Assessment 1

Due: Thursday – 4pm, Gradescope entry code: `948XVG`

- Please sign the honor code set up on Gradescope.
- Please upload your answers to Gradescope before 4pm.
- Start a new page for every problem.
- The only allowable aids are a double-sided sheet of notes and a calculator.
- Justify all our answers except for Question 1.
- Questions are weighted differently. The total number of points is 104.

Good luck!

1. (15 points) Which of these matrices are covariance matrices? No justification needed.

a) $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

b) $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

c) $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$

d) $\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$

e) $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$

Solution:

a) Yes, this is a covariance matrix. It is positive semi-definite with eigenvalues 0,0.

For example, take the random vector $\mathbf{X} = (X_1, X_2)$ with $X_1 = 0$ with probability 1 and $X_2 = 0$ with probability 1 (deterministic variables). Then $\mathbf{X}$ has this covariance matrix since $\text{Var}(X_1) = \text{Var}(X_2) = 0$ and $\text{Cov}(X_1, X_2) = 0$.

b) Yes, this is a covariance matrix. It is positive semi-definite with eigenvalues 2,0.

For example, take the random vector $\mathbf{X} = (X_1, X_2)$ with $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 = X_1$. Then $\mathbf{X}$ has this covariance matrix since $\text{Var}(X_1) = \text{Var}(X_2) = 1$ and $\text{Cov}(X_1, X_2) = 1$.

c) No, this is not a covariance matrix since it is not even a symmetric matrix.

d) Yes, this is a covariance matrix. It is positive semi-definite with eigenvalues 2,0.

For example, take the random vector $\mathbf{X} = (X_1, X_2)$ with $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 = -X_1$. Then $\mathbf{X}$ has this covariance matrix since $\text{Var}(X_1) = \text{Var}(X_2) = 1$ and $\text{Cov}(X_1, X_2) = -1$.

e) No, this is not a covariance matrix. You can check that the determinant is $-3 < 0$. The eigenvalues are $3, -1$, so it is not positive semi-definite. Also, $\text{Cov}(X_1, X_2) = 2$, while $\text{Var}(X_1)\text{Var}(X_2) = 1$. So this does not satisfy $\text{Cov}(X_1, X_2)^2 \leq \text{Var}(X_1)\text{Var}(X_2)$.

2. (30 points) A fraction f of the U.S. voters will vote for Trump in the 2020 election. You would like to estimate f via polling.

   a) (10 points) Suppose you randomly poll n voters. The voters are chosen uniformly and independently from the entire U.S. population of voters. Can you construct an estimator for f such that the estimate converges to f as $n \to \infty$? Make precise mathematically what your notion of convergence means.

   b) (10 points) Find an n to guarantee that your estimate is within ± 0.03 of the true fraction f with probability at least 95%. Your answer should be an actual number.

   c) (10 points) The country is divided into red states and blue states. Now suppose you poll randomly $n/2$ voters from the red states and randomly $n/2$ voters from the blue states. With only this data and no other information, can you construct an estimator for f such that the estimate converges to f? If yes, provide such an estimator. If not, state what additional information you need.

Solution:

a) For $i = 1, \dots, n$, let $X_i = \mathbb{I}\{\text{voter } i \text{ votes for Trump}\}$. Then $X_1, \dots, X_n$ are mutually independent, and

$$\mathbb{E}[X_i] = f, \qquad \mathrm{Var}(X_i) = f - f^2$$

for $i = 1, \dots, n$. Construct the estimator $\widehat{f}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

$$\mathbb{E}[\widehat{f}_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = f$$

$$\mathrm{Var}(\widehat{f}_n) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_i] = \frac{f(1-f)}{n}$$

Thus, using the law of large numbers, we know that $\widehat{f}_n$ converges to f as

$$\lim_{n\to\infty} \Pr\left(|\widehat{f}_n - f| > \epsilon\right) = 0$$

b) Using the Chebyshev inequality, we know that

$$\begin{aligned}\Pr\left(|\widehat{f}_n - f| > \epsilon\right) &\le \frac{\mathrm{Var}(\widehat{f}_n)}{\epsilon^2} \\ &= \frac{f(1-f)}{n\epsilon^2} \\ &\le \frac{1}{4n\epsilon^2}\end{aligned}$$

where the last inequality is because the maximum value of $f(1-f)$ is $1/4$.

We need to set $\epsilon = 0.03$, and make the right side of the above bound smaller than 0.05. Therefore,

$$\frac{1}{4n(0.03)^2} \leq 0.05$$
$$\implies n \geq \frac{1}{4 \times 0.05 \times (0.03)^2} \geq 5555.$$

You can also use the Chernoff bound to derive a better bound on n.

c) Using just this information, we cannot design an estimator for f. We can do so if we also knew the fraction of total voters who are from red states.

