Stat 200A, Fall 2018
A. Adhikari and Z. Bartha

WEEK 12 EXERCISES

You are expected to do all these problems, but for **Homework 12** please turn in **only Problems 2, 3, 4, and 7** on **Thursday November 15 at the start of lecture**.

1. Warm-up

(a) For $i = 1, 2$, let V_i be such that $E(V_i) = 0$, $Var(V_i) = \sigma^2$ for some $\sigma > 0$, and $Cov(V_1, V_2) = 0$. Let $W = V_1 + V_2$. Find the best linear predictor of V_1 based on W, and find the MSE of the predictor.

The best linear predictor is

$$\hat{V_1} = \frac{Cov(V_1, W)}{Var(W)} \cdot (W - E[W]) + E[V_1] = \frac{\sigma^2}{2\sigma^2} \cdot W = \frac{W}{2}.$$

The MSE of this predictor is

$$(1 - Corr(V_1, W)^2) \cdot Var(V_1) = \left(1 - \left(\frac{1}{\sqrt{2}}\right)^2\right) \cdot \sigma^2 = \frac{\sigma^2}{2}.$$

(b) Let Z_1, Z_2, Z_3 be i.i.d. standard normal variables. Let $X_1 = Z_1 + Z_2$ and $X_2 = Z_2 + Z_3$. Find the conditional distribution of X_2 given $X_1 = x$.

Since (X_1, X_2) is bivariate normal, the conditional distribution of X_2 given $X_1 = x_1$ is normal with expectation (the linear regression estimate)

$$E[X_2 \mid X_1 = x_1] = \frac{Cov(X_1, X_2)}{Var(X_1)} \cdot (x_1 - E[X_1]) + E[X_2] = \frac{x_1}{2}$$

and variance (the MSE of the linear regression estimate)

$$Var[X_2 \mid X_1 = x_1] = (1 - Corr(X_1, X_2)^2) \cdot Var(X_2) = \left(1 - \left(\frac{1}{2}\right)^2\right) \cdot 2 = \frac{3}{2}.$$

(c) The *rms error* of regression is short for *root mean squared error*, defined as the square root of the MSE of the regression estimate. Let the joint distribution of height and weight be bivariate normal with correlation 0.6. Suppose height has mean 68 inches and SD 3 inches, and weight has mean 150 pounds and SD 15 pounds. Find the equation of the regression line for estimating height based on weight, and find the rms error of regression.

The joint distribution of the weight (X) and the height (Y) is $N(150, 68, 15^2, 3^2, 0.6)$, so the regression line is

$$\hat{Y} = Corr(X, Y) \cdot \frac{SD(Y)}{SD(X)} \cdot (X - E[X]) + E[Y] = 0.6 \cdot \frac{3}{15} \cdot (X - 150) + 68 = 0.12 \cdot X + 50.$$

The rms is

$$\sqrt{1 - Corr(X, Y)^2} \cdot SD(Y) = 0.8 \cdot 3 = 2.4$$

(d) How would your answers to Part (c) change if the five summary statistics (two means, two SDs, and correlation) were the same but the joint distribution were not bivariate normal?

The answers would not change, since the regression line and the MSE of the estimate only depend on these statistics.

2. Correlations

Let X and Y be jointly distributed random variables and let $\hat{Y} = a^*X + b^*$ be the regression prediction of Y based on X. Let ρ be the correlation between X and Y. In terms of ρ, find the correlation between

(a) X and the prediction $\hat{Y}$

$a^* = \rho\frac{\sigma_Y}{\sigma_X}$. Assuming $\rho \neq 0$, $Corr(X, \hat{Y}) = Corr(X, a^*X + b^*) = Corr(X, \rho X) = \begin{cases} 1 & \text{if } \rho > 0 \\ -1 & \text{if } \rho < 0 \end{cases}$

(b) X and the *residual* or prediction error $Y - \hat{Y}$

$Cov(X, Y - \hat{Y}) = Cov(X, Y) - Cov(X, a^*X + b^*) = \rho\sigma_X\sigma_Y - Cov(X, \rho\frac{\sigma_Y}{\sigma_X}X) = \rho\sigma_X\sigma_Y - \rho\frac{\sigma_Y}{\sigma_X}\sigma_X^2 = 0$, so $Corr(X, Y - \hat{Y}) = 0$.

(c) Y and the prediction $\hat{Y}$

$Corr(Y, \hat{Y}) = Corr(Y, a^*X + b^*) = Corr(Y, \rho X) = |\rho|$, assuming $\rho \neq 0$.

