Remerciement

Je tiens tout d'abord à remercier notre encadreur **F. Maref**, je le remercie pour tous ses conseils lors de la rédaction de ce mémoire. Son expérience m'a été tout à fait profitable.

Je tiens à exprimer ma reconnaissance aux membres du jury de m'avoir fait l'honneur de participer à ma soutenance.

Je me permets également de remercier mes parents, mes frères pour leur soutient moral et leur encouragement tout au long de se mémoire.

je n'oublies pas d'adresser notre gratitude à tous nos amis et collègues pour leurs soutiens moral.

Je tiens enfin à remercier toutes les personnes non citées qui auraient contribué d'une manière ou d'une autre à la réalisation de ce travail.

Table des matières

Table des figures

Introduction

Le terme de durée de survie désigne le temps écoulé jusqu'à la survenue d'un événement précis. L'événement étudié (communément appelé "décès") est le passage irréversible entre deux états (communément nommé "vivant" et "décès"). L'événement terminal n'est pas forcément la mort, il peut s'agir de l'apparition d'une maladie (par exemple, le temps avant une rechute ou un rejet de greffe), d'une guérison (temps entre le diagnostic et la guérison), la panne d'une machine (durée de fonctionnement d'une machine, en fiabilité) ou la survenue d'un sinistre (temps entre deux sinistres, en actuariat).

L'analyse des données durées de survie est l'étude du délai de la survenue de cet événement. Dans le domaine biomédical, on étudie ces durées dans le contexte des études longitudinales comme les enquêtes de cohorte (suivi de patients dans le temps) ou les essais thérapeutiques (tester l'efficacité d'un médicament). On cherche alors dans ce mémoire à estimer la distribution des temps de survie dans le cas paramétrique et à comparer les fonctions de survie de deux groupes .

Une caractéristique importante de l'analyse de la survie est la présence des données censurées. Cette caractéristique, source de difficulté, a nécessité le développement de techniques alternatives à l'inférence usuelle. Les données censurées sont des observations pour lesquelles la valeur exacte d'un événement n'est pas toujours connue. Cependant, nous disposons tout de même d'une information partielle permettant de fixer une borne inférieure (censure à droite) ou une borne supérieure (censure à gauche). Les raisons de cette censure peuvent être le fait que le patient soit toujours vivant ou non malade à la fin de l'étude, ou qu'il se soit retiré de l'étude pour des raisons personnelles (immigration, mutation professionnelle).

En 1912, au moment où Ronald Fisher rédige son premier article consacré au maximum de vraisemblance, les deux méthodes statistiques les plus utilisées sont la méthode des moindres carrés et la méthode des moments. Dans son article de 1912, il prend l'exemple d'une loi normale. En 1921, il applique la même méthode à l'estimation d'un coefficient de corrélation.

Weibull (1951), propose un modèle paramétrique pour calculer la fiabilité d'un système non réparable. Il aborde notamment la présence de données tronquées ou censurées.

Le travail présenté dans ce mémoire a pour objectif d'étudier l'estimation paramétrique de la fonction de survie. Cette approche paramétrique stipule l'appartenance de la loi de probabilité réelle des observations à une classe particulière de lois, qui dépendent d'un certain nombre (fini) de paramètres. L'avantage de cette approche est la facilitation attendue de la phase d'estimation des paramètres, ainsi que de l'obtention d'intervalles de confiance et de la construction de tests. L'inconvénient de la méthode paramétrique est l'inadéquation pouvant exister entre le phénomène étudié et le modèle retenu.

Ce mémoire est partagé en trois chapitres. Dans le premier chapitre, nous rappelons des préliminaires sur les modèles de survie. Nous introduisons les principales fonctions en analyse de survie : fonction de survie, taux de survie et les différentes formes du taux de risque ect .... Nous donnons aussi les différents types de censure (censure à droite, censure à gauche, censure par intervalle, troncature ect ...). Le deuxièmme chapitre est consacré à l'estimation paramétrique de la fonction de survie, on utilise la méthode de maximum de vraisemblance. Dans le dernier chapitre, on fait une comparaison entre les taux de hasard de deux groupes, on utilise le test de wald, test du rapport de vraisemblance et le test de score.

Chapitre 1

Généralités sur l'analyse de survie

L'analyse des données de survie étudie le délai jusqu'à l'apparition d'un évènement pour un ensemble d'individus. A l'origine, cet événement désignait le décés" mais d'autres événements peuvent être considérés tels que la survenue d'une maladie en épidiemiologie, la survenue d'une panne dans les applications industrielles, l'acceptation d'une offre d'emploi pour une personne au chômage en économie. Dans ce chapitre nous rappellerons quelques définitions et notations utiles pour la suite de mémoire.

1.1 Définitions

Quelques définitions sont couramment utilisées dans les études de survie.

- ⋆ **Date de début d'étude** : elle correspond à la date des début de rassembler des informations sur un sujet
- ⋆ **Date d'origine** : elle correspond à l'origine de la durée étudiée. Elle peut être la date de naissance, le début d'une exposition à un facteur de risque, la date d'une opération chirurgicale, la date de début d'une maladie ou la date d'entrée dans l'étude. Chaque individu peut donc avoir une date d'origine différente (pas important car c'est la durée qui nous intéresse).
- ⋆ **Date de point** : c'est la date au-delà de laquelle on arrêtera l'étude et on ne tiendra plus compte des informations sur les sujets.
- ⋆ **Date des dernières nouvelles** : c'est la date la plus récente où des informations sur un sujet ont été recueillies.

1.2 Distributions de la durée de survie

Supposons que la durée de survie X soit une variable aléatoire positive ou nulle et absolument continue, alors sa loi de probabilité peut être définie par l'une des cinq fonctions équivalentes suivantes (chacune des fonctions ci-dessous peut être obtenue à partir de l'une des autres fonctions) :

1.2.1 Fonction de survie S

On appelle fonction de survie, notée S, la fonction définie par :

$$S(t) \quad = \quad \mathbb{P}(X > t), \qquad t \geqslant 0.$$

$S(t)$ s'interprète comme la probabilité de survivre au moins une durée t.

1.2.2 Fonction de répartition F

La fonction de répartition (ou c.d.f. pour "cumulative distribution function") représente, pour t fixé, la probabilité de mourir avant l'instant t, c'est-à-dire

$$F(t) \quad = \quad \mathbb{P}(X \leqslant t) = 1 - S(t).$$

Remarque 1 Il est arbitraire de décider que $S(t) = \mathbb{P}(X \geqslant t)$ où $S(t) = \mathbb{P}(X > t)$ Cela n'a aucune importance quand la loi de X est continue car $\mathbb{P}(X > t) = \mathbb{P}(X \geqslant t)$. Dans les cas où F a des sauts (quand le temps est discret, par exemple, compté en mois ou semaine), on utilise les notations suivantes :

$$F^-(t) = \mathbf{P}(X < t) \quad et \quad F^+(t) = \mathbf{P}(X \leqslant t)$$

où F^- est la limite à gauche et F^+ la limite à droite de F (définitions et notations sont identiques pour la fonction S). Remarquons que $F^- \leqslant F^+$ et $S^- \geqslant S^+$.

1.2.3 Densité de probabilité f

C'est la fonction $f(t) \geqslant 0$, telle que pour, tout $t \geqslant 0$,

$$F(t) \quad = \quad \int_0^t f(u)du.$$

Si la fonction de répartition F admet une dérivée au point t alors

$$f(t) = \lim_{h\to 0} \frac{\mathbb{P}(t \leqslant X < t+h)}{h} = F^{'}(t) = -S^{'}(t).$$

Pour t fixé, la densité de probabilité représente la probabilité de mourir dans un petit intervalle de temps après l'instant t.

