

# Econométrie des modèles de durée

Guillaume Horny\*

\*Banque de France et UCLouvain

Master 2 MOSEF

# Chapitre 4: Modèles à hasards proportionnels

# Bienvenu dans un monde semi-paramétrique!

hasards proportionnels  $\Rightarrow$  semi-paramétrique devient possible

**Semi-paramétrique** car une partie du modèle peut ne **pas** être spécifiée avec une forme fonctionnelle

**Approches semi-paramétriques :**

- + flexibles que les modèles paramétriques
- moins souples que les approches non-paramétriques
- + peuvent prendre en compte des variables explicatives

# Plan

- 1 Modèles à hasards proportionnels
- 2 Modèle exponentiel à hasard constant par morceaux
- 3 Modèle de Cox
- 4 Estimation paramétrique
- 5 Estimation semi-paramétrique
- 6 Modèle de Cox pour données groupées
- 7 Modèle de Cox pour durées discrètes

# Modèles à hasards proportionnels

# Chapitre 4

- 1 Modèles à hasards proportionnels
- 2 Modèle exponentiel à hasard constant par morceaux
- 3 Modèle de Cox
- 4 Estimation paramétrique
- 5 Estimation semi-paramétrique
- 6 Modèle de Cox pour données groupées
- 7 Modèle de Cox pour durées discrètes

# Forme générale

$$\lambda(t|X, \theta) = \lambda_0(t; \alpha)\lambda_1(X; \beta); \quad (1)$$

où  $\theta = (\alpha, \beta)$ ,  $\lambda_0$  et  $\lambda_1$  sont deux fonctions positives.

- $\lambda_0$  ne dépend que du temps et est appelée **hasard de base**
- $\lambda_1$  ne dépend que de  $X$ . En général, on pose  $\lambda_1(X; \beta) = \exp(X'\beta)$

**Note:** les modèles exponentiels et Weibull sont à hasards proportionnels.

# Conséquence de la proportionnalité

Soit  $X$  une indicatrice:

$$\lambda_i(t|X, \theta) = \begin{cases} \lambda_0(t|\alpha) & \text{si } X_i = 0 \\ \lambda_0(t|\alpha)\lambda_1(\beta) & \text{si } X_i = 1 \end{cases} \quad (2)$$

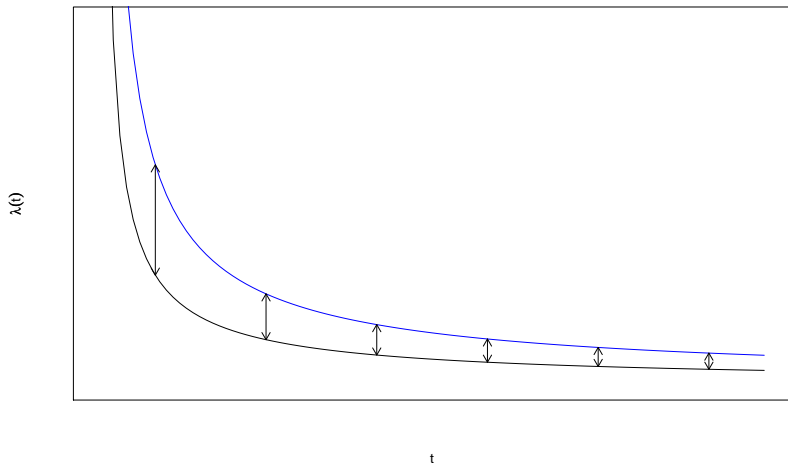
$\lambda_0(t)$  est le risque commun à tous les individus, dont dévient les individus pour lesquels  $X_i = 1$  d'un facteur  $\lambda_1(\beta)$ .



# Conséquence de la proportionnalité (suite)

Les hasards de tous les individus  $i$  sont proportionnels au hasard de base

Fonctions de hasards proportionnels



# Modèle exponentiel à hasard constant par morceaux

# Chapitre 4

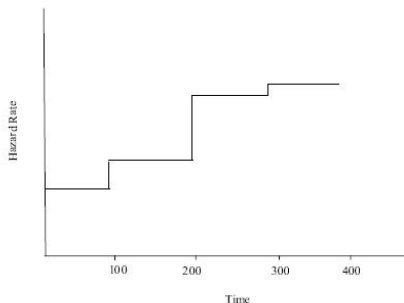
- 1 Modèles à hasards proportionnels
- 2 Modèle exponentiel à hasard constant par morceaux**
- 3 Modèle de Cox
- 4 Estimation paramétrique
- 5 Estimation semi-paramétrique
- 6 Modèle de Cox pour données groupées
- 7 Modèle de Cox pour durées discrètes

## Modèle exponentiel à hasard constant par morceaux

On peut spécifier un **modèle exponentiel à hasard constant par morceaux** où:

$$\lambda_0(t; \alpha) = \exp(\alpha_j), \alpha_{j-1} \leq t < \alpha_j, j = 1, \dots, J, \quad (3)$$

où  $\alpha_0 = 0$  et  $\alpha_{J+1} = \infty$ .



# Modèle exponentiel à hasard constant par morceaux (suite)

## Propriétés:

- modèle paramétrique potentiellement **très** flexible ( $J$  paramètres!), car la forme du hasard n'est pas contrainte par une forme fonctionnelle aussi rigide que dans les modèles paramétriques du chapitre 3
- choix du nombre d'intervalles de temps?  
⇒ construire des intervalles contenant au moins une transition (sinon  $\alpha_j \rightarrow -\infty$ )
- modèle est semi-paramétrique lorsqu'il y a autant d'indicatrices que de durées

## Ré-écriture

Pour  $\lambda_1 = \exp$ , on peut écrire le hasard:

$$\lambda(t; \theta) = \exp(\alpha_j + X'\beta) \quad (4)$$

Sur chaque intervalle de temps, on a un hasard constant (=exponentiel)  
⇒ équivalent à un modèle avec  $J$  indicatrices!

