

Faculté des sciences

Bases techniques de l'assurance obsèques

Auteure : Mariam Leila Zatio SANOGO

Promoteur : Michel DENUIT

Lecteur : Karim BARIGOU

Année académique 2024-2025

Master en sciences actuarielles, à finalité spécialisée

Résumé

La mortalité aux âges extrêmes représente un enjeu technique majeur pour la tarification des assurances obsèques. Si les tables réglementaires, comme la table XK, assurent une bonne représentativité aux âges jeunes, elles deviennent moins fiables au-delà de 95, voire 100 ans, où les données se raréfient. Elles s'avèrent alors trop prudentes pour ce type de garantie, notamment en raison d'un chargement de sécurité incorporé, y compris aux âges avancés.

Ce mémoire, à travers une série d'analyses fondées sur des données belges récentes issues de la base IDL, mobilise des techniques issues de la théorie des valeurs extrêmes, en particulier l'approche proposée par Gbari, Poulain, Dal et Denuit (2017), afin d'extrapoler la mortalité au-delà des âges observés.

Il en résulte la construction de tables de mortalité d'expérience, genrées et unisexe, établies en partie par raccordement aux statistiques de la BNB sur la période 2018–2022. Ces tables, plus proches des profils assurés, constituent une base technique mieux adaptée aux besoins concrets des assureurs obsèques. Elles permettent une tarification plus adéquate et une évaluation plus précise des engagements à très long terme, indispensables dans un marché concurrentiel.

Abstract

Mortality at extreme ages presents a significant technical challenge for the pricing of funeral insurance. While regulatory tables such as the XK table offer reliable estimates at younger ages, their accuracy declines beyond age 95 or 100 due to the scarcity of available data. These tables tend to be overly conservative for this type of coverage, particularly due to an embedded safety margin, even at advanced ages.

This thesis, through a series of analyses based on recent Belgian data from the IDL database, applies techniques from extreme value theory, especially the approach introduced by Gbari, Poulain, Dal and Denuit in 2017, to extrapolate mortality beyond observed ages.

The result is the construction of gender-specific and unisex experience tables, partly aligned with observed mortality rates in the Belgian market between 2018 and 2022. These tables, which better reflect the actual profiles of insured individuals, constitute a more appropriate technical basis for funeral insurers. They enable more accurate pricing and a more precise evaluation of long-term liabilities, which is essential in a competitive market.

Remerciements

Je tiens tout d'abord à exprimer ma profonde gratitude à mon promoteur, Monsieur Michel DENUIT, pour sa disponibilité, sa rigueur scientifique et la clarté de ses conseils tout au long de l'élaboration de ce mémoire. Son exigence méthodologique et sa bienveillance ont été une source constante de motivation.

Je remercie également l'ensemble du corps enseignant du master en sciences actuarielles de l'UCLouvain, dont la qualité de l'enseignement m'a permis d'acquérir les bases solides et les compétences nécessaires à la réalisation de ce travail.

Un merci tout particulier à ma famille et à mes amis pour leur soutien indéfectible, leur patience et leurs encouragements, même dans les moments de doute. Vous avez toujours cru en moi.

J'exprime toute ma gratitude à COULIBALY Yoh Fatim, KOMARA Ben Ahmed et ESSEHIN Derre, qui ont été formidables tout au long de ce parcours. Leur présence et leur soutien constant ont compté plus que je ne saurais le dire.

Je dédie enfin ce mémoire à mon père, qui aurait sûrement été fier de moi.

Table des figures

3.1	Histogrammes des âges au décès – Femmes (Groupe 1 à gauche, Groupe 2 à droite)	24
3.2	Histogrammes des âges au décès – Hommes (Groupe 1 à gauche, Groupe 2 à droite) . . .	25
3.3	Boxplots comparés des âges au décès – (Groupe 1 à gauche, Groupe 2 à droite)	25
3.4	Histogrammes des âges exacts – Femmes (Groupe 1 à gauche, Groupe 2 à droite)	27
3.5	Histogrammes des âges exacts – Hommes (Groupe 1 à gauche, Groupe 2 à droite)	28
3.6	Boxplots comparés des âges exacts au décès (Groupe 1 à gauche, Groupe 2 à droite) . . .	28
3.7	Histogrammes des excès au-delà de 105 ans – Femmes (Groupe 1 à gauche, Groupe 2 à droite)	31
3.8	Découpage en 4 classes : $[0, 1[$, $[1, 2[$, $[2, 3[$, $[3, 17, 73[$, Groupe 1 (à gauche) – Groupe 2 (à droite).	33
3.9	Découpage en 5 classes : $[0, 1[$, $[1, 2[$, $[2, 3[$, $[3, 4[$, $[4, 17, 73[$, Groupe 1 (à gauche) – Groupe 2 (à droite).	33
3.10	Découpage en 8 classes : $[0-0,5[$, $[0,5-1[$, $[1-1,5[$, $[1,5-2[$, $[2-2,5[$, $[2,5-3[$, $[3-4[$, $[4-17,73[$, Groupe 1 (à gauche) – Groupe 2 (à droite).	33
3.11	QQ-plots des excès féminins vs GPD ($\xi = -0,092$, $\beta = 1,641$) pour les Groupes 1, 2 et 2 restreint	36
3.12	QQ-plots des excès masculins vs GPD ($\xi = -0,132$, $\beta = 1,291$) pour les Groupes 1, 2 et 2 restreint	36
3.13	QQ-plots : excès IDL vs GPD ($u = 105$) – Groupe 1 ; femmes à gauche, hommes à droite	38
3.14	QQ-plots : excès IDL vs GPD ($u = 105$) – Groupe 2 ; femmes à gauche, hommes à droite	39
3.15	Groupe 1 : $\xi = -0,126$, $\beta = 1,111$	41
3.16	Groupe 2 : $\xi = -0,069$, $\beta = 1,385$	41
4.1	Probabilités de décès brutes entre 0 et 99 ans : à gauche selon le sexe (femmes, hommes, mixte), à droite avec ajout de la table réglementaire XK.	44
4.2	Tables d'expérience (féminine, masculine, unisexe).	47
4.3	Évolution des réserves mathématiques pour un assuré de 65 ans selon la table utilisée. . .	49

Liste des tableaux

2.1	Paramètres estimés de la loi GPD pour un seuil commun $x^* = 101$	19
2.2	Comparaison des probabilités de décès q_x issues des différentes tables.	19
3.1	Statistiques descriptives (âges en années révolues) pour les deux groupes	24
3.2	Indicateurs par cohorte – Groupe 1	25
3.3	Indicateurs par cohorte – Groupe 2	25
3.4	Indicateurs par cohorte – Groupe 1	26
3.5	Indicateurs par cohorte – Groupe 2	26
3.6	Statistiques descriptives (âges exacts au décès) pour les deux groupes	27
3.7	Groupe 1 : échantillon restreint (206 femmes, 10 hommes)	29
3.8	Groupe 2 : échantillon élargi (556 femmes, 45 hommes)	29
3.9	Synthèse des résultats des tests du χ^2 sur les excès IDL femmes - Groupe 1	32
3.10	Synthèse des résultats des tests du χ^2 sur les excès IDL Femmes – Groupe 2	32
3.11	Synthèse des résultats des tests du χ^2 sur les excès féminins – Groupe 2 restreint (485 femmes)	34
3.12	Paramètres estimés, IC 95 % de ξ , test KS et âge ultime – seuil $u = 105$	38
3.13	Ajustement d’un modèle exponentiel au-delà de 110 ans (groupe 2)	42
4.1	Primes uniques nettes pour un capital de 1 euro, selon l’âge d’entrée et la table utilisée . .	48
A.1	Liste des doublons sur l’âge exact au décès (même durée en jours vécus)	iii
B.1	Récapitulatif des probabilités de décès brutes q_x entre 60 et 99 ans.	iv
B.2	Récapitulatif des probabilités de décès q_x des différentes tables.	v

Table des matières

Résumé	i
Abstract	ii
Remerciements	iii
Table des figures	iv
Liste des tableaux	v
Introduction	1
1 L'assurance obsèques	3
1.1 Introduction	3
1.2 Présentation du produit	4
1.2.1 La garantie	4
1.2.2 Le capital assuré	4
1.2.3 Les conditions d'adhésion	5
1.2.4 Les cotisations	5
1.2.5 Les bénéficiaires	5
1.2.6 La fiscalité	5
1.2.7 Le rachat	6
1.3 Importance de la mortalité aux âges avancés	6
1.3.1 Tarification des contrats d'assurance obsèques	6
1.3.2 La table de mortalité : un paramètre essentiel	6
1.3.3 Table réglementaire et table marché : quelle base pour la tarification ?	7
1.3.4 Enjeux de la modélisation et défis aux âges avancés	8
2 Modélisation de la mortalité aux âges extrêmes	10
2.1 Introduction	10
2.2 Cadre théorique et principes du modèle	10

2.2.1	Limites des modèles classiques de mortalité aux âges avancés	11
2.2.2	Apport de la théorie des valeurs extrêmes (EVT) dans la modélisation de la mortalité	11
2.2.3	Données et approche méthodologique	12
2.3	Méthodologie et hypothèses	13
2.3.1	Modélisation des âges extrêmes et distribution des durées de vie résiduelles	13
2.3.2	Choix du seuil pour l'application de l'EVT	15
2.3.3	Estimation des paramètres de la loi de Pareto généralisée .	16
2.4	Application du modèle et résultats de l'extrapolation	18
2.5	Conclusion	20
3	Validation du modèle GPD (2017) sur les données IDL	21
3.1	Introduction	21
3.2	Présentation des données utilisées	21
3.2.1	Base IDL et origine des données	21
3.2.2	Constitution des groupes étudiés	22
3.2.3	Analyse basée sur les âges en années révolues	24
3.2.4	Analyse basée sur les âges exacts au décès	27
3.3	Loi de survie au-delà de 105 ans	30
3.3.1	Propriété de stabilité de la GPD	30
3.3.2	Application à notre étude	30
3.4	Résumé et analyse des tests du χ^2	31
3.5	Résultats des tests de Kolmogorov–Smirnov	34
3.5.1	Principe du test	35
3.5.2	Application aux excès IDL — Femmes	35
3.5.3	Application aux excès IDL — Hommes	36
3.6	Réajustement de la GPD sur les données IDL	37
3.6.1	Réestimation des paramètres à partir du seuil de 105 ans .	38
3.6.2	Modèle unisexe : validation empirique	39
3.6.3	Modèle GPD ajusté sur les deux sexes	41
3.6.4	Exploration d'un modèle exponentiel au-delà de 110 ans .	42
4	Bases techniques de l'assurance obsèques	43
4.1	Introduction	43
4.2	Construction des tables d'expérience	43
4.2.1	Présentation des données BNB	43
4.2.2	Ajustement par le modèle de Makeham	44
4.2.3	Extrapolation au-delà de 105 ans avec la loi GPD	45

4.2.4 Raccordement entre les modèles Makeham et GPD	45
4.2.5 Présentation des tables construites	46
4.3 Illustration numérique	48
4.3.1 Hypothèses et cadre de calcul	48
4.3.2 Primes uniques et réserves mathématiques	48
Conclusion	50
Bibliographie	i
A Doublons sur l'âge exact au décès	iii
B Tables détaillées des probabilités de décès q_x	iv
B.1 Probabilités de décès brutes (BNB)	iv
B.2 Probabilités de décès issues des différentes tables	v
C Code R	vi
C.1 Analyse exploratoire des âges au décès	vi
C.2 Tests du χ^2	xi
C.3 Tests de Kolmogorov–Smirnov	xiii
C.4 Réestimation des paramètres GPD au seuil de 105 ans	xvii
C.5 Modèle unisexe : validation empirique	xxi
C.6 Modèle GPD ajusté sur les deux sexes	xxiv
C.7 Exploration d'un modèle exponentiel au-delà de 110 ans	xxvi
C.8 Présentation des données BNB	xxvii
C.9 Construction des tables d'expérience	xxix
C.10 Illustration numérique	xxxii

Introduction

Dans le domaine de l'assurance vie, l'assurance obsèques occupe une place particulière. Elle garantit, à vie, le versement d'un capital au décès de l'assuré, destiné à couvrir les frais liés à ses funérailles. Ce type de contrat engage l'assureur sur un horizon incertain, potentiellement très long. Il nécessite donc une évaluation rigoureuse du risque de décès à chaque âge, y compris aux âges les plus avancés.

C'est à ces âges extrêmes que la modélisation de la mortalité révèle des enjeux spécifiques. La raréfaction des données empiriques, l'incertitude croissante des comportements de survie et les limites des outils traditionnels complexifient l'évaluation. Les bases techniques utilisées pour tarifier et provisionner les contrats doivent pourtant rester cohérentes et robustes, même dans cette zone peu documentée. Dans ce contexte, la modélisation de la mortalité aux âges extrêmes devient un levier essentiel pour garantir la justesse des engagements pris par l'assureur.

Or, les tables de mortalité actuellement utilisées présentent des limites notables. Les tables réglementaires s'achèvent à un âge fixé par convention et intègrent des marges de sécurité. Les tables issues de l'expérience du marché, quant à elles, s'arrêtent généralement autour de 99 ans, faute de volumes suffisants. Dans les deux cas, l'absence de données fiables au-delà de ces bornes réduit leur capacité à décrire avec précision le comportement de la mortalité aux âges très avancés.

Pour combler cette lacune, plusieurs travaux ont proposé des méthodes d'extrapolation de la mortalité aux grands âges, notamment en s'appuyant sur la théorie des valeurs extrêmes. Parmi ces contributions, l'étude de Gbari et al. (2017) se distingue en introduisant un cadre particulièrement rigoureux fondé sur cette théorie. Leur approche repose sur l'ajustement d'une loi de Pareto généralisée (GPD) aux durées de vie résiduelles observées au-delà d'un seuil statistiquement déterminé. Appliquée à des données belges, elle a permis d'estimer un modèle paramétrique stable, sans effet de génération détecté, et de prolonger les tables de mortalité de manière cohérente jusqu'à un âge ultime théorique.

Notre étude s’inscrit dans la continuité de ces travaux. Elle vise à tester la validité du modèle de 2017 à partir de données plus récentes, extraites de l’International Database on Longevity (IDL). L’objectif est de déterminer si les nouveaux individus observés peuvent encore être décrits par le modèle initial, autrement dit si celui-ci reste applicable dans le temps.

Au-delà de cette validation, l’ambition est d’exploiter les résultats obtenus pour construire des tables d’expérience, adaptées aux besoins concrets des assureurs obsèques. Une illustration numérique vient éclairer cette démarche, en montrant comment ces tables peuvent renforcer la rigueur et la cohérence des bases techniques utilisées en pratique.

L’analyse proposée s’organise en quatre chapitres, allant de la compréhension du produit à la construction des tables d’expérience. Le premier présente les caractéristiques de l’assurance obsèques et les enjeux spécifiques liés à la mortalité aux grands âges. Le deuxième développe la méthode d’extrapolation fondée sur la théorie des valeurs extrêmes, telle que proposée par Gbari et al. (2017). Le troisième confronte ce modèle aux données récentes de la base IDL. Enfin, le quatrième chapitre met en œuvre cette approche pour élaborer des tables d’expérience extrapolées, et en montre l’utilité pratique pour la tarification et l’évaluation des réserves en assurance obsèques.

Chapitre 1

L'assurance obsèques

1.1 Introduction

Ce chapitre traite de l'assurance obsèques en tant qu'assurance vie entière relevant de la branche 21 en Belgique. D'un point de vue actuariel, ce contrat garantit le versement d'un capital au décès de l'assuré, contribuant ainsi à la couverture des frais funéraires. Cependant, cette approche technique ne reflète pas pleinement l'expérience du souscripteur.

Pour la clientèle, l'assurance obsèques se distingue des assurances décès classiques car elle ne se limite pas à une simple garantie financière. Elle est surtout perçue comme un service complet. Ce que recherchent les assurés, c'est une solution qui permet de tout organiser à l'avance, selon leurs souhaits. Le contrat comprend ainsi une panoplie de services, avec pour objectif de soulager les proches de toutes les charges, qu'elles soient financières, matérielles ou émotionnelles.

Dans un premier temps, les principales caractéristiques de l'assurance obsèques seront examinées, en abordant les aspects clés de son fonctionnement et ses particularités au sein de la branche 21. Par la suite, la question de la mortalité aux âges avancés, un élément fondamental pour la tarification de ce type de contrat, sera abordée. En effet, comprendre et modéliser la mortalité à ces âges est essentiel, car elle impacte directement le montant des primes et des provisions nécessaires pour assurer la viabilité du contrat sur le long terme.

Cette section présentera donc une revue succincte de la littérature sur les spécificités de la couverture obsèques et son lien avec la mortalité aux âges avancés, un facteur déterminant dans le calcul des primes relatives à cette couverture.

1.2 Présentation du produit

En Belgique, le secteur des assurances est structuré en 27 branches, dont 18 sont consacrées aux assurances non-vie et 8 aux assurances vie. Faisant partie de ces dernières, la branche 21 regroupe les assurances vie qui ne sont pas liées à des fonds d'investissement, à l'exception des contrats de nuptialité et de natalité. Cette branche comprend divers produits tels que l'assurance en cas de vie, l'assurance décès, l'assurance mixte et l'assurance de rente.

L'assurance obsèques, relevant de la branche 21, est une assurance vie-entière, régie par le droit belge. Elle est spécifiquement destinée à couvrir, en partie ou en totalité, les frais funéraires de l'assuré.

1.2.1 La garantie

Le contrat d'assurance obsèques garantit, à vie, le versement d'un capital au(x) bénéficiaire(s) désigné(s) en cas de décès de l'assuré, afin de couvrir les frais funéraires dans la limite du capital assuré.

Outre le financement, de nombreux contrats obsèques incluent des services d'accompagnement, comme le rapatriement du corps, un soutien psychologique pour les proches ou encore une aide dans les démarches administratives. Le rapatriement peut concerner le retour du corps en Belgique si le décès survient à l'étranger, par exemple pendant un voyage. Dans certains cas, le corps peut aussi être transféré vers le pays d'origine de l'assuré si les obsèques doivent s'y dérouler. Ces prestations peuvent être incluses dans l'offre de base ou proposées en option, selon les conditions propres à chaque assureur.

Malgré l'étendue de la couverture, certaines situations restent exclues, notamment le suicide dans la première année du contrat ou la participation volontaire à des actes de violence, de terrorisme ou de guerre, sauf si cela survient dans un cadre professionnel.

1.2.2 Le capital assuré

Le souscripteur a la liberté de choisir le montant du capital à assurer selon son budget ou le coût estimé des obsèques. En Belgique, les frais funéraires ou de crémation avoisinent généralement les 7 000 €, selon le site guide-épargne.be (11 décembre 2023). Par conséquent, le capital décès proposé varie entre 2 250 € et 15 000 €, selon les offres des compagnies d'assurance. Au moment du décès, ce capital est versé à la personne responsable de l'exécution ou du paiement des obsèques de l'assuré.

1.2.3 Les conditions d'adhésion

Les conditions de souscription sont spécifiques à chaque compagnie d'assurance. En général, l'assurance obsèques peut être souscrite dès l'âge de 18 ans et, dans certains cas, les enfants peuvent bénéficier de la couverture de leurs parents, bien que cela ne soit pas systématique et dépende des modalités de chaque assureur.

En ce qui concerne l'âge maximal de souscription, certaines compagnies fixent cette limite à 69 ans, tandis que d'autres acceptent des souscriptions jusqu'à 75 ans, voire même jusqu'à 99 ans.

La souscription de l'assurance obsèques peut également être conditionnée à une acceptation médicale, généralement fondée sur un examen médical ou une déclaration de santé.

1.2.4 Les cotisations

L'un des avantages majeurs de l'assurance obsèques est la couverture à vie dès le premier versement de la prime, offrant ainsi une protection immédiate. Cependant, les primes peuvent être élevées, car elles dépendent de facteurs tels que l'âge à la souscription, l'acceptation médicale, le capital-décès choisi ou encore la durée de paiement.

Le souscripteur peut opter pour un paiement unique, souvent obligatoire à partir d'un certain âge, ou alors échelonner les paiements sur une période définie. Cette période peut être limitée à un âge précis ou répartie sur un nombre d'années déterminé à l'avance.

1.2.5 Les bénéficiaires

Le preneur d'assurance désigne librement le ou les bénéficiaires qui recevront le capital assuré après son décès. Ce bénéficiaire peut être un membre de la famille, un ami ou l'entrepreneur de pompes funèbres chargé des obsèques.

1.2.6 La fiscalité

Le traitement fiscal des assurances obsèques en Belgique prévoit une taxation de 2 % sur les primes versées, applicable à toutes les primes d'assurance vie de la branche 21. Contrairement à d'autres produits de cette branche, les primes des assurances obsèques ne bénéficient d'aucun avantage fiscal particulier. En revanche, les capitaux versés au décès de l'assuré sont exonérés d'impôt sur le revenu, à l'exception d'un éventuel solde versé aux bénéficiaires, qui pourrait être soumis à imposition. Dans les faits, les dispositions

fiscales belges en matière de droits de succession s'appliquent.

1.2.7 Le rachat

Conformément aux termes du contrat, l'assuré peut, à tout moment, procéder à un rachat partiel ou total de sa police d'assurance. Pour cela, il doit adresser une demande écrite, datée et signée, à l'assureur. La valeur de rachat est calculée à la date de réception de cette demande. En cas de rachat total, le contrat est immédiatement résilié, mettant fin aux garanties prévues, et la compagnie d'assurance verse la valeur de rachat théorique, diminuée de l'indemnité de rachat.

1.3 Importance de la mortalité aux âges avancés

1.3.1 Tarification des contrats d'assurance obsèques

La tarification d'un contrat d'assurance obsèques repose sur l'évaluation des engagements réciproques entre l'assuré et l'assureur. D'un côté, l'assuré s'engage à payer des primes pour bénéficier de la couverture. De l'autre, l'assureur s'engage à verser un capital au(x) bénéficiaire(s) désigné(s) dès le décès de l'assuré, permettant ainsi de couvrir, partiellement ou totalement, les frais liés aux obsèques.

Pour déterminer la prime pure, il est essentiel de calculer la valeur actuelle moyenne des engagements pris de part et d'autre au moment de la souscription. Ce calcul repose sur deux paramètres fondamentaux : un paramètre financier, le taux d'intérêt technique, et un paramètre viager, la table de mortalité. Le taux d'intérêt technique sert à actualiser les flux financiers associés aux primes et aux prestations, tandis que la table de mortalité fournit les probabilités de décès à chaque âge, indispensables pour estimer le moment où le capital décès sera versé.

En somme, l'établissement des primes repose sur l'actualisation des engagements financiers à l'aide du taux d'intérêt technique et sur les probabilités de décès fournies par la table de mortalité. Ces deux éléments combinés garantissent une prime en adéquation avec les risques couverts, tout en assurant la capacité de l'assureur à respecter ses engagements.

1.3.2 La table de mortalité : un paramètre essentiel

La table de mortalité est un outil technique incontournable pour les assureurs vie, car elle permet de déterminer les probabilités de décès à chaque âge et de calculer les primes et provisions nécessaires. En assurance vie en Belgique, elle forme, avec le taux d'intérêt et les chargements, l'une des bases

techniques fondamentales. Partant d'un nombre donné de personnes à la naissance, la table de mortalité suit la survie de cette population au fil des années, jusqu'à un âge ultime au-delà duquel personne ne survit.

Dans le cadre des contrats d'assurance obsèques, la table de mortalité joue un rôle particulièrement crucial, puisqu'elle permet d'estimer la durée de vie restante des assurés en fonction de leur âge au moment de la souscription. Cette estimation aide l'assureur à anticiper le moment où le capital décès sera versé, et à ajuster les primes en conséquence. Ce rôle est d'autant plus déterminant que la majorité des prestations sont versées aux âges avancés, où les données sont souvent limitées.

Ainsi, la table de mortalité impacte directement la tarification en fournissant, selon le type de table utilisée (table marché de la BNB ou table réglementaire XK), des estimations plus ou moins prudentes des probabilités de survie et de décès des assurés. Elle permet ainsi de fixer des primes cohérentes avec le risque, tout en contribuant à la solidité financière de l'assureur.

1.3.3 Table réglementaire et table marché : quelle base pour la tarification ?

La table réglementaire XK : une exigence en cas de prime garantie

En assurance vie de la branche 21, le choix de la table de mortalité dépend du type d'engagement pris par l'assureur envers l'assuré. Si la prime est garantie dès la souscription, l'assureur doit utiliser une table réglementaire fixée par les autorités. Dans le cas de l'assurance obsèques, il s'agit de la table XK. Élaborée de manière prudente, cette table fixe des niveaux minimaux de primes et de provisions afin d'assurer une marge de sécurité suffisante.

Cependant, cette prudence marquée, en particulier aux âges avancés, peut rendre la table XK moins adaptée lorsqu'il s'agit d'analyser plus finement le passif et d'estimer le coût réel de la couverture. C'est pourquoi, même lorsqu'elle est utilisée, il peut être utile de la comparer à des tables fondées sur l'expérience observée sur le marché.

Les tables marché : une alternative encadrée

Certains assureurs choisissent ainsi d'utiliser une table marché, construite à partir des données publiées par la Banque nationale de Belgique (BNB). Ces tables, déduites des feuilles Excel disponibles sur le site de la BNB, sont établies à partir des observations de mortalité des cinq dernières années sur un portefeuille représentatif des affaires directes en assurance vie en Belgique, déclaré par les assureurs soumis au contrôle de la BNB. Elles permettent de

formuler des hypothèses plus proches de la réalité, mais leur utilisation reste encadrée par la réglementation.

En effet, selon l'arrêté royal relatif à l'activité d'assurance sur la vie (Moniteur belge du 15 novembre 2003), un assureur ne peut garantir des taux de mortalité ou de survie inférieurs à ceux des tables officielles que pour une durée maximale de trois ans. Autrement dit, une table marché peut être utilisée pour la tarification, mais la prime ne peut alors être garantie que temporairement. Cette contrainte limite leur usage dans les contrats à long terme, sauf si l'assureur prévoit explicitement la possibilité de réviser les conditions au fil du temps.

En pratique, ces tables présentent également une limite technique : elles s'arrêtent souvent à un âge relativement bas (99 ans), ce qui les rend insuffisantes pour estimer la mortalité aux âges très avancés. Or, c'est précisément à ces âges que la plupart des prestations des contrats obsèques sont versées. Les statistiques marché, bien qu'intéressantes, ne permettent donc pas à elles seules de couvrir l'ensemble de la période pertinente.

Un défi commun : la fermeture des tables aux âges extrêmes

Ce manque de données au-delà d'un certain âge, connu sous le nom de fermeture de la table, impose le recours à une extrapolation des taux de mortalité. Il s'agit de prolonger la table au-delà de sa borne supérieure de manière cohérente. Ce besoin se pose quel que soit le cas de figure : même lorsqu'on se base sur la table réglementaire XK, une analyse plus fine du passif peut nécessiter des hypothèses moins prudentes et plus réalistes ; et lorsqu'on utilise la table BNB, il devient indispensable de la prolonger pour couvrir les âges extrêmes.

1.3.4 Enjeux de la modélisation et défis aux âges avancés

Pour les assureurs, l'estimation précise de la mortalité aux âges avancés est capitale, notamment pour les produits comme les rentes viagères et l'assurance obsèques. En effet, les contrats obsèques sont conçus pour couvrir les frais funéraires, un besoin qui se manifeste surtout aux âges élevés. Cependant, la rareté et la qualité variable des données pour ces tranches d'âge rendent cette estimation complexe, nécessitant souvent une extrapolation des tables de mortalité au-delà des âges habituellement observés. Ce procédé, connu sous le nom de « fermeture » de la table de mortalité, est indispensable aux calculs actuariels car il permet d'établir des projections fiables même en l'absence de données suffisantes.

