

UNIVERSITÉ ASSANE SECK DE ZIGUINCHOR



L'excellence, ma référence

UFR SCIENCES ET TECHNOLOGIES
DÉPARTEMENT DE MATHÉMATIQUES

MÉMOIRE DE MASTER

DOMAINE : SCIENCES ET TECHNOLOGIES
MENTION : MATHÉMATIQUES ET APPLICATIONS
SPÉCIALITÉ : MATHÉMATIQUES APPLIQUÉES
OPTION : STATISTIQUE

Présenté par :

WASSA TOURE

TITRE

**Différentes procédures d'estimation de la loi
géométrique exponentielle étendue appliquée à
des données médicales**

Sous la direction de : Dr. Emmanuel Nicolas CABRAL

Soutenu publiquement le 09 Mars 2024 devant le jury ci-après :

Prénom(s) et Nom	Grade	Qualité	Établissement
Clément MANGA	Professeur Assimilé	Président	UASZ
Alassane DIEDHIOU	Professeur Titulaire	Examineur	UASZ
Mor NDONGO	Maître de Conférences Titulaire	Examineur	UASZ
Emmanuel N. CABRAL	Professeur Assimilé	Directeur	UASZ

Année Universitaire 2022-2023

DÉDICACES

Ce travail de recherche est dédié

À **Alhassane Ndiaye** qui m'a donné l'amour paternel et qui n'a ménagé aucun effort pour ma réussite, paix à son âme et que firdaws soit sa demeure éternelle.

À ma mère **Amy samaké**.

À mon homonyme **Wassa Touré**.

À mes frères, plus particulièrement **Omar Keita, Lamine Bâ et Hamidou Bâ** et à ma sœur **Mbarcké Bâ**.

À **Babacar Diouf** et à ma fille chérie **Nandy sankharé**.

À mes cousins et cousines, plus particulièrement **Fatou Touré**.

À ma grand-mère bien aimée **Khady coulibaly**.

À mes oncles et mes tantes.

À **Ababacar Ndiaye** un oncle que j'apprécie beaucoup.

À ma marraine à l'université **Ndeye Awa Ndiaye**.

À **khady Sarr** Boucotte à Ziguinchor et ses enfants plus particulièrement **Mariama Thior**.

À l'amicale des étudiants ressortissants de la commune de Ndiaffate à Ziguinchor.

À mes voisins du campus des chambres 46D pavillon D, 33E et 4E pavillon E plus particulièrement **Famboj Boh, Nogaye Bâ, Mouhsinatou Mbacké Thiaw et Penda Bâ**.

À tous ceux qui se soucient mon bien-être et ma réussite et aussi ceux qui m'ont soutenus et assistés.

REMERCIEMENTS

Je rends grâce à Allah qui m'a donné la force, le courage et la santé d'arriver au bout de mon travail.

Je tiens à remercier au plus profond de mon cœur mon encadreur, **Pr. Emmanuel Nicolas CABRAL**, qui a accepté de m'accompagner tout au long de ce travail grâce à sa disponibilité, ses encouragements, ses conseils et ses suggestions. Ces derniers m'ont vraiment motivé. Merci encore une fois de plus, ça était un réel plaisir de travailler avec vous.

Je tiens à remercier vivement les membres du jury, **Pr. Clément MANGA** d'accepter d'être le président du jury, **Pr. Alassane DIÉDHIOU** et **M. Mor NDONGO** pour avoir acceptés d'évaluer ce travail de recherche.

Je remercie mes parents plus particulièrement ma mère **Amy Samaké** et mon homonyme **Wassa touré** de m'avoir soutenus et aidés ainsi que toute ma famille.

Je voudrais remercier sincèrement ces deux personnes **Ababacar NDIAYE** et **Babacar DIOUF** de m'avoir soutenu et de m'encourager.

Je remercie également toutes personnes de près ou de loin qui m'ont soutenu et aidée durant tout ce travail.

Mes remerciements à l'association des élèves et étudiants musulmans du Sénégal (AEEMS), plus particulièrement la section universitaire de Ziguinchor (SECUZ) pour m'avoir assisté et soutenu.

J'adresse mes remerciements à mes aînés, **Mariama Thior**, Fatou DIENG et Lamine MANE qui m'ont apportés leurs aides, à mes camarades de promotion, Seydou BA, Ndeye Khady GNINGUE, Fatou MBATHIE, Isseu THIAM, Fatoumata Bineta BA, Mariama DIOP, Babacar saddik NDIOR, Salif DIALLO, Cheikh FALL, Cheikh FAYE, Fallou DIAGNE, Omar PANÉ, Siaka DIÉMÉ, Abdourahmane KONÉ, Ibrahima DIALLO, etc... pour ne citer que ceux-la.

Résumé

Tout d'abord, nous avons considéré sept(7) différentes méthodes d'estimation pour les paramètres inconnus de la loi géométrique exponentielle étendue. Nous avons introduit ces différentes méthodes que sont l'estimateur du maximum de vraisemblance, la méthode du percentile, la méthode des moindres carrés ordinaires et pondérés, la méthode du produit maximal des espacements, la méthode de Cramer-Von-Mise qui est un estimateur de distance minimale, et la méthode d'Anderson-Darling. Mais avant d'introduire ces méthodes, nous avons donné quelques notions nécessaires de la géométrie exponentielle étendue telles que la densité de probabilité, la fonction de répartition, la fonction de survie, la fonction de hasard ou de risque ainsi que l'espérance, la variance, le mode, la médiane et les relations qui existent entre eux. Ensuite, nous avons comparé en terme d'erreurs relatives moyennes (ERM) et d'erreurs quadratiques moyennes (EQM) les différents estimateurs à l'aide d'une étude de simulation numérique qui est effectué grâce au logiciel R. Les résultats de simulation nous ont montré que l'estimateur du produit maximal des espacements présente les erreurs relatives moyennes (ERM) et erreurs quadratiques moyennes (EQM) les plus faibles pour les deux paramètres. Ce qui s'avère être la méthode la plus efficace par rapport aux autres méthodes. Donc on a conclu que c'est la meilleure méthode pour estimer les paramètres de la loi géométrique exponentielle étendue par rapports aux autres. Et enfin, pour une illustration, nous avons considéré deux ensembles de données importants appliqués à cette méthodologie proposée. Et c'est ce qui nous a montré que la loi géométrique exponentielle étendue est le modèle parfait à utiliser pour les données relatives à la durée de vie.

Table des matières

Dédicaces	i
Remerciements	ii
Résumé	iii
Abréviations	vi
Liste des tableaux	vii
Table des figures	viii
Introduction Générale	1
1 Rappels	3
1.1 Notion de probabilité	3
1.1.1. Variable aléatoire discrète	3
1.1.2. Variable aléatoire réelle continue	6
1.1.3. Couple de v.a. continues	8
1.2 Les outils statistiques	9
1.2.1. Estimateur	9
1.2.2. Quelques méthodes d'estimation	10
1.3 Quelques Modèles de durée de vie	11
1.3.1. Loi de Weibull	11
1.3.2. Loi Gamma	12
1.3.3. Loi Log-normale	12
1.3.4. Loi Exponentielle généralisée	13
2 Loi géométrique exponentielle étendue	15
2.1 Loi géométrique exponentielle étendue	15
2.1.1. La fonction densité de probabilité	17
2.1.2. La fonction de répartition	18

2.1.3.	La fonction de survie(également appelée fiabilité)	19
2.1.4.	La fonction de risque ou de hasard	19
2.1.5.	Relation entre fonctions de répartition, de densité et de risque	21
2.2	Fonction génératrice des moments	22
2.2.1.	Espérance de la loi GEE	23
2.2.2.	Variance de la GEE	23
2.2.3.	Mode de la GEE	23
2.2.4.	Médiane de la GEE	24
2.2.5.	Relations entre Mode, Médiane et Espérance de la GEE	25
2.2.6.	L'entropie de Shannon	25
3	Méthodes d'estimation	27
3.1	Méthode du Maximum de Vraisemblance	27
3.2	Méthode des percentiles	30
3.3	Méthode des moindres carrés ordinaires	31
3.4	Méthode des moindres carrés pondérés	32
3.5	Méthode du produit maximal des espacements	33
3.6	Méthode de distance minimale de Cramer-Von Mises	35
3.7	Méthodes d'Anderson-Darling	36
4	Simulation Numérique	39
4.1	Les résultats de la simulation	39
4.2	Applications	43
4.2.1.	Les données de Boag	45
4.2.2.	Les enfants exposés à la transmission du VIH	48
4.3	Conclusion	50
	Conclusion générale	51
	Bibliographie	52

Abréviations & Notations

GEE	: Géométrie exponentielle étendue
EQM	: Erreur quadratique moyenne
ERM	: Erreur relative moyenne
GEB	: Géométrie exponentielle Bêta
GEEX	: Géométrie exponentielle exponentielle
GEEC	: Géométrie exponentielle exponentielle complémentaire
GEG	: géométrie exponentielle généralisée
MV	: Méthode du maximum de vraisemblance
PC	: Méthode du percentile
MCO	: Méthode des moindres carrés ordinaires
MCP	: Méthode des moindres carrés pondérés
PME	: Méthode du produit maximal des espacements
CVM	: Méthode de Cramer-Von-Mise
AD	: Méthode d'Anderson-Darling
MIF	: Matrice information de Fisher
FDP	: Fonction densité de probabilité
FR	: Fonction de répartition
FH	: Fonction de hasard
FS	: Fonction de survie
WE	: Loi de Weibull
GA	: Loi de Gamma
LN	: Loi de Log-normal
EG	: Loi exponentielle généralisée
$L_2(\cdot)$: Fonction dilogarithme
KS	: Test de Kolmogorov-Smirnov
i.e	: c'est-à-dire
v.a	: Variable aléatoire

Liste des tableaux

4.1.1 Les estimations de $\gamma = 0.5$ et $\lambda = 2$ des différentes méthodes	41
4.1.2 Les estimations de $\gamma = 2$ et $\lambda = 4$ des différentes méthodes	43
4.2.3 Données relatives à l'âge (en mois) de 18 patients décédés d'autres causes que le cancer	45
4.2.4 Résultats des estimations des données du tableau 4.2.3	45
4.2.5 Comparaison des ERM des estimations obtenues à partir des différentes méthodes à celle de PME en utilisant les données du tableau 4.2.3	46
4.2.6 Résultats du test KS (p-value), de l'AIC, de l'AICc, du HQIC et du CAIC pour les différentes lois de probabilités en utilisant les données du tableau 4.2.3	47
4.2.7 Données relatives au temps de retournement du sérum (en jours) de 143 enfants nés de mères infectées par le VIH.	48
4.2.8 Résultats des estimations des données du tableau 4.2.7	48
4.2.9 Comparaison des ERM des estimations obtenues à partir des différentes méthodes à celle du PME en utilisant les données du tableau 4.2.7	49
4.2.10 Résultats du test KS (p-value), de l'AIC, de l'AICc, du HQIC et du CAIC pour les différentes lois de probabilités en utilisant les données du tableau 4.2.7	50

Table des figures

1.3.1 FDP des modèles	13
1.3.2 FR des modèles	14
2.1.1 densité de GEE	17
2.1.2 La fonction de répartition de GEE	18
2.1.3 La fonction de survie de GEE	20
2.1.4 Fonction de hasard de GEE.	21
4.1.1 ERMes et EQMs des estimations de $\gamma=0.5$ et $\lambda=2$ pour N échantillons simulés, en considérant différentes valeurs de n :AD, CVM, MV, MCO, MCP, PC et PME	40
4.1.2 ERMes et EQMs des estimations de $\gamma=2$ et $\lambda=4$ pour N échantillons simu- lés, en considérant différentes valeurs de n :AD, CVM, MV, MCO, MCP, PC et PME	42
4.2.3 Fonctions de survies des différents modèles des 18 patients décédés d'autres causes que le cancer	46
4.2.4 fonctions de survies des différents modèles en utilisant les données du tableau 4.2.7	49

Introduction Générale

L'étude de la durée de vie des organismes, des dispositifs, des structures, des matériaux, etc., est d'une importance majeure dans les sciences biologiques et les sciences de l'ingénieur. Une part importante de cette étude est consacrée à la description mathématique de la durée de vie par une distribution de défaillance. De nombreux Mathématiciens s'intéressent aux lois de probabilités (c'est-à-dire les distributions de défaillance) qui peuvent être utilisées pour analyser les données relatives à la durée de vie. C'est dans ce sens que les dérivées de la loi exponentielle standard telles que la loi de Gamma, la loi de Weibull et la loi Exponentielle Généralisée (voir dans [16]) ont été introduites. Une autre dérivée utile de la loi exponentielle standard est connue sous le nom de loi géométrique exponentielle étendue. Adamidis et Loukas (voir dans [1]) ont été les premiers à proposer une distribution géométrique exponentielle à deux paramètres, dans laquelle la fonction de risque (défaillance) pouvait être décroissante. Dans ce travail, nous nous intéressons uniquement à la loi géométrique exponentielle étendue (GEE). Une des particularités de cette loi est que la croissance ou la décroissance de sa fonction de risque dépend des valeurs de ses paramètres. L'importance de la distribution géométrique exponentielle étendue réside dans sa polyvalence pour la modélisation de divers phénomènes du monde réel (voir dans [2],[24],[13]), en particulier ceux qui présentent des taux d'aléas (risques) qui changent avec le temps. Contrairement à la distribution géométrique exponentielle standard, la GEE peut modéliser des taux d'aléas monotones décroissants. Elle est donc bien adaptée aux données où le taux de défaillance diminue initialement, puis augmente avec le temps, ce qui représente une période de "vieillesse" avant la défaillance. La GEE introduit un paramètre supplémentaire par rapport à la distribution géométrique exponentielle standard, ce qui permet un contrôle plus fin de la forme du taux de hasard. Cela offre une plus grande flexibilité dans l'adaptation du modèle à différents types de données. La GEE trouve une application dans l'analyse de fiabilité, en particulier pour la modélisation de la durée de vie des composants ou des systèmes avec des taux de défaillance variables. Cette méthode est utile dans des domaines tels que l'ingénierie, la fabrication et le contrôle de la qualité. Dans l'analyse de survie, la GEE peut être utilisée pour modéliser le temps qui s'écoule avant un événement tel que le décès ou l'apparition d'une maladie, en particulier lorsque le taux d'aléa suit un modèle non monotone. Cela a des implications pour la recherche médicale, l'épidémiologie et la science actuarielle. La GEE peut être utile pour modéliser des données financières

présentant des risques variables dans le temps, comme les délais de défaillance de crédit ou la durée de vie des actifs. Cela trouve des applications dans la gestion des risques et la modélisation financière. La GEE est généralement efficace sur le plan informatique, ce qui le rend adapté aux grands ensembles de données et aux applications en temps réel. Dans l'ensemble, la loi géométrique exponentielle étendue offre un outil précieux pour l'analyse des données avec des taux de hasard non monotones. Sa flexibilité, la diversité de ses applications et son efficacité informatique en font un choix judicieux pour divers domaines, notamment l'analyse de fiabilité, l'analyse de survie, la finance et l'économie. En raison de son importance, certaines dérivées de la loi GEE ont été proposées, telles que la loi géométrique exponentielle Bêta GEB (voir [27]), la loi géométrique exponentielle exponentielle GEEX (voir [20]), la loi géométrique exponentielle exponentielle complémentaire GEEC (voir [17]), et la loi géométrique exponentielle généralisé GEG (voir [5]). Nous avons deux objectifs pour ce travail. Premièrement, c'est de trouver la méthode d'estimation la plus efficace parmi les sept (7) méthodes d'estimation de la loi GEE, ce qui serait intéressant pour les statisticiens. Deuxièmement, c'est de montrer que la loi GEE est une alternative simple à utiliser dans les applications médicales. Dans ce travail, nous allons faire d'abord un petit rappel sur les notions statistiques et probabilistes. Ensuite, nous introduisons la loi géométrique exponentielle étendue en donnant sa fonction densité, de répartition, de survie et de risque et aussi ses propriétés c'est-à-dire son espérance, sa variance, son mode et sa médiane et leurs relations. Nous étudierons aussi les différentes méthodes d'estimation suivantes : la méthode du maximum de vraisemblance, de percentile, des moindres carrés ordinaires et pondérés, du produit maximal des espacements, de Cramer-Von-Mises et de Anderson-Darling. Et enfin, l'efficacité de ces différentes méthodes d'estimation sera comparée à l'aide de simulations numériques.

Rappels

La théorie des probabilités est l'étude mathématique des phénomènes caractérisés par le hasard et l'incertitude ; la Statistique est l'activité qui consiste à recueillir, traiter et interpréter un ensemble de données. Il existe des interconnexions entre ces deux domaines des sciences de l'aléatoire. Pour commencer cette partie, nous allons définir quelques notions de bases en probabilité et en statistique qui seront d'une importance capitale pour la suite de notre travail.

1.1 Notion de probabilité

Le calcul des probabilités prend réellement naissance en 1654 lors d'un échange de correspondance entre deux mathématiciens, Blaise Pascal (à Paris) et Pierre Simon de Fermat (à Toulouse) au sujet de problèmes posés par le chevalier de Méré (un ami à Pascal). Toutefois, aucun des deux mathématiciens n'a publié ses résultats et c'est Christian Huygens qui rassembla et publia ces problèmes déjà résolus avec des nouveaux. Leibniz élargit le champs des calculs des probabilités en l'appliquant à tous les domaines : sociaux, moraux etc (pour plus de détails voir [37]).

Donnons quelques définitions qui seront utiles pour la suite.

Définition 1.1.1 Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et (E, \mathcal{E}) un espace mesurable, une application mesurable $X : \Omega \rightarrow E$ est appelée variable aléatoire à valeurs dans E . Lorsque $E \subset \mathbb{R}$, on parle de variable aléatoire réelle. On appelle support d'une variable aléatoire X , l'ensemble de ses valeurs possibles. On le note $S(X)$ ou $X(\Omega)$ et est inclus dans E .

On distingue deux types de variables aléatoires :

1.1.1. Variable aléatoire discrète

Définition 1.1.2 On appelle une v.a discrète définie sur (Ω, \mathcal{A}) une application $X : \Omega \rightarrow \mathbb{R}$ avec $X(\Omega)$ dénombrable, i.e en correspondance bijective avec \mathbb{N} , et pour tout x réel :

$$X^{-1}(x) = \{\omega \in \Omega / X(\omega) = x\} \in \mathcal{A}$$

ce qui exprime que $X^{-1}(x)$ est un évènement.

Loi de probabilité

La probabilité notée P_X est définie sur $X(\Omega)$ par :

$$P_X(x) = P \{X^{-1}(x)\} = P \{\omega \in \Omega / X(\omega) = x\}$$

s'appelle la probabilité image de P par X . La loi de probabilité est alors définie par les probabilités individuelles :

$$p_i = P(X = x_i) = P \{X^{-1}(x_i)\}, \quad i \in \mathbb{N}$$

On appelle alors distribution ou loi de probabilité de X l'ensemble des couples $(x_i, p_i)_{i \in \mathbb{N}}$.

Fonction de répartition

On appelle fonction de répartition de la v.a X , la fonction définie pour tout réel par :

$$F(x) = P \{X \leq x\} = \sum_{\{p_i / x_i \leq x\}} p_i$$

Cette valeur représente la probabilité de toutes les réalisations strictement inférieures au réel x .

