

**EDF**

**Direction des Etudes et Recherches**

**Electricité  
de France**

SERVICE ENSEMBLES DE PRODUCTION  
Département Retour d'Expériences Mesures-Essais

Juillet 1993

LANNOY A.

**METHODES FREQUENTISTES DE TRAITEMENT  
DES DONNEES DE DEFAILLANCE**

***FREQUENTIST METHODS FOR FAILURE DATA  
ANALYSIS***

Pages : 51

93NB00119

Diffusion : J.-M. Lecœuvre  
EDF-DER  
Service IPN. Département SID  
1, avenue du Général-de-Gaulle  
92141 Clamart Cedex

© Copyright EDF 1993

ISSN 1161-0611

## **SYNTHÈSE :**

Cette note présente quelques méthodes de calcul fréquentistes d'une loi de fiabilité d'un matériel, prenant en compte les données de défaillance et les données de dégradation issues du retour d'expérience archivé dans une banque de données de défaillances.

## **EXECUTIVE SUMMARY :**

This note presents some frequentist methods for calculation of the reliability law of a component, taking into account failure data and degradation data, extracted from operation feedback data banks.

## SOMMAIRE

1.	LES DONNEES DE SURVIE	8
1.1.	Les types de données	8
1.2.	Le recensement des données nécessaires et les effets de la maintenance préventive	9
1.2.1.	L'effet de la maintenance préventive	9
1.2.2.	La maintenance corrective «As good as new» («aussi bon que neuf»)	10
1.2.3.	La maintenance corrective «As bad as old» («aussi mauvais que vieux»)	11
1.2.4.	En bref	11
2.	QUELQUES NOTIONS PRELIMINAIRES	12
2.1.	Généralités	12
2.2.	Quelques lois de survie	13
2.2.1.	la loi exponentielle	14
2.2.2.	La loi de Weibull à deux paramètres	14
2.2.3.	La loi de Weibull à trois paramètres	15
2.2.4.	Une autre forme de la loi de Weibull à deux paramètres	16
2.2.5.	La distribution Gamma	16
3.	LES DIFFICULTES LIEES A LA DEFINITION DE LA DEFAILLANCE	18
4.	LES METHODES NON PARAMETRIQUES	18
4.1.	La méthode de Wayne-Nelson	18
4.2.	La méthode de «Johnson» ou des rangs médians	19
4.3.	La méthode «Life Table» ou méthode actuarielle	22
4.4.	L'estimateur de Kaplan-Meier ou «Product Limit»	22
5.	LES METHODES PARAMETRIQUES	25
5.1.	Le choix a priori d'un modèle	25
5.2.	La méthode du maximum de vraisemblance	25
5.2.1.	Généralités	25
5.2.2.	Application à la loi exponentielle	26
5.2.3.	Application à la loi de Weibull à deux paramètres $\beta$ et $\eta$	27
5.2.4.	Exemples d'application	28
6.	LES INTERVALLES DE CONFIANCE	30
6.1.	Le cas de la loi exponentielle	30
6.2.	La méthode de la normalité asymptotique	32
6.3.	La méthode du rapport des vraisemblances	34

7.	LA DISTRIBUTION DES PARAMETRES DE FIABILITE - LA METHODE DU BOOTSTRAP	36
7.1.	Le principe du Bootstrap	36
7.1.1.	L'algorithme du Bootstrap	36
7.1.2.	Le cas particulier des données censurées	37
7.1.3.	Intervalle de confiance	37
7.2.	Exemple d'application : calcul des distribution des paramètres d'une loi	37
8.	LES METHODES DE DETECTION D'UNE EVOLUTION	39
8.1.	La méthode TTT (Total Time on Test)	39
8.2.	Le test de comparaison	41
8.3.	Le test de constance du taux de défaillance	41
8.4.	La méthode des «PP-plots»	43
9.	LE MODELE DES RISQUES PROPORTIONNELS	44
10.	L'INFLUENCE DE LA MAINTENANCE PREVENTIVE	47
10.1.	Détermination de la fiabilité intrinsèque	47
10.2.	Influence de la périodicité de la maintenance préventive	47
10.3.	Evolution de la fiabilité d'un matériel à la suite des opérations de maintenance préventive	49
11.	CONCLUSION	50
	REFERENCES BIBLIOGRAPHIQUES	51



## METHODES FREQUENTISTES DE TRAITEMENT DES DONNEES DE DEFAILLANCE

La fiabilité d'un matériel ou d'un équipement correspond à sa probabilité de bon fonctionnement, c'est-à-dire à la probabilité d'accomplir la fonction pour laquelle il est requis. A l'origine, il s'agit d'une notion qualitative.

Cependant, de plus en plus, la fiabilité devient un concept quantitatif et il convient de développer des méthodes afin de la mesurer.

Ces méthodes nécessitent des données de défaillances issues des banques de données de retour d'expérience. Ces méthodes permettent ainsi de déterminer à partir des observations de défaillance la loi de fiabilité observée du matériel. Cette loi de fiabilité observée prend en compte implicitement la politique de maintenance préventive réalisée sur ce matériel. Si l'on souhaite calculer la loi de fiabilité vraie (ou fiabilité intrinsèque) du matériel, on conçoit qu'il y a lieu de prendre en compte cette maintenance préventive et d'en estimer l'influence.

Le présent texte traite ces problèmes. En particulier, il met en évidence que, la plupart du temps, les données de défaillances sont censurées. Faut-il utiliser des méthodes paramétriques ou non paramétriques ? L'estimateur de Kaplan-Meier est la méthode à recommander dans le sens qu'elle ne nécessite aucune hypothèse quant au choix d'une loi de fiabilité. Au contraire, les méthodes paramétriques qui consistent à ajuster les données de retour d'expérience sur un modèle mathématique, le plus utilisé étant le modèle de Weibull, doivent permettre d'estimer les paramètres de ce modèle. Dans ces conditions, l'estimation des paramètres ne suffit pas. Il convient également de déterminer un intervalle de confiance ou plutôt une loi de distribution de ces paramètres. Le texte présente donc également ces méthodes.

L'utilisation de ces dernières peut être considérée comme un luxe. Les modèles de fiabilité de systèmes ne savent pas prendre en compte des valeurs variables avec le temps de fonctionnement. Une valeur de facteur d'erreur suffit bien souvent et la sensibilité des résultats est étudiée en prenant un facteur multiplicatif important, souvent calculé sur le taux de défaillance. Toutefois, il est clair que l'objectif du retour d'expérience est d'abord l'amélioration des matériels et que, pour cet objectif, le traitement des données brutes de défaillances est indispensable. De même, avant tout calcul, il est important de qualifier et valider les données de retour d'expérience : il s'agit de la tâche la plus difficile et la plus critique de l'analyse du retour d'expérience, mais on ne peut obtenir des résultats crédibles sans la réaliser.

Il faut distinguer enfin les matériels non réparables, pour lesquels la variable à étudier est la durée de vie d'un individu, des matériels réparables où l'on s'intéresse plutôt à la chronologie des défaillances d'un système. Cette distinction, faite par quelques auteurs de la littérature, n'est guère évidente. En effet, le taux de défaillance d'un individu et son évolution sont des notions utiles, que le matériel soit réparable ou pas.

La loi de fiabilité d'un matériel ne dépend pas que de son âge ou de son temps de fonctionnement. Elle dépend d'autres paramètres comme les conditions de fonctionnement, les caractéristiques de construction... Quelques mots seront dits sur des méthodes permettant le calcul d'une loi de fiabilité à plusieurs paramètres mais de nombreux progrès restent à réaliser sur ce point et il s'agit d'un sujet prioritaire pour la recherche.

*Note.* - Toutes les méthodes signalées dans ce chapitre sont disponibles dans des progiciels statistiques du commerce.

## 1. LES DONNEES DE SURVIE

Pour pouvoir déterminer une loi de fiabilité et ajuster ses paramètres, il faut constituer des échantillons de « temps »  $t$ .

### 1.1. Les types de données

Les échantillons sont formés de deux grands types de données :

- **les temps de défaillance** : le dispositif  $x$  a une défaillance à la date  $t$  (référéncée par rapport à une origine),  $t$  sera une donnée de défaillance de l'échantillon,

- **les temps de censure** (on dit aussi : les temps de troncature) : il existe trois types de données censurées : les données censurées à droite, à gauche ou par intervalle.

#### . les données censurées à droite

On décide d'arrêter l'observation à la date  $d$ . A cette date, le dispositif n'a pas eu de défaillance. La date  $d$  sera une donnée censurée à droite (ou tronquée à droite).

#### . les données censurées à gauche

On décide d'observer l'état du dispositif  $x$  à partir de la date  $g$ . On constate que le dispositif est défaillant mais l'on ne sait pas à quelle date  $t$  a eu lieu cette défaillance. On a seulement l'information :  $t$  est inférieure à  $g$ .

#### . les données censurées par intervalle

Le dispositif  $x$  a eu une défaillance entre deux dates  $g$  et  $d$  connues ; il s'agit d'une donnée censurée par intervalle.

Dans le cas du retour d'expérience, on a le plus souvent affaire à des données censurées à droite : à la date d'observation, le matériel n'a pas eu de défaillance. Remarquons que les opérations de maintenance préventive qui consistent à remplacer un composant non défaillant (mais pouvant être dégradé) génèrent des données censurées à droite.

#### *Remarques importantes :*

Il faut signaler que, dans tous les cas, les dates sont référencées par rapport à une **origine connue** (mise en service industriel ou date de remplacement du composant défaillant).

Les temps obtenus sont donc la différence entre deux dates. Le terme le plus exact n'est pas temps de défaillance ou temps de censure mais plutôt âge du dispositif au moment de la défaillance ou de la censure.



De plus, l'unité d'âge est en fait une unité d'usage qui peut, par exemple, s'exprimer en nombre de cycles, en nombre de kilomètres, en heures de fonctionnement, en nombre de sollicitations, etc.

## **1.2. Le recensement des données nécessaires et les effets de la maintenance préventive**

La prise en compte des données de retour d'expérience introduit certaines difficultés. En effet, l'exploitation d'un matériel peut être assimilée à un plan d'essais multicensurés:

- plusieurs dispositifs sont mis en service à une date donnée et ceux-ci ne sont pas obligatoirement défectueux à la fin de l'observation (censure de type I),
- plusieurs dispositifs sont mis en service, on arrête l'observation lorsqu'on a observé un nombre suffisamment important de défaillances (censure de type II).

Dans le cas des matériels suivis par un retour d'expérience :

- les matériels sont mis en service à des instants différents,
- à la date de fin de collecte ou de fin d'observation, tous les matériels ne sont pas ou n'ont pas été défectueux,
- certains matériels sont remplacés ou remis à neuf ou subissent une opération de maintenance préventive avant la date de fin de collecte ou d'observation.

La connaissance de l'unité d'âge est une grande difficulté. Souvent, celle-ci n'est pas connue ou mal connue. L'idéal serait la connaissance de l'âge exact du matériel au moment de sa défaillance.

Bien souvent, cette information n'est pas donnée, même dans les grandes banques de retour d'expérience, donc elle n'est qu'approchée (interpolée, ou moyennée ou estimée).

Dans le cadre de la collecte des données de survie, il est indispensable de distinguer trois types de maintenance : la maintenance préventive et deux types de maintenance corrective.

### **1.2.1. L'effet de la maintenance préventive**

De façon très sommaire, la maintenance préventive comprend la maintenance conditionnelle et la maintenance préventive systématique (périodique). Elle est effectuée dans l'intention de réduire la probabilité de défaillance d'un bien ou la dégradation d'un service rendu, avec l'objectif :

- d'augmenter la fiabilité d'un équipement qui se dégrade, donc de réduire les défaillances en service,
- d'augmenter la durée de vie de l'équipement,
- de réduire l'indisponibilité éventuelle due à une panne.

L'analyse statistique des défaillances survenues sur un matériel peut orienter une politique de maintenance préventive en conduisant à remplacer périodiquement un composant sensible : c'est, par exemple, le cas du presse-étoupe monté sur une vanne, que l'on remplace périodiquement par un presse-étoupe neuf, ce qui a pour effet de réduire la probabilité d'occurrence d'une défaillance (d'une fuite) sur la vanne, si l'opération de maintenance est parfaitement réalisée.

En réalité, une intervention sur un composant peut fragiliser le composant neuf et les composants voisins (erreur humaine, mauvais montage, mauvais serrage, etc.) et donc provoquer d'autres défaillances. Elle peut conduire à un remplacement prématuré d'un composant, par un autre susceptible de se dégrader rapidement pendant sa période de jeunesse. L'étude de Procaccia-Lannoy (1989) montre bien ces effets positifs (améliorer la fiabilité dans environ 70 % des cas pour les pompes ou les vannes) et négatifs (dégradation de la fiabilité dans environ 30 % des cas) dus aux opérations de maintenance.

Dans le cadre d'une opération de maintenance préventive ou d'une inspection, l'ouverture d'un dispositif peut abîmer un composant. Ce composant subira alors également une maintenance préventive.

Généralement, et il s'agit d'une hypothèse fondamentale, l'opération de maintenance préventive est supposée n'amener aucune dégradation aux divers composants constituant le matériel.

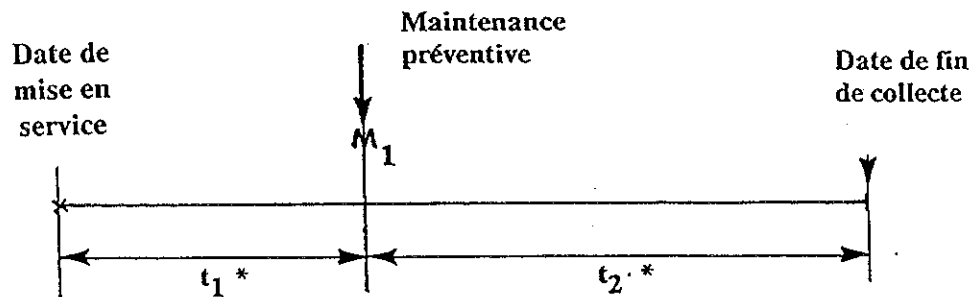


Figure 1. - Illustration de la maintenance préventive.

Sur l'exemple illustré par la figure 1., deux données censurées à droite  $t_1^*$  et  $t_2^*$  sont générées :

- l'une due au remplacement par un composant neuf, le composant initial n'étant pas défaillant à l'âge  $t_1^*$ ,
- l'autre due à la date de fin d'observation, le composant n'étant pas non plus défaillant à la date  $t_2^*$ .

### 1.2.2. La maintenance corrective « As good as new » (« aussi bon que neuf »)

Exemple :

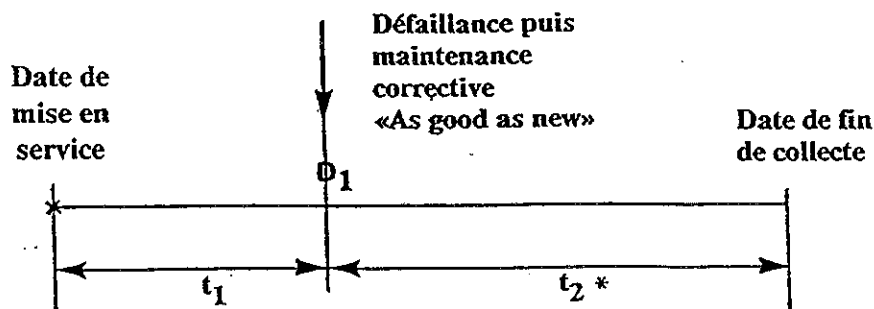


Figure 2. - Illustration de la maintenance « As good as new ».

A la date  $D_1$ , le matériel a une défaillance. A la suite de cette défaillance, une maintenance « as good as new » remet parfaitement à neuf ce matériel : le matériel est aussi performant que s'il était neuf. On obtient donc pour ce matériel un temps à la défaillance  $t_1$  (différence des dates  $D_1$  et Date de mise en service) et un temps censuré  $t_2^*$  (différence des dates Date de fin de collecte et de la nouvelle origine  $D_1$ ) car, à la date de fin de collecte, le matériel est encore en vie.

### 1.2.3. La maintenance corrective « As bad as old » (« aussi mauvais que vieux »)

Exemple

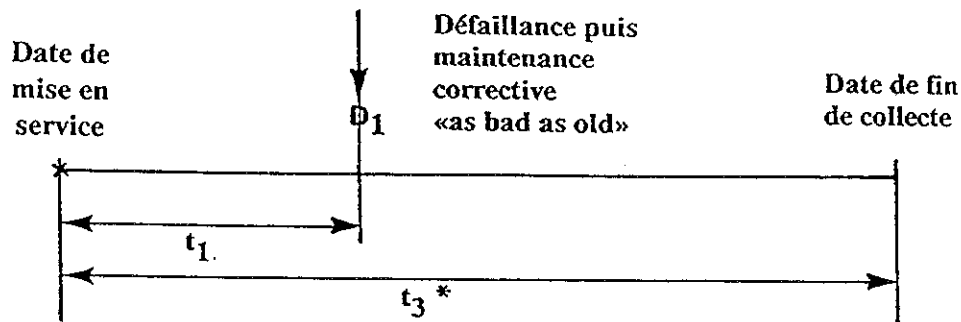


Figure 3. - Illustration de la maintenance « as bad as old ».

