

Modèles de durée

M1 DUAS - Université de Strasbourg

Pierre-O Goffard

updated on April 12, 2023

Contents

1	Introduction	3
1.1	Un premier exemple: présentation des données	4
1.2	Distribution d'une durée	6
1.2.1	Formulation continue	6
1.2.2	Formulation discrète	6
1.2.3	Le passage du continu au discret et inversement	7
1.3	Modélisation statistique	7
1.3.1	Approche non paramétrique	7
1.3.2	L'approche paramétrique	9
1.4	Ajout de facteurs explicatifs	10
2	Approche paramétrique	12
2.1	Fonction de vraisemblance en présence de censure et de troncature	12
2.1.1	Troncature	13
2.2	Lois paramétriques usuelles	14
2.2.1	Loi exponentielle	14
2.2.2	Loi Gamma	14
2.2.3	Loi de Weibull	14
2.3	Comparaison et sélection d'un modèle	14
2.3.1	Critère d'information	15
2.3.2	Test d'adéquation	15
3	Approche non paramétrique	17
3.1	Estimateur de Kaplan-Meier et Nelson-Aalen	17
3.2	Variance et intervalle de confiance	19
3.2.1	Rappel des propriétés de l'estimateur du maximum de vraisemblance	19
3.2.2	Application à l'estimateur de Kaplan-Meier	20
3.3	Comparaison de la courbe de survie dans deux populations	21
4	Modèles à hasard proportionnel	22
4.1	Définition du modèle	22

4.2	Inférence des paramètres	23
4.3	Test d'hypothèse	24
4.3.1	Significativité des coefficients: le test de Wald	24
4.3.2	Significativité du modèle: le test du rapport de vraisemblance	24
4.3.3	Vérification de l'hypothèse de proportionalité	25
5	Table de mortalité	27
5.1	Estimation des probabilités de décès et des taux de mortalité	28
5.1.1	Notations et données	28
5.1.2	Estimation	29
5.2	Lissage et fermeture de la table	30
5.2.1	Lissage paramétrique	31
5.2.2	Lissage non paramétrique	32
5.2.3	Fermeture de la table	34
5.3	Evolution temporelle de la mortalité et effet cohorte	35
5.3.1	Diagramme de Lexis	36
5.3.2	Taux de mortalité par période	38
5.3.3	Taux de mortalité par cohorte	41
5.3.4	Modèle de projection de la mortalité	42

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

L'analyse de survie vise à répondre à des questions du type

- Quel proportion de la population sera toujours en vie après une certaine date?
- Parmi ceux qui ont survécu, à quel taux vont-ils nous quitter?
- Plusieurs causes de sortie peuvent-elles être prise en compte? Par exemple dans le cadre d'un contrat d'assurance vie de type épargne un assuré peut sortir du portefeuille pour décès ou en cas de rachat.
- Comment certaines circonstances de l'événement ou caractéristiques individuelles impactent-elles la durée de vie?
 - le genre
 - l'âge
 - être fumeur ou non

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. La parcours assuré et les hypothèse sont résumées sur la [Figure 1.1](#).

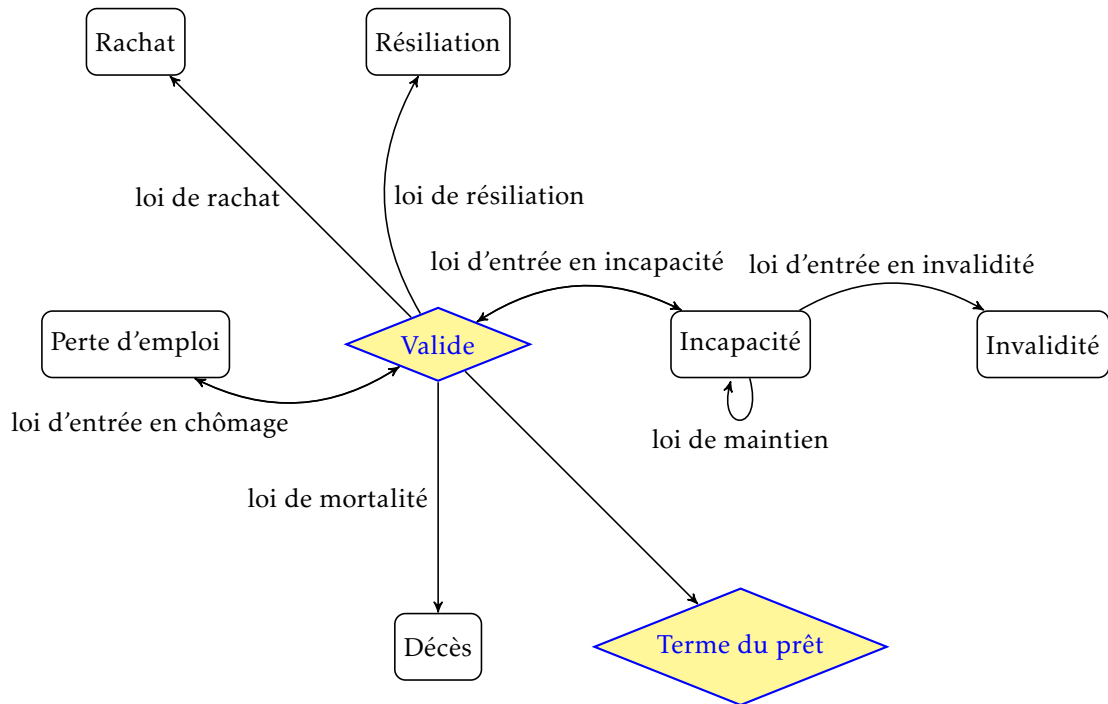


Figure 1.1: Parcours assuré dans le cadre d'un contrat d'assurance emprunteur

1.1 Un premier exemple: présentation des données

On étudie l'entrée et le maintien en incapacité au sein d'une population d'assuré sont étudié sur la période allant du 01/01/2003 au 31/12/2008. Le tableau [Table 1.1](#) montre les 5 premières observations du jeu de données Ce jeu de données comprends 560,725 lignes, chaque observations

Sexe	CSP	ModeSouscription	TypeArret	DateNaissance	DateSurvenance	DateEntree	DateSortie	MotifReprise
H	NCA	NONCSPEC	Accident du travail	1962-12-06	2006-10-07	2006-11-04	2006-11-13	3 - Termine
F	ENP	NONCSTAN	Maladie	1969-01-01	2006-08-05	2006-08-05	2006-09-10	3 - Termine
F	ENP	NONCSTAN	Maladie	1969-01-01	2005-01-14	2005-01-14	2005-01-30	3 - Termine
F	ENP	NONCSTAN	Maladie	1969-01-01	2005-11-18	2006-01-01	2006-01-01	3 - Termine
F	CAD	NONCSTAN	Maladie	1958-01-02	2004-06-01	2005-10-26	2005-11-30	2 - Invalide

Table 1.1: Extrait du jeu de incapacité

correspond à une entrée en incapacité avec potentiellement plusieurs évènements par individus. Le jeu de données contient 9 variables

