INSA de Toulouse
Département Génie Mathématique et Modélisation
5 ModIA
2024-2025

An introduction to the Mathematics of Actuarial Sciences

Anthony Réveillac[1]

[1] anthony.reveillac@insa-toulouse.fr, bureau GMM 111 - CSH B10

Contents

Chapter 1

Actuarial Sciences : an overview and what about Insurance in an Engineering curriculum ?

What is your relationship with luck ? In the very first years of our lives we get to know several forces of nature like the light, the wind, the gravity and another one rather more tricky to capture but definitely part of our life : randomness ! Indeed randomness is the generic word that gather our contingent experiences of uncertain outcomes; whatever you call it *luck, hasard*[1] *or risk.* Although all these terms refer to the notion of randomness; but are they precise synonyms ? Let's figure out.

1.1 Randomness, Risk : is it six of one and half of a dozen of the other ?

Not really ! It is important for us to make a clear difference between these two terms.

Definition 1.1.1 (Randomness)**.** *Randomness refers to a situation for which given (apparent) exact same conditions can lead to different outcomes.*

Example 1.1.2. *Tossing a coin falls into that description. Maybe a precise evaluation of the initial speed and precise move of the toss would lead to getting Tail or Head but in practice one does not have access to these initial conditions and so the getting of Head or Tail is considered as random.*

What about Risk then ? In English it is commonly confused with the notion of randomness or more precisely with a random situation where there are negative outcomes; and also positive otherwise the situation is simply seen as a hazard.

Definition 1.1.3 (Risk (mathematically))**.** *The risk within a probabilistic framework is the evaluation of the probabilities of each outcomes for a random situation.*

[1] Be cautious hasard is a false cognate as in English it translates to "danger" and not to "hasard" as in French

Remark 1.1.4. *Note that an immediate consequence is that risk stands for the quantification of the probability of a given outcome (or all of them). So it is by essence a mathematical notion that calls for a probabilistic framework.*

Example 1.1.5. *Indeed coming back to the tossing a coin it can easily be modeled using a probability framework in which one would evaluate the probability of getting Head (and thus Tail). If one looses 10€ in case of Tail and wins 15€ in case of Head then the "risk" in this situation will be to determine the probability of loosing (or winning) and maybe other quantities like the expectation of gains (or losses).*

Another example that helps to make precise our point is the Casino. Indeed, games in a Casino like the Jackpot for instance are contingent systems. But in order to be a sustainable business each of the machines (or games) at play are precisely calibrated so that the probability of a Jackpot is known and accurate (and can easily be checked by experience). So as a head of a Casino you have some risks in the sense that all the games have a probability of success and fails which are known to you.

So this is what it is about : given a random situation handling the risk of it. For a customer (we will say later one a contractor) the notion of risk is the one we use in our everyday life : *the fear to loose something or some money.*

1.2 Risk management in a nutshell : who one can manage risk ?

We established that Risk was mostly a concern and needed to be handled. Once again, in our everyday life understand of Risk we may mean Randomness that is uncertain outcomes. Many theories in Finance, Economics or Insurance are tagged with the label "Risk Management" but they all originate from natural principles that we list below.

1.2.1 Cautiousness

You want to avoid risk then just be cautious ! These are wise words, that we learn from our early times in life. Parents learn this very wise principle to us. Avoiding to swim with sharks covered with fish blood is indeed one way to avoid the risk of a hazardous swim but there are many risks in our everyday life that one cannot completely cancel. A very good example for our discussion that we will use throughout this section is the situation of driving a car. Once one accept to drive a car then he or she puts himself or herself in a situation at risk. It is not always easy to avoid this situation or to choose to avoid it. Let's note though that even if one accepts this situation different behaviors (being careful or at the opposite wreckless) have a huge impact on our exposure to negative events. In fact this impacts the risk, that is the evaluation of the probability of an accident and of the value of the damages. We will come back on this idea later on.

1.2.2 Mutualisation of Risk : *"The contributions of the many to the misfortunes of the few"*

We have seen that one way (not always realistic though) to prevent risks is to avoid risky situations. This is also called *self-insurance*; so what is a *non self-insurance* ?

