MTH6312 - Méthodes statistiques d'apprentissage

Devoir no 1

Gervais Presley Koyaweda 2305686

September 27, 2024

Question N°1

$$\mathcal{D} = \{x_i, y_i\} = i = 1, \dots, n \text{ et } p = 1$$

1-a) Détermination de l'estimateur de maximum de vraisemblance

$$f(y_i|\theta, x_i) = \begin{cases} 2(x_i\theta)^2 y_i^3 \exp\{-\theta x_i y_i^2\} & \text{si } y_i \geq 0, \ x_i \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

$$L(\theta) = \prod_{i=1}^{n} f(y_i|\theta, x_i)$$

$$= \prod_{i=1}^{n} 2(x_i\theta)^2 y_i^3 \exp\{-\theta x_i y_i^2\}$$

$$\mathcal{L}(\theta) = \ln \prod_{i=1}^{n} 2(x_i\theta)^2 y_i^3 e^{-\theta x_i y_i^2}$$

$$= \sum_{i=1}^{n} \left[\ln(2) + 2\ln(x_i\theta) + 3\ln y_i - \theta x_i y_i^2\right.$$

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = \sum_{i=1}^{n} \left[\frac{2}{\theta} - x_i y_i^2\right]$$

$$= \frac{2n}{\theta} - \sum_{i=1}^{n} x_i y_i^2$$

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = 0 \Rightarrow \boxed{\hat{\theta}_{emv} = \frac{2n}{\sum_{i=1}^{n} x_i y_i^2}}$$

$$\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta^2} = -\frac{2n}{\theta^2} < 0$$; nous somme donc à un maximum.

• Information de Fisher

$$I(\theta) = -E\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta^2}\right]$$

$$= -E\left[-\frac{2n}{\theta^2}\right] = \frac{2n}{\theta^2}$$

Donc $\boxed{I(\theta) = \frac{2n}{\theta^2}}$

La variance asymptotique de θ est

$$V(\hat{\theta}_{evm}) = \frac{1}{I(\theta)} = \frac{1}{\frac{2n}{\theta^2}} = \frac{\theta^2}{2n}$$

Alors lorsque n est grand on a:

$$\boxed{\hat{\theta}_{evm} \rightsquigarrow \mathcal{N}\left(\theta x_i, \frac{\theta^2}{2n}\right)}$$

1-b) Détermination de l'estimateur de maximum de vraisemblance θ_{evm} de θ.

$$D = \{x_i ; y_i\} \sim P(x_i, \theta)$$

$$P(y_i | \theta, x_i) = \begin{cases} \binom{y_i - 1}{x_i - 1} \theta^{x_i} (1-\theta)^{y_i - x_i} & \text{si } y_i = x_i, x_{i+1}, \dots \\ 0 & \text{sinon.} \end{cases}$$

$$L(\theta) = \prod_{i=1}^{n} \binom{y_i - 1}{x_i - 1} \theta^{x_i} (1-\theta)^{y_i - x_i}$$

$$\mathcal{L}(\theta) = \ln \prod_{i=1}^{n} \binom{y_i - 1}{x_i - 1} \theta^{x_i} (1-\theta)^{y_i - x_i}$$

$$= \sum_{i=1}^{n} \left[\ln \binom{y_i - 1}{x_i - 1} + x_i \ln \theta + (y_i - x_i) \ln (1-\theta) \right]$$

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = \sum_{i=1}^{n} \left[\frac{x_i}{\theta} - \frac{(y_i - x_i)}{(1-\theta)} \right]$$

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = 0 \Rightarrow \sum_{i=1}^{n} \frac{x_i}{\theta} - \sum_{i=1}^{n} \frac{(y_i - x_i)}{(1-\theta)} = 0$$

Après transformation et résolution d'equation $\Rightarrow$ $\boxed{\hat{\theta}_{evm} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i}}$

$$\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta^2} = \sum_{i=1}^{n} \left[-\frac{x_i}{\theta^2} - \frac{(y_i - x_i)}{(1-\theta)^2} \right] < 0$$

