STT 3260
Devoir 2

Mathieu Lemire

10 Octobre 2025

Exercice 1

a)

Loi Log-Normale

La vraisemblance de la loi Log-Normale est:

$$\mathcal{L}(\mu, \sigma) = \prod_{i=1}^{n} f_T(t_i \mid \mu, \sigma) = \prod_{i=1}^{n} \frac{1}{t_i \sigma} \phi\left(\frac{\log(t_i) - \mu}{\sigma}\right)$$

Sa log-vraisemblance est:

$$l(\mu, \sigma) = \sum_{i=1}^{n} \left[-\log(t_i) - \log(\sigma) + \log\left(\phi\left(\frac{\log(t_i) - \mu}{\sigma}\right)\right)\right]$$

On trouve l'EVM de μ :

$$\begin{aligned}
\frac{\partial}{\partial \mu} l(\mu, \sigma) &= \sum_{i=1}^{n} \frac{\left(\phi\left(\frac{\log(t_i)-\mu}{\sigma}\right)\right)'}{\phi\left(\frac{\log(t_i)-\mu}{\sigma}\right)} = 0 \\
&= \sum_{i=1}^{n} \frac{-(\log(t_i) - \mu)}{\sigma} \frac{-1}{\sigma} \frac{\phi\left(\frac{\log(t_i)-\mu}{\sigma}\right)}{\phi\left(\frac{\log(t_i)-\mu}{\sigma}\right)} = 0 \\
&= \sum_{i=1}^{n} (\log(t_i) - \mu) = 0 \\
\hat{\mu} &= \frac{\sum_{i=1}^{n} \log(t_i)}{n}
\end{aligned}$$

On obtient

$$\hat{\mu} \approx 2.137294$$

On calcule l'EVM de σ:

$$\begin{aligned}
\frac{\partial}{\partial\sigma} l(\mu,\sigma) &= \sum_{i=1}^{n}\left[-\frac{1}{\sigma}+\frac{\left(\phi\left(\frac{\log(t_i)-\mu}{\sigma}\right)\right)'}{\phi\left(\frac{\log(t_i)-\mu}{\sigma}\right)}\right]=0\\
&= -\frac{n}{\sigma}+\sum_{i=1}^{n}\left[\frac{-\frac{\log(t_i)-\mu}{\sigma}\phi\left(\frac{\log(t_i)-\mu}{\sigma}\right)}{\phi\left(\frac{\log(t_i)-\mu}{\sigma}\right)}\frac{\log(t_i)-\mu}{-\sigma^2}\right]=0\\
&= -\frac{n}{\sigma}+\sum_{i=1}^{n}\left[\frac{(\log(t_i)-\mu)^2}{\sigma^3}\right]=0\\
n\sigma^2 &= \sum_{i=1}^{n}(\log(t_i)-\mu)^2\\
\hat{\sigma} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(t_i)-\hat{\mu})^2}
\end{aligned}$$

On obtient

$$\hat{\sigma}\approx 0.7455794$$

Donc, on obtient un modèle log-normal pour les données avec paramètres estimés $\hat{\mu}=2.137$ et $\hat{\sigma}=0.746$

Loi Gamma

La vraisemblance de la loi Gamma est:

$$\mathcal{L}(\lambda,\beta)=\prod_{i=1}^{n} f_T(t_i\mid\lambda,\beta)=\prod_{i=1}^{n}\frac{\lambda^{\beta}t_i^{\beta-1}}{\Gamma(\beta)}\exp(-\lambda t_i)$$

Sa log-vraisemblance est:

$$l(\lambda,\beta)=\sum_{i=1}^{n}\left[\beta\log(\lambda)+(\beta-1)\log(t_i)-\log(\Gamma(\beta))-\lambda t_i\right]$$

On calcule la dérivé partiel de la log-vraisemblance par rapport à λ:

$$\begin{aligned}
\frac{\partial}{\partial\lambda} l(\lambda,\beta) &= \sum_{i=1}^{n}\left[\frac{\beta}{\lambda}-t_i\right]=0\\
n\frac{\beta}{\lambda} &= \sum_{i=1}^{n} t_i\\
\hat{\lambda} &= \frac{n\hat{\beta}}{\sum_{i=1}^{n} t_i}
\end{aligned}$$

