

UNE ÉTUDE COMPARATIVE ENTRE L'APPROCHE BAYÉSIENNE ET L'APPROCHE CLASSIQUE DANS LA MODÉLISATION DES DONNÉES DE SURVIE

Saliha BENAMIROUCHE ¹

benamirouchesaliha2018@gmail.com

Faculté des Sciences Economiques, Commerciales et de Gestion - Université d'Alger 3

Ahmed Hamimes ²

ahmed.hamimes@univ-constantine3.dz

Faculté de Médecine, Laboratoire de Biostatistique, Bioinformatique et méthodologie Mathématique Appliquées aux Sciences de la Santé (BIOSTIM), Université de Constantine 3

Reçu le: 19/01/2023 **Accepté le :** 27/05/2023 **Publication en ligne le:** 01/06/2023

ABSTRACT : In a context of comparison between the two statistical approaches (the frequentist approach and the Bayesian approach), we try to test these hypotheses: i) the best statistical analysis of clinical trials is carried out via the Bayesian paradigm; ii) the choice between the two statistical approaches (Bayesian and classical) is likely to generate a multitude of decisions; iii) the Bayesian design enriches the statistical method used. For this purpose, survival functions were estimated in a clinical study for two pharmaceutical products (placebo and prednisolone), and a comparison procedure based on standard error was used to evaluate the performance between a Bayesian method and three methods of classical nonparametric estimation.

Keywords : Survival analysis, Bayesian analysis; frequentist analysis.

JEL Classification : C11, C13, C24.

RESUME : Dans un contexte de comparaison entre les deux approches de la statistique (l'approche fréquentiste et l'approche bayésienne), nous nous essayons de tester ces hypothèses : *i*) la meilleure analyse statistique des essais cliniques est réalisée via le paradigme Bayésien ; *ii*) le choix entre les deux approches statistiques (Bayésienne et classique) est susceptible de générer une multitude de décisions ; *iii*) la conception Bayésienne² permet d'enrichir la méthode statistique. Pour cette finalité, on a estimé les fonctions de survie dans une étude clinique pour deux produits pharmaceutiques (placebo et prednisolone), et on utilise une procédure de comparaison basée sur l'erreur standard pour évaluer la performance entre une méthode Bayésienne et trois méthodes d'estimation classique non paramétriques.

Mots-clés : Analyse de Survie, analyse bayésienne ; analyse fréquentiste.

JEL Classification : C11 , C13, C24.

I. INTRODUCTION

L'analyse de durées de vie censurées est utilisée dans des domaines d'application variés (comme la santé, l'économie...) et différentes possibilités ont été proposées pour la modélisation de telles données. Dans l'inférence bayésienne, l'analyse de survie connaît au cours des dernières années une attention croissante, mais encore limitée à cause de la rareté de logiciels spécialisés (une des causes de cette rareté est la difficulté à automatiser les analyses bayésiennes par rapport à l'approche fréquentiste), et aussi la force de l'habitude et la difficulté à accepter une conception différente de la statistique. Un grand intérêt de l'inférence bayésienne par rapport aux méthodes fréquentistes est sa grande cohérence et sa méthodologie unifiée dans le plan théorique, qui permet de déduire des résultats d'interprétations plus riches et plus directes que ceux fournis par l'approche classique.

Plusieurs auteurs ont discuté de l'utilisation de l'inférence bayésienne en analyse de survie dans le cas de modèles semi-paramétriques (Ferguson [9] ; Kalbfleisch [16] ; Kalbfleisch et Prentice [17] ; Clayton [6]) dans lesquels les a priori pour le taux de base et pour les coefficients de régression des covariables sont spécifiés de manières différentes. Certaines de ces méthodes bayésiennes ont utilisé la vraisemblance totale plutôt que la vraisemblance partielle. En effet, un des avantages à utiliser les méthodes bayésiennes pour modéliser conjointement les coefficients de régression des covariables et le taux de base

de décès est que l'on peut calculer de manière précise les distributions a posteriori du modèle ainsi que leurs écarts types en utilisant les techniques MCMC. Néanmoins, le problème de la spécification rationnelle des fonctions de distributions a priori ainsi que la réalisation de calculs intensifs restent toujours posés. Les résultats de ces études bayésiennes de risque proportionnelle ont montré la précision des estimations et les avantages potentiels à utiliser ces méthodes pour analyser des données de survie (Giorgi [12]).

Les modèles paramétriques ont compris une place considérable en analyse de survie bayésienne, la modélisation paramétrique offre une modélisation directe. La littérature statistique dans l'approche bayésienne paramétrique en analyse et tests de survie est très vaste, nous donnons ici quelques références dans le domaine médical ou dans la santé publique (Grieve [13]; Achcar et coll [2]; Achcar et coll [1]; Chen et coll [5]; Dellaportas et Smith [8]; Kim et Ibrahim [18]), le livre de Ibrahim et coll [15] fournit un bon aperçu des modèles de survie bayésiens en général et spécialement aux modèles paramétriques. On peut citer aussi dans la modélisation paramétrique la méthode Bayésienne pour les modèles avec points de rupture introduits par Carlin, Gelfand et Smith [4].

Dans la modélisation non paramétrique, Florens et Rolin [10] ont démontré principalement que les estimations par simulation du processus de Dirichlet et l'inférence Bayésienne non paramétrique donnent de bons résultats par rapport aux méthodes classiques. Ces différents travaux ont été développés dans des contextes méthodologiques différents, en utilisant des lois a priori différentes et/ou en modélisant la fonction cumulative de risque ou directement la fonction de risque instantané (Giorgi, 2002). Dans l'approche classique, différents modèles paramétriques, semi-paramétriques et non-paramétriques ont été développés, (citer ici, K-M, Gamma, Log-logistique, Cox (version classique)...

