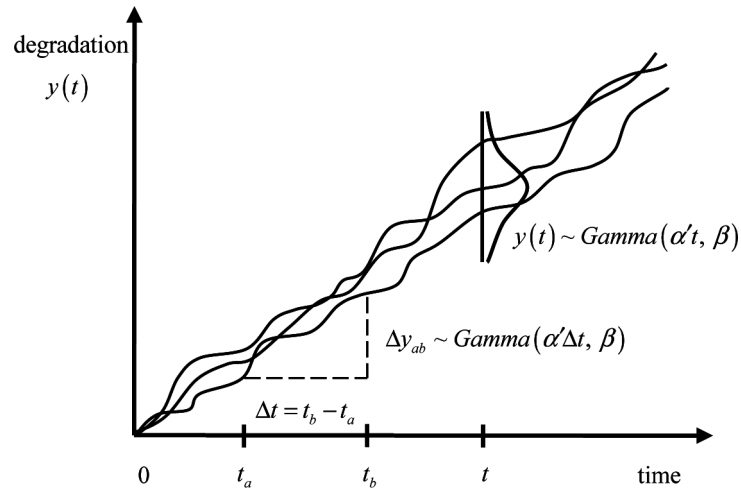


# Rapport - Projet de spécialité

Modèles aléatoires et inférence statistique pour les données de dégradation



Corentin Alary-Ferrandiz | Aya Elarbaoui | Soufiane Lemrabet | Dillane Mahan | Abdoul Kader



# Table des matières

<b>I.Introduction</b>	<b>3</b>
<b>II.Organisation Générale</b>	<b>3</b>
<b>III.Etude du processus de Wiener</b>	<b>3</b>
III.1 : Estimateurs du maximum de vraisemblance . . . . .	4
Estimateur du maximum de vraisemblance de $\mu, \hat{\mu}$ . . . . .	4
Estimateur du maximum de vraisemblance de $\sigma^2, \hat{\sigma}^2$ . . . . .	4
III.2 : Estimateurs des moments . . . . .	4
Estimateur des moments de $\mu, \tilde{\mu}$ . . . . .	5
Estimateur des moments de $\sigma^2, \tilde{\sigma}^2$ . . . . .	5
III.3 : Temps de dépassement du seuil . . . . .	5
III.4 : Simulation et résultats pratiques pour le processus de Wiener . . . . .	5
<b>IV.Etude du processus Gamma</b>	<b>9</b>
IV.1 : Estimateurs du maximum de vraisemblance . . . . .	10
IV.2 : Estimateurs des moments . . . . .	10
IV.3 : Temps de dépassement du seuil . . . . .	10
IV.4 : Simulation et résultats pratiques pour le processus Gamma . . . . .	11
<b>V. Etude de données réelles</b>	<b>12</b>
<b>VI. Biais des Estimateurs</b>	<b>13</b>
<b>VII. Annexes</b>	<b>14</b>
VII.1 : Calcul des estimateurs du maximum de vraisemblance dans le cas où $t_{i,j} = t_{1,j}$ et $m_i = m_1$ pour le processus de Wiener . . . . .	14
VII.2 : Calcul des estimateurs des moments dans le cas où $t_{i,j} = t_{1,j}$ et $m_i = m_1$ pour le processus de Wiener . . . . .	17
VII.3 : Calcul des estimateurs du maximum de vraisemblance dans le cas où $t_{i,j} = t_j$ et $m_i = m_1$ pour le processus Gamma . . . . .	18
VII.4 : Calcul des estimateurs des moments dans le cas où $t_{i,j} = t_j$ et $m_i = m_1$ pour le processus Gamma . . . . .	18

## I.Introduction

L'objet de ce projet était d'étudier différents processus de dégradation. Avant de présenter notre travail, nous allons donc expliquer brièvement ce qu'est un «processus de dégradation» ainsi qu'explicitier l'intérêt de leur étude.

Le terme «Processus de dégradation» fait référence à un ensemble de modèles probabilistes qui correspondent à différents types de processus stochastiques. De nombreux phénomènes de dégradations comme par exemple les processus d'usure de composants matériels, la corrosion ou encore la croissance de fissures peuvent se modéliser par des processus de dégradation. Les étudier permet de mieux les comprendre et de prédire leur évolution. Dans le cadre industriel cela permet notamment de prévoir la date à laquelle un composant atteindra son niveau d'usure critique afin de faire de la maintenance prédictive. Dans la suite, c'est dans cette perspective que nous avons étudié le processus Gamma et le processus de Wiener en les simulant ainsi qu'en estimant les paramètres les caractérisants à la fois sur des données simulées et réelles.

Le support qui a été choisi lors de ce projet afin de visualiser nos résultats est une application web Shiny.

Par la suite nous étudierons  $n$  trajectoires de dégradations dans le cas du processus de Wiener comme dans le cas du processus Gamma.

Pour chacun des  $i$ ème systèmes nous relevons les niveaux de dégradation pour les temps  $t_{i,1}$  à  $t_{i,m_i}$ .

Nous adopterons également les notations suivantes :

- $X_{t_{i,j}}^{(i)}$  est la variable aléatoire modélisant le niveau de dégradation du  $i$ ème système au temps  $t_{i,j}$  et  $x_{t_{i,j}}^{(i)}$  son observation
- $Y_{i,j} = X_{t_{i,j}}^{(i)} - X_{t_{i,j-1}}^{(i)}$  et  $y_{i,j} = x_{t_{i,j}}^{(i)} - x_{t_{i,j-1}}^{(i)}$
- $s_{i,j} = t_{i,j} - t_{i,j-1}$
- $\bar{m} = \frac{1}{n} \sum_{i=1}^n m_i$

Enfin il est à noter que nous avons fait le choix de nous restreindre au cas où les trajectoires de dégradations commencent toutes au temps 0 et à la valeur 0.

## II.Organisation Générale

Afin de bien identifier la source de nos éventuelles erreurs et de permettre de faire évoluer facilement notre application Shiny nous avons décidé de cloisonner au maximum notre travail. Nous avons ainsi travaillé en séparant trois grandes parties.

La première partie était celle de l'application web en elle-même composé des fichiers server.R, consacré aux calculs côté serveur de l'application, et ui.R, consacré à l'interface graphique de l'application.

La seconde partie concernait la simulation des processus de dégradation et l'estimation des paramètres les caractérisant sur des données simulées puis réelles. En dehors de server.R et ui.R, tous les fichiers .R de notre projet se situent dans cette partie. Dans celle-ci se trouve également des fichiers.csv contenant des données réelles de dégradation.

La troisième partie quant à elle était la partie purement mathématique avec tous les calculs permettant d'estimer les paramètres des différents processus de dégradation.

## III.Etude du processus de Wiener

Le premier processus que nous avons étudié est le processus de Wiener qui modélise une dégradation continue avec une tendance croissante mais avec une trajectoire non monotone.

Voici sa définition :

Le processus stochastique  $X = (X_t)_{t \geq 0}$  est un processus de Wiener si  $\forall t \geq 0, X_t = \mu t + \sigma B_t$  où  $\mu \in \mathbb{R}, \sigma > 0$  et  $(B_t)_{t \geq 0}$  est un mouvement brownien.

Dans ce cas  $\mu$  est un paramètre de “drift” et  $\sigma^2$  un paramètre de variance (ou de volatilité).

Un processus de Wiener a les mêmes propriétés qu’un mouvement brownien, on note en particuliers que les increments sont indépendants. La différence est que  $\forall s < t$ , l’increment  $W_t - W_s$  suit une loi gaussienne et on a donc  $\mathbb{E}(W_t - W_s) = \mu(t - s)$  et  $Var(W_t - W_s) = \sigma^2(t - s)$ .

Nous nous plaçons ici dans le cas où nous étudions n trajectoires de dégradations modélisées par des processus de Wiener.

### III.1 : Estimateurs du maximum de vraisemblance

Dans un premier temps avons calculé les estimateurs du maximum de vraisemblance pour les paramètres  $\mu$  et  $\sigma^2$  dans le cas où les temps d’observation des niveaux de dégradation sont identiques pour les n systèmes. Autrement dit dans le cas où  $\forall i \in [[1, n]], t_{i,j} = t_j$ .

