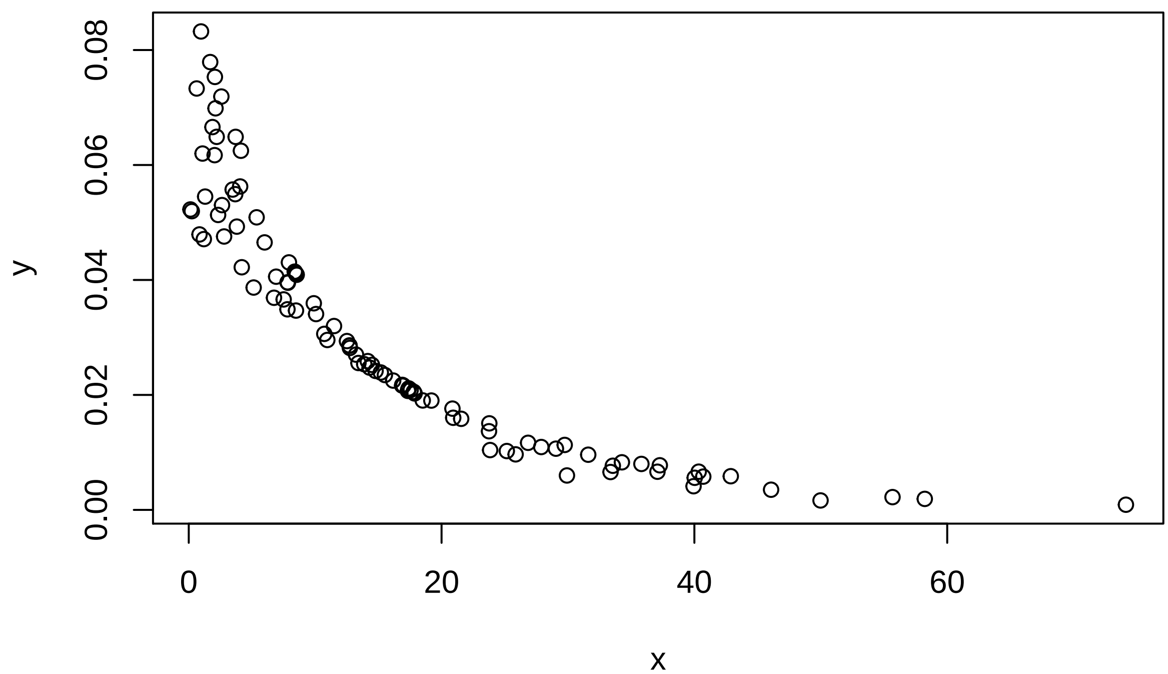
Recherche du paramètre optimal

Olivier Turcotte

Objectif



L'objectif de ce présent document est de tenter de trouver un θ optimal afin de prédire les prochaines valeurs en Y à partir de X en supposant que la distribution suit une loi exponentielle.

