TP1_IARD1

Ruffin Adja,Teffery Makondé

15/09/2020

Question 1

a) Simulation de 100 observations de la loi exponentielle de paramètre $\theta = 1$ et calcul du coefficient d'asymétrie

Pour simuler les observations de la loi exponentielle,nous avons utilisé la méthode de l'inverse.

En effet, la densité d'une variable aléatoire X d'une loi expontentielle de paramètre $\theta = 1$ est :

$$f_X(x) = \theta e^{-\theta x}$$

La fonction de répartition est:

$$F_X(x) = 1 - e^{-\theta x}$$

et la fonction de répartition inverse est:

$$F_X^{-1}(u) = -\frac{ln(1-u)}{\theta}$$

Si nous appliquons cette fonction de répartition inverse aux observations de la loi uniforme entre 0 et 1, nous obtenons les observations de la loi exponentielle en question.

Mais on sait que si u appartient à $[0, 1]$ alors $1 - u$ appartient aussi à $[0, 1]$. Par conséquent,nous pouvons appliquer la fonction

$$F_X^{-1}(u) = -\frac{ln(u)}{\theta}$$

aux observations de la loi uniforme entre 0 et 1 est suffisant pour simuler les observations de la loi exponentielle.

La moyenne empirique des observations de l'échantillon simulé est:

$$\bar{X} = \frac{1}{100}\sum_{i=1}^{100} x_i$$

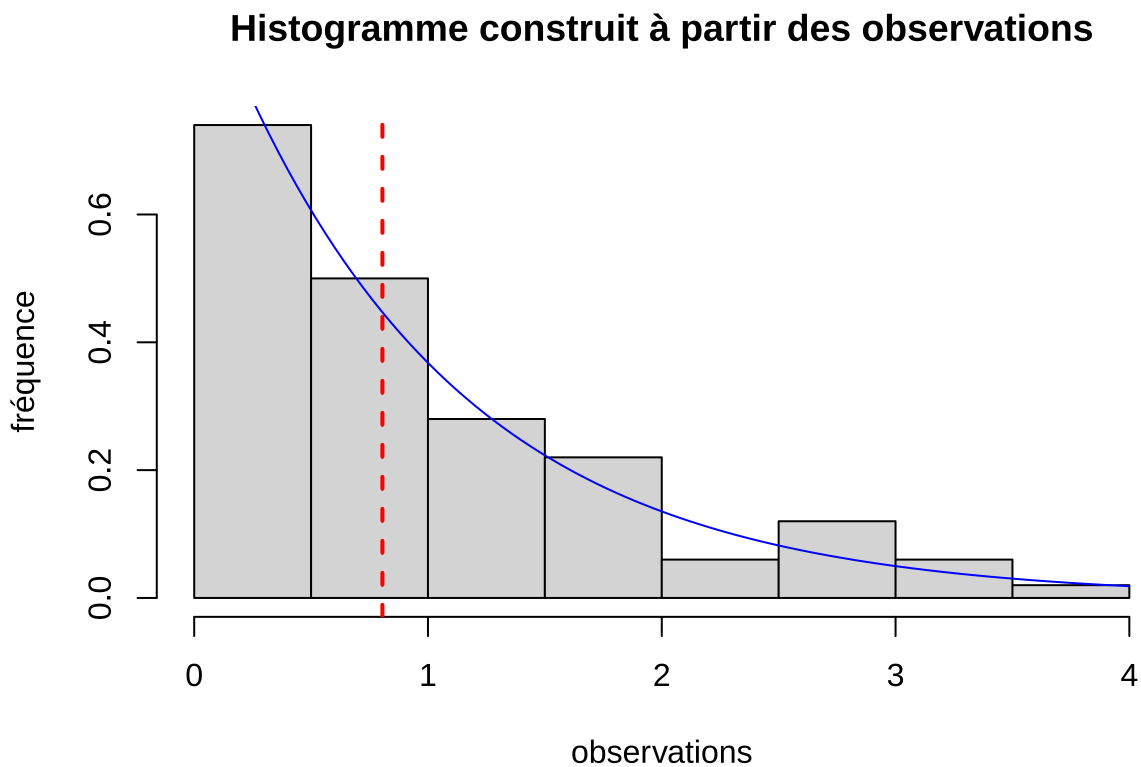
et est égale à : 0.9988561.

