JOURNAL DE LA SOCIÉTÉ FRANÇAISE DE STATISTIQUE

GILLES CELEUX

Situations de maintenance à structure de données incomplètes

Journal de la société française de statistique, tome 141, n° 3 (2000), p. 43-59

<http://www.numdam.org/item?id=JSFS_2000__141_3_43_0>

SITUATIONS DE MAINTENANCE À STRUCTURE DE DONNÉES INCOMPLÈTES

Gilles CELEUX [1]

RÉSUMÉ

Nous montrons à travers deux exemples comment l'interprétation de problèmes de maintenance préventive dans les termes de modèles à structure cachée peut faciliter leur résolution par le maximum de vraisemblance. Le premier exemple concerne l'élimination d'un biais pessimiste dans l'évaluation de la durée de vie d'un matériel dû à un comportement trop préventif lors de sa maintenance. Le deuxième exemple concerne l'estimation d'un modèle de vieillissement prenant en compte l'instant de démarrage du vieillissement Par ailleurs, nous évoquons, en conclusion, les problèmes statistiques posés dans des situations courantes où les données disponibles sont faiblement informatives Et, nous indiquons, notamment, comment alors l'inférence bayésienne peut être préférée au maximum de vraisemblance. De plus, une annexe donne une présentation synthétique de l'algorithme EM.

Mots-clés . données censurées à gauche et à droite, facteur humain, vieillissement, algorithme EM, loi de Weibull, inférence bayésienne.

ABSTRACT

We hightlight the interest of embedding some problems of preventive maintenance in the framework of hidden structure models with two examples. This point of view facilitates the maximum likelihood estimation of the parameters of interest via the EM algorithm The first example concerns the reduction of a pessimist bias due to a too cautious behavior during a maintenance process. The second example concerns the estimation of an ageing model incorporating a change point in the lifetime distribution. Moreover, we discuss the statistical difficulties involved by ill-posed maintenance problems and we point how Bayesian inference can be useful in such a context Finally we give in an appendix a synthetic presentation of the EM algorithm.

Key words : left and right censored lifetimes, human factor, ageing, EM algorithm, Weibull distribution, Bayesian inference.

1. Inria Rhône-Alpes, ZIRST 655 avenue de l'Europe, Montbonnot Saint Martin, F38334 Saint ISMIER CEDEX, e-mail Gilles.Celeux@inria.fr

1. INTRODUCTION

Les opérations de maintenance préventive sur un matériel se décrètent souvent à partir de l'analyse des données de retour d'expérience. Cette analyse lorsqu'elle est effectuée à partir de données fiables et nombreuses permet de bien cerner la loi de durée de vie du matériel et doit ainsi conduire à une politique de maintenance pertinente. Malheureusement, de par leur nature même les données de retour d'expérience sont souvent peu nombreuses, lacunaires, imprécises voire partiellement erronées. Elles sont ainsi d'une exploitation difficile. Pourtant, elles sont porteuses d'informations objectives et utiles, et le défi posé au statisticien est de parvenir à en tirer des éléments judicieux et bénéfiques pour organiser la maintenance du matériel malgré leurs déficiences.

Dans ce texte, nous illustrons des outils d'analyse statistique pour le traitement de données de retour d'expérience faiblement informatives. Tout d'abord nous montrons à travers deux exemples, présentés dans les deux sections qui suivent, comment la considération de modèles à structure de données incomplètes, identifiés par l'algorithme EM (Dempster, Laird et Rubin 1977), permet de répondre correctement à des problèmes susceptibles de se poser en maintenance préventive. Puis, dans une dernière section, nous indiquons comment le paradigme bayésien peut être utile pour obtenir des résultats d'analyse statistique fiables malgré des données de retour d'expérience rares et peu informatives, et comment il est parfois préférable de se rabattre sur un point de vue qualitatif en de telles circonstances. Enfin, vu son importance dans notre démarche, nous donnons en annexe une présentation synthétique de l'algorithme EM.

2. UN COMPORTEMENT TROP PRÉVENTIF

Nous considérons ici la situation de maintenance suivante. Des matériels identiques, constitués chacun de différents composants, sont contrôlés périodiquement. Les composants sont susceptibles de subir des maladies ne provoquant pas de défaillance, mais entraînant un fonctionnement en mode dégradé du matériel. Ainsi lors d'un contrôle, on remplace les composants malades et à son issue le matériel est considéré *aussi bon que neuf*. Le but est d'étudier la loi de dégradation du matériel. On suppose qu'il s'agit d'une loi exponentielle caractérisée par un taux de dégradation λ et une moyenne $\eta = 1/\lambda$ à estimer.

Pour ce faire, on ne dispose que de données censurées à droite (si aucun composant n'a été déclaré malade lors d'un contrôle) ou à gauche (dans le cas contraire). On ne dispose jamais du temps exact de dégradation. Mais il y a plus grave. Le matériel est de conception relativement nouvelle, et les ingénieurs de maintenance avaient, avant une date connue d_0, une tendance avérée à mettre au rebut le matériel par précaution. Par contre, ils se sont départis de cette attitude trop préventive après la date d_0. Cette possibilité

de mise au rebut par précaution entâche sérieusement les données disponibles et s'il n'est pas pris en compte induit un biais important (pessimiste) dans l'estimation du taux de défaillance λ. La figure 1 donne un exemple d'un tel comportement trop préventif. Ici la date d_0 est 1992.

