

Modèles de durée

M1 DUAS - Université de Strasbourg

Pierre-O Goffard

updated on November 13, 2025

Contents

1	Introduction	3
1.1	Distribution d'une durée	3
1.1.1	Formulation continue	3
1.1.2	Formulation discrète	5
1.1.3	Le passage du continu au discret et inversement	5
1.2	Modélisation statistique	5
2	Approche paramétrique	8
2.1	Fonction de vraisemblance en présence de censure et de troncature	8
2.1.1	Censure	8
2.1.2	Troncature	9
2.2	Lois paramétriques usuelles	10
2.2.1	Loi exponentielle	10
2.2.2	Loi Gamma	10
2.2.3	Loi de Weibull	10
2.3	Comparaison et sélection d'un modèle	11
2.3.1	Critère d'information	11
2.3.2	Test d'adéquation	12
3	Approche non paramétrique	13
3.1	Estimateur de Kaplan-Meier et Nelson-Aalen	13
3.2	Variance et intervalle de confiance	16
3.2.1	Rappel des propriétés de l'estimateur du maximum de vraisemblance	16
3.2.2	Application à l'estimateur de Kaplan-Meier	17
3.3	Comparaison de la courbe de survie dans deux populations	18
4	Modèles à hasard proportionnel	19
4.1	Définition du modèle	19
4.2	Inférence des paramètres	20
4.3	Test d'hypothèse	21
4.3.1	Significativité des coefficients: le test de Wald	21

4.3.2	Significativité du modèle: le test du rapport de vraisemblance	22
4.3.3	Vérification de l'hypothèse de proportionalité	22
5	Table de mortalité	24
5.1	Estimation des probabilités de décès et des taux de mortalité	25
5.1.1	Notations et données	25
5.1.2	Estimation	26
5.2	Lissage et fermeture de la table	27
5.2.1	Lissage paramétrique	28
5.2.2	Lissage de Whittaker-Henderson	29
5.2.3	Fermeture de la table	31
5.2.4	Validation de l'ajustement	32
5.3	Evolution temporelle de la mortalité et effet cohorte	33
5.3.1	Diagramme de Lexis	33
5.3.2	Taux de mortalité par période	33
5.3.3	Taux de mortalité par cohorte	37
5.3.4	Modèle de projection de la mortalité	37
5.4	Annexe: Lissage par moyennes mobiles	42
5.5	Annexe: Whittaker-Henderson vu comme un lissage bayésien	43

Chapter 1

Introduction

Les modèles de durée permettent l'étude statistique du temps avant qu'un événement ne se produise, par exemple la mort d'un organisme en biologie ou la défaillance d'un composant dans un système mécanique. La modélisation de la durée est un problème récurrent en assurance, branche au sein de laquelle on s'intéresse par exemple à la

- Durée de vie humaine
- Durée d'un arrêt de travail
- Durée entre deux sinistres
- Durée dans un état incapacité ou d'invalidité
- Temps avant le défaut d'un emprunteur

Les tables de mortalité, de rachat ou de maintien en capacité sont issus d'analyse de survie et sont en entrée des modèles actuarielles de valorisation des portefeuilles et de gestion des risques.

Exemple 1. Prenons l'exemple d'un contrat d'assurance emprunteur. Ce contrat permet une indemnisation du prêteur (la banque) en cas de défaut de paiement de l'assuré. Il permet également une indemnisation de l'emprunteur en apportant un complément en cas de perte de revenu par exemple en cas d'arrêt de travail. L'assuré peut également résilier son contrat ou racheter son prêt. Ces différentes situations doivent être prise en compte par l'actuaire et font intervenir plusieurs modèles de durée. Le parcours de l'assuré et les hypothèses sont résumées sur la [Figure 1.1](#).

1.1 Distribution d'une durée

1.1.1 Formulation continue

Une durée T est une v.a. sur \mathbb{R}_+ caractérisée par sa densité de probabilité f . On s'intéresse en particulier à la fonction de répartition

$$F(t) = \mathbb{P}(T \leq t) = \int_0^t f(s)ds,$$

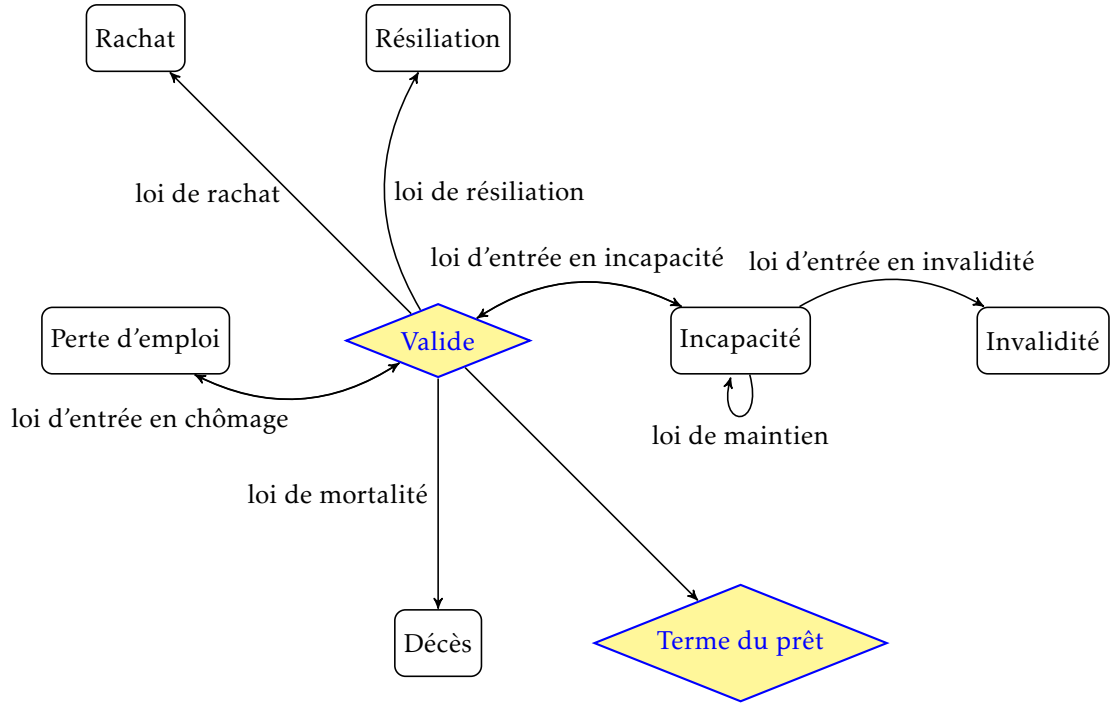


Figure 1.1: Parcours assuré dans le cadre d'un contrat d'assurance emprunteur

et à sa fonction de survie $S(t) = \mathbb{P}(T > t)$. La moyenne de T est donnée par

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

La fonction de survie conditionnellement à la survie jusqu'à l'instant u est donnée par

$$S_u(t) = \mathbb{P}(T > t + u | T > u) = \frac{\mathbb{P}(T > t + u)}{\mathbb{P}(T > u)} = \frac{S(u + t)}{S(u)}.$$

La fonction de hasard h^1 est donnée par

$$h(t) = \frac{\mathbb{P}[T \in (t, t + dt) | T > t]}{dt} = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \ln S(t).$$

Il s'agit de la probabilité que l'évènement se produise à t sachant qu'il ne s'est pas produit avant.

On définit également la fonction de hasard cumulé par

$$H(t) = \int_0^t h(s) ds.$$

On note que

$$S(t) = \exp[-H(t)].$$

La fonction de survie conditionnelle est donnée en terme de la fonction de hasard cumulée par

$$S_u(t) = \exp \left[- \int_u^{t+u} h(s) ds \right] = \exp [-(H(u + t) - H(u))].$$

¹ aussi appelé taux de panne, taux de décès, taux de défaillance suivant le contexte

1.1.2 Formulation discrète

Il arrive que T est exprimé en nombre d'heure, nombre de jour, ou nombre d'année. Cela rend possible la modélisation par une v.a. à valeur dans \mathbb{N} , caractérisée par sa loi de probabilité

$$p(k) = \mathbb{P}(T = k), \quad k \in \mathbb{N}.$$

La fonction de survie est donnée par

$$S(k) = \sum_{l \geq k+1}^{\infty} p(l).$$

On définit, de manière équivalente au cas continu, la fonction de hasard avec

$$h(k) = \mathbb{P}(T = k | T > k - 1) = \frac{p(k)}{S(k-1)}, \quad k \geq 1.$$

On note que

$$\begin{aligned} S(k) &= \mathbb{P}(T > k) \\ &= \mathbb{P}(T > k | T > k - 1) S(k - 1) \\ &= [1 - \mathbb{P}(T \leq k | T > k - 1)] S(k - 1) \\ &= [1 - \mathbb{P}(T = k | T > k - 1)] S(k - 1) \\ &= [1 - h(k)] S(k - 1) \\ &\vdots \\ &= \prod_{l=0}^k [1 - h(l)]. \end{aligned}$$

Remark 1. Dans l'étude de la mortalité, la variable T correspond à l'âge de décès et on note

$$q_x = h(x) = \mathbb{P}(T = x | T > x - 1) = \frac{\mathbb{P}(T = x)}{S(x-1)},$$

la probabilité de décès à l'âge x .

1.1.3 Le passage du continu au discret et inversement

Pour une variable T continue, on peut discrétiser en considérant que $T = k\epsilon$ lorsque $k\epsilon \leq T < (k+1)\epsilon$, ce qui revient à remplacer T par $\epsilon \lfloor T/\epsilon \rfloor$. ϵ est le pas de discrétisation. Le passage du discret au continu se fait par interpolation des fonctions de survie ou de hasard.

1.2 Modélisation statistique

Soient t_1, \dots, t_n des réalisations i.i.d. de la variable T . Les données disponibles pour l'analyse de survie sont souvent incomplètes, on est ainsi confronté au problème de censure et de troncature comme illustré sur la [Figure 1.2](#).

Chaque ligne de la [Figure 1.2](#) correspond à un évènement et nous observons l'occurrence et la durée de chaque évènement. Nous avons

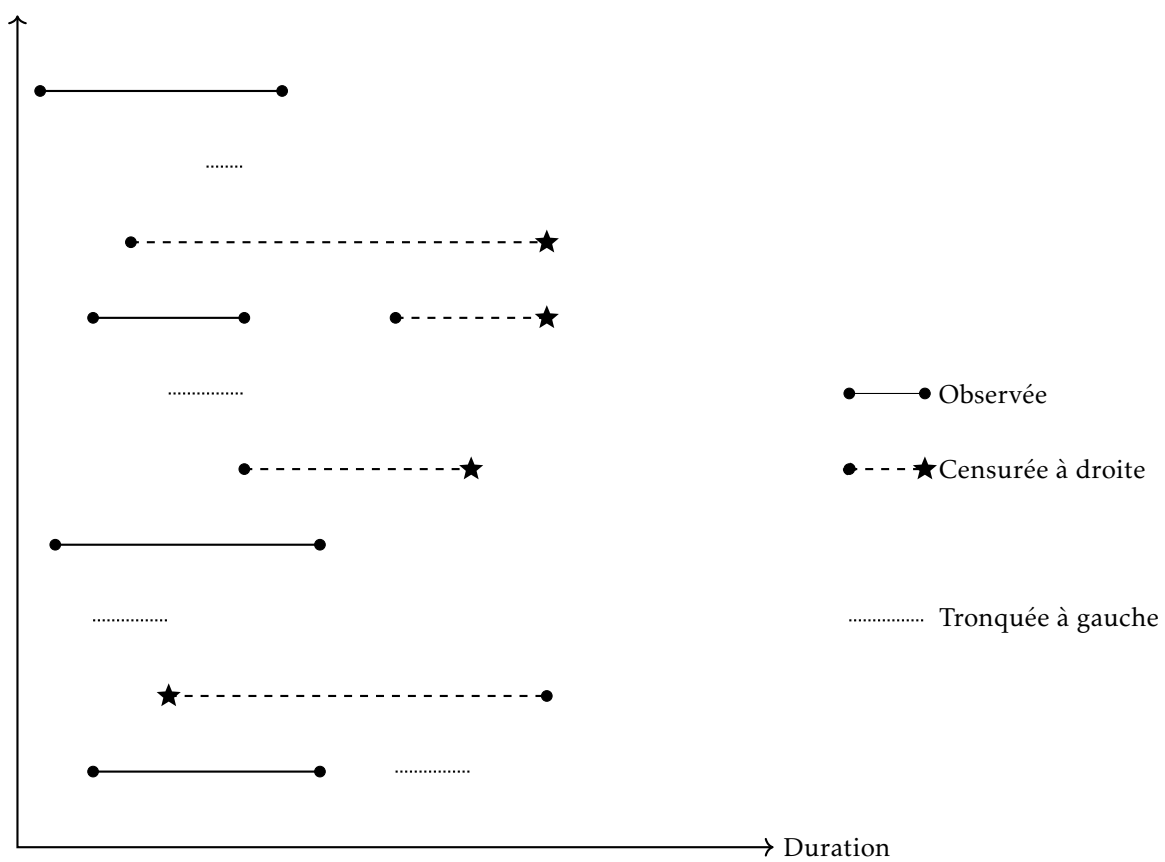


Figure 1.2: Illustration des phénomènes de censure, de troncature et de données répétées

- des données répétées, c'est à dire potentiellement plusieurs évènements pour un même individu
- des données censurées droite, c'est à dire des évènements observé jusqu'à une certaine date
- des données censurées à gauche, c'est à dire des évènements observé depuis une certaine date
- des données tronqués, ici des évènements qui n'ont pas duré assez longtemps pour être observés
- des covariables, qui peuvent influencer la durée d'un évènement

L'objectif du cours est de proposer des méthodes d'estimation pour la fonction de survie S . Le [Chapter 2](#) détaille l'approche paramétrique consiste à supposer que la distribution de T appartient à une famille de distributions paramétriques, c'est à dire que $S(t) := S(t; \theta)$. L'objectif est de trouver la valeur du paramètre $\hat{\theta}$ permettant le meilleur ajustement du modèle aux données. N'importe quelle loi de probabilité sur \mathbb{R}_+ ou sur \mathbb{N} peut convenir. Le [Chapter 3](#) présente les approches non paramétriques dans lesquelles on ne fait pas l'hypothèse d'un modèle sous-jacent, l'estimateur $\hat{S}(t)$ est totalement *data-driven*. Le [Chapter 4](#) présente des méthodes pour étudier la distribution de T conditionnellement à des covariable Z . Les évènements étudiés sont caractérisé par un vecteur Z d'information externe ayant un impact supposé sur la durée de l'évènement. Le [Chapter 5](#) introduit des modèles actuariels (aussi étudié en démographie) pour étudier la mortalité des individus au sein d'une population (ou d'un portefeuille d'assurés).

Chapter 2

Approche paramétrique

Soit t_1, \dots, t_n un échantillon de réalisations i.i.d. de T . L'estimation paramétrique suppose que la loi de T appartient à une famille de distribution caractérisée par un paramètre (vecteur de paramètres) θ . la loi de T est caractérisée par sa densité $f(t; \theta)$ si T est continue ou sa loi $p(t; \theta)$ si T est discrète. l'inférence s'effectue usuellement en maximisant la fonction de vraisemblance définie par

$$L(\mathcal{D}; \theta) = L(t_1, \dots, t_n; \theta) = \begin{cases} \prod_i^n f(t_i; \theta), & \text{dans le cas continu,} \\ \prod_i^n p(t_i; \theta), & \text{dans le cas discret.} \end{cases}$$

Cette écriture correspond au cas où les données sont complètes. comme vu dans le chapitre précédent, ce n'est pas toujours le cas dans les analyses de survie. Nous devons distinguer plusieurs cas de données incomplètes et modifier en conséquence l'écriture de la vraisemblance. Dans la pratique, on cherche à maximiser la log-vraisemblance

$$l(\mathcal{D}; \theta) = \log L(\mathcal{D}; \theta),$$

via des algorithmes d'optimisation¹.

