DSE 210 (Probability and Statistics Using Python)
Homework 3, due Friday 2/19 9:00am

Instructions

Please follow these instructions when completing your assignment:

- Please upload your written answers to Gradescope by the due date. **Late submissions will not be graded**.
- You can write up your answers using pencil and paper or using document editing software (LaTeX, Word, etc...). If you write your answers using pencil and paper, you can scan your answers and upload the resulting file to Gradescope or take pictures of each page and upload those.
- Please associate each problem with a page on your gradescope submission
- For written answers you are not required to show work. However, showing work will enable better feedback.
- Collaboration is encouraged, but all submissions should be in your own writing and completed with your own understanding.

Problems

- **Worksheet 6**: $1, 2, 3, 4, 5, 6$
- **Worksheet 7**: All

Worksheet 6 Solutions

1. (a) You could make a reasonable argument for positive or negative correlation. Out of curiosity I computed to correlation for Seattle over the last year and found it was 0.42 which is a moderate positive correlation.

   (b) These are negatively correlated. If it's warm and sunny people will go to the beach but there won't be snow for skiing.

   (c) Depending on how you think social security numbers work, these could be negatively correlated or uncorrelated. If SSNs are just assigned randomly, then they would be uncorrelated. If they are assigned sequentially, then older people would have a lower SSN and so they would be negatively correlated.

2. (a) Let's compute the marginal distributions:

$X \backslash Y$	-1	0	-1	$\Pr(X = x)$
-1	$-1/8$	0	$1/8$	$1/4$
0	0	$1/2$	0	$1/2$
1	$1/8$	0	$1/8$	$1/4$
$\Pr(Y = y)$	$1/4$	$1/2$	$1/4$	

   Now we can see $\Pr(X = 0, Y = 0) = 1/2$, but $\Pr(X = 0)\Pr(Y = 0) = 1/4$ so the two random variables are not independent.

   (b) We know $\texttt{Corr}(X, Y) = \texttt{Cov}(X, Y)/\texttt{Std}(Y)\texttt{Std}(Y)$. Let's start by computing:

$$\texttt{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[XY]$$

Since $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. Now:

$$\mathbb{E}[XY] = \sum_{x \in X} \sum_{y \in Y} \Pr(X = x, Y = y)(xy) = 0$$

Since $\texttt{Cov}(X, Y) = 0$, we know $\texttt{Corr}(X, Y) = 0$. Therefore, the random variables are dependent but have a correlation of zero!

3. (a) By definition $\texttt{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. But we know $Y = 2X$. Therefore: $\mathbb{E}[XY] = \mathbb{E}[X \cdot 2X] = 2\mathbb{E}[X]^2$, and $\mathbb{E}[X]\mathbb{E}[2X] = 2\mathbb{E}[X]^2$. So:

$$\texttt{Cov}(X, Y) = 2\mathbb{E}[X^2] - 2\mathbb{E}[X]^2 = 2\texttt{Var}(X) = 200.$$

(b) Recall $\texttt{Corr}(X, Y) = \texttt{Cov}(X, Y)/\texttt{Std}(X)\texttt{Std}(Y)$. We know $\texttt{Std}(X) = 10$ so $\texttt{Std}(Y) = \texttt{Std}(2X) = 20$. Then: $\texttt{Corr}(X, Y) = 200/200 = 1$ which makes sense because Y is a function of X.

4. (a) In general, a bivariate Gaussian over a pair of variables X and Y is specified by a mean and a covariance matrix:

$$\mu = \begin{pmatrix} \mu_x & \mu_y \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} \texttt{Var}(X) & \texttt{Cov}(X, Y) \\ \texttt{Cov}(X, Y) & \texttt{Var}(Y) \end{pmatrix}$$

We know $\mu_x = 2$, $\texttt{Var}(X) = 1$ and $\mu_y = 4$, $\texttt{Var}(Y) = 0.25$. So it remains to compute:

$$\texttt{Cov}(X, Y) = \texttt{Corr}(X, Y) \cdot (\texttt{Std}(X)\texttt{Std}(Y)).$$

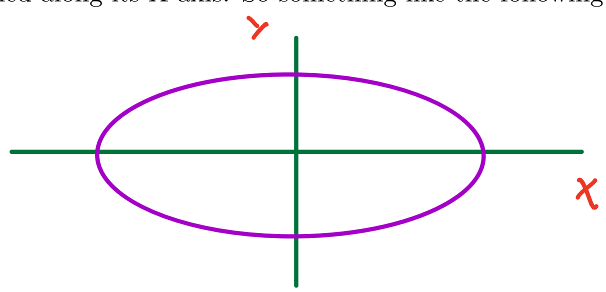
Since we know $\texttt{Corr}(X, Y) = -0.5$ we can determine $\texttt{Cov}(X, Y) = -0.25$. So, we have:

$$\mu = \begin{pmatrix} 2 & 4 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} 1 & -0.25 \\ -0.25 & 0.25 \end{pmatrix}.$$

