TP SdF N° 31

Application du modèle de Cox à la maintenance prédictive

Ce TP porte sur le modèle de Cox et son utilisation dans le cadre d'une maintenance prédictive pour optimiser l'occurrence des interventions sur un matériel à partir de divers indicateurs d'état de dégradation.

1 - Présenter le modèle de Cox avec des covariables fixes ou évolutives en fonction du temps.

2 - Simuler l'occurrence de 100 défaillances d'un type de matériel dont le fonctionnement est régi par un modèle de COX à 3 covariables couplé à une loi de Weibull à 3 paramètres (on concidérera des covariables uniformément réparties entre 0 et 10 et une loi de Weibull de paramètres $\beta_0 = 1,5$, $\sigma_0 = 2000$ et $\gamma_0 = 100$).

3 - Effectuer un ajustement à partir des données simulées pour tenter de retrouver les 6 paramètres du modèle de deux manières différentes :

- par la méthode du maximum de vraisemblance,
- en maximisant la fonction de vraisemblance partielle de Cox.

4 - Trouver la date de maintenance optimale d'un équipement du type de ceux préalablement simulés, sachant que :

- L'évolution des covariables ont été observées sur l'équipement pendant une certaine durée (500 heures avec un pas de 50 heures),

- L'équipement doit respecter une exigence de sécurité exprimée sous la forme d'un taux de défaillance par heure à ne pas dépasser ($5\ 10^{-4}$).

On simulera les valeurs des covariables observées avec des lois de type $Xi(t) = ai\ t^{bi}$, auxquelles sera ajouté un bruit uniforme entre 0 et 0,05, en utilisant les paramètres suivants :

a_1 : 0,000005	a_2 : -0,02	a_3 : 0,01
b_1 : 1,5	b_2 : 0,3	b_3 : 0,5

1 - Le modèle de Cox

Le modèle PH (Proportional Hazards) de régression à risques proportionnels proposé par Cox[1], est un modèle de vie accélérée dans lequel le taux de défaillance de base $\lambda_0(t)$ est affecté par différentes covariables X_i favorisant ($\beta_i > 0$) ou s'opposant ($\beta_i < 0$) à l'apparition des défaillances :

$$\lambda(t) = \lambda_0(t) \times \exp(\sum \beta_i X_i)$$

Utilisé notamment en épidémiologie pour prendre en compte plusieurs facteurs de risque dans la survenue de maladies ou de décès, ce modèle est parfois employé en fiabilité pour tenir compte de certains facteurs de stress (charge, sollicitation, température, etc.).

A partir de données expérimentales, sa forme particulière permet d'estimer les coefficients β_i indépendamment de $\lambda_0(t)$, en maximisant la fonction de vraisemblance partielle de Cox qui ne s'intéresse qu'à l'ordre d'apparition des défaillances.

$$L(y_1,\dots y_n, \beta_i) = \prod_{k=1 \text{à } n} (\exp(\sum \beta_i X_{ki}) / \sum_{j=k \text{ à } n} \exp(\sum \beta_i X_{ji}))$$

En effet la probabilité que l'un des équipements survivant k soit concerné par l'une des défaillances est égale à :

$$p_k / \sum_{j \text{ survivants}} p_j = \lambda_0(t) \times \exp(\sum \beta_i X_{ki}) \times \Delta t / \sum_{j \text{ survivants}} \lambda_0(t) \times \exp(\sum \beta_i X_{ji}) \times \Delta t$$

$$\text{soit} \quad \exp((\sum \beta_i X_{ki}) / \sum_{j \text{ survivants}} \exp(\sum \beta_i X_{ji})$$

Ce modèle peut être également utilisé avec des covariables évolutives dans le temps $X_i(t)$. Dans le cadre d'une maintenance prédictive, il peut ainsi servir à optimiser l'occurrence des interventions sur un matériel à partir de divers indicateurs d'état de dégradation (niveau vibratoire, qualité du liquide de lubrification, échauffement du matériel, impédance d'un circuit électrique, etc.)

Mais il ne prend pas alors en compte un éventuel effet cumulatif des covariables. Ainsi, le taux de défaillance d'une machine pourra être affecté par la qualité de son huile à l'instant courant mais pas par l'évolution antérieure de celle-ci. Le modèle GPH (Generalized Proportional Hazards) de Bagdonavicius-Nikulin, qui ne fait pas l'objet du présent TP, permet de pallier cette limitation[2].

Les coefficients β_i peuvent encore être estimés par la fonction de vraisemblance partielle de Cox si la valeur des covariables Xi (t) de tous les équipements est connue au moment de chacune des défaillances. Mais, cette condition est évidemment difficile à respecter dans le cadre du suivi opérationnel d'une flotte de matériels.

