

PRONOSTIC

RAPPORT DE STAGE

Présenté dans le cadre de l'obtention du diplôme



SEPTEMBRE 2011

Étude réalisée par M. Raouf HAMOUDA

Travaux encadrés par :

– M. Vincent Feuillard et Melle Ariane Lorton, Tuteurs EADS IW

Et

– M. Marc Wouts, Tuteur École

Visa du responsable de stage	Visa du service de sécurité
Date et signature	Date et signature

Ce document n'a pas fait l'objet d'une décision de classification.

Ce rapport contient 63 pages numérotées n° 1/63 à 63/63

3 exemplaires ont été édités sous format papier

Remerciements

Je tiens à remercier dans un premier temps, toute l'équipe pédagogique de *Ingénieurs Sup Galilée* de l'université Paris-Nord.

Je tiens à remercier tout particulièrement monsieur *Olivier LAFITTE*, le directeur de la filière "Mathématiques appliquées et Calcul Scientifique" qui m'a soutenu pendant mes années d'étude.

Je tiens à témoigner toute ma reconnaissance à monsieur *Marc WOUTS* de m'avoir fait confiance.

J'exprime toute ma gratitude à l'égard de mes encadrants en entreprise, pour m'avoir suivi et conseillé pendant ces six mois de stage.

Table des matières

1	Introduction	7
1.1	Cadre	7
1.2	Maintenance conditionnelle	8
1.3	Objectifs	9
1.4	Plan	9
2	Présentation de l'exemple	11
2.1	L'Exemple	11
2.2	Modélisation	11
2.3	Données de simulations	12
3	Présentation des quantités d'intérêt	14
3.1	Cas Général	14
3.1.1	Système non réparable	14
3.1.2	Système réparable	16
3.2	Cas markovien	17
4	Approche Globale	18
4.1	Contexte	18
4.2	Estimation non-paramétrique	19
4.2.1	Estimation non paramétrique de la Fiabilité	19
4.2.2	Taux de défaillance à partir de la fiabilité	22
4.2.3	MRL	24
4.3	Estimation paramétrique	25
4.3.1	Modèle HPP	25
4.3.2	Modèle NHPP	34
4.4	Tests statistiques et validation de modèle	36
4.4.1	Test de Comparaison de Kolmogorov-Smirnov	37
4.4.2	Test de Comparaison entre IC	37
4.4.3	Test d'adéquation au HPP :	39
4.4.4	Test d'hypothèse sur λ dans le modèle HPP	39
4.4.5	Test de tendance	41
4.4.6	Tests de tendance appliqué au cas Weibull	43
4.4.7	Tests de tendance appliqué aux données du système	48
4.4.8	Test β du modèle PLP	48
4.5	Conclusion	50
5	Approche structurelle	52
5.1	Contexte	52
5.2	Estimation de paramètres	52
5.2.1	Estimation de λ_1 via l'estimateur de NA	53
5.2.2	Estimation de λ_1 via EMV	54
5.2.3	Retour sur l'approche Globale	57
5.2.4	Estimateur de Kaplan-Meier	58
5.3	Conclusion	59

6 Conclusion	60
Références	63

1 Introduction

1.1 Cadre

European Aeronautic Defense and Space Company (EADS) est un groupe industriel du secteur de l'industrie aéronautique et spatiale civile et militaire. La division Airbus, grand constructeur d'avions, applique un soin tout particulier à la recherche et à l'innovation dans tous les domaines associés à son activité. A ce titre, l'organisation et la méthode de la maintenance des appareils est une des cibles de la recherche menée au sein du centre de recherche d'EADS Innovation Works.

Actuellement, Airbus offre aux acheteurs de ses avions son savoir-faire technique en termes de maintenance de la flotte. Airbus fournit à cet effet à ses clients une expertise dans le domaine de la maintenance technique des matériels, des recommandations concernant la fiabilité des avions ainsi que l'analyse des coûts de maintenance et les bonnes solutions techniques au bon moment. Par ailleurs pour minimiser l'immobilisation au sol des appareils pour cause de réparations, Airbus met également à disposition 24h/24, des ingénieurs spécialistes via un "call-center". Malgré la qualité des services proposés, la recherche d'Airbus se concentre activement sur de *nouveaux modèles de gestion de la maintenance* et donc de la fiabilité de la flotte aéronautique.

L'activité d'Airbus implique une importance considérable de la maintenance des appareils. En effet, une bonne maintenance des avions permet de s'assurer de la sécurité et ainsi de bénéficier d'une bonne image de constructeurs fiables et soucieux de la qualité de leurs appareils.

La maintenance assure deux fonctions principales et indispensables dans le cadre de l'industrie. D'une part, elle permet de garantir la bonne marche et la longévité des avions délivrés aux clients tels que les grandes compagnies aériennes. Elle est également une assurance de la sécurité de leurs appareils. Ainsi, un bon suivi technique des avions entraîne une confiance d'autant plus grande accordée à leur caractère opérationnel.

La maintenance constitue alors un point clé de la construction aéronautique : il s'agit de suivre au mieux les pistes d'améliorations de ce domaine, aujourd'hui plutôt subi que maîtrisé.

La maintenance classique peut emprunter deux voies différentes dans le cadre de l'industrie. Une première orientation est de profiter des immobilisations des matériels pour inspecter entièrement ceux-ci et remplacer tout ce qui doit l'être : il s'agit de la maintenance corrective. Une seconde vision de la maintenance est la réaction à une défaillance, une panne ou autre dégradation du matériel.

Les deux approches de la maintenance des avions engagent des coûts considérables :

- Immobilisation du matériel
- Conséquences en termes d'image de la marque
- Mise à disposition d'équipes de maintenance réactives
- Constitution de stocks de pièces détachées au cas où.

Une telle conception de la maintenance des avions n'est en aucun cas optimisée, et présente un manque à gagner en termes de fiabilité, efficacité et coûts engendrés. Pour augmenter la disponibilité et la sûreté de la flotte d'avions tout en réduisant les coûts associés, de nouvelles formes de maintenance doivent être développées. Pour mieux résister à la concurrence et aux circonstances économiques, la maintenance a aussi son rôle à jouer dans la compétitivité. C'est pourquoi on envisage un nouveau type de maintenance : préventive et conditionnelle.

1.2 Maintenance conditionnelle

Une nouvelle façon de gérer la maintenance et surtout de l'optimiser est la maintenance conditionnelle. Cette forme de maintenance se fonde principalement sur la mesure et l'analyse de données issues de capteurs, pour déterminer et prédire une défaillance. La maintenance conditionnelle peut être qualifiée de prévisionnelle car elle permet de mieux gérer les interventions, en fonction de l'état réel des organes d'une machine, de leur usure ou dégradation. C'est la forme la plus aboutie de maintenance.

En d'autres termes la maintenance conditionnelle est définie comme l'anticipation des pannes avec suffisamment de précision pour déclencher une action préventive avant l'immobilisation des avions, et ce à moindre coût, donc de façon optimisée.

Le développement de ce type de maintenance provient du compromis qu'il faut établir entre l'ensemble des coûts induits par une politique d'attente de panne et de réparation, et une politique de maintenance préventive systématique.

Ainsi la démarche de maintenance conditionnelle est constituée de 3 étapes :

- Le *monitoring du système* considéré : il faut observer le système et pouvoir en retirer un maximum d'informations.
- L'*établissement d'un diagnostic* issu de ces informations : on cherche à obtenir grâce à ces données une courbe de fiabilité de l'équipement (La fiabilité sera définie dans une prochaine section)
- *Pronostic du système puis décision* : Grâce au diagnostic précédent, on va pouvoir prendre une décision concernant la programmation d'une réparation, remplacement de composants ou non ; sur la fiabilité précédemment obtenue, on délimite une zone critique du système pour laquelle une réparation sera indispensable.

Le processus de pronostic est l'élément clé de la stratégie de maintenance prévisionnelle car il supporte la notion d'anticipation qui caractérise une approche proactive. Une telle gestion de la maintenance demande un changement radical des méthodes habituelles et une étude très approfondie des systèmes considérés. En effet, connaître parfaitement le comportement du système est indispensable pour l'établissement de diagnostics. Une fois effectuées ces recherches amont, à la fois au niveau du comportement des systèmes des avions Airbus et du cadre mathématique, EADS aurait pour ambition de fournir au client un service de suivi sous forme de maintenance conditionnelle de leurs avions : suivi et analyse de données, diagnostics et mise en place d'interventions planifiées. Il ne s'agira plus seulement de vendre du matériel mais des heures de vols, garanties par le service de maintenance conditionnelle associé. La R&D est très active sur ce sujet où il s'agit de travaux amonts qui permettraient de définir un cadre de modélisation mathématique, plus précisément probabiliste et statistique pour la mise en oeuvre de la maintenance conditionnelle.

Aux prémices de cette recherche, il faut donc tenter de développer des méthodes simples, et mathématiquement robustes de pronostic de systèmes qui soient applicables aux avions. Ces travaux font l'objet d'une thèse menée par l'une des personnes encadrant mon stage au sein du département de mathématiques appliquées de EADS IW. Le lecteur pourra consulter [10] pour plus d'information sur les enjeux de cette nouvelle forme de maintenance.

1.3 Objectifs

Le but du stage consiste donc à :

- mettre en place des méthodes statistiques pour le traitement de l'information disponible, c'est-à-dire celle obtenue à l'issue de la première étape de monitoring du système considéré
- établir un diagnostic issu de ces informations, c'est-à-dire dans le cadre de ce stage, à estimer certaines quantités qui caractériseront le comportement du système du point de vue de sa fiabilité, et qui seront indispensables à la mise en place d'un pronostic pour notre équipement
- construire des méthodes permettant de faire le choix d'un modèle simple qui soit en même temps capable de modéliser notre système avec un seuil de confiance donné
- trouver des méthodes qui appliqueraient également ces considérations aux composants du système (approche structurelle) au lieu du système tout entier (approche globale).

Afin de mettre en place ces méthodes, nous allons considérer un système simple qu'on modélisera par un processus markovien de saut. L'étude menée se fera donc à un niveau d'application très réduit puisque réalisée sur un exemple simple pour faciliter la compréhension de la construction, mais s'inscrit dans un processus global de recherche dans le domaine de la modernisation de la maintenance dans l'aéronautique.

1.4 Plan

Nous présenterons dans un premier temps le système qui nous permettra de construire notre processus de pronostic (section 2.1) ainsi que la façon dont celui-ci se comporte (section 2.2).

Nous définirons dans un cadre général, les quantités caractérisant la façon dont un système tombe en panne (section 3.1), puis donnerons les résultats correspondant dans le cas de notre système modélisé par un processus markovien de sauts (section 3.2).

Nous définirons l'espace des observables (section 4.1), c'est-à-dire les quantités qu'on est susceptible d'observer à l'issue du monitoring du système considéré. Ces quantités permettront l'estimation des paramètres qui nous intéressent, d'abord dans un cadre non-paramétrique (section 4.2) puis dans un cadre paramétrique (section 4.3).

Le cadre paramétrique suppose un choix de modèle. On sera amené (section 4.3.1) à quantifier l'erreur qu'on fait quand on choisit un modèle en particulier, qu'on appellera par la suite *modèle de défaillance*.

Nous développerons ensuite des méthodes statistiques de tests permettant de quantifier le seuil de confiance qu'on s'autorise suite au choix de tel ou tel modèle de défaillance pour notre

système (section 4.4).

L'approche adoptée dans les points précédents est une approche de type global puisqu'elle s'intéresse au système dans son ensemble, sans considération des éléments qui le composent. Nous développerons en fait également une approche de type structurel (section 5), où l'on s'intéressera à l'estimation de la propension que chacun des composants a de tomber en panne (section 5.2). Cela nous conduira à définir la notion de censure, qui découle du fait que l'observation d'un composant donné peut s'avérer incomplète à cause de son interaction avec les autres composants du système. Nous proposerons dans ce cadre d'autres estimateurs qui tiennent compte de ce type de données (sections 5.2.1 et 5.2.2).

2 Présentation de l'exemple

2.1 L'Exemple

Tout au long de ce stage, les résultats numériques seront donnés dans le cadre de la simulation d'un système formé de deux générateurs 1 et 2 montés en parallèle, alimentant une lampe 3. Chaque composant sera caractérisé par son taux de défaillance ainsi que son taux de réparation. Intuitivement le taux de défaillance d'un composant à un instant t représente sa propension à avoir une défaillance juste après cet instant, sachant combien de défaillances il a subi avant t et quand elles se sont produites.

Pour les générateurs montés en parallèle, il est raisonnable de supposer que la panne du générateur 1, par exemple, augmente la propension du générateur 2 à tomber en panne puisque ce dernier supportera toute la charge de travail pour alimenter la lampe 3. Ceci revient à dire que le taux de défaillance du générateur 2 sera plus important en cas de panne du générateur 1. Dans ces conditions nous considérerons les processus suivants :

λ_1 , taux de défaillance du composant 1 sachant que le composant 2 est en état de marche

λ'_1 , taux de défaillance du composant 1 sachant que le composant 2 est en panne

λ_2 , taux de défaillance du composant 2 sachant que le composant 1 est en état de marche

λ'_2 , taux de défaillance du composant 2 sachant que le composant 1 est en panne.

Ceci conduit au montage Figure 1.

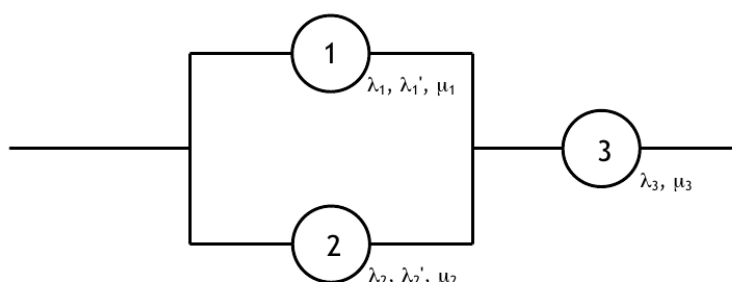


FIGURE 1 – Montage du système étudié

2.2 Modélisation

Un composant en état de panne sera dans un état identifié par 0. Il sera dans un état égal à 1 sinon. L'état du composant i sera modélisé par le processus noté Z_i . A tout instant t et pour chaque composant i , on a donc $Z_{i,t} \in \{0, 1\}$. L'état du système sera modélisé par le processus $Z = (Z_1, Z_2, Z_3)$ à valeurs dans l'espace des états $\{0, 1\}^3$. Ordonnons cet espace de façon à ranger les états de marche en premier. Notons \mathcal{M} l'ensemble des états de marche du système considéré, c'est à dire l'ensemble ordonné $\mathcal{M} = \{e_1, e_2, e_3\} = \{(1, 1, 1), (1, 0, 1), (0, 1, 1)\}$. Soit $\mathcal{P} = \{e_4, e_5, e_6, e_7, e_8\} = \{(1, 1, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (0, 0, 0)\}$ l'ensemble ordonné des états de pannes. On peut alors poser $\mathcal{E} = \mathcal{M} \amalg \mathcal{P}$ l'espace ordonné des états du système.

Nous ferons l'hypothèse que le système passe d'un état à l'autre à des instants aléatoires, les transitions se faisant par sauts successifs. On supposera également que l'état du système à

un instant t ne dépend de son "histoire" qu'à travers l'état correspondant au dernier saut (celui juste avant l'instant t). Ceci justifie la modélisation par un processus markovien de sauts, ce qui revient à supposer que les taux de défaillance et de réparation des composants sont constants. Nous pouvons alors considérer (Figure 2) le graphe des états du système.

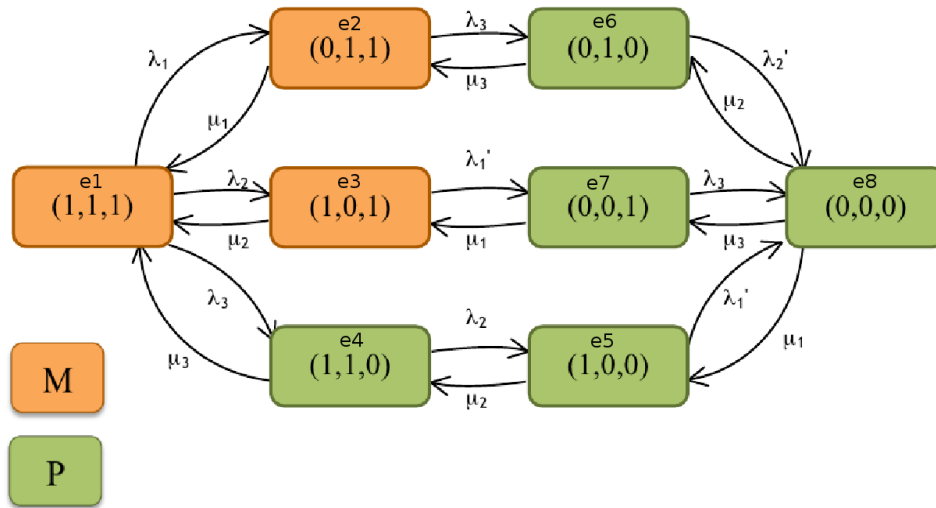


FIGURE 2 – Graphe des états

Notations : Fixons les notations utilisées dans ce rapport :

- n le nombre des états du système : $n = \text{card}(\mathcal{E}) (= 8)$
- $F_{\mathbb{L}}^{-1}$ l'inverse de la fonction de répartition d'une variable aléatoire suivant une loi \mathbb{L}
- A la matrice génératrice du système
- ν la loi du processus Markovien Z : pour tout instant t , $\mathbb{P}(Z_t = e_i) = \nu_t(e_i)$
- $B = A|_{\mathcal{M}}$ la sous-matrice restreinte aux états de marche \mathcal{M}
- On identifiera ν_t au vecteur ligne $(\nu_t(e_i))_i$
- π la restriction de la loi initiale à \mathcal{M} : $\pi = \nu_0|_{\mathcal{M}}$
- $\mathbf{1}_{\mathcal{M}}$ le vecteur colonne de $\mathbb{R}^{\text{card}(\mathcal{M})}$ de composantes toutes égales à 1 :

$$\mathbf{1}_{\mathcal{M}} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

2.3 Données de simulations

On supposera pour les simulations que $Z_0 \stackrel{ps}{=} e_1$, c'est-à-dire $\nu_0 = (1, 0, 0, 0, 0, 0, 0, 0)$, d'où $\pi = (1, 0, 0)$. Cela signifie qu'à l'instant initial, tous les composants fonctionnent.

Pour les applications numériques, on prendra :

$$\lambda_1 = 0.001, \lambda'_1 = 0.0015, \lambda_2 = 0.002, \lambda'_2 = 0.0025, \lambda_3 = 0.00005$$

$$\mu_1 = 0.01, \mu_2 = 0.002, \mu_3 = 0.5$$

Et vu le graphe des états (Figure 2), on obtient la matrice génératrice :

$$A = \begin{bmatrix} a_{1,1} & \lambda_1 & \lambda_2 & \lambda_3 & 0 & 0 & 0 & 0 \\ \mu_1 & a_{2,2} & 0 & 0 & 0 & \lambda_3 & \lambda'_2 & 0 \\ \mu_2 & 0 & a_{3,3} & 0 & \lambda_3 & 0 & \lambda'_1 & 0 \\ \mu_3 & 0 & 0 & a_{4,4} & \lambda_2 & \lambda_1 & 0 & 0 \\ 0 & 0 & \mu_3 & \mu_2 & a_{5,5} & 0 & 0 & \lambda'_1 \\ 0 & \mu_3 & 0 & \mu_1 & 0 & a_{6,6} & 0 & \lambda'_2 \\ 0 & \mu_2 & \mu_1 & 0 & 0 & 0 & a_{7,7} & \lambda_3 \\ 0 & 0 & 0 & 0 & \mu_1 & \mu_2 & \mu_3 & a_{8,8} \end{bmatrix}$$

avec

$$a_{i,i} = - \sum_{j \neq i} a_{i,j}$$

d'où :

$$B = \begin{bmatrix} a_{1,1} & \lambda_1 & \lambda_2 \\ \mu_1 & a_{2,2} & 0 \\ \mu_2 & 0 & a_{3,3} \end{bmatrix}$$

3 Présentation des quantités d'intérêt

Nous allons définir dans ce paragraphe les quantités que nous allons estimer dans le cadre de ce stage. Nous donnerons par la suite les expressions analytiques correspondantes, dans notre hypothèse d'un système suivant un modèle markovien de sauts.

3.1 Cas Général

On se place dans le cas du système régi par le processus Z non nécessairement markovien. On fera l'hypothèse que $Z_0 \in \mathcal{M}$. Une notion importante est celle de *fiabilité* à un instant t :

Définition 3.1. *La fiabilité à l'instant t d'un système, est la probabilité qu'il fonctionne sans défaillance jusqu'à cet instant. On notera cette fiabilité $R(t)$ et on a par définition :*

$$R(t) = \mathbb{P}(\forall s \in [0, t], Z_s \in \mathcal{M})$$

Dans le cadre de la fiabilité, l'objectif est de s'assurer qu'à tout instant t , la quantité $R(t)$ ne passe pas en dessous d'une certaine valeur $\underline{R}(t)$ avec un seuil de confiance α petit, c'est-à-dire $\mathbb{P}(\forall t, \underline{R}(t) < R(t)) = 1 - \alpha$. Par la suite, nous allons distinguer le cas d'un système non réparable et celui d'un système réparable.

3.1.1 Système non réparable

Dans ce cas, seule la date de panne appelée aussi instant de défaillance est à prendre en considération. Cette date étant imprévisible, il convient donc de la modéliser par une variable aléatoire que nous noterons T par la suite.

Définition 3.2. *La fiabilité à l'instant t d'un système non réparable, est la probabilité que le système fonctionne sans défaillance jusqu'à cet instant. On a alors dans ce cas particulier :*

$$R(t) = \mathbb{P}(T > t)$$

Si F est la fonction de répartition de T on a $R(t) = 1 - F(t)$, ce qui montre que la connaissance de la fiabilité R détermine entièrement la loi de T . Cette relation montre aussi la décroissance de la fiabilité, traduisant que l'aptitude au bon fonctionnement d'un système non réparable diminue avec le temps. Une autre notion importante qu'on va pouvoir lier à la fiabilité est celle de taux de défaillance :

Définition 3.3. *Le taux de défaillance, ou taux de panne ou taux de hasard d'un système non réparable est la fonction du temps définie par :*

$$\forall t \geq 0 \quad \lambda(t) = \lim_{dt \rightarrow 0} \frac{\mathbb{P}(t < T < t + dt | T > t)}{dt}$$

On remarque que lorsque dt est petit, $\mathbb{P}(t < T < t + dt | T > t)$ est approximativement égal à $\lambda(t) dt$, qui vaut donc la probabilité que le système tombe en panne *juste après l'instant t* , sachant qu'il n'est pas tombé en panne avant cet instant.

