
Travaux dirigés : modèles de durée
Séance n°3 - Corrigé

Exercice 1 Modèle binomiale et ajustement logistique.

On considère un modèle pour la durée de la vie humaine en temps discret pour les âges entiers $x \in \{x_{\min}, \dots, x_{\max}\}$. On suppose que le nombre de décès D_x à l'âge x suit une loi Binomiale $\mathcal{B}(n_x, q_x)$, où n_x correspond à l'effectif observé à l'âge x .

1. Pour une population donnée où n_x et d_x peuvent être calculés, rappeler l'estimateur du maximum de vraisemblance de q_x . Ces estimateurs des probabilités conditionnelles \hat{q}_x sont généralement appelés probabilités de décès brutes.
2. On se place à présent dans le cadre d'un modèle linéaire généralisé, et on suppose que q_x s'écrit tel que

$$q_x = \frac{\exp(\eta_x)}{1 + \exp(\eta_x)},$$

avec le prédicteur linéaire $\eta_x = \sum_{s=0}^p \beta_s x^s$, où $\beta = (\beta_0, \dots, \beta_p)$ est un vecteur de paramètres à estimer.

Écrire la vraisemblance du modèle et indiquer comment obtenir un estimateur du vraisemblance pour β .

3. Comment contrôler la qualité de l'ajustement réalisé.
4. Une autre spécification classique consiste à analyser le modèle de régression suivant pour chaque âge

$$y_x = \sum_{s=0}^p \alpha_s x^s + \epsilon_x,$$

avec ϵ_x un bruit blanc et $y_x = \text{logit}(\hat{q}_x)$. Comparer le biais des estimateurs obtenus avec cette approche. Commenter ce résultat.

Réponse de l'exercice 1.

1. L'estimateur "naturel" des taux bruts est

$$\hat{q}_x = \frac{d_x}{n_x}.$$

On retrouve aisément ce résultat en considérant l'estimateur du maximum de vraisemblance du paramètre de fréquence d'une loi binomiale.

2. En utilisant des notations évidentes, la vraisemblance du modèle s'écrit en regroupant par âge entier

$$\mathcal{L}(\beta) = \prod_{x=x_{\min}}^{x_{\max}} \binom{n_x}{d_x} q_x^{d_x} (1 - q_x)^{n_x - d_x}$$

Ainsi, la log-vraisemblance du modèle est égale (à une constante près) à

$$\sum_{x=x_{\min}}^{x_{\max}} d_x \eta_x - n_x \ln(1 + \exp(\eta_x)).$$

En dérivant par rapport à β_s , le système à résoudre pour obtenir les paramètres du modèle est, pour tout $s = 0, \dots, p$

$$\sum_{x=x_{\min}}^{x_{\max}} x^s \left(d_x - n_x \frac{\exp(\eta_x)}{1 + \exp(\eta_x)} \right) = \sum_{x=x_{\min}}^{x_{\max}} x^s (d_x - q_x n_x) = 0,$$

La première relation pour $s = 0$ permet d'assurer la reproduction du nombre total de décès par le modèle, i.e. $\sum_{x=x_{\min}}^{x_{\max}} d_x = \sum_{x=x_{\min}}^{x_{\max}} \hat{q}_x n_x$. En posant le résidu

$$r_x = d_x - n_x \frac{\exp(\eta_x)}{1 + \exp(\eta_x)},$$

les équations suivantes apparaissent comme une relation d'orthogonalité entre les résidus et les variables x^s . Ces équations peuvent être résolues numériquement en appliquant un algorithme de Newton-Raphson.