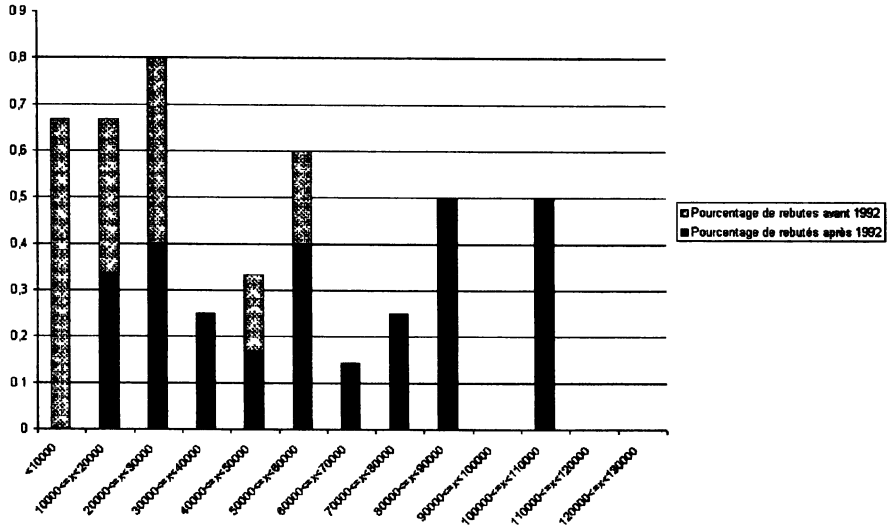

FIG 1. — Pourcentage des rebutés / censurés à droite au cours du temps.

Nous allons voir qu'il est possible de s'affranchir de ce problème de maintenance par un modèle où le facteur humain est envisagé comme une variable cachée. On pourra alors le résoudre de manière satisfaisante par l'algorithme EM qui fournit l'estimateur du maximum de vraisembance par maximisation itérative de l'espérance conditionnelle de la log-vraisemblance des données complétées sachant les données observées et une valeur courante des paramètres. Plus précisément dans le cas présent, les donnés complètes sont la date d_0 et $(t_i, \delta_i, z_i), i = 1, \ldots, n$ où n désigne le nombre total de contrôles, et au $i$ème contrôle, t_i désigne le temps du contrôle (c'est-à-dire le temps écoulé depuis le précédent contrôle ou la mise en service du matériel), $\delta_i \in \{0, 1\}$ est l'indication de rebut ($\delta_i = 1$ en cas de rebut), $z_i \in \{0, 1\}$ est l'indication d'un rebut par précaution ($z_i = 1$ si le rebut n'est pas justifié). La structure de données incomplètes est la suivante :

- Les données observées sont la date d_0, les temps de contrôle t_i et l'indication δ_i de rebut pour $i = 1, \ldots, n$.

- Les données manquantes ou cachées sont les indications z_i d'un rebut par précaution, i représentant un contrôle avant la date d_0 ayant donné lieu à un rebut.

La vraisemblance complète du paramètre η de la loi exponentielle régissant la durée entre deux dégradations s'écrit, en notant n_0 le nombre de contrôles

avant la date d_0 et en remarquant que ce sont les n_0 premiers contrôles,

$$L(\eta;(t_i,\delta_i,z_i)_i) = \left(\prod_{i=n_0+1}^{n} F(t_i \mid \eta)^{\delta_i} R(t_i \mid \eta)^{1-\delta_i}\right) \left(\prod_{i=1}^{n_0} \left\{R(t_i \mid \eta)^{z_i} F(t_i \mid \eta)^{1-z_i}\right\}^{\delta_i} R(t_i \mid \eta)^{1-\delta_i}\right),$$

les fonctions F et R désignant respectivement la fonction de répartition et la fonction de survie de la loi de durée de vie. Dans le cas exponentiel qui nous occupe, on a $F(x \mid \eta) = 1 - e^{-\frac{x}{\eta}}$ et $R(x \mid \eta) = e^{-\frac{x}{\eta}}$.

Il est important de remarquer qu'un rebut par précaution revient à remplacer une censure à droite par une censure à gauche :

$$P(z_i = 1 | t_i, \delta_i = 1, \eta) = R(t_i|\eta).$$

Ainsi, les deux étapes de l'algorithme EM prennent la forme suivante :

- L'étape E consiste à calculer l'espérance conditionnelle de la log-vraisemblance complétée sachant les données observées et la valeur courante du paramètre, c'est-à-dire (r désignant l'indice d'itération)

$$Q(\eta \mid \eta^r) = \sum_{i=1}^{n_0} \{-\delta_i R(t_i \mid \eta^r)\frac{t_i}{\eta} + \delta_i \left(1 - R(t_i \mid \eta^r)\right) \log(1 - e^{-\frac{t_i}{\eta}}) - (1-\delta_i)\frac{t_i}{\eta}\} + \sum_{i=n_0+1}^{n} \{\delta_i \log(1 - e^{-\frac{t_i}{\eta}}) - (1-\delta_i)(\frac{t_i}{\eta})\}.$$

- L'étape M consiste à maximiser en η la fonction $Q(\eta \mid \eta^r)$, ce qui se fait par la résolution itérative de l'équation de vraisemblance, obtenue par annulation de la dérivée de $Q(\eta \mid \eta^r)$ par rapport à $\lambda = 1/\eta$,

$$\left.\begin{aligned} &\sum_{i=1}^{n_0} \{-\delta_i R(t_i \mid \eta^r) t_i + \left(1 - R(t_i \mid \eta^r)\right) \delta_i \left(\frac{t_i e^{-\frac{t_i}{\eta}}}{1 - e^{-\frac{t_i}{\eta}}}\right) - (1-\delta_i) t_i\} \\ &+ \sum_{i=n_0+1}^{n} \{\delta_i \left(\frac{t_i e^{-\frac{t_i}{\eta}}}{1 - e^{-\frac{t_i}{\eta}}}\right) - (1-\delta_i) t_i\} \end{aligned}\right\} = 0.$$

À la convergence, l'algorithme EM fournit une estimation de η débarrassée de l'effet néfaste du biais de précaution. De plus, on peut considérer comme

sous-produit que la quantité

$$Q_{\hat{\eta}} = \frac{\sum_{i=1}^{n_0} R(t_i \mid \hat{\eta}, \delta_i = 1)}{\text{card}\{i \mid i \leqslant n_0,\ \delta_i = 1\}}$$

constitue une évaluation du pourcentage de rebuts par précaution.