D'où l'estimateur du coefficient d'asymétrie est obtenu comme suit:

$$\widehat{\gamma} = \frac{(\sum_{i=1}^{100}(x_i - \bar{X})^3)/100}{((\sum_{i=1}^{100}(x_i - \bar{X})^2)/100)^{1.5}}$$

et est égal à : 1.1557096

L'histogramme construit à partir des données de l'échantillon est présenté ci-après. Le tracé bleu est la densité d'une loi exponentielle de paramètre $\theta = 1$, tandis que le tracé rouge est la médiane empririque.

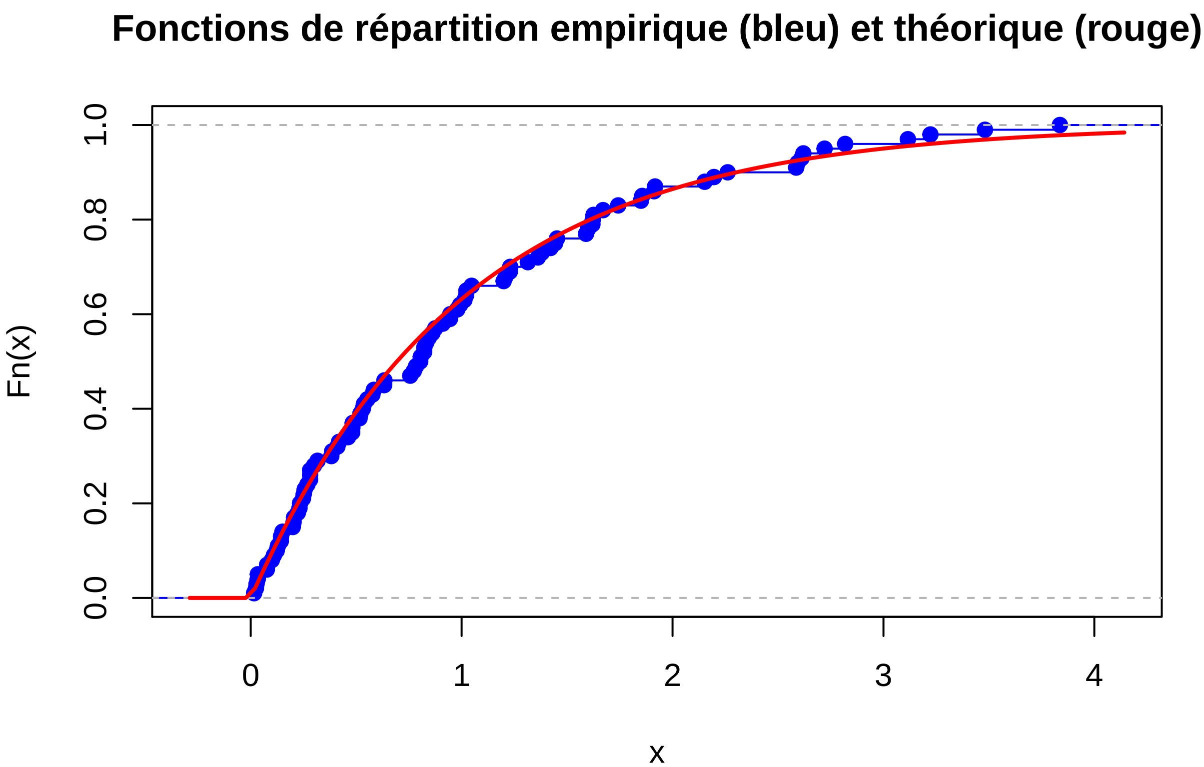


Nous pouvons faire trois constats à la suite du graphique ci-dessus:

-L'histogramme s'adapte parfaitement à la courbe de la densité théorique de la loi exponentielle.

-L'estimateur du coefficient d'asymétrie est proche de 2,ce qui est la valeur théorique du coefficient d'asymétrie de la loi expnentielle.

-Le coefficient d'asymétrie est positif donc la distribution empirique est décalée à gauche de la médiane et nous obtenons une queue de distribution étalée vers la droite.



Le tracé bleu représente la courbe de la fonction de répartition empirique de la loi exponentielle de paramètre $\theta = 1$ alors que le tracé rouge est la courbe de la fonction de répartition théorique de la même loi.