## Spécifications du hasard de base

- le hasard constant ne peut décrire que la forme générale du hasard
- d'autant plus que le hasard est "pentu"  
⇒ fixer des intervalles d'autant plus petits que le hasard est pentu
- une alternative est de laisser  $\lambda_0$  non-spécifiée par une forme paramétrique, et le modèle est alors semi-paramétrique. Il est alors appelé **modèle de Cox** (plus loin dans ce chapitre).

# Reformulation du modèle EHCPM

**Rappel:** Si les durées suivent une loi exponentielle, alors le nombre d'événements se produisant dans un intervalle de temps suit une loi de Poisson.



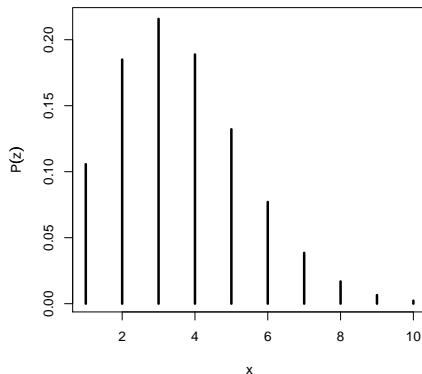
## Le retour de Siméon Denis...

**Loi de Poisson:**

$$Z \sim \mathcal{P}(\lambda)$$

$$\Leftrightarrow \Pr(Z = z) = \frac{\lambda^z \exp(-\lambda)}{z!}, \forall z \in \mathbb{N}.$$

Densité d'une loi de Poisson ( $\lambda = 3.5$ )



## Vraisemblance d'un modèle de Poisson

Soit un échantillon  $(z_i)_{i=1,\dots,n}$ , où  $z_i \sim \mathcal{P}(\lambda)$ . La vraisemblance s'écrit:

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^n \Pr(Z_i = z_i) \\ &= \prod_{i=1}^n \frac{\lambda^{z_i} \exp(-\lambda)}{z_i!} \\ \Rightarrow \ln L(\lambda) &= \sum_{i=1}^n [z_i \ln \lambda - \lambda - \ln(z_i!)] . \end{aligned}$$

Pour le modèle de régression  $\ln \lambda = X'\beta$ , on a:

$$\ln L(\lambda) = \sum_{i=1}^n [z_i X_i' \beta - \exp(X_i' \beta)] . \quad (5)$$

## Vraisemblance d'un modèle de hasard

$$\begin{aligned}\ln L(\theta) &= \sum_{i=1}^n [\delta_i \ln f(t_i; \theta) + (1 - \delta_i) \ln S(t_i; \theta)] \\ &= \sum_{i=1}^n [\delta_i \ln \lambda(t_i; \theta) + \ln S(t_i; \theta)] \\ &= \sum_{i=1}^n [\delta_i \ln \lambda(t_i; \theta) - \Lambda(t_i; \theta)].\end{aligned}$$

## Vraisemblance d'un modèle HCPM

On note  $j(i)$  l'indice de l'intervalle au cours duquel l'individu  $i$  effectue une transition.

On a :

$$\lambda(t_i; \theta) = \exp(\alpha_{j(i)} + X_i' \beta)$$

$$\Lambda(t_i; \theta) = \sum_{k \leq j(i)} (y_k - \alpha_{k-1}) \exp(\alpha_k + X_i' \beta),$$

où  $y_k = \min(\alpha_k, t_i)$ .

## Vraisemblance d'un modèle HCPM (suite)

D'où:

$$\begin{aligned}
 \ln L(\theta) &= \sum_{i=1}^n \left[ \delta_i(\alpha_{j(i)} + X_i' \beta) - \sum_{k \leq j(i)} (y_k - \alpha_{k-1}) \exp(\alpha_k + X_i' \beta) \right]. \\
 &= \sum_{i=1}^n \left[ \sum_{k \leq j(i)} \delta_k(\alpha_k + X_i' \beta) - \sum_{k \leq j(i)} (y_k - \alpha_{k-1}) \exp(\alpha_k + X_i' \beta) \right]. \\
 &= \sum_{i=1}^n \sum_{k \leq j(i)} [\delta_k(\alpha_k + X_i' \beta) - (y_k - \alpha_{k-1}) \exp(\alpha_k + X_i' \beta)].
 \end{aligned}$$

## Vraisemblances Poisson et HCPM

$$\ln L^P(\lambda) = \sum_{i=1}^n [z_i X_i' \beta - \exp(X_i' \beta)],$$

$$\ln L^{HCPM}(\theta) = \sum_{i=1}^n \sum_{k \leq j(i)} [\delta_k (\alpha_k + X_i' \beta) - (\alpha_k - \alpha_{k-1}) \exp(\alpha_k + X_i' \beta)].$$

la vraisemblance d'un modèle à HCPM est équivalente à celle d'un modèle de Poisson pour  $\delta_k$  lorsque:

- les durées sont subdivisées en intervalles,
- on a une indicatrice  $\alpha_k$  par intervalle de temps,
- la longueur de chaque intervalle intervient comme un paramètre de décentrage (offset).