On calcule la dérivé partiel de la log-vraisemblance par rapport à β:

$$\begin{aligned}
\frac{\partial}{\partial \beta} l(\lambda, \beta) &= \sum_{i=1}^{n} \left[\log(\lambda) + \log(t_i) - \frac{(\Gamma(\beta))'}{\Gamma(\beta)} \right] = 0 \\
&= \log(\hat{\lambda}) + \frac{1}{n} \sum_{i=1}^{n} \log(t_i) - \frac{(\Gamma(\hat{\beta}))'}{\Gamma(\hat{\beta})} = 0 \\
&= \log\left(\frac{n\hat{\beta}}{\sum_{i=1}^{n} t_i} \right) + \frac{1}{n} \sum_{i=1}^{n} \log(t_i) - \frac{(\Gamma(\hat{\beta}))'}{\Gamma(\hat{\beta})} = 0
\end{aligned}$$

Il faut donc calculer numériquement $\hat{\beta}$, on obtient:

$$\hat{\beta} \approx 2.399356$$

On trouve à présent l'estimateur de λ:

$$\hat{\lambda} \approx 0.2265646$$

Donc, on obtient un modèle Gamma pour les données avec paramètres estimés $\hat{\lambda} = 0.227$ et $\hat{\beta} = 2.399$

```r
data <- data.frame(
  c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21),
  c(1, 22, 3, 12, 8, 17, 2, 11, 8, 12, 2, 5, 4, 15, 8, 23, 5, 11, 4, 1, 9)
)
colnames(data) <- c("Temps", "Nombre de rechutes")

n <- sum(data$`Nombre de rechutes`)
EVM_mu_lognormale <- sum(log(data$Temps) * data$`Nombre de rechutes`) / n

EVM_sigma_lognormale <- sqrt(sum( ((log(data$Temps) - EVM_mu_lognormale)^2)
                                  * data$`Nombre de rechutes`)/n)

# Fonction à résoudre
fonction_beta <- function(beta) {
  log(n * beta / sum(data$Temps * data$`Nombre de rechutes`)) +
    sum(log(data$Temps)* data$`Nombre de rechutes`)/n -
    digamma(beta)
}

EVM_beta_gamma <- uniroot(fonction_beta,
                          interval = c(0.001, 100))$root
EVM_lambda_gamma <- n * EVM_beta_gamma / (sum(data$Temps * data$`Nombre de rechutes`))
```

b)

i.

La densité d'un modèle gamma généralisé étant

$$f(t) = \frac{\alpha\lambda^\beta}{\Gamma(\beta)} t^{\alpha\beta-1} \exp(-\lambda t^\alpha)$$

On cherche la densité de $Y = \log T$. On a $T = \exp(Y)$ et le jacobien de la transformation $\text{Jac} = \exp(Y)$. On a donc:

$$\begin{aligned} f_Y(y) = f_T(\exp(y))\ \text{Jac} &= \frac{\alpha\lambda^\beta}{\Gamma(\beta)} \exp(y\alpha\beta - 1)\exp(-\lambda\exp(y\alpha))\exp(y) \\ &= \frac{\alpha\lambda^\beta}{\Gamma(\beta)} \exp\big(y\alpha\beta - \lambda\exp(y\alpha)\big) \end{aligned}$$

On veut

$$Y = -\frac{\log\lambda - \log\beta}{\alpha} + \frac{1}{\alpha\sqrt{\beta}} W$$

En posant $\mu = -\frac{\log\lambda - \log\beta}{\alpha}$ et $\sigma = \frac{1}{\alpha\sqrt{\beta}}$, on a $Y = \mu + \sigma W$.

$$\begin{aligned} f_Y(\mu + \sigma w) &= \frac{\alpha\lambda^\beta}{\Gamma(\beta)} \exp\big((\mu+\sigma w)\alpha\beta - \lambda\exp((\mu+\sigma w)\alpha)\big) \\ &= \frac{\alpha\lambda^\beta}{\Gamma(\beta)} \exp\big(\mu\alpha\beta + \sigma w\alpha\beta - \lambda\exp(\mu\alpha + \sigma w\alpha)\big) \end{aligned}$$

On calcule ces valeurs:

- $\mu\alpha = -\frac{\log\lambda - \log\beta}{\alpha}\alpha = \log\beta - \log\lambda$
- $\sigma\alpha = \frac{1}{\alpha\sqrt{\beta}}\alpha = \beta^{-\frac{1}{2}}$