1.2.4 Risque instantané h (ou taux de hasard)

Le risque instantané (ou taux d'incidence), pour t fixé, caractérise la probabilité de mourir dans un petit intervalle de temps après t, conditionnellement au fait d'avoir survécu après temps t (c'est-à-dire le risque de mort instantané pour ceux qui ont survécu) :

$$h(t) = \lim_{h\to 0} \frac{\mathbb{P}(t \leqslant X < t+h | X \geqslant t)}{h} = \frac{f(t)}{S(t)} = -(ln(S(t)))^{'}$$

En effet :

$$\begin{aligned} h(t) &= \lim_{h\to 0} \frac{\mathbb{P}(t \leqslant X < t+h | X \geqslant t)}{h} = \lim_{h\to 0} \frac{\mathbb{P}(t \leqslant X < t+h \cap X \geqslant t)}{h \times \mathbb{P}(X \geqslant t)} \\ &= \lim_{h\to 0} \frac{\mathbb{P}(t \leqslant X < t+h)}{h \times S(t)} = \frac{f(t)}{S(t)} \\ &= \frac{-S^{'}(t)}{S(t)} = -(ln(S(t)))^{'} \end{aligned}$$

1.2.5 Taux de hasard cumulé H

Le taux de hasard cumulé est l'intégrale du risque instantané h,

$$H(t) = \int_0^t h(u)du = -ln(S(t)).$$

On peut déduire de cette équation une expression de la fonction de survie en fonction du taux de hasard cumulé (ou du risque instantané) :

$$S(t) = exp(-H(t)) = exp\left(-\int_0^t h(u)du\right).$$

On en déduit que

$$f(t) = h(t)exp\left(-\int_0^t h(u)du\right).$$

1.2.6 Quantités associées à la distribution de survie

Moyenne et variance de la durée de survie

Le temps moyen de survie $\mathbb{E}(X)$ et la variance de la durée de survie $\mathbb{V}(X)$ sont définis par les quantités suivantes :

$$\mathbb{E}(X) = \int_0^\infty S(t)dt$$

et

$$\mathbb{V}(X) = 2\int_0^\infty tS(t)dt - (\mathbb{E}(X)^2).$$

En effet

$$\begin{aligned} \mathbb{E}(X) &= \int_0^\infty tf(t)dt \\ &= \int_0^\infty -tS^{'}(t)dt, \end{aligned}$$

on utilise l'intégrale par partie, on obtient,

$$\begin{aligned} \int_0^\infty -tS^{'}(t)dt &= [-tS(t)]_0^\infty - \int_0^\infty -S(t)dt \\ &= \int_0^\infty S(t)dt. \end{aligned}$$

Alors,

$$\mathbb{E}(X) = \int_0^\infty S(t)dt.$$

On ce qui conserne la variance, on a

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \\ &= \int_0^\infty -t^2S^{'}(t)dt - (\mathbb{E}(X))^2, \end{aligned}$$

l'intégration par partie donne,

$$\begin{aligned}\int_0^\infty -t^2 S^{'}(t)dt &= [-t^2 S(t)]_0^\infty - \int_0^\infty -2tS(t)dt \\ &= 2\int_0^\infty tS(t)dt,\end{aligned}$$

donc

$$\mathbb{V}(X) = \int_0^\infty 2tS(t)dt - (\mathbb{E}(X)^2).$$

Ainsi on peut déduire l'espérance et la variance à partir de n'importe laquelle des fonctions F, S, f, h, H (mais pas l'inverse).

Quantiles de la durée de survie

- La médiane de la durée de survie est le temps t pour lequel la probabilité de survie $S(t)$ est égale à 0.5, c'est-à-dire, la valeur t_m qui satisfait $S(t_m) = 0.5$.
- La fonction quantile de la durée de survie est définie par

$$\begin{aligned}q(p) &= inf(t : F(t) \geq p), \qquad 0 < p < 1, \\ &= inf(t : S(t) \leq 1 - p).\end{aligned}$$

  Lorsque la fonction de répartition F est strictement croissante et continue alors

$$\begin{aligned}q(p) &= F^{-1}(p) \\ &= S^{-1}(1 - p).\end{aligned}$$

  Le quantile $q(p)$ est le temps où une proportion p de la population a disparu .

1.3 Censure et troncature

Une des caractéristiques des données de survie est l'existence d'observations incomplètes. En effet, les données sont souvent recueillies partiellement, notamment, à cause des processus de censure et de troncature. Les données censurées ou tronquées proviennent du fait qu'on n'a pas accès à toute l'information : au lieu d'observer des réalisations indépendantes et identiquement distribuées (i.i.d.) de durées X, on observe la réalisation de la variable X soumise à diverses perturbations, indépendantes ou non du phénomène étudié.

1.3.1 Censure

La censure est le phénomène le plus couramment rencontré lors du recueil de données de survie. Pour l'individu i ; considérons
- son temps de survie X_i,
- son temps de censure C_i,
- la durée réellement observée T_i.

Censure à droite

La durée de vie est dite censurée à droite si l'individu n'a pas subi l'événement à sa dernière observation. En présence de censure à droite, les durées de vie ne sont pas toutes observées ; pour certaines d'entre elles, on sait seulement qu'elles sont supérieures à une certaine valeur connue.

1. La censure de type I
   Soit C une valeur fixée, au lieu d'observer les variables $X_1, \ldots, X_n$ qui nous intéressent, on n'observe X_i uniquement lorsque $X_i \leq C$; sinon on sait uniquement que $X_i > C$. On utilise la notation suivante :$T_i = X_i \wedge C = min(X_i, C)$. Ce mécanisme de censure est fréquemment rencontré dans les applications industrielles. Par exemple, Dans l'apprentissage d'une langue par un groupe d'étudiants durant un stage de période fixée. On note T la durée d'apprentissage de cette langue. Pour certains étudiants nous allons observer leurs durées T_i d'apprentissage de la langue par contre pour d'autres leurs T_i ne seront pas observées car le stage est limité dans le temps.
2. La censure de type II
   Elle est présente quand on décide d'observer les durées de survie des n patients jusqu'à ce que k d'entre eux soient décédés et d'arrêter l'étude à ce moment là. Soient $X_{(i)}$ et $T_{(i)}$ les statistiques d'ordre des variables X_i et T_i : La date de censure est donc $X_{(k)}$ et on observe les variables suivantes

$$\begin{aligned} T_{(1)} &= X_{(1)} \\ &\vdots \\ T_{(k)} &= X_{(k)} \\ T_{(k+1)} &= X_{(k)} \\ &\vdots \\ T_{(n)} &= X_{(k)} \end{aligned}$$

3. La censure de type III (ou censure aléatoire de type I) :
   Soient $C_1, \ldots, C_n$ des variables aléatoires i.i.d. On observe les variables

$$T_i \quad = \quad X_i \wedge C_i.$$

   L'information disponible peut être résumée par :
   – la durée réellement observée T_i ;

   – un indicateur $\delta_i = \mathbb{I}_{X_i \leq C_i}$

   – $\delta_i = 1$ si l'événement est observé (d'où $T_i = X_i$). On observe les "vraies" durées ou les durées complètes.

   – $\delta_i = 0$ si l'individu est censuré (d'où $T_i = C_i$). On observe des durées incomplètes (censurées).

La censure aléatoire est la plus courante. Par exemple, lors d'un essai thérapeutique, elle peut être engendrée par

1. la perte de vue : le patient quitte l'étude en cours et on ne le revoit plus (à cause d'un déménagement, le patient décide de se faire soigner ailleurs). Ce sont des patients "perdus de vue".
2. l'arrêt ou le changement du traitement : les effets secondaires ou l'inecacité du traitement peuvent entraîner un changement ou un arrêt du traitement. Ces patients sont exclus de l'étude.
3. la fin de l'étude : l'étude se termine alors que certains patients sont toujours vivants (ils n'ont pas subi l'événement). Ce sont des patients "exclus-vivants". Les "perdus de vue" (et les exclusions) et les "exclus-vivants" correspondent à des observations censurées mais les deux mécanismes sont de nature différente (la censure peut être informative chez les "perdus de vue").

Censure à gauche

La censure à gauche correspond au cas où l'individu a déjà subi l'événement avant que l'individu soit observé. On sait uniquement que la date de l'événement est inférieure à une certaine date connue. Pour chaque individu, on peut associer un couple de variables aléatoires (T, δ) :

$$\begin{aligned} T &= X \vee C = max(X, C), \\ \delta &= \mathbb{I}_{X \geq C}. \end{aligned}$$

Comme pour la censure à droite, on suppose que la censure C est indépendante de X. Un des premiers exemples de censure à gauche rencontré dans la littérature considère le cas d'observateurs qui s'intéressent à l'horaire où les babouins descendent de leurs arbres pour aller manger (les babouins passent la nuit dans les arbres). Le temps d'événement (descente de l'arbre) est observé si le babouin descend de l'arbre après l'arrivée des observateurs. Par contre, la donnée est censurée si le babouin est descendu avant l'arrivée des observateurs : dans ce cas on sait uniquement que l'horaire de descente est inférieur à l'heure d'arrivée des observateurs. On observe donc le maximum entre l'heure de descente des babouins et l'heure d'arrivée des observateurs (l'heure correspond à une durée).

