Économétrie avancée :
Introduction aux modèles de durée

Pr. Mokhtar KOUKI

Université de Carthage (ESSAI)
mokhtar.kouki@essai.u-carthage.tn

Décembre 2020

Contents

On considère une expérience dans laquelle on observe des patients pour lesquels on administre un traitement et on observe la **"durée jusqu'à ce que un événement survienne"** (décès, guérison, rechute, ...). On parle ainsi de **"temps de survie"**.

On considère une expérience dans laquelle on observe des patients pour lesquels on administre un traitement et on observe la **"durée jusqu'à ce que un événement survienne"** (décès, guérison, rechute, ...). On parle ainsi de **"temps de survie"**.
Quatre (04) éléments sont à considérer :

- T : le temps de survie (variable aléatoire)
- t : une réalisation de T
- F : la période de suivie (follow-up)
- δ : un évènement (Oui/Non) (variable aléatoire)

On considère une expérience dans laquelle on observe des patients pour lesquels on administre un traitement et on observe la **"durée jusqu'à ce que un événement survienne"** (décès, guérison, rechute, ...). On parle ainsi de **"temps de survie"**.
Quatre (04) éléments sont à considérer :

- T : le temps de survie (variable aléatoire)
- t : une réalisation de T
- F : la période de suivie (follow-up)
- δ : un évènement (Oui/Non) (variable aléatoire)

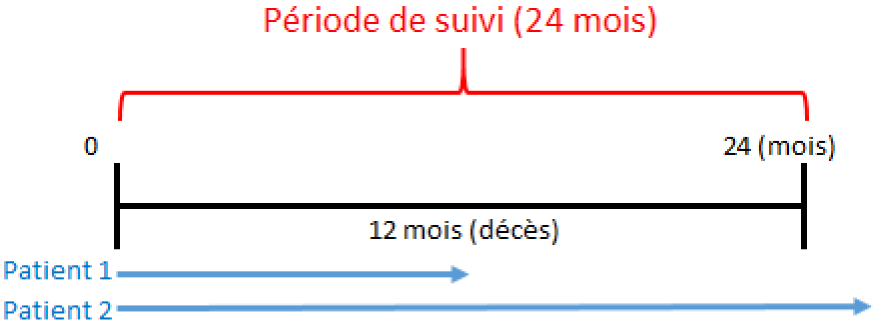
Exemples :

- Emploi : Délai d'insertion des diplômés
- Téléphonie mobile : Délai avant changement

En observant des durées, 2 configurations sont envisageables :

En observant des durées, 2 configurations sont envisageables :
Considérons, le suivie de patients sur une période de 24 mois. On observe le temps de survie avant décès :

En observant des durées, 2 configurations sont envisageables :
Considérons, le suivie de patients sur une période de 24 mois. On observe le temps de survie avant décès :



Remarque : Pour le "Patient 2" on observe pas le temps de survie avant décès. On parle ainsi de **censure**.

"Censure à gauche" et "Censure à droite" : Considérons l'exemple de suivi d'individus sur **12 mois** et on observe la durée jusqu'à la réalisation d'un événement donné (δ 1 si oui,0 sinon) :

i	t_i	δ_i	
1	7	1	
2	12	0	Censure à droite
3	10	0	
4	7	1	
5	5	0	Censure à gauche

Censure à gauche : Perte de suivi, Retrait, ...

Définition

Fonction de survie : La fonction de survie correspond à la probabilité qu'un individu survive au delà d'une date t :

$$S(t) = P(T \geq t) = 1 - P(T < t) = 1 - F(t),$$

avec $S(0) = 1$ et $S(\infty) = 0$

Définition

Fonction de survie : La fonction de survie correspond à la probabilité qu'un individu survive au delà d'une date t :

$$S(t) = P(T \geq t) = 1 - P(T < t) = 1 - F(t),$$

avec $S(0) = 1$ et $S(\infty) = 0$

Exemple : $T \rightsquigarrow Exp(\lambda)$

$$S(t) = \int_t^{\infty} f(x)dx = 1 - \int_0^t f(x)dx$$

Définition

Fonction de survie : La fonction de survie correspond à la probabilité qu'un individu survive au delà d'une date t :

$$S(t) = P(T \geq t) = 1 - P(T < t) = 1 - F(t),$$

avec $S(0) = 1$ et $S(\infty) = 0$

Exemple : $T \rightsquigarrow Exp(\lambda)$

$$\begin{aligned} S(t) &= \int_t^{\infty} f(x)dx = 1 - \int_0^t f(x)dx \\ &= 1 - \int_0^t \lambda \exp(-\lambda x)dx \end{aligned}$$

