Outils Modernes de Conception

Ingénierie de la Fiabilité

(Extraction de Connaissances & Fiabilité)

Ecole Centrale de Lyon - Métiers 3A

2009-2010

Alexander Saidi

Propos

- Cycle de Vie de produits industriels

- Fiabilité et défaillance par des **méthodes et outils informatiques modernes**

Plan

1- Aperçu des méthodes générales de l'extraction de connaissances

➥ A travers des exemples simples

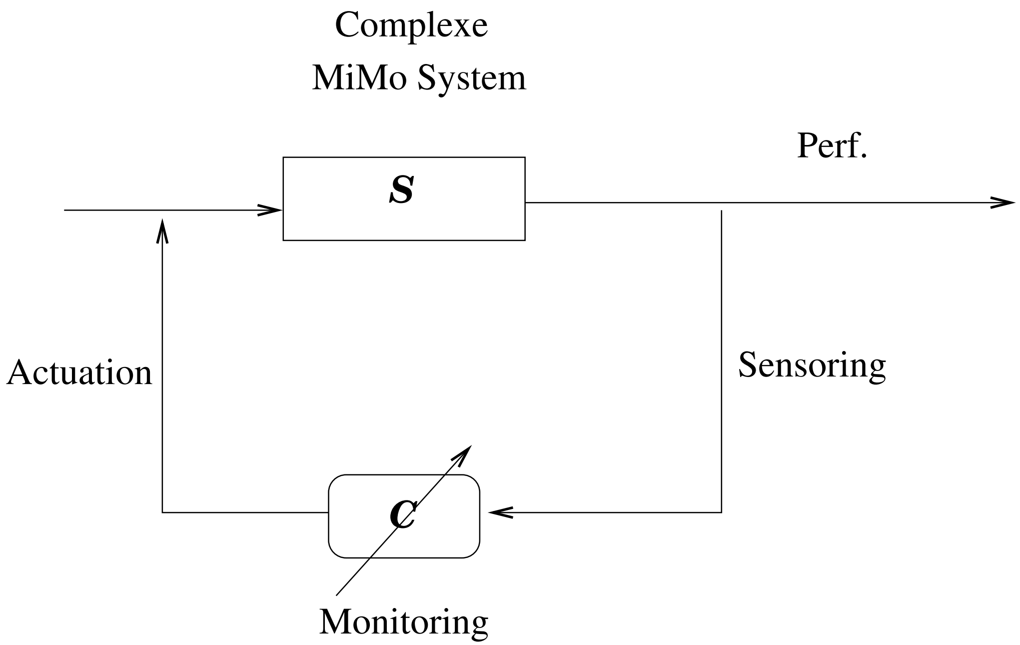
2- Fiabilité et défaillances des processus complexes

➥ Fiabilité et Cycle de vie de produits

➥ Métriques et Distributions

➥ Utilisation des Réseaux Bayésiens, Réseaux de Neurones, etc.

Rappel : une vue Synthétique :



- Analyse Fiabiliste : principalement **S** (cf. Chapitre 2)

- Synthèse Fiabiliste : Fiabilité de la fonction Perf. (difficile)

➥ Exemples, ABS, Suspension Hydro-active, etc.

• Habituellement, une **fonction** englobe les **modes** de défaillances (produit des fiab)

1 $\boxed{\text{Séance 2 : Introduction}}$

L'apport de l'analyse Bayesienne ?

Règle de Bayes : $$\boxed{\mathrm{Pr[H \mid E]} = \frac{Pr[E|H] \times Pr[H]}{Pr[E]}}$$

➥ $Pr[E]$: la croyance = une idée de la proba **a priori** (de préférence sans les données)

➥ $Pr[E|H]$: la **vraisemblance** du modèle observé (avec les données)

➥ On obtient la probabilité **a posteriori** (permet de réviser la proba a priori).

1.0.1 *Propos et Motivations*

• Utilisation des données de Fiabilité issues de retour d'expérience (REX) dans des fonctions de sécurité, fonctionnement, survie, ...

• (Aperçu des) outils et méthodes pour utiliser ces informations de manière optimum dans un soucis de sécurité

• <u>On procédera par :</u>

➙ Rappel des notions fondamentales de sûreté de fonctionnement

➙ Proposition de modèles utiles au calcul du taux de défaillance en fonction des données disponibles

➙ Approche globale pour la détermination des facteurs influents

1.0.2 Différentes approches

• On peut calculer la fiabilité par des techniques non Bayesiennes :

1. **Fault Tress (FT)** : la plus utilisée parmi les non Bayesiens

   ➙ mais ne prend pas en compte la démendance temporelles et fonctionnelle des composantes.

   ➙ Markov Chains et FT dynamiques apportent une réponse à la dépendance temporelle.

   ➥ Mais restent d'une combinatoire très grandes pour les systèmes complexes.

2. Diagramme de Blocks : permet de tester des config diverses (1 block par composant).

3. Fuzzy Bayesiens

4. Réseaux de Petri

5. Méthode des moments (math) : $\mu_k = E[(X - E[X])^k]$ approche bien les distributions.

6. etc...

1.0.3 2 exemples de FT

• Un arbre de défaillance (Fault Tree) pour la protection incendie avec 3 événements basiques, un intermédiaire et un évt principal.

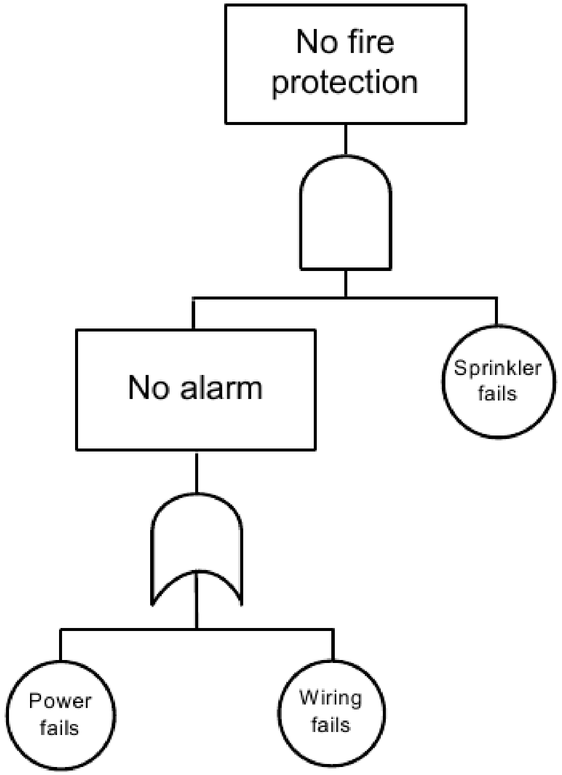
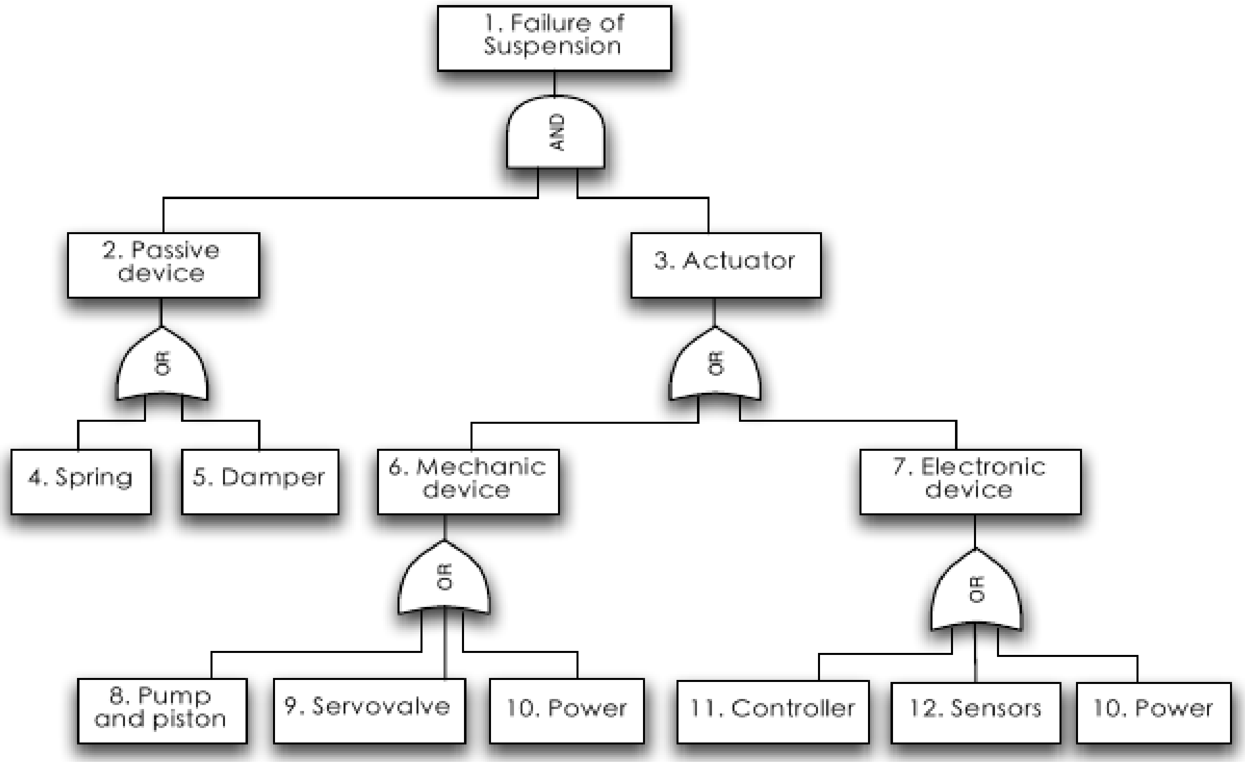


FIGURE 1.1 – Arbre de défaillance de la protection incendie

Exemple 2 : Arbre de défaillance du dispositif *suspension active* :



1.0.4 *Contexte de la fiabilité*

• **Maîtrise des risques industriels** : les règlements de tous genres poussent les industriels à s'assurer de la sûreté de fonctionnement (cf. : soupape de sécurité)

➥ Protection d'environnement, les directives **Seveso**s diverses

• **Frein ou Atout ?** : atout si bien utilisée :

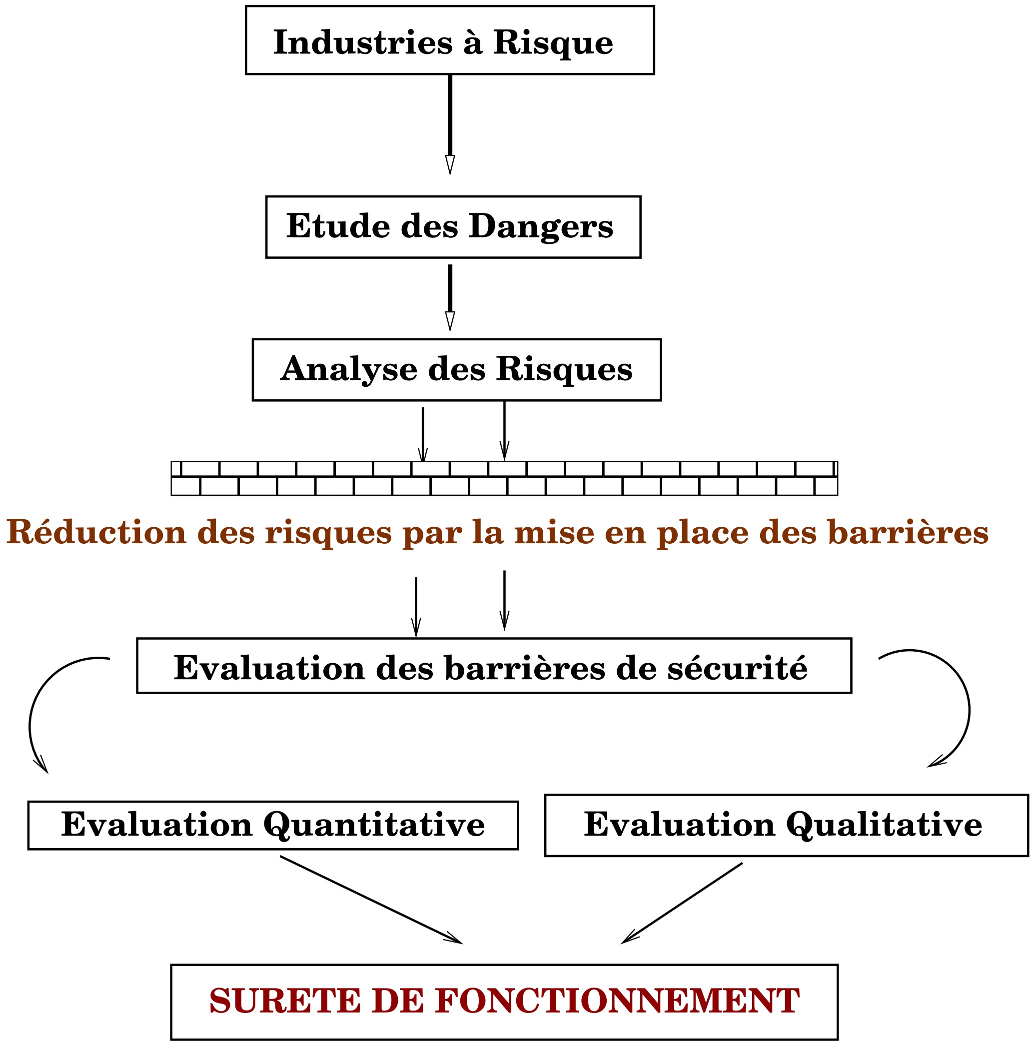
☞ Gestion / prévision des risques ➙ prévision des risques de non fonctionnement ➙ amélioration de la **productivité**.

☞ Mise en place des **barrières de sécurité** (de nature humaine ou techniques) pour assurer une prévention efficace des risques dans l'entreprise.

☞ Maîtrise des risques des installation industrielles.

➙ ..

Sureté de fonctionnement :



• Evaluer la fiabilité d'une barrière de sécurité (qualitative et quantitative)

➥ **il faut connaître sa fiabilité**

Constat : pendant long temps, les industries ont négligé cela (sauf dans certains secteurs comme l'*aéronautique*, plus généralement en transport).

➥ Le **REX** interne se fait rare et négligé,

➥ **Peu diffusé** (si présent) du fait des enjeux économiques (coûts) engendrés.

• Il existe tout de même des BDs (BDF) sur le marché qui permettent de trouver des informations intéressantes (mais le contexte d'utilisation est très important)

1.0.5 *L'apport des BDs*

• En **électronique** : bonnes BDFs

➥ données REX et des modèles de fiabilité prévisionnels bien établis

➥ car : production en masse, standardisation de la production

• En Systèmes **mécaniques** :

➥ BDs de synthèse des données de fiabilité d'équipements **similaires**.

• Systèmes **électro-mécaniques** : problèmes !

➥ **aucune** méthode commune de prédiction de fiabilité n'existe

➥ **Grande dispersion** des taux de défaillance (même pour des composants apparemments similaires)

1.0.6 *Difficultés inhérentes à la mécanique*

• Un équipement mécanique est **composé** de plusieurs composants

➥ Utilisés de manières **diverses** : multitude des modes (de *défaillance*)

➥ Assurent souvent une **fonction** ➙ utilisés pour des objectifs <u>différents</u>.

➧ Selon leur utilisation, les défaillances ne seront pas les mêmes.

• <u>Rareté de données</u> de défaillance pour des composants moins "std" utilisés dans des applications spécifiques ➙ <u>récupération de REX difficile</u>

• Le taux de défaillance des **composants** mécaniques rarement modélisables par des taux **constantes** à cause des mécanismes de **dégradation** auxquels ils sont soumis

➥ **conditions** de fonctionnement, fatigue, vibrations et autres stress.

➥ le **mélange** de divers taux rend la prédiction difficile.

Obstacles pour une modélisation de la fiabilité mécanique :

• Contrairement aux composants électroniques, la fiabilité des matériels mécaniques est **sensible** au régime, au taux d'utilisation, au mode de fonctionnement.

• La seule info du taux de bon fonctionnement est insuffisante (pour une prédiction).

• Le taux de défaillance est très lié à l'application (seule la fonction importe !)

• **Données de qualité rares** : temps de défaillances individuels + temps de bon fonctionnement + nombre de défaillance de l'ensemble, ... en quantité suffisante.

Autres obstacles : le coût supérieur des composants mécaniques (vs. éléctroniques) et leur production restreinte ➞ REX a moins de conséquence.

• Dans les installations à risque, les industriels procèdent souvent à une **maintenance préventive** efficace

➞ les défaillances pendant une phase de fonctionnement sont moins fréquemment constatées.

Bilan pour la suite :

• Données de qualité **rares**

• Il existent des outils d'analyse (cf. **AMEDEC** : analyse des Modes de Défaillance, leurs effets et leurs Criticités).

• Mais le problème récurrent reste toujours le même :

déterminer la probabilité d'occurrence des modes de défaillance observées .

• La difficulté de prévision d'une fiabilité méca est **croissante selon le niveau** :

➥**Facile** pour un composant simple soumis à un phénomène phys. de dégradation

➥ Bien plus **difficile** pour un matériel complexe soumis à l'ensemble des phénomènes de dégradation de chaque composant (➙ vers la *fonction* du système ?)

Peut on décomposer un système complexe ?

• La **décomposition naïve** d'un équipement complexe n'est pas une bonne méthode de prédiction de la fiabilité de l'ensemble :

➥ Car les composants ne suivent pas tous le même **régime de dégradation**,

➥ N'ont pas la même **distribution** en terme de fiabilité.

➥ De plus, connaître toutes les fiabilités des différents composants devient rapidement trop conséquent en temps et donc en **coût**.

▶ **Et par mimétisme ?**

• L'étude des matériels similaires est une indication mais ne donne pas de résultats précis.

• Un bonne méthode d'analyse doit en tenir compte sans être trop dépendant des données en entrée. On peut néanmoins utiliser les données REX.

On y va quand même ? :

Objectif : calculer des taux de **défaillances** spécifiques au matériel mécanique et définir les facteurs **influents** sur la fiabilité.

• On utilisera des données REX, plus <u>pessimistes</u> que les données constructeurs

➥ car testés dans des conditions "saines" .

• On manipule donc des données du "terrain"

➥ ce qui rend les choses plus réalistes mais plus difficiles.

• Les modèles de décomposition peuvent aider (un peu, voir plus loin)

➥ MCL ?

<u>Ad augusta, per Angusta</u>

1.0.7 Plan sommaire de la suite

- Quelques notions indispensables à la compréhension du reste ...
  - ➥ Définitions des domaines, concepts, notions relatives au REX
  - ➥ Différentes notions nécessaires à la modélisation de la fiabilité

- Méthodes d'évaluation d'un équipement : Modèle statistique de **vie accélérée**
- Données de Fiabilité disponibles et leur utilisation
- Modélisation Bayésienne
- D'autres méthodes (Hasard proportionnel et RN)
- Application à un exemple : **soupape de sûreté**.

1.1 Généralités

• Un ingénieur en sûreté de fonctionnement tente d'évaluer les performances des barrières de sécurité à travers l'analyse de 3 critères :

➥ **Efficacité**, **Temps de réponse** et le **Niveau de confiance**

• La plus difficile : quantification du niveau de confiance car :

➥ Il faut savoir quantifier la proba d'occurrences des états de **défaillance**

➥ Il faut une étude de **fiabilité** plus générale : sûreté de fonctionnement (SdF)

• FDMS = Fiabilité, Maintenabilité, Disponibilité, Sécurité

➙ Ce sont les 4 objectifs de la **sûreté de fonctionnement** que l'ingénieur doit modéliser.

• Il peut disposer des données précises en entrée comme le taux de défaillance (obtenu par les calculs d'un modèle sur la base de REX + BDFs).

1.1.1 *Quelques définitions*

• **Fiabilité** : *aptitude à bien accomplir une fonction dans des conditions données, sur un intervalle de temps donné*

➥ Notion de probabilité de Fiabilité = probabilité de ce comportement.

