

**Université de Rennes I**

- UFR Institut Mathématique de Rennes I -

**D.E.A. de Mathématiques Fondamentales et Applications  
Option Statistiques**

année 1998-1999

**Analyse de fiabilité à partir de données censurées à  
droite et à gauche**

**Franck CORSET**

---

Mémoire de Stage de DEA

Réalisé au sein du projet IS2 - Inria - et la Direction des Études et  
Recherches d'Électricité De France sous la direction de Gilles CELEUX



## Remerciements

Je tiens à exprimer ma sincère gratitude à Gilles CELEUX pour m'avoir permis de réaliser ce stage de DEA au sein de son équipe, mais aussi pour avoir dirigé ce travail. Je remercie également toute l'équipe IS2 pour leur soutien et leur agréable compagnie, sans oublier les autres projets, avec qui j'ai pu m'entretenir d'une manière très profitable.

Enfin, j'adresse mes vifs remerciements à Marie-Agnès GARNERO, qui a suivi mon travail à Électricité de France, et qui par son entière disposition, m'a permis de parfaire celui-ci.

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>Les données</b>	<b>6</b>
2.1	Description . . . . .	6
2.2	Méthodologie . . . . .	7
<b>3</b>	<b>Présentation des modèles statistiques</b>	<b>8</b>
3.1	Définitions . . . . .	8
3.2	Rappel : Le cas de données censurées à droite . . . . .	9
3.3	Autres possibilités, méthodes actuariales . . . . .	10
3.4	Fiabilité dans un contexte doublement censuré . . . . .	11
3.5	Approche bayésienne . . . . .	11
3.6	Estimation non paramétrique de la fonction de survie . . . . .	13
3.6.1	L'estimateur empirique . . . . .	13
3.6.2	L'estimateur de Turnbull . . . . .	14
<b>4</b>	<b>Propriétés des estimateurs</b>	<b>16</b>
4.1	Le cas censurés à droite . . . . .	16
4.1.1	Modélisation par une loi exponentielle . . . . .	16
4.1.2	Modélisation par une loi de Weibull à deux paramètres . . . . .	17
4.2	Le cas doublement censuré . . . . .	18
4.2.1	Consistance de l'estimateur du maximum de vraisemblance . . . . .	18
4.2.2	Normalité asymptotique de l'estimateur du maximum de vraisemblance . . . . .	19
4.2.3	Modélisation par une loi exponentielle . . . . .	20
4.2.4	Modélisation par une loi de Weibull à deux paramètres . . . . .	24
4.3	Estimation dans un cadre bayésien . . . . .	27
4.4	Simulation de Monte-Carlo . . . . .	27
4.4.1	Loi a priori informative . . . . .	28
4.4.2	Avec un a priori non informatif . . . . .	31
4.4.3	Avec un a priori "semi informatif" . . . . .	32
<b>5</b>	<b>Applications et simulations</b>	<b>35</b>
5.1	Simulations . . . . .	35
5.1.1	Estimateurs du maximum de vraisemblance . . . . .	35
5.1.2	Estimateurs bayésiens . . . . .	42
5.1.3	Estimateurs non paramétrique . . . . .	44
5.2	Applications . . . . .	46
5.2.1	Estimateurs du maximum de vraisemblance . . . . .	46
5.2.2	Estimateurs bayésiens . . . . .	50
5.2.3	Estimateurs non paramétriques . . . . .	52
<b>6</b>	<b>Conclusions et perspectives</b>	<b>53</b>



# 1 Introduction

Le département Surveillance, Diagnostique, Maintenance (SDM) de la Direction des Études et Recherches d'Électricité de France s'est engagé dans une nouvelle politique de maintenance sur les pompes primaires 900 MW de tout le parc nucléaire français. La fiabilité de ces composants étant très grande, la politique de maintenance d'EDF consistait jusqu'à présent en des inspections régulières (environ tous les 6 ans), c'est-à-dire en une démarche purement déterministe coûtant cher à l'entreprise.

Cette étude a pour objectif d'introduire des modèles statistiques analysant le vieillissement des composants qui constituent ces pompes, afin de déterminer une politique optimale de maintenance.

Pour ce problème on ne peut appliquer des modèles classiques. En effet on n'observe aucune défaillance, mais uniquement des dégradations survenues avant la date d'inspection (données censurées à gauche), ou aucune dégradation (données censurées à droite).

Ainsi, après une présentation des données, nous introduisons différents modèles paramétriques pour des données censurées à gauche et à droite. L'estimation des paramètres s'est faite par maximum de vraisemblance, donnant lieu à des propriétés asymptotiques, et donc à des intervalles de confiance.

Puis une approche bayésienne a été adoptée, grâce aux avis d'experts nous donnant un a priori sur les paramètres à estimer, mais aussi dans le but de régulariser les estimateurs.

Ensuite, nous avons voulu comparer ces deux méthodes avec une étude non paramétrique, n'imposant aucune contrainte sur la loi du processus.

De plus, nous nous sommes intéressés aux propriétés des estimateurs définis précédemment.

Enfin, nous avons appliqué ces méthodes sur des données réelles, puis sur des données simulées, afin d'étudier la stabilité, et la validité de nos modèles.

## 2 Les données

### 2.1 Description

Les données nous ont été fournies par le département Surveillance, Diagnostic et Maintenance d'EDF. Ce département souhaite modéliser statistiquement le vieillissement des principaux composants qui constituent les pompes primaires. Ces composants sont les suivants :

- \* la glace tournante du joint 2 (noté  $c_1$ ),
- \* la glace flottante du joint 2 (noté  $c_2$ ),
- \* les douilles flottantes du joint 2 (noté  $c_3$ ),
- \* la glace tournante du joint 3 (noté  $c_4$ ),
- \* la glace flottante du joint 3 (noté  $c_5$ ),
- \* les douilles flottantes du joint 3 (noté  $c_6$ ),
- \* la cartouche (noté  $c_7$ ).

Lors des précédentes inspections sur ces composants, une importante quantité d'information a été enregistrée dans des bases de données (fichiers Excel). Les variables étaient au nombre de 9 et se présentaient sous la forme suivante :

- \* la provenance des pompes primaires (le site de la centrale nucléaire),
- \* la tranche nucléaire (numérotée de 1 à 4 pour la plupart des centrales),
- \* la date d'arrivée sur le site des pompes primaires,
- \* la date d'expertise,
- \* le nombre d'heures de fonctionnement,
- \* le composant,
- \* observations lors de l'expertise (décollement, fissures, griffures, ...),
- \* propositions pour éventuellement remettre en état le composant (pierrer les griffes, adoucir, remplacer,...),
- \* état du composant (accepté ou rebuté).

Il est à noter que pour l'arrêt d'une pompe primaire, tous les composants sont inspectés en même temps. Ainsi, ils auront le même nombre d'heures de fonctionnement, notés  $T$ , sauf dans le cas de valeurs manquantes.

## 2.2 Méthodologie

Nous avons voulu privilégier, durant ce stage, des modèles à deux états sans tenir compte de la gravité de la dégradation. Le composant est donc considéré comme défaillant si son état a été déclaré “rebuté” lors de l’inspection.

Pour modéliser le vieillissement du composant, il nous a fallu considérer des modèles avec des données censurées à gauche et censurées à droite. En effet pour ce type de matériel très fiable, nous n’observons jamais de défaillance durant la visite de maintenance, mais uniquement des dégradations.

Le tableau suivant représente les caractéristiques des composants étudiés. Les différentes tailles d’échantillons s’expliquent par les valeurs manquantes, assez importantes dans certains cas (comme pour le composant 6 et le composant 7).

Composant	Nom du composant	taux de censures à gauche	taille de l’échantillon
c1	Glace tournante du joint 2	40,5 %	n=168
c2	Glace flottante du joint 2	26,4 %	n=72
c3	Douille flottante du joint 2	1,4 %	n=71
c4	Glace tournante du joint 3	59,8 %	n=72
c5	Glace flottante du joint 3	44,6 %	n=74
c6	Douille flottante du joint 3	0 %	n=39
c7	Cartouche	17,8 %	n=28

TAB. 1: Composants et caractéristiques des échantillons

Ainsi, une censure à gauche correspond à une dégradation détectée lors de la visite (c’est-à-dire survenue avant l’inspection). Une censure à droite correspond, quant à elle, à un matériel non dégradé (mais susceptible de se dégrader ultérieurement).

Nous avons alors introduit une variable indicatrice  $\delta$  signalant la présence ou non d’une censure à gauche. le nombre de censures à gauche. Il est important de signaler que l’information est contenue principalement dans ces censures à gauche, et donc que l’estimation sera d’autant meilleure que le nombre de données censurées à gauche sera conséquent.

### 3 Présentation des modèles statistiques

Dans toute la suite  $(\Omega, \mathcal{F}, \mathbb{P})$  désigne un espace probabilisé.

On note  $\mathcal{L}_n$  la vraisemblance du modèle et  $\ell_n$  son logarithme népérien.  $\dot{\ell}_n$  et  $\ddot{\ell}_n$  désignent respectivement la dérivée première et seconde de la log-vraisemblance.

#### 3.1 Définitions

On note  $(X_1, \dots, X_n)$  les durées de vie, que l'on suppose indépendantes et identiquement distribuées, de fonction de répartition  $F$  et de densité  $f$  par rapport à la mesure de Lebesgue, et  $(T_1, \dots, T_n)$  les temps d'observations de distribution  $G$  et de densité  $g$ . On suppose que ces deux variables aléatoires sont indépendantes entre elles.

**Définition 1** On appelle *fiabilité* ou *fonction de survie* la fonction  $R(x)$  telle que :

$$\begin{aligned} R(x) = 1 - F(x) &= P(X > x) \\ &= \text{Proba}(\text{le composant ait fonctionné sur } [0, x]). \end{aligned}$$

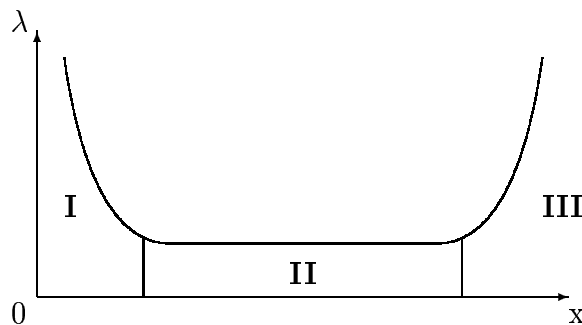
**Définition 2** Dans le cas où  $X$  est de loi absolument continue on peut définir le *taux de panne instantané*  $\lambda$  par

$$\lambda(x) = \lim_{h \rightarrow 0^+} \frac{1}{h} P(X \in ]x, x+h] / X > x) = \lim_{h \rightarrow 0^+} \frac{F(x+h) - F(x)}{hR(x)} = \frac{f(x)}{R(x)}$$

et le *taux de panne cumulé*  $\Lambda$  par

$$\Lambda(x) = \int_0^x \lambda(s) ds.$$

Exemple : La courbe en baignoire décrit le comportement classique du taux de panne instantané au cours de la vie d'un matériel.



Dans ce graphe la période **I** correspond à la jeunesse du composant pendant laquelle le taux de panne instantané décroît, **II** correspond à la période de vie utile où  $\lambda$  est constant, et **III** à la période de vieillesse ou d'usure (la probabilité de tomber en panne augmente avec le temps).

### 3.2 Rappel : Le cas de données censurées à droite

Afin de se ramener à notre cas d'étude, nous avons considéré les censures à gauche (dégradation avant le temps d'inspection) comme une défaillance au temps d'observation, notées  $T_i$ . Les censures à droites, quant à elles, restent inchangées et sont dites *censures progressives à droite de type I*, i.e. elles sont déterministes et connues mais peuvent être différentes suivant les observations.

Les variables sont donc les durées de vie notées  $X_i$ , de densité  $f$ , la variable indicatrice  $\delta$  sur la présence d'une dégradation, et les censures (à droite) de densité  $g$ . La fonction de répartition des durées de vie est notée  $F$ , celle des censures est notée  $G$ .

On cherche à modéliser l'apparition d'une dégradation, en supposant que les  $n$  composants observés ont un comportement indépendant.

On observe donc les couples

$$(T_i, \delta_i) \text{ avec } \delta_i = \mathbb{I}_{\{X_i \leq T_i\}}, \text{ pour } i = 1, \dots, n$$

$$\delta_i = \begin{cases} 0 & \text{s'il y a une censure} \\ 1 & \text{s'il y a une panne} \end{cases}$$

On désire obtenir les densités correspondants à l'observation d'une panne et à l'observation d'une censure à droite. Pour cela, on calcule la probabilité de constater une panne :

$$\begin{aligned} P(T \leq t, \delta = 1) &= \int_{x \leq t, x \leq c} f(x)g(c) dx dc \\ &= \int_{x \leq t} f(x) \left( \int_x^\infty g(c) dc \right) dx \\ &= \int_{x \leq t} f(x)\bar{G}(x) dx \end{aligned}$$

pour tout  $t \geq 0$ , et où  $\bar{G} = 1 - G$ . La densité correspondant à l'observation d'une panne est alors  $f(x)\bar{G}(x)$ . De même la probabilité d'obtenir une censure à droite est donnée par :

$$\begin{aligned} P(T \leq t, \delta = 0) &= \int_{c \leq x, x \leq t} f(x)g(c) dx dc \\ &= \int_{c \leq t} g(c) \left( \int_c^\infty f(x) dx \right) dc \\ &= \int_{c \leq t} g(c)R(c) dc \end{aligned}$$

Ainsi, la densité correspondant à l'observation d'une censure est alors  $g(x)R(x)$ . On peut donc déterminer la densité conjointe du couple  $(T, \delta)$  par rapport à la mesure  $\lambda \otimes (\delta_0 + \delta_1)$ , où  $\lambda$  est la mesure de Lebesgue sur  $\mathbb{R}^+$  et  $\delta_x$  la mesure de Dirac :

$$[f(x)\bar{G}(x)]^\delta [R(x)g(x)]^{1-\delta} \mathbb{I}_{\{x \in \mathbb{R}^+, \delta_x \in \{0,1\}\}}$$

La vraisemblance des observations est alors :

$$\mathcal{L}_n((t_1, \delta_1), \dots, (t_n, \delta_n); \theta) = \prod_{i=1}^n [f_\theta(t_i)\bar{G}(t_i)]^{\delta_i} [R_\theta(t_i)g(t_i)]^{1-\delta_i}$$

De plus, on suppose que la loi de la censure à droite ne dépend pas du paramètre d'intérêt de la loi de durée de vie ; on peut donc travailler uniquement sur la vraisemblance informative ou pertinente [10], i.e.

$$\mathcal{L}_n((t_1, \delta_1), \dots, (t_n, \delta_n); \theta) = \prod_{i=1}^n [f_\theta(t_i)]^{\delta_i} [R_\theta(t_i)]^{1-\delta_i}$$

Ainsi, le paramètre d'intérêt peut être estimé par le maximum de vraisemblance.

L'intérêt de cette approche classique et connue, mais quelque peu simpliste dans notre cas (doubles censures), est qu'elle fournit un ordre de grandeur des paramètres. Pour cette raison, elle a été le point de départ de tout ce travail. Cependant, nous devons souligner que cette approximation (censures à gauche = panne) est assez grossière, et induit un biais. Les résultats de cette approche nous permettrons cependant de contrôler le biais de méthode mieux fondées.

### 3.3 Autres possibilités, méthodes actuariales

Dans la méthode précédente, on remplaçait les censures à gauche par des temps de panne. D'autres méthodes dites actuariales consistent à remplacer les censures à gauche par un tirage d'une loi uniforme sur  $[t_j/2, t_j]$ , ou sur  $[0, t_j]$ , et de les considérer comme des vrais temps de défaillance. Cela revient donc à utiliser un nouveau jeu de données (avec des nouveaux temps de défaillance).

Pour résumer dans le cas d'un tirage sur  $[0, t_j]$  :

Pour la  $j^{\text{eme}}$  donnée :

- si c'est une censure à droite alors on garde le temps  $t_j$ ,
- si c'est une censure à gauche alors on remplace  $t_j$  par  $t'_j$ , tiré uniformément sur  $[0, t_j]$ .

Le paramètre sera donc estimé par le maximum de vraisemblance par la même méthode que précédemment. Ces deux dernières méthodes nous permettent a priori d'obtenir une meilleure modélisation que la première citée. Cependant, les temps d'observation étant de l'ordre de 20000 heures de fonctionnement, cette méthode reste très approximative, car elle consiste à tirer des temps suivant une uniforme  $U[0, 20000]$ .

En résumé, on pourrait dire que, de façon simpliste, la première méthode est très optimiste (remplacer les censures à gauche par des temps de panne), que la seconde

(remplacer par une loi uniforme sur  $[t_j/2, t_j]$ ) l'est moins, et que la dernière méthode (remplacer par une loi uniforme sur  $[0, t_j]$ ) est celle qui laisse le plus de liberté, mais aussi, vu l'ordre de grandeur des  $t_j$ , qui est la plus pessimiste.

### 3.4 Fiabilité dans un contexte doublement censuré

On s'intéresse à présent à des données censurées à droite et à gauche. Ce qui est le cas pour nos données réelles. On observe les couples

$$(T_i, \delta_i) \text{ avec } \delta_i = \mathbb{I}_{\{X_i \leq T_i\}}, \text{ pour } i = 1, \dots, n$$

$$\delta_i = \begin{cases} 0 & \text{s'il y a censure à droite (le composant ne présente pas de défaut)} \\ 1 & \text{s'il y a censure à gauche (le composant présente un défaut)} \end{cases}$$

Ainsi, en gardant les mêmes notations que précédemment, on peut écrire la vraisemblance pour  $F$  et  $G$  [6] :

$$\prod_{i=1}^n g(T_i) [F(T_i)]^{\delta_i} [1 - F(T_i)]^{1-\delta_i}$$

La vraisemblance pour les durées de vie étant égale quant à elle à :

$$\mathcal{L}_n = \prod_{i=1}^n [F(T_i)]^{\delta_i} [1 - F(T_i)]^{1-\delta_i} \quad (1)$$

La log-vraisemblance que l'on cherche à maximiser est de la forme :

$$l_n = \sum_{i=1}^n \delta_i \log [F(T_i)] + (1 - \delta_i) \log [1 - F(T_i)] \quad (2)$$

Ce critère est du type M-estimateur, c'est-à-dire que  $\theta$  est solution d'un problème du type :

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} \sum_{i=1}^n \psi_i(X_i, \theta)$$

où  $\psi$  est une fonction réelle.