Puis nous avons cherché à analyser la qualité de ces estimateurs en calculant leur biais. Le détail des calculs que nous avons ici effectué se situe en annexe VII.1.

Par la suite nous avons cherché à généraliser ces formules pour le cas où les temps d’observation des niveaux de dégradation ne sont pas nécessairement identiques pour les n systèmes. En suivant la même démarche nous avons obtenu les résultats suivants :

**Estimateur du maximum de vraisemblance de  $\mu, \hat{\mu}$**

$$\hat{\mu} = \frac{\sum_{i=1}^n x_{t_{i,m_i}}^{(i)}}{\sum_{i=1}^n t_{i,m_i}}$$

**Estimateur du maximum de vraisemblance de  $\sigma^2, \hat{\sigma}^2$**

$$\hat{\sigma}^2 = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{(y_{i,j} - \hat{\mu} s_{i,j})^2}{s_{i,j}}$$

### III.2 : Estimateurs des moments

Après avoir obtenu des estimateurs par maximum de vraisemblance, nous avons cherché à obtenir de nouveaux estimateurs grâce à la méthode des moments. Nous sommes premièrement penché sur le cas où les temps d’observation des niveaux de dégradation sont identiques pour les n systèmes. Autrement dit dans le cas où  $\forall i \in [[1, n]], t_{i,j} = t_j$ .

Le détail des calculs que nous avons effectué se situe en annexe VII.2.

Par la suite nous avons cherché à généraliser ces formules pour le cas où les temps d’observation des niveaux de dégradation ne sont pas nécessairement identiques pour les n systèmes.

Dans notre application les  $m_i$  sont identiques pour toutes les courbes (autrement dit il y a le même nombre de points pour toutes les trajectoires de dégradation), nous adopterons donc ici cette hypothèse en notant simplement ce nombre  $m$ .

Dans ce cadre et en appliquant la même démarche que pour les calculs de l’annexe VII.2 nous avons obtenu les résultats suivants :

Estimateur des moments de  $\mu$ ,  $\tilde{\mu}$

$$\tilde{\mu} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \frac{Y_{i,j}}{s_{i,j}}$$

Estimateur des moments de  $\sigma^2$ ,  $\tilde{\sigma}^2$

$$\tilde{\sigma}^2 = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \left( \frac{Y_{i,j}}{\sqrt{s_{i,j}}} - \frac{1}{m} \sum_{j=1}^m \frac{Y_{i,j}}{\sqrt{s_{i,j}}} \right)^2$$

### III.3 : Temps de dépassement du seuil

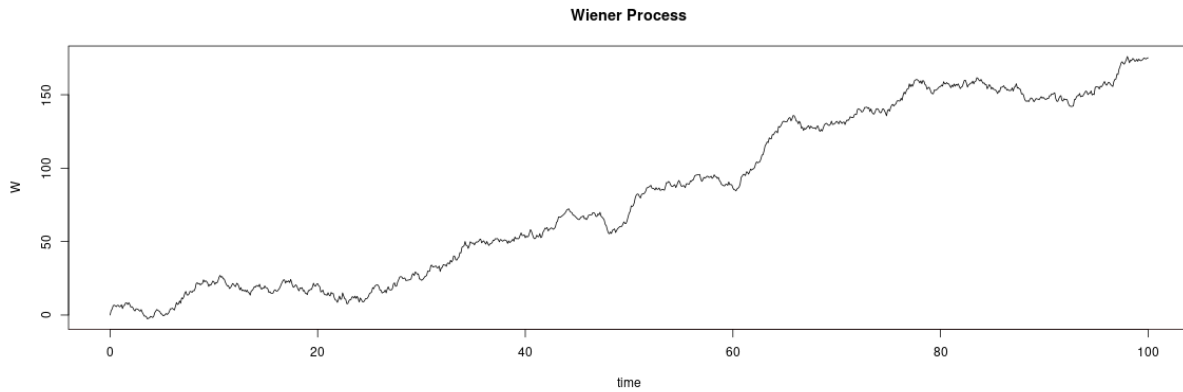
Un autre aspect de notre étude du processus de Wiener a consisté à étudier le temps de dépassement d'un seuil donné. Concrètement nous avons cherché à déterminer l'instant  $T_s$  où le processus de Wiener franchit pour la première fois un seuil  $s$  donné.

En nous plongeant dans la littérature scientifique nous avons appris que cette variable aléatoire  $T_s$  suit, dans le cadre d'un processus de Wiener, la loi inverse-gaussienne de paramètres  $\frac{s}{\mu}$  et  $\frac{s^2}{\sigma^2}$ .

Ainsi comme  $T_s \sim \mathcal{IG}(\frac{s}{\mu}, \frac{s^2}{\sigma^2})$  nous savons également que  $\mathbb{E}[T_s] = \frac{s}{\mu}$ .

### III.4 : Simulation et résultats pratiques pour le processus de Wiener

Avant de présenter en détail la page "Simulation" de notre application dans le cas du processus de Wiener commençons simplement par montrer un exemple de simulation d'une trajectoire de dégradation suivant un processus de Wiener de paramètre  $\mu = 1$  et  $\sigma^2 = 20$  :



Comme nous pouvons l'observer sur cette simulation, on est bien ici dans le cas d'une dégradation continue avec une tendance croissante mais qui à certains moments peut diminuer.

Maintenant détaillons l'interface de notre application pour la page "Simulation" en commençant par la partie gauche nommée "Parameters :".

## Partie “Parameters”

Parameters :

Select a model

Wiener Process

mu :

1

sigma2:

20

Type of distribution for the time

Périodique

Period:

0,1

Number of paths :

1

☒ different T

Number of points by path:

1000

Threshold:

10

OK

“Select a model” permet de choisir entre une simulation du processus de Wiener ou bien une simulation du processus Gamma, ici nous nous placerons dans le cas du processus de Wiener.

“mu” et “sigma2” correspondent aux paramètres  $\mu$  et  $\sigma^2$  de notre processus.

“Number of paths” permet de choisir le nombre de systèmes suivant un processus de Wiener que nous souhaitons simuler. “Number of points by path” permet de choisir le nombre de points simulés pour chacun des systèmes.

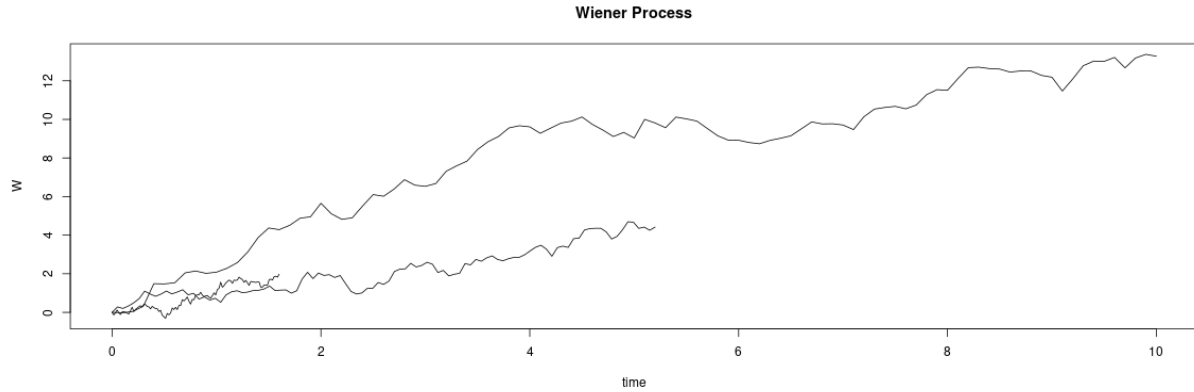
“Threshold” permet de choisir le seuil pour lequel nous souhaitons étudier le temps  $T_s$  de dépassement.

“Type of distribution for the time” permet de choisir comment doivent être simulés les temps pour chacun des processus. Par défaut tous les processus sont simulés aux mêmes temps  $t_1$  à  $t_{1,m_1}$ .

