



# UNIVERSITÉ PARIS DAUPHINE

Département MIDO<sup>(\*)</sup>

MASTER MIDO  
MENTION MMD<sup>(\*\*)</sup>

SPÉCIALITÉ ACTUARIAT

Année Universitaire : 2009-2010

Mémoire d'Actuariat présenté en novembre 2010 devant l'Université Paris  
Dauphine et l'Institut des Actuaire

Par : *Thuy-Linh NGUYEN*

Tuteur : *Riada DJEBBAR*  
*Bruno VIAL*

Sujet : *Etude et modélisation de la sévérité des sinistres en assurance crédit*  
Entreprise d'accueil : *COFACE*

**CONFIDENTIEL - 5 ans**

## JURY

Jean Marie Nessi  
Pierre Mathoulin  
Stéphane Ménart

(\*) MIDO : Mathématiques, Informatique, Décision, Organisation

(\*\*) MMD : Mathématiques, Modélisation, Décision



## *Remerciements*

En préambule à ce mémoire, je souhaite adresser mes remerciements aux personnes qui ont contribué au bon déroulement de mon stage.

Je tiens à remercier **M. Bruno Vial**, responsable du service Sesame et **M<sup>me</sup> Riada Djebbar**, ma tutrice de stage pour leur suivi et pour l'aide et le temps qu'ils m'ont accordés tout au long du stage.

Je remercie également **M<sup>me</sup> Séverine Outreville** et **M<sup>me</sup> Laëtitia Aka**, membres de Sesame pour leur disponibilité et leurs compétences et pour les précisions qu'elles m'ont apportées.

Je remercie **M<sup>me</sup> Jennifer Forest**, membre de Sesame pour sa sympathie mais aussi pour ses conseils : ils me seront précieux dans ma future vie professionnelle.

Mes remerciements à **M. Frantz Bouchereau** et **M. Severin Resch** pour leur gentillesse et leur professionnalisme.

Merci à mes homologues stagiaires **Djeumbe Sambe**, **Bruno Tallard** et **Hugues Rattaire** pour les moments partagés. Je leur souhaite de la réussite dans leur vie professionnelle.

Enfin, je tiens à remercier l'équipe des professeurs de l'Université Paris-Dauphine et en particulier **M. Christian Hess** pour leur enseignement durant la formation universitaire d'actuariat.



## Résumé

Les risques portés par Coface proviennent principalement de son activité en assurance crédit. Il est donc important pour Coface d'avoir un modèle de risque de crédit qui décrit au mieux la réalité et qui lui permette une bonne gestion de ses risques. C'est dans ce cadre que s'intègre le stage de fin d'études.

Ce mémoire d'actuariat s'intitule *Etude et modélisation de la sévérité des sinistres en assurance crédit*. La sévérité est un paramètre qui mesure l'intensité d'un sinistre lorsque celui-ci survient. Elle est schématiquement mesurée par le rapport entre le montant de sinistre et le montant de l'exposition. Elle constitue un des modules du modèle actuel du risque de crédit Coface. L'objectif du mémoire a été d'étudier et de modéliser cette variable au niveau de détail acheteur.

La première étape a consisté à donner une définition de la sévérité de niveau acheteur et à élaborer une base de données fiable. Cette étape a été suivie d'une analyse descriptive de la sévérité de niveau acheteur qui nous a permis de détecter quelques caractéristiques : par exemple, l'importance du facteur nombre de polices exposées, la distinction entre défaut total et partiel ou encore le fait que la majorité des sinistres ne porte que sur une seule police même pour les acheteurs exposés à travers plusieurs polices. Ce sont sur ces deux étapes que la calibration statistique a reposé: l'analyse descriptive a permis de cibler correctement le modèle tandis que la fiabilité des données a fourni de bonnes estimations des paramètres du modèle. L'étape suivante a été de simuler cette sévérité notamment grâce à des méthodes de Monte Carlo. Enfin, une application directe a été d'intégrer la sévérité dans le modèle du risque de crédit afin d'estimer les pertes.



## *Abstract*

The risks of Coface are mainly due to its credit insurance activity. That is why it is essential to Coface to have a reliable model which describe the credit risk and which enable Coface to have an efficient risk management. The work placement has taken place within this context.

The title of the report is Study and Modeling of the severity of losses in credit insurance. The severity is a parameter which gives the intensity of a loss when this one arises. We can simply define it as the ratio of loss amount to the exposure amount. It is one of the components of the current Coface risk credit model. The work placement aimed at studying and modeling this variable on the debtor level.

The first step has consisted in giving a definition of the severity and in working out a reliable database. This step has been followed by a descriptive analysis which showed several phenomena like the significance of the number of policies, the difference between a part default and a total default or the fact of most of losses concern only one policy even if the debtors are exposed by several policies. The statistical adjustment depended on these two previous steps. Indeed the descriptive analysis enabled us to well target the model and the reliability has been of the database brought us good estimations of the parameters. The next step was the simulation of the severity by using Monte Carlo methods. Finally, the severity has been inserted in the current risk credit model in order to assess the losses.



# Sommaire

Introduction.....	11
Chapitre I Introduction.....	13
I.1. Présentation de l'entreprise .....	13
I.1.1. Présentation de Coface .....	13
I.1.2. Présentation de SESAME.....	13
I.2. Contexte et problématique de l'étude .....	14
I.3. L'assurance crédit.....	15
I.3.1. Terminologie utilisée .....	15
I.3.2. Notation interne et probabilité de défaut .....	18
I.3.3. Chronologie d'un sinistre .....	20
I.3.4. Les modèles du risque de crédit.....	21
Chapitre II Les données.....	27
II.1. Architecture des systèmes d'informations Coface .....	27
II.2. Présentation des données disponibles .....	28
II.3. Définition retenue de la sévérité .....	29
Chapitre III Analyse descriptive.....	33
III.1. Analyse descriptive.....	33
III.1.1. Caractéristiques de la variable Severite de niveau acheteur .....	33
III.1.2. Dépendance de la sévérité avec les autres variables .....	36
III.2. Modèle ciblé .....	47
Chapitre IV Ajustement et simulation de lois.....	49
IV.1. Lois utilisées .....	49
IV.2. Procédés d'ajustement.....	52
IV.2.1. Procédés classiques .....	52
IV.2.2. Procédé « intuitif » .....	54
IV.2.3. Adéquation à une loi.....	55
IV.3. Procédés de simulation .....	56
IV.3.1. Simulation d'une variable aléatoire.....	56
IV.3.2. Simulation d'un mélange .....	57
IV.3.3. Simulation de variables corrélées.....	58

Chapitre V	Application au modèle de sévérité de niveau acheteur.....	61
V.1.	Modèle de sévérité au global .....	61
V.1.1.	Forme de l'histogramme.....	61
V.1.2.	Ajustement .....	62
V.1.3.	Simulation .....	65
V.2.	Modèle de sévérité en fonction des facteurs discriminants.....	65
V.2.1.	Modèle de sévérité sur un agrément.....	67
V.2.2.	Loi de la proportion de polices sinistrées.....	70
V.2.3.	Loi du montant garanti.....	73
V.2.4.	Résultats.....	74
Chapitre VI	Intégration du modèle de sévérité dans la modélisation du risque de crédit.....	79
Conclusion.....		83
Références bibliographiques .....		85
Annexe.....		87

## Introduction

Ce rapport est un mémoire d'actuariat qui s'inscrit dans le cadre de la formation du M2 Actuariat de Paris-Dauphine. Il est présenté dans le but d'être validé par l'Institut des Actuaire.

Le mémoire s'intitule Etude et modélisation de la sévérité des sinistres en assurance crédit. Il a fait l'objet d'un stage de six mois chez Coface<sup>1</sup> au sein du service SESAME<sup>2</sup> de la direction générale sous la responsabilité de Riada DJEBBAR, membre de l'équipe Sesame et de Bruno Vial, responsable du service.

Coface est le 2<sup>ème</sup> assureur crédit mondial. Les risques portés par Coface proviennent principalement de son activité en assurance crédit. La quantification et la gestion des risques de crédit représentent alors un besoin important pour Coface. Sesame a développé dans ce but un modèle de risque de crédit qui prend en compte la probabilité de défaut, le montant d'exposition et l'intensité du sinistre.

C'est dans ce contexte que s'intègre mon stage car la sévérité des sinistres représente un des modules dans le modèle actuel du risque de crédit de Coface. La sévérité mesure l'intensité d'un sinistre quand celui-ci intervient. Elle est définie schématiquement comme le rapport du montant d'indemnisation sur le montant d'exposition. Nous verrons qu'il existe plusieurs niveaux d'agrégation des sinistres. Mon sujet abordera la sévérité de niveau acheteur. Son étude et modélisation est un axe d'amélioration du modèle de risque.

Le premier chapitre complète l'introduction en présentant Coface et ses différents métiers. Il décrit la problématique du sujet de mémoire. Ce chapitre reflète le travail de compréhension des mécanismes de l'assurance crédit qui a été effectué dans un premier temps.

Les chapitres suivants développent les différents travaux menés pour l'élaboration du mémoire.

---

<sup>1</sup> *Compagnie Française d'Assurance pour le Commerce Extérieur*

<sup>2</sup> *Service des Etudes Statistiques, Actuarielles, Mathématiques et Econométriques*

Avant toute étude ou modélisation, il est nécessaire d'avoir des données, le deuxième chapitre est consacré à la présentation des données disponibles et à l'élaboration d'une base de données fiable.

L'analyse de ses données dans le troisième chapitre permettra de détecter les facteurs discriminants de la sévérité de niveau acheteur et différents phénomènes. Cette analyse nous guidera vers le type de modèle sélectionné.

Le quatrième chapitre est un chapitre théorique qui décrit les méthodes utilisées pour la modélisation. Les lois, les méthodes d'ajustement et les méthodes de simulation utilisées y seront présentées.

Dans le cinquième chapitre, nous appliquerons la théorie à la modélisation de la sévérité. Dans une première partie, nous élaborerons un modèle de la sévérité au global et dans une seconde partie, la sévérité sera modélisée en fonction des facteurs discriminants.

Enfin le dernier chapitre est une application directe du modèle de sévérité. Nous intégrerons la sévérité modélisée dans le modèle de risque de crédit afin d'estimer les pertes.

# Chapitre I Introduction

Ce premier chapitre a pour objectif de présenter le cadre et le contexte dans lesquels a été réalisée l'étude qui fait le sujet de ce mémoire d'actuariat. L'ensemble des notions utilisées en assurance crédit et le modèle d'assurance crédit y seront présentés.

## I.1. Présentation de l'entreprise

### I.1.1. Présentation de Coface

Coface est créée au lendemain de la seconde guerre mondiale en 1946 dans un contexte de développement du commerce extérieur. Elle représente à l'origine une compagnie d'assurance spécialisée dans l'assurance-crédit à l'exportation française. Elle diversifie par la suite son activité pour pratiquer de nos jours, outre l'assurance-crédit, l'information et la notation d'entreprise, la gestion de créances et l'affacturage. Ces quatre métiers interviennent dans sa principale mission, à savoir faciliter les échanges entre les entreprises partout dans le monde. Coface assure également la gestion des garanties publiques à l'exportation pour le compte de l'Etat. Depuis 2002, elle est une filiale de Natixis, banque de financement et de services financiers.

Coface met à la disposition de ses 130 000 clients, plus de 6800 collaborateurs répartis dans 67 pays où elle est directement présente. De plus, Coface fait partie de CreditAlliance, premier réseau mondial dans les métiers du Poste clients ce qui lui permet d'être présente dans 32 pays supplémentaires.

### I.1.2. Présentation de SESAME

Sesame est le Service des Etudes Statistiques, Actuarielles, Mathématiques et Econométriques de Coface. Le service est intégré à la direction générale et est sous la responsabilité de Bruno Vial. L'équipe Sesame comprend une dizaine de collaborateurs. L'activité de Sesame est axée autour de trois pôles : le scoring, la tarification et la modélisation des risques.

Le scoring est le pôle en charge de la notation des entreprises. Le système de notation d'entreprises françaises (score @Rating) se base sur les informations disponibles

sur les entreprises et fournit une échelle exprimée en probabilité de défaut. Plus de 6,5 millions d'entreprises françaises sont notées par ce système.

Le pôle tarification a pour rôle de créer des outils permettant la tarification des produits Coface.

Le pôle risque a pour mission la modélisation et la quantification des risques portés par Coface en assurance crédit. Il établit des modèles et des outils d'aide à la décision en termes de gestion des risques.

## 1.2. Contexte et problématique de l'étude

L'étude réalisée pour le mémoire est intégrée dans l'activité risque. Les travaux ont été menés dans le cadre de la modélisation du risque de crédit. Plus précisément, nous nous intéresserons à une variable présente dans le modèle de risque de crédit : la sévérité. La sévérité nous informe sur l'intensité du sinistre provoqué par un défaut.

Le mémoire est structuré de façon à répondre aux questions suivantes :

1. Quelle est la définition de la sévérité ?
2. Comment la mesurer ?
3. Quels en sont les facteurs discriminants ?
4. Comment la modéliser et la simuler ?
5. Qu'apporte sa modélisation-simulation au modèle du risque de crédit?

Le chapitre II répondra aux questions 1 et 2, le chapitre III répondra à la question 3, le chapitre V répondra à la question 4 et le chapitre VI répondra à la dernière question à travers des applications directes.

Le portefeuille de risques de Coface, sur lequel porte l'étude, est un portefeuille international qui regroupe donc des risques provenant de pays différents. La notion de défaut pouvant différer selon le pays, il s'agit d'un portefeuille hétérogène. Par ailleurs, l'univers de l'assurance crédit est un univers pauvre compte tenu du peu d'informations « a priori ».

L'objectif de ce mémoire est donc d'avoir une vision globale du comportement de la sévérité et d'identifier les phénomènes qui l'expliquent.

## I.3. L'assurance crédit

L'assurance crédit fait partie de l'assurance non-vie. Coface en est un des acteurs majeurs. Il s'agit d'une forme particulière d'assurance qui ne touche pas la totalité de la population. En effet, les produits d'assurance crédit sont destinés aux professionnels. L'assurance crédit a pour objet de protéger ses assurés contre le risque de crédit, le risque de crédit correspondant au risque de non remboursement des créances commerciales.

### I.3.1. Terminologie utilisée

Coface accepte d'assurer une entreprise en lui délivrant une ou plusieurs **polices** d'assurance. Ainsi Coface s'engage à indemniser cet assuré en cas de non remboursement de créances commerciales qu'il détient sur ses entreprises clientes.

Le terme **assureur** désigne Coface. Le terme **assuré** désigne les clients de Coface. Le terme **acheteur** désigne les entreprises créancières, clientes de l'assuré. Inversement, l'assuré représente le **fournisseur** de ces entreprises créancières.

Le terme **encours** désigne le montant de crédit consenti par un assuré à un de ses acheteurs à un moment donné. Un assuré va demander à Coface de couvrir une partie ou la totalité de l'encours commercial sur chacun des acheteurs. Nous disons qu'il fait une demande d'**agrément** pour chacun de ses acheteurs. Coface peut répondre positivement à cette demande : elle accepte de couvrir l'assuré à hauteur d'un certain montant que nous appelons **montant de l'agrément** ou **montant garanti**. Il est enregistré à la souscription de la police mais il peut évoluer à l'initiative de Coface ou de l'assuré. De plus, un autre élément vient caractériser l'agrément. Il s'agit de la **quotité garantie QG** qui représente la part que Coface prend à sa charge dans l'indemnisation des sinistres, la **quotité non garantie QNG** étant la part restant à la charge de l'assuré. Cette clause a pour objectif d'intéresser l'assuré à la gestion de ses risques sachant que sa couverture n'est pas totale.

La police d'assurance d'un assuré contient donc plusieurs agréments. D'autre part, la police peut contenir une **franchise** : il y a indemnisation si le montant cumulé des sinistres sur les différents acheteurs est supérieure à la valeur seuil de la franchise.

L'**exposition** désigne un montant en risque pour Coface. Elle correspond à une somme de montants garantis qui dépend du choix d'agrégation.

Le **risque** auquel Coface est exposée est celui de non paiement des créances des acheteurs à leurs fournisseurs. Il ne porte pas sur une défaillance de l'assuré mais sur une défaillance de ses acheteurs. Cette remarque est fondamentale et est prise en compte dans la modélisation et la quantification des risques.

Le terme **groupe** désigne un ensemble d'acheteurs qui ont la même maison mère, le nombre d'acheteurs étant supérieur à deux.

Par ailleurs, il est fréquent que des agréments soient délivrés à plusieurs assurés sur un même acheteur ou sur des acheteurs d'un même groupe : il y a une accumulation du risque sur ces entreprises et qu'il existe une corrélation entre les risques non négligeable. L'analyse de risque ne peut donc se limiter à une analyse police par police.

Sur le schéma ci-dessous, les agréments A2 et B1 portent sur le même acheteur (acheteur 2) bien qu'ils soient issus de deux polices d'assurance distinctes (police 1 et police B).

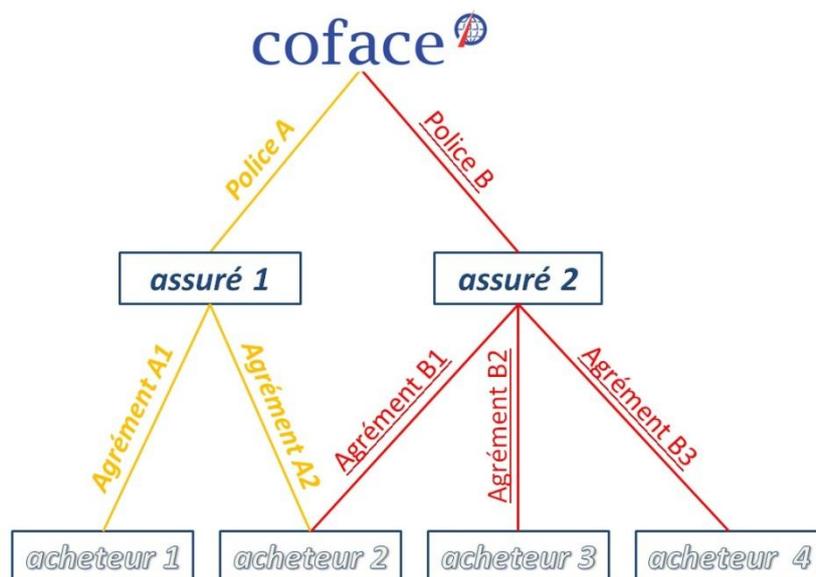


Schéma des interactions entre les agréments

Les schémas ci-dessous nous illustrent l'assurance crédit en situation de prévention puis en situation d'impayé.

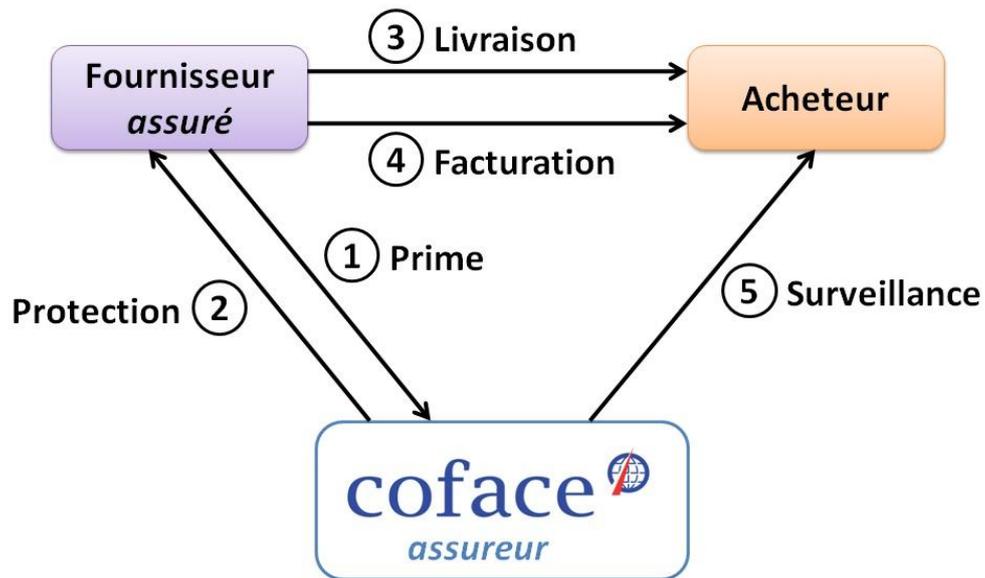


Schéma de l'assurance crédit en situation de prévention

Le versement de la prime par l'assuré lui garantit la protection de Coface contre les impayés. La livraison de marchandises est accompagnée d'une émission de facture.

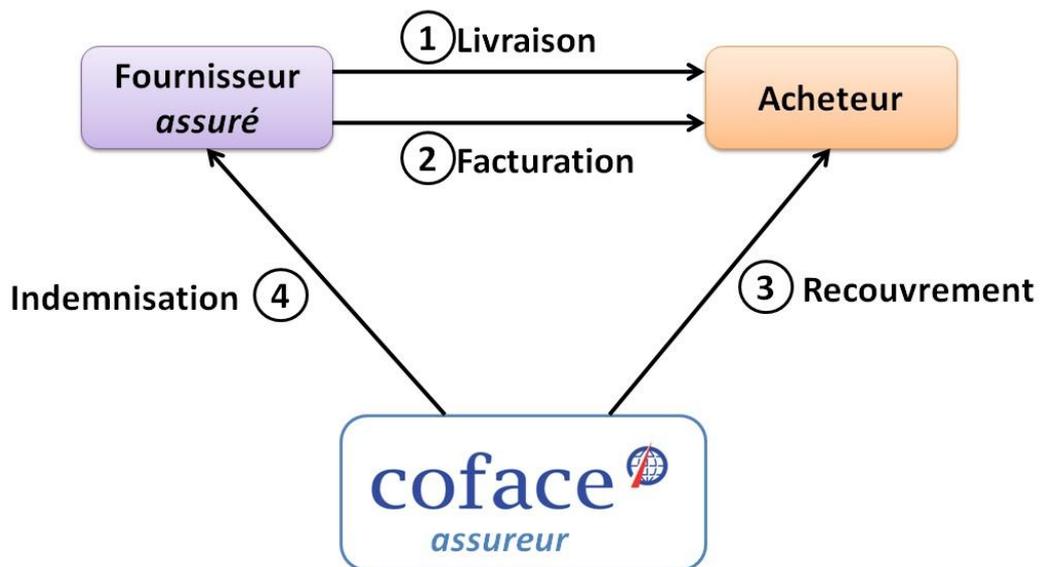


Schéma de l'assurance crédit en situation d'impayés

S'il y a un impayé suite à l'émission de la facture, Coface procède au recouvrement puis indemnise l'assuré.

S'il n'y a pas d'impayé, l'assuré a versé une prime inutilement. Au contraire, si l'assuré a un montant d'impayés important alors il est protégé par Coface. Les entreprises assurées n'ont pas besoin de constituer de réserves importantes en prévention des risques et elles peuvent alors entièrement se consacrer à leur activité.

### 1.3.2. Notation interne et probabilité de défaut

Coface protège ses assurés dans leurs créances commerciales. Il est important pour Coface d'avoir une idée de la situation financière des entreprises commerciales clientes de ses assurés à la fois pour déterminer le montant maximum que Coface peut garantir et pour évaluer la probabilité que les acheteurs ne remboursent pas leurs factures. Des renseignements sur l'état de santé d'une entreprise peuvent être obtenus grâce aux agences de notation telles que Standard & Poor's ou Moody's et Fitch. Cependant cette notation ayant un coût, la totalité des acheteurs du portefeuille Coface n'a pas systématiquement de note.

Coface a alors mis en place deux systèmes internes de notation : le CCR et le G-Score.

- le Coface Credit Rating CCR

Ce système de notation a été mis en place pour les risques les plus importants du portefeuille Coface. Il s'agit d'un système de notation qui s'appuie sur une analyse financière semi-automatisée et qui permet des corrections manuelles. Il est disponible pour plusieurs milliers d'acheteurs et de groupes. Il fournit une échelle de cent niveaux allant de 0 à 10 (des mauvais risques aux bons risques) avec une décimale. Des études menées par la direction des risques groupe ont permis de faire la correspondance entre le CCR et la probabilité de défaut à horizon un an. De plus, des classes homogènes en termes de probabilité de défaut ont été formées.

Le système de notation peut se présenter sous la forme du tableau suivant :

Classe	Min	Max	Probabilité de défaut
1	7.5	10.0	0,01%
2	7.0	7.4	0,1%
3	6.5	6.9	0,15%
4	6.0	6.4	0,24%
5	5.6	5.9	0,36%
6	5.2	5.5	0,85%
7	4.8	5.1	0,85%
8	4.4	4.7	1%
9	4.0	4.3	1,7%
10	3.6	3.9	2,4%
11	3.2	3.5	3,4%
12	2.5	3.1	4,9%
13	1.8	2.4	6,8%
14	1.1	1.7	10%
15	0.1	1	11,1%

La notation CCR ne peut pas être appliquée à l'ensemble du portefeuille Coface du fait du nombre important de risques et du niveau d'information insuffisant pour les entreprises de petite ou moyenne taille. Un système universel a donc été créé :

- le score générique G-Score

Il a été établi par une procédure statistique de scoring. Cette procédure se base sur des données du bilan des entreprises, dites **renseignements structurés** et sur un historique de sinistralité assurance crédit. Le système fournit une échelle de dix niveaux allant de 0 à 10 (des mauvais risques aux bons risques). De même que pour la notation CCR, une classe contient des entreprises homogènes en termes de probabilité de défaut à un an.

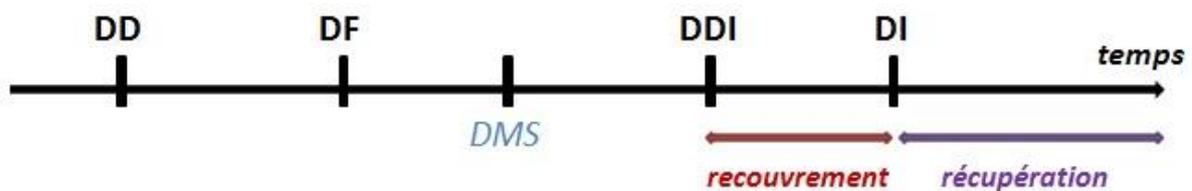
Le système de notation peut se présenter sous la forme du tableau suivant :

Classe	Probabilité de défaut
1	25%
2	10%
3	4%
4	2%
5	1,3%
6	0,7%
7	0,4%
8	0,15%
9	0,05%
10	0,01%

Dans le système G-Score, le défaut correspond à une date de déclaration de menace sinistre DMS. C'est pour cette raison que les probabilités de défaut sont plus élevées (une DMS n'étant pas systématiquement suivie d'une indemnisation).

### I.3.3. Chronologie d'un sinistre

Un sinistre peut être représenté par une suite de dates :



La **Date de Début** de facture (DD) et la **Date de Fin** de facture (DF) correspondent respectivement à la date d'émission et la date d'échéance d'une facture.

Lorsque l'assuré sait ou pense qu'une facture va donner lieu à un impayé, il signale cet éventuel impayé à Coface : nous avons une date **Déclaration de Menace de Sinistre** (DMS). Sur la chronologie ci-dessus la date DMS est postérieure à la DF mais il est possible qu'elle soit antérieure.

Après la DMS, nous n'aboutissons pas systématiquement à un sinistre : il peut par exemple s'agir d'un retard. La DDI représente la **Date de Demande d'Intervention** : l'assuré demande officiellement à Coface d'intervenir.

Coface procède alors au recouvrement des créances impayées. Le **recouvrement** correspond à une démarche amiable effectuée par Coface pour obtenir le paiement des créances impayées de l'assuré. Cette période dure en général entre 2 à 5 mois.

Après le recouvrement, Coface indemnise l'assuré à la **Date d'Indemnisation** (DI) : son intervention auprès de l'assuré est terminée.

Si pendant la période de recouvrement, Coface n'a pas réussi à obtenir le paiement total des créances impayées, elle continue cette démarche : c'est la phase de **recupération**. La récupération peut durer plusieurs années.

#### 1.3.4. Les modèles du risque de crédit

La modélisation permet de formaliser la sinistralité du risque de crédit. La sinistralité de Coface peut s'entendre comme le montant des pertes de Coface. Il est nécessaire de définir la notion de perte. S'il n'y a pas de défaut, la perte est nulle. Sinon la perte est définie à partir du montant en risque (l'exposition) au moment du défaut. A ce montant, nous pouvons appliquer la sévérité qui correspond à l'intensité du sinistre. Après application de la sévérité, nous obtenons la **perte sèche**. Enfin la **perte ultime** est obtenue après récupération.