A la date  $D_1$ , le matériel a une défaillance. A la suite de cette défaillance, une maintenance « as bad as old » remet le matériel dans l'état qui précédait la défaillance : il continue à vieillir après sa réparation. On obtient donc pour ce matériel un temps à la défaillance  $t_1$  et un temps censuré  $t_3^*$  (différence des dates date de fin de collecte et date de mise en service, l'origine n'a pas été modifiée par la maintenance).

### 1.2.4. En bref

Les données nécessaires pour établir un échantillon sont donc :

- l'origine qui est, soit la Mise en Service Industriel (MSI) du matériel, soit la date de la dernière maintenance « as good as new »,
- les dates des défaillances ou les dates de maintenance corrective,
- les dates de censure, à savoir la date de fin de collecte ou de fin d'observation, et les dates des opérations de maintenance préventive.

Les banques de données de retour d'expérience doivent permettre l'extraction de ces données. La connaissance du fonctionnement du matériel est indispensable. L'incertitude liée au calcul des paramètres de fiabilité peut être en grande partie due à la non connaissance des temps de fonctionnement ou à leur estimation. Elle peut également être due aux hypothèses faites sur les conditions de réparation.

## 2. QUELQUES NOTIONS PRELIMINAIRES

### 2.1. Généralités

Considérons la variable aléatoire  $T$ , durée à la défaillance d'un dispositif. Cette durée  $T$  est positive, elle peut être un temps réel de fonctionnement ou une durée calendaire ou...

La fonction de répartition de cette variable est :

$$F(t) = \text{Prob} (T < t) \quad (1)$$

et la fonction fiabilité s'écrit :

$$R(t) = \text{Prob} (T \geq t) = 1 - F(t) \quad (2)$$

La densité de probabilité  $f(t)$  dérivée de la fonction  $F(t)$  est :

$$f(t) = \frac{dF(t)}{dt} = - \frac{dR(t)}{dt} \quad (3)$$

de telle sorte que :

$$\text{Prob} (t \leq T \leq t + dt) \sim f(t) \cdot dt$$

pour un intervalle de temps  $dt$ .

$f(t)$  exprime la probabilité de défaillance dans l'intervalle de temps  $dt$ .

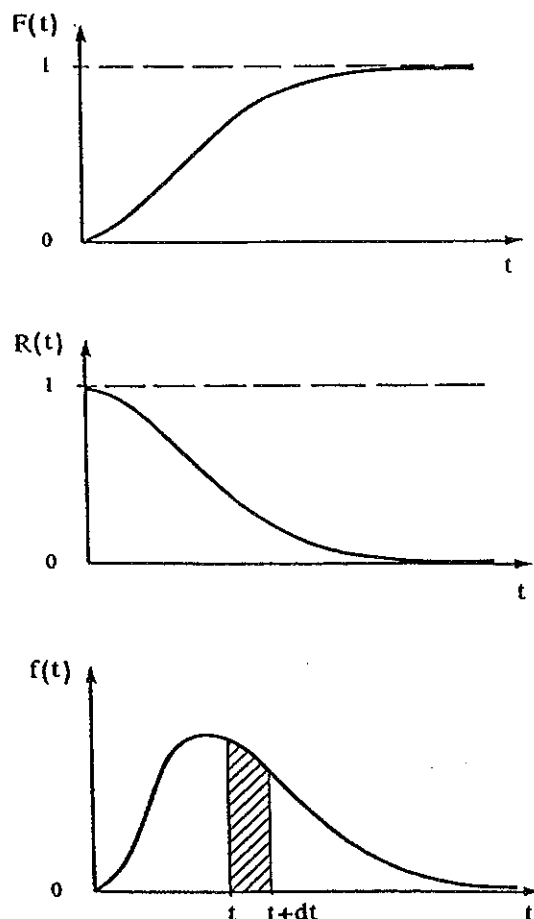


Figure 4. - Fonction de répartition, fonction de survie et densité de probabilité.

Si l'on désigne par  $\lambda(t)$  la probabilité conditionnelle de défaillance pendant  $dt$ , sachant que le matériel a vécu jusqu'à  $t$ , on a :

$$f(t) dt = R(t) \cdot \lambda(t) \cdot dt$$

soit  $\lambda(t)$ , taux de défaillance instantané,

$$\lambda(t) = \frac{f(t)}{R(t)} \quad (4)$$

et

$$\frac{dR(t)}{R(t)} = - \lambda(t) dt$$

$$R(t) = \exp \left( - \int_0^t \lambda(u) du \right) \quad (5)$$

$$R(t) = \exp (- H(t)) \quad (6)$$

où

$$H(t) = \int_0^t \lambda(u) du \quad (7)$$

s'appelle la fonction de hasard cumulé.

Notons que  $\lambda(t)$  est un indicateur de fiabilité, indiquant l'imminence d'une défaillance à l'instant  $t$ .

*Quelques remarques :*

-  $\lambda$  peut être décroissant (on dit : DFR « decreasing failure rate »), c'est l'amélioration du processus d'exploitation en début de vie des matériels; c'est très souvent le cas des composants électroniques,

-  $\lambda$  peut être constant, la fonction de survie est alors exponentielle; il n'y a pas de dégradation, c'est la période de vie utile, la partie plate de la courbe en baignoire,

-  $\lambda$  peut être croissant avec l'âge (on dit : IFR « increasing failure rate »), le matériel se dégrade.

## 2.2. Quelques lois de survie

Seules les lois les plus fréquemment utilisées dans le traitement du retour d'expérience sont évoquées.

### 2.2.1. La loi exponentielle

$$\begin{aligned}
 R(t) &= e^{-\lambda t} \\
 f(t) &= \lambda e^{-\lambda t} \\
 H(t) &= \lambda t
 \end{aligned}
 \tag{8}$$

$\lambda$  est le taux de défaillance, constant.

### 2.2.2. La loi de Weibull à deux paramètres

$$\begin{aligned}
 R(t) &= e^{-\left(\frac{t}{\eta}\right)^\beta} \\
 \lambda(t) &= \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} \\
 f(t) &= \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\eta}\right)^\beta\right] \\
 H(t) &= \left(\frac{t}{\eta}\right)^\beta
 \end{aligned}
 \tag{9}$$

-  $\beta$  est le paramètre de forme. C'est le paramètre le plus important de la loi de Weibull, car il joue sur la variation du taux de défaillance et permet ainsi de modéliser alternativement les trois phases de la courbe en baignoire.

- si  $\beta < 1$ , le taux instantané de défaillance décroît (DFR),
- si  $\beta = 1$ , le taux instantané de défaillance est constant, on retrouve la loi exponentielle,
- si  $\beta > 1$ , le taux instantané de défaillance croît (IFR),
- si  $\beta \rightarrow 3.6$ , la loi de Weibull a une allure voisine mais distincte de la loi normale.

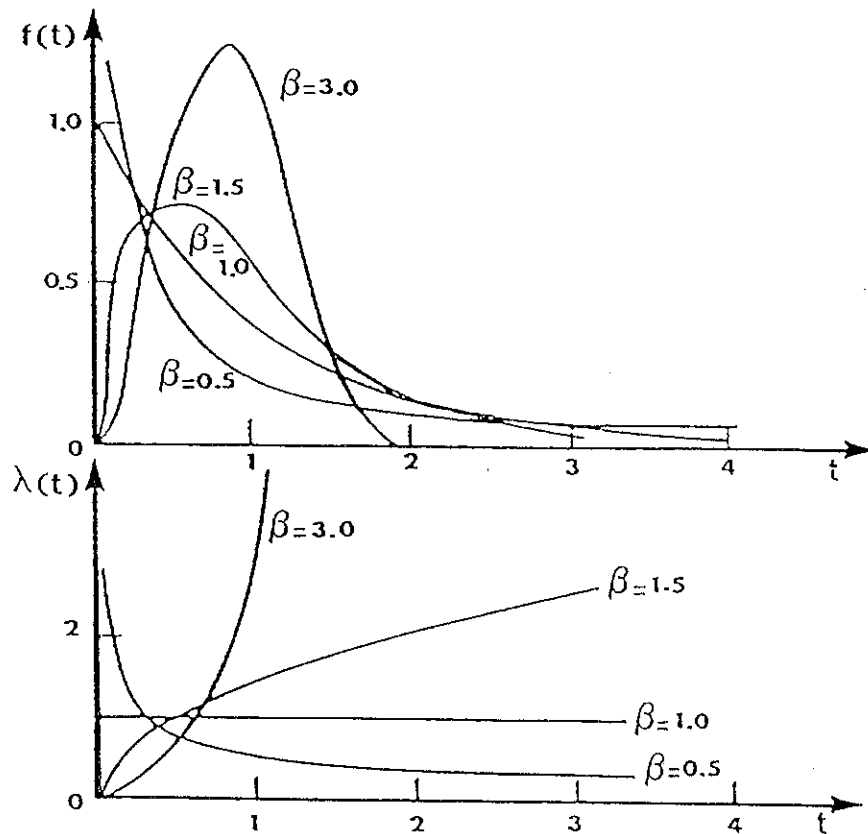


Figure 5. - Modèle de Weibull à deux paramètres ( $\eta = 1$ ) Influence du paramètre  $\beta$ .

-  $\eta$  est un simple paramètre d'échelle de temps appelé aussi durée de vie caractéristique. Lorsque  $t = \eta$ ,  $F(t) = 63\%$ .

### 2.2.3. La loi de Weibull à trois paramètres

$$R(t) = e^{-\left(\frac{t-\gamma}{\eta}\right)^\beta} \quad (10)$$

$\gamma$  est le paramètre de localisation, il correspond à une simple translation.

Le problème d'estimation est plus difficile, mais également celui de l'interprétation. On peut dire :

- si  $\gamma < 0$ , le matériel a subi une dégradation avant sa mise en service industriel. Dans ce cas, c'est un problème de qualité. Il y a eu usure lors des essais (essais trop longs), du rodage ou un problème de montage. Une autre interprétation consiste à dire que le gamma négatif n'existe pas, c'est un mélange de populations et par manque de données, on a l'impression que gamma est négatif,

- si  $\gamma > 0$ , le matériel ne commence à se dégrader qu'à la date  $t = \gamma$ , après une période d'incubation. Cela peut vouloir dire également qu'il y a un problème d'observabilité. Le matériel peut se dégrader dès  $t = 0$  mais cela n'a pas de conséquence apparente sur son bon fonctionnement.

#### 2.2.4. Une autre forme de la loi de Weibull à deux paramètres

Cette forme, plus pratique, est très utilisée dans les progiciels statistiques.

$$R(t) = \exp \left[ - \exp \left( \frac{\ln t - \mu}{\sigma} \right) \right] \quad (11)$$

On a donc :

$$\beta = \frac{1}{\sigma}$$

$$\eta = \exp(\mu)$$

#### 2.2.5. La distribution Gamma

Sa densité de probabilité s'écrit :

$$f(t) = \frac{\lambda^a t^{a-1} \exp(-\lambda t)}{\Gamma(a)} \quad (12)$$

$1/\lambda$  est le paramètre d'échelle et  $a$  est le paramètre de forme.

- si  $a = 1$ , on retrouve la loi exponentielle,
- si  $a < 1$ , c'est la période DFR,
- si  $a > 1$ , c'est la période IFR.

Si  $a$  est entier, alors  $R(t)$  s'écrit :

$$R(t) = e^{-\lambda t} \left\{ 1 + (\lambda t) + \frac{(\lambda t)^2}{2!} + \dots + \frac{(\lambda t)^{a-1}}{(a-1)!} \right\}$$

La distribution Gamma peut être considérée comme une généralisation de la loi exponentielle et une alternative à la distribution de Weibull. Comme cette dernière, la distribution Gamma peut représenter toutes les phases de vie d'un matériel.



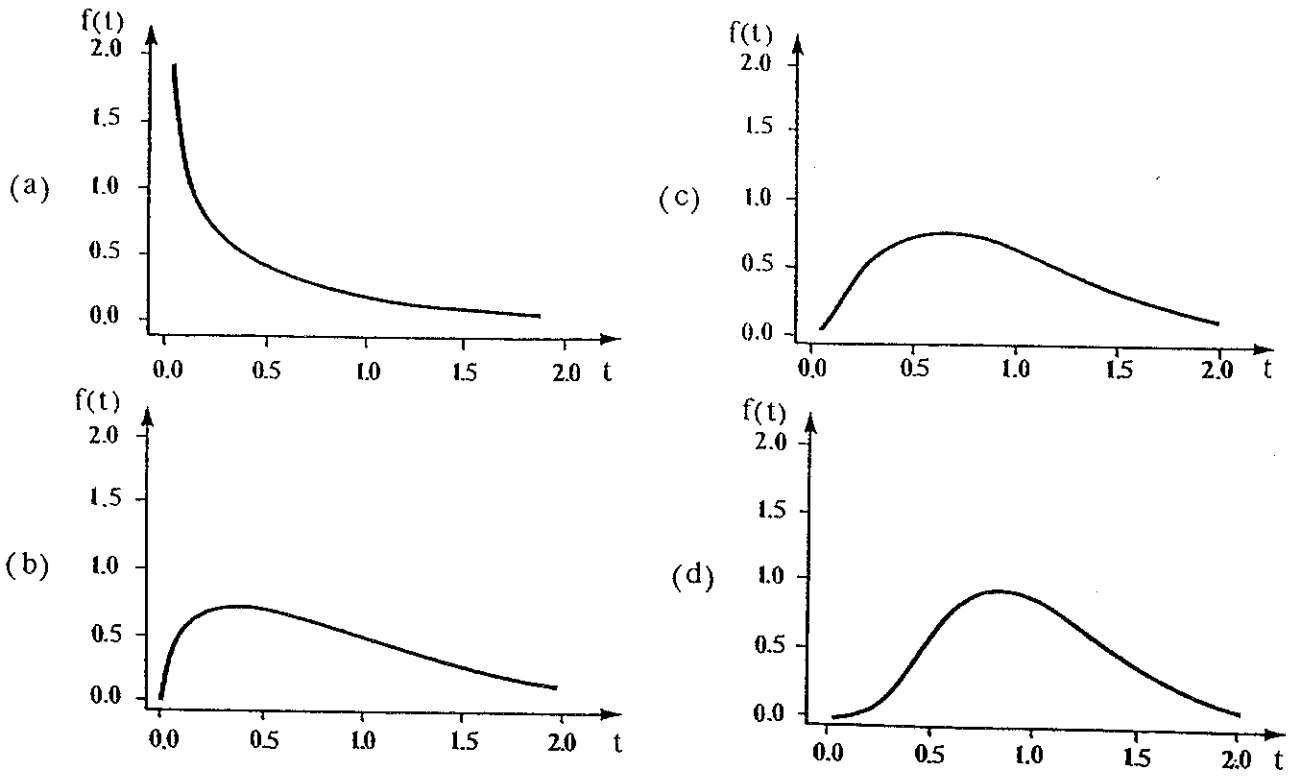


Figure 6. - Densités de probabilité ( $f(t)$ ) pour 4 distributions Gamma de moyenne  $a/\lambda = 1$  et (a)  $a = 0,5$  (b)  $a = 1,5$  (c)  $a = 2,5$  (d)  $a = 5$ .