Dans ce contexte, les assureurs recourent à des modèles qui prolongent

les estimations de mortalité de manière réaliste, en tenant compte des particularités de la survie aux âges extrêmes. L'importance de la mortalité aux âges avancés dans la tarification des assurances obsèques souligne ainsi la nécessité de modèles précis pour estimer les probabilités de décès au-delà des âges habituels. Une modélisation adaptée garantit non seulement la viabilité financière des contrats, mais assure également que les primes reflètent les risques réels associés à ces âges avancés.

Dans le chapitre suivant, l'étude de Gbari et al. sera présentée en détail, apportant une approche innovante pour modéliser la fin de la table de mortalité. Cette étude permet d'approfondir la compréhension de la mortalité aux âges extrêmes et d'affiner les projections, contribuant ainsi à des estimations plus robustes pour les assurances obsèques et une tarification plus adéquate pour ces risques spécifiques.

Chapitre 2

Modélisation de la mortalité aux âges extrêmes : Approche de Gbari et al. (2017)

2.1 Introduction

L'extrapolation de la mortalité aux âges extrêmes représente un défi majeur en assurance vie, en particulier pour la tarification des contrats d'assurance obsèques. La rareté des données empiriques au-delà d'un certain âge rend difficile l'estimation des probabilités de décès, nécessitant ainsi des modèles adaptés pour prolonger les tables de mortalité classiques.

L'étude menée par Gbari et al. (2017) propose une approche fondée sur la théorie des valeurs extrêmes (EVT) afin de modéliser la distribution des âges au décès au-delà d'un seuil critique. Cette approche permet d'affiner les estimations de mortalité aux âges avancés et constitue un cadre méthodologique pertinent pour les analyses actuarielles.

Ce chapitre a pour objectif de décrire le modèle de Gbari et al. (2017) en mettant en lumière ses fondements théoriques et sa méthodologie. Après une présentation du cadre théorique et des principes du modèle, les hypothèses clés et la méthodologie adoptée seront détaillées.

2.2 Cadre théorique et principes du modèle

L'approche développée par Gbari et al. (2017) s'appuie sur la théorie des valeurs extrêmes (Extreme Value Theory, EVT), un cadre statistique permettant d'analyser le comportement des valeurs situées dans la queue d'une distribution. En actuariat, l'EVT est couramment utilisée pour modéliser des

événements rares, comme les sinistres extrêmes en assurance dommages ou les risques financiers extrêmes. Son application à la mortalité vise à extrapoler les probabilités de décès aux âges avancés, là où les données empiriques deviennent insuffisantes pour une estimation directe.

2.2.1 Limites des modèles classiques de mortalité aux âges avancés

Historiquement, la modélisation de la mortalité repose sur des modèles paramétriques tels que ceux de Gompertz (1825) et Makeham (1860). Le modèle de Gompertz postule une croissance exponentielle du taux de mortalité avec l'âge, ce qui s'avère pertinent pour les adultes et les personnes âgées. Cependant, plusieurs études empiriques ont montré que cette hypothèse devient moins réaliste aux âges extrêmes, où l'on observe souvent un ralentissement de l'augmentation du taux de mortalité, voire un plateau.

Ce phénomène, qualifié de décélération de la mortalité, est attribué à une sélection naturelle progressive : les individus les plus robustes d'une cohorte survivent plus longtemps, créant une hétérogénéité dans la mortalité observée.

2.2.2 Apport de la théorie des valeurs extrêmes (EVT) dans la modélisation de la mortalité

Face aux limites des modèles classiques, l'EVT offre une approche plus flexible en s'intéressant spécifiquement aux valeurs extrêmes. En appliquant cette théorie à la mortalité, l'objectif est de modéliser la queue de distribution des âges au décès, permettant ainsi d'estimer les probabilités de décès pour les âges les plus avancés.

Dans ce cadre, Gbari et al. (2017) utilisent la méthode des excès au-delà d'un seuil (Peaks Over Threshold, POT), qui consiste à sélectionner un seuil au-delà duquel la distribution des âges au décès est approximée par une loi de Pareto généralisée (Generalized Pareto Distribution, GPD). Cette approche repose sur trois principes clés :

- Adéquation aux valeurs extrêmes : selon l'EVT, lorsque l'on analyse les âges au décès dépassant un certain seuil, la distribution des excédents converge asymptotiquement vers une loi limite décrite par la GPD, permettant ainsi une extrapolation robuste.
- Estimation ciblée sur les valeurs extrêmes : plutôt que d'imposer une structure rigide sur l'ensemble de la distribution des âges au décès, la loi de Pareto généralisée s'ajuste uniquement aux observations les plus élevées, réduisant ainsi les biais des modèles traditionnels.

- Extrapolation probabiliste au-delà des données observées : L'EVT ne se limite pas aux observations disponibles, mais repose sur un cadre probabiliste décrivant le comportement des valeurs extrêmes. Cela permet d'extrapoler les taux de mortalité aux âges avancés en s'appuyant sur des fondements théoriques solides.

2.2.3 Données et approche méthodologique

L'étude repose sur des statistiques de mortalité en cohortes complètes, exploitant des données issues du Registre National de Belgique. Celles-ci incluent les âges exacts au décès des individus ayant atteint 95 ans et plus, couvrant les cohortes nées entre 1886 et 1904. Cette sélection permet une modélisation plus fine et réduit l'impact des biais d'agrégation liés aux données par tranche d'âge.

Le choix du seuil de sélection des données est une étape clé dans l'application de l'EVT. Dans cette étude, les auteurs sélectionnent un seuil optimal à l'aide de plusieurs méthodes statistiques (Pickands, Reiss-Thomas, modèle de mélange) afin d'identifier la zone où la loi de Pareto généralisée constitue une approximation adéquate de la distribution des âges au décès. Ce choix de seuil permet d'isoler les valeurs les plus élevées, où l'hypothèse de convergence vers une loi de Pareto généralisée est testée et validée empiriquement.

Enfin, la méthode utilisée repose sur une estimation par maximum de vraisemblance (ML) des paramètres de la distribution des âges extrêmes. Toutefois, d'autres estimateurs (moments, moments pondérés, Hill) sont également testés afin d'évaluer la robustesse des résultats. Ces paramètres permettent d'ajuster la loi de Pareto généralisée aux données empiriques et d'extrapoler la mortalité aux âges encore plus avancés, offrant ainsi une base solide pour des projections actuarielles.

En somme, l'étude de Gbari et al. (2017) s'appuie donc sur la théorie des valeurs extrêmes pour proposer une alternative aux modèles classiques de mortalité aux âges avancés. Grâce à l'utilisation de la méthode POT et de la loi de Pareto généralisée, elle permet une extrapolation des taux de mortalité fondée sur un cadre probabiliste solide.

Cependant, la pertinence de cette approche dépend des choix méthodologiques effectués, en particulier la sélection du seuil et l'estimation des paramètres, qui impactent directement la fiabilité des projections. Le point suivant détaillera ces aspects ainsi que la méthodologie adoptée pour estimer les paramètres du modèle et valider son adéquation aux données étudiées.

2.3 Méthodologie et hypothèses

2.3.1 Modélisation des âges extrêmes et distribution des durées de vie résiduelles

L'approche de Gbari et al. (2017) repose sur la théorie des valeurs extrêmes (EVT) pour modéliser la mortalité aux âges avancés. Elle vise à capter le comportement des âges extrêmes et à extrapoler la mortalité au-delà des observations. Deux aspects clés sont considérés : la distribution des âges extrêmes et celle des durées de vie résiduelles.

Distribution des âges extrêmes et convergence vers une loi limite

Les durées de vie totales sont modélisées comme une séquence de variables aléatoires indépendantes $T_1, T_2, \dots, T_n$. Pour étudier les âges les plus avancés, on considère le maximum des durées de vie observées dans un échantillon de taille n :

$$M_n = \max(T_1, T_2, \dots, T_n).$$

L'EVT établit que, sous certaines conditions, ce maximum, après normalisation, converge en distribution vers une loi limite :

$$\frac{M_n - b_n}{a_n} \xrightarrow[n \rightarrow \infty]{d} H_\xi(x), \quad \text{avec } a_n > 0 \text{ et } b_n \in \mathbb{R},$$

où $H_\xi(x)$ est la fonction de répartition associée à une loi de valeurs extrêmes généralisée (GEV) donnée par :

$$H_\xi(x) = \begin{cases} \exp\left(-(1 + \xi x)_+^{-1/\xi}\right), & \text{si } \xi \neq 0, \\ \exp(-\exp(-x)), & \text{si } \xi = 0. \end{cases}$$

Le paramètre ξ joue un rôle clé dans la modélisation de la distribution des âges extrêmes :

- Si $\xi > 0$: queue lourde (loi de Fréchet), impliquant une diminution du risque de décès avec l'âge, ce qui contredit les observations empiriques en mortalité humaine.
- Si $\xi = 0$: queue exponentielle (loi de Gumbel), caractérisant un taux de mortalité constant.
- Si $\xi < 0$: queue bornée (loi de Weibull), ce qui signifie qu'il existe un âge maximal ω fixant une limite supérieure à la longévité humaine. Les

études empiriques confirment généralement cette hypothèse, indiquant une espérance de vie finie.

Distribution des durées de vie résiduelles et lien avec la GPD

L'objectif est d'estimer la probabilité de survie d'un individu ayant déjà atteint un âge avancé x , ce qui revient à modéliser la durée de vie résiduelle $T - x$ conditionnellement à $T > x$. Cette fonction de répartition correspond aux quotients de mortalité, notés :

$${}_s q_x = P(T - x \leq s \mid T > x).$$

La loi de Pareto généralisée (GPD) qui en résulte est définie par :

$$G_\xi(s) = \begin{cases} 1 - (1 + \xi s)_+^{-1/\xi}, & \text{si } \xi \neq 0, \\ 1 - \exp(-s), & \text{si } \xi = 0. \end{cases}$$

Pour tenir compte de la dispersion des durées de vie résiduelles, on généralise la loi en introduisant le paramètre $\beta > 0$:

$$G_{\xi,\beta}(s) = G_\xi\left(\frac{s}{\beta}\right).$$

Cette loi est paramétrée par l'indice de queue ξ et un paramètre d'échelle β , qui décrit l'évolution de la mortalité aux âges extrêmes et constitue la base de l'extrapolation effectuée par Gbari et al. (2017).

D'après le théorème de Pickands–Balkema–de Haan, lorsque x est suffisamment grand, la distribution des excès au-delà d'un seuil x suit approximativement une loi de Pareto généralisée (GPD). Ce résultat justifie que, pour les âges avancés, la distribution des durées de vie résiduelles peut être bien approximée par une GPD, comme établi par la relation suivante :

$$\lim_{x \rightarrow +\infty} \sup_{s \geq 0} |{}_s q_x - G_{\xi,\beta(x)}(s)| = 0.$$

Le lien entre la loi des valeurs extrêmes et la distribution des durées de vie résiduelles permet donc d'estimer les taux de mortalité au-delà des observations disponibles. En particulier, si $\xi < 0$, l'existence d'un âge maximal ω est assurée et peut être estimé à partir des données. Gbari et al. (2017) exploitent cette propriété pour définir un cadre d'extrapolation cohérent avec les observations empiriques.

Cette première étape pose les bases du modèle en caractérisant le comportement des âges extrêmes. L'étape suivante consiste à déterminer un seuil

optimal à partir duquel la distribution des excès peut être approximée par une loi de Pareto généralisée, ce qui sera abordé dans la section suivante.

2.3.2 Choix du seuil pour l'application de l'EVT

L'identification d'un seuil optimal est une étape clé dans l'application de la théorie des valeurs extrêmes (EVT). Ce seuil délimite la région où les observations sont suffisamment extrêmes pour être bien décrites par une loi de Pareto généralisée (GPD). Un seuil trop bas inclurait des valeurs qui ne suivent pas le comportement asymptotique attendu, tandis qu'un seuil trop élevé réduirait le nombre d'observations disponibles, augmentant l'incertitude des estimations.

Méthodes de sélection du seuil

Pour déterminer le seuil optimal, Gbari et al. (2017) ont testé plusieurs méthodes statistiques :

— Méthode graphique par espérance de vie restante

Cette approche analyse la fonction d'excès moyen, qui mesure l'espérance de vie restante après un âge donné x . Si la distribution suit bien une GPD, cette fonction devient approximativement linéaire à partir d'un certain seuil, facilitant son identification visuelle.

$$\hat{e}_x = \frac{\sum_{i=1}^n (t_i - x) I[t_i > x]}{\#\{t_i : t_i > x\}}. \quad (2.1)$$

Ici :

- $t_1, t_2, \dots, t_n$ désignent les âges au décès observés dans l'échantillon ;
- x est l'âge à partir duquel on mesure l'espérance de vie restante ;
- L'indicatrice $I[t_i > x]$ vaut 1 si l'individu i est décédé après l'âge x , et 0 sinon.

Un tracé linéaire décroissant de $\hat{e}_x$ indique une queue bornée ($\xi < 0$), ce qui est cohérent avec la mortalité humaine. À l'inverse, une fonction constante correspondrait à un taux de mortalité constant, caractéristique du cas $\xi = 0$.

Outre cette méthode, les auteurs utilisent également des techniques, telles que :

- l'**estimateur de Pickands** (1975), basé sur la comparaison entre la fonction de survie empirique et celle de la GPD ;
- la **méthode de Reiss-Thomas** (1997), qui sélectionne le seuil en minimisant la variance des estimations du paramètre de forme ξ ;

- le **modèle flexible**, combinant une estimation non paramétrique avant le seuil et une GPD au-delà.

Seuil retenu

Chaque méthode ayant produit un seuil différent, Gbari et al.(2017) ont retenu, pour chaque sexe, la valeur la plus élevée parmi celles obtenues :

$$\mathbf{x}^* = \max(\mathbf{x}_{\text{Pickands}}^*, \mathbf{x}_{\text{Reiss-Thomas}}^*, \mathbf{x}_{\text{modèle flexible}}^*). \quad (2.2)$$

Ce choix est justifié par la **propriété de stabilité** de la loi de Pareto généralisée. En effet, si $\mathbf{X} \sim \text{GPD}(\xi, \beta)$, alors tout excès au-delà d'un seuil $\mathbf{u} > 0$, sachant que ce niveau a été atteint, suit également une GPD avec les mêmes paramètres de forme :

$$\mathbf{X} - \mathbf{u} \mid \mathbf{X} > \mathbf{u} \sim \text{GPD}(\xi, \beta + \xi \mathbf{u}). \quad (2.3)$$

L'indice de queue ξ restant inchangé, cette propriété garantit la cohérence de la modélisation, quel que soit le seuil retenu. Ainsi, en sélectionnant le seuil le plus élevé parmi ceux obtenus par les différentes méthodes, on s'assure que la distribution des excès conserve la structure asymptotique imposée par l'EVT.

Ce seuil constitue le point à partir duquel la modélisation par EVT est appliquée et servira à l'estimation des paramètres de la loi de Pareto généralisée dans la suite de l'étude.

2.3.3 Estimation des paramètres de la loi de Pareto généralisée

L'extrapolation de la mortalité aux âges avancés repose sur l'estimation des paramètres de la loi de Pareto généralisée (GPD), notamment l'indice de queue ξ et le paramètre d'échelle β . Une fois le seuil $\mathbf{x}^*$ déterminé, ces paramètres sont estimés à partir des durées de vie restantes observées au-delà de ce seuil.

Estimation par Maximum de Vraisemblance (ML)

Gbari et al. (2017) estiment ξ et β par maximisation de la vraisemblance des durées de vie résiduelles au-delà de $\mathbf{x}^*$:

$$\mathcal{L}(\xi, \beta) = \prod_{i \mid t_i > \mathbf{x}^*} \left(\frac{1}{\beta} \left(1 + \frac{\xi}{\beta} (t_i - \mathbf{x}^*) \right)^{-\frac{1}{\xi} - 1} \right).$$

et sa version logarithmique :

$$L(\xi, \beta) = -\ln \beta L_{x^*} - \left(1 + \frac{1}{\xi}\right) \sum_{i|t_i > x^*} \ln \left(1 + \frac{\xi}{\beta}(t_i - x^*)\right),$$

où L_{x^*} représente le nombre d'individus survivants à x^* . L'optimisation de cette fonction permet d'obtenir les estimateurs $\hat{\xi}_{ML}$ et $\hat{\beta}_{ML}$ à l'aide d'algorithmes numériques.

Les auteurs proposent une initialisation des paramètres à l'aide de la méthode des moments, donnée par :

$$\hat{\xi}_{MOM} = \frac{1}{2} \left(1 - \frac{\bar{x}^2}{s^2}\right), \quad \hat{\beta}_{MOM} = \frac{1}{2} \bar{x} \left(\frac{\bar{x}^2}{s^2} + 1\right),$$

où $\bar{x}$ et s^2 sont la moyenne et la variance des durées de vie résiduelles $t_i - x^*$.

Une autre méthode consiste à ajuster une régression linéaire sur les espérances de vie restante observées.

Validation et interprétation des résultats

Une fois les paramètres estimés, plusieurs tests permettent d'évaluer la qualité de l'ajustement :

- **Visualisation graphique** : les histogrammes et les Q-Q plots confirment un bon ajustement de la distribution GPD aux données observées.
- **Test de Kolmogorov-Smirnov (KS)** : les p-values obtenues sont élevées ($p > 0.5$), confirmant l'adéquation du modèle.

Les résultats indiquent un indice de queue négatif $\hat{\xi} < 0$, validant ainsi l'hypothèse d'un âge maximal ω .

Impact sur l'extrapolation de la mortalité

L'estimation des paramètres ξ et β permet d'extrapoler la mortalité au-delà des âges observés. Plus particulièrement, lorsque $\xi < 0$, un âge maximal théorique ω peut être estimé par :

$$\hat{\omega} = x^* - \frac{\hat{\beta}}{\hat{\xi}}.$$

Ce paramètre est essentiel pour prolonger les tables de mortalité et estimer les probabilités de décès aux âges extrêmes.

Comparaison avec d'autres méthodes d'estimation

Gbari et al. (2017) évaluent la robustesse de leurs résultats en comparant l'estimateur du maximum de vraisemblance (ML) à :

- **Estimateur des moments (MOM)** : basé sur l'égalité entre moments empiriques et théoriques de la GPD.
- **Estimateur des moments pondérés (PMOM)** : proposé par *Beirlant et al. (2005a)*, affine MOM en pondérant les moments pour améliorer la précision.
- **Estimateur de Hill étendu (H)** : proposé par *Beirlant et al. (2005b)*, mieux adapté aux distributions bornées ($\xi < 0$).
- **Estimateur de type moments (M)** : proposé par *Dekkers et al. (1989)*, variante améliorée du MOM pour limiter le biais.

Gbari et al. (2017) ne privilégient pas un estimateur unique mais soulignent que le maximum de vraisemblance (ML) est largement utilisé en actuariat pour sa cohérence statistique. Toutes ces méthodes confirment $\xi < 0$, validant ainsi l'existence d'un âge maximal ω , bien que sa valeur exacte varie selon l'estimateur employé.

2.4 Application du modèle et résultats de l'extrapolation

L'approche de Gbari et al. (2017), fondée sur l'ajustement d'une loi de Pareto généralisée (GPD) aux âges avancés, a été appliquée aux données belges dans le but d'extrapoler la mortalité au-delà des âges où les effectifs deviennent trop faibles pour garantir des estimations fiables. Les seuils initiaux, estimés à 98,89 ans pour les hommes et 100,89 ans pour les femmes, ont été déterminés à l'aide de procédures automatiques de sélection (Pickands, Reiss-Thomas), puis arrondis respectivement à 99 et 101 ans dans la note technique de DetraLytics (2017), destinée à un public de praticiens.

Cette démarche s'inscrit dans une volonté de dépasser les limites des tables réglementaires existantes, notamment la table XK. Cette dernière, utilisée pour les opérations en cas de décès, s'arrête à 106 ans et présente des irrégularités marquées autour de 105 ans. Ces limites s'expliquent par les faibles effectifs de survivants à ces âges et les arrondis opérés sur les effectifs publiés. En outre, la table incorpore un important chargement de sécurité, ce qui tend à surestimer les taux de mortalité aux âges extrêmes.

Pour harmoniser l'extrapolation entre les sexes et faciliter son application, un seuil commun de 101 ans a été retenu, en s'appuyant sur la propriété de

stabilité de la loi GPD.

Les paramètres estimés à partir de ce seuil sont présentés dans le tableau suivant.

Paramètres	x^*	$\hat{\xi}$	s.e($\hat{\xi}$)	IC _{95%} ξ	$\hat{\beta}$	s.e($\hat{\beta}$)	$\frac{-\hat{\beta}}{\hat{\xi}}$	$\hat{\omega}$
Hommes	101	-0,13	0,02	[-0,16; -0,10]	1,82	0,04	13,71	114,71
Femmes	101	-0,09	0,01	[-0,12; -0,07]	2,01	0,04	21,57	122,57

TABLE 2.1 – Paramètres estimés de la loi GPD pour un seuil commun $x^* = 101$.

Sur cette base, des quotients de mortalité ont été extrapolés au-delà des âges observés. Ils ont permis la construction de deux tables distinctes : GP2D-H pour les hommes et GP2D-F pour les femmes, acronyme formé d’après les noms des auteurs Gbari, Poulain, Dal et Denuit. Ces tables sont comparées dans le tableau suivant à la table réglementaire XK (disponible jusqu’à 106 ans) et aux tables brutes issues de l’expérience belge observée (Brut-H et Brut-F).

x	ℓ_x XK	XK	Brut-H	GP2D-H	Brut-F	GP2D-F
95	10 942	0,386	0,314	–	0,270	–
96	6 723	0,418	0,334	–	0,286	–
97	3 910	0,453	0,345	–	0,302	–
98	2 139	0,489	0,388	–	0,316	–
99	1 093	0,527	0,376	–	0,341	–
100	517	0,567	0,406	–	0,352	–
101	224	0,603	0,451	0,434	0,388	0,395
102	89	0,652	0,493	0,460	0,422	0,413
103	31	0,677	0,540	0,489	0,431	0,431
104	10	0,700	0,517	0,521	0,478	0,452
105	3	0,667	0,548	0,558	0,458	0,475
106	1	1,000	0,737	0,600	0,457	0,499
107	0	–	0,400	0,648	0,492	0,527
108	0	–	0,333	0,703	0,508	0,557
109	0	–	0,000	0,765	0,500	0,590
110	0	–	0,500	0,834	0,667	0,628

TABLE 2.2 – Comparaison des probabilités de décès q_x issues des différentes tables.

Une irrégularité marquée apparaît dans la table réglementaire XK autour des âges 104–105. La probabilité de décès y décroît entre 104 ans ($q_{104}^{XK} = 0,700$) et 105 ans ($q_{105}^{XK} = 0,667$), ce qui contredit la progression attendue de la mortalité aux âges élevés. Cette anomalie provient des arrondis appliqués aux valeurs publiées dans l’arrêté royal. Si ces arrondis sont négligeables à des âges jeunes, ils peuvent engendrer des distorsions importantes lorsque les effectifs deviennent faibles (par exemple, $\ell_{105}^{XK} = 3$ et $\ell_{106}^{XK} = 1$).

Ainsi, bien que les tables XK soient issues des lois de Makeham (MK pour les hommes, FK pour les femmes), qui assurent théoriquement une progression régulière, l’application d’arrondis réglementaires introduit ici un comportement erratique.

Cette analyse des résultats met en évidence l'intérêt de l'extrapolation par la GPD. Contrairement à la table XK, qui tend à majorer les probabilités aux âges les plus avancés, les tables GP2D fournissent des estimations plus cohérentes avec les observations empiriques. Elles s'inscrivent dans les intervalles de confiance des données brutes, renforçant ainsi leur crédibilité statistique, tout en assurant une progression lisse jusqu'à un âge limite théorique.

Enfin, cette extrapolation constitue une base solide pour la tarification des contrats d'assurance obsèques, qui nécessitent une évaluation fiable de la mortalité aux âges très élevés. Elle permet de pallier les biais liés aux arrondis ou à l'absence de données dans les tables classiques, et d'assurer une meilleure adéquation entre tarification technique et réalité démographique.

2.5 Conclusion

L'application de la théorie des valeurs extrêmes (EVT) à la mortalité permet de pallier les limites des modèles traditionnels en offrant une extrapolation plus robuste des taux de mortalité aux âges avancés. L'approche retenue repose sur l'utilisation de la loi de Pareto généralisée (GPD), nécessitant l'identification d'un seuil optimal et l'estimation de ses paramètres pour modéliser la queue de la distribution des âges au décès.

L'étude de Gbari et al. (2017) adopte une approche rigoureuse fondée sur l'EVT et confirme l'existence d'un âge ultime, dont la valeur dépend de la méthode d'estimation employée. Les résultats indiquent un indice de queue négatif ($\xi < 0$), suggérant une limite supérieure à la longévité humaine. L'âge ultime estimé est de 114,82 ans pour les hommes et 122,73 ans pour les femmes, des valeurs cohérentes avec les âges les plus élevés observés dans leur base de données et proches des records mondiaux de longévité.

Par ailleurs, cet âge ultime ne constitue pas une limite biologique stricte, mais représente un outil opérationnel en actuariat, notamment pour la construction des tables de mortalité et l'évaluation du risque de longévité. Son estimation permet d'encadrer la fermeture des tables, bien que certaines approches supposent une mortalité constante au-delà d'un certain âge.

Dans le chapitre suivant, nous testerons la robustesse de l'approche de Gbari et al. (2017) en confrontant leur modèle aux données belges issues de la *International Database on Longevity* (IDL). Cette validation permettra d'évaluer si un recalibrage est nécessaire pour mieux s'adapter aux nouvelles données.

Chapitre 3

Validation du modèle GPD (2017) sur les nouvelles données IDL

3.1 Introduction

Le chapitre précédent a présenté l’approche de Gbari et al. (2017) pour prolonger les tables de mortalité aux âges très avancés à l’aide de la loi de Pareto généralisée (GPD). Ce modèle estime les décès au-delà d’un certain seuil, afin de couvrir les âges extrêmes de manière cohérente.

Dans ce chapitre, nous testons la robustesse de cette approche à partir d’un jeu de données indépendant : les données belges de la *International Database on Longevity* (IDL), mises à jour en 2022. L’objectif est d’évaluer si ces nouvelles observations sont compatibles avec le modèle établi en 2017, sans réestimer les paramètres dans un premier temps.

Cette validation externe est essentielle. Elle permet de savoir si les paramètres de l’étude initiale peuvent encore être utilisés pour projeter ou tarifier, ou s’il faut ajuster le modèle. Pour cela, nous appliquons des tests statistiques sur les excès observés, comparons les résultats aux quantiles empiriques, et analysons la cohérence globale avec les données actuelles.

La première section décrit brièvement les données et les hypothèses retenues. Les sections suivantes présentent la méthodologie utilisée, puis les résultats obtenus pour les hommes et pour les femmes.

3.2 Présentation des données utilisées

3.2.1 Base IDL et origine des données

L’*International Database on Longevity* (IDL) est une base internationale de référence développée pour l’étude de la longévité extrême. Elle rassemble

des données individuelles validées sur des personnes ayant atteint un âge très avancé (au moins 105 ans), dans plusieurs pays, dont la Belgique.

Les données utilisées dans cette étude proviennent de la dernière mise à jour disponible pour la Belgique, datée du 4 avril 2022. Elles sont extraites du Registre National, un registre centralisé, informatisé et continuellement mis à jour, qui collecte les événements démographiques majeurs (naissances, décès, mariages, migrations, etc.) enregistrés par les communes belges. Ce registre est reconnu pour son haut niveau de fiabilité, notamment pour les événements vitaux tels que les naissances et les décès.