- Sexe
- CSP
- ModeSouscription
- TypeArret = S'agit-t-il d'un accident du travail ou d'un arrêt maladie
- dateNaissance
- DateSurvenance = date de déclenchement de l'entrée en incapacité
- DateEntree = date de début d'indemnisation
- DateSortie = date de fin d'indemnisation
- MotifReprise = raison de fin d'incapacité

On peut représenter les données comme sur la [Figure 1.2](#). Chaque ligne de la [Figure 1.2](#)

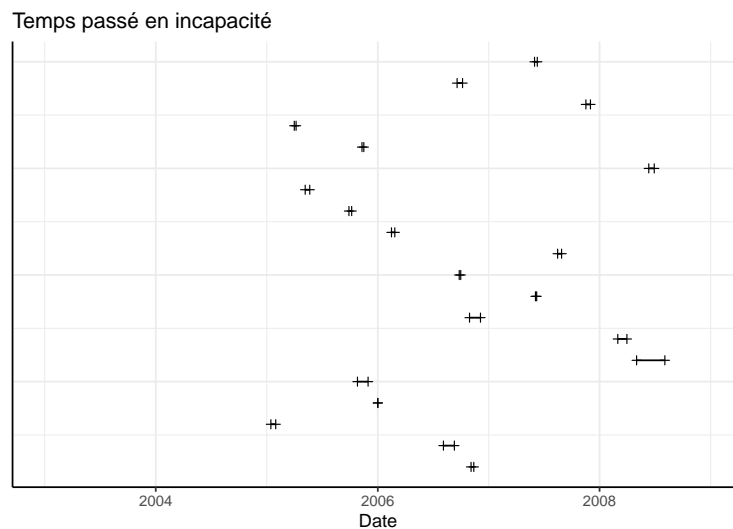


Figure 1.2: Occurrence et longueur des arrêts de travail par individu

correspond à une entrée en incapacité et nous observons l'occurrence et la temps passé en incapacité. Nous avons

- des données répétées, c'est à dire potentiellement plusieurs entrées en incapacité pour un même individu
- des données censurées droite, c'est à dire des arrêts de travail qui se sont poursuivis après la date de fin d'observation (ici au 31/12/2008)
- des données censurées à gauche, des arrêts de travail qui ont débuté avant la date de début d'observations (ici au 04/01/2003)
- des données tronquées la période d'observation comprend 2188 jours, les temps passés en incapacité sont donc compris entre 0 et 2188 jours.

- des covariables, qui peuvent influencer le temps passé en incapacité par exemple est ce que l'entrée en incapacité est due à un accident du travail ou une maladie.

1.2 Distribution d'une durée

1.2.1 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

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

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

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

1.2.2 Formulation discrète

Dans le cas des données d'entrée et de maintien en incapacité, T est exprimé en nombre de jour. 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

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

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

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

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

le taux de mortalité à l'âge x .

1.2.3 Le passage du continu au discret et inversement

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

1.3 Modélisation statistique

Soit t_1, \dots, t_n des réalisations i.i.d. de la variable T . En prenant l'exemple des données d'arrêt de travail ce correspond à ignorer le caractère répété des observations et la potentielle censure. Le tableau [Table 1.2](#) donne quelques statistiques descriptives.

	n	mean	sd	q25	q50	q75	min	max
T	560,725	25	62	4	10	23	0	1440

Table 1.2: Statistiques descriptives des temps passé en incapacité

1.3.1 Approche non paramétrique

La distribution empirique d'une v.a. est souvent visualiser graphiquement au travers d'un histogramme et d'une boîte à moustache comme sur la [Figure 1.3](#). La hauteur des barres de l'histogramme sont données par $\sum_i \sum_j \mathbb{I}_{[kh, (k+1)h)}(t_{i,j})$, en divisant par n , par h et en prenant h

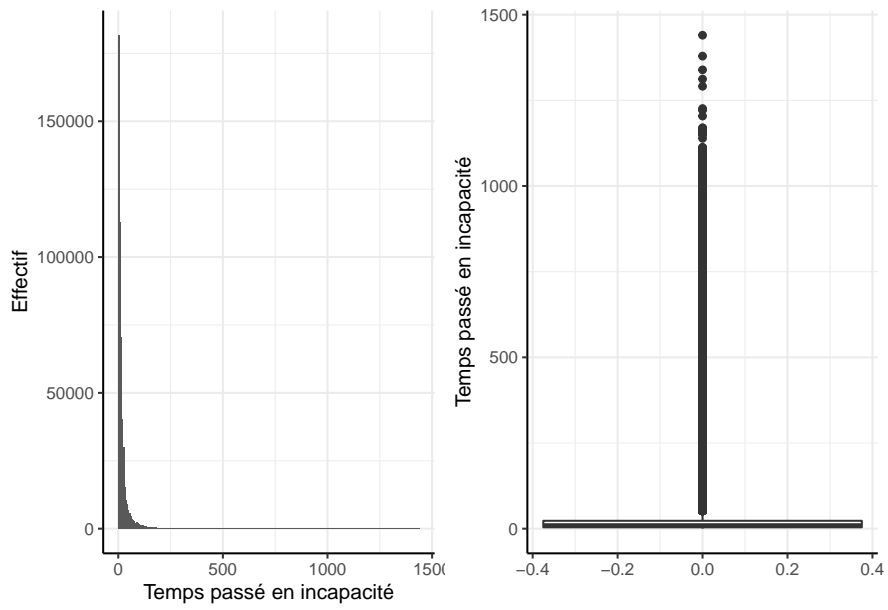


Figure 1.3: Histogramme et boîte à moustache des $t_{i,j}$

petit, on retrouve une estimation de la densité de probabilité. La fonction de survie peut être estimée par sa contre partie empirique avec

$$\widehat{S}(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{t_i > t}. \quad (1.1)$$

De même pour la fonction de hasard cumulé avec

$$\widehat{H}(t) = -\ln \widehat{S}(t).$$

Les estimations obtenues sont reportées sur la [Figure 1.4](#). Nous verrons plus tard comment