- Schéma de fonctionnement du modèle de risque

Le modèle global de risque de crédit possède trois composantes :

- l'occurrence ou non du défaut avec une probabilité de défaut (**Probability of Default PD**)
- l'exposition au moment du défaut (**Exposure At Default EAD**)
- l'intensité du sinistre au moment du défaut (**Loss Given Default LGD**)

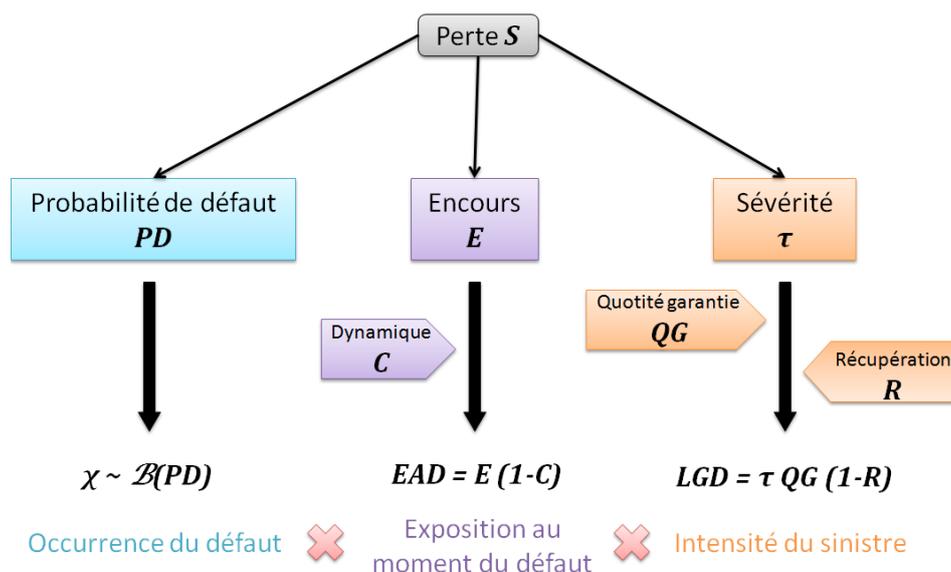


Schéma de fonctionnement du modèle de risque

Notons que le modèle a une composante binaire. Il est important de préciser la notion de défaut (défaillance légale, impayé ou versement d'indemnités).

Le schéma de fonctionnement ayant été présenté, il est nécessaire de décrire les variables qui entre dans sa composition.

- Variables du modèle

Concernant la probabilité de défaut  $PD$ , nous avons vu dans la partie I.3.2 qu'elle était associée à un rating ou un score.

Dans le modèle, nous nous intéressons au montant d'exposition au moment du défaut. La variable  $C$  permet de décrire les variations du **montant d'exposition**  $E$  entre la date de la signature du contrat et la date du défaut. Différents facteurs peuvent conduire à ces variations de  $E$  :

- un agrément possède une vie : naissance, mort, suspension, augmentation/diminution de son montant. Autrement dit, les montants garantis sont variables.
- le nombre des montants garantis pris en compte dans le montant de l'exposition varie également
- au niveau groupe, la composition d'un groupe peut évoluer par exemple à la suite de fusion ou d'acquisition

Il s'agit de la **dynamique du montant d'exposition**.

La **sévérité** en cas de sinistre  $\tau$  correspond à un taux de perte mesuré par rapport au montant d'exposition. Plus le sinistre est sévère, plus son coût sera élevé. Cette variable fait l'objet du mémoire. Nous en donnerons une définition plus précise dans le chapitre II.

La quotité garantie  $QG$ , décrite dans le paragraphe I.3.1 est la part prise en charge par Coface lors de l'indemnisation d'un sinistre. Cette  $QG$  se situe généralement autour de 80% – 90%.

Nous avons vu au paragraphe I.3.3. que la phase de récupération débutait après l'indemnisation. Le **taux de récupération  $R$**  correspond donc au pourcentage que Coface récupère après l'indemnisation. Comme cette phase peut durer plusieurs années, l'intensité du défaut ne sera parfois connue que très tardivement. Dans le modèle, il est supposé constant et égal à 10%.

- Classification des risques

Les risques du portefeuille Coface sont répartis dans deux catégories de risque : les **Grands Risques GR** et les **Petits Risques PR** dont nous allons faire la description ci-dessous.

Les grands risques sont caractérisés par :

- l'appartenance à un groupe
- un montant d'exposition de niveau groupe supérieur à 25M€
- le CCR comme système de notation

Il s'agit de risques avec une forte intensité potentielle de sinistre dans la mesure où le risque encouru est très important. Le montant d'exposition de cette catégorie est environ de 50G€ pour environ 1700 groupes.

Les risques n'appartenant pas à la première catégorie sont automatiquement classés dans la catégorie des petits risques. Ils sont caractérisés par :

- la non-appartenance à un groupe
- un montant d'exposition de niveau groupe inférieur à 25M€
- le G-Score comme système de notation

Il s'agit de risques à forte fréquence de sinistres.

Cette catégorie regroupe environ 1,4 millions d'acheteurs.

Nous dirons que les grands risques représentent le risque d'intensité et que les petits risques représentent le risque de fréquence. Ces deux catégories de risque ne peuvent donc pas être modélisées de la même façon au vu de leurs caractéristiques. Un modèle a été développé par Sesame pour chacune des catégories.

- Les modèles de risque

Les notations précédentes pour les différentes variables intervenant dans le fonctionnement du modèle sont reprises et nous désignons par  $S$  la perte ultime.

Le modèle du risque unitaire se formalise ainsi :

$$S = \chi E(1 - C) \tau QG(1 - R) \quad \text{Où} \quad \chi \sim \mathcal{B}(PD)$$

Le modèle est binaire du fait de la présence d'une composante binaire :

- avec probabilité  $PD$ , il y a un défaut
- avec probabilité  $1 - PD$ , il n'y a pas de défaut

Sesame a retenu le **modèle individuel** pour les grands risques. Dans ce cas, le risque unitaire désigne le risque sur un groupe. Une simulation de ce modèle fournit un montant de sinistre d'un seul groupe. Si nous désignons par  $N$  le nombre total de groupe, nous avons :

$$\forall i = 1, \dots, N \quad S_i = \chi_i E_i(1 - C) \tau_i QG(1 - R) \quad \text{Où} \quad \chi_i \sim \mathcal{B}(PD_i)$$

La dynamique du montant d'exposition, la quotité garantie et la récupération sont des variables communes à tous les groupes. Dans le modèle actuel, elles sont déterministes. Plus précisément elles sont constantes :  $C = 0$ ,  $QG = 90\%$  et  $R = 10\%$ .

Le montant d'exposition, la probabilité de défaut sont spécifiques à un groupe  $i$ . La probabilité de défaut est celle associée à la CCR.

Concernant la sévérité, Sesame a deux approches. La première approche consiste à la considérer comme déterministe et égale à 17% qui correspondent à la moyenne des sévérités groupes dont nous disposons un échantillon. La deuxième approche consiste à tirer une observation de l'échantillon pour chaque simulation. La modélisation de la sévérité de niveau groupe est un objectif que Sesame s'est donné. Elle permettrait d'obtenir des résultats plus précis.

La perte totale associée au portefeuille GR peut alors s'écrire :

$$S = \sum_{i=1}^N S_i = (1 - C) QG(1 - R) \sum_{i=1}^N \chi_i E_i \tau_i$$

Concernant les petits risques, le **modèle collectif** a été choisi du fait de la taille importante de cette catégorie de risque rendant impossible toute application du modèle individuel. Les risques sont segmentés par classes de risque. Chaque classe de risque est homogène du point de vue de G-Score/probabilité de défaut, de la tranche de montant d'exposition et du secteur d'activité. Nous avons :

- 10 classes de G-Score
- 5 tranches de montant d'exposition
- 7 classes de secteur d'activité

Autrement dit, nous avons segmenté les petits risques en  $K = 10 \times 5 \times 7 = 350$  classes de risque chacune caractérisée par un triplet unique (G-Score, tranche de montant d'exposition, secteur d'activité).

Pour chacune des classes de risques  $k$  ( $k = 1, \dots, K$ ), nous désignons par  $n_k$  le nombre de risque, par  $p_k$  la probabilité de défaut de la classe et par  $N_k$  le nombre d'occurrences de défauts. De plus, nous supposons que les défauts surviennent indépendamment au sein d'une même classe de risque. Par conséquent la variable  $N_k$  suit une loi binomiale de paramètres  $n_k$  et  $p_k$  :  $N_k \sim \mathcal{B}(n_k, p_k)$ .

Notons  $E_k$  le montant d'exposition moyen de la classe  $k$ . Nous supposons que les défauts proviennent de risques dont le montant d'exposition est égal à  $E_k$ . En effet, Sesame ne dispose pas d'étude concernant la répartition de l'exposition.

Sous les hypothèses précédentes, le montant des pertes pour la classe  $k$  s'écrit :

$$\forall k = 1, \dots, K \quad S(k) = N_k E_k \tau QG(1 - R)$$

Dans le modèle actuel et comme pour les grands risques la sévérité est soit déterministe et égale à 31% qui correspond à la moyenne d'un échantillon de valeurs de sévérité de niveau acheteur soit elle est tirée dans cet échantillon de valeurs de sévérités. **L'objectif du mémoire est dans un premier temps de produire un nouvel échantillon de valeurs de sévérité et dans un second temps d'améliorer cet aspect en modélisant la variable sévérité de niveau acheteur.**

La quotité garantie est supposée constante et égale 90%. Le taux de récupération est supposé constant et égal 10% qui est le taux moyen de récupération.

La perte totale associée au portefeuille PR peut alors s'écrire :

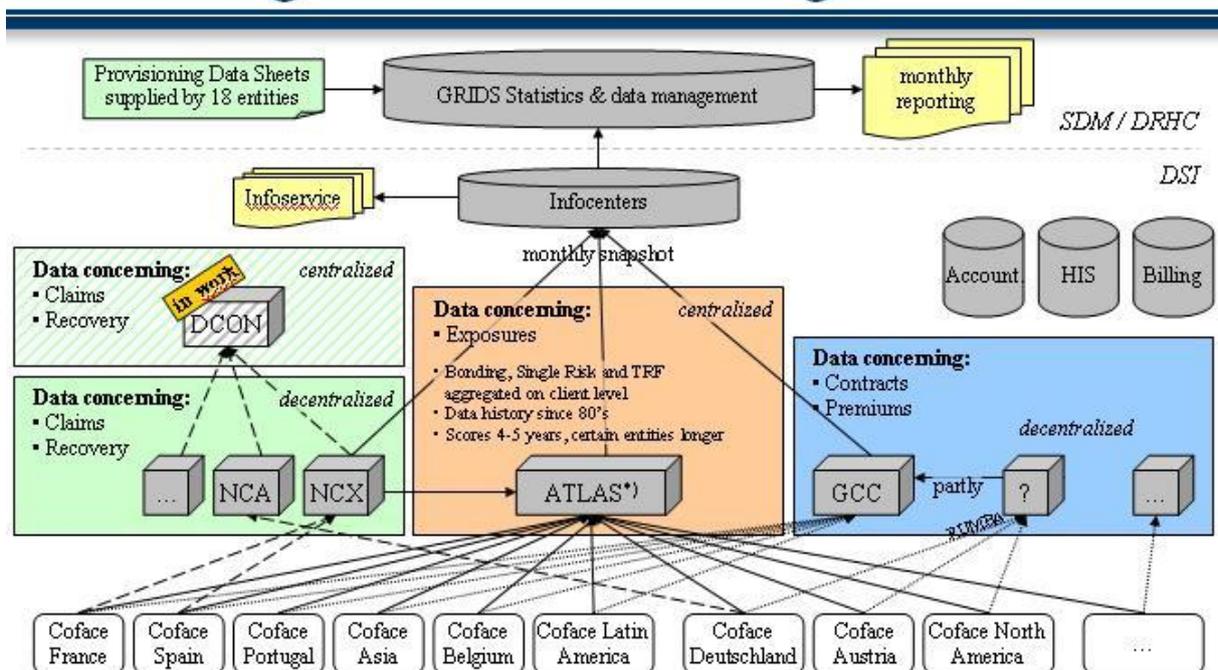
$$S = \sum_{k=1}^K S(k) = \tau Q G (1 - R) \sum_{k=1}^K N_k E_k$$

## Chapitre II Les données

Une base de données a été créée en fonction de nos besoins pour l'étude de la sévérité de niveau acheteur. La fiabilité des données est un facteur indispensable à l'obtention de résultats significatifs.

### II.1. Architecture des systèmes d'informations Coface

## Data Storage and Data Processing



Source : internal presentation of Sesame

Pour l'étude de la sévérité de niveau acheteur, nous avons besoin de données sur les expositions et de données sur les sinistres. Toutes les données dont nous disposons nous ont été fournies par la direction des risques qui gère la base de données ATLAS. Les sources de données sur les sinistres de la base ATLAS sont NCX et NCA. Nous disposons donc de façon exhaustive, de données sur les sinistres pour les entités Coface France, Coface Spain et Coface Deutschland.

## II.2. Présentation des données disponibles

Nous disposons d'une table de données pour les sinistres et d'une autre pour les expositions. Le croisement de ces deux tables permettra de mesurer la sévérité.

La première difficulté rencontrée est dans le niveau d'agrégation des données dans les tables. En effet, dans la table des sinistres, nous sommes de niveau agrément : une ligne correspond à un sinistre sur un agrément. Or dans la table des expositions, le niveau d'agrégation est le niveau acheteur : une ligne est le résultat d'une agrégation d'agréments d'un même acheteur.

Nous voulons effectuer notre étude au niveau acheteur c'est la raison pour laquelle des opérations d'agrégation devront être effectuées dans la table des sinistres.

La table des sinistres de niveau agrément nous renseigne entre autres sur le numéro de police, le type de police, le code pays de l'acheteur ainsi que son identifiant, la quotité garantie, la date de demande d'intervention, la date d'effet du sinistre, l'entité et sur différents montants tels que le montant de l'agrément, celui des factures garanties, le montant de recouvrement, celui de la franchise, celui des frais honoraires et celui versé à l'assuré.

La table des expositions donne un historique du montant garanti pour chaque acheteur. Elle contient les identifiants des acheteurs et des informations sur la situation géographique (pays, ville, rue). Elle contient également le code nace qui correspond à un code du secteur d'activité des acheteurs, leur montant garanti, la date d'enregistrement du montant garanti, leur nombre de polices, leur score et leur notation CCR.

Par ailleurs, nous disposons aussi de tables qui permettent de faire la correspondance entre le code nace et le libellé du secteur d'activité (groupe 15). Enfin, nous avons une table qui associe à un pays une note pays. La note pays créée par Coface indique le niveau de risque moyen présenté par les entreprises d'un pays dans le cadre de leurs transactions commerciales à court terme.

La sévérité étant définie de façon schématique comme un rapport entre le montant du sinistre sur le montant d'exposition, des valeurs de la sévérité au niveau acheteur se

déduisent des données sur le montant des sinistres dus à l'acheteur et sur le montant de l'exposition de Coface sur l'acheteur.

Le montant de sinistre sur un agrément, plusieurs facteurs doivent être pris en compte :

- le montant de recouvrement
- la franchise de la police à laquelle appartient l'agrément
- les frais honoraires accordés aux tiers pour leurs prestations dans le cadre du litige
- l'indemnisation versée à l'assuré
- le montant de récupération

Le montant de sinistre au niveau acheteur résulte d'une agrégation des montants de sinistre des agréments sur l'acheteur.

Le montant d'exposition ayant une dynamique, nous nous interrogeons sur le montant à choisir. Par exemple, lorsqu'il fait ou qu'il est fort probable que l'acheteur fasse défaut, Coface baisse le montant garanti. Coface peut également revoir à la hausse le montant garanti si le besoin en est exprimé par le client et si le risque demeure acceptable.

Il est nécessaire de formuler des hypothèses concernant ces montants, la manière dont nous agrégeons les montants de sinistres pour fournir une définition de la sévérité et pour en obtenir un échantillon. Ces hypothèses auront de manière certaine un impact sur nos résultats.

### II.3. Définition retenue de la sévérité

- Le montant d'exposition

Lorsqu'un acheteur a eu un sinistre ou lorsque Coface prédit qu'il va faire défaut, son montant garanti est diminué. L'exposition à la date de sinistre d'un acheteur ne serait pas représentative du réel engagement de Coface. Nous prendrons le montant d'exposition le plus élevé de l'année précédant celle de la date du sinistre.

- Le défaut

La notion de défaut diffère selon les entités. Concernant les entités AKC ou Coface Deutschland, le défaut correspond à la liquidation judiciaire. Le défaut pour les autres entités correspond à un impayé.

- Le sinistre

Un sinistre est provoqué par un défaut. Au niveau acheteur, la difficulté réside dans la distinction des sinistres qui concernent un même acheteur. Il est important de définir si deux agréments sinistrés sont relatifs à un même sinistre ou non. Il est par exemple impossible de considérer qu'un sinistre s'étale sur deux ans. Le critère retenu est le suivant : le terme **sinistre acheteur** désigne l'ensemble des agréments sinistrés sur le même acheteur d'une même année.

- La période des sinistres

Afin que la manipulation de la base de données soit possible, nous devons tenir compte de la taille des tables de données dans le choix de la période de sinistres. C'est pour cette raison que la période des sinistres débute le 01/01/2006. Par ailleurs, Coface dispose d'un délai pour procéder à l'indemnisation. Ce délai est de deux mois pour les petits sinistres et de cinq mois pour les petites créances et les classiques. La période des sinistres s'arrête au 31/12/2009 afin de n'avoir que des dossiers indemnisés. Cette période représente quatre ans d'historique. Nous sélectionnons donc les agréments dont la date de sinistre est comprise dans cette période. Il est donc nécessaire de déterminer ce qu'est la date de sinistre pour un agrément.

- La date de sinistre

Nous savons que plusieurs dates pouvaient représenter un sinistre : les dates de facture, la date de demande de menace de sinistre, la date de demande d'intervention ou la date d'indemnisation. De plus, la définition du défaut dépend de l'entité. La date d'effet du sinistre sera prise en compte pour les entités AKC et Coface Deutschland tandis que la date de demande d'intervention sera prise en compte pour les autres entités. Au niveau acheteur, la date associée à un sinistre acheteur est la plus ancienne des dates de sinistre

agrément. L'exposition étant choisie par rapport à cette date, il est certain que la date d'enregistrement de l'exposition sélectionnée sera antérieure à tous les sinistres agrément.

- Le montant du sinistre

Nous voulons modéliser la sinistralité dans le sens de la perte ultime. Le montant du sinistre correspond donc au montant des indemnités versées à l'assuré (elles sont donc nettes de recouvrement, de QG, de frais honoraires, de franchise et brutes de récupération). Le montant de sinistre pour un acheteur est obtenu en sommant les montants des sinistres agrément d'une même année.

- Le rating

Les données concernant le G-Score ou le CCR étant incomplètes, la combinaison des deux types de données permettrait d'obtenir un rating pour chaque acheteur. Dans un premier temps, nous convertissons les notes CCR en score à l'aide de la table de correspondance ci-dessous. Pour un acheteur, nous sélectionnons cette dernière note si elle est renseignée sinon nous sélectionnons son score s'il est renseigné (dans le cas contraire, l'acheteur n'aura pas de rating). La fiabilité de cette donnée peut alors être discutée.

CCR	Score
0.1	1
0.2 – 1.0	2
1.1 – 3.1	3
3.2 – 3.9	4
4.0 – 4.3	5
4.4 – 4.7	6
4.8 – 5.1	7
5.2 – 5.5	8
5.6 – 6.9	9
7.0 – 10	10

- La sévérité

Du fait de la franchise, des frais honoraires et du recouvrement, il est possible qu'un défaut ne soit pas indemnisé (ces défauts et le recouvrement font l'objet d'une autre étude). Du fait du caractère binaire du recouvrement, la sévérité sera définie en cas d'indemnisation. La sévérité niveau acheteur est donc le rapport du montant de sinistre sur l'exposition majoré par division par la quotité garantie. La sévérité est une variable comprise entre 0 et 1. Du fait de la dynamique de l'encours, il est possible d'obtenir des valeurs de sévérité supérieures à 1. Plus généralement, entre la date à laquelle nous sélectionnons le montant garanti et la date du sinistre, le montant garanti peut augmenter ce qui expliquerait un montant de sinistre plus élevé. Dans ce cas, nous avons rapproché un montant de sinistre à un montant d'exposition qui n'est plus représentatif du montant réellement sous risque. Une étude sur ces valeurs anormales a été effectuée (Voir la note *Etude des valeurs anormales de la sévérité – Niveau acheteur* en annexe).

## Chapitre III Analyse descriptive

A l'issue du croisement des sinistres avec les expositions et du traitement, nous obtenons une table de 45532 observations. Une observation est constituée de 12 valeurs. En effet il y a 12 variables qui sont :

variables qualitatives		variables quantitatives	
nominales	ordinales	discrètes	continues
Libelle pays	Rating	Nombre polices	Indemnites
Identifiant acheteur	Note Pays	Nombre polices sinistrees	Exposition
Libelle groupe 15			Severite
Date sinistre			
Année sinistre			

Dans l'optique de la modélisation de la variable Severite de niveau acheteur, il est important de la décrire par des statistiques et d'en étudier ses liaisons avec les autres variables présentes. Selon leur type, les variables ne seront pas traitées de la même manière. La variable Identifiant acheteur est un code, il n'y a pas de liaison. Concernant les variables, Date sinistre, Année sinistre, Nombre polices sinistrees et Indemnites, elles ne sont pas connues à priori et ne peuvent alors pas apparaître dans le modèle.

### III.1. Analyse descriptive

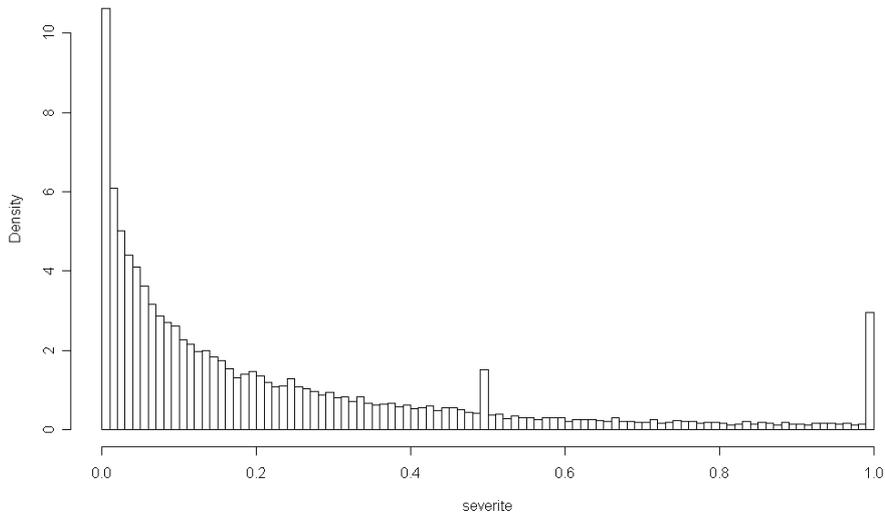
#### III.1.1. Caractéristiques de la variable Severite de niveau acheteur

La variable Severite est une variable quantitative continue.

- Caractéristiques graphiques

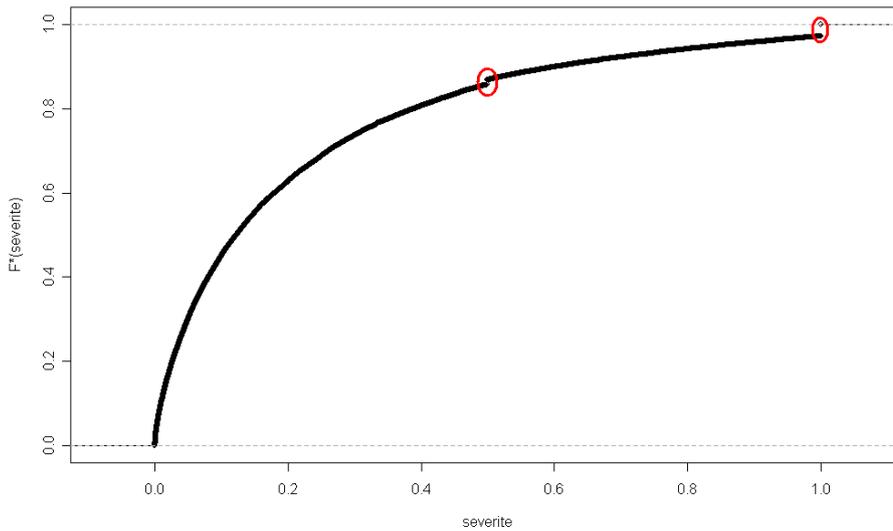
A l'aide du logiciel R, nous visualisons la distribution empirique de la sévérité :

distribution empirique de la severite



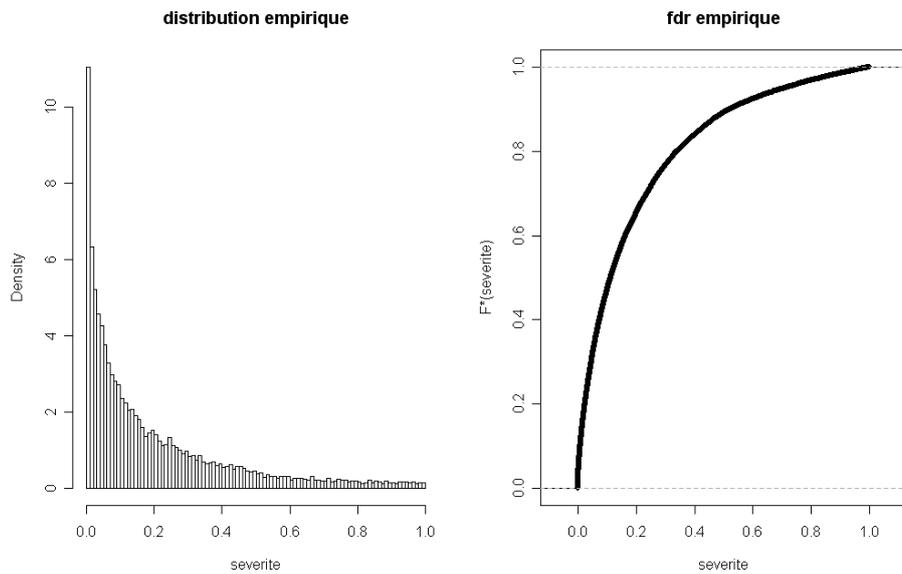
ainsi que sa fonction de répartition empirique :

fdr empirique de la severite



Nous remarquons que la distribution semble avoir deux « pics » : un vers la valeur 50% et un autre vers la valeur 100%. Etudions ces pics de plus près. Les observations de la sévérité égales à 50% sont au nombre de 491 (1,1% des observations) et celles égales à 100% sont au nombre de 1290 (2,8% des observations).

Si nous retirons ces observations de notre échantillon, nous obtenons graphiquement :



La distribution ne présente plus de saut, nous pouvons considérer que les sauts étaient aux valeurs précises de 50% et 100%. Une sévérité à 50% nous indique que la moitié du montant d'exposition a été « utilisé ». Une sévérité à 100% nous indique que le montant du sinistre a été au moins égal au montant d'exposition : nous avons alors un défaut total.

- Caractéristiques de position

min	0%
10è centile	1%
20è centile	3%
30è centile	5%
40è centile	8%
médiane	12%
60è centile	18%
70è centile	26%
80è centile	39%
90ècentile	60%
max	100%
VaR 95%	85%
VaR 99%	100%
moyenne	22%
ESF 95%	97%
ESF 99%	100%

Nous remarquons que 80% des acheteurs ont une sévérité inférieure à 40%.

Nous définissons deux mesures de la sévérité :

$$\text{moyenne des sévérités} = \frac{1}{n} \sum_{i=1}^n \frac{\text{indemnités}_i}{\text{exposition}_i} = \frac{1}{n} \sum_{i=1}^n \text{sévérité}_i$$

$$\text{sévérité moyenne} = \frac{\sum_{i=1}^n \text{indemnités}_i}{\sum_{i=1}^n \text{exposition}_i} = \frac{\sum_{i=1}^n \text{exposition}_i \times \text{sévérité}_i}{\sum_{i=1}^n \text{exposition}_i}$$

où  $n$  est le nombre d'acheteurs

Dans notre cas :

$$\text{sévérité moyenne} = 4\% < 22\% = \text{moyenne des sévérités}$$

Nous pouvons penser qu'il y a un nombre important d'acheteurs qui ont une sévérité élevée mais une exposition faible.

- Caractéristiques de dispersion

étendue	100%
intervalle interquartile	28%
écart type	26%
variance	7%

L'écart type est de 28%. Nous avons une forte dispersion autour de la moyenne.

### III.1.2. Dépendance de la sévérité avec les autres variables

L'objectif de cette partie est d'étudier les liaisons entre la sévérité et les autres variables. Nous allons calculer la moyenne des sévérités, l'écart type des sévérités et la sévérité moyenne en fonction des autres variables. Lors de cette étude de dépendance, nous avons trouvé que le pays de provenance de l'acheteur et son secteur d'activité ne sont pas des facteurs significatifs, les résultats se situent alors en annexe.

Le premier facteur que nous allons étudier est le montant d'exposition. Le tableau suivant montre quelques statistiques descriptives du montant d'exposition des acheteurs sinistrés.

min	1000
10è centile	10000
20e centile	15000
30e centile	20000
40e centile	32000
médiane	50000
60e centile	80000
70è centile	140000
80è centile	276000
90ècentile	740000
Max	776392000
moyenne	617789
ecart type	7453451

Nous pouvons affirmer que le montant d'exposition est très volatile autour de la moyenne. Sur la majorité des acheteurs, le montant d'exposition est faible ( $\leq 50K\text{€}$ ) et sur plus de 80% des acheteurs, le montant d'exposition est inférieur à la moyenne.