Propriété 1.1.1 *La fonction de répartition ainsi défini vérifie les propriétés suivantes :*

1. Pour tout $x \in \mathbb{R}$, $0 \leq F(x) \leq 1$.
2. F est croissante (au sens large).
3. $\lim_{x \rightarrow +\infty} F(x) = 1$, $\lim_{x \rightarrow -\infty} F(x) = 0$.
4. F est continue à droite en tout réel x_0 avec limite à gauche :

$$F(x_0) = \lim_{x \rightarrow x_0^+} F(x) \quad \text{et} \quad \lim_{x \rightarrow x_0^-} F(x) = P(X < x_0).$$

Moments d'une v.a discrète

► Espérance mathématique

Définition 1.1.3 *On appelle espérance mathématique de la v.a. X la quantité, si elle existe :*

$$\mathbb{E}(X) = \sum_{i \in \mathbb{N}} p_i x_i$$

Propriété 1.1.2 ♣ *Si on ajoute une constante à une v.a., il est de même pour son espérance :*

$$\mathbb{E}(X + a) = \mathbb{E}(X) + a \quad \forall a \in \mathbb{R}$$

♣ *Si on multiplie une v.a. par une constante, il en sera de même pour son espérance :*

$$\mathbb{E}(aX) = a\mathbb{E}(X) \quad \forall a \in \mathbb{R}$$

♣ *L'espérance mathématique d'une somme de deux v.a. est la somme des espérances :*

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

On résume ses trois propriétés en disant que l'espérance est un opérateur linéaire :

$$\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}(X) + \beta\mathbb{E}(Y) \quad \forall \alpha \in \mathbb{R}, \quad \forall \beta \in \mathbb{R}$$

► Variance

Il s'agit d'un indicateur mesurant la dispersion des valeurs x_i autour de $\mathbb{E}(X)$:

$$\mathbb{V}(X) = \sum_{i \in \mathbb{N}} p_i [x_i - \mathbb{E}(X)]^2$$

lorsque cette quantité existe. Elle s'écrit aussi :

$$\mathbb{V}(X) = \mathbb{E}[X - \mathbb{E}(X)]^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Cette quantité $\mathbb{V}(X)$ est notée aussi par σ_X^2 , σ_X désigne l'écart-type de la v.a. X .

Propriété 1.1.3 *Par définition :*

$$\mathbb{V}(X) \geq 0.$$

Pour tout réel a , on a :

$$\mathbb{V}(X + a) = \mathbb{V}(X) \quad \text{et} \quad \mathbb{V}(aX) = a^2\mathbb{V}(X)$$

Si X et Y sont deux v.a. indépendantes, alors

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$$

On définit la covariance de deux v.a. X et Y par $Cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ et on a dans le cas général :

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2Cov(X, Y)$$

► Moments centrés et non centrés

On appelle moment non centré d'ordre $r \in \mathbb{N}^*$ la quantité :

$$m_r(X) = \sum_{i \in \mathbb{N}} p_i x_i^r = \mathbb{E}(X^r)$$

L'espérance est un cas particulier du moment non centré. Le moment centré d'ordre $r \in \mathbb{N}^*$ est :

$$\mu_r(X) = \sum_{i \in \mathbb{N}} p_i [x_i - \mathbb{E}(X)]^r = \mathbb{E}[X - \mathbb{E}(X)]^r$$

Le moment centré d'ordre 2 coïncide avec la variance.

1.1.2. Variable aléatoire réelle continue

Définition 1.1.4 On appelle une v.a réelle définie sur (Ω, \mathcal{A}) une application $X : \Omega \rightarrow \mathbb{R}$ telle que pour tout intervalle $I \subset \mathbb{R}$ on ait :

$$X^{-1}(I) = \{\omega \in \Omega / X(\omega) \in I\} \in \mathcal{A}$$

Cette condition est vérifiée si pour tout réel x , on a $X^{-1}(]-\infty, x]) \in \mathcal{A}$.

Loi de probabilité

Elle est déterminée par la FR F , définie pour tout réel x par :

$$F(x) = P_X(]-\infty, x]) = P\{X^{-1}(]-\infty, x])\} = P\{\omega \in \Omega / X(\omega) \leq x\}$$

Propriété 1.1.4 La fonction de répartition ainsi défini vérifie les propriétés suivantes :

1. Pour tout $x \in \mathbb{R}$, $0 \leq F(x) \leq 1$.
2. F est croissante (au sens large).
3. $\lim_{x \rightarrow +\infty} F(x) = 1$, $\lim_{x \rightarrow -\infty} F(x) = 0$.
4. F est continue à droite en tout réel x_0 avec limite à gauche :

$$F(x_0) = \lim_{x \rightarrow x_0^+} F(x) \quad \text{et} \quad \lim_{x \rightarrow x_0^-} F(x) = P(X < x_0).$$

Loi continue

Si la fonction F est continue, on dit que X est une v.a. réelle continue.

Loi absolument continue

Dans le cas où la fonction F admet une dérivée f , celle-ci est appelée densité de probabilité de la v.a. X , de loi qualifiée d'absolument continue. La FR est alors déterminée par l'intégrale :

$$F(x) = \int_{-\infty}^x f(t)dt$$

Une densité est et positive intégrable, d'intégrale égale à 1 :

$$\int_{-\infty}^{+\infty} f(t)dt = 1$$

La probabilité d'un intervalle s'obtient en intégrant la densité sur cet intervalle :

$$P_X \{X \in [x_1, x_2]\} = \int_{x_1}^{x_2} f(t)dt$$

Moments d'une v.a. absolument continue

► Espérance mathématique

Elle est définie par :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

lorsque cette intégrale existe. Pour tout réel a :

$$\mathbb{E}(X + a) = \mathbb{E}(X) + a \quad \text{et} \quad \mathbb{E}(aX) = a\mathbb{E}(X)$$

Si X et Y deux v.a. admettent une espérance alors on a :

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

► Variance

La variance, elle est définie par :

$$\mathbb{V}(X) = \mathbb{E}[X - \mathbb{E}(X)]^2 = \int_{-\infty}^{+\infty} [x - \mathbb{E}(X)]^2 f(x)dx = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \sigma_X^2$$

lorsque cette intégrale existe. Pour tout réel a :

$$\mathbb{V}(X + a) = \mathbb{V}(X) \quad \text{et} \quad \mathbb{V}(aX) = a^2\mathbb{V}(X)$$

Si X et Y deux v.a. indépendantes admettant une variance, alors on a :

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$$

► Moments centrés et non centrés

Le **moment non centré** d'ordre $k \in \mathbb{N}^*$ de X est défini par :

$$m_k = E(X^k) = \int_{-\infty}^{+\infty} x^k f(x) dx.$$

Le **moment centré d'ordre k** est défini aussi par :

$$\mu_k = E[X - E(X)]^k = \int_{-\infty}^{+\infty} [x - \mathbb{E}(X)]^k f(x) dx.$$

1.1.3. Couple de v.a. continues

Loi du couple

Soit X et Y sont deux v.a continues, la loi de (X, Y) est donnée par sa FR :

$$F(x, y) = P(X < x, Y < y).$$

Si F est dérivable par rapport à x et y , la loi de (X, Y) admet une densité f donnée par :

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}.$$

Lois marginales

Les fonctions de répartition marginales de la v.a. X et de la v.a. Y sont définies par :

$$F_X(x) = P(X \leq x) = \lim_{y \rightarrow +\infty} F(x, y) \quad \text{et} \quad F_Y(y) = P(Y \leq y) = \lim_{x \rightarrow +\infty} F(x, y)$$

Si la loi du couple est définie par sa densité, les densités marginales sont données par :

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy \quad \text{et} \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

Lois conditionnelles

Pour une loi de densité f , les lois conditionnelles sont définies par les densités :

$$f_X(x|Y = y) = \frac{f(x, y)}{f_Y(y)} \quad \text{et} \quad f_Y(y|X = x) = \frac{f(x, y)}{f_X(x)}$$

Indépendance

Les v.a. X et Y sont dites indépendantes, si pour tout réel x et y :

$$f(x, y) = f_X(x) f_Y(y).$$

1.2 Les outils statistiques

Le terme "statistique" est tiré d'un mot grec qui signifie "constater". La statistique est la science ayant pour objet le groupement méthodique des faits sociaux qui se prêtent à une évaluation numérique, comme les impôts, la production industrielle et agricole, la population, la religion, etc. On fait aussi dériver le mot du latin "status" qui veut dire "état", et à l'origine on appelait statisticien celui qui s'occupait des affaires d'État (voir [36]). Nous allons donner quelques définitions qui seront utiles pour la suite de ce travail.

Définition 1.2.1 : Soit X une variable aléatoire qui suit une loi de probabilité, notée P_θ . On appelle un **échantillon** un ensemble (X_1, \dots, X_n) de n variables aléatoires indépendantes qui suivent la même loi que X .

1.2.1. Estimateur

Définition 1.2.2 Un **estimateur** de θ est une application T_n de E^n dans F , qui à un échantillon (X_1, \dots, X_n) de la loi P_θ associe une variable aléatoire réelle (ou plusieurs dans le cas d'un paramètre multidimensionnel) dont on peut déterminer la loi de probabilité.

Biais d'un estimateur

Définition 1.2.3 Soit (X_1, \dots, X_n) un échantillon indépendant tiré sur le modèle (X, P_θ) . L'estimateur $T_n(X_1, X_2, \dots, X_n)$ de θ a pour **biais** la quantité

$$b_n(\theta) = E_\theta(T_n) - \theta.$$

Définition 1.2.4 On dit qu'un estimateur T_n de θ est **sans biais** si pour $\theta \in \Theta$ et tout entier on a

$$E_\theta(T_n) = \theta \iff b_n(\theta) = 0.$$

Un estimateur T_n de θ est dit **asymptotiquement sans biais** si pour $\theta \in \Theta$ et tout entier on a

$$E_\theta(T_n) \rightarrow \theta \quad \text{quand } n \rightarrow \infty.$$

Convergence d'un estimateur

Définition 1.2.5 Un estimateur T_n est **convergent** si la suite de v.a (T_n) converge en probabilité vers la valeur paramètre, i.e pour tout :

$$T_n \rightarrow \theta \iff P_\theta(|T_n - \theta| < \varepsilon) \rightarrow 1, \quad \forall \varepsilon > 0, \quad n \rightarrow \infty$$

Théorème 1.2.1 Tout estimateur sans biais, ou asymptotiquement sans dont la variance tend vers 0, quand n tend vers ∞

Qualité d'un estimateur : elle se mesure par l'erreur quadratique moyenne, définie pour tout θ par

$$EQM(T_n) = \mathbb{E}_\theta (T_n - \theta)^2 = \mathbb{V}_\theta(T_n) + b_n^2(\theta).$$

1.2.2. Quelques méthodes d'estimation

Il existe plusieurs méthodes d'estimation pour estimer les paramètres d'une distribution. En particulier, on peut citer la méthode du maximum de vraisemblance.

Définition 1.2.6 Soit (X, P_θ) , $\theta \in \Theta$ un modèle statistique. On appelle **Vraisemblance** (likelihood) de l'échantillon X_1, \dots, X_n la loi de probabilité de cet échantillon notée $L(x_1, x_2, \dots, x_n; \theta)$ et définie par :

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n P_\theta(X_i) = \prod_{i=1}^n P(X_i = x_i | \theta);$$

si X est une variable aléatoire discrète, et par :

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta);$$

si X est une variable aléatoire continue de densité $f(x_i; \theta)$.

La quantité d'information de Fisher est définie par :

$$I_n(\theta) = E_\theta \left(\frac{\partial \ln L}{\partial \theta} \right)^2.$$

Définition 1.2.7 On appelle un estimateur du **maximum de vraisemblance** toute fonction $\hat{\theta}_n$ de (x_1, \dots, x_n) qui vérifie :

$$\log(L(x_1, \dots, x_n; \hat{\theta}_n)) = \max_{\theta \in \Theta} \log L(x_1, \dots, x_n; \theta)$$

c'est-à-dire $\hat{\theta}_n$ est solution de

$$\frac{\partial}{\partial \theta} \log(L(x_1, \dots, x_n; \hat{\theta}_n)) = 0$$

avec

$$\frac{\partial^2}{\partial \theta^2} \log(L(x_1, \dots, x_n; \hat{\theta}_n)) < 0.$$

Remarque 1.2.1 – Dans le cas où le paramètre à estimer est $\theta = E_\theta(X)$, moyenne théorique de la loi, l'estimateur naturel est la moyenne empirique, ou moyenne de l'échantillon :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

De même, pour estimer le paramètre $\theta = V_\theta(X)$, variance de la loi, nous retenons logiquement comme estimateur la variance empirique :

$$S_n'^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

(la variance empirique modifiée S_n^2 est obtenue en divisant par $n-1$ au lieu de n). Plus généralement, si l'un des moments d'ordre $k \in \mathbb{N}$, non centré $m_k = E_\theta(X^k)$ ou centré $\mu_k = E_\theta(X - m_1)^k$, dépend de θ , nous allons chercher un estimateur par résolution de l'équation en θ obtenue en égalant moment théorique et moment empirique correspondant, soit :

$$m_{kn} = \frac{1}{n} \sum_{i=1}^n X_i^k = m_k(\theta);$$

ou

$$\mu_{kn} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^k.$$

La solution de l'équation retenue, si elle existe et est unique, sera appelée estimateur obtenu par la **méthode des moments**.

– etc...

1.3 Quelques Modèles de durée de vie

Dans cette section, nous décrivons brièvement quelques modèles qui seront utiles pour la suite de notre travail.

1.3.1. Loi de Weibull

Définition 1.3.1 La fonction densité de probabilité de la loi Weibull est définie sous la forme suivante :

$$f_{We}(x; \beta, \eta) = \frac{\beta}{\eta} \left(\frac{x}{\eta}\right)^{\beta-1} \exp\left(-\left(\frac{x}{\eta}\right)^\beta\right); \quad x > 0 \quad (1.1)$$

en mettant en évidence les deux paramètres :

β : paramètre de forme, $\beta > 0$

η : paramètre d'échelle (vie caractéristique), $\eta > 0$

Si, pour simplifier la présentation des calculs on pose $\alpha = \frac{1}{\eta^\beta}$, la fonction densité de probabilité de la loi s'écrit :

$$f_{We}(x; \beta, \alpha) = \alpha \beta x^{\beta-1} \exp(-\alpha x^\beta) \quad (1.2)$$

Fonction de répartition : Elle est définie par la formule suivante :

$$F_{We}(x; \beta, \alpha) = 1 - \exp(-\alpha x^\beta) \quad (1.3)$$

Fonction de survie : Elle est définie par :

$$S_{We}(x; \beta, \alpha) = \exp(-\alpha x^\beta) \quad (1.4)$$

1.3.2. Loi Gamma

Définition 1.3.2 La fonction densité de probabilité de la loi Gamma est définie sous la forme suivante :

$$f_{Ga}(x; \beta, \alpha) = \frac{\alpha^\beta}{\Gamma(\beta)} x^{\beta-1} \exp(-\alpha x); \quad x > 0 \quad (1.5)$$

en mettant en évidence les deux paramètres :

β : paramètre de forme, $\beta > 0$

α : paramètre d'échelle (vie caractéristique), $\alpha > 0$

avec $\Gamma(\beta)$ la fonction de gamma incomplète définie par :

$$\Gamma(\beta) = \int_0^\infty x^{\beta-1} \exp(-x) dx$$

Fonction de survie : Elle est définie par :

$$S_{Ga}(x; \beta, \alpha) = 1 - \Gamma(\alpha x, \beta) \quad (1.6)$$

où

$$\Gamma(\alpha x, \beta) = \frac{1}{\Gamma(\beta)} \int_0^{\alpha x} u^{\beta-1} \exp(-u) du$$

Fonction de répartition : Elle est définie par :

$$F_{Ga}(x; \beta, \alpha) = 1 - S_{Ga}(x; \beta, \alpha) = \Gamma(\alpha x, \beta). \quad (1.7)$$

1.3.3. Loi Log-normale

Fonction de densité : La fonction de densité de la loi log-normale à deux paramètres avec le paramètre d'échelle σ et le paramètre de forme μ est la suivante :

$$f_{Ln}(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right); \quad -\infty < \mu < \infty, \sigma > 0, x \geq 0. \quad (1.8)$$

Fonction de survie : Elle est donnée par :

$$S_{Ln}(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_x^\infty \frac{1}{t} \exp\left(-\frac{(\log t - \mu)^2}{2\sigma^2}\right) dt = 1 - \Phi\left(\log\left(\frac{x \exp(-\mu)}{\sigma}\right)\right) \quad (1.9)$$

où $\Phi(\cdot)$ est la fonction de répartition de loi normale standard.

1.3.4. Loi Exponentielle généralisée

Définition 1.3.3 La distribution exponentielle généralisée à deux paramètres a pour fonction de densité :

$$f_{Ge}(x; \alpha, \lambda) = \alpha \lambda (1 - \exp(-\lambda x))^{\alpha-1} \exp(-\lambda x); \quad \alpha, \lambda, x > 0. \quad (1.10)$$

Fonction de répartition : Elle est définie par :

$$F_{Ge}(x; \alpha, \lambda) = (1 - \exp(-\lambda x))^{\alpha} \quad (1.11)$$

Fonction de survie : Elle est donnée par :

$$S_{Ge}(x; \alpha, \lambda) = 1 - (1 - \exp(-\lambda x))^{\alpha} \quad (1.12)$$

Les figures 1.3.1 et 1.3.2 suivantes représentent respectivement la FDP et la de FR des différents modèles.