• **Maintenabilité** : *aptitude à être maintenu ou rétabli, sur un intervalle de temps donnée dans un état où l'on peut accomplir une fonction requise.*

➥ Conditions de maintenance, procédure et moyens

• **Disponibilité** : *aptitude d'un matériel, sous les aspects de sa fiabilité et d'organisation de maintenance, à être en état d'accomplir une fonction requise dans des conditions de temps déterminé.*

Pour la **Sûreté** : deux approches (de plus en plus complémentaires)

➥ Approche déterministe : on choisit un scénario parmi N qui englobe l'ensemble des contraintes.

➥ Approche probabiliste (**Très utilisé en France**) : évaluation prévisionnelle de la SdF en attachant une proba à chaque scénario. On évalue donc les risques.

- Dans les 2 cas : définir la fréquence des dangers identifiés

- Pour évaluer la SdF, il faut connaître :

  - L'architecture du système & les Conditions et l'environnement de son utilisation

  - Le modèle de défaillance du système ou de ses composants

    ➥ loi de fiabilité du système + proba pour chaque composant + loi de composition

- **REX** : retour d'expérience : collecter, archiver, analyser....

1.1.2 *Divers phases de Fiabilité et les méthodes associées*

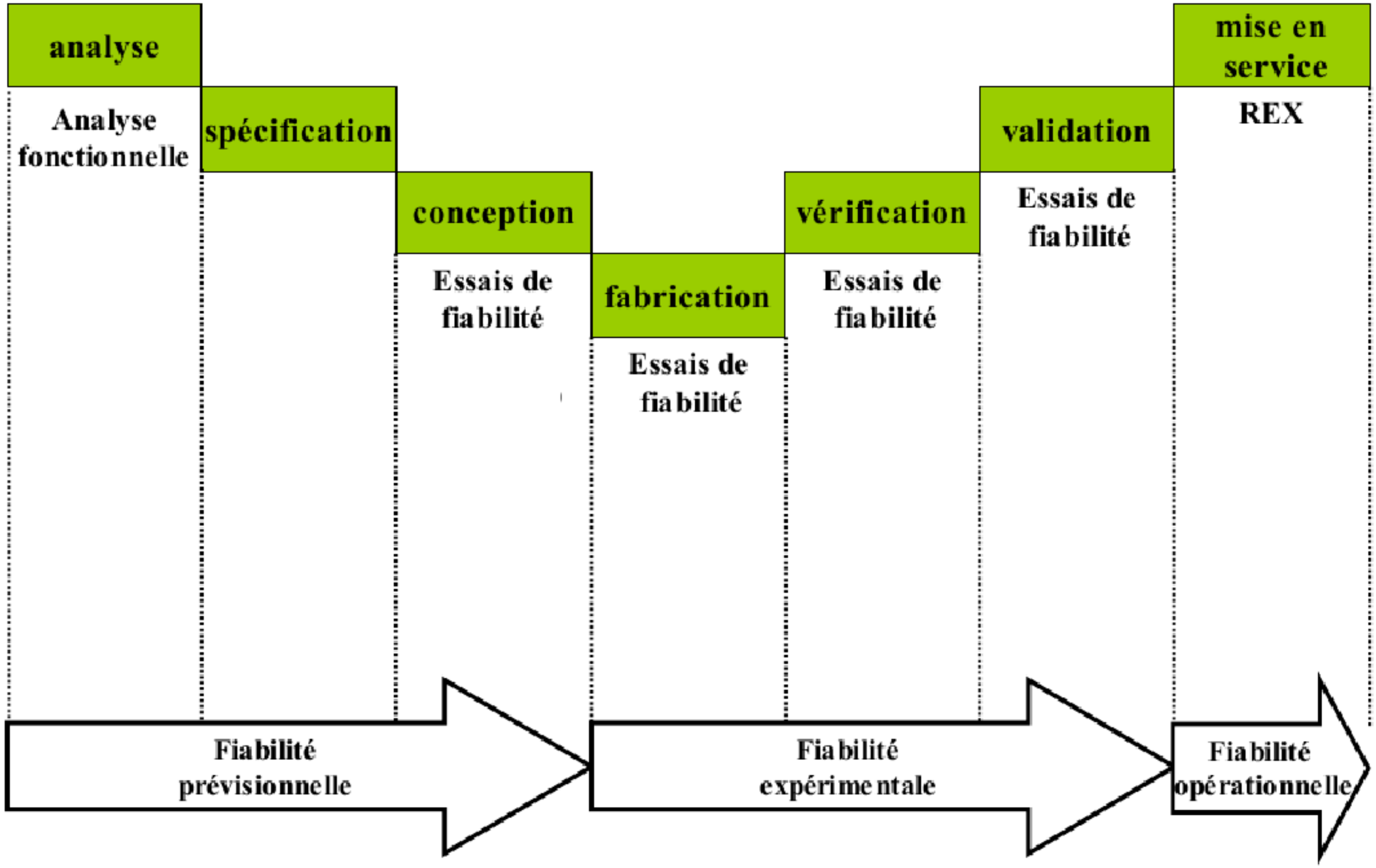


FIGURE 1.2 – Méthodes d'évaluation pendant les trois différents niveaux de l'évaluation de la fiabilité tout au long du cycle de développement.

N.B. : domaines **Logiciel**, **Electronique**, **Mécanique** , **Mécatronique**

1.1.3 *Fiabilité et le cycle V*

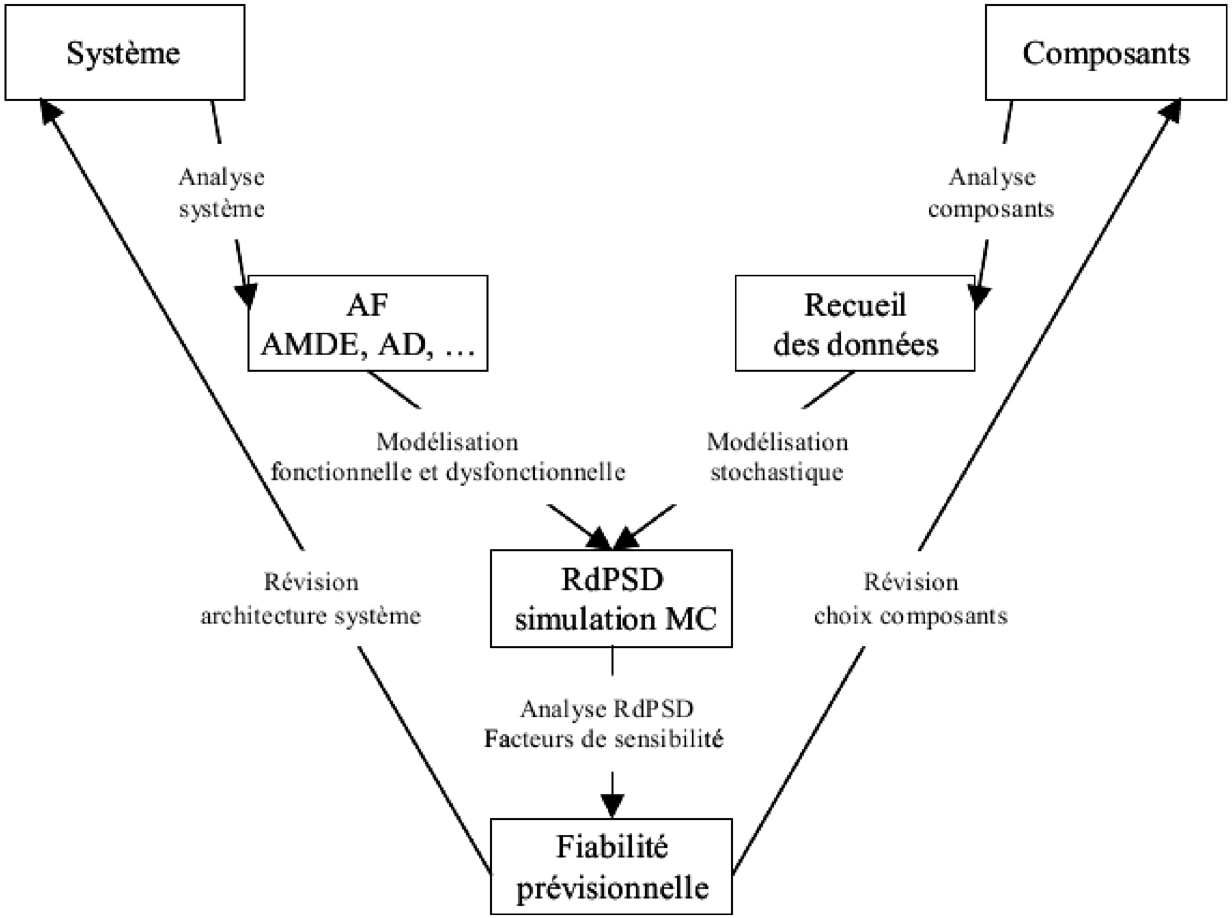


FIGURE 1.3 – Démarche de la fiabilité Prévisionnelle

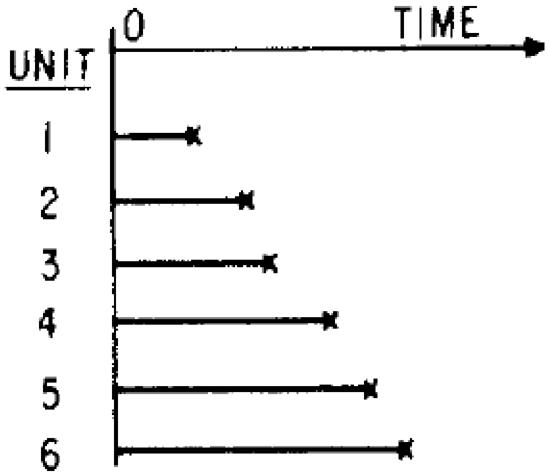
1.1.4 *Différents types de données (REX ou tests)*

• Les données peuvent être complètes ou censurées

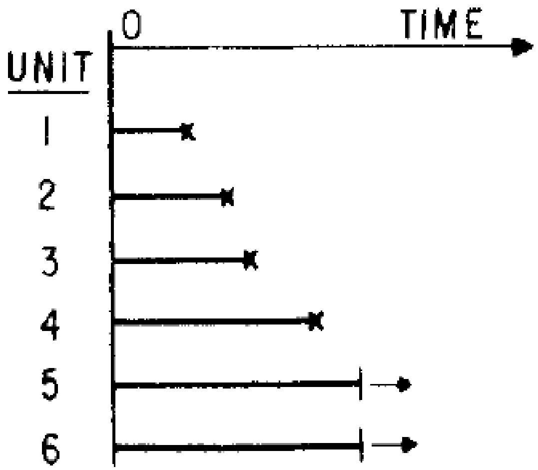
▶ **Complète** : on relève tout, jusqu'à la panne de tous matériels (peut être long)

➥ les instants de défaillance de chaque unité de l'échantillon sont connues

▶ **Censurés** : lorsque **toutes** les unités testées ne sont pas tombées en panne.



Données complètes (tous en panne)



Censurées avec 4 pannes (arrêt au temps t)

1- Censurée à droite : on décide d'arrêter l'observation à la date Td.

➥ A cette date, le dispositif n'a pas eu de défaillance.

➥ La date T est alors une donnée censurée à droite (ou tronquée à droite).

➥ type de données récolotées très fréquent pour les équipements de sécurité qui tombent peu en panne.

Deux types de données censurées à droite : (fig. ci-dessus à droite)

✔ **Type I** : un test se termine à un moment t donné.

➥ appelé simplement **censure à droite** car le temps de panne à droite est inconnu.

➥ On connaît alors le nombre de matériel défaillants, les moments des défaillances, le type de panne, ...

➥ Le temps censuré est fixe, le nbr. des défaillances= une variable aléatoire

✔ **Type II** (peu courant, dit aussi **pourcentage** = percentile) :

➙ le test se poursuit jusqu'à ce qu'une proportion fixe de matériel échoue.

➥ Ici, le nbr de matériel ayant échoué est fixe et le temps est une var aléatoire.

➥ Très utile car on sait d'avance combien de pannes auront lieu

➥ Facilite la planification des tests mais impraticable si le temps n'est pas borné.

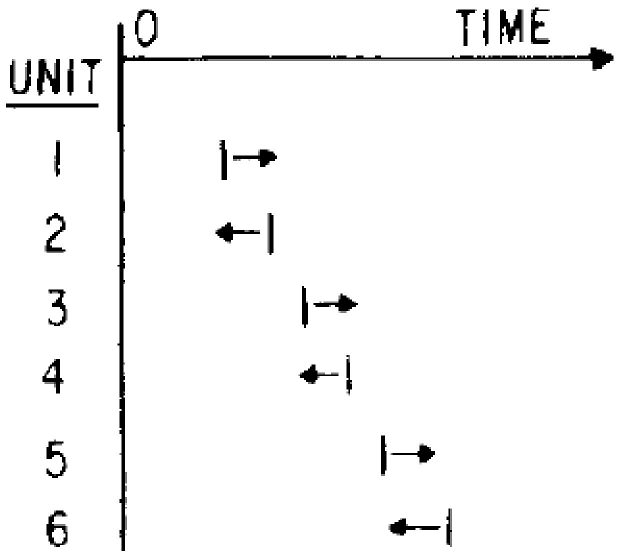


FIGURE 1.4 – |→ : continue à fonctionner, ←| : panne a eu lieu avant

2- Censuré à gauche : On observe le matériel à partir de la date Tg.

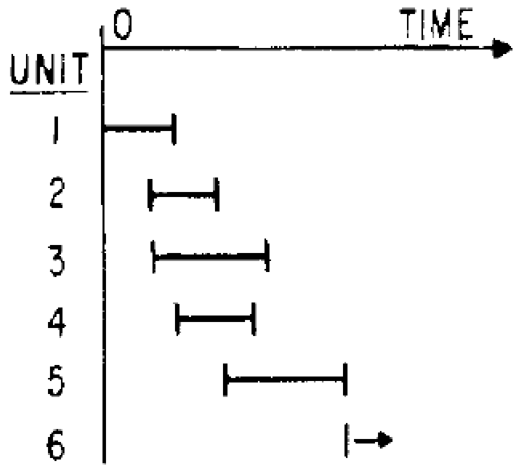
➥ On constate des pannes mais on ne connaît pas la date t de la défaillance

➥ On sait seulement que "t est inférieur à Tg".

Exemple : en médecine, un patient entre à l'hôpital à une date Tg et survit un certain temps. Mais l'on ne sait pas à quel moment les symptômes se sont présentés et ont été diagnostiqués la première fois.

3- Censuré par intervalle :

Le dispositif a eu une défaillance entre Tg et Td connues (mais pas le moment exact de la panne).



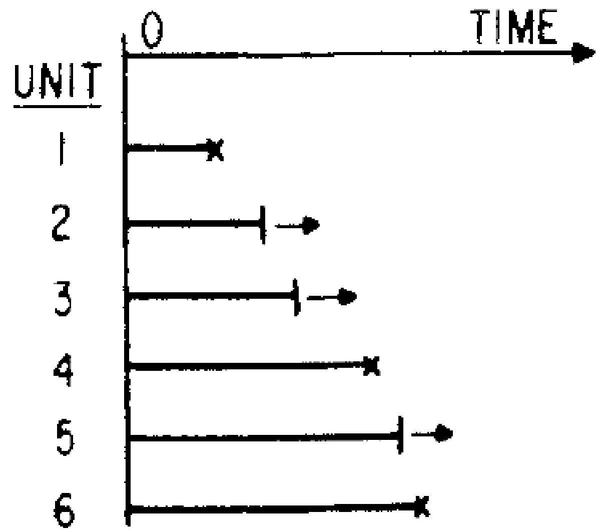
4- Censure multiple : la censure peut se produire à différents moments avec les situations :

➥ pas de panne pendant l'observation ou

➥ panne + le moment exact de la panne ou

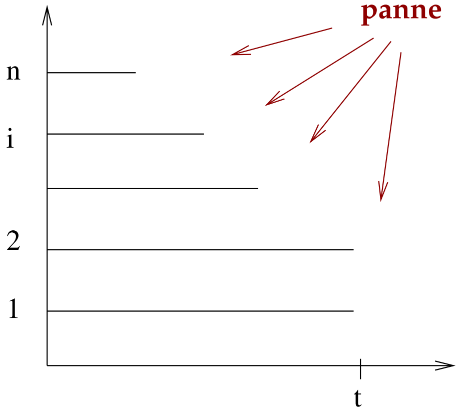
➥ l'intervalle pendant lequel il y a eu défaillance

➥ on peut donc avoir des temps ou des intervalles différents pour différentes unités

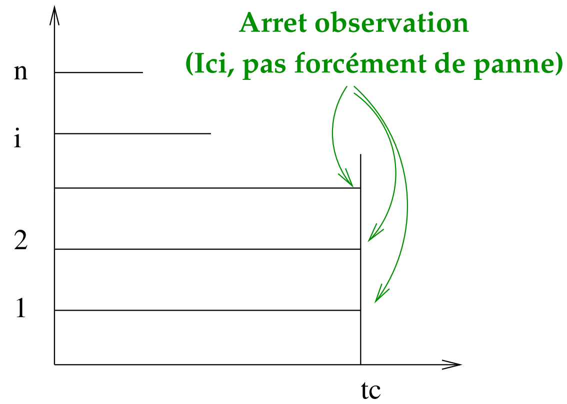


5- Censure Simple : la censure se produit à un moment donné

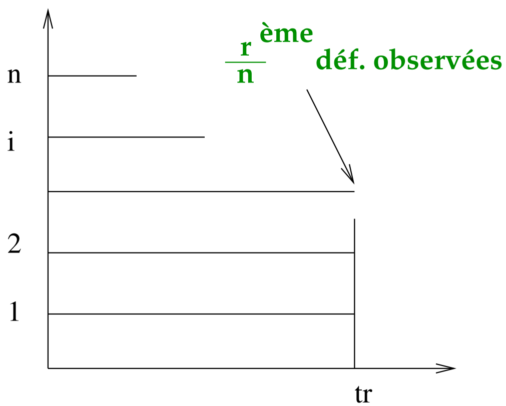
• Quelques types importants de données :



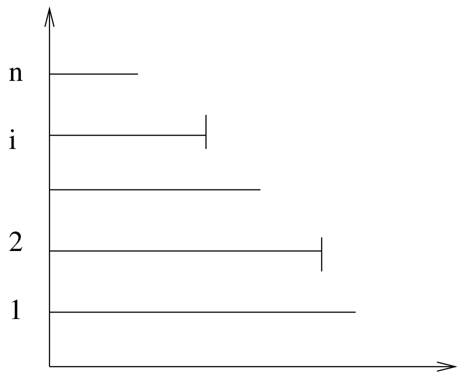
(a) Ecantillon complet



(b) Echantillon censuré sur tc à droite



(c) Censuré sur rème défaillance



(d) Censuré aléatoirement à droite

Dans le cas du REX

• On a souvent affaire à des données **censurées à droite** :

➥ à la date d'observation, le matériel a eu / n'a pas eu de défaillance.

• N.B. : la maintenance préventive (remplacement du matériel supposé être dégradé mais dégradation pas forcément observée) génère des données censurées à droite.

• Les dates sont souvent référencées par rapport à une origine connue :

➥ **MSI** (mise en service indust), date de remplacement, etc.

➥ Cette date n'est pas forcément la date de naissance réelle du matériel.

➥ Les temps obtenus sont en général la différence entre deux dates.

• Autre unités : une **unité d'usage** : nombre de cycles, nbr de sollicitaions, etc...

1.2 La maintenance

• Pose problème car modifie les données et les caractéristiques de fiabilité (temps, nature, occurrence, gravité, etc).

➥ Les données de REX deviennent <u>censurées à droite</u> (maintenances préventives).

• 2 types de maintenances : **corrective** et **préventive**

Maintenance Corrective : réparation de panne.

➥ Ne pose pas de problème majeur mais la nature de la réparation devient importante.

➥ On distingue ABAO (As Bad As Old) et AGAN (As Good As New).

• **Suite : Préventive**

1.2.1 *Maintenance Corrective*

ABAO : la maintenance remet le système dans son état d'avant la panne.

• Le processus aléatoire correspondant est un processus **Poisson** *Non Homogène* (NHPP) :

➥ l'intensité (λ, Poisson) de la défaillance est une fonction du temps et ne dépend pas du passé du processus : $\forall t \geq 0, \lambda = \lambda(t)$.