Estimations

Première approche

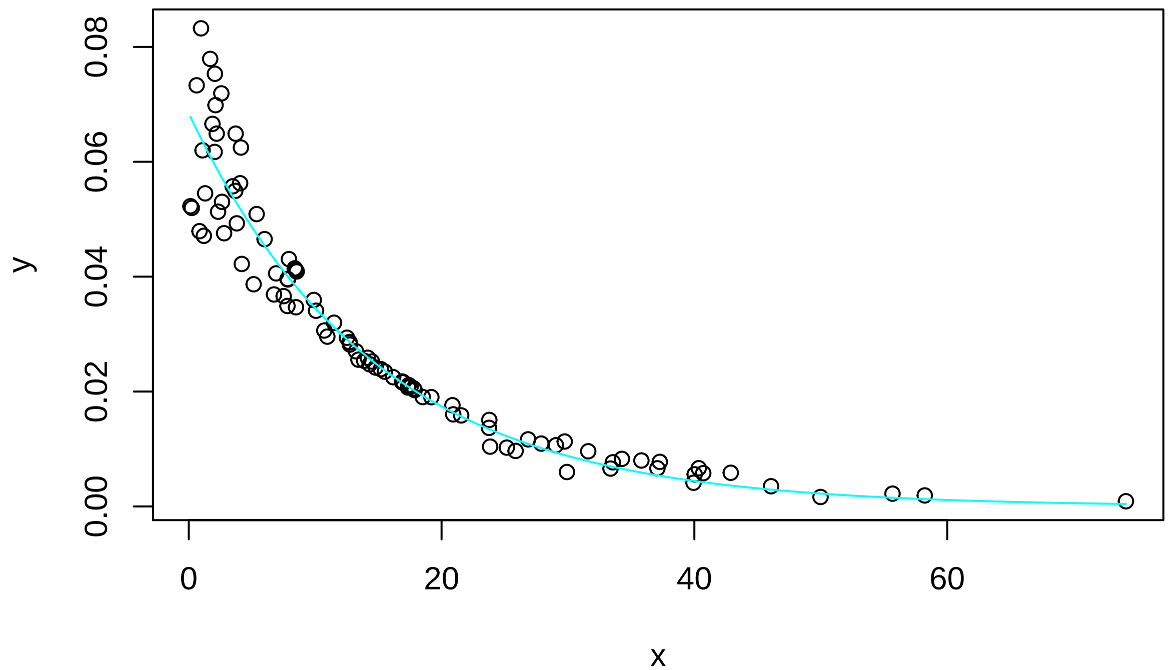
Mon idée première : réduire la distance quadratique entre mes valeurs de X et Y en posant l'hypothèse $h_\theta(x_i) = \theta e^{\theta x_i}$. Plus cette fonction est près de 0, plus les valeurs de $h_\theta(x_i)$ et y_i sont proches.

Ainsi, la 'loss function' à optimiser est donc

$$\begin{aligned}
J(\theta) &= \frac{1}{2m} * \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2 \\
\Rightarrow \frac{d}{d\theta} J(\theta) &= \frac{d}{d\theta} \frac{1}{2m} * \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2 \\
&= \frac{2}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i) * \frac{d}{d\theta} (h_\theta(x_i) - y_i) \\
&= \frac{1}{m} \sum_{i=1}^{m} (\theta e^{\theta x_i} - y_i) * \frac{d}{d\theta} (\theta e^{\theta x_i} - y_i) \\
&= \frac{1}{m} \sum_{i=1}^{m} (\theta e^{\theta x_i} - y_i)(x_i \theta - 1) * -e^{\theta x_i} = 0 \\
\Rightarrow 0 &= \sum_{i=1}^{m} (\theta e^{-\theta x_i} - y_i)(x_i \theta - 1) e^{-\theta x_i}
\end{aligned}$$

```
theta[1] <- uniroot(function(a) sum((a*exp(-a*x)-y)*(x*a-1)*exp(-a*x)) - 0, c(0,1))$root
```

En effectuant ce calcul, on obtient $\theta_1 = 0.0684386$.



Seconde approche

La seconde approche est une mise en oeuvre de la méthode du gradient basé sur les mêmes principes que la première approche. Ainsi, on a toujours $J(\theta) = \frac{1}{2m} * \sum_{i=1}^{m}(\theta e^{\theta x_i} - y_i)^2$.