2.1 Fonction de vraisemblance en présence de censure et de troncature

2.1.1 Censure

Definition 1. La v.a. T est censurée à droite (resp. à gauche) si

$$T = C \text{ si } T \geq C \text{ (resp. } T \leq C).$$

Exemple 2. Prenons l'exemple du maintien en incapacité

1. Les assurés qui rentrent en portefeuille en état d'incapacité vont générer une observation censurée à droite si la date de début d'arrêt de travail n'est pas connue exactement.

¹ Les fonctions `nlm` et `optim` en R

2. Les assurés qui vont sortir du poretefeuille alors qu'ils sont encore en incapacité vont générée une observation censuré à droite
3. Un arrêt de travail mal reporté au sens où on enregistre (erreur de reporting) un arrêt de travail plus long que sa durée effective engendre une donnée censurée à gauche.

Dans les applications actuarielles, on fait souvent face à des cas de censure à droite. Supposons que les niveaux de censure soient des réalisations i.i.d. c_1, \dots, c_n d'un v.a. positive C de densité $f_C(\cdot; \theta)$. On suppose que les v.a. T et C sont indépendantes. Les données disponibles sont

$$\mathcal{D} = (x_k, \delta_k)_{k=1, \dots, n} = (t_k \wedge c_k, \mathbb{I}_{t_k \leq c_k})_{k=1, \dots, n}.$$

La vraisemblance s'écrit

$$L(\mathcal{D}; \theta) = \prod_{k=1}^n [f(x_k; \theta) \cdot S_C(x_k; \theta)]^{\delta_k} [f_C(x_k; \theta) \cdot S(x_k; \theta)]^{1-\delta_k}.$$

Il est courant que la censure n'apporte aucune information sur le paramètre du modèle θ . Cela implique que $f_C(\cdot; \theta) = f_C(\cdot)$ et $S_C(\cdot; \theta) = S_C(\cdot)$. On a alors

$$L(\mathcal{D}; \theta) = A \prod_{k=1}^n f(x_k; \theta)^{\delta_k} S(x_k; \theta)^{1-\delta_k} = A \prod_{k=1}^n h(x_k; \theta)^{\delta_k} S(x_k; \theta),$$

où A est une constante en fonction de θ qui peut être négligé dans le cadre d'une procédure d'optimisation de la vraisemblance.

2.1.2 Troncature

Definition 2. La v.a. T est tronquée à droite (resp. à gauche) si T n'est pas observée si $T > C$ (resp. $T \leq C$).

Example 3. Toujours dans l'exemple du maintien en incapacité

1. Si la compagnie d'assurance ne s'intéresse pas aux arrêt de travail de plus de 3 ans car l'incapacité peut être requalifiée en invalidité alors les données sont tronqués à droite car on ne prend pas en compte ces observations.
2. La sécurité social met en place un délai de carence de 3 jours avant le versement d'indemnités journalières donc la sécurité sociale peut exclure les arrêts de travail inférieur à 3 jours. Les données sont tronquées à gauche.

Dans le cas de la troncature, une observation t_k n'est disponible que si $t_k \in [c_1, c_2]$ avec $0 \leq c_1 < c_2 \leq \infty$. La loi observée est alors la loi conditionnelle $T|T \in [c_1, c_2]$. Il faut donc remplacer f , h et S par

$$f_{[c_1, c_2]}(t) = \frac{f(t)}{S(c_1) - S(c_2)} \mathbb{I}_{[c_1, c_2]}(t),$$

$$S_{[c_1, c_2]}(t) = \begin{cases} 1, & t \leq c_1 \\ \frac{S(t) - S(c_2)}{S(c_1) - S(c_2)}, & c_1 \leq t \leq c_2, \\ 0, & t > c_2, \end{cases}$$

et

$$h_{[c_1, c_2]}(t) = \frac{S(t)h(t)}{S(t) - S(c_2)} \mathbb{I}_{[c_1, c_2]}(t),$$

dans l'expression de la vraisemblance.

2.2 Lois paramétriques usuelles

2.2.1 Loi exponentielle

Une v.a. T de loi exponentielle $\text{Exp}(\beta)$ a pour densité

$$f(t) = \frac{e^{-t/\beta}}{\beta} \mathbb{I}_{(0, \infty)}(t),$$

fonction de survie

$$S(t) = e^{-t/\beta},$$

et fonction de hasard

$$h(t) = \frac{1}{\beta} \mathbb{I}_{(0, \infty)}(t).$$

2.2.2 Loi Gamma

Une v.a. T de loi gamma $\text{Gamma}(\alpha, \beta)$ a pour densité

$$f(t) = \frac{t^{\alpha-1} e^{-\beta t}}{\beta^\alpha \Gamma(\alpha)} \mathbb{I}_{(0, \infty)}(t),$$

où $\Gamma(\alpha) = \int_0^\infty e^{-s} s^{\alpha-1} ds$ désigne la fonction gamma. La fonction de survie est donnée par

$$S(t) = \frac{\Gamma_u(t/\beta, \alpha)}{\Gamma(\alpha)},$$

où $\Gamma_u(t, \alpha) = \int_t^\infty e^{-s} s^{\alpha-1} ds$ est la fonction gamma incomplète. La fonction de hasard est donnée par

$$h(t) = \frac{t^{\alpha-1} e^{-\beta t}}{\Gamma_u(t/\beta, \alpha) \beta^\alpha} \mathbb{I}_{(0, \infty)}(t).$$

2.2.3 Loi de Weibull

Une v.a. T de loi de Weibull $\text{Weib}(\alpha, \beta)$ a pour densité

$$f(t) = \frac{\alpha}{\beta} \left(\frac{t}{\beta} \right)^{\alpha-1} e^{-(t/\beta)^\alpha} \mathbb{I}_{(0, \infty)}(t),$$

fonction de survie

$$S(t) = e^{-(t/\beta)^\alpha},$$

et fonction de hasard

$$h(t) = \frac{\alpha}{\beta} \left(\frac{t}{\beta} \right)^{\alpha-1} \mathbb{I}_{(0, \infty)}(t).$$

2.3 Comparaison et sélection d'un modèle

Les trois modèles ont des caractéristiques très différentes. Les fonctions de densité, de survie, de hasard et de hasard cumulée des modèles $\text{Exp}(\beta = 1)$, $\text{Gamma}(\alpha = 2, \beta = 1/2)$ et $\text{Weib}(\alpha = 1/2, \beta = 1/2)$ sont données sur la [Figure 2.1](#). L'approche paramétrique consiste à calibrer plusieurs

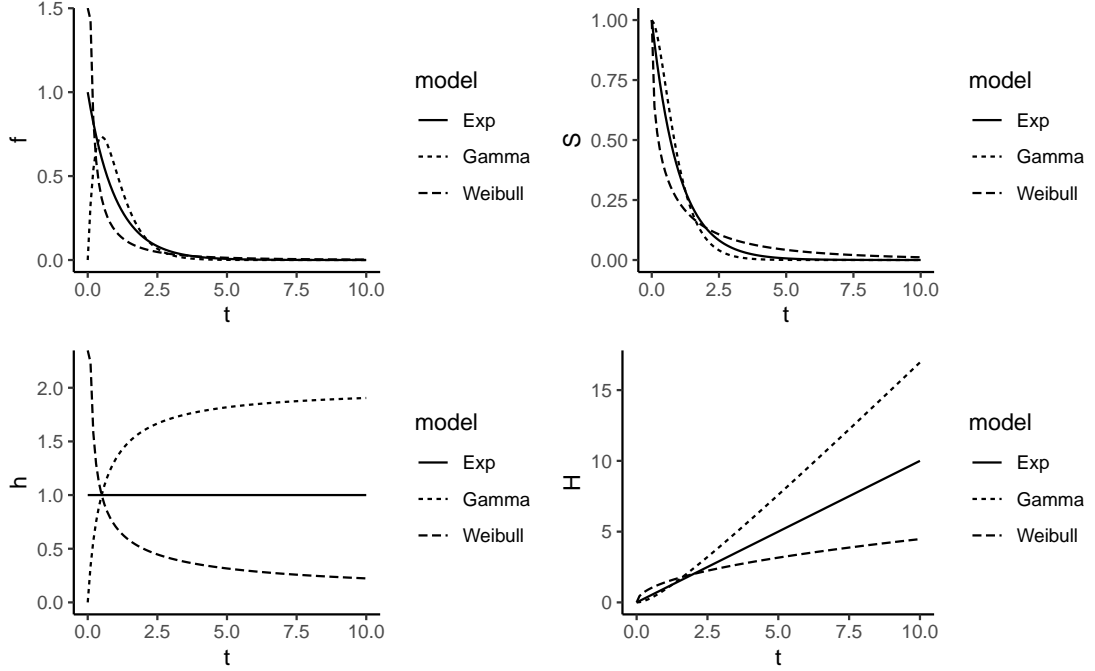


Figure 2.1: Fonction de densité, de survie, de hasard et de hasard cumulée des modèles $\text{Exp}(\beta = 1)$, $\text{Gamma}(\alpha = 2, \beta = 1/2)$ et $\text{Weib}(\alpha = 1/2, \beta = 1/2)$.

modèles puis à choisir le plus adapté grâce à des critères d'information.

2.3.1 Critère d'information

Les critères d'informations mesurent l'ajustement du modèle aux données. Ils sont basés sur la vraisemblance et pénalisent la complexité du modèle (c'est à dire le nombre de paramètres). L'*Akaike Information Criterion* (AIC) est défini par

$$\text{AIC} = 2k - 2l(\mathcal{D}; \hat{\theta}),$$

où k désigne le nombre de paramètre du modèle, voir [Akaike \[1998, 1974\]](#). Le Bayesian Information Criteria (BIC) est défini par

$$\text{BIC} = k \ln(n) - 2l(\mathcal{D}; \hat{\theta}),$$

où n désigne la taille de l'échantillon voir [Schwarz \[1978\]](#). le meilleur modèle est caractérisé par la plus petite valeur d'AIC et de BIC. le terme $-2l(\mathcal{D}; \hat{\theta})$ correspond à la déviance du modèle par rapport au modèle qui aurait généré les données.

2.3.2 Test d'adéquation

Une validation des modèles peut s'effectuer via des test d'adéquation statistique. Les hypothèses du test sont

$$\begin{cases} H_0 & \mathbb{P}(T \leq t) = F(t; \theta) \\ H_1 & \mathbb{P}(T \leq t) \neq F(t; \theta) \end{cases}$$

les statistique de test sont définies par des écarts entre la fonction de répartition du modèle considéré $F(\cdot; \hat{\theta})$ et la fonction de répartition empirique

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{t_i \leq t}.$$

Le test de Kolmogorov-Smirnov s'appuie sur la distance

$$D_n = \sup_{t \geq 0} |F_n(t) - F(t; \theta)|,$$

estimée concretement par

$$D_n = \max(D_-, D_+)$$

où

$$D_- = \max_{i=1, \dots, n} i/n - F(t_{(i)}; \theta), \text{ et } D_+ = \max_{i=1, \dots, n} F(t_{(i+1)}; \theta) - i/n,$$

où $t_{(1)} < \dots < t_{(n)}$ sont ls statistiques d'ordres. On a le résultat de convergence suivant

$$\sqrt{n}D_n \rightarrow K \text{ en loi}, \quad (2.1)$$

où K est une v.a. de loi de Kolmogorov. La fonction de répartition n'a pas une forme explicite mais cette distribution a été tabulé et encodé dans les logiciel statistiques. Soit $K_{1-\alpha}$ tel que

$$\mathbb{P}(K > K_{1-\alpha}) = \alpha.$$

On rejette l'hypothèse H_0 si

$$\sqrt{n}D_n > K_{1-\alpha}.$$

Definition 3. 1. α est le niveau du test, c'est à dire la probabilité de rejeter l'hypothèse nulle alors qu'elle est vraie

$$\mathbb{P}(\sqrt{n}D_n > K_{1-\alpha} | H_0)$$

2. La puissance du test, est la probabilité de rejeter l'hypothèse nulle alors qu'elle est fausse, soit

$$\beta = \mathbb{P}(\sqrt{n}D_n > K_{1-\alpha} | H_1)$$

3. La p -value est définie par

$$p\text{-value} = \mathbb{P}(K > \sqrt{n}D_n),$$

On rejette l'hypothèse nulle au niveau 5% si $p\text{-value} \leq 0.05$.

Remark 2. Le résultat de convergence (2.1) n'est pas valide lorsque les paramètre de la distribution doivent être estimés. Des méthodes de Monte Carlo de type bootstrap paramétrique doivent être mis en oeuvre pour déterminé la valeur critique du test (le $K_{1-\alpha}$).

Chapter 3

Approche non paramétrique

Soit T une variable aléatoire positive et un échantillon t_1, \dots, t_n de réalisations i.i.d. de T . L'estimateur naturelle de la fonction de survie de T est donnée par

$$\widehat{S}_n(t) = \frac{1}{n} \sum_{k=1}^n \mathbb{I}_{t_k > t}. \quad (3.1)$$

La loi des grands nombre indique qu'il s'agit d'un estimateur convergent de la fonction de survie. Dans une analyse de survie, nous devons prendre en compte le phénomène de censure. Pour chaque observation $k = 1, \dots, n$, nous observons

$$\mathcal{D} = (x_k, \delta_k) = (t_k \wedge c_k, \mathbb{I}_{t_k < c_k}),$$

où $(c_k)_{k \geq 1}$ est une suite de réels positives. Nous savons si l'observation est censurée via δ_k , auquel cas la valeur censurée c_k remplace la réalisation t_k . L'utilisation des x_k brutalement dans l'estimateur (3.1) biaise l'estimation. Une alternative qui exploite mieux l'information disponible est l'estimateur de Kaplan-Meier.

3.1 Estimateur de Kaplan-Meier et Nelson-Aalen

Ce problème de valeurs observées censurées a été contournée par [Kaplan and Meier \[1958\]](#) qui ont proposé un estimateur "produit-limite". Supposons que T soit discrète à valeur dans $\{v_0, v_1, \dots\}$ tel que $0 = v_0 < v_1 < \dots$. Notons

$$h_0 = \mathbb{P}(T = v_0), \text{ et } h_k = \mathbb{P}(T = v_k | T > v_{k-1}) \text{ pour } k \geq 1.$$

Remark 3. Si T est un v.a. continu alors on peu discrétiser la loi de T en approchant T par $h \lfloor T/h \rfloor$. Comme $h \lfloor T/h \rfloor \rightarrow T$ lorsque $h \rightarrow 0$ et l'approximation est d'autant meilleur que le pas de discrétisation h est petit.

Theorem 1. L'estimateur de Kaplan-Meier est donnée par

$$\widehat{S}_n^{KP}(t) = \prod_{k: v_k \leq t} \left[1 - \frac{d(v_k)}{n(v_k)} \right], \quad (3.2)$$

où $d(v_k) = \sum_{i=1}^n \delta_i \mathbb{I}_{x_i \in [v_k, v_{k+1})}$ et $n(v_k) = \sum_{i=1}^n \mathbb{I}_{x_i \geq v_k}$.