• La forme la plus utilisé de NHPP est le modèle PLP (*Power Law Process ou Douan*) où l'intensité est une puissance du temps : $\lambda_t = \alpha.\beta.t^{\beta-1}$ avec $\alpha, \beta > 0$

• Ce modèle est aux systèmes réparables ce que **Weibull** est aux systèmes non réparables : les paramètres caractérisent l'usure du système :

$\beta < 1$: amélioration ou rajeunissement

$\beta = 1$: stabilité $\beta > 1$: usure ou vieillissement

• ABAO : l'origine du temps n'est pas modifiée et on ne prend pas en compte le temps de censure.

$AGAN$: le matériel est remis à neuf suite à une défaillance.

- ➥ L'origine du temps change : comme une date de mise en service (MSI)
- ➥ Le processus aléatoire correspondant est le processus de renouvellement (Perfect Renewal Processes PRP) :
  - ➞ l'intensité de la défaillance est de la forme : $\forall t \geq 0, \lambda_t = \lambda(t - T_{D_t})$.

Avec D_t : le nbr de défaillance observées entre l'instant initial et t

T_{D_t} : le temps entre la D^{eme} défaillance et l'instant t.

• Maintenance imparfaite et **âge virtuel** :

➥ entre les deux modèles ABAO et AGAN : le modèle d'âge virtuel.

➥ on considère que la **ième maintenance** rajeunit le système au sens où son intensité juste après T_i est égale à l'intensité initiale à l'instant A_i, $A_i < T_i$.

A_i est **l'âge virtuel du système** après la ième maintenance.

➥ On a : $\forall t \geq 0, \lambda_t = \lambda(t - T_{D_t} + A_{D_t})$.

1.2.2 *Maintenance Préventive*

• Elle est faite pour réduire les défaillance et augmenter la fiabilité

➥ Rend la *collecte de données encore plus compliquée*.

• Augmente la durée de vie *mais fausse* les résultats observés.

• 2 type de maintenance préventives : **conditionnelle et périodique**.

• Ce type de maintenance peut amener à d'autres défaillances non intrinsèques :

➥ mauvais démontage/remontage, détérioration des pièces, ...

• Pour une analyse, il faut obtenir des infos sur les maintenances (type, lieu, date, raisons, détails déroulement, etc.) ➙ Définitions

1.2.3 Liens avec la Modélisation

Pour la modélisation, il faut pouvoir obtenir des maintenances les infos suivantes :

- Les dates de défaillances et des maintenances correctives,

- Les dates de censure (fin collecte ou fin observation)

- Date opération maintenance préventive

- Les dates d'origine : MSI ou les dernières AGAN

- Détails des opérations pour toute opération de maintenance

• Souvent, on s'attend à des estimations **approximatives** du modèle si le temps exact de fonctionnement n'est pas connu (à cause de cette maintenance).

Remarque importante : les maintenances posent des problèmes aux données REX.

➥ Elle n'est jamais réalisée selon les cas idéaux ABAO ou AGAN.

➥ A la limite , on préfère ne pas maintenir un système à l'étude !

1.3 Notions relatives à SdF : MTTF, MTBF, MUT, MDT, MTTR

• Notions de temps relatives à la maintenance

MTTF : mean time to failure = temps moyen de bon fonctionnement jusqu'à l'apparition de la 1e défaillance (ou la défaillance pour un produit non réparable comme une ampoule).

MTBF : mean time between failures : moyenne du temps de bon fonctionnement entre deux pannes (pour un produit réparable).

MTTR : Mean time to repair : moyen du temps de réparation

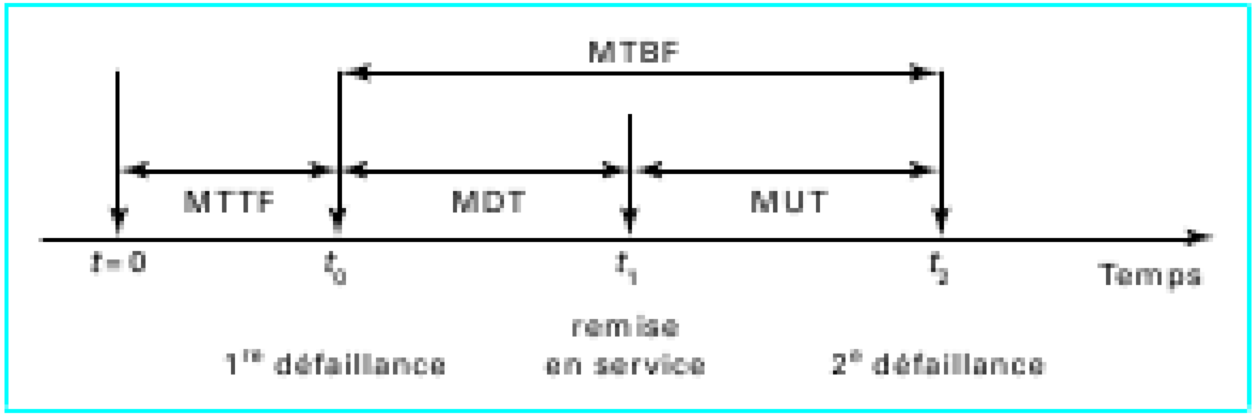
= durée moyenne jusqu'à la réparation d'une unité réparable.

MDT : Mean down time : la moyenne des temps d'arrêt

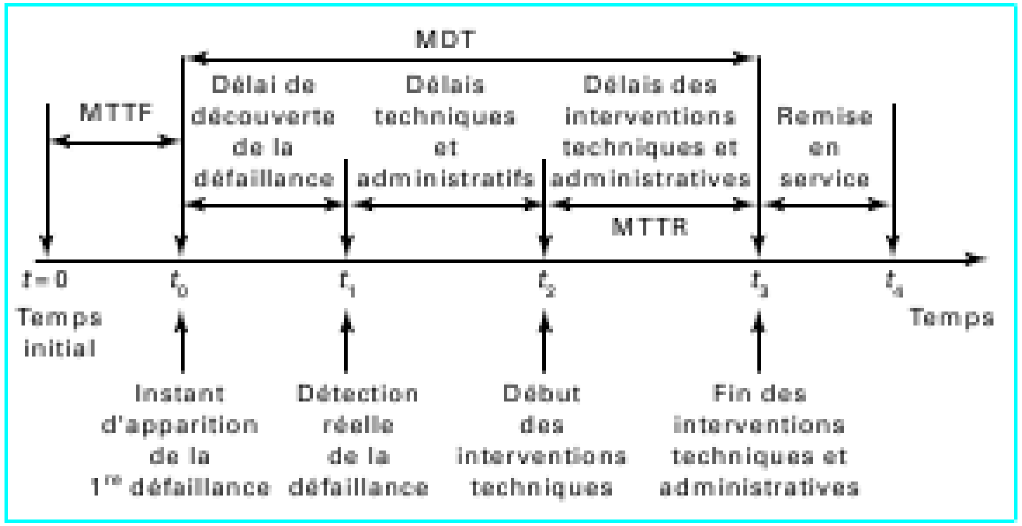
MUT : mean Up time : moyenne des temps de (re) fonctionnement

• **La taux de défaillance constante = 1/MTTF**.

• Les schémas suivants montrent ces relations (MTTR dans la 2e figure) :



Détails **MDT** :



Remarque : ne pas confondre MTTF (mean time to failure) et la durée de vie

➙ mais comme la réciproque d'un taux de défaillance cst.

➙ I.e. Si le MTBF (θ) est connue, alors le taux de défaillance ($\lambda = 1/\theta$)

• MTTF est plus adapté à un matériel non réparable.

Exemple : soit pour un produit (eg. ampoule électrique) , MTTF=1000 heures,

➥ Signifie qu'en moyenne, si 1000 produits sont en circulation, l'un d'entre eux tombe en panne après *une heure* (indép. de la loi de distribution de la défaillance F) !

➥ Si les défaillances du produit devait réellement suivre une distribution exponentielle, alors en moyenne, 63% des produits seront en panne après 1000 heures !

➙ Car (on verra plus loin la loi de fiabilité exponentielle $R(t) = exp(-t/\theta)$)

$R(t) = e^{-\left(\frac{t}{MTBF}\right)}$ et avec $t = MTBF$, $R(t) = e^{-1} = 0.3677$, d'où $F(t) = 1 - R(t) = 0.63$.

Un autre exemple :

• Supposons que 10 dispositifs sont testés pendant 500 heures.

➙ Pendant ces tests, deux des produits tombent en panne.

• Une estimation de MTBF est $\theta = \dfrac{10 * 500}{2} = 2,500$ heures / panne.

➙ Si les 10 étaient tombés en panne après 500h, le MTBF aura été = 500.

• Alors que le MTTF est $\gamma = \dfrac{10 * 500}{10} = 500$ heures / panne

• Si le MTBF est connu, on peut calculer le taux de défaillance comme l'inverse de MTBF.

➙ Ce qui donne pour (le taux de défaillance λ) : $\lambda = \dfrac{1}{\theta} = \dfrac{r}{T}$ avec r=nbr de défaillances.

➙ Pour les 2 pannes ci-dessus, $\lambda = \dfrac{1}{2500} = 0,0004$

1.4 Distribution de survie

1.4.1 Fonction de répartition cumulée de défaillance

• Notée $F(t)$ = la proba d'avoir au moins une défaillance avant le temps t.

Si T = var aléatoire caractérisant l'instant de défaillance : $\boxed{F(t) = Pr(T = t)}$

• On peut obtenir une estimation de la fonction de répartition cumulée (quelque soit la répartition) en affectant un rang aux observations.

Exemple de la méthode **Rang médian** : $F(t) = \dfrac{(j - 0.3)}{(n + 0.4)}$

où j=le rang (ordre temps croissant de pannes depuis le début des observations)

n = nbr total d'observations.

Pour des données multi-censurées, j=un ordre moyen pondéré de défaillance.

• Autre méthode courante : **Rang moyen** $F(t) = \dfrac{j}{n+1}$ et $F(t) = \dfrac{(j - 3/8)}{(n + 1/4)}$

1.4.2 Exemples (Exponentielle)

Consiste en des valeurs $t \geq 0$ et une densité de probabilité : $f(t) = \frac{1}{\theta}\, exp(-\dfrac{t}{\theta}) \quad t \geq 0$.

• Le paramètre $\theta \geq 0$ (de la même nature que t) est appelé **la moyenne de la distribution**.

• La fig svte donne une densité de proba de la distribution continue exponentielle. On a :

➙ $f(t) \geq 0$

➙ $\displaystyle\int_0^{\infty} \frac{1}{\theta}\, exp(-\frac{t}{\theta}) dy = 1$

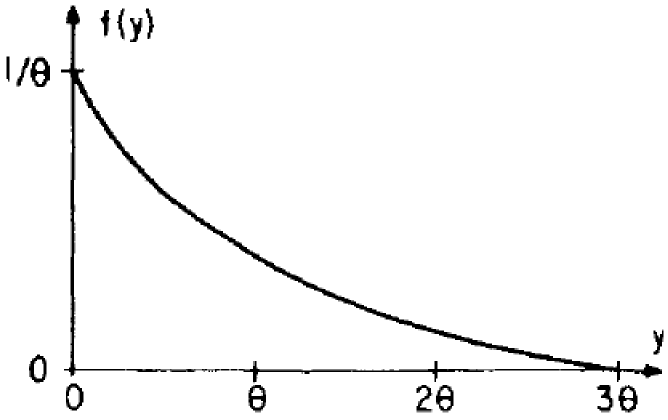


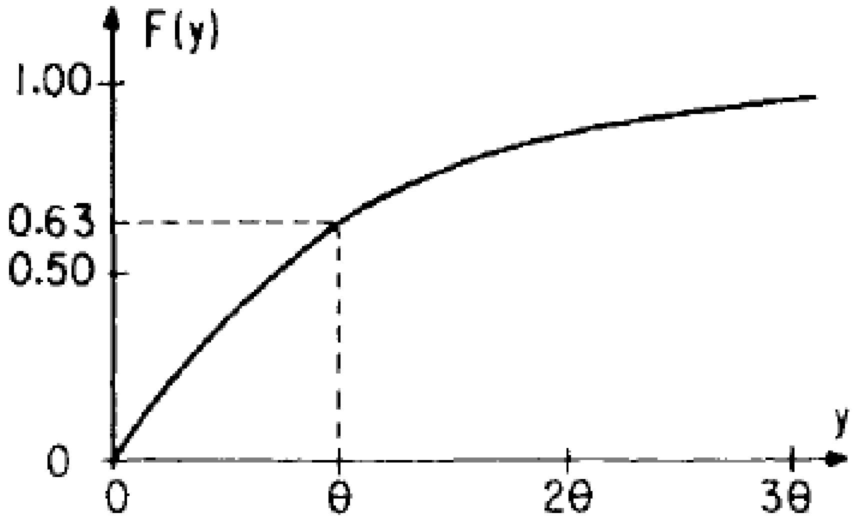
FIGURE 1.5 – Densité de proba Exponentielle

CDF : Fonction de distribution cumulative est

- $F(t)$ pour une distribution continue avec une densité de proba $f(t)$ estimeée par :

$$\boxed{F(t) = P(T \leq t) = \int_{-\infty}^{t} f(u)d(u) \qquad -\infty < t < \infty}$$

Exemple : La figure suivante donne la CDF exponentielle :



1.4.3 *Exemple exponentielle : ventilateurs*

• Une distribution exponentielle avec la moyenne $\theta = 28,700$ heurs a été utilisée pour décrire le nombre d'heures jusqu'à la panne des ventilateurs des moteurs Diesels.

➙ Le taux de défaillance est $\lambda = 1/\theta = 1/28700 = 34.8$ défaillances par millions d'heures.

➙ La probabilité pour un ventilateur de tomber en panne sur 8000 heurs de garantie est

$$F(8000) = P(Y \leq 8000) = \int_0^{8000} \frac{1}{28700} exp(-t/28700) dy = 0.24$$

➥ I.e. 24% de ces ventilos tomberont en panne pendant la garanti.

• Cette information (taux élevé) a permis au constructeur de décider de remplacer les ventilo par un modèle plus fiable.

1.4.4 Fonction de Fiabilité

• Notée $R(t)$, appelée aussi **survivance** ou **fonction de survie**

➥= la probabilité de fonctionnement sans défaillance pendant la période $[0, t]$

➥ = la probabilité de ne pas échouer (ou de survivre) jusqu'à t

• C'est un complément de la fonction de répartition cumulée ($F(t)$).

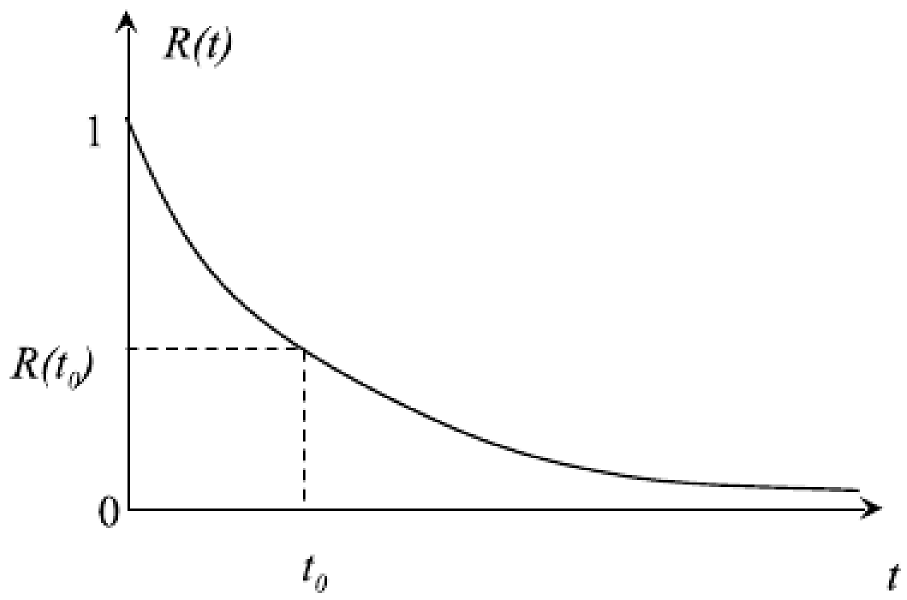
• Si T= la var aléatoire caractérisant l'instant de défaillance, on a : $\boxed{R(t) = 1 - F(t) = Pr(T \geq t)}$

• Pour une loi de défaillance exponentielle :

$$R(t) = 1 - F(t) = exp(-\frac{t}{\theta}) \quad t \geq 0$$

• Pour l'exemple des ventilos, la fiabilité pour 8000 heures est $R(8000) = exp(-8000/28700) = 0.76$.

I.e. 76% de ces ventilos survivent au moins 8000 h.



Corbe de survie ou de Fiabilité (loi exponentielle)

1.4.5 *La densité de probabilité*

• Représente la **densité de défaillance** (à ne pas confondre avec le *taux de défaillance* $\lambda(t)$)

• Notée $f(t)$ correspond à la dérivée de $F(t)$ et représente la proba de la défaillance d'un élément à l'instant t. On a :

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

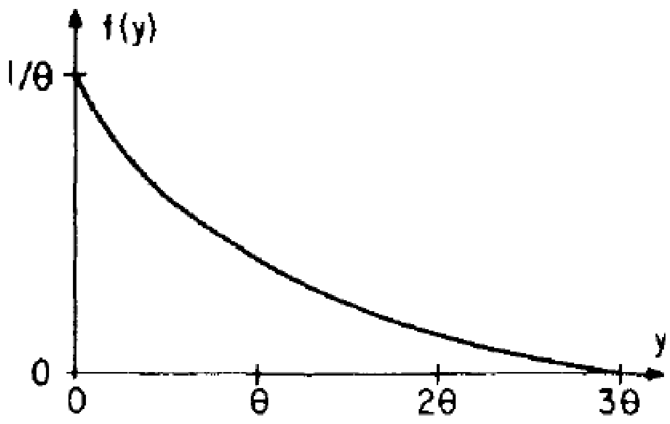
• Habituellement, on n'emploie pas directement $f(t)$ mais son intégrale $F(t)$ qui est la fonction cumulée de densité de proba (CDF).

$f(t)$: PDF of failure ,

$F(t)$: CDF of failure ,

$R(t) = 1 - F(t)$: CDF of survival

Densité de probabilité Exponentielle



1.4.6 *Le taux de défaillance ou la fonction Hasard* $\lambda(t)$

• Permet d'observer la fonction de probabilité de défaillance au cours de <u>très petites</u> <u>périodes</u> (sous l'hypothèse qu'aucune défaillance ne s'est produite avant cette période).

$$\lambda(t)dt = \frac{\textit{Prob(défaillant sur [t ; t + dt] sans défaillance sur [0 ; t])}}{\textit{Prob(non défaillant sur [0 ; t])}}$$

$$\lambda(t)dt = \frac{\textit{Prob(défaillant sur [0 ; t + dt]) - Prob(défaillant sur [0 ; t])}}{\textit{Prob(non défaillant sur [0 ; t])}}$$

• C'est la fonction du taux de **défaillance instantané** $\lambda(t)$ ou **fonction de risque**.

➥ C'est la proba d'être défaillant pendant un instant Δt, à cond. de ne pas l'avoir été jusqu'à t.

On a $$\boxed{\lambda(t) = \frac{F(t+dt)}{R(t)} = \frac{f(t)}{R(t)}} = \frac{-1}{R(t)}.\frac{dR(t)}{d(t)}$$

On en déduit : $R(t) = exp\left[-\int_0^t \lambda(u).d(u)\right]$

• **Pour le cas exponentiel** : $\lambda(t) = \dfrac{f(t)}{R(t)} = \dfrac{\lambda\ exp(-\frac{\lambda}{t})}{exp(-\frac{\lambda}{t})} = \lambda$ constante.

La norme NF *EN 13306* définit le taux de défaillance λ :

"la fréquence d'apparition de la défaillance d'un élément, par période individuelle successives, après que l'élément ait fonctionné jusqu'à un certain temps".

• On distingue :

λ : taux de défaillance en fonctionnement

λ_a : taux de défaillance à l'arrêt

γ : taux de défaillance à la sollicitation

μ : taux de réparation

• On distingue également 3 variantes de λ : défaillance dangereuse (λ^D), sûre (λ^S) et sans incidence sur la sécurité (λ^0).

➥ Ces 3 taux sont déclinés en version détectée / non détectée.