If I do not insure myself somebody has to do it then ! Yes and...no. If your risk is not acceptable to you it will not be to somebody else. Let's take again the car insurance example. You are risk averse to a possible accident with your car. To simplify and not make the discussion depressing assume the possible damages are material. You are afraid of facing material damages and more specifically the expanses associated to them. So who would accept to pay for these possible expanses in the futures in exchange of the paiement of a premium that has to be considerably low compared the expanses (you will not pay 70€for being insured of paying a 100 €).

But ... *you are not alone* and this is indeed the famous quote by Edward Lloyd *"The contributions of the many to the misfortunes of the few"*. The idea of mutualisation lies in the fact that if many agents share a risk of common nature (but *uncorrelated* among agents) which is *quite unlikely* then facing this risk as a group rather than as an individual is an advantage. This is called the mutualisation of the risk.

Discussion on mutualisation

This topic deserves from the point of view of the author a focus.

Is it always working ?

This is indeed the first question that comes to mind as we introduced this risk management method as an alternative to the "Insure yourself" but is it really efficient ? There are two many words in its definition to make it efficient : *dispersion* and *independence*.

Definition 1.2.1 (Dispersion)**.** *Dispersion for a risk to be mutualized refers to the fact that the number of agents impacted by a damage is "small". The last word is not very quantitative but it is at the same time meaningful. Imagine a group of agents who are individually exposed to a given random loss which is quite likely, then how can the community be of help ? In fact we are then back to the situation of one individual facing a given risk ! A good example of mutualisation is car insurance. As an individual one is quite unlikely to have a car accident; this is what we will call dispersion.*

Another important point is the notion of independence which guarantees that there is no contagion effect.

Definition 1.2.2 (Independence)**.** *Independence for a risk to be mutualized refers to the fact that possible damages to different agents are independent. Once again for car accidents usually contractors have accidents independently of each others. This is of course a model feature that can be sometimes relaxed. The main point here is to exclude contagion effects that will lead to a different analysis.*

Is it fair ?

Once community is at play the question of fairness seems unavoidable. You'll be soon in a company, in some places people contribute to the common coffee bank. But sometimes you see questions rising like "Should everybody contribute equally ? Anthony is drinking far more coffee than the rest of the team he should pay more !" Coming back to our insurance problem, one can simply say that we share the same risk but do we all behave the same way ? We moved from individual self-insurance to mutualisation but the latter does not prevent us from trying to minimize our exposure to risk. And this is an effort, and as we learned as kids efforts call for a reward, or symmetrically those who do not make effort should be penalized. This is exactly why one pays its car insurance premium depending on his driver profile. This question is at the same time very important and delicate in Economics. Part of this discussion falls into the notion of *Principal - Agent* problems[2] This collects several interesting models in insurance like the *Moral Hazard* which deals with uncertainty of the actions of the agent or the so-called *adverse selection* model where the agents have different types that are unknown to you and that impacts the actions of the agent. Considering again the car insurance model, the insurer would like to be able to vary the premium depending on how cautious the driver is but usually these information are not public to him.

Some limitation to this method ?

We have already mentioned several points which are crucial to this method and mainly it is worth focusing on a very current and important issue that is brought by climate changes. Indeed, several climate risks are a threat to many contractors (soils changes, hurricanes, heavy heats, drought, diseases, higher mortality...).

1.2.3 Towards financial markets : risk transfer

At the frontier of Economics and Finance lies some interesting theories related to the so-called risk transfer. Let us try to consider an example that will help us to understand what it is about.

Example 1.2.3 (ENSO (El Niño Southern Oscillation))**.** *ENSO is the name given to the pseudo-random natural phenomenon (which occurs approximately every 3 to 8 years) also known as El Niño which consists of a sudden increase of the temperature at the surface of East Pacific Ocean. This phenomenon entails for instance an increase of rainfall nearby South America. This consequence is profitable to some farmers (those who cultivate cotton for instance) in dry regions. In the same time, the increase of surface ocean leads to less fish catches for fishermen in this area. Thus El Niño creates a situation where two types of economic agents: farmers Fa and fishermen Fi are exposed to an opposite risk.*

The goal of this exercise, is to propose a very simple contract which enables the farmers and the fishermen to reduce their exposure to El Niño.