On a le choix entre une distribution périodique, exponentielle, Gamma et normale pour lesquelles on spécifie ses paramètres. Si on a  $n$  points par trajectoire de dégradation, dans le cas de la distribution périodique on spécifie la période  $T$  qui correspond au pas de temps entre chacun des temps de la simulation. Par exemple, si on choisit la période  $T=1$  et si  $n=10$  alors on va simuler ces systèmes aux temps 1, 2, 3, 4, 5, 6, 7, 8, 9 et 10. Dans le cas des autres distributions on va simuler chaque nouveau temps pour les  $n$  points en additionnant le

temps précédent à un tirage de la distribution choisit. Par exemple si on prend la loi exponentielle de paramètre 1, on va faire une tirage de loi exponentielle de paramètre 1 pour obtenir le premier temps puis on va additionner à celui-ci un nouveau tirage de loi exponentielle de paramètre 1 pour obtenir le 2ème temps puis on va faire de même jusqu’à obtenir  $n$  temps.

Enfin il est possible de cocher la case “different T” qui permet de simuler pour toutes les trajectoires de dégradation des temps différents suivant la distribution choisie avec “Type of distribution for the time”. Dans le cas de la distribution périodique si cette case a été cochée alors la première trajectoire aura la période spécifiée en paramètre et les autres auront une période choisie au hasard entre la période spécifiée en paramètre divisée par 100 et la période spécifiée en paramètre. Ainsi si l’utilisateur sélectionne 10 trajectoires et sélectionne la période 1, la première trajectoire aura la période 1 et les 9 autres auront des périodes choisis au hasard et comprises entre  $1/100$  et 1. Dans le cas des autres distributions on applique la même démarche que lorsque la case n’est pas cochée mais on recalcule pour chacune des trajectoires un nouveau vecteur de temps.



Ci-dessus voici un exemple de simulation pour laquelle les temps sont différentes pour chacune des 3 trajectoires.

### Partie Centrale de l'application

Au centre de l'écran sur la partie haute voici comment se présente notre application :

#### Parameters Estimates :

Estimations	$\mu$	$\sigma^2$
EMV	1.00744	1.01578
EMM	1.00643	1.01578

Sous la partie “Parameters Estimates :” se trouve un tableau avec les estimations des paramètres  $\mu$  et  $\sigma^2$  selon les estimateurs du maximum de vraisemblance (“EMV”) puis selon les estimateurs des moments (“EMM”).

Pour obtenir ces valeurs nous avons simplement implémenté en R les formules des estimateurs des parties III.1 et III.2.

En manipulant l'application nous constatons qu'au delà d'une centaine de points, c'est-à-dire une trajectoire de dégradation de 100 points ou bien 10 trajectoires de 10 points, la précision des estimateurs dépend principalement du paramètre  $\sigma^2$  du processus simulé.

En effet plus on fait tendre la valeur de  $\sigma^2$  vers de grandes valeurs plus les erreurs de précision des estimateurs à la fois pour  $\mu$  et pour  $\sigma^2$  deviennent conséquentes.

Faire varier la valeur du paramètre  $\mu$  simulé n'a en revanche pas un réel impacte sur la qualité de ces estimateurs.

Dans la suite de l'application se trouvent les éléments suivants :

True mean of the time of exceeding the threshold:

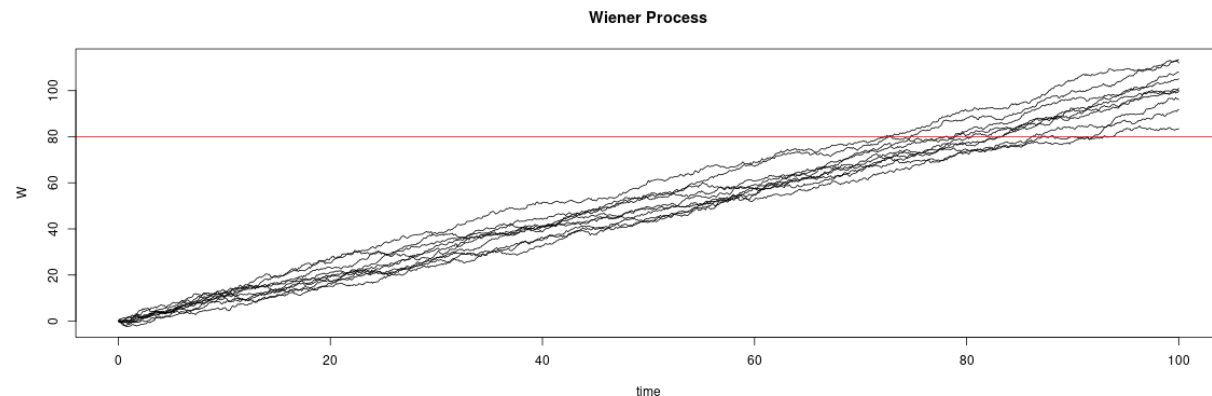
80

Estimated mean of the time of exceeding the threshold:

79.16217

Empirical mean of the time of exceeding the threshold:

81.42



“True mean of the time of exceeding the threshold :” correspond à l’espérance du premier instant  $T_s$  de dépassement du seuil  $s$  calculé à l’aide des véritables paramètres  $\mu$  et  $\sigma^2$ . En d’autres terme c’est le temps théorique moyen de dépassement du seuil  $s$ .

“Estimated mean of the time of exceeding the threshold :” correspond à l’espérance du premier instant  $T_s$  de dépassement du seuil  $s$  calculé à l’aide des estimateurs du maximum de vraisemblance de  $\mu$  et  $\sigma^2$ . En d’autres terme c’est une estimation du temps théorique moyen de dépassement du seuil  $s$ .

Concrètement pour calculer ces deux valeurs nous nous sommes appuyés sur la formule de l’espérance de la partie III.3. Dans le cas de “Estimated mean of the time of exceeding the threshold :” nous avons remplacé  $\mu$  par son estimateur de maximum de vraisemblance.

“Empirical mean of the time of exceeding the threshold :” correspond quant à lui à la moyenne des premiers instants de dépassement du seuil  $s$  présents dans nos simulations pour chacune des trajectoires.

Cette valeur seuil  $s$  dont il est question ici se visualise directement dans notre application avec la barre rouge que l’on peut observe sur la simulation ci-dessus.

Tout comme pour les estimateurs de  $\mu$  et  $\sigma^2$  nous observons cette même tendance à obtenir des résultats de moins en moins précis lorsque la valeur du  $\sigma^2$  simulée augmente pour le premier instant de dépassement théorique. Cela est cohérent puisque ce dernier est calculé à partir des estimateurs de  $\mu$  et  $\sigma^2$ , qui deviennent de moins en moins précis lorsque la valeur de  $\sigma^2$  augmente.

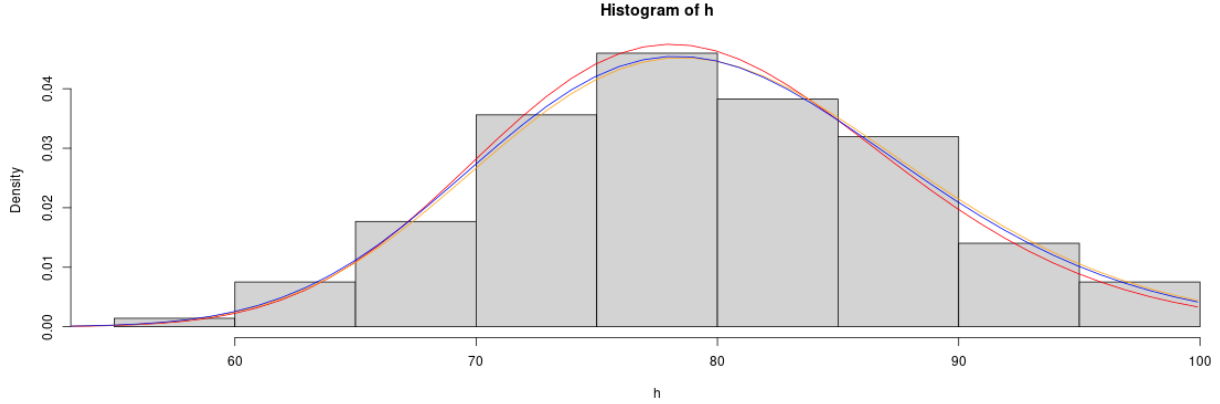
Ce qui nous permet ici de conclure à une perte de précision c’est le constat que les valeurs théoriques s’écartent de plus en plus de la valeur moyenne du premier instant de dépassement observé dans les données.

