a4paper, total=170mm,257mm, left=20mm, top=20mm,

Coursework #1

Alexandre Allani
aa4719@ic.ac.uk

October 31, 2019

1 Statistics and Probabilities

1.1 Question 1

We have the following dataset:

$$D = \left\{ d_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, d_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, d_3 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\}$$

Then, the sample mean :

$$\mu = E[D] = \frac{1}{3} \sum_{1}^{3} d_n = \frac{1}{3} \begin{bmatrix} -4 \\ 6 \\ 5 \end{bmatrix}$$

For the covariance matrix we have :

$$\begin{aligned}
\Sigma_D &= \frac{1}{3} \sum_{1}^{3} (d_n - \mu)(d_n - \mu)^T \\
&= \frac{1}{3}[\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} -4 \\ 6 \\ 5 \end{bmatrix} \right) \left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} -4 \\ 6 \\ 5 \end{bmatrix} \right)^T \\
&\quad + \left(\begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} -4 \\ 6 \\ 5 \end{bmatrix} \right) \left(\begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} -4 \\ 6 \\ 5 \end{bmatrix} \right)^T \\
&\quad + \left(\begin{bmatrix} -4 \\ 4 \\ 2 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} -4 \\ 6 \\ 5 \end{bmatrix} \right) \left(\begin{bmatrix} -4 \\ 4 \\ 2 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} -4 \\ 6 \\ 5 \end{bmatrix} \right)^T] \\
&= \frac{1}{3} \left[\begin{bmatrix} 64/9 & 48/9 & 8/9 \\ 48/9 & 4 & 6/9 \\ 8/9 & 6/9 & 1/9 \end{bmatrix} + \begin{bmatrix} 1/9 & 6/9 & 5/9 \\ 6/9 & 4 & 30/9 \\ 5/9 & 30/9 & 25/9 \end{bmatrix} + \begin{bmatrix} 49/9 & 0 & 28/9 \\ 0 & 0 & 0 \\ 28/9 & 0 & 16/9 \end{bmatrix} \right] \\
&= \frac{1}{3} \begin{bmatrix} 38/3 & -6 & 5/3 \\ -6 & 8 & 4 \\ 5/3 & 4 & 14/3 \end{bmatrix} \\
&= \begin{bmatrix} 38/9 & -2 & 5/9 \\ -2 & 8/3 & 4/3 \\ 5/9 & 4/3 & 14/9 \end{bmatrix}
\end{aligned} \tag{1}$$

Hence the sample mean and covariance are :

$$\boxed{\mu = \frac{1}{3} \begin{bmatrix} -4 \\ 6 \\ 5 \end{bmatrix} \text{ and } \Sigma_D = \begin{bmatrix} 38/9 & -2 & 5/9 \\ -2 & 8/3 & 4/3 \\ 5/9 & 4/3 & 14/9 \end{bmatrix}}$$

1.2 Question 2

Let's take one dataset $D = \{(x_1, x_2)\}$ We know that it have the following mean vector:

$$mu = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Then $\mu_1 = E[X_1] = -1$ and $\mu_2 = E[X_2] = 1$. Moreover we know that the marginal variances are $\sigma_1^2 = 2$ and $\sigma_2^2 = 0.5$. This means that $Var[X_1] = 2$ and $Var[X_2] = 0.5$. We want to generate two datasets with different shapes. To do that, we choose to generate one dataset with a uniform distribution, and the second with a gaussian distribution.

For the uniform distribution, we generate 100 points $X = (x_1, ;x_{100})$ randomly on the interval [-10,20]. We can then compute its sample mean μ and it variance σ^2.
Then $Z_1 = \frac{\sigma_1}{\sigma}(X - \mu) + \mu_1$ has a mean of μ_1 and a variance of σ_1. With the same transformation, we can have another set of points Z_2 with a mean equals to μ_2 and a variance of σ_2. For the gaussian distribution we do the same thing

This gives us the following plot (in order to see the two different shapes, I generated 1000 points for each dataset)

