Midterm #2 - Probability and Statistics II

1. (30 points) Suppose that X_1, X_2, X_3 are i.i.d., normal random variables with mean 0 and variance 1. Let

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}.$$

(a) What is the covariance matrix for the random vector X?

> **Solution:** Since all of the X_i's are independent, $\text{Cov}(X_i, X_j) = 0$, unless $i = j$. So all of the entries of the covariance matrix are zero, except for the entries of the diagonal. The entries on the diagonal are all one since each X_i has variance equal to 1. Therefore Σ_X is the 3-by-3 identity matrix
>
> $$\Sigma_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

(b) What is the covariance matrix for the random vector Y?

> **Solution:** $\Sigma_Y = A\Sigma_X A^T = AA^T$ where A is the matrix in the problem. Calculating AA^T gives:
>
> $$\Sigma_Y = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

(c) Are Y_1 and Y_2 independent random variables?

> **Solution:** No, because $\text{Cov}(Y_1, Y_2) = -1 \neq 0$. The covariance of Y_1 and Y_2 is the entry in row 1, column 2 of Σ_Y.

(d) What is the variance of Y_1?

> **Solution:** $\text{Var}(Y_1) = \text{Cov}(Y_1, Y_1) = 2$. The variance of Y_1 is the entry in row 1, column 1 of Σ_Y.

2. (20 points) The following linear equation has no solutions.

$$\begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}.$$

Show that $b = \begin{bmatrix} 3/2 \\ 3/2 \end{bmatrix}$ is the least squares solution.

Solution: One way to show this is to show that b satisfies the normal equations $X^T X b = X^T y$. Calculating both sides gives:

$$X^T X b = \begin{bmatrix} 3 & -1; -1 & 1 \end{bmatrix} \begin{bmatrix} 3/2 \\ 3/2 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix},$$

which is the same thing you get if you calculate:

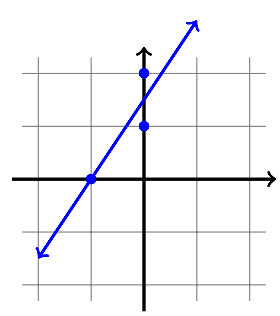
$$X^T y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}.$$

3. (20 points) The least squares solution $b = \begin{bmatrix} 3/2 \\ 3/2 \end{bmatrix}$ to the system of equations

$$\begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

corresponds to a line that is the best fit regression line for three points in $\mathbb{R}^2$. Draw and label a graph showing the line and the three points. Does the line intersect any of the three points?

Solution:



Yes, the line $y = \frac{3}{2} + \frac{3}{2}x$ does intersect the point $(-1, 0)$.

4. (15 points) Suppose that x and y are two orthogonal vectors in $\mathbb{R}^n$ and $\|x\| = \|y\| = 1$. Let $P = xx^T + yy^T$. Calculate P^2 and simplify as much as you can.

Solution:

$$\begin{aligned} P^2 = (xx^T + yy^T)(xx^T + yy^T) &= xx^T xx^T + xx^T yy^T + yy^T xx^T + yy^T yy^T \\ &= x(x^T x)x^T + x(x^T y)y^T + y(y^T x)x^T + y(y^T y)y^T \end{aligned}$$

Since x and y are orthogonal, $x^T y = y^T x = 0$, so the two middle terms are zero. Since $\|x\| = \|y\| = 1$, the end terms become: $xx^T + yy^T$, which means that $P^2 = P$.

5. (15 points) Suppose that you roll an ordinary six-sided die. Let X be the number it lands on and let $Y = 6$ if the result is a six and 0 otherwise. What is the covariance of X and Y?

Solution: $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$. We know that $E(X) = 3.5$ and $E(Y) = 1$. Also, $E(XY) = 6$, since XY is either 36 if the die lands on six, or 0 otherwise. Therefore,

$$\mathrm{Cov}(X, Y) = 6 - 3.5 = 2.5 = \frac{5}{2}.$$

Formulas

- **Normal Equations** To find the least squares solution of $Xb = y$, solve

$$X^T X b = X^T y.$$

- **Inverse of 2-by-2 Matrix** If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

- **Covariance** For random variables X and Y,

$$\text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y)).$$

  Also,

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

- **Covariance Matrix** For a random vector $X = (X_1, \ldots, X_n)^T$,

$$\Sigma_X = \begin{bmatrix} \text{Cov}(X_1, X_1) & \ldots & \text{Cov}(X_1, X_n) \\ \vdots & & \vdots \\ \text{cov}(X_n, X_1) & \ldots & \text{Cov}(X_n, X_n) \end{bmatrix}.$$

- **Covariance Matries & Linear Transformations** If X and Y are random vectors such that $Y = AX$ where $A \in \mathbb{R}^{m \times n}$, then

$$\Sigma_Y = A \Sigma_X A^T.$$

- **Multivariate Normal Density**

$$f_X(x) = \frac{1}{(2\pi)^{n/2} |\Sigma_X|^{1/2}} \exp\left(-\tfrac{1}{2}(x - \mu_X)^T \Sigma_X^{-1} (x - \mu_X)\right).$$