## Exemple: estimation sur données simulées (1/5)

```
rm(list=ls())
library(survival)

# paramètres de la simulation
N <- 5000
id <- 1:N

# simulation durées Weibull ( $S(t) = \exp(-\alpha t^\theta)$ )
theta0 <- 0.8
alpha0 <- 1
t <- (-log(runif(N))/alpha0)^(1/theta0)
delta <- seq(1,1, length=N) # pas de censure
cut1 <- c(1.5,3,6)
```

## Exemple: estimation sur données simulées (2/5)

```

# mise en forme des données par intervalles
data <- data.frame(t, delta)
data2 <- survSplit(data, cut=cut1, start="start", end="t", \\
  event="delta", id ="id",episode = "interval")
data3 <- data2[order(data2$id), ] # trie par individu
data3$duree <- data3$t - data3$start
data3
table(data3$interval)

#création des indicatrices d'intervalle (alpha_j)
data3$t1 <- ifelse(data3$interval==0,1,0)
data3$t2 <- ifelse(data3$interval==1,1,0)
data3$t3 <- ifelse(data3$interval==2,1,0)
data3$t4 <- ifelse(data3$interval==3,1,0)

```



## Exemple: estimation sur données simulées (3/5)

```

# maximisation de la vraisemblance d'un modèle exponentiel
loglikeli <- fonction(theta){ -1 * sum(data3$delta * \
(theta[1] * data3$t1 \
+ theta[2] * data3$t2 + theta[3] * data3$t3 \
+ theta[4] * data3$t4 - data3$duree * exp(theta[1] * data3$t1 \
+ theta[2] * data3$t2 + theta[3] * data3$t3 \
+ theta[4] * data3$t4)))}

inits <- c(0,0,0,0)
est <- optim(inits, loglikeli, method = "BFGS")
fit <- glm(delta ~ t1 + t2 + t3 + t4 -1, offset=log(duree), \
family=poisson, data = data3)

```

## Exemple: estimation sur données simulées (4/5)

```
# hasard prédit pour l'individu moyen à chaque date
```

```
has_pred <- c(0,0,0,0)
```

```
has_pred[1] <- exp(est$par[1])
```

```
has_pred[2] <- exp(est$par[2])
```

```
has_pred[3] <- exp(est$par[3])
```

```
has_pred[4] <- exp(est$par[4])
```

```
# vrai hasard de l'individu moyen
```

```
has_vrai <- function(t){alpha0 * theta0 * t^(theta0 - 1)}
```

```
plot(has_vrai, xlim=c(0,max(t)), ylim=c(0,1.3),\
```

```
  main="Vrai hasard et hasard estimé",ylab=expression(lambda(t))
```

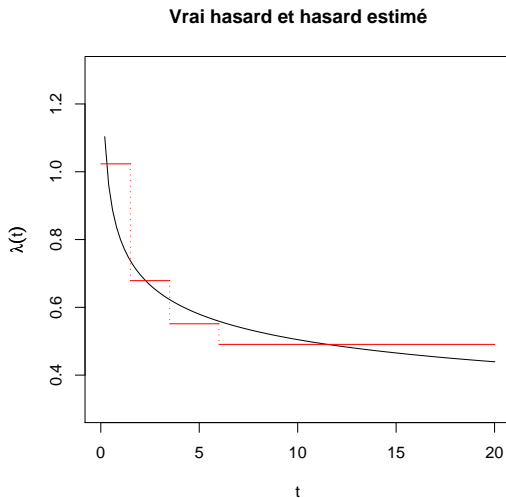
```
segments(0,has_pred[1],cut1[1], has_pred[1], col=2)
```

```
segments(cut1[1],has_pred[2],cut1[2], has_pred[2], col=2)
```

```
segments(cut1[2],has_pred[3],cut1[3], has_pred[3], col=2)
```

```
segments(cut1[3],has_pred[4],max(t), has_pred[4], col=2)
```

## Exemple: estimation sur données simulées (5/5)



## Exemple: mortalité infantile (1/)

Données issues du World Fertility Survey de 1976, concernant la mortalité infantile en Colombie.

### Echantillon:

- 3 cohortes (1941-59, 1961-67 et 1968-76),
- données groupées:
  - ▶ mesures effectuées à 1,3 et 6 mois, puis à 1,2,5 et 10 ans,
  - ▶ pas de durée individuelles: les transitions sont supposées se produire au milieu de l'intervalle.
- les données sont déjà découpées en intervalles.

## Exemple: mortalité infantile (2/)

Aperçu des données

interval	cohort	exposure	events
1	1	278.4	168
2	1	538.8	48
3	1	794.4	63
4	1	1550.8	89
5	1	3006.0	102
6	1	8743.5	81

## Exemple: mortalité infantile (2/)

### Variables:

- cohort : 1 pour 1941-59, 2 pour 1961-67 et 3 pour 1968-76,
- interval :  $1, \dots, 7$ ,
- exposure : durée de vie dans l'intervalle,
- event : indicatrice de transition sur l'intervalle.

## Exemple: mortalité infantile (3/)

## Notations:

- $i$  est l'indice de cohorte ( $i = 1, \dots, 3$ ) et  $j$  l'indice de l'intervalle de temps ( $j = 1, \dots, 7$ ).
- $\alpha_j \in \{1/12, 3/12, 6/12, 1, 2, 5\}$
- $\delta_{ij}$  est le nombre de décès pour la cohorte  $i$  dans l'intervalle  $j$
- $\delta_{ij} \sim \mathcal{P}(\lambda_{ij})$  où

$$\lambda_{ij} = \exp(\alpha_j + X' \beta) \quad (6)$$

## Exemple: mortalité infantile (4/)

```
rm(list=ls())
library(survival)

# chargement des données
infant <- read.table("X:/.../Colombia.txt", header = T)
head(infant)

infant$interval <- as.factor(infant$interval)
infant$cohort <- as.factor(infant$cohort)
```