### 3.5 Approche bayésienne

En pratique, on dispose de quelques informations sur le comportement des matériaux, sur les pannes, ou sur les premières expérimentations. L'approche bayésienne fournit une méthode pour inclure ces informations dans notre modèle. Les durées de vie  $(X_1, \dots, X_n)$  ont pour densité  $f(x; \theta)$ , où  $\theta$  est un élément de  $\Theta$ , sous espace de  $\mathbb{R}^d$ . L'approche

bayésienne consiste à supposer que le paramètre d'intérêt  $\theta$  est aléatoire, de loi à densité  $g(\theta)$  par rapport à la mesure de Lebesgue dans  $\mathbb{R}^d$ . Cette loi du paramètre est dite loi *a priori*. Ainsi, la densité marginale de  $X$ , appelé densité prédictive, est alors :

$$f^*(x) = \int_{\Theta} f(x; \theta) g(\theta) d\theta$$

La densité conditionnelle de  $\theta$ , appelée loi *a posteriori*, est la densité de  $\theta$  sachant  $X = x$ , c'est-à-dire d'après le théorème de Bayes :

$$g(\theta/x_1, \dots, x_n) = \frac{g(x; \theta)}{f^*(x)} = \frac{\prod_{k=1}^n f(x_k; \theta) g(\theta)}{\int_{\Theta} \prod_{k=1}^n f(x_k; \theta) g(\theta) d\theta} \quad (3)$$

Si on note  $t_1, \dots, t_l$  les  $l$  censures à gauche et  $t_{l+1}, \dots, t_n$  les  $n - l$  censures à droite, on peut écrire la densité *a posteriori* comme suit :

$$g(\theta/t_1, \dots, t_n) = \frac{g(\theta) \prod_{k=1}^l F(t_k; \theta) \prod_{k=l+1}^n (1 - F(t_k; \theta))}{K} \quad (4)$$

où

$$K = \int_{\Theta} g(\theta) \prod_{k=1}^l F(t_k; \theta) \prod_{k=l+1}^n (1 - F(t_k; \theta)) d\theta \text{ est une constante indépendante de } \theta.$$

Il nous faut maintenant définir une fonction de perte  $L(\hat{\theta}, \theta)$  représentant le coût de décider

*hattheta* alors que la vraie valeur du paramètre est  $\theta$ . Dans la majorité des cas, on choisit une fonction de perte comme une distance entre l'estimateur et la vraie valeur :

$$L(\hat{\theta}, \theta) = D(\hat{\theta} - \theta)$$

En prenant  $D(x) = x^2$  ou  $D(x) = |x|$ , on parle respectivement d'erreur quadratique ou d'erreur absolue.

**Définition 3** On définit le risque *a posteriori* comme la perte moyenne par rapport à la distribution *a posteriori*, i.e

$$R(\hat{\theta}, t_1, \dots, t_n) = \int_{\Theta} L(\hat{\theta}, \theta) g(\theta/t_1, \dots, t_n) d\theta.$$

L'estimateur de Bayes, associé à la fonction de perte  $L(\hat{\theta}, \theta)$ , est alors la valeur  $\hat{\theta}_{\text{Bayes}}$  qui minimise le risque *a posteriori* sachant  $t_1, \dots, t_n$ .

*Exemple1.* Si on utilise la fonction de perte quadratique, l'estimateur de Bayes de  $\theta$  est alors la moyenne de la densité a posteriori. En effet, le risque a posteriori est :

$$R(\hat{\theta}, t_1, \dots, t_n) = \int_{\Theta} (\hat{\theta} - \theta)^2 g(\theta/t_1, \dots, t_n) d\theta$$

et son minimum est atteint pour

$$\hat{\theta}_{Bayes} = \mathbb{E}[\theta/t_1, \dots, t_n] = \int_{\Theta} \theta g(\theta/t_1, \dots, t_n) d\theta$$

*Exemple2.* Si on utilise la fonction de perte absolue, l'estimateur de Bayes de  $\theta$  est alors la médiane de la densité a posteriori. En effet, le risque a posteriori devient :

$$R(\hat{\theta}, t_1, \dots, t_n) = \int_{\Theta} |\hat{\theta} - \theta| g(\theta/t_1, \dots, t_n) d\theta$$

et son minimum est atteint pour  $\hat{\theta}_{bayes}$  qui vérifie

$$\int_0^{\hat{\theta}_{Bayes}} g(\theta/t_1, \dots, t_n) d\theta = \int_{\hat{\theta}_{Bayes}}^{+\infty} g(\theta/t_1, \dots, t_n) d\theta = \frac{1}{2}.$$

Dans notre étude, nous nous intéresserons à la moyenne a posteriori.

### 3.6 Estimation non paramétrique de la fonction de survie

L'objectif de ce paragraphe est d'estimer la fonction de survie d'un composant dans le cadre de données doublement censurées. Les notations restants les mêmes, on observe toujours les couples  $(T_i, \delta_i)$  avec  $\delta_i = \mathbb{I}_{X_i \leq T_i}$ , pour  $i = 1, \dots, n$ , où  $X_i$  sont les durées de vie. Le premier estimateur proposé est empirique. Le second proposé par Turnbull [13] consiste à revenir dans un cas mono-censuré.

#### 3.6.1 L'estimateur empirique

Une des méthodes pour estimer la fonction de survie est celle décrite dans l'article de Andersen et Rønn [6]. Nous allons tout d'abord ordonner les temps d'observations et les noter :  $T_{(1)} < T_{(2)} < \dots < T_{(n)}$  et renommer les  $\delta_i$  correspondants. Soit alors la fonction :

$$H_i = \sum_{k=1}^i \delta_k \quad \text{pour } i = 1, \dots, n.$$

qui calcule le nombre de censures à gauche au temps  $T_{(i)}$ .  $H_n$  est donc égale au nombre totale de censures à gauche présentes dans l'échantillon.

Ainsi, on prend comme estimateur de la fonction de survie l'expression suivante :

$$\hat{R}(T_i) = 1 - \frac{H_i}{H_n}.$$

En prenant de plus  $t_0 = 0$  et  $t_{n+1} = +\infty$ , on pose  $R_n(t) = R_n(T_i)$  pour  $t \in [T_i, T_{i+1}]$ .

Cet estimateur, assez grossier, est une fonction en escalier, pouvant être lissée, par des noyaux gaussiens par exemple.

### 3.6.2 L'estimateur de Turnbull

Ce paragraphe fait référence à Turnbull [13], qui a proposé un estimateur non paramétrique pour la fonction de survie avec des données doublement censurées. La seule différence par rapport à notre cas, est qu'il considère observer toujours au moins une défaillance, ce qui entraîne une expression différente de la vraisemblance ([2]).

On suppose qu'il existe une échelle de temps discret (par exemple nos temps d'observations ordonnés)  $0 < t_1 < t_2 < \dots < t_m$ . On note  $\lambda_i$  le nombre de défaillances au temps  $t_i$ ,  $\delta_i$  le nombre de censures à gauche, et  $\mu_i$  le nombre de censures à droite.

Le nombre moyen de censures à gauche qui vont tomber défaillant dans la période  $[t_{j-1}, t_j]$  sachant qu'il a été défaillant avant  $t_i$  est égale à  $\delta_i \alpha_{ij}$  où

$$\begin{aligned} \alpha_{ij} &= \left( \hat{P}_{j-1} - \hat{P}_j \right) / \left( 1 - \hat{P}_i \right), \\ P_i &= R(t_i), \\ R(t) &= 1 - F(t). \end{aligned}$$

Ainsi, il est possible de se ramener à un cas de données censurées uniquement à droite en se plaçant dans un contexte "ajusté". On va remplacer les défaillances  $\lambda_j$  par  $\lambda'_j = \lambda_j + \sum_{i=j}^m \delta_i \alpha_{ij}$ , laisser les  $\mu_j$  inchangés et remplacer les  $\delta_j$  par 0.

Nous sommes maintenant en mesure d'appliquer la méthode de Kaplan-Meier [7], qui fournit l'estimateur du maximum de vraisemblance consistant. Il reste à trouver les nombres  $1 \geq \hat{P}_1 \geq \hat{P}_2 \geq \dots \geq \hat{P}_m \geq 0$ , satisfaisants les équations :

$$\begin{aligned} \hat{P}_1 &= q_1 \\ \hat{P}_j &= q_j \hat{P}_{j-1} \quad (j = 2, 3, \dots, m) \end{aligned}$$

où

$$\begin{aligned} q_j &= 1 - \lambda'_j / n'_j \quad , \quad n'_j = \sum_{i=j}^m \left[ \mu_i + \lambda'_i \right] \\ \lambda'_j &= \lambda_j + \sum_{i=j}^m \delta_i \alpha_{ij} \end{aligned}$$

et

$$\alpha_{ij} = \left( \hat{P}_{j-1} - \hat{P}_j \right) / \left( 1 - \hat{P}_i \right) \quad \text{pour } j \leq i.$$

Cette procédure proposée par Turnbull peut s'interpréter comme un algorithme EM (cf annexe Algorithme 1). En effet, les données observées sont les défaillances  $\lambda = (\lambda_1, \dots, \lambda_m)$ , les censures à droite  $\mu = (\mu_1, \dots, \mu_m)$ , et les censures à gauche  $\delta = (\delta_1, \dots, \delta_m)$ .

La log-vraisemblance observée est

$$\mathcal{L}(P_1, \dots, P_m; \lambda, \mu, \delta) = \sum_{i=1}^m [\lambda_i \log(P_{i-1} - P_i) + \mu_i \log P_i + \delta_i \log(1 - P_i)]$$

On va considérer que seuls les censures à gauche sont des valeurs manquantes. En effet, s'il n'y avait pas de données censurées à gauche, l'estimateur de Kaplan-Meier serait l'estimateur du maximum de vraisemblance (qui plus est consistant cf. [5] chapitre 4). Dans ce cadre les données complètes sont  $\lambda, \mu$  et  $\nu = (\nu_1, \dots, \nu_n)$  avec

$$\nu_j = \sum_{i=j}^m \sum_{l=1}^{\delta_i} a_{lj}^i \quad (5)$$

et

$$a_{lj}^i = \begin{cases} 1 & \text{si } l \text{ est un individu censuré à gauche en } t_i \text{ et s'il est tombé en panne entre } [t_{j-1}, t_j], \\ 0 & \text{sinon.} \end{cases}$$

Les données manquantes sont donc les  $a_{lj}^i$ ,  $n$  étant le nombre d'individus pris en compte dans l'expérience. La vraisemblance des données complètes s'écrit

$$\mathcal{L}(P_1, \dots, P_m; \lambda, \mu, \nu) = \sum_{i=1}^m [(\lambda_i + \nu_i) \log(P_{i-1} - P_i) + \mu_i \log P_i] \quad (6)$$

D'après (5) et (6), il est clair que l'étape E de EM revient à calculer

$$P(a_{lj}^i = 1 | P_1, \dots, P_m, \lambda, \mu, \delta)$$

qui est le  $\alpha_{ij} = (\hat{P}_{j-1} - \hat{P}_j) / (1 - \hat{P}_i)$  de la procédure de Turnbull.

L'étape M revient à rechercher  $P_1^{r+1}, \dots, P_m^{r+1}$  tel que

$$\mathbb{E}[\mathcal{L}(P_1, \dots, P_m; \lambda, \mu, \nu) | P_1^r, \dots, P_m^r] = \sum_{i=1}^m [(\lambda_i + \mathbb{E}[\nu_i | P_1^r, \dots, P_m^r]) \log(P_{i-1} - P_i) + \mu_i \log P_i]$$

soit maximal avec

$$\mathbb{E}[\nu_j | P_1^r, \dots, P_m^r] = \sum_{i=j}^m \delta_j \alpha_{ij}^r,$$

ce qui revient à utiliser l'estimateur de Kaplan-Meier sur les données complètes où  $\lambda_j$  est remplacé par  $\lambda_j + \sum_{i=j}^m \delta_j \alpha_{ij}$  et  $\delta_j$  remplacé par 0, comme indiqué par Turnbull.

## 4 Propriétés des estimateurs

### 4.1 Le cas censurés à droite

#### 4.1.1 Modélisation par une loi exponentielle

On suppose que la distribution des durées de vie du composant est exponentielle de paramètre  $\theta$ , appelé intensité ou taux de défaillance. On a alors :

$$F(t) = 1 - \exp(-\theta t)$$

La vraisemblance pertinente s'écrit alors :

$$\begin{aligned} \mathcal{L}_n(t_1, \dots, t_n; \delta_1, \dots, \delta_n \theta) &= \prod_{i=1}^n \left[ (\theta \exp(-\theta t_i))^{\delta_i} (\exp(-\theta t_i))^{1-\delta_i} \right] \\ &= \prod_{i=1}^n (\theta)^{\delta_i} \exp(-\theta \sum_{i=1}^n t_i) \\ &= (\theta)^{\sum_{i=1}^n \delta_i} \exp(-\theta \sum_{i=1}^n t_i) \end{aligned}$$

Le logarithme de la vraisemblance est

$$\ell_n = \log \mathcal{L}_n(t_1, \dots, t_n; \delta_1, \dots, \delta_n; \theta) = \log \theta \sum_{i=1}^n \delta_i - \theta \sum_{i=1}^n t_i$$

D'où

$$\dot{\ell}_n = \frac{\partial \log \mathcal{L}_n(t_1, \dots, t_n; \delta_1, \dots, \delta_n; \theta)}{\partial \theta} = \frac{\sum_{i=1}^n \delta_i}{\theta} - \sum_{i=1}^n t_i.$$

La dérivée en  $\theta$  de ce dernier terme étant négative (et la fonction log-vraisemblance concave), l'estimateur du maximum de vraisemblance du taux de défaillance est donc :

$$\hat{\theta}_n = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n t_i} = \frac{\text{nombre de défaillants}}{\text{temps total d'observation}}.$$

*Remarque* : Il est important de rappeler que pour notre étude, cette modélisation revient à remplacer les censures à gauche par des pannes. Ainsi, cette méthode a tendance à sous-estimer le taux de défaillance. Cependant cette première étude nous permet d'estimer l'ordre de grandeur du paramètre d'intérêt.

### 4.1.2 Modélisation par une loi de Weibull à deux paramètres

On suppose ici que la durée de vie du composant suit une loi de Weibull notée  $W(\eta, \beta)$  de fonction de répartition :

$$F(t) = 1 - \exp \left\{ - \left( \frac{t}{\eta} \right)^\beta \right\} \quad \text{pour } t > 0.$$

La vraisemblance est alors donnée par :

$$\mathcal{L}_n(t_1, \dots, t_n; \delta_1, \dots, \delta_n; \eta, \beta) = \prod_{i=1}^n \left[ \frac{\beta}{\eta} \left( \frac{t_i}{\eta} \right)^{\beta-1} \exp \left\{ - \left( \frac{t_i}{\eta} \right)^\beta \right\} \right]^{\delta_i} \left[ \exp \left\{ - \left( \frac{t_i}{\eta} \right)^\beta \right\} \right]^{1-\delta_i}$$

Ainsi, la log-vraisemblance vaut :

$$\ell_n = \log \mathcal{L}_n(t_1, \dots, t_n; \delta_1, \dots, \delta_n; \eta, \beta) = \sum_{i=1}^n \left\{ \delta_i \left[ \log \frac{\beta}{\eta} + (\beta - 1) \log \left( \frac{t_i}{\eta} \right) \right] - \left( \frac{t_i}{\eta} \right)^\beta \right\}$$

Les dérivées partielles donnent :

$$\begin{cases} \dot{\ell}_n^\eta = \frac{\partial \mathcal{L}_n}{\partial \eta} = - \sum_{i=1}^n \delta_i + \sum_{i=1}^n \left( \frac{t_i}{\eta} \right)^\beta \\ \dot{\ell}_n^\beta = \frac{\partial \mathcal{L}_n}{\partial \beta} = \frac{\sum_{i=1}^n \delta_i}{\beta} + \sum_{i=1}^n \delta_i \log \left( \frac{t_i}{\eta} \right) - \sum_{i=1}^n \left[ \log \left( \frac{t_i}{\eta} \right) \left( \frac{t_i}{\eta} \right)^\beta \right] \end{cases}$$

Le maximum de vraisemblance est alors donné par les équations implicites suivantes :

$$\begin{cases} \hat{\eta}_n = \frac{1}{\left[ \frac{\sum_{i=1}^n t_i^{\hat{\beta}_n}}{n} \right]^{\hat{\beta}_n}} \\ \frac{1}{\hat{\beta}_n} + \frac{\sum_{i=1}^n \delta_i \log(t_i)}{\sum_{i=1}^n \delta_i} = \frac{\sum_{i=1}^n \log(t_i) (t_i)^{\hat{\beta}_n}}{\sum_{i=1}^n (t_i)^{\hat{\beta}_n}} \end{cases}$$

Ainsi, le taux de défaillance est égal à :

$$\lambda_n(x) = \frac{f(x)}{R(x)} = \frac{\hat{\beta}_n}{\hat{\eta}_n} \left( \frac{x}{\hat{\eta}_n} \right)^{\hat{\beta}_n - 1}.$$

## 4.2 Le cas doublement censuré

Nous allons rechercher, dans ce paragraphe, les principales caractéristiques asymptotiques de l'estimateur du maximum de vraisemblance dans le cas doublement censuré.