En revanche, si la valeur des covariables n'est mémorisée au moment des défaillances que sur les équipements concernés, un modèle entièrement paramétrique, de type Cox couplé à une loi de Weibull par exemple, peut être ajusté globalement par la méthode du maximum de vraisemblance, sous réserve d'utiliser un outil d'optimisation performant.

Par ailleurs, la fiabilité d'un équipement peut s'exprimer de la manière suivante :

$$R(t) = e^{-\int_0^t \lambda(\tau) d\tau} = e^{-\int_0^t \lambda_0(\tau) \exp(\sum \beta_i X_i(\tau)) d\tau}$$ soit en intégrant par partie :

[1] Cox D.R., « Regression models and life tables », journal of the Royal Statistical Society, series B, vol 34, n°2, 1972, pages 187-220.

[2] Mikhail Nikulin, Vincent Couallier, Leo Gerville-Reache, « Accelerated Life and Degradation Models in Reliabilty and Safety: An Engineering Perspective », EA 2961, Statistique Mathematique et ses Applications, University Victor Segalen Bordeaux 2, Bordeaux, France

$$R(t) = e^{-[H_0(\tau)\exp(\sum \beta_i X_i(\tau))]_0^t + \int_0^t H_0(\tau)\exp(\sum \beta_i X_i(\tau))' d\tau}$$ avec H_0 la primitive de λ_0

ou $R(t) = R_0(t)^{\sum \beta_i X_i(t)} e^{\int_0^t H_0(\tau)\exp(\sum \beta_i X_i(\tau)) \sum \beta_i X_i'(\tau) d\tau}$

Soit $R(t) = R_0(t)^{\sum \beta_i X_i}$ si les covariables sont indépendantes du temps

2 - Simulation d'un modèle de COX couplé à une loi de Weibull

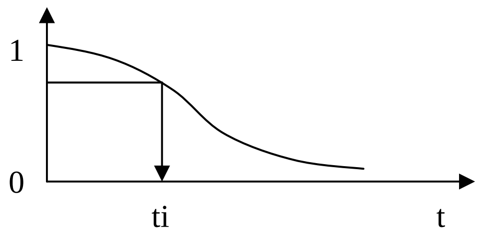
Dans le cas d'un modèle de COX couplé à une loi de Webull, les différentes covariables X_i n'ont un effet que sur la valeur du paramètre σ :

$$\lambda(t) = \lambda_0(t) \times \exp(\sum\beta_i X_i) = \beta_0(t-\gamma_0)^{\beta 0-1}/\sigma_0^{\beta 0} \times \exp(\sum\beta_i X_i) = \beta(t-\gamma)^{\beta-1}/\sigma^{\beta}$$

$$\text{avec } \gamma = \gamma_0,\ \beta = \beta_0 \text{ et } 1/\sigma^{\beta} = 1/\sigma_0^{\beta 0} \times \exp(\sum\beta_i X_i)$$

$$\text{soit } \sigma = \sigma_0 \times \exp(-\sum\beta_i X_i/\beta)$$

Une durée de fonctionnement avant défaillance peut être simulée en tirant une valeur aléatoire entre 0 et 1 et en lui appliquant l'inverse de la fonction de répartition (ou de fiabilité) de la loi de Weibull qui le caractérise.



$$R(t) = 1\text{- } F(t) = \exp(-[(t-\gamma)/\sigma]^{\beta}) \qquad \text{Fonction inverse : } t = \gamma\text{-}\sigma\text{*}\ln(R(t))^{1/\beta}$$

$$\text{soit } t_i = \gamma + \sigma\text{*(-LN(ALEA())}\wedge^{1/\beta} \text{ ou}$$

$$\text{soit } t_i = \gamma_0 + \sigma_0\text{*(-LN(ALEA()/exp(}\sum\beta_i X_i\text{))}\wedge^{1/\beta 0} \text{ sous Excel}$$

Un jeu de 100 défaillances peut être ainsi simulé en considérant 3 covariables uniformément réparties entre 0 et 10 et une loi de Weibull de paramètres $\beta_0 = 1{,}5$, $\sigma_0 = 2000$ et $\gamma_0 = 100$.

3 - Ajustement à partir des données simulées

3.1 Méthode du maximum de vraisemblance

La méthode du maximum de vraisemblance consiste à rechercher la configuration de paramètres du modèle théorique qui donne la plus forte probabilité à l'occurrence d'un jeu de données et donc la densité de probabilité maximale pour les valeurs de celui-ci. On maximise alors le produit des densités (vraisemblance) ou plutôt la somme des logarithmes des densités (LNV) pour s'affranchir de problèmes numériques (produit de très petites valeurs).