2 Graphical Models

According to the question, the likelihood follow a binomial distribution $B(20, \mu)$, the prior that the student decided to choose is the conjugate distribution $\beta(\mu; \alpha, \beta)$ where $\beta = \alpha = 2$. We want to compute the posterior distribution based on these assumptions. According to Bayes theorem we know that:

$$posterior \propto prior \times likelihood$$

Let $h = 6$ the number of success, $N = 20$ the number of tries. Hence:

$$\begin{aligned} P(\mu | x = h, N, \alpha, \beta) &\propto p(x|\mu) \times p(\mu|\alpha, \beta) \\ &\propto \mu^h (1-\mu)^{N-h} \mu^{\alpha-1} (1-\mu)^{\beta-1} \\ &\propto \mu^{h+\alpha-1} (1-\mu)^{N-h+\beta-1} \end{aligned} \tag{2}$$

We need to find the normalization factor of this distribution. However we know that the Beta distribution $\beta(\mu; h+\alpha, N-h+\beta)$ is normalized and also proportional to $\mu^{h+\alpha-1}(1-\mu)^{N-h+\beta-1}$

Hence the posterior distribution follow a beta distribution

$$\boxed{\beta(\mu; h+\alpha, N-h+\beta) = \beta(\mu; 8, 16)}$$

We obtain then the following plot We can see that the mean and the variance have changed between the two distribution. The posterior distribution is less uncertain than the prior distribution, since the variance is smaller. We are more certain about where in its range μ lies.

3 Graphical Model

3.1 Question 1

The graphical model of (3) is chown in **Figure 3.1**

$$p(a,b,c,d,e,f) = p(a|b,c)p(c|b)p(d)p(e|d,a)p(f|c,d,e)p(b) \tag{3}$$

3.2 Question 2

A REFAIRE FLEMME LA

(a) All path going from a to g are blocked by e, because arrows meet head to head on e, and none of e's descendant are observed. Hence:

$$\boxed{a \perp\!\!\!\perp g|k}$$

(b) The path $d \to j \to h$ is not blocked. Hence:

$$\boxed{d \not\perp\!\!\!\perp h|i}$$

(c) All path going from g to a are blocked by e. Arrows meet head to head on e, and e is not observed. Hence:

$$\boxed{g \perp\!\!\!\perp a|c}$$

(d) The path $e \to d \to j$ is not blocked because k is an observed descendant of d. Hence:

$$\boxed{e \not\perp\!\!\!\perp j|k}$$

(e) The path $b \to d \to e$ is not blocked. Hence

$$\boxed{b \not\perp\!\!\!\perp e|j}$$

(f) The path $j \to h \to i \to c$ is not blocked. Hence:

$$\boxed{j \not\perp\!\!\!\perp c|\{k,g\}}$$

(g) The path $a \to b \to d \to k$ is not blocked. Hence:

$$\boxed{a \not\perp\!\!\!\perp k}$$

(h) The path $a \to b \to d \to k$ is not blocked. Hence:

$$\boxed{a \not\perp\!\!\!\perp k | e}$$

(i) The path $h \to i \to d$ is not blocked Hence:

$$\boxed{h \not\perp\!\!\!\perp d | j, e}$$

(j) The path $b \to j \to h$ is not blocked (because e is a descendant of j). Hence:

$$\boxed{b \not\perp\!\!\!\perp h | e}$$

(k) The path $h \to i \to c$ is not blocked. Hence:

$$\boxed{h \not\perp\!\!\!\perp c | d}$$

(l) The path $a \to b \to j \to f$ is not blocked (because k is a descendant of j). Hence:

$$\boxed{a \not\perp\!\!\!\perp f | k}$$

(m) The path $i \to h \to j \to b \to a$ is not blocked. Hence:

$$\boxed{i \not\perp\!\!\!\perp a | j}$$

(n) The path $g \to e \to j \to d \to h$ is not blocked. Hence:

$$\boxed{g \not\perp\!\!\!\perp h | e}$$

(o) Each path from g to f is going through d. The arrows meet head to tail there, and d is observed, so all paths are blocked. Hence:

$$\boxed{g \perp\!\!\!\perp f | d}$$

(p) Each path from g to h is going through e. Arrows meet head to head on e, and none of e's descendant are observed. All paths are blocked, hence:

$$\boxed{g \perp\!\!\!\perp h}$$

(q) Each path from a to i are going through j or d. Arrows meet head to head on both nodes, and both nodes don't have any observed descendant.Hence:

$$\boxed{a \perp\!\!\!\perp i | g}$$

(r) Each path from b to h are going through j or d. Arrows meet head to head on both nodes, and both nodes don't have any observed descendant. Hence:

$$\boxed{b \perp\!\!\!\perp h}$$

(s) Each path from b to h are going through j or d. Arrows meet head to head on both nodes, and both nodes don't have any observed descendant. Hence:

$$\boxed{b \perp\!\!\!\perp h | g}$$

(t) Each path from i to a are going through j or d. Arrows meet head to head on both nodes, and both nodes don't have any observed descendant.Hence:

$$\boxed{i \perp\!\!\!\perp a}$$

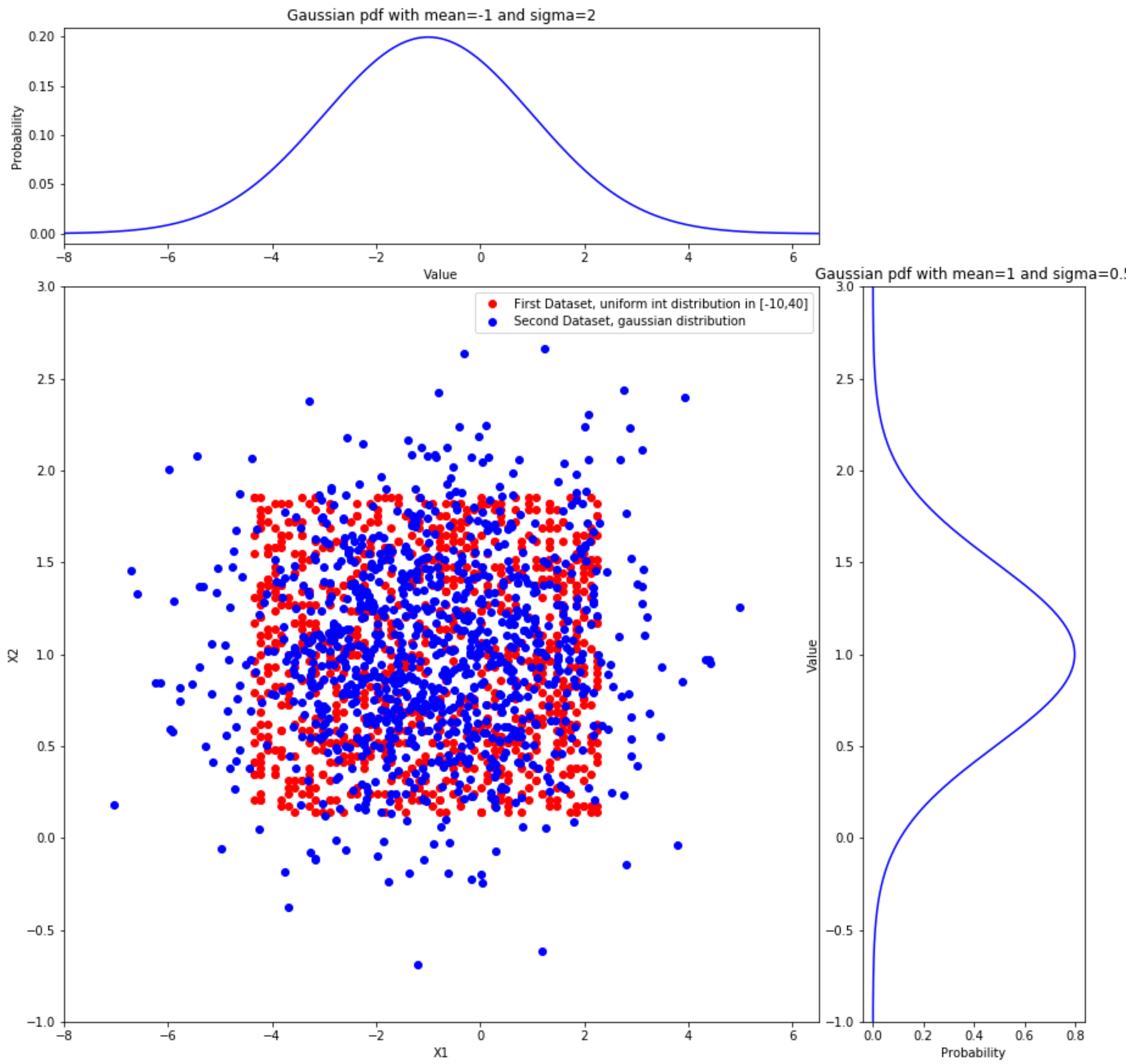

Figure 1: Generated Dataset

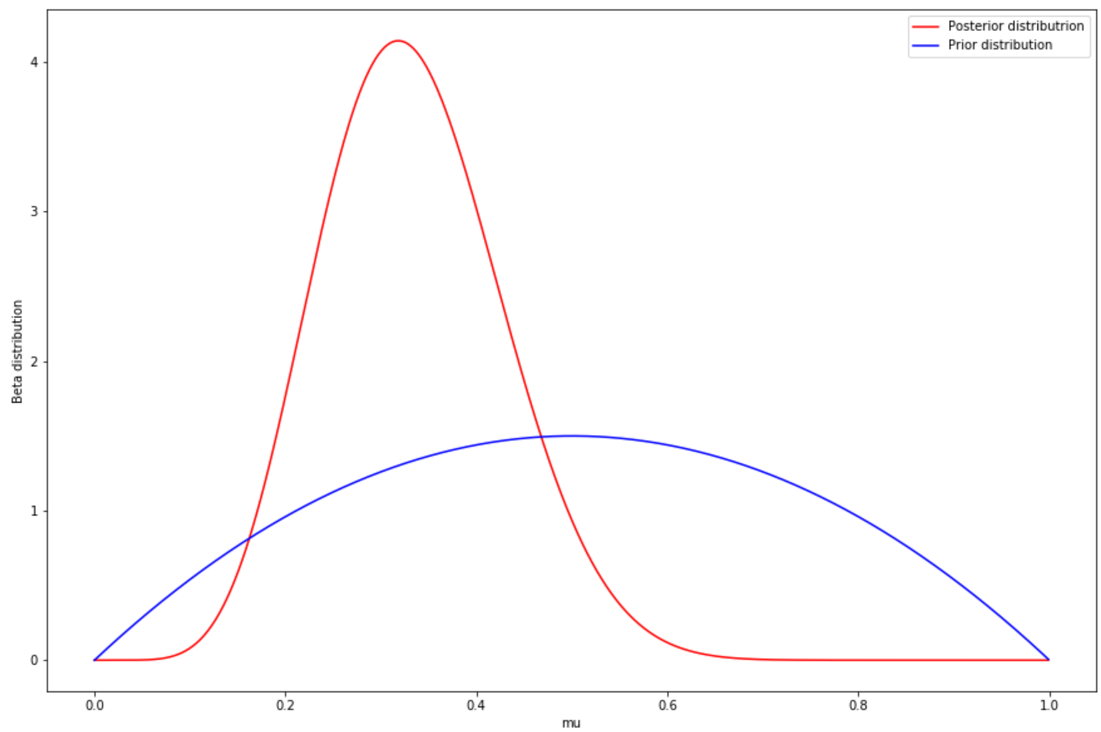

Figure 2: Posterior and Prior distributions

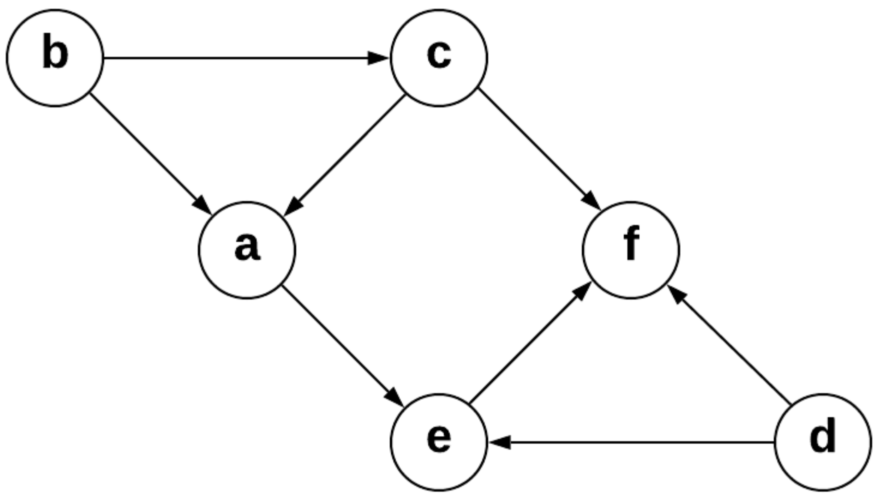

Figure 3: Posterior and Prior distributions