### 4.2.1 Consistance de l'estimateur du maximum de vraisemblance

Sous des conditions de régularité du modèle, l'estimateur du maximum de vraisemblance possède des propriétés de convergence vers la vraie valeur. On se propose donc dans ce paragraphe d'étudier toutes ces propriétés.

Pour commencer, nous définissons tout d'abord ce qu'est un estimateur :

**Définition 4** *L'espace  $\Omega$  des observations est supposé muni d'une  $\sigma$ -algèbre  $\mathcal{A}$ ; l'espace  $\Theta$  des valeurs du paramètre est supposé muni d'une  $\sigma$ -algèbre  $\mathcal{B}$ .*

*Un estimateur  $\hat{\theta}$  est une application mesurable de  $(\Omega, \mathcal{A})$  dans  $(\Theta, \mathcal{B})$ , c'est-à-dire est telle que :*

$$\hat{\theta}^{-1}(B) \in \mathcal{A}, \forall B \in \mathcal{B}$$

L'existence d'une telle application est assurée par la proposition suivante (cf. [9]).

**Proposition 1** *Soit  $\Theta$  une partie compacte de  $\mathbb{R}^p$  et  $\mathcal{B} = \mathcal{B}(\Theta)$  sa tribu borélienne. Soit  $\mathcal{Q}$  une application de  $\Omega \times \Theta$  dans  $\mathbb{R}$ , satisfaisant les deux conditions suivantes :*

- (i) *l'application  $\mathcal{Q}(\cdot, \theta) : \omega \longrightarrow \mathcal{Q}(\omega, \theta)$  est mesurable de  $(\Omega, \mathcal{A})$  dans  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  et ceci pour tout  $\theta \in \Theta$ ;*
- (ii) *l'application partielle  $\mathcal{Q}(\omega, \cdot) : \theta \longrightarrow \mathcal{Q}(\omega, \theta)$  est continue pour tout  $\omega \in \Omega$ .*

*Alors il existe une application mesurable  $\hat{\theta}$  de  $(\Omega, \mathcal{A})$  dans  $(\Theta, \mathcal{B}(\Theta))$  telle que :*

$$\forall \omega \quad \mathcal{Q}(\omega, \hat{\theta}(\omega)) = \max_{\theta \in \Theta} \mathcal{Q}(\omega, \theta)$$

Remarque : À propos de la compacité de  $\Theta$ , on peut se restreindre à une hypothèse plus faible, à savoir que  $\Theta$  est d'intérieur non vide, et que la vraie valeur du paramètre  $\theta_0$  est à l'intérieur de  $\Theta$  (ce que l'on supposera être le cas dans cette étude).

**Proposition 2** *Si en plus des hypothèses de la proposition 1, on suppose que, pour toute vraie loi  $P_0$  du processus :*

- *la suite  $\mathcal{Q}_n(\omega, \theta)$  converge uniformément en  $\theta$  vers  $\mathcal{Q}_\infty(\theta)$ ,  $P_0$  presque sûrement (c'est-à-dire  $\sup_{\theta \in \Theta} |\mathcal{Q}_n(\omega, \theta) - \mathcal{Q}_\infty(\theta)| \rightarrow 0$  pour  $P_0$  presque tout  $\omega$ )*
- *$\mathcal{Q}_\infty(\theta)$  a un maximum unique en  $\theta_0$*

*alors il existe une suite d'estimateurs  $\hat{\theta}_n$ , solutions de  $\sup_{\theta \in \Theta} \mathcal{Q}_n(\omega, \theta)$  qui converge  $P_0$  presque sûrement vers  $\theta_0$ .*

Sous de bonnes hypothèses de régularité de la fonction à maximiser, on est donc en mesure de montrer la consistance de l'estimateur du maximum de vraisemblance.

### 4.2.2 Normalité asymptotique de l'estimateur du maximum de vraisemblance

Dans ce paragraphe, nous allons montrer la convergence normale de l'estimateur du maximum de vraisemblance.

Avant cela, nous définissons différentes matrices d'information de Fisher intervenant dans les théorèmes asymptotiques :

**Définition 5** *On appelle :*

- **information incrémentale**, la variation quadratique du score notée

$$J_n(\theta) = \sum_{i=1}^n \left( \dot{\ell}_i(\theta) - \dot{\ell}_{i-1}(\theta) \right) \left( \dot{\ell}_i(\theta) - \dot{\ell}_{i-1}(\theta) \right)';$$

- **information de Fisher conditionnelle**, la variation quadratique prévisible du score notée

$$I_n(\theta) = \sum_{i=1}^n \mathbb{E}_\theta \left[ \left( \dot{\ell}_i(\theta) - \dot{\ell}_{i-1}(\theta) \right) \left( \dot{\ell}_i(\theta) - \dot{\ell}_{i-1}(\theta) \right)' \mid \mathcal{F}_{i-1} \right];$$

- **information de Fisher moyenne**, la matrice semi-définie positive notée

$$i_n(\theta) = \mathbb{E}_\theta [I_n(\theta)] = \mathbb{E}_\theta [J_n(\theta)];$$

- **information de Fisher observée**, la matrice notée

$$j_n(\theta) = -\ddot{\ell}_n.$$

**Proposition 3** [9] *Si, en plus des hypothèses de la proposition 2, on suppose que,*

- \* la vraie valeur  $\theta_0$  du paramètre est à l'intérieur de  $\Theta$ ,
- \*  $\mathcal{Q}_n(\omega, \theta)$  est deux fois continuellement différentiable par rapport à  $\theta$ ,
- \*  $\frac{\partial^2 \mathcal{Q}_n}{\partial \theta \partial \theta'}(\omega, \theta)$  converge,  $P_0$ -p.s., uniformément en  $\theta$ , vers une limite  $-J(\theta)$ ,
- \*  $J(\theta_0)$  est inversible,
- \* la suite  $\sqrt{n} \frac{\partial \mathcal{Q}_n}{\partial \theta}(\omega, \theta_0)$  converge en loi vers une loi normale  $\mathcal{N}(0, I(\theta_0))$ ,

Alors la suite d'estimateurs  $\hat{\theta}_n$  est telle que :

$$\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, J^{-1}(\theta_0) I(\theta_0) J^{-1}(\theta_0) \right)$$

Les convergences presque sûres peuvent être vérifiées grâce à la loi forte des grands nombres. L'uniformité de la convergence peut être donnée grâce au théorème générique d'Andrews (cf annexe proposition 7).

Pour montrer que la fonction score converge en loi, on utilise un théorème de convergence sur les accroissements de martingales. Il restera ensuite à montrer dans les cas particulier que la fonction score est une différence de martingale.

**Définition 6** *On dit que  $((Y_n)_{n \geq 1}, (\mathcal{F}_n)_{n \geq 0})$  est une suite d'accroissements de martingale si :*

(i)  $(Y_n)_{n \geq 1}$  est adapté à la filtration  $(\mathcal{F}_n)_{n \geq 1}$  ;

(ii)  $\mathbb{E}[|Y_n|] < +\infty$ , pour tout  $n \geq 1$  ;

(iii)  $\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = 0$ , pour tout  $n \geq 0$ .

**Théorème 1** [9] *Soit  $((Y_n)_{n \geq 1}, (\mathcal{F}_n)_{n \geq 0})$  une suite d'accroissements de martingale de carré intégrable. Pour tout  $n \geq 1$ , nous posons :*

$$S_n = Y_1 + \dots + Y_n, \quad \sigma_n^2 = \sum_{i=1}^n \mathbb{E}[Y_i^2] \quad \text{et} \quad V_n = \sum_{i=1}^n \mathbb{E}[Y_i^2|\mathcal{F}_{i-1}]$$

*Supposons que les deux conditions suivantes soient satisfaites :*

(i) *la suite des v.a.r.  $(\sigma_n^{-2}V_n)_{n \geq 1}$  converge en probabilité vers 1.*

(ii) *pour tout  $\epsilon > 0$ ,  $\lim_{n \rightarrow +\infty} \sigma_n^{-2} \sum_{i=1}^n \mathbb{E}[Y_i^2 \mathbb{I}_{\{|Y_i| > \epsilon \sigma_n\}}] = 0$ .*

*Alors :*

$$\lim_{n \rightarrow +\infty} \sup_{\alpha \in \mathbb{R}} \left| \mathbb{P} \left[ \frac{S_n}{\sigma_n} \leq \alpha \right] - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-\frac{x^2}{2}} dx \right| = 0.$$

### 4.2.3 Modélisation par une loi exponentielle

Si on suppose que la loi des durées de vie est exponentielle de paramètre  $\theta$  alors la vraisemblance s'écrit sous la forme :

$$\mathcal{L}_n = \prod_{i=1}^n [1 - \exp(-\theta t_i)]^{\mathbb{I}_{\{x_i \leq t_i\}}} [\exp(-\theta t_i)]^{\mathbb{I}_{\{x_i > t_i\}}}$$

où  $\theta \in \Theta = \mathbb{R}_+^*$ . On suppose que la vraie valeur du paramètre est à l'intérieur de  $\Theta$ , ce qui est réaliste.

La log-vraisemblance s'écrit alors :

$$\ell_n = \log \mathcal{L}_n = \sum_{i=1}^n \left\{ \mathbb{I}_{\{x_i \leq t_i\}} \log [1 - \exp(-\theta t_i)] - (\mathbb{I}_{\{x_i > t_i\}}) \theta t_i \right\}.$$

Soit  $\mathcal{Q}_n = \frac{1}{n} \ell_n$  la fonction à maximiser :

$$\mathcal{Q}_n = \frac{1}{n} \sum_{i=1}^n \psi_i = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{x_i \leq t_i\}} \log(1 - e^{\theta t_i}) - \mathbb{I}_{\{x_i > t_i\}} \theta t_i$$

\* Cette fonction est mesurable en  $\omega = (t, \delta) \in \Omega = (\mathbb{R}_+^*, (0, 1))$ , et ce pour tout  $\theta$  : (i) de la proposition 1,

\*  $\mathcal{Q}_n$  est continue en  $\theta$  pour tout  $\omega$  : (ii) de la proposition 1.

Il existe alors, d'après la proposition 1, un estimateur qui maximise  $\mathcal{Q}_n$ , et donc la vraisemblance.

Pour prouver la consistance de cet estimateur, il nous faut montrer la convergence uniforme presque sûre de la fonction  $\mathcal{Q}_n(\omega, \theta)$  vers une fonction  $\mathcal{Q}_\infty$ .

Pour la convergence presque sûre, il nous suffit de supposer que l'on connaît la suite infinie  $(t_1, \dots, t_\infty)$ , et qu'il existe un intervalle qui contient tous les  $t_i$  (cf. [4]).

Alors il est possible de trouver  $T$  une variable aléatoire réelle (avec  $\mathbb{E}[T] < +\infty$ ), telle que ses réalisations soient les  $t_i$  (pour nous déterministes et connus). Ainsi, en supposant que  $X$  et  $T$  sont deux variables aléatoires indépendantes, la convergence presque sûre vient de la loi forte des grands nombres.

On a donc

$$\mathcal{Q}_\infty = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n (1 - e^{\theta t_i}) \log(1 - e^{\theta t_i}) + e^{\theta t_i} \log e^{\theta t_i}$$

Pour l'uniformité de la convergence, nous devons utiliser le théorème d'Andrews (cf annexe proposition 7) :

Comme auparavant, l'hypothèse **(i)** du théorème d'Andrews sur  $\Theta$  peut être "allégée" en un ensemble d'intérieur non vide, et tel que la vraie valeur soit dans cet intérieur.

L'hypothèse **(ii)** est immédiatement vérifiée, par définition de  $\mathcal{Q}_\infty$ .

On calcule pour tout  $\theta, \theta' \in \Theta$

$$\begin{aligned} & \left| \psi_i(X_i, \theta) - \psi_i(X_i, \theta') \right| \\ &= \left| \mathbb{I}_{\{x_i \leq t_i\}} \log \left( \frac{1 - e^{-\theta t_i}}{1 - e^{-\theta' t_i}} \right) + \mathbb{I}_{\{x_i > t_i\}} t_i (\theta' - \theta) \right| \end{aligned}$$

Or  $\forall \epsilon > 0, \exists k$  tel que  $\psi_i(\theta)$  soit  $k$ -lipschitzienne sur  $[\epsilon, +\infty[$ , car la dérivée est bornée en  $\theta$ .

Ainsi, la condition **(iii)** est remplie avec la fonction  $h(x) = x$ , et  $B_i(X_i) = kt_i$ . Les deux dernières conditions sont alors immédiatement satisfaites.

On peut donc conclure que  $\mathcal{Q}_n$  converge uniformément en  $\theta$ . De plus l'inégalité de Kullback nous dit que

$$\mathbb{E}_{\theta_0}[\ell_n(\theta_0)] \geq \mathbb{E}_{\theta_0}[\ell_n(\theta)] \quad \forall \theta \in \Theta$$

Donc en passant à la limite,  $\mathcal{Q}_\infty(\theta)$  a un maximum unique en  $\theta_0$ . L'unicité est donnée par la concavité de la log-vraisemblance ( $\ddot{\ell}_n < 0$ ).

**En conclusion, l'estimateur du maximum de vraisemblance pour le modèle exponentiel noté  $\hat{\theta}_n$  est consistant.**

On va maintenant s'intéresser à la normalité asymptotique de cet estimateur consistant. Il faut s'assurer que les hypothèses de la proposition 3 sont satisfaites.

On supposera que la vraie valeur est à l'intérieur de  $\Theta$ . De plus, il est clair que  $\mathcal{Q}_n$  est deux fois continuellement différentiable par rapport à  $\theta$ .

La fonction score peut s'écrire

$$\dot{\ell}_n = \frac{\partial \log \mathcal{L}_n}{\partial \theta} = \sum_{i=1}^n t_i \left\{ \mathbb{I}_{\{x_i \leq t_i\}} \frac{\exp(-\theta t_i)}{1 - \exp(-\theta t_i)} - \mathbb{I}_{\{x_i > t_i\}} \right\},$$

et la dérivée seconde

$$\ddot{\ell}_n = - \sum_{i=1}^n \mathbb{I}_{\{x_i \leq t_i\}} t_i^2 \frac{\exp(-\theta t_i)}{(1 - \exp(-\theta t_i))^2} < 0.$$

Pour les mêmes raisons que précédemment, en considérant la variable aléatoire  $T \frac{\partial^2 \mathcal{Q}_n}{\partial \theta \partial \theta'}$  converge uniformément vers une limite notée  $-J(\theta)$  avec

$$J(\theta) = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{x_i \leq t_i\}} t_i^2 \frac{\exp(-\theta t_i)}{(1 - \exp(-\theta t_i))^2}$$

et  $J(\theta_0)$  inversible.

Il nous reste maintenant à montrer que  $\sqrt{n} \frac{\partial \mathcal{Q}_n}{\partial \theta}(\omega, \theta_0)$  converge vers une loi normale.

**Proposition 4**  $(\dot{\psi}_n, \mathcal{F}_{n-1})_{n \geq 1}$  est une différence de martingale sous  $\mathbb{P}_\theta, \forall \theta$ .

**Preuve :** La dérivée de la log-vraisemblance est égale à

$$\dot{\ell}_n = \sum_{i=1}^n \dot{\psi}_i = \sum_{i=1}^n \mathbb{I}_{\{x_i < t_i\}} \frac{t_i \exp(-\theta t_i)}{1 - \exp(-\theta t_i)} - \mathbb{I}_{\{x_i > t_i\}} t_i$$

Soit  $\mathcal{F}_n = \sigma(\dot{\psi}_1, \dots, \dot{\psi}_n)$  la filtration naturelle de  $\dot{\psi}_n$ , c'est-à-dire la tribu engendrée par  $(X_1, \dots, X_n)$ . Par définition  $\dot{\ell}_n$  est adaptée à  $\mathcal{F}_n$ .

De plus,

$$\mathbb{E} \left[ |\dot{\psi}_n| \right] = |t_i \exp(-\theta t_i) - t_i \exp(-\theta t_i)| = 0 \quad \forall n > 0$$

Enfin,

$$\begin{aligned} \mathbb{E} \left[ \dot{\psi}_{n+1} | \mathcal{F}_n \right] &= t_{n+1} \exp(-\theta t_{n+1}) - t_{n+1} * \exp(-\theta t_{n+1}) \\ &= 0 \end{aligned}$$

■

Pour appliquer le théorème 1 à la fonction score, on pose  $Y_i = \psi_i$ ,  $S_n = \dot{\ell}_n$ ,  $\sigma_n^2 = i_n(\theta)$ , et  $V_n = I_n(\theta)$ , où  $i_n(\theta)$  est l'information de Fisher moyenne, et  $I_n(\theta)$  est l'information de Fisher conditionnelle (cf définition 5).

C'est-à-dire

$$\begin{aligned} J_n(\theta) = I_n(\theta) &= \sum_{i=1}^n \left( \mathbb{I}_{\{x_i \leq t_i\}} \frac{t_i \exp(-\theta t_i)}{1 - \exp(-\theta t_i)} - \mathbb{I}_{\{x_i > t_i\}} t_i \right)^2 \\ &= \sum_{i=1}^n \left( \mathbb{I}_{\{x_i \leq t_i\}} \frac{t_i^2 \exp(-2\theta t_i)}{(1 - \exp(-\theta t_i))^2} + \mathbb{I}_{\{x_i > t_i\}} t_i^2 \right) \end{aligned}$$

et

$$\begin{aligned} i_n(\theta) &= \sum_{i=1}^n t_i^2 \left( \frac{\exp(-2\theta t_i)}{1 - \exp(-\theta t_i)} + \exp(-\theta t_i) \right) \\ &= \sum_{i=1}^n t_i^2 \frac{\exp(-\theta t_i)}{1 - \exp(-\theta t_i)} \end{aligned}$$

Ainsi, la condition (i)  $V_n/\sigma_n^2 \rightarrow 1$  est vérifiée, grâce à la loi forte des grands nombres. De plus, comme  $i_n(\theta)$  tend vers  $+\infty$ , et que les  $\psi_i$  sont finis, la condition (ii) se trouve elle aussi vérifiée.

