# STA2001 Assignment 5 Solution

(4.2-5). Let $X$ and $Y$ be random variables with respective means $\mu_X$ and $\mu_Y$, respective variances $(\sigma_x)^2$, $(\sigma_Y)^2$ and correlation coefficient $\rho$. Fit the line $y = a + bx$ by the method of least squares to the probability distribution by minimizing the expectation

$$K(a,b) = E\left[(Y - a - bX)^2\right]$$

with respect to $a$ and $b$. Hint: Consider $\partial k/\partial a = 0$ and $\partial k/\partial b = 0$, and solve simultaneously.

**Solution:**

We first rewrite the expectation in order to express it in terms of $a$ and $b$ explicitly,

$$\begin{aligned} K(a,b) &= E\left[(Y - a - bX)^2\right] \\ &= E\left[Y^2 - 2aY - 2bXY + a^2 + 2abX + b^2X^2\right] \\ &= E(Y^2) - 2aE(Y) - 2bE(XY) + a^2 + 2abE(X) + b^2E(X^2) \end{aligned}$$

By fist order condtion (i.e. set the first partial derivative equal to 0),

$$\frac{\partial K(a,b)}{\partial a} = 2a - 2E(Y) + 2bE(x) = 0$$

which yields

$$a = \mu_Y - b\mu_X$$

Another partial derivative is given by

$$\begin{aligned} \frac{\partial K(a,b)}{\partial b} &= 2aE(x) - 2E(XY) + 2bE(X^2) \\ &= 2a\mu_X - 2(\mu_X\mu_Y + \rho\sigma_X\sigma_Y) + 2b(\sigma_X^2 + \mu_X^2) \end{aligned}$$

Again by first order condtion we set $\frac{\partial K(a,b)}{\partial b} = 0$ and note that we have obtained $a = \mu_Y - b\mu_X$, and thus we can solve for $b$,

$$b = \rho\frac{\sigma_X}{\sigma_Y}$$

and therefore

$$a = \mu_Y - b\mu_X = \mu_Y - \left(\rho\frac{\sigma_X}{\sigma_Y}\right)\cdot\mu_X$$

With $a$ and $b$, we can write the equation for the line,

$$y = a + bx = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

(4.2-3). Roll a fair four-sided die twice. Let $X$ equal the outcome on the first roll, and let $Y$ equal the sum of the two rolls.

(a) Determine $\mu_X, \mu_Y, \sigma_x^2, \sigma_Y^2, \mathrm{Cov}(X,Y)$, and $\rho$.

(b) Find the equation of the least squares regression line and draw it on your graph. Does the line make sense to you intuitively?

**Solution:**

(a) According to given information, we can easily write down the joint pmf and the marginal pmfs,

$$f(x,y) = \frac{1}{16}, \quad x = 1,2,3,4, \; y = x+1, x+2, x+3, x+4.$$
$$f_X(x) = \frac{1}{4}, \quad x = 1,2,3,4.$$
$$f_Y(y) = \begin{cases} \frac{1}{16}, & y = 2,8 \\ \frac{1}{8}, & y = 3,7 \\ \frac{3}{16}, & y = 4,6 \\ \frac{1}{4}, & y = 5 \end{cases}$$

Then we can compute the following,

$$\mu_X = \sum_{x=1}^{4} x f_X(x) = 1 \times \frac{1}{4} + 2 \times \frac{1}{4} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} = \frac{10}{4} = \frac{5}{2}$$
$$\mu_Y = \sum_{y=2}^{8} y f_Y(y) = 2 \cdot \frac{1}{16} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{3}{16} + 5 \cdot \frac{1}{4} + \frac{1}{16} + 7 \cdot \frac{1}{8} + 6 \times \frac{3}{4} = 5$$
$$\mu_X^2 = \sum_{x=1}^{4} x^2 f_X(x) = 1^2 \cdot \frac{1}{4} + 2^2 \cdot \frac{1}{4} + 3^2 \cdot \frac{1}{4} + 4^2 \cdot \frac{1}{4} = 7.5$$
$$\mu_Y^2 = \sum_{y=2}^{8} y^2 f_Y(y) = 2^2 \cdot \frac{1}{16} + 3^2 \cdot \frac{1}{8} + 4^2 \cdot \frac{3}{16} + 5^2 \cdot \frac{1}{4} + +6^2 \cdot \frac{3}{16} + 7^2 \frac{1}{8} + 8^2 \cdot \frac{1}{16} = 27.5$$
$$\text{Var}(X) = E(X^2) - (EX)^2 = 7.5 - 2.5^2 = 1.25$$
$$\text{Var}(Y) = E(Y^2) - (EY)^2 = 27.5 - 5^2 = 2.5$$
$$\text{Cov}(X,Y) = E(XY) - E(X)E(Y) = \sum_{(x,y)\in S} xy f(x,y) - \mu_X \mu_Y = 1.25, \text{ where } S \text{ is the support of the joint pmf.}$$
$$\rho = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\sqrt{2}}{2}$$