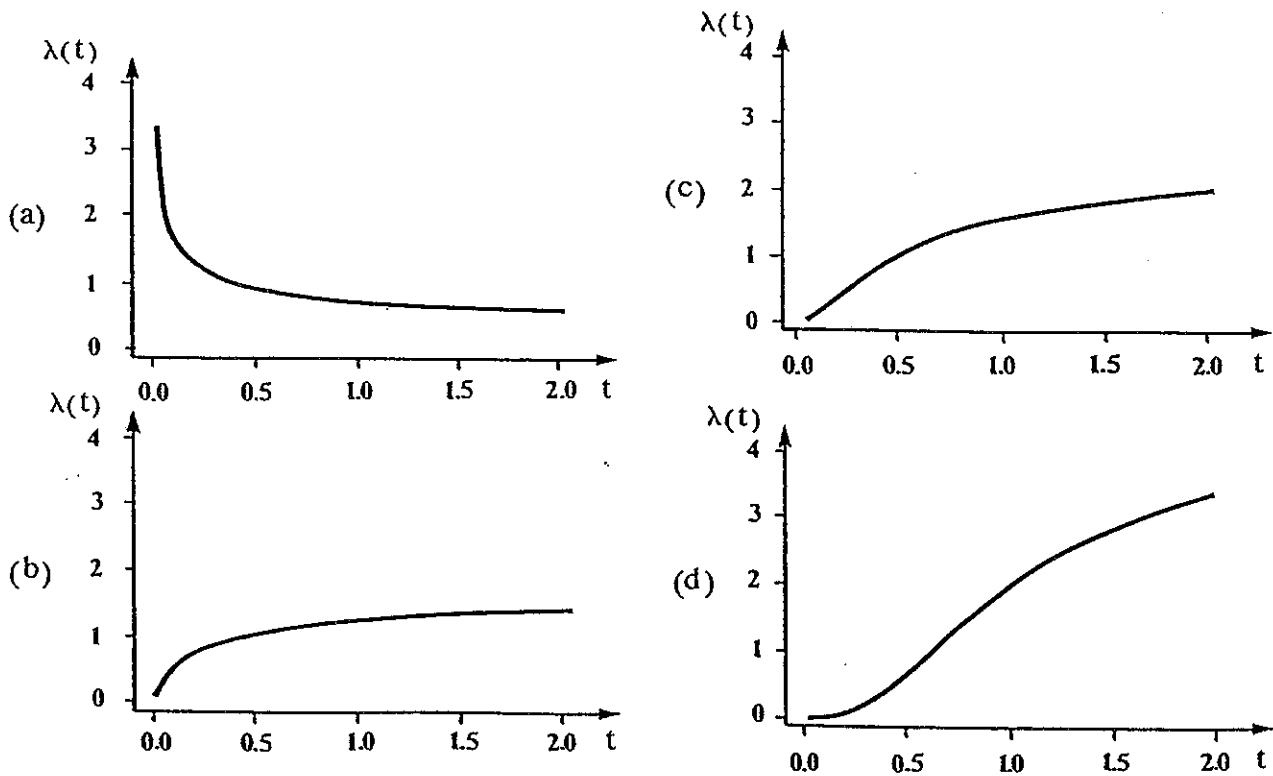


Figure 6 (suite). - Taux de défaillance pour 4 distributions Gamma de moyenne  $a/\lambda = 1$  et (a)  $a = 0,5$  (b)  $a = 1,5$  (c)  $a = 2,5$  (d)  $a = 5$ .

### 3. LES DIFFICULTES LIEES A LA DEFINITION DE LA DEFAILLANCE

Le premier paragraphe a permis de mettre en évidence les difficultés de détermination du temps de fonctionnement (le « dénominateur » de l'indicateur  $\lambda$ ). Il convient de signaler les difficultés inhérentes à la détermination du « numérateur » de l'indicateur  $\lambda$ .

Ces difficultés sont les suivantes :

- Qu'est qu'une défaillance ? Faut-il considérer la défaillance (aussi appelée défaillance complète), cessation d'un matériel à accomplir sa fonction requise ou faut-il plutôt considérer la dégradation (appelée aussi défaillance partielle), altération d'un matériel à accomplir cette fonction ?

- Comment la définir ? Comment apprécier son caractère « critique » ? Vis-à-vis de la sécurité, de la disponibilité, de la maintenance préventive, des coûts, de la dosimétrie, ...

- Quel est l'état du matériel : fonctionnement, sollicitation, attente ? Quel est son mode de défaillance ? ...

Ces difficultés sont nombreuses et on constate qu'on ne peut traiter des données de retour d'expérience sans élaborer une démarche difficile, fastidieuse, exigeant des compétences, bref structurée de validation et d'expertise des données de retour d'expérience.

Notons enfin que le traitement exige l'accès à plusieurs bases de données, de façon simultanée : les bases de défaillances et les historiques de maintenance, les données de fonctionnement et les statistiques d'exploitation.

### 4. LES METHODES NON PARAMETRIQUES

Si l'on dispose d'un échantillon non censuré  $(t_1; \dots; t_n)$  de défaillances, la fonction de répartition empirique  $F_n(t)$  est définie par :

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n I(t_i \leq t), \quad (13)$$

où  $I(A)$  désigne la fonction indicatrice de l'événement  $A$ .

Si maintenant l'échantillon  $((t_1; \dots; t_n)$  est censuré, on cherche à obtenir un équivalent de la fonction de répartition empirique.

En général, ce n'est pas à  $F$  mais à  $R = 1-F$ , la fonction de fiabilité que l'on s'intéresse et c'est elle que l'on cherche à estimer. Quatre méthodes sont exposées ci-après. Chacune présente un intérêt mais seul l'estimateur de Kaplan-Meier (appelé aussi « Product limit ») fait référence.

#### 4.1. La méthode de Wayne-Nelson (Lawless, 1982)

La méthode de Wayne-Nelson ne s'applique qu'à du matériel non réparable. Elle consiste à calculer l'évolution de la fonction  $H(t)$  de hasard cumulé.

Les événements (défaillances ou suppressions) sont classés chronologiquement.

Si  $N$  est l'effectif,  $n_{t_i}$  le nombre d'éléments défaillants à l'instant  $t_i$ , le taux de défaillance instantané vaut alors :

$$\lambda(t_i) = \frac{n_{t_i}}{N - \sum_{t_j < t_i} n_{t_j}} \quad (14)$$

$\sum_{t_j < t_i} n_{t_j}$  désigne l'ensemble des éléments défaillants ou censurés avant l'instant  $t_i$ .

Il est alors facile d'en déduire la fonction  $H(t)$  par :

$$H(t_i) = \sum_{t_j < t_i} \lambda(t_j) \quad (15)$$

Si l'on attribue un modèle  $H(t)$ , par exemple un modèle de Weibull, il est alors facile de déterminer, par régression, les paramètres de ce modèle.

#### 4.2. La méthode de « Johnson » ou des rangs médians

Cette méthode calcule l'ordre d'apparition des défaillances en donnant un certain incrément de l'ordre en fonction du nombre de données censurées entre deux défaillances.

Si  $I$  est l'incrément et  $N$  l'effectif de l'échantillon, on a :

$$I = \frac{N + 1 - \text{ordre de la défaillance précédente}}{1 + \text{Nombre d'individus ayant fonctionné plus longtemps que le dernier censuré}} \quad (16)$$

Si  $\theta_i$  est l'ordre d'apparition de la défaillance (ordre effectif + incrément), le rang médian s'écrit :

$$\text{Rang médian} = \frac{\theta_i - 0,3}{N + 0,4} \quad (17)$$

et la fiabilité est estimée par :

$$R_n(t_i) = 1 - \text{Rang médian} \quad (18)$$

*Exemple d'application de la méthode de Johnson*

Il s'agit du clapet principal des vannes de décharge au condenseur. Son comportement est observé sur 16 vannes pendant cinq années (31-12-1980 au 31-12-1985). L'échantillon est composé de 44 données dont 37 sont censurées, l'unité de temps est le jour calendaire.

Les informations ont été collectées en centrale sur les sites de Bugey et Dampierre.

La table 1. donne les temps collectés (temps de défaillance et temps censurés à droite)..

La table 2. utilise la méthode du rang médian et calcule la fiabilité R du matériel. Cette fiabilité est représentée sur la figure 7.

Table 1. - Données collectées (temps exprimés en jours calendaires)  
sur les sites de Bugey et Dampierre, pour le  
clapet principal des vannes de décharge au condenseur  
- Période 1980-1985

107 116* 116* 151* 151* 177* 177* 177* 183 284* 325* 325* 325* 345*
---

349 357* 365* 374* 374* 374* 406 406* 406* 406* 418* 418* 456 660*
--

745* 745* 757 931* 1273* 1273* 1276* 1336* 1336* 1336* 1976* 2057* 2148* 2299 2414*
---

Bilan : N = 44 (7 défaillants, 37 censurés)

\* = donnée censurée

Table 2. - Calcul du rang médian.

Rang de la défaillance	Nombre de jours (t)	Incrément I	$\theta_i$	$R_n(t)$
1 <sup>er</sup>	107	0	1	$\frac{1 - 0,3}{44 + 0,4} = 0,984$
2 <sup>e</sup>	183	$\frac{44 + 1 - 1}{1 + 36} = 1,189$	$1 + 1,189 = 2,189$	$1 - \frac{\theta_2 - 0,3}{44 + 0,4} = 0,957$
3 <sup>e</sup>	349	$\frac{44 + 1 - 2,189}{1 + 30} = 1,381$	3,570	0,926
4 <sup>e</sup>	406	$\frac{44 + 1 - 1,381}{1 + 23}$	5,387	0,885
5 <sup>e</sup>	456	$\frac{44 + 1 - 5,387}{1 + 17}$	7,588	0,836
6 <sup>e</sup>	757	$\frac{44 + 1 - 7,588}{1 + 13}$	10,260	0,776
7 <sup>e</sup>	2299	$\frac{44 + 1 - 10,260}{1 + 2}$	21,840	0,515

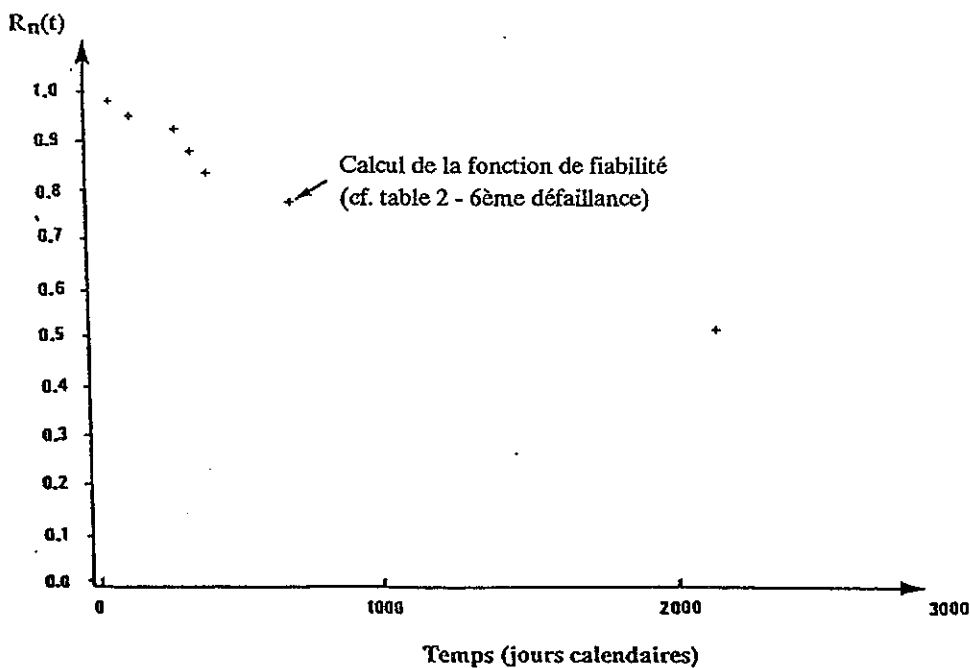


Figure 7. - Estimation de la fonction de survie - Méthode de Johnson  
Exemple du clapet principal des vannes de décharge au condenseur  
(7 défaillances, 37 censures).

#### 4.3. La méthode « Life Table » ou méthode actuarielle

Cette méthode est généralement développée dans les progiciels statistiques.

On désigne par :

$d_j$  le nombre de matériels défectueux à la date  $t_j$ ,

$w_j$  le nombre d'éprouvettes censurées dans l'intervalle de temps  $[t_{j-1} ; t_j]$ ,

$r_j$  le nombre d'éprouvettes encore en test à la date  $t_j$  (éprouvettes ni mortes ni censurées).

La fiabilité est alors calculée par la relation :

$$R_n(t_i) = \prod_{j=1}^i \left( 1 - \frac{d_j}{r_j - \frac{w_j}{2}} \right) \quad (19)$$

Cet estimateur tient compte des données censurées dans l'estimation de la proportion de défectueux dans l'intervalle  $[t_{j-1}, t_j]$ . On affecte le poids  $\frac{1}{2}$  à ces individus censurés.

#### 4.4. L'estimateur de Kaplan-Meier ou « Product Limit »

L'estimateur de Kaplan-Meier est considéré comme l'estimateur non paramétrique le plus direct de la fonction de survie. Il est l'unique estimateur cohérent de la fonction de survie..

Un estimateur  $R$  de la fonction de survie est dit cohérent si pour tout âge  $t$ , il vérifie :

$$R(t) = \frac{1}{n} \left[ \sum_{i=1}^n 1(t_i \geq t) + \sum_{i=1}^n 1(t_i^* < t) \frac{R(t)}{R(t_i)} \right] \quad (20)$$

$t_i^*$  : instant de censure

L'estimateur de Kaplan-Meier vérifie cette propriété.

Cette équation est constituée de la manière suivante : les survivants au-delà de l'instant  $t$  sont ceux qui n'ont été avant cette date ni morts, ni censurés, et dont l'effectif constitue la première somme dans le crochet, et d'autre part ceux qui, ayant été censurés à un instant  $t_i^*$  antérieur à  $t$ , survivent au-delà de  $t$  avec la probabilité conditionnelle

$\frac{R(t)}{R(t_i)}$  qui pondère chacun d'eux (Droesbeke et al, 1989).

Soit :

$d_j$  le nombre de défaillances au temps  $t_j$ ,

$r_j$  le nombre de dispositifs soumis à risque à  $t_j$ , c'est-à-dire ni morts ni censurés.

On a :

$$R(t_i) = \prod_{j=1}^i \left( 1 - \frac{d_j}{r_j} \right) \quad (21)$$

Table 3. - Méthode de l'estimateur de Kaplan-Meier  
Exemple d'application.

(7 défaillances, 37 censures)

Temps	Fonction de survie	Calcul réalisé
0.	1.0000	
107.	0.9773	1-1/44
116.*	.	
116.*	.	
151.*	.	
151.*	.	
177.*	.	
177.*	.	
177.*	.	
183.	0.9501	0,9773 (1-1/36)
284.*	.	
325.*	.	
325.*	.	
325.*	.	
345.*	.	
349.	0.9185	(0,9501) (1-1/30)
357.*	.	
365.*	.	
365.*	.	
374.*	.	
374.*	.	
374.*	.	
406.*	0.8785	0,9185 (1-1/23)
406.*	.	
406.*	.	
406.	.	
418.*	.	
418.*	.	
456.	0.8268	0,8785 (1-1/17)
660.*	.	
745.*	.	
745.*	.	
757.	0.7632	0,8268 (1-1/13)
931.*	.	
1273.*	.	
1273.*	.	
1276.*	.	
1336.*	.	
1336.*	.	
1336.*	.	
1976.*	.	
2057.*	.	
2148.*	.	
2299.	0.3816	0,7632 (1-1/2)
2414.*		

\* donnée censurée

*Exemple :*

Il s'agit du même exemple que celui utilisé pour l'estimation par la méthode de Johnson (cf. § 4.1). Un progiciel statistique a permis les calculs. La table 3. et la figure 8. présentent les résultats obtenus.

Un grand nombre des données de retour d'expérience traitées tout au long de ces paragraphes sont des données censurées à droite. Il faut donc faire un choix entre la méthode dite de « Johnson » et l'estimateur de Kaplan-Meier. Les deux estimateurs donnent des résultats comparables sur les  $(n-1)$  premières défaillances. Par contre, l'estimation de la fonction de survie à l'instant de la dernière défaillance est quelque peu différente. La méthode de Johnson a tendance à surestimer cette valeur.

Par sa structure non paramétrique indépendante du modèle, l'estimateur de Kaplan-Meier est devenu un outil classique et universel d'estimation (Droesbeke et al, 1989).

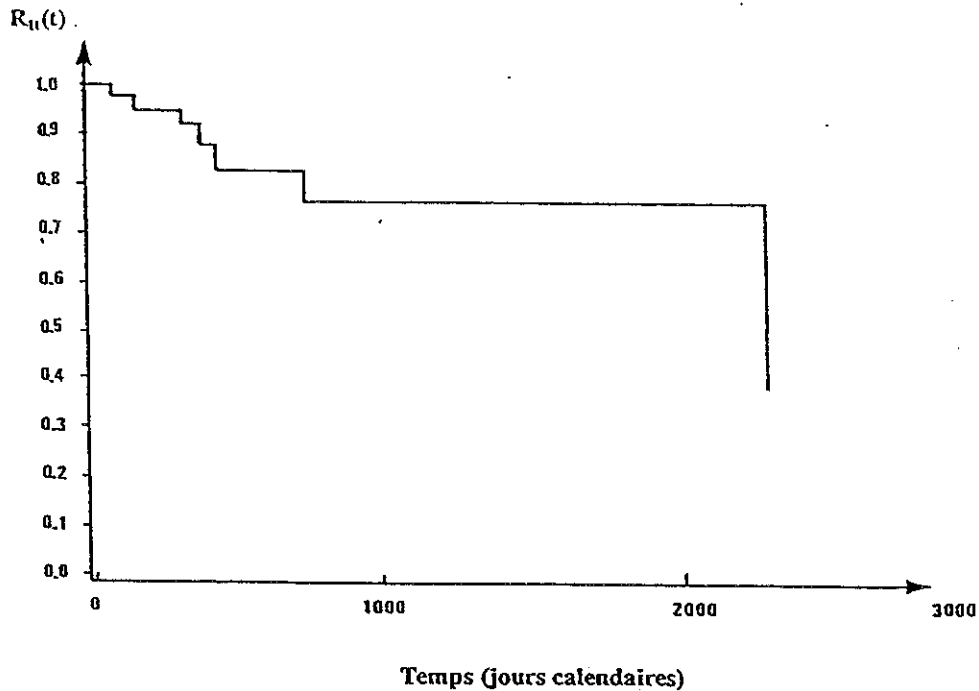


Figure 8. - Estimation de la fonction de survie - méthode de Kaplan-Meier  
(Exemple du clapet principal des vannes de décharge du condenseur, à comparer avec la figure 7).