## Exemple: mortalité infantile (5/)

```
# estimation
fitN <- glm(events ~ 1, offset = log(exposure), \\
family = poisson, data = infant)
fitA <- glm(events ~ -1 + interval, offset = log(exposure),\\
family = poisson, data = infant)
fitC <- glm(events ~ -1 + cohort, offset = log(exposure),\\
family = poisson, data = infant)
fitAC <- glm(events ~ -1 + interval + cohort, offset =\\
log(exposure), family = poisson, data = infant)
```

## Exemple: mortalité infantile (6/)

```
# comparaison des modèles
anova(fitN, fitA, fitC, fitAC)

# prévisions
t <- c(0, 1/12, 3/12, 6/12, 1, 2, 5, 10)
M <- length(t) - 1

haz41 <- fitAC$coeff[1:M]
haz60 <- fitAC$coeff[1:M] + fitAC$coeff[M + 1]
haz68 <- fitAC$coeff[1:M] + fitAC$coeff[M + 2]
```

## Exemple: mortalité infantile (7/)

```

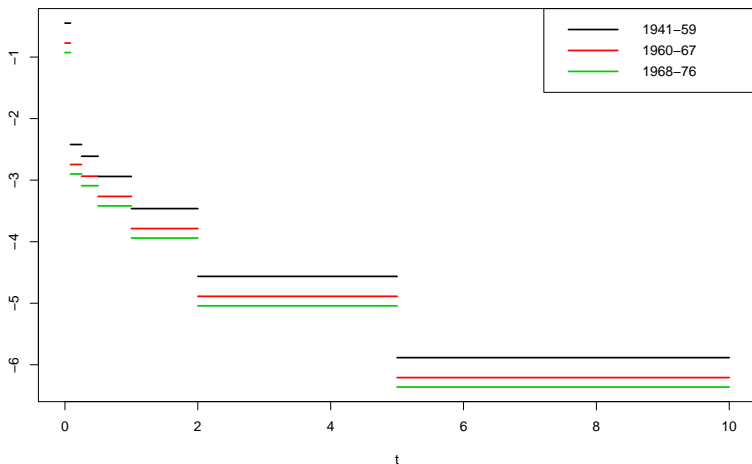
# graphique des prévisions
plot(1, 1, t = "n", ylim = range(haz41,haz60, haz68), xlim = ra

for (i in 1:length(haz41)) {
segments(x0 = t[i], y0 = haz41[i],x1 = t[i + 1], y1 = haz41[i],
segments(x0 = t[i], y0 = haz60[i],x1 = t[i + 1], y1 = haz60[i],
segments(x0 = t[i], y0 = haz68[i],x1 = t[i + 1], y1 = haz68[i],
}
legend("topright", legend = c("1941-59","1960-67", "1968-76"),

```

## Exemple: mortalité infantile (8/)

Hasards de base prédits



# Modèle de Cox

# Chapitre 4

- 1 Modèles à hasards proportionnels
- 2 Modèle exponentiel à hasard constant par morceaux
- 3 Modèle de Cox**
- 4 Estimation paramétrique
- 5 Estimation semi-paramétrique
- 6 Modèle de Cox pour données groupées
- 7 Modèle de Cox pour durées discrètes

# Fonction de hasard

$$\lambda(t|X, \theta) = \lambda_0(t) \exp(X'\beta), \quad (7)$$

où  $\lambda_0(t)$  n'est pas spécifiée avec une forme paramétrique

## Sir David Cox

**David Cox** (1924-), professeur à Oxford, auteur de nombreuses contributions aux statistiques et probabilités. La plus reconnue est sûrement son modèle de hasard proportionnel (publié en 1972). Bien qu'ayant plus de 80 ans aujourd'hui, il continue à publier, à participer aux congrès...





## Interprétation d'estimations d'un modèle de Cox

On a:

$$\ln \lambda(t) = \ln \lambda_0(t) + X' \beta. \quad (8)$$

D'où:

- si  $X_k$  est une variable continue en log, son coefficient est l'élasticité du hasard par rapport à  $X_k$ .
- si  $X_k$  est une variable binaire, le coefficient mesure environ le taux de croissance du hasard lorsque  $X_k$  passe de 0 à 1:

$$\begin{aligned} \ln \lambda(t|X_k = 1) - \ln \lambda(t|X_k = 0) &= \beta_k, \\ \Leftrightarrow \beta_k &= \ln \left( 1 + \frac{\lambda(t|X_k = 1) - \lambda(t|X_k = 0)}{\lambda(t|X_k = 0)} \right) \\ &\approx \frac{\ln \lambda(t|X_k = 1) - \ln \lambda(t|X_k = 0)}{\ln \lambda(t|X_k = 0)} \end{aligned}$$

## Ré-écriture du modèle de Cox

On a vu que le hasard intégré suit une loi exponentielle unitaire (chap 1):

$$f_{\Lambda(T)}[\Lambda(t)] = \exp[-\Lambda(t)] \quad (9)$$

La densité de  $X = \ln \Lambda(T)$  est donc (changement de variable):

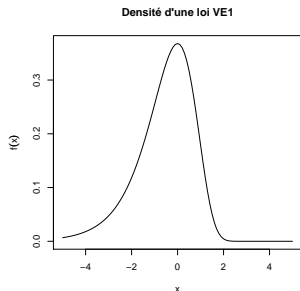
$$f_X(x) = \exp[x - \exp(x)], \quad (10)$$

où  $x = \ln \Lambda(t)$ . Ainsi,  $\ln[\Lambda(T)]$  suit une loi VE1.