Dans le cas belge, la base IDL se compose initialement d'un *fichier brut unique*, contenant l'ensemble des individus validés. Ce fichier a ensuite été ventilé en deux volets distincts dans la présentation officielle :

- une liste de **975 individus** décédés entre 105 et 109 ans, nés entre 1891 et 1915, sans restriction sur le lieu de naissance, et décédés en Belgique entre 1996 et 2021 ;
- une liste de **27 individus** décédés à 110 ans ou plus, nés en Belgique entre 1892 et 1910, et également décédés en Belgique entre 2003 et 2021.

Dans le cadre de cette étude, l'analyse repose directement sur le *fichier brut*, qui agrège les deux volets décrits précédemment. Ce fichier contient plusieurs variables individuelles (sexe, lieu de naissance et de décès, année et date complète de naissance et de décès lorsque disponibles), offrant ainsi une base riche pour étudier la mortalité à des âges très avancés.

3.2.2 Constitution des groupes étudiés

La version brute du fichier IDL utilisée pour cette étude comptait **1 002 individus** âgés d'au moins 105 ans au moment de leur décès. Afin d'évaluer la robustesse de nos résultats, deux échantillons distincts ont été constitués.

Groupe 1 : échantillon restreint (206 femmes, 10 hommes). Ce premier groupe reprend les critères de sélection stricts adoptés dans l'étude de Gbari et al. (2017) afin d'en permettre une comparaison directe. Deux filtres ont donc été appliqués :

- exclusion des individus **nés entre 1886 et 1904**, déjà inclus dans l'analyse de 2017 ;
- exclusion de toutes les personnes **nées hors de Belgique**, afin de garantir la cohérence géographique.

L'échantillon ainsi obtenu comprend 216 observations, réparties entre 206 femmes et 10 hommes.

Groupe 2 : échantillon élargi (556 femmes, 45 hommes). Le second groupe reprend uniquement le premier filtre, à savoir l'exclusion des individus nés entre 1886 et 1904, mais conserve tous les individus des autres cohortes, indépendamment du lieu de naissance. Cette approche repose sur le fait que les données de l'IDL ont été validées par des démographes, et permet d'intégrer notamment les générations 1911-1915, qui étaient absentes du groupe 1 en raison d'un manque d'information géographique.

Le groupe ainsi obtenu constitue un échantillon élargi, composé de 601 individus, dont 556 femmes et 45 hommes.

Absence des cohortes récentes (1911-1915) dans le groupe 1

Dans le fichier original de la base IDL, plusieurs individus ayant atteint l'âge de 105 ans appartiennent aux cohortes de naissance 1911-1915. Toutefois, après application des critères de sélection décrits précédemment, aucun d'entre eux n'a été retenu dans l'échantillon final. Cette exclusion s'explique principalement par deux raisons : d'une part, certains de ces individus sont nés hors de Belgique, ce qui les rend incompatibles avec notre périmètre d'analyse ; d'autre part, un grand nombre d'observations ne disposent d'aucune information sur le pays de naissance et de décès.

Au total, on recense **376 individus nés entre 1905 et 1915** pour lesquels le **pays de naissance et le pays de décès est manquant**, ce qui rend leur utilisation impossible dans le cadre d'une sélection cohérente avec l'étude de référence. Par conséquent, bien que les cohortes 1911-1915 soient représentées dans la base brute par des survivants à 105 ans et plus, elles apparaissent vides dans nos tableaux filtrés, non pas en raison d'un manque de données, mais à cause du filtrage méthodologique strict appliqué. Cette étape s'avère indispensable pour garantir la comparabilité avec l'étude de référence et la fiabilité des analyses.

Deux niveaux d'analyse des âges au décès

Les données disponibles permettent de mener deux niveaux d'analyse :

- une première approche, fondée sur les âges au décès exprimés en **années révolues**, tels que fournis directement dans la base ;
- une seconde approche, plus fine, reposant sur le calcul des **âges exacts en années décimales**, rendu possible par la disponibilité des dates complètes de naissance et de décès pour l'ensemble des individus retenus.

Cette double lecture permet d'approfondir l'étude de la mortalité aux âges extrêmes et de tester la robustesse des résultats en fonction du degré de précision des données utilisées.

3.2.3 Analyse basée sur les âges en années révolues

Nous avons d’abord analysé les âges au décès tels que fournis dans la base IDL, exprimés en **années révolues**. Cette approche permet une première lecture synthétique des données, sans retraitement des dates pour les deux groupes étudiés.

Les tableaux ci-dessous présentent un résumé statistique par sexe pour les échantillons retenus.

TABLE 3.1 – Statistiques descriptives (âges en années révolues) pour les deux groupes

Groupe 1 (échantillon restreint)								
Sexe	n	Min	Moy.	Méd.	Q ₁	Q ₃	Max	Sd
F	206	105	105,59	105	105	106,00	109	0,87
H	10	105	105,50	105	105	105,75	108	0,97

Groupe 2 (échantillon élargi)								
Sexe	n	Min	Moy.	Méd.	Q ₁	Q ₃	Max	Sd
F	556	105	105,86	106	105	106	112	1,18
H	45	105	105,73	105	105	106	109	1,03

La distribution des âges au décès est illustrée ci-dessous à l’aide d’histogrammes et de boxplots, séparément pour les femmes et pour les hommes. Pour chaque sexe, les résultats sont présentés côte à côte afin de comparer visuellement les deux groupes d’étude.

On observe dans les deux groupes une forte concentration autour de 105 ans, seuil minimal d’entrée dans la base IDL. Chez les femmes, la dispersion est systématiquement plus élevée, en particulier dans le groupe élargi, où les âges peuvent atteindre 112 ans. Cette variabilité accrue reflète la surreprésentation féminine parmi les personnes les plus âgées. Chez les hommes, les distributions restent plus resserrées autour de la moyenne.

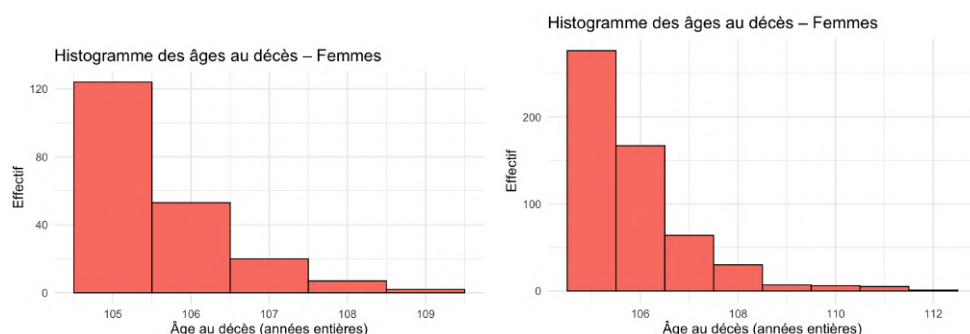


FIGURE 3.1 – Histogrammes des âges au décès – Femmes (Groupe 1 à gauche, Groupe 2 à droite)

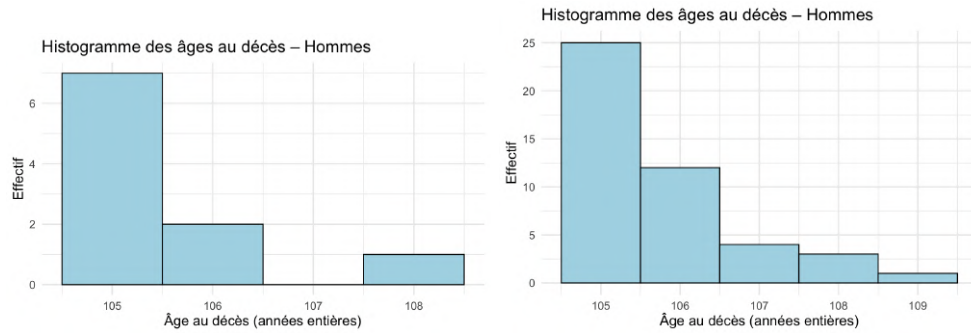


FIGURE 3.2 – Histogrammes des âges au décès – Hommes (Groupe 1 à gauche, Groupe 2 à droite)

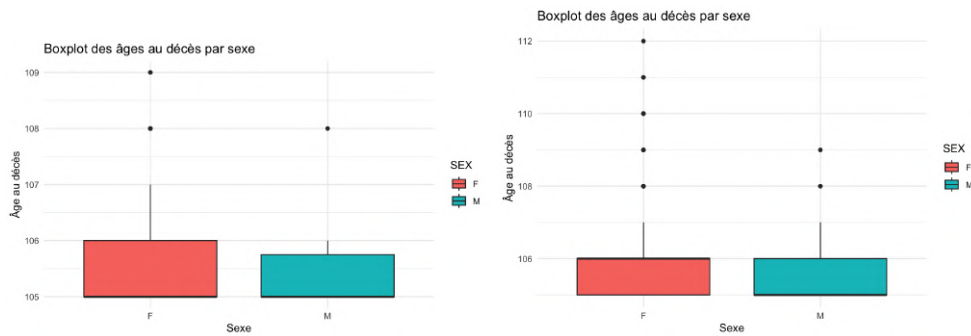


FIGURE 3.3 – Boxplots comparés des âges au décès – (Groupe 1 à gauche, Groupe 2 à droite)

Nous terminons par une analyse par cohorte (1905–1915), en croisant le nombre de survivants à 105 ans (L_{105}), l'âge maximal observé (m) et l'âge moyen conditionnel ($105 + e_{105}$). Cela permet d'évaluer la progression inter-générationnelle de la longévité dans chaque groupe.

TABLE 3.2 – Indicateurs par cohorte – Groupe 1

Année	L_{105}^F	m^F	$105 + e_{105}^F$
1905	52	109	105.75
1906	44	108	105.77
1907	41	108	105.76
1908	41	107	105.34
1909	27	106	105.14
1910	1	105	105.00
1911	0	—	—
1912	0	—	—
1913	0	—	—
1914	0	—	—
1915	0	—	—

TABLE 3.3 – Indicateurs par cohorte – Groupe 2

Année	L_{105}^F	m^F	$105 + e_{105}^F$
1905	58	112	105.97
1906	50	111	105.86
1907	51	110	106.02
1908	57	111	105.98
1909	64	111	106.08
1910	42	111	106.12
1911	64	109	106.00
1912	50	108	105.68
1913	49	108	105.47
1914	49	107	105.59
1915	22	106	105.32

Légende :

- L_{105} : survivants à l'âge de 105 ans dans la cohorte
- m : âge maximal observé
- $105 + e_{105}$: âge moyen conditionnel au-delà de 105 ans

TABLE 3.4 – Indicateurs par cohorte – Groupe 1

Année	L_{105}^H	m^H	$105 + e_{105}^H$
1905	1	108	108.00
1906	2	105	105.00
1907	1	106	106.00
1908	4	105	105.00
1909	2	106	105.50
1910	0	–	–
1911	0	–	–
1912	0	–	–
1913	0	–	–
1914	0	–	–
1915	0	–	–

TABLE 3.5 – Indicateurs par cohorte – Groupe 2

Année	L_{105}^H	m^H	$105 + e_{105}^H$
1905	2	108	107.00
1906	3	107	105.67
1907	1	106	106.00
1908	5	105	105.00
1909	5	108	106.20
1910	5	105	105.00
1911	8	108	106.00
1912	3	109	106.67
1913	5	107	105.40
1914	6	107	105.67
1915	2	106	105.50

Retrait des cohortes récentes (1914–1915) dans le groupe 2

Le modèle de Gbari et al. (2017) repose sur l’analyse de générations entièrement éteintes. Cette condition garantit que tous les décès ont été observés, évitant ainsi les biais liés à la censure à droite.

Or, la base IDL ne précise pas si une cohorte est totalement éteinte. Pour contourner cette limite, nous avons vérifié, pour chaque génération, l’âge maximum enregistré, puis croisé ces informations avec des sources externes.

Deux cas récents¹ ont remis en cause l’inclusion des générations 1914 et 1915 :

- **Fernande Courtoy**, doyenne des Belges jusqu’à son décès en mai 2024, était née en juillet 1914 et est décédée à 109 ans. Or, la base IDL n’enregistrait aucun décès au-delà de 107 ans pour cette cohorte.
- **Maria Waelbers**, née en juillet 1915, est devenue la nouvelle doyenne à 109 ans, ce qui confirme que la cohorte 1915 est encore partiellement vivante.

Ces éléments montrent que les générations 1914 et 1915 n’étaient pas entièrement éteintes au moment de la dernière mise à jour de l’IDL (avril 2022). Par souci de rigueur, elles seront donc exclues du groupe 2, bien qu’elles respectaient les filtres initiaux.

Néanmoins, afin de mesurer l’impact de cette décision, une analyse de sensibilité sera menée. Deux variantes du groupe 2 sont donc considérées dans la suite de l’étude :

- le groupe 2 complet, tel que défini initialement, incluant les cohortes 1914 et 1915 ;
- le groupe 2 restreint, excluant ces deux générations, ce qui ramène la

1. Voir, les articles de RTL Info (mai 2024) et L’Avenir (mars 2025) en bibliographie.

taille de l'échantillon de **601 à 522 individus**.

Cette double analyse permettra d'évaluer si la présence éventuelle de données censurées dans les dernières cohortes influe de manière significative sur les résultats obtenus.

3.2.4 Analyse basée sur les âges exacts au décès

Dans un second temps, nous avons affiné les données de longévité en calculant l'âge exact de chaque individu (en années décimales), à partir des dates complètes de naissance et de décès. Cette approche permet une représentation plus précise de la répartition des âges extrêmes.

Les tableaux suivants présentent un résumé statistique des âges exacts au décès, par sexe et par groupe.

TABLE 3.6 – Statistiques descriptives (âges exacts au décès) pour les deux groupes

Groupe 1 (échantillon restreint)

Sexe	n	Min	Q ₁	Médiane	Moy.	Q ₃	Max	Sd
F	206	105,01	105,29	105,66	105,99	106,53	109,16	0,90
H	10	105,08	105,20	105,62	105,93	106,12	108,97	1,16

Groupe 2 (échantillon élargi)

Sexe	n	Min	Q ₁	Médiane	Moy.	Q ₃	Max	Sd
F	556	105,01	105,42	106,01	106,31	106,80	112,06	1,22
H	45	105,00	105,45	105,90	106,18	106,55	109,10	1,03

Les figures suivantes illustrent les distributions des âges exacts au décès pour les deux sexes et les deux groupes. Les histogrammes permettent d'observer la forme des distributions, tandis que les boxplots facilitent la comparaison des dispersions.

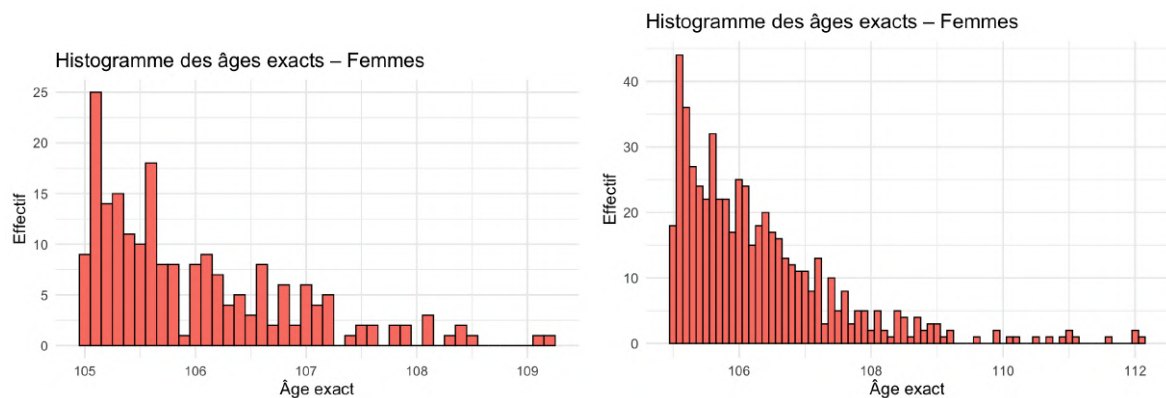


FIGURE 3.4 – Histogrammes des âges exacts – Femmes (Groupe 1 à gauche, Groupe 2 à droite)

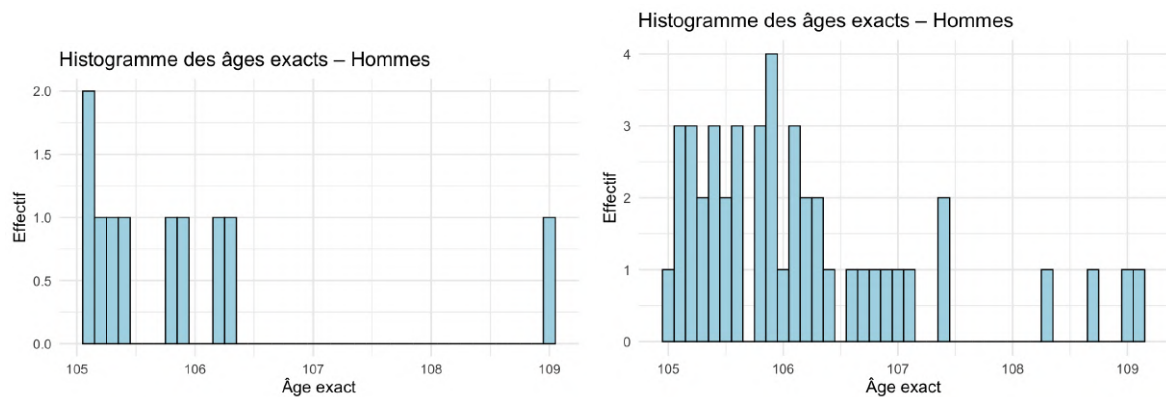


FIGURE 3.5 – Histogrammes des âges exacts – Hommes (Groupe 1 à gauche, Groupe 2 à droite)

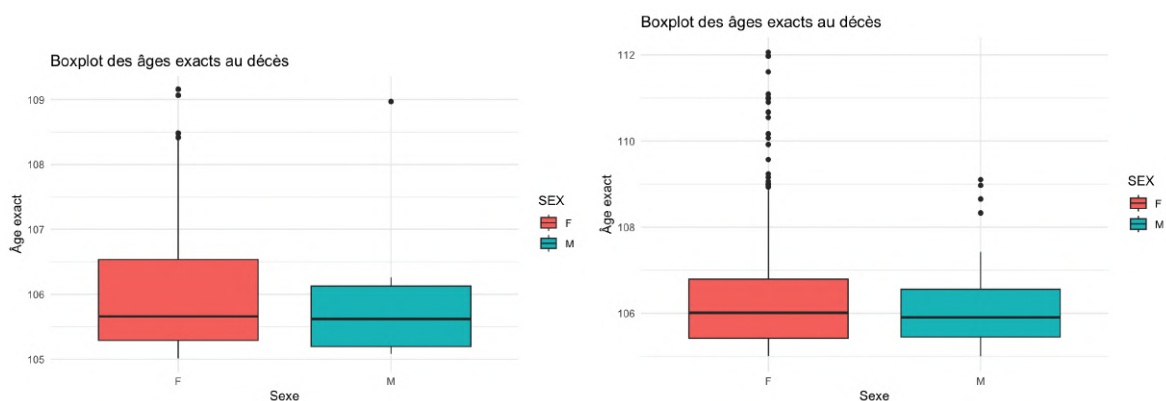


FIGURE 3.6 – Boxplots comparés des âges exacts au décès (Groupe 1 à gauche, Groupe 2 à droite)

Remarque sur les doublons. Dans notre échantillon de 216 individus, l'âge au décès a été calculé avec précision à partir des dates de naissance et de décès, grâce à la fonction `time_length(interval(...), "years")` du package `lubridate`. Cette méthode permet d'obtenir un âge exact exprimé en années décimales, avec une précision proche du jour.

Cependant, malgré cette précision, nous avons détecté **15 doublons exacts** dans les âges calculés, c'est-à-dire des cas où plusieurs individus ont exactement la même valeur numérique d'âge au décès (par exemple 105,0493). Ces doublons ne résultent pas d'une erreur ni d'un manque de précision dans les données : ils s'expliquent par le fait que plusieurs personnes ont vécu exactement le même nombre de jours. En effet, deux individus nés à des dates différentes peuvent parfaitement décéder à des dates différentes tout en ayant la même durée de vie en jours (voir tableau A.1). Ces 15 doublons concernent **27 individus** : **25 femmes** et **2 hommes**. Les cas les plus fréquents se concentrent autour de l'âge de 105 ans.

Afin de pouvoir utiliser ces âges dans certains tests statistiques nécessitant des observations distinctes (comme les QQ-plots ou le test de Kolmogorov–Smirnov), nous avons construit une version modifiée de la variable,

appelée **AgeExact_unique**. Celle-ci est obtenue en ajoutant une très faible quantité proportionnelle au rang d'apparition ($+ \text{row_number}() * 1\text{e-}10$) **uniquement pour les observations en double**, ce qui garantit l'unicité des âges exacts sans modifier les autres valeurs ni l'ordre initial. Cette correction est purement technique et n'affecte en rien les résultats des analyses globales.

La même méthode a été appliquée au groupe 2, afin d'assurer la compatibilité des données avec les tests statistiques fondés sur des distributions continues. Aucune transformation n'a été effectuée sur les observations non concernées. Cette correction est purement technique et n'affecte en rien les résultats des analyses globales.

Enfin, les tableaux ci-dessous reprennent la structure précédente par cohorte de naissance, en s'appuyant cette fois sur les **âges exacts au décès**. Ils permettent d'évaluer l'évolution des indicateurs de longévité entre 1905 et 1915, séparément pour chaque groupe étudié.

TABLE 3.7 – Groupe 1 : échantillon restreint (206 femmes, 10 hommes)

Année	$L_{105}^{(F)}$	$m^{(F)}$	$105 + e_{105}^{(F)}$	$L_{105}^{(M)}$	$m^{(M)}$	$105 + e_{105}^{(M)}$
1905	52	109,16	106,14	1	108,97	108,97
1906	44	108,48	106,16	2	105,28	105,21
1907	41	108,13	106,18	1	106,26	106,26
1908	41	107,01	105,74	4	105,95	105,50
1909	27	106,46	105,53	2	106,18	105,82
1910	1	105,11	105,11	0	—	—
1911	0	—	—	0	—	—
1912	0	—	—	0	—	—
1913	0	—	—	0	—	—
1914	0	—	—	0	—	—
1915	0	—	—	0	—	—

TABLE 3.8 – Groupe 2 : échantillon élargi (556 femmes, 45 hommes)

Année	$L_{105}^{(F)}$	$m^{(F)}$	$105 + e_{105}^{(F)}$	$L_{105}^{(M)}$	$m^{(M)}$	$105 + e_{105}^{(M)}$
1905	58	112,06	106,37	2	108,97	107,76
1906	50	111,01	106,24	3	107,09	105,84
1907	51	110,99	106,48	1	106,26	106,26
1908	57	111,97	106,44	5	105,95	105,53
1909	64	111,98	106,55	5	108,65	106,65
1910	42	111,09	106,57	5	105,89	105,55
1911	64	109,92	106,48	8	108,33	106,43
1912	50	108,86	106,17	3	109,10	107,03
1913	49	108,10	105,91	5	107,08	105,80
1914	49	107,67	105,97	6	107,40	106,24
1915	22	106,68	105,76	2	106,07	105,54

Légende :

- L_{105} : survivants à l'âge de 105 ans dans la cohorte
- m : âge maximal observé

— $105 + e_{105}$: âge moyen conditionnel au-delà de 105 ans

3.3 Loi de survie au-delà de 105 ans

Dans cette section, nous rappelons le principe fondamental qui permet d’obtenir la loi de survie au-delà d’un seuil donné, ici fixé à 105 ans, tout en conservant la structure de la loi de Pareto généralisée (GPD) utilisée dans l’étude de Gbari et al. (2017).

3.3.1 Propriété de stabilité de la GPD

Soit une variable aléatoire X suivant une loi GPD de paramètres (ξ, β) au-dessus d’un seuil u . La propriété de stabilité stipule que, conditionnellement à l’évènement $X > u$, l’excès $X - u$ suit également une loi GPD, de même indice de queue ξ mais avec un paramètre de dispersion modifié :

$$X - u \mid X > u \sim \text{GPD}(\xi, \beta + \xi u).$$

Autrement dit, si l’on augmente le seuil, le paramètre ξ reste inchangé, mais β devient :

$$\beta_{\text{ajustée}} = \beta_{\text{initiale}} + \xi(u_{\text{nouveau}} - u_{\text{ancien}}).$$

3.3.2 Application à notre étude

Dans le cadre de l’étude de Gbari et al. (2017), les paramètres initiaux avaient été estimés pour des seuils $x^* = 100,89$ (femmes) et $x^* = 98,89$ (hommes). Pour tester l’adéquation du modèle à nos données, nous devons recalculer les lois au-delà de 105 ans en utilisant cette propriété de stabilité.

Les paramètres recalculés sont présentés dans le tableau ci-dessous :

Paramètre	Femmes ($x^* = 100,89 \rightarrow 105$)	Hommes ($x^* = 98,89 \rightarrow 105$)
$\hat{\xi}$	−0,092	−0,132
$\text{IC}_{95} \hat{\xi}$	[−0,12 ; −0,07]	[−0,16 ; −0,10]
$\hat{\beta}$ initiale	2,019	2,098
β ajustée	1,641	1,291
$\hat{\omega} = x^* - \beta/\hat{\xi}$	122,73	114,82

Ces nouvelles valeurs seront utilisées dans les tests d’ajustement (Kolmogorov–Smirnov, Khi²) développés dans les sections suivantes. L’objectif est d’évaluer dans quelle mesure les observations issues de notre base IDL sont compatibles avec le modèle GPD initialement établi dans l’article de 2017.

3.4 Résumé et analyse des tests du χ^2

Dans cette section, nous évaluons l'ajustement des excès féminins issus de la base IDL au modèle de loi GPD recalibré, en fixant le seuil à $u = 105$. Les paramètres retenus sont $\xi = -0,092$ et $\beta = 1,641$, ce qui conduit à un âge ultime estimé de $\omega = 122,73$, soit un excès maximal théorique de $x_{\max} = 17,73$ ans.

Pour chaque groupe d'analyse (restreint et élargi), nous avons extrait les excès au-delà de 105 ans et appliqué un test du χ^2 afin d'évaluer la compatibilité avec la loi GPD susmentionnée. Les histogrammes ci-dessous illustrent la distribution des excès observés dans chacun des deux échantillons féminins.

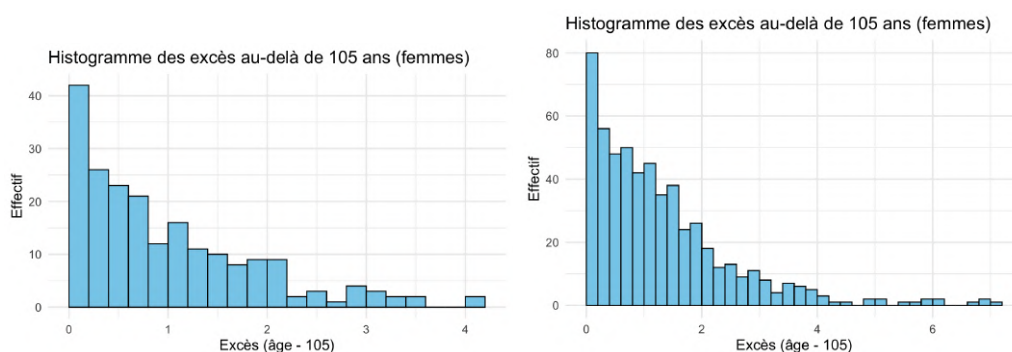


FIGURE 3.7 – Histogrammes des excès au-delà de 105 ans – Femmes (Groupe 1 à gauche, Groupe 2 à droite)

Dans les deux cas, la majorité des excès sont concentrés dans les premières tranches d'âge au-delà du seuil, ce qui reflète une forte mortalité peu après 105 ans. Le groupe élargi présente toutefois une dispersion plus marquée, en lien avec l'inclusion des cohortes plus jeunes (1911–1915) et de cas exceptionnels atteignant 112 ans.