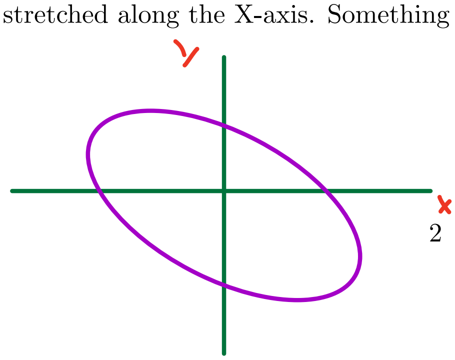
(b) Following the same process as above, and noting that $Y = X$ which implies $\texttt{Corr}(X, Y) = 1 \Rightarrow \texttt{Cov}(X, Y) = 1$ since $\texttt{Var}(X) = \texttt{Var}(Y) = 1$:

$$\mu = \begin{pmatrix} 1 & 1 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

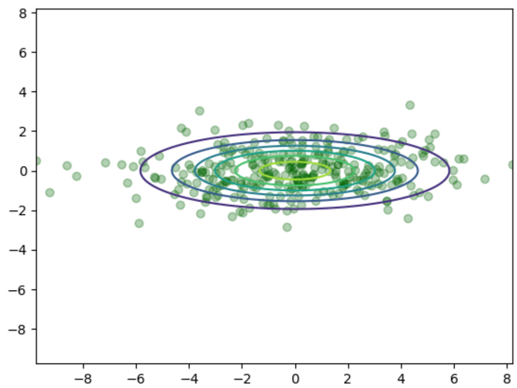
1. (a) Since $\texttt{Cov}(X, Y) = 0$, the ellipse is axis-aligned. Since $\texttt{Var}(X) > \texttt{Var}(Y)$, we know the distribution is more stretched along its X-axis. So something like the following would be reasonable:

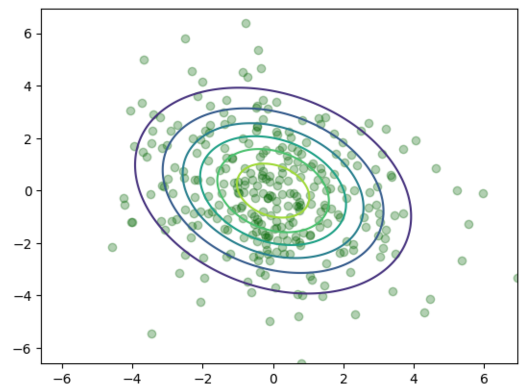


(b) Now, we know $\texttt{Corr}(X, Y) = -1/4$, so the ellipse should have a shallow downward slope and be more stretched along the X-axis. Something like the following would be reasonable:



2. Here's are some example plots:





And the code I used:

```python
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt


def main():
    μ = np.array([0,0])
    Σ1 = np.array([[9,0],[0,1]])
    plot_gaussian(μ, Σ1, "1")

    Σ2 = np.array([[4,-1], [-1,4]])
    plot_gaussian(μ, Σ2, "2")


def plot_gaussian(μ, Σ, stub=""):
    dist = stats.multivariate_normal(mean=μ, cov=Σ)
    samples = dist.rvs(size=300)
    
    xi = np.linspace(samples.min(),samples.max(),300)
    X,Y = np.meshgrid(xi, xi)
    Z = dist.pdf(np.c_[X.ravel(), Y.ravel()]).reshape(X.shape)

    plt.close()
    plt.scatter(samples[:,0], samples[:,1], c="darkgreen", alpha=0.3)
    plt.contour(X, Y, Z)
    plt.savefig("gaussians_{}.png".format(stub), bbox_inches="tight")
```

Worksheet 7 Solutions

1. For some vector $x \in \mathbb{R}^d$, we can turn it into a unit vector by dividing each entry by its norm: $x = x/\|x\|$. In this case, $\|x\|^2 = 1 + 4 + 9 = 14$. So the unit length vector in the direction of x is:

$$x = (\frac{1}{\sqrt{14}}, \frac{2}{\sqrt{14}}, \frac{3}{\sqrt{14}}) = \frac{1}{\sqrt{14}}(1, 2, 3)$$

2. Let $x = (1, 1)$. We're looking for vectors u_i such that $u_i \cdot x = 0$ and $\|u_i\| = 1$. Here are two candidates: $\hat{u}_1 = (-1, 1)$ and $\hat{u}_2 = (1, -1)$. We want unit vectors though so let's normalize them: $u_1 = \frac{1}{\sqrt{2}}(1, -1), u_2 = \frac{1}{\sqrt{2}}(-1, 1)$. For a d-dimensional Euclidean space, a set of d orthogonal vectors

spans the entire space, and so these are the only two orthogonal unit vectors, since any other vectors are a linear combination of these two.

