

Power Law

Berrie, Michel, Galuppini, Laurent

2025-03-07

1. Résultats théoriques

Nous étudions le processus de Weibull (power law process dans la terminologie américaine). C'est un processus de Poisson d'intensité : $\lambda(s) = \frac{\beta}{\alpha} (\frac{s}{\alpha})^{\beta-1}$ où les paramètres α et β sont inconnus. Il modélise les instants successifs de défaillance d'un matériel dont le taux de défaillance est de type Weibull et qui subit de petites réparations. Notre travail se base essentiellement sur le livre "Processus Stochastiques et fiabilité des systèmes" par C. Coccozza-Thivent paru en 1997. De plus, nous nous plaçons dans le cas où T, le temps maximum est fixé. Le nombre d'observation est quant à lui variable.

Nous allons commencer par estimer les paramètres du processus α et β , puis nous regarderons leurs lois, ensuite nous déterminerons leurs intervalles de confiance asymptotiques et nous finirons par construire un test d'adéquation pour déterminer si un échantillon peut être modélisé par un processus de Weibull.

1. 1 Estimateurs des paramètres inconnus

Nous allons estimer les paramètres α et β par leurs estimateurs de maximum vraisemblance (MLE), $\hat{\alpha}$ et $\hat{\beta}$.

Dans un premier temps nous allons calculer $L(N, \theta)$ tel que :

On note $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$, $\hat{\theta} \in \arg \max_{\theta \in \mathbb{R}^+} L(N, \theta)$

D'après la proposition 4.14 du cours, la vraisemblance vaut dans le cas d'un processus inhomogène :

$$L((N_t)_{t \in [0, T]}; \theta) = \left(\prod_{i=1}^{N_t} \lambda(T_i) \right) \exp \left(- \int_0^T \lambda(x) dx \right)$$

On remplace λ par la fonction du Weibull ainsi :

$$L((N_t)_{t \in [0, T]}; \theta) = \left(\prod_{i=1}^{N_t} \frac{\beta}{\alpha} \left(\frac{T_i}{\alpha} \right)^{\beta-1} \right) \exp \left(\frac{T}{\alpha} \right)^{\beta}$$

puisque en effet :

$$\left(- \int_0^T \lambda(x) dx \right) = \left(- \int_0^T \frac{\beta}{\alpha} \left(\frac{x}{\alpha} \right)^{\beta-1} dx \right) = \frac{-\beta}{\alpha} \frac{1}{\alpha^{\beta-1}} \left[\frac{1}{\beta} x^{\beta} \right]_0^T = \left(\frac{T}{\alpha} \right)^{\beta}$$

Nous passons au log pour calculer la logvraisemblance (nous travaillons sur \mathbb{R}^+) afin de passer d'un produit à une somme et de nous faciliter les prochains calculs:

$$l((N_t)_{t \in [0, T]; \theta}) = \log(L((N_t)_{t \in [0, T]; \theta})) = N_t((\log(\beta) - \beta \log(\alpha)) + (\beta - 1) \sum_{i=1}^{N_t} \log(T_i) - (\frac{t}{\alpha})^\beta$$

Ensuite nous calculons le gradient de la logvraisemblance pour trouver les estimateurs qui maximisent la log vraisemblance.

$$\nabla l = \begin{bmatrix} \frac{\partial l}{\partial \alpha} \\ \frac{\partial l}{\partial \beta} \end{bmatrix} = \begin{bmatrix} -\frac{\beta}{\alpha} N_t + \beta \frac{1}{\alpha} (\frac{T}{\alpha})^\beta \\ N_t((\frac{1}{\beta} - \log(\alpha)) + \sum_{i=1}^{N_t} \log(T_i) - \log(\frac{T}{\alpha}) \exp(\beta \log(\frac{T}{\alpha}))) \end{bmatrix}$$

Dans notre cas, on cherche à trouver le point où le gradient s'annule (présence de maximum ou de minimum) ainsi $\nabla l = 0$ soit :

$$\begin{bmatrix} \log(\frac{\beta}{\alpha} (\frac{T}{\alpha})^\beta) = \log(\frac{\beta}{\alpha} N_t) \quad (0) \\ \frac{1}{\beta} = \log(\alpha) - \frac{1}{N_t} \sum_{i=1}^{N_t} \log(T_i) + \frac{1}{N_t} \log(\frac{T}{\alpha}) \exp(\beta \log(\frac{T}{\alpha})) \end{bmatrix}$$

or d'après (0) on peut en déduire que :

$$\begin{aligned} \log(\beta) - \log(\alpha) + \beta \log(\frac{T}{\alpha}) &= \log(\beta) - \log(\alpha) + \log(N_t) \\ \beta \log(\frac{T}{\alpha}) &= \log(N_t) \end{aligned}$$

ainsi :

$$\begin{bmatrix} \log(\alpha) = \log(T) - \frac{1}{\beta} \log(N_t) \\ \frac{1}{\beta} = \log(\alpha) - \frac{1}{N_t} \sum_{i=1}^{N_t} \log(T_i) + \frac{1}{N_t} \log(\frac{T}{\alpha}) \exp(\log(N_t)) \end{bmatrix}$$

et finalement :

$$\begin{bmatrix} \log(\alpha) = \log(T) - \frac{1}{\beta} \log(N_t) \\ \frac{1}{\beta} = \log(T) - \frac{1}{N_t} \sum_{i=1}^{N_t} \log(T_i) \end{bmatrix}$$

De plus, pour vérifier que nous sommes bien sur un maximum (vérification de la concavité) nous allons analyser le signe de la Hessienne localement en ce point en partant de ∇l calculé plus haut :

Premièrement calculons les dérivées partielles de notre gradient (donc les dérivées secondes) :

$$\begin{bmatrix} \frac{\partial^2 l}{\partial \alpha^2} = \frac{\beta}{\alpha^2} N_t - \frac{\beta}{\alpha^2} (\frac{T}{\alpha})^\beta - \frac{\beta^2}{\alpha^2} (\frac{T}{\alpha})^\beta \\ \frac{\partial^2 l}{\partial \alpha \partial \beta} = -\frac{N_t}{\alpha} + \frac{1}{\alpha} \exp(\beta \log(\frac{T}{\alpha})) + \frac{\beta}{\alpha} \log(\frac{T}{\alpha}) \exp(\beta \log(\frac{T}{\alpha})) \\ \frac{\partial^2 l}{\partial \beta^2} = -\frac{N_t}{\beta^2} - \log(\frac{T}{\alpha})^2 \exp(\beta \log(\frac{T}{\alpha})) \\ \frac{\partial^2 l}{\partial \beta \partial \alpha} = -\frac{N_t}{\alpha} + \frac{1}{\alpha} \exp(\beta \log(\frac{T}{\alpha})) + \frac{\beta}{\alpha} \log(\frac{T}{\alpha}) \exp(\beta \log(\frac{T}{\alpha})) \end{bmatrix}$$

Or localement on a pour rappel :

$$\frac{\partial l}{\partial \alpha} = 0$$

donc

$$-\frac{\beta}{\alpha} N_t + \beta \frac{1}{\alpha} (\frac{T}{\alpha})^\beta = 0$$

ainsi

$$(\frac{T}{\alpha})^\beta = N_t$$

d'où :

$$\exp(\beta \log(\frac{T}{\alpha})) = (\frac{T}{\alpha})^\beta = N_t$$

Ce qui nous permet de dire que localement, (sachant $N_t > 0$) :

$$\begin{aligned}\frac{\partial^2 l}{\partial \alpha^2} &= \frac{\beta}{\alpha^2} N_t - \frac{\beta}{\alpha^2} N_t - \frac{\beta^2}{\alpha^2} N_t = -\frac{\beta^2}{\alpha^2} N_t < 0 \\ \frac{\partial^2 l}{\partial \beta^2} &= -\frac{N_t}{\beta^2} - N_t \log(\frac{T}{\alpha})^2 < 0\end{aligned}$$

On remarque que $\frac{\partial^2 l}{\partial \alpha^2}$ et $\frac{\partial^2 l}{\partial \beta^2}$ sont toutes deux négatives or pour rappel :

$$\text{Tr}(H) = \frac{\partial^2 f}{\partial \alpha^2} + \frac{\partial^2 f}{\partial \beta^2}$$

ainsi $\text{Tr}(H(\hat{\alpha}, \hat{\beta})) < 0$ ce qui signifie que la somme des valeurs propres de $H(\hat{\alpha}, \hat{\beta})$ est négative.