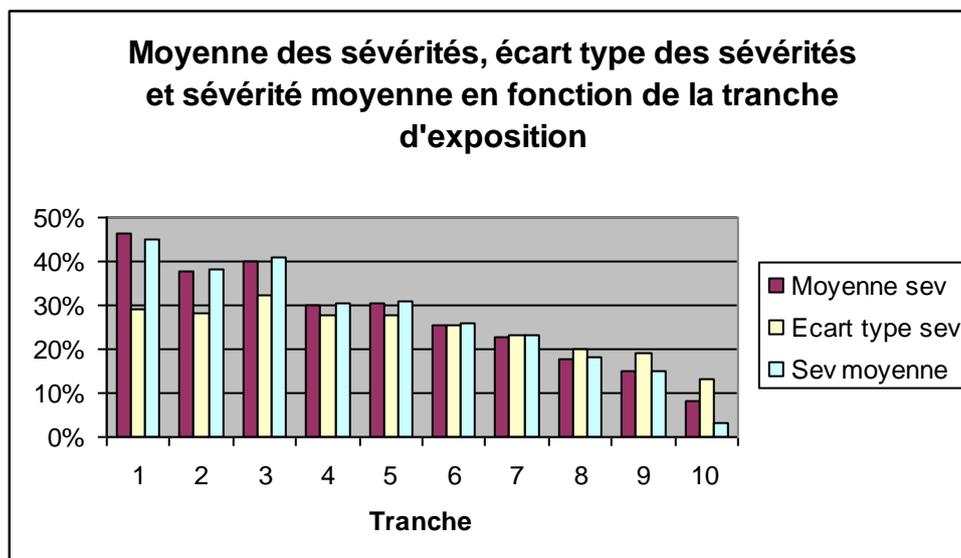
Nous obtenons un coefficient de corrélation de Pearson empirique de -6% entre la sévérité et l'exposition qui traduit une faible corrélation négative linéaire entre la sévérité et l'exposition.

Nous avons étudié la sévérité de niveau acheteur sur différentes tranches de montant d'exposition. Pour cela, nous avons pris en compte toutes les expositions de notre échantillon à savoir celles sur les acheteurs qui ont sinistré mais aussi celles sur les acheteurs qui n'ont pas sinistré. Nous avons ensuite calculé les quantiles afin de déterminer les bornes des dix tranches. Sur ces tranches, la moyenne des sévérités, l'écart type des sévérités et la sévérité moyenne ont été calculés.

Nous obtenons le tableau suivant :

tranche	borne incluse	inf borne exclue	borne sup	Frequence	Pourcentage	Moyenne sev	Ecart type sev	Sev moyenne	
1	0	7000		1727	4%	46%	63%	45%	min
2	7000	10000		933	2%	38%	81%	38%	max
3	10000	11000		4667	10%	40%	35%	41%	
4	11000	19000		2980	7%	30%	37%	31%	
5	19000	22000		4236	9%	30%	28%	31%	
6	22000	35000		3971	9%	25%	25%	26%	
7	35000	55000		5225	11%	23%	19%	23%	
8	55000	100000		4975	11%	18%	17%	18%	
9	100000	262000		7453	16%	15%	11%	15%	
10	262000			9365	21%	8%	7%	3%	
				45532	100%	22%	26%	4%	

Graphiquement :

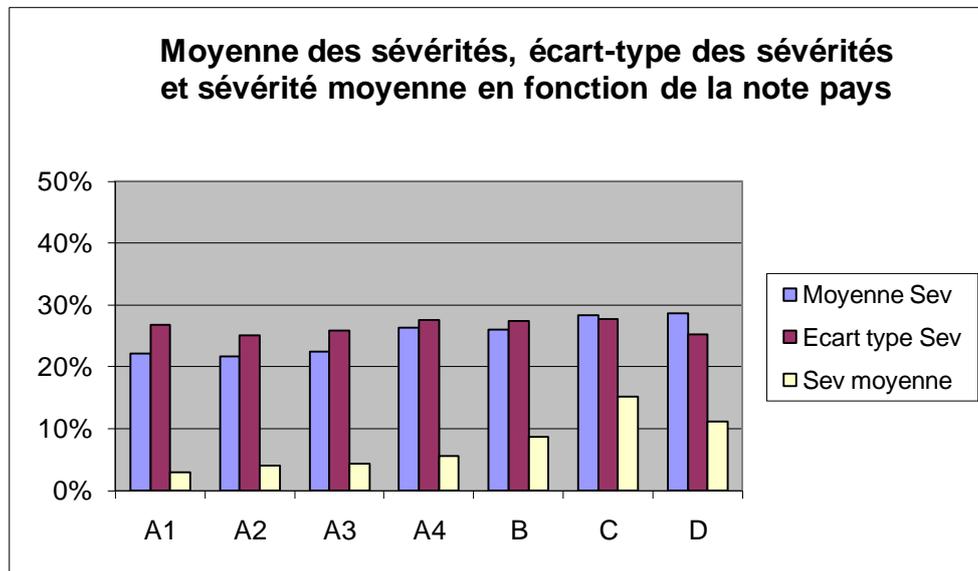


Nous pouvons remarquer qu'en établissant des tranches d'exposition, nous réconcilions la sévérité moyenne et la moyenne des sévérités et qu'elles sont décroissantes du montant d'exposition. Ce comportement est le phénomène de dilution : pour un sinistre de même coût, la sévérité sera plus faible lorsque le montant d'exposition est élevé.

Le tableau et le graphique suivants nous indique le comportement de la sévérité en fonction de la note pays. La note pays comporte 6 modalités.

		Données				
Note Pays	Frequence	Pourcentage	Moyenne Sev	Ecart type Sev	Sev moyenne	
A1	660	1%	22%	27%	3%	min
A2	23561	52%	22%	25%	4%	max
A3	18308	40%	22%	26%	4%	
A4	1474	3%	26%	28%	6%	
B	873	2%	26%	27%	9%	
C	519	1%	28%	28%	15%	
D	137	0%	29%	25%	11%	
Total	45532	100%	22%	26%	4%	

Nous observons que plus de 90% des acheteurs sinistrés proviennent d'un pays noté A2 ou A3.

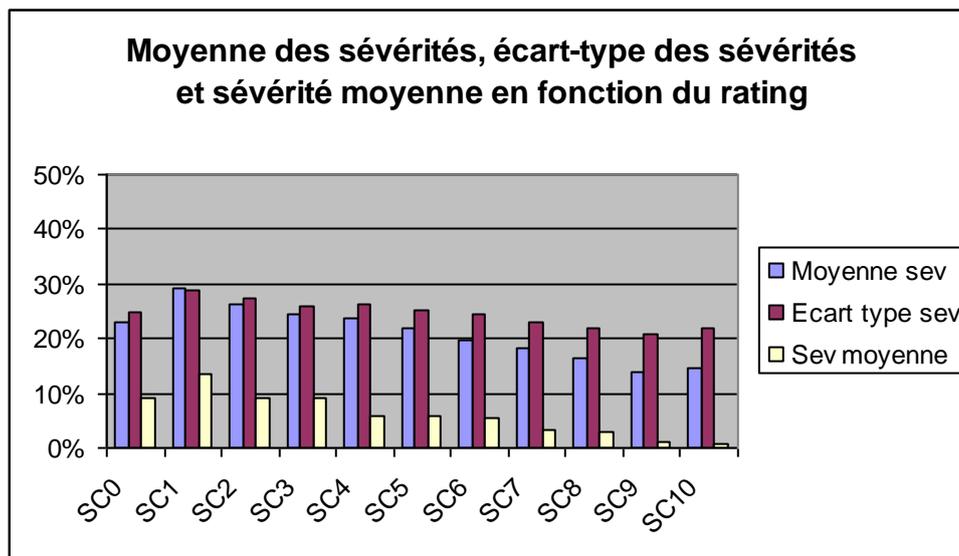


Nous pouvons observer une légère augmentation de la moyenne des sévérités et de la sévérité moyenne lorsque la note pays se dégrade mais ces variations sont faibles comparées à celles en fonction de la tranche d'exposition.

Nous procédons de même pour le rating. Nous avons 11 modalités pour le rating. 7646 valeurs sont manquantes soit 17%.

Nous obtenons :

Rating	Données					
	Frequence	Pourcentage	Moyenne sev	Ecart type sev	Sev moyenne	
SC0	152	0%	23%	25%	9%	min
SC1	948	3%	29%	29%	14%	max
SC2	1 692	4%	26%	27%	9%	
SC3	4 162	11%	24%	26%	9%	
SC4	7 865	21%	24%	26%	6%	
SC5	8 272	22%	22%	25%	6%	
SC6	6 639	18%	20%	24%	5%	
SC7	4 268	11%	18%	23%	3%	
SC8	2 229	6%	16%	22%	3%	
SC9	1 302	3%	14%	21%	1%	
SC10	357	1%	15%	22%	1%	
Total	37 886	100%	21%	25%	4%	



Nous pouvons remarquer que les trois mesures sont décroissantes quand la note s'améliore mais les variations sont faibles.

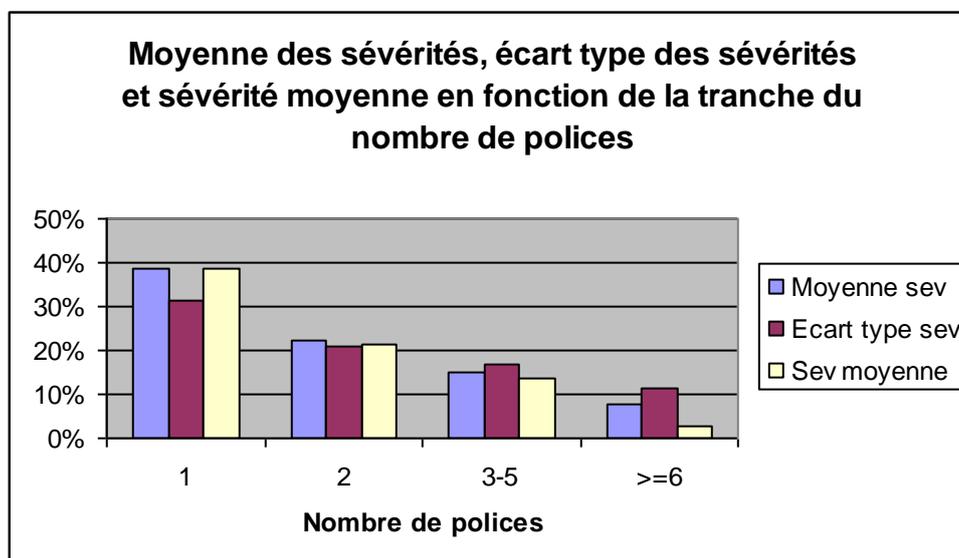
Le dernier facteur étudié sera le nombre de polices exposées. La variable Nombre de polices est discrète et possède un grand nombre de modalités.

Au total 234234 polices relatives aux acheteurs sinistrés ont été identifiées pour le calcul de la sévérité (en moyenne 5 polices par acheteur). Mais l'écart type de la variable Nombre de polices est de 11 donc il y a une forte disparité.

Nous obtenons un coefficient de corrélation de Pearson de -21% entre le nombre de polices et la sévérité ce qui n'est pas négligeable.

Du fait du nombre important de modalités, il est préférable d'effectuer des regroupements afin que l'analyse ou le graphique soit lisible. Nous avons formé quatre tranches de nombre de polices afin d'avoir des groupes d'effectifs similaires.

		Données					
Nbpol2	Nbpol	Frequence	Pourcentage	Moyenne sev	Ecart type sev	Sev moyenne	
1	1	14735	32%	39%	32%	39%	min
2	2	9598	21%	22%	21%	21%	max
3-5		10266	23%	15%	17%	13%	
>=6		10933	24%	8%	11%	3%	
Total		45532	100%	22%	26%	4%	



La moyenne des sévérités et la sévérité moyenne diminuent quand le nombre de polices augmente. La sévérité est moins volatile quand le nombre de polices augmente. Nous pouvons affirmer que la liaison entre le nombre de polices et la sévérité est particulièrement significative.

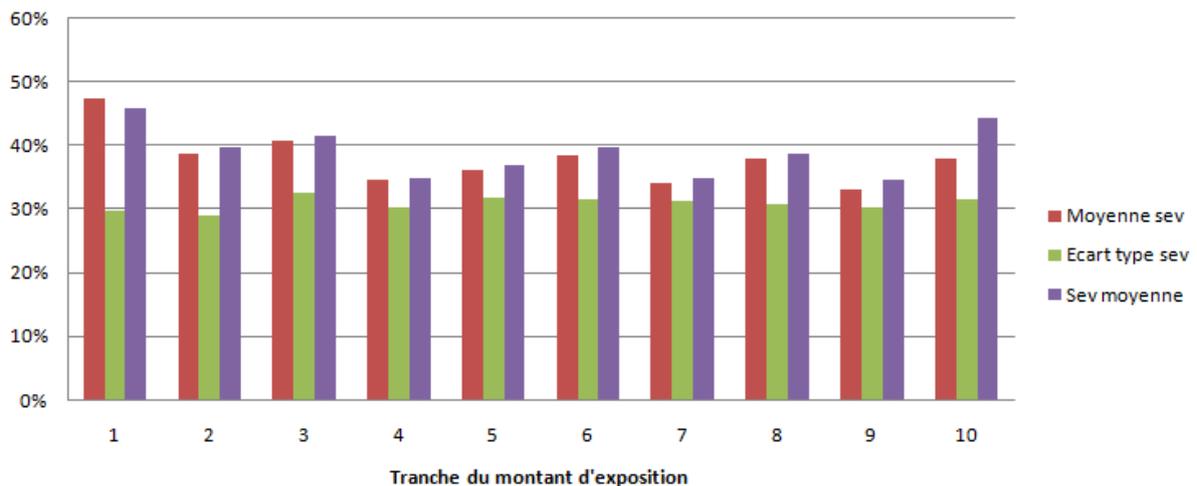
A l'issue de ces analyses, il nous apparaît que l'exposition et le nombre de polices sont des facteurs significatifs pour la sévérité. Nous pouvons penser que ces facteurs sont liés. Etudions l'effet de l'interaction nombre de polices – exposition sur la sévérité.

Quand le nombre de polices est égal à un :

Nombre polices	1
----------------	---

TrExpo	Données					
	Frequence	Pourcentage	Moyenne sev	Ecart type sev	Sev moyenne	
1	1559	11%	47%	30%	46%	min
2	817	6%	39%	29%	40%	max
3	4473	30%	41%	33%	41%	
4	1859	13%	34%	30%	35%	
5	2204	15%	36%	32%	37%	
6	1218	8%	38%	32%	40%	
7	1254	9%	34%	31%	35%	
8	450	3%	38%	31%	39%	
9	744	5%	33%	30%	35%	
10	157	1%	38%	31%	44%	
Total général	14735	100%	39%	32%	39%	

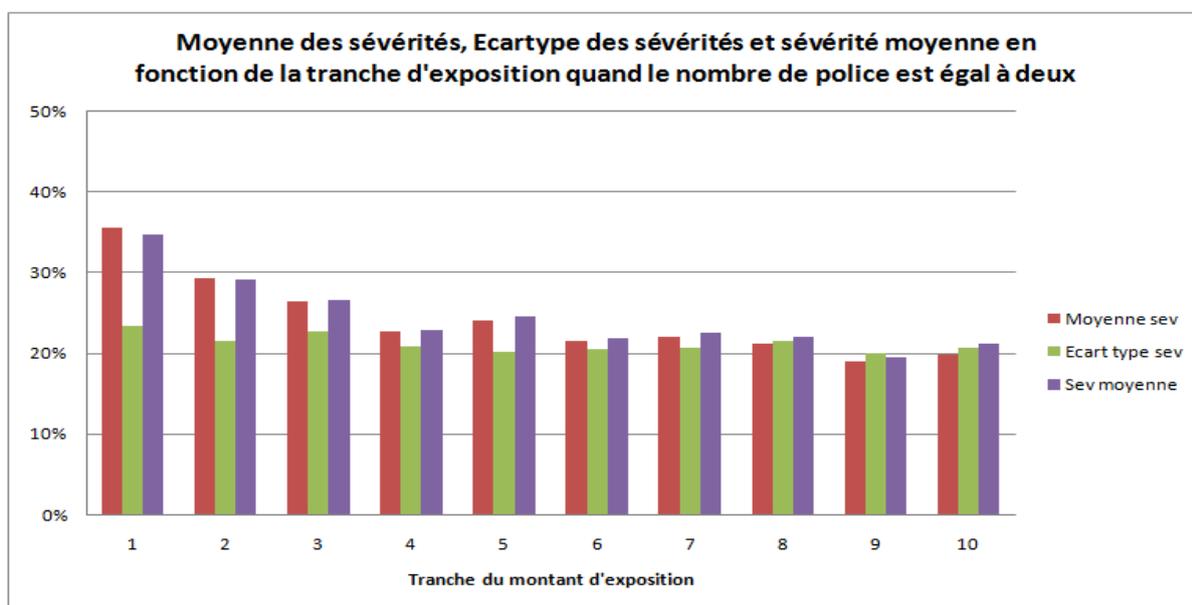
**Moyenne des sévérités, Ecartype des sévérités et sévérité moyenne en fonction de la tranche d'exposition quand le nombre de police est égal à un**



Nous pouvons affirmer que les grandeurs sont stables. L'effet du montant d'exposition disparaît quand nous raisonnons par nombre de polices. Vérifions ce résultat quand le nombre de polices est égal deux.

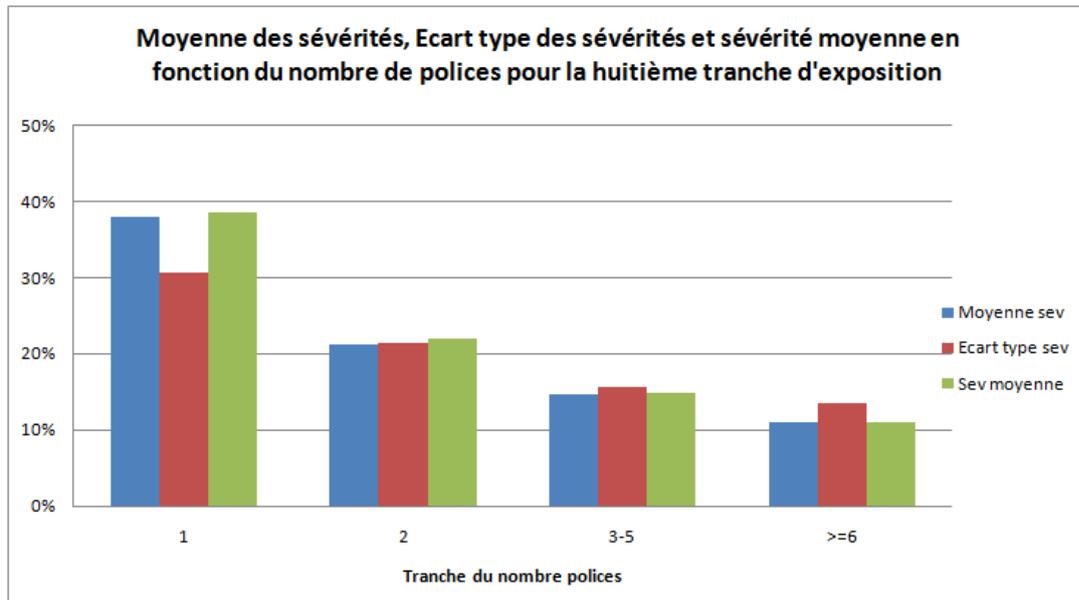
Nombre polices	2
----------------	---

TrExpo	Données					min
	Frequence	Pourcentage	Moyenne sev	Ecart type sev	Sev moyenne	
1	153	2%	36%	23%	35%	min
2	93	1%	29%	21%	29%	max
3	170	2%	26%	23%	27%	
4	947	10%	23%	21%	23%	
5	1791	19%	24%	20%	25%	
6	1612	17%	21%	21%	22%	
7	1979	21%	22%	21%	23%	
8	1184	12%	21%	21%	22%	
9	1329	14%	19%	20%	20%	
10	340	4%	20%	21%	21%	
Total général	9598	100%	22%	21%	21%	



Ces chiffres confirment la première conclusion.

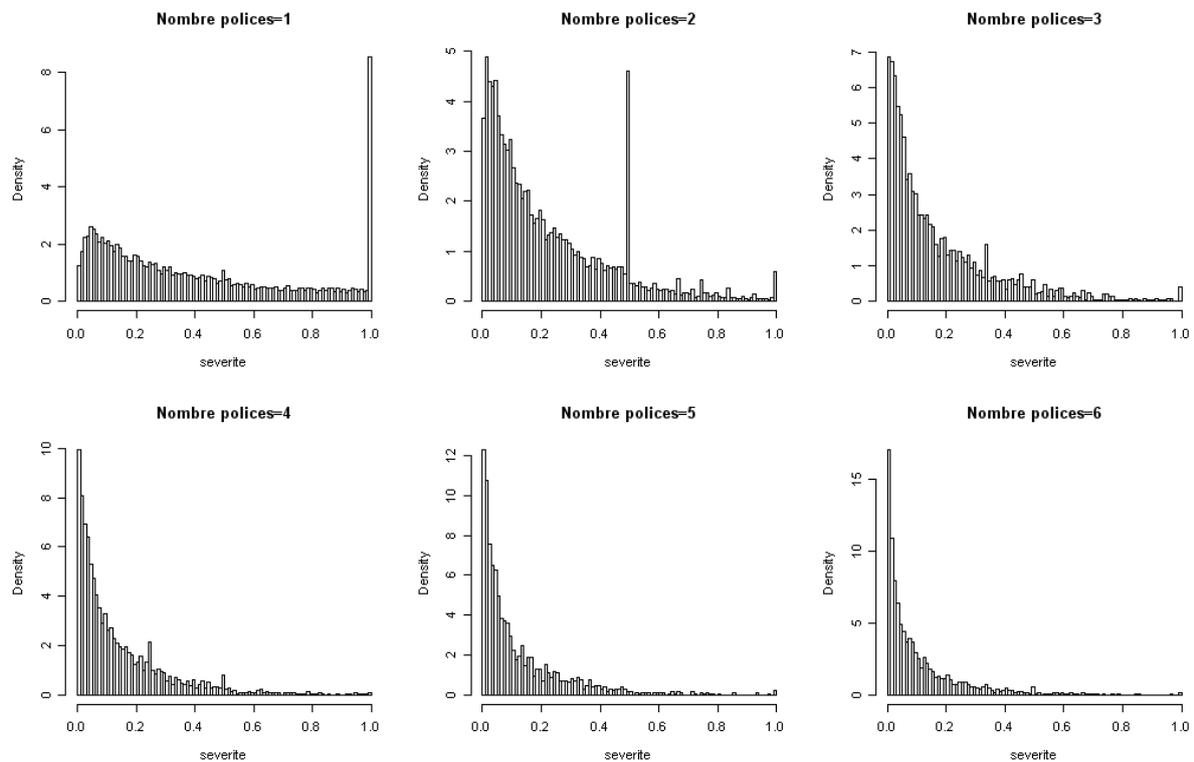
A l'inverse, pour une tranche de police, étudions l'effet du nombre de polices. Par exemple, pour la huitième tranche, nous obtenons :



Lorsque nous étudions l'interaction nombre de polices – exposition, nous n'observons plus de décroissance en fonction du montant d'exposition. De plus, il y a une forte corrélation positive (54%) entre l'exposition et le nombre de polices. L'effet de décroissance en fonction du montant d'exposition provient de celui en fonction du nombre de polices.

La variable nombre de polices exposées est donc la variable la plus discriminante. Nous aimerions modéliser la sévérité en fonction du nombre de polices.

En traçant la distribution empirique de la sévérité de niveau acheteur en fonction du nombre de polices, nous obtenons :



Nous pouvons observer différents phénomènes :

- la probabilité d’avoir un défaut total est particulièrement forte lorsque le nombre de polices est égal à un (pic à 100%)
- quand le nombre de polices est égal à deux, nous avons en plus un pic à 50%  
quand il est égal à trois, il y en a un à 33%  
quand il est égal à quatre, il y en a un à 25%.

Intéressons nous au pic à 100%. La valeur 100% est atteinte par la sévérité lorsque le montant d’indemnisation est égal au montant d’exposition. Cela implique que toutes les polices ont été touchées par un sinistre. D’après le premier graphique, nous pouvons affirmer que la probabilité d’avoir un défaut total sur un agrément est forte (8,2%). La probabilité d’avoir un défaut total est inférieure à 1% quand le nombre de polices est différent de un.

Une sévérité égale à 50% est possible lorsque le montant d’indemnisation représente la moitié du montant d’exposition.

Lorsque nous observons ces valeurs exactement égales à 50% et lorsque le nombre de polices est égal à deux, nous pouvons formuler deux remarques :

1. dans 100% des cas, seulement un des deux agréments est touché
2. dans 99%, les montants garantis sont égaux

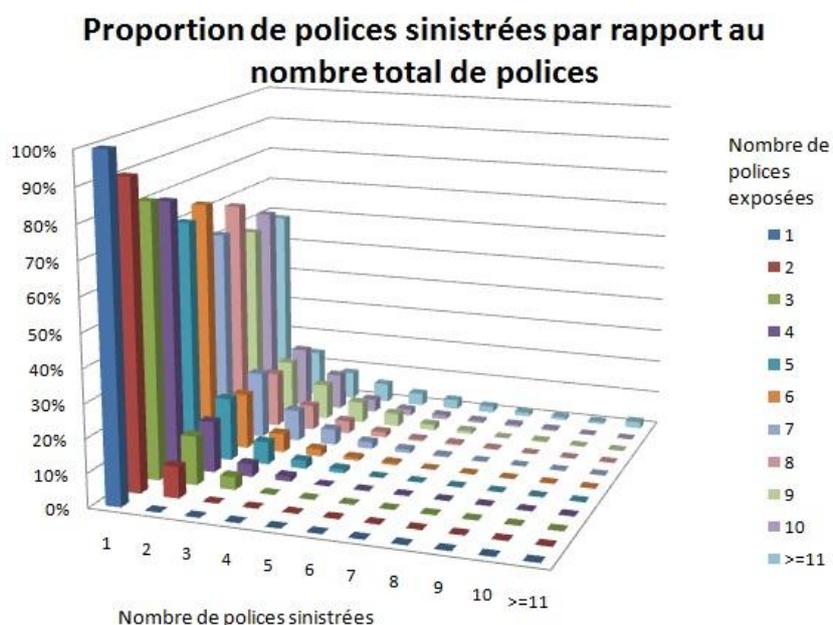
Autrement dit, nous pouvons dire que le pic à 50% est expliqué par deux événements simultanés : « un seul agrément touché » et « montants garantis égaux ».

Nous pouvons nous demander si nous observons les mêmes événements quand le nombre de polices est égal à trois ou quatre. Nous avons :

Evénement Observation	« un agrément sur deux touché »	« montants garantis égaux »	les deux événements
pic 50% (Nbpol=2)	100%	99%	99%
pic 33% (Nbpol=3)	100%	97%	97%
pic 25% (Nbpol=4)	98%	82%	80%

Quand le nombre de polices est égal à deux, trois ou quatre, les différents pics sont en grande partie expliqués par ces deux événements. Nous allons donc nous intéresser à ces événements.

Nous allons étudier la relation entre le nombre total de polices sur un même acheteur et le nombre de polices sinistrées sur l'acheteur.



La probabilité de n'avoir qu'une seule police sinistrée est particulièrement forte quelque soit le nombre de polices exposées : elle est au moins de 50%. Celle d'avoir deux polices sinistrées est environ de 10%. Nous avons ici le phénomène de dilution en nombre de polices : si un acheteur a un montant d'exposition important provenant de plusieurs agréments et s'il n'est sinistré que sur une seule police alors sa sévérité sera plus faible qu'un acheteur qui n'a qu'un seul agrément sur lequel il a fait sinistre.

Concernant les montants garantis, le tableau suivant nous indique le pourcentage d'acheteurs qui ont tous leurs montants garantis de leurs polices exposées égaux.

<b>Nombre de polices exposées</b>	<b>Pourcentage d'acheteurs avec des montants garantis égaux</b>
<b>2</b>	50%
<b>3</b>	12%
<b>4</b>	12%
<b>5</b>	4%
<b>6</b>	5%
<b>7</b>	3%
<b>8</b>	3%
<b>9</b>	2%
<b>10</b>	2%
<b>&gt;=11</b>	0%
<b>Total</b>	46%

Nous constatons que la probabilité d'avoir des montants garantis diminue quand le nombre d'agréments augmente. Cela pourrait expliquer l'absence de pic dans la distribution de la sévérité empirique quand le nombre de polices est supérieur à quatre.

### III.2. Modèle ciblé

L'analyse descriptive nous permet d'affirmer que la variable Nombre de polices est particulièrement discriminante.

Par l'analyse descriptive, nous avons pu identifier différents phénomènes :

- la probabilité d'avoir un défaut total sur un agrément est forte
- lorsque le nombre de polices est inférieur à quatre, nous observons un nombre non négligeable d'acheteurs qui ont des montants garantis égaux
- la probabilité de n'avoir qu'une seule police sinistrée est très forte  
la proportion de polices sinistrées diminue en fonction du nombre de polices sinistrées  
⇒ il s'agit du phénomène de dilution en nombre de polices

Ces trois phénomènes expliquent en grande partie la présence des différents pics étudiés précédemment lorsque le nombre de polices exposées est inférieur à quatre.

A l'issue de ces observations, nous optons pour une modélisation de la sévérité en fonction du nombre de polices. Plus précisément, à partir de la distribution de la sévérité sur un unique agrément (donc quand le nombre de polices est égal à un), nous allons essayer de modéliser la distribution de la sévérité quand le nombre de polices est supérieur à un. Ainsi nous voulons comprendre la logique d'agrégation.

