–Texte–
Fiabilité d'un système complexe

Mots-clefs : fonction de répartition, estimation de paramètre, intervalle de confiance

1 Le problème et sa modélisation

On considère un système complexe constitué de composants électroniques associés les uns aux autres, et supposés en état de fonctionnement à l'instant initial. En cas de défaillance du système, le temps de réparation est très court, et on peut considérer qu'il est négligeable par rapport à la durée de fonctionnement sans défaillance. Enfin, chaque réparation est mineure, de sorte que l'amélioration du système n'est pas fonction des pannes observées.

On veut modéliser les instants successifs de défaillance du système. Ces instants sont représentés par des v.a.r. $0 < T_1 < T_2 < \cdots$. Avec la convention $T_0 = 0$, les égalités en loi suivantes :

$$\forall n \geqslant 1 \quad : \quad \mathcal{L}\big(T_n - T_{n-1} | T_0, \cdots, T_{n-1}\big) = \mathcal{L}\big(T_n - T_{n-1} | T_{n-1}\big)$$

constituent, d'après les observations mentionnées dans le paragraphe précédent, une approximation raisonnable de la réalité. De plus, on supposera que ces lois conditionnelles possèdent une densité.

2 Taux de hasard, et taux de défaillance

Il est pratique de décrire la fiabilité d'un système par son *taux de hasard* :

Définition 1.1 *Soit T une v.a. à valeurs dans $\mathbb{R}_+$ de fonction de répartition F, qui possède une densité f. Le taux de hasard de T est la fonction h définie par*

$$h(t) = \begin{cases} \frac{f(t)}{1-F(t)} & \textit{si } F(t) \neq 1\,; \\ 0 & \textit{sinon.} \end{cases}$$

Dans notre étude, ce taux de hasard s'interprète comme un *taux de défaillance*. La terminologie de "taux" est justifiée par la remarque qui suit.

Remarque 1.1 *Soit T une v.a. à valeurs dans $\mathbb{R}_+$ de fonction de répartition F, qui possède une densité continue f. Son taux de défaillance h est tel que, si $F(t) \neq 1$:*

$$h(t) = \lim_{\varepsilon \searrow 0} \frac{1}{\varepsilon} \mathbb{P}\big(t < T \leqslant t + \varepsilon \big| T > t\big).$$

Typiquement, la courbe du taux de défaillance est une courbe en forme de baignoire. Pendant une première période, le taux de défaillance est décroissant : c'est la période de "rodage". Puis, le taux de défaillance est approximativement constant : c'est la période de "vie utile". Enfin, le taux de défaillance est croissant : c'est la période de "vieillissement".

Théorème 1.1 *Soit T une v.a. à valeurs dans $\mathbb{R}_+^*$ de fonction de répartition F, qui possède une densité continue f. La v.a. T admet h pour taux de hasard si, et seulement si, pour tout $t \geqslant 0$ tel que $F(t) \neq 1$:*

$$\mathbb{P}(T > t) = \exp\Big(- \int_0^t h(s)ds \Big).$$

Dans ce cas, on a aussi :

$$f(t) = h(t) \exp\Big(- \int_0^t h(s)ds \Big).$$

Si la v.a. T désigne la durée de fonctionnement d'un système, un taux de défaillance constant signifie que le système est sans mémoire. De fait, T a un taux de défaillance constant égal à $c > 0$ si, et seulement si $T \sim \mathcal{E}(c)$.

Etant donné un taux de défaillance, on peut simuler une réalisation de la loi correspondante.

Remarque 2.1 *Soit h le taux de hasard d'une loi à densité continue sur $\mathbb{R}_+$. On suppose que la fonction H définie par $H(t) = \int_0^t h(u)du$ est inversible. Si $U \sim \mathcal{U}[0,1]$, alors la v.a. $H^{-1}(-\ln U)$ a pour taux de hasard h.*

L'état du système complexe qui nous intéresse est décrit par son taux de défaillance : connaissant l'histoire du matériel jusqu'à l'instant t, la durée jusqu'à l'apparition de la prochaine défaillance est une v.a. dont le taux de défaillance vaut :

$$s \mapsto \lambda(s + t),$$

où λ est une fonction à valeurs positives et localement intégrable. Autrement dit, pour tout $n \geqslant 1$ et $0 = t_0 < t_1 < \cdots < t_{n-1}$, la loi conditionnelle de $T_n - T_{n-1}$ sachant $T_0 = t_0, \cdots, T_{n-1} = t_{n-1}$ possède pour taux de défaillance la fonction $s \mapsto \lambda(s + t_{n-1})$.