De plus on a pour rappel :

$$\det(H) = \frac{\partial^2 f}{\partial \alpha^2} \cdot \frac{\partial^2 f}{\partial \beta^2} - \left(\frac{\partial^2 f}{\partial \alpha \partial \beta} \right)^2$$

ainsi :

$$\det(H(\hat{\alpha}, \hat{\beta})) = \frac{\partial^2 f}{\partial \alpha^2} \Big|_{\alpha=\hat{\alpha}, \beta=\hat{\beta}} \cdot \frac{\partial^2 f}{\partial \beta^2} \Big|_{\alpha=\hat{\alpha}, \beta=\hat{\beta}} - \left(\frac{\partial^2 f}{\partial \alpha \partial \beta} \Big|_{\alpha=\hat{\alpha}, \beta=\hat{\beta}} \right)^2 = \frac{1}{\alpha^2} N_t^2 + \frac{\beta^2}{\alpha^2} N_t^2 \log(\frac{T}{\alpha})^2 - \frac{\beta^2}{\alpha^2} N_t^2 \log(\frac{T}{\alpha})^2 = \frac{N_t^2}{\alpha^2} > 0$$

Ainsi $\det(H(\hat{\alpha}, \hat{\beta})) > 0$ ce qui signifie que le produit des valeurs propres de $H(\hat{\alpha}, \hat{\beta})$ est positif.

Nous pouvons donc en déduire que les valeurs propres de $H(\hat{\alpha}, \hat{\beta})$ sont toutes négatives ce qui signifie que localement nous nous trouvons sur un maximum.

Finalement on a bien $\hat{\theta} \in \arg \max_{\theta \in \mathbb{R}^+} L(N, \theta)$ et nos estimateurs sont bien des estimateurs de maximum vraisemblance.

Ces estimateurs vont nous permettre de travailler ensuite sur la construction d'intervalle de confiance de nos paramètres inconnus.

C'est ce que nous allons faire ci-dessous.

1. 2 Loi de l'estimateur

Lemme Soit $X \sim \Gamma(n, \beta)$. Alors la variable

$$Y = \frac{2}{\beta} X$$

suit la loi $\Gamma(n, 2)$, c'est-à-dire la loi du χ^2 à $2n$ degrés de liberté.

Preuve Pour montrer que $Y = \frac{2}{\beta} X$ suit $\Gamma(n, 2)$, nous allons étudier l'espérance d'une fonction de Y puis utiliser un changement de variable dans l'intégrale.

Rappel de la densité de X .

Comme $X \sim \Gamma(n, \beta)$, pour toute fonction mesurable ϕ telle que $\mathbb{E}[|\phi(Y)|] < \infty$, on écrit :

$$\mathbb{E}[\phi(Y)] = \int_{-\infty}^{\infty} \phi\left(\frac{2}{\beta}x\right) f_X(x) dx.$$

Or

$$f_X(x) = \frac{1}{\Gamma(n) \beta^n} x^{n-1} \exp\left(-\frac{x}{\beta}\right), \quad x > 0.$$

Changement de variable

Posons

$$y = \frac{2}{\beta} x \implies x = \frac{\beta}{2} y, \quad dx = \frac{\beta}{2} dy.$$

Alors,

$$\mathbb{E}[\phi(Y)] = \int_0^{\infty} \phi\left(\frac{2}{\beta}x\right) \frac{1}{\Gamma(n) \beta^n} x^{n-1} \exp\left(-\frac{x}{\beta}\right) dx.$$

En substituant $x = \frac{\beta}{2} y$ et $dx = \frac{\beta}{2} dy$, nous obtenons :

$$\begin{aligned} \mathbb{E}[\phi(Y)] &= \int_0^{\infty} \phi(y) \frac{1}{\Gamma(n) \beta^n} \left(\frac{\beta}{2} y\right)^{n-1} \exp\left(-\frac{\beta}{2} y\right) \frac{\beta}{2} dy \\ &= \int_0^{\infty} \phi(y) \frac{1}{\Gamma(n) \beta^n} \left(\frac{\beta}{2}\right)^{n-1} y^{n-1} \exp\left(-\frac{y}{2}\right) \frac{\beta}{2} dy. \end{aligned}$$

Factorisons les termes constants :

$$\frac{1}{\Gamma(n) \beta^n} \left(\frac{\beta}{2}\right)^{n-1} \frac{\beta}{2} = \frac{1}{\Gamma(n) \beta^n} \frac{\beta^n}{2^n} = \frac{1}{2^n \Gamma(n)}.$$

Ainsi,

$$\mathbb{E}[\phi(Y)] = \int_0^{\infty} \phi(y) \frac{1}{2^n \Gamma(n)} y^{n-1} \exp\left(-\frac{y}{2}\right) dy.$$

Interprétation

L'expression

$$\frac{1}{2^n \Gamma(n)} y^{n-1} \exp\left(-\frac{y}{2}\right), \quad y > 0$$

est exactement la densité d'une loi $\Gamma(n, 2)$, laquelle coïncide aussi avec une loi χ_{2n}^2 .

Par conséquent, pour toute fonction ϕ , l'espérance de $\phi(Y)$ coïncide avec celle que l'on obtiendrait si Y avait la densité $\Gamma(n, 2)$. L'unicité de la loi impliquée par cette égalité d'intégrales pour toutes les fonctions test ϕ permet de conclure que

$$Y = \frac{2}{\beta} X \sim \Gamma(n, 2).$$

Puisque $\Gamma(n, 2)$ est une χ^2 à $2n$ degrés de liberté, on peut également écrire :

$$\frac{2}{\beta} X \sim \chi_{2n}^2.$$

Proposition Dans le cas de l'observation d'un processus de Weibull sur $[0, t]$, on a

$$Z_n = \frac{\beta}{\hat{\beta}} \mid \{N(t) = n\} \sim \Gamma\left(n, \frac{1}{n}\right) \quad (n \neq 0),$$

et par conséquent,

$$\frac{2n\beta}{\hat{\beta}} \mid \{N(t) = n\} \sim \chi^2(2n).$$

Preuve Sachant $\{N(t) = n\}$, on a :

$$Z_n = \frac{\beta}{\hat{\beta}} = \beta \ln(t) - \frac{\beta}{n} \sum_{i=1}^n \ln(T_i).$$

