$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \mathbb{1}_{]t;\infty[}(T_x)$$

GRANDE ÉCOLE D'ACTUARIAT ET DE GESTION DES RISQUES
INSTITUT DE SCIENCE FINANCIÈRE ET D'ASSURANCES

MODÈLES DE DURÉE

Support de cours 2025-2026

Statistique des modèles paramétriques et semi-paramétriques

Frédéric PLANCHET

Version 5.6

Septembre 2025

SOMMAIRE

1. La prise en compte de censure dans les modèles de durée

L'objet de cette section est de déterminer la forme générale de la vraisemblance d'un modèle de durée censuré en fonction du type de censure et d'illustrer, dans le cas de la distribution exponentielle, l'impact sur la vraisemblance des phénomènes de censure[1].

En pratique on peut être confronté à une censure droite (si *X* est la variable d'intérêt, l'observation de la censure C indique que $X \geq C$) ou à une censure à gauche (l'observation de la censure C indique que $X \leq C$) ; les deux types de censure peuvent s'observer de manière concomitante. L'exemple classique est donné par la situation suivante : on veut savoir à quel âge *X* les enfants d'un groupe donné sont capables d'effectuer une certaine tâche. Lorsque l'expérience débute, certains enfants d'âge C sont déjà capables de l'accomplir, et pour eux $X \leq C$: il s'agit d'une censure gauche ; à la fin de l'expérience, certains enfants ne sont pas encore capables d'accomplir la tâche en question, et pour eux $X \geq C$: il s'agit d'une censure droite.

Dans la suite on s'intéressera à la censure droite, courante dans les situations d'assurance.

1.1. Censure de type I : censure fixe

Soit un échantillon de durées de survie $(X_1,\ldots,X_n)$ et $C > 0$ fixé; la vraisemblance du modèle associé aux observations $(T_1,D_1),\ldots,(T_n,D_n)$ avec :

$$T_i = X_i \wedge C \text{ et } D_i = \begin{cases} 1 \; si \; X_i \leq C \\ 0 \; si \; X_i > C \end{cases}$$

possède une composante continue et une composante discrète ; elle s'écrit :

$$L(\theta) = \prod_{i=1}^{n} f_\theta(T_i)^{D_i} \, S_\theta(C)^{1-D_i}$$

en d'autres termes lorsqu'on a observé la sortie avant la censure, c'est le terme de densité qui intervient dans la vraisemblance, et dans le cas contraire on retrouve le terme discret, avec comme valeur la fonction de survie à la date de censure. La distribution est donc continue par rapport à T_i et discrète par rapport à D_i.

Pour démontrer cette formule, il suffit de calculer $P\left(T_i \in \left[t_i, t_i + dt_i\right], D_i = d_i\right)$. Comme D_i ne peut prendre que les valeurs 0 et 1, on calcule, sur $[0, C]$:

$$\begin{aligned} P\left(T_i \in \left[t_i, t_i + dt_i\right], D_i = 1\right) &= P\left(X_i \wedge C \in \left[t_i, t_i + dt_i\right], X_i \leq C\right) \\ &= P\left(X_i \in \left[t_i, t_i + dt_i\right]\right) = f_\theta(t_i)\, dt_i \end{aligned}$$

(on peut toujours supposer dt_i suffisamment petit pour que $t_i + dt_i \leq C$) et

[1] Et, marginalement, de troncature, qui seront mentionnés pour mémoire mais pas développés.

$$P\left(T_i \in \left[t_i, t_i + dt_i\right], D_i = 0\right) = P\left(X_i \wedge C \in \left[t_i, t_i + dt_i\right], X_i \geq C\right) = P\left(X_i \geq C\right) = S_\theta\left(C\right)$$

Ces deux cas peuvent se résumer en :

$$P\left(T_i \in \left[t_i, t_i + dt_i\right], D_i = d_i\right) = f_\theta\left(t_i\right)^{d_i} S_\theta\left(C\right)^{1-d_i}$$

On peut retrouver cette expression également en observant que :

$$P\left(T_i > t_i, D_i = 1\right) = P\left(X_i > t_i, X_i \leq C\right) = \int_{t_i}^{C} f_\theta\left(u\right) du$$

et dans le cas où $D_i = 0$ comme alors $T_i = C$ il n'y a pas de densité, mais simplement la probabilité de cet événement est égale à $S_\theta\left(C\right)$.

Comme pour une observation censurée, par définition, $T_i = C$, l'expression ci-dessus peut se réécrire :

$$L\left(\theta\right) = \prod_{i=1}^{n} f_\theta\left(T_i\right)^{D_i} S_\theta\left(T_i\right)^{1-D_i}$$

En se souvenant que la densité peut s'écrire en fonction de la fonction de hasard et de la fonction de survie $f_\theta\left(t\right) = h_\theta\left(t\right) S_\theta\left(t\right)$ on peut également écrire la vraisemblance sous la forme (à une constante multiplicative près) :

$$L\left(\theta\right) = \prod_{i=1}^{n} S_\theta\left(T_i\right) h_\theta\left(T_i\right)^{D_i}$$

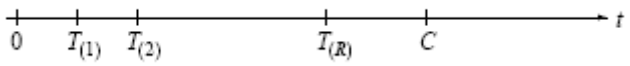
Cette expression est donc simplement le produit des valeurs de la fonction de survie (qui traduit le fait que les individus sont observés au moins jusqu'en T_i), pondérée pour les sorties non censurées par la valeur de la fonction de hasard (qui traduit le fait que pour ces observations la sortie a effectivement lieu à l'instant T_i). On utilise en général la log-vraisemblance, égale, à une constante additive près, à :

$$\mathbf{ln}\, L\left(\theta\right) = \sum_{i=1}^{n} \left[D_i \,\mathbf{ln}\left(h_\theta\left(T_i\right)\right) + \mathbf{ln}\left(S_\theta\left(T_i\right)\right)\right].$$

À titre d'illustration, on détaille ci-après les cas de l'estimation ponctuelle et de l'estimation par intervalle dans l'exemple de la loi exponentielle.

1.1.1. Estimation ponctuelle

On considère donc maintenant le cas où la distribution sous-jacente est exponentielle, de paramètre θ ; on pose $R = \sum_{i=1}^{n} D_i$ le nombre de décès observés :



Comme $f_\theta(t) = \theta e^{-\theta t}$, la vraisemblance s'écrit donc $L(\theta) = \prod_{i=1}^{n} \left(\theta e^{-\theta T_i}\right)^{D_i} \left(e^{-\theta C}\right)^{1-D_i}$, ce qui devient :

$$L(\theta) = \theta^R \exp\left(-\theta \sum_{i=1}^{n} T_i\right)$$

On peut incidemment remarquer que la loi de *R* est discrète, et est une loi binomiale de paramètres $\left(n, 1-e^{-\theta C}\right)$: le nombre de sorties non censurées correspond à un tirage dans *n* valeurs, la probabilité de succès étant égale à $1-e^{-\theta C} = P_\theta(T \le C)$.

Si $T = \sum_{i=1}^{n} T_i$ désigne l'« exposition au risque » totale[2], on a ici $T = \sum_{i=1}^{R} T_{(i)} + (n-R)C$; en annulant la dérivée première de la log-vraisemblance $l(\theta) = R\ln(\theta) - \theta\left(\sum_{i=1}^{R} T_{(i)} + (n-R)C\right)$ par rapport à θ, on trouve que l'estimateur du maximum de vraisemblance (EMV) de θ est $\hat{\theta} = \frac{R}{T}$. La statistique exhaustive est donc bi-dimensionnelle, (T, R).

L'estimateur de θ est donc le rapport du nombre de décès observés à l'exposition au risque ; dans un modèle non censuré (obtenu comme cas limite du modèle censuré lorsque $C \to +\infty$), l'expression $\hat{\theta} = \frac{R}{T}$ devient $\hat{\theta} = \frac{1}{\bar{X}}$; en effet, on observe alors tous les décès, et l'estimateur est le classique « inverse de la moyenne empirique des durées de vie ».

1.1.2. Estimation par intervalle

On peut utiliser l'efficacité asymptotique de l'estimateur du maximum de vraisemblance pour déterminer un intervalle de confiance pour l'estimateur. Dans le cas de la loi exponentielle on peut également remarquer que, si $m_C(\theta)$ et $\sigma_C(\theta)$ désignent l'espérance et l'écart-type de *T*, alors par le théorème central-limite on a $\sqrt{n}\frac{T - m_C(\theta)}{\sigma_C(\theta)}$ qui converge en loi vers une loi normale centrée réduite. En effet, les variables aléatoires

[2] *T* est parfois appelé le « temps global de fonctionnement au cours des essais ».

$T_i = X_i \wedge C$ sont iid, puisque les X_i le sont. Les expressions de $m_C(\theta)$ et $\sigma_C(\theta)$ peuvent être obtenues par quelques calculs :

- ✓ $m_C(\theta) = \int_0^C u\theta e^{-\theta u} du + Ce^{-\theta C} = \frac{1-e^{-\theta C}}{\theta}$
- ✓ $\sigma_C^2(\theta) = E\left(T_i^2\right) - \left(m_C(\theta)\right)^2 = \frac{1}{\theta^2}\left(1 - 2\theta C e^{-\theta C} - e^{-2\theta C}\right)$

Dans l'hypothèse où la durée de l'expérience C est petite devant la durée de vie *a priori* de chaque individu $\frac{1}{\theta}$, on a θC qui est petit devant 1 et on peut donc faire un développement limité des exponentielles à l'ordre 3 en θC, qui conduit à : $\sigma_C^2(\theta) = \frac{\theta C^3}{3}$. On obtient ainsi une forme relativement simple de région de confiance pour le paramètre θ.

1.2. Censure de type III : censure aléatoire[3]

1.2.1. Le cas d'un échantillon iid

La censure de type III généralise la censure de type I au cas où la date de censure est une variable aléatoire ; plus précisément, soient un échantillon de durées de survie $(X_1,\ldots,X_n)$ et un second échantillon indépendant composé de variables positives $(C_1,\ldots,C_n)$; on dit qu'il y a censure de type III pour cet échantillon si au lieu d'observer directement $(X_1,\ldots,X_n)$ on observe $(T_1,D_1),\ldots,(T_n,D_n)$ avec :

$$T_i = X_i \wedge C_i \text{ et } D_i = \begin{cases} 1 \; si \; X_i \le C_i \\ 0 \; si \; X_i > C_i \end{cases}$$

La vraisemblance de l'échantillon $(T_1,D_1),\ldots,(T_n,D_n)$ s'écrit, avec des notations évidentes :

$$L(\theta) = \prod_{i=1}^{n} \left[f_X(T_i,\theta) S_C(T_i,\theta)\right]^{D_i} \left[f_C(T_i,\theta) S_X(T_i,\theta)\right]^{1-D_i}$$

La forme de la vraisemblance ci-dessus se déduit par exemple du fait que $(T_1,\ldots,T_n)$ est un échantillon de la loi $S_T(\theta,.)$ avec :

$$S_T(\theta,t) = P_\theta(T_i > t) = P_\theta(X_i \wedge C_i > t) = P_\theta(X_i > t) P_\theta(C_i > t) = S_X(t,\theta) S_C(t,\theta).$$

On écrit comme en 1.1 que :

[3] Ces modèles peuvent s'analyser comme des modèles à 2 risques concurrents indépendants.

$$P\left(T_i \in \left[t_i, t_i + dt_i\right], D_i = 1\right) = P\left(X_i \wedge C_i \in \left[t_i, t_i + dt_i\right], X_i \leq C_i\right)$$
$$= P\left(X_i \in \left[t_i, t_i + dt_i\right], t_i < C_i\right) = f_X\left(\theta, t_i\right) S_C\left(\theta, t_i\right) dt_i$$

et

$$P\left(T_i \in \left[t_i, t_i + dt_i\right], D_i = 0\right) = P\left(X_i \wedge C_i \in \left[t_i, t_i + dt_i\right], X_i \geq C_i\right)$$
$$= P\left(C_i \in \left[t_i, t_i + dt_i\right], X_i > t_i\right) = S_X\left(\theta, t_i\right) f_C\left(\theta, t_i\right) dt_i$$

Ces expressions sont directement obtenues de celles vues en 1.1 en conditionnant par rapport à la censure, puis en intégrant par rapport à la loi de celle-ci. Plus précisément, on écrit :

$$P\left(T_i > t_i, D_i = 1\right) = P\left(X_i \wedge C_i > t_i, X_i \leq C_i\right) = P\left(t_i < X_i \leq C_i\right)$$
$$= \int_{t_i}^{+\infty} P\left(t_i < X_i \leq c\right) f_C\left(\theta, c\right) dc = \int_{t_i}^{+\infty} \left(\int_{t_i}^{c} f_X\left(\theta, x\right) dx \right) f_C\left(\theta, c\right) dc$$

puis par Fubini on inverse les intégrales pour obtenir :

$$P\left(T_i > t_i, D_i = 1\right) = \int_{t_i}^{+\infty} f_X\left(\theta, x\right) \left(\int_{x}^{+\infty} f_C\left(\theta, c\right) dc \right) dx$$
$$= \int_{t_i}^{+\infty} f_X\left(\theta, x\right) S_C\left(\theta, x\right) dx$$

et finalement $P\left(T_i \in \left[t_i, t_i + dt_i\right], D_i = 1\right) = -\frac{d}{dt_i} P\left(T_i > t_i, D_i = 1\right) = f_X\left(\theta, t_i\right) S_C\left(\theta, t_i\right) dt_i$. On fait alors l'hypothèse que la censure est non informative, c'est à dire que la loi de censure est indépendante du paramètre θ. La vraisemblance se met dans ce cas sous la forme :

$$L\left(\theta\right) = const \prod_{i=1}^{n} f_\theta\left(T_i\right)^{D_i} S_\theta\left(C_i\right)^{1-D_i}$$

Le terme *const* regroupe les informations en provenance de la loi de la censure, qui ne dépend pas du paramètre. Cette dernière expression peut s'écrire comme en 1.1 ci-dessus :

$$L\left(\theta\right) = \prod_{i=1}^{n} S_\theta\left(T_i\right) h_\theta\left(T_i\right)^{D_i}$$

On observe ici simplement le fait que la censure fixe est un cas particulier de la censure aléatoire non informative dans laquelle la loi de censure est une loi de Dirac au point C. L'expression établie dans le cas particulier de la censure fixe se généralise donc aisément.

1.2.2. La prise en compte de covariables

Lorsque le modèle comporte *p* variables explicatives (covariables) $Z = (Z_1, \ldots, Z_p)$, on fait l'hypothèse que la loi conditionnelle de *X* sachant *Z* dépend d'un paramètre θ.