Censure par intervalle

Une date est censurée par intervalle si au lieu d'observer avec certitude le temps de l'événement, la seule information disponible est qu'il a eu lieu entre deux dates connues. Par exemple, dans le cas d'un suivi de cohorte, les personnes sont souvent suivies par intermittence (pas en continu), on sait alors uniquement que l'événement s'est produit entre ces deux temps d'observations. On peut noter que pour simplifier l'analyse, on fait souvent l'hypothèse que le temps d'événement correspond au temps de la visite pour se ramener à de la censure à droite.

1.3.2 Troncature

Nous parlons de troncature à droite (respectivement à gauche) lorsque la variable d'intérêt n'est pas observable quand elle est supérieure (respectivement inférieure) à un seuil C fixé. Dans le cadre de la censure, la variable C est observée alors que dans le cas de la troncature à droite (respectivement à gauche) l'analyse porte uniquement sur la loi de T conditionnellement à l'événement $T < C$ (respectivement $T > C$) et une donnée tronquée ne peut faire partie de l'échantillon. Si une maison de retraite n'accepte que des personnes âgées d'au moins soixante ans, aucun individu décédé avant cet âge n'a la possibilité d'y avoir été admis et est de ce fait tronqué à gauche.

Exemple

Durée de vie après la retraite : on étudie la durée de vie après la retraite de sujets qui entrent dans l'enquête à la suite d'un tirage au sort dans une caisse de retraite et l'instant de l'enquête, la durée de vie après la retraite est donc tranquée à gauche par ce délai. Elle peut être censurée à droite si la fin de l'enquête à lieu alors que le sujet est toujours vivant.

Chapitre 2

Estimation paramétrique

Lors d'une approche paramétrique, on fait l'hypothèse que la distribution des données de survie est caractérisée par une fonction connue qui est entièrement déterminée par un vecteur de paramètres réels de dimension finie. Les modèles les plus utilisés en analyse paramétrique de la survie sont les modèles exponentiels et Weibull. Les paramètres sont obtenus en maximisant la vraisemblance du modèle. Cette méthode a été développée par le statisticien Ronald Aylmer Fisher en 1921. Ce chapitre est divisé en deux sections, dans le premier section, on présente la méthode de maximum de vraisemblance en absence de censure. Dans le deuxième section, nous expliquons comment la méthode de maximum de vraisemblance peut être appliquée au cas de données censurées.

2.1 Quelques rappels sur les estimateurs

On considère un modèle statistique $(\mathcal{X}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ dominé par une mesure λ (Lebesgue). $\mathcal{X}$ est l'espace des observations, $(\mathbb{P}_\theta)_{\theta\in\Theta}$ est une famille de lois de probabilités dépendant du paramètre θ à valeurs dans Θ et absolument continues par rapport à λ (i.e $\mathbb{P}_\theta$ admet une densitè f_θ par rapport à λ).

Definition 2.1.1 *Dans un modèle statistique $(\mathcal{X}, (\mathbb{P}_\theta)_{\theta\in\Theta})$, un estimateur du paramétre θ est une fonction des données à valeurs dans Θ. On parle de statistique $\theta_n = \varphi(X_1, \cdots, X_n)$.*

- *Le biais d'un estimateur θ_n est défini par $b_n(\theta) = \mathbb{E}(\theta_n) - \theta$.*
- *Si $\forall\theta \in \Theta$, $b_n(\theta) = 0$ on dit que l'estimateur θ_n est sans biais.*
- *Un estimateur θ_n est dit convergent si $\theta_n \to \theta$ en probabilité, $\forall\theta \in \Theta$.*
- *Un estimateur sans biais tel que $\mathbb{V}(\theta_n) \to 0$ lorsque $n \to +\infty$ est convergent.*

2.2 Maximum de vraisemblance (Données complètes)

Definition 2.2.1 *Soit $(T_1, \cdots, T_n)$ un échantillon indépendant et identiquement distribué (i.i.d.) de densité f_θ. On appelle vraisemblance du modèle $(\mathcal{X}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ (ici $\mathcal{X} = \mathbb{R}^+$), l'application : $L : \mathcal{X} \times \Theta \to \mathbb{R}^+$, définie par :*

$$L(t_1, \cdots, t_n, \theta) \;=\; \prod_{i=1}^{n} f_\theta(t_i).$$

Estimateur du maximum de vraisemblance

Definition 2.2.2 *Soit $(T_1, \cdots, T_n)$ un échantillon de $T \sim P_\theta$, $\theta \in \Theta$. Un estimateur du maximum de vraisemblance (EMV) du paramètre θ est un estimateur $\widehat{\theta}_n$ qui maximise la vraisemblance :*

$$L(t_1, \cdots, t_n, \hat{\theta}_n) \;=\; \sup_{\theta\in\Theta} L(t_1, \cdots, t_n, \theta)$$

En pratique, on cherche plutôt à maximiser la log-vraisemblance (qui transforme le produit en somme) :

$$\log L(t_1, \cdots, t_n, \theta) \;=\; \log\left(\prod_{i=1}^{n} f_\theta(t_i)\right) = \sum_{i=1}^{n} \log f_\theta(t_i).$$

à maximiser par rapport à θ.

2.2.1 Information de Fisher

Dans un modèle statistique $((\mathcal{X}, \mathbb{P}_\theta)_{\theta\in\Theta}))$,

Definition 2.2.3 *On appelle score la fonction*

$$U(\theta) \;=\; \frac{\partial}{\partial\theta} \log L(\cdot, \theta).$$

Definition 2.2.4 *On appelle information de Fisher la quantité :*

$$I(\theta) \;=\; \mathbb{V}\left(\frac{\partial \log L(\cdot, \theta)}{\partial\theta}\right) = \mathbb{V}(U(\theta)).$$

On cherche l'EMV en annulant la fonction de score $U(\theta) = 0$.

Exemples

Distribution uniforme $\mathcal{U}[a,b]$.

On suppose que l'échantillon $X_1, \cdots, X_n$ est tiré de manière uniforme entre a et b, mais a et b sont inconnus. On modélise donc le problème par une loi uniforme $\mathcal{U}[a,b]$ dont la densité est $f_{(a,b)} = \dfrac{1}{b-a}\mathbb{I}_{[a,b]}$ et on va chercher un estimateur de $\theta = (a,b)$ par la méthode du maximum de vraisemblance. La vraisemblance de l'échantillon $x_1, \cdots, x_n$ est donc $L_n(x_1, \cdots, x_n, \theta) = \prod_{i=1}^{n} f_\theta(x_i) = \left(\dfrac{1}{b-a}\right)^n \mathbb{I}_{[a,b]}(x_i)$ si tous les $x_i \notin [a,b]$ et vaut 0 si un des $x_i \in [a,b]$. On voit donc que $L_n(x_1, \cdots, x_n, \theta)$ est maximal si

$$\theta = \theta^* = (a^*, b^*) = (Min\{x_1, \cdots, x_n\}, Max\{x_1, \cdots, x_n\}),$$

puisque ceci nous donne la plus petite valeur de $(b-a)$ sans annuler la vraisemblance. Ceci nous conduit à considérer l'estimateur

$$\widehat{\theta} \;=\; (\widehat{a}, \widehat{b}) = (Min\{x_1, \cdots, x_n\}, Max\{x_1, \cdots, x_n\}).$$

Il reste à montrer que si $X_1, \cdots, X_n$ est un n-échantillon de loi $\mathcal{U}[a,b]$, alors $Min\{x_1, \cdots, x_n\}$ converge bien, en probabilité, vers (a) et que $Max\{x_1, \cdots, x_n\}$ converge en probabilité vers (b). Considérons par exemple le cas de $Min\{x_1, \cdots, x_n\}$. On a

$$\{a + \varepsilon < Min\{x_1, \cdots, x_n\}\} = \{a + \varepsilon < X_1, \cdots, a + \varepsilon < X_n\},$$

d'où, comme les X_i sont indépendants,

$$\begin{aligned}
\mathbb{P}(\{Min\{x_1, \cdots, x_n\}\}) &= \mathbb{P}(\{a + \varepsilon < X_1, \cap \cdots \cap \{a + \varepsilon < X_n\}) \\
&= \mathbb{P}(\{a + \varepsilon < X_1\}) \cdots \mathbb{P}(\{a + \varepsilon < X_n\}) \\
&= \left(\frac{b - a - \varepsilon}{b - a}\right)^n
\end{aligned}$$

qui tend bien vers 0 lorsque n tend vers $+\infty$. On montrerais de même que $Max\{x_1, \cdots, x_n\}$ converge en probabilité vers b.

