Régression pour les évènements récurrents : application à la maintenance imparfaite des systèmes réparables

Vincent Couallier, Genia Babykina

Régression pour les évènements récurrents : application à la maintenance imparfaite des systèmes réparables.

Vincent Couallier[(1)] & Genia Babykina[(1),(2)]

(1) : Université Victor Segalen Bordeaux 2, IMB UMR 5251
(2) : Cemagref, 50 Avenue de Verdun, 33610 Cestas

Résumé : Un modèle pour les évènements récurrents adapté à l'analyse de la maintenance imparfaite de systèmes réparables a été défini dans Le Gat (2009). Dans ce modèle, l'intensité stochastique du processus de comptage dépend du temps mais aussi du nombre d'évènements passés. Le modèle peut prendre en compte la donnée de covariables dépendantes du temps. Les données tronquées ou censurés peuvent être envisagés ainsi que des fenêtres d'observations variant d'un individu à l'autre. L'estimation statistique dans un cadre paramétrique permet la prédiction du nombre d'évènements à venir pour un individu particulier ou pour la population à risque.

Dans cette présentation, nous décrirons le modèle et donnerons un résultat de convergence pour l'estimateur du maximum de vraisemblance. Les techniques utilisent une forme particulière du processus de vraisemblance de Jacod adapté à l'observation filtrée de processus de comptage ainsi que les conditions de Borgan pour la normalité asymptotique de l'estimateur.

Abstract : A model for recurrent events adapted to the analysis of imperfect repair of reparable systems has been defined in (Le gat 2009). In this model the stochastic intensity is driven by the cumulated number of previous failures. The model accounts for time-dependent covariates and thus is adapted to dynamic environment. Truncated or censored data can be envisaged as well as several systems with different start and discharge times. The statistical parameter estimation allows predicting the number of future events of a system or of a set of systems. In this presentation, we shall describe this model and give some results of convergence for the maximum likelihood estimator. The technics use a particular form of the Jacod's likelihood process adapted to a filtered counting process and well-known Borgan's conditions for convergence of the MLE.

1 Le modèle LEYP

Le modèle proposé est basé sur une écriture de l'intensité stochastique du processus de comptages des défaillances qui ne correspond ni à un processus de Poisson non homogène, ni à un processus de renouvellement. De nombreux modèles pour la maintenance imparfaite de systèmes réparables existent dans la littérature (Pham et Wang 1996, Doyen et

Gaudoin 2004, Kristov 2007), certains d'entre eux utilisant la notion d'âge virtuel : à un temps t donné, l'intensité dépend d'une fonction de t qui définit l'âge virtuel du système.

Supposons que, pour un système réparable donné, la suite des évènements récurrents reflètent les dates de défaillance, immédiatement suivies d'une action corrective dont la durée est négligée, et notons $(T_i)_{i\geq 1}$ les dates d'évènements comptés dans une échelle temporelle propre au système (i.e. $t = 0$ est la date de "pose" de l'individu). On note $N(t) = \sum_{i\geq 1} \mathbb{1}_{T_i \leq t}$ le processus ponctuel associé, avec $N(0) = 0$.

La fenêtre d'observation des évènements pouvant être réduite, on suppose que ne sont observés que les évènements arrivant dans l'intervalle de temps $[a, T]$. On note alors

$$N^*(t) = [N(t) - N(a)]1_{[a,T]}(t)$$

le processus de comptage des défaillances observées.

L'intensité stochastique d'un processus de comptage (par rapport à une filtration $\mathcal{H}$) est définie par Andersen et al (1993)

$$\lambda(t) = \lim_{dt\to 0} \frac{1}{dt}\mathbb{P}(N(t + dt) - N(t) = 1 \,|\, \mathcal{H}_{t-})$$

où $\mathcal{H}_{t-}$ est l'histoire du processus jusqu'au temps t non inclus. De façon équivalente, on peut noter

$$\mathbb{E}\left[dN(t) \,|\, \mathcal{H}_{t^-}\right] = \lambda(t)dt$$

et $\mathcal{H}$ est très souvent la filtration naturelle engendrée par le processus lui même : $\mathcal{H}(t) = \sigma(N(s), s \in [0, t])$

Dans la présentation, nous nous intéressons au modèle LEYP (Linear Extended Yule Process) introduit par Le Gat (2009) dans lequel l'action corrective a un effet dit "Worse Than Old" (WTO, l'accroissement de l'intensité est en partie due au nombre de défaillances passées). Si on note $Z(t)_{t>0}$ un vecteur de covariables éventuellement dépendent du temps qui décrit par exemple l'environnement dans lequel le système est plongé, le model LEYP avec covariables pose pour intensité (par rapport à la filtration $\sigma(N(s), Z(s), 0 \leq s \leq t)$) le processus suivant

$$\mathbb{E}\left[dN(t) \,|\, N(t-),\, Z(s),\, 0 \leq s < t\right] = (1 + \alpha N(t-))\lambda_0(t, \delta)e^{Z'(t)\beta}\, dt \tag{1}$$

Dans ce modèle purement paramétrique, λ_0 est une fonction non décroissante du temps que nous supposerons dans la suite vérifier $\lambda_0(t, \delta) = \delta t^{\delta-1}$ pour simplifier. Les paramètres à estimer sont donc $\alpha \geq 0$, $\delta > 0$ et $\beta \in \mathbb{R}^p$ et forment le paramètre $\theta = (\alpha, \delta, \beta)$.