3. Comme pour un modèle linéaire généralisé classique, la qualité du modèle peut être analysée par le biais de :

- la déviance (avec $\ln(\mathcal{L}(d))$, la log-vraisemblance du modèle saturé, i.e. pour lequel les observations sont exactement prédites, et $\ln(\mathcal{L}(\hat{d}))$, la log-vraisemblance du modèle ajusté)

$$\begin{aligned} \text{Dev}(d, \hat{d}) &= \sum_{x=x_{\min}}^{x_{\max}} \text{Dev}_x(d, \hat{d}) = 2 \left(\ln(\mathcal{L}(d)) - \ln(\mathcal{L}(\hat{d})) \right) \\ &= 2 \sum_{x=x_{\min}}^{x_{\max}} d_x \ln \left(\frac{d_x}{n_x \hat{q}_x} \right) - (n_x - d_x) \ln \left(\frac{n_x - d_x}{n_x - n_x \hat{q}_x} \right) \end{aligned}$$

- un test d'adéquation du chi-2 de Pearson

$$\chi^2 = \sum_{x=x_{\min}}^{x_{\max}} \frac{(d_x - n_x \hat{q}_x)^2}{n_x \hat{q}_x (1 - \hat{q}_x)}$$

- le pseudo- R^2

$$\text{pseudo-}R^2 = 1 - \frac{\sum_{x=x_{\min}}^{x_{\max}} (d_x/n_x - \hat{q}_x)^2}{\sum_{x=x_{\min}}^{x_{\max}} d_x/n_x - \left(n^{-1} \sum_{x=x_{\min}}^{x_{\max}} d_x/n_x \right)^2}.$$

4. Cette approche correspond à une régression linéaire effectuée sur le logit des taux bruts, alors que le premier modèle est un modèle linéaire généralisé binomial avec un lien logit. L'estimateur obtenu avec cette approche est asymptotiquement sans biais, alors que l'approche s'appuyant sur une régression linéaire fait apparaître systématiquement un biais conduisant à sous estimer les décès théoriques. Pour montrer ce résultat, il faut appliquer l'inégalité de Jensen, en rappelant que la fonction logit est concave pour des $q_x < 0,5$ (cf. cours sur

les statistique des modèles paramétriques et semi-paramétriques, section 5.4). En effet, on a (puisque les estimateurs \hat{q}_x sont sans biais)

$$\hat{y}_x = \mathbb{E}[\text{logit}(\hat{q}_x)] \leq \text{logit}(\mathbb{E}[\hat{q}_x]) = \text{logit}(q_x).$$

Les modèles utilisant les logits des données de mortalité doivent être utilisés avec prudence dans le cas d'un contrat où le risque assuré est un risque décès puisqu'ils sous-estiment les taux de mortalité, parfois avec des proportions importantes. Cependant, ils entraînent une marge de sécurité dans le cas du risque de survie.

Remarque 1. *Le prédicteur linéaire est ici constitué d'un polynôme de degré quelconque, fonction de l'âge. Dans les modèles relationnels, les q_x d'une table de référence (information exogène) sont ajoutés et constituent des covariables supplémentaires du modèle. Le modèle de Brass par exemple est un modèle de régression linéaire où les logit des probabilités de décès pour la population d'intérêt dépendent des logits des probabilités de décès pour une population de référence.*

Remarque 2. *Dans cet exercice, nous avons considérer des probabilités conditionnelles de décès. Il aurait été possible de travailler avec les fonctions de hasard (modèle GLM Poisson). Une approche en deux dimensions (âge \times temps) suit la même logique de base, sauf que des effets "années" et "cohortes" peuvent être ajoutés. La complexité de la mise en œuvre est toutefois accrue.*

Exercice 2 Analyse du modèle à hasard proportionnel de Cox.

On considère des données de survie (avec censure aléatoire à droite indépendante et non-informative). Elles sont influencées par un certain nombres de covariables observées \mathbf{X} (non touchées par la censure). Ces données sont analysées au moyen d'un modèle à hasard proportionnel, pour lequel une fonction de base non spécifiée est utilisée (modèle semi-paramétrique de Cox).

1. En notant β le vecteur de paramètres du modèle, rappeler l'expression de la fonction de hasard $h(t | \mathbf{X}; \beta)$ d'un modèle à hasard proportionnel où $h_0(t)$ est la fonction hasard de base. En déduire l'expression de la fonction de survie, puis celle des probabilités de décès conditionnelles $q(t | \mathbf{X}; \beta)$.
2. On considère deux individus i et j de covariables \mathbf{X}_i et \mathbf{X}_j . Commenter le ratio de leur fonction de hasard.
3. On analyse au moyen d'un tel modèle la durée de vie résiduelle d'individus malades en fonction du sexe* et de l'âge auquel la maladie s'est développée. Les résultats obtenus sous le logiciel R sont les suivants.

```
Call:
coxph(formula = Surv ~ age + sex, data = t)

n= 228, number of events= 165

            coef  exp(coef)   se(coef)      z Pr(>|z|)    
age    0.017045  1.017191  0.009223  1.848  0.06459 .  
sex   -0.513219  0.598566  0.167458 -3.065  0.00218 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95
```