Remark 4. Pour une v.a. T continue et un pas de discrétisation petit, on a

$$d(v_k) = d(t_k) = 1 \text{ pour } k = 1, \dots, n.$$

Proof. Nous allons considérer que les h_k sont les paramètres du modèle, la vraisemblance des données s'écrit

$$\mathcal{L}(\mathcal{D}; \theta) = \prod_{i=1}^n f(x_i; \theta)^{\delta_i} S(x_i; \theta)^{1-\delta_i},$$

où $\mathcal{D} = (x_i, \delta_i)$, $i = 1, \dots, n$ et $\theta = h_0, h_1, \dots$. Dans le cadre de notre modèle discret nous avons

$$\mathcal{L}(\mathcal{D}; \theta) = \prod_{i=1}^n \prod_{k=0}^{\infty} \left[p_k^{\mathbb{I}_{x_i \in [v_k, v_{k+1})}} \right]^{\delta_i} \left[S_k^{\mathbb{I}_{x_i \in [v_k, v_{k+1})}} \right]^{1-\delta_i},$$

où

$$p_k = \mathbb{P}(T = v_k), \text{ et } S_k = \mathbb{P}(T > v_k).$$

On a

$$p_k = h_k S_{k-1} \text{ et } S_k = \prod_{j=0}^k (1 - h_j).$$

On peut ré-écrire la vraisemblance

$$\begin{aligned} \mathcal{L}(\mathcal{D}; \theta) &= \prod_{i=1}^n \prod_{k=0}^{\infty} h_k^{\mathbb{I}_{x_i \in [v_k, v_{k+1})} \delta_i} \left[\prod_{j=0}^{k-1} (1 - h_j) \right]^{\mathbb{I}_{x_i \in [v_k, v_{k+1})} \delta_i} \left[\prod_{j=0}^k (1 - h_j) \right]^{\mathbb{I}_{x_i \in [v_k, v_{k+1})} (1-\delta_i)} \\ &= \prod_{i=1}^n \prod_{k=0}^{\infty} h_k^{\mathbb{I}_{x_i \in [v_k, v_{k+1})} \delta_i} (1 - h_k)^{-\mathbb{I}_{x_i \in [v_k, v_{k+1})} \delta_i} \left[\prod_{j=0}^k (1 - h_j) \right]^{\mathbb{I}_{x_i \in [v_k, v_{k+1})}} \\ &= \prod_{k=0}^{\infty} h_k^{d(v_k)} (1 - h_k)^{-d(v_k)} \left[\prod_{j=0}^k (1 - h_j) \right]^{\sum_{i=1}^n \mathbb{I}_{x_i \in [v_k, v_{k+1})}} \end{aligned}$$

On passe au log pour obtenir

$$\begin{aligned}
l(\mathcal{D}; \theta) &= \log \mathcal{L}(\mathcal{D}; \theta) \\
&= \sum_{k=0}^{\infty} \left\{ d(v_k) \log(h_k) - d(v_k) \log(1 - h_k) + \left[\sum_{j=0}^k \sum_{i=1}^n \mathbb{I}_{x_i \in [v_k, v_{k+1})} \log(1 - h_j) \right] \right\} \\
&= \sum_{k=0}^{\infty} \{ d(v_k) \log(h_k) - d(v_k) \log(1 - h_k) \} + \sum_{k=0}^{\infty} \left[\sum_{j=0}^k \sum_{i=1}^n \mathbb{I}_{x_i \in [v_k, v_{k+1})} \log(1 - h_j) \right] \\
&= \sum_{k=0}^{\infty} \{ d(v_k) \log(h_k) - d(v_k) \log(1 - h_k) \} + \sum_{j=0}^{\infty} \left[\sum_{k=j}^{\infty} \sum_{i=1}^n \mathbb{I}_{x_i \in [v_k, v_{k+1})} \log(1 - h_j) \right] \\
&= \sum_{k=0}^{\infty} \{ d(v_k) \log(h_k) - d(v_k) \log(1 - h_k) \} + \sum_{j=0}^{\infty} \log(1 - h_j) \sum_{i=1}^n \sum_{k=j}^{\infty} \mathbb{I}_{x_i \in [v_k, v_{k+1})} \\
&= \sum_{k=0}^{\infty} \{ d(v_k) \log(h_k) - d(v_k) \log(1 - h_k) \} + \sum_{j=0}^{\infty} \log(1 - h_j) \sum_{i=1}^n \sum_{k=j}^{\infty} \mathbb{I}_{x_i \in [v_k, v_{k+1})} \\
&= \sum_{k=0}^{\infty} \{ d(v_k) \log(h_k) - d(v_k) \log(1 - h_k) \} + \sum_{j=0}^{\infty} \log(1 - h_j) \sum_{i=1}^n \mathbb{I}_{x_i \geq v_j} \\
&= \sum_{k=0}^{\infty} \{ d(v_k) \log(h_k) - d(v_k) \log(1 - h_k) \} + \sum_{j=0}^{\infty} \log(1 - h_j) n(v_j) \\
&= \sum_{k=0}^{\infty} d(v_k) \log(h_k) - [n(v_k) - d(v_k)] \log(1 - h_k).
\end{aligned}$$

On recherche les h_k qui maximisent la log vraisemblance, c'est à dire solution de

$$\frac{\partial}{\partial h_k} l(\mathcal{D}; \theta) = 0 \Leftrightarrow \frac{d(v_k)}{h_k} - \frac{n(v_k) - d(v_k)}{1 - h_k} = 0.$$

On en déduit que

$$\widehat{h}_k = \frac{d(v_k)}{n(v_k)}.$$

On vérifie également que

$$\frac{\partial^2 l}{\partial h_k \partial h_l}(\mathcal{D}; \theta) = \begin{cases} 0, & \text{si } l \neq k \\ -\frac{d(v_k)}{h_k^2} - \frac{n(v_k) - d(v_k)}{(1 - h_k)^2}, & \text{sinon.} \end{cases}$$

La matrice Hessienne est définie négative. On approche la fonction de survie par

$$S(t) \approx S_k, \text{ pour } t \in [v_k, v_{k+1}),$$

puis on estime par

$$\widehat{S}_n(t) = \prod_{k: v_k \leq t} \left[1 - \frac{d(v_k)}{n(v_k)} \right].$$

□

Pour proposer un estimateur de la fonction de hasard cumulée, on pourrait s'appuyer sur la relation entre la fonction de hasard cumulé et la fonction de survie

$$\widehat{H}_n^{\text{KP}}(t) = -\ln[\widehat{S}_n^{\text{KP}}(t)] = -\sum_{k: v_k \leq t} \ln \left[1 - \frac{d(v_k)}{n(v_k)} \right].$$

L'estimateur de Nelson-Aalen (Nelson [1972] et Aalen [1978]) est un estimateur de la fonction de hasard cumulé qui se base sur l'estimation des taux de hasard instantanée.

Corollary 1.

$$\widehat{H}_n^{NA}(t) = \sum_{k:v_k \leq t} \frac{d(v_k)}{n(v_k)}.$$

Proof. Pour une v.a. T discrète à valeur dans $\{v_0, v_1, \dots\}$, on a

$$H(t) = \sum_{k:v_k \leq t} h_k.$$

On remplace les h_k par leur contre-partie empirique $\widehat{h}_k = d(v_k)/n(v_k)$ déterminé précédemment. □

Remark 5. Pour $d(v_k)/n(v_k)$ suffisamment petit, on a

$$\ln \left[1 - \frac{d(v_k)}{n(v_k)} \right] \approx -d(v_k)/n(v_k),$$

et

$$\widehat{H}_n^{NA}(t) \approx \widehat{H}_n^{KP}(t).$$

3.2 Variance et intervalle de confiance

3.2.1 Rappel des propriétés de l'estimateur du maximum de vraisemblance

Soit $\mathcal{D} = (x_1, \dots, x_n)$ un échantillon i.i.d. de réalisations de X v.a. de densité $f(x; \theta)$. On note $l(\mathcal{D}; \theta)$ la log-vraisemblance du modèle et $\widehat{\theta} = (\widehat{\theta}_1, \dots, \widehat{\theta}_k)$ l'estimateur du maximum de vraisemblance.

Theorem 2. Sous certaines conditions de régularité, on a

$$\widehat{\theta} \sim \text{Normal}(\theta, I_n(\widehat{\theta})^{-1}), \text{ pour } n \rightarrow \infty,$$

où

$$I_n(\widehat{\theta})^{-1} = \left(-\frac{\partial^2}{\partial \theta^2} l(\mathcal{D}; \widehat{\theta}) \right)^{-1},$$

est la matrice d'information de Fisher. De plus,

$$\widehat{\theta}_j \sim \text{Normal}(\theta_j, I_n(\widehat{\theta})^{-1}(j, j)),$$

où $I_n(\widehat{\theta})^{-1}(j, j)$ est le $j^{\text{ème}}$ terme diagonale de la matrice $I_n(\widehat{\theta})^{-1}$. A noter que le terme $I_n(\widehat{\theta})^{-1}(i, j)$ correspond à la covariance entre $\widehat{\theta}_i$ et $\widehat{\theta}_j$.

L'estimateur du maximum de vraisemblance est asymptotiquement normal ce qui permet de construire des intervalles de confiance et d'approcher sa matrice de variance-covariance par

$$\mathbb{V}(\widehat{\theta}) = I_n(\widehat{\theta})^{-1}.$$

Une transformation régulière de l'estimateur du maximum de vraisemblance est aussi asymptotiquement normal.

Proposition 1. Soit $\widehat{\theta}$ l'estimateur du maximum de vraisemblance de θ et g une fonction dérivable, on note ${}^t\nabla g(\theta) = \left(\frac{\partial g}{\partial \theta_1}(\theta), \dots, \frac{\partial g}{\partial \theta_k}(\theta) \right)$ alors

$$g(\widehat{\theta}) \sim \text{Normal}\left(g(\theta), {}^t\nabla g(\widehat{\theta}) I_n(\widehat{\theta})^{-1} \nabla g(\widehat{\theta})\right).$$

Ce résultat est connu sous le nom de méthode delta. Pour plus de détails sur ces résultats voir par exemple [Wasserman, 2013, Chapitre 9].

3.2.2 Application à l'estimateur de Kaplan-Meier

On a $\theta = (h_0, h_1, \dots)$ et

$$\frac{\partial^2}{\partial h_k \partial h_l} l(\mathcal{D}; \theta) = \begin{cases} 0, & \text{si } l \neq k \\ -\frac{d(v_k)}{h_k^2} - \frac{n(v_k) - d(v_k)}{(1 - h_k)^2}, & \text{sinon.} \end{cases} \quad (3.3)$$

La matrice d'information de Fisher est diagonale, ce qui implique une indépendance asymptotique des \widehat{h}_k . En substituant h_k par $\widehat{h}_k = d(v_k)/n(v_k)$, on obtient

$$\mathbb{V}(\widehat{h}_k) \rightarrow \frac{d(v_k)[n(v_k) - d(v_k)]}{n(v_k)^3}, \text{ pour } n \rightarrow \infty.$$

L'estimateur de Kaplan-Meier vérifie

$$\ln \widehat{S}_n(t) = \sum_{k: v_k \leq t} \ln[1 - \widehat{h}_k].$$

On applique la méthode delta avec la fonction $g : x \mapsto \ln(1 - x)$. On a

$$g(\widehat{h}_k) - g(h_k) \sim \text{Normal}(0, \mathbb{V}(\widehat{h}_k) g'(\widehat{h}_k)^2).$$

On en déduit que

$$\mathbb{V}[\ln(1 - \widehat{h}_k)] \rightarrow \frac{d(v_k)}{n(v_k)[n(v_k) - d(v_k)]}, \text{ lorsque } n \rightarrow \infty.$$

L'estimateur de Kaplan Meier vérifie

$$\ln \widehat{S}_n(t) \sim \text{Normal}[\ln S(t), \mathbb{V} \ln \widehat{S}_n(t)]$$

où

$$\mathbb{V} \ln \widehat{S}_n(t) \rightarrow \sum_{k: v_k \leq t} \frac{d(v_k)}{n(v_k)[n(v_k) - d(v_k)]} \text{ pour } n \rightarrow \infty,$$

car les \widehat{h}_k sont asymptotiquement indépendants. On applique une deuxième fois la méthode delta sur $\ln \widehat{S}_n(t)$ avec $g : x \mapsto e^x$ pour aboutir à

$$\widehat{S}_n(t) - S(t) \sim \text{Normal}\left(0, \widehat{S}_n(t)^2 \mathbb{V} \ln \widehat{S}_n(t)\right).$$

L'estimateur de la variance $\widehat{S}_n(t)^2 \mathbb{V} \ln \widehat{S}_n(t)$ de l'estimateur de Kaplan-Meier est parfois appelé estimateur de Greenwood.

3.3 Comparaison de la courbe de survie dans deux populations

L'objectif est de tester statistiquement l'égalité des taux de hasard instantanées au sein de deux populations. Sous (H_0) , on $h_k^1 = h_k^2$ pour tout k . On définit

$$d_j(v_k), \text{ et } n_j(v_k) \text{ pour } j = 1, 2.$$

les décomptes d'évènement au sein des groupes 1 et 2. Sous H_0 , $d_j(v_k)$ suit une loi hypergéométrique $H - \text{Geom}[n(v_k), n_j(v_k), d(v_k)]$.

Définition 4. Une urne contient N boules dont K boules blanches et $N - K$ boules noires. On réalise h tirage aléatoire sans remise dans cette urne et on note X le nombre de boules blanches tirées. La v.a. $X \sim H - \text{Geom}(N, K, h)$ admet une loi de probabilité donnée par

$$\mathbb{P}(X = x) = \frac{\binom{K}{x} \binom{N-K}{h-x}}{\binom{N}{h}}, \text{ pour } \max(0, h - (N - K)) \leq x \leq \min(K, h).$$

Les moyenne et variance ont données par

$$\mathbb{E}(X) = h \frac{K}{N}, \text{ et } \mathbb{V}(X) = h \frac{K}{N} \frac{N-K}{N} \frac{N-h}{N-1}.$$

La statistique du log-rang compare la valeur observée $O_{j,k} = d_j(v_k)$ à la valeur attendue $E_{j,k} = d(v_k) \frac{n_j(v_k)}{n(v_k)}$ sous H_0 avec

$$Z_j = \sum_{k:d(v_k) \geq 1} \frac{(O_{j,k} - E_{j,k})}{\sqrt{\sum_{k:d(v_k) \geq 1} V_{j,k}}} \sim \text{Normal}(0, 1), \text{ pour } j = 1, 2, \quad (3.4)$$

où

$$V_{j,k} = d(v_k) \frac{n_j(v_k)}{n(v_k)} \frac{n(v_k) - n_j(v_k)}{n(v_k)} \frac{n(v_k) - d(v_k)}{n(v_k) - 1}.$$

La convergence (3.4) est une application du théorème centrale limite¹, d'autant plus valide que $\#\{k : d(v_k) \geq 1\}$ est grand. Un autre test du log-rang consiste à élever la statistique de test au carré et alors $Z_j^2 \sim \chi^2(1)$. Une meilleure approximation de la distribution asymptotique de la statistique de test a été obtenu par Peto and Peto [1972]. On peut également pondérer les observation en ajoutant des poids w_j avec

$$Z_j = \sum_{k:d(v_k) \geq 1} w_j \frac{(O_{j,k} - E_{j,k})}{\sqrt{\sum_{k:d(v_k) \geq 1} w_j^2 V_{j,k}}} \sim \text{Normal}(0, 1), \text{ pour } j = 1, 2. \quad (3.5)$$

¹une généralisation du TCL de Lyapunov ou Lindeberg-Feller, voir https://en.wikipedia.org/wiki/Central_limit_theorem

Chapter 4

Modèles à hasard proportionnel

Nous étudions des données t_1, \dots, t_n censurés à droite (censure non informative)

$$(x_i, \delta_i) = (t_i \wedge c_i, \mathbb{I}_{t_i \leq c_i}), \quad i = 1, \dots, n.$$

Chaque évènement i est associé à des caractéristiques ou variable explicatives

$$z_{i1}, \dots, z_{ip} \quad i = 1, \dots, n.$$

L'objectif est d'étudier la loi conditionnelle de la durée T sachant Z .