L'initialisation de l'algorithme EM est souvent une étape importante et délicate, cet algorithme produisant souvent des résultats sensibles à sa position initiale. Ici, cette initialisation ne pose pas de problème, car le choix consistant à partir de η^0 estimateur du maximum de vraisemblance sous l'hypothèse que tous les z_i sont nuls s'impose. En effet, il consiste à supposer que tous les rebuts sont justifiés, hypothèse qui est bien celle de référence. De plus, cette position initiale peut être sans inconvénient grossièrement approchée par

$$\eta^0 = \frac{\sum_{i=1}^{n} t_i}{\sum_{i=1}^{n} \delta_i}.$$

Les résultats pour les données de la figure 1 sont résumés dans le tableau 1. On voit que la mise au rebut avant 1992 est systématique. Ce tableau donne les valeurs de l'estimation η_m lorsque tous les rebuts avant 1992 sont considérés comme de vrais rebuts et η_M qui correspond à l'estimation obtenu si tous les rebuts d'avant 1992 sont considérés comme rebuts par précaution. L'estimateur η_{EM} fourni par l'algorithme EM produit une durée de vie moyenne sans dégradation presque double de l'estimation prenant en compte tous les rebuts. De plus, le pourcentage de rebuts d'avant 1992 considérés mal fondés est important (plus de 90 %).

TABLEAU 1. — Estimation de η pour les données de la figure 1.

d_0	% rebuts avant d_0	% rebuts après d_0	η_m	η_M	η_{EM}	Proba. rebut par précaution
1992	100	20	16800	35169	32332	0.917

Nous avons également évalué les performances de notre procédure sur des données simulées. Nous avons considéré des échantillons de taille $n = 70$, les dates de contrôle étant les mêmes que celles de l'exemple réel traité ci-dessus. Nous avons choisi la date de changement de comportement d_0 de telle sorte que $n_0 = 30$. Dans ce cadre, nous avons simulé des réalisations d'une loi exponentielle avec différentes valeurs de η et nous avons simulé des rebuts par précaution de la manière suivante, à partir des indicatrices h_i que la réalisation de la loi exponentielle est plus petite que le temps de contrôle associé pour $i = 1, \ldots, n$:

- si $i \leqslant n_0$ et $h_i = 0$, on tire la variable z_i selon une loi de Bernoulli de paramètre $R(t_i \mid \eta)$, et on pose $\delta_i = z_i$.

- si $i \leqslant n_0$ et $h_i = 1$, ou si $i > n_0$, on pose $z_i = 0$ et $\delta_i = h_i$.

Chaque situation a été répétée 100 fois. Le tableau 2 donne les estimations moyennes obtenues pour le paramètre η. La première colonne donne la valeur de η utilisée pour simuler les données. Les résultats sont assez satisfaisants même si notre procédure semble avoir tendance à surestimer la durée de vie moyenne de la loi exponentielle.

TABLEAU 2. — Estimation de η obtenu à partir de données simulées.

η	η_{EM}	η_m	η_M
20000	21473	15282	61994
30000	32737	23298	70817
40000	44869	31911	81592
60000	71202	48578	108458
100000	122192	79454	160900

3. UN MODÈLE DE VIEILLISSEMENT DIFFÉRÉ

L'un des grands problèmes à résoudre pour optimiser une maintenance préventive est de diagnostiquer à temps un vieillissement apparaissant sur un matériel pour y remédier sans retard. Dans le cadre des matériels non réparables la loi de Weibull est souvent utilisée pour modéliser un éventuel vieillissement (cf. par exemple Bacha *et al.* 1998). Cependant elle est mal adaptée pour détecter l'instant de démarrage d'un vieillissement car elle suppose que le vieillissement commence dès le début de vie du matériel. Dans cette section, nous présentons un modèle qui permet de décrire des matériels dont les défaillances sont accidentelles jusqu'à une date t_0, et peuvent soit être accidentelles soit provenir du vieillissement du matériel si elles ont lieu après cette date (cf. Bertholon 2000).

La durée de vie X du matériel est modélisée par une loi dont le taux de défaillance, représenté figure 2, est choisi de la manière suivante : il est constant, égal à $\frac{1}{\eta_0}$, jusqu'à un instant de vieillissement noté t_0 (ce qui traduit une absence de vieillissement jusqu'à cette date), puis il augmente suivant une fonction de la forme $\frac{1}{\eta_0} + \frac{\beta}{\eta_1}(\frac{x-t_0}{\eta_1})^{(\beta-1)}$ ce qui traduit alors un vieillissement. On doit noter que le second terme correspond au taux de défaillance d'une loi de Weibull de paramètre de forme β et de paramètre d'échelle η_1.

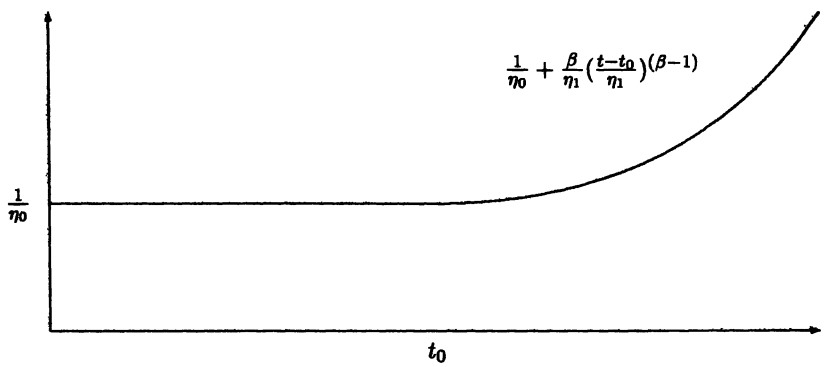

FIG 2. — Taux de défaillance associé au modèle de vieillissement différé.

Ce taux de défaillance ainsi défini est donc la somme de deux fonctions : l'une constante égale à $\frac{1}{\eta_0}$ et l'autre nulle jusqu'à t_0 puis ensuite égale à $\frac{\beta}{\eta_1}(\frac{x-t_0}{\eta_1})^{(\beta-1)}$.
Il s'ensuit que X peut s'écrire sous la forme $X = \min(E, W)$ où

- E est une variable aléatoire de loi exponentielle de paramètre d'échelle η_0 (qui correspond à un taux de défaillance constant égal à $\frac{1}{\eta_0}$),

- W est une variable aléatoire, indépendante de E, de loi de Weibull de paramètre de forme β, de paramètre d'échelle η_1 et de paramètre de position t_0.