Définition

Fonction de survie : La fonction de survie correspond à la probabilité qu'un individu survive au delà d'une date t :

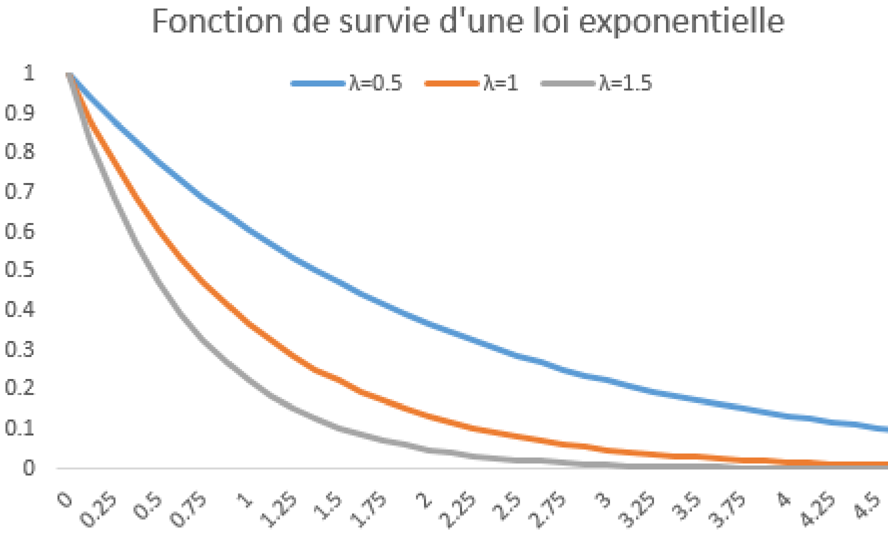
$$S(t) = P(T \geq t) = 1 - P(T < t) = 1 - F(t),$$

avec $S(0) = 1$ et $S(\infty) = 0$

Exemple : $T \rightsquigarrow Exp(\lambda)$

$$\begin{aligned} S(t) &= \int_t^{\infty} f(x)dx = 1 - \int_0^t f(x)dx \\ &= 1 - \int_0^t \lambda \exp(-\lambda x)dx \\ &= 1 - [-\exp(-\lambda x)]_0^t = \exp(-\lambda t) \end{aligned}$$

Fonctions de survie exponentielles



Définition

Fonction de hasard : La fonction de hasard correspond au risque, **par unité de temps**, pour que l'événement survienne sachant que l'individu a survécu jusqu'à la date t :

$$h(t) = \lim_{\Delta t \to 0} \frac{P\left(t \leq T < t + \Delta t / T \geq t\right)}{\Delta t}$$

Définition

Fonction de hasard : La fonction de hasard correspond au risque, **par unité de temps**, pour que l'événement survienne sachant que l'individu a survécu jusqu'à la date t :

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{P\left(t \leq T < t + \Delta t \,/\, T \geq t\right)}{\Delta t} \\ &= \frac{1}{P(T \geq t)} \lim_{\Delta t \to 0} \frac{P\left(t \leq T < t + \Delta t\right)}{\Delta t} \end{aligned}$$

Définition

Fonction de hasard : La fonction de hasard correspond au risque, **par unité de temps**, pour que l'événement survienne sachant que l'individu a survécu jusqu'à la date t :

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{P\left(t \leq T < t + \Delta t / T \geq t\right)}{\Delta t} \\ &= \frac{1}{P(T \geq t)} \lim_{\Delta t \to 0} \frac{P\left(t \leq T < t + \Delta t\right)}{\Delta t} \\ &= \frac{f(t)}{P(T \geq t)} \end{aligned}$$