1) *Let e and e' be the two events:*

$$e = \{\textit{ENSO occurs the year } X\}, \quad e' = \{\textit{ENSO does not occur the year } X\}$$

The wealth (in a given unit of numéraire) of the agents Fa and Fi are given in the table

[2] As students of N7 and INSA Toulouse it is worth to point that Toulouse and the Toulouse School of Economics is an historical research leader on this part of Economics whose ultimate recognition was the "Nobel Memorial Prize in Economic Sciences" (aka Nobel price in Economics) awarded to Jean Tirole in 2014.

below.

	Fa	Fi
e	3	2
e'	1	4

Plot in a graph (whose horizontal axis represents the wealth in case of e and whose vertical axis represents the wealth in case of e') the points Fa and Fi. Plot as well the bisectrix and give an interpretation of the points on it.
2) *The 1st of december of year X, Fa and Fi subscribe to the following contract :*
"On December 31st of the year X, Fa pays $\frac{1}{2}$ to Fi in case of e; otherwise Fi pays $\frac{1}{2}$ to A." We denote by Fa^c and Fi^c the new coordinates of the wealths of Fa and Fi with this contract. Add these points Fa^c and Fi^c on the graph and give an interpretation of the positions compared to the bisectrix.
3) *We now replace the table of 1) by the following one:*

	Fa	Fi
e	3	1
e'	1	3

Design a contract between Fa and Fi so that the risk of Fa and Fi vanishes.

Note that this simple exercise provides an example of a securitization contract where no cash-flows are exchanged at the signature of the contract but only at maturity. Note also that we make here a very elementary analysis of this problem but the real market based on this situation calls for a much deeper analysis which goes beyond the objectives of this course.

More generally it is a usual question to point out the difference between Finance and Actuarial Sciences. For instance many techniques in Risk Management are used in Finance. A very big picture can be given by distinguishing Finance and Actuarial sciences by the fact that in Finance agents consider products (called financial products such as bonds, options, futures, swaps...) that are complicated contracts based on underlying that are accessible on a financial market. In other words the paradigmatic situation in Finance is to consider risk which are *endogeneous* to the market. Insurers mostly (once again this is a big picture) for which the randomness is not accessible to the insurer (climate events, car accidents, diseases

Chapter 2

A probabilistic model of insurance contracts

2.1 Some definitions and modeling features

We gave a very short panorama of the notion of risk. Providing the first notions of insurance in Europe dates back to the XIV th century when merchants wanted to insure the commodities carried out on the ship through the Atlantic ocean. It is maybe interesting to note that the first contracts on these materials are also the first financial products which are known as Call Options or Put Options. Yet these contracts did not take their nowadays form; banks financed ships and cargos and covered the risk in case of a shipwreck, otherwise banks got profits out of the sell of the cargo by the merchants. This kind of insurance contract focus on a good. Another type of insurance contract ("tontine") was developed by an italian banker Lorenzo Tonti in 1652 and was pretty much similar to a life-insurance contract. This opened the way to other life-insurance contracts appeared in London in 1698. In the meantime one of the most famous insurance to individuals the "fire insurance" appeared in London again in 1676 ten years after the great fire of London.

It is time to give a very generic definition of an insurance contract.

Definition 2.1.1 (Insurance contract)**.** *An insurance contract is a contract in which in exchange of a payment (called the premium) at some prescribed time(s), the buyer receives cashflows depending on the occurence of random events (that are called sinisters or claims). We distinguish two main type of insurance contracts :*

- *Non-life insurance contracts : in which the life or the death of the buyer is not in play for the payment of claims*
- *Life-insurance contracts : in which the life or death of the buyer triggers some payment to the buyer or to one or several beneficiaries.*

An important feature of an insurance contract is the notion of inversion of the production cycle. If one aims in building and selling a chair, then he or she can price all the commodities involved and the workforce to determine the (intrinsic) price of the chair. It is then possible to choose the commercial margin. An insurance contract does not follow this line of actions as the seller has to determine the price of the contract before knowing the claims that will have to be paid in the future.