On a donc

$$\begin{aligned} f_Y(\mu + \sigma w) &= \frac{\alpha\lambda^\beta}{\Gamma(\beta)} \exp\big((\log\beta - \log\lambda)\beta + (\beta^{-\frac{1}{2}})w\beta - \lambda\exp(\log\beta - \log\lambda + \beta^{-\frac{1}{2}}w)\big) \\ &= \frac{\alpha\lambda^\beta}{\Gamma(\beta)} \beta^\beta\lambda^{-\beta} \exp\big(\beta^{\frac{1}{2}}w - \lambda\beta\lambda^{-1}\exp(\beta^{-\frac{1}{2}}w)\big) \\ &= \frac{\alpha}{\Gamma(\beta)} \beta^\beta \exp\big(\beta^{\frac{1}{2}}w - \beta\exp(\beta^{-\frac{1}{2}}w)\big) \end{aligned}$$

En multipliant par le jacobien de la transformation, Jac $= \dfrac{1}{\alpha\sqrt{\beta}}$. On retrouve la densité de W.

$$
\begin{aligned}
f_W(w) &= f_Y(\mu + \sigma w)\text{Jac} \\
&= \frac{\beta^{-\frac{1}{2}}}{\Gamma(\beta)}\beta^{\beta} \exp\left(\beta^{\frac{1}{2}} w - \beta \exp(\beta^{-\frac{1}{2}} w)\right) \\
&= \frac{Q}{\Gamma(Q^{-2})}(Q^{-2})^{Q^{-2}} \exp\left(Q^{-1} w - Q^{-2} \exp(Qw)\right) \\
&= \frac{Q}{\Gamma(Q^{-2})}(Q^{-2})^{Q^{-2}} \exp\left(\frac{Qw - \exp(Qw)}{Q^2}\right)
\end{aligned}
$$

ii.

On a:

- $\sigma = \dfrac{1}{\alpha\sqrt{\beta}} \iff \alpha = \dfrac{Q}{\sigma}$
- $\mu = -\dfrac{\log\lambda - \log\beta}{\alpha} \iff \lambda = \exp(\log\beta - \mu\alpha) = \exp\left(\log(Q^{-2}) - \mu\dfrac{Q}{\sigma}\right)$

Il suffit à présent de remplacer dans $f(t) = \dfrac{\alpha\lambda^{\beta}}{\Gamma(\beta)} t^{\alpha\beta - 1}\exp(-\lambda t^{\alpha})$

$$
\begin{aligned}
f_T(t,\mu,\sigma,Q) &= \frac{\frac{Q}{\sigma}\left(\exp\left(\log(Q^{-2}) - \mu\frac{Q}{\sigma}\right)\right)^{Q^{-2}}}{\Gamma(Q^{-2})} t^{\frac{Q}{\sigma}Q^{-2}-1} \exp\left(-\left(\exp\left(\log(Q^{-2}) - \mu\frac{Q}{\sigma}\right)\right) t^{\frac{Q}{\sigma}}\right) \\
&= \frac{\frac{Q}{\sigma}\exp\left(Q^{-2}\left(\log(Q^{-2}) - \mu\frac{Q}{\sigma}\right)\right)}{\Gamma(Q^{-2})} t^{\frac{Q^{-1}}{\sigma}-1} \exp\left(-Q^{-2}\exp\left(-\mu\frac{Q}{\sigma}\right) t^{\frac{Q}{\sigma}}\right) \\
&= \frac{\frac{Q}{\sigma}\exp\left(Q^{-2}\log(Q^{-2}) - \mu\frac{Q^{-1}}{\sigma}\right)}{\Gamma(Q^{-2})} t^{\frac{Q^{-1}}{\sigma}-1} \exp\left(-Q^{-2}\exp\left(-\mu\frac{Q}{\sigma}\right) t^{\frac{Q}{\sigma}}\right) \\
&= \frac{Q(Q^{-2})^{Q^{-2}}\exp\left(-\mu\frac{Q^{-1}}{\sigma}\right)}{\sigma\Gamma(Q^{-2})} t^{\frac{Q^{-1}}{\sigma}-1} \exp\left(-Q^{-2}\exp\left(-\mu\frac{Q}{\sigma}\right) t^{\frac{Q}{\sigma}}\right)
\end{aligned}
$$

