Inférence statistique pour le processus de Wiener avec période d'initiation aléatoire

Christian Paroissin [1]

[1] *Université de Pau et des Pays de l'Adour, Laboratoire de Mathématiques et de leurs Applications - UMR CNRS 5142, Avenue de l'Université, 64013 Pau cedex, France,*
cparoiss@univ-pau.fr

Résumé. Quand on s'intéresse à un phénomène de dégradation, il est généralement supposé que celui-ci démarre dès que le composant est mise en service. Cependant, en pratique, la dégradation peut ne commencer qu'après une période dite d'initiation. On considère ici le processus de Wiener comme modèle de dégradation, commençant à une date aléatoire. On suppose qu'on observe plusieurs trajectoires d'un modèle aux mêmes instants réguliers. On cherche à estimer les paramètres de ce modèle pour ce schéma d'échantillonnage.

Mots-clés. Modèle de dégradation, modèle à retard, taille d'échantillon aléatoire, temps de panne

Abstract. When considering a degradation model, it is often assumed that the degradation begins as soon as the device is put in service. However, in practice, the degradation occurs only after a certain delay called initiation time. We consider here the Wiener process, as a degradation model, starting at a random times. Sample paths are assumed to be observed at the same regular instants. The statistical inference under this sampling scheme is studied here.

Keywords. Degradation processes, Delayed model, Random sample size, Time-to-failure distribution

1 Introduction et modèle

Depuis quelques décennies, des modèles de dégradation ont été proposés et étudiés afin de mieux comprendre le vieillissement de composants (ou de systèmes). Ces modèles sont particulièrement utiles par rapport à des modèles de durées de vie pour des composants/systèmes hautement fiables, par exemple. Par ailleurs, les modèles de dégradation permettent de construire des politiques de maintenance plus sophistiquées que celle basées sur les durées de vie.

Dans la plupart des modèles, il est supposé, implicitement ou pas, que le composant se dégrade dès qu'il est mis en service. Or, en pratique, il arrive, dans certaines situations, que le composant ne commence à se dégrader qu'après une certaine période, appelée initiation. Les articles considérant un modèle de dégradation incluant une période d'initiation sont rares. On peut citer, par exemple, l'article de Guo *et al.* (2013) et celui de Nelson (2010).

Nous considérons ici le modèle de degradation $(X(t))_{t\geq 0}$ défini ainsi:

$$X(t) = [\mu(t-S) + \sigma B(t-S)]\, \mathbb{I}_{t\geq S}.$$

où l'instant $t = 0$ correspond à la date de mise en service du composant, S est une variable aléatoire positive absolument continue indépendante du mouvement brownien $(B(t))_{t\geq 0}$.

Quand on considère un modèle de dégradation, on peut alors définir le temps de panne du composant comme étant le premier instant où le composant dépasse un certain seuil c déterministe et connu. Dans le modèle ci-dessus, le temps de panne du composant est donc la somme de la période d'initiation S et d'une variable aléatoire de loi inverse-gaussienne. Un cas particulier, obtenu en supposant que S est de loi exponentielle, a été étudié par Schwarz (2012) avec une application en psychologie (Schwarz, 2001).

2 Estimation des paramètres du modèle

On suppose qu'on observe n copies $X_1, \ldots, X_n$ du processus stochastique X décrit ci-dessus, à des instants réguliers $0, \delta, 2\delta, \ldots, m\delta = \tau$.

2.1 Notations

Pour le i-ème individu, on note R_i la variable aléatoire à valeurs dans $\mathbb{N}^*$ telle que $(R_i - 1)\delta < S_i \leq R_i\delta$. On distingue alors trois cas qui se révéleront utiles par la suite.

1. Si $R_i > m$, alors $S_i > m\delta = \tau$ et donc $X_i(j\delta) = 0$ pour tout $j \in \{0, \ldots, m\}$: cet individu apporte de l'information uniquement sur la loi de S (censure à droite).

2. Si $R_i = m$, alors $(m-1)\delta < S \leq m\delta$ et $X_i(j\delta) = 0$ pour tout $j \in \{0, \ldots, m-1\}$ mais $X_i(m\delta) \neq 0$. Seule, une mesure non-nulle de dégradation est observée, mais elle ne peut pas être utilisée pour l'estimation de μ et σ^2 : cet individu apporte de l'information uniquement sur la loi de S (censure par intervalle).

