Probability and Statistics for Data Analysis

Assignment 2

Vasileios Galanos
MSc in Data Science(PT)
p3351902

November 17, 2019

Exercise 1.

We know that the joint probability density function of X and Y is given by

$$f_{X,Y}(x,y) = \frac{2}{c^2} I_A(x,y)$$

1.1

Obviously it holds that $f_{X,Y}(x,y) > 0 \qquad \forall (x,y) \in A$, cause of the indicator function I_A and $c > 0$.
Next, we need to check whether $\int \int_A f_{X,Y}(x,y) = 1$.

Indeed,

$$\begin{aligned}\iint_A f_{X,Y}(x,y) &= \int_0^c \left(\int_0^y \frac{2}{c^2} dx \right) dy = \frac{2}{c^2} \int_0^c \left(\int_0^y (x)' dx \right) dy \\ &= \frac{2}{c^2} \int_0^c (y-0) dy = \frac{2}{c^2} \frac{1}{2} \int_0^c (y^2)' dy = \frac{1}{c^2} [y^2]|_0^c \quad = \frac{c^2}{c^2} = 1\end{aligned}$$

So $f_{X,Y}(\cdot,\cdot)$ is a probability density function.

1.2

By definition,

$$\begin{aligned}
P(X \leq \frac{c}{2}, Y \leq \frac{c}{2}) &= \int_0^{\frac{c}{2}} \left(\int_y^{\frac{c}{2}} \frac{2}{c^2} dx \right) dy = \frac{2}{c^2} \int_0^{\frac{c}{2}} \left(\int_y^{\frac{c}{2}} (x)' dx \right) dy \\
&= \frac{2}{c^2} \int_0^{\frac{c}{2}} \left(\frac{c}{2} - y \right) dy = \frac{2}{c^2} \left[\int_0^{\frac{c}{2}} \frac{c}{2} dy - \int_0^{\frac{c}{2}} y \, dy \right] \\
&= \frac{2}{c^2} \left[\frac{c}{2} \int_0^{\frac{c}{2}} (y)' dy - \frac{1}{2} \int_0^{\frac{c}{2}} (y^2)' dy \right] = \frac{2}{c^2} \left[\frac{c}{2} (\frac{c}{2} - 0) - \frac{1}{2} (\frac{c^2}{4} - 0) \right] \\
&= \frac{2}{c^2} \frac{c^2}{8} = \frac{1}{4} = 0.25
\end{aligned}$$

1.3

For $x \in (0, c)$ we have that $x < y < c$. Thus, the marginal distribution of X is given by

$$f_X(x) = \int_{y \in \mathbb{R}} f_{X,Y}(x, y) = \int_x^c \frac{2}{c^2} dy = \frac{2}{c^2} \left[y \right]_x^c = \frac{2(c - x)}{c^2}$$

For $y \in (0, c)$ we have that $0 < x < y$. Thus, the marginal distribution of Y is given by

$$f_Y(y) = \int_{x \in \mathbb{R}} f_{X,Y}(x, y) = \int_0^y \frac{2}{c^2} dx = \frac{2}{c^2} \left[x \right]_0^y = \frac{2y}{c^2}$$

Now, if X and Y are independent random variables then

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

However,

$$f_X(x) f_Y(y) = \frac{2(c - x)}{c^2} \frac{2y}{c^2} = \frac{4(c - x)y}{c^4} \neq f_{X,Y}(x, y)$$

so X and Y are not independent. Obviously they are not identical as well.

1.4

Previously, we calculated the marginal distributions of X and Y. We will use them to calculate the marginal probabilities $P(X \leq c/2)$ and $P(Y \leq c/2)$ as follows:

$$\begin{aligned} P(X \leq c/2) &= \int_0^{\frac{c}{2}} f_X(x)dx = \int_0^{\frac{c}{2}} \frac{2(c-x)}{c^2}dx = \frac{2}{c^2}\int_0^{\frac{c}{2}} (c-x)dx \\ &= \frac{2}{c^2}\left[c\int_0^{\frac{c}{2}} dx - \int_0^{\frac{c}{2}} xdx\right] = \frac{2}{c^2}\left[c[x]|_0^{\frac{c}{2}} - [x^2]|_0^{\frac{c}{2}}\right] \\ &= \frac{2}{c^2}\left[\frac{c^2}{2} - \frac{1}{2}\frac{c^2}{4}\right] = \frac{2}{c^2} \times \frac{3c^2}{8} = \frac{3}{4} = 0.75 \end{aligned}$$

Similarly for $P(Y \leq c/2)$,

$$\begin{aligned} P(Y \leq c/2) &= \int_0^{\frac{c}{2}} f_Y(y)dy = \frac{2}{c^2}\int_0^{\frac{c}{2}} y\, dy = \frac{2}{c^2} \times \frac{1}{2}[y]^2|_0^{\frac{c}{2}} \\ &= \frac{1}{c^2} \times \frac{c^2}{4} = \frac{1}{4} = 0.25 \end{aligned}$$

1.5

The covariance function is given by

$$Cov(X,Y) = E(XY) - E(X)E(Y) \tag{1}$$

The first term on the right hand side is equal to

$$\begin{aligned} E(XY) &= \iint_A xy f_{X,Y}(x,y)dxdy = \int_0^c \left(\int_0^y xy\frac{2}{c^2}dx\right)dy \\ &= \frac{2}{c^2}\int_0^c y\left(\int_0^y x\, dx\right)dy = \frac{2}{c^2}\int_0^c y\frac{1}{2}\left([x^2]|_0^y\right)dy \\ &= \frac{1}{c^2}\int_0^c y^3\, dy = \frac{1}{4c^2}[y^4]|_0^c = \frac{c^4}{4c^2} = \frac{c^2}{4} \end{aligned} \tag{2}$$

The expectation of X is

$$\begin{aligned} E(X) &= \int_{x\in A} x f_X(x)dx = \int_0^c x\frac{2}{c^2}(c-x)dx \\ &= \frac{2}{c^2}\int_0^c (cx - x^2)dx = \frac{2}{c^2}\left[\frac{c}{2}[x^2]|_0^c - \frac{1}{3}[x^3]|_0^c\right] \\ &= \frac{2}{c^2}\left[\frac{c}{2}c^2 - \frac{1}{3}c^3\right] = \frac{2}{c^2}\times\frac{c^3}{6} = \frac{c}{3} \end{aligned} \tag{3}$$

In a similar manner, we have that

$$\begin{aligned} E(Y) &= \int_{y\in A} y f_Y(y)dy = \int_0^c y\frac{2y}{c^2}dy = \frac{2}{c^2}\int_0^c y^2 dy \\ &= \frac{2}{c^2}\frac{1}{3}[y^3]|_0^c = \frac{2}{3c^2}c^3 = \frac{2c}{3} \end{aligned} \tag{4}$$

Substituting Equations (2), (3) and (4) into (1) we obtain that

$$Cov(X,Y) = \frac{c^2}{4} - \frac{c}{3}\times\frac{2c}{3} = \frac{c^2}{36}$$

The Correlation function is given by

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \tag{5}$$

The variance of X is

$$\begin{aligned} Var(X) &= \int_{x\in A}\left[x - E(X)\right]^2 f_X(x)dx = \int_0^c (x-\frac{c}{3})^2\frac{2}{c^2}(c-x)dx \\ &= \frac{2}{c^2}\int_0^c -x^3 + \frac{5c}{3}x^2 - \frac{7c^2}{9}x + \frac{c^3}{9}\,dx \\ &= \frac{2}{c^2}\left[-\frac{1}{4}[x^4]|_0^c + \frac{5c}{9}[x^3]|_0^c - \frac{7c^2}{18}[x^2]|_0^c + \frac{c^3}{9}[x]|_0^c\right] \\ &= \frac{2}{c^2}\left(-\frac{c^4}{4} + \frac{5c^4}{9} - \frac{7c^4}{18} + \frac{c^4}{9}\right) = \frac{c^2}{18} \end{aligned} \tag{6}$$

The variance of Y is

$$\begin{aligned} Var(Y) &= \int_{y\in A} [y - E(Y)]^2 f_Y(y)dy = \int_0^c (y - \frac{2c}{3})^2 \frac{2y}{c^2} dy \\ &= \frac{2}{c^2}\int_0^c y^3 - \frac{4c}{3}y^2 + \frac{4c^2}{9}y\, dy = \frac{2}{c^2}\left[\frac{1}{4}[y^4]|_0^c - \frac{4c}{9}[y^3]|_0^c + \frac{4c^2}{18}[y^2]|_0^c\right] \\ &= \frac{2}{c^2}\left(\frac{c^4}{4} - \frac{4c^4}{9} + \frac{4c^4}{18}\right) = \frac{c^2}{18} \end{aligned} \tag{7}$$

