Exercice 6

Les données pour cet exercice sont stockées dans le fichier `RM04data.csv`. Commençons par charger ces données. Nous ne prennons que les lignes 5 et 6.

```r
# La fonction read_csv2 est issue du package readr qui permet des imports
# exports facilités dans R
df <- read.csv("../data/RM04data.csv", sep = ',', header = FALSE)
df <- df[5:6,]

columns_names = c("xi","ui")
observations_labels <- paste('x',as.character(1:200),sep = '')

rownames(df) <- columns_names
colnames(df) <- observations_labels

df <- t(df)

head(df)
```

```
##              xi ui
## x1 0.003407508  1
## x2 0.014933541  1
## x3 0.009207314  1
## x4 0.650709545  1
## x5 0.486458950  1
## x6 0.017409099  0
```

Dans ce jeu de données, on a pour chaque observations (200) :

- la durée inter-panne : x_i,
- le type de maintenance : $u_i = \{0, 1\}$ avec $u_i = 1$: Maintenance préventive et $u_i = 0$: Maintenance corrective.

On suppose que les durées inter-pannes suivent une loi exponentielle avec une maintenance parfaite.

Vérification de l'hypothèse exponentielle

Si les maintenance sont parfaite, on observera une absence de tendance sur les durées inter-panne.

Dans ce cas les X_i sont *iid*.

On peut vérifier cette hypothèse avec un test de Spearman.

Le test se présente ainsi :

$$\begin{cases} H_0 : \text{les données ne présentent pas de tendance} \\ H_1 : \text{les données présentent une tendance} \end{cases}$$

On calcul la statistique de test :

$$T = \sqrt{n-2}\frac{\rho_s}{\sqrt{1-\rho_s^2}} \sim St(n-2)\rho_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n}$$

On va refuser la présence de tendance si

$$|T| < F^{-1}_{St(n-2)}\left(1 - \frac{\alpha}{2}\right)$$

Avec : $d_i = i - R_i$, et R_i le rang des observations

On peut d'abord construire les vecteurs R_i, i, et d_i.

```
n <- length(df[,"xi"])
i <- 1:n
Ri <- match(df[,"xi"], sort(df[,"xi"]))
di <- i - Ri
```

On calcul ensuite le statistique de test

```
ps <- 1 - (6*sum(di^2))/(n^3-3)
T <- sqrt(n-2) * ((ps)/(sqrt(1-ps^2)))
sprintf("Valeur de la statistique de test : %f", T)
```

```
## [1] "Valeur de la statistique de test : -0.630558"
```

On regarde le quantile de la loi de Student :

```
qt(1-0.05/2,n-2)
```

```
## [1] 1.972017
```

Dans notre cas on a : $|T| < F^{-1}_{St(n-2)}\left(1 - \frac{\alpha}{2}\right)$, donc on va rejeter la présence d'une tendance, conformément à l'hypothèse initiale.

On peut ensuite tester l'hypothèse de la loi exponentielle avec un test de Bartlett.

```
compute_bartlett_stat <- function(tn, xi) {

  # Fonction qui calcule la statistique de test du test de Bartlett

  n <- length(xi)

  2*n*(log(tn/n)-sum(log(xi))/n)/(1+((n+1)/(6*n))) %>% return()
}
```

On ne dispose que des x_i il faut donc calculer les t_i :

```
xi <- df[,"xi"]
ti <- cumsum(xi)
```

```
compute_bartlett_stat(as.vector(ti[n]), xi)
```

```
## [1] 237.8812
```

```
qchisq(0.05/2, n-1)
```

```
## [1] 161.8262
```

```
qchisq(1-0.05/2, n-1)
```

```
## [1] 239.9597
```