## Loi Valeur Extrême de type 1

Une variable  $X$  suit une loi VE (parfois aussi appelée Gumbel) ssi elle a :

- pour fonction de répartition  $F(x) = 1 - \exp[-\exp(x)]$
- pour densité  $f(x) = \exp[x - \exp(x)]$ .



La loi VE1 ci-dessus est la loi asymptotique du min de  $n$  VA indépendantes normales. Elle admet  $E(\epsilon) \approx -0.5722$  et  $V(\epsilon) = \pi^2/6$ .

## Ré-écriture du modèle de Cox (suite)

Or:

$$\Lambda(t) = \Lambda_0(t) \exp(X'\beta). \quad (11)$$

D'où l'écriture:

$$\ln[\Lambda(t)] = \ln[\Lambda_0(t)] + X'\beta = \epsilon, \quad (12)$$

La loi de  $\epsilon$  est la loi VE1, qui ne dépend d'aucun paramètre!  
 $\Rightarrow$  toute la flexibilité du modèle de Cox provient de la forme non-paramétrique de  $\lambda_0$ .

# Estimation paramétrique

# Chapitre 4

- 1 Modèles à hasards proportionnels
- 2 Modèle exponentiel à hasard constant par morceaux
- 3 Modèle de Cox
- 4 Estimation paramétrique**
- 5 Estimation semi-paramétrique
- 6 Modèle de Cox pour données groupées
- 7 Modèle de Cox pour durées discrètes

# Ré-écriture des modèles à hasards proportionnels

Pour  $\lambda_1 = \exp$ , on a :

$$\begin{aligned}\Lambda(t) &= \Lambda_0(t) \exp(X'\beta) \\ \Leftrightarrow \ln [\Lambda_0(t)] + X'\beta &= \epsilon,\end{aligned}$$

où  $\epsilon \sim \text{VE1}$ , avec  $E(\epsilon) \approx -0.5722$  et  $V(\epsilon) = \pi^2/6$ .

Comme  $\ln [\Lambda(t)]$  n'est pas observable (mais on connaît sa loi!), on peut écrire :

$$\begin{aligned}\ln [\Lambda(t)] - X'\beta &= \ln [\Lambda_0(t)], \\ \Leftrightarrow \ln [\Lambda_0(t)] &= -X'\beta + \epsilon,\end{aligned}\tag{13}$$

L'équation (13) est un modèle de régression lorsque  $\Lambda_0(t)$  est connu

# Estimation par régression linéaire des modèles HP

La théorie usuelle des MCO ne s'applique pas directement à (13):

- la loi de  $\epsilon$  n'est pas centrée:  $E(\epsilon) \approx -0.5722$
- $\Lambda_0$  n'est en général pas connu (et c'est tout l'intérêt du modèle, qui le distingue du HP paramétrique!)
- on observe généralement des durées censurées



# Estimation par régression linéaire des modèles HP (suite)

- on résout le premier problème en posant  $\epsilon = \epsilon_0 + \epsilon_1$ , où  $\epsilon_0$  est une constante (la constante du modèle)
- on suppose que  $\lambda_0$  est connu,
- on suppose qu'il n'y a pas de censure;

⇒ pour  $\beta \in \mathbb{R}^p$ , on peut estimer  $\beta$  par moindres carrés linéaire ou non-linéaires avec la relation (13) (exercice!). L'intérêt d'une telle procédure est toutefois purement académique, tant les conditions requises sont restrictives!

# Estimation par MV des modèles HP

Soit  $\lambda_0$  connue et définie par un paramètre de dimension finie.  
La log-vraisemblance s'écrit:

$$\begin{aligned} \ln L_n(\theta) &= \sum_{i=1}^n [\delta_i \ln f(t_i; \theta) + (1 - \delta_i) \ln S(t_i; \theta)] \\ &= \sum_{i=1}^n [\delta_i \ln \lambda(t_i; \theta) + \ln S(t_i; \theta)] \\ &= \sum_{i=1}^n \left\{ \delta_i [X_i' \beta + \ln \lambda_0(t_i; \alpha)] - \exp(X_i' \beta) \Lambda_0(t_i; \alpha) \right\}. \end{aligned}$$

## Estimation par MV des modèles HP (suite)

$$\begin{aligned}\frac{\partial}{\partial \beta} \ln L_n(\beta) &= \sum_{i=1}^n X_i [\delta_i - \exp(X_i' \beta) \Lambda_0(t_i; \alpha)] \\ &= - \sum_{i=1}^n X_i \exp(X_i' \beta) r_i^*(\beta),\end{aligned}$$

où  $r_i^*(\beta) = \Lambda_0(t_i; \alpha) - \delta_i \exp(-X_i' \beta)$  est le résidu généralisé. Sans censure, ( $\delta_i = 1, \forall i$ ) il serait le résidu théorique du modèle régressant  $\Lambda_0(t_i; \alpha)$  sur  $\exp(-X_i' \beta)$ .

L'EMV est donc caractérisé par la condition d'orthogonalité entre  $X_i \exp(X_i' \beta)$  et le résidu généralisé  $r_i^*(\beta)$ .

## Calcul de la matrice d'information

$$\begin{aligned}
 I_n(\theta) &= -E_\theta \left[ \frac{\partial^2 \ln L_n(\theta)}{\partial \beta \partial \theta'} \middle| X \right] \\
 &= \sum_{i=1}^n X_i X_i' \exp(X_i' \beta) E_\theta [\Lambda_0(T_i) | X_i]
 \end{aligned}$$

La quantité  $E_\theta [\Lambda_0(T_i) | X_i]$  fait intervenir la distribution de la censure, on ne peut donc pas l'évaluer sans poser une loi.

