrow \text{ML estimator} \quad \widehat{\theta} = \bar{x}
\end{aligned}$$

To show this estimator to be unbiased, we need to find its expected value of $\bar{x}$ and show that it is equal to θ. The expected value of $\bar{x}$ is:

$$E(\bar{x}) = E(\frac{1}{n}\sum_{i=1}^{n} x_i) = \frac{1}{n} \cdot E(\sum_{i=1}^{n} x_i) = \frac{1}{n} \cdot \sum_{i=1}^{n} \theta = \frac{1}{n} \cdot n \cdot \theta = \theta$$

Therefore, we have shown that the ML estimator $\widehat{\theta} = \bar{x}$ is unbiased.
(b) Using the computed estimator in part **(a)**, we have

$$\widehat{\theta} = \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{1}{10} \cdot (9 + 7 + 9 + 15 + 10 + 13 + 11 + 7 + 2 + 12) = 9.5$$

Therefore, based on our estimated result, we may assume that there were roughly 9-10 customers who entered the store between 9:00 a.m. and 10:00 a.m.

■