L'échantillon observé devient une séquence de triplets (T_i, D_i, Z_i) ; on reprend l'hypothèse de censure non informative ; on suppose de plus que *X* et *C* sont indépendantes conditionnellement à *Z* et que *C* est non informative pour les paramètres de la loi conditionnelle de *X* sachant *Z*. On suppose enfin que *Z* admet une densité qui dépend d'un paramètre ϕ, $f_Z(z, \phi)$.

Dans ces conditions, l'expression de la vraisemblance vue en 1.2 ci-dessus devient :

$$L(\theta) = \prod_{i=1}^{n} h_{\theta|Z}(T_i)^{D_i} S_{\theta|Z}(T_i) f_Z(Z_i, \phi)$$

Lorsque la loi de *T* sachant *Z* et la loi de *Z* n'ont pas de paramètre en commun, on retrouve simplement l'expression de 1.2, dans laquelle la loi de *X* est remplacée par la loi conditionnelle de *X* sachant *Z*. Ce raisonnement se généralise sans difficulté au cas de covariables dépendant du temps.

1.3. Un autre type de censure : « arrêt au $r^{ième}$ décès » (censure de type II)

On se place maintenant dans le cas où la date de fin d'observation n'est pas définie à l'avance, mais où l'on convient d'arrêter l'observation lors de la survenance de la $r^{ième}$ sortie. La date de fin de l'expérience est donc aléatoire et est égale à $X_{(r)}$.

De manière plus formelle, soit un échantillon de durées de survie $(X_1, \ldots, X_n)$ et $r > 0$ fixé; on dit qu'il y a censure de type II pour cet échantillon si au lieu d'observer directement $(X_1, \ldots, X_n)$ on observe $(T_1, D_1), \ldots, (T_n, D_n)$ avec :

$$T_i = X_i \wedge X_{(r)} \text{ et } D_i = \begin{cases} 1 \; si \; X_i = T_i \\ 0 \; si \; X_i \neq T_i \end{cases}$$

avec $X_{(r)}$ la $r^{ième}$ statistique d'ordre de l'échantillon $(X_1, \ldots, X_n)$. La définition de l'indicatrice de censure peut se réécrire $D_i = \begin{cases} 1 \; si \; X_i \leq X_{(r)} \\ 0 \; si \; X_i > X_{(r)} \end{cases}$, qui est une forme analogue au cas de la censure fixe avec $C = X_{(r)}$.

La vraisemblance a une forme proche du cas de la censure de type I ; on remarque pour l'écrire que, dans la partie discrète de la distribution, il convient de choisir les instants des *r* sorties parmi les *n* observations. Cela conduit à écrire :

$$L(\theta) = \frac{n!}{(n-r)!}\left[\prod_{i=1}^{r} f_\theta\left(X_{(i)}\right)\right] S_\theta\left(X_{(r)}\right)^{n-r}$$

$$= \frac{n!}{(n-r)!}\prod_{i=1}^{n} f_\theta(T_i)^{D_i}\, S_\theta(T_i)^{1-D_i} = \frac{n!}{(n-r)!}\prod_{i=1}^{n} h_\theta(T_i)^{D_i}\, S_\theta(T_i)$$

Si la loi de référence est la loi exponentielle, on trouve ainsi que :

$$L(\theta) = \frac{n!}{(n-r)!}\theta^r \exp(-\theta T)$$

avec $T = \sum_{i=1}^{r} T_{(i)} + (n-r)T_{(r)}$; la statistique T est donc exhaustive pour le modèle. L'estimateur du maximum de vraisemblance se déduit facilement de l'expression ci-dessus : $\hat{\theta} = \frac{r}{T}$. En fait on peut dans ce cas déterminer complètement la loi de T ; précisément :

Proposition : $2\theta T$ suit une loi du Khi-2 à 2r degrés de liberté ou, de manière équivalente, T suit une loi $\gamma(r,\theta)$ puisque la loi du Khi-2 à 2r degrés de liberté est une loi Gamma de paramètres $\left(r, \frac{1}{2}\right)$

Démonstration : On veut montrer que $P(T \le x) = P\left(\chi^2_{2r} \le 2\theta x\right)$; comme la loi du Khi-2 à 2r degrés de liberté est une loi Gamma de paramètre $\left(r, \frac{1}{2}\right)$, sa densité est :

$$f(x) = \frac{1}{2^r \Gamma(r)} x^{r-1} e^{-\frac{x}{2}}.$$

On écrit :

$$P(T \le x) = \frac{n!}{(n-r)!}\theta^r \int_{A_x} \exp\left(-\theta\left(\sum_{i=1}^{r} t_i + (n-r)t_r\right)\right) dt_1 \ldots dt_r ,$$

avec $A_x = \left\{0 < t_1 \ldots < t_r \,/\, \sum_{i=1}^{t} t_i + (n-r)t_r \le x\right\}$. On fait le changement de variable :

$$t_1 = u_1; t_2 = u_1 + u_2 ; \ldots ; t_{r-1} = u_1 + \ldots + u_{r-1} ; \sum_{i=1}^{r-1} t_i + (n-r+1)t_r = u .$$

On vérifie que le déterminant de la matrice jacobienne de terme générique $\frac{\partial t_i}{\partial u_j}$ vaut $\frac{1}{n-r+1}$, ce qui conduit à :

$$P(T \le x) = \frac{n!}{(n-r)!}\theta^r \int_{B_x} \frac{1}{n-r+1} e^{-\theta u} du_1..du_{r-1}du$$

avec $B_x = \left\{ u_1 > 0,\ldots,u_{r-1} > 0 ; \sum_{i=1}^{t-1}(r-i)u_i < u \le x \right\}$. Le nouveau changement de variable :

$$v_i = (r-i)\times u_i, 1 \le i \le r-1 ; v = u$$

permet finalement d'obtenir :

$$P(T \le x) = \frac{n!}{(n-r+1)!}\theta^r \int_0^x \left[\int_{C_v} \frac{1}{(r-1)!} dv_1..dv_{r-1} \right] e^{-\theta v} dv$$

avec $C_v = \left\{ v_1 > 0,..,v_{r-1} > 0 ; \sum_{i=1}^{t-1} v_i \le v \right\}$; en observant que l'intégrale multiple sur C_x est de la forme $cste \times x^{t-1}$ on en conclut finalement que :

$$P(T \le x) = \frac{1}{\Gamma(r)} \int_0^{\theta x} u^{r-1} e^{-u} du = P\left(\chi^2_{2r} \le 2\theta x\right).$$

On déduit en particulier de cette proposition que l'estimateur EMV est biaisé et que $E(\hat{\theta}) = \frac{r}{r-1}\theta$: en effet, si T suit une loi gamma de paramètre (r,λ) alors $E\left(T^p\right) = \lambda^{-p}\frac{\Gamma(r+p)}{\Gamma(r)}$ pour tout $p > -r$ et donc :

$$E(\hat{\theta}) = 2\theta r E\left(\frac{1}{2\theta T}\right) = 2\theta r \frac{1}{2}\frac{\Gamma(r-1)}{\Gamma(r)} = \theta\frac{r}{r-1}.$$

Le meilleur estimateur sans biais pour θ est donc $\tilde{\theta} = \frac{r-1}{T}$. On montre de même que la variance de $\tilde{\theta}$ est $V(\tilde{\theta}) = \frac{\theta^2}{r-2}$.

Ce résultat peut être obtenu plus simplement. On utilise pour cela le fait que la loi conjointe de la statistique d'ordre $\left(X_{(1)},\ldots,X_{(n)}\right)$ est $\bar{f}(x_1,\ldots,x_n) = n!\prod_{i=1}^{n} f(x_i)1_{\{x_1<\ldots<x_n\}}$. Par un changement de variable, on montre alors que les variables aléatoires $Y_i = (n-i+1)\left(X_{(i)} - X_{(i-1)}\right)$ sont indépendantes et de loi commune la loi exponentielle de paramètre θ.

Comme $T = \sum_{i=1}^{r} Y_i$ on a immédiatement le résultat en observant que la somme de *r* variables exponentielles de paramètre θ a une loi $\gamma(r,\theta)$. On en déduit également très facilement la durée moyenne de l'expérience : puisque $T_{(r)} = \sum_{i=1}^{r} \frac{Y_i}{n-i+1}$, on a $E(T_{(r)}) = \frac{1}{\theta}\sum_{i=1}^{r} \frac{1}{n-i+1}$.

1.4. Troncature

On dit qu'il y a troncature gauche (resp. droite) lorsque la variable d'intérêt n'est pas observable lorsqu'elle est inférieure à un seuil $c > 0$ (resp. supérieure à un seuil $C > 0$).

Le phénomène de troncature est très différent de la censure, puisque dans ce cas on perd complètement l'information sur les observations en dehors de la plage : dans le cas de la censure, on a connaissance du fait qu'il existe une information, mais on ne connaît pas sa valeur précise, simplement le fait qu'elle excède un seuil ; dans le cas de la troncature on ne dispose pas de cette information.

La distribution observée dans ce cas est donc la loi conditionnelle à l'événement $\{c < T < C\}$. La fonction de survie tronquée s'écrit donc :

$$S(t \mid c < T < C) = \begin{cases} 1 \; si \; t < c \\ \frac{S(t) - S(C)}{S(c) - S(C)} \; si \; c \le t \le C \\ 0 \; si \; t > C \end{cases}$$

La fonction de hasard a également le support $\{c < t < C\}$ et s'écrit $h(t \mid c < T < C) = h(t)\frac{S(t)}{S(t) - S(C)}$, ce qui montre que l'expression de *h* ne dépend pas de *c*.

La troncature droite augmente la fonction de hasard, et s'il n'y a que de la troncature gauche $(C = +\infty)$ alors la fonction de hasard n'est pas modifiée.

La troncature peut s'observer par exemple dans le cas d'une migration informatique au cours de laquelle n'auraient été repris dans la nouvelle base que les sinistres encore en cours au moment de la bascule ; les informations sur les sinistres de durée plus courte, pour les mêmes survenances, sont alors perdues. La troncature s'observe également dans le cas d'un contrat d'arrêt de travail avec une franchise : les arrêts de durée inférieure à la franchise ne sont pas observés, et on ne dispose donc sur eux d'aucune information.

1.5. Synthèse : troncature gauche et censure aléatoire droite

Dans les applications actuarielles (construction de tables de mortalité, de lois d'incidence et de maintien en arrêt de travail, etc.), le cadre de modélisation issu de l'expérience est le suivant.

Les individus ne sont en général pas observés depuis l'origine, mais depuis l'âge (ou l'ancienneté) atteint au début de la période d'observation, qu'on notera E_i.

La censure C_i peut être inférieure à l'âge atteint en fin de période d'observation si la sortie a lieu de manière anticipée (résiliation par exemple).

1.5.1. Expression générale de la vraisemblance

Dans ces conditions, l'expression de la vraisemblance du modèle est :

$$L(\theta) = \prod_{i=1}^{n} h_{\theta,E}(t_i)^{d_i} \, S_{\theta,E}(t_i)$$

ce qui s'écrit :

$$\ln L(\theta) = cste + \sum_{i=1}^{n} d_i \ln\left(h_\theta(t_i)\right) + \ln S_\theta(t_i) - \ln S_\theta(e_i).$$

comme $h_{\theta,E}(t_i) = h_\theta(t_i)$ et $S_{\theta,E}(t_i) = \dfrac{S_\theta(t_i)}{S_\theta(e_i)}$.

Si tous les individus sont observés depuis l'origine, $e_i = 0$ et on retrouve l'expression classique :

$$\ln L(\theta) = cste + \sum_{i=1}^{n} d_i \ln\left(h_\theta(t_i)\right) + \ln S_\theta(t_i).$$

Exemple : on considère le modèle à hasard proportionnel de Weibull (*cf.* 3.2) dans lequel :

$$h(x\,|z;\theta,\alpha) = \exp(-z'\theta)\,\alpha x^{\alpha-1}.$$

La log-vraisemblance de ce modèle s'écrit d'après l'expression générale rappelée *supra* :

$$\ln L(y\,|z;\theta,\alpha) = d\ln(\alpha) + (\alpha-1)\sum_{i=1}^{n} d_i \ln(t_i) - \sum_{i=1}^{n} d_i z_i'\theta - \sum_{i=1}^{n} \exp\left(-z_i'\theta\right)(t_i - e_i)^{\alpha}$$

où on a noté $d = \sum_{i=1}^{d} d_i$ le nombre de sorties non censurées.

1.5.2. Cas particulier du modèle exponentiel, lien avec le modèle de Poisson

On considère *n* individus pour lesquels on fait l'hypothèse que la fonction de hasard sous-jacente est constante sur un intervalle $[x, x+1[$; à l'aide de ce qui précède on trouve que la log-vraisemblance du modèle est, à une constante près :

$$\ln L(\theta) = \sum_{i=1}^{n} \left[d_i \ln(\theta) + \theta \times (t_i - e_i) \right] = d_x \times \ln(\theta) + \theta \times E_x$$

avec $d_x = \sum_{i=1}^{d} d_i$ et $E_x = \sum_{i=1}^{d} (t_i - e_i)$. On remarque alors que tout se passe comme si la variable D_x qui compte le nombre de sorties sur l'intervalle $[x, x+1[$ était une loi de Poisson de paramètre $\theta \times E_x$; en effet, dans ce cas $\ln\left(P(D_x = d)\right) = cste + d_x \times \ln(\theta) - \theta \times E_x$.

1.5.3. Maximum de vraisemblance discrétisé

Les praticiens utilisent peu l'approche décrite ci-dessus et raisonnent le plus souvent de la manière suivante.

Dans le cadre du modèle binomial[4], le nombre de décès observés à l'âge x, D_x, suit une loi binomiale de paramètres $(N_x, q_x(\theta))$ et la vraisemblance associée à la réalisation d'un nombre d_x de décès est donc égale à :

$$P(D_x = d_x) = C_{N_x}^{d_x} \, q_x^{d_x} \, (1 - q_x)^{N_x - d_x}.$$

Pour l'ensemble des observations on obtient donc la log-vraisemblance suivante (à une constante indépendante du paramètre près) :

$$\ln L(\theta) = \sum_x d_x \ln q_x(\theta) + \sum_x (N_x - d_x) \ln(1 - q_x(\theta)).$$

Cette expression n'est pas très aisée à manipuler (par exemple dans le cadre du modèle de Makeham on montrera que $q_x(\theta) = 1 - s \times g^{c^x(c-1)}$), quoique numériquement la recherche du maximum ne pose pas de problème majeur. Afin de parvenir à un problème de moindres carrés pondérés, on réalise toutefois plutôt en général l'approximation de la loi de $\hat{q}_x$ par une loi normale :

$$\hat{q}_x \approx N\left(q_x(\theta); \sigma^2(\theta) = \frac{q_x(\theta)(1 - q_x(\theta))}{N_x} \right).$$

[4] On peut en pratique souvent se ramener à ce modèle modulo une détermination adaptée de l'effectif soumis au risque.