La fonction de densité de probabilité d'une loi de Weibull s'exprime de la manière suivante :

$$f(t) = R(t) * \lambda(t) \text{ avec } R(t) = \exp(-[(t-\gamma)/\sigma]^{\beta}) \quad \text{et} \quad \lambda(t) = \beta(t-\gamma)^{\beta-1} / \sigma^{\beta} \qquad t \geq \gamma$$

Afin de s'affranchir de la condition $t \geq \gamma$, on peut exprimer la fiabilité et le taux de défaillance, sous Excel, au moyen des formules suivantes:

λ(T) =β*MAX(0;T-γ)^(β-1)/σ^β R(T) =EXP(-((MAX(0;T-γ)/σ)^σ)) avec $\sigma = \sigma_0 \times \exp(-\sum \beta_i X_i/\beta)$

Par ailleurs, le calcul peut conduire à des résultats erronés durant l'optimisation si l'on cherche à calculer le logarithme d'une valeur nulle ou 0^0 qui n'est pas défini sous Excel. Aussi est-il nécessaire d'effectuer un test sur les valeurs des densités avant de sommer leurs logarithmes.

Réalisé par l'outil d'optimisation GENCAB, l'ajustement au jeu des 100 données préalablement simulées donne les résultats suivants :

Modèle de Cox

	Lois simulées		Lois ajustées
Bêta :	1,5	Bêta :	1,61
Sigma :	2000	Sigma :	2049
Gamma :	100	Gamma :	99
Bêta X :	0,5	Bêta X :	0,54
Bêta Y :	-0,2	Bêta Y :	-0,19
Bêta Z :	0,3	Bêta Z :	0,33

LNV : -659,853

	Xi	Yi	Zi	T
Simu :	**5,50**	**0,78**	**8,71**	**137**
	2,81	3,14	2,89	965
	4,31	1,01	2,39	245
	4,71	8,75	5,44	261
	8,37	1,93	2,01	199

Lbd1	R1	f(T)		LN(f(T))
0,003016	0,19678	0,000593	0,000593	-7,4295
0,002929	0,766472	0,002245	0,002245	-6,09892
0,002381	0,786882	0,001874	0,001874	-6,27986
0,01529	0,384311	0,005876	0,005876	-5,13688

On retrouve approximativement les 6 paramètres du modèle utilisés pour simuler les données.

Le fichier Excel correspondant est accessible en cliquant sur l'icône ci après :

Cox

3.1 Utilisation de la fonction de vraisemblance partielle de Cox

A partir du même jeu de données préalablement trié, les paramètre β_i ont été calculés en maximisant la fonction de vraisemblance partielle de Cox.

Présentés ci-dessous, les résultats obtenus sont peu différents des précédents.

Modèle de Cox (Vraissemblance partielle)

	Lois simulées
Bêta :	1,5
Sigma :	2000
Gamma :	100
Bêta X :	0,5
Bêta Y :	-0,2
Bêta Z :	0,3

	Paramètres ajustés
Bêta X :	0,52
Bêta Y :	-0,19
Bêta Z :	0,33

LNV : -321,147

	Xi	Yi	Zi	T
Simu :	**9,71**	**7,87**	**4,64**	**218**
	6,81	2,53	9,37	100
	8,95	8,52	6,58	106
	9,00	8,45	8,46	108
	8,65	5,87	9,54	109

Exp(Σβi Xi)	Proba		LN(Proba)
437,9545	0,033638	0,033638	-3,39209
171,3414	0,013618	0,013618	-4,29633
329,1065	0,026519	0,026519	-3,62989
635,7265	0,052622	0,052622	-2,94463

Fichier Excel :

Vraissemblance partielle de Cox

4 – Maintenance prédictive

Les valeurs des 3 covariables, observées sur l'équipement pendant 500 heures avec un pas de 50 heures, ont été simulées en considérant des lois de type $X_i(t) = a_i\, t^{b_i}$, avec les paramètres ci-dessous, plus un bruit uniforme entre 0 et 0,05.

a_1 : 0,000005	a_2 : -0,02	a_3 : 0,01
b_1 : 1,5	b_2 : 0,3	b_3 : 0,5

A partir des données simulées, ces mêmes lois ont été ajustées par la méthode des moindres carrés afin de pouvoir prédire la valeur des covariables au-delà de la durée d'observation de 500 heures.