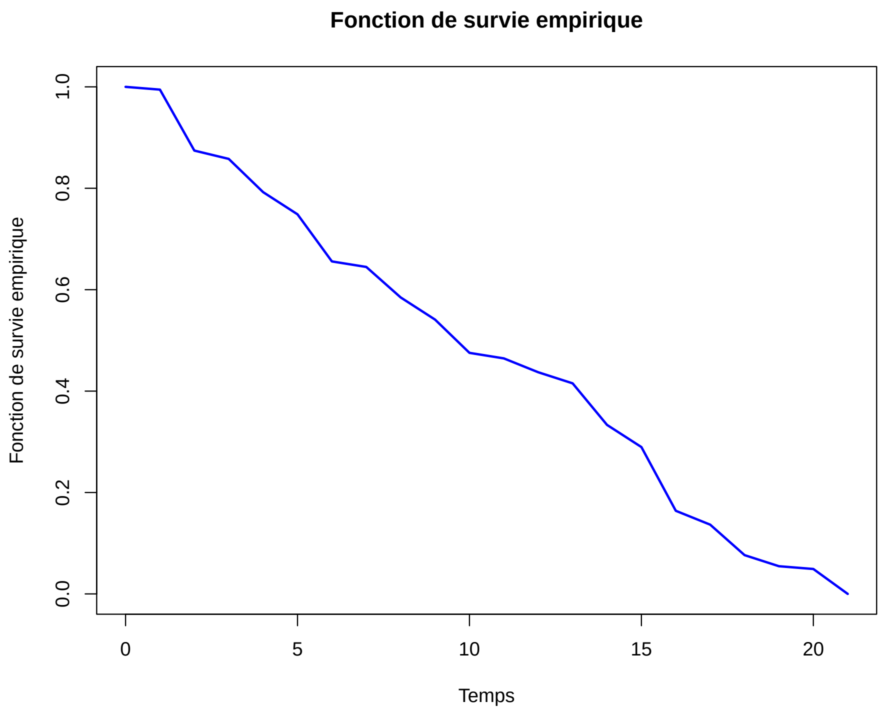
iii.

```
library(flexsurv)
d <- read.table("vers/donneesRechute.txt", header=TRUE)

modele_gengamma <- flexsurvreg(Surv(Temps) ~ 1,data= d ,dist
     = "gengamma")

modele_gengamma$loglik
[1] -561.2214

modele_gengamma$coefficients
       mu      sigma         Q
 2.850828 -1.384425  3.437482
```

La log-vraisemblance maximale du modèle est $\ell(\mu, \sigma, Q) \approx -561.2214$ et la vraisemblance est donc $L(\mu, \sigma, Q) \approx \exp(-561.2214)$. Les estimateurs des paramètres sont $\hat{\mu} \approx 2.850828 \quad \hat{\sigma} \approx -1.384425 \quad \hat{Q} \approx 3.437482$.

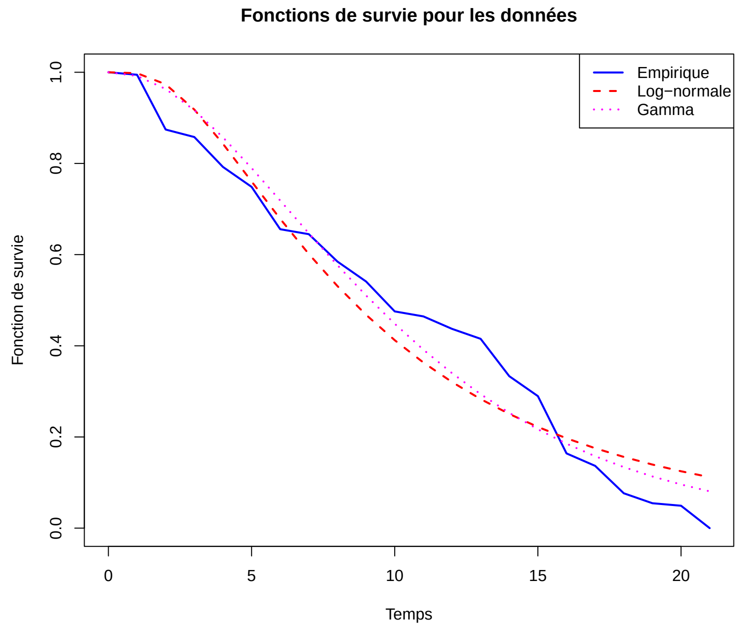
c)

```
fonction_repartion_empirique <- function(t, data){
  return( sum((data <= t)) / length(data) )
}
fonction_survie_empirique <- function(t, data){
  return(1-  fonction_repartion_empirique(t, data) )
}

t <- seq(0, max(d$Temps) , 1)
s <- numeric(max(d$Temps) + 1)
for (i in 1 :max(d$Temps)){
  s[i] <- fonction_survie_empirique(i - 1, d$Temps)
}
rm(i)
plot(t, s, pch=19, col = "blue", main="Fonction de survie
    empirique",
     ylab = "Fonction de survie empirique",
     xlab="Temps")
```



d)

