



**HAL**  
open science

# Some statistical and game-theoretic models with an actuarial perspective

Christophe Dutang

► **To cite this version:**

Christophe Dutang. Some statistical and game-theoretic models with an actuarial perspective. Statistics [math.ST]. Université Paris Dauphine - PSL, 2021. tel-03581539

**HAL Id: tel-03581539**

**<https://theses.hal.science/tel-03581539>**

Submitted on 19 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paris-Dauphine, PSL University  
Centre de Recherche en Mathématiques de la Décision

Mémoire d'habilitation à diriger des recherches  
Spécialité : Mathématiques Appliquées  
présenté par  
Christophe DUTANG

---

**Some statistical and game-theoretic models  
with an actuarial perspective**

---

coordonné par :

M. Marc **HOFFMANN** Université Paris-Dauphine

soutenu publiquement le Vendredi 10 Décembre 2021 après avis de :

Mme Katrien **ANTONIO** Katholieke Universiteit Leuven  
M. Arthur **CHARPENTIER** Université du Québec à Montréal  
M. Michel **DENUIT** Université Catholique de Louvain

devant le jury composé de :

Mme Katrien <b>ANTONIO</b>	Katholieke Universiteit Leuven	Rapporteure
M. Arthur <b>CHARPENTIER</b>	Université du Québec à Montréal	Rapporteur
M. Michel <b>DENUIT</b>	Université Catholique de Louvain	Rapporteur
Mme Caroline <b>HILLAIRET</b>	Ecole Nationale de la Statistique et de l'Administration Economique	Examinatrice
M. Marc <b>HOFFMANN</b>	Université Paris-Dauphine	Coordinateur
M. Vincent <b>RIVOIRARD</b>	Université Paris-Dauphine	Examineur



# Remerciements

En premier lieu, je tiens à remercier mes rapporteurs : Katrien Antonio, Arthur Charpentier et Michel Denuit. Mon rapport d'habilitation n'étant pas spécialement court, sa relecture et l'écriture de leur rapport ont dû bien les occuper. En second lieu, je remercie également Marc Hoffmann qui a joué le rôle de coordinateur de mon habilitation au sein de Paris-Dauphine et qui m'a toujours fourni de très bons conseils. Enfin je suis également très reconnaissant à Caroline Hillairet et Vincent Rivoirard d'avoir accepté de faire partie du jury de soutenance.

La plupart de mes travaux n'ont pas été réalisés seul mais en collaboration avec d'autres chercheurs de différentes universités de France et d'ailleurs. C'est l'occasion de les remercier pour leur collaboration : j'ai beaucoup appris à leurs côtés. Notamment, je tiens à remercier Stéphane Loisel à Lyon, Hansjoerg Albrecher à Lausanne, Armelle Guillou à Strasbourg, Alexandre Brouste au Mans et Quentin Guibert à Paris.

Je suis aussi extrêmement reconnaissant à tout ceux qui ont relu tout ou partie de mon manuscrit, notamment, Marc, Stéphane, Hansjoerg et Quentin m'ont apporté de précieux conseils sur ce document. Merci à Benoit pour la relecture rigoureuse de mon introduction. Merci à Patrice pour sa relecture détaillée du manuscrit. L'écriture de ce rapport est l'occasion aussi de remercier tous les enseignants-chercheurs et les personnels administratifs de Dauphine. Paris-Dauphine propose un formidable environnement pour travailler qui sait s'adapter très rapidement à n'importe quelle situation, même en cas de crise sanitaire. Et même si j'ai pris des responsabilités de master et encadrer un grand nombre de mémoires d'actuels, le décompte des heures est largement en faveur des enseignants-chercheurs afin qu'ils aient du temps pour leur recherche.

Le mot de la fin pour ma femme Isabelle qui est mon soutien perpétuel, et mes enfants Alexandre, Maxime et Alice dont l'énergie débordante me remplit toujours de joie.



*À Pierre*



# Contents

<b>Remerciements</b>	<b>3</b>
<b>Contents</b>	<b>7</b>
<b>Résumé / Abstract</b>	<b>9</b>
<b>Introduction</b>	<b>11</b>
<b>1 Parametric regression models</b>	<b>17</b>
1.1 Generalized linear models . . . . .	17
1.2 A closed-form maximum likelihood estimator . . . . .	20
1.3 Closed-form likelihood-based estimators . . . . .	23
1.4 Generalized Linear Model Trees . . . . .	24
1.5 Other regression models . . . . .	27
1.6 R implementation . . . . .	31
<b>2 Parametric models without explanatory variables</b>	<b>33</b>
2.1 Independent and identically distributed models . . . . .	33
2.2 Loss and risk models in $\mathbb{R}$ . . . . .	34
2.3 Destruction rate models . . . . .	35
2.4 Le Cam's one-step estimation . . . . .	37
2.5 TL-Moment matching estimation . . . . .	38
<b>3 Univariate and bivariate extreme models</b>	<b>41</b>
3.1 Known results for univariate and multivariate models . . . . .	41
3.2 Estimation of the tail dependence coefficient . . . . .	45
3.3 Failure set estimation . . . . .	47



3.4	R implementation . . . . .	48
<b>4</b>	<b>Ruin theory and reserving methods</b>	<b>49</b>
4.1	A brief review of ruin theory . . . . .	49
4.2	A risk model with dependence . . . . .	51
4.3	Interconnections between reserving models and ruin theory . . . . .	53
4.4	R implementation . . . . .	59
<b>5</b>	<b>Customer models and non-cooperative games</b>	<b>61</b>
5.1	Customer behavior models . . . . .	61
5.2	Games literature review . . . . .	64
5.3	A static game for non-life insurance competition . . . . .	70
5.4	A one-period repeated game for non-life insurer strategies . . . . .	75
	<b>Bibliography</b>	<b>81</b>
<b>A</b>	<b>Additional examples of distributions in the exponential family</b>	<b>97</b>
A.1	Chapter 2 . . . . .	97

## Résumé

---

Ce rapport d'habilitation à diriger des recherches traite de modèles statistiques et de théorie des jeux ainsi que de leur usage en actuariat. Il se divise en 5 chapitres. Le chapitre 1 traite de l'estimation de modèles linéaires généralisés. Des formules explicites de l'estimateur de maximum de vraisemblance sont proposées dans le cas de variables explicatives catégorielles. Ces formules sont ensuite utilisées pour les arbres de régression et pour proposer des formules explicites dans un cadre plus général. Le chapitre 2 se concentre sur les modèles statistiques paramétriques sans variable explicative. Des lois paramétriques telle que la famille Feller-Pareto sont étudiées, implémentées dans le paquet **actuar** et mises en pratique via le paquet **fidistrplus**. Le chapitre traite aussi de méthodes alternatives à la méthode de maximum de vraisemblance tel que la méthode de Le Cam et la méthode des moments rognés linéairement (TL-moments). Le chapitre 3 étudie quelques modèles de théorie de valeurs extrêmes dans le cas bivarié. Un nouvel estimateur du coefficient de dépendance de la queue de distribution est proposé et possède des propriétés de robustesse et d'absence de biais. Cet estimateur permet de construire un estimateur de probabilité d'excès bivarié. Le chapitre 4 présente un modèle de ruine avec dépendance entre les montants ou les temps d'attente de sinistre pour lequel de nouvelles formules asymptotiques de probabilité de ruine sont proposées. Un autre modèle de théorie de la ruine est aussi proposé à des fins de provisionnement et propose une alternative aux méthodes traditionnelles. Le chapitre 5 propose des modèles statistiques de comportement client ainsi des jeux non-coopératifs pour les marchés d'assurance non-vie. Plusieurs modèles linéaires généralisés et modèles de survie de régression sont utilisés pour modéliser la résiliation, la conversion et le rachat de police. Enfin, un jeu statique et un jeu répété sont proposés pour modéliser la stratégie des assureurs et leurs propriétés théoriques sont étudiées.

---

Mots clés : *modèles linéaires généralisés, loi de Feller-Pareto, TL-moments, modèle de ruine avec dépendance, théorie des jeux.*

## Abstract

---

This habilitation thesis to supervise research deals with statistical and game-theoretic models and their use for actuarial science. It is composed of five chapters. Chapter 1 focuses on the estimation of generalized linear models. New closed-form solutions are proposed in the case of categorical explanatory variables. Based on them, an explicit fitting procedure is proposed regression trees as well as alternative estimators for generalized linear models. Chapter 2 is devoted to parametric statistical models without explanatory variables. Parametric distributions of the Feller-Pareto are studied and implemented in the R package **actuar** and fitted with R package **fidistrplus**. Le Cam's one-step estimation and trimmed-linear moments are also studied in this chapter. Chapter 3 is devoted to extreme values models in bivariate setting. A new estimator of the tail dependence coefficient that exhibits asymptotic unbiasedness and robustness is proposed. We then derive an estimator for small tail probabilities with similar properties. Chapter 4 deals with a ruin model integrating dependence among claim severity or claim waiting times for which asymptotics of ruin probability are derived. Then another risk process is used to compute reserve for non-life insurance. Chapter 5 is dedicated to statistical models of customer behaviors as well as non-cooperative game for non-life insurance markets. Generalized linear models and survival regression models are used for lapse, conversion and surrender of insurance policies. A static game and a repeated game are proposed to model insurer strategies and their theoretical properties are derived.

---

Keywords: *generalized linear models; Feller-Pareto distribution; TL-moments; ruin models with dependence, game theory.*



# Introduction

This manuscript presents my research activity since my Ph.D. defense in May 2012. It is composed of the research I have conducted at Université de Strasbourg between September 2012 and August 2013, at Le Mans Université between September 2013 and August 2017, and finally at Université Paris-Dauphine since September 2017.

Most of my research has been dedicated to actuarial science and related fields which aims to propose mathematical models in order to assess risk in insurance and finance. This document contains statistical models which are generally used for insurance pricing and reserving, probabilistic models devoted to the analysis long-term insolvency risk, known as ruin theory, as well as mathematical games modeling the behavior of insurers and policyholders. For a general introduction to actuarial science, we recommend, e.g., Fromenteau and Petauton (2017) for life insurance, Charpentier and Denuit (2004a,b) for non-life insurance and also the valuable Encyclopedia of Actuarial Science by Teugels and Sundt (2004). In my research, R (R Core Team, 2021) packages have been developed. Therefore, I take special care in this document to emphasize my contributions in that respect and refer to Charpentier (2014) for a detailed overview of actuarial science with R. This document is organized in five chapters described as follows.

## Chapter 1 – Parametric regression models

Chapter 1 deals with regression models used in statistics and in particular in actuarial science in many situations. Static regression models such as generalized linear models (GLM) are used in pricing of non-life insurance since the 1990s, where individual features of the policyholder and the insured good (house, car, ...) can be taken into account very easily. Typically, two regressions are performed: one for claim count and one for claim severity, e.g., Denuit et al. (2019a), Frees (2009), Jong and Heller (2008), and Ohlsson and Johansson (2010). GLM are also used for binary models such as assessing credit risk or nominal/ordinal choice models such as estimating deductible level (Lee, 2017). Another relevant application of GLM in actuarial science is their use for claim reserving by predicting incremental claim amounts, e.g., Wuethrich and Merz (2008).

Due to their efficiency property, generalized linear models are fitted by maximum likelihood estimation which generally does not have a closed-form expression and requires the use of numerical methods. On a sub class of models, where there is a single categorical explanatory variable or there are multiple crossed categorical variables, Brouste et al. (2020) provide a new closed-form solution of maximum likelihood estimators of GLM. There are two advantages with such a situation. First, closed-form estimators are far quicker to compute than using a numerical optimization method. Second, the exact distribution of closed-form estimators can be derived explicitly for many distributions allowing to quantify the bias of the MLE and to determine of exact confidence intervals.

Thanks to Brouste et al. (2020)'s result, Brouste et al. (2021a) propose an alternative closed-form estimator of GLM valid for any situation of categorical explanatory variables. In non-life insurance pricing, it is very common to discretize continuous variables such as policyholder age in order to make a rate table rather than a fully personalized premium. As the alternative closed-form estimator is not necessarily a maximum likelihood estimator, its asymptotic properties are studied in Brouste et al. (2021a).

Dutang and Guibert (2021) also benefit from Brouste et al. (2020) and put forward a new explicit split point procedure in model-based trees. Using the general framework of Zeileis et al. (2008), model-based partitioning trees can use non-Gaussian non-binary distribution for the response variable, generalizing CART models of (Breiman et al., 1984). The new procedure of Dutang and Guibert (2021) greatly accelerates the fitting procedure of generalized linear model-based trees and makes possible the use of GLM forests.

Chapter 1 ends up with a presentation of other regression models used in Chapter 5. Generalized Additive Models (GAM) (Hastie and Tibshirani, 1990) extends GLM by considering a smooth non-linear predictor to estimate the expectation of the response variable. GAM are also particularly used in non-life insurance since they avoid discretizing continuous variables (Denuit et al., 2019a; Denuit and Lang, 2004; Frees, 2009). Another class of regression models presented in this manuscript is the class of survival regression models which have obvious applications in life insurance but also for other topics.

## Chapter 2 – Parametric models without explanatory variables

Chapter 2 deals with estimation methods of parametric models without explanatory variables. Since the development of mathematical statistics, many estimation methods have been proposed in the literature to fit probability distributions to a sample, e.g., Casella and Lehmann (1998) and Shao (2003). The maximum likelihood estimation is probably the most used method due to its appealing property, but other concurrent methodologies exist such as moment matching estimation, quantile matching estimation, generalized method of moments. . . Delignette-Muller and Dutang (2015) details how the **fitdistrplus** package helps for fitting univariate distributions to different types of data (continuous censored or non-censored data and discrete data) via different estimation methods: maximum likelihood, moment matching, quantile matching and maximum goodness-of-fit estimation.

As in many other fields, fitting distribution plays a central role in actuarial science: typically the modeling of claim amount distributions for ratemaking, loss reserving in non-life insurance, e.g., (Charpentier and Denuit, 2004a; Frees, 2009) as well as modeling survival curves in life insurance, e.g., Dickson et al. (2013) and Macdonald et al. (2018). In R, thanks to Vincent Goulet, the **actuar** package is among the first to provide probability distributions. Dutang et al. (2008) make a comprehensive overview of **actuar** by presenting four usual topics of actuarial science: loss distributions modeling, risk theory, simulation of compound hierarchical models and credibility theory.

The **actuar** package is particularly used in actuarial science, for instance in recent contributions by Denuit (2020a,b), Papush et al. (2021), and Zariņa et al. (2021), but also it is used in other fields of statistics for its wide class of probability distribution, for example recent works by Carter et al. (2021), Mohamed Nor et al. (2021), and Wang et al. (2021a). Dutang (2014b) makes a review of loss models and associated parametric inference. Using both **fitdistrplus** and **actuar**, some applications on insurance datasets are provided in Dutang (2014b) and show the relevance of such approaches via classical measures of adequacy.

In many books, e.g., Hogg and Klugman (1984) and Klugman et al. (2012), distributions used for claim amounts come from the beta-transformed, the gamma-transformed or the Pareto families. However, the Feller–Pareto distribution popularized by Arnold (2015) is a more general class of positive distributions including beta-transformed and Pareto families which reveals relevant for loss modeling. Dutang et al. (2021) discuss the first implementation of support functions for the Feller–Pareto distribution and present solutions to deal with numerical difficulties associated with this distribution.

When modeling destruction rate models, a mixed-type distribution proposed by Bernegger (1997) called Maxwell Boltzmann Bose Einstein Fermi Dirac distribution (MBBEFD) is generally used for exposure rating. Dutang and Spedicato (2021b) study the convergence properties of the maximum likelihood estimator for one-inflated distributions and MBBEFD distribution. Applications are carried out thanks to the `mbbefd` package.

As an alternative to maximum likelihood estimation, Le Cam (1956) propose to use only a single step of Newton method on the log-likelihood estimation rather than the computation of a full sequence of optimized steps. Brouste and Dutang (2021b) implement Le Cam’s one-step estimator, which is asymptotically efficient and proves to be much faster than the maximum likelihood estimator.

Dutang (2017b) closes Chapter 2 by studying trimmed linear (TL) moments. New closed-form formulas of L-moments and TL-moments are derived for continuous probability distributions and apply it for the exponential and the uniform distributions. New formulas could be used in conjunction with moment matching estimation to provide reliable methods for heavy-tailed distributions for which ordinary moments do not exist.

### Chapter 3 – Univariate and bivariate extreme models

Chapter 3 is devoted to extreme value models, that is the estimation of the tail of univariate probability distributions or the estimation of the tail of a multivariate distribution function based on a random sample. Domains of applications of extreme value theory are numerous such as hydrology (e.g. flood discharge), meteorology (e.g. rainfall) and insurance (e.g. large claim), see Beirlant et al. (2004). Modeling the largest claim amounts both in frequency and in severity is needed for non-life (re)insurance pricing, e.g., Albrecher et al. (2017), Cebrián et al. (2003), and Charpentier and Denuit (2004b).

Univariate extreme value models have been well studied in the literature since the Fisher-Gnedenko-Tippett theorem (Fisher and Tippett, 1928; Gnedenko, 1943). However, multivariate extreme value theory is still growing with various approaches to deal with the tail estimation of multivariate distributions. In this document, we consider only bivariate models. Dutang et al. (2014) introduce a robust and asymptotically unbiased estimator for the coefficient of tail dependence in bivariate extreme value statistics. In a bivariate framework, the estimator is obtained by fitting a second-order model to the data by means of the minimum density power divergence criterion. In addition a simulation analysis to assess the estimation uncertainty, this model is then applied to losses due to permanent/partial disability claims for workers compensation insurance in North America.

Pursuing this work, Dutang et al. (2016) propose a bias-corrected and robust estimator for small tail probabilities based again on a second-order model. After a unit-Pareto transformation, bivariate observations are used to estimate the tail probability again by means of the minimum density power divergence technique. A numerical illustration aims to assess the ruin probability of a non-life insurer operating in Australia through the estimation of a high loss ratio and a high expense ratio.

Chapter 3 ends up with a review of R packages dedicated to extreme value statistics referenced by Dutang and Jaunatre (2020). Following reviews of Gilleland et al. (2013) and Stephenson and Gilleland (2006), Raymond-Belzile et al. (2021) survey recent development in software implementation of extreme methods with a view towards numerical challenges arising when performing extreme value analyses.

## Chapter 4 – Ruin theory and reserving methods

Chapter 4 focuses on ruin theory and reserving methods, which both study to the long-term commitment of insurance carriers to fulfill their obligation, but with different approaches. Ruin theory aims to assess the long-term ruin (or by complement solvency) probability of an insurer underwriting non-life policies and indemnify the corresponding claims. A central question is how to determine the initial capital at the origin to mitigate the ruin probability, e.g., Asmussen and Albrecher (2010).

Claim reserving methods are statistical and probabilistic models dedicated to the study of long-term management of claims. As soon as insurance products are underwritten, some events raise insured losses which subsequently become claims for the insurer. For multiple reasons, even with modern computers and internet, delays may occur at the different times of a claim management. Typically, there might be delays in noticing damages by the policyholder, delays in reporting loss to insurers, delays in assessing the final claim amount. Obviously, claim reserving is mandatory in most insurance markets and follows a guideline depending on the line of business, e.g., Wuehrich and Merz (2008).

Firstly, Dutang et al. (2013b) investigate a wide class of dependent risk processes, in continuous or discrete time in order to point out that an asymptotic rule  $A + B/u$  for the ultimate ruin probability applies. That dependence is incorporated through a mixing model in the individual claim amount distributions. Several special mixing distributions are examined in detail and some close-form formulas and asymptotics of the ruin probability are derived. Claim tail distributions and the dependence structure are also investigated. Avram et al. (2014) explore some particular aspects and recent results in ruin theory.

Secondly, Dutang and Brouste (2016) present an efficient way to compute all the key indicators in an unified approach of the ruin theory and claim reserving methods. The proposed framework allows to derive closed-form formulas for both ruin theory and claim reserves indicators. Chapter 4 ends up with a discussion on R packages specialized for claim reserving and ruin theory. As already mentioned, **actuar** is a pioneer package in providing ruin-related functions, whereas **ChainLadder** (Gesmann et al., 2021) provides many functions for reserving methods.

## Chapter 5 – Customer models and non-cooperative games for insurance markets

Chapter 5 presents customer behavior models and theoretical games for insurer strategies. Both topics are necessarily discussed in actuarial teaching programs or in actuarial textbooks, but they are of importance for insurers. Customer modeling is an area of applied statistics and econometrics aiming to model customer decisions, e.g., Gupta et al. (2006). In the insurance industry, this consists mainly in analyzing policyholder's decision to convert or not its policy offer, to renew or to lapse its policy renewal in non-life insurance. For life insurance, the decision by a policyholder to terminate its contract is known as surrender.

Based on policyholder models, numerous game-theoretic models have been proposed in the literature dating back to Borch (1960a). For instance, cooperative game theory is used to model risk transfer between a group of insured/insurers or between an insured and the insurer, e.g., Lemaire (1991); or allocation of risk-based capital among lines of business, e.g., Panjer (2002). Non-cooperative game theory help in pricing insurance market competition or in supervising an insurance market, e.g., Rees et al. (1999).

Regarding customer modeling for non-life insurance, Dutang (2012b) tackles the issue of price elasticity from various points of view: we focus on price elasticity of different markets, check the impact of distribution channels, investigate the use of market proxies and test for evidence of adverse selection. Dutang and Spedicato (2018) explore the applicability of recent machine learning techniques such as boosted tree models and neural networks in order to optimize the proposed premium on prospective policyholders. Given their predictive gain over generalized linear models, they carefully analyse both the advantages and disadvantages induced by their use.

Focusing on surrender in life insurance, Dutang and Milhaud (2018) deals with the challenging problem of modeling policyholders' behaviors in life insurance by considering the Fine & Gray model, a semi-parametric regression model presented in Chapter 1. This framework is quite efficient and recovers the empirical lapse rate trajectory by aggregating individual predicted lifetimes and particularly useful to design future insurance product.

Chapter 5 continues with a brief literature review of game theory and its use for actuarial topics. A review of R packages implementing game-theoretic concepts is also presented. Then, based on Dutang (2012b) and Dutang and Spedicato (2018)'s contributions of customer models, Dutang et al. (2013a) formulate a new non-cooperative game to model competition for policyholders among non-life insurance companies, taking customer behavior into account, market premium, solvency level, market share and underwriting results. Based on Nash equilibria and Stackelberg equilibria, the proposed game shows that the ability of an insurer to sell contracts is essential for its survival. Dutang (2013b) studies theorems guaranteeing existence of (generalized) Nash equilibria and analyze the assumptions on practical parametric feasible sets.

Since the static nature of Dutang et al. (2013a)'s insurance game is not realistic, Dutang (2014a) propose a repeated one-period game which analyze through Monte-Carlo simulations. Mouminoux et al. (2021b) carefully study the repeated one-period game of Dutang (2014a) based on Dutang et al. (2013a) and investigate convergence properties of the market using a Markov chain approach. Deviations from a regulated market situation are analyzed to provide some insights on the effect of competition in the repeated game.





# Chapter 1

## Parametric regression models

In this chapter, we present our contributions to parametric regression models. Section 1.1 presents Generalized Linear Models (GLM) for which closed-form estimators are proposed in Sections 1.2 and 1.3. Applications to non-life insurance are also proposed in Sections 1.2 and 1.3. Section 1.5 introduces survival regression models and generalized additive models applied in Section 5.1. Finally, Section 1.6 makes a review of existing R packages available for GLM and position our contribution in that respect.

### Notation

For the sake of clarity, bold notations are reserved for vector of  $\mathbb{R}^p$  and  $\mathbb{R}^n$  where  $n$  is the observation number and  $p$  is the number of explanatory variables or the number of dummies. We consider deterministic exogenous variables  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , with  $\mathbf{x}_i = (x_i^{(1)}, \dots, x_i^{(p)}) \in \mathbb{R}^p$  for  $i = 1, \dots, n$ . The index  $i \in I = \{1, \dots, n\}$  is reserved for the observations, while the indexes  $j, k, l$  are used for the explanatory variables. Response variables are denoted by  $Y_i$  and the sample by  $\mathbf{Y} = (Y_1, \dots, Y_n)$ .

### 1.1 Generalized linear models

The assumption of identical distributions for random variables in an observation sample is relaxed for regression models by considering explanatory variables. GLM were introduced by Nelder and Wedderburn (1972) and popularized in McCullagh and Nelder (1989). GLM rely on probability distribution functions of exponential type for the response variable which include most of the light and medium tailed distributions (such as normal, gamma or inverse Gaussian).

Precisely, the likelihood  $L$  associated to the statistical experiment generated by  $Y_i$ ,  $i \in I$ , verifies

$$\log L(\boldsymbol{\theta} | \mathbf{y}_i) = \frac{\lambda_i(\boldsymbol{\theta})y_i - b(\lambda_i(\boldsymbol{\theta}))}{a(\phi)} + c(y_i, \phi), \quad y_i \in \mathbb{Y} \subset \mathbb{R}, \quad (1.1)$$

and  $-\infty$  if  $y_i \notin \mathbb{Y}$ , where  $a : \mathbb{R} \rightarrow \mathbb{R}$ ,  $b : \Lambda \rightarrow \mathbb{R}$  and  $c : \mathbb{Y} \times \mathbb{R} \rightarrow \mathbb{R}$  are known real-valued measurable functions and  $\phi$  is the dispersion parameter, e.g., McCullagh and Nelder (1989, Section 2.2). Table 1.1 gives classic examples of probability distribution in the exponential family characterized by  $a$ ,  $b$ ,  $c$  and  $\mathbb{Y}$ .

In Equation (1.1), the parameters  $\lambda_1, \dots, \lambda_n$  depend on a finite-dimensional parameter  $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$  through the expectation of the response variable. That is we assume

$$b'(\lambda_i(\boldsymbol{\theta})) = E(Y_i) \quad \text{and} \quad b''(\lambda_i(\boldsymbol{\theta}))a(\phi) = \text{Var}(Y_i). \quad (1.2)$$

Distribution	$\lambda$	$\phi$	$a(x)$	$b(x)$	$c(x, \phi)$	$b'(x)$	$\bar{b}(x)$	$b(\bar{b}(x))$	$V(x)$
Bernoulli $\mathcal{B}(p)$	$\log(\frac{p}{1-p})$	1	$x$	$\log(1 + e^x)$	0	$\frac{e^x}{1+e^x}$	$\log(\frac{x}{1-x})$	$-\log(1+x)$	$x(1-x)$
Binomial $\mathcal{B}(m, p)$	$\log(\frac{p}{1-p})$	$\frac{1}{m}$	$x$	$\log(1 + e^x)$	$\log\left(\frac{1/\phi}{x/\phi}\right)$	$\frac{e^x}{1+e^x}$	$\log(\frac{x}{1-x})$	$-\log(1+x)$	$x(1-x)$
Gaussian $\mathcal{N}(\mu, \sigma^2)$	$\mu$	$\sigma^2$	$x$	$x^2/2$	$\frac{x^2/\phi}{-\frac{1}{2}\log(2\pi\phi)}$	$x$	$x$	$x^2/2$	1
Gamma $\mathcal{G}(\nu, \mu)$	$\frac{-1}{\mu}$	$1/\nu$	$x$	$-\log(-x)$	$\frac{\log(x/\phi) - \log(x)}{-\log(\Gamma(\frac{1}{\phi}))}$	$-1/x$	$-1/x$	$\log(x)$	$x^2$
Poisson $\mathcal{P}(\mu)$	$\log(\mu)$	1	$x$	$e^x$	$-\log(x!)$	$e^x$	$\log(x)$	$x$	$x$
Inv. Gauss. $\mathcal{IG}(\mu, \sigma^2)$	$-1/(2\mu^2)$	$1/\sigma^2$	$x$	$-\sqrt{-2x}$	$\frac{-\frac{1}{2}\log(2\pi\phi x^3)}{-1/(2\phi x)}$	$1/\sqrt{-2x}$	$-1/(2x^2)$	$-1/x$	$x^3$

Table 1.1 – Usual distributions in the exponential family

Using a twice continuously differentiable and bijective function  $g$  from  $b'(\Lambda)$  to  $\mathbb{R}$ , the GLM are defined by assuming the following relation between the expectation  $E(Y_i)$  (1.2) and the predictor

$$g(b'(\lambda_i(\boldsymbol{\theta}))) = \langle \mathbf{x}_i, \boldsymbol{\theta} \rangle = \eta_i, \quad \text{for all } \boldsymbol{\theta} \in \Theta, \quad (1.3)$$

where  $\eta_i$  are the linear predictors and  $\langle \cdot, \cdot \rangle$  denotes the scalar product.  $g$  is called the link function in the regression framework. In other words, the bijective function  $\ell = (b')^{-1} \circ g^{-1}$  is setted; then we have

$$\lambda_i(\boldsymbol{\theta}) = \ell(\eta_i). \quad (1.4)$$

We summarize with the following relations

$$X \times \Theta \xrightarrow{\langle \cdot, \cdot \rangle} D \xrightleftharpoons[\ell]{\ell^{-1}} \Lambda,$$

where  $D$  is the space of linear predictor and  $X$  the possible set of value of  $\mathbf{x}_i$  for  $i \in I$ . Here  $\ell$  is chosen and, consecutively  $\Theta$ ,  $\Lambda$  and  $D$  must be set\*.

The parameter  $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$  is to be estimated. Let us compute the log-likelihood of  $\mathbf{y} = (y_1, \dots, y_n)$ :

$$\log L(\boldsymbol{\theta} | \mathbf{y}) = \sum_{i=1}^n \frac{y_i \ell(\eta_i) - b(\ell(\eta_i))}{a(\phi)} + \sum_{i=1}^n c(y_i, \phi), \quad (1.5)$$

with  $b$ ,  $h$  and  $\ell$  being respectively defined in (1.1) and (1.4). If the model is identifiable, it can be shown that the sequence of MLE  $(\hat{\boldsymbol{\theta}}_n)_{n \geq 1}$  defined by  $\hat{\boldsymbol{\theta}}_n = \arg \max_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta} | \mathbf{y})$  asymptotically exists and is consistent (e.g. Fahrmeir and Kaufmann, 1985, Theorems 2 and 4).

The MLE  $\hat{\boldsymbol{\theta}}_n$ , if it exists, is the solution of the non linear system

$$S_j(\boldsymbol{\theta}) = 0, \quad j = 1, \dots, p, \quad (1.6)$$

with  $S_j(\boldsymbol{\theta})$  are the component of the score vector defined by

$$S_j(\boldsymbol{\theta}) = \frac{1}{a(\phi)} \sum_{i=1}^n x_i^{(j)} \ell'(\eta_i) (y_i - b'(\ell(\eta_i))).$$

It is worth mentioning that for a small data set or large number of explanatory variables, the existence of the MLE is not guaranteed. Note that the MLE  $\hat{\boldsymbol{\theta}}_n$  does not depend on the value of the dispersion

---

\*We talk of canonical link function, when  $\ell$  is the identity function. By default, softwares such as R use canonical links  $g = b'$ . Otherwise we use link functions preserving the parameter domain  $\Theta$  such as  $\log(x)$ .

parameter  $\phi$ . Indeed, the dispersion parameter is estimated in a second step using the sum of square residuals or the log-likelihood (McCullagh and Nelder, 1989, e.g. Chap. 9). Model selection can be performed through various metrics such as the deviance, the Aikake's Information criterion as well as chi-square statistical tests (McCullagh and Nelder, 1989, e.g. Chap. 2). Variable selection is carried out via statistical tests using the Wald statistics generally or via penalization procedures such as elastic net regression (Denuit et al., 2019a, e.g. Chap. 4).

In a general setting, the system (1.6) does not have a closed-form solution and GLM are generally fitted using a Newton-type method, such as an Iteratively re-Weighted Least Square (IWLS or IRLS) algorithm. McCullagh and Nelder (1989, Section 2.5) propose a quick overview of the IWLS algorithm which consists in using the expected Hessian matrix instead of the true Hessian when searching for the next iterate of the Newton method. There are two arguments in favor of using this trick: the fact that the expected Hessian is negative definite under regularity conditions on  $b$  and  $g$ ; two out of three terms cancel out in the expected Hessian. This variation was first proposed by Fisher in the context of probit regression and thus also known as the Fisher Scoring algorithm. The IWLS algorithm is typically used in statistical softwares such as R. Recently, Dutang (2017a) proposes further practical explanations of IWLS such as the initial guess of  $\theta$ , the working weights and responses, the globalization scheme.

### 1.1.1 Logistic regression

We focus only on binary regression, where the response variable is either 1 or 0, respectively for success and failure. Writing the mass probability function as  $f_{Y_i}(y) = \pi_i^y(1 - \pi_i)^{1-y}$  emphasizes the exponential family feature. So the link function for a binary model is expressed as follows

$$\pi_i = g^{-1}(\langle \mathbf{x}_i, \boldsymbol{\theta} \rangle).$$

As indicating in Fox (2010), the link function and the response variable can be reformulated as an unobserved variable  $\pi_i = P(Y_i = 1) = P(\mathbf{x}_i^T \boldsymbol{\beta} - \epsilon_i > 0)$ . If  $\epsilon_i$  follows one of the distribution of Table 1.2, then  $\pi_i$  uses the corresponding distribution function  $g^{-1}$ . In addition of being the canonical link function, the logit link is generally preferred because of its simple interpretation as the logarithm of the odds ratio and a faster fitting time.

Link	$g(x)$	$g^{-1}$	associated distribution
logit	$\ln\left(\frac{x}{1-x}\right)$	$\frac{1}{1+e^{-x}}$	standard logistic
probit	$\Phi^{-1}(x)$	$\Phi(x)$	standard normal
complementary log-log	$\ln(-\ln(1-x))$	$1 - e^{-e^x}$	standard Gumbel II

Table 1.2 – Common link functions for logistic regression

### 1.1.2 Count regression

Count regression consists in modeling a response variable belonging to  $\mathbb{N}$ . The most common distribution in the exponential family is the Poisson distribution with the density  $f_Y(y) = \lambda^y e^{-\lambda} / (y!)$ , see Table 1.1.

In practice, count regressions need to take exposures into account. For non-life insurance, exposure is measured as the percentage of year where the property (vehicle, house, . . .) is insured by the insurer. For the  $i$ th policy,  $e_i$  is the observed exposure such that  $e_i = 1$  indicates a fully observed risk over a

year. The logarithm of  $e_i$  is used in the linear predictor as an offset. Indeed assuming the claim count  $Y_i$  follows a Poisson distribution  $\mathcal{P}(e_i e^{\eta_i})$ , the expectation is given by

$$\ln(E(Y_i)) = \ln(e_i) + \eta_i = 1 \times \ln(e_i) + \langle \mathbf{x}_i, \boldsymbol{\theta} \rangle.$$

More advanced models for count regression use the negative binomial distribution or zero-inflated or zero-modified distributions, see, e.g., Denuit et al. (2007). There is a trick to use the IWLS algorithm for the negative binomial distribution by considering a parametrization through its expectation for a fixed dispersion parameter and by fitting its dispersion parameter in a second step. Zero-inflated or zero-modified distributions do not enter in the GLM framework and require a full likelihood maximization.

### 1.1.3 Severity regression

Count regression consists in modeling a response variable belonging to  $(0, +\infty)$ . Common distribution in the exponential family are the gamma and the inverse Gaussian distributions, see Table 1.1. The IWLS algorithm works correctly for these distributions.

Another type of models is to model the log claim amount  $\ln(Y)$  which allow to use other distributions of the exponential family. Typically, one can deal with a lognormal model if  $\ln(Y_i)$  follows a Gaussian distribution  $\mathcal{N}(\mu_i, \sigma^2)$ . The fitting procedure consists in applying a GLM on the log claim amount  $\ln(Y)$ . However care is needed for moments prediction as  $E(Y) \geq \exp(E(\ln(Y)))$ . Indeed, we have

$$E[Y_i] = \exp(\mu_i + \sigma^2/2), \text{Var}[Y_i] = \exp(2\mu_i + \sigma^2)(\exp(\sigma^2) - 1).$$

A similar procedure can be carried out for the Pareto 1 by assuming  $\ln(Y_i)$  follows an exponential distribution  $\mathcal{E}(\beta_i)$ . One can even consider a gamma distribution for the log claim amount  $\ln(Y)$ . It is equivalent to assuming a log-gamma density

$$f_Y(x) = \left(\frac{\nu}{\mu_i}\right)^\nu \frac{(\ln(x))^{\nu-1}}{\Gamma(\nu)x^{\nu/\mu_i+1}},$$

Expectation and variance are finite when

$$\nu > \mu_i \Rightarrow E[Y_i] = \left(\frac{\nu}{\nu - \mu_i}\right)^\nu, \nu > 2\mu_i \Rightarrow \text{Var}[Y_i] = \left(\frac{\nu}{\nu - 2\mu_i}\right)^\nu - \left(\frac{\nu}{\nu - \mu_i}\right)^{2\nu}.$$

see, e.g., Hogg and Klugman (1984).

## 1.2 Situations where a closed-form maximum likelihood estimator exists

Brouste et al. (2020) identify practical situations where a closed-form MLE exists for any distribution of the exponential family and any link function. These situations arise when explanatory variables are only categorical, that is no continuous variables are present as regressors. The case of categorical explanatory variables (only) in GLM is relevant in many practical situations. For instance, in the insurance industry, policy pricing uses a finite number of risk group relying on categorical explanatory variables. Typically, motor insurance ratings rely on vehicle classification with a large number of modalities, e.g., the dataset used for the 2017 pricing game by Charpentier (2017) exhibits this feature with 1023 vehicle models for 101 vehicle brands.

### 1.2.1 A single explanatory variable

Consider the case where  $x_i^{(1)} = 1$  is the intercept and  $x_i^{(2)}$  takes values in a set of  $d_2$  modalities  $\{v_1, \dots, v_{d_2}\}$  with  $d_2 > 2$ . To perform the estimation, an incidence matrix is derived

$$\left(x_i^{(2),j}\right)_{i,j} = \left(\mathbf{1}_{x_i^{(2)}=v_j}\right)_{i,j},$$

where  $x_i^{(2),j}$  is the binary dummy of the  $j$ th category for  $i \in I$  and  $j \in J = \{1, \dots, d_2\}$ . By construction, this incidence matrix has rows that sum to 1. If we use the combination of the incidence matrix with a 1-column for the intercept  $(x_i^{(1)}, x_i^{(2),j})_{i,j}$ : a redundancy appears. Hence, a linear constraint is imposed on  $\boldsymbol{\theta}$

$$\langle \mathbf{R}, \boldsymbol{\theta} \rangle = 0. \quad (1.7)$$

with  $\mathbf{R} = (r_0, r_1, \dots, r_{d_2})$  any real vector of size  $d_2 + 1$ . We get the usual set-ups (*no intercept, no first modality* or *zero-sum condition*) for some specific value of  $\mathbf{R}$ . Let  $\bar{Y}_n^{(j)}$  be the average empirical response

$$\bar{Y}_n^{(j)} = \frac{1}{m_j} \sum_{i=1}^n Y_i \times x_i^{(2),j},$$

of individuals having  $v_j$  and the frequency  $m_j = \sum_i x_i^{(2),j}$  of level  $v_j$ .