Notant $\theta = (\eta_0, \beta, \eta_1)$, la densité de X s'écrit

$$f(x \mid \theta, t_0) = \begin{cases} \frac{1}{\eta_0} e^{-\frac{1}{\eta_0}x} \text{ si } x \leqslant t_0 \\ \left[\frac{1}{\eta_0} + \frac{\beta}{\eta_1}\left(\frac{x-t_0}{\eta_1}\right)^{\beta-1}\right] e^{-\frac{1}{\eta_0}x - \left(\frac{x-t_0}{\eta_1}\right)^{\beta}} \text{ si } x \geqslant t_0. \end{cases}$$

Lorsque les données ne sont pas censurées, la vraisemblance du paramètre vectoriel (θ, t_0) s'écrit

$$L(\theta, t_0) = \prod_{i=1}^{n} f(x_i \mid \theta, t_0).$$

Ici la densité n'ayant pas la même forme avant et après t_0, il faut distinguer la position de t_0 par rapport aux n durées de vie observées $x_1, ..., x_n$. Pour des raisons de simplicité de calcul, t_0 est pris sur l'une de ces durées x_i. Notant $x_{(1)} < \ldots < x_{(n)}$ l'échantillon ordonné, cela donne n formes possibles pour la

vraisemblance, notées $L_i(\theta)$ lorsque $t_0 = x_{(i)}$. $L_i(\theta)$ s'écrit alors :

$$L_i(\theta) = \left(\frac{1}{\eta_0}\right)^i \times e^{-\frac{1}{\eta_0}\sum_{j=1}^{i} x_{(j)}} \times \left[\prod_{j=i+1}^{n}\left(\frac{1}{\eta_0} + \frac{\beta}{\eta_1}\left(\frac{x_{(j)} - x_{(i)}}{\eta_1}\right)^{\beta-1}\right)\right]$$
$$\times e^{-\frac{1}{\eta_0}\sum_{j=i+1}^{n} x_{(j)} - \sum_{j=i+1}^{n}\left(\frac{x_{(j)}-x_{(i)}}{\eta_1}\right)^{\beta}}.$$

L'estimation du maximum de vraisemblance de (θ, t_0) consiste à résoudre $\max_i \max_\theta L_i(\theta)$. La maximisation en i est bien sûr triviale, mais la maximisation en θ est difficile. À l'instar de Bacha *et al.* (1998) pour des systèmes à risques masqués fonctionnant en série, nous proposons d'utiliser l'algorithme EM pour maximiser $L_i(\theta)$ en θ.

Ici, les données manquantes sont les indicateurs z_i qu'une défaillance est accidentelle (réalisation de la loi E, $z_i = 1$) ou qu'elle soit due au vieillissement (réalisation de la variable W, $z_i = 2$). Ainsi, l'itération $r+1$ de l'algorithme EM prend la forme suivante :

Étape E : Pour tout i, calcul de $p_j(t_i|\theta^r, t_0), j = 1, 2$, probabilité que le composant 1 (E) ou 2 (W) soit responsable de la défaillance :

$$p_2(t_i|\theta^r, t_0) = \begin{cases} 0 & \text{si } t_i \leqslant t_0 \\ \dfrac{r(t_i|\theta^r, t_0)}{1/\eta_0^r + r(t_i|\theta^r, t_0)} & \text{si } t_i > t_0. \end{cases}$$

avec $r(t_i|\theta, t_0) = \dfrac{\beta}{\eta_1}\left(\dfrac{t_i - t_0}{\eta_1}\right)^{\beta-1}$. Et, $p_1(t_i|\theta^r, t_0) = 1 - p_2(t_i|\theta^r, t_0)$.

Étape M :

$$\eta_0^{r+1} = \frac{\sum_{i=1}^{n} t_i}{\sum_{i=1}^{n} p_1(t_i|\theta^r, t_0)}$$

$$\frac{1}{\beta^{r+1}} + \frac{\sum_{i=1}^{n} p_2(t_i|\theta^r, t_0)\log(t_i - t_0)}{\sum_{i=1}^{n} p_2(t_i|\theta^r, t_0)} - \frac{\sum_{i=1}^{n}(t_i - t_0)^{\beta^{r+1}}\log(ti - t_0)}{\sum_{i=1}^{n}(t_i - t_0)^{\beta^{r+1}}} = 0$$

$$\eta_1^{r+1} = \left[\frac{\sum_{i=1}^{n}(t_i - t_0)^{\beta^{r+1}}}{\sum_{i=1}^{n} p_2(t_i|\theta^r, t_0)}\right]^{\frac{1}{\beta^{r+1}}}$$

Remarque : ces formules restent valables avec des données censurées à droite avec les particularités suivantes.

- *Étape E* : les probabilités $p_j(t_i|\theta^r, t_0), j = 1, 2$ ne sont pas à calculer pour les données censurées.
- *Étape M* : les sommes faisant intervenir les probabilités $p_j(t_i|\theta^r, t_0), j = 1, 2$ sont à considérer uniquement sur les i non censurées.

La vision de ce modèle de vieillissement comme un modèle à structure cachée rend son estimation beaucoup plus facile. Une estimation directe de la vraisemblance est difficile et en général nécessite de fixer l'un des paramètres (par exemple le paramètre de forme β) pour donner des résultats fiables, ce qui est une limitation importante (cf. Bertholon 2000).

Cependant, en pratique, l'estimation de ce modèle par le maximum de vraisemblance est mise en danger par des censures à droite fortes et des données rares et, dans de tels cas, il vaut mieux se placer dans un cadre bayésien analogue à celui décrit dans Bacha *et al.* (1998) pour des systèmes à risques masqués. Mais là aussi, l'interprétation de ce modèle comme un modèle à structure cachée facilite grandement les calculs bayésiens, comme on l'évoque dans la discussion qui suit.