La fonction de vraisemblance s'écrit alors, en faisant l'hypothèse d'indépendance entre les âges :

$$L(\theta) = \prod_x \frac{1}{\sigma(\theta)\sqrt{2\pi}} \exp\left(\frac{1}{2}\frac{\left(q_x(\theta) - \hat{q}_x\right)^2}{\sigma^2(\theta)}\right),$$

d'où la log-vraisemblance :

$$\ln\left(L(\theta)\right) = \sum_x \ln\left(\frac{1}{\sigma(\theta)\sqrt{2\pi}}\right) - \sum_x \frac{1}{2}\frac{\left(q_x(\theta) - \hat{q}_x\right)^2}{\sigma^2(\theta)}.$$

La fonction objectif est là encore complexe et le paramètre intervient à la fois dans l'espérance et dans la variance de la loi normale ; cela peut engendrer une certaine instabilité des algorithmes de recherche de l'optimum ; aussi, on va utiliser la vraisemblance approchée dans laquelle on remplace la variance théorique par la variance estimée. La maximisation de la vraisemblance est alors équivalente à la minimisation de :

$$\varphi(\theta) = \sum_x \frac{1}{2}\frac{\left(q_x(\theta) - \hat{q}_x\right)^2}{\hat{\sigma}^2} = \sum_x \frac{N_x}{\hat{q}_x(1-\hat{q}_x)}\left(q_x(\theta) - \hat{q}_x\right)^2.$$

Le problème est ainsi ramené à un problème de moindres carrés pondérés dans le cas non linéaire ; il peut être résolu numériquement dans la plupart des logiciels statistiques spécialisés.

Dans les lignes qui précèdent, à aucun moment n'interviennent la censure et la troncature. Celles-ci se trouvent en fait incluses dans effectif de référence N_x de l'expérience binomiale. Il reste donc à spécifier correctement ce coefficient.

Il apparaît raisonnable de souhaiter qu'en moyenne le modèle soit sans biais, ce qui se traduit par $E(D_x) = q_x \times N_x$.

En l'absence de troncature et de censure, on choisit donc $N_x = S(x)$.

En présence de troncature et / ou de censure, il faut prendre en compte ces phénomènes dans le calcul. On peut montrer qu'il est alors raisonnable de retenir l'exposition au risque $N_x = E_x$ où $E_x = \sum_{i \in I} \tau_i(x)$ avec $\tau_i(x)$ la durée de présence à risque de l'individu *i*.

Ce résultat sera justifié de la manière suivante : le nombre de sorties estimé par le modèle est simplement l'espérance de la variable aléatoire « nombre de sorties » ; avec des notations évidentes, on a, en se limitant à l'intervalle[5] $[x, x+1[$:

[5] Ce qui signifie que l'on travaille avec la loi de *X* tronquée à gauche en *x* et la censure est au maximum égale à *x*+1.

$$P(D_i = 1) = \int_e^{x+1} \left(1 - \frac{S_X(c)}{S_X(e)}\right) F_C(dc)$$

En supposant la constance de la fonction de hasard sur cet intervalle, $1 - \frac{S_X(c)}{S_X(e)} = 1 - e^{-\mu_x \times (c-e)}$; pour éviter l'intervention explicite de la loi de la censure $F_C(dc)$ dans le calcul, on suppose μ_x assez petit pour que $1 - e^{-\mu_x \times (c-e)} \approx \mu_x \times (c-e)$, alors

$$P(D_i = 1) = \mu_x \times \int_e^{x+1} (c-e) F_C(dc).$$

Le nombre de sorties espéré sur l'intervalle $[x, x+1[$ s'écrit donc, en supposant la constance de la fonction de hasard sur cet intervalle et en utilisant une approximation affine de la fonction de survie, valide lorsque le taux de hasard μ_x est petit, $E(D) = \sum_{i=1}^{n} P(D_i = 1) = \mu_x \times E(x)$, avec $E(x) = \sum_{i=1}^{n} \left[t_i \wedge (x+1) - e_i \vee x \right]^+$; à l'aide de la probabilité conditionnelle de sortie ajustée, on a donc

$$E(D) = -\ln(1 - q(x)) \times E(x).$$

On peut éviter de faire l'hypothèse que μ_x est petit et n'utiliser donc que l'hypothèse de constance de la fonction de hasard en utilisant

$$E(D) \approx \sum_{i=1}^{n} \left(1 - e^{-\mu_x \times (T_i - E_i)}\right) = \sum_{i=1}^{n} \left(1 - (1 - q(x))^{(T_i - E_i)}\right).$$

Lorsque la probabilité conditionnelle de sortie est petite, cette expression devient $E(D) \approx q(x) \times E(x)$, qui est la formule souvent utilisée par les praticiens.

2. Vraisemblances latente et observable en présence de censure

Dans ce paragraphe, on considère des observations de durées $(t_1, \ldots, t_n)$, censurées par une censure de type I (censure fixe) ou III (censure aléatoire non informative), dépendant de l'observation[6] ; c'est en effet le type de censure que l'on rencontre le plus souvent dans les problèmes d'assurance. On note $(c_1, \ldots, c_n)$ les valeurs observées de la censure. Enfin, on suppose que les durées de vie observées dépendent également de *p* variables explicatives[7] $(z_1, \ldots, z_p)$. On a déterminé dans la partie précédente la forme de la vraisemblance générale, et on souhaite maintenant réaliser l'estimation des paramètres par maximisation de cette vraisemblance, en intégrant la prise en compte de ces variables explicatives. On

[6] Cela revient au même qu'une censure aléatoire en raisonnant conditionnellement à la valeur de la censure.

[7] z_j est donc un vecteur composé des *n* valeurs de l'explicative pour les individus de l'échantillon.

s'attachera ainsi à exprimer la relation entre le score latent et le score observable, et à obtenir l'information de Fisher du modèle observable. On observe donc comme précédemment :

$$T_i = X_i \wedge C_i \text{ et } D_i = \begin{cases} 1 \; si \; X_i \leq C_i \\ 0 \; si \; X_i > C_i \end{cases}$$

et les variables $Y_i = (T_i, D_i)$ sont indépendantes. Lorsque la censure est connue, Y_i est une fonction de la variable latente X_i ; le modèle observable est donc un modèle qui fournit une information incomplète sur X_i. Cette relation fonctionnelle entre variables latentes et variables observables entraîne des conséquences sur la forme de la vraisemblance observable. Plus précisément, on a une relation fonctionnelle de la forme $Y = \phi(X)$; les densités respectives de *Y* et *X* sont notées[8] $l(\theta)$ et $l^*(\theta)$; l'observation de *Y* fournit une information sur la loi de *X*, et il est naturel de s'intéresser à la loi conditionnelle de $X \,|\, Y = y$; on a :

$$l^*(x,\theta) = l(y,\theta) l(x \,|\, y,\theta)$$

et en passant à la log-vraisemblance on peut donc écrire[9] :

$$\ln l^*(x,\theta) = \ln l(y,\theta) + \ln l(x \,|\, y,\theta)$$

En dérivant cette expression par rapport à θ, puis en intégrant par rapport à la loi de $X \,|\, Y = y$, on trouve[10] :

$$E\left[\frac{\partial \ln l^*(x,\theta)}{\partial \theta} \,|\, y\right] = \frac{\partial \ln l(y,\theta)}{\partial \theta} + E\left[\frac{\partial \ln l(x \,|\, y,\theta)}{\partial \theta} \,|\, y\right].$$

Mais $E\left[\frac{\partial \ln l(x \,|\, y,\theta)}{\partial \theta} \,|\, y\right] = \int \frac{\partial l(x \,|\, y,\theta)}{\partial \theta} dx$ puisque la loi conditionnelle de $X \,|\, Y = y$ a pour densité $l(x \,|\, y,\theta)$; en inversant dérivation et intégrale, comme l'intégrale de la densité est égale à un, on trouve que $\int \frac{\partial l(x \,|\, y,\theta)}{\partial \theta} dx = 0$, et donc le score s'écrit :

$$\frac{\partial \ln l(y,\theta)}{\partial \theta} = E\left[\frac{\partial \ln l^*(x,\theta)}{\partial \theta} \,|\, y\right]$$

[8] On notera *l* la vraisemblance pour une observation et *L* la vraisemblance d'un échantillon.
[9] On se place dans le cas d'une censure non informative et on n'indique pas la censure dans ces expressions.
[10] Les espérances dépendent du paramètre θ qui est omis dans les notations pour alléger les écritures.

Le score observable est donc la meilleure prédiction du score latent, conditionnellement aux observations. En dérivant 2 fois l'expression de la log-vraisemblance on obtient de même :

$$-\frac{\partial^2 \ln l^*(x,\theta)}{\partial\theta\partial\theta'} = -\frac{\partial^2 \ln l(y,\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 \ln l(x|y,\theta)}{\partial\theta\partial\theta'}$$

puis en prenant l'espérance on trouve que les informations de Fisher des modèles latent et observable sont liées par la relation :

$$I_X^*(\theta) = I_Y(\theta) + E\left[E\left(-\frac{\partial^2 \ln l(x|y,\theta)}{\partial\theta\partial\theta'}|y\right)\right]$$

Remarque : la notation $\frac{\partial^2 f(\theta)}{\partial\theta\partial\theta'}$ désigne la matrice Hessienne associée à *f*, de terme courant $\frac{\partial^2 f(\theta)}{\partial\theta_i\partial\theta_j}$.

2.1. Application de la méthode du maximum de vraisemblance

On présente dans cette section les liens entre vraisemblance observable et vraisemblance latente dans un modèle général, avant de spécifier les écritures dans le cas d'un modèle de durée.

2.1.1. Généralités

On suppose l'indépendance des observations conditionnellement aux variables explicatives et aux censures ; la log-vraisemblance du modèle s'écrit :

$$\ln L(y|z,c;\theta) = \sum_{i=1}^{n} \ln l(y_i|z_i,c_i;\theta)$$

et dès lors que la log-vraisemblance est dérivable, l'estimateur du maximum de vraisemblance annule le vecteur des scores : $\frac{\partial \ln L(y|z,c;\hat{\theta})}{\partial\theta} = 0$.

Sous des conditions techniques de régularité la plupart du temps satisfaites en pratique, on sait alors qu'il existe un maximum local de la log-vraisemblance convergeant presque sûrement vers la vraie valeur du paramètre et que, de plus, l'estimateur du maximum de vraisemblance est asymptotiquement efficace et gaussien, *i.e.* :

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

avec l'information de Fisher définie par $I(\theta) = \lim_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} E\left(-\frac{\partial^2 \ln l(y_i | z_i, c_i)}{\partial\theta \partial\theta'} | z_i, c_i \right)$, la limite étant en probabilité. La variance asymptotique de l'estimateur peut être estimée par :

$$\hat{V}(\hat{\theta}) = -n \times \left[\frac{\partial^2 \ln L(y | z, c; \hat{\theta})}{\partial\theta \partial\theta'} \right]^{-1}$$

On dispose ainsi d'un cadre assez général pour estimer le paramètre par maximum de vraisemblance en présence de censure[11] et de variables explicatives.

2.1.2. Vraisemblance latente et vraisemblance observable

La vraisemblance du modèle complet, latent, n'est pas observable ; on a toutefois une relation simple entre le score latent et le score observable, au sens où le score observable est la prévision optimale du score latent à partir des variables observables, soit de manière formelle :

$$\frac{\partial \ln L(y | z, c; \theta)}{\partial\theta} = E\left[\frac{\partial \ln L^*(x | z, c; \theta)}{\partial\theta} | y, z, c \right]$$

Cette propriété découle directement de la relation établie pour une observation en introduction : $\frac{\partial \ln l(y, \theta)}{\partial\theta} = E\left[\frac{\partial \ln l^*(x, \theta)}{\partial\theta} | y \right]$.

En ce qui concerne l'information de Fisher, l'information du modèle latent peut être décomposée en la somme de l'information du modèle observable et d'un terme mesurant la perte d'information due à la présence de la censure. On a le résultat suivant :

Proposition : $I^*(\theta) = I(\theta) + J(\theta)$, avec :

$$J(\theta) = \lim_{n\to+\infty} E\left[\frac{1}{n} \sum_{i=1}^{n} V\left(\frac{\partial \ln l^*(x_i | z_i, c_i; \theta)}{\partial\theta} | y_i, z_i, c_i \right) | z, c \right],$$

la limite étant prise en probabilité.

Pour prouver ce résultat on applique l'équation de décomposition de la variance $V[A] = E(V[A|B]) + V(E[A|B])$ à $A = \frac{\partial \ln l^*(x_i | z_i, c_i; \theta)}{\partial\theta} | z_i, c_i$ et $B = Y$.

[11] La forme de la vraisemblance dans le cas d'un modèle de durée est précisée en 2.2.

2.2. Écritures particulières aux modèles de durée

Dans le cas d'un modèle de durée, on calcule la vraisemblance en fonction du taux de hasard et de la fonction de survie, plutôt que de la densité ; comme on a $f(t) = S(t)h(t)$, on obtient :

$$\ln L^*(x|z;\theta) = \sum_{i=1}^{n} \ln h(x_i|z_i;\theta) + \sum_{i=1}^{n} \ln S(x_i|z_i;\theta)$$

La log-vraisemblance observable est calculée conditionnellement à (z,c) et s'exprime par[12] :

$$\ln L(y|z,c;\theta) = \sum_{i=1}^{n} d_i \ln h(t_i|z_i;\theta) + \sum_{i=1}^{n} \ln S(t_i|z_i;\theta)$$

On retrouve donc, comme on l'avait établi en 1.2 ci-dessus que la log-vraisemblance observable s'écrit de la même manière que dans le modèle latent, mais en remplaçant la durée réelle par la durée tronquée et en ne conservant la fonction de hasard que pour les informations complètes (repérées par $d_i = 1$).

Les équations de vraisemblance n'ont toutefois pas d'expression simple dans le cas général ; on utilisera les algorithmes usuels pour déterminer l'EMV de manière approchée : NEWTON-RAPHSON, BHHH (BERNDT, HALL, HALL, HAUSMAN) et algorithme EM, ce dernier étant particulièrement bien adapté au cas des données incomplètes. Ces algorithmes sont présentés en 2.4 *infra*.

Cependant, dans certaines classes de modèles une approche directe reste possible : cela est notamment le cas des modèles à hasard proportionnel, étudiés ci-après.

2.3. Exemple : le modèle de Weibull

On a vu en introduction l'estimation des paramètres du modèle de Weibull dans le cas non censuré. On traite maintenant à titre d'exemple le cas d'une censure droite. On considère donc le modèle :

$$f(x) = \frac{\alpha}{l^{\tau}} x^{\alpha-1} \exp\left\{-\left(\frac{x}{l}\right)^{\alpha}\right\},\ S(x) = \exp\left\{-\left(\frac{x}{l}\right)^{\alpha}\right\}$$

pour lequel on observe un échantillon censuré $(t_i, d_i)_{i\in\{1,\ldots,n\}}$ où $d_i = \begin{cases} 1 & \text{si } t_i = x_i \\ 0 & \text{si } t_i < x_i \end{cases}$ est l'indicatrice d'une information non censurée.