On a également que

$$(T_1, \dots, T_n) \stackrel{\text{loi}}{=} (X_{(1)}, \dots, X_{(n)}),$$

où $(X_{(1)}, \dots, X_{(n)})$ désigne l'ordre statistique d'un échantillon i.i.d. (X_1, \dots, X_n) de densité

$$f(x) = \frac{\lambda(x)}{\int_0^t \lambda(s) ds} \mathbf{1}_{\{0 < x < t\}},$$

avec

$$\lambda(x) = \frac{\beta}{\alpha^\beta} x^{\beta-1},$$

et

$$\int_0^t \lambda(s) ds = \int_0^t \frac{\beta}{\alpha^\beta} s^{\beta-1} ds = \frac{\beta}{\alpha^\beta} \frac{t^\beta}{\beta} = \frac{t^\beta}{\alpha^\beta}.$$

Donc,

$$f(x) = \frac{\frac{\beta}{\alpha^\beta} x^{\beta-1}}{\frac{t^\beta}{\alpha^\beta}} \mathbf{1}_{\{0 < x < t\}} = \frac{\beta x^{\beta-1}}{t^\beta} \mathbf{1}_{\{0 < x < t\}}.$$

Par ailleurs, puisque (T_1, \dots, T_n) et (X_1, \dots, X_n) coïncident en loi (à réordonnancement près), on a que :

$$\sum_{i=1}^n \ln(T_i) = \sum_{i=1}^n \ln(X_i) \quad (\text{en loi}).$$

On considère la fonction génératrice des moments (FGM) de Z_n :

$$\mathbb{E}[e^{u Z_n}] = \mathbb{E}\left[\exp\left(u \left[\beta \ln(t) - \frac{\beta}{n} \sum_{i=1}^n \ln(T_i)\right]\right)\right].$$

En factorisant $\exp\{u \beta \ln(t)\} = t^{\beta u}$, on obtient :

$$\mathbb{E}[e^{u Z_n}] = t^{\beta u} \times \mathbb{E}\left[\exp\left(-\frac{\beta u}{n} \sum_{i=1}^n \ln(T_i)\right)\right].$$

En substituant $\sum_{i=1}^n \ln(T_i)$ par $\sum_{i=1}^n \ln(X_i)$ (même loi) et en utilisant l'indépendance des X_i , nous obtenons :

$$\begin{aligned}
\mathbb{E}[e^{u Z_n}] &= t^{\beta u} \times \mathbb{E}\left[\exp\left(-\frac{\beta u}{n} \sum_{i=1}^n \ln(X_i)\right)\right] \\
&= t^{\beta u} \times \mathbb{E}\left[\underbrace{\prod_{i=1}^n \exp\left(-\frac{\beta u}{n} \ln(X_i)\right)}_{X_i^{-\frac{\beta u}{n}}}\right] \\
&= t^{\beta u} \times \mathbb{E}\left[\prod_{i=1}^n X_i^{-\frac{\beta u}{n}}\right] \\
&= t^{\beta u} \times \left(\mathbb{E}[X_1^{-\frac{\beta u}{n}}]\right)^n \quad (\text{car les } X_i \text{ sont i.i.d.}).
\end{aligned}$$

On calcule $\mathbb{E}[X_1^{-\frac{\beta u}{n}}]$:

$$\begin{aligned}
\mathbb{E}[X_1^{-\frac{\beta u}{n}}] &= \int_{\mathbb{R}} x^{-\frac{\beta u}{n}} f(x) dx \\
&= \int_{\mathbb{R}} x^{-\frac{\beta u}{n}} \frac{\beta x^{\beta-1}}{t^{\beta}} \mathbf{1}_{\{0 < x < t\}} dx \\
&= \frac{\beta}{t^{\beta}} \int_0^t x^{\beta-1-\frac{\beta u}{n}} dx \\
&= \frac{\beta}{t^{\beta}} \left[\frac{x^{\beta-\frac{\beta u}{n}}}{\beta-\frac{\beta u}{n}} \right]_0^t \\
&= \frac{\beta}{t^{\beta}} \frac{t^{\beta-\frac{\beta u}{n}}}{\beta-\frac{\beta u}{n}} \\
&= \frac{\beta t^{-\frac{\beta u}{n}}}{\beta(1-\frac{u}{n})} \\
&= \frac{t^{-\frac{\beta u}{n}}}{1-\frac{u}{n}}.
\end{aligned}$$

On injecte l'expression obtenue pour $\mathbb{E}[X_1^{-\frac{\beta u}{n}}]$ dans l'expression de $\mathbb{E}[e^{u Z_n}]$, ce qui donne :

$$\begin{aligned}
\mathbb{E}[e^{u Z_n}] &= t^{\beta u} \times \left(\mathbb{E}[X_1^{-\frac{\beta u}{n}}]\right)^n \\
&= t^{\beta u} \times \left(\frac{t^{-\frac{\beta u}{n}}}{1-\frac{u}{n}}\right)^n \\
&= t^{\beta u} \times \frac{t^{-\beta u}}{\left(1-\frac{u}{n}\right)^n} \\
&= \frac{1}{\left(1-\frac{u}{n}\right)^n}.
\end{aligned}$$

On reconnaît que cette expression est la fonction génératrice des moments d'une loi $\Gamma\left(n, \frac{1}{n}\right)$. Par conséquent, on a

$$Z_n = \frac{\beta}{\hat{\beta}} \sim \Gamma\left(n, \frac{1}{n}\right).$$

D'après le lemme précédent :

$$X \sim \Gamma(n, \beta) \Rightarrow \frac{2X}{\beta} \sim \chi_{2n}^2,$$

donc dans notre cas :

$$Z_n = \frac{\beta}{\hat{\beta}} \sim \Gamma\left(n, \frac{1}{n}\right) \Rightarrow \frac{2Z_n}{1/n} = 2n Z_n \sim \chi_{2n}^2.$$

Ce qui conclut la preuve.

1. 3. Intervalle de confiance Nous avons donc

$$\frac{2n\beta}{\hat{\beta}}$$

suivant une loi de khi deux à $2n$ degrés de liberté, $\chi^2(2n)$.

Construisons l'intervalle de confiance bilatéral de niveau γ sur β associé.

$$\mathbb{P}\left(x_{\frac{1-\gamma}{2}, 2n} < \frac{2n\beta}{\hat{\beta}} < x_{\frac{1+\gamma}{2}, 2n} \middle| N(t) = n\right) = \gamma$$

avec $x_{\frac{1-\gamma}{2}, 2n}$ le quantile de niveau $\frac{1-\gamma}{2}$ d'une $\chi^2(2n)$.

D'où

$$\mathbb{P}\left(\beta \in \left[\frac{\hat{\beta}}{2n} x_{\frac{1+\gamma}{2}, 2n}, \frac{\hat{\beta}}{2n} x_{\frac{1-\gamma}{2}, 2n}\right] \middle| N(t) = n\right) = \gamma$$

Ainsi un premier intervalle de confiance sur β est :

$$\left[\frac{\hat{\beta}}{2n} x_{\frac{1+\gamma}{2}, 2n}, \frac{\hat{\beta}}{2n} x_{\frac{1-\gamma}{2}, 2n}\right]$$

Intervalles de confiance asymptotiques Nous allons désormais nous intéresser au comportement asymptotique des estimateurs $\hat{\beta}$ et $\hat{\alpha}$, soit lorsque t tend vers l'infini.