2.1.1 A general model

We are now ready to provide a first general model. As any general model it gives an overview of the main features but of course does not take into account more particular situations that need to be inspected in a second time. The wealth of an insurance company will be modeled as a stochastic process $(R_t)_{t\geq 0}$ where R_t models the wealth at time t. Why is it random ? This wealth only models the "insurance" activity so to say of the company (here we only aim in modeling any cashflow resulting from the insurance activity of the company so paying wages or real estate income is not taken into consideration) so it results over time of the difference of inwards cashflows and negative cashflows.

1. **Modeling of the inward cashflows**: we will assume that the positive cashflows will be represented by the premium that will be given by two parts : a positive initial capital $u \geq$ and a continuous time premium of the form pt with $p > 0$. Both coefficients u, p are assumed to be deterministic. This might lead to questions but if we come back to the very initial aim of an insurance contract (that is to transfer risk in exchange of one or several cashflows) this assumption seems to satisfy the basic requirement of an insurance contract.

2. **Modeling of the outward cashflows**: these cashflows will represent the sinisters (or claims) to be paid to the buyer of the contracts. They are the source of the randomness and uncertainty in an insurance contract. When one thinks about it there are two components in this uncertainty : one is called the *severity* and refers of the cost of a given claim and the other one the *frequency* refers to the random number of the claims. These two components of the risk are in practice essential. Each claim will be modeled as a positive random variable, more precisely we note X_i the ith claim (so X_i represents the cost the insurer will have to pay due to the ith sinister). We also need to model the random number of claims over time : this will be done using a counting process $N := (N_t)_{t\geq 0}$.

We now have all the ingredients to describe our model :

Definition 2.1.2 (A Cramér-Lundberg type model[1])**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which there exist a family of positive random variables $(X_i)_{i\ \geq 1}$ and $N := (N_t)_{t\geq 0}$ a counting process. Let $u \geq 0$ and $p > 0$. The wealth of the insurer (also called surplus process) is modeled by the stochastic process $R := (R_t)_{t\geq 0}$ with*

$$R_t = u + pt - \sum_{i=1}^{N_t} X_i, \quad t \geq 0. \tag{2.1.1}$$

This model rewrites as :

$$R_t = P_t - S_t, \quad t \geq 0$$

where $P_t := u + pt$ denotes the premium process (u is called the initial capital and p the premium rate) and $S_t := \sum_{i=1}^{N_t} X_i$ the total claim amount process (in French: processus de la charge sinistre agrégée).

[1]This model is named after its introduction by the swedish actuary Filip Lundberg in 1903; revisited by the swedish Mathematician Haral Cramér in the 30's.

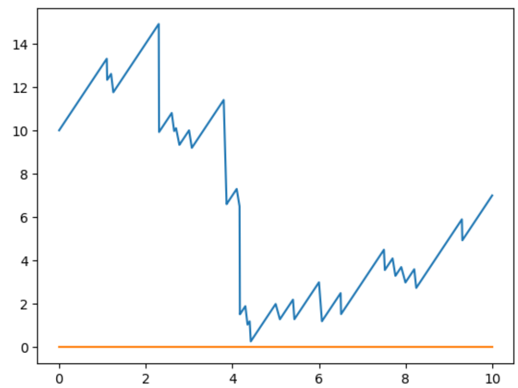

Figure 2.1: A sample path of the risk process

Maybe it is important to make precise the choice of a continuous model for the premium part. Indeed, it is common sense that an insurance contract will involve cashflows during the lifetime of the contract but we usually think of it as a lump sum paiement (one pays for example annuities of a given value). We will see that this formulation as a continuous time payment is equivalent to the one of a discrete time one. The main drawback of a discrete time model relies in the fact that the times of interest are not given deterministic dates but the random one at which a sinister occurs. We need some notations to make this point more precise.

Definition 2.1.3. *We denote for any $k \geq 1$, T_k the kth jump time of N, that is :*

$$T_k := \inf t > T_{k-1}; N_t = k; \quad T_0 := 0$$

and we write $W_k := T_k - T_{k-1}$ the inter-jump time. This is the time between sinister $k-1$ and sinister k.