Distribution normal $\mathcal{N}(\mu, \sigma^2)$.

On suppose à présent que l'échantillon $x_1, \cdots, x_n$ est tiré de manière normale avec une espérance μ et un écart-type σ, mais μ et σ sont inconnus. On modélise donc le problème par une loi normale $\mathcal{N}(\mu, \sigma^2)$ dont la densité est $f_{(\mu,\sigma)}(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$ et on va chercher un estimateur de $\theta = (\mu, \sigma)$ par la méthode du maximum de vraisemblance. La vraisemblance de l'échantillon $x_1, \cdots, x_n$ est donc

$$\begin{aligned} L_n(x_1, \cdots, x_n, \theta) &= \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\dfrac{\sum_{i=1}^{n}(x-\mu)^2}{2\sigma^2}}. \end{aligned}$$

Ici, il est une nouvelle fois plus agréable de considérer la log-vraisemblance

$$\begin{aligned} l_n(x_1, \cdots, x_n, \theta) &= \log(L_n(x_1, \cdots, x_n, \theta)) \\ &= -n(\log(\sigma) + \log(\sqrt{2\pi})) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2. \end{aligned}$$

Pour que $\theta^* = (\mu^*, \sigma^*)$ soit un extremum sur $\mathbb{R} \times \mathbb{R}_*^+$ il faut que les deux dérivées

$$\frac{\partial}{\partial\mu} L_n(x_1, \cdots, x_n, \theta) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = \frac{1}{\sigma^2}(s - n\mu)$$

où

$$s = \sum_{i=1}^{n} x_i$$

et

$$\frac{\partial}{\partial\sigma} l_n(x_1, \cdots, x_n, \theta) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2$$

s'annulent pour $\theta = \theta^*$, ce qui implique que

$$\mu = \frac{s}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{et} \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Ceci nous conduit donc à envisager l'estimateur

$$\widehat{\theta} \;=\; (\widehat{\mu},\widehat{\sigma}) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i, \left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 \right)^{\frac{1}{2}} \right).$$

En ce qui concerne la première composante $\hat{\mu}$,nous retrouvons une nouvelle fois la moyenne comme estimateur de l'espérance $\mu = \mathbb{E}(X_i)$, quant-à la seconde composante, nous trouvons

$$\widehat{\sigma^2} \;=\; \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

dont nous verrons qu'il s'agit bien, pour toute loi, d'un estimateur de la variance σ^2.

Proposition 2.2.5 *[5] Sous la condition de permutation de dérivé par rapport a θ et intégral, on a*

$$\mathbb{E}(U(\theta)) \;=\; 0$$

et

$$I(\theta) \;=\; -\mathbb{E}\left[\frac{\partial^2 logL(\cdot,\theta)}{\partial\theta^2}\right]$$

Proposition 2.2.6 *[5] Pour un échantillon $(T_1,\cdots,T_n)$ i.i.d. de loi $\mathbb{P}_\theta$, l'information de Fisher est :*

$$I_n(\theta) \;=\; nI_1(\theta)$$

où $I_1(\theta)$ correspond à l'information de Fisher du modèle "à 1 observation T_1".

Remarque 2 L'information de Fisher est proportionnelle à n la taille de l'èchantillon : plus n augmente et plus l'information du modèle augmente.

Proposition 2.2.7 *[5] Pour un échantillon $(T_1,\cdots,T_n)$ i.i.d. de loi $\mathbb{P}_\theta$, si $\hat{\theta}_n$ est un estimateur sans biais de θ alors*

$$\mathbb{V}(\widehat{\theta}_n) \;\geq\; \frac{1}{nI_1(\theta)}$$

Remarque 3 La quantité $1/I_n(\theta) = 1/(nI_1(\theta))$ est un minorant de la variance d'un estimateur sans biais de θ : c'est la borne de Cramer-Rao.

Definition 2.2.8 *Un estimateur $\hat{\theta}$ de θ sans biais dont la variance est égale à la borne de Cramer-Rao est dit efficace.*

Theorem 2.2.9 *Il existe un estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ tel que (sous des conditions de régularité) :*

- $\widehat{\theta}_n \to \theta$ *p.s lorsque* $n \to +\infty$
- *L'estimateur $\hat{\theta}_n$ est dit asymptotiquement efficace i.e.*

$$\sqrt{n}(\widehat{\theta}_n - \theta) \quad \xrightarrow{\mathcal{L}} \quad \mathcal{N}(0, \frac{1}{I_1(\theta)})$$

Ce résultat est admis. Il permet de construire des intervalles de confiance (asymptotiques) pour θ.

2.2.2 Exemples

Nous allons présenter dans ces exemples quelques lois qui peuvent être utilisées dans le cadre d'une étude sur des données de survie. Bien que n'importe quelle distribution d'une variable aléatoire continue non-négative puisse être utilisée, nous ne parlons ici que des lois les plus usitées, telles que présentées dans [5].

Distribution exponentielle

La fonction de densité d'une loi exponentielle de paramétre $\lambda > 0$ est donnée par

$$f(t, \lambda) \quad = \quad \lambda e^{-\lambda t}, \qquad\qquad t \geq 0$$

Sa fonction de survie est $S(t) = e^{-\lambda t}$ et sa fontion de taux de hasard est $h(t) = \lambda$, une constante indépendante de t.

Soit $X_1, \cdots, X_n \overset{iid}{\sim} \varepsilon(\lambda)$, calculons l'estimateur du maximum de vraisemblance pour cette loi. La vraisemblance est donnée par

$$L(\lambda) \quad = \quad \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}.$$

En prenant le logarithme, nous obtenons

$$l(\lambda) \quad = \quad ln(L(\lambda)) = nln\lambda - \lambda \sum_{i=1}^{n} x_i = nln\lambda - n\lambda\overline{x}$$

Nous cherchons la valeur $\widehat{\lambda}$ qui maximise cette expression. Nous trouvons

$$\frac{\partial l(\lambda)}{\partial \lambda} \quad = \quad 0 \Leftrightarrow n\left(\frac{1}{\lambda} - \overline{x}\right) = 0 \Leftrightarrow \widehat{\lambda} = \frac{1}{\overline{x}}$$

La loi exponentielle a été largement utilisée dans les premiers travaux de fiabilité, par exemple, de composantes électroniques, mais a eu une extension limitée dans le domaine médical. Ceci est principalement du au fait que cette distribution ne posséde qu'un seul paramètre.

Distribution gamma

La loi gamma comporte deux paramètres, $\lambda > 0$ et $\gamma > 0$. λ est appelé paramétre d'echelle et γ est le paramétre de forme. La fonction de densité de cette loi est donnée par

$$f(t, \lambda, \gamma) \quad = \quad \frac{\lambda}{\Gamma(\gamma)}(\lambda t)^{\gamma-1}e^{-\lambda t}, \qquad t > 0,$$

où $\Gamma(\gamma) = \displaystyle\int_0^\infty x^{\gamma-1}e^{-x}dx$ est la fonction Gamma. La fonction de survie s'exprime comme

$$S(t) \quad = \quad \int_t^\infty \frac{\lambda}{\Gamma(\gamma)}(\lambda x)^{\gamma-1}e^{-\lambda x}dx.$$

En choisissant le paramétre γ entier, nous obtenons la distribution dite de Erlang. Pour celle-ci, nous obtenons comme fonction de taux de hasard La figure 2.1 montre comment varie

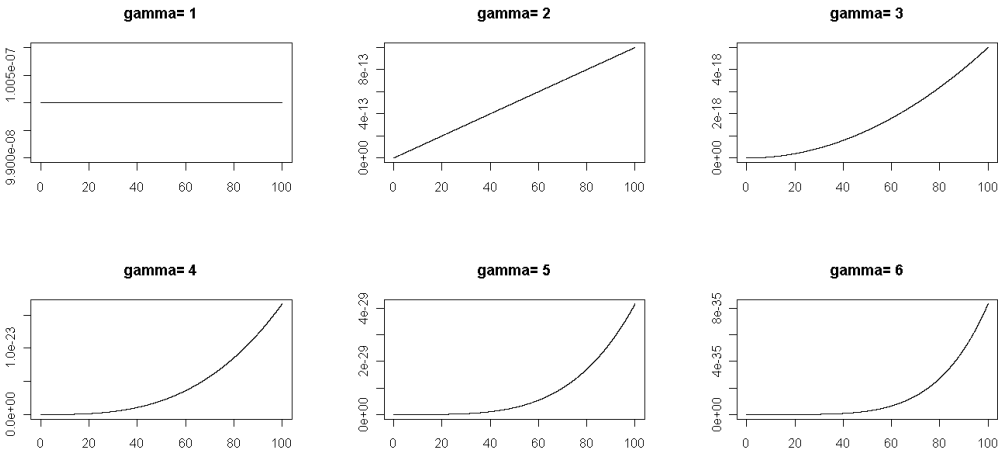

FIGURE 2.1 – Fonction de taux de hasard d'une loi $gamma(\gamma, 10^{-7})$ pour différentes valeurs de γ, $\gamma \in (\{1, 2, \cdots, 6\})$

.