Nous somme donc à un maximum

- Information de Fisher

$$I(\theta) = -E\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta^2}\right]$$

$$= -E\left[\sum_{i=1}^{n}\left[-\frac{x_i}{\theta^2} - \frac{(y_i - x_i)}{(1-\theta)^2}\right]\right]$$

$$= \sum_{i=1}^{n}\left[E\left[\frac{x_i}{\theta^2}\right] + E\left[\frac{y_i - x_i}{(1-\theta)^2}\right]\right]$$

$$E(Y_i) = \frac{x_i}{\theta}$$

Alors $I(\theta) = \sum_{i=1}^{n}\left[\frac{x_i}{\theta^2} + \frac{1}{(1-\theta)^2} \times \left(\frac{x_i}{\theta} - x_i\right)\right]$

$$\boxed{I(\theta) = \sum_{i=1}^{n} \frac{x_i}{\theta^2(1-\theta)}} \quad V(\hat{\theta}_{evm}) = \frac{1}{n\,I(\theta)}$$

lors n devient grand,

$$\boxed{\hat{\theta}_{evm} \leadsto N\left(\theta, \frac{1}{n\sum_{i=1}^{n}\frac{x_i}{\theta^2(1-\theta)}}\right)}$$

1-c) On donne :

$$\pi(\theta) \begin{cases} 2(1-\theta) & \text{si } 0<\theta<1 \\ 0 & \text{sinon} \end{cases}$$

Déterminons l'estimateur MAP de θ

$$L(\theta|y_i) = \prod_{i=1}^{n} \binom{y_i-1}{x_i-1} \theta^{x} (1-\theta)^{y_i-x_i}$$

* Recherchons π à postériori

$$\pi(\theta|y_i) = \frac{P(y_i|\theta, x_i)\,\pi(\theta)}{\int_{-\infty}^{\infty} P(y_i|\theta, x_i)\,d\theta}$$

or $\int_{-\infty}^{\infty} P(y_i|\theta, x_i)\,d\theta = \text{Conste} = c$

Posons $\frac{1}{c} = K = \text{constante}$

Alors $\pi(\theta|y_i) = K \times P(y_i|\theta, x_i)\,\pi(\theta)$

$$= K \times \binom{y_i-1}{x_i-1} \theta^{x_i} (1-\theta)^{y_i-x_i} \times 2(1-\theta)$$

Or $\hat{\theta}_{MAP} = \text{Arg}\max \{\pi(\theta|y_i)\}$

$$= \operatorname{Arg\,max}\left\{ \ln\left[2k\binom{y_i-1}{x_i-1}\theta^{x_i}(1-\theta)^{y_i-x_i+1}\right]\right.$$

$$= \sum_{i=1}^{n} \operatorname{Arg\,max}\left\{C^{ste} + x_i\ln\theta + (y_i - x_i + 1)\ln(1-\theta)\right\}$$

$$\frac{\partial \mathcal{L}(\theta)}{\partial\theta = 0} \Rightarrow \frac{1}{\theta}\sum_{i=1}^{n} x_i - \frac{1}{1-\theta}\sum_{i=1}^{n}(y_i - x_i + 1) = 0$$

$$\Longleftrightarrow \frac{1}{\theta}\sum_{i=1}^{n} x_i = \frac{1}{1-\theta}\sum_{i=1}^{n}(y_i - x_i + 1)$$

$$(1-\theta)\sum_{i=1}^{n} x_i = \theta\sum_{i=1}^{n}(y_i - x_i + 1)$$

$$\sum_{i=1}^{n} x_i = \theta\left(\sum_{i=1}^{n} y_i + n\right)$$

D'où $$\boxed{\hat{\theta}_{MAP} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i + n}}$$

C-2) Estimateur de Bayes

Considérons la fonction de perte quadratique.

$$\hat{\theta}_\pi = \arg\min E_\pi\left[(\theta - \hat{\theta})^2 \mid Y_1 \ldots Y_n\right]$$

$$= E_\pi(\theta \mid Y_1 \ldots Y_n)$$

$$= \int_{-\infty}^{+\infty} \theta \, \pi(\theta \mid y_i) \, d\theta$$

On sait aussi que :

$$L(\theta) = \prod_{i=1}^{n} \binom{y_i - 1}{x_i - 1} \theta^{x_i} (1-\theta)^{y_i - x_i}$$ qui

est la fonction de vraissemblance

Sa fonction à posteriori est

$$\pi(\theta) = \text{Conste} \times \theta^{\sum x_i} \times (1-\theta)^{\sum_{i=1}^{n}(y_i - x_i) + 1}$$

poson $\alpha - 1 = \sum_{i=1}^{n} x_i$ et $\beta - 1 = \sum_{i=1}^{n}(y_i - x_i) + 1$

$$\Rightarrow \alpha = \sum_{i=1}^{n} x_i + 1 \text{ et } \beta = \sum_{i=1}^{n}(y_i - x_i) + 2$$