2.3.1. Estimation des paramètres

La vraisemblance de ce modèle s'écrit :

[12] Voir 1.1. Cette égalité est vraie à une constante additive prêt dépendant de la distribution de la censure.

$$L(\alpha,l) \propto \prod_{i=1}^{n} f(t_i)^{d_i} S(t_i)^{1-d_i}$$

En notant $d_\bullet = \sum_{i=1}^{n} d_i$ le nombre de sorties observées non censurées, il vient :

$$L(\alpha,l) \propto \left(\frac{\alpha}{l^\alpha}\right)^{d_\bullet} \prod_{i=1}^{n} t_i^{(\alpha-1)d_i} \exp\left\{-d_i\left(\frac{t_i}{l}\right)^\alpha\right\} \exp\left\{-(1-d_i)\left(\frac{t_i}{l}\right)^\alpha\right\},$$

$$L(\alpha,l) \propto \left(\frac{\alpha}{l^\alpha}\right)^{d_\bullet} \exp\left\{-l^{-\alpha}\sum_{i=1}^{n} t_i^{\alpha}\right\} \exp\left\{(\alpha-1)\sum_{i=1}^{n} d_i \ln t_i\right\}$$

d'où l'on déduit la log-vraisemblance :

$$\ln L(\alpha,l) = \ln k + d_\bullet(\ln\alpha - \alpha\ln l) - l^{-\alpha}\sum_{i=1}^{n} t_i^{\alpha} + (\alpha-1)\sum_{i=1}^{n} d_i \ln t_i$$

Les équations aux dérivés partielles s'écrivent donc :

$$\begin{cases} \dfrac{\partial}{\partial l}\ln L(\alpha,l) = -\dfrac{d_\bullet}{l} + \alpha l^{-\alpha-1}\displaystyle\sum_{i=1}^{n} t_i^{\alpha} \\ \dfrac{\partial}{\partial \alpha}\ln L(\alpha,l) = d_\bullet\left(\dfrac{1}{\alpha} - \ln l\right) + l^{-\tau}\left(\ln l\displaystyle\sum_{i=1}^{n} t_i^{\alpha} - \sum_{i=1}^{n} t_i^{\alpha}\ln t_i\right) + \displaystyle\sum_{i=1}^{n} d_i \ln t_i \end{cases}$$

On cherche donc les solutions du système suivant :

$$\begin{cases} l = \left(\dfrac{1}{d_\bullet}\displaystyle\sum_{i=1}^{n} t_i^{\alpha}\right)^{1/\alpha} \\ \dfrac{1}{\alpha} = \dfrac{\displaystyle\sum_{i=1}^{n} t_i^{\alpha}\ln t_i}{\displaystyle\sum_{i=1}^{n} t_i^{\alpha}} - \dfrac{1}{d_\bullet}\displaystyle\sum_{i=1}^{n} d_i \ln t_i \end{cases}$$

La deuxième équation définit un algorithme qui converge vers $\hat{\alpha}$ pour autant qu'on lui fournisse une valeur initiale pas trop éloignée. En pratique, cette valeur pourra être l'estimateur obtenu par la méthode des quantiles sur l'ensemble des observations complètes (*cf.* le support d'introduction). Une fois $\hat{\alpha}$ obtenu, $\hat{l}$ s'en déduit grâce à la première équation.

2.3.2. Application numérique

On propose une illustration dans laquelle 1 000 observations ont été simulées dont 47 % censurées.

Une première estimation des paramètres a été effectuée sur les 1 000 réalisations du risque principal afin d'obtenir des estimations qui serviront d'étalon pour la comparaison des estimations obtenues dans le cas censuré.

Remarque : Il faut définir un critère d'arrêt pour les algorithmes permettant d'obtenir l'e.m.v. $\hat{\alpha}$. Dans cette application, on s'est arrêté lorsque la variation relative de la valeur lors d'une itération devenait inférieure (en valeur absolue) à 0,01 %.

Il convient de remarquer qu'avec le critère d'arrêt qui a été choisi, l'algorithme qui fournit $\hat{\alpha}$ est nettement plus rapide (facteur 10 en nombre d'itérations) dans le cas où l'on ne conserve que des données complètes que dans la situation où l'on dispose de données censurées.

Le tableau suivant reprend les différentes estimations des paramètres effectuées et indique l'espérance et la variance correspondant à ces estimations. Les simulations ont été effectuées en prenant comme valeur théorique pour les paramètres $\alpha = 2,5$ et $l = 45$.

Tab. 1. Synthèse des estimations

	Données complètes	Données incomplètes	
		Pas de prise en compte des censures	Prise en compte des censures
α	2,43	2,26	2,48
l	44,78	38,31	44,65
Moyenne	39,71	33,93	39,61
Variance	302,97	252,29	291,72

Le tableau suivant reprend les erreurs relatives d'estimation en référence à la situation dans laquelle toutes les observations sont complètes.

Tab. 2. Erreurs relatives d'estimation

	Pas de prise en compte des censures	Prise en compte des censures
α	-7,1 %	1,9 %
L	-14,5 %	-0,3 %
Moyenne	-14,5 %	-0,2 %
Variance	-16,7 %	-3,7 %

L'utilisation des *toutes* les données disponibles, même incomplètes, s'avère essentielle. En particulier, ne pas prendre en compte les censures conduit à sous-estimer de 15 % la durée

de survie. Dans le même esprit, en présence de censures de type I ou II, ne pas prendre la totalité des observations disponibles conduit à estimer un modèle dans lequel la durée de survie maximale est le niveau de la censure.

2.4. Les algorithmes numériques de maximisation de la vraisemblance

Comme on l'a vu en 2.2 ci-dessus, l'expression analytique de la log-vraisemblance ne rend que rarement possible un calcul direct de l'estimateur du maximum de vraisemblance. Bien entendu, les algorithmes standards de type Newton-Raphson peuvent être utilisés dans ce contexte. Toutefois, des méthodes spécifiques peuvent s'avérer mieux adaptées.

Le lecteur intéressé par une introduction aux méthodes numériques d'optimisation pourra consulter CIARLET [1990].

2.4.1. L'algorithme de Newton-Raphson

On utilise ici pour résoudre l'équation $f(x_0) = 0$ un algorithme construit à partir d'une linéarisation au voisinage de la solution, sur la base du développement de Taylor à l'ordre un ; en notant que $f(x_{k+1}) = f(x_k) + (x_{k+1} - x_k)\frac{df}{dx}(x_k) + o(x_{k+1} - x_k)$, on propose ainsi la récurrence définie par $f(x_{k+1}) = 0$, qui conduit à :

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}.$$

Dans le cas d'un modèle de durée, on utilise comme fonction f la dérivée de la log-vraisemblance par rapport au paramètre (le score), ce qui conduit à l'expression :

$$\theta_{k+1} = \theta_k - \left[\frac{\partial^2}{\partial\theta\partial\theta'} \ln L(y|z,c;\theta_k)\right]^{-1} \frac{\partial \ln L(y|z,c;\theta_k)}{\partial\theta}.$$

L'écriture ci-dessus est une écriture matricielle, valable pour un θ multidimensionnel.

Afin que cet algorithme converge il convient de partir d'une valeur initiale « proche » de la valeur théorique. Il possède une propriété intéressante : si l'on dispose d'un estimateur convergent, pas nécessairement asymptotiquement efficace, on peut l'utiliser comme valeur initiale de l'algorithme de Newton-Raphson. On obtient alors l'efficacité asymptotique dès la première itération[13].

Il existe une variante de l'algorithme de Newton-Raphson, appelée algorithme BHHH (BERNDT, HALL, HALL, HAUSMAN), qui consiste à remplacer dans l'expression itérative ci-dessus la matrice d'information de Fischer par son expression ne faisant appel qu'à la dérivée première de la log-vraisemblance. On obtient ainsi :

[13] Dans ce cas l'estimateur obtenu n'est pas du maximum de vraisemblance, mais il est tout de même asymptotiquement efficace.

$$\theta_{k+1} = \theta_k - \left[\sum_{i=1}^{n} \frac{\partial \ln l(y_i | z_i, c_i; \theta_k)}{\partial \theta} \frac{\partial \ln l(y_i | z_i, c_i; \theta_k)}{\partial \theta'} \right]^{-1} \sum_{i=1}^{n} \frac{\partial \ln l(y_i | z_i, c_i; \theta_k)}{\partial \theta}$$

Cette version de l'algorithme de Newton-Raphson a les mêmes propriétés que la précédente.

2.4.2. L'algorithme Espérance-Maximisation (EM)

Cet algorithme a été imaginé plus spécifiquement dans le cadre de données incomplètes ; il s'appuie sur la remarque que, si les variables $(x_1, \ldots, x_n)$ étaient observables, l'estimation serait effectuée simplement en maximisant la log-vraisemblance latente $\ln L(x | z, c; \theta)$; comme on ne dispose pas de ces observations, l'idée est de remplacer la fonction objectif par sa meilleure approximation connaissant les variables observables $(y_1, \ldots, y_n)$. Il a été proposé initialement par DEMPSTER et al. [1977].

On introduit, pour $(\theta, \hat{\theta})$ fixé, la fonction $q(\theta, \hat{\theta}) = E_{\hat{\theta}}\left[\ln L^*(x | z, c; \theta) | y, z, c \right]$; l'algorithme EM est alors défini par la répétition des étapes suivantes :

- calcul de $q(\theta, \theta_k)$;
- maximisation en θ de $q(\theta, \theta_k)$, dont la solution est θ_{k+1}

En pratique cet algorithme est intéressant lorsque le calcul de $q(\theta, \theta_k)$ est sensiblement plus simple que le calcul direct de $\ln L(y | z, c; \theta)$; dans le cas contraire, on peut être conduit à utiliser un algorithme de Newton-Raphson pour l'étape d'optimisation de $q(\theta, \theta_k)$, ce qui alourdit la démarche.

L'algorithme EM possède sous certaines conditions de régularité qui ne seront pas détaillées ici les « bonnes propriétés » suivantes :

Proposition : L'algorithme EM est croissant, au sens où $\ln L(y | z, c; \theta_{k+1}) \geq \ln L(y | z, c; \theta_k)$; de plus toute limite θ_∞ d'une suite de solutions (θ_k) satisfait la condition du premier ordre :

$$\frac{\partial \ln L(y | z, c; \theta_\infty)}{\partial \theta} = 0$$

Démonstration : voir DROESBEKE et *al.* [1989].

2.4.3. Les autres méthodes

D'autres méthodes peuvent s'avérer utiles dans le cas d'échantillons fortement censurés ; en effet dans ce cas, l'estimation « fréquentielle » usuelle utilisée jusqu'ici peut s'avérer mal adaptée ; on peut alors se tourner vers des algorithmes d'échantillonnage pondéré bayésiens, notamment les algorithmes MCMC.

Cette situation étant peu courante en assurance ne sera pas développée ici ; le lecteur intéressé pourra consulter ROBERT [1996].

3. Les modèles à hasard proportionnel

Dans ces modèles la fonction de hasard est écrite $h(x|z;\theta) = \exp(-z'\theta) h_0(x)$ avec h_0 la fonction de hasard de base, qui est une donnée. Cette situation se rencontre par exemple lorsque l'on veut positionner la mortalité d'un groupe spécifique par rapport à une mortalité de référence, connue, représentée par h_0. On peut par exemple imaginer que l'on a ajusté la mortalité d'un groupe important selon un modèle de Makeham[14] et que l'on s'intéresse au positionnement de la mortalité de certaines sous-populations : hommes / femmes, fumeurs / non-fumeurs, *etc*. Dans cette approche, on s'attachera essentiellement à définir le positionnement d'une population par rapport à une autre, sans chercher toujours le niveau absolu du risque. L'expression de la fonction de hasard d'un modèle proportionnel peut s'écrire :

$$\ln \frac{h(x|z;\theta)}{h_0(x)} = -z'\theta,$$

ce qui exprime que le logarithme du taux de risque instantané, exprimé relativement à un taux de base, est une fonction linéaire des variables explicatives. Les variables explicatives sont au nombre de *p*, ce qui implique que $z'\theta = \sum_{i=1}^{p} z_i \theta_i$. On vérifie aisément que la fonction de survie du modèle est de la forme :

$$S(x|z;\theta) = \exp\left(-\exp(-z'\theta) H_0(x)\right),$$

avec H_0 la fonction de hasard cumulée de base[15]. Compte tenu de la forme de la fonction de survie, il est naturel de s'intéresser à la variable transformée $V = \ln(H_0(X))$; en effet si on considère le modèle suivant :

$$v = z'\theta + \varepsilon$$

(en d'autres termes on pose $\varepsilon = v - z'\theta$) on trouve que

$$P(\varepsilon > t|z;\theta) = P(\ln H_0(x) - z'\theta > t|z;\theta) = P(H_0(x) > \exp(z'\theta)\exp(t)|z;\theta),$$

soit :

$$P(\varepsilon > t|z;\theta) = S\left(H_0^{-1}\left[\exp(z'\theta)\exp(t)\right]|z;\theta\right) = \exp(-\exp(t))$$

[14] Voir la section 5.

[15] En utilisant la relation $S(t) = \exp\left(-\int_0^t h(s)ds\right)$.

La loi (conditionnelle) du résidu ε est donc une loi de Gumbel[16], qui vérifie $E(\varepsilon) = -\gamma$ et $V(\varepsilon) = \frac{\pi^2}{6}$, γ étant la constante d'Euler[17].

On reconnaît dans l'équation $v = z'\theta + \varepsilon$ une formulation formellement équivalente à celle d'un modèle linéaire, dans lequel les résidus ne sont toutefois ni gaussiens, ni centrés, puisque $E(\varepsilon) = -\gamma$:

$$E(V|z;\theta) = -\gamma + z'\theta$$

Le point important ici est que la loi de ε ne dépend pas du paramètre. Si on souhaite obtenir un modèle avec des résidus centrés on considère la transformation $V = H_0(X)$. On a $P(V > t) = P\left(X > H_0^{-1}(t)\right) = S\left(H_0^{-1}(t)\right)$ et donc :

$$P(V > t) = \mathbf{exp}\left(-\mathbf{exp}(-z'\theta) \times t\right).$$

V suit donc une loi exponentielle de paramètre $\mathbf{exp}(-z'\theta)$, ce qui conduit à poser le modèle non linéaire :

$$v = \mathbf{exp}(z'\theta) + \varepsilon$$

avec $E(\varepsilon) = 0$ et $V(\varepsilon) = \mathbf{exp}(2z'\theta)$, et $E(V|z;\theta) = \mathbf{exp}(z'\theta)$. On note que les résidus de ce modèle sont hétéroscédastiques.