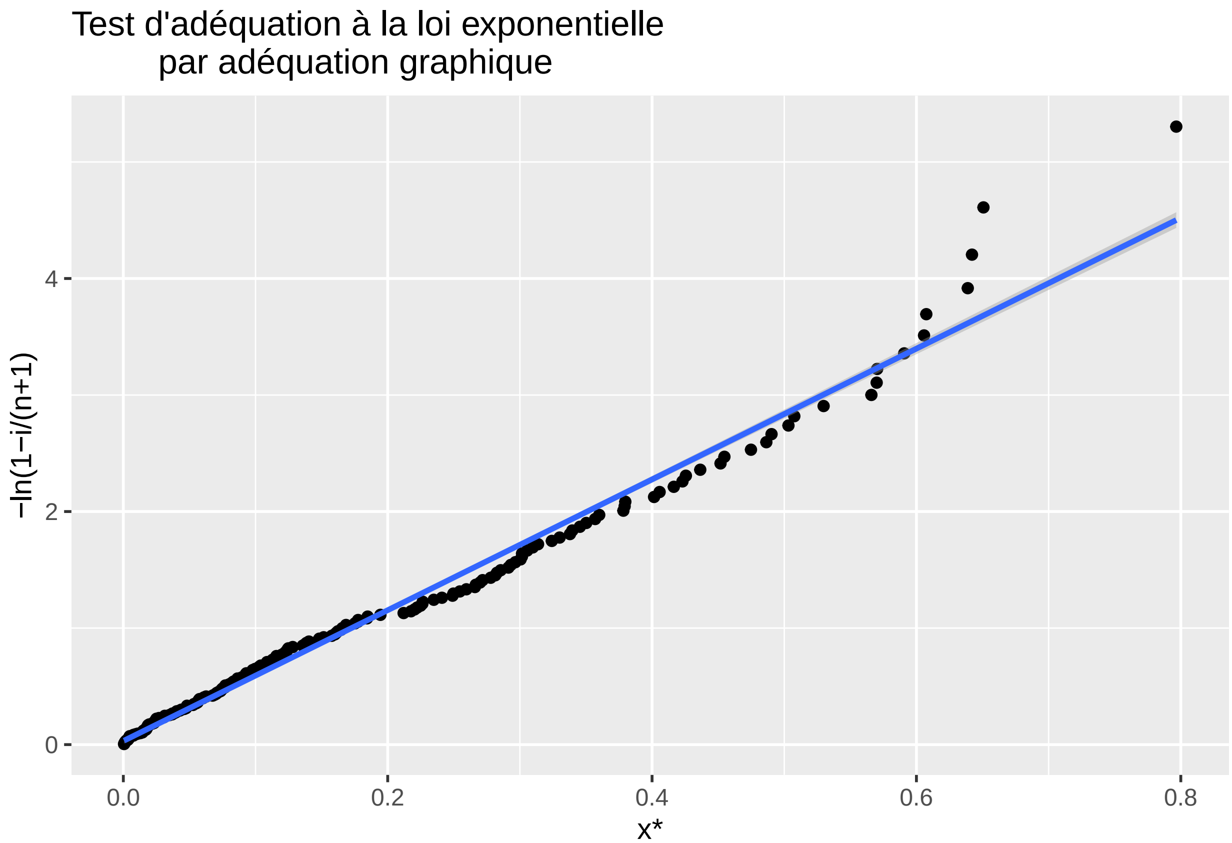
Dans notre cas, on va accepter H_0 de justesse car on a $237.8812 \in [161.8262; 239.9597]$.

On peut vérifier à l'aide d'un test par adéquation graphique :

```
ggplot() +
    geom_point(aes(x = sort(xi), y = -log(1-i/(n+1)))) +
    geom_smooth(aes(x = sort(xi), y = -log(1-i/(n+1))),
                method = lm) +
    labs(x = "x*",
```

```
        y = "-ln(1-i/(n+1)",
        title = "Test d'adéquation à la loi exponentielle
        par adéquation graphique")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



On trouve une droite de régression. On peut chercher les paramètres de la régression linéaire pour déterminer les paramètres de la loi exponentielle.

```
lm(-log(1-i/(n+1)) ~ sort(xi)) %>% summary()
```

```
##
## Call:
## lm(formula = -log(1 - i/(n + 1)) ~ sort(xi))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.20628 -0.08132  0.01726  0.03967  0.92745
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.03121    0.01259    2.48    0.014 *
## sort(xi)     5.61157    0.05258  106.72   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1251 on 198 degrees of freedom
```

```
## Multiple R-squared:  0.9829, Adjusted R-squared:  0.9828
## F-statistic: 1.139e+04 on 1 and 198 DF,  p-value: < 2.2e-16
```

On a $R^2 = 0.9828$, et $p_{value} = 2.2 \times 10^{-16}$, ce qui nous conforte dans l'adéquation des durées inter-pannes à une loi exponentielle.

On a donc une estimation $\lambda_{gph} = 5.61157$.

On peut aussi chercher une estimation par méthode des moments :