1.4.7 Relations entre R, F, f et λ

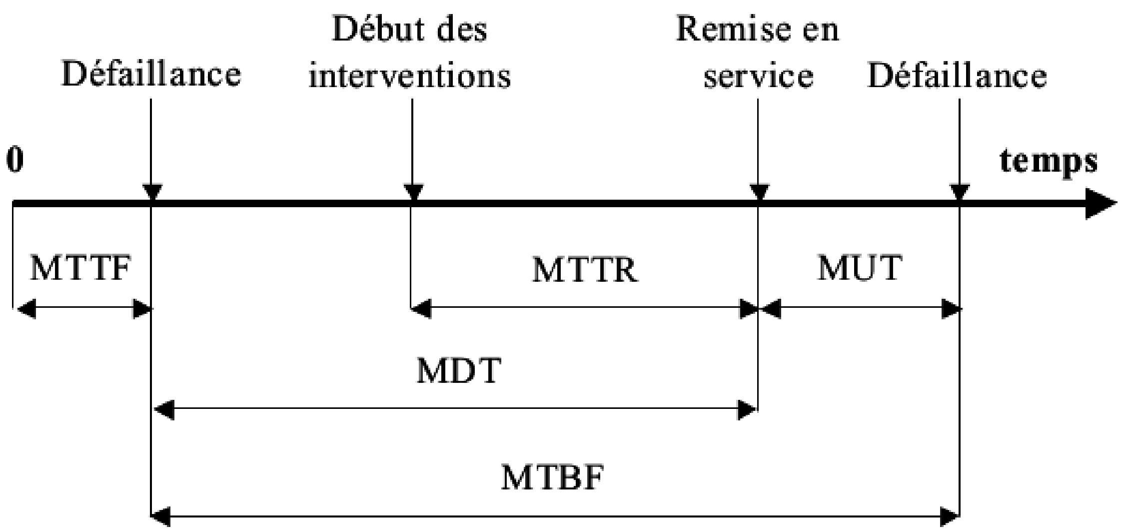
Fonction	$F(t)$	$R(t)$	$f(t)$	$\lambda(t)$
$F(t)$		$1 - R(t)$	$\int_0^t f(u)du$	$1 - \exp - \int_0^t \lambda(u)du$
$R(t)$	$1 - F(t)$		$\int_t^{\infty} f(u)du$	$\exp - \int_0^t \lambda(u)du$
$f(t)$	$\frac{dF(t)}{dt}$	$-\frac{dR(t)}{dt}$		$\lambda(t) \exp \int_0^t \lambda(u)du$
$\lambda(t)$	$\frac{\frac{dF(t)}{dt}}{1 - F(t)}$	$-\frac{R'(t)}{R(t)}$	$\frac{f(t)}{\int_t^{+\infty} f(u)du}$	

1.4.8 *Relations avec les métriques de SdF*

• **MTTF** (Mean time to Failure) : $MTTF = \int_0^{\infty} R(t)d(t)$

• **MTTR** (Mean time to Repair)

• Relation entre : **MUT** (Mean Up time), **MDT** (Mean Down time), **MTBF** (Mean time Bet. Failure) : MTBF = MDT + MUT



1.5 Deux modèles de défaillance courants

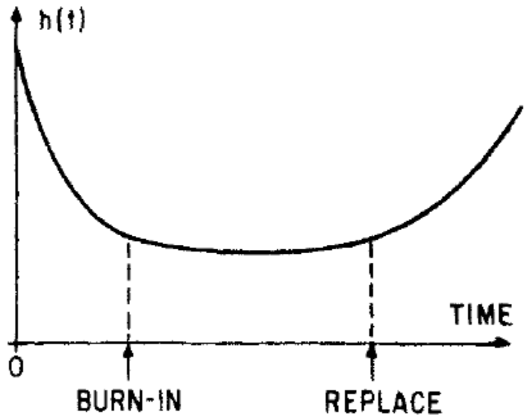
• Exemples de modèles de défaillance couramment utilisés.

NB : β est le taux employé pour **Weibull**.

Loi	Exponentielle	Weibull
$f(t)$	$\lambda \exp - \lambda t$	$\beta \frac{(t-\gamma)^{\beta-1}}{\eta^{\beta}} \exp - \left(\frac{t-\gamma}{\eta}\right)^{\beta}$
$\lambda(t)$	λ	$\frac{\beta}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\beta-1}$

1.6 Cycle de Vie

Décrit le taux de défaillance sous une forme dite "baignoire" :



• **Mortalité infantile** : Un taux **décroissasnt** de la fonction hasard pendant le début de la vie

➙ Un taux décroissant représente la qualité du designe faible ou des défauts de fabrication.

• Un taux **croissasnt** de la fonction du hasard pendant la phase tardive de la vie d'un produit est appelé **défaillance d'usure** .

➙ Un tel taux indique les défaillances dues au viellissement (**usure/fatigu**e) du produit.

➥ C'est le cas de beaucoup de produits.

1.7 Prédiction de Fiabilité de structure

• Prédiction de Fiabilité est basée sur la **fonction de structure** :

➥ l'état d'un système est le produit des états de ses composants, en particulier, si :

$$x_i = \begin{cases} = 1 & \text{Si la composante } e_i \text{ est en bon état} \\ = 0 & \text{Sinon.} \end{cases}$$

alors l'état du système est

$$y = \prod_{i=1}^{n} x_i \qquad \text{pour une structure en série et}$$

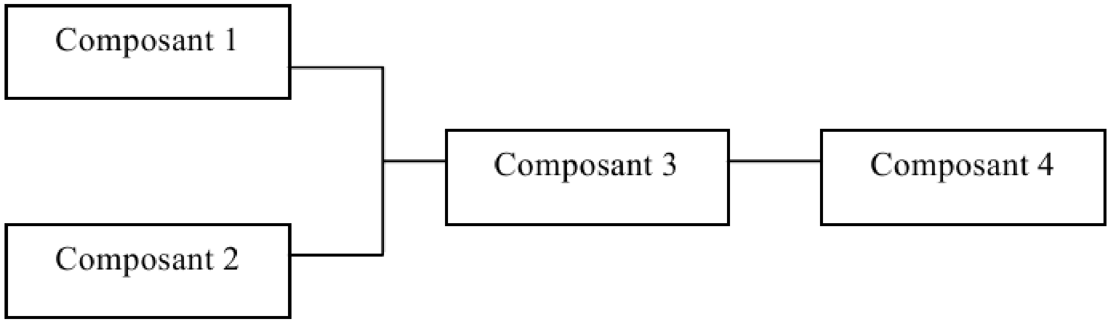
$$y = 1 - \prod_{i=1}^{n} (1 - x_i) \qquad \text{pour une structure en parallèle .}$$

• Le comportement du système est basé sur celui de ses composantes

➥ En tirer les fonctions de Fiabilité

1.7.1 *Système en série / Parallèle*

• Fiabilité et Structure (architecture)



$$R_{sys} = [1 - \prod_{i=1}^{2}(1 - R_i)] \times R_3 \times R_4 = [1 - (1 - R_1) \times (1 - R_2)] \times R_3 \times R_4$$

Si on a un taux de défaillance constant et une fiabilité modélisée par une loi exponentielle, alors : $R_{sys} = [1 - (1 - e^{-\lambda_1 t}) \times (1 - e^{-\lambda_2 t})] \times e^{-\lambda_3 t} \times e^{-\lambda_4 t}$

• Rappel : $F_{sys} = 1 - R_{sys}$. ➙ Suite

• Si les λ_i sont des variables aléatoires, alors R_{sys} devient aléatoire (par composition des lois de probabilités).

➥ Par exemple, pour le cas ci-dessus (si λ_i aléatoires), on peut utiliser la Convolution pour calculer la somme de 2 variables aléatoires (avec défaillance connue).

➥ Si $Z = A + B$, alors la densité de proba de Z sera $g_z(y) = \int_{-\infty}^{+\infty} f_A(u) \times f_B(y-u)\, du$

• Et si $Z = A.B$, alors la densité de proba de Z sera $g_z(y) = \int_{-\infty}^{+\infty} f_A(u) \times f_B(\frac{y}{u}) \times \frac{du}{|u|}$

➥ Intégrales difficiles à calculer, sauf pour additionner des lois Normales (pour A et B) ou multiplier des lois log-normales.

➥ Si les lois différentes, alors calsuls complexes !

• D'ou la simulation (MCL) .

1.7.2 *Estimation de Fiabilité des systèmes complexes*

• Connaissant tous les instants de défaillance, on peut faire une estimation de la fiabilité :

- du système ;

- de chaque fonction ;

- de chaque composant en fonction du temps système.

• Une fois que les lois de fiabilité sont connues, il est possible d'estimer la fiabilité du système en considérant sa structure.

• Le diagramme série correspond au cas le plus pessimiste du calcul de la **fiabilité**.

$$R_{syst}(t) = \prod_{i=1}^{c} \prod_{j=1}^{k_i} R_{ij}(t)$$

où c : le nombre des composants ;

k_i : le nombre des **modes de défaillance** pour le composant i.

Fiabilité d'une fonction :

• La loi de **fiabilité pour une fonction** est donnée par la relation :

$$R_{fonc}(t) = \prod_{i=1}^{c} \prod_{j=1}^{k_i} R_{ij}(t)^{\Psi(i)}$$

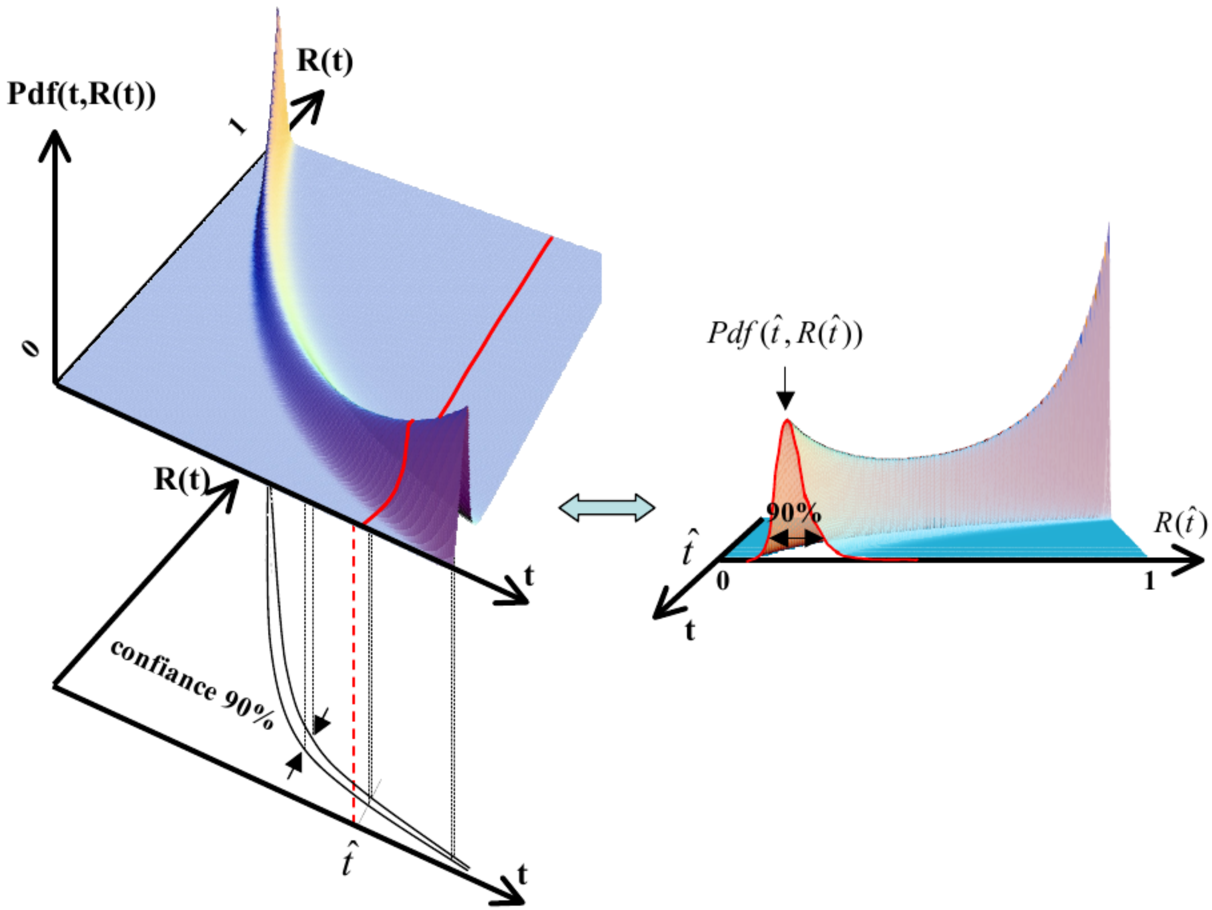
où c : le nombre des composants ;

k_i : le nombre des **modes de défaillance** pour le composant i.

$$\Psi(i) = \begin{cases} = 1 & \underline{\text{Si la composante}}\ i\ \underline{\text{participe à la fonction}} \\ = 0 & \underline{\text{Sinon.}} \end{cases}$$

• On estime en général les courbes d'intervalle de confiance sur la fonction de **fiabilité**.

• Un exemple :



1.8 Quelques distributions courantes

Pourquoi une distribution ? :

Rappel : on dispose de données de survie et l'on souhaite y ajuster un modèle paramétrique

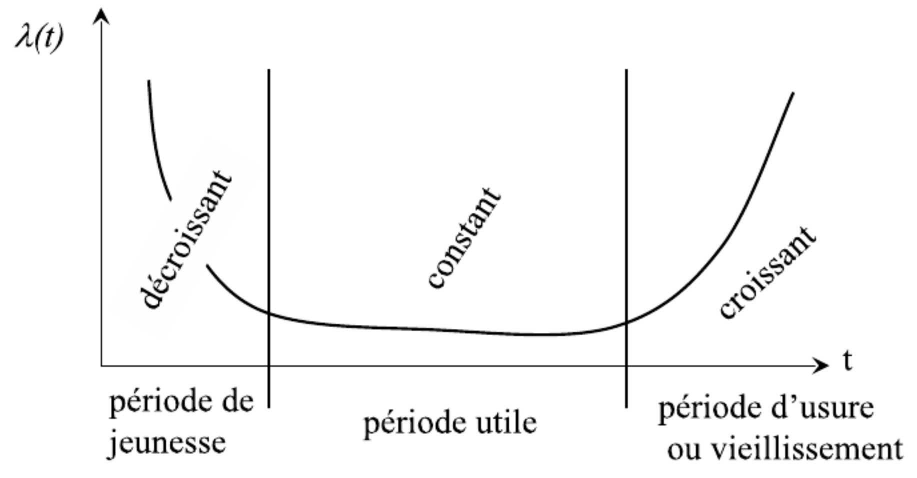
➞ la question du choix d'une distribution se pose.

• A partir de la fonction de **densité de défaillance** ($f(t)$), on obtient la fonction de **répartition** ($F(t)$), la fonction de **Fiabilité** ($R(t)$) et le **taux** de défaillance(λ_t).

➞ Exemple : Courbe en *baignoire*

• Il existe de multiples lois de fiabilité

➡ On cite les plus utiles.



1.8.1 *Loi de Gompertz : historique**

• Une des premières lois, portée sur les populations humaines (a plus d'un siècle).

• Utilisée pour d'autres populations statistiques et en Fiabilité (électroniques, corrosion, etc).

• Gompertz faisait l'hypothèse que la mort peut être provoquée par 2 causes distinctes :

➙ *Un hasard* : accident subit avec une intensité indépendante du temps

➙ *Affaiblissement* (usure) dans le temps.

➥ Elle correspond alors (la force de mortalité = la fonction du hasard h) à une quantité φ qui s'accroît dans le temps d'une quantité $d\varphi$ proportionnelle à φ.

• Ce modèle établit un taux de défaillance qui se compose de termes indépendants de l'âge (négligeable et abandonné ici) et des termes dépendants de l'âge (= fonction de Gompertz).

➥ Le taux de défaillance (λ en fiabilité) :

$$\frac{1}{\varphi}.\frac{d\varphi}{dx} = q = Cste$$

➥ d'où $\varphi = a.e^{q.x}$

• On exprime ensuite cette loi sous sa forme la plus courante par

$V_x = k.g^{c^x}$: nombre de survivants d'âge x à un instant donné

Où x = l'âge, k = nbr d'individus à l'origine

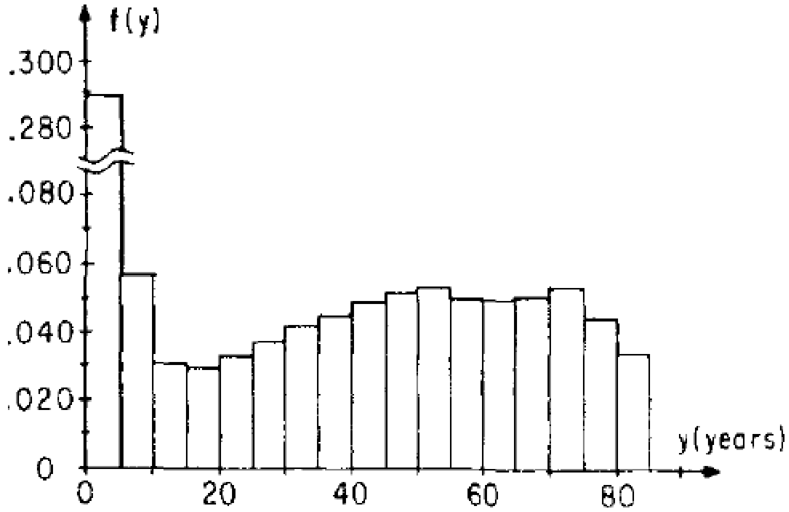
g et c caractérisent la loi de survie.

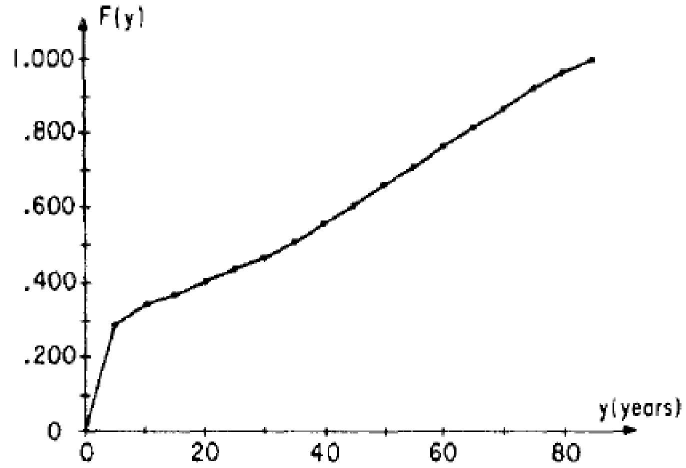
• Dans le tableau suivant, la colonne y représente des tranches d'age, $f(y)$ est la probabilité de cette tranche. Pour les autres colonnes, voir la suite.