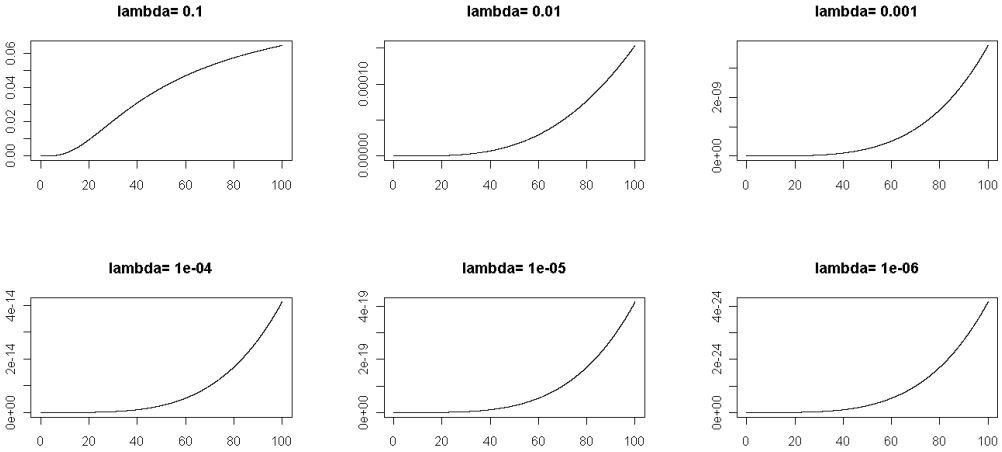

FIGURE 2.2 – Fonction de taux de hasard d'une loi de $gamma(5, \lambda)$ pour différentes valeurs de λ ($\lambda \in \{10^{-1}, 10^{-2}, \cdots, 10^{-6}\}$)

.

la fonction de taux de hasard d'une loi gamma en fonction du paramétre d'échelle λ.

$$h(t) = \frac{\lambda(\lambda t)^{\gamma-1}}{(\gamma-1)!\sum_{k=0}^{\gamma-1}\frac{1}{k!}(\lambda t)^k}.$$

Le logarithme de la vraisemblance d'un échantillon issu d'une loi gamma est donné par

$$l(\lambda,\gamma) = n\lambda\log\lambda - n\log\Gamma(\gamma) + (\gamma-1)\sum_{i=1}^{n} lnt_i - n\lambda\bar{t}.$$

En dérivant par rapport à λ et en appelant $\widehat{\gamma}$ l'estimateur du maximum de vraisemblance pour γ, nous obtenons

$$\widehat{\lambda} = \frac{\widehat{\gamma}}{t}$$

Par contre, le calcul exact de $\widehat{\gamma}$ n'est pas possible, ainsi, nous pouvons seulement exprimer un estimateur en fonction de l'autre.

Distribution de Weibull

La distribution de Weibull est également une généralisation de l'exponentielle. Elle est caractérisée par deux paramétres, $\gamma > 0$ et $\lambda > 0$, qui sont les paramétres de forme et d'échelle respectivement. La fonction de densité d'une telle loi est donnée par

$$f(t;\gamma,\lambda) = \lambda\gamma(\lambda t)^{\gamma-1}e^{(-\lambda t)^\gamma}, \quad t > 0.$$

Ainsi nous remarquons que si $\gamma = 1$ nous obtenons l'exponentielle. La fonction de survie est $S(t) = e^{(-\lambda t)^\gamma}$ et la fonction de taux de hasard vaut $h(t) = \lambda\gamma(\lambda t)^{\gamma-1}$ qui est donc de l'ordre de $t^{\gamma-1}$. Ceci se voit sur la figure (2.3), qui représente la fonction de taux de hasard lorsque le paramètre γ varie.
Nous constatons également sur la figure (2.4) comment cette fonction varie lorsque le paramètre d'échelle λ change.

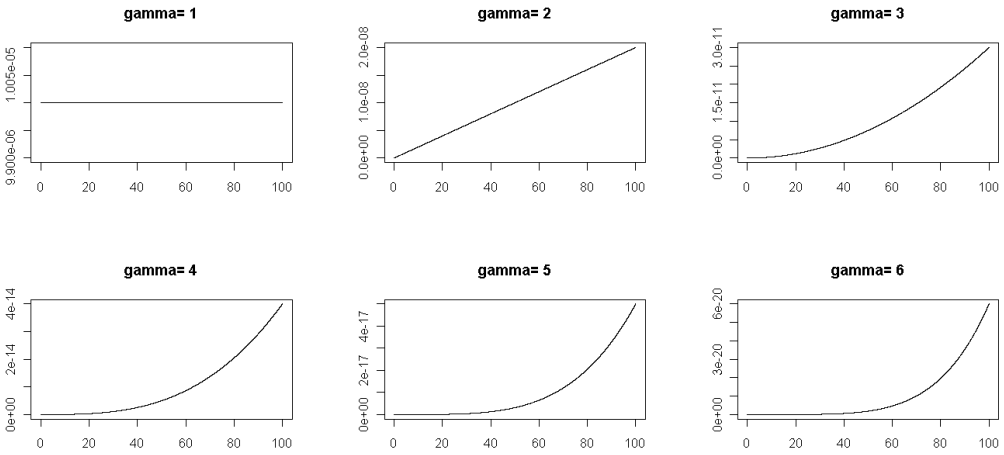

FIGURE 2.3 – Fonction de taux de hasard d'une loi de $Weibull(\gamma, 10^{-5})$ pour différentes valeurs de γ ($\gamma \in \{1, 2, \cdots, 6\}$)

.

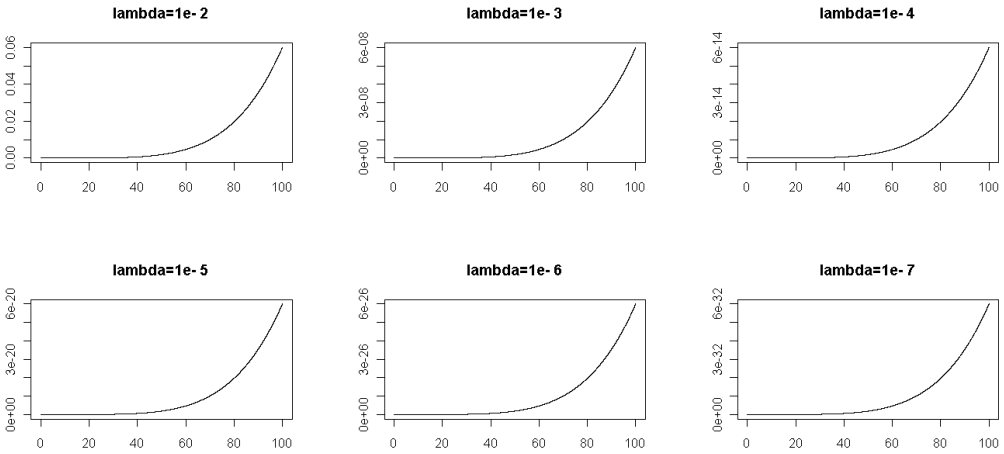

FIGURE 2.4 – Fonction de taux de hasard d'une loi de $Weibull(2, \lambda)$ pour différentes valeurs de λ, ($\lambda \in \{10^{-2}, 10^{-3}, \cdots, 10^{-7}\}$)

.

