REVUE DE STATISTIQUE APPLIQUÉE

OLIVIER GAUDOIN
JEAN-LOUIS SOLER

Modèles pour l'étude de la fiabilité des systèmes présentant des fautes de conception. Application à l'évaluation de la fiabilité des logiciels

Revue de statistique appliquée, tome 40, n° 2 (1992), p. 91-98
<http://www.numdam.org/item?id=RSA_1992__40_2_91_0>

Rev. Statistique Appliquée, 1992, **XXXX** (2), 91-98

MODÈLES POUR L'ÉTUDE DE LA FIABILITÉ DES SYSTÈMES PRÉSENTANT DES FAUTES DE CONCEPTION. APPLICATION A L'ÉVALUATION DE LA FIABILITÉ DES LOGICIELS

Olivier GAUDOIN, Jean-Louis SOLER

L.M.C.-I.M.A.G., BP 53 X, 38041 Grenoble Cedex

RÉSUMÉ

L'objet de ce travail est le problème de l'évaluation de la fiabilité d'un système présentant des fautes de conception à un instant donné de son cycle de vie, au vu de ses défaillances successives et des corrections qui lui sont apportées. Dans ce but, nous proposons une modélisation stochastique du comportement au cours du temps d'un tel système, où l'on montre que l'évolution de ce système résulte de l'interaction complexe de trois processus aléatoires ponctuels, cette interaction étant entièrement résumée par la donnée de l'intensité conditionnelle de défaillances. Ce résultat permet de construire une classe de modèles mathématiquement exploitables, les modèles proportionnels, qui donnent lieu à d'intéressants problèmes de statistique appliquée. Cet article est volontairement concentré sur les aspects de modélisation et les résultats qu'ils peuvent engendrer. Les développements techniques détaillés pourront être trouvés dans Gaudoin [90].

Mots-clés : *Fiabilité de systèmes réparables, fiabilité des logiciels, modélisation probabiliste, statistique de processus aléatoires ponctuels.*

I. Introduction

Partant d'une origine de temps, on observe le processus ponctuel $T = \{T_i\}_{i\geqslant 0}(T_0 = 0)$, des instants successifs de défaillance d'un système. On définit de même le processus des temps inter-défaillances successifs $X = \{X_i\}_{i\geqslant 1}$, où $\forall i \geqslant 1$, $X_i = T_i - T_{i-1}$, ainsi que le processus $N = \{N_t\}_{t\geqslant 0}$, qui compte à chaque instant t le nombre de défaillances survenues jusque-là. A chaque défaillance, le système est corrigé ou pas, les temps de correction étant supposés négligeables ou non comptabilisés.

Le système étant réparable et améliorable, sa fiabilité est susceptible d'évoluer au cours du temps en fonction des corrections apportées. Sa fiabilité à l'instant t est définie comme la probabilité qu'il exécute sa tâche sans défaillances pendant encore une durée τ, dans un environnement spécifié. C'est donc la fonction $R_t(.)$

définie par :

$$\forall \tau \geqslant 0,\ R_t(\tau) = P(T_{N_t+1} - t > \tau).$$

Il s'agit d'évaluer cette fonction à chaque instant t à partir de l'observation faite jusque-là du processus T (ou X ou N), en utilisant un modèle de loi de probabilité pour ce processus censé être en adéquation avec les données d'observation.

Dans le §II, nous établirons des bases pour la modélisation stochastique des processus observés ; le §III sera consacré à la description du profil opérationnel poissonnien homogène dans lequel nous nous placerons dans la suite ; enfin, dans le §IV, nous décrirons la classe naturelle des modèles proportionnels et nous évoquerons leur étude statistique.

II. Modélisation stochastique du processus des défaillances et corrections d'un système présentant des fautes de conception

On considère que la fonction du système étudié consiste à transformer des **données d'entrée** en données de sortie (par l'intermédiaire d'un programme dans le cas d'un logiciel).

Soit E l'ensemble de toutes les éventualités d'entrée du système (que l'on pourra éventuellement assimiler à l'ensemble de toutes les données d'entrée admissibles), muni d'une tribu $\mathcal{A}$ d'événements d'entrée de référence.