Définition

Fonction de hasard : La fonction de hasard correspond au risque, **par unité de temps**, pour que l'événement survienne sachant que l'individu a survécu jusqu'à la date t :

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{P\left(t \leq T < t + \Delta t \middle/ T \geq t\right)}{\Delta t} \\ &= \frac{1}{P(T \geq t)} \lim_{\Delta t \to 0} \frac{P\left(t \leq T < t + \Delta t\right)}{\Delta t} \\ &= \frac{f(t)}{P(T \geq t)} \end{aligned}$$

Observation : $h(t) \in]0, +\infty[$

Pour une loi exponentielle de paramètre λ $(Exp(\lambda))$:

$$h(t) = \frac{f(t)}{S(t)} = \lambda$$

Exemple : Unité de temps.
Considérons deux individus ayant la même probabilité $P\left(t \leq T < t + \Delta t \diagup T \geq t\right) = 0.25)$ avec $\Delta t_1 = \Delta t_2 = 12h$. Pour le premier individu l'unité de temps le jour alors que pour le deuxième individu l'unité de temps la semaine :

$$\Delta t_1 = \frac{1}{2} j \text{ et } \Delta t_2 = \frac{1}{14} \text{semaine}$$

Exemple : Unité de temps.
Considérons deux individus ayant la même probabilité $P\left(t \leq T < t + \Delta t \middle/ T \geq t\right) = 0.25)$ avec $\Delta t_1 = \Delta t_2 = 12h$. Pour le premier individu l'unité de temps le jour alors que pour le deuxième individu l'unité de temps la semaine :

$$\Delta t_1 = \frac{1}{2} j \text{ et } \Delta t_2 = \frac{1}{14}\text{semaine}$$

$$h_1(t) = \frac{0.25}{\frac{1}{2}} = 0.5/jour$$

Exemple : Unité de temps.
Considérons deux individus ayant la même probabilité $P\left(t \leq T < t + \Delta t \diagup T \geq t\right) = 0.25)$ avec $\Delta t_1 = \Delta t_2 = 12h$. Pour le premier individu l'unité de temps le jour alors que pour le deuxième individu l'unité de temps la semaine :

$$\Delta t_1 = \frac{1}{2} j \text{ et } \Delta t_2 = \frac{1}{14} \text{semaine}$$

$$h_1(t) = \frac{0.25}{\frac{1}{2}} = 0.5/jour \text{ et } h_2(t) = \frac{0.25}{\frac{1}{14}} = 3.5/semaine$$

Exemple : Unité de temps.
Considérons deux individus ayant la même probabilité $P\left(t \leq T < t + \Delta t \middle/ T \geq t\right) = 0.25)$ avec $\Delta t_1 = \Delta t_2 = 12h$. Pour le premier individu l'unité de temps le jour alors que pour le deuxième individu l'unité de temps la semaine :

$$\Delta t_1 = \frac{1}{2} j \text{ et } \Delta t_2 = \frac{1}{14} \text{semaine}$$

$$h_1(t) = \frac{0.25}{\frac{1}{2}} = 0.5/jour \text{ et } h_2(t) = \frac{0.25}{\frac{1}{14}} = 3.5/semaine$$

Relation entre fonction de survie et fonction de hasard

$$S(t) = P(T \geq t) = 1 - F(t) \Rightarrow$$

Relation entre fonction de survie et fonction de hasard

$$S(t) = P(T \geq t) = 1 - F(t) \Rightarrow S'(t) = -f(t)$$

Relation entre fonction de survie et fonction de hasard

$$
\begin{aligned}
S(t) &= P(T \geq t) = 1 - F(t) \Rightarrow S'(t) = -f(t) \\
h(t) &= \frac{f(t)}{S(t)} =
\end{aligned}
$$

Relation entre fonction de survie et fonction de hasard

$$\begin{aligned} S(t) &= P(T \geq t) = 1 - F(t) \Rightarrow S'(t) = -f(t) \\ h(t) &= \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = \end{aligned}$$