Substituting Equations (6) and (7) into (5) we obtain that

$$\rho(X,Y) = \frac{\frac{c^2}{36}}{\sqrt{\frac{c^2}{18}}\sqrt{\frac{c^2}{18}}} = \frac{\frac{c^2}{36}}{\frac{c^2}{18}} = \frac{1}{2} = 0.5$$

1.6

The conditional distribution of X given Y = y is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\frac{2}{c^2}}{\frac{2y}{c^2}} = \frac{1}{y} \quad , 0 < x < y$$

The conditional distribution of Y given X = x is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{\frac{2}{c^2}}{\frac{2(c-x)}{c^2}} = \frac{1}{c-x} \quad , x < y < c$$

1.7

The situation is that we observe $X = x$ and we want to predict Y. If we will attemp to predict Y by h(X)The best predictor under squared error loss is the conditional expectation of Y given X.

$$\begin{aligned} E[Y|X = x] &= \int_x^c y f_{Y|X}(y|x)dy = \int_x^c \frac{y}{c-x}dy \\ &= \frac{1}{2(c-x)}[y^2]|_x^c = \frac{x^2 - x^2}{2(c-x)} = \frac{c+x}{2} \end{aligned}$$

So, if Shooter 1 informs Shooter 2 that the target is spotted at coordinate $X = x$ then the best position for Shooter 2 in order to minimize the square error loss from the target is $\frac{c+x}{2}$

Exercise 2.

We want to calculate the probability $P(\bar{X}_n \leq 2)$, where $\bar{X}_n = \sum_{i=1}^{n} X_i/n$. Thus,

$$P(\bar{X}_n \leq 2) = P(\sum_{i=1}^{n} X_i \leq 2n) \qquad \text{,for } n = 2, 5, 10, 20, 40$$

We also know that $X_1, \ldots\ldots X_n$ are independent and identically distributed. We can also calculate the probability using the Central Limit Theorem if our sample follows a distribution with a finite variance.

The CLT states that,

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \overset{n\to\infty}{\longrightarrow} N(0,1)$$

Applying that in our case, we have that:

$$P(\bar{X}_n \leq 2) = P(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq \frac{2-\mu}{\sigma/\sqrt{n}}) = P(Z \leq \frac{2-\mu}{\sigma/\sqrt{n}}) \qquad \text{,for } n = 2, 5, 10, 20, 40$$

2.1

Using exact calculations

We know that the sample follows the Binomial distribution, $X_i \sim \mathcal{B}(10, 0.1)$ and because the sample is independent we can use the property:

$$\sum_{i=1}^{n} X_i \sim \mathcal{B}(\sum_{i=1}^{n} v_i, p)$$

in order to calculate the probability as follows:

$$\begin{aligned} P(\sum_{i=1}^{n} X_i \leq 2n) &= P\left(\mathcal{B}(\sum_{i=1}^{n} 10, 0.1) \leq 2n\right) = P\left(\mathcal{B}(10n, 0.1) \leq 2n\right) \\ &= \sum_{k=0}^{2n} \binom{10n}{k} 0.1^k (1-0.1)^{10n-k} \end{aligned}$$

For $n = 2$, $P(\bar{X}_n \leq 2) \approx 0.95683$
For $n = 5$, $P(\bar{X}_n \leq 2) \approx 0.99065$
For $n = 10$, $P(\bar{X}_n \leq 2) \approx 0.99919$
For $n = 20$, $P(\bar{X}_n \leq 2) \approx 0.99999$
For $n = 40$, $P(\bar{X}_n \leq 2) \approx 1.00000$

Using Central Limit Theorem approximation

The Binomial distribution $\mathcal{B}(n, p)$ has,

$$E[x] = np \qquad Var[X] = np(1-p)$$

Using CLT for an *iid* sample $X_i \sim \mathcal{B}(10, 0.1)$ we get that

$$\mu = E[x] = 1 \qquad \sigma = \sqrt{Var[X]} = \sqrt{0.9}$$

Thus,

$$P(\bar{X}_n \leq 2) = P(Z \leq \frac{2-1}{\sqrt{0.9}/\sqrt{n}}) \qquad \text{,for } n = 2, 5, 10, 20, 40$$

For $n = 2$, $P(\bar{X}_n \leq 2) \approx 0.93198$
For $n = 5$, $P(\bar{X}_n \leq 2) \approx 0.99079$
For $n = 10$, $P(\bar{X}_n \leq 2) \approx 0.99957$
For $n = 20$, $P(\bar{X}_n \leq 2) \approx 1.00000$
For $n = 40$, $P(\bar{X}_n \leq 2) \approx 1.00000$

2.2

Using exact calculations

The sample follows a Poisson distribution, $X_i \sim \mathcal{P}(1.9)$ and because the sample is independent we can use the property:

$$\sum_{i=1}^{n} X_i \sim \mathcal{P}(\sum_{i=1}^{n} \lambda_i)$$

in order to calculate the probability as follows:

$$P(\sum_{i=1}^{n} X_i \leq 2n) = P\left(\mathcal{P}(\sum_{i=1}^{n} 1.9) \leq 2n\right) = P\left(\mathcal{P}(1.9n) \leq 2n\right)$$
$$= \sum_{k=0}^{2n} \frac{e^{-1.9n} 1.9n^k}{k!}$$

For $n = 2$, $P(\bar{X}_n \leq 2) \approx 0.66784$
For $n = 5$, $P(\bar{X}_n \leq 2) \approx 0.64533$
For $n = 10$, $P(\bar{X}_n \leq 2) \approx 0.64717$
For $n = 20$, $P(\bar{X}_n \leq 2) \approx 0.66569$
For $n = 40$, $P(\bar{X}_n \leq 2) \approx 0.70200$

Using Central Limit Theorem approximation

The Poisson distribution $\mathcal{P}(\lambda)$ has,

$$E[x] = \lambda \qquad Var[X] = \lambda$$

Using CLT for an *iid* sample $X_i \sim \mathcal{P}(1.9)$ we get that

$$\mu = E[x] = 1.9 \qquad \sigma = \sqrt{Var[X]} = \sqrt{1.9}$$

Thus,

$$P(\bar{X}_n \leq 2) = P(Z \leq \frac{2 - 1.9}{\sqrt{1.9}/\sqrt{n}}) \qquad \text{,for } n = 2, 5, 10, 20, 40$$

For $n = 2$, $P(\bar{X}_n \leq 2) \approx 0.54086$
For $n = 5$, $P(\bar{X}_n \leq 2) \approx 0.56443$
For $n = 10$, $P(\bar{X}_n \leq 2) \approx 0.59073$
For $n = 20$, $P(\bar{X}_n \leq 2) \approx 0.62720$
For $n = 40$, $P(\bar{X}_n \leq 2) \approx 0.67682$

2.3

Using exact calculations

The sample follows a Negative Binomial distribution, $X_i \sim \mathcal{NB}(5, 0.65)$ and because the sample is independent we can use the property:

$$\sum_{i=1}^{n} X_i \sim \mathcal{NB}(\sum_{i=1}^{n} v_i, p)$$

in order to calculate the probability as follows:

$$P(\sum_{i=1}^{n} X_i \leq 2n) = P\left(\mathcal{NB}(\sum_{i=1}^{n} 5, 0.65) \leq 2n\right) = P\left(\mathcal{NB}(5n, 0.65) \leq 2n\right)$$
$$= \sum_{k=0}^{2n} \binom{5n+k-1}{k} 0.65^{5n}(1-0.65)^k$$

For $n = 2$, $P(\bar{X}_n \leq 2) \approx 0.42272$
For $n = 5$, $P(\bar{X}_n \leq 2) \approx 0.27160$
For $n = 10$, $P(\bar{X}_n \leq 2) \approx 0.15803$
For $n = 20$, $P(\bar{X}_n \leq 2) \approx 0.06436$
For $n = 40$, $P(\bar{X}_n \leq 2) \approx 0.01320$

Using Central Limit Theorem approximation

The Negative Binomial distribution $\mathcal{NB}(r, p)$ has,

$$E[x] = \frac{r(1-p)}{p} \qquad Var[X] = \frac{r(1-p)}{p^2}$$

Using CLT for an *iid* sample $X_i \sim \mathcal{NB}(5, 0.65)$ we get that

$$\mu = E[x] = \frac{1.75}{0.65} \qquad \sigma = \sqrt{Var[X]} = \frac{\sqrt{1.75}}{0.65}$$