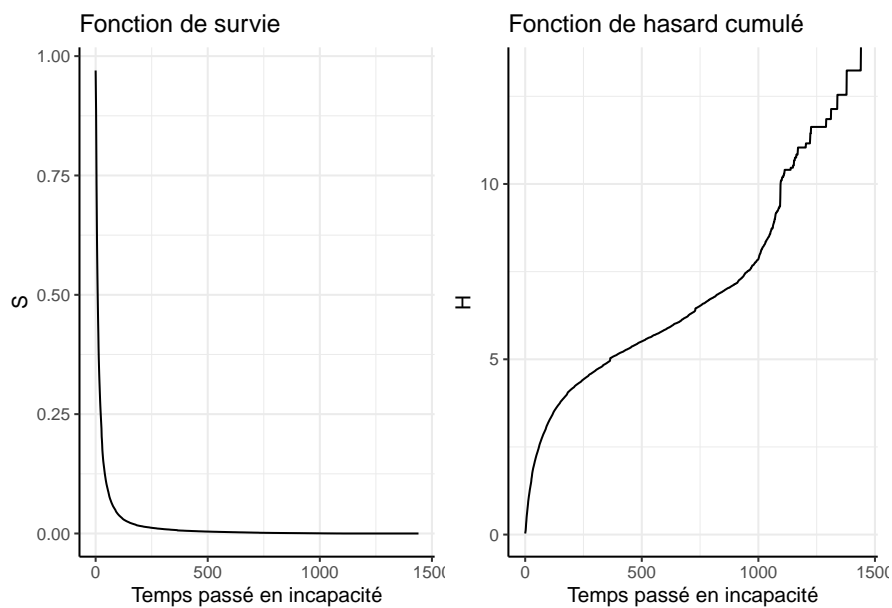


Figure 1.4: Fonction de survie et de hasard cumulée estimées non paramétriquement.

prendre en compte l'occurrence de valeur censurée / tronquée dans l'écriture des estimateurs. L'utilisation de l'estimateur (1.1) nécessiterait de retirer les valeurs censurées de l'étude. Si la variable aléatoire T est considérée comme discrètes, on peut estimer les taux de sortie de l'état d'incapacité par

$$\widehat{h}(t) = \frac{\sum_{i=1}^n \mathbb{I}_{t_i=t}}{\sum_{i=1}^n \mathbb{I}_{t_i \geq t}}.$$

La série des taux de sortie est donnée sur la Figure 1.5. L'aspect un peu erratique de ces taux de

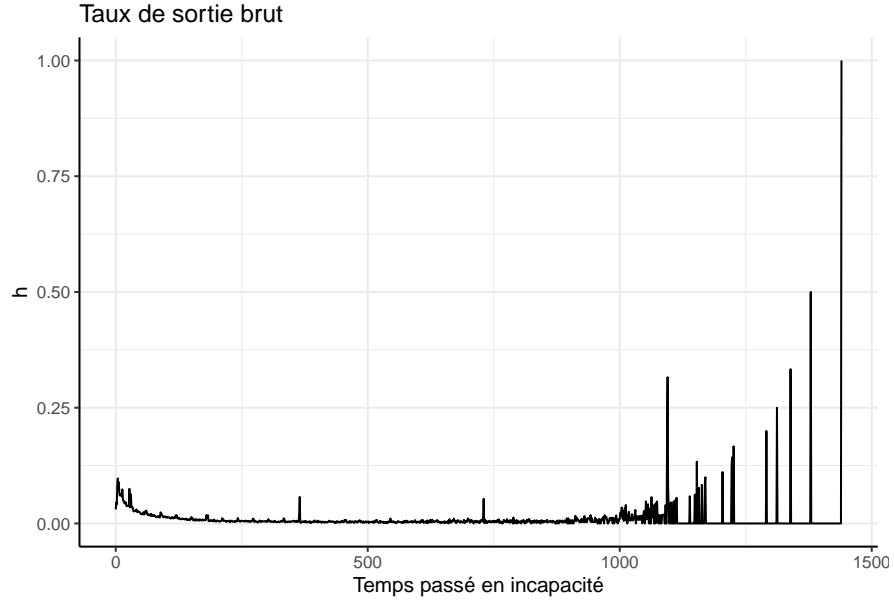


Figure 1.5: Fonction de hasard estimée non paramétriquement via l'estimateur empirique.

sortie "brut" est souvent corrigé par l'application d'une méthode de lissage.

1.3.2 L'approche paramétrique

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 $\widehat{\theta}$ permettant le meilleur ajustement du modèle aux données. N'importe quelle loi de probabilité sur \mathbb{R}_+ ou sur \mathbb{N} peut convenir. Pour l'exemple, prenons la loi exponentielle $\text{Exp}(1/\mu)$ (telle que $\mathbb{E}[\text{Exp}(\mu)] = \mu$) et la loi de poisson $\text{Pois}(\lambda)$. L'estimation est très facile pour ces lois comme

$$\widehat{\mu} = \widehat{\lambda} = \frac{1}{n} \sum_{i=1}^n t_i.$$

Nous pouvons vérifier l'adéquation en superposant les graphiques des fonctions théoriques et empiriques comme sur la Figure 1.6. L'adéquation n'est pas optimale, nous étudierons des modèles paramétrique plus adaptés et nous verrons également des modèles dit semi-paramétriques permettant un arbitrage entre des hypothèses fortes (paramétrique) et une grande flexibilité (modèle non paramétrique). L'intérêt de ces modèles réside aussi dans l'ajout de covariables qui est discuté ci-après.

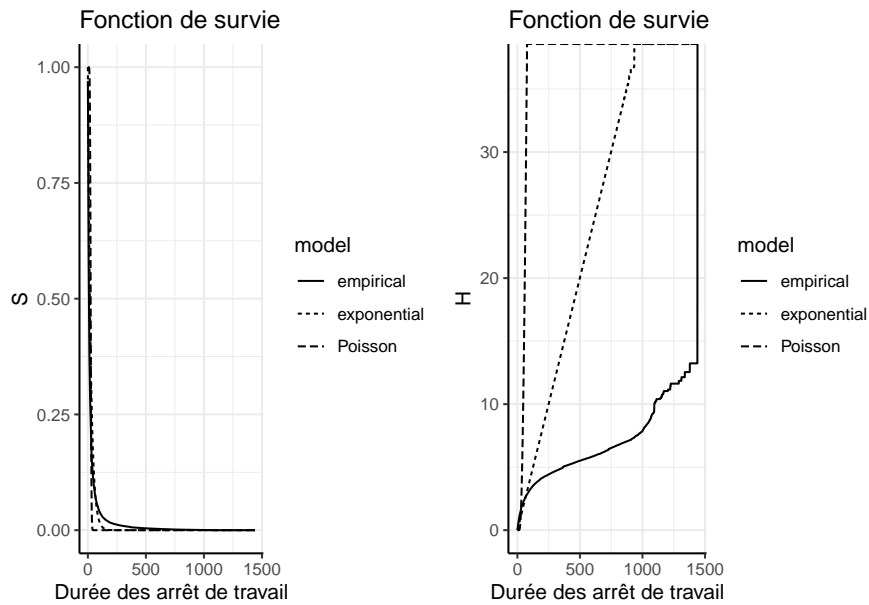


Figure 1.6: Fonction de survie et de hasard cumulée estimées non paramétriquement.