Proposition 1.1 *Notons $\Lambda(t) = \int_0^t \lambda(s)ds$, $t \geqslant 0$. Pour tout $n \geqslant 1$, le vecteur aléatoire $(T_1, \cdots, T_n)$ possède une densité qui est*

$$(t_1, \cdots, t_n) \mapsto \exp\big(- \Lambda(t_n)\big)\lambda(t_1) \cdots \lambda(t_n) \mathbf{1}_{\{0<t_1<\cdots<t_n\}}.$$

La remarque 2.1 fournit une méthode pour simuler une réalisation de $(T_1, \cdots, T_n)$. Soient $V_1, \cdots, V_n$ des v.a. indépendantes et de même loi $\mathcal{U}[0,1]$. Notons $\Lambda_x(t) = \int_0^t \lambda(s+x)ds$ -supposée inversible- et, par récurrence, $L_0 = 0$ et pour tout $n \geqslant 1$:

$$L_n - L_{n-1} = \Lambda_{L_{n-1}}^{-1}(-\ln V_n).$$

On vérifie avec la proposition 1.1 que $(L_1, \cdots, L_n) \sim (T_1, \cdots, T_n)$.

3 La loi de Weibull

Soient $\alpha, \beta > 0$. La loi de densité

$$\frac{\beta}{\alpha}\Big(\frac{t}{\alpha}\Big)^{\beta-1} \exp\Big(-\Big(\frac{t}{\alpha}\Big)^{\beta}\Big), \quad t \in \mathbb{R}_+,$$

est la *loi de Weibull* de paramètres (α, β), notée $W(\alpha, \beta)$. Le paramètre α est un paramètre d'échelle, et le paramètre β est un paramètre de forme. La loi de Weibull apparaît fréquemment dans les problèmes de fiabilité de systèmes complexes. Le *théorème des valeurs extrêmes* -que nous admettrons- fournit une explication mathématique à ce phénomène qui a par ailleurs été aussi constaté de manière empirique.

Théorème 2.1 *Soient $X_1, X_2, \cdots$ des v.a.r. indépendantes et de même fonction de répartition F. Supposons qu'il existe $\theta_1 > 0$ et $x_0 \in \mathbb{R}$ tels que $F(x_0) = 0$ et pour tout $x > x_0$, $F(x) > 0$ et*

$$\lim_{t \searrow 0} \frac{F(xt + x_0)}{F(t + x_0)} = x_1^{\theta}.$$

Alors, il existe $\theta_2 > 0$ et des suites de réels $(a_n)_n$ et $(b_n)_n$ telles que

$$Z_n := \inf_{k=1,\cdots,n} \big(a_n X_k + b_n\big) \to W(\theta_2, \theta_1),$$

en loi, lorsque $n \to \infty$.

La loi exponentielle, et de manière plus générale toute loi gamma de paramètre de forme $\theta > 0$, vérifie les hypothèses de ce théorème (auquel cas, le paramètre de forme de la loi de Weibull est θ).

Dans notre étude, la v.a. Z_n peut représenter la durée de vie d'un système composé de n éléments, le système étant défaillant dès que l'un des éléments le constituant est défaillant. Dans le cadre du problème qui nous intéresse, on peut donc considérer que le modèle statistique est un modèle paramétrique, dont les paramètres strictement positifs (α, β) sont ceux d'une loi $W(\alpha, \beta)$. On supposera dorénavant que le taux de défaillance λ s'écrit :

$$\lambda(t) = \frac{\alpha}{\beta}\Big(\frac{t}{\alpha}\Big)^{\beta-1}.$$

4 Estimation des paramètres

4.1 Les estimateurs

La log-vraisemblance de l'observation $(t_1, \cdots, t_n)$ du vecteur aléatoire $(T_1, \cdots, T_n)$ est

$$L_n(t_1, \cdots, t_n; \alpha, \beta) = n\big(\ln\beta - \beta\ln\alpha\big) + (\beta-1)\sum_{i=1}^n \ln t_i - \Big(\frac{t_n}{\alpha}\Big)^\beta.$$

Les estimateurs du maximum de vraisemblance $\hat\alpha$ et $\hat\beta$ de α et β vérifient donc :

$$\frac{1}{\hat\beta} = \ln T_n - \frac{1}{n}\sum_{i=1}^n \ln T_i \quad \text{et} \quad \ln\hat\alpha = \ln T_n - \frac{1}{\hat\beta}\ln n.$$

Dans toute la suite, on note pour $i = 1, \cdots, n$: $U_i = \Lambda(T_i) = (T_i/\alpha)^\beta$.