On observe donc que la distribution à priori est de type Beta

$$\beta = \left(\alpha = \sum_{i=1}^{n} x_i + 1 \text{ et } \beta = \sum_{i=1}^{n}(y_i - x_i) + 2\right)$$

Pour une loi Beta(α, β), l'esperance est :

$$E(\pi(\theta \mid y_i)) = \frac{\alpha}{\alpha + \beta}$$

$$\Rightarrow E\left[\pi(\theta \mid y_i)\right] = \frac{\sum_{i=1}^{n} x_i + 1}{\left(\sum_{i=1}^{n} x_i + 1\right) + \left(\sum_{i=1}^{n} (y_i - x_i) + 2\right)}$$

$$= \frac{\sum_{i=1}^{n} x_i + 1}{\sum_{i=1}^{n} x_i + 1 + \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i + 2}$$

$$E\left[\pi(\theta \mid y_i)\right] = \frac{\sum_{i=1}^{n} x_i + 1}{\sum_{i=1}^{n} y_i + 3}$$

Alors $$\boxed{\hat{\theta}_{Bayes} = \frac{\sum_{i=1}^{n} x_i + 1}{\sum_{i=1}^{n} y_i + 3}}$$

Comparaison entre $\hat{\theta}_{MAP}$ et $\hat{\theta}_{Bayes}$

$$\hat{\theta}_{map} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i + 1}$$

$$\hat{\theta}_{Bayes} = \frac{\sum_{i=1}^{n} x_i + 1}{\sum_{i=1}^{n} y_i + 3}$$

- $\hat{\theta}_{map}$ maximise la densité à postériori, tandis que $\hat{\theta}_{Bayes}$ est l'espérance de la distribution à postériori

- Par ailleur l'estimateur de Bayes est légèrement plus régularisé, car l'ajout de (+1) au numérateur et (+2) au dénominateur, ce qui correspond à une forme de régularisation introduite par la distribution à priori.

En fin θ_{map} est plus concentré sur la valeur qui maximise la probabilité, alors que Bayes prend en compte la forme complete de la distribution.

Les deux estimateurs tendent à donner des valeurs proches lorsque le nombre d'observation "n" est grand, car l'influence de la distribution à priori diminue par rapport à la vraissemblance basée sur les données.

Question 2

$\{(x_i, y_i); i=1,\dots,n\}$ $\quad p=1 \quad \sum_{i=1}^{n} u_i = 1$

2-a) Détermination de l'estimateur des moindres carrés $\hat{\beta}$ definit par

$$\hat{\beta} = \arg\min RSS_1(\beta) \text{ où } RSS_1(\beta) = \sum_{i=1}^{n} (y_i - \beta x_i)^2$$

$$\frac{\partial}{\partial \beta} RSS_1(\beta) = \frac{\partial}{\partial \beta}\left[\sum_{i=1}^{n} (y_i - \beta x_i)^2\right]$$

$$= 2\left[\sum_{i=1}^{n} \left[(-x_i)(y_i - \beta x_i)\right]\right]$$

$$= 2\sum_{i=1}^{n} (x_i y_i - \beta x_i^2)$$

$$= 2\sum_{i=1}^{n} x_i y_i + 2\beta \sum_{i=1}^{n} x_i^2$$

$$\frac{\partial}{\partial \beta} RSS_1(\beta) = 0 \Rightarrow \beta \sum_{i=1}^{n} x^2 = \sum_{i=1}^{n} x_i y_i$$

$$\text{D'où } \boxed{\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}}$$

2-b) Détermination de $\hat{\beta}_0$ et $\hat{\beta}_1$

$$(\hat{\beta}_1, \hat{\beta}_2) = \arg\min RSS_2(\beta_0, \beta_1)$$

$$RSS_2(\beta_0, \beta_1) = \sum_{i=1}^{n} w_i (y_i - \beta_0 - \beta_1 x_i)^2$$

$$\frac{\partial RSS_2}{\partial \beta_0} = -2 \sum_{i=1}^{n} w_i (y_i - \beta_0 - \beta_1 x_i) = 0 \quad (1)$$

$$\frac{\partial RSS_2}{\partial \beta_1} = -2 \sum_{i=1}^{n} w_i x_i (y_i - \beta_0 - \beta_1 x_i) = 0 \quad (2)$$