With this definition we have of course that $T_n = \sum_{i=1}^{n} W_k$ and $T_{N_t} = \sum_{i=1}^{N_t} W_k$ giving that the model rewrites as :

$$R_t = u + pt - \sum_{i=1}^{N_t} X_i = u + \sum_{i=1}^{N_t}(X_i - W_i) + (t - T_{N_t})$$

in other words at a time $t = T_j$ a claim then the wealth at this time R_{T_j} equals $u + \sum_{i=1}^{j}(X_i - W_i)$ the sum of the initial premium and of of the values of the claims minus the time lap between the previous claim.

Definition 2.1.4 (Cramér-Lundberg model)**.** *We name the Cramér-Lundberg model a model of the form (2.1.1A Cramér-Lundberg type model[2]equation.2.1.1) where we assume that :*

1. *N is an homogeneous Poisson process with intensity $\lambda > 0$;*
2. *The random variables $(X_i)_{i\geq 1}$ are iid random variables and are independent of N.*

Additionally we will assume for convenience that $\mathbb{E}[|X_1|^2] < +\infty$ (the modeling of the so-called large claims case does not fit with this assumption).

[2]This model is named after its introduction by the swedish actuary Filip Lundberg in 1903; revisited by the swedish Mathematician Haral Cramér in the 30's.

2.1.2 Mutualisation and segmentation

One point might call for additional explanation : what is the wealth process representing ? Is it related to one contract or the overall wealth resulting from all the contracts ? To answer that question we simply might consider a set of n agents (buyers) together with $(X_i)^j_{i\geq 1}$ the claims related to the jth agent and N^j a counting process of these claims. Then the wealth associated to the jth insured is R^j with

$$R^j_t := u^j + p^j t - \sum_{i=1}^{N^j_t} X^j_i, \quad t \geq 0$$

and this the total wealth of the firm is R with

$$R_t := \sum_{j=1}^{n} R^j_t, \quad t \geq 0.$$

But what is exactly the risk the agents are facing ? Do they share the same risk ? Do they have the same risk profile ? The answer is yes and no! There are situations where the agents are somehow independent of each others and have the same exposure to a given risk : they form a somehow homogeneous community. This idea is at the heart of the concept of mutualisation. For instance the French state health insurance system is based on this assumption (people will not get sick at the same time; of course pandemics are period during which this assumption fails). Remember that to face a random risk the insurance firm is asking deterministic payments as a counterpart. This makes sense because of the mutualisation assumption, il which the sum of all the contract play the role of a large number of outcomes of the underlying randomness; the premium is then a mean in a similar fashion of the law of large numbers. In that case, Mathematically we can assume that all the claims are iid and that the counting processes as well. You have seen that for a Poisson process, a sump of independent Poisson processes is also a Poisson process with the sum of intensities as intensity. Hence under that assumption of mutualisation the risk model takes the form of the one we have introduced ($u = \sum_{j=1}^n u^j$; $p = \sum_{j=1}^n p^j$; $N = \sum_{j=1}^n N^j$ and we merge all the claims $(X^j_i)_{i,j}$.
But naturally the previous assumption of somehow homogeneity of the agents regarding the risk might not fit into some situations. A very famous case is the situations of car insurance in which one clearly identifies different profiles of drivers (risky drivers, safe drivers, occasional drivers, everyday drivers....). In that case the firm has to adapt the contract to the profile of the driver : this is called segmentation. To avoid any form of discrimination, the firm has to design a menu of contracts; each of them adapted to a specific profile of driver and he/she has to do so so that a given driver sees advantages in choosing the contract fitting with his/her profile. This leads to another field of the insurance industry related to behavioral Economics known under the name of Adverse selection which is a particular case of so-called Principal-Agents problem. We will not deal with this aspect and stick to the mutualisation assumption.

Beyond mutualisation and segmentation ? The insurance industry has to handle (and Mathematicians are welcome here!) new challenges where all the individuals are facing the same risk : this is called systemic risk. It includes global warming consequences (droughts, floods, fires...) and also the so-called Cyber-Risk.

2.2 Analysis of the model

We start with a discussion regarding what could be a convenient notion of premium[3].