Trois tests du χ^2 sont menés en variant les bornes de classes, tout en respectant les règles suivantes :

- couverture complète de l'intervalle des excès $[0, x_{\max}[$;
- effectifs théoriques suffisants dans chaque classe (≥ 5) ;
- égalité entre la somme des observés et celle des attendus, par construction.

Synthèse. Dans les deux groupes étudiés (restreint et élargi), les tests du χ^2 révèlent une incompatibilité marquée entre les excès féminins observés dans la base IDL et la loi GPD calibrée à partir de la propriété de stabilité au-delà du seuil de 105 ans ($\xi = -0,092$, $\beta = 1,641$).

Pour chacun des groupes, trois configurations de découpage ont été testées, couvrant intégralement l'intervalle $[0, 17,73[$, avec des effectifs théo-

riques supérieurs à 5 et une correspondance exacte entre les totaux observés et attendus (206 et 556 femmes respectivement). Malgré ces conditions méthodologiquement solides, les résultats obtenus sont significatifs dans tous les cas : les p-values vont de $3,1 \times 10^{-5}$ à $4,0 \times 10^{-4}$ dans le groupe restreint, et de $1,08 \times 10^{-3}$ à $1,35 \times 10^{-2}$ dans le groupe élargi.

Dans les deux échantillons, les histogrammes montrent un **excès systématique d'observations dans les premières classes** (notamment $[0-1[$), accompagné d'un déficit progressif dans les classes plus éloignées du seuil. Ce déséquilibre suggère que les décès surviennent plus précocement que ce que prévoit le modèle GPD considéré.

Test	Intervalles utilisés	Probab théoriques (par classe)	Effectifs Observés / Attendus	Résultat test χ^2
1	$[0,1[$, $[1,2[$, $[2,3[$, $[3,17.73[$	0.4659, 0.2596, 0.1394, 0.1351	Obs : 124, 53, 20, 9 Att : 95.98, 53.48, 28.72, 27.83	$\chi^2 = 23.57$, ddl = 3, $p = 3.1 \times 10^{-5}$
2	$[0,1[$, $[1,2[$, $[2,3[$, $[3,4[$, $[4,17.73[$	0.4659, 0.2596, 0.1394, 0.0718, 0.0633	Obs : 124, 53, 20, 7, 2 Att : 95.98, 53.48, 28.72, 14.79, 13.03	$\chi^2 = 24.27$, ddl = 4, $p = 7.0 \times 10^{-5}$
3	$[0,0.5[$, $[0.5,1[$, $[1,1.5[$, $[1.5,2[$, $[2,2.5[$, $[2.5,3[$, $[3,4[$, $[4,17.73[$	0.2659, 0.2000, 0.1492, 0.1103, 0.0808, 0.0586, 0.0718, 0.0633	Obs : 78, 46, 29, 24, 12, 8, 7, 2 Att : 54.77, 41.21, 30.74, 22.73, 16.65, 12.07, 14.79, 13.03	$\chi^2 = 26.69$, ddl = 7, $p = 4.0 \times 10^{-4}$

TABLE 3.9 – Synthèse des résultats des tests du χ^2 sur les excès IDL femmes - Groupe 1 .

Test	Intervalles utilisés	Probab théoriques (par classe)	Effectifs Observés / Attendus	Résultat test χ^2
1	$[0,1[$, $[1,2[$, $[2,3[$, $[3,17.73[$	0.4659, 0.2596, 0.1394, 0.1351	Obs : 276, 167, 64, 49 Att : 259.04, 144.33, 77.52, 75.11	$\chi^2 = 16.11$, ddl = 3, $p = 1.08 \times 10^{-3}$
2	$[0,1[$, $[1,2[$, $[2,3[$, $[3,4[$, $[4,17.73[$	0.4659, 0.2596, 0.1394, 0.0718, 0.0633	Obs : 276, 167, 64, 30, 19 Att : 259.04, 144.33, 77.52, 39.93, 35.18	$\chi^2 = 16.94$, ddl = 4, $p = 1.99 \times 10^{-3}$
3	$[0,0.5[$, $[0.5,1[$, $[1,1.5[$, $[1.5,2[$, $[2,2.5[$, $[2.5,3[$, $[3,4[$, $[4,17.73[$	0.2659, 0.2000, 0.1492, 0.1103, 0.0808, 0.0586, 0.0718, 0.0633	Obs : 159, 117, 99, 68, 40, 24, 30, 19 Att : 147.82, 111.22, 82.98, 61.35, 44.94, 32.58, 39.93, 35.18	$\chi^2 = 17.67$, ddl = 7, $p = 1.35 \times 10^{-2}$

TABLE 3.10 – Synthèse des résultats des tests du χ^2 sur les excès IDL Femmes – Groupe 2 .

Conclusion. Que l'on considère l'échantillon restreint (206 femmes nées entre 1905 et 1910) ou élargi (556 femmes nées entre 1905 et 1915), les excès d'âge au-delà de 105 ans ne semblent pas suivre la loi GPD initialement proposée. Ces résultats invitent à reconsidérer la validité de ce modèle pour cette population. Toutefois, les tests du chi carré, sensibles au découpage

en classes et fondés sur des approximations asymptotiques, doivent être interprétés avec prudence. Cela justifie le recours complémentaire au test de Kolmogorov–Smirnov, plus adapté aux faibles effectifs en queue.

Visualisation des écarts

Les graphiques suivants comparent les effectifs observés et attendus sous la loi GPD, pour chacun des trois découpages testés. Pour chaque découpage, les résultats sont présentés côte à côte pour le groupe restreint (gauche) et le groupe élargi (droite).

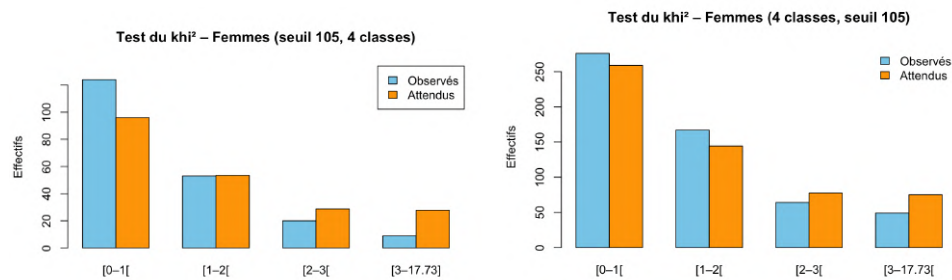


FIGURE 3.8 – Découpage en 4 classes : $[0, 1[$, $[1, 2[$, $[2, 3[$, $[3, 17.73[$, Groupe 1 (à gauche) – Groupe 2 (à droite).

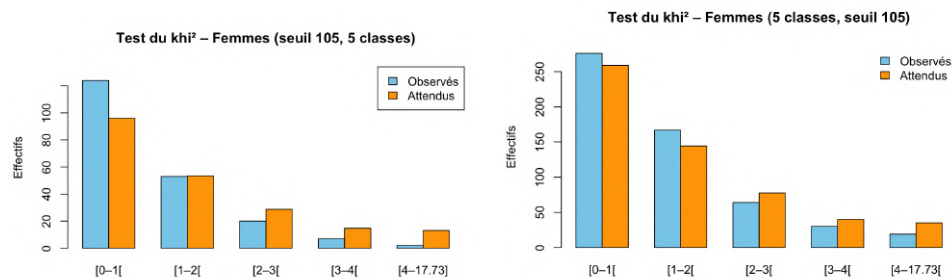


FIGURE 3.9 – Découpage en 5 classes : $[0, 1[$, $[1, 2[$, $[2, 3[$, $[3, 4[$, $[4, 17.73[$, Groupe 1 (à gauche) – Groupe 2 (à droite).

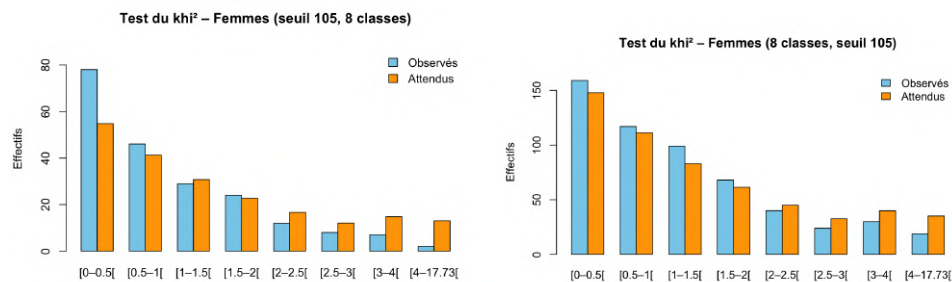


FIGURE 3.10 – Découpage en 8 classes : $[0, 0.5[$, $[0.5, 1[$, $[1, 1.5[$, $[1.5, 2[$, $[2, 2.5[$, $[2.5, 3[$, $[3, 4[$, $[4, 17.73[$, Groupe 1 (à gauche) – Groupe 2 (à droite).

Afin de compléter cette analyse, nous avons considéré une variante du groupe 2 dans laquelle les individus nés en 1914 et 1915 sont exclus, conformément aux remarques formulées plus haut sur l’extinction incomplète de ces générations. Cette version restreinte permet de tester la robustesse des résultats en supprimant les observations potentiellement censurées.

Analyse de sensibilité – Groupe 2 restreint (485 femmes)

Nous réitérons ici les trois tests du χ^2 sur les excès féminins, en conservant les mêmes configurations d’intervalles, cette fois appliqués à la variante restreinte du groupe 2.

Test	Intervalles utilisés	Probab théoriques (par classe)	Effectifs Observés / Attendus	Résultat test χ^2
1	[0,1[, [1,2[, [2,3[, [3,17.73[	0.4659, 0.2596, 0.1394, 0.1351	Obs : 239, 135, 62, 49 Att : 225.96, 125.90, 67.62, 65.52	$\chi^2 = 6.04$, ddl = 3, p = 0.110
2	[0–1[, [1–2[, [2–3[, [3–4[, [4–17.73[	0.4659, 0.2596, 0.1394, 0.0718, 0.0633	Obs : 239, 135, 62, 30, 19 Att : 225.96, 125.90, 67.62, 34.83, 30.69	$\chi^2 = 7.00$, ddl = 4, p = 0.136
3	[0–0.5[, [0.5–1[, [1–1.5[, [1.5–2[, [2–2.5[, [2.5–3[, [3–4[, [4–17.73[	0.2659, 0.2000, 0.1492, 0.1103, 0.0808, 0.0586, 0.0718, 0.0633	Obs : 138, 101, 74, 61, 39, 23, 30, 19 Att : 128.95, 97.02, 72.38, 53.52, 39.20, 28.42, 34.83, 30.69	$\chi^2 = 8.04$, ddl = 7, p = 0.329

TABLE 3.11 – Synthèse des résultats des tests du χ^2 sur les excès féminins – Groupe 2 restreint (485 femmes)

Interprétation. les résultats obtenus sur cet échantillon restreint diffèrent nettement des tendances observées dans les groupes précédents.

Dans les trois configurations testées, les *p-values* sont toutes largement supérieures au seuil usuel de 5%, ce qui ne permet pas de rejeter l’hypothèse d’un bon ajustement au modèle GPD. Ces résultats indiquent qu’après exclusion des cohortes 1914 et 1915, la loi de Pareto généralisée, avec les paramètres $\xi = -0,092$ et $\beta = 1,641$, devient compatible avec les excès observés chez les femmes.

Autrement dit, le rejet du modèle observé dans le groupe 2 initial semble provenir de la présence de générations non totalement éteintes. Leur retrait améliore significativement la qualité de l’ajustement statistique.

3.5 Résultats des tests de Kolmogorov–Smirnov

Dans cette section, nous évaluons l’ajustement des excès IDL (femmes et hommes) au modèle GPD en utilisant les tests de Kolmogorov-Smirnov.

3.5.1 Principe du test

Le test de Kolmogorov–Smirnov (KS) est un test non paramétrique qui permet de comparer la fonction de répartition empirique $F_n(x)$ d'un échantillon à une fonction de répartition théorique $F(x)$, en l'occurrence celle de la loi GPD spécifiée. La statistique du test est définie comme suit :

$$D_n = \sup_x |F_n(x) - F(x)|.$$

Elle mesure l'écart maximal entre les deux fonctions de répartition. Plus D_n est grand, plus l'ajustement est considéré comme mauvais. La p-value est alors calculée sous l'hypothèse nulle H_0 : « les données suivent la loi F », selon deux approches :

- la **loi asymptotique** de D_n , utilisée pour les grands échantillons, reposant sur la distribution limite de Kolmogorov ;
- la **loi exacte** de D_n , applicable aux petits échantillons ($n \leq 100$), à condition que la loi soit entièrement spécifiée (paramètres fixés) et que les données soient continues (sans doublons).

3.5.2 Application aux excès IDL — Femmes

Les excès sont calculés au-delà du seuil de 105 ans, à partir des âges exacts au décès (en années décimales), obtenus à l'aide des dates de naissance et de décès. Dans les groupes féminins, plusieurs individus présentent le même âge exact. Afin d'éviter les égalités (doublons), un léger décalage individualisé a été appliqué à chaque valeur (**AgeExact_unique**), assurant leur unicité sans altérer la distribution globale.

Les excès sont comparés à une loi GPD spécifiée par les paramètres fixés $\xi = -0,092$ et $\beta = 1,641$. Le test de Kolmogorov–Smirnov est ici utilisé en version asymptotique, en raison de la taille élevée des échantillons ($n > 100$).

Résultats – Femmes (test asymptotique) :

- Groupe 1 : $D = 0,1699$, p-value = $1,368 \times 10^{-5}$
- Groupe 2 : $D = 0,0746$, p-value = $0,0041$
- Groupe 2 restreint : $D = 0,0521$, p-value = $0,144$

Interprétation. Les deux premiers groupes conduisent à un rejet clair de l'hypothèse de conformité à la loi GPD recalibrée. En revanche, l'ajustement devient acceptable dans la version restreinte du groupe 2, après retrait des générations 1914 et 1915.

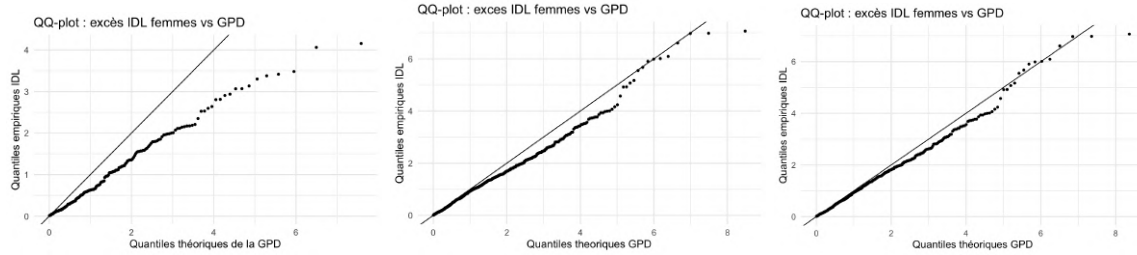


FIGURE 3.11 – QQ-plots des excès féminins vs GPD ($\xi = -0,092$, $\beta = 1,641$) pour les Groupes 1, 2 et 2 restreint

3.5.3 Application aux excès IDL — Hommes

L'analyse repose sur les excès masculins (au-delà de 105 ans), calculés à partir des âges exacts au décès (en années décimales), comme précédemment. Aucun doublon n'est présent dans les observations.

Étant donné le faible effectif dans chaque groupe (10, 45 et 37 individus), nous utilisons la **version exacte** du test de Kolmogorov–Smirnov, via la commande `ks.test(..., exact = TRUE)` dans R. Cette version repose sur la distribution exacte de la statistique D_n , telle que décrite par Marsaglia, Tsang et Wang (2003), et évite les approximations asymptotiques peu fiables dans les petits échantillons.

Son application suppose deux conditions, toutes deux remplies ici :

- les paramètres de la loi GPD (ici $\xi = -0,132$ et $\beta = 1,291$) sont fixés a priori ;
- les données ne présentent aucun doublon.

Résultats – Hommes (test exact) :

- Groupe 1 : $D = 0,2509$, p-value = 0,4793
- Groupe 2 : $D = 0,0835$, p-value = 0,8864
- Groupe 2 restreint : $D = 0,0695$, p-value = 0,9885

Interprétation. Dans les trois groupes masculins, les p-values élevées ne permettent pas de rejeter l'hypothèse que les excès suivent la loi GPD spécifiée. Cette conclusion doit cependant être interprétée avec prudence, compte tenu du faible nombre d'observations.

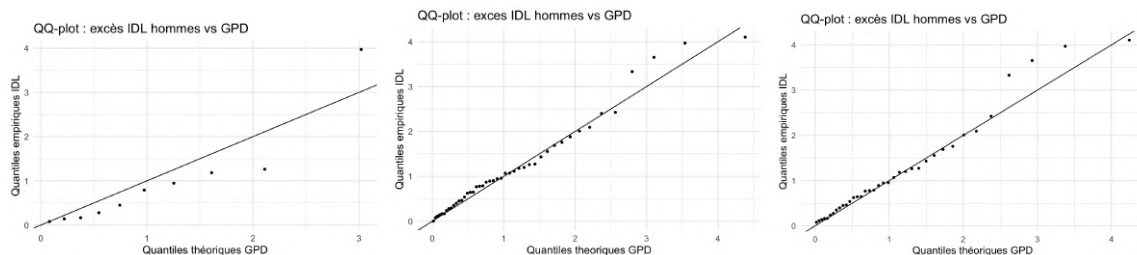


FIGURE 3.12 – QQ-plots des excès masculins vs GPD ($\xi = -0,132$, $\beta = 1,291$) pour les Groupes 1, 2 et 2 restreint

Conclusion. Ce travail avait pour objectif de tester si le modèle GPD proposé en 2017 reste adapté aux données belges récentes issues de la base IDL, pour les âges au-delà de 105 ans. Les tests de Kolmogorov–Smirnov, complétés par l’analyse des QQ-plots, ont révélé des ajustements contrastés selon le sexe et la composition des échantillons.

Chez les femmes, le modèle GPD recalibré ($\xi = -0,092$, $\beta = 1,641$) est rejeté dans les groupes 1 et 2, avec des p-values très faibles. Les QQ-plots montrent une surestimation systématique des quantiles, en particulier dans le groupe 1. En revanche, lorsque les générations 1914 et 1915 sont exclues (groupe 2 restreint), l’ajustement devient acceptable : le test ne rejette plus la GPD et la concordance visuelle est globalement satisfaisante, malgré quelques écarts pour les quantiles les plus élevés.

Chez les hommes, aucun rejet n’est observé dans les trois groupes. Les p-values sont élevées et les QQ-plots confirment une bonne adéquation, notamment dans les groupes élargi et restreint. Toutefois, les effectifs masculins restent faibles, ce qui limite la fiabilité des conclusions.

En résumé, la GPD utilisée avec les paramètres de 2017 ne permet pas de modéliser correctement la mortalité féminine au-delà de 105 ans dans les données IDL, sauf après filtrage strict de l’échantillon. Pour les hommes, le modèle reste compatible, mais les données disponibles sont trop limitées pour conclure avec certitude.

La suite du travail consistera désormais à réestimer les paramètres du modèle à partir des données IDL récentes, afin de proposer une version actualisée et mieux adaptée à la réalité belge, en particulier pour la population féminine.

3.6 Réajustement de la GPD sur les données IDL

Après avoir constaté l’inadéquation du modèle GPD de 2017 avec les données IDL les plus récentes, nous procédons ici à une réestimation des paramètres de la loi, directement à partir des excès observés dans la base. L’objectif est de vérifier si une mise à jour des paramètres ξ et β permettrait un meilleur ajustement aux données belges, en fonction du sexe, du groupe étudié et du seuil retenu.

Cette réestimation est particulièrement justifiée chez les **femmes**, pour lesquelles le modèle initial a été rejeté dans les groupes 1 et 2. Elle permet également de compléter l’analyse pour les **hommes**, même si les effectifs réduits appellent à une certaine prudence dans l’interprétation.

3.6.1 Réestimation des paramètres à partir du seuil de 105 ans

Dans cette série d'ajustements, le seuil u est fixé à 105 ans pour tous les groupes. Ce choix s'impose naturellement, car la base IDL ne contient que des individus ayant atteint cet âge, ce qui rend impossible l'exploration de seuils plus bas. Ce seuil reste néanmoins pertinent, car il dépasse largement les seuils obtenus par Gbari et al. (2017), ce qui assure une certaine stabilité de l'ajustement au modèle GPD, généralement meilleure lorsque le seuil augmente.

Si les résultats s'avèrent satisfaisants, la loi ainsi ajustée pourra ensuite être utilisée pour extrapoler la mortalité à des âges plus avancés.

Les ajustements ont été réalisés à l'aide du logiciel R, en utilisant la fonction `fitgpd` du package `POT`, avec une estimation par maximum de vraisemblance (`est = "mle"`). Les quatre cas sont analysés séparément : femmes et hommes, pour les groupes 1 et 2.

TABLE 3.12 – Paramètres estimés, IC 95 % de ξ , test KS et âge ultime – seuil $u = 105$

Résultats	Groupe 1		Groupe 2	
	Femmes	Hommes	Femmes	Hommes
$\hat{\xi}$ (shape)	−0,150	0,209	−0,066	−0,185
IC 95 % pour ξ	[−0,30 ; −0,004]	[−0,57 ; 0,99]	[−0,14 ; 0,01]	[−0,49 ; 0,12]
$\hat{\beta}$ (scale)	1,139	0,738	1,391	1,398
Âge ultime $\hat{\omega}$	112,60	$+\infty$	126,11	112,57
p-value (KS)	0,6646	–	0,7038	–

Les graphiques ci-dessous permettent d'évaluer visuellement la qualité de l'ajustement dans chaque cas, à travers la comparaison entre les quantiles observés et ceux issus de la loi GPD estimée.

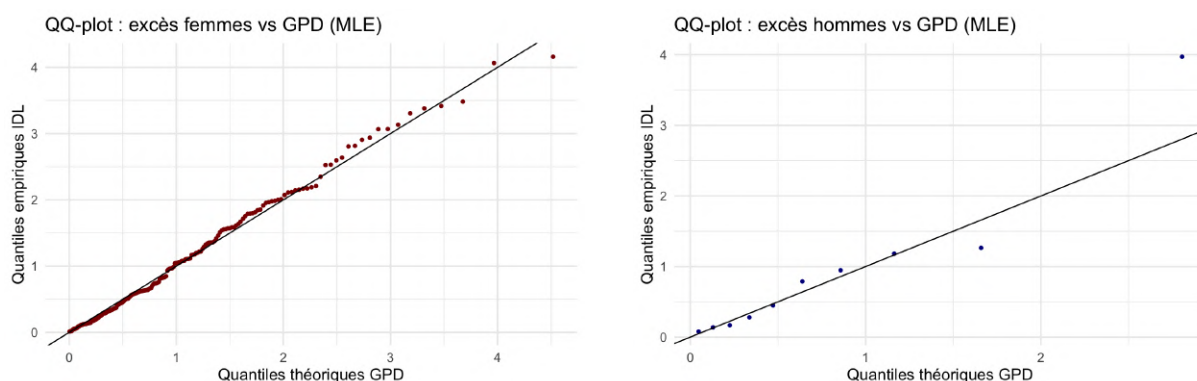


FIGURE 3.13 – QQ-plots : excès IDL vs GPD ($u = 105$) – Groupe 1 ; femmes à gauche, hommes à droite

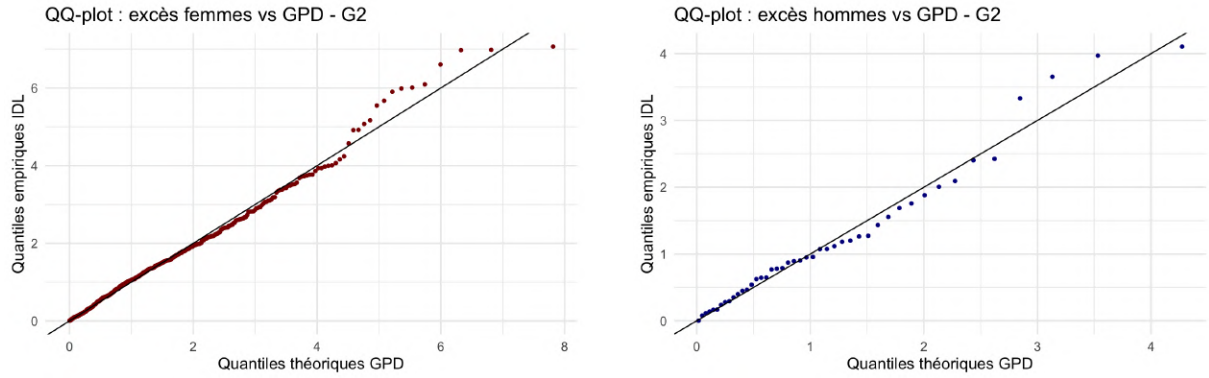


FIGURE 3.14 – QQ-plots : excès IDL vs GPD ($u = 105$) – Groupe 2 ; femmes à gauche, hommes à droite

Remarque importante. Il est crucial de souligner qu’aucun test de Kolmogorov–Smirnov n’a été réalisé pour les hommes ici, car les paramètres de la loi GPD ont été estimés directement sur leurs données. Or, la version exacte de ce test n’est valide que si les paramètres de la loi sont fixés à l’avance et non dérivés de l’échantillon. Cette restriction s’applique également en théorie aux grands échantillons, comme chez les femmes, mais son impact y est généralement limité, permettant d’utiliser le test de manière conservatrice.

Conclusion. La réestimation avec un seuil unique à 105 ans permet d’obtenir un bon ajustement pour les femmes, dans les deux groupes. Les QQ-plots montrent une concordance visuelle satisfaisante, et le test de Kolmogorov–Smirnov ne rejette pas l’hypothèse d’une loi GPD. Le paramètre ξ reste négatif dans le groupe 1, avec un intervalle de confiance à 95 % entièrement négatif, ce qui indique une distribution bornée. Dans le groupe 2, en revanche, l’intervalle inclut des valeurs positives, ce qui introduit une incertitude sur l’existence d’un âge ultime.

Chez les hommes, aucun test statistique n’a été réalisé. Les résultats doivent donc être interprétés avec prudence. Les intervalles de confiance de ξ sont très larges et englobent des valeurs positives, notamment dans le groupe 1 où l’estimation centrale est elle-même strictement positive. Cela suggère une distribution non bornée, mais l’effectif très limité rend ces résultats peu fiables.

3.6.2 Modèle unisexe : validation empirique

Afin d’évaluer la pertinence d’un modèle unisexe au-delà de 105 ans, nous avons comparé les excès féminins et masculins dans les groupes 1 et 2. Deux tests complémentaires ont été réalisés dans chaque groupe : un test de Kolmogorov–Smirnov pour comparer les distributions empiriques des excès, et un test de rapport de vraisemblance pour confronter les paramètres

estimés.

Comme précédemment, une correction des doublons a été appliquée aux vecteurs d'excès afin de respecter les conditions du test de Kolmogorov–Smirnov, qui suppose des observations indépendantes et strictement continues.

Test de Kolmogorov–Smirnov bilatéral

Ce test permet de comparer directement les distributions empiriques des excès chez les femmes et les hommes.

- H_0 : les excès féminins et masculins suivent la même loi de probabilité.
- H_1 : les lois suivies par les excès diffèrent selon le sexe.