Thus,

$$P(\bar{X}_n \leq 2) = P(Z \leq \frac{2 - \frac{1.75}{0.65}}{\frac{\sqrt{1.75}}{0.65}/\sqrt{n}}) \qquad \text{,for } n = 2, 5, 10, 20, 40$$

For $n = 2$, $P(\bar{X}_n \leq 2) \approx 0.31523$
For $n = 5$, $P(\bar{X}_n \leq 2) \approx 0.22344$
For $n = 10$, $P(\bar{X}_n \leq 2) \approx 0.14103$
For $n = 20$, $P(\bar{X}_n \leq 2) \approx 0.06410$
For $n = 40$, $P(\bar{X}_n \leq 2) \approx 0.01572$

2.4

Using exact calculations

We know that the sample follows the Gamma distribution, $X_i \sim \mathcal{G}(5, 0.5)$ and because the sample is independent we can use the property:

$$\sum_{i=1}^{n} X_i \sim \mathcal{G}(\sum_{i=1}^{n} \alpha_i, \beta)$$

in order to calculate the probability as follows:

$$\begin{aligned} P(\sum_{i=1}^{n} X_i \leq 2n) &= P\left(\mathcal{G}(\sum_{i=1}^{n} 5, 0.5) \leq 2n\right) = P\left(\mathcal{G}(5n, 0.5) \leq 2n\right) \\ &= \int_0^{2n} \frac{1}{\Gamma(5n) 0.5^{5n}} k^{5n-1} e^{-\frac{k}{0.5}} dk \end{aligned}$$

For $n = 2$, $P(\bar{X}_n \leq 2) \approx 0.28338$
For $n = 5$, $P(\bar{X}_n \leq 2) \approx 0.15677$
For $n = 10$, $P(\bar{X}_n \leq 2) \approx 0.07034$
For $n = 20$, $P(\bar{X}_n \leq 2) \approx 0.01711$
For $n = 40$, $P(\bar{X}_n \leq 2) \approx 0.00127$

Using Central Limit Theorem approximation

The Gamma distribution $\mathcal{G}(\alpha, \beta)$ has,

$$E[x] = \alpha\beta \qquad Var[X] = \alpha\beta^2$$

Using CLT for an *iid* sample $X_i \sim \mathcal{G}(5, 0.5)$ we get that

$$\mu = E[x] = 2.5 \qquad \sigma = \sqrt{Var[X]} = \sqrt{1.25}$$

Thus,

$$P(\bar{X}_n \leq 2) = P(Z \leq \frac{2 - 2.5}{\sqrt{1.25}\sqrt{n}}) \qquad \text{,for } n = 2, 5, 10, 20, 40$$

For $n = 2$, $P(\bar{X}_n \leq 2) \approx 0.26354$
For $n = 5$, $P(\bar{X}_n \leq 2) \approx 0.15866$
For $n = 10$, $P(\bar{X}_n \leq 2) \approx 0.07865$
For $n = 20$, $P(\bar{X}_n \leq 2) \approx 0.02275$
For $n = 40$, $P(\bar{X}_n \leq 2) \approx 0.00234$

2.5

Using exact calculations

We know that the sample follows the Chi-Squared distribution, $X_i \sim \mathcal{X}_1^2$. We also know that the X_v^2 distribution is a special case of the Gamma distribution $\mathcal{G}(v/2, 2)$.
In our case, $\mathcal{X}_1^2$ is a special case of $\mathcal{G}(0.5, 2)$. and because the sample is independent we can use the property:

$$\sum_{i=1}^{n} X_i \sim \mathcal{G}(\sum_{i=1}^{n} \alpha_i, \beta)$$

in order to calculate the probability as follows:

$$P(\sum_{i=1}^{n} X_i \leq 2n) = P\left(\mathcal{G}(\sum_{i=1}^{n} 0.5, 2) \leq 2n\right) = P\left(\mathcal{G}(0.5n, 2) \leq 2n\right)$$
$$= \int_0^{2n} \frac{1}{\Gamma(0.5n)2^{0.5n}} k^{0.5n-1} e^{-\frac{k}{2}} dk$$

For $n = 2$, $P(\bar{X}_n \leq 2) \approx 0.86466$
For $n = 5$, $P(\bar{X}_n \leq 2) \approx 0.92476$
For $n = 10$, $P(\bar{X}_n \leq 2) \approx 0.97075$
For $n = 20$, $P(\bar{X}_n \leq 2) \approx 0.99500$
For $n = 40$, $P(\bar{X}_n \leq 2) \approx 0.99982$

Using Central Limit Theorem approximation

The Chi-Squared distribution $\mathcal{X}_v^2$ has,

$$E[x] = v \qquad Var[X] = 2v$$

Using CLT for an *iid* sample $X_i \sim \mathcal{X}_1^2$ we get that

$$\mu = E[x] = 1 \qquad \sigma = \sqrt{Var[X]} = \sqrt{2}$$

Thus,

$$P(\bar{X}_n \leq 2) = P(Z \leq \frac{2-1}{\sqrt{2}\sqrt{n}}) \qquad \text{,for } n = 2, 5, 10, 20, 40$$

For $n = 2$, $P(\bar{X}_n \leq 2) \approx 0.84134$
For $n = 5$, $P(\bar{X}_n \leq 2) \approx 0.94308$
For $n = 10$, $P(\bar{X}_n \leq 2) \approx 0.98733$
For $n = 20$, $P(\bar{X}_n \leq 2) \approx 0.99922$
For $n = 40$, $P(\bar{X}_n \leq 2) \approx 1.00000$

Exercise 3.

3.1

To show that the joint probability $f_{\mathbf{X}}(\mathbf{x}|\theta)$ is member of the exponential family of distributions we must show that:

$$f_{\mathbf{X}}(\mathbf{x}|\theta) = h(x)c(\theta)e^{\sum_{i=1}^{k} w_i(\theta)t_i(x)}$$

In our case, $X_i \sim \mathcal{P}(\theta a_i)$, $\theta \in \Theta = (0, \infty)$ and $a_i > 0$. We also know that $\mathbf{X} = (X_1, X_2, \ldots, X_n)$. Indeed, it is a member of the exponential family because

$$\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}|\theta) &= f_{X_1}(x|\theta) f_{X_2}(x|\theta) \ldots f_{X_n}(x|\theta) \\
&= \frac{e^{-\theta a_1}(\theta a_1)^{x_1}}{x_1!} \frac{e^{-\theta a_2}(\theta a_2)^{x_2}}{x_2!} \cdots \frac{e^{-\theta a_n}(\theta a_n)^{x_n}}{x_n!} \\
&= \frac{e^{-\theta \sum_{i=1}^{n} a_i} \prod_{i=1}^{n} \theta^{x_i} \prod_{i=1}^{n} a_i^{x_i}}{\sum_{i=1}^{n} x_i!} \\
&= \frac{e^{log(\theta^{\sum_{i=1}^{n} x_i})} \prod_{i=1}^{n} a_i^{x_i} e^{-\theta \sum_{i=1}^{n} a_i}}{\sum_{i=1}^{n} x_i!} \\
&= \frac{e^{\sum_{i=1}^{n} x_i log(\theta)} \prod_{i=1}^{n} a_i^{x_i} e^{-\theta \sum_{i=1}^{n} a_i}}{\sum_{i=1}^{n} x_i!} \\
&= \frac{\prod_{i=1}^{n} a_i^{x_i}}{\sum_{i=1}^{n} x_i!} e^{-\theta \sum_{i=1}^{n} a_i} e^{\sum_{i=1}^{n} x_i log(\theta)}, \qquad where
\end{aligned}$$

- $h(x) = \dfrac{\prod_{i=1}^{n} a_i^{x_i}}{\sum_{i=1}^{n} x_i!} \geq 0$,a real valued function of observation x
- $t_1(x) = x_1, \ldots t_n(x) = x_n$,real valued functions of observation x
- $c(\theta) = e^{-\theta \sum_{i=1}^{n} a_i}$, a real valued function of the parameter θ
- $w_1(\theta) = w_2(\theta) = \cdots = w_n(\theta) = log(\theta)$ real valued functions of the parameter θ.

