# Statistics 341

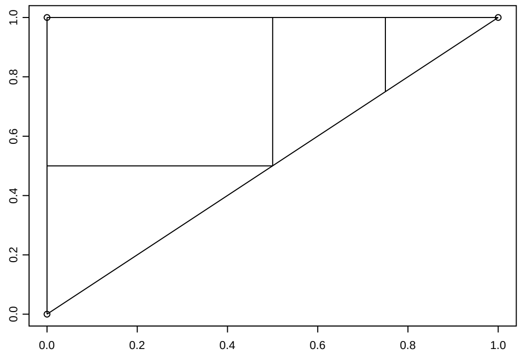
## Fall 2008 - Homework Assignment #8 Answers

## We will go over some of these problems during the last week of classes

1. Let $Y_1$ and $Y_2$ have the joint probability density function given by

$$f(y_1, y_2) = \begin{cases} 6(1-y_2) & 0 \le y_1 \le y_2 \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

   (a) Find $P(Y_1 \le 3/4, Y_2 \ge 1/2)$.
   The region where $Y_1 \le 0.75$ and $Y_2 \ge 0.5$ within the support of $Y_1$ and $Y_2$ is shown in Figure 1. This region must be split into two parts. The two parts I chose are shown in the graph. You could choose to break this region up in a different manner.




First region

$$\begin{aligned} \int_{0.5}^{1}\int_{0}^{0.5} 6(1-y_2)dy_1dy_2 &= \int_{0.5}^{1} 6(1-y_2)\left(y_1|_0^{0.5}\right)dy_2 \\ &= \int_{0.5}^{1} 3(1-y_2)dy_2 \\ &= \left(3y_2 - 1.5y_2^2\right)|_{0.5}^{1} \\ &= 1.5 - 1.125 = 0.375 \end{aligned}$$

Second region

$$\begin{aligned} \int_{0.5}^{0.75}\int_{y_1}^{1} 6(1-y_2)dy_2dy_1 &= \int_{0.5}^{0.75}\left(6y_2 - 3y_2^2|_{y_1}^{1}\right)dy_1 \\ &= \int_{0.5}^{0.75} 3 - 6y_1 + 3y_1^2 dy_1 \\ &= \left(3y_1 - 3y_1^2 + y_1^3\right)|_{0.5}^{0.75} \\ &= 0.984375 - 0.875 = 0.109375 \end{aligned}$$

Adding the two regions together gives $0.375 + 0.109375 = 0.484375$.

(b) Find the marginal distributions for both $Y_1$ and $Y_2$.

The marginal distribution of $Y_1$ is

$$\begin{aligned} f_1(y_1) &= \int_{y_1}^{1} 6(1-y_2)dy_2 \\ &= \left(6y_2 - 3y_2^2|_{y_1}^{1}\right) \\ &= 3y_1^2 - 6y_1 + 3 = 3(1-y_1)^2 \end{aligned}$$

for $0 \le y_1 \le 1$ and 0 otherwise.

The marginal distribution of $Y_2$ is

$$\begin{aligned} f_2(y_2) &= \int_{0}^{y_2} 6(1-y_2)dy_1 \\ &= (6(1-y_2)y_1|_0^{y_2}) \\ &= 6y_2(1-y_2) \end{aligned}$$

for $0 \le y_2 \le 1$ and 0 otherwise.

(c) Find the conditional density function of $Y_1$ when $Y_2 = y_2$.

$$f(y_1|Y_2 = y_2) = \frac{f(y_1, y_2)}{f_2(y_2)} = \frac{6(1-y_2)}{6y_2(1-y_2)} = \frac{1}{y_2}$$

for $0 \le y_1 \le y_2$ and 0 elsewhere.

(d) Find the conditional density function of $Y_2$ when $Y_1 = y_1$.

$$f(y_2|Y_1 = y_1) = \frac{f(y_1, y_2)}{f_1(y_1)} = \frac{6(1-y_2)}{3(1-y_1)^2} = \frac{2(1-y_2)}{(1-y_1)^2}$$

for $y_1 \le y_2 \le 1$ and 0 elsewhere.

(e) Find $P(Y_2 \ge 3/4|Y_1 = 1/2)$.

$$\begin{aligned} P(Y_2 \ge 0.75|Y_1 = 0.5) &= \int_{0.75}^{1} \frac{2(1-y_2)}{(1-0.5)^2} dy_2 \\ &= \int_{0.75}^{1} 8(1-y_2)dy_2 \\ &= \left(8y_2 - 4y_2^2|_{0.75}^{1}\right) \\ &= 4 - (8(3/4) - 4(3/4)^2) \\ &= 0.25 \end{aligned}$$

(f) Are $Y_1$ and $Y_2$ independent random variables? Explain your answer.

