Projet de recherche : Modélisation d'une fiabilité par Polyà inverse, application à la maintenance d'équipements industriels

M2 2024

1 Contexte

L'assimilation de la durée de vie de systèmes industriels ou humains avec des tirages d'urnes de Pólya a été appliquée avec succès à des problèmes de contagion. Voir par exemple [1, 4, 3]. On trouve plus de précisions et d'autres références dans [6]. On assimile ici le composant à une urne contenant initialement N boules, dont k noires et $N - k$ blanches. Chaque sollicitation est considérée comme un tirage : si l'on tire une boule blanche, il n'y a pas défaillance ; si l'on tire une boule noire, il y a défaillance. La modélisation d'une usure régulière du composant est faite de la façon suivante : tant que l'on tire une boule blanche, on rajoute ν boules noires dans l'urne, en sus de la boule blanche tirée. Ainsi, on augmente régulièrement la probabilité d'obtenir une boule noire, soit une défaillance à la sollicitation.

D'après [5], ce type de modèle peut être utile pour représenter la durée de vie d'un composant industriel de type "interrupteur" (données de fonctionnement discrètes = comptage) dont le vieillissement est décéléré au cours du temps, ce qui est réaliste pour bon nombre d'applications. Son intérêt pour le monde industriel et la fiabilité n'a toutefois été étudié que dans un cadre fréquentiste classique.

L'objectif de ce travail est de l'étudier dans un cadre bayésien, en produisant des propositions de modèles complets, de calcul bayésien et de sensibilité décisionnelle. Pour cela, on pourra par exemple proposer d'étudier la référence récente [2] qui pourra donner des idées, mais on pourra chercher à développer une méthodologie propre sur des jeux de simulations et des données réelles.

2 Application

Trois jeux de données réelles sont téléchargeables :

— un jeu indiquant le nombre de cycles de fonctionnement d'un diesel jusqu'à défaillance : `https://perso.lpsm.paris/~bousquet/projets/DieselSollicitation.txt`
— un jeu indiquant le nombre de sollicitations d'un disjoncteur jusqu'à défaillance : `https://perso.lpsm.paris/~bousquet/projets/Disjoncteurs.txt`

— un jeu NIFE indiquant le nombre de démarrages d'une batterie jusqu'à défaillance : `https://perso.lpsm.paris/~bousquet/projets/NIFE.txt`

Ces jeux de données comprennent des vraies données de défaillance (labellisées par 1) et des données de censure à droite (labellisées par 0). Ces jeux de données sont accompagnées d'une expertise qui s'exprime sous la forme d'un estimateur et un écart-type de l'espérance du nombre de défaillance N_d (voir plus bas), fournie dans le tableau ci-dessous :

Jeu	moyenne	écart-type
Diesel	150	35
Disjoncteur	900	100
Batteries NIFE	5	1

TABLE 1 – Expertise sur le nombre de sollicitations avant défaillance.

Dans les trois cas, on souhaite pouvoir estimer des durées de vies moyennes avant défaillance (MTTF, voir plus bas), calculer une gamme de probabilités de tomber en panne avant des temps (nombres de cycles ou de sollicitation) $n_0, n_1, \ldots$ et proposer une règle de maintenance fondée sur une fonction de coÃ»t à élaborer. Le tout peut prendre la forme d'un code Python ou R, possiblement sous la forme d'une interface graphique accompagnée d'un petit document méthodologique.

Toutefois, on souhaite aussi savoir si la loi *a posteriori* prédictive permet d'expliquer ces données. Le modèle de Polya inverse bayésien est-il un bon modèle ? Comment peut-on le simuler ?

On portera un soin particulier à expliquer les choix *a priori*.

3 Formalisation

Pour une bonne introduction au modèle de Polyà inverse, voir [5].

Soit N_d le nombre de défaillances. On peut paramétriser le modèle par

$$\begin{aligned} \theta &= \frac{k}{N}, \\ \delta &= \frac{\nu}{N}. \end{aligned}$$