Cependant, lorsque la valeur de  $\sigma^2$  est raisonnable (de l’ordre de la dizaine), qu’importe la valeur de  $\mu$  ou bien du nombre de points on observe que les valeurs théoriques du premier instant de dépassement “True mean of the time of exceeding the threshold :” et “Estimated mean of the time of exceeding the threshold :” sont très proches de la valeur moyenne empirique de ces instants “Empirical mean of the time of exceeding the threshold :”.

## Partie Histogramme



Enfin tout en bas de la partie centrale si trouve cette partie :



Sur ce graphique se trouve un histogramme qui représente la répartition des seuils observés dans les simulations pour les différentes trajectoires de dégradation. A celui-ci se superposent également deux courbes.

La première courbe en orange représente la densité de la loi théorique du temps de dépassement  $T_s$  avec les véritables paramètres de la loi inverse-gaussienne vu à la partie III.3 obtenus à partir des valeurs  $\mu$  et  $\sigma^2$  choisis pour la simulation.

La deuxième courbe en rouge représente également la densité de la loi théorique du temps de dépassement  $T_s$  mais obtenue en utilisant l'espérance et la variance observées dans les valeurs des simulations.

Enfin la troisième courbe en bleu représente la densité de la loi théorique du temps de dépassement  $T_s$  mais cette fois en utilisant les valeurs des estimateurs du maximum de vraisemblance de  $\mu$  et  $\sigma^2$ .

Nous observons que lorsque le nombre de trajectoires de dégradation et le nombre de points par trajectoire est grand l'histogramme ainsi que ces trois courbes se superposent parfaitement.

## IV. Etude du processus Gamma

Le second processus que nous avons étudié est le processus Gamma. Ce processus modélise des détériorations qui ne diminuent pas au cours du temps mais qui au contraire s'accumulent au travers de nombreux petits incréments.

Sa définition est la suivante :

Le processus  $(X_t)_{t \geq 0}$  est appelé un processus Gamma (non homogène) de fonction "shape"  $A(t)$  et de paramètre  $b$  si :

- $X_0 = 0$  presque sûrement
- $(X_t)_{t \geq 0}$  est un processus d'incrémentes indépendants :  $\forall n \in \mathbb{N}$ , tous les  $(t_1, \dots, t_{n+1}) \in \mathbb{R}^{n+1}$  tels que  $0 \leq t_1 < t_2 < \dots < t_{n+1}$ , les incréments  $X_{t_{i+1}} - X_{t_i}$  avec  $1 \leq i \leq n$  sont indépendants ;
- $\forall 0 < s < t$ , la variable aléatoire  $X_t - X_s \sim \mathcal{G}(A(t) - A(s), b)$ .

Dans le cas où  $\forall t \geq 0$ ,  $A(t) = at$  avec  $a > 0$ , le processus est appelé processus Gamma homogène de paramètres  $a$  et  $b$ . C'est sur ce dernier cas que portera notre étude.

Nous nous plaçons également ici dans le cas où nous étudions  $n$  trajectoires de dégradation modélisées par des processus Gamma.

## IV.1 : Estimateurs du maximum de vraisemblance

En premier lieu tout comme pour le processus de Wiener nous avons cherché à calculer les estimateurs du maximum de vraisemblance du processus Gamma dans le cas où les temps d'observation des niveaux de dégradation sont identiques pour les  $n$  systèmes. Autrement dit dans le cas où  $\forall i \in [[1, n]], t_{i,j} = t_j$ .

Nous avons détaillé dans l'annexe VII.3 les calculs que nous avons mené pour ces estimateurs.

Ces calculs nous ont permis d'obtenir une expression simple pour l'estimateur de maximum de vraisemblance du premier paramètre  $a$  en fonction du second paramètre  $b$  du processus.

Pour l'estimateur du maximum de vraisemblance de  $b$  nous avons obtenu une expression beaucoup plus complexe dépendante des données de dégradation et qui nécessite une résolution numérique. Pour effectuer celle-ci nous nous sommes appuyé sur la fonction R "uniroot". Cependant la méthode de changement de signe sur laquelle se base cette fonction ne permet pas de trouver une solution pour toutes les simulations. Ne trouvant pas d'alternatives, nous avons donc décidé pour notre application d'abandonner ces estimateurs pour nous focaliser sur les estimateurs des moments pour le processus Gamma.

## IV.2 : Estimateurs des moments

Contrairement aux estimateurs précédents, les estimateurs des moments nous ont permis d'obtenir des expressions exploitables pour toutes nos simulations.

Les calculs que nous avons effectué pour ces estimateurs dans le cas où les temps d'observation des niveaux de dégradation sont identiques pour les  $n$  systèmes se trouvent dans l'annexe VII.4.

Dans le cas où les temps peuvent être différents pour chacun des systèmes, en suivant le même raisonnement, nous avons obtenu les formules suivantes :

$$\hat{a} = \frac{\left[ \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{Y_{i,j}}{s_{i,j}} \right]^2}{\frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{(Y_{i,j} - s_{i,j} [\frac{1}{nm} \sum_{k=1}^n \sum_{l=1}^{m_k} \frac{Y_{k,l}}{s_{k,l}}])^2}{s_{i,j}}}$$

$$\hat{b} = \frac{\frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{Y_{i,j}}{s_{i,j}}}{\frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{(Y_{i,j} - s_{i,j} [\frac{1}{nm} \sum_{k=1}^n \sum_{l=1}^{m_k} \frac{Y_{k,l}}{s_{k,l}}])^2}{s_{i,j}}}$$

## IV.3 : Temps de dépassement du seuil

Tout comme pour le processus de Wiener nous avons cherché à obtenir l'instant  $T_s$  où le processus Gamma franchit pour la première fois un seuil  $s$  donné.

Cependant en cherchant dans la littérature scientifique nous n'avons cette fois pas obtenu de résultats aussi facilement exploitables.

Nous n'avons ainsi pas trouvé de loi connue pour le temps de dépassement d'un processus Gamma.

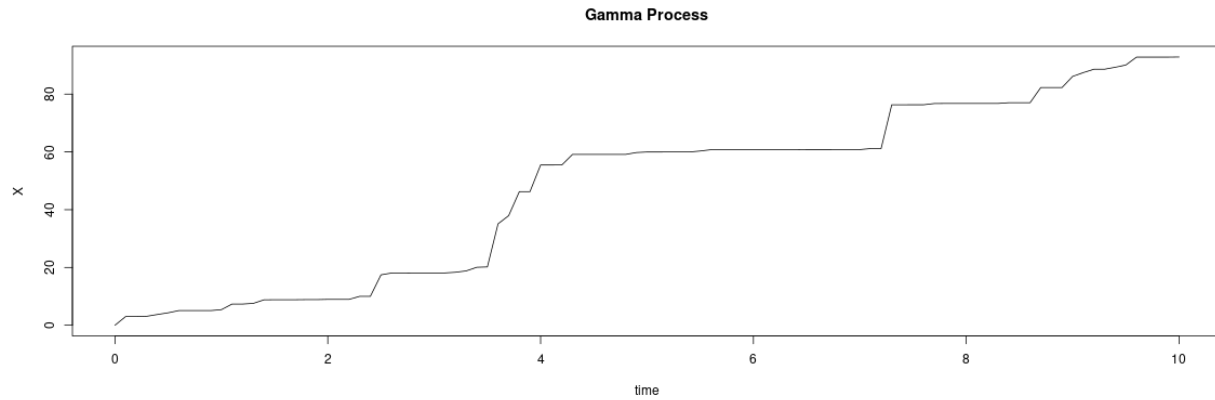
Nous avons trouvé une expression de l'espérance du temps de dépassement dans le cas d'un processus Gamma homogène mais là encore, celle-ci est beaucoup plus complexe que dans le cas du processus de Wiener :

$$\mathbb{E}[T_s] = \frac{1}{a} \int_0^h U(x) dx \quad \text{avec} \quad U(x) = \int_0^{+\infty} f_{t,1}(x) dt \quad \text{et où } f_{a,b} \text{ est la densité d'un loi Gamma de paramètres } a \text{ et } b.$$

Malheureusement pris par le temps et dû à la difficulté du calcul numérique de cette double intégrale nous n'avons pas pu exploiter cette formule.