Le **processus de sollicitation** est une suite de couples de variables aléatoires (S_n, Z_n), où S_n est l'instant de la $n^{ème}$ sollicitation et Z_n la donnée d'entrée correspondante. On peut aussi le représenter par le processus ponctuel spatial $\{\mathbb{S}_U\}_{U \in \mathcal{B}(\mathbb{R}^+) \otimes \mathcal{A}}$, où $\mathbb{S}_U$ est le nombre aléatoire de couples (S_n, Z_n) tombant dans la région U de l'espace produit $\mathbb{R}^+ \times E$.

Le **profil opérationnel**, qui détermine les conditions d'utilisation du système, sera défini comme étant la loi de probabilité de ce processus.

Une **défaillance** est une différence entre le résultat fourni par le système et le résultat prévu par les spécifications du système.

Une **faute de conception** est un défaut du système, qui, sous certaines conditions de sollicitation, entraînera une défaillance.

On découvre l'existence d'une faute quand, pour une donnée d'entrée précise, le résultat en sortie n'est pas conforme aux spécifications. On pourra donc confondre une faute de conception avec l'ensemble des données d'entrée qui conduiront à sa manifestation, c'est-à-dire à l'événement d'entrée correspondant. Ainsi, la **faute totale** à l'instant t, F_t, est l'ensemble des données d'entrée pouvant provoquer une défaillance à cet instant, ou encore l'événement d'entrée correspondant à la manifestation de l'une quelconque des fautes existant dans le système à cet instant.

Une défaillance se produit suite à la $n^{ème}$ sollicitation si l'entrée correspondante Z_n appartient à la faute totale F_{S_n}. On confondra l'instant d'une défaillance et l'instant de la sollicitation qui a entrainé cette défaillance, ce qui suppose le

temps d'exécution négligeable ou non décompté. Le processus $\{T_i\}_{i\geqslant 1}$ est donc une sous-suite du processus $\{S_n\}_{n\geqslant 1}$.

Une **correction** est une transformation qui, à une faute totale avant correction F, fait correspondre une faute totale après correction F'.

Comme on a défini le processus des instants de défaillance successifs T, on définit le processus des instants de correction successifs $C = \{C_i\}_{i\geqslant 1}$, et le processus qui compte à chaque instant le nombre de corrections effectuées jusque-là : $K = \{K_t\}_{t\geqslant 0}$. A chaque instant de correction C_i est associée une **marque** $F_i \in \mathcal{A}$ qui est la faute totale après la $i^{ème}$ correction. Le fait qu'il soit naturel d'admettre que la correction d'une faute à un instant donné ne dépend que de l'état de la faute à cet instant, et non de ses états passés, implique que le processus $\{F_i\}_{i\geqslant 1}$ est **markovien**.

L'évolution du système au cours du temps résulte alors de l'interaction complexe des trois processus aléatoires ponctuels introduits : celui des sollicitations, celui des instants de défaillances et celui des instants de correction et de leur effet sur l'état des fautes de conception à chaque instant. Le processus résultant sera appelé processus des défaillances-corrections du système, et sera noté : $\{\mathbb{S}_{[0,t]\times A},\ N_t,\ K_t,\ F_t\}_{t\geqslant 0, A\in\mathcal{A}}$.

La loi de probabilité du processus de défaillance est entièrement déterminée par la donnée de l'**intensité conditionnelle de défaillance**, définie par :

$$\lambda_t = \lim_{dt\to 0} \frac{1}{dt} P(N_{t+dt} - N_t = 1 | \mathcal{H}_t) = \lim_{dt\to 0} \frac{1}{dt} P(\mathbb{S}_{]t,t+dt]\times F_{K_t}} = 1 | \mathcal{H}_t),$$

si ces limites existent, où $\mathcal{H}_t$ est la tribu, dans l'espace probabilisé de référence, engendrée par le passé du processus des défaillances-corrections à l'instant t. C'est cette intensité qu'il s'agit de modéliser.

On remarque que tous les modèles classiques de fiabilité des logiciels (Jelinski-Moranda [72], Littlewood-Verral [73], Goel-Okumoto [79], Musa-Okumoto [84], Kanoun-Laprie [85], etc...) s'intègrent parfaitement dans ce cadre, chaque modèle correspondant en fait à une intensité conditionnelle de défaillance particulière. Le processus de défaillance est souvent modélisé par un processus ponctuel auto-excité et, dans le cas le plus simple, par un processus de Poisson non homogène.