4. DISCUSSION

Nous avons donné deux exemples de situations intervenant en fiabilité où il est utile de considérer un modèle à structure cachée pour résoudre par le maximum de vraisemblance le problème à traiter. Dans ces deux cas, la structure cachée n'était pas apparente. Il existe bien des situations où la structure cachée n'est pas sous-jacente, mais apparaît directement dans la définition du problème. C'est par exemple le cas pour les modèles à risques masqués (*competing risk models*) considérés dans Bacha *et al.* (1998). Ainsi, l'algorithme EM constitue un outil intéressant pour estimer des modèles statistiques utiles dans une optique d'optimisation de la maintenance préventive. Cependant l'usage de cet algorithme et plus généralement du maximum de vraisemblance présentent, dans le contexte de la fiabilité, des limitations que nous voudrions maintenant évoquer.

Simulation des données censurées. De manière presque systématique dans un contexte industriel, beaucoup de données de durées de vie sont censurées à droite et parfois, comme on l'a vu dans la section 2, des données censurées à gauche sont présentes. De la sorte, on se trouve intrinséquement dans un contexte à structure de données incomplètes, les données manquantes étant les durées de vie au-delà du temps de censure pour les données censurées à droite, ou les temps entre l'occurence d'une défaillance et le temps de censure pour les données censurées à gauche. En conséquence, on pourrait utiliser le formalisme de l'algorithme EM, voire simuler les durées de vie

manquantes par sa variante stochastique, l'algorithme SEM, pour estimer les paramètres du maximum de vraisemblance de différents modèles. En réalité, ce type de stratégie ne s'avère pas bénéfique. Soit, elle ne fait que retomber sur des procédures d'estimation déjà connues, soit elle produit des procédures d'estimation induisant une trop grande variabilité des résultats. On pourra trouver dans Bacha (1996) de nombreuses illustrations de ce fait.

Inférence bayésienne. Dans un contexte industriel, non seulement les données sont souvent censurées, mais de plus elles sont souvent rares. Dans de telles circonstances, le maximum de vraisemblance est susceptible de fournir des estimateurs peu fiables vu leur grande variabilité. Par contre, l'absence de données objectives est souvent palliée par des avis d'experts qui synthétisent les expériences passées. Cet état de fait invite à placer les problèmes d'estimation dans le cadre de l'inférence bayésienne (cf. Robert 1992, ou Proccacia 1992 pour une présentation des méthodes bayésiennes dans le contexte de la fiabilité). L'utilisation de l'analyse statistique bayésienne est de plus facilitée par l'émergence des méthodes de Monte-Carlo par chaînes de Markov, dites MCMC, qui autorisent par des simulations de chaînes de Markov l'approximation de lois de probabilité difficilement réalisable par des moyens numériques traditionnels (cf. Gilks, Richardson et Spiegelhalter 1996 ou Robert 1996). On trouvera ainsi, dans Bacha *et al.* (1998) des exemples d'utilisation de l'inférence bayésienne pour estimer des systèmes à risques masqués, correspondant à des composants montés en série suivant des lois de Weibull, pour lesquels le composant responsable de la panne est inconnu. Ces exemples montrent nettement qu'une inférence bayésienne, fondée sur des lois a priori raisonnables pour les paramètres des lois de Weibull des composants, conduit à des résultats hautement préférables à une approche par maximum de vraisemblance dans la plupart des cas. Un autre exemple d'application où l'inférence bayésienne via les méthodes MCMC s'est montrée particulièrement utile est celui de l'analyse de défauts intervenants sur les cuves à eau pressurisée des centrales nucléaires (Celeux *et al.* 1999). Dans un but de prévision, il s'agissait de modéliser la loi de la hauteur de ces défauts. Mais le contexte était extrêmement difficile : peu de défauts, mal mesurés, parfois non détectés, les défauts en deçà d'un certain seuil n'étant pas mesurés, mais juste notifiés. Les contraintes étaient telles que l'inférence par le maximum de vraisemblance était numériquement impossible sans introduire des hypothèses très restrictives. Par contre, l'utilisation des méthodes MCMC, et notamment de l'échantillonnage de Gibbs, qui s'appuie sur la simulation des lois conditionnelles des paramètres et des données manquantes en jeu, autorisent une inférence bayésienne de bonne qualité, même si l'évaluation de la convergence de l'échantillonneur de Gibbs doit être menée avec soin et constitue un point délicat (cf. Robert 1998). Par ailleurs, il est important de souligner que, dans le cadre de la fiabilité, l'inférence bayésienne n'est pas seulement utile dans un contexte où des opinions d'experts sont disponibles. L'analyse bayésienne peut être vue comme une procédure cohérente de régularisation et de stabilisation des résultats. Par exemple, nous avons montré (Celeux, Lavergne et Vernaz 2000) qu'il

était possible d'obtenir une estimation des paramètres d'une loi de Weibull, uniquement à partir de données censurées à droite ou à gauche, beaucoup plus fiable avec une inférence bayésienne fondée sur des lois a priori non informatives (cf. Sun 1997) que par le maximum de vraisemblance.

Absence dramatique de données. En réalité, il existe des cas où même l'inférence bayésienne n'est d'aucun secours. Ces cas se produisent lorsque les données de retour d'expérience sont très peu nombreuses et très censurées. Souvent, il paraît alors illusoire de vouloir utiliser les modèles classiques de durées de vie et il est préférable de se rabattre sur la recherche de réponses qualitatives aux questions qui se posent. On trouve par exemple dans Celeux, Lavergne et Vernaz (2000), un exemple d'une telle situation. Comme dans la section 2, on ne dispose que de données censurées à droite et à gauche obtenues lors de contrôles périodiques. Le but est de déterminer, dans un souci d'optimisation de maintenance, si un vieillissisement est intervenu entre deux contrôles. De par la nature même des données (les instants des défaillances sont inconnus), une inférence fondée sur la loi de Weibull, classiquement utilisée pour modéliser un vieillissement, a peu de chance d'atteindre son but. Aussi, il est montré dans Celeux, Lavergne et Vernaz (2000) que la recherche d'un effet vieillissement par un test dans un modèle linéaire généralisé fondé sur le processus de Poisson associé au nombre de défaillances entre deux contrôles donne des résultats bien plus fiables : on détecte avec plus de sûreté l'apparition d'un vieillissement même si on ne peut le quantifier.