2.2.1 What kind of premium ?

From now we model the surplus process (wealth of the firm) as in (2.1.1A Cramér-Lundberg type model[4]equation.2.1.1). More specifically we will assume that we are working within the Cramér-Lundberg model (see Definition **??**). In practice given the data one estimates the law of the common X_i's and the parameter λ of the Poisson process. But then how to determine the premium parameters u and p ? In contradistinction with Mathematical Finance there is not a canonical notion of price[5] resulting from the fundamental theorems of asset pricing; that is there is no notion of arbitrage price or risk measure here since we assume the risk is exogenous to any form of market as there is no financial market (one cannot sell and buy car accidents, fortunately!) So we need to provide an notion of premium. In fact this is an important feature of the insurance industry. Insurance firms follow strict regulations on how to perform their risk analysis; however historically Insurance perform their own analysis which usually results in more careful conclusions. There are very specific analyses of risks for very specific contracts but there are three main concepts which are related to the more general theory of risk management. Consider just for notation that we write $p(Y)$ the amount of cash that would allow one to cover a risk carried out by a random variable Y. We distinguish the notion of

1. Pure premium : $P(Y) := \mathbb{E}[Y]$

[3]We follow mostly [2, 3]

[4]This model is named after its introduction by the swedish actuary Filip Lundberg in 1903; revisited by the swedish Mathematician Haral Cramér in the 30's.

[5]one has to keep in mind though that many different notion of price can be designed in the context of pricing and hedging financial derivatives

2. Premium Charge (Fr. : "prime chargée") : $P(Y) := \mathbb{E}[Y](1+\eta)$ for some $\eta \geq 0$

3. Premium that satisfy some Economic axioms like :

   (a) Security margin : $P(Y) \geq \mathbb{E}[Y]$

   (b) No excess margin : $Y = y$, $\mathbb{P}-a.s. \Rightarrow P(Y) = y$

   (c) Additivity : if Y_1 and Y_2 are independent then $P(Y_1+Y_2) = P(Y_1) + P(Y_2)$

   (d) Subadditivity : for any Y_1 and Y_2, $P(Y_1+Y_2) \leq P(Y_1) + P(Y_2)$

   (e) Scale invariance : for any $a > 0$, $P(aY) = aP(Y)$

   (f) Translation invariant : for any $a > 0$, $P(Y+a) = P(Y) + a$

   (g) Maximum principle : for any $a > 0$, $Y \leq a$, $\mathbb{P}-a.s. \Rightarrow P(Y) \leq a$.

   This last concept is highly related tot the notion of Risk measures that we will study in the Chapter 4An introduction to Risk Measureschapter.4.

2.2.2 A first ordre approach : pure premium and safety loading (Fr.: "chargement de sécurité")

We will focus on the so-called pure premium analysis that is in a very intuitive approach we just try to choose u, p in order to control $\mathbb{E}[R_t]$. To do we first compute it and for this we rely on the so-called Wald's identity.

Lemma 2.2.1 (Wald's identity). *Let $(X_i)_{i\geq 1}$ be iid random variables with $\mathbb{E}[|X_1|] < +\infty$ and $N \sim \mathcal{P}(\mu)$ independent on $(X_i)_{i\geq 1}$, $\mu > 0$. Then*

$$\mathbb{E}\left[\sum_{i=1}^{N} X_i\right] = \mu\mathbb{E}[X_1].$$

Proof. We make use of the tower property to get

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{N} X_i\right] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{N} X_i \middle| \sigma(N)\right]\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{N} \mathbb{E}\left[X_i \middle| \sigma(N)\right]\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{N} \mathbb{E}\left[X_i\right]\right] \\
&= \mathbb{E}\left[X_1\right]\mathbb{E}\left[N\right] = \mu\mathbb{E}\left[X_1\right].
\end{aligned}$$

□

Proposition 2.2.2. *Let R the surplus process in the Cramér-Lundberg model, that is*

$$R_t = u + pt - \sum_{i=1}^{N_t} X_i, \quad t \geq 0,$$

with $(X_i)_{i\geq 1}$ iid random variables independent of $N \sim \mathcal{PP}(\lambda)$, $\lambda > 0$ and