*. Codage : 1 pour les hommes et 2 pour les femmes.

age	1.0172	0.9831	0.9990	1.0357
sex	0.5986	1.6707	0.4311	0.8311

Concordance= 0.603 (se = 0.026)
 Rsquare= 0.06 (max possible= 0.999)
 Likelihood ratio test= 14.12 on 2 df, p=0.0008574
 Wald test = 13.47 on 2 df, p=0.001187
 Score (logrank) test = 13.72 on 2 df, p=0.001048

Commenter ces résultats (significative et interprétation des effets).

4. Rappeler comment peut-être analysée l'hypothèse de proportionnalité. Commenter les résultats ci-dessous et la Figure 1

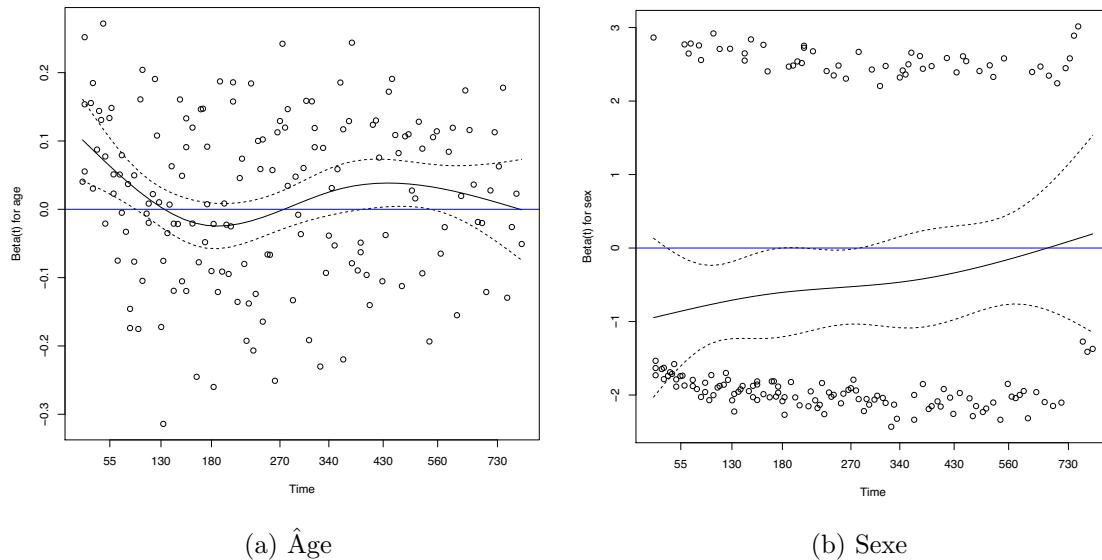


FIGURE 1 – Analyse graphique des résidus de

	rho	chisq	p
age	-0.0275	0.129	0.719
sex	0.1236	2.452	0.117
GLOBAL	NA	2.651	0.266

Réponse de l'exercice 2.

1. On a

$$h(t | \mathbf{X}; \boldsymbol{\beta}) = h_0(t) \exp(\mathbf{X}^\top \boldsymbol{\beta}).$$

D'où

$$S(t | \mathbf{X}; \boldsymbol{\beta}) = \exp\left(-\exp(\mathbf{X}^\top \boldsymbol{\beta}) H_0(t)\right) = S_0(t)^{\exp(\mathbf{X}^\top \boldsymbol{\beta})},$$

et

$$q(t | \mathbf{X}; \boldsymbol{\beta}) = 1 - \left(\frac{S_0(t+1)}{S_0(t)}\right)^{\exp(\mathbf{X}^\top \boldsymbol{\beta})} = 1 - (1 - q_0(t))^{\exp(\mathbf{X}^\top \boldsymbol{\beta})},$$

2. Le ratio des fonctions de hasard est

$$\frac{h(t | \mathbf{X}_i; \boldsymbol{\beta})}{h(t | \mathbf{X}_j; \boldsymbol{\beta})} = \exp\left((X_i - X_j)^\top \boldsymbol{\beta}\right).$$

Ce ratio ne dépend pas du temps et uniquement de la "différence" de caractéristiques entre deux individus.

