HAL open science

Genèse et estimation d'un modèle de fiabilité en baignoire à taux de défaillance borné

Edwige Idee, Lambert Pierrat

To cite this version:

Edwige Idee, Lambert Pierrat. Genèse et estimation d'un modèle de fiabilité en baignoire à taux de défaillance borné. 42èmes Journées de Statistique, Société Française de Statistique, May 2010, Marseille, France. inria-00494734

HAL Id: inria-00494734

https://inria.hal.science/inria-00494734v1

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.

GENESE ET ESTIMATION D'UN MODELE DE FIABILITE EN BAIGNOIRE A TAUX DE DEFAILLANCE BORNE

Edwige Idée (1) & Lambert Pierrat (2)

(1) LAMA, Université de Savoie, Chambéry, France, edwige.idee@univ-savoie.fr
(2) LJK - STAT , M3S, Université de Grenoble, France, e_zainescu@yahoo.com

RESUME
La pertinence d'un modèle de fiabilité en baignoire résulte d'un compromis entre sa représentativité et sa complexité. Nous présentons la genèse et la procédure d'estimation d'un nouveau modèle, ayant à la fois une relative souplesse d'adaptation et un nombre minimal de paramètres. Partant d'un échantillon de 60 données de défaillance relatif à un dispositif technique, on compare les performances de ce modèle simple à 2 paramètres à celles d'un modèle plus classique à 5 paramètres, basé sur un mélange additif de deux lois de Weibull.
MOTS CLES: fiabilité, ingéniérie, genèse modèle, courbe en baignoire, loi exponentielle, loi de Weibull;
ABSTRACT
Genesis and estimation of a bathtub curve reliability model having a bounded failure rate.
The relevance of a "bathtub" reliability model results from a compromise between representativity and complexity. We present the genesis and the estimation procedure of a new model having both a sufficient adaptability and a minimal number of parameters. Starting from a statistical sample of 60 failure data relative to a technical device, we compare the performances of this simple two parameters model to those of a more classical model, i.e. a five parameters model based on an additive mixture of two Weibull distributions.
KEY-WORDS: reliability, engineering, model genesis, bathtub curve, exponential distribution, Weibull distribution;

1. GENESE DU MODELE

Une loi exponentielle de paramètre d'échelle (η) est une loi de Weibull particulière dont le paramètre de forme est unitaire (β=1). Le modèle proposé initialement par Idée (2006), consiste à introduire une perturbation de ce paramètre en posant (β=1-z). Dans cette expression, (z) est une variable aléatoire dont le support de la densité s'écrit [0,a] ⊆[0,1] .
Sa moyenne (ε=E(z)<1) existe donc, ce qui permet, en utilisant un développement limité du premier ordre, puis une intégration, de faire apparaître un mélange additif de deux lois. L'une est la loi exponentielle de poids (1- ε), l'autre de poids (ε) et de densité:

$$f(t,\eta) = \frac{(-1+t/\eta)\ln(t/\eta)e^{-t/\eta}}{\eta} t \geq 0, \eta > 0. \qquad (1)$$

Sa moyenne s'écrit:

$$E(T) = \eta(1+\Psi'(2)) \approx \eta(1.42)$$

$$\Psi'(x) = (\ln(\Gamma(x)))'$$

expression dans laquelle intervient la fonction Psi, dérivée du logarithme népérien de la fonction Gamma, indifféremment nommée digamma par Davis (1972).
Le moment d'ordre deux s'écrit quant à lui:

$$E(T^2) = 2\eta^2(1+2\Psi'(3)) \approx 2\eta^2(2.84)$$

Ce qui permet d'obtenir la variance en fonction du seul paramètre d'échelle de la loi exponentielle:

$$V(T) \approx \eta^2(3.66)$$

La fonction de survie est alors donnée par la relation explicite:

$$S(t,\eta) = e^{-t/\eta}(1+\frac{t}{\eta}\ln(\frac{t}{\eta})), t \geq 0, \eta > 0 \qquad (2)$$

Nota: la fonction de survie S(t) est identique à la fonction de fiabilité R(t).