On peut montrer qu'en l'absence de censure, la matrice serait:

$$I_n^*(\theta) = \sum_{i=1}^n X_i X_i' \gg I_n(\theta) \quad (14)$$

**Conclusion:** la censure introduit une perte d'information, donc des estimations moins précises

# Estimation semi-paramétrique

# Chapitre 4

- 1 Modèles à hasards proportionnels
- 2 Modèle exponentiel à hasard constant par morceaux
- 3 Modèle de Cox
- 4 Estimation paramétrique
- 5 Estimation semi-paramétrique**
- 6 Modèle de Cox pour données groupées
- 7 Modèle de Cox pour durées discrètes

# Introduction

Cox propose une méthode d'estimation de  $\beta$  qui ne requiert pas d'estimer simultanément  $\lambda_0(t)$  (on peut toutefois le déduire ensuite avec  $\hat{\beta}$ ). Pour simplifier, la présentation va reprendre les intuitions de modèles en temps discret. Toutefois, la méthode est conçue en temps continu.

## Intuition

Soit  $R(t_i) = \{j : t_j \geq t_i\}$  l'ensemble des individus au risque. On a :

$$\begin{aligned} \Pr[T_i = t_i | R(t_i)] &= \frac{\Pr[T_i = t_i | T_i \geq t_i]}{\sum_{j \in R(t_i)} \Pr[T_j = t_j | T_j \geq t_j]} \\ &= \frac{\lambda_i(t_i | X_i, \beta)}{\sum_{j \in R(t_i)} \lambda_j(t_j | X_j, \beta)} \\ &= \frac{\exp(X_i' \beta)}{\sum_{j \in R(t_i)} \exp(X_j' \beta)}. \end{aligned}$$



## Exemple

Individu	Durée	$X_i$
1	2	4
2	3	1
3	6	3
4	12	2

**Note:** le durées ont été classées par ordre croissant avant que les individus soient indicés

- $R(t_1) = R(2) = \{1, 2, 3, 4\}$ , d'où:

$$\Pr[T_1 = t_1 | R(t_1)] = \frac{\exp(X_1\beta)}{\exp(X_1\beta) + \exp(X_2\beta) + \exp(X_3\beta) + \exp(X_4\beta)}.$$

- $R(t_2) = R(3) = \{2, 3, 4\}$

$$\Pr[T_2 = t_2 | R(t_2)] = \frac{\exp(X_2\beta)}{\exp(X_2\beta) + \exp(X_3\beta) + \exp(X_4\beta)}.$$

## Exemple (suite)

- $R(t_3) = R(6) = \{3, 4\}$

$$\Pr[T_3 = t_3 | R(t_3)] = \frac{\exp(X_3\beta)}{\exp(X_3\beta) + \exp(X_4\beta)}.$$

- $R(t_4) = R(12) = \{4\}$

$$\Pr[T_4 = t_4 | R(t_4)] = \frac{\exp(X_4\beta)}{\exp(X_4\beta)} = 1.$$

## Vraisemblance partielle sans censure

Cox définit la vraisemblance partielle:

$$L_p(\beta) = \prod_{i=1}^n \Pr[T_i = t_i | R(t_i)] = \prod_{i=1}^n \frac{\exp(X_i' \beta)}{\sum_{j \in R(t_i)} \exp(X_j' \beta)}. \quad (15)$$

D'où la log-vraisemblance partielle:

$$\ln L_p(\beta) = \sum_{i=1}^n \left( X_i' \beta - \ln \left[ \sum_{j \in R(t_i)} \exp(X_j' \beta) \right] \right). \quad (16)$$

## Vraisemblance partielle avec censure

Les observations censurées ne contribuent pas aux transitions observées mais à l'ensemble au risque, d'où:

$$L_p(\beta) = \prod_{i=1}^n \left[ \frac{\exp(X_i' \beta)}{\sum_{j \in R(t_i)} \exp(X_j' \beta)} \right]^{\delta_i}. \quad (17)$$

D'où la log-vraisemblance partielle:

$$\ln L_p(\beta) = \sum_{i=1}^n \delta_i \left( X_i' \beta - \ln \left[ \sum_{j \in R(t_i)} \exp(X_j' \beta) \right] \right). \quad (18)$$

# Score

Les conditions de premier ordre sont:

$$\frac{\partial}{\partial \beta} \ln L_p(\beta) = \sum_{i=1}^n \delta_i \left[ X_i - \frac{\sum_{j \in R(t_i)} X_j \exp(X_j' \beta)}{\sum_{j \in R(t_i)} \exp(X_j' \beta)} \right]. \quad (19)$$

Le dernier terme s'interprète comme une moyenne pondérée de  $X_j$  sur les individus au risque.

## Comparaison avec la vraisemblance complète

La vraisemblance partielle n'utilise que l'information sur l'ordre des sorties, donc moins d'information que la vraisemblance complète. On peut toutefois montrer que:

$$\widehat{\beta}_p \stackrel{n \rightarrow \infty}{\sim} \mathcal{N} \left[ \beta, \left( -E \left[ \frac{\partial^2 \ln L_p(\beta)}{\partial \beta \partial \beta'} \right] \right)^{-1} \right]. \quad (20)$$

Bien que  $\widehat{\beta}_p$  ne soit pas efficace, la perte de précision par rapport à des approches paramétriques est généralement limitée et constitue un faible coût à payer pour éviter le risque d'une spécification erronée de  $\lambda_0$ .

