STAT 511 su 2020 hw 8

covariance, correlation, independence, hierarchical models

1. Show that for any two rvs X and Y, $\text{Cov}(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$.

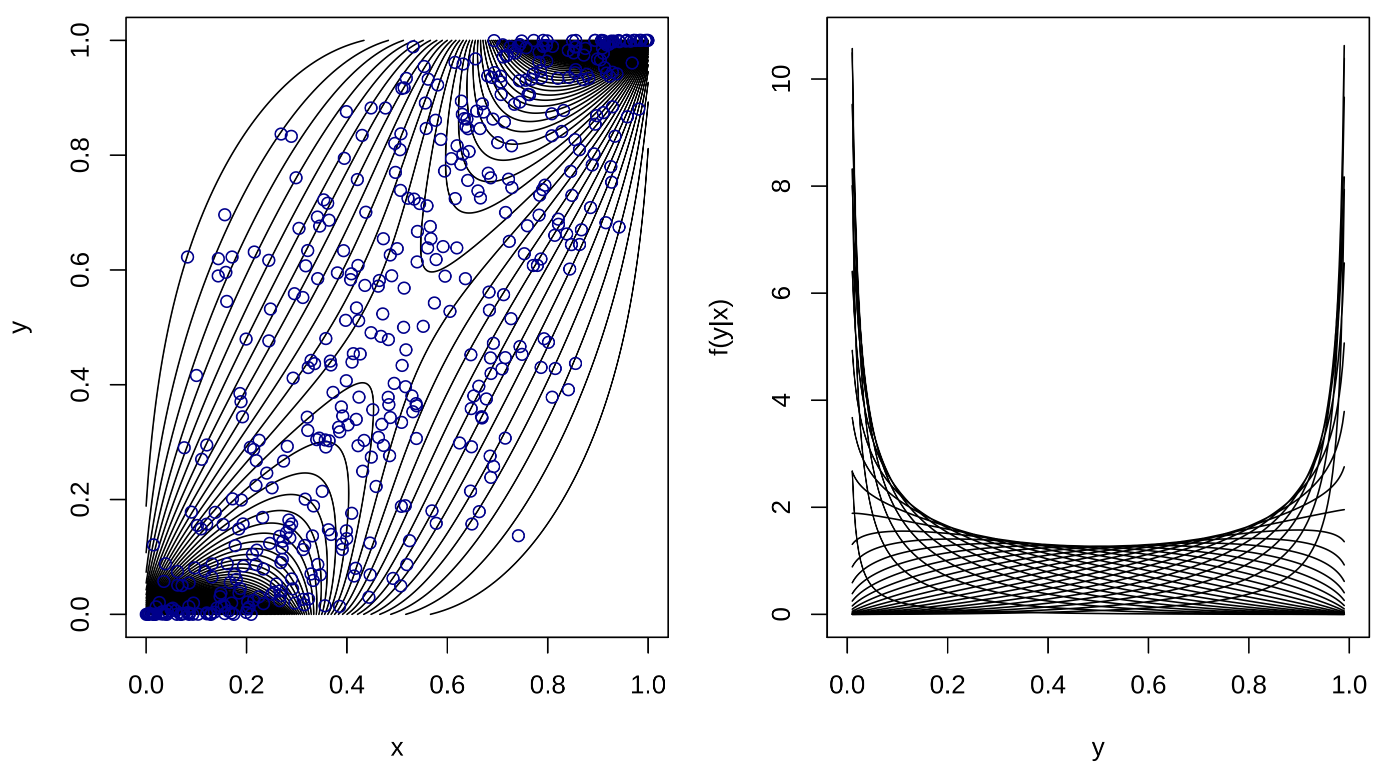
Writing $\mathbb{E}X = \mu_X$ and $\mathbb{E}Y = \mu_Y$, we have

$$\begin{aligned}\text{Cov}(X,Y) &= \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]\\ &= \mathbb{E}[XY - \mu_X Y - X\mu_Y + \mu_X\mu_Y]\\ &= \mathbb{E}XY - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y\\ &= \mathbb{E}XY - \mu_X\mu_Y.\end{aligned}$$

2. Let (X,Y) be a pair of random variables such that

$$\begin{aligned}Y|X &\sim \text{Beta}(3X, 3(1-X))\\ X &\sim \text{Uniform}(0,1).\end{aligned}$$

The plots below show contours of the joint density of X and Y, with 500 realizations of (X,Y) overlaid, as well as the conditional pdfs of $Y|X = x$ for several values of x.