On peut donc conclure que

$$\frac{\dot{\ell}_n}{\sqrt{\sum_{i=1}^n t_i^2 \frac{e^{-\theta t_i}}{1 - e^{-\theta t_i}}}} \rightarrow \mathcal{N}(0, 1) \quad (7)$$

Pour garder les mêmes notations que la proposition 3, posons

$$I(\theta_0) = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n t_i^2 \frac{e^{-\theta_0 t_i}}{1 - e^{-\theta_0 t_i}}$$

Ainsi

$$\sqrt{n} \frac{J(\theta_0)}{\sqrt{I(\theta_0)}} (\hat{\theta}_n - \theta_0) \rightarrow \mathcal{N}(0, 1)$$

Pour un seuil à  $\alpha$ , on peut en déduire un intervalle de confiance pour  $\theta_0$  :

$$\left[ \hat{\theta}_n - q_{1-\alpha/2} \frac{\sqrt{I(\theta_0)}}{\sqrt{n} J(\theta_0)}, \hat{\theta}_n + q_{1-\alpha/2} \frac{\sqrt{I(\theta_0)}}{\sqrt{n} J(\theta_0)} \right] \quad (8)$$

où  $q_{1-\alpha/2}$  est le quantile d'ordre  $\alpha$  de la loi normale centrée réduite (à 95 %,  $q = 1.96$ ).

#### 4.2.4 Modélisation par une loi de Weibull à deux paramètres

On va supposer ici que les durées de vie des matériaux suivent une loi de Weibull à deux paramètres  $W(\beta, \eta)$ ,  $\beta$  étant appelé paramètre de forme et  $\eta$  paramètre d'échelle. Cette modélisation nous permettra de savoir si le matériel vieillit ou pas :

$$\begin{aligned} \beta < 1 & \quad \text{si rajeunissement du matériel} \\ \beta > 1 & \quad \text{si vieillissement du matériel} \\ \beta = 1 & \quad \text{si il n'y a pas de vieillissement} \end{aligned}$$

En effet le taux de défaillance instantané est égal à :

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{\beta}{\eta} \times \left(\frac{t}{\eta}\right)^{\beta-1}$$

La vraisemblance est alors donnée par :

$$\mathcal{L}_n = \prod_{i=1}^n \left\{ 1 - \exp \left[ - \left(\frac{t_i}{\eta}\right)^\beta \right] \right\}^{\delta_i} \left\{ \exp \left[ - \left(\frac{t_i}{\eta}\right)^\beta \right] \right\}^{1-\delta_i}$$

D'où la log-vraisemblance :

$$\ell_n = \sum_{i=1}^n \left[ \mathbb{I}_{\{x_i \leq t_i\}} \log \left\{ 1 - \exp \left[ - \left(\frac{t_i}{\eta}\right)^\beta \right] \right\} - \mathbb{I}_{\{x_i > t_i\}} \left(\frac{t_i}{\eta}\right)^\beta \right]$$

La fonction à maximiser sera donc, comme dans le cas précédent,  $\mathcal{Q}_n = \frac{1}{n} \ell_n$ . Nous supposons que cette fonction possède les mêmes caractéristiques que pour le modèle exponentiel. **Ainsi, nous supposons que l'estimateur du maximum de vraisemblance pour le modèle de Weibull noté  $\hat{\theta}_n$  est consistant.**

Remarque importante : Ce résultat n'a pas été démontré théoriquement. Certaines conditions peuvent être vérifiées par analogie avec le cas exponentiel. Cependant des conditions, comme celles sur  $\Theta$  (de dimension 2), posent des difficultés à être vérifiées. Ainsi, comme nous pourrons le voir dans les applications, l'estimateur du maximum de vraisemblance peut se trouver sur la frontière de l'espace des paramètres (typiquement  $\eta = +\infty$  et  $\beta = 0$ ). Cependant nous poursuivons les calculs ci après.

De plus les dérivées partielles sont égales à :

$$\left\{ \begin{array}{l} \dot{\ell}_n^\eta = \frac{\partial \ell_n}{\partial \eta} = \sum_{i=1}^n \left(\frac{t_i}{\eta}\right)^\beta \frac{\beta}{\eta} \left\{ \mathbb{I}_{\{x_i > t_i\}} - \mathbb{I}_{\{x_i \leq t_i\}} \frac{\exp\left[-\left(\frac{t_i}{\eta}\right)^\beta\right]}{1 - \exp\left[-\left(\frac{t_i}{\eta}\right)^\beta\right]} \right\} \\ \dot{\ell}_n^\beta = \frac{\partial \ell_n}{\partial \beta} = \sum_{i=1}^n \left(\log \frac{t_i}{\eta}\right) \left(\frac{t_i}{\eta}\right)^\beta \left\{ \mathbb{I}_{\{x_i \leq t_i\}} \frac{\exp\left[-\left(\frac{t_i}{\eta}\right)^\beta\right]}{1 - \exp\left[-\left(\frac{t_i}{\eta}\right)^\beta\right]} - \mathbb{I}_{\{x_i > t_i\}} \right\} \end{array} \right.$$

Elles peuvent s'écrire sous la forme  $\frac{\partial \ell_n}{\partial \theta} = \sum_{i=1}^n \psi_i(x_i, \theta)$ , où  $\theta = (\eta, \beta)$  et

$$\psi_i = \begin{pmatrix} \mathbb{I}_{\{x_i \leq t_i\}} \frac{e^{-\left(\frac{t_i}{\eta}\right)^\beta}}{1 - e^{-\left(\frac{t_i}{\eta}\right)^\beta}} - \mathbb{I}_{\{x_i > t_i\}} \\ \left(\frac{t_i}{\eta}\right)^\beta \frac{\log t_i}{\eta} \end{pmatrix} \begin{bmatrix} -\left(\frac{t_i}{\eta}\right)^\beta \frac{\beta}{\eta} \\ \left(\frac{t_i}{\eta}\right)^\beta \frac{\log t_i}{\eta} \end{bmatrix}$$

est une différence de martingale.

En effet,

$$\begin{aligned} \mathbb{E}[\mathbb{I}_{\{x_i \leq t_i\}}] &= 1 - e^{-\left(\frac{t_i}{\eta}\right)^\beta} \\ \mathbb{E}[\mathbb{I}_{\{x_i > t_i\}}] &= e^{-\left(\frac{t_i}{\eta}\right)^\beta} \end{aligned}$$

Calculons les différentes matrices d'information de Fisher. Comme dans le cas exponentiel, on a :

$$\begin{aligned} I_n(\theta) &= J_n(\theta) = \sum_{i=1}^n \psi_i \psi_i' \\ &= \sum_{i=1}^n \left( \mathbb{I}_{\{x_i \leq t_i\}} \frac{e^{-2\left(\frac{t_i}{\eta}\right)^\beta}}{\left(1 - e^{-\left(\frac{t_i}{\eta}\right)^\beta}\right)^2} + \mathbb{I}_{\{x_i > t_i\}} \right) \left(\frac{t_i}{\eta}\right)^{2\beta} \begin{bmatrix} \beta^2/\eta^2 & -\beta/\eta \log t_i/\eta \\ -\beta/\eta \log t_i/\eta & (\log(t_i/\eta))^2 \end{bmatrix} \end{aligned}$$

et

$$\begin{aligned} i_n(\theta) &= \mathbb{E}[I_n(\theta)] \\ &= \sum_{i=1}^n \left(\frac{t_i}{\eta}\right)^{2\beta} \frac{e^{-\left(\frac{t_i}{\eta}\right)^\beta}}{1 - e^{-\left(\frac{t_i}{\eta}\right)^\beta}} \begin{bmatrix} \beta^2/\eta^2 & -\beta/\eta \log t_i/\eta \\ -\beta/\eta \log t_i/\eta & (\log(t_i/\eta))^2 \end{bmatrix} \end{aligned}$$

Posons, comme dans la proposition 1,  $Y_i = \psi_i$ ,  $\sigma_n^2 = \sum_{i=1}^n \mathbb{E}[Y_i^2] = i_n(\theta)$ ,  $V_n = \sum_{i=1}^n \mathbb{E}[Y_i^2 | \mathcal{F}_{i-1}] = I_n(\theta)$ ,  $S_n = \sum_{i=1}^n Y_i = \sum_{i=1}^n \psi_i = \dot{\ell}_n$ . Comme dans le cas exponentiel, les deux conditions du théorème 1 sont vérifiées.

On peut donc en déduire que la fonction score converge en loi :

$$\frac{\dot{\ell}_n}{\sqrt{\sum_{i=1}^n \left(\frac{c_i}{\eta}\right)^{2\beta} \frac{e^{-(c_i/\eta)^\beta}}{1 - e^{-(c_i/\eta)^\beta}}}} \begin{bmatrix} \beta^2/\eta^2 & -\beta/\eta \log c_i/\eta \\ -\beta/\eta \log c_i/\eta & (\log c_i/\eta)^2 \end{bmatrix}^{-1/2} \longrightarrow \mathcal{N}(0, 1)$$

On est maintenant en mesure d'énoncer le théorème central limite pour l'estimateur du maximum de vraisemblance avec

$$I(\theta) = \lim_{n \rightarrow \infty} \frac{1}{n} i_n(\theta)$$

et

$$J(\theta) = - \lim_{n \rightarrow \infty} \frac{1}{n} \begin{bmatrix} \ddot{\ell}_n^\eta & \ddot{\ell}_n^{\eta\beta} \\ \ddot{\ell}_n^{\beta\eta} & \ddot{\ell}_n^\beta \end{bmatrix}$$

$$\text{où } \ddot{\ell}_n^\eta = \frac{\partial^2 \ell_n}{\partial \eta^2}, \ddot{\ell}_n^\beta = \frac{\partial^2 \ell_n}{\partial \beta^2}, \ddot{\ell}_n^{\beta\eta} = \frac{\partial^2 \ell_n}{\partial \beta \partial \eta} \text{ et } \ddot{\ell}_n^{\eta\beta} = \frac{\partial^2 \ell_n}{\partial \eta \partial \beta}.$$

avec

$$\begin{aligned} \ddot{\ell}_n^\eta &= \sum_{i=1}^n \left(\frac{t_i}{\eta}\right)^\beta \frac{\beta(\beta+1)}{\eta^2} \left( \mathbb{I}_{\{x_i > t_i\}} - \mathbb{I}_{\{x_i \leq t_i\}} \frac{e^{-\left(\frac{t_i}{\eta}\right)^\beta}}{1 - e^{-\left(\frac{t_i}{\eta}\right)^\beta}} \right) \\ &\quad - \left(\frac{t_i}{\eta}\right)^{2\beta} \left(\frac{\beta}{\eta}\right)^2 \mathbb{I}_{\{x_i \leq t_i\}} \frac{e^{-\left(\frac{t_i}{\eta}\right)^\beta}}{\left(1 - e^{-\left(\frac{t_i}{\eta}\right)^\beta}\right)^2} \end{aligned}$$

$$\begin{aligned} \ddot{\ell}_n^\beta &= \sum_{i=1}^n \left(\log \frac{t_i}{\eta}\right)^2 \left(\frac{t_i}{\eta}\right)^\beta \left( \mathbb{I}_{\{x_i \leq t_i\}} \frac{e^{-\left(\frac{t_i}{\eta}\right)^\beta}}{1 - e^{-\left(\frac{t_i}{\eta}\right)^\beta}} - \mathbb{I}_{\{x_i > t_i\}} \right) \\ &\quad - \left(\frac{t_i}{\eta}\right)^{2\beta} \left(\log \frac{t_i}{\eta}\right)^2 \mathbb{I}_{\{x_i \leq t_i\}} \frac{e^{-\left(\frac{t_i}{\eta}\right)^\beta}}{\left(1 - e^{-\left(\frac{t_i}{\eta}\right)^\beta}\right)^2} \end{aligned}$$

et

$$\begin{aligned} \ddot{\eta}^\beta = \ddot{\beta}^\eta &= \sum_{i=1}^n \left( \frac{t_i}{\eta} \right)^\beta \frac{1}{\eta} \left( \beta \log \frac{t_i}{\eta} + 1 \right) \left( \mathbb{I}_{\{x_i > t_i\}} - \mathbb{I}_{\{x_i \leq t_i\}} \frac{e^{-(\frac{t_i}{\eta})^\beta}}{1 - e^{-(\frac{t_i}{\eta})^\beta}} \right) \\ &+ \left( \frac{t_i}{\eta} \right)^{2\beta} \frac{\beta}{\eta} \log \frac{t_i}{\eta} \mathbb{I}_{\{x_i \leq t_i\}} \frac{e^{-(\frac{t_i}{\eta})^\beta}}{\left( 1 - e^{-(\frac{t_i}{\eta})^\beta} \right)^2} \end{aligned}$$

On a donc

$$\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \longrightarrow \mathcal{N} \left( 0, J^{-1}(\theta_0) I(\theta_0) J^{-1}(\theta_0) \right)$$

On peut donc en déduire un intervalle de confiance au seuil  $\alpha$  pour le paramètre  $\theta_0 = (\eta_0, \beta_0)$  :

$$\left[ \hat{\theta}_n - \frac{1}{\sqrt{n}} J^{-1}(\theta_0) I^{1/2}(\theta_0) Q_{1-\alpha/2}, \hat{\theta}_n + \frac{1}{\sqrt{n}} J^{-1}(\theta_0) I^{1/2}(\theta_0) Q_{1-\alpha/2} \right] \quad (9)$$

où  $Q_{1-\alpha/2} = [q_{1-\alpha/2} \ q_{1-\alpha/2}]'$  est le vecteur quantile d'ordre  $\alpha$  de la loi normale centrée réduite (à 95 %,  $q = 1.96$ ).

### 4.3 Estimation dans un cadre bayésien

Le possible mauvais comportement de l'estimateur de vraisemblance nous a amené à considérer un cadre bayésien, permettant notamment l'apport d'information d'experts. Nous présentons tout d'abord la simulation de Monte-Carlo, usuel en estimation bayésienne.

### 4.4 Simulation de Monte-Carlo

Dans un cadre bayésien, il nous faut déterminer la loi a posteriori. Cette loi étant implicite avec nos types de données, une approximation de la moyenne peut être faite à l'aide de simulation de Monte-Carlo, dont nous présentons la méthode ci-dessous.

Pour déterminer l'estimateur de Bayes, dans le cas où la fonction de perte est quadratique, il faut calculer la moyenne de la loi marginale a posteriori. Si la dimension de  $\Theta$  est égale à 1, on trouve :

$$\hat{\theta} = \mathbb{E}[\theta/t_1, \dots, t_n] = \int_{\Theta} \theta g(\theta/t_1, \dots, t_n) d\theta$$

Il en va de même si la dimension est strictement supérieure à 1. Par exemple pour la loi de Weibull à deux paramètres, on a :

$$\begin{aligned} \hat{\beta}_{Bayes} &= \mathbb{E}_{\beta} [\beta/t_1, \dots, t_n] = \int_a^b \beta g_{\beta}(\beta/t_1, \dots, t_n) d\beta \\ \hat{\eta}_{Bayes} &= \mathbb{E}_{\eta} [\eta/t_1, \dots, t_n] = \int_0^{+\infty} \eta g_{\eta}(\eta/t_1, \dots, t_n) d\eta \end{aligned}$$

où

$$g_\beta(\beta/t_1, \dots, t_n) = \int_0^{+\infty} g(\eta, \beta/t_1, \dots, t_n) d\eta$$

$$g_\eta(\eta/t_1, \dots, t_n) = \int_a^b g(\eta, \beta/t_1, \dots, t_n) d\beta$$

et donc,

$$\hat{\beta}_{Bayes} = \int_a^b \int_0^{+\infty} \beta g(\eta, \beta/t_1, \dots, t_n) d\beta d\eta$$

$$\hat{\eta}_{Bayes} = \int_0^{+\infty} \int_a^b \eta g(\eta, \beta/t_1, \dots, t_n) d\eta d\beta$$

Pour calculer les intégrales de la forme :

$$\int_{\Theta} g(\theta) f(\theta) d\theta,$$

on va simuler des données  $X_1, \dots, X_n$  suite i.i.d. de loi à densité  $f(x)$  par rapport à la mesure de Lebesgues.

Ainsi en calculant  $g(X_i)$  et grâce à la loi des grands nombres, il vient :

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n g(X_i) = \mathbb{E}[g(X)] = \int_{\Theta} g(x) f(x) dx.$$

#### 4.4.1 Loi a priori informative

En estimation bayésienne, il est commun de prendre des lois a priori qui facilitent le calcul de la loi a posteriori. C'est le cas des lois dites conjuguées d'un modèle pour lequel les lois a priori et a posteriori sont de même nature. Citons, par exemple, les lois gamma et beta pour la plupart des modèles classiques. Cependant, dans notre contexte (données doublement censurées), ces lois ne sont plus conjuguées.

En revanche, elles ont l'avantage de présenter différentes formes suivant le choix des paramètres.