## 5. LES METHODES PARAMETRIQUES

### 5.1. Le choix a priori d'un modèle

Le problème du choix a priori d'un modèle se pose, si l'on veut ajuster un modèle paramétrique sur des données de survie. A l'incertitude des données, on ajoutera l'incertitude due au choix du modèle.

La loi de Weibull est généralement choisie et ceci pour plusieurs raisons :

- elle décrit alternativement les trois phases de la vie d'un matériel, en fonction de son paramètre de forme  $\beta$  ; elle couvre la loi exponentielle ( $\beta = 1$ ) et approche la loi normale ( $\beta \sim 3,6$ ),

- elle se justifie par la théorie des valeurs extrêmes : la panne précoce d'un composant détermine le bon fonctionnement du matériel, c'est le maillon (l'élément) le plus faible qui détermine le bon fonctionnement d'un dispositif.

### 5.2. La méthode du maximum de vraisemblance

La méthode du maximum de vraisemblance est généralement préférée à d'autres méthodes car l'estimateur du maximum de vraisemblance est convergent, exhaustif, asymptotiquement normal et efficace.

#### 5.2.1. Généralités

Si nous avons un échantillon  $t_1, t_2, \dots, t_n$  de données de survie non censurées, et si  $\theta$  est le vecteur des paramètres à estimer,  $f(t, \theta)$  est la densité de probabilité, de forme connue mais de vecteur paramètre  $\theta$  inconnu, la vraisemblance s'écrit alors :

$$\mathcal{L}(\theta) = \prod_{i=1}^n f(t_i, \theta) \quad (22)$$

Dans le cas des données de retour d'expérience, où les données sont censurées à droite, comme nous l'avons vu plus haut, la vraisemblance des observations s'écrit :

$$\mathcal{L}(\theta) = \left\{ \prod_i f(t_i, \theta) \right\} \left\{ \prod_j R(t_j, \theta) \right\} \quad (23)$$

l'indice i	l'indice j
est relatif	est relatif
aux défaillances	aux données
	censurées à droite

Si les données avaient été censurées à gauche, la densité aurait été remplacée par la fonction de répartition.

Généralement, on pose :

$$\ell(\theta) = \ell_n \mathcal{L}(\theta)$$

et les estimateurs du maximum de vraisemblance, vecteur  $\hat{\theta}$ , sont ceux qui maximisent la vraisemblance, c'est-à-dire :

$$\frac{\partial \ell}{\partial \theta_j} = 0 \quad \text{pour } j = 1, 2, \dots, m \quad (24)$$

Cette estimation nécessite le recours à des algorithmes numériques comme ceux de Newton ou Newton-Raphson. Ces algorithmes sont généralement disponibles dans les bibliothèques scientifiques.

Ils permettent le calcul de la matrice  $(m, m)$  de variance-covariance dont le terme, évalué à  $\hat{\theta}$ , est :

$$-\frac{\partial^2 \ell}{\partial \theta_j \partial \theta_k} \quad \text{pour } j = 1, 2, \dots, m ; k = 1, 2, \dots, m, \quad (25)$$

ce qui permet le calcul des écarts types.

La méthode du maximum de vraisemblance présente de nombreux avantages :

- elle est générale ; d'autres méthodes simples existent mais on ne peut les généraliser,
- elle a la propriété d'invariance;
- elle est asymptotiquement normale : si toute la population est observée, le maximum de vraisemblance donnera la réponse exacte et l'estimateur sera distribué selon une loi normale (approximativement), pour de grands échantillons.

### 5.2.2. Application à la loi exponentielle

Nous supposons  $n$  observations  $t_1, t_2, \dots, t_n$  dont :

$r$  observations non censurées

$(n-r)$  observations censurées à droite.

On pose :  $x_i = \ell_n t_i$

On en déduit :

$$\ell(\lambda) = r \ell_n \lambda - \lambda \sum_{i=1}^n t_i$$

$$\frac{\partial \ell}{\partial \lambda} = \frac{r}{\lambda} - \sum_{i=1}^n t_i$$

d'où :

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^n t_i} \quad (26)$$

Le dénominateur est le TTT (« total time on test »).

On a :

$$-\frac{\partial^2 \ell}{\partial \lambda^2} = \frac{r}{\lambda^2}$$

d'où l'on déduit l'écart type sur  $\hat{\lambda}$  égal à :

$$\frac{\hat{\lambda}}{\sqrt{r}}$$

$r$  doit donc être différent de 0, il convient donc qu'une durée de vie au moins soit non censurée.

Si toutes les observations sont non censurées et si les  $r$  plus petites durées de vie ont été observées et classées  $t_1 < t_2 < \dots < t_r$ , alors :

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^r t_i + (n-r) t_r} \quad (27)$$

### 5.2.3. Application à la loi de Weibull à deux paramètres $\beta$ et $\eta$

La vraisemblance s'écrit :

$$\ell(\beta, \eta) = r \ln \beta - r\beta \ln \eta + (\beta-1) \sum_{\text{défaillances}} \ln t_i - \eta^{-\beta} \sum_{i=1}^n t_i^{\beta} \quad (28)$$

Si l'on utilise la notation habituellement utilisée (cf. paragraphe 2.2.4) dans les progiciels statistiques pour la loi de Weibull à deux paramètres,  $\mu$  et  $\sigma$ , on obtient :

$$\ell(\mu, \sigma) = -r \ln \sigma + \sum_{\text{défaillances}} \frac{x_i}{\sigma} - \frac{r\mu}{\sigma} - \sum_{i=1}^n \exp\left(\frac{x_i - \mu}{\sigma}\right)$$

d'où :

$$\sigma \frac{\partial \ell}{\partial \mu} = -r + \sum_{i=1}^n \exp\left(\frac{x_i - \mu}{\sigma}\right)$$

et

$$\sigma^2 \frac{\partial \ell}{\partial \sigma} = -r\sigma - \sum_{\text{défaillances}} x_i + r\mu + \sum_{i=1}^n \exp\left(\frac{x_i - \mu}{\sigma}\right) \cdot (x_i - \mu)$$

On en déduit :

$$\hat{\mu} = \hat{\sigma} \ln \left[ \frac{1}{r} \sum_{i=1}^n \exp\left(\frac{x_i}{\hat{\sigma}}\right) \right] \quad (29)$$

$$\frac{1}{r} \sum_{\text{défaillances}} x_i + \hat{\sigma} - \sum_{i=1}^n x_i \exp \left( \frac{x_i}{\hat{\sigma}} \right) \sum_{i=1}^n \exp \left( \frac{x_i}{\hat{\sigma}} \right) = 0$$

Ces deux équations implicites et non linéaires sont résolues par des algorithmes numériques.

Notons qu'en  $\hat{\mu}, \hat{\sigma}$ , les dérivées secondes peuvent être calculées :

$$-\frac{\partial^2 \ell}{\partial \mu^2} = \frac{r}{\hat{\sigma}^2}$$

$$-\frac{\partial^2 \ell}{\partial \mu \partial \sigma} = \sum_{i=1}^n \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) \exp \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) \quad (30)$$

$$-\frac{\partial^2 \ell}{\partial \sigma^2} = r + \sum_{i=1}^n \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) \exp \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \right)$$

Signalons qu'une variante existe pour cette méthode. Cette variante donne de bons résultats, même si l'échantillon est faible et si les censures sont importantes. Partant d'une estimation a priori  $\hat{\beta}_i$  et  $\hat{\eta}_i$ , pour les paramètres d'une loi de Weibull, on simule les observations censurées suivant la loi de Weibull ayant pour paramètres  $\hat{\beta}_i$  et  $\hat{\eta}_i$ , c'est-à-dire qu'on complète l'échantillon par tirage d'une réalisation pour chaque censure existante, au delà du temps de censure. On calcule ensuite les estimateurs  $\hat{\beta}_{i+1}$  et  $\hat{\eta}_{i+1}$ , estimateurs du maximum de vraisemblance de l'échantillon ainsi complété. L'arrêt de la procédure se fait lorsque chaque paramètre tend vers une loi stationnaire. Ce principe est un peu analogue à celui du bootstrap, décrit au paragraphe 7.

#### 5.2.4. Exemples d'application

Ces exemples concernent les méthodes décrites dans les paragraphes 4 et 5.

a) Les trois méthodes Wayne-Nelson, Johnson et maximum de vraisemblance sont appliquées à l'ensemble des défaillances et dégradations observées sur la pivoterie du multiplicateur d'une pompe (75 observations dont 53 censures à droite). Les résultats sont donnés sur la figure 9. Insistons sur le fait que l'application de ces méthodes doit toujours être précédée et suivie d'une interprétation physique par l'expert du matériel. Elle nécessite de plus de disposer d'un ensemble de données suffisant, de bonne qualité et validé par l'expert. Le calcul de la loi de fiabilité permet l'optimisation des intervalles de temps entre opérations de maintenance préventive.

b) Sur la figure 10., la méthode du maximum de vraisemblance est comparée à l'estimateur de Kaplan-Meier pour le sous-ensemble fonctionnel «confinement eau» d'une pompe (4 défaillances et 52 données censurées à droite ont été prises en compte). On constate un bon accord entre les deux méthodes.

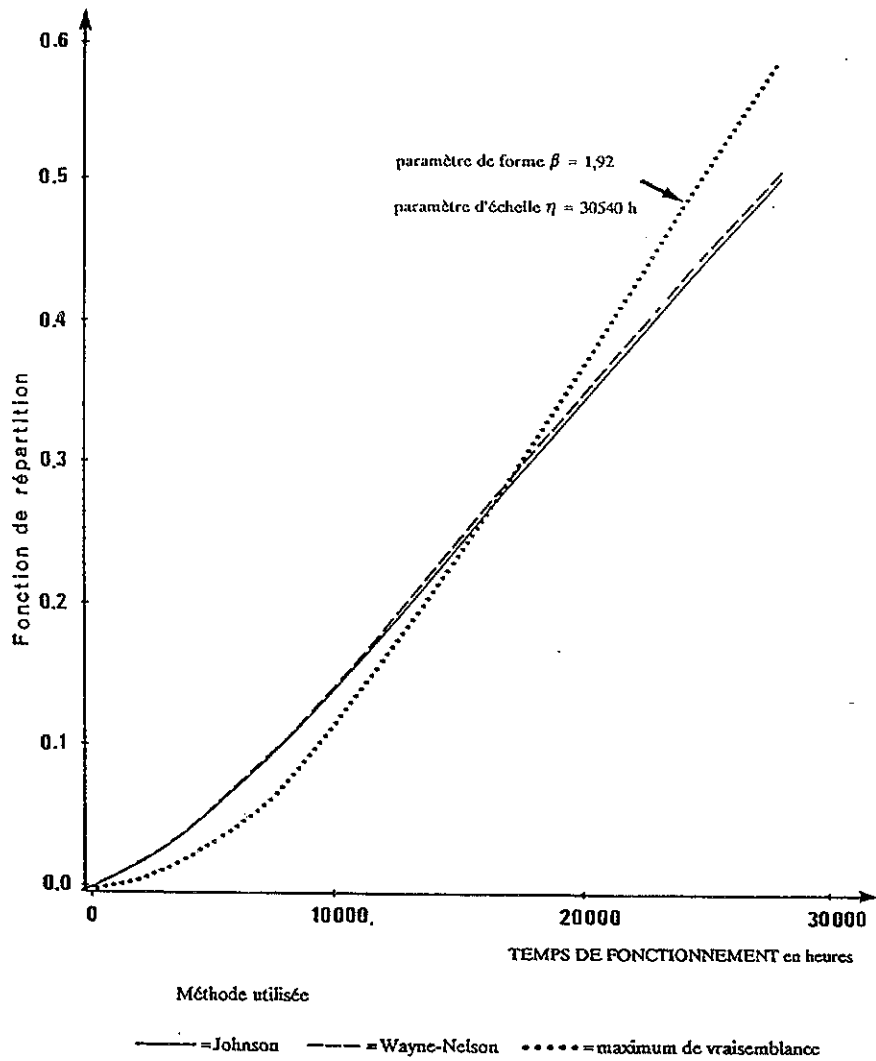


Figure 9. - Calcul de l'évolution de la fiabilité de la pivoterie du multiplicateur d'une pompe du système de contrôle volumétrique et chimique, pour un ensemble de défaillances et de dégradations. Comparaison de trois méthodes.

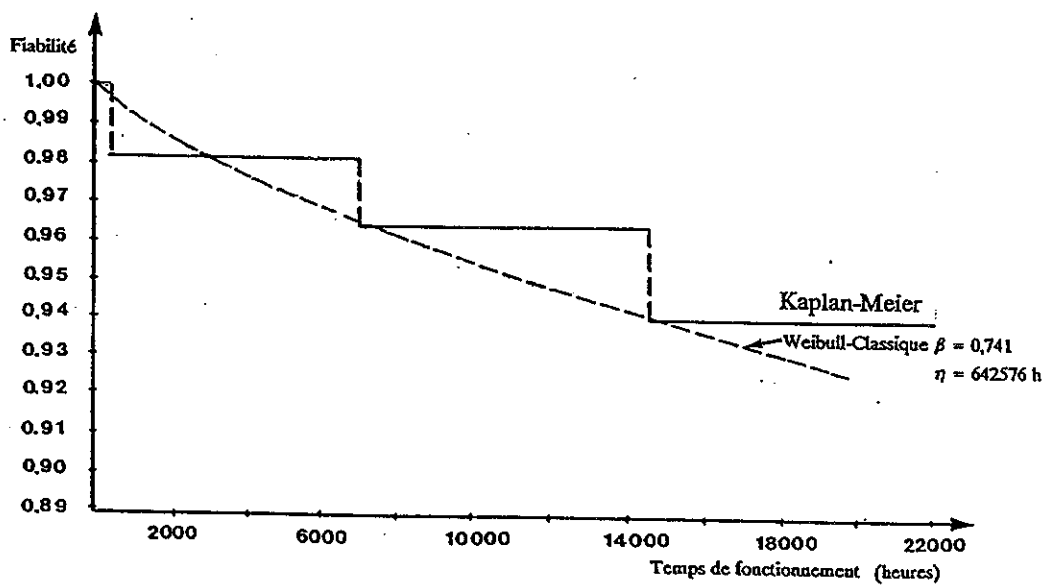


Figure 10. - Comparaison des estimations obtenues par la méthode du maximum de vraisemblance et l'estimateur de Kaplan-Meier. Sous-ensemble fonctionnel « confinement eau » des pompes du système de contrôle volumétrique et chimique.

## 6. LES INTERVALLES DE CONFIANCE

Une fois les paramètres des modèles de fiabilité estimés, il est d'usage de calculer des intervalles de confiance. Les algorithmes de la méthode du maximum de vraisemblance sont particulièrement utiles.

### 6.1. Le cas de la loi exponentielle

Rappelons que la loi exponentielle est très utilisée en fiabilité et qu'elle correspond à la partie plate de la courbe en baignoire où le taux  $\lambda$  est constant. Ceci correspond au processus de Poisson qui exprime la fréquence d'un nombre d'événements en un temps donné, lorsqu'à tout instant la probabilité d'occurrence d'un événement est la même.

L'intégration de la densité de probabilité de la loi du  $\chi^2$  donne la fonction de répartition de la loi de Poisson, ce qui permet de calculer la valeur  $\lambda_{sup}$  de  $\lambda$  telle que la probabilité d'avoir au plus  $k$  événements soit égale à une valeur  $\alpha_1$  et de même la valeur  $\lambda_{inf}$  de  $\lambda$  telle que la probabilité d'avoir au moins  $k$  événements soit égale à une valeur faible  $\alpha_2$ .

La valeur moyenne est donnée par  $\lambda = k/T$ ,  $T$  est la taille de l'échantillon observé, et l'intervalle de confiance sur  $\lambda$  a pour limites, au niveau  $1-\alpha_1-\alpha_2$  :

$$\lambda_{inf} = \frac{\chi^2_{\alpha_2} (2k)}{2T} \quad \text{et} \quad \lambda_{sup} = \frac{\chi^2_{1-\alpha_1} (2k+2)}{2T} \quad (31)$$

Le niveau de confiance  $1-\alpha_1-\alpha_2$  signifie que l'intervalle limité par les deux variables  $\lambda_{inf}$  et  $\lambda_{sup}$  a la probabilité  $1-\alpha_1-\alpha_2$  d'encadrer effectivement la vraie valeur, fixe, inconnue de  $\lambda$ .