To see this, suppose we know that g fraction of the total voters are from red states. Let f^R be the fraction of red-state voters who will vote for Trump, and let f^B be the fraction of blue-state voters who will vote for Trump. Then, $f = gf^R + (1-g)f^B$. Define the estimators

$$\widehat{f}_n^R = \frac{2}{n}\sum_{i=1}^{n/2} \mathbb{I}\{\text{voter } i \text{ from red states votes for Trump}\}$$
$$\widehat{f}_n^B = \frac{2}{n}\sum_{i=1}^{n/2} \mathbb{I}\{\text{voter } i \text{ from blue states votes for Trump}\}$$

From the law of large numbers we know that $\widehat{f}_n^R \to f^R$ and $\widehat{f}_n^B \to f^B$. Thus the estimator $g\widehat{f}_n^R + (1-g)\widehat{f}_n^B \to f$.

3. (25 points) Let $\mathbf{X}$ be a Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix K given by

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix} \quad \text{and} \quad K = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 9 \end{bmatrix}.$$

a) (5 points) What is the distribution of X_1?

b) (7 points) What is the distribution of $X_1 + X_2$?

c) (7 points) What is the distribution of X_3 given $X_1 = x_1$ and $X_2 = x_2$?

d) (6 points) We wish to project $\mathbf{X}$ onto 1 dimension. Find the unit-norm vector along which the projection of $\mathbf{X}$ has the largest variance.

Solution:

a) $X_1 \sim \mathcal{N}(1, 3)$.

b) Since X_1, X_2 are jointly Gaussian, $X_1 + X_2$ is also Gaussian.

$$\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 6$$
$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2) = 8.$$

Therefore $X_1 + X_2 \sim \mathcal{N}(6, 8)$.

c) (X_1, X_2, X_3) are jointly Gaussian, and X_3 is uncorrelated with X_1 and X_2. Therefore X_3 is independent of (X_1, X_2). The distribution of X_3 given $X_1 = x_1$ and $X_2 = x_2$ is the same as the marginal distribution of X_3, that is $\mathcal{N}(2, 9)$.

d) The eigenvalues of K are 2,4,9. The largest eigenvalue is 9 and the corresponding eigenvector is $(0, 0, 1)$. This is the required vector.

4. (34 points) Let the signal S be a random variable defined as follows:

$$S = \begin{cases} 0 & \text{with probability } \frac{1}{2} \\ +1 & \text{with probability } \frac{1}{2}. \end{cases}$$

The signal is sent over a channel with additive exponential noise Z_1, i.e., Z_1 is an exponential random variable with pdf with parameter $\lambda > 0$:

$$f_{Z_1}(z) = \begin{cases} 0 & z < 0 \\ \lambda e^{-\lambda z} & z \geq 0 \end{cases}$$

The signal S and the noise Z_1 are assumed to be independent and the channel output is their sum $Y_1 = S + Z_1$.

a) (6 points) Find $f_{Y_1|S}(y|s)$ for $s = 0, +1$. Sketch the conditional pdfs on the same graph.

b) (8 points) Find the optimal decoding rule for deciding whether S is 0 or $+1$. Give your answer in terms of ranges of values of Y_1.

c) (6 points) Find the probability of decoding error of your optimal rule in terms of λ.

d) Suppose now there is an additional observation $Y_2 = S + Z_2$, where Z_2 has the same distribution as Z_1, and S, Z_1, Z_2 are mutually independent.

   i. (8 points) Find the optimal decoding rule for S given Y_1 and Y_2.

   ii. (6 points) What is the probability of decoding error of this rule? Is there any improvement over the one when there is a single observation Y_1?

Solution:

a) If $S = 0$, $Y_1 = Z_1$. Hence

$$f_{Y_1|S}(y \mid S = 0) = f_{Z_1}(y) = \begin{cases} 0 & y < 0 \\ \lambda e^{-\lambda y} & y \geq 0 \end{cases}$$