##### Application pour une loi exponentielle

Si on choisit comme loi *a priori* une loi de gamma inverse de paramètre  $(r, s)$ , notée  $\mathcal{IG}(r, s)$  [11]. La densité de  $\theta$  est alors :

$$g(\theta) = \frac{1}{s^r \Gamma(r) \theta^{r+1}} \exp\left(-\frac{1}{s\theta}\right) \quad \theta > 0.$$

avec

$$\Gamma(r) = \int_0^{+\infty} \theta^{r-1} \exp(-\theta) d\theta,$$

et on a :

$$\begin{aligned}\mathbb{E}[\theta] &= \frac{1}{s(r-1)} \\ \mathbb{V}[\theta] &= 1/s^2 (r-1)^2 (r-2)\end{aligned}$$

Ainsi, la densité *a posteriori* peut se mettre sous la forme :

$$\mathcal{L}_n = \frac{\exp\left(-\frac{1}{s\theta}\right)}{\theta^{r+1}K} \prod_{k=1}^l [1 - \exp(-\theta t_k)] \exp\left(-\theta \sum_{k=l+1}^n t_k\right)$$

En choisissant la fonction de perte quadratique, on peut calculer la moyenne de cette loi *a posteriori*, ainsi que sa variance, par des méthodes numériques types Monte-Carlo.

*Remarque* : Il est important de constater que nous ne sommes pas en mesure d'exprimer analytiquement la loi *a posteriori* avec des données doublement censurées. La proposition suivante fournit malgré tout une bonne approximation de l'estimateur bayésien :

**Proposition 5** : Soit  $(X_1, \dots, X_n)$  les durées de vie, suivant une loi exponentielle de paramètre  $\theta$ , doublement censurées, telles que l'on observe  $(t_1, \dots, t_l)$  correspondant à des censures à gauche et  $(t_{l+1}, \dots, t_n)$  correspondant à des censures à droite.

On note  $\ell_n$  la vraisemblance de ce modèle. On suppose comme loi *a priori* sur  $\theta$  une loi de gamma inverse notée  $\mathcal{IG}(r, s)$  on choisit la fonction de perte quadratique. Si on simule  $(\theta_1, \dots, \theta_N) \rightsquigarrow \mathcal{IG}(r-1, s)$  et  $(\tilde{\theta}_1, \dots, \tilde{\theta}_N) \rightsquigarrow \mathcal{IG}(r, s)$  et si on choisit la fonction de perte quadratique alors on a :

$$\frac{1}{s(r-1)} \frac{\sum_{i=1}^N \ell_n(\theta_i)}{\sum_{i=1}^N \ell_n(\tilde{\theta}_i)} \xrightarrow{N \rightarrow +\infty} \hat{\theta}_{Bayes} \quad p.s.$$

Remarque : Le terme  $\frac{1}{s(r-1)}$  correspond à la moyenne de la loi *a priori*.

**Preuve** : On calcule la moyenne *a posteriori* avec une simulation de type Monte-Carlo :

$$\hat{\theta}_{Bayes} = \frac{\int_0^{+\infty} \frac{\exp\left(-\frac{1}{s\theta}\right)}{\theta^r} \prod_{k=1}^l [1 - \exp(-\theta t_k)] \exp\left(-\sum_{k=l+1}^n \theta t_k\right) d\theta}{\int_0^{+\infty} \frac{\exp\left(-\frac{1}{s\theta}\right)}{\theta^{r+1}} \prod_{k=1}^l [1 - \exp(-\theta t_k)] \exp\left(-\sum_{k=l+1}^n \theta t_k\right) d\theta}$$

Soit maintenant  $\theta_1, \dots, \theta_n$  suivant une loi inverse gamma  $\mathcal{IG}(r-1, s)$  de densité  $f_{r-1,s}$ , alors d'après la loi des grands nombres :

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{i=1}^N \ell_n(\theta_i) = s^{r-1} \Gamma(r-1) \int_0^{+\infty} f_{r-1,s}(\theta) \ell_n(\theta) d\theta$$

où le membre de droite correspond au numérateur de la moyenne a posteriori.

Pour calculer le dénominateur de  $\hat{\theta}_{Bayes}$ , on va procéder de la même manière avec des  $\tilde{\theta}_1, \dots, \tilde{\theta}_n$  suivant une loi de gamma inverse  $\mathcal{IG}(r, s)$  de densité  $f_{r,s}$  :

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{i=1}^N \ell_n(\tilde{\theta}_i) = s^r \Gamma(r) \int_0^{+\infty} f_{r,s}(\theta) \ell_n(\theta) d\theta$$

D'où le résultat. ■

Remarque : Il est aisé de calculer la variance de la loi a posteriori par la même méthode.

### Application pour une loi de Weibull

Nous avons choisi, pour  $\beta$ , la loi Bêta de paramètres  $p$  et  $q$  et de support  $[\beta_g, \beta_d]$ , notée  $\mathcal{B}(p, q, [\beta_g, \beta_d])$ . La densité est alors :

$$f(\beta) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} * \frac{(\beta - \beta_g)^{p-1} (\beta_d - \beta)^{q-1}}{(\beta_d - \beta_g)^{p+q-1}}$$

Ses caractéristiques sont :

$$\begin{aligned} \mathbb{E}[\beta] &= \beta_g + \frac{(p-1)(\beta_d - \beta_g)}{p+q} \\ \mathbb{V}[\beta] &= \frac{pq(\beta_d - \beta_g)^2}{(p+q+1)(p+q)^2} \end{aligned}$$

Nous avons choisi  $\beta_g = 0.5$ ,  $\beta_d = 4$ ,  $p = 3.6$  et  $q = 14.6$  (ce qui correspond à une moyenne a priori égale à 1, et à une variance a priori égale à 0.1)(cf. [3]). Il est à noter que si  $x \rightsquigarrow \mathcal{B}(p, q, [0, 1])$ , alors  $(\beta_d - \beta_g)x + \beta_g \rightsquigarrow \mathcal{B}(p, q, [\beta_g, \beta_d])$ .

La loi sur  $\eta$  est une inverse gamma  $\mathcal{IG}(r, s)$  (comme dans le cas exponentiel)

$$g(\eta) = \frac{1}{s^r \Gamma(r) \eta^{r+1}} \exp\left(-\frac{1}{s\eta}\right) \quad \eta > 0.$$

Ainsi, les deux estimateurs bayésiens sont

$$\hat{\beta}_{Bayes} = \frac{\int_{\beta_g}^{\beta_d} \int_0^{+\infty} \beta f(\beta) g(\eta) \mathcal{L}_n d\eta d\beta}{K}$$

et

$$\hat{\eta}_{Bayes} = \frac{\int_0^{+\infty} \int_{\beta_g}^{\beta_d} \eta f(\beta) g(\eta) \mathcal{L}_n d\beta d\eta}{K}$$

où  $K = \int_{\beta_g}^{\beta_d} \int_0^{+\infty} f(\beta) g(\eta) \mathcal{L}_n d\eta d\beta$ .

En pratique, on procédera de la façon suivante

\* On simule  $\beta \rightsquigarrow \mathcal{B}(p, q, [\beta_g, \beta_d])$

\* et  $\eta \rightsquigarrow \mathcal{IG}(r, s)$

Alors

$$\frac{\sum_{i=1}^N \eta_i \mathcal{L}_n(\eta_i, \beta_i)}{N} \xrightarrow{N \rightarrow +\infty} \hat{\eta}_{Bayes} \quad \text{et} \quad \frac{\sum_{i=1}^N \beta_i \mathcal{L}_n(\eta_i, \beta_i)}{N} \xrightarrow{N \rightarrow +\infty} \hat{\beta}_{Bayes}.$$

$$\frac{\sum_{i=1}^N \mathcal{L}_n(\eta_i, \beta_i)}{\sum_{i=1}^N \mathcal{L}_n(\eta_i, \beta_i)}$$

#### 4.4.2 Avec un a priori non informatif

Il arrive parfois de ne pas avoir d'information sur les composants sur lesquels on cherche à modéliser le vieillissement. Une approche bayésienne peut, tout de même être retenue, en implémentant un a priori non informatif.

Dans ce paragraphe, nous allons donc nous intéresser à une famille de loi *a priori* non informative introduite par Martz et Waller (1982) [8] :

$$g(\theta) = \frac{1}{\theta^c} \quad c > 0.$$

Il s'agit d'un *a priori* impropre (i.e.  $\int_0^{+\infty} \frac{1}{\theta^c} d\theta = +\infty$ ) et qui correspond à une loi de Jeffreys si  $c = 1$ .

#### Pour la loi exponentielle :

La loi a priori choisie est la suivante :

$$g(\theta) = \frac{1}{\theta}.$$

D'où

$$\hat{\theta}_{Bayes} = \frac{\int_0^{+\infty} \mathcal{L}_n d\theta}{\int_0^{+\infty} \frac{1}{\theta} \mathcal{L}_n d\theta}.$$

En pratique, nous avons simulé une loi inverse gamma pour le paramètre, puis nous avons calculé ces intégrales par simulation de Monte-Carlo, en multipliant la vraisemblance par l'inverse de la loi a priori.

#### Pour la loi de Weibull :

La loi a priori est

$$g(\eta, \beta) = \frac{1}{\eta\beta}$$

En pratique, nous avons simulé deux lois inverse gamma pour  $\eta$  et  $\beta$ . Les paramètres de ces lois ont été ajusté suivant l'ordre de grandeur de  $\eta$  et  $\beta$ . Ainsi, les méthodes numériques restent les mêmes que précédemment.

#### 4.4.3 Avec un a priori “semi informatif”

On va supposer dans ce paragraphe que les durées de vie des composants  $(X_1, \dots, X_n)$  sont régies par une loi de Weibull, notée  $W(\eta, \beta)$ . On observe toujours  $((t_1, \delta_1), \dots, (t_n, \delta_n))$  où  $\delta_i = \mathbb{I}_{\{X_i \leq T_i\}}$ . De plus, on supposera que le paramètre d'échelle de la loi de Weibull,  $\eta$  suit une loi de Jeffreys, que le paramètre de forme  $\beta$  suit une loi uniforme sur  $[1, 4]$  et que ces deux paramètres sont indépendants. Par cette méthode, nous imposons que le paramètre de forme soit plus grand que 1, c'est-à-dire que le matériel vieillisse. Ce sont ces deux a priori conjugués que l'on appelle a priori “semi informatif” : nous n'apportons aucune information sur le paramètre  $\eta$ , mais nous imposons au paramètre  $\beta$  d'appartenir à  $[1, 4]$ . Ceci peut se justifier par le fait que l'on essaye de modéliser le comportement d'un composant, subissant des contraintes mécaniques et thermiques. Il y a alors tout lieu de penser que ce composant vieillit, et que sa probabilité de tomber en panne, ou de présenter une dégradation, augmente avec le temps.

Cependant la loi uniforme est très peu informative. En effet, celle-ci impose “uniquement” que  $\beta$  appartienne à l'intervalle  $[1, 4]$ , mais laisse celui-ci parfaitement libre à l'intérieur de cet intervalle.

Ainsi, la loi a priori est la suivante :

$$g(\eta, \beta) = \frac{1}{3\eta} \mathbb{I}_{\{\beta \in [1, 4]\}}$$

En remplaçant dans la formule de vraisemblance (1), on trouve :

$$\mathcal{L}_n(\eta, \beta / (t_1, \delta_1), \dots, (t_n, \delta_n)) = \frac{\prod_{k=1}^l \left\{ 1 - \exp \left[ - \left( \frac{t_k}{\eta} \right)^\beta \right] \right\} \prod_{k=l+1}^n \exp \left[ - \left( \frac{t_k}{\eta} \right)^\beta \right]}{K * \eta} \mathbb{I}_{\{\beta \in ]1, 4[ \}} \quad (10)$$

**Proposition 6 :** *Si on suppose qu'il existe au moins une censure à gauche et une censure à droite ( $l \geq 1$  et  $l < n$ ), alors la distribution a posteriori (10) de  $(\eta, \beta)$  en utilisant la loi a priori de Jeffreys pour  $\eta$  et une loi uniforme sur  $[1, 4]$  n'est pas une loi impropre.*

**Preuve :** On cherche à savoir si cette double intégrale est finie :

$$K = \int_1^4 \int_0^{+\infty} \frac{1}{\eta} \prod_{k=1}^l \left\{ 1 - \exp \left[ - \left( \frac{t_k}{\eta} \right)^\beta \right] \right\} \prod_{k=l+1}^n \exp \left[ - \left( \frac{t_k}{\eta} \right)^\beta \right] d\eta d\beta$$

La fonction à intégrer est continue sur le domaine  $]0, +\infty[ \times ]1, 4[$ . Il faut donc s'intéresser à l'intégrale impropre en  $\eta$ . On a :

$$\lim_{\eta \rightarrow 0} \mathcal{L}_n(\eta, \beta / (t_1, \delta_1), \dots, (t_n, \delta_n)) = 0$$

Le premier produit étant équivalent à 1 et le deuxième tendant vers 0, grâce à la prédominance de l'exponentielle sur la puissance, la fonction  $\mathcal{L}_n$  est prolongeable par continuité en 0. Le comportement asymptotique de la distribution a posteriori est :

$$\mathcal{L}_n(\eta, \beta / (t_1, \delta_1), \dots, (t_n, \delta_n)) \sim \frac{\prod_{k=1}^l t_k^\beta}{\eta^{\beta l + 1}}$$

Or  $\beta l > 1$ , l'intégrale est donc convergente, et par conséquent la loi a posteriori n'est pas impropre. ■

On est maintenant en mesure de calculer les lois marginales a posteriori, et donc de déterminer les estimateurs de Bayes.

$$\hat{\eta}_{bayes} = \frac{\int_1^4 \int_0^{+\infty} \prod_{k=1}^l F(t_k) \prod_{k=l+1}^n R(t_k) d\eta d\beta}{\int_1^4 \int_0^{+\infty} \frac{1}{\eta} \prod_{k=1}^l F(t_k) \prod_{k=l+1}^n R(t_k) d\eta d\beta}$$

$$\hat{\beta}_{bayes} = \frac{\int_1^4 \int_0^{+\infty} \frac{\beta}{\eta} \prod_{k=1}^l F(t_k) \prod_{k=l+1}^n R(t_k) d\eta d\beta}{\int_1^4 \int_0^{+\infty} \frac{1}{\eta} \prod_{k=1}^l F(t_k) \prod_{k=l+1}^n R(t_k) d\eta d\beta}$$

Ces intégrales seront aussi calculées par simulation de Monte-Carlo, comme dans les cas non informatif.

## 5 Applications et simulations

Les méthodes présentées précédemment ont été appliquées pour des données simulées, pour se donner une idée de la qualité de ces estimateurs. Puis, nous les avons appliqués sur les composants des pompes primaires. Les programmes d'optimisation ont été effectués sous Matlab, ainsi que le calcul des informations de Fisher et des intervalles de confiance. Dans tout la suite  $\tau$  désigne le pourcentage de censures à gauche présentes dans l'échantillon.

### 5.1 Simulations

Afin d'être dans le même cadre que nos données réelles, nous nous sommes servis des temps d'observations  $t_1, \dots, t_n$  du composant 1,  $n = 168$ , et du composant 7,  $n = 28$ .

#### 5.1.1 Estimateurs du maximum de vraisemblance

##### Pour la loi exponentielle :

Nous avons simulé plusieurs lois exponentielles, afin de vérifier la validité de nos méthodes d'estimation.

Les étapes sont les suivantes :

- \* On simule  $X_1, \dots, X_n$ , représentant les durées de vie suivant une loi exponentielle de paramètre  $\theta_0$  connu,
- \* on compare les  $x_i$  et les  $t_i$  pour tout  $i \in \ll 1, n \gg$ ,
- \* si  $x_i \leq t_i$  alors  $\delta_i = 1$ , sinon  $\delta_i = 0$ .

Les observations sont  $(t_i, \delta_i)_{1 \leq i \leq n}$ . Nous sommes alors dans les mêmes conditions que les données réelles. Les résultats sont donnés dans le tableau ci-après :

Simulations	$EMV_1$	$EMV_2$	$EMV_3$	$EMV$	$\tau$
$Exp(10^{-5})$	$0,92 \times 10^{-5}$	$0,99 \times 10^{-5}$ $(0,77 \times 10^{-7})$	$1,08 \times 10^{-5}$ $(2,00 \times 10^{-7})$	$1,10 \times 10^{-5}$	24,4 %
$Exp(1,4 \cdot 10^{-5})$	$1,00 \times 10^{-5}$	$1,09 \times 10^{-5}$ $(0,86 \times 10^{-7})$	$1,19 \times 10^{-5}$ $(2,05 \times 10^{-7})$	$1,21 \times 10^{-5}$	26,8 %
$Exp(2 \cdot 10^{-5})$	$1,43 \times 10^{-5}$	$1,61 \times 10^{-5}$ $(1,61 \times 10^{-7})$	$1,84 \times 10^{-5}$ $(4,16 \times 10^{-7})$	$1,90 \times 10^{-5}$	38,1 %
$Exp(2,6 \cdot 10^{-5})$	$1,45 \times 10^{-5}$	$1,64 \times 10^{-5}$ $(1,62 \times 10^{-7})$	$1,89 \times 10^{-5}$ $(4,45 \times 10^{-7})$	$1,96 \times 10^{-5}$	38,7 %
$Exp(3,4 \cdot 10^{-5})$	$2,17 \times 10^{-5}$	$2,59 \times 10^{-5}$ $(3,11 \times 10^{-7})$	$3,21 \times 10^{-5}$ $(9,83 \times 10^{-7})$	$3,54 \times 10^{-5}$	57,7 %

TAB. 2: EMV et méthodes actuariales pour des lois exponentielles simulées avec n=168.

On estime le paramètre d'intérêt, qui pour une loi exponentielle, représente le taux de défaillance.

On notera  $EMV_1$ , l'estimateur du maximum de vraisemblance dans le cas uniquement censuré à droite ;  $EMV_2$  lorsque les censures sont remplacées par des temps de défaillance tirés sur  $[t_k/2, t_k]$  ;  $EMV_3$  pour des temps tirés sur  $[0, t_k]$  ; et enfin  $EMV$  pour le cas doublement censuré.