Le facteur d'erreur vaut :

$$\text{facteur d'erreur} = \max \left( \frac{\lambda}{\lambda_{inf}}, \frac{\lambda_{sup}}{\lambda} \right) \quad (32)$$

La table 4. donne les intervalles de confiance recommandés. La figure 11. donne les intervalles de confiance en fonction du nombre de défaillances observées pour différents niveaux de confiance.

La table 5. donne quelques exemples extraits d'un recueil de données concernant des pompes, pour des défaillances considérées critiques après expertise.

Table 5. - Quelques résultats concernant les défaillances critiques au sens de la sûreté.  
(intervalle de confiance bilatéral de 90 %)

Matériel	Cumul des défaillances en fonctionnement	Temps cumulé de fonctionnement (10 <sup>6</sup> h)	$\lambda$ 10 <sup>-6</sup> /h		
			Min	Moyenne	Max
Pompe de circulation	6	1,5	1,7	4,0	7,9
Pompe de réfrigération intermédiaire	14	1,1	7,7	12,7	19,9
Pompe de charge	11	3,9	1,6	2,8	4,7

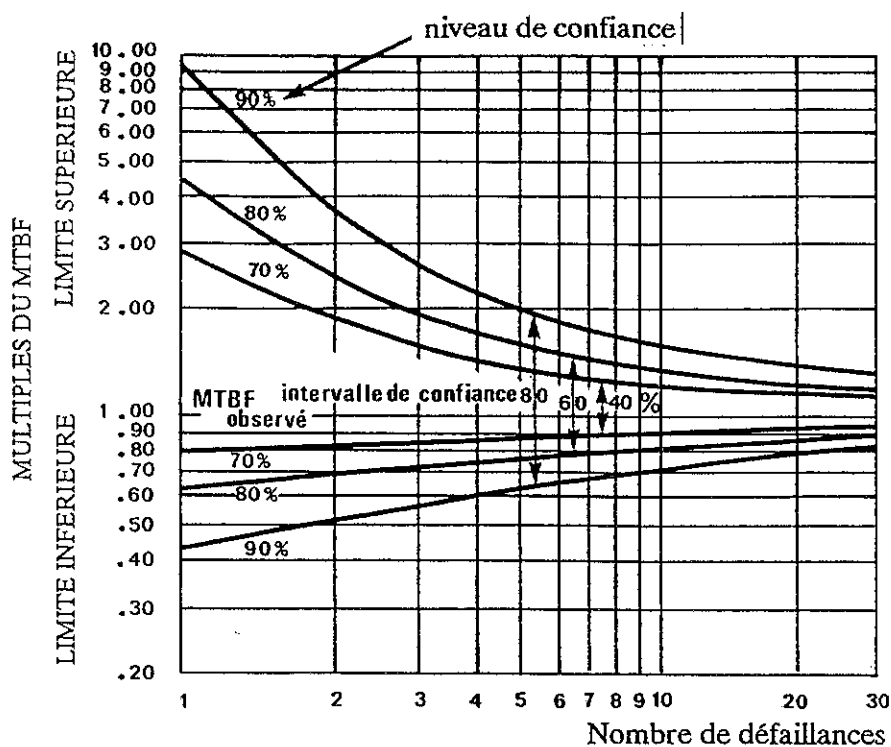


Figure 11. - Intervalles de confiance pour les MTBF -  
Hypothèses : taux de défaillance constant, loi exponentielle,  
calcul réalisé sur la base de la distribution du  $\chi^2$   
(MTBF =  $1/\lambda$ ).

Table 4. - Intervalles de confiance recommandés.

Effectif de défaillances	Valeur minimale	Valeur maximale
0	-	80 %
< 5	20 %	80 %
5 < effectif ≤ 10	10 %	90 %
> 10	5 %	95 %

### Remarques sur l'intervalle de confiance :

1 - Lorsqu'il n'y a pas de défaillance observée ( $k = 0$ ), les taux de défaillance sont calculés par une loi du Khi-deux pour un taux de confiance de 50 % selon la relation :

$$\text{valeur médiane} = \frac{\chi^2_{50\%}(2)}{2T} = \frac{0,7}{T} \quad (33)$$

La moyenne est prise égale à la médiane.

2 - Cette notion d'intervalle de confiance est très critiquable. En effet, si  $g(\lambda)$  est la densité de probabilité du taux  $\lambda$ , l'intervalle de confiance est défini par :

$$\int_{\lambda_{\text{inf}}}^{\lambda_{\text{sup}}} g(\lambda) \cdot d\lambda = \alpha,$$

qui exprime que la probabilité qu'un estimateur  $\lambda^*$  de  $\lambda$  tombe en dehors des limites  $[\lambda_{\text{inf}}, \lambda_{\text{sup}}]$  est égale à  $1-\alpha$ . Par conséquent, les valeurs  $\lambda_{\text{inf}}$  et  $\lambda_{\text{sup}}$  ne peuvent être uniques, elles dépendent de la forme de la distribution de  $\lambda$  et en particulier des queues de cette distribution. On voit donc que pour déterminer l'intervalle de confiance, il faut disposer de la densité  $g(\lambda)$  et donc d'un nombre suffisant de déterminations de  $\lambda$ . Lorsque l'on détermine un nouveau  $\lambda$  (c'est-à-dire un nouveau taux de défaillance d'un matériel critique), on peut estimer à quel quantile ce nouveau  $\lambda$  correspond. La pratique correspond en réalité à ce qui est écrit au début de ce paragraphe : on effectue une mesure du taux  $\lambda$  de laquelle on déduit l'intervalle de confiance, ce qui, clairement, n'est pas très rigoureux. Il apparaît par conséquent important de disposer de méthodes permettant de déterminer les lois de distribution des paramètres des lois de fiabilité.

### 6.2. La méthode de la normalité asymptotique

La méthode du maximum de vraisemblance possède des propriétés asymptotiques intéressantes. Asymptotiquement (quand  $n \rightarrow \infty$ ), l'estimation des paramètres est non biaisée (l'espérance de l'estimateur est égale au paramètre estimé), efficace (la variance de l'estimateur est faible) et surtout de distribution d'échantillonnage normale. Par conséquent, il est fort simple de construire un intervalle de confiance sur un paramètre estimé par la méthode du maximum de vraisemblance (Chevalier, 1991).

#### Application

Cet exemple est extrait d'une étude sur les presse-étoupe des vannes d'injection de sécurité des centrales de Fessenheim et de Bugey, à partir des données extraites d'une banque de données de défaillances, le S.R.D.F.



La table 6. précise l'échantillon utilisé dans cet exemple.

Table 6. - Caractéristiques des deux échantillons.

	Echantillon 1	Echantillon 2
Type de données	Défaillances + données censurées à droite, dues à la fin d'observation	Défaillances + données de maintenance préventive + données censurées à droite, dues à la fin d'observation
Population mère	9 vannes RIS VP suivies depuis la MSI jusqu'au 31.12.86  6 tranches 900 MW	5 tranches 900 MW (1 tranche a été écartée)
Données traitées	Censurées : 56 Défaillances : 42 Total : 98	Censurées : 330 Défaillances : 37 Total : 367

L'échantillon 2 inclut toutes les opérations de maintenance préventive réalisées sur ces vannes. Ceci explique le nombre important de données censurées à droite.

Les paramètres estimés des modèles de Weibull sont donnés dans les tables 7. et 8.

Table 7. - Paramètres de la loi de Weibull estimés à partir de l'échantillon 1 du tableau.

Paramètres (notation du paragraphe 2.2.4.)				Paramètres (notation classique)		
Para-mètres	Valeur estimée	Ecart type	Intervalle de confiance à 95 %	Para-mètre	Valeur estimée	Intervalle de confiance à 95 %
$\mu$ (mois)	4,775	0,171	[4,44 ; 5,11]	$\eta$ (jours)	3607	[2577 ; 5037]
$\sigma$	0,971	0,131	[0,714 ; 1,228]	$\beta$	1,03	[0,81 ; 1,40]

L'intervalle de confiance de  $\beta$  est par exemple calculé ainsi. Comme  $\sigma = \beta^{-1}$ , l'intervalle de confiance asymptotique à 95 % sur le paramètre  $\beta$  s'écrit pour les valeurs de la table 7. :

$$\text{De } I_{c\sigma} = [a, b] \quad \text{on obtient } I_{c\beta} = \left[ \frac{1}{b} ; \frac{1}{a} \right]$$

Si  $x$  suit une loi normale réduite, on a :

$$P(-1,96 \leq x \leq 1,96) = 0,95$$

donc :

$$P (m - 1,96 \sigma \leq x \leq m + 1,96 \sigma) = 0,95$$

et :

$$Ic \beta_{95} \% = \left[ \frac{1}{0,971 + 1,96 \times 0,131} ; \frac{1}{0,971 - 1,96 \times 0,131} \right]$$

Soit :

$$Ic \beta_{95} \% = [0,81 ; 1,40]$$

Table 8. - Paramètres de la loi de Weibull estimés à partir de l'échantillon 2 du tableau.

Paramètres (notation du paragraphe 2.2.4.)				Paramètres (notation classique)		
Para- mètres	Valeur estimée	Ecart type	Intervalle de confiance à 95 %	Para- mètre	Valeur estimée	Intervalle de confiance à 95 %
$\mu$ (mois)	3,558	0,157	[3,250 ; 3,866]	$\eta$ (jours)	1065	[784 ; 1452]
$\sigma$	0,447	0,0657	[3,319 ; 0,576]	$\beta$	2,24	[1,74 ; 3,13]

### 6.3. La méthode du rapport des vraisemblances

$f(t, \theta)$  est la densité de probabilité dont on cherche à estimer le vecteur  $\theta$  de ses paramètres.

$\theta_0$  est le vecteur de paramètres qui coïncide avec le maximum de la fonction de vraisemblance. Soit  $\Lambda$  la statistique :

$$\Lambda = - 2 \ln \frac{\mathcal{L}(\theta_0)}{\mathcal{L}(\hat{\theta})} = - 2 (\ell(\theta_0) - \ell(\hat{\theta})) \quad (34)$$

où  $\hat{\theta}$  est une estimation de  $\theta$ .

$\Lambda$  suit asymptotiquement une loi du  $\chi^2(p)$  à  $p$  degrés de liberté,  $p$  étant la dimension du vecteur  $\theta$  à estimer.

Par exemple, pour une loi de Weibull à deux paramètres, on peut estimer les paramètres  $\beta_0$  et  $\eta_0$  qui maximisent la fonction de vraisemblance et on souhaite construire un intervalle de confiance sur le paramètre  $\beta$  de la loi de Weibull, au niveau de confiance  $(1-\alpha)$ . Il suffit alors de résoudre :

$$- 2 \ln \frac{\mathcal{L}(\beta_0)}{\mathcal{L} \left[ \begin{array}{c} \beta_{inf} \\ \eta_0 \end{array} \right]} = \chi^2_{\alpha}(1) \quad (35)$$

et

$$- 2 \ln \frac{\mathcal{L}(\beta_0)}{\mathcal{L} \left[ \begin{array}{c} \beta_{sup} \\ \eta_0 \end{array} \right]} = \chi^2_{\alpha}(1)$$

le degré de liberté étant 1, car seul  $\beta_0$  est à estimer, si  $\eta_0$  est fixé.

### Application

L'évolution de la fiabilité des tiges des vannes de décharge au condenseur d'une centrale a été recherchée. 16 vannes ont été observées pendant 5 années : 25 données ont été relevées dont 20 sont censurées. L'application de la méthode du maximum de vraisemblance a permis d'estimer :

- le logarithme du maximum de vraisemblance = - 13,507,
- le paramètre  $\sigma_0 = 0,601$  (d'où  $\beta = 1,66$ ),
- le paramètre  $\mu_0 = 8,099$ .

On souhaite construire un intervalle de confiance à 95 %.

L'utilisation de la méthode du rapport des vraisemblances permet la construction de la figure 12.

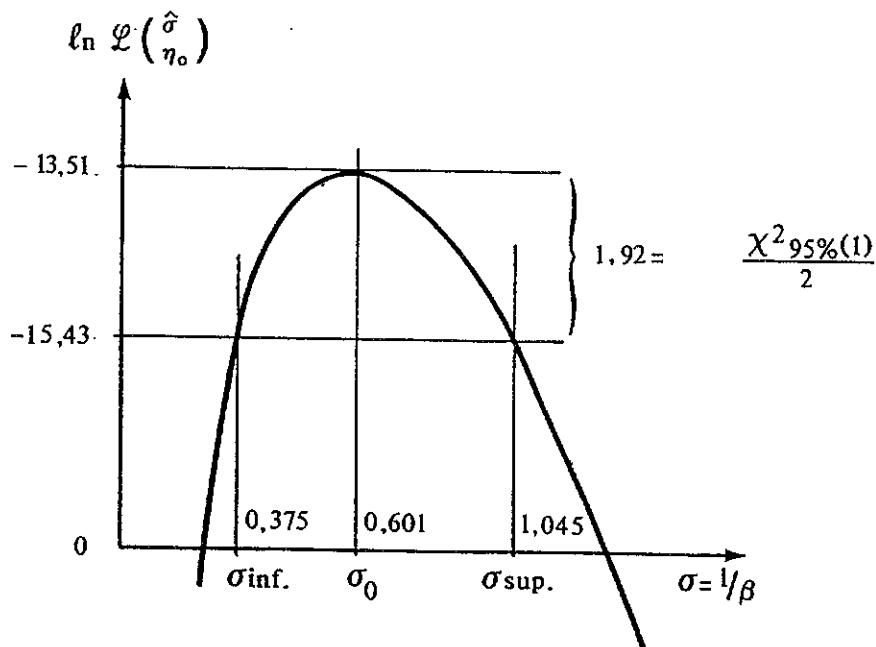


Figure 12. - Profil du logarithme de la fonction de vraisemblance d'une loi de Weibull pour  $\eta_0$  fixé ( $\mu_0$  fixé).

D'où l'on déduit l'intervalle de confiance à 95 % sur  $\beta$  :

$$I_{c\beta} 95 \% = \left[ \frac{1}{1,045} ; \frac{1}{0,375} \right] = [0,96 ; 2,67]$$

## 7. LA DISTRIBUTION DES PARAMETRES DE FIABILITE - LA METHODE DU BOOTSTRAP

Cette méthode permet non seulement de construire un intervalle de confiance sur les paramètres d'une loi mais aussi d'obtenir leur distribution.

Cette méthode du bootstrap, introduite par B. Efron (en 1979) a fortement bénéficié de la capacité de plus en plus grande des moyens de calcul, et fait partie des méthodes utilisant des calculs statistiques intensifs sur ordinateur. Le bootstrap apparaît comme la plus ingénieuse, la mieux fondée théoriquement, et potentiellement la plus féconde, des méthodes d'estimation d'échantillons (Chevalier, Deheuvels, 1991).

### 7.1. Le principe du Bootstrap

Le principe du bootstrap consiste à créer, à partir de l'échantillon original, un grand nombre d'échantillons par tirage aléatoire. Sur chacun de ces échantillons, le (ou les) paramètre(s) sont calculés permettant d'approcher au mieux leur variance ou leur loi.

#### 7.1.1. L'algorithme du Bootstrap

L'algorithme du bootstrap peut être résumé ainsi.

Soit une variable aléatoire  $T$  (par exemple le temps à la défaillance) de fonction de répartition  $F$  dépendant d'un vecteur paramètre  $\theta$ . On possède un échantillon  $\varepsilon = (t_1, \dots, t_n)$ .

##### Phase 1 :

$\varepsilon$  sert de population de base.  $\varepsilon$  est muni de la fonction de répartition empirique  $F_n$ .

##### Phase 2 :

On procède dans  $\varepsilon$  à  $N$  tirages équiprobables avec remise

$\varepsilon^* = (t_1^*, \dots, t_N^*)$  est l'échantillon « bootstrappé » ainsi obtenu, avec

$$\forall i, 1 \leq i \leq N \quad \exists j, 1 \leq j \leq n / t_i^* = t_j$$

En pratique, on prendra souvent  $N = n$ . On pourrait tout aussi bien remplacer  $N$  par  $mN$  avec  $m$  fixé. Pour des raisons de convergence, il convient que le rapport  $n/N$  soit borné supérieurement et inférieurement par des nombres positifs fixés.

##### Phase 3 :

On approche le comportement de  $\theta(\varepsilon, F)$  (c'est-à-dire la valeur de  $\theta$  que l'on aurait pu obtenir à partir de l'échantillon  $\varepsilon$ ) issu directement de  $\varepsilon$  et  $F$  par celui de  $\theta(\varepsilon^*, F_n) = \theta^*$  issu de  $\varepsilon^*$  et  $F_n$ .