3. No-Intercept Regression

Let (X, Y) have a joint distribution that makes $Var(X)$, $Var(Y)$, and $Cov(X, Y)$ well defined and finite.

(a) Find the least squares estimate of Y among all functions of X that are linear and pass through the origin.

The estimate that we are looking for is $\hat{Y} = aX$ with the constant a that minimizes the following MSE:

$$E(Y - aX)^2 = EY^2 - 2aE[XY] + a^2EX^2.$$

By equating the derivative with respect to a to 0, we get $a = \frac{E[XY]}{EX^2} = \frac{Cov(X,Y)+(EX)(EY)}{Var(X)+(EX)^2}$.

(b) Find the MSE of the estimate in Part (a).

The MSE of this estimate by the formula above is $EY^2 - \frac{(E[XY])^2}{(EX)^2}$

(c) Use Part (b) to prove the Cauchy-Schwarz inequality: $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$.

Since the MSE is nonnegative, we have

$$\begin{aligned} 0 &\leq EY^2 - \frac{(E[XY])^2}{(EX)^2} \\ (E[XY])^2 &\leq (EY^2)(EX^2) \\ |E(XY)| &\leq \sqrt{E(X^2)E(Y^2)} \end{aligned}$$

4. Satellite Signal

This is Pitman 6.5.12.

Suppose that the magnitude of a signal received from a satellite is $S = a + bV + W$ where V is a voltage that the satellite is measuring, a and b are constants, and W is a noise term. Suppose V and W are independent and normally distributed with means 0 and variances σ_V^2 and σ_W^2.

(a) Find the correlation between S and V.

$$Corr(S,V) = \frac{Cov(S,V)}{\sigma_S \sigma_V} = \frac{Cov(a+bV+W,V)}{SD(a+bV+W)\sigma_V} = \frac{b\sigma_V^2}{\sqrt{b^2\sigma_V^2+\sigma_W^2}\cdot\sigma_V} = \frac{b}{\sqrt{b^2+\sigma_W^2/\sigma_V^2}}$$

(b) Given that $S = s$, what is the distribution of V?

Since (V,W) is bivariate normal, $(V,S) = (V, a+bV+W)$ is also bivariate normal, and thus the conditional distribution of V given $S = s$ is normal with mean

$$Corr(S,V)\cdot\frac{\sigma_V}{\sigma_S}(s-\mu_S)+\mu_V = \frac{b}{\sqrt{b^2+\sigma_W^2/\sigma_V^2}}\cdot\frac{\sigma_V}{\sqrt{b^2\sigma_V^2+\sigma_W^2}}\cdot(s-a) = \frac{b}{b^2+\sigma_W^2/\sigma_V^2}\cdot(s-a)$$

and variance

$$(1-Corr(S,V)^2)\sigma_V^2 = \left(1-\frac{b^2}{b^2+\sigma_W^2/\sigma_V^2}\right)\sigma_V^2 = \frac{\sigma_V^2\sigma_W^2}{b^2\sigma_V^2+\sigma_W^2}.$$

(c) What is the least squares estimate of V given $S = s$?

The least squares estimate is the conditional mean, which is $\frac{b}{b^2+\sigma_W^2/\sigma_V^2}\cdot(s-a)$.

(d) If the estimate is used repeatedly for different values of S coming from a sequence of independent values of V and W with the given normal distributions, what is the long run average absolute value of the error of estimation?

The error of the estimation (ε) has normal distribution with mean 0 and variance $\frac{\sigma_V^2\sigma_W^2}{b^2\sigma_V^2+\sigma_W^2}$, so the long run average of its absolute value is

$$E|\varepsilon| = \sqrt{\frac{\sigma_V^2\sigma_W^2}{b^2\sigma_V^2+\sigma_W^2}}\cdot E|Z| = \sqrt{\frac{\sigma_V^2\sigma_W^2}{b^2\sigma_V^2+\sigma_W^2}}\cdot\sqrt{\frac{2}{\pi}},$$

where we used the formula for the expected absolute value of Z, a standard normal variable.

5. Heights of Fathers and Sons

Suppose that heights of fathers and sons have a bivariate normal distribution with parameters $(68, 68, 2^2, 2^2, 0.5)$.

(a) Of the sons on the 90th percentile of heights, what percent have fathers who are above the 90th percentile of heights?