## Chapitre IV Ajustement et simulation de lois

Ce chapitre est un chapitre théorique dans lequel nous allons décrire les lois utilisées pour la modélisation de la sévérité de niveau acheteur avant de décrire les méthodes d'ajustement et de simulation de ces lois.

### IV.1. Lois utilisées

Nous allons brièvement décrire les lois utilisées, la justification de leur utilisation sera donnée dans le chapitre suivant qui correspond à l'application à notre sévérité.

- Loi gamma

La variable aléatoire  $X$  suit une loi gamma de paramètres  $\alpha > 0$  et  $\beta > 0$ , ce que nous notons  $X \sim \text{Gamma}(\alpha, \beta)$ , si sa densité est donnée par :

$$\forall x \geq 0, \quad f_X(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{x}{\beta}} x^{\alpha-1}$$

où  $\Gamma$  est la fonction gamma définie par :

$$\forall z > 0, \quad \Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt$$

Ses moments sont :

$$E(X) = \alpha\beta$$

$$V(X) = \alpha\beta^2$$

- Loi Weibull

La variable aléatoire  $X$  suit une loi de Weibull de paramètres  $\beta > 0$  et  $\theta > 0$ , ce que nous notons  $X \sim W(\beta, \theta)$ , si sa densité est donnée par :

$$\forall x \geq 0, \quad f_X(x) = \frac{\beta}{\theta} \left(\frac{x}{\theta}\right)^{\beta-1} e^{-\left(\frac{x}{\theta}\right)^\beta}$$

où  $\beta$  est le paramètre de forme  
 $\theta$  est le paramètre d'échelle

Ses moments sont :

$$E(X) = \theta \Gamma\left(1 + \frac{1}{\beta}\right)$$

$$V(X) = \theta^2 \left[ \Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right) \right]$$

- Loi exponentielle

La variable aléatoire  $X$  suit une loi exponentielle d'intensité  $\lambda > 0$ , ce que nous notons  $X \sim \mathcal{E}(\lambda)$ , si sa densité est donnée par :

$$\forall x \geq 0, \quad f_X(x) = \lambda e^{-\lambda x}$$

Ses moments sont :

$$E(X) = \frac{1}{\lambda}$$

$$V(X) = \frac{1}{\lambda^2}$$

- Loi lognormale

La variable aléatoire  $X$  suit une loi lognormale de paramètres  $\mu \in \mathbb{R}$  et  $\sigma > 0$ , ce que nous notons  $X \sim LN(\mu, \sigma)$ , si sa densité est donnée par :

$$\forall x \geq 0, \quad f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}$$

Ses moments sont :

$$E(X) = e^{\mu + \frac{\sigma^2}{2}}$$

$$V(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$$

- Loi bêta

La variable aléatoire  $X$  suit une loi bêta de paramètres  $\mu \in \mathbb{R}$  et  $\sigma > 0$ , ce que nous notons  $X \sim LN(\mu, \sigma)$ , si sa densité est donnée par :

$$\forall x \geq 0, \quad f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}$$

Ses moments sont :

$$E(X) = e^{\mu+\frac{\sigma^2}{2}}$$

$$V(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$$

- Loi tronquée

Soit  $X$  une variable aléatoire à support réel. On note  $f_X$  sa fonction de densité et  $F_X$  sa fonction de répartition. Il est possible d'étudier la distribution de la variable aléatoire  $X$  sur un intervalle  $]a, b[$ . Nous définissons la loi de probabilité tronquée de fonction de densité :

$$f_{X|a < X < b}(x) = \frac{f_X(x)}{\int_a^b f_X(x) dx} \mathbb{I}_{]a, b[}(x)$$

Sa fonction de répartition est

$$F_{X|a < X < b}(t) = \frac{\int_a^t f_X(x) dx}{\int_a^b f_X(x) dx}$$

- Mélange de lois

Un mélange de lois est une combinaison convexe de plusieurs lois. Nous avons :

$$X \sim P_X = \sum_{j=1}^J \alpha_j P_j$$

avec :

$$J \in \mathbb{N}^*$$

$$(\alpha_1, \dots, \alpha_j, \dots, \alpha_J) \in [0,1]^J$$

$$\sum_{j=1}^J \alpha_j = 1$$

## IV.2. Procédés d'ajustement

Soit  $\underline{x} = (x_1, x_2, \dots, x_i, \dots, x_n)$  une réalisation d'un  $n$ -échantillon de variables aléatoires indépendantes de fonction de densité  $f$  de paramètre  $\Theta = (\theta_1, \theta_2, \dots, \theta_l, \dots, \theta_L)$ . Les procédés ci-dessous ont pour objectif de trouver les valeurs de  $\Theta$  telles que la densité  $f(\cdot; \Theta)$  s'ajuste au mieux des données.  $f$  est connue et  $\Theta$  est inconnu.

### IV.2.1. Procédés classiques

- Estimation des paramètres par la méthode des moments

Cette méthode nécessite de calculer des moments empiriques de l'échantillon. Le moment empirique d'ordre  $l$   $m_l$  s'écrit :

$$m_l = \frac{1}{n} \sum_{i=1}^n x_i^l$$

Le moment d'ordre  $l$  de la distribution est  $E(X^l)$ . En général les moments théoriques s'écrivent en fonction des paramètres.

La méthode des moments consiste à trouver les valeurs de  $\Theta$  pour lesquelles nous avons égalité entre les moments empiriques et les moments théoriques de la distribution. Pour  $L$  paramètres à déterminer, nous prenons en général les  $L$  premiers moments. Il est donc nécessaire que les  $L$  premiers moments existent.

Nous cherchons  $\Theta = (\theta_1, \theta_2, \dots, \theta_l, \dots, \theta_L)$  tel que :

$$\begin{cases} m_1 = E(X) \\ m_2 = E(X^2) \\ \vdots \\ m_l = E(X^l) \\ \vdots \\ m_L = E(X^L) \end{cases}$$

Le même raisonnement peut être fait avec les moments centrés.

La résolution du système peut être difficile lorsque le nombre de paramètres à estimer est grand (mélange de lois) ou lorsque les moments théoriques n'ont pas de forme analytique (loi Weibull).

- Estimation du maximum de vraisemblance (version continue)

La fonction de vraisemblance correspond au produit de la fonction de densité aux points  $x_i$ . Elle s'écrit :

$$L_n(\underline{x}; \Theta) = \prod_{i=1}^n f(x_i; \Theta)$$

L'estimateur du maximum de vraisemblance de  $\Theta$  est la valeur de  $\Theta$  qui maximise la vraisemblance. En effet nous cherchons les paramètres qui font du fait d'observer ces valeurs  $x_1, \dots, x_n$  le plus vraisemblable possible.

En pratique, nous cherchons à maximiser la fonction de log-vraisemblance qui correspond au logarithme de la fonction de la vraisemblance :

$$l_n(\underline{x}; \Theta) = \log[L_n(\underline{x}; \Theta)] = \sum_{i=1}^n \log[f(x_i; \Theta)]$$

L'estimateur est :

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} l_n(\underline{x}; \Theta)$$

Parfois il est impossible de résoudre le problème analytiquement, nous pouvons alors procéder numériquement.

## IV.2.2. Procédé « intuitif »

Sans ses méthodes, intuitivement, pour ajuster une loi aux données, nous pouvons chercher les paramètres de la loi qui permettent de minimiser la somme quadratique des écarts entre la distribution empirique et la distribution théorique. Nous décrivons cette méthode induite de la pratique.

Soit

$$x_{min} = \min(x_1, x_2, \dots, x_i, \dots, x_n)$$

$$x_{max} = \max(x_1, x_2, \dots, x_i, \dots, x_n)$$

$$x_{min} = t_0 < t_1 < \dots < t_k < \dots < t_K = x_{max}, \text{ une subdivision de } [x_{min}, x_{max}].$$

Nous calculons la probabilité empirique et la probabilité théorique d'appartenir sur chacun des intervalles  $[t_k, t_{k+1}[$   $k \in \{0, 1, \dots, K-1\}$  :

$$P_k^E = P^E(X \in [t_k, t_{k+1}[) = \frac{\text{Card}(\{x_i | t_k \leq x_i < t_{k+1}\})}{n}$$

$$P_k^T = P^T(X \in [t_k, t_{k+1}[) = F_X(t_{k+1}) - F_X(t_k)$$

La fonction à minimiser est :

$$S(\theta) = \sum_{k=0}^{K-1} (P_k^E - P_k^T)^2$$

L'estimateur est :

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} S(\theta)$$

Nous nous sommes demandé si par ce raisonnement « par tranches », nous pouvions retrouver la méthode du maximum de vraisemblance.

Notons  $dt_k = t_{k+1} - t_k$ .

Si  $dt_0 = dt_1 = \dots = dt_k = \dots = dt_K = dt$  sont infiniment petits et si  $t_k \leq x_i < t_{k+1}$

alors  $P^T(X \in [t_k, t_{k+1}[) \approx f(x_i; \theta)dt$

La fonction à maximiser est :

$$S'(\theta) = \sum_{k=0}^{K-1} nP_k^E \log(P_k^T)$$

Nous retrouvons alors la méthode du maximum de vraisemblance.

### IV.2.3. Adéquation à une loi

Nous nous baserons sur des méthodes d'adéquation qui compare la fonction de répartition de l'échantillon  $F$  avec une fonction de répartition entièrement spécifiée  $G$ .

Le lemme de Glivenko-Cantelli montre une forte proximité asymptotique entre la fonction de répartition de l'échantillon  $F$  et sa fonction de répartition empirique  $F^*$ .

$$\sup |F^*(x) - F(x)| \xrightarrow[n \rightarrow +\infty]{} 0 \text{ ps}$$

La fonction de répartition  $F$  de l'échantillon étant inconnue, nous la substituons à la fonction de répartition empirique  $F^*$ .

Soit  $(x_1, \dots, x_i, \dots, x_n)$  un  $n$ -échantillon d'une variable aléatoire  $X$  positive à support réel.

Nous avons :

$$\forall t > 0, \quad F^*(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{x_i \leq t}$$

- Probability – Probability plot PP-plot

Il s'agit d'un diagramme qui permet une appréciation graphique de la concordance entre la fonction de répartition observée et une fonction de répartition théorique. L'axe des ordonnées porte les probabilités cumulées observées et celui des abscisses porte les probabilités cumulées théorique. Le nuage de points  $[G(x_i), F^*(x_i)]$  s'aligne sur la première bissectrice si la loi théorique s'ajuste bien aux données. Cette méthode peut également s'appliquer avec les quantiles. Nous pouvons affirmer que l'appréciation de l'alignement des points le long de la première bissectrice est subjective.

- Test de Kolmogorov

Nous voulons tester :  $\mathcal{H}_0: F^* = G$  contre  $\mathcal{H}_1: F^* \neq G$  pour un seuil  $\alpha > 0$ .

La statistique du test est :

$$K = \sup_x |G(x) - F^*(x)|$$

Méthodologie :

0. Ordonner les observations  $(x_1, \dots, x_i, \dots, x_n) : (x_{[1]}, \dots, x_{[i]}, \dots, x_{[n]})$

1. Calculer  $F^*(x_{[i]}) \forall i \in \llbracket 1, n \rrbracket$

2. Calcul opérationnel de  $K$  :

$$k = \max_{1 \leq i \leq n} [F^*(x_{[i]}) - G(x_{[i]}), G(x_{[i]}) - F^*(x_{[i-1]})]$$

4. Comparer à la valeur critique  $D_n$  lue dans la table de la loi du  $\Delta$  de Kolmogorov.

Si  $k < D_n$  alors on accepte  $\mathcal{H}_0$ .

Le test s'applique aux fonctions de répartition continues.

### IV.3. Procédés de simulation

#### IV.3.1. Simulation d'une variable aléatoire

- Inversion générique

La méthode de simulation par inverse généralisée s'appuie sur la propriété suivante :  
Si  $U$  est une variable aléatoire uniforme et  $F_X$  est la fonction de répartition d'une variable  $X$   
alors  $F_X^{-1}(U)$  a même loi que  $X$ .

Cette méthode nécessite donc de connaître l'inverse généralisée de la fonction de répartition. L'algorithme de simulation pour un échantillon de taille  $n$  est alors le suivant :

Pour  $i = 1, \dots, n$

1. Générer  $U_i \sim \mathcal{U}([0,1])$
2.  $X_i = F_X^{-1}(U_i)$

- Acceptation-rejet

Certaines lois sont connues à une constante multiplicative près. C'est le cas des lois tronquées. Dans ce cas, une solution serait de simuler ces lois à partir d'une loi de proposition  $g$  plus simple que nous savons simuler puis d'appliquer l'algorithme d'acceptation-rejet.

Soit  $f$  la fonction de densité de la loi à simuler et  $g$  celle de la loi de proposition telles que :

$$\forall x \in \text{Supp}(f), \quad f(x) \leq Mg(x)$$

Nous pouvons simuler la loi  $f$  avec l'algorithme suivant :

Pour  $i = 1, \dots, n$

1. Générer  $X_i \sim g$  et  $U_i \sim \mathcal{U}([0,1])$

2. Tant que : (rejet)

$$U_i > \frac{f(X_i)}{Mg(X_i)}$$

Générer  $X_i \sim g$  et  $U_i \sim \mathcal{U}([0,1])$

3.  $Y_i = X_i$  (acceptation)

### IV.3.2. Simulation d'un mélange

$$X \sim P_X = \sum_{j=1}^J \alpha_j P_j$$

avec :

$$J \in \mathbb{N}^*$$

$$(\alpha_1, \dots, \alpha_j, \dots, \alpha_J) \in [0,1]^J$$

$$\sum_{j=1}^J \alpha_j = 1$$

Dans un premier temps, il nous faut modéliser une variable qui nous indiquera quelle loi générer parmi les  $J$  lois. Par exemple, nous la notons  $Loi$  et nous supposons qu'elle prend les valeurs suivantes :  $loi1, \dots, loi_j, \dots, loiJ$ . Nous créons un vecteur qui comporte les valeurs

$loi_1, \dots, loi_j, \dots, loi_J$  dans les proportions  $\alpha_1, \dots, \alpha_j, \dots, \alpha_J$ . Une simulation de cette variable consiste alors à tirer une valeur dans le vecteur.

L'algorithme pour la simulation d'un mélange est le suivant :

Pour  $i = 1, \dots, n$

1. Générer  $Loi_i$
2. Si  $Loi_i = loi_1$  alors générer  $X_i \sim P_1$   
 $\vdots$   
 Si  $Loi_i = loi_j$  alors générer  $X_i \sim P_j$   
 $\vdots$   
 Si  $Loi_i = loi_J$  alors générer  $X_i \sim P_J$

### IV.3.3. Simulation de variables corrélées

Le but de cette méthode est de simuler un vecteur aléatoire  $X = (X^1, \dots, X^k, \dots, X^K)'$  qui a les propriétés suivantes :

- $X^k \sim F_{X^k}$   
 où  $k = 1, \dots, K$   
 $F_{X^k}$  est la fonction de répartition d'une loi ayant une espérance et une variance finies
- $\Sigma_X$  est la matrice de corrélation

Soit  $Z = (Z^1, \dots, Z^k, \dots, Z^K)'$  un vecteur gaussien et  $\Sigma_Z$  sa matrice de corrélation.

La méthode consiste en une transformation du vecteur gaussien. La difficulté est de trouver  $\Sigma_Z$  qui permet d'avoir  $\Sigma_X$ .

Les étapes sont les suivantes :

1. effectuer la décomposition de Cholesky de  $\Sigma_Z$   
 $\rightarrow$  matrice triangulaire inférieure  $C$   
 $\Sigma_Z = CC^t$
2. pour  $i = 1, \dots, n$ 
  - générer  $W_i = (W_i^1, \dots, W_i^k, \dots, W_i^K)'$   
 où  $W_i^k \text{ iid } \sim \mathcal{N}(0,1)$
  - $Z_i = CW_i$  (variables gaussiennes corrélées)

- $U_i = \Phi(Z_i)$   
( $\Phi$  fonction de répartition d'une normale centrée réduite variables uniformes corrélées)
- $X_i = [F_{X^1}(U_i^1), \dots, F_{X^k}(U_i^k) \dots, F_{X^K}(U_i^K)]$  (variables corrélées)



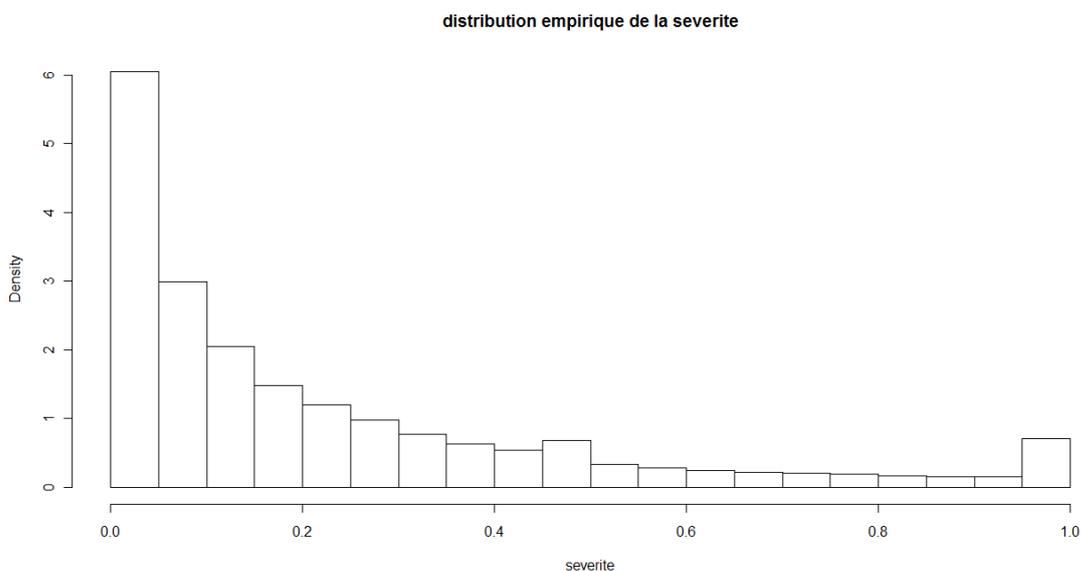
# Chapitre V Application au modèle de sévérité de niveau acheteur

Dans cette partie, nous allons appliquer les méthodes vues dans le chapitre précédent afin de modéliser la sévérité de niveau acheteur. Compte tenu des résultats de l'analyse descriptive, le modèle ciblé est un modèle en fonction du nombre de polices. Dans un premier temps, nous allons tout de même chercher un modèle de la sévérité de niveau acheteur au global.

## V.1. Modèle de sévérité au global

### V.1.1. Forme de l'histogramme

Nous visualisons à nouveau la distribution empirique de la sévérité acheteur à l'aide du logiciel R. Cette distribution a été tracée de façon à pouvoir superposer une fonction de densité lors des ajustements. Plus précisément, le support de la sévérité niveau acheteur a été divisé en intervalles. Pour chaque intervalle, nous pouvons calculer la probabilité empirique d'y appartenir : elle est égale au nombre d'observations contenues dans l'intervalle rapporté au nombre total d'observations. Enfin, la hauteur d'un bâton pour un intervalle est égale à cette probabilité empirique divisée par la taille de l'intervalle.



L'histogramme possède une forme dissymétrique (la skewness est de 1.55476416). Nous pouvons alors exclure les lois qui possèdent des propriétés de symétrie : la loi normale, la loi de Cauchy et la loi de Student.

Les lois testées sont : la loi gamma, la loi weibull, la loi log-normale, la loi exponentielle et la loi bêta.

Par ailleurs, nous avons pu constater que la distribution empirique est multimodale. Le « pic » en fin de distribution est particulièrement significatif. Nous pouvons affirmer qu'il serait difficile d'obtenir un bon ajustement avec une loi unique. Nous testons des mélanges de lois. Plus précisément, nous ajustons une loi sur  $]0\%, 100\%[$ . Nous devons alors utiliser une loi tronquée pour que son support coïncide avec  $]0\%, 100\%[$ . Finalement nous optons pour mélange d'une loi tronquée et d'une masse de Dirac.

### V.1.2. Ajustement

Le mélange d'une loi tronquée sur  $]0\%, 100\%[$  et d'une masse de Dirac à 100% peut s'écrire ainsi :

$$\tau = B + (1 - B)X | X < 1 \quad \text{où} \quad B \sim \mathcal{B}(p)$$

$X$  suit une des lois citées précédemment

Les paramètres à déterminer sont  $p$  et les paramètres de la loi de  $X$   $p'$ .

$$\Theta = (p, p')$$

Nous utilisons la méthode du maximum de vraisemblance pour obtenir une estimation des paramètres. Nous l'avons appliqué pour toutes les lois citées précédemment. Nous n'afficherons que le meilleur résultat, le reste se situant en annexe. La vraisemblance s'écrit :

$$\begin{aligned} L_n(\underline{x}; p, p') &= \prod_{i: x_i < 1} (1 - p) f_{X|X < 1}(x_i; p') \times \prod_{i: x_i = 1} P(\tau = 1) \\ &= p^{\text{Card}\{x_i | x_i = 1\}} \left( \frac{1 - p}{F_X(1; p')} \right)^{\text{Card}\{x_i | x_i < 1\}} \prod_{i: x_i < 1} f_X(x_i; p') \end{aligned}$$

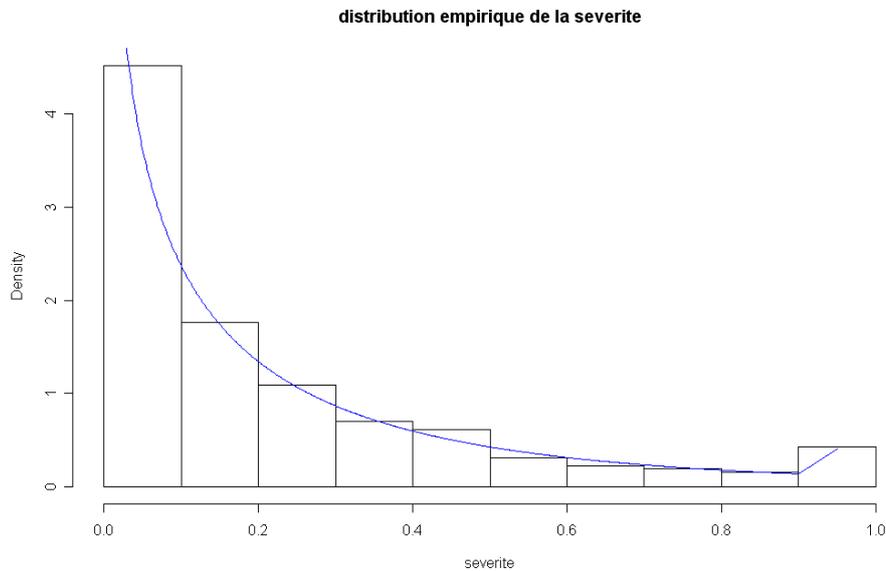
La log-vraisemblance s'écrit alors :

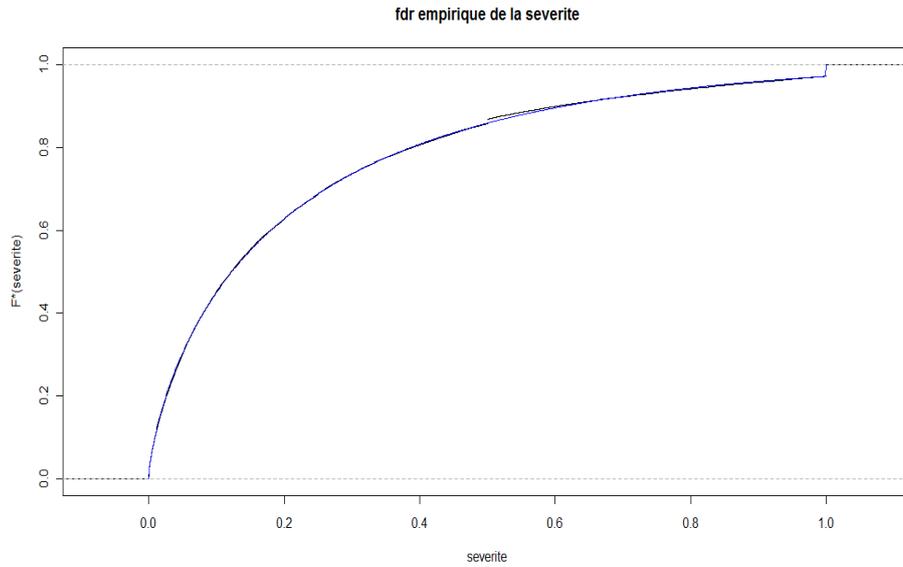
$$l_n(\underline{x}; p, p') = \text{Card}\{x_i | x_i = 1\} \log(p) + \text{Card}\{x_i | x_i < 1\} \log(1 - p) \\ - \text{Card}\{x_i | x_i < 1\} \log[F_X(1; p')] + \sum_{i: x_i < 1} \log[f_X(x_i; p')]$$

$$(\hat{p}, \hat{p}') = \underset{(\hat{p}, \hat{p}')}{\text{argmin}} l_n(\underline{x}; p, p')$$

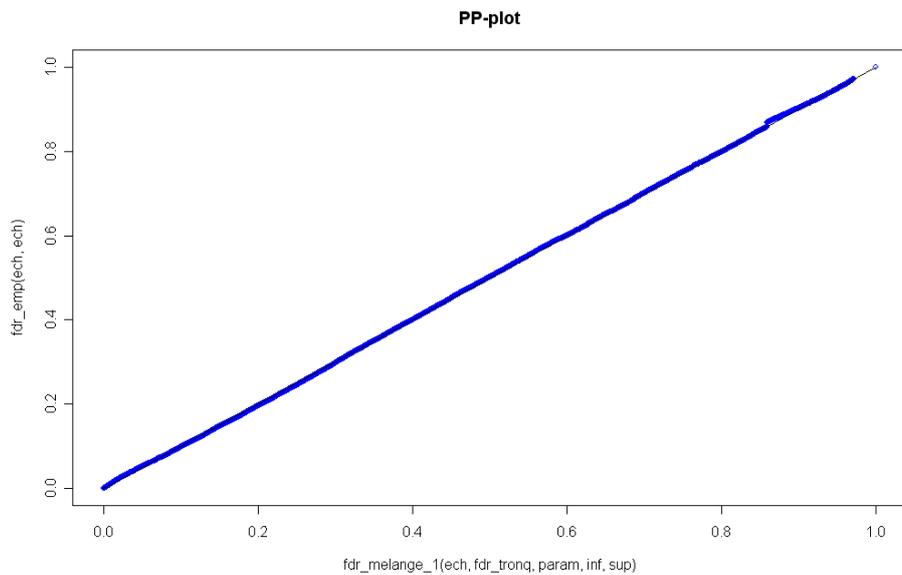
Nous avons ensuite tracé les diagrammes PP puis effectuer le test de Kolmogorov. La fonction de répartition n'étant pas continue, nous avons appliqué le test de Kolmogorov sur la loi tronquée. L'hypothèse nulle du test de Kolmogorov a été rejetée pour chacune des lois tronquées. Les lois tronquées, le mélange de lois, le diagramme PP et le test de Kolmogorov ont dus être implémentés (le code sous R se situe en annexe).

D'après le test de Kolmogorov, le meilleur résultat est celui obtenu avec la loi Weibull : l'écart entre la loi théorique et la loi empirique a été le plus petit. L'écart était de  $9,45 \times 10^{-3}$  pour une valeur critique de  $6,47 \times 10^{-3}$ . Nous pouvons affirmer que le test a été rejeté de justesse. Nous obtenons les résultats suivants :





Nous pouvons tracer un Probability-Probability plot (PP-plot) afin de visualiser la proximité entre la fonction de répartition empirique et celle du mélange.



Les points sont alignés sur la première bissectrice.

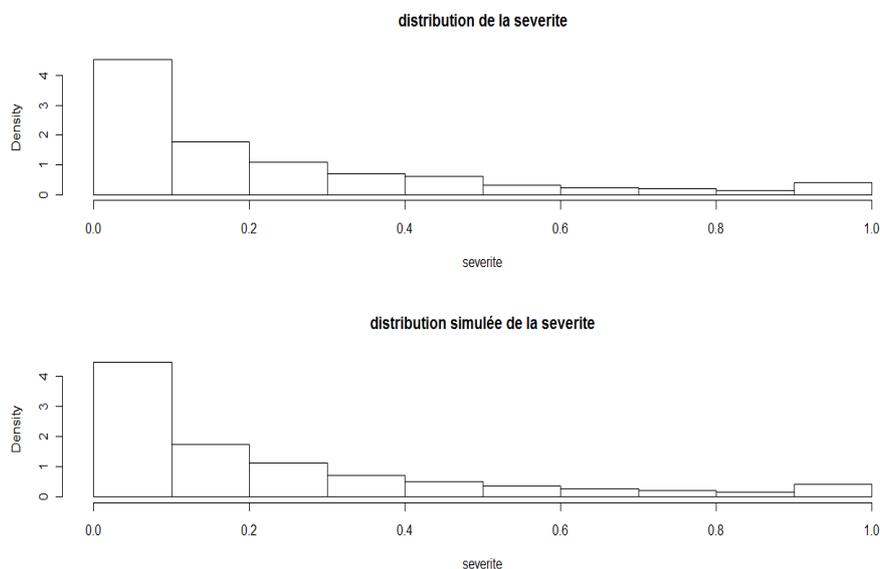
En conclusion, le meilleur modèle obtenu par ce mélange de loi est le suivant :

- avec une probabilité de 3%, la sévérité vaut un 100%
- avec une probabilité de 97%, elle suit une loi Weibull tronquée de support  $]0\%,100\%[$  et de paramètres  $(0,71 ; 0,21)$

Dans notre modélisation au global, nous aurions pu prendre en compte le pic à 50%.