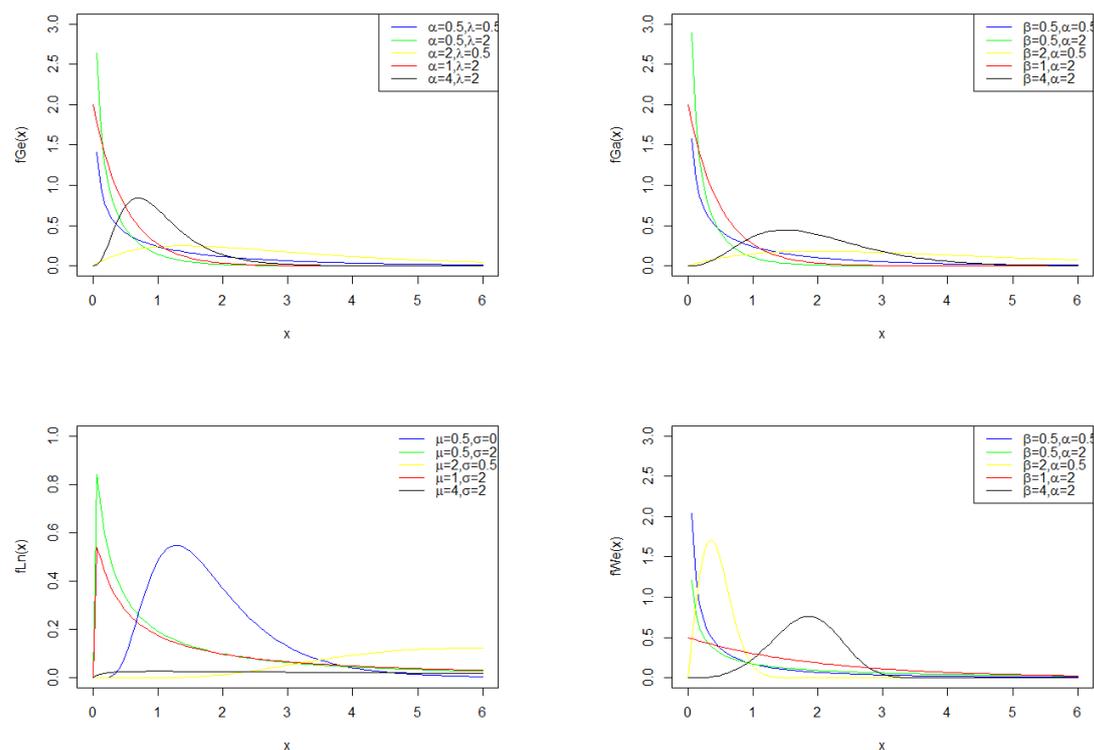


FIGURE 1.3.1 – FDP des modèles

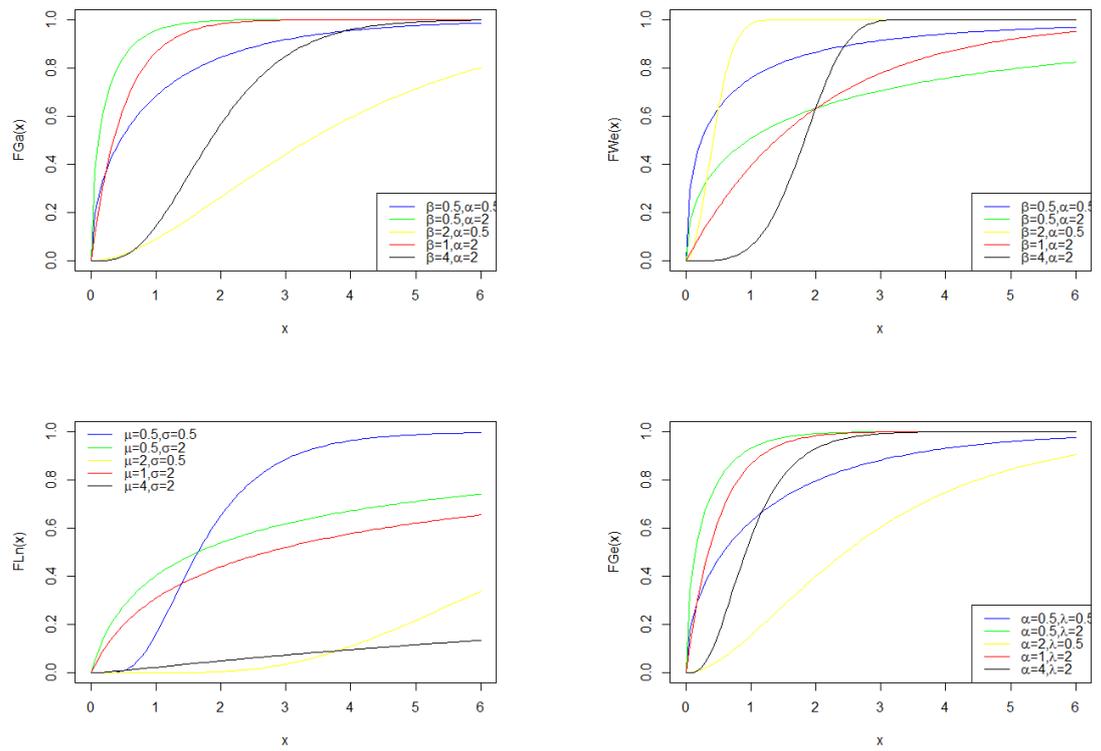


FIGURE 1.3.2 – FR des modèles

Loi géométrique exponentielle étendue

Dans ce chapitre nous allons d'abord introduire la distribution (loi) géométrique exponentielle étendue (GEE), ensuite définir quelques caractéristiques de cette distribution : fonction densité de probabilité (FDP), fonction de répartition (FR), fonction de hasard (FH), fonction de survie (FS) et les relations qui existent entre elles et enfin les moments et fonction génératrice des moments .

2.1 Loi géométrique exponentielle étendue

Supposons que la distribution conditionnelle de X étant donné que $\lambda = \beta Z$ est exponentielle avec une fonction densité de probabilité (FDP)

$$f(x; \lambda) = \lambda e^{-\lambda x}, x > 0, \lambda > 0; \quad (2.1)$$

où Z est une variable aléatoire géométrique avec fonction de probabilité

$$P(Z = z) = (1 - p)^{z-1} p; \quad z = 1, 2, \dots; \quad 0 < p < 1. \quad (2.2)$$

La distribution inconditionnelle de X est alors appelée distribution géométrique exponentielle (GE) et la FDP est la suivante

$$f(x; \beta, p) = \beta p e^{-\beta x} \left(1 - (1 - p)e^{-\beta x}\right)^{-2}, x > 0, \beta > 0, 0 < p < 1. \quad (2.3)$$

Cette distribution de X a été introduite pour la première fois (voir [23]) via une méthode permettant d'ajouter un paramètre de forme à une fonction de survie de la distribution exponentielle, et a été appelée distribution géométrique exponentielle (voir [1]). Une extension de la distribution EG a été introduit (voir [2]) . Cette distribution possède à la fois une fonction de risque décroissante et une fonction de risque croissante. Supposons que la distribution conditionnelle de $X | A$ soit modifiée comme une distribution de valeurs extrêmes avec FDP

$$f(x | a; \lambda) = a\lambda \exp(\lambda x + a(1 - \exp(\lambda x))), x > 0, \lambda > 0, a > 0. \quad (2.4)$$

Et que A soit distribuée comme une loi exponentielle avec FDP

$$f(a; \gamma) = \gamma e^{-\gamma a}, \gamma > 0, a > 0. \quad (2.5)$$

Pour obtenir la loi inconditionnelle de X , nous devons effectuer une intégration marginale de a sur la loi $f(a; \gamma)$ dans l'équation (2.4). Pour cela, nous multiplions simplement $f(x | a; \lambda)$ par $f(a; \gamma)$ et intégrons par rapport à a .

La FDP de la distribution inconditionnelle de X , notée $f(x; \gamma, \lambda)$, est donnée par

$$\begin{aligned} f(x; \gamma, \lambda) &= \int_0^{\infty} f(x | a; \lambda) f(a; \gamma) da \\ &= \int_0^{\infty} a\lambda \exp(\lambda x + a(1 - \exp(\lambda x))) \gamma e^{-\gamma a} da \end{aligned}$$

Cette intégrale peut être résolue en utilisant des techniques d'intégration appropriées. Cependant, avant de procéder, remarquons que cette intégrale n'a pas de forme fermée simple. Par conséquent, nous allons d'abord résoudre l'intégrale, puis simplifier autant que possible.

$$\begin{aligned} f(x; \gamma, \lambda) &= \int_0^{\infty} a\lambda \exp(\lambda x + a(1 - \exp(\lambda x))) \gamma e^{-\gamma a} da \\ &= \lambda\gamma \int_0^{\infty} a \exp(\lambda x + a(1 - \exp(\lambda x)) - \gamma a) da \\ &= \lambda\gamma e^{\lambda x} \int_0^{\infty} a \exp(a(1 - \exp(\lambda x) - \gamma)) da \end{aligned}$$

Nous pouvons désormais résoudre cette intégrale en utilisant des techniques appropriées, telles que le changement de variable ou l'intégration par parties, pour obtenir la forme finale de $f(x; \gamma, \lambda)$. Après l'intégration, nous avons la distribution inconditionnelle de X donnée par :

$$f(x; \gamma, \lambda) = \lambda\gamma e^{-\lambda x} \left(1 - (1 - \gamma) e^{-\lambda x}\right)^{-2}; x > 0, \lambda > 0, \gamma > 0. \quad (2.6)$$

On appelle la distribution inconditionnelle de X la loi géométrique exponentielle étendue (GEE) (voir [2]). Ce modèle met en évidence des scénarios de risques naturellement concurrents. Soit $X = \min(T_1, T_1, \dots, T_M)$, où M est une loi géométrique et T_i sont indépendants de M et sont supposés être indépendants et identiquement distribués selon la loi exponentielle; alors la variable aléatoire X suit une loi GEE avec $0 < \gamma < 1$, également connue sous le nom de la loi géométrique exponentielle GE (voir [1]). En considérant les mêmes hypothèses et $X = \max(T_1, T_1, \dots, T_M)$, la variable aléatoire X suit une loi GEE avec $\gamma > 1$, également connue sous le nom de la loi géométrique exponentielle complémentaire GEC (voir [19]).

2.1.1. La fonction densité de probabilité

L'équation (2.6) représente la fonction densité de probabilité de la distribution géométrique exponentielle étendue (GEE). On peut prouver que pour $0 < \gamma \leq 2$, la fonction densité de la GEE est monotone décroissante avec la valeur modale $\frac{\lambda}{\gamma}$ à $x = 0$. Pour $\gamma > 2$, la fonction densité augmente de $\frac{\lambda}{\gamma}$ à sa valeur modale $\frac{\lambda\gamma}{4(\gamma-1)}$ à $x = \frac{1}{\lambda} \ln(\gamma-1)$ et diminue ensuite pour s'approcher de sa valeur asymptotique de zéro. Cette densité est représentée dans la figure 2.1.1 pour certaines valeurs de γ et λ .

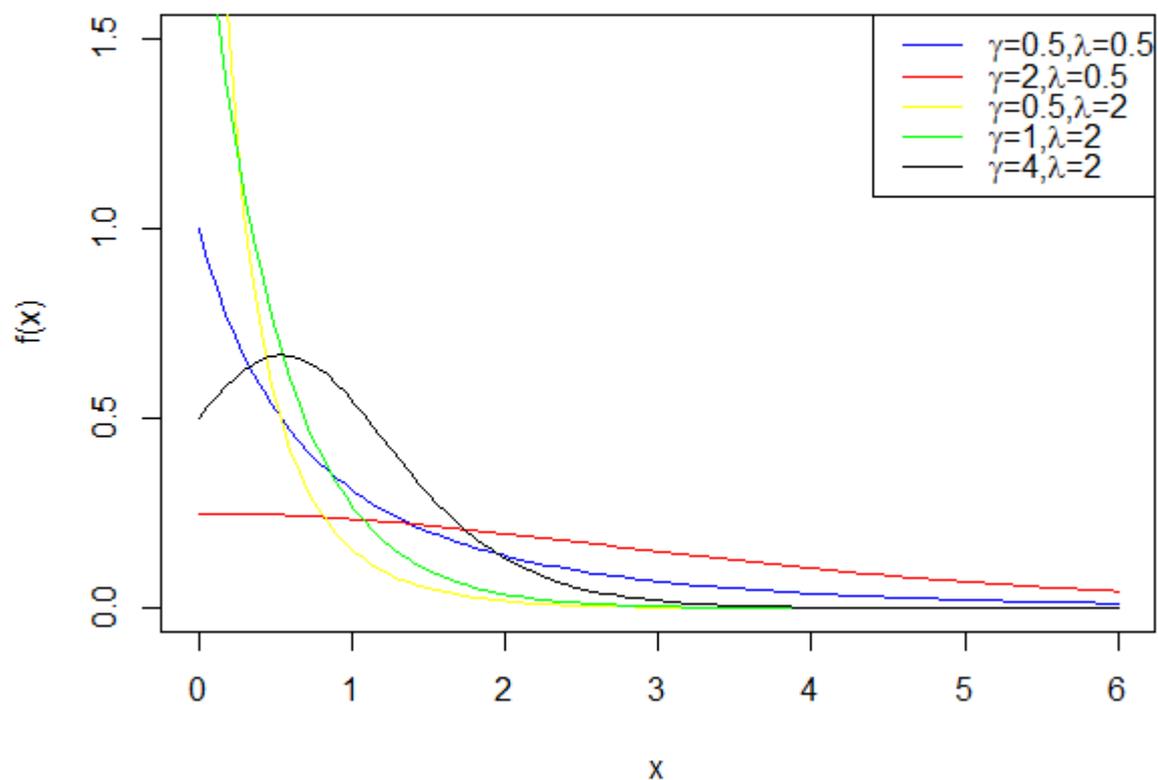


FIGURE 2.1.1 – densité de GEE

2.1.2. La fonction de répartition

Par intégration directe, nous trouvons que la fonction de répartition de la distribution EEG est donnée par :

$$\begin{aligned} F(x; \gamma, \lambda) &= \int_0^x f(t; \gamma, \lambda) dt \\ &= \int_0^x \frac{\lambda \gamma e^{-\lambda t}}{(1 - (1 - \gamma) e^{-\lambda t})^2} dt \\ &= \frac{1 - e^{-\lambda x}}{1 - (1 - \gamma) e^{-\lambda x}}; x > 0, \lambda > 0, \gamma > 0. \end{aligned} \quad (2.7)$$

Cette fonction de répartition est représentée dans la figure 2.1.2.

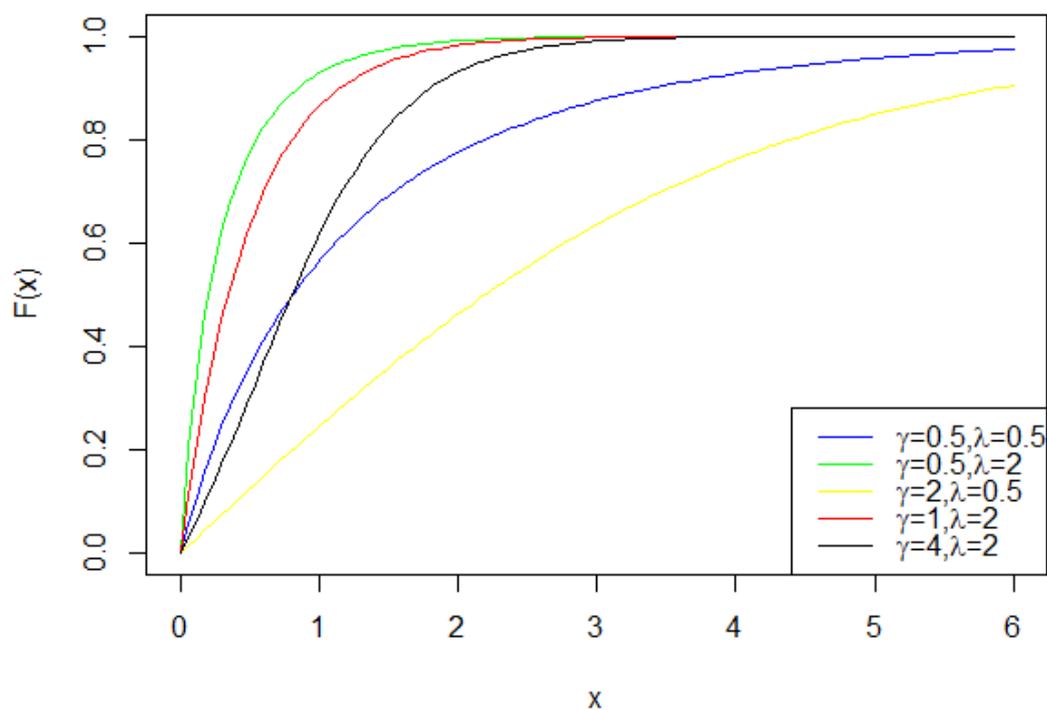


FIGURE 2.1.2 – La fonction de répartition de GEE

Remarque 2.1.1 *On observe que la fonction de répartition d'un GEE est similaire à la famille Gamma et à la famille Weibull correspondantes pour certains paramètres de forme spécifiés. Si le paramètre de forme est égal à 1, toutes les distributions coïncident avec la distribution exponentielle. Par conséquent, elles sont toutes des extensions ou des généralisations de la distribution exponentielle, mais de manière différente.*

2.1.3. La fonction de survie(également appelée fiabilité)

La fonction de fiabilité (ou de survie) est la fraction de la population qui dépasse une valeur x . En utilisant (2.7), la fonction de survie est donnée par la formule suivante :

$$\begin{aligned} S(x; \gamma, \lambda) &= 1 - F(x; \gamma, \lambda) \\ &= \frac{\gamma e^{-\lambda x}}{1 - (1 - \gamma) e^{-\lambda x}}; x > 0, \lambda > 0, \gamma > 0 \end{aligned} \quad (2.8)$$

La fonction de survie, $S(x; \gamma, \lambda)$, donne la probabilité qu'une personne survive plus longtemps qu'un certain temps x . Elle donne aussi la probabilité que la variable aléatoire X dépasse le temps x . La fonction de survie est fondamentale pour une analyse de survie.

2.1.4. La fonction de risque ou de hasard

Définition 2.1.1 *La fonction de hasard de la distribution géométrique exponentielle étendue joue un rôle crucial dans la compréhension du comportement des défaillances ou des événements dans le temps. Elle reflète la probabilité instantanée qu'un événement se produise s'il ne s'est pas déjà produit jusqu'à présent. La fonction de risque est donnée par la formule suivante.*

$$\begin{aligned} h(x; \gamma, \lambda) &= \frac{f(x; \gamma, \lambda)}{S(x; \gamma, \lambda)} \\ &= \frac{\lambda}{1 - (1 - \gamma) e^{-\lambda x}}; x > 0, \lambda > 0, \gamma > 0 \end{aligned} \quad (2.9)$$

La fonction de risque $h(x; \gamma, \lambda)$ donne le potentiel instantané par unité de temps pour que l'événement se produise, étant donné que l'individu a survécu jusqu'au temps x . Il s'agit de la probabilité de défaillance dans une période de temps infiniment petite. La fonction de hasard a sa propre importance, car elle donne un aperçu des taux d'échec conditionnels; elle peut être utilisée pour identifier une forme de modèle spécifique et c'est le véhicule par lequel la modélisation mathématique des données de survie est effectuée.

Le comportement de la fonction de risque dépend de la valeur du paramètre γ :

- $0 < \gamma < 1$: La fonction de risque est **monotone décroissante**, ce qui signifie que la probabilité d'un événement diminue avec le temps. Elle convient à la modélisation de situations présentant des périodes initiales d'usure suivies de taux de défaillance décroissants, comme les composants présentant des défauts initiaux.

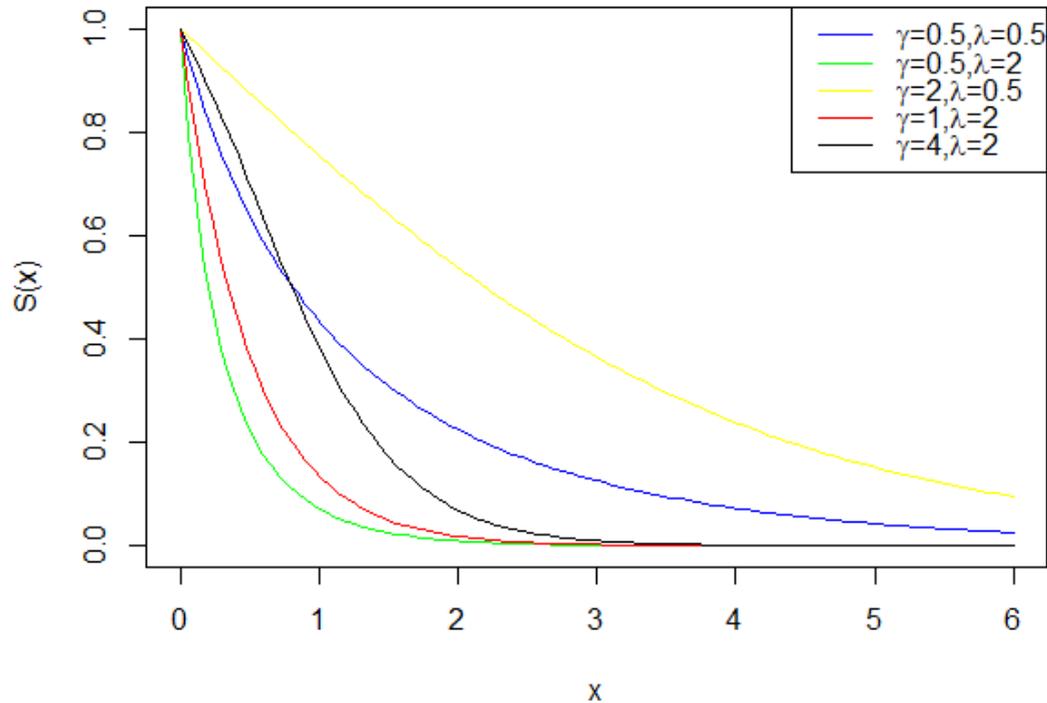


FIGURE 2.1.3 – La fonction de survie de GEE

- $\gamma > 1$: La fonction de hasard est **monotone croissante**, ce qui indique que la probabilité d'un événement augmente avec le temps. Cela s'applique aux systèmes dont les composants vieillissent et dont la probabilité de défaillance augmente avec le temps.
- $\gamma = 1$: La fonction de risque est **constante**, ce qui représente un taux de défaillance constant dans le temps. Ce scénario s'applique à des processus tels que la désintégration radioactive, où la probabilité d'un événement reste inchangée indépendamment du temps.