On peut noter que le taux de décès d'une sous-population s'exprime simplement à l'aide du taux de décès de base :

$$q(x|z;\theta) = 1 - \left(\frac{S(x+1|z;\theta)}{S(x|z;\theta)}\right) = 1 - \left(\frac{S_0(x+1)}{S_0(x)}\right)^{\mathbf{exp}(-z'\theta)} = 1 - \left(1 - q_0(x)\right)^{\mathbf{exp}(-z'\theta)}.$$

Lorsque $q_0(x)$ est petit on retrouve comme on pouvait s'y attendre :

$$q(x|z;\theta) \approx q_0(x) \times \mathbf{exp}(-z'\theta).$$

3.1. Cas où la fonction de hasard de base est connue[18]

On s'intéresse dans un premier temps au cas de données non censurées dans le cadre du modèle linéaire défini ci-dessus.

On cherche à estimer θ en supposant H_0 connue ; l'équation ci-dessus peut être utilisée pour construire un estimateur convergent du paramètre, mais cet estimateur est non asymptotiquement efficace ; on peut imaginer de l'utiliser comme valeur d'initialisation

[16] Cf. l'introduction consacrée à la loi de Weibull et http://fr.wikipedia.org/wiki/Loi_de_Gumbel
[17] Dont la valeur est approximativement 0,577215665.
[18] Dans le modèle de Cox la fonction de hasard de base est supposée inconnue, alors qu'ici elle est supposée connue.

d'un algorithme de maximisation de la log-vraisemblance. Toutefois, l'expression du modèle sous la forme d'un modèle linéaire conduit naturellement à proposer l'estimateur des moindres carrés ordinaires (MCO) :

$$\hat{\theta}_{MCO} = \left[\sum_{i=1}^{n} z_i' z_i\right]^{-1} \sum_{i=1}^{n} z_i' \ln H_0(x_i).$$

Dans l'écriture ci-dessus $z_i = (z_{i1},\ldots,z_{ip})$ est le vecteur ligne de taille *p* composé des valeurs des variables explicatives pour l'individu n°*i*. Si on suppose que le modèle intègre une constante et que celle-ci est la première composante de θ, alors on peut montrer que $\hat{\theta} - (\gamma,0,\ldots,0)'$ est un estimateur convergent de θ. La transposition directe du cas du modèle linéaire conduit donc à un estimateur simple à calculer et disposant *a priori* de « bonnes propriétés » pour θ.

Dans le cadre du modèle $v = \exp(z'\theta) + \varepsilon$, qui présente l'avantage d'avoir des résidus centrés, l'estimateur EMV est solution du programme de moindres carrés non linéaires :

$$\mathbf{Min} \sum_{i=1}^{n} \left[H_0(x_i) - \exp(z_i'\theta)\right]^2.$$

Cet estimateur peut être aisément calculé ; toutefois, les estimateurs ci-dessus sont utilisables pour des données complètes, mais pas dans le cas de données censurées.

En effet, en présence de censure, l'estimateur $\hat{\theta}_{MCO}$ restreint aux données complètes est asymptotiquement biaisé. Le biais étant toutefois peu important en pratique, cet estimateur pourra servir de valeur initiale pour des algorithmes numériques.

En présence de données incomplètes, on revient aux équations de vraisemblance du modèle.

3.1.1. Équations de vraisemblance

D'après les équations générales déterminées en 2.1.2 ci-dessus, on a :

$$\ln L^*(x|z;\theta) = \sum_{i=1}^{n} \left(-z_i'\theta + \ln h_0(x_i)\right) - \sum_{i=1}^{n} \exp(-z_i'\theta) H_0(x_i)$$

pour la vraisemblance latente et :

$$\ln L(y|z,c;\theta) = \sum_{i=1}^{n} d_i \left(-z_i'\theta + \ln h_0(t_i)\right) - \sum_{i=1}^{n} \exp(-z_i'\theta) H_0(t_i)$$

pour la vraisemblance observable. Par dérivation on trouve le vecteur des scores latent :

$$\frac{\partial \ln L^*(x|z;\theta)}{\partial \theta} = -\sum_{i=1}^{n} z_i' + \sum_{i=1}^{n} z_i' \exp\left(-z_i'\theta\right) H_0(x_i) = \sum_{i=1}^{n} z_i' \exp\left(-z_i'\theta\right) \varepsilon_i$$

Le score latent est donc le produit scalaire entre les erreurs $\varepsilon_i = H_0(x_i) - \exp\left(z_i'\theta\right)$ et les variables explicatives, pour la métrique définie par les poids $\exp\left(-z_i'\theta\right)$. En ce qui concerne le vecteur des scores observable, on a :

$$\frac{\partial \ln L(y|z,c;\theta)}{\partial \theta} = \sum_{i=1}^{n} z_i' \exp\left(-z_i'\theta\right) \tilde{\varepsilon}_i$$

avec $\tilde{\varepsilon}_i = E\left(\varepsilon_i | y_i, z_i, c_i\right)$. Comme le résidu du modèle non censuré est défini par $\varepsilon_i = H_0(x_i) - \exp\left(z_i'\theta\right)$, il s'agit de montrer que $E\left(\varepsilon_i | y_i, z_i, c_i\right) = H_0(t_i) - d_i \exp\left(z_i'\theta\right)$.

Les équations de vraisemblance s'assimilent donc à une condition d'orthogonalité entre variables explicatives et erreurs prévues, comme dans le cas d'un modèle linéaire classique.

3.1.2. Information de Fisher

L'information de Fisher a ici une expression particulièrement simple :

$$I(\theta) = \sum_{i=1}^{n} z_i' z_i p_i$$

avec $p_i = E(d_i | z_i, c_i) = P(X_i < c_i)$ la probabilité que l'observation soit complète. On écrit pour cela que $\frac{\partial^2 \ln L(y|z,c;\theta)}{\partial \theta \partial \theta'} = -\sum_{i=1}^{n} z_i' z_i \exp\left(-z_i'\theta\right) H_0(t_i)$ puis on prend l'espérance en observant que le vecteur des scores est, dans ce modèle, centré. La décomposition de l'information de Fisher présentée en 2.1.2 ci-dessus s'écrit ici :

$$\sum_{i=1}^{n} z_i' z_i = \sum_{i=1}^{n} z_i' z_i p_i + \sum_{i=1}^{n} z_i' z_i (1 - p_i)$$

3.2. Cas d'un hasard de base paramétrique : le modèle de Weibull

On a examiné en 2.3 le modèle de Weibull sans variables explicatives ; on souhaite ici généraliser ce modèle dans le cadre d'un modèle à hasard proportionnel. La fonction de hasard de base n'est plus supposée connue et est supposée suivre une loi de Weibull ; elle dépend d'un paramètre, qui devra donc être estimé et le modèle comporte donc un paramètre supplémentaire par rapport à la version précédente.

3.2.1. Présentation générale

Ce modèle est défini par la spécification[19] :

$$h(x|z;\theta,\alpha) = \exp(-z'\theta)\,\alpha x^{\alpha-1}$$

D'après ce qui précède la log-vraisemblance du modèle s'écrit[20] :

$$\ln L(y|z,c;\theta,\alpha) = d\ln(\alpha) + (\alpha-1)\sum_{i=1}^{n} d_i \ln(t_i) - \sum_{i=1}^{n} d_i z_i'\theta - \sum_{i=1}^{n} \exp(-z_i'\theta)\, t_i^{\alpha}$$

où on a noté $d = \sum_{i=1}^{d} d_i$ le nombre de sorties non censurées. Les équations de vraisemblance sont donc :

$$\frac{\partial \ln L(y|z,c;\hat{\theta},\hat{\alpha})}{\partial\theta} = -\sum_{i=1}^{n} d_i z_i' + \sum_{i=1}^{n} z_i' \exp(-z_i'\hat{\theta})\, t_i^{\hat{\alpha}} = 0$$

$$\frac{\partial \ln L(y|z,c;\hat{\theta},\hat{\alpha})}{\partial\alpha} = \frac{d}{\hat{\alpha}} + \sum_{i=1}^{n} d_i \ln(t_i) + \sum_{i=1}^{n} \exp(-z_i'\hat{\theta})\, t_i^{\hat{\alpha}} \ln(t_i) = 0\,.$$

Comme dans le cas où la fonction de hasard de base est connue, la première équation s'interprète comme un produit scalaire, entre les variables explicatives et les résidus généralisés $\tilde{\varepsilon}_i = t_i^{\hat{\alpha}} - d_i \exp(z_i'\hat{\theta})$, comme en 3.1.1 ci-dessus, mais après estimation de la fonction de hasard de base. La seconde équation n'admet pas d'interprétation particulière.

Ces équations doivent être résolues par des méthodes numériques.

Les termes de la matrice d'information de Fisher s'obtiennent en dérivant une seconde fois, et on trouve :

$$\frac{\partial^2 \ln L(y|z,c;\theta,\alpha)}{\partial\theta^2} = \sum_{i=1}^{n} z_i' z_i \exp(-z_i'\theta)\, t_i^{\alpha}$$

$$\frac{\partial^2 \ln L(y|z,c;\theta,\alpha)}{\partial\theta\partial\alpha} = \sum_{i=1}^{n} z_i' \exp(-z_i'\theta)\, t_i^{\alpha} \ln(t_i)$$

$$\frac{\partial^2 \ln L(y|z,c;\hat{\theta},\hat{\alpha})}{\partial\alpha^2} = -\frac{d}{\alpha^2} - \sum_{i=1}^{n} \exp(-z_i'\theta)\, t_i^{\alpha} (\ln(t_i))^2$$

[19] On fixe le paramètre d'échelle de la loi de Weibull à 1.

[20] On pourra rapprocher cette expression de celle établie en 2.3 dans le modèle sans variables explicatives.

3.2.2. Cas particulier du modèle exponentiel

Lorsque dans le modèle précédent le paramètre α est contraint à être égal à 1, on obtient le cas d'une fonction de hasard de base exponentielle, dont on note λ le paramètre (égal à la valeur de la fonction de hasard[21]). En 1.1.1 ci-dessus on a étudié ce cas et montré que l'estimateur du maximum de vraisemblance était égal[22] à $\frac{d}{\sum_{i=1}^{n} d_i t_i + (n-d)c}$.

On prend maintenant comme paramètre $\theta = \frac{1}{\lambda}$; dans le cas non censuré, l'estimateur de θ est la moyenne empirique de l'échantillon, qui est sans biais. En présence de censure, l'estimateur EMV de θ est l'inverse de l'estimateur ci-dessus (par invariance fonctionnelle de l'estimateur du maximum de vraisemblance), $\hat{\theta} = \frac{\sum_{i=1}^{n} d_i t_i + (n-d)c}{d}$, qui est un estimateur biaisé. L'existence de censure introduit donc du biais dans le modèle. On peut montrer[23] que le biais a pour expression :

$$E\left(\hat{\theta}\right) - \theta = \frac{c \exp\left(-\frac{c}{\theta}\right)}{n\left[1 - \exp\left(-\frac{c}{\theta}\right)\right]^2} + O\left(n^2\right),$$

et que la variance asymptotique s'écrit :

$$V\left(\hat{\theta}\right) \approx \frac{\theta^2}{n\left[1 - \exp\left(-\frac{c}{\theta}\right)\right]}.$$

On en déduit l'approximation normale usuelle.

3.3. Cas où la fonction de hasard de base n'est pas spécifiée : le modèle de Cox[24]

On ne suppose plus maintenant de forme particulière pour la fonction de hasard de base ; celle-ci devient alors un paramètre de nuisance, de dimension infinie.

En effet, spécifier complètement un modèle paramétrique peut s'avérer trop restrictif dans certains cas ; de plus, on peut n'être intéressé que par la mesure de l'effet des covariables,

[21] En d'autres termes on réintroduit ici la paramètre d'échelle dont on n'avait pas tenu compte dans le modèle de Weibull.
[22] En supposant les censures toutes égales à *c*.
[23] Voir BARTHOLOMEW [1957] et BARTHOLOMEW [1963].
[24] Pour un traitement détaillé du modèle de Cox on pourra se reporter à DUPUY [2002], dont on reprend ici les notations et la logique de présentation.

et alors la spécification de la fonction de hasard de base n'apporte rien au modèle (à part des contraintes). En d'autres termes, on se place dans un contexte où l'objectif est le positionnement de différentes populations les unes par rapport aux autres, sans considération du niveau absolu du risque. Cela motive l'intérêt pour une spécification partielle, étudiée ici.

On part donc de la formulation suivante :

$$h(x|z;\theta) = \exp(-z'\theta) h_0(x)$$

avec h_0 inconnue.

3.3.1. Estimation des paramètres

Pour mener l'inférence statistique dans ce modèle, COX [1972] a proposé de s'appuyer sur une vraisemblance partielle dans laquelle le paramètre de nuisance h_0 n'intervient pas. Cette approche est un cas particulier d'une démarche plus générale consistant à déterminer une vraisemblance partielle lorsque le modèle contient un paramètre de nuisance de grande dimension. Le principe de cette démarche, décrite dans COX [1975], est présenté ci-après, puis appliqué au cas du modèle de Cox.

On considère ici un vecteur *X* de densité $f_X(x,\beta)$. On suppose qu'il est possible de décomposer *X* en une paire (V,W) telle que :

$$f_X(x,\beta) = f_{W|V}(w|v,\beta) f_V(v,\beta)$$

Un exemple d'une telle décomposition est fourni par le vecteur *V* des valeurs de *X* ordonnées par ordre croissant et *W* le vecteur des rangs. On suppose de plus que le paramètre β est de la forme $\beta = (\theta, h_0)$, θ étant le paramètre d'intérêt. L'idée est que, si, dans la décomposition ci-dessus, l'un des termes de dépend pas de h_0, on peut l'utiliser pour estimer θ. La simplification occasionnée par cette approximation doit compenser la perte d'information.

On rappelle que le modèle de base considéré est toujours le suivant :

$$T_i = X_i \wedge C_i \text{ et } D_i = \begin{cases} 1 \text{ si } X_i \le C_i \\ 0 \text{ si } X_i > C_i \end{cases}$$

avec $h(x|z;\theta) = \exp(-z'\theta) h_0(x)$. D'après l'expression générale de la vraisemblance d'un modèle censuré en présence de covariables (*cf.* 1.2.2 ci-dessus), on peut écrire la vraisemblance complète du modèle de Cox :

$$L(\theta, h_0) = \prod_{i=1}^{n} \left[h_0(t_i) \exp(-\theta' z_i) \exp\left(-H_0(t_i) \exp(-\theta' z_i)\right) \right]^{d_i} \left[\exp\left(-H_0(t_i) \exp(-\theta' z_i)\right) \right]^{1-d_i}$$

Dans l'expression ci-dessus, la fonction de hasard de base intervient de deux manières :

directement, et au travers de la fonction de hasard cumulée H_0. On peut montrer qu'il n'existe pas de maximum à la vraisemblance si on n'impose pas de restriction à la fonction de hasard de base.