ANNEXE : L'ALGORITHME EM

De nombreux problèmes statistiques comportent intrinsèquement des données manquantes. Dans beaucoup de cas, l'existence de données manquantes fait que les paramètres du modèle n'admettent pas d'estimateurs explicites du maximum de vraisemblance. L'identification du modèle nécessite alors l'usage d'algorithmes itératifs. Traditionnellement, dans de tels cas, on a recours à des algorithmes de type Newton-Raphson (NR). Si on note $L(\theta)$ la fonction de vraisemblance, l'algorithme NR construit une suite (θ^m) définie par

$$\theta^{m+1} = \theta^m - [D^2 L(\theta^m)]^{-1} DL(\theta^m)$$

qui converge vers le maximum de vraisemblance si la fonction de vraisemblance est concave et unimodale. Mais ces algorithmes sont difficilement applicables si les paramètres à estimer sont nombreux. Ils ne peuvent prétendre être une réponse générale aux problèmes d'estimation pour des modèles à données incomplètes, car ils ne prennent pas en compte la structuration particulière des données en données observées et en données manquantes. L'algorithme EM vise à rechercher le maximum de vraisemblance en tirant parti de cette caractéristique des données.

La structure des données incomplètes

On considère un modèle statistique dépendant d'un paramètre θ pour lequel l'échantillon complet $\mathbf{x}$, défini sur un espace mesurable C muni d'une mesure σ-finie de référence γ, n'est pas totalement observé mais se décompose en données observées $\mathbf{y}$, définies sur un espace mesurable O muni d'une mesure σ-finie de référence ω et en données manquantes $\mathbf{z}$, définies sur un espace mesurable M muni d'une mesure de référence σ-finie μ. (En général, les mesures γ, ω et μ sont soit la mesure de Lebesgue, soit la mesure de comptage, soit une mesure produit de ces deux mesures.) Les données observées sont l'image par une application π de l'échantillon complet : $\mathbf{y} = \pi(\mathbf{x})$. On suppose que $\mathbf{x}$ suit une loi appartenant à une famille paramétrée $P(d\mathbf{x}|\theta)$. Les données observées suivent une loi $Q(d\mathbf{y}|\theta)$ définie comme la probabilité image par π de $P(d\mathbf{x}|\theta)$: pour tout ensemble mesurable A de O

$$Q(A|\theta) = \int_{\pi^{-1}(A)} P(d\mathbf{x}|\theta). \tag{1}$$

La loi conditionnelle des données complètes sachant $\mathbf{y}$ s'écrit $R(d\mathbf{x}|\mathbf{y},\theta)$. On a

$$P(d\mathbf{x}|\theta) = \int_O R(d\mathbf{x}|\mathbf{y},\theta) Q(d\mathbf{y}|\theta). \tag{2}$$

En général, $\mathbf{x}$ admet une densité $f(\mathbf{x}|\theta)$ par rapport à la mesure γ. Les données observées admettent une densité $g(\mathbf{y}|\theta)$ par rapport à la mesure ω. On a donc pour tout ensemble mesurable A de O, $d\mathbf{x}$ (resp. $d\mathbf{y}$) désignant la mesure $\gamma(d\mathbf{x})$ (resp. $\omega(d\mathbf{y})$),

$$\int_A g(\mathbf{y}|\theta) d\mathbf{y} = \int_{\pi^{-1}(A)} f(\mathbf{x}|\theta) d\mathbf{x}.$$

Ce que, à l'instar de Dempster, Laird et Rubin (1977), on notera de manière informelle mais suggestive

$$g(\mathbf{y}|\theta) = \int_{\pi^{-1}(y)} f(\mathbf{x}|\theta) d\mathbf{x}. \tag{3}$$

De plus, dans le cas courant où C = O × M, la densité conditionnelle des données manquantes, par rapport à la mesure μ, s'écrit

$$k(\mathbf{z}|\mathbf{y},\theta) = f(\mathbf{x}|\theta)/g(\mathbf{y}|\theta). \tag{4}$$

Pour les log-vraisemblances, (4) se traduit par l'égalité

$$\ell(\theta|\mathbf{x}) = \ell(\theta|\mathbf{y}) + \log k(\mathbf{z}|\mathbf{y},\theta) \tag{5}$$

$\ell(\theta|\mathbf{x})$ et $\ell(\theta|\mathbf{y})$ représentant respectivement la log-vraisemblance de l'échantillon complet et des données observées.

Le principe de l'information manquante

Le but est, bien sûr, de maximiser la vraisemblance observée $\ell(\theta|\mathbf{y})$ mais ce n'est pas souvent chose aisée. Une idée intuitive consisterait à remplacer les valeurs manquantes par leurs espérances sachant les données observées et une valeur courante de θ. Mais cette stratégie conduit souvent à des estimateurs biaisés de θ. Cet état de fait a été relevé par Orchard et Woodbury (1972) qui ont émis le *principe de l'information manquante* exprimant que l'information observée est la différence de l'observation complète et de l'information manquante. Ce principe invite à rechercher l'estimateur du paramètre θ qui maximise l'espérance conditionnelle par rapport à la densité $k(\mathbf{z}|\mathbf{y},\theta)$ de la vraisemblance des données complètes. Il est à la base de l'algorithme EM.

L'algorithme EM

L'algorithme EM va donc essayer de tirer parti de l'information manquante en considérant l'espérance conditionnelle de la vraisemblance de l'échantillon complet $\ell(\theta|\mathbf{x})$ sachant les données observées.