Dans cet article, on estime la fonction de survie dans une étude clinique pour deux produits pharmaceutiques (placébo et prednisolone¹), cet exemple utilise les durées de survie de 42 patients atteints d'une hépatite chronique active (par mois). Ces patients ont été randomisés en deux groupes égaux l'un a été traité par prednisolone, l'autre a reçu un placebo (voir Held [14]). On utilise une procédure de comparaison basée sur l'erreur standard pour évaluer la performance entre une méthode Bayésienne (la méthode semi-paramétrique de Kalbfleisch) et trois méthodes d'estimation classique non paramétriques (la méthode de Kaplan – Meier, la méthode actuarielle et la méthode de Fleming et Harrington).

II. MODELE STATISTIQUE BAYESIEN

Le concept de « statistique Bayésienne » part du néologisme « Bayésien », tiré du nom de Thomas Bayes², qui introduisit le théorème qui porte à présent son nom dans un article posthume de 1763, il y a 250 ans. Ce théorème exprime une probabilité conditionnelle en termes des probabilités conditionnelles inverses, si A et E sont des événements tels que $P(E) \neq 0$, $P(A/E)$ et $P(E/A)$ sont reliés par :

$$P(A/E) = \frac{P(E/A)P(A)}{P(E/A)P(A) + P(E/\bar{A})P(\bar{A})} = \frac{P(E/A)P(A)}{P(E)} \quad (1)$$

Bayes utilisait la version continue de ce théorème, en prenant deux variables aléatoires x et y . La distribution conditionnelle de y sachant x est donnée par :

$$h(y/x) = \frac{f(x/y) \times f(y)}{\int f(x/y) \times f(y) dy} \quad (2)$$

Le résultat (2) permet de mener l'inférence à partir de la loi du paramètre θ conditionnellement aux observations x , appelée loi a posteriori et définie par :

$$\pi(\theta/x) = \frac{f(x/\theta) \times \pi(\theta)}{\int_{\theta} f(x/\theta) \times \pi(\theta) d\theta} = \frac{f(x/\theta) \times \pi(\theta)}{m(x)} \quad (3)$$

Cette loi a posteriori est la combinaison de :

- $f(x/\theta)$ la fonction de densité de x sachant la valeur de la variable aléatoire θ .
- $\pi(\theta)$ modélise la fonction de densité a priori sur θ .
- $m(x)$ la distribution marginale de x .

Une fois que l'on dispose des données, la quantité $m(x)$ est une constante de normalisation qui garantit que $\pi(\theta/x)$ est bien une distribution de probabilité.

On peut écrire :

¹La prednisolone peut être administrée dans les cas d'affections suivantes : affections allergiques, affections rhumatismales, collagénoses, maladies de la peau...

²Thomas Bayes (1702 ?-1761) était un prêtre presbytérien non-conformiste, membre de la Royal Society, qui publia à titre posthume « An Essay to wards solving a Problem in the Doctrine of Chances ». On connaît très peu de détails sur sa vie et le seul portrait de lui dont on dispose demeure incertain.

$$\pi(\theta/x) \propto f(x/\theta) \times \pi(\theta) \quad (4)$$

III. METHODE D'ECHANTILLONNAGE DE GIBBS³

L'échantillonnage de Gibbs représente l'algorithme Bayésien le plus utilisé dans l'inférence statistique, cette méthode a été utilisée par Geman et Geman [11] en 1984 pour générer des observations à partir d'une distribution de Gibbs (distribution de Boltzmann). Parmi les propriétés de cet algorithme, on trouve :

- l'échantillonnage de Gibbs est un cas particulier de l'algorithme M-H tel que la probabilité d'acceptation soit toujours égal à 1 ;
- l'échantillonnage de Gibbs tire profit des structures hiérarchiques d'un modèle ;
- dans certains algorithmes on trouve que les simulations peuvent être rejetées, par contre l'échantillonnage de Gibbs où toutes les simulations sont prises en comptes ;
- cet algorithme est bien adapté dans le cadre d'un modèle avec information cachée que cela soit avec une censure ou avec une variable latente (Cyprien [7], Boreux et al [3]).

L'algorithme se décompose en les points suivants :

- Initialiser $\theta_0 = \theta_0^{(1)}, \theta_0^{(2)}, \dots, \theta_0^{(k)}$, soit le premier vecteur d'éléments de la chaîne.
- Poser $t \leftarrow 0$.
- Pour passer de l'étape t à l'étape $t + 1$:

Générer $\theta_{t+1}^{(1)}$ en simulant selon la loi $\pi_1(\theta_t^{(1)}/\theta_t^{(2)}, \dots, \theta_t^{(1)}, \dots, \theta_t^{(k)})$

Générer $\theta_{t+1}^{(2)}$ en simulant selon la loi $\pi_2(\theta_t^{(2)}/\theta_{t+1}^{(1)}, \theta_t^{(3)}, \dots, \theta_t^{(k)})$

⋮

Générer $\theta_{t+1}^{(k)}$ en simulant selon la loi $\pi_k(\theta_t^{(k)}/\theta_{t+1}^{(1)}, \theta_{t+1}^{(2)}, \dots, \theta_{t+1}^{(k-1)})$

- Changer la valeur de t à $t \leftarrow t + 1$, et aller en 3.

IV. CONCEPTS DE BASE DANS L'ANALYSE DE SURVIE

a. Caractérisation de la loi de durée

Soit X une variable aléatoire positive ou nulle, et absolument continue définie sur un espace probabilisé $(\Omega, \mathcal{A}, IP)$ qui décrit le temps qui s'écoule entre deux événements. La loi de probabilité de cette variable qu'on appelle durée de survie peut être définie par l'une des cinq fonctions équivalentes suivantes :

Définition 1. (La fonction de survie).

La fonction de survie S de la variable aléatoire X est par définition le complément à un de la fonction de répartition :

$$S(t) = P(X > t); \quad t \geq 0$$

pour t fixé c 'est la probabilité de survivre jusqu'à l'instant t .

Les différentes propriétés de cette fonction sont :

- S est continue à gauche ;
- S est monotone et décroissante ;
- $S(0) = 1$;
- $\lim_{t \rightarrow +\infty} S(t) = 0$.