Let (X,Y) be the pair of heights of a son and his father, and let (X^*,Y^*) be the heights in standard units. Using that $Y^* = 0.5\cdot X^* + \sqrt{1-0.5^2}\cdot Z$ with $Z \sim N(0,1)$ being independent of X and that the 90th percentile of a standard normal distribution is 1.28, the probability in question is

$$\begin{aligned}\mathbb{P}(Y^* > 1.28 \mid X^* = 1.28) &= \mathbb{P}(0.5\cdot 1.28 + \sqrt{1-0.5^2}\cdot Z > 1.28)\\ &= \mathbb{P}\left(Z > \frac{0.5\cdot 1.28}{\sqrt{1-0.5^2}}\right)\\ &= 1-\Phi\left(\frac{0.5\cdot 1.28}{\sqrt{1-0.5^2}}\right) \approx 1-\Phi(0.739) \approx 0.23 = 23\%.\end{aligned}$$

(b) Of the sons of above average height, what percent are taller than their fathers?

Using again that $Y^* = 0.5 \cdot X^* + \sqrt{1-0.5^2} \cdot Z$ and that $\mu_X = \mu_Y$ and $\sigma_X = \sigma_Y$ we have

$$\begin{aligned}
\mathbb{P}(X > Y \mid X > \mu_X) &= \mathbb{P}(\sigma_X X^* + \mu_X > \sigma_Y Y^* + \mu_Y \mid X^* > 0) \\
&= \mathbb{P}(X^* > Y^* \mid X^* > 0) \\
&= \mathbb{P}(X^* > 0.5 \cdot X^* + \sqrt{1-0.5^2} \cdot Z \mid X^* > 0) \\
&= \mathbb{P}\left(\frac{0.5}{\sqrt{1-0.5^2}} \cdot X^* > Z \mid X^* > 0\right).
\end{aligned}$$

By the rotational symmetry of the joint density of (X^*, Z), this probability is

$$\frac{\arctan(0.5/\sqrt{1-0.5^2}) + \pi/2}{\pi} = \frac{2}{3}.$$

6. Random Linear Combination

This is from Stat 201A, Fall 2018.

Let $\mathbf{X}$ be an $n \times 1$ multivariate normal vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

(a) Let $\mathbf{y}$ be any non-zero $n \times 1$ vector of numbers. Show that

$$\frac{\mathbf{y}^T(\mathbf{X}-\boldsymbol{\mu})}{\sqrt{\mathbf{y}^T\boldsymbol{\Sigma}\mathbf{y}}}$$

has the standard normal distribution.

The mean vector of $\mathbf{X}-\boldsymbol{\mu}$ is $\mathbf{0}$, and hence the variance of $\mathbf{y}^T(\mathbf{X}-\boldsymbol{\mu})$ is

$$E[\mathbf{y}^T(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T\mathbf{y}] = \mathbf{y}^T E[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T]\mathbf{y} = \mathbf{y}^T\Sigma\mathbf{y}.$$

These imply that $\frac{\mathbf{y}^T(\mathbf{X}-\boldsymbol{\mu})}{\sqrt{\mathbf{y}^T\boldsymbol{\Sigma}\mathbf{y}}}$ has mean 0 and variance 1. It is also normal, since it is a linear combination of the entries of the multinomial vector $\mathbf{X}$. Therefore, $\frac{\mathbf{y}^T(\mathbf{X}-\boldsymbol{\mu})}{\sqrt{\mathbf{y}^T\boldsymbol{\Sigma}\mathbf{y}}}$ is standard normal.

(b) Now let $\mathbf{Y}$ be any non-zero $n \times 1$ random vector that is independent of X. Show that

$$\frac{\mathbf{Y}^T(\mathbf{X}-\boldsymbol{\mu})}{\sqrt{\mathbf{Y}^T\boldsymbol{\Sigma}\mathbf{Y}}}$$

has the standard normal distribution.

For any real number z

$$\mathbb{P}\left(\frac{\mathbf{Y}^T(\mathbf{X}-\boldsymbol{\mu})}{\sqrt{\mathbf{Y}^T\boldsymbol{\Sigma}\mathbf{Y}}} < z\right) = E\left[\mathbb{P}\left(\frac{\mathbf{Y}^T(\mathbf{X}-\boldsymbol{\mu})}{\sqrt{\mathbf{Y}^T\boldsymbol{\Sigma}\mathbf{Y}}} < z\right) \mid \mathbf{Y}\right] \overset{\text{part (a)}}{=} E[\Phi(z) \mid \mathbf{Y}] = \Phi(z),$$

so $\frac{\mathbf{Y}^T(\mathbf{X}-\boldsymbol{\mu})}{\sqrt{\mathbf{Y}^T\boldsymbol{\Sigma}\mathbf{Y}}}$ is standard normal.