**Theorem 1.2.1 (Brouste et al. (2020))** *Suppose that for all  $i \in I$ ,  $Y_i$  takes values in  $b'(\Lambda)$ . If the vector  $\mathbf{R}$  is such that*

$$\sum_{j=1}^{d_2} r_j - r_0 \neq 0,$$

*then there exists a unique, consistent and explicit MLE  $\hat{\boldsymbol{\theta}}_n$  of  $\boldsymbol{\theta}$  for GLM  $g(E(Y_i)) = \theta_{(1)} + \sum_{j=1}^{d_2} x_i^{(2),j} \theta_{(2),j}$ , given by*

$$\hat{\theta}_{n,(1)} = \frac{\sum_{k=1}^{d_2} r_k g(\bar{Y}_n^{(k)})}{\sum_{k=1}^{d_2} r_k - r_0}, \hat{\theta}_{n,(2),j} = g(\bar{Y}_n^{(j)}) - \frac{\sum_{k=1}^{d_2} r_k g(\bar{Y}_n^{(k)})}{\sum_{k=1}^{d_2} r_k - r_0}, j \in J. \quad (1.8)$$

*Note that if  $\bar{Y}_n^{(j)}$  does not belong to  $b'(\Lambda)$ ,  $g(\bar{Y}_n^{(j)})$  and hence  $\hat{\theta}_{n,(l),j}$  are not defined.*

The proof of Theorem 1.2.1 relies on the explicit computation of the score equations (1.6) where simplifications occur in this special setting. We get the following linear system

$$M_{d_2} \times \boldsymbol{\theta} = \begin{pmatrix} \mathbf{g}(\bar{\mathbf{Y}}) \\ 0 \end{pmatrix}, \text{ with } M_{d_2} = \begin{pmatrix} \mathbf{1}_{d_2} & I_{d_2} \\ r_0 & \mathbf{r} \end{pmatrix}.$$

The matrix  $M_{d_2}$  can be inverted under the condition  $\sum_{j=1}^{d_2} r_j - r_0 \neq 0$  stated in Theorem 1.2.1. Three usual examples of contrasts are given in Table 1.3 where the matrix  $M_{d_2}$  is given. Equation (1.8) further simplifies in that case.

The case when there is no explanatory variable, i.e.  $Y_i$  are identically distributed, cannot be obtained with Equation (1.8). But in that case, we get with similar arguments  $\hat{\boldsymbol{\theta}}_{n,(1)} = g(\bar{Y}_n)$ . Despite the distribution of  $\bar{Y}_n^{(j)}$  still belongs to the exponential family, the bias of  $\hat{\theta}_{n,(1)}$  and  $\hat{\theta}_{n,(2),j}$

name	no-intercept	no first-level	zero-sum
$r_0$	1	0	0
$\mathbf{r}$	$\mathbf{0}_{d_2}^T$	$(1, \mathbf{0}_{d_2-1}^T)$	$\mathbf{1}_{d_2}^T$
$M_{d_2}^{-1}$	$\begin{pmatrix} \mathbf{0}_{d_2}^T & 1 \\ I_{d_2} & -\mathbf{1}_{d_2} \end{pmatrix}$	$\begin{pmatrix} 1 & \mathbf{0}_{d_2-1}^T & -1 \\ 0 & \mathbf{0}_{d_2-1}^T & 1 \\ -\mathbf{1}_{d_2} & I_{d_2-1} & \mathbf{1}_{d_2-1} \end{pmatrix}$	$-\frac{1}{d_2} \begin{pmatrix} -\mathbf{1}_{d_2}^T & 1 \\ \mathbf{1}_{d_2 \times d_2} - d_2 I_{d_2} & -\mathbf{1}_{d_2} \end{pmatrix}$

Table 1.3 – Three well known examples of contrasts

cannot be determined for a general link function  $g$ . However, we can show the consistency of the estimator and an asymptotic confidence interval, see Appendix A.1 of Brouste et al. (2020).

An important corollary giving an explicit fitted log-likelihood can be derived from Theorem 1.2.1 which will be used later. Let us note that the fitted log-likelihood does not depend on the link function but on the  $a$ ,  $b$ ,  $c$  functions and related derivatives, see Table 1.1.

**Corollary 1.2.1 (Brouste et al. (2020))** *The value of the log-likelihood defined in (1.5) taken on the exact MLE  $\hat{\boldsymbol{\theta}}_n$  (if it exists) given by (1.8), under constraint (1.7), does not depend on the link function  $g$ . More precisely, we have  $\forall i \in I$ ,  $\ell(\hat{\eta}_i) = (b')^{-1}(\bar{y}_n^{(j)})$  for  $j \in J$  such that  $x_i^{(2),j} = 1$  and*

$$\log L(\hat{\boldsymbol{\theta}}_n | \mathbf{y}) = \frac{1}{a(\phi)} \sum_{j=1}^d \sum_{i, x_i^{(2),j}=1} \left( y_i \tilde{b}(\bar{y}_n^{(j)}) - b(\tilde{b}(\bar{y}_n^{(j)})) \right) + \sum_{i=1}^n c(y_i, \phi),$$

with  $\tilde{b} = (b')^{-1}$ . Therefore, the criteria AIC and BIC are also independent of the link function  $g$ . The estimator of  $\phi$  is obtained by maximizing  $\log L(\hat{\boldsymbol{\theta}}_n | \mathbf{y})$  with respect to  $\phi$  given  $a, b, c$  functions.

## 1.2.2 Two explanatory variables

Let us now consider the two variable case where  $x_i^{(1)} = 1$  is the intercept and  $x_i^{(2)}, x_i^{(3)}$  are the two explanatory variables taking values in  $\{v_{j1}, \dots, v_{jd_j}\}$  with  $d_2$  and  $d_3$  modalities respectively.

Dummy	Frequency	Mean	Index
$x_i^{(2),k} = 1_{x_i^{(2)}=v_{2k}}$	$m_k^{(2)} = \sum_{i=1}^n x_i^{(2),k}$	$\bar{y}_n^{(2),k} = \frac{1}{m_k^{(2)}} \sum_{i=1}^n y_i x_i^{(2),k}$	$k \in K = \{1, \dots, d_2\}$
$x_i^{(3),l} = 1_{x_i^{(3)}=v_{3l}}$	$m_l^{(3)} = \sum_{i=1}^n x_i^{(3),l}$	$\bar{y}_n^{(3),l} = \frac{1}{m_l^{(3)}} \sum_{i=1}^n y_i x_i^{(3),l}$	$l \in L = \{1, \dots, d_3\}$
$x_i^{(k,l)} = x_i^{(2),k} x_i^{(3),l}$	$m_{k,l} = \sum_{i=1}^n x_i^{(k,l)}$	$\bar{y}_n^{(k,l)} = \frac{1}{m_{k,l}} \sum_{i=1}^n y_i x_i^{(k,l)}$	$(k, l) \in K \times L$

Table 1.4 – Dummies, frequencies and averages w.r.t explanatory variables

We define the matrix  $Q$  by

$$Q = (Q_1, I_{d_2 d_3}), \quad Q_1 = (\mathbf{1}_{d_2 d_3}, \mathbf{1}_{d_3} \otimes I_{d_2}, I_{d_3} \otimes \mathbf{1}_{d_2}), \quad (1.9)$$

where  $\otimes$  is the Kronecker product. The user usually imposes linear constraints  $R\boldsymbol{\theta} = 0$  characterized by the contrast matrix  $R = (R_1, R_2)$  with

$$R_1 = \begin{pmatrix} r_{0,1} & r_{1,1}^{(2)} & \dots & r_{d_2,1}^{(2)} & r_{1,1}^{(3)} & \dots & r_{d_3,1}^{(3)} \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ r_{0,q} & r_{1,q}^{(2)} & \dots & r_{d_2,q}^{(2)} & r_{1,q}^{(3)} & \dots & r_{d_3,q}^{(3)} \end{pmatrix}, \quad R_2 = \begin{pmatrix} r_{11,1} & \dots & r_{d_2 d_3,1} \\ \vdots & \dots & \vdots \\ r_{11,q} & \dots & r_{d_2 d_3,q} \end{pmatrix}.$$

**Theorem 1.2.2 (Brouste et al. (2020))** Suppose that for all  $i \in \{1, \dots, n\}$ ,  $Y_i$  takes values in  $b'(\Lambda)$ . Under constraint  $\mathbf{R}\boldsymbol{\theta} = \mathbf{0}_q$ , and if  $\mathbf{R}$  such that  $(\mathbf{Q}', \mathbf{R}')$  is of rank  $d_2 d_3$ , there exists a unique, consistent and explicit MLE  $\hat{\boldsymbol{\theta}}_n$  of  $\boldsymbol{\theta}$  for GLM

$$g(E(Y_i)) = \theta_1 + \sum_{k=1}^{d_2} x_i^{(2),k} \theta_{(2),k} + \sum_{l=1}^{d_3} x_i^{(3),l} \theta_{(3),l} + \sum_{(k,l) \in K \times L} x_i^{(k,l)} \theta_{k,l},$$

given by

$$\hat{\boldsymbol{\theta}}_n = (\mathbf{Q}'\mathbf{Q} + \mathbf{R}'\mathbf{R})^{-1} \mathbf{Q}'g(\bar{\mathbf{Y}}), \quad (1.10)$$

where the vector  $g(\bar{\mathbf{Y}})$  is  $((g(\bar{Y}_n^{(k,l)})))_{l,k}$ .

In the case where some interactions are not observed, i.e.,  $m_{k,l} = 0$  the design matrix  $Q$  and contrasts matrix  $R$  need to be appropriately reduced. The MLE of the model with only main effects for two explanatory variables defined as  $g(E(Y_i)) = \theta_1 + \sum_{k=1}^{d_2} x_i^{(2),k} \theta_{(2),k} + \sum_{l=1}^{d_3} x_i^{(3),l} \theta_{(3),l}$  does not present such an explicit formula whatever the matrix  $\mathbf{R}$  of rank 2. In that case, the MLE does not solve a least square problem under a linear constraint, see Appendix A.2 of Brouste et al. (2020). A similar corollary to Corollary 1.2.1 can also be obtained where an explicit log-likelihood is obtained.

### 1.2.3 Special cases of probability distribution

Brouste et al. (2020) analyze two very well known distributions: Pareto 1 and shifted log-normal distributions. They consider a transformation  $T$  such that the transformed response  $Z_i = T(Y_i)$  belongs to the exponential family (1.1), see Section 1.1.3. Table 1.5 gives the transform as well as the  $a$ ,  $b$ ,  $c$  functions of the exponential family.

Brouste et al. (2020, Section 4) study in details the exact distribution of the MLE in the Pareto 1 case, while Brouste et al. (2020, Section 5) focuses on the exact distribution of the MLE in the lognormal case. The latter case is interesting since a dispersion has to be estimated possibly by maximizing the log-likelihood. In both, model diagnostic can be performed via the study of specific residuals (respectively from a standard exponential or from a standard normal distributions).

Name	$T(x)$	$a(x)$	$b(x)$	$c(x, \phi)$
Pareto 1	$-\log(x/\mu)$	1	$-\log(\lambda)$	0
shifted lognormal	$\log(x - \mu)$	$x$	$x^2/2$	$-\frac{1}{2}(x^2/\phi + \log(2\pi\phi))$

Table 1.5 – log-transformed distributions

## 1.3 Closed-form likelihood-based estimators

In the spirit of the closed-form MLE of previous section, Brouste et al. (2021a) propose a closed-form estimator for GLM with multiple categorical explanatory variables. They define  $\tilde{\boldsymbol{\theta}}_n$  as

$$\tilde{\boldsymbol{\theta}}_n = (\mathbf{Q}^T \mathbf{Q} + \mathbf{R}^T \mathbf{R})^{-1} \mathbf{Q}^T g(\bar{\mathbf{Y}}), \quad g(\bar{\mathbf{Y}}) = \left( g(\bar{Y}_n^{(j)}) \right)_{j=1, \dots, d}, \quad (1.11)$$

where  $Q$  is defined by modeler's choice,  $g(\bar{\mathbf{Y}}) \in \mathbb{R}^p$  is the vector of  $g$ -transformed average response variable for each cross-category of categorical explanatory variables, and  $R$  the contrast matrix.



In general situations with multiple explanatory variables, typically taken as single effects in the linear predictor,  $\tilde{\boldsymbol{\theta}}_n$  is not the MLE but has the advantage to be explicit. Theorems 1 and 2 of Brouste et al. (2021a) give also the asymptotic distribution of MLE.

A simulation analysis is performed in order to benchmark the proposed estimator against the IWLS algorithm in terms of computation time and to compare the asymptotic variances on a Gamma distributed GLM. Brouste et al. (2021a) observe that the IWLS algorithm has a computation time increasing almost linearly with the modality number  $d$ . In comparison, the proposed estimator's computation time is almost constant and significantly lower. Asymptotic variances between the MLE and the proposed estimator are very closed in almost every cases.

## 1.4 Generalized Linear Model Trees

Dutang and Guibert (2021) propose GLM-based trees with an explicit likelihood cut-off. GLM-based trees are algorithms splitting dataset recursively where at each node the best split is searched among GLM with each explanatory variable taken individually in a two-step procedure : (i) the selection of the splitting variable  $j^*$  and (ii) search for the best split point  $s^*$ .

### Model-based (MOB) partitioning tree

The MOB model introduced by Zeileis et al. (2008) carries the integration of a parametric model fitted at each leaf of a tree, based on least squares, maximum likelihood or more broadly M-estimation. With this approach, the parametric model  $\mathcal{M}_{\mathcal{B}_b}(\mathbf{Y}, \{\mathbf{x}_i\}, \{\lambda\})$ ,  $b = 1, \dots, B$ ,  $i \in I$ , is fitted locally using vectors of covariates  $\{\mathbf{x}_i\}$  on each subset of a partition  $\{\mathcal{B}_b\}$  of  $B$  segments. The partition is determined when growing a tree based on partitioning variables  $\mathbf{z}_i = (z_{i,1}, \dots, x_{i,q}) \in \mathbb{R}^q$ , and an objective function is maximized to obtain the generic collection of parameters  $\{\lambda_b\}$ .

Dutang and Guibert (2021) consider the GLM as local model similarly to Rusch and Zeileis (2013) where parameters  $\boldsymbol{\theta}_b$ ,  $b = 1, \dots, B$ , are estimated by maximizing the log-likelihood. Their algorithm for GLM tree is given in Algorithm 1. Contrary to the CART, the MOB doesn't require a post-pruning procedure of the tree. A pre-pruning step can be applied for avoiding, e.g., that the size of a node becomes too small.

Firstly, the GLM is fitted on all observations of the current node  $b$  ( $i \in b$ ) with all explanatory variables available. If no control is assumed, there are  $p$  variables but it might be less if a random selection is performed. Unless explanatory variables come from a single categorical variable, for which an explicit MLE exists (Brouste et al., 2020),  $\hat{\boldsymbol{\theta}}_b$  is computed by the IWLS algorithm, see Section 1.1.

Secondly, variable selection is performed using a M-fluctuation test in order to avoid variable selection bias. Let  $W_j(t)$  be an empirical fluctuation process. Zeileis and Hornik (2007) show that under regularity conditions, a normalized  $W_j(t)$  converges toward a standard Brownian bridge, which allows to perform M-fluctuation tests for testing the null hypothesis of parameter stability and thus to select partitioning variables. For GLM, the  $i$ -th score contribution is explicit and given by

$$s_{i,j}(\boldsymbol{\theta}) = \frac{\partial \log L(\lambda_i, y_i)}{\partial \beta_j} = \frac{y_i - \mu_i}{V(\mu_i)} h'(\eta_i) z_{i,j}, \quad (1.12)$$

where  $\mu_i = h(\eta_i)$ ,  $h = g^{-1}$  the inverse link function,  $\eta_i = \langle \mathbf{x}_i, \boldsymbol{\theta} \rangle$ , and Table 1.1 in Section 1.1. The corresponding covariance matrix  $J(\boldsymbol{\theta})$  is also explicit so that the empirical fluctuation process has the

**Data :** Response vector  $\mathbf{y}$ , explanatory covariates  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , partitioning variables  $\mathbf{z}_1, \dots, \mathbf{z}_n$   
**while** Loop over node  $b$  until no significant instability is detected **do**

    Compute the observation number  $n_b = \#b$  for node  $b$ .

**if**  $n_b$  is too small **then**

        | Stop the process for that node.

**end**

    1. Fit the local model:

    Fit GLM for observations  $i \in b$  to obtain fitted parameter  $\hat{\theta}_b$ .

    2. Assess param. instab. with M-fluctuation tests:

**for**  $j = 1, \dots, q$  **do**

        Compute the  $i$ -th score contribution as  $\hat{s}_{i,j} = s_{i,j}(\hat{\theta}_b)$  in (1.12) for all  $i \in b$ .

**if**  $j$  is a numerical variable **then**

            Compute parameter instability using (1.13) as

$$\lambda_j = \max_{i=\underline{i}, \dots, \bar{i}} \frac{(n_b)^2}{i(n_b - i)} \left\| W_j \left( i/n_b, \hat{\theta}_b \right) \right\|_2^2,$$

            where  $[\underline{i}, \bar{i}]$  is the interval of potential instability.

**else**

            Compute parameter instability using (1.13) as

$$\lambda_j = \frac{1}{n_b} \sum_{c=1}^{l_j} (\#I_{v_{j,c}})^{-1} \left\| \Delta_{v_{j,c}} W_j \left( i/n_b, \hat{\theta}_b \right) \right\|_2^2,$$

            where  $I_{v_{j,c}} = \{i \in b, z_{i,j} = v_{j,c}\}$  is the set of observation indices in category  $v_{j,c}$ .

**end**

**end**

    Compute the  $p$ -value of the fluctuation test and assess the significance.

**if** there is at least one significant instable variable **then**

        Select the most unstable variable

$$j^* = \arg \max_{j \in \{1, \dots, q\}} \lambda_j.$$

    3. Choose the best splitting point  $s$ :

**if**  $j^*$  is a numerical variable **then**

        | Search for the optimal split point  $s^* \in (\min_i z_{i,j^*}, \max_i z_{i,j^*})$  based on (1.16).

**else**

        | Search for the optimal set  $s^* \subset \{v_{j^*,1}, \dots, v_{j^*,l_{j^*}}\}$  based on (1.16).

**end**

**end**

**end**

**Algorithm 1 :** Recursive partition algorithm for GLMs for a binary tree

following expression for GLM

$$W_j(t, \hat{\theta}) = \hat{J}^{-1/2} \frac{1}{\sqrt{n_b}} \sum_{i \in b} \hat{s}_{\sigma(i),j}, 0 \leq t \leq 1, \quad (1.13)$$

where  $\hat{s}_{i,j} = s_{i,j}(\hat{\theta})$ ,  $\sigma(i)$  is the ordering permutation giving the anti-rank observation of  $z_{i,j}$  and

$\hat{J} = J(\hat{\theta})$  the fitted covariance matrix. The function  $t \mapsto W_j(t, \hat{\theta})$  is a step function and we denote by  $\Delta_v W_j(t, \hat{\theta})$  the increment for variable  $j$  in category  $v$ . The splitting variable with the highest significant instability is selected, i.e., the lowest  $p$ -value satisfying the significant level adjusted with Bonferroni correction.

Thirdly, given the splitting variable, the best split point is chosen based on the objective function calculated on the  $B \geq 2$  daughter nodes of the splitting variable

$$O(\mathbf{y}, \phi, \hat{\theta}_1, \dots, \hat{\theta}_B) = \sum_{b=1}^B \log L(\hat{\theta}_b, \phi, y_i 1_{\{i \in L_b(j^*)\}}), \quad (1.14)$$

where  $L_b(j^*)$  corresponds to the  $b$ -th segment with respect to values taken by the variable  $j^*$ . For binary tree ( $B = 2$ ), only one split point for a continuous variable or one subset for a categorical variable, hereafter noted  $s$ , should be exhibited. In general, objective function (1.14) is not explicit because parameter vector  $\theta_b$  has to be estimated by IWLS. Dutang and Guibert (2021) focus on this last step and propose closed-form formula for  $(\hat{\theta}_1, \dots, \hat{\theta}_B)$  leading to an explicit objective function.

### Explicit likelihood cut-off for constant binary trees

Whatever the  $j^*$  variable is categorical or numerical, we use a generic notation for the linear predictor

$$\eta_i = \theta_L \times 1_{\{i \in L(j^*, s)\}} + \theta_R \times 1_{\{i \in R(j^*, s)\}}, \quad (1.15)$$

where  $L(j, s)$  and  $R(j, s)$  are the children subsets resulting from the split. When the variable is numerical, the subsets are  $i \in L(j, s) \Leftrightarrow z_{i,j} \in ]-\infty, s]$  and  $i \in R(j, s) \Leftrightarrow z_{i,j} \in ]s, +\infty[$ . Otherwise when the variable is categorical, the subsets are defined as  $i \in L(j, s) \Leftrightarrow i \in s$  and  $i \in R(j, s) \Leftrightarrow i \notin s$ .

	Left node	Right node
Frequency	$m_j^L(s) = \sum_{i=1}^n 1_{\{i \in L(j, s)\}}$	$m_j^R(s) = \sum_{i=1}^n 1_{\{i \in R(j, s)\}}$
Average	$\bar{Y}_j^L(s) = \frac{1}{m_j^L(s)} \sum_{i=1}^n y_i 1_{\{i \in L(j, s)\}}$	$\bar{Y}_j^R(s) = \frac{1}{m_j^R(s)} \sum_{i=1}^n y_i 1_{\{i \in R(j, s)\}}$

Table 1.6 – Notations for conditional frequencies and averages

Let  $\bar{\mathbf{y}}_{j^*}(s) = (\bar{y}_{j^*}^L(s), \bar{y}_{j^*}^R(s))$  be the vector of conditional average responses and  $\mathbf{m}_{j^*}(s) = (m_{j^*}^L(s), m_{j^*}^R(s))$  be the vector of absolute frequencies. Using Theorem 1.2.1 and Corollary 1.2.1 of Brouste et al. (2020), an explicit objective function  $s \mapsto O(\bar{\mathbf{y}}_{j^*}(s), \mathbf{m}_{j^*}(s))$  is obtained for the splitting variable  $j^*$  in order to find the best split point or the best subset  $s^*$  with

$$\begin{aligned} O(\bar{\mathbf{y}}_{j^*}(s), \mathbf{m}_{j^*}(s)) &= \tilde{b}(\bar{y}_{j^*}^L(s)) m_{j^*}^L(s) \bar{y}_{j^*}^L(s) - b\left(\tilde{b}(\bar{y}_{j^*}^L(s))\right) m_{j^*}^L(s) \\ &\quad + \tilde{b}(\bar{y}_{j^*}^R(s)) m_{j^*}^R(s) \bar{y}_{j^*}^R(s) - b\left(\tilde{b}(\bar{y}_{j^*}^R(s))\right) m_{j^*}^R(s), \end{aligned} \quad (1.16)$$

where  $\tilde{b} = (b')^{-1}$  is the inverse of  $b'$ , see Table 1.1.

This framework includes many tree-based algorithms such as the Bernoulli distribution corresponding to CART algorithm for classification, the normal distribution corresponding to CART algorithm for regression. However, we can use other well-known distributions of the exponential family: gamma, Poisson, inverse Gaussian, binomial (taken into account weights), log-transformed distributions, see Table 1.1.

Dutang and Guibert (2021) present a comprehensive numerical analysis in order to demonstrate that this approach significantly reduces the computation time of the GLM-based tree model compared to the features originally offered by the R package `partykit`. Furthermore, their numerical applications on continuous simulated datasets confirm the effective out-of-sample performance of GLM-trees compared to other tree-based approaches such as `rpart` or `ctree`, both with categorical and continuous splitting variables. Due to the explicit objective functions, GLM forests are also tested by Dutang and Guibert (2021).

It remains an open question if the Tweedie distribution could be used directly in the MOB approach or needs further study. Recent articles (Delong et al., 2021; Denuit et al., 2021; Denuit and Trufin, 2021) shows the relevancy of Tweedie models in actuarial models. Tree-based methods are used by practitioners for pricing (Henckaerts et al., 2021) and reserving purposes (Wüthrich, 2018a).

## 1.5 Other regression models

We briefly present here two other types of regression model that will be used in Chapter 5. Indeed, Section 5.1 focuses on customer modeling with Generalized Additive Models (GAM) and Survival Regression Models (SRM).

### 1.5.1 Additive regression models

As for GLM, the response variable  $Y_i$  belongs to the exponential family (1.1), however the predictor is no-longer linear as in (1.3) but of the form

$$\eta_i = \alpha_0 + \sum_{j=1}^p f_j(x_{i,j}),$$

where  $f_j$  are smooth non-parametric functions based on piecewise polynomials and  $\alpha_0$  is the intercept. In practice, categorical variables are included in the predictor as

$$\eta_i = \alpha_0 + \sum_{j=1}^{p_1} \alpha_j x_{i,j} + \sum_{j=p_1+1}^{p_1+p_2} f_j(x_{i,j}),$$

where  $(x_{i,j})_{j \leq p_1}$  are categorical variables and  $(x_{i,j})_{j > p_1}$  are continuous variables.

Fitting smooth non-parametric functions is a challenging task because the user will control the shape of the smooth functions only via a smoothing parameter but not the piecewise polynomials fitted for functions  $f_j$ 's, called smoothers. We present here only the main idea of the fitting algorithms, see Hastie and Tibshirani (1990) or Venables and Ripley (2002) for details.

All smoothers have a smoothing parameter  $\lambda$  (the polynom degree, the bandwidth or the span). A first concern is how to choose a criterion on which to optimize  $\lambda$  in an adaptive way. Then, a second concern is to find a reliable estimate of the parameters  $\alpha$  and the smooths coefficients given a smoothing value  $\lambda$ .

Firstly, assuming a value of  $\lambda$ , we present an algorithm to fit the model. Wood (2008) proposes a reliable method: the Penalized Iteratively Reweighted Least Square method (PIRLS). Unsurprisingly,

PIRLS is an iterative method aiming to minimize the following (theoretical) penalized deviance

$$\tilde{D} = D(f_1, \dots, f_p) + \sum_{j=1}^p \lambda_j \int f_j''(x_j)^2 dx_j,$$

where the first term corresponds to the deviance for the given distribution family and the second term penalizes the wiggly behavior of the smooth functions by approximating the curvature of  $f_j$ . In practice, a set of basis functions  $(b_{j,k})_{j,k}$  is chosen so that the smooth function  $f_j$  as

$$f_j(x) = \sum_{k=1}^{K_j} b_{j,k}(x) \beta_{j,k} \Rightarrow \sum_{j=p_1+1}^{p_1+p_2} f_j(x_i) = \sum_{j=p_1+1}^{p_1+p_2} \mathbf{b}_j^T(x_i) \boldsymbol{\beta}_j,$$

with  $\mathbf{b}_j$  is a vector of known coefficients and  $\boldsymbol{\beta}_j$  a vector of unknown coefficients

$$\mathbf{b}_j(x) = \begin{pmatrix} b_{j,1}(x) \\ \vdots \\ b_{j,K_j}(x) \end{pmatrix}, \quad \boldsymbol{\beta}_j = \begin{pmatrix} \beta_{j,1} \\ \vdots \\ \beta_{j,K_j} \end{pmatrix}.$$

Typical basis functions are B-splines, P-splines, see, e.g., Yee (2015, Chapter 2). In other words, the GAM is represented as a GLM with  $\eta_i = \langle \tilde{\mathbf{x}}_i, \boldsymbol{\theta} \rangle$  with  $\tilde{\mathbf{x}}_i$  containing the basis functions evaluated at the covariate values and  $\boldsymbol{\theta}$  containing linear parameter  $\alpha$  and coefficients  $\beta_{j,k}$ 's. Hence, the (empirical) penalized deviance to be minimized is given by

$$\tilde{D}(\boldsymbol{\theta}) = D(\boldsymbol{\theta}) + \sum_{j=1}^{p_1+p_2} \lambda_j \boldsymbol{\theta}^T S_j \boldsymbol{\theta},$$

where  $S_j$  contains known coefficients formed with  $\mathbf{b}_j$  and  $D(\boldsymbol{\theta})$  the GLM version of the deviance. PIRLS algorithm solves the problem  $\min \tilde{D}(\boldsymbol{\theta})$ , see Wood (2008) for details.

Secondly, PIRLS algorithm gives for any  $\lambda$  the corresponding fitted coefficient  $\hat{\boldsymbol{\theta}}(\lambda)$ . In the literature, there are many criteria to select the smoothing parameter\*: likelihood measures such as Restricted Maximum Likelihood (REML), Maximum Likelihood (ML) and cross validation measures such as Generalized Cross Validation (GCV), Generalized Approximate Cross Validation (GACV). These methods differ whether the smoothing parameter is treated as a random effect or not. We maximize either a likelihood metric (ML/REML) or minimize a prediction error (GCV/GACV). In both cases, a nested optimization method is needed where outer iterations optimizes  $\lambda$  and inner iterations provide  $\boldsymbol{\theta}(\lambda)$ , see Wood (2010) for details. GAM are used by actuaries and academics for insurance pricing, e.g., (Denuit and Lang, 2004; Henckaerts et al., 2018)

## 1.5.2 Survival regression models

In survival analysis, the variable of interest is represented by the random variable  $T$  denoting a lifetime of a specie, a human or a policy. There are three widely used distributions for  $T$ : the Weibull, the loglogistic and the lognormal distribution. Each of them can be characterized equivalently on  $T$  or  $\ln(T)$ . Generally, survival distributions are characterized by survival function  $S_T(t) = P(T > t)$ . Table 1.7 summarizes the relationship between the distributions<sup>†</sup> of  $T$  and  $\ln(T)$ . We change parametrizations from  $T$  to  $Y$  with  $\sigma = 1/\alpha$ ,  $\mu = -\ln(\lambda)$  and  $y = \ln(t)$ .

\*We cannot choose the smoothing parameter  $\lambda$  minimizing the deviance, because the model will overfit the data.

<sup>†</sup>Note that with a shape  $\alpha = 1$ , the distribution of Weibull is the exponential distribution.

$T$	Weibull	loglogistic	lognormal
$S_T(t)$	$e^{-(\lambda t)^\alpha}$	$\frac{1}{1+(\lambda t)^\alpha}$	$1 - \Phi(\alpha \ln(\lambda t))$
$h_T(t)$	$\alpha \lambda^\alpha t^{\alpha-1}$	$\frac{\alpha \lambda^\alpha t^{\alpha-1}}{1+(\lambda t)^\alpha}$	$\frac{\phi(t)}{1-\Phi(t)}$
$Y = \ln(T)$	extreme (min.) value	logistic	normal
$S_Y(y)$	$e^{-e^{\frac{y-\mu}{\sigma}}}$	$\frac{1}{1+e^{\frac{y-\mu}{\sigma}}}$	$1 - \Phi\left(\frac{y-\mu}{\sigma}\right)$

$\Phi$  denotes the distribution function of the standard normal distribution  $\mathcal{N}(0, 1)$

Table 1.7 – Survival distributions

Hidden in those expressions, we have the three link functions for binomial GLM, see Section 1.1, logit link:  $g(\pi) = \ln\left(\frac{\pi}{1-\pi}\right)$ , probit link:  $g(\pi) = \Phi^{-1}(\pi)$ , complementary log-log link:  $g(\pi) = \ln(-\ln(1-\pi))$ , whose inverse are the distribution function of standard distributions. Let us note  $Z$  the variable characterized by the distribution function  $g^{-1}$ , and then  $Y = \ln(T) = \mu + \sigma Z$ . To take explanatory variables  $\mathbf{x}_i$  of individual  $i$  into account, we change the constant location parameter  $\mu$  to a linear predictor  $\mu + \eta_i$

$$Y_i = \mu + \langle \mathbf{x}_i, \boldsymbol{\theta} \rangle + \sigma Z,$$

with  $\boldsymbol{\theta}$  an unknown coefficient vector. This implies that  $S_T(t) = S_{T_0}(e^{\langle \mathbf{x}_i, \boldsymbol{\theta} \rangle} t)$ , where  $T_0$  is a baseline distribution (i.e. one of the distributions in Table 1.7). From the last equation, we get the name of that type of model : accelerated / decelerated failure time model, since the coefficient  $e^{\langle \mathbf{x}_i, \boldsymbol{\theta} \rangle}$  changes the scale of time implying a decrease / increase of the survival function.

The estimation of the accelerated failure time model is done simply by maximizing the log-likelihood. From the asymptotic normal behavior of maximum likelihood estimators, we can deduce confidence interval, hypothesis test for the  $\boldsymbol{\theta}$ 's components. Therefore a p-value is available for each coefficient of the regression, which help to keep only the most significant variable.

In practice, full parametric models may not lead to reliable results due to the strong assumption on the baseline distribution. To circumvent this idea, one can use the Cox proportional hazard (PH) model. The Cox PH model can be seen as an extension of the accelerated failure time model. Let us recall that in the accelerated failure time model, the hazard function has the following form  $h_T(t|x_i) = e^{\langle \mathbf{x}_i, \boldsymbol{\theta} \rangle} h_{T_0}(t)$ , where  $T_0$  is a baseline distribution in Table 1.7. A natural extension is therefore to consider models with hazard function

$$h_T(t|x_i) = \beta(\mathbf{x}_i, \boldsymbol{\theta}) \times h_0(t),$$

where  $\beta$  models the effect of covariates on the response (with an unknown parameter  $\boldsymbol{\theta}$ ) and  $h_0$  is a non-parametric function. The name comes from the fact that  $h(t|x_i)/h(t|x_j) = \beta(\mathbf{x}_i, \boldsymbol{\theta})/\beta(\mathbf{x}_j, \boldsymbol{\theta})$  is constant with respect to time  $t$ , so the hazard functions are “proportional”. Unlike the previous framework, the baseline hazard rate is estimated non-parametrically by a step function, see, e.g., Therneau and Grambsch (2000).

Yet not stressed out here, these two frameworks of survival analysis can deal with censorship and/or truncation. But they cannot model multiple type of exit, that is to say, when the survival time  $T$  represent the (first) exit time due to multiple causes. When modeling human mortality, one may consider the different causes of death (e.g. infectious disease, respiratory diseases, heart diseases,...) and typically use the HCDatabase by INED and Mack Planck Institute (2021). In a multi-cause framework, (Beyersmann et al., 2012; Martinussen and Scheike, 2006), the failure cause is assumed to be known at the failure time. We define  $J_T$  the type of failure among  $\{1, \dots, J\}$ . The process  $(J_t)_t$  starting with  $J_0 = 0$  is a continuous-time random process that jumps at  $T$  into a state in  $\{1, \dots, J\}$ , see

Figure 1.1. This new framework is called *competing risk models* as any of the cause among  $\{1, \dots, J\}$  stops the process  $(J_t)_t$ .

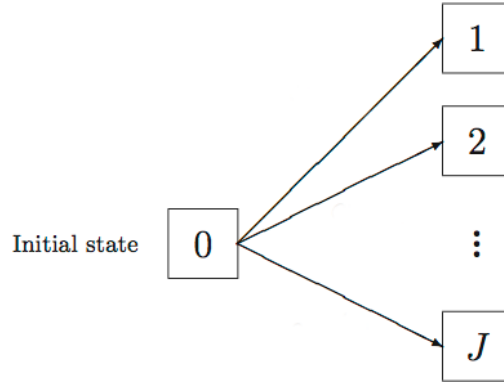


Figure 1.1 – Competing risk model and multistate representation, for  $j = 1, 2, \dots, J$ .

Let us define *cause-specific* hazard rates for  $j \in \{1, \dots, J\}$  as

$$\lambda_{T,j}(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt, J_T = j \mid T \geq t)}{dt}. \quad (1.17)$$

By the formula of total probability, we can retrieve the overall hazard rate by summation in (1.17):  $\lambda_{T,1}(t) + \dots + \lambda_{T,J}(t) = \lambda_T(t)$ , and recover the overall survival distribution of  $T$  by

$$P(T > t) = 1 - F_T(t) = S_T(t) = \exp\left(-\int_0^t (\lambda_{T,1}(s) + \dots + \lambda_{T,J}(s)) ds\right).$$

In practice, we are not interested in  $S_T$  but in the following probability  $F_{T,j}(t) = P(T \leq t, J_T = j)$ , the so-called *cumulative incidence function* (CIF) for cause  $j$ . Due to the event  $J_T = j$ , this is not a proper cumulative distribution function since  $F_{T,j}(t) \rightarrow P(J_T = j)$  as  $t \rightarrow +\infty$ . For a continuous distribution of  $T$ , we characterize it as the integral of a (improper) density. In other words,  $f_{T,j}(t)$  is the product of the cause-specific hazard rate and the probability to survive up to time  $t$ . Therefore, the CIF is

$$F_{T,j}(t) = \int_0^t \lambda_{T,j}(s) \exp\left(-\int_0^s \lambda_T(u) du\right) ds. \quad (1.18)$$

Hence, the CIF of cause  $j$  depends on **all other causes** via the global survival function, which makes the interpretation of the effects of covariates quite tricky since some effects come from the overall hazard rate  $\lambda_T(t)^*$ . This approach, called the *cause-specific approach*, thus requires to estimate the hazard rates (1.17) of all causes so as to estimate the CIF (1.18) of cause  $j$ .

The concurrent methodology to estimate the CIF of a single cause is obtained by considering a new competing risk process. Let us assume that cause 1 is our cause of interest. We define  $\tau$  as

$$\tau = T \times \mathbb{1}_{J_T=1} + \infty \times \mathbb{1}_{J_T \neq 1}.$$

The distribution of  $\tau$  is the same as  $T$  for  $J_T = 1$ ,  $P(\tau \leq t) = F_{T,1}(t)$  and a mass point at infinity  $1 - F_{T,1}(\infty)$ , probability to observe other causes ( $J_T \neq 1$ ) or not to observe any failure. The hazard rate of  $\tau$  is such that the CIF for cause 1 is computed as

$$F_{T,1}(t) = 1 - \exp\left(-\int_0^t \lambda_\tau(s) ds\right). \quad (1.19)$$

---

\*The CIF has the good property to be interpretable and summable  $P(T \leq t) = F_{T,1}(s) + \dots + F_{T,J}(s)$ , unlike to the function  $1 - \exp\left(-\int_0^t \lambda_{T,j}(u) du\right)$ .