3.2

We have already shown that $X_1, X_2, \ldots, X_n$ is a random sample from a pdf $f_{\mathbf{X}}(\mathbf{x}|\theta)$ that belongs to an exponential family. Thus,

$$T(\mathbf{X}) = \left(\sum_{j=1}^{n} t_1(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j) \right)$$

But in our case, $t_i(X_j) = X_j$ so we have that,

$$T(\mathbf{X}) = \left(\sum_{j=1}^{n} X_j, \ldots, \sum_{j=1}^{n} X_j \right) = \sum_{j=1}^{n} X_j$$

is a sufficient statistic for θ.

3.3

Firstly, we find that our estimator $\delta(\mathbf{X})$ is unbiased because:

$$\begin{aligned} E[\delta(\mathbf{X})] = E\left[\frac{\sum X_i}{\sum \alpha_i}\right] &= \frac{\sum E[X_i]}{\sum \alpha_i} \overset{X_i \sim \mathcal{P}(\theta a_i)}{=\!=\!=} \\ &= \frac{\sum \theta a_i}{\sum \alpha_i} = \theta \frac{\sum a_i}{\sum \alpha_i} = \theta \end{aligned} \tag{1}$$

Next we calculate tha variance of our estimator:

$$\begin{aligned} Var[\delta(\mathbf{X})] = Var\left[\frac{\sum X_i}{\sum \alpha_i}\right] &\overset{Var(cX)=c^2 Var(X)}{=\!=\!=} \\ &= \frac{\sum Var(X_i)}{(\sum \alpha_i)^2} = \frac{\sum \theta a_i}{(\sum \alpha_i)^2} = \frac{\theta}{\sum a_i} \end{aligned} \tag{2}$$

Now that we know that our estimator is unbiased we will check whether its variance achieves the Cramer-Rao lower bound or not.

The Cramer-Rao inequality states that,

$$Var_\theta[\delta(\mathbf{X}] \geq \frac{(\frac{\partial}{\partial \theta} E_\theta[\delta(\mathbf{X}])^2}{E_\theta\left[(\frac{\partial}{\partial \theta} log(f_{\mathbf{X}}(\mathbf{x}|\theta)))^2\right]}$$

We first calculate the numerator,

$$(\frac{\partial}{\partial\theta}E_\theta[\delta(\mathbf{X}])^2 \overset{(1)}{=} (\frac{\partial}{\partial\theta}1)^2 = 1$$

Next for the denominator,

- $$log(f_{\mathbf{X}}(\mathbf{x})) = log\left(\frac{\prod_{i=1}^n a_i^{x_i}}{\sum_{i=1}^n x_i!}e^{-\theta\sum_{i=1}^n a_i}e^{\sum_{i=1}^n x_i log(\theta)}\right)$$
$$= log\left(\frac{\prod_{i=1}^n a_i^{x_i}}{\sum_{i=1}^n x_i!}\right) + log(\theta)\sum_{i=1}^n x_i - \theta\sum_{i=1}^n a_i$$

- $$\frac{\partial}{\partial\theta}log(f_{\mathbf{X}}(\mathbf{x}|\theta)) = 0 + \frac{\sum x_i}{\theta} - \sum a_i$$

- $$(\frac{\partial}{\partial\theta}log(f_{\mathbf{X}}(\mathbf{x}|\theta)))^2 = \frac{(\sum x_i)^2}{\theta^2} - 2\frac{\sum x_i\sum a_i}{\theta} + (\sum a_i)^2$$

Thus,

$$E_\theta\left[(\frac{\partial}{\partial\theta}log(f_{\mathbf{X}}(\mathbf{x}|\theta)))^2\right] = E_\theta\left[\frac{(\sum x_i)^2}{\theta^2} - 2\frac{\sum x_i\sum a_i}{\theta} + (\sum a_i)^2\right]$$
$$= \frac{E[(\sum x_i)^2]}{\theta^2} - 2\frac{\sum a_i}{\theta}E[\sum x_i] + (\sum a_i)^2 \quad (3)$$

We will use the property of Variance: $Var[X] = E[X^2] - E^2[X]$ In our case,

$$E[(\sum x_i)^2] = Var[\sum x_i] + E^2[\sum x_i] = \sum\theta a_i + \theta^2(\sum a_i)^2 \quad (4)$$

$$\overset{(3),(4)}{\Rightarrow} E_\theta\left[(\frac{\partial}{\partial\theta}log(f_{\mathbf{X}}(\mathbf{x}|\theta)))^2\right] = \frac{\theta\sum a_i + \theta^2(\sum a_i)^2}{\theta^2} - 2\frac{\sum a_i}{\theta}\theta\sum a_i + (\sum a_i)^2$$
$$= \frac{\sum a_i}{\theta}$$

Finally we get the Cramer-Rao lower bound:

$$Var_\theta[\delta(\mathbf{X}] \geq \frac{\theta}{\sum a_i}$$

and from (2) we confirm that our unbiased (1) estimator achieves it, so the estimator $\delta(\mathbf{X})$ is the UMVUE of θ.

3.4

The likelihood function of X is

$$l(\theta) = \frac{\prod_{i=1}^{n} a_i^{x_i}}{\sum_{i=1}^{n} x_i!} e^{-\theta \sum_{i=1}^{n} a_i} e^{\sum_{i=1}^{n} x_i log(\theta)}$$

Now in order to find the MLE of θ we must find the maximum of $l(\theta), \quad \theta \in \Theta$. Because logarithms are strictly increasing functions, maximizing the likelihood is equivalent to maximizing the log-likelihood.

$$log(l(\theta)) = log\left(\frac{\prod_{i=1}^{n} a_i^{x_i}}{\sum_{i=1}^{n} x_i!}\right) + log(\theta) \sum_{i=1}^{n} x_i - \theta \sum_{i=1}^{n} a_i$$

Now we calculate the first derivative and equate it to zero.

$$\begin{aligned} \frac{\partial}{\partial \theta} log(l(\theta)) =& 0 + \frac{\sum x_i}{\theta} - \sum a_i = 0 \\ \Leftrightarrow \quad & \hat{\theta} = \frac{\sum x_i}{\sum a_i}, \quad \forall \theta \in \Theta \end{aligned}$$

To check whether this is a maximum or a minimum we must calculate the second derivative as follows:

$$\frac{\partial^2}{\partial^2 \theta} log(l(\theta)) = -\frac{\sum x_i}{\theta^2} < 0, \quad \forall \theta \in \Theta$$

We conclude that $\hat{\theta} = \frac{\sum x_i}{\sum a_i}$ maximizes $log(l(\theta))$, so it is the MLE of $\theta, \quad \forall \theta \in \Theta$

Interestingly,if $\sum x_i = 0$ and because we know that $\sum a_i > 0, \quad \forall \alpha_i > 0$ then

$$\frac{\partial}{\partial \theta} log(l(\theta)) = -\sum a_i < 0, \quad \forall \theta \in \Theta$$

So $log(l(\theta))$ has a maximum for $\theta = 0$. But from our definition $0 \notin \Theta = (0, \infty)$, so the MLE cannot be defined there.

However though, this is a highly unlikely scenario as we can calculate the probability of $\sum x_i = 0 \Rightarrow X_i = 0$ where X_i are independent random variables, thus

$$P(\textit{MLE not defined}) = P(\sum x_i = 0) = \prod_i^n P(X_i = 0) = \prod_i^n \frac{e^{-\theta a_i}(\theta a_i)^0}{0!} = e^{-\theta \sum_i^n a_i}$$

Exercise 4.