L'utilisation de cette intensité de défaillance permet de proposer une classification simple des modèles de fiabilité des logiciels (voir tableau 1), en mettant en évidence les différentes hypothèses qui les caractérisent respectivement (Gaudoin [90], deuxième partie).

III. Le profil opérationnel poissonnien homogène (POPH)

On se place dans le cas simple mais vraisemblable du profil opérationnel particulier où les instants de sollicitation $\{S_n\}_{n\geqslant 1}$ forment un processus de Poisson homogène sur $\mathbb{R}^+$ d'intensité $\mu > 0$, et les données d'entrée $\{Z_n\}_{n\geqslant 1}$ sont indépendantes entre elles, de même loi de probabilité Q sur $(E, \mathcal{A})$, et

TABLEAU 1
Tableau récapitulatif de la classification des modèles de fiabilité des logiciels

Modèles ND $\lambda_t = \lambda(N_t)$	Jelinski-Moranda (72) Shooman (73) Musa (75) première version Moranda (75) Jelinski-Moranda bayésien (87) Modèle Proportionnel Déterministe (90)
Modèles NDT $\lambda_t = \lambda(N_t, t)$	Shantikumar (81) Littlewood (81) Abdel-Ghali (86)
Modèles NDTE $\lambda_t = \lambda(N_t, t - T_{N_t})$	Schick-Wolverton (73) Littlewood-Verral (73) Keiller-Littlewood (83)
Modèles T ou NHPP $\lambda_t = \lambda(t)$	Duane (64) Crow (74) Musa (75) deuxième version Goel-Okumoto (79) Yamada-Ohba-Osaki (83) Musa-Okumoto (84) Kanoun-Laprie (85) Wightman-Bendell (85) Littlewood NHPP (86)

indépendantes des instants de sollicitation. De ce fait, $\mathbb{S}$ est un processus de Poisson spatial sur $\mathbb{R}^+ \times E$ d'intensité la mesure $\sigma = \mu L \otimes Q$, où L est la mesure de Lebesgue sur $\mathbb{R}^+$, et l'intensité conditionnelle de défaillance dans ce profil opérationnel est alors donnée par :

$$\lambda_t = \mu Q(F_{K_t}).$$

On en déduit la propriété suivante (Soler [88]) :

Dans le POPH et en supposant la correction immédiate ($N_t = K_t$ à tout instant t), il existe un processus de Markov $\Lambda = \{\Lambda_i\}_{i \geqslant 1}$, constitué de v.a.r. positives, tel que, conditionnellement à $\{\Lambda_i = \lambda_i\}_{i \geqslant 1}$, les temps inter-défaillances X_i sont indépendants et de loi exponentielle de paramètre λ_i respectivement.

Λ_i s'appellera naturellement taux de défaillance du système à la $i^{ème}$ étape (c'est-à-dire après la $(i-1)^{ème}$ correction).

Finalement, construire un modèle pour le processus de défaillance revient, dans ces conditions opérationnelles, à proposer un modèle pour le processus des

taux de défaillances successifs Λ, et donc à exprimer l'effet d'une correction sur le taux de défaillance du système.

IV. Les modèles proportionnels et leur utilisation

Une correction bien faite peut éliminer une partie des fautes de conception, tandis qu'une correction mal faite peut en ajouter de nouvelles. En général, une correction sera en partie de bonne qualité et en partie de mauvaise qualité. On montre (Gaudoin [90], troisième partie) que, si on se place dans le POPH, il existe deux suites de variables aléatoires à valeurs dans [0, 1], $\{\alpha_i\}_{i\geqslant 1}$ et $\{\beta_i\}_{i\geqslant 1}$, représentant des taux de bonne et de mauvaise correction à chaque étape, telles que la relation entre deux taux de défaillance successifs s'écrive :

$$\Lambda_{i+1} = (1 - \alpha_i - \beta_i)\Lambda_i + \mu\beta_i \qquad \forall i \geqslant 1.$$