```r
1/mean(xi)
```

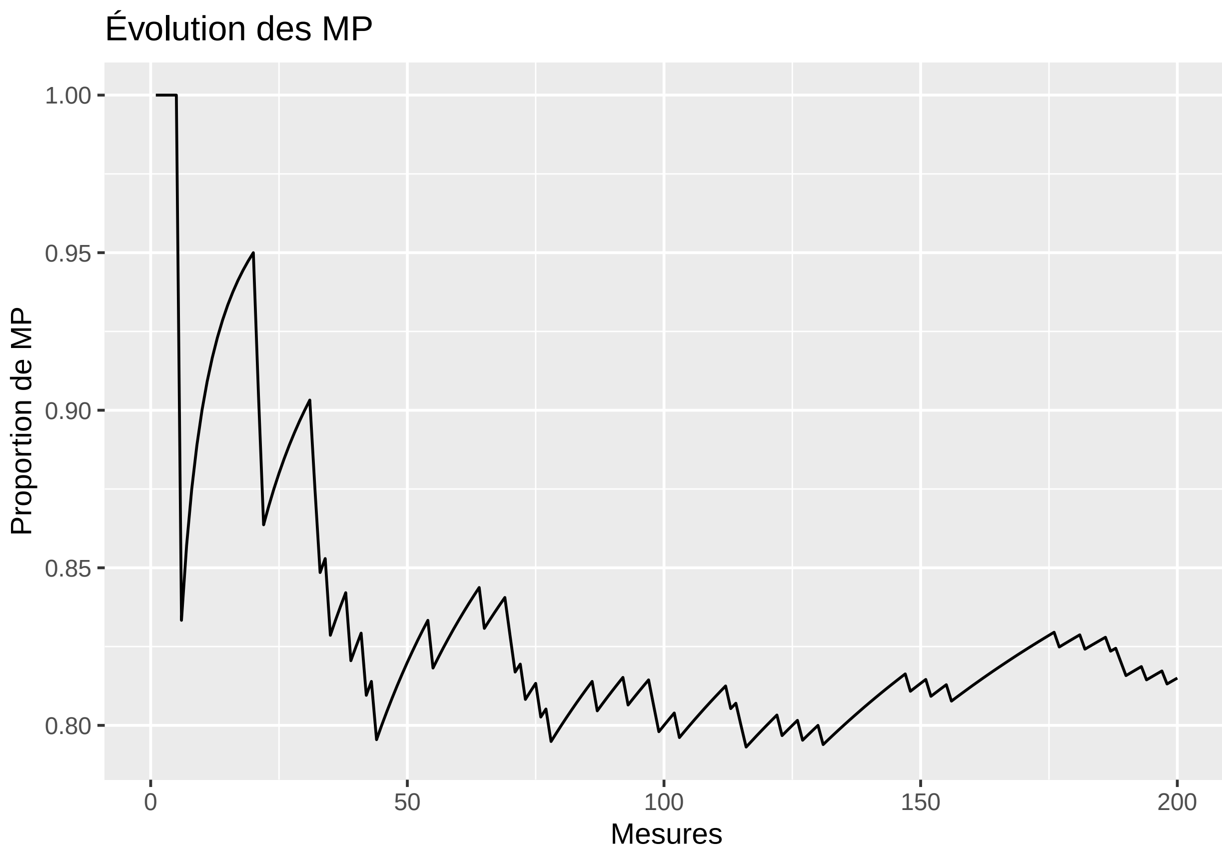
```
## [1] 5.870253
```

Dans ce cas on a : $\lambda_{gph} = 5.870253$.

Donc on peut sans problème conserver l'hypothèse selon laquelle les x_i suivent une loi exponentielle.

On peut tracer la courbe de la proportion de MP :

```r
ggplot() +
    geom_line(aes(x = i, y = cumsum(df[,"ui"])/i)) +
    labs(x = "Mesures",
         y = "Proportion de MP",
         title = "Évolution des MP")
```



Après quelques mesure, la proportion de MP semble se stabiliser vers 80%. On effectue donc 80% de maintenances préventives d'après les données.

Politique de maintenance

Après une maintenance, on suppose que la durée jusqu'à une MP est Y et la durée jusqu'à une MC est Z avec Y et Z indépendants.

Dans ce cas la durée effective jusqu'à la prochaine maintenance est $W = min(Y, Z)$.

On suppose maintenant que $Y \sim Exp(\mu)$ et $Z \sim Exp(\lambda)$.

Dans ce cas on cherche : $P(W \leq w)$:

$$P(w \leq w) = P(W \leq w|z \leq w)P(z \leq w) + P(W \leq w|z > w)P(z > w)$$

Avec :

$P(z \leq w) = 1 - e^{-\lambda w} P(z > w) = e^{-\lambda w} P(W \leq w|z \leq w) = 1$en effet quelque soit la valeur de y alors$min(y, z) < w$ si $z \leq wP$

Dans ce cas :

$$P(W \leq w) = (1 - e^{-\lambda w}) + (1 - e^{-\mu w})e^{-\lambda w} = 1 - e^{-(\mu+\lambda)w}$$

Donc $W \sim Exp(\mu + \lambda)$ d'espérance $\frac{1}{\mu+\lambda}$

On cherche ensuite la probabilité π que la prochaine maintenance soit une MP :

$$\pi = P(y < z) = \int_{z \in \Omega_Z} F_Y(z) f_Z(z) dz = \int_{z \in \Omega_Z} (1 - e^{-\mu z}) \lambda e^{-\lambda z} dz = \int_{z \in \Omega_Z} \lambda e^{-\lambda z} - \lambda e^{-\lambda z} e^{-\mu z} = \frac{\lambda}{\lambda} - \frac{\lambda}{\lambda + \mu} = 1 - \frac{\lambda}{\lambda + \mu} = \frac{\mu}{\lambda +}$$

Dans notre jeu de données, la durée moyenne inter-pannes observée est de :