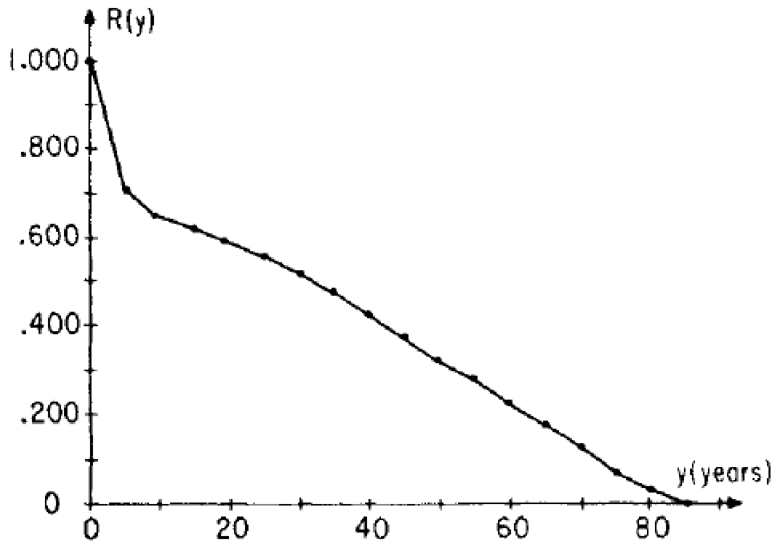
y	f(y)	F(y)	R(y)	h(y)
0	-	0	1.000	
0-5	.290	.290	.710	.058
5-10	.057	.347	.653	.016
10-15	.031	.378	.622	.010
15-20	.030	.408	.592	.010
20-25	.032	.440	.560	.011
25-30	.037	.477	.523	.013
30-35	.042	.519	.481	.016
35-40	.045	.564	.436	.019
40-45	.049	.613	.387	.022
45-50	.052	.665	.335	.027
50-55	.053	.718	.282	.032
55-60	.050	.768	.232	.035
60-65	.050	.818	.182	.043
65-70	.051	.869	.131	.056
70-75	.053	.922	.078	.081
75-80	.044	.966	.034	.113
80-85	.034	1.000	0	.200

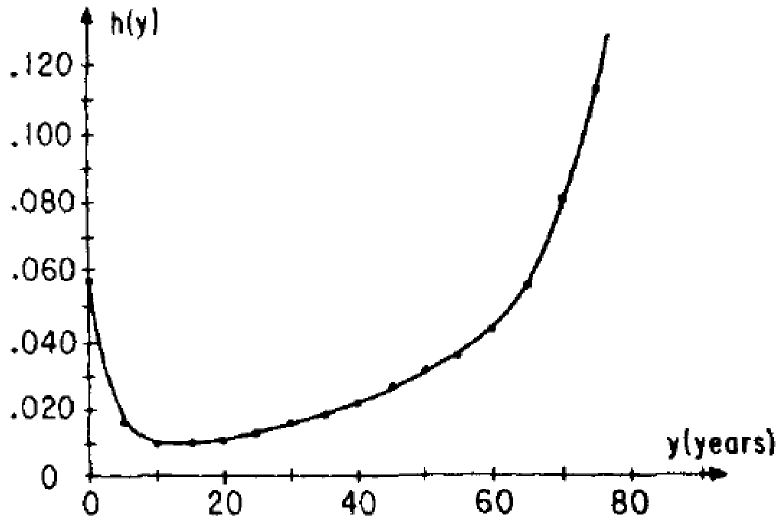
FIGURE 1.6 – Table des données de mortalité humaines de **Halley-1693** (tous les 5 ans)

• L'histogramme et la CDF $F(t)$, $R(t)$ et $h(t)$ de la table ci-dessus.









1.8.2 Loi Normale (ou Laplcae-Gauss)

- Représente bien la **fin de vie** des dispositifs subissant un phénomène de **vieillessement**.
- Loi continue et symétrique $N(\mu, \sigma^2)$
- On a :

$$f(t) = \frac{1}{\sigma\sqrt{2.\pi}}.e^{-\frac{1}{2}.\left(\frac{t-\mu}{\sigma}\right)^2} \qquad F(t) = \int_{-\infty}^{t} f(t).d(t)$$

$$E(t) = \int_{-\infty}^{\infty} t.f(t)dt = \int_{0}^{\infty} R(t)dt = \mu \qquad Var(t) = E[(t-\mu)^2] = E[t^2] - E^2[t] = \sigma^2$$

<u>Remarque</u> : si t suit une loi Normale $N(\mu, \sigma)$, $u = \frac{t-\mu}{\sigma}$ suit une loi normale centrée réduite dont la fonction de répartition ϕ est donnée par : $\phi(u) = \frac{1}{\sqrt{2.\pi}} \int_{-\infty}^{u} .e^{-\frac{1}{2}x^2}\, dx$

1.8.3 *Loi Log-Normale ou Galton ou Gibrat*

• Représente bien la **mortalité** ou la durée de réparation des matériels.

• S'applique bien lorsque les observations (de volume important) faites sont les conséquences d'un effet **multiplicatif de différentes causes indépendantes et aléatoires**.

• Représente bien les phénomènes de **fatigue et d'usure** en mécanique

$$f(t) = \frac{1}{t.\sigma\sqrt{2.\pi}}.e^{-\frac{1}{2}.\left(\frac{ln(t)-\mu}{\sigma}\right)^2} \qquad \lambda(t) = \frac{e^{-\frac{1}{2}\left(\frac{ln(t)-\mu}{\sigma}\right)^2}}{t\int_0^{\infty}\sigma\sqrt{2\pi}\ f(t)d(t)}$$

$$E(t) = \int_0^{\infty} t.f(t)dt = e^{(\mu+\frac{\sigma^2}{2})} \qquad Var(t) = \int_0^{+\infty}[t.E(T)]^2.f(t).dt = e^{(2\mu+\sigma^2)}.(e^{\sigma^2}-1)$$

Remarque : une variable aléatoire continue et positive t est distribuée selon une loi **lognormale** si son logarithme est distribué suivant une loi **normale**.

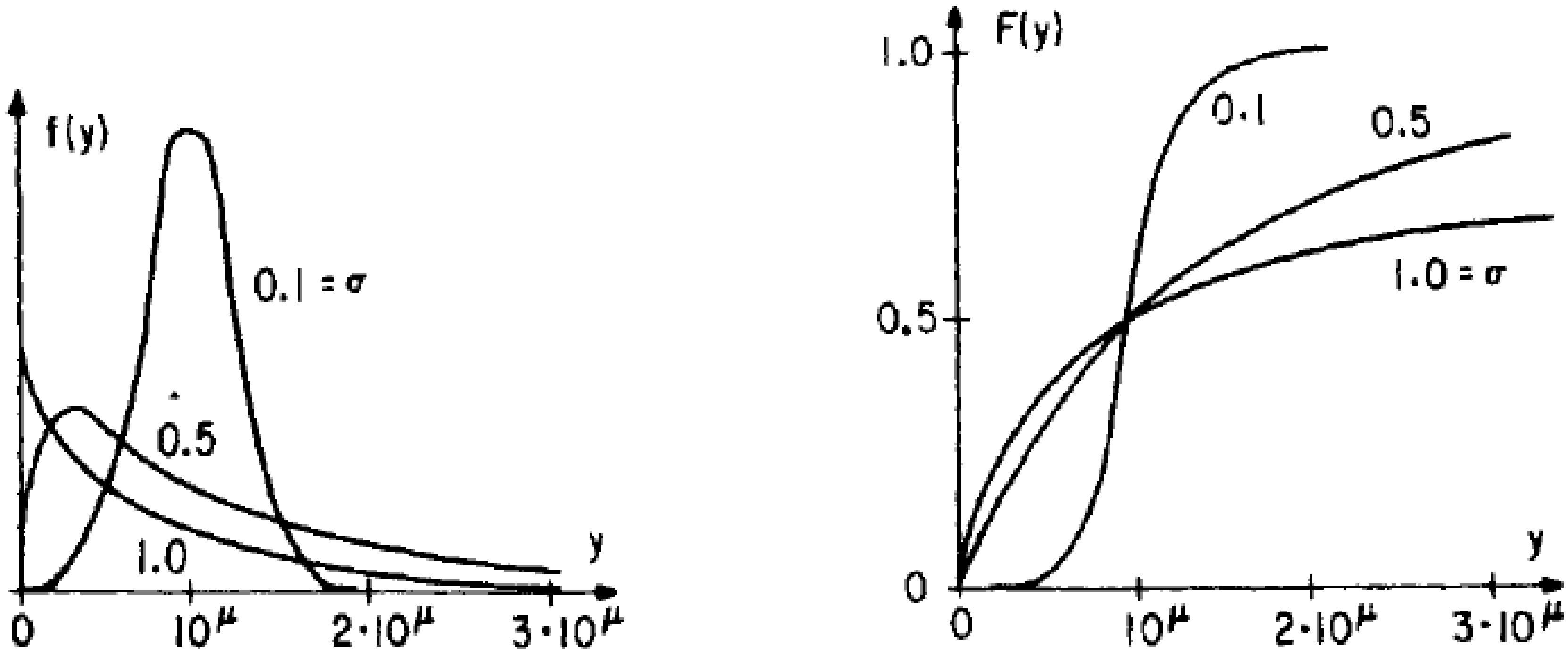


FIGURE 1.7 – $f(t)$ et $F(t)$ de la loi log-Normale.

1.8.4 Loi Gamma

- Représente la **loi d'occurrence de β événements dans un processus Poissonien**.
- Souvent utilisée pour modéliser le temps de défaillance
  - ☞ Si t_i est le temps entre les défaillances successives d'un système, et t_i suit une distribution exponentielle, le temps cumulé d'apparition des défaillances suit une loi Gamma (conjuguaison Exp(=PDF) – Gamma(=CDF)).
- Sera utilisée comme une distribution a priori dans l'approche Bayesienne
- Représente aussi le **comportement réel** d'un matériel depuis sa mise en service jusqu'à son rebut ➙ représente les 3 phases de vie (voir Weibull).

• Les 3 paramètres habituels de la loi (α, β, η) souvent restreints à 2 en posant $\eta = 1/\alpha$) où α représente la forme (shape) et β le paramètre échelle. ➙ donnent de la souplesse à la loi.

$$f(t) = \frac{1}{\alpha^{\beta}.\Gamma(\beta)} t^{\beta-1}.e^{-\frac{t}{\alpha}} \qquad \texttt{avec} \qquad \Gamma(\beta) = \int_0^{\infty} t^{\beta-1}.e^{-t}dt = (\beta-1)! \quad si\ \beta\ entier$$

NB. : pour $\alpha = 2, \beta = v/2$, v entier= degrés de liberté, Gamma devient χ^2

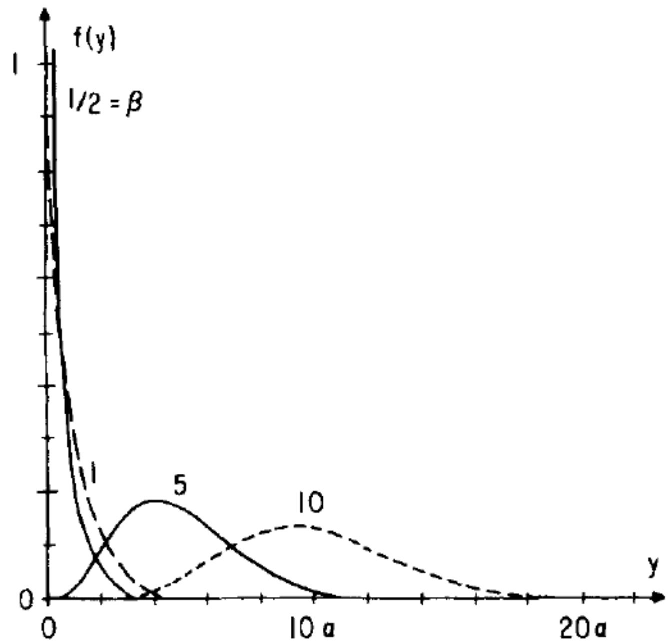
➙ E.g. pour $v = 2$, Gamma est une distribution Exponentielle avec $\mu = 2$

• $\lambda(t) = \dfrac{1}{\alpha^{\beta}.\displaystyle\int_t^{\infty} \Gamma(\beta).f(u).du}$

• $E(T) = \dfrac{\alpha}{\beta}$

• $Var(T) = \dfrac{\alpha}{\beta^2}$

• Pour $\beta = 1$, la loi Gamma devient Exponentielle



1.8.5 *Loi Exponentielle*

• Très **fréquemment** utilisée en Fiabilité (mécanique, électronique).

• En particulier pendant la période de **vie utile** où le taux de défaillance instantané λ est constante (comme en électronique) et les défaillances sont aléatoires (associé au processus poissonien) .

➥ Très utilisée en électronique, décrit le temps écoulé jusqu'à une défaillance, ou l'intervalle de temps entre deux défaillances.

• Elle est utilisée également dans le cas de <u>défaillance brutale</u>, ou pour des <u>systèmes</u> <u>complexes</u> <u>composés</u> avec des lois de fiabilité <u>élémentaires</u> différentes.

$$f(t) = \lambda . e^{-\lambda . t} \qquad \lambda(t) = \lambda \qquad R(t) = e^{-\lambda . t} \qquad E(T) = \frac{1}{\lambda} \qquad Var(T) = \frac{1}{\lambda^2}$$

• Remarque : la loi exponentielle représente un cas particulier (conjugué) de Gamma.

1.8.6 *Loi Beta*

• Loi à 2 paramètres, très utile dans l'évaluation de la durée des **essais de fiabilité**

➥ Représente la proba pour qu'un matériel survive au moins jusqu'à un temps t, lorsque plusieurs matériels sont testés simultanément.

• Très utilisée en statistique **bayesienne** comme distribution a priori pour la proba d'un événement suivant une distribution **Binomiale** (ainsi qu'en contrôle de qualité).

$$f(t) = \frac{1}{\beta(n,p)}.t^{n-1}.\,(1-t)^{p-1} \quad = \quad \frac{\Gamma(x+y)}{\Gamma(x).\Gamma(y)}.t^{x-1}(1-t)^{y-1}$$

$$E(t) = \frac{n}{n+p}$$

$$Var(t) = \frac{n.p}{(n+p+1)(n+p)^2}$$

• Avec la fonction

$$\beta(x,y) = \int_0^1 t^{x-1}.(1-t)^{y-1}.dt = \frac{\Gamma(x).\Gamma(y)}{\Gamma(x+y)}$$

Rappel : $$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1}.e^{-t}dt = (\alpha-1)! \quad pour\ \alpha\ entier.$$

➥ la fonction $\Gamma(c) = (c-1)!$

• Voir exemple d'introduction en séance 1.

1.8.7 Loi Weibull

• Distribution générale et puissante qui permet de décrire le temps d'échec.

• Les paramètre γ, β, η :

γ est la position (ou localisation) : le décalage entre le début de l'observation (de panne) et du processus = le plus petit temps de défaillance possible (souvent = 0).

β est le paramètre de la forme (shape). C'est le plus important car il joue sur la variation du taux de défaillance et permet ainsi la modélisation des 3 phases :

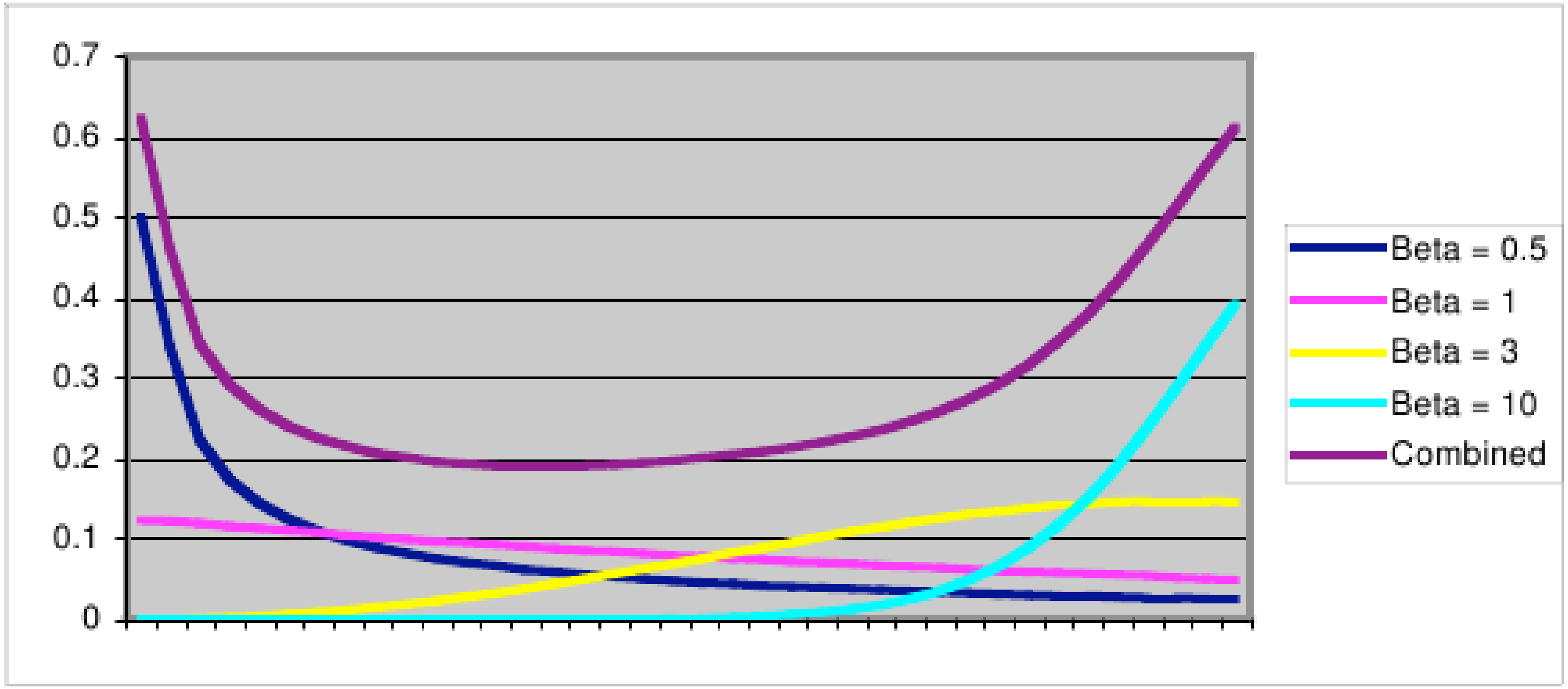
➥ $\beta < 1$: période de jeunesse. La fiabilité décroît très rapidement puis lentement.

➥ $\beta = 1$: période de maturité (➝ loi exponentielle)

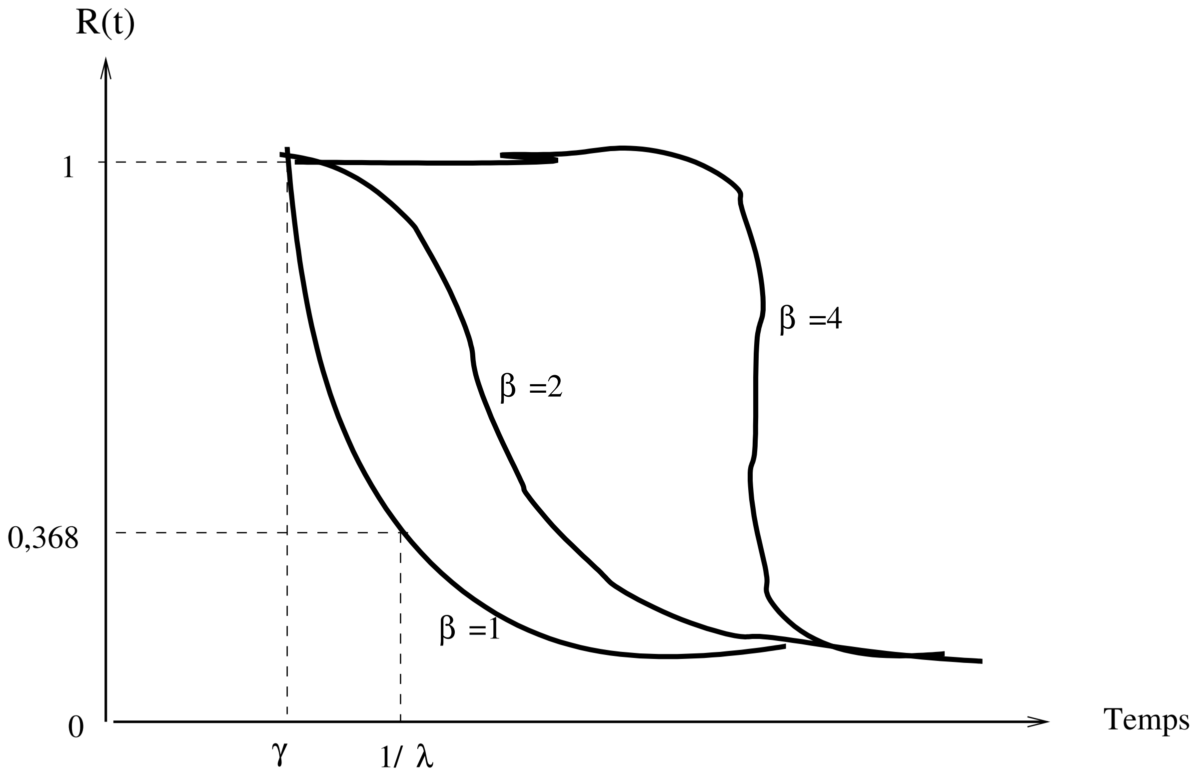
➥ $\beta > 1$: période de vieillesse ou d'usure.