(b) The equation is given by

$$y = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X) = x + 2.5$$

(4.2-9). A car dealer sells $X$ cars each day and always tries to sell an extended warranty on each of these cars. (In our opinion, most of these warranties are not good deals.) Let $Y$ be the number of extended warranties sold; then $Y \leq X$. The joint pmf of $X$ and $Y$ is given by

$$f(x,y) = c(x+1)(4-x)(y+1)(3-y)$$

$x = 0,1,2,3$, $y = 0,1,2$, with $y \leq x$.

(a) Find the value of $c$.
(b) Sketch the support of $X$ and $Y$.
(c) Record the marginal pmfs $f_X(x)$ and $f_Y(y)$ in the margins.
(d) Are $X$ and $Y$ independent?
(e) Compute $\mu_X$ and $\sigma_X^2$.
(f) Compute $\mu_Y$ and $\sigma_Y^2$.
(g) Compute $\text{Cov}(X,Y)$
(h) Determine $\rho$, the correlation coefficient.
(i) Find the best-fitting line and draw it on your figure.

**Solution:**

(a) Let $S$ be the support of the joint distribution (i.e. the space that the joint pmf takes positive values), so we must have

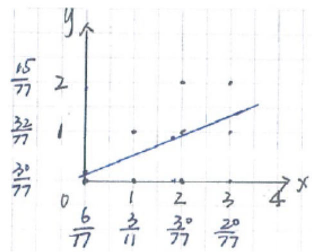
$$\sum_{(x,y)\in S} f(x,y) = 1$$

Therefore, given this condition and the joint pmf $f(x,y)$ we could write an equation with respect to $c$,

$$f(0,0)+f(1,1)+f(1,0)+f(2,0)+f(2,1)+f(2,2)+f(3,0)+f(3,1)+f(3,2)=1$$

Solve this equation, we have $c = \frac{1}{154}$.

(b)



(c)

$$f_X(x) = \begin{cases} \frac{6}{77}, & x=0 \\ \frac{21}{77}, & x=1 \\ \frac{30}{77}, & x=2 \\ \frac{20}{77}, & x=3 \end{cases} \qquad f_Y(y) = \begin{cases} \frac{30}{77}, & y=0 \\ \frac{32}{77}, & y=1 \\ \frac{15}{77}, & y=2 \end{cases}$$

(d) Since the space is not a rectangular, $X$ and $Y$ are not independent.

(e) $\mu_x = \sum_{x=0}^{3} x f_X(x) = \frac{141}{77}$

$\sigma_X^2 = E\left(X^2\right) - (EX)^2 = \frac{4836}{5929}$

(f) $\mu_Y = \sum_{y=0}^{2} y f_Y(y) = \frac{62}{77}$

$\sigma_Y^2 = E\left(Y^2\right) - (EY)^2 = \frac{3240}{5929}$

(g) $\mathrm{Cov}(X,Y) = E(XY) - \mu_X\mu_Y = \frac{1422}{5929}$

(h) $\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{79}{24180}\sqrt{12090} \approx 0.3592$

(i) Recall that

$$y = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)$$

Fill in the known values:

$$y = \frac{62}{77} + \frac{7\sqrt{12090}}{24180}\frac{\sqrt{3240/5929}}{\sqrt{4836/5929}}\left(x - \frac{141}{77}\right) = \frac{21}{806}x + \frac{47011}{62062}$$

(4.3-4). The alleles for eye color in a certain male fruit fly are $(R,W)$. The alleles for eye color in the mating female fruit fly are $(R,W)$. Their offspring receive one allele for eye color from each parent. If an offspring ends up with either $(W,W)$, $(R,W)$, or $(W,R)$, its eyes will look white. Let $X$ equal the number of offspring having white eyes. Let $Y$ equal the number of white-eyed offspring having $(R,W)$ or $(W,R)$ alleles.

(a) If the total number of offspring is $n=400$, how is $X$ distributed?
(b) Give the values of $E(X)$ and $\mathrm{Var}(X)$.
(c) Given that $X=300$, how is $Y$ distributed?
(d) Give the value of $E(Y|X=300)$ and the value of $\mathrm{Var}(Y|X=300)$.