Elle se rapproche de $\frac{\lambda}{\gamma}$ lorsque $x \rightarrow 0$, et de λ lorsque $x \rightarrow \infty$. Cette fonction de hasard est représentée dans la figure 2.1.4.

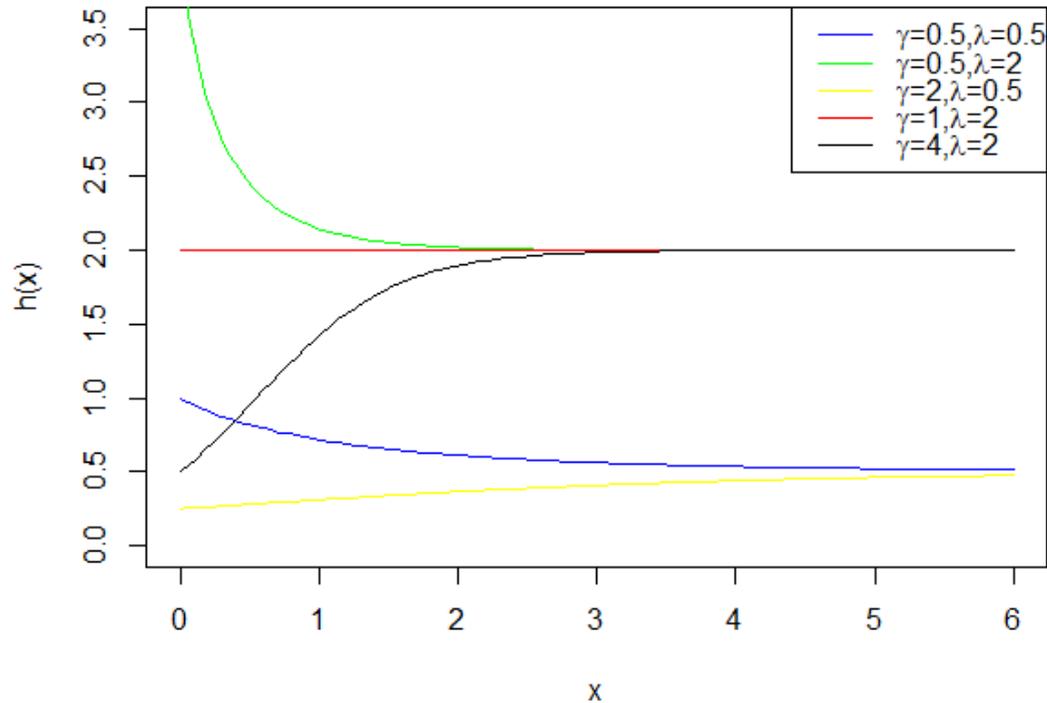


FIGURE 2.1.4 – Fonction de hasard de GEE.

2.1.5. Relation entre fonctions de répartition, de densité et de risque

Les fonctions de répartition, de densité et de hasard correspondantes sont données respectivement comme suit.

$$\begin{aligned}
 F(x; \gamma, \lambda) &= 1 - S(x; \gamma, \lambda) \\
 &= 1 - \frac{\gamma e^{-\lambda x}}{1 - (1 - \gamma) e^{-\lambda x}} \\
 &= \frac{1 - e^{-\lambda x} + \gamma e^{-\lambda x} - \gamma e^{-\lambda x}}{1 - (1 - \gamma) e^{-\lambda x}} \\
 &= \frac{1 - e^{-\lambda x}}{1 - (1 - \gamma) e^{-\lambda x}}; x > 0, \lambda > 0, \gamma > 0.
 \end{aligned} \tag{2.10}$$

$$\begin{aligned}
f(x; \gamma, \lambda) &= \frac{dF(x; \gamma, \lambda)}{dx} \\
&= \frac{(1 - (1 - \gamma)e^{-\lambda x})(\lambda e^{-\lambda x}) - (1 - e^{-\lambda x})(\lambda(1 - \gamma)e^{-\lambda x})}{(1 - (1 - \gamma)e^{-\lambda x})^2} \\
&= \frac{\lambda \gamma e^{-\lambda x}}{(1 - (1 - \gamma)e^{-\lambda x})^2}.
\end{aligned} \tag{2.11}$$

$$\begin{aligned}
h(x; \gamma, \lambda) &= \frac{f(x; \gamma, \lambda)}{S(x; \gamma, \lambda)} \\
&= \frac{\lambda \gamma e^{-\lambda x}}{(1 - (1 - \gamma)e^{-\lambda x})^2} \frac{1 - (1 - \gamma)e^{-\lambda x}}{\gamma e^{-\lambda x}} \\
&= \frac{\lambda}{1 - (1 - \gamma)e^{-\lambda x}}; x > 0, \lambda > 0, \gamma > 0.
\end{aligned} \tag{2.12}$$

2.2 Fonction génératrice des moments

Si $X \sim GEE(\gamma, \lambda)$ avec une FDP donnée par (2.6), la fonction génératrice de moments (voir [13]) peut être vérifiée comme étant

$$M_X(t) = 1 + \frac{t\gamma}{\lambda} \Phi\left(1 - \gamma, 1, 1 - \frac{t}{\lambda}\right); t < \lambda; \tag{2.13}$$

où

$$\begin{aligned}
\Phi(z, s, a) &= \frac{1}{\Gamma(s)} \int_0^x \frac{u^{s-1} e^{-au}}{1 - ze^{-u}} du \\
&= \sum_{k=0}^{\infty} \frac{z^k}{(a+k)^s}; \text{ pour } z < 1; a, s > 0;
\end{aligned} \tag{2.14}$$

la fonction spéciale $\Phi(z, s, a)$ est appelée fonction **transcendante de Lerch** (voir [9]). Le $k^{\text{ème}}$ moment de X peut être obtenu en évaluant la dérivée d'ordre k de $M_X(t)$ à $t = 0$ comme suit :

$$\begin{aligned}
m_k &= E(X^k) = M_X^{(k)}(0) \\
&= \frac{k! \gamma}{\lambda^k} \Phi(1 - \gamma, k, 1); k = 1, 2, 3, \dots
\end{aligned} \tag{2.15}$$

2.2.1. Espérance de la loi GEE

La moyenne de la distribution GEE est donnée par

$$\begin{aligned} E(X | \gamma, \lambda) &= \frac{\gamma}{\lambda} \Phi(1 - \gamma, 1, 1) \\ &= \frac{\gamma}{\lambda} \sum_{k=0}^{\infty} \frac{(1 - \gamma)^k}{(1 + k)} \end{aligned}$$

en posant $k = n - 1$, on a

$$\begin{aligned} E(X | \gamma, \lambda) &= \frac{\gamma}{\lambda} \sum_{n=1}^{\infty} \frac{(1 - \gamma)^{n-1}}{n} \\ &= \frac{\gamma}{\lambda} \sum_{n=1}^{\infty} \frac{(1 - \gamma)^{(n-1)}}{n} \frac{(1 - \gamma)}{(1 - \gamma)} \\ &= \frac{\gamma}{\lambda(1 - \gamma)} \sum_{n=1}^{\infty} \frac{(1 - \gamma)^n}{n} \\ &= -\frac{\gamma}{\lambda(1 - \gamma)} \ln(\gamma). \end{aligned} \tag{2.16}$$

2.2.2. Variance de la GEE

La variance de cette distribution géométrique exponentielle étendue est donnée par :

$$\begin{aligned} V(X | \gamma, \lambda) &= \frac{\gamma}{\lambda^2} \left(2\Phi(1 - \gamma, 2, 1) - \gamma\Phi^2(1 - \gamma, 1, 1) \right) \\ &= \frac{\gamma}{\lambda^2} \left(\frac{2L_2(1 - \gamma)}{(1 - \gamma)} - \frac{\gamma \log^2(\gamma)}{(1 - \gamma)^2} \right); \end{aligned} \tag{2.17}$$

où $L_2(z)$ est la fonction dilogarithme donnée par

$$\begin{aligned} L_2(z) &= \sum_{k=1}^{\infty} \frac{z^k}{k^2} = -\int_0^z \frac{\log(1 - t)}{t} dt \\ &= -\int_0^1 \frac{\log(1 - zt)}{t} dt. \end{aligned} \tag{2.18}$$

2.2.3. Mode de la GEE

Soit X une variable aléatoire absolument continue, de densité f continue sur \mathbb{R} . On appelle **Mode** de X tout réel x_0 où f atteint son maximum.

Donc pour trouver le maximum, on dérive $\log f(x; \gamma, \lambda)$ en x .

On a

$$\begin{aligned} \log f(x; \gamma, \lambda) &= \log \left(\frac{\lambda \gamma e^{-\lambda x}}{(1 - (1 - \gamma)e^{-\lambda x})^2} \right) \\ &= \log(\lambda \gamma e^{-\lambda x}) - 2 \log(1 - (1 - \gamma)e^{-\lambda x}). \end{aligned} \tag{2.19}$$

$$\begin{aligned}
\frac{\partial \log f(x; \gamma, \lambda)}{\partial x} &= \frac{-\lambda^2 \gamma e^{-\lambda x}}{\lambda \gamma e^{-\lambda x}} - \frac{2\lambda(1-\gamma)e^{-\lambda x}}{1 - (1-\gamma)e^{-\lambda x}} \\
&= -\lambda - \frac{2\lambda(1-\gamma)e^{-\lambda x}}{1 - (1-\gamma)e^{-\lambda x}} \\
&= \frac{-\lambda + \lambda(1-\gamma)e^{-\lambda x} - 2\lambda(1-\gamma)e^{-\lambda x}}{1 - (1-\gamma)e^{-\lambda x}} \\
&= \frac{-\lambda(1 + (1-\gamma)e^{-\lambda x})}{1 - (1-\gamma)e^{-\lambda x}}. \tag{2.20}
\end{aligned}$$

Donc la fonction $\log f(x; \gamma, \lambda)$ est convexe pour $0 < \gamma < 1$ et concave pour $\gamma \geq 1$.
Donc pour $\gamma \geq 1$ déterminons le réel x_0 .

Pour trouver le réel x_0 , on pose $\frac{\partial \log f(x; \gamma, \lambda)}{\partial x} = 0$. On sait que $\lambda > 0$ donc

$$\begin{aligned}
1 + (1-\gamma)e^{-\lambda x} &= 0 \\
\Leftrightarrow (1-\gamma)e^{-\lambda x} &= -1 \\
\Leftrightarrow e^{-\lambda x} &= \frac{1}{\gamma-1} \\
\Leftrightarrow -\lambda x &= \log\left(\frac{1}{\gamma-1}\right) \Leftrightarrow -\lambda x = -\log(\gamma-1) \\
x_0 &= \frac{\log(\gamma-1)}{\lambda}, \text{ pour } \gamma > 2
\end{aligned}$$

et

$$x_0 = 0, \text{ pour } \gamma \leq 2.$$

Donc pour $\gamma \geq 1$, $f(x; \gamma, \lambda)$ est uni-modale.

$$\text{D'où } Mode(x; \gamma, \lambda) = \begin{cases} 0 & \text{si } \gamma \leq 2 \\ \frac{\log(\gamma-1)}{\lambda} & \text{si } \gamma > 2 \end{cases}.$$

2.2.4. Médiane de la GEE

La **médiane** (Me) est le quantile d'ordre $\frac{1}{2}$. Or le quantile d'ordre y est donné par :

$$q_y = F^{-1}(y)$$

On a :

$$\begin{aligned}
F(x; \gamma, \lambda) &= y. \\
\frac{1 - e^{-\lambda x}}{1 - (1-\gamma)e^{-\lambda x}} &= y.
\end{aligned}$$

Si on démontre, on trouve que

$$q_y = \frac{-\log\left(\frac{1-y}{1-y(1-\gamma)}\right)}{\lambda}.$$

Donc on a :

$$Me(x; \gamma, \lambda) = q_{\frac{1}{2}}.$$

D'où on a :

$$Me(x; \gamma, \lambda) = \frac{\log(1+\gamma)}{\lambda}.$$

2.2.5. Relations entre Mode, Médiane et Espérance de la GEE

Il est facile de voir que $Me(x; \gamma, \lambda)$, $Mode(x; \gamma, \lambda)$ et $E(x; \gamma, \lambda)$ sont tous croissants dans γ et décroissants dans le paramètre d'échelle λ . De la monotonie de $\log x$ et du fait que $\log(x) \leq x - 1 (x > 0)$, on a :

$$\begin{aligned} \log(\gamma - 1) &\leq \log(\gamma + 1) \leq \gamma \\ \implies \frac{\log(\gamma - 1)}{\lambda} &\leq \frac{\log(\gamma + 1)}{\lambda} \leq \frac{\gamma}{\lambda} \end{aligned}$$

Donc on a

$$Mode(x; \gamma, \lambda) \leq Me(x; \gamma, \lambda) \leq \frac{\gamma}{\lambda}.$$

On sait aussi que

$$\begin{aligned} \frac{\gamma}{\lambda} &\leq \frac{\gamma \log(\gamma)}{\lambda(\gamma - 1)} \\ \implies \frac{\gamma}{\lambda} &\leq E(x; \gamma, \lambda). \end{aligned}$$

D'où on a

$$Mode(x; \gamma, \lambda) \leq Me(x; \gamma, \lambda) \leq \frac{\gamma}{\lambda} \leq E(x; \gamma, \lambda).$$

Mais notons aussi que $\lim_{\gamma \rightarrow +\infty} \frac{Mode(x; \gamma, \lambda)}{E(x; \gamma, \lambda)} = 1$. Pour plus de détails voir ([23]).

2.2.6. L'entropie de Shannon

L'entropie a été utilisée dans diverses situations en science et en ingénierie. L'entropie d'une variable aléatoire X est une mesure de la variation de l'incertitude. L'entropie de

Shannon, voir [25] est définie comme $E \{-\log [f(x; \gamma, \lambda)]\}$. Notons $H_s(x; \gamma, \lambda)$ l'entropie de Shannon et est donnée par la formule suivante :

$$\begin{aligned}
H_s(x; \gamma, \lambda) &= E \{-\log [f(x; \gamma, \lambda)]\} \\
&= E \left[-\log \left(\frac{\lambda \gamma e^{-\lambda x}}{(1 - (1 - \gamma)e^{-\lambda x})^2} \right) \right] \\
&= E \left[-\log (\lambda \gamma e^{-\lambda x}) + \log (1 - (1 - \gamma)e^{-\lambda x})^2 \right] \\
&= -\log(\lambda \gamma) + \lambda E[X] + 2E \left[\log (1 - (1 - \gamma)e^{-\lambda x}) \right] \\
&= \log(\gamma \lambda) + \frac{2 \log(\gamma) - \gamma^2 \log(\gamma) \lambda^{-1} - 2\gamma + 2}{\gamma - 1}. \tag{2.21}
\end{aligned}$$

Pour plus de détails voir [24] et [25].

En guise de conclusion, nous avons dans ce chapitre donné quelques fonctions qui caractérisent la distribution géométrique exponentielle étendue. Ces fonctions, nous permettrons d'étudier les différentes méthodes d'estimations.

Méthodes d'estimation

Un estimateur est une statistique particulière, en tant que fonction de l'échantillon, qui permet d'attribuer une valeur au paramètre θ à estimer. Dans ce chapitre nous allons donner quelques méthodes d'estimation des paramètres de la distribution géométrique exponentielle étendue (GEE). Il existe plusieurs méthodes d'estimation. Pour ce travail, nous nous limiterons à sept (7) méthodes d'estimation pour obtenir les estimations des paramètres γ et λ de la distribution de la GEE.

3.1 Méthode du Maximum de Vraisemblance

Un estimateur est une fonction de l'échantillon, tandis qu'une estimation est la valeur réalisée d'un estimateur (c'est-à-dire un nombre) qui est obtenue lorsqu'un échantillon est effectivement prélevé. Le principe de l'estimateur du maximum de vraisemblance (MV), développé à l'origine par Fisher dans les années 1920 (voir [26]), stipule que la distribution de probabilité souhaitée est celle qui rend les données observées "les plus probables". Dans le cadre de la méthode du maximum de vraisemblance, les estimateurs sont obtenus en maximisant la fonction de vraisemblance voir [7]. Soit X_1, X_2, \dots, X_n un échantillon aléatoire de taille n de la distribution géométrique exponentielle étendue (GEE) de densité donnée par la relation (2.7). La fonction de vraisemblance pour les paramètres γ et λ notée $L(x; \gamma, \lambda)$ est donnée par

$$\begin{aligned}
 L(x; \gamma, \lambda) &= \prod_{i=1}^n f(x_i; \gamma, \lambda) \\
 &= \prod_{i=1}^n \left(\frac{\gamma \lambda e^{-\lambda x_i}}{(1 - (1 - \gamma)e^{-\lambda x_i})^2} \right) \\
 &= \frac{(\gamma \lambda)^n \exp(-\lambda \sum_{i=1}^n x_i)}{\prod_{i=1}^n (1 - (1 - \gamma)e^{-\lambda x_i})^2}.
 \end{aligned} \tag{3.1}$$

La fonction de la log-vraisemblance notée $l(x; \gamma, \lambda)$ est donnée par

$$\begin{aligned}
l(x; \gamma, \lambda) &= \log L(x; \gamma, \lambda) \\
&= \log \left((\gamma\lambda)^n e^{-\lambda \sum_{i=1}^n x_i} \right) - \log \left(\prod_{i=1}^n (1 - (1 - \gamma)e^{-\lambda x_i})^2 \right) \\
&= n \log(\gamma) + n \log(\lambda) - \lambda \sum_{i=1}^n x_i - 2 \sum_{i=1}^n \log(1 - (1 - \gamma)e^{-\lambda x_i}). \quad (3.2)
\end{aligned}$$

Pour trouver l'estimateur du maximum de vraisemblance (MV) $\hat{\lambda}_{MV}$ et $\hat{\gamma}_{MV}$, nous allons déterminer les dérivées partielles de la log-vraisemblance par rapport à γ et λ et de l'égaliser à 0, nous obtenons le résultat suivant :

$$\frac{\partial l(x; \gamma, \lambda)}{\partial \gamma} = \frac{n}{\gamma} - 2 \sum_{i=1}^n \frac{e^{-\lambda x_i}}{1 - (1 - \gamma)e^{-\lambda x_i}} = 0. \quad (3.3)$$

$$\frac{\partial l(x; \gamma, \lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^n x_i - 2(1 - \gamma) \sum_{i=1}^n \frac{x_i e^{-\lambda x_i}}{1 - (1 - \gamma)e^{-\lambda x_i}} = 0. \quad (3.4)$$

La solution des deux équations non linéaires peut être obtenue en utilisant des méthodes numériques. Dans le cadre des méthodes numériques statistiquement orientées avec une convergence stable garantie. La solution de (3.3) et (3.4) est obtenue à travers le théorème suivant.