(a) Give the joint pdf of (X, Y).

We have $f(x,y) = f(y|x)f_X(x)$, so that

$$f(x,y) = \frac{2}{\Gamma(3x)\Gamma(3(1-x))} y^{3x}(1-y)^{3(1-x)} \cdot \mathbf{1}(0 < x < 1)\mathbf{1}(0 < y < 1),$$

noting that $\Gamma(3x + 3(1-x)) = \Gamma(3) = 2$.

(b) Check whether X and Y are independent.

Since we cannot factor the joint pdf $f(x,y)$ into the product of a function of only x and a function of only y, the random variables X and Y are not independent.

(c) Write down the integral which would yield the marginal pdf of Y.

The marginal pdf f_Y of Y is given by

$$f_Y(y) = \int_0^1 \frac{2}{\Gamma(3x)\Gamma(3(1-x))} y^{3x}(1-y)^{3(1-x)} dx \cdot \mathbf{1}(0 < y < 1).$$

(d) Find $\mathbb{E}[Y|X]$.

We have

$$\mathbb{E}[Y|X] = \frac{3X}{3X + 3(1-X)} = X.$$

(e) Find $\mathbb{E}Y$.

Using the iterated expectation result, we have

$$\mathbb{E}Y = \mathbb{E}(\mathbb{E}[Y|X]) = \mathbb{E}(X) = 1/2.$$

(f) Find $\text{Var}[Y|X]$.

We have

$$\text{Var}[Y|X] = \frac{3X \cdot 3(1-X)}{(3X + 3(1-X))^2(3X + 3(1-X) + 1)} = \frac{9X(1-X)}{9 \cdot 4} = \frac{X(1-X)}{4}.$$

(g) Find $\text{Var}\, Y$.

Using the iterated variance result, we have

$$\begin{aligned}\operatorname{Var} Y &= \operatorname{Var}(\mathbb{E}[Y|X]) + \mathbb{E}(\operatorname{Var}[Y|X]) \\ &= \operatorname{Var} X + \mathbb{E}[X(1-X)/4] \\ &= \frac{1}{12} + \frac{1}{4}[\mathbb{E}X - \mathbb{E}X^2] \\ &= \frac{1}{12} + \frac{1}{4}\left[\frac{1}{2} - \left(\frac{1}{12} + \frac{1}{4}\right)\right] \\ &= \frac{1}{8}.\end{aligned}$$

(h) Use the following code to generate 50,000 realizations of (X, Y). Report the mean and the variance of the Y values (check whether these support your results for $\mathbb{E}Y$ and $\operatorname{Var} Y$).

```
n <- 50000
X <- runif(n,0,1)
Y <- rbeta(n, shape1 = 3*X, shape2 = 3*(1-X))
```

I obtained

```
> var(Y)
[1] 0.124702
> mean(X)
[1] 0.5010046
```

which supports $\mathbb{E}Y = 1/2$ and $\operatorname{Var} Y = 1/8 = 0.125$.

3. Consider a game in which a player shoots 3 free throws; if the player makes i free throws, she draws one bill at random from a bag containing $i+1$ ten-dollar bills and $5-(i+1)$ one-dollar bills. Let X be the number of free throws she makes and Y be the amount of money she wins and assume that she makes free-throws with probability $1/2$.

(a) Tabulate the marginal probabilities $P(X = x)$ for $x \in \mathcal{X}$.

We have

x	0	1	2	3
$P(X = x)$	1/8	3/8	3/8	1/8

(b) Tabulate the joint probabilities $P(X = x, Y = y)$ for $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

The joint probabilities are

		$\mathcal{Y}$	
		1	10
$\mathcal{X}$	0	4/40	1/40
	1	9/40	6/40
	2	6/40	9/40
	3	1/40	4/40

We find these as

$$\begin{aligned} P(X = x, Y = y) &= P(Y = y|X = x)P(X = x) \\ &= \begin{cases} (i+1)/5 \cdot \binom{3}{x}(1/2)^x(1-1/2)^{3-x}, & y = 1 \\ (5-(i+1))/5 \cdot \binom{3}{x}(1/2)^x(1-1/2)^{3-x}, & y = 10 \end{cases} \end{aligned}$$

(c) Tabulate the marginal probabilities $P(Y = y)$ for $y \in \mathcal{Y}$.