3. La population de base est constitué des hommes.

Un test de Wald ($H_0 : \beta_k = 0$ pour $k \in \{1, 2\}$) est effectué pour analyser la significativité de chaque coefficient. Le sexe est une variable significative à 5% (p-valeur de 0.2%), alors que l'âge ne l'est quasiment pas (p-valeur de 6.5%). L'effet des coefficients peut être simplement interprété compte tenu du caractère proportionnel du modèle (effet multiplicatif). La colonne $\exp(\text{coef})$ permet de donner l'accroissement du risque pour une année supplémentaire en termes d'âge (soit une augmentation de l'ordre de 1,7%), et la baisse du risque pour les femmes (la mortalité des femmes est ici 40% inférieure à celle des hommes). **Pour que cette intéprétation soit validée, il faut que l'hypothèse de proportionnalité soit satisfaite.**

Un test global $H_0 : \beta_1 = \beta_2 = 0$ est également effectué. Il est satisfait à 5%.

Un intervalle de confiance à 95% (lower .95 et upper .95) pour les exponentiels des coefficients.

Le R^2 apparaît très mauvais, mais cela est courant lorsque l'on utilise le modèle de Cox. Ceci n'est en réalité pas très informatif.

4. Cette hypothèse peut être analysée graphiquement à partir des résidus de Schoenfeld. En notant $Y_i(t)$, l'indicatrice de présence de l'individu i en t , on introduit pour la k -ième covariable

$$\bar{X}_k(t) = \frac{\sum_{i=1}^n Y_i(t) X_{ik} \exp(\mathbf{X}_i \boldsymbol{\beta}^\top)}{\sum_{i=1}^n Y_i(t) \exp(\mathbf{X}_i \boldsymbol{\beta}^\top)},$$

puis le résidu de Schoenfeld ($X_{ik} - \bar{X}_k(t)$). Pour chaque covariable, il est attendu que la valeur des résidus soit nulle pour toute date t . Lorsque l'on représente graphiquement ces résidus, l'existence d'une tendance tend donc à indiquer que l'hypothèse de proportionnalité n'est pas satisfaite.

Cette hypothèse peut également être testée ($H_0 : \beta_k(t) = \beta_k$ pour $k \in \{1, 2\}$). Ce test est construit en posant $\beta_k(t) = \beta_k + \theta_k t$, puis consiste à tester si θ_k est nul pour tout $k = 1, 2$.

Le test effectué sur ces résidus du modèle indique que l'hypothèse de proportionnalité est globalement respectée. Le modèle semble adapté. Le test graphique des résidus permet d'avoir une idée plus précise sur le respect de l'hypothèse de proportionnalité, une pente différente de 0 pour la courbe lissée étant un argument en défaveur de cette hypothèse.

Exercice 3 Modèle de Hannerz.

De nombreux modèles paramétriques ont été proposés dans la littérature. Beaucoup sont souvent peu paramétrés et ne sont applicables qu'à certaines tranches d'âge de la population. Le modèle de Hannerz * tente de rendre compte de la mortalité féminine suédoise dans son ensemble. Ce modèle s'écrit en fonction de 5 paramètres (a_0, \dots, a_3, c) strictement positifs

$$G(x) = \ln\left(\frac{1 - S(x)}{S(x)}\right) = a_0 - \frac{a_1}{x} + \frac{a_2}{2}x^2 + \frac{a_3}{c}e^{cx}.$$

*. Hannerz, H. (2001). Presentation and derivation of a five-parameter survival function intended to model mortality in modern female populations. Scandinavian Actuarial Journal, 2001(2), 176-187.

