Modèle de stress par paliers appliqué à des données censurées par intervalles

Christelle Breuils, Jean-François Masfaraud

▶ **To cite this version:**

Christelle Breuils, Jean-François Masfaraud. Modèle de stress par paliers appliqué à des données censurées par intervalles. 42èmes Journées de Statistique, 2010, Marseille, France, France. inria-00494825

HAL Id: inria-00494825
https://inria.hal.science/inria-00494825

Submitted on 24 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modèle de stress par paliers appliqué à des données censurées par intervalles

Christelle Breuils (1) & J.-F. Masfaraud (2)

(1) LMAM - Département STID - IUT de Metz - Ile du Saulcy 57045 Metz cedex
(2) LIEBE - UPV-M - CNRS UMR 7146 Campus Bridoux - Rue du Général Delestraint - 57070 METZ

Résumé

Le caractère épisodique de la pollution est rarement pris en compte dans les études de toxicité lorsqu'il s'agit d'évaluer la durée de vie des composants organiques soumis à des expositions toxiques. Cependant, l'exposition à la toxicité des composants en pratique met en évidence des niveaux d'exposition fluctuants au cours du temps. C'est dans l'optique d'exhiber la spécificité des durées de vie des composants exposés par intermittence à la toxicité que les expériences ont été menées au sein du LIEBE dans le cadre de la thèse de Beaunoir (2001) encadrée par J.-F. Masfaraud.

Une étude menée au préalable nous permet d'affirmer que les modèles usuels de durées de vie accélérée ne sont pas adaptés à ce type de données.

L'étude met en oeuvre un succession de périodes d'exposition à un composant toxique (le chrome) et de repos. Ainsi le composant biologique (ceriodaphnia dubia) est soumis dans un premier temps à un stress puis dans un second temps à une période de repos.

Supposant sa durée de vie exponentielle, nous avons adapté le modèle de durée de vie accéléré (Bagdonavičius *et al.*, 2000) soumis à un schéma de stress semblable à celui étudié dans Balakhrishnan (2009) mais dans le cas de censures par intervalles. L'estimateur des paramètres de deux lois exponentielles en jeu est obtenu par un maximum de vraisemblance, le calcul de sa loi permet ensuite d'obtenir un intervalle de confiance.

Ce modèle est ensuite appliqué aux données réelles dans le cas premier d'une seule séquence exposition / repos.

Abstract

Episodical character of toxicity phenomenons is rarely taken into account as modelling lifetime function of organical compounds. Indeed experimental studies have underlined a fluctuating time effect in terms of duration or intensity. Experiments have been led in the LIEBE laboratory in order to exhaust and model this effects and was one of the goals of the PhD thesis of V. Beaunoir directed by J.-F. Masfaraud (2001).

A first study shows that traditional lifetimes models are inadequate for this data.

Experiments consist in a succession of couples of exposition and rest periods. The biological compound (ceriodaphnia dubia) is stressed during τ_1 and then rests during τ_2.

Supposing the exponential model within each interval, we adapt the accelerated lifetime model (Bagdonavičius *et al.*, 2000) with a stress scheme similar to Balakhrishnan

(2009) with interval censorship. We estimate the exponential parameters using a maximum likelihood estimator. Computing its law allow us to obtain a confidence interval.

We finally apply this model for our dataset in the case of one sequence exposure/rest and conclude.

Mots-clés : Données de survie et données censurées, biostatistique

Dans les études de toxicité usuelles, le caractère épisodique de la pollution est rarement pris en compte pour évaluer le risque de contamination de l'environnement. En effet, on considère souvent des expériences qui font intervenir des niveaux constants d'exposition alors qu'ils fluctuent au cours du temps en pratique.

Des études ont été menées au sein du laboratoire LIEBE de Metz dans le but de prendre en compte ces spécificités de la toxicité pour mieux la mettre en évidence voire de la modéliser. Ces travaux ont fait l'objet de la thèse de V. Beaunoir (Beaunoir(2001), Masfaraud *et al.* (2002)).

Dans le domaine de la fiabilité, il est possible d'utiliser le modèle de durées de vie pour des composants soumis à des stress. Le but premier est de soumettre le composant à un stress supérieur à celui constaté en utilisation normale (Bagdonavičius *et al.*, 2000) et de modéliser sa durée de vie lorsqu'elle est soumise à un stress usuel.

Il s'agit ici d'un cas particulier de modèle de vie accéléré (voir par exemple Balakhrishnan (2009)) qui consiste à faire varier le niveau de stress au cours du temps à des instants fixés préalablement par l'utilisateur.