Ainsi le taux de défaillance instantané du mélange des deux lois s'écrit:

$$\lambda(t,\varepsilon,\eta) = \frac{1}{\eta}(1-\varepsilon\frac{1+\ln(t/\eta)}{1+\varepsilon(t/\eta)\ln(t/\eta)}), \eta > 0, \varepsilon \in [0,1] \qquad (3)$$

Son graphe a une allure dite « en baignoire » dépendant de la valeur du paramètre (ε) et sa valeur asymptotique est bornée par l'inverse du paramètre d'échelle, comme indiqué sur la Figure 1.

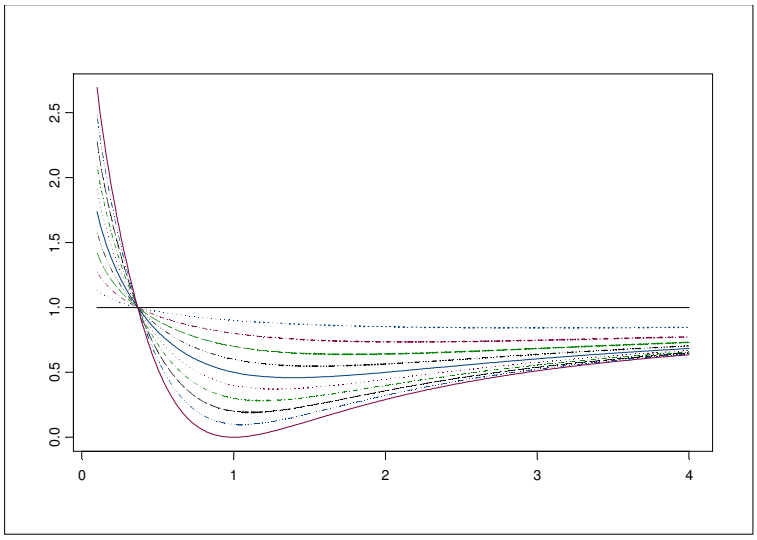

Figure1 : Taux de défaillance pour η=1, ε=0(noir), 1(trait rouge) et t compris entre 0 et 4. Les courbes intermédiaires correspondent aux valeurs 0.1(0.1)0.9 du paramètre (ε).

L'instant correspondant au minimum de la courbe est toujours supérieur au paramètre d'échelle, ce qui dans certains cas d'application peut conduire à des limitations. Afin de ramener cet instant à une valeur inférieure au paramètre d'échelle et d'accroître l'acuité du minimum, ce modèle initial est modifié de la façon suivante.
On fait d'abord disparaître le paramètre de perturbation en le fixant à une valeur unitaire, puis on mélange la loi correspondante à une loi gamma. On montre que si le paramètre de forme de cette loi gamma est un entier positif fixé (ici nous le fixerons à 2), on obtient un modèle explicite, défini par deux paramètres: le paramètre d'échelle et un paramètre de mélange, le taux de défaillance étant encore asymptotiquement borné à l'infini.

2. PROPRIETES DU MODELE

En conservant les mêmes notations, mais en notant que la signification de (ε) est celle du paramètre de mélange, la densité de ce nouveau mélange dans lequel on a introduit les modifications précédentes s'écrit comme suit:

$$f(t,\eta,\varepsilon)=(1-\varepsilon)\frac{t}{\eta}\frac{e^{-t/\eta}}{\eta}+\varepsilon\frac{(-1+t/\eta)\ln(t/\eta)e^{-t/\eta}}{\eta} \tag{4}$$

$t\geq 0,\eta>0,0\leq\varepsilon\leq 1.$

La fonction de fiabilité est alors donnée par la relation:

$$S(t,\eta,\varepsilon)=e^{-t/\eta}(1+(1-\varepsilon)\frac{t}{\eta}+\varepsilon\frac{t}{\eta}\ln(\frac{t}{\eta})) \tag{5}$$

$t\geq 0,\eta>0,\varepsilon\in[0,1]$

D'où la moyenne:

$E(T)=\eta(2+\varepsilon(\Psi'(2)-1))\approx\eta(2-0.58\varepsilon)$

$\Psi'(x)=(\ln(\Gamma(x)))'$

et le moment d'ordre deux:

$E(T^2)=2\eta^2(3+\varepsilon(2\Psi'(3)-2))\approx 2\eta^2(3-0.15\varepsilon)$

qui permettent d'obtenir la variance:

$V(T)\approx\eta^2(2+2.02\varepsilon-0.34\varepsilon^2)$

Finalement, le taux de défaillance a l'expression suivante qui peut être comparée à l'expression (3) relative au modèle intial:

$$\lambda(t,\eta,\varepsilon)=\frac{1}{\eta}(1-\frac{1+\varepsilon\ln(t/\eta)}{1+(1-\varepsilon)(t/\eta)+\varepsilon(t/\eta)\ln(t/\eta)}),\eta>0,\varepsilon\in[0,1] \tag{6}$$

Son graphe a toujours une allure dite « en baignoire », dépendant de la valeur du paramètre de forme (ε) et comme le montre la Figure 2, l'instant du minimum de la courbe est ramené à une valeur inférieure à celle du paramètre d'échelle.

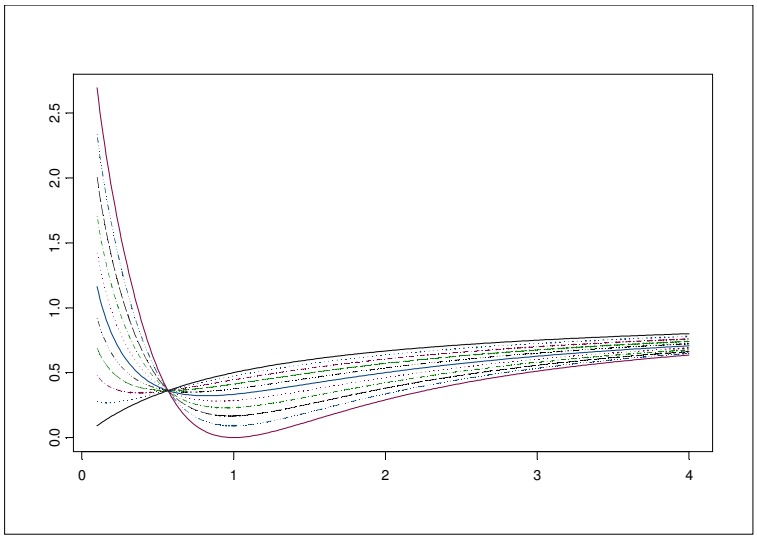

Figure 2 : Taux de défaillance pour η=1, ε=0(noir), 1(trait rouge) et t compris entre 0 et 4. Les courbes intermédiaires correspondent aux valeurs 0.1(0.1)0.9 du paramètre de forme

Ce minimum, quand il existe(ε>0) est solution unique de l'équation :

$$\frac{1}{u}=\varepsilon\ln^2(u)+(2-\varepsilon)\ln(u)+\frac{1}{\varepsilon}-1+\varepsilon$$

avec ($u=\frac{t}{\eta}$)

3. ESTIMATION DES PARAMETRES

Nous disposons d'un échantillon de n observations indépendantes et complètes notées $t_1, t_2, \ldots t_n$..
La vraisemblance de l'échantillon s'écrit:

$$V(\eta, \varepsilon / ech) = \frac{e^{-\sum_1^n t_i/\eta}}{\eta^n} \prod_1^n ((1-\varepsilon)(t_i/\eta) + \varepsilon(t_i/\eta - 1)\ln(t_i/\eta)) \tag{7}$$

et la log vraisemblance :

$$L(\eta, \varepsilon / ech) = -n\ln(\eta) - \frac{\sum_1^n t_i}{\eta} + \sum_1^n \ln((1-\varepsilon)\frac{t_i}{\eta} + \varepsilon(\frac{t_i}{\eta} - 1)\ln(\frac{t_i}{\eta})) \tag{8}$$