Définition 2. (La fonction de répartition).

La fonction de répartition (f. r. ou c.d.f en anglais pour « cumulative distribution function ») est :

$$F(t) = P\{X \leq t\} = 1 - S(t)$$

Pour t fixé, c 'est la probabilité de mourir avant l'instant t .

La fonction de répartition est croissante et on a :

$$F(0) = 0 \text{ et } \lim_{t \rightarrow +\infty} F(t) = 1$$

Définition 3. (La densité de probabilité).

³ L'algorithme de Gibbs tire son nom d'un physicien et mathématicien américain du 19^{ème} siècle Josiah Willard Gibbs, considéré comme l'un des fondateurs de la thermodynamique moderne et de la mécanique statistique.

La densité de probabilité de X , noté $f(t)$ c'est une fonction $f(t) \geq 0$ et pour tout $t \geq 0$:

$$F(t) = \int_0^t f(s) ds$$

$f(t)$ est définie par :

$$f(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq X < t + dt)}{dt} = F'(t) = -S'(t)$$

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

Définition 4. (Taux de hasard).

On appelle taux de hasard⁴ (ou fonction de risque, taux de panne, taux de défaillance, taux de décès, risque instantané, etc.) au temps t la fonction :

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{1}{dt} P(t \leq X < t + dt / X \geq t), t > 0$$

En temps continu, la fonction de risque est la probabilité limite qu'une unité effectue une transition sachant qu'elle n'en a pas effectuée au préalable. On peut montrer la relation suivante :

$$\lambda(t) = \frac{\text{La probabilité de mourir dans un petite intervalle du temps}}{\text{La probabilité de survivre jusqu'à l'instant } t} = \frac{f(t)}{S(t)}$$

Lorsque t fixé, caractérise la probabilité de mourir dans un petit intervalle de temps après l'instant t , conditionnellement au fait d'avoir survécu jusqu'à l'instant t , aussi cela signifie le risque de mort instantané pour ceux qui ont survécu.

Définition 5. (Le taux de hasard cumulé).

Le taux de hasard cumulé Λ c'est l'intégrale du taux de hasard λ , et lorsque :

$$f(t) = -\frac{d(1 - F(t))}{dt} = -\frac{d(S(t))}{dt}$$

On trouve :

$$\begin{aligned} \Lambda(t) &= \int_0^t \lambda(u) du = \int_0^t \frac{f(u)}{S(u)} du = \int_0^t \frac{1}{S(u)} f(u) du \\ &= -\int_0^t \frac{1}{S(u)} d(S(u)) = -\ln S(t). \end{aligned}$$

On peut déduire la fonction de survie du taux de hasard cumulé grâce à la relation $-\ln S(t) = \Lambda(t)$, on trouve :

$$\ln S(t) = -\Lambda(t) \Rightarrow S(t) = \exp\{-\Lambda(t)\} = \exp\left\{-\int_0^t \lambda(u) du\right\}$$

donc $S(t) = \exp\left\{-\int_0^t \lambda(u) du\right\}$.

Définition 6. (La fonction de survie conditionnelle).

$$S(t/t_0) = P(X > t + t_0 / X > t_0) = \frac{S(t + t_0)}{S(t_0)} = \exp\left(-\int_{t_0}^t \lambda(u) du\right); \forall t > t_0$$

Définition 7. (Moyenne et variance de la durée de survie).

La variance et le temps moyen de survie données respectivement par :

$$\begin{aligned} E(X) &= \int_0^{\infty} S(t) dt \\ V(X) &= 2 \left(\int_0^{\infty} tS(t) dt \right) - (E(X))^2 \end{aligned}$$

⁴Cette fonction n'est pas une densité de probabilité.

V. METHODES FREQUENTISTE D'ANALYSE DE SURVIE

a. L'estimateur de Kaplan Meier

L'estimateur de Kaplan- Meier (1958) est une méthode fonctionnelle pour estimer la fonction de survie s'écrit

$$\hat{S}(t) = \begin{cases} \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) & \text{si } t \geq t_1 \\ 1 & \text{si } t < t_1 \end{cases} \quad (5)$$

t_i représente le temps de suivi depuis l'inclusion dans l' étude pour chaque individu i .

d_i est le nombre de décès au temps t_i .

n_i est le nombre de sujets à risque de présenter l'évènement étudié à l'instant t_i , i.e. le nombre de patients n'ayant pas encore subi l'évènement ni la censure juste avant t_i .

b. L'estimateur actuariel

Relativement aux méthodes de calcul employées dans l'estimateur KM, une seule différence notable intervient dans le traitement des censures au sein d'un intervalle. Ceci revient à supposer que la censure se produit de manière uniforme sur l'intervalle. Si l'effectif d'individus risqués est n_i au début de l'intervalle $[t_i; t_{i+1}[$ est t_{i+1} ; le nombre d'individus à risque est de $n_i - c_i$.

L'effectif risqué moyen sur $[t_i; t_{i+1}[$ est donc $n_i^* = n_i - \frac{c_i}{2}$. En conséquence la probabilité de survenue de l'évènement au cours de ce $i^{\text{ème}}$ intervalle pour les individus qui ne l'ont pas connu au cours des $(i - 1)$ intervalles précédents est donnée par :

$$q_i^* = \frac{d_i}{n_i - \frac{c_i}{2}} = \frac{d_i}{n_i^*}$$

L'estimateur de la fonction de survie est donné comme suit :

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i^*}\right)$$

La formule de Greenwood permet d'obtenir une estimation de la variance,

$$Var[\hat{S}(t)] = \hat{S}(t)^2 \times \sum_{i | t_i \leq t} \frac{d_i}{(n_i^* - d_i)n_i^*}$$

c. L'estimateur du hasard intégré : Nelson-Aalen¹

L'estimation du taux de risque cumulé $\Lambda(t)$ à travers la méthode de Nelson et Aalen est basé sur l'équation suivante :

$$\begin{aligned} \Lambda(u + du) - \Lambda(u) &\approx \lambda(u)du & (6) \\ &= P(\text{événement dans}(u, u + du)|\text{à risque en } u) \end{aligned}$$