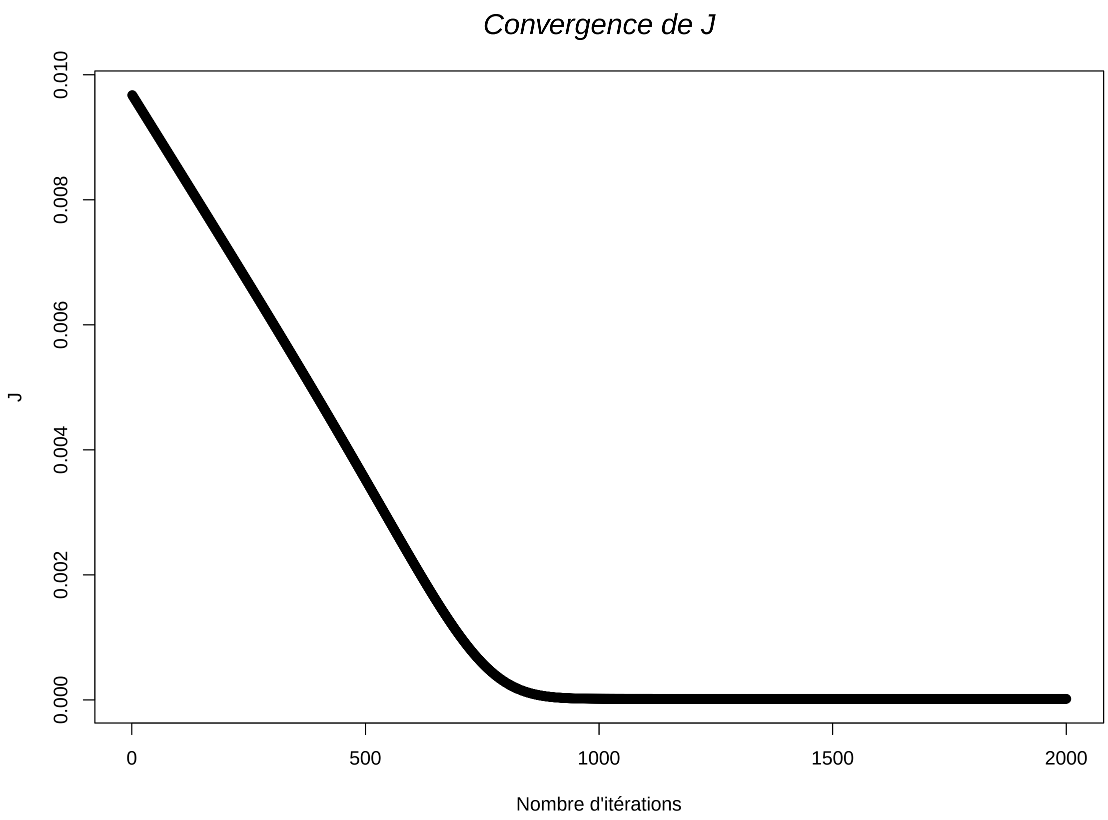
L'algorithme du gradient:

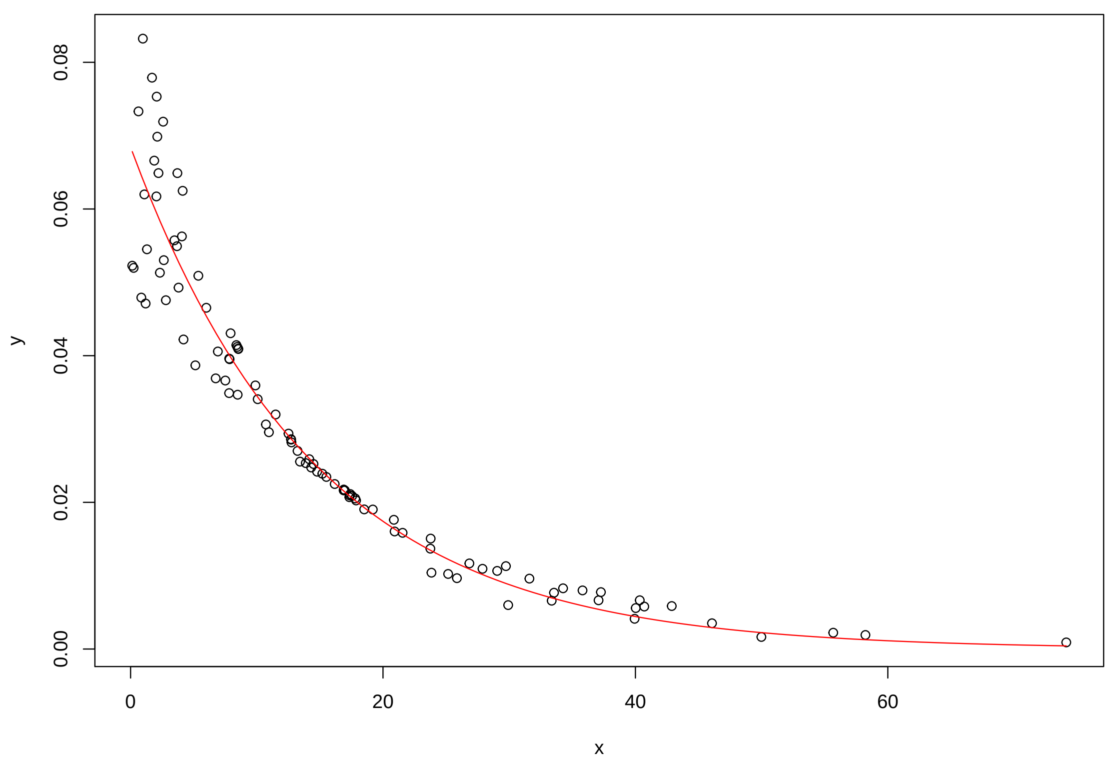
- Effectuer jusqu'à convergence: $\theta = \theta + \alpha * \frac{d}{d\theta}J(\theta)$