4.1

First we calculate the joint probability function $f_{\mathbf{X}}(\mathbf{x}|\mu_1, \sigma^2)$ of the independent random sample $X_i \sim \mathcal{N}(\mu_1, \sigma^2)$

$$f(\mathbf{x}|\mu_1, \sigma^2) = f(x_1, x_2, \ldots, x_n|\mu_1, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu_1)^2}{2\sigma^2}}$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu_1)^2}{2\sigma^2}}$$

Similarly, the joint probability function $f_{\mathbf{Y}}(\mathbf{y}|\mu_2, \sigma^2)$ of the independent random sample $Y_i \sim \mathcal{N}(\mu_2, \sigma^2)$

$$f(\mathbf{y}|\mu_2, \sigma^2) = f(y_1, y_2, \ldots, y_n|\mu_2, \sigma^2) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\mu_2)^2}{2\sigma^2}}$$
$$= (2\pi\sigma^2)^{-\frac{m}{2}} e^{-\frac{\sum_{i=1}^{m}(y_i-\mu_2)^2}{2\sigma^2}}$$

Because the two samples are independent, the probability joint function can be calculated as $f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}|\mu_1, \mu_2, \sigma^2) = f_{\mathbf{X}}(\mathbf{x}|\mu_1, \sigma^2) f(\mathbf{y}|\mu_2, \sigma^2)$. Thus,

$$f(\mathbf{x},\mathbf{y}|\mu_1, \mu_2, \sigma^2) = (2\pi\sigma^2)^{-\frac{m+n}{2}} e^{-\frac{[\sum_{i=1}^{n}(x_i-\mu_1)^2+\sum_{i=1}^{m}(y_i-\mu_2)^2]}{2\sigma^2}}$$

So the likelihood function is equal to

$$l(\theta) = (2\pi\sigma^2)^{-\frac{m+n}{2}} e^{-\frac{[\sum_{i=1}^{n}(x_i-\mu_1)^2+\sum_{i=1}^{m}(y_i-\mu_2)^2]}{2\sigma^2}} \quad , \theta = (\mu_1, \mu_2, \sigma^2)$$

Because logarithms are strictly increasing functions, maximizing the likelihood is equivalent to maximizing the log-likelihood.

$$log(l(\theta)) = log\left((2\pi\sigma^2)^{-\frac{m+n}{2}} e^{-\frac{[\sum_{i=1}^{n}(x_i-\mu_1)^2+\sum_{i=1}^{m}(y_i-\mu_2)^2]}{2\sigma^2}}\right)$$
$$= -\frac{(m+n)}{2} log(2\pi) - \frac{(m+n)}{2} log(\sigma^2) - \frac{[\sum_{i=1}^{n}(x_i-\mu_1)^2 + \sum_{i=1}^{m}(y_i-\mu_2)^2]}{2\sigma^2}$$

Let $\theta^* = (\mu_1^*, \mu_2^*, \sigma^{2*}) \in \Theta$ a value of θ maximizing $log(l(\theta))$. Assuming that we have already found the values μ_2^*, σ^{2*} and we wish to compute μ_1^*.

This value will maximize the function

$$h_1(\mu_1) := log(l(\mu_1, \mu_2^*, \sigma^{2*}))$$
$$= -\frac{(m+n)}{2}log(2\pi) - \frac{(m+n)}{2}log(\sigma^{2*}) - \frac{[\sum_{i=1}^{n}(x_i-\mu_1)^2 + \sum_{i=1}^{m}(y_i-\mu_2^*)^2]}{2\sigma^{2*}}$$

Differentiating with respect to μ_1 and setting the derivative equal to zero, we obtain that

$$h_1'(\mu_1) = -(-\frac{2\sum_{i=1}^{n}(x_i-\mu_1)^2}{2\sigma^{2*}}) = \frac{\sum_{i=1}^{n}(x_i-\mu_1)^2}{\sigma^{2*}} = 0$$
$$\Rightarrow \hat{\mu_1} = \bar{X}$$

The second derivative is

$$h_1''(\mu_1) = -\frac{\sum_{i=1}^{n}(x_i-\mu_1)}{\sigma^{2*}} < 0 \qquad \forall \mu_1 \in \Re \times (0,\infty)$$

Hence, regardless of the values μ_2^*, σ^{2*}, the log-likelihood will be maximized for $\hat{\mu_1} = \bar{X}$. In the same way,

$$h_2'(\mu_2) = -(-\frac{2\sum_{i=1}^{m}(y_i-\mu_2)^2}{2\sigma^{2*}}) = \frac{\sum_{i=1}^{m}(y_i-\mu_2)^2}{\sigma^{2*}} = 0$$
$$\Rightarrow \hat{\mu_2} = \bar{Y}$$

$$h_2''(\mu_2) = -\frac{\sum_{i=1}^{m}(y_i-\mu_2)}{\sigma^{2*}} < 0 \qquad \forall \mu_2 \in \Re \times (0,\infty)$$

Given that we know that $\mu_1^* = \bar{X}, \mu_2^* = \bar{Y}$, σ^{2*} corresponds to the value that maximizes (with respect to σ^2) the function

$$h_3'(\sigma^2) = -\frac{(m+n)}{2\sigma^2} + \frac{[\sum_{i=1}^{n}(x_i-\bar{X})^2 + \sum_{i=1}^{m}(y_i-\bar{Y})^2]}{2\sigma^4} = 0$$
$$\Rightarrow \frac{\sum_{i=1}^{n}(x_i-\bar{X})^2 + \sum_{i=1}^{m}(y_i-\bar{Y})^2}{2\sigma^4} = \frac{(m+n)}{2\sigma^2}$$
$$\Rightarrow \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{X})^2 + \sum_{i=1}^{m}(y_i-\bar{Y})^2}{m+n}$$

The second derivative is

$$h_3''(\sigma^2) = \frac{m+n}{2\sigma^4} - \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2 + \sum_{i=1}^{m}(y_i - \bar{Y})^2}{\sigma^6}, \quad \sigma^2 \in (0, \infty)$$

and we can conclude that $h_3''(\sigma^2) < 0$ for $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2 + \sum_{i=1}^{m}(y_i - \bar{Y})^2}{m+n}$. Hence this value corresponds to a maximum.

So, we derived that the log-likelihood function is maximed at

$$\hat{\theta} = (\bar{X}, \bar{Y}, \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2 + \sum_{i=1}^{m}(y_i - \bar{Y})^2}{m+n})$$

Because $\hat{\theta} \in \Theta = \mathcal{R}^2 \times (0, \infty)$ for all $x \in \mathcal{R}^n, y \in \mathcal{R}^m$ we conclude that the MLE of $\theta = (\mu_1, \mu_2, \sigma^2)$ is

$$\hat{\theta}(\mathbf{X}) = (\bar{X}, \bar{Y}, \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2 + \sum_{i=1}^{m}(y_i - \bar{Y})^2}{m+n})$$

4.2

Since $\hat{\mu}_1$ is the sample mean it is unbiased estimator of the corresponding population mean, that is $E_\theta(\hat{\mu}_1) = \mu_1$, for all $\theta \in \Theta$. Similarly, $\hat{\mu}_2$ is an unbiased estimator of μ_2.

Let $S_X^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$ denote the sample variance for $\mathbf{X}$ and $S_Y^2 = \frac{\sum_{i=1}^{m}(Y_i - \bar{Y})^2}{m-1}$ denote the sample variance for $\mathbf{Y}$. We also know that $E_\theta[S_X^2] = E_\theta[S_Y^2] = \sigma^2$. We have that

$$\begin{aligned}
E_\theta(\hat{\sigma}^2) &= E_\theta\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{m}(Y_i - \bar{Y})^2}{m+n})\right] \\
&= \frac{E_\theta[\sum_{i=1}^{n}(X_i - \bar{X})^2] + E_\theta[\sum_{i=1}^{m}(Y_i - \bar{Y})^2]}{m+n} \\
&= \frac{(n-1)E_\theta[S_X^2] + (m-1)E_\theta[S_Y^2]}{m+n} = \frac{(n-1)\sigma^2 + (m-1)\sigma^2}{m+n} \\
&= \frac{m+n-2}{m+n}\sigma^2
\end{aligned}$$

Thus, $\hat{\sigma}^2$ is a biased estimator of σ^2

4.3

Since $X_1, X_2, \ldots, X_n$ are observations of a random sample of size n from $\mathcal{N}(\mu_1, \sigma^2)$ population, then the sample mean $\bar{X}$ is normally distributed with mean μ_1 and variance $\frac{\sigma}{n}$ Thus,

$$\hat{\mu_1} \sim \mathcal{N}(\mu_1, \frac{\sigma^2}{n})$$

Similarly,

$$\hat{\mu_2} \sim \mathcal{N}(\mu_2, \frac{\sigma^2}{m})$$

Now for the difference of sample means, we know that the two samples are independent and normally distributed. So, the set of differences between the sample means is normally distributed as well with mean

$$E[\bar{X} - \bar{Y}] = \mu_{\bar{X}-\bar{Y}} = \mu_1 - \mu_2$$

.