On remarque que les deux courbes sont très proches et se superposent par moment.Ainsi, on confirme que la fonction de répartition empirique calculée, $F_n(x)$ est un bon estimateur de la fonction de répartition théorique $F(x)$.

b) Calcul de la variance de l'estimateur du coefficient d'asymétrie et d'un intervalle de confiance de niveau 95%

Par la méthode du bootstrap, nous allons simuler 50 échantillons de taille 100 d'une loi exponentielle de paramètre $\theta = 1$.

La moyenne des coefficients d'asymétrie des 50 échantillons simulés est: 1.8260431 et sa variance est: 0.2876344.

Si a et b sont les bornes d'un intervalle de confiance de niveau 95% de

$$Z = \frac{\widehat{\gamma} - \mu}{\sigma}$$

qui suit une loi normale standard alors on peut écrire que:

$$P(a < Z < b) = 0.95$$

et il s'ensuit que $P(Z < a) = 0.025$ et $P(Z > b) = 0.975$. Finalement, on a

$$a = \phi^{-1}(0.025)$$

et

$$b = \phi^{-1}(0.975)$$

Et on peut tirer que

$$P(a\sigma + \mu < \widehat{\gamma} < b\sigma + \mu) = 0.95$$

Ainsi, on a $P($ 0.7748838 $< \widehat{\gamma}<$2.8772025$) = 0.95$.

L'intervalle de confiance de niveau 95% de $\widehat{\gamma}$ est : [0.7748838 , 2.8772025]

c) Calcul du coefficient d'asymétrie théorique de la loi exponentielle de paramètre $\theta = 1$

On sait que si X est une variable aléatoire de la loi exponenentielle de moyenne $\theta = 1$ et de densité

$$f_X(x) = \theta e^{-\theta x}, x > 0$$

,

Le coefficient d'asymétrie est :

$$\gamma = \frac{E((X-\mu)^3)}{\sigma^3}$$

En développant cette expression, on obtient :

$$\gamma = \frac{1}{\sigma^3}(E(X^3) - 3\mu\sigma^2 - \mu^3)$$

Calculons chacun des éléments de cette égalité.

$$E(X^3) = \int_0^\infty x^3 e^{-x} dx$$

qui est encore égal à

$$E(X^3) = \Gamma(4) \int_0^\infty \frac{x^{4-1}}{\Gamma(4)} e^{-x} dx$$

D'où

$$E(X^3) = 6$$

On sait aussi que

$$\mu = \frac{1}{\theta} = 1$$

et

$$Var(X) = \sigma^2 = \frac{1}{\theta^2} = 1$$

On peut donc calculer le coefficient d'asymétrie: $\gamma = 6 - 3 - 1 = 2$.

On peut noter que la valeur de l'estimateur du coefficient d'asymétrie est proche de sa valeur théorique. Ici, l'écart entre les deux valeurs est de 42.2145205%

Question 2

a) Estimation de $E(min(X, u))$ de façon non paramétrique

Un estimateur de $E(min(X, u))$ est $\frac{\sum_{i=1}^{100} min(x_i, u)}{100}$

Les résultats sont présentées dans le tableau ci-dessous.

u	Valeur de u	Estimateur de la perte limitée
$u = F^{-1}(0,25)$	0.2876821	0.2535318
$u = F^{-1}(0,35)$	0.4307829	0.3542954
$u = F^{-1}(0,50)$	0.6931472	0.5099046
$u = F^{-1}(0,65)$	1.0498221	0.6193727
$u = F^{-1}(0,75)$	1.3862944	0.7777432
$u = F^{-1}(0,85)$	1.89712	0.8829425

On note que, de même que u augmente avec la probabilité, l'estimateur augmente aussi avec la valeur de u. Ce résultat est logique. En effet, lorsque le déductible (ordinaire ou avec franchise)augmente, il faut que $E(min(X,u))$, la perte limitée, augmente également afin que la prime (par perte ou par paiement) diminue.

De façon théorique,

$$E(min(X,u)) = \int_0^u \theta x e^{-\theta x} dx + \int_u^\infty u\theta e^{-\theta x} dx$$

qui est égal, après calculs à $E(min(X,u)) = \frac{1}{\theta}(1 - e^{-\theta u})$.