Cette dernière phase sera itérée B fois pour donner lieu à une approximation par la méthode de Monte Carlo. Dans ce cas, l'échantillon  $\varepsilon$  engendre B échantillons  $\varepsilon_k^*$ ,  $k = 1$  à B avec  $\varepsilon_k^* = (t_{1k}^*, \dots, t_{Nk}^*)$ . On observe donc B valeurs de  $\theta_k$  ( $\varepsilon_k^*$ ,  $F_n$ ) (qui serviront pour la construction de l'intervalle de confiance).

### 7.1.2. Le cas particulier des données censurées

Le bootstrap n'est pas affecté par les données censurées et en particulier la censure à droite, typique des données de retour d'expérience.

La variable aléatoire T dont on possède un échantillon  $\varepsilon = (t_1, \dots, t_n)$  n'est pas modifiée. Mais l'échantillon  $\varepsilon$  pour des données censurées à droite s'écrit :

$$\varepsilon = \{(t_1, d_1); (t_2, d_2); \dots; (t_n, d_n)\}$$

$$d_j = \begin{cases} 0 & \text{si } t_j \text{ n'est pas censuré} \\ 1 & \text{si } t_j \text{ est censuré} \end{cases}$$

La justification du bootstrap d'Efron n'en est donc pas modifiée. La seule différence est l'ensemble sur lequel F est définie, soit  $\mathbb{R} \times \{0,1\}$ .

### 7.1.3. Intervalle de confiance

La méthode la plus classique pour construire un intervalle de confiance basé sur le bootstrap serait :

$\hat{\theta} \pm c \sigma_{\text{Boot}}$  où c est une constante issue des tables de la loi normale ou de Student suivant que B est grand ou petit et où  $\sigma_{\text{Boot}}$  est l'écart type calculé sur les  $\theta_k^*$ .

Efron propose une méthode simple et plus ambitieuse pour construire un intervalle de confiance basé sur la distribution du bootstrap de  $\theta^*$ . Elle permet en particulier l'obtention d'un intervalle de confiance asymétrique particulièrement pertinent dans le cas des petits échantillons.

$\hat{\theta}_{\text{inf}}$  est la borne inférieure de l'intervalle de confiance à  $(1-2\alpha)\%$ , et  $\hat{\theta}_{\text{sup}}$  la borne supérieure de l'intervalle de confiance à  $(1-2\alpha)\%$ . On a :

$$\hat{\theta}_{\text{sup}} = \text{CDF}^{-1}(\alpha) \text{ et } \hat{\theta}_{\text{inf}} = \text{CDF}^{-1}(1-\alpha)$$

où CDF est la fonction de répartition empirique des  $\theta^*$ . Cette méthode s'appelle de façon assez naturelle la méthode des percentiles.

## 7.2. Exemple d'application : calcul des distributions des paramètres d'une loi

On considère le clapet principal des vannes de décharge au condenseur. Sur 16 vannes, pendant 5 ans, ont été observées 44 données dont 37 censurées à droite.

L'estimateur du maximum de vraisemblance donne :

$$\sigma = 0,93, \text{ écart type} = 0,27 \text{ (soit } \beta = 1,08),$$

$$\mu = 8,29, \text{ écart type} = 0,53 \text{ (soit } \eta = 3977).$$

Cet échantillon est « bootstrappé » 100 fois et les paramètres  $\sigma^*$  et  $\mu^*$  sont à chaque fois estimés. Nous obtenons la table 9. Cette table donne la moyenne, l'écart type et les quantiles empiriques 5 % et 95 % des paramètres  $\sigma^*$  et  $\mu^*$ . Les cinq valeurs extrêmes inférieures et supérieures de  $\sigma^*$  et  $\mu^*$  sur les 100 valeurs issues des 100 échantillons « bootstrappés » sont consignées respectivement par ordres croissant et décroissant dans les deux dernières colonnes de la table 9.

Table 9. - Paramètres  $\sigma$  et  $\mu$  d'une loi de Weibull issus de 100 échantillons « bootstrappés ».

	Moyenne	Ecart type	Quantiles		Extrêmes	
			5 %	95 %	Inf	Sup.
$\sigma^*$	1,04	0,23	0,68	1,42	0,00	1,74
					0,62	1,64
					0,64	1,48
					0,68	1,45
					0,68	1,42
$\mu^*$	8,80	0,81	7,80	10,44	7,45	11,84
					7,55	11,10
					7,64	10,91
					7,67	10,67
					7,80	10,44

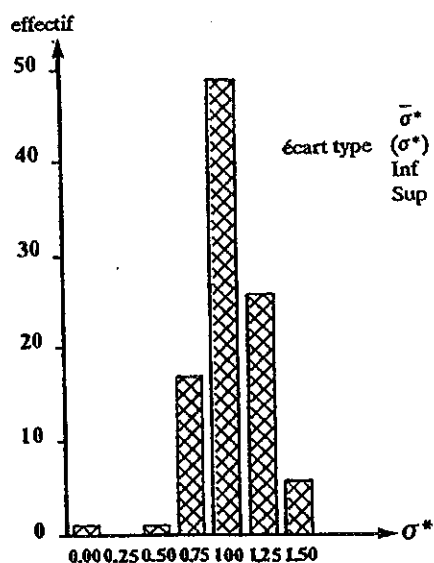


Figure 13. - Distribution du paramètre  $\sigma^*$  « bootstrappé » 100 passages Clapet principal des vannes de décharge du condenseur.

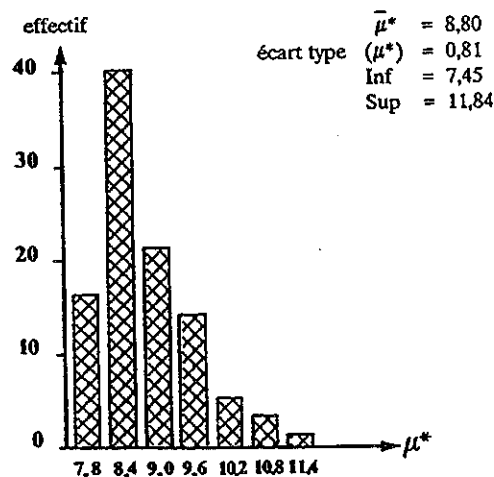


Figure 14. - Distribution du paramètre  $\mu^*$  « bootstrappé » 100 passages Clapet principal des vannes de décharge du condenseur.

La figure 13. montre la distribution du paramètre  $\sigma^*$  ajusté sur 100 échantillons « bootstrappés ». On remarque que la distribution de  $\sigma^*$  semble approximativement normale. La moyenne de  $\sigma^*$  ( $= 1,04$ ) est un peu différente de la valeur du paramètre  $\sigma$  estimé par la méthode du maximum de vraisemblance ( $= 0,93$ ).

L'intervalle de confiance à 90 % sur le paramètre  $\sigma$  obtenu par la méthode de la normalité asymptotique donne :

$$Ic \sigma_{90 \%} = [0,486 ; 1,374]$$

et la méthode du bootstrap :

$$Ic \sigma^*_{90 \%} = [0,68 ; 1,42]$$

L'intervalle de confiance obtenu par la méthode du bootstrap est légèrement plus resserré mais la méthode de la normalité asymptotique (du fait de la forme de la distribution de  $\sigma^*$ ) semble rester ici très valable.

La figure 14. montre la distribution du paramètre  $\mu^*$  ajusté sur 100 échantillons bootstrappés. On remarque que la distribution de  $\mu^*$  est plus proche d'une distribution de type lognormale que normale. La moyenne de  $\mu^*$  ( $= 8,80$ ) est peu différente de la valeur  $\mu$  ( $= 8,29$ ), estimée par la méthode du maximum de vraisemblance.

L'intervalle de confiance à 90 % sur le paramètre  $\mu$  obtenu par la méthode de la normalité asymptotique donne :

$$Ic \mu_{90 \%} = [7,418 ; 9,162]$$

et par la méthode du bootstrap :

$$Ic \mu^*_{90 \%} = [7,80 ; 10,44]$$

L'intervalle de confiance obtenu par la méthode du bootstrap n'est pas symétrique du fait de la distribution à l'allure lognormale de  $\mu^*$ , il est légèrement différent de l'intervalle de confiance de  $\mu$  obtenu par la normalité asymptotique, et semble plus adapté.

## 8. LES METHODES DE DETECTION D'UNE EVOLUTION

La méthode du maximum de vraisemblance, utilisée sur un modèle de Weibull, permet déjà de détecter des évolutions favorables ou défavorables de la fiabilité, par l'estimation du paramètre de forme  $\beta$ .

Le présent paragraphe propose quelques méthodes qui ont montré leur efficacité pour détecter une éventuelle évolution. Le problème de la constance du taux de défaillance est un leitmotiv du traitement des données de défaillance. Les méthodes ci-après permettent de tester cette constante.

### 8.1. La méthode TTT (Total Time on Test)

Cette méthode est graphique. Elle permet, et c'est son grand avantage, de connaître l'évolution du taux de défaillance avec un minimum de calcul.

On classe les temps de défaillance tels que :

$$t_1 < t_2 < \dots < t_n$$

La  $i^{\text{ème}}$  statistique TTT,  $t_{i,n}$  est définie par :

$$t_{i,n} = t_1 + t_2 + \dots + t_{i-1} + (n-i+1) t_i \quad (36)$$

et la  $i^{\text{ème}}$  statistique TTT réduite s'écrit :

$$U_{(i)} = \frac{t_{i,n}}{t_{n,n}} \quad (37)$$

Le graphique TTT est obtenu en traçant  $U_{(i)}$  en fonction de  $i/n$ . Si la distribution est IFR (DFR), le graphique sera concave (convexe). Ce graphique est difficile à interpréter et il est possible qu'il conduise à des analyses erronées lorsque les données censurées sont nombreuses.

### Application

Un exemple est donné pour quelques composants d'une pompe d'un système de contrôle volumétrique et chimique d'une centrale : ligne d'arbre, étanchéité interne et dispositif interne (au total 351 défaillances, extraites du Système de Recueil de Données de Fiabilité, observées sur 81 pompes). La figure 15 trace les courbes obtenues, indiquant une évolution croissante de la fiabilité pour les trois types de composants observés.

statistique (%)

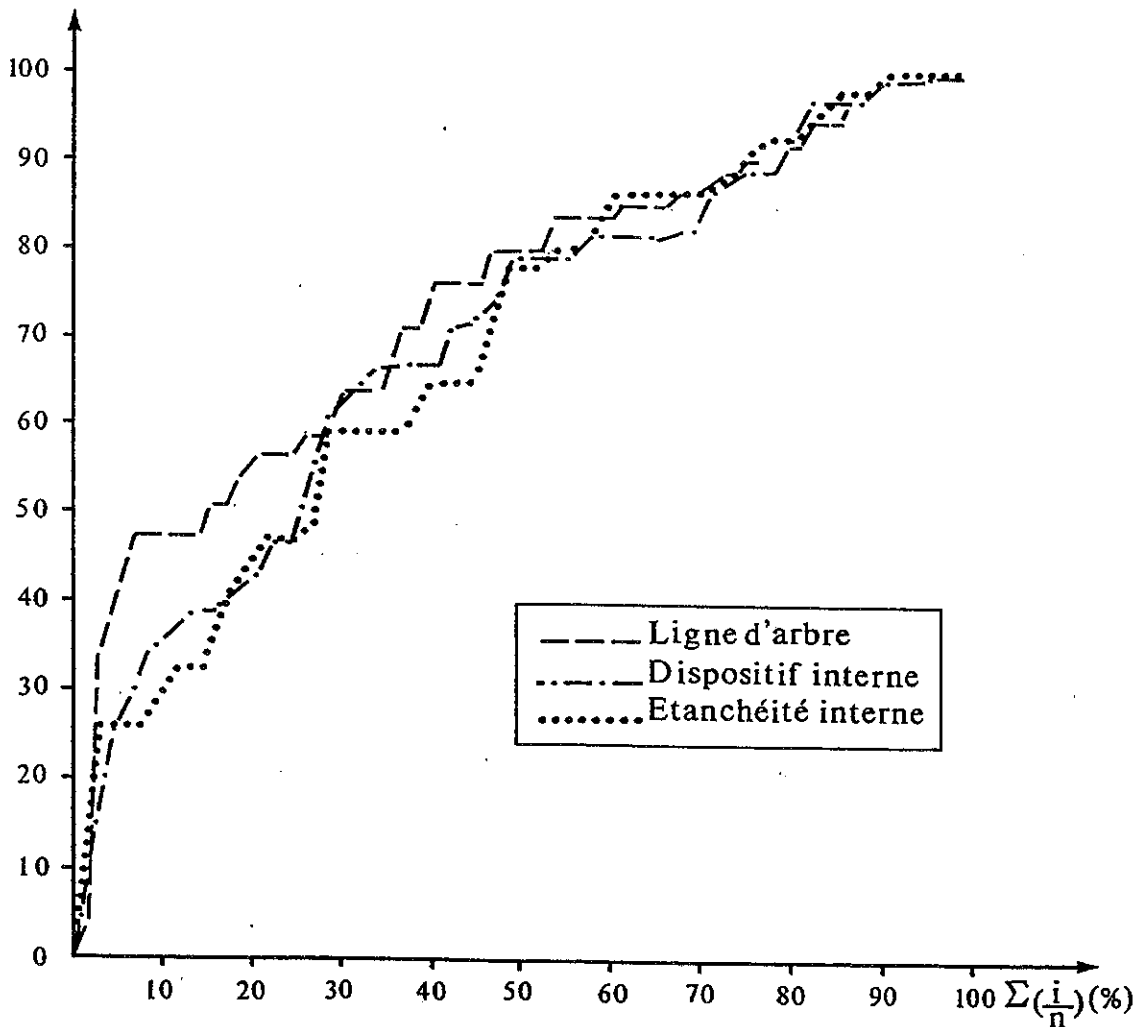


Figure 15. - Application de la méthode TTT à des composants de pompes (Thepault, 1990).



### 8.2. Le test de comparaison

Ce test recherche s'il y a une différence statistiquement significative entre deux échantillons.

Il calcule le nombre d'écart types (valeur Z), entre les fréquences des deux échantillons comparés, en supposant une loi normale de distribution.

Si  $Z > 0$ , la fréquence du premier échantillon est supérieure à celle du second.

Si  $Z < 0$ , la fréquence est inférieure à celle du second échantillon.

Si  $|Z| > 1,645$  (quantile 5 %), les fréquences sont considérées significativement différentes.

Z est calculé par la formule suivante :

$$Z = \frac{P_1 - P_2}{\sqrt{P_3(1-P_3) \left( \frac{1}{T_1} + \frac{1}{T_2} \right)}} \quad (38)$$

où :

$P_1$  est la fréquence observée sur le premier échantillon choisi a priori  $\left( \frac{n_1}{T_1} \right)$ ,

$P_2$  est la fréquence observée sur le second échantillon choisi a priori  $\left( \frac{n_2}{T_2} \right)$ ,

$P_3$  est la fréquence moyenne observée sur l'ensemble des individus des deux échantillons

$$P_3 = \frac{n_1 + n_2}{T_1 + T_2},$$

$n_1, n_2$  les effectifs observés sur  $T_1, T_2$ ,

$T_1, T_2$  les périodes d'observation (exprimées par exemple en heures de marche des individus).

### 8.3. Le test de constance du taux de défaillance

On compare la fréquence de défaillance sur deux périodes différentes, mais de taille sensiblement identique, afin de vérifier la constance du taux de défaillance. Si la fréquence est constante, ceci signifie qu'il n'y a pas eu évolution, c'est-à-dire dégradation ou, au contraire, amélioration du matériel.

On définit :

- T la période d'observation,
- k le nombre de défaillances observées sur T,
- $\hat{\lambda}$  la fréquence de défaillances estimée,
- $\nu$  le nombre de degrés de liberté ( $= 2k$ ),
- $\alpha$  le niveau de confiance,
- 1 et 2 désignant l'indice de la période.

On a :

$$\hat{\lambda} = \frac{\chi^2_{\alpha}(\nu)}{2T} \quad (39)$$

On peut calculer la fréquence de défaillance, observée sur les périodes 1, et 2,

soit  $\frac{k_1}{T_1}$  et  $\frac{k_2}{T_2}$ , ainsi que le rapport :

$$F_0 = \frac{k_2 T_1}{k_1 T_2} \quad (\text{rapport des fréquences de défaillance}).$$

Le rapport de deux  $\chi^2_{\alpha}(\nu)$  pondérés par leurs degrés de liberté suit une loi de Fisher-Snedecor  $F_{\alpha}(\nu_1, \nu_2)$  :

$$F_{\alpha}(\nu_1, \nu_2) = \frac{\chi^2_{\alpha}(\nu_1)/\nu_1}{\chi^2_{\alpha}(\nu_2)/\nu_2} = \frac{1}{F_{1-\alpha}(\nu_2, \nu_1)} \quad (40)$$

On compare  $F_0$  à :

- la valeur inférieure :

$$F_i = F_{1-\alpha}(\nu_1, \nu_2) = \frac{1}{F_{\alpha}(\nu_2, \nu_1)}$$

- la valeur supérieure :

$$F_s = F_{\alpha}(\nu_1, \nu_2)$$

- Si  $F_i < F_0 < F_s$ , on ne peut pas rejeter l'hypothèse ; c'est-à-dire sur les deux périodes examinées, la fréquence de défaillance peut être considérée constante, au seuil  $\alpha$  de signification,

-  $F_0 \leq F_i$  ou  $F_s \leq F_0$ , on refuse l'hypothèse de constance ; la valeur de  $F_0$  n'est pas contenue dans l'intervalle ; il y a une différence notable des valeurs de fréquence et donc une dégradation (ou une amélioration).