L'équation (6) est asymptotique à la probabilité de mourir dans un petit intervalle de temps après l'instant u , elle est donnée par :

$$P(u < X \leq u + du | X > u) = \frac{P(u < X \leq u + du)}{P(X > u)} \quad (7)$$

Il est naturel d'estimer le rapport (7) par :

$$\frac{[N(u + du) - N(u)]}{Y(u)} \quad (8)$$

Ainsi, la probabilité (8) est approchée par la fréquence :

$$\frac{[N(u + du) - N(u)]}{Y(u)} = \frac{dN(u)}{Y(u)} \quad (9)$$

où $dN(u)$ c'est le nombre "d'évènements" observés dans l'intervalle $]u, u + du]$ et $Y(u)$ est le nombre "d'individus" à risque à l'instant u .

En sommant les quantités(9) sur tous les sous intervalles formant $(0, t]$ et en faisant tendre ces intervalles vers 0, de telle sorte que chacun ne contienne qu'un seul évènement. Ainsi, l'estimateur de Nelson-Aalen est défini par :

$$\hat{\Lambda}_{NA}(t) = \int_0^t \frac{dN(u)}{Y(u)}$$

¹ Cet estimateur a été proposé par Nelson dans l'année 1972 et ensuite par Aalen en 1978.

qui peut aussi s'écrire, si on note les instants t_1, t_2, \dots, t_N ordonnés de survenue de l'évènement pour les N individus (processus de sauts) :

$$\begin{aligned} \int_0^t \frac{dN(u)}{Y(u)} &= \sum_{\{i, t_i \leq t\}} \int_{t_{i-1}}^{t_i} \frac{dN(u)}{Y(u)} = \sum_{\{i, t_i \leq t\}} \frac{N(t_i) - N(t_{i-1})}{Y(t_i)} \\ &= \sum_{\{i, t_i \leq t\}} \frac{\Delta N(t_i)}{Y(t_i)} = \widehat{\lambda}_{NA}(t) \end{aligned}$$

où sur $]t_{i-1}, t_i]$ le processus $Y(u)$ est constant égal à $Y(t_i)$ et $\Delta N(t_i)$ est le nombre d'évènements observés dans l'intervalle $]t_{i-1}, t_i]$ i.e. en t_i :

$$\Delta N(t_i) = N(t_i) - N(t_{i-1})$$

Selon la dernière équation, il est possible de trouver un estimateur pour le taux de hasard instantané $\lambda(t_i)$ à l'instant t_i , tel que l'estimateur $\widehat{\lambda}_{NA}(t)$ de Nelson- Aalen (1972,1978), s'écrit comme suit :

$$\widehat{\lambda}_{NA}(t) = \sum_{\{i, t_i \leq t\}} \widehat{\lambda}(t_i)$$

on pose :

$$\widehat{\lambda}(t_i) = \frac{d_i}{n_i}$$

on trouve la forme finale de cet estimateur comme suit :

$$\widehat{\lambda}_{NA}(t) = \sum_{\{i, t_i \leq t\}} \widehat{\lambda}(t_i) = \sum_{\{i, t_i \leq t\}} \frac{d_i}{n_i} \quad (10)$$

Proposition 1. Un estimateur de la variance de $\widehat{\lambda}_{NA}(t)$ est :

$$var[\widehat{\lambda}_{NA}(t)] = \sum_{i: t_i \leq t} \frac{d_i}{n_i^2}$$

Théorème 1.

Pour $t \geq 0$ on a :

$$|\widehat{\lambda}_{NA}(t) - \lambda(t)| \xrightarrow[n \rightarrow \infty]{P} 0$$

Ce théorème donne la convergence en probabilité de l'estimateur de Nelson-Aalen.

d. Estimateur de Harrington et Fleming de la survie

À partir de la relation $S(t) = \exp(-\lambda(t))$ et de l'estimateur de Nelson-Aalen, on peut en déduire un autre estimateur de la fonction de survie :

$$\begin{aligned} \widehat{S}_{HF}(t) &= \exp(-\widehat{\lambda}(t)) = \exp\left(-\sum_{\{i, t_i \leq t\}} \frac{d_i}{n_i}\right) = \prod_{i: T_i \leq t}^n e^{-\frac{d_i}{n_i}} \\ &\approx \prod_{i: T_i \leq t}^n \left(1 - \frac{d_i}{n_i}\right), \text{ si } \frac{d_i}{n_i} \rightarrow 0 \quad (11) \end{aligned}$$

où d_i et n_i sont le nombre de décès et d'individus à risque en T_i (la durée réellement observée). En appliquant un développement limité, on retrouve l'estimateur de Kaplan-Meier. En utilisant la delta-méthode ($Var(f(Z)) \sim [f'(E(Z))]^2 Var(Z)$), on peut obtenir un estimateur de la variance de cet estimateur :

$$\begin{aligned} \widehat{var}(\widehat{S}_{HF}(t)) &= (\widehat{S}_{HF}(t))^2 \widehat{var}(\widehat{\lambda}(t)) = \exp\left(-2 \sum_{i: T_i \leq t}^n \frac{d_i}{n_i}\right) \\ &\quad \times \left(\sum_{i: T_i \leq t}^n \frac{d_i}{n_i^2}\right) \quad (12) \end{aligned}$$

VI. MODELE BAYESIENNE D'ANALYSE DE SURVIE

Définition 8.