Comme $\theta = 1$, $E(min(X,u)) = 1 - e^{-u}$

Le tableau ci-après récapitule le résultat des calculs.

u	Estimat. perte limitée	Valeur théorique perte limitée
$u = F^{-1}(0,25)$	0.2535318	0.25
$u = F^{-1}(0,35)$	0.3542954	0.35
$u = F^{-1}(0,50)$	0.5099046	0.5
$u = F^{-1}(0,65)$	0.2535318	0.6
$u = F^{-1}(0,75)$	0.2535318	0.75
$u = F^{-1}(0,85)$	0.2535318	0.85

A partir des valeurs de la perte limitée (estimées et théoriques), on peut calculer les valeurs estimées et théorique de la prime par par perte, comme suit:

$$Estimateur(prime_u) = \frac{\sum_{i=1}^{100} x_i}{100} - Estimateur(PerteLimitée_u)$$

et

$$PrimeThéorique_u = e^{-u}$$

Ces valeurs sont présentées dans le tabeau ci-dessous.

u	Estimat.prime par perte	Valeur théorique prime par perte
$u = F^{-1}(0,25)$	0.7453242	0.75
$u = F^{-1}(0,35)$	0.6445607	0.65
$u = F^{-1}(0,50)$	0.4889515	0.5
$u = F^{-1}(0,65)$	0.3794833	0.4
$u = F^{-1}(0,75)$	0.2211129	0.25
$u = F^{-1}(0,85)$	0.1159135	0.15

b) Bootstrap et intervalle de confiance pour les primes loss avec déductibles ordinaires $u = F^{-1}(0,50)$ et $u = F^{-1}(0,75)$

On sait que

$$E(Y^L) = E(X) - E(min(X,u))$$

Donc,

$$\widehat{E(Y^L)} = \widehat{E(X)} - E(\widehat{min(X},u))$$

Ce qui revient à

$$\widehat{E(Y^L)} = \frac{\sum_{m=1}^{50}\sum_{i=1}^{100} x_{i,m}}{50 * 100} - \frac{\sum_{m=1}^{50}\sum_{i=1}^{100} min(x_{i,m}, u)}{50 * 100}$$

Les primes par perte et leur variance en fonction du deductible sont:

u	Valeur de u	Estimateur de la prime par perte	Variance Estimateur
$u = F^{-1}(0,50)$	0.6931472	0.4589056	0.0485308
$u = F^{-1}(0,75)$	1.3862944	0.1948053	0.2001934

De ces valeurs,on peut tirer les bornes de l'intervalles de confiance de niveau 95% de la prime par perte selon les valeurs de u, le déductible simple.

Les primes par perte et leur variance en fonction du deductible sont:

u	Est. prime par perte	IC à 95%
$u = F^{-1}(0,50)$	0.4589056	[0.0271315, 0.8906797]
$u = F^{-1}(0,75)$	0.1948053	[-0.682141, 1.0717517]

c) Comparaisom des valeurs

u	Est. perte limitée	Val théo.perte limitée
$u = F^{-1}(0,50)$	0.2535318	0.25
$u = F^{-1}(0,75)$	0.3542954	0.35

u	Est.prime par perte	Val théo.prime par perte	IC 95%
$u = F^{-1}(0,50)$	0.4589056	0.75	[0.0271315, 0.8906797]
$u = F^{-1}(0,75)$	0.1948053	0.65	[-0.682141, 1.0717517]

On remarque que le montant du déductible ordinaire est positivement correlé à la perte limitée, mais il est négativement correlé à la prime par perte. En effet, lorsque le déductible a augmenté en passant de $F^{-1}(0,50)$ à $F^{-1}(0,75)$, la perte limitée a également augmenté, alors que la prime par perte a chuté. Ce résultat est cohérent avec celui obtenu à la question a.

Question 3

a) Estimateur de Kaplan-Meier pour les données: 30; 40; 57; 65; 65; 84*; 90; 92; 98; 101

Soient s_i le nombre de d'occurence de la i^e observation (ou encore le nombre de décès $(---, y_i])$, b_i le nombre d'observations censurées entre $[y_i, y_{i+1}]$, et r_t le nombre d'observations à risque. On a $s(y_1) = s(30) = 1$ car il y a 1 décès entre $(0, 30]$. $b(30) = 0$ car il n'y a pas de données censurées entre $[30, 40)$. Enfin, $r(y_1) = r(30) = 10$ car $r(i)$ est définie comme suit:

$$r(i) = \begin{cases} n & \text{si } i = 1 \\ r_{i-1} - s_{i-1} - b_{i-1} & \text{sinon.} \end{cases}$$