En décomposant la vraisemblance de sorte que l'on isole dans un terme que l'on négligera l'incidence de la fonction de hasard de base, on obtient (après une série de développements fastidieux qui ne sont pas repris ici, *cf.* DUPUY [2002]) l'expression suivante de la vraisemblance partielle (valable avec ou sans *ex-æquo*) :

$$L_{Cox}(\theta) = \prod_{i=1}^{n} \left[\frac{\exp(-\theta' z_i)}{\sum_{j=1}^{n} \exp(-\theta' z_j) 1_{\{T_i \leq T_j\}}} \right]^{d_i}$$

On peut toutefois donner une justification heuristique simple de la formule ci-dessous ; on observe en effet que dans le dénominateur de la fraction ci-dessus intervient $R_i = \sum_{j=1}^{n} 1_{\{T_i \leq T_j\}}$, qui n'est autre que l'effectif soumis au risque au moment de la sortie de l'individu *i* (si celle-ci est observée). En conditionnant par les instants de survenance des décès $0 < t_1 \ldots < t_k$ (avec donc $k \leq n$ correspondant aux sorties non censurées), on considère les évènements (ordonnés) suivants : C_i est l'ensemble des censures intervenues entre t_{i-1} et t_i et D_i l'ensemble des sorties non censurées (décès) intervenues en t_i. On se ramène ainsi à un problème d'analyse combinatoire consistant à compter les configurations des sorties conduisant à la séquence observée, les dates de décès étant connues. En d'autres termes, on retrouve ici le fait que l'on n'est pas intéressé par le niveau absolu de la sinistralité, mais simplement par le positionnement des individus les uns par rapport aux autres, en fonction des valeurs prises par les variables explicatives pour chacun d'entre eux. On peut alors décomposer la probabilité d'observer la séquence (C_i, D_i) selon :

$$P\left[(C_i, D_i), 1 \leq i \leq k\right] = P(C_1) P(D_1 | C_1) P(C_2 | C_1, D_1) \ldots P(D_k | C_1 \ldots C_k D_1 \ldots D_{k-1}).$$

En regroupant les évènements relatifs aux décès d'une part et ceux relatifs aux censures d'autre part on met l'expression ci-dessus sous la forme :

$$P\left[(C_i, D_i), 1 \leq i \leq k\right] = \prod_{i=1}^{k} P(D_i | C_1 \ldots C_i D_1 \ldots D_{i-1}) \times \prod_{i=1}^{k} P(C_i | C_1 \ldots C_{i-1} D_1 \ldots D_{i-1})$$

On remarque l'analogie de la formule ci-dessus avec l'expression générale de la vraisemblance donnée *supra*. On peut alors noter que le complémentaire R_i de $\{C_1 \ldots C_i D_1 \ldots D_{i-1}\}$ décrit la population sous risque juste avant l'instant t_i. L'idée de base de la vraisemblance partielle de Cox consiste à ignorer dans la vraisemblance le terme associé aux censures pour ne conserver que :

$$P\left[\left(C_i, D_i\right), 1 \le i \le k\right] \approx \prod_{i=1}^{k} P\left(D_i \,|R_i\right).$$

Il reste à évaluer $P\left(D_i \,|R_i\right)$; on suppose pour simplifier l'absence d'*ex-aequo*, ce qui revient à dire que l'ensemble D_i est un singleton : $D_i = \{j_i\}$. On trouve alors que :

$$P\left(D_i \,|R_i\right) = \frac{h\left(t_i, z_{j_i}\right)}{\sum_{j \in R_i} h\left(t_i, z_j\right)} = \frac{\exp\left(-\theta' z_{j_i}\right)}{\sum_{j \in R_i} \exp\left(-\theta' z_j\right)},$$

ce qui conduit finalement à l'expression cherchée.

L'expression de la vraisemblance partielle se généralise sans difficulté au cas de covariables dépendant du temps ; dans le cas de covariables fixes, on peut montrer (*cf.* FLEMING et HARRINGTON [1991]) que cette expression est égale à la loi du vecteur des rangs associé à $\left(T_1, \ldots, T_n\right)$. En pratique la résolution du système d'équation $\frac{\partial}{\partial \theta_i} \ln L_{Cox}(\theta) = 0$ est effectuée *via* un algorithme numérique (*cf. infra*).

L'intérêt de l'estimateur $\hat{\theta}$ ainsi obtenu est légitimé par le fait qu'il est convergent et asymptotiquement normal, comme un estimateur du maximum de vraisemblance standard[25].

3.3.2. Tests du modèle

Deux types de tests peuvent être menés dans le cadre du modèle de Cox :

- La validation de l'hypothèse de hasard proportionnel ;
- La nullité globale des coefficients, *i.e.* $\theta = 0$.

La validation globale du modèle peut être effectuée en s'appuyant sur un test, dont le principe est étudié en détail par THERNEAU et GRAMBSCH [2000], basé sur les résidus de Schoenfeld. Ces derniers sont définis pour chaque individu *i* et chaque covariable *j* comme la différence entre la valeur, à la date T_i de sortie de *i*, de la covariable pour cet individu, $z_i = \left(z_{i1}, \ldots, z_{ip}\right)$ et sa valeur attendue :

$$r_i = d_i \times \left(z_i - \frac{\sum_{j \in R_i} \exp\left(-\theta' z_j\right) z_j}{\sum_{j \in R_i} \exp\left(-\theta' z_j\right)} \right).$$

En introduisant alors le produit de l'inverse de la matrice de variance-covariance des résidus de Schoenfeld pour l'individu *i* avec le vecteur de ces mêmes résidus, appelé résidu de

[25] Ce résultat est démontré par ANDERSEN et GILL [1982].

Schoenfeld réduit, on peut construire un test de l'hypothèse de hasard proportionnel. Ce test sera étudié en détails ultérieurement.

La nullité globale des coefficients peut être testée *via* un test classique de type Wald ou score[26] (*cf.* la section 4).

4. Les tests fondés sur la vraisemblance

On se propose ici de tester une hypothèse de la forme $g(\theta)=0$, où *g* est une fonction à valeurs dans $\Re^r$, contre l'alternative $g(\theta)\neq 0$. Trois tests asymptotiques faisant appel à la vraisemblance sont classiquement utilisés : le rapport des maxima de vraisemblance, le test de Wald et le test du score. On peut en fait montrer qu'ils sont équivalents, au sens où les statistiques associées diffèrent d'un infiniment petit en probabilité. On choisira donc celui dont la mise en œuvre est la plus simple.

On note $\hat{\theta}$ l'estimateur du maximum de vraisemblance dans le modèle non contraint et $\hat{\theta}^0$ son équivalent dans le modèle contraint. $g(\theta)$ est un vecteur de dimension *r* (une matrice $(r,1)$) et on suppose que la matrice $\frac{\partial g'}{\partial \theta}=\left(\frac{\partial g_j}{\partial \theta_i}\right)$ qui est de dimension (p,r) est de rang *r*.

4.1. Rapport des maxima de vraisemblance

L'idée est ici de comparer les vraisemblances contraintes et non contraintes et d'accepter l'hypothèse nulle si ces 2 valeurs sont proches. On utilise donc la statistique :

$$\xi^R = 2\left(\ln L\left(\hat{\theta}\right) - \ln L\left(\hat{\theta}^0\right)\right)$$

qui converge sous l'hypothèse nulle vers un $\chi^2(r)$, d'où un test dont la région critique est donnée par $W=\left\{\xi^R > \chi^2_{1-\alpha}(r)\right\}$.

4.2. Test de Wald

L'idée du test de Wald est que, si $g\left(\hat{\theta}\right)\approx 0$, alors on accepte l'hypothèse nulle. De manière formelle la statistique :

$$\xi^W = ng'\left(\hat{\theta}\right)\left[\frac{\partial g\left(\hat{\theta}\right)}{\partial \theta'} I\left(\hat{\theta}\right)^{-1} \frac{\partial g'\left(\hat{\theta}\right)}{\partial \theta}\right]^{-1} g\left(\hat{\theta}\right)$$

converge sous l'hypothèse nulle vers un $\chi^2(r)$, d'où un test dont la région critique est donnée par $W=\left\{\xi^W > \chi^2_{1-\alpha}(r)\right\}$.

[26] Voir l'exemple détaillé dans http://larmarange.github.io/analyse-R/analyse-de-survie.html.

4.3. Test du score

On s'intéresse ici à la condition du premier ordre du modèle contraint, qui fait apparaître le Lagrangien $\ln L(\theta) + g'(\theta)\lambda$. La condition du premier ordre s'écrit donc :

$$\frac{\partial \ln L(\hat{\theta}^0)}{\partial \theta} + \frac{\partial g'(\hat{\theta}^0)}{\partial \theta}\hat{\lambda} = 0$$

et on utilise la statistique :

$$\xi^S = \frac{1}{n}\frac{\partial \ln L(\hat{\theta}^0)}{\partial \theta'} I(\hat{\theta}^0)^{-1} \frac{\partial \ln L(\hat{\theta}^0)}{\partial \theta}$$

qui converge sous l'hypothèse nulle vers un $\chi^2(r)$, d'où un test dont la région critique est donnée par $W = \{\xi^S > \chi^2_{1-\alpha}(r)\}$.

5. Illustrations : ajustement de taux de mortalité bruts

Dans ce paragraphe on illustre la mise en œuvre d'une démarche paramétrique dans le cas de la construction d'une table de mortalité. On dispose pour différents âges, $x_0 \le x \le x_1$ d'observations constituées d'une part des effectifs sous risque en début de période[27], notés N_x et, d'autre part, des décès observés pendant la période de référence, D_x. Le nombre de décès à l'âge *x* est une variable aléatoire binomiale de paramètres N_x et q_x, où q_x désigne le taux de mortalité à l'âge *x*. Il est naturel d'estimer ce taux par l'estimateur empirique $\hat{q}_x = \frac{D_x}{N_x}$, qui est sans biais, convergent et asymptotiquement normal[28]. On supposera que l'on dispose de suffisamment de données pour considérer que l'approximation gaussienne est valide. On pourra par exemple utiliser le critère de Cochran, qui consiste à vérifier que $N_x \times \hat{q}_x \ge 5$ et $N_x \times (1-\hat{q}_x) \ge 5$.

D'après ce qui précède, la méthode la plus directe pour estimer les paramètres d'un modèle paramétrique dans ce contexte consiste, une fois la forme de la fonction de hasard fixée, à écrire la log-vraisemblance :

$$\ln L(y_1,\ldots,y_n;\theta) = \sum_{i=1}^{n} d_i \times \ln h_\theta(t_i) + \sum_{i=1}^{n} \ln S_\theta(t_i) - \sum_{i=1}^{n} \ln S_\theta(e_i)$$

puis à résoudre les équations normales $\frac{\partial}{\partial \theta}\ln L(y_1,\ldots,y_n;\theta) = 0$. C'est ce qui a été fait dans l'exemple 1.1.1 ci-dessus. Toutefois, en pratique ces équations peuvent être délicates à

[27] En général la période de temps sera l'année.
[28] En pratique souvent on obtiendra le taux de décès brut dans un cadre non paramétrique (Kaplan-Meier) puis on déduira l'exposition au risque de ce taux et du nombre de décès observés à l'âge considéré.

résoudre. Ainsi, si l'on souhaite utiliser le modèle de Makeham, la log-vraisemblance d'un échantillon censuré[29] a la forme suivante :

$$\ln L(y_1,\ldots,y_n;\theta) = \sum_{i=1}^{n} d_i \times \ln\left(a + b \times c^{t_i}\right) + \sum_{i=1}^{n}\left(-at_i - \frac{b}{\ln(c)}\left(c^{t_i} - 1\right)\right).$$

La résolution du système d'équations $\frac{\partial}{\partial a}\ln L = 0$, $\frac{\partial}{\partial b}\ln L = 0$, $\frac{\partial}{\partial c}\ln L = 0$ est fastidieuse, lorsqu'elle est possible. En effet, d'une part les sommes intervenant dans les expressions ci-dessus comportent potentiellement un très grand nombre de termes. Aussi, on est conduit à proposer une démarche en deux temps :

- ✓ on commence par calculer des taux de décès bruts $\hat{q}_x$ par une méthode intégrant les éventuelles censures (et tenant compte du degré de précision associé aux données individuelles),
- ✓ puis on ajuste dans un second temps le modèle paramétrique retenu à ces taux bruts. Pour cela on utilise la « formule de passage » entre l'expression du modèle à temps continu et les taux bruts suivante :

$$q_x = 1 - \exp\left[-\int_x^{x+1} \mu(y)\,dy\right]$$

Cette relation entre le taux de mortalité discret q_x et la fonction de hasard[30] μ_x exprime simplement le fait que la probabilité de survie entre *x* et *x*+1, conditionnellement au fait que l'individu est vivant à l'âge *x*, est égale à $\frac{S(x+1)}{S(x)}$.

La recherche d'un ajustement est justifiée par le fait que la courbe des taux bruts présente des irrégularités en fonction de l'âge et que l'on peut supposer que ces variations assez brusques ne sont pas dues à des variations de l'incidence réelle du risque, mais à une insuffisance de données. Un ajustement par une fonction modélisant le risque sous-jacent constitue un moyen de lisser ces fluctuations d'échantillonnage[31]. Parmi les lois les plus souvent utilisées figure la loi de Makeham, que l'on appliquera ci-dessous, après avoir présenté l'approche générale.

5.1. Le modèle de Makeham

La loi Makeham vérifie la relation : $\mu_x = a + b \times c^x$ où μ_x représente le taux instantané de décès à l'âge *x*. Le paramètre *a* peut s'interpréter comme une incidence accidentelle ; le coefficient $b \times c^x$, correspondant à un vieillissement de la population, fait croître le taux de

[29] Supposé non tronqué à gauche pour simplifier l'écriture.
[30] La fonction de hasard *h* est traditionnellement notée μ en démographie.
[31] Pour des arguments plus développés, voir le support sur les « lissages et ajustements ».

décès de manière exponentielle. Compte tenu de la croissance des taux de décès avec l'âge, on doit avoir une constance *c* supérieure à 1 et un *b* positif. On a alors :

$$p_x = \exp\left[-\int_x^{x+1} \mu_y dy\right] = \exp\left[-\int_x^{x+1} \left(a + b \times c^y\right) dy\right] = \exp(-a) \exp\left[-\frac{b}{\ln(c)} c^x (c-1)\right].$$

Posons $s = \exp(-a)$ et $g = \exp\left(-\frac{b}{\ln(c)}\right)$, la fonction utilisée pour l'ajustement des taux de décès discrets est donc :

$$q_x = 1 - p_x = 1 - s \times g^{c^x(c-1)}.$$

C'est sur la base de cette version discrétisée du modèle que nous allons dorénavant nous appuyer.