De (1), on a :

$$\sum_{i=1}^{n} w_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} w_i y_i - \beta_0 \sum_{i=1}^{n} w_i - \beta_1 \sum_{i=1}^{n} w_i x_i = 0 \quad (1)$$

De (2), on a :

$$\sum_{i=1}^{n} w_i x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\sum_{i=1}^{n} w_i x_i y_i - \beta_0 \sum_{i=1}^{n} w_i x_i - \beta_1 \sum_{i=1}^{n} w_i x_i^2 = 0$$

$\sum_{i=1}^{n} w_i = 1$ alors posons :

$$\overline{X_w} = \sum_{i=1}^{n} w_i x_i \quad \text{et} \quad \overline{Y_w} = \sum_{i=1}^{n} w_i y_i$$

Donc $\overline{Y}_{\omega} - \beta_0 - \beta_1 \overline{X}_{\omega} = 0$

De ②, on a aussi

$$\sum_{i=1}^{n} \omega_i x_i y_i - \beta_0 \overline{X}_{\omega} - \beta_1 \sum_{i=1}^{n} \omega_i X_i^2 = 0$$

si $\beta_0 = \overline{Y}_{\omega} - \beta_1 \overline{X}_{\omega}$ alors

$$\sum_{i=1}^{n} \omega_i x_i y_i - (\overline{Y}_{\omega} - \beta_1 \overline{X}_{\omega}) \overline{X}_{\omega} + \beta_1 \sum_{i=1}^{n} \omega_i X_i^2$$

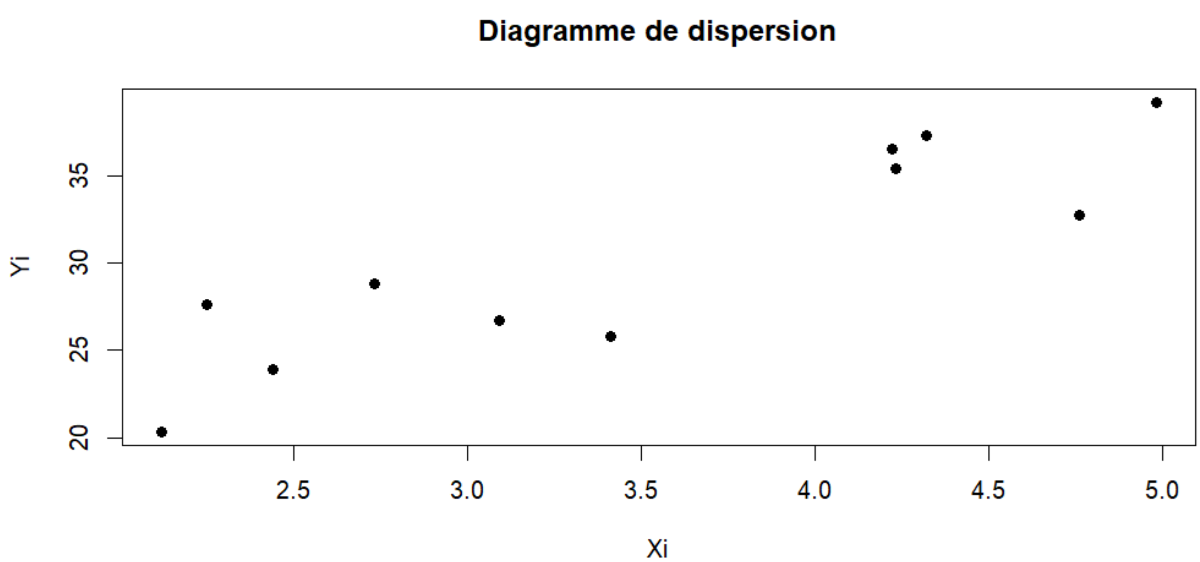
$$\sum_{i=1}^{n} \omega_i x_i y_i - \overline{Y}_{\omega} \overline{X}_{\omega} = \beta_1 \left(\sum_{i=1} \omega_i x_i^2 - \overline{X}_{\omega}^2 \right)$$

$$\boxed{\beta_1 = \frac{\sum_{i=1}^{n} \omega_i x_i y_i - \overline{Y}_{\omega} \overline{X}_{\omega}}{\sum_{i=1}^{n} \omega_i X_i^2 - \overline{X}_{\omega}^2}}$$