Considérons une estimation courante θ^r du paramètre et écrivons l'espérance conditionnelle de l'équation (5) pour la loi conditionnelle $k(\mathbf{z}|\mathbf{x},\theta^r)$. Il vient

$$\ell(\theta|\mathbf{y}) = Q(\theta|\theta^r) - H(\theta|\theta^r) \tag{6}$$

où

$$Q(\theta|\theta^r) = \int k(\mathbf{z}|\mathbf{y},\theta^r)\ell(\theta|\mathbf{y},\mathbf{z})d\mathbf{z} \tag{7}$$

et

$$H(\theta|\theta^r) = \int k(\mathbf{z}|\mathbf{x},\theta^r)\log k(\mathbf{z}|\mathbf{x},\theta)d\mathbf{z}. \tag{8}$$

D'après une inégalité classique de la théorie de l'information, conséquence directe de l'inégalité de Jensen (cf. Rao 1973 pp. 47), on a

$$H(\theta|\theta^r) \leqslant H(\theta^r|\theta^r) \tag{9}$$

de sorte que, si $Q(\theta|\theta^r) \geqslant Q(\theta^r|\theta^r)$, on a $\ell(\theta|\mathbf{y}) \geqslant \ell(\theta^r|\mathbf{y})$. Partant de là, l'algorithme EM se définit ainsi : on se donne une position initiale θ^0 et on effectue sucessivement les deux étapes E et M.

Étape E : calcul de l'espérance conditionnelle $Q(\theta|\theta^r)$.

Étape M : calcul de θ^{r+1} qui maximise $Q(\theta|\theta^r)$.

REMARQUE 1. — *En pratique, il est la plupart du temps inutile de calculer explicitement $Q(\theta|\theta^r)$ et l'étape E se ramène au calcul de la loi conditionnelle $k(\mathbf{z}|\mathbf{y},\theta^r)$ comme il est fait dans le modèle de vieillissement différé présenté à la section 3.*

REMARQUE 2. — *Si θ est le paramètre d'une loi de la famille exponentielle, alors l'algorithme EM revient à attribuer aux statistiques exhaustives canoniques leur espérance conditionnelle sachant les données obervées et la valeur courante de θ.*

Comportement théorique

Convergence. La principale caractéristique de l'algorithme EM est donc de faire croître la vraisemblance $\ell(\theta|\mathbf{y})$ à chaque itération. Plus précisément, on a

THÉORÈME 1. — *Toute suite* (θ^r) *engendrée par EM vérifie* $\ell(\theta^{r+1}|\mathbf{y}) \geqslant \ell(\theta^r|\mathbf{y})$ *avec l'égalité ssi* $Q(\theta^{r+1}|\theta^r) = Q(\theta^r|\theta^r)$.

Le corollaire suivant implique que l'estimateur du maximum de vraisemblance de θ est un point fixe de l'algorithme EM. Notons T l'opérateur de l'algorithme EM qui à θ^m associe θ^{m+1}.

COROLLAIRE 1. — *Pour tout* θ^* *qui produit la valeur maximum de* $\ell(\theta^*|y)$, *on a, presque sûrement,*

$$\ell(T(\theta^*)|\mathbf{y}) = \ell(\theta^*|\mathbf{y})$$
$$Q(T(\theta^*)|\theta^*) = Q(\theta^*|\theta^*)$$

et

$$k(\mathbf{z}|\mathbf{y}, T(\theta^*)) = k(\mathbf{z}|\mathbf{y}, \theta^*)$$

si, de plus, θ^* *est unique on a* $T(\theta^*) = \theta^*$.

Le théorème suivant indique que, sous des conditions assez générales, la suite engendrée par l'algorithme EM converge vers une valeur stationnaire de la vraisemblance.

THÉORÈME 2. — *Si une suite* (θ^r) *engendrée par EM vérifie* $D^{10}Q(\theta^{r+1}, \theta^r) = 0$ *(*D^{10} *signifiant la dérivée partielle par rapport au premier argument de la fonction Q), si* $k(\mathbf{z}|\mathbf{y}, \theta)$ *est suffisamment régulière pour que l'on puisse intervertir les ordres d'intégration et de dérivation et si* (θ^r) *converge vers* θ^*, *alors*

$$\frac{\partial}{\partial\theta}\ell(\theta|\mathbf{y}) \mid_{\theta=\theta^*} = 0.$$

Remarquons que la convergence vers un point stationnaire ne garantit pas la convergence vers un maximum local de la vraisemblance : la suite (θ^r) peut converger vers un col de la vraisemblance.

D'autres résultats sur la convergence de EM peuvent être trouvés dans Wu (1983). Nous résumons les plus importants.

- Si $g(\mathbf{x}|\theta)$ appartient à une famille exponentielle régulière et si θ appartient à un compact, alors (θ^r) converge vers une composante connexe compacte de l'ensemble des points stationnaires de la vraisemblance $\ell(\theta|\mathbf{y})$
- Si, de plus, $\ell(\theta|\mathbf{y})$ est unimodal et possède un seul point stationnaire, alors (θ^r) converge vers le maximum unique θ^r de $\ell(\theta|\mathbf{y})$.

Vitesse de convergence. Voici un théorème qui permet de préciser la vitesse de convergence de l'algorithme EM.

THÉORÈME 3. — *Sous des conditions de régularité classiques, on a pour tout point fixe θ^* de l'algorithme EM, la relation*

$$DT(\theta^*) = (D^{20}Q(\theta^*|\theta^*))^{-1} D^{20}H(\theta^*|\theta^*) \tag{10}$$

où $DT(\theta^*)$ représente la matrice jacobien de T au point θ^*, $D^{20}Q(\theta^*|\theta^*)$ (resp. $D^{20}H(\theta^*|\theta^*)$) représente la matrice dérivée seconde par rapport au premier argument de $Q(\theta^*|\theta^*)$ (resp. $H(\theta^*|\theta^*)$) évaluée au point θ^*.

La vitesse de convergence de l'algorithme EM vers un point fixe θ^* va donc être pilotée par les valeurs propres de DT. De plus, ce théorème s'interprète de manière suggestive en termes d'information ; en dérivant 2 fois (7) par rapport à θ, il vient

$$I(\theta|\mathbf{y}) = -D^{20}Q(\theta^*|\theta^*) - D^{20}H(\theta^*|\theta^*) \tag{11}$$

où $I(\theta|\mathbf{y})$ représente l'information observée empirique, $-D^{20}Q(\theta^*|\theta^*)$ représente l'information complète et $-D^{20}H(\theta^*|\theta^*)$ l'information manquante. Ainsi, l'équation (11) justifie le principe de l'information manquante d'Orchard et Woodbury. De plus, on déduit de l'équation (10) que $DT(\theta^*)$ s'interprète comme le « rapport » de l'information manquante sur l'information complète : plus l'information manquante sera importante par rapport à l'information complète, plus la convergence de EM vers θ^* sera lente. Notons, enfin, que les relations (10) et (11) ont été exploitées par plusieurs auteurs pour accélérer la convergence de l'algorithme EM et pour évaluer la variance des estimateurs qu'il fournit (cf. McLachlan et Krishnan 1997, chapitre 4).