We have found the marginal distributions as

$$\begin{aligned} f_1(y_1) &= 3(1-y_1)^2 \qquad \text{for } 0 \le y_1 \le 1 \\ f_2(y_2) &= 6y_2(1-y_2) \qquad \text{for } 0 \le y_2 \le 1 \end{aligned}$$

Since the product of the marginal distributions does not equal to joint distribution, these random variables are dependent.

(g) Find the covariance between $Y_1$ and $Y_2$.
Using the marginal distributions of $Y_1$ and $Y_2$, we have

$$\begin{aligned}
E(Y_1) &= \int_0^1 y_1 3(1-y_1)^2 dy_1 \\
&= 3\int_0^1 y_1 - 2y_1^2 + y_1^3 dy_1 \\
&= 3\left(\frac{y_1^2}{2} - \frac{2y_1^3}{3} + \frac{y_1^4}{4}\right)|_0^1 \\
&= 3\left(\frac{1}{2} - \frac{2}{3} + \frac{1}{4}\right) \\
&= 0.25
\end{aligned}$$

$$\begin{aligned}
E(Y_2) &= \int_0^1 y_2 6y_2(1-y_2) dy_2 \\
&= 6\int_0^1 y_2^2 - y_2^3 dy_2 \\
&= 6\left(\frac{y_2^3}{3} - \frac{y_2^4}{4}\right)|_0^1 \\
&= 6\left(\frac{1}{3} - \frac{1}{4}\right) \\
&= 0.5
\end{aligned}$$

$$\begin{aligned}
E(Y_1Y_2) &= \int_0^1 \int_0^{y_2} 6y_1y_2(1-y_2) dy_1 dy_2 \\
&= \int_0^1 6y_2(1-y_2)\left(\frac{y_1^2}{2}\right)|_0^{y_2} dy_2 \\
&= \int_0^1 3y_2^3(1-y_2) dy_2 \\
&= \int_0^1 3y_2^3 - 3y_2^4 dy_2 \\
&= \left(\frac{3}{4}y_2^4 - \frac{3}{5}y_2^4\right)|_0^1 \\
&= \frac{3}{4} - \frac{3}{5} = 0.15
\end{aligned}$$

$$\begin{aligned}
Cov(Y_1, Y_2) &= E(Y_1Y_2) - E(Y_1)E(Y_2) \\
&= 0.15 - (0.25)(0.5) = 0.025
\end{aligned}$$

(h) Find the correlation between $Y_1$ and $Y_2$.

$$
\begin{aligned}
E(Y_1^2) &= 3\int_0^1 y_1^2 - 2y_1^3 + y_1^4 dy_1 \\
&= 3\left(\frac{y_1^3}{3} - \frac{y_1^4}{2} + \frac{y_1^5}{5}\right)|_0^1 \\
&= 3\left(\frac{1}{3} - \frac{1}{2} + \frac{1}{5}\right) \\
&= 0.1
\end{aligned}
$$

$$
\begin{aligned}
V(Y_1) &= E(Y_1^2) - (E(Y_1))^2 \\
&= 0.1 - 0.25^2 = 0.037
\end{aligned}
$$

$$
\begin{aligned}
E(Y_2^2) &= 6\int_0^1 y_2^3 - y_2^4 dy_2 \\
&= 6\left(\frac{y_2^4}{4} - \frac{y_2^5}{5}\right)|_0^1 \\
&= 6\left(\frac{1}{4} - \frac{1}{5}\right) \\
&= 0.3
\end{aligned}
$$

$$
\begin{aligned}
V(Y_2) &= E(Y_2^2) - (E(Y_2))^2 \\
&= 0.3 - 0.5^2 = 0.05
\end{aligned}
$$

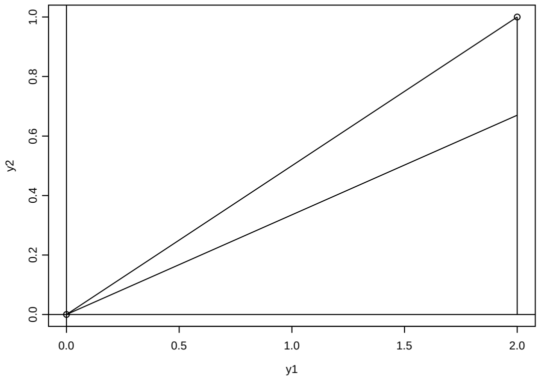
$$
\begin{aligned}
Corr(Y_1, Y_2) &= \frac{Cov(Y_1, Y_2)}{\sqrt{V(Y_1)V(Y_2)}} \\
&= \frac{0.025}{\sqrt{0.0375(0.05)}} \\
&= 0.5774
\end{aligned}
$$