### V.1.3. Simulation

Nous utilisons l'algorithme décrit précédemment pour les mélanges de lois (le code se situe en annexe).



Nous parvenons à reconstituer la distribution observée.

## V.2. Modèle de sévérité en fonction des facteurs discriminants

Rappelons que nous souhaitons modéliser la sévérité de niveau acheteur en fonction du nombre de polices exposées. Plus précisément, nous aimerions, à partir de la sévérité sur un seul agrément (quand le nombre de polices est égal à un), modéliser la sévérité de niveau acheteur lorsque le nombre total de polices relatives à un même acheteur est supérieur à un.

Notons  $\tau$  la sévérité lorsque le nombre de polices exposées est égal à un. Cette sévérité est donc une sévérité de niveau agrément.

Notons  $J$  le nombre total de polices relatives à un même acheteur.  $J \geq 1$

Nous voulons donc modéliser la sévérité de niveau acheteur lorsque  $J \geq 2$  à partir de la loi de  $\tau$ .

Nous utilisons les notations suivantes :

$\tau^J$	la sévérité relative à un acheteur sur lequel il y a $J$ agréments
$\tau_1, \dots, \tau_J$	les sévérités sur chaque police (même loi que $\tau$ )
$e_1, \dots, e_J$	les montants garantis des différents agréments
$(w_1, \dots, w_J) = \left( \frac{e_1}{\sum_{j=1}^J e_j}, \dots, \frac{e_J}{\sum_{j=1}^J e_j} \right)$	les proportions des montants garantis par rapport au montant total d'exposition
$i_1, \dots, i_J$	les montants d'indemnisation sur chaque police
$a^J$	le nombre de polices sinistrées $1 \leq a^J \leq J$

Nous avons alors :  $\tau^1 = \tau$  et  $a^1 = 1$

Nous rappelons que la sévérité de niveau acheteur est définie comme le rapport du montant d'indemnisation sur le montant d'exposition.

$$\tau^J = \frac{i_1 + \sum_{j=2}^J i_j \mathbb{I}_{a^J \geq j}}{\sum_{j=1}^J e_j}$$

Nous avons alors :

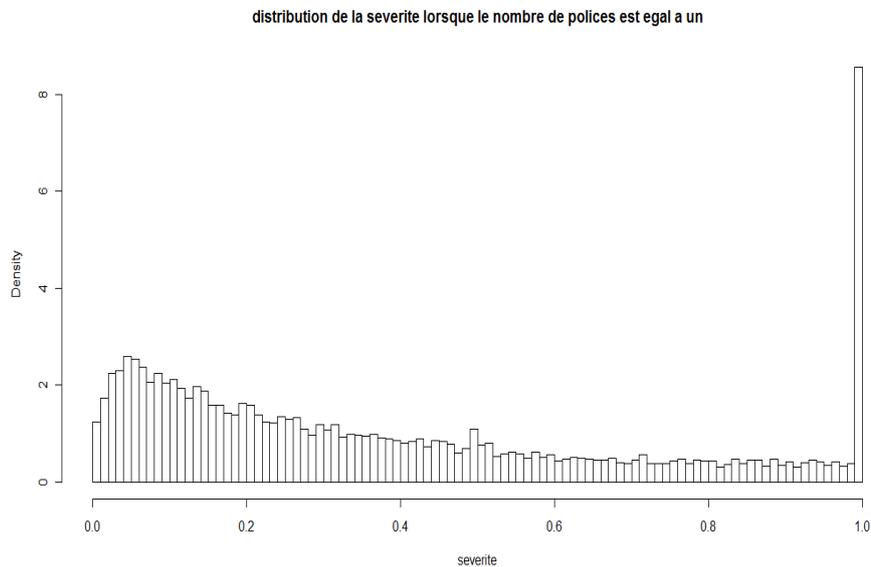
$$\tau^J = w_1 \tau_1 + \sum_{j=2}^J w_j \tau_j \mathbb{I}_{a^J \geq j}$$

Afin de modéliser cette sévérité, nous avons besoin de :

- la loi des  $\tau_j$  qui est celle de  $\tau$
- la loi de  $a^J$
- la loi des  $w_j$

## V.2.1. Modèle de sévérité sur un agrément

La distribution empirique de la sévérité lorsqu'il n'y a qu'un seul agrément est la suivante :

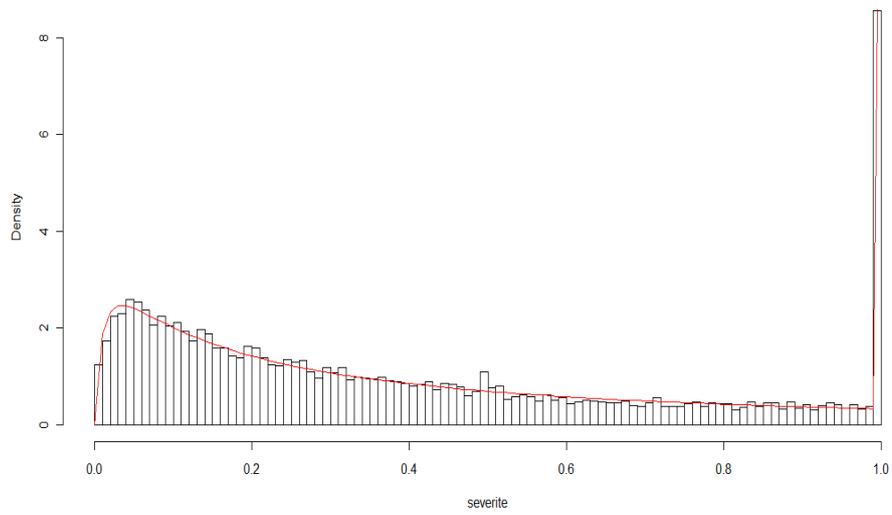


La distribution présente un pic en fin de distribution. Nous avons vu dans le paragraphe III.1.2 que la probabilité d'avoir un défaut total est environ de 8%. Comme pour la modélisation de la sévérité au global, nous optons pour un mélange d'une loi tronquée et d'une masse de Dirac. Nous procédons alors comme pour la modélisation de la sévérité au global. Tous les ajustements et les tests d'adéquation se situent en annexe.

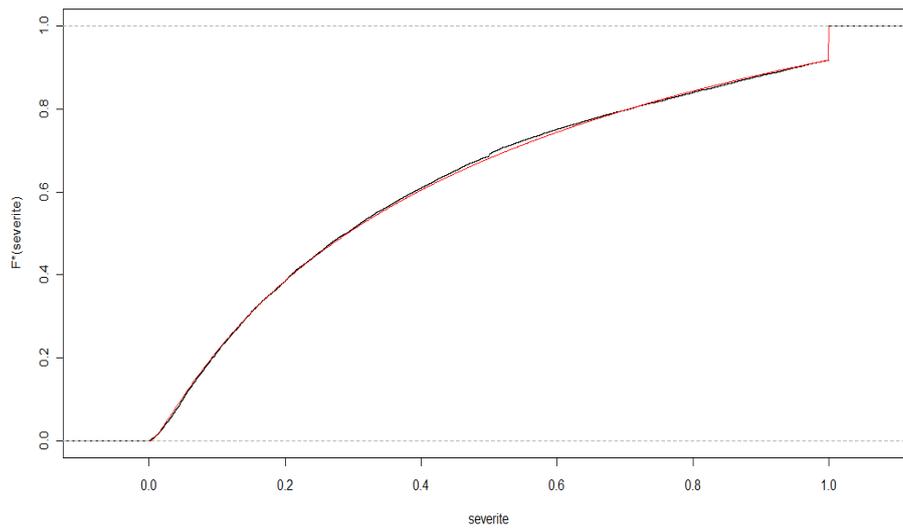
Nous n'afficherons que le meilleur résultat qui est le suivant :

- avec une probabilité de 8%, la sévérité vaut un 100%
- avec une probabilité de 92%, elle suit une log-normale tronquée de support  $]0\%,100\%[$  et de paramètres  $(-0,55 ; 1,68)$

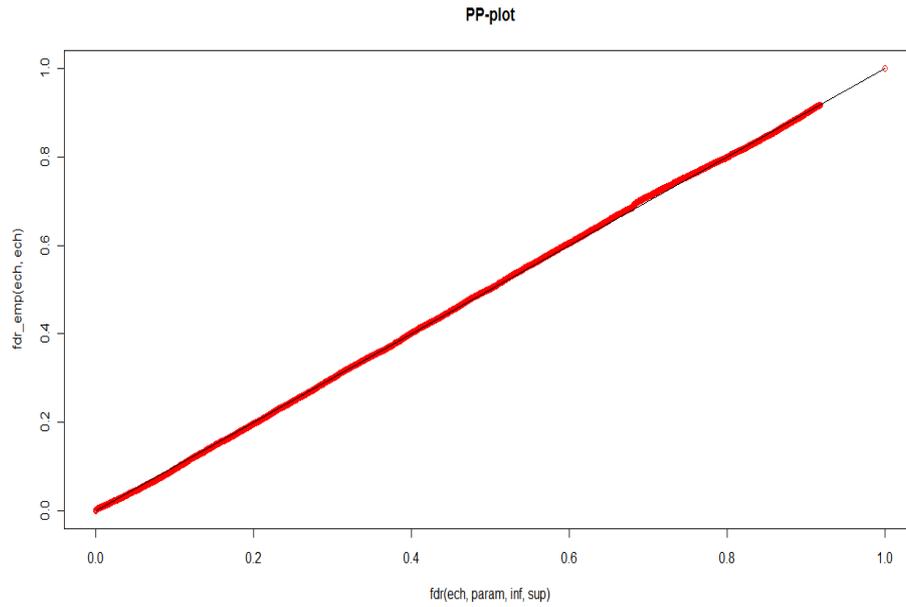
distribution de la severite lorsque le nombre de polices est egal a un



fdr empirique de la severite lorsque le nombre de polices est egal a un

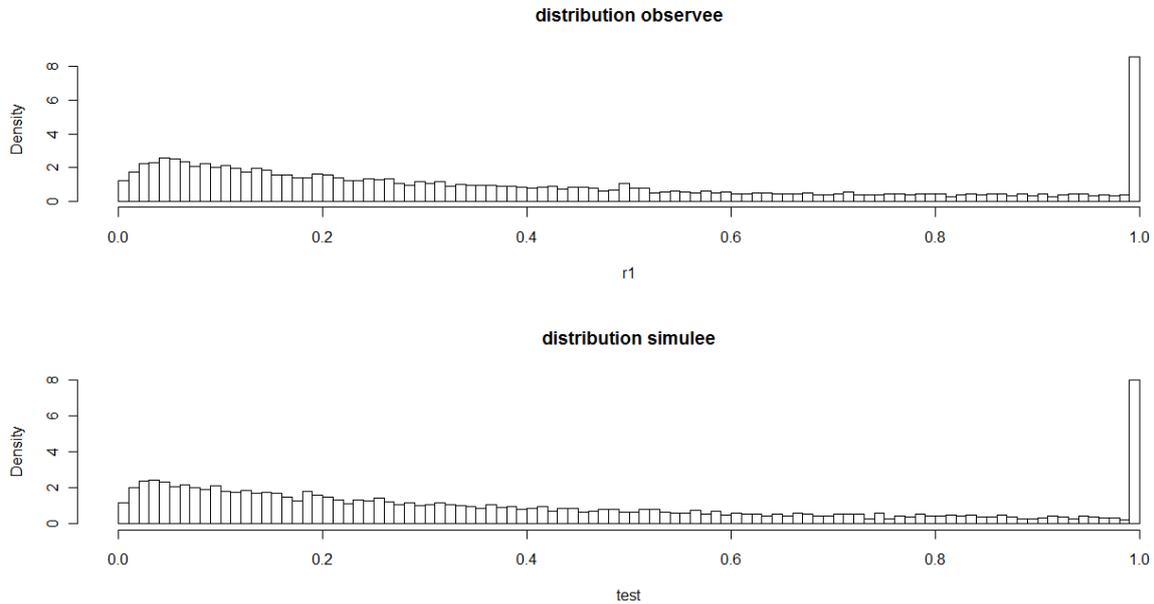


Nous pouvons affirmer que le modèle s'ajuste bien aux données.



Les points sont alignés sur la première bissectrice ce qui confirme la conclusion précédente.

De même que pour la sévérité au global, l'algorithme de simulation d'un mélange a été utilisé.



Nous parvenons à reconstituer la distribution observée.

## V.2.2. Loi de la proportion de polices sinistrées

Nous avons noté  $a^J$  le nombre de polices sinistrées lorsque le nombre total de polices exposées était égal à  $J$ . Nous pouvons affirmer que les valeurs possibles de cette variable sont  $\{j | 1 \leq j \leq J\}$ . Autrement dit, il s'agit d'une variable discrète sur un ensemble fini qui dépend de la valeur  $J$ . Nous aimerions l'approximer par une variable continue (plus facile à manipuler). Nous raisonnons alors avec la fonction de répartition qui a l'intérêt d'exister pour les lois discrètes mais aussi les lois continues.

Par ailleurs, nous raisonnons en pourcentage de polices sinistrées ce qui nous permet d'obtenir un support commun  $]0\%, 100]$ .

Notons  $\alpha^J$  la variable pourcentage de polices sinistrées quand le nombre total de polices est de  $J$ .

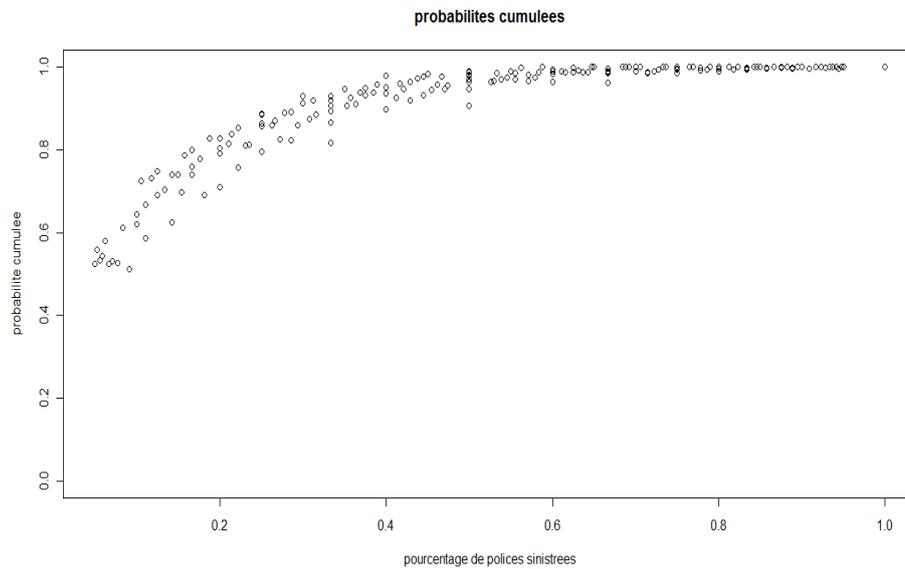
Le tableau suivant nous donne la répartition des acheteurs en fonction du nombre de polices sinistrées sachant le nombre total de polices exposées.

		Nombre de polices sinistrées																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Nombre total polices	1	100%																			
	2	91%	9%																		
	3	82%	15%	4%																	
	4	80%	15%	4%	1%																
	5	71%	19%	7%	2%	1%															
	6	74%	17%	6%	2%	1%	1%														
	7	62%	20%	9%	5%	2%	1%	0%													
	8	69%	17%	8%	4%	2%	1%	1%	0%												
	9	59%	17%	11%	6%	4%	2%	1%	0%	0%											
	10	62%	18%	11%	4%	2%	1%	1%	1%	0%	0%										
	11	51%	18%	13%	9%	3%	3%	1%	1%	0%	0%	0%									
	12	61%	15%	10%	7%	3%	2%	1%	1%	0%	0%	0%	0%								
	13	53%	17%	11%	7%	6%	2%	1%	2%	1%	0%	0%	0%	0%							
	14	53%	21%	10%	5%	3%	4%	2%	0%	1%	0%	1%	0%	0%	0%						
	15	52%	18%	9%	8%	2%	4%	4%	1%	0%	0%	0%	0%	0%	0%	0%					
	16	58%	17%	8%	6%	3%	3%	2%	2%	1%	0%	0%	0%	0%	0%	0%	0%				
	17	54%	19%	5%	3%	5%	5%	2%	2%	2%	3%	0%	0%	0%	0%	0%	0%	0%			
	18	53%	14%	13%	5%	3%	3%	4%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
	19	56%	17%	6%	3%	4%	3%	5%	1%	1%	1%	1%	2%	1%	0%	0%	0%	0%	0%	0%	0%
	20	52%	12%	10%	9%	6%	4%	2%	3%	1%	1%	0%	1%	1%	0%	0%	0%	0%	0%	0%	0%

Le tableau ci-dessous correspond aux probabilités cumulées :

		Nombre de polices sinistrées																				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Nombre total polices	1	100%																				
	2	91%	100%																			
	3	82%	96%	100%																		
	4	80%	95%	99%	100%																	
	5	71%	90%	96%	99%	100%																
	6	74%	91%	96%	99%	99%	100%															
	7	62%	82%	92%	97%	99%	100%	100%														
	8	69%	86%	93%	97%	99%	99%	100%	100%													
	9	59%	76%	87%	93%	97%	99%	100%	100%	100%												
	10	62%	80%	91%	95%	97%	98%	99%	100%	100%	100%											
	11	51%	69%	82%	91%	94%	98%	99%	99%	99%	100%	100%										
	12	61%	76%	86%	93%	96%	98%	99%	99%	100%	100%	100%	100%									
	13	53%	70%	81%	88%	94%	96%	97%	99%	100%	100%	100%	100%	100%								
	14	53%	74%	84%	89%	93%	96%	98%	98%	99%	99%	99%	100%	100%	100%	100%						
	15	52%	70%	79%	87%	89%	94%	98%	99%	99%	100%	100%	100%	100%	100%	100%	100%					
	16	58%	75%	83%	89%	92%	95%	97%	99%	100%	100%	100%	100%	100%	100%	100%	100%	100%				
	17	54%	73%	78%	81%	86%	91%	93%	95%	97%	100%	100%	100%	100%	100%	100%	100%	100%	100%			
	18	53%	67%	80%	85%	89%	92%	96%	98%	98%	98%	99%	99%	99%	99%	100%	100%	100%	100%	100%		
	19	56%	73%	79%	81%	86%	88%	94%	95%	96%	96%	97%	99%	100%	100%	100%	100%	100%	100%	100%	100%	
	20	52%	64%	74%	83%	89%	93%	95%	98%	98%	99%	99%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

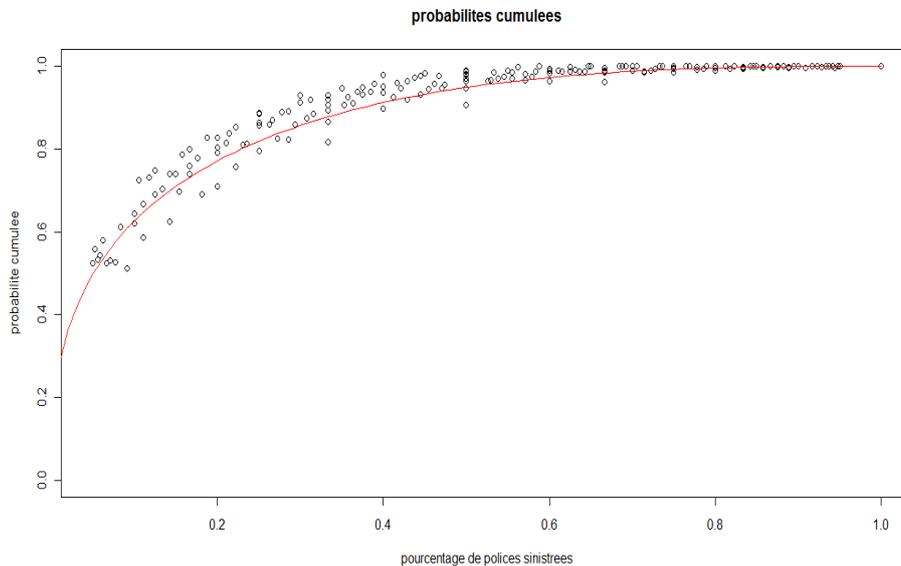
Nous remarquons que toutes les probabilités d'avoir au plus la moitié de polices sinistrées (en rose) sont peu variables. Nous observons la même chose pour toutes celles d'avoir au plus un tiers de polices sinistrées (en bleu). Si nous traçons le nuage de points correspondant aux probabilités cumulées en fonction du pourcentage de polices sinistrées, nous obtenons :



Nous considérons que les points ne sont pas trop dispersés. Autrement dit, nous émettons l'hypothèse que la loi du pourcentage de polices sinistrées ne dépend pas du nombre total de polices et est donc unique. Cette hypothèse est discutable.

Nous tentons de caler la fonction de répartition d'une loi connue. Pour chaque valeur du nombre total de polices  $J \geq 1$ , nous calculons la somme quadratique des écarts entre la probabilité cumulée constaté et la probabilité cumulée théorique. Nous affectons un poids à chaque somme quadratique, ce poids étant la proportion d'acheteurs ayant  $J$  polices au total. Enfin, nous cherchons les paramètres qui minimisent la somme des quantités précédentes.

Les meilleurs résultats ont été obtenus avec une loi bêta de paramètres (0,34 ; 2,49) :



Nous considérons que la proportion de polices sinistrées suit une loi bêta de paramètres (0,34 ; 2,49) :

$$\alpha^2, \alpha^3, \dots \sim \alpha \sim \text{Beta}(0,34 ; 2,49)$$

La sévérité se réécrit :

$$\tau^J = w_1 \tau_1 + \sum_{j=2}^J w_j \tau_j \mathbb{I}_{\alpha \geq \frac{j}{J}}$$

### V.2.3. Loi du montant garanti

Il nous reste à chercher une distribution pour les montants garantis :

$$e_1, \dots, e_J$$

Plus précisément nous allons modéliser leur proportion par rapport au montant total d'exposition. Nous les avons notés  $w_1, \dots, w_J$ .

Nous rappelons que nous avons les montants d'exposition au niveau acheteur. Autrement dit, nous ne possédons pas le détail des montants d'agrément. Sans cette information, il est difficile de poursuivre la modélisation.

Afin de continuer notre étude de la sévérité, nous avons eu à émettre des hypothèses sur la distribution sur des montants garantis. Nous espérons ainsi comprendre la logique d'agrégation.

#### V.2.4. Résultats

La sévérité s'exprime :

$$\tau^J = w_1 \tau_1 + \sum_{j=2}^J w_j \tau_j \mathbb{I}_{\alpha \geq \frac{j}{J}}$$

Nous rappelons : pour  $J \geq 2$

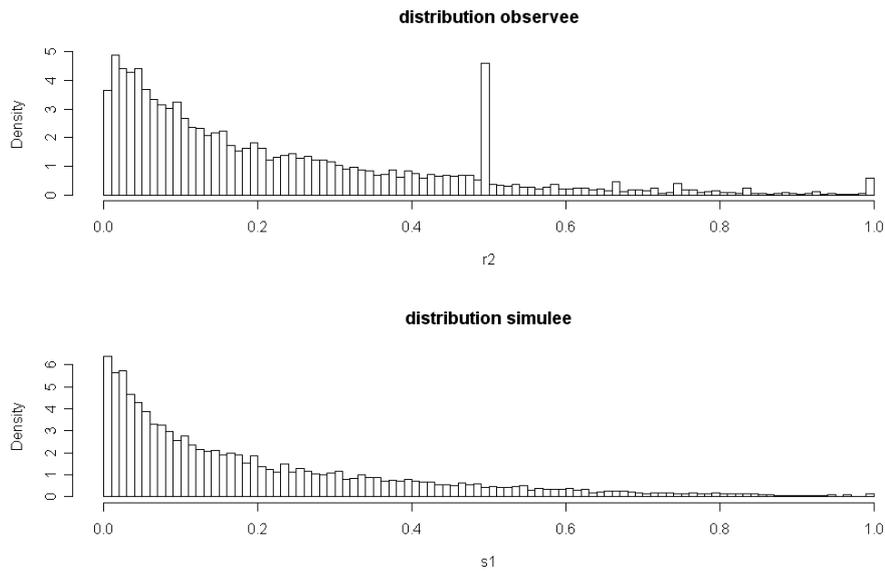
- $\tau_1, \tau_2, \dots, \tau_J \sim B + (1 - B)X | X < 1$  (égalité en lois)  
avec  $B \sim \mathcal{B}(8\%)$   
 $X \sim \mathcal{LN}(-0,55 ; 1,68)$
  
- $\alpha \sim \text{Beta}(0,34 ; 2.49)$

Dans cette partie, nous allons montrer graphiquement les résultats de la simulation de la sévérité en fonction des hypothèses que nous émettrons sur les montants garantis. L'implémentation des simulations se situent en annexe.

Nous nous intéressons au cas où le nombre de polices est égal à deux. Dans un premier temps, nous émettons l'hypothèse que les proportions des montants d'agréments sont uniformes et indépendants :

$$w_1, w_2, \dots, w_J \text{ iid } \sim \mathcal{U}([0,1])$$

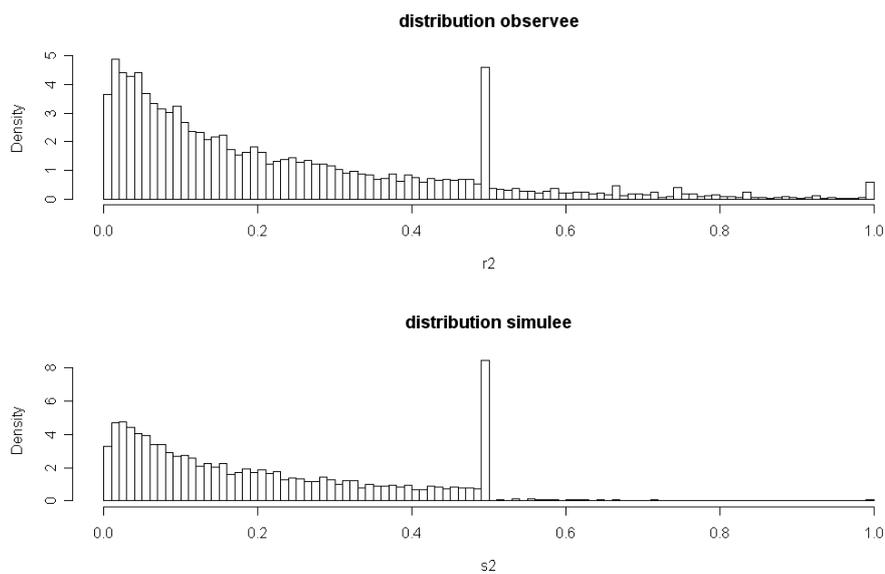
Nous obtenons :



Les pics à 50% et à 100% n'apparaissent pas. La forme du reste de la distribution est ressemblante. Nous rappelons que le pic à 50% s'explique par le fait d'avoir une seule police sur les deux qui a sinistré avec des montants garantis sont égaux.

Supposons maintenant que les proportions des montants d'agréments sont égales :

$$w_1, w_2, \dots, w_j = 0,5$$



Le pic central est bien apparent mais sa hauteur est trop importante si nous la comparons à celle observée.

Intuitivement, nous combinons les deux distributions. Par ailleurs, nous nous demandons quels phénomènes impliqueraient l'égalité des montants garantis sur un même acheteur. Nous supposons qu'il y a un effet de plafonnement du montant accordé : le montant demandé par un client sera accordé à hauteur d'un montant de référence. L'égalité des montants garantis serait atteinte lorsque les montants demandés sont supérieurs au montant de référence. Cette hypothèse reste à vérifier.