(c) Hence show that if $\mathbf{X} = [X_1 \ X_2 \ X_3]^T$ where X_1, X_2, and X_3 are i.i.d. standard normal, then

$$\frac{X_1 e^{X_3} + X_2 \log(|X_3|)}{\sqrt{e^{2X_3} + (\log(|X_3|)^2}}$$

has the standard normal distribution.

This follows by using part (b) with $n = 2$, $\mathbf{X} = (X_1,\ X_2)^T$, $\mathbf{Y} = (e^{X_3},\ \log(|X_3|))^T$, $\boldsymbol{\mu} = \mathbf{0}$, and Σ being the 2×2 identity matrix.

7. Product of Centered Bivariate Normals

This is from Stat 201A, Fall 2016.

Let X and Y be bivariate normal $(0, 0, \sigma_X^2, \sigma_Y^2, \rho)$.

(a) Find a number a such that X and $Y - aX$ are independent.

$(X, Y - aX)$ is bivariate normal, so it is enough to make the covariance zero.

$$Cov(X, Y - aX) = Cov(X, Y) - a \cdot Cov(X, X) = \rho\sigma_X\sigma_Y - a\sigma_X^2,$$

so this is zero when $a = \rho\frac{\sigma_Y}{\sigma_X}$.

(b) Use Part (a) to find $E(XY)$ and $Var(XY)$ in terms of σ_X, σ_Y, and ρ.

$$E[XY] = E[X(Y - aX) + aX^2] = (EX)(E[Y - aX]) + aEX^2 = a\sigma_X^2 = \rho\sigma_X\sigma_Y$$

and

$$\begin{aligned}
Var(XY) &= E(XY)^2 - (E[XY])^2 = E(X(Y - aX) + aX^2)^2 - (\rho\sigma_X\sigma_Y)^2 \\
&= E[X^2(Y - aX)^2] + 2aE[X^3(Y - aX)] + a^2EX^4 - (\rho\sigma_X\sigma_Y)^2 \\
&= (EX^2)(E(Y - aX)^2) + 2a(EX^3)(E[Y - aX]) + a^2EX^4 - (\rho\sigma_X\sigma_Y)^2 \\
&= \sigma_X^2 Var(Y - aX) + a^2EX^4 - (\rho\sigma_X\sigma_Y)^2 \\
&= \sigma_X^2(1 - \rho^2)\sigma_Y^2 + a^2\sigma_X^4 \cdot 3 - (\rho\sigma_X\sigma_Y)^2 \\
&= \sigma_X^2(1 - \rho^2)\sigma_Y^2 + 3\rho^2\sigma_X^2\sigma_Y^2 - \rho^2\sigma_X^2\sigma_Y^2 \\
&= (1 + \rho^2)\sigma_X^2\sigma_Y^2
\end{aligned}$$

8. Bivariate Normal?

Let X be standard normal. Define a random variable Y by

$$Y = \begin{cases} X & \text{if } |X| \le 2 \\ -X & \text{if } |X| > 2 \end{cases}$$

(a) Show that Y is standard normal.

We will show that the cdf of Y is Φ. We will repeatedly use that $-X$ is standard normal, so as X. First, for $z < -2$

$$\mathbb{P}(Y \le z) = \mathbb{P}(-X \le z) = \Phi(z).$$

Then, for $-2 \le z \le 2$,

$$\begin{aligned}
\mathbb{P}(Y \le z) &= \mathbb{P}(Y < -2) + \mathbb{P}(-2 \le Y \le z) = \mathbb{P}(-X < -2) + \mathbb{P}(-2 \le X \le z) \\
&= \mathbb{P}(X < -2) + \mathbb{P}(-2 \le X \le z) = \mathbb{P}(X \le z) = \Phi(z).
\end{aligned}$$

Finally, for $z > 2$

$$\mathbb{P}(Y \le z) = 1 - \mathbb{P}(Y > z) = 1 - \mathbb{P}(-X > z) = \mathbb{P}(-X \le z) = \Phi(z).$$

(b) Is the joint distribution of X and Y bivariate normal? Why or why not?

No, because the joint distribution of X and Y is concentrated on the half lines $(y = -x, x < -2)$, $(y = -x, x > 2)$ and on the segment $(y = x, -2 \leq x \leq 2)$, which is impossible in the bivariate normal case.