Therefore the estimation of the CIF (1.19) does not depend on the estimation of other causes' hazard rates. This second approach is called the *subdistribution approach*, and often leads to different effects of covariates on the cause-specific hazard function on one side and on the corresponding CIF on the other side. Fine and Gray (1999) propose the so-called Fine–Gray model based the subdistribution approach. As we will later see in Section 5.1, competing risk models will prove useful to model the surrender risk in life insurance.

## 1.6 R implementation

Previous works of this section use extensively the R statistical software R Core Team, 2021. Generalized linear models are available in core package **stats**, generalized additive models in **mgcv** by Wood (2010). Numerous packages implements extensions or additional tools for GLM. Just to name few of them: **aglm** (Kondo, 2021) for regularized GLM, **cglm** (Sjolander, 2019) for conditional GLM, **dglm** (Dunn and Smyth, 2020) for double GLM, **glmnet** (Friedman et al., 2010) for lasso and elastic-net regularized GLM, **glmx** (Zeileis et al., 2015) for GLM with extra parameter distributions. The full list can be obtained by **RWsearch** (Kiener, 2021) package.

Closed-form estimators for GLM are implemented in **glmtools** by Brouste et al. (2021b), GLM trees and forests in **GLMsplittingTree** by Guibert and Dutang (2021). Classic survival regression models are performed thanks to the recommended package **survival** by Therneau (2021), whereas competing risk models are implemented in packages **timereg** (Martinussen and Scheike, 2006) and **cmprsk** (Gray, 2014).





## Chapter 2

# Parametric models without explanatory variables

This chapter is concerned with usual parametric models where observations come from independent and identically distributed (i.i.d.) models. Section 2.1 reviews quickly classic theorems for regular i.i.d. models, Section 2.2 provides an overview of loss models, Section 2.3 presents destruction rate models. Section 2.4 deals with the Le Cam's one-step estimation. Finally, Section 2.5 presents new explicit formulas for trimmed linear moments.

### 2.1 Independent and identically distributed models

Let  $x_1, \dots, x_n$  be the sample of observations: i.e. realizations of the random variables  $X_1, \dots, X_n$ . Assuming  $X_1, \dots, X_n$  are i.i.d. from a generic random variable  $X$ . We are interested in estimating the unknown parameter  $\theta$  of a parametric distribution  $F_X(\cdot; \theta)$ . Casella and Lehmann (1998) propose general methods to derive estimators of  $\theta$  based on some key features of  $F_X$ . Namely, one can maximize the log-likelihood of the sample  $x_1, \dots, x_n$  given a parametric distribution but other methods can be used such as moment matching estimation (2.8) or quantile matching estimation.

Under regularity conditions, e.g., Casella and Lehmann (1998, Theorem 6.5.1), the maximum likelihood estimator of  $\theta$  is consistent, asymptotically Gaussian and efficient. Among probability distributions verifying the regularity conditions, there is the multi-parameter exponential family (generalizing the one-parameter presented in (1.1) in Section 1.1) defined as the following density

$$f(x; \eta) = \exp \left( \sum_{i=1}^s \eta_i T_i(x) - A(\eta) \right) h(x), \quad x \in \mathbb{X}, \quad \eta \in \Xi, \quad (2.1)$$

where  $\mathbb{X}$  is the support of the distribution,  $\Xi$  the parameter set for which  $f$  is a density w.r.t. a measure,  $T : \mathbb{X} \mapsto \mathcal{T}$  is a vector-valued function,  $A : \Xi \mapsto \mathbb{R}$  is a real-valued function,  $h : \mathbb{X} \mapsto \mathbb{R}$  a base measure and  $\eta$  denote the natural parameter vector. We refer to Tables A.1 for univariate random variables and A.2 for random vectors in Appendix A.1. The multi-parameter exponential family is further used in Section 2.3 for one-inflated distributions.

Delignette-Muller and Dutang (2015) propose an R package **fitdistrplus** dedicated to fitting parametric models for any distribution created by users both for non-censored and interval-censored data. Fitting methods include maximum likelihood estimation, moment matching estimation or quantile



Figure 2.1 – Interrelations between distributions of the transformed gamma and inverse transformed gamma families. Diagrams derived from Figure 5.3 of Klugman et al. (2012).

matching estimation as well as maximum spacing estimation. Utility functions are implemented through S3 generic methods allowing to define usual functions (e.g. `plot`, `summary`,...) for fitting outputs. Concurrent packages use a more sophisticated approach (S4 or R6 classes), e.g., **DistributionFitR** (Geier et al., 2020). Both **OneStep** and **mbbfd** packages, respectively described in Section 2.4 and Section 2.3, are built upon **fitdistrplus** which allows inheriting S3 generic methods from the latter.

## 2.2 Loss and risk models in R

One important task of actuaries is the modeling of claim amount and claim count distributions for ratemaking, loss reserving or other risk evaluation purposes. Package **actuar** by Goulet et al. (2021a) features many support functions for loss distributions modeling. Firstly, **actuar** provides functions for heavy tail continuous distributions, Klugman et al. (2012); for phase-type distributions; for zero-truncated and zero-modified extensions of the discrete distributions commonly used in loss frequency modeling, Denuit et al. (2007). Secondly, it allows to compute raw moments, limited moments and the moment generating function (when it exists) of continuous distributions; and their empirical counterpart. Thirdly, credibility models and ruin theory are also implemented but are not discussed here. Dutang (2014b) shows an application of loss models of **actuar** in conjunction with **fitdistrplus**.

Regarding probability distributions, **actuar** provides a numerically reliable and fast implementation of additional distributions not supported in base R distributions, see Dutang et al. (2008). That is, **actuar** provides all the probability distributions found in Appendix A of Klugman et al. (2012) excluding the log- $t$ , but including the loggamma distribution (Hogg and Klugman, 1984), as well as for the Feller–Pareto distribution. Positive loss distributions used for loss modeling can be grouped in two families: the transformed (inverse) gamma family (Figure 2.1) and the Feller–Pareto family and related Pareto distributions (Figure 2.2).

The implementation of the transformed gamma family does not pose any difficulty. However, the Feller–Pareto distribution introduced by Feller (1968) and popularized by Arnold (2015) needs great care. It is defined as the distribution of the random variable

$$X = \mu + \theta \left( \frac{U}{V} \right)^{1/\gamma}, \quad \gamma, \theta > 0, \mu \in \mathbb{R}, \quad (2.2)$$

where  $U$  and  $V$  are two independent Gamma distributions with shape parameter  $\tau > 0$  and  $\alpha > 0$ , respectively, and common scale parameter 1. As displayed in Figure 2.2, this family contains for instance the Pareto IV and the Burr distributions. Dutang et al. (2021) make a reliability study of the implementation of the Feller–Pareto. Indeed, the computation of the distribution function or

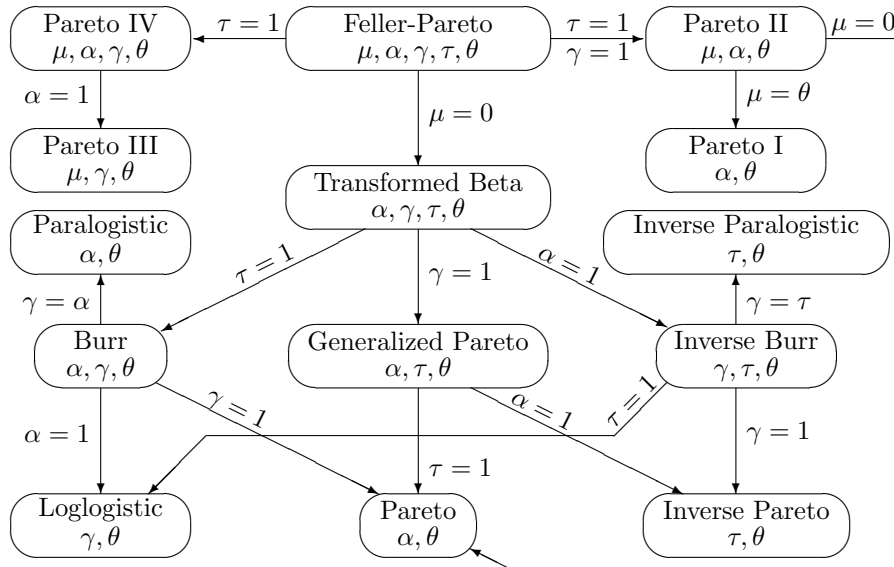


Figure 2.2 – Interrelations between distributions of the Feller–Pareto family. This diagram is an extension of Figure 5.2 of Klugman et al. (2012).

the limited expected values of the Feller–Pareto distribution may lead to numerical issues for certain parameter values of the shape parameters  $\alpha, \gamma, \tau$ .

In practice, there are also useful distributions not implemented in **actuar** but provided by other R packages. Dutang and Kiener (2021) provides a comprehensive list of probability distributions implemented in R packages. In particular, **tsallisqexp** by Cosma and Dutang (2021) proposes the Tsallis distribution also known as the q-exponential family distribution (Naudts, 2009). Distribution and density functions are based on the q-deformed exponential function. Another possibility for loss modeling is to use kernel-based estimation, e.g., (Buch-Larsen et al., 2005; Charpentier and Oulidi, 2010).

## 2.3 Destruction rate models

In the actuarial field, destruction rate models consist in studying random variable valued in  $(0, 1]$  representing the percentage of loss against an exposure measure, typically the total sum insured, the maximum probable loss or the known limit of the policy. These random variables have intrinsically a mixed-type distribution.

Due to the presence of a positive probability mass at 1, a common solution is to use one-inflated distributions for modeling such observations. Unlike probability distributions valued in  $(0, 1)$  which are particularly well-known (see, e.g., Casella and Lehmann (1998)), probability distributions valued in  $(0, 1]$  are not particularly studied. Only, Ospina and Ferrari (2010) study parameter estimation for the beta distribution inflated at zero and/or one.

One-inflated distributions can also be seen through the point-of-view of mixtures. In fact, it is a 2-component nonstandard mixture of a continuous distributions  $F_0$  and a degenerate distribution at 1. As pointed in McLachlan and Peel (2000), a particular popular nonstandard mixture is the mixture with excess of zeros where the first component is the degenerate distribution at 0 and the second component is a discrete distribution such as the Poisson distribution, see, e.g., Lachenbruch

(2002). Another particular nonstandard mixture is the mixture of a degenerate distribution at 0 and a continuous distribution, see, e.g., Panel on Nonstandard Mixtures of Distributions (1989) for a survey and Mills (2013) for a recent article. Despite its popularity, nonstandard mixtures are generally not studied in books of finite mixtures (Everitt and Hand, 1981; Titterington et al., 1985). Due to the nonstandard nature of our mixture, the use of the expectation-maximization (EM) algorithm is not relevant in our context, yet this algorithm is generally used when fitting mixtures, see Ridner and Walker (1984) for a survey.

The first part of Dutang and Spedicato (2021b) studies in details one-inflated defined by the following distribution function

$$F_1(x) = (1 - p_1)F_0(x) + p_1 \mathbb{1}_{[1, +\infty[}(x). \quad (2.3)$$

where  $F_0$  is the distribution function valued in  $(0, 1)$ . Assuming the original distribution  $F_0$  belongs to the multi-parameter exponential family, they show that the resulting distribution  $F_1$  remains in this family.

**Proposition 2.3.1 (Dutang and Spedicato (2021b))** *Assume that  $f_0$  belongs to the exponential family distribution (2.1) w.r.t. the Lebesgue measure  $\lambda$  with  $\mathbb{X} = (0, 1)$  and  $A, T, h$  given. The resulting one-inflated distribution also belongs the exponential family with a density w.r.t. the measure  $\mu(x) = \lambda(x) + \delta_1(x)$  with a parameter transform  $\tilde{\eta} : (p_1, \theta) \in (0, 1) \times \Theta \mapsto (\eta_0, \eta) \in \mathbb{R} \times H$  defined as*

$$\tilde{\eta} = \begin{pmatrix} \text{logit}(p_1) + A(\eta) \\ \eta \end{pmatrix}, \quad \tilde{A}(\tilde{\eta}) = \ln(1 - \text{logit}^{-1}(\eta_0 - A(\eta))) - A(\eta),$$

where  $\text{logit}^{-1}(x) = 1/(1 + e^{-x})$ , a data transformation  $\tilde{T}$  and a base measure  $\tilde{h}$  are defined as

$$\tilde{T}(x) = \begin{pmatrix} \tilde{T}_0(x) \\ \vdots \\ \tilde{T}_s(x) \end{pmatrix} = \begin{pmatrix} \mathbb{1}_{x=1} \\ \mathbb{1}_{x \neq 1} T(x) \end{pmatrix}, \quad \tilde{h}(x) = h(x) \mathbb{1}_{x \neq 1}.$$

Sufficient statistics, moments, covariances and moment generating function of transformed variables  $\tilde{T}(X)$  can be obtained by differentiating  $\tilde{A}$ , see, e.g., Theorems 1.5.8 and 1.5.10 of Casella and Lehmann, 1998.

Using Proposition 2.3.1, Dutang and Spedicato (2021b) establish the convergence and the asymptotic distribution of the MLE, in particular the parameter  $p_1$  is estimated by the empirical proportion of observations equalling one. Dutang and Spedicato (2021b) illustrate this theorem with the following distributions: one-inflated uniform, one-inflated shifted truncated Pareto, one-inflated beta distributions. If the original distribution  $F_0$  does not belong to the exponential family but is sufficiently regular, then the one-inflated distribution still preserves the regularity conditions. Dutang and Spedicato (2021b) illustrates this for one-inflated generalized beta distribution.

The second part of Dutang and Spedicato (2021b) considers a distribution defined through the exposure curve. The exposure curve function of  $X$  is defined as the ratio of the limited expected value and the expectation

$$G_X(d) = \frac{E(\min(X, d))}{E(X)} = \frac{\int_0^d (1 - F_X(x)) dx}{\int_0^1 (1 - F_X(x)) dx}, \quad (2.4)$$

where  $d \in [0, 1]$ .  $G_X$  have the desirable properties to be continuous, bounded, differentiable, concave and increasing on  $[0, 1]$ . One can derive distribution function, density, quantile and moments from an

exposure curve. Introduced by Bernegger (1997), the MBBEFD distribution is characterized by the following exposure curve

$$G_X(x) = \frac{\ln\left(\frac{a+b^x}{a+1}\right)}{\ln\left(\frac{a+b}{a+1}\right)},$$

with  $(a, b) \in \mathcal{D}_{a,b}$  where the parameter domain  $\mathcal{D}_{a,b}$  is

$$\mathcal{D}_{a,b} = \{(a, b) \in \mathbb{R}^2, a > -1, b > 0, a(1-b) \geq 0\} \cup \{(a, b), a = +\infty, 0 < b < 1\} \cup \{(a, b), a = -1, b > 1\}.$$

From the exposure curve, one can derive the distribution function, the quantile function, the expectation. Notably, the probability mass in 1 (i.e. total destruction) is given by

$$P(X = 1) = \frac{(a+1)b}{a+b} = h(a, b). \quad (2.5)$$

In particular,  $h(a, b)$  equals to 1 when  $a(1-b) = 0$ . Using (2.5), a density is defined w.r.t. the measure  $\mu(x) = \lambda(x) + \delta_1(x)$ . Dutang and Spedicato (2021b) study the parameter constraints so that the maximum likelihood estimator is asymptotically Gaussian and efficient. Dutang and Spedicato (2021a) provides an R package **mbbefd** of destruction rate models which allow users to fit by maximum likelihood estimation or by total-loss-moment matching estimation as well as to assess parameter uncertainty.

## 2.4 Le Cam's one-step estimation

As soon as the Fisher information matrix is sufficiently regular with respect to the parameter to be estimated, Le Cam's one-step estimation procedures can be used (Le Cam, 1956). They are based on an initial sequence of guess estimators  $\boldsymbol{\theta}_n^*$  and a single Newton step or a Fisher scoring step on the log-likelihood function. Let  $\ell(\boldsymbol{\theta})$  be the log-likelihood function,  $\dot{\ell}(\boldsymbol{\theta})$  be the gradient with respect to  $\boldsymbol{\theta}$  and  $\mathcal{I}(\boldsymbol{\theta}) = -E\left(\ddot{\ell}(\boldsymbol{\theta})\right)$  be the Fisher information matrix. Given an initial sequence of guess estimators  $(\boldsymbol{\theta}_n^*, n \geq 1)$ , Le Cam's one-step estimators write as

$$\bar{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_n^* + \mathcal{I}(\boldsymbol{\theta}_n^*)^{-1} \cdot \frac{1}{n} \sum_{j=1}^n \dot{\ell}(\boldsymbol{\theta}_n^*, X_j), \quad n \geq 1, \quad (2.6)$$

for the Fisher scoring type procedure and

$$\bar{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_n^* + \widehat{\mathcal{I}}_n(\boldsymbol{\theta}_n^*)^{-1} \cdot \frac{1}{n} \sum_{j=1}^n \dot{\ell}(\boldsymbol{\theta}_n^*, X_j), \quad n \geq 1, \quad (2.7)$$

for the Newton type procedure where  $\widehat{\mathcal{I}}_n(\boldsymbol{\theta}) = -\frac{1}{n} \sum_{j=1}^n \ddot{\ell}(\boldsymbol{\theta}, X_j)$  is the opposite of the Hessian for the log-likelihood function.

Brouste and Dutang (2021b) analyze in details one-step estimators with different type of initial guess as well as different form of Hessian matrix. For classical distributions with tractable gradient and Hessian, the Fisher information matrix  $\mathcal{I}(\boldsymbol{\theta})$  is used, whereas for other distributions the empirical average  $\widehat{\mathcal{I}}_n(\boldsymbol{\theta})$  is preferred.

Let us note that the initial guess may come from moment matching estimation, see Section 2.5, (e.g. for gamma or beta distributions), quantile matching estimation (e.g. for the Cauchy distribution), from ordinary least square (e.g. for the Weibull distribution) or maximum likelihood estimation on

a subsample of size  $\sqrt{n}$  (e.g. for the Pareto 2 distribution). Brouste and Dutang (2021b) show the computation gain in using the Le Cam's one-step estimation over the usual maximum likelihood estimation, even when initialization is carried out on a sub-sample.

The R package **OneStep** by Brouste et al. (2021c) is dedicated to one-step fitting procedure. In future research, Brouste and Dutang (2021a) will study the situation where the initial guess is based on the quantile matching estimation and the distribution belongs to the location-scale family.

## 2.5 Moment matching estimation and linear/trimmed-linear moments

Moment matching estimation consists in equalling empirical raw moments against theoretical moments of the parametric distribution  $F_X(\cdot; \theta)$  by solving

$$\begin{cases} E(X^1; \theta) = \frac{1}{n} \sum_{i=1}^n x_i, \\ \vdots \\ E(X^p; \theta) = \frac{1}{n} \sum_{i=1}^n x_i^p, \end{cases} \quad (2.8)$$

where  $E(X^j; \theta)$  is the  $j$ th raw moment depending on the parameter  $\theta \in \Theta \in \mathbb{R}^p$ . This approach may be unstable for the following reasons: (i)  $E(X^j; \theta)$  may not be tractable leading to a numerical evaluation of it which is time consuming when solving (2.8); (ii)  $E(X^j; \theta)$  may not exist for some values of  $\theta$  leading to a high uncertainty of corresponding empirical averages. Typically, the Cauchy distribution has an infinite expectation irrespectively of the value of its parameters, whereas the shifted Gompertz distribution does not have tractable moments.

Hosking (1990) proposes to use linear moments, so called L-moments in the following, as new measures of the location, scale and shape of probability distributions. To circumvent this drawback of non finite moments, Elamir and Seheult (2003) propose to use trimmed linear-moments (so-called TL-moment), which are a natural extension of L-moments. The TL-moments are of two types: either symmetric  $\lambda_m^{(t)}$  or asymmetric  $\lambda_m^{(s,t)}$ . They are defined as

$$\lambda_m^{(t)} = \frac{1}{m} \sum_{j=0}^{m-1} (-1)^j \binom{m-1}{j} E(X_{m+t-j, m+2t}), \quad (2.9)$$

$$\lambda_m^{(s,t)} = \frac{1}{m} \sum_{j=0}^{m-1} (-1)^j \binom{m-1}{j} E(X_{m+s-j, m+s+t}). \quad (2.10)$$

L-moments are obtained for  $s = t = 0$ . Table 2.1 lists the L/TL-moments for some common two-parameter distributions. As for centered moments,  $\lambda_3/\lambda_2$  and  $\lambda_3^{(1)}/\lambda_2^{(1)}$  are measures of skewness respectively L-skewness and TL-skewness, while  $\lambda_4/\lambda_2$  and  $\lambda_4^{(1)}/\lambda_2^{(1)}$  are measures of kurtosis respectively L-kurtosis and TL-kurtosis. TL-moments exist even if the corresponding L-moments do not, e.g., for the Cauchy distribution.

Dutang (2017b) is able to derive new closed-form formulas of L-moments and TL-moments for continuous probability distributions based on combinatorial identities and finite operators. Assuming a continuous distribution, the  $j$ th order statistic  $X_{j,n}$  has the following density

$$f_{X_{j,n}}(x) = \frac{n!}{(j-1)!(n-j)!} F(x)^{j-1} (1-F(x))^{n-j} f(x). \quad (2.11)$$

Distribution	Quantile $Q(p)$	L-moments				TL-moments			
		$\lambda_1$	$\lambda_2$	$\frac{\lambda_3}{\lambda_2}$	$\frac{\lambda_4}{\lambda_2}$	$\lambda_1^{(1)}$	$\lambda_2^{(1)}$	$\frac{\lambda_3^{(1)}}{\lambda_2^{(1)}}$	$\frac{\lambda_4^{(1)}}{\lambda_2^{(1)}}$
Uniform $\mathcal{U}(a, b)$	$a + (b - a)p$	$\frac{a+b}{2}$	$\frac{b-a}{6}$	0	0	$\frac{a+b}{2}$	$\frac{b-a}{10}$	0	0
Exponential $\mathcal{E}(\lambda)$	$-\lambda \log(1 - p)$	$\lambda$	$\frac{\lambda}{2}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{5\lambda}{6}$	$\frac{\lambda}{4}$	$\frac{2}{9}$	$\frac{1}{12}$
Normal $\mathcal{N}(\mu, \sigma^2)$	$\mu + \sigma\Phi(p)$	$\mu$	$\frac{\sigma}{\pi}$	0	0.1226	$\mu$	$0.297\sigma$	0	0.062
Cauchy $\mathcal{C}(\mu, \sigma^2)$	$\mu + \sigma \tan(\pi p - \frac{\pi}{2})$	$+\infty$	$+\infty$			$\mu$	$0.698\sigma$	0	0.343

Table 2.1 – L/TL-moments of common distributions.

By simple manipulations, the expectation can be written as

$$E(X_{j,n}) = j \binom{n}{j} \sum_{k=0}^{n-j} \binom{n-j}{k} (-1)^{n-j-k} I_F(n-k-1), \quad (2.12)$$

where  $I_F$  is defined as

$$I_F(k) = \int_a^b x F(x)^k f(x) dx = \int_0^1 Q(p) p^k dp, \quad (2.13)$$

with  $Q = F^{-1}$  the quantile function. The following proposition simplifies the computation of TL-moments (2.9) and (2.10).

**Proposition 2.5.1 (Dutang (2017b))** *Let  $m, s, t \in \mathbb{N}$ . The TL-moment of  $X$  can be expressed as*

$$\lambda_m^{(s,t)} = \frac{m+t+s}{m} \sum_{l=0}^{m+t-1} (-1)^{m+l+1} I_F(s+l) \binom{m+s+t-1}{s+l} \binom{m+s+l-1}{l}. \quad (2.14)$$

The proof of Proposition 2.5.1 is based on the absorption and the symmetry rules, the trinomial revision and the Vandermonde identity of combinatorial identities, see, e.g., Graham et al. (1994). This is an alternative proof compared to the original proof of Hosking (2007) based on shifted Legendre polynomials. The symmetric TL-moments are obtained by setting  $s = t$  in (2.14) and L-moments by setting  $s = t = 0$  in (2.14).

Dutang (2017b) applies this result to derive any TL-moment of two classical distributions : uniform distribution  $\mathcal{U}(a, b)$  in Propositions 11 and 12; and exponential distribution  $\mathcal{E}(\lambda)$  in Propositions 13 and 14. The latter case uses results proved by finite operators based on Harmonic numbers. Replacing ordinary moments in (2.8) by TL-moments (2.14) leads to TL-moment matching estimation. This allows to fit heavy-tailed distributions, e.g., Section 2.2, for which ordinary moments do not exist.





## Chapter 3

# Univariate and bivariate extreme models

This chapter is dedicated to extreme models in univariate and multivariate setting. Univariate extreme models focus on the estimation of the tail of the probability distribution. Generally, this is studied by looking at the distribution of the observed maximum of a finite sample or the distribution excesses over a high threshold. Multivariate extreme value statistics deals with the estimation of the tail of a multivariate distribution function based on a random sample. Of particular interest is the estimation of the extremal dependence between two or more variables. Modeling tail dependence is a crucial problem in actuarial science (see, e.g., Joe (2010)), firstly, because of the forthcoming Solvency II regulation framework that requires insurers and mutuals to compute 99.5% quantiles. Secondly, tail dependence can be used in the daily work of actuaries.

For instance, modeling tail dependence is of particular interest for pricing an excess-of-loss reinsurance treaty (see Goegebeur et al. (2021) for a recent article on reinsurance premium and Albrecher et al. (2017) for a review of statistical extremes in reinsurance), and for approximating very large quantiles of the distribution of the sums of possibly dependent risks (Arendarczyk et al., 2018; Barbe et al., 2006). In finance, obvious applications also arise, particularly since the financial crisis where academics and regulators were requested to further investigate the systemic risk, e.g., (Cai et al., 2015; Charpentier and Juri, 2006; Poon et al., 2004). Therefore, an accurate modeling of extremal events is needed to better understand the relationship of possibly dependent risks at the tail. Section 3.1 makes a short literature review of univariate and multivariate extreme value theory. Section 3.2 provides a robust and asymptotically unbiased estimation procedure of the tail dependence coefficient. Section 3.3 pursues the methodology of Section 3.2 with the estimation of extreme failure sets. Finally, Section 3.4 reviews R packages for extreme value analysis.

### 3.1 Known results for univariate and multivariate models

As in Chapter 2, we consider again a sample of size  $n$  of independent and identically distributed (i.i.d.)  $X_1, \dots, X_n$  from a generic random variable  $X$  having a distribution  $F$ . Extreme value analysis studies situations where the central limit theorem cannot apply since the variance or sometimes the expectation do not exist, and/or the targeted probability  $P(X > x)$  is smaller than  $1/n$ : the empirical survival function cancels out after the observed maximum.

In the univariate setting, there are multiple approaches to derive the extreme value index. We choose the maximum approach and refer to Beirlant et al. (2004) for a comprehensive overview.

**Theorem 3.1.1 (Fisher and Tippett (1928))** *Let  $(X_{i,n})_{1 \leq i \leq n}$  be the order statistics of a i.i.d. sample  $X_1, \dots, X_n$ . If  $F$  is continuous, there exists two normalizing series  $a_n$  et  $b_n > 0$  such that*

$$P\left(\frac{X_{n,n} - a_n}{b_n} \leq x\right) \xrightarrow{n \rightarrow +\infty} H_\gamma(x) \triangleq \exp\left(- (1 + \gamma x)_+^{-\frac{1}{\gamma}}\right),$$

where  $y_+ = \max(y, 0)$  and the limiting case  $H_0(x) = \exp(-\exp(-x))$ .  $H_\gamma$  is known as the generalized extreme value distribution.

Therefore, there are three types of distribution, generally called domain of attractions, depending on the sign of the extreme value index  $\gamma$ . Table 3.1 gives the three domains of attractions, the distribution support as well as some examples of distributions in the three domains. For a more comprehensive list, we refer to Beirlant et al. (2004) with Table 2.1 for the Fréchet case, Table 2.2 for the Weibull case, Table 2.3 for the Gumbel case.

Attraction domain	Values of $\gamma$	Distr. support	Examples	$P(X > x) = 1 - F(x)$	$\gamma$
Weibull	$\gamma < 0$	$(0, -1/\gamma)$	$\mathcal{U}(0, 1)$	$1 - x$	-1
			$\mathcal{RBurr}(\beta, \tau, \lambda, x_F)$ $\beta, \tau, \lambda > 0$	$\left(\frac{\beta}{\beta + (x_F - x)^{-\tau}}\right)^\lambda$	$-\frac{1}{\lambda\tau}$
Gumbel	$\gamma = 0$	$\mathbb{R}$	$\mathcal{LN}(\mu, \sigma)$ $\mu \in \mathbb{R}, \sigma > 0$	$\int_x^\infty \frac{1}{u} \frac{\exp(-\frac{1}{2\sigma^2}(\log u - \mu)^2)}{\sqrt{2\pi}\sigma} du$	0
			$\mathcal{G}(m, \lambda)$ $m > 0, \lambda > 0$	$\frac{\lambda^m}{\Gamma(m)} \int_x^\infty u^{m-1} e^{-\lambda u} du$	0
Fréchet	$\gamma > 0$	$(0, +\infty)$	$\mathcal{Pa}_1(1, \alpha)$ $\alpha > 0$	$x^{-\alpha}$	$\frac{1}{\alpha}$
			$\mathcal{Burr}(\beta, \tau, \lambda)$ $\beta > 0, \tau > 0, \lambda > 0$	$\left(\frac{\beta}{\beta + x^\tau}\right)^\lambda$	$\frac{1}{\lambda\tau}$
			$\mathcal{Fre}\left(\frac{1}{\alpha}\right)$ $\alpha > 0$	$1 - \exp(-x^{-\alpha})$	$\frac{1}{\alpha}$

Table 3.1 – Examples of extreme value index for usual distributions

The estimation of the extreme value index is obviously a well studied topic. Two usual estimators are based on the mean excess value of the log-transformed data. In the Fréchet case, Hill (1975) proposes to estimate  $\gamma$  by

$$\hat{\gamma}_{H,k} = \frac{1}{k-1} \sum_{i=1}^{k-1} \log\left(\frac{X_{n-i+1,n}}{X_{n-k+1,n}}\right). \quad (3.1)$$

Dekkers and de Haan (1989) propose a moment estimator defined as

$$\hat{\gamma}_{M,k} = \hat{\gamma}_{H,k} + 1 - \frac{1}{2} \left(1 - \frac{\hat{\gamma}_{H,k}^2}{S_{X,k,n}}\right)^{-1}, \quad \text{with } S_{X,k,n} = \frac{1}{k} \sum_{j=1}^k \log^2\left(\frac{X_{n-j+1,n}}{X_{n-k,n}}\right). \quad (3.2)$$

The performance of the Hill estimator strongly depends on the number of observations kept to estimate the tail index:  $\hat{\gamma}_{H,k}$  has a large variance if  $k$  is too small, whereas the Pareto-type tail behavior might not be verified for the selected  $k$  largest values if  $k$  is too large. A trade-off must be done: either by an empirical rule to find the area where the estimator is “stable” or by minimizing the asymptotic mean squared error, e.g., (Bader et al., 2018; Caeiro and Gomes, 2016).

A full characterization of the extremal dependence between variables can be obtained from functions like the spectral distribution function or the Pickands dependence function. The theorem below gives the limiting distribution of the component wise maximum of a bivariate sample.

**Theorem 3.1.2 (Beirlant et al. (2004))** *Let  $(X_i, Y_i) \stackrel{i.i.d.}{\sim} (X, Y)$  such that  $F_X(x) = F_Y(x) = \exp(-1/x)$ . Let  $(M_{x,n}, M_{y,n})$  be component-wise maximum  $M_{x,n} = \max_{i=1, \dots, n} \{X_i\}$  and  $M_{y,n} = \max_{i=1, \dots, n} \{Y_i\}$ . If*

$$P(M_{x,n} \leq x, M_{y,n} \leq y) \xrightarrow{n \rightarrow +\infty} G(x, y),$$

where  $G$  is a non-degenerated distribution, then  $G$  has the following form

$$G(x, y) = \exp(-V(x, y)), \quad x > 0, y > 0,$$

where  $V(x, y) = 2 \int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH(\omega)$  and  $H$  is a distribution on  $[0, 1]$  verifying  $\int_0^1 \omega dH(\omega) = 1/2$ .

An usual example is obtained by considering  $H$  as a discrete measure s.t.  $H(0) = \frac{1}{2}$  and  $H(1) = 1$ , leading to  $V(x, y) = x^{-1} + y^{-1}$ . Hence

$$G(x, y) = \exp(-x^{-1}) \exp(-y^{-1}), \quad x > 0, y > 0.$$

In other words, we get the independence copula with Fréchet marginals. Another usual example is the Dirac measure at 0.5, i.e.  $H(0.5) = 1$ , leading to  $V(x, y) = \max\left(\frac{1}{x}, \frac{1}{y}\right)$ . Hence

$$G(x, y) = \exp\left(-\max\left(\frac{1}{x}, \frac{1}{y}\right)\right) = \min\left(\exp\left(-\frac{1}{x}\right), \exp\left(-\frac{1}{y}\right)\right), \quad x > 0, y > 0.$$

This is the comonotone copula with Fréchet marginals.

Alternatively, similar to classical statistics one can try and summarize the extremal dependency in a number of well chosen coefficients that give a representative picture of the full dependency structure. We consider the estimation of the coefficient of tail dependence. Indeed, the extremal dependence between the components of a continuous random vector  $(X, Y)$  with unit Fréchet margins (note that this can be assumed without loss of generality) can be analyzed with the model of Ledford and Tawn (1997)

$$\mathbb{P}(X > x, Y > y) = x^{-c_1} y^{-c_2} \ell(x, y), \quad x, y > 0,$$

where  $c_1, c_2 > 0$  and  $\ell$  is a bivariate slowly varying function, i.e.

$$\frac{\ell(tx, ty)}{\ell(t, t)} \rightarrow \zeta(x, y) \text{ as } t \rightarrow \infty, \text{ for all } x, y > 0,$$

and the function  $\zeta$  is homogenous of order zero. The parameter  $\eta \triangleq (c_1 + c_2)^{-1}$  is called the coefficient of tail dependence. It satisfies  $\eta \in (0, 1]$ , and larger values of  $\eta$  indicate a stronger extremal dependence. Ledford and Tawn (1997) identify three types of asymptotic dependence in Table 3.2, where  $(X^\perp, Y^\perp)$  is a couple of independent variables but with the same marginal distribution as  $(X, Y)$ .

Extremal dependence	$\eta$ values	tail behavior
weak	$\eta \in (0, 1/2)$	$P(X > x, Y > x) < P(X^\perp > x, Y^\perp > x)$
independent	$\eta = 1/2$	$P(X > x, Y > x) = P(X^\perp > x, Y^\perp > x)$
strong	$\eta \in (1/2, 1]$	$P(X > x, Y > x) > P(X^\perp > x, Y^\perp > x)$

Table 3.2 – Extremal dependence

As we can imagine, several attempts have been made to estimate  $\eta$  from data. Since

$$\mathbb{P}(\min(X, Y) > z) = \mathbb{P}(X > z, Y > z) = z^{-1/\eta} \ell(z, z),$$

the transformed variable  $\min(X, Y)$  follows a Pareto-type distribution with index  $1/\eta$ . Therefore, one can estimate  $\eta$  with classical estimators for the extreme value index like the Hill (3.1) or moment estimator (3.2) of the univariate setting.

Beirlant et al. (2011) consider a couple  $(X, Y)$  with Fréchet marginals s.t.

$$P(X > x, Y > y) \underset{x, y \rightarrow +\infty}{\sim} x^{-c_1} y^{-c_2} g_1(x, y) (1 + g_2(x, y) x^{\rho_1} y^{\rho_2}), \quad (3.3)$$

where  $g_1, g_2$  are homogeneous function of order 0. Hence, the tail dependence coefficient  $\eta = 1/(c_1 + c_2)$  is such that the following function  $q$

$$q(t) = P(1 - F_X(X) < t, 1 - F_Y(Y) < t) = P(F_X^{-1}(1 - t) < X, F_Y^{-1}(1 - t) < Y),$$

is a regularly varying function  $q \in RV_{1/\eta}$ . In that context,  $\tau \triangleq \rho_1 + \rho_2$  is called the second-order parameter. Beirlant et al. (2011) adopt the following procedure to estimate  $\eta$ . They simplify the bivariate problem to an univariate setting

$$P(X > x, Y > y) = P(\min(X, \frac{\omega}{1-\omega} Y) > x) = P(Z > x),$$

where  $\omega = x/(x + y)$  is the radial parameter assumed to be fixed. The excess probability is approximated by

$$\frac{P(Z > ux)}{P(Z > u)} \approx \bar{F}_{EPD}(x/u; \eta, \delta_\omega, \tau),$$

where  $F_{EPD}$  is the cumulative distribution function of Extended Pareto Distribution defined as

$$F_{EPD}(z; \eta, \delta, \tau) = \left( 1 - \frac{1}{[z(1 + \delta - \delta z^{-\tau})]^{1/\eta}} \right) \mathbb{1}_{z > 1}. \quad (3.4)$$

Parameters  $\eta, \delta$  are estimated by maximum likelihood estimators whereas the second-order parameter is fixed at -1, linked to the Feller-Pareto family, see Section 2.2. Beirlant et al. (2011) illustrates their procedure on simulated framework by considering the FGM distribution with Fréchet marginals

$$P(X > x, Y > y) \underset{x, y \rightarrow +\infty}{\sim} \frac{1}{xy} (1 + \beta) \left[ 1 - \frac{1 + 3\beta}{2(1 + \beta)} \left( \sqrt{\frac{x}{y}} + \sqrt{\frac{y}{x}} \right) \frac{1}{\sqrt{xy}} \right], \quad (3.5)$$

Model (3.5) satisfies Equation (3.3) with

$$g_1(x, y) = (1 + \beta), \quad g_2(x, y) = -\frac{1 + 3\beta}{2(1 + \beta)} \left( \sqrt{\frac{x}{y}} + \sqrt{\frac{y}{x}} \right)$$

and  $c_1 = c_2 = 1, \rho_1 = \rho_2 = -1/2$  leading to  $\eta = 1/2, \rho = -1$ .

### 3.2 Estimation of the tail dependence coefficient

Classical estimators for the extreme value index like the Hill (3.1) or moment estimator (3.2) typically suffer from bias and also they are not robust with respect to outliers. The estimators proposed by Beirlant et al. (2011) are bias-reduced but are not robust with respect to outliers. These issues are addressed by Dutang et al. (2014). In order to obtain a bias-corrected estimator we, as usual in extreme value statistics, invoke a second order condition. In particular, we work under the following condition from Draisma et al. (2004), which can be seen as an extension of the above discussed Ledford and Tawn condition. The framework uses a random vector  $(X, Y)$  with joint distribution function  $F$  and continuous marginal distribution functions  $F_X$  and  $F_Y$  such that

$$\lim_{t \downarrow 0} q_1(t)^{-1} \left( \frac{P(1 - F_X(X) < tx, 1 - F_Y(Y) < ty)}{P(1 - F_X(X) < t, 1 - F_Y(Y) < t)} - c(x, y) \right), \quad (3.6)$$

exists for all  $x \geq 0, y \geq 0$  with  $x + y > 0$ , a function  $q_1$  tending to zero as  $t \downarrow 0$ , and  $c_1$  a function neither constant nor a multiple of  $c$ . Moreover, we assume that the convergence is uniform on  $\{(x, y) \in [0, \infty)^2 | x^2 + y^2 = 1\}$ .