```
sprintf("Moyenne de durée inter-pannes : %f", mean(xi))
```

```
## [1] "Moyenne de durée inter-pannes : 0.170350"
```

Avec une proportion de MP de :

```
sprintf("Proportion de MP : %f", sum(df[,"ui"])/n)
```

```
## [1] "Proportion de MP : 0.815000"
```

Dans ce cas on a : $\pi = 0.815000$ et $E[W] = 0.170350$, soit :

$$\begin{cases} \pi = \frac{\mu}{\lambda+\mu} \\ E[W] = \frac{1}{\lambda+\mu} \end{cases}$$

Donc : $\pi = \mu * E[W]$ donc $\mu = \frac{\pi}{E[W]}$, et $\lambda = \frac{1}{E[W]}(1 - E[W]\mu)$.

On peut alors estimer les paramètres :

```
sprintf("mu = %f", 0.815000/0.170350)
```

```
## [1] "mu = 4.784268"
```

```
sprintf("lambda = %f", (1-0.170350*0.815000)/0.170350)
```

```
## [1] "lambda = 5.055267"
```

Donc on a $Y \sim Exp(4.784268)$ et $Z \sim Exp(5.055267)$.

Indépendance entre U et W

On va ensuite chercher à montrer que les variables aléatoires U et W sont indépendante. On a :

- U, le type de maintenance avec $P(U = 0) = 1 - \pi$ et $P(U = 1) = \pi$
- W, la durée inter panne, de loi exponentielle de paramètre $\mu + \lambda$.

Deux variables aléatoires A et B sont indépendantes si on a : $P(A = a \cap B = b) = P(A = a)P(B = b), \forall a \in \Omega_A \forall b \in \Omega_B$

On cherche donc à montrer que par exemple $P(W > w|U = 1) = P(W > w)P(U = 1)$. Avec :

- $P(W > w) = e^{-(\mu+\lambda)w}$
- $P(U = 1) = \pi$

$$P(W > w|U = 1) = P(min(y, z) > w|y < z) = \int_{z\in\Omega_z} P(y > w|Z = z, y < z)f_Z(z)dz$$

$$z < w, y < z \Rightarrow y < z < w \Rightarrow P(y > w|Z = z, y < z) = 0 z > w, y < z \Rightarrow P(y > w|Z = z, y < z) = P(y \in [w, z]) = R_Y(w) - R$$

Donc :

$$\begin{aligned} P(W > w|U = 1) &= \int_{z\in\Omega_z} P(y > w|Z = z, y < z)f_Z(z)dz \\ &= \int_z^w 0 f_Z(z)dz + \int_w^{+\infty} (R_Y(w) - R_Y(z))f_Z(z)dz \\ &= \int_w^{+\infty} \left(e^{-\mu w} - e^{-\mu z}\right) \lambda e^{-\lambda z}dz \\ &= \lambda e^{-\mu w} \int_w^{+\infty} e^{-\lambda z}dz - \lambda \int_w^{+\infty} e^{-(\mu+\lambda)z}dz \\ &= e^{-\mu w}e^{-\lambda w} - e^{-(\mu+\lambda)w} \tfrac{\lambda}{\lambda+\mu} = \tfrac{\mu}{\lambda+\mu} e^{-(\mu+\lambda)w} \\ &= P(U = 1)P(W > w) \end{aligned}$$

Donc, la probabilité d'avoir une maintenance préventive est indépendante de la valeur de W.