Lorsque t tend vers l'infini, nous admettons que les variables aléatoires suivantes convergent en loi vers une loi normale centrée réduite, $\mathcal{N}(0, 1)$:

Pour α :

$$\frac{\sqrt{\Lambda(t)}}{\log(t)} \log\left(\frac{\hat{\alpha}}{\alpha}\right) \quad (a)$$

$$\frac{\sqrt{\Lambda(t)}}{\log(t)} \left(\frac{\hat{\alpha}}{\alpha} - 1\right) \quad (b)$$

Pour β :

$$\sqrt{\Lambda(t)} \left(\frac{\hat{\beta}}{\beta} - 1\right) \quad (c)$$

$$\sqrt{\Lambda(t)} \left(\frac{\beta}{\hat{\beta}} - 1\right) \quad (d)$$

Nous admettons également , d'après la référence Coccozza_Thivent proposition 2.20, lorsque t tend vers $+\infty$:

$$\left(\frac{N(t)}{\Lambda(t)} \right) \xrightarrow{\text{p.s.}} 1$$

La convergence presque sure entraine la convergence en probabilité. De plus sachant que la fonction $f : x \rightarrow \sqrt{x}$ est continue sur son ensemble de définition.

Nous pouvons écrire:

$$\left(\frac{\sqrt{N(t)}}{\sqrt{\Lambda(t)}} \right) \xrightarrow{\text{P}} \sqrt{1} = 1$$

Nous appliquons cela sur (a) :

$$\frac{\sqrt{N(t)}}{\log(t)} \log \left(\frac{\hat{\alpha}}{\alpha} \right) = \underbrace{\left(\frac{\sqrt{N(t)}}{\sqrt{\Lambda(t)}} \right)}_{\xrightarrow{\text{P}} 1} \underbrace{\frac{\sqrt{\Lambda(t)}}{\log(t)} \log \left(\frac{\hat{\alpha}}{\alpha} \right)}_{\xrightarrow{\text{L}} \mathcal{N}(0,1)}$$

D'après le lemme de Slutsky nous avons que:

$$\frac{\sqrt{N(t)}}{\log(t)} \log \left(\frac{\hat{\alpha}}{\alpha} \right) \xrightarrow{\text{L}} \mathcal{N}(0,1)$$

Ainsi, nous obtenons que (1) tend vers une normale centrée réduite.

Nous pouvons faire la même chose sur (b), (c) et (d) pour obtenir que (2), (3) et (4) convergent aussi vers une normale centrée réduite quand t tend vers l'infini.

En résumé, les statistiques pivotales suivantes convergent vers une normale centrée réduite quand t tend vers l'infini :

Pour α :

$$\frac{\sqrt{N(t)}}{\log(t)} \log \left(\frac{\hat{\alpha}}{\alpha} \right) \quad (1)$$

$$\frac{\sqrt{N(t)}}{\log(t)} \left(\frac{\hat{\alpha}}{\alpha} - 1 \right) \quad (2)$$

pour β :

$$\sqrt{N(t)} \left(\frac{\hat{\beta}}{\beta} - 1 \right) \quad (3)$$

$$\sqrt{N(t)} \left(\frac{\beta}{\hat{\beta}} - 1 \right) \quad (4)$$

Construisons les intervalles de confiance asymptotiques bilatéraux associés.

Soit n_γ le quantile d'ordre γ d'une $\mathcal{N}(0,1)$,

pour α :

En reprenant (1), nous avons,

$$\mathbb{P} \left(n_{\frac{1-\gamma}{2}} < \frac{\sqrt{N(t)}}{\log(t)} \log \left(\frac{\hat{\alpha}}{\alpha} \right) < n_{\frac{1+\gamma}{2}} \right) \xrightarrow{t \rightarrow +\infty} \gamma$$

ce qui donne après calcul :

$$\mathbb{P}\left(\hat{\alpha} \exp\left(\frac{-\log(t)}{\sqrt{N(t)}} n^{\frac{1+\gamma}{2}}\right) < \alpha < \hat{\alpha} \exp\left(\frac{-\log(t)}{\sqrt{N(t)}} n^{\frac{1+\gamma}{2}}\right)\right) \xrightarrow{t \rightarrow +\infty} \gamma$$

Ainsi

$$IC_{\gamma}(\alpha) = \left[\hat{\alpha} \exp\left(\frac{-\log(t)}{\sqrt{N(t)}} n^{\frac{1+\gamma}{2}}\right), \hat{\alpha} \exp\left(\frac{-\log(t)}{\sqrt{N(t)}} n^{\frac{1+\gamma}{2}}\right) \right]$$

est un intervalle de confiance bilatéral asymptotique pour α .

Pour β :

En reprenant (3), nous avons

$$\mathbb{P}\left(n_{(1-\gamma)/2} < \sqrt{N(t)}\left(\frac{\hat{\beta}}{\beta} - 1\right) < n_{(1-\gamma)/2}\right) \xrightarrow{t \rightarrow +\infty} \gamma$$

Ainsi lorsque t tend vers $+\infty$

$$IC_{\gamma}(\beta) = \left[\hat{\beta}\left(1 - \frac{n_{(1-\gamma)/2}}{\sqrt{N(t)}}\right), \hat{\beta}\left(1 + \frac{n_{(1-\gamma)/2}}{\sqrt{N(t)}}\right) \right]$$

1.4 Test d'adéquation Le but de cette partie est de construire un test permettant de vérifier si un échantillon est issu d'un processus de Weibull.

Nous allons nous appuyer sur le théorème suivant, tiré de la référence Moller 1976:

Théorème Pour tout $n > 2$, le vecteur stochastique

$$(-\ln(W_{n-1}/W_n), -\ln(W_{n-2}/W_n), \dots, -\ln(W_1/W_n))$$

suit la même loi que $n-1$ observations ordonnées indépendantes d'une distribution exponentielle de paramètre γ^{-1} si et seulement si les événements enregistrés proviennent d'un processus Rasch-Weibull avec un paramètre de tendance γ .

Construction du test Soit un échantillon d'observations de temps d'arrivée indépendants W_1, W_2, \dots, W_n .

Appliquons la transformation suivante aux W_i et notons Z_i les variables transformées:

$$Z_i = -\ln\left(\frac{W_i}{W_n}\right), \quad i = 1, 2, \dots, n-1.$$

Nous ordonnons les Z_i et par construction nous pouvons supposer que les Z_i sont indépendants.

Dans notre cas, nous souhaitons tester si notre échantillon provient d'une loi exponentielle paramètre γ^{-1} , mais le paramètre de cette distribution n'est pas connu et doit être estimé directement à partir des données observées.

Le test classique de Kolmogorov-Smirnov repose sur l'hypothèse d'une distribution complètement spécifiée, où tous les paramètres sont connus préalablement. Or, comme indiqué dans la référence Lilliefors (1969), l'estimation du paramètre à partir de l'échantillon testé affecte significativement la distribution théorique de la statistique de test, rendant les seuils critiques standards non valides.

Pour pallier ce problème, nous optons pour le test Lilliefors-corrected Kolmogorov-Smirnov, qui détermine la statistique de test avec une méthode de Monte Carlo. Cette correction garantit une meilleure validité statistique, en maintenant des niveaux de significativité corrects lorsque le paramètre de la loi exponentielle est estimé.

En supposant qu'un temps d'arrivée n'impacte pas les autres, l'échantillon W_1, W_2, \dots, W_n est iid et de fonction de répartition continue.