$$\boxed{\beta_0 = \overline{Y} - \beta_1 \overline{X}_{\omega}}$$

2-c) Diagramme de dispersion

```
R 4.4.1 · ~/
> Xi<-c(2.12,2.25,4.23,4.98,4.76,4.22,3.09,2.73,2.44,4.32,3.41)
> Yi<-c(20.3,27.6,35.4,39.2,32.7,36.5,26.7,28.8,23.9,37.3,25.8)
> Wi<-c(1/24,1/12,1/48,7/48,1/24,1/12,1/6,5/48,1/12,7/48,1/12)
>
>
> #COntruction du diagramme de dispersion
>
> plot(Xi,Yi, main = "Diagramme de dispersion",xlab ="Xi",ylab = "Yi", pch=19)
>
```



Les valeurs des estimateurs Beta, Beta0 et Beta1

Beta = 8.41

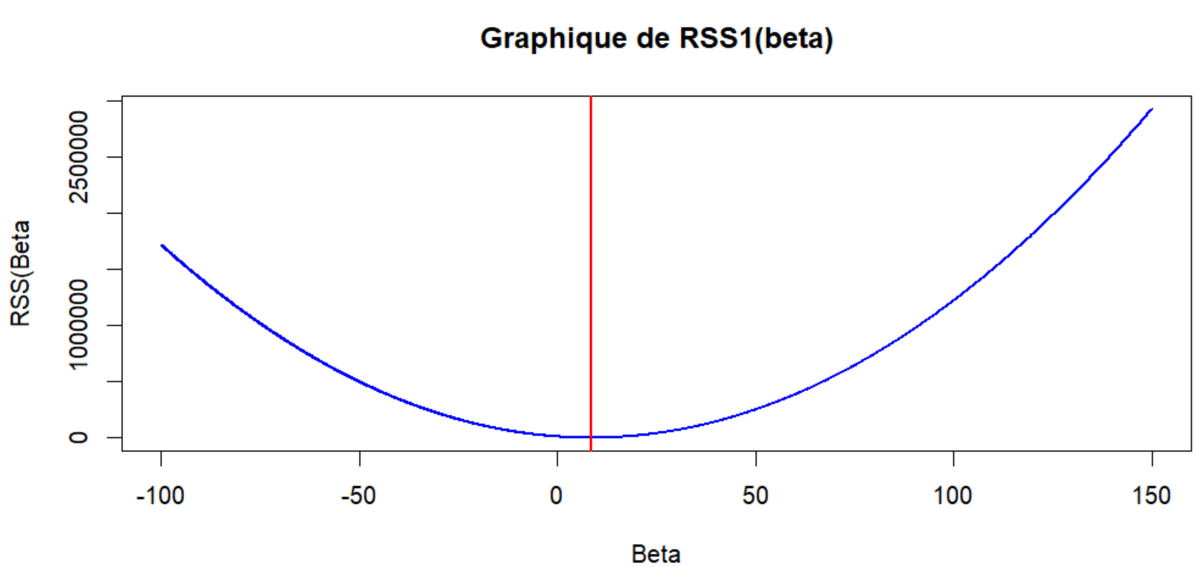
Beta_1 = 5.50

Beta_0 = 11.44

Les détails du calcul à la page suivante

```
> #COntruction du diagramme de dispersion
>
> plot(Xi,Yi, main = "Diagramme de dispersion",xlab ="Xi",ylab = "Yi", pch=19)
>
>
> #Valeur de numériques des estimateurs obtenus precedemment
>
> ## Calcul de la somme des produits pondérés Xi*Yi et Xi^2
> sum_w_Xi_Yi <- sum(Wi * Xi * Yi)
> sum_w_Xi <- sum(Wi * Xi)
> sum_w_Yi <- sum(Wi * Yi)
> sum_w_Xi2 <- sum(Wi * Xi^2)
> sum_w <- sum(Wi)
>
> # Calcul de Beta_1
> Beta_1 <- (sum_w_Xi_Yi - (sum_w_Xi * sum_w_Yi) / sum_w) / (sum_w_Xi2 - (sum_w_Xi^
2) / sum_w)
>
> # Calcul de Beta_0
> Beta_0 <- (sum_w_Yi - Beta_1 * sum_w_Xi) / sum_w
>
> # Affichage des résultats
> print(paste("Beta_1 =", Beta_1))
[1] "Beta_1 = 5.50799160304272"
> print(paste("Beta_0 =", Beta_0))
[1] "Beta_0 = 11.4430315353435"
>
> Calcul de Beta
Erreur : symbole inattendu dans "Calcul de"
> #Calcul de Beta
> Beta<-sum(Xi*Yi)/sum(Xi*Xi)
> print(paste("Beta = ",Beta))
[1] "Beta =  8.41547040208603"
>
```