Un processus de comptage (de survenue de l'événement) est définie par :

$$N_i(t) = 1\{T_i \leq t, \delta_i = 1\}$$

À chaque individu correspond un indicateur $Y_i(t)$ de présence à risque à l'instant t :

$$Y_i(t) = 1\{T_i \geq t\}$$

On a un n -échantillon. On note $Y(t)$ la somme des processus $Y_i(t)$, compte le nombre total d'individus qui sont toujours dans l'étude au temps t , et $N(t)$ la somme des $N_i(t)$, le processus de comptage qui compte le nombre total de décès jusqu'à l'instant t :

$$Y(t) = \sum_{i=1}^n Y_i(t) \quad \text{et} \quad N(t) = \sum_{i=1}^n N_i(t)$$

Théorème 2. Soit $N(t)$ un processus de comptage. Les propriétés suivantes sont équivalentes :

- i. $N(t)$ est un processus de poisson homogène ;
- ii. l'intensité de $N(t)$ est constante ;
- iii. $N(t)$ est un processus à accroissements indépendants et stationnaires.

Dans ce cas, il existe $I \geq 0$ tel que $N(t) \sim \text{Poisson}(It)$ pour tout $t \in \mathbb{R}_+$ et I est le paramètre des assertions (i) et (ii).

Définition 9

On s'intéresse à la modélisation des occurrences dans le temps d'événements aléatoires. On étudiera particulièrement le cas où les événements ont lieu en temps continue et sont de nature discrète (ils appartiennent à un espace d'états dénombrable). On considère donc un intervalle de temps continu :

$$\mathcal{T} = [0, \tau[$$

où τ temps terminal tel que $0 \leq \tau \leq \infty$.

Soit un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, une filtration ou histoire $\{\mathcal{F}_t; t \in \mathcal{T}\}$ est une famille croissante et continue à droite de sous-tribus de \mathcal{A} . On supposera que la filtration est complète dans le sens où $\forall t \in \mathcal{T}, \mathcal{F}_t$ contient tous les ensembles négligeables de \mathcal{A} . Les filtrations $\{\mathcal{F}_t; t \in \mathcal{T}\}$ considérées dans ce mémoire satisferont les conditions suivantes :

- $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{A}, \forall s \leq t$ (croissance) ;
- $\mathcal{F}_s = \bigcap_{t>s} \mathcal{F}_t, \forall s$ (càd) ;
- $A \subset B \subset \mathcal{A}, IP(B) = 0 \Rightarrow A \in \mathcal{A}$ (complète).

Une filtration permet de définir ce qui appartient passé du processus (qui est observé et connu) et ce qui appartient au future (pas encore accessible).

Définition 10

Soit $\{\mathcal{F}_t; t \in \mathcal{T}\}$ une filtration sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Un processus de comptage $N(\cdot)$ sur \mathcal{T} est un processus continu à droite avec une limite à gauche, \mathcal{F}_t – adaptée, nul en zéro, croissant, à trajectoire constante par morceaux, ayant des sauts d'amplitude 1 et tel que $N(t)$ est p.s fini pour tout $t \in \mathcal{T}$.

Proposition 2.

Soit $N = \{N(t), t \geq 0\}$ un processus de comptage adapté à \mathcal{F}_t tel que $E(N(t)) < \infty$ pour tout t . Alors il existe un unique processus $\Lambda(\cdot)$ \mathcal{F}_t -prévisible, croissant, càdlàg et nul en zéro p.s. vérifiant $E(\Lambda(t)) < \infty$ pur tout t , tel que :

$$M(t) = N(t) - \Lambda(t), t \in \mathcal{T}$$

soit une martingale continue à droite.

Théorème 3. (La décomposition de Doob-Meyer).

Soit N une sous martingale⁵ positive continue à droite adaptée à la filtration $\mathcal{F} = \{\mathcal{F}_t; t \geq 0\}$. Il existe une \mathcal{F} – martingale continue à droite M et un processus prédictible croissant continue à droite Λ tel que $E(\Lambda(t)) < \infty$ et :

$$N(t) = M(t) + \Lambda(t)$$

Pour tout $t \geq 0$. Si $\Lambda(0) = 0$ presque sûrement et $N = M' + \Lambda'$ est une autre décomposition avec $\Lambda'(0) = 0$, alors pour tout $t \geq 0$:

$$P(M(t) \neq M'(t)) = 0 = P(\Lambda(t) \neq \Lambda'(t))$$

Proposition 3

soient $N(t)$ un processus de comptage et Λ son **compensateur** (ou encore son processus d'intensité cumulée) obtenu par la décomposition de Doob-Meyer. On dit que $N(t)$ a une **intensité** I si I est prévisible et si :

⁵ Comme un processus de comptage (croissant, nul en zéro et borné) est une sous –martingale positive, on déduit de la décomposition de Doob –Meyer.

$$\Lambda(t) = \int_0^t I(u)du, \forall t.$$

Si $N(t)$ intégrable et I càdlàg on montre que :

$$I(u^+) = \lim_{\varepsilon \rightarrow 0} IE(N(u + \varepsilon) - N(u)/\mathcal{F}_u)$$

où $\mathcal{F}_u = \sigma\{N(s) : s \leq u\}$. Le paramètre I représente, sachant le passé, la probabilité infinitésimale d'avoir un saut du processus de comptage.

Il est possible de réécrire la fonction de hasard en termes de processus de comptage. Comme l'incrément du processus de comptage $dN_i(t)$ ne peut prendre que la valeur 0 ou 1, on écrit :

$$E(dN_i(t) / \mathcal{F}_{t-}) \approx P(dN_i(t) = 1 / \mathcal{F}_{t-})$$

Dans un cas sans variable explicative :

$$P(dN_i(t) = 1 / \mathcal{F}_{t-}) = P(dN_i(t) = 1 / Y_i(t))$$

Le cas $Y_i(t) = 1$, correspond à une situation où l'individu est à la fois observé et susceptible d'effectuer une transition dans l'intervalle de temps $[t, t + dt]$, on a donc

$$P(dN_i(t) = 1 / Y_i(t) = 1) = P(t \leq X_i < t + dt / t \leq X_i, t \leq C_i)$$