Furthermore, the variance of the difference between independent random variables is equal to the sum of the individual variances. We already know that $\sigma^2_{\bar{X}} = \sigma^2/n$ and $\sigma^2_{\bar{Y}} = \sigma^2/m$, so the variance of the difference of sample means is

$$Var[\bar{X} - \bar{Y}] = \sigma^2_{\bar{X}-\bar{Y}} = \sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}} = \frac{\sigma^2}{n} + \frac{\sigma^2}{m} = \sigma^2(\frac{1}{n} + \frac{1}{m})$$

Thus,

$$\hat{\mu_1} - \hat{\mu_1} \sim \mathcal{N}\left(\mu_1 - \mu_2, \sigma^2(\frac{1}{n} + \frac{1}{m})\right)$$

Moreover, we have that

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2}{m+n} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{m+n} + \frac{\sum_{i=1}^m (Y_i - \bar{Y})^2}{m+n} \\
&= \frac{\sigma^2}{m+n} \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} + \frac{\sigma^2}{m+n} \frac{\sum_{i=1}^m (Y_i - \bar{Y})^2}{\sigma^2} \\
&= \frac{\sigma^2}{m+n} \left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} + \frac{\sum_{i=1}^m (Y_i - \bar{Y})^2}{\sigma^2} \right)
\end{aligned}$$

Now, from the definition of Chi-Square distribution we know that,

$$\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} \sim \mathcal{X}^2_{n-1}$$

and

$$\frac{\sum_{i=1}^{m}(Y_i-\bar{Y})^2}{\sigma^2} \sim \mathcal{X}^2_{m-1}$$

We can now use a known property of Chi-square distributions:
if X, Y independent random variables with $X \sim \mathcal{X}^2_n$ and $Y \sim \mathcal{X}^2_m$ then $X+Y \sim \mathcal{X}^2_{m+n}$. So in our case:

$$\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} + \frac{\sum_{i=1}^{m}(Y_i-\bar{Y})^2}{\sigma^2} \sim \mathcal{X}^2_{m+n-2}$$

Though, the $\mathcal{X}^2_{m+n-2}$ distribution is a special case of the $\mathcal{G}(\frac{m+n-2}{2}, 2)$ distribution. so we can use the property that states for some $c > 0$ then $cX \sim \mathcal{G}(m/2, 2c)$. So, finally for $c = \frac{\sigma^2}{m+n}$

$$\hat{\sigma}^2 = \frac{\sigma^2}{m+n}\left(\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} + \frac{\sum_{i=1}^{m}(Y_i-\bar{Y})^2}{\sigma^2}\right) \sim \mathcal{G}\left(\frac{m+n-2}{2}, \frac{2\sigma^2}{n+m}\right)$$

4.4

Previously we proved that:

$$\bar{X}-\bar{Y} \sim \mathcal{N}\left(\mu_1-\mu_2, \sigma^2(\frac{1}{n}+\frac{1}{m})\right)$$

and

$$\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2 + \sum_{i=1}^{m}(Y_i-\bar{Y})^2}{\sigma^2} \sim \mathcal{X}^2_{m+n-2}$$

Thus we have that

$$\begin{aligned}\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} &= \frac{\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{1}{n}+\frac{1}{m}}}}{\sqrt{\frac{\sum(X_i-\bar{X})^2+\sum(Y_i-\bar{Y})^2}{m+n-2}}} \\ &= \frac{\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}}{\sqrt{\frac{\sum(X_i-\bar{X})^2+\sum(Y_i-\bar{Y})^2}{\sigma^2(m+n-2)}}} \sim \frac{\mathcal{N}(0,1)}{\sqrt{\mathcal{X}^2_{m+n-2}/(m+n-2)}}\end{aligned}$$

and that by definition is the Student's t distribution with $m+n-2$ degrees of freedom t_{m+n-2}.

4.5

Observe that the quantity $\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}}$

(a) depends on the unknown parameters only through $\mu_1 - \mu_2$

(b) its distribution

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{m+n-2}$$

does not depend on $\mu_1 - \mu_2, \quad \forall \mu_1, \mu_2 \in \mathcal{R}$

Thus, $\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}}$ can serve as a pivotal quantity for obtaining a confidence interval for $\mu_1 - \mu_2$. We have to define constants $c_1 < c_2$ such that

$$P_\theta\left(c_1 \leq \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \leq c_2\right) = 1-\alpha, \quad \forall\theta \in \Theta$$

$$\Leftrightarrow P_\theta\left(c_1 S_p\sqrt{\frac{1}{n}+\frac{1}{m}} \leq \bar{X}-\bar{Y}-(\mu_1-\mu_2) \leq c_2 S_p\sqrt{\frac{1}{n}+\frac{1}{m}}\right) = 1-\alpha$$

$$P_\theta\left(\bar{X}-\bar{Y}+c_1 S_p\sqrt{\frac{1}{n}+\frac{1}{m}} \leq \mu_1-\mu_2 \leq \bar{X}-\bar{Y}+c_2 S_p\sqrt{\frac{1}{n}+\frac{1}{m}}\right) = 1-\alpha$$

From the last expression we conclude that the general form of an $100(1-\alpha)\%$ confidence interval of the mean difference is

$$\left[(\bar{X}-\bar{Y}+c_1 S_p\sqrt{\frac{1}{n}+\frac{1}{m}}, \bar{X}-\bar{Y}+c_2 S_p\sqrt{\frac{1}{n}+\frac{1}{m}}\right]$$

In order to obtain an $100(1-\alpha)\%$ equally-tailed confidence interval we set $c_1 = t_{m+n-2;1-\alpha/2} = -t_{m+n-2;\alpha/2}$ and $c_2 = t_{m+n-2;\alpha/2}$. Substituting in the previous general form we get that the $100(1-\alpha)\%$ equally-tailed confidence interval of the difference of means $\mu_1 - \mu_2$ is

$$\left[(\bar{X}-\bar{Y}-t_{m+n-2;\alpha/2} S_p\sqrt{\frac{1}{n}+\frac{1}{m}}, \bar{X}-\bar{Y}+t_{m+n-2;\alpha/2} S_p\sqrt{\frac{1}{n}+\frac{1}{m}}\right]$$

4.6

We have already proved that $\hat{\sigma}^2 \sim \mathcal{G}\left(\frac{m+n-2}{2}, \frac{2\sigma^2}{n+m}\right)$ and from the properties of the Gamma distribution we have that

$$E_\theta(\hat{\sigma}^2) = \frac{m+n-2}{m+n}\sigma^2$$

and

$$Var_\theta(\hat{\sigma}^2) = \frac{2(m+n-2)}{(m+n)^2}\sigma^4$$

So the MSE is equal to:

$$MSE(\hat{\sigma}^2, \sigma^2) = (E_\theta(\hat{\sigma}^2) - \sigma^2)^2 + Var_\theta(\hat{\sigma}^2) = \frac{4\sigma^4}{(m+n)^2} + \frac{2(m+n-2)}{(m+n)^2}\sigma^4$$
$$= \frac{2\sigma^4}{m+n}$$

For S_p^2 we have that

$$S_p^2 = \frac{\sigma^2}{m+n-2}\frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2}{\sigma^2}$$

We know that

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2}{\sigma^2} \sim \mathcal{X}^2_{m+n-2} = \mathcal{G}(\frac{m+n-2}{2}, 2)$$

from before, so using the property of Gamma distributions

$$S_p^2 \sim \mathcal{G}(\frac{m+n-2}{2}, \frac{2\sigma^2}{m+n-2})$$

Similarly we get that

$$E_\theta(S_p^2) = \sigma^2, \quad Var_\theta(S_p^2) = \frac{2\sigma^4}{m+n-2}$$

and the MSE is equal to:

$$MSE(S_p^2, \sigma^2) = (E_\theta(S_p^2) - \sigma^2)^2 + Var_\theta(S_p^2) = \frac{2\sigma^4}{m+n-2}$$

Now in order to compare the two MSE we take the following factor

$$\begin{aligned}\frac{MSE(S_p^2, \sigma^2)}{MSE(\hat{\sigma}^2, \sigma^2)} &= \frac{\frac{2\sigma^4}{m+n-2}}{\frac{2\sigma^4}{m+n}} \\ &= \frac{m+n}{m+n-2} > 1, \quad \forall m \geq 2, n \geq 2\end{aligned}$$

Thus, S_p^2 has a greater Mean Squared Error than $\hat{\sigma}^2$, so we choose $\hat{\sigma}^2$ as the best estimator of σ^2 under the MSE.

Exercise 5.