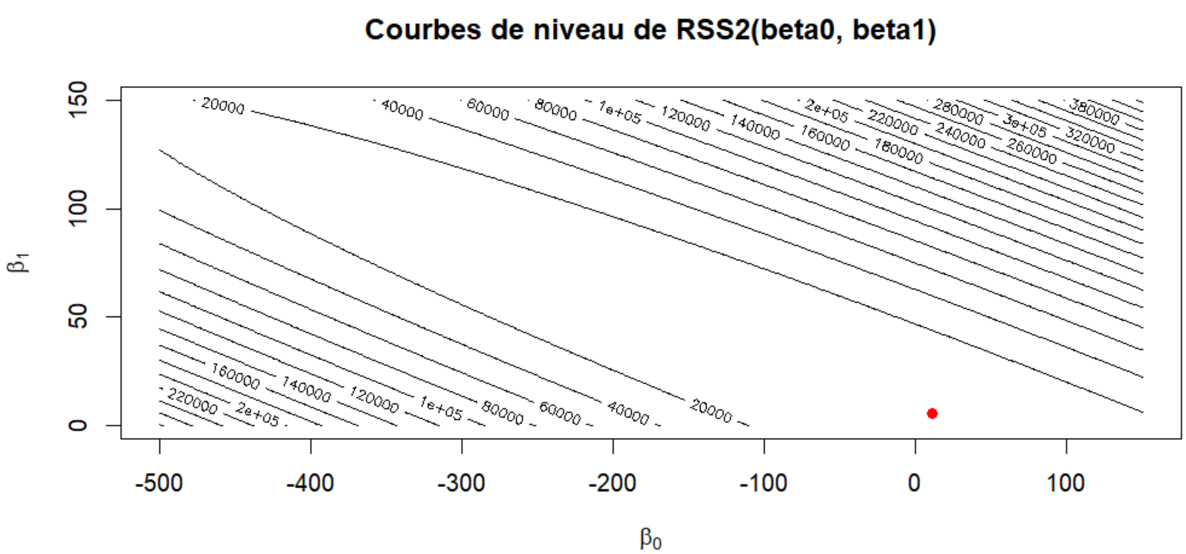
2-d) Graphique de RSS1 en fonction de Beta



```
> #Graphe de RSS1 en fonction de Beta
>
> # Fonction RSS1
> RSS1 <- function(beta) {
+     sum((Yi - beta * Xi) ^ 2)
+ }
> # Calcul des valeurs de RSS1 pour différents beta
> beta_values <- seq(-100, 150, by = 0.1)
> rss_values <- sapply(beta_values, RSS1)
>
> #Tracer du graphe
> plot(beta_values,rss_values, type = 'l', col="blue", lwd=2,
+      xlab = "Beta", ylab = "RSS(Beta",main = "Graphique de RSS1(beta)")
> abline(v = 8.41, col = 'red', lwd = 2)
> 
```

La valeur de Beta=8.41 calculée algébriquement, correspond presque exactement au minimum de la courbe observée graphiquement, qui est autour de Beta = 8. Cela montre que le calcul algébrique et le graphique sont en accord, avec une différence négligeable.