La dérivée partielle de la log vraisemblance par rapport à (η) fournit l'équation suivante :

$$\frac{\sum_1^n t_i}{\eta} = n + \sum_1^n \frac{(t_i/\eta) - \varepsilon + \varepsilon(t_i/\eta)\ln(t_i/\eta)}{(1-\varepsilon)(t_i/\eta) + \varepsilon(t_i/\eta - 1)\ln(t_i/\eta)} \tag{9}$$

Pour maximiser la log vraisemblance, on adopte la procédure suivante:

- pour (ε) fixé, la valeur de (η) est estimée en résolvant l'équation obtenue en annulant la dérivée partielle de la log vraisemblance par rapport à (η),
- on calcule ensuite la valeur de la log vraisemblance qui n'est plus qu'une fonction de (ε),
- enfin, on maximise cette log vraisemblance sur l'intervalle [0,1].

4. APPLICATION

En vue d' apprécier les performances du modèle (5), nous utilisons un échantillon issu de Lawless (2003), formé des 60 données complètes suivantes, relatives aux instants de défaillance d'un dispositif technique:

14 34 59 61 69 80 123 142 165 210 381 464 479 556 574 839 917 969 991 1064 1088 1091 1174 1270 1275 1355 1397 1477 1578 1649 1702 1893 1932 2001 2161 2292 2326 2337 2628 2785 2811 2886 2993 3122 3248 3715 3790 3857 3912 4100 4106 4116 4315 4510 4584 5267 5299 5583 6065 9701

Ces données correspondent à un taux de défaillance empirique « en baignoire » que l'auteur a proposé d'identifier à un mélange additif pondéré de deux lois de Weibull ayant pour fonction de fiabilité:

$$R(t, \alpha_1, \beta_1, \alpha_2, \beta_2, p) = pe^{-(t/\eta_1)^{\beta_1}} + (1-p)e^{-(t/\eta_2)^{\beta_2}} \tag{10}$$

Par maximisation de la vraisemblance, il a obtenu les estimations suivantes:

$$\beta_{1MV} = 1.66(0.49), \alpha_{1MV} = 95.4(25.8), \beta_{2MV} = 1.40(0.18), \alpha_{2MV} = 2774.5(314.2), p_{MV} = 0.137(0.051)$$

(les valeurs entre parenthèses sont les écart-types associés aux estimateurs).

La distance de Kolmogorov –Smirnov entre la fiabilité empirique et la fiabilité estimée par ce modèle de mélange à 5 paramères vaut: $D_{KM} = 0.0417$

Comme alternative, nous proposons un modèle de mélange à 2 paramètres du type (5) en remarquant le premier terme de la densité (4) traduit le fait qu'un petit nombre de composants de la

population présente des défauts dits de "jeunesse".

La maximisation de la vraisemblance fournit les estimations suivantes:

$\varepsilon_{MV} = 0.291(0.087), \eta_{MV} = 1190.4(78.7)$

La distance de Kolmogorov –Smirnov entre la fiabilité empirique et la fiabilité estimée par ce mélange à 2 paramètres vaut: $D_{KS} = 0.0460$.

Les deux distances de Kolmogorov –Smirnov sont très proches (0.046 vs 0.042) ce qui traduit la grande proximité des fonctions de fiabilité estimées et le fait qu'elles sont pratiquement indiscernables sue la Figure 3.