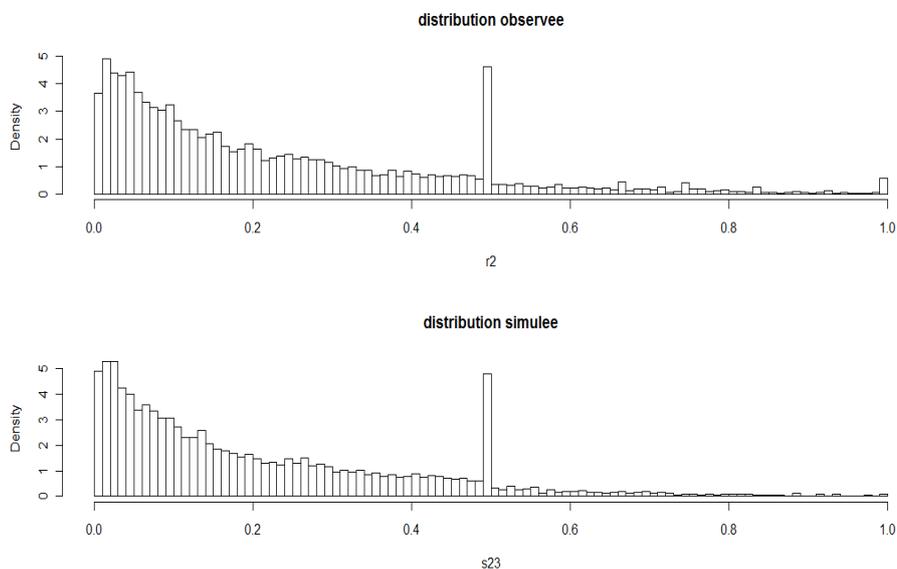
Nous avons alors :

$$\frac{e_i}{e_{ref}} = I + (1 - I)U([0,1])$$

avec  $I \sim \mathcal{B}(p)$

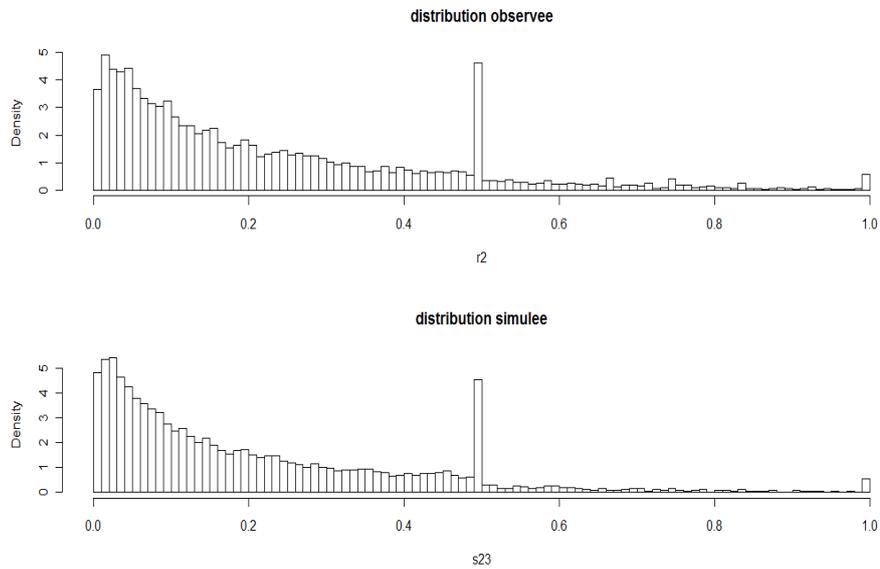
Lorsque le nombre de polices exposées est égal à deux, nous avons vu qu'environ 50% des cas correspondaient à des montants garantis égaux. Fixons alors  $p$  à  $71\% = \sqrt{50\%}$ .

Nous constatons :



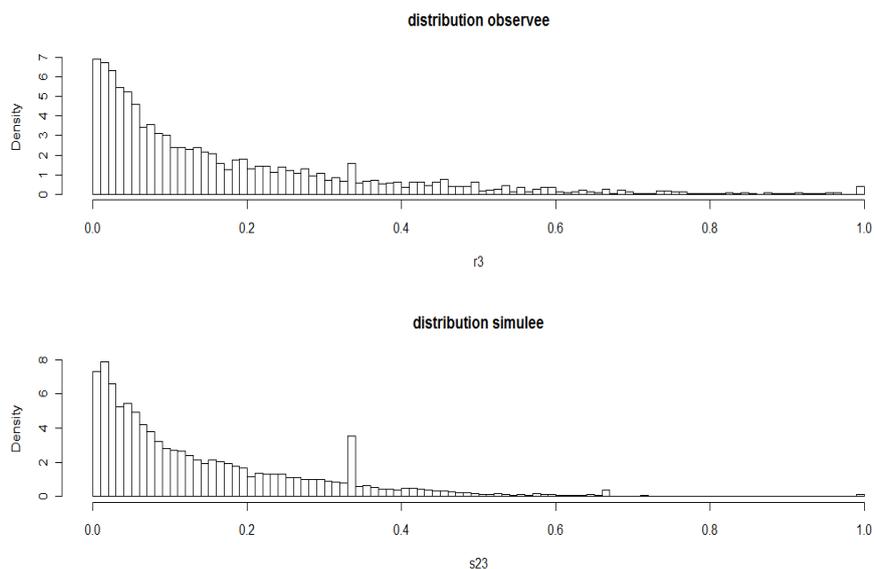
Le pic à 50% est bien reconstitué. Cependant, le pic à 100% n'apparaît pas du tout. Nous incorporons de la corrélation entre les sévérités de niveau agrément grâce à une copule gaussienne dont la simulation a été décrite dans le chapitre III.

Avec une forte corrélation nous avons :



Une forte corrélation entre les sévérités signifie que le comportement de l'acheteur est semblable d'un fournisseur à un autre. Graphiquement, nous pouvons dire que nous reconstituons la distribution de la sévérité lorsque le nombre de polices est égal à deux.

Si nous gardons les mêmes paramètres à savoir une probabilité de 71% d'avoir le montant garanti limite et une forte corrélation entre les sévérités, nous obtenons graphiquement pour un nombre total de polices égal à trois :



Nous constatons que le pic à 33% est trop important. La probabilité d'avoir un montant garanti limite est donc trop élevé. Si nous diminuons cette probabilité, nous ne reconstituerons plus le pic à 50% quand le nombre de polices est égal à deux. La reconstitution de ce pic serait possible en générant des montants garantis corrélés. Cependant cette corrélation est difficile à interpréter. Elle est d'autant plus difficile que le nombre de polices exposées est élevé. Comme pour la sévérité, nous utilisons une copule gaussienne pour générer des Bernoulli corrélées.

En jouant sur les différents paramètres : corrélation entre les montants garantis et la probabilité d'avoir le montant garanti limite, nous avons trouvé le modèle suivant pour la sévérité :

$$\tau^J = w_1 \tau_1 + \sum_{j=2}^J w_j \tau_j \mathbb{I}_{\alpha \geq \frac{j}{J}}$$

- $\tau_1, \tau_2, \dots, \tau_J \sim B + (1 - B)X | X < 1$  (égalité en lois)  
avec  $B \sim \mathcal{B}(8\%)$   
 $X \sim \mathcal{LN}(-0,55 ; 1,68)$

$\tau_j$  fortement corrélés

- $\alpha \sim \text{Beta}(0,34 ; 2.49)$
- Si  $J \geq 5$  alors les  $w_i$  sont indépendants et suivent une loi uniforme  
Si  $3 \leq J \leq 5$  alors il y a un effet de plafonnement

$$\frac{e_i}{e_{ref}} = I + (1 - I)\mathcal{U}([0,1])$$

avec  $I \sim \mathcal{B}(0,6)$

Si  $J = 2$  alors il y a en plus une corrélation entre les montants garantis

Dans cette partie, la calibration du modèle a été effectuée à l'aide de simulations dépendantes des différentes hypothèses émises et à l'aide de la visualisation des résultats de ces simulations. Par ces travaux nous avons pu détecter des phénomènes tels que la corrélation des sévérités ou l'effet de plafonnement du montant garanti. Cependant ces phénomènes peuvent être discutés. En effet, il est nécessaire d'obtenir les données permettant la validation ou non du modèle car un nombre important d'hypothèses a dû être émis.

## Chapitre VI Intégration du modèle de sévérité dans la modélisation du risque de crédit

Ce chapitre correspond à une application directe de l'étude et la modélisation de la sévérité des sinistres de niveau acheteur. Nous allons intégrer le modèle de sévérité dans un modèle simplifié de risque de crédit. Nous allons appliquer ce modèle à un portefeuille d'un client de Coface donné.

Le modèle est le suivant :

$$S_i = \chi E_i \tau$$

Où  $S_i$  désigne la perte liée à un acheteur  $i$

$\chi \sim \mathcal{B}(PD)$  avec  $PD = 4\%$

$E_i$  désigne le montant d'exposition au moment du défaut sur l'acheteur  $i$

$\tau$  désigne la sévérité des sinistres de niveau acheteur.

La perte totale liée à cette police sera la somme des pertes individuelles. Le portefeuille contient 265 acheteurs.

$$S = \sum_{i=1}^{265} S_i$$

Nous rappelons que ces acheteurs peuvent apparaître dans d'autres portefeuilles/polices. Nous avons vu l'importance du nombre total de polices. Le tableau suivant nous donne quelques statistiques descriptives du nombre total de polices :

<i>Nb polices</i>	
Moyenne	65
Médiane	36
Mode	2
Écart-type	77
Variance	5868
Minimum	1
Maximum	483
Somme	17113
Nombre	265

Nous voulons étudier l'impact du comportement de la sévérité sur les pertes. Nous allons évaluer les pertes pour trois scénarios :

1. sévérité constante et égale à la moyenne des sévérités  $\tau = 22\%$
2. sévérité empirique
3. sévérité en fonction du nombre de polices

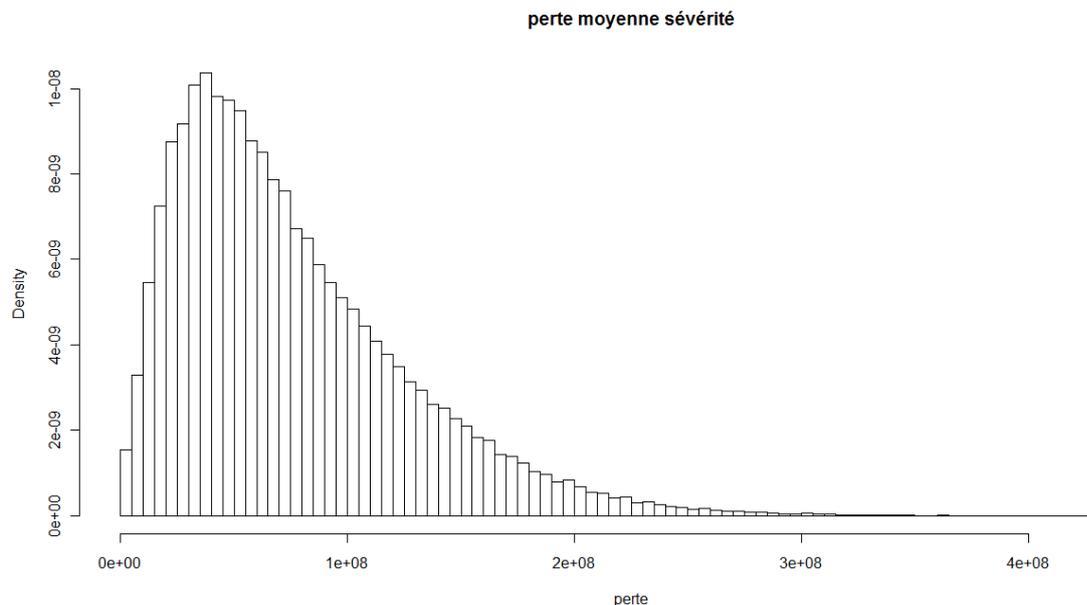
Nous obtenons :

Scénario	Min	Max	VaR99%	Moyenne	Ecart type
1	0	428 100 000	233 001 934	75 480 000	51 006 311
2	0	951 301 166	386 646 359	76 360 768	78 982 463
3	0	523 638 059	118 697 671	16 234 313	24 400 740

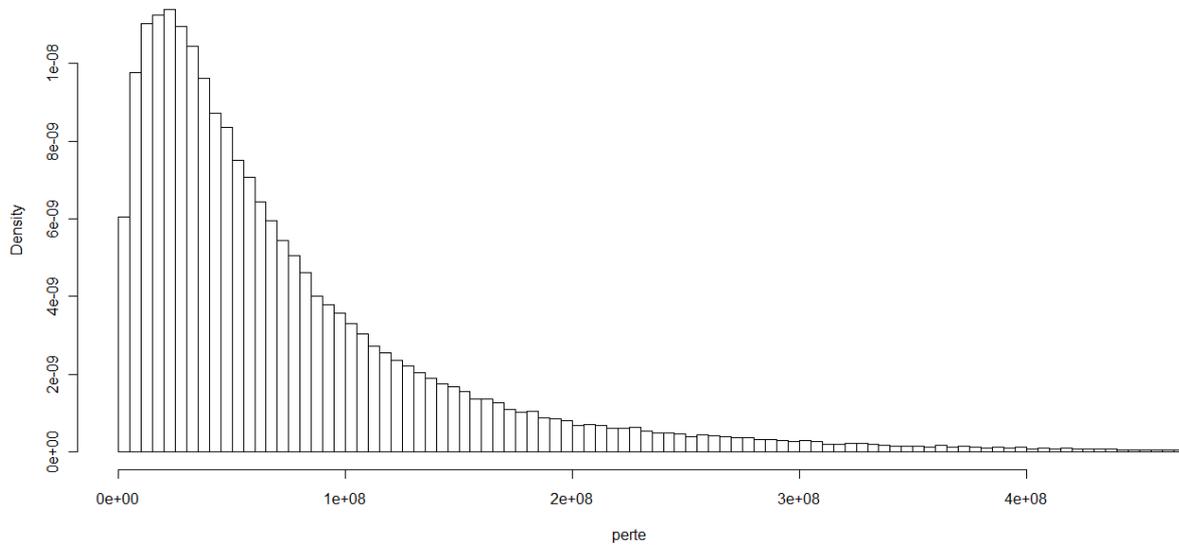
Le modèle le plus volatile est celui avec la sévérité empirique. L'écart type et donc le maximum et la VaR 99% sont plus élevés. Nous trouvons la même moyenne lorsque la sévérité est égale à la moyenne et lorsque la sévérité est empirique. Cette remarque constitue un point de validation de la méthode de simulation de Monte Carlo.

Concernant la moyenne des pertes trouvée avec le modèle de sévérité en fonction du nombre de polices, elle est très inférieure aux deux autres moyennes.

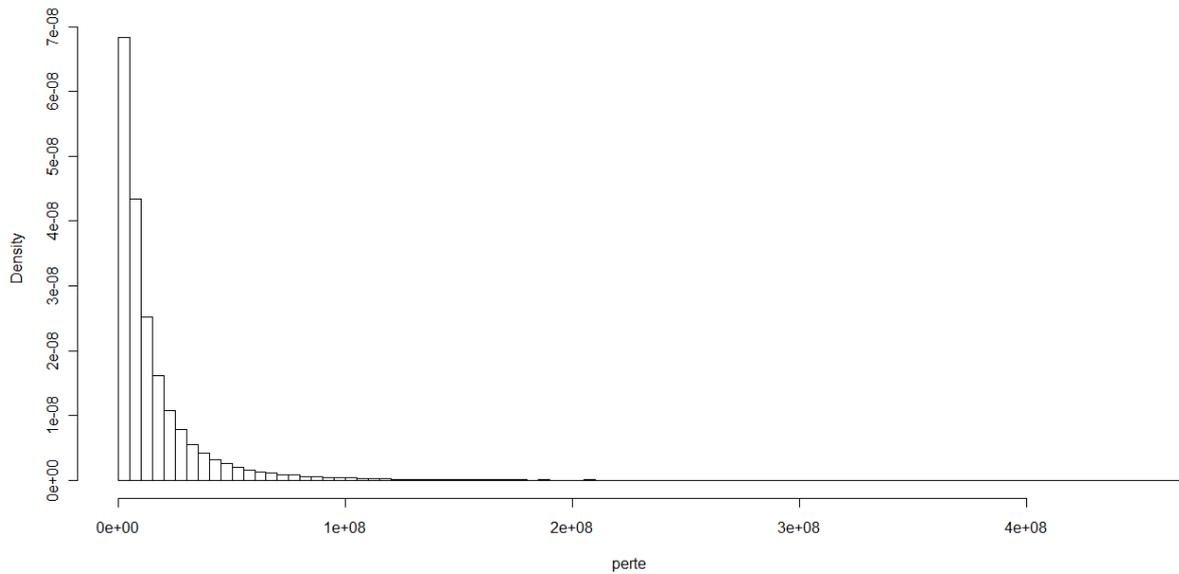
Graphiquement :



perte sévérité empirique



perte sévérité modélisée



Nous pouvons affirmer que l'utilisation du modèle de sévérité en fonction du nombre de polices conduit à des estimations de pertes plus faibles. Mais nous manquons d'information pour juger de la qualité des résultats trouvés. Nous rappelons que des hypothèses ont dus être émises pour ce modèle.



## Conclusion

Les études menées pour le mémoire concernent l'étude et la modélisation de la sévérité des sinistres au niveau acheteur.

Les travaux réalisés ont d'abord permis de comprendre le contexte et le fonctionnement de l'assurance crédit. Des méthodes d'analyse descriptive ont été utilisées pour déterminer les facteurs significatifs de la sévérité. Des ajustements ont été ensuite effectués afin de déterminer les différents paramètres du modèle : nous avons procédé avec des méthodes classiques mais aussi avec des méthodes intuitives. De même des méthodes classiques de simulation ont été utilisées notamment pour la modélisation de la sévérité au global. Concernant la sévérité en fonction du nombre de polices, le manque d'information nous a amenés à émettre des hypothèses justifiées et à en constater les conséquences sur la simulation. Cette manière de procéder nous a permis de détecter des phénomènes qui seront vérifiés et pris en compte. Un axe d'amélioration est de vérifier ces hypothèses pour la validation du modèle.

Pour ce mémoire, un travail important d'implémentation a été effectué : les différentes lois (tronquées, mélanges), les différentes méthodes d'ajustement et de simulation ont dues êtes implémentées sous le logiciel R.

Enfin, nous nous sommes intéressés à la sévérité de niveau acheteur. Par le même raisonnement, la sévérité de niveau agrément ou la sévérité de niveau groupe peut être modélisée.



## Références bibliographiques

- [1] Assurance non-vie : Modélisation, Simulation – Christian PARTRAT, Jean-Luc BESSON, ECONOMICA, 2005
  
- [2] Probabilités, Analyse des données et Statistiques – Gilbert SAPORTA, Editions TECHNIP, 2006
  
- [3] La gestion des risques financiers – Thierry RONCALLI, ECONOMICA, 2009
  
- [4] Théorie du risque et de la réassurance – Christian HESS, cours Master 2 Actuariat Paris-Dauphine, 2009-2010
  
- [5] Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix – Marne C. Cario, Barry L. Nelson, Steven D. Roberts, James R. Wilson, 2000
  
- [6] Modélisation et mesures du risque de crédit du groupe Coface : Application à la détermination du capital économique – Anne GENEST, mémoire, 2005
  
- [7] Etudes actuarielles en assurance crédit et applications dans les méthodes d'allocation du capital économique pour le groupe Coface – Iuliana COVACI VASILESCU, mémoire, 2006
  
- [8] Data mining et statistique décisionnelle – Stéphane TUFFERY, Editions TECHNIP, 2005



# ÉTUDE DES VALEURS ANORMALES DE LA SEVERITE DES SINISTRES

## *NIVEAU ACHETEUR*

### INTRODUCTION

La sévérité est définie comme le taux de sinistre en cas de sinistre. De façon schématique, il s'agit du rapport entre le montant de sinistre et le montant d'exposition. Elle est comprise entre 0 et 1.

$$\tau = \frac{\text{montant du sinistre}}{\text{montant d'exposition}}$$

Suite au rapprochement entre le montant du sinistre et le montant de l'exposition, nous obtenons des valeurs de sévérité supérieures à 1. Or Coface n'indemnise pas plus que ce qu'elle garantit.

Plus généralement, entre la date à laquelle nous sélectionnons l'exposition et la date du sinistre, le montant garanti peut augmenter ce qui expliquerait un montant de sinistre plus élevé. Dans ce cas, nous avons rapproché un montant de sinistre à un montant d'exposition qui n'est plus représentatif du montant réellement en risque.

Le sujet de cette note porte sur l'étude de ces valeurs anormales.

## 1. DETECTION DES VALEURS ANORMALES DE SEVERITE

Nous avons un échantillon de 55232 valeurs de sévérité.

Nous cherchons les observations pour lesquelles le montant d'exposition a augmenté entre la date de l'exposition sélectionnée et la date du sinistre.

Pour 14592 acheteurs de notre échantillon initial, nous n'avons pas retrouvé de montant d'exposition entre la date de l'exposition sélectionnée et la date du sinistre. Appelons *groupe 1* cet ensemble d'acheteurs.

Pour les autres acheteurs, nous comparons le montant d'exposition initial avec ceux retrouvés. Nous rappelons que nous possédons les montants d'exposition aux dates du 01/03, 01/06, 01/09 et 01/12 de chaque année.

Autrement dit, si une augmentation puis une diminution ont lieu entre 2 dates, l'augmentation du montant d'exposition ne sera pas détectée.

Sur les 40640 autres acheteurs, 10843 acheteurs ont eu une augmentation de l'encours (*groupe 2*). Donc 29797 acheteurs n'ont pas eu d'augmentation perceptible de leur encours. (*groupe 3*)

## 2. TRAITEMENT DES VALEURS DU GROUPE 2 : RETRAIT 1

Etudions les valeurs du *groupe 2*. Nous allons observer si le nombre de polices a augmenté et nous allons calculer le pourcentage d'augmentation du montant d'exposition.

Si nous prenons la dernière augmentation en compte, nous obtenons :

Frequence %_augmentation_expo	augmentation_nbpol		Total
	non	oui	
0-0.1	635	656	1291
0.1-0.2	411	692	1103
0.2-0.3	322	688	1010
0.3-0.4	210	496	706
0.4-0.5	164	407	571
0.5-0.6	202	499	701
0.6-0.7	143	324	467
0.7-0.8	87	232	319
0.8-0.9	86	232	318
0.9-1	53	110	163
>1	768	3426	4194
	3081	7762	10843

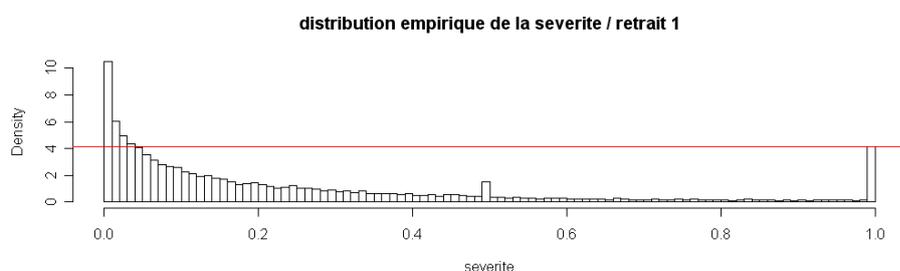
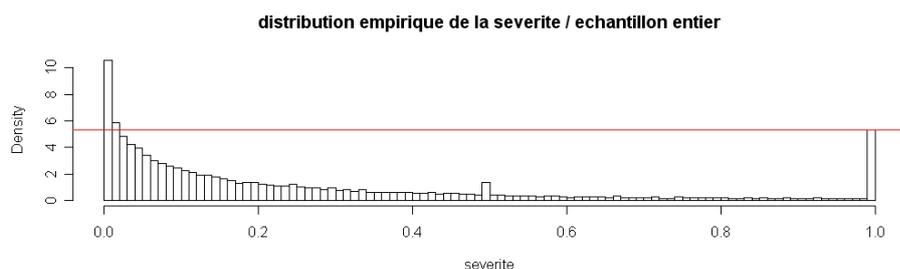
Nous remarquons que dans 72% des cas, l'augmentation de l'exposition est due à une augmentation du nombre de polices.

Nous gardons dans notre échantillon les acheteurs qui répondent à l'un des critères suivants :

- soit le nombre de police n'a pas augmenté, nous gardons alors les acheteurs pour lesquels l'augmentation de l'exposition est strictement inférieure à 20%
- soit le nombre de police a augmenté, nous sommes plus sévères et nous gardons les acheteurs pour lesquels l'augmentation de l'exposition est strictement inférieure à 10%

Au total, 9141 acheteurs sur les 10843 (84% du *groupe 2*, 17% de l'échantillon total) sont retirés de notre échantillon.

Concernant les 1702 autres valeurs de sévérité du groupe, elles sont bornées par 1 car Coface n'indemnise pas plus que ce qu'elle garantit. Traçons alors la distribution empirique de la sévérité.



Le pic à 100% a diminué. En effet la probabilité d’avoir un sinistre total passe de 5% à 4%.

### 3. TRAITEMENT DES VALEURS DU GROUPE 1 ET DU GROUPE 3 : RETRAIT 2

---

Pour ces données appartenant à ces groupes, nous n’avons pas détecté d’augmentation de l’encours. Nous ne pouvons expliquer les valeurs anormales

de la sévérité. Cependant les valeurs de sévérités très supérieures à 1 sont aberrantes.

La taille de ce groupe est de 44389.

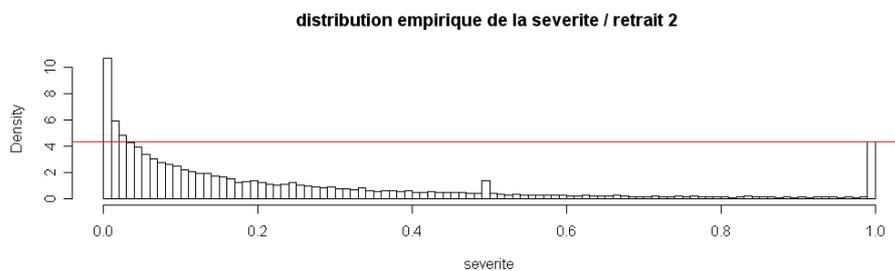
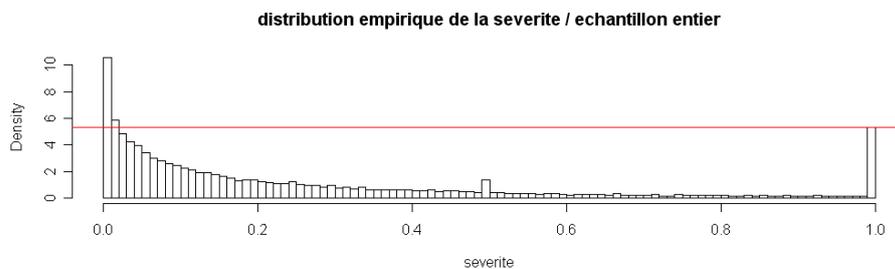
Nous avons :

Tranches sévérité	Fréquence	Pourcentage
<=100%	43 607	98.2%
]100%,110%[	125	0.3%
[110%,120%[	98	0.2%
[120%,130%[	94	0.2%
[130%,140%[	70	0.2%
[140%,150%[	58	0.1%
[150%,160%[	46	0.1%
[160%,170%[	33	0.1%
[170%,180%[	30	0.1%
[180%,190%[	17	0.0%
[190%,200%[	12	0.0%
>=200%	199	0.4%
	<b>44 389</b>	<b>100%</b>

La proportion de valeurs de sévérité supérieures à 1 est de 2%.

Nous décidons de retirer les valeurs de sévérité supérieures à 120%. Elles sont au nombre de 559 soit 1%.

Traçons la distribution empirique de la sévérité afin de visualiser l'effet de ces retraits.



De même que précédemment ces retraits ont pour effet de diminuer le pic à 100%. En effet la probabilité d'avoir un sinistre total passe de 5% à 4%.

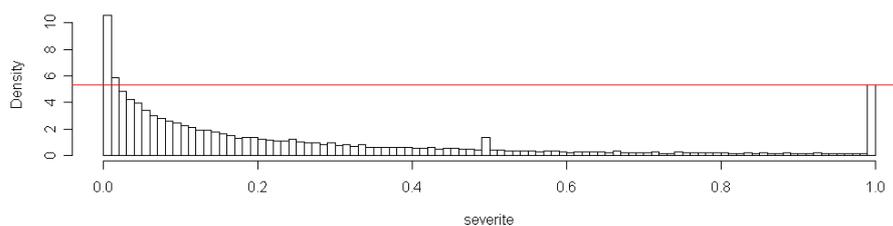
## 4. RESULTATS DE LA COMBINAISON DES 2 RETRAITS

---

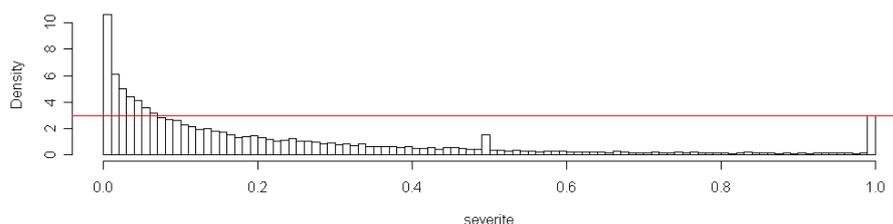
Si nous retirons toutes les valeurs selon les critères précédants simultanément, nous obtenons un échantillon de taille 45532.

La distribution empirique est :

distribution empirique de la severite / echantillon entier



distribution empirique de la severite / retrait 1 & retrait 2



La probabilité d’avoir un sinistre total passe de 5% à 3%.

▪ Pays – Severite

Nous avons 159 modalités pour le pays. Considérons les pays pour lesquels le nombre d’observations est supérieur à 200 pour obtenir des statistiques significatives.