2-e) Graphique de RSS2 en fonction de Beta0 et Beta1



Les courbes de niveau montrent que le point estimé (Beta0 et Beta1) est bien situé près du minimum global de la fonction RSS2, confirmant que les valeurs calculées sont proches de la solution optimale. Les lignes de contour resserrées autour de ce point indiquent que nous sommes dans une région de faible erreur résiduelle.

```
> # Fonction RSS2
> RSS2 <- function(params) {
+     beta0 <- params[1]
+     beta1 <- params[2]
+     sum(Wi * (Yi - beta0 - beta1 * Xi) ^ 2)
+ }
> # Grille de valeurs pour beta0 et beta1
> beta0_values <- seq(-500, 150, by = 1)
> beta1_values <- seq(0, 150, by = 1)
>
> # Calcul de RSS2 pour chaque combinaison de beta0 et beta1
> rss1_grid <- outer(beta0_values, beta1_values, Vectorize(function(b0, b1) RSS2(c
(b0, b1))))
>
> # Tracer des courbes de niveau
> # Tracer les courbes de niveau
> contour(beta0_values, beta1_values, rss_grid, nlevels = 30,
+         xlab = expression(beta[0]), ylab = expression(beta[1]),
+         main = "Courbes de niveau de RSS2(beta0, beta1)")
Erreur : objet 'rss_grid' introuvable
> # Tracer les courbes de niveau
> contour(beta0_values, beta1_values, rss1_grid, nlevels = 30,
+         xlab = expression(beta[0]), ylab = expression(beta[1]),
+         main = "Courbes de niveau de RSS2(beta0, beta1)")
> points(11.44, 5.51, col = 'red', pch = 19)  # Point optimal algébrique trouvé
>
```

2-f) Optimisation des valeurs de valeurs des estimateurs pour le RSS, RSS1 et RSS2

```
>
> # Optimisation numérique pour RSS1
> result_RSS1 <- optim(par = 0, fn = RSS1, method = "Brent", lower = -100, upper =
150)
> beta_optimal_RSS1 <- result_RSS1$par
>
>
> #Optimisation de RSS2
> init_params_RSS2 <- c(0, 0)
> result_RSS2 <- optim(init_params_RSS2, RSS2, method = "BFGS")
> beta0_optimal_RSS2 <- result_RSS2$par[1]
> beta1_optimal_RSS2 <- result_RSS2$par[2]
>
>
> #Optimisation de RSS2
> RSS3 <- function(params) {
+     beta0 <- params[1]
+     beta1 <- params[2]
+     beta2 <- params[3]
+     sum(Wi * (Yi - beta0 - beta1 * Xi - beta2 * Xi^2) ^ 2)
+ }
> init_params_RSS3 <- c(11.44, 5.51, 0)
> result_RSS3 <- optim(init_params_RSS3, RSS3, method = "L-BFGS-B", lower = c(-50
0, -500, -500), upper = c(500, 500, 500))
> beta0_optimal_RSS3 <- result_RSS3$par[1]
> beta1_optimal_RSS3 <- result_RSS3$par[2]
> beta2_optimal_RSS3 <- result_RSS3$par[3]
>
```

```
> cat("Résultat optimal pour RSS1:\n")
Résultat optimal pour RSS1:
> cat("Beta optimal pour RSS1:", beta_optimal_RSS1, "\n\n")
Beta optimal pour RSS1: 8.41547

> cat("Résultat optimal pour RSS2:\n")
Résultat optimal pour RSS2:
> cat("Beta0 optimal pour RSS2:", beta0_optimal_RSS2, "\n")
Beta0 optimal pour RSS2: 11.44301
> cat("Beta1 optimal pour RSS2:", beta1_optimal_RSS2, "\n\n")
Beta1 optimal pour RSS2: 5.507998

> cat("Résultat optimal pour RSS3:\n")
Résultat optimal pour RSS3:
> cat("Beta0 optimal pour RSS3:", beta0_optimal_RSS3, "\n")
Beta0 optimal pour RSS3: 19.50651
> cat("Beta1 optimal pour RSS3:", beta1_optimal_RSS3, "\n")
Beta1 optimal pour RSS3: 0.6870021
> cat("Beta2 optimal pour RSS3:", beta2_optimal_RSS3, "\n\n")
Beta2 optimal pour RSS3: 0.6689546
```

Comparaison des valeurs pour RSS1 et RSS2 :

RSS1(β) :

Valeur algébrique : beta = 8.41

Valeur optimisée : beta = 8.41

Comparaison : La différence entre la valeur algébrique et la valeur optimisée est très faible (environ 0.005). Cela montre que les deux méthodes sont en très bon accord, confirmant que le calcul algébrique fournit une solution presque identique à celle obtenue par optimisation numérique.

RSS2 :

Valeurs algébriques : Beta_0 = 11.44 et Beta_1 = 5.50

Valeurs optimisées : Beta_0 = 11.44 Beta_1 = 5.50

Comparaison : Ici aussi, la différence est très faible. Pour beta_0, la différence est de l'ordre de 0.003, et pour beta_1, elle est de l'ordre de 0.008. Ces écarts mineurs peuvent être attribués à des approximations algébriques ou des différences dans les méthodes d'optimisation, mais elles restent négligeables.