On remarque que les résultats sont cohérents. En effet, on a la relation suivante :

$$EMV_1 < EMV_2 < EMV_3,$$

ce que nous avons présentés auparavant. En revanche, l'estimateur du maximum de vraisemblance dans le cas doublement censuré est supérieur à l'estimateur que nous avons désigné de plus "pessimiste".

L'estimateur du maximum de vraisemblance semble être le meilleur estimateur des quatre présentés ici. Les méthodes actuariales présente donc peu d'intérêt dans ce cas là.

Ci-après, les intervalles de confiances pour les mêmes lois simulées :

Simulations	$EMV$	Intervalle de confiance	$\tau$
$Exp(10^{-5})$	$1, 10 \times 10^{-5}$	$[0, 75 \times 10^{-5}; 1, 44 \times 10^{-5}]$	24,4 %
$Exp(1, 4.10^{-5})$	$1, 21 \times 10^{-5}$	$[0, 85 \times 10^{-5}; 1, 56 \times 10^{-5}]$	26,8 %
$Exp(2.10^{-5})$	$1, 90 \times 10^{-5}$	$[1, 42 \times 10^{-5}; 2, 39 \times 10^{-5}]$	38,1 %
$Exp(2, 6.10^{-5})$	$1, 96 \times 10^{-5}$	$[1, 46 \times 10^{-5}; 2, 45 \times 10^{-5}]$	38,7 %
$Exp(3, 4.10^{-5})$	$3, 54 \times 10^{-5}$	$[2, 79 \times 10^{-5}; 4, 29 \times 10^{-5}]$	57,7 %

TAB. 3: Simulations et Intervalles de confiance pour la loi exponentielle avec  $n=168$ .

L'estimation ponctuel du paramètre semble correct dans la plupart des cas, excepté pour le cinquième modèle ( $Exp(2, 6.10^{-5})$ ). En revanche, les intervalles de confiance ne sont pas satisfaisants. En effet, pour le dernier modèle simulé, cela correspond pour le temps moyen de bon fonctionnement avant défaillance (ici dégradation), noté MTBF, à un intervalle [22000 heures, 40000 heures] pour un seuil de confiance à 95 % , malgré une taille d'échantillon honorable ( $n = 168$ ).

Remarque : La variance asymptotique théorique de l'estimateur est croissante avec la valeur de l'estimateur. Ainsi, plus le taux de censures à gauche sera important et plus la

variance de l'estimateur du maximum de vraisemblance sera grande.

Pour vérifier la stabilité de l'estimateur du maximum de vraisemblance pour des données doublement censurées, nous avons itéré 100 fois le procédé. Ainsi, ce bootstrap paramétrique nous donne un écart type empirique que l'on pourra comparer à l'écart type asymptotique théorique.

Simulations	$EMV$ moyen	Écart type asymptotique	$\tau$ moyen
$Exp(10^{-5})$	$1,03 \times 10^{-5}$ ( $1,63 \times 10^{-6}$ )	$1,78 \times 10^{-6}$	23,3 %
$Exp(1,4 \cdot 10^{-5})$	$1,39 \times 10^{-5}$ ( $1,80 \times 10^{-6}$ )	$1,79 \times 10^{-6}$	29,9 %
$Exp(2 \cdot 10^{-5})$	$2,00 \times 10^{-5}$ ( $2,40 \times 10^{-6}$ )	$2,45 \times 10^{-6}$	39,5 %
$Exp(2,6 \cdot 10^{-5})$	$2,63 \times 10^{-5}$ ( $2,84 \times 10^{-6}$ )	$2,55 \times 10^{-6}$	47,8 %
$Exp(3,4 \cdot 10^{-5})$	$3,43 \times 10^{-5}$ ( $4,01 \times 10^{-6}$ )	$3,83 \times 10^{-6}$	57,7 %

TAB. 4: Simulations et écarts type asymptotiques de loi exponentielle pour  $n=168$ .

On peut remarquer d'après ce tableau que l'estimateur du maximum de vraisemblance possède de bonnes propriétés, avec une taille d'échantillon honorable ( $n = 168$ ). L'écart type calculé à partir des estimations est très proche de l'écart type théorique. Ce qui nous fait penser que nous pouvons considérer que les calculs asymptotiques sont valables pour cette valeur de  $n$ .

Nous allons maintenant étudier le comportement de l'estimateur du maximum de vraisemblance en fonction de la taille de l'échantillon.

Simulations	$EMV$ moyen	Écart type asymptotique	$\tau$ moyen
$Exp(10^{-5})$	$1,02 \times 10^{-5}$ ( $5,08 \times 10^{-6}$ )	$3,22 \times 10^{-6}$	17,9 %
$Exp(1,4 \cdot 10^{-5})$	$1,40 \times 10^{-5}$ ( $5,47 \times 10^{-6}$ )	$3,66 \times 10^{-6}$	23,7 %
$Exp(2 \cdot 10^{-5})$	$2,03 \times 10^{-5}$ ( $6,79 \times 10^{-6}$ )	$1,22 \times 10^{-5}$	32,4 %
$Exp(2,6 \cdot 10^{-5})$	$2,69 \times 10^{-5}$ ( $7,97 \times 10^{-6}$ )	$8,04 \times 10^{-6}$	40,2 %
$Exp(3,4 \cdot 10^{-5})$	$3,41 \times 10^{-5}$ ( $0,90 \times 10^{-5}$ )	$1,29 \times 10^{-5}$	47,7 %

TAB. 5: Simulations et écarts type asymptotiques de loi exponentielle pour  $n=28$ .

Avec la baisse de la taille de l'échantillon ( $n = 28$ ), on remarque que l'estimateur du maximum de vraisemblance est toujours performant. Ceci étant dit, l'écart type empirique et théorique est bien supérieur au cas où  $n = 168$ . Les intervalles de confiance ne peuvent donc être pris en compte dans ce cas précis (nous ne sommes en aucun cas dans des conditions asymptotiques!).

### Quelques conclusions sur l'estimateur du maximum de vraisemblance pour le modèle exponentiel :

L'estimation ponctuel semble, pour ce modèle, assez stable, et peu perturbé par une faible taille d'échantillon. En revanche, l'écart type asymptotique empirique ou théorique peut être assez important, soit pour des valeurs du paramètres importantes, soit pour une taille d'échantillon faible. En conclusion, ce modèle peut nous satisfaire, en ayant tout de fois quelques réserves.

Cependant en fiabilité industrielle, l'intérêt d'une étude comme la notre, est la connaissance du vieillissement ou non du composant, ce que le modèle exponentielle ne prend pas en compte. Nous allons donc maintenant étudier le modèle de Weibull.

**Pour une loi de Weibull :**

Nous avons simulé plusieurs lois de Weibull à deux paramètres, et procédé de la même façon que pour le modèle exponentiel. Les résultats sont donnés dans le tableau ci-après :

$W(\eta_0; \beta_0)$	$EMV_1$	$EMV_2$	$EMV_3$	$EMV$	$\tau$
$\eta_0 = 80000$ $\beta_0 = 0,6$	$\hat{\eta} = 45238$ $\hat{\beta} = 2,51$	$\hat{\eta} = 44539 (407)$ $\hat{\beta} = 1,87 (0.05)$	$\hat{\eta} = 57655 (3827)$ $\hat{\beta} = 1,00 (0.10)$	$\hat{\eta} = 217767$ $\hat{\beta} = 0,35$	37,5 %
$\eta_0 = 60000$ $\beta_0 = 0,8$	$\hat{\eta} = 44356$ $\hat{\beta} = 2,64$	$\hat{\eta} = 42054 (509)$ $\hat{\beta} = 2,03 (0.06)$	$\hat{\eta} = 52501 (3307)$ $\hat{\beta} = 1,06 (0.10)$	$\hat{\eta} = 77461$ $\hat{\beta} = 0,63$	38,7 %
$\eta_0 = 40000$ $\beta_0 = 1$	$\hat{\eta} = 41222$ $\hat{\beta} = 2,86$	$\hat{\eta} = 36186 (569)$ $\hat{\beta} = 2,33 (0.07)$	$\hat{\eta} = 36686 (1395)$ $\hat{\beta} = 1,18 (0.11)$	$\hat{\eta} = 37663$ $\hat{\beta} = 1,16$	46,4 %
$\eta_0 = 30000$ $\beta_0 = 1,5$	$\hat{\eta} = 40002$ $\hat{\beta} = 2,86$	$\hat{\eta} = 34521 (524)$ $\hat{\beta} = 2,34 (0.07)$	$\hat{\eta} = 34585 (972)$ $\hat{\beta} = 1,20 (0.11)$	$\hat{\eta} = 33376$ $\hat{\beta} = 1,19$	50,6 %
$\eta_0 = 20000$ $\beta_0 = 2$	$\hat{\eta} = 35304$ $\hat{\beta} = 2,82$	$\hat{\eta} = 27662 (532)$ $\hat{\beta} = 2,56 (0.11)$	$\hat{\eta} = 21415 (763)$ $\hat{\beta} = 1,38 (0.11)$	$\hat{\eta} = 19808$ $\hat{\beta} = 2,26$	72 %

TAB. 6: EMV et méthodes actuariales pour des lois de Weibull simulées avec n=168.

L'estimateur du maximum de vraisemblance est, comme pour le modèle exponentiel, le meilleur des quatre estimateurs proposés ici. Néanmoins, il existe une difficulté pour cet estimateur, à bien estimer le paramètre de forme  $\beta$ , lorsque celui-ci est inférieur à 1. Lorsque  $\beta > 1$ , l'estimateur du maximum de vraisemblance a un bon comportement.

Nous donnons ci-après les intervalles de confiance au seuil 95 % :

Simulations	$EMV$	Intervalle de confiance	$\tau$
$\eta_0 = 80000$ $\beta_0 = 0,6$	$\hat{\eta} = 217767$ $\hat{\beta} = 0,35$	$IC = [61145; 374389]$ $IC = [0,26; 0,43]$	37,5 %
$\eta_0 = 60000$ $\beta_0 = 0,8$	$\hat{\eta} = 77461$ $\hat{\beta} = 0,63$	$IC = [44352; 110571]$ $IC = [0,55; 0,71]$	38,7 %
$\eta_0 = 40000$ $\beta_0 = 1$	$\hat{\eta} = 37663$ $\hat{\beta} = 1,16$	$IC = [27623; 47703]$ $IC = [1,12; 1,19]$	46,4 %
$\eta_0 = 30000$ $\beta_0 = 1,5$	$\hat{\eta} = 33376$ $\hat{\beta} = 1,19$	$IC = [26380; 40374]$ $IC = [0,99; 1,38]$	50,6 %
$\eta_0 = 20000$ $\beta_0 = 2$	$\hat{\eta} = 19808$ $\hat{\beta} = 2,26$	$IC = [18462; 21155]$ $IC = [1,53; 2,98]$	72 %

TAB. 7: Intervalle de confiance pour l'EMV pour la loi de Weibull avec n=168.

Le modèle exponentiel étant un cas particulier d'un modèle de Weibull ( $\beta = 1$  et  $\eta = 1/\theta$ ), il est possible de faire un parallèle. Ainsi, la variance asymptotique sur  $\eta$ , comme dans le cas exponentiel, croît avec la valeur de celui-ci. Pour  $\eta = 20000$ , l'intervalle de confiance est bon, alors que pour  $\eta = 80000$ , il devient catastrophique. Pour  $\beta$ ,

sachant qu'il est très important pour nous de connaître la position de  $\beta$  par rapport à 1, les intervalles de confiances nous donnent une réponse correcte dans les 5 cas simulés.

Nous allons maintenant étudier la variance asymptotique empirique de l'estimateur du maximum de vraisemblance pour le modèle de Weibull :

Simulations	$EMV$ moyen	Écart type asymptotique	$\tau$ moyen
$\eta_0 = 80000$ $\beta_0 = 0,6$	n'existe pas $\hat{\beta} = 0,57 (0,27)$	$+\infty$ 0.07	38,5 %
$\eta_0 = 60000$ $\beta_0 = 0,8$	$\hat{\eta} = 99148 (175443)$ $\hat{\beta} = 0,81 (0,31)$	17027 0.14	39,4 %
$\eta_0 = 40000$ $\beta_0 = 1$	$\hat{\eta} = 45000 (24179)$ $\hat{\beta} = 0,81 (0.28)$	6660 0,20	46,5 %
$\eta_0 = 30000$ $\beta_0 = 1,5$	$\hat{\eta} = 30267 (2764)$ $\hat{\beta} = 1,48 (0.27)$	3867 0,27	52,8 %
$\eta_0 = 20000$ $\beta_0 = 2$	$\hat{\eta} = 19955 (1208)$ $\hat{\beta} = 2,13 (0.45)$	931 0,28	71,3 %

TAB. 8: Simulations et écarts type asymptotiques pour la loi de Weibull pour  $n=168$ .

Il existe une grande difficulté d'estimer correctement les paramètres lorsque  $\beta < 1$ . En effet, il existe des cas où la méthode d'estimation donne  $\eta = +\infty$  et  $\beta = 0$  (4 % des cas pour la première loi simulée). Les méfiances théoriques que nous avons pour ce modèle sur la normalité asymptotique de l'estimateur (et notamment sur la frontière de  $\Theta$ ) semble se confirmer ici. Pour  $\beta > 1$ , l'estimation est correct.

Pour étudier l'effet de la taille de l'échantillon, nous nous sommes placés dans le même cas que le composant 7, à savoir  $n=28$ .

Simulations	$EMV$ moyen	Écart type asymptotique	$\tau$ moyen
$\eta_0 = 80000$ $\beta_0 = 0,6$	n'existe pas $\hat{\beta} = 0,78 (0,96)$	515218 0.07	34 %
$\eta_0 = 60000$ $\beta_0 = 0,8$	n'existe pas $\hat{\beta} = 1,32 (1,29)$	7953 0.31	32,6 %
$\eta_0 = 40000$ $\beta_0 = 1$	n'existe pas $\hat{\beta} = 1,24 (1,14)$	31449 0,42	38,25 %
$\eta_0 = 30000$ $\beta_0 = 1,5$	n'existe pas $\hat{\beta} = 2,40 (4,24)$	118788 0,53	41,8 %
$\eta_0 = 20000$ $\beta_0 = 2$	$\hat{\eta} = 19810 (3182)$ $\hat{\beta} = 2,39 (1,38)$	3582 1,02	60,8 %

TAB. 9: Simulations et écarts type asymptotiques pour la loi de Weibull avec  $n=28$ .

À  $n=28$ , l'estimation du paramètre d'échelle devient quasiment impossible (excepté pour des faibles valeurs de  $\eta$  et un  $\beta$  plus grand que 1).  $\beta$  est légèrement surestimé, mais ne peut constituer, eu égard à son écart type, une estimation valable (par exemple pour  $W(30000, 1.5)$ , l'écart type sur  $\beta$  est de 4,24).

### Quelques conclusions sur l'estimateur du maximum de vraisemblance pour le modèle de Weibull :

Pour ce modèle, l'estimateur du maximum de vraisemblance a montré plusieurs difficultés à bien estimer les paramètres, principalement  $\eta$ . L'estimation sur  $\beta$  est meilleur quand celui-ci est plus grand que 1. Ce modèle a également semblé moins stable quand la taille de l'échantillon diminuait. Cependant, le point positif est que ce modèle nous permet a priori de déterminer correctement si le composant vieillit ou pas.

### 5.1.2 Estimateurs bayésiens

#### Pour le modèle exponentiel :

Ces calculs ont été effectués par des simulations de Monte-Carlo avec  $N = 2000$ .

Simulations	m et $\sigma$ a priori	Bayes informatif	Bayes non informatif	tau
$Exp(10^{-5})$	$m = 1,6 \times 10^{-5}$ $\sigma = 0,4 \times 10^{-5}$	$1,19 \times 10^{-5}$ $(3,94 \times 10^{-6})$	$1,47 \times 10^{-5}$ $(1,71 \times 10^{-6})$	22,6 %
$Exp(1,4 \cdot 10^{-5})$	$m = 1,6 \times 10^{-5}$ $\sigma = 0,4 \times 10^{-5}$	$1,49 \times 10^{-5}$ $(3,24 \times 10^{-6})$	$1,92 \times 10^{-5}$ $(2,16 \times 10^{-6})$	32,1 %
$Exp(2 \cdot 10^{-5})$	$m = 1,6 \times 10^{-5}$ $\sigma = 0,4 \times 10^{-5}$	$1,78 \times 10^{-5}$ $(5,61 \times 10^{-7})$	$2,22 \times 10^{-5}$ $(2,35 \times 10^{-6})$	37,5 %
$Exp(2,6 \cdot 10^{-5})$	$m = 1,6 \times 10^{-5}$ $\sigma = 0,4 \times 10^{-5}$	$2,05 \times 10^{-5}$ $(9,70 \times 10^{-6})$	$2,81 \times 10^{-5}$ $(3,02 \times 10^{-6})$	46,4 %
$Exp(3,4 \cdot 10^{-5})$	$m = 1,6 \times 10^{-5}$ $\sigma = 0,4 \times 10^{-5}$	$3,35 \times 10^{-5}$ $(6,08 \times 10^{-6})$	$3,16 \times 10^{-5}$ $(1,45 \times 10^{-6})$	58,3 %

TAB. 10: Estimateurs bayésien pour une loi exponentielle avec  $n = 168$ .

On peut remarquer tout d'abord la cohérence des résultats. En effet, l'estimateur de Bayes informatif est toujours compris entre l'estimateur du maximum de vraisemblance et la moyenne a priori donnée par l'expert. L'estimateur de Bayes non informatif a tendance à surestimer le paramètre d'intérêt. Pour ces deux estimateurs l'écart type de la loi a posteriori, donné entre parenthèse, semble convenable pour ce modèle. Néanmoins, comme l'estimateur du maximum de vraisemblance possède de bonnes propriétés pour ce modèle, l'intérêt du bayésien ne subsiste que dans l'apport d'information d'experts.