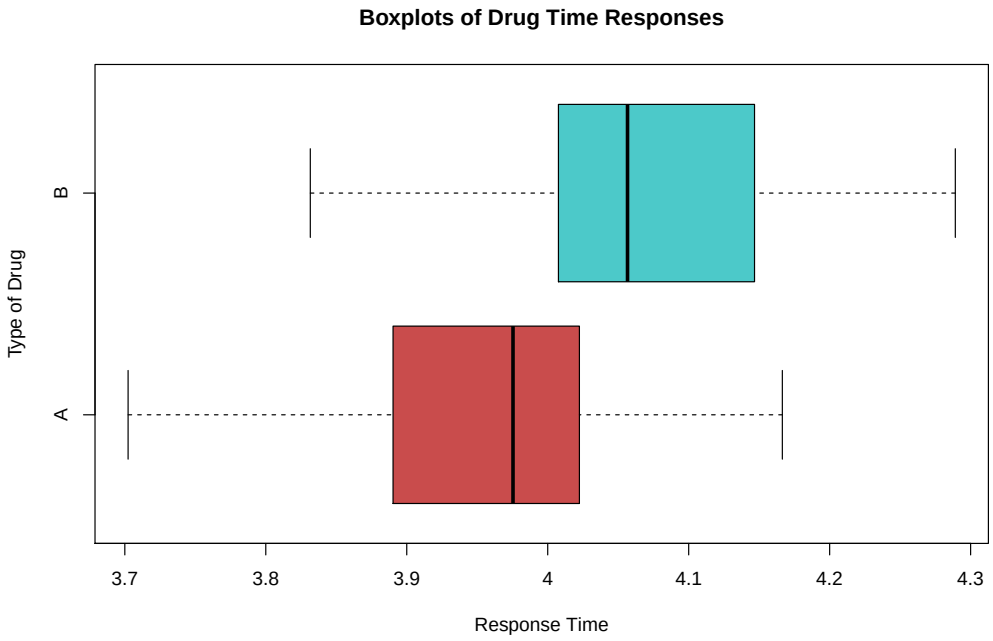
5.1

Using R we can view calculate summary statistics for our data:

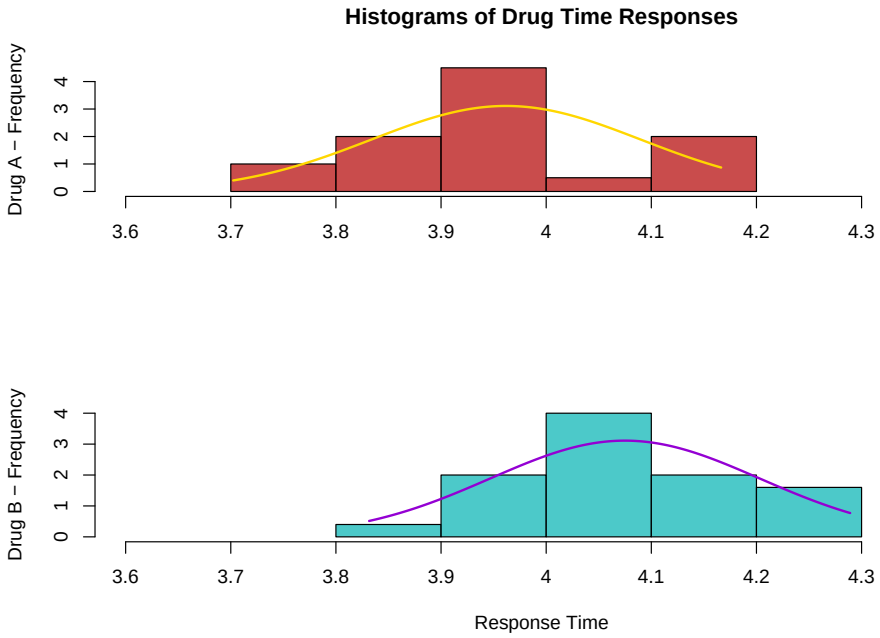
```
> tapply(drug_response$time, drug_response$drug, summary)
$A
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.702   3.895   3.975   3.962   4.009   4.166

$B
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.832   4.008   4.057   4.075   4.147   4.289
```

Also we can produce a figure with two boxplots, one for each drug:



and a figure with two histograms as well:



5.2

We can see clearly that we have a bigger response time for Drug B than Drug A. That can be deduced from all the summary statistics like median and mean. This difference is clearly visualized in both figures. The IQR range is the same for both drugs. Also, we assume that the sample means follow a normal distribution with means μ_1 and μ_2 respectively and a common standard deviation σ.

5.3

We calculate the MLE estimates of mean response times within each group as follows

```
> cat('MLE-drug A :',sum(drug_A) / length(drug_A),'\n'
+     ,'MLE-drug B :',sum(drug_B) / length(drug_B),'\n'
+     )
MLE-drug A : 3.961835
  MLE-drug B : 4.074892
```

5.4

We calculate the MLE estimate of (common) variance of response time $\hat{\sigma}^2$:

```
> cat('MLE-variance biased:',(sum((drug_A - mean(drug_A)) ^ 2)
+  + sum((drug_B - mean(drug_B)) ^ 2))
/(length(drug_A)+length(drug_B)))
MLE-variance biased: 0.01291238
```

and the MLE estimate of the unbiased S_p^2:

```
> cat('MLE-variance unbiased:',(sum((drug_A - mean(drug_A)) ^ 2)
+  + sum((drug_B - mean(drug_B)) ^ 2))
/ (length(drug_A) + length(drug_B) - 2))
MLE-variance unbiased: 0.01351295
```

5.5

Now we calculate the equally-tailed 95%, 99% and 99.9% confidence intervals for the difference of mean response time between groups:

```
> #Confidence Intervals
> Sp <- (sum((drug_A - mean(drug_A)) ^ 2)
+ sum((drug_B - mean(drug_B)) ^ 2)) / (length(drug_A)
+ length(drug_B) - 2)
> alpha_seq <- c(0.05, 0.01, 0.001)
> for (alpha in alpha_seq) {
drug_t <- qt(p = alpha / 2, df = length(drug_A) + length(drug_B) - 2
, lower.tail = F)
CI_lower <- mean(drug_A) - mean(drug_B)
- drug_t * sqrt(Sp) * sqrt((1 / length(drug_A))
+ (1 / length(drug_B)))
CI_upper <- mean(drug_A) - mean(drug_B)
+ drug_t * sqrt(Sp) * sqrt((1 / length(drug_A))
+ (1 / length(drug_B)))
cat('Confidence Interval for mean(A) - mean(B): ', 100 * (1 - alpha)
, "\%  [", CI_lower, ' , ', CI_upper, ']','\n')
+ }
CI for mean(A) - mean(B):  95 %   [ -0.1833867  ,  -0.04272819 ]
```

```
CI for mean(A) - mean(B):  99 %   [ -0.2070453  ,  -0.01906962 ]
CI for mean(A) - mean(B):  99.9 %   [ -0.2362178  ,  0.01010295 ]
```

For the first two confidence intervals 0 is not included, so there is a difference in mean response times, drug B has a greater response time.
But for the last confidence interval, 0 is included, so we can conclude that with 99.9% condifence that the two drugs do not differ in means reponse times.

5.6

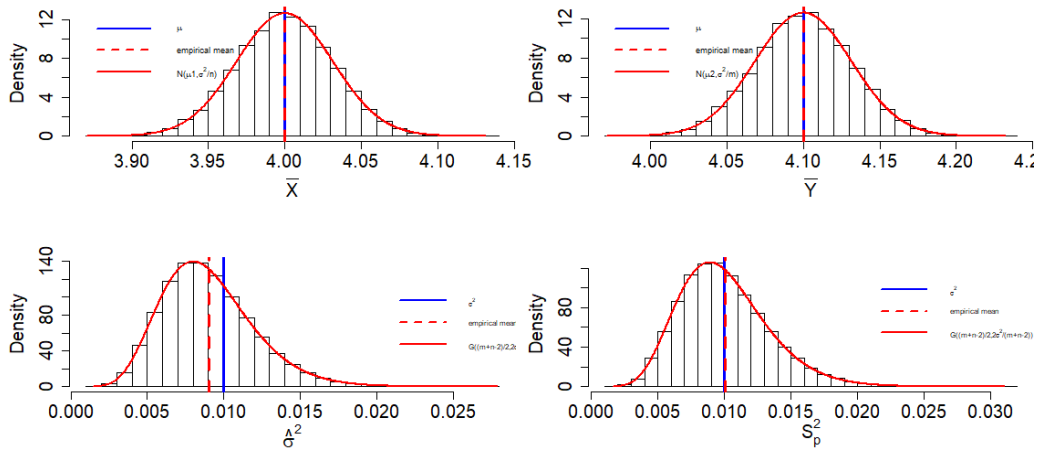
We can validate the previously computed confidence intervals using R's built-in function *t.test*:

```
> # Confirmation with t.test
> for (alpha in alpha_seq) {
+ cat('CI with t.test: ',100*(1-alpha),'%\t'
+     ,'[',as.numeric(t.test(drug_A, drug_B
, var.equal = TRUE, conf.level = 1-alpha)$conf.int),']','\n')
+ }
CI with t.test:  95 %  [ -0.1833867 -0.04272819 ]
CI with t.test:  99 %  [ -0.2070453 -0.01906962 ]
CI with t.test:  99.9 %  [ -0.2362178 0.01010295 ]
```

Exercise 6.