Relation entre fonction de survie et fonction de hasard

$$\begin{aligned} S(t) &= P(T \geq t) = 1 - F(t) \Rightarrow S'(t) = -f(t) \\ h(t) &= \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\ln(S(t))' \end{aligned}$$

En conclusion :

$$S(t) = \exp\left(-\int_0^t h(x)dx\right)$$

X suit une loi de Weibull ssi :

$$f(x, \beta, \lambda) = \frac{\beta}{\lambda}\left(\frac{x}{\lambda}\right)^{\beta-1} \exp -\left(\frac{x}{\lambda}\right)^{\beta}$$

$x > 0$, $\beta > 0$, paramètre de forme et $\lambda > 0$, paramètre d'échelle.

X suit une loi de Weibull ssi :

$$f(x, \beta, \lambda) = \frac{\beta}{\lambda}\left(\frac{x}{\lambda}\right)^{\beta-1} \exp - \left(\frac{x}{\lambda}\right)^{\beta}$$

$x > 0$, $\beta > 0$, paramètre de forme et $\lambda > 0$, paramètre d'échelle.

Fonction de survie et fonction de hasard

$$S(t) = P(T \geq t) = \int_t^{\infty} \frac{\beta}{\lambda}\left(\frac{x}{\lambda}\right)^{\beta-1} \exp - \left(\frac{x}{\lambda}\right)^{\beta} dx$$

X suit une loi de Weibull ssi :

$$f(x, \beta, \lambda) = \frac{\beta}{\lambda}\left(\frac{x}{\lambda}\right)^{\beta-1} \exp - \left(\frac{x}{\lambda}\right)^{\beta}$$

$x > 0$, $\beta > 0$, paramètre de forme et $\lambda > 0$, paramètre d'échelle.

Fonction de survie et fonction de hasard

$$\begin{aligned} S(t) = P(T \geq t) &= \int_t^{\infty} \frac{\beta}{\lambda}\left(\frac{x}{\lambda}\right)^{\beta-1} \exp - \left(\frac{x}{\lambda}\right)^{\beta} dx \\ &= \left[- \exp - \left(\frac{x}{\lambda}\right)^{\beta}\right]_t^{\infty} \end{aligned}$$

X suit une loi de Weibull ssi :

$$f(x,\beta,\lambda) = \frac{\beta}{\lambda}\left(\frac{x}{\lambda}\right)^{\beta-1} \exp -\left(\frac{x}{\lambda}\right)^{\beta}$$

$x > 0$, $\beta > 0$, paramètre de forme et $\lambda > 0$, paramètre d'échelle.

Fonction de survie et fonction de hasard

$$\begin{aligned} S(t) = P(T \geq t) &= \int_t^{\infty} \frac{\beta}{\lambda}\left(\frac{x}{\lambda}\right)^{\beta-1} \exp -\left(\frac{x}{\lambda}\right)^{\beta} dx \\ &= \left[-\exp -\left(\frac{x}{\lambda}\right)^{\beta}\right]_t^{\infty} \\ &= \exp -\left(\frac{t}{\lambda}\right)^{\beta} \end{aligned}$$

X suit une loi de Weibull ssi :

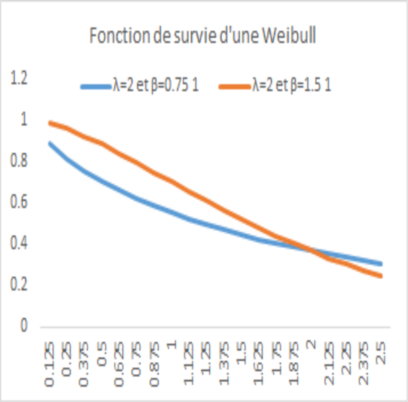
$$f(x,\beta,\lambda) = \frac{\beta}{\lambda}\left(\frac{x}{\lambda}\right)^{\beta-1} \exp - \left(\frac{x}{\lambda}\right)^{\beta}$$

$x > 0$, $\beta > 0$, paramètre de forme et $\lambda > 0$, paramètre d'échelle.