1.4 Ajout de facteurs explicatifs

La survie des individus ou des composants peut s'expliquer par les caractéristiques des individus ou des circonstances exogènes qui peuvent être connu grâce à des covariables ou variables explicatives, notées X . La variable T est la variable à expliquer et on s'intéresse à la loi conditionnelle de T sachant X . Dans notre exemple, nous nous servons de la variable `TypeArret` qui indique le motif de l'entrée en incapacité. Le tableau [Table 1.3](#) reporte les statistiques descriptives de la durée des arrêts de travail en fonction du traitement reçu. Les distributions semblent

TypeArret	n	mean	sd	q25	q50	q75	min	max
Accident du travail	64295	26	60	5	12	25	0	1379
Maladie	496107	25	62	4	9	23	0	1440
Maternité	323	52	52	10	29	98	0	237

Table 1.3: Statistiques descriptives de la durée de l'incapacité en fonction du motif d'entrée en incapacité

similaires pour les incapacités liées à une maladie ou un accident du travail, nous laissons de côté les entrées en incapacité pour congé de maternité. La [Figure 1.7](#) permet de comparer les estimations de la fonction de survie et de hasard cumulés. Nous verrons dans ce cours comment prendre en compte de manière appropriées (ou de manière plus sophistiquée) les covariables.

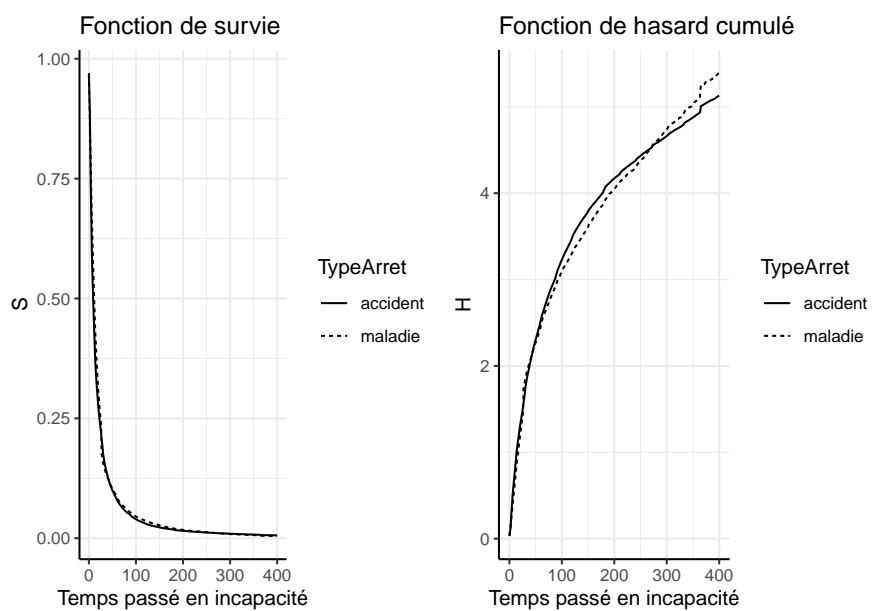


Figure 1.7: Fonction de survie et de hasard cumulée pour chacune des cohortes.

Chapter 2

Approche paramétrique

Soit t_1, \dots, t_n un échantillon de réalisation 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 caractérisé 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'optimisation numérique¹.

2.1 Fonction de vraisemblance en présence de censure et de troncation

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

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

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

Example 2. 1. L'étude porte sur l'âge au quel les enfant apprenne une compétence. Lorsque l'étude débute, certains enfant d'âge C maitrise déjà la compétence, pour eux $T \leq C$ (censure à gauche). A la fin de l'étude d'autre enfant d'âge C ne maitrise pas encore la compétence alors $T \geq C$ (censure à droite)

¹ Les fonctions `nlm` et `optim` en R

2. L'étude porte sur la durée de survie T de patients atteints d'une certaine maladie. Pour les patients perdus de vue au bout du temps C alors qu'ils étaient encore vivants, C censure T à droite puisque, pour eux, T est inconnue mais supérieure à C .
3. L'étude porte sur la durée de vie après la retraite de sujets qui entrent dans l'enquête à la suite d'un tirage au sort dans une caisse de retraite. Un sujet n'est donc observé que si sa durée de vie après la retraite excède le délai entre sa prise de retraite et l'instant de l'enquête. La durée de vie après la retraite est donc tronquée à gauche par ce délai.

Pour plus de détails, on pourra consulter le rapport suivant http://www.numdam.org/item/JSFS_1994__135_4_3_0.pdf. 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.

2.1.1 Troncature

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 h et S par

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

et

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

2.2 Lois paramétriques usuelles

2.2.1 Loi exponentielle

Definition 2. 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 $\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} dt$ désigne la fonction gamma. La fonction de survie est donnée par

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

où $\Gamma_u(t, \alpha) = \int_t^\infty e^{-t} t^{\alpha-1} dt$ 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 modèles puis à choisir le plus adapté grâce à des critères d'information.

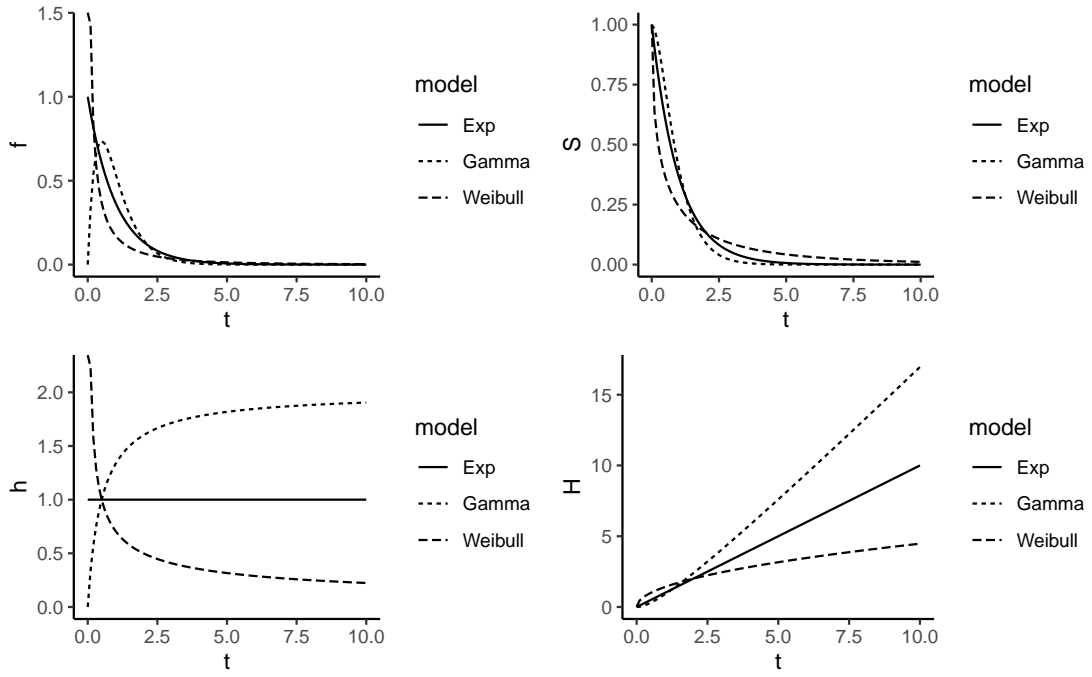


Figure 2.1: Fonction de densité, de survie, de hasard et de hasard cumulée des modèles $\text{Exp}(\beta = 1)$, $\text{Gamma}(\alpha = 2, \beta = 1/2)$ et $\text{Weib}(\alpha = 1/2, \beta = 1/2)$.