1. Montrer que $g(x) = \frac{dG(x)}{dx} = \frac{h(x)}{1 - S(x)} = a_1x^{-2} + a_2x + a_3e^{cx}$.
2. Le terme a_1x^{-2} permet de capturer le pic de mortalité aux âges jeunes. Expliquer pourquoi ?
3. Sur la population féminine, le terme a_2x correspond à un accroissement de $g(x)$ sur la période comprise entre 16 et 64 (période d'activité) dans le modèle original de Hannerz. Cette spécification vous semblerait-elle adaptée à la population masculine ?
4. Que pensez-vous du dernier terme ?
5. Bien souvent les données disponibles en assurance ne permettent pas de construire un tableau complet. Un actuari cherchera donc plutôt à utiliser une logique de positionnement. Le modèle relationnel construit sur le modèle de Hannerz prend la forme

$$G(x) - G^{\text{ref}}(x) = b_0 - \frac{b_1}{x} + \frac{b_2}{2}x^2 + \frac{b_3}{c}e^{cx}.$$

Cette relation vous paraît-elle facilement interprétable ?

6. Comment estimeriez-vous ce modèle ?

Réponse de l'exercice 3.

1. Immédiat.
2. Il s'agit d'un terme dont l'intensité se réduit rapidement avec l'âge. Le paramètre a_1 modélise l'intensité du pic.
3. Un accroissement linéaire de $g(x)$ est constaté sur cette tranche d'âge. Cette relation ne convient pas aux hommes, compte tenu de la "bosse des accidents" qui est observée chez les jeunes adultes. Hannerz capture cet effet par le mélange de deux lois de probabilité.
4. Il s'agit d'un accroissement exponentiel de la mortalité aux âges élevés.
5. Le modèle est très facilement interprétable : b_1 , b_2 et b_3 capturent des différences entre les deux populations à différents âges de la vie. Le paramètre b_0 correspond à une différence de niveau pour la mortalité globale.
6. Il s'agit d'une simple régression linéaire.

Exercice 4 Modélisation prospective de la mortalité.

Le risque de longévité d'un régime de retraite est usuellement analysé au moyen d'un modèle de mortalité prospective. Ces analyses s'appuient généralement sur les données d'une population nationale. On suppose que la fonction de hasard à l'âge x et pour la date t , notée $\mu_x(t)$ est constante sur un carré $[x, x+1] \times [t, t+1]$ du diagramme de Lexis.

On s'intéresse au modèle de Lee-Carter

$$\ln(\hat{\mu}_x(t)) = \alpha_x + \beta_x \kappa_t + \epsilon_{xt},$$

avec ϵ_{xt} des erreurs centrées, indépendantes et de variance σ^2 .

1. Interpréter les coefficients du modèle.
2. Les contraintes

$$\sum_{x=x_{\min}}^{x_{\max}} \beta_x = 1 \text{ et } \sum_{t=t_{\min}}^{t_{\max}} \kappa_t = 0,$$

sont appliquées sur les coefficients. Expliquer pourquoi ?

3. Poser le problème d'optimisation considéré dans le modèle traditionnel de Lee-Carter pour déterminer les paramètres (α, β, κ) . Quel est l'inconvénient majeur de cette approche ?
4. Proposer une reformulation du modèle pour corriger cet inconvénient.
5. Dans le cadre de cette seconde approche, nous avons tracé les résidus de la déviance (cf. Figure 2) du modèle en fonction de l'âge et de l'année calendaire à partir des données nationales correspondant à la mortalité des hommes en Angleterre et au Pays de Galles. Quelle limite importante voyez-vous apparaître ? Comment la corrigeeriez-vous ?