Proposition 3.1. *La fiabilité R et le taux de défaillance λ vérifient :*

$$\forall t \geq 0 \quad R(t) = \exp\left(-\int_0^t \lambda(s) \, ds\right) \quad (1)$$

Ceci montre que la fiabilité est une fonction de $\int_0^t \lambda(s) \, ds$ appelée taux de défaillance cumulé :

Définition 3.4. *Le taux de défaillance cumulé d'un système non réparable est la fonction du temps Λ définie par :*

$$\forall t \geq 0 \quad \Lambda(t) = \int_0^t \lambda(s) \, ds = -\ln R(t) \quad (2)$$

Nous avons donc :

$$\lambda(t) = -\frac{d}{dt} \ln R(t) = -\frac{R'(t)}{R(t)} = \frac{f(t)}{R(t)} \quad (3)$$

En tenant compte de (1), il vient que :

$$\forall t \geq 0 \quad f(t) = \lambda(t) \exp\left(-\int_0^t \lambda(s) \, ds\right)$$

Ceci entraîne que la donnée du taux de défaillance λ détermine entièrement la loi de T . Par conséquent, en pratique, la construction d'un modèle de fiabilité des systèmes non réparables revient à se donner une forme particulière pour le taux de défaillance λ .

Définition 3.5. *Le MTTF (Mean Time to Failure) d'un système non réparable est la durée moyenne de bon fonctionnement avant sa défaillance :*

$$MTTF = \mathbb{E}(T)$$

Sous l'hypothèse que $t R(t) \xrightarrow[t \rightarrow \infty]{} 0$, on montre que :

$$MTTF = \int_0^\infty R(t) \, dt$$

La prise en compte des différents types d'usure rencontrés dans l'industrie ne peut se faire par la simple connaissance de la fiabilité. Une quantité intéressante est celle de la fiabilité en t sur un horizon de temps τ , qui fait apparaître la notion de durée de vie résiduelle d'un système non réparable :

Définition 3.6. *On appelle temps résiduel ou durée de vie résiduelle ou RUL (Remaining Useful Life), à l'instant t d'un système non réparable, la variable aléatoire $T - t$.*

Cette quantité apparaît dans :

Définition 3.7. *La fiabilité en t sur un horizon de temps τ est donnée par*

$$R_t(\tau) = \mathbb{P}(T - t > \tau | T > t) = \frac{R(t + \tau)}{R(t)}$$

Mais également dans :

Définition 3.8. On appelle moyenne de la durée de vie résiduelle ou MRL (Mean Residual Life), la variable aléatoire définie par :

$$\forall t \geq 0 \quad m(t) = \mathbb{E}(T - t | T > t)$$

L'étude de cette quantité a permis l'obtention de nombreux résultats ([7] et [2]). Certain auteurs l'appellent MEF (Mean Excess Functions) dans le domaine du risk management, de l'assurance ou dans l'étude des valeurs extrêmes [11]. La MRL détermine entièrement la loi de T , au même titre que la fiabilité et le taux de défaillance [4]. Le champ d'applications de la MRL est très large : finance, assurance, analyse de survie dans le domaine biomedical [6].

Nous remarquons que $m(0) = \mathbb{E}(T)$ est le MTTF du système réparable et que :

$$m(t) = \int_0^\infty R_t(s) ds = \frac{\int_t^\infty R(s) ds}{R(t)} = \frac{m(0) - \int_0^t R(s) ds}{R(t)} \quad (4)$$

3.1.2 Système réparable

Définition 3.9. La fiabilité à l'instant t du système réparable est donnée par :

$$R(t) = \mathbb{P}(X_1 > t)$$

X_1 étant la première durée de bon fonctionnement ou premier instant de panne dont on continuera à noter f la densité et F la fonction de répartition.

Dans le cas du système réparable, deux configurations sont envisageables :

- a) On tient compte du temps passé à réparer le système qui inclura le temps de détection de la panne et la durée de remise en service. Le système passe par une succession de pannes et de marches entrelacées par des durées de réparations et de fonctionnements. Dans ce cas les aléas régissant le système étudié peuvent se résumer en la famille de variables aléatoires $(X_n)_{n \geq 1}$ représentant les durées de bon fonctionnement successives et $(Y_n)_{n \geq 1}$ correspondant aux durées de réparations (Figure3).

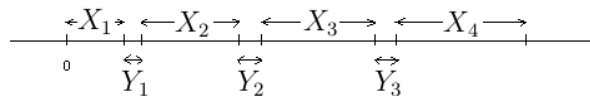


FIGURE 3 – Durée de bon fonctionnement et temps de réparation dans le cas réparable

On peut dès lors définir le taux de réparation :

Définition 3.10. Si Y est une durée de réparation, on appelle taux de réparation associé à Y la fonction du temps μ définie par :

$$\forall t \geq 0 \quad \mu(t) = \lim_{dt \rightarrow 0} \frac{\mathbb{P}(t < Y < t + dt | Y > t)}{dt}$$

- b) On ne tiens pas compte de ces durées de réparation. C'est ce que nous ferons tout le long de ce stage.

3.2 Cas markovien

Dans le cas où le système est régi par un processus markovien de sauts, les quantités d'intérêt s'expriment en fonction de π , B et $\mathbf{1}_{\mathcal{M}}$ (voir [3]) :

Proposition 3.2. *Les agrégats de fiabilité sont alors (voir les notations introduites dans le § 2.2) :*

$$i) R(t) = \pi e^{tB} \mathbf{1}_{\mathcal{M}} \text{ d'où la densité : } f(t) = -R'(t) = -\pi e^{tB} B \mathbf{1}_{\mathcal{M}}$$

$$ii) \lambda(t) = -\frac{\pi e^{tB} B \mathbf{1}_{\mathcal{M}}}{\pi e^{tB} \mathbf{1}_{\mathcal{M}}}$$

$$iii) m(t) = -\frac{\pi B^{-1} e^{tB} \mathbf{1}_{\mathcal{M}}}{\pi e^{tB} \mathbf{1}_{\mathcal{M}}}$$

Dans toute la suite du rapport, nous conserverons ces notations pour les valeurs théoriques des quantités d'intérêt. Comme nous l'avons indiqué, nous allons chercher à estimer ces quantités dans le cadre de l'exemple (§ 1) avec les données de simulations déjà introduites (§ 2.3). Avec ces valeurs, nous avons tracé (Figure 4) la fiabilité, la densité, le taux de défaillance théorique ainsi que la moyenne du temps résiduel de survie en fonction du temps.

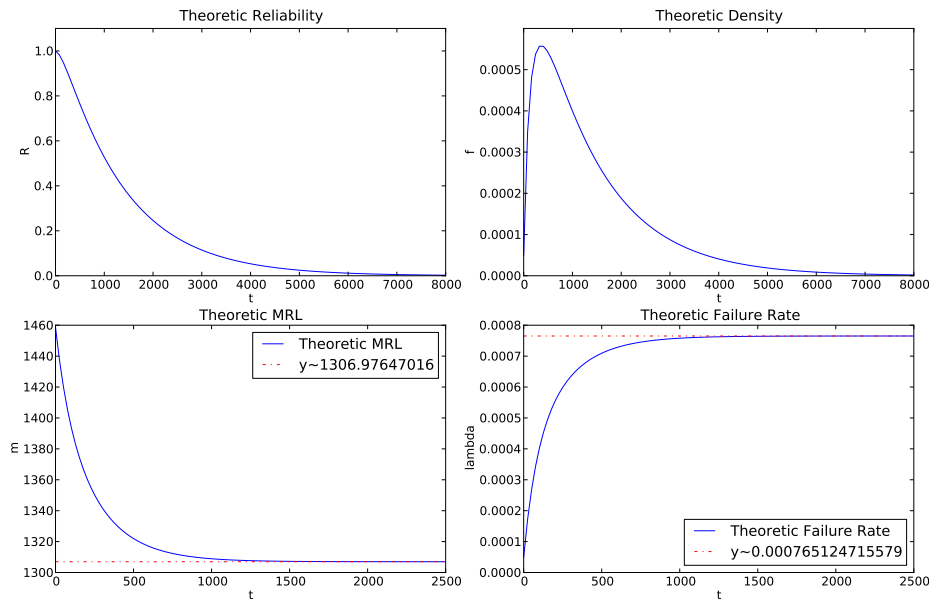


FIGURE 4 – Courbes des quantités d'intérêt théoriques : *fiabilité, densité, moyenne du temps résiduel de survie, et taux de défaillance* du système modélisé par un processus markovien de sauts (les expressions des quantités d'intérêt théoriques étant données par la proposition 3.2). Les valeurs de λ_{∞} et $m_{\infty} = \frac{1}{\lambda_{\infty}}$ ont été calculée grâce au résultat (5) et la valeur de B choisie dans le § 2.3.

Nous remarquons en particulier que le taux de défaillance tend vers une valeur limite que nous noterons λ_∞ dans toute la suite. On montre que (chapitre 9 de [3]) :

$$\lambda_\infty = \min\{|\lambda|; \lambda \text{ valeur propre de } B\} \quad (5)$$

Nous avons également tracé (Figure 5) le graphe de la fonction $t \mapsto m(t)\lambda(t)$. Nous remarquons que $m(t)\lambda(t) \xrightarrow[t \rightarrow \infty]{} 1$, c'est-à-dire que $m_\infty = \frac{1}{\lambda_\infty}$. Ces derniers résultats se démontrent à partir de la proposition 3.2 et des valeurs de π et $1_{\mathcal{M}}$ choisies. Les valeurs de λ_∞ et m_∞ ont été calculé grâce à (5) et la valeur de B choisie dans le § 2.3. Ces valeurs ont été reportées sur les courbes de la Figure 4.

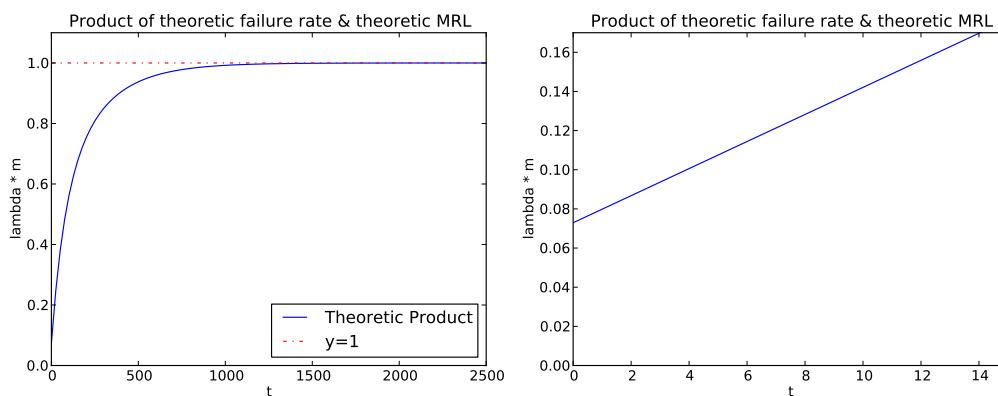


FIGURE 5 – Produit de m et λ théoriques. Sur la courbe de gauche on remarque la convergence de ce produit vers 1 pour t de plus en plus grand. On met en évidence une valeur non nulle de ce produit en zéro sur la courbe de droite, contrairement à ce que laisserait penser la courbe de gauche.

4 Approche Globale

4.1 Contexte

Dans l'approche globale, on considérera le système de l'exemple § 1 comme une sorte de boîte noire, et on s'intéressera à la suite de ses défaillances et réparations.

Dans le cadre de la fiabilité, l'observation de la suite des durées de pannes et de réparations d'un système quelconque a pour objectif de conduire à une modélisation simple de sa fiabilité, de façon que le modèle choisi soit proche de la fiabilité réelle du système observé, c'est-à-dire celle donnée par la proposition 3.2 dans le cas de notre système. Ce choix de modèle ne peut se faire en dehors d'une analyse intégrant des tests statistiques permettant d'affirmer avec un seuil de confiance donné, si ce choix est pertinent au vu de ces observations. Par ailleurs, le fait que nous ayons modélisé le système par un processus markovien de sauts permettra de calculer l'erreur qu'on fait quand on choisi tel ou tel modèle pour la fiabilité.

1. Dans un premier temps on supposera que l'on ne peut observer que les instants de défaillance de N copie du système. Cela revient à se donner (T_1, \dots, T_N) , N instants de

défaillance indépendants. On notera $(T_{(1)}, \dots, T_{(N)})$ la statistique d'ordre, qui dans le cadre du modèle étudié vérifie $T_{(1)} < \dots < T_{(N)}$ (il y a absence d'ex-aequo). Nous proposerons dans ce contexte un estimateur non paramétrique de la fiabilité et mettrons en évidence la convergence de cet estimateur en introduisant une mesure de l'erreur, c'est-à-dire d'une distance entre la fiabilité théorique donnée par la proposition 3.2, et l'estimateur proposé pour une taille N de plus en plus grande de l'échantillon (T_1, \dots, T_N) . Ceci sera traité en utilisant des méthodes de type Monte Carlo. Une fois que nous avons estimé la fiabilité, nous pourrions dès lors estimer le taux de défaillance grâce à :

$$\lambda(t) = -\frac{R'(t)}{R(t)}$$

Nous proposerons de même un estimateur non paramétrique de la MRL.

Par suite, nous suggérerons une modélisation du processus de défaillance en passant par la fiabilité, et chercherons à construire des méthodes statistiques qui pourront s'apparenter à des techniques de validation de modèle. Pour ce faire, nous proposerons des estimateurs paramétriques du taux de défaillance, donc de la fiabilité grâce à $R(t) = \exp(-\Lambda(t))$.

2. *Dans un second temps* on supposera avoir accès à N durées de bon fonctionnement du système supposé réparable, passant alternativement par N états de marche, et N états de réparation. Dans ce cas réparable, on simulera une trajectoire du système markovien, enregistrant alternativement N durées de bon fonctionnement et N durées de réparation. Comme nous l'avons indiqué, on s'intéressera exclusivement aux N durées de bon fonctionnement (X_1, \dots, X_N) , ce qui revient à négliger les N durées de réparation (Y_1, \dots, Y_N) . De même que dans le cas non réparable, nous suggérerons une modélisation de la fiabilité suivi d'une étape de validation de modèle.

On cherchera à estimer le taux de défaillance, en nous interrogeant à chaque fois sur la validité de ces estimations et leurs cadres d'application. Comme nous venons de l'indiquer, une fois ce taux estimé, nous pourrions alors estimer la fiabilité grâce à $R(t) = \exp(-\Lambda(t))$. Nous émettrons ensuite des hypothèses de modèle en supposant une forme particulière pour le taux de défaillance afin de pouvoir l'estimer dans ce cadre particulier. On procédera par la suite à des tests statistiques à partir des observations de notre système.

4.2 Estimation non-paramétrique

Comme nous venons d'annoncer, nous pouvons estimer certaines quantités d'intérêt, comme la fiabilité, grâce au N -échantillon (T_1, \dots, T_N) vérifiant $T_{(1)} < \dots < T_{(N)}$ dans notre modèle. On posera par convention $T_0 = 0$. Nous nous intéresserons également à l'estimation de l'erreur quadratique entre la fiabilité théorique et celle estimée, ce qui se fera par simulation de K N -échantillons $(T_1^i, \dots, T_N^i)_{i=1, \dots, K}$.

4.2.1 Estimation non paramétrique de la Fiabilité

Par définition, la fiabilité dans le cas non réparable $R(t) = \mathbb{P}(T > t) = \mathbb{E}(\mathbb{1}_{T>t})$ suggère une estimation par *Monte Carlo* (voir [9]) en posant :

$$\hat{R}_N(t) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{T_i > t} \quad (6)$$

D'où par absence d'ex-aequo :

$$\forall t \in [T_{(k)}, T_{(k+1)}[, \quad \hat{R}_N(t) = \frac{1}{N} \sum_{i=k+1}^N \mathbb{1}_{T_{(i)} > t} = \frac{N-k}{N} \quad (7)$$

La variance empirique de la variable aléatoire $\mathbb{1}_{T > t}$ que nous noterons $\hat{\sigma}_N^2(t)$ vérifie :

$$\forall t \in [T_{(k)}, T_{(k+1)}[, \quad \hat{\sigma}_N^2(t) = \frac{N-k}{N} - \left(\frac{N-k}{N} \right)^2$$

D'où les intervalles de confiance bilatère asymptotiques (ICA) à 95% : $\left[\hat{R}_N(t) \pm \frac{1.96 \hat{\sigma}_N(t)}{\sqrt{N}} \right]$.

Dans le cadre de la fiabilité, l'objectif est de s'assurer qu'à tout instant t , la quantité $R(t)$ ne passe pas en dessous d'une certaine variable aléatoire $\underline{R}(t)$ avec un seuil de confiance α petit, c'est-à-dire $\mathbb{P}(\forall t, \underline{R}(t) < R(t)) = 1 - \alpha$. On s'intéresse donc à des intervalles de confiance unilatères du type :

$$\left[\hat{R}_N(t) - \frac{z_\alpha \hat{\sigma}_N(t)}{\sqrt{N}}, 1 \right] \quad (8)$$

avec $z_\alpha = F_{N(0,1)}^{-1}(1 - \alpha)$. Voici (Figure 6) les courbes des fiabilités, théorique et estimée, avec mise en évidence de la borne inférieure de l'ICA à 95% : $\alpha = 0.05 \Rightarrow z_\alpha \sim 1.6448$.

Estimation non paramétrique de l'erreur : Nous remarquons sur la figure 6, que la taille N de notre échantillon doit être importante pour avoir une estimation satisfaisante de la fiabilité. Nous allons quantifier en fonction de N l'erreur entre la fiabilité théorique R et l'estimateur paramétrique de la fiabilité.

Nous pouvons mettre en évidence de façon ponctuelle l'erreur entre \hat{R}_N et R en un point t_0 donné. Nous avons tracé (Figure 7) l'évolution de $\hat{R}_N(t_0)$ vers $R(t_0)$ pour N de plus en plus grand.

Nous pouvons aussi mettre en évidence l'erreur en fonction de N en considérant l'erreur quadratique :

$$d(\hat{R}_N, R) = \mathbb{E} \left[\int_0^\infty \left(\hat{R}_N(t) - R(t) \right)^2 dt \right] \quad (9)$$

Cette distance s'appelle IMSE (*Integrated Mean Square Error*) entre \hat{R}_N et R . Nous avons tracé (Figure 8) l'évolution de $d(\hat{R}_N, R)$ en fonction de N . Ceci montre que nous avons une estimation satisfaisante pour $N > 2000$.

Méthodologie de calcul : En tenant compte de (7), il vient que :

$$d(\hat{R}_N, R) = \sum_{k=0}^{N-1} \mathbb{E} \left[\int_{T_{(k)}}^{T_{(k+1)}} \left(\frac{N-k}{N} - R(t) \right)^2 dt \right] + \epsilon_N$$

avec

$$\epsilon_N = \mathbb{E} \left[\int_{T_{(N)}}^\infty \left(\hat{R}_N(t) - R(t) \right)^2 dt \right] = \mathbb{E} \left[\int_{T_{(N)}}^\infty R(t)^2 dt \right]$$

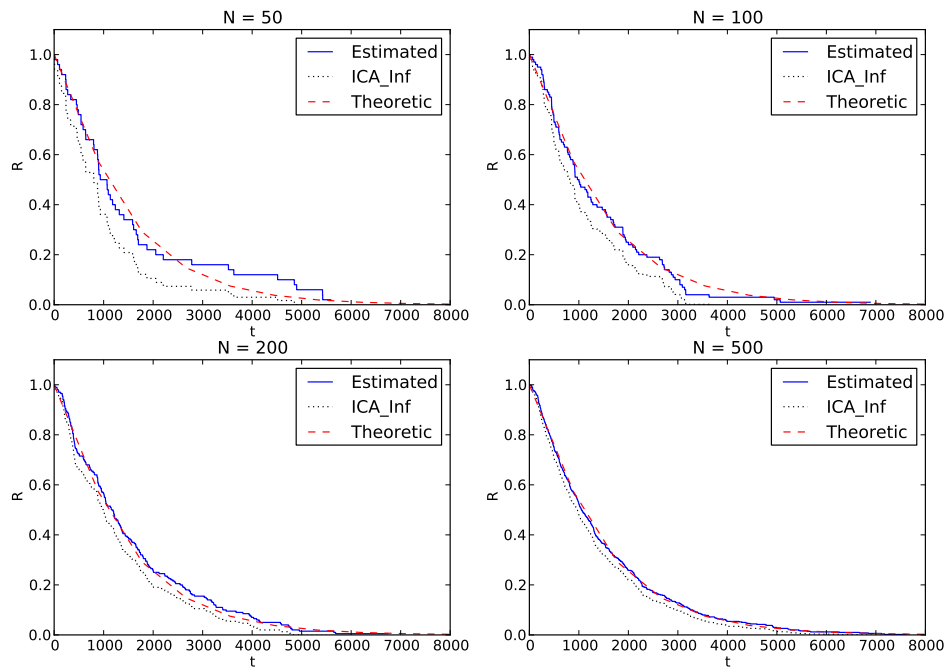


FIGURE 6 – Estimation non paramétrique de la Fiabilité avec ICA(95%) pour différentes tailles d'échantillons : $N = 50, 100, 200, 500$. Sur ces figures apparait la borne inférieure de l'intervalle (8) ainsi que la courbe théorique de la fiabilité donnée par la proposition (3.2) dans le cadre de notre exemple, section 2. Nous remarquons la convergence de l'estimateur non-paramétrique de la fiabilité (6) vers la fiabilité théorique R pour N de plus en plus grand.

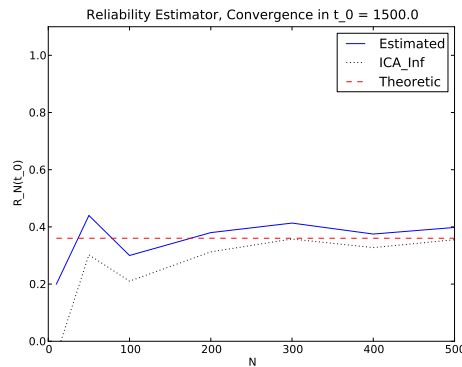


FIGURE 7 – Convergence ponctuelle au point $t_0 = 1500$ de l'estimateur non paramétrique (6) de la fiabilité. Nous avons tracé la valeur théorique $R(t_0)$ et celles de $\hat{R}_N(t_0)$ pour N de plus en plus grand avec mise en évidence de l'ICA(95%) (8). Nous vérifions $\hat{R}_N(t_0) \xrightarrow{N \rightarrow \infty} R(t_0)$.

car $\hat{R}_N(t) = 0$ pour $t > T_{(N)}$ d'après (6).