Théorème 3.1.1 Soit $g(x_1, \dots, x_n; \gamma, \lambda)$ la fonction de l'équation (3.4) et $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. Alors, pour $\gamma \in (0, 1)$ la racine de $g(x_1, \dots, x_n; \gamma, \lambda) = 0$ se trouve dans l'intervalle $\left[\frac{\gamma}{\bar{x}(2 - \gamma)}, \frac{1}{\bar{x}} \right]$, quand $\gamma \in (1, 2)$ la racine de $g(x_1, \dots, x_n; \gamma, \lambda) = 0$ se trouve dans l'intervalle $\left[\frac{1}{\bar{x}}, \frac{\gamma}{\bar{x}(2 - \gamma)} \right]$, et quand $\gamma \in (2, \infty)$ la racine de $g(x_1, \dots, x_n; \gamma, \lambda) = 0$ se trouve dans l'intervalle $\left[\frac{1}{\bar{x}}, \infty \right]$. Les racines sont uniques .

Pour la preuve du théorème 3.1.1 (dans [1] et [2]).

L'algorithme de l'estimateur maximal (EM) (voir[8]) a été utilisé (voir [1]) pour résoudre simultanément les deux équations non linéaires ((3.3) et (3.4)). Le résultat est

$$\hat{\gamma}_{MV} = \hat{\gamma}^{(j+1)} = n \left[\sum_{i=1}^n \frac{1 + (1 - \hat{\gamma}^{(j)}) \exp(-\hat{\lambda}^{(j)} x_i)}{1 - (1 - \hat{\gamma}^{(j)}) \exp(-\hat{\lambda}^{(j)} x_i)} \right]^{-1}; j = 1, 2 \dots \quad (3.5)$$

$$\hat{\lambda}_{MV} = \hat{\lambda}^{(j+1)} = n \left[\sum_{i=1}^n \frac{x_i (1 + (1 - \hat{\gamma}^{(j)}) \exp(-\hat{\lambda}^{(j)} x_i))}{1 - (1 - \hat{\gamma}^{(j)}) \exp(-\hat{\lambda}^{(j)} x_i)} \right]^{-1}; j = 1, 2 \dots \quad (3.6)$$

Si les estimateurs du maximum de vraisemblance (MV) $\hat{\lambda}_{MV}$ et $\hat{\gamma}_{MV}$ de (3.5) et de (3.6) existent, alors la matrice d'information observée $I(\hat{\theta}) = I(\lambda, \gamma) = \begin{bmatrix} i_{11} & i_{12} \\ i_{21} & i_{22} \end{bmatrix}$ peut être obtenue en différenciant (3.2) par rapport aux paramètres (γ, λ) . Les éléments de la matrice d'information observée du second ordre sont les suivants :

$$\begin{aligned}
i_{11} &= -\frac{\partial^2 l(x; \gamma, \lambda)}{\partial \gamma^2} \\
&= \frac{n}{\gamma^2} - 2 \sum_{i=1}^n \frac{e^{-2\lambda x_i}}{(1 - (1 - \gamma)e^{-\lambda x_i})^2} \\
i_{22} &= -\frac{\partial^2 l(x; \gamma, \lambda)}{\partial \lambda^2} \\
&= \frac{n}{\lambda^2} - 2(1 - \gamma) \sum_{i=1}^n \frac{x_i^2 e^{-\lambda x_i}}{(1 - (1 - \gamma)e^{-\lambda x_i})^2} \\
i_{12} = i_{21} &= -\frac{\partial^2 l(x; \gamma, \lambda)}{\partial \gamma \partial \lambda} = -\frac{\partial^2 l(x; \gamma, \lambda)}{\partial \lambda \partial \gamma} \\
&= -2 \sum_{i=1}^n \frac{x_i e^{-\lambda x_i}}{(1 - (1 - \gamma)e^{-\lambda x_i})^2}
\end{aligned} \tag{3.7}$$

La matrice (ou quantité) d'information de FISHER (MIF) (voir [13]) est donnée par

$$I_{MIF}(\gamma, \lambda) = \begin{bmatrix} E(i_{11}) & E(i_{12}) \\ E(i_{21}) & E(i_{22}) \end{bmatrix} = \begin{bmatrix} \frac{n}{3\gamma^2} & \frac{n(\gamma - 1 - \gamma^2 \log(\gamma))}{3\lambda\gamma(\gamma - 1)^2} \\ \frac{n(\gamma - 1 - \gamma^2 \log(\gamma))}{3\lambda\gamma(\gamma - 1)^2} & E(i_{22}) \end{bmatrix} \tag{3.8}$$

avec

$$E(i_{22}) = \begin{cases} \frac{n[3(1 - \gamma) - 2((1 - \gamma) - \gamma L_2(1 - \gamma))]}{3\lambda^3(1 - \gamma)} & \text{si } 0 < \gamma < 1, \\ \frac{n \left[1 - \gamma \left(1 + \frac{\pi^2}{3} + \ln^2 \gamma - 2L_2\left(\frac{1}{\gamma}\right) \right) \right]}{3\lambda^3(1 - \gamma)} & \text{si } \gamma > 1 \end{cases}$$

où $L_r(p) = \sum_{i=1}^{\infty} \frac{p^i}{i^r}$, ($0 < p < 1$).

Les estimations du maximum de vraisemblance pour γ et λ sont biaisées pour les problèmes de petits échantillons. Dans le cas de grands échantillons, elles sont sans biais et asymptotiquement efficaces. Ces estimations sont asymptotiquement distribuées selon la loi normale avec une distribution conjointe donnée par

$$(\hat{\gamma}_{MV}, \hat{\lambda}_{MV}) \sim N_2 \left[(\gamma, \lambda), I_{MIF}^{-1}(\gamma, \lambda) \right].$$

3.2 Méthode des percentiles

La méthode des percentiles a été proposée à l'origine par Kao (voir [12] et [11]). Cette méthode est couramment utilisée pour estimer les paramètres inconnus à partir des fonctions de distribution qui ont une forme fermée de la fonction quantile. Les estimateurs des percentiles (PC) peuvent être obtenus en minimisant. Par rapport aux paramètres inconnus, la distance euclidienne entre les points de l'échantillon ordonné et les points théoriques ordonnés, calculés tout au long de la fonction quantile. On a la fonction de répartition de la GEE donnée par

$$F(x; \gamma, \lambda) = \frac{1 - e^{-\lambda x}}{1 - (1 - \gamma) e^{-\lambda x}}.$$

La fonction quantile est donc donnée par

$$Q(p; \gamma, \lambda) = \frac{1}{\lambda} \log \left(\frac{1 - (1 - \gamma)p}{1 - p} \right). \quad (3.9)$$

Soit $X_{(i)}$ les statistiques du $i^{\text{ème}}$ ordre, alors que $X_{(1)} < X_{(2)} < \dots < X_{(n)}$. Si p_i désigne certains estimateurs de $F(x_i; \gamma, \lambda)$, alors les estimateurs $\hat{\gamma}_{PC}$ et $\hat{\lambda}_{PC}$ de γ et λ peuvent être obtenus en minimisant la fonction notée PC suivante :

$$\begin{aligned} PC(x; \gamma, \lambda) &= \sum_{i=1}^n (x_i - Q(p_i; \gamma, \lambda))^2 \\ &= \sum_{i=1}^n \left(x_i - \frac{1}{\lambda} \log \left(\frac{1 - (1 - \gamma)p_i}{1 - p_i} \right) \right)^2. \end{aligned} \quad (3.10)$$

par rapport à γ et λ .

La dérivée première de PC par rapport à γ est :

$$\begin{aligned} \frac{\partial PC}{\partial \gamma} &= \frac{\partial}{\partial \gamma} \left(\sum_{i=1}^n \left(x_i - \frac{1}{\lambda} \log \left(\frac{1 - (1 - \gamma)p_i}{1 - p_i} \right) \right)^2 \right) \\ &= 2 \sum_{i=1}^n \left[x_i - \frac{1}{\lambda} \log \left(\frac{1 - (1 - \gamma)p_i}{1 - p_i} \right) \right] \left(\frac{-p_i}{\lambda(1 - (1 - \gamma)p_i)} \right). \end{aligned} \quad (3.11)$$

La dérivée première de PC par rapport à λ est donnée par :

$$\begin{aligned} \frac{\partial PC}{\partial \lambda} &= \frac{\partial}{\partial \lambda} \left(\sum_{i=1}^n \left(x_i - \frac{1}{\lambda} \log \left(\frac{1 - (1 - \gamma)p_i}{1 - p_i} \right) \right)^2 \right) \\ &= 2 \sum_{i=1}^n \left[x_i - \frac{1}{\lambda} \log \left(\frac{1 - (1 - \gamma)p_i}{1 - p_i} \right) \right] \left(\frac{1}{\lambda^2} \log \left(\frac{1 - (1 - \gamma)p_i}{1 - p_i} \right) \right). \end{aligned} \quad (3.12)$$

En posant (3.11) et (3.12) égaux à 0, les estimateurs de percentile $\hat{\gamma}_{PC}$ et $\hat{\lambda}_{PC}$ peuvent être obtenus en résolvant les équations non linéaires suivantes :

$$\sum_{i=1}^n \left[x_i - \frac{1}{\lambda} \log \left(\frac{1 - (1 - \gamma)p_i}{1 - p_i} \right) \right] \left(\frac{p_i}{1 - (1 - \gamma)p_i} \right) = 0 \quad (3.13)$$

$$\sum_{i=1}^n \left[x_i - \frac{1}{\lambda} \log \left(\frac{1 - (1 - \gamma)p_i}{1 - p_i} \right) \right] \left(\frac{1}{\lambda^2} \log \left(\frac{1 - (1 - \gamma)p_i}{1 - p_i} \right) \right) = 0. \quad (3.14)$$

Dans ce document, nous considérons l'estimateur de p_i comme $p_i = \frac{i}{n+1}$. Cependant, d'autres estimateurs peuvent être utilisés, voir par exemple ([31]).

3.3 Méthode des moindres carrés ordinaires

Dans cette section, nous fournissons des estimateurs des paramètres inconnus basés sur la méthode des moindres carrés, proposée à l'origine par Swain, Venkatraman et Wilson (voir dans [34]) pour estimer les paramètres des distributions bêta. Cette méthode peut également être utilisée dans d'autres cas. Soit X_1, X_2, \dots, X_n est un échantillon aléatoire de taille n à partir d'une fonction de distribution $F(X_i)$ et X_i désigne l'échantillon ordonné. La méthode proposée utilise la distribution de $F(X_i)$. Pour un échantillon de taille n , nous avons

$$E(F(X_i)) = \frac{i}{n+1}$$

$$V(F(X_i)) = \frac{i(n-i+1)}{(n+1)^2(n+2)}.$$

Les estimateurs des moindres carrés ordinaires(MCO) peuvent être obtenus en minimisant la fonction

$$LS = \sum_{i=1}^n \left(F(x_i) - \frac{i}{n+1} \right)^2. \quad (3.15)$$

Par conséquent, dans le cas de la distribution GEE, les estimateurs des moindres carrés ordinaires $\hat{\gamma}_{MCO}$ et $\hat{\lambda}_{MCO}$ des paramètres γ et λ s'obtiennent en minimisant la fonction notée LS suivante :

$$LS(\gamma, \lambda) = \sum_{i=1}^n \left(F(x_i; \gamma, \lambda) - \frac{i}{n+1} \right)^2$$

$$= \sum_{i=1}^n \left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma)e^{-\lambda x_i}} - \frac{i}{n+1} \right)^2. \quad (3.16)$$

La dérivée première de $LS(\gamma, \lambda)$ par rapport à γ est :

$$\frac{\partial LS(\gamma, \lambda)}{\partial \gamma} = \frac{\partial}{\partial \gamma} \left(\sum_{i=1}^n \left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma)e^{-\lambda x_i}} - \frac{i}{n+1} \right)^2 \right),$$

$$= 2 \sum_{i=1}^n \left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma)e^{-\lambda x_i}} - \frac{i}{n+1} \right) \left(\frac{1 - e^{\lambda x_i}}{(e^{\lambda x_i} - 1 + \gamma)^2} \right). \quad (3.17)$$

La dérivée première de $LS(\gamma, \lambda)$ par rapport à λ est donnée par :

$$\begin{aligned} \frac{\partial LS(\gamma, \lambda)}{\partial \lambda} &= \frac{\partial}{\partial \lambda} \left(\sum_{i=1}^n \left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma) e^{-\lambda x_i}} - \frac{i}{n+1} \right)^2 \right), \\ &= 2 \sum_{i=1}^n \left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma) e^{-\lambda x_i}} - \frac{i}{n+1} \right) \left(\frac{\gamma x_i e^{\lambda x_i}}{(e^{\lambda x_i} - 1 + \gamma)^2} \right). \end{aligned} \quad (3.18)$$

En égalisant (3.17) et (3.18) à zéro, on obtient les équations non linéaires suivantes :

$$\sum_{i=1}^n \left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma) e^{-\lambda x_i}} - \frac{i}{n+1} \right) \Delta_1(x_i; \gamma, \lambda) = 0, \quad (3.19)$$

$$\sum_{i=1}^n \left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma) e^{-\lambda x_i}} - \frac{i}{n+1} \right) \Delta_2(x_i; \gamma, \lambda) = 0, \quad (3.20)$$

avec

$$\Delta_1(x_i; \gamma, \lambda) = \frac{e^{\lambda x_i} - 1}{(e^{\lambda x_i} - 1 + \gamma)^2}. \quad (3.21)$$

et

$$\Delta_2(x_i; \gamma, \lambda) = \frac{\gamma x_i e^{\lambda x_i}}{(e^{\lambda x_i} - 1 + \gamma)^2}. \quad (3.22)$$

Ces estimations peuvent également être obtenues en résolvant les équations (3.19) et (3.20).

3.4 Méthode des moindres carrés pondérés

La méthode des moindres carrés pondérés est obtenue en multipliant la fonction LS des moindres carrés ordinaires par l'inverse de la variance de $F(X_i)$. Les estimateurs des moindres carrés pondérés peuvent être obtenus en minimisant l'équation notée WLS suivante :

$$\begin{aligned} WLS &= \sum_{i=1}^n \frac{1}{V(F(X_i))} \left(F(x_i) - \frac{i}{n+1} \right)^2 \\ &= \sum_{i=1}^n \frac{(n+1)^2(n+2)}{i(n-i+1)} \left(F(x_i) - \frac{i}{n+1} \right)^2. \end{aligned} \quad (3.23)$$

Les estimateurs des moindres carrés pondérés(MCP) $\hat{\gamma}_{MCP}$ et $\hat{\lambda}_{MCP}$ des paramètres γ et λ de la GEE s'obtiennent en minimisant la fonction suivante :

$$\begin{aligned} WLS(\gamma, \lambda) &= \sum_{i=1}^n \frac{1}{V(F(X_i))} \left(F(x_i; \gamma, \lambda) - \frac{i}{n+1} \right)^2 \\ &= \sum_{i=1}^n \frac{(n+1)^2(n+2)}{i(n-i+1)} \left(\frac{1 - e^{-\lambda x_i}}{1 - (1-\gamma)e^{-\lambda x_i}} - \frac{i}{n+1} \right)^2. \end{aligned} \quad (3.24)$$

La dérivée première de $WLS(\gamma, \lambda)$ par rapport à γ est donnée par :

$$\begin{aligned} \frac{\partial WLS(\gamma, \lambda)}{\partial \gamma} &= \frac{\partial}{\partial \gamma} \left(\sum_{i=1}^n \frac{(n+1)^2(n+2)}{i(n-i+1)} \left(\frac{1 - e^{-\lambda x_i}}{1 - (1-\gamma)e^{-\lambda x_i}} - \frac{i}{n+1} \right)^2 \right), \\ &= 2 \sum_{i=1}^n \frac{(n+1)^2(n+2)}{i(n-i+1)} \left(\frac{1 - e^{-\lambda x_i}}{1 - (1-\gamma)e^{-\lambda x_i}} - \frac{i}{n+1} \right) \left(\frac{1 - e^{\lambda x_i}}{(e^{\lambda x_i} - 1 + \gamma)^2} \right). \end{aligned} \quad (3.25)$$

La dérivée première de $LS(\gamma, \lambda)$ par rapport à λ est donnée par :

$$\begin{aligned} \frac{\partial WLS(\gamma, \lambda)}{\partial \lambda} &= \frac{\partial}{\partial \lambda} \left(\sum_{i=1}^n \frac{(n+1)^2(n+2)}{i(n-i+1)} \left(\frac{1 - e^{-\lambda x_i}}{1 - (1-\gamma)e^{-\lambda x_i}} - \frac{i}{n+1} \right)^2 \right), \\ &= 2 \sum_{i=1}^n \frac{(n+1)^2(n+2)}{i(n-i+1)} \left(\frac{1 - e^{-\lambda x_i}}{1 - (1-\gamma)e^{-\lambda x_i}} - \frac{i}{n+1} \right) \left(\frac{\gamma x_i e^{\lambda x_i}}{(e^{\lambda x_i} - 1 + \gamma)^2} \right). \end{aligned} \quad (3.26)$$

En égalisant (3.25) et (3.26) à 0, on obtient les équations non linéaires suivantes :

$$\sum_{i=1}^n \frac{(n+1)^2(n+2)}{i(n-i+1)} \left(\frac{1 - e^{-\lambda x_i}}{1 - (1-\gamma)e^{-\lambda x_i}} - \frac{i}{n+1} \right) \Delta_1(x_i; \gamma, \lambda) = 0, \quad (3.27)$$

$$\sum_{i=1}^n \frac{(n+1)^2(n+2)}{i(n-i+1)} \left(\frac{1 - e^{-\lambda x_i}}{1 - (1-\gamma)e^{-\lambda x_i}} - \frac{i}{n+1} \right) \Delta_2(x_i; \gamma, \lambda) = 0, \quad (3.28)$$

où $\Delta_1(x_i; \gamma, \lambda)$ et $\Delta_2(x_i; \gamma, \lambda)$ sont donnés par (3.21) et (3.22). Ces estimateurs peuvent aussi être obtenus en résolvant les équations non linéaires (3.27) et (3.28).

3.5 Méthode du produit maximal des espaces

Cheng et Amin, dans [3] et [4], ont introduit la méthode du produit maximal d'espacements (MPS) comme alternative à la méthode MLE pour l'estimation des paramètres

de distributions continues uni-variées. Ranneby, dans [28], a développé indépendamment la même méthode comme approximation de la mesure d'information de Kullback-Leibler. Les espacements uniformes d'un échantillon aléatoire de la distribution géométrique exponentielle étendue(GEE) sont définis comme suit :

$$D_i(\gamma, \lambda) = F(x_i; \gamma, \lambda) - F(x_{i-1}; \gamma, \lambda), i = 1, 2, \dots, n + 1 \quad (3.29)$$

où $F(x_0; \gamma, \lambda) = 0$, $F(x_{n+1}; \gamma, \lambda) = 1$ et $\sum_{i=1}^{n+1} D_i(\gamma, \lambda) = 1$.