3. Si $R_i < m$, alors au moins deux mesures non-nulles de dégradation sont observées et cet individu apporte de l'information utilisable pour estimer tous les paramètres du modèle.

Comme indiqué ci-dessus, les paramètres μ et σ ne peuvent être estimés qu'en utilisant les individus i tels que $R_i < m$. Alors, nous allons également considérer les variables aléatoires $\tilde{R}_i$ telles que la loi $\tilde{R}_i$ est égale à la loi de R_i conditionnellement à $R_i < m$.

On introduit maintenant trois sous-ensembles aléatoires de $\{1, \ldots, n\}$ correspondant aux trois cas considérés ci-dessus. On note $\mathcal{N}_0$ l'ensemble des individus pour lesquels on ne dispose que de mesure nulle de dégradation :

$$\mathcal{N}_0 = \{i; R_i > m\} \subseteq \{1, \ldots, n\}.$$

De même, on note $\mathcal{N}_1$ (resp. $\mathcal{N}_{2+}$) l'ensemble des individus pour lesquels on a exactement une (resp. au moins deux) mesure(s) de degradation non-nulle(s) :

$$\mathcal{N}_1 = \{i; R_i = m\} \subseteq \{1, \ldots, n\}$$

et

$$\mathcal{N}_{2+} = \{i; R_i < m\} \subseteq \{1, \ldots, n\}.$$

On a donc $\mathcal{N}_0 \cup \mathcal{N}_1 \cup \mathcal{N}_{2+} = \{1, \ldots, n\}$. On note N_0, N_1 et N_{2+} les cardinaux de ces trois sous-ensembles.

Puis, on introduit le vecteur aléatoire $\underline{\mathcal{K}} = (\mathcal{K}_r)_{r \in \mathbb{N}^*}$ tel que, pour tout $r \in \mathbb{N}^*$,

$$\mathcal{K}_r = \sum_{i=1}^{n} \mathbb{I}_{(r-1)\delta < R_i \leq r\delta} = \sum_{i=1}^{n} \mathbb{I}_{R_i = r}.$$

On a $|\underline{\mathcal{K}}| = \sum_{r \in \mathbb{N}^*} \mathcal{K}_r = n$. De même que précédemment, on note $\underline{\tilde{\mathcal{K}}} = (\tilde{\mathcal{K}}_r)_{r \in \{1,\ldots,m\}}$ le vecteur aléatoire tel que, pour tout $r \in \{1, \ldots, m\}$,

$$\tilde{\mathcal{K}}_r = \sum_{i=1}^{n} \mathbb{I}_{(r-1)\delta < \tilde{R}_i \leq r\delta} = \sum_{i=1}^{n} \mathbb{I}_{\tilde{R}_i = r}.$$

On a $|\underline{\tilde{\mathcal{K}}}| = N_1 + N_{2+} \leq n$ et $N_0 = n - |\underline{\tilde{\mathcal{K}}}|$.

Pour finir, on considère le nombre aléatoire Q_n de l'ensemble des accroissements non-nuls pour l'ensemble des individus du sous-ensemble $\mathcal{N}_{2+}$:

$$Q_n = \sum_{i \in \mathcal{N}_{2+}} (m - R_i).$$

Cette variable aléatoire discrète est à valeurs dans $\{1, \ldots, (m-1)n\}$.

2.2 Phase d'initiation

Tous les individus apportent une information sur la loi de S, mais la nature de cette information peut être différente : les individus du sous-ensemble aléatoire $\mathcal{N}_0$ vont correspondre à des durées censurées à droite alors que les autres individus vont correspondre à des durées censurées par intervalle.