We have

y	1	10
$P(Y = y)$	1/2	1/2

(d) Check whether X and Y are independent random variables.

We have $P(X = 1, Y = 1) = 4/40 \neq P(X = 1)P(Y = 1) = 1/8(1/2) = 1/16$, so X and Y are not independent.

(e) Compute $\mathbb{E}X$.

Since $X \sim \text{Binomial}(3, 1/2)$, we have $\mathbb{E}X = 3/2$.

(f) Compute $\mathbb{E}Y$.

We have $\mathbb{E}Y = 1(1/2) + 10(1/2) = 11/2 = 5.5$.

(g) Compute $\text{Cov}(X, Y)$.

We have

$$\begin{aligned} \mathbb{E}XY &= (1\cdot 1)(9/40) + (1\cdot 10)(6/40) + (2\cdot 1)(6/40) + (2\cdot 10)(9/40) + (3\cdot 1)(1/40) + (3\cdot 10)(4/40) \\ &= 48/5 \\ &= 9.6. \end{aligned}$$

So $(X, Y) = 48/5 - (3/2)(11/2) = 27/20 = 1.35$.

(h) Compute $\mathbb{E}[Y|X=1]$.

We may tabulate the conditional probability distribution of $Y|X=1$ as

y	1	10
$P(Y=y\|X=1)$	$9/15$	$6/15$

So we have

$$\mathbb{E}[Y|X=1] = 1\cdot\frac{9}{15} + 10\cdot\frac{6}{15} = \frac{69}{15} = \frac{23}{5}.$$

4. Let (Z_1, Z_2) be a pair of rvs with the standard bivariate Normal distribution with correlation ρ, so that their joint pdf is given by

$$f(z_1,z_2) = \frac{1}{2\pi}\frac{1}{\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2}\frac{1}{1-\rho^2}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\right] \quad \text{for all } z_1, z_2 \in \mathbb{R}.$$

(a) Show that Z_1 and Z_2 are independent if $\rho = 0$.

If $\rho = 0$ then the joint pdf of (Z_1, Z_2) becomes

$$\begin{aligned} f(z_1,z_2) &= \frac{1}{2\pi}\exp\left[-\frac{1}{2}(z_1^2+z_2^2)\right] \\ &= \frac{1}{\sqrt{2\pi}}e^{-z_1^2/2}\cdot\frac{1}{\sqrt{2\pi}}e^{-z_2^2/2}, \end{aligned}$$

so that it can be factored into the product of a function of only z_1 and a function of only z_2, implying independence of Z_1 and Z_2.

(b) Show that the marginal pdf of Z_1 is the $\text{Normal}(0,1)$ distribution.

The marginal pdf f_{Z_1} is given by

$$\begin{aligned} f_{Z_1}(z_1) &= \int_{-\infty}^{\infty}\frac{1}{2\pi}\frac{1}{\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2}\frac{1}{1-\rho^2}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\right]dz_2 \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2}\frac{1}{1-\rho^2}\left((z_2^2 - 2\rho z_1)^2 + z_2^2 - \rho^2 z_2^2\right)\right]dz_2 \\ &= \frac{1}{\sqrt{2\pi}}e^{-z_1^2/2}\underbrace{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2}\frac{1}{1-\rho^2}(z_2^2 - 2\rho z_1)^2\right]dz_2}_{=1,\text{ integral over Normal}(\rho z_1, 1-\rho^2)\text{ pdf}} \\ &= \frac{1}{\sqrt{2\pi}}e^{-z_1^2/2}, \end{aligned}$$

which is the pdf of the $\text{Normal}(0,1)$ distribution.

(c) Show that $Z_2 | Z_1 = z_1 \sim \text{Normal}(\rho z_1, 1-\rho^2)$.