Les résultats sont les suivants :

- **Groupe 1** : $D = 0,201$, p-value = 0,7652 ;
- **Groupe 2** : $D = 0,112$, p-value = 0,6712 .

Dans tous les cas, la p-value est largement supérieure au seuil de 5 %, ce qui signifie que l'on ne rejette pas H_0 . Aucune différence significative n'est détectée entre les distributions d'excès selon le sexe.

Test du rapport de vraisemblance

Ce test permet de comparer deux modèles imbriqués : un modèle restreint dans lequel les deux sexes partagent les mêmes paramètres (ξ, β) , et un modèle général où les paramètres sont estimés séparément pour les femmes et les hommes. L'objectif est de vérifier si l'on peut adopter une loi commune au-delà de 105 ans sans perte significative d'ajustement.

La statistique de test est donnée par :

$$\Lambda = -2 [\ell_{\text{restreint}} - \ell_{\text{général}}] = -2 [\ell_G - (\ell_F + \ell_M)] ,$$

où :

- ℓ_F est la log-vraisemblance obtenue sur les excès féminins ;
- ℓ_M est la log-vraisemblance obtenue sur les excès masculins ;
- ℓ_G est la log-vraisemblance globale sur l'ensemble des excès combinés.

Sous l'hypothèse nulle (modèle commun), la statistique Λ suit asymptotiquement une loi du χ^2 à 2 degrés de liberté. Ce nombre correspond à la différence de paramètres entre les deux modèles (quatre dans le modèle général contre deux dans le modèle restreint). Aucun ajustement n'est nécessaire car les conditions classiques sont remplies : modèles entièrement paramétriques, estimation par maximum de vraisemblance, et inclusion stricte du modèle restreint dans le modèle général.

Résultats.

- **Groupe 1** : $\Lambda = 1,197$, p-value = 0,550.
- **Groupe 2** : $\Lambda = 1,095$, p-value = 0,578.

Dans les deux cas, la p-value est nettement supérieure à 5 %, ce qui signifie qu'il n'existe pas de différence significative entre les paramètres estimés chez les femmes et les hommes. Ces résultats suggèrent qu'un modèle commun de longévité au-delà de 105 ans est envisageable, bien que les limites liées à la taille des échantillons, notamment chez les hommes, doivent être prises en compte.

Conclusion. Ces deux tests apportent des arguments favorables à la construction d'une table unisexe au-delà de 105 ans. Toutefois, ce résultat doit être nuancé par le déséquilibre des effectifs (par exemple, seulement 45 hommes contre 556 femmes dans le groupe 2), qui donne un poids statistique beaucoup plus élevé à la population féminine. La bonne concordance observée pourrait donc refléter principalement les caractéristiques féminines.

3.6.3 Modèle GPD ajusté sur les deux sexes

Un ajustement a été réalisé sur l'ensemble des individus de chaque groupe (femmes et hommes réunis), en considérant une population homogène pour tester l'hypothèse du modèle unisexe. L'estimation a été effectuée à l'aide de la fonction `fitgpd` du package `POT` dans le logiciel R.

Dans le **groupe 1**, les paramètres estimés sont :

$$\hat{\xi} = -0,126, \quad \text{IC}_{95\%} = [-0,271 ; 0,020], \quad \hat{\beta} = 1,111, \quad \hat{\omega} = 113,84 \text{ ans.}$$

Dans le **groupe 2**, les 601 individus sont également combinés dans un modèle unique. L'ajustement donne :

$$\hat{\xi} = -0,069, \quad \text{IC}_{95\%} = [-0,141 ; 0,003], \quad \hat{\beta} = 1,385, \quad \hat{\omega} = 124,98 \text{ ans.}$$

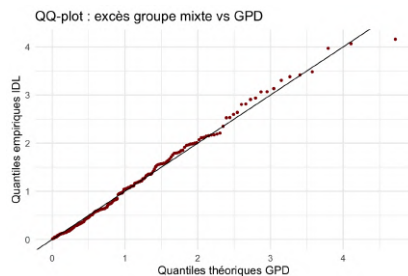


FIGURE 3.15 – Groupe 1 : $\xi = -0,126$, $\beta = 1,111$

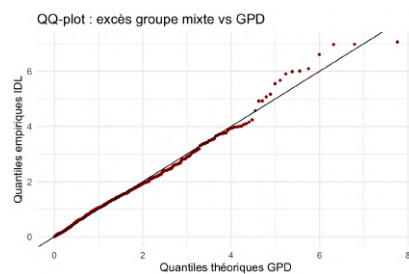


FIGURE 3.16 – Groupe 2 : $\xi = -0,069$, $\beta = 1,385$

Dans les deux groupes, l'ajustement visuel ci-dessus du modèle combiné basé sur la loi GPD est satisfaisant. Toutefois, les intervalles de confiance de ξ

recouvrant 0 laissent une incertitude sur la présence d’une borne supérieure.

3.6.4 Exploration d’un modèle exponentiel au-delà de 110 ans

Nous testons ici la pertinence d’un modèle exponentiel pour les excès observés après 110 ans, en nous inspirant de l’approche de Rootzén et Zholud (2017). Cette analyse est limitée au **groupe 2**, aucun individu du groupe 1 n’ayant atteint cet âge.

Deux variantes sont considérées :

- **Modèle libre** : le paramètre λ est estimé par maximum de vraisemblance ;
- **Modèle contraint** : λ est fixé à $\frac{1}{1,34} \approx 0,746$, comme dans l’étude de Rootzén et Zholud (2017).

Dans chaque cas, un test exact de Kolmogorov–Smirnov est réalisé, après correction des doublons comme précédemment.

TABLE 3.13 – Ajustement d’un modèle exponentiel au-delà de 110 ans (groupe 2)

Méthode	Paramètre λ	p-value du test KS
Estimation par MLE	0,918	0,481
Valeur fixée ($\lambda = 1/1,34$)	0,746	0,568

Conclusion. Le modèle exponentiel n’est pas rejeté par les données IDL au-delà de 110 ans. Toutefois, cette conclusion doit être nuancée compte tenu de la taille très réduite de l’échantillon, qui ne comprend que 12 femmes. Dans le groupe 2, l’ajustement GPD donne $\hat{\xi} = -0,069$, avec un intervalle de confiance $[-0,141 ; 0,003]$. Comme cet intervalle contient la valeur $\xi = 0$, propre au modèle exponentiel, cette hypothèse demeure compatible avec les données observées. En pratique, ce type de modèle a longtemps été utilisé en actuariat pour les rentes viagères, où surestimer la longévité est une approche prudente. À l’inverse, dans le cadre d’une assurance obsèques, il convient d’éviter de sous-estimer la mortalité. Retenir la valeur ponctuelle $\hat{\xi} < 0$ reste donc plus conservateur car elle borne la longévité maximale et conduit à une tarification plus sûre.

Ce travail d’ajustement et de comparaison des modèles fournit désormais une base solide pour la suite de l’étude. Le chapitre suivant portera sur la construction des bases techniques de l’assurance obsèques, en raccordant la table marché BNB (valable jusqu’à 99 ans) aux extrapolations obtenues avec nos modèles GPD, qu’ils soient genrés ou unisexe. L’objectif sera de proposer des tables complètes et cohérentes, utiles à la tarification pratique des contrats.

Chapitre 4

Bases techniques de l'assurance obsèques

4.1 Introduction

Ce chapitre clôture notre étude en mobilisant les outils développés jusqu'ici pour atteindre un objectif concret : **proposer des tables d'expérience adaptées à la tarification de contrats obsèques**. Pour cela, nous construisons trois tables distinctes : une pour les femmes, une pour les hommes et une unisexe. Ces tables sont élaborées à partir de la mortalité observée sur le marché belge, à savoir la table BNB valable jusqu'à 99 ans, et du modèle de Pareto généralisée validé précédemment, extrapolé au-delà de 105 ans grâce aux données IDL.

Ces tables serviront ensuite de base à des illustrations numériques. Nous les utiliserons pour le calcul des primes, la détermination des réserves et une comparaison avec la table réglementaire XK, afin d'évaluer l'impact des hypothèses de mortalité retenues sur les résultats techniques.

4.2 Construction des tables d'expérience

4.2.1 Présentation des données BNB

Pour construire les tables d'expérience, nous avons utilisé les données publiées par la Banque nationale de Belgique (BNB), couvrant la période 2018 à 2022. Cette base contient des statistiques individuelles et agrégées ventilées par âge (de 0 à 99 ans), par sexe, et par type d'assurance. Elle distingue les contrats en cas de décès et les contrats en cas de vie, pour lesquels sont renseignés les effectifs assurés et les décès observés.

Dans le cadre de cette étude, seule la composante liée aux décès a été ex-

plottée, en cohérence avec la nature du produit examiné, à savoir l'assurance obsèques. L'analyse se concentre sur les quotients de mortalité observés entre 60 et 99 ans, calculés en rapportant le nombre de décès au nombre d'assurés à chaque âge.

Les probabilités de décès ainsi calculées présentent toutefois d'importantes irrégularités, en particulier aux âges élevés, où les produits décès (hors obsèques) sont très peu souscrits. Ces fluctuations, dues à la faiblesse des effectifs dans certaines tranches d'âge, nuisent à la stabilité des taux et rendent leur utilisation directe difficile dans un cadre actuariel.

Afin d'obtenir des valeurs cohérentes et plus lissées, nous avons donc choisi d'ajuster la loi de Makeham sur l'intervalle 60 à 99 ans. Le graphique ci-après illustre la variabilité des taux bruts selon le sexe (hommes, femmes, mixte) avant modélisation.

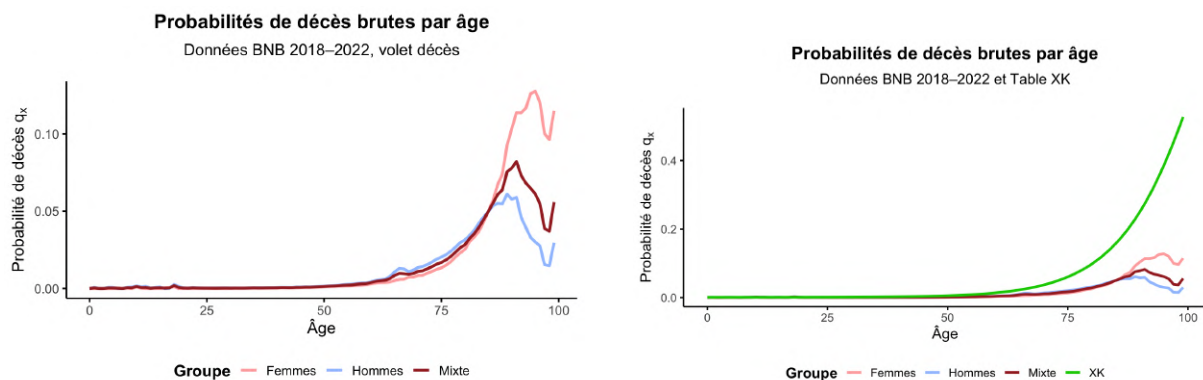


FIGURE 4.1 – Probabilités de décès brutes entre 0 et 99 ans : à gauche selon le sexe (femmes, hommes, mixte), à droite avec ajout de la table réglementaire XK.

Les valeurs numériques correspondantes, de 60 à 99 ans, sont reprises en annexe (voir tableau B.1).

4.2.2 Ajustement par le modèle de Makeham

Pour stabiliser les probabilités brutes de décès observées entre 60 et 99 ans, nous avons retenu le modèle de Makeham, couramment utilisé en modélisation actuarielle. Ce modèle approxime directement q_x , la probabilité annuelle de décès à l'âge x , par la formule :

$$q_x = 1 - \exp \left(-A - \frac{B}{\log(c)} (c^{x+1} - c^x) \right)$$

où A , B et c sont des paramètres à estimer.

L'ajustement a été réalisé séparément pour les femmes, les hommes et le groupe mixte, à l'aide de la méthode des moindres carrés non linéaires (fonction `nls` dans R), sur la base des taux bruts effectivement observés.

Dans les trois cas, le paramètre A a été estimé à zéro, ce qui revient à retenir une forme exponentielle de la mortalité, telle que décrite par le modèle de Gompertz.

Les probabilités ajustées selon cette modélisation sont conservées jusqu'à l'âge de 94 ans. À partir de 105 ans, la mortalité est déterminée par le modèle GPD validé précédemment. Afin d'assurer une jonction fluide et une progression cohérente entre ces deux régimes, l'intervalle intermédiaire [95 ; 104] fait l'objet d'un raccordement progressif par spline cubique.

4.2.3 Extrapolation au-delà de 105 ans avec la loi GPD

L'extrapolation de la mortalité au-delà de 105 ans repose sur le modèle GPD, validé au chapitre précédent. Pour l'élaboration de nos tables, nous avons choisi d'utiliser exclusivement le groupe 2 de la base IDL, comme présenté au chapitre 3. Ce choix s'appuie sur plusieurs éléments. Le groupe 2 est plus récent, car il intègre les cohortes de naissance de 1911 à 1915, absentes du groupe 1. De plus, les tests statistiques menés ont montré une meilleure adéquation au modèle GPD et des résultats plus stables sur ce sous-échantillon. La GPD ajustée sur ce groupe sera donc utilisée pour prolonger nos tables aux âges extrêmes et assurer le raccordement avec la table BNB observée entre 60 et 99 ans.

La probabilité de décès entre un âge x et l'âge suivant $x + 1$, pour $x \geq 105$, est calculée à partir de la fonction de survie :

$$q_x = \frac{S(x) - S(x+1)}{S(x)} \quad \text{avec} \quad S(x) = \left(1 + \xi \frac{x - 105}{\beta}\right)^{-1/\xi}$$

où ξ et β sont les paramètres estimés sur le groupe 2.

Cette formulation permet de prolonger la table de mortalité au-delà de 105 ans, jusqu'à l'âge ultime ω défini pour chaque groupe (hommes, femmes, mixte).

4.2.4 Raccordement entre les modèles Makeham et GPD

Pour assurer une transition fluide et continue entre les probabilités de décès obtenues par le modèle de Makeham (utilisé jusqu'à 94 ans) et celles extrapolées par la loi GPD à partir de 105 ans, nous utilisons une spline cubique de Hermite sur l'intervalle [95, 105].

Cette méthode est retenue car elle garantit non seulement la continuité de la fonction q_x , mais aussi celle de sa dérivée première. Elle permet ainsi d'éviter toute rupture abrupte ou fluctuation non réaliste dans l'évolution des probabilités de décès, assurant un profil de mortalité régulier et cohérent.

Les conditions imposées aux bornes de l'intervalle sont les suivantes :

- À 95 ans, on utilise la valeur q_{95} issue du modèle de Makeham, avec une pente locale estimée par $m_0 = q_{96} - q_{95}$;
- À 105 ans, on retient la valeur q_{105} issue de la GPD, avec une pente approchée par $m_1 = q_{106} - q_{105}$.

La spline est définie par l'expression suivante :

$$q_x = h_{00}(t) y_0 + h_{10}(t) (x_1 - x_0) m_0 + h_{01}(t) y_1 + h_{11}(t) (x_1 - x_0) m_1$$

où $x_0 = 95$, $x_1 = 105$, $y_0 = q_{95}$, $y_1 = q_{105}$, et $t = \frac{x - x_0}{x_1 - x_0}$.

Les fonctions de base cubiques sont données par :

$$h_{00}(t) = 2t^3 - 3t^2 + 1$$

$$h_{10}(t) = t^3 - 2t^2 + t$$

$$h_{01}(t) = -2t^3 + 3t^2$$

$$h_{11}(t) = t^3 - t^2$$

Cette approche assure un raccordement harmonieux et mathématiquement rigoureux entre les deux modèles sur l'ensemble de l'intervalle [95, 105].

4.2.5 Présentation des tables construites

À l'issue des ajustements réalisés, trois tables de mortalité sont obtenues : une table féminine, une table masculine et une table unisexe. Elles partagent toutes une structure commune, combinant les quotients de mortalité issus du modèle de Makeham jusqu'à 94 ans (ajusté sur les données BNB 2018–2022), une interpolation par spline de Hermite construite sur l'intervalle [95, 105] et utilisée pour les âges intermédiaires (96 à 104 ans), puis une extrapolation à partir de 105 ans selon le modèle GPD.

Les courbes de mortalité correspondantes, présentées ci-dessous, illustrent l'évolution des probabilités de décès q_x sur l'ensemble des âges. Elles mettent en évidence la cohérence des ajustements retenus et la continuité assurée entre les différents segments du raccordement.

Les valeurs numériques associées, de 60 ans jusqu'à l'âge ultime, sont reprises en annexe pour les différentes tables (féminine, masculine et unisexe), ainsi que pour la table réglementaire XK actuellement utilisée pour la tarification des contrats obsèques en Belgique (voir tableau B.2).

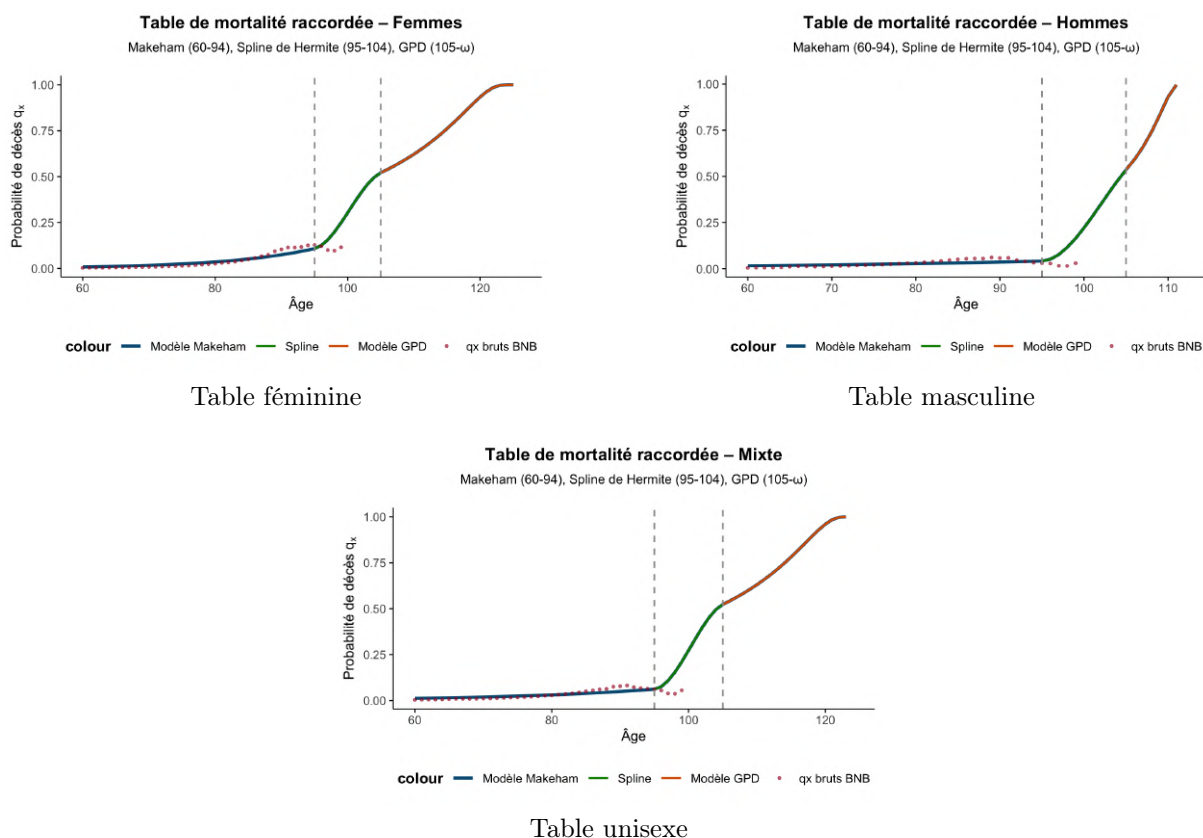


FIGURE 4.2 – Tables d'expérience (féminine, masculine, unisexe).

Les trois tables ainsi établies remplissent des fonctions distinctes. Les tables genrées, féminine et masculine, permettent une évaluation plus fidèle des passifs, en intégrant les différences de mortalité entre hommes et femmes. Elles sont donc particulièrement adaptées aux estimations internes des provisions techniques, lorsque l'objectif est d'obtenir une vision réaliste des engagements futurs.

En matière de tarification, en revanche, l'usage d'une table unisexe s'impose. Cette obligation découle de l'article 10, § 4 de la loi du 10 mai 2007 tendant à lutter contre la discrimination entre hommes et femmes (modifiée par la loi du 19 décembre 2012), qui interdit toute différenciation tarifaire fondée sur le sexe, y compris lorsque l'on recourt à une table d'expérience.

Par ailleurs, l'article 24, § 7 de l'arrêté royal du 14 novembre 2003 relatif à l'activité de l'assurance sur la vie encadre l'usage de tables d'expérience pour la tarification : les taux retenus ne peuvent être garantis que pour une durée maximale de trois ans, et uniquement s'ils ne sont pas inférieurs à ceux observés sur le marché belge au cours des cinq dernières années. La table unisexe obtenue dans cette étude, construite à partir des données BNB 2018–2022 et prolongée par un modèle validé, satisfait à ces exigences. Elle constitue donc une base conforme pour le calcul de primes non garanties.

4.3 Illustration numérique

Cette section examine l'impact des tables construites sur les résultats techniques. Nous illustrons leur usage à travers le calcul de primes et de réserves mathématiques. Les résultats sont comparés à ceux issus de la table réglementaire XK.

4.3.1 Hypothèses et cadre de calcul

Nous considérons un contrat d'assurance vie entière individuelle à prime unique, versant un capital constant de 1 euro au moment du décès de l'assuré. Le paiement du capital intervient en fin d'année, avec un taux d'intérêt technique fixé à $i = 1\%$.

Les calculs sont effectués pour différents âges d'entrée dans le contrat : 60, 65 et 70 ans. Pour chacun de ces âges, nous déterminons la prime unique nette correspondant à un capital assuré de 1 euro et la réserve mathématique à différentes échéances suivant l'entrée dans le contrat.

4.3.2 Primes uniques et réserves mathématiques

La prime unique nette est calculée pour un capital versé en fin d'année, en actualisant par v^{k+1} , avec $v = \frac{1}{1+i}$.

Cette prime correspond à la valeur actuelle des prestations futures. Pour un assuré âgé de x à l'entrée dans le contrat, elle est donnée par :

$$\text{Prime}_x = \sum_{k=0}^{\omega-x-1} {}_k p_x \cdot q_{x+k} \cdot v^{k+1}$$

où :

- q_{x+k} est le quotient de mortalité à l'âge $x+k$, selon la table utilisée ;
- ${}_k p_x$ est la probabilité de survie de x à $x+k$, donnée par $\prod_{j=0}^{k-1} (1 - q_{x+j})$.

Les calculs sont réalisés à l'aide du logiciel R, pour chaque combinaison d'âge d'entrée (60, 65 et 70 ans) et de table utilisée (féminine, masculine, unisexe, XK). Les résultats sont synthétisés dans le tableau ci-dessous.

Âge	T ^F	T ^H	T ^U	T ^{XK}
60	0,7745	0,7685	0,7710	0,8354
65	0,8037	0,7893	0,7958	0,8653
70	0,8320	0,8104	0,8201	0,8929

TABLE 4.1 – Primes uniques nettes pour un capital de 1 euro, selon l'âge d'entrée et la table utilisée

Afin de comparer l'impact des différentes tables de mortalité, nous avons également calculé la réserve mathématique, définie comme la valeur actualisée

des engagements futurs. Ce calcul repose sur l'hypothèse d'une prime unique versée à l'âge de 65 ans.

Dans ce cadre, comme l'assuré ne paie plus aucune prime après cette date, la réserve à un âge donné correspond à la prime unique qu'un nouvel assuré du même âge devrait verser pour bénéficier de la même couverture.

Le graphique ci-dessous montre l'évolution de cette réserve pour chacune des quatre tables.

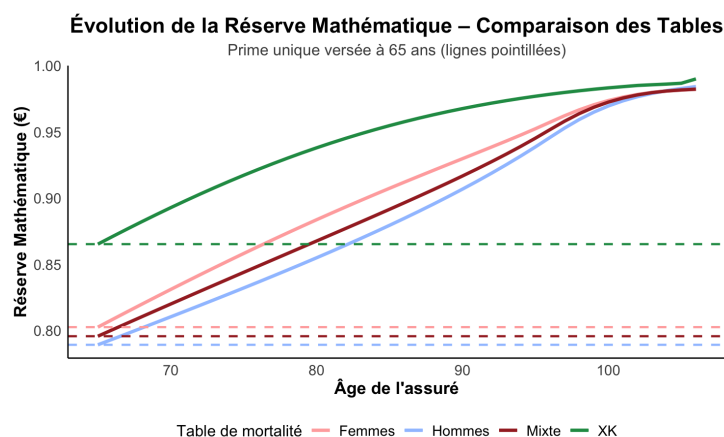


FIGURE 4.3 – Évolution des réserves mathématiques pour un assuré de 65 ans selon la table utilisée.

Conclusion. Les résultats montrent que les primes et les réserves issues des tables d'expérience sont systématiquement inférieures à celles de la table réglementaire XK.

L'écart est matériel, avec une prime unique plus faible de 7 à 10,% selon l'âge d'entrée et la table d'expérience utilisée. Cette différence s'explique à la fois par l'absence de marge de sécurité dans les tables construites et par leur horizon de projection, étendu bien au-delà des 106 ans de la table réglementaire XK.

Les tables genrées, extrapolées jusqu'à 126 ans pour les femmes et 112 ans pour les hommes, conduisent elles aussi à des primes et des réserves plus faibles. Elles permettent donc d'ajuster plus finement l'évaluation des passifs au profil de mortalité de l'assuré, sans recourir à des hypothèses excessivement conservatrices.

Dans l'ensemble, les tables d'expérience construites dans cette étude, qu'elles soient unisexe ou genrées, offrent une base plus réaliste pour la tarification et l'évaluation des réserves. Elles permettent de proposer des tarifs plus compétitifs et de mieux ajuster les engagements de l'assureur, en s'écartant du cadre trop prudent imposé par la table XK.

Conclusion

Ce mémoire s’est intéressé à la problématique cruciale de la modélisation de la mortalité aux âges avancés, spécifiquement dans le cadre de l’assurance obsèques. Face aux limites avérées des tables réglementaires classiques, qui peinent à refléter fidèlement la dynamique de la mortalité à ces âges extrêmes, l’objectif principal était d’affiner l’évaluation du risque de décès, composante indispensable pour une tarification adéquate des produits de type vie entière.

Notre démarche a d’abord exploré le modèle de Pareto généralisée proposé par Gbari et al. (2017). Ce cadre théorique, reconnu pour sa robustesse et bâti sur des cohortes belges complètes, offrait une base prometteuse pour l’extrapolation de la mortalité. Toutefois, son application directe aux données IDL les plus récentes s’est heurtée à la présence de cohortes incomplètes, rendant impossible une validation rigoureuse du modèle tel quel. Cette contrainte a conduit à une adaptation méthodologique, consistant à réestimer les paramètres du modèle à partir du seuil de 105 ans, afin d’en préserver la structure tout en garantissant une cohérence statistique.

Cette démarche a permis de construire trois tables d’expérience féminine, masculine et unisexe, obtenues par raccordement aux quotients de mortalité observés sur le marché belge (BNB 2018–2022). Ces tables prolongent la mortalité jusqu’à des âges ultimes réalistes et offrent une représentation plus nuancée du risque de décès aux grands âges. Leur comparaison avec la table réglementaire XK a mis en évidence l’existence d’un chargement de sécurité significatif dans cette dernière, qui conduit à une surévaluation systématique du risque et à une sur-tarification des opérations en cas de décès aux âges avancés. Elle doit, dans ce contexte, être utilisée avec prudence, car elle ne reflète pas fidèlement la mortalité des assurés.