## IV.4 : Simulation et résultats pratiques pour le processus Gamma

Commençons par présenter une simulation d'une trajectoire de dégradation d'un processus Gamma de paramètre  $a = 1$  et  $b = 0.1$  :



Cette simulation nous permet bien ici d'observer une dégradation qui ne diminue pas au cours du temps et qui semble s'accumuler au travers de petits "sauts".

### Partie "Parameters :"

Cette partie est la même que dans le cas du processus de Wiener (cf III.4) à ceci près que maintenant les paramètres à saisir au début par l'utilisateur sont  $a$  et  $b$  et non plus  $\mu$  et  $\sigma^2$ .

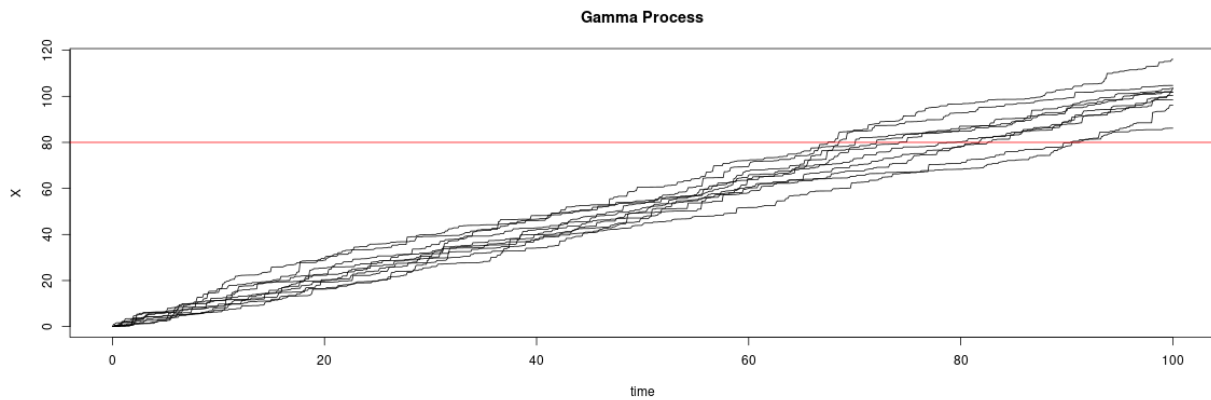
### Partie Centrale de l'application

Tout en haut de cette partie en dessous de "Parameters Estimates :" se trouvent les valeurs des estimateurs des moments du paramètre  $a$  et du paramètres  $b$ . Ces valeurs sont simplement l'implémentation en R des formules vu à la partie IV.2.

En dessous de celle-ci on retrouve les éléments suivants :

Empirical mean of the time of exceeding the threshold:

77.79



Contrairement au processus de Wiener nous n'obtenons pas de valeurs pour "True mean of the time of

exceeding the threshold :”ou bien pour “Estimated mean of the time of exceeding the threshold :” dû aux difficultés évoquées à la partie IV.3.

“Empirical mean of the time of exceeding the threshold :” correspond quant à lui toujours à la moyenne des premiers instants de dépassement du seuil  $s$  présents dans les simulations pour chacune des trajectoires.

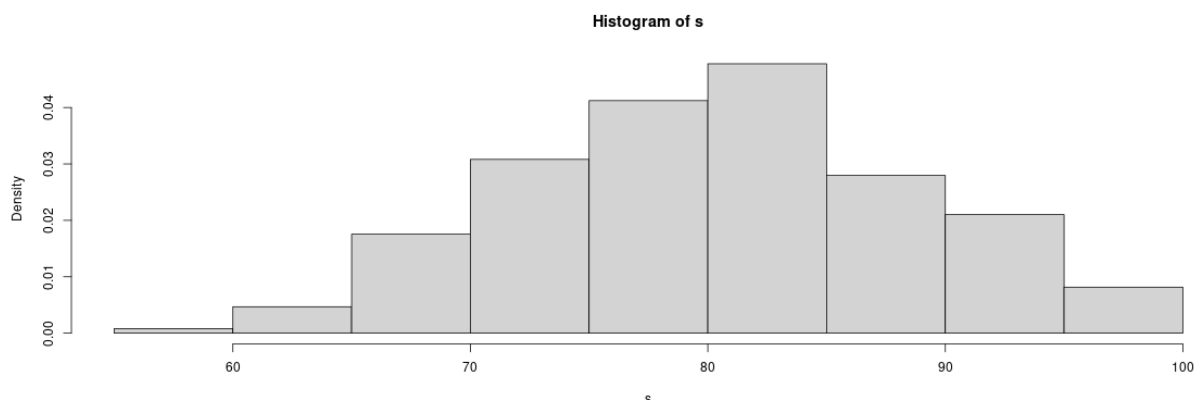
De même le seuil  $s$  se visualise directement dans notre application avec la barre rouge que l’on peut observer sur la simulation ci-dessus.

Enfin en manipulant l’application on remarque que les estimateurs sont nettement moins précis que dans le cas du processus de Wiener. Ainsi pour estimer correctement  $a$  et  $b$  ces estimateurs nécessitent d’avoir un grand nombre de points (environs une centaine) pour être fiable.

La finalité de l’étude de ces estimateurs est de les appliquer à des données réelles, en ce sens les estimateurs des moments de  $a$  et de  $b$  semblent ici problématique puisque dans la réalité nous ne disposons généralement pas de centaines de données sur des processus de dégradation.

Enfin on observe que plus la valeur de  $a$  s’écarte de celle de  $b$  (ou bien  $b$  de celle de  $a$ ) pour les simulations plus les estimateurs perdent en précision.

## Partie Histogramme



Enfin tout comme pour le processus de Wiener nous avons tracé l’histogramme représentant la répartition des seuils observés dans les simulations pour les différentes trajectoires de dégradation.

Cependant ici nous n’avons pas tracé la courbe de la densité de la loi du temps de dépassement car celle-ci n’est pas connu dans le cas d’un processus Gamma. Par conséquent nous n’avons donc logiquement pas non plus tracer de courbes estimant cette loi.

## V. Etude de données réelles

La finalité de notre application était de pouvoir modéliser des données réelles de dégradation par les processus que nous avons étudié dans les partie précédente.

Pour cela nous avons ajouté l’onglet “data” à notre application :

Estimations :

$\mu$  :  
0.00203717

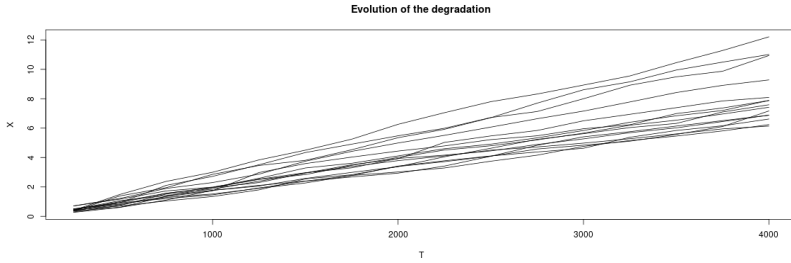
$\sigma^2$  :  
0.00014761

Prediction:  
4908.779

Threshold:

Choose CSV File

laser3.csv



Sur celui-ci se trouve un bouton “Browse” qui permet de charger des données de dégradation au format .csv. Une fois chargées ces données sont affichées sur un graphique à droite.

A gauche sont affichés les estimateurs du maximum de vraisemblance des paramètres  $\mu$  et  $\sigma^2$  des processus de Wiener modélisant ces données.