Le modèle LEYP inclut le processus de Poisson non homogène ($\alpha = 0$) et le cas particulier cité avec $\lambda_0(t, \delta) = \delta t^{\delta-1}$ se ramène alors au NHPP avec intensité en puissance. Ce dernier est une modélisation bien connue des réparations dites ABAO (As Bad As Old) pour lequel la réparation ne modifie par l'intensité du processus.

2 Estimation par maximum de vraisemblance pour des données filtrées

A partir d'un échantillon d'individus de taille n, tous observés entre t=0 et t=T, l'estimation par maximum de vraisemblance dans le cadre d'un processus de comptage multivarié $N(t) = (N_1(t), \ldots, N_n(t))$, où $N_i(t)$ est le processus de comptage des évènements récurrents de l'individu i, se fait classiquement en maximisant la fonction de vraisemblance (Andersen et al, 1993, sec. II7, ou Lindqvist, 2006)

$$L_{[0,T]}(\alpha, \delta, \beta) = \prod_{i=1}^{n} \Big\{ \pi_{t \in [0,T]} \lambda(t, \theta, Z_i(t))^{dN_i(t)} \exp\big(-\int_0^t \lambda(s, \theta, Z_i(s)) ds\big) \Big\} \quad (2)$$

Cette expression utilise le fait que l'intensité est un processus prédictible dont l'intégrale $\Lambda(t) = \int_0^t \lambda(s)ds$ forme le compensateur du processus de comptage et que ce dernier est entièrement observé sur l'intervalle [0,T]. Cependant, dans notre cas, $L_{[0,T]}$ ne peut être notre fonction de vraisemblance car les processus de comptage ne sont pas observés dans leur totalité.

On suppose donc pouvoir observer les évènements dans un intervalle de temps "calendaire" $[S, E]$ appelé la fenêtre d'observation. Pour chaque système i ($i = 1, \ldots, n$) la date d'installation, la date de dépose, supposée indépendante de la succession des évènements, l'ensemble des covariables Z_i et les m_i (aléatoire) défaillances dans la fenêtre d'observation, sont observées (voir Fig. 1). Notons que dans un premier temps, on suppose pouvoir observer le vecteur de covariables (éventuellement temporelles) Z sur l'intervalle complet $[0, E]$, i.e. la censure n'affecte que les dates de défaillances.

Soit a_i et b_i les âges de l'individu i aux dates S et E. Les conventions de notation sont :

- $a_i = 0$ si le système i est installé dans l'intervalle $[S, E]$;
- $t_{i(m_i+1)} = b_i$;
- $t_{i0} = a_i$.

La fonction de vraisemblance doit être écrite en fonction des processus de comptage $N_i^*(t) = [N_i(t) - N_i(a_i)] 1_{t \in [a_i, b_i]}$ et de leur intensité. On utilise pour cela un théorème d'innovation qui permet de calculer le processus d'intensité λ^* à partir de la filtration "d'observation", $\mathcal{H}^*$ engendrée par la réunion des filtrations individuelles $\sigma(N_i(s) - N_i(a_i), s \in [a_i, b_i])$. Notons que (par exemple pour les cas 1 et 3 de la Fig.1), les ages a_i ne sont pas connus, mais $N_i^*(t) = N_i(t) - N_i(a_i)$ est bien observable pour $t \in [a_i, b_i]$ puisque c'est le nombre de défaillances dans l'intervalle $]a_i, t]$.

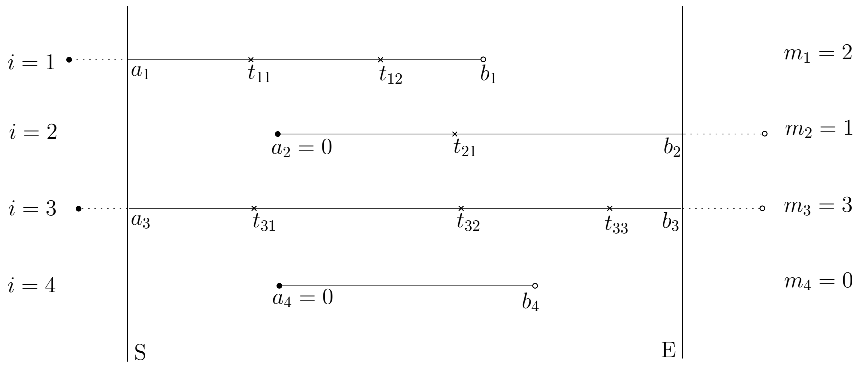

Figure 1: Data collection schema.