Les tables construites dans cette étude fournissent aux assureurs un outil rigoureux et opérationnel pour mieux piloter leurs engagements à très long terme. Elles permettent de garantir, dans les limites réglementaires de trois ans, des tarifs plus compétitifs fondés sur une mortalité extrapolée de manière réaliste, contribuant ainsi à renforcer la capacité concurrentielle des entreprises.

En définitive, ce mémoire démontre qu'il est possible, à partir de données empiriques disponibles, de construire des tables viagères extrapolées fiables, cohérentes, et mieux adaptées aux besoins spécifiques de l'assurance obsèques. Il souligne également l'importance de maîtriser cette composante technique pour garantir une tarification pertinente et durable. Bien que certaines réserves subsistent, notamment en raison du faible effectif masculin et de l'incomplétude des cohortes les plus récentes, cette approche offre des perspectives prometteuses. Elle concilie rigueur actuarielle et adaptation aux réalités observées, tout en s'affranchissant partiellement d'un cadre prudentiel parfois trop conservateur. À l'avenir, il serait pertinent d'étudier l'impact de ces nouvelles tables de mortalité d'expérience sur les exigences de solvabilité.

Bibliographie

- [1] Gbari, S., Poulain, M., Dal, L. et Denuit, M. *Extreme value analysis of mortality at the oldest ages : A case study based on individual ages at death*. North American Actuarial Journal, 21(3), 397–416, 2017.
- [2] Denuit, M., Hainaut, D. et Trufin, J. *Effective Statistical Learning Methods for Actuaries. Volume 1 : GLM and Extensions*. Springer Actuarial Lecture Notes Series, 2019. (Chapitre consacré à la théorie des valeurs extrêmes.)
- [3] Gbari, S., Denuit, M. et Trufin, J. *Quand Vilfredo Pareto rencontre Jeanne Calment : analyse de la mortalité aux âges extrêmes en Belgique*. Note technique Detralytics 2017-1, janvier 2017.
- [4] Hoang Khanh Dang, L., Ouellette, N., Meslé, F. et Poulain, M. *Trajectoire de mortalité aux âges extrêmes de la vie : une représentation paramétrique des données récentes de la France, de la Belgique et du Québec*. Revue Quetelet, 9(3), 2021. <https://doi.org/10.14428/rqj2021.09.01.03>
- [5] Reiss, R.-D. et Thomas, M. *Statistical Analysis of Extreme Values : With Applications to Insurance, Finance, Hydrology and Other Fields*. 3rd edition, Birkhäuser, Basel, 2007.
- [6] Rootzén, H. et Zholud, D. *Human life is unlimited – but short*. Extremes, 20(4), 713–728, 2017. <https://doi.org/10.1007/s10687-017-0305-5>
- [7] Huang, F., Maller, R. et Ning, X. *Modelling Life Tables with Advanced Ages : An Extreme Value Theory Approach*. Insurance : Mathematics and Economics, 93, 95–115, 2020.
- [8] Denuit, Michel. *Bases techniques de l’assurance-vie individuelle en Belgique*. Revue générale des assurances et des responsabilités, vol. 72, no. 4, pp. 13078/1–13078/11, 1999.
- [9] Denuit, M. *LACTU2230 – Actuariat des assurances de personnes*. Notes de cours, année académique 2023–2024, UCLouvain, Louvain-la-Neuve.

- [10] Yan, L. *Extreme Value Analysis of Mortality at Age Beyond 105*. Mémoire de master, Faculté des sciences, Université catholique de Louvain, 2020. Promoteur : M. Denuit. <http://hdl.handle.net/2078.1/thesis:27132>
- [11] Hariri, A. *Aménagement d'un produit d'assurance obsèques*. Mémoire de fin d'études, Institut de Statistique de l'Université de Paris (ISUP), présenté le 22 mai 2015.
- [12] International Database on Longevity (IDL). Données relatives à la Belgique (1996–2021), extraites en octobre 2024 depuis <https://www.supercentenarians.org/en/data-and-metadata/>
- [13] Banque nationale de Belgique. *Statistique concernant le nombre des personnes assurées et le nombre de décès*. Données extraites en Mars 2025 depuis <https://www.nbb.be/fr/supervision-financiere/controle-prudentiel/domaines-de-control-e/entreprises-dassurance-ou-de-7>
- [14] RTL Info. *Après le décès de Fernande Courtoy, voici l'identité de la nouvelle doyenne des Belges*. Consulté le 04/07/25. <https://www.rtl.be/actu/belgique/societe/apres-le-deces-de-fernande-courtoy-voici-lidentite-de-la-nouvelle-doyenne-des/2024-05-30/article/674265>
- [15] L'Avenir. *On a retrouvé la trace de la doyenne des Belges en Corse : voici son portrait*. Consulté le 04/07/25. <https://www.lavenir.net/actu/belgique/2025/03/03/on-a-retrouve-la-trace-de-la-doyenne-des-belges-en-corse-voici-son-portrait-II0IVM2QQBBQFHTPGQUCYJQUCM/>

Annexe A

Doublons sur l'âge exact au décès

IDL_ID	Sexe	Naissance	Décès	Jours	Âge
BEL21905201000417	F	1905-08-03	2010-08-21	38369	105.0493
BEL21905201000424	F	1905-09-09	2010-09-27	38369	105.0493
BEL21909201500590	F	1909-12-13	2015-01-02	38371	105.0548
BEL21909201400601	F	1909-02-08	2014-02-28	38371	105.0548
BEL21907201200524	F	1907-09-28	2012-10-27	38381	105.0795
BEL21908201300965	M	1908-01-21	2013-02-19	38381	105.0795
BEL21905201000370	F	1905-10-24	2010-12-09	38397	105.1260
BEL21909201400632	F	1909-07-06	2014-08-21	38397	105.1260
BEL21909201400639	F	1909-08-26	2014-10-11	38397	105.1260
BEL21908201300544	F	1908-02-27	2013-04-14	38398	105.1260
BEL21906201100454	F	1906-02-19	2011-04-12	38403	105.1425
BEL21909201400626	F	1909-06-29	2014-08-20	38403	105.1425
BEL21908201400532	F	1908-11-30	2014-01-22	38404	105.1452
BEL21908201300555	F	1908-03-31	2013-05-23	38404	105.1452
BEL21909201500587	F	1909-11-25	2015-01-25	38412	105.1671
BEL21908201300964	M	1908-01-19	2013-03-21	38413	105.1671
BEL21909201400642	F	1909-09-14	2014-11-16	38414	105.1726
BEL21908201300537	F	1908-01-25	2013-03-29	38415	105.1726
BEL21908201300568	F	1908-06-08	2013-08-18	38422	105.1945
BEL21907201300484	F	1907-12-01	2013-02-10	38423	105.1945
BEL21905201000401	F	1905-04-22	2010-08-05	38456	105.2877
BEL21907201300486	F	1907-12-17	2013-04-01	38457	105.2877
BEL21907201200502	F	1907-03-31	2012-07-14	38457	105.2877
BEL21905201100409	F	1905-06-17	2011-07-04	38733	106.0464
BEL21906201200434	F	1906-01-10	2012-01-27	38733	106.0464
BEL21908201500533	F	1908-11-03	2015-03-13	38846	106.3562
BEL21908201400560	F	1908-05-10	2014-09-17	38846	106.3562

TABLE A.1 – Liste des doublons sur l'âge exact au décès (même durée en jours vécus)

Annexe B

Tables détaillées des probabilités de décès q_x

B.1 Probabilités de décès brutes (BNB)

Âge	q_x^F	q_x^H	q_x^U	Âge	q_x^F	q_x^H	q_x^U
60	0.0031	0.0044	0.0039	80	0.0253	0.0313	0.0283
61	0.0035	0.0051	0.0046	81	0.0303	0.0344	0.0324
62	0.0037	0.0056	0.0049	82	0.0332	0.0377	0.0355
63	0.0038	0.0061	0.0053	83	0.0370	0.0423	0.0396
64	0.0041	0.0077	0.0064	84	0.0429	0.0463	0.0446
65	0.0051	0.0107	0.0084	85	0.0500	0.0497	0.0499
66	0.0059	0.0129	0.0097	86	0.0571	0.0536	0.0553
67	0.0060	0.0127	0.0095	87	0.0672	0.0550	0.0608
68	0.0073	0.0109	0.0092	88	0.0740	0.0548	0.0638
69	0.0073	0.0119	0.0097	89	0.0927	0.0611	0.0755
70	0.0085	0.0136	0.0111	90	0.1035	0.0577	0.0779
71	0.0085	0.0142	0.0114	91	0.1138	0.0590	0.0822
72	0.0099	0.0155	0.0128	92	0.1136	0.0460	0.0732
73	0.0111	0.0171	0.0141	93	0.1167	0.0395	0.0682
74	0.0124	0.0190	0.0157	94	0.1260	0.0331	0.0648
75	0.0134	0.0203	0.0168	95	0.1277	0.0301	0.0613
76	0.0151	0.0219	0.0185	96	0.1204	0.0276	0.0549
77	0.0173	0.0242	0.0207	97	0.1001	0.0154	0.0388
78	0.0204	0.0267	0.0235	98	0.0963	0.0146	0.0371
79	0.0230	0.0297	0.0263	99	0.1151	0.0296	0.0558

TABLE B.1 – Récapitulatif des probabilités de décès brutes q_x entre 60 et 99 ans.

B.2 Probabilités de décès issues des différentes tables

$\hat{\text{Age}}$	q_x^F	q_x^H	q_x^U	q_x^{XK}	$\hat{\text{Age}}$	q_x^F	q_x^H	q_x^U	q_x^{XK}
60	0.0081	0.0156	0.0121	0.0140	93	0.0929	0.0390	0.0565	0.3257
61	0.0087	0.0160	0.0127	0.0154	94	0.1001	0.0401	0.0592	0.3548
62	0.0094	0.0165	0.0133	0.0169	95	0.1078	0.0412	0.0621	0.3856
63	0.0101	0.0169	0.0140	0.0186	96	0.1245	0.0510	0.0757	0.4184
64	0.0108	0.0174	0.0146	0.0205	97	0.1558	0.0764	0.1078	0.4529
65	0.0117	0.0179	0.0153	0.0227	98	0.1982	0.1149	0.1542	0.4890
66	0.0126	0.0184	0.0161	0.0250	99	0.2483	0.1638	0.2107	0.5270
67	0.0135	0.0189	0.0168	0.0275	100	0.3028	0.2207	0.2729	0.5667
68	0.0146	0.0195	0.0176	0.0304	101	0.3583	0.2830	0.3367	0.6027
69	0.0157	0.0200	0.0185	0.0335	102	0.4112	0.3482	0.3979	0.6517
70	0.0169	0.0206	0.0193	0.0369	103	0.4583	0.4138	0.4522	0.6774
71	0.0182	0.0211	0.0203	0.0408	104	0.4962	0.4771	0.4954	0.7000
72	0.0196	0.0217	0.0212	0.0450	105	0.5213	0.5356	0.5232	0.6667
73	0.0211	0.0224	0.0222	0.0496	106	0.5389	0.5910	0.5419	1.0000
74	0.0227	0.0230	0.0233	0.0547	107	0.5577	0.6575	0.5619	–
75	0.0245	0.0236	0.0244	0.0603	108	0.5778	0.7374	0.5834	–
76	0.0264	0.0243	0.0256	0.0665	109	0.5991	0.8313	0.6063	–
77	0.0284	0.0250	0.0268	0.0733	110	0.6219	0.9306	0.6308	–
78	0.0306	0.0257	0.0281	0.0808	111	0.6463	0.9958	0.6572	–
79	0.0329	0.0264	0.0294	0.0890	112	0.6723	–	0.6854	–
80	0.0355	0.0272	0.0308	0.0980	113	0.7001	–	0.7155	–
81	0.0382	0.0279	0.0323	0.1079	114	0.7297	–	0.7477	–
82	0.0411	0.0287	0.0339	0.1187	115	0.7610	–	0.7819	–
83	0.0443	0.0295	0.0355	0.1306	116	0.7942	–	0.8178	–
84	0.0477	0.0304	0.0372	0.1436	117	0.8287	–	0.8550	–
85	0.0514	0.0312	0.0389	0.1578	118	0.8643	–	0.8924	–
86	0.0553	0.0321	0.0408	0.1733	119	0.8997	–	0.9285	–
87	0.0596	0.0330	0.0427	0.1902	120	0.9336	–	0.9605	–
88	0.0642	0.0339	0.0448	0.2085	121	0.9633	–	0.9846	–
89	0.0691	0.0349	0.0469	0.2285	122	0.9854	–	0.9972	–
90	0.0744	0.0359	0.0492	0.2501	123	0.9972	–	1.0000	–
91	0.0801	0.0369	0.0515	0.2735	124	0.9999	–	–	–
92	0.0863	0.0379	0.0540	0.2987	125	1.0000	–	–	–

TABLE B.2 – Récapitulatif des probabilités de décès q_x des différentes tables.

Annexe C

Code R

C.1 Analyse exploratoire des âges au décès

Analyse basée sur les âges révolus — Groupe 1

```
# *****  
#   Analyse stat ages revolus G1  
# *****  
  
library(readxl)  
library(dplyr)  
library(ggplot2)  
library(tidyr)  
  
# Chargement des donnees  
df <- read_excel("/Users/leilasanogo/Documents/memoires moi/idl_bel_death_105_  
  analyse.xlsx", sheet = "new")  
  
# Statistiques descriptives globales par sexe  
desc_stats_sex <- df %>%  
  group_by(SEX) %>%  
  summarise(  
    n = n(),  
    min_age = min(AGEYEARS),  
    q1 = quantile(AGEYEARS, 0.25),  
    median = median(AGEYEARS),  
    mean = mean(AGEYEARS),  
    q3 = quantile(AGEYEARS, 0.75),  
    max_age = max(AGEYEARS),  
    sd = sd(AGEYEARS),  
    .groups = "drop"  
  )  
  
# Histogrammes par sexe  
ggplot(df %>% filter(SEX == "F"), aes(x = AGEYEARS)) +  
  geom_histogram(binwidth = 1, fill = "salmon", color = "black") +  
  labs(title = "Histogramme des ages au deces - Femmes",  
    x = "Age au deces (annees entieres)", y = "Effectif") +  
  theme_minimal(base_size = 14)  
  
ggplot(df %>% filter(SEX == "M"), aes(x = AGEYEARS)) +  
  geom_histogram(binwidth = 1, fill = "lightblue", color = "black") +  
  labs(title = "Histogramme des ages au deces - Hommes",  
    x = "Age au deces (annees entieres)", y = "Effectif") +  
  theme_minimal(base_size = 14)
```

```

# Boxplot
ggplot(df, aes(x = SEX, y = AGEYEARS, fill = SEX)) +
  geom_boxplot() +
  labs(title = "Boxplot des ages au deces par sexe", x = "Sexe", y = "Age au deces")
  +
  theme_minimal()

# Tableau des cohortes 1905 a 1915
cohort_table <- df %>%
  filter(BIRTH_YEAR >= 1905, BIRTH_YEAR <= 1915) %>%
  group_by(BIRTH_YEAR, SEX) %>%
  summarise(
    L105 = n(),
    m105 = max(AGEYEARS),
    `105+e105` = mean(AGEYEARS),
    .groups = "drop"
  ) %>%
  pivot_wider(
    names_from = SEX,
    values_from = c(L105, m105, `105+e105`),
    names_glue = "{.value}_{SEX}"
  ) %>%
  arrange(BIRTH_YEAR)

```

Analyse basée sur les âges exacts — Groupe 1

```

# *****
#   Analyse stat ages exacts G1
# *****

library(readxl)
library(dplyr)
library(ggplot2)
library(tidyr)
library(lubridate)

# Chargement des donnees
df <- read_excel("/Users/leilasanogo/Documents/memoires moi/idl_bel_death_105_
  analyse.xlsx", sheet = "new")

# Creation des dates de naissance et de deces
df <- df %>%
  mutate(
    DateNaissance = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),
    DateDeces     = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),
    AgeExact = time_length(interval(DateNaissance, DateDeces), unit = "years")
  )

# Correction des doublons
df <- df %>%
  arrange(AgeExact, IDL_ID) %>%
  group_by(AgeExact) %>%
  mutate(
    n = n(),
    AgeExact_unique = if_else(n > 1, AgeExact + row_number() * 1e-10, AgeExact)
  ) %>%
  ungroup() %>%
  select(-n)

# Statistiques descriptives globales par sexe
desc_stats_exact <- df %>%
  group_by(SEX) %>%

```

```

summarise(
  n = n(),
  min_age = min(AgeExact),
  q1 = quantile(AgeExact, 0.25),
  median = median(AgeExact),
  mean = mean(AgeExact),
  q3 = quantile(AgeExact, 0.75),
  max_age = max(AgeExact),
  sd = sd(AgeExact),
  .groups = "drop"
)

# Histogrammes
ggplot(df %>% filter(SEX == "F"), aes(x = AgeExact)) +
  geom_histogram(binwidth = 0.1, fill = "salmon", color = "black") +
  labs(title = "Histogramme des ages exacts - Femmes", x = "Age exact", y = "
    Effectif") +
  theme_minimal(base_size = 14)

ggplot(df %>% filter(SEX == "M"), aes(x = AgeExact)) +
  geom_histogram(binwidth = 0.1, fill = "lightblue", color = "black") +
  labs(title = "Histogramme des ages exacts - Hommes", x = "Age exact", y = "
    Effectif") +
  theme_minimal(base_size = 14)

# Boxplot
ggplot(df, aes(x = SEX, y = AgeExact, fill = SEX)) +
  geom_boxplot() +
  labs(title = "Boxplot des ages exacts au deces", x = "Sexe", y = "Age exact") +
  theme_minimal()

# Tableau des cohortes 1905 a 1915
cohort_exact <- df %>%
  filter(BIRTH_YEAR >= 1905, BIRTH_YEAR <= 1915) %>%
  group_by(BIRTH_YEAR, SEX) %>%
  summarise(
    L105 = n(),
    m105 = max(AgeExact),
    `105+e105` = mean(AgeExact),
    .groups = "drop"
  ) %>%
  pivot_wider(
    names_from = SEX,
    values_from = c(L105, m105, `105+e105`),
    names_glue = "{.value}_{SEX}"
  ) %>%
  arrange(BIRTH_YEAR)

```

Analyse basée sur les âges révolus — Groupe 2

```

# *****
# Analyse ages revolus version filttree G2
# *****

library(readxl)
library(dplyr)
library(ggplot2)
library(tidyr)

# Chargement des donnees depuis la feuille "IDL_BEL_Death_105"
df <- read_excel(
  "/Users/leilasanogo/Documents/memoires moi/idl_bel_death_105_analyse.xlsx",
  sheet = "IDL_BEL_Death_105"
)

```

```

# Suppression des individus nes entre 1886 et 1904
df <- df %>% filter(BIRTH_YEAR < 1886 | BIRTH_YEAR > 1904)

# Statistiques descriptives globales par sexe
desc_stats_sex <- df %>%
  group_by(SEX) %>%
  summarise(
    n = n(),
    min_age = min(AGEYEARS),
    q1 = quantile(AGEYEARS, 0.25),
    median = median(AGEYEARS),
    mean = mean(AGEYEARS),
    q3 = quantile(AGEYEARS, 0.75),
    max_age = max(AGEYEARS),
    sd = sd(AGEYEARS),
    .groups = "drop"
  )

# Histogrammes par sexe
ggplot(df %>% filter(SEX == "F"), aes(x = AGEYEARS)) +
  geom_histogram(binwidth = 1, fill = "salmon", color = "black") +
  labs(title = "Histogramme des ages au deces - Femmes",
    x = "Age au deces (annees entieres)", y = "Effectif") +
  theme_minimal(base_size = 14)

ggplot(df %>% filter(SEX == "M"), aes(x = AGEYEARS)) +
  geom_histogram(binwidth = 1, fill = "lightblue", color = "black") +
  labs(title = "Histogramme des ages au deces - Hommes",
    x = "Age au deces (annees entieres)", y = "Effectif") +
  theme_minimal(base_size = 14)

# Boxplot
ggplot(df, aes(x = SEX, y = AGEYEARS, fill = SEX)) +
  geom_boxplot() +
  labs(title = "Boxplot des ages au deces par sexe", x = "Sexe", y = "Age au deces")
  +
  theme_minimal()

# Tableau des cohortes 1905 a 1915
cohort_table <- df %>%
  filter(BIRTH_YEAR >= 1905, BIRTH_YEAR <= 1915) %>%
  group_by(BIRTH_YEAR, SEX) %>%
  summarise(
    L105 = n(),
    m105 = max(AGEYEARS),
    `105+e105` = mean(AGEYEARS),
    .groups = "drop"
  ) %>%
  pivot_wider(
    names_from = SEX,
    values_from = c(L105, m105, `105+e105`),
    names_glue = "{.value}_{SEX}"
  ) %>%
  arrange(BIRTH_YEAR)

```

Analyse basée sur les âges exacts — Groupe 2

```

# *****
# Analyse ages exacts version filtree G2
# *****

library(readxl)

```

```

library(dplyr)
library(ggplot2)
library(tidyr)
library(lubridate)

# Chargement des donnees (feuille IDL_BEL_Death_105)
df <- read_excel(
  "/Users/leilasanogo/Documents/memoires moi/idl_bel_death_105_analyse.xlsx",
  sheet = "IDL_BEL_Death_105"
)

# Retrait des individus nes entre 1886 et 1904
df <- df %>% filter(BIRTH_YEAR < 1886 | BIRTH_YEAR > 1904)

# Creation des dates + age exact
df <- df %>%
  mutate(
    DateNaissance = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),
    DateDeces      = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),
    AgeExact = time_length(interval(DateNaissance, DateDeces), unit = "years")
  )

# Correction doublons
df <- df %>%
  arrange(AgeExact, IDL_ID) %>%
  group_by(AgeExact) %>%
  mutate(
    n = n(),
    AgeExact_unique = if_else(n > 1, AgeExact + row_number() * 1e-10, AgeExact)
  ) %>%
  ungroup() %>%
  select(-n)

# Statistiques descriptives par sexe
desc_stats_exact <- df %>%
  group_by(SEX) %>%
  summarise(
    n = n(),
    min_age = min(AgeExact),
    q1 = quantile(AgeExact, 0.25),
    median = median(AgeExact),
    mean = mean(AgeExact),
    q3 = quantile(AgeExact, 0.75),
    max_age = max(AgeExact),
    sd = sd(AgeExact),
    .groups = "drop"
  )

# Histogrammes
ggplot(df %>% filter(SEX == "F"), aes(x = AgeExact)) +
  geom_histogram(binwidth = 0.1, fill = "salmon", color = "black") +
  labs(title = "Histogramme des ages exacts - Femmes", x = "Age exact", y = "Effectif") +
  theme_minimal(base_size = 14)

ggplot(df %>% filter(SEX == "M"), aes(x = AgeExact)) +
  geom_histogram(binwidth = 0.1, fill = "lightblue", color = "black") +
  labs(title = "Histogramme des ages exacts - Hommes", x = "Age exact", y = "Effectif") +
  theme_minimal(base_size = 14)

# Boxplot
ggplot(df, aes(x = SEX, y = AgeExact, fill = SEX)) +
  geom_boxplot() +
  labs(title = "Boxplot des ages exacts au deces", x = "Sexe", y = "Age exact") +

```

```

theme_minimal()

# Tableau par cohorte
cohort_exact <- df %>%
  filter(BIRTH_YEAR >= 1905, BIRTH_YEAR <= 1915) %>%
  group_by(BIRTH_YEAR, SEX) %>%
  summarise(
    L105 = n(),
    m105 = max(AgeExact),
    `105+e105` = mean(AgeExact),
    .groups = "drop"
  ) %>%
  pivot_wider(
    names_from = SEX,
    values_from = c(L105, m105, `105+e105`),
    names_glue = "{.value}_{SEX}"
  ) %>%
  arrange(BIRTH_YEAR)

```

C.2 Tests du χ^2

Groupe 3 (Groupe 2 restreint)

La logique du test de chi2 est identique pour les trois groupes étudiés. Seule la partie de chargement et de filtrage des données varie d'un groupe à l'autre. On présente ci-dessous le code complet pour le **groupe 3**, à savoir les femmes nées entre 1905 et 1913, avec excès calculés au-delà de 105 ans. Les mêmes étapes sont appliquées aux groupes 1 et 2.

```

# *****
#   Analyse Khi2 FEMMES -- sans 1914 et 1915 G2 restreint
# *****

library(readxl)
library(dplyr)
library(lubridate)
library(ggplot2)
library(POT)

# Parametres GPD
xi <- -0.092
beta <- 1.64088
seuil <- 105
omega <- 122.73
x_max <- omega - seuil # = 17.73

# Chargement et filtrage des donnees
data <- read_excel(
  "/Users/leilasanogo/Documents/memoires moi/idl_bel_death_105_analyse.xlsx",
  sheet = "IDL_BEL_Death_105"
)

data_femmes <- data %>%
  filter(SEX == "F", BIRTH_YEAR > 1904, BIRTH_YEAR < 1914) %>%
  mutate(
    BIRTH_Date = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),
    DEATH_Date = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),
    AgeExact = time_length(interval(BIRTH_Date, DEATH_Date), unit = "years"),
    excess = AgeExact - seuil
  ) %>%
  filter(excess >= 0)

# =====

```

```

#   Histogramme general
# =====
ggplot(data_femmes, aes(x = excess)) +
  geom_histogram(binwidth = 0.2, fill = "skyblue", color = "black", boundary = 0) +
  labs(title = "Histogramme des excès au-delà de 105 ans (femmes)", x = "Excès (age
    - 105)", y = "Effectif") +
  theme_minimal(base_size = 14)

# =====
#   Fonction generique pour Khi2
# =====
test_khi2 <- function(bins, labels, titre_plot) {
  data_femmes_cut <- data_femmes %>%
    mutate(classe = cut(excess, breaks = bins, labels = labels, right = FALSE))

  effectifs_observes <- table(data_femmes_cut$classe)
  observed <- as.numeric(effectifs_observes)
  total_observed <- sum(observed)

  probs <- numeric(length(labels))
  for (i in 1:length(probs)) {
    p_inf <- POT::pgpd(bins[i], scale = beta, shape = xi)
    p_sup <- POT::pgpd(bins[i + 1], scale = beta, shape = xi)
    probs[i] <- p_sup - p_inf
  }

  expected <- round(probs * total_observed, 2)

  cat("\n=== Test du chi2 :", titre_plot, " ===\n")
  cat("\nEffectifs observes par classe :\n")
  print(effectifs_observes)

  cat("\nEffectifs attendus (arrondis) :\n")
  names(expected) <- labels
  print(expected)

  cat("\nSomme observes :", total_observed, "\n")
  cat("Somme attendus :", sum(expected), "\n")

  chi2_stat <- sum((observed - expected)^2 / expected)
  ddl <- length(observed) - 1
  p_value <- pchisq(chi2_stat, df = ddl, lower.tail = FALSE)

  resultats <- data.frame(
    Classe = labels,
    Observees = observed,
    Attendus = expected,
    Proba_theorique = round(probs, 4)
  )
  cat("\nTableau recapitulatif :\n")
  print(resultats)

  cat("\nStatistique du chi2 :", round(chi2_stat, 4), "\n")
  cat("Degrés de liberte :", ddl, "\n")
  cat("p-value :", round(p_value, 6), "\n")

  barplot_heights <- rbind(observed, expected)
  colnames(barplot_heights) <- labels
  barplot(
    barplot_heights,
    beside = TRUE,
    col = c("skyblue", "orange"),
    names.arg = labels,
    ylim = c(0, max(barplot_heights) + 10),
    ylab = "Effectifs",