Essentially, this condition is a second-order multivariate regular variation condition on the function  $R(x, y) \triangleq P(1 - F_X(X) < x, 1 - F_Y(Y) < y)$ . It can be shown that  $R(t, t)$  is regularly varying at zero with index  $1/\eta$ ,  $|q_1|$  is regularly varying at zero with index  $\tau \geq 0$ , and that the function  $c$  is homogeneous of order  $1/\eta$ , that is  $c(tx, ty) = t^{1/\eta}c(x, y)$ . Dutang et al. (2014) assume that the marginal distributions are unit Pareto and

$$P(X > x, Y > y) = x^{-c_1}y^{-c_2}g^*(x, y) \left( 1 + \frac{1}{\eta}\delta^*(x, y) \right), \quad (3.7)$$

where  $c_1, c_2$  are positive constants,  $\eta = (c_1 + c_2)^{-1} \in (0, 1)$  is the tail dependence coefficient,  $g$  is a continuous function that is homogeneous of order 0 and  $\delta$  is a function of constant sign in the neighborhood of zero,  $g^*(x, y) \triangleq g(1/x, 1/y)$  and  $\delta^*(x, y) \triangleq \delta(1/x, 1/y)$ . Furthermore, we assume that  $|\delta|$  being a bivariate regularly varying function\* We assume additionally that  $\xi$  is continuous, homogeneous of order  $\tau > 0$ , and that the convergence is uniform on  $\{(x, y) \in [0, \infty)^2 | x^2 + y^2 = 1\}$ .

Model (3.7) satisfies the second-order condition (3.6) and is a slight generalization of the second-order model considered in Beirlant et al. (2011). As in Beirlant et al. (2011), Dutang et al. (2014) reduce the problem dimension using a

$$P(X > x, Y > y) = P\left(X > x, \frac{\omega}{1-\omega}Y > x\right),$$

where  $\omega = x/(x+y) \in (0, 1)$  being the radial parameter. The transformed variable  $Z_\omega \triangleq \min(X, \frac{\omega}{1-\omega}Y)$  has the following survival function

$$P(Z_\omega > z) = C_\omega z^{-1/\eta} \left( 1 + \frac{1}{\eta}\delta_\omega(z) \right), \quad z > 0, \quad (3.8)$$

where  $C_\omega \triangleq (\omega/(1-\omega))^{d_2}g^*(1, (1-\omega)/\omega)$ , and  $|\delta_\omega|$  is a function of regular variation with index  $-\tau$ . This second-order condition (3.8) is identical to the one used in the univariate framework in Beirlant

---

\* $\delta$  is a bivariate regularly varying function when there exists a function  $\xi$  such that for all  $x, y \geq 0$

$$\lim_{t \downarrow 0} \frac{|\delta(tx, ty)|}{|\delta(t, t)|} = \xi(x, y).$$

et al. (2009). Again, the distribution of  $Z_\omega/u$  conditionally on  $Z_\omega > u$  is approximated, for large  $u$ , by the extended Pareto distribution (3.4).

In order to obtain robust estimates, Dutang et al. (2014) use the minimum density power divergence (MDPD) criterion. The density power divergence criterion was originally introduced by Basu et al. (1998) for the purpose of developing a robust estimation method. The density power divergence between density functions  $f$  and  $h$  is given by

$$\Delta_\alpha(f, h) = \begin{cases} \int_{\mathbb{R}} [h^{1+\alpha}(z) - (1 + \frac{1}{\alpha}) h^\alpha(z) f(z) + \frac{1}{\alpha} f^{1+\alpha}(z)] dz, & \alpha > 0, \\ \int_{\mathbb{R}} \log \frac{f(z)}{h(z)} f(z) dz, & \alpha = 0. \end{cases} \quad (3.9)$$

Note that for  $\alpha = 0$  one recovers the Kullback-Leibler divergence, whereas setting  $\alpha = 1$  leads to the  $L_2$  divergence. Assume that the density function  $h$  depends on a parameter vector  $\theta$ , and let  $f$  be the true density function of the random variable under consideration. The idea is then to estimate  $\theta$  by minimizing an empirical version of (3.9) based on a random sample  $Z_1, \dots, Z_n$  from  $f$

$$\widehat{\Delta}_\alpha(\theta) = \int_{\mathbb{R}} h^{1+\alpha}(z) dz - \left(1 + \frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^n h^\alpha(Z_i), \alpha > 0, \quad \widehat{\Delta}_0(\theta) = -\frac{1}{n} \sum_{i=1}^n \log h(Z_i). \quad (3.10)$$

For  $\alpha = 0$  in (3.10), one fits the model  $h$  to the data using the maximum likelihood method. The parameter  $\alpha$  controls the trade-off between efficiency and robustness of the MDPD estimator: the estimator becomes more efficient but less robust against outliers as  $\alpha$  gets closer to zero, whereas for increasing  $\alpha$  the robustness increases and the efficiency decreases.

In summary, for a sample  $(X_1, Y_1), \dots, (X_n, Y_n)$  of independent random vectors, one transforms into unit Pareto margins by using the empirical distribution functions. This gives

$$\widetilde{Z}_{\omega,i} \triangleq \min \left( \frac{n+1}{n+1-R_i^X}, \frac{\omega}{1-\omega} \frac{n+1}{n+1-R_i^Y} \right), \quad (3.11)$$

with  $R_i^X$  and  $R_i^Y$  denoting the rank of  $X_i$  and  $Y_i$ ,  $i = 1, \dots, n$ , in the respective samples. The parameters  $\eta$  and  $\delta_\omega$  of the extended Pareto distribution are estimated using (3.10) on the relative excesses  $Z_j = \widetilde{Z}_{\omega, n-m+j, n} / \widetilde{Z}_{\omega, n-m, n}$ ,  $j = 1, \dots, m$ , where  $1 \leq m \leq n-1$ , and  $\widetilde{Z}_{\omega, 1, n} \leq \dots \leq \widetilde{Z}_{\omega, n, n}$  are the order statistics of  $\widetilde{Z}_{\omega, 1}, \dots, \widetilde{Z}_{\omega, n}$ .

Let us denote the true value of  $\eta$  and  $\rho$  by  $\eta_0$  and  $\rho_0$ , respectively. We denote by  $\widehat{\eta}_m$  and  $\widehat{\delta}_{\omega, n}$  obtained by minimizing (3.10) for a given  $\alpha$  value.

**Theorem 3.2.1 (Dutang et al. (2014))** *Let  $\delta_{\omega, n} \triangleq \delta_\omega(\widetilde{Z}_{\omega, n-m, n})$ . Under regularity conditions, we have*

$$\sqrt{m} \begin{bmatrix} \widehat{\eta}_m - \eta_0 \\ \widehat{\delta}_{\omega, n} - \delta_{\omega, n} \end{bmatrix} \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}_2(\mathbf{0}, C^{-1}(\rho_0) B(\rho_0) D(\rho_0) B'(\rho_0) C^{-1}(\rho_0)),$$

where

$$B(\rho_0) = \eta_0^{-\alpha-2} \begin{bmatrix} \eta_0 & 0 & -1 \\ \eta_0 & -\eta_0(1-\rho_0) & 0 \end{bmatrix},$$

the symmetric matrix  $C(\rho_0)$  has elements

$$\begin{aligned} c_{11}(\rho_0) &= \eta_0^{-\alpha-2} \frac{1 + \alpha^2(1 + \eta_0)^2}{[1 + \alpha(1 + \eta_0)]^3}, \\ c_{12}(\rho_0) &= \eta_0^{-\alpha-2} \frac{\rho_0(1-\rho_0)[1 + \alpha(1 + \eta_0) + \alpha^2(1 + \eta_0)^2] + \alpha^3 \rho_0(1 + \eta_0)^3}{[1 + \alpha(1 + \eta_0)]^2 [1 - \rho_0 + \alpha(1 + \eta_0)]^2}, \\ c_{22}(\rho_0) &= \eta_0^{-\alpha-2} \frac{(1-\rho_0)\rho_0^2 + \alpha\rho_0^2(1 + \eta_0)[\alpha(1 + \eta_0) - \rho_0]}{[1 + \alpha(1 + \eta_0)][1 - \rho_0 + \alpha(1 + \eta_0)][1 - 2\rho_0 + \alpha(1 + \eta_0)]}, \end{aligned}$$

and the symmetric matrix  $D(\rho_0)$  has elements

$$\begin{aligned} d_{11}(\rho_0) &= \frac{\alpha^2(1+\eta_0)^2}{[1+\alpha(1+\eta_0)]^2[1+2\alpha(1+\eta_0)]}, \\ d_{21}(\rho_0) &= \frac{\alpha(1+\eta_0)[\alpha(1+\eta_0)-\rho_0]}{[1+\alpha(1+\eta_0)][1-\rho_0+\alpha(1+\eta_0)][1-\rho_0+2\alpha(1+\eta_0)]}, \\ d_{22}(\rho_0) &= \frac{[\alpha(1+\eta_0)-\rho_0]^2}{[1-\rho_0+\alpha(1+\eta_0)]^2[1-2\rho_0+2\alpha(1+\eta_0)]}, \\ d_{31}(\rho_0) &= \eta_0 \left( \frac{1}{[1+2\alpha(1+\eta_0)]^2} - \frac{1}{[1+\alpha(1+\eta_0)]^3} \right), \\ d_{32}(\rho_0) &= \eta_0 \left( \frac{1}{[1-\rho_0+2\alpha(1+\eta_0)]^2} - \frac{1}{[1+\alpha(1+\eta_0)]^2[1-\rho_0+\alpha(1+\eta_0)]} \right), \\ d_{33}(\rho_0) &= \eta_0^2 \left( \frac{2}{[1+2\alpha(1+\eta_0)]^3} - \frac{1}{[1+\alpha(1+\eta_0)]^4} \right). \end{aligned}$$

In practice, the true value of  $\rho$  is unknown. However, Dutang et al. (2014) study the asymptotic behavior of the MDPD estimator when the parameter  $\rho$  is possibly mis-specified. A similar convergence result can also be obtained in that case. Dutang et al. (2014) show the relevancy of this approach on simulated datasets from the FGM copula (3.5) and the Frank copula as well as an actuarial dataset of workers' compensation claims.

### 3.3 Failure set estimation

Dutang et al. (2016) pursue the work of Dutang et al. (2014) by considering the estimation of failure sets. That is, we want to estimate the tail probability

$$p_n = P(X > z_n, Y > y_n),$$

with  $X$  and  $Y$  being unit Pareto random variables, and where  $z_n \rightarrow \infty$  as  $n \rightarrow \infty$ , and  $y_n = \tilde{\omega}z_n$  for some  $\tilde{\omega} > 0$ . That is, we estimate a tail probability along a ray  $p_n = \overline{F}_{Z_\omega}(z_n)$ , where  $\omega = (1 + \tilde{\omega})^{-1}$ . Using the standard Landau notation  $\mathcal{O}()$ ,  $o()$  and  $\sim$  (Jones, 1997), let  $m$  be an intermediate sequence, i.e.  $m \rightarrow \infty$  as  $n \rightarrow \infty$  with  $m = o(n)$ . Assume that  $p_n$  satisfies

$$\frac{n}{m} \times p_n \xrightarrow{n \rightarrow +\infty} \beta \in [0, 1). \quad (3.12)$$

We denote again by  $\hat{\eta}_m$  and  $\hat{\delta}_{\omega,n}$  obtained by minimizing (3.10) for a given  $\alpha$  value. We can construct an estimator of  $p_n$  based on the approximation by the extended Pareto distribution (3.4) of  $Z_\omega/u$ , conditionally on  $Z_\omega > u$ , for large  $u$ . Setting  $u = \tilde{Z}_{\omega,n-m,n}$  as the  $m$ th order statistics of the transformed sample  $\tilde{Z}_{\omega,1}, \dots, \tilde{Z}_{\omega,n}$  defined in (3.11) and replacing  $\overline{F}_{Z_\omega}(u)$  by the empirical proportion  $m/n$ , we define the probability estimator  $\hat{p}_n$  as

$$\hat{p}_n = \frac{m}{n} \overline{F}_{EPD} \left( \frac{z_n}{\tilde{Z}_{\omega,n-m,n}}; \hat{\eta}_m, \hat{\delta}_{\omega,n}, \tilde{\rho} \right), \quad (3.13)$$

where  $\tilde{\rho}$  is either the true value of  $\rho$  or a mis-specified one.



**Theorem 3.3.1 (Dutang et al. (2016))** *Under regularity conditions with  $\beta \in (0, 1)$ , we have*

$$\sqrt{m} \begin{pmatrix} \hat{p}_n \\ p_n \end{pmatrix} - 1 \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \overline{W} \begin{pmatrix} 1 \\ c_\omega \end{pmatrix} - \ln \frac{\beta}{\eta_0} \Gamma - \frac{1 - \beta^{-\tilde{\rho}}}{\eta_0} \Delta + \lambda \frac{\eta_0 \xi_\omega c_\omega^{\rho_0}}{\rho_0} (\beta^{-\rho_0} - \beta^{-\tilde{\rho}}),$$

where the vector  $\overline{W}, Z, \Gamma$  follows a multivariate Gaussian distribution\*.

The previous theorem indicates that the tail probability estimator is asymptotically unbiased if one uses the correct value for  $\rho$ . We now consider the case where  $np_n/m \rightarrow 0$ , corresponding to a more extreme failure set. Let  $d_n = m/(np_n)$ .

**Theorem 3.3.2 (Dutang et al. (2016))** *Under regularity conditions and assuming additionally  $np_n = o(m)$ ,  $\ln(n)p_n = o(\sqrt{m})$  we have that*

$$\frac{\eta_0 \sqrt{m}}{\ln d_n} \begin{pmatrix} \hat{p}_n \\ p_n \end{pmatrix} - 1 \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \Gamma,$$

where the variable  $\Gamma$  follows a Gaussian distribution which is centered when the correct value  $\tilde{\rho} = \rho_0$ .

Dutang et al. (2016) propose a simulation analysis from the FGM copula (3.5) and the Frank copula as well as a bivariate actuarial dataset of (gross) loss ratios  $X$  and expense ratios  $Y$  of 173 private insurance companies operating in Australia. Both numerical illustrations show that the estimator  $\hat{p}_{n,0.5}$  remains stable both in the situation with and without outliers.

### 3.4 R implementation

The CRAN task view by Dutang and Jaunatre (2020) lists most of R packages related to extreme value analysis. Among the R packages listed in the task view or the review paper, the core packages are **evd** (Stephenson, 2002) and **evir** (Pfaff and McNeil, 2018) providing functions for univariate and bivariate models. There is also the **POT** package (Ribatet and Dutang, 2019) providing functions useful to perform a Peak Over Threshold analysis in univariate and bivariate cases. Finally, the **RTDE** package (Dutang, 2020b) provide robust tail dependence estimation corresponding to Sections 3.2 and 3.3.

The software review by Raymond-Belzile et al. (2021) surveys the recent development in the software implementation of extreme methods with a view towards the numerical challenges arising when performing extreme value analyses. They provide a comparative review by topic and highlight the differences between the routines provided by different softwares and packages, as well as, list areas where software development has been lacking. In particular, the **ReIns** (Reynkens and Verbelen, 2020) package covers statistical aspects of reinsurance thanks to splicing methods developed in Reynkens et al. (2017).

We do not mention it in Section 3.1 but for bivariate and multivariate modeling of extremes, one can use the class of copula satisfying the maximum-stable property. In that respect, the **copula** package (Hofert et al., 2020) provides the following max-stable copulas: Gumbel, Galambos and Husler-Reiss, whereas the **gumbel** package (Dutang, 2018) is a stand-alone for the Gumbel copula.

---

\*For sake brevity, parameters of this multivariate Gaussian distribution is omitted here.

# Chapter 4

## Ruin theory and reserving methods

This chapter focuses on the risk theory which includes ruin theory and reserving methods. Section 4.1 sets up the ruin theory problem of an insurance company and provides an overview of classical results. Section 4.2 proposes a risk model with dependence for the surplus process of an insurance company. Section 4.3 makes interconnections between ruin theory and reserving risk. Section 4.4 closes this chapter with a review of R packages dealing with risk theory.

### 4.1 A brief review of ruin theory

In this section, we study the solvency of an insurance company for a given portfolio size in the long run. As we do not identify the insurance policy of the claim provoking the insurer bankruptcy, claims are aggregated as they occur without distinguishing their policy origin. Therefore, the surplus  $(U_t)_{t \geq 0}$  of an insurance company at time  $t$  is represented by

$$U_t = u + ct - \sum_{i=1}^{N_t} X_i, \quad (4.1)$$

where  $u$  is the initial surplus,  $c$  is the premium rate,  $(X_i)_{i \geq 1}$  are the successive claim amounts<sup>†</sup> and  $(N_t)_{t \geq 0}$  is the claim arrival process (the claim associated waiting times are denoted by  $(T_i)_{i \geq 1}$ ).

In the Cramér-Lundberg model (Cramér, 1930; Lundberg, 1903),  $(N_t)_{t \geq 0}$  is modeled by an homogeneous Poisson process<sup>‡</sup>,  $(X_i)_{i \geq 1}$  are independent and identically distributed (i.i.d.) random variables and claim severities  $(X_i)_{i \geq 1}$  are independent of the claim waiting times  $(T_i)_{i \geq 1}$ . Andersen (1957) generalized this model by proposing a renewal process for the claim arrival process  $(N_t)_{t \geq 0}$  leading to the so-called Sparre Andersen model. The ultimate ruin probability

$$\psi(u) = P(\exists t > 0, U_t < 0 \mid U_0 = u), \quad (4.2)$$

is a major risk measure and has received considerable attention in the literature.

Since pioneer works of Cramér (1930) and Lundberg (1903), extensions have been proposed in many directions. Asmussen and Rolski (1991) studied ruin models with phase-type distributions for

---

<sup>†</sup>Once occurred, claim amounts are assumed to be fully determined. Few papers consider incurred but not reported claims directly the risk process, e.g., Trufin et al., 2011.

<sup>‡</sup>In other words,  $T_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{E}(\lambda)$ .

both claim severities  $X_i$  and claim waiting times  $T_i$ : a wide framework which includes mixture of exponentials and mixture of Erlang distributions. Gerber and Shiu (1998) unified the analysis of ruin measures in the Cramér-Lundberg model, including the deficit at ruin, the claim causing the ruin or the ruin probability, by introducing a so-called discounted penalty function. Gerber and Shiu (2005) and many others extended the Gerber-Shiu approach to a wider class of risk models.

Many variants of the Sparre Andersen model have also been proposed such as for non-homogeneous claim arrivals, e.g., (Albrecher and Asmussen, 2006; Lu and Garrido, 2005), reinsurance treaties, e.g., (Centeno, 2002a; Dimitrova and Kaishev, 2010), multivariate risks, e.g., (Cai and Li, 2005; Collamore, 1996; Picard et al., 2003) and capital injection, e.g., (Yao et al., 2011; Zhao et al., 2015). Many dependent risk models are also proposed to get rid of the independence assumption, e.g., (Albrecher and Boxma, 2004; Albrecher and Teugels, 2006; Boudreault et al., 2006; Constantinescu et al., 2019; Eryilmaz and Gebizlioglu, 2017; Ignatov and Kaishev, 2012; Lefèvre and Loisel, 2009; Lefèvre and Picard, 2011; Yang and Konstantinides, 2015)).

For a comprehensive review of ruin theory, we refer to Asmussen and Albrecher (2010) for a reference book and to Avram et al. (2014) for a survey of recent results in risk theory. We give below only two useful results. In the Sparre Andersen model with light-tailed claim amounts, with the aggregate claim increment  $Y = X - cT$  and  $x_0$  defined as the supremum of the set  $\{x, F_Y(x) < 1\}$ , we have

$$b_- e^{-\gamma u} \leq \psi(u) \leq b_+ e^{-\gamma u}, \quad u \geq 0,$$

where  $\gamma$  is the positive root of  $M_X(r)M_T(-rc) = 1$ , and constants  $b_-, b_+$  are defined by

$$b_- = \inf_{x \in [0, x_0[} \frac{e^{\gamma x} \bar{F}_Y(x)}{\int_x^{+\infty} e^{\gamma y} d\bar{F}_Y(y)} \quad \text{and} \quad b_+ = \sup_{x \in [0, x_0[} \frac{e^{\gamma x} \bar{F}_Y(x)}{\int_x^{+\infty} e^{\gamma y} d\bar{F}_Y(y)}.$$

$\gamma$  is called the adjustment coefficient. In the Sparre Andersen model with heavy-tailed claim amounts, where the expectation of claim amounts and claim waiting times are finite such that  $E(X) < cE(T)$ . Let  $F_{X,0}(x) = \int_0^x \bar{F}_X(y) dy / E(X)$ . If  $F_{X,0}$  belongs to the sub-exponential class, we have

$$\psi(u) \underset{u \rightarrow +\infty}{\sim} \frac{1}{cE(T) - E(X)} \int_u^{+\infty} \bar{F}_X(y) dy.$$

In particular for a Pareto-type tailed distribution  $P(X > x) = (k/x)^\alpha$  with  $\alpha > 1$ , it yields to

$$\psi(u) \underset{u \rightarrow +\infty}{\sim} \frac{k}{cE(T)(\alpha - 1) - \alpha k} \left(\frac{k}{u}\right)^{\alpha-1}.$$

Hence, the asymptotic behavior of  $\psi(u)$  depends merely on the tail behavior of the claim amount. A certain number of explicit results for (4.2) are available when the claim severity distribution belongs to the phase-type family, see Asmussen and Albrecher (2010, Chapter 9).

For models with dependence, Albrecher and Teugels (2006) investigated the case where the claim amounts and claim waiting times,  $(X_i, T_i)_{i \geq 1}$ , are correlated. They obtained an exponential decrease for  $\psi(u)$  in the case of light-tailed claim sizes. Albrecher et al. (2011) discussed the ruin probability when there is dependence by mixing in the claim amounts  $(X_i)_{i \geq 1}$  or the claim waiting times  $(T_i)_{i \geq 1}$ ; see also Constantinescu et al. (2011).

## 4.2 A risk model with dependence

Dutang et al. (2013b) consider the same insurance context as Albrecher et al. (2011) where the claim amounts  $(X_i)_{i \geq 1}$  are dependent through a mixing model. Such a form of dependence can translate some uncertainty on the model parameters due to incomplete available information. For instance, due to reporting or claim settlement delays, it might take a long time for the insurer to realize that claims are more adverse than expected. While an instantaneously informed insurer could manage the parameter uncertainty with credibility techniques, e.g., (Trufin and Loisel, 2013), an insurer suffering from information delays would in practice undergo high losses due to unfavorable parameter values before being able to react. In some markets, the only solution could be to stop the business. Thus,  $\psi(u)$  must then be interpreted as the probability of being ruined or of stopping the business if no premium adjustment is possible. Now, another possible situation of mixing is when the model parameters are not univocally fixed but depend on exogenous socio-economic factors. Another problem of interest would consist in accounting for the influence of an exterior environment that modifies the parameters in the course of time.

Precisely, the free surplus of an insurance company at time  $t$  is modeled by (4.1) where  $(N_t)_{t \geq 0}$  is the Poisson claim arrival process with intensity  $\lambda$  and claim amounts  $X_i$ 's are independent of the claim arrival process. However this time, we assume that the  $X_i$ 's are i.i.d. random variables conditionally on a latent random variable  $\Theta$ : given  $\Theta = \theta$ , they are conditionally distributed as  $X | \Theta = \theta$ . The variable  $\Theta$  can be used to translate, for example, the uncertainty in the claim amounts or their heterogeneity, see Denuit et al. (2007) bonus-malus models based on a latent variable approach. Its distribution is assumed to be continuous here.

Ruin occurs as soon as the surplus process becomes negative. Given  $\Theta = \theta$ , the ruin probability is thus defined as

$$\psi(u, \theta) = P(\exists t > 0 : U_t < 0 | U_0 = u, \Theta = \theta). \quad (4.3)$$

When  $X | \Theta = \theta \sim \mathcal{E}(\theta)$ , one gets from (4.3) the well-known formula

$$\psi(u, \theta) = \min \left( \frac{\lambda}{\theta c} e^{-u(\theta - \lambda/c)}, 1 \right),$$

where  $\min(\dots) < 1$  under the net profit condition  $\theta > \lambda/c \equiv \theta_0$ . Integrating over the parameter  $\theta$  yields for the (global) ruin probability

$$\psi(u) = F_{\Theta}(\theta_0) + I(u, \theta_0), \quad (4.4)$$

where  $F_{\Theta}$  is the distribution function of  $\Theta$  and

$$I(u, \theta_0) = \int_{\theta_0}^{\infty} \frac{\theta_0}{\theta} e^{-u(\theta - \theta_0)} dF_{\Theta}(\theta). \quad (4.5)$$

This expression for  $\psi(u)$  can also be interpreted from a regulatory point of view. Suppose that a regulator supervises a set of insurers who face exponentially distributed claims, each insurer having its proper parameter. Then, the term independent of  $u$ ,  $F_{\Theta}(\theta_0)$ , may be viewed as the proportion of insurers that will bankrupt irrespectively of their initial reserves, while the  $u$ -dependent term  $I(u, \theta_0)$  corresponds to the proportion of insurers that will bankrupt with  $u$  as initial reserves, the distribution of  $\Theta$  translating the heterogeneity among insurers.

Using the standard Landau notation  $\mathcal{O}()$ ,  $o()$  and  $\sim$  (Jones, 1997), Dutang et al. (2013b) obtain an upper bound and an asymptotic for the ruin probability.

**Theorem 4.2.1 (Dutang et al. (2013b))** Consider the continuous time model with conditional exponentially distributed claims ( $X|\Theta = \theta \sim \mathcal{E}(\theta)$ ) with  $\theta_0 = \lambda/c$ .

(i) If the maximum of the function  $g(\theta) \equiv F_\Theta(\theta)(1/\theta^2 + u/\theta)$ ,  $\theta \geq \theta_0$ , is attained at  $\theta = \theta_0$ , then an upper bound for  $\psi(u)$  with  $u > 0$  is

$$\psi(u) \leq F_\Theta(\theta_0) + \frac{1}{u} \frac{F_\Theta(\theta_0)}{\theta_0}. \quad (4.6)$$

(ii) If  $\Theta$  has a continuous distribution of density  $f_\Theta$  such that for some integer  $k \geq 1$ ,  $f_\Theta$  is  $C^{k-1}$  a.e. on  $[\theta_0, \infty)$  and  $f_\Theta^{(k)}$  is Lebesgue integrable on  $[\theta_0, \infty)$ , then

$$\psi(u) = F_\Theta(\theta_0) + \sum_{i=0}^{k-1} \frac{h^{(i)}(0)}{u^{i+1}} + o\left(\frac{1}{u^k}\right), \quad (4.7)$$

where  $h(x) \equiv \theta_0 f_\Theta(x + \theta_0)/(x + \theta_0)$  so that

$$h^{(i)}(0) = \sum_{j=0}^i (-1)^j \frac{i!}{(i-j)! \theta_0^j} f_\Theta^{(i-j)}(\theta_0).$$

For a first-order asymptotic ( $k = 1$ ), formula (4.7) gives the asymptotic rule  $A + B/u$

$$\psi(u) = F_\Theta(\theta_0) + \frac{f_\Theta(\theta_0)}{u} + o\left(\frac{1}{u}\right),$$

when  $f_\Theta$  is  $C^0$  a.e. on  $[\theta_0, +\infty)$  with  $f_\Theta'$  Lebesgue integrable. For  $k = \infty$ ,

$$\psi(u) \sim F_\Theta(\theta_0) + \sum_{i=0}^{+\infty} \frac{h^{(i)}(0)}{u^{i+1}} \quad \text{as } u \rightarrow +\infty,$$

when  $f_\Theta$  is  $C^\infty$  a.e. on  $[\theta_0, +\infty)$ . Note that this is related to an asymptotic result (2.3.2) of Olver et al., 2010 since  $I(u, \theta_0)$  is a Laplace transform.

A similar approach is possible when mixing the waiting times ( $T_1, T_2, \dots$ ) instead of the claim amounts. Following Section 3 of Albrecher et al. (2011), we get here

$$\psi(u) = \bar{F}_\Lambda(\lambda_0) + \int_0^{\lambda_0} \psi_u(\lambda) dF_\Lambda(\lambda), \quad \text{with } \psi_u(\lambda) = \frac{\lambda}{\lambda_0} e^{-u/\theta(1-\lambda/\lambda_0)},$$

where  $\lambda_0 = \theta c$ . A result similar can be derived too: in particular, a first order asymptotic expansion gives

$$\psi(u) = \bar{F}_\Lambda(\lambda_0) + \frac{f_\Lambda(\lambda_0)}{cu} + o\left(\frac{1}{u}\right).$$

Closed-form expression of the ruin probabilities are available for some specific distributions of  $\Theta$ . Table 4.1 from Dutang et al. (2013b) provides two examples of continuous distributions for  $\Theta$  for which explicit computations are possible using the incomplete lower gamma function  $\gamma(\cdot)$ , the incomplete upper gamma function  $\Gamma(\cdot)$  and the complementarity error function  $\text{erfc}$ , respectively for the gamma distribution  $\mathcal{G}a(\alpha, \lambda)$  and the Lévy distribution  $\mathcal{L}e(\alpha)$ . Dutang et al. (2013b) also study the tail of the resulting claim distribution under mixing by  $\Theta$  (their Proposition 4.1).

Dutang et al. (2013b) also propose a similar reasoning for the discrete-time analog of the continuous time ruin problem introduced by Gerber (1988) where the insurance portfolio is now examined at times  $t \in \mathbb{N}_0 = \{0, 1, \dots\}$ . A similar  $A + B/u$  for the ruin probability (Theorem 3.9 of Dutang et al. (2013b)) is possible in this setting as well as closed-form expressions for some particular cases of  $\Theta$ . Finally, a discussion of the dependence structure involved by the mixing approach is done, which for discrete-time yields to the use of the interpolated copula proposed by Genest and Nešlehová (2007).

$\Theta$	exact $\psi(u)$	asymptotic $\psi(u)$
$\mathcal{G}a(\alpha, \lambda)$	$\frac{\gamma(\alpha, \lambda\theta_0)}{\Gamma(\alpha)} + \frac{\lambda^\alpha \theta_0}{\Gamma(\alpha)} \frac{\Gamma(\alpha-1, \theta_0(\lambda+u))}{(\lambda+u)^{\alpha-1}} e^{\theta_0 u}$	$\frac{\gamma(\alpha, \lambda\theta_0)}{\Gamma(\alpha)} + \frac{\lambda^\alpha \theta_0^{\alpha-1} e^{-\lambda\theta_0}}{u\Gamma(\alpha)} + o(\frac{1}{u})$
$\mathcal{L}e(\alpha)$	$\operatorname{erfc}\left(\frac{\alpha}{2\sqrt{\theta_0}}\right) + \frac{\theta_0\sqrt{u}}{\alpha} e^{u\theta_0} \left[ \left(1 - \frac{1}{\alpha\sqrt{u}}\right) \tilde{u}_+ + \left(1 + \frac{1}{\alpha\sqrt{u}}\right) \tilde{u}_- - \frac{\tilde{u}}{\sqrt{\pi u\theta_0}} \right]$	$\operatorname{erfc}\left(\frac{\alpha}{2\sqrt{\theta_0}}\right) + \frac{\alpha e^{-\alpha^2/4\theta_0}}{2u\sqrt{\pi\theta_0^3}} + o(\frac{1}{u})$
where $\alpha > 1$ , $\tilde{u}_+ = e^{\alpha\sqrt{u}} \operatorname{erfc}(\sqrt{u\theta_0} + \alpha/(2\sqrt{\theta_0}))$ , $\tilde{u}_- = e^{-\alpha\sqrt{u}} \operatorname{erfc}(\sqrt{u\theta_0} - \alpha/(2\sqrt{\theta_0}))$ , $\tilde{u} = 2e^{-u\theta_0 - \alpha^2/(4\theta_0)}$ .		

Table 4.1 – Special cases of mixing distribution  $\Theta$

### 4.3 Interconnections between reserving models and ruin theory

Insurance reserving is another well-known topic for both actuaries and academics, while the ruin theory remains mainly the field of academics. The computation of insurance reserves being mandatory whereas ruin-related indicators are not is one of the main reasons to explain why practitioners neglect the use of ruin theory in their daily business. Nevertheless, with the upcoming risk-based regulatory requirements, the computation of solvency probabilities at different levels and different time horizons is increasingly popular in the past ten years.

Another important factor explaining the disaffection of practitioners for ruin theory when assessing reserves is the type of data to be used: the data granularity for classic reserving methods is line-of-business aggregated datasets whereas in ruin theory, individual loss level is needed, (Asmussen and Albrecher, 2010) and the references therein. Reserving methods are in fact mainly for aggregated data triangles, (Wuethrich and Merz, 2008). As pointed out by Wuethrich and Merz (2008), “most of the classical claims reserving methods do not distinguish reported claims from not-reported claims.” However, there is a growing literature for micro-level or individual claim-level reserving methods, and also advanced aggregated models are still studied, e.g., Denuit and Trufin (2017) propose to use compound Poisson distribution with 2-distribution mixture for summands.

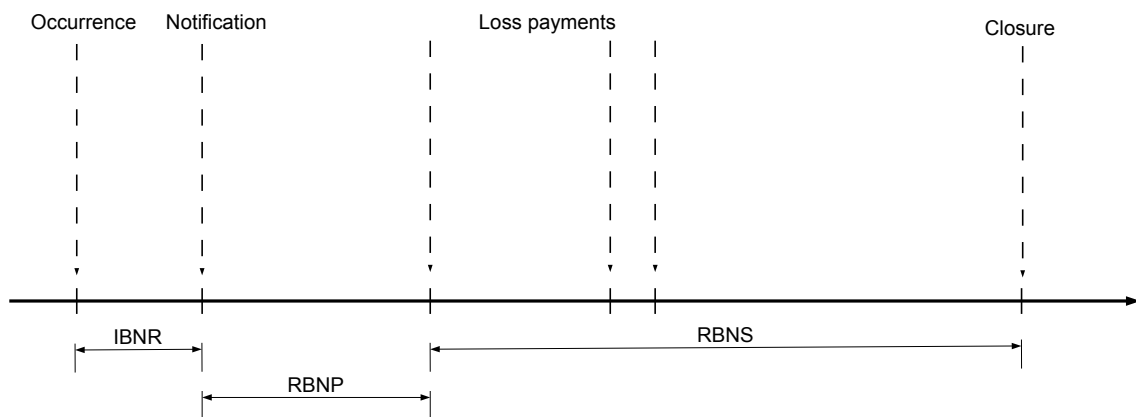


Figure 4.1 – Claim development process (IBNR: incurred but not reported, RBNP: reported but not paid, RBNS: reported but not settled)

Reserving in a continuous time perspective dates back to pioneer works of Karlsson (1976), Jewell (1989) and Arjas (1989). Few years after these papers, Norberg (1993) first formulates the reserving problem in a continuous time probabilistic setting by considering marked Poisson processes, see e.g. subsequent extensions Haastруп and Arjas (1996). In these works, the full claim process described in Figure 4.1 is considered. That is, the  $i$ th claim is characterized by a quintuple  $(T_i, U_i, V_i, Y_i, Y'_i(v)_{v \in [0, V_i]})$  where  $T_i$  denotes the occurrence time,  $U_i$  the reporting delay,  $V_i$  the settlement time,  $Y_i$  the total claim amount and  $Y'_i(\cdot)$  the payment process. The time between occurrence

and notification (i.e.  $t \in [T_i, T_i + U_i]$ ) corresponds to the reporting delay (IBNR in Figure 4.1) by the policyholder. And the time between notification and closure (i.e.  $t \in [T_i + U_i, T_i + U_i + V_i]$ ) corresponds to the settlement time, which can be further subdivided into the waiting time of first payment (RBNP in Figure 4.1) and the payment process (RBNS in Figure 4.1). The claim process  $(T_1, T_2, \dots)$  is governed by a non-homogeneous Poisson process  $(N_t, t \geq 0)$ .

A certain number of recent papers deal with individual claim reserving models with different approaches. A first approach is the statistical approach in discrete-time. Pigeon et al. (2013) model paid claim random vectors with a multivariate skew symmetric distribution, claim's development time/delays with a multi-modified discrete distribution (among the  $(a, b, 3)$  class). Parametric distribution allows to derive tractable conditional moments of ultimate claim amounts. Pigeon et al. (2014) extends Pigeon et al. (2013) by modeling paid and incurred claim random vectors by further assuming that at ultimate year both paid and incurred amounts are equal. Again parametric distribution allows to derive tractable conditional moments. For both articles, the predicted distribution of claim reserves from the micro-level method is narrower than the corresponding distribution of the macro-level method. Charpentier and Pigeon (2016) provide some clues why micro-level methods might be more robust. Some authors also use non-parametric machine learning methods, e.g., (Wüthrich, 2018a,b).

A second approach is the probabilistic approach in continuous time. Larsen (2007) revisits the marked Poisson process of Norberg (1993) by including claim type when fitting this model under a GLM framework. Antonio and Plat (2014) consider a marked Poisson process based on the framework of Norberg (1993, 1999) using a set of parametric distributions for reporting delay and claim amount and a piecewise constant intensity for claim arrivals. Boumezoued and Devineau (2017) work with similar assumptions and derive MSEF and other uncertainty measures of the reserve. Currently, there exists almost one alternative to marked Poisson processes in the actuarial literature: the Poisson shot noise processes of Klueppelberg and Mikosch (1995), further developed in Matsui (2014, 2015) and Matsui and Mikosch (2010). They consider that the  $i$ th claim is a pair  $(T_i, L_i(\cdot))$  where  $L_i$  may represent the loss process, typically independent Lévy processes.