La proposition suivante permet de préciser le statut des points fixes de l'algorithme EM par rapport aux points stationnaires de la vraisemblance.

THÉORÈME 4. — *Pour tout point fixe θ^* de EM, on a*

$$D^2\ell(\theta|\mathbf{y}) = D^{20}Q(\theta^*|\theta^*)\,[I - DT(\theta^*)]\,. \tag{12}$$

COROLLAIRE . — *Si $D^{20}Q(\theta^*|\theta^*)$ est définie positive, alors*

θ^* est un point fixe attractif de EM (toutes les valeurs propres de $DT(\theta^*)$ sont comprises entre 0 et 1) si et seulement si θ^* est un maximum local de $\ell(\theta|\mathbf{y})$;

θ^* est un point fixe répulsif de EM (toutes les valeurs propres de $DT(\theta^*)$ sont plus grandes que 1) si et seulement si θ^* est un minimum local de $\ell(\theta|\mathbf{y})$;

θ^* est un point fixe hyperbolique de EM (certaine valeurs propres de $DT(\theta^*)$ sont comprises entre 0 et 1, d'autres sont plus grandes que 1) si et seulement si θ^* est un col de $\ell(\theta|\mathbf{y})$.

On doit insister sur le fait que tous ces résultats ne décrivent le comportement de EM qu'au voisinage des points fixes, ce qui rend compte imparfaitement du comportement réel de cet algorithme.

Comportement pratique

Le nombre de publications, dans des domaines très divers, sur l'algorithme EM témoigne de sa souplesse et de son efficacité. Néanmoins, bon nombre d'articles analysent des défauts bien réels de l'algorithme EM que l'on peut résumer en deux points :

(i) forte dépendance parfois par rapport à sa position initiale et risque de convergence vers des solutions hautement sous-optimales,

(ii) situations de convergence catastrophiquement lentes.

Le deuxième aspect (ii) a conduit au cours des années 1990 au développement de nombreux algorithmes qui visent et souvent réussissent à pallier cet inconvénient de EM tout en gardant sa simplicité et notamment sa caractéristique agréable de faire augmenter la vraisemblance à chaque itération. On trouvera une description synthétique de ces algorithmes dans le chapitre 5 de McLachlan et Krishnan (1997). L'autre aspect (i), de forte dépendance à la position initiale, a reçu moins d'attention. Une solution qui donne souvent de bons résultats consiste à utiliser une version stochastique de EM, l'algorithme SEM (cf. McLachlan et Krishnam 1997, chapitre 6) qui complète à chaque itération l'échantillon $\mathbf{x}$ par tirages aléatoires des données manquantes $\mathbf{z}$ selon la loi $k(\mathbf{z}|\mathbf{y},\theta)$. Cet algorithme évite ainsi d'être capturé par le premier point fixe rencontré. Au minimum, il convient en tout cas de faire tourner l'algorithme EM à partir de plusieurs positions initiales tirées au hasard et de choisir la solution fournissant la plus grande vraisemblance observée.

Remerciements. Ce texte a profité des années de travail sur des problèmes d'analyse statistique de données de retour d'expérience de différents partenaires. Je suis en particulier redevable à Marie-Agnès Garnero, André Lannoy, Matthieu Persoz et Benjamin Villain du département des études et recherche d'EDF, ainsi qu'à mes collègues du projet IS2 de l'Inria, Henri Bertholon, Christelle Breuils, Franck Corset, Christian Lavergne et Yann Vernaz.

BIBLIOGRAPHIE

BACHA, M. (1996) *Inférence statistique pour des modèles de durées de vie et applications.* Thèse de l'université de Rouen.

BACHA M., CELEUX G., IDÉE E., LANNOY A. et VASSEUR D. (1998) *Estimation de modèles de durées de vie fortement censurées.* Eyrolles.

BERTHOLON H. (2000) A change point ageing model. *MMR'2000*, 199-201.

CELEUX G., PERSOZ M., NGATCHOU-WANDJI J. et PERROT F. (1999) Bayesian modelling of PWR vessels flaw distributions. *Reliability Engineering and System Safety*, **66**, 243-252.

CELEUX G., LAVERGNE C. et VERNAZ Y. (2000) Assessing material aging from doubly censored data : Weibull distribution vs. Poisson process. Rapport de recherche Inria 3857.

DEMPSTER A. P., LAIRD N. M. et RUBIN D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal. Roy. Statist. Soc. (Ser. B)*, **39**, 1-38.

GILKS W. R., RICHARDSON S. et SPIEGELHALTER D. J. (1996) *Markov Chain Monte Carlo in Practice.* Chapman and Hall.

MCLACHLAN G. J. et KRISHNAN T. (1997) *The EM algorithm and Extensions.* New York Wiley.

ORCHARD T. et WOODBURY M. A. (1972) A missing Information Principle : Theory and Application. *Procceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability*, (vol. 1), 697-715.

PROCACCIA H. (1992) *Fiabilité des équipements et théorie de la décision statistique fréquentielle et bayésienne.* Eyrolles.

RAO C. R. (1973) *Linear Statistical Inference and its Applications.* (2d. edition) New York, Wiley.

ROBERT C. P. (1992) *Analyse statistique bayésienne.* Economica.

ROBERT C. P. (1996) *Méthodes de Monte-Carlo par chaînes de Markov.* Economica.

ROBERT C. P. (1998) *Discretization and MCMC convergence assessment.* Lexture Notes in Statistics 135, Springer-Verlag.

SUN D. (1997) A note on non informative priors for Weibull distributions. *Journal of Statistical Planning and Inference*, **61**, 319-338.

WU C. F. (1983) On the Convergence Properties of the EM Algorithm. *Annals of Statistics* **11**, 95–103.