➥ La chute de fiabilité est d'abord lente puis monte fort après un certain moment.

η est le paramètre d'échelle de temps (MTBF). Si $t - \gamma = \eta$ alors $F(t) = 63\%$.



On note que si toutes les courbes sont **combinées**, le résultat est proche d'une courbe **Baignoire** (Bath Hub)



• Remarque : si $\beta = 1$ et $\gamma = 0$, on a une loi exponentielle.

• Le paramètre de forme β permet également de donner des indications sur le type de défaillance.

Par ex, β compris entre 2 et 3 nous place dans un phénomène d'usure.

• Loi très souple, représentative d'une très grande variété de phénomènes aléatoires.

• Souvent utilisée en fiabilité mécanique du fait de sa capacité à prendre en compte le vieillissement.

$$f(t) = \frac{\beta}{\eta}.\left(\frac{t-\gamma}{\eta}\right)^{\beta-1}.e^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}} \qquad \lambda = \frac{\beta}{\eta}.\left(\frac{t-\gamma}{\eta}\right)^{\beta-1}$$

$$F(t) = 1 - e^{-(t/\eta)^{\beta}} \qquad R(t) = e^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}}$$

$$E(T) = \gamma + \eta.\Gamma\left(1 + 1/\beta\right)$$

$$Var(T) = \eta^2.\left[\Gamma\left(1+2/\beta\right) - \Gamma^2\left(1+1/\beta\right)\right]$$

• Remarque générale sur les lois : t n'est pas toujours le temps, peut être KM ou cycle.

➥ On devrait plutôt utiliser <u>période ou cycle</u>.

1.8.8 Birnbaum-Saunders

• **Birnbaum-Saunders** ou la **loi de la fatigue** très utilisée :

➙ bien adaptée à un mode de défaillance (fissure) dû au phénomène de fatigue.

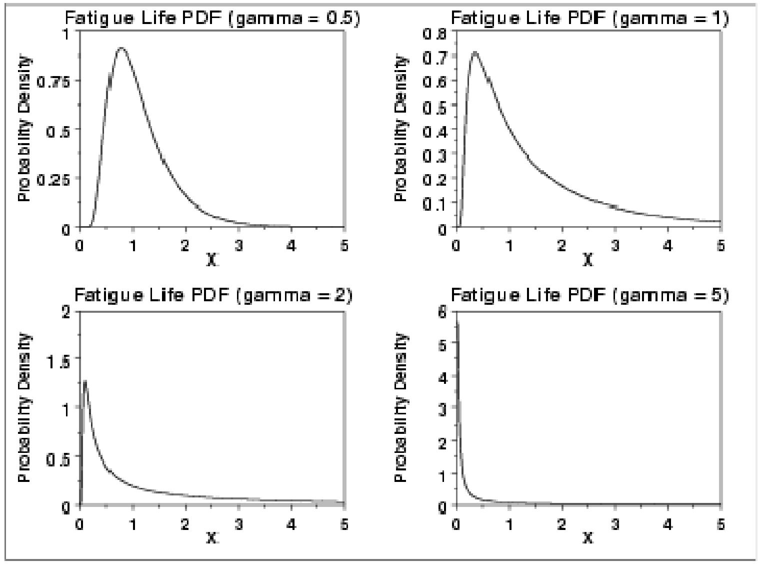
➥ Cela la rend spécifique et réduit son champs d'action.

• Existent sous différentes formulations.

• La **forme standard** de la loi Birnbaum-Saunders :

$$f(t) = \frac{\sqrt{x}+\sqrt{1/x}}{2\gamma x}\,\phi\left(\frac{\sqrt{x}-\sqrt{1/x}}{\gamma}\right) \quad x>0, \gamma>0$$

γ : shape, ϕ = PDF de la loi Normale



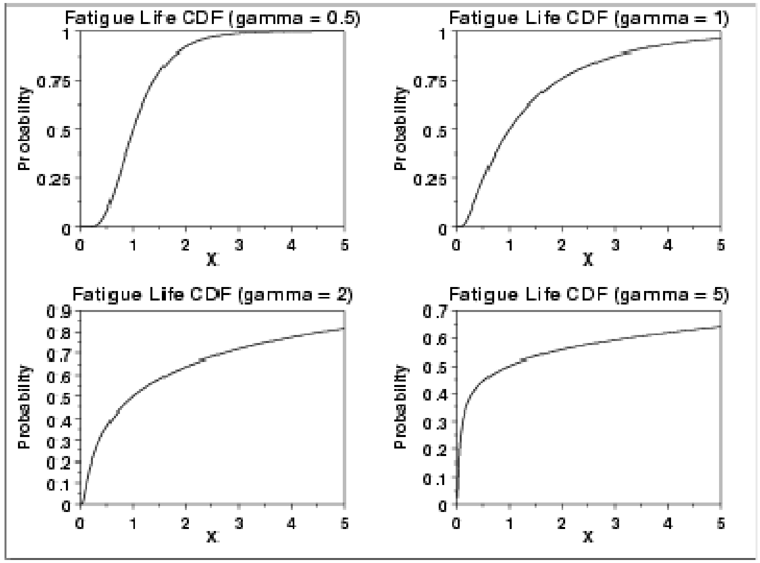
PDF de la loi Birnbaum-Saunders avec différentes valeurs de γ

• La CDF de la loi Birnbaum-Saunders :

$$F(x) = \Phi\left(\frac{\sqrt{x} - \sqrt{1/x}}{\gamma}\right) \quad x > 0, \gamma > 0$$

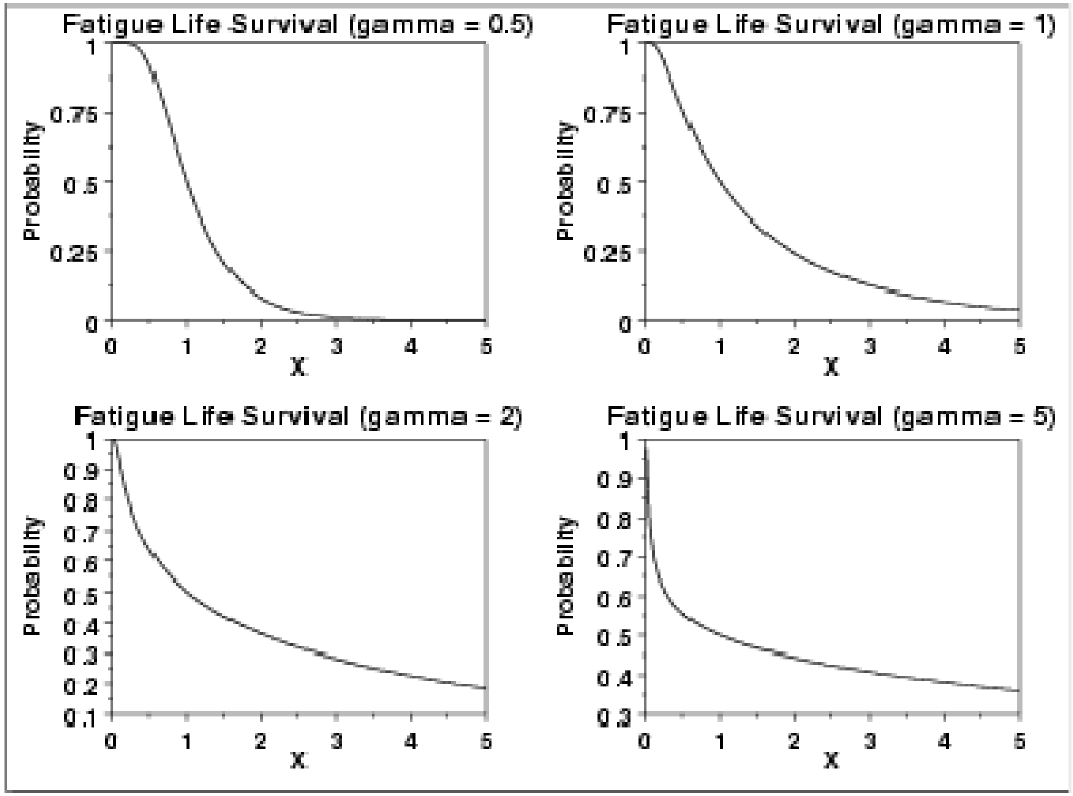
Φ = CDF de la loi Normale

γ : shape



CDF de la loi Birnbaum-Saunders avec différentes valeurs de γ

• La loi de survie ($R(t)$ pour différentes valeur de γ :



1.8.9 *Autres lois*

• La loi **uniforme** : souvent utilisée en fiabilité pour les essais bayésiens en l'absence connaissances pour construire l'information *a priori* (voir la suite).

Cette loi peut prendre toute valeur dans un intervalle (a, b) avec une densité de probabilité constante.

➥ la fonction de répartition : $F(t) = \dfrac{t-a}{b-a}$

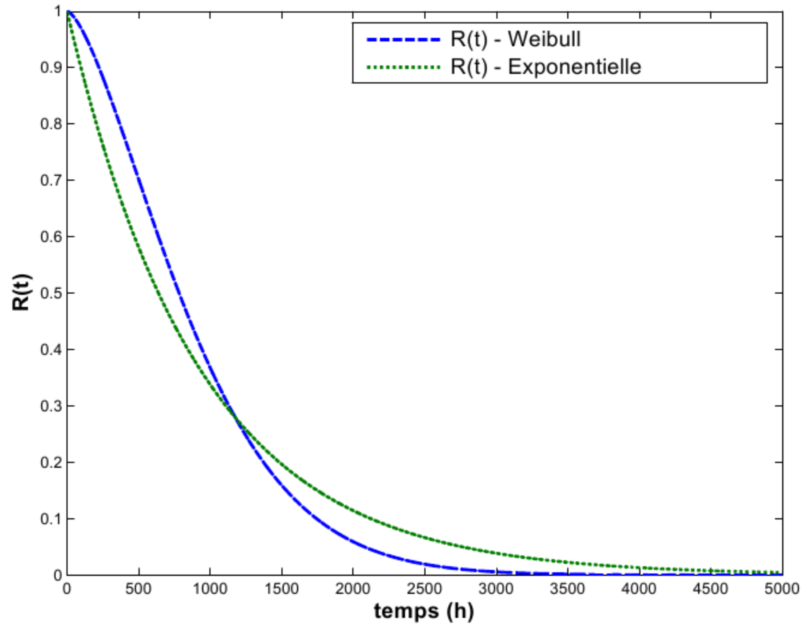
➥ la densité de probabilité : $f(t) = \dfrac{1}{b-a}$

• Autres lois existent pour des cas particuliers...

1.8.10 *Ex : Simulation de modèles dysfonctionnels (Weibull et Exp)*

• Tirage (MCL) de 1000 valeurs aléatoires d'instants de défaillance selon une loi de Weibull ($\beta = 1.5, \eta = 1000h$) (loi caractéristique) des composants mécaniques.

• A partir des instants de défaillance, on estime les paramètres de Weibull et la valeur moyenne du taux de défaillance.



• On utilise ces estimations pour représenter les lois de fiabilité Weibull et Exponentielle (avec le même taux de défaillance moyen λ_{moy}).

• La figure ci-dessus montre l'écart entre les deux lois.

➥ Weibull est plus générale et ici on a une simulation des paramètres.

Correction de l'estimation : on définit les paramètres à l'aide des recueils de données (et de l'AMDEC, voir ci-dessous), η recalculé par REX.

➥ Le taux de défaillance moyen donne une information sur le paramètre d'échelle η et l'AMDEC sur le paramètre de forme β qui est associé au mode de défaillance.

➥ Expertise : pour β, on peut utiliser la correspondance suivante entre les valeurs de β et modes de défaillance :

Mode de défaillance	Valeurs de β
Fatigue	$\beta \in [1.5; 2.5]$
Usure, corrosion	$\beta \in [3; 4]$

N.B. **AMDEC** : Méthode Prévisionnelle d'Analyse qualitative des Modes de Défaillance de leurs criticités dans le but d'Évaluer les Conséquences des défaillances

• Le paramètre d'échelle η est estimé en considérant le **taux de défaillance moyen** qui est **l'inverse du MTBF** : $$\eta = \frac{1}{\lambda_{moy}.\Gamma(1+\frac{1}{\beta})}$$

➥ N.B. : ici γ = le décalage =0

• En reprenant l'exemple précédent, on obtient une loi de Weibull estimée correcte.

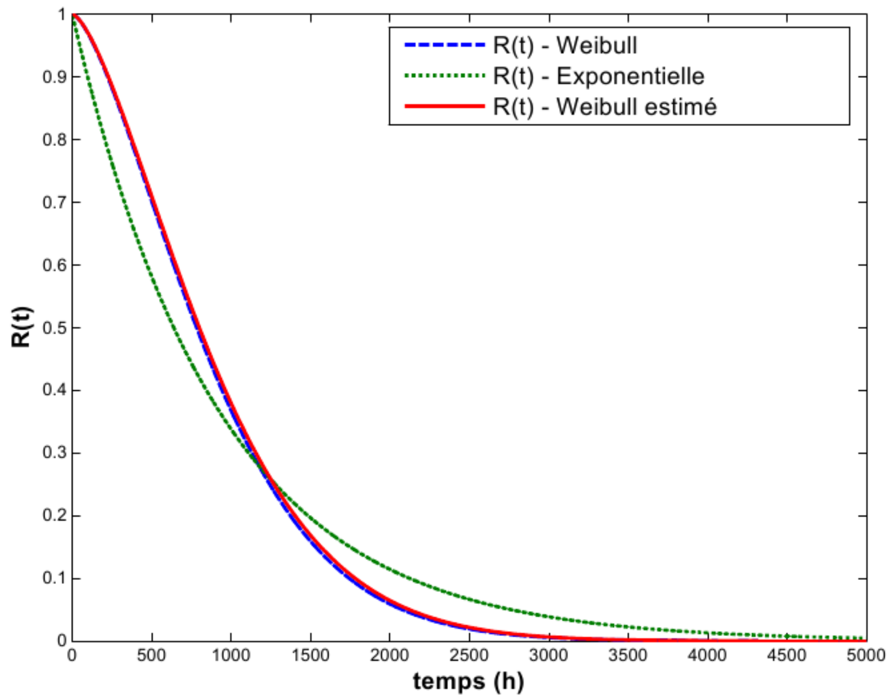


FIGURE 1.8 – Comparaison des lois de Weibull, Exponentielle et Weibull estimé

1.8.11 *Bilan : Choix d'une distribution**

Rappel : on dispose de données de survie et l'on souhaite y ajuster un modèle paramétrique

➙ la question du choix d'une distribution se pose.

Bilan des distributions :

• **Gamma** : adaptée à un taux de défaillance monotone mais sa fonction de répartition ne s'obtient pas analytiquement.

- Permet de modéliser toutes les phases, comme Weibull.

- Elle est **conjugué** (pour l'approche Bayesienne) à une loi exponentiel : facilite grandement les calculs (inférence, itérations, intégrales, ..., comme Beta avec Binomiale).

• **Weibull** : souvent bien adaptée mais complexe (donc pas la plus utilisée).

- Elle a l'avantage de décrire alternativement les **3 phases, à l'aide de** β.

- **Couvre la loi Exponentielle** et approche la loi Normale pour β proche de 3,6.

- De plus, Weibull couvre un grand nombre de distributions empiriques.

• **Log-Normale** : pas pratique car associée à un taux de défaillance nul à l'origine puis croissant au maximum puis tend vers 0 pour de grandes valeurs de t,

➙ ce qui n'est pas vrai en pratique, ni en électronique, ni en mécanique.

• **Birnbaum-Saunders** ou la **loi de la fatigue** très utilisée :

➙ bien adaptée à un mode de défaillance (fissure) dû au phénomène de fatigue.

1.8.12 Où en sommes nous ? : la suite

• Une fois le modèle choisi, le problème d'estimation de ses paramètres se pose.

➥ 2 possibilités :

- La valeur ponctuelle du paramètre recherché : facile à manipuler

- **L'intervalle de confiance** qui contient le paramètre recherché et reflète la confiance accordée à l'estimation.

1.9 Estimateurs : Méthode ponctuelle

- Deux méthodes :
  - ➙ maximum de vraisemblance (optimisation de la vraisemblance) et
  - ➙ la méthode de moindre carrés (minimisation de l'erreur quadratique).
    - ➥ Pas traité ici.

1.9.1 *Estimation par la méthode du max de vraisemblance*

• Méthode académique, efficace au sens statistique

• Est asymptotiquement sans biais : l'espérance de l'estimateur est égale au paramètre estimé de manière efficace (la variance de l'estimateur est faible).

• La distribution de l'échantillonage est <u>normale</u>.

➥ Il est donc facile de donner un intervalle de conf avec le paramètre estimé par cette méthode.

• Ci-dessous, on étudie cette méthode avec les différents type d'échantillons

1.9.2 *Maximum de vraisemblance avec échantillon complet*

• Si on dispose d'un échantillon T de n variables aléatoires, indépendantes de durée de vie T (donc, tous les n éléments ont été défaillants), la loi de proba du n-échantillon

$$V_{(t_1,t_2,...,t_n,\theta)} = \prod_{i=1}^{i=n} f_{(t,\Theta)} = V(T,\Theta)$$

• On choisit ensuite l'estimateur $\hat{\Theta}$ de Θ qui rend maximale la densité de probabilité du n-uple $(T_1, T_2, ..., T_3)$.

d'où
$$\frac{\partial V(T,\Theta)}{\partial \Theta} = 0$$

• Si plusieurs paramètres Θ : $(\Theta_1, \Theta_1, ...)$ à estimer, ils seront les solutions du système

$$\frac{\partial V(T,\Theta_1)}{\partial \Theta_1} = 0, ..., \frac{\partial V(T,\Theta_i)}{\partial \Theta_i} = 0$$

../..

• On vérifiera que $\dfrac{\partial V^2(T,\Theta_i)}{\partial\Theta_i^2} \neq 0$ et $\dfrac{\partial V^2(T,\Theta_i)}{\partial\Theta_i^2} < 0$ pour un maximum.

• L'estimateur obtenu est Asymptotiquement non biaisé, efficace, obtenu par la règle du maximum de vraisemblance (si elle existe).

• NB : pour faciliter les calculs, on peut utiliser le *log-vraisemblance* :

$$L_{(T,\theta)} = ln\, V_{(T,\theta)}$$

Avec

$$\frac{\partial L(T,\Theta_1)}{\partial\Theta_1} = 0, ..., \frac{\partial L(T,\Theta_i)}{\partial\Theta_i} = 0$$

1.9.3 *Maximum de vraisemblance avec échantillon incomplet*

La cas précédent : n données sur des éléments <u>tous défaillants</u>.

➥ Ce cas <u>n'est pas le cas le plus courant dans l'industrie</u>.