Dutang and Brouste (2016) consider a process closed to the marked Poisson process of Norberg (1993). The surplus of an insurance company at time  $t$  is represented by the risk process  $R_t = u + ct - S_t$ , where  $S_t$  denotes the aggregate claim amount,  $u$  is the initial surplus,  $c$  is the premium rate. Traditionally, the aggregate claim amount  $S_t$  is the sum of claim amounts  $X_1, X_2, \dots$  arrived before time  $t$ , i.e.  $S_t = \sum_{i=1}^{N_t} X_i$ . By considering settlement times, we assume that

$$S_t = \sum_{i=1}^{N_t} Z_i(t), \quad \text{with } Z_i(t) = \frac{X_i}{V_i}(t - T_i) \mathbb{1}_{[T_i, T_i + V_i)}(t) + X_i \mathbb{1}_{[T_i + V_i, \infty)}(t),$$

where  $V_i, T_i$  denote respectively the settlement time and the occurrence of the  $i$ th claim. We denote by  $R_t$  the corresponding risk process. In other words,  $Z_i(t)$  corresponds to the claim amount paid at time  $t$  and  $X_i - Z_i(t)$  is the outstanding claim amount. As the  $i$ th claim is represented by  $(T_i, V_i, X_i)$  and the implicit assumption that the payment process is an affine function of time  $t$ , we have a simplified version of Norberg's model.

Let the risk process with no-delay and no-settlement be

$$\tilde{R}_t = u + ct - \sum_{i=1}^{N_t} X_i.$$

Comparing both risk processes  $R_t$  and  $\tilde{R}_t$ , we remark that  $R_t \geq \tilde{R}_t$  a.s.. Therefore, the corresponding ruin probability of the considered model is always lower than in the classical setting, which provides

an upper bound. For the following study, we introduce a deterministic settlement function

$$g(t, w, v) = \frac{t - w}{v} \mathbb{1}_{[w, w+v)}(t) + \mathbb{1}_{[w+v, \infty)}(t). \quad (4.8)$$

representing the percentage of the claim paid at time  $t$  such that  $Z_i(t) = X_i g(t, T_i, V_i)$ . Keeping in mind that we want to derive explicit formulas, we make the following assumptions

- A1. the claim arrival process  $(N_t, t \geq 0)$  is a homogeneous Poisson process with intensity  $\lambda$ ,
- A2. the settlement times are independent and identically distributed  $((V_i)_i \stackrel{\text{i.i.d.}}{\sim} V)$ ,
- A3. the claim amounts are independent and identically distributed  $((X_i)_i \stackrel{\text{i.i.d.}}{\sim} X$  with finite variance),
- A4. there is independence between waiting times, settlement times and claim amounts  $(T_i - T_{i-1} \perp V_i \perp X_i)$ .

Dutang and Brouste (2016) assume that no reporting delay occurs: only the time of settlement is assumed random.

In the sequel, we will need claim index sets defined as follows

$$\mathcal{C}_t^{ns} = \{i \in \{1, \dots, N_t\}, T_i \leq t < T_i + V_i\}, \quad \mathcal{C}_t^s = \{i \in \{1, \dots, N_t\}, T_i + V_i \leq t\},$$

representing respectively not-settled claims and settled claims. These sets are a disjoint partition of the set of claims occurred before time  $t$ , i.e.  $\mathcal{C}_t^{ns}, \mathcal{C}_t^s \subset \{1, \dots, N_t\}$ . We introduce filtrations depending for the claim arrival process and the knowledge up to time  $t$

$$\mathcal{F}_t^N = \sigma(N_s, 0 \leq s \leq t), \quad \mathcal{F}_t^{N,V} = \sigma((N_s, 0 \leq s \leq t), V_1, \dots, V_{N_t}),$$

$$\mathcal{F}_t^{N,C} = \sigma((N_s, 0 \leq s \leq t), (V_i)_{i \in \mathcal{C}_t^s}), \quad \mathcal{F}_t^{N,C,X} = \sigma((N_s, 0 \leq s \leq t), (V_i)_{i \in \mathcal{C}_t^s}, X_1, \dots, X_{N_t}).$$

Dutang and Brouste (2016) present an efficient procedure to compute the conditional first two moments of the aggregate claim process  $(S_t, t \geq 0)$  based on two operators  $A_k$  and  $A_k^*$ .  $A_k(\cdot)(s, t)$  is defined as

$$A_k(G)(s, t) = \sum_{j=1}^k \int_s^{t_{k+1}} \dots \int_s^{t_{j+1}} G(t, t_j) \frac{(t_j - s)^{j-1}}{(j-1)!} dt_j \dots dt_k, \quad t_{k+1} = t, \quad s < t, \quad (4.9)$$

and  $A_k^*(\cdot)(s, t)$  is defined by

$$A_k^*(\mathbf{G})(s, t) = \sum_{i=1}^{k-1} \int_s^{t_{k+1}} \dots \int_s^{t_{i+1}} \frac{(t_i - s)^{i-1}}{(i-1)!} G^{i,k}(t, t_i, \dots, t_k) dt_i \dots dt_k \quad t_{k+1} = t, \quad s < t, \quad (4.10)$$

with  $\mathbf{G} = (G^{1,k}, \dots, G^{k,k})$  and  $G^{i,k}(t, w_i, \dots, w_k) = \sum_{m=i+1}^k G(t, w_i) G(t, w_m)$ .

The two following propositions give the conditional moments, unconditional moments are obtained by setting  $s = 0$ .

**Proposition 4.3.1 (Dutang and Brouste (2016))** *The conditional expectation at time  $t$  of the aggregate claim amount knowing the information up to time  $s < t$  is*

$$E(S_t | \mathcal{F}_s^{N,C,X}) = \sum_{i=1, i \in \mathcal{C}_s^s}^{N_s} X_i + \sum_{i=1, i \in \mathcal{C}_s^{ns}}^{N_s} X_i G(t, T_i) + E(X) e^{-\lambda(t-s)} \sum_{k=1}^{\infty} \lambda^k A_k(G)(s, t), \quad (4.11)$$

where  $A_k$  is defined in (4.9) and  $G$  is the bivariate function defined as

$$G(t, w) = E(g(t, w, V)). \quad (4.12)$$

Here  $g$  is the settlement function defined in (4.8).



**Proposition 4.3.2 (Dutang and Brouste (2016))** *The conditional second-order moment at time  $t$  of the aggregate claim amount knowing the information up to time  $s < t$  is*

$$\begin{aligned}
E(S_t^2 | \mathcal{F}_s^{N,C,X}) &= \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} X_i X_j \mathbb{1}_{i,j \in \mathcal{C}_s^s} + 2 \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} X_i X_j G(t, T_i) \mathbb{1}_{i \in \mathcal{C}_s^{n_s}, j \in \mathcal{C}_s^s} \\
&+ \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} X_i X_j G(t, T_i) G(t, T_j) \mathbb{1}_{i,j \in \mathcal{C}_s^{n_s}} \\
&+ 2E(X) e^{-\lambda(t-s)} \sum_{k=1}^{\infty} \lambda^k A_k(G)(s, t) \sum_{i=1}^{N_s} (X_i \mathbb{1}_{i \in \mathcal{C}_s^s} + X_i \mathbb{1}_{i \in \mathcal{C}_s^{n_s}} G(t, T_i)) \\
&+ e^{-\lambda(t-s)} \sum_{k=1}^{\infty} \lambda^k \left( E(X^2) A_k(G_2)(s, t) + 2E(X)^2 A_k^*(\mathbf{G})(s, t) \right),
\end{aligned}$$

where  $A_k^*$  is defined in (4.10),  $G(t, w) = E(g(t, w, V))$ ,  $G_2(t, w) = E(g(t, w, V)^2)$ , and  $g$  the settlement function in (4.8).

It is worth emphasizing that  $A_k(G)(s, t)$  defined in (4.9) is only a particular case of the operator  $A_k^*(\mathbf{G})(s, t)$  with a family of bivariate functions, namely  $\mathbf{G} = \mathbf{G}_{bi} = (G_{bi}^{1,k}, \dots, G_{bi}^{k,k}) = (G(t, t_1), \dots, G(t, t_k))$ . From the first two moments, the computation of the conditional variance is immediate

$$\text{Var}(S_t | \mathcal{F}_s^{N,C,X}) = E((S_t)^2 | \mathcal{F}_s^{N,C,X}) - E(S_t | \mathcal{F}_s^{N,C,X})^2.$$

Direct computation of (4.12) leads to

$$G(t, w) = (t - w) \int_{t-w}^{\infty} \frac{dF_V(x)}{x} + F_V(t - w),$$

for  $t \geq w \geq 0$ . In order to compute the second-order moment, similar computations lead to

$$G_2(t, w) = (t - w)^2 \int_{t-w}^{\infty} \frac{dF_V(x)}{x^2} + F_V(t - w).$$

The  $i$ th component of  $\mathbf{G} = (G^{1,k}, \dots, G^{k,k})$  consists in summing  $G$  functions, namely

$$G^{i,k}(t, w_i, \dots, w_k) = \sum_{m=i+1}^k G(t, w_i) G(t, w_m).$$

Below, we present two settlement functions studied in Dutang and Brouste (2016).

### 4.3.1 Immediate settlement

Let us start with the usual case of immediate settlement. If  $V = 0$  a.s., then  $G(t, w) = \mathbb{1}_{\{w \leq t\}}$ , leading to  $A_k(G)(t) = t^k / (k - 1)!$ . Therefore, we get back to a well known result  $E(S_t) = \lambda t E(X)$ . Consequently, we also have  $G_2(t, w) = \mathbb{1}_{\{w \leq t\}}$ . Then,  $A_k(G_2)(t) = t^k / (k - 1)!$ . Furthermore,

$$G^{i,k,\Sigma}(t, w_i, \dots, w_k) = \mathbb{1}_{\{w_i \leq t\}} \sum_{m=i+1}^k \mathbb{1}_{\{w_m \leq t\}} \Rightarrow A_k^*(\mathbf{G})(0, t) = \sum_{i=1}^{k-1} (k - i) \frac{t^k}{k!} = \frac{t^k}{2(k - 2)!}.$$

Thus we retrieve another well-known result of a compound Poisson process

$$\begin{aligned}
E(S_t^2) &= e^{-\lambda t} \sum_{k=1}^{\infty} \lambda^k \left( E(X^2) \frac{t^k}{(k-1)!} + E(X)^2 \frac{t^k}{(k-2)!} \right) \\
&= E(X^2) \lambda t e^{-\lambda t} \sum_{k=1}^{\infty} \frac{(\lambda t)^{k-1}}{(k-1)!} + e^{-\lambda t} E(X)^2 (\lambda t)^2 \sum_{k=1}^{\infty} \lambda^{k-2} \frac{t^{k-2}}{(k-2)!} \\
&= E(X^2) \lambda t + E(X)^2 (\lambda t)^2.
\end{aligned}$$

### 4.3.2 Zero-inflated settlement

Short-tailed business (such as material damages for motor and household insurance with settlement generally within four or five years) corresponds to line of business where the settlement time is either quick or immediate. As the opposite, long-tailed business such as third-party liability (especially medical malpractice or liability for lawyers) experiences very long development of claims (generally more than to 20 years, see, e.g., Partrat et al. (2008)). Hence, modeling the settlement process depends heavily on the studied guarantee.

Dutang and Brouste (2016) first attempt to model such a process using the zero-inflated exponential distribution. In the numerical section, we will split the dataset between long and short tailed guarantees leading to distinct values of parameters of the two situations. The use of more complex distributions (such as Gamma or Weibull and their zero-inflated version) is postponed to future research.

Considering a zero-inflated exponential distribution for  $V$  (i.e. a mixture of a geometric distribution and a Dirac distribution at 0) yields to

$$G_{ZIE}(x) = (p + (1-p)(1 - e^{-\mu x})) \mathbb{1}_{[0,+\infty)}(x).$$

In other words with probability  $p$ , the claim is settled immediately, otherwise (with probability  $1-p$ ) the settlement time is strictly positive. Hence, for  $t > w$ ,

$$\begin{aligned}
G_{ZIE}(t, w) &= p + (1-p)(t-w) \int_{t-w}^{\infty} \frac{\mu e^{-\mu x}}{x} dx + (1-p)(1 - \exp(-\mu(t-w))) \\
&= 1 - (1-p) \exp(-\mu(t-w)) + (1-p)\mu(t-w)E_1(\mu(t-w)),
\end{aligned}$$

where  $E_1$  denotes the exponential integral, see, e.g., Olver et al. (2010, Chap. 6).

$$\begin{aligned}
G_{ZIE,2}(t, w) &= p + (1-p)(t-w)^2 \int_{t-w}^{\infty} \frac{\mu e^{-\mu x} dx}{x^2} + (1-p)(1 - \exp(-\mu(t-w))) \\
&= 1 - (1-p) \exp(-\mu(t-w)) + (1-p)\mu(t-w)E_2(\mu(t-w)),
\end{aligned}$$

where  $E_2$  denotes the generalized exponential integral, see, e.g., Olver et al. (2010, Chap. 8). Finally,

$$G_{ZIE}^{i,k,\Sigma}(t, w_i, \dots, w_k) = \sum_{m=i+1}^k G_{ZIE}(t, w_i) G_{ZIE}(t, w_m).$$

Of course, the case of the exponential distribution is obtained by setting  $p = 0$  in the previous expressions of  $G_{ZIE}$  and  $G_{ZIE,2}$ .

### 4.3.3 Reserving topics

From a reserving perspective, we now ignore the initial capital  $u$  and the premium rate  $c$  and focus on the aggregate claim amount  $S_t$  at time  $t$ . Classical methods for claim reserving are designed for aggregated data for which claim amounts are aggregated per accident year and per development year, e.g., (Wuethrich and Merz, 2008). Therefore, claims are sorted per accident year and cumulated per development year to get a so-called claims development triangle.

At individual claim level, the accident year  $k$  of a claim occurred at time  $T$  is the year of occurrence, i.e.  $k = \lfloor T \rfloor$  (where  $\lfloor \cdot \rfloor$  denotes the integer part). The  $j$ th development year of a claim occurred at time  $T$  corresponds to payments done in interval  $(\lfloor T \rfloor + j - 1, \lfloor T \rfloor + j)$ . Let  $k = 0, \dots, K$  be an accident year and  $j = 0, \dots, J$  a development year. As before, we want to deal with reserving topics, and we introduce the claim set of accident year  $k$  reported at time  $t$

$$\mathcal{C}_{t,k} = \{i \in \{1, \dots, N_t\}, k = \lfloor T_i \rfloor\}.$$

Note that the current time is  $k + j + 1$  since both  $k$  and  $j$  starts from 0. Let us define the aggregate (paid) claim amount for accident year  $k$  and development year  $j$

$$S_{k,j} = \sum_{i \in \mathcal{C}_{k+j+1,k}} Z_i(j+k+1) = \sum_{i=1}^{N_{j+k+1}} Z_i(j+k+1) \mathbb{1}_{i \in \mathcal{C}_{k+j+1,k}}.$$

The sum can be simplified using  $Z_i(t) = X_i g(t, T_i, V_i)$

$$S_{k,j} = \sum_{i=1}^{N_{j+k+1}} X_i g(j+k+1, T_i, V_i) \mathbb{1}_{\{k \leq T_i < k+1\}}.$$

Denoting  $\tilde{g}_k(y, t, v) = g(y, t, v) \mathbb{1}_{\{k \leq t < k+1\}}$ , we get back to a sum similar the aggregate claim  $S_t$  at time  $t = j + k + 1$  with a new settlement function  $\tilde{g}_k(y, t, v)$ .

In order to deal with conditional expectation, we split the claim set into two subsets

$$\begin{aligned} \mathcal{C}_{t,k}^{ns} &= \{i \in \{1, \dots, N_{j+k+1}\}, k = \lfloor T_i \rfloor, T_i \leq t < T_i + V_i\}, \\ \mathcal{C}_{t,k}^s &= \{i \in \{1, \dots, N_{j+k+1}\}, k = \lfloor T_i \rfloor, T_i + V_i \leq t\}. \end{aligned}$$

They represent claims of accident year  $k$  not-settled and settled at time  $t$ . Let us define the aggregate (paid) claim amount for accident year  $k$  and development year  $j + m$  given that the current time is  $k + j + 1$

$$S_{k,j+m} = \sum_{i=1}^{N_{j+k+1}} X_i g(j+m+k+1, T_i, V_i) \mathbb{1}_{\{k \leq T_i < k+1\}}.$$

Within this notation, a reserving triangle looks like (for  $s = 3$ )

AY $k \setminus$ DY $j$	0	1	2
0	$S_{0,0}$	$S_{0,1}$	$S_{0,2}$
1	$S_{1,0}$	$S_{1,1}$	$E\left(S_{1,2} \mid \mathcal{F}_3^{N,C,X}\right)$
2	$S_{2,0}$	$E\left(S_{2,1} \mid \mathcal{F}_3^{N,C,X}\right)$	$E\left(S_{2,2} \mid \mathcal{F}_3^{N,C,X}\right)$

Using Proposition 4.3.1, the conditional expectation of the aggregate claim after  $j + m$  development years amount knowing the information up to time  $s = k + j + 1$  is

$$E\left(S_{k,j+m} \mid \mathcal{F}_s^{N,C,X}\right) = \sum_{i=1, i \in \mathcal{C}_{s,k}^s}^{N_{k+j+1}} X_i + \sum_{i=1, i \in \mathcal{C}_{s,k}^{ns}}^{N_{k+j+1}} X_i \tilde{G}_k(s+m, T_i), \quad (4.13)$$

where  $\tilde{G}_k(t, u) = E(g(y, t, V) \mathbb{1}_{\{k \leq t < k+1\}})$ . Using Proposition 4.3.2, the conditional second-order moment of the aggregate claim after  $j + m$  development years amount knowing the information up to time  $s = k + j + 1$  is

$$E(S_{k,j+m}^2 | \mathcal{F}_s^{N,C,X}) = \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} (X_i X_j \mathbb{1}_{i,j \in \mathcal{C}_{s,k}^s} + 2X_i X_j \tilde{G}_k(s+m, T_i) \mathbb{1}_{i \in \mathcal{C}_{s,k}^{n_s}, j \in \mathcal{C}_{s,k}^s}) + \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} X_i X_j \tilde{G}_k(s+m, T_i) \tilde{G}_k(s+m, T_j) \mathbb{1}_{i,j \in \mathcal{C}_{s,k}^{n_s}}, \quad (4.14)$$

where  $\tilde{G}_k(t, u) = E(g(y, t, V) \mathbb{1}_{\{k \leq t < k+1\}})$ .

Dutang and Brouste (2016) illustrate these methods on a real insurance dataset from an unknown private insurer on a portfolio of general third-party liability policies for private individuals. These numerical applications reveal that the proposed framework underestimates the ultimate claim charges (assuming the Chain-Ladder method is the most appropriate method). On a one-year time horizon, the backtesting procedure shows that the new method to estimate claim charges performs reasonably well.

Future research of Dutang and Brouste (2016) should to take reporting delays into account as well as random claim charge in order to better assess the reserving risk. By considering random reporting delays, the observed claim process is no longer a Poisson process. This could be better tackled with a general renewal process for the claim process, see Asmussen and Albrecher (2010, Chapter 6), an inhomogeneous Poisson process, see Wuethrich and Merz (2008, Chapter 10), or a marked Poisson process as in Norberg (1993).

## 4.4 R implementation

As presented in Dutang et al. (2008), **actuar** contains functions related to the risk theory problems: two for the calculation of the aggregate claim amount distribution and two for infinite-time ruin probability calculations, see Section 2.2 for loss modeling. Currently, five methods are supported for the computation of the aggregate claim distribution function: recursive calculation using the algorithm of Panjer, exact calculation by numerical convolutions, Normal and Normal Power II approximations, simulation, see, e.g., Klugman et al. (2012).

Regarding ruin considerations, **actuar** propose a function to compute closed-form expressions of the ruin probability (4.2) based on phase-type distributions (Neuts, 1981), which includes the Sparre Andersen model and requires the computation of matrix exponentials. In that respect, the **expm** (Goulet et al., 2021b) package implements matrix operations (exponential or logarithm). **actuar** has also a function to determine the adjustment coefficient, see Section 4.1, in the Sparre Andersen model as well as with proportional or excess-of-loss reinsurance (Centeno, 2002a,b). Competitors packages on ruin topics are few. **bootruin** (Baumgartner and Gatto, 2010) provides a single function to compute ruin probabilities or estimate ruin probabilities from data in the Cramér-Lundberg model. **finiteruinprob** (Gatto and Baumgartner, 2016) provides a function to compute finite-time ruin probabilities in the Cramér-Lundberg pertubated by a Brownian motion. Finally, **ruin** (Rudnytskyi, 2018) provides functions to simulate random paths of Sparre Andersen model in order to estimate ruin probabilities.

Regarding reserving methods in non-life insurance, only the R package **ChainLadder** (Gesmann et al., 2021) provides functions for reserving methods. Currently, only aggregated methods are implemented in this package.



## Chapter 5

# Customer models and non-cooperative games for insurance markets

This chapter investigates models of policyholders behavior and insurers strategy. Section 5.1 deals with customer behavior models for both life and non-life insurance. Section 5.2 makes a review of game theory useful for the subsequent sections. Section 5.3 proposes a one-period game for non-life insurers for which policyholders behavior is based on Section 5.1. Finally, Section 5.4 presents a multi-period game based on a repeated version of the one-period game of Section 5.3.

### 5.1 Customer behavior models

Every year, insurers face the recurring question of adjusting premiums for renewal or of setting new rates for new business. In price elasticity studies, one analyzes how customers react to price changes. Where is the trade-off between increasing premium to favor higher projected profit margins and decreasing premiums to obtain a greater market share? We must strike a compromise between these contradictory objectives. The price elasticity and customer modeling are therefore a factor to contend with in actuarial and marketing departments of every insurance company.

In order to target new market shares (new business) or to retain customers in the portfolio (in-force policyholders), it is essential to assess the impact of pricing on the whole portfolio. To avoid a portfolio-based approach, we must take into account the individual policy features. Moreover, the methodology to estimate the price elasticity of an insurance portfolio must be sufficiently refined to identify customer segments.

Customer modeling is not a new topic in actuarial literature. In non-life insurance, two ASTIN workshops (Bland et al., 1997; Kelsey et al., 1998) were held in the 90's to analyze customer retention and price/demand elasticity topics. Shapiro and Jain (2003) also devote two chapters of their book to price elasticity: Guillen et al. (2003) use logistic regressions, whereas Yeo and Smith (2003) consider neural networks. Brockett et al. (2008) should also be mentioned for their use of survival regression models to study non-life insurance lapse. Jeong et al. (2018) find empirical evidences of possible association between policyholder switching after a claim and the associated premium change using association rules. Le Faou (2019) is devoted to survival models with censorship in order to determine lapse in health insurance. We refer also to Leiria et al. (2020) for a review.

In the context of life insurance, the topic is even more complex as the lapse can occur at any time, whereas for non-life policies, most lapses occur at renewal dates\*. There are some trigger effects due to contractual constraints: penalties are enforced when lapses occur at the beginning of the policy duration, while after that period, penalties no longer apply. Another influential feature is the profit benefit option of some life insurance policies allowing insurers to distribute part of benefits to customers in a given year. This benefit option stimulates customers to shop around for policies with higher profit benefits.

In terms of models, Atkins and Gallop (2007) and Kagraoka (2005) use counting process to model surrenders of life insurance, while Cox and Lin (2006) and Kim (2005) use a logistic regression to predict the lapse. Others (Bacinello, 2005; Kuen, 2005) developed financial methods to price the surrender option embedded in life insurance policies. Milhaud et al. (2011) point out relevant customer segments when using Classification And Regression Trees models (CART) and logistic regression. Loisel and Milhaud (2011) study the copycat behavior of insureds during correlation crises. We refer to Eling and Kochanski (2013) for a review.

### 5.1.1 Lapse models for non-life insurance products

Dutang (2012b) deals with lapse risk for non-life insurance products, i.e., binary model where the response variable indicates if a current policyholder lapses or cancels its policy. We refer to Section 1.1.1 for logistic regression. Dutang (2012b) applies GLM on two datasets from continental Europe: a dataset with few explanatory variables and a large dataset.

Predictions are analyzed through the estimation of additional lapse rates. Indeed, the price elasticity defined as the customer's sensitivity to price changes relative to their current price, is computed as the normed derivative  $e_r(p) = \frac{dr(p)}{dp} \times \frac{p}{r(p)}$ , where  $r(p)$  denotes lapse rate as a function of the price  $p$ . However, in practice, we focus on the additional lapse rate  $\Delta_{dp} = r(p+dp) - r(p)$  rather  $e_r(p)$  since the results are more robust and easier to interpret<sup>†</sup>.

Getting outputs is easy but having reliable estimates is harder. After a naive application, the first dataset reveals to have too few variables<sup>‡</sup> to get realistic estimate of additional lapse rates  $\Delta_{5\%}$ . The second dataset with additional variables (distribution channel, coverage type, ...<sup>§</sup>) also leads to unrealistic estimate, even when splitting data by distribution channel and coverage types.

GLM estimates of central lapse rates<sup>¶</sup> begin to make sense when new price-related variables are introduced. That is the rebate level, a proxy of the market premium defined as the tenth lowest price. Incorporating new key variables in the GLM regression substantially improves the lapse rate predictions in the different premium scenarios. The rebate level partially reveals the agent or the broker actions on the customer decisions, while the use of market proxies illustrates how decisive the competition level is when studying customer price-sensitivity.

Secondly, Dutang (2012b) propose to use generalized additive models (GAM) which consists in assuming non-smooth function for predictor  $\eta_i$  rather than a linear function, see Section 1.5.1 for

---

\*In France, most non-life policies are subjected to tacit renewal, yet the new law Hamon allows policyholders to cancel their contract when they want.

<sup>†</sup>The base premium considered is the premium by the policyholder in the previous so that  $dp = 0$  means that no premium increase nor decrease at renewal.

<sup>‡</sup>Only policy age, lapse age in addition to past and proposed premium are available.

<sup>§</sup>Three channels are available Direct, Broker and Tied agent; three covers are sold Full comprehensive, partial comprehensive and third-part liability.

<sup>¶</sup>Lapse rates when the proposed premium is the same as last year.

details. Dutang (2012b) uses the REML criterion to determine the appropriate  $\lambda$  and thin plate basis regression. The thin plate regression uses a basis of thin plate (also known as polyharmonic functions) functions  $\phi_{md}(r) = \alpha_{md}r^{2m-d}\log(r)$  if  $d$  is even and  $\alpha_{md}r^{2m-d}$  if  $d$  is odd. The smooth function is defined as  $s(x) = \sum_{i=1}^n \delta_i \phi_{md}(\|x - x_i\|_2)$  is approximated during the fitting procedure to decrease the computational burden. This method avoids the knot placement problems of traditional regression spline models.

Using GAM on the large dataset reveals useful for the tied-agent channel where highly nonlinear functions are estimated for the following variables: premium ratio, driver age, vehicle age, the difference to technical premium and the car class. The additional complexity coming with additive modeling compared to GLM permit to really fit the data as long as the dataset has a sufficiently large size.

Thirdly, Dutang (2011) consider the use of survival regression models, presented in Section 1.5.2, on a third dataset over four years. Unlike the latter static regression models such as GLM and GAM, we no longer consider the lapse rate as the target random variable but derive it from the life span of the policy. That is, the lapse rate of policies of age  $t$  is

$$r_t = P(T < t + 1 | T \geq t),$$

where  $T$  denotes the life span of the policy. Dutang (2011) considers full parametric survival regression models (SRM) such as accelerated failure time models as well as semi-parametric Cox-PH regression models. Cox models have the advantage to take into account the dynamic aspects of a policy life along with dynamic explanatory variables, unlike GLM and GAM. On the third dataset, a long data pre-processing is needed to transform to a survival dataset. We observe that lapse rate estimates of Cox-PH are relatively similar to others models. Part of this study was published in Dutang (2012b) and pursue the work initiated by Brockett et al. (2008) to use survival models for customers behavior models.

### 5.1.2 Conversion models for non-life insurance products

In a similar spirit as for renewal topics, conversion models a binary response variable indicating if the quote is converted or not. Dutang and Spedicato (2018) aim to investigate how machine learning methodologies can improve policyholder retention and conversion estimation over that of classical GLM. They use a dataset of 1.2 million quotes for private motor insurance with several explanatory variables. A certain number of models (in addition to GLM) are used Gradient Boosting Models (GBM), Extreme Gradient Boosting Models (XGBoost), Random Forests and neural networks, see, e.g., Denuit et al. (2019b,c).

Ranking the models according to log-loss metrics, it is clear that boosted models (GBM and XGBoost) show the highest performance in terms of predictive accuracy\*. Dutang and Spedicato (2018) also conduct an individual pricing optimization which consists in maximizing the expected profit. In their application, the company's knowledge of consumer behavior is represented by a risk premium model that estimates the expected cost of the insurance cover that will be provided, as well as by a conversion model, which estimates the probability of a prospect entering the portfolio. Also, information on the competitive environment such as the distance between market price and the company's premium should be taken into account. Interestingly, Dutang and Spedicato (2018) found that both GLM and XGBoost approaches produced very similar results in terms of optimized premium volume.

---

\*This is also true for Area Under the Curve metric.



Surprisingly, neural networks do not perform well on this dataset. This might be due the software implementation. Dutang and Spedicato (2018) use the **h2o** package (LeDell et al., 2021) which is not the best implementation of neural networks. Mahdi et al. (2021)'s benchmark of neural networks in R packages shows that **h2o** obtains only the 11th RMSE score and the 51st time score out of 60 package-algorithms tested.

### 5.1.3 Surrender models for life insurance products

Dutang and Milhaud (2018) deals with the crucial problem of modeling policyholders' behaviors in life insurance. They focus on the surrender behaviors and model the contract lifetime through the use of survival regression models. Standard models may fail at giving acceptable forecasts for the timing of surrenders because of too much heterogeneity. The competing risk framework (see Section 1.5.2) provides interesting insights and more accurate predictions by considering the single cause of interest when estimating the hazard rate (and not all causes).

Numerical results are carried out on an insurance portfolio embedding Whole Life contracts in US. This dataset contains three type of exit causes: the (voluntary) surrender by the policyholder, the death of the policyholder, other causes mixed together including payment default. Regarding the cause of interest, the estimation via the subdistribution approach (see Section 1.5.2) is slightly better than with the cause-specific approach. This is probably due to uncertain estimations of hazard rates for other causes (death and other), which are fewly observed on this dataset.

Through backtests, this framework reveals to be quite efficient and recovers the empirical lapse rate trajectory by aggregating individual predicted lifetimes. The overall quality of the Fine & Gray model is rather satisfying when considering the timing of individual surrender decisions. In an Assets & Liabilities Management perspective, this allows to give precise information about the future insurer's cash-flows, still keeping in mind that portfolio composition evolves as time flies and thus the surrender risk has to be updated on a regular basis. At individual levels, the Fine & Gray model generally looks much more robust than other modeling strategies belonging to competing risk models.

Premium and reserve computation associated to life insurance products can be carried out in R thanks to packages **lifecontingencies** (Spedicato et al., 2021), **DetLifeInsurance** (Auza and Alvarez, 2020), and **LifeInsuranceContracts** (Kainhofer, 2021) see Spedicato (2013) for an overview. **valuer** (Zoccolan, 2018) implements the pricing frameworks taking into account surrenders described in Bacinello et al. (2011) and is currently only the one package implementing surrender risk.

## 5.2 Games literature review

Game theory focuses on the study of complex interactions between agents (men, companies, animals, ...) modeled by tailored mathematical models which are enabled to catch decision making of agents for a given topic. The fundamental principles underlining game theory is that agents, also known as players, take into account other players' behaviors, as opposed to individualist point of view of optimal control.

Game theory dates back to the study of economic oligopolies studied by Bertrand (1883), Cournot (1838), and Edgeworth (1881). It has been popularized and becomes a scientific discipline unto itself thanks to von Neumann and Morgenstern (1944), who studied zero-sum multiplayer games both cooperative and non-cooperative. Few years later, John F. Nash made a breakthrough in a series of papers (Nash, 1950a,b, 1951, 1953) to propose new game concepts and to study their existences. Since

then, game theory grows in multiple direction and is no longer only applied to economy. We refer to Basar and Olsder (1999), Fudenberg and Tirole (1991), and Osborne and Rubinstein (2006) for a general introduction.

A game is a formal description of interactions between players and is composed of a set of players  $E = \{1, \dots, I\}$ , an objective or cost function for each player  $O_i : X \mapsto \mathbb{R}$ , and a set of possible actions per player  $X_i \subset \mathbb{R}^{n_i}$  for  $i \in E$ , where  $X = X_1 \times \dots \times X_I^*$ . A set of actions per player is denoted  $\mathbf{x} = (x_1, \dots, x_I)$ . Game concepts consists in specifying a set of actions  $\mathbf{x}$  which is preferable to  $\mathbf{y}$  for a given player. There exists multiple class of games allowing different type of player interactions: are action played sequentially or simultaneously; can players form alliances or do they play against each other; does information perfectly known ; does objective function depend on a random factor? We present with non-cooperative deterministic games in Section 5.2.1, cooperative games in Section 5.2.2. This section ends with a review of actuarial games in Section 5.2.3.

### 5.2.1 One-shot non-cooperative games

We assume that  $X_i$  is a compact convex nonempty set representing all possible actions of Player  $i$ , say, e.g., a price interval or a quantity interval<sup>†</sup>. Given  $i \in E$  a player,  $x_i$  denotes his action, while  $x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_I)$  stands for other players' actions. Both Stackelberg and Nash equilibria are used in Section 5.3. A Nash equilibrium is defined as follows.

**Definition 5.2.1 (Nash equilibrium)** *For  $I$ -player game with  $O_i, i \in E$  the objective function of Player  $i$  to maximize, a vector of strategies  $(x_1^*, \dots, x_I^*) \in X$  is a Nash equilibrium if for all  $i \in E$ , we have*

$$O_i(x_i^*, x_{-i}^*) \geq O_i(x_i, x_{-i}^*), \text{ for all } x_i \in X_i. \quad (5.1)$$

In other words, a Nash equilibrium can hence be interpreted as a point at which no player has an incentive to deviate, given the actions of the other players. To better understand existence theorem, the Nash equilibrium problem (NEP) (5.1) is reformulated as  $x^*$  verifying for all  $i \in E$

$$x_i^* \in \arg \max_{x_i \in X_i} O_i(x_i, x_{-i}^*).$$

Now, we state formally the existence theorem for NEP (5.1).

**Theorem 5.2.1 (Nikaido and Isoda (1955))** *For a  $I$ -player game, where  $X_i$  are non-empty convex and compact, if for all objective functions  $O_i : X \mapsto \mathbb{R}$  are continuous and  $x_i \mapsto O_i(x_i, x_{-i})$  are quasiconvex, see Table 5.1, there exists a Nash equilibrium.*

Concept	Definition
$f$ quasiconvex	$f(\lambda x + (1 - \lambda)y) \leq \max(f(x), f(y))$
$f$ convex	$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$
$f$ strictly convex	$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y)$
$\forall x, y \in \mathbb{R}, \forall \lambda \in ]0, 1[$	

Table 5.1 – Different types of convexity, e.g., Diewert et al. (1981)

To guarantee unicity of the Nash equilibrium, we need much stricter conditions than quasiconvexity of objective functions. The following theorem of Rosen (1965) gives us some sufficient conditions.

\*Note that  $X_i$  is not necessarily finite nor discrete.

<sup>†</sup>For a discrete set of actions, it is common to consider mixed strategies by playing randomly among pure strategies.

**Theorem 5.2.2 (Rosen (1965))** For a  $I$ -player game, where  $X_i$  are non-empty convex and compact, assume that for all objective functions  $O_i : X \mapsto \mathbb{R}$  are continuous and  $x_i \mapsto O_i(x_i, x_{-i})$  are convex. Let  $g_O : \mathbb{R}^n \times \mathbb{R}^I \mapsto \mathbb{R}^n$  defined by

$$g_O(\mathbf{x}, \mathbf{r}) = \begin{pmatrix} r_1 \nabla_{x_1} O_1(\mathbf{x}) \\ \vdots \\ r_I \nabla_{x_I} O_I(\mathbf{x}) \end{pmatrix}.$$

If  $g_O$  verifies for some  $\mathbf{r}$  with positive components

$$(\mathbf{x} - \mathbf{y})^T g_O(\mathbf{y}, \mathbf{r}) + (\mathbf{y} - \mathbf{x})^T g_O(\mathbf{x}, \mathbf{r}) > 0, \quad (5.2)$$

then there exists a unique Nash equilibrium.

In practice, the action set  $X_i$  may be derived from a univariate constraint function  $g_i$ , that is

$$X_i = \{x_i \in \mathbb{R}, g_i(x_i) \geq 0\}.$$

The closedness and the convexity of  $X_i$  is verified when  $g_i$  is continuous and quasiconcave. The nonemptiness is harder to verify in general.

Now, we turn to the definition of another type of game solution: a Stackelberg equilibrium. Stackelberg equilibrium focuses on the situation where players do not play at the same time: some (the leaders) play before others (the followers). For simplicity, we assume there is one unique leader and other players  $j \in \{2, \dots, I\}$  are followers.

**Definition 5.2.2 (Stackelberg equilibrium)** For a game with one leader and  $I - 1$  followers, with payoff functions  $O_j$  and action set  $X_j$ , a Stackelberg equilibrium is a vector  $\mathbf{x}^* = (x_1^*, \dots, x_I^*)$  such that  $x_1^*$  solves the problem

$$\sup_{x_1 \in X_1} O_1(x_1, x_{-1}^*(x_1)),$$

and  $x_{-1}^*(x_1)$  is a Nash equilibrium for the subgame with the  $I - 1$  followers and given action  $x_1$  for Player 1.

We refer to Lucchetti et al. (1987) for existence theorems of a Stackelberg equilibrium. Finally, we close our brief presentation of non-cooperative games with generalized Nash equilibrium. A generalized Nash equilibrium (GNE) is defined as follows.

**Definition 5.2.3 (generalized Nash equilibrium)** For  $I$ -player game with  $O_i, i \in E$  the objective function of Player  $i$  to maximize, a vector of strategies  $(x_1^*, \dots, x_I^*) \in X$  is a Nash equilibrium if for all  $i \in E$ , we have

$$O_i(x_i^*, x_{-i}^*) \geq O_i(x_i, x_{-i}^*), \text{ such that } g_i(x_i, x_{-i}^*) \geq 0. \quad (5.3)$$

In the following, the set of possible actions of Player  $i$  given other players' action  $x_{-i}$  is derived using a correspondence

$$C_i(x_{-i}) = \{x_i \in X_i, g_i(x_i, x_{-i}) \geq 0\}, \quad (5.4)$$

where  $g_i$  is the constraint function. We give below the existence theorem for GNE, sometimes called Kakutani's theorem, since the proof is based on the fixed-point theorem for correspondences by Kakutani (1941).

