IC252 Lab 7

April 25, 2019

1. Three fair coins of value 10, 20 and 50 are tossed. Only one face of the coin has the denomination displayed. Thus after tossing, some coins show the denomination, while the others do not show it. The visible denominations are then added to give the total amount A.

   (a) What is the expected value of A? Determine the answer by simulation using the code snippet below. The answer is 40.

   (b) What is the expected value of A given that the denominations of exactly two coins are visible?

      i. Work out the answer by hand calculation. The answer is $\frac{60+70+30}{3} = 53.3$.

      ii. Determine the answer by simulation.

```
import pandas as pd
import numpy as np

# Generate a DataFrame to hold the data
d = pd.DataFrame(columns = ['50','20','10'])

numTrials = 10
d.x1 = np.random.randint(0,2,numTrials)
d.x2 = np.random.randint(0,2,numTrials)
d.x5 = np.random.randint(0,2,numTrials)

# print the DataFrame
print(d)
```

2. Casinos design gambling games such that there is always a built-in statistical advantage for the casino over players. Consider the following games (a simplified version of the game called *craps*):

   (a) A fair die is rolled. If an odd number turns up, the player wins an equivalent amount. If an even number turns up, the player loses an equivalent amount. Is this game good for the player or for the casino?

   (b) The game is same as earlier, with odd and even reversed i. e. even number means player wins an equivalent amount and odd means the player loses. Is this game good for the player or for the casino?

   (c) The game is same as in part (b), but the die is now biased as follows: $p(1) = p(3) = p(5) = 0.19$, $p(2) = p(4) = p(6) = 0.143$.

   Simulate the three games by playing it $N = 10,000$ times. In each case, determine the amount the player will earn (a negative earning means the player loses money.)

3. The provided file `data.txt` has 2000 2-D data points. The data is termed mutivariate when more than one variable is being observed. For this question, each observation has two variables (the data is

2-dimensional.) For the multivariate datapoint $\mathbf{x}$, having d dimensions, the Gaussian density function is

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

Here $\boldsymbol{\mu}$ is the d-dimensional mean vector, and $\boldsymbol{\Sigma}$ is the $d\times d$ covariance matrix. The diagonal elements of the covariance matrix is the variance of each variable, and the off-diagonal elements are the covariances between the variables.

(a) Plot the data using a scatter plot.

(b) Compute the sample mean and sample covariance matrix.

Determine the probability that each data point was sampled from the following:

(c) G1: A Gaussian distribution with mean $\begin{bmatrix}0\\0\end{bmatrix}$ and covariance matrix $\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$

(d) G2: A Gaussian distribution with mean $\begin{bmatrix}2.5\\2.5\end{bmatrix}$ and covariance matrix $\begin{bmatrix}0.5 & 0\\0 & 0.5\end{bmatrix}$

Plot the probability of all the data points for both cases. The plot for a subset of the data is shown below.