La fonction de risque est monotone croissante si $\gamma > 1$, monotone d'écroissante si $\gamma < 1$ et constante pour $\gamma = 1$, c'est pourquoi ce paramètre est appelé paramètre de forme. Comme λ est un paramètre d'échelle, différentes valeurs de λ changent seulement l'échelle sur l'axe horizontal, et non pas la forme de base du graphe. Le modèle est assez flexible, et on a montré qu'il constitue une bonne description de plusieurs types de données de survie. Le fait que les fonctions de densité, de survie et de risque aient une forme relativement simple explique également la popularité du modèle.

Distribution log-normale

Si le temps de survie T est tel que $ln(T)$ suit une loi normale avec moyenne μ et variance σ^2, alors on dit que T suit une distribution log-normale. Sa fonction de densité est donnée par

$$f(t;\mu,\sigma^2) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(\log t - \mu)^2\right\},$$

où μ est le paramètre d'échelle et σ est le paramètre de forme. Contrairement à la loi normale, les paramètres ne donnent pas la moyenne et la variance de la loi. En posant $a = e^{-\mu}$, alors $-\mu = \log a$ et nous obtenons

$$f(t;a,\sigma^2) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(lnat)^2\right\}.$$

La fonction de survie d'une variable suivant une loi log-normale est donnée par

$$S(t) = 1 - \Phi\left(\log\left(\frac{at}{\sigma}\right)\right),$$

où $\Phi(y) = \dfrac{1}{\sqrt{2\sqrt{\pi}}} \displaystyle\int_{-\infty}^{y} e^{-u^2/2} du$ est la fonction de distribution d'une loi normale standard centrée réduite. La fonction de taux de hasard est de la forme

$$h(t) = \frac{\dfrac{1}{t\sigma\sqrt{2\pi}} exp\left(-\dfrac{(lnt-\mu)^2}{2\sigma^2}\right)}{1 - \Phi\left(\log\left(\dfrac{at}{\sigma}\right)\right)}$$

Nous pouvons montrer que $h(t) = 0$ pour $t = 0$, que $h(t)$ croît jusqu'à un maximum et ensuite décroît et tend vers 0 lorsque $t \to \infty$. Comme la fonction de risque décroît pour de grandes

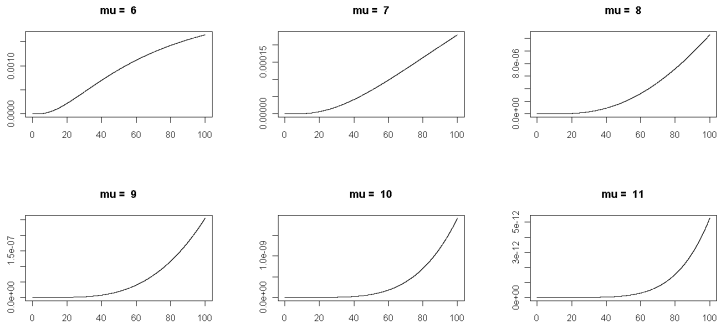

FIGURE 2.5 – Fonction de taux de hasard d'une loi log-normale $(\mu, 1)$ pour différentes valeurs de μ, $(\mu \in \{6, 7, \cdots, 11\})$

.

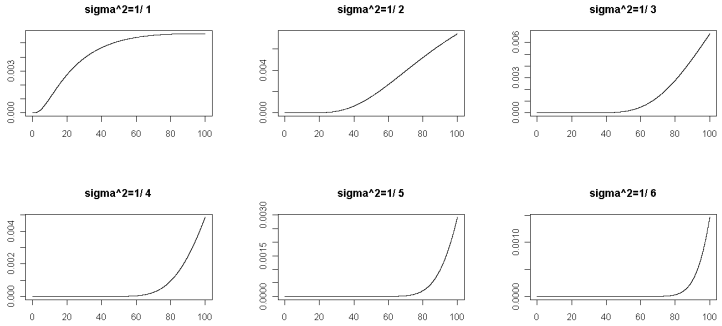

FIGURE 2.6 – Fonction de taux de hasard d'une loi log-normale $(5, \sigma^2)$ pour différentes valeurs de σ^2 $(\sigma^2 \in \{\frac{1}{2}, \cdots, \frac{1}{6}\})$

.

valeurs de t, la distribution ne paraît pas plausible comme modèle de vie dans la plupart des situations. Malgré cela, ce modèle peut être intéressant lorsque de très grandes valeurs de t ne sont pas d'un intérêt particulier. Les figures (2.5) et (2.6) montrent comment se comporte la fonction de taux de hasard en fonction des paramètres μ et σ^2 respectivement. Pour la loi log-normale, les estimateurs de maximum de vraisemblance peuvent être calculés facilement. Nous obtenons,

$$\widehat{\mu} \;=\; \frac{1}{n}\sum_{i=1}^{n} ln(t_i)$$

et

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(lnt_i - \widehat{\mu})^2.$$

La distribution log-logistique

La distribution des temps d'événement est log-logistique si la densité de x est :

$$f(t) = \frac{\alpha\gamma t^{\gamma-1}}{(1+\alpha t^{\gamma})^2}.$$

Les fonctions de survie et de risque associées sont respectivement :

$$S(t) = \frac{1}{(1+\alpha t^{\gamma})}$$

$$h(t) = \frac{\alpha\gamma t^{\gamma-1}}{(1+\alpha t^{\gamma})}.$$

On rappelle que si X est une variable aléatoire log-logistique, alors $Y = \log X$ a une distribution logistique. Des exemples de fonctions de survie obtenues pour diverses valeurs du paramètre de forme γ sont présentés dans le graphique (2.7). Rappelez-vous également que des évolutions similaires de la fonction de risque peuvent être obtenues avec la distribution log-normale.

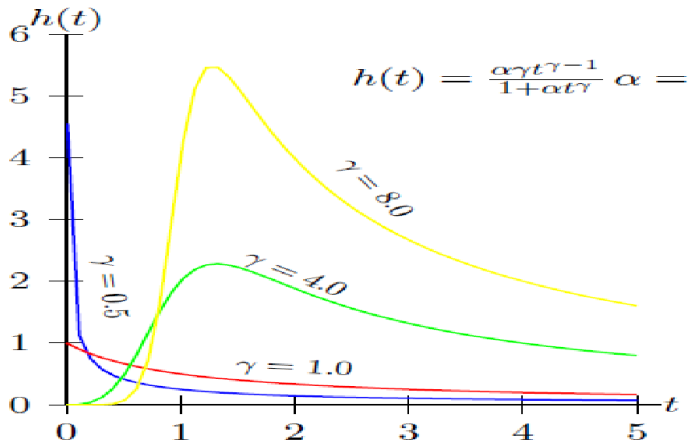

FIGURE 2.7 – Exemples de fonctions de risque sous distribution Loglogistique

.

2.3 Maximum de vraisemblance (Données censurées)

Nous rappelons brièvement la méthode du maximum de vraisemblance pour estimer, au vu de données censurées, les paramétres réels d'un modèle d'analyse des durées de vie.

Nous considérons le cas de la censure aléatoire à droite. Les observations sont les couples $(Z_1, \delta_1), \cdots, (Z_n, \delta_n)$ où

$$Z_i = \min(T_i, C_i) \quad et \quad \delta_i = \mathbb{I}_{(T_i \leq C_i)}.$$

- Si $\delta_i = 1$ alors $Z_i = T_i$: on observe la durée de vie.
- Si $\delta_i = 0$ alors $Z_i = C_i$: il y a censure.

Hypothèse Fondamentale : On suppose que le délai de censure C_i de l'individu i est une variable aléatoire indépendante de la durée de vie T_i .

Proposition 2.3.1 *Sous l'hypothése fondamentale d'indépendance $T_i \coprod C_i$, pour $i = 1, \cdots, n$ la vraisemblance s'écrit :*

$$L((z_1, \delta_1), \cdots, (z_n, \delta_n), \theta) \quad = \quad \prod_{i=1}^{n} f_\theta(z_i)^{\delta_i} S_\theta(z_i)^{1-\delta_i}.$$

où f_θ est la densité commune des T_i et S_θ la fonction de survie associée.