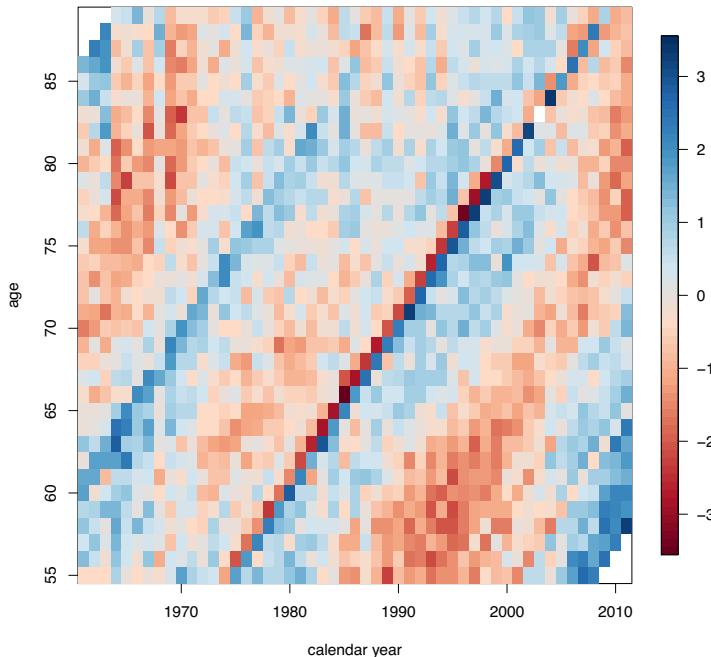


FIGURE 2 – Analyse graphique des résidus de la déviance en fonction de l'âge et de l'année calendaire.

6. Expliquer comment procéder à la projection des taux de mortalité pour des dates t futures ?

Réponse de l'exercice 4.

Les applications numériques de cet exemple sont repris des exemples présentés dans le package R StMoMo *.

1. Pour la population nationale d'Angleterre et du Pays de Galles observée entre 1961 et 2011 pour la tranche d'âge 55-89, l'évolution de ces coefficients sont présentés sur la Figure 3. Le paramètre α_x est le niveau moyen du log des taux instantanés à chaque âge. Ils sont donc généralement croissants (sauf aux âges jeunes, mais qui ne sont pas présentés ici). Les κ_t décrivent l'évolution, en log, de la mortalité au court du temps et les β_x correspondent à la sensibilité par âge à cette tendance d'évolution (coefficients le plus souvent positifs dans les applications). Le coefficients κ_t sont globalement décroissants et suivent une tendance

*. Villegas, A. M., Kaishev, V. K., et Millossovich, P. (2015). StMoMo : An R Package for Stochastic Mortality Modelling. In *7th Australasian Actuarial Education and Research Symposium*.

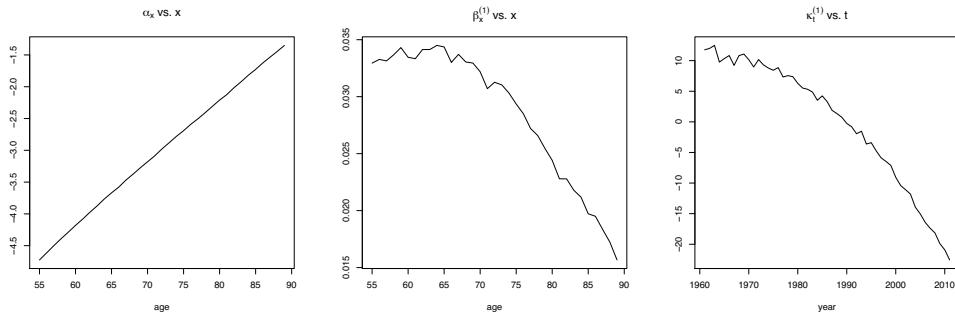


FIGURE 3 – Évolution des coefficients du modèle de Lee-Carter en fonction de l’âge et du temps.

approximativement linéaire. Cette évolution du coefficient traduit l’accroissement régulier de l’espérance de vie avec le temps (amélioration de l’hygiène, du système de santé, ...).

2. Sous sa formulation non contrainte, le modèle de Lee-Carter n’est pas identifiable, i.e. qu’il est possible de trouver deux jeux de paramètres qui renvoient la même valeur de $\ln(\mu_x(t))$. Par exemple, la modification

$$(\alpha_x, \beta_x, \kappa_t) \rightarrow \left(\alpha_x + c\beta_x, \frac{\beta_x}{d}, d(\kappa_t - c) \right),$$

n’affecte pas la valeur de $\ln(\mu_x(t))$. Les contraintes imposées permettent d’éviter ce problème.