```



```

    main = paste("Test du chi2 --", titre_plot),
    legend.text = c("Observes", "Attendus"),
    args.legend = list(x = "topright", bty = "n")
  )
}

# =====
# 4 classes
# =====
bins_4 <- c(0, 1, 2, 3, 17.73)
labels_4 <- c("[0-1[", "[1-2[", "[2-3[", "[3-17.73[")
test_khi2(bins_4, labels_4, "Femmes (4 classes, seuil 105)")

# =====
# 5 classes
# =====
bins_5 <- c(0, 1, 2, 3, 4, 17.73)
labels_5 <- c("[0-1[", "[1-2[", "[2-3[", "[3-4[", "[4-17.73[")
test_khi2(bins_5, labels_5, "Femmes (5 classes, seuil 105)")

# =====
# 8 classes
# =====
bins_8 <- c(0, 0.5, 1, 1.5, 2, 2.5, 3, 4, 17.73)
labels_8 <- c("[0-0.5[", "[0.5-1[", "[1-1.5[", "[1.5-2[", "[2-2.5[", "[2.5-3[", "[3-4[", "[4-17.73[")
test_khi2(bins_8, labels_8, "Femmes (8 classes, seuil 105)")

```

C.3 Tests de Kolmogorov–Smirnov

Groupe 1 (hommes et femmes)

```

# KS HOMMES G1
library(readxl)
library(dplyr)
library(lubridate)
library(POT)
library(ggplot2)

xi <- -0.132
beta <- 1.29148
seuil <- 105

data <- read_excel("idl_bel_death_105_analyse.xlsx", sheet = "new")

data_hommes <- data %>%
  filter(SEX == "M") %>%
  mutate(
    BIRTH_Date = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),
    DEATH_Date = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),
    AgeExact = time_length(interval(BIRTH_Date, DEATH_Date), "years")
  )

cat("Doublons sur AgeExact :", sum(duplicated(data_hommes$AgeExact)), "\n")

data_hommes <- data_hommes %>%
  mutate(excess = AgeExact - seuil)

exces_hommes <- data_hommes %>% filter(excess > 0) %>% pull(excess)

ks_test_h <- ks.test(exces_hommes, function(x) pgpd(x, scale = beta, shape = xi),
  exact = TRUE)

```

```

cat("=== Test KS exact pour les hommes G1 ===\n")
print(ks_test_h)

th_quantiles_h <- qgpd(ppoints(length(exces_hommes)), scale = beta, shape = xi)
df_qq_h <- data.frame(theorique = sort(th_quantiles_h), empirique = sort(exces_
  hommes))

ggplot(df_qq_h, aes(x = theorique, y = empirique)) +
  geom_point(size = 1.2) +
  geom_abline(slope = 1, intercept = 0, color = "black") +
  labs(title = "QQ-plot : exces IDL hommes vs GPD",
    x = "Quantiles theoriques GPD", y = "Quantiles empiriques IDL") +
  theme_minimal(base_size = 14)

```

```

# KS FEMMES G1
xi <- -0.092
beta <- 1.64088

data_femmes <- data %>%
  filter(SEX == "F") %>%
  mutate(
    BIRTH_Date = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),
    DEATH_Date = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),
    AgeExact = time_length(interval(BIRTH_Date, DEATH_Date), "years")
  )

cat("Doublons sur AgeExact :", sum(duplicated(data_femmes$AgeExact)), "\n")

data_femmes <- data_femmes %>%
  arrange(AgeExact, IDL_ID) %>%
  group_by(AgeExact) %>%
  mutate(n = n(),
    AgeExact_unique = if_else(n > 1, AgeExact + row_number() * 1e-10, AgeExact)
  ) %>%
  ungroup() %>%
  select(-n)

cat("Doublons sur AgeExact_unique :", sum(duplicated(data_femmes$AgeExact_unique)),
  "\n")

data_femmes <- data_femmes %>%
  mutate(excess = AgeExact_unique - seuil)

exces <- data_femmes %>% filter(excess > 0) %>% pull(excess)

ks_test_f <- ks.test(exces, function(x) pgpd(x, scale = beta, shape = xi), exact =
  FALSE)

cat("=== Test KS asymptotique pour les femmes G1 ===\n")
print(ks_test_f)

th_quantiles <- qgpd(ppoints(length(exces)), scale = beta, shape = xi)
df_qq <- data.frame(theorique = sort(th_quantiles), empirique = sort(exces))

ggplot(df_qq, aes(x = theorique, y = empirique)) +
  geom_point(size = 1.2) +
  geom_abline(slope = 1, intercept = 0, color = "black") +
  labs(title = "QQ-plot : exces IDL femmes vs GPD",
    x = "Quantiles theoriques GPD", y = "Quantiles empiriques IDL") +
  theme_minimal(base_size = 14)

```

Groupe 2 (hommes et femmes)

```
# KS HOMMES G2
```

```

data <- read_excel("idl_bel_death_105_analyse.xlsx", sheet = "IDL_BEL_Death_105")

data_hommes <- data %>%
  filter(SEX == "M", BIRTH_YEAR > 1904) %>%
  mutate(
    BIRTH_Date = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),
    DEATH_Date = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),
    AgeExact = time_length(interval(BIRTH_Date, DEATH_Date), "years")
  )

cat("Doublons exacts sur AgeExact :", sum(duplicated(data_hommes$AgeExact)), "\n")

data_hommes <- data_hommes %>% mutate(excess = AgeExact - seuil)

exces_hommes <- data_hommes %>% filter(excess > 0) %>% pull(excess)

ks_test_h <- ks.test(exces_hommes, function(x) pgpd(x, scale = beta, shape = xi),
  exact = TRUE)

cat("=== Test KS exact pour les hommes G2 ===\n")
print(ks_test_h)

th_quantiles_h <- qgpd(ppoints(length(exces_hommes)), scale = beta, shape = xi)
df_qq_h <- data.frame(theorique = sort(th_quantiles_h), empirique = sort(exces_
  hommes))

ggplot(df_qq_h, aes(x = theorique, y = empirique)) +
  geom_point(size = 1.2) +
  geom_abline(slope = 1, intercept = 0, color = "black") +
  labs(title = "QQ-plot : exces IDL hommes vs GPD",
    x = "Quantiles theoriques GPD", y = "Quantiles empiriques IDL") +
  theme_minimal(base_size = 14)

```

```

# KS FEMMES G2
xi <- -0.092
beta <- 1.64088

data_femmes <- data %>%
  filter(SEX == "F", BIRTH_YEAR > 1904) %>%
  mutate(
    BIRTH_Date = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),
    DEATH_Date = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),
    AgeExact = time_length(interval(BIRTH_Date, DEATH_Date), "years")
  )

cat("Doublons sur AgeExact :", sum(duplicated(data_femmes$AgeExact)), "\n")

data_femmes <- data_femmes %>%
  arrange(AgeExact, IDL_ID) %>%
  group_by(AgeExact) %>%
  mutate(n = n(),
    AgeExact_unique = if_else(n > 1, AgeExact + row_number() * 1e-10, AgeExact)
  ) %>%
  ungroup() %>%
  select(-n)

cat("Doublons sur AgeExact_unique :", sum(duplicated(data_femmes$AgeExact_unique)),
  "\n")

data_femmes <- data_femmes %>%
  mutate(excess = AgeExact_unique - seuil)

exces <- data_femmes %>% filter(excess > 0) %>% pull(excess)

ks_test_f <- ks.test(exces, function(x) pgpd(x, scale = beta, shape = xi), exact =

```

```

FALSE)

cat("=== Test KS asymptotique pour les femmes G2 ===\n")
print(ks_test_f)

th_quantiles <- qgpd(ppoints(length(exces)), scale = beta, shape = xi)
df_qq <- data.frame(theorique = sort(th_quantiles), empirique = sort(exces))

ggplot(df_qq, aes(x = theorique, y = empirique)) +
  geom_point(size = 1.2) +
  geom_abline(slope = 1, intercept = 0, color = "black") +
  labs(title = "QQ-plot : exces IDL femmes vs GPD",
       x = "Quantiles theoriques GPD", y = "Quantiles empiriques IDL") +
  theme_minimal(base_size = 14)

```

Groupe 3 (Groupe 2 restreint)

```

# KS HOMMES G3
data_hommes <- data %>%
  filter(SEX == "M", BIRTH_YEAR > 1904, BIRTH_YEAR < 1914) %>%
  mutate(
    BIRTH_Date = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),
    DEATH_Date = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),
    AgeExact = time_length(interval(BIRTH_Date, DEATH_Date), "years")
  )

cat("Doublons sur AgeExact :", sum(duplicated(data_hommes$AgeExact)), "\n")

data_hommes <- data_hommes %>% mutate(excess = AgeExact - seuil)

exces_hommes <- data_hommes %>% filter(excess > 0) %>% pull(excess)

ks_test_h <- ks.test(exces_hommes, function(x) pgpd(x, scale = beta, shape = xi),
  exact = TRUE)

cat("=== Test KS exact pour les hommes G3 ===\n")
print(ks_test_h)

th_quantiles_h <- qgpd(ppoints(length(exces_hommes)), scale = beta, shape = xi)
df_qq_h <- data.frame(theorique = sort(th_quantiles_h), empirique = sort(exces_
  hommes))

ggplot(df_qq_h, aes(x = theorique, y = empirique)) +
  geom_point(size = 1.2) +
  geom_abline(slope = 1, intercept = 0, color = "black") +
  labs(title = "QQ-plot : exces IDL hommes vs GPD",
       x = "Quantiles theoriques GPD", y = "Quantiles empiriques IDL") +
  theme_minimal(base_size = 14)

```

```

# KS FEMMES G3
xi <- -0.092
beta <- 1.64088

data_femmes <- data %>%
  filter(SEX == "F", BIRTH_YEAR > 1904, BIRTH_YEAR < 1914) %>%
  mutate(
    BIRTH_Date = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),
    DEATH_Date = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),
    AgeExact = time_length(interval(BIRTH_Date, DEATH_Date), "years")
  )

cat("Doublons sur AgeExact :", sum(duplicated(data_femmes$AgeExact)), "\n")

data_femmes <- data_femmes %>%

```

```

  arrange(AgeExact, IDL_ID) %>%
  group_by(AgeExact) %>%
  mutate(n = n(),
         AgeExact_unique = if_else(n > 1, AgeExact + row_number() * 1e-10, AgeExact)
        ) %>%
  ungroup() %>%
  select(-n)

cat("Doublons sur AgeExact_unique :", sum(duplicated(data_femmes$AgeExact_unique)),
    "\n")

data_femmes <- data_femmes %>% mutate(excess = AgeExact_unique - seuil)

exces <- data_femmes %>% filter(excess > 0) %>% pull(excess)

ks_test_f <- ks.test(exces, function(x) pgpd(x, scale = beta, shape = xi), exact =
  FALSE)

cat("=== Test KS asymptotique pour les femmes G3 ===\n")
print(ks_test_f)

th_quantiles <- qgpd(ppoints(length(exces)), scale = beta, shape = xi)
df_qq <- data.frame(theorique = sort(th_quantiles), empirique = sort(exces))

ggplot(df_qq, aes(x = theorique, y = empirique)) +
  geom_point(size = 1.2) +
  geom_abline(slope = 1, intercept = 0, color = "black") +
  labs(title = "QQ-plot : exces IDL femmes vs GPD",
       x = "Quantiles theoriques GPD", y = "Quantiles empiriques IDL") +
  theme_minimal(base_size = 14)

```

C.4 Réestimation des paramètres GPD au seuil de 105 ans

Groupe 1 (hommes et femmes)

```

# Ajustement GPD Femmes G1
library(readxl)
library(dplyr)
library(lubridate)
library(POT)
library(ggplot2)

seuil <- 105
df <- read_excel("idl_bel_death_105_analyse.xlsx", sheet = "new")

data_femmes <- df %>%
  filter(SEX == "F") %>%
  mutate(BIRTH_Date = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),
         DEATH_Date = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),
         AgeExact = time_length(interval(BIRTH_Date, DEATH_Date), unit = "years"),
         excess = AgeExact - seuil) %>%
  filter(excess > 0)

fit_femmes <- fitgpd(data_femmes$excess, threshold = 0, est = "mle")

cat("=== Parametres estimees (MLE) ===\n")
print(fit_femmes$fitted.values)

std_err <- fit_femmes$std.err
est <- fit_femmes$fitted.values

```

```

IC_shape <- c(est["shape"] - 1.96 * std_err["shape"],
             est["shape"] + 1.96 * std_err["shape"])
IC_scale <- c(est["scale"] - 1.96 * std_err["scale"],
             est["scale"] + 1.96 * std_err["scale"])

cat("IC 95% pour shape :", round(IC_shape, 4), "\n")
cat("IC 95% pour scale :", round(IC_scale, 4), "\n")

n <- length(data_femmes$excess)
qq_theorique <- qgpd(ppoints(n), scale = est["scale"], shape = est["shape"])
qq_empirique <- sort(data_femmes$excess)
df_qq <- data.frame(theorique = qq_theorique, empirique = qq_empirique)

ggplot(df_qq, aes(x = theorique, y = empirique)) +
  geom_point(color = "darkred", size = 1.2) +
  geom_abline(slope = 1, intercept = 0, color = "black") +
  labs(title = "QQ-plot : excès femmes vs GPD (MLE)",
       x = "Quantiles théoriques GPD", y = "Quantiles empiriques IDL") +
  theme_minimal(base_size = 14)

# Test KS
data_femmes <- data_femmes %>%
  arrange(excess) %>%
  group_by(excess) %>%
  mutate(n = n(), excess_unique = if_else(n > 1, excess + row_number() * 1e-10,
                                           excess)) %>%
  ungroup() %>% select(-n)

ks_femmes_groupe1 <- ks.test(data_femmes$excess_unique,
                             function(x) pgpd(x, scale = est["scale"], shape = est["shape"]),
                             exact = FALSE)

cat("\n=== Test KS - Femmes (feuille 'new') ===\n")
print(ks_femmes_groupe1)

# Age ultime
xi <- est["shape"]
beta <- est["scale"]
if (xi < 0) {
  omega <- seuil - (beta / xi)
  cat("\n>>> Age ultime estime :", round(omega, 6), "ans\n")
} else {
  cat("\n>>> xi >= 0 -> pas de borne supérieure (omega infini)\n")
}

```

```

# Ajustement GPD Hommes G1
seuil <- 105
df <- read_excel("idl_bel_death_105_analyse.xlsx", sheet = "new")

data_hommes <- df %>%
  filter(SEX == "M") %>%
  mutate(BIRTH_Date = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),
         DEATH_Date = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),
         AgeExact = time_length(interval(BIRTH_Date, DEATH_Date), unit = "years"),
         excess = AgeExact - seuil) %>%
  filter(excess > 0)

fit_hommes <- fitgpd(data_hommes$excess, threshold = 0, est = "mle")

cat("=== Paramètres estimés (MLE) - HOMMES ===\n")
print(fit_hommes$fitted.values)

std_err <- fit_hommes$std.err
est <- fit_hommes$fitted.values

```

```

IC_shape <- c(est["shape"] - 1.96 * std_err["shape"],
             est["shape"] + 1.96 * std_err["shape"])
IC_scale <- c(est["scale"] - 1.96 * std_err["scale"],
             est["scale"] + 1.96 * std_err["scale"])

cat("IC 95% pour shape :", round(IC_shape, 4), "\n")
cat("IC 95% pour scale :", round(IC_scale, 4), "\n")

n <- length(data_hommes$excess)
qq_theorique <- qgpd(ppoints(n), scale = est["scale"], shape = est["shape"])
qq_empirique <- sort(data_hommes$excess)
df_qq <- data.frame(theorique = qq_theorique, empirique = qq_empirique)

ggplot(df_qq, aes(x = theorique, y = empirique)) +
  geom_point(color = "darkblue", size = 1.2) +
  geom_abline(slope = 1, intercept = 0, color = "black") +
  labs(title = "QQ-plot : excès hommes vs GPD (MLE)",
       x = "Quantiles théoriques GPD", y = "Quantiles empiriques IDL") +
  theme_minimal(base_size = 14)

# Age ultime
xi <- est["shape"]
beta <- est["scale"]
if (xi < 0) {
  omega <- seuil - (beta / xi)
  cat("\n>>> Age ultime estime :", round(omega, 6), "ans\n")
} else {
  cat("\n>>> xi >= 0 -> pas de borne supérieure (omega infini)\n")
}

```

Groupe 2 (hommes et femmes)

```

# Ajustement GPD Femmes G2
library(readxl)
library(dplyr)
library(lubridate)
library(POT)
library(ggplot2)

seuil <- 105

df <- read_excel("idl_bel_death_105_analyse.xlsx", sheet = "IDL_BEL_Death_105")
df <- df %>% filter(BIRTH_YEAR < 1886 | BIRTH_YEAR > 1904)

data_femmes <- df %>%
  filter(SEX == "F") %>%
  mutate(BIRTH_Date = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),
         DEATH_Date = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),
         AgeExact = time_length(interval(BIRTH_Date, DEATH_Date), unit = "years"),
         excess = AgeExact - seuil) %>%
  filter(excess > 0)

fit_femmes <- fitgpd(data_femmes$excess, threshold = 0, est = "mle")

cat("=== Paramètres estimés (MLE) - FEMMES - Feuille IDL_BEL_Death_105 ===\n")
print(fit_femmes$fitted.values)

std_err <- fit_femmes$std.err
est <- fit_femmes$fitted.values

IC_shape <- c(est["shape"] - 1.96 * std_err["shape"],
             est["shape"] + 1.96 * std_err["shape"])
IC_scale <- c(est["scale"] - 1.96 * std_err["scale"],
             est["scale"] + 1.96 * std_err["scale"]),

```

```

        est["scale"] + 1.96 * std_err["scale"]])

cat("IC 95% pour shape :", round(IC_shape, 4), "\n")
cat("IC 95% pour scale :", round(IC_scale, 4), "\n")

n <- length(data_femmes$excess)
qq_theorique <- qgpd(ppoints(n), scale = est["scale"], shape = est["shape"])
qq_empirique <- sort(data_femmes$excess)
df_qq <- data.frame(theorique = qq_theorique, empirique = qq_empirique)

ggplot(df_qq, aes(x = theorique, y = empirique)) +
  geom_point(color = "darkred", size = 1.2) +
  geom_abline(slope = 1, intercept = 0, color = "black") +
  labs(title = "QQ-plot : excès femmes vs GPD - G2",
        x = "Quantiles théoriques GPD", y = "Quantiles empiriques IDL") +
  theme_minimal(base_size = 14)

# Test KS
data_femmes <- data_femmes %>%
  arrange(excess) %>%
  group_by(excess) %>%
  mutate(n = n(),
         excess_unique = if_else(n > 1, excess + row_number() * 1e-10, excess)) %>%
  ungroup() %>% select(-n)

ks_femmes_groupe2 <- ks.test(data_femmes$excess_unique,
                             function(x) pgpd(x, scale = est["scale"], shape = est["shape"]),
                             exact = FALSE)

cat("\n=== Test KS - Femmes (feuille 'IDL_BEL_Death_105') ===\n")
print(ks_femmes_groupe2)

# Age ultime
xi <- est["shape"]
beta <- est["scale"]

if (xi < 0) {
  omega <- seuil - (beta / xi)
  cat("\n>>> Age ultime estime :", round(omega, 6), "ans\n")
} else {
  cat("\n>>> xi >= 0 -> pas de borne supérieure (omega infini)\n")
}

```

```

# Ajustement GPD Hommes G2
df <- read_excel("idl_bel_death_105_analyse.xlsx", sheet = "IDL_BEL_Death_105")
df <- df %>% filter(BIRTH_YEAR < 1886 | BIRTH_YEAR > 1904)

data_hommes <- df %>%
  filter(SEX == "M") %>%
  mutate(BIRTH_Date = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),
         DEATH_Date = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),
         AgeExact = time_length(interval(BIRTH_Date, DEATH_Date), unit = "years"),
         excess = AgeExact - seuil) %>%
  filter(excess > 0)

fit_hommes <- fitgpd(data_hommes$excess, threshold = 0, est = "mle")

cat("=== Paramètres estimés (MLE) - HOMMES - Feuille IDL_BEL_Death_105 ===\n")
print(fit_hommes$fitted.values)

std_err <- fit_hommes$std.err
est <- fit_hommes$fitted.values

IC_shape <- c(est["shape"] - 1.96 * std_err["shape"],

```



```

        est["shape"] + 1.96 * std_err["shape"])
IC_scale <- c(est["scale"] - 1.96 * std_err["scale"],
             est["scale"] + 1.96 * std_err["scale"])

cat("IC 95% pour shape :", round(IC_shape, 4), "\n")
cat("IC 95% pour scale :", round(IC_scale, 4), "\n")

n <- length(data_hommes$excess)
qq_theorique <- qqpd(ppoints(n), scale = est["scale"], shape = est["shape"])
qq_empirique <- sort(data_hommes$excess)
df_qq <- data.frame(theorique = qq_theorique, empirique = qq_empirique)

ggplot(df_qq, aes(x = theorique, y = empirique)) +
  geom_point(color = "darkblue", size = 1.2) +
  geom_abline(slope = 1, intercept = 0, color = "black") +
  labs(title = "QQ-plot : excès hommes vs GPD - G2",
       x = "Quantiles theoriques GPD", y = "Quantiles empiriques IDL") +
  theme_minimal(base_size = 14)

# Age ultime
xi <- est["shape"]
beta <- est["scale"]

if (xi < 0) {
  omega <- seuil - (beta / xi)
  cat("\n>>> Age ultime estime :", round(omega, 2), "ans\n")
} else {
  cat("\n>>> xi >= 0 -> pas de borne superieure (omega infini)\n")
}

```

C.5 Modèle unisexe : validation empirique

Groupe 1 (femmes et hommes)

```

# Verification des doublons dans les excès - Groupe 1

# Pour les femmes
doublons_femmes <- data_femmes %>%
  group_by(excess) %>%
  filter(n() > 1) %>%
  arrange(excess)

cat("\n--- Doublons dans excess_femmes (groupe 1) ---\n")
print(doublons_femmes)

# Pour les hommes
doublons_hommes <- data_hommes %>%
  group_by(excess) %>%
  filter(n() > 1) %>%
  arrange(excess)

cat("\n--- Doublons dans excess_hommes (groupe 1) ---\n")
print(doublons_hommes)

# Recherche des valeurs communes
valeurs_communes <- intersect(data_femmes$excess, data_hommes$excess)

cat("\n>>> Nombre de valeurs communes entre femmes et hommes :", length(valeurs_
  communes), "\n")
print(valeurs_communes)

# Correction des ex-aequo
excess_femmes_corrige <- data_femmes %>%

```

```

  arrange(excess) %>%
  group_by(excess) %>%
  mutate(n = n(), excess_corrige = if_else(n > 1, excess + row_number() * 1e-10,
    excess)) %>%
  ungroup() %>%
  pull(excess_corrige)

nb_doublons_f <- sum(duplicated(excess_femmes_corrige))
cat("\n>>> Nombre de doublons restants chez les femmes :", nb_doublons_f, "\n")

excess_hommes_corrige <- data_hommes$excess
indices_h <- which(excess_hommes_corrige %in% valeurs_communes)
excess_hommes_corrige[indices_h] <- excess_hommes_corrige[indices_h] + seq_along(
  indices_h) * 1e-10

# Test KS bilateral
group1_ks <- ks.test(excess_femmes_corrige, excess_hommes_corrige)

cat("\n=== Test KS bilateral Femmes vs Hommes - Groupe 1 ===\n")
print(group1_ks)

if (group1_ks$p.value > 0.05) {
  cat("\n-> Conclusion : Pas de difference significative (p-value =", round(group1_ks$p.value, 4), ")\n")
} else {
  cat("\n-> Conclusion : Difference significative (p-value =", round(group1_ks$p.value, 4), ")\n")
}

# Test de vraisemblance
fit_femmes_1 <- fitgpd(data_femmes$excess, threshold = 0, est = "mle")
loglik_femmes <- sum(dgpd(data_femmes$excess, scale = fit_femmes_1$fitted.values["scale"], shape = fit_femmes_1$fitted.values["shape"], log = TRUE))

fit_hommes_1 <- fitgpd(data_hommes$excess, threshold = 0, est = "mle")
loglik_hommes <- sum(dgpd(data_hommes$excess, scale = fit_hommes_1$fitted.values["scale"], shape = fit_hommes_1$fitted.values["shape"], log = TRUE))

excess_mixes <- c(data_femmes$excess, data_hommes$excess)
fit_mixes <- fitgpd(excess_mixes, threshold = 0, est = "mle")
loglik_mixes <- sum(dgpd(excess_mixes, scale = fit_mixes$fitted.values["scale"], shape = fit_mixes$fitted.values["shape"], log = TRUE))

stat_LRT <- -2 * (loglik_mixes - (loglik_femmes + loglik_hommes))
ddl <- 2
p_val_LRT <- pchisq(stat_LRT, df = ddl, lower.tail = FALSE)

cat("\n=== Test de vraisemblance Femmes vs Hommes - Groupe 1 ===\n")
cat("Log-vraisemblance femmes :", round(loglik_femmes, 4), "\n")
cat("Log-vraisemblance hommes :", round(loglik_hommes, 4), "\n")
cat("Log-vraisemblance globale :", round(loglik_mixes, 4), "\n")
cat("Statistique LRT :", round(stat_LRT, 4), "\n")
cat("Degrés de liberte :", ddl, "\n")
cat("p-value :", round(p_val_LRT, 6), "\n")

if (p_val_LRT > 0.05) {
  cat("\n-> Conclusion : Pas de difference significative entre les parametres.\n")
} else {
  cat("\n-> Conclusion : Difference significative entre les parametres estimes.\n")
}

```

```

# Groupe 2 - Verification et tests

```

```

# Doublons femmes
doublons_femmes <- data_femmes %>% group_by(excess) %>% filter(n() > 1) %>% arrange(

```

```

    excess)
cat("\n--- Doublons dans excess_femmes ---\n")
print(doublons_femmes)

# Doublons hommes
doublons_hommes <- data_hommes %>% group_by(excess) %>% filter(n() > 1) %>% arrange(
  excess)
cat("\n--- Doublons dans excess_hommes ---\n")
print(doublons_hommes)

valeurs_communes <- intersect(data_femmes$excess, data_hommes$excess)
cat("\n>>> Nombre de valeurs communes :", length(valeurs_communes), "\n")

# Correction
excess_femmes_corrige <- data_femmes %>%
  arrange(excess) %>%
  group_by(excess) %>%
  mutate(n = n(), excess_corrige = if_else(n > 1, excess + row_number() * 1e-10,
    excess)) %>%
  ungroup() %>% pull(excess_corrige)

nb_doublons <- sum(duplicated(excess_femmes_corrige))
cat("\n>>> Doublons restants chez les femmes :", nb_doublons, "\n")

excess_hommes_corrige <- data_hommes$excess
indices_corr <- which(excess_hommes_corrige %in% valeurs_communes)
excess_hommes_corrige[indices_corr] <- excess_hommes_corrige[indices_corr] + seq_
  along(indices_corr) * 1e-10

ks_test_sexes <- ks.test(excess_femmes_corrige, excess_hommes_corrige)

cat("\n=== Test KS bilatéral Femmes vs Hommes - Groupe 2 ===\n")
print(ks_test_sexes)

if (ks_test_sexes$p.value > 0.05) {
  cat("\n-> Conclusion : Pas de différence significative (p-value =", round(ks_test_
    sexes$p.value, 4), ").\n")
} else {
  cat("\n-> Conclusion : Différence significative (p-value =", round(ks_test_sexes$p
    .value, 4), ").\n")
}

# Test de vraisemblance
fit_femmes <- fitgpd(data_femmes$excess, threshold = 0, est = "mle")
loglik_femmes <- sum(dgpd(data_femmes$excess, scale = fit_femmes$fitted.values["
  scale"], shape = fit_femmes$fitted.values["shape"], log = TRUE))

fit_hommes <- fitgpd(data_hommes$excess, threshold = 0, est = "mle")
loglik_hommes <- sum(dgpd(data_hommes$excess, scale = fit_hommes$fitted.values["
  scale"], shape = fit_hommes$fitted.values["shape"], log = TRUE))

excess_mixes <- c(data_femmes$excess, data_hommes$excess)
fit_mixes <- fitgpd(excess_mixes, threshold = 0, est = "mle")
loglik_mixes <- sum(dgpd(excess_mixes, scale = fit_mixes$fitted.values["scale"],
  shape = fit_mixes$fitted.values["shape"], log = TRUE))

stat_LRT <- -2 * (loglik_mixes - (loglik_femmes + loglik_hommes))
ddl <- 2
p_val_LRT <- pchisq(stat_LRT, df = ddl, lower.tail = FALSE)

cat("\n=== Test de vraisemblance Femmes vs Hommes - Groupe 2 ===\n")
cat("Log-vraisemblance femmes :", round(loglik_femmes, 4), "\n")
cat("Log-vraisemblance hommes :", round(loglik_hommes, 4), "\n")
cat("Log-vraisemblance globale :", round(loglik_mixes, 4), "\n")
cat("Statistique LRT :", round(stat_LRT, 4), "\n")

```

```

cat("Degrés de liberté :", ddl, "\n")
cat("p-value :", round(p_val_LRT, 6), "\n")

if (p_val_LRT > 0.05) {
  cat("\n-> Conclusion : Pas de différence significative entre les paramètres
      estimés.\n")
} else {
  cat("\n-> Conclusion : Paramètres significativement différents entre les deux
      sexes.\n")
}

```

Groupe 2 (femmes et hommes)

Le même protocole est appliqué au groupe 2 afin de tester la validité d'un modèle unisexe. Les étapes de correction des ex-aequo, de comparaison des distributions d'excès (via le test de Kolmogorov–Smirnov) et de comparaison des paramètres estimés (par un test de vraisemblance) sont strictement identiques à celles décrites pour le groupe 1. Seul le jeu de données change, avec un filtrage spécifique sur les cohortes et l'utilisation de la feuille *IDL_BEL_Death_105*.