**Theorem 5.2.3 (Arrow and Debreu (1954))** *Let  $N$  players be characterized by an action space  $X_i$ , a constraint correspondence  $C_i$  and an objective function  $O_i : X \rightarrow \mathbb{R}$ . Assume for all players, we have*

- (i)  $X_i$  is nonempty, convex and compact subset of a Euclidean space,
- (ii)  $C_i$  is both upper and lower semi-continuous in  $X_{-i}$ ,
- (iii)  $\forall x_{-i} \in X_{-i}$ ,  $C_i(x_{-i})$  is nonempty, closed, convex,
- (iv)  $O_i$  is continuous on the graph  $Gr(C_i)$ ,
- (v)  $\forall x \in X$ ,  $x_i \rightarrow O_i(x_i, x_{-i})$  is quasiconcave on  $C_i(x_{-i})$ ,

*Then there exists a generalized Nash equilibrium.*

Dutang (2013b) provide an in-depth study of theorems guaranteeing existence of generalized Nash equilibria and analyze the assumptions on parametric feasible sets (5.4), which is common for most games. For parametric feasible sets, the upper and lower-semicontinuity of  $C_i$  is satisfied when  $g_i$  has continuous components and is concave with respect to  $x_i$ .

In practice, many softwares propose packages to compute game solution concepts. In R, various packages implement game theoretic tools. The **CoopGame** (Staudacher and Anwander, 2019) provides a large collection of cooperative games, like e.g. bankruptcy games, cost sharing games and weighted voting games. The **GameTheory** (Cano-Berlanga et al., March 2015) proposes functions to compute of two solutions concepts, namely the Shapley value and the nucleolus, whereas **GameTheoryAllocation** (Saavedra-Nieves, 2016) proposes more game concepts. The **GNE** package (Dutang, 2020a) provides functions to compute standard and generalized Nash Equilibria. Dutang (2013a) deals with optimization methods for solving GNE problems and use the **GNE** package for benchmarking purposes.

## 5.2.2 Cooperative games

Cooperatives games (with  $n$  players) are characterized by a pair of  $(N, v)$  with a set of players  $N = \{1, \dots, n\}$  and a function  $v : 2^N \mapsto \mathbb{R}$  where  $2^N$  denotes the set of all subsets of  $N$  and  $v(S)$  the power or the value of coalition  $S^*$ . Best coalitions have simply the supremum value of  $v$ . In general, we assume that  $v$  is superadditive since cooperation creates savings.

Assuming a best coalition is found, cooperative games focuses on how savings are redistributed among players of the best coalition. To that purpose, the concept of (individual) payoffs is introduced with a  $n$ -dimensional vector  $\alpha = (\alpha_1, \dots, \alpha_n)$  where  $\alpha_i$  is the redistributed value to Player  $i$  after the coalition. A list of the minimal desirable properties of payoffs  $\alpha$  is: *efficiency*  $\sum_{i \in S} \alpha_i = v(S)$ ; *individual rationality*  $\forall i, \alpha_i \geq v(\{i\})$ ; *collective rationality*  $\forall S \in N, \sum_{i \in S} \alpha_i \geq v(S)$ . The individual rationality and the collective rationality form the so-called stability condition. Defining an *imputation* as an individually efficient rational payoff, the *core* of a game is the set of all collectively rationally payoffs.

There are some other desirable properties (or axioms) of payoff: *symmetry* for all permutations  $\Pi, v(\Pi(S)) = v(S) \Rightarrow \forall i, \alpha_{\Pi(i)} = \alpha_i$ ; *inefficient player*  $v(S) = v(S \setminus \{i\}) + v(\{i\}) \Rightarrow \alpha_i = v(i)$ ; *additivity on game* for two games  $(N, v)$  and  $(N, w)$ ,  $\forall i, \alpha_i(v + w) = \alpha_i(v) + \alpha_i(w)$ ; *monotonicity*  $\forall S \subset T, v(S) \leq v(T)$ . Table 5.2 lists common solution concepts. The properties verified by solution

---

\*The dual representation of the game  $(N, v)$  uses a cost function  $c$  instead of a value function  $v$ , they are related one another by  $v(S) = \sum_{s \in S} c(s) - c(S)$ .

concepts of Table 5.2 are given in Table 5.3: the Shapley’s value is the only one solution satisfying the last three properties.

Solution concept	Optimal payoff $\alpha_i^*$
Moriati method	$\frac{v(\{i\})}{\sum_{i=1}^n v(\{i\})} v(N)$
Shapley’s value	$\sum_{S \subset N} \frac{(\text{Card}(S)-1)!(n-\text{Card}(S))!}{n!} (v(S) - v(S \setminus \{i\}))$
method of residual profits	$v(\{i\}) + \frac{v(\{i\})+MV(\{i\})}{\sum_{j=1}^n (v(\{j\})+MV(\{j\}))} \left( v(N) - \sum_{j=1}^n v(\{j\}) \right)$

$MV$  is the marginal value  $MV(\{i\}) = v(N) - v(N \setminus \{i\})$ .

Table 5.2 – Solution concepts

Property	Moriati method	Shapley’s value	residual profit
efficiency	yes	yes	yes
individual rationality	yes	yes	yes
collective rationality	no	yes	yes
monotonicity	yes	yes	yes
symmetry	yes	yes	yes
inefficient player	no	yes	yes
additivity on game	no	yes	no

Table 5.3 – Properties of solution concepts

Another type of games of interest are repeated games which consists in repeating a one-shot game over and over either in a finite or infinite horizon. Three versions of repeated games could be considered: (i) infinitely repeated games with no discounting, (ii) infinitely repeated games with discounting, and (iii) finitely repeated games. Different solution concepts are studied in the literature, typically punishing strategies, subgame perfect equilibrium, perfect equilibrium, which are characterized through their payoff vectors, see Osborne and Rubinstein (2006, Chapter 8) or Gossner and Tomala (2020). We consider repeated games as not valuable for modeling insurer strategies since the set of player’s actions is usually finite, it is very questionable which horizon to choose and which discount factor is appropriate. For these reasons, Sections 5.3 and 5.4 do not consider concepts from repeated games theory.

### 5.2.3 Review of actuarial game models

We make a review of actuarial game models in the past sixty years in order to position our models of Sections 5.3 and 5.4 in relation to the literature. We summarize main ideas below starting with contributions of Karl Borch, Jean Lemaire, Hans Gerber.

#### Borch – Lemaire – Gerber’s contributions

A series of paper by Borch (1960a,b,c) works risk transfer between two insurers (or 1 insurer - 1 reinsurer) and identify optimal situations with Pareto optimality conditions on expected utility. This is generalized later to  $n$  insurers by Borch (1962b, 1975) In parallel, a use of cooperative game to model risk transfer between  $n$  insurers based on the core and the Shapley value is studied by Borch (1962a), Gerber (1978), and Lemaire (1975, 1977, 1991). Baton and Lemaire (1981a,b) and Lemaire (1979)

introduce a less restrictive model of non transferable utility game (linear invariance, recursiveness, independence of irrelevant alternatives and partial symmetry).

Chains of (re)insurance was also considered by Lemaire and Quairière (1986) using non-cooperative bargaining theory. As a consequence of optimal risk transfer, Briegleb and Lemaire (1982) and Bühlmann (1980, 1984) study the application of game theory to premium principles by noticing that the model of risk transfer between insurers can also be applied to the situation of insurers selling a product to an insured. Briegleb and Lemaire (1982) uses Nash and Kalai-Smorodinsky solutions, whereas Bühlmann (1980, 1984) use Pareto optimality.

Finally, Lemaire (1983) works on cost allocation of insurance expenses among line of businesses by starting to review arbitrary accounting methods. Using cooperative game theory, he derives new allocation methods based on the Shapley value and nucleolus' derivative methods.

### Recent advances on risk transfer between (re)insurers

A series of papers studied risk transfer between insurers and/or reinsurers using cooperative game theory. Boonen et al. (2017a) focus on the redistribution of longevity risk in a 2-insurer cooperative game. Anthropolos and Boonen (2020) investigate a insurer-reinsurer game with risk-measure utility functions. Jiang et al. (2019) study a insurer-reinsurer game of risk transfer thanks to Nash and Kalai-Smorodinsky solutions. Jiang et al. (2021) extends this work by considering distortion risk measure and rank dependent expected utility. Zeng and Luo (2013) consider a differential stochastic cooperative game (i.e. a dynamic game in continuous time) between an insurer and a reinsurer where the surplus process is approximated by a diffusion model and the insurer may purchase quota-share or excess-of-loss covers.

However, many papers studied risk transfer between insurers and/or reinsurers using non-cooperative game theory. They consider differential games to model the insurer-reinsurer risk transfer where the aggregate loss model is approximated by a diffusion model and game solutions are computing using the associated Hamilton-Jacobi-Bellman equations. Chen et al. (2019), Meng et al. (2015), and Pun et al. (2016) determine Nash equilibrium of 2-insurer expected-utility games. Asmussen et al. (2019a,b) consider a 2-insurer game where insurers try to minimize a reserve process either via a Stackelberg equilibrium or via Nash equilibrium, respectively. Kwok et al. (2016) study a 2-insurer game where insurers choose an investment strategy between a risk-free asset and a longevity security. Chen and Shen (2019) determine Stackelberg equilibrium of a insurer-reinsurer game mean-variance criteria. Wang et al. (2021b) extends an extension of the latter where players choose also an investment strategy in a risky asset.

Unlike Chen and Shen (2019) and Wang et al. (2021b), some papers do not consider the reinsurer as a player in the differential game, that is by considering a 2-insurer game where the two insurers are allowed to purchase reinsurance covers from a non-playing reinsurance company, see (Jin et al., 2013; Luo et al., 2021; Pun and Wong, 2016; Taksar and Zeng, 2011; Wang et al., 2019, 2021c; Zeng, 2010; Zhang et al., 2019).

Finally, some papers consider a non-cooperative game with more than 2 insurers, such as differential games of Emms (2012), Li et al. (2021), and Ramasubramanian (2006) or static games of Warren et al. (2012) and Wu and Pantelous (2017).

### Recent advances on the risk transfer from policyholder(s) to (re)insurers

Generally, insured-insurer games consider static strategies. Indeed, Huang et al. (2013), Suijs et al. (1998), and Viaene et al. (2002) use cooperative insured-insurer games for non-life insurance, de Janvry et al. (2014) and Liu (2010) for agricultural insurance, Driessen et al. (2011) for catastrophic risk.

Non-cooperative games are also considered by Bensalem et al. (2020) with an emphasis of prevention and self-insurance, Li and Young (2021) with a mean-variance criterion. As between insurance carriers, the demand for insurance is studied via insured-insurer differential games: Boonen et al. (2018) and Mourdoukoutas et al. (2021) compute Nash equilibrium strategies of non-cooperative games over a finite horizon.

Some papers also study the optimal number of insurance carriers: using the law of large numbers, as in Borch (1962a), Powers and Shubik (1998) found that optimal quantity and price are concave functions of the number of insurers. With similar arguments, Powers and Shubik (2006) show that the optimal number of reinsurers is almost a square root function of the number of insurers.

### Recent advances on capital allocation

Following Lemaire (1983), Alegre and Claramunt (1995) compare actuarial principles and cooperative game theory. Denault (2001) sets up a concept of coherent allocation as the direct application of cooperative game theory concepts. Panjer (2002) studies the problem of capital allocation among line of businesses in an insurance group using on coherent measure (Artzner et al., 1999) and coherent allocation (Denault, 2001) and applies the Shapley value for the banking industry.

Furthermore, he also studies fractional player game, from which he derives a less complex solution, the Aumann-Shapley value, not requiring the additivity axiom. This study was further extended to distortion risk measure in a static framework by Tsanakas and Barnett (2003) and a dynamic framework by Tsanakas (2004) using Brownian motions to model risk process. Boonen et al. (2017b) generalize the latter model by considering Lévy processes for the risk process.

Finally, the problem of insurance loss redistribution can be seen an allocation problem and Denuit and Dhaene (2012) analyze the benefit of conditional mean risk sharing through convex order. Denuit and Robert (2021a,b) consider a peer-to-peer insurance scheme where the higher layer of the aggregate claim amount is transferred to an insurance carrier and retained losses are distributed among insured according to the conditional mean risk sharing rule.

## 5.3 A static game for non-life insurance competition

Dutang et al. (2013a) consider  $J$  insurers competing in a market of  $N$  policyholders with one-year contracts\*. The policyholders are assumed to react to price changes (either stay with the present insurer or switch to one of the competitors), but do not have any other influence on the premium level. This is a realistic assumption, in particular for personal lines of business such as compulsory third-party motor liability, see Section 5.1. In view of the one-year time horizon and the randomness of claim sizes, this model focuses on non-life insurance products.

The “game” for insurers is to set the premium for which policies are sold to the policyholders.

---

\* $N$  is assumed constant.

Let  $(x_1, \dots, x_J) \in \mathbb{R}^J$  be a price vector, with  $x_j$  representing the premium of Insurer  $j$ . Once the premium is set by all insurers, the policyholders choose to renew or to lapse from their current insurer. Then, insurers pay occurring claims during the coverage year. At the end of the period, underwriting results are determined, and the insurer capital is updated: some insurers may be bankrupt. As we deal with a one-period model, investment results is not considered for simplicity. This game differs from other static games proposed in the literature in the fact that policyholders are not players of the game, insurers do not cooperate, bankruptcy can occur either because of large losses or because of losing customers.

In the next subsections, the four components of the game are: (i) a lapse model, (ii) a loss model, (iii) an objective function and (iv) a solvency constraint function. These four components are critical factors for the analysis of the non-life insurance market, see, e.g., IASB (2008). In the sequel, a subscript  $j \in \{1, \dots, J\}$  will always denote an insurer index, whereas a subscript  $i \in \{1, \dots, N\}$  denotes policyholder index. In the sequel, “insurer” is used when referring to players of the insurance game.

### 5.3.1 Lapse model

We start with our lapse model which is designed as a compromise between reflecting the policyholders’ behavior in a reasonable way, yet keeping mathematical tractability. Let  $n_j$  be the initial portfolio size of Insurer  $j$  (such that  $\sum_{j=1}^J n_j = N$ ). It seems natural that the choice of policyholders for an insurer is highly influenced by the choice of the previous period. We assume that the dispatch (among the  $J$  insurers) of the  $n_j$  policyholders of Insurer  $j$  follows an  $J$ -dimensional multinomial distribution  $\mathcal{M}_J(n_j, p_{j \rightarrow}(\mathbf{x}))$  with probability vector  $p_{j \rightarrow}(\mathbf{x}) = (p_{j \rightarrow 1}(\mathbf{x}), \dots, p_{j \rightarrow J}(\mathbf{x}))$ . The probability  $p_{j \rightarrow k}(\mathbf{x})$  to move from Insurer  $j$  to Insurer  $k$  naturally depends on the price vector  $x$ , (concretely, the difference of premiums). Empirically, the probability to lapse  $p_{j \rightarrow k}(\mathbf{x})$  (with  $k \neq j$ ) is generally much lower than the probability to renew  $p_{j \rightarrow j}(\mathbf{x})$ .

In the economics literature,  $p_{j \rightarrow k}$  is considered in the framework of discrete choice utility models, e.g., McFadden (1981). In that respect, policyholders could be seen as followers of a leader-follower game where leaders are insurers. In the sake of simplicity, we choose a multinomial logit model\*. Working with unordered choices, we arbitrarily set the insurer reference category for  $p_{j \rightarrow k}$  to  $j$ , the current insurer. We define the probability for a customer to go from Insurer  $j$  to Insurer  $k$  given the price vector  $\mathbf{x}$  by the multinomial logit model

$$p_{j \rightarrow k}(\mathbf{x}) = \begin{cases} \frac{1}{1 + \sum_{l \neq j} e^{f_j(x_j, x_l)}} & \text{if } j = k, \\ \frac{e^{f_j(x_j, x_k)}}{1 + \sum_{l \neq j} e^{f_j(x_j, x_l)}} & \text{if } j \neq k, \end{cases} \quad (5.5)$$

where the sum is taken over the set of insurers  $\{1, \dots, J\}$  and  $f_j$  is a price-sensitivity function. Two types of price functions are studied by Dutang et al. (2013a)

$$\bar{f}_j(x_j, x_l) = \bar{\mu}_j + \bar{\alpha}_j \frac{x_j}{x_l} \quad \text{and} \quad \tilde{f}_j(x_j, x_l) = \tilde{\mu}_j + \tilde{\alpha}_j (x_j - x_l). \quad (5.6)$$

The first function  $\bar{f}_j$  assumes a price-sensitivity according to the ratio of proposed premium  $x_j$  and competitor premium  $x_l$ , whereas  $\tilde{f}_j$  works with the premium difference  $x_j - x_l$ . Parameters  $\mu_j, \alpha_j$

---

\*the probit link function, based on the multivariate normal distribution, would not significantly change the shape of the lapse function.



represent a base lapse level and price-sensitivity, respectively. We assume that insurance products display positive price elasticity of demand  $\alpha_j > 0$ .

The vector of (random) portfolio sizes  $\mathbf{N}(\mathbf{x})$  is obtained by summing the (independent) choices of each Insurers' customers

$$\mathbf{N}(\mathbf{x}) = \sum_{j=1}^J \bar{\mathbf{C}}_j(\mathbf{x}) = \begin{pmatrix} \bar{C}_{1,1}(\mathbf{x}) \\ \vdots \\ \bar{C}_{1,J}(\mathbf{x}) \end{pmatrix} + \cdots + \begin{pmatrix} \bar{C}_{J,1}(\mathbf{x}) \\ \vdots \\ \bar{C}_{J,J}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} N_1(\mathbf{x}) \\ \vdots \\ N_J(\mathbf{x}) \end{pmatrix},$$

where  $\bar{\mathbf{C}}_j(\mathbf{x})$  follows  $\mathcal{M}_J(n_j, p_{j \rightarrow}(\mathbf{x}))$ . For Insurer  $j$ , his portfolio size  $N_j$  is a sum of independent binomial variables with parameters  $n_k, p_{k \rightarrow j}(\mathbf{x})$ . It is important to note that the insurers' portfolio sizes are not independent, since the total market size remains constant. This assumption is in contrast to the standard models in classical ruin theory, where the portfolio size is assumed constant over time (Asmussen and Albrecher, 2010). With this insurer choice model, the expected portfolio size of insurer  $j$  reduces to

$$\mathbb{E}(N_j(\mathbf{x})) = n_j \times p_{j \rightarrow j}(\mathbf{x}) + \sum_{l \neq j} n_l \times p_{j \rightarrow l}(\mathbf{x}). \quad (5.7)$$

### 5.3.2 Loss model

Let  $Y_i$  be the aggregate loss of policy  $i$  during the coverage period. We assume no adverse selection among policyholders of any insurers, i.e.  $Y_i$  are independent and identically distributed (i.i.d.) random variables, for all  $i = 1, \dots, n$ . Let us assume a simple frequency – average severity loss model, e.g., (Klugman et al., 2012)

$$Y_i = \sum_{l=1}^{M_i} Z_{i,l},$$

where the claim number  $M_i$  is Policyholder  $i$ . The claim severities  $(Z_{i,l})_l$  are i.i.d. as some generic random variable  $Z$ . The aggregate claim amount for Insurer  $j$  is then

$$S_j(\mathbf{x}) = \sum_{i=1}^{N_j(\mathbf{x})} Y_i = \sum_{i=1}^{N_j(\mathbf{x})} \sum_{l=1}^{M_i} Z_{i,l},$$

where  $N_j(\mathbf{x})$  is the portfolio size. Two claim number distributions are studied: (i)  $M_i$  follows a Poisson distribution  $\mathcal{P}(\lambda)$  and (ii)  $M_i$  follows a negative binomial distribution  $\mathcal{NB}(r, p)$ . These instances of the frequency – average severity model are such that the aggregate claim amount  $S_j(\mathbf{x}) = \sum_{i=1}^{N_j(\mathbf{x})} Y_i$  given  $N_j(\mathbf{x}) = n$  is still a compound distribution of the same kind. Hence, the aggregate claim amount  $S_j(\mathbf{x})$  of Insurer  $j$  is a compound distribution

$$S_j(\mathbf{x}) = \sum_{l=1}^{\widetilde{M}_j(\mathbf{x})} Z_l,$$

where all claim severities  $Z_l \stackrel{i.i.d.}{\sim} Z$  and the claim number  $\widetilde{M}_j(\mathbf{x})$  follows either a Poisson distribution  $\mathcal{P}(n\lambda)$  or a negative-binomial distribution  $\mathcal{NB}(nr, p)$  given  $N_j(\mathbf{x}) = n$ , e.g., (Klugman et al., 2012). For the distribution of  $Z$ , we choose the lognormal distribution, which is due to its heavy-tailedness and at the same time pleasant statistical properties is a popular modeling assumption in many lines of non-life insurance. This choice is by no means crucial, as any other can be handled in an analogous way.

### 5.3.3 Objective function

Two components of the insurance markets have been presented the lapse model (how policyholders react to premium changes) and the loss model (how policyholders face claims). We now turn our attention to the underwriting strategy of insurers, i.e., on how they set premiums. In Section 5.3.1, we assume that price elasticity of demand for the insurance product is positive. Thus, if the whole market underwrites at a loss, any actions of a particular insurer to get back to profitability will result in a reduction of his business volume. This has two consequences for the choice of the objective function: (i) it should involve a decreasing demand function of price  $x_j$  given the competitors price vector  $x_{-j} = (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_J)$  and (ii) it should depend on an assessment of the insurer break-even premium  $\pi_j$  per unit of exposure.

The parameter  $\pi_j$  corresponds to the estimated mean but depends on the assessment of loss expectation by insurer  $j$ . We thus define  $\pi_j$  as

$$\pi_j = \omega_j \bar{a}_{j,0} + (1 - \omega_j) \bar{m}_0,$$

where  $\bar{a}_{j,0}$  is the actuarial premium based on the past loss experience of Insurer  $j$ ,  $\bar{m}_0$  is the market premium, available for instance, via rating bureaus or through insurer associations and  $\omega_j \in [0, 1]$  is the credibility factor of Insurer  $j$ .  $\omega_j$  reflects the confidence of Insurer  $j$  in its own loss experience: the closer to 1, the more confident Insurer  $j$  is. Note that  $\pi_j$  takes into account expenses implicitly via the actuarial and the market premiums.

Dutang et al. (2013a) choose the demand function as

$$D_j(\mathbf{x}) = \frac{n_j}{N} \left( 1 - \beta_j \left( \frac{x_j}{m_j(\mathbf{x})} - 1 \right) \right), \quad (5.8)$$

where  $\beta_j > 0$  is the elasticity parameter and  $m_j(\mathbf{x})$  is a market premium proxy. The demand  $D_j(\mathbf{x})$  is not restricted to  $[0, n_j/n]$ , and thus  $D_j$  targets both renewal and new business. In this form,  $D_j(\mathbf{x})$  approximates the expected market share  $\mathbb{E}(N_j(\mathbf{x}))/N$  (5.7) presented in Section 5.3.1. As the elasticity parameter  $\beta_j$  is positive, a premium increase (of insurer  $j$ ) will result in a decrease of the demand for insurance. The market proxy used in Equation (5.8) is the mean price of the other competitors

$$m_j(\mathbf{x}) = \frac{1}{J-1} \sum_{k \neq j} x_k.$$

The market proxy aims to assess other insurer premiums without specifically targeting one competitor. It can be interpreted as the premium of an ideal medium competitor.

Now we can state our objective function. We suppose that Insurer  $j$  maximizes the expected profit of next year policies which we here define in the multiplicative form

$$O_j(\mathbf{x}) = \frac{n_j}{N} \left( 1 - \beta_j \left( \frac{x_j}{m_j(\mathbf{x})} - 1 \right) \right) (x_j - \pi_j), \quad (5.9)$$

i.e. the product of the demand  $D_j$  and the expected profit per policy, representing a company-wide expected profit. Thus, maximizing the objective function  $O_j$  leads to a trade-off between increasing premium to favor higher projected profit margins and decreasing premium to defend the current market share.

### 5.3.4 Solvency constraint function

Another key feature of the model is a solvency constraint the goal of which is to require insurers to hold a certain amount of capital in order to protect policyholders against adverse collective claim experience. Therefore, in addition to maximizing a certain objective function, insurers must satisfy a solvency constraint imposed by the regulator. A reasonable criterion to find the minimum capital requirement is linked to deviations of the aggregate losses from its expected value, concretely the difference of a high-level quantile and the mean of the loss distribution. For simplicity, this quantity is supposed to be a linear function of the standard deviation of the loss distribution as in Borch (1962a) (see Section 5.2.3). In practice, the solvency capital is also required on a prospective basis; we take here the simplifying assumption to use only the in-force policy number. The solvency constraint function is defined as

$$g_j^1(x_j) = \frac{K_j + n_j(x_j - \pi_j)(1 - e_j)}{k\sigma(Y)\sqrt{n_j}} - 1, \quad (5.10)$$

where  $k$  is the solvency coefficient chosen to approximate a 99.5% quantile and  $e_j$  denotes the expense rate. The numerator corresponds to the sum of the current capital  $K_j$  and the expected profit on the in-force portfolio, whereas the denominator approximates the required capital. The constraint  $g_j^1(x_j) \geq 0$  is equivalent to  $K_j + n_j(x_j - \pi_j)(1 - e_j) \geq k\sigma(Y)\sqrt{n_j}$ .

In addition to the solvency constraint, we need to impose bounds on the possible premium. A first choice could be simple linear constraints as  $x_j - \underline{x} \geq 0$  and  $\bar{x} - x_j \geq 0$ , where  $\underline{x}$  and  $\bar{x}$  represent the minimum and the maximum premium, respectively. The following equivalent reformulation is numerically more stable  $g_j^2(x_j) = 1 - e^{-(x_j - \underline{x})} \geq 0$  and  $g_j^3(x_j) = 1 - e^{-(\bar{x} - x_j)} \geq 0$ .

The bounds  $\underline{x}$  and  $\bar{x}$  could for instance be justified by a prudent point of view of. Precisely, we set  $\underline{x} = \mathbb{E}(Y)/(1 - e_{min}) < \bar{x} = 3\mathbb{E}(Y)$ , where  $e_{min}$  is the minimum expense rate. Summarizing, the constraint function  $g_j(x_j) = (g_j^l(x_j))_{1 \leq l \leq 3}$  for Insurer  $j$  is

$$\{x_j, g_j(x_j) \geq 0\} = \{x_j \in [\underline{x}, \bar{x}], K_j + n_j(x_j - \pi_j)(1 - e_j) \geq k_{995}\sigma(Y)\sqrt{n_j}\}. \quad (5.11)$$

### 5.3.5 Game sequence

For our game with objective function  $O_j$  in (5.9) and constraint function  $g_j$  in (5.11), the game sequence is given as follows.

**Game 1** *Initiate positive capital levels  $K_j$  and positive portfolio sizes  $n$ .*

- (i) *Insurers set their premium according to a Nash or a Stackelberg equilibrium  $\mathbf{x}^*$ .*
- (ii) *Policyholders randomly choose their new insurer according to probabilities  $p_{k \rightarrow j}(\mathbf{x}^*)$  in order to set  $N_j(\mathbf{x}^*)$ .*
- (iii) *For the one-year coverage, claims are random according to the frequency – average severity model relative to the portfolio size  $N_j(\mathbf{x}^*)$ .*
- (iv) *Finally the underwriting result is determined by  $UW_j(\mathbf{x}^*) = N_j(\mathbf{x}^*)x_j^*(1 - e_j) - S_j(\mathbf{x}^*)$  and new capital is  $K_j + UW_j(\mathbf{x}^*)$ , where  $e_j$  denotes the expense rate and  $K_j$  the initial capital value.*

A certain number of theoretical properties of Game 1 has been established by Dutang et al. (2013a). Existence and uniqueness of the Nash equilibrium is given in their Proposition 2.1, while existence of the Stackelberg equilibrium is proved in their Proposition 2.4. Sensitivities of the premium equilibrium with respect to parameters is studied in the following proposition

**Proposition 5.3.1 (Dutang et al. (2013a))** *Let  $x^*$  be the Nash premium equilibrium of the insurance game with  $J$  insurers. For each Insurer  $j$ , the insurer equilibrium  $x_j^*$  with  $x_j^* \in ]\underline{x}, \bar{x}[$  depends on the parameters in the following way: it increases with break-even premium  $\pi_j$ , solvency coefficient  $k_{995}$ , loss standard deviation  $\sigma(Y)$ , expense rate  $e_j$  and decreases with sensitivity parameter  $\beta_j$  and capital  $K_j$ . When  $x_j^* = \underline{x}$  or  $\bar{x}$ , the premium equilibrium is independent of those parameters.*

The proposed Game 1 models a rational behavior of insurers in setting premiums taking into account other insurers. The ability of an insurer to sell contracts is essential for its survival. In terms of equilibrium concepts, the Nash equilibrium is a natural concept when there is no strong leadership, e.g. in private lines, whereas the Stackelberg equilibrium is relevant for lines manifesting oligopolistic tendencies, typically corporate lines. Numerical illustrations show that the Stackelberg premiums are higher than the Nash premiums and that these premium levels become lower when the number of insurers in the market increases.

If we consider a more complex constraint function in Game 1, where the denominator in (5.10) uses the expected portfolio size of Insurer  $j$ , the constraint function becomes a function of  $\mathbf{x}$  (not only  $x_j$ ). This leads to a new type of equilibrium: the generalized Nash equilibrium (5.3). Tested in Dutang (2012a, Chapter 2), existence is still guaranteed but we lose uniqueness. To deal with non-uniqueness, Rosen (1965) studies a subset of generalized Nash equilibrium, where Lagrange multipliers resulting from the Karush-Kuhn-Tucker (KKT) conditions are normalized. Such a normalized equilibrium is unique given a scale of the Lagrange multiplier when the constraint function verifies additional assumptions. Other authors such as von Heusinger and Kanzow (2009) or Facchinei et al. (2007) define normalized equilibrium when Lagrange multipliers are set equal. Hence, complexifying the constraint function (5.10) is not further studied.

## 5.4 A one-period repeated game for non-life insurer strategies

A certain number of extensions of Dutang et al. (2013a) have been proposed by Albrecher and Daily-Amir (2017), who consider a Bayesian framework in order to investigate the effects of asymmetric information as well as Daily-Amir (2019, Chapter 4), which considers different objective functions to model market shares in health insurance studied in Daily-Amir et al. (2019). Moreover, Battulga et al. (2018) consider a  $m$ -period version of the objective function but still investigates the premium equilibrium in a static framework.

Mouminoux et al. (2021b) consider the repeated version of the one-shot Game 1 considered by Dutang et al. (2013a) in Section 5.3. The “game” for each insurer  $j \in \{1, \dots, J\}$  consists in defining a premium  $x_{j,t}$  at the beginning of each year  $t$ , in order to maximize the profit by selling identical policies to the insured market of size  $N$ . Let  $\mathbf{x}_t = (x_{1,t}, \dots, x_{J,t}) \in \mathbb{R}^J$  be the insurers’ price vector, with  $x_{j,t}$  representing premium of Insurer  $j$  for year  $t$ . We denote by  $\mathbf{x}_{-j,t} = (x_{1,t}, \dots, x_{j-1,t}, x_{j+1,t}, \dots, x_{J,t})$  the vector  $\mathbf{x}_t$  without the  $j$ th component. This repeated game differs from other dynamic games proposed in the literature in the fact that policyholders are random agents, insurers set price in a repeated non-cooperative way, bankruptcy can occur either because of large losses or because of losing customers.

### 5.4.1 Extensions and modifications of the one-period game

Slight extensions have been considered by Mouminoux et al. (2021b). Firstly, most parameters are time-dependent. We define the estimated mean of overall costs including handling costs and claims of Insurer  $j$  as

$$\pi_{j,t} = \omega_j \bar{a}_{j,t-1} + (1 - \omega_j) \bar{m}_{t-1}, \quad (5.12)$$

where  $\omega_j \in [0, 1]$  is the credibility factor of Insurer  $j$  and the average market premium is determined as the mean of last  $d$  market premiums.

$$\bar{m}_{t-1} = \frac{1}{d} \sum_{u=1}^d m_{t-u}, \quad \text{with } m_{t-u} = \frac{\sum_{j \in J_t} \text{GWP}_{j,t-u} \times x_{j,t-u}^*}{\sum_{j \in J_t} \text{GWP}_{j,t-u}}.$$

Insurer  $j$  computes its actuarially based premium as the empirical mean of individual loss averages

$$\bar{a}_{j,t-1} = \frac{1}{1 - e_{j,t}} \frac{1}{d} \sum_{u=1}^d \frac{S_{j,t-u}}{n_{j,t-u}},$$

where  $S_{j,t}$  denotes the observed aggregate loss of Insurer  $j$  during year  $t$  and  $e_{j,t}$  denotes the expense rate as a percentage of gross written premium.

Secondly, an alternative market proxy value is considered, where firms do not attribute the same weight to each competitor, and we assume here that firms put more importance on the biggest competitors' price. We analyze the following market proxy (a weighted mean of other competitors' prices)

$$m_{j,t}(\mathbf{x}, \mathbf{n}) = \frac{1}{N - n_j} \sum_{k \in J_t \setminus \{j\}} n_k x_k. \quad (5.13)$$

Thirdly, objective and constraint functions are also time-dependent. The market proxy used in Equation (5.8) is the mean of other competitors' premium

$$m_{j,t}(\mathbf{x}) = \frac{1}{J_t - 1} \sum_{k \in J_t \setminus \{j\}} x_k. \quad (5.14)$$

The market proxy aims to assess other insurers' premiums. Assuming a positive price elasticity, the objective function, defined as the product of a demand function and an expected profit per policy representing a company-wide expected profit, is generalized to

$$O_{j,t}(\mathbf{x}_t) = \frac{n_{j,t}}{N} \left( 1 - \beta_j \left( \frac{x_{j,t}}{m_{j,t}(\mathbf{x}_t, \mathbf{n}_t)} - 1 \right) \right) (x_{j,t} - \pi_{j,t}), \quad (5.15)$$

where  $\pi_{j,t}$  is the break-even premium  $j$  in (5.12) and  $m_{j,t}(\mathbf{x}_t, \mathbf{n}_t)$  is the market proxy. The (tractable) solvency constraint function is generalized to

$$g_{j,t}^1(x_{j,t}) = \frac{K_{j,t-1} + n_{j,t-1}(x_{j,t} - \pi_{j,t})}{k_{995} \sigma(Y) \sqrt{n_{j,t-1}}} - 1, \quad (5.16)$$

Note that (5.16) and (5.11) are not strictly identical since the expense term disappears in (5.11).

Mouminoux et al. (2021b), we define the repeated game as the iteration of one-period games over  $T$  years, where  $K_{j,t}$  stands for the capital of Insurer  $j$  at time  $t$ ,  $J_t$  for the set of active insurers and  $S_{j,t}$  for the aggregate claim amount of Insurer  $j$ .

**Game 2** At time  $t = 0$ , initiate  $J_0 = \{1, \dots, J\}$ , positive capital levels  $K_{j,0}$  and positive portfolio sizes  $\mathbf{n}_0$ . For period  $t = 1, \dots, T$ , repeat

1. The insurers among  $J_t$  maximize their objective function subject to the solvency constraint:

$$\sup_{\mathbf{x}_{j,t} \in [\underline{\mathbf{x}}, \bar{\mathbf{x}}]} O_{j,t}(\mathbf{x}_{j,t}, \mathbf{x}_{-j,t}) \text{ s.t. } g_{j,t}(\mathbf{x}_{j,t}) \geq 0.$$

2. Once the premium equilibrium vector  $\mathbf{x}_t^*$  is determined, customers randomly lapse or renew, so we get a  $N_{j,t}(\mathbf{x}_t^*)$ .
3. Aggregate claim amounts  $S_{j,t}$  are randomly drawn according to the chosen loss model and the portfolio size  $N_{j,t}(\mathbf{x}_t^*)$ .
4. The underwriting result for Insurer  $j$  is computed by  $UW_{j,t} = N_{j,t}(\mathbf{x}_t^*) \times x_{j,t}^* \times (1 - e_{j,t}) - S_{j,t}$ , where  $e_{j,t}$  corresponds to the rate of handling costs of Insurer  $j$  at time  $t$ .
5. The capital is updated via  $K_{j,t+1} = K_{j,t} + UW_{j,t}$ .
6. The set of competitors  $J_{t+1}$  is updated by removing bankrupted insurers, tiny insurers and insolvent insurers.

Since the regulator asks insurers to be solvent, insurers are removed from the market when they have a negative capital level  $K_{j,t} < 0$  (bankrupted insurers). We also remove small insurers with a market share below 0.1%,  $n_{j,t}^* < 0.1\%N$ , assuming that in such a case an insurer will not be able to face future losses and handling costs. Indeed, such an insurer will not benefit from mutualisation concepts, essential for the insurance market and thus will decide to run-off the business. Finally, we remove players from the game when the capital is below the minimum capital requirement (MCR), whereas we keep them if capital is between MCR and solvency capital requirement (SCR). As a reminder, MCR can be defined as a percentage of the SCR computed in the solvency constraint. In general, in the non-life insurance retail market MCR is between 25% and 45% of the SCR set via the constant  $k$  in the constraint function (5.10).

### 5.4.2 Theoretical properties

The following assumptions on customers are made

- A1: Customer behavior is identical across the market and over time, i.e.  $(C_{i,t})_t$  are identically distributed for period  $t$  per insurer but depends on  $\mathbf{x}$ .
- A2: Customers are independent, i.e.  $(C_{i,t})_i$  are independent.
- A3: The customer choice of insurer at time  $t$  depends only on the previous choice at time  $t - 1$  and  $C_{i,t} \sim \mathcal{M}_J(1, \mathbf{p}_{j \rightarrow}(\mathbf{x}_t))$ .
- A4: No customer can enter or exit the market, the total market size  $N$  is constant.

The following assumptions on  $Y_{i,t}$  the aggregate loss of policy  $i$  during the period  $t$  are made

- A5: There is no adverse selection, i.e.  $Y_{i,t}$  are independent and identically distributed (i.i.d.) random variables,  $\forall i = 1, \dots, N$ .
- A6: Catastrophic events are excluded and  $Y_i$  follows a frequency – average severity loss model

$$Y_{i,t} = \sum_{l=1}^{M_{i,t}} Z_{i,l,t}, \quad (5.17)$$

where the claim number  $M_{i,t}$  is independent of the claim severity  $Z_{i,l,t}$ .

- A7: The insurance business is short-tailed, i.e. the loss  $Y_i$  is paid in total on December 31 of each year.

Mouminoux et al. (2021b) give general properties of the lapse and loss models used in Game 2 given a price vector value.