3. Since $x \cdot x = 25$, we know $\|x\| = \sqrt{x \cdot x} = 5$. Therefore, this is the set of all points at a distance of 5 from the origin. So these points form the surface of the d-dimensional ball of radius 5. For example, in $\mathbb{R}^2$ this would be a circle of radius 5 and in $\mathbb{R}^3$ it would be a sphere of radius 5.

4. $w = (2, -1, 6)$

5. For A and B to be conformable (e.g. can be multiplied) we know: $\underset{r\times c}{C} = \underset{r\times n}{A}\underset{n\times c}{B}$. Therefore, A is 10×30 and B is 30×20.

6. (a) We're stacking up n vectors, each of which have d coordinates into a matrix X, so $X \in \mathbb{R}^{n\times d}$
   (b) $C = \underset{n\times d}{X}\underset{d\times n}{X^T}$, so $C \in \mathbb{R}^{n\times n}$.
   (c) Let $C = AB$ for a pair of conformable matrices A, B. In general, C_{ij} is the dot-product of the i-th row of A with the j-th column of B. So in this case, let $A = X$ and $B = X^T$. So $C_{ij} = x_i \cdot x_j$.

7. By the associative property of matrix-multiplication: $x^T x x^T x x^T x = (x^T x)(x^T x)(x^T x) = \|x\|^6 = 10^6$.

8. By convention, we typically assume x is a column vector. Under this assumption: $x^T x = \|x\|^2 = 1 + 9 + 25 = 37$. Now:
$$xx^T = \begin{pmatrix} 1 & 3 & 5 \\ 3 & 9 & 15 \\ 5 & 15 & 25 \end{pmatrix}.$$

9. Let θ be the angle between x and y. We know $\cos(\theta) = (x \cdot y)/(\|x\|\|y\|)$. We know $x^T y = x \cdot y = 2$ and $\|x\| = \|y\| = 2$. Therefore:
$$\cos(\theta) = \frac{2}{4} = \frac{1}{2} \Rightarrow \theta = 1.0472$$
using the inverse cosine function.

10. We're looking for a matrix M such that $x^T M x = f(x)$. In general, for some $M \in \mathbb{R}^3$
$$x^T M x = M_{11}x_1^2 + M_{22}x_2^2 + M_{33}x_3^2 + (M_{21} + M_{12})x_1x_2 + (M_{31} + M_{13})x_1x_3 + (M_{32} + M_{33})x_1x_3$$
But we know M must be symmetric so $M_{ij} = M_{ji}$. Then the above simplifies to:
$$x^T M x = M_{11}x_1^2 + M_{22}x_2^2 + M_{33}x_3^2 + 2M_{21}x_1x_2 + 2M_{31}x_1x_3 + 2M_{32}x_1x_3$$
So we should pick:
$$M = \begin{pmatrix} 3 & 1 & -2 \\ 1 & 0 & 0 \\ -2 & 0 & 6 \end{pmatrix}$$

11. Let A_i denote the i-th row of A and A^j be the j-th column.
   (a) We first recall the following useful property: for a pair of conformable matrices A, B, $(AB)^T = B^T A^T$. Now let $Z = X^T X$ for some arbitrary matrix $X \in \mathbb{R}^{r\times c}$. Then $Z^T = (X^T X)^T = X^T X = Z$. Therefore, Z **is symmetric**.
   (b) Using the same property as before: let $Z = XX^T$. Then $Z^T = (XX^T)^T = XX^T = Z$. Therefore, Z **is symmetric**.
   (c) $Z_{ij} = A_{ij} + A_{ji} = A_{ji} + A_{ij} = Z_{ji}$. Therefore Z **is symmetric**.

(d) Counterexample:

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

But:

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

12. (a) The determinant of a diagonal matrix is just the product of the elements on its diagonal. Therefore $|A| = 8! = 40320$

    (b) A matrix A^{-1} is the inverse of A if $AA^{-1} = A^{-1}A = I$. Let $A^{-1} = \text{diag}((1, 1/2, 1/3, 1/4, 1/5, 1/6, 1/7, 1/8))$. Then we can verify that the previous property holds. In general, to invert a diagonal matrix we just need to invert the individual elements on the diagonal.

13. (a) Let $W = UU^T$. Then $W_{ij} = u_i \cdot u_j$. Since $\{u_i\}_{i=1}^d$ forms an orthonormal set (a collection of vectors that are mutually orthogonal and unit-length), we know $W_{ij} = 1$ if $i = j$ and 0 otherwise. Therefore $W = I_d$, the $d \times d$ identity matrix.

    (b) Since $UU^T = U^TU = I$, $U^{-1} = U^T$ (a useful property of orthonormal matrices)

14. A matrix is singular if its determinant is zero. For a 2×2 matrix A:

$$\left| \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \right| = a_{11}a_{22} - a_{21}a_{22}$$

Then in this case $|A| = z - 6$, so we should pick $z = 6$.