Les estimateurs du produit maximal d'espacements(PME) $\hat{\gamma}_{PME}$ et $\hat{\lambda}_{PME}$ des paramètres γ et λ peuvent être obtenus en maximisant la moyenne géométrique des espacements notés $G(\gamma, \lambda)$. La fonction $G(\gamma, \lambda)$ est donnée par la formule suivante :

$$G(\gamma, \lambda) = \left[\prod_{i=1}^{n+1} D_i(\gamma, \lambda) \right]^{\frac{1}{n+1}} \quad (3.30)$$

ou, de manière équivalente, en maximisant la fonction notée $H(\gamma, \lambda)$ suivante :

$$\begin{aligned} H(\gamma, \lambda) &= \ln G(\gamma, \lambda) \\ &= \frac{1}{n+1} \sum_{i=1}^{n+1} \ln D_i(\gamma, \lambda) \\ &= \frac{1}{n+1} \sum_{i=1}^{n+1} \ln (F(x_i; \gamma, \lambda) - F(x_{i-1}; \gamma, \lambda)) \\ &= \frac{1}{n+1} \left(\ln(F(x_1; \gamma, \lambda)) + \sum_{i=2}^n \ln (F(x_i; \gamma, \lambda) - F(x_{i-1}; \gamma, \lambda)) + \ln(1 - F(x_n)) \right) \end{aligned} \quad (3.31)$$

La dérivée première de $H(\gamma, \lambda)$ par rapport à γ est :

$$\begin{aligned} \frac{\partial}{\partial \gamma} H(\gamma, \lambda) &= \frac{\partial}{\partial \gamma} \left(\frac{1}{n+1} \sum_{i=1}^{n+1} \ln D_i(\gamma, \lambda) \right), \\ &= \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{\partial}{\partial \gamma} \ln D_i(\gamma, \lambda), \\ &= \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{1}{D_i(\gamma, \lambda)} (\Delta_1(x_i; \gamma, \lambda) - \Delta_1(x_{i-1}; \gamma, \lambda)). \end{aligned} \quad (3.32)$$

La dérivée première de $H(\gamma, \lambda)$ par rapport à λ est donnée comme suit :

$$\begin{aligned} \frac{\partial}{\partial \lambda} H(\gamma, \lambda) &= \frac{\partial}{\partial \lambda} \left(\frac{1}{n+1} \sum_{i=1}^{n+1} \ln D_i(\gamma, \lambda) \right), \\ &= \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{\partial}{\partial \lambda} \ln D_i(\gamma, \lambda), \\ &= \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{1}{D_i(\gamma, \lambda)} (\Delta_2(x_i; \gamma, \lambda) - \Delta_2(x_{i-1}; \gamma, \lambda)). \end{aligned} \quad (3.33)$$

avec

$$D_i(\gamma, \lambda) = \left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma) e^{-\lambda x_i}} - \frac{1 - e^{-\lambda x_{i-1}}}{1 - (1 - \gamma) e^{-\lambda x_{i-1}}} \right) \quad (3.34)$$

où $\Delta_1(x_i; \gamma, \lambda)$ et $\Delta_2(x_i; \gamma, \lambda)$ sont donnés par (3.21) et (3.22).

En égalisant (3.32) et (3.33) à zéro, les estimateurs $\hat{\gamma}_{PME}$ et $\hat{\lambda}_{PME}$ des paramètres γ et λ peuvent également être obtenus en résolvant les équations non linéaires (3.32) et (3.33).

Cheng et Amin, dans [3] ont montré que la maximisation de H comme méthode d'estimation des paramètres est aussi efficace que l'estimation du maximum de vraisemblance (MV) et que les estimateurs PME sont cohérents dans des conditions plus générales que les estimateurs MV. Par conséquent, pour la distribution EEG, les estimateurs PME sont asymptotiquement normalement distribués (voir [32] pour plus de détails) avec une distribution normale bivariée commune donnée par

$$(\hat{\gamma}_{PME}, \hat{\lambda}_{PME}) \sim N_2 \left[(\gamma, \lambda), I_{MIF}^{-1}(\gamma, \lambda) \right]$$

où I_{MIF} est donnée par (3.8).

3.6 Méthode de distance minimale de Cramer-Von Mises

L'estimateur de Cramer-von Mises est l'un des estimateurs de qualité d'ajustement. Cette méthode est basée sur la différence entre l'estimation de la fonction de distribution cumulative et la fonction de distribution empirique (voir [29],[33],[21]). Pour motiver notre choix des estimateurs de distance minimale de type Cramer-von-Mise, MacDonald, dans [22], a fourni des preuves empiriques que le biais de l'estimateur est plus faible que les autres estimateurs de distance minimale. Ainsi, dans notre cas, les estimateurs de Cramer -Von-Mise (CVM) $\hat{\gamma}_{CVM}$ et $\hat{\lambda}_{CVM}$ des paramètres γ et λ de la GEE sont obtenus en minimisant la fonction notée $C(\gamma, \lambda)$ suivante :

$$C(\gamma, \lambda) = n \int_{-\infty}^{+\infty} (F(x; \gamma, \lambda) - S_n(x))^2 dF(x),$$

En considérant que $S_n(x)$ est une fonction en escalier avec des sauts au niveau des statistiques d'ordre (voir [22]), on peut écrire $C(\gamma, \lambda)$ comme suit :

$$\begin{aligned} C(\gamma, \lambda) &= n \left(\frac{1}{12n^2} + \frac{1}{n} \sum_{i=1}^n \left(F(x_i; \gamma, \lambda) - \frac{(i - \frac{1}{2})}{n} \right)^2 \right) \\ &= \frac{1}{12n} + \sum_{i=1}^n \left(F(x_i; \gamma, \lambda) - \frac{2i - 1}{2n} \right)^2 \\ &= \frac{1}{12n} + \sum_{i=1}^n \left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma) e^{-\lambda x_i}} - \frac{2i - 1}{2n} \right)^2. \end{aligned} \quad (3.35)$$

La dérivée première de $C(\gamma, \lambda)$ par rapport à γ est :

$$\begin{aligned} \frac{\partial C(\gamma, \lambda)}{\partial \gamma} &= \frac{\partial}{\partial \gamma} \left(\sum_{i=1}^n \left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma) e^{-\lambda x_i}} - \frac{2i - 1}{2n} \right)^2 \right), \\ &= 2 \sum_{i=1}^n \left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma) e^{-\lambda x_i}} - \frac{2i - 1}{2n} \right) \left(\frac{1 - e^{\lambda x_i}}{(e^{\lambda x_i} - 1 + \gamma)^2} \right). \end{aligned} \quad (3.36)$$

La dérivée première de $C(\gamma, \lambda)$ par rapport à λ est donnée par :

$$\begin{aligned} \frac{\partial C(\gamma, \lambda)}{\partial \lambda} &= \frac{\partial}{\partial \lambda} \left(\sum_{i=1}^n \left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma) e^{-\lambda x_i}} - \frac{2i - 1}{2n} \right)^2 \right), \\ &= 2 \sum_{i=1}^n \left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma) e^{-\lambda x_i}} - \frac{2i - 1}{2n} \right) \left(\frac{\gamma x_i e^{\lambda x_i}}{(e^{\lambda x_i} - 1 + \gamma)^2} \right). \end{aligned} \quad (3.37)$$

En égalant (3.36) et (3.37) à 0, on obtient les équations suivantes :

$$\sum_{i=1}^n \left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma) e^{-\lambda x_i}} - \frac{2i - 1}{2n} \right) \Delta_1(x_i; \gamma, \lambda) = 0, \quad (3.38)$$

$$\sum_{i=1}^n \left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma) e^{-\lambda x_i}} - \frac{2i - 1}{2n} \right) \Delta_2(x_i; \gamma, \lambda) = 0. \quad (3.39)$$

où $\Delta_1(x_i; \gamma, \lambda)$ et $\Delta_2(x_i; \gamma, \lambda)$ sont donnés par (3.21) et (3.22).

Les estimateurs $\hat{\gamma}_{CVM}$ et $\hat{\lambda}_{CVM}$ peuvent également être obtenus en résolvant les équations non linéaires (3.38) et (3.39).

3.7 Méthodes d'Anderson-Darling

L'estimateur d'Anderson-Darling (AD) est un autre type d'estimateur à distance minimale et est basé sur une statistique d'Anderson-Darling. L'utilisation de cette méthode

a été motivée par Luceño, dans [21]. Les estimateurs d'Anderson-Darling $\hat{\gamma}_{AD}$ et $\hat{\lambda}_{AD}$ des paramètres γ et λ de la distribution géométrique exponentielle étendue sont obtenus en minimisant la fonction notée $A(\gamma, \lambda)$ suivante :

$$A(\gamma, \lambda) = n \int_{-\infty}^{\infty} \frac{(F(x; \gamma, \lambda) - S_n(x))^2}{F(x; \gamma, \lambda) (1 - F(x; \gamma, \lambda))} dF(x).$$

En considérant que $S_n(x)$ est une fonction en escalier avec des sauts au niveau des statistiques d'ordre (voir [21]), on peut réécrire $A(\gamma, \lambda)$ comme suit :

$$\begin{aligned} A(\gamma, \lambda) &= n \left(-1 - \frac{1}{n} \sum_{i=1}^n \frac{(2i-1)}{n} (\ln F(x_i; \gamma, \lambda) + \ln (1 - F(x_{n+1-i}; \gamma, \lambda))) \right) \\ &= -n - \frac{1}{n} \sum_{i=1}^n (2i-1) (\ln F(x_i; \gamma, \lambda) + \ln (1 - F(x_{n+1-i}; \gamma, \lambda))) \\ &= -n - \frac{1}{n} \sum_{i=1}^n (2i-1) (\ln F(x_i; \gamma, \lambda) + \ln S(x_{n+1-i}; \gamma, \lambda)) \\ &= -n - \frac{1}{n} \sum_{i=1}^n (2i-1) \left(\ln \left(\frac{1 - e^{-\lambda x_i}}{1 - (1-\gamma) e^{-\lambda x_i}} \right) + \ln \left(\frac{\gamma e^{-\lambda x_{n+1-i}}}{1 - (1-\gamma) e^{-\lambda x_{n+1-i}}} \right) \right). \end{aligned} \quad (3.40)$$

La dérivée première de $A(\gamma, \lambda)$ par rapport à γ est comme suit :

$$\begin{aligned} \frac{\partial A(\gamma, \lambda)}{\partial \gamma} &= \frac{\partial}{\partial \gamma} \left(-n - \frac{1}{n} \sum_{i=1}^n (2i-1) (\ln (F(x_i; \gamma, \lambda)) + \ln (S(x_{n+1-i}; \gamma, \lambda))) \right) \\ &= \frac{1}{n} \sum_{i=1}^n (2i-1) \left(\frac{\Delta_1(x_i; \gamma, \lambda)}{\left(\frac{1 - e^{-\lambda x_i}}{1 - (1-\gamma) e^{-\lambda x_i}} \right)} - \frac{\Delta_1(x_{n+1-i}; \gamma, \lambda)}{\left(\frac{\gamma e^{-\lambda x_{n+1-i}}}{1 - (1-\gamma) e^{-\lambda x_{n+1-i}}} \right)} \right) \end{aligned} \quad (3.41)$$

La dérivée première de $A(\gamma, \lambda)$ par rapport à λ est donnée par :

$$\begin{aligned} \frac{\partial A(\gamma, \lambda)}{\partial \lambda} &= \frac{\partial}{\partial \lambda} \left(-n - \frac{1}{n} \sum_{i=1}^n (2i-1) (\ln (F(x_i; \gamma, \lambda)) + \ln (S(x_{n+1-i}; \gamma, \lambda))) \right) \\ &= -\frac{1}{n} \sum_{i=1}^n (2i-1) \left(\frac{\Delta_2(x_i; \gamma, \lambda)}{\left(\frac{1 - e^{-\lambda x_i}}{1 - (1-\gamma) e^{-\lambda x_i}} \right)} + \frac{\Delta_2(x_{n+1-i}; \gamma, \lambda)}{\left(\frac{\gamma e^{-\lambda x_{n+1-i}}}{1 - (1-\gamma) e^{-\lambda x_{n+1-i}}} \right)} \right) \end{aligned} \quad (3.42)$$

avec $F(x_i; \gamma, \lambda) = \frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma) e^{-\lambda x_i}}$ et $S(x_{n+1-i}; \gamma, \lambda) = \frac{\gamma e^{-\lambda x_{n+1-i}}}{1 - (1 - \gamma) e^{-\lambda x_{n+1-i}}}$.

En égalant (3.41) et (3.42) à 0, on obtient les équations suivantes :

$$\sum_{i=1}^n (2i - 1) \left(\frac{\Delta_1(x_i; \gamma, \lambda)}{\left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma) e^{-\lambda x_i}} \right)} - \frac{\Delta_1(x_{n+1-i}; \gamma, \lambda)}{\left(\frac{\gamma e^{-\lambda x_{n+1-i}}}{1 - (1 - \gamma) e^{-\lambda x_{n+1-i}}} \right)} \right) = 0 \quad (3.43)$$

$$\sum_{i=1}^n (2i - 1) \left(\frac{\Delta_2(x_i; \gamma, \lambda)}{\left(\frac{1 - e^{-\lambda x_i}}{1 - (1 - \gamma) e^{-\lambda x_i}} \right)} - \frac{\Delta_2(x_{n+1-i}; \gamma, \lambda)}{\left(\frac{\gamma e^{-\lambda x_{n+1-i}}}{1 - (1 - \gamma) e^{-\lambda x_{n+1-i}}} \right)} \right) = 0 \quad (3.44)$$

où $\Delta_1(x_i; \gamma, \lambda)$ et $\Delta_2(x_i; \gamma, \lambda)$ sont donnés par (3.21) et (3.22).

Les estimateurs $\hat{\gamma}_{AD}$ et $\hat{\lambda}_{AD}$ peuvent également être obtenus en résolvant les équations non linéaires (3.43) et (3.44).

Dans ce chapitre nous avons déterminé les estimateurs des différentes méthodes d'estimation proposées. Les équations normales ci-dessus ne peuvent pas être résolues analytiquement, c'est pourquoi nous utiliserons les méthode numériques pour obtenir la solution de ces équations.

Simulation Numérique

Dans ce chapitre, nous allons faire une étude de simulation à l'aide de la méthode de monte-carlo. Cette dernière est une méthode algorithmique visant à calculer une valeur numérique approchée en utilisant des procédés aléatoires, c'est-à-dire des techniques probabilistes. L'objectif principal de ces simulations est de comparer l'efficacité des différentes méthodes d'estimation. Pour étudier les performances des estimateurs $\hat{\theta} = (\hat{\gamma}, \hat{\lambda})$, nous considérons sept méthodes différentes d'estimation des paramètres : MV, PC, MCO, MCP, PME, CVM et AD. Nous choisissons $\theta = ((0.5, 2), (2, 4))$ et générons des échantillons aléatoires de taille $n = (15, 20, 25, 30, 35, 40, 45, 50, 100, 150)$ à partir de la loi géométrique exponentielle étendue (GEE) en utilisant la méthode d'inversion et nous répétons ce processus pour $N = 10000$ fois. Nous calculons la valeur estimée pour N fois avec les méthodes d'estimation considérées en utilisant la formule suivante :

$$\hat{\theta}_j = \frac{1}{N} \sum_{i=1}^N \hat{\theta}_{ij}, j = 1, 2, \quad (4.1)$$

Et on calcule les erreurs quadratiques moyennes (EQM) et les erreurs relatives moyennes (ERM) en utilisant les formules suivantes :

$$EQM = \frac{1}{N} \sum_{i=1}^N (\hat{\theta}_{ij} - \theta_j)^2, j = 1, 2. \quad (4.2)$$

et

$$ERM = \frac{1}{N} \sum_{i=1}^N (\hat{\theta}_{ij}/\theta_j), j = 1, 2. \quad (4.3)$$

Le choix de θ est fait en suivant respectivement la décroissance et la croissance de la fonction de risque (hasard).

4.1 Les résultats de la simulation

Les figures 4.1.1 et 4.1.2 illustrent les résultats de cette simulation. Quatre graphiques sont présentés pour chacun des deux cas , c'est-à-dire le cas où $\theta = (0.5, 2)$ et le cas où

$\theta = (2, 4)$. Le premier et le troisième des quatre graphiques représentent l'erreur relatives moyennes de λ et de γ en fonction de la taille de l'échantillon. Les meilleurs estimateurs en termes d'ERM sont les plus bas sur ces deux graphiques. Le deuxième et le quatrième graphes représentent l'erreur quadratiques moyennes de λ et de γ en fonction de la taille de l'échantillon. Les meilleurs estimateurs en termes d'EQM sont les plus bas sur ces deux graphiques. Ainsi, les meilleurs estimateurs sont ceux qui présentent les ERMs et les EQMs les plus bas de ces quatre graphiques.

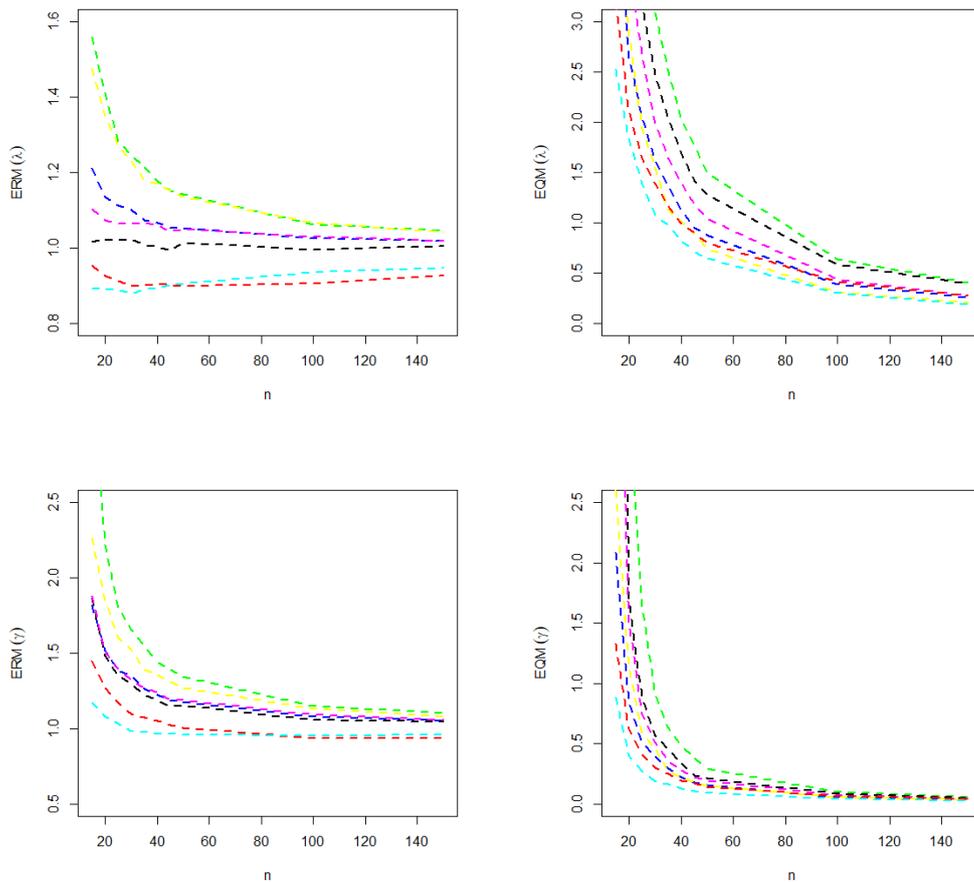


FIGURE 4.1.1 – ERMs et EQMs des estimations de $\gamma=0.5$ et $\lambda=2$ pour N échantillons simulés, en considérant différentes valeurs de n :AD, CVM, MV, MCO, MCP, PC et PME

La figure 4.1.1 présente les résultats de $GEE(0.5, 2)$. Les EQMs de toutes les estimations de γ et λ tendent vers 0 au fur et à mesure que la taille de l'échantillon augmente, les valeurs des ERM aussi tendent vers 1. Autrement dit, les estimations sont asymptotiquement sans biais pour les paramètres γ et λ . L'ERM des estimations de γ pour MV, AD, CVM, MCO et MCP est beaucoup plus importante que l'ERM de $\hat{\gamma}_{PC}$ et $\hat{\gamma}_{PME}$. L'ERM de $\hat{\gamma}_{PC}$ est plus importante que celui de $\hat{\gamma}_{PME}$ pour des échantillons de taille petite c'est-à-dire pour n petit. Mais pour n plus grand les courbes des ERM de ces deux coïncident. L'ERM des estimations de λ pour MV, AD, CVM, MCO et MCP est beaucoup plus importante que l'ERM de $\hat{\lambda}_{PC}$ et $\hat{\lambda}_{PME}$. L'ERM de $\hat{\lambda}_{PC}$ est plus importante que celui de $\hat{\lambda}_{PME}$ pour des échantillons de taille n petite et cette dernière devient plus importante que γ_{PC} à partir de grande taille. Ainsi, en termes d'ERM, on peut dire que PME est la meilleure méthode pour des échantillons de petite taille et PC pour de grande taille. Donc, on peut dire que la meilleure méthode en terme d'ERM dépend de la taille de l'échantillon qu'on a choisi. Les EQMs des estimations de γ et de λ pour MV, AD, CVM, MCO, MCP et PC sont beaucoup plus importantes que les EQMs de $\hat{\gamma}_{PME}$ et de $\hat{\lambda}_{PME}$. Ainsi, en termes de l'EQM, PME est la meilleure méthode. Les estimateurs PME présentent les EQM les plus faibles et les ERM les plus proches de un pour les deux paramètres, ce qui prouve qu'il s'agit de la procédure la plus efficace pour estimer les paramètres inconnus. Par conséquent, nous concluons que les estimateurs PME devraient être utilisés pour estimer les paramètres de la loi GEE pour $0 < \gamma < 1$. Pour $\theta = (0.5, 2)$, nous avons des résultats analogues pour la distribution géométrique exponentielle (voir [1]).