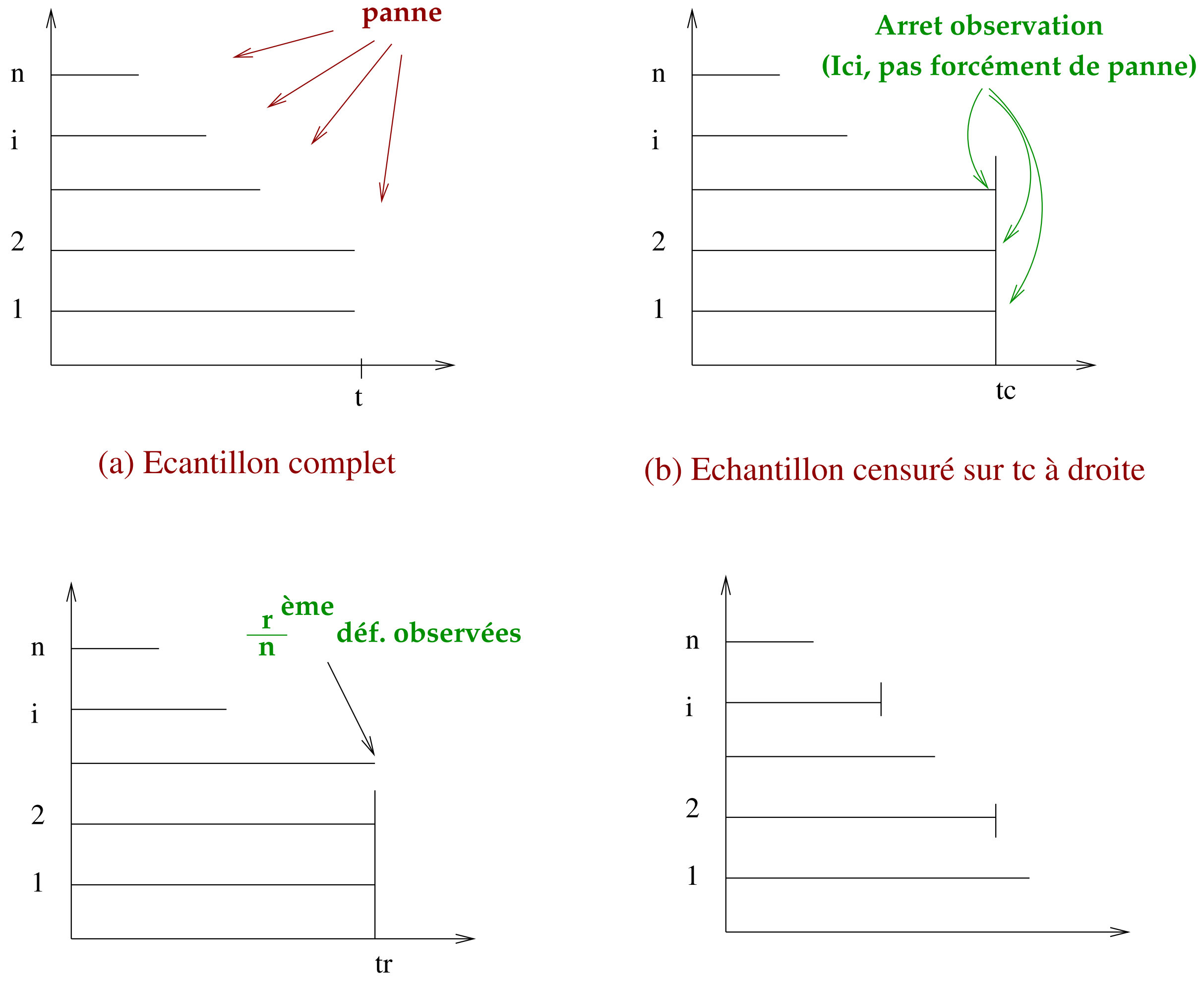
• Il arrive plus fréquemment que l'on arrête un essai au bout d'un certain temps t_c, sans que tous les éléments soient totalement défaillants.

➥ On a donc un échantillon censuré (**à droite type I**)

• Les données REX sont souvent censurées, avec des instants de censure différents.

➥ Voir figure suivante pour un rappel des différents types de censure.

• Rappel : différents type de données :



(a) Ecantillon complet

(b) Echantillon censuré sur tc à droite

(c) Censuré sur rème défaillance

(d) Censuré aléatoirement à droite

1.9.4 *Vraisemblance et Données censurées sur le temps tc, type I (b)*

• On arrête l'observation au temps tc précisé d'avance.

- Sur un n-échantillon, on a $k < n$ défaillants.

- Correspond aux démos de fiabilité où les essais s'arrêtent à une date fixe.

- La densité conjointe s'écrit alors (couples indépendants) :

$$V_{((y_1,\delta_1),(y_2,\delta_2),\ldots,(y_n,\delta_n))\theta)} = \prod_{i=1}^{i=n} R^{1-\delta_i}(y_c, \Theta).f^{\delta_i}(y_i, \Theta)$$

Où $R = 1 - F$ est la fiabilité (des éléments <u>non défaillants</u>)

$\delta_i = 0$ si l'élément est censuré (non tombé en panne)

$\delta_i = 1$ si l'élément est non censuré (tombé en panne)

1.9.5 *Données censurées sur la r^{eme} défaillance, type I (c)*

• Arrête des observations au bout de r défaillances, r précisé d'avance.

• Correspond en pratique aux essais dont le but est de démontrer un % de défaillances

• La fonc. de vraisemblance des éléments défaillants est réalisée par la PDF ($f(t)$) ;

➥ pour les non défaillants, au temps t_r, on utilise la fonction de fiabilité ($R(t)$) .

• La fonction de vraisemblance = probabilité conjointe d'observer le n-uple d'échantillon : $\{(t_1, \delta_1 = 1), ..., (t_r, \delta_r = 1), (t_r, \delta_{r+1} = 0), ..., (t_r, \delta_{r+2} = 0), (t_r, \delta_n = 0)\}$

Avec δ = indicateur de censure (0 : censuré).

On a la vraisemblance (statistique d'ordre) :

$$V_{(t_1,t_2,...,t_n,\theta)} = \frac{n!}{(n-r)!}(R(t_r))^{n-r}\prod_{i=1}^{i=r} f_{(t_i,\Theta)}$$

Autres cas de censures :

• Cas censure aléatoire : compliqué,

• Méthode d'estimation par Intervalle ... compliquée ... pas en métiers !

• Est-ce fini ? non !

• Voyons un exemple concrêt de prévision

Puis la méthode statistique Bayesienne

1.10 Modélisation Bayesienne

• **Constat** : difficulté d'obtenir des données fiables en Fiabilité ainsi que les infos sur les équipements.

• **Motivations** : obtenir un taux de défaillance pertinent à partir du matériel similaire (renseigné par des BDF) + calculs (REX).

• La méthode Bayesienne permet de combiner des données hétérogènes .

• Rappels : dans une démarche bayesienne, on a besoin de :

- **A priori** : donnée par les experts du domaine mais avec une marge d'incertitude.

- **La vraisemblance** à partir des données recueillies

➥ Donneront $\boxed{\textbf{A posteriori} \propto \text{A priori} \times \text{Vraisemblance}}$.

• La méthode permet de prédire, inférer, faire une meilleure estimation de l'*a priori*.

Dans cette démarche :

• Ce qui est **innovant** ici (pour déterminer un taux de défaillance) :

Utiliser des données REX des BDs diverses, élaborées mondialement.

➥ Les données locales (à pondérer) reflètent des conditions d'utilisation.

• Certaines données ont besoin d'être pondérées, par exemple selon le degré de similarité du matériel.

➥ Il faut alors pouvoir quantifier l'information apportée par chaque source.

• Le calcul du taux de défaillance est complété par un intervalle de tolérance.

1.10.1 Hypothèses faites pour la Bayesienne

1. Taux de défaillance constante du fait de la méconnaissance de la chronologie des défaillances.
2. Les maintenances préventives sont supposés être efficaces.
3. Les sources, une fois pondérées, ont la même qualité d'information

• En mécanique, les temps des défaillances matériels suivent souvent une **loi exponentiell**e, directement associée à un processus **Poissonien** (défaillance aléatoires, à taux constant).

➥ La loi Weibull est plus complexe à mettre en oeuvre . ../..

• On différencie les matériels en **fonctionnement continu** du matériel **fonctionnant à la sollicitation**.

• Pour les matériels dont la sollicitation est continue (= fonctionnement continu), la **vraisemblance** est modélisée par une loi **exponentielle**.

• Pour les systèmes qui fonctionnent à la sollicitaion (à la demande), la **vraisemblance** est modélisée par une loi **Binomiale** (cf. l'exemple précédent)

➥ Cette loi décrit un phénomène ayant des occurrences s'excluant mutuellement (états défaillant ou non défaillant), donc **Binomiale**.

Rappels du théorème Bayesien : **a** posteriori $\propto$ a priori $\times$ vraisemblance

L'a priori ne devrait pas être déduite des données (si possible).

➥ *A priori* sera la loi de distribution donnée par les experts (constatée).

• Eléments importants (cas réels) :

On distingue les cas suivants et selon les données disponibles pour équilibrer les calculs entre la vraisemblance et l'a priori ➙ *A posteriori* :

I) Sources peu informatives : on privilégie la vraisemblance.

II) Sources plus informatives : on augmente l'importance de l' a priori

III) Sources très informatives : on peut trouver tous les paramètres de la distribution effective.

1.10.2 *Résumé du calcul Bayésien*

On étudie les différents modèles utilisés en fonction du type de fonctionnement et le degré d'information des sources :

• Fonctionnement en continu :

➙ distribution vraisemblance **Exponentielle**

➙ Calcul de la probabilité **a priori** :

➥ Méconnaissance (source) : uniforme non informative

➥ Peu de connaissance : uniforme informative (deux valeurs extrêmes)

➥ Bonne connaissance : **Gamma**

N.B. utilisation de principe de Conjuguaison Exponentielle - Gamma

../..

- Fonctionnement à la sollicitaion :

  ➙ distribution vraisemblance **Binomiale**

  ➙ Calcul de la proba **a priori** :

  ➥ Méconnaissance (source) : uniforme non informative

  ➥ Peu de connaissance : uniforme informative (deux valeurs extrêmes)

  ➥ Bonne connaissance : loi **Bêta**

  Conjuguaison Binomiale - Bêta

Les deux exemples suivants examinent les 2 cas. ../..

1.11 Vers le calcul Bayesien

1.11.1 Vraisemblance (L) de défaillance en fonctionnement continu

Selon les hypothèses de départ, la vraisemblance modélisée par une loi **exponentielle** :

$$L(n,t|\lambda) = \prod_{i=1}^{n} f(t_i,\lambda) = \prod_{i=1}^{n} \lambda.e^{-\lambda t_i} = \quad \lambda^n.e^{-\lambda T}$$

Où n = nombre de défaillances,

λ = taux de défaillance en continue,

T = temps de bon fonctionnement cumulé ($T = \sum_{i=1}^{n} t_i$).

• On distingue maintenant le degré d'information disponible ➝ **a priori**

1.11.1.1 Méconnaissance : non informative

• Les différentes sources de données n'apportent pas d'info sur le taux de défaillance.

➥ On utilise la loi uniforme non informative :

$$g(\lambda) = \begin{cases} \dfrac{1}{\lambda_0}; & \lambda_0 \in]0; \infty] \\ 0, & \lambda_0 = 0 \end{cases}$$

Le théorème de Bayes s'écrit alors : (Rappel $g(\lambda|T) = \dfrac{g(T|\lambda).g(\lambda)}{\sum ... = g(T)}$)

$$g(\lambda|T) = \frac{\lambda^k.e^{-\lambda.T}}{\displaystyle\int_0^{\lambda_0} \lambda^k.e^{-\lambda.T}.d\lambda}$$

Après simplifications, on obtient

$$g(\lambda|T) = \frac{T^{k+1}.\lambda^k.e^{-\lambda.T}}{\Gamma(k+1, \lambda_0.T)}$$

L'espérance de cette distribution = **l'estimateur du taux de défaillance** :

$$\hat{\lambda} = \frac{\Gamma(k+2, \lambda_0.T)}{T.[\Gamma(k+1, \lambda_0.T)]}$$

Et sa variance :

$$V(\lambda|T) = \frac{\Gamma(k+3, \lambda_0.T).\Gamma(k+1, \lambda_0.T) - \Gamma^2(k+2, \lambda_0.T)}{T^2.\Gamma^2(k+1, \lambda_0.T)}$$

1.11.1.2 Peu de connaissance a priori

• Le déficite de connaissances impose de modéliser l'a priori par une loi uniforme informative.

• On est dans le cas d'un intervalle borné et toutes les valeurs possibles de la probabilité sont équiprobables sur cet intervalle.

$$g(\lambda) = \begin{cases} \dfrac{1}{\lambda_1 - \lambda_0} & ; \lambda \in [\lambda_0, \lambda_1] \\ 0 & ; \lambda \notin [\lambda_0, \lambda_1] \end{cases}$$

• La forme générale de Bayes :

$$g(\lambda|T) = \frac{\dfrac{1}{\lambda_1 - \lambda_0}.\lambda^k.e^{-\lambda.T}}{\displaystyle\int_{\lambda_0}^{\lambda_1} \dfrac{1}{\lambda_1 - \lambda_0}.\lambda^k.e^{-\lambda.T}.d\lambda} = \frac{T^{k+1}.\lambda^k.e^{-\lambda.T}}{\Gamma(k+1, \lambda_1.T) - \Gamma(k+1, \lambda_0.T)}$$

Et $\hat{\lambda} = \dfrac{\Gamma(k+2, \lambda_1.T) - \Gamma(k+2, \lambda_0.T)}{T.[\Gamma(k+1, \lambda_1.T) - \Gamma(k+1, \lambda_0.T)]}$

1.11.1.3 Bonne connaissance a priori

• On est dans le cas le plus informatif :

➥ Modèle de la distribution **Gamma** pour le taux de défaillance final en continu.

• La loi Gamma est intéressante dans ce cas car elle permet de représenter les 3 phases (courbe en baignoire) d'un système :

• L'hypothèse d'une vraisemblance **exponentielle** permet de pratiquer le principe de conjuguaison.

$$g(\lambda) = \frac{\lambda^{\alpha-1}.e^{-\dfrac{\lambda}{\beta}}}{\beta^{\alpha}.\Gamma(\alpha)}$$

La défaillance a posteriori en obtenant les paramètres de la loi Gamma finale :

$$g(\lambda|T) = (\frac{1}{\beta} + T)^{\alpha+k}.\lambda^{\alpha+k-1}.e^{-\lambda(\dfrac{1}{\beta}+T)}.\frac{1}{\Gamma(\alpha+k)}$$

1.12 Un exemple Poisson-Gamma

Un exemple pour le fonctionnement continue avec comptage du nombre de défaillances.

• On compte le nombre de défaillances pendant une période de temps spécifiée.

➙ On pourra ensuite réparer la panne ou remplacer l'élément défaillant

• On peut avoir recours à ce type de données si l'on a peu de capteurs ou bien le système n'enregistre que des pannes récentes (par ex. sur un mois sulement).

• Pour ce type de situation, le modèle *Poisson* de base est :

$$f(y|\lambda) = \frac{(\lambda t)^y \, exp(-\lambda t)}{y!} \quad \lambda > 0, y = 0, 1, 2...$$

➙ y est le nbr de défaillances

➙ λ est le nbr moyen de défaillances par unité (intervalle) de temps

➙ t est le temps de l'observation (spécifié d'avance).

➥ Pour les modules réparables, λ sera une intensité.

• On notera que l'égalité de la moyenne et de la variance (ici λt) est une limitation pour l'application de la distribution Poissonienne.

• Pour le model Posonnien, λ (le nbr moyen de défaillances par intervalle de temps) est le paramètre à estimer.

• L'équation ci-dessus qui est une fonction de λ sachant le nbr de défaillance y observé est une fonction de **vraisemblance** appropriée pour le décompte du nbr de défaillances.

• S'il y a les décomptes $y_1, y_2, ..., y_n$ pendant les périodes de temps $t_1, t_2, ..., t_n$, alors, sous la condition de l'indépendance et un λ constante, la fonction de vraisemblance consiste en le produit des m fonctions de vraissemblance individuelles spécifiées par l'équation ci-dessus.

• Pour compléter l'analyse, on doit préciser la distribution *a priori* pour λ.

Distribution ***a priori*** **pour des données poissoniennes**

• Une distribution souvent utilisée pour λ le nbr moyen de défaillances par unité de temps vu dans l'équation ci-dessus est la distribution **Gamma**.

➙ Une raison principale : elle est la conjuguée de la distribution a priori pour λ et possède un support positif : la distrib. a priori Gamma et la vraissemblance Poisonnienne ont la même forme de telle sorte que la distrib. résultant a posteriori pour λ est aussi une distribution de Poisson.

• Si on a $Y_i \sim Poisson(\lambda t_i),\ i = 1, ..., n$ où y_i sont les différents décomptes de défaillances

Et que la distribution a priori pour λ est $\lambda \sim Gamma(\alpha, \beta)$

Alors la distribution posteriori pour λ est $\lambda|y \sim Gamma(\alpha + \sum_{i=1}^{n} y_i\ ,\ \beta + \sum_{i=1}^{n} t_i)$

Avec $y = (y_1, y_2, ..., y_n)$

➙ β est la taille a priori (prior sample size) vs. taille des données (data sample size) qui est $\sum_{i=1}^{n} t_i$;

➙ α est le nbr a priori des défaillances vs. le nombre observé des défaillances $\sum_{i=1}^{n} y_i$.

Exemple : Modèle de Poisson appliqué aux pannes des supercomputers

• Soit la modélisation du nbr mensuel des pannes des composantes des super ordinateurs du laboratoire *Blue Mountain* du centre*Los Alamos* par une distribution de Poisson.

• Un super ordi possède 47 processeurs identiques et la table ci-dessous donne le nbr mensuel des défaillances pour le premier mois du fonctionnement.

1	5	1	4	2	3	1	3	6	4	4	4
2	3	2	2	4	5	5	2	5	3	2	2
3	1	1	2	5	1	4	1	1	1	2	1
3	2	5	3	5	2	5	1	1	5	2	

• Soit $y_1, y_2, ..., y_{47}$ le nbr de défaillances meneuelles enregistrés pour les processeurs.

➙ Avec $t_i = 1$ mois, on modélise le décompte du nbr de défaillance par l'équation vue ci-dessus :

$Y_i \sim Poisson(\lambda t) = Poisson(\lambda),\ i = 1, ..., 47$ avec λ : nbr moyen de défaillances mensuelle.

A priori : les ingénieurs estiment qu'il ne doit pas y avoir plus de 10 défaillances par composante dans le 1er mois.

➙ Une manière de représenter cette connaissance *a priori* est de supposer une distribution *a priori* **Gamma** pour $\lambda = 5$ (la moyenne de 0 et 10).

➙ On représentera cette info a priori par $\lambda \sim Gamma(\alpha = 5, \beta = 1)$.

➙ Noton qu'avec la distribution $Gamma(5, 1)$, la proba pour que λ dépasse 10 est de 0.03.

• En utilisant l'équation $\lambda|y$ ci-dessus pour la distribution a posteriori de λ sachant les données $y = y_1, y_2, ..., y_{47}$, on a

$$\lambda|y \sim Gamma(\alpha + \sum_{i=1}^{n} y_i \ , \ \beta + \sum_{i=1}^{n} t_i) = Gamma(5 + 132 \ , \ 1 + 47) = Gamma(137 \ , \ 48)$$

• La figure suivante représente les 2 distributions pour la moyenne du nbr de défaillances λ.

• On remarque que la *a priori* (pointillé) est diffuse alors que la *a posteriori* (solide) est piquée et indique que le décompte des défaillances apporte une <u>évidence substentielle</u> pour une moyenne inféiruere (à 5) estimée par les ingénieurs.

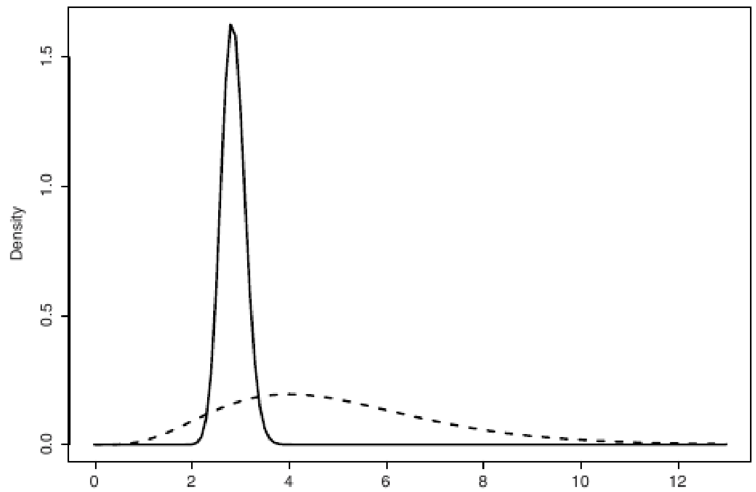


FIGURE 1.9 – pointillé : a priori et solide : a posteriori pour la moyenne λ

• Le nbr moyen de défaillances a posteriori est $E(\lambda|y) = \alpha^*/\beta^* = 137/48 = 2.85$

• L'écart type posterior est $\sqrt{Var(\lambda|y)} = \sqrt{\alpha^*/\beta^{*2}} = 0.24$

• L'intervalle de confiance à 95% est $(2.40, 3.35)$ pour le nbr de défaillances mensuelles.

• NB. : pour confirmer que la distribution Poisson modélise bien le problème, on pourrait appliquer un test χ^2 de goodness-of-fit.

1.13 Probabilité de défaillance à la sollicitation

• La distribution recherchée est celle du nombre k de défaillances sur N sollicitaions.