Les variables aléatoires X_i et C_i sont indépendantes dans le cas de censure non-informative, on peut écrire :

$$P(dN_i(t) = 1 / Y_i(t) = 1) = \lambda(t)dt$$

Le cas $Y_i(t) = 0$, correspond à une situation où processus de comptage est incrémenté avec probabilité nulle. On en déduit la reformulation du hasard dans le cadre des processus de comptage :

$$E(dN_i(t) / \mathcal{F}_{t-}) \approx P(dN_i(t) = 1 / Y_i(t)) \approx Y_i(t)\lambda(t)dt$$

Pour les individus $i = 1, \dots, n$, nous observons le processus $N_i(t)$ qui compte le nombre de décès jusqu'à l'instant t . Le processus d'intensité $I_i(t)$ est donné par :

$$I_i(t)dt = E(dN_i(t) / \mathcal{F}_{t-})$$

où $dN_i(t)$ est l'augmentation de N_i sur le petit intervalle de temps $[t, t + dt]$, et \mathcal{F}_{t-} représente les données disponibles (ou bien les informations) juste avant l'instant t . Si l'objet i décède pendant cet intervalle de temps, $dN_i(t)$ prend la valeur 1; autrement $dN_i(t) = 0$. Donc $E(dN_i(t) / \mathcal{F}_{t-})$ est la probabilité de décès dans l'intervalle $[t, t + dt]$ pour l'individu i . Comme $dt > 0$ (en supposant que le temps soit continu), cette probabilité devient le risque instantané à l'instant t pour le sujet i . Ceci est supposé avoir la forme des risques proportionnels :

$$I_i(t) = Y_i(t)\lambda_0(t) \exp(\beta Z_i) \quad (13)$$

où $Y_i(t)$ est connu sous le nom de processus de risque, indiquant si un individu est susceptible de subir un événement au temps t , en prenant la valeur 1 ou 0 selon si oui ou non l'objet i est observé à l'instant t , et $\lambda_0(t) \exp(\beta Z_i)$ est le modèle familier de régression de Cox. Ainsi les données que nous avons observé $D = \{N_i(t), Y_i(t), Z_i, i = 1, 2, \dots, n\}$, où $Z_i : i = 1, \dots, n$ est le vecteur des covariables, et β est le vecteur des paramètres inconnus, et $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ estimée d'une méthode non-paramétrique.

La distribution a posteriori conjointe pour le modèle ci-dessus est définie par :

$$\pi(\beta, \Lambda_0(t) / D) \propto l(D / \beta, \Lambda_0(t)) \pi(\beta) \pi(\Lambda_0(t)) \quad (14)$$

Nous devons préciser la forme de la vraisemblance $l(D / \beta, \Lambda_0(t))$ et les distributions a priori pour β et $\Lambda_0(t)$.

Sous la censure non informative, la vraisemblance des données est proportionnelle à :

$$\prod_{i=1}^n \left[\prod_{t \geq 0} I_i(t)^{dN_i(t)} \right] \exp \left(- \int_{t \geq 0} I_i(t) dt \right)$$

Le processus de comptage incrémente $dN_i(t)$ dans l'intervalle de temps $[t, t + dt)$ est i.i.d d'une distribution de Poisson avec des moyennes $I_i(t)dt$:

$$dN_i(t) \sim \text{Poisson}(I_i(t)dt)$$

Nous pouvons écrire :

$$I_i(t)dt = E(dN_i(t) / \mathcal{F}_{t-}) = \exp(\beta Z_i) Y_i(t) d \Lambda_0(t)$$

où $d \Lambda_0(t) = \lambda_0(t)dt$ représente l'accroissement du risque instantané de décès pendant l'intervalle de temps $[t, t + dt]$.

Comme la distribution a priori conjuguée de Poisson est une distribution de gamma, Kalbfleisch [16] considère que $\Lambda_0(t)$ est un processus gamma avec $E[\Lambda_0(t)] = \Lambda^*_0(t)$, qui est une fonction connue, et $V[\Lambda_0(t)] = \Lambda^*_0(t)/c$ où c est un nombre réel positif. Les incréments $d \Lambda_0(t)$ sur $[t, t + dt]$ sont censés être indépendants et suivent les lois Gamma $(c \Lambda^*_0(t), c)$ indépendantes, où les paramètres $\Lambda^*_0(t)$ et c représentent respectivement l'estimation a priori de la fonction de risque instantané inconnue et le degré de confiance que l'on attache à cette loi a priori. Quand c est grand la croyance a priori sur $\Lambda^*_0(t)$ est forte,

et inversement. Kalbfleisch [16] choisit comme loi a priori, $\Lambda^*_0(t) = rdt$ où r estime le taux d'événements par unité de temps, et dt est l'intervalle de temps considéré (voir Khribi (2007)).

Les coefficients de régression à effets fixes β sont attribués selon une densité a priori vague :

$$\beta \sim \mathcal{N}(0; 0,000001)$$

selon cette distribution, la loi a priori pour les composantes de β est considérée comme étant faible informative.

Donc pour un temps $t_j (j = 1, \dots, n)$. L'estimation Bayésienne de la fonction de survie par la méthode de Kalbfleisch est de la forme suivante :

$$SB(t_j) = \exp\left(-\sum_{k=1}^j \Lambda_0(t_k)\right)^{\exp(\beta Z_k)} \quad (15)$$

Ce modèle peut être implanté avec le logiciel WinBUGS ou OpenBUGS et l'estimation des paramètres est réalisée en utilisant la méthode d'échantillonneur de Gibbs.

Avant la comparaison numérique, nous effectuons une comparaison graphique entre les courbes de survie des deux approches statistiques (Bayésienne et classique).