Nous obtenons :

Libelle pays	Données						
	Frequence	Pourcentage	Moyenne sev	Ecart type sev	Sev moyenne		
FRANCE	11 114	27%	22%	26%	3%	min	
SPAIN	8 135	20%	24%	26%	5%	max	
GERMANY	6 446	16%	21%	23%	6%	min-max	
ITALY	5 008	12%	19%	24%	4%		
UNITED KINGDOM	1 928	5%	20%	25%	3%		
NETHERLANDS	1 608	4%	22%	25%	8%		
PORTUGAL	1 518	4%	25%	27%	5%		
BELGIUM	1 287	3%	20%	25%	2%		
USA	931	2%	19%	24%	4%		
GREECE	482	1%	28%	29%	12%		
POLAND	362	1%	22%	25%	9%		
BRAZIL	361	1%	35%	28%	16%		
SWITZERLAND	330	1%	19%	25%	4%		
AUSTRIA	305	1%	16%	21%	4%		
IRELAND	298	1%	22%	26%	3%		
TURKEY	290	1%	27%	29%	13%		
DENMARK	277	1%	24%	28%	2%		
HUNGARY	232	1%	26%	28%	7%		
AUSTRALIA	209	1%	35%	28%	19%		
SWEDEN	207	1%	24%	27%	2%		

Total	41 328	100%	22%	25%	4%
			19%	8%	17%

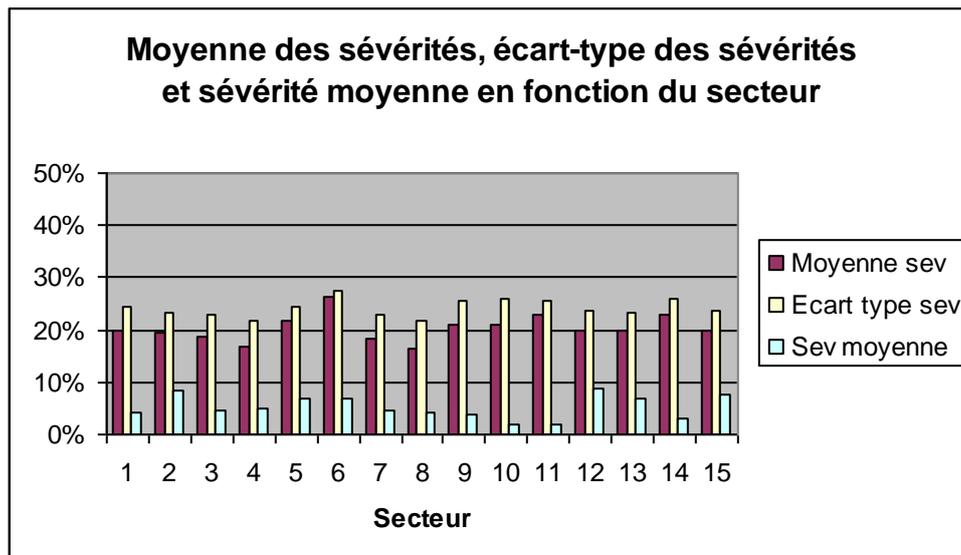
▪ Secteur – Severite

Nous avons 15 modalités pour le secteur d'activité. 4422 valeurs sont manquantes soit environ 10%.

Nous obtenons :

Libelle groupe 15	Données					
	Frequence	Pourcentage	Moyenne sev	Ecart type sev	Sev moyenne	
01 - Agriculture, viande, agro et vins	4750	12%	20%	24%	4%	min
02 - Bois & meubles	887	2%	19%	23%	8%	max
03 - Minéraux, chimie, pétrole, plastique, pharma et verre	1805	4%	19%	23%	5%	max-min
04 - Métaux	1474	4%	17%	22%	5%	
05 - Papiers, emballages et imprimerie	1145	3%	22%	25%	7%	
06 - Construction	10107	25%	26%	27%	7%	
07 - Textile, cuir et habillement	6421	16%	18%	23%	4%	
08 - Mécanique et Mesure	1743	4%	16%	22%	4%	
09 - Biens d'équipements électriques, électronique, informatique et télécoms	1935	5%	21%	25%	4%	
10 - Autos & Cycles, autres véhicules et transport	2695	7%	21%	26%	2%	
11 - Commerces non spécialisés ou nca	4698	11%	23%	26%	2%	
12 - Services collectifs	315	1%	20%	24%	9%	
13 - Services financiers	294	1%	20%	23%	7%	
14 - Services aux entreprises & particuliers	2296	6%	23%	26%	3%	
15 - Divers	545	1%	20%	23%	8%	
Total	41110	100%	22%	25%	4%	

	10%	6%	7%
--	-----	----	----



#### CODE R

```

decoupage_nbpol=fonction(table,min_inclus=-
Inf,max_inclus=+Inf,num_col,nums_col_sortie=c(1:dim(table)[2])){
  res=matrix(ncol=length(nums_col_sortie));
  for (i in 1:dim(table)[1]){

```

```

    if (table[i,num_col]>=min_inclus & table[i,num_col]<=max_inclus){
      res=rbind(res,table[i,nums_col_sortie]);
    }
  }
  res[-1,];
}

```

```

decoupage_encours=function(table,min_inclus=-
Inf,max_exclus=+Inf,num_col,nums_col_sortie=c(1:dim(table)[2])){
  res=matrix(ncol=length(nums_col_sortie));
  for (i in 1:dim(table)[1]){
    if (table[i,num_col]>=min_inclus & table[i,num_col]<max_exclus){
      res=rbind(res,table[i,nums_col_sortie]);
    }
  }
  res[-1,];
}

```

### **#fonction de repartition empirique**

```

fdr_emp=function(t,ech){
  res=c()
  for(i in 1:length(t)){
    res=c(res,length(ech[ech<=t[i]])/length(ech))
  }
  res
}

```

### **#val critique pour test Kolmogorov**

```

val_critique=function(ech,risque){
  if(length(ech)<=100){
    print("Regardez dans la table de Kolmogorov")
  }else{
    if(risque==0.01){1.629/sqrt(length(ech))}
    else{
      if(risque==0.05){1.358/sqrt(length(ech))}
      else{
        if(risque==0.1){1.223/sqrt(length(ech))}
        else{print("Choisissez un risque de niveau 1%, 5% ou 10%")}
      }
    }
  }
}

```

### **#PP-plot**

```

ppplot=function(ech,fdr,param,inf=-Inf,sup=+Inf,col){
  ech=sort(ech)
  plot(fdr(ech,param,inf,sup),fdr_emp(ech,ech),main="PP-plot",col=col)
  lines(seq(0,1),seq(0,1))
}

```

### **#test Kolmogorov**

```

kolmogorov=function(ech,fdr,param,inf=-Inf,sup=+Inf){
  ech=sort(ech)
  res=max(fdr_emp(ech[1],ech)-
fdr(ech[1],param,inf,sup),fdr(ech[1],param,inf,sup))
  for(i in 2:length(ech)){
    res=c(res,max(fdr_emp(ech[i],ech)-
fdr(ech[i],param,inf,sup),fdr(ech[i],param,inf,sup)-fdr_emp(ech[i-1],ech)))
  }
}

```

```
c(max(res),val_critique(ech,0.05),ifelse(max(res)>val_critique(ech,0.05),"r
ejet H0","ok"))
}
```

-----loi tronquee-----

### #densite

```
dens_tronquee=function(x,dens_complete,fdr_complete,param,inf=-
Inf,sup=+Inf){
res=c()
if(length(param)==1){
for(i in 1:length(x)){
if(x[i]>inf&x[i]<sup){
res=c(res,dens_complete(x[i],param)/(fdr_complete(sup,param)-
fdr_complete(inf,param)))
}else{res=c(res,0)}
}
}else{
for(i in 1:length(x)){
if(x[i]>inf&x[i]<sup){
res=c(res,dens_complete(x[i],param[1],param[2])/(fdr_complete(sup,param[1],
param[2])-fdr_complete(inf,param[1],param[2])))
}else{res=c(res,0)}
}
}
res
}
```

### #moins log-vraisemblance

```
moins_logvrais_tronquee=function(ech,dens_tronq,param,inf=-Inf,sup=+Inf){
-sum(log10(dens_tronq(echantillon,param,inf,sup)))
}
```

### #fonction de repartition

```
fdr_tronquee=function(t,fdr_complete,param,inf=-Inf,sup=+Inf){
res=c()
if(length(param)==1){
for(i in 1:length(t)){
if(t[i]<inf){res=c(res,0)}
else{
if(t[i]>=sup){res=c(res,1)}
else{
res=c(res,(fdr_complete(t[i],param)-
fdr_complete(inf,param))/(fdr_complete(sup,param)-fdr_complete(inf,param)))
}
}
}
}else{
for(i in 1:length(t)){
if(t[i]<inf){res=c(res,0)}
else{
if(t[i]>=sup){res=c(res,1)}
else{
res=c(res,(fdr_complete(t[i],param[1],param[2])-
fdr_complete(inf,param[1],param[2]))/(fdr_complete(sup,param[1],param[2])-
fdr_complete(inf,param[1],param[2])))
}
}
}
}
res
}
```

```

#simulation
simul_tronquee=function(taille,rloi_complete,param,inf=-Inf,sup=+Inf) {
if(length(param)==1) {
x=rloi_complete (2*taille,param)
z=x[(x>inf) & (x<sup)]
while(length(z)<taille) {
x=rloi_complete (2*taille,param)
z=c(z,x[(x>inf) & (x<sup)])
}
}else{
x=rloi_complete (2*taille,param[1],param[2])
z=x[(x>inf) & (x<sup)]
while(length(z)<taille) {
x=rloi_complete (2*taille,param[1],param[2])
z=c(z,x[(x>inf) & (x<sup)])
}
}
z[1:taille]
}

dgamma=function(x,param,inf=-Inf,sup=+Inf) {
dens_tronquee(x,dgamma,pgamma,param,inf,sup)
}

pgamma=function(t,param,inf=-Inf,sup=+Inf) {
fdr_tronquee(t,pgamma,param,inf,sup)
}

dweibull=function(x,param,inf=-Inf,sup=+Inf) {
dens_tronquee(x,dweibull,pweibull,param,inf,sup)
}

pweibull=function(t,param,inf=-Inf,sup=+Inf) {
fdr_tronquee(t,pweibull,param,inf,sup)
}

rweibull=function(taille,param,inf=-Inf,sup=+Inf) {
simul_tronquee(taille,rweibull,param,inf,sup)
}

dlnormt=function(x,param,inf=-Inf,sup=+Inf) {
dens_tronquee(x,dlnorm,plnorm,param,inf,sup)
}

plnormt=function(t,param,inf=-Inf,sup=+Inf) {
fdr_tronquee(t,plnorm,param,inf,sup)
}
rlnormt=function(taille,param,inf=-Inf,sup=+Inf) {
simul_tronquee(taille,rlnorm,param,inf,sup)
}

dexp=function(x,param,inf=-Inf,sup=+Inf) {
dens_tronquee(x,dexp,pexp,param,inf,sup)
}

pexp=function(t,param,inf=-Inf,sup=+Inf) {
fdr_tronquee(t,pexp,param,inf,sup)
}

dbetat=function(x,param,inf=-Inf,sup=+Inf) {

```

```

dens_tronquee(x,dbeta,pbeta,param,inf,sup)
}

pbetat=function(t,param,inf=-Inf,sup=+Inf){
fdr_tronquee(t,pbeta,param,inf,sup)
}

dpareto=function(x,a,alpha){
if(a>0&alpha>0){
res=c()
for(i in 1:length(x)){
if(x[i]>a){
res=c(res,alpha*(a^alpha)/(x[i]^(alpha+1)))
}else{res=c(res,0)}
}
res
}else{print("erreur : les coefficients doivent être strictement positifs")}
}

ppareto=function(t,a,alpha){
if(a>0&alpha>0){
res=c()
for(i in 1:length(t)){
if(t[i]>a){
res=c(res,1-(a/t[i])^alpha)
}else{res=c(res,0)}
}
res
}else{print("erreur : les coefficients doivent être strictement positifs")}
}

dparetot=function(x,param,inf=-Inf,sup=+Inf){
dens_tronquee(x,dpareto,ppareto,param,inf,sup)
}

pparetot=function(t,param,inf=-Inf,sup=+Inf){
fdr_tronquee(t,ppareto,param,inf,sup)
}

dlgamma=function(x,alpha,beta){
if(alpha>0&beta>0){
res=c()
for(i in 1:length(x)){
if(x[i]>1){
res=c(res,(beta^alpha)/gamma(alpha)*(log(x[i])^(alpha-1))/(x[i]^(beta+1)))
}else{res=c(res,0)}
}
res
}else{print("erreur : les coefficients doivent être strictement positifs")}
}

integrande=function(y,alpha){
res=c()
for(i in 1:length(y)){
res=c(res,y[i]^(alpha-1)*exp(-y[i]))
}
res
}

gamma_incomplete=function(z,inf=0,sup=+Inf){
res=c()

```

```

for(i in 1:length(z)){
res=c(res,integrate(integrande,inf,sup,100))
}
res
}

plgamma=function(t,alpha,beta){
if(alpha>0&beta>0){
res=c()
for(i in 1:length(t)){
if(t[i]>=1){
res=c(res,gamma_incomplete(alpha,sup=beta*log(t[i]))/gamma(alpha))
}else{res=c(res,0)}
}
res
}else{print("erreur : les coefficients doivent être strictement positifs")}
}

dlgamma=function(x,param,inf=-Inf,sup=+Inf){
dens_tronquee(x,dlgamma,plgamma,param,inf,sup)
}

plgamma=function(t,param,inf=-Inf,sup=+Inf){
fdr_tronquee(t,plgamma,param,inf,sup)
}

dnorm=function(x,param,inf=-Inf,sup=+Inf){
dens_tronquee(x,dnorm,pnorm,param,inf,sup)
}
pnorm=function(t,param,inf=-Inf,sup=+Inf){
fdr_tronquee(t,pnorm,param,inf,sup)
}

-----melange de lois-----
#loi tronquee + masse de Dirac

#moins log-vraisemblance
moins_logvrais_melange=function(param,ech,dens_tronq,inf=-Inf,sup=+Inf){
s=length(ech[ech<sup])*log10(1-
param[1])+length(ech[ech==sup])*log10(param[1])
if(length(param)==2){
for(i in 1:length(ech)){
if(ech[i]<sup){s=s+log10(dens_tronq(ech[i],param[2],sup=sup))}
}
}else{
for(i in 1:length(ech)){
if(ech[i]<sup){s=s+log10(dens_tronq(ech[i],c(param[2],param[3]),sup=sup))}
}
}
-s
}

#fonction de repartition
fdr_melange=function(t,fdr_tronq,param,inf=-Inf,sup=+Inf){
res=c()
if(length(param)==2){
for(i in 1:length(t)){
res=c(res,ifelse(t[i]<sup,(1-param[1])*fdr_tronq(t[i],
param[2],sup=sup),1))
}
}else{

```

```

for(i in 1:length(t)){
res=c(res,ifelse(t[i]<sup,(1-
param[1])*fdr_tronq(t[i],c(param[2],param[3]),sup=sup),1))
}
}
res
}

fdr_melange_gamma=function(t,param,inf=-Inf,sup=+Inf){
fdr_melange(t,pgamma, param, inf, sup)
}

fdr_melange_weibull=function(t,param,inf=-Inf,sup=+Inf){
fdr_melange(t,pweibull, param, inf, sup)
}

fdr_melange_lnorm=function(t,param,inf=-Inf,sup=+Inf){
fdr_melange(t,plnorm, param, inf, sup)
}

fdr_melange_exp=function(t,param,inf=-Inf,sup=+Inf){
fdr_melange(t,pexp, param, inf, sup)
}

fdr_melange_beta=function(t,param,inf=-Inf,sup=+Inf){
fdr_melange(t,pbet, param, inf, sup)
}

simul_melange=function(taille,rloi_complete,param,inf=-Inf,sup=+Inf){
if(length(param)==1){
x=rloi_complete(2*taille,param)
z=x[(x>inf)&(x<sup)]
while(length(z)<taille){
x=rloi_complete(2*taille,param)
z=c(z,x[(x>inf)&(x<sup)])
}
}else{
x=rloi_complete(2*taille,param[1],param[2])
z=x[(x>inf)&(x<sup)]
while(length(z)<taille){
x=rloi_complete(2*taille,param[1],param[2])
z=c(z,x[(x>inf)&(x<sup)])
}
}
z[1:taille]
}

simul_bornee=function(taille,rloi_complete,param,inf=-Inf,sup=+Inf){
if(length(param)==1){
x=max(inf,min(sup,rloi_complete(2*taille,param)))
while(length(x)<taille){
x=c(x,max(inf,min(sup,rloi_complete(2*taille,param))))
}
}else{
x=max(inf,min(sup,rloi_complete(2*taille,param[1],param[2])))
while(length(x)<taille){
x=c(x,max(inf,min(sup,rloi_complete(2*taille,param[1],param[2])))
}
}
x[1:taille]
}

```

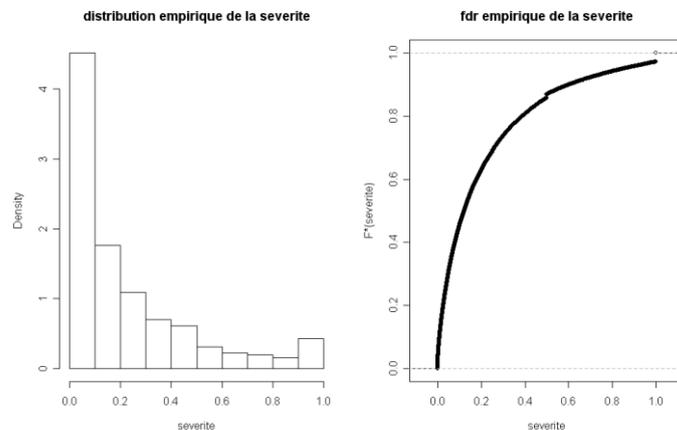
```
rlnormb=function(taille,param,inf=-Inf,sup=+Inf) {
simul_bornee(taille,rlnorm,param,inf,sup)
}
```

### #ajustement de la sévérité au global

```
donnees=read.table("C:\\Users\\Linh\\Desktop\\COFACE\\SEVERITE\\donneesR\\s
ev.txt",header=T)
```

```
sev=donnees[,1]
```

```
par(mfrow=c(1,2))
hist(sev,breaks=10,probability=T,main="distribution empirique de la
severite",xlab="severite")
plot(ecdf(sev),main="fdr empirique de la
severite",xlab="severite",ylab="F*(severite)")
```



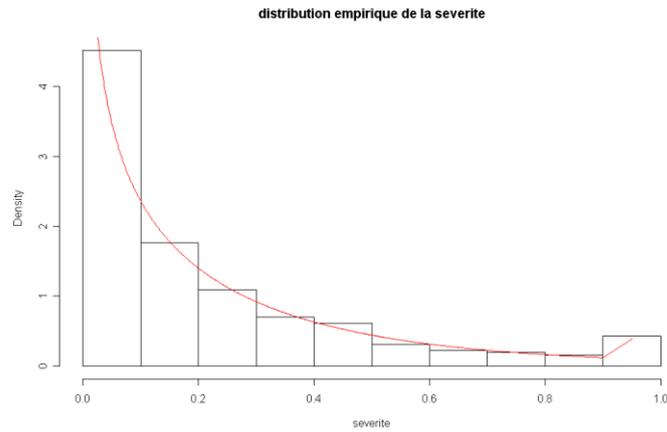
```
part1=sev[sev<1]
p1=length(part1)/length(sev)
part2=sev[(sev==1)]
p2=length(part2)/length(sev)
p1+p2          #1
p1              #0.9716683
p2              #0.02833172
```

```
x=seq(0,0.9,0.001)
```

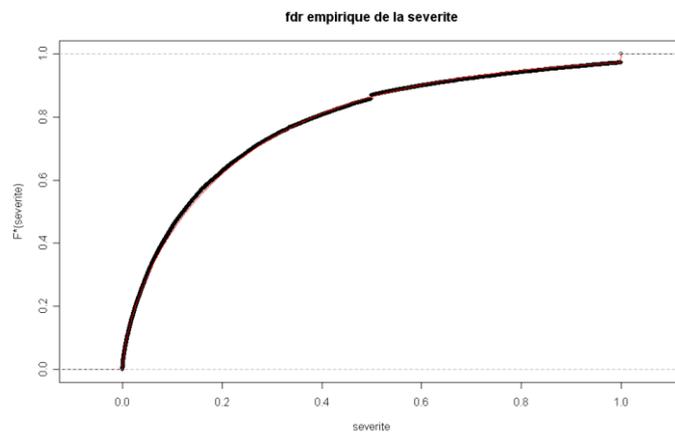
```
#-----
# melange gamma tronquee + masse de Dirac

#sortie_gamma=nlminb(c(p2,0.6,2.7),moins_logvrais_melange,ech=sev,dens_tron
q=dgammam,sup=1,lower=c(0,0,0),upper=c(1,+Inf,+Inf))
#sortie_gamma$par
#0.02833174 0.64035428 2.73929298
```

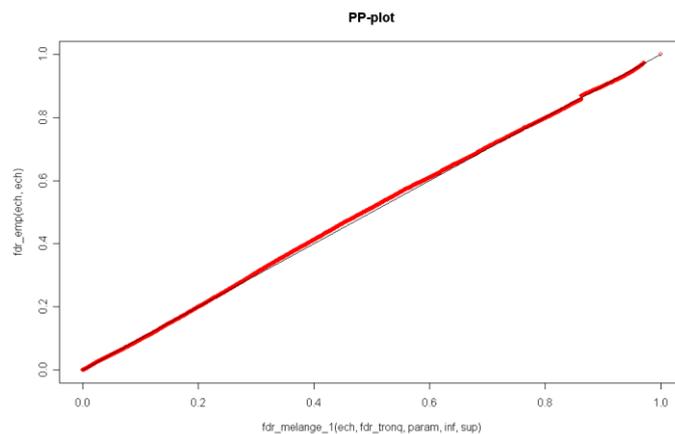
```
hist(sev,breaks=10,probability=T,main="distribution empirique de la
severite",xlab="severite")
lines(c(x,0.95),c(dgammam(x,c(0.64035428,2.73929298),sup=1)*(1-
0.02833174),pgammam(1,c(0.64035428,2.73929298),sup=1)-
pgammam(0.9,c(0.64035428,2.73929298),sup=1)+0.02833174)/0.1),col="red")
```



```
plot(ecdf(sev),main="fdr empirique de la
severite",xlab="severite",ylab="F*(severite)")
lines(seq(0,1,0.001),fdr_melange_gamma(seq(0,1,0.001),c(0.02833174,0.640354
28,2.73929298)),sup=1,col="red")
```



```
ppplot(sev,fdr_melange_gamma,c(0.02833174,0.64035428,2.73929298),sup=1,col=
"red")
```



```
#kolmogorov(sev,fdr_melange_gamma,c(0.02833174,0.64035428,2.73929298),sup=1
)
#"0.0283317227444435" "0.00636416464148603" "rejet H0"

#kolmogorov(part1,pgammat,c(0.64035428,2.73929298),sup=1)
#"0.0147308020528976" "0.00645628055535586" "rejet H0"
```

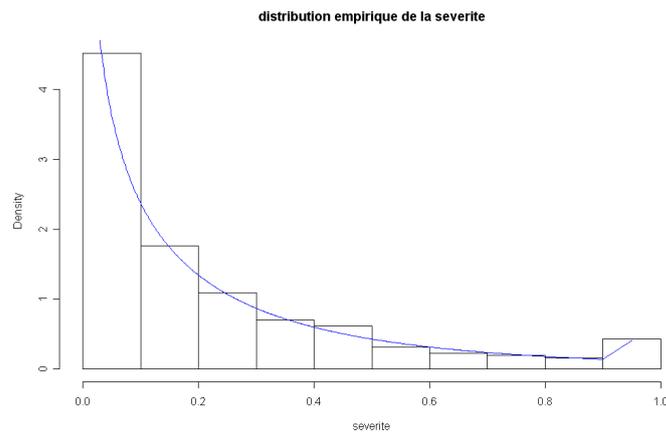
```

#-----
# melange weibull tronquee + masse de Dirac

#sortie_weibull=nlminb(c(p2,0.7,0.2),moins_logvrais_melange,ech=sev,dens_tr
ong=dweibullt,sup=1,lower=c(0,0,0),upper=c(1,+Inf,+Inf))
#sortie_weibull$par
#0.02833169 0.71574984 0.21403802

hist(sev,breaks=10,probability=T,main="distribution empirique de la
severite",xlab="severite")
lines(c(x,0.95),c(dweibullt(x,c(0.71574984,0.21403802),sup=1)*(1-
0.02833169),(pweibullt(1,c(0.71574984,0.21403802),sup=1)-
pweibullt(0.9,c(0.71574984,0.21403802),sup=1)+0.02833169)/0.1),col="blue")

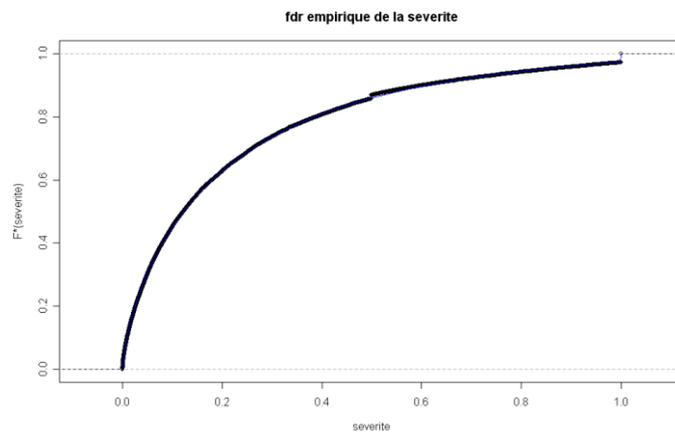
```



```

plot(ecdf(sev),main="fdr empirique de la
severite",xlab="severite",ylab="F*(severite)")
lines(seq(0,1,0.001),fdr_melange_weibull(seq(0,1,0.001),c(0.02833169,0.7157
4984,0.21403802),sup=1),col="blue")

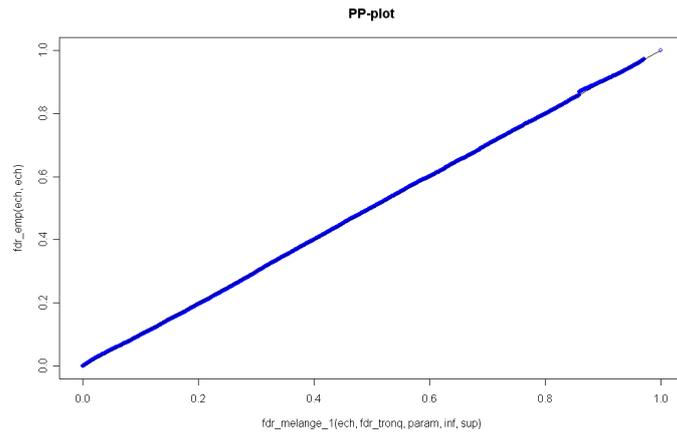
```



```

ppplot(sev,fdr_melange_weibull,c(0.02833169,0.71574984,0.21403802),sup=1,col="blue")

```



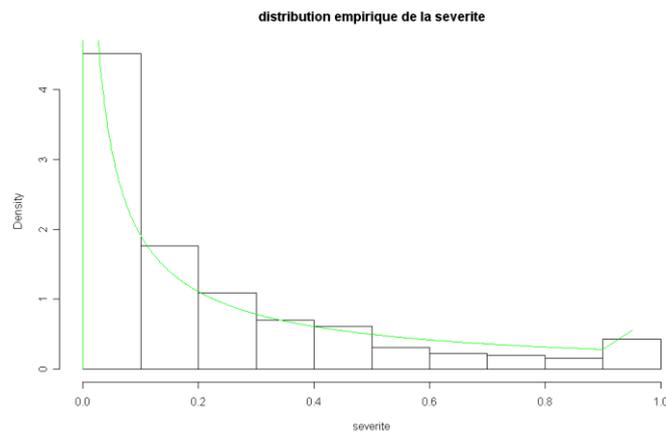
```
#kolmogorov_(sev, fdr_melange_weibull, c(0.02833169, 0.71574984, 0.21403802), sup=1)
"0.0283317227444435" "0.00636416464148603" "rejet H0"
```

```
kolmogorov(part1, pweibullt, c(0.71574984, 0.21403802), sup=1)
"0.0094549379086638" "0.0064562805535586" "rejet H0"
```

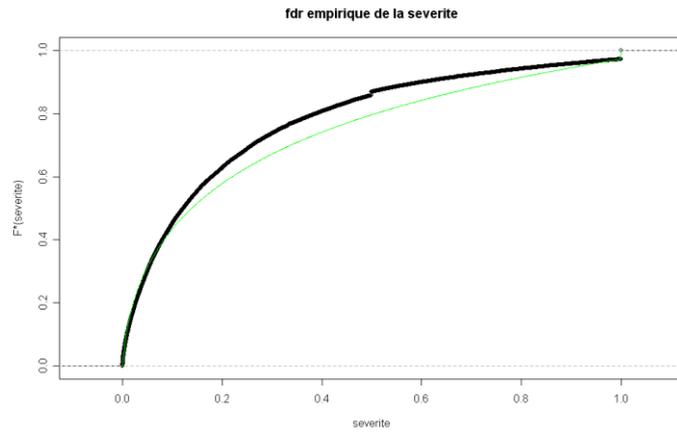
```
#-----
# melange lognormale tronquee + masse de Dirac

#sortie_lognormale=nlminb(c(p2, 0.6, 3.1), moins_logvrais_melange, ech=sev, dens
_tronq=dlnormt, sup=1, lower=c(0, 0.000001, 0), upper=c(1, +Inf, +Inf))
#sortie_lognormale$par
# 0.02833172 0.00000100 3.03192968
```