In our work towards finding the marginal pdf f_{Z_1} of Z_1, we rewrote the joint pdf of Z_1 and Z_2 as

$$f(z_1, z_2; \rho) = \underbrace{\frac{1}{\sqrt{2\pi}} e^{-z_1^2/2}}_{f_{Z_1}(z_1)} \cdot \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}\frac{1}{1-\rho^2}(z_2^2 - 2\rho z_1)^2\right].$$

We see from here that the conditional pdf $f(z_2|z_1)$ of $Z_2 | Z_1 = z_1$ is given by

$$f(z_2|z_1) = \frac{f(z_1, z_2)}{f_{Z_1}(z_1)} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}\frac{1}{1-\rho^2}(z_2^2 - 2\rho z_1)^2\right],$$

which is the pdf of the $\text{Normal}(\rho z_1, 1-\rho^2)$ distribution.

(d) Show that $\text{Cov}(Z_1, Z_2) = \rho$. *Hint: Obtain $\mathbb{E}Z_1Z_2$ via iterated expectation.*

We have

$$\mathbb{E}Z_1Z_2 = \mathbb{E}(\mathbb{E}[Z_1Z_2|Z_1]) = \mathbb{E}(Z_1\mathbb{E}[Z_2|Z_1]) = \mathbb{E}(Z_1^2\rho) = \rho\mathbb{E}Z_1^2 = \rho,$$

using the fact that $\mathbb{E}Z_1^2 = 1$. Since Z_1 and Z_2 both have mean 0, $\text{Cov}(Z_1, Z_2) = \mathbb{E}Z_1Z_2 = \rho$.

(e) Show that $\text{corr}(Z_1, Z_2) = \rho$.

Since Z_1 and Z_2 both have variance equal to 1, we have

$$\text{corr}(Z_1, Z_2) = \frac{\text{Cov}(Z_1, Z_2)}{\sqrt{\text{Var } Z_1}\sqrt{\text{Var } Z_2}} = \text{Cov}(Z_1, Z_2) = \rho.$$

5. A random spectator is to be selected from the audience of a basketball game and given the chance to shoot 10 free throws. Let Y be the number of free throws made by the selected spectator and let P be the probability with which the selected spectator makes a free throw on any attempt. Assume that P and Y follow the hierarchical model

$$\begin{aligned} Y|P &\sim \text{Binomial}(10, P) \\ P &\sim \text{Beta}(2, 2). \end{aligned}$$

(a) Run a Monte Carlo simulation to generate many realizations of Y. Use the following R code:

```
S <- 10000
P <- rbeta(S,2,2)
Y <- rbinom(S,10,P)
```

Use the realizations of Y to get approximate values for

i. $\mathbb{E}Y$.

I obtained `5.009`.

ii. $\operatorname{Var} Y$.

I obtained `7.00502`.

(b) Find $\mathbb{E}[Y|P]$

We have $\mathbb{E}[Y|P] = 10P$

(c) Find $\mathbb{E}[Y]$.

We have $\mathbb{E}Y = \mathbb{E}(\mathbb{E}[Y|P]) = \mathbb{E}(10 \cdot P) = 10 \cdot 2/(2+2) = 5$.

(d) Find $\operatorname{Var}[Y|P]$.

We have $\operatorname{Var}[Y|P] = 10P(1-P)$.

(e) Find $\operatorname{Var}[Y]$.

We have

$$\begin{aligned}
\operatorname{Var} Y &= \operatorname{Var}(\mathbb{E}[Y|P]) + \mathbb{E}(\operatorname{Var}[Y|P]) \\
&= \operatorname{Var}(10 \cdot P) + \mathbb{E}(10 \cdot P(1-P)) \\
&= 100 \cdot \frac{2 \cdot 2}{(2+2)^2(2+2+1)} + 10(\mathbb{E}P - \mathbb{E}P^2) \\
&= 100 \cdot \frac{1}{20} + 10\left[\frac{1}{2} - \left(\frac{1}{20} + \frac{1}{4}\right)\right] \\
&= 7.
\end{aligned}$$

(f) Find values of α and β such that $\mathbb{E}Y = 4$ when $P \sim \text{Beta}(\alpha, \beta)$.

The value $\alpha = 4$ and $\beta = 6$ satisfy the equation

$$\mathbb{E}Y = 10 \cdot \frac{\alpha}{\alpha + \beta} = 4.$$