Avec les données de simulations choisies, nous remarquons sur les courbes de la Figure 6 que

pour $N > 100$, $R(t)$ est quasiment nulle quand $t > T_{(N)}$, il en sera donc *a priori* de même pour ϵ_N qu'on cherchera également à estimer. Les espérances :

$$\mathbb{E} \left[\int_{T_{(k)}}^{T_{(k+1)}} \left(\frac{N-k}{N} - R(t) \right)^2 dt \right]$$

pourront être estimées par Monte Carlo. Pour cela, il suffit de considérer le K -échantillon $(T_1^i, \dots, T_N^i)_{i=1, \dots, K}$ en simulant de façon indépendante, K fois la panne de N systèmes identiques, K étant fixé et N choisi de plus en plus grand (on prendra $K = 100$). Afin d'estimer ces espérances, on considérera alors l'estimateur de la moyenne empirique :

$$\frac{1}{K} \sum_{i=1}^K \int_{T_{(k)}^i}^{T_{(k+1)}^i} \left(\frac{N-k}{N} - R(t) \right)^2 dt$$

les intégrales, y compris celle dans ϵ_N , étant calculées en utilisant une formule de quadrature (méthode *quad* de la classe *scipy.integrate* dans python).

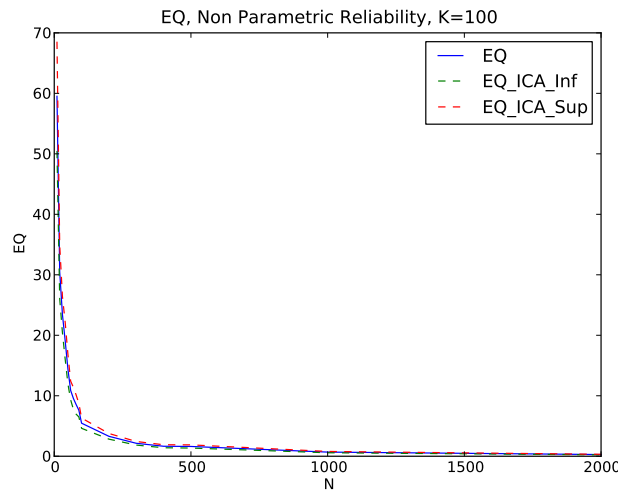


FIGURE 8 – Estimation non paramétrique de l'IMSE relative à la Fiabilité du système. Nous mettons en évidence la convergence vers zéro de l'erreur quadratique entre l'estimateur non paramétrique (6) de la fiabilité, et la fiabilité théorique R donnée par la proposition 3.2, avec les données de l'exemple étudié (§2.3).

4.2.2 Taux de défaillance à partir de la fiabilité

Grâce à l'expression de R donnée par la proposition 3.2 et la formule

$$\lambda(t) = -\frac{R'(t)}{R(t)} \sim -\frac{R(t+dt) - R(t)}{dt \times R(t)} \quad (10)$$

Soit alors une subdivision régulière de l'intervalle sur lequel le processus de défaillance est observé de pas de temps dt fixé "raisonnablement" : $t_1 < t_2 < \dots$, $dt = t_{k+1} - t_k$. En remplaçant R par l'estimateur non paramétrique \hat{R}_N , nous pouvons suggérer l'estimateur

$$\hat{\lambda}_N(t_k) = \frac{\hat{R}_N(t_{k+1}) - \hat{R}_N(t_k)}{dt \times \hat{R}_N(t_k)}$$

comme estimateur de taux de défaillance. Nous avons tracé le résultat graphique sur la Figure 9. Nous obtenons pour un choix judicieux du pas de temps dt , un aperçu très sommaire du taux de défaillance. Nous remarquons que la dépendance de cet estimateur au pas de temps dt le rend dans la pratique peu exploitable dans la réalité.

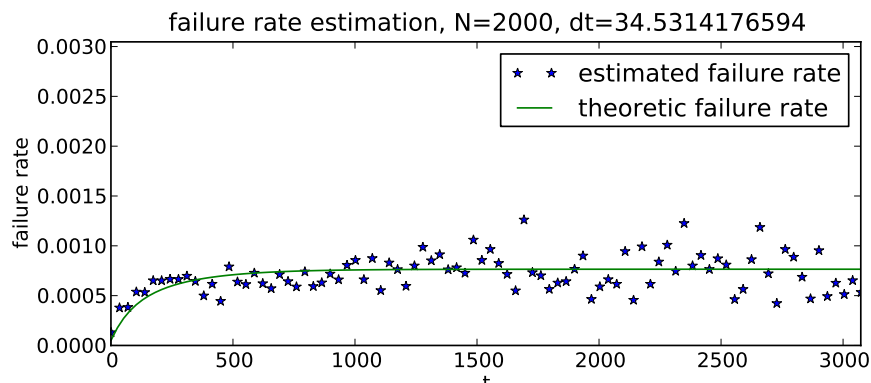


FIGURE 9 – Estimation du taux de défaillance à partir de l'estimateur non paramétrique de la fiabilité. Nous obtenons pour un choix judicieux du pas de temps dt , un aperçu très sommaire du taux de défaillance. Nous remarquons que la dépendance de cet estimateur au pas de temps dt le rend dans la pratique peu exploitable.

4.2.3 MRL

Par définition, la MRL à l'instant t vaut $m(t) = \mathbb{E}(T - t | T > t)$. Donc avec les notations précédentes, en posant :

$$\forall t \in [T_{(k)}, T_{(k+1)}[\quad \hat{m}_N(t) = \frac{1}{N - k} \sum_{j=k+1}^N (T_{(j)} - t) \quad (11)$$

nous obtenons un estimateur consistant de $m(t)$ pour $t \in [T_{(k)}, T_{(k+1)}[$. De même que pour la fiabilité, la variance empirique permettant de construire des intervalles de confiance asymptotiques est donnée par :

$$\forall t \in [T_{(k)}, T_{(k+1)}[, \quad \hat{\sigma}_N^2(t) = \frac{1}{N - k} \sum_{j=k+1}^N (T_{(j)} - t)^2 - \hat{m}_N^2(t)$$

Cela donne les intervalles de confiance asymptotiques à 95% : $\left[\hat{m}_N(t) \pm \frac{1.96 \hat{\sigma}_N(t)}{\sqrt{N}} \right]$ et les courbes (Figure 10) de MRL estimés et théoriques pour $N = 100, 200$.

Avec les données simulées, les tracés (Figure 11) permettent par exemple de dire que si le système est arrivé à mille heures sans tomber en panne, en moyenne il pourra encore fonctionner 1310 heures avant qu'il ne tombe effectivement en panne.

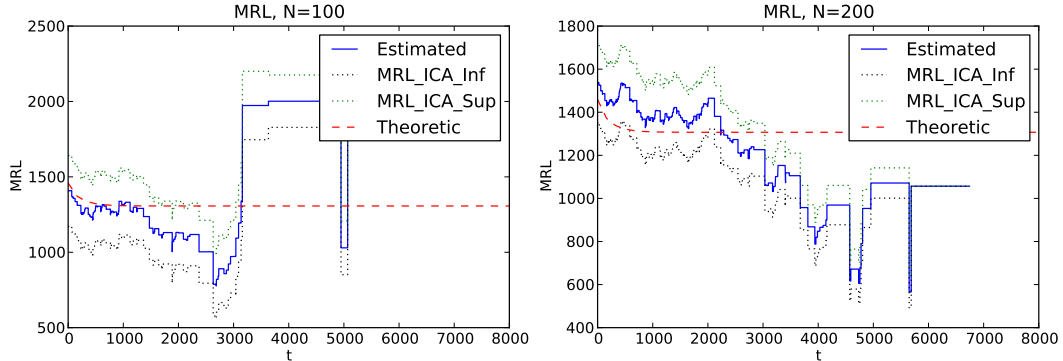


FIGURE 10 – Estimation non paramétrique de la MRL avec mise en évidence de l'ICA(95%). Nous remarquons que les simulations donnent des résultats médiocres sur les grands horizons de temps. Ceci est dû au fait qu'il y a peu de données pour ces horizons de temps.

Nous constatons que l'estimateur non-paramétrique de la MRL est plutôt mauvais, à la fois pour les petites valeurs de N et pour les grands horizons de temps. Il serait possible d'améliorer cet estimateur pour les horizons de temps importants, par exemple à l'aide de la théorie des valeurs extrêmes. Ceci n'est pas l'objectif du rapport dans la mesure où nous cherchons à estimer nos quantités d'intérêt pour les moyens horizons de temps.

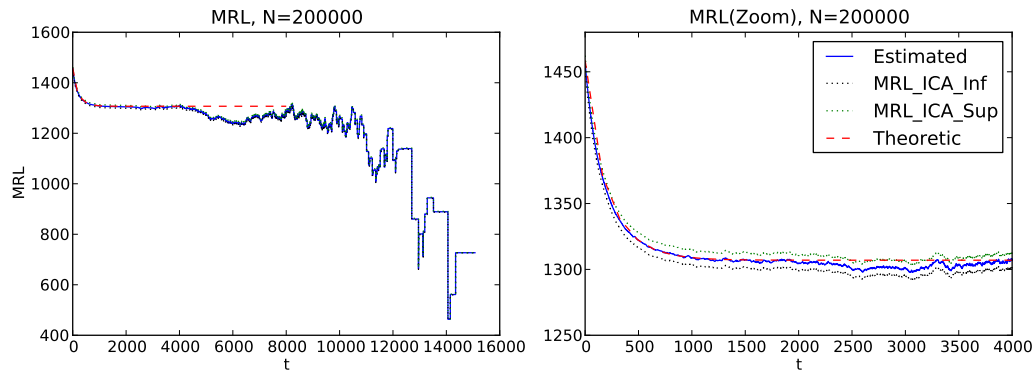


FIGURE 11 – Estimation non paramétrique de la MRL avec mise en évidence de l'ICA(95%). Nous remarquons que les simulations donnent des résultats médiocres sur les grands horizons de temps. Ceci est dû au fait qu'il y a peu de données pour ces horizons de temps. Les tracés permettent par exemple de dire que si le système est arrivé à mille heures sans défaillance, en moyenne il pourra encore fonctionner 1310 heures avant de tomber en panne.

4.3 Estimation paramétrique

Comme nous l'avons indiqué au § 4.1, nous allons estimer le taux de défaillance avec un *a priori* sur sa forme.

1. On cherchera à l'estimer dans un premier temps dans le cadre du modèle HPP (Homogeneous Poisson Process : processus de Poisson homogène) pour lequel il est constant. Cela peut se faire *a priori* indifféremment grâce aux instants de défaillance (T_1, \dots, T_N) dans le cas non réparable, ou bien grâce à ce qu'on a appelé durée de bon fonctionnement (X_1, \dots, X_N) dans le cas réparable. En effet, le fait de supposer que l'on soit dans le cadre d'un modèle HPP, revient à dire que les (X_1, \dots, X_N) sont iid de loi exponentielle de paramètre le taux de défaillance. Or $T_1 = X_1$, et la façon d'obtenir les (T_1, \dots, T_N) supposent qu'ils sont iid, donc de même loi que la loi de X_1 , c'est à dire exponentielle de paramètre le taux de défaillance. Donc (T_1, \dots, T_N) et (X_1, \dots, X_N) sont dans le modèle HPP de même loi.
2. Le modèle NHPP (Non Homogeneous Poisson Process : processus de Poisson non homogène) correspond au cas où λ est une fonction du temps. Sa mise en œuvre nécessitera la prise en compte des durées de bon fonctionnement (X_1, \dots, X_N) exclusivement.

4.3.1 Modèle HPP

Cette hypothèse équivaut à dire que le taux de défaillance du système est une constante λ .

Estimation du taux de défaillance grâce à (T_1, \dots, T_N) , système non réparable :

On posera dans ce paragraphe $S_N = T_1 + \dots + T_N$. On a $T_1 \sim \mathcal{E}(\lambda) \Rightarrow \mathbb{E}(T_1) = \frac{1}{\lambda}$.

Les (T_i) étant iid, estimer λ peut se faire en considérant l'inverse de la moyenne empirique :

$$\hat{\lambda}_N = \frac{N}{T_1 + \dots + T_N} = \frac{N}{S_N} \quad (12)$$

Ce qui nous intéresse dans le cadre de la fiabilité, est de garantir que le taux de défaillance ne dépasse pas une certaine valeur. Donc on souhaite garantir que λ soit dans un certain intervalle du type $[0, \bar{\lambda}]$ avec un seuil de confiance α petit : $\mathbb{P}(\lambda \in [0, \bar{\lambda}]) = 1 - \alpha$. Pour un seuil de confiance α fixé, on cherche donc $\bar{\lambda}$ tel que

$$1 - \alpha = \mathbb{P}(\lambda \in [0, \bar{\lambda}]) = \mathbb{P}(\lambda < \bar{\lambda}) = \mathbb{P}(2\lambda S_N < 2\bar{\lambda} S_N)$$

Or on montre que dans le cadre du modèle HPP :

Proposition 4.1.

$$2\lambda S_N \sim \chi^2_{(2N)}$$

En effet, comme les (T_i) sont iid de loi $\mathcal{E}(\lambda) = \Gamma(1, \lambda)$, c'est que $S_N \sim \Gamma(N, \lambda)$, c'est-à-dire $2\lambda S_N \sim \Gamma(N, \frac{1}{2}) = \chi^2_{(2N)}$.

On a alors :

$$1 - \alpha = \mathbb{P}(2\lambda S_N < 2\bar{\lambda} S_N) = F_{\chi^2_{(2N)}}(2\bar{\lambda} S_N) \iff \bar{\lambda} = \frac{1}{2S_N} F_{\chi^2_{(2N)}}^{-1}(1 - \alpha)$$

Ceci entraîne qu'un intervalle de confiance unilatéral gauche de seuil α pour λ est :

$$[0, \frac{1}{2S_N} F_{\chi^2_{(2N)}}^{-1}(1 - \alpha)]$$

Nous avons graphiquement (Figure 12) mis en évidence la convergence de l'estimateur $\hat{\lambda}_N$ pour N de plus en plus grand (avec $\ln(N)$ en abscisse) vers $1/\mathbb{E}(T) = 1/m(0)$; nous noterons désormais $\lambda_m = 1/m(0)$. Nous avons également reporté la valeur théorique λ_∞ .

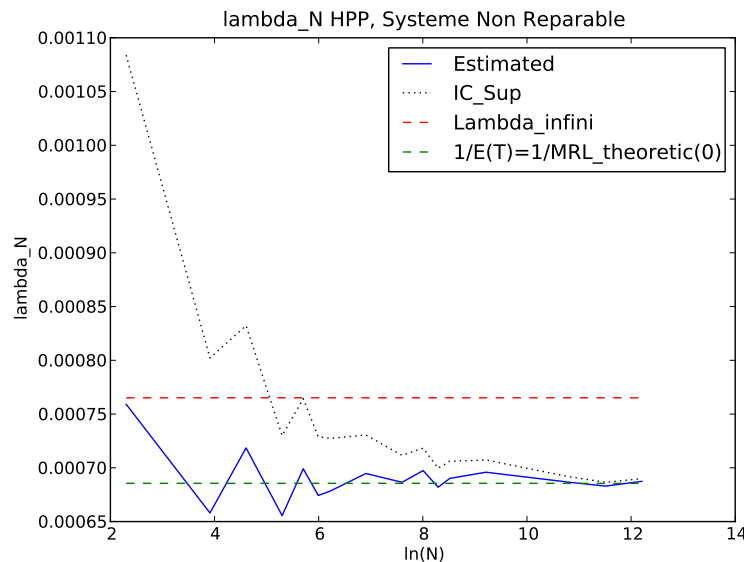


FIGURE 12 – Convergence de $N \mapsto \hat{\lambda}_N$ dans le cas HPP avec IC(95%) dans le cas d'un système non réparable. Nous remarquons que nous avons bien convergence de l'inverse de la moyenne empirique $\hat{\lambda}_N$ vers $\lambda_m = 1/\mathbb{E}(T) = 1/m(0)$, avec $m(0)$ donné par la proposition (3.2).

Remarque : Plus généralement on montre que ([5] proposition 7.4) :

Proposition 4.2. Des intervalles de confiance de seuil α de λ sont :

$$\left[0, \frac{1}{2S_N} F_{\chi_{(2N)}^2}^{-1}(1 - \alpha)\right], \left[\frac{1}{2S_N} F_{\chi_{(2N)}^2}^{-1}\left(\frac{\alpha}{2}\right), +\infty\right], \left[\frac{1}{2S_N} F_{\chi_{(2N)}^2}^{-1}\left(\frac{\alpha}{2}\right), \frac{1}{2S_N} F_{\chi_{(2N)}^2}^{-1}\left(1 - \frac{\alpha}{2}\right)\right]$$

Conséquence pour la fiabilité Maintenant que nous avons estimé le taux de défaillance dans le cadre du modèle HPP, il est *a priori* possible d'estimer la fiabilité en utilisant le fait que $R(t) = \exp(-\Lambda(t))$. En effet, on peut poser

$$\hat{R}_N(t) = \exp(-t\hat{\lambda}_N) \quad (13)$$

avec $\hat{\lambda}_N$ l'estimateur du taux de défaillance, constant dans le cadre HPP. Nous allons comparer cet estimateur avec celui étudié dans le paragraphe 7.2.1.4 de [5] :

Proposition 4.3. Soit $t \geq 0$. Dans le cadre du modèle HPP, un estimateur sans biais et de variance minimale de $R(t)$ est donné par :

$$\hat{R}_N(t) = \begin{cases} \left(1 - \frac{t}{S_N}\right)^{N-1} & , \text{ si } S_N > t \\ 0 & , \text{ sinon} \end{cases} \quad (14)$$

Un intervalle de confiance unilatéral droit au niveau α pour (13) se déduit de l'intervalle de confiance unilatéral gauche de λ :

$$\left[\exp\left(-\frac{t}{2S_N} F_{\chi_{(2N)}^2}^{-1}(1 - \alpha)\right), 1\right] \quad (15)$$

En tenant compte de l'intervalle de confiance unilatéral gauche pour λ , et du fait que

$$\left(1 - \frac{t}{S_N}\right)^{N-1} = \left(1 - \frac{t}{N} \hat{\lambda}_N\right)^{N-1}$$

on déduit l'intervalle de confiance *asymptotique* unilatéral droit au niveau α pour (14) :

$$\left[\left(1 - \frac{t}{2NS_N} F_{\chi_{(2N)}^2}^{-1}(1 - \alpha)\right)^{N-1}, 1\right] \quad (16)$$

Nous avons tracé (Figure 13) les courbes de la fiabilité théorique, celles estimées par (13) et (14), et les bornes inférieures des intervalles (15) et (16) ci-dessus (avec $\alpha = 0.05$).

Nous remarquons que les estimateurs (13) et (14) sont graphiquement "très proches", et qu'il y a un écart important entre la fiabilité théorique et les estimateurs considérés, même quand N est très grand.

Estimation paramétrique de l'erreur : Afin de quantifier rigoureusement l'erreur qu'on fait quand on remplace la fiabilité théorique par celle estimée, nous allons utiliser l'IMSE entre \hat{R}_N et R . Par exemple si $\hat{R}_N(t) = \exp(-\hat{\lambda}_N t)$, on aura :

$$d(\hat{R}_N, R) = \mathbb{E} \left[\int_0^\infty \left(\exp(-\hat{\lambda}_N t) - R(t) \right)^2 dt \right] \quad (17)$$

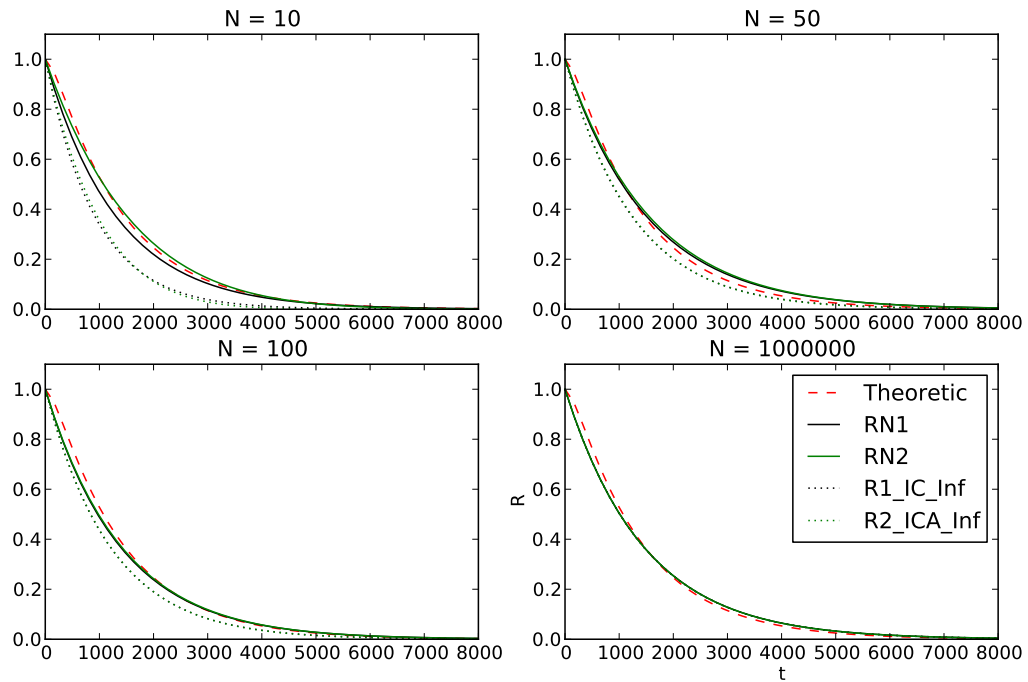


FIGURE 13 – Estimation paramétrique de la fiabilité dans le cas non réparable avec ICA(95%) pour différentes valeurs de N . Nous remarquons que les estimateurs (13) et (14), respectivement représentés par RN1 et RN2 sur la figure, sous-estiment la fiabilité théorique donnée par la proposition 3.2 pour les temps courts. Estimer la fiabilité du système pour les temps courts grâce à ces deux estimateurs permet d’avoir une borne inférieure raisonnable du point de vue de la sûreté de fonctionnement. Les estimateurs considérés sur-estiment la fiabilité pour les temps longs.

L’espérance précédente peut être estimée par Monte Carlo en considérant un K -échantillon $(\hat{\lambda}_N^i)_{i=1,\dots,K}$ de l’estimateur $\hat{\lambda}_N$ et en posant :

$$\hat{d}_K(\hat{R}_N, R) = \frac{1}{K} \sum_{i=1}^K \int_0^\infty \left(\exp(-\hat{\lambda}_N^i t) - R(t) \right)^2 dt$$

L’intégrale est calculée en utilisant une formule de quadrature. Les courbes d’erreurs sont reportées sur la Figure 14.

L’IMSE estimée décroît et converge vers l’*erreur de modèle* définie par :

$$\int_0^\infty (\exp(-\lambda_m t) - R(t))^2 dt \quad (18)$$

et qui vaut numériquement 2.7225141865.

Nous pouvons aussi nous intéresser à l’*erreur d’approximation* définie par :

$$\mathbb{E} \left[\int_0^\infty \left(\exp(-\hat{\lambda}_N t) - \exp(-\lambda_m t) \right)^2 dt \right] \quad (19)$$

qui tend vers zéro (Figure 15) de part le mode de convergence de $\hat{\lambda}_N$ vers λ_m .