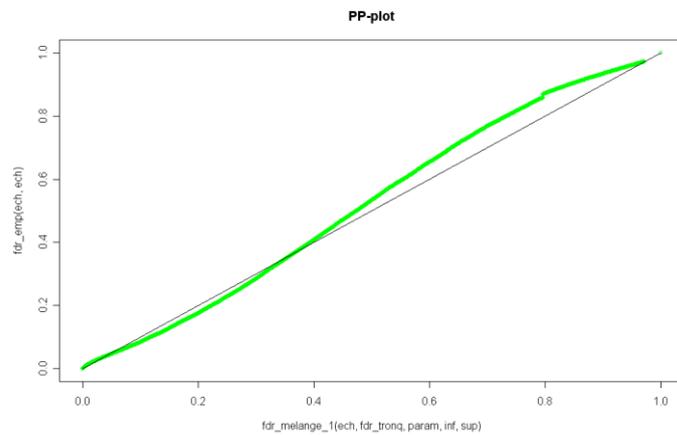
```
hist(sev, breaks=10, probability=T, main="distribution empirique de la
severite", xlab="severite")
lines(c(x, 0.95), c(dlnormt(x, c(0.000001, 3.03192968), sup=1) * (1 -
0.02833172), (plnormt(1, c(0.000001, 3.03192968), sup=1) -
plnormt(0.9, c(0.000001, 3.03192968), sup=1) + 0.02833172) / 0.1), col="green")
```



```
plot(ecdf(sev), main="fdr empirique de la
severite", xlab="severite", ylab="F*(severite)")
lines(seq(0, 1, 0.001), fdr_melange_lnorm(seq(0, 1, 0.001), c(0.02833172, 0.000001
00, 3.03192968), sup=1), col="green")
```



```
ppplot(sev, fdr_melange_lnorm, c(0.02833172, 0.00000100, 3.03192968), sup=1, col="green")
```



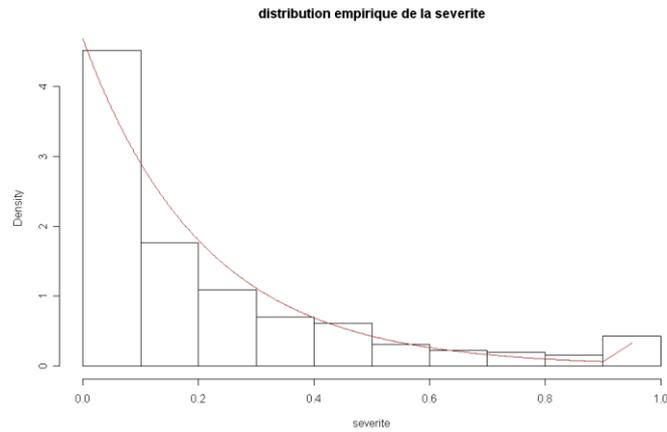
```
#kolmogorov(sev, fdr_melange_lnorm, c(0.02833172, 0.00000100, 3.03192968), sup=1)
#"0.0719580289506075" "0.00636416464148603" "rejet H0"
```

```
kolmogorov(part1, plnormt, c(0.00000100, 3.03192968), sup=1)
```

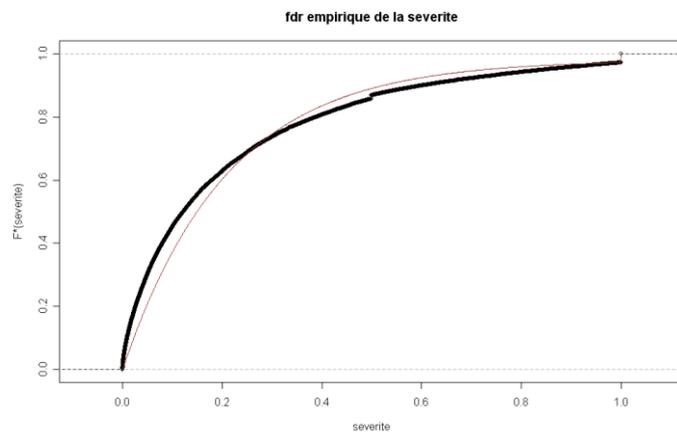
```
#-----
# melange exponentielle tronquee + masse de Dirac

#sortie_exponentielle=nlminb(c(p2, 6), moins_logvrais_melange, ech=sev, dens_tronq=dexpt, sup=1, lower=c(0, 0), upper=c(1, +Inf))
#sortie_exponentielle$par
#0.02833172 4.77928923
```

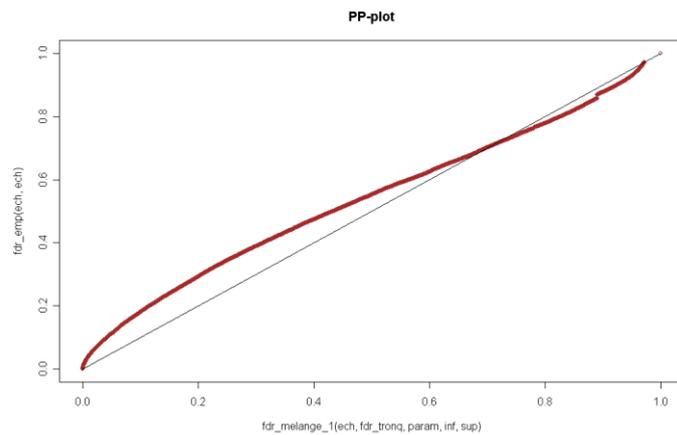
```
hist(sev, breaks=10, probability=T, main="distribution empirique de la severite", xlab="severite")
lines(c(x, 0.95), c(dexpt(x, 4.77928923, sup=1) * (1 - 0.02833172), (pexpt(1, 4.77928923, sup=1) - pexpt(0.9, 4.77928923, sup=1) + 0.02833172) / 0.1), col="brown")
```



```
plot(ecdf(sev),main="fdr empirique de la
severite",xlab="severite",ylab="F*(severite)")
lines(seq(0,1,0.001),fdr_melange_exp(seq(0,1,0.001),c(0.02833172,4.77928923
),sup=1),col="brown")
```



```
ppplot(sev,fdr_melange_exp,c(0.02833172,4.77928923),sup=1,col="brown")
```



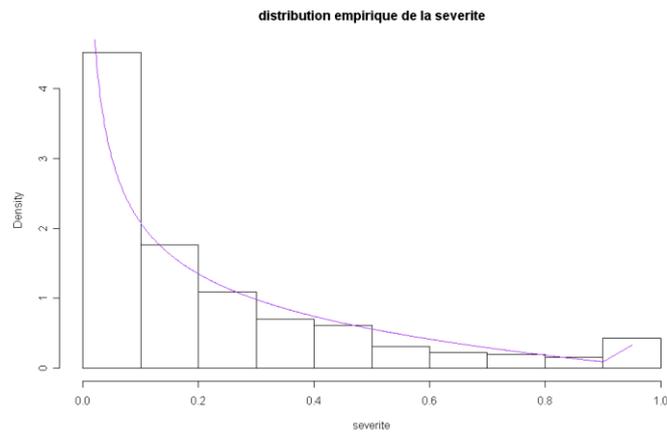
```
#kolmogorov(sev,fdr_melange_exp,c(0.02833172,4.77928923),sup=1)
#"0.0949412588035906" "0.00636416464148603" "rejet H0"
```

```
kolmogorov(part1,pexpt,4.77928923,sup=1)
```

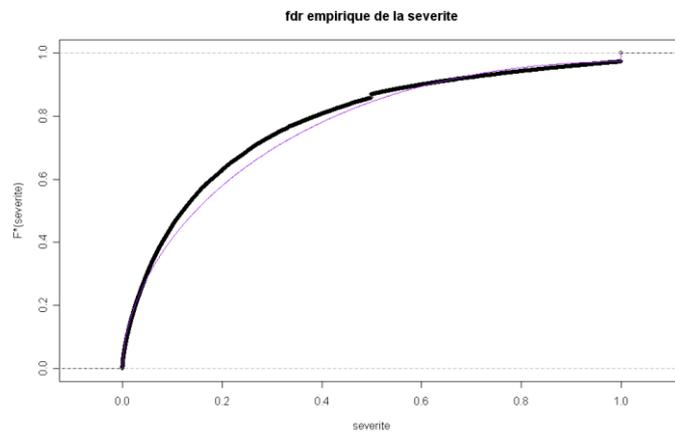
```
#-----
# melange beta tronquee + masse de Dirac
```

```
#sortie_beta=nlminb(c(p2,0.6,3.4),moins_logvrais_melange,ech=sev,dens_tronq
=dbetat,sup=1,lower=c(0,0,0),upper=c(1,+Inf,+Inf))
#sortie_beta$par
# 0.02833172 0.53856333 1.96382279
```

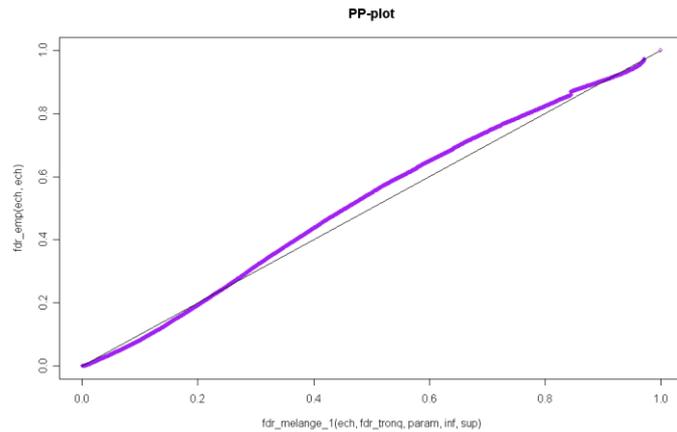
```
hist(sev,breaks=10,probability=T,main="distribution empirique de la
severite",xlab="severite")
lines(c(x,0.95),c(dbetat(x,c(0.53856333,1.96382279),sup=1)*(1-
0.02833172),(pbetat(1,c(0.53856333,1.96382279),sup=1)-
pbetat(0.9,c(0.53856333,1.96382279),sup=1)+0.02833172)/0.1),col="purple")
```



```
plot(ecdf(sev),main="fdr empirique de la
severite",xlab="severite",ylab="F*(severite)")
lines(seq(0,1,0.001),fdr_melange_beta(seq(0,1,0.001),pbetat,c(0.02833172,0.
53856333,1.96382279),sup=1),col="purple")
```



```
ppplot(sev,fdr_melange_beta,c(0.02833172,0.53856333,1.96382279),sup=1,col="
purple")
```



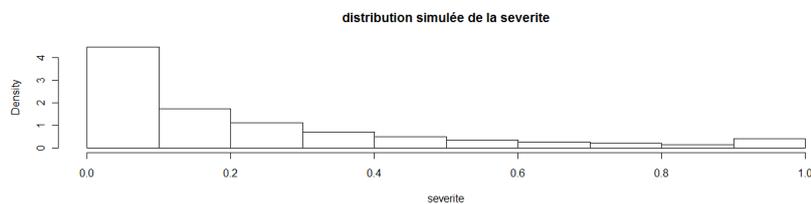
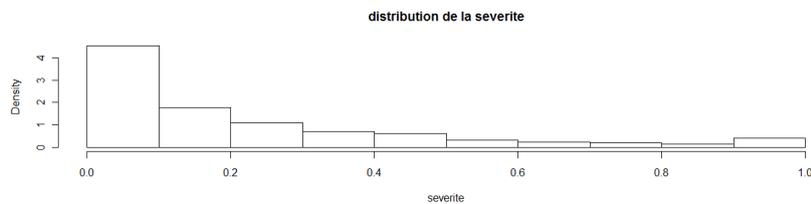
```
#kolmogorov(sev, fdr_melange_beta, c(0.02833172, 0.53856333, 1.96382279), sup=1)
"0.0505103184743694" "0.00636416464148603" "rejet H0"
```

```
kolmogorov(part1, pbetat, c(0.53856333, 1.96382279), sup=1)
```

```
-----

simul_sev_globale=function(taille,param){
B=rbinom(taille,1,param[1])
(B==1)*1+(B==0)*rweibullt(taille,c(param[2],param[3]),sup=1)
}
```

```
s_simul=simul_sev_globale(10000,c(0.02833169,0.71574984,0.21403802))
par(mfrow=c(2,1))
hist(sev,breaks=10,probability=T,main="distribution de la
severite",xlab="severite")
hist(s_simul,breaks=10,probability=T,main="distribution simulée de la
severite",xlab="severite")
```



```
#observations severite qd nbpol = 1
R1=read.table("C:\\Users\\Linh\\Desktop\\COFACE\\SEVERITE\\donneesR\\R1.txt",
header=T, sep="\t")
```

```
#observations severite qd nbpol = 2
R2=read.table("C:\\Users\\Linh\\Desktop\\COFACE\\SEVERITE\\donneesR\\R2.txt",
header=T, sep="\t")
```

```
#observations severite qd nbpol = 3
```

```

R3=read.table("C:\\Users\\Linh\\Desktop\\COFACE\\SEVERITE\\donneesR\\R3.txt",header=T,sep="\t")

#observations severite qd nbpol = 4
R4=read.table("C:\\Users\\Linh\\Desktop\\COFACE\\SEVERITE\\donneesR\\R4.txt",header=T,sep="\t")

#observations severite qd nbpol = 5
R5=read.table("C:\\Users\\Linh\\Desktop\\COFACE\\SEVERITE\\donneesR\\R5.txt",header=T,sep="\t")

#observations severite qd nbpol = 6
R6=read.table("C:\\Users\\Linh\\Desktop\\COFACE\\SEVERITE\\donneesR\\R6.txt",header=T,sep="\t")

#observations severite qd nbpol = 7
R7=read.table("C:\\Users\\Linh\\Desktop\\COFACE\\SEVERITE\\donneesR\\R7.txt",header=T,sep="\t")

#observations severite qd nbpol = 8
R8=read.table("C:\\Users\\Linh\\Desktop\\COFACE\\SEVERITE\\donneesR\\R8.txt",header=T,sep="\t")

#observations severite qd nbpol = 9
R9=read.table("C:\\Users\\Linh\\Desktop\\COFACE\\SEVERITE\\donneesR\\R9.txt",header=T,sep="\t")

#observations severite qd nbpol = 10
R10=read.table("C:\\Users\\Linh\\Desktop\\COFACE\\SEVERITE\\donneesR\\R10.txt",header=T,sep="\t")

r1=R1[,1]
r2=R2[,1]
r3=R3[,1]
r4=R4[,1]
r5=R5[,1]
r6=R6[,1]
r7=R7[,1]
r8=R8[,1]
r9=R9[,1]
r10=R10[,1]

#severite qd npol = 1 :
#taul=(1-B)X+B          X=LNt(m, sigma)          B=Bernouilli(q)

#estimation param
sortie_lnorm=nlm(moins_logvrais_melange,ech=r1,dens_trongq=dlnormt,sup=1,c(.1,-0.6,1.7))
#sortie_lnorm$estimate
#0.08225316 -0.54504305 1.68234128

q=0.08225316
m=-0.54504305
sigma=1.68234128
p=plnorm(1,m,sigma)

x=seq(0,0.99,0.01)
hist(r1,breaks=100,probability=T,main="distribution de la severite lorsque le nombre de polices = un",xlab="severite")

```

```

lines(c(x,0.995),c(dlnormt(x,c(-0.54504305,1.68234128),sup=1)*(1-
0.08225316), (plnormt(1,c(-0.54504305,1.68234128),sup=1)-plnormt(0.99,c(-
0.54504305,1.68234128),sup=1)+0.08225316)/0.01),col="red")

#fdr tau1
fdr_tau1=function(t,q,m,sigma){
fdr_melange_lnorm(t,c(q,m,sigma),sup=1)
}

plot(ecdf(r1),main="fdr empirique de la severite lorsque le nombre de
polices est egal a un",xlab="severite",ylab="F*(severite)")
lines(seq(0,1,0.001),fdr_tau1(seq(0,1,0.001),0.08225316,-
0.54504305,1.68234128),col="red")

#PP-plot tau1
ppplot_tau1=function(ech,q,m,sigma,col){
ppplot(ech,fdr_melange_lnorm,c(q,m,sigma),sup=1,col=col)
}

#ppplot_tau1(r1,0.08225316,-0.54504305,1.68234128,"red")

#fdr reciproque tau1
fdr_reciproque_tau1=function(u,q,m,sigma){
res=c()
for(i in 1:length(u)){
if(u[i]==0){
res=c(res,-Inf)
}else{
if((u[i]>0)&(u[i]<1-q)){
res=c(res,exp(sigma*qnorm(u[i]*plnorm(1,m,sigma)/(1-q))+m))
}else{
res=c(res,1)
}
}
}
res
}

#QQ-plot tau1
qqplot_tau1=function(ech,q,m,sigma,col){
ech=sort(ech)
u=seq(0,1,0.001)
plot(fdr_reciproque_tau1(u,q,m,sigma),quantile(ech,u),col=col)
lines(seq(0,1),seq(0,1))
}

#qqplot_tau1(r1,q,m,sigma,"red")

#kolmogorov(r1[r1<1],plnormt,c(m,sigma),sup=1)
#"0.0135269088245105" "0.0116778595030679" "rejet H0"

#E(X^n)
moment_lnormt=function(ordre,m,sigma){
exp(m*ordre+(ordre^2*sigma^2)/2)*pnorm(-m/sigma-
ordre*sigma,0,1)/plnorm(1,m,sigma)
}

#E(X)
mean_lnormt=function(m,sigma){
moment_lnormt(1,m,sigma)
}

```

```

#E(tau1)
mean_tau1=function(q,m,sigma){
  (1-q)*mean_lnormt(m,sigma)+q
}

mean_tau1(q,m,sigma)
#0.3868356

#V(X)
var_lnormt=function(m,sigma){
  moment_lnormt(2,m,sigma)-moment_lnormt(1,m,sigma)^2
}

#V(tau1)
var_tau1=function(q,m,sigma){
  (1-q)*moment_lnormt(2,m,sigma)+q-((1-q)*mean_lnormt(m,sigma)+q)^2
}

var_tau1(q,m,sigma)
#0.09944906

simul_tau1=function(taille,q,m,sigma){
  u=runif(taille)
  fdr_reciproque_tau1(u,q,m,sigma)
}

test=simul_tau1(10000,q,m,sigma)
par(mfrow=c(2,1))
hist(r1,breaks=100,probability=T,main="distribution observee")
hist(test,breaks=100,probability=T,main="distribution simulee")

```

```
#severite qd nb > 1
```

```

#o=1 → sev corr à 100% ; o="i" → sev indép
simul_sev=function(taille,nvar,q,m,sigma,o){
  if(o=="i"){
    matrix(simul_tau1(nvar*taille,q,m,sigma),ncol=nvar)
  }else{
    if(o==1){
      matrix(rep(simul_tau1(taille,q,m,sigma),nvar),ncol=nvar)
    }else{
      sigmaZ=matrix(c(rep(c(1,rep(o,nvar)),nvar-1),1),nrow=nvar)
      C=chol(sigmaZ)
      Z=t(C)%*%matrix(rnorm(nvar))
      U=pnorm(Z)
      A=fdr_reciproque_tau1(U,q,m,sigma)
      for(i in 1:(taille-1)){
        Z=t(C)%*%matrix(rnorm(nvar))
        U=pnorm(Z)
        A=rbind(A,fdr_reciproque_tau1(U,q,m,sigma))
      }
      A
    }
  }
}

```

```

#rho=1 → indic corr à 100% ; rho="i" → indic indép
simul_bernouilli=function(taille,nvar,p,rho){
  if(rho=="i"){

```

```

matrix(rbinom(nvar*taille,1,p),ncol=nvar)
}else{
  if(rho==1){
    matrix(rep(rbinom(taille,1,p),nvar),ncol=nvar)
  }else{
    sigmaZ=matrix(c(rep(c(1,rep(rho,nvar)),nvar-1),1),nrow=nvar)
    C=chol(sigmaZ)
    Z=t(t(C)%*%matrix(rnorm(nvar)))
    U=pnorm(Z)
    I=ifelse(U<=p,1,0)
    if(taille>1){
      for(i in 1:(taille-1)){
        Z=t(t(C)%*%matrix(rnorm(nvar)))
        U=pnorm(Z)
        I=rbind(I,ifelse(U<=p,1,0))
      }
    }
    I
  }
}

```

#type="e" → expo égales ; type="i" → expo indépendantes ;  
#type="p" → expo plafonnées

```

simul_expo=function(taille,nvar,p,rho,type){
  if(type=="i"){
    E=matrix(runif(nvar*taille),ncol=nvar)
    sum=0
    for(i in 1:nvar){
      sum=sum+E[,i]
    }
    E/sum
  }else{
    if(type=="e"){
      matrix(rep(1/nvar,nvar*taille),ncol=nvar)
    }else{
      I=simul_bernouilli(taille,nvar,p,rho)
      U=matrix(runif(nvar*taille),ncol=nvar)
      E=I+(1-I)*U
      sum=0
      for(i in 1:nvar){
        sum=sum+E[,i]
      }
      E/sum
    }
  }
}

```

```

a=0.35858022
#0.29200533
b=2.48838943
#2.20823466

```

```

tau=function(taille,q,m,sigma,nvar,o,p,rho,type,a,b){
  S=simul_sev(taille,nvar,q,m,sigma,o)
  W=simul_expo(taille,nvar,p,rho,type)
  alpha=rbeta(taille,a,b)
  res=c()
  for(i in 1:taille){
    j=1

```

```

sum=W[i,1]*S[i,1]
while (alpha[i]>j/nvar) {
sum=sum+W[i,j+1]*S[i,j+1]
j=j+1
}
res=c(res,sum)
}
res
}

#expo uniformes indépendantes, sev indépendantes
par(mfrow=c(2,1))
s21=tau(10000,q,m,sigma,2,o="i",type="i",a=a,b=b)
hist(r2,breaks=100,probability=T,main="distribution observee")
hist(s21,breaks=100,probability=T,main="distribution simulee")

#expo egales, sev indépendantes
par(mfrow=c(2,1))
s22=tau(10000,q,m,sigma,2,o="i",type="e",a=a,b=b)
hist(r2,breaks=100,probability=T,main="distribution observee")
hist(s22,breaks=100,probability=T,main="distribution simulee")

#expo egales ou uniformes independantes, sev indépendantes
par(mfrow=c(2,1))
s23=tau(10000,q,m,sigma,2,o="i",type="p",p=4776/9598,rho="i",a=a,b=b)
hist(r2,breaks=100,probability=T,main="distribution observee")
hist(s23,breaks=100,probability=T,main="distribution simulee")

#expo egales ou uniformes corrélées, sev indépendants
par(mfrow=c(2,1))
s24=tau(10000,q,m,sigma,2,o="i",type="p",p=4776/9598,rho=1,a=a,b=b)
hist(r2,breaks=100,probability=T,main="distribution observee")
hist(s24,breaks=100,probability=T,main="distribution simulee")

#expo egales ou uniformes corrélées, sev corrélées
par(mfrow=c(2,1))
s25=tau(10000,q,m,sigma,2,o=1,type="p",p=4776/9598,rho=1,a=a,b=b)
hist(r2,breaks=100,probability=T,main="distribution observee")
hist(s25,breaks=100,probability=T,main="distribution simulee")

par(mfrow=c(2,1))
s26=tau(10000,q,m,sigma,2,o=1,type="p",p=0.6,rho=0.85,a=a,b=b)
hist(r2,breaks=100,probability=T,main="distribution observee")
hist(s26,breaks=100,probability=T,main="distribution simulee")

par(mfrow=c(2,1))
s31=tau(10000,q,m,sigma,3,o="i",type="i",a=a,b=b)
hist(r3,breaks=100,probability=T,main="distribution observee")
hist(s31,breaks=100,probability=T,main="distribution simulee")

par(mfrow=c(2,1))
s32=tau(10000,q,m,sigma,3,o="i",type="e",a=a,b=b)
hist(r3,breaks=100,probability=T,main="distribution observee")
hist(s32,breaks=100,probability=T,main="distribution simulee")

par(mfrow=c(2,1))
s33=tau(10000,q,m,sigma,3,o="i",type="p",p=0.6,rho="i",a=a,b=b)
hist(r3,breaks=100,probability=T,main="distribution observee")
hist(s33,breaks=100,probability=T,main="distribution simulee")

par(mfrow=c(2,1))

```

```

s34=tau(10000,q,m,sigma,3,o="i",type="p",p=0.6,rho=1,a=a,b=b)
hist(r3,breaks=100,probability=T,main="distribution observee")
hist(s34,breaks=100,probability=T,main="distribution simulee")

par(mfrow=c(2,1))
s35=tau(10000,q,m,sigma,3,o="i",type="p",p=0.5,rho=0.2,a=a,b=b)
hist(r3,breaks=100,probability=T,main="distribution observee")
hist(s35,breaks=100,probability=T,main="distribution simulee")

par(mfrow=c(2,1))
s36=tau(10000,q,m,sigma,3,o=1,type="p",p=0.6,rho="i",a=a,b=b)
hist(r3,breaks=100,probability=T,main="distribution observee")
hist(s36,breaks=100,probability=T,main="distribution simulee")

par(mfrow=c(2,1))
s4=tau(10000,q,m,sigma,4,o=1,type="p",p=0.6,rho="i",a=a,b=b)
hist(r4,breaks=100,probability=T,main="distribution observee")
hist(s4,breaks=100,probability=T,main="distribution simulee")

par(mfrow=c(2,1))
s5=tau(10000,q,m,sigma,5,o=1,type="p",p=0.6,rho="i",a=a,b=b)
hist(r5,breaks=100,probability=T,main="distribution observee")
hist(s5,breaks=100,probability=T,main="distribution simulee")

par(mfrow=c(2,1))
s6=tau(10000,q,m,sigma,6,o=1,type="p",p=0.6,rho="i",a=a,b=b)
hist(r6,breaks=100,probability=T,main="distribution observee")
hist(s6,breaks=100,probability=T,main="distribution simulee")

par(mfrow=c(2,1))
s7=tau(10000,q,m,sigma,7,o=1,type="p",p=0.6,rho="i",a=a,b=b)
hist(r7,breaks=100,probability=T,main="distribution observee")
hist(s7,breaks=100,probability=T,main="distribution simulee")

par(mfrow=c(3,1))
s81=tau(10000,q,m,sigma,8,o=1,type="p",p=0.6,rho="i",a=a,b=b)
s82=tau(10000,q,m,sigma,8,o=1,type="i",p=0.6,rho="i",a=a,b=b)

hist(r8,breaks=100,probability=T,main="distribution observee")
hist(s81,breaks=100,probability=T,main="distribution simulee")
hist(s82,breaks=100,probability=T,main="distribution simulee")

par(mfrow=c(2,1))
s5=tau(10000,q,m,sigma,5,o=1,type="p",p=0.5,rho="i",a=a,b=b)
hist(r5,breaks=100,probability=T,main="distribution observee")
hist(s5,breaks=100,probability=T,main="distribution simulee")

#encours tirés,tau indépendants
tau2_6=function(taille,q,m,sigma,a=0){
s=simul_var_correlees(taille,q,m,sigma,a)
res=c()
for(i in 1:taille){
a1=sample(w1,1)
indic1=rbinom(1,1,pbeta(1,0.29200533,2.20823466)-
pbeta(0.5,0.29200533,2.20823466))
res=c(res,a1*s[1,i]+(1-a1)*s[2,i]*indic1)
}
res
}

```

```

par(mfrow=c(2,1))
s6=tau2_6(10000,q,m,sigma)
s6=tau2_6(10000,q,m,sigma,0.5)

hist(r2,breaks=100,probability=T,main="distribution observee")
hist(s6,breaks=100,probability=T,main="distribution simulee",xlim=c(0,1))

donnees=read.table("C:\\Users\\Linh\\Desktop\\COFACE\\SEVERITE\\donneesR\\p
ortefeuille.txt",header=T,sep="\t")

s=read.table("C:\\Users\\Linh\\Desktop\\COFACE\\SEVERITE\\donneesR\\sevemp.
txt",header=T)

expo=donnees[,1]
nbpol=donnees[,2]
sev_emp=s[,1]

perte_sev_emp=function(taille,PD){
res=c()
for(i in 1:taille){
res=c(res,sum(rbinom(length(expo),1,PD)*expo*sample(sev_emp,length(expo),re
p=T)))
}
res
}

perte_sev_mod=function(taille,PD){
res=c()
for(i in 1:taille){
sum=0
for(j in 1:length(expo)){
if(nbpol[j]==1){
sum=sum+rbinom(1,1,PD)*expo[j]* simul_tau1(1,q,m,sigma)
}else{
if(nbpol[j]==2){
sum=sum+rbinom(1,1,PD)*expo[j]*tau(1,q,m,sigma,2,o=1,type="p",p=0.6,rho=0.8
5,a=a,b=b)
}else{
if((nbpol[j]>=3)&(nbpol[j]<=5)){
sum=sum+rbinom(1,1,PD)*expo[j]*tau(1,q,m,sigma,nbpol[j],o=1,type="p",p=0.6,
rho="i",a=a,b=b)
}else{
sum=sum+rbinom(1,1,PD)*expo[j]*tau(1,q,m,sigma,nbpol[j],o=1,type="i",p=0.6,
rho="i",a=a,b=b)
}
}
}
}
res=c(res,sum)
}
res
}

perte_sev_moy=function(taille,PD){
res=c()
for(i in 1:taille){
res=c(res,sum(rbinom(length(expo),1,PD)*expo*0.22))
}
res
}

```