C.6 Modèle GPD ajusté sur les deux sexes

Groupe 1

```

# Ajustement GPD Groupe mixte (femmes + hommes) Groupe 1
library(readxl)
library(dplyr)
library(lubridate)
library(POT)
library(ggplot2)

seuil <- 105

df <- read_excel("idl_bel_death_105_analyse.xlsx", sheet = "new")

data_mixte <- df %>%
  mutate(
    BIRTH_Date = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),
    DEATH_Date = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),
    AgeExact = time_length(interval(BIRTH_Date, DEATH_Date), unit = "years"),
    excess = AgeExact - seuil
  ) %>%
  filter(excess > 0)

fit_mixte <- fitgpd(data_mixte$excess, threshold = 0, est = "mle")
cat("=== Paramètres estimés (MLE) Groupe mixte ===\n")
print(fit_mixte$fitted.values)

std_err <- fit_mixte$std.err
est <- fit_mixte$fitted.values

IC_shape <- c(est["shape"] - 1.96 * std_err["shape"], est["shape"] + 1.96 * std_err[
  "shape"])
IC_scale <- c(est["scale"] - 1.96 * std_err["scale"], est["scale"] + 1.96 * std_err[
  "scale"])

cat("IC 95% pour shape :", round(IC_shape, 4), "\n")
cat("IC 95% pour scale :", round(IC_scale, 4), "\n")

n <- length(data_mixte$excess)
qq_theorique <- qgpd(ppoints(n), scale = est["scale"], shape = est["shape"])

```

```

qq_empirique <- sort(data_mixte$excess)
df_qq <- data.frame(theorique = qq_theorique, empirique = qq_empirique)

ggplot(df_qq, aes(x = theorique, y = empirique)) +
  geom_point(color = "darkred", size = 1.2) +
  geom_abline(slope = 1, intercept = 0, color = "black") +
  labs(title = "QQ-plot : excès groupe mixte vs GPD",
       x = "Quantiles théoriques GPD", y = "Quantiles empiriques IDL") +
  theme_minimal(base_size = 14) + ylim(0, NA)

data_mixte <- data_mixte %>%
  arrange(excess) %>%
  group_by(excess) %>%
  mutate(n = n(),
         excess_unique = if_else(n > 1, excess + row_number() * 1e-10, excess)) %>%
  ungroup() %>%
  select(-n)

ks_mixte <- ks.test(data_mixte$excess_unique,
                    function(x) pgpd(x, scale = est["scale"], shape = est["shape"]),
                    exact = FALSE)

cat("\n=== Test KS Groupe mixte ===\n")
print(ks_mixte)

xi <- est["shape"]
beta <- est["scale"]
if (xi < 0) {
  omega <- seuil - (beta / xi)
  cat("\n>>> Age ultime estime :", round(omega, 6), "ans\n")
} else {
  cat("\n>>> xi >= 0 -> pas de borne supérieure (omega infini)\n")
}

```

Groupe 2

```

# Ajustement GPD Groupe mixte (femmes + hommes) Groupe 2

df <- read_excel("idl_bel_death_105_analyse.xlsx", sheet = "IDL_BEL_Death_105")
df <- df %>% filter(BIRTH_YEAR < 1886 | BIRTH_YEAR > 1904)

data_mixte <- df %>%
  mutate(
    BIRTH_Date = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),
    DEATH_Date = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),
    AgeExact = time_length(interval(BIRTH_Date, DEATH_Date), unit = "years"),
    excess = AgeExact - seuil
  ) %>%
  filter(excess > 0)

fit_mixte <- fitgpd(data_mixte$excess, threshold = 0, est = "mle")
cat("=== Paramètres estimés (MLE) Groupe mixte ===\n")
print(fit_mixte$fitted.values)

std_err <- fit_mixte$std.err
est <- fit_mixte$fitted.values

IC_shape <- c(est["shape"] - 1.96 * std_err["shape"], est["shape"] + 1.96 * std_err["shape"])
IC_scale <- c(est["scale"] - 1.96 * std_err["scale"], est["scale"] + 1.96 * std_err["scale"])

cat("IC 95% pour shape :", round(IC_shape, 4), "\n")
cat("IC 95% pour scale :", round(IC_scale, 4), "\n")

```

```

n <- length(data_mixte$excess)
qq_theorique <- qgpd(ppoints(n), scale = est["scale"], shape = est["shape"])
qq_empirique <- sort(data_mixte$excess)
df_qq <- data.frame(theorique = qq_theorique, empirique = qq_empirique)

ggplot(df_qq, aes(x = theorique, y = empirique)) +
  geom_point(color = "darkred", size = 1.2) +
  geom_abline(slope = 1, intercept = 0, color = "black") +
  labs(title = "QQ-plot : exces groupe mixte vs GPD",
       x = "Quantiles theoriques GPD", y = "Quantiles empiriques IDL") +
  theme_minimal(base_size = 14) + ylim(0, NA)

data_mixte <- data_mixte %>%
  arrange(excess) %>%
  group_by(excess) %>%
  mutate(n = n(),
         excess_unique = if_else(n > 1, excess + row_number() * 1e-10, excess)) %>%
  ungroup() %>%
  select(-n)

ks_mixte <- ks.test(data_mixte$excess_unique,
                    function(x) pgpd(x, scale = est["scale"], shape = est["shape"]),
                    exact = FALSE)

cat("\n=== Test KS Groupe mixte ===\n")
print(ks_mixte)

xi <- est["shape"]
beta <- est["scale"]
if (xi < 0) {
  omega <- seuil - (beta / xi)
  cat("\n>>> Age ultime estime :", round(omega, 6), "ans\n")
} else {
  cat("\n>>> xi >= 0 -> pas de borne superieure (omega infini)\n")
}

```

C.7 Exploration d'un modèle exponentiel au-delà de 110 ans

Groupe 2 (femmes et hommes)

```

# =====
# GROUPE 2
# =====

library(readxl)
library(dplyr)
library(lubridate)
library(MASS)

# =====
# Parametre general
# =====
seuil <- 110

# =====
# Chargement des donnees (feuille IDL_BEL_Death_105)
# =====
df <- read_excel("idl_bel_death_105_analyse.xlsx", sheet = "IDL_BEL_Death_105")

# Exclusion des cohortes 1886-1904

```

```

df <- df %>% filter(BIRTH_YEAR < 1886 | BIRTH_YEAR > 1904)

# Construction des exces
data_mixte <- df %>%
  mutate(
    BIRTH_Date = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),
    DEATH_Date = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),
    AgeExact = time_length(interval(BIRTH_Date, DEATH_Date), unit = "years"),
    excess = AgeExact - seuil
  ) %>%
  filter(excess > 0)

# =====
# Correction des doublons
# =====
data_mixte <- data_mixte %>%
  arrange(excess) %>%
  group_by(excess) %>%
  mutate(
    n = n(),
    excess_corrige = if_else(n > 1, excess + row_number() * 1e-10, excess)
  ) %>%
  ungroup() %>%
  dplyr::select(-n)

# =====
# Methode 1 : lambda estime (via MLE)
# =====
fit_exp <- fitdistr(data_mixte$excess_corrige, "exponential")
lambda_mle <- fit_exp$estimate["rate"]

ks_mle <- ks.test(data_mixte$excess_corrige, "pexp", rate = lambda_mle)

cat("\n=== Test KS - Loi exponentielle (lambda estime) ===\n")
cat("lambda estime :", round(lambda_mle, 5), "\n")
print(ks_mle)

# =====
# Methode 2 : lambda fixe a 1/1.34
# =====
lambda_rz <- 1/1.34

ks_fix <- ks.test(data_mixte$excess_corrige, "pexp", rate = lambda_rz)

cat("\n=== Test KS - Loi exponentielle (lambda = 1/1.34 fixe) ===\n")
cat("lambda fixe :", lambda_rz, "\n")
print(ks_fix)

# Statistiques complementaires
cat("Nombre d'exces au-dela de 110 ans :", nrow(data_mixte), "\n")
table(data_mixte$SEX)

```

C.8 Présentation des données BNB

```

library(readxl)
library(dplyr)
library(tidyr)
library(ggplot2)

# Lecture des donnees
fichier_bnb <- "201822globalmortalityon5years.xlsx"
bnb_raw <- read_excel(fichier_bnb, sheet = 1, skip = 1)

# Renommage des colonnes

```

```

colnames(bnb_raw) <- c("AGE", "ASSURE_H", "DECES_H", "ASSURE_F", "DECES_F")

# Calcul des qx (tous ages)
bnb_clean <- bnb_raw %>%
  mutate(across(everything(), as.numeric)) %>%
  filter(!is.na(AGE), AGE >= 0, AGE <= 99) %>%
  mutate(
    qx_homme = DECES_H / ASSURE_H,
    qx_femme = DECES_F / ASSURE_F,
    qx_mixte = (DECES_H + DECES_F) / (ASSURE_H + ASSURE_F)
  )

# Passage au format long pour graphique
bnb_long <- bnb_clean %>%
  dplyr::select(AGE, qx_homme, qx_femme, qx_mixte) %>%
  pivot_longer(cols = starts_with("qx_"),
    names_to = "Groupe", values_to = "qx") %>%
  mutate(Groupe = recode(Groupe,
    qx_homme = "Hommes",
    qx_femme = "Femmes",
    qx_mixte = "Mixte"))

# Trace des probabilités de décès
ggplot(bnb_long, aes(x = AGE, y = qx, color = Groupe)) +
  geom_line(size = 1.2) +
  labs(
    title = "Probabilités de décès brutes par âge",
    subtitle = "Données BNB 2018--2022, volet décès",
    x = "Age",
    y = expression("Probabilité de décès " * q[x]),
    color = "Groupe"
  ) +
  theme_classic(base_size = 14) +
  theme(
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    plot.title = element_text(hjust = 0.5, face = "bold", margin = margin(b = 10)),
    plot.subtitle = element_text(hjust = 0.5, margin = margin(b = 20)),
    axis.line = element_line(color = "black"),
    axis.ticks = element_line(color = "black"),
    panel.background = element_rect(fill = "white", colour = NA),
    plot.background = element_rect(fill = "white", colour = NA),
    legend.position = "bottom",
    legend.title = element_text(face = "bold"),
    legend.background = element_rect(fill = "white", colour = NA)
  ) +
  scale_color_manual(values = c("Femmes" = "#FFADAD", "Hommes" = "#A0C4FF", "Mixte" = "brown"))

# Calcul des qx pour tableau (ages 60+)
bnb_clean <- bnb_raw %>%
  mutate(across(everything(), as.numeric)) %>%
  filter(!is.na(AGE), AGE >= 60, AGE <= 99) %>%
  mutate(
    qx_homme = DECES_H / ASSURE_H,
    qx_femme = DECES_F / ASSURE_F,
    qx_mixte = (DECES_H + DECES_F) / (ASSURE_H + ASSURE_F)
  )

# Extraction du tableau des qx
table_qx_bnb <- bnb_clean %>%
  dplyr::select(AGE, qx_femme, qx_homme, qx_mixte)

print(table_qx_bnb)

```


C.9 Construction des tables d'expérience

```
# =====  
# FEMME  
# =====  
  
# 1. Parametrage et chargement des librairies  
library(readxl)  
library(dplyr)  
library(ggplot2)  
library(lubridate)  
library(POT)  
  
# Chemins de fichiers  
fichier_bnb <- "201822globalmortalityon5years.xlsx"  
fichier_idl <- "idl_bel_death_105_analyse.xlsx"  
  
# 2. Preparation des donnees BNB  
bnb_raw <- read_excel(fichier_bnb, sheet = 1, skip = 1)  
colnames(bnb_raw) <- c("AGE", "ASSURE_H", "DECES_H", "ASSURE_F", "DECES_F")  
data_f <- bnb_raw %>%  
  mutate(across(everything(), as.numeric)) %>%  
  filter(!is.na(AGE), AGE >= 60 & AGE <= 99) %>%  
  mutate(qx = DECES_F / ASSURE_F)  
  
# 3. Ajustement du modele de Makeham  
makeham_fn <- function(age, A, B, c) {  
  1 - exp(-(A + (B / log(c)) * (c^(age + 1) - c^age)))  
}  
fit_makeham <- nls(qx ~ A + B * c^AGE, data = data_f,  
  start = list(A = 0.0001, B = 1e-7, c = 1.1),  
  algorithm = "port",  
  lower = c(0, 0, 1.0001), upper = c(1, 1, 1.5))  
params <- coef(fit_makeham)  
summary(fit_makeham)  
  
# 4. Donnees IDL et ajustement GPD  
seuil <- 105  
df <- read_excel(fichier_idl, sheet = "IDL_BEL_Death_105") %>%  
  filter(BIRTH_YEAR < 1886 | BIRTH_YEAR > 1904)  
data_femmes_idl <- df %>%  
  filter(SEX == "F") %>%  
  mutate(  
    BIRTH_Date = make_date(BIRTH_YEAR, BIRTH_MONTH, BIRTH_DAY),  
    DEATH_Date = make_date(DEATH_YEAR, DEATH_MONTH, DEATH_DAY),  
    AgeExact = time_length(interval(BIRTH_Date, DEATH_Date), unit = "years"),  
    excess = AgeExact - seuil  
  ) %>%  
  filter(excess > 0)  
fit_gpd <- fitgpd(data_femmes_idl$excess, threshold = 0, est = "mle")  
xi <- fit_gpd$fitted.values["shape"]  
beta <- fit_gpd$fitted.values["scale"]  
omega <- seuil - beta / xi  
  
gpd_fn <- function(age, seuil, xi, beta) {  
  Sx <- (1 + xi * (age - seuil) / beta)^(-1 / xi)  
  Sx1 <- (1 + xi * (age + 1 - seuil) / beta)^(-1 / xi)  
  (Sx - Sx1) / Sx  
}  
  
# 5. Spline cubique de Hermite  
x0 <- 95  
x1 <- 105  
y0 <- makeham_fn(x0, params["A"], params["B"], params["c"])  
m0 <- (makeham_fn(x0 + 1, params["A"], params["B"], params["c"]) - y0)
```

```

y1 <- gpd_fn(x1, seuil, xi, beta)
m1 <- (gpd_fn(x1 + 1, seuil, xi, beta) - y1)

spline_hermite <- function(x) {
  t <- (x - x0) / (x1 - x0)
  h00 <- 2 * t^3 - 3 * t^2 + 1
  h10 <- t^3 - 2 * t^2 + t
  h01 <- -2 * t^3 + 3 * t^2
  h11 <- t^3 - t^2
  h00 * y0 + h10 * (x1 - x0) * m0 + h01 * y1 + h11 * (x1 - x0) * m1
}

ages_spline <- x0:x1
qx_spline <- spline_hermite(ages_spline)
df_spline <- data.frame(Age = ages_spline, qx = round(qx_spline, 6))

# 6. Construction table finale
ages_makeham_pure <- 60:(x0 - 1)
qx_makeham_pure <- makeham_fn(ages_makeham_pure, params["A"], params["B"], params["c"])
df_makeham_pure <- data.frame(Age = ages_makeham_pure, qx = round(qx_makeham_pure, 6))

ages_gpd_pure <- (x1):floor(omega)
qx_gpd_pure <- gpd_fn(ages_gpd_pure, seuil, xi, beta)
df_gpd_pure <- data.frame(Age = ages_gpd_pure, qx = round(qx_gpd_pure, 6))

table_finale_f <- bind_rows(df_makeham_pure, df_spline, df_gpd_pure) %>%
  distinct(Age, .keep_all = TRUE)

# 7. Visualisation
library(ggplot2)
library(dplyr)

ggplot() +
  geom_line(data = table_finale_f, aes(x = Age, y = qx, color = "Modele Makeham"),
    size = 1.4) +
  geom_line(data = df_spline, aes(x = Age, y = qx, color = "Spline"), size = 1) +
  geom_line(data = df_gpd_pure, aes(x = Age, y = qx, color = "Modele GPD"), size = 1) +
  geom_point(data = data_f, aes(x = AGE, y = qx, color = "qx bruts BNB"), alpha = 0.6, size = 1) +
  geom_vline(xintercept = 95, linetype = "dashed", color = "grey60", size = 0.7) +
  geom_vline(xintercept = 105, linetype = "dashed", color = "grey60", size = 0.7) +
  labs(
    title = "Table de mortalite raccordee - Femmes",
    subtitle = "Makeham (60-94), Spline de Hermite (95-104), GPD (105-omega)",
    x = "Age",
    y = expression("Probabilite de deces " * q[x])
  ) +
  theme_classic(base_size = 14) +
  theme(
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    plot.title = element_text(hjust = 0.5, face = "bold", size = 16, margin = margin(b = 10)),
    plot.subtitle = element_text(hjust = 0.5, size = 12, margin = margin(b = 20)),
    axis.title = element_text(size = 13),
    axis.text = element_text(size = 11),
    axis.line = element_line(color = "black", size = 0.5),
    axis.ticks = element_line(color = "black", size = 0.5),
    panel.background = element_rect(fill = "white", colour = NA),
    plot.background = element_rect(fill = "white", colour = NA),
    legend.position = "bottom",
    legend.title = element_text(face = "bold"),
  )

```

```

legend.background = element_rect(fill = "white", colour = NA),
legend.key.size = unit(0.8, "cm")
) +
scale_color_manual(
  values = c(
    "Modele Makeham" = "#005B82",
    "Spline" = "green4",
    "Modele GPD" = "#D95F02",
    "qx bruts BNB" = "#A60026"
  ),
  breaks = c("Modele Makeham", "Spline", "Modele GPD", "qx bruts BNB")
)

```

Hommes et groupe mixte

La construction des tables d'expérience pour les hommes et pour le groupe mixte suit exactement la même logique que celle décrite pour les femmes. Seuls changent les jeux de données utilisés pour l'ajustement des modèles Makeham et GPD.

Construction de la table XK et table unique comparée

```

# =====
# TABLE XK
# =====

# Ages et lx de la table XK
ages <- 60:106
lx <- c(
  842198, 830436, 817670, 803829, 788839, 772629, 755129, 736271, 715996, 694252,
  670997, 646205, 619869, 592004, 562652, 531885, 499814, 466586, 432395, 397476,
  362112, 326628, 291389, 256792, 223253, 191193, 161023, 133118, 107802, 85320,
  65824, 49360, 35861, 25150, 16958, 10942, 6723, 3910, 2139, 1093,
  517, 224, 89, 31, 10, 3, 1
)

# Calcul des qx
qx <- numeric(length(lx))
for (i in 1:(length(lx) - 1)) {
  qx[i] <- (lx[i] - lx[i + 1]) / lx[i]
}
qx[length(lx)] <- 1 # a 106 ans, probabilite de deces = 1

# Creation du data frame
xk_table <- data.frame(Age = ages, lx = lx, qx = qx)

# Affichage
print(xk_table)

# =====
# TABLE UNIQUE qx FEMMES, HOMMES, MIXTE + XK
# =====

# Renommer les colonnes pour distinguer les groupes
table_f <- table_finale_f %>% rename(qx_femme = qx)
table_h <- table_finale_h %>% rename(qx_homme = qx)
table_m <- table_finale_m %>% rename(qx_mixte = qx)

# Creer une table XK avec colonnes Age et qx_XK
table_XK <- data.frame(Age = 60:106, qx_XK = xk_table$qx)

# Jointure sur l'age pour obtenir une table unique
table_qx <- table_f %>%
  full_join(table_h, by = "Age") %>%

```

```

full_join(table_m, by = "Age") %>%
full_join(table_XK, by = "Age") %>%
arrange(Age)

# Affichage
print(table_qx)

```

C.10 Illustration numérique

Calcul des primes uniques selon la table

```

# =====
#   TABLEAU RECAP DES PRIMES
# =====

library(dplyr)
library(tidyr)

# Parametres globaux
i <- 0.01
v <- 1 / (1 + i)
capital <- 1
ages <- c(60, 65, 70)

# Ages ultimes pour chaque table
omega_femme <- 126
omega_homme <- 112
omega_mixte <- 124
omega_xk <- 107

# Fonction de calcul
calcul_prime <- function(qx_vect, omega, x) {
  qx_vect <- qx_vect[!is.na(qx_vect)]
  qx_vect <- qx_vect[1:(omega - x)]
  px_vect <- 1 - qx_vect
  kpx <- cumprod(c(1, px_vect[-length(px_vect)]))
  k <- 0:(length(qx_vect) - 1)
  sum(kpx * qx_vect * v^(k + 1) * capital)
}

# Calcul des primes par age et par table
resultats <- data.frame()

for (x in ages) {
  qx_femme <- table_qx %>% filter(Age >= x) %>% pull(qx_femme)
  qx_homme <- table_qx %>% filter(Age >= x) %>% pull(qx_homme)
  qx_mixte <- table_qx %>% filter(Age >= x) %>% pull(qx_mixte)
  qx_xk <- table_qx %>% filter(Age >= x) %>% pull(qx_XK)

  resultats <- rbind(
    resultats,
    data.frame(
      Age = x,
      Femmes = calcul_prime(qx_femme, omega_femme, x),
      Hommes = calcul_prime(qx_homme, omega_homme, x),
      Mixte = calcul_prime(qx_mixte, omega_mixte, x),
      XK = calcul_prime(qx_xk, omega_xk, x)
    )
  )
}

# Affichage final du tableau
print(resultats)

```

Calcul des réserves à partir de 65 ans

```
# =====  
#   CALCUL RESERVE A 65 ANS POUR CHAQUE TABLE  
# =====  
  
library(dplyr)  
library(tidyr)  
  
# Parametres  
x <- 65  
i <- 0.01  
v <- 1 / (1 + i)  
capital <- 1  
durees <- 0:41  
  
# Ages ultimes  
omega_femme <- 126  
omega_homme <- 112  
omega_mixte <- 124  
omega_xk <- 107  
  
# Fonction de calcul de la reserve  
calcul_reserve <- function(qx_vect, age_debut, omega) {  
  if (age_debut >= omega) return(0)  
  qx_vect <- qx_vect[!is.na(qx_vect)]  
  qx_vect <- qx_vect[1:(omega - age_debut)]  
  px_vect <- 1 - qx_vect  
  kpx <- cumprod(c(1, px_vect[-length(px_vect)]))  
  k <- 0:(length(qx_vect) - 1)  
  sum(kpx * qx_vect * v^(k + 1) * capital)  
}  
  
# Calcul  
reserves <- data.frame()  
  
for (t in durees) {  
  age_t <- x + t  
  
  qx_femme <- table_qx %>% filter(Age >= age_t) %>% pull(qx_femme)  
  qx_homme <- table_qx %>% filter(Age >= age_t) %>% pull(qx_homme)  
  qx_mixte <- table_qx %>% filter(Age >= age_t) %>% pull(qx_mixte)  
  qx_xk <- table_qx %>% filter(Age >= age_t) %>% pull(qx_XK)  
  
  reserves <- rbind(  
    reserves,  
    data.frame(  
      Age = age_t,  
      Femmes = calcul_reserve(qx_femme, age_t, omega_femme),  
      Hommes = calcul_reserve(qx_homme, age_t, omega_homme),  
      Mixte = calcul_reserve(qx_mixte, age_t, omega_mixte),  
      XK = calcul_reserve(qx_xk, age_t, omega_xk)  
    )  
  )  
}  
  
# Affichage final  
print(reserves)
```

Visualisation graphique

```
library(ggplot2)  
library(tidyr)  
library(dplyr)
```

```

# Long format
reserves_long <- reserves %>%
  pivot_longer(cols = c(Femmes, Hommes, Mixte, XK),
               names_to = "Table", values_to = "Reserve")

# Filtrage selon la table
reserves_long <- reserves_long %>%
  filter(
    (Table == "Femmes") |
    (Table == "Hommes" & Age <= 111) |
    (Table == "Mixte" & Age <= 123) |
    (Table == "XK" & Age <= 106)
  )

# Primes a 65 ans
primes_ref <- data.frame(
  Table = c("Femmes", "Hommes", "Mixte", "XK"),
  Prime = c(
    reserves %>% filter(Age == 65) %>% pull(Femmes),
    reserves %>% filter(Age == 65) %>% pull(Hommes),
    reserves %>% filter(Age == 65) %>% pull(Mixte),
    reserves %>% filter(Age == 65) %>% pull(XK)
  )
)

# Graphique
ggplot(reserves_long, aes(x = Age, y = Reserve, color = Table)) +
  geom_line(size = 1.2) +
  geom_hline(data = primes_ref, aes(yintercept = Prime, color = Table),
            linetype = "dashed", size = 0.8) +
  scale_color_manual(values = c(
    "Femmes" = "#FFADAD",
    "Hommes" = "#A0C4FF",
    "Mixte" = "brown",
    "XK" = "#229954"
  )) +
  labs(
    title = "Evolution de la Reserve Mathematique Comparaison des Tables",
    subtitle = "Prime unique versee a 65 ans (lignes pointillees)",
    x = "Age de l'assure",
    y = "Reserve Mathematique (euro)",
    color = "Table de mortalite"
  ) +
  theme_minimal() +
  theme(
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    panel.background = element_rect(fill = "white", colour = NA),
    plot.background = element_rect(fill = "white", colour = NA),
    axis.line = element_line(colour = "black"),
    plot.title = element_text(hjust = 0.5, face = "bold", size = 16),
    plot.subtitle = element_text(hjust = 0.5, size = 12, color = "gray30"),
    axis.title = element_text(size = 13, face = "bold"),
    axis.text = element_text(size = 11),
    legend.title = element_text(size = 12),
    legend.text = element_text(size = 11),
    legend.position = "bottom"
  )

```