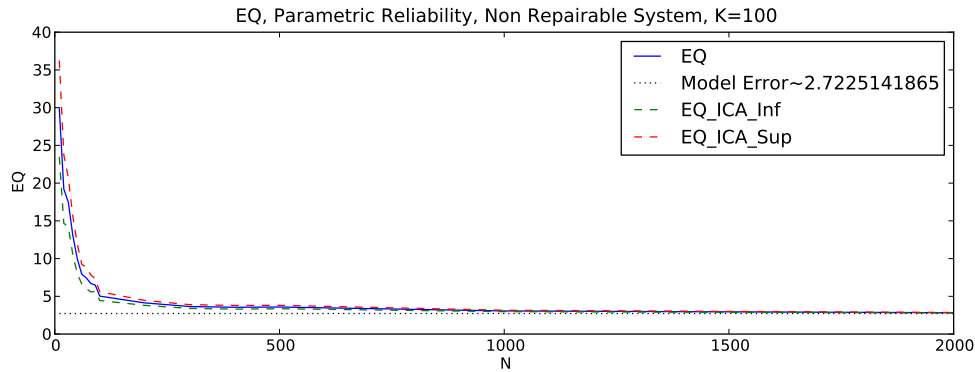


FIGURE 14 – Estimation paramétrique dans le cas du système non réparable de l’IMSE entre l’estimateur (13) de la fiabilité et la fiabilité théorique R donnée par la proposition 3.2. Nous remarquons que l’erreur tend vers l’erreur de modèle donnée par (18).

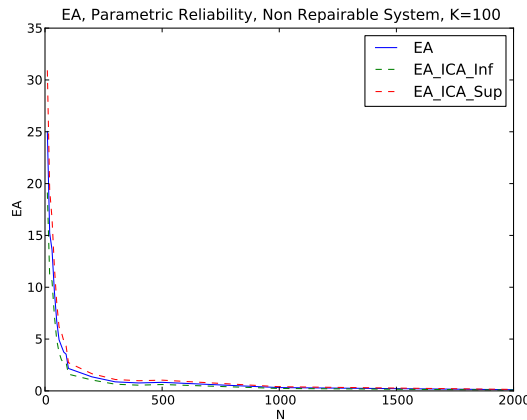


FIGURE 15 – Erreur d’approximation dans le cas de l’hypothèse d’un modèle HPP et d’un système non réparable. Nous mettons bien en évidence la convergence vers zéro de l’erreur d’approximation pour N de plus en plus grand.

Estimation du taux de défaillance grâce à (X_1, \dots, X_N) , système réparable :

En toute rigueur, l’estimation du taux de défaillance traitée dans le paragraphe précédent aurait dû être faite en considérant les durées de bon fonctionnement dans le cas réparable (X_1, \dots, X_N) , tout en négligeant les durées de réparation ([5] chapitre 7). C’est pourquoi nous effectuerons les simulations précédentes dans ce nouveau cas de figure. Nous allons montrer en particulier que le processus (X_1, \dots, X_N) n’est pas un HPP.

Nous posons $S_i = X_1 + \dots + X_i$ et imposerons par convention $S_0 = T_0 = 0$. L’estimateur du taux de défaillance dans le cas HPP est alors :

$$\hat{\lambda}_N = \frac{N}{X_1 + \dots + X_N} = \frac{N}{S_N} \quad (20)$$

Le § 9.2.3 de [3] montre que cet estimateur a une limite théorique que l’on peut calculer dans notre cas. Le processus markovien de sauts Z est irréductible car tous les états de \mathcal{E}

communiquent. En particulier Z admet une loi invariante que nous noterons ρ et qui vérifie $\rho A = 0 \Leftrightarrow A^T \rho^T = 0$. La matrice A n'étant pas inversible, nous allons remplacer la dernière équation du système à résoudre $A^T \rho^T = 0$ par $\sum_i \rho(i) = 1$. Notons pour cela

$$D = \begin{cases} A_{i,j}^T & \text{si } i < n \\ 1 & \text{si } i = n \end{cases}$$

On obtient par conséquent

$$\rho^T = D^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

Soit ρ_1 la restriction de ρ aux états de marche, ρ_2 sa restriction aux états de panne, et posons

$$\lambda_v = \frac{\rho_1 \mathbb{1}_{\mathcal{M}}}{\rho_2 A_{21} \mathbb{1}_{\mathcal{M}}} \quad (21)$$

avec A_{21} la sous-matrice inférieure gauche de

$$A = \begin{bmatrix} B & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

Nous avons alors le résultat de convergence suivant (proposition 9.17 de [3]) :

Proposition 4.4.

$$\hat{\lambda}_N \xrightarrow[N \rightarrow \infty]{} \lambda_v \quad (22)$$

λ_v est appelé *taux de Vesely*.

Nous avons mis en évidence (Figure 16) la convergence de l'estimateur $\hat{\lambda}_N$ vers λ_v pour N de plus en plus grand (avec $\ln(N)$ en abscisse), et avons reporté les valeurs théoriques λ_v , λ_∞ , et λ_m . Nous constatons que dans le cas réparable, la valeur limite λ_v de cet estimateur est différente de la limite λ_m de l'estimateur (12) du cas de système non réparable qui tient compte des instants de défaillance (T_1, \dots, T_N) . Ceci montre que (X_1, \dots, X_N) et (T_1, \dots, T_N) ne peuvent pas être égales en loi, et de fait, que le processus (X_i) n'est pas un HPP.

Conséquence pour la fiabilité : Nous obtenons (Figure 17) les courbes d'évolution des estimateurs paramétriques (13) et (14) de la fiabilité dans le cas d'un système réparable. Nous constatons un écart important entre la fiabilité théorique et les estimateurs paramétriques proposés. Nous avons quantifié la distance entre R et (13) au sens de (17) pour N de plus en plus grand. Nous avons reporté la valeur numérique de l'*erreur de modèle* défini dans ce cas par :

$$\int_0^\infty (\exp(-t \lambda_v) - R(t))^2 dt \quad (23)$$

Nous constatons que l'erreur quadratique dans le cas où les données utilisées sont les instants de premières pannes est plus grande que celle que nous avons simulé dans le cas où les données simulées sont les durées de bon fonctionnements. Cette erreur converge vers l'erreur

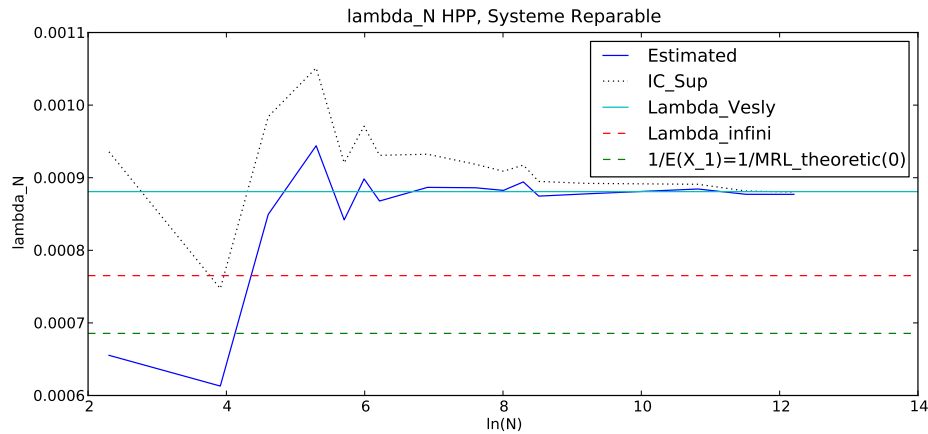


FIGURE 16 – Evolution de l'estimateur $\hat{\lambda}_N$ dans le cas d'un système réparable avec IC(95%). La valeur limite λ_v de cet estimateur est différente de la limite λ_m de l'estimateur (12) du cas de système non réparable qui tient compte des instants de défaillance (T_1, \dots, T_N) . Ceci montre que (X_1, \dots, X_N) et (T_1, \dots, T_N) ne peuvent pas être égales en loi, et de fait que le processus (X_i) n'est pas un HPP.

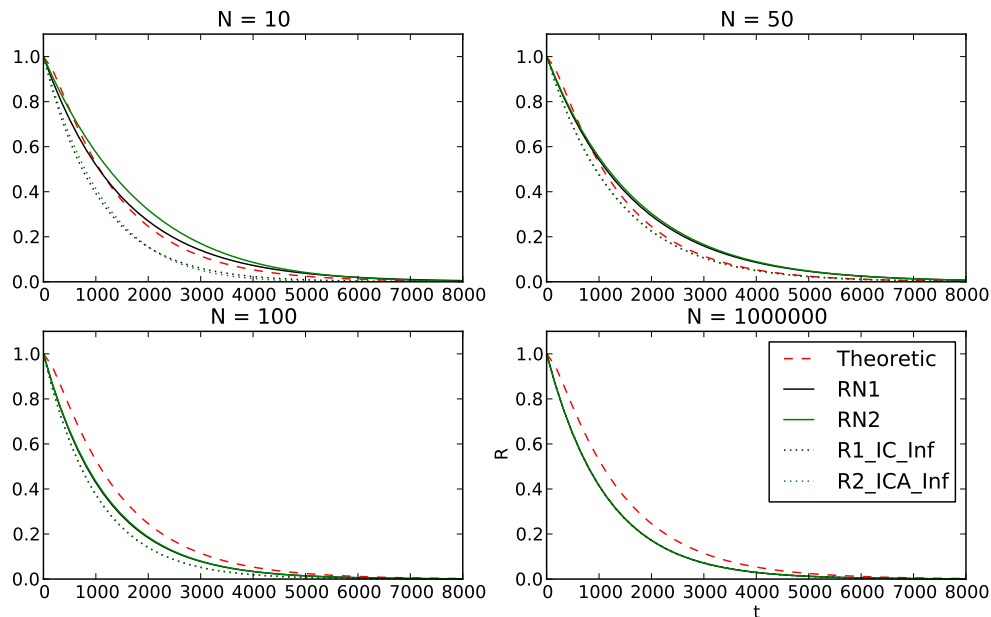


FIGURE 17 – Estimation paramétrique de la fiabilité dans le cas réparable avec IC(95%) dans le cas de l'estimateur (13) et ICA(95%) dans le cas de l'estimateur (14). Nous constatons que les estimateurs considérés sous estiment la fiabilité théorique R à partir de $N = 100$ et la sur-estiment pour N inférieur à 50 et les temps longs.

de modèle (23) dont la valeur numérique (~ 25.29) est plus importante que celle (~ 2.72) trouvée quand on a pris en compte les instants de défaillance (T_1, \dots, T_N) . Estimer le taux de

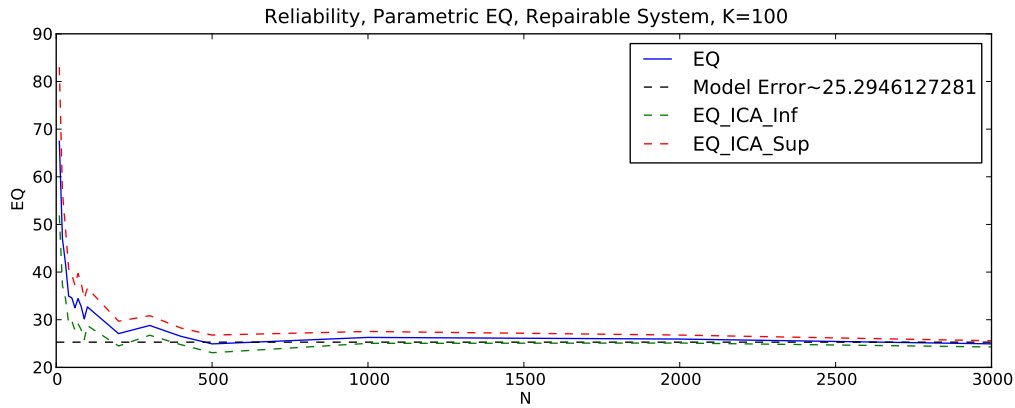


FIGURE 18 – Estimation de l’erreur quadratique entre l’estimateur (14) et la fiabilité théorique dans le cas du système réparable avec ICA(95%). Nous constatons que l’erreur quadratique dans le cas réparable est plus grande que celle que nous avons simulé dans le cas non réparable. Cette erreur converge vers l’erreur de modèle (23) dont la valeur numérique ~ 25.29 est plus importante que celle trouvée quand on a pris en compte les instants de défaillance (T_1, \dots, T_N) . Estimer le taux de défaillance grâce à (20) conduit donc à une moins bonne estimation de la fiabilité que lorsqu’on estime ce taux avec l’estimateur (12).

défaillance grâce à (20) conduit donc à une moins bonne estimation de la fiabilité que lorsqu’on estime ce taux en utilisant la relation (12).

Nous nous sommes également intéressés à l’*erreur d’approximation* définie par :

$$\mathbb{E} \left[\int_0^\infty \left(\exp(-\hat{\lambda}_N t) - \exp(-\lambda_v t) \right)^2 dt \right] \quad (24)$$

qui tend vers zéro de part le mode de convergence de $\hat{\lambda}_N$ vers λ_v . Nous avons mis en évidence cette convergence sur la Figure 19.

Erreur de modèle : Nous avons défini dans (18) et (23) deux erreurs qu’on a appelé erreurs de modèle, respectivement dans le cas non réparable et dans le cas réparable. Nous pouvons aussi nous demander s’il n’existe pas une constante λ_M qui minimise l’erreur qu’on fait quand on remplace la modélisation de notre système régie par un processus markovien de sauts, par un modèle pour lequel le taux de défaillance est constant et de valeur λ_M . Ceci revient finalement à résoudre le problème d’optimisation :

$$\text{Trouver } \lambda_M > 0, \text{ tel que, } \lambda_M = \operatorname{argmin} \left\{ \int_0^\infty (\exp(-\lambda t) - R(t))^2 dt ; \lambda > 0 \right\} \quad (25)$$

Considérons l’application J définie sur \mathbb{R}_+^* par :

$$J(\lambda) = \int_0^\infty (\exp(-\lambda t) - R(t))^2 dt, \lambda \in \mathbb{R}_+^*$$

Nous avons tracé (Figure 20) la courbe représentative de la fonction J . Afin de résoudre le problème de minimisation précédent, nous avons utilisé l’algorithme de Newton en cherchant

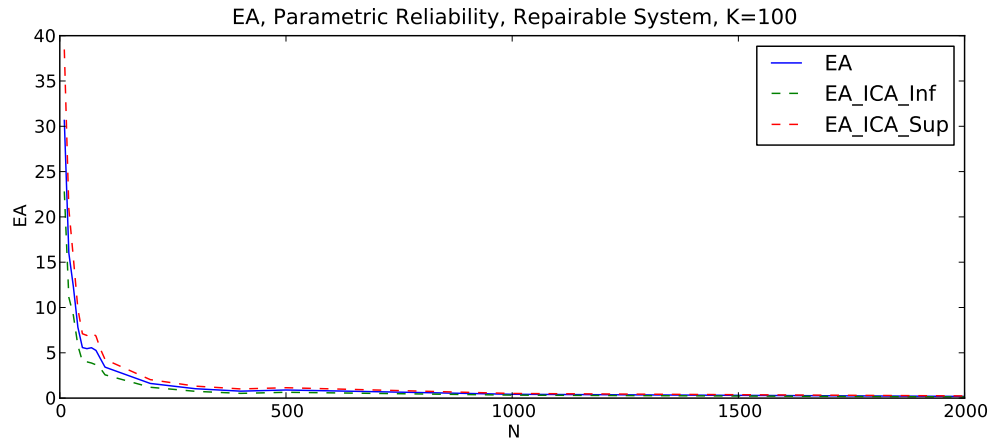


FIGURE 19 – Erreur d'approximation dans le cas de l'hypothèse d'un modèle HPP et d'un système réparable. Nous mettons bien en évidence la convergence vers zéro de l'erreur d'approximation pour N de plus en plus grand.

la valeur auquel J' s'annule. Nous avons obtenu au bout de 4 itérations la solution du problème de minimisation à 10^{-12} près, soit $\lambda_4 \sim 0.00068$, le minimum étant dans ce cas 2.7052.

Algorithme de Newton :

Afin d'implémenter l'algorithme de Newton nous avons dû calculer :

$$J'(\lambda) = \int_0^{\infty} 2t e^{-\lambda t} R(t) dt - \frac{1}{2\lambda^2}$$

puis

$$J''(\lambda) = - \int_0^{\infty} 2t^2 e^{-\lambda t} R(t) dt + \frac{1}{\lambda^3}$$

Reste alors à considérer la suite (λ_n) définie par :

$$\begin{cases} \lambda_0 & \text{choisie} \\ \lambda_{n+1} & = \lambda_n - \frac{J'(\lambda_n)}{J''(\lambda_n)} \end{cases} \quad (26)$$

Remarquons que le choix de λ_0 est important dans la mesure où $J'(\lambda) \xrightarrow{\lambda \rightarrow \infty} 0$.

Validation Croisée : L'inconvénient avec le calcul de l'erreur de modèle est bien sûr le fait qu'on ne connaît pas l'expression de la fiabilité d'un système donné. Une façon de contourner ce problème est de passer par une estimation de l'erreur qui consiste à faire intervenir l'estimateur non paramétrique (6) de la fiabilité. Cette méthode que nous allons décrire est appelé "validation croisée". Soit (T_1, \dots, T_N) , les instants de panne de N systèmes identiques et K un diviseur entier de N qu'on fera varier. Pour chaque valeur de K , on trie aléatoirement l'échantillon (T_1, \dots, T_N) qu'on divise en K paquet (T_1^k, \dots, T_M^k) , $k \in \{1, \dots, K\}$ avec M l'entier valant N/K . On calcule dès lors :

$$d_{N,K}^k = \int_0^{\infty} \left(\hat{R}_N^k(t) - \exp(-t\hat{\lambda}_N^k) \right)^2$$

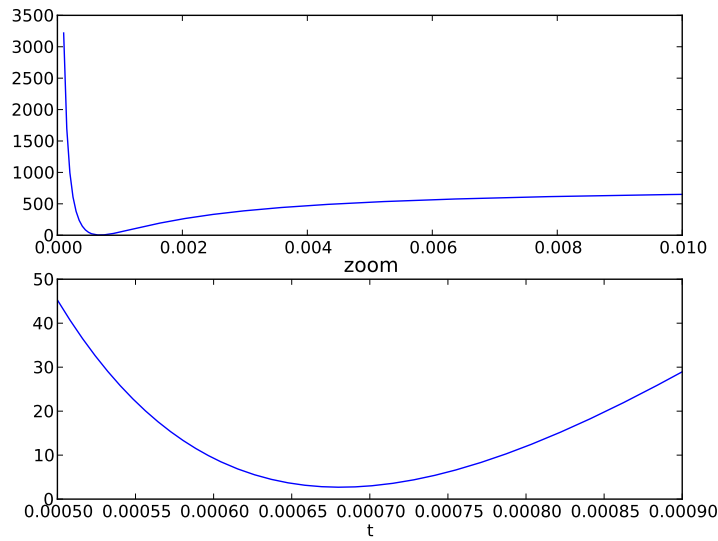


FIGURE 20 – Mise en évidence du minimum dans le problème (25)

avec $\hat{R}_N^k(t)$ et $\hat{\lambda}_N^k$ obtenus grâce à (T_1^k, \dots, T_M^k) . On pose alors :

$$d_{N,K} = \frac{1}{K} \sum_{k=1}^K d_{N,K}^k$$

Si on considère le cas $N = 5000$, les simulations permettent alors de dresser le tableau :

K	500	250	100	50	125	10	5
$d_{N,K}$	56.20	32.33	14.13	8.78	5.71	3.86	3.46

ε représente à chaque fois l'erreur que *python* fait dans le calcul des intégrales précédentes.

La théorie montre que pour N grand et K de l'ordre de 5, $d_{N,K}$ peut être considéré comme une bonne approximation de l'erreur quadratique (17). Nous remarquons en effet que l'erreur calculée dans ce cas est de l'ordre de 3.46 pas très loin de l'erreur que nous avons trouvé par un calcul direct en passant par (17). L'avantage étant ici qu'on a une estimation de l'erreur de modèle sans la connaissance de l'expression analytique de R .

4.3.2 Modèle NHPP

Ce cas de figure est le plus courant puisqu'un matériel sollicité s'use et que son taux de défaillance est amené à varier. Ici on suppose donc que ce taux est une fonction du temps. Modéliser le système consiste à choisir une forme particulière de cette fonction. Plusieurs formes ont dans ce cadre été proposées, nous présentons par la suite le modèle du à Duane pour lequel le taux de défaillance est une fonction puissance du temps, appelé processus de puissance (PLP, *Power-Low-Process*) :

$$\lambda(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta-1} \quad (27)$$

avec β et η des paramètres réels positifs qu'on sera amené à estimer.

Nous remarquons que :

- si $\beta > 1$, le système aura un taux de défaillance croissant, et de fait il y a aura usure et donc décroissance de la fiabilité
- Pour $\beta < 1$, le taux est décroissant, ce qui est synonyme d'amélioration du système et donc de croissance de la fiabilité
- pour $\beta = 1$, le taux de défaillance est constant, la fiabilité est stabilisée, et l'on retombe dans le cas du modèle HPP

Le paramètre β peut donc être interprété comme le degré d'amélioration ou de dégradation du système, d'où l'importance de pouvoir l'estimer ([5] proposition 9.8) :

Proposition 4.5. Soit $(X_j)_{j=1,\dots,N}$ les durées de bon fonctionnement du système dont on a négligé les durées de réparation, et posons $S_i = \sum_{j=1}^i X_j$. Dans le cadre NHPP, les estimateurs de maximum de vraisemblance des paramètres du PLP sont :

$$\hat{\beta}_N = \frac{N}{\sum_{i=1}^{N-1} \ln(\frac{S_N}{S_i})} \quad \text{et} \quad \hat{\eta}_N = \frac{S_N}{N^{1/\hat{\beta}_N}}$$

Nous résumons les propriétés de convergence en loi associées à ces deux estimateurs dans la proposition suivante (proposition 3.22 et 3.23 de [3]) :

Proposition 4.6. Avec les notations précédentes,

$$\begin{aligned} i) \quad & \sqrt{N} \left(\frac{\beta}{\hat{\beta}_N} - 1 \right) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathbb{N}(0, 1) \\ ii) \quad & \frac{\hat{\beta}_N \sqrt{N}}{\ln(N)} \ln \frac{\hat{\eta}_N}{\eta} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathbb{N}(0, 1) \end{aligned}$$

Cette proposition permet de construire des intervalles de confiance asymptotiques pour η et β . Si $z_\alpha = F_{\mathbb{N}(0,1)}^{-1} \left(1 - \frac{\alpha}{2} \right)$ alors on peut écrire :

$$\mathbb{P} \left(\beta \in \hat{\beta}_N \left[1 \pm \frac{z_\alpha}{\sqrt{N}} \right] \right) \quad \text{et} \quad \mathbb{P} \left(\eta \in \hat{\eta}_N \left[\exp\left(\pm \frac{z_\alpha \ln(N)}{\hat{\beta}_N \sqrt{N}} \right) \right] \right) \xrightarrow[N \rightarrow \infty]{} 1 - \alpha$$

Nous avons mis en évidence la convergence des estimateurs $\hat{\beta}_N$ et $\hat{\eta}_N$ et tracé (Figure 21) les intervalles de confiance asymptotiques bilatéraux à 95%.