6.1

First of all, we plot histograms of the relative frequencies of the sampled values: In each histogram we superimpose the true value of the relevant parameter, the empirical mean of the sampled values and the distribution of the sample estimate.



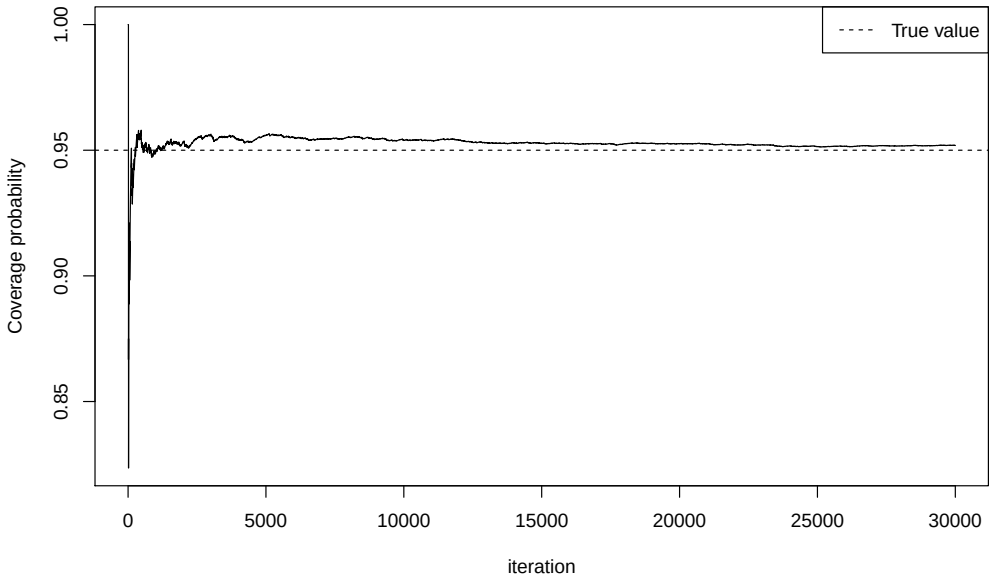
We can see clearly that our biased sample estimator estimate of the variance $\hat{\sigma}^2$ differs from the true value of the parameter despite the fact that it had a smaller Mean Squared Error from the unbiased estimator S_p^2. The unbiased estimator of the sample variance estimates the true parameter exactly.

6.2

The probability that the random interval contains the true value of the difference of means is

```
> sum(ci_probability)/nIter
[1] 0.9528453
```

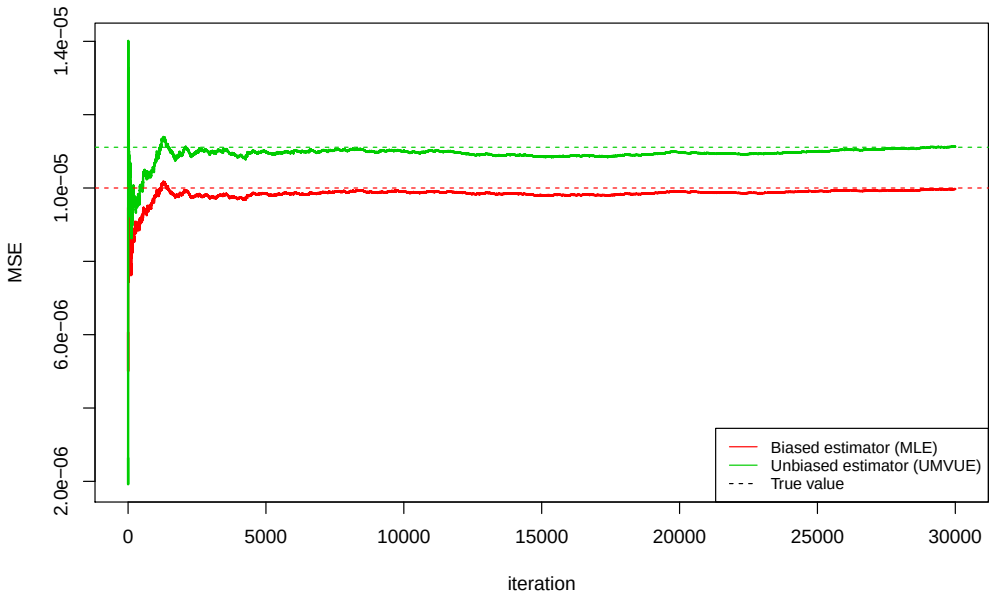
,which is expected for $\alpha = 0.05$.
Now we can produce a plot showing the estimated coverage probability versus the number of interations and sumperimpose the true coverage probability of the confidence interval.



We can see clearly the the coverage probability converges to the true coverage probability (95%)

6.3

Finally, we estimate the value of Mean Square Error of $\hat{\sigma}^2$ and S_p^2 for each iteration and plot it in a figure where we sumperimpose the true mean square errors that we calculated in 4.6:



Exercise 7

Let X_i be a voter of a given region with a population $n = 200$. The experiment has a binary random variable with two possible outcomes $X_i \in 0, 1($ the voter chooses to vote for the specific party, or to not vote for it). Now, we assume that the voters vote independently so this follows a Binomial distribution $X_i \sim \mathcal{B}(n, p)$ where n is the number of trials and p is the probability of success with $f(x|n, p) = P(X = x|n, p) = \binom{n}{p} p^x (1-p)^{(n-x)}$. We are trying to estimate the true population proportion p that voted for a specific party using the sample proportion $\hat{p}$ that we have.

If X is the number of successes in n trials then ,

$$E[X] = n\hat{p} \Rightarrow \hat{p} = \frac{X}{n} = \frac{80}{200} = \frac{2}{5} = 0.4$$

and from the definition of the Bernoulli distribution we can also calculate the variance of our random sample as

$$\hat{\sigma}^2 = Var[X] = n\hat{p}(1-\hat{p}) = \frac{2}{5} \times \frac{3}{5} = \frac{6}{25} = 0.24$$

Assuming that p is not close to 0 or 1 and our sample size is large enough, we can use the CLT, so the following follows a normal distribution:

$$\frac{\hat{p} - p}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\hat{p} - p}{\frac{\sigma}{\sqrt{n}}} = Z \sim \mathcal{N}(0, 1)$$

Furthermore, it is known that:

$$W = \frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \mathcal{X}^2_{n-1}$$

with W being independent of Z. Then we will have:

$$T = \frac{Z}{\sqrt{\frac{W}{n-1}}} = \frac{\frac{\hat{p}-p}{\sqrt{\frac{\sigma^2}{n}}}}{\sqrt{\frac{\frac{(n-1)\hat{\sigma}^2}{\sigma^2}}{n-1}}} = \frac{\hat{p} - p}{\frac{\hat{\sigma}}{\sqrt{n}}} \sim t_{n-1}$$

Then we bound the statistic T between the quantiles $\pm t_{n-1;a/2}$

$$-t_{n-1;a/2} \leq \frac{\hat{p}-p}{\frac{\hat{\sigma}}{\sqrt{n}}} \leq t_{n-1;a/2} \Rightarrow$$
$$\Rightarrow -t_{n-1;a/2}\frac{\hat{\sigma}}{\sqrt{n}} \leq \hat{p}-p \leq t_{n-1;a/2}\frac{\hat{\sigma}}{\sqrt{n}} \Rightarrow$$
$$\Rightarrow \hat{p}-t_{n-1;a/2}\frac{\hat{\sigma}}{\sqrt{n}} \leq p \leq \hat{p}+t_{n-1;a/2}\frac{\hat{\sigma}}{\sqrt{n}}$$

We want to calculate a 99% asymptotic confidence interval, thus $100(1-\alpha) = 99 \Rightarrow \alpha = 0.01$

The (upper) 0.005 percentile of the student's t distribution with 199 degrees of freedom is $t_{199;0.005} = 2.60076$. Thus the realization of the 99% confidence interval for the true population proportion will be given by:

$$\hat{p} \pm t_{n-1;a/2}\frac{\hat{\sigma}}{\sqrt{n}} = [0.309907, 0.490093]$$