Hypothèses

$$H_0 : Z_1, \dots, Z_{n-1} \stackrel{\text{i.i.d.}}{\sim} \text{Exp}(\gamma^{-1}) \quad \gamma \text{ inconnu}$$

$$H_1 : \exists i \in \{1, \dots, n-1\} \quad Z_i \not\sim \text{Exp}(\gamma^{-1})$$

D'après la référence (Lilliefors, 1969) le test Lilliefors-corrected Kolmogorov-Smirnov estime le paramètre γ par l'estimateur du maximum de vraisemblance:

$$\gamma_{\text{MLE}} = \frac{n-1}{\sum_{i=1}^{n-1} Z_i} = \frac{n-1}{\sum_{i=1}^{n-1} -\ln\left(\frac{W_i}{W_n}\right)}$$

De plus, notons \hat{F}_n la fonction de répartition empirique de l'échantillon Z_1, Z_2, \dots, Z_n que nous observons.

Pour tout $x \in \mathbb{R}$, elle est définie par :

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i \leq x\}}$$

$F_0(x)$ est la fonction de répartition théorique de la loi exponentielle de paramètre inconnu γ^{-1} .

Soit

$$F_0(x) = \begin{cases} 0, & \text{si } x < 0 \\ 1 - e^{-\gamma^{-1}x}, & \text{si } x \geq 0 \end{cases}$$

La statistique du test de Lilliefors-corrected Kolmogorov-Smirnov est définie par :

$$D_n = \sup_{x \in \mathbb{R}} \left| F_{\gamma_{\text{MLE}}}(x) - \hat{F}_n(x) \right|.$$

Sous l'hypothèse H_0 , D_n est indépendante de $F_{\gamma_{\text{MLE}}}$ et est tabulée. On rejettera H_0 dès lors que l'écart entre $F_{\gamma_{\text{MLE}}}(x)$ et $F_n(x)$ est significativement grand. Nous obtenons ainsi la zone de rejet suivante:

$$R_\alpha = \{D_n \geq d_{1-\alpha}\}$$

où $d_{1-\alpha}$ est le $1-\alpha$ quantile de D_n

Dans le cas où nous acceptons H_0 , d'après le théorème nous savons que les événements de notre échantillon suivent une répartition de Weibull de tendance γ .

2. Résultats numériques

Simulation d'un processus de poisson inhomogène par méthode de changement de temps :

```

simulPPi <- function(Tmax, alpha, beta) {
  Lambda<-function(x){(x/alpha)**beta}
  invLambda<-function(x){alpha*x**(1/beta)}
  S=0      # Temps actuel
  t=invLambda(0)
  #c'est le temps à chaque n
  events=c()
  # Boucle while pour générer les temps d'arrivée
  while (invLambda(t) <= Tmax) {
    t=t+rexp(1,rate=1)
    events<-c(events,invLambda(t))
  }
  return (events[1:length(events)-1])
}

```

Fonction pour afficher nos simulations :

```

plot.PP<- function(PP)
{
  # plot the counting process (with jumps of size 1 (starting at point (0,0))):
  plot(c(0,PP),0:length(PP),type="s",xlab="time t",ylab="Number of events by time t")
  # add the arrival times on the horizontal axis:
  points(PP,0*PP,type="p",pch=16, col='red')
  # link arrival times with the counts:
  lines(PP,0:(length(PP)-1),type="h",lty=2, col='blue')
}

```

Testons la simulation pour les paramètres suivant :

```

Tmax=100
alphatest = 3
betatest= 1

```

```

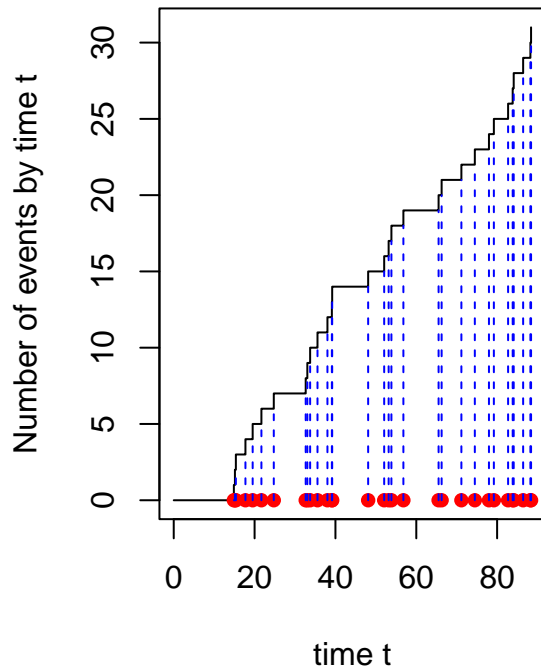
PPitest = simulPPi(Tmax,alphatest,betatest)
par(mfrow=c(1,2))
#curve(lambda_fct(alpha,beta),from=0,to=Tmax,n=1000)
plot.PP(PPitest)
print(PPitest)

```

```

## [1] 14.87856 15.14069 15.36917 17.73632 19.52497 21.69008 24.74911 32.64526
## [9] 33.05015 33.74394 35.56310 38.02370 39.11731 39.19700 48.09767 52.06038
## [17] 53.16022 53.84623 56.80701 65.51730 66.27193 71.15882 74.48465 77.98052
## [25] 79.19341 82.71514 83.84700 84.10113 86.43399 88.23203 88.39423

```



Bien, maintenant que nous savons comment simuler un PPI, estimons les paramètres β et α et étudions les intervalles de confiance asymptotiques.

Estimateurs de alpha et beta Fonction pour estimer beta et alpha :

```
MLE<-function(PPi) {
  N_t=length(PPi)
  Tm=PPi[N_t]
  beta_hat=1/(log(Tm)-1/N_t*sum(log(PPi)))
  alpha_hat=Tm*exp(-1/beta_hat*log(N_t))
  return(c(alpha_hat,beta_hat))
}
```

Fonctions pour tester si α_{MLE} et β_{MLE} appartiennent aux intervalles de confiance

$$\left[\hat{\alpha} \exp\left(\frac{-\log(t)}{\sqrt{N(t)}} n^{\frac{1+\gamma}{2}}\right), \hat{\alpha} \exp\left(\frac{-\log(t)}{\sqrt{N(t)}} n^{\frac{1+\gamma}{2}}\right) \right] \times \left[\hat{\beta} \left(1 - \frac{n^{(1-\gamma)/2}}{\sqrt{N(t)}}\right), \hat{\beta} \left(1 + \frac{n^{(1-\gamma)/2}}{\sqrt{N(t)}}\right) \right]$$

lorsque t tend vers $+\infty$

Nous prenons $\gamma=0.05$.

```
testintervallesalpha <- function(PPi, MLE_alpha, alpha) {
  z <- qnorm(0.975)
  N=length(PPi)
  inside<-logical(1)
```

```

t_i <- PPi[N] # on prend le t correspondant à alpha
inside <- ((MLE_alpha * exp(-log(t_i) / sqrt(N) * z) < alpha) && (MLE_alpha * exp(log(t_i) / sqrt(N) * z) < alpha)
# print(paste(MLE_alpha, "appartient à l'IC", MLE_alpha * exp(-log(t_i) / sqrt(N) * z), (MLE_alpha * exp(log(t_i) / sqrt(N) * z) < alpha))
return(inside)
}

testintervallesbeta <- function(PPi, MLE_beta, beta) {
  z <- qnorm(0.975)
  inside <- logical(length(MLE_beta))
  N=length(PPi)
  t_i <- PPi[N] # on prend le t correspondant à beta
  inside <- (MLE_beta * (1 - z / sqrt(N)) < beta && MLE_beta * (1 + z / sqrt(N)) > beta)
  return(inside)
}