En fait on montre que l'on peut trouver un intervalle de confiance non asymptotique pour β grâce à ([5] proposition 9.9) :

Proposition 4.7. La variable aléatoire $2N \frac{\beta}{\hat{\beta}_N}$ est indépendante de T_N et a pour loi $\chi_{(2(N-1))}^2$.

Ceci entraîne que

$$\mathbb{P} \left(\beta \in \frac{\hat{\beta}_N}{2N} \left[F_{\chi_{(2(N-1))}^2}^{-1} \left(\frac{\alpha}{2} \right), F_{\chi_{(2(N-1))}^2}^{-1} \left(1 - \frac{\alpha}{2} \right) \right] \right) = 1 - \alpha \quad (28)$$

et fournit un intervalle de confiance bilatéral au niveau α pour β . Ceci permet de mettre en évidence graphiquement (Figure 21) la convergence de $\hat{\beta}$ vers β avec cette fois un *intervalle de confiance* à 95%.

Nous remarquons que $\hat{\beta}_N$ converge vers 1, c'est-à-dire que si on se place dans le cadre d'un modèle PLP, l'observation du système nous ramène au modèle HPP. La dualité entre intervalles de confiance et tests d'hypothèse fait que l'on peut facilement construire des tests portants sur la valeur de β . En particulier, comme " $\beta = 1$ " correspond au HPP, on peut tester l'hypothèse qu'un PLP est en fait un HPP en testant $H_0 : \beta = 1$ contre $H_1 : \beta \neq 1$. Nous développerons ce test dans la section suivante.

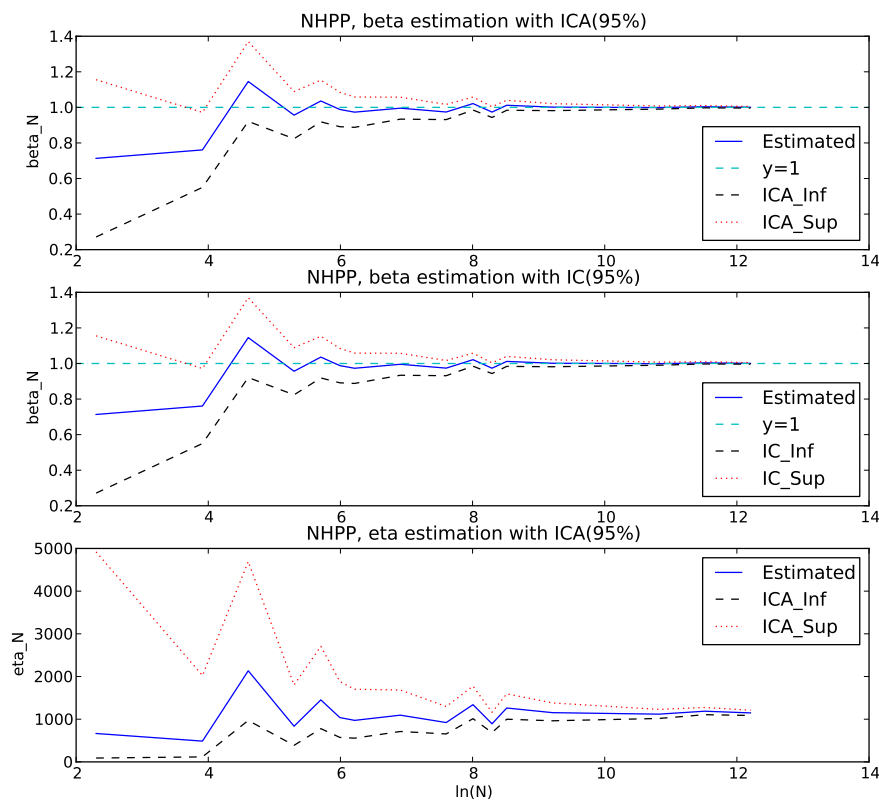


FIGURE 21 – Estimations de β et η dans le modèle PLP (27). Dans le premier graphique nous avons mis en évidence ce vers quoi converge l'estimateur $\hat{\beta}$ avec un aperçu de l'intervalle de confiance asymptotique à 95%. Dans le second graphique nous avons tracé l'intervalle de confiance exact à 95% issu du résultat (28). Le dernier graphique met en évidence le comportement de l'estimateur $\hat{\eta}$. Ceci montre que si on se place dans le cadre d'un modèle PLP, l'observation du système nous ramène au modèle HPP.

4.4 Tests statistiques et validation de modèle

Nous allons maintenant passer aux tests statistiques permettant de valider la modélisation.

4.4.1 Test de Comparaison de Kolmogorov-Smirnof

L'un des objectifs du stage est de trouver des outils permettant de faire des choix de modèles, c'est-à-dire pouvoir affirmer avec un certain seuil de confiance que le système étudié peut être modélisé par un modèle pour lequel la loi du processus de défaillance est connue. Nous avons étudié le cas du modèle HPP ainsi que le modèle NHPP. Par exemple, peut-on modéliser notre système par un processus de défaillance de loi $\mathcal{E}(\lambda)$, ce qui correspond au modèle HPP. Comme nous l'avons indiqué au § 4.3, (T_1, \dots, T_N) et (X_1, \dots, X_N) sont égaux en loi dans le modèle HPP. On rejettera donc l'hypothèse " (X_1, \dots, X_N) est un HPP" au seuil de confiance α , si on peut affirmer que les réalisations de (T_1, \dots, T_N) et (X_1, \dots, X_N) ne sont pas identiquement distribuées. On peut donc tester l'hypothèse nulle $H_0 : "(T_1, \dots, T_N) \text{ et } (X_1, \dots, X_N) \text{ ont même loi}"$, contre l'hypothèse alternative $H_1 : "(T_1, \dots, T_N) \text{ et } (X_1, \dots, X_N) \text{ ne sont pas de même loi}"$. Nous avons utilisées le test de comparaison de Kolmogorov-Smirnof (KS) pour trancher cette question. Nous avons calculé la statistique de KS, ainsi que la p-value (résultat fournit par la commande *ks_2samp* de la librairie python *scipy.stats*). Voici les résultats pour différentes valeurs de N :

N	20	100	300	500
KS	0.4499	0.23	0.21	0.20
p-value	0.02321	0.0082	2.71e-6	2e-9

Ce test est construit de telle façon à ne pas rejeter l'hypothèse nulle dans le cas où la statistique de KS est petite *ou* la p-value est grande. Pour toutes les valeurs de N considérées, on est conduit à rejeter l'hypothèse nulle $H_0 : "(T_1, \dots, T_N) \text{ et } (X_1, \dots, X_N) \text{ sont égaux en loi}"$. Cela confirme le fait que le système markovien de sauts étudié, ne peut être modélisé par un HPP.

4.4.2 Test de Comparaison entre IC

Remarquons que :

$$\text{HPP} \Rightarrow H_0 : "(T_1, \dots, T_N) \text{ et } (X_1, \dots, X_N) \text{ sont égaux en loi}" \Rightarrow \lambda_m = \lambda_v.$$

Soit $S_{N_T}^T = T_1 + \dots + T_{N_T}$ et $S_{N_X}^X = X_1 + \dots + X_{N_X}$ (on considérera N_T instants de pannes, et N_X durées de bon fonctionnement avec N_T et N_X non nécessairement égaux) et posons :

$$\lambda_T = \lim_{N \rightarrow \infty} \frac{N}{T_1 + \dots + T_N} \text{ et } \lambda_X = \lim_{N \rightarrow \infty} \frac{N}{X_1 + \dots + X_N}$$

On pourra dès lors considérer la zone de rejet :

$$W = \{IC(\lambda_T) \cap IC(\lambda_X) = \emptyset\}$$

avec d'après la proposition 4.2 $IC(\lambda_T) = [\frac{1}{2S_{N_T}^T} F_{\chi_{(2N_T)}^2}^{-1}(\frac{\alpha}{2}), \frac{1}{2S_{N_T}^T} F_{\chi_{(2N_T)}^2}^{-1}(1 - \frac{\alpha}{2})]$ et de même pour $IC(\lambda_X)$.

Proposition 4.8. Soit $G_{(n,p)}$ la fonction de répartition de la loi de Fisher $\mathcal{F}(n, p)$. Alors le niveau de confiance du test de comparaison entre intervalles de confiance s'écrit :

$$\mathbb{P}_{H_0}(W) = 2 - G_{(N_X, N_T)} \left(\frac{N_X F_{\chi^2(2N_T)}^{-1} \left(1 - \frac{\alpha}{2}\right)}{N_T F_{\chi^2(2N_X)}^{-1} \left(\frac{\alpha}{2}\right)} \right) - G_{(N_T, N_X)} \left(\frac{N_T F_{\chi^2(2N_X)}^{-1} \left(1 - \frac{\alpha}{2}\right)}{N_X F_{\chi^2(2N_T)}^{-1} \left(\frac{\alpha}{2}\right)} \right)$$

En effet, nous remarquons que :

$$W = \left\{ \frac{1}{2S_{N_T}^T} F_{\chi^2(2N_T)}^{-1} \left(\frac{\alpha}{2}\right) > \frac{1}{2S_{N_X}^X} F_{\chi^2(2N_X)}^{-1} \left(1 - \frac{\alpha}{2}\right) \right\} \cup \left\{ \frac{1}{2S_{N_X}^X} F_{\chi^2(2N_X)}^{-1} \left(\frac{\alpha}{2}\right) > \frac{1}{2S_{N_T}^T} F_{\chi^2(2N_T)}^{-1} \left(1 - \frac{\alpha}{2}\right) \right\}$$

L'union ensembliste précédente étant disjointe, nous pouvons alors écrire :

$$\begin{aligned} \mathbb{P}_{H_0}(W) &= \mathbb{P}_{H_0} \left\{ \frac{1}{2S_{N_T}^T} F_{\chi^2(2N_T)}^{-1} \left(\frac{\alpha}{2}\right) > \frac{1}{2S_{N_X}^X} F_{\chi^2(2N_X)}^{-1} \left(1 - \frac{\alpha}{2}\right) \right\} \\ &+ \mathbb{P}_{H_0} \left\{ \frac{1}{2S_{N_X}^X} F_{\chi^2(2N_X)}^{-1} \left(\frac{\alpha}{2}\right) > \frac{1}{2S_{N_T}^T} F_{\chi^2(2N_T)}^{-1} \left(1 - \frac{\alpha}{2}\right) \right\} \end{aligned}$$

Par exemple on a :

$$\mathbb{P}_{H_0} \left\{ \frac{1}{2S_{N_T}^T} F_{\chi^2(2N_T)}^{-1} \left(\frac{\alpha}{2}\right) > \frac{1}{2S_{N_X}^X} F_{\chi^2(2N_X)}^{-1} \left(1 - \frac{\alpha}{2}\right) \right\} = \mathbb{P}_{H_0} \left(\frac{2\lambda S_{N_X}^X}{2\lambda S_{N_T}^T} > \frac{F_{\chi^2(2N_X)}^{-1} \left(1 - \frac{\alpha}{2}\right)}{F_{\chi^2(2N_T)}^{-1} \left(\frac{\alpha}{2}\right)} \right)$$

Or sous H_0 , $2\lambda S_{N_X}^X \sim \chi^2_{(2N_X)}$, de même $2\lambda S_{N_T}^T \sim \chi^2_{(2N_T)}$ et on a de plus d'après la proposition 4.1 :

$$\frac{2N_T \lambda S_{N_X}^X}{2N_X \lambda S_{N_T}^T} \sim \mathcal{F}(N_X, N_T), \text{ loi de fisher à } N_X \text{ et } N_T \text{ degrés de liberté}$$

Si on note $G_{(n,p)}$ la fonction de répartition de la loi $\mathcal{F}(n, p)$, cela conduit à :

$$\mathbb{P}_{H_0}(W) = 2 - G_{(N_X, N_T)} \left(\frac{N_X F_{\chi^2(2N_T)}^{-1} \left(1 - \frac{\alpha}{2}\right)}{N_T F_{\chi^2(2N_X)}^{-1} \left(\frac{\alpha}{2}\right)} \right) - G_{(N_T, N_X)} \left(\frac{N_T F_{\chi^2(2N_X)}^{-1} \left(1 - \frac{\alpha}{2}\right)}{N_X F_{\chi^2(2N_T)}^{-1} \left(\frac{\alpha}{2}\right)} \right)$$

Nous avons tracé (Figure 22) l'évolution du niveau de confiance ci-dessus pour $\alpha = 0.01$ et $\alpha = 0.05$. Nous constatons que le niveau tend vers α quand N est de plus grand. On conclut alors qu'on est amené à rejeter l'hypothèse H_0 au seuil α .

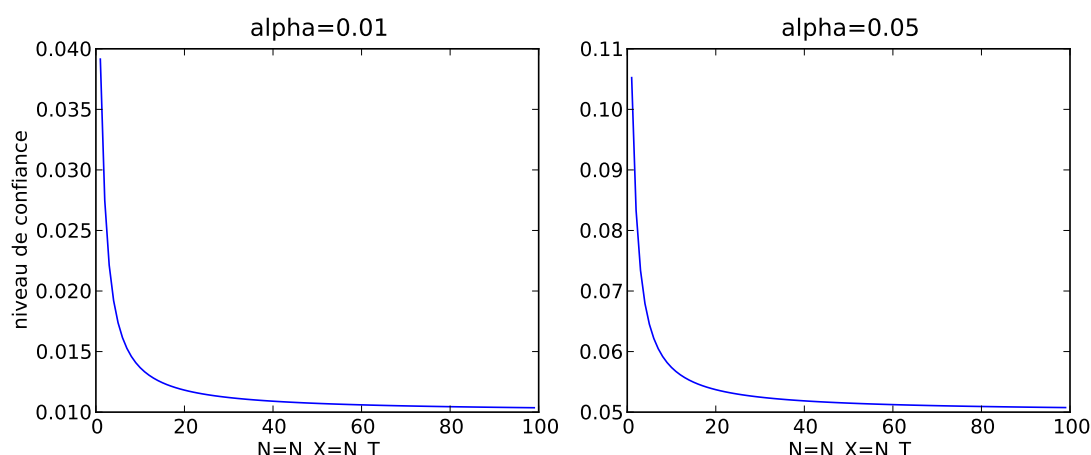


FIGURE 22 – Niveau de confiance asymptotiques du test de comparaison entre intervalles de confiance suggéré.

4.4.3 Test d'adéquation au HPP :

L'estimation du taux de défaillance a été réalisée en faisant l'hypothèse d'un modèle HPP. Néanmoins, peut-on au vu des observations admettre l'adéquation des données observées avec celle qui proviendrait d'un vrai modèle HPP pour lequel les durées de bon fonctionnement (X_1, \dots, X_N) sont iid de loi $\mathcal{E}(\lambda)$ pour un seuil de confiance fixé ?

Répondre à cette question revient à faire un test d'adéquation de niveau α de :

L'hypothèse nulle H_0 : (X_1, \dots, X_N) sont iid de loi $\mathcal{E}(\lambda)$

contre l'hypothèse alternative H_1 : (X_1, \dots, X_N) ne sont pas iid de loi $\mathcal{E}(\lambda)$

Le test basé sur la statistique d'Anderson-Darling (AD) permet de répondre à cette question. Nous avons utilisé pour cela la fonction *anderson* de la librairie python *scipy.stats*. Nous résumons les résultats obtenus pour différentes valeurs de N avec $\alpha = 0.05$:

N	20	100	300	500
AD	0.8273	1.4763	7.7141	12.7707
valeur critique	0.02321	1.333	1.338	1.339

Ce test est construit de telle façon à rejeter l'hypothèse nulle H_0 : " (X_1, \dots, X_N) sont iid de loi $\mathcal{E}(\lambda)$ ", dans le cas où la statistique de AD dépasse la valeur critique comme indiquée dans le tableau, ce qui est le cas pour toutes les valeurs de N considérées. Cela confirme le fait que le système markovien de sauts étudié, ne peut être modélisé par un HPP.

4.4.4 Test d'hypothèse sur λ dans le modèle HPP

Dans la sous-section 4.3.1, on a vu qu'on pouvait traduire des objectifs de fiabilité du système en exigeant que son taux de défaillance soit plus petit qu'une certaine quantité $\bar{\lambda}$ avec un seuil de confiance α . On peut d'un autre côté être amené à déterminer si les objectifs de fiabilité sont atteints ou pas, en gardant à l'esprit qu'il faut éviter de conclure à tort qu'ils le sont. Dans

ce paragraphe, le taux de défaillance est constant (HPP). Ceci conduit à procéder au test de seuil α :

de l'hypothèse nulle $H_0 : \lambda \geq \lambda_0$ contre l'hypothèse alternative $H_1 : \lambda < \lambda_0$

Afin de déterminer la région critique associée à ce test, on considère qu'il est raisonnable de conclure que $\lambda < \lambda_0$ si $\hat{\lambda}_N$ est nettement plus petit que λ_0 . Ceci conduit à poser comme région critique :

$$W = \{\hat{\lambda}_N < k_\alpha\}$$

k_α étant donné par :

$$\alpha = \sup_{\lambda \geq \lambda_0} \mathbb{P}(W) = \sup_{\lambda \geq \lambda_0} \mathbb{P}(\hat{\lambda}_N < k_\alpha) = \sup_{\lambda \geq \lambda_0} \mathbb{P}\left(S_N > \frac{N}{k_\alpha}\right) = \sup_{\lambda \geq \lambda_0} \mathbb{P}\left(2\lambda S_N > 2\lambda \frac{N}{k_\alpha}\right)$$

Ce qui entraîne d'après la proposition (4.1) puis la croissance de $F_{\chi^2_{(2N)}}$, que k_α doit vérifier :

$$\alpha = \sup_{\lambda \geq \lambda_0} \left[1 - F_{\chi^2_{(2N)}}\left(\frac{2\lambda N}{k_\alpha}\right)\right] = 1 - F_{\chi^2_{(2N)}}\left(\frac{2\lambda_0 N}{k_\alpha}\right)$$

Soit :

$$k_\alpha = \frac{2\lambda_0 N}{F_{\chi^2_{(2N)}}^{-1}(1 - \alpha)}$$

Conclusion :

$$W = \left\{S_N > \frac{F_{\chi^2_{(2N)}}^{-1}(1 - \alpha)}{2\lambda_0}\right\}$$

Simulation : En prenant $\lambda_0 = \lambda_\infty$, $N = 1000$ et $\alpha = 0.05$, les simulations fournissent :

$$S_N = 1128917.71296 \quad \text{et} \quad \frac{F_{\chi^2_{(2N)}}^{-1}(1 - \alpha)}{2\lambda_0} = 1375693.52636$$

c'est-à-dire,

$$S_N < \frac{F_{\chi^2_{(2N)}}^{-1}(1 - \alpha)}{2\lambda_0}$$

Donc on ne rejette pas l'hypothèse nulle $\lambda \geq \lambda_\infty$ contre l'hypothèse alternative $\lambda < \lambda_\infty$ au seuil $\alpha = 0.05$.

Bien entendu ce test a supposé un taux de défaillance λ constant, ce qui n'est pas le cas du système considéré, d'où le caractère surprenant du résultat obtenu car nous savons qu'à tout instant t on a $\lambda(t) < \lambda_\infty$.

Par ailleurs le fait d'avoir essayé $\lambda_0 = \lambda_\infty$ repose sur la connaissance du modèle markovien de sauts fournissant la valeur de λ_∞ qu'on ne peut pas estimer contrairement aux quantités λ_v et λ_m . Comme $\lambda_\infty > \lambda_m$ on aboutit à la même conclusion de non rejet en considérant $\lambda_0 = \lambda_m$.

En prenant $\lambda_0 = \lambda_v$, $N = 1000$ et $\alpha = 0.05$, les simulations fournissent avec les mêmes données :

$$S_N = 1128917.71296 \quad \text{et} \quad \frac{F_{\chi^2_{(2N)}}^{-1}(1 - \alpha)}{2\lambda_v} = 1194971.98314$$

c'est-à-dire toujours,

$$S_N < \frac{F_{\chi^2(2N)}^{-1}(1 - \alpha)}{2\lambda_v}$$

Donc on ne rejette pas l'hypothèse nulle $\lambda \geq \lambda_v$ contre l'hypothèse alternative $\lambda < \lambda_v$ au seuil $\alpha = 0.05$.

4.4.5 Test de tendance

La théorie montre qu'étant donné λ_0 , on ne peut pas trouver un test UPP de l'hypothèse nulle $H_0 : \lambda = \lambda_0$ contre l'hypothèse alternative $\lambda \neq \lambda_0$, qui permettrait de trancher avec un certain seuil de confiance que le taux de défaillance est égal à λ_0 . Nous allons donc nous ramener à un test de tendance.

Nous allons présenter deux tests de tendance : le *test de Laplace* qui est asymptotique, et le *test logarithmique* qui est exact.

Test de Laplace

Le modèle HPP est celui pour lequel le taux de défaillance est constant, ce qui correspond à l'absence de tendance. L'hypothèse nulle H_0 : "Il y a absence de tendance" est équivalente à l'hypothèse "le processus des durées de bon fonctionnement (X_1, \dots, X_N) est un HPP".

Il nous faut aussi traduire statistiquement l'existence de tendance, c'est-à-dire la croissance ou la décroissance de la fiabilité :

1. On peut raisonnablement affirmer qu'il y a *croissance de la fiabilité* dans le cas où les X_i sont de plus en plus grand, c'est-à-dire que le système met de plus en plus de temps avant de tomber en panne après réparation. Donc il y a croissance de la fiabilité dans le cas où les $(S_i)_{i=1, \dots, N-1}$ sont plus petits comparés à S_N , voir que $\sum_{i=1}^{N-1} S_i$ est "plutôt petit" ([5] page 270).
2. *A contrario*, il y a *décroissance de la fiabilité* quand les X_i sont de plus en plus petits, c'est-à-dire que les $(S_i)_{i=1, \dots, N-1}$ et *a fortiori* $\sum_{i=1}^{N-1} S_i$ sont plutôt grands.
3. L'existence de tendance peut donc se traduire par

$$\sum_{i=1}^{N-1} S_i \text{ est "soit trop petit soit trop grand".}$$

Ceci conduit à considérer les trois hypothèses alternatives suivantes :

- H_1 : $\sum_{i=1}^{N-1} S_i$ est "soit trop petit soit trop grand" ((X_i) n'est pas un HPP)
- H_2 : $\sum_{i=1}^{N-1} S_i$ est "petit" (*croissance de la fiabilité*)
- H_3 : $\sum_{i=1}^{N-1} S_i$ est "grand" (*décroissance de la fiabilité*)

Nous avons besoin de construire les régions critiques correspondants aux tests de seuil α de l'hypothèse nulle H_0 , contre respectivement ces trois hypothèses alternatives. Pour cela nous aurons besoin du résultat de la proposition suivante (G[5], § 7.1.6) :

Proposition 4.9. *Sous H_0 , le vecteur aléatoire $(\frac{S_1}{S_N}, \dots, \frac{S_{N-1}}{S_N})$ est égale en loi au vecteur $(U_{(1)}, \dots, U_{(N-1)})$, avec U_i iid de loi $\mathcal{U}(0, 1)$.*

En utilisant les remarques précédentes et en tenant compte de cette proposition, l'idée est d'exprimer les régions critiques à l'aide de la *statistique de Laplace* :

$$V_N = \frac{\sum_{i=1}^{N-1} \frac{S_i}{S_N} - (N-1)\frac{1}{2}}{\sqrt{(N-1)\frac{1}{12}}} = \sqrt{\frac{12}{(N-1)S_N^2}} \left(\sum_{i=1}^{N-1} S_i - (N-1)\frac{S_N}{2} \right) \quad (29)$$

qui converge en loi vers une variable aléatoire de loi $N(0, 1)$.