5.1.1. Adéquation de la courbe au modèle de Makeham

Avant de réaliser l'ajustement proprement dit, on cherche à valider l'adéquation de ce type de fonction à la situation proposée. Pour cela on observe que l'on a $\ln(1-q_x) = \ln(s) + c^x(c-1)\ln(g)$. Pour les q_x proches de zéro[32], on peut faire l'approximation $\ln(1-q_x) \approx -q_x$, et donc :

$$-q_x = \ln(s) + c^x(c-1)\ln(g)$$

Il en résulte que $q_x - q_{x+1} = c^x(c-1)^2 \ln(g)$, ce qui conduit à remarquer en prenant le logarithme de cette expression que :

$$\ln(q_{x+1} - q_x) = x \ln(c) + \ln\left((c-1)^2 \ln\left(\frac{1}{g}\right)\right).$$

Sous l'hypothèse que les taux de mortalité suivent une loi de Makeham, les points $(x, y = \ln(q_{x+1} - q_x))$ sont donc alignés sur une droite de pente $\ln(c)$. L'idée est donc de faire une régression linéaire et de produire une analyse de la régression sur le modèle suivant :

Tab. 3. Analyse de variance

	Degré de liberté	Somme des carrés	Moyennes des carrés	F	Valeur critique de F
Régression	1				
Résidus	*n*-1				
Total	*n*				

Tab. 1 - Analyse de variance

[32] On peut retenir que le taux de mortalité à 60 ans est en France de l'ordre de 0,50% pour les femmes, et de 1,20% pour les hommes (source : TV/TD 99/01).

On conclut éventuellement à l'ajustement par une droite sur l'intervalle $x_0 \le x \le x_1$ en effectuant un test de Fisher (à un seuil à définir, par exemple 5%). On rappelle que la statistique de test de Fisher utilisée pour tester la significativité globale d'un modèle de régression linéaire[33] $y_i = \beta_0 + \beta_1 x_1 + .. + \beta_{p-1} x_{p-1} + \varepsilon_i$ est $F_{p-1} = \frac{R^2}{1-R^2} \frac{n-p}{p-1}$ avec $R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$. Cette statistique suit une loi de Fisher $(1, p-1)$.

5.1.2. Ajustement par la méthode du maximum de vraisemblance

Une fois validé le fait qu'un ajustement de type Makeham peut s'avérer pertinent, on cherche à en estimer les paramètres par la méthode du maximum de vraisemblance. On notera incidemment que le maximum de vraisemblance déterminé dans le modèle discrétisé étudié ici n'est pas identique au maximum de vraisemblance direct que l'on obtient à partir du modèle de base continu.

On note $\theta = (s, g, c)$ le vecteur des paramètres à déterminer et $q_x(\theta) = 1 - s \times g^{c^x (c-1)}$ la fonction de Makeham à ajuster. On cherche le vecteur de paramètre qui donne la fonction ajustant au mieux la courbe des $\hat{q}_x$ (taux d'incidence bruts observés).

On peut également simplement utiliser le solveur d'Excel. Dans tous les cas, l'algorithme ne converge vers la vraie valeur du paramètre qu'à la condition de partir d'une valeur initiale θ_0 assez proche de θ.

Il convient donc de déterminer des valeurs initiales acceptables des paramètres. On peut utiliser pour cela la propriété établie en 5.1.1 ci-dessus sur l'alignement des points $(x, y = \ln(q_{x+1} - q_x))$; l'ordonnées à l'origine et la pente de la droite déterminent g et c, et on peut trouver à partir de la relation $\ln(p_x) - c^x (c-1) \ln(g) = \ln(s)$[34].

Afin de tester si les coefficients de la fonction de Makeham ainsi déterminés ne sont pas significativement égaux à zéro, on effectue un test de Student qui consiste à comparer le ratio (estimation/écart type) à une loi de Student à *m* degrés de liberté (*m* = nombre d'âges observés - 3 paramètres estimés). On réalise enfin des tests du Khi-2, sur la base de la statistique $W = \sum N_x \frac{(\hat{q}_x - q_x)^2}{q_x}$, q_x étant le taux de décès théorique du modèle à l'âge *x*. La loi asymptotique de *W* est une loi $\chi^2(p - 3 - 1)$, où *p* désigne le nombre d'âges intervenants dans la somme. Il convient en pratique de manipuler avec précaution le test

[33] C'est à dire pour valider le fait que les coefficients de régression soient non tous nuls.
[34] Le membre de gauche de l'égalité ne doit donc que peu dépendre de *x*.

du Khi-2, la loi asymptotique n'étant un $\chi^2(p-k-1)$, *p* étant le nombre de classes et *k* le nombre de paramètres du modèle que parce qu'ici l'estimateur est du maximum de vraisemblance. Pour d'autres méthodes de détermination du paramètre, ce résultat n'est plus vrai en général (voir FISCHER [1924]).

Le graphique suivant reprend l'ajustement Makeham réalisé par pseudo-maximum de vraisemblance (en normant les effectifs sous risque à chaque âge) sur la tranche d'âges 40-105 ans de la TF 00-02.

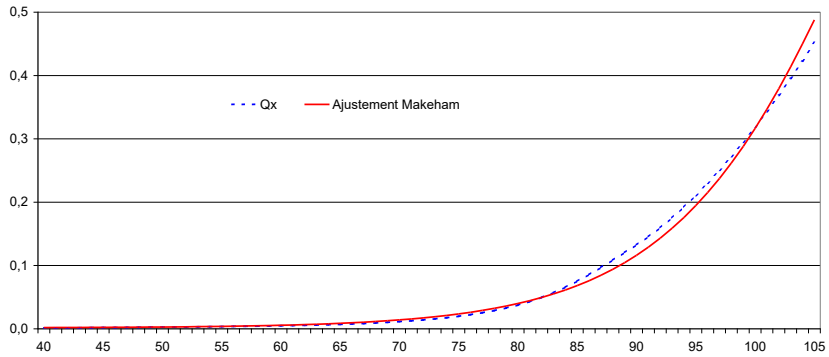


Figure 1 - *Ajustement de la TH00-02 à une loi de Makeham*

5.2. Le modèle de Thatcher

En pratique, le modèle de Makeham conduit à une surestimation des taux de décès conditionnels aux âges élevés. Afin de corriger cet surestimation, THATCHER [1999] proposé un modèle proche en posant $\mu(t) = \alpha + \frac{\beta \times e^{\gamma t}}{1+\beta \times e^{\gamma t}}$. En posant $v_{\beta,\gamma}(u) = 1+\beta \exp(\gamma u)$ on remarque que $\frac{\beta \exp(\gamma u)}{1+\beta \exp(\gamma u)} du = \frac{1}{\gamma}\frac{dv}{v}$, ce qui conduit après quelques manipulations à $S_\theta(t) = e^{-\alpha t} v_{\beta,\gamma}(t)^{-\frac{1}{\gamma}}$. On en déduit notamment :

$$E(T_\theta) = \int_0^{+\infty} e^{-\alpha t} v_{\beta,\gamma}(t)^{-\frac{1}{\gamma}} dt = \int_0^{+\infty} e^{-\alpha t} \left(1+\beta e^{\gamma t}\right)^{-\frac{1}{\gamma}} dt .$$

Il reste à calculer $q_x = 1 - \exp\left(-\int_x^{x+1} \mu(y)\, dy\right)$, qui conduit à :

$$q_x = 1 - e^{-\gamma}\left(\frac{v_{\beta,\gamma}(x+1)}{v_{\beta,\gamma}(x)}\right)^{-\frac{1}{\beta}} .$$

On obtient des ajustements proches de ceux obtenus avec le modèle de Makeham, mais avec des taux légèrement plus faibles :

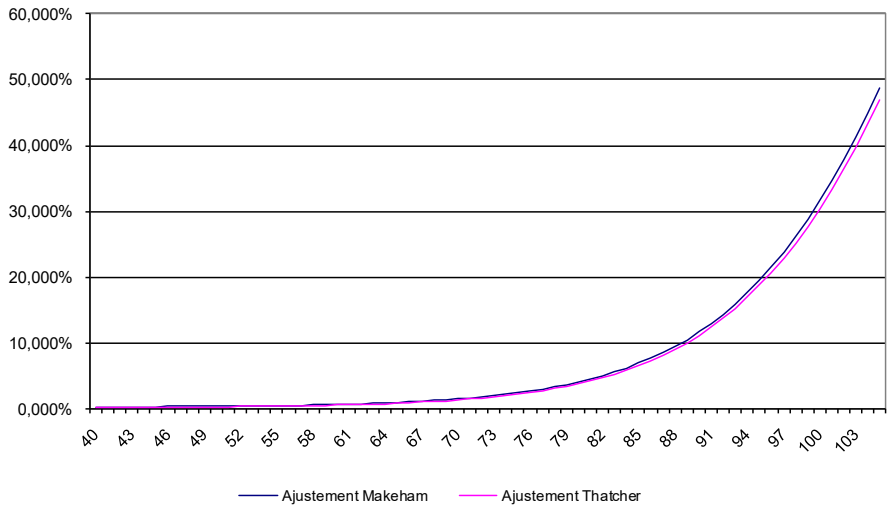


Figure 2 - *Comparaison des ajustements Makeham et Thatcher de la TH00-02*

5.3. Ajustement des taux bruts sur la base des Logits

L'estimation des taux de mortalité q_x est contrainte par le fait que l'on doit avoir $q_x \in [0,1]$; en posant $\mathbf{lg}(x) = \mathbf{ln}\left(q_x / (1-q_x)\right)$, le logit du taux de décès, on est ramené à une valeur « libre » dans $]-\infty,+\infty[$ et on peut alors utiliser les techniques de régression linéaire sur des variables explicatives. Les variables explicatives candidates les plus simples peuvent être l'âge et le logit des taux de décès d'une table de référence.

5.3.1. La fonction logistique

La fonction logistique est par définition $\mathbf{lg}(x) = \mathbf{ln}\left(\frac{x}{1-x}\right)$ est définie sur $]0,1[$; elle est croissante sur cet intervalle :

$$\frac{d}{dx}\mathbf{lg}(x) = \frac{1}{x(1-x)}.$$

On a par ailleurs :

$$\frac{d^2}{dx^2}\mathbf{lg}(x) = \frac{2x-1}{x^2(1-x)^2}.$$

Sur l'intervalle $]0,1/2[$ la fonction $\mathbf{lg}(x)$ est donc concave. Rappelons que selon l'inégalité de Jensen, si *f* est convexe, alors $\mathbf{E}\, f(X) \geq f\left(\mathbf{E}(X)\right)$. On en déduit que, dans une zone où les taux de décès sont petits, et si l'on a estimé le taux de décès par $\hat{q}_x$ supposé dans biais, alors :

$$\mathbf{E}\,\mathbf{lg}(\hat{q}_x) \leq \mathbf{lg}(q_x).$$

En d'autres termes, les logits empiriques ainsi obtenus sont biaisés négativement (ils sous-estiment les vrais logits). Comme la fonction $\mathbf{lg}(x)$ (et son inverse) est croissante, en sous-estimant les logits théoriques, cette démarche sous-estime les taux de décès théoriques. La conclusion est inverse pour des taux de sortie supérieurs à ½.

Dans le cadre d'un ajustement des $\hat{y}_x = \mathbf{lg}(\hat{q}_x)$, on obtient les taux de décès ajustés par la transformation inverse $y \to \frac{e^y}{1+e^y}$. La présence d'exponentielles dans cette expression conduit à une amplification importante du biais d'estimation évoqué ci-dessus. Ainsi, dans le cas d'un risque décès, un modèle d'ajustement des logits des taux de décès conduit à sous-estimer dans des proportions qui peuvent être importantes (typiquement de 5 % à 10 %) les taux de décès.

Les modèles utilisant les logits des taux de décès doivent donc être utilisés avec prudence dans le cas d'un risque en cas de décès. Ils intègrent au contraire une marge de sécurité dans le cas d'un risque en cas de vie.

L'utilisation des régressions logistiques dans le cadre de variables qualitative est de plus « légitimée » par la remarque suivante : la quantité $c_x = \frac{q_x}{1-q_x}$ est le rapport de la probabilité de « succès » à la probabilité d'« échec » dans le cadre d'une expérience de Bernoulli ; cette grandeur s'interprète donc en disant qu'il y a « c_x fois plus de chances que le décès survienne qu'il ne survienne pas ». Il est alors relativement naturel de chercher à expliquer le niveau atteint par c_x à l'aide de variables explicatives, et du fait de la positivité de c_x le modèle le plus simple que l'on puisse imaginer est obtenu en posant $c_x = \mathbf{exp}\left({}^t\theta z_x\right)$, avec z_x le vecteur des variables explicatives.

On se trouve alors dans le contexte d'un modèle linéaire généralisé[35] avec une fonction de lien logistique :

$$\mathbf{lg}(q_x) = {}^t\theta z_x + \varepsilon_x,$$

ce qui permet d'utiliser les procédures standards d'estimation disponibles dans la plupart des logiciels spécialisés (une fois spécifiée la loi de ε_x). On peut également noter que ce modèle peut s'écrire sous la forme :

$$q_x(\theta) = \frac{e^{\theta' z_x}}{1+e^{\theta' z_x}}.$$

On peut donc rechercher la solution par la méthode décrite ci-dessus de maximum de vraisemblance discret.

[35] Voir NELDER et WEDDERBURN [1972] pour la présentation originale et PLANCHET et *al.* [2011] pour une introduction.

5.3.2. Ajustements logistiques

Le modèle de base d'ajustement logistique part du constat que sur une large plage le logit des taux de décès présente une tendance linéaire ; on propose alors la modélisation suivante, version la plus simple du module présenté *infra* si l'on suppose que l'âge constitue une variable explicative pertinente :

$$\mathbf{lg}\left(\hat{q}_x\right) = a + bx + \varepsilon_x$$

où ε est un bruit gaussien iid ; on régresse donc simplement les logits des taux de décès sur l'âge. La transformation inverse du logit étant $y \to \frac{e^y}{1+e^y}$, le modèle $\mathbf{lg}\left(q_x\right) = a + bx$ s'écrit de manière équivalente :

$$q_x = \frac{ce^{dx}}{1+ce^{dx}}$$

en posant $c = e^a$ et $d = b$. Une approche alternative à la régression linéaire $\mathbf{lg}\left(\hat{q}_x\right) = a + bx + \varepsilon$ consiste donc à effectuer une estimation par maximum de vraisemblance dans le modèle paramétrique $q_x = \frac{ce^{dx}}{1+ce^{dx}}$. Cette approche évite *a priori* l'effet de sous-estimation des taux de mortalité associée à l'approche par régression linéaire, le taux de décès étant la variable modélisée (mais l'estimateur du maximum de vraisemblance n'a toutefois pas de raison d'être sans biais).