```

Testons la convergence de nos estimateurs lorsque t tend vers $+\infty$ avec $\alpha=1$ et $\beta=2$:

```

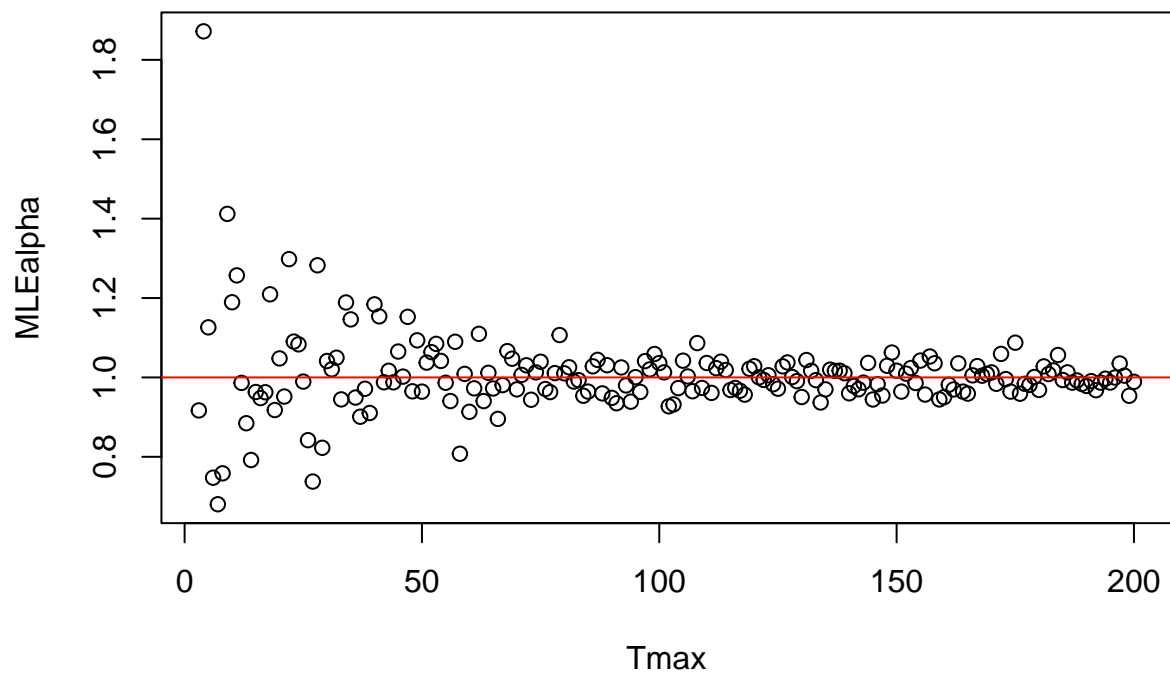
alpha=1
beta=2
Tillustr=3:200
MLEalpha=c()
MLEbeta=c()
interalpha=c()
interbeta=c()

for (Tmax in Tillustr){
  PPi=simulPPi(Tmax,alpha,beta)
  MLE_results=MLE(PPi)

  MLEalpha=c(MLEalpha,MLE_results[1])
  MLEbeta=c(MLEbeta,MLE_results[2])

  interalpha=c(interalpha,testintervallesalpha(PPi,MLE_results[1],alpha))
  interbeta=c(interbeta,testintervallesbeta(PPi,MLE_results[2],beta))
}
plot(Tillustr,MLEalpha,xlab="Tmax",ylab="MLEalpha")
abline(h=alpha,col="red",lwd=alpha)

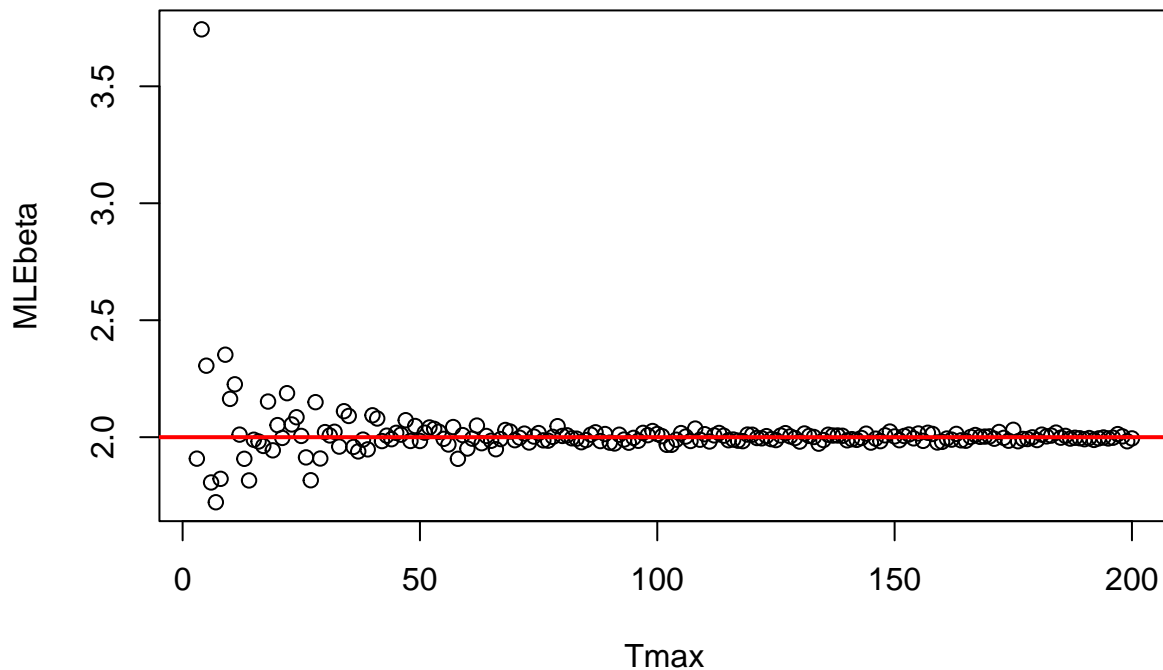
```



```
print(paste("Proportion de alpha_MLE qui appartiennent à l'IC :", mean(interalpha)))
```

```
## [1] "Proportion de alpha_MLE qui appartiennent à l'IC : 0.97979797979798"
```

```
plot(Tillustr,MLEbeta,xlab="Tmax",ylab="MLEbeta")
abline(h=beta,col="red",lwd=beta)
```



```
print(paste("Proportion de beta_MLE qui appartiennent à l'IC :", mean(interbeta)))
```

```
## [1] "Proportion de beta_MLE qui appartiennent à l'IC : 0.984848484848485"
```

Numériquement, nous confirmons bien nos résultats théoriques sur les intervalles de confiance asymptotiques. On observe une proportion d'environ 95% d'IC contenant véritablement α et β lorsque t est grand.