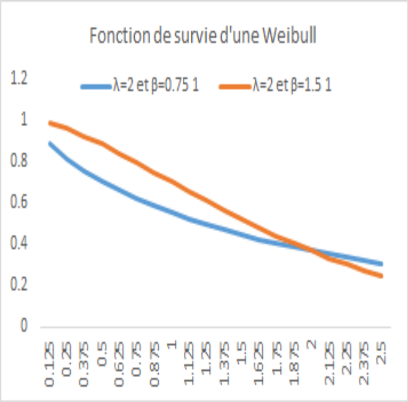
Fonction de survie et fonction de hasard

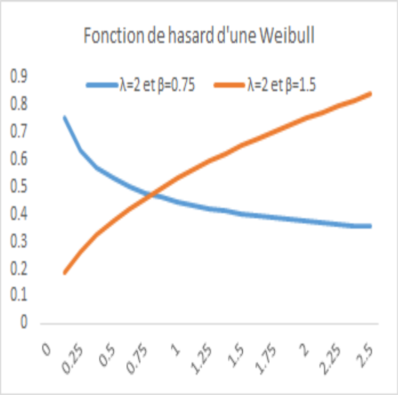
$$\begin{aligned}
S(t) = P(T \geq t) &= \int_t^{\infty} \frac{\beta}{\lambda}\left(\frac{x}{\lambda}\right)^{\beta-1} \exp - \left(\frac{x}{\lambda}\right)^{\beta} dx \\
&= \left[- \exp - \left(\frac{x}{\lambda}\right)^{\beta} \right]_t^{\infty} \\
&= \exp - \left(\frac{t}{\lambda}\right)^{\beta} \\
h(t) = \frac{f(t)}{S(t)} &= \frac{\beta}{\lambda}\left(\frac{t}{\lambda}\right)^{\beta-1}
\end{aligned}$$

Fonction de survie et fonction de hazard d'une Weibull



Fonction de survie et fonction de hazard d'une Weibull





X suit une loi log-normale de paramètre m et σ^2 ssi :

X suit une loi log-normale de paramètre m et σ^2 ssi :

$$f(x) = \frac{1}{x\sqrt{2\pi}\sigma} \exp -\frac{1}{2}\frac{(\ln(x)-m)^2}{\sigma^2}, x > 0$$

X suit une loi log-normale de paramètre m et σ^2 ssi :

$$
\begin{aligned}
f(x) &= \frac{1}{x\sqrt{2\pi}\sigma} \exp -\frac{1}{2}\frac{(\ln(x)-m)^2}{\sigma^2}, x > 0 \\
&= \frac{1}{x}\phi\left(\frac{\ln(x)-m}{\sigma}\right)
\end{aligned}
$$

X suit une loi log-normale de paramètre m et σ^2 ssi :

$$
\begin{aligned}
f(x) &= \frac{1}{x\sqrt{2\pi}\sigma} \exp -\frac{1}{2}\frac{(\ln(x)-m)^2}{\sigma^2}, x > 0 \\
&= \frac{1}{x}\phi\left(\frac{\ln(x)-m}{\sigma}\right) \\
S(t) &= \int_t^{\infty} \frac{1}{x\sqrt{2\pi}\sigma} \exp -\frac{1}{2}\frac{(\ln(x)-m)^2}{\sigma^2} dx
\end{aligned}
$$

X suit une loi log-normale de paramètre m et σ^2 ssi :

$$
\begin{aligned}
f(x) &= \frac{1}{x\sqrt{2\pi}\sigma} \exp -\frac{1}{2}\frac{(\ln(x)-m)^2}{\sigma^2}, x > 0 \\
&= \frac{1}{x}\phi\left(\frac{\ln(x)-m}{\sigma}\right) \\
S(t) &= \int_t^{\infty} \frac{1}{x\sqrt{2\pi}\sigma} \exp -\frac{1}{2}\frac{(\ln(x)-m)^2}{\sigma^2} dx
\end{aligned}
$$