**Solution:**

(a) $X \sim b(400, 0.75)$, that is, follows a binomial distribution with parameters $n=400$ and $p=0.75$.

(b) Since $X \sim b(400, 0.75)$, we can easily compute

$$E(X) = 400 \times 0.75 = 300, \quad \mathrm{Var}(X) = 400 \times 0.75 \times (1 - 0.75) = 75$$

(c) The conditional random variable $Y|X = 300$ follows $b(300, 2/3)$.

(d) From the answer to (c), we can easily compute

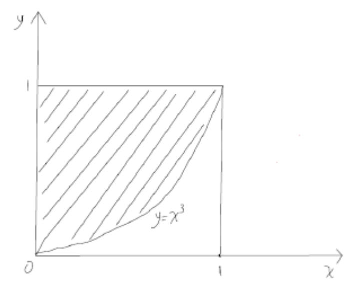
$$E(Y|X = 300) = 300 \times 2/3 = 200$$

$$\mathrm{Var}(Y|X = 300) = 300 \times 2/3 \times 1/3 = 200/3$$

(4.4-7). Let $f(x, y) = 4/3$, $0 < x < 1$, $x^3 < y < 1$, zero elsewhere.
(a) Sketch the region where $f(x, y) > 0$.
(b) Find $P(X > Y)$.

**Solution:**

(a)



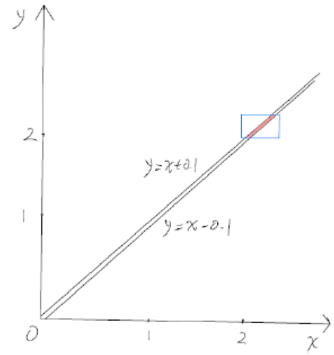

(b)

$$\begin{aligned} P(X > Y) &= 1 - P(X \le Y) \\ &= 1 - \int_0^1 \int_x^1 \frac{4}{3}\, dy\, dx \\ &= 1 - \int_0^1 \frac{4}{3}(1 - x)\, dx \\ &= 1 - \frac{2}{3} \\ &= \frac{1}{3} \end{aligned}$$

(4.4-9). Two construction companies make bids of $X$ and $Y$ (in \$100,000 's) on a remodeling project. The joint pdf of $X$ and $Y$ is uniform on the space $2 < x < 2.5$, $2 < y < 2.3$. If $X$ and $Y$ are within 0.1 of each other, the companies will be asked to rebid; otherwise, the low bidder will be awarded the contract. What is the probability that they will be asked to rebid?

**Solution:**

We sketch the region as follows. The entire blue rectangular is the region where $f(x, y) > 0$, and the region in pink is the desired probability.




We are given the joint pdf,

$$f(x,y) = \frac{20}{3}, \quad 2 < x < 2.5,\ 2 < y < 2.3$$

As we are going to compute $P(|X-Y| < 0.1)$, we must be careful about the support of $Y$ (i.e. partition the support of $Y$).

$$\begin{aligned}
P(|X-Y|<0.1) &= P(|X-Y|<0.1, 2<Y<2.1) + P(|X-Y|<0.1, 2.1<Y<2.3) \\
&= \int_2^{2.1}\int_2^{y+0.1} \frac{20}{3}\,dx\,dy + \int_{2.1}^{2.3}\int_{-0.1+y}^{0.1+y} \frac{20}{3}\,dx\,dy \\
&= \frac{1}{10} + \frac{8}{30} \\
&= \frac{11}{30}
\end{aligned}$$

(4.4-15). An automobile repair shop makes an initial estimate $X$ (in thousands of dollars) of the amount of money needed to fix a car after an accident. Say $X$ has the pdf

$$f(x) = 2e^{-2(x-0.2)}, \quad 0.2 < x < \infty$$

Given that $X = x$, the final payment $Y$ has a uniform distribution between $x - 0.1$ and $x + 0.1$. What is the expected value of $Y$ ?

**Solution:**

From the given information, the conditional density is given by

$$f(y|x) = \frac{1}{(x+0.1)-(x-0.1)} = 5, \quad x - 0.1 < y < x + 0.1$$

By Baye's theorem, the joint density is given by

$$f(x,y) = f(y|x)\cdot f(x) = 10e^{-2(x-0.2)}, \quad 0.2 < x < \infty,\ x - 0.1 < y < x + 0.1$$