Afin d'être certain, testons avec d'autres valeurs de α et β . Testons la convergence de nos estimateurs lorsque t tend vers $+\infty$ avec $\alpha=0.5$ et $\beta=0.9$:

```
alpha=1
beta=2
Tillustr=3:200
MLEalpha=c()
MLEbeta=c()
interalpha=c()
interbeta=c()

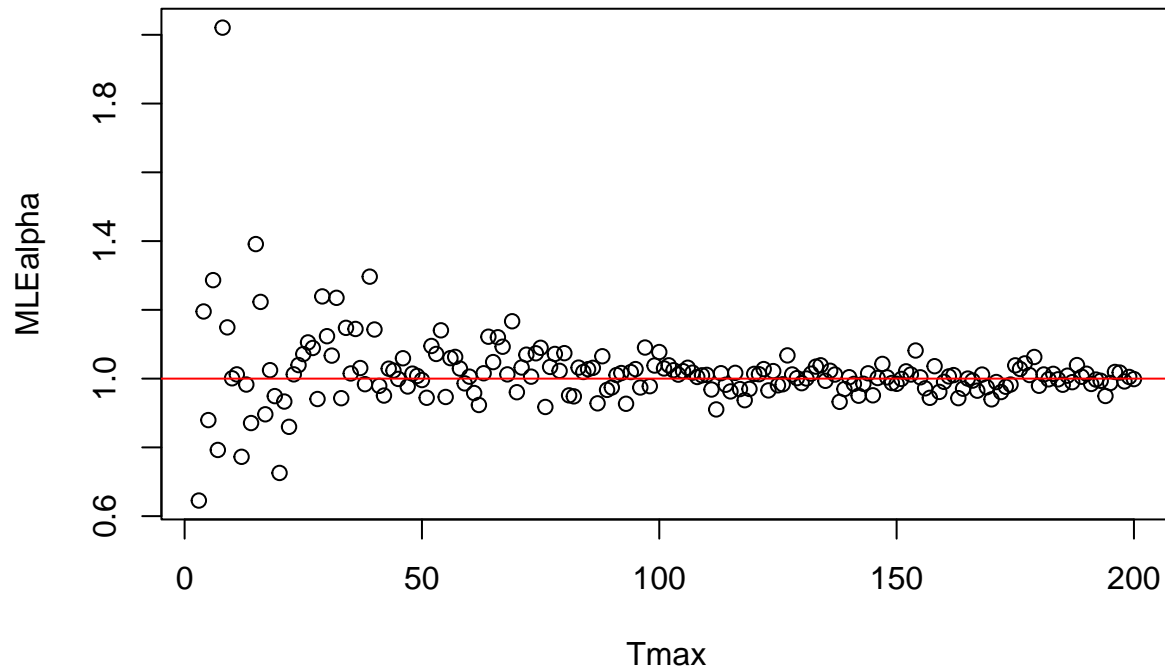
for (Tmax in Tillustr){
  PPi=simulPPi(Tmax,alpha,beta)
  MLE_results=MLE(PPi)

  MLEalpha=c(MLEalpha,MLE_results[1])
  MLEbeta=c(MLEbeta,MLE_results[2])
}
```

```

interalpha=c(interalpha,testintervallesalpha(PPi,MLE_results[1],alpha))
interbeta=c(interbeta,testintervallesbeta(PPi,MLE_results[2],beta))
}
plot(Tillustr,MLEalpha,xlab="Tmax",ylab="MLEalpha")
abline(h=alpha,col="red",lwd=alpha)

```



```

print(paste("Proportion de alpha_MLE qui appartiennent à l'IC :", mean(interalpha)))

```

```

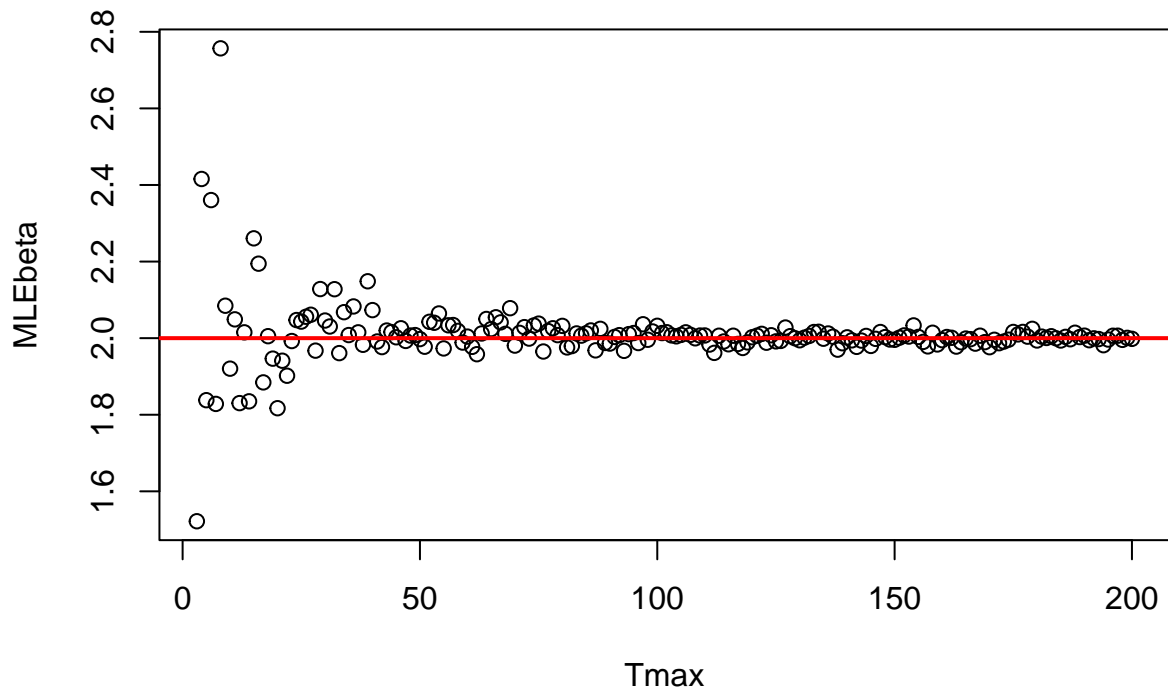
## [1] "Proportion de alpha_MLE qui appartiennent à l'IC : 0.95959595959596"

```

```

plot(Tillustr,MLEbeta,xlab="Tmax",ylab="MLEbeta")
abline(h=beta,col="red",lwd=beta)

```

```
print(paste("Proportion de beta_MLE qui appartiennent à l'IC :", mean(interbeta)))
```

```
## [1] "Proportion de beta_MLE qui appartiennent à l'IC : 0.95959595959596"
```

Test d'adéquation sur le jeu de données Boeing 720.

Dans la partie suivante, nous allons illustrer la réalisation numérique de la phase de test d'adéquation à l'aide d'un jeu de données représentant les intervalles entre les pannes de l'air conditionné dans une flotte d'avions Boeing 720.

Nous considérons donc ici notre échantillon comme représentant les instants où les pannes interviennent. Il faut légèrement modifier le jeu de données afin de ne plus avoir les intervalles entre les pannes, mais les différents temps d'occurrence des pannes.

```
library(KScorrect)
library(readxl)
df <- read_excel("data.xlsx", sheet = 1)
head(df)
```

```
## # A tibble: 6 x 13
##   B7907 B7908 B7909 B7910 B7911 B7912 B7913 B7914 B7915 B7916 B7917 B8044 B8045
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1   194   413    90    74    55    23    97    50   359    50   130   487   102
## 2   209   427   100   131   375   284   151    94   368   304   623   505   311
```

## 3	250	485	160	179	431	371	162	196	380	309	NA	605	325
## 4	279	522	346	208	535	378	166	268	650	592	NA	612	382
## 5	312	622	407	710	755	498	307	290	1253	627	NA	710	436
## 6	493	687	456	722	994	512	325	329	1256	639	NA	715	468

Pour pouvoir faire un test Lilliefors-corrected Kolmogorov-Smirnov il faut avoir un échantillon iid de loi continue.

On suppose alors que lorsqu'une panne intervient, cela n'a pas d'impact sur le temps des pannes suivantes.

Suivons le plan du test construit dans la partie théorique.