		$\gamma = 0.5 \quad \lambda = 2$									
n		15	20	25	30	35	40	45	50	100	150
AD	$\hat{\gamma}$	0.904	0.771	0.695	0.667	0.626	0.606	0.594	0.588	0.543	0.528
	$\hat{\lambda}$	2.389	2.297	2.222	2.19	2.145	2.124	2.109	2.116	2.055	2.032
CVM	$\hat{\gamma}$	1.562	1.098	0.927	0.823	0.773	0.733	0.691	0.675	0.582	0.552
	$\hat{\lambda}$	3.136	2.789	2.603	2.491	2.44	2.369	2.308	2.288	2.15	2.088
MV	$\hat{\gamma}$	1.148	0.903	0.828	0.745	0.71	0.673	0.649	0.635	0.561	0.542
	$\hat{\lambda}$	2.972	2.689	2.579	2.454	2.383	2.328	2.288	2.255	2.123	2.087
MCO	$\hat{\gamma}$	0.918	0.752	0.683	0.644	0.596	0.588	0.584	0.571	0.537	0.524
	$\hat{\lambda}$	2.094	2.056	2.045	2.041	1.986	2.003	2.026	2.007	2.014	2.007
MCP	$\hat{\gamma}$	0.933	0.756	0.698	0.65	0.621	0.619	0.598	0.592	0.548	0.526
	$\hat{\lambda}$	2.181	2.176	2.146	2.109	2.088	2.106	2.089	2.095	2.057	2.027
PC	$\hat{\gamma}$	0.716	0.633	0.582	0.553	0.528	0.514	0.51	0.501	0.466	0.467
	$\hat{\lambda}$	1.887	1.846	1.816	1.805	1.795	1.786	1.802	1.798	1.808	1.851
PME	$\hat{\gamma}$	0.587	0.529	0.503	0.489	0.485	0.482	0.478	0.473	0.476	0.477
	$\hat{\lambda}$	1.809	1.76	1.768	1.749	1.771	1.784	1.796	1.784	1.863	1.887

TABLE 4.1.1 – Les estimations de $\gamma = 0.5$ et $\lambda = 2$ des différentes méthodes

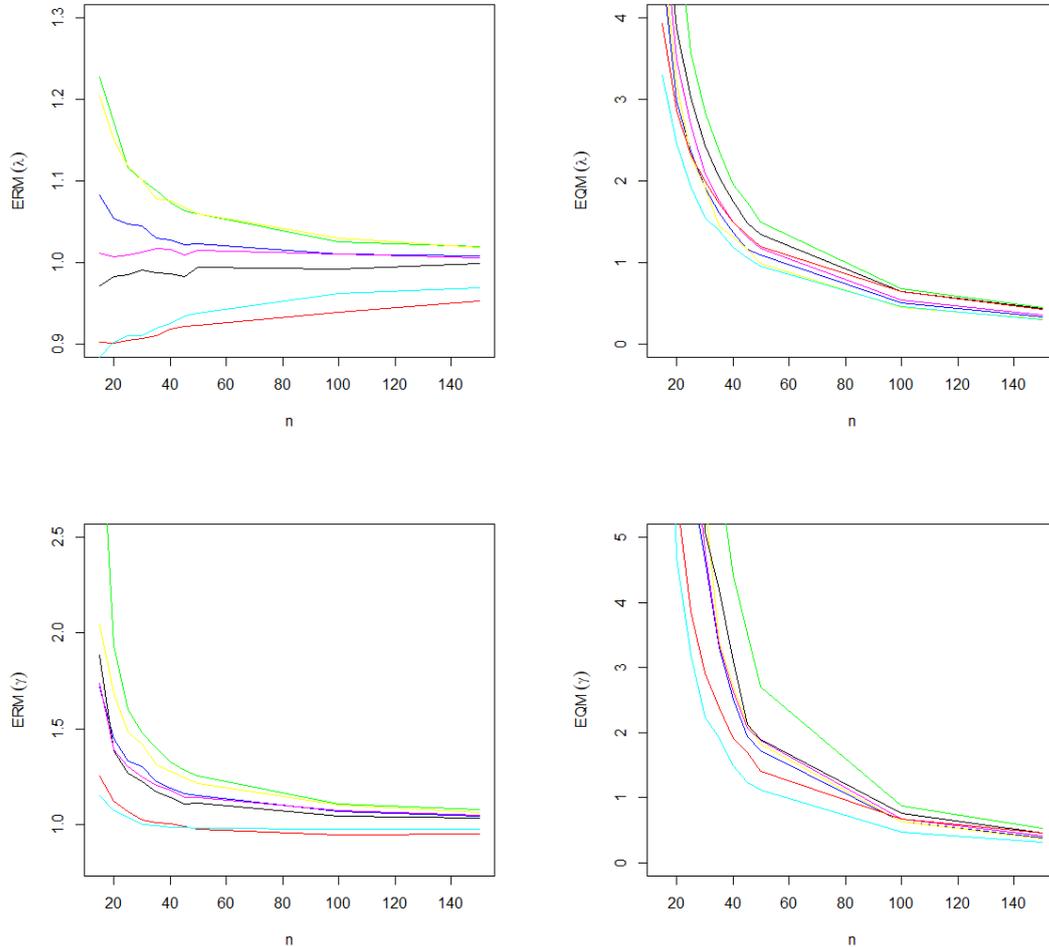


FIGURE 4.1.2 – ERM et EQM des estimations de $\gamma=2$ et $\lambda=4$ pour N échantillons simulés, en considérant différentes valeurs de n : AD, CVM, MV, MCO, MCP, PC et PME

La figure 4.1.2 présente les résultats de $GEE(2, 4)$. L'ERM et l'EQM des paramètres γ et λ diminuent au fur et à mesure que la taille l'échantillon augmente. De plus, les ERM tendent vers 1 et les EQM vers 0. Les ERM des estimations de γ et λ pour MV et CVM sont beaucoup plus importantes que les ERM des estimations γ et λ pour AD, MCO, MCP, PC et PME. $\hat{\gamma}_{PC}$, $\hat{\gamma}_{PME}$, $\hat{\lambda}_{PC}$ et $\hat{\lambda}_{PME}$ présentent des ERM plus faibles. En terme d'ERM, on peut dire que les méthodes de PC et du PME sont les meilleures pour estimer les paramètres de la loi GEE. De plus, on peut dire que la meilleure mé-

thode dépend de la taille de l'échantillon. En effet, pour des échantillons de petite taille, l'ERM de $\hat{\gamma}_{PME}$ et $\hat{\lambda}_{PME}$ sont en dessous de l'ERM de $\hat{\gamma}_{PC}$ et de $\hat{\lambda}_{PC}$. Ces derniers sont dessous pour des échantillons de grande taille. Les EQMs des estimation de γ et λ de CVM sont plus importantes que les EQMs des autres méthodes. Les EQMs de $\hat{\gamma}_{PME}$ et $\hat{\lambda}_{PME}$ sont les plus faibles. Ainsi, en terme d'EQM on peut dire que la méthode du PME est la meilleure. Nous concluons que la méthode du PME est la plus performante pour estimer les paramètres de la GEE. Les résultats de $GEE(2, 4)$ sont analogues pour la distribution géométrique exponentielle complémentaire (voir [19]).

		$\gamma = 2 \quad \lambda = 4$										
		n	15	20	25	30	35	40	45	50	100	150
AD	$\hat{\gamma}$		3.469	2.914	2.672	2.61	2.454	2.381	2.319	2.3	2.134	2.088
	$\hat{\lambda}$		4.391	4.243	4.204	4.186	4.123	4.116	4.086	4.093	4.04	4.031
CVM	$\hat{\gamma}$		6.599	3.869	3.205	2.96	2.799	2.649	2.569	2.502	2.212	2.154
	$\hat{\lambda}$		4.928	4.69	4.466	4.404	4.354	4.293	4.252	4.238	4.099	4.078
MV	$\hat{\gamma}$		4.104	3.374	2.975	2.843	2.625	2.56	2.492	2.425	2.204	2.12
	$\hat{\lambda}$		4.842	4.62	4.486	4.41	4.315	4.306	4.275	4.24	4.12	4.072
MCO	$\hat{\gamma}$		3.784	2.772	2.532	2.448	2.341	2.285	2.213	2.226	2.089	2.069
	$\hat{\lambda}$		3.921	3.941	3.941	3.963	3.947	3.945	3.931	3.977	3.967	3.993
MCP	$\hat{\gamma}$		3.48	2.783	2.605	2.493	2.412	2.363	2.284	2.286	2.141	2.091
	$\hat{\lambda}$		4.059	4.03	4.036	4.051	4.069	4.065	4.037	4.059	4.041	4.022
PC	$\hat{\gamma}$		2.52	2.247	2.132	2.047	2.021	2.004	1.978	1.951	1.886	1.897
	$\hat{\lambda}$		3.637	3.612	3.622	3.627	3.642	3.673	3.685	3.693	3.754	3.812
PME	$\hat{\gamma}$		2.395	2.205	2.104	2.019	2.005	1.98	1.979	1.969	1.953	1.948
	$\hat{\lambda}$		3.683	3.695	3.684	3.675	3.702	3.716	3.747	3.759	3.846	3.877

TABLE 4.1.2 – Les estimations de $\gamma = 2$ et $\lambda = 4$ des différentes méthodes

4.2 Applications

Dans cette section, nous considérons deux données réelles. La première est présentée par Boag (voir dans [6]) et concerne les âges (en mois) de 18 patients décédés d'autres causes que le cancer (tableau 4.2.3). La seconde donnée est présentée par Silva (voir dans [30]) et se réfère au temps d'inversion du sérum (en jours) de 143 enfants nés de mères infectées par le VIH qui n'ont pas reçu de traitement anti-VIH (tableau 4.2.7).

Nos résultats de la section précédente nous indique que les estimateurs du produit maximal d'espacements (PME) devraient être utilisés pour estimer les paramètres de loi GEE. Tout d'abord, nous allons comparer les estimations obtenues par les différentes méthodes d'estimations avec l'estimateur du PME en termes d'ERM. Ensuite, on compare les résultats obtenus à partir de la loi de GEE ajustée par les estimateurs du PME

avec certains modèles de durée de vie, comme les lois Weibull, Gamma, Lognormale et Exponentielle généralisée.

Nous avons envisagé de faire le test de Kolmogorov-Smirnov (KS) pour vérifier la qualité des ajustements. Cette procédure est basée sur la statistique D_n de KS. Cette statistique est donnée par

$$D_n = \sup_x |F_n(x) - F(x; \gamma, \lambda)| \quad (4.4)$$

avec n le nombre d'observations X , $F_n(x)$ la fonction de répartition empirique (observée) et $F(x; \gamma, \lambda)$ la fonction de répartition théorique de X .

Dans ce cas, nous avons considéré deux hypothèses

- H_0 : est l'hypothèse nulle selon laquelle les données proviennent de $F(x; \gamma, \lambda)$ avec un niveau de confiance $\alpha = 5\%$.
- H_1 : est l'hypothèse selon laquelle les données ne proviennent pas de $F(x; \gamma, \lambda)$

Autrement dit,

- $H_0 : F_X = F(x; \gamma, \lambda)$
- $H_1 : F_X \neq F(x; \gamma, \lambda)$

Nous rejetons l'hypothèse nulle si la valeur du p -value est inférieure à 0,05. C'est-à-dire avec une confiance de 95%.

On a déterminé aussi les intervalles de confiance des paramètres . En effet, l'intervalle de confiance asymptotique $100(1 - \alpha)\%$ est donnée par la formule suivante :

$$IC_{95\%}(\theta) = \left(\hat{\theta} - z_{\alpha/2} \sqrt{J(r, r)/n}, \hat{\theta} + z_{\alpha/2} \sqrt{J(r, r)/n} \right) \quad (4.5)$$

avec $z_{\alpha/2}$ est le quantile $1 - \alpha/2$ de la loi normale standard, n la taille de l'échantillon, $J(r, r)$ représente les éléments de la diagonale de l'inverse de matrice d'information de Fisher et $r = 1, 2$.

Nous avons considéré 4 méthodes comme critère de discrimination pour ce travail. Ces méthodes sont : AIC (Critère d'information d'Akaike), AICc (Critère d'information d'Akaike corrigé), HQIC (Critère d'information de Hannan-Quinn) et CAIC (Critère d'information d'Akaike Cohérent).

Le critère d'information d'Akaike est calculé comme suit :

$$AIC = -2l(\hat{\Theta}, x) + 2k. \quad (4.6)$$

Le critère d'information d'Akaike corrigé est donné par la relation suivante :

$$AICc = AIC + \frac{2k(k+1)}{(n-k-1)}. \quad (4.7)$$

Le critère d'information de Hannan-Quinn et Critère d'information d'Akaike Cohérent sont donnés respectivement par les formules suivantes

$$HQIC = -2l(\hat{\Theta}, x) + 2k \log(\log(n)) \quad (4.8)$$

et

$$CAIC = -2l(\hat{\Theta}, x) + k \log(n) + 1. \quad (4.9)$$

où k représente le nombre paramètres à ajuster, $\hat{\Theta}$ l'estimation de Θ , n la taille de l'échantillon et $l(\hat{\Theta}, x)$ la fonction de vraisemblance. Étant donné qu'on a plusieurs modèles candidats, on choisit le modèle qui fournit des valeurs minimales pour les critères de discrimination.

4.2.1. Les données de Boag

Les données du tableau ci-dessous sont extraites de Boag (voir [6]), qui a considéré la loi lognormale pour décrire de telles données. Ces dernières représentent l'âge (en mois) de 18 patients décédés d'autres causes que le cancer.

0.3	4	7.4	15.5	23.4	46	46	51	65
68	83	88	96	110	111	112	132	162

TABLE 4.2.3 – Données relatives à l'âge (en mois) de 18 patients décédés d'autres causes que le cancer

Tout d'abord, nous avons déterminé $\hat{\gamma}_{PME} = 3.18655762$ et $\hat{\lambda}_{PME} = 0.02391472$ du PME. En remplaçant $\hat{\gamma}_{PME}$ et $\hat{\lambda}_{PME}$ dans la formule de l'intervalle de confiance, on obtient les résultats suivants :

$$IC_{95\%}(\gamma) = (2.242537; 4.130578)$$

et

$$IC_{95\%}(\lambda) = (0.02118844; 0.026641).$$

Et on a calculé aussi les estimations de γ et λ des différentes méthodes d'estimations en utilisant les données du tableau 4.2.3. Les résultats de ces calculs sont représentés dans le tableau ci-dessous.

$\hat{\theta}$	MV	MCO	MCP	PC	CVM	AD
$\hat{\lambda}$	0.0261104	0.02025308	0.02166199	0.02563183	0.02341081	0.02075455
$\hat{\gamma}$	3.9025563	2.55356454	2.75946472	3.97831396	3.40571427	2.40363670

TABLE 4.2.4 – Résultats des estimations des données du tableau 4.2.3

Nous avons utilisé les estimations de γ et λ de la méthode du produit maximal d'espacements pour comparer les ERM des différentes méthodes en considérant les données du tableau 4.2.4, étant donné que cette méthode est la meilleure. Ces résultats sont présentés :

$\hat{\theta}_{PME}/\hat{\theta}$	MV	MCO	MCP	PC	CVM	AD
λ	0.9159079	1.180794	1.103995	0.9330086	1.0215249	1.152264
γ	0.8165309	1.247886	1.154774	0.8009819	0.9356503	1.325723

TABLE 4.2.5 – Comparaison des ERM des estimations obtenues à partir des différentes méthodes à celle de PME en utilisant les données du tableau 4.2.3

Les résultats du tableau ci-dessus vient confirmer les résultats de la simulation de la section précédente. C'est-à-dire pour des échantillons de petite taille les ERM dépendent de la méthode d'estimation considérée. Ensuite, nous avons calculé et tracé les courbes des fonctions de survies de la GEE et des différents modèles de l'âge des 18 patients décédés d'autres causes que le cancer. Ces courbes sont représentées dans la figure suivante :

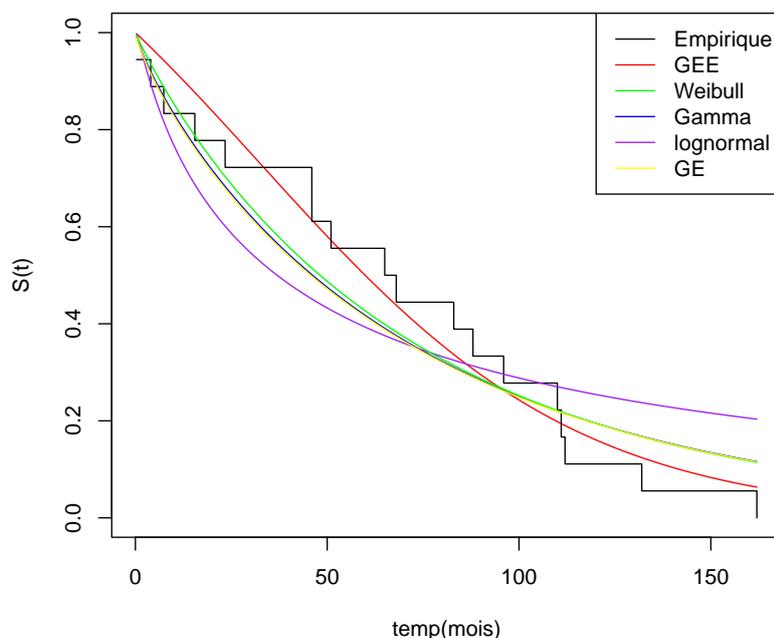


FIGURE 4.2.3 – Fonctions de survies des différents modèles des 18 patients décédés d'autres causes que le cancer

En observant la figure 4.2.3, on voit que la courbe de la fonction de survie empirique suit l'allure des courbes des différentes lois. De plus, on constate qu'elle est plus proche

de la courbe de GEE. Donc on peut supposer que la GEE est le modèle qu'il faut parmi les modèles choisis. Et enfin, nous avons calculé la p-value, le AIC, le AICc, le HQIC et le CAIC de la GEE et des différents modèles en utilisant les estimations du PME. Le tableau suivant présente le résultat de ces calculs.