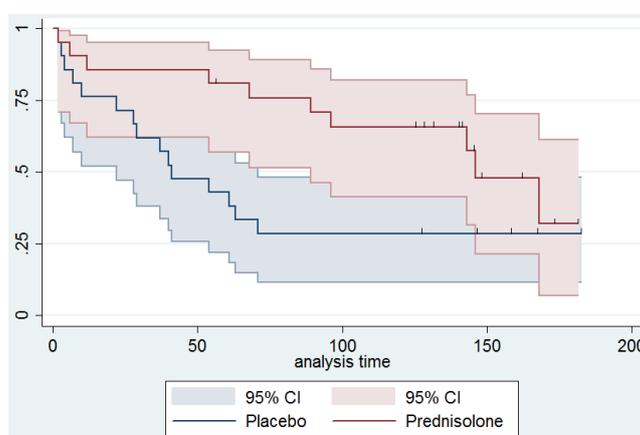


Fig.1 Les courbes de survie estimées selon la méthode de Kaplan-Meier.

À partir des courbes de survie estimées selon la méthode de Kaplan-Meier, on trouve les résultats suivants :

1. la divergence entre les deux fonctions se produit graduellement dans le début d'expérience, ce qui signifie que l'efficacité du prednisolone nécessite une période (25 mois) pour apparaître ;
2. la fonction de survie des deux produits pharmaceutiques démontre que les patients de groupe placebo présentent une probabilité de survie plus faible que ceux du groupe prednisolon, et selon ce résultat le prednisolone possède une efficacité par rapport au comparateur (placebo) ;
3. le graphe est de forme « transient difference », où la convergence des deux fonctions est après le 168^{ème} mois, ce qui signifie que le prednisolone va perdre son efficacité après cette durée.

On souhaite maintenant dépasser la comparaison graphique des deux groupes de traitement pour effectuer un test statistique approprié. Le test de Student ne s'y prête pas, car il demande d'ignorer que chez une partie des patients, l'événement n'a pas encore eu lieu. Pour cela, on utilise les tests du tableau 1.

Statistique	Valeur observée	Valeur critique	p-value	Alpha
Log-Rank	4,028	3,841	0,045	0,05
Wilcoxon	5,686	3,841	0,017	0,05
Tarone -Ware	5,255	3,841	0,022	0,05

Tab. 1 Les tests d'égalité des fonctions de survie.

A partir des résultats récapitulés dans le Tab(1), la valeur p des tests log-rank, Wilcoxon et Taron-Ware (voir le tableau (1)) est inférieure au seuil (5%), ce qui signifie que l'on a prouvé qu'il existe une différence statistiquement significative entre les probabilités de survie des deux groupes.

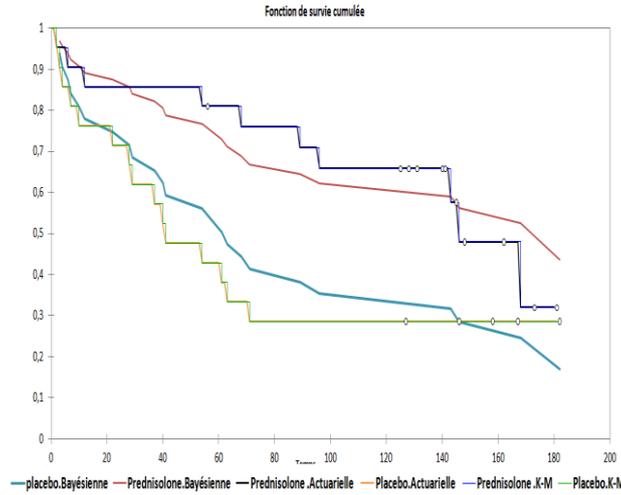


Fig.2 La comparaison entre l’approche Bayésienne et classique dans l’estimation des courbes de survie.

Selon le graphe précédent (figure(2)), on remarque bien que le troisième point de l’interprétation classique n’est pas juste selon une vision Bayésienne c.à.d. la convergence des deux fonctions de survie après le 168^{ème} mois n’est pas valable. La méthode Bayésienne de Kalbfleish estime le coefficient de régression β , qui nous révèle qu’un patient qui fait partie du groupe avec un placebo a un risque de mourir de $e^{0,8245} = 2,28$ fois plus élevé qu’un patient du groupe prednisolone. La masse de probabilité a postériori de β située à droite de « 0 », donc le risque relatif montre l’efficacité absolue de prednisolone par rapport au placebo.

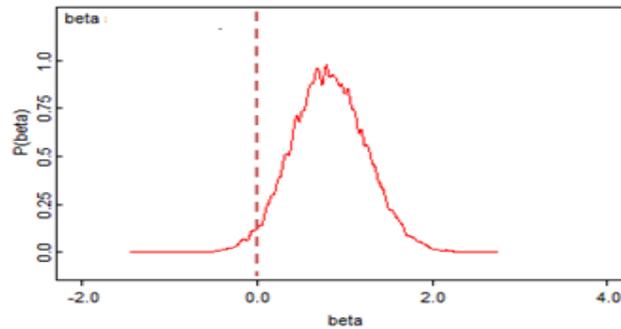


Fig.3 L’estimateur du noyau de Kernel pour la densité marginale de β .

Pour la comparaison numérique entre les approches, on utilise les erreurs standards moyennes, donnée par :

$$ESM = \frac{\sum_{i=1}^n [ES(t_i)]}{n} = \frac{TES}{n}$$

n est le nombre de points de durées estimés par chaque méthode. Dans l’estimation de Kaplan Meier est le nombre de durée non censurées.

Méthode d’estimation	Placebo		Prednisolone		Placebo+ Prednisolone	
	TES	ESM	TES	ESM	TES	ESM
Kalbfleisch⁶	2.02	0,084	1,61 6	0,067	3.6	0.15
Kaplan- Meier	1,41	0,094	0,98 3	0,0983	2,3	0,19
Actuarielle	1,71	0,093	2 ,25 1	0,071	3.9	0,16
Fleming-Harrington	1 ,8 9	0,094	2,23 6	0,106	4,1	0 ,20

Tab.2 La comparaison entre les résultats d’estimation.

L’approche Bayésien (la méthode de Kalbfleisch) dans le tableau précédent est la plus performante, car elle donne les meilleures valeurs de précision.