#### 8.4. La méthode des « PP-plots »

Cette méthode a été introduite et expérimentée par Deheuvels (1992). Elle nécessite encore d'être validée.

Elle permet de trouver une différence possible entre deux lois de probabilité sous-jacentes de deux échantillons de variables aléatoires, censurées à droite, en particulier des durées de vie.

Deux classes d'objets, ayant évolué dans des environnements distincts, sont observées. On souhaite comparer les lois de probabilités réelles  $F$  et  $G$  des durées de vie dans chacune des deux classes.

La méthode de Kaplan-Meier permet de déterminer les estimateurs  $F_n$  et  $G_n$  de ces deux classes.

La méthode des «PP-plots» peut consister à tracer, pour  $0 \leq s \leq 1$  :

-  $F_n^{\text{inv}}(G_n^{\text{inv}}(s))$  et  $G_n^{\text{inv}}(F_n^{\text{inv}}(s))$  qui, sous l'hypothèse  $F = G$ , convergent vers la première bissectrice ( $y = s$ ),

-  $F_n^{\text{inv}}(G_n^{\text{inv}}(s)) + G_n^{\text{inv}}(F_n^{\text{inv}}(s))$  qui, sous l'hypothèse  $F = G$ , convergent vers la droite d'équation ( $y = 2s$ ),

$F_n^{\text{inv}}$  et  $G_n^{\text{inv}}$  sont les fonctions de quantiles.  $F_n^{\text{inv}}(G_n^{\text{inv}}(s))$  est la probabilité qu'une durée de vie du premier échantillon  $F$  soit inférieure ou égale au quantile d'ordre  $s$  du second échantillon.

#### Application

On s'intéresse aux défaillances du système des grappes longues d'une tranche nucléaire, ayant conduit à des arrêts automatiques ou à des arrêts fortuits. Les durées de fonctionnement sans panne issues d'un fichier d'événements ont été calculées et séparées en deux classes égales. Chaque classe comprend 63 valeurs dont 21 valeurs censurées pour la première et 32 censurées pour la seconde.

La figure 16. donne l'estimation de Kaplan-Meier,  $F_n$  et  $G_n$  pour les deux classes. Les durées de vie sont plus petites sur la première période que sur la seconde. La figure 17. montre que la probabilité de défaillance  $G_n$  est plus petite que la probabilité de défaillance de  $F_n$ , c'est-à-dire que les durées de fonctionnement sans panne de  $F_n$  sont plus petites que celles relatives à  $G_n$ . Par exemple, 50 % des durées de vie de  $F_n$  (médiane) correspondent à 23 % des durées de vie de  $G_n$ . Il y a une augmentation des durées de fonctionnement sans panne pour le second échantillon, prouvant ainsi que les modifications qui ont été opérées sur ce système ont permis l'augmentation des durées de vie.

Signalons cependant que les graphiques issus de cette méthode ne sont pas aisés à interpréter sans l'aide d'un expert du domaine.

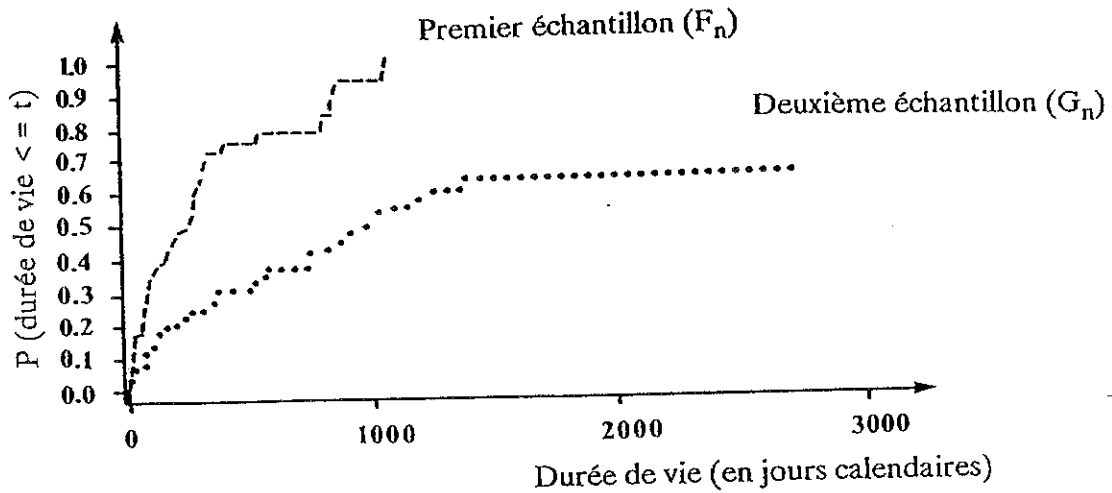


Figure 16. - Estimation des probabilités de défaillance par la méthode de Kaplan-Meier.

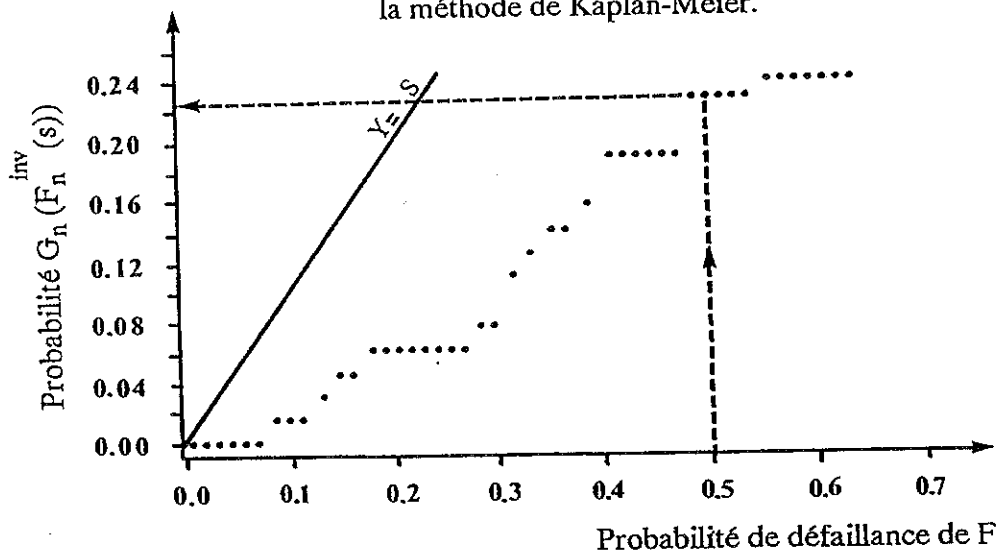


Figure 17. - Graphe du P-P plot  $G_n$  contre  $F_n$ .

## 9. LE MODELE DES RISQUES PROPORTIONNELS

Ce modèle, également appelé modèle de Cox, a été utilisé dans le domaine médical pour analyser des données de survie.

Le modèle des risques proportionnels introduit les effets, sur le taux de défaillance instantané, de certaines variables explicatives ou facteurs d'influence de la durée de vie des matériels. Ce modèle suppose que le taux de défaillance s'écrit :

$$\lambda(t, Z) = \lambda_0(t) \cdot \exp \left( \sum_i \beta_i Z_i \right) \quad (41)$$

où :

$\lambda_0$  est le taux de défaillance, fonction du temps de fonctionnement,

$Z_i$  sont les variables explicatives ( $Z_i = 0$  ou  $1$ ),

$\beta_i$  sont les paramètres inconnus d'un vecteur, qu'il convient de déterminer ; les paramètres  $\beta_i$  déterminent les effets de chaque variable  $Z_i$  sur le taux de défaillance.

On remarque que si  $Z_i = 1$ , le taux  $\lambda$  est multiplié par  $\exp(\beta_i)$ . Les variables explicatives agissent de façon multiplicative sur le taux de défaillance, d'où le nom de risques proportionnels.

On aura :

$$R(t, Z) = \exp \left[ - \int_0^t \lambda(t, Z) \cdot dt \right] \quad (42)$$

La détermination des paramètres  $\beta_i$  par l'analyse statistique dépend de la forme de la fonction  $\lambda(t, Z)$ . Les procédures statistiques permettent le calcul des  $\beta_i$ , par la méthode du maximum de vraisemblance prenant en compte les données censurées à droite.

Si la distribution est une loi de Weibull à deux paramètres, on a :

$$R_0(t) = \exp \left[ - \left( \frac{t}{\eta} \right)^\beta \right]$$

soit :

$$R(t, Z) = R_0(t) \cdot \exp \left( \sum_i \beta_i Z_i \right) \quad (43)$$

On peut estimer les facteurs d'importance des différentes variables explicatives. On peut vérifier l'hypothèse de proportionnalité en traçant  $\ln(-\ln R(t, Z))$  en fonction de  $\ln t$  et en vérifiant que les droites obtenues sont parallèles.

En prenant l'autre forme de la loi de Weibull,

$$R_0(t) = \exp \left[ \exp \left( \frac{\ln t - \mu}{\sigma} \right) \right]$$

on obtient :

$$R(t, Z) = \exp \left\{ - \left[ t^{\frac{1}{\sigma}} \exp \left( - \frac{\mu}{\sigma} + \sum_i \beta_i Z_i \right) \right] \right\} \quad (44)$$

Notons que d'autres modèles de fiabilité à plusieurs paramètres existent : les modèles multiplicatifs, les modèles de régression logistique, ... mais qu'un travail de recherche et de développement reste à réaliser.

#### Application (Pitner, 1990)

Cette application consiste à calculer la probabilité de défaillance d'un tube de générateur de vapeur en fonction de quatre variables (le temps de fonctionnement, le type de dimensionnement mécanique (type I ou type II), l'origine de l'acier et l'origine du tube).

Ce tube est soumis au phénomène de corrosion sous tension. Cet exemple correspond à l'examen de 74 générateurs de vapeur : 244 200 tubes dont 17 280 considérés défectueux soit un taux de censure de 93 % environ.

Les résultats sont donnés sur la figure 18. Ils permettent de déterminer la probabilité de défaillance d'un tube, à un certain temps de fonctionnement, en fonction de son origine.

Table 10. - Probabilité de défaillance à  $t = 50\ 000$  h.

Tubiste	Aciériste	Dimensionnement mécanique	Probabilité de défaillance
V	C	type I	0,22
U	B	type I	0,22
V	B	type I	0,08
V	C	type II	0,004

Il est également possible d'estimer les effets de chaque variable, puisque  $\exp(\beta_i Z_i)$  est le risque relatif associé à la variable explicative  $Z_i$ .

Ainsi l'aciériste C a un taux de défaillance plus grand que l'aciériste B d'un facteur multiplicatif,  $\exp \beta_i = 3,07$ . Le tubiste U a un taux de défaillance plus grand que le tubiste V, d'un facteur multiplicatif  $\exp \beta_i = 3,11$ .

densité de probabilité

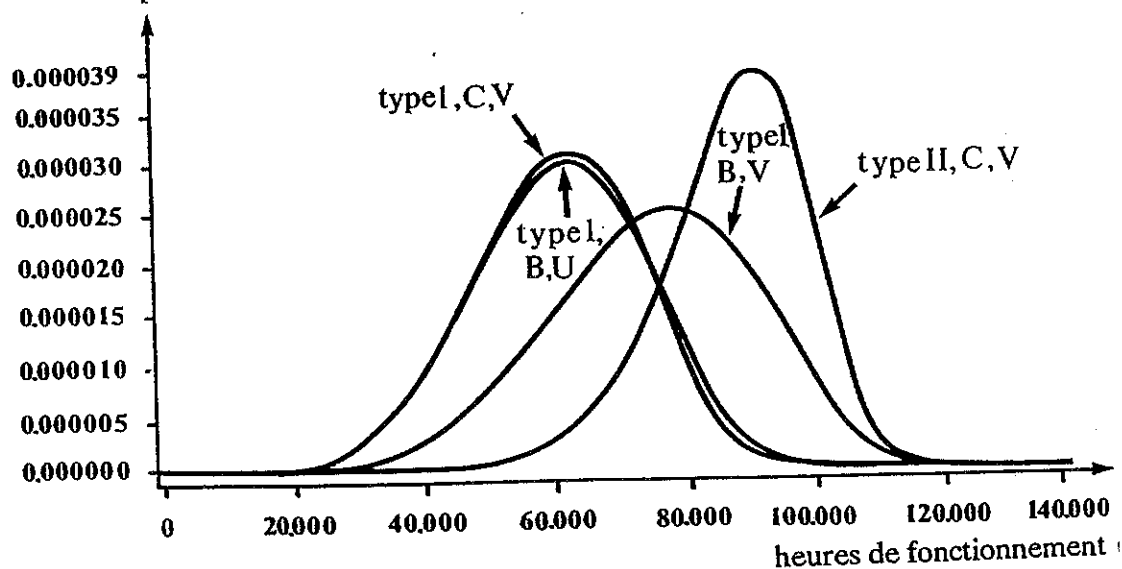


Figure 18. - Modèle de risques proportionnels - Densité de la probabilité de défaillance d'un tube de générateur de vapeur en fonction de trois variables (temps de fonctionnement, type de dimensionnement mécanique, origine de l'acier, origine du tube).

## 10. L'INFLUENCE DE LA MAINTENANCE PREVENTIVE

Cette influence va être montrée sur l'exemple du paragraphe 6.2.

Lorsqu'on détermine, à partir d'une banque de retour d'expérience, les paramètres de fiabilité d'un matériel, on prend en compte implicitement la politique de maintenance de l'exploitant du matériel. La détermination de la fiabilité intrinsèque du matériel nécessite de prendre en compte ces opérations de maintenance préventive, qui contiennent les observations et par conséquent les calculs de taux de défaillance.

### 10.1. Détermination de la fiabilité intrinsèque

Si l'on considère l'exemple de la table 6, on souhaite calculer la fiabilité du presse-étoupe d'un robinet. On suppose que le presse-étoupe est remis à neuf après chaque rechargement (tous les 415 jours environ). Les calculs ont été réalisés par la méthode du maximum de vraisemblance, sur une loi de Weibull à deux paramètres. Les résultats sont donnés sur la figure 19. On observe trois courbes pour le taux de défaillance  $\lambda(t)$  :

- la première correspond à une loi exponentielle ( $\beta = 1,03$ ), c'est la valeur qu'aurait donné l'observation stricte d'une banque de défaillances, sans prendre en compte les opérations de maintenance préventive réalisées sur le matériel (échantillon 1),

- la seconde correspond à une loi de dégradation ( $\beta = 2,24$ ), c'est la loi de fiabilité vraie du presse-étoupe ; le taux de défaillance est fortement croissant ; cette courbe intéresse l'ingénieur de conception ; c'est également elle qui permettra l'optimisation des périodes de maintenance préventive systématique, si l'on adjoint des critères de coût ou de sécurité,

- on constate sur la troisième courbe que le taux de défaillance est représenté par une courbe en dent de scie ; la moyenne de  $\lambda(t)$  sur cette période est sensiblement égale au taux de défaillance de l'échantillon 1 ; on conclut que la réfection du presse-étoupe réinitialise parfaitement le taux de défaillance intrinsèque du presse-étoupe. Ceci veut aussi dire que les actions de maintenance sont effectuées correctement sans être elles-mêmes sources de défaillance.

Le gain de fiabilité apporté par les opérations de maintenance périodique est mesuré par l'écart entre la loi de fiabilité initiale (échantillon n° 2) et la loi de fiabilité intégrant la maintenance (échantillon n° 1). La probabilité de ne pas avoir de défaillance avec la politique actuelle de maintenance, à chaque arrêt pour rechargement, est de 0,79 à la fin du second cycle (cf. figure 19.). Si aucune maintenance n'était réalisée, cette probabilité chuterait à 0,55 (cf. la figure 19 au point d'abscisse 830 jours), d'où un gain de fiabilité égal à 0,24.

Le tracé de ces figures suppose de plus que l'évolution du taux de défaillance intrinsèque du presse-étoupe peut être extrapolée et que la fiabilité  $R(t)$  est observée au temps  $t = 0$  : en effet, le bon fonctionnement du presse-étoupe à l'âge  $t$  dépend de la probabilité conditionnelle de bon fonctionnement dans les âges précédant  $t$ .