**Pour le modèle de Weibull :**

Nous avons simulé plusieurs lois de Weibull à deux paramètres, et procédé de la même façon que le modèle exponentiel. Les moyennes a priori pour l'estimation dite informative sont de 30000, pour  $\eta$ , et de 1,18 pour  $\beta$ . On n'impose donc très peu de contrainte sur le paramètre d'échelle. Le paramètre de forme possède une moyenne a priori supérieure à 1, ce qui impose un léger vieillissement a priori. Les résultats sont donnés dans le tableau ci-dessous :

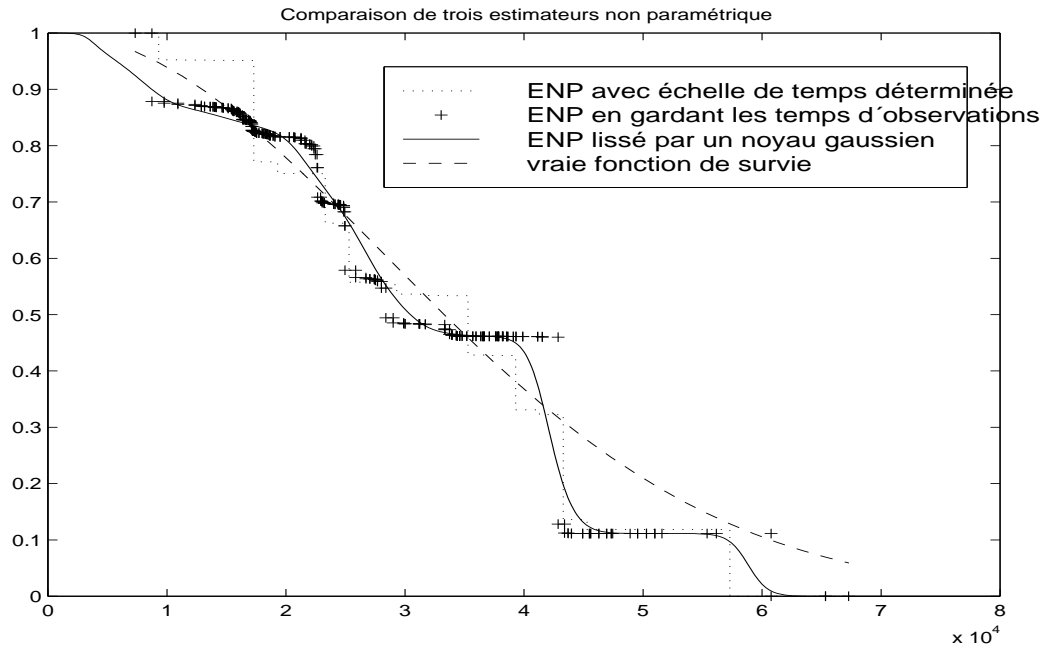
Simulation	Bayes non informatif	Bayes semi informatif	Bayes informatif	$\tau$
$\eta_0 = 80000$ $\beta_0 = 0,6$	$\hat{\eta} = 52996$ (30528) $\hat{\beta} = 0,78$ (0,13)	$\hat{\eta} = 42782$ (14909) $\hat{\beta} = 1,12$ (0,10)	$\hat{\eta} = 47094$ (28989) $\hat{\beta} = 0,88$ (0,16)	44 %
$\eta_0 = 40000$ $\beta_0 = 1$	$\hat{\eta} = 38637$ (14988) $\hat{\beta} = 1,25$ (0,30)	$\hat{\eta} = 36746$ (38907) $\hat{\beta} = 1,40$ (0,24)	$\hat{\eta} = 37269$ (14387) $\hat{\beta} = 1,26$ (0,22)	45,8 %
$\eta_0 = 20000$ $\beta_0 = 2$	$\hat{\eta} = 19410$ (34148) $\hat{\beta} = 1,98$ (0,35)	$\hat{\eta} = 19527$ (51281) $\hat{\beta} = 2,06$ (0,35)	$\hat{\eta} = 18955$ (43536) $\hat{\beta} = 1,65$ (0,27)	34 %

TAB. 11: Estimateurs bayésiens pour des lois simulées de Weibull avec n=168.

L'estimateur de Bayes non informatif semble avoir le meilleur comportement. En effet, même si l'écart type de la loi a posteriori donné entre parenthèse est élevé, il estime correctement les deux paramètres. Le Bayes semi informatif ne permet guère d'améliorer l'estimation. Ces trois estimateurs ne semblent donc pas être meilleur que l'estimateur du maximum de vraisemblance, le bayésien ne subsistera alors que dans l'apport d'information a priori.

### 5.1.3 Estimateurs non paramétrique

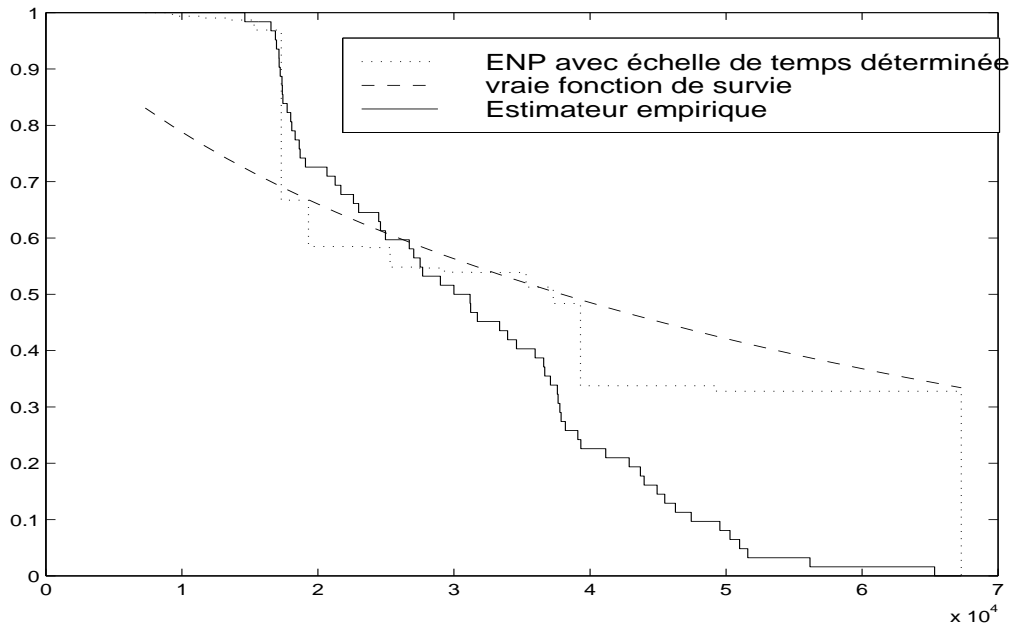
On étudie ici l'estimateur de Turnbull, puis cet estimateur lissé par un noyau gaussien (de fenêtre  $h = 2000$ ) avec une taille d'échantillon égale à 168 :



On a simulé une Weibull  $W(40000; 2)$ . On veut comparer trois estimateurs non paramétriques de la fonction de survie. Nous avons pu construire deux estimateurs de Turnbull. Pour l'un, on a gardé les temps d'observation du composant 1. Pour l'autre, on s'est fixé une échelle de temps déterminé. Cette dernière méthode présente l'inconvénient d'une perte d'information. L'estimateur, en gardant les temps d'observation, a été lissé par un noyau gaussien, afin de régulariser l'estimateur.

On peut remarquer que pendant la vie utile du composant (jusqu'à 60000 heures) ces trois estimateurs sont proches de la vraie fonction de survie.

On étudie ici l'estimateur de Turnbull, et l'estimateur empirique avec une taille d'échantillon égale à 28 :



On a simulé une Weibull  $W(60000; 0, 8)$ . On remarque que comme dans le cas paramétrique, l'estimation se détériore lorsque le paramètre  $\beta$  est inférieur à 1. De plus, l'estimateur non paramétrique en gardant les temps d'observation n'existe pas (l'algorithme ne converge pas!).

## 5.2 Applications

### 5.2.1 Estimateurs du maximum de vraisemblance

Pour le modèle exponentiel :

Nous allons ici donner les applications numériques sur les composants des pompes primaires. Nous étudions tout d'abord le maximum de vraisemblance dans le cas simplement censuré, puis doublement censuré, et enfin les deux méthodes actuariales. Pour les deux méthodes actuariales ( $EMV_2$  et  $EMV_3$ ), nous itérons 500 fois le procédé, pour nous donner 500 estimateurs. Nous ne donnons ici que la moyenne de ces estimateurs, ainsi que l'écart type sur cet échantillon d'estimateurs. Ce dernier nous renseigne qualitativement sur la stabilité de cette méthode d'estimation. Il n'est en aucun cas l'écart type théorique ou empirique de l'estimateur.

Composant	$EMV_1$	$EMV_2$	$EMV_3$	$EMV$	$\tau$
c1	$1,52 \times 10^{-5}$	$1,71 \times 10^{-5}$ ( $1,66 \times 10^{-7}$ )	$1,96 \times 10^{-5}$ ( $4,36 \times 10^{-7}$ )	$2,03 \times 10^{-5}$	40,5 %
c2	$1,23 \times 10^{-5}$	$1,32 \times 10^{-5}$ ( $1,33 \times 10^{-7}$ )	$1,43 \times 10^{-5}$ ( $3,06 \times 10^{-7}$ )	$1,44 \times 10^{-5}$	26,4 %
c3	$6,56 \times 10^{-7}$	$6,59 \times 10^{-7}$ ( $1,30 \times 10^{-9}$ )	$6,61 \times 10^{-7}$ ( $2,59 \times 10^{-9}$ )	$6,61 \times 10^{-7}$	1.4 %
c4	$2,79 \times 10^{-5}$	$3,30 \times 10^{-5}$ ( $5,62 \times 10^{-7}$ )	$4,06 \times 10^{-5}$ ( $1,77 \times 10^{-7}$ )	$4,40 \times 10^{-5}$	59.8 %
c5	$2,07 \times 10^{-5}$	$2,33 \times 10^{-5}$ ( $3,18 \times 10^{-7}$ )	$2,67 \times 10^{-5}$ ( $7,88 \times 10^{-7}$ )	$2,75 \times 10^{-5}$	44.6 %
c6	0	0	0	0	0 %
c7	$8,92 \times 10^{-6}$	$9,38 \times 10^{-6}$ ( $1,31 \times 10^{-7}$ )	$9,90 \times 10^{-6}$ ( $2,85 \times 10^{-7}$ )	$9,93 \times 10^{-6}$	17.8 %

TAB. 12: Estimateurs du maximum de vraisemblance pour la loi exponentielle.

Le tableau nous rappelle en dernière colonne, par  $\tau$ , le taux de censures à gauche dans l'échantillon.

Comme pour les simulations, l'ordre entre ces quatre estimateurs est bien respecté ( $EMV_1 \leq EMV_2 \leq EMV_3 \leq EMV$ ). De plus, on peut remarquer que l'EMV crît avec le pourcentage de censures à gauche. En effet, l'estimateur représente, pour ce modèle, le taux de défaillance, qui est grand quand le nombre de dégradations présentes dans l'échantillon est élevé, et est nulle (composant 6) quand le taux est nul. Ce modèle est donc satisfaisant pour toutes ces raisons. Cela étant dit, nous devons nous intéresser à un modèle plus complexe qu'est le modèle de Weibull.

D'après l'équation (8), nous donnons les intervalles de confiance pour l'estimateur du maximum de vraisemblance dans le cas doublement censuré, pour un niveau de confiance à 95 %.

Composant	$EMV$	Intervalle de confiance	$\tau$
c1	$2,03 \times 10^{-5}$	$[1,54 \times 10^{-5}; 2,52 \times 10^{-5}]$	40,5 %
c2	$1,44 \times 10^{-5}$	$[0,79 \times 10^{-5}; 2,09 \times 10^{-5}]$	26,4 %
c3	$6,61 \times 10^{-7}$	$[0; 19,56 \times 10^{-7}]$	1.4 %
c4	$4,40 \times 10^{-5}$	$[3,02 \times 10^{-5}; 5,78 \times 10^{-5}]$	59.8 %
c5	$2,75 \times 10^{-5}$	$[1,80 \times 10^{-5}; 3,69 \times 10^{-5}]$	44.6 %
c6	0	$[0; +\infty]$	0 %
c7	$9,93 \times 10^{-6}$	$[1,19 \times 10^{-6}; 18,66 \times 10^{-6}]$	17.8 %

TAB. 13: Intervalle de confiance pour l'EMV pour la loi exponentielle.

Les intervalles de confiances sont relativement bons lorsque le taux de censures à gauche est assez important, c'est-à-dire lorsque l'information apporté par les observations est suffisantes. Pour les composants ne présentant que très peu de dégradations, le taux de défaillance est très inférieur au autres composants.

### Pour le modèle de Weibull

Nous allons dans ce paragraphe présenter les résultats pour les mêmes méthodes pour la loi de Weibull. Nous gardons ici les même notations.

	$EMV_1$	$EMV_2$	$EMV_3$	$EMV$	$\tau$
c1	$\hat{\eta} = 43525$ $\hat{\beta} = 2,68$	$\hat{\eta} = 40504 (492)$ $\hat{\beta} = 2,09 (0.06)$	$\hat{\eta} = 48611 (2594)$ $\hat{\beta} = 1,08 (0.10)$	$\hat{\eta} = 62067$ $\hat{\beta} = 0,72$	40,5 %
c2	$\hat{\eta} = 31945$ $\hat{\beta} = 4,65$	$\hat{\eta} = 36066 (938)$ $\hat{\beta} = 2,57 (0.20)$	$\hat{\eta} = 67482 (16237)$ $\hat{\beta} = 1,10 (0.23)$	$\hat{\eta} = 98226$ $\hat{\beta} = 0,77$	26,4 %
c3	$\hat{\eta} = 68080$ $\hat{\beta} = 4,06$	$\hat{\eta} = 201844 (137550)$ $\hat{\beta} = 2,38 (0.77)$	n'existe pas n'existe pas	$\hat{\eta} \geq \times 10^{16}$ $\hat{\beta} = 0,15$	1,4 %
c4	$\hat{\eta} = 26805$ $\hat{\beta} = 4,52$	$\hat{\eta} = 23460 (376)$ $\hat{\beta} = 2,97 (0.16)$	$\hat{\eta} = 23170 (678)$ $\hat{\beta} = 1,24 (0.16)$	$\hat{\eta} = 22759$ $\hat{\beta} = 0,97$	59,8 %
c5	$\hat{\eta} = 28900$ $\hat{\beta} = 4,20$	$\hat{\eta} = 28725 (216)$ $\hat{\beta} = 2,45 (0.14)$	$\hat{\eta} = 36005 (2572)$ $\hat{\beta} = 1,10 (0.16)$	$\hat{\eta} = 237214$ $\hat{\beta} = 0,22$	44,6 %
c6	n'existe pas n'existe pas	n'existe pas n'existe pas	n'existe pas n'existe pas	n'existe pas n'existe pas	0 %
c7	$\hat{\eta} = 32418$ $\hat{\beta} = 4,63$	$\hat{\eta} = 39564 (3426)$ $\hat{\beta} = 2,64 (0.43)$	$\hat{\eta} = 113936 (110304)$ $\hat{\beta} = 1,21 (0.56)$	$\hat{\eta} = 211782$ $\hat{\beta} = 0,68$	17,8 %

TAB. 14: Estimateur du maximum de vraisemblance pour la loi de Weibull.

On peut remarquer que, comme dans le cas exponentiel, l'estimation semble plus réaliste, lorsque le taux de censures à gauche est important. Il apparaît que pour la loi de Weibull, la difficulté à estimer le paramètre de forme  $\beta$  est très importante. De plus, les méthodes actuariales nous montrent l'instabilité de l'estimation par maximum de vraisemblance.

Dans tous les cas présentés ci-dessus, l'EMV nous affirme que le paramètre de forme est inférieure à 1. Il nous semble cependant peu probable voire impossible que les composants se rajeunissent au cours du temps. D'après cette méthode, un modèle exponentiel semble être la meilleure modélisation.

Comme pour la loi exponentielle, nous donnons ci-après les intervalles de confiance au seuil 95 %.

Composant	$EMV$	Intervalle de confiance	$\tau$
c1	$\hat{\eta} = 62067$ $\hat{\beta} = 0,72$	$IC = [36690; 87444]$ $IC = [0, 67; 0, 77]$	40,5 %
c2	$\hat{\eta} = 98226$ $\hat{\beta} = 0,77$	$IC = [46229; 150223]$ $IC = [0, 65; 0, 88]$	26,4 %
c3	$\hat{\eta} \geq 10^{16}$ $\hat{\beta} = 0,15$	$IC = [0; +\infty]$ $IC = [0, 08; 0, 21]$	1.4 %
c4	$\hat{\eta} = 22759$ $\hat{\beta} = 0,97$	$IC = [15322; 30196]$ $IC = [0, 07; 1, 86]$	59.8 %
c5	$\hat{\eta} = 237314$ $\hat{\beta} = 0,22$	$IC = [0; 611938]$ $IC = [0, 09; 1, 33]$	44.6 %
c6	n'existe pas n'existe pas	n'existe pas n'existe pas	0 %
c7	$\hat{\eta} = 211782$ $\hat{\beta} = 0,68$	$IC = [0; 437385]$ $IC = [0, 45; 0, 91]$	17.8 %

TAB. 15: Intervalle de confiance pour l'EMV pour la loi de Weibull.