Malheureusement, même dans le cas le plus simple, où les deux taux de correction et le premier taux de défaillance sont supposés déterministes et constants, ce modèle est mathématiquement très complexe et pratiquement inutilisable ; aussi, nous restreignons-nous à supposer que l'effet de la correction est simple, c'est-à-dire qu'à chaque étape, l'un des deux taux précédents est nul. On montre alors qu'il existe une suite de variables aléatoires positives $\{\gamma_i\}_{i\geqslant 1}$ telle que la relation entre les taux de défaillance successifs s'écrive :

$$\Lambda_{i+1} = \gamma_i \Lambda_i \quad \forall i \geqslant 1.$$

γ_i représente l'efficacité de la $i^{ème}$ correction. Λ_1 et les γ_i sont indépendantes. Pour des raisons évidentes, les modèles reposant sur cette relation seront appelés **modèles proportionnels**, et nous en étudions deux cas particuliers.

1) Le modèle proportionnel déterministe

Il consiste à supposer que les variables γ_i sont déterministes et toutes égales à un réel positif γ, et que Λ_1 est également déterministe, égale à un réel positif λ. On retrouve le modèle «géométrique» de Moranda [75], qui ne semble pas avoir retenu l'attention qu'il mérite, tant sur le plan pratique que théorique.

Pour des raisons de commodité, on pose $\gamma = e^{-\theta}$, $\theta \in \mathbb{R}$.

Les v.a.r. X_i sont indépendantes, de loi $\Gamma\left(1, \frac{1}{\lambda e^{-(i-1)\theta}}\right)$, pour tout $i \geqslant 1$.

Le paramètre λ représente le taux de défaillance initial, tandis que le paramètre θ représente la qualité de la correction effectuée.

Une étude statistique approfondie de ce modèle est effectuée dans Gaudoin [90] et Gaudoin-Soler [92].

Les estimateurs de maximum de vraisemblance de θ et λ après l'observation de n défaillances, $\widehat{\theta}(X_1, ..., X_n)$ et $\widehat{\lambda}(X_1, ..., X_n)$ sont définis de manière unique par :

$$\widehat{\lambda} = \frac{n}{\sum_{i=1}^{n} e^{-(i-1)\widehat{\theta}} X_i} \qquad \text{et} \qquad \sum_{i=1}^{n} (n - 2i + 1) e^{-(i-1)\widehat{\theta}} X_i = 0.$$

On démontre que $\widehat{\theta}$ est un estimateur sans biais de θ, libre par rapport à λ.

Par ailleurs, on se ramène à un modèle linéaire d'ordre $(n, 2)$:

$$Y = A\beta + \varepsilon$$

avec $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$, en posant $Y_i = \text{Log}(X_i) + \gamma_E$, où γ_E est la constante d'Euler,

$A = \begin{pmatrix} 1 & 0 \\ \vdots & 1 \\ \vdots & \vdots \\ 1 & n-1 \end{pmatrix}$, $\beta = \begin{pmatrix} -\text{Log}\lambda \\ \theta \end{pmatrix}$, et ε est un vecteur aléatoire centré de matrice des covariances $\frac{\pi^2}{6} I_n$, I_n désignant la matrice identité d'ordre n. Ce procédé fournit des estimateurs des moindres carrés. Celui de θ, asymptotiquement gaussien, est une combinaison convexe de variables aléatoires de loi logistique et sa variance est en O (n^{-3}).

2) Le modèle proportionnel lognormal

Le défaut majeur du modèle proportionnel déterministe est évidemment l'hypothèse très irréaliste que la qualité de la correction est constante au cours du temps (égale à θ). Il est probablement plus juste de considérer que celle-ci est variable. Un second modèle consiste à supposer que les qualités des corrections successives sont des variables aléatoires Θ_i indépendantes. Faute d'informations supplémentaires, nous prendrons pour tous les Θ_i une même loi normale $\mathcal{N}(\theta, \sigma^2)$, et $\Lambda_1 = \lambda$. Alors la variable aléatoire $\gamma_i = e^{-\Theta_i}$ est de loi lognormale.

Une étude statistique approfondie de ce modèle est effectuée dans Gaudoin [90] et Gaudoin-Lavergne-Soler [92].

