Homework 1

Zhaoxia Yu

2023-04-11

Due on April 17 2023

Problem 1

- Suppose X_1, X_2, Y_1, Y_2 are mutually independent.
  - X_1 and X_2 are iid from $N(\mu = 0, \sigma_x^2 = 2^2)$
  - Y_1 and Y_2 are iid from $N(\mu = 0, \sigma_y^2 = 1^2)$ Consider the two pairs (X_1, X_2) and (Y_1, Y_2). Which pair tends to have a larger difference? To answer the question, please calculate and estimate the following two probabilities:

$$P(|X_1 - X_2| > 4), P(|Y_1 - Y_2| > 4)$$

- The hints for calculating/estimating $P(|X_1 - X_2| > 4)$ can be found in the two slides. Using similar strategies, you can calculate/estimate $P(|Y_1 - Y_2| > 4)$

Calculate $P(|X_1 - X_2| > 4)$

- Hints for calculating $P(|X_1 - X_2| > 4)$.
  - First find the distribution of $X_1 - X_2$. Then standard it to have mean 0 and SD 1.
  - Second, express the probability to $P(|Z| > z)$, where $Z \sim N(0, 1)$.
  - Next, expression the probability in terms of $\Phi(\cdot)$, the CDF of the standard normal distribution.
  - Last, use the "pnorm" function in R to find the numerical value.

Estimate $P(|X_1 - X_2| > 4)$

- The probability can be estimated by doing simulations/sampling.
- If you sample many (say 10,000) pairs of X_1 and X_2, count how many pairs satisfying $|X_1 - X_2| > 4$. The probability can be used to estimate $P(|X_1 - X_2| > 4)$

Problem 2

- Find a matrix A such that AY gives the difference of mean vectors between iris setosa and iris versicolor
- Find a matrix B such that YB is column-standardized, i.e., the standard deviation of each column/feature is 1.
- Check the following
  - Let $C = \mathbf{I}_{150} - \frac{1}{150}J$, where $J_{150\times 150}$ is an all-ones matrices . Use R to verify that CY centers each column/feature. The R code for C is "

```
C=diag(1,150) - \frac{150}matrix(1, 150, 150)
```

– Let S be the sample covariance matrix. Use R to verify that each column of $CYS^{-1/2}$ has been centered and standardized (in fact, the columns have also been de-correlated). Hints:

```
S=cov(Y)
```

To compute $S^{-1/2}$, you may need an R package, such as "qtl2pleio".

Problem 3

- Choose a picture you like and conduct approximations using singular value decomposition (SVD).