2. An environmental engineer measures the amount (by weight) of particulate pollution in air samples of a certain volume collected over two smokestacks at a coal-operated power plant. One of the stacks is equipped with a cleaning device. Let $Y_1$ denote the amount of pollutant per sample collected above the stack that has no cleaning device and let $Y_2$ denote the amount of pollutant per sample collected above the stack that is equipped with the cleaning device. Suppose that the relative frequency behavior of $Y_1$ and $Y_2$ can be modeled by

$$
f(y_1, y_2) = \begin{cases} 1 & 0 \leq y_1 \leq 2, 0 \leq y_2 \leq 1, 2y_2 \leq y_1 \\ 0 & \text{elsewhere} \end{cases}
$$

(a) Find $P(Y_1 \geq 3Y_2)$. This is the probability the cleaning device reduces the amount of pollutant by one-third or more.
The region where $Y_1 \geq 3Y^2$ within the support of $Y_1$ and $Y_2$ is given in Figure 2.

$$
P(Y_1 \geq 3Y_2) = \int_0^{2/3}\int_{3y_2}^{2} dy_1 dy_2
$$




$$
\begin{aligned}
&= \int_0^{2/3} (2 - 3y_2) dy_2 \\
&= \left(2y_2 - \frac{3}{2}y_2^2)|_0^{2/3}\right) \\
&= 2/3
\end{aligned}
$$

(b) Find the marginal distributions for both $Y_1$ and $Y_2$.
The marginal distribution for $Y_1$ is

$$f_1(y_1) = \int_0^{0.5y_1} dy_2 = 0.5y_1 \qquad \text{for } 0 \leq y_1 \leq 2$$

The marginal distribution for $Y_2$ is

$$f_2(y_2) = \int_{2y_2}^{2} dy_1 = 2(1 - y_2) \qquad \text{for } 0 \leq y_2 \leq 1$$

(c) Find the conditional density function of $Y_1$ when $Y_2 = y_2$.

$$f(y_1|Y_2 = y_2) = \frac{f(y_1, y_2)}{f_2(y_2)} = \frac{1}{2(1 - y_2)}$$

for $2y_2 \leq y_1 \leq 2$ and 0 elsewhere.

(d) Find the conditional density function of $Y_2$ when $Y_1 = y_1$.

$$f(y_2|Y_1 = y_1) = \frac{f(y_1, y_2)}{f_1(y_1)} = \frac{1}{0.5y_1} = \frac{2}{y_1}$$

for $0 \leq y_2 \leq 0.5y_1$ and 0 elsewhere.

(e) Find the probability that the amount of pollutant in a sample taken above the stack without the cleaner ($Y_1$) is 1.5 or greater given the amount of pollutant in a sample taken above the stack with the cleaner ($Y_2$) is 0.5.
We need to use $f(y_1|Y_2 = 0.5) = \frac{1}{2(1-y_2)} = \frac{1}{2(1-0.5)} = 1$.

$$P(Y_1 \geq 1.5|Y_2 = 0.5) = \int_{1.5}^{2} dy_1 = 2 - 1.5 = 0.5$$

(f) Are $Y_1$ and $Y_2$ independent random variables? Explain your answer.

We have found the marginal distributions as

$$\begin{aligned} f_1(y_1) &= 0.5y_1 \\ f_2(y_2) &= 2(1-y_2) \end{aligned}$$

The joint distribution is not equal to the product of the marginal distributions and so the two variables are dependent.

(g) Find the covariance between $Y_1$ and $Y_2$.