⁶Les résultats du tableau sont obtenus après 100000 itérations (on utilise l’échantillonnage de Gibbs).

Dans la méthode actuarielle, les probabilités conditionnelles sont estimées sur des intervalles fixés par l'utilisateur et non déterminés par les temps d'événements, il est possible d'utiliser cette propriété pour une comparaison directe entre les deux approches.

Le graphe suivant illustre l'évolution de l'erreur standard dans les deux méthodes où il est remarquable que la trajectoire Bayésienne soit plus proche au centre, ce qui signifie l'efficacité absolue de cette méthode par rapport à d'autres alternatives classiques.

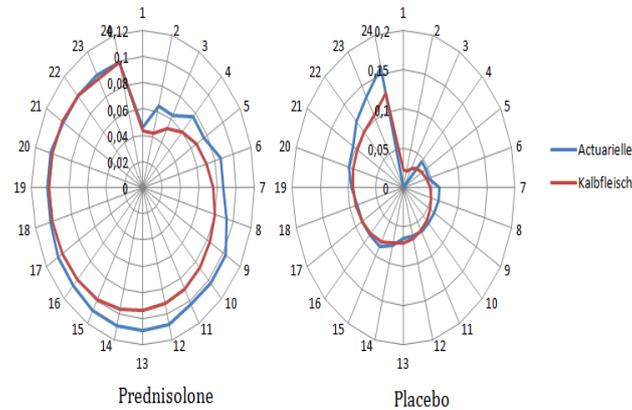


Fig.4 Les courbes spirales pour les erreurs standards dans les deux méthodes.

VII. CONCLUSION

La pertinence et l'efficacité de l'approche Bayésienne comme guide du raisonnement scientifique face à l'incertitude sont reconnues depuis longtemps, nous avons montré cette efficacité dans l'exemple de prednisolone avec l'application de l'échantillonnage de Gibbs dans le modèle de Kalbfleisch comparativement à l'approche classique des estimateurs de Kaplan Meir, actuarielle et de Fleming et Harrington. La pratique de cet article nous permet aussi de tester l'approche d'échantillonnage de Gibbs et de donner les résultats suivants :

- échantillonnage de Gibbs tire profit des structures hiérarchiques d'un modèle ;
- dans les choix non restreints des lois instrumentales, on trouve des problèmes de convergence ;
- la vitesse de convergence est très élevée dans plusieurs cas.

La statistique Bayésienne donne des résultats plus riches que les estimateurs classiques. Les techniques Bayésiennes permettent d'obtenir la loi jointe des paramètres du modèle (le modèle Bayésien dans l'analyse de survie permet de modéliser conjointement le taux de base de décès et les coefficients de régression des covariables), donc de prendre en compte simultanément l'effet de l'incertitude globale sur l'ensemble des paramètres inconnus.

VII. REFERENCES

- [1] ACHCAR, J.A., BOLFARINE, H., PERICCHI, L.R., "Transformation of survival data to an extreme value distribution, the statistician", 36, 1986, 229-234.
- [2] ACHCAR, J.A., BROOKMEYER, R., HUNTER, W.G., "An application of Bayesian analysis to medical follow-up data", Statistics in Medicine, 4, 1985, 509-520.
- [3] BOREUX, J., PARENT, E., BERNIER, J., Pratique du calcul Bayésien, Springer, 2010.
- [4] CARLIN, B.P., GELFAND, A.E., SMITH, A.F.M., "Hierarchical Bayesian Analysis of Change point Problems", Appl. Statist. 41(2), 1992, 389-405.
- [5] CHEN, W.C., HILL, B.M., GREENHOUSE, J.B., FAYOS, J.V., "Bayesian analysis of survival curves for cancer patients following treatment", In Bayesian Statistics 2 (Eds. J.O. Berger, J. Bernardo. A.F.M. Smith), Amsterdam: North-Holland, 1985, 299-328.
- [6] CLAYTON, D.G., "A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence", Biometrika, 65, 1978, 141 – 151.
- [7] CYPRIEN, M.T., Inférence dans les modèles conjoints et de mélange non-linéaires à effets mixtes, Thèse de doctorat, Universités Paris-Sud XI et Yaounde1, 2012.
- [8] DELLAPORTAS, P., SMITH, A.F.M., "Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling", Applied Statistics, 42, 443-459, 1993.
- [9] FERGUSON, T.S., "A Bayesian analysis of some nonparametric problems", The Annals of Statistics, 1(2), 209-230, 1973.
- [10] FLORENS, J.P., ROLIN, J.M., "Simulation of posterior distributions in nonparametric censored analysis", International Statistical Review, 67(2), 187-210, 2001.

- [11] GEMAN, S., GEMAN, D., "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of image". IEEE Trans. Pattern Anal. Mach. Intell., 6, 1984, 721-741.
- [12] GIORGI, R., Analyses comparatives des méthodes de survie et extensions d'un modèle régressif de survie relative : prise en compte de la non-proportionnalité des risques par des fonctions B-splines et développement d'une méthode d'analyse bayésienne, Thèse de doctorat, Université Aix-Marseille II, France, 2002.
- [13] GRIEVE, A.P., "Applications of Bayesian software: Two examples", The Statistician, 36, 283-288, 1987.
- [14] HELD, U., «Représentation graphique et comparaison de courbes de survie », Forum Med Suisse, 10 (33), 2010, 548-550.
- [15] IBRAHIM, J. G., CHEN, M, H., SINHA, D., "Bayesian survival analysis", Springer, 2001.
- [16] KALBFLEISCH, J.D., "Nonparametric bayesian analysis of survival time data", Journal of the Royal Statistical Society, Series B, 40(2), 214-221, 1978, 214-221.
- [17] KALBFLEISCH, J.D., PRENTICE, R.I., "The Statistical analysis of failure time data", John Wiley & Sons, New York, 1980.
- [18] KIM, S.W., IBRAHIM, J.G., "On Bayesian inference for proportional hazards models using noninformative priors", Lifetime Data Analysis, to appear, 2001.