Therefore, the expectation is computed as follows

$$\begin{aligned}
E(Y) &= \int_{0.2}^{+\infty}\int_{x-0.1}^{x+0.1} 10ye^{-2(x-0.2)}\,dy\,dx \\
&= \int_{0.2}^{+\infty}\left(5y^2e^{-2(x-0.2)}\right)\Big|_{x-0.1}^{x+0.1}\,dx \\
&= \int_{0.2}^{+\infty} 5(0.4x)e^{-2(x-0.2)}\,dx \\
&= \int_{0.2}^{+\infty} 2xe^{-2(x-0.2)}\,dx \\
&= \left(-\frac{1}{2}(2x+1)e^{-2(x-0.2)}\right)\Big|_{0.2}^{+\infty} \\
&= 0.7
\end{aligned}$$

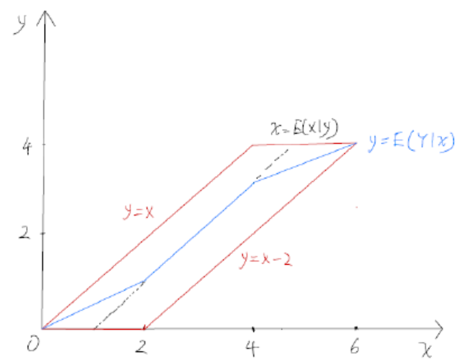
(4.4-18). Let $f(x,y) = 1/8$, $0 \le y \le 4$, $y \le x \le y+2$, be the joint pdf of $X$ and $Y$.

(a) Sketch the region for which $f(x,y) > 0$.
(b) Find $f_X(x)$, the marginal pdf of $X$.
(c) Find $f_Y(y)$, the marginal pdf of $Y$.
(d) Determine $h(y|x)$, the conditional pdf of $Y$, given that $X = x$.
(e) Determine $g(x|y)$, the conditional pdf of $X$, given that $Y = y$.
(f) Compute $E(Y|x)$, the conditional mean of $Y$, given that $X = x$.
(g) Compute $E(X|y)$, the conditional mean of $X$, given that $Y = y$.
(h) Graph $y = E(Y|x)$ on your sketch in part (a). Is $y = E(Y|x)$ linear?
(i) Graph $x = E(X|y)$ on your sketch in part (a). Is $x = E(X|y)$ linear?

**Solution:**

(a) & (h) & (i)

The entire red parallelogram is the region where $f(x,y) > 0$, the blue line is the conditional mean required by (h), and the dot line in black and the blue straight line connected them is the conditional mean required by (i).




(b)

$$f_X(x) = \begin{cases} \int_0^x \frac{1}{8}\,dy = \frac{x}{8}, & 0 \le x \le 2 \\ \int_{x-2}^{x} \frac{1}{8}\,dy = \frac{1}{4}, & 2 < x < 4 \\ \int_{x-3}^{4} \frac{1}{8}\,dy = \frac{6-x}{8}, & 4 \le x \le 6 \end{cases}$$

(c)

$$f_Y(y) = \int_y^{y+2} \frac{1}{8}\,dx = \frac{1}{4}, \quad 0 \le y \le 4$$

(d)

$$h(y|x) = \frac{f(x,y)}{f_X(x)} = \begin{cases} 1/x, & 0 \le y \le x,\ 0 \le x \le 2 \\ 1/2, & x-2 \le y \le x,\ 2 < x < 4 \\ 1/(6-x), & x-2 \le y \le 4,\ 4 \le x \le 6 \end{cases}$$

(e)

$$g(x|y) = \frac{f(x,y)}{f_Y(y)} = 1/2, \quad y \le x \le y+2,\ 0 \le y \le 4$$

(f)

$$E(Y|x) = \begin{cases} \int_0^x y\frac{1}{x}\,dy = \frac{x}{2}, & 0 \le x \le 2 \\ \int_{x-2}^{x} y\frac{1}{2}\,dy = x-1, & 2 < x < 4 \\ \int_{x-2}^{4} y\frac{1}{6-x}\,dy = \frac{x+2}{2}, & 4 <\le x \le 6 \end{cases}$$

(g)

$$E(X|y) = \int_y^{y+2} x\frac{1}{2}\,dx = y+1, \quad 0 < y < 4$$

(h) No.

(i) Yes.

(4.5-1). Let $X$ and $Y$ have a bivariate normal distribution with para.meters $\mu_X = -3$, $\mu_Y = 10$, $\sigma_X^2 = 25$, $\sigma_Y^2 = 9$, and $\rho = 3/5$. Compute
(a) $P(-5 < X < 5)$
(b) $P(-5 < X < 5|Y = 13)$
(c) $P(7 < Y < 16)$
(d) $P(7 < Y < 16|X = 2)$

**Solution:**

Given a bivariate normal distribution, we can prove that the marginal distributions and conditional distributions ( $X|y$ and $Y|x$ ) also follow normal distributions (check this on the textbook or lecture note). Based on this fact, we can answer the following questions.