On adopte le changement de variable :

$$
U = \frac{\ln(x)-m}{\sigma} \Rightarrow du = \frac{dx}{\sigma x}
$$

X suit une loi log-normale de paramètre m et σ^2 ssi :

$$
\begin{aligned}
f(x) &= \frac{1}{x\sqrt{2\pi\sigma}} \exp -\frac{1}{2}\frac{(\ln(x)-m)^2}{\sigma^2}, x > 0 \\
&= \frac{1}{x}\phi\left(\frac{\ln(x)-m}{\sigma}\right) \\
S(t) &= \int_t^{\infty} \frac{1}{x\sqrt{2\pi\sigma}} \exp -\frac{1}{2}\frac{(\ln(x)-m)^2}{\sigma^2} dx
\end{aligned}
$$

On adopte le changement de variable :

$$U = \frac{\ln(x)-m}{\sigma} \Rightarrow du = \frac{dx}{\sigma x}$$

Ce qui donne comme fonction de survie et fonction de hasard :

$$S(t) = \int_{\frac{\ln(t)-m}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2})du$$

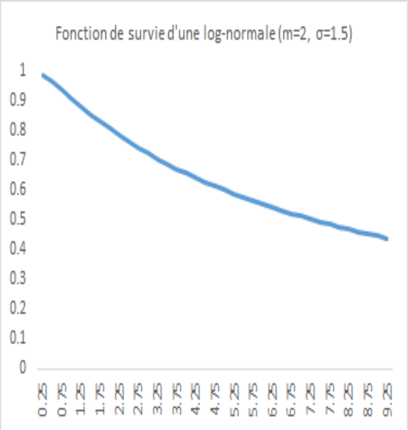
Ce qui donne comme fonction de survie et fonction de hasard :

$$\begin{aligned} S(t) &= \int_{\frac{\ln(t)-m}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \\ &= 1 - \Phi\left(\frac{\ln(t)-m}{\sigma}\right) \end{aligned}$$

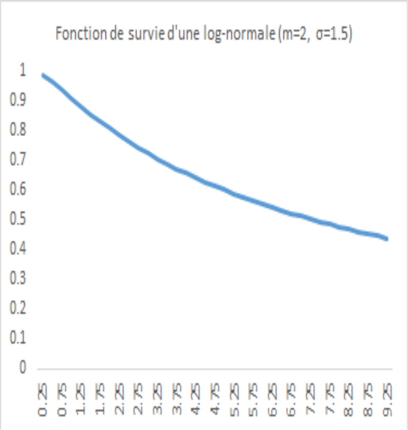
Ce qui donne comme fonction de survie et fonction de hasard :

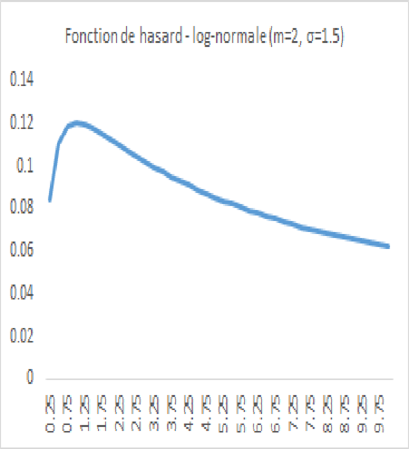
$$
\begin{aligned}
S(t) &= \int_{\frac{\ln(t)-m}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2}) du \\
&= 1 - \Phi\left(\frac{\ln(t)-m}{\sigma}\right) \\
h(t) &= \frac{f(t)}{S(t)} = \frac{\phi\left(\frac{\ln(t)-m}{\sigma}\right)}{t\left[1-\Phi\left(\frac{\ln(t)-m}{\sigma}\right)\right]}
\end{aligned}
$$

Fonction de survie et fonction de hasard d'une log-normale



Fonction de survie et fonction de hasard d'une log-normale





On considère les données d'une l'expérience sur deux traitements sur 100 patients (50 dans chaque groupe) (fichier Excel).