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 \[1998\]](#). 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 statitique. 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 définie 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

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

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. Un estimateur de la fonction de survie qui tient compte de la censure est le suivant

$$\widehat{S}_n(t) = \frac{\sum_{k=1}^n \mathbb{I}_{x_k > t}}{\sum_{k=1}^n \mathbb{I}_{x_k > t} + \sum_{k=1}^n \mathbb{I}_{x_k \leq t} \delta_k}, \quad (3.2)$$

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

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

Remark 3. Si T est un v.a. continu alors on peut 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

$$\widehat{S}_n^{KP}(t) = \prod_{k: v_k \leq t} \left[1 - \frac{d(v_k)}{n(v_k)} \right], \quad (3.3)$$

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

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

On peut ré-écrire la vraisemblance

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

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.

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

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

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

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 $\widehat{\theta}_n$ l'estimateur du maximum de vraisemblance.

Theorem 2. sous certaines conditions de régularité, on a

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

où

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

est la matrice d'information de Fisher.

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

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

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

Proposition 1. Soit $\widehat{\theta}_n$ l'estimateur du maximum de vraisemblance de θ et g une fonction dérivable alors

$$g(\widehat{\theta}_n) \sim \text{Normal}(g(\theta), {}^t g'(\theta_n) I_n(\widehat{\theta}_n)^{-1} g'(\theta_n)).$$

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

$$\frac{\partial^2}{\partial h_k \partial h_l} l(\mathcal{D}; \theta) = \begin{cases} 0, & \text{si } l \neq k \\ -\frac{d(v_k)}{h_k^2} - \frac{n(v_k) - d(v_k)}{(1-h_k)^2}, & \text{sinon.} \end{cases} \quad (3.4)$$

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

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

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

$$\ln \widehat{S}_n(t) \sim \text{Normal}[0, \mathbb{V} \ln \widehat{S}_n(t)]$$

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

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

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

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

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

Définition 4. Une urne contient N boules dont K boules blanches et $N - K$ boules noires. On réalise h tirage 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

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

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 $O_{j,k} = d_i(v_k)$ à la valeur attendue $E_{j,k} = n_i(v_k) \frac{d(v_k)}{n(v_k)}$ sous H_0 avec

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

où

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

La convergence (3.5) est une application du théorème centrale limite, d'autant plus valide que $\#\{k : d(v_k) \geq 1\}$ est grand. Une meilleure approximation de la distribution asymptotique de la statistique de test a été obtenu par [Peto and Peto \[1972\]](#).

Chapter 4

Modèles à hasard proportionnel

Nous étudions des données t_1, \dots, t_n censurés à droite

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

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

$$h(t) = \begin{cases} h_0(t), & \text{si } Z_0 = 0 \\ h_0(t)e^\beta, & \text{sinon.} \end{cases}$$

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 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 $\mathcal{D} = (x_k, \delta_k)_{k=1, \dots, n}$ s'écrit

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

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 covariables 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 continue par morceaux, voir Breslow [1975].

Soit

$$0 = v_0 < v_1 < v_2 < \dots < v_K < v_{K+1} = \infty,$$

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

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

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

$$\begin{aligned} l(\mathcal{D}; \theta) &= \sum_{i=1}^n \delta_i \{ \log[h_0(x_i)] + z_i \beta \} - e^{z_i \beta} \int_0^{x_i} h_0(s) ds \\ &= \sum_{i=1}^n \delta_i \sum_{k=1}^K [\log(\alpha_k) + \delta_i z_i \beta] \mathbb{I}_{x_i = v_k} - e^{z_i \beta} \sum_{k=1}^K \alpha_k \int_0^{x_i} \mathbb{I}_{(v_{k-1}, v_k]}(s) ds \\ &= \sum_{k=1}^K \log(\alpha_k) d(v_k) + s_k \beta - \alpha_k \sum_{i=1}^n \int_0^{x_i} e^{z_i \beta} \mathbb{I}_{(v_{k-1}, v_k]}(s) ds \\ &= \sum_{k=1}^K \log(\alpha_k) d(v_k) + s_k \beta - \alpha_k \left\{ \sum_{i=1}^n e^{z_i \beta} \left[(v_k - v_{k-1}) \mathbb{I}_{x_i \geq v_k} + (x_i - v_{k-1}) \mathbb{I}_{(v_{k-1}, v_k]}(x_i) \right] \right\} \\ &= \sum_{k=1}^K \log(\alpha_k) d(v_k) + s_k \beta - \alpha_k E_k(\beta), \end{aligned}$$

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

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

où

$$E_k(\beta) = \sum_{i=1}^n e^{z_i \beta} \left[(v_k - v_{k-1}) \mathbb{I}_{x_i \geq v_k} + (x_i - v_{k-1}) \mathbb{I}_{(v_{k-1}, v_k]}(x_i) \right],$$

est l'exposition.

Remark 7. Dans le papier original de [Breslow \[1975\]](#), les données censurées entre v_{k-1} et v_k sont remplacées par v_{k-1} (l'observation non censurée la plus proche). Cela simplifie l'exposition qui devient

$$E_k(\beta) = \sum_{i=1}^n e^{z_i \beta} (v_k - v_{k-1}) \mathbb{I}_{x_i \geq v_k}.$$

On recherche β qui annule

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

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

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

La fonction de hasard cumulée de base est estimée par

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

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 θ dont l'estimateur $\widehat{\theta}$ est asymptotiquement normal. On peut effectuer un test de Wald pour tester sa nullité

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

La statistique de test est donnée par

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

où \widehat{se} est une estimation de l'écart-type type de $\widehat{\theta}$. Ici la racine carrée 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 teste des hypothèses du type

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

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

$$\lambda \sim \chi_q^2.$$

La significativité du modèle de Cox teste l'hypothèse de nullité simultanée de tous 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 proportionnalité

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

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

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 observation t_i par les observations censurées à droite x_i . Soient $v_1 < \dots < v_K$ les temps d'occurrence des événements distincts et non censurés.

On peut comparer les caractéristiques moyenne observée et attendu des évènement s'étant produit 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

$$\mathbb{E}(r_k^*) = \gamma g(v_k).$$

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 $g(v_k)$. 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 reposent sur l'utilisation de tables de mortalités. 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 viagères 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 auditées par un actuinaire 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 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 1.1](#) 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 8. 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

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

qui permet de construire des intervalle de confiance.

Modèle de Poisson: estimation des μ_x

Le temps de survie résiduelle T_x à l'âge $x \in \mathbb{R}_+$ (Approche dite continue) suit une loi exponentielle $\text{Exp}(1/\mu_x)$. Cela implique que

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

Il est usuel de supposer que $s \mapsto \mu_s$ constant entre deux âge x et $x+1$, c'est à dire

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

Ce modèle revient à supposer que le nombre de décès D_x à l'âge x suit une loi de Poisson $D_x \sim \text{Poisson}(E_x^c \mu_x)$ où E_x^c est l'exposition centrale (par opposition à l'exposition initiale l_x). 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 9. 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 taux de mortalité sont estimés par

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

via le maximum de vraisemblance. On en déduit que

$$\widehat{q}_x = 1 - e^{-d_x / E_x^c}.$$

L'intervalle de confiance pour les \widehat{q}_x est obtenu par la méthode Delta avec

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

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 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. Cela permet aussi de réduire

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](#).