La détermination de la fonction de survie et de la fonction de hasard, liées l'une à l'autre part la relation $S(t) = \mathbf{exp}\left(-\int_0^t \mu(s)ds\right)$ nécessite de faire des hypothèses. En effet, la relation $q(x) = 1 - \frac{S(x+1)}{S(x)}$ conduit dans le cas général à la contrainte sur la fonction de hasard :

$$-\mathbf{ln}\left(1-q_x\right) = \int_x^{x+1} \mu(s)ds$$

Dans le modèle discret spécifié jusqu'alors *x* est *a priori* entier. Il faut donc une règle de passage du temps discret au temps continu. On peut utiliser différentes approches (Balducci, constance des taux de hasard par morceau, *etc.*). Si on choisit l'hypothèse de constance de la fonction de hasard entre deux valeurs entières, on trouve que la fonction de hasard est une fonction en escalier avec aux points entiers :

$$\mu_x = \frac{cde^{dx}}{1+ce^{dx}}.$$

En pratique il peut apparaître que la courbe des taux bruts $\hat{q}_x$ présente un décrochage à partir d'un âge pivot qui indique une accélération de l'incidence. Dans ce contexte, on est amené à rechercher un ajustement via des modèles de type logistique construits sur des ajustements de $\ln\left(\hat{q}_x/\left(1-\hat{q}_x\right)\right)$ sur l'âge, qui jouera donc le rôle de variable explicative.

On cherche à ajuster les taux bruts sur une fonction de la forme :

$$\ln\left(\hat{q}_x/\left(1-\hat{q}_x\right)\right) = ax+b+c\times\left(0\vee\left(x-x_c\right)\right)$$

où x_c est un « âge charnière » au-delà duquel la mortalité s'accélère (modèle *logit* standard). En d'autres termes, on écrit le modèle de régression logistique suivant :

$$\ln\left(\hat{q}_x/\left(1-\hat{q}_x\right)\right) = ax+b+c\times 0\wedge\left(x-x_c\right)$$

où les (ε_X) forment un bruit blanc gaussien. On peut généraliser ces modèles en écrivant :

$$\ln\left(q_x/\left(1-q_x\right)\right) = ax+b+c\times 0\wedge\left(x-x_c\right)^{\lambda}+\varepsilon_x$$

Si on ne dispose pas de données suffisantes pour structurer correctement la table complète, on peut imaginer d'utiliser la structure d'une table de référence existante et de simplement positionner la mortalité du groupe considéré par rapport à cette référence. Lorsque l'on souhaite positionner une table par rapport à une autre, il peut apparaître naturel d'effectuer la régression des logits des taux bruts sur les logits de la table de référence, ce qui conduit au modèle suivant :

$$\ln\left(\hat{q}_x/\left(1-\hat{q}_x\right)\right) = a\ln\left(q_x^{ref}/\left(1-q_x^{ref}\right)\right)+b+\varepsilon_x$$

$$\ln\left(q_x/\left(1-q_x\right)\right) = a\ln\left(q_x^{ref}/\left(1-q_x^{ref}\right)\right)+b$$

5.3.3. Estimation des paramètres

Dans le cas du modèle de régression sur l'âge, l'estimation peut être effectuée selon la procédure suivante : avant l'âge charnière x_c, on effectue une régression linéaire de $\ln\left(\hat{q}_x/\left(1-\hat{q}_x\right)\right)$ sur *x*, puis au-delà on fait une seconde régression (non linéaire) de $\ln\left(\hat{q}_x/\left(1-\hat{q}_x\right)-\left(ax+b\right)\right)$.

Dans le cas d'une régression des logits des taux bruts sur les logits d'une table de référence, l'estimation est une estimation des moindres carrés ordinaires classique.

5.4. Intervalles de confiance pour les taux bruts

La première étape de la construction de la table de mortalité est constituée par l'estimation des taux bruts à chaque âge. Il convient, au-delà de l'estimation ponctuelle, d'avoir une idée de la précision de l'estimation effectuée. Celle-ci dépend de deux facteurs :

- ✓ l'effectif sous risque, N_x,
- ✓ le niveau du taux de mortalité à estimer, q_x.

En effet, la précision sera d'autant meilleure que N_x est grand, et que q_x est grand. La précision sera mesurée par la largeur de l'intervalle de confiance. Pour déterminer celui-ci, deux méthodes sont possibles :

- ✓ l'utilisation de l'approximation gaussienne, si l'on dispose de suffisamment d'observations ;
- ✓ le calcul de l'intervalle à distance finie, qui est *a priori* possible puisque la loi de $\hat{q}_x$ est connue.

Dans un premier temps, on cherche donc quel type d'intervalle de confiance utiliser. Pour cela on remarque qu'une relation lie l'incertitude de l'estimation, le nombre d'observations et le niveau de confiance de l'intervalle désiré :

$$\Delta p = u_{a/2}\sqrt{\frac{f(1-f)}{N}}$$

où *f* est la valeur autour de laquelle est construit l'intervalle. (*i.e. f* est égale à la valeur estimée $\hat{q}_x$) et u_p désigne le quantile d'ordre *p* de la loi normale centrée réduite.

Exemple

Si la valeur à estimer q_x vaut 0,2, si l'on souhaite un intervalle à 95 % pour une précision de l'ordre de 0,01. Il est nécessaire de disposer de :

$$N_x = \frac{f(1-f)}{\Delta p^2} u_{a/2}{}^2 = \frac{0,2 \times 0,8}{0,01^2} \times 1,96^2$$

soit environ :

$$N_{\approx} \approx 6\,150$$

Si l'on ne dispose que de 3 000 observations on se tournera vers l'intervalle de confiance à distance finie. Au niveau de 95 %, en se plaçant dans le cas le plus défavorable d'une fréquence égale à ½, on obtient un majorant (assez large) du nombre d'observations nécessaires à l'obtention de la précision Δp par $N = \frac{1}{\Delta p^2}$.

5.4.1. Intervalles de confiance asymptotiques

N_x désigne l'exposition au risque à l'âge *x* D_x le nombre de décès dans l'année des personnes d'âge *x*, et on a estimé q_x par $\hat{q}_x$. D'après le théorème central-limite :

$$\sqrt{N_x}\frac{q_x - \hat{q}_x}{\sqrt{\hat{q} \times (1-\hat{q})}} \underset{N \to \infty}{\to} N(0,1)$$

L'intervalle de confiance asymptotique de niveau α pour q_x est donc donné par :

$$I_\alpha = \left[\hat{q}_x - u_{\alpha/2} \sqrt{\frac{\hat{q}_x(1-\hat{q}_x)}{N_x}} , \hat{q}_x + u_{\alpha/2} \sqrt{\frac{\hat{q}_x(1-\hat{q}_x)}{N_x}} \right]$$

La limite de cette approche est qu'elle ne permet de construire que des intervalles de confiances ponctuels, pour un âge fixé, mais ne permet pas d'encadrer les taux de décès sur une plage d'âges fixées à un niveau de confiance connu. On souhaite désormais encadrer les taux de décès simultanément sur tous les âges *x* d'une plage d'âges $[x_0, x_0+n]$ (où *n* est un nombre entier positif). L'encadrement des taux de décès correspond donc désormais à une bande de confiance, et non plus à un intervalle de confiance ponctuel.

On souhaite ici construire des bandes de confiance pour les taux de décès, et non pour des fonctions de survie. En pratique, on cherche ainsi $t(\hat{q}_x)$ tel que $P\left(q_x \in \hat{q}_x \pm t(\hat{q}_x), \forall x \in [x_0, x_0+n]\right) = 1-\alpha$. À cet effet, on s'appuie sur la méthode d'estimation de Sidak, qui repose sur le principe d'inflation du seuil du test lorsque le nombre de tests augmente (*cf.* par exemple ABDI [2007]).

Pour mémoire, une bande de confiance au niveau de confiance $1-\alpha$ sur la plage d'âges $[x_0, x_0+n]$ peut être présentée comme une collection d'intervalles de confiance pour les différents âges $x \in [x_0, x_0+n]$ construits de manière à avoir un intervalle simultané de probabilité égal à $1-\alpha$. Soit donc $P\left(q_x \in \hat{q}_x \pm t(\hat{q}_x), x = x_0\right) = 1-\beta$ l'intervalle de probabilité de niveau $1-\beta$ (avec $\beta \in]0,1[$) pour q_x à l'âge $x = x_0$. La probabilité simultanée d'encadrer les taux de décès q_x aux deux âges $x = x_0$ et $x = x_0+1$ est alors $(1-\beta)^2$, en supposant l'encadrement indépendant sur ces deux âges. En répétant l'opération de manière à inclure tous les âges de $[x_0, x_0+n]$, il apparaît alors, toujours sous l'hypothèse d'indépendance, que la probabilité simultanée d'encadrer les taux de décès q_x pour les différents âges $x \in [x_0, x_0+n]$ est $(1-\beta)^{n+1}$.

Sur ces bases, on peut ainsi construire une bande de confiance au seuil α sur la tranche d'âges $[x_0, x_0+n]$, en constituant des intervalles de confiance ponctuels pour chaque âge $x \in [x_0, x_0+n]$ au seuil :

$$\beta = 1-(1-\alpha)^{1/(n+1)},$$

puisque dans ce cas on a bien $(1-\alpha) = (1-\beta)^{n+1}$. Aussi, une approximation de la bande de confiance permettant d'encadrer simultanément les taux de décès sur tous les âges $[x_0, x_0+n]$ à partir de la méthode de Sidak est :

$$P\left(q_x \in \hat{q}_x \pm u_{\beta/2} \sqrt{\frac{\hat{q}_x(1-\hat{q}_x)}{R_x}}, \forall x \in [x_0, x_0+n] \right) = 1-\alpha,$$

avec $\beta = 1-(1-\alpha)^{1/(n+1)}$. Les intervalles et bandes de confiance ci-dessus permettent d'encadrer les taux de décès bruts au titre des fluctuations d'échantillonnage, respectivement pour un âge donné ou sur une tranche d'âges. Les bandes de confiance sont par construction plus larges que les intervalles de confiance.

5.4.2. Intervalles de confiance à distance finie

Ici on considère le cas où N_x n'est pas assez grand pour pouvoir utiliser le théorème de la limite centrale, on s'appuie sur le fait que $P(D_x = k) = C_{N_x}^k \times q_x^k \times (1-q_x^k)$ pour calculer l'intervalle de confiance exacte à distance finie. On cherche donc m_α tel que :

$$P[\hat{q}_x - m_\alpha \le q_x \le \hat{q}_x + m_\alpha] \ge 1-\alpha$$

En multipliant par N_x les termes de l'inégalité dont on veut calculer la probabilité on trouve que l'on doit avoir :

$$\sum_{k=\left\lfloor N(\hat{q}_x - m_a)\right\rfloor}^{\left\lfloor N(\hat{q}_x + m_a)\right\rfloor + 1} P[D_x = k] \ge P[\hat{q}_x - m_\alpha \le q_x \le \hat{q}_x + m_\alpha] \ge 1-\alpha$$

On peut imaginer une procédure itérative pour trouver m_α :

Étape n°o

On calcule $P(D_x = k)$ avec $k = N_x \hat{q}_x$ que l'on compare à $1-\alpha$, et si $P(D_x = k) < 1-\alpha$, on passe à l'étape suivante.

Étape n°j

On calcule $\sum_{k=N\hat{q}_x - j}^{N\hat{q}_x + j} P[D_x = k]$ que l'on compare à $1-\alpha$. Si $\sum_{k=N\hat{q}_x - j}^{N\hat{q}_x + j} P[D_x = k] < 1-\alpha$, on passe à l'étape j+1.

Étape finale

Lorsque ce processus itératif s'arrête on pose $m_\alpha = \frac{j}{N_x}$.

6. Références

ABDI H. [2007] « Bonferroni and Sidak corrections for multiple comparisons », *N. J. Salkind (ed.). Encyclopedia of Measurement and Statistics*, Thousand Oaks, CA: Sage.

ANDERSEN P.K, GILL R.D. [1982], « Cox's regression model for counting processes : a large sample study », *The Annals of Statistics*, 10, 1100-1120.

BARTHOLOMEW D. [1957], « A problem in life testing », *J. Amer. Statist. Assoc.*, 52, 350-355

BARTHOLOMEW D. [1963], « The sampling distribution of an estimate arising in life testing », *Technometrics*, 5, 361-374.

CIARLET P.G. [1990] *Introduction à l'analyse numérique matricielle et à l'optimisation*. Paris : Dunod.

COX D.R. [1972] « Regression models and life-tables (with discussion) ». *J. R. Statist. Soc. Ser. B*, pages 187-220.

COX D.R. [1975] « Partial likelihood ». *Biometrika*, 62, pages 269-276.

DEMPSTER, A. P., LAIRD, N. M. RUBIN, D. B. [1977], « Maximum likelihood from incomplete data via the EM algorithm ». *Journal of the Royal Statistical Society*, B, 39, 1-38.

DROESBEKE J.J., FICHET B., TASSI P. [1989], *Analyse statistique des durées de vie*, Paris : Economica

DUPUY J.F. [2002], « Modélisation conjointe de données longitudinales et de durées de vie », Université Paris V, Thèse de doctorat.

FISHER R.A. [1924], « The conditions under wich χ^2 mesasures the discrepancy between observation and hypothesis ». *J. Roy. Statist. Soc.*, 87, 442-450.

FLEMING T.R., HARRINGTON D.P. [1991] *Counting processes and survival analysis*, Wiley Series in Probability and Mathematical Statistics. New-York : Wiley.

HILL C., COM-NOUGUÉ C., KRAMAR A., MOREAU T., O'QUIGLEY J., SENOUSSI R., CHASTANG Cl. [2000] *Analyse Statistique des Données de Survie*, Paris : Flammarion Sciences

NELDER J., WEDDERBURN R. [1972] « Generalized linear models », *Journal of Roy. Stat. Soc. B*, vol. 135, 370-384.

JUILLARD M., PLANCHET F., THÉROND P.E. [2011] *Modèles financiers en assurance. Analyses de risques dynamiques - seconde édition revue et augmentée*, Paris : Economica (première édition : 2005).

ROBERT C.P. [1996] *Méthodes de Monte-Carlo par chaînes de Markov*, Paris : Economica

SAPORTA G., [1990] *Probabilités, analyse des données et statistiques*, Paris : Editions Technip

THATCHER A.R. [1999] « *The Long-term Pattern of Adult Mortality and the Highest Attained Age.* », Journal of the Royal Statistical Society 162 Part 1: 5-43.

THERNEAU T. M., GRAMBSCH P. M., FLEMING T. R. [1990] « Martingale-based residuals for survival models », Biometrika, vol. 77, n°1, pp. 147-160.

THERNEAU T. M., GRAMBSCH P. M. [1990] *Modeling Survival Data: Extending the Cox Model*, Series: Statistics for Biology and Health, New-York: Springer