Commençons par transformer les échantillons en calculant le vecteur Z pour l'avion 7907:

```
W <- df[[1]]
W <- na.omit(W)
# Supprime les valeurs NA
W_n <- tail(W, 1)
Z <- -log(W[1:(length(W) - 1)] / W_n)
Z <- sort(Z)
print(W)
```

```
## [1] 194 209 250 279 312 493 502
## attr(,"na.action")
## [1] 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
## attr(,"class")
## [1] "omit"
```

Maintenant nous pouvons effectuer le test de Lilliefors-corrected Kolmogorov-Smirnov pour vérifier si notre échantillon est bien distribué selon une loi exponentielle.

```
LcKS_result <- LcKS(Z, "pexp")
cat("Statistique de test :", LcKS_result$D.obs, "\n")
```

```
## Statistique de test : 0.3801748
```

```
cat("P-valeur :", LcKS_result$p.value, "\n")
```

```
## P-valeur : 0.0846
```

D'après les résultats, on obtient une p-valeur d'environ 0,086, soit supérieure à 0,05. Au seuil de 5 %, nous ne rejetons donc pas l'hypothèse nulle, soit nous acceptons que l'échantillon suit une loi exponentielle. En revanche, la p-valeur est très proche de 0,05 ; cela est probablement dû à la faible taille de l'échantillon (seulement 7 observations).

D'après le théorème présenté dans la partie théorique, nous pouvons modéliser le processus de comptage des pannes de l'air conditionné de l'avion 7907 par un processus de Weibull.

Nous pouvons répéter la même opération pour les autres avions de la flotte :

```
for (i in 1:ncol(df)) {
  avion_id <- colnames(df)[i] # Nom de l'avion
```

```

W_avion <- df[[i]]

W_avion <- na.omit(W_avion)

if (length(W_avion) > 1) {
  W_n_avion <- tail(W_avion, 1)
  Z_avion <- -log(W_avion[1:(length(W_avion) - 1)] / W_n_avion)
  Z_avion <- sort(Z_avion)

  #Estimateur du paramètre
  g_avion <- mean(Z_avion)

  LcKS_result_avion <- LcKS(Z_avion, "pexp")

  cat("\n===== \n")
  cat("Résultats pour l'avion", avion_id, "\n")
  cat("----- \n")
  cat("Paramètre gamma estimé :", g_avion, "\n")
  cat("Statistique D de Kolmogorov-Smirnov :", LcKS_result_avion$statistic, "\n")
  cat("P-valeur :", LcKS_result_avion$p.value, "\n")
  cat("===== \n\n")
} else {
  cat("\n===== \n")
  cat("L'avion", avion_id, "ne contient pas assez de données exploitables.\n")
  cat("===== \n\n")
}
}

```

```

##
## =====
## Résultats pour l'avion B7907
## -----
## Paramètre gamma estimé : 0.6008705
## Statistique D de Kolmogorov-Smirnov :
## P-valeur : 0.0912
## =====
##
##
## =====
## Résultats pour l'avion B7908
## -----
## Paramètre gamma estimé : 0.6006449
## Statistique D de Kolmogorov-Smirnov :
## P-valeur : 0.0982
## =====
##
##
## =====
## Résultats pour l'avion B7909
## -----

```

```

## Paramètre gamma estimé : 1.149863
## Statistique D de Kolmogorov-Smirnov :
## P-valeur : 0.3708
## =====
##
##
## =====
## Résultats pour l'avion B7910
## -----
## Paramètre gamma estimé : 1.147749
## Statistique D de Kolmogorov-Smirnov :
## P-valeur : 0.6642
## =====
##
##
## =====
## Résultats pour l'avion B7911
## -----
## Paramètre gamma estimé : 0.8231338
## Statistique D de Kolmogorov-Smirnov :
## P-valeur : 0.4966
## =====
##
##
## =====
## Résultats pour l'avion B7912
## -----
## Paramètre gamma estimé : 0.6869462
## Statistique D de Kolmogorov-Smirnov :
## P-valeur : 0.062
## =====
##
##
## =====
## Résultats pour l'avion B7913
## -----
## Paramètre gamma estimé : 1.047523
## Statistique D de Kolmogorov-Smirnov :
## P-valeur : 0.8206
## =====
##
##
## =====
## Résultats pour l'avion B7914
## -----
## Paramètre gamma estimé : 0.9590572
## Statistique D de Kolmogorov-Smirnov :
## P-valeur : 0.5492
## =====
##
##
## =====
## Résultats pour l'avion B7915
## -----

```

```

## Paramètre gamma estimé : 0.8818567
## Statistique D de Kolmogorov-Smirnov :
## P-valeur : 0.276
## =====
##
##
## =====
## Résultats pour l'avion B7916
## -----
## Paramètre gamma estimé : 0.8225354
## Statistique D de Kolmogorov-Smirnov :
## P-valeur : 0.3902
## =====
##
##
## =====
## Résultats pour l'avion B7917
## -----
## Paramètre gamma estimé : 1.567012
## Statistique D de Kolmogorov-Smirnov :
## P-valeur : 0.0696
## =====
##
##
## =====
## Résultats pour l'avion B8044
## -----
## Paramètre gamma estimé : 0.5486662
## Statistique D de Kolmogorov-Smirnov :
## P-valeur : 0.14
## =====
##
##
## =====
## Résultats pour l'avion B8045
## -----
## Paramètre gamma estimé : 0.8259296
## Statistique D de Kolmogorov-Smirnov :
## P-valeur : 0.6924
## =====

```

De manière analogue à l'avion B7907, nous pouvons modéliser le processus de comptage de panne de l'air conditionné de tous les avions par un processus de Weibull de paramètre de tendance égal au gamma estimé.

En revanche pour les avions B8044 B7917 B7912 B7908 B7907 la p-valeur reste proche de 0.05 cela est probablement dû à la faible taille des échantillons.

Nous allons finir par déterminer les intervalles de confiance pour les paramètres de l'avion B7909 (la construction est identique pour les autres avions en adaptant le temps et le nombre de panne).

Le paramètre de tendance que nous avons estimé correspond à la quantité que nous avons notée β dans la partie théorique.

En reprenant les intervalles de confiance asymptotiques présentés dans la section 1.3, nous obtenons :

pour β :

```

hat_beta <- 1.149863
N_t <- 28
gamma <- 0.05

z <- qnorm(1 - gamma / 2)

lower_bound <- hat_beta * (1 - z / sqrt(N_t))
upper_bound <- hat_beta * (1 + z / sqrt(N_t))

IC <- c(lower_bound, upper_bound)
names(IC) <- c("Lower", "Upper")
IC

```

```

##      Lower      Upper
## 0.7239556 1.5757704

```

Pour α , il faut commencer par estimer le paramètre, puis déterminer son intervalle de confiance.

Notons que la quantité t est fixée à 2422, car il s'agit du temps d'observation des pannes pour l'avion B7909. Ce temps correspond au moment de la dernière panne observée.

```

hat_beta <- 1.149863
t <- 2422
hat_alpha = t * exp(-1/hat_beta * log(N_t))
N_t <- 28
gamma <- 0.05

q <- qnorm((1 + gamma) / 2)
lower_bound2 <- hat_alpha * exp(-log(t) / sqrt(N_t) * q)
upper_bound2 <- hat_alpha * exp(log(t) / sqrt(N_t) * q)

IC <- c(lower_bound2, upper_bound2)
names(IC) <- c("Lower", "Upper")
IC

```

```

##      Lower      Upper
## 121.7648 146.4637

```