# Ex-aequo

- on peut approcher la vraisemblance partielle de plusieurs manières en cas d'*ex-aequo* : approximation de Breslow, d'Efron, vraisemblance exacte, moyenne de vraisemblance ...
- en pratique, on recommande d'utiliser la vraisemblance exacte ou une moyenne de vraisemblances lorsqu'il y a peu de durées égales, et l'approximation d'Efron dans le cas contraire. Toutes deux sont généralement disponibles dans les routines d'estimations des logiciels.
- la vraisemblance partielle en présence de durées groupées est étudiée plus loin

## Exemple d'estimation sous R

```

# temps continu
test1 <- list(time= c(4,3,2,1,5,6,7),
              status=c(1,NA,1,0,1,1,0),
              x=      c(0, 2,1,1,1,0,0))
coxph( Surv(time, status) ~ x, test1)

# ex-aequo
test2 <- list(time= c(4,3,1,1,2,2,3),
              status=c(1,1,1,0,1,1,1),
              x=      c(0,2,1,1,1,0,0))

coxph( Surv(time, status) ~ x, test2) # Efron par défaut
coxph( Surv(time, status) ~ x, test2, method='breslow')
coxph( Surv(time, status) ~ x, test2, method='exact')

```



# Estimation de la survie

On cherche un estimateur non-paramétrique de la survie. On reprend donc ici une approche similaire à celles conduisant à l'estimateur de Kaplan-Meier (chapitre 2)

**Notation:**

- $d_i$ , le nombre de transitions dans l'intervalle  $]t_{i-1}, t_i]$

## Vraisemblance approchée

Sans censure, on a:

$$L_{app}(S) = \prod_{i=1, \dots, n} [S(t_{i-1}|X) - S(t_i|X)]^{d_i}. \quad (21)$$

Or, pour le modèle de Cox:

$$\begin{aligned} S(t_i) &= \exp[-\Lambda(t|X)] \\ &= \exp \left[ - \exp(X_i' \beta) \int_0^t \lambda_0(u) du \right] \\ &= \exp \left[ - \int_0^t \lambda_0(u) du \right]^{\exp(X_i' \beta)} \\ &= S_0(t)^{\exp(X_i' \beta)}. \end{aligned}$$

## Vraisemblance approchée (suite)

D'où:

$$\begin{aligned}
 S(t_{i-1}|X) - S(t_i|X) &= S_0(t_{i-1})^{\exp(X'_{i-1}\beta)} - S_0(t_i)^{\exp(X'_i\beta)} \\
 &= \left[ \alpha_j^{-1} S_0(t_i) \right]^{\exp(X'_i\beta)} - S_0(t_i)^{\exp(X'_i\beta)} \\
 &= \left[ \alpha_j^{-\exp(X'_i\beta)} - 1 \right] S_0(t_i)^{\exp(X'_i\beta)},
 \end{aligned}$$

que l'on substitue dans l'équation (21). On a également:

$$S_0(t)^{\exp(X'\beta)} = \prod_{j|t_j \leq t} \alpha_j^{\exp(X'_j\beta)}, \quad (22)$$

## Vraisemblance approchée (suite)

On peut montrer que la fonction maximisant  $L_{app}(S_0)$  sur l'espace des fonctions de survie est constante par morceaux, avec des sauts aux dates de sortie:

$$S_0(t|X) = \prod_{j|t_j \leq t} \alpha_j, \quad (23)$$

où les  $\alpha_j$  maximisent  $L_{app}(S_0)$ . Lorsque  $\hat{\beta}$  est connu, les  $\alpha_j$  sont solution de:

$$\frac{\partial}{\partial \alpha_j} L_{app}(\alpha, \hat{\beta}) = 0. \quad (24)$$

D'où, en l'absence ex-aequo:

$$\hat{\alpha}_j = 1 - \frac{\exp(X_j' \beta)}{\sum_{l \in R(t_j)} \exp(X_l' \beta)} \quad (25)$$

## Remarques

- on déduit  $\widehat{S}(t|X)$  par  $\widehat{S}(t|X) = \widehat{S}_0(t)^{\exp(X'\widehat{\beta})}$
- lorsqu'il  $\widehat{\beta} = 0$ , on retrouve l'estimateur de Kaplan-Meier du hasard intégré
- Attention à l'interprétation:  $\widehat{S}(t|X) \neq \widehat{S}_{KM}(t)$ , même si  $\widehat{\beta} = 0$ , car  $\widehat{S}(t|X)$  change lorsqu'on ajoute ou retire des colonnes dans  $X$ .

# Modèle de Cox pour données groupées

# Chapitre 4

- 1 Modèles à hasards proportionnels
- 2 Modèle exponentiel à hasard constant par morceaux
- 3 Modèle de Cox
- 4 Estimation paramétrique
- 5 Estimation semi-paramétrique
- 6 Modèle de Cox pour données groupées**
- 7 Modèle de Cox pour durées discrètes

# Notations et hypothèses

On observe de intervalles de temps  $I_j = ]a_{j-1}, a_j]$ ,  $\forall j \in 2, \dots, J$ , où  $I_1 = [0, a_1]$  et  $I_{J+1} = ]a_J, +\infty[$ , à l'intérieur desquels des transitions ou des censures se produisent à des dates inobservées.