Le calcul de la fonction de vraisemblance étant ici inextricable, on peut se ramener à un modèle linéaire mixte d'ordre $(n, 2)$:

$$Y = A\beta + WZ + \varepsilon$$

avec $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$, en posant $Y_i = \mathrm{Log}(X_i) + \gamma_E$, où γ_E est la constante d'Euler,

$A = \begin{pmatrix} 1 & 0 \\ \vdots & 1 \\ \vdots & \vdots \\ 1 & n-1 \end{pmatrix}$, $\beta = \begin{pmatrix} -\mathrm{Log}\lambda \\ \theta \end{pmatrix}$, $W = \begin{pmatrix} 0 & \dots & \dots & 0 \\ 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 1 & \dots & 1 & 0 \end{pmatrix}$, Z est un vecteur aléatoire gaussien de $\mathbb{R}^n$, de loi $\mathcal{N}(0, \sigma^2 I_n)$, et ε est un vecteur aléatoire centré de $\mathbb{R}^n$, non gaussien, de matrice des covariances $\frac{\pi^2}{6} I_n$, indépendant de Z. Ce procédé permet de construire des estimateurs des moindres carrés de θ et de σ^2, ainsi que des estimateurs MINQUE ou de type Gauss-Markov de σ^2, grâce auxquels on peut d'ailleurs tester l'hypothèse «$\sigma^2 = 0$», pour mettre ce dernier modèle en concurrence avec le précédent.

Ces deux modèles ont été utilisés sur plusieurs jeux de données réelles de défaillances de logiciels, en concurrence avec d'autres modèles. Les tests de validité prédictive, basés sur un indicateur de Kolmogorov-Smirnov (méthode de «U-plot», voir Keiller et al [83]), ont démontré leur pertinence et parfois leur supériorité (Gaudoin [90]).

V. Conclusion

Nous proposons une modélisation stochastique des phénomènes de défaillances des sytèmes présentant des fautes de conception, qui se veut la plus large possible. Elle prend en compte un maximum de facteurs régissant le comportement de ces systèmes vis-à-vis de la défaillance, à savoir le processus de sollicitation et l'effet des corrections des fautes. Elle englobe tous les modèles de fiabilité des logiciels existant, et permet de les comparer. Enfin, elle permet de construire une nouvelle classe de modèles basée sur des hypothèses réalistes, qui conduisent à des problèmes de statistique mathématique très intéressants et dont l'application à des données réelles donne des résultats satisfaisants.

Références

Gaudoin 0. (1990) «Outils statistiques pour l'évaluation de la fiabilité des logiciels», Thèse, Université Joseph Fourier, Grenoble.

Gaudoin 0. , Lavergne C. & Soler J.L. (1992) «A Bayesian proportional model for software reliability», soumis à *IEEE Transactions on Reliability*.

Gaudoin 0. & Soler J.L. (1992) «Statistical analysis of a software reliability model and a new trend test», *IEEE Transactions on Reliability*, sept. 92.

Goel A.L. & Okumoto K. (1979) «Time dependent error detection rate model for software reliability and other performance measures», *IEEE Transactions on Reliability*, R 28, **3**, 206-211.

Jelinski Z. & Moranda P.B. (1972) «Software reliability research», Statistical computer performance evaluation, pp 465-484. New-York : Academic Press.

Kanoun K. & Laprie J.C. (1985) «Modeling software reliability and availability from development-validation up to operation», Rapport de Recherche L.A.A.S., n° 85-042, Toulouse.

Keiller P.A., Littlewood B., Miller D.R., Sofer A. (1983) : «Comparison of software reliability predictions», *IEEE FCTS*, **13**, pp 128-134.

Littlewood B. & Verral J. (1973) «A bayesian reliability growth model for computer software», *Journal of the Royal Statistical Society*, C, **22**, 332-336.

Moranda P.B. (1975) «Event altered rate models for general reliability analysis», *IEEE Transactions on Reliability*, R 28, **5**, 376-381.

Musa J.D. & Okumoto K. (1984), «A logarithmic Poisson execution time model for software reliability measurement», *Proc. 7th Int. Conf. on software engineering*, Orlando, 3-7 Oct. 1988, pp 230-238.

Soler J.L. (1988) «Modélisation des processus de risque, de défaillance et de correction des systèmes présentant des fautes de conception. Application à la fiabilité des logiciels», *Proc. 6th Int. Conf. on reliability and maintainability*, Strasbourg, pp 647-650.