- classer, par ordre croissant, les données selon les durées de survie
- enlever les durées correspondant à une censure
- pour chaque groupe, calculer les colonnes suivantes

  i. $a = \#(\delta_i = 1)$ à la date $t_{[i]}$

  ii. $b = \#(\delta_i = 0)$ à la date $t_{[i]}$

  iii. $c = \#$ survivants à la date $t_{[i]} = N - a - b$

  iv $S(t_{[i]}) = c/N$

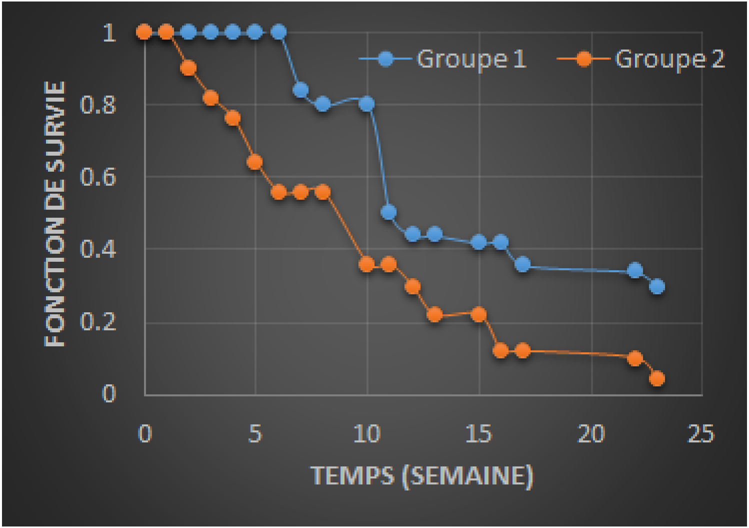


Figure – Fonction de survie

On considère les durées ordonnées par ordre croissant, $t_1, t_2, \cdots, T_{max}$. L'estimateur de la probabilité de survie à l'instant t_i est défini par :

$$\hat{S}(t_i) = \prod_{j=1}^{i} \hat{P}(T > t_j / T \geq t_j) = \prod_{j=1}^{i} \left(\frac{N(t_j) - D(t_j)}{N(t_j)} \right)$$

avec :

- $N(t_j)$: Nombre d'individus à risque avant t_j
- $D(t_j)$: Nombre d'événements (i.e. Décès) survenus à la date t_j
- $S(t) = 1$ si $t < t_1$

Exemple sous R : Package survival : leucémie : temps de survie (semaines), groupe, censure, âge, sexe.

$>$ *surv1*¡-*survfit*(Surv(time, event) 1, data=survie)

$>$ *plot*(surv1)

On considère 2 groupes d'individus (patients). A chaque date t_i on observe les quantités suivantes :

- $N_g(t_j)$: Nombre d'individus à risque avant t_j du groupe g, $(g = 1, 2)$
- $D_g(t_j)$: Nombre d'événements (i.e. Décès) à la date t_j du groupe g, $(g = 1, 2)$
- Le nombre d'événements estimés à l date t_j est donné par :

$$e_g(t_j) = \frac{N_g(t_i)}{N_1(t_i) + N_2(t_i)} \times (D_1(t_i) + D_2(t_i))$$

La statistique de comparaison des deux courbe de survies (Statistique de Log-Rank) est définie par :

$$log - rank = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \rightsquigarrow \chi^2(1)$$

avec O_g le nombre total d'événements dans le groupe g et $E_g = \sum_{t_i} e_g(t_i)$

instructions R

$>$ *survfit*(*Surv*(time, censure)   Group, data = survie)
$>$ *plot*(*survfit*(*Surv*(time, censure)   Group, data = survie))
$>$ *survdiff* (*Surv*(time, censure)   Group, data = survie)

Le modèle de Cox est défini la fonction de hasard comme suit :

$$h(t) = h_0(t) \exp \left(\sum_{l=1}^{k} \beta_l X_l \right)$$

avec X_1, X_2, $\cdots$, X_k les variables explicatives et $h_0(t)$ une fonction qui ne dépend que du temps (risque de base).
Ratio de hasard : Hazard ratio : Considérons deux groupes d'individus pour lesquels on observe les X_{gl}, $g = 1, 2$. Le ratio de hasard (risque) est défini par :

$$HR = \exp \left(\sum_{l=1}^{k} \beta_l \left(X1l - X2l \right) \right) l$$

Instructions R

$>$ *coxph*(*Surv*(time, censure)   Group + Log_wbc+sexe+age , data = survie)

MERCI POUR VOTRE ATTENTION