Démonstration :
Soit la suite des délais de censure $C_1, \cdots, C_n$ i.i.d. de densité commune g et G la survie associée

i.e. $G(c) = \mathbb{P}(C_1 > c)$.
Le couple de v.a. (Z_i, Δ_i) admet pour densité :

$$\begin{cases} g(z_i)S_\theta(z_i), & \text{si } \delta_i = 0 \quad (observations \ \ censures); \\ f_\theta(z_i)G(z_i), & \text{si } \delta_i = 1 \quad (dures \ \ t_i = z_i \ \ observes) \end{cases}$$

que l'on peut aussi écrire de façon équivalente :

$$[f_\theta(z_i)G(z_i)]^{\delta_i}[g(z_i)S_\theta(z_i)]^{1-\delta_i}.$$

de plus les couples $(Z_1, \triangle_1), \cdots, (Z_n, \triangle_n)$ sont indépendants donc la vraisemblance des observations s'écrit :

$$\prod_{i=1}^{n}[f_\theta(z_i)G(z_i)]^{\delta_i}[g(z_i)S_\theta(z_i)]^{1-\delta_i}.$$

Comme la loi des C_i ne fait pas intervenir le paramètre θ, la partie utile de la vraisemblance se réduit à :

$$L(\theta) \ = \ \prod_{i=1}^{n} f_\theta(z_i)^{\delta_i} S_\theta(z_i)^{1-\delta_i}.$$

Ecriture de la log-vraisemblance en fonction du risque :

Proposition 2.3.2 *Sous l'hypothèse fondamentale d'indépendance $T_i \coprod C_i$, pour $i = 1, \cdots, n$*

$$\log L(\theta) \ = \ \sum_{i=1}^{n} \delta_i \log h_\theta(z_i) + \sum_{i=1}^{n} \log S_\theta(z_i).$$

où $h_\theta(\cdot)$ désigne la fonction de risque instantané.
En effet :

$$\begin{aligned} logL(\theta) \ &= \ \log[\prod_{i=1}^{n} f_\theta(z_i)^{\delta_i} S_\theta(z_i)^{1-\delta_i}] \\ &= \ \sum_{i=1}^{n} \log[f_\theta(z_i)^{\delta_i} S_\theta(z_i)^{1-\delta_i}] \\ &= \ \sum_{i=1}^{n} \delta_i \log[f_\theta(z_i) + 1 - \delta_i \sum_{i=1}^{n} S_\theta(z_i)] \\ &= \ \sum_{i=1}^{n} \delta_i \log h_\theta(z_i) + \sum_{i=1}^{n} logS_\theta(z_i). \end{aligned}$$

Notation : On note

$$L(\theta) \quad = \quad L((z_1, \delta_1), \cdots, (z_n, \delta_n), \theta).$$

De façon analogue au cas non censuré, on définit l'EMV $\widehat{\theta}_n$ *de* θ *et on peut montrer sous certaines hypothèses :*

Theorem 2.3.3 ([2]) *L'EMV* $\widehat{\theta}_n$ *suit approximativement une loi normale de moyenne* θ *et de variance* $nI_1(\theta)$ *;* $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \frac{1}{I_1(\theta)})$*. On peut généraliser ce résultat au cas où le paramètre* θ *est un vecteur de* R^p*. On a alors une matrice* $p \times p$ *de variance-covariance* $nI_1(\theta)$*.*

Exemple1 : loi exponentielle

On a $f(t) = \lambda \exp(-\lambda t), S(t) = \exp(-\lambda t)$.

Nous avons

$$\begin{aligned} \log L(\theta) &= \sum_{i=1}^{n} \delta_i \log h_\theta(z_i) + \sum_{i=1}^{n} logS_\theta(z_i) \\ &= \sum_{i=1}^{n} \delta_i log\lambda - \lambda \sum_{i=1}^{n} \delta_i z_i - \lambda \sum_{i=1}^{n} (1 - \delta_i) z_i \\ &= (\sum_{i=1}^{n} \delta_i) log\lambda - \lambda \sum_{i=1}^{n} z_i. \end{aligned}$$

On en conclue que $\widehat{\lambda} = \dfrac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} z_i}$ (au lieu de $\widehat{\lambda} = \dfrac{1}{\sum_{i=1}^{n} z_i}$ pour les données sans censure)).

Exemple2 : Loi weibull

On a vu dans la partie précédente l'estimation des paramètres du modèle de Weibull dans le cas non censuré. On traite maintenant à titre d'exemple le cas d'une censure droite. On considère donc le modèle suivant : $f(t) = \gamma \lambda^\gamma t^{\gamma-1} e^{-(\lambda t)^\gamma}$ et $S(t) = e^{(-\lambda t)^\gamma}$.

La vraisemblance de ce modèle s'écrit :

$$L(\theta) = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}.$$

D'où l'on déduit la log-vraisemblance

$$\log L(\theta) = \gamma \log \lambda \sum_{i=1}^{n} \delta_i + \log \gamma \sum_{i=1}^{n} \delta_i + (\gamma - 1) \sum_{i=1}^{n} \delta_i \log t_i - \sum_{i=1}^{n} (\lambda t_i)^{\gamma}.$$

Les équations aux dérivés partielles s'écrivent donc :

$$\begin{aligned}
\frac{\partial \log L(\theta)}{\partial \lambda} &= \frac{\gamma \sum_{i=1}^{n} \delta_i}{\lambda} - \lambda^{\gamma-1} \gamma \sum_{i=1}^{n} t_i^{\gamma} \\
\frac{\partial \log L(\theta)}{\partial \gamma} &= \sum_{i=1}^{n} \delta_i (\frac{1}{\gamma} - \log \lambda) + \lambda^{\gamma} \left[\log \lambda \sum_{i=1}^{n} t_i^{\gamma} - \sum_{i=1}^{n} t_i^{\gamma} \log t_i \right] + \sum_{i=1}^{n} \delta_i \log t_i.
\end{aligned}$$

On cherche donc les solutions du système suivant :

$$\begin{aligned}
\lambda &= \left(\frac{1}{\sum_{i=1}^{n} t_i^{\gamma}} \sum_{i=1}^{n} \delta_i \right)^{1/\gamma} \\
\frac{1}{\gamma} &= \frac{\sum_{i=1}^{n} t_i^{\gamma} \log t_i}{\sum_{i=1}^{n} t_i^{\gamma}} - \frac{1}{\sum_{i=1}^{n} t_i^{\gamma}} \sum_{i=1}^{n} \log t_i.
\end{aligned}$$

Chapitre 3

Tests de comparaison

En statistiques, un test d'hypothèse est une démarche consistant à rejeter ou à ne pas rejeter une hypothèse statistique, appelée hypothèse nulle, en fonction d'un jeu de données (échantillon). Il s'agit de statistique inférentielle : à partir de calculs réalisés sur des données observées, nous émettons des conclusions sur la population, en leur rattachant des risques de se tromper. Pour tester l'hypothèse nulle dégalité des survies dans les deux groupes, on dispose de trois tests asymptotiquement équivalents :

1. le test de Wald
2. le test du rapport de vraisemblance
3. le test de Rao ou test du score

3.1 Comparaison de deux groupes dans un modèle exponentiel

Pourquoi le modèle exponentiel ? Ce modèle suppose que la fonction de risque instantané $h(t)$ est une constante indépendante du temps. Son avantage est l'existence de solutions explicites au maximum de vraisemblance : on dispose alors d'estimateurs et de tests faciles à calculer, ce qui permet une première approche des données.
Nous allons voir comment comparer deux échantillons exponentiels.
Cadre :
Soient A et B deux groupes d'individus dont on veut comparer la survie. On suppose que les durées de vie T_A et T_B suivent une loi exponentielle dans chacun des deux groupes.
Les densités de probabilités de T_A et T_B s'écrivent :

$$f_A(t) = h_A \exp(-h_A t) \quad et \quad f_B(t) = h_B \exp(-h_B t)$$

où h_A et h_B sont les risques instantanés de décès dans les deux groupes supposés constants au cours du temps.
On note $\exp(b)$ le rapport des risques instantanés, donnant le risque relatif du groupe B par rapport au groupe A :

$$h_B = h_A \exp(b)$$

Pour comparer les deux groupes (c.à.d. les survies S_A et S_B)), il faut estimer b et tester l'hypothèse nulle :

$$H_0 : h_A = h_B \quad \text{ou de façon équivalente} \quad H_0 : b = 0.$$

Soient n_A et n_B, le nombre d'individus dans chacun des groupes ($n = n_A + n_B$). Dans le groupe A, on observe $(Z_{A,i}, \delta_{A,i})_{i=1}^{n_A}$, et dans le groupe B, on observe $(Z_{B,i}, \delta_{B,i})_{i=1}^{n_B}$.
La vraisemblance des observations s'écrit :

$$\begin{aligned} L(h_A, h_B) &= L(h_A)L(h_B) \\ &= \prod_{i=1}^{n_A} f_A^{\delta_i} S_A^{1-\delta_i} \prod_{i=1}^{n_B} f_B^{\delta_i} S_B^{1-\delta_i} \end{aligned}$$

et la log-vraisemblance :

$$\log L(h_A, h_B) = \log(h_A) \sum_{i=1}^{n_A} \delta_{A,i} - h_A \sum_{i=1}^{n_A} z_{A,i} + log(h_B) \sum_{i=1}^{n_B} \delta_{B,i} - h_B \sum_{i=1}^{n_B} z_{B,i}$$

talle que $r_A = \sum_{i=1}^{n_A} \delta_{A,i}$ et $r_B = \sum_{i=1}^{n_B} \delta_{B,i}$