4.1 Définition du modèle

Le modèle à hasard proportionnel spécifie la fonction de hasard par

$$h(t; z, \beta) = h_0(t) e^{z\beta} = h_0(t) e^{z_1\beta_1 + \dots + z_p\beta_p}, \quad (4.1)$$

où la fonction h_0 est le risque de base, modifié par les caractéristiques individuelles suivant les coefficients de regressions $\beta = (\beta_1 \dots \beta_p)$. L'impact multiplicatif des variables sur le taux de hasard explique le nom du modèle.

Exemple 4. 1. Prenons l'exemple d'un modèle ne comprenant qu'une variable explicative $Z \in \{0, 1\}$ alors le risque instantané s'écrit

$$h(t) = \begin{cases} h_0(t), & \text{si } Z = 0 \\ h_0(t)e^\beta, & \text{sinon.} \end{cases}$$

Le risque est multiplié par e^β lorsque $Z_1 = 1$. Le signe de β indique si le risque augmente $\beta > 0$ ou diminue $\beta < 0$

2. Pour une variable explicative catégorielle $Z \in E$, on crée autant de variables $\mathbb{I}_{Z=z}$ que de modalité $z \in E$. Chacune de ces modalités est alors associée à un coefficient β_z de telle sorte que

$$h(t) = h_0(t) e^{\beta_z}, \quad \text{lorsque } Z = z.$$

3. pour une variable continue $Z \in \mathbb{R}$, le coefficient de regression donne l'impact d'une variation d'amplitude 1. En effet,

$$\frac{h_0(t)e^{\beta \cdot (z+1)}}{h_0(t)e^{\beta \cdot z}} = e^\beta.$$

4.2 Inférence des paramètres

La vraisemblance du modèle pour les données $\mathcal{D} = (x_i, \delta_i, z_i)_{i=1, \dots, n}$ s'écrit

$$\mathcal{L}(\mathcal{D}; \theta) = \prod_{i=1}^n \left[h_0(x_i) e^{t_{z_i} \cdot \beta} \right]^{\delta_i} \cdot \exp \left(-e^{t_{z_i} \cdot \beta} \int_0^{x_i} h_0(s) ds \right), \quad (4.2)$$

où $\theta = (h_0, \beta)$. Il n'est pas possible d'inférer les paramètres avec la vraisemblance (4.2) sans donner une structure au risque de base.

Remark 6. Le modèle (4.1) est le modèle de Cox [1972]. L'étude se focalise plus sur l'impact des covariable Z_1, \dots, Z_p que sur le risque de base h_0 qui peut prendre une forme arbitraire. Cox [1972] définit une vraisemblance partielle qui correspond à la probabilité que les événements se produisent dans l'ordre observé. Nous souhaitons également estimer le risque de base, nous allons donc supposer que h_0 est une fonction constante par morceaux, voir Breslow [1975].

Soit

$$0 < v_1 < v_2 < \dots < v_K < v_{K+1} = \infty,$$

tels que v_1, \dots, v_K correspondent aux observations ordonnées, distinctes et non censuré. Le risque de base est donnée par

$$h_0(t) = \sum_{k=1}^K \alpha_k \mathbb{I}_{[v_k, v_{k+1})}(t), t \geq 0.$$

Les paramètres du modèle sont données par $\theta = (\alpha, \beta)$. La vraisemblance s'écrit

$$\mathcal{L}(\mathcal{D}; \theta) = \prod_{i=1}^n \prod_{k=1}^K \left[\alpha_k e^{t_{z_i} \cdot \beta} \right]^{\delta_i \mathbb{I}_{x_i=v_k}} \cdot \exp \left(-e^{t_{z_i} \cdot \beta} \alpha_k \int_0^{x_i} \mathbb{I}_{[v_k, v_{k+1})}(s) ds \right),$$

On passe au log

$$\begin{aligned} l(\mathcal{D}; \theta) &= \sum_{i=1}^n \sum_{k=1}^K \left[\log(\alpha_k) + t_{z_i} \beta \right] \delta_i \mathbb{I}_{x_i=v_k} - e^{t_{z_i} \beta} \sum_{k=1}^K \alpha_k \int_0^{x_i} \mathbb{I}_{[v_k, v_{k+1})}(s) ds \\ &= \sum_{k=1}^K \log(\alpha_k) d(v_k) + s_k \beta - \alpha_k \sum_{i=1}^n \int_0^{x_i} e^{t_{z_i} \beta} \mathbb{I}_{[v_k, v_{k+1})}(s) ds \\ &= \sum_{k=1}^K \log(\alpha_k) d(v_k) + s_k \beta - \alpha_k \left\{ \sum_{i=1}^n e^{t_{z_i} \beta} \left[(v_{k+1} - v_k) \mathbb{I}_{x_i \geq v_{k+1}} + (x_i - v_k) \mathbb{I}_{[v_k, v_{k+1})}(x_i) \right] \right\} \\ &= \sum_{k=1}^K \log(\alpha_k) d(v_k) + s_k \beta - \alpha_k E_k(\beta), \end{aligned}$$

où $d(v_k)$ est le nombre d'évènements se produisant exactement à l'instant v_k et s_k la somme des caractéristiques $(\sum_i \delta_i z_i)$ de ces évènements. L'estimateur du maximum de vraisemblance $\widehat{\theta}$ est obtenu en annulant les dérivées premières, il vient

$$\widetilde{\alpha}_k = \frac{d(v_k)}{E_k(\beta)},$$

où

$$\begin{aligned} E_k(\beta) &= \sum_{i=1}^n e^{t_i \beta} \left[(v_{k+1} - v_k) \mathbb{I}_{x_i \geq v_{k+1}} + (x_i - v_k) \mathbb{I}_{[v_k, v_{k+1})}(x_i) \right], \\ &= \sum_{i=1}^n e^{t_i \beta} [v_{k+1} \wedge x_i - v_k]_+. \end{aligned}$$

est l'exposition. On recherche β qui annule

$$\frac{\partial}{\partial \beta} l[\mathcal{D}; (\widetilde{\alpha}, \beta)] = \sum_{k=1}^K s_k - d(v_k) \frac{E'_k(\beta)}{E_k(\beta)}.$$

Le programme d'optimisation alterne les mises à jour des $\widetilde{\alpha}$ et β (numériquement pour β). Les estimateurs de α et β sont convergents et asymptotiquement normaux puisque issus du maximum de vraisemblance, la matrice de variance-covariance est donnée par la matrice Hessienne de la log vraisemblance

$$I_n(\theta) = \left[-\frac{\partial^2}{\partial \theta^2} l(\mathcal{D}; \widehat{\theta}) \right].$$

La fonction de hasard cumulée de base est estimée par

$$\widehat{H}_0(t) = \sum_{k=1}^K \frac{d(v_k)}{E_k(\widetilde{\beta})} \mathbb{I}_{v_k \leq t}.$$

Il est possible d'obtenir une loi asymptotique pour cette estimateur, voir [Spiekerman and Lin \[1998\]](#), afin de construire des intervalles de confiance.

4.3 Test d'hypothèse

4.3.1 Significativité des coefficients: le test de Wald

Pour un paramètre univarié β dont l'estimateur $\widehat{\beta}$ est asymptotiquement normal. On peut effectuer un test de Wald pour tester sa nullité

$$(H_0) : \beta = 0 \quad (H_1) : \beta \neq 0.$$

La statistique de test est donnée par

$$\frac{\widehat{\beta}}{\widehat{se}} \sim \text{Normal}(0, 1),$$

où \widehat{se} est une estimation de l'écart-type type de $\widehat{\beta}$. Ici la racine carré du coefficient diagonal correspondant de l'inverse de la matrice de Fisher.

4.3.2 Significativité du modèle: le test du rapport de vraisemblance

Le test du rapport de vraisemblance test des hypothèses du type

$$(H_0) : \theta \in \Theta_0. (H_1) : \theta \notin \Theta_0,$$

où Θ_0 est une restriction de l'espace des paramètres Θ , avec $\Theta_0 \subset \Theta$. La statistique du test correspond est un ratio de log vraisemblance donnée par

$$\lambda = 2 \log \frac{\sup_{\theta \in \Theta} \mathcal{L}(\mathcal{D}; \theta)}{\sup_{\theta \in \Theta_0} \mathcal{L}(\mathcal{D}; \theta)}.$$

L'ensemble Θ_0 consiste souvent en l'ensemble Θ pour lequel certains paramètres sont fixés. Supposons que $\theta = (\theta_1, \dots, \theta_p)$ et que l'on souhaite tester l'hypothèse

$$(H_0) : \theta \in \Theta_0 = \{\theta : \theta_1 = \theta_1^0, \dots, \theta_q = \theta_q^0\}, \text{ tel que } q \leq p,$$

alors

$$\lambda \sim \chi_q^2,$$

où q est différence entre la dimension de Θ et Θ_0 . La significativité du modèle de Cox test l'hypothèse de nullité simultanée de tout les coefficients. Concrètement

$$(H_0) : \beta_1 = 0, \dots, \beta_p = 0 \quad (H_1) : \beta_1 \neq 0, \dots, \beta_p \neq 0,$$

et $\lambda \sim \chi_p^2$.

4.3.3 Vérification de l'hypothèse de proportionalité

Dans le cadre du modèle de Cox, on s'intéresse à la loi jointe du couple (T, Z) . Supposons que Z soit un vecteur aléatoire sur un espace d'état discret E et T soit également une v.a. discrète. Le modèle de Cox donne la fonction de hasard de la durée T sachant le vecteur de covariable Z , c'est à dire

$$\mathbb{P}(T = t | Z = z, T \geq t) = h(t|z) = h_0(t)e^{z\beta}.$$

La loi de $Z|T = t$ est obtenue par

$$\begin{aligned} \mathbb{P}(Z = z | T = t) &= \frac{\mathbb{P}(Z = z, T = t)}{\mathbb{P}(T = t)} \\ &= \frac{\mathbb{P}(Z = z, T = t, T \geq t)}{\sum_{z \in E} \mathbb{P}(T = t, Z = z, T \geq t)} \\ &= \frac{\mathbb{P}(T = t | Z = z, T \geq t) \mathbb{P}(Z = z, T \geq t)}{\sum_{z \in E} \mathbb{P}(T = t | Z = z, T \geq t) \mathbb{P}(Z = z, T \geq t)} \\ &= \frac{h_0(t)e^{z\beta} \mathbb{P}(Z = z, T \geq t)}{\sum_{z \in E} h_0(t)e^{z\beta} \mathbb{P}(Z = z, T \geq t)} \\ &= \frac{\mathbb{E}(e^{\beta Z} \mathbb{I}_{\{Z=z, T \geq t\}})}{\mathbb{E}(e^{\beta Z} \mathbb{I}_{T \geq t})} \end{aligned}$$

En présence d'un échantillon i.i.d. $(t_i, z_i)_{i=1, \dots, n}$, cette loi peut être estimée par

$$\widehat{\mathbb{P}}(Z = z|T = t) = \frac{\sum_{i=1}^n e^{\widehat{\beta} z_i} \mathbb{I}_{\{z_i=z, t_i \geq t\}}}{\sum_{i=1}^n e^{\widehat{\beta} z_i} \mathbb{I}_{t_i \geq t}}.$$

cette estimateur demeure valide en remplaçant les observations t_i par les observations censurées à droite x_i . Soient $v_1 < \dots < v_K$ les temps d'occurrence des évènements distincts et non censurés. On peut comparer les caractéristiques moyennes observées et attendues des évènements s'étant produits aux temps $v_k, k = 1, \dots, K$ en calculant la différence

$$r_k = Z(v_k) - \mathbb{E}(Z|T = v_k) = \frac{\sum_{i=1}^n z_i \mathbb{I}_{x_i=v_k}}{d(v_k)} - \sum_{i=1}^n z_i \frac{e^{\widehat{\beta} z_i} \mathbb{I}_{\{x_i \geq v_k\}}}{\sum_{i=1}^n e^{\widehat{\beta} z_i} \mathbb{I}_{x_i \geq v_k}}.$$

Les r_k sont appelés résidus de [Schoenfeld \[1982\]](#). Nous allons tester l'hypothèse de proportionnalité séparément pour chacune des variables explicatives, cela revient à considérer Z et β comme des scalaires. On définit une fonction de hasard alternative avec

$$h^{\text{alt}}(t|z) = h_0(t) \exp[z\beta(t)] = h_0(t) \exp[z(\widehat{\beta} + \gamma g(t))],$$

pour $\gamma \in \mathbb{R}$ et $g : \mathbb{R}_+ \rightarrow \mathbb{R}$. On test la significativité du coefficient γ avec

$$(H_0) : \gamma = 0.$$

Les résidus se ré-écrivent

$$r_k = Z(v_k) - \mathbb{E}^{\text{alt}}(Z|T = v_k) + [\mathbb{E}^{\text{alt}}(Z|T = v_k) - \mathbb{E}(Z|T = v_k)] = Z(v_k) - M(\widehat{\beta}(v_k), v_k) + [M(\widehat{\beta}(v_k), v_k) - M(\widehat{\beta}, v_k)]$$

On effectue un développement limité de $M(\widehat{\beta}(v_k), v_k)$ au voisinage de $\widehat{\beta}$. Il vient

$$r_k = Z(v_k) - M(\widehat{\beta}(v_k), v_k) + V(\widehat{\beta}, v_k) \gamma g(v_k),$$

où

$$V(\widehat{\beta}, v_k) = \left. \frac{\partial}{\partial x} M(x, v_k) \right|_{x=\widehat{\beta}} \approx \mathbb{E}(Z^2|T = v_k) - \mathbb{E}(Z|T = v_k)^2.$$

L'espérance sous H_0 des résidus standardisés $r_k^* = r_k / V(\widehat{\beta}, v_k)$ vérifie

$$\mathbb{E}(r_k^*) = \gamma g(v_k) \Leftrightarrow r_k^* + \hat{\beta} = \beta(v_k) \text{ (Voir le graphique de l'application en R).}$$

Cela suggère une relation linéaire entre les r_k^* et les $g(v_k)$. On peut choisir pour g une fonction arbitraire. Un choix commun est l'identité. La significativité du coefficient γ est établi par un test de student sur le coefficient γ vu comme la pente de la regression linéaire entre les série r_k^* et $g(t_k)$. La significativité de toutes les covariables simultanément est possible également, grâce aux travaux de [GRAMBSCH and THERNEAU \[1994\]](#).

Chapter 5

Table de mortalité

En assurance-vie, les modèles de projections des cash-flows permettant d'évaluer les engagements de l'assureur repose sur l'utilisation de tables de mortalité. Il s'agit d'un tableau comprenant pour chaque âge x la probabilité de décès avant d'atteindre l'âge $x + 1$. Les probabilités de décès peuvent varier en fonction du genre des individus. la réglementation (Article A132-1 à A132-19 du code des assurances¹) impose l'utilisation de tables réglementaires TH et TF 00 – 02 pour l'évaluation des engagements liés aux garanties décès. Les tables de mortalités peuvent inclure

- l'année calendaire pour prendre en compte une situation particulière liée à une année (par exemple une pandémie mondiale)
- l'année de naissance des individus (génération ou cohortes) pour capter l'allongement progressif de la durée de vie des individus.