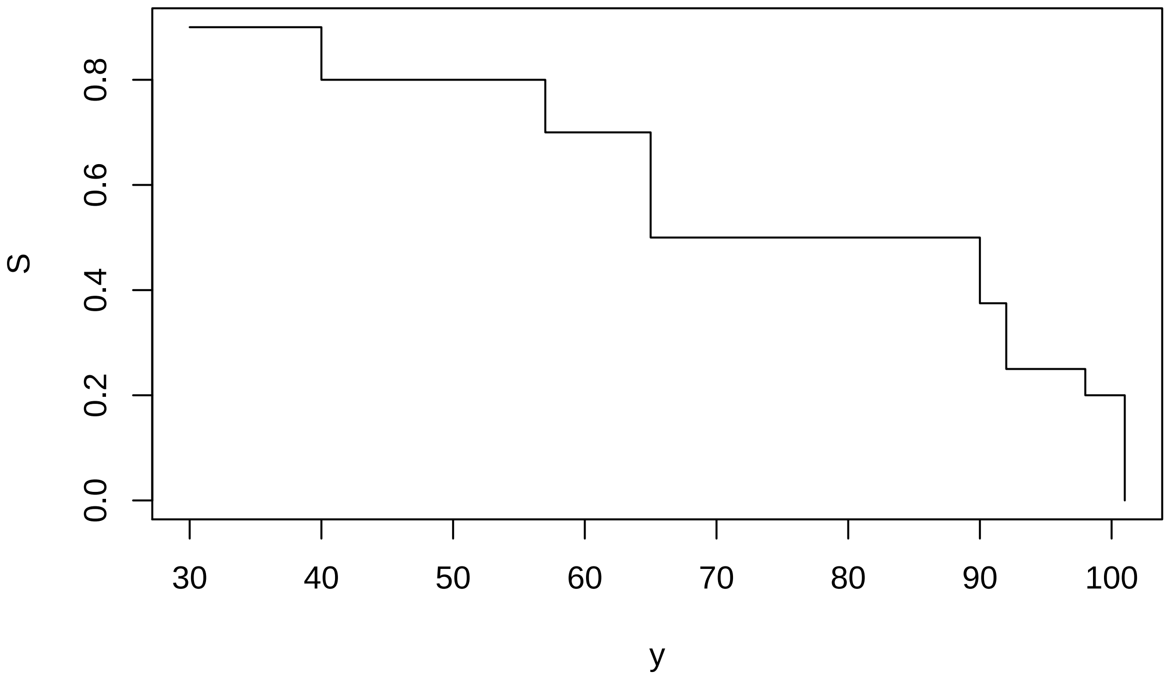
où n est le nombre d'observations. Et on a:

$$\widehat{S(y_i)} = \prod_{j<=i} (1 - \frac{s_j}{r_j})$$

Pour résumer, on obtient le tableau suivant:

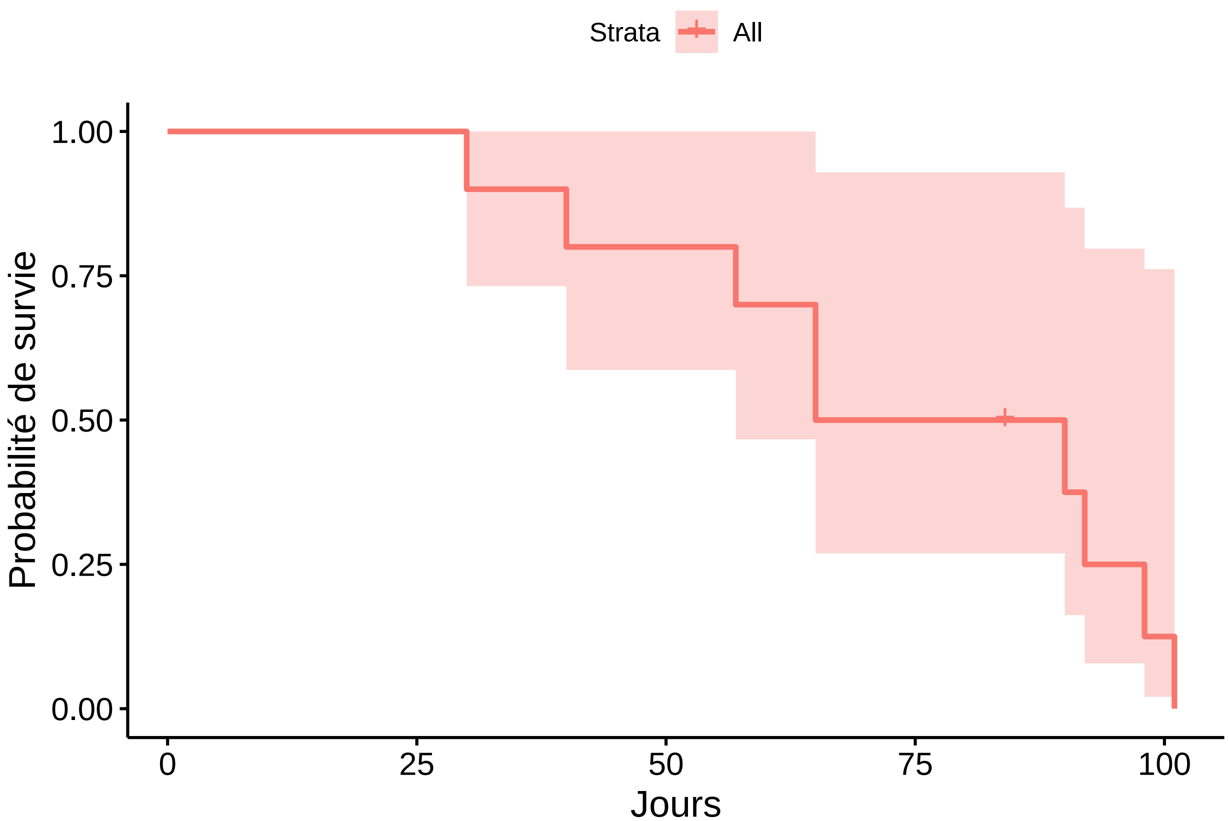
i	y_i	$s(y_i)$	$b(y_i)$	$r(y_i)$	$\widehat{S(y_i)}$
1	30	1	0	10	0.9
2	40	1	0	9	0.8
3	57	1	0	8	0.7
4	65	2	1	7	0.5
5	90	1	0	4	0.375
6	92	1	0	3	0.25
7	98	1	0	2	0.2
8	101	1	0	1	0

L'estimateur de Kaplan-Meier $\widehat{S}(91) = 0.375$



#

Deuxième méthode pour la question a



L'estimateur de Kaplan-Meier 91 est égale à 0.375.

b) Utilisons la formule de Greenwood pour calculer un intervalle de confiance au niveau 0.95 pour $S(91)$

La variance de l'estimateur de Kaplan-Meier, selon la formule de Greenwood est :

$$\widehat{V}(\widehat{S}(y)) = (\widehat{S}(y))^2 * \sum_{i|y_i<=y} \frac{\widehat{q_i}}{\widehat{p_i} * r_i}$$

où $\widehat{p_i} = 1 - \widehat{q_i}$ et $\widehat{q_i} = \frac{s_i}{r_i}$. On peut alors réécrire la variance comme suit:

$$\widehat{V}(\widehat{S}(y)) = (\widehat{S}(y))^2 * \sum_{i|y_i<=y} \frac{s_i}{r_i * (r_i - s_i)}$$

Ainsi,

$$\widehat{V}(\widehat{S}(91)) = 0,375^2 * (\frac{1}{10*(10-1)} + \frac{1}{9*(9-1)} + \frac{1}{8*(8-1)} + \frac{2}{7*(7-2)} + \frac{1}{4*(4-1)})$$

$$\widehat{V}(\widehat{S}(91)) = 0.0257812$$

Un intervalle de confiance de niveau 95% pour $S(91)$ est:

$$\widehat{S}(91) \pm Z_{1-\frac{\alpha}{2}} * \sqrt{\widehat{V}(\widehat{S}(91))}$$

où $Z_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale centrée réduite.
Ce qui donne

$$[0.0602976, 0.6897024]$$

c) Calcul d'un intervalle de confiance de niveau 0,95 par la méthode log(-log) pour l'estimateur de Kaplan-Meier

Cet intervalle est de la forme:

$$S(91) \in [\widehat{S}(91)^{\frac{1}{U}}, \widehat{S}(91)^U]$$

où

$$exp[\frac{Z_{1-\frac{\alpha}{2}} * \sqrt{\widehat{V}(\widehat{S}(91))}}{\widehat{S}(91) * ln(\widehat{S}(91))}]$$

Nous avons l'intervalle de confiance :

$$[0.099486, 0.6591068]$$

On constate que l'intervalle de confiance log log est plus précis que l'intervalle de confiance linéaire au niveau de confiance 95%.