La date de maintenance optimale de l'équipement a alors été calculée afin que son taux de défaillance ne dépasse pas le seuil de 5 10^{-4} pannes par heure de fonctionnement.

La valeur de T = 815 heures a été obtenue en résolvant par optimisation l'équation suivante.

$$\lambda(T) = \mathrm{EXP}(\beta_1 * a_1 T^{b1} + \beta_2 * a_2 T^{b2} + \beta_3 * a_3 T^{b3}) * \beta * \mathrm{MAX}(0; T-\gamma)^{\beta-1} / \sigma^{\beta} = 5\ 10^{-4}$$

En fait, l'ajustement des lois d'évolution des covariables et l'estimation de la valeur de T, présentés ci-dessous, ont été réalisés simultanément par une optimisation globale à 7 paramètres afin de simplifier les traitements.

Maintenance prédictive (date optimale avec taux de défaillance > seuil)

Lois théorique

a_1 :	5E-06
b_1 :	1,5
a_2 :	-0,02
b_2 :	0,3
a_3 :	0,01
b_3 :	0,5

Bruit uniforme

entre +/- 0,05

	Modèle Théorique			Simulation			Jeu de données			Moindres carrés			
T	X_1	X_2	X_3	X_1	X_2	X_3	X1	X2	X3	X1	X2	X3	$\Sigma\varepsilon^2$
0	0,00	0,00	0,00	0,04	0,02	0,02	0,02	0,04	0,02	0,00	0,00	0,00	0,002
50	0,00	-0,06	0,07	0,02	-0,04	0,07	0,03	-0,02	0,08	0,01	-0,04	0,09	0,001
100	0,01	-0,08	0,10	0,02	-0,06	0,10	0,02	-0,07	0,13	0,02	-0,06	0,12	2E-04
150	0,01	-0,09	0,12	0,03	-0,05	0,15	0,01	-0,06	0,17	0,03	-0,07	0,14	0,001
200	0,01	-0,10	0,14	0,03	-0,05	0,17	0,03	-0,10	0,15	0,04	-0,08	0,16	5E-04
250	0,02	-0,10	0,16	0,02	-0,07	0,21	0,05	-0,10	0,17	0,05	-0,08	0,18	3E-04
300	0,03	-0,11	0,17	0,07	-0,08	0,22	0,07	-0,10	0,20	0,05	-0,09	0,19	4E-04
350	0,03	-0,12	0,19	0,07	-0,08	0,20	0,07	-0,08	0,20	0,06	-0,10	0,20	4E-04
400	0,04	-0,12	0,20	0,09	-0,09	0,22	0,06	-0,11	0,20	0,07	-0,10	0,21	3E-04
450	0,05	-0,13	0,21	0,07	-0,08	0,24	0,07	-0,09	0,24	0,08	-0,11	0,23	7E-04
500	0,06	-0,13	0,22	0,09	-0,09	0,23	0,08	-0,11	0,23	0,08	-0,11	0,23	4E-05

$\Sigma\varepsilon^2$: 0,0074

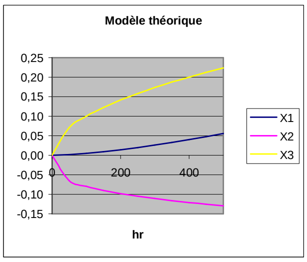
Ajustement

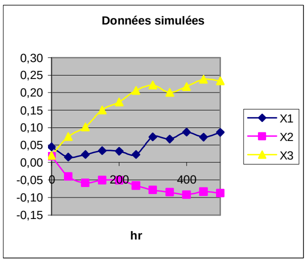
a_1 :	0,00056
b_1 :	0,80324
a_2 :	-0,00884
b_2 :	0,40979
a_3 :	0,01933
b_3 :	0,40188

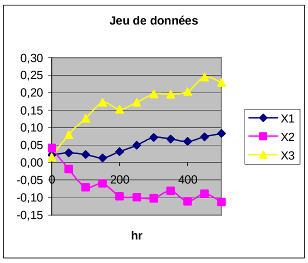
Loi de Cox

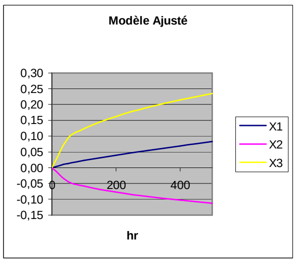
β :	1,61
σ :	2049
γ :	99
$\beta 1$:	0,54
$\beta 2$:	-0,19
$\beta 3$:	0,33

	T	λ	Objectif
T_M :	815	5,00E-04	5,00E-04
			7,21E-11
		ε :	7,39E-03









Fichier Excel : Maintenance prédictive