On peut dès lors construire les régions critiques de seuil α de l'hypothèse nulle H_0 contre l'hypothèse alternative :

- H_1 : " (X_i) n'est pas un HPP"
en posant

$$W = \{V_N < k_{1,\alpha}\} \cup \{V_N > k_{2,\alpha}\}$$

avec les $k_{i,\alpha}$ vérifiant

$$\alpha = \mathbb{P}_{H_0}(W) = \mathbb{P}_{H_0}(V_N < k_{1,\alpha}) + \mathbb{P}_{H_0}(V_N > k_{2,\alpha}) \xrightarrow{N \rightarrow \infty} F_{N(0,1)}(k_{1,\alpha}) + 1 - F_{N(0,1)}(k_{2,\alpha}) \quad (30)$$

Il est usuel de choisir les $k_{i,\alpha}$ en imposant de plus :

$$F_{N(0,1)}(k_{1,\alpha}) = \frac{\alpha}{2} \quad \text{et} \quad F_{N(0,1)}(k_{2,\alpha}) = 1 - \frac{\alpha}{2}$$

c'est-à-dire $k_{1,\alpha} = F_{N(0,1)}^{-1}\left(\frac{\alpha}{2}\right)$ et $k_{2,\alpha} = F_{N(0,1)}^{-1}\left(1 - \frac{\alpha}{2}\right) = -F_{N(0,1)}^{-1}\left(\frac{\alpha}{2}\right)$, d'où :

$$W = \left\{ |V_N| > F_{N(0,1)}^{-1}\left(1 - \frac{\alpha}{2}\right) \right\} \quad (31)$$

- H_2 : Il y a *croissance de la fiabilité*
 $\sum_{i=1}^{N-1} S_i$ est "trop petit" si et seulement V_N est "trop petit". Donc le test H_0 contre H_2 admettra comme région critique

$$W = \{V_N < k_\alpha\} \quad (32)$$

avec k_α donné par :

$$\alpha = \mathbb{P}_{H_0}(W) = \mathbb{P}_{H_0}(V_N < k_\alpha) \quad (33)$$

D'après ce qui précède, cette quantité tend vers $F_{N(0,1)}(k_\alpha)$

- H_3 : Il y a *décroissance de la fiabilité*
 $\sum_{i=1}^{N-1} S_i$ est "trop grand" si et seulement V_N est "trop grand". Donc le test H_0 contre H_3 admettra

$$W = \{V_N > l_\alpha\} \quad (34)$$

comme région critique, avec l_α donné par :

$$\alpha = \mathbb{P}_{H_0}(W) = \mathbb{P}_{H_0}(V_N > l_\alpha) \quad (35)$$

qui tend vers $1 - F_{N(0,1)}(l_\alpha)$

Test logarithmique

Pour le *test logarithmique*, on utilise le résultat de la proposition suivante :

Proposition 4.10. Sous H_0 ,

$$K_N = -2 \sum_{i=1}^{N-1} \ln \left(\frac{S_i}{S_N} \right) \sim \chi^2_{(2(N-1))}$$

Or sous H_0 , les S_i sont petits, c'est-à-dire que K_N est grand. Le même raisonnement que dans la construction du test de Laplace conduit au test logarithmique de seuil α , de la même hypothèse nulle H_0 , contre les hypothèses alternatives :

a) H_1 : "le processus (X_i) n'est pas un HPP",

$$W = \left\{ K_N < F_{\chi^2_{(2(N-1))}}^{-1} \left(\frac{\alpha}{2} \right) = k_{1,\alpha} \right\} \cup \left\{ K_N > F_{\chi^2_{(2(N-1))}}^{-1} \left(1 - \frac{\alpha}{2} \right) = k_{2,\alpha} \right\} \quad (36)$$

b) H_2 : "Il y a croissance de la fiabilité",

$$W = \left\{ K_N > F_{\chi^2_{(2(N-1))}}^{-1} (1 - \alpha) = l_\alpha \right\} \quad (37)$$

c) H_3 : "il y a décroissance de la fiabilité",

$$W = \left\{ K_N < F_{\chi^2_{(2(N-1))}}^{-1} (\alpha) = k_\alpha \right\} \quad (38)$$

4.4.6 Tests de tendance appliqué au cas Weibull

Afin de nous assurer de la validé des tests de tendance, nous avons simulé les durées de bon fonctionnement dans le cas d'un système réparable en supposant que :

- le système est réparé instantanément dès qu'il tombe en panne (on néglige les temps de réparation)
- le système est remis dans le même état d'usure que celui juste avant la panne (hypothèse de réparation minimale)
- son taux de défaillance suit une loi de Weibull de paramètre (η, β) avec $\eta = 1$

$$\lambda(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta-1} \quad (39)$$

Nous remarquons que :

- Si $\beta > 1$, le système aura un taux de défaillance croissant, et de fait il y a aura usure et donc décroissance de la fiabilité
- Pour $\beta < 1$, le taux est décroissant, ce qui est synonyme d'amélioration du système et donc de croissance de la fiabilité
- Pour $\beta = 1$, le taux de défaillance est constant, la fiabilité est stabilisée, et l'on retombe dans le cas du modèle HPP

Le paramètre β peut donc être interprété comme le degré d'amélioration ou de dégradation du système. Nous présenterons les résultats des tests de tendances dans les cas $\beta = 0.1$, 1 et 10.

Simulation des durées de bon fonctionnements :

La méthode d'inversion de la fonction de répartition [9] permet de construire par récurrence N durées de bon fonctionnement. En notant (X_1, X_2, \dots) la suite des durées de bon fonctionnement successives, la loi de X_{n+1} sachant que $S_n = X_1 + \dots + X_n = t_n$ est donnée par sa fonction de répartition définie pour $\tau > 0$ par (voir § 6.3 de [5]) :

$$F_{X_{n+1}|S_n=t_n}(\tau) = 1 - \exp\left(-\frac{1}{\eta^\beta}((t_n + \tau)^\beta - t_n^\beta)\right)$$

On écrit alors :

$$F_{X_{n+1}|S_n=t_n}(\tau) = u \Leftrightarrow \tau = (t_n^\beta - \ln(1 - u))^{\frac{1}{\beta}} - t_n$$

D'après la méthode d'inversion, obtenir une réalisation τ de $X_{n+1}|S_n = t_n$ revient à simuler une réalisation u de loi uniforme sur $[0, 1]$ et à faire $\tau = (t_n^\beta - \ln(1 - u))^{\frac{1}{\beta}} - t_n$, t_n étant obtenu lors des précédentes itérations, comme la somme de toutes les réalisations précédentes de τ .

Résultat suite au test de laplace :

Reste donc à calculer pour chaque valeur de $\beta = 0.1, 1$ et 10 une réalisation de V_N pour N grand (test asymptotique), et à effectuer les tests au seuil α , c'est-à-dire selon les besoins, confronter l'hypothèse nulle H_0 contre les différentes hypothèses alternatives :

- Pour ce qui est des tests de H_0 contre H_2 et H_0 contre H_3 , nous allons tracer pour différentes valeurs de N , les positions des réalisations V_N par rapport aux quantiles $k_\alpha = F_{\mathbb{N}(0,1)}^{-1}(\alpha)$ et $l_\alpha = F_{\mathbb{N}(0,1)}^{-1}(1 - \alpha)$ définis dans (33) et (35). Le test étant asymptotique, on doit tenir compte des grandes valeurs de N .
La zone de rejet (32) du test H_0 contre H_2 correspond aux valeurs des réalisations de V_N se trouvant en dessous de $k_\alpha = F_{\mathbb{N}(0,1)}^{-1}(\alpha)$. La zone de rejet (34) du test H_0 contre H_3 correspond aux valeurs des réalisations de V_N se trouvant au dessus de $l_\alpha = F_{\mathbb{N}(0,1)}^{-1}(1 - \alpha)$
- Enfin s'agissant du test de H_0 contre H_1 , nous allons tracer pour différentes valeurs de N , les positions des réalisations $|V_N|$ par rapport aux quantiles $k_{2,\alpha} = F_{\mathbb{N}(0,1)}^{-1}(1 - \frac{\alpha}{2})$, la zone de rejet étant telle que $|V_N| > k_{2,\alpha}$

Considérons les valeurs suivantes de β et prenons $\alpha = 0.05$:

- $\beta = 0.1$: nous avons reporté (Figure 23) les résultats des tests décrits ci-dessus. Nous constatons que les réalisations de V_N sont situés dans la zone de rejet de l'hypothèse H_0 au profit des hypothèses H_2 (croissance de la fiabilité) et de H_1 ((X_1, \dots, X_N) n'est pas un HPP).
- $\beta = 1$: Nous constatons (Figure 24) que les réalisations des V_N sont situés en dehors des zones de rejet de H_0 contre $H_{1,2,3}$, ce qui est attendu puisque le cas $\beta = 1$ correspond au modèle HPP.
- $\beta = 10$: Nous constatons (Figure 25) que les réalisations de V_N sont situés dans la zone de rejet de l'hypothèse H_0 au profit des hypothèses H_3 (décroissance de la fiabilité) et de H_1 ((X_1, \dots, X_N) n'est pas un HPP).

Ces résultats sont tout à fait conformes à nos attentes d'après l'interprétation qu'on a pu faire des valeurs prises par β suite à la formule (39) du taux de défaillance.

Résultat pour le test logarithmique :

Pour le test logarithmique il faut simuler les réalisations de K_N pour $0.1, 1$ et 10 de β et effectuer les tests au seuil $\alpha = 0.05$:

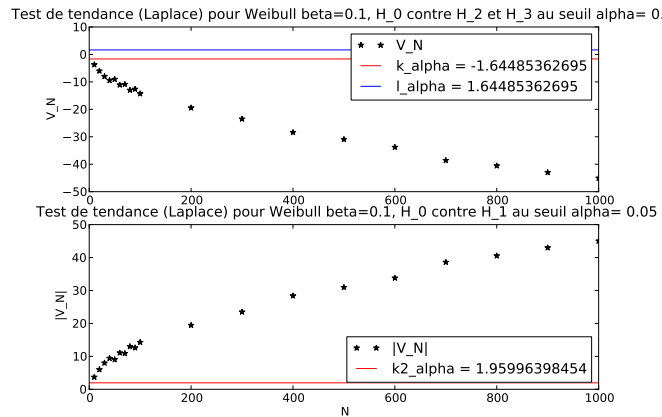


FIGURE 23 – Test de Laplace pour une loi de Weibull ($\eta = 1, \beta = 0.1$) (croissance de la fiabilité). Nous constatons que les réalisations de V_N sont situés dans la zone de rejet de l'hypothèse H_0 au profit des hypothèses H_2 (croissance de la fiabilité) et de H_1 ((X_1, \dots, X_N) n'est pas un HPP).

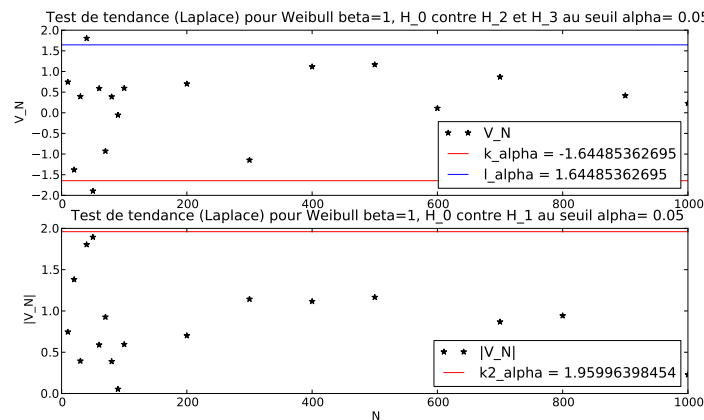


FIGURE 24 – Test de Laplace pour une loi de Weibull ($\eta = 1, \beta = 1$) (Absence de tendance). Nous constatons que les réalisations des V_N sont situés en dehors des zones de rejet de H_0 contre $H_{1,2,3}$, ce qui est attendu puisque le cas $\beta = 1$ correspond au modèle HPP. Nous constatons que les réalisations des V_N sont situés en dehors des zones de rejet de H_0 contre $H_{1,2,3}$, ce qui est attendu puisque le cas $\beta = 1$ correspond au modèle HPP.

- Pour ce qui est des tests de H_0 contre H_2 et H_0 contre H_3 , nous allons tracer pour différentes valeurs de N , les positions des réalisations K_N par rapport aux quantiles $k_{\alpha} = F_{\chi^2_{2(N-1)}}^{-1}(\alpha)$ et $l_{\alpha} = F_{\chi^2_{2(N-1)}}^{-1}(1 - \alpha)$. La zone de rejet (37) du test H_0 contre H_2 correspond aux valeurs des réalisations de K_N se trouvant au dessus de l_{α} et la zone de rejet du test H_0 contre H_3 correspond aux valeurs des réalisations de K_N se trouvant en dessous de k_{α} .
- Enfin s'agissant du test de H_0 contre H_1 , nous allons tracer pour différentes valeurs de

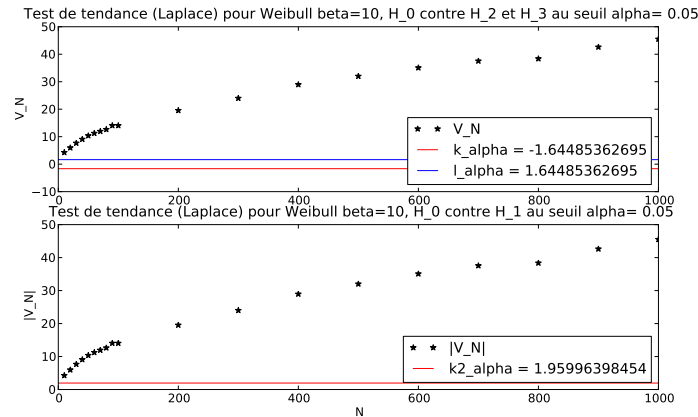


FIGURE 25 – Test de Laplace pour une loi de Weibull ($\eta = 1$, $\beta = 10$) (décroissance de la fiabilité).

N , les positions des réalisations K_N par rapport aux quantiles $k_{1,\alpha} = F_{\chi^2_{(2(N-1))}}^{-1}\left(\frac{\alpha}{2}\right)$ et $k_{2,\alpha} = F_{\chi^2_{(2(N-1))}}^{-1}\left(1 - \frac{\alpha}{2}\right)$, la zone de rejet de l'hypothèses H_0 contre H_1 correspond aux réalisations de K_N situé soit en dessous de $k_{1,\alpha}$, soit au dessus de $k_{2,\alpha}$ d'après la région critique (36).

Considérons les valeurs suivantes de β en prenant $\alpha = 0.05$:

- $\beta = 0.1$: nous avons reporté (Figure 26) les résultats des tests décrits ci-dessus. Nous constatons que les réalisations de V_N sont situés dans la zone de rejet de l'hypothèse H_0 au profit des hypothèses H_2 (croissance de la fiabilité) et de H_1 ((X_1, \dots, X_N) n'est pas un HPP).

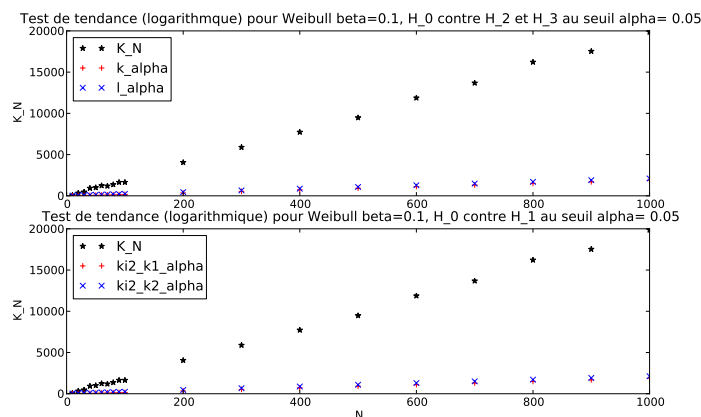


FIGURE 26 – Test logarithmique pour une loi de Weibull ($\eta = 1$, $\beta = 0.1$) (croissance de la fiabilité). Nous constatons que les réalisations de K_N sont situés dans la zone de rejet de l'hypothèse H_0 au profit des hypothèses H_2 (croissance de la fiabilité) et de H_1 ((X_1, \dots, X_N) n'est pas un HPP).

- $\beta = 1$: Nous constatons (Figure 27) que les réalisations des V_N sont situés en dehors des zones de rejet de H_0 contre $H_{1,2,3}$, ce qui est attendu puisque le cas $\beta = 1$ correspond au modèle HPP.

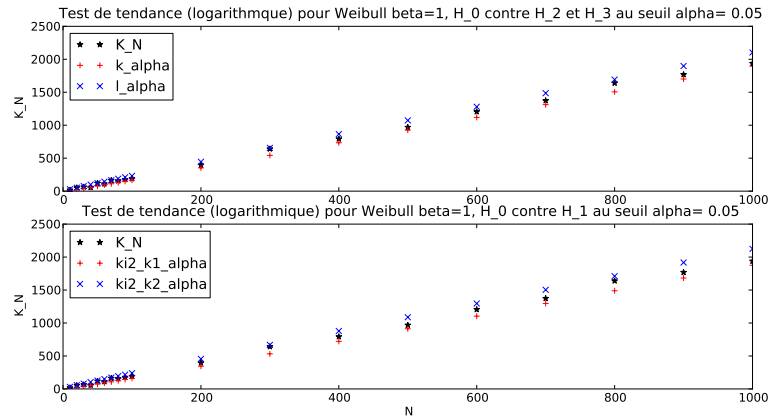


FIGURE 27 – Test logarithmique pour une loi de Weibull ($\eta = 1$, $\beta = 1$) (Absence de tendance). Nous constatons que les réalisations des V_N sont situés en dehors des zones de rejet de H_0 contre $H_{1,2,3}$, ce qui est attendu puisque le cas $\beta = 1$ correspond au modèle HPP. Nous constatons que les réalisations des V_N sont situés en dehors des zones de rejet de H_0 contre $H_{1,2,3}$, ce qui est attendu puisque le cas $\beta = 1$ correspond au modèle HPP.

- $\beta = 10$: Nous constatons (Figure 28) que les réalisations de V_N sont situés dans la zone de rejet de l'hypothèse H_0 au profit des hypothèses H_3 (décroissance de la fiabilité) et de H_1 ((X_1, \dots, X_N) n'est pas un HPP).

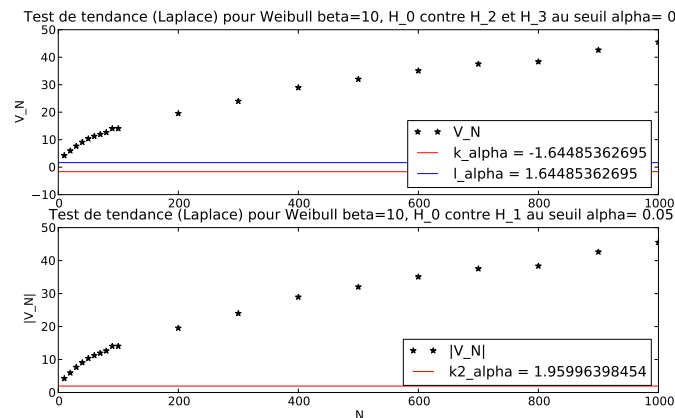


FIGURE 28 – Test logarithmique pour une loi de Weibull ($\eta = 1$, $\beta = 10$) (décroissance de la fiabilité).

Ces résultats sont donc les mêmes que ceux obtenus dans le cas du test de Laplace, ce qui valide les tests de tendance suggérés.

4.4.7 Tests de tendance appliqué aux données du système

Nous pouvons maintenant appliquer ces tests sur les données du système simulé. Nous avons résumé les résultats obtenus des tests de tendance sur les Figures 29 et 30 avec les mêmes considérations que dans le cas d'un système régi par une loi de Weibull.

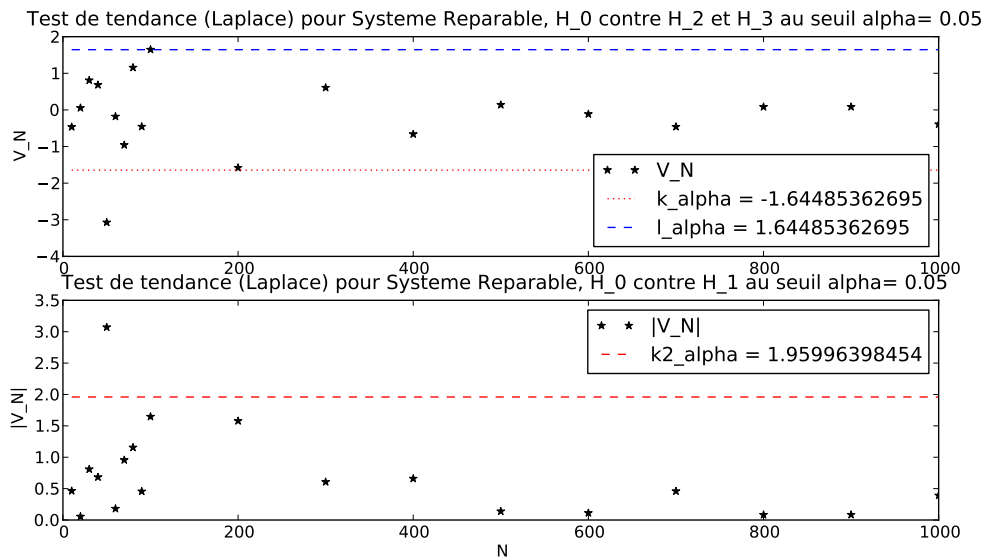


FIGURE 29 – Test asymptotique de Laplace au seuil $\alpha = 0.05$, de l'hypothèse nulle H_0 ("il y a absence de tendance") contre les hypothèses H_2 : "il y a *croissance de la fiabilité*" puis H_3 : "il y a *décroissance de la fiabilité*". La zone de rejet (32) du premier test (H_0 contre H_2) correspond aux valeurs des réalisations de V_N se trouvant en dessous de $k_{\alpha} = F_{N(0,1)}^{-1}(\alpha)$. Vu la position inverse des étoiles représentant ces réalisations par rapport à k_{α} , nous pouvons conclure qu'on ne peut pas rejeter H_0 au profit de l'hypothèse alternative H_2 avec le seuil de confiance α choisi. Nous arrivons à la même conclusion pour ce qui est du test H_0 contre H_3 . Puis test asymptotique de Laplace au seuil $\alpha = 0.05$, de l'hypothèse nulle H_0 ("il y a absence de tendance") contre l'hypothèse alternative H_1 (" X_i n'est pas un HPP"). La zone de rejet (31) correspond aux valeurs des réalisations de $|V_N|$ se trouvant au dessus de $k_{2,\alpha} = F_{N(0,1)}^{-1}(1 - \frac{\alpha}{2})$. Vu la position inverse des étoiles représentant ces réalisations pour les grandes valeurs de N par rapport à $k_{2,\alpha}$, nous pouvons conclure qu'on ne peut pas rejeter l'hypothèse H_0 au profit de l'hypothèse alternative H_1 avec le seuil de confiance α choisi.