### 10.2. Influence de la périodicité de la maintenance préventive

La maintenance préventive de période  $\pi$  a pour effet de remettre à neuf le composant (réinitialisation du taux de défaillance). Le taux de défaillance  $\lambda_1(t)$  d'un composant maintenu périodiquement est égal à la moyenne de son taux de défaillance réel  $\lambda_2(t)$  (fiabilité initiale ou intrinsèque) sur la période d'observation. Le taux de défaillance  $\lambda_1(t)$  prend en compte les effets des remises à neuf séquentielles toutes les périodes  $\pi$ .

$$t \in [0 ; \pi], \lambda_1(t) = \frac{1}{\pi} \int_0^\pi \lambda_2(t) dt = \lambda_1$$

Si  $\lambda_2(t)$  suit une loi de Weibull, on a :

$$\lambda_1 = \frac{\pi^{\beta-1}}{\eta^\beta} \Rightarrow \frac{d\lambda_1}{d\pi} = (\beta - 1) \frac{\pi^{\beta-2}}{\eta^\beta}$$

$d\lambda_1/d\pi$  est la variation de  $\lambda_1$  en fonction de  $\pi$ , de degré  $(\beta - 2)$ . Par conséquent :

- si  $0 < \beta < 1$ , le taux de défaillance est décroissant. Toute opération de maintenance diminue la fiabilité. Il s'agit de la période de jeunesse,

- si  $\beta = 1$ , le taux de défaillance est constant, la maintenance préventive n'a aucun effet. Quel que soit l'âge du composant, la probabilité de défaillance est la même,

- si  $\beta = 2$ , le taux de défaillance est linéairement croissant, les variations de la période de maintenance sont proportionnelles aux variations de la fiabilité,

- si  $1 < \beta < 2$ , tout allongement de la période  $\pi$  produit une variation moindre, de sens opposé, sur la fiabilité. Le composant vieillit légèrement. Un allongement de  $\pi$  provoque une légère augmentation du taux de défaillance  $\lambda_1$ ,

- si  $\beta > 2$ , l'allongement de la période  $\pi$  produit une variation plus importante de sens opposé sur la fiabilité. Le composant vieillit plus vite que précédemment. Tout allongement de  $\pi$  provoque une augmentation bien plus importante que précédemment du taux de défaillance  $\lambda_1$ .

La loi de fiabilité intrinsèque du presse-étoupe a un paramètre de forme  $\beta$  égal à 2,24 (cf. figure 20.). Le cas  $\beta > 2$  s'applique donc. L'allongement de la période  $\pi$  de la maintenance préventive produit une variation plus importante de sens opposé sur la fiabilité. Dans le cas particulier du presse-étoupe, on peut regarder l'effet d'un doublement de la période entre deux maintenances. Sur la figure 20, les trois courbes illustrent : la fiabilité intrinsèque du presse-étoupe (qui décroît très rapidement), la fiabilité du presse-étoupe maintenu à chaque rechargement (au-dessus des deux autres courbes), la fiabilité du presse-étoupe maintenu tous les deux rechargements. La partie haute de la figure 20 décrit les taux de défaillance du presse-étoupe. Le taux de défaillance moyen pour un presse-étoupe maintenu à chaque rechargement est  $2,82 \cdot 10^{-4}$  défaillance/jour, tandis que le taux de défaillance moyen pour un presse-étoupe maintenu tous les deux rechargements est  $8,12 \cdot 10^{-4}$  défaillance/jour. Pour un doublement de la période entre deux maintenances, la fiabilité du presse-étoupe diminue de :

$$65 \% = \frac{(8,12 - 2,82) \cdot 10^{-4}}{8,12 \cdot 10^{-4}}$$

Il faut signaler que le doublement peut être accepté. Seule une analyse économique peut en réalité permettre de choisir. Cela dépend du rapport coût de réparation/coût d'indisponibilité du matériel.



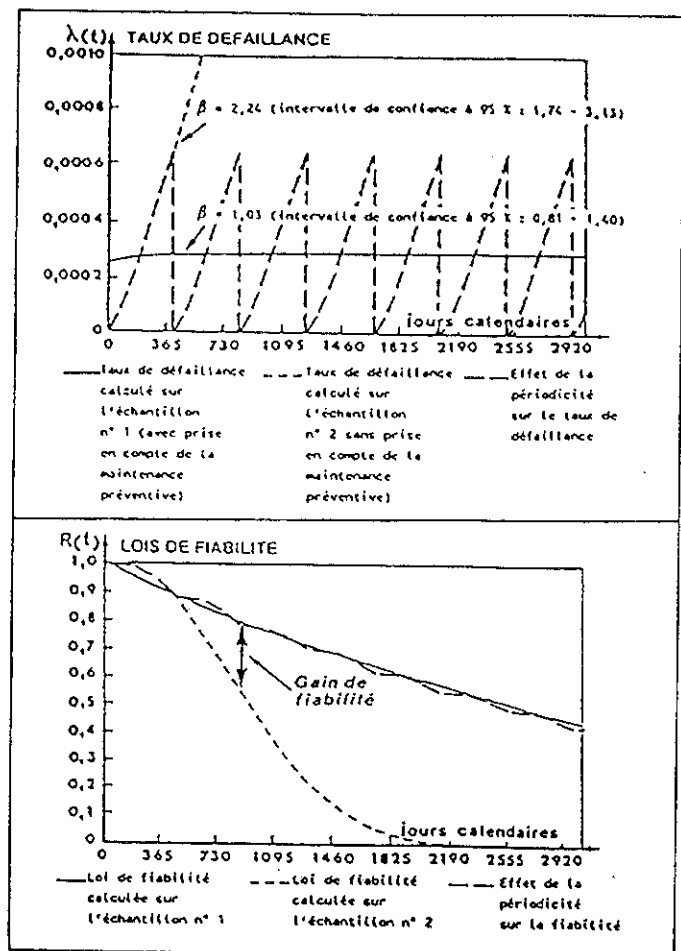


Figure 19. - Comparaison de la fiabilité des deux échantillons. Prise en compte de l'effet de la maintenance périodique du presse-étoupe au bout de 415 jours.

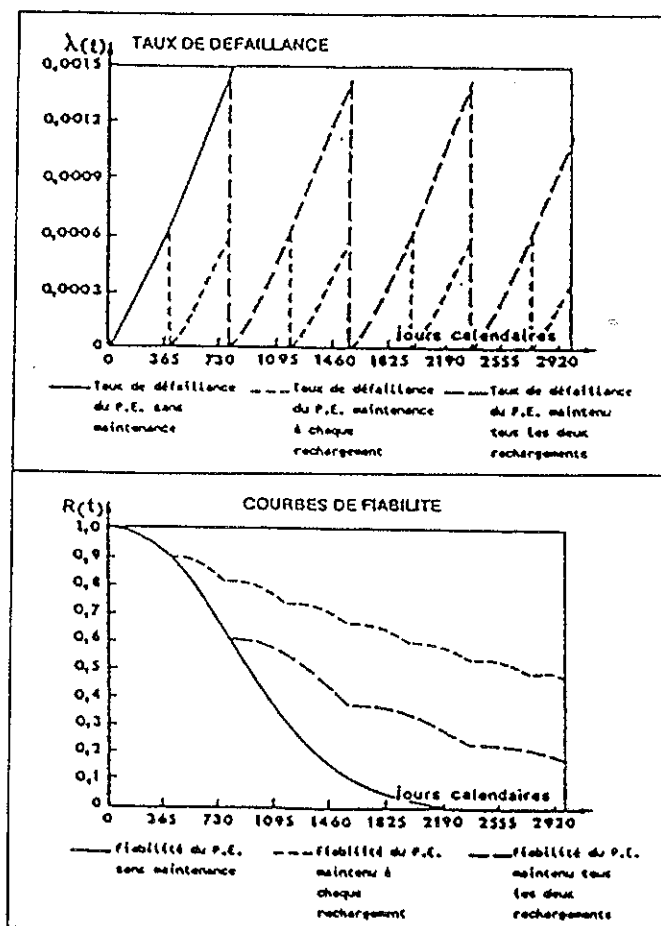


Figure 20. - Comparaison des diverses politiques de maintenance du presse-étoupe (aucune maintenance, maintenance tous les rechargements, maintenance tous les deux rechargements).

### 10.3. Evolution de la fiabilité d'un matériel à la suite des opérations de maintenance préventive

Un matériel, ici un robinet du circuit vapeur de contournement de la turbine est composé de huit sous-composants, pour lesquels il est possible, compte tenu de l'historique des maintenances préventives réalisées et des défaillances observées, de calculer le taux de défaillance. La figure 21 montre l'évolution des taux de défaillance de ces composants. On constate que le clapet pilote a un taux de défaillance décroissant : la maintenance va donc augmenter le taux de défaillance. Au contraire, les opérations de maintenance sur le clapet principal, la tige, le diffuseur et surtout le presse-étoupe feront chuter leur taux de défaillance.

Le taux de défaillance résultant de la vanne est très influencé par les opérations de maintenance préventive et en particulier par le remplacement du presse-étoupe qui, bien qu'étant le composant le plus maintenu, reste le composant le plus pénalisant de cette vanne. La figure 22 montre l'évolution du taux de défaillance de la vanne. On remarque que les variations les plus importantes de ce taux de défaillance sont dues aux maintenances effectuées sur le presse-étoupe. Sans les opérations de maintenance effectuées sur le presse-étoupe, le taux de défaillance de la vanne serait très fortement croissant.

TAUX DE DEFAILLANCE

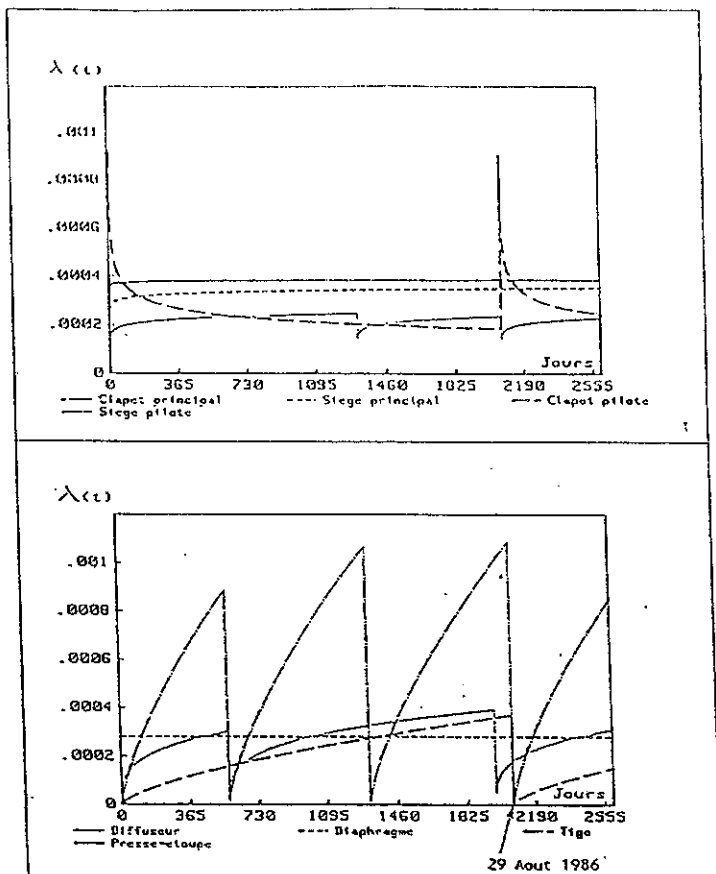


Figure 21. - Taux de défaillance des composants d'une vanne vapeur du circuit de contournement (7 années d'observation).

TAUX DE DEFAILLANCE

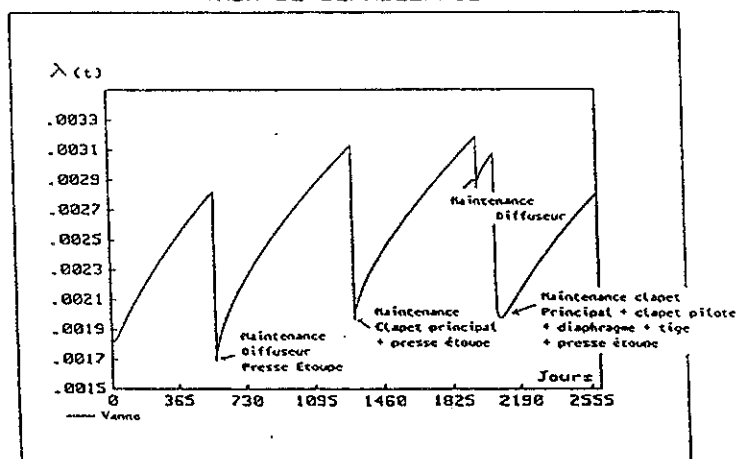


Figure 22. - Evolution du taux de défaillance résultant de cette vanne vapeur du circuit de contournement - Effet de la maintenance préventive.

## 11. CONCLUSION

Après avoir rappelé les caractéristiques principales des données contenues dans les banques de données de défaillances, notamment le fait que ces données sont censurées à droite, ce texte présente les méthodes qui ont paru les plus appropriées et les plus efficaces pour traiter ces données de défaillances. Ces méthodes (du moins la plupart) sont disponibles dans les progiciels statistiques. On a vu que, si les défaillances permettent le calcul de la fiabilité opérationnelle observée, elles ne permettent pas le calcul de la fiabilité vraie du matériel, calcul qui nécessite en outre la connaissance des opérations de maintenance préventive réalisées sur le matériel.

Toutefois, les principales difficultés du traitement du retour d'expérience ne résident ni dans l'accès à ces données de défaillance et de maintenance ni dans le choix et l'utilisation d'une méthode mathématique. Les difficultés se situent dans la validation et l'expertise des défaillances avant leur prise en compte. Cette expertise est la phase la plus difficile et la plus importante, la plus ingrate car il faut bien souvent analyser visuellement un grand nombre de défaillances mais aussi la plus gratifiante car notre souci n'est pas de posséder une belle banque de données et de la traiter mais bien d'améliorer la fiabilité des matériels et des

## REFERENCES BIBLIOGRAPHIQUES

- J.C. LIGERON, C. MARCOVICI. - *Utilisation des techniques de fiabilité en mécanique*. Technique et Documentation, 1974.
- B. EFRON. - *Bootstrap methods: another look to the jackknife*. Ann. Stat., Vol. 7, 1979.
- J.F. LAWLESS. - *Statistical models and methods for lifetime data*. John Wiley & Sons, 1982.
- A. LANNOY, H. PROCACCIA. - *Impact of maintenance actions on the reliability and availability of components*. Euredata Seminar of Maintenance, Chamonix, 19-20 janvier 1989.
- J.J. DROESBEKE, B. FICHET, P. TASSI. - *Analyse statistique des durées de vie*. Economica, 1989.
- S. THEPAULT. - *Etude de vieillissement de pompes*. Rapport interne Framatome, octobre 1990.
- P. PITNER. - *Maintenance programming as a function of parameters influencing ageing*. Euredata meeting « Ageing and maintenance ». Chamonix, 4-5 octobre 1990.
- C. CHEVALIER. - *Modélisation de la fiabilité des matériels mécaniques nucléaires à partir de données issues du retour d'expérience (applications industrielles)*. Thèse de l'Université PARIS VI, décembre 1991.
- M.J. CRAWDER, A.C. KIMBER, R.L. SMITH et T.J. SWEETING. - *Statistical analysis of reliability data*. Chapman & all. Première édition, 1991.
- P. DEHEUVELS, A. LANNOY, V. VERON. - *Trois tests non paramétriques d'égalité des lois de probabilité sous-jacentes à deux échantillons de durées de vie connues*. Rapport de DEA, Université de PARIS VI - EDF-REME, août 1992.
- C. CHEVALIER, P. DEHEUVELS, A. LANNOY. - *Deux exemples d'estimation de la fiabilité de matériels à partir d'informations fournies par les maintenances corrective et préventive*. Revue Statistique appliquée, 1992, XXXX(2), pp. 77-90.
- J.L. BOUCHET, J. HAVART, J.L. JACQUOT, A. LANNOY, D. VASSEUR. - *Development of operation feedback analysis methods for the RCM approach*. ESREDA « Equipment ageing and maintenance ». Chamonix, 14-15 octobre 1992.
- P. DEHEUVELS (Université de PARIS VI), J. EINMAHL (Eindhoven University of Technology). - *Approximation and two sample tests based on P-P and Q-Q plots of the Kaplan-Meier estimators of life time distributions*. Journal of multivariate analysis - Vol. 43, N° 2, Novembre 1992, pp. 200-217.
- H. PROCACCIA, C.A. CLAROTTI, L. PIEPSZOWNIK. - *Fiabilité des équipements et théorie de la décision statistique fréquentielle et bayésienne*. Editions Eyrolles. Collection de la Direction des Etudes et Recherches, 1992.