Enfin on trouve marqué sous “Prediction :” se trouve la valeur de l’espérance du premier instant  $T_s$  de dépassement du seuil  $s$  spécifié dans sous la case “Threshold :”. Celui-ci est calculé à l’aide des estimateurs du maximum de vraisemblance de  $\mu$  et  $\sigma^2$ .

### Forme attendu des données du fichier au format csv :

Nous avons fait le choix pour notre application d’imposer à l’utilisateur que les données de dégradation soient sous la forme suivante : Première colonne nommée “indice” qui contient l’indice de la trajectoire de dégradation considérée, deuxième colonne nommée “T” avec les temps d’observations des dégradations pour la trajectoire de l’indice spécifiée, enfin dernière colonne nommée “X” qui contient les valeurs de dégradations aux temps et indices spécifiées dans les colonnes précédentes.

### Application à des données laser :

Dans le livre “Degradation Processes in Reliability” se trouvent des données de dégradation de laser. Ces données nous les avons récupéré et le résultat que nous obtenu est celui que nous avons présenté au début de cette partie(laser3.csv).

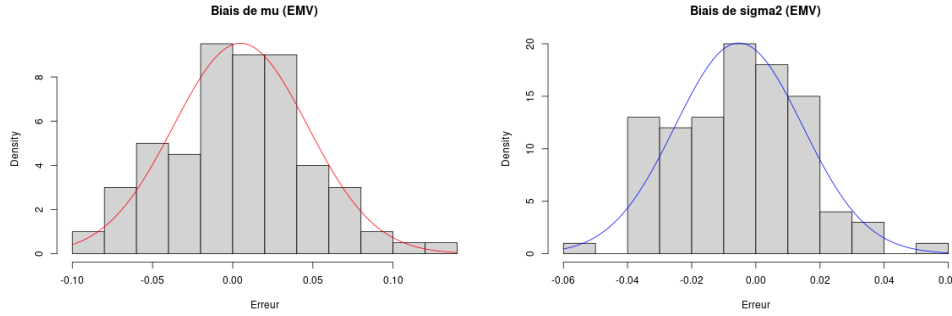
On observe que  $\mu$  est estimé à environs 0.00203717 et  $\sigma^2$  à 0.00014761.

Ces résultats sont cohérents malgré une légère imprécision pour  $\sigma^2$  puisque les auteurs du livre quant eux ont obtenu l’intervalle de confiance à 95% de  $\mu$  suivant : [0.002028; 0.002042] et cet intervalle de confiance à 95% pour  $\sigma^2$  : [0.000156; 0.000161].

## VI. Biais des Estimateurs

Voici comment se présente le tout dernier onglet de notre application, l’onglet “Biais” :

Biais :



Cet onglet a pour but de représenter le biais des estimateurs de maximum de vraisemblance des paramètres du processus de Wiener ainsi que le biais des estimateurs des moments des paramètres du processus Gamma.

En effet nous faisons plusieurs simulations des paramètres puis faisons la différence entre ces estimations et les vraies valeurs des paramètres au bout d'un certain nombre d'itération pour vérifier à quel point notre estimateur donne des valeurs admissibles.

Sur cet onglet se trouve tout d'abord à gauche une partie qui permet de paramétrer finement le processus de Wiener ou bien le processus Gamma pour lequel nous souhaitons analyser le biais.

Au centre se trouvent les histogrammes représentant la répartition des valeurs des erreurs obtenues pour le modèle considéré ainsi qu'une courbe gaussienne.

L'erreur correspond simplement à la différence entre la valeur de l'estimation obtenue et la valeur réelle du paramètre.

Enfin nous avons simplement ajouté une courbe gaussienne car nous avons remarqué que les erreurs semblent suivre une répartition gaussienne.

## VII. Annexes

### VII.1 : Calcul des estimateurs du maximum de vraisemblance dans le cas où $t_{i,j} = t_{1,j}$ et $m_i = m_1$ pour le processus de Wiener

**Calcul des estimateurs du maximum de vraisemblance :** Calculons ainsi la fonction de vraisemblance pour les  $n$  échantillons.

$$\begin{aligned}
 \mathcal{L}(y_{1,1}, \dots, y_{1,m_1}, \dots, y_{n,1}, \dots, y_{n,m_1}, \mu, \sigma^2) &= f_{Y_{1,1}, \dots, Y_{1,m_1}, \dots, Y_{n,1}, \dots, Y_{n,m_1}}(y_{1,1}, \dots, y_{1,m_1}, \dots, y_{n,1}, \dots, y_{n,m_1}, \mu, \sigma^2) \\
 &= \prod_{i=1}^n f_{Y_{i,1}, \dots, Y_{i,m_1}}(y_{i,1}, \dots, y_{i,m_1}) \\
 &= \prod_{i=1}^n \prod_{j=1}^{m_1} f_{Y_{i,j}}(y_{i,j}, \mu, \sigma^2) \\
 &= \prod_{i=1}^n \prod_{j=1}^{m_1} \frac{1}{\sigma \sqrt{s_{1,j}} \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{(y_{i,j} - \mu s_{1,j})^2}{\sigma^2 s_{1,j}} \right)} \\
 &= \left( \frac{1}{\sigma \sqrt{2\pi}} \right)^{nm_1} \left( \prod_{j=1}^{m_1} \frac{1}{\sqrt{s_{1,j}}} \right)^n e^{-\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{m_1} \left( \frac{(y_{i,j} - \mu s_{1,j})^2}{\sigma^2 s_{1,j}} \right)}
 \end{aligned}$$

D'où :

$$\ln(\mathcal{L}(y_{1,1}, \dots, y_{1,m_1}, \dots, y_{n,1}, \dots, y_{n,m_1}, \mu, \sigma^2)) = nm_1 \times \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - n \times \ln\left(\prod_{j=1}^{m_1} \frac{1}{\sqrt{s_{1,j}}}\right) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{m_1} \left(\frac{(y_{i,j} - \mu s_{1,j})^2}{\sigma^2 s_{1,j}}\right)$$

**Estimateur de maximum de vraisemblance de  $\mu$ ,  $\hat{\mu}$  :** Notons le résultat suivant :

$$\frac{\partial(\mu \rightarrow \frac{(y_{i,j} - \mu s_{1,j})^2}{\sigma^2 s_{1,j}})}{\partial \mu} = \frac{-2(y_{i,j} - \mu s_{1,j})}{\sigma^2}$$

Ainsi :

$$\begin{aligned} \frac{\partial \ln(\mathcal{L}(y_{1,1}, \dots, y_{1,m_1}, \dots, y_{n,1}, \dots, y_{n,m_1}, \mu, \sigma^2))}{\partial \mu} &= -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{m_1} \frac{-2(y_{i,j} - \mu s_{1,j})}{\sigma^2} \\ &= \sum_{i=1}^n \left[ \frac{1}{\sigma^2} \sum_{j=1}^{m_1} y_{i,j} - \frac{\mu}{\sigma^2} \sum_{j=1}^{m_1} s_{1,j} \right] \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n \sum_{j=1}^{m_1} y_{i,j} - \frac{\mu}{\sigma^2} \sum_{i=1}^n \sum_{j=1}^{m_1} s_{1,j} \\ \frac{\partial \ln(\mathcal{L}(\mathcal{L}(y_{1,1}, \dots, y_{1,m_1}, \dots, y_{n,1}, \dots, y_{n,m_1}, \mu, \sigma^2)))}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n \sum_{j=1}^{m_1} y_{i,j} - \frac{\mu \times n}{\sigma^2} \sum_{j=1}^{m_1} s_{1,j} \end{aligned}$$

Ainsi on obtient que :

$$\frac{\partial \ln(\mathcal{L}(\mathcal{L}(y_{1,1}, \dots, y_{1,m_1}, \dots, y_{n,1}, \dots, y_{n,m_1}, \mu, \sigma^2)))}{\partial \mu} = 0$$

$\Leftrightarrow$

$$\frac{\mu \times n}{\sigma^2} \sum_{j=1}^{m_1} s_{1,j} = \frac{1}{\sigma^2} \sum_{i=1}^n \sum_{j=1}^{m_1} y_{i,j}$$