## Hypothèses:

- les variables explicatives sont constante sur chaque intervalle de temps
- les données sont continues et leur distribution est donnée par le hasard  $\lambda(t|X) = \lambda_0(t) \exp(X'\beta)$



## Taux de hasard HP pour données groupées

Le taux de hasard en temps discret s'écrit:

$$\begin{aligned}
 \Pr(T \in I_j | T > a_{j-1}, X) &= \frac{\Pr(T \in I_j | X)}{\Pr(T > a_{j-1} | X)} \\
 &= \frac{S(a_{j-1} | X) - S(a_j | X)}{S(a_{j-1} | X)} \\
 &= 1 - \frac{S(a_j | X)}{S(a_{j-1} | X)} \\
 &= 1 - \frac{\exp \left[ - \int_0^{a_j} \lambda(u | X) du \right]}{\exp \left[ - \int_0^{a_{j-1}} \lambda(u | X) du \right]}
 \end{aligned}$$

## Taux de hasard HP pour données groupées (suite)

$$\begin{aligned}
 \Pr(T \in I_j | T > a_{j-1}, X) &= 1 - \exp\left(-\int_{a_j}^{a_{j-1}} \lambda(u|X) du\right) \\
 &= 1 - \exp\left[-\exp(X'\beta) \int_{a_j}^{a_{j-1}} \lambda_0(u) du\right] \\
 &= 1 - \exp\left[-\lambda_0^j \exp(X'\beta)\right] \\
 &= 1 - \exp\left(-\exp\left[\ln \lambda_0^j + X'\beta\right]\right),
 \end{aligned}$$

où  $\lambda_0^j = \int_{a_{j-1}}^{a_j} \lambda_0(u) du$ . D'où:

$$\lambda^{GR}(a_j|X) = 1 - \exp\left(-\exp\left[\ln \lambda_0^j + X'\beta\right]\right). \quad (26)$$

## Fonction de survie HP pour données groupées

$$\begin{aligned} S^{GR}(a_j|X) &= \Pr[T > a_{j-1}|X] \\ &= \prod_{k=1}^{a-1} \left[ 1 - \lambda^{GR}(k|X) \right] \\ &= \prod_{j=1}^{a-1} \left[ \exp \left( - \exp \left[ \ln \lambda_0^j + X' \beta \right] \right) \right]. \end{aligned}$$

## Vraisemblance du modèle HP pour données groupées

La contribution à la vraisemblance de l'individu  $i$  s'écrit:

$$L_i(\beta, \lambda_0^1, \dots, \lambda_0^J) = \prod_{j=1}^{J_i} \Pr(T_i \in ]a_{j-1}, a_j] | X)^{\delta_{ij}} \Pr(T_i > a_j | X)^{1-\delta_{ij}}, \quad (27)$$

où  $J_i = \sup\{j | T_i \geq a_j\}$ . D'où:

$$L_i(\beta, \lambda_0^1, \dots, \lambda_0^J) = \prod_{j=1}^{J_i} [\Pr(T_i \in ]a_{j-1}, a_j] | T_i > a_{j-1}, X) \Pr(T_i > a_{j-1} | X)]^{\delta_{ij}} \Pr(T_i > a_j | X)^{1-\delta_{ij}},$$

## Vraisemblance du modèle HP pour données groupées (suite)

$$\begin{aligned}
 L_i(\beta, \lambda_0^1, \dots, \lambda_0^J) &= \prod_{j=1}^{J_i} \left[ 1 - \exp \left( - \exp \left[ \ln \lambda_0^j + X' \beta \right] \right) \right]^{\delta_{ij}} \\
 &\quad \Pr(T_i > a_{j-1} | X)^{\delta_{ij}} \Pr(T_i > a_j | X)^{1 - \delta_{ij}} \\
 &= \prod_{j=1}^{J_i} \left[ 1 - \exp \left( - \exp \left[ \ln \lambda_0^j + X' \beta \right] \right) \right]^{\delta_{ij}} \\
 &\quad \left[ \exp \left( - \exp \left[ \ln \lambda_0^j + X' \beta \right] \right) \right]^{1 - \delta_{ij}} \\
 &\quad \exp \left( - \exp \left[ \ln \lambda_0^{j-1} + X' \beta \right] \right).
 \end{aligned}$$

# Maximisation de la vraisemblance du modèle HP pour données groupées

- la gradient de la log-vraisemblance est non-linéaire  
⇒ optimisation *via* des procédures numériques
- les propriétés de l'EMV s'appliquent pour  $\widehat{\beta}, \widehat{\lambda}_0^1, \dots, \widehat{\lambda}_0^J$ .

# Modèle de Cox pour durées discrètes

# Chapitre 4

- 1 Modèles à hasards proportionnels
- 2 Modèle exponentiel à hasard constant par morceaux
- 3 Modèle de Cox
- 4 Estimation paramétrique
- 5 Estimation semi-paramétrique
- 6 Modèle de Cox pour données groupées
- 7 Modèle de Cox pour durées discrètes**



## Extension aux durées discrètes

Cox (1972) propose une extension du modèle HP à des durées discrètes:

$$\frac{\lambda^d(a_j|X)}{1 - \lambda^d(a_j|X)} = \frac{\lambda_0^d(a_j|X)}{1 - \lambda_0^d(a_j|X)} \exp(X'\beta). \quad (28)$$

En log:

$$\ln \left( \frac{\lambda^d(a_j|X)}{1 - \lambda^d(a_j|X)} \right) = \alpha_j + X'\beta, \quad (29)$$

où  $\alpha_j = \ln \left( \frac{\lambda_0^d(a_j|X)}{1 - \lambda_0^d(a_j|X)} \right)$ . Ce qui est encore exprimé:

$$\lambda^d(a_j|X) = \frac{\exp(\alpha_j + X'\beta)}{1 + \exp(\alpha_j + X'\beta)} \quad (30)$$

# Estimation

Un tel modèle peut-être estimé avec les routines standards pour des modèles logistiques, où on introduit des indicatrices pour chaque intervalle de temps.

# Conclusion

- si les durées sont continues: vraisemblance partielle,
- si les durées sont groupées: “cloglog”,
- si les durées sont vraiment discrètes: “logit”,
- si les durées sont continues mais que vous voulez un hasard constant par morceaux (ça peut arriver parfois): modèle à hasard constant par morceaux.