Ici θ correspond à la probabilité de tomber en panne à la première sollicitation (par accident, pourrait-on dire). Ainsi $\theta \in [0,1]$. Au contraire, δ est un indicateur du vieillissement ou plutàÂ´t de la *défiabilité* du composant, et celui-ci n'est pas forcément restreint à $[0,1]$. Le *taux de défaillance* est alors

$$h(n) \;=\; P(N_d = n | N_d > n-1) \;=\; \frac{\theta + (n-1)\delta}{1 + (n-1)\delta},$$

la probabilité de défaillance ("densité" du modèle) et la survie s'écrivent

$$\begin{aligned}
P(N_d = n) &= \frac{(1-\theta)^{n-1}\,(\theta + (n-1)\delta)}{\prod_{i=1}^{n}(1+(i-1)\delta)},\\
P(N_d > n) &= \frac{(1-\theta)^{n}}{\prod_{i=1}^{n}(1+(i-1)\delta)}.
\end{aligned}$$

Le MTTF (*Mean Time To Failure*) vaut

$$\begin{aligned}
MTTF &= \mathrm{E}[N_p] \;=\; \sum_{n=0}^{\infty} nP(N_d = n|\theta,\delta)\\
&= \frac{(1-\delta)\delta^{(1/\delta)-2}}{(1-\theta)^{\frac{1-\delta}{\delta}}}\exp\left(\frac{1-\theta}{\delta}\right)\,\gamma\left(\frac{1-\delta}{\delta},\frac{1-\theta}{\delta}\right)
\end{aligned}$$

où $\gamma(a,x)$ est la fonction Gamma incomplète ("inférieure")

$$\gamma(a,x) \;=\; \int_0^x t^{a-1}exp(-t)\,dt \;=\; \Gamma(a) - \Gamma(a,x).$$

Remarque. Dans le cas où le vieillissement est nul, on obtient un modèle *géométrique* d'espérance

$$\begin{aligned}
\mathrm{E}[N_p] &= \sum_{n=0}^{\infty} n\theta(1-\theta)^{n-1} \;=\; \theta\left(\sum_{n=0}^{\infty} x^n\right)'(x = 1-\theta),\\
&= \frac{\theta}{\{1-(1-\theta)\}^2} \;=\; \theta^{-1}.
\end{aligned}$$

Ce type de modèle est l'équivalent "discret" d'un modèle exponentiel.

4 Quelques questions pour guider ce travail

1. Ecrivez la vraisemblance complète du modèle en utilisant la notation

$$\alpha \;=\; \sum_{i=1}^{s} k_i n_i + \sum_{j=s+1}^{s+r} n_j - r,$$

2. Construisez un simulateur de ce modèle, afin de produire des données synthétiques qui permettront de tester le bien-fondé des algorithmes d'estimation bayésienne.
3. Proposez des mesures *a priori* non informative ou peu informatives, puis un ou des choix de lois *a priori* explicables et enfin produisez des algorithmes de calcul bayésien, dont il faudra tester le bon comportement (par exemple grâce aux données simulées). Les lois non informatives (ex : Jeffreys) ne sont peut-être pas explicites, mais cela ne doit pas vous empêcher d'essayer de les mettre en oeuvre dans les algorithmes de calcul bayésien.
   — Pour calibrer une loi *a priori* sur (δ, θ) ou mener une analyse de sensibilité, il peut être intéressant de déterminer un domaine de valeurs possible en considérant les cas extrêmes $\delta = 0$ (défaillance purement accidentelle) ou $\theta = 0$ (usure sans accident initial).
   — Il est peut-être intéressant de supposer que θ, qui est une probabilité, suit une loi bêta (éventuellement restreinte à un sous-ensemble de $[0, 1]$. D'une manière générale, il est utile de s'interroger sur la signification des paramètres (θ, δ) pour construire des lois *a priori*.

Références

[1] F. Alajaji and T. Fuja. A communication channel modeled on contagion. *Institute for Systems Research Tech. Rep. TR 93-78*, 1993.

[2] C. Glynn, J. He, N.G. Polson, and J. Xu. Bayesian inference for polya inverse gamma models. *arXiv :1905.12141*, 2019.

[3] M. Hayhoe, F. Alajaji, and B. Gharesifard. A polya contagion model for networks. *IEEE Transactions on Control of Network Systems*, 2017.

[4] A.W. Marshall and I. Olkin. Bivariate life distribution from polya's urn model for contagion. *J. Appl. Prob.*, 30 :497–508, 1993.

[5] A. Pasanisi, C. Roero, E. Remy, and N. Bousquet. On the practical interest of discrete inverse pólya and weibull-1 models in industrial reliability studies. *Quality and Reliability Engineering International*, page in press, 2015.

[6] M. Xie, O. Gaudoin, and C. Bracquemond. Redefining failure rate function for discrete distributions. *Journal of Reliability, Quality and Safety Engineering*, 9 :275–285, 2002.