Using the two marginal distributions, we have

$$\begin{aligned} E(Y_1) &= \int_0^2 y_1 * 0.5y_1 dy_1 \\ &= \int_0^2 0.5y_1^2 dy_1 \\ &= \frac{1}{6} y_1^3|_0^2 \\ &= \frac{1}{6}(8-0) = \frac{4}{3} \end{aligned}$$

$$\begin{aligned} E(Y_2) &= \int_0^1 y_2 * 2(1-y_2) dy_2 \\ &= \int_0^1 2y_2 - 2y_2^2 dy_2 \\ &= y_2^2 - \frac{2}{3} y_2^3|_0^1 \\ &= 1 - \frac{2}{3} = \frac{1}{3} \end{aligned}$$

$$\begin{aligned} E(Y_1Y_2) &= \int_0^2 \int_0^{0.5y_1} y_1y_2 dy_2 dy_1 \\ &= \int_0^2 y_1 \frac{1}{2} y_2^2|_0^{0.5y_1} \\ &= \int_0^2 y_1 \frac{1}{8} y_1^2 \\ &= \int_0^2 \frac{1}{8} y_1^3 \\ &= \frac{1}{32} y_1^4|_0^2 \\ &= \frac{1}{32}(16-0) \\ &= 0.5 \end{aligned}$$

$$Cov(Y_1, Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2)$$

$$= 0.5 - \left(\frac{4}{3}\right)\left(\frac{1}{3}\right)$$
$$= \frac{1}{18}$$

(h) Find the correlation between $Y_1$ and $Y_2$.
For the correlation, we need $V(Y_1)$ and $V(Y_2)$.

$$\begin{aligned}
E(Y_1^2) &= \int_0^2 y_1^2 * 0.5y_1 dy_1 \\
&= \int_0^2 0.5y_1^3 dy_1 \\
&= \frac{1}{8}y_1^4|_0^2 \\
&= \frac{1}{8}(16-0) = 2
\end{aligned}$$

$$\begin{aligned}
V(Y_1) &= E(Y_1^2) - (E(Y_1))^2 \\
&= 2 - \left(\frac{4}{3}\right)^2 \\
&= \frac{2}{9}
\end{aligned}$$

$$\begin{aligned}
E(Y_2^2) &= \int_0^1 y_2^2 * 2(1-y_2) dy_2 \\
&= \int_0^1 2y_2^2 - 2y_2^3 dy_2 \\
&= \frac{2}{3}y_2^3 - \frac{1}{2}y_2^4|_0^1 \\
&= \frac{2}{3} - \frac{1}{2} \\
&= \frac{1}{6}
\end{aligned}$$

$$\begin{aligned}
V(Y_2) &= E(Y_2^2) - (E(Y_2))^2 \\
&= \frac{1}{6} - \left(\frac{1}{3}\right)^2 \\
&= \frac{1}{18}
\end{aligned}$$

$$\begin{aligned}
Corr(Y_1, Y_2) &= \frac{Cov(Y_1, Y_2)}{\sqrt{V(Y_1)V(Y_2)}} \\
&= \frac{\frac{1}{18}}{\sqrt{\left(\frac{2}{9}\right)\left(\frac{1}{18}\right)}} \\
&= 0.5
\end{aligned}$$

(i) Find the expected value and variance of $Y_1 - Y_2$, the amount of pollutant removed by the cleaner.

$$\begin{aligned} E(Y_1 - Y_2) &= E(Y_1) - E(Y_2) \\ &= \frac{4}{3} - \frac{1}{3} \\ &= 1 \end{aligned}$$

$$\begin{aligned} V(Y_1 - Y_2) &= V(Y_1) + V(Y_2) + 2(1)(-1)Cov(Y_1, Y_2) \\ &= \frac{2}{9} + \frac{1}{18} - 2\left(\frac{1}{18}\right) \\ &= \frac{1}{6} \end{aligned}$$

3. Let $Y_1, Y_2, \ldots, Y_n$ be independent Poisson random variables with means $\lambda_i$ for $i = 1, 2, \ldots, n$. Define the mean of these $n$ random variables to be

$$\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

Find the expected value and variance of $\overline{Y}$.

$$\begin{aligned} E(\overline{Y}) &= E\left(\frac{\sum Y_i}{n}\right) \\ &= \frac{1}{n} E\left(\sum Y_i\right) \\ &= \frac{1}{n} \sum E(Y_i) \\ &= \frac{1}{n} \sum_{i=1}^{n} \lambda_i \\ V(\overline{Y}) &= V\left(\frac{\sum Y_i}{n}\right) \\ &= \frac{1}{n^2} V\left(\sum Y_i\right) \\ &= \frac{1}{n^2}\left(\sum_{i=1}^{n} V(Y_i) + 2\sum\sum Cov(Y_i, Y_j)\right) \\ &= \frac{1}{n^2} \sum_{i=1}^{n} V(Y_i) \qquad \text{since the } Y_i \text{ variables are all independent} \\ &= \frac{1}{n^2} \sum_{i=1}^{n} \lambda_i \end{aligned}$$