Sous forme de code :

```
m <- length(y)
alpha <- 0.1
n <- 2000
theta[2] = 1
j <- numeric(0)
for (i in seq_len(n)) {
    theta[2] <- theta[2] + alpha / m *
        sum((theta[2] * exp(-theta[2] * x) - y) *
        (x * theta[2] -1) * exp(-theta[2] * x))

    j[i] <- costCPT(theta[2]) ## Utilisé à des fins de graphiques
}
```


Convergence de J
J
0.000
0.002
0.004
0.006
0.008
0.010
0
500
1000
1500
2000
Nombre d'itérations


y
0.00
0.02
0.04
0.06
0.08
0
20
40
60
x

On obtient alors $\theta_2 = 0.0684402$.

Cependant, cet algorithme à quelques faiblesses en comparaison au premier:

1. Lenteur d'exécutions
2. Nécessite une paramétrisations initiale de α et θ.
    - Dépendemment des paramètre initiales utilisés, l'algorithme peut converger vers un minimum local et non global.

Troisième approche

La troisième approche consiste à prendre le logarithme de chacun des côtés de l'équation $Y = h_\theta(x)$:

$$
\begin{aligned}
Y &= \theta * e^{-\theta * x} \\
\Rightarrow \ln(Y) &= \ln(\theta) - \theta x \\
\Rightarrow Z &= \theta_a + \theta_b x, \theta_a = \ln(\theta), \theta = -\theta_b
\end{aligned}
$$

On se retrouve donc dans une belle situation de régression linéaire. En me basant sur les équations retrouvés dans ce *document*, on peut définir les paramètres θ_a et θ_b comme ceci:

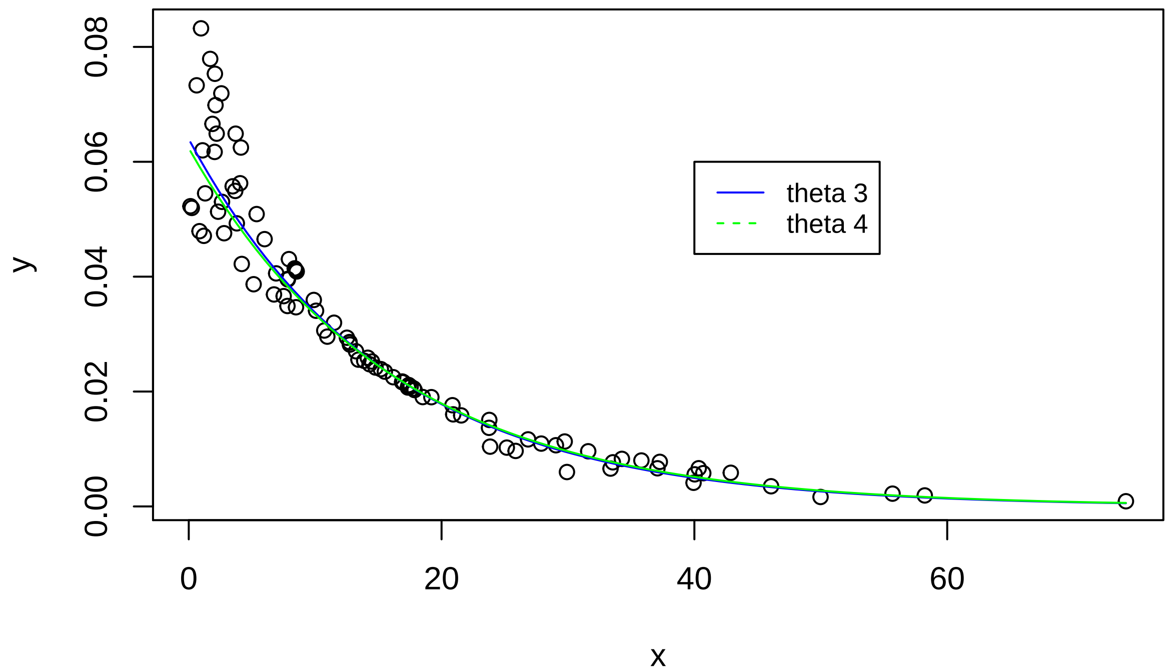
$$
\begin{aligned}
\theta_a &= \bar{Z} - \theta_b * \bar{X} \\
\theta_b &= \frac{\sum_{i=1}^{n} X_i Z_i - n\bar{X}\bar{Z}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}}
\end{aligned}
$$