L'évaluation des engagements pour les garanties viagère se fait en utilisant des tables réglementaires par génération TGH et TGF 05 qui fournissent une table de mortalité pour chaque année de naissance. Les assureurs peuvent aussi utiliser une table de mortalité d'expérience s'ils considèrent que le risque de mortalité pour leur portefeuille d'assurés diffère significativement du risque pour la population générale. Les probabilités de décès, la méthode de calculs et les données utilisées doivent être audités par un actuair indépendant certifié. Les méthodes vues dans ce chapitre sont présentées dans le cadre de la mortalité mais peuvent s'appliquer dans d'autres contextes pour construire des lois de rachat, d'entrée en incapacité, en invalidité ou en dépendance, des lois de maintien en incapacité.

¹https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000035514715

5.1 Estimation des probabilités de décès et des taux de mortalité

5.1.1 Notations et données

Soit T la durée de vie d'un individu. Supposons que l'on dispose d'information sur un nombre N d'individus

- $\mathcal{D} = \{t_i\}$: âge exacte de décès
- $\mathcal{D} = \{x_i, \delta_i\}$: âge exacte de décès censuré
- $\mathcal{D} = \{E_x^0, d_x\}$: nombre de survivant (initial) et nombre de décès à l'âge x
- $\mathcal{D} = \{E_x^C, d_x\}$: nombre de survivant (central) et nombre de décès à l'âge x

Nous pouvons estimer la fonction de survie $\widehat{S}(t)$. Dans l'étude de la mortalité on s'intéresse à la durée de vie résiduelle à l'âge x défini par

$$T_x \sim T - x | T > x.$$

On s'intéresse aux quantités présentées dans le [Table 5.1](#) Les estimateurs des quantités d'intérêts

E_x^0				Exposition initiale Nombre de survivant jusqu'à l'âge x
E_x^C	$(E_x^0 + E_{x+1}^0)/2$ $\sum_{i=1}^N \tau_{x,i}$			Exposition centrale Cumul des temps d'observation d'individu d'âge x
d_x	$E_x^0 - E_{x+1}^0$			Nombre de décès à l'âge x
q_x	$\mathbb{P}(T_x \leq 1)$	$\mathbb{P}(T \leq x+1 T > x)$	d_x/E_x^0 $1 - e^{-d_x/E_x^C}$	Probabilité de décès d'un individu d'âge x avant d'atteindre l'âge $x+1$
p_x	$\mathbb{P}(T_x > 1)$	$\mathbb{P}(T > x+1 T > x)$	$1 - q_x$	Probabilité de survie d'un individu d'âge x jusqu'à l'âge $x+1$
${}_t p_x$	$\mathbb{P}(T_x > t)$	$\mathbb{P}(T > x+t T > x)$	$\prod_{s=1}^t p_{x+s}$	Probabilité de survie d'un individu d'âge x jusqu'à l'âge $x+t$
${}_t q_x$	$\mathbb{P}(T_x \leq t)$	$\mathbb{P}(T \leq x+t T > x)$	$1 - {}_t p_x$	Probabilité de décès d'un individu d'âge x avant d'atteindre l'âge $x+t$
μ_x	$\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T_x \in [x, x+h])$		$-\log(p_x)$	Taux de décès instantanée à l'âge x
e_x	$\mathbb{E}(T_x)$	$\mathbb{E}(T - x T > x)$	$\sum_{t \geq 0} {}_t p_x$	Espérance de vie résiduelle à l'âge x

Table 5.1: Récapitulatif des notations des tables de mortalités

q_x, p_x et μ_x utilisent les expositions et les nombre de décès d_x . Leur justification est donnée dans la section suivante.

5.1.2 Estimation

La justification de l'expression des estimateurs du [Table 5.1](#) repose sur la maximisation de la vraisemblance des données. Les données de mortalités se limitent aux effectifs par âge E_x^0 , avec $\mathcal{D} = \{E_x^0\}_{x \in \mathbb{N}}$. Un extrait de la table réglementaire TF00 – 02 est donné par [Table 5.2](#).

Age	l_x
0	100,000
1	99,616
2	99,583
3	99,562
4	99,545
\vdots	\vdots
111	4
112	1

Table 5.2: Extrait de la table de mortalité TF00 – 02.

Remark 7. l_x correspond à un nombre de survivant jusqu'à l'âge x . La taille initiale de la population (le radix) est noté l_0 . Il s'agit d'une version normalisée des tailles de population observées telle que

$$l_x = \mathbb{P}(T > x) \cdot l_0$$

Modèle binomial: estimation des q_x

Le nombre décès D_x à l'âge $x \in \mathbb{N}$ (Approche dite discrète) est une variable aléatoire binomial $D_x \sim \text{Bin}(E_x^0, q_x)$. L'application du maximum de vraisemblance conduit à

$$\widehat{q}_x = \frac{d_x}{E_x^0}.$$

Pour $E_x^0 \rightarrow \infty$, nous avons l'approximation suivante

$$\widehat{q}_x \sim \text{Normal}\left(q_x, \frac{\widehat{q}_x(1 - \widehat{q}_x)}{E_x^0}\right),$$

qui permet de construire des intervalle de confiance.

Modèle de Poisson: estimation des μ_x

Le temps de survie T admet une fonction de hasard $s \mapsto \mu_s$ constante entre deux âge x et $x + 1$, c'est à dire

$$\mu_s = \sum_{x \geq 0} \mu_x \mathbb{I}_{[x, x+1)}(s).$$

La probabilité de décès est donnée par

$$q_x = 1 - p_x = 1 - \exp\left(-\int_x^{x+1} \mu_s ds\right).$$

Le taux de décès est estimé par

$$\widehat{\mu}_x = \frac{d_x}{E_x^c}$$

où E_x^c est l'exposition centrale. L'exposition centrale est estimée par

$$E_x^c = \sum_{i=1}^N \tau_{x,i}$$

où les $\tau_{x,i} \in [0, 1]$ sont les durées d'observations de chacun des individus $i = 1, \dots, n$ entre les âges x et $x + 1$.

Remark 8. Pour la plupart des individus

- $\tau_{x,i} = 0$: décès avant l'âge x
- $\tau_{x,i} = 1$: survie jusqu'à l'âge $x + 1$

Si $\tau_{x,i} \in (0, 1)$ alors décès ou sortie (censure) pour l'observation i . En l'absence de mesures précises des observations, on se contentera de

$$E_x^c = \frac{E_x^0 + E_{x+1}^0}{2}.$$

Les $\widehat{\mu}_x$ sont asymptotiquement normaux. L'intervalle de confiance pour les $\widehat{q}_x = 1 - e^{-\widehat{\mu}_x}$ est obtenu par la méthode Delta avec

$$\widehat{q}_x \sim \text{Normal}\left(q_x, \frac{d_x}{(E_x^c)^2} e^{-2d_x/E_x^c}\right).$$

Ce modèle revient à supposer que le nombre de décès D_x à l'âge x suit une loi de Poisson $D_x \sim \text{Poisson}(E_x^c \mu_x)$.

Interpolation aux âges non entiers

Dans l'étude de la mortalité, l'unité de temps privilégiée est l'année. En pratique, on spécifie une répartition des décès au cours de l'année. Soit $t \in (0, 1)$.

1. On peut supposer que les taux de décès instantanée sont constant $\mu_{x+t} = \mu_x$, il s'agit de l'hypothèse de la section précédente, qui conduit à

$${}_t q_x = 1 - (1 - q_x)^t, \quad t \in [0, 1]$$

2. On peut supposer une répartition uniforme des décès sur une année ce qui conduit à

$${}_t q_x = t \cdot q_x, \quad t \in [0, 1]$$

5.2 Lissage et fermeture de la table

La construction d'une table d'expérience comprend deux étapes à savoir

1. Estimation des taux bruts, voir [Section 5.1](#)

2. *Post-processing* des taux bruts

Les probabilités de décès \widehat{q}_x ou les taux de mortalité $\widehat{\mu}_x$ estimées dans la section précédente forme la série des taux bruts. Le *post-processing* a pour objet de limiter l'aspect erratique de la série des taux brut en lissant la courbe. Aux grands âges, les taux bruts obtenus peuvent être très volatile

- Exposition trop faible
- Nombre de décès trop faible

Nous pouvons réduire le bruit inhérent à ces taux bruts via des méthodes de lissage et de fermeture des tables aux grands âges.

5.2.1 Lissage paramétrique

Les méthodes de lissage paramétrique consiste à spécifier une forme paramétrique pour les taux instantanée de décès ou les probabilité de décès. Pour les taux instantanées de décès on trouve

- Le modèle de [Gompertz \[1825\]](#)

$$\mu_x = b \cdot c^x$$

- les lois de [Makeham \[1860\]](#)

$$\mu_x = a + b \cdot c^x, \text{ et } \mu_x = a + h \cdot x + b \cdot c^x.$$

- La loi de Weibull

$$\mu_x = \frac{\alpha}{\beta} \left(\frac{x}{\beta} \right)^{\alpha-1}$$

- Le modèle de [Heligman and Pollard \[1980\]](#)

$$\frac{q_x}{p_x} = A^{(x+B)^C} + D e^{E(\ln x - \ln F)} + G H^x$$

Il s'agit d'un modèle flexible permettant la prise en compte de l'ensemble des caractéristique de la courbe des probabilité de décès aux différents âge. En effet,

- La première composante prend en compte la mortalité infantile et sa décroissance en fonction du temps.
- La deuxième composante comprend le décès accidentel pour prendre en compte le comportement à risques des jeunes homme ou la sur-mortalité des femmes liés à l'accouchement.
- La dernière composante traduit l'augmentation du risque de décès avec l'âge (similaire au modèle de Gompertz)

L'effet de ces trois composantes est résumé sur la [Figure 5.1](#).

La calibration de ces modèles se fait de deux façons

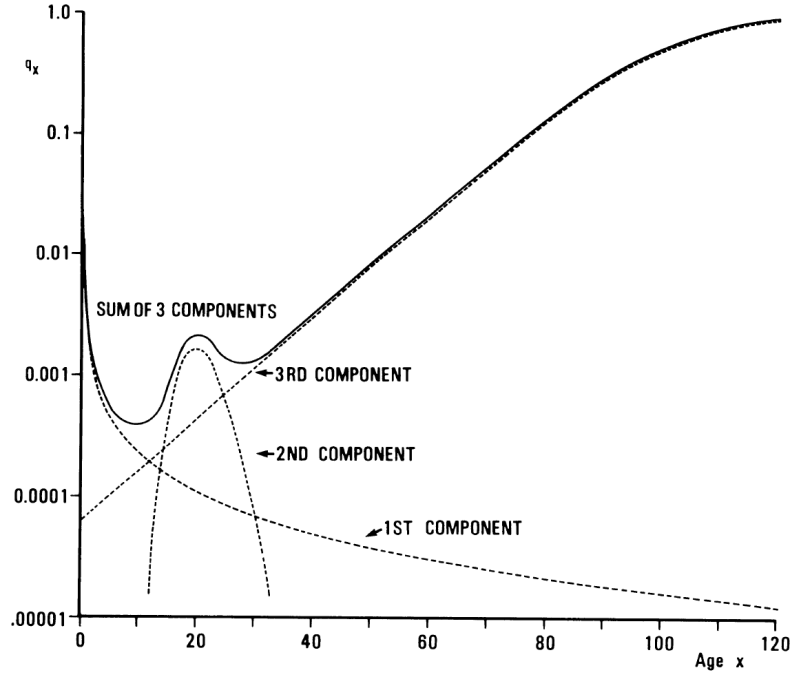


Figure 1. The graduated q_x curve and its three components: Australian national mortality, 1970-72 (males).

Figure 5.1: Illustration de l'impact de chacune des composantes du modèle de Heligman et Pollard.

- Par le maximum de vraisemblance sans utiliser les taux bruts
- Par les moindres carrés ordinaires, en minimisant l'écart entre les taux bruts et les taux du modèle paramétrique avec

$$\arg \min_{\theta \in \Theta} \sum_{x \geq 0} w_x (\hat{q}_x - q_x(\theta))^2 \text{ ou } \arg \min_{\theta \in \Theta} \sum_{x \geq 0} w_x (\hat{\mu}_x - \mu_x(\theta))^2$$

Remark 9. Pour le lissage des probabilités de décès, on peut par exemple choisir les poids inversement proportionnel de la variance des taux brut c'est à dire

$$w_x = \frac{E_x^0}{q_x(1 - q_x)}, \text{ ou } w_x = E_x^c \cdot \mu_x$$

Cela va pénaliser les taux bruts aux grands âges.

5.2.2 Lissage de Whittaker-Henderson

Il est aussi possible de lisser la série des taux bruts sans faire d'hypothèse paramétrique. On cherche à remplacer une série de donnée $y = (y_i)_{i=1, \dots, n}$ (les probabilités de décès, les taux de décès ou une fonction de ces derniers), par $\theta^\lambda = (\theta_i^\lambda)_{i=1, \dots, n}$. Le lissage de Whittaker-Henderson, expliqué par exemple dans le papier de Biessy [2023], consiste en un arbitrage entre la fidélité au données y et la régularité des taux lissés θ^λ . Les taux lissés sont solution du problème d'optimisation

$$\theta^\lambda = \arg \min_{\theta} F(y, w, \theta) + \lambda R(\theta),$$

où $w = (w_i)_{i=1,\dots,n}$ est un vecteur de poids et $\lambda > 0$ qui traduit le compromis entre fidélité F et régularité R . Le critère de fidélité s'écrit

$$F(y, \theta) = \sum_{i=1}^n w_x (y_i - \theta_i)^2.$$

le critère de régularité est donnée par

$$R(\theta) = \sum_{i=0}^{n-z} (\Delta^z \theta)_i^2,$$

où $(\Delta^z \theta)_i$ est l'opérateur de différence progressive d'ordre z défini par

$$(\Delta^z \theta)_i = \sum_{j=0}^z \binom{z}{j} (-1)^{z-j} \theta_{i+j}, \text{ pour } 0 \leq i \leq n-z.$$

On note que

$$\Delta^1 \theta_i = \theta_{i+1} - \theta_i, \Delta^2 \theta_i = \Delta^1 \Delta^1 \theta_i = \theta_i \theta_{i+2} - 2\theta_{i+1} + \theta_i, \dots$$

Remark 10. L'ordre z de l'opérateur de différence progressive indique le niveau de lissage souhaité. Un polynôme de degré $z-1$ est ce que l'on peut obtenir de plus régulier.

- Pour $\lambda = 0$, on retrouve les taux bruts
- Pour $\lambda \rightarrow \infty$, on obtient le polynômes d'ordre $z-1$ qui ajuste le mieux la courbe des taux bruts suivant le critère des moindres carrés.