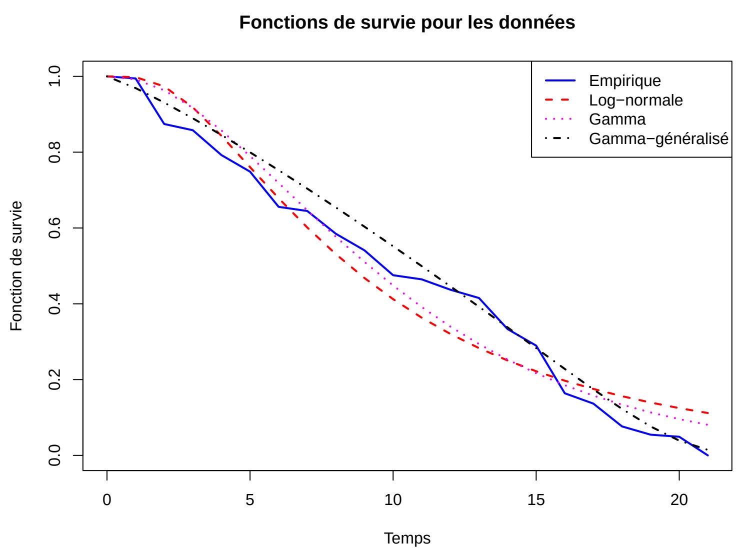
```r
plot(t, s, pch=19, col = "blue", main="Fonctions de survie
    pour les données",
      ylab = "Fonction de survie",
      xlab="Temps", type="l", lwd=2)

# Log-normale
survie_log_norm <- 1 - pnorm((log(t) - EVM_mu_lognormale) /
    EVM_sigma_lognormale)
lines(t, survie_log_norm, lwd =2, col="red" , lty=2)

# Gamma
survie_gamma <- 1 - pgamma(t, shape = EVM_beta_gamma, rate =
     EVM_lambda_gamma)
lines(t, survie_gamma, col="magenta", lty=3, lwd = 2)

legend("topright",
       legend=c("Empirique", "Log-normale", "Gamma"),
       col=c("blue","red","magenta"), lty=c(1,2,3), lwd=2)
```



Les deux modèles semblent appropriés, ils suivent bien la fonction de survie empirique. Je trouve personnellement que le modèle Gamma est légèrement meilleur, il semble en général mieux suivre la fonction de survie empirique.

e)

```r
plot(t, s, pch=19, col = "blue", main="Fonctions de survie pour les données",
     ylab = "Fonction de survie",
     xlab="Temps", type="l", lwd=2)

# Normale
lines(t, survie_log_norm, lwd =2, col="red" , lty=2)

# Gamma
lines(t, survie_gamma, col="magenta", lty=3, lwd = 2)

# Gamma-généralisé
#         mu      sigma          Q
#     2.850828 -1.384425   3.437482

survie_gengamma <- 1 - pgengamma(t, mu = 2.850828, sigma = exp(-1.384425), Q = 3.437482)
lines(t, survie_gengamma, col="black", lty= 4, lwd=2)

legend("topright",
       legend=c("Empirique", "Log-normale", "Gamma","Gamma-généralisé"),
       col=c("blue","red","magenta","black"), lty=c(1,2,3,4), lwd=2)
```



Le modèle gamma généralisé semble mieux représenter les données car elle suit mieux la tendance globale de la fonction de survie empirique. Il est naturel que la gamma-généralisé représente mieux les données que la Gamma car cette dernière est un modèle emboîté dans la gamma-gnénéralisé et la gamma semblait déjà mieux représenter les données que la log-normale. D'après le graphique la gamma-généralisé semble représenter significativement mieux les données. Toutefois, il faudrait faire des tests statistiques pour vérifier cette hypothèse.

f)