Sous forme de code :

```
n <- length(y)
x_b <-  sum(x)/n
z_b <- sum(log(y))/n
a <- sum(x*log(y))-n*x_b*z_b
b <- sum(x^2)-n*x_b^2

theta[4] <- -a/b
theta[3] <- exp(z_b+theta[4]*x_b)
```

Bon, n'ayant pas encore couvert la régression linéaire, je ne suis pas très à l'aise avec cette matière. J'avais prévu que les deux valeurs retournés seraient identique, mais ce n'est pas le cas. Néanmoins, je vais procéder et voir ce que ça donne graphiquement.

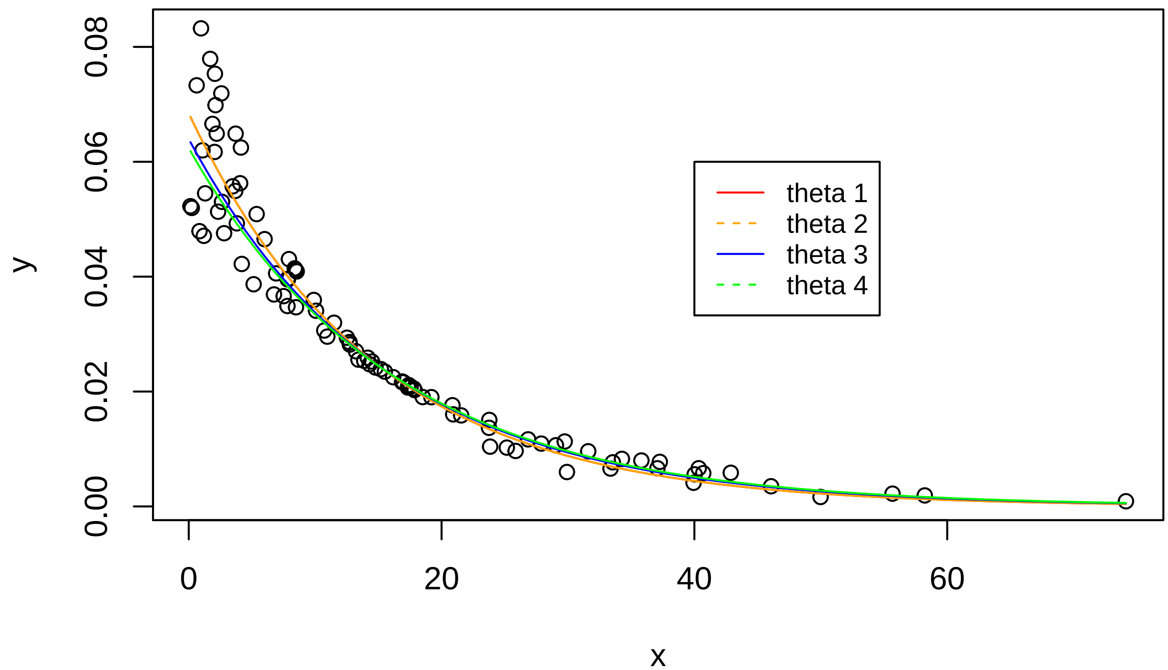


Pas mal du tout! Il ne reste plus qu'à comparer les 4 valeurs de theta que j'ai jusqu'à présent:

Theta	Valeurs
θ_1	0.0684386
θ_2	0.0684402
θ_3	0.0639454
θ_4	0.0623672

Je constate que θ_1 et θ_2 sont très près. Mon explication: les deux étant basé sur l'optimisation de la fonction de distance quadratique, il est normal que leurs valeurs soient très près.