3. Sous les contraintes évoquées à la question précédente qui garantissent l’unicité de la solution, il s’agit de résoudre le problème des moindres carrés suivant

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\kappa}) = \arg \min_{(\boldsymbol{\alpha}, \boldsymbol{\beta}, \kappa)} \sum_{x=x_{\min}}^{x_{\max}} \sum_{t=x_{\min}}^{x_{\max}} (\ln(\hat{\mu}_x(t) - \alpha_x + \beta_x \kappa_t))^2.$$

Une approche classique pour le résoudre consiste à procéder à une décomposition en valeurs singulières. Une approche de type Newton-Raphson est également possible. Un des inconvénients majeur de cette approche est qu’elle s’appuie une l’hypothèse d’homoskédasticité des erreurs. Cette hypothèse n’est généralement pas satisfaite dans la pratique.

4. L’approche proposée dans le modèle de Brouhns *et al.* (2002) consiste à utiliser un modèle GLM de Poisson. Une approche par maximum de vraisemblance peut alors être envisagée, ce qui permet de s’affranchir de l’hypothèse d’homoskédasticité des erreurs. En utilisant un lien log, on a donc

$$D_{xt} \sim \mathcal{P}(E_{xt}\mu_x(t)),$$

avec $\mu_x(t) = \exp(\alpha_x + \beta_x \kappa_t)$ et E_{xt} , l’exposition au risque sur le carré (x, t) du diagramme de Lexis. Cette reformulation ne modifie pas la signification des coefficients. Un algorithme de Newton-Raphson multivarié est utilisé pour l’estimation des paramètres.

5. Les résidus laissent apparaître des tendances diagonales, alors qu’ils devraient être répartis aléatoirement. Cette effet est corrigé (partiellement) en introduisant un effet cohorte γ_{t-x} , i.e.

$$\ln(\mu_x(t)) = \alpha_x + \beta_x^{(0)} \kappa_t + \beta_x^{(1)} \gamma_{t-x}.$$

6. La projection du modèle s’effectue traditionnellement en supposant, pour le modèle de Lee-Carter, que la tendance κ_t suit un modèle de série temporelle. Dans la très grande majorité des cas, un modèle ARIMA est considéré (une marche aléatoire le plus souvent).

Exercice 5 Modèle à risques concurrents.

Soit T la durée de vie d'un individu que l'on suppose soumise à K causes de sortie (ex : causes de décès). On définit la *fonction de hasard spécifique au risque i*

$$h^{(i)}(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \mathbb{P}(t < T \leq t + \Delta t, V = i | T > t),$$

où V est la variable aléatoire qui désigne la cause de sortie.

1. Donner l'expression de :

- la fonction de hasard h (associée à T) en fonction des $h^{(i)}$;
- la fonction de survie S (associée à T) en fonction des $h^{(i)}$;
- la *fonction d'incidence cumulée* $F^{(i)}(t) = \mathbb{P}(T \leq t, V = i)$;
- le taux de décès entre les dates t et $t + 1$ pour la cause $V = i$.

2. On introduit les durées de vie latentes par cause T_1, T_2, \dots, T_K . Exprimer $S(t)$ en fonction des fonctions de survie de chacune des lois latentes en faisant l'hypothèse qu'elles sont indépendantes. Que peut-on en déduire s'agissant de l'expression des taux de hasard de chacune de ces lois latentes ?
3. En notant $S_{(T_1, \dots, T_K)}(t_1, \dots, t_K)$ la fonction de survie jointe de T_1, T_2, \dots, T_K et sans supposer que ces lois sont nécessairement indépendantes, montrer que

$$h^{(i)}(t) = -\frac{1}{S(t)} \left. \frac{\partial S_{(T_1, \dots, T_K)}(t_1, \dots, t_K)}{\partial t_i} \right|_{t_1, \dots, t_K=t}.$$

4. Pour $K = 2$, nous faisons à présent l'hypothèse que les durées de vie latentes sont corrélées et reliées par la fonction de survie jointe

$$S_{T_1, T_2}(t_1, t_2) = (1 + \theta(\lambda_1 t_1 + \lambda_2 t_2))^{-\frac{1}{\theta}},$$

avec $\lambda_1, \lambda_2, \theta \geq 0$. Fournir l'expression de la fonction de hasard $h^{(i)}(t)$ pour $i = 1, 2$, puis celle de $F^{(i)}(t)$.