$\Leftrightarrow$

$$\mu = \frac{1}{n \sum_{j=1}^{m_1} s_{1,j}} \times \sum_{i=1}^n \sum_{j=1}^{m_1} y_{i,j}$$

Ainsi :  $\boxed{\hat{\mu} = \frac{1}{n \sum_{j=1}^{m_1} s_{1,j}} \times \sum_{i=1}^n \sum_{j=1}^{m_1} Y_{i,j}}$

**Estimateur du maximum de vraisemblance de  $\sigma^2$ ,  $\hat{\sigma}^2$  :**

$$\ln(\mathcal{L}(y_{1,1}, \dots, y_{1,m_1}, \dots, y_{n,1}, \dots, y_{n,m_1}, \mu, \sigma^2)) = m_1 n \times \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - n \times \ln\left(\prod_{j=1}^{m_1} \frac{1}{\sqrt{s_{1,j}}}\right) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{m_1} \left(\frac{(y_{i,j} - \mu s_{1,j})^2}{\sigma^2 s_{1,j}}\right)$$

Notons le résultat suivant :

$$\frac{\partial(\sigma^2 \rightarrow \frac{(y_{i,j} - \mu s_{1,j})^2}{\sigma^2 s_{1,j}})}{\partial \sigma^2} = -\frac{(y_{i,j} - \mu s_{1,j})^2}{(\sigma^2)^2 s_{1,j}}$$

Notons également le second résultat suivant :

$$\frac{\partial(\sigma^2 \rightarrow nd \times \ln(\frac{1}{\sigma\sqrt{2\pi}}))}{\partial \sigma^2} = \frac{-nd}{2\sigma^2}$$

Ainsi on obtient :

$$\begin{aligned}\frac{\partial \ln(\mathcal{L}(y_{1,1}, \dots, y_{1,m_1}, \dots, y_{n,1}, \dots, y_{n,m_1}, \mu, \sigma^2))}{\partial \sigma^2} &= \frac{-nm_1}{2\sigma^2} - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{m_1} -\frac{(y_{i,j} - \mu s_{1,j})^2}{(\sigma^2)^2 s_{1,j}} \\ &= \frac{1}{(\sigma^2)^2} (-m_1 n \times \sigma^2 + \sum_{i=1}^n \sum_{j=1}^{m_1} \frac{(y_{i,j} - \mu s_{1,j})^2}{s_{1,j}})\end{aligned}$$

Par conséquent :

$$\frac{\partial \ln(\mathcal{L}(y_{1,1}, \dots, y_{1,m_1}, \dots, y_{n,1}, \dots, y_{n,m_1}, \mu, \sigma^2))}{\partial \sigma^2} = 0$$

$\Leftrightarrow$

$$nd \times \sigma^2 = \sum_{i=1}^n \sum_{j=1}^{m_1} \frac{(y_{i,j} - \mu s_{1,j})^2}{s_{1,j}}$$

Ainsi :

$$\boxed{\hat{\sigma}^2 = \frac{1}{m_1 n} \times \sum_{i=1}^n \sum_{j=1}^{m_1} \frac{(y_{i,j} - \hat{\mu} s_{1,j})^2}{s_{1,j}}}$$

**Biais des estimateurs du maximum de vraisemblance :**

**Biais de  $\hat{\mu}$  :**  $\hat{\mu} = \frac{\sum_{j=1}^{m_1} X_{t_{j,m_1}}^{(1)}}{n \times t_{1,m_1}}$

Or  $X_{t_{j,m_1}}^{(1)} \sim \mathcal{N}(\mu t_{1,m_1}, \sigma^2 t_{1,m_1})$

Donc :

$$\begin{aligned}\mathbb{E}[\hat{\mu}] &= \mathbb{E}\left[\frac{\sum_{i=1}^n X_{t_{1,m_1}}^{(i)}}{n \times t_{1,m_1}}\right] \\ &= \frac{\sum_{i=1}^n \mathbb{E}[X_{t_{1,m_1}}^{(i)}]}{n \times t_{1,m_1}} \\ &= \frac{d \times \mu t_{1,m_1}}{n \times t_{1,m_1}} \\ \mathbb{E}[\hat{\mu}] &= \mu\end{aligned}$$

Par conséquent  $\hat{\mu}$  est sans biais

**Biais de  $\hat{\sigma}^2$  :** • On sait que  $y_{i,j} \sim \mathcal{N}(\mu s_{1,j}, \sigma^2 s_{1,j})$

Donc :

$$\begin{aligned}\mathbb{E}[(y_{i,j})^2] &= \text{Var}[y_{i,j}] + \mathbb{E}[y_{i,j}]^2 \\ &= \sigma^2 s_{1,j} + \mu^2 s_{1,j} \\ \mathbb{E}[(y_{i,j})^2] &= s_{1,j}(\sigma^2 + \mu^2 s_{1,j})\end{aligned}$$

• On sait que  $X_{t_{1,m_1}}^{(i)} \sim \mathcal{N}(\mu t_{1,m_1}, \sigma^2 t_{1,m_1})$  :

Donc  $\sum_{i=1}^n X_{t_{1,m_1}}^{(i)} \sim \mathcal{N}(\mu n \times t_{1,m_1}, \sigma^2 n \times t_{1,m_1})$

$$\hat{\mu} = \frac{\sum_{i=1}^n X_{t_{1,m_1}}^{(i)}}{n \times t_{1,m_1}} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n \times t_{1,m_1}}\right)$$



Par conséquent :

$$\begin{aligned}
\mathbb{E}[\hat{\mu}^2 \times \frac{t_{1,m_1}}{m_1}] &= \frac{t_{1,m_1}}{m_1} \mathbb{E}[\hat{\mu}^2] \\
&= \frac{t_{1,m_1}}{m_1} [\text{Var}[\hat{\mu}] + \mathbb{E}[\hat{\mu}]^2] \\
&= \frac{t_{1,m_1}}{m_1} [\frac{\sigma^2}{n \times t_{1,m_1}} + \mu^2] \\
\mathbb{E}[\hat{\mu}^2 \times \frac{t_{1,m_1}}{m_1}] &= \frac{\sigma^2}{m_1 n} + \mu^2 \frac{t_{1,m_1}}{m_1}
\end{aligned}$$

- Finalement à partir de ces résultats intermediaires on calcule le biais de  $\hat{\sigma}^2$

$$\begin{aligned}
\mathbb{E}[\hat{\sigma}^2] &= \mathbb{E}[\frac{1}{nm_1} \sum_{i=1}^n \sum_{j=1}^{m_1} \frac{(y_{i,j})^2}{s_{1,j}} - \hat{\mu}^2 \frac{t_{1,m_1}}{n}] \\
&= \frac{1}{nd} \sum_{i=1}^n \sum_{j=1}^{m_1} \frac{\mathbb{E}[(y_{i,j})^2]}{s_{1,j}} - \mathbb{E}[\hat{\mu}^2 \frac{t_{1,m_1}}{n}] \\
&= \frac{1}{nm_1} \sum_{i=1}^n \sum_{j=1}^{m_1} (\sigma^2 + \mu^2 s_{1,j}) - (\frac{\sigma^2}{nm_1} + \mu^2 \frac{t_{1,m_1}}{n}) \\
&= \sigma^2 + \frac{\mu^2 t_{1,m_1}}{n} - \frac{\sigma^2}{nm_1} - \mu^2 \frac{t_{1,m_1}}{n} \\
\mathbb{E}[\hat{\sigma}^2] &= \sigma^2 (\frac{nm_1 - 1}{nm_1})
\end{aligned}$$

Par conséquent  $\hat{\sigma}^2$  est biaisé mais il est simple de le débiaisé pour obtenir  $\hat{\sigma}^{2'}$  :

$\hat{\sigma}^{2'} = \frac{nm_1}{nm_1 - 1} \hat{\sigma}^2$  Ainsi :

$$\boxed{\hat{\sigma}^{2'} = \frac{1}{nm_1 - 1} \times \sum_{i=1}^n \sum_{j=1}^{m_1} \frac{(y_{i,j} - \hat{\mu} s_{1,j})^2}{s_{1,j}}}$$