(a) $P(-5 < X < 5) = \Phi\left(\frac{5-(-3)}{\sqrt{25}}\right) - \Phi\left(\frac{-5-(-3)}{\sqrt{25}}\right) = 0.6006$

(b)

$$\mu_{X|Y=13} = \mu_X + \rho\left(\frac{\sigma_X}{\sigma_Y}\right)(13-\mu_Y) = -3 + \frac{3}{5}\cdot\frac{5}{3}\cdot(13-10) = 0$$

$$\sigma^2_{X|Y=13} = \sigma_X^2\left(1-\rho^2\right) = 25\times\left(1-\left(\frac{3}{5}\right)^2\right) = 16$$

$$P(-5 < X < 5|Y = 13) = \Phi\left(\frac{5-0}{\sqrt{16}}\right) - \Phi\left(\frac{-5-0}{\sqrt{16}}\right) = 0.7888$$

(c) $P(7 < Y < 16) = \Phi\left(\frac{16-10}{\sqrt{9}}\right) - \Phi\left(\frac{7-10}{\sqrt{9}}\right) = 0.8185$

(d)

$$\mu_{Y|X=2} = \mu_Y + \rho\left(\frac{\sigma_Y}{\sigma_X}\right)(2-\mu_X) = 10 + \frac{3}{5}\times\frac{3}{5}(2-1-3) = \frac{59}{5}$$

$$\sigma^2_{Y|X=2} = \sigma_Y^2\cdot\left(1-\rho^2\right) = 9\cdot\left(1-\left(\frac{3}{5}\right)^2\right) = \frac{144}{25}$$

$$\Phi(7 < X < 16|X = 2) = \Phi\left(\frac{16-59/5}{12/5}\right) - \Phi\left(\frac{7-59/5}{12/5}\right) = 0.9371$$

(4.5-6). For a freshman taking introductory statistics and majoring in psychology, let $X$ equal the student's ACT mathematics score and $Y$ the students ACT verbal score. Assume that $X$ and $Y$ have a bivariate normal distribution with $\mu_X = 22.7$, $\sigma_X^2 = 17.64$, $\mu_Y = 22.7, \sigma_Y^2 = 12.25$, and $\rho = 0.78$.
(a) Find $P(18.5 < Y < 25.5)$.
(b) Find $E(Y|x)$.
(c) Find $\mathrm{Var}(Y|x)$.
(d) Find $P(18.5 < Y < 25.5|X = 23)$.
(e) Find $P(18.5 < Y < 25.5|X = 25)$.
(f) For $x = 21, 23$, and $25$, draw a graph of $z = h(y|x)$ similar to Figure 4.5-1.

**Solution:**

(a) $P(18.5 < Y < 25.5) = \Phi(0.8) - \Phi(-1.2) = 0.6730$

(b) $E(Y|x) = 22.7 + 0.78\left(\sqrt{\frac{12.25}{17.64}}\right)(x-22.7) = 0.65x + 7.945$

(c) $\mathrm{Var}(Y|x) = 12.25\left(1-0.78^2\right) = 4.7971$

(d)

$$P(18.5 < Y < 25.5|X = 23) = \Phi\left(\frac{25.5-0.65\times 23-7.945}{\sqrt{4.7971}}\right) - \Phi\left(\frac{18.5-0.65\times 23-7.945}{\sqrt{4.7971}}\right)$$

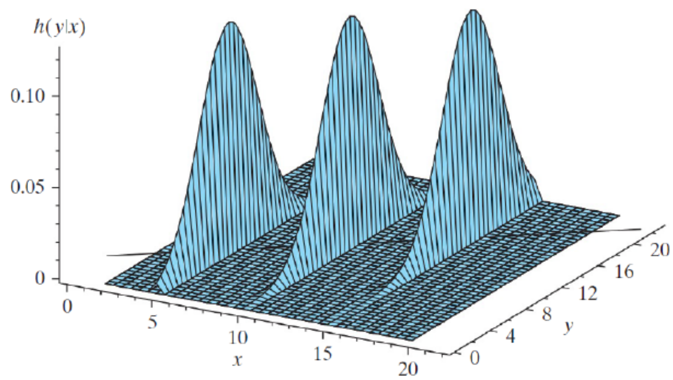



**Figure 4.5-1** Conditional pdf of $Y$, given that $x = 5, 10, 15$

$$= \Phi(1.189) - \Phi(-2.007)$$
$$= 0.8604$$

(e) $P(18.5 < Y < 25.5|X = 25) = \Phi(0.596) - \Phi(-2.60) = 0.7197$