On suppose généralement que la durée de vie est distribuée exponentiellement. Balakhrishnan *et al.* (2009) ont obtenu des techniques pour aboutir à un intervalle de confiance du paramètre des lois exponentielles en jeu.

Notre but lors de cette étude est d'appliquer ce modèle à notre cas. Une première remarque ici est la finalité qui diffère légèrement de celle habituellement considérée pour les modèles de stress : en effet, plutôt que de faire varier le niveau de stress, nous considèrerons deux périodes : la période de stress puis celle de repos. On suppose que les composants (ici organismes biologiques) ont une durée de vie qui dépend uniquement de la période.

Les données considérées font l'objet de censure par intervalles : en effet, c'est lors du changement (repos ou stress) que l'expérimentateur a observé les morts des organismes biologiques considérés.

Dans un premier temps, nous adaptons le modèle issu de Balakhrishnan (2009) à notre cas (la censure considérée dans l'article ainsi que le schéma d'exposition au stress sont différents) puis nous en déduisons un estimateur du paramètre de la loi exponentielle.

Dans une seconde partie, le calcul de la loi exacte de cet estimateur nous permettra de considérer un intervalle de confiance pour les paramètres qui répondra à la double mission : avoir une idée de sa précision bien sûr mais aussi pouvoir juger empiriquement du bien-fondé du modèle.

Enfin, nous appliquons ce modèle aux données proposées.

1 Estimateur du maximum de vraisemblance

On suppose que la durée de vie des composants est exponentielle et que le paramètre de cette loi dépend de la phase d'exposition en cours (exposition ou repos). Le composant est soumis successivement à une phase dite d'exposition (sa durée de vie T est supposée exponentielle de paramètre θ_1, de fonction de répartition F_1) puis de repos (T est exponentielle de paramètre θ_2 de fonction de répartition F_2). L'expérience prend fin lors d'une période de repos, à un temps T_0 fixé par l'expérimentateur. Ainsi, les composants vivants à la fin de l'expérience font l'objet d'une censure à droite.

Nous considérons le modèle d'exposition cumulée qui suppose que la durée de vie des composants dépend uniquement de la quantité de stress accumulée à l'instant t sans mémoire de la fonction d'exposition. La continuité de la fonction de répartition permet d'obtenir h tel que $F_2(\tau_1 + h) = F_1[\tau_1)$, où τ_1 est la durée d'exposition.

La séquence exposition-repos est répétée r fois. À chaque changement de phase, l'expérimentateur observe le nombre de composants morts. En supposant que la phase d'exposition i a une durée τ_{2i+1} et la phase de repos i a une durée τ_{2i}, les données sont censurées par intervalles et les bornes de ces derniers sont les $t_i, i = 1 \dots r$ définis par $t_0 = 0, t_1 = \tau_1, t_2 = \tau_1 + \tau_2, \dots, t_{2r} = \tau_1 + \dots + \tau_{2r}$.

La fonction de vraisemblance en les observations est alors donnée par

$$\begin{aligned} L(\theta_1, \theta_2, n_1, \dots, n_{2r}) &= \prod_{i=0}^{r} \left(\exp\left(\frac{-t_{2i}}{\theta_2}\right) \left(1 - \exp\left(\frac{-\tau_{2i+1}}{\theta_1}\right)\right)\right)^{N_{2i}} \\ &\dots \left(\exp\left(\frac{-t_{2i+1}}{\theta_1}\right) \left(1 - \exp\left(\frac{-\tau_{2i+2}}{\theta_2}\right)\right)\right)^{N_{2i+1}} \exp\left(\frac{-t_{r+1}}{\theta_1}\right)^{N_{r+1}}, \end{aligned}$$

où N_i réprésente le nombre de composants qui défaillent dans l'intervalle $[t_{i-1}, t_i[$.

Annulant les dérivées partielles de la log-vraisemblance, on obtient une équation analytique pour le point critique $(\hat{\theta_1}, \hat{\theta_2})$. Supposant les durées d'exposition constantes égales à τ_1 ainsi que les durées de repos constantes égales à τ_2, cela nous permet d'obtenir une formule explicite pour l'estimateur du maximum de vraisemblance pour le paramètre $\theta = (\theta_1, \theta_2)$.

Ainsi, on obtient :

$$(\hat{\theta_1}, \hat{\theta_2}) = \left(\frac{-\tau_1}{\ln\left(1 - \frac{\sum_{i=1}^{r} N_{2i}}{\sum_{i=1}^{r} N_{2i}\tau_1 + N_{2i+1}t_{2i+1} + N_{r+1}t_{r+1}}\right)}, \frac{-\tau_2}{\ln\left(1 - \frac{\sum_{i=0}^{r} N_{2i+1}\tau_2}{\sum_{i=0}^{r} N_{2i}t_{2i}}\right)} \right),$$

sous réserve d'existence. Les conditions d'existence de cet estimateur supposent que les morts sont réparties dans chaque intervalle.