4.4.8 Test β du modèle PLP

Comme nous l'avons indiqué dans la sous-section 4.3.2, nous remarquons que l'estimateur $\hat{\beta}_N$ de β dans l'expression 27 du taux de défaillance converge vers 1, c'est-à-dire que si on se place dans le cadre d'un modèle PLP, l'observation du système nous ramène au modèle HPP. La dualité entre intervalles de confiance et tests d'hypothèse fait que l'on peut facilement construire des tests portant sur la valeur de β . En particulier, comme " $\beta = 1$ " correspond au HPP, on peut tester l'hypothèse qu'un PLP est en fait un HPP en testant $H_0 : \beta = 1$ contre

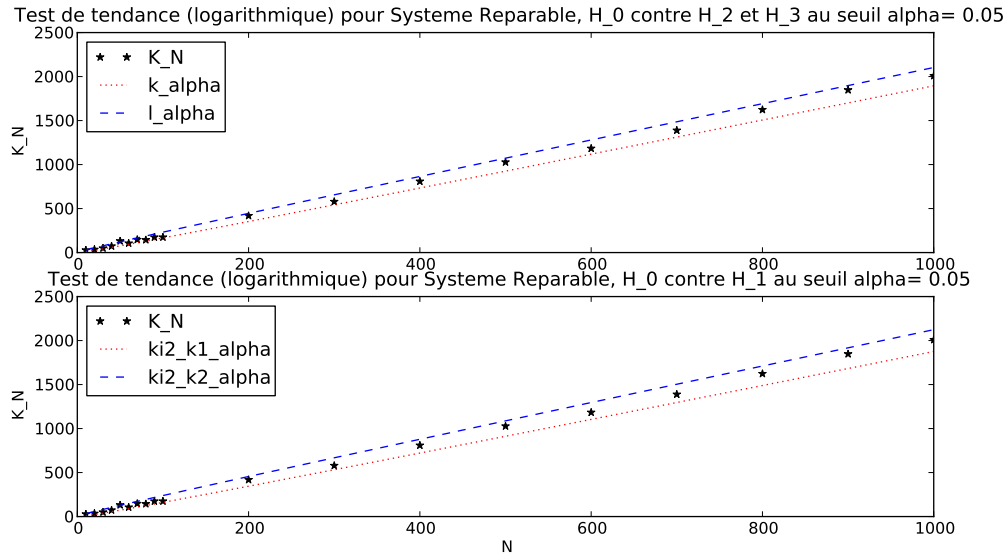


FIGURE 30 – Test logarithmique au seuil $\alpha = 0.05$ de l'hypothèse nulle H_0 : "il y a absence de tendance" contre les hypothèses H_2 : "il y a croissance de la fiabilité" puis H_3 : "il y a décroissance de la fiabilité". La zone de rejet (37) du premier test H_0 contre H_2 , correspond aux valeurs des réalisations de K_N se trouvant au dessus de $l_\alpha = F_{\chi^2_{(2(N-1))}}^{-1}(1 - \alpha)$. Vu la position inverse des étoiles représentant ces réalisations par rapport à l_α , nous pouvons conclure qu'on ne peut pas rejeter H_0 au profit de l'hypothèse alternative H_2 avec le seuil de confiance α choisi. Nous arrivons à la même conclusion pour ce qui est du test H_0 contre H_3 . Puis test logarithmique au seuil $\alpha = 0.05$ de l'hypothèse nulle H_0 : "il y a absence de tendance" contre l'hypothèse alternative H_1 : " (X_i) n'est pas un HPP". La zone de rejet (36) correspond aux valeurs des réalisations de K_N se trouvant en dessous de $k_{1,\alpha} = F_{\chi^2_{(2(N-1))}}^{-1}(\frac{\alpha}{2})$ ou au dessus de $k_{2,\alpha} = F_{\chi^2_{(2(N-1))}}^{-1}(1 - \frac{\alpha}{2})$. Vu la position inverse des étoiles représentants ces réalisations par rapport à $k_{1,\alpha}$ et $k_{2,\alpha}$, nous pouvons conclure qu'on ne peut pas rejeter H_0 au profit de l'hypothèse alternative H_1 avec le seuil de confiance α choisi.

H_1 : " $\beta \neq 1$ ". Le test consiste à rejeter " $\beta = 1$ " au seuil α si la valeur 1 n'appartient pas à l'intervalle de confiance bilatéral pour β de seuil α donné par (28).

En posant $k_{1,\alpha} = F_{\chi^2_{(2(N-1))}}^{-1}(\frac{\alpha}{2})$, $k_{2,\alpha} = F_{\chi^2_{(2(N-1))}}^{-1}(1 - \frac{\alpha}{2})$ et $Q_N = 2 \sum_{i=1}^{N-1} \ln(\frac{S_N}{S_i})$, la région critique de ce test est :

$$\begin{aligned} W &= \left\{ 1 \notin \left[\frac{\hat{\beta}}{2N} F_{\chi^2_{(2(N-1))}}^{-1}(\frac{\alpha}{2}), \frac{\hat{\beta}}{2N} F_{\chi^2_{(2(N-1))}}^{-1}(1 - \frac{\alpha}{2}) \right] \right\} \\ &= \left\{ Q_N < k_{1,\alpha} \right\} \cup \left\{ Q_N > k_{2,\alpha} \right\} \end{aligned} \quad (40)$$

On peut montrer que ce test est optimal au sens où il est uniformément le plus puissant parmi les tests sans biais (voir fin du § 9.2.2.1 de [5]).

Simulation : En prenant $\alpha = 0.05$ et des échantillons (X_1, \dots, X_N) de plus en plus grand, on a illustré sur la Figure 31 le résultat du test de l'hypothèse H_0 : " $\beta = 1$ " contre H_1 : " $\beta \neq 1$ ".

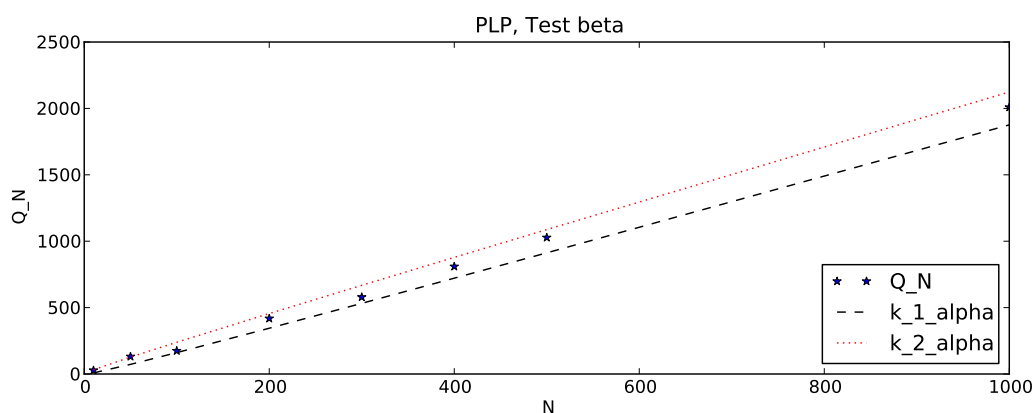


FIGURE 31 – Test β du modèle PLP. Nous remarquons que les réalisations de la variable aléatoire Q_N sont situées en dehors de la zone de rejet définie par (40). Nous ne pouvons donc pas rejeter l'hypothèse nulle $H_0 : \beta = 1$ contre l'hypothèse alternative $H_1 : \beta \neq 1$ au seuil de confiance $\alpha = 0.05$, c'est-à-dire qu'on ne peut pas rejeter l'hypothèse que le PLP est un HPP.

4.5 Conclusion

Cette partie a permis de mettre en place des méthodes de validation de modèles dans le cadre du pronostic d'un système dont on ne connaît pas les caractéristiques. Ces considérations peuvent être appliquées à un système quelconque étant donnée que les méthodes développées n'utilisent pas les propriétés intrinsèques du système considéré dans son approche boîte noire. L'intérêt d'une telle approche est de pouvoir se demander si on peut modéliser un système complexe donné par un processus de défaillance simple avec un soucis constant que ce choix de modélisation n'altère pas les impératifs de la sûreté de fonctionnement.

Nous nous sommes intéressés à ce qu'on a appelé *erreur de modèle* qu'on a appliqué à la fiabilité, et qui correspond à l'erreur qu'on fait quand on remplace le système étudié par celui, plus simple, dont le processus de défaillance est un processus de Poisson homogène. Nous avons pu calculer l'erreur minimale (~ 2.70) dans ce cadre, et avons remarqué que l'utilisation des instants de défaillance du système non réparable conduisait à une erreur plus proche (~ 2.72) que celle qu'on obtient (~ 25.29) quand on considère les durées de bon fonctionnement.

Les nombreux tests statistiques réalisés dans le cadre du stage ont permis de discuter, la validité de la modélisation du système par un processus de Poisson homogène, étant donné un seuil de confiance α . En fonction des données fournies (Instants de défaillance ou durées de bon fonctionnement) nous disposons de plusieurs tests. Nous avons résumé les tests pouvant être réalisés dans le *tableau de choix* suivant :

Données	Tests	Résultats
(T_1, \dots, T_N) ou (X_1, \dots, X_N)	Test de Anderson Darling	Rejet de H_0
(T_1, \dots, T_N) et (X_1, \dots, X_N)	Test de comparaison de KS	Rejet de H_0
	Test de comparaison des IC	Rejet de H_0
(X_1, \dots, X_N)	Test de tendance	Non Rejet de H_0

Les résultats obtenus sont cohérents dans les trois premiers cas. *A contrario* le dernier n'est pas satisfaisant. Nous avons pu l'interpréter compte tenu de l'allure (Figure 32) du véritable taux de défaillance associé au système étudié.

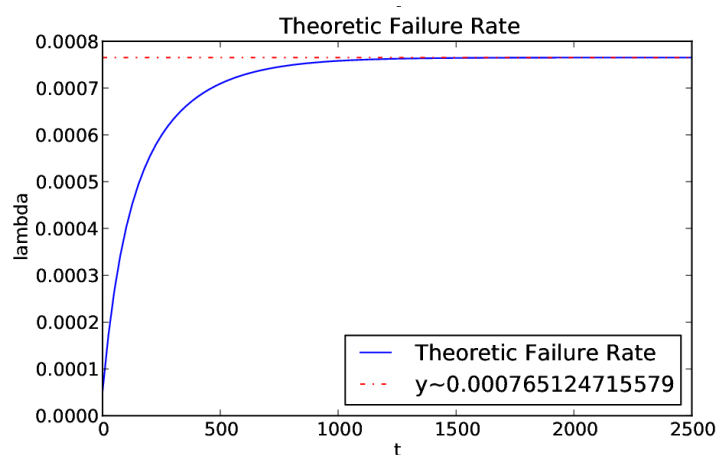


FIGURE 32 – Courbe représentant le taux de défaillance théorique du système étudié.

Nous constatons en effet que le taux de défaillance tend rapidement vers une constante, ce qui fait tendre le processus de défaillance vers un HPP et explique le résultat obtenu dans le cas des tests de tendance. Nous pouvons également émettre des réserves quant au respect des conditions d'application de ces tests qui supposent des réparations minimales, ce qui n'est pas le cas de notre système d'étude. Ce que l'on peut retenir, est que lorsqu'on dispose d'un système dont le taux de défaillance est monotone et tend vers une constante, alors le test de tendance ne fournit pas un outil de validation de rejet de l'hypothèse selon laquelle le système n'est pas un HPP. Le test le plus approprié de validation du modèle HPP est le test de Anderson-Darling qui ne nécessite qu'un seul type de donnée ((T_1, \dots, T_N) ou (X_1, \dots, X_N)). Les tests de comparaison souffrent de la nécessité d'avoir les deux types d'information, ce qui n'est pas donné dans la pratique.

5 Approche structurelle

5.1 Contexte

Les formules de la proposition 3.2 permettent en fait de faire des calculs de pronostic du système modélisé par un processus markovien de sauts, si on connaît la sous matrice des états de marche que nous avons noté B , et la loi initiale du processus markovien de sauts noté Z . De fait, la loi initiale étant une donnée du problème, si on parvient à estimer les coefficients de la sous matrice B , on pourra dès lors faire des calculs de pronostic.

On reprend donc ici le système formé des trois composants de la Figure 1, les deux générateurs et la lampe. On reprend également les notations de la partie introductive. Le système considéré étant modélisé par un processus markovien de sauts, les paramètres λ_i , μ_i , λ'_1 , λ'_2 sont constants. Avant d'introduire les estimateurs utilisés, nous allons définir la notion importante de censure, utile à la construction de l'estimateur de Nelson-Aalen qui doit être adapté à notre cas. Enfin nous allons construire l'estimateur de Kaplan-Meier qui permet d'estimer la fiabilité, y compris dans le cas d'observations incomplètes. On déduira ensuite des estimateurs des coefficients de B en utilisant la formule $R(t) = \exp(-\lambda t)$.

Notion de Censure : Nous allons nous intéresser à l'estimation des paramètres λ_i et λ'_i pour $i = 1, 2$. Pour fixer les idées, nous nous proposons l'estimation de λ_1 dans le cas où le système n'est pas réparable. Nous allons considérer l'expérience suivante :

1. tous les composants sont en état de marche à l'instant $t = 0$
2. le générateur 2 tombe en panne à un instant t_2
3. le générateur 1 tombe en panne à un instant $t_1 > t_2$

L'expérience observée se termine à l'instant t_1 , vu l'hypothèse de système non réparable. Si on avait $t_1 < t_2$, t_1 serait l'instant de panne observé du composant 1 permettant l'estimation de son taux de défaillance λ_1 . Le fait d'avoir ici $t_1 > t_2$, entraîne que la date t_2 censure l'observation de la date de panne t_1 servant à estimer le paramètre λ_1 , le taux de défaillance du composant 1 passant à cette date de λ_1 à λ'_1 . La censure consiste ici à ce que la seule observation du composant 1 induira une information incomplète du fait que suite à la panne du composant 2, son taux de défaillance doit passer à l'instant t_2 de λ_1 à λ'_1 . Le fait que le composant 2 soit tombé en panne à l'instant $t_2 < t_1$, entraîne que l'instant de panne t_1 n'est pas la seule date pertinente à prendre en compte pour l'estimation du paramètre λ_1 : il convient également de tenir compte de la date t_2 qui correspond à une date dite de censure pour le composant 1.

Ce type de censure est appelé censure aléatoire droite.

5.2 Estimation de paramètres

Nous allons dans cette section chercher à estimer les coefficients de B . L'estimateur de Nelson-Aalen (NA) permettra l'estimation du taux de défaillance cumulé. Nous introduirons également l'estimateur du maximum de vraisemblance (EMV) de la matrice génératrice développé dans [1]. Afin d'illustrer les méthodes d'estimation, nous allons nous consacrer à estimer le taux de défaillance λ_1 du composant 1. Les explications fournies dans ce qui suit pourront être transposées à l'estimation de tous les paramètres de B . Nous chercherons par la suite à

comparer indirectement l' EMV de A avec celui fourni par l'estimateur de NA compte tenu de la relation :

$$\Lambda_1(t) = \int_0^t \lambda_1(s) ds = t\lambda_1$$

5.2.1 Estimation de λ_1 via l'estimateur de NA

Nous allons construire cet estimateur de façon heuristique en considérant N copies d'un système non réparable quelconque dont on cherche à estimer le taux de défaillance cumulé Λ . Nous donnerons par la suite l'expression de cet estimateur dans le cas d'un système réparable puis transposerons cette construction au cas où ce système est notre composant 1. Nous supposons que les N copies sont mis en marche au même instant initial.

Cas où l'on dispose des premiers instants de panne : Cette construction repose sur la remarque suivante : pour $dt \rightarrow 0$,

$$\lambda(t) dt \sim P(t \leq T < t + dt \mid T \geq t)$$

avec T la variable aléatoire représentant le premier instant de panne.

Soit $(0, t_1, t_2, \dots)$ une subdivision régulière de notre intervalle d'observation $[0, S]$ de pas "très petit". Notons d_i le nombre de copies tombant en panne à l'instant t_i , et n_i le nombre de copies non encore tombées en panne *juste avant* l'instant t_i . Nous pouvons considérer l'estimation :

$$P(t_i \leq T < t_{i+1} \mid T \geq t_i) \sim \frac{d_i}{n_i}$$

En sommant nous obtenons alors :

$$\Lambda(t) \sim \sum_{\{i; t_i < t\}} \lambda(t_i)(t_{i+1} - t_i) \sim \sum_{\{i; t_i < t\}} \frac{d_i}{n_i} \quad (41)$$

Une autre construction est celle développée dans [12] et qui consiste à prendre les t_i représentant les instants de pannes rangés par ordre croissant. Cette construction, utilisant des données aléatoires est certes maladroite, mais permettra une comparaison facile entre NA et EMV surtout dans le cas de données censurées.

Cas où l'on dispose des durées de bon fonctionnement : La formule précédente se généralise au cas d'un système réparable (d'abord avec $N = 1$), en remplaçant les instants ordonnés de premières pannes (t_i) par les durées ordonnées de bon fonctionnement (x_i). Elle peut aussi se généraliser en considérant N systèmes réparables, en ordonnant les durées de bon fonctionnement observées sur tout l'intervalle $[0, S]$. Nous allons présenter cette construction dans le cas où le système est le composant 1. Reconsidérons donc le composant 1 et cherchons à exprimer son estimateur de Nelson-Aalen à un instant t en tenant compte des durées censurées. Nous allons traiter le cas de N copies du système modélisé par le processus markovien de sauts $(Z)_{0 \leq t \leq S}$ soumis à l'observation jusqu'à l'instant S . Rappelons que dans le cas étudié, il y a absence d'ex-aequo. Donc dans la formule (41) on a $d_i = 1$. Ensuite il convient dans cette somme, de ne tenir compte que des instants où le composant 1 est effectivement tombé en panne, supprimant de fait les termes correspondant aux instants qui censurent l'estimation

du paramètre λ_1 . Ces durées de censure correspondent aux transitions $e_1 \rightarrow e_3$ et $e_1 \rightarrow e_4$ (Nous avons ajouté la transition $e_1 \rightarrow e_4$ par rapport au paragraphe qui a introduit la notion de censure, car nous nous plaçons ici dans le cas réparable). Dans cette somme l'entier n_i ne représente plus le nombre de composants non encore tombé en panne, mais les composants à risque, c'est-à-dire ni morts, ni censurés. Notons $(x_i)_{i \in \{1, \dots, K\}}$ la suite triée par ordre croissant des durées de séjour dans l'état 1 ; ceci définit l'entier K comme étant le nombre de transitions correspondants à la panne du composant 1, c'est-à-dire celle permettant l'estimation de λ_1 , et des transitions censurant l'estimation de ce taux de défaillance). Ces durées correspondent aux transitions $e_1 \rightarrow e_i$ avec $i = 2, 3, 4$. Si on définit $(\varepsilon_i)_{i \in \{1, \dots, K\}}$ telle que :

- $\varepsilon_i = 1$ si x_i correspond à une transition $e_1 \rightarrow e_2$ (panne de 1)
- $\varepsilon_i = 0$ si x_i correspond à une transition $e_1 \rightarrow e_3$ ou e_4 (censure de 1)

alors l'estimateur de Nelson-Aalen du composant 1 s'écrit (voir aussi [12]) :

$$\hat{\Lambda}_{S,N}^1(t) = \sum_{\{i; x_i \leq t\}} \frac{\varepsilon_i}{K - i + 1} \quad (42)$$

$K - i + 1$ représentant le nombre de composant 1 ni morts ni censurés juste avant l'instant x_i . On a alors :

$$\text{Pour } N \text{ fixé, } \hat{\Lambda}_{S,N}^1(t) \xrightarrow{f \rightarrow \infty} \Lambda_1(t) = t\lambda_1$$

$$\text{Pour } S \text{ fixé, } \hat{\Lambda}_{S,N}^1(t) \xrightarrow{N \rightarrow \infty} \Lambda_1(t) = t\lambda_1$$

Donc pour $t > 0$ fixé, l'estimateur de λ_1 issue de NA pourrait être :

$$\hat{\lambda}_{S,N}^1 = \frac{\hat{\Lambda}_{S,N}^1(t)}{t} \quad (43)$$

Résultat graphique : Afin de mettre en évidence la convergence de l'estimateur (43) ci-dessous vers le taux de défaillance λ_1 , nous avons choisi de tracer l'estimateur (42) au regard de la droite de pente λ_1 (Nous utiliserons néanmoins l'estimateur (43) afin de le comparer avec l'*EMV* traité ci-dessous). Nous avons tracé la Figure 33 en considérant le cas de N copies d'un système non réparable. Nous remarquons que l'estimateur de NA (42, en remplaçant les x_i par les t_i) suit bien la pente donnée par le paramètre λ_1 . Nous constatons que l'estimation est bonne même pour $N = 1000$. Nous avons également tracé (Figure 34) l'estimateur de NA dans le cas où on a fait vivre le processus Z pendant 20000 heures (cas réparable) en considérant un seul système puis $N = 500$ systèmes. Nous remarquons que l'estimation est vraiment mauvaise dans le premier cas compte tenu du peu de données disponibles et qu'on a intérêt à jouer aussi sur le nombre de systèmes disponibles. On a aussi fait vivre un seul système 1 million d'heures et avons obtenu le résultat Figure 35.

Ces résultats graphiques mettent bien en évidence la convergence de l'estimateur de NA et montre qu'on peut l'utiliser pour l'estimation du taux de défaillance du composant 1, mais uniquement pour des instants t pas très grand (typiquement de l'ordre de 1000 avec les données utilisées).