Puisque lorsque Q tend vers 0, la gamma-généralisé converge en loi vers la log-normale, il faut tester l'hypothèse nulle $H_0 : Q = 0$ (On prend le modèle log-normale) contre l'alternative $H_1 : Q \neq 0$ (On garde le modèle gamma-généralisé. On fait le test du rapport de vraisemblance:

$$RV = -2(\ell(\hat{\mu}_0, \hat{\sigma}_0, Q = 0) - \ell(\hat{\mu}, \hat{\sigma}, \hat{Q}))$$

Puisque $\ell(\hat{\mu}_0, \hat{\sigma}_0, Q = 0)$ est égale à la log-vraisemblance de la log-normale on calcule

```
modele_log_normale <- flexsurvreg(Surv(Temps) ~ 1,data= d ,
    dist = "lnorm")
modele_log_normale$loglik

stat_vraisemblance <- -2*(modele_log_normale$loglik - modele
    _gengamma$loglik)

> stat_vraisemblance
[1] 71.68299
```

On a un paramètre de différent entre les modèles donc on utilise un χ^2_1. On test au niveau 0.01

```
> qchisq(0.99,1)
[1] 6.634897

> stat_vraisemblance > qchisq(0.99,1)
[1] TRUE
```

On rejette donc l'hypothèse nulle. On garde le modèle gamma-généralisé.

Exercice 2

a)

Les temps de décès $T_i \sim \lambda_i \exp(-\lambda_i t_i)$ où $\lambda_i = e^{\beta x_i}$. La varisemblance du modèle est

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} \lambda_i \exp(-\lambda_i t_i) = \prod_{i=1}^{n} e^{\beta x_i} \exp(-e^{\beta x_i} t_i)$$

b)

On trouve d'abord la log-vraisemblance

$$\ell(\beta) = \sum_{i=1}^{n} \left[\beta x_i - e^{\beta x_i} t_i\right]$$

On a donc

$$\frac{\partial}{\partial \beta}\ell(\beta) = \sum_{i=1}^{n} \left[x_i - e^{\beta x_i} x_i t_i\right] = 0$$

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} e^{\hat{\beta} x_i} x_i t_i$$

Qui est l'équation qui doit être satistfaite.

c)

On calcule l'information de Fisher de l'échantillion:

$$\begin{aligned}
I(\beta) &= -\mathbb{E}_\beta \left[\frac{\partial^2}{\partial \beta^2}\ell(\beta)\right] \\
&= -\mathbb{E}_\beta \left[\frac{\partial^2}{\partial \beta^2} \sum_{i=1}^{n} \left[\beta x_i - e^{\beta x_i} t_i\right]\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} x_i^2 t_i e^{\beta x_i}\right] \\
&= \sum_{i=1}^{n} x_i^2 e^{\beta x_i} \mathbb{E}[t_i] \\
&= \sum_{i=1}^{n} x_i^2 e^{\beta x_i} e^{-\beta x_i} \\
&= \sum_{i=1}^{n} x_i^2
\end{aligned}$$

La variance asymptotique de $\hat{\beta}$ est donc

$$Var(\hat{\beta}) \xrightarrow{n \to \infty} \frac{1}{\sum_{i=1}^{n} x_i^2}$$

d)

Les hyptothèses sont

$$H_0 : \beta = 1 \qquad H_1 : \beta \neq 1$$

La statistique de Wald basé sur la normalité asymptotique de l'estimateur de vraisemblance maximale et les est:

$$W = \frac{\hat{\beta} - 1}{\sqrt{Var(\hat{\beta})}}$$

On rejette H_0 au seuil α si $|W| \geq Z_{\alpha/2}$

e)

On calcule la statistique du rapport de vraisemblance

$$\begin{aligned}
RV &= -2(\ell(\beta = 1) - \ell(\beta = \hat{\beta})) \\
&= -2\left(\sum_{i=1}^{n} [x_i - e^{x_i} t_i] - \sum_{i=1}^{n} \left[\hat{\beta} x_i - e^{\hat{\beta} x_i} t_i\right]\right) \\
&= 2\left(\sum_{i=1}^{n} \left[(\hat{\beta} - 1) x_i\right] - \sum_{i=1}^{n} \left[(e^{\hat{\beta} x_i} - e^{x_i}) t_i\right]\right)
\end{aligned}$$

On rejette l'hypothèse nulle au seuil α si $RV > \chi^2_{1;\alpha}$

Source

1. Note de cours de Alejandro Murua
2. ChatGPT:
   - Chat 1
   - Chat 2
   - Chat 3
   - Chat 4