Test	GEE	Weibull	Gamma	Lognormale	GE
KS	0.9824	0.4198	0.3424	0.1416	0.3245
AIC	189.496	192.2781	192.2712	202.4442	192.2471
AICc	190.296	193.0781	193.0712	203.2442	193.0471
HQIC	189.7415	192.5237	192.5167	202.6897	192.4927
CAIC	192.2767	195.0589	195.0519	205.2249	195.0279

TABLE 4.2.6 – Résultats du test KS (p-value), de l'AIC, de l'AICc, du HQIC et du CAIC pour les différentes lois de probabilités en utilisant les données du tableau 4.2.3

On voit que toutes les p-values du test de Kolmogorov sont inférieures à α . C'est-à-dire que $KS > 0.05$. Donc on accepte l'hypothèse nulle à 95% pour la GEE et tous les modèles choisis. De plus on ne peut rien dire sur quel modèle choisir. Donc nous allons nous intéresser au critère de discrimination. On constate que le AIC, le AICc, le HQIC et le CAIC de la GEE sont inférieures à ceux des autres modèles. Donc, les résultats du tableau 4.2.6 viennent confirmer les observations de la figure 4.2.3. Donc on peut en déduire que la GEE est le modèle adéquat parmi les modèles qu'on a choisi pour cette donnée de Boag. Dans ce cas, chacune des causes du décès peut être décrite par une loi exponentielle; étant donné que la durée de vie associée à un risque particulier n'est pas observable (variables latentes), nous n'observons que la valeur maximale de la durée de vie ($\gamma > 1$) parmi tous les risques, tandis que le nombre de causes suit une loi géométrique.

4.2.2. Les enfants exposés à la transmission du VIH

Dans cette sous section, nous utilisons les données de Silva. Ces dernières représentent le temps d'inversion du sérum (en jours) de 143 enfants nés de mères infectées par le VIH. Ces données sont présentés dans le tableau ci-dessous :

2	2	2	5	9	9	19	32	32	46	50	56	56	78	91	95
106	129	129	148	149	156	175	176	191	192	204	209	211	225	229	230
238	254	271	274	276	290	291	292	297	297	322	334	334	334	344	346
353	353	359	365	366	367	370	378	378	382	382	385	398	400	402	414
422	424	428	434	435	440	443	446	448	448	451	454	459	460	461	473
480	481	484	487	493	497	498	502	511	511	513	514	516	521	524	526
537	538	541	543	544	544	545	549	551	553	553	554	556	559	571	576
577	578	582	588	590	596	609	610	615	619	626	627	648	653	678	680
687	696	729	744	748	777	847	848	867	874	894	901	907	974	1021	

TABLE 4.2.7 – Données relatives au temps de retournement du sérum (en jours) de 143 enfants nés de mères infectées par le VIH.

Tout d'abord, nous avons calculé $\hat{\gamma}_{PME} = 14.809056598$ et $\hat{\lambda}_{PME} = 0.006581982$ du PME. Les intervalles de confiance des paramètres γ et λ sont donnés comme suit :

$$IC_{95\%}(\gamma) = (14.1017; 15.51641)$$

et

$$IC_{95\%}(\lambda) = (0.006491866; 0.006672098).$$

Et on a déterminé aussi les estimations de γ et λ pour les différentes méthodes d'estimations avec les données du tableau 4.2.7. Voici les résultats de ces calculs :

	MV	MCO	MCP	PC	CVM	AD
$\hat{\lambda}$	0.00672806	0.007035631	0.006956492	0.006964211	0.007132088	0.006507967
$\hat{\gamma}$	15.57051010	19.170756335	17.733421554	16.664791442	20.052704007	14.048103724

TABLE 4.2.8 – Résultats des estimations des données du tableau 4.2.7

Ensuite, nous avons calculé et comparé les ERM des estimations de ces différentes méthodes d'estimation avec PME. Les résultats de ces calculs sont présentés dans le tableau suivant.

Le tableau ci-dessous montre que pour des échantillons de grande taille, les estimations sont très proches indépendamment de la méthode choisie. De plus, en raison de la grande taille des échantillons, les estimations de la méthode PME et les estimations de la méthode MV sont presque identiques. La figure ci-dessous présente les courbes des fonctions de survies ajustées des différents modèles choisis et empirique du tableau 4.2.7.

$\hat{\theta}_{PME}/\hat{\theta}$	MV	MCO	MCP	PC	CVM	AD
λ	0.9782881	0.9355211	0.946164	0.9451152	0.9228688	1.011373
γ	0.9510964	0.7724816	0.835093	0.8886434	0.7385067	1.054168

TABLE 4.2.9 – Comparaison des ERM des estimations obtenues à partir des différentes méthodes à celle du PME en utilisant les données du tableau 4.2.7

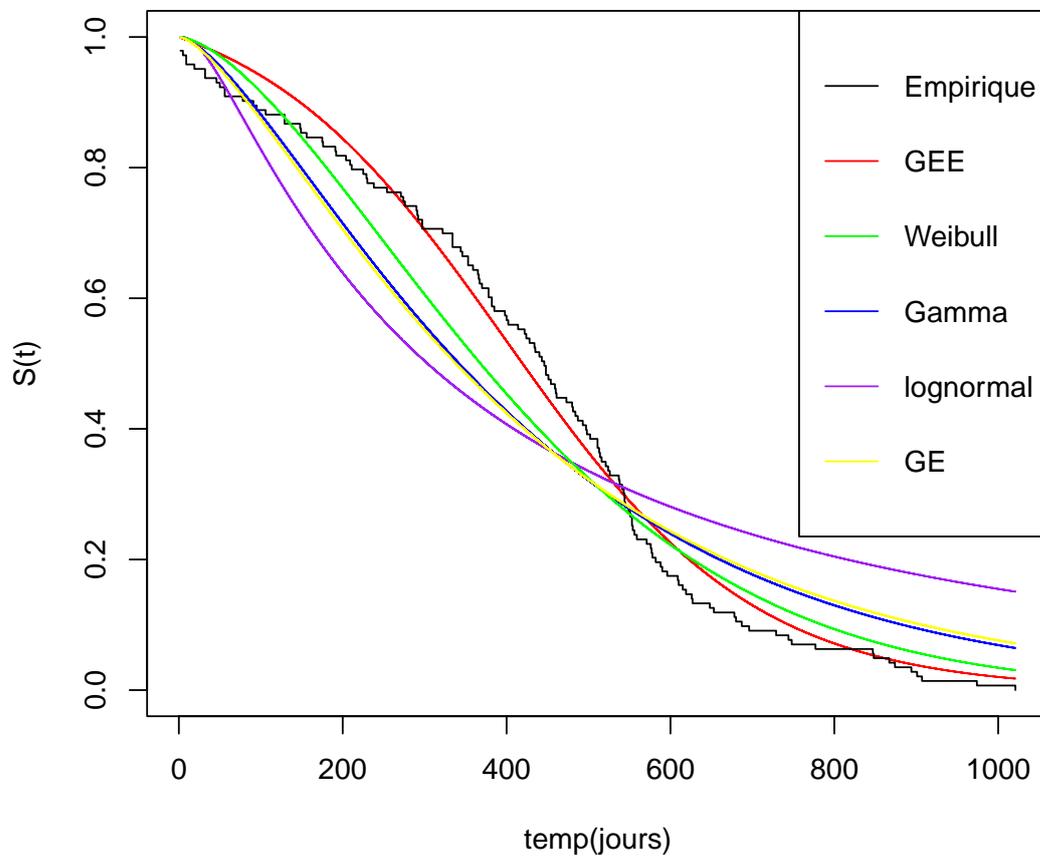


FIGURE 4.2.4 – fonctions de survies des différents modèles en utilisant les données du tableau 4.2.7

On observe que la courbe de la fonction de survie empirique a la même allure que celle de la GEE. Et en comparant la fonction de survie empirique avec celles des autres lois de la figure 4.2.4, on constate que la loi géométrique exponentielle étendue est la plus adéquate pour les données du tableau 4.2.7 .

Et enfin , nous avons calculé la p-value, le AIC, le AICc, le HQIC et CAIC de la GEE et les différentes modèles choisis en considérant les estimations du PME. Voici, les résultats de ces calculs :

Test	GEE	Weibull	Gamma	Lognormale	GE
KS	0.6312	0.004033	$7.688.10^{-05}$	$4.241.10^{-07}$	$4.24.10^{-05}$
AIC	1950.801	1981.592	2002.121	2088.911	2005.529
AICc	1950.887	1981.677	2002.206	2088.997	2005.615
HQIC	1953.209	1984	2004.529	2091.319	2007.937
CAIC	1957.727	1988.517	2009.046	2095.837	2012.455

TABLE 4.2.10 – Résultats du test KS (p-value), de l’AIC, de l’AICc, du HQIC et du CAIC pour les différentes lois de probabilités en utilisant les données du tableau 4.2.7

On constate que les p-values de Weibull, de Gamma, de lognormale et de exponentielle généralisée sont toutes inférieures à 0.05 tandis que la p-value de GEE est supérieure à 0.05. Donc, on rejette l’hypothèse nulle pour les modèles choisis et on accepte l’hypothèse nulle pour la GEE. Pour le critère de discrimination, on observe que les valeurs de AIC, de AICc, de HQIC et de CAIC de la GEE sont minimales à celles des autres modèles qu’on a choisi. Les résultats du tableau 4.2.10 confirment les observations de la figure 4.2.4. Donc, on peut dire que la GEE est la meilleure alternative simple pour les données du tableau 4.2.7.

4.3 Conclusion

D’après les résultats de la simulation, nous avons remarqué que l’ estimateur du PME présente une faible erreur quadratique moyenne et aussi une faible erreur relative moyenne. C’est ce qui nous a permis de conclure que l’estimateur PME est meilleur pour estimer les paramètres. De plus, on a fait une étude comparative en utilisant l’estimateur PME et en calculant le p-value, le AIC, le AICc, le HQIC et CAIC des différents modèles. De ce fait, on a conclu que GEE est le modèle adéquat parmi les autres modèles.

Conclusion générale

Dans ce travail, nous avons eu à comparer par le biais d'une étude de simulation, les estimations des paramètres de la loi géométrique exponentielle étendue. Cette étude a été faite à l'aide de sept méthodes d'estimation. Les estimations des paramètres de ces différentes méthodes ont été obtenues par la fonction optim du logiciel r. Ce pendant, les erreurs relatives moyennes des estimateurs de la méthode de Cramer-Von-Mise et du maximum de vraisemblance sont plus importantes parmi tous les sept méthodes considérées. De plus, l'estimateur de CVM a la plus importante erreur quadratique moyenne de tous. Tandis que, les estimateurs des paramètres du produit maximal des espacements ont respectivement les erreurs relatives moyennes (ERM) et erreurs quadratiques moyennes (EQM) les plus faibles. C'est ce qui prouve qu'il s'agit de la méthode la plus efficace par rapport aux autres méthodes pour estimer les paramètres inconnus. De même, en tenant compte des propriétés de la méthode telles que la cohérence, l'efficacité asymptotique, la normalité et l'invariance, nous avons conclu que les estimateurs du PME sont les meilleurs pour estimer les paramètres de la loi géométrique exponentielle étendue par rapport à leurs concurrents.

Et enfin, nous avons appliqué cette méthode du PME à deux ensembles de données réelles. Cette application nous a permis de démontrer que la géométrique exponentielle étendue est la meilleure alternative simple à utiliser pour les applications de la durée de vie parmi les modèles choisis.

Comme perspectives, nous pouvons élargir ce travail en utilisant d'autres méthodes d'estimation différentes de ce que nous avons étudié en les comparant à la méthode du PME avec les mêmes modèles choisis. Nous pouvons aussi élargir ce travail en comparant le modèle de la loi GEE à ceux des lois des valeurs extrêmes.

Bibliographie

- [1] Adamidis K. , Loukas S.(1998)., " A lifetime distribution with decreasing failure rate ", *Statistics & Probability Letters*, vol. 39, no 1, p. 35-42, juill. 1998.
- [2] Adamidis K., Loukas S., Dimitrakopoulou T.(2005)., " On an extension of the exponential-geometric distribution ", *Statistics & Probability Letters*, vol. 73, no. 3, p. 259-269, juill. 2005.
- [3] Amin N. A. K., Cheng R. C. H.(1983)., "Estimating parameters in continuous univariate distributions with a shifted origin," *Journal of the Royal Statistical Society, Series B : Methodological*, vol.45,no.3,pp.394-403.
- [4] Amin N. A. K., Cheng R. C. H.(1979)., "Maximum product of spacings estimation with application to the lognormal distributions", *Math Report 79-1*, Department of Mathematics, UWIST, Cardiff, UK.
- [5] Bidram H., Behboodian J., Towhidi M.(2013)., " A New Generalized Exponential Geometric Distribution ", *Communications in Statistics - Theory and Methods*, vol. 42, no. 3, p. 528-542, févr. 2013.
- [6] Boag J. W.(1949)., "Maximum likelihood estimates of the proportion of patients cured by cancer therapy", *Journal of the Royal Statistical Society, Series B*,vol.11,pp.15-53.
- [7] Casella G., Berger R.(2002)., *Statistical Inference*, Duxbury, Belmont, Calif, USA, 2nd edition.
- [8] Dempster.(1977)., " Maximum Likelihood from Incomplete Data Via the EM Algorithm ",*Journal of the Royal Statistical Society : Series B (Methodological)*, Wiley Online Library .
- [9] Erdelyi A., Oberhettinger F., Tricomi F.G., Maguns W.(1953)., *Higher Transcendental Functions*, McGraw-Hill, NewYork , NY, USA.
- [10] Hosking J. R. M.(1990)., "L-Moments : Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics" , *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 52, no 1, p. 105-124.
- [11] Kao J. H. K.(1959)., " A Graphical Estimation of Mixed Weibull Parameters in Life-Testing of Electron Tubes ", *Technometrics*, vol. 1, no 4, p. 389-407, nov. 1959.

- [12] Kao J. H. K.(1958)., " Computer Methods for Estimating Weibull Parameters in Reliability Studies ", *IRE Transactions on Reliability and Quality Control*, vol.13, p. 15-22, juill. 1958.
- [13] Kitidamrongsuk P.(2010)., *Discriminating between the extended exponential geometric distribution and the Gamma distribution Doctor of philosophy(statistics)* , National Institute of Development Administration, bangkok, Thailand.
- [14] Kundu D., Dey S., Dey T.(2014)., " Two-Parameter Rayleigh Distribution : Different Methods of Estimation ", *American Journal of Mathematical and Management Sciences*, vol. 33, no 1, p. 55-74, janv. 2014.
- [15] Kundu D., Gupta R. D.(2001)., " Generalized exponential distribution : different method of estimations ", *Journal of Statistical Computation and Simulation*, vol. 69, no. 4, p. 315-337, juill. 2001.
- [16] Kundu D., Gupta R. D.(1999). , " Generalized exponential distributions ", *Australian & New Zealand Journal of Statistics*, vol. 41, no. 2, p. 173-188, juin 1999.
- [17] Louzada F., Carpenter J., Marchi V.(2013)., "The Complementary Exponentiated Exponential Geometric Lifetime Distribution ", *Journal of Probability and Statistics*, vol. 2013, p. 1-12, 2013.
- [18] Louzada F., Mazucheli J. ,Ghitany M. E.(2013)., " Comparison of estimation methods for the parameters of the weighted Lindley distribution ", *Applied Mathematics and Computation*, vol. 220, p. 463-471, sept. 2013.
- [19] Louzada F., Roman M., Cancho V. G.(2011)., "The complementary exponential geometric distribution : Model, properties, and a comparison with its counterpart ", *Computational Statistics & Data Analysis*, vol. 55, no. 8, p. 2516-2524, août 2011.
- [20] Louzada F., Roman M.,Marchi V.(2014)., " The exponentiated exponential-geometric distribution : a distribution with decreasing, increasing and unimodal failure rate ", *Statistics*, vol. 48, no 1, p. 167-181, janv. 2014.
- [21] Luceño A.(2006)., "Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators", *Computational Statistics & Data Analysis*, vol.51,no.2,pp.904-917.
- [22] Macdonald P.D. M.(1971)., "Comment on 'an estimation procedure for mixtures of distributions' by Choi and Bulgren", *Journal of the Royal Statistical Society, Series B : Methodological*, vol.33,no. 2, pp. 326-329.
- [23] Marshall A. W., Olkin I.(1997)., " A New Method for Adding a Parameter to a Family of Distributions with Application to the Exponential and Weibull Families ", *Biometrika*, vol. 84, no 3, p. 641-652.
- [24] Moala F. A., Achcar J. A., Ramos P. L.(2014)., " Objective Priors for Estimation of Extended Exponential Geometric Distribution ", *Journal of Modern Applied Statistical Methods*, vol. 13, no 2, p. 226-243, nov. 2014.
- [25] Morais A. L., Cordeiro G. M., Barreto-Souza W.(2011)., " The Weibull-Geometric distribution ", *Journal of Statistical Computation and Simulation*, vol. 81, no 5, p. 645-657, mai 2011.

- [26] Myung I. J.(2003), " Tutorial on maximum likelihood estimation ", *Journal of Mathematical Psychology*, vol. 47, no 1, p. 90-100, févr. 2003.
- [27] Nassar M.,Nada N. (2012)., " A new generalization of the exponential-geometric distribution ", *Journal of Statistics : Advances in Theory and Applications*, vol. 7, janv. 2012.
- [28] Ranney B.(1984)., "The maximum spacing method : an estimation method related to the maximum likelihood method", *Scandinavian Journal of Statistics*, vol.11,no.2,pp.93-112.
- [29] Saracoglu B., Tanis C.(2019)., " Comparisons of six different estimation methods for log-Kumaraswamy distribution ", *Thermal science*, vol. 23, no Suppl. 6, p. 1839-1847.
- [30] Silva A.N.F.(2004)., *Estudo evolutivo das crianças expostas ao VIH e notificadas pelo Núcleo de Vigilância epidemiológica do HCFMRP-USP*,dissertação de Mestrado , Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo,Brazil.
- [31] Singpurwalla N. D., Mann N. R., Schafer R. E.(1974)., *Methods for Statistical Analysis of Reliability and Life Data*, John Wiley & Sons, New York, NY, USA.
- [32] Stephens M. A., Cheng R. C. H.(1989)., "A goodness-of-fit test using Moran's statistic with estimated parameters", *Biometrika*, vol.76, no. 2, pp. 385-392.
- [33] Stephens M., D'Agostino R.(1986). , *Goodness-of-Fit Techniques*, Marcel Dekker, New York, NY, USA.
- [34] Swain J. J., Wilson J. R., Venkatraman S.(1988)., Least-squares estimation of distribution functions in johnson's translation system , *Journal of Statistical Computation and Simulation*, vol. 29, no 4, p. 271-297, juin 1988.
- [35] Teimouri M., Nadarajah S., Hoseini S. M.(2013)., " Comparison of estimation methods for the Weibull distribution ", *Statistics*, vol. 47, no 1, p. 93-109, févr. 2013.
- [36] "L'utilité des statistiques ", RBC.Vol. 37, N° 1 , Janvier 1956 . [En ligne]. Disponible sur :<https://www.rbc.com/fr/notre-compagnie/histoire/bulletin/vol-37-n-1-janvier-1956-lutilite-des-statistiques>.
- [37] "L'histoire des probabilités", <https://blogpeda.ac-bordeaux.fr/labomaths-redon-pauillac/files/2020/11/HISTOIRE-DES-PROBABILIT%C3%89S.pdf>