On définit $W = \mathbf{Diag}(w)$ la matrice diagonale contenant les poids. L'opérateur de différence progressive Δ^z peut être représenté sous forme matricielle avec $D_{n,z} \theta = (D_{n,z} \theta)_i$ où $D_{n,z}$ est une matrice de taille $n-z \times n$. Pour $z=1$ et $z=2$, nous avons respectivement

$$D_{n,1} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}, \text{ et } D_{n,2} = \begin{pmatrix} -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 2 & -1 \end{pmatrix}.$$

le problème d'optimisation s'écrit

$$\theta^\lambda = \arg \min_{\theta} \mathcal{L}(\theta) = \arg \min_{\theta} {}^t(y - \theta)W(y - \theta) + \lambda {}^t(D_{n,z} \theta)D_{n,z} \theta.$$

En notant $P_\lambda = \lambda {}^t D_{n,z} D_{n,z}$ (matrice carrée de taille n^2), on en déduit que²

$$\frac{d}{d\theta} \mathcal{L}(\theta) = -2W(y - \theta) + 2P_\lambda \theta$$

Le système d'équation

$$\frac{d}{d\theta} \mathcal{L}(\theta) = 0$$

²Voici un petit formulaire de dérivation matricielle <https://www.di.ens.fr/~fbach/courses/fall2009/formulaire.pdf>

a pour solution

$$\theta^\lambda = (W + P_\lambda)^{-1} W y.$$

On peut appliquer le lissage brutalement sur les probabilités de décès ou les taux brut à l'aide de la fonction **whittaker** du package **pracma**. L'article de [Biessy \[2023\]](#) rationalise le choix de λ . Pour appliquer cette méthode, des observations y , des poids w , et du paramètre de lissage λ en introduisant un modèle de statistique Bayésienne expliquer dans la [Section 5.5](#). on retiendra que

- Les données a utiliser sont

$$y = \log(d/E_c), \text{ et } w = d.$$

- On obtient un intervalle de crédibilité suivant pour le log des taux de mortalités

$$\ln \mu|d, E^c \in \left[\theta^\lambda \pm \phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{\mathbf{diag}\{(\mathbf{Diag}(d) + P_\lambda)^{-1}\}} \right],$$

où l'opérateur **diag** extrait la diagonale et **Diag** transforme un vecteur en une matrice diagonale.

- Le paramètre de lissage λ est obtenu en maximisant la vraisemblance des données.

On peut utiliser le package **WH** qui comprend également une fonction permettant d'étendre la méthode en dimension 2. C'est utile pour lisser les taux de d'entrée en incapacité ou invalidité qui prennent en compte l'ancienneté en plus de l'âge.

5.2.3 Fermeture de la table

Le problème du manque (voir l'absence) de données aux grands âges oblige parfois d'avoir recours à des méthodes d'extrapolation. Deux méthodes sont décrites ci-après, l'idée est assez proche du lissage paramétrique. On utilise ces méthodes pour obtenir les q_x aux grands âges.

Fermeture de Denuit-Goderniaux

La méthode de [Denuit and Goderniaux \[2005\]](#) consiste à ajuster un modèle linéaire de la forme

$$\log(\widehat{q}_x) = a + bx + cx^2 + \epsilon_x, \text{ pour } x \in [x_{\min}, x_{\max}].$$

où $\epsilon_x \stackrel{i.i.d.}{\sim} \text{Normal}(0, \sigma^2)$. Les bornes x_{\min} et x_{\max} sont fixées de manière arbitraire par le modélisateur en fonction de la fiabilité des probabilités de décès aux grands âges. Deux conditions sont imposées

1. la condition de fermeture $q_{130} = 1$
2. la condition de concavité $q'_{130} = 0$. La probabilité de décès doit augmenter avec l'âge, cette croissance doit ralentir aux grands âges.

Ces deux contraintes reviennent à imposer la relation suivante

$$a + bx + cx^2 = c(130 - x)^2, \text{ pour } x \in [x_{\min}, x_{\max}].$$

Fermeture de Kanisto

La méthode de [Thatcher et al. \[1998\]](#) consiste à ajuster un modèle logistique de la forme

$$\text{logit}(\widehat{q}_x) = \log \frac{q_x}{1 - q_x} = \log(a) + bx, \text{ pour } x \in [x_{\min}, x_{\max}].$$

par les moindres carrés ordinaires.

Remark 11. 1. *L'ordre dans lequel on applique le lissage ou la fermeture est à la discrétion du praticien. On peut décider de lisser puis de fermer ou inversement. La fermeture peut entraîner une discontinuité dans la courbe d'où l'intérêt d'appliquer la procédure de lissage après la fermeture. L'utilisation de la fermeture sur les taux bruts peut nuire à l'estimation des paramètres d'où l'intérêt de lisser d'abord. On peut tout à fait lisser puis fermer et lisser à nouveau.*

2. *La procédure de fermeture permet en premier lieu d'extrapoler les probabilités de décès aux âges non accessibles. On peut également décider de remplacer les probabilités de décès des âges au delà de x_{\min} ou de x_{\max} ou encore prendre une moyenne entre taux brut et taux issus de la fermeture. Voir la note de la SOA³.*

5.2.4 Validation de l'ajustement

Pour s'assurer de la qualité des taux révisés \widehat{y} versus les taux bruts y , pouvant correspondre aux probabilités de décès ou aux taux de décès, on a recours à des tests statistiques. On effectue un test binomial (dit test des signes) pour vérifier que les taux révisés ne sont pas systématiquement au dessus ou en dessous des taux brut. On suppose que l'ajustement pour chaque âge est indépendant et on estime la probabilité π que les taux révisés soient au dessus des taux brut par

$$Z = \sum_{i=1}^n \mathbb{I}_{\widehat{y}_i > y_i}.$$

Sous

$$(H_0): \pi = 1/2$$

La statistique de test suit une loi binomiale $Z \sim (n, 0.5)$. On utilise l'approximation normale

$$\frac{Z - np}{\sqrt{np(1-p)}} \sim \text{Normal}(0, 1).$$

dès que n est assez grand ($n \approx 25$). Il arrive que le *post-processing* des taux brut conduisent à, par exemple, une sous-estimation aux âges jeunes et une sur-estimation aux âges plus avancées. Une telle situation permet de passer le test des signes. On peut alors procéder à un autre test, appelé test des runs de Wald-Wolfowitz voir l'article wikipédia https://en.wikipedia.org/wiki/Wald%E2%80%93Wolfowitz_runs_test.

On peut comparer plusieurs méthodes de lissages en terme d'erreurs aux taux bruts en calculant par exemple des écarts moyens quadratiques

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n \frac{(y_i - \widehat{y}_i)^2}{\widehat{y}_i}.$$

³<https://www.soa.org/globalassets/assets/files/resources/essays-monographs/2005-living-to-100/m-li05-1-ix.pdf>

5.3 Evolution temporelle de la mortalité et effet cohorte

Les tables de mortalité générationnelles sont indexés sur l'année de calendrier t à laquelle les données ont été recoltées ou l'année de naissance des individus $a = t - x$.

$$\mathcal{D} = \bigcup_t \{E_{x,t}, D_{x,t}\} \text{ ou } \bigcup_a \{E_{x,a}, D_{x,a}\}$$

On peut construire autant de table de mortalité "statique" que d'année calendrier ou de cohortes disponibles. Les tables générationnelles TGH et TGF 05 comprennent des projections permettant la tarification des produits d'assurance vie. Ces tables prennent la forme de tableau à double entrée indexé sur l'âge et l'année de naissance, comme sur le [Table 5.3](#). La législation impose

Age	1996	1997	1998	...
0	100000	100000	100000	...
1	99607	99617	99626	...
2	99487	99499	99510	...
3	99435	99448	99460	...
4	99406	99419	99432	...
\vdots	\vdots	\vdots	\vdots	...

Table 5.3: Extrait de la table de mortalité générationnelle TGH 05.

l'utilisation de ces tables dans le cadre des contrats d'assurance vie comprenant des garanties du type rente viagère. La construction des tables générationnelles à partir des données bruts nécessite l'introduction d'un outil graphique: le diagramme de Lexis.

5.3.1 Diagramme de Lexis

Le parcours des individus d'une population est représenté sur un diagramme de Lexis

- Abscisse: $t = \text{année calendrier}$,
- Ordonnée: $x = \text{âge}$,

5.3.2 Taux de mortalité par période

Considérons les évènements se produisant sur la période $[t, t + 1]$, nous intéressons au carré de Lexis de la [Figure 5.2](#)

- $D_{x,t}^U = 1$ compte le nombre de décès d'individu d'âge x au début de la période $[t, t + 1]$
- $D_{x,t}^L = 0$ compte le nombre de décès d'individu d'âge $x - 1$ atteignant l'âge x durant la période $[t, t + 1]$.
- Le nombre de décès au cours de la période est donné par

$$D_{x,t} = D_{x,t}^U + D_{x,t}^L = 1.$$

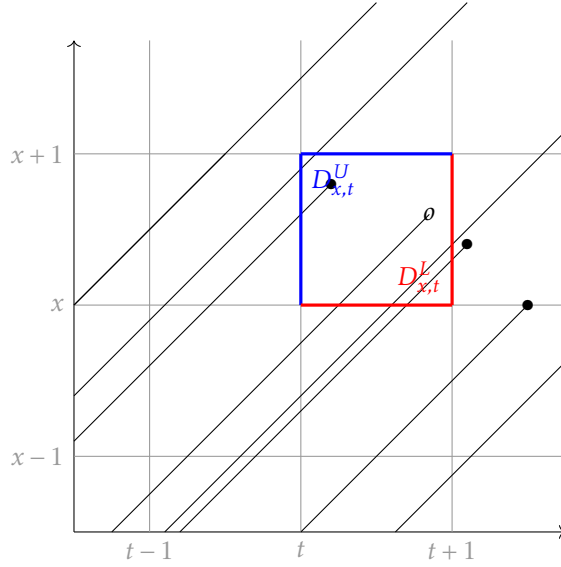


Figure 5.2: Parcours des individus sur le diagramme de Lexis. ● = décès et o = fin d'observation de l'individu (censure à droite)

L'exposition exacte est donnée par la somme des longueurs des segments (divisé par $\sqrt{2}$) dans le carré de Lexis, ici

$$E_{x,t} = 0.08 + 0.3 + 0.6 + 0.42 + 0.33 = 1.73.$$

Le taux de décès d'un individu d'âge x durant la période $[t, t + 1]$ est égale à

$$\mu_{x,t} = \frac{1}{1.73} = 0.57.$$

Remark 12. L'estimateur prenant en compte la longueur de segment pour le calcul de l'exposition est l'estimateur de [Hoem \[1971\]](#). Il est très utilisé en sciences actuarielle, une comparaison avec l'estimateur de Kaplan-Meier dans le cadre d'une application à la mortalité est proposé dans [Guibert and Planchet \[2017\]](#). La méthodologie ci-après suit le protocole de calcul établi dans le cadre du projet HMD⁴ voir [Wilmoth et al. \[2007\]](#).

Si les longueurs de segments ne sont pas accessible l'exposition est approchée en sommant l'exposition des triangles supérieurs et inférieur avec

$$E_{x,t} = E_{x,t}^L + E_{x,t}^U.$$

L'exposition du triangle inférieur comprend $P_{x,t+1}$ la population d'âge x au début de l'année $t + 1$ auquel nous devons ajouter la contribution des décès du triangle inférieur. On a

$$E_{x,t}^L = l_1 P_{x,t+1} + l_2 D_{x,t}^L.$$

L'exposition du triangle supérieur comprend $P_{x,t}$ la population d'âge x au début de l'année t à laquelle nous devons soustraire la contribution des décès du triangle supérieur.

$$E_{x,t}^U = u_1 P_{x,t} - u_2 D_{x,t}^U.$$

⁴<https://www.mortality.org/>

Les coefficients l_1, l_2, u_1 et u_2 peuvent être inférés à l'aide de la distribution des naissances pour les cohortes $t - x - 1$ et $t - x$ comme sur la Figure 5.3. Soit $B_1, B_2 \in [0, 1]$ les v.a. qui indiquent la

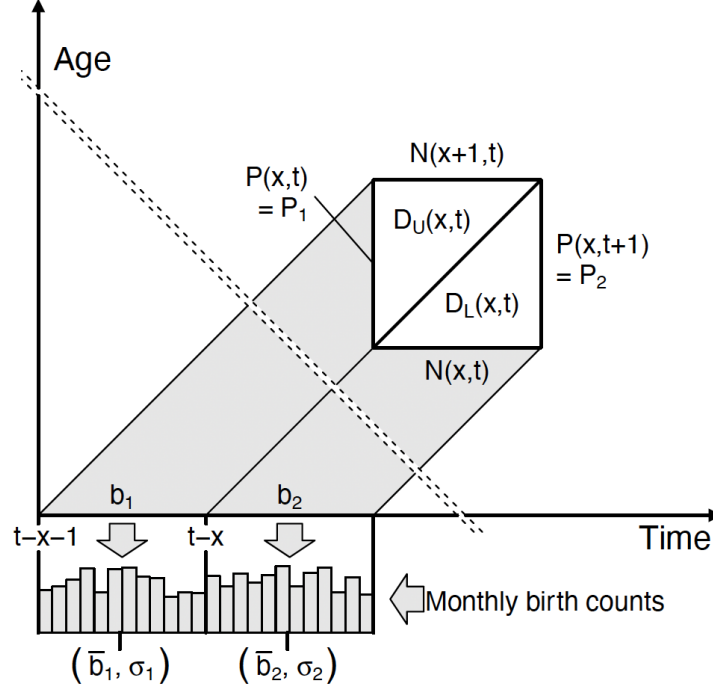


Figure 5.3: Calcul d'exposition central sur un diagramme de Lexis grâce à la distribution des naissances pour les cohortes concernées. (Source: HMD Documentation)

répartition des naissances au sein de la cohorte $t - x - 1$ et $t - x$ respectivement. Soient f_{B_1} et f_{B_2} les densités de B_1 et B_2 . On note également

$$\bar{b}_i = \mathbb{E}(B_i), \text{ et } \sigma_i^2 = \mathbb{V}(B_i), \text{ pour } i = 1, 2.$$

Dans le triangle inférieur, la contribution à l'exposition des individus d'âge x au temps $t + 1$ est donnée par

$$\mathbb{E}[1 - B_2] = (1 - \bar{b}_2), \quad (5.1)$$

La contribution des décès est un peu plus subtil. Un décès dans le triangle inférieur est un point $(t + U, x + V)$ avec (U, V) un couple de variable aléatoire telle que $0 \leq V \leq U \leq 1$. On suppose que la densité jointe de (U, V) est donnée par

$$f_{(U,V)}(u, v) = C_L f_{B_2}(u - v) \mathbb{I}_{0 \leq v \leq u \leq 1}(u, v), \quad (5.2)$$

où C_L est une constante de normalisation. Ce choix entraîne que

- La probabilité que le décès survienne est constante le long de la ligne de vie
- La densité est proportionnelle à la densité de probabilité de B_2

En intégrant (5.2), on identifie $C_L = 1/(1 - \bar{b}_2)$. Pour un décès qui a lieu au point $(t + U, x + V)$, la perte d'exposition est donnée par

$$\int_0^1 \int_0^u (1 - u) \frac{f_{B_2}(u - v)}{1 - \bar{b}_2} dv du = \frac{1 - \bar{b}_2}{2} + \frac{\sigma_2^2}{2(1 - \bar{b}_2)} \quad (5.3)$$

On déduit de (5.1) et (5.3) l'exposition dans le triangle inférieur avec

$$\begin{aligned} E_{x,t}^L &= (1 - \bar{b}_2)(P_{x,t+1} + D_{x,t}^L) - \left(\frac{1 - \bar{b}_2}{2} + \frac{\sigma_2^2}{2(1 - \bar{b}_2)} \right) D_{x,t}^L \\ &= (1 - \bar{b}_2)P_{x,t+1} + \left(\frac{1 - \bar{b}_2}{2} - \frac{\sigma_2^2}{2(1 - \bar{b}_2)} \right) D_{x,t}^L \\ &= l_1 P_{x,t+1} + l_2 D_{x,t}^L \end{aligned}$$

Dans le triangle supérieur, la contribution des individus d'âge x à l'année t à l'exposition est donnée par

$$\mathbb{E}(B_1) = \bar{b}_1. \quad (5.4)$$

Un décès dans le triangle supérieur est un point $(t + U, x + V)$ avec $0 \leq U \leq V \leq 1$. La densité jointe de (U, V) est donnée par (pour les mêmes raison que précédemment)

$$f_{(U,V)}(u, v) = C_U f_{B_1}(u + 1 - v) \mathbb{I}_{0 \leq u \leq v \leq 1}(u, v), \quad (5.5)$$

En intégrant (5.5), on identifie $C_U = 1/\bar{b}_1$. Pour un décès qui a lieu au point $(t + U, x + V)$, le gain d'exposition est donnée par

$$\int_0^1 \int_u^1 u \frac{f_{B_1}(u + 1 - v)}{\bar{b}_1} dv du = \frac{\bar{b}_1}{2} + \frac{\sigma_1^2}{2\bar{b}_1} \quad (5.6)$$

On déduit de (5.4) et (5.6) l'exposition dans le triangle inférieur avec

$$\begin{aligned} E_{x,t}^U &= \bar{b}_1(P_{x,t} - D_{x,t}^U) + \left(\frac{\bar{b}_1}{2} + \frac{\sigma_1^2}{2\bar{b}_1} \right) D_{x,t}^U \\ &= \bar{b}_1 P_{x,t} - \left(\frac{\bar{b}_1}{2} - \frac{\sigma_1^2}{2\bar{b}_1} \right) D_{x,t}^U \\ &= u_1 P_{x,t} - u_2 D_{x,t}^U \end{aligned}$$

En l'absence d'information sur la distribution des naissances au cours des l'années au sein des cohortes, on suppose que $B_1, B_2 \sim \text{Unif}(0, 1)$. Il vient alors

$$E_{x,t} \approx \frac{1}{2}[P_{x,t} + P_{x,t+1}] + \frac{1}{6}[D_{x,t}^L - D_{x,t}^U] = 1.7.$$

Le taux de décès d'un individu d'âge x durant la période $[t, t + 1]$ est égale à

$$\widehat{\mu}_{x,t} = \frac{1}{1.7} = 0.59.$$

Remark 13. Si nous n'avons accès qu'au nombre de décès $D_{x,t}$ d'individus d'âge x pendant l'année t alors on peut considérer que $D_{x,t}^U = D_{x,t}^L = D_{x,t}/2$. Pour plus de détail sur les calculs d'exposition en prenant en compte les données de natalité, voir le travail de Boumezoued [2020].

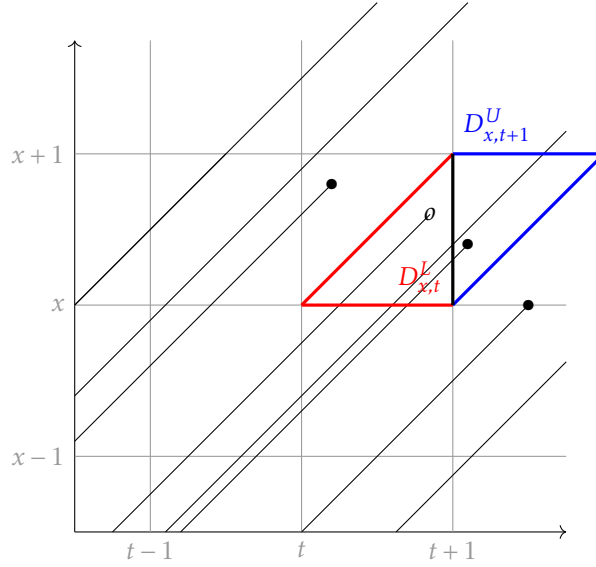


Figure 5.4: Parcours des individus sur le diagramme de Lexis. • = décès et o = fin d'observation de l'individu (censure à droite)

5.3.3 Taux de mortalité par cohorte

Considérons les événements concernant les individus d'âge x nés l'année $a = t - x$, nous intéressons au parallélogramme de Lexis de la Figure 5.4 L'exposition est donnée par la somme des longueurs des segments dans le parallélogramme, soit

$$E_{x,a} = 0.8 + 1 + 0.4 = 2.2.$$

Le taux de décès des individus d'âge x né l'année a est donnée par

$$q_{x,a} = 1/2.2 = 0.45.$$

Si la longueur des segments est inconnue alors au vu des calculs effectués pour les taux de mortalité par période, il vient

$$E_{x,a} = P_{x,t+1} + \left(\frac{1 - \bar{b}}{2} + \frac{\sigma^2}{2(1 - \bar{b})} \right) D_{x,t}^L - \left(\frac{\bar{b}}{2} + \frac{\sigma^2}{2\bar{b}} \right) D_{x,t+1}^U$$

avec $a = t - x$, voir Figure 5.5. Si l'information sur les moments de la distribution des naissances n'est pas disponible alors on suppose une distribution uniforme qui mène à la simplification suivante

$$E_{x,a} = P_{x,t+1} + \frac{1}{3}(D_{x,t}^L - D_{x,t+1}^U)$$

puis

$$\widehat{q}_{x,a} = 1/2 = 0.5.$$

5.3.4 Modèle de projection de la mortalité

L'objectif est de rendre compte de la tendance des taux de mortalité $\mu_{x,t}$ et probabilités de décès $q_{x,t}$ au cours du temps. Dans la section précédente, nous avons estimé des expositions dites

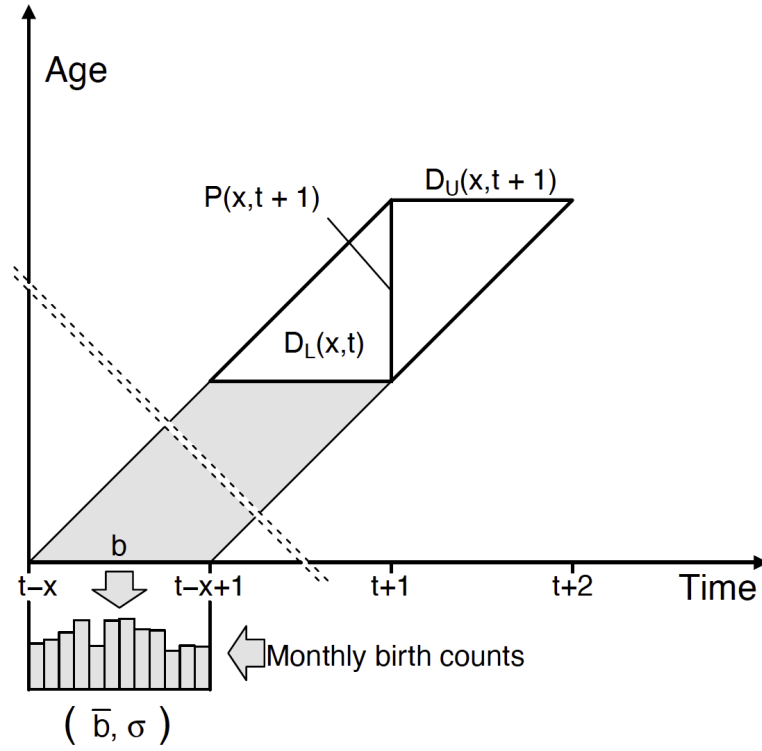


Figure 5.5: Calcul d'exposition central sur un diagramme de Lexis grâce à la distribution des naissances pour la cohortes concernée. (Source: HMD Documentation)

centrales $E_{x,t}^c$. Les taux de mortalités sont estimés sur la base de l'exposition centrale avec

$$\mu_{x,t} = \frac{D_{x,t}}{E_{x,t}^c}.$$

Les probabilité de décès requiert usuellement l'exposition initiale $E_{x,t}^0$ avec

$$q_{x,t} = \frac{D_{x,t}}{E_{x,t}^0}.$$

Les deux expositions sont approximativement lié par $E_{x,t}^0 \approx E_{x,t}^c + D_{x,t}/2$. On utilisera la notation $E_{x,t}$ pour l'exposition centrale ou initiale lorsque le contexte est clair. Les probabilité de décès et taux de mortalité sont liés par

$$q_{x,t} = 1 - \exp(-\mu_{x,t}).$$

Les modèles de mortalité stochastiques, aussi appelé GAPC (*Generalized Age-Period-Cohort*) comprennent 4 composantes.

1. Une composante aléatoire

$$D_{x,t} \sim \text{Poisson}(\mu_{x,t} \cdot E_{x,t}^c), \text{ avec } \mathbb{E}(D_{x,t}) = E_{x,t}^c \cdot \mu_{x,t},$$

ou

$$D_{x,t} \sim \text{Bin}(E_{x,t}^0, q_{x,t}), \text{ avec } \mathbb{E}(D_{x,t}) = E_{x,t}^0 \cdot q_{x,t},$$

2. Une composante systématique

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^N \beta_x^{(i)} \kappa_t^{(i)} + \beta_x^{(0)} \gamma_{t-x},$$

où

- Le terme α_x caractérise l'effet statique de l'âge sur la mortalité
- N est le nombre de terme de type age-période. Les termes $\kappa_t^{(i)}$, $i = 1, \dots, N$ décrivent la tendance de la mortalité au cours du temps modulé par l'âge via les termes $\beta_x^{(i)}$
- Le terme γ_{t-x} tient compte d'un possible effet cohorte modulé par l'âge via le terme $\beta_x^{(0)}$.

Les termes age-dépendant β_x^i peuvent être des fonctions prédéterminé de l'âge $\beta_x^i = f^i(x)$ ou bien des coefficients sans structures préalables. Les termes périodes-dépendant κ_t^i et cohortes-dépendants γ_{t-x} sont considérés comme des processus stochastiques. Les méthodes d'études des série chronologiques s'appliquent pour effectuer les prévisions de l'évolution de la mortalité. Cette structure permet de prendre en compte la plupart des modèles de mortalités, voir [Hunt and Blake \[2020\]](#).

3. La fonction de lien g entre la composante aléatoire et la composante systématique avec

$$g \left[\mathbb{E} \left(\frac{D_{x,t}}{E_{x,t}} \right) \right] = \eta_{x,t}.$$

Dans le cadre du modèle de Poisson, on utilise la fonction \log^5 . Pour le modèle binomial la fonction logit^6 . Ce sont les fonctions de lien canonique des modèles linéaires généralisés, voir [Currie \[2014\]](#) pour une discussion sur les fonctions de liens dans le cadre des modèles de mortalités stochastiques.

4. Les contraintes sur les paramètres permettant de rendre le modèle identifiable. Si le prédicteur est invariant lorsqu'on applique une transformation v définie par

$$v(\theta) := \tilde{\theta} = \left(\tilde{\alpha}_x, \tilde{\beta}_x^{(1)}, \dots, \tilde{\beta}_x^{(N)}, \tilde{\kappa}_t^{(1)}, \dots, \tilde{\kappa}_t^{(N)}, \tilde{\beta}^{(0)}, \tilde{\gamma}_{t-x} \right).$$

au vecteur de paramètres

$$\theta = \left(\alpha_x, \beta_x^{(1)}, \dots, \beta_x^{(N)}, \kappa_t^{(1)}, \dots, \kappa_t^{(N)}, \beta^{(0)}, \gamma_{t-x} \right)$$

alors des contraintes sont nécessaires. La conséquence est une perte de degré de liberté sans changer le prédicteur $\eta_{x,t}$.

Exemple 5. Voici quelques exemples de modèles de type GAPC.

⁵ $\log(\mu_{x,t})$

⁶ $\log(q_{x,t}/(1 - q_{x,t}))$

1. Dans le modèle de [Lee and Carter \[1992\]](#), on utilise le modèle de Poisson et le prédicteur est donnée par

$$\eta_{x,t} = \alpha_x + \beta_x \kappa_t.$$

La composante temporelle est modélisée par une marche aléatoire avec une tendance, c'est à dire

$$\kappa_t = \delta + \kappa_{t-1} + \epsilon_t, \text{ avec } \epsilon_t \sim \text{Normal}(0, \sigma^2).$$

Le modèle de Lee-Carter n'est pas identifiable au sens où la transformation

$$(\alpha_x, \beta_x, \kappa_t) \mapsto \left(\alpha_x + c_1 \beta_x, \frac{\beta_x}{c_2}, c_2(\kappa_t - c_1) \right)$$

n'a aucun effet sur les $\eta_{x,t}$. On impose donc les conditions suivante

$$\sum_x \beta_x = 1 \text{ et } \sum_t \kappa_t = 0,$$

ce qui revient à fixer c_1 et c_2 de la façon suivante

$$c_1 = \frac{1}{n} \sum_t \kappa_t, \text{ et } c_2 = \sum_x \beta_x,$$

où n est le nombre d'année d'historique.

2. Le modèle APC, voir [Clayton and Schiffers \[1987\]](#), s'appuie aussi sur le modèle de Poisson et suppose que

$$\eta_{x,t} = \alpha_x + \kappa_t + \gamma_{t-x}.$$

Le modèle APC n'est pas identifiable car il est invariant par rapport aux deux transformations suivantes

$$(\alpha_x, \kappa_t, \gamma_{t-x}) \mapsto (\alpha_x + \phi_1 - \phi_2 x, \kappa_t + \phi_2 t, \gamma_{t-x} - \phi_1 - \phi_2(t-x))$$

et

$$(\alpha_x, \kappa_t, \gamma_{t-x}) \mapsto (\alpha_x + c_1, \kappa_t - c_1, \gamma_{t-x}),$$

où c_1, ϕ_1 et ϕ_2 sont des constantes dans \mathbb{R} . L'identifiabilité est obtenue en imposant les contraintes suivantes

$$\sum_t \kappa_t = 0, \sum_{a=t_1-x_k}^{t_n-x_1} \gamma_a = 0, \sum_{a=t_1-x_k}^{t_n-x_1} a \gamma_a = 0,$$

où $x_1 = \min x$, $t_1 = \min t$, $x_k = \max x$ et $t_n = \max t$. L'effet cohorte fluctue autour de 0 sans faire apparaître de tendance. Les contraintes sont imposées d'abord par une regression linéaire de γ_{t-x} sur $t-x$ avec

$$\gamma_{t-x} = \phi_1 + \phi_2(t-x) + \epsilon_{t-x}, \epsilon_{t-x} \sim \text{Normal}(0, \sigma^2).$$

puis la contrainte sur la composante temporelle

$$c_1 = \frac{1}{n} \sum_t \kappa_t,$$

3. Le modèle CBD de Cairns et al. [2006] s'appuie sur le modèle binomiale et un prédicteur donné par

$$\eta_{x,t} = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)}.$$

Les composantes temporelles sont modélisées par une marche aléatoire bivariée. Ce modèle est identifiable donc aucune contrainte n'est imposée.