**Proposition 5.4.1 (Mouminoux et al. (2021b))** *The choice  $(C_{i,t})_t$  of Customer  $i$  at time  $t$  is a time-inhomo-geneous Markov chain with transition matrix  $P_{\rightarrow}^{(t)} = P_{\rightarrow}(\mathbf{x}_1) \times \cdots \times P_{\rightarrow}(\mathbf{x}_t)$ . The Markov chain  $(C_{i,t})_t$  has an invariant measure.*

*The choice vector  $(C_{1,t}, \dots, C_{N,t})_t$  of all customers at time  $t$  is a time-inhomogeneous Markov chain with transition matrix  $P_{\rightarrow}(\mathbf{x}_t) \otimes^n$  (the  $n$ -times Kronecker product of the matrix  $P_{\rightarrow}(\mathbf{x}_t)$ ).*

**Proposition 5.4.2 (Mouminoux et al. (2021b))** *The insurer portfolio size vector  $(\mathbf{N}_t)_t$  is a time-inhomo-geneous Markov chain with state space  $\mathcal{S}_{ms}$ . The probability generating function of  $\mathbf{N}_t \mid \mathbf{N}_{t-1} = \mathbf{n}$  is given by*

$$G_{\mathbf{N}_t \mid \mathbf{N}_{t-1} = \mathbf{n}}^P(\mathbf{z}) = (\mathbf{z}^T \mathbf{p}_{1 \rightarrow}(\mathbf{x}_t))^{n_1} \times \cdots \times (\mathbf{z}^T \mathbf{p}_{J \rightarrow}(\mathbf{x}_t))^{n_J},$$

where  $\mathbf{z} \in \mathbb{R}^J$  and  $^T$  denotes the matrix transpose.

*Let  $\mu$  be the invariant measure of  $(C_{i,t})_t$ .  $\mathbf{N}_t \mid \mathbf{N}_0 = \mathbf{n}$  tends to a multinomial distribution  $\mathcal{M}_J(N, \mu)$  and the invariant measure of  $(\mathbf{N}_t)_t$  is the vector with all probabilities of that multinomial distribution  $\mathcal{M}_J(N, \mu)$ .*

**Proposition 5.4.3 (Mouminoux et al. (2021b))** *Under Assumptions A5-A7, the moment generating function of the aggregate claim amount per insurer  $S_{j,t}$  at period  $t$  is given by*

$$G_{S_{j,t}}^M(z) = G_{N_{j,t}}^P \left( G_{\widetilde{M}_{j,t}}^P (G_Z^M(z)) \right),$$

where  $G^P$  stands for the probability generating function and  $G^M$  for the moment generating function. In particular, the insurer's aggregate claim amount  $S_{j,t}(\mathbf{x}_t)$  given that  $N_{j,t} = n_{j,t}$  is a compound distribution of the same kind as the individual loss amount  $Y_{i,t}$ .

Mouminoux et al. (2021b) focus then on properties of the repeated Game 2 in the case of a strongly regulated market. In this particular setting, a tractable expression of the invariant measure is available.

**Theorem 5.4.1 (Mouminoux et al. (2021b))** *The choice  $(C_{i,t})_t$  of Customer  $i$  at time  $t$  is a time-homogeneous Markov chain when  $\mathbf{x}_t = \mathbf{x}$ . In particular,  $P_{\rightarrow}^{(t)} = (P_{\rightarrow}(\mathbf{x}))^t$ . There exists a unique invariant measure  $\mu$  for  $(C_{i,t})_t$  given by*

$$\mu = \left( \frac{c_1^\Pi}{c_1^\Pi + \cdots + c_J^\Pi}, \dots, \frac{c_J^\Pi}{c_1^\Pi + \cdots + c_J^\Pi} \right) \text{ with } c_i^\Pi = \prod_{j=1, j \neq i}^J p_{j \neq i}. \quad (5.18)$$

*If in addition the choice probabilities  $p_{j \rightarrow k}$  are identical for all insurers, then  $\mu = (1/J, \dots, 1/J)$ .*

*The portfolio size vector  $(\mathbf{N}_t)_t$  at time  $t$  is a time-homogeneous Markov chain with state space  $\mathcal{S}_{ms}$  for which the invariant measure is the vector with all probabilities of that multinomial distribution  $\mathcal{M}_J(N, \mu)$ .*

Also, a tractable expression of the survival function of insurer loss  $S_{j,t}$  is given in the following proposition.

**Proposition 5.4.4 (Mouminoux et al. (2021b))** *Consider the invariant measure  $\mu$  given in (5.18). The survival function of the aggregate claim amount is given by*

$$P(S_{j,t} > s) = \sum_{0 \leq m \leq N} \binom{N}{m} (\mu_j)^m (1 - \mu_j)^{N-m} \cdot \sum_{0 \leq k} P\left(\widetilde{M}_{j,t} = k | N_{j,t} = m\right) P\left(\sum_{l=1}^k Z_l > s\right),$$

where the distribution of the total claim number  $\widetilde{M}_{j,t}$ .

Mouminoux et al. (2021b) consider the case of a deviation from a regulated price or a market-accepted level by one insurer, yet the other competitors remain at the same level. That is, we study  $\mathbf{x}_t = (x, \rho x, \dots, \rho x)$  with  $\rho > 0$  a fixed parameter.

**Theorem 5.4.2 (Mouminoux et al. (2021b))** *The choice  $(C_{i,t})_t$  of Customer  $i$  at time  $t$  is a time-homogeneous Markov chain. There exists a unique invariant measure  $\mu$  for  $(C_{i,t})_t$  given by*

$$\mu_1 = \frac{d_{-1}^{\Pi} - \sum_{j=2}^J d_{-1,-j}^{\Pi} p_{j \neq}}{d_{-1}^{\Pi} + \sum_{j=2}^J d_{-1,-j}^{\Pi} (p_{1 \neq} - p_{j \neq})}, \quad \mu_j = \frac{d_{-1,-j}^{\Pi} p_{1 \neq}}{d_{-1}^{\Pi} + \sum_{j=2}^J d_{-1,-j}^{\Pi} (p_{1 \neq} - p_{j \neq})}, \quad j = 2, \dots, J. \quad (5.19)$$

with  $d_l = (J-1)p_{l \neq} + p_{l \rightarrow 1}$  and

$$d_{-1,-j}^{\Pi} = \prod_{l=2, l \neq j}^J d_l, \quad d_{-1}^{\Pi} = \prod_{l=2}^J d_l.$$

The portfolio sizes  $(\mathbf{N}_t)_t$  at time  $t$  is a time-homogeneous Markov chain with state space  $\mathcal{S}_{ms}$  for which the invariant measure is the vector with all probabilities of that multinomial distribution  $\mathcal{M}_J(N, \mu)$ .

Let us analyze the case  $\rho > 1$ , i.e. Insurer 1 is the cheapest insurer. We study the stochastic ordering of the empirical average loss of insurers, e.g., (Shaked and Shanthikumar, 2007).

**Proposition 5.4.5 (Mouminoux et al. (2021b))** *If Insurer 1 is the cheapest insurer with  $\rho > 1$ , then the loss average by policy of Insurer 1, at any time  $t$ , is stochastically smaller than the one of the others in the following sense:*

$$\frac{1}{N_{1,t}(\tilde{\mathbf{x}})} \sum_{i=1}^{N_{1,t}(\tilde{\mathbf{x}})} Y_i \leq_{cx} \frac{1}{N_{k,t}(\tilde{\mathbf{x}})} \sum_{i=1}^{N_{k,t}(\tilde{\mathbf{x}})} Y_i, \quad \forall k \neq 1,$$

where  $\tilde{\mathbf{x}} = (x, \rho x, \dots, \rho x)$ .



**Proposition 5.4.6 (Mouminoux et al. (2021b))** *If in addition to Insurer 1 being the cheapest, for all  $k \neq 1$ ,  $x_1(1 - e_1) \leq x_k(1 - e_k)$ , then the underwriting result by policy is ordered  $UW_{1,t} \leq_{icx} UW_{k,t}$ , where  $UW_{j,t}$  is the random variable*

$$UW_{j,t} = x_j(1 - e_j) - \frac{1}{N_{j,t}(\mathbf{x})} \sum_{i=1}^{N_{j,t}(\mathbf{x})} Y_i,$$

where  $\tilde{\mathbf{x}} = (x, \rho x, \dots, \rho x)$ .

Mouminoux et al. (2021b) investigate some long-run properties of the repeated Game 2 after ensuring the existence and uniqueness of the premium equilibrium. The fact that the Nash equilibrium problem reduces to a linear system in some situations was already observed in Dutang et al. (2013a, Proposition 2.3).

**Proposition 5.4.7 (Mouminoux et al. (2021b))** *If there are at least two non-bankrupted insurers at time  $t$ , the repeated game with objective function (5.9) and solvency constraint (5.11) admits a unique (Nash) premium equilibrium. If in addition no constraint function is active, the premium equilibrium solves a linear system of equations.*

Due to the long-term behavior of portfolio size and loss amounts, Mouminoux et al. (2021b) derives the leadership probability as well as the end of the game.

**Proposition 5.4.8 (Mouminoux et al. (2021b))** *For the repeated insurance game, the probability that there are at least two non-bankrupt insurers at time  $t$  decreases geometrically as  $t$  increases.*

### 5.4.3 Numerical outputs

In a Monte-Carlo analysis, Dutang (2014a) analyze some random paths of the repeated Game 2 using the following packages: **actuar** for simulating loss models, **GNE** for computing Nash equilibrium repeatedly and **NLIG** for the repeated game. This game-theoretic approach allows to account for the effect of competition on insurer solvency. The proposed rational game shows that the most significant part of solvency relies on the ability of insurers to sell contracts (i.e. premium risk). This is opposite to classic risk theory where the collection of premiums is fixed per unit of time and the main risk is the randomness of losses.

Secondly, Game 2 also sheds new light on the presence of cycles in non-life insurance markets. Since for a range of parameters the market premium appears to be cyclical, we add a new argument in favor of a rational explanation (i.e. competition and loss uncertainty) for the presence of insurance cycles. This was also observed in Boonen et al. (2018) with open-loop Nash equilibrium strategies as well as in Lazar and Denuit (2012) with an econometric approach.

Mouminoux et al. (2021a) further study economics aspects of insurer solvency in Game 2 facing both the insurance loss uncertainty and the effect of competition. They apply the repeated Game 2 on the French motor market by considering the TOP-5 insurers in order to compute ruin and leadership probabilities.

# Bibliography

- Albrecher, H., Constantinescu, C., and Loisel, S. (2011). Explicit ruin formulas for models with dependence among risks. In: *Insurance: Mathematics and Economics* 48.2, pp. 265–270 (cit. on pp. 50–52).
- Albrecher, H. and Asmussen, S. (2006). Ruin probabilities and aggregate claims distributions for shot noise Cox processes. In: *Scandinavian Actuarial Journal* 2006.2, pp. 86–110 (cit. on p. 50).
- Albrecher, H. and Boxma, O. (2004). A ruin model with dependence between claim sizes and claim intervals. In: *Insurance: Mathematics and Economics* 35.2, pp. 245–254 (cit. on p. 50).
- Albrecher, H. and Daily-Amir, D. (2017). On Effects of Asymmetric Information on Non-Life Insurance Prices under Competition. In: *International Journal of Data Analysis Techniques and Strategies* 9.4, pp. 287–299 (cit. on p. 75).
- Albrecher, H. and Teugels, J. L. (2006). Exponential behavior in the presence of dependence in risk theory. In: *Journal of Applied Probability* 43.1, pp. 265–285 (cit. on p. 50).
- Albrecher, H., Beirlant, J., and Teugels, J. L. (2017). Reinsurance: actuarial and statistical aspects. John Wiley & Sons (cit. on pp. 13, 41).
- Alegre, A. and Claramunt, M. M. (1995). Allocation of solvency cost in group annuities: Actuarial principles and cooperative game theory. In: *Insurance: Mathematics and Economics* 17.1, pp. 19–34 (cit. on p. 70).
- Andersen, S. (1957). On the collective theory of risk in case of contagion between claims. In: *Bulletin of the Institute of Mathematics and its Applications* 12, pp. 2775–279 (cit. on p. 49).
- Anthropelos, M. and Boonen, T. J. (2020). Nash equilibria in optimal reinsurance bargaining. In: *Insurance: Mathematics and Economics* 93, pp. 196–205 (cit. on p. 69).
- Antonio, K. and Plat, R. (2014). Micro-level stochastic loss reserving for general insurance. In: *Scandinavian Actuarial Journal* 2014.7, pp. 649–669 (cit. on p. 54).
- Arendarczyk, M., Kozubowski, T. J., and Panorska, A. K. (2018). The joint distribution of the sum and maximum of dependent Pareto risks. In: *Journal of Multivariate Analysis* 167, pp. 136–156 (cit. on p. 41).
- Arjas, E. (1989). The claims reserving problem in non-life insurance: some structural ideas. In: *ASTIN Bulletin* 19.2, pp. 139–152 (cit. on p. 53).
- Arnold, B. C. (2015). Pareto Distributions. 2nd ed. CRC Press (cit. on pp. 13, 34).
- Arrow, K. and Debreu, G. (1954). Existence of an equilibrium for a competitive economy. In: *Econometrica* 22, pp. 265–290 (cit. on p. 67).
- Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. In: *Mathematical Finance* 9.3, pp. 203–228 (cit. on p. 70).
- Asmussen, S. and Albrecher, H. (2010). Ruin Probabilities. 2nd. World Scientific New Jersey (cit. on pp. 14, 50, 53, 59, 72).
- Asmussen, S., Christensen, B. J., and Thøgersen, J. (2019a). Nash equilibrium premium strategies for push–pull competition in a frictional non-life insurance market. In: *Insurance: Mathematics and Economics* 87, pp. 92–100 (cit. on p. 69).

- Asmussen, S., Christensen, B. J., and Thøgersen, J. (2019b). Stackelberg equilibrium premium strategies for push-pull competition in a non-life insurance market with product differentiation. In: *Risks* 7.2, p. 49 (cit. on p. 69).
- Asmussen, S. and Rolski, T. (1991). Computational methods in risk theory: A matrix algorithmic approach. In: *Insurance: Mathematics and Economics* 10.4, pp. 259–274 (cit. on p. 49).
- Atkins, D. C. and Gallop, R. J. (2007). Re-thinking how family researchers model infrequent outcomes: A tutorial on count regression and zero-inflated models. In: *Journal of Family Psychology* 21.4, pp. 726–735 (cit. on p. 62).
- Auza, J. and Alvarez, M. S. (2020). DetLifeInsurance: Life Insurance Premium and Reserves Valuation. R package version 0.1.3. URL: <https://CRAN.R-project.org/package=DetLifeInsurance> (cit. on p. 64).
- Avram, F., Biard, R., Dutang, C., Loisel, S., and Rabehasaina, L. (2014). A survey of some recent results on Risk Theory. In: *ESAIM: Proc.* 44, pp. 322–337 (cit. on pp. 14, 50).
- Bacinello, A. R. (2005). Endogenous model of surrender conditions in equity-linked life insurance. In: *Insurance: Mathematics and Economics* 37.2, pp. 270–296 (cit. on p. 62).
- Bacinello, A. R., Millossovich, P., Olivieri, A., and Pitacco, E. (2011). Variable annuities: A unifying valuation approach. In: *Insurance: Mathematics and Economics* 49.3, pp. 285–297 (cit. on p. 64).
- Bader, B., Yan, J., and Zhang, X. (2018). Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. In: *The Annals of Applied Statistics* 12.1, pp. 310–329 (cit. on p. 42).
- Barbe, P., Fougères, A.-L., and Genest, C. (2006). On the tail behavior of sums of dependent risks. In: *ASTIN Bulletin: The Journal of the IAA* 36.2, pp. 361–373 (cit. on p. 41).
- Basar, T. and Olsder, G. (1999). Dynamic Noncooperative Game Theory. SIAM (cit. on p. 65).
- Basu, A., Harris, I. R., Hjort, N. L., and Jones, M. (1998). Robust and efficient estimation by minimising a density power divergence. In: *Biometrika* 85.3, pp. 549–559 (cit. on p. 46).
- Baton, B. and Lemaire, J. (1981a). The bargaining set of a reinsurance market. In: *ASTIN Bulletin* 12.2, pp. 101–114 (cit. on p. 68).
- Baton, B. and Lemaire, J. (1981b). The core of a reinsurance market. In: *ASTIN Bulletin* 12.1, pp. 57–71 (cit. on p. 68).
- Battulga, G., Altangerel, L., and Battur, G. (2018). An extension of one-period Nash equilibrium model in non-life insurance markets. In: *Applied Mathematics* 9, pp. 1339–1350 (cit. on p. 75).
- Baumgartner, B. and Gatto, R. (2010). A Bootstrap Test for the Probability of Ruin in the Compound Poisson Risk Process. In: *ASTIN Bulletin* 40.1, pp. 241–255 (cit. on p. 59).
- Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004). Statistics of extremes: Theory and applications. Wiley & Sons (cit. on pp. 13, 42, 43).
- Beirlant, J., Dierckx, G., and Guillou, A. (2011). Bias-reduced estimators for bivariate tail modelling. In: *Insurance: Mathematics and Economics* 49.1, pp. 18–26 (cit. on pp. 44, 45).
- Beirlant, J., Joossens, E., and Segers, J. (2009). Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. In: *Journal of Statistical Planning and Inference* 139.8, pp. 2800–2815 (cit. on p. 45).
- Bensalem, S., Santibáñez, N. H., and Kazi-Tani, N. (2020). Prevention efforts, insurance demand and price incentives under coherent risk measures. In: *Insurance: Mathematics and Economics* 93, pp. 369–386 (cit. on p. 70).
- Bernegger, S. (1997). The Swiss Re Exposure Curves and the MBBEFD Distribution Class. In: *ASTIN Bulletin* 27.1, pp. 99–111 (cit. on pp. 13, 37).
- Bertrand, J. (1883). Théorie mathématique de la richesse sociale. In: *Journal des Savants* (cit. on p. 64).
- Beyersmann, J., Allignol, A., and Schumacher, M. (2012). Competing risks and multistate models with R. Use R! Springer (cit. on p. 29).

- Bland, R., Carter, T., Coughlan, D., Kelsey, R., Anderson, D., Cooper, S., and Jones, S. (1997). Workshop - Customer Selection and Retention. In: *General Insurance Convention & ASTIN Colloquium* (cit. on p. 61).
- Boonen, T. J., De Waegenaere, A., and Norde, H. (2017a). Redistribution of longevity risk: The effect of heterogeneous mortality beliefs. In: *Insurance: Mathematics and Economics* 72, pp. 175–188 (cit. on p. 69).
- Boonen, T. J., Pantelous, A. A., and Wu, R. (2018). Non-cooperative dynamic games for general insurance markets. In: *Insurance: Mathematics and Economics* 78. Longevity risk and capital markets: The 2015–16 update, pp. 123–135 (cit. on pp. 70, 80).
- Boonen, T. J., Tsanakas, A., and Wüthrich, M. V. (2017b). Capital allocation for portfolios with non-linear risk aggregation. In: *Insurance: Mathematics and Economics* 72, pp. 95–106 (cit. on p. 70).
- Borch, K. (1960a). Reciprocal reinsurance treaties. In: *ASTIN Bulletin* 1.4, pp. 170–191 (cit. on pp. 15, 68).
- Borch, K. (1960b). Reciprocal reinsurance treaties seen as a two-person cooperative game. In: *Skandinavisk Aktuarietidskrift* 1960.1-2, pp. 29–58 (cit. on p. 68).
- Borch, K. (1960c). The safety loading of reinsurance premiums. In: *Skandinavisk Aktuarietidskrift* 1960.3-4, pp. 163–184 (cit. on p. 68).
- Borch, K. (1962a). Application of game theory to some problems in automobile insurance. In: *ASTIN Bulletin* 2.2, pp. 208–221 (cit. on pp. 68, 70, 74).
- Borch, K. (1962b). Equilibrium in a reinsurance market. In: *Econometrica* 30.3, pp. 424–444 (cit. on p. 68).
- Borch, K. (1975). Optimal insurance arrangements. In: *ASTIN Bulletin* 8.3, pp. 284–290 (cit. on p. 68).
- Boudreault, M., Cossette, H., Landriault, D., and Marceau, E. (2006). On a risk model with dependence between interclaim arrivals and claim sizes. In: *Scandinavian Actuarial Journal* 2006.5, pp. 265–285 (cit. on p. 50).
- Boumezoued, A. and Devineau, L. (2017). Individual claims reserving: a survey. HaL paper 01643929. URL: <https://hal.archives-ouvertes.fr/hal-01643929> (cit. on p. 54).
- Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (Jan. 1984). Classification and Regression Trees. New Ed. Boca Raton: Chapman and Hall/CRC (cit. on p. 12).
- Briegleb, D. and Lemaire, J. (1982). Calcul des primes et marchandage. In: *ASTIN Bulletin* 13.2, pp. 115–131 (cit. on p. 69).
- Brockett, P. L., Golden, L. L., Guillen, M., Nielsen, J. P., Parner, J., and Perez-Marin, A. M. (2008). Survival analysis of a household portfolio insurance policies: How much time do you have to stop total customer defection? In: *Journal of Risk and Insurance* 75.3, pp. 713–737 (cit. on pp. 61, 63).
- Brouste, A. and Dutang, C. (2021a). One-step quantile matching estimation. Working paper (cit. on p. 38).
- Brouste, A. and Dutang, C. (2021b). OneStep - Le Cam's one-step estimation procedure. In: *R journal* 13.1, pp. 366–377 (cit. on pp. 13, 37, 38).
- Brouste, A., Dutang, C., and Rohmer, T. (2020). Closed form Maximum Likelihood Estimator for Generalized Linear Models in the case of categorical explanatory variables: Application to insurance loss modelling. In: *Computational Statistics* 35.2 (cit. on pp. 11, 12, 20–24, 26).
- Brouste, A., Dutang, C., and Rohmer, T. (2021a). A closed-form alternative estimator for GLM with categorical explanatory variables. submitted (cit. on pp. 12, 23, 24).
- Brouste, A., Dutang, C., and Rohmer, T. (2021b). Tools to fit generalized linear models (GLM) with explicit expressions. R package version 1.0.0 (cit. on p. 31).
- Brouste, A., Mieniedou, D. N., and Dutang, C. (2021c). OneStep: One-step Estimation. R package version 0.9.1. URL: <https://CRAN.R-project.org/package=OneStep> (cit. on p. 38).

- Buch-Larsen, T., Nielsen, J. P., Guillén, M., and Bolancé, C. (2005). Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. In: *Statistics* 39.6, pp. 503–516 (cit. on p. 35).
- Bühlmann, H. (1980). An economic premium principle. In: *ASTIN Bulletin* 11.1, pp. 52–60 (cit. on p. 69).
- Bühlmann, H. (1984). The general economic premium principle. In: *ASTIN Bulletin* 14.1, pp. 13–21 (cit. on p. 69).
- Caeiro, F. and Gomes, M. I. (2016). Threshold Selection in Extreme Value Analysis. In: *Extreme Value Modeling and Risk Analysis: Methods and Applications*. Ed. by Dey, D. K. and Yan, J. Boca Raton, FL: CRC Press, pp. 69–86 (cit. on p. 42).
- Cai, J. and Li, H. (2005). Multivariate risk model of phase type. In: *Insurance: Mathematics and Economics* 36.2, pp. 137–152 (cit. on p. 50).
- Cai, J.-J., Einmahl, J. H., de Haan, L., and Zhou, C. (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme. In: *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pp. 417–442 (cit. on p. 41).
- Cano-Berlanga, S., Gimenez-Gomez, J. M., and Vilella, C. (March 2015). Enjoying cooperative games: The R package GameTheory. In: *Working Paper No. 06; CREIP; Spain* (cit. on p. 67).
- Carter, Z. T., Lumley, T., Bodey, T. W., and Russell, J. C. (2021). The clock is ticking: Temporally prioritizing eradications on islands. In: *Global Change Biology* 27.7, pp. 1443–1456 (cit. on p. 12).
- Casella, G. and Lehmann, E. (1998). *Theory of Point Estimation*. Springer-Verlag (cit. on pp. 12, 33, 35, 36).
- Cebrián, A. C., Denuit, M., and Lambert, P. (2003). Generalized pareto fit to the society of actuaries' large claims database. In: *North American Actuarial Journal* 7.3, pp. 18–36 (cit. on p. 13).
- Centeno, M. d. L. (2002a). Excess of loss reinsurance and Gerber's inequality in the Sparre Anderson model. In: *Insurance: Mathematics and Economics* 31.3, pp. 415–427 (cit. on pp. 50, 59).
- Centeno, M. d. L. (2002b). Measuring the effects of reinsurance by the adjustment coefficient in the Sparre Anderson model. In: *Insurance: Mathematics and Economics* 30.1, pp. 37–49 (cit. on p. 59).
- Charpentier, A. (2017). Third Actuarial Pricing Game. Tech. rep. ACTINFO chair of Institut Louis Bachelier and Institut des Actuaire. URL: <https://freakonometrics.hypotheses.org/50078> (cit. on p. 20).
- Charpentier, A. and Denuit, M. (2004a). *Mathématiques de l'assurance non vie. Tome I: Principes Fondamentaux de Théorie du Risque*. Economica (cit. on pp. 11, 12).
- Charpentier, A. and Denuit, M. (2004b). *Mathématiques de l'assurance non vie. Tome II: Tarification et Provisionnement*. Economica (cit. on pp. 11, 13).
- Charpentier, A., ed. (2014). *Computational Actuarial Science with R*. Chapman and Hall-CRC (cit. on p. 11).
- Charpentier, A. and Juri, A. (2006). Limiting dependence structures for tail events, with applications to credit derivatives. In: *Journal of Applied Probability* 43.2, pp. 563–586 (cit. on p. 41).
- Charpentier, A. and Oulidi, A. (2010). Beta kernel quantile estimators of heavy-tailed loss distributions. In: *Statistics and computing* 20.1, pp. 35–55 (cit. on p. 35).
- Charpentier, A. and Pigeon, M. (2016). Macro vs. micro methods in non-life claims reserving (an econometric perspective). In: *Risks* 4.2, p. 12 (cit. on p. 54).
- Chen, L. and Shen, Y. (2019). Stochastic Stackelberg differential reinsurance games under time-inconsistent mean–variance framework. In: *Insurance: Mathematics and Economics* 88, pp. 120–137 (cit. on p. 69).
- Chen, S., Liu, Y., and Weng, C. (2019). Dynamic risk-sharing game and reinsurance contract design. In: *Insurance: Mathematics and Economics* 86, pp. 216–231 (cit. on p. 69).
- Collamore, J. (1996). Hitting probabilities and large deviations. In: *The Annals of Probability* 24.4, pp. 2065–2078 (cit. on p. 50).

- Constantinescu, C., Hashorva, E., and Ji, L. (2011). Archimedean copulas in finite and infinite dimensions— with application to ruin problems. In: *Insurance: Mathematics and Economics* 49.3, pp. 487–495 (cit. on p. 50).
- Constantinescu, C. D., Kozubowski, T. J., and Qian, H. H. (2019). Probability of ruin in discrete insurance risk model with dependent Pareto claims. In: *Dependence Modeling* 7.1, pp. 215–233 (cit. on p. 50).
- Cosma, S. and Dutang, C. (2021). tsallisqexp: Tsallis distribution. R package version 0.9-3. URL: <https://CRAN.R-project.org/package=tsallisqexp> (cit. on p. 35).
- Cournot, A. (1838). *Recherches sur les Principes Mathématiques de la Théorie des Richesses*. Paris: Hachette (cit. on p. 64).
- Cox, S. H. and Lin, Y. (2006). Annuity lapse rate modeling: Tobit or not tobit? In: *Society of actuaries* (cit. on p. 62).
- Cramér, H. (1930). On the Mathematical Theory of Risk. Skandia Jubilee Volume, Stockholm (cit. on p. 49).
- Daily-Amir, D. (2019). Market Share Analysis and Non-cooperative Game Theory with Applications in Swiss Health Insurance. PhD thesis. Université de Lausanne, Faculté des hautes études commerciales (cit. on p. 75).
- Daily-Amir, D., Albrecher, H., Bladt, M., and Wagner, J. (2019). On Market Share Drivers in the Swiss Mandatory Health Insurance Sector. In: *Risks* 7.4, p. 114 (cit. on p. 75).
- de Janvry, A., Dequiedt, V., and Sadoulet, E. (2014). The demand for insurance against common shocks. In: *Journal of Development Economics* 106, pp. 227–238 (cit. on p. 70).
- Dekkers, A. and de Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation. In: *Annals of Statistics* 17, pp. 1795–1832 (cit. on p. 42).
- Delignette-Muller, M. and Dutang, C. (2015). fitdistrplus: An R Package for Fitting Distributions. In: *Journal of Statistical Software* 64.1, pp. 1–34 (cit. on pp. 12, 33).
- DeLong, L., Lindholm, M., and Wüthrich, M. V. (2021). Making Tweedie’s compound Poisson model more accessible. In: *European Actuarial Journal*, pp. 1–42 (cit. on p. 27).
- Denault, M. (2001). Coherent allocation of risk capital. In: *Institute For Operations Research* 4.1-34 (cit. on p. 70).
- Denuit, M., Hainaut, D., and Trufin, J. (2019a). Effective Statistical Learning Methods for Actuaries I: GLMs and extensions. Springer Actuarial Lecture Notes. Springer (cit. on pp. 11, 12, 19).
- Denuit, M., Hainaut, D., and Trufin, J. (2019b). Effective Statistical Learning Methods for Actuaries II: Tree-Based Methods and Extensions. Springer Actuarial Lecture Notes. Springer (cit. on p. 63).
- Denuit, M., Hainaut, D., and Trufin, J. (2019c). Effective Statistical Learning Methods for Actuaries III: Neural Networks and Extensions. Springer Actuarial Lecture Notes. Springer (cit. on p. 63).
- Denuit, M., Maréchal, X., Pitrebois, S., and Walhin, J. (2007). *Actuarial Modelling of Claim Counts*. John Wiley & Sons (cit. on pp. 20, 34, 51).
- Denuit, M. (2020a). Investing in your own and peers’ risks: The simple analytics of P2P insurance. In: *European Actuarial Journal* 10, pp. 335–359 (cit. on p. 12).
- Denuit, M. (2020b). Size-biased risk measures of compound sums. In: *North American Actuarial Journal* 24.4, pp. 512–532 (cit. on p. 12).
- Denuit, M., Charpentier, A., and Trufin, J. (2021). Autocalibration and Tweedie-dominance for insurance pricing with machine learning. In: *arXiv preprint arXiv:2103.03635* (cit. on p. 27).
- Denuit, M. and Dhaene, J. (2012). Convex order and comonotonic conditional mean risk sharing. In: *Insurance: Mathematics and Economics* 51.2, pp. 265–270 (cit. on p. 70).
- Denuit, M. and Lang, S. (2004). Non-life rate-making with Bayesian GAMs. In: *Insurance: Mathematics and Economics* 35.3, pp. 627–647 (cit. on pp. 12, 28).
- Denuit, M. and Robert, C. Y. (2021a). From risk sharing to pure premium for a large number of heterogeneous losses. In: *Insurance: Mathematics and Economics* 96, pp. 116–126 (cit. on p. 70).

- Denuit, M. and Robert, C. Y. (2021b). Stop-loss protection for a large P2P insurance pool. In: *Insurance: Mathematics and Economics* 100, pp. 210–233 (cit. on p. 70).
- Denuit, M. and Trufin, J. (2017). Beyond the Tweedie reserving model: the collective approach to loss development. In: *North American actuarial journal* 21.4, pp. 611–619 (cit. on p. 53).
- Denuit, M. and Trufin, J. (2021). Generalization error for Tweedie models: decomposition and error reduction with bagging. In: *European Actuarial Journal*, pp. 1–7 (cit. on p. 27).
- Dickson, D., Hardy, M., and Waters, H. (2013). *Actuarial Mathematics for Life Contingent Risks*. 2nd. Cambridge University Press (cit. on p. 12).
- Diewert, W. E., Avriel, M., and Zang, I. (1981). Nine kinds of quasiconcavity and concavity. In: *Journal of Economic Theory* 25.3 (cit. on p. 65).
- Dimitrova, D. S. and Kaishev, V. K. (2010). Optimal joint survival reinsurance: An efficient frontier approach. In: *Insurance: Mathematics and Economics* 47.1, pp. 27–35 (cit. on p. 50).
- Draisma, G., Drees, H., Ferreira, A., and de Haan, L. (2004). Bivariate tail estimation: dependence in asymptotic independence. In: *Bernoulli* 10.2, pp. 251–280 (cit. on p. 45).
- Driessen, T. S., Fragnelli, V., Katsev, I. V., and Khmelnitskaya, A. B. (2011). On 1-convexity and nucleolus of co-insurance games. In: *Insurance: Mathematics and Economics* 48.2, pp. 217–225 (cit. on p. 70).
- Dunn, P. K. and Smyth, G. K. (2020). dglm: Double Generalized Linear Models. R package version 1.8.4. URL: <https://CRAN.R-project.org/package=dglm> (cit. on p. 31).
- Dutang, C. (2011). Regression models of price elasticity in non-life insurance. MA thesis. ISFA, Université Claude Bernard Lyon 1 (cit. on p. 63).
- Dutang, C. (2012a). Etude des marchés d’assurance non-vie à l’aide d’équilibres de Nash et de modèles de risques avec dépendance. PhD thesis. ISFA, Université Claude Bernard Lyon 1 (cit. on p. 75).
- Dutang, C. (2012b). The customer, the insurer and the market. In: *Bulletin Français d’Actuariat* 12.24, pp. 35–85 (cit. on pp. 15, 62, 63).
- Dutang, C. (2013a). A survey of GNE computation methods: theory and algorithms. Working paper, IRMA (cit. on p. 67).
- Dutang, C. (2013b). Existence theorems for generalized Nash equilibrium problems. In: *Journal of Nonlinear Analysis and Optimization: Theory and Applications* 4.2 (cit. on pp. 15, 67).
- Dutang, C. (2014a). A game-theoretic approach to non-life insurance market cycles. In: *SCOR paper* 29 (cit. on pp. 15, 80).
- Dutang, C. (2014b). Standard Statistical Inference. In: ed. by Charpentier, A. Chapman and Hall/CRC. Chap. 2 (cit. on pp. 12, 34).
- Dutang, C. (2017a). Some explanations about the IWLS algorithm to fit generalized linear model. Working paper (cit. on p. 19).
- Dutang, C. (2017b). Theoretical L-moments and TL-moments using combinatorial identities and finite operators. In: *Communications in Statistics - Theory and Methods* 46.8, pp. 3801–3828 (cit. on pp. 13, 38, 39).
- Dutang, C. (2020a). GNE: computation of generalized Nash Equilibria. R package version 0.99-3. URL: <https://cran.r-project.org/package=GNE> (cit. on p. 67).
- Dutang, C. (2020b). RTDE: Robust Tail Dependence Estimation. R package version 0.2-1. URL: <https://CRAN.R-project.org/package=RTDE> (cit. on p. 48).
- Dutang, C., Albrecher, H., and Loisel, S. (2013a). Competition between non-life insurers under Solvency constraints: a game-theoretic approach. In: *European Journal of Operation Research* 231.3 (cit. on pp. 15, 70, 71, 73–75, 80).
- Dutang, C. and Brouste, A. (2016). Closed-form and numerical computations of actuarial indicators in ruin theory and claim reserving. In: *Bulletin Français d’Actuariat* 16.31, pp. 111–137 (cit. on pp. 14, 54–57, 59).

- Dutang, C., Goegebeur, Y., and Guillou, A. (2014). Robust and bias-corrected estimation of the coefficient of tail dependence. In: *Insurance: Mathematics and Economics* 57, pp. 46–57 (cit. on pp. 13, 45–47).
- Dutang, C., Goegebeur, Y., and Guillou, A. (2016). Robust and Bias-Corrected Estimation of the Probability of Extreme Failure Sets. In: *Sankhya A* 78.1, pp. 52–86 (cit. on pp. 13, 47, 48).
- Dutang, C., Goulet, V., and Langevin, N. (2021). Numerical implementation of the Feller-Pareto distribution and related distributions. In: *Journal of Statistical Software* (cit. on pp. 13, 34).
- Dutang, C., Goulet, V., and Pigeon, M. (2008). actuar: An R Package for Actuarial Science. In: *Journal of Statistical Software* 25.1, pp. 1–37 (cit. on pp. 12, 34, 59).
- Dutang, C. and Guibert, Q. (2021). An explicit split point procedure in model-based trees allowing for a quick fitting of GLM trees and GLM forests. In: *Statistics and Computing* 32.1 (cit. on pp. 12, 24, 26, 27).
- Dutang, C. and Jaunatre, K. (2020). CRAN Task View: Extreme Value Analysis. URL: <https://CRAN.R-project.org/view=Distributions> (cit. on pp. 14, 48).
- Dutang, C. and Kiener, P. (2021). CRAN Task View: Probability Distributions. URL: <https://CRAN.R-project.org/view=Distributions> (cit. on p. 35).
- Dutang, C., Lefèvre, C., and Loisel, S. (2013b). On an asymptotic rule  $A+B/u$  for ultimate ruin probabilities under dependence by mixing. In: *Insurance: Mathematics and Economics* 53.3 (cit. on pp. 14, 51, 52).
- Dutang, C. and Milhaud, X. (2018). Lapse tables for lapse risk management in insurance: a competing risk approach. In: *European Actuarial Journal* 8.1 (cit. on pp. 15, 64).
- Dutang, C. and Spedicato, G. (2018). Machine Learning methods to perform pricing optimization. A comparison with standard GLMs. In: *Variance* 12.1 (cit. on pp. 15, 63, 64).
- Dutang, C. and Spedicato, G. (2021a). mbbefd: Maxwell Boltzmann Bose Einstein Fermi Dirac Distribution and Destruction Rate Modelling. R package version 0.8.10. URL: <https://cran.r-project.org/package=mbbefd> (cit. on p. 37).
- Dutang, C. and Spedicato, G. (2021b). Statistical aspects of destruction rate models: one-inflated and MBBEFD distributions. Working paper (cit. on pp. 13, 36, 37).
- Dutang, C. (2018). gumbel: package for Gumbel copula. R package version 1.10-2. URL: <https://cran.r-project.org/package=gumbel> (cit. on p. 48).
- Edgeworth, F. (1881). *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. London: Kegan Paul (cit. on p. 64).
- Elamir, E. and Scheult, A. (2003). Trimmed L-moments. In: *Computational Statistics & Data Analysis*, pp. 299–314 (cit. on p. 38).
- Eling, M. and Kochanski, M. (2013). Research on lapse in life insurance: what has been done and what needs to be done? In: *The Journal of Risk Finance* (cit. on p. 62).
- Emms, P. (2012). Equilibrium pricing of general insurance policies. In: *North American Actuarial Journal* 16.3, pp. 323–349 (cit. on p. 69).
- Eryilmaz, S. and Gebizlioglu, O. L. (2017). Computing finite time non-ruin probability and some joint distributions in discrete time risk model with exchangeable claim occurrences. In: *Journal of Computational and Applied Mathematics* 313, pp. 235–242 (cit. on p. 50).
- Everitt, B. and Hand, D. (1981). *Finite Mixture Models*. Chapman and Hall (cit. on p. 36).
- Facchinei, F., Fischer, A., and Piccialli, V. (2007). On generalized Nash games and variational inequalities. In: *Operations Research Letters* 35.2, pp. 159–164 (cit. on p. 75).
- Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. In: *The Annals of Statistics*, pp. 342–368 (cit. on p. 18).
- Feller, W. (1968). *An Introduction to Probability Theory and its Applications*. Third. John Wiley & Sons (cit. on p. 34).