On reparamètre la log-vraisemblance en remplaçant (h_A, h_B) par (h_A, b) où $b = \log(h_B/h_A)$. La log-vraisemblance s'écrit alors :

$$r \log(h_A) + b r_B - h_A \left[\sum_{i=1}^{n_A} z_{A,i} + \exp(b) \sum_{i=1}^{n_B} z_{B,i} \right]$$

où $r = r_A + r_B$ représente le nombre total d'individus non censurés (ou le nombre de décès observés).

On calcule le vecteur de score (dérivées partielles de $\log L$) :

$$U(h_A, b) \quad = \quad \begin{pmatrix} \partial \log L/ & \partial h_A \\ \partial \log L/ & \partial b \end{pmatrix}$$

Les estimateurs du maximum de vraisemblance $\widehat{h}_A$ de h_A et $\widehat{b}$ de b sont les solutions du système d'équations $U(h_A, b) = \overrightarrow{0}$. On obtient :

$$\widehat{h}_A = \frac{r_A}{\sum_{i=1}^{n_A} z_{A,i}} \quad et \quad \widehat{\exp(b)} = \frac{r_B / \sum_{i=1}^{n_B} z_{B,i}}{r_A / \sum_{i=1}^{n_A} z_{A,i}}.$$

Remarque 4 On retrouve bien $\widehat{h}_B = \widehat{h}_A \widehat{exp(b)} = r_A / \sum_{i=1}^{n_A} z_{A,i}$

3.2 Le test de Wald

Proposition 3.2.1 *Sous H_0 : $b = 0$, la loi de l'EMV $\widehat{b}$ est asymptotiquement normale de moyenne nulle et*

$$\widehat{V}(\widehat{b}) \quad = \quad \frac{r}{r_A r_B}$$

(c'est le terme ($U(h_A, b)$) de l'inverse de la matrice d'Information de Fisher $I^{-1}(\widehat{h}_A, \widehat{b})$.)
Il en découle

Proposition 3.2.2 *[2] La statistique du test de Wald pour comparer b à 0 est $\chi^2_W = \frac{r_A r_B}{r} \widehat{b}^2$ suit asymptotique une loi $\chi^2(1)$ sous H_0.*

3.3 Test du rapport de vraisemblance

Proposition 3.3.1 *[2] La statistique du logarithme du rapport de vraisemblance est définie par :*

$$\chi^2_L \quad = \quad 2 \log \left(\frac{L(\widehat{h}_A, \widehat{b})}{L(\widehat{h}, 0)} \right)$$

où $L(\widehat{h}_A, \widehat{b})$ est la vraisemblance maximisée c.à.d calculée sans restriction en :

$$h_A = \widehat{h}_A \quad et \quad b = \widehat{b}$$

et $L(\widehat{h}, 0)$ est la vraisemblance restreinte sous H_0, maximisée en :

$$h \;=\; \widehat{h} = \frac{r}{\sum_{i=1}^{n_A} z_{A,i} + \sum_{i=1}^{n_B} z_{B,i}} \quad et \quad b = 0$$

Proposition 3.3.2 *[2] La statistique du rapport de vraisemblance s'écrit :*

$$\chi^2_L \;=\; 2\left[r_A \log\left(\frac{r_A}{\sum_{i=1}^{n_A} z_{A,i}} \right) + r_B \log\left(\frac{r_B}{\sum_{i=1}^{n_B} z_{B,i}} \right) - r \log\left(\frac{r}{\sum_{i=1}^{n_A} z_{A,i} + \sum_{i=1}^{n_B} z_{B,i}} \right) \right]$$

Sous $H_O : b = 0$, χ^2_L suit asymptotiquement une loi $\chi^2(1)$.

3.4 Test de Rao ou test du score

Proposition 3.4.1 *[2] La statistique de test du score est donnée par :*

$$\chi^2_S \;=\; \frac{r_A r_B}{r} \frac{(\exp \widehat{b} - 1)^2}{\exp \widehat{b}}$$

Sous $H_O : b = 0$, χ^2_S suit asymptotiquement une loi $\chi^2(1)$.

3.5 La pratique des procédures de tests

- Les procédures de tests conduisent à rejeter $H_0 : b = 0$ pour des valeurs élevées de la statistique de test, c. à. d. pour des valeurs supérieures au quantile d'ordre $1-\alpha$ de $\chi^2(1)$ pour un risque de première espèce α.
- En pratique, le test de Wald peut donner parfois des résultats assez différents du test du score ou du rapport de vraisemblance qui sont assez proches en général.
- Le test du rapport de vraisemblance est le plus robuste et donc le plus fiable des trois.

Annexe

Voici le code utilisé avec le programme R pour produire les figures présentées dans ce mémoire. Nous avons tout d'abord représenté les fonctions de taux de hasard des différentes distributions paramétriques (figures 2.1 à 2.6).

♯*Gamma*♯

```
tt <- seq(0,100,by=0.01)
k <- 5
split.screen(c(2,3))
for (i in c(1:6)) {
screen(i)
lambda[i] <- 10^ (-i)
plot(tt,dgamma(tt,k,lambda[i])/(1-pgamma(tt,k,lambda[i])),
col=1,type="l",xlim=range(tt),xlab="",ylab="",
main=paste("lambda=",lambda[i]))
}
lambda <- 1e-7
split.screen(c(2,3))
for (i in c(1:6)) {
screen(i)
plot(tt,dgamma(tt,i,lambda)/(1-pgamma(tt,i,lambda)),col=1,
"l",main=paste("gamma=",i),xlab="",ylab="")
}
```

♯*Weibull*♯

```
gamma <- 6
lambda <- 0
split.screen(c(2,3))
for (i in c(1:6)) {
screen(i)
lambda[i] <- 10^ (i+1)
```

```
plot(tt,dweibull(tt,gamma,lambda[i])/(1-pweibull(tt,gamma,
lambda[i])),col=1,type="l",xlim=range(tt),xlab="",
ylab="",main=paste("lambda=1e-",i+1))
}
lambda <- 1e5
gam <- 0
split.screen(c(2,3))
for (i in c(1:6))
screen(i)
gam[i] <- i
plot(tt,dweibull(tt,gam[i],lambda)/(1-pweibull(tt,gam[i],
lambda)),col=1,type="l",xlim=range(tt),xlab="",
ylab="",main=paste("gamma=",gam[i]))
}
```

$\sharp Log - Normale\sharp$

```
mu <- 5
sigma <- 0
split.screen(c(2,3))
for (i in c(1:6)) {
screen(i)
sigma[i] <- 1/i
plot(tt,dlnorm(tt,mu,sigma[i])/(1-plnorm(tt,mu,sigma[i])),
col=1,type="l",xlim=range(tt),xlab="",ylab="",
main=paste("sigma2=1/",i))
}
mu <- 0
sigma <- 1
split.screen(c(2,3))
for (i in c(1:6))
screen(i)
mu[i] <- i+5
plot(tt,dlnorm(tt,mu[i],sigma)/(1-plnorm(tt,mu[i],sigma)),
col=1,type="l",xlim=range(tt),xlab="",ylab="",
main=paste("mu = ",i+5))
}
```

Bibliographie

[1] Elodie BRUNEL. Fiabilité et Survie. 2010. Licence MASS. FLMA 610.

[2] Gilbert Colletaz. 2012. Modèles de survie. Notes de cours. Master 2 ESA.

[3] Olivier Gaudoin. 2008. Statistique inférentielle Avancée.

[4] Philipe Saint Pierre. 2013. Introduction à l'analyse des durée de survie.

[5] Vuistiner Philippe. 2008. Analyse de données de survie.