4.2 Lois des statistiques

Proposition 3.1 *Les v.a. U_n et $\beta/\hat\beta$ sont indépendantes et de lois gamma de paramètres respectifs $(n, 1)$ et $(n-1, n)$.*

Preuve. La loi conditionnelle de $(U_1, \cdots, U_{n-1})$ sachant $U_n = t$ est celle de la statistique d'ordre de $(n-1)$ v.a. indépendantes et de même loi uniforme sur $[0, t]$. Comme

$$\frac{\beta}{\hat\beta} = \ln U_n - \frac{1}{n}\sum_{i=1}^n \ln U_i,$$

on en déduit que pour tout $u \in \mathbb{R}$:

$$\mathbb{E}\Big(\exp\Big(iu\frac{\beta}{\hat\beta}\Big)\Big| U_n = t\Big) = \Big(1 - \frac{iu}{n}\Big)^{1-n},$$

d'où le résultat. •

Ainsi, $2n\beta/\hat\beta \sim \chi^2_{2n-2}$. En notant $\chi^2_{2n-2}(\theta)$ le quantile d'ordre θ de la loi χ^2_{2n-2}, on a donc montré que

$$\Big[\frac{\hat\beta}{2n}\chi^2_{2n-2}\Big(\frac{1-\gamma}{2}\Big), \frac{\hat\beta}{2n}\chi^2_{2n-2}\Big(\frac{1+\gamma}{2}\Big)\Big]$$

est un intervalle de confiance pour β au niveau γ.

D'après la proposition 3.1, $2U_n \sim \chi^2_{2n}$. On en déduit le résultat :

Proposition 3.2 *Soit F_{2n} la fonction de répartition de la loi χ^2_{2n}. Alors, pour tout $x \in \mathbb{R}$:*

$$\mathbb{P}\Big(\hat\beta\ln\frac{\hat\alpha}{\alpha} \leqslant x\Big) = \int_0^\infty F_{2n}\Big(2\exp\big(x + \ln n\big)t\Big)\frac{n^{n-1}}{(n-2)!}t^{n-2}\exp(-tn)dt.$$

En particulier, la statistique $\hat\beta\ln(\hat\alpha/\alpha)$ est libre, mais sa loi n'est pas tabulée. Il faut donc calculer numériquement les quantiles de la loi de $\hat\beta\ln(\hat\alpha/\alpha)$ afin de construire des intervalles de confiance pour α. En pratique, on pourra par exemple utiliser l'égalité

$$\hat\beta\ln\frac{\hat\alpha}{\alpha}=\frac{\ln U_n}{\ln U_n-\frac{1}{n}\sum_{i=1}^n\ln U_i}-\ln n.$$

afin de donner des valeurs approchées des quantiles.

4.3 Lois asymptotiques des statistiques

Proposition 3.3 *Lorsque $n\to\infty$, on a les convergences en loi suivantes :*

$$\sqrt{n}\Big(\frac{\beta}{\hat\beta}-1\Big)\to N(0,1)\quad et\quad \hat\beta\,\frac{\sqrt{n}}{\ln n}\Big(\frac{\hat\alpha}{\alpha}-1\Big)\to N(0,1).$$

Preuve. La première partie s'obtient à partir de la proposition 3.1. Pour la deuxième partie, il faut remarquer que

$$\ln\hat\alpha=\ln\alpha+\frac{1}{\beta}\ln\frac{U_n}{n}-\Big(\frac{1}{\hat\beta}-\frac{1}{\beta}\Big)\ln n.$$

Comme $2U_n\sim\chi^2_{2n}$, le théorème de la limite centrale et l'égalité

$$\frac{\sqrt{n}}{\ln n}\ln\frac{U_n}{n}=\frac{\sqrt{n}}{\ln n}\Big(\frac{U_n}{n}-1\Big)+O\bigg(\frac{\sqrt{n}}{\ln n}\Big(\frac{U_n}{n}-1\Big)^2\bigg)$$

nous montrent que $(\sqrt{n}/\ln n)\ln(U_n/n)$ tend vers 0 en probabilité. Le résultat annoncé en découle. •

5 Suggestions de développement

(a.) Présenter et commenter le modèle. Quels sont les influences des paramètres de forme et d'échelle sur la fiabilité du système ?

(b.) Démontrer certains résultats du texte. Le cas échéant, compléter les preuves.

(c.) Pour différents taux de défaillance, comparer des réalisations de $(T_1,\cdots,T_n)$. Graphiquement, $(T_1,\cdots,T_n)$ peut être représenté de la manière suivante : l'axe des abscisses contient les valeurs prises par $T_1,\cdots,T_n$, et l'axe des ordonnées contient le nombre de défaillances du système. Pour faciliter une comparaison visuelle, on pourra décaler (un peu) les trajectoires selon l'axe des ordonnées.

(d.) Illustrer par des simulations la proposition 3.3. Comparer numériquement les intervalles de confiance pour α (resp. β) obtenus avec les propositions 3.1-3.3.