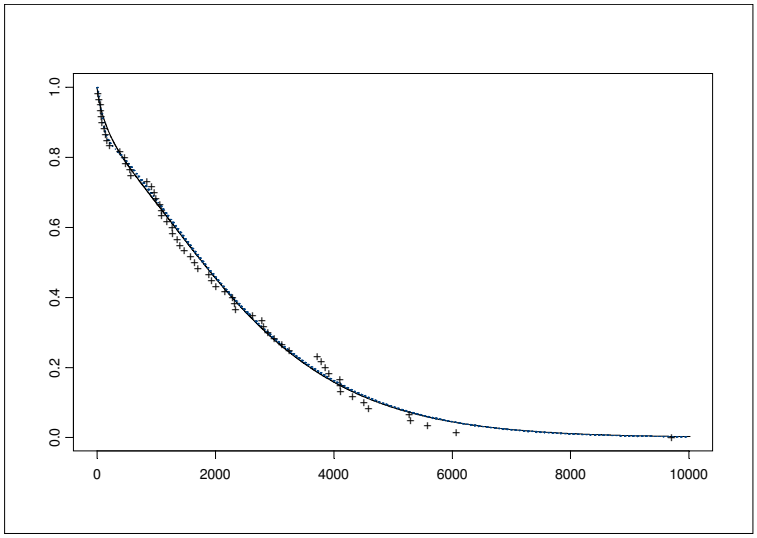

Figure 3: Fiabilités empirique (+), estimée à partir du modèle proposé (trait noit) et estimée à partir du mélange de 2 lois de Weibull (pointillés bleus).

La Figure 4 permet de comparer les taux de défaillance estimés à partir des deux modèles. On constate que le modèle proposé fait apparaître, d'une part un minimum moins accentué, d'autre part une tendance asymptotique bornée. Ces particularités résultent des structures différentes des deux modèles considérés.
En effet, le modèle de mélange additif de Weibull postule l'existence au sein de la population, de deux sous populations distinctes. Ce n'est pas le cas du modèle proposé dont la tendance se rapprocherait plus de celle d'un modèle multiplicatif de Weibull, caractérisé par une population homogène soumise à deux modes de défaillance différents.
La justification de tel ou tel modèle impliquerait de disposer d'informations complémentaires, en particulier de nature physique, mais ceci est en dehors du contexte statistique qui nous intéresse.

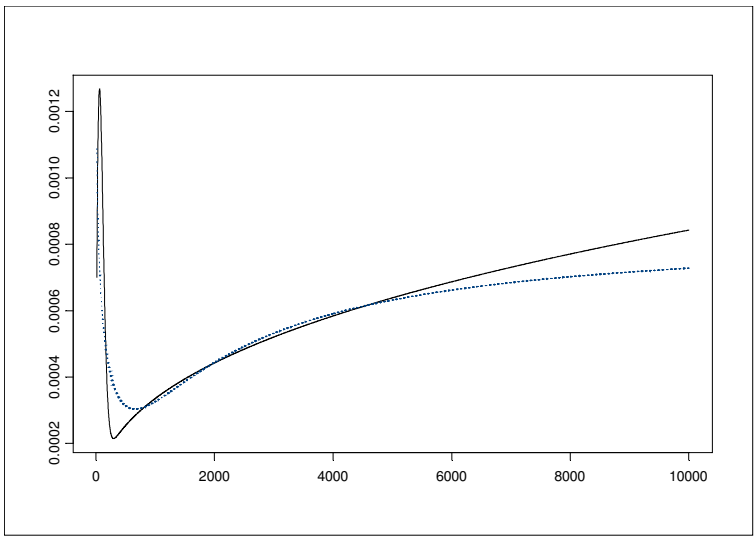

Figure 4: taux de défaillance estimé à partir du modèle proposé (pointillés) et estimé à partir du mélange de 2 lois de Weibull (trait noir).

5. CONCLUSIONS

Dans le cas étudié, le modèle proposé présente une bonne adéquation aux données pour un nombre minimal de paramètres. Il est évident que sa souplesse d'adaptation est intrinséquement plus faible que celle d'autres modèles de mélanges (additif ou multiplicatif) de lois de Weibull. Cependant, la complexité minimale de ce modèle le rend d'autant plus compétitif que la qualité des données est insuffisante, ce qui est fréquent dans le cas d'analyse de données de fiabilité issues du domaine technique. En effet, à quantité d'information égale, lorsque le nombre de paramètres augmente, la variabilité de leurs estimateurs augmente corrélativement.

Bibliographie

[1] Idée, E. (2006), Loi de probabilité obtenue par perturbation aléatoire simple de l'un des paramètres de la loi de référence , *Prépublication du LAMA-Université de Savoie*, N° 07-06a.

[2] Davis, P.J. (1972), Gamma function and related functions, Chapter 6, in Handbook of Mathematical Functions, M. Abramowitz & I. Stegun, NBS series 55

[3] Lawless, G.J.(2003), *Statistical Models and Methods for Lifetime Data*, Wiley Series in Probability and Statistics.