On obtient ainsi un estimateur du paramètre de la loi exponentiel.

2 Calcul de la loi exacte

Dans ce paragraphe, nous nous restreignons au cas où $r = 1$. Cela nous permet d'obtenir une première formulation pour l'estimateur de θ :

$$(\hat{\theta_1}, \hat{\theta_2}) = \left(\frac{-\tau_2}{\ln\left(\frac{n-N_1-N_2}{n-N_1}\right)}, \frac{-\tau_1}{\ln\left(\frac{n-N_1}{n}\right)} \right),$$

sous réserve d'existence , c'est-à-dire conditionnellement à la survenue de $A = \{1 \leq N_1 \leq n-1; 0 \leq N_2 \leq n-N_1\}$. Les conditions d'existence sont relativement semblables à celles obtenues dans Balakrishnan (2009).

Nos données étant censurées par intervalles, il apparaît que les temps de défaillance n'apparaissent pas dans l'expression de notre estimateur. Cela nous permet ainsi de constater naturellement que seule la loi des N_i nous permet d'obtenir celle de $\hat{\theta}$. On remarque que les N_i suivent la même loi que celle obtenue dans Balakrishnan et Qihao (2009), même si le schéma de censure diffère. Les N_i suivent ainsi une loi trinomiale de paramètre p_1, p_2, p_3, les p_i correspondant aux probabilités de défaillance dans chacun des intervalles successifs (le dernier étant celui de la censure à droite).

Notre but est d'obtenir un intervalle de confiance pour θ à partir de la loi obtenue et nous présenterons les applications aux données expérimentales.

Nous avons donc obtenu un modèle permettant de prendre en compte les différentes phases caractéristiques des données de toxicité dont nous disposons. Ces données ont la particularité d'être censurées par intervalles, ce qui rend l'observation exacte de la durée de vie d'un composant impossible. Notre but à venir serait de généraliser l'intervalle de confiance obtenu au cas de plusieurs séquences exposition-repos.

3 Application aux données expérimentales

L'organisme biologique considéré est le ceriodaphnia dubia et les expérimentations ont été réalisées selon un protocole garantissant leur longévité (qui doit excéder la durée de l'expérience), ce qui implique que les morts peuvent être considérées comme due à l'apport toxique apporté par le chrome. Chaque expérience a une durée de 168 heures durant lesquelles les composants organiques sont soumis à une succession de périodes d'exposition au chrome puis de repos. La concentration de composant toxique ainsi que les durées des plages d'exposition et de repos varient. Considérant ces trois paramètres fixés, on dispose de 4 expériences mettant en jeu 20 organismes. Notre objectif sera double : il s'agit de savoir si le paramètre de la loi exponentielle au cours de la période de stress est supérieur à celui de la période de repos mais aussi de voir s'il varie avec le temps.

Notons que la durée de vie d'un organisme fait systématiquement l'objet d'une censure (par intervalle ou à droite si l'organisme considéré survit après l'étude). Les durées des périodes d'exposition étant faibles par rapport aux durées des périodes de repos mais

aussi à la durée de totale de l'expérience, les données peuvent être considérées comme relativement précises dans le cas où l'organisme est défaillant au cours d'un intervalle d'exposition, ce qui est le cas le plus fréquent.

Bibliographie

[1] Balakhrishnan, N. (2009) A synthesis of exact inferential results for exponential step-stress models ans associated optimal accelerated life-tests. *Metrika*, 69, 351–396.
[2] Balakhrishnan, N., Qihao, X. et Kundu, D. (2009) Exact inference for a simple step-stress model from the exponential distribution under time constraint. *Ann. Inst. Stat. Math.*, 61, 251–274.
[3] Masfaraud, J.-F., Beaunoir, V., Cotelle, S. et Ferard, J.-F. (2002) Ceriodaphnia dubia survival analysis related to intermittent exposures to chromium. *SETAC Europe 12th Annual meeting*, 12-16 may 2002, Vienne, Austria.
[4] Beaunoir, V. (2001) Mobilisation des polluants du sol et pollution épisodique : modélisation des relations entre exposition intermittente et effets sur deux modèles biologiques : Lemna minor et Ceriodaphnia dubia. *Thèse de l'Université de Metz.*
[5] Bagdonavičius, V., Gerville-Réache, L., Nikoulina, V. et Nikulin, M. (2000) Expériences accélérées : analyse statistique du modèle standard de vie accélérée. *Revue de statistique appliquée*, tome 48, 3, 5–38.