5.2.2 Estimation de λ_1 via *EMV*

L'estimation de λ_1 rentre dans le cadre plus général de l'estimation de la matrice génératrice du système considéré. Présentons le cas général et supposons que l'on veuille estimer le

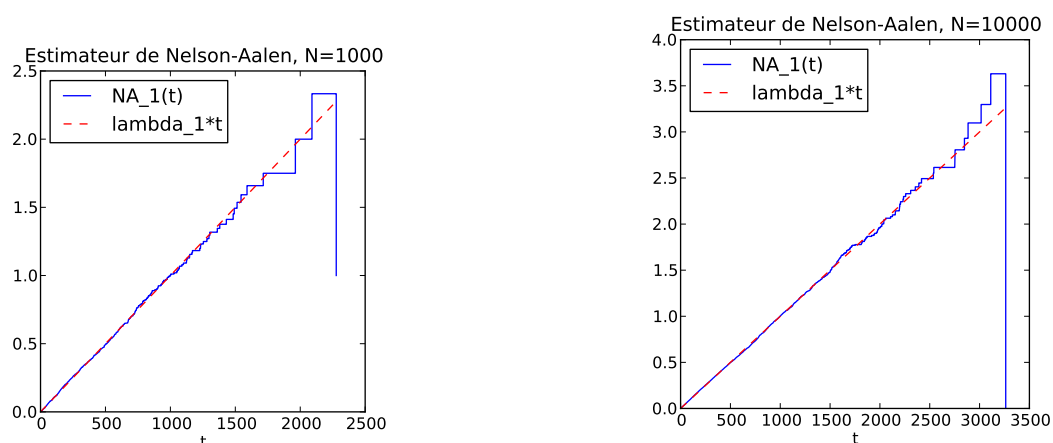


FIGURE 33 – Evolution de l'estimateur de Nelson-Aalen du composant 1.

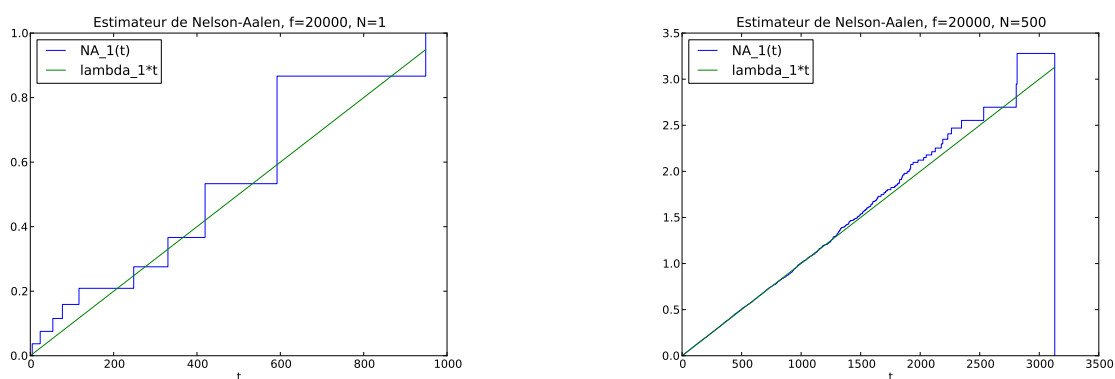


FIGURE 34 – Evolution de l'estimateur de Nelson-Aalen du composant 1 pour un durée d'observation fixée.

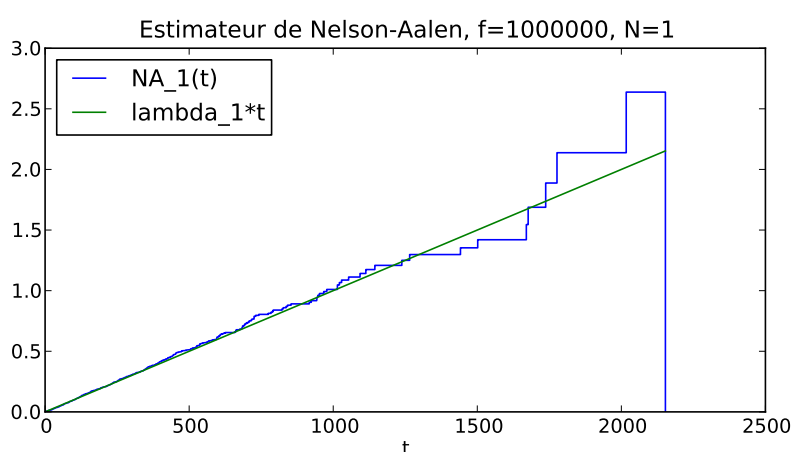


FIGURE 35 – Evolution de l'estimateur de Nelson-Aalen du composant 1.

paramètre $a_{i,j}$ correspondant à la transition $i \rightarrow j$. Soit $I = [0, S]$ l'intervalle d'observation

de N systèmes modélisés par le processus markovien de sauts Z , $n_{i,j}$ le nombre de transitions $i \rightarrow j$ sur tout l'intervalle I et v_i la durée de séjour passé dans l'état i . Alors d'après [1], un estimateur de $a_{i,j}$ est donnée par :

$$\hat{a}_{i,j}(S, N) = \begin{cases} \frac{n_{i,j}}{v_i} & , \text{ si } i \neq j \\ -\sum_{k \neq i} \hat{a}_{i,k} & , \text{ sinon} \end{cases} \quad (44)$$

et on montre que pour N fixé, sous certaines conditions :

$$\hat{a}_{i,j}(S, N) \xrightarrow{S \rightarrow \infty} a_{i,j}$$

et pour S fixé :

$$\hat{a}_{i,j}(S, N) \xrightarrow{N \rightarrow \infty} a_{i,j}$$

Il s'ensuit que calculer l'EMV de $\lambda_1 = a_{1,2}$ du composant 1 de notre système, revient à compter les transitions $e_1 \rightarrow e_2$ et le temps de séjour dans l'état 1 sur toute la durée d'observation $[0, S]$.

Nous avons tracé (Figure 36) l'évolution de $\hat{a}_{1,2}(S, 1)$ en fonction de S (avec $N = 1$).

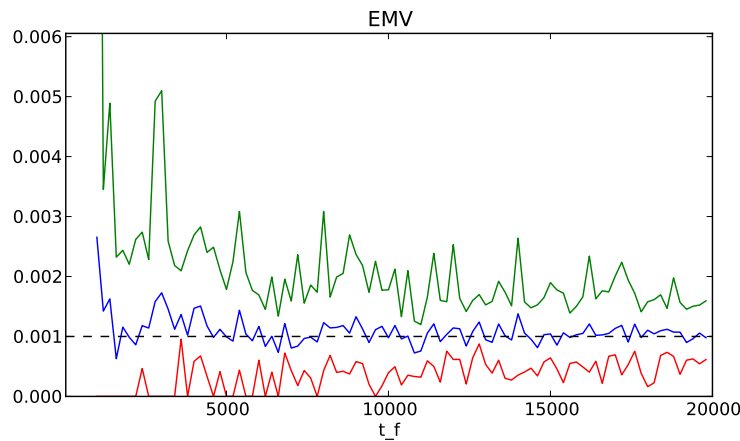


FIGURE 36 – Evolution de l'estimateur du maximum de vraisemblance de λ_1 .

Comparaison entre EMV et NA Nous allons nous intéresser à l'éventuelle relation algébrique entre l'estimateur donnée par (42) et celui dans (44). En reprenant les notations des deux paragraphes précédents consacré à l'estimation du taux de défaillance $\lambda_1 = a_{1,2}$ par *NA* puis par *EMV*, nous pouvons écrire :

$$n_{1,2} = \sum_{i=1}^K \varepsilon_i \quad \text{et} \quad v_1 = \sum_{i=1}^K x_i$$

Donc l'EMV de λ_1 est donnée par :

$$\frac{n_{1,2}}{v_1} = \frac{\sum_{i=1}^K \varepsilon_i}{\sum_{i=1}^K x_i}$$

et celui de donnée par NA est :

$$\frac{\sum_{\{i; x_i \leq t\}} \frac{\varepsilon_i}{K-i+1}}{t}$$

Donc la question est : y a-t-il une relation entre :

$$\frac{\sum_{i=1}^K x_i}{t} \quad \text{et} \quad \frac{\sum_{i=1}^K \varepsilon_i}{\sum_{\{i; x_i \leq t\}} \frac{\varepsilon_i}{K-i+1}} \quad ?$$

Le cas d'égalité en tout cas est complètement exclu.

Résultat Graphique :

Nous avons considéré le cas $N = 1$, avons fait évoluer la durée d'observation S afin de tracer (Figure 37) l'évolution de l'estimateur (43) du taux de défaillance issu de NA (avec $t = 1000$ fixé) et l' EMV de λ_1 . Nous avons mis en évidence des intervalles de confiance empiriques afin de comparer les deux estimateurs. Nous remarquons que l' EMV est meilleur vu la longueur de son intervalle de confiance plus petite que dans le cas de NA .

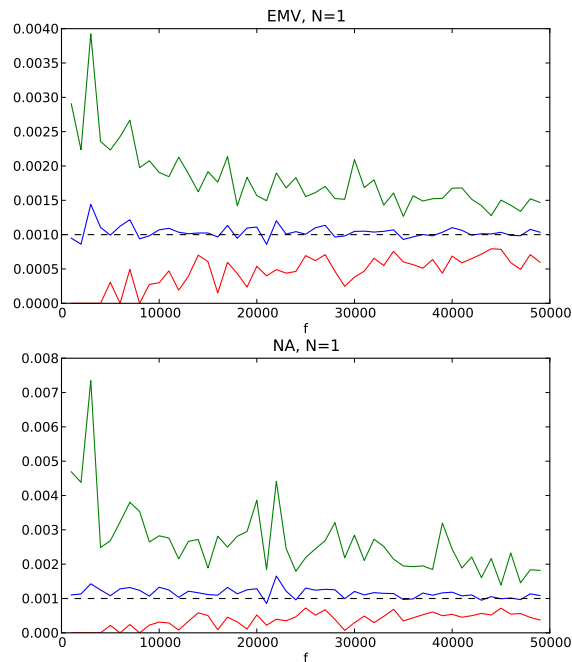


FIGURE 37 – Comparaison de l' EMV de λ_1 et l'estimateur de NA .

5.2.3 Retour sur l'approche Globale

La formule (42) s'applique en fait au cas du système vu comme une boîte noire. De fait, nous pouvons *a priori* estimer le taux de défaillance du système à partir de l'estimateur de NA en remarquant que :

$$\lambda(t) = \Lambda'(t) \sim \frac{\Lambda(t + dt) - \Lambda(t)}{dt}$$

Néanmoins, la discussion que nous avons eu au § 4.2.2, nous a montré que cette estimation dépendait du pas de temps dt , qui plus est s'était avérée de mauvaise qualité. Nous avons pu dans l'approche structurelle estimer le taux de défaillance du composant 1 à partir de l'estimateur de NA du fait que λ_1 est dans notre cas constant.

5.2.4 Estimateur de Kaplan-Meier

Le texte fondateur de cet estimateur est l'article [8]. Notons T la variable aléatoire représentant le premier instant de panne du composant 1 et $R^1(t) = \mathbb{P}(T > t) = \exp(-\lambda_1 t)$ sa fiabilité à un instant t . Soit (t_1, \dots, t_N) , N réalisations indépendantes de T tels que $t_1 < \dots < t_N$. Si on suppose qu'il n'y a pas de censure, une réalisation de $R^1(t)$ peut être la quantité :

$$\tilde{R}_N^1(t) = \frac{1}{N} \sum_{k=1}^N \mathbb{1}_{t_k \geq t}$$

Il s'agit d'adapter cet estimateur au cas de données censurées. Plusieurs constructions ayant été proposées depuis l'article de Kaplan et Meier, nous allons nous contenter de fournir une construction heuristique empruntée de [12]. Notons n_k le nombre de composants observés juste avant l'instant t_k .

Définition 5.1. Avec ces notations une réalisation de l'estimateur de Kaplan-Meier de $R^1(t)$ est défini par :

$$\hat{R}_N^1(t) = \prod_{\{k; t_k < t\}} \left(1 - \frac{1}{n_k}\right) \quad (45)$$

En effet pour vivre jusqu'à l'instant t le composant 1 doit d'abord vivre jusqu'à l'instant t_1 . Il doit ensuite rester en vie entre t_1 et t_2 sachant qu'il a vécu jusqu'à l'instant t_1 . Et ainsi de suite. La probabilité conditionnelle de tomber en panne à l'instant t_k sachant qu'il était en marche juste avant est $1/n_k$. Donc la probabilité conditionnelle de vivre à t_k est $1 - 1/n_k$. La probabilité de survivre jusqu'à l'instant t s'obtient par multiplication de toutes ces probabilités conditionnelles, d'où la formule (45).

En l'absence de censure, on aura $n_k = N - k + 1$, d'où, pour $t \in [t_i, t_{i+1}[$:

$$\hat{R}_N^1(t) = \prod_{\{k; t_k < t\}} \left(1 - \frac{1}{n_k}\right) = \prod_{\{k; t_k < t\}} \left(\frac{N - k}{N - (k - 1)}\right) = \frac{N - i}{N} = \tilde{R}_N^1(t)$$

On retrouve donc l'estimateur non paramétrique de la fiabilité dans le cas de données non censurées.

En tenant compte des données censurées, l'estimateur (45) se réécrit alors, en reprenant la notation ε_k du paragraphe 5.2.1 :

$$\hat{R}_N^1(t) = \prod_{\{k; t_k < t\}} \left(1 - \frac{\varepsilon_k}{N - k + 1}\right) \quad (46)$$

Nous pouvons dès lors poser comme estimateur du taux de défaillance du composant 1 :

$$\hat{\lambda}_N^1 = -\frac{1}{t} \ln(\hat{R}_N^1(t)) = -\frac{1}{t} \sum_{\{k; t_k < t\}} \ln \left(1 - \frac{\varepsilon_k}{N - k + 1}\right) \underset{\text{N grand}}{\sim} \frac{1}{t} \sum_{\{k; t_k < t\}} \frac{\varepsilon_k}{N - k + 1}$$

faisant apparaître l'estimateur de NA ! On retrouve dans un sens la formule (43).

5.3 Conclusion

Dans cette section, nous avons développé des calculs de pronostic en tenant compte des caractéristiques propres au système modélisé par un processus markovien de sauts. Dans cette approche structurelle, nous avons cherché à estimer la sous matrice B des états de marche du système qui apparaît dans les quantités permettant de faire des calculs de pronostic et résumées dans la proposition 3.2. Nous nous sommes rendu compte que la meilleur façon d'estimer les coefficients de B passait par l'estimateur du maximum de vraisemblance de la matrice génératrice A qui ne nécessite pas que l'on tienne compte des observations censurées. L'estimateur de NA estime le taux de défaillance cumulé et perd en précision pour les grands horizons de temps. Cela est un frein à une bonne estimation du taux de défaillance. L'estimateur de Kaplan-Meier nous a quant à lui ramené à l'estimateur de NA .

6 Conclusion

Le stage réalisé a permis de mettre en place des méthodes de calcul de pronostic dans l'objectif de prédire des pannes ou défaillance nécessitant une action de maintenance. Les méthodes utilisées s'appuyant sur des données, des simulations ont été effectuées afin d'obtenir les données correspondant à un simple système parallèle-série.

Nous avons adopté deux approches :

- une approche globale où nous avons assimilé le système à une boîte noire
- une approche structurelle où notre attention a été portée sur les composants du système

Dans l'*approche globale*, en fonction du type de données disponibles, nous avons construit des méthodes statistiques qui s'apparentent à des techniques de validation de modèle. Ces méthodes permettent d'affirmer avec un seuil de confiance fixé, si les données sont celles issues d'un système dont le processus de défaillance est connu. Concrètement on peut être amené à se demander si le processus de défaillance étudié est un processus de Poisson homogène (HPP). Le but est de pouvoir représenter un système compliqué par un modèle de défaillance simple sans compromettre les exigences de la sûreté de fonctionnement et la sécurité des usagers du système.

Nous remarquons qu'en fonction du type de données observées, nous pouvons adopter différentes stratégies de modélisation. Nous distinguons trois cas de figure :

- Si nous avons à notre disposition uniquement les premiers instants de panne, nous pouvons procéder à deux types d'estimation. Des estimations non paramétriques qui passent essentiellement par des calculs de types Monte-Carlo, et des estimations paramétriques qui supposent de fait un choix de modèle. Dans ce dernier cas nous avons développé le cas du modèle HPP pour lequel le processus de défaillance est exponentiel. Nous avons introduit une erreur de modèle calculée par validation croisée. Celle-ci permet d'approcher l'erreur lorsqu'on remplace le processus de défaillance réel du système par un HPP. Des tests statistiques d'adéquation de loi (comme le test d'Anderson Darling), peuvent être également utilisés pour affirmer ou infirmer si le système peut être modélisé par un HPP.
- Si nous disposons uniquement des durées de bon fonctionnement, nous pouvons effectuer des estimations paramétriques des quantités d'intérêt, et encore une fois, cela suppose le choix d'un modèle. Nous avons considéré deux types de modèles : le modèle HPP puis NHPP. Nous avons développé les tests de tendance qui permettent de dire avec un seuil de confiance donnée si la fiabilité du système est croissante ou décroissante. Le test de tendance a été accepté (à tort) dans le cas du système étudié. Nous avons expliqué ce phénomène par le fait que le taux de défaillance du système tend vers *une constante*. Dans le cas où le test tendance est accepté, il nous reste néanmoins la possibilité d'estimer le taux de défaillance cumulé $\Lambda(t)$ en utilisant l'estimateur de NA. Cependant, cela ne permet pas d'avoir un estimateur du taux de défaillance de "bonne qualité" en considérant uniquement l'approximation :

$$\lambda(t) = \Lambda'(t) \sim \frac{\Lambda(t + dt) - \Lambda(t)}{dt}$$

étant donnée la dépendance au choix du pas de temps dt .

- Si nous disposons des deux types de données, difficilement accessibles en pratique, nous

pouvons les comparer. En effet, si leurs distributions ne sont pas identiques, on peut affirmer que le processus de défaillance considéré n'est pas un HPP. Nous avons résumé sur la Figure 38 la discussion qu'on vient d'avoir.

Dans l'*approche structurelle*, nous avons pu estimer les paramètres de fiabilité de chaque composant en tenant compte des données de retour d'expérience issues d'observations censurées. Nous avons en effet remarqué que l'interaction que peut avoir un composant avec un autre, peut induire des observations incomplètes. Le choix de l'estimateur de Nelson-Aalen, doit passer par l'observation de chaque composant en diagnostiquant tous les cas éventuels de censure. La connaissance *a priori* qu'on a eu du système modélisé par un processus markovien de sauts, nous a permis grâce à l'estimateur du maximum de vraisemblance de la matrice génératrice, d'estimer les paramètres sans nous soucier de ces cas de censure.

Ce stage m'a permis un travail de fond sur les outils statistiques de test, de faire preuve de créativité dans la façon d'estimer ce qu'on appelé *erreur de modèle*. Il a également été l'occasion d'une immersion enrichissante dans le milieu de l'entreprise et l'industrie aéronautique en particulier, où j'ai eu la chance de participer activement au processus global de recherche dans le domaine de la modernisation de la maintenance chez EADS Innovation Works.

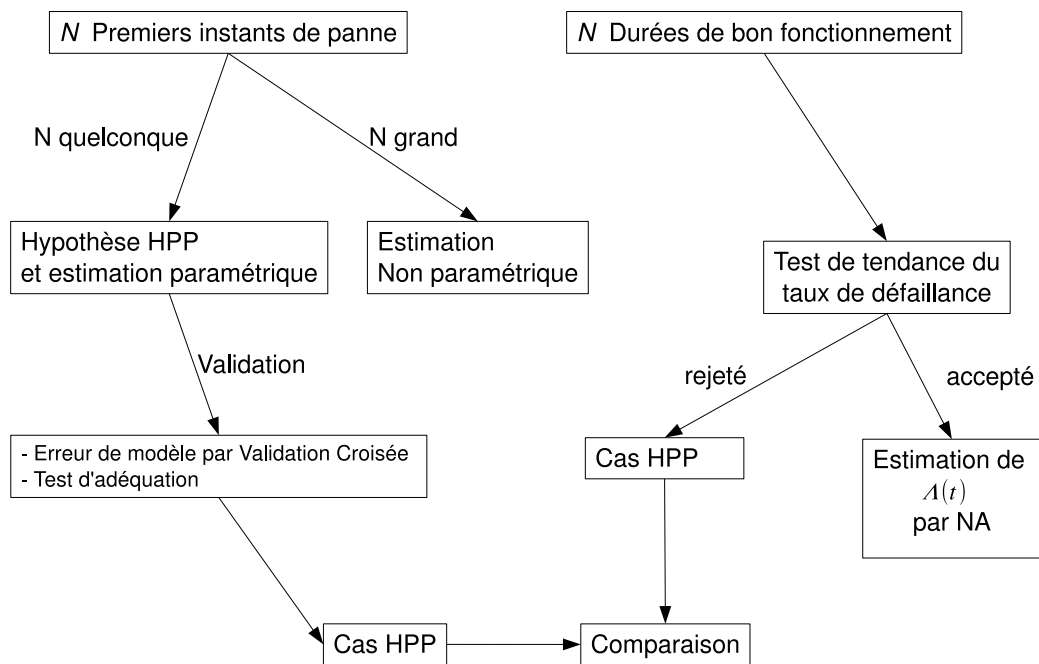


FIGURE 38 – Traitement en fonction des données de retour d'expérience.

Références

- [1] Nikolaos Limnios Amr Sadek. *Nonparametric estimation of reliability and survival function for continuous-time finite Markov*. Hermes Science, Lavoisier, 2007.
- [2] Dragan Banjevic. Remaining useful life in theory and practice. *Metrika*, 69 :337–349, décembre 2009.
- [3] Christiane Coccozza-Thivent. *Processus stochastiques et fiabilité des systèmes*. Springer, 1997.
- [4] Cox. *Renewal Theory*. Methuen, 1962.
- [5] Olivier Gaudoin. *Modélisation aléatoire en fiabilité des logiciels*. Hermes Science, Lavoisier, 2007.
- [6] Gross and Clark. *Survival Distribution : Reliability Application in the Biomedical Sciences*. John Wiley and Sons Inc, 1975.
- [7] Frank Guess and Frank Proschan. Mean residual life : theory and applications. *Handbook of Statistics, Quality Control and Reliability*, 7, June 1985.
- [8] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, pages 457–481, June 1958.
- [9] Annie Millet. Méthodes de monte carlo.
- [10] Alenxendre Muller. *Contribution à la maintenance prévisionnelle des systèmes de production par la formalisation d'un processus de pronostic*. PhD thesis, Université Joseph Fourier, 2003.
- [11] Thomas Mikosch Paul Embrechts, Claudia Kluppelberg. *Modelling for Extremal Events for Insurance and Finance*. Springer, 1997.
- [12] Germán Rodríguez. *Non-Parametric Estimation in Survival Models*. Springer, 2005.