## VII.2 : Calcul des estimateurs des moments dans le cas où $t_{i,j} = t_{1,j}$ et $m_i = m_1$ pour le processus de Wiener

**Estimateur des moments de  $\mu$ ,  $\tilde{\mu}$  :** On sait que  $X_{t_{1,m_1}}^{(i)} \sim \mathcal{N}(\mu t_{1,m_1}, \sigma^2 t_{1,m_1})$  et donc  $\frac{X_{t_{1,m_1}}^{(i)}}{t_{1,m_1}} \sim \mathcal{N}(\mu, \frac{\sigma^2}{t_{1,m_1}})$

Ainsi  $\mathbb{E}[\frac{X_{t_{1,m_1}}^{(i)}}{t_{1,m_1}}] = \mu$

Donc  $\tilde{\mu} = \frac{1}{n} \sum_{i=1}^n \frac{X_{t_{1,m_1}}^{(i)}}{t_{1,m_1}}$

Finalement :

$$\boxed{\tilde{\mu} = \frac{1}{nt_{1,m_1}} \sum_{i=1}^n X_{t_{1,m_1}}^{(i)} = \hat{\mu}}$$

**Estimateur des moments de  $\sigma^2$ ,  $\tilde{\sigma}^2$  :** On sait que  $y_{i,j} \sim \mathcal{N}(\mu s_{1,j}, \sigma^2 s_{1,j})$

Donc  $\frac{y_{i,j} - \mu s_{1,j}}{\sqrt{s_{1,j}}} \sim \mathcal{N}(0, \sigma^2)$

Finalement :  $\boxed{\tilde{\sigma}^2 = \frac{1}{nm_1} \sum_{i=1}^n \sum_{j=1}^{m_1} \frac{(y_{i,j} - \tilde{\mu} s_{1,j})^2}{s_{1,j}}}$

### VII.3 : Calcul des estimateurs du maximum de vraisemblance dans le cas où $t_{i,j} = t_j$ et $m_i = m_1$ pour le processus Gamma

**Notations :**

- $\hat{a}$  : l'estimateur de  $a$
- $\hat{b}$  : l'estimateur de  $b$

**Méthode de maximum de vraisemblance :** Pour calculer les estimateurs, on procède par méthode du maximum de vraisemblance en partant du fait que :  $\Delta Y_{i,j} \sim \Gamma(as_{1,j}, b), \forall 1 \leq i \leq n, \forall 1 \leq j \leq m_1$  La fonction de vraisemblance est alors :

$$\begin{aligned} \mathcal{L}(y_{1,1}, \dots, y_{1,n}, y_{1,m_1}, \dots, y_{n,m_1}, a, b) &= \prod_{i=1}^n \prod_{j=1}^{m_1} f_{Y_{i,j}}(y_{i,j}, a, b) \\ &= \prod_{i=1}^n \prod_{j=1}^{m_1} y_{i,j}^{s_{1,j}a-1} e^{-ay_{i,j}} \frac{a^{s_{1,j}a}}{\Gamma(s_{1,j}a)} \\ &= e^{-a \sum_{i=1}^n \sum_{j=1}^{m_1} y_{i,j}} \prod_{j=1}^{m_1} \left( \frac{a^{s_{1,j}a}}{\Gamma(s_{1,j}a)} \right)^n \prod_{i=1}^n y_{i,j}^{s_{1,j}a-1} \\ \ln(\mathcal{L}(y_{1,1}, \dots, y_{1,n}, y_{1,m_1}, \dots, y_{n,m_1}, a, b)) &= -a \sum_{i=1}^n \sum_{j=1}^{m_1} y_{i,j} + \sum_{i=1}^n \sum_{j=1}^{m_1} \ln(y_{i,j})(s_{1,j}a-1) - n \sum_{j=1}^{m_1} \ln(\Gamma(as_{1,j})) + na \sum_{j=1}^{m_1} \ln(a)s_{1,j} \\ \frac{\partial \ln(\mathcal{L}(y_{1,1}, \dots, y_{1,n}, y_{1,m_1}, \dots, y_{n,m_1}, a, b))}{\partial a} &= - \sum_{i=1}^n \sum_{j=1}^{m_1} y_{i,j} + \frac{na}{b} \sum_{j=1}^{m_1} s_{1,j} \end{aligned}$$

Donc :  $\hat{b} = \frac{n\hat{a} \sum_{j=1}^{m_1} s_{1,j}}{\sum_{i=1}^n \sum_{j=1}^{m_1} Y_{i,j}}$

$$\frac{\partial \ln(\mathcal{L}(y_{1,1}, \dots, y_{1,n}, y_{1,m_1}, \dots, y_{n,m_1}, a, b))}{\partial a} = \sum_{i=1}^n \sum_{j=1}^{m_1} \ln(y_{i,j})s_{1,j} - n \sum_{j=1}^{m_1} s_{1,j} \frac{\Gamma'(as_{1,j})}{\Gamma(as_{1,j})} + n \sum_{j=1}^{m_1} \ln(b)s_{1,j}$$

En remplace  $b$  par  $\frac{n\alpha t_{1,m_1}}{\sum_{i=1}^n Y_{n,j}}$ , on obtient que  $\hat{a}$  est la solution de l'équation suivante :

$$\sum_{i=1}^n \sum_{j=1}^{m_1} \ln(Y_{i,j})s_{1,j} - n \sum_{j=1}^{m_1} s_{1,j} \frac{\Gamma'(as_{1,j})}{\Gamma(as_{1,j})} + n \times \ln\left(\frac{nat_{1,m_1}}{\sum_{i=1}^n Y_{n,j}}\right) \sum_{j=1}^{m_1} s_{1,j} = 0$$

### VII.4 : Calcul des estimateurs des moments dans le cas où $t_{i,j} = t_j$ et $m_i = m_1$ pour le processus Gamma

**Notations :**

- $\hat{a}$  : l'estimateur de  $a$
- $\hat{b}$  : l'estimateur de  $b$

**Méthode des moments :** On sait que :  $\frac{Y_{i,j}}{s_{1,j} \sim \Gamma(as_{1,j}, bs_{1,j})}$ , et que  $\frac{Y_{i,j}}{\sqrt{s_{1,j}}} \sim \Gamma(as_{1,j}, b\sqrt{s_{1,j}}), \forall 1 \leq i \leq n, \forall 1 \leq j \leq m_1$

Donc :  $\mathbb{E}\left[\frac{Y_{i,j}}{s_{1,j}}\right] = \frac{a}{b}$

et  $Var\left[\frac{Y_{i,j}}{\sqrt{s_{1,j}}}\right] = \frac{a}{b^2}$

Alors

$$b = \frac{\mathbb{E}\left[\frac{Y_{i,j}}{s_{1,j}}\right]}{\text{Var}\left[\frac{Y_{i,j}}{\sqrt{s_{1,j}}}\right]}$$

Et

$$a = \frac{(\mathbb{E}\left[\frac{Y_{i,j}}{s_{1,j}}\right])^2}{\text{Var}\left[\frac{Y_{i,j}}{\sqrt{s_{1,j}}}\right]}$$

Ainsi :

$$\hat{b} = \frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^{m_1} \frac{Y_{n,j}}{s_{1,j}}}{\sum_{j=1}^{m_1} \left( \frac{Y_{i,j}}{\sqrt{s_{1,j}}} - \frac{1}{m_1} \sum_{j=1}^{m_1} \frac{Y_{i,j}}{\sqrt{s_{1,j}}} \right)^2}$$

Et :

$$\hat{a} = \frac{m_1}{n} \sum_{i=1}^n \frac{[\sum_{j=1}^{m_1} \frac{Y_{n,j}}{s_{1,j}}]^2}{\sum_{i=1}^n \left( \frac{Y_{i,j}}{\sqrt{s_{1,j}}} - \frac{1}{m_1} \sum_{j=1}^{m_1} \frac{Y_{i,j}}{\sqrt{s_{1,j}}} \right)^2}$$