- Fine, J. and Gray, R. (1999). A Proportional Hazards Model for the Subdistribution of a Competing Risk. In: *Journal of the American Statistical Association* 94.446, pp. 496–509 (cit. on p. 31).
- Fisher, R. and Tippett, L. (1928). On the estimation of the frequency distributions of the largest or smallest member of a sample. In: *Proceedings of the Cambridge Philosophical Society* (cit. on pp. 13, 42).
- Fox, J. (2010). Logit and Probit models. Tech. rep. York SPIDA (cit. on p. 19).
- Frees, E. (2009). Regression Modeling with Actuarial and Financial Applications. International Series on Actuarial Science. Cambridge University Press (cit. on pp. 11, 12).
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. In: *Journal of Statistical Software* 33.1, pp. 1–22 (cit. on p. 31).
- Fromenteau, M. and Petauton, P. (2017). Théorie et Pratique de l'Assurance Vie. Cinquième. Dunod (cit. on p. 11).
- Fudenberg, D. and Tirole, J. (1991). Game Theory. The MIT Press (cit. on p. 65).
- Gatto, R. and Baumgartner, B. (2016). Saddlepoint approximations to the probability of ruin in finite time for the compound Poisson risk process perturbed by diffusion. In: *Methodology and Computing in Applied Probability* 18.1, pp. 217–235 (cit. on p. 59).
- Geier, B. et al. (2020). DistributionFitR: Fitting Multiple Distributions. R package version 0.1. URL: <https://github.com/YCartes/DistributionFitr/tree/master> (cit. on p. 34).
- Genest, C. and Nešlehová, J. (2007). A primer on Copulas for count data. In: *ASTIN Bulletin* 37.2, pp. 475–515 (cit. on p. 52).
- Gerber, H. U. (1988). Mathematical fun with compound binomial process. In: *ASTIN Bulletin* 18.2, pp. 161–168 (cit. on p. 52).
- Gerber, H. U. (1978). Pareto-optimal risk exchanges and related decision problems. In: *ASTIN Bulletin* 10.1, pp. 25–33 (cit. on p. 68).
- Gerber, H. U. and Shiu, E. S. (1998). On the time value of ruin. In: *North American Actuarial Journal* 2.1, pp. 48–78 (cit. on p. 50).
- Gerber, H. U. and Shiu, E. S. (2005). The time value of ruin in a Sparre Andersen model. In: *North American Actuarial Journal* 9.2, pp. 49–84 (cit. on p. 50).
- Gesmann, M., Murphy, D., Zhang, Y. W., Carrato, A., Wuthrich, M., Concina, F., and Dal Moro, E. (2021). ChainLadder: Statistical Methods and Models for Claims Reserving in General Insurance. R package version 0.2.12. URL: <https://CRAN.R-project.org/package=ChainLadder> (cit. on pp. 14, 59).
- Gilleland, E., Ribatet, M., and Stephenson, A. G. (2013). A software review for extreme value analysis. In: *Extremes* 16.1, pp. 103–119 (cit. on p. 14).
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. In: *Annals of mathematics*, pp. 423–453 (cit. on p. 13).
- Goegebeur, Y., Guillou, A., and Qin, J. (2021). Extreme value estimation of the conditional risk premium in reinsurance. In: *Insurance: Mathematics and Economics* 96, pp. 68–80 (cit. on p. 41).
- Gossner, O. and Tomala, T. (2020). Repeated games with complete information. In: *Complex Social and Behavioral Systems: Game Theory and Agent-Based Models*, pp. 139–156 (cit. on p. 68).
- Goulet, V., Auclair, S., Dutang, C., Langevin, N., Milhaud, X., Ouellet, T., Parent, A., Pigeon, M., and Pouliot, L.-P. (2021a). actuar: Actuarial Functions and Heavy Tailed Distributions. R package version 3.1-2. URL: <https://CRAN.R-project.org/package=actuar> (cit. on p. 34).
- Goulet, V., Dutang, C., Maechler, M., Firth, D., Shapira, M., and Stadelmann, M. (2021b). expm: Matrix Exponential, Log, 'etc'. R package version 0.999-6. URL: <https://CRAN.R-project.org/package=expm> (cit. on p. 59).
- Graham, R., Knuth, D., and Patashnik, O. (1994). Concrete Mathematics: A Foundation for Computer Science. Addison-Wesley (cit. on p. 39).

- Gray, B. (2014). *cmprsk*: Subdistribution Analysis of Competing Risks. R package version 2.2-7. URL: <https://CRAN.R-project.org/package=cmprsk> (cit. on p. 31).
- Guibert, Q. and Dutang, C. (2021). Exponentially family regression tree. R package version 0.1-5 (cit. on p. 31).
- Guillen, M., Parner, J., Densgoe, C., and Perez-Marin, A. M. (2003). Using Logistic Regression Models to Predict and Understand Why Customers Leave an Insurance Company. In: *Intelligent and Other Computational Techniques in Insurance*. Vol. 6. Innovative Intelligence. World Scientific Publishing. Chap. 13 (cit. on p. 61).
- Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., Ravishanker, N., and Sriram, S (2006). Modeling customer lifetime value. In: *Journal of service research* 9.2, pp. 139–155 (cit. on p. 14).
- Haastrup, S. and Arjas, E. (1996). Claims reserving in continuous time: a nonparametric Bayesian approach. In: *ASTIN Bulletin* 26.2, pp. 139–164 (cit. on p. 53).
- Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall (cit. on pp. 12, 27).
- Henckaerts, R., Antonio, K., Clijsters, M., and Verbelen, R. (2018). A data driven binning strategy for the construction of insurance tariff classes. In: *Scandinavian Actuarial Journal* 2018.8, pp. 681–705 (cit. on p. 28).
- Henckaerts, R., Côté, M.-P., Antonio, K., and Verbelen, R. (2021). Boosting insights in insurance tariff plans with tree-based machine learning methods. In: *North American Actuarial Journal* 25.2, pp. 255–285 (cit. on p. 27).
- Hill, B. (1975). A simple general approach to inference about the tail of a distribution. In: *Annals of Statistics* 3, pp. 1163–1174 (cit. on p. 42).
- Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J. (2020). *copula*: Multivariate Dependence with Copulas. R package version 1.0-1. URL: <https://CRAN.R-project.org/package=copula> (cit. on p. 48).
- Hogg, R. V. and Klugman, S. A. (1984). *Loss distributions*. John Wiley & Sons (cit. on pp. 13, 20, 34).
- Hosking, J. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. In: *Journal of the Royal Statistical Society: Series B* 52.1, pp. 105–124 (cit. on p. 38).
- Hosking, J. (2007). Some theory and practical uses of trimmed L-moments. In: *Journal of Statistical Planning and Inference* 137.9, pp. 3024–3039 (cit. on p. 39).
- Huang, R. J., Huang, Y.-C., and Tzeng, L. Y. (2013). Insurance bargaining under ambiguity. In: *Insurance: Mathematics and Economics* 53.3, pp. 812–820 (cit. on p. 70).
- IASB (2008). Non-life insurance contracts. In: *IASB Meeting, Insurance Working Group* (cit. on p. 71).
- Ignatov, Z. G. and Kaishev, V. K. (2012). Finite time non-ruin probability for Erlang claim inter-arrivals and continuous inter-dependent claim amounts. In: *Stochastics An International Journal of Probability and Stochastic Processes* 84.4, pp. 461–485 (cit. on p. 50).
- INED and Mack Planck Institute (2021). Human Cause-of-Death Database. French Institute for Demographic Studies (France) and Max Planck Institute for Demographic Research (Germany). URL: [www.causeofdeath.org](http://www.causeofdeath.org) (cit. on p. 29).
- Jeong, H., Gan, G., and Valdez, E. A. (2018). Association rules for understanding policyholder lapses. In: *Risks* 6.3, p. 69 (cit. on p. 61).
- Jewell, W. (1989). Predicting IBNYR events and delays. I. Continuous time. In: *ASTIN Bulletin* 19.1, pp. 25–55 (cit. on p. 53).
- Jiang, W., Hong, H., and Ren, J. (2021). Pareto-optimal reinsurance policies with maximal synergy. In: *Insurance: Mathematics and Economics* 96, pp. 185–198 (cit. on p. 69).

- Jiang, W., Ren, J., Yang, C., and Hong, H. (2019). On optimal reinsurance treaties in cooperative game under heterogeneous beliefs. In: *Insurance: Mathematics and Economics* 85, pp. 173–184 (cit. on p. 69).
- Jin, Z., Yin, G., and Wu, F. (2013). Optimal reinsurance strategies in regime-switching jump diffusion models: Stochastic differential game formulation and numerical methods. In: *Insurance: Mathematics and Economics* 53.3, pp. 733–746 (cit. on p. 69).
- Joe, H. (2010). Tail dependence in vine copulae. In: *Dependence modeling: Vine copula handbook*. World Scientific, pp. 165–187 (cit. on p. 41).
- Jones, D. S. (1997). Introduction to Asymptotics: a Treatment using Nonstandard Analysis. World Scientific (cit. on pp. 47, 51).
- Jong, P. D. and Heller, G. (2008). Generalized Linear Models for Insurance Data. Cambridge University Press (cit. on p. 11).
- Kagraoka, Y. (2005). Modeling Insurance Surrenders by the Negative Binomial Model. Working Paper 2005 (cit. on p. 62).
- Kainhofer, R. (2021). LifeInsuranceContracts: Framework for Traditional Life Insurance Contracts. R package version 0.0.3. URL: <https://CRAN.R-project.org/package=LifInsuranceContracts> (cit. on p. 64).
- Kakutani, S. (1941). A generalization of Brouwer’s fixed point theorem. In: *Duke Math. J.* 8.3, pp. 457–459 (cit. on p. 66).
- Karlsson, J. (1976). The expected value of IBNR-claims. In: *Scandinavian Actuarial Journal* 1976.2, pp. 108–110 (cit. on p. 53).
- Kelsey, R., Anderson, D., Beauchamp, R., Black, S., Bland, R., Klauke, P., and Senator, I. (1998). Workshop - Price/Demand Elasticity. In: *General Insurance Convention & ASTIN Colloquium* (cit. on p. 61).
- Kiener, P. (2021). RWsearch: Lazy Search in R Packages, Task Views, CRAN, the Web. All-in-One Download. R package version 4.9.3. URL: <https://CRAN.R-project.org/package=RWsearch> (cit. on p. 31).
- Kim, C. (2005). Modeling Surrender and Lapse Rates with Economic Variables. In: *North American Actuarial Journal* 9.4, pp. 56–70 (cit. on p. 62).
- Clueppelberg, C. and Mikosch, T. (1995). Explosive Poisson shot noise processes with applications to risk reserves. In: *Bernoulli* 1.1/2, pp. 125–147 (cit. on p. 54).
- Klugman, S. A., Panjer, H. H., and Willmot, G. (2012). Loss Models: From Data to Decisions. 4th ed. New York: Wiley (cit. on pp. 13, 34, 35, 59, 72).
- Kondo, K. (2021). aglm: Accurate Generalized Linear Model. R package version 0.4.0. URL: <https://CRAN.R-project.org/package=aglm> (cit. on p. 31).
- Kuen, S. T. (2005). Fair valuation of participating policies with surrender options and regime switching. In: *Insurance: Mathematics and Economics* 37.3, pp. 533–552 (cit. on p. 62).
- Kwok, K. Y., Chiu, M. C., and Wong, H. Y. (2016). Demand for longevity securities under relative performance concerns: Stochastic differential games with cointegration. In: *Insurance: Mathematics and Economics* 71, pp. 353–366 (cit. on p. 69).
- Lachenbruch, P. (2002). Analysis of data with excess zeros. In: *Statistical Methods in Medical Research* 11, pp. 297–302 (cit. on p. 35).
- Larsen, C. (2007). An individual claims reserving model. In: *ASTIN Bulletin* 37.1, pp. 113–132 (cit. on p. 54).
- Lazar, D. and Denuit, M. (2012). New evidence for underwriting cycles in US property-liability insurance. In: *The Journal of Risk Finance* (cit. on p. 80).
- Le Cam, L. (1956). On the Asymptotic Theory of Estimation and Testing Hypotheses. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contri-*

- Contributions to the Theory of Statistics*. Berkeley, Calif.: University of California Press, pp. 129–156 (cit. on pp. 13, 37).
- Le Faou, Y. (2019). Contributions à la modélisation des données de durée en présence de censure: application à l'étude des résiliations de contrats d'assurance santé. PhD thesis. Sorbonne université (cit. on p. 61).
- LeDell, E. et al. (2021). h2o: R Interface for the 'H2O' Scalable Machine Learning Platform. R package version 3.32.1.3. URL: <https://CRAN.R-project.org/package=h2o> (cit. on p. 64).
- Ledford, A. and Tawn, J. (1997). Modelling dependence within joint tail regions. In: *Journal of the Royal Statistical Society. Series B. (Methodological)* 59, pp. 457–499 (cit. on p. 43).
- Lee, G. Y. (2017). General insurance deductible ratemaking. In: *North American Actuarial Journal* 21.4, pp. 620–638 (cit. on p. 11).
- Lefèvre, C. and Loisel, S. (2009). Finite-time ruin probabilities for discrete, possibly dependent, claim severities. In: *Methodology and Computing in Applied Probability* 11.3, pp. 425–441 (cit. on p. 50).
- Lefèvre, C. and Picard, P. (2011). A new look at the homogeneous risk model. In: *Insurance: Mathematics and Economics* 49.3, pp. 512–519 (cit. on p. 50).
- Leiria, M., Matos, N., and Rebelo, E. (2020). Non-life insurance cancellation: a systematic quantitative literature review. In: *The Geneva Papers on Risk and Insurance-Issues and Practice*, pp. 1–21 (cit. on p. 61).
- Lemaire, J. (1975). Sur l'emploi des fonctions d'utilité en assurance. In: *Bulletin de l'Association Royale des Actuaires Belges* (cit. on p. 68).
- Lemaire, J. (1977). Echange de risques et théorie des jeux. In: *ASTIN Bulletin* 9.1-2, pp. 165–180 (cit. on p. 68).
- Lemaire, J. (1979). A non symmetrical value for games without transferable utilities: application to reinsurance. In: *ASTIN Bulletin* 10.2, pp. 195–214 (cit. on p. 68).
- Lemaire, J. (1983). An application of game theory: cost allocation. In: *ASTIN Bulletin* 14.1, pp. 61–81 (cit. on pp. 69, 70).
- Lemaire, J. (1991). Cooperative game theory and its insurance applications. In: *ASTIN Bulletin* 21.1, pp. 1–40 (cit. on pp. 15, 68).
- Lemaire, J. and Quairière, J.-P. (1986). Chains of reinsurance revisited. In: *ASTIN Bulletin* 16.2, pp. 77–88 (cit. on p. 69).
- Li, D., Li, B., and Shen, Y. (2021). A dynamic pricing game for general insurance market. In: *Journal of Computational and Applied Mathematics* 389, p. 113349 (cit. on p. 69).
- Li, D. and Young, V. R. (2021). Bowley solution of a mean–variance game in insurance. In: *Insurance: Mathematics and Economics* 98, pp. 35–43 (cit. on p. 70).
- Liu, Z. (2010). Interaction Relationship between Agricultural Insurance and Transfer of Land Contract Rights Based on Game Analysis of Farmer Selection. In: *Agriculture and Agricultural Science Procedia* 1. International Conference on Agricultural Risk and Food Security 2010, pp. 187–192 (cit. on p. 70).
- Loisel, S. and Milhaud, X. (2011). From deterministic to stochastic surrender risk models: Impact of correlation crises on economic capital. In: *European Journal of Operational Research* 214.2 (cit. on p. 62).
- Lu, Y. and Garrido, J. (2005). Doubly periodic non-homogeneous Poisson models for hurricane data. In: *Statistical Methodology* 2.1, pp. 17–35 (cit. on p. 50).
- Lucchetti, R., Mignanego, F., and Pieri, G. (1987). Existence theorems of equilibrium points in stackelberg. In: *Optimization* 18.6, pp. 857–866. eprint: <https://doi.org/10.1080/02331938708843300> (cit. on p. 66).
- Lundberg, F. (1903). 1. Approximerad framställning af sannolikhetsfunktionen: 2. Återförsäkring af kollektivrisker. Akademisk avhandling, Almqvist och Wiksell, Uppsala. (cit. on p. 49).

- Luo, S., Wang, M., and Zhu, W. (2021). Stochastic differential reinsurance games in diffusion approximation models. In: *Journal of Computational and Applied Mathematics* 386, p. 113252 (cit. on p. 69).
- Macdonald, A., Richard, S., and Currie, I. (2018). *Modelling Mortality with Actuarial Applications*. Cambridge University Press (cit. on p. 12).
- Mahdi, S., Verma, A., Dutang, C., Kiener, P., and Nash, J. (2021). A Review of R Neural Network Packages (with NNbenchmark): Accuracy and Ease of Use. in revision (cit. on p. 64).
- Martinussen, T. and Scheike, T. (2006). *Dynamic Regression Models for Survival Data*. Springer (cit. on pp. 29, 31).
- Matsui, M. (2014). Prediction in a non-homogeneous Poisson cluster model. In: *Insurance: Mathematics and Economics* 55, pp. 10–17 (cit. on p. 54).
- Matsui, M. (2015). Prediction in a Poisson cluster model with multiple cluster processes. In: *Scandinavian Actuarial Journal* 2015.1, pp. 1–31 (cit. on p. 54).
- Matsui, M. and Mikosch, T. (2010). Prediction in a Poisson cluster model. In: *Journal of Applied Probability* 47.2, pp. 350–366 (cit. on p. 54).
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. 2nd. Chapman and Hall (cit. on pp. 17, 19).
- McFadden, D. (1981). Econometric Models of Probabilistic Choice. In: *Structural Analysis of Discrete Data with Econometric Applications*. The MIT Press. Chap. 5 (cit. on p. 71).
- McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley Series in Probability and Statistics (cit. on p. 35).
- Meng, H., Li, S., and Jin, Z. (2015). A reinsurance game between two insurance companies with nonlinear risk processes. In: *Insurance: Mathematics and Economics* 62, pp. 91–97 (cit. on p. 69).
- Milhaud, X., Maume-Deschamps, V., and Loisel, S. (2011). Surrender triggers in Life Insurance: What main features affect the surrender behavior in a classical economic context? In: *Bulletin Français d'Actuariat* 22.11 (cit. on p. 62).
- Mills, E. (2013). Adjusting for covariates in zero-inflated gamma and zero-inflated log-normal models for semicontinuous data. PhD thesis. Graduate College of The University of Iowa (cit. on p. 36).
- Mohamed Nor, N. H., Kooi, M., Diepens, N. J., and Koelmans, A. A. (2021). Lifetime accumulation of microplastic in children and adults. In: *Environmental science & technology* 55.8, pp. 5084–5096 (cit. on p. 12).
- Mouminoux, C., Dutang, C., Loisel, S., and Albrecher, H. (2021a). A game-theoretic approach to non-life insurance market cycles. Working paper (cit. on p. 80).
- Mouminoux, C., Dutang, C., Loisel, S., and Albrecher, H. (2021b). On a Markovian game model for competitive insurance pricing. In: *Methodology and Computing in Applied Probability* (cit. on pp. 15, 75, 76, 78–80).
- Mourdoukoutas, F., Boonen, T. J., Koo, B., and Pantelous, A. A. (2021). Pricing in a competitive stochastic insurance market. In: *Insurance: Mathematics and Economics* 97, pp. 44–56 (cit. on p. 70).
- Nash, J. (1950a). Equilibrium points in N-person games. In: *Proc. Nat. Acad. Sci. U.S.A.* 36.1, pp. 48–49 (cit. on p. 64).
- Nash, J. (1950b). The Bargaining Problem. In: *Econometrica* 18.2, pp. 155–162 (cit. on p. 64).
- Nash, J. (1951). Non-Cooperative Games. In: *The Annals of Mathematics* 54.2, pp. 286–295 (cit. on p. 64).
- Nash, J. (1953). Two Person Cooperative Games. In: *Econometrica* 21.1, pp. 128–140 (cit. on p. 64).
- Naudts, J. (2009). The q-exponential family in statistical physics. In: *Central European Journal of Physics* 7.3, pp. 405–413 (cit. on p. 35).
- Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized Linear Models. In: *Journal of the Royal Statistical Society. Series A* 135.3, pp. 370–384 (cit. on p. 17).

- Neuts, M. (1981). *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. The John Hopkins University Press (cit. on p. 59).
- Nikaido, H. and Isoda, K. (1955). Note on non-cooperative convex games. In: *Pacific J. Math.* 5.1, pp. 807–815 (cit. on p. 65).
- Norberg, R. (1993). Prediction of outstanding liabilities in non-life insurance. In: *ASTIN Bulletin* 23.1, pp. 95–115 (cit. on pp. 53, 54, 59).
- Norberg, R. (1999). Prediction of outstanding liabilities II. Model variations and extensions. In: *ASTIN Bulletin* 29.1, pp. 5–25 (cit. on p. 54).
- Ohlsson, E. and Johansson, B. (2010). *Non-Life Insurance Pricing with Generalized Linear Models*. Springer (cit. on p. 11).
- Olver, F. W. J., Lozier, D. W., Boisvert, R. F., and Clark, C. W., eds. (2010). *NIST Handbook of Mathematical Functions*. Cambridge University Press (cit. on pp. 52, 57).
- Osborne, M. and Rubinstein, A. (2006). *A Course in Game Theory*. Massachusetts Institute of Technology (cit. on pp. 65, 68).
- Ospina, R. and Ferrari, S. L. (2010). Inflated beta distributions. In: *Statistical Papers* 51.1, pp. 111–126 (cit. on p. 35).
- Panel on Nonstandard Mixtures of Distributions (1989). Statistical Models and Analysis in Auditing: Panel on Nonstandard Mixtures of Distributions. In: *Statistical Science* 4.1, pp. 2–33 (cit. on p. 36).
- Panjer, H. (2002). Measurement of risk, solvency requirements and allocation of capital within financial conglomerates. In: *Cancun AFIR/ICA conference* (cit. on pp. 15, 70).
- Papush, D. E., Popelyukhin, A. S., and Zhang, J. G. (2021). Approximating the Aggregate Loss Distribution. In: *Variance Journal*, p. 26120 (cit. on p. 12).
- Partrat, C., Lecoecur, E., Nessi, J.-M., Nisipasu, E., and Reiz, O. (2008). Provisionnement technique en assurance non-vie : Perspectives actuarielles modernes. *Economica* (cit. on p. 57).
- Pfaff, B. and McNeil, A. (2018). *evir: Extreme Values in R*. R package version 1.7-4 (cit. on p. 48).
- Picard, P., Lefevre, C., and Coulibaly, I. (2003). Multirisks model and finite-time ruin probabilities. In: *Methodology and Computing in Applied Probability* 5.3, pp. 337–353 (cit. on p. 50).
- Pigeon, M., Antonio, K., and Denuit, M. (2013). Individual loss reserving with the multivariate skew normal framework. In: *ASTIN Bulletin* 43.3, pp. 399–428 (cit. on p. 54).
- Pigeon, M., Antonio, K., and Denuit, M. (2014). Individual loss reserving using paid–incurred data. In: *Insurance: Mathematics and Economics* 58, pp. 121–131 (cit. on p. 54).
- Poon, S.-H., Rockinger, M., and Tawn, J. (2004). Extreme value dependence in financial markets: Diagnostics, models, and financial implications. In: *The Review of Financial Studies* 17.2, pp. 581–610 (cit. on p. 41).
- Powers, M. R. and Shubik, M. (1998). On the tradeoff between the law of large numbers and oligopoly in insurance. In: *Insurance: Mathematics and Economics* 23.2, pp. 141–156 (cit. on p. 70).
- Powers, M. and Shubik, M. (2006). A “square- root rule” for reinsurance. In: *Cowles Foundation Discussion Paper No. 1521*. (cit. on p. 70).
- Pun, C. S., Siu, C. C., and Wong, H. Y. (2016). Non-zero-sum reinsurance games subject to ambiguous correlations. In: *Operations Research Letters* 44.5, pp. 578–586 (cit. on p. 69).
- Pun, C. S. and Wong, H. Y. (2016). Robust non-zero-sum stochastic differential reinsurance game. In: *Insurance: Mathematics and Economics* 68, pp. 169–177 (cit. on p. 69).
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <http://www.R-project.org> (cit. on pp. 11, 31).
- Ramasubramanian, S (2006). An insurance network: Nash equilibrium. In: *Insurance: Mathematics and Economics* 38.2, pp. 374–390 (cit. on p. 69).
- Raymond-Belzile, L., Dutang, C., Northrop, P., and Opitz, T. (2021). A modeler’s guide to extreme-value software. Working paper (cit. on pp. 14, 48).

- Rees, R., Gravelle, H., and Wambach, A. (1999). Regulation of insurance markets. In: *The Geneva Paper on Risk and Insurance Theory* 24.1, pp. 55–68 (cit. on p. 15).
- Reynkens, T. and Verbelen, R. (2020). **ReIns**: Functions from "Reinsurance: Actuarial and Statistical Aspects". R package version 1.0.10 (cit. on p. 48).
- Reynkens, T., Verbelen, R., Beirlant, J., and Antonio, K. (2017). Modelling censored losses using splicing: A global fit strategy with mixed Erlang and extreme value distributions. In: *Insurance: Mathematics and Economics* 77, pp. 65–77 (cit. on p. 48).
- Ribatet, M. and Dutang, C. (2019). **POT**: Generalized Pareto Distribution and Peaks Over Threshold. R package version 1.1-7. URL: <https://CRAN.R-project.org/package=POT> (cit. on p. 48).
- Ridner, R. and Walker, H. (1984). Mixture densities, maximum likelihood and the EM algorithm. In: *SIAM Review* 26.2, pp. 195–239 (cit. on p. 36).
- Rosen, J. (1965). Existence and uniqueness of equilibrium points for concave N-person games. In: *Econometrica* 33.3, pp. 520–534 (cit. on pp. 65, 66, 75).
- Rudnytskyi, I. (2018). **ruin**: Simulation of Various Risk Processes. R package version 0.1.1. URL: <https://CRAN.R-project.org/package=ruin> (cit. on p. 59).
- Rusch, T. and Zeileis, A. (July 2013). Gaining insight with recursive partitioning of generalized linear models. In: *Journal of Statistical Computation and Simulation* 83.7, pp. 1301–1315 (cit. on p. 24).
- Saavedra-Nieves, A. (2016). **GameTheoryAllocation**: Tools for Calculating Allocations in Game Theory. R package version 1.0. URL: <https://CRAN.R-project.org/package=GameTheoryAllocation> (cit. on p. 67).
- Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer (cit. on p. 79).
- Shao, J. (2003). *Mathematical Statistics*. Springer Science+Business Media (cit. on p. 12).
- Shapiro, A. F. and Jain, L. C. (2003). *Intelligent and Other Computational Techniques in Insurance*. World Scientific Publishing (cit. on p. 61).
- Sjolander, A. (2019). **cglm**: Fits Conditional Generalized Linear Models. R package version 1.1. URL: <https://CRAN.R-project.org/package=cglm> (cit. on p. 31).
- Spedicato, G., Dutang, C., Kainhofer, R., Owens, K., Schirmacher, E., Clemente, G., and Williams, I. (2021). **lifecontingencies**: Financial and Actuarial Mathematics for Life Contingencies. R package version 1.3.7. URL: <https://CRAN.R-project.org/package=lifecontingencies> (cit. on p. 64).
- Spedicato, G. A. (2013). The lifecontingencies Package: Performing Financial and Actuarial Mathematics Calculations in R. In: *Journal of Statistical Software* 55.10, pp. 1–36 (cit. on p. 64).
- Staudacher, J. and Anwander, J. (2019). Using the R package **CoopGame** for the analysis, solution and visualization of cooperative games with transferable utility. R Vignette. URL: <https://cran.r-project.org/package=CoopGame> (cit. on p. 67).
- Stephenson, A. G. (2002). **evd**: Extreme Value Distributions. In: *R News* 2.2 (cit. on p. 48).
- Stephenson, A. G. and Gilleland, E. (2006). Software for the analysis of extreme events: The current state and future directions. In: *Extremes* 8, pp. 87–109 (cit. on p. 14).
- Suijs, J., De Waegenaere, A., and Borm, P. (1998). Stochastic cooperative games in insurance. In: *Insurance: Mathematics and Economics* 22.3, pp. 209–228 (cit. on p. 70).
- Taksar, M. and Zeng, X. (2011). Optimal non-proportional reinsurance control and stochastic differential games. In: *Insurance: Mathematics and Economics* 48.1, pp. 64–71 (cit. on p. 69).
- Teugels, J. and Sundt, B. (2004). *Encyclopedia of Actuarial Science*. Vol. 1. John Wiley & Sons (cit. on p. 11).
- Therneau, T. and Grambsch, P. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag (cit. on p. 29).
- Therneau, T. M. (2021). **survival**: A Package for Survival Analysis in R. R package version 3.2-11. URL: <https://CRAN.R-project.org/package=survival> (cit. on p. 31).
- Titterton, D., Smith, A., and Makov, U. (1985). *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons (cit. on p. 36).

- Trufin, J., Albrecher, H., and Denuit, M. (2011). Ruin problems under IBNR dynamics. In: *Applied Stochastic Models in Business and Industry* 27.6, pp. 619–632 (cit. on p. 49).
- Trufin, J. and Loisel, S. (2013). Ultimate ruin probability in discrete time with Bühlmann credibility premium adjustments. In: *Bulletin Français d'Actuariat* (cit. on p. 51).
- Tsanakas, A. (2004). Dynamic capital allocation with distortion risk measures. In: *Insurance: Mathematics and Economics* 35.2. Papers presented at the 7th IME Conference, pp. 223–243 (cit. on p. 70).
- Tsanakas, A. and Barnett, C. (2003). Risk capital allocation and cooperative pricing of insurance liabilities. In: *Insurance: Mathematics and Economics* 33.2. Papers presented at the 6th IME Conference, Lisbon, 15–17 July 2002, pp. 239–254 (cit. on p. 70).
- Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. 4th. Springer (cit. on p. 27).
- Viaene, S., Veugelers, R., and Dedene, G. (2002). Insurance bargaining under risk aversion. In: *Economic Modelling* 19.2, pp. 245–259 (cit. on p. 70).
- Von Heusinger, A. and Kanzow, C. (2009). Optimization reformulations of the generalized Nash equilibrium problem using the Nikaido-Isoda type functions. In: *Computational Optimization and Applications* 43.3 (cit. on p. 75).
- Von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton Classic Editions. Princeton: Princeton University Press (cit. on p. 64).
- Wang, J., Reyes-Gibby, C. C., and Shete, S. (2021a). An Approach to Analyze Longitudinal Zero-Inflated Microbiome Count Data Using Two-Stage Mixed Effects Models. In: *Statistics in Biosciences* 13.2, pp. 267–290 (cit. on p. 12).
- Wang, N., Zhang, N., Jin, Z., and Qian, L. (2019). Robust non-zero-sum investment and reinsurance game with default risk. In: *Insurance: Mathematics and Economics* 84, pp. 115–132 (cit. on p. 69).
- Wang, N., Zhang, N., Jin, Z., and Qian, L. (2021b). Reinsurance–investment game between two mean–variance insurers under model uncertainty. In: *Journal of Computational and Applied Mathematics* 382, p. 113095 (cit. on p. 69).
- Wang, N., Zhang, N., Jin, Z., and Qian, L. (2021c). Stochastic differential investment and reinsurance games with nonlinear risk processes and VaR constraints. In: *Insurance: Mathematics and Economics* 96, pp. 168–184 (cit. on p. 69).
- Warren, R., Yao, J., Rourke, T., and Iwanik, J. (2012). Game Theory in General Insurance: How to outdo your adversaries while they are trying to outdo you. In: *GIRO Conf* (cit. on p. 69).
- Wood, S. N. (2008). Fast stable direct fitting and smoothness selection for Generalized Additive Models. In: *Journal of the Royal Statistical Society: Series B* 70.3 (cit. on pp. 27, 28).
- Wood, S. N. (2010). Fast stable REML and ML estimation of semiparametric GLMs. In: *Journal of the Royal Statistical Society: Series B* 73.1, pp. 3–36 (cit. on pp. 28, 31).
- Wu, R. and Pantelous, A. A. (2017). Potential games with aggregation in non-cooperative general insurance markets. In: *ASTIN Bulletin: The Journal of the IAA* 47.1, pp. 269–302 (cit. on p. 69).
- Wüthrich, M. and Merz, M. (2008). *Stochastic Claims Reserving Methods in Insurance*. Wiley Finance (cit. on pp. 11, 14, 53, 58, 59).
- Wüthrich, M. V. (2018a). Machine learning in individual claims reserving. In: *Scandinavian Actuarial Journal* 2018.6, pp. 465–480 (cit. on pp. 27, 54).
- Wüthrich, M. V. (2018b). Neural networks applied to chain–ladder reserving. In: *European Actuarial Journal* 8.2, pp. 407–436 (cit. on p. 54).
- Yang, Y. and Konstantinides, D. G. (2015). Asymptotics for ruin probabilities in a discrete-time risk model with dependent financial and insurance risks. In: *Scandinavian Actuarial Journal* 2015.8, pp. 641–659 (cit. on p. 50).



- Yao, D., Yang, H., and Wang, R. (2011). Optimal dividend and capital injection problem in the dual model with proportional and fixed transaction costs. In: *European Journal of Operational Research* 211.3, pp. 568–576 (cit. on p. 50).
- Yee, T. (2015). *Vector Generalized Linear and Additive Models With an Implementation in R*. Springer (cit. on p. 28).
- Yeo, A. C. and Smith, K. A. (2003). An integrated Data Mining Approach to Premium Pricing for the Automobile Insurance Industry. In: *Intelligent and Other Computational Techniques in Insurance*. Vol. 6. Innovative Intelligence. World Scientific Publishing. Chap. 5 (cit. on p. 61).
- Zariņa, I., Voronova, I., and Pettere, G. (2021). Alternative capital requirement for insurers: possibilities and issues. In: *International Journal of Economics and Business Research* 21.1, pp. 41–61 (cit. on p. 12).
- Zeileis, A. and Hornik, K. (2007). Generalized M-fluctuation tests for parameter instability. In: *Statistica Neerlandica* 61.4, pp. 488–508 (cit. on p. 24).
- Zeileis, A., Hothorn, T., and Hornik, K. (June 2008). Model-Based Recursive Partitioning. In: *Journal of Computational and Graphical Statistics* 17.2, pp. 492–514 (cit. on pp. 12, 24).
- Zeileis, A., Koenker, R., and Doebler, P. (2015). *glm: Generalized Linear Models Extended*. R package version 0.1-1. URL: <https://CRAN.R-project.org/package=glm> (cit. on p. 31).
- Zeng, X. (2010). A stochastic differential reinsurance game. In: *Journal of Applied Probability* 47.2, pp. 335–349 (cit. on p. 69).
- Zeng, X. and Luo, S. (2013). Stochastic Pareto-optimal reinsurance policies. In: *Insurance: Mathematics and Economics* 53.3, pp. 671–677 (cit. on p. 69).
- Zhang, N., Jin, Z., Qian, L., and Fan, K. (2019). Stochastic differential reinsurance games with capital injections. In: *Insurance: Mathematics and Economics* 88, pp. 7–18 (cit. on p. 69).
- Zhao, Y., Wang, R., Yao, D., and Chen, P. (2015). Optimal dividends and capital injections in the dual model with a random time horizon. In: *Journal of Optimization Theory and Applications* 167.1, pp. 272–295 (cit. on p. 50).
- Zoccolan, I. (2018). *valuer: Pricing of Variable Annuities*. R package version 1.1.2. URL: <https://CRAN.R-project.org/package=valuer> (cit. on p. 64).

# Appendix A

## Additional examples of distributions in the exponential family

### A.1 Chapter 2

Name	$\mathcal{X}$	$\theta$	$\Theta$	$\eta$	$T(x)$	$B(\theta)$	$h$
Normal	$\mathbb{R}$	$(\mu, \sigma^2)$	$\mathbb{R} \times \mathbb{R}_+^*$	$(\frac{\mu}{\sigma^2}, \frac{-1}{\sigma^2})$	$(x, x^2)$	$\frac{\mu^2}{2\sigma^2} + \frac{\log(2\pi\sigma^2)}{2}$	$\frac{\mathbb{1}_{\mathbb{R}}(x)}{x}$
Gamma	$\mathbb{R}_+^*$	$(\alpha, \lambda)$	$(\mathbb{R}_+^*)^2$	$(\frac{-1}{\lambda}, \alpha)$	$(x, \log(x))$	$\alpha \log \lambda + \log \Gamma(\alpha)$	$\frac{\mathbb{1}_{\mathbb{R}_+^*}(x)}{x}$
Beta	$(0, 1)$	$(a, b)$	$(\mathbb{R}_+^*)^2$	$(a, b)$	$(\log(x), \log(1-x))$	$\log \beta(a, b)$	$\frac{\mathbb{1}_{(0,1)}(x)}{x(1-x)}$

Table A.1 – Some examples for  $d = 1$

Nom	$\mathcal{X}$	$\theta$	$\Theta$	$\eta$	$T(x)$	$B(\theta)$	$h$
Normal $\mu, \Sigma$	$\mathbb{R}^d$	$(\mu, \Sigma)$	$\mathbb{R}^d, \mathbb{R}^{d \times d}$	$\Sigma^{-1}\mu, -\frac{1}{2}\Sigma^{-1}$	$x, xx^T$	$\frac{\mu^T \Sigma^{-1} \mu}{2} + \frac{\log \det(\Sigma)}{2}$	
Multinomial $\mathbf{n}; \mathbf{p}$	$x \in \mathbb{N}^d$ $\sum_i x_i = n$	$\begin{pmatrix} p_1 \\ \vdots \\ p_{d-1} \end{pmatrix}$	$\mathbf{p} \in [0, 1]^{d-1}$ $p_d = 1 - \sum_i p_i$	$\begin{pmatrix} \log(\frac{p_1}{p_d}) \\ \vdots \\ \log(\frac{p_{d-1}}{p_d}) \end{pmatrix}$	$\begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$	$n \log(p_d)$	$\frac{n! \mathbb{1}_{\mathcal{X}}(x)}{\prod_{j=1}^d x_j!}$

Table A.2 – Some examples for  $d \geq 2$