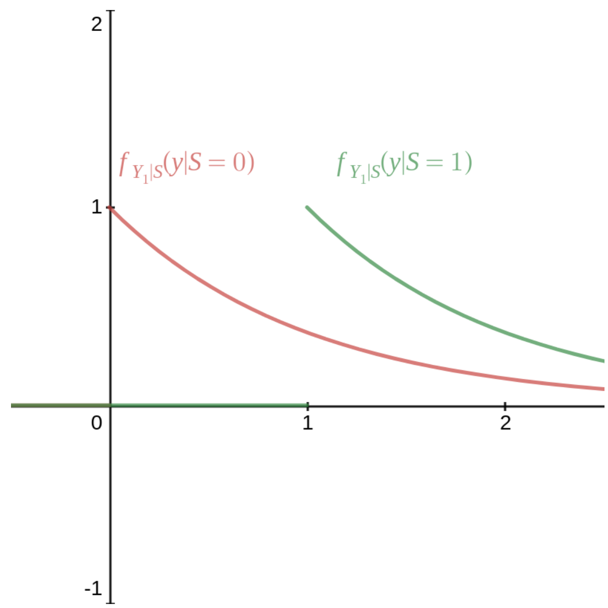
If $S = 1$, $Y_1 = 1 + Z_1$. Hence

$$f_{Y_1|S}(y \mid S = 1) = f_{Z_1}(y - 1) = \begin{cases} 0 & y < 1 \\ \lambda e^{-\lambda(y-1)} & y \geq 1 \end{cases}$$

The two pdfs are sketched in the figure below (with $\lambda = 1$ as an example).

b) Suppose we observe $Y_1 = y$. The MAP decoding rule is

$$\frac{f_{Y_1|S}(y \mid S = 1)}{f_{Y_1|S}(y \mid S = 0)} \; \substack{\geq^{\widehat{S}(y)=1} \\ <_{\widehat{S}(y)=0}} \; 1$$



From the sketch in part a), $f_{Y_1|S}(y \mid S=1)$ is larger whenever $y \geq 1$ and $f_{Y_1|S}(y \mid S=0)$ is larger whenever $y < 1$. So the MAP rule becomes

$$y \underset{<_{\widehat{S}(y)=0}}{\overset{\geq^{\widehat{S}(y)=1}}{}} 1$$

c) When $S = 1$, $Y_1 = 1 + Z_1 \geq 1$, so we always declare $\widehat{S}(Y_1) = 1$. Therefore $\Pr(\text{Error} \mid S = 1) = 0$. When $S = 0$, there is an error when $Y_1 \geq 1$. Therefore

$$\begin{aligned}
\Pr(\text{Error}) &= \Pr(\text{Error} \mid S=0)\Pr(S=0) + \Pr(\text{Error} \mid S=1)\Pr(S=1) \\
&= \frac{1}{2}\Pr(Y_1 \geq 1 \mid S=0) \\
&= \frac{1}{2}\int_1^\infty f_{Y_1|S}(y \mid S=0)dy \\
&= \frac{1}{2}\int_1^\infty \lambda e^{-\lambda y} dy \\
&= \frac{1}{2}e^{-\lambda}
\end{aligned}$$

d) If $S = 0$, $Y_1 = Z_1$ and $Y_2 = Z_2$. The joint pdf of $Y_1, Y_2 \mid S = 0$ is

$$f_{Y_1,Y_2|S}(y_1, y_2 \mid S=0) = f_{Z_1}(y_1)f_{Z_2}(y_2) = \begin{cases} \lambda^2 e^{-\lambda(y_1+y_2)} & y_1 \geq 1, y_2 \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

If $S = 1$, $Y_1 = 1 + Z_1$ and $Y_2 = 1 + Z_2$. Then

$$f_{Y_1,Y_2|S}(y_1, y_2 \mid S = 1) = f_{Z_1}(y_1 - 1) f_{Z_2}(y_2 - 1) = \begin{cases} \lambda^2 e^{-\lambda(y_1+y_2-2)} & y_1 \geq 1, y_2 \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

i. Suppose we observe $Y_1 = y_1, Y_2 = y_2$. The MAP decoding rule is

$$\frac{f_{Y_1,Y_2|S}(y_1, y_2 \mid S = 1)}{f_{Y_1,Y_2|S}(y_1, y_2 \mid S = 0)} \begin{array}{c} \geq^{\widehat{S}(y_1,y_2)=1} \\ <_{\widehat{S}(y_1,y_2)=0} \end{array} 1$$

Using the joint pdfs derived above, this rule becomes

$$\widehat{S}(y_1, y_2) = \begin{cases} 1 & y_1 \geq 1, y_2 \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

ii. When $S = 1$, $Y_1 \geq 1$ and $Y_2 \geq 1$. Again $\Pr(\text{Error} \mid S = 1) = 0$. When $S = 0$, we get an error when $Y_1 \geq 1$ and $Y_2 \geq 1$. Therefore

$$\begin{aligned} \Pr(\text{Error}) &= \Pr(\text{Error} \mid S = 0) \Pr(S = 0) + \Pr(\text{Error} \mid S = 1) \Pr(S = 1) \\ &= \frac{1}{2} \Pr(Y_1 \geq 1, Y_2 \geq 1 \mid S = 0) \\ &= \frac{1}{2} \Pr(Y_1 \geq 1 \mid S = 0) \Pr(Y_2 \geq 1 \mid S = 0) \\ &= \frac{1}{2} e^{-2\lambda} \end{aligned}$$

The probability of error has decreased when we have taken two independent observations.