➥ La probabilité de défaillance p suit une distribution **Binomiale** que l'on utilisera pour la fonction de **vraisemblance**.

$$f(k|p) = \binom{N}{k}.p^k.(1-p)^{N-k} \quad \text{avec N connu et } p \in [0,1]$$

• Les 3 cas informatifs précédents se présentent.

➥ Pour obtenir l'a priori.

1.13.1 *Méconnaissance : loi uniforme non informative*

$$g(p) = \begin{cases} 1; & p \in [0,1] \\ 0; & p \notin [0,1] \end{cases}$$

On obtient la formule de Bayes :

$$g(p|k) = \frac{p^k.(1-p)^{N-k}}{\displaystyle\int_0^1 p^k.(1-p)^{N-k}.dp} = ... = \frac{p^k.(1-p)^{N-k}}{\beta(k+1, N-k+1)}$$

➥ C'est donc une loi **Bêta** de paramètres $b = N - k + 1$ et $a = k + 1$

L'espérance : $E(P|k) = \dfrac{k+1}{N+2}$ et sa variance $Var(p|k) = \dfrac{(k+1).(N-k+1)}{(N+3).(N+2)^2}$

1.13.2 *Peu de connaissance a priori*

Similaire au fonctionnement continu :

$$g(p) = \begin{cases} \dfrac{1}{p_1 - p_0} & ; p \in [p_0, p_1] \\ 0; & p \notin [p_0, pa_1] \end{cases}$$

On obtient la formule de Bayes :

$$g(p|k) = \frac{p^k.(1-p)^{N-k}}{\displaystyle\int_{p_0}^{p_1} p^k.(1-p)^{N-k}.dp}$$

L'intégration donnera .....(exprimée par Gamma Incomplète)

1.13.3 *Bonne connaissance a priori*

• Loi **Bêta** pour l'apriori

➥ Elle est conjuguée de la même famille que la vraisemblance **Binomiale**.

$$g(p) = \frac{p^{a-1}.(1-p)^{b-1k}}{\beta(a,b)}$$

On obtient la formule de Bayes :

$$g(p|k) = \frac{p^{a+k-1}.(1-p)^{b+N-(k+1)}}{\beta(a+k,b+N-k)}$$

1.14 Un exemple Binomiale-Beta

Analyse des données de fiabilité du type succès/échec. (rappel de la séance 1)

➙ Succès/échec des composants d'un système plus complexe.

• L'objectif : données + hypothèse (*a priori*) ➙ prédiction (*a posteriori*)

• Dans cet exemple : données disponibles à propos de générateurs d'urgence (diesel) en dépannage dans une centrale nucléaire.

➙ On a observé leur comportement (démarrage ou non) lors d'une sollicitation (en urgence).

☞ N.B. : un autre exemple similaire (succès/échec) peut être les résultats à l'allumage de missiles (militaires) construits par différents fabricants. Voir plus loin.

• Ce type de données est traitée avec une distribution **Binomiale**.

☞ La *Binomiale* est appropriée dans le cas de test d'un nombre fixe n de composants testés, où les tests sont supposés indépendants étant donné une proba de succès π .

• La fonction de densité de probabilité dans ce cas : $f(x|n,\pi) = \begin{pmatrix} n \\ x \end{pmatrix} \pi^x \, (1-\pi)^{n-x}$

Où $0 \leq \pi \leq 1$ = la proba de succès.

☞ NB. 1 : la loi binomiale n'est pas adaptée si les tests ne sont pas indép ; De plus, elle ne s'applique que si tous les événements ont la même probabilité de succès.

☞ NB. 2 : la distrib de **Bernoulli** est un cas spécial de la Binomial pour $n = 1$.

☞ NB. 3 : Parfois, on utilise le temps d'échec (défaillance) comme une donnée de type succès/échec par rapport à un temps spécifié t.

➙ C'est à dire : x sera le nbr de défaillances parmi n items testés avant le délai t .

➙ La raison principale pour cela est d'éviter un modèle de temps défaillance (continue temporel) ;

Mais dans ce cas, il y a beaucoup de perte d'information

➥ on ne l'utilise pas comme cas général.

Précisions :

• Pour le model Binomial, la probabilité π est le paramètre inconnu que l'on veut estimer.

• Dans un model succès/échec, l'équation $f(x|n,\pi) = \begin{pmatrix} n \\ x \end{pmatrix} \pi^x \, (1-\pi)^{n-x}$

est comme une fonction de π pour x succès observés,

➙ Elle est la fonction de **vraissemblance** (au sens Bayesian) pour les données binomiales.

• S'il y a m BDs binomiales (soit $x_1, ..., x_m$ succès sur $n_1, ..., n_m$ tests), alors sous l'hypothèse d'indépendance et de taux π constante, la fonction de vraissemblance sera le **produit** des m vraissemblances individuelles (chacune donnée par l'équation ci-dessus).

On a donc la vraissemblance. ../..

• Pour compléter le modèle et faire une prédiction, on doit spécifier une distribution a priori pour π (cf. Bayes).

• La distribution adaptée pour π est celle qui est conjugée.

➙ Une loi a priori conjuguée est celle que a **la même forme** que la distribution a posteriori.

• La *distribution a priori conjuguée* pour une donnée binomiale est une distribution **Beta** :

$$p(\pi|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\pi^{\alpha-1}(1-\pi)^{\beta-1} \qquad 0 \le x \le 1,\ \alpha > 0,\ \beta > 0$$

➙ Avec α interprété comme le nombre *a priori* de succès du composant testé et

➙ β comme le nombre *a priori* des échecs et

➙ Donc $\alpha + \beta$ est comme la taille *a priori* des données (la BD de référence).

• La distribution *Beta* est un choix naturel pour la loi a priori de π d'autant que son support est l'intervalle $(0, 1)$.

• On a donc notre vraisemblance + a priori. ../..

La prédiction (ajustement de l'*a priori*) :

• Après calculs, la distribution *a posteriori* de π (conditionnée par x succès observés parmi n tests)

est de la forme : $p(\pi|x) \propto \pi^{\alpha+x-1}\,(1-\pi)^{\beta+n-x-1}$

➞ Cela veut dire que la distribution a posteriori de π sachant x est :

$$\pi|x \sim Beta(\alpha+x, \beta+n-x)$$

Un exemple : les générateurs de secours

• Les données (*USA nuclear power plants*) sur une centrale nucléaire (no 63) sont :

– $x = 212$ succès au démarrage sur 212 sollicitaions de génrétauers (tests).

– Cependant, les données de la centrale no 62 (référence) donnent 273 démarrages avec succès sur 278 (donc 5 échecs)

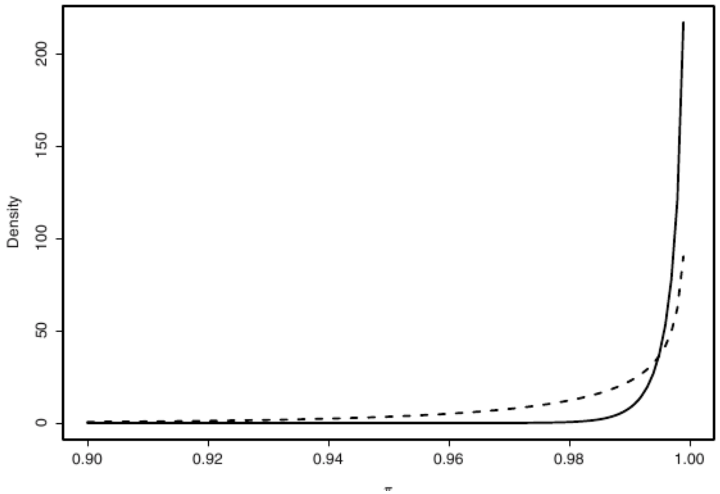
– mais on dispose seulement des 10% des données de la 62.

• Dans ce cas, la 62 (la référence) génère une distribution *a priori* pour la probabilité de succès π :

$$Beta(\alpha = (273/278)/27.8,\ \beta = (1 - 273/278)/27.8)$$

• De la distribution a posteriori précédente, la probabilité π pour la centrale 63 sera :

$$Beta(\alpha + x, \beta + n - x) = Beta(239.3 = (273/278)/28.3 + 212,\ 0.5 = (1 - 5/278)/28.3 + 0)$$



La figure montre les distributions *a priori* et *a posteriori* pour la probabilité de succès π

La distribution a posteriori (lignes pleines) est plus piquée que l'a priori (pointillée).

Les données démontrent davantage d'évidence en faveur d'un taux de succès élevé.

1.15 Résumé du calcul Bayésien

Rappel : on va vu (dans les pages précédentes) les différents modèles utilisés en fonction du type de fonctionnement et le degré d'information des sources disponibles :

- Fonctionnement en continu ➙ distribution vraisemblance **Exponentielle/Poisson**

  à la sollicitaion ➙ distribution vraisemblance **Binomiale**

- Calcul de la proba a priori :

  ➥ Méconnaissance (source) : uniforme non informative

  ➥ Peu de connaissance : uniforme informative (deux valeurs extrêmes)

  ➥ Bonne connaissance : en continu ➙ **Gamma**

  à la sollicitaion ➙ **Bêta**

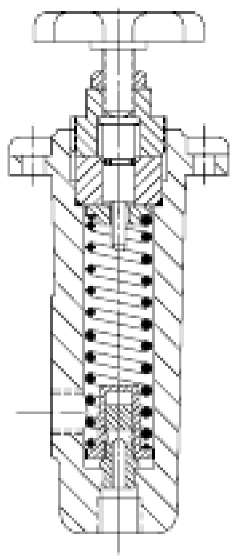
- N.B. utilisation de principe de Conjuguaison (pour *a posteriori*)

  Exponentielle/Poisson - Gamma ou *Binomiale - Bêta*

1.16 Application aux Soupapes de sûreté

• Dispositif de sécurité

Vue en coupe



• Modes / Causes de défaillance :

➥ Les manipulations dégradant les réglages (transport, maintenance).

➥ La corrosion

➥ Surface des sièges endommagées (usure)

➥ Défaillance ressort (fatigue)

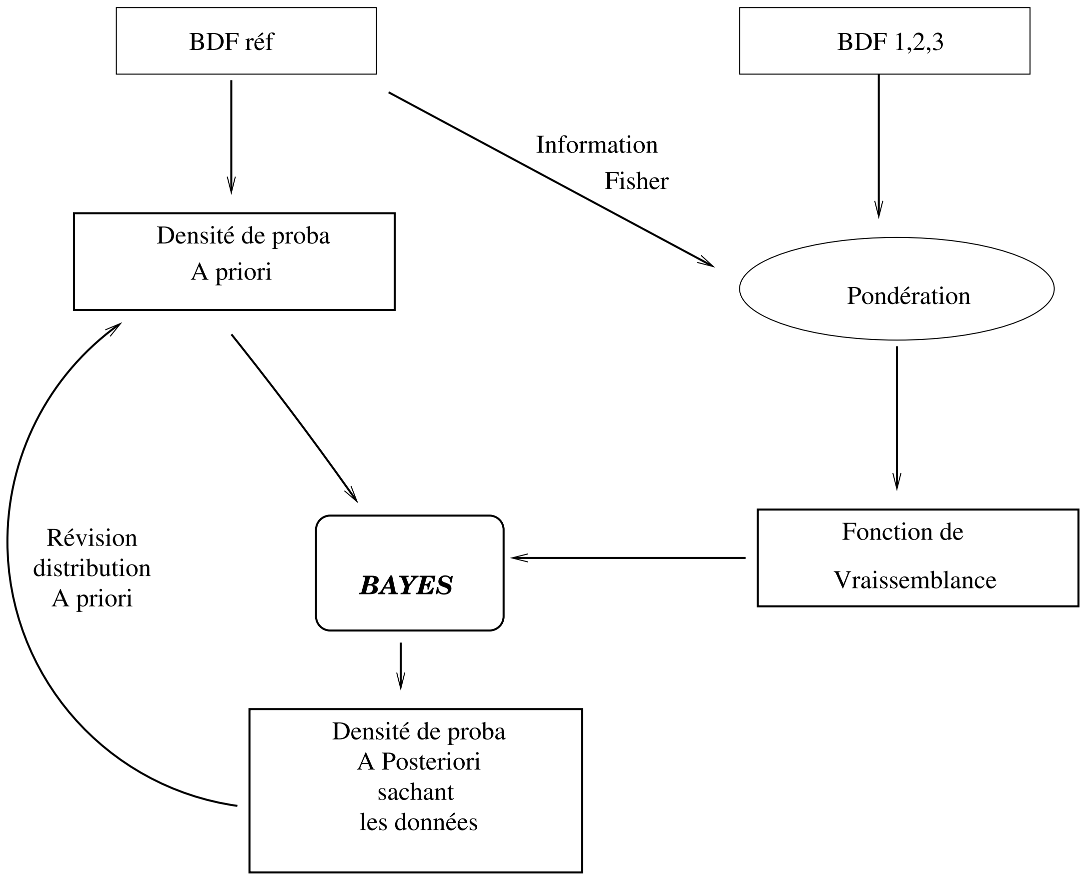
➥ Colmatage et blocage

• Ces défaillances provoquent de la fuite, vibration en fonctionnement, manque d'ouverture à la pression indiquée, manque de fermeture....

• D'autres défaillances peuvent être provoquées par le milieu d'utilisation : température, milieu chimique, support vibrant, chocs, immersion...

1.16.1 *Application de la méthode Bayesien*

• Données REX : on a disposé de 4 BDs dont une de **référence**.



• Les BDFs REX : ../..

Fonctionnement continu	**OREDA**	**EIReDA**	**NPRD95**	**INERIS**
Tps opérationnel cumulé $\times 10^6$ heures	0,69		61,64	31,79
nbr défaillances	19		2001	1479
Taux défaillance moyen donné ($\times 10^6$)	29,35	6,1	32,46	46,52
Intervalle de tolérance ($\times 10^6$) Ou Facteur d'err : $= Max(\frac{\lambda}{\lambda_{min}}, \frac{\lambda_{max}}{\lambda})$	[11,11 ; 54,66]	1,15		[45 ; 48,1]
degré de confiance	90%	60%		80%

1.16.2 *La modélisation soupape*

Une fois les données recueillies, il faut trouver le modèle de **la loi de survie** et la fonction de **vraisemblance**.

Qualité des données : bonne connaissances

➥ La loi de survie (a priori) : mode continue, loi Gamma

La fonction de vraisemblance :

- On a ici un mode de fonctionnement continu .

  ➥ La loi de survie est de type exponentiel et le paramètre à estimer λ.

1.16.3 *La vraisemblance soupape (Exp)*

• Pour la base de données EIREDA, on a seulement un taux de défaillance moyen avec un intervalle de tolérance.

➥ On a : Facteur d'erreur $= Max\left(\dfrac{\lambda}{\lambda_{min}}, \dfrac{\lambda_{max}}{\lambda}\right)$

➥ Dans l'hypothèse d'un intervalle bilatéral, on aura $\dfrac{\lambda}{\lambda_{min}} = \dfrac{\lambda_{max}}{\lambda} = 1,15$

➥ D'où $\lambda \in [5,3 \times 10^{-6}\ ;\ 7 \times 10^{-7}]$

➥ Et avec un intervalle de confiance à 60%, on aura (sous l'hypothèse de distribution normale de λ) : $\lambda_{max} - \lambda_{min} = 2 \times 0.84.\sigma$

• Les deux premiers moments (paramètres) de la loi Gamma (finale) se calculent par :

$$\alpha = \beta.\lambda = 36 \qquad \text{et} \qquad \beta = \frac{\lambda}{\sigma^2} = 5,96 \times 10^6$$

NB : on peut aussi obtenir α et β par la distribution χ^2 par approximations successives.

1.16.4 *La distribution a priori soupape (Gamma)*

• Matériel fonctionnant en continu,

• On dispose (sources) du temps cumulé de fonctionnement et du nombre de défaillances

➥ Cas *bonne* information : on a vu que le modèle *a priori* est la distribution **Gamma**.

➥ On utilisera la BDF d'INERIS à cet effet (BDF de référence).

• Si l'on considère que $\pi(\lambda)$ suit une distribution **Gamma** de paramères (α, β), on obtient (voir le tableau des sources) :

$\alpha = k = 1479$ (nbr de défaillances) et $\beta = T = 31,79 \times 10^6$ (temps cumulés).

Rappel : pour une variable aléatoire X (distribution Gamma) de paramètres **a** (entier) et **b** (MTBF), la densité de proba est la somme de **a** variables aléatoires exponentielles indépendantes de probabilité $f(x_i) = \frac{1}{b}.e^{-\frac{x_i}{b}}$ (cf. conjuguaison Gamma - Exp.)

1.16.5 *Pondérations*

• A cette étape, après avoir modélisé la distribution selon la BDF de référence (INERIS), on procède à la pondération des autres BDFs pour obtenir de ces BDfs secondaires un échantillon comparable à celle de la BDF de référence (pour les cumuler et procéder à l'étape de calculs suivante).

• NB : Pour déterminer les coefficients de pondération, on fait l'hypothèse que le taux ou la proba des défaillances sont **constantes** et que les distributions sont **conjuguées** (ce qui est le cas des BDFs utilisées où l'on utilise pour distribution des densités Exponentielle-Gamma ou Binomiale-Bêta, deux à deux conjuguées).

• On laisse les détails de cette pondération et on utilise ses résultats.

1.16.6 *Taux de défaillance*

La dernière étape : calculer la probabilité *a posteriori* du taux de défaillance (Bayes)

➥ pour obtenir les paramètres de la distribution a posteriori ainsi que son espérance et son intervalle de crédibilité.

• Ici, la fonction de **vraisemblance** est *Exponentielle* : $L(n, T|\lambda) = \lambda^n.e^{-\lambda.T}$

Et la distribution a priori (Gamma) a donné : $g(\lambda) = \dfrac{\lambda^{\alpha-1}.e^{-\frac{\lambda}{\beta}}}{\beta^\alpha.\Gamma(\alpha)}$

• Le taux de défaillance pertinent a posteriori obtenu est : $\hat{\lambda} = \dfrac{\alpha_0 + n_{vrais}}{\beta_0 + T_{vrais}} = 23,66 \times 10^{-6}$

➥ Les taux vrais sont les résultats après toutes les pondération des BDFs.

• L'intervalle de confiance calculé (par χ^2) sera :

$\lambda_{min} = 23,28 \times 10^{-6}$ et $\lambda_{max} = 24 \times 10^{-6}$

• NB : sans l'étape de pondération, on aurait obtenu $\hat{\lambda} = 35,42 \times 10^{-6}$

1.16.7 Analyse des résultats

• On procède toujours à une comparaison des courbes *a priori* et *a posteriori*.

➥ Si les données sont informatives mais les deux courbes très différentes, on peut en déduire que la modélisation n'a pas été correcte ou bien a été peu précise.

➥ Plus l'écart entre les deux courbes est important, plus il y a un risque d'erreur dans la modélisation d'a priori.

Dans ce cas, on procède à une modélisation simple fréquentielle classique (à la place de Byesienne).

➥ Si a priori est mal estimée, a posteriori le sera aussi .

Plus exactement, cela veut dire que les matériels sont très différents et une combinaison des données n'a alors pas de sens.

1.17 Autres méthodes

- Autres méthodes (non abordés ici) :

  Modélisation paramétrique a Hasard proportionnel

  Par les RN.

  Par les RdP

  ....

1.18 A mettre - Quelques références Bibliographiques + Pour la fiabilité

• Ce cours s'est inspiré de plusieurs documents sur l'Extraction de Connaissances ainsi que les outils employés (outils Mathématiques, Probabilité et Statistiques, etc.).

• Aussi, un nombre important d'articles et des rapports de recherche peuvent être consultés (*http ://citeseer.ist.psu.edu/cs*).

• A titre d'indication, les ouvrages suivants peuvent être approfondis :

Data Mining : Introductory and Advanced Topics : Margaret H. Dunham. Prentice Hall 2002.

Data Mining : Concepts, Models, Methods and Algorithms. M. Kantardzic. IEEE 2001

Data Mining : A tutorial based primer. R.J. Roiger , M.W.Geatz

Principles of Data Mining : D. Hand, H. Manila, P. Smyth. MIT Press, Cambridge, 2001

ET bien d'autres !

Table des matières