Il ne reste plus qu'à tout mettre sur un même graphique pour voir ce que ça donne.

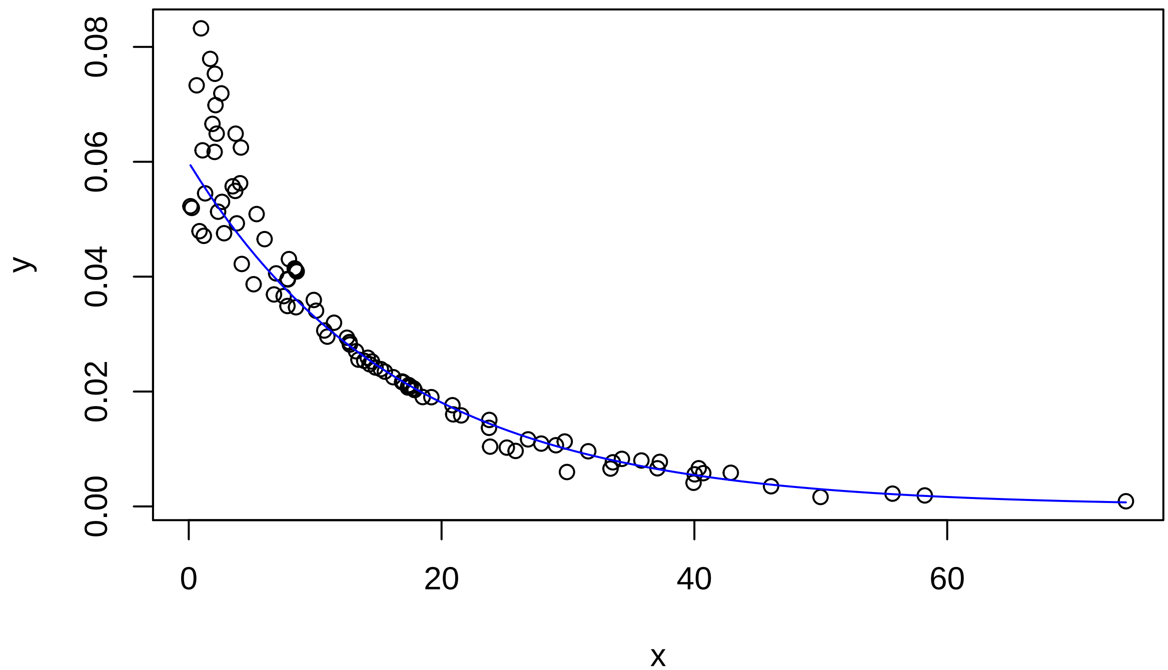


Quatrième approche

Ici, j'utilise la méthode du maximum de vraisemblance afin d'estimer θ.

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\
&= \prod_{i=1}^{n} \theta * e^{-\theta x_i} \\
&= \theta^n * e^{-\theta * \sum_{i=1}^{n} x_i} \\
\Rightarrow \ln(L(\theta)) &= n\ln(\theta) - \theta \sum_{i=1}^{n} x_i \\
\Rightarrow \frac{d}{d\theta}\ln(L(\theta)) &= \frac{n}{\theta} - \sum_{i=1}^{n} x_i = 0 \\
\Rightarrow \theta &= \frac{n}{\sum_{i=1}^{n} x_i} \\
\theta &= 0.0598772
\end{aligned}
$$

```
theta[5] <- n/sum(x)
```



Theta optimal

Désormais, j'ai 5 valeurs de theta différentes. La question est, lequel dois-je utiliser afin d'effectuer des prédictions les plus précises possibles?

Theta	Valeurs
θ_1	0.0684386
θ_2	0.0684402
θ_3	0.0639454
θ_4	0.0623672
θ_5	0.0598772

Dois-je me baser sur la variance de mes estimateurs? Si oui, comment trouver la variance des estimateurs utilisés dans la première et seconde approche? Si non, quel piste dois-je emprunter pour déclarer quel theta utiliser?