Ainsi, d'après Andersen et al (1993) et Martinussen et Scheike (2006, chap.3), on peut écrire le processus d'intensité d'un processus de comptage N filtré par le processus d'observation $C(t) = 1_{[a,b]}(t)$, qui s'écrit $N^*(t) = \int_0^t C(s)dN(s) = N(t) - N(a)$, pour $t \in [a,b]$ comme étant égale à

$$\begin{aligned} \lambda^*(t) &= E\Big[\lambda(t,\theta,Z(t))|N(t)-N(a)\Big] \\ &= \lambda_0(t,\delta)E\Big[(1+\alpha N(t)|N(t)-N(a)\Big]e^{Z'(t)\beta} \end{aligned}$$

Cette écriture permet d'écrire une fonction de vraisemblance cohérente avec les observations en utilisant les processus de comptages N_i^* et leur intensité:

$$L_{[S,E]}(\alpha,\delta,\beta) = \prod_{i=1}^{n}\Big\{\pi_{t\in[a_i,b_i]}\lambda^*(t,\theta,Z_i(t))^{dN_i^*(t)}\exp\Big(\int_0^t \lambda^*(s,\theta,Z_i(s))ds\Big)\Big\} \qquad (3)$$

Pour la maximisation de la fonction de vraisemblance, on peut utiliser un résultat assez remarquable qui permet d'écrire explicitement la log-vraisemblance en fonction des dates de défaillances observées et des paramètres (Le Gat, 2009, Babykina et Couallier, 2010):

$$\begin{aligned} \ln L_{[S,E]}(\alpha,\delta,\beta) &= \sum_{i=1}^{n} m_i \ln\alpha + \ln\Gamma(\alpha^{-1}+m_i) - \ln\Gamma(\alpha^{-1}) \\ &- (\alpha^{-1}+m_i)\ln(e^{\alpha\Lambda_{i0}(b_i)} - e^{\alpha\Lambda_{i0}(a_i)}+1) \\ &+ \sum_{j=1}^{m_i}(\ln\lambda_{i0}(t_{ij}) + \alpha\Lambda_{i0}(t_{ij})) \end{aligned}$$

où $\Lambda_{i0}(t) = \int_0^t \lambda_0(s,\delta)e^{Z_i(s)'\beta}ds$

3 Normalité asymptotique de l'estimateur du maximum de vraisemblance

Un résultat de normalité asymptotique de l'estimateur du maximum de vraisemblance existe dans le cas des processus de comptage totalement observés jusqu'au temps t (voir Borgan 1984, ou Andersen et al, 1993, section VI.1.2). Ici, nous pouvons montrer que, en utilisant les processus de comptages des défaillances *observées* et en les interprétant comme des processus filtrés, on peut écrire le processus d'intensité des processus filtrés à partir de l'intensité donnée en (1) et vérifier une suite de conditions du type de celles trouvées dans Borgan (1984) qui garantissent la convergence en loi quand le nombre d'individu tend vers l'infini.

Ce résultat permet la construction d'intervalles de confiance pour les paramètres ainsi que des tests statistiques d'effet des covariables sur l'intensité des défaillances de type Wald ou rapport de vraisemblance.

Bibliographie

[1] Y. Le Gat. (2009) *Une extension du processus de Yule pour la modélisation stochastique des évènements récurrents.* PhD thesis, ENGREF, 2009.

[2] Pham, H. et Wang, H. (1996) Imperfect maintenance, *European Journal of Operational Research*, 94, 3, 425–438.

[3] Doyen, L. et Gaudoin, O. (2004) Classes of imperfect repair models based on reduction of failure intensity or virtual age, *Reliability Engineering and System Safety*, 84, 1, 45–56.

[4] Krivtsov, V. (2007) Recent advances in theory and applications of stochastic point process models in reliability engineering, *Reliability Engineering and System Safety*

[5] P.K. Andersen, O. Borgan, R.D. Gill and N. Keiding (1993) *Statistical models based on counting process*, Springer-Verlag, New York.

[6] Lindqvist, B.H. (2006) On the Statistical Modeling and Analysis of Repairable Systems, *Statist. Sci.*, 21, 532-551.

[7] Martinussen et Scheike (2006), *Dynamic Regression Models for Survival Data*, Springer New York.

[8] Babykina, G. et Couallier, V. (2010), Modeling Recurrent Events for Repairable Systems Under Worse Than Old Assumption, in *Advances in Degradation Modeling Applications to Reliability, Survival Analysis, and Finance*, Birkhäuser Boston. p. 339-354.

[9] Borgan, O. (1984) Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scand. J. Statist.* 11, 1, 16.