Ces résultats ne sont dans l'ensemble pas acceptables. En effet, l'estimateur du maximum de vraisemblance est biaisé pour des faibles échantillons, ou pour des échantillons faiblement censurés à gauche. Ainsi, les intervalles de confiance, calculés à partir de cet estimateur, ne peuvent être corrects. De plus, on remarque la difficulté de cette méthode à bien estimer les deux paramètres simultanément. Une bonne estimation de l'un entre eux se fait apparemment au détriment de l'autre (composant 4). Cependant, on constate, comme dans le cas exponentiel, que l'estimation semble stable quand le nombre de censures à gauche est important, comme par exemple pour le composant 1. Enfin, pour tous les composants, l'estimation du paramètre de forme  $\beta$  semble se faire avec très peu de liberté. Il serait peut être bon d'effectuer une optimisation composante par composante, c'est-à-dire d'optimiser d'abord suivant  $\beta$ , puis suivant  $\eta$ .

### 5.2.2 Estimateurs bayésiens

#### Pour la loi exponentielle :

On veut comparer l'estimateur du maximum de vraisemblance, l'estimateur de Bayes pour un a priori informatif, et l'estimateur de Bayes non informatif. Les lois a priori informatives sont des inverses gamma dont on définit dans la deuxième colonne la moyenne et l'écart type entre parenthèse.

Pour le calcul des estimateurs de Bayes, nous avons pris  $N = 2000$  simulations.

	moyenne a priori	Bayes informatif	Bayes non informatif	EMV	$\tau$
c1	$1,6 \times 10^{-5}$ ( $0,4 \times 10^{-5}$ )	$1,82 \times 10^{-5}$ ( $5,62 \times 10^{-6}$ )	$2,38 \times 10^{-5}$ ( $2,69 \times 10^{-6}$ )	$2,03 \times 10^{-5}$	40,5 %
c2	$1,6 \times 10^{-5}$ ( $0,4 \times 10^{-5}$ )	$1,48 \times 10^{-5}$ ( $3,52 \times 10^{-6}$ )	$2,37 \times 10^{-5}$ ( $3,80 \times 10^{-6}$ )	$1,44 \times 10^{-5}$	26,4 %
c3	$0,1 \times 10^{-5}$ ( $0,05 \times 10^{-5}$ )	$8,88 \times 10^{-7}$ ( $3,48 \times 10^{-7}$ )	$1,39 \times 10^{-5}$ ( $2,13 \times 10^{-6}$ )	$6,61 \times 10^{-7}$	1,4 %
c4	$1,6 \times 10^{-5}$ ( $0,4 \times 10^{-5}$ )	$5,48 \times 10^{-5}$ ( $3,14 \times 10^{-5}$ )	$3,72 \times 10^{-5}$ ( $1,13 \times 10^{-6}$ )	$4,40 \times 10^{-5}$	59,8 %
c5	$1,6 \times 10^{-5}$ ( $0,4 \times 10^{-5}$ )	$2,28 \times 10^{-5}$ ( $6,34 \times 10^{-6}$ )	$3,25 \times 10^{-5}$ ( $4,09 \times 10^{-6}$ )	$2,75 \times 10^{-5}$	44,6 %
c6	$0,1 \times 10^{-5}$ ( $0,05 \times 10^{-5}$ )	$0,1 \times 10^{-5}$ ( $0,05 \times 10^{-5}$ )	$3,77 \times 10^{-5}$ ( $2,19 \times 10^{-6}$ )	0	0 %
c7	$1,6 \times 10^{-5}$ ( $0,4 \times 10^{-5}$ )	$1,37 \times 10^{-5}$ ( $4,14 \times 10^{-6}$ )	$2,79 \times 10^{-5}$ ( $4,30 \times 10^{-6}$ )	$9,93 \times 10^{-6}$	17,8 %

TAB. 16: Comparaisons EMV et estimateurs bayésiens pour la loi exponentielle.

La première remarque sera la cohérence des résultats présentés. En effet, l'estimateur de Bayes informatif est toujours compris entre l'estimateur du maximum de vraisemblance et la moyenne a priori (excepté pour le composant 4). De plus, pour le composant 6 (qui ne présente aucune défaillance!) l'estimateur de Bayes est égal à la moyenne a priori. Ainsi, comme aucune information n'est apportée pour ce composant, l'estimateur "fait une entière confiance" à l'a priori que l'expert lui donne. On peut aussi remarquer que pour le composant 3 (1,4 % de censures à gauche, donc peu d'information sur le comportement du composant), il est très proche de la moyenne a priori donnée par l'expert.

L'estimateur de Bayes non informatif semble uniformiser les résultats. En effet, il est anormal de trouver  $\hat{\theta} = 3,77 \times 10^{-5}$  pour le taux de défaillance du composant 6, qui ne présente aucune dégradation. À ce propos, l'estimateur de Bayes informatif a un meilleur comportement.

### Pour la loi de Weibull

Le tableau ci-après représente les trois estimateurs bayésiens et l'estimateur du maximum de vraisemblance. La moyenne a priori sur  $\eta$  a été prise à 30000 pour tous les composants. La moyenne a priori sur  $\beta$  a, quant à elle, été prise à 1,22 pour tous les composants pour l'estimateur de Bayes informatif.

	Bayes non informatif	Bayes semi informatif	Bayes informatif	$\tau$
c1	$\eta = 59354$ (32207) $\beta = 0,82$ (0,17)	$\eta = 47175$ (27283) $\beta = 1,14$ (0,12)	$\eta = 51480$ (19261) $\beta = 0,95$ (0,96)	40,5 %
c2	$\eta = 96886$ (90854) $\beta = 0,98$ (0,39)	$\eta = 52135$ (12698) $\beta = 1,53$ (0,44)	$\eta = 60713$ (41322) $\beta = 1,18$ (1,21)	26,4 %
c3	$\eta = 181181$ (177240) $\beta = 2,71$ (1,54)	$\eta = 184670$ (177589) $\beta = 2,45$ (0,77)	$\eta = 213279$ (37019) $\beta = 1,61$ (1,63)	1,4 %
c4	$\eta = 23819$ (33595) $\beta = 0,93$ (0,35)	$\eta = 22864$ (46060) $\beta = 1,42$ (0,34)	$\eta = 22927$ (32291) $\beta = 1,07$ (1,10)	59,8 %
c5	$\eta = 42018$ (19630) $\beta = 0,88$ (0,22)	$\eta = 34417$ (33133) $\beta = 1,26$ (0,23)	$\eta = 37163$ (37034) $\beta = 0,97$ (1,00)	44,6 %
c6	$\eta = 73425$ (61440) $\beta = 3,48$ (2,74)	$\eta = 256663$ (250301) $\beta = 2,50$ (0,85)	$\eta = 29704$ (28962) $\beta = 1,19$ (0,85)	0 %
c7	$\eta = 281261$ (278922) $\beta = 1,08$ (0,66)	$\eta = 63701$ (37807) $\beta = 1,88$ (0,72)	$\eta = 67994$ (38880) $\beta = 1,29$ (0,31)	17,8 %

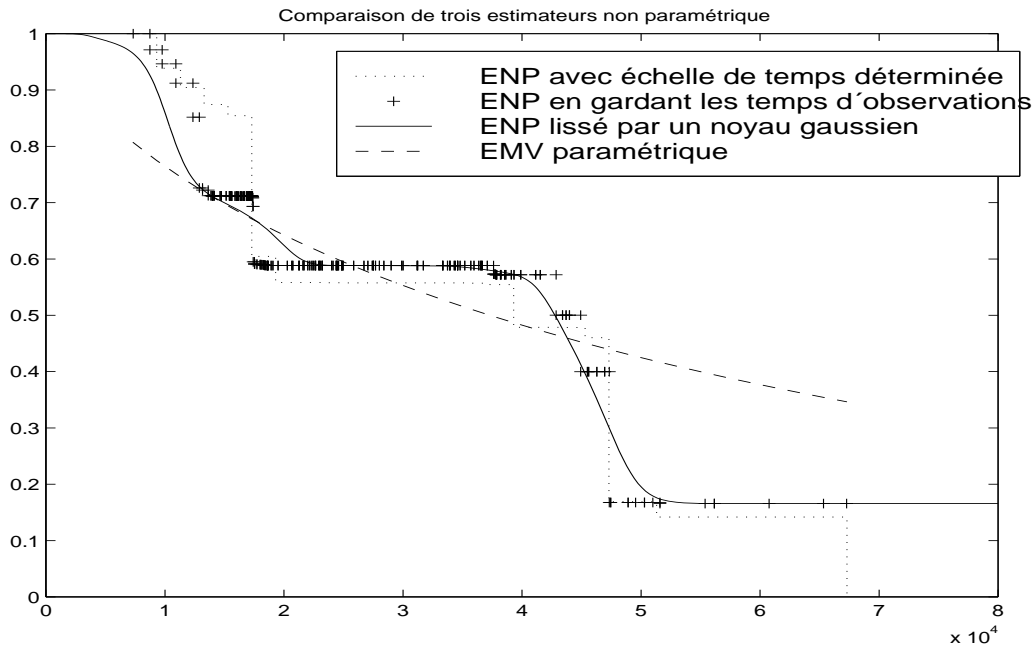
TAB. 17: Estimateurs bayésiens pour la loi de Weibull.

On peut remarquer que l'estimateur de Bayes semi informatif (on impose le  $\beta > 1$ , donc un vieillissement du composant) est assez stable pour les composants présentant un taux de censure à gauche non négligeable. En effet l'écart type de la loi a posteriori du  $\beta$  (le chiffre entre parenthèse) est faible. De plus comme pour le modèle exponentiel le paramètre de forme du composant 6 (ne présentant aucune défaillance) est égal à la moyenne a priori (la moyenne d'une loi uniforme sur  $[1, 4]$  étant égale à 2.50), et est très proche (2.45) pour le composant 3 qui présente un faible taux de censures. En revanche, ceci n'est plus vrai pour le paramètre d'échelle.

L'estimateur de Bayes non informatif semblent tout d'abord biaisé pour des composants ayant des faibles taux de censures à gauche. Pour les autres, le paramètre de forme est la plupart du temps très proche de 1. Il apparaît donc que l'estimateur du maximum de vraisemblance sera retenu, et que la valeur de  $\beta$  sera égale à 1 (on retiendra donc un modèle exponentiel). De nouvelles connaissances a priori peuvent tout de fois nous permettre d'utiliser l'estimateur de Bayes informatif.

### 5.2.3 Estimateurs non paramétriques

On étudie ici l'estimateur non paramétrique empirique, puis l'estimateur de Turnbull, et enfin ce dernier lissé par un noyau gaussien (de fenêtre  $h=2000$ ) :



Si on compare les estimateurs non paramétriques avec la fonction de survie (donnée avec l'estimateur du maximum de vraisemblance de Weibull), on remarque qu'entre 10000 et 50000 heures les quatre courbes sont très proches. Cet estimateur pourra donc être retenu, avec néanmoins quelques retenues lorsqu'il existe de grands sauts.

## 6 Conclusions et perspectives

Nous avons mis en évidence les difficultés d'estimation dans un contexte doublement censuré en fiabilité industrielle, et notamment pour le modèle de Weibull, que ce soit d'un point de vue théorique ou pratique. Une estimation bayésienne peut permettre dans certains cas d'améliorer les estimations. Dans le même but, l'approche non paramétrique a semblé assez stable pour les simulations. En résumé, on peut dire que les modèles de Weibull n'ont pas apporté d'information sur le vieillissement des composants. Le modèle exponentiel nous permettra alors de nous fournir tous les indices sur le comportement de ces composants.

Toute cette étude s'est faite en considérant l'état du composant comme une variable binaire (0 ou 1). Une perspective est d'étudier des modèles multi-états (comme cela se fait en médecine), et notamment grâce à l'élaboration de graphes généraux, connectés avec des probabilités de passage données par cette présente étude.

Le but, in fine, étant d'optimiser le plan de maintenance sur ces composants, il est à envisager d'étudier des problèmes de type optimisation combinatoire.

## 7 Annexe

### Algorithme 1 L'algorithme EM

L'algorithme EM (*Expectation Maximisation*) vise à rechercher le maximum de vraisemblance avec des données incomplètes. On considère un modèle statistique paramétrique, dont le paramètre est noté  $\theta$ . On note  $x$  l'échantillon complet, défini sur un espace  $C$ , qui se décompose en données effectivement observées  $y$ , définies sur un espace  $O$ , et en données manquantes  $z$ , définies sur  $M$ . Ainsi, on a  $x = (y, z)$  et  $C = (O \times M)$ .  $X$  suit de loi à densité  $f(x|\theta)$ ,  $Y$  suit une loi à densité  $g(y|\theta)$  avec

$$g(y|\theta) = \int f(y, z|\theta) dz$$

La densité des valeurs manquantes s'écrit alors

$$h(z|y, \theta) = f(x|\theta)/g(y|\theta)$$

Ainsi, les log-vraisemblances sont :

$$\ell(\theta, x) = \ell(\theta, y) + \log h(z|y, \theta) \quad (11)$$

où  $\ell(\theta, x)$  et  $\ell(\theta, y)$  représente respectivement la log-vraisemblance de l'échantillon complet et des données observées. Or, on veut maximiser la vraisemblance observée. L'algorithme EM consiste alors à rechercher la valeur de  $\theta$ , qui maximise l'information manquante en considérant l'espérance conditionnelle de la vraisemblance de l'échantillon complet  $\ell(\theta, x)$  sachant les données observées.

Soit  $\theta^r$ , un estimateur du paramètre. L'espérance conditionnelle de l'équation (11), pour la loi conditionnelle  $h(z|y, \theta)$ , c'est-à-dire sachant les données observées, est

$$\ell(\theta|y) = Q(\theta|\theta^r) - H(\theta|\theta^r)$$

avec

$$Q(\theta|\theta^r) = \int h(z|y, \theta^r) \ell(\theta, y, z) dz$$

et

$$H(\theta|\theta^r) = \int h(z|y, \theta^r) \log h(z|y, \theta) dz$$

D'après la théorie de l'information, on a

$$H(\theta|\theta^r) \leq H(\theta^r|\theta^r)$$

Ainsi, si  $Q(\theta|\theta^r) \geq Q(\theta^r|\theta^r)$ , on aura  $\ell(\theta, y) \geq \ell(\theta^r, y)$ .

L'algorithme EM se décompose alors en deux étapes :

- \* **Etape E** : Calcul de l'espérance conditionnelle  $Q(\theta|\theta^r)$ .
- \* **Etape M** : Calcul de  $\theta^{r+1}$  qui maximise  $Q(\theta|\theta^r)$ .

Ainsi, toute suite  $(\theta^r)_r$  engendrée par EM vérifie  $\ell(\theta^{r+1}, y) \geq \ell(\theta^r, y)$ .

**Proposition 7 Théorème générique d'Andrews**

On suppose que

(i)  $\Theta$  est un ensemble relativement compact (de fermeture compact) dans un espace métrique,

(ii)  $\frac{1}{n} \sum_{i=1}^n (\psi_i(X_i, \theta) - \mathbb{E}[\psi_i(X_i, \theta)]) \xrightarrow{p.s.} 0, \forall \theta \in \Theta,$

(iii)  $\forall \theta, \theta' \in \Theta, |\psi_i(X_i, \theta) - \psi_i(X_i, \theta')| \leq B_i(X_i)h(d(\theta, \theta'))$ , où  $h$  est une fonction réelle satisfaisant  $\lim_{x \rightarrow 0} h(x) = 0$ , où  $d$  est la distance dans  $\Theta$  et où  $(B_n)_{n \geq 1}$  est une suite de fonctions réelles mesurables,

(iv)  $\sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^n \mathbb{E}[B_i(X_i)] < +\infty,$

(v)  $\frac{1}{n} \sum_{i=1}^n (B_i(X_i) - \mathbb{E}[B_i(X_i)]) \xrightarrow{p.s.} 0,$

Alors

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^n (\psi_i(X_i, \theta) - \mathbb{E}[\psi_i(X_i, \theta)]) \right| \xrightarrow{p.s.} 0 \text{ quand } n \rightarrow +\infty$$

et  $\frac{1}{n} \sum_{i=1}^n \mathbb{E}[\psi_i(X_i, \theta)]$  est continue en  $\theta$

**Preuve :** cf. Andrews [1].

## Références

- [1] D.W.K. Andrews. Generic uniform convergence. *Econometric Theory*, (8), 1982.
- [2] Ewing G.M Reid W.T Ayer M., Brunk H.D and Silverman E. An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics*, pages 641–647, 1955.
- [3] M. Bacha. *Inférence statistique pour des modèles de durées de vie et applications*. PhD thesis, Université de Rouen, Mars 1996.
- [4] A.P. Basu and J.K. Ghosh. Asymptotic properties of a solution to the likelihood equation with life-testing applications. *Journal of the American Statistical Association*, 75(370) :410–414, June 1980.
- [5] P. Tassi J-J. Dreesbeke, B. Fichet (éditeurs). *Analyse Statistique des durées de vie*. Association pour la Statistique et ses Utilisations. Economica, 1989.
- [6] B. Rønn K. Andersen. A nonparametric test for comparing two samples where all observations are either left- or right-censored. *Biometrics*, pages 323–329, March 1995.
- [7] E.L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, pages 457–481, June 1958.
- [8] Martz and Waller. *Bayesian Reliability Analysis*. New York Wiley, 1982.
- [9] A. Monfort. *Cours de l’X*, chapter 6 et 11.
- [10] W. Nelson. *Statistical Applied Life Data Analysis*. Wiley, 1982.
- [11] Christian P. Robert. *The Bayesian Choice*. Springer-Verlag, 1994.
- [12] G. Saporta. *Probabilités, Analyse des données et Statistique*. Paris technip edition, 1996.
- [13] B.W. Turnbull. Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association*, 69 :169–173, 1974.