Réponse de l'exercice 5.

1. La fonction de hasard s'obtient par

$$h(t) = \sum_{i=1}^K h^{(i)}(t).$$

La fonction de survie s'écrit

$$S(t) = \exp \left(- \int_0^t h(u) du \right) = \exp \left(- \sum_{i=1}^K \int_0^t h^{(i)}(u) du \right).$$

La fonction d'incidence cumulée s'écrit

$$F^{(i)}(t) = \int_0^t h^{(i)}(u) S(u) du = \int_0^t h^{(i)}(u) \exp \left(- \sum_{j=1}^K \int_0^u h^{(j)}(s) ds \right) du.$$

Le taux de décès est

$$\begin{aligned} {}_1q_t^{(i)} &= \mathbb{P}(t < T \leq t+1, V = i | T > t) \\ &= \frac{\int_t^{t+1} h^{(i)}(u) S(u) du}{S(t)} = \frac{F^{(i)}(t+1) - F^{(i)}(t)}{S(t)}. \end{aligned}$$

2. On a $T = \min_{i=1,\dots,K} (T_i)$. Par conséquent sous l'hypothèse d'indépendance, la fonction de survie s'écrit en fonction des fonctions de survie des lois marginales $S_i(t)$

$$S(t) = \prod_{i=1}^K \mathbb{P}(T_i > t) = \prod_{i=1}^K S_i(t).$$

On montre également sous cette hypothèse que

$$\begin{aligned} F^{(i)}(t) &= \mathbb{P}\left(T_i \leq t, T_i \leq \min_{j \neq i} (T_j)\right) \\ &= \mathbb{E}\left[\mathbb{1}_{\{T_i \leq t\}} \mathbb{E}\left[\mathbb{1}_{\{T_i \leq \min_{j \neq i} (T_j)\}} | T_i\right]\right] \\ &= \int_0^t \prod_{j \neq i} S_j(u) f_i(u) du, \end{aligned}$$

avec f_i la densité de T_i .

Dans ces conditions, la fonction de hasard spécifique au risque i s'écrit

$$\begin{aligned} h^{(i)}(t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \mathbb{P}(t < T \leq t + \Delta t, V = i | T > t) \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \Delta t q_t^{(i)} \\ &= \frac{1}{S(t)} \frac{dF^{(i)}(t)}{dt} = \frac{1}{S(t)} \prod_{j \neq i} S_j(t) f_i(t) \\ &= \frac{f_i(t)}{S_i(t)} = h_i(t) \end{aligned}$$

avec h_i la fonction de hasard de T_i .

3. On écrit que (avec des notations évidentes)

$$\begin{aligned} &\frac{\partial S_{(T_1, \dots, T_K)}(t_1, \dots, t_K)}{\partial t_i} \Big|_{t_1, \dots, t_K=t} \\ &= -f_i(t) \int_t^\infty \cdots \int_t^\infty dF_{(T_1, \dots, T_{i-1}, T_{i+1}, \dots, T_K) | T_i}(t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_K | t) \\ &= -\frac{dF^{(i)}(t)}{dt}. \end{aligned}$$

D'où le résultat.

4. En reprenant le résultat précédent et comme $S(t) = S_{(T_1, T_2)}(t, t)$, on a pour $i = 1, 2$

$$h^{(i)}(t) = -\frac{1}{S(t)} \frac{\partial S_{(T_1, T_2)}(t_1, t_2)}{\partial t_i} \Big|_{t_1, t_2=t} = \frac{\lambda_i}{(1 + \theta t(\lambda_1 + \lambda_2))},$$

et

$$\begin{aligned} F^{(i)}(t) &= \int_0^t \frac{\lambda_i}{(1 + \theta u (\lambda_1 + \lambda_2))} (1 + \theta u (\lambda_1 + \lambda_2))^{-\frac{1}{\theta}} du \\ &= \frac{\lambda_i}{(\lambda_1 + \lambda_2)} \left(1 - (1 + \theta t (\lambda_1 + \lambda_2))^{-\frac{1}{\theta}} \right). \end{aligned}$$