Considérons que la loi de S est donnée par une famille paramétrique dépendant d'un paramètre inconnu $\theta \in \Theta \subseteq \mathbb{R}^p$. On note $F_S(\cdot;\theta)$ la fonction de répartition de S. La fonction de vraisemblance est alors donnée par :

$$\begin{aligned} L(\theta|\text{data}) &= \prod_{i\in\mathcal{N}_0} \overline{F}_S(\tau;\theta) \prod_{r=1}^{m} \left[\overline{F}_S((r-1)\delta) - \overline{F}_S(r\delta)\right]^{\tilde{\mathcal{K}}_r} \\ &= \left[\overline{F}_S(\tau;\theta)\right]^{N_0} \prod_{r=1}^{m} \left[\overline{F}_S((r-1)\delta) - \overline{F}_S(r\delta)\right]^{\tilde{\mathcal{K}}_r} \end{aligned}$$

On peut alors déterminer (numériquement) l'estimateur du maximum de vraisemblance (il n'y a pas d'expression explicite dans le cas général) :

$$\widehat{\theta}_n = \text{argmax}_{\theta\in\Theta} L(\theta|\text{data}).$$

Pour une approche non-paramétrique, on pourra considérer l'estimateur obtenu par l'algorithme de Turnbull (1976).

2.3 Phase de dégradation

Pour tout $i \in \mathcal{N}_{2+}$ et pour tout $j \in \{1,\ldots,m-R_i\}$, soit $\Delta X_{i,j} = X_i((R_i+j)\delta) - X_i((R_i+j-1)\delta)$. Ces accroissements sont des variables aléatoires i.i.d. de loi normale d'espérance $\mu\delta$ et de variance $\sigma^2\delta$. Un estimateur naturel de μ est le suivant:

$$\widehat{\mu}_n = \frac{\displaystyle\sum_{i\in\mathcal{N}_{2+}} \sum_{j=1}^{m-R_i} \Delta X_{i,j}}{\displaystyle\delta \sum_{i\in\mathcal{N}_{2+}} (m-R_i)} = \frac{1}{\delta Q_n} \sum_{h=1}^{Q_n} Z_h,$$

où $Z_1,\ldots,Z_{Q_n}$ sont les accroissements rangés en utilisant l'ordre lexico-graphique (par exemple), en rappelant que Q_n est le nombre d'accroissements non nuls pour les individus dans $\mathcal{N}_{2+}$. De même, un estimateur naturel de σ^2 est donnée par:

$$\widehat{\sigma}_n^2 = \frac{1}{\delta(Q_n-1)} \sum_{h=1}^{Q_n} (Z_h - \delta\widehat{\mu}_n)^2.$$

Moyennant des résultats sur le comportement asymptotique de Q_n, on montre que ces deux estimateurs possèdent les propriétés asymptotiques usuelles.

3 Estimation de la loi du temps de panne

Le temps jusqu'à panne, défini en introduction, est égal en loi à la convolution de la loi de S et d'une loi inverse-gaussienne correspondant au premier temps de passage du niveau c

par un processus de Wiener (indépendant de S). On peut donc alors facilement estimer par plug-in la loi du temps jusqu'à panne et le temps moyen jusqu'à panne (MTTF).

4 Conclusion

Nous avons présenté ici un modèle de dégradation intégrant une période d'initiation aléatoire. Pour ce modèle, nous avons proposé une estimation des paramètres pour laquelle il est possible d'obtenir des propriétés asymptotiques. Ce modèle reste relativement simple et des extensions plus complexes et réalistes peuvent être proposées. Par exemple, on pourrait supposer que la tendance du processus de Wiener est aléatoire et dépend la période d'initiation et/ou que la dégradation initiale n'est pas nulle (comme supposée dans notre modèle) et est aléatoire pouvant dépendre de S.

Bibliographie

[1] Guo, G.H. et Gerokostopoulos, A. et Liao, H. et Pengying, N. (2013), Modeling et analysis for degradation with an initiation time, *Reliability et Maintainability Symposium (RAMS)*, 1–6.

[2] Nelson, W.B. (2010), Defect initiation, growth, et failure – A general statistical model et data analyses, In M.S. Nikulin, N. Limnios, N. Balakrishnan, W. Kahle et C. Huber-Carol (Eds), *Advances in degradation modeling. Applications to reliability, survival analysis, et finance*, Birkhäuser-Basel.

[3] Schwarz, W. (2001), The ex-Wald distribution as a descriptive model of response time, *Behavior Research Methods, Instruments et Computers*, 33(4), 457–469.

[4] Schwarz, W. (2002), On the convolution of inverse Gaussian et exponential random variables, *Communications in Statistics - Theory et Methods*, 31(12), 2113–2121.

[4] Turnbull, B.W. (1976), The empirical distribution function with arbitrarily grouped, censored et truncated data, *Journal of the Royal Statistical Society B*, 38, 290–295.