Les modèles sont ajustées au données via le maximum de vraisemblance via le package StMoMo, voir Villegas et al. [2018]. Soit $\mathcal{D} = (D_{x,t}, E_{x,t})_{x,t}$ les données disponibles et $\theta = (\mu_{x,t})_{x,t}$ ou $(q_{x,t})_{x,t}$ les paramètres. La log-vraisemblance s'écrit

$$l(\mathcal{D}; \theta) = \sum_x \sum_t w_{x,t} \{D_{x,t} \log(\mu_{x,t} E_{x,t}) - E_{x,t} \mu_{x,t} - \log(D_{x,t}!)\}$$

dans le modèle de Poisson et

$$l(\mathcal{D}; \theta) = \sum_x \sum_t w_{x,t} \left\{ \log \left(\frac{E_{x,t}}{D_{x,t}} \right) + D_{x,t} \log(q_{x,t}) + (E_{x,t} - D_{x,t}) \log(1 - q_{x,t}) \right\}.$$

dans le modèle binomial, avec

$$w_{x,t} = \begin{cases} 1, & \text{si l'observation } (x, t) \text{ est utilisée pour la calibration,} \\ 0, & \text{sinon.} \end{cases}$$

L'ajustement du modèle est mesuré à l'aide des résidus définis, pour chaque observation (x, t) par

$$r_{x,t} = \text{sign}(D_{x,t} - \widehat{D}_{x,t}) \sqrt{\frac{\text{dev}(x, t)}{\phi}},$$

où

- $\widehat{D}_{x,t} = \widehat{\mu}_{x,t} E_{x,t}^c$ ou $\widehat{q}_{x,t} E_{x,t}^0$ est la prédiction du modèle
- La déviance en chaque point (x, t) est donnée par

$$\text{dev}(x, t) = \begin{cases} 2 \left[D_{x,t} \log \left(\frac{D_{x,t}}{\widehat{D}_{x,t}} \right) - (D_{x,t} - \widehat{D}_{x,t}) \right] & \text{pour le modèle de Poisson,} \\ 2 \left[D_{x,t} \log \left(\frac{D_{x,t}}{\widehat{D}_{x,t}} \right) + (E_{x,t}^0 - D_{x,t}) \log \left(\frac{E_{x,t}^0 - D_{x,t}}{E_{x,t}^0 - \widehat{D}_{x,t}} \right) \right] & \text{pour le modèle Binomial.} \end{cases}$$

- $\phi = \text{Dev}/(K - \nu)$ avec

$$\text{Dev} = \sum_x \sum_t w_{x,t} \text{dev}(x, t), \quad K = \sum_x \sum_t w_{x,t},$$

est le nombre d'observation utilisée pour la calibration et ν est le nombre de paramètre effectif du modèle (eu égard au jeu de contraintes).

L'ajustement global du modèle est mesuré via les critère d'information standard comme l'AIC et le BIC. La capacité prédictive du modèle s'évalue au moyen d'une erreur moyenne absolue ou quadratique calculée le cadre d'une procedure de validation croisée

$$\text{MAE} = \frac{1}{\bar{K}} \sum_x \sum_t (1 - w_{x,t}) |D_{x,t} - \widehat{D}_{x,t}|,$$

avec $\bar{K} = \sum_x \sum_t (1 - w_{x,t})$. La comparaison de la mortalité au sein de deux populations s'articule autour d'un outil graphique: le *Standardized Mortality Ratio*. Il s'agit du graphique représentant le ratio des probabilités de décès sur des probabilités de décès de référence. On compare ainsi

- Les probabilité de décès dans deux pays
- Les probabilité de décès pendant une pandémie VS en temps normal
- Les probabilités de décès au sein d'un portefeuille de contrats d'assurance vie VS les probabilités de décès en population générale pour justifier la pertinence de l'usage d'une table d'expéirenece plutôt que réglementaire
- les probabilités de décès attendues Vs les probabilités de décès observées.

Remark 14. *Il est aussi possible de réaliser une inférence Bayésienne via le package StanMoMo décrit dans Barigou et al. [2022].*

5.4 Annexe: Lissage par moyennes mobiles

La technique de lissage la plus simple consiste à prendre la moyenne des taux brut au voisinage de chaque âge. On remplace les taux bruts \widehat{q}_x par

$$q_x^h = \sum_{k=-h}^h w_x^k \widehat{q}_{x+k},$$

avec $w_x^k \geq 0$, $k = -h, \dots, h$ et $\sum_{k=-h}^h w_x^k = 1$ pour tout $x \geq 0$. Une façon naïve de choisir les poids consiste à prendre

$$w_x^k = \frac{1}{2h+1}.$$

Une méthode plus élaborée fait intervenir un noyau de lissage $K : \mathbb{R}_+ \mapsto \mathbb{R}_+$ qui est une fonction décroissante. Les taux lissé sont donné par

$$q_x^h = \sum_{y \geq 0} \frac{K[(x-y)/h]}{\sum_{y \geq 0} K[(x-y)/h]} \widehat{q}_y.$$

Les noyaux usuels incluent

- Le noyau uniforme

$$K(u) = \frac{1}{2} \mathbb{I}_{|u| < 1}$$

- Le noyau triangle

$$K(u) = (1 - |u|) \mathbb{I}_{|u| < 1}$$

- Le noyau Epanechnikov

$$K(u) = \frac{3}{4} (1 - u^2) \mathbb{I}_{|u| < 1}$$

- Le noyau gaussien

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

Le lissage par noyau nécessite de choisir une fenêtre de lissage h .

5.5 Annexe: Whittaker-Henderson vu comme un lissage bayésien

La paramétrisation du lissage de Whittaker-Henderson s'établit en définissant un modèle de statistique Bayésienne. On suppose que les observations ont une loi normale

$$y|\theta \sim \text{Normal}(\theta, W^{-1}),$$

et le vecteur de paramètre admet une loi a priori

$$\theta \sim \text{Normal}(0, P_\lambda^{-1}).$$

La loi a posteriori du paramètre a une densité vérifie

$$\begin{aligned} f_{\theta|y}(\theta) &= \frac{f(y|\theta)f(\theta)}{f(y)} \\ &\propto \exp\left(-\frac{1}{2}\left[{}^t(y-\theta)W(y-\theta) + {}^t\theta P_\lambda\theta\right]\right). \end{aligned}$$

L'estimateur ponctuel du maximum a posteriori coïncide avec celui des moindres carrés trouvé précédemment. En effet, on a

$$\theta^\lambda = \arg \max_{\theta} f_{\theta|y}(\theta) = (W + P_\lambda)^{-1} W y.$$

On effectue un développement limité de $\ln f_{\theta|y}(\theta)$ au voisinage de $\theta = \theta^\lambda$, il vient

$$\ln f_{\theta|y}(\theta) = \ln f_{\theta|y}(\theta^\lambda) + {}^t \frac{\partial \ln f_{\theta|y}(\theta)}{\partial \theta} \Big|_{\theta=\theta^\lambda} (\theta - \theta^\lambda) + \frac{1}{2} {}^t (\theta - \theta^\lambda) \frac{\partial^2 \ln f_{\theta|y}(\theta)}{\partial \theta^2} \Big|_{\theta=\theta^\lambda} (\theta - \theta^\lambda),$$

où

$$\frac{\partial \ln f_{\theta|y}(\theta)}{\partial \theta} \Big|_{\theta=\theta^\lambda} = 0 \text{ et } \frac{\partial^2 \ln f_{\theta|y}(\theta)}{\partial \theta^2} \Big|_{\theta=\theta^\lambda} = -(W + P_\lambda)$$

On en déduit, en passant le développement à l'exponentiel que

$$\begin{aligned} f_{\theta|y}(\theta) &\propto \exp\left(\ln f_{\theta|y}(\theta^\lambda) - {}^t(\theta - \theta^\lambda)(W + P_\lambda)(\theta - \theta^\lambda)\right) \\ &\propto \exp\left(-{}^t(\theta - \theta^\lambda)(W + P_\lambda)(\theta - \theta^\lambda)\right) \end{aligned}$$

On reconnaît la densité d'une loi normale multivariée $M\text{-Normal}(\theta^\lambda, (W + P_\lambda)^{-1})$ et on en déduit un intervalle de crédibilité de la forme

$$\theta|y \in \left[\theta^\lambda \pm \phi\left(1 - \frac{\alpha}{2}\right) \sqrt{\text{diag}\{(W + P_\lambda)^{-1}\}} \right]$$

avec probabilité $1 - \alpha$. Contrairement à l'opérateur Diag , l'opérateur diag extrait la diagonale d'une matrice carrée. Pour choisir les données et le vecteur de poids, on considère le modèle de Poisson pour lequel la fonction de hasard est constante entre deux âges et on suppose que

$$\mu = \exp(\theta)$$

où $\mu = (\mu_x)_{x \in \mathcal{X}}$ les taux de décès aux âges $x \in \mathcal{X}$. Dans le cadre de ce modèle la log-vraisemblance (sans les termes ne dépendant pas de θ) s'écrit

$$l(\theta; \mathcal{D}) = {}^t \theta d - {}^t \exp(\theta) E^c,$$

où $\mathcal{D} = (d, E^c) = \{(d_x, E_x^c)\}_{x \in \mathcal{X}}$, et $\exp(\cdot)$ désigne la fonction exponentielle appliquée composante par composante. On rappelle que les nombre de décès d_x suivent des lois de Poisson de paramètres $\mu_x E_x^c$. Les dérivées d'ordre 1 et 2 de la log vraisemblance sont données par

$$\frac{\partial}{\partial \theta} l(\theta; \mathcal{D}) = d - \exp(\theta) \odot E^c, \quad \frac{\partial^2}{\partial \theta^2} l(\theta; \mathcal{D}) = -\text{diag}\{\exp(\theta) \odot E^c\}.$$

On en déduit que

$$\widehat{\theta} = \ln(d/E^c) \sim \text{Normal}(\ln \mu, \text{diag}(d)^{-1})$$

Cela justifie d'appliquer le lissage de Whittaker-Henderson aux vecteurs d'observations $y = \ln(d/E^c)$ avec le vecteur de poids $w = d$. L'interprétation bayésienne précédente permet d'en déduire un intervalle de crédibilité avec

$$\ln \mu | d, E^c \in \left[\widehat{\theta} \pm \phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{\text{diag}\{(\text{Diag}(d) + P_\lambda)^{-1}\}} \right],$$

où

$$\widehat{\theta} = (\text{Diag}(d) + P_\lambda)^{-1} \text{Diag}(d) (\ln(d) - \ln(E^c)).$$

Le paramètre de lissage λ a été fixé jusqu'à présent dans la calibration. Il est possible d'en déterminer une valeur "optimale" en le considérant comme un paramètre. La vraisemblance des données sachant le paramètre λ est donnée par

$$l(\mathcal{D}; \lambda) = f(y|\lambda) = \int f(y, \theta|\lambda) d\theta = \int f(y|\theta) f(\theta|\lambda) d\theta, \quad (5.7)$$

avec

$$\theta|\lambda \sim M - \text{Normal}(0, P_\lambda^{-1}).$$

On cherche λ qui maximise (5.7), il faut employer des méthodes numériques et des approximation via par exemple des développements de Taylor. Les détails sont fournis dans [Biessy \[2023\]](#). Le package [WH](#) effectue le lissage de Whittaker-Henderson en suivant ces principes sans fine-tuning de la part de l'utilisateur.

Bibliography

- Odd Aalen. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6(4), jul 1978. doi: 10.1214/aos/1176344247.
- Hirotougu Akaike. Information theory and an extension of the maximum likelihood principle. In *Springer Series in Statistics*, pages 199–213. Springer New York, 1998. doi: 10.1007/978-1-4612-1694-0_15.
- Hirotougu Akaike. *A New Look at the Statistical Model Identification*, pages 215–222. Springer New York, 1974. ISBN 9781461216940. doi: 10.1007/978-1-4612-1694-0_16.
- Karim Barigou, Pierre-Olivier Goffard, Stéphane Loisel, and Yahia Salhi. Bayesian model averaging for mortality forecasting using leave-future-out validation. *International Journal of Forecasting*, mar 2022. doi: 10.1016/j.ijforecast.2022.01.011.
- Guillaume Biessy. Revisiting whittaker-henderson smoothing, 2023.
- Alexandre Boumezoued. Improving HMD mortality estimates with HFD fertility data. *North American Actuarial Journal*, 25(sup1):S255–S279, jan 2020. doi: 10.1080/10920277.2019.1672567.
- N. E. Breslow. Analysis of survival data under the proportional hazards model. *International Statistical Review / Revue Internationale de Statistique*, 43(1):45, apr 1975. doi: 10.2307/1402659.
- Andrew J. G. Cairns, David Blake, and Kevin Dowd. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, 73(4): 687–718, dec 2006. doi: 10.1111/j.1539-6975.2006.00195.x.
- D. Clayton and E. Schifflers. Models for temporal variation in cancer rates. II: Age–period–cohort models. *Statistics in Medicine*, 6(4):469–481, jun 1987. doi: 10.1002/sim.4780060406.
- D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, jan 1972. doi: 10.1111/j.2517-6161.1972.tb00899.x.
- Iain D. Currie. On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal*, 2016(4):356–383, jul 2014. doi: 10.1080/03461238.2014.928230.

- Michel Denuit and Anne-Cécile Goderniaux. Closing and projecting lifetables using log-linear models. *Bulletin of the Swiss Association of Actuaries*, page 29, 2005.
- Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. <https://www.jstor.org/stable/41140508> <https://doi.org/10.1017/s0020268100040257>. *Philosophical Transactions of the Royal Society of London*, 115:513–583, dec 1825. doi: 10.1098/rstl.1825.0026.
- PATRICIA M. GRAMBSCH and TERRY M. THERNEAU. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526, 1994. doi: 10.1093/biomet/81.3.515.
- Quentin Guibert and Frédéric Planchet. UTILISATION DES ESTIMATEURS DE KAPLAN-MEIER PAR GÉNÉRATION ET DE HOEM POUR LA CONSTRUCTION DE TABLES DE MORTALITÉ PROSPECTIVES. working paper or preprint, April 2017. URL <https://hal.science/hal-01509483>.
- L. Heligman and J. H. Pollard. The age pattern of mortality. *Journal of the Institute of Actuaries*, 107(1):49–80, jan 1980. doi: 10.1017/s0020268100040257.
- Jan M. Hoem. Point estimation of forces of transition in demographic models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 33(2):275–289, jul 1971. doi: 10.1111/j.2517-6161.1971.tb00879.x.
- Andrew Hunt and David Blake. On the structure and classification of mortality models. *North American Actuarial Journal*, 25(sup1):S215–S234, feb 2020. doi: 10.1080/10920277.2019.1649156.
- E. L. Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, jun 1958. doi: 10.1080/01621459.1958.10501452.
- Ronald D. Lee and Lawrence R. Carter. Modeling and forecasting u.s. mortality. *Journal of the American Statistical Association*, 87(419):659–671, sep 1992. doi: 10.1080/01621459.1992.10475265.
- William Matthew Makeham. On the law of mortality and the construction of annuity tables. *The Assurance Magazine and Journal of the Institute of Actuaries*, 8(6):301–310, jan 1860. doi: 10.1017/s204616580000126x.
- Wayne Nelson. Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14(4):945–966, nov 1972. doi: 10.1080/00401706.1972.10488991.
- Richard Peto and Julian Peto. Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society. Series A (General)*, 135(2):185, 1972. doi: 10.2307/2344317.

- David Schoenfeld. Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1):239–241, 1982. doi: 10.1093/biomet/69.1.239.
- Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2), mar 1978. doi: 10.1214/aos/1176344136.
- C. F. Spiekerman and D. Y. Lin. Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association*, 93(443):1164–1175, sep 1998. doi: 10.1080/01621459.1998.10473777.
- A Roger Thatcher, Väinö Kannisto, and James W Vaupel. The force of mortality at ages 80 to 120. 1998.
- Andrés M. Villegas, Vladimir K. Kaishev, and Pietro Millossovich. Stmomo: An r package for stochastic mortality modeling. *Journal of Statistical Software*, 84(3), 2018. doi: 10.18637/jss.v084.i03.
- Larry Wasserman. *All of Statistics A Concise Course in Statistical Inference*. Springer, 2013. ISBN 9780387217369.
- John R Wilmoth, Kirill Andreev, Dmitri Jdanov, Dana A Glei, C Boe, M Bubenheim, D Philipov, V Shkolnikov, and P Vachon. Methods protocol for the human mortality database. *University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock*. URL: <http://mortality.org> [version 31/05/2007], 9:10–11, 2007.