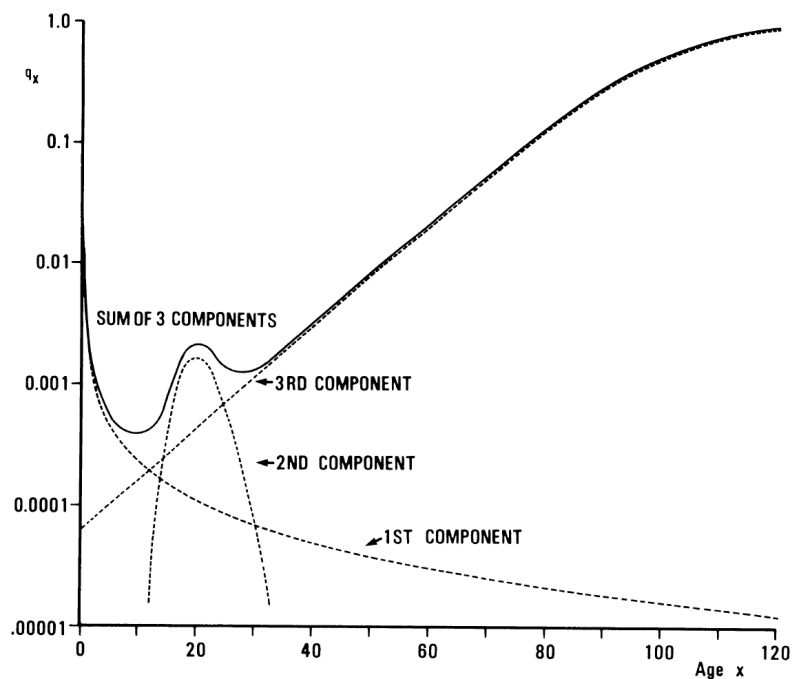


Figure 1. The graduated q_x curve and its three components: Australian national mortality, 1970-72 (males).

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

La calibration de ces modèles se fait de deux façons

- 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

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

Remark 10. Pour le lissage des probabilités de décès, on peut par exemple choisir les poids inversement proportionnel de la variance des taux bruts 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 non paramétrique

Il est aussi possible de lisser la série des taux bruts sans faire d'hypothèse paramétrique.

Les 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

$$q_x^h = \sum_{k=-h}^h w_x^k \widehat{q}_x,$$

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 lissés 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 .

Le lissage de Whittaker-Henderson

Le lissage de Whittaker-Henderson, expliqué par exemple dans le papier de [Weinert \[2007\]](#), consiste en un arbitrage entre la fidélité au taux brut \widehat{q}_x et la régularité des taux lissés \widehat{q}_x^λ . Les taux lissés sont solution du problème d'optimisation

$$\arg \max_{\widehat{q}_x^\lambda} \sum_{x=0}^{x^*} F(\widehat{q}_x, q_x^\lambda) + \lambda R(q_x^\lambda),$$

où $x^* = \sup\{x \in \mathbb{N} ; q_x > 0\}$ et $\lambda > 0$ qui traduit le compromis entre fidélité F et régularité R . Le critère de fidélité est donnée par

$$F(\widehat{q}_x, q_x^\lambda) = \sum_{x=0}^{x^*} w_x (\widehat{q}_x - q_x^\lambda)^2.$$

le critère de régularité est donnée par

$$R(q_x) = \sum_{x=0}^{x^*-z} (\Delta^z q_x^\lambda)^2,$$

où $\Delta^z q_x^\lambda$ est l'opérateur de différence progressive d'ordre z défini par

$$\Delta^z q_x^\lambda = \sum_{j=0}^z \binom{z}{j} (-1)^{z-j} q_{x+j}^\lambda.$$

On note que

$$\Delta^1 q_x^\lambda = q_{x+1}^\lambda - q_x^\lambda, \Delta^2 q_x^\lambda = \Delta^1 \Delta^1 q_x^\lambda = q_x^\lambda q_{x+2}^\lambda - 2q_{x+1}^\lambda + q_x^\lambda, \dots$$

Remark 11. 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.

L'implémentation passe par une représentation matricielle des différentes quantités. On a

$$W = \text{diag} \begin{pmatrix} w_1 & \dots & w_{x^*} \end{pmatrix}, \widehat{q} = \begin{pmatrix} \widehat{q}_1 & \dots & \widehat{q}_{x^*} \end{pmatrix}, \text{ et } q^\lambda = \begin{pmatrix} q_1^\lambda & \dots & q_{x^*}^\lambda \end{pmatrix}.$$

Pour $z = 2$, on définit

$$K = \begin{pmatrix} -1 & 2 & -1 & 0 & \dots & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & -1 & 2 & -1 \end{pmatrix}.$$

le problème d'optimisation devient

$$\arg \min_{q^\lambda} \mathcal{L}(q^\lambda) = \arg \min_{q^\lambda} {}^t(\widehat{q} - q^\lambda)W(\widehat{q} - q^\lambda) + \lambda {}^t(Kq^\lambda)Kq^\lambda.$$

On en déduit que

$$\frac{d}{dq^\lambda} \mathcal{L}(q^\lambda) = 2Wq^\lambda - 2W\widehat{q} + 2\lambda {}^tKKq^\lambda$$

Le système d'équation

$$\frac{d}{dq^\lambda} \mathcal{L}(q^\lambda) = 0$$

a pour solution

$$q^\lambda = (W + \lambda {}^tKK)^{-1} W\widehat{q}.$$

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 12.** 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 SOA².

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 $a = t - x$.

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

²<https://www.soa.org/globalassets/assets/files/resources/essays-monographs/2005-living-to-100/m-li05-1-ix.pdf>

On peut construire autant de table de mortalité "statique" que d'année calendaire ou de cohortes disponibles. Les tables générationnelles TGH et TGF 05 comprennent des projections permettant la traification 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. On peut estimer les probabilités de décès par période $t_0 \in [t, t + 1]$ avec

$$q_{x,t_0} = \frac{d_{x,t_0}}{E_{x,t_0}},$$

par génération $a_0 \in [t - x, t - x + 1]$ avec

$$q_{x,a_0} = \frac{d_{x,a_0}}{E_{x,a_0}},$$

ou par période et par cohorte avec

$$q_{x,t_0,a_0} = \frac{d_{x,t_0,a_0}}{E_{x,t_0,a_0}},$$

suivant le type de données disponibles. 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 = année calendaire ,
- Ordonnée: x = âge,

voir la [Figure 5.3](#). Lorsque les données sont exhaustives on peut représenter précisément le parcours de chaque individu sur le diagramme. l'estimation des taux de mortalité se résume à compter le nombre de décès (●) dans les triangles, les carrés ou les parallélogramme du diagramme de Lexis suivant que l'on calcule des taux de décès pour

- une période $t_0 \in [t, t + 1]$: carré de Lexis
- cohorte $a_0 \in [t - x, t - x + 1]$: parallélogramme de Lexis

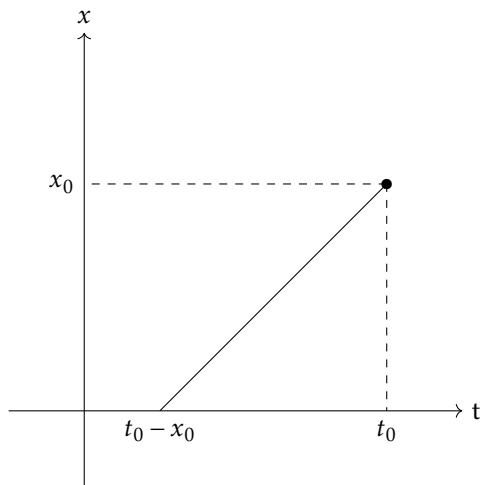


Figure 5.2: Positionnement d'un individu qui décède à l'âge x_0 durant l'année t_0 sur le diagramme de Lexis

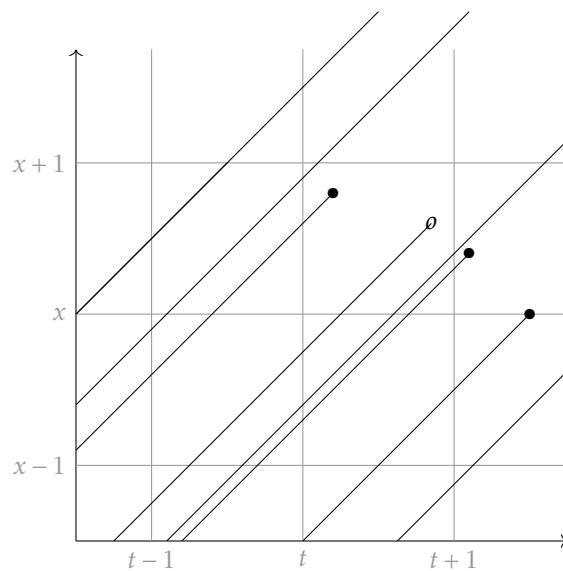


Figure 5.3: Parcours des individus sur le diagramme de Lexis. • = décès et o = fin d'observation de l'individu (censure à droite)

- période et cohorte $(t_0, a_0) \in [t, t+1] \times [t-x, t-x+1]$: triangle de Lexis

puis à rapporter ce nombre sur l'exposition au risque. Lorsque l'on connaît la date exacte des décès, l'exposition est le cumul des longueurs des segments des parcours passant au travers du carré, triangle ou parallélogramme étudié, voir la Figure 5.4. Les individus de la cohorte nés

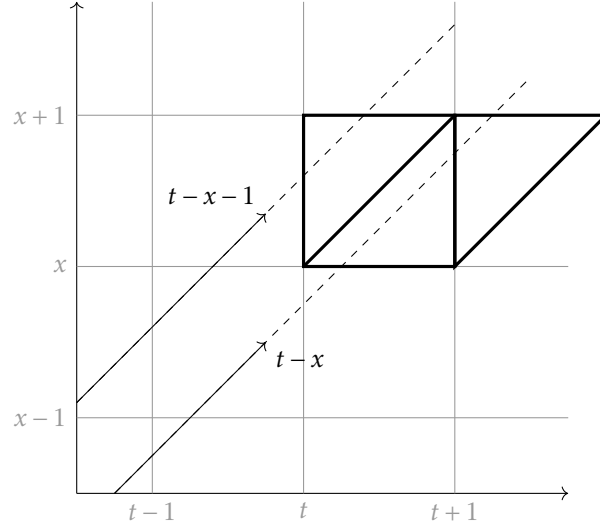


Figure 5.4: Géométrie de Lexis

l'année $t-x$ atteindrons l'âge x durant les années calendaires t et $t+1$.

Remark 13. 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 HMD³ voir Wilmoth et al. [2007].

5.3.2 Taux de mortalité par période

Considérons les événements se produisant sur la période $[t, t+1]$, nous intéressons au carré de Lexis de la Figure 5.5

- $P_{x,t} = 2$ est le nombre d'individu d'âge x au début de la période t . Il s'agit du nombre de segment intersectant le côté $(x, t) - (t, x+1)$.
- $D_{x,t}^U = 1$ 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 atteignant l'âge x durant la période $[t, t+1]$.
- Le nombre de décès au cours de la période est donné par

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

³<https://www.mortality.org/>

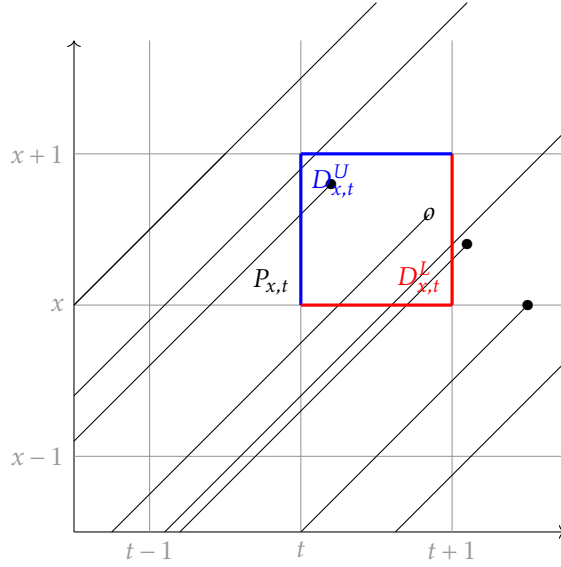


Figure 5.5: Parcours des individus sur le diagramme de Lexis. ● = décès et o = fin d'observation de l'individu (censure à droite)

L'exposition exacte est donnée par la somme des longueurs des segments dans le carré de Lexis, ici

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

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

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

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.

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

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 5.6. Soit $B_1, B_2 \in [0, 1]$ les v.a. qui indiquent la 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

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

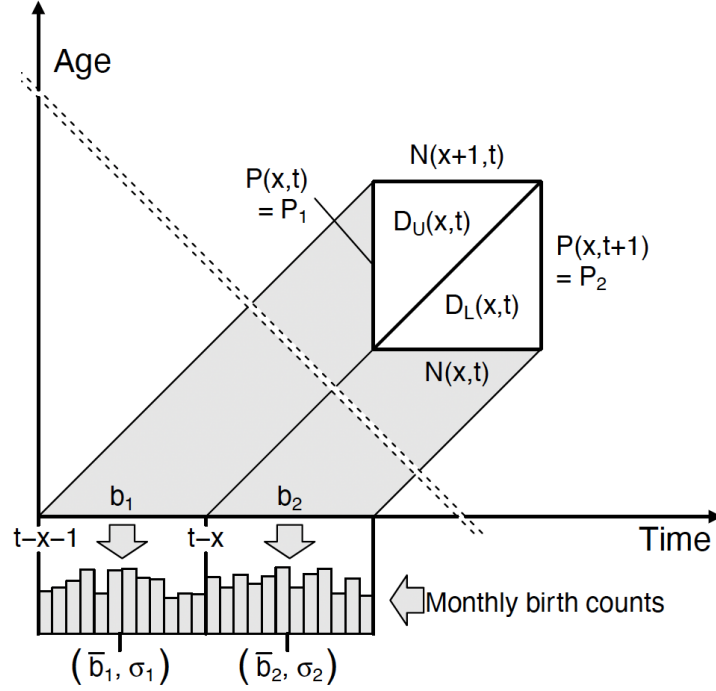


Figure 5.6: Calcul d'exposition central sur un diagramme de Lexis grâce à la distribution des naissances pour les cohortes concernées. (Source: HMD Documentation)

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

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

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

$$\mathbb{E}(B_1) = \bar{b}_1. \quad (5.2)$$

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

$$f_{(U,V)}(u, v) = C_L f_{B_2}(u - v) \mathbb{I}_{0 \leq v \leq u \leq 1}(u, v), \quad (5.3)$$

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 (5.3), 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

$$\int_0^1 \int_0^u (1 - u) f_{B_2}(u - v) dv du = \frac{1 - \bar{b}_2}{2} + \frac{\sigma_2^2}{2(1 - \bar{b}_2)} \quad (5.4)$$

On déduit de (5.1) et (5.4) l'exposition dans le triangle inférieur avec

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

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 $(t + U, x + V)$, la perte d'exposition est donnée par

$$\int_0^1 \int_u^1 u f_{B_2}(u + 1 - v) dv du = \frac{\bar{b}_1}{2} + \frac{\sigma_2^2}{2\bar{b}_1} \quad (5.6)$$

On déduit de (5.2) 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

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

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

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

Remark 14. 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 natatilité, voir le travail de Boumezoued [2020].

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 5.7 L'exposition est donnée par la somme des longueurs des segments dans le parallélogramme, soit

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

La probabilité de décès des individus d'âge x né l'année a est donnée par

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

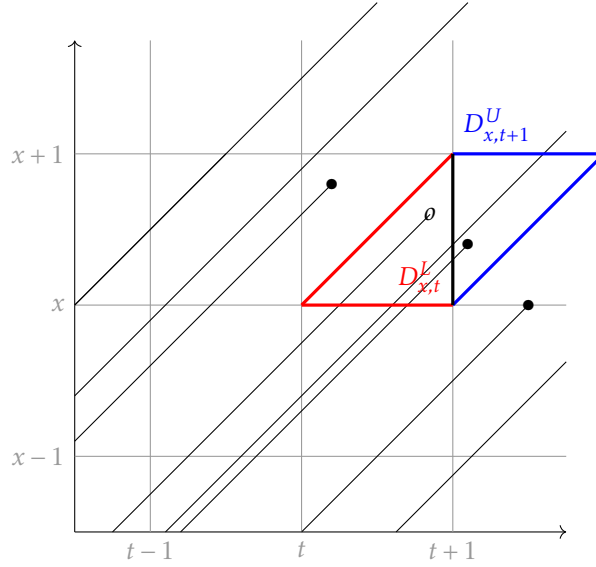


Figure 5.7: Parcours des individus sur le diagramme de Lexis. ● = décès et o = fin d'observation de l'individu (censure à droite)

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 5.8. 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 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

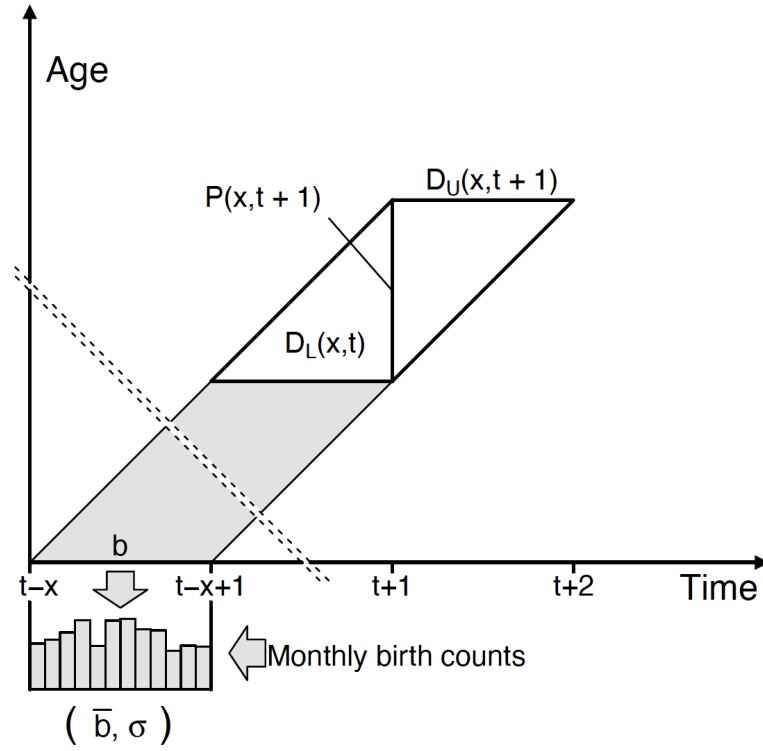


Figure 5.8: Calcul d'exposition central sur un diagramme de Lexis grâce à la distribution des naissances pour la cohortes concernée. (Source: HMD Documentation)

taux de mortalité sont liés par

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

Les modèles de mortalité stochastique, 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^4 . Pour le modèle binomial la fonction logit^5 . 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. On applique au vecteur de paramètres

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

une transformation v avec

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

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

Exemple 4. Voici quelques exemples de modèles de type GAPC.

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

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

⁴ $\log(\mu_{x,t})$

⁵ $\log(q_{x,t}/(1 - q_{x,t}))$

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 Schifflers \[1987\]](#), 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 modélisés 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 vraisemblance s'écrit

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

dans le modèle de Poisson et

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

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 critères 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 procédure 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'expérience plutôt que réglementaire
- les probabilités de décès attendues Vs les probabilités de décès observées.

Remark 15. Il est aussi possible de réaliser une inférence Bayésienne via le package StanMoMo décrit dans [Barigou et al. \[2022\]](#).

Bibliography

- Odd Aalen. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6(4), jul 1978. doi: 10.1214/aos/1176344247.
- Hiroto 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.
- 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.
- 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.
- Howard L. Weinert. Efficient computation for whittaker-henderson smoothing. *Computational Statistics & Data Analysis*, 52(2):959–974, oct 2007. doi: 10.1016/j.csda.2006.11.038.
- John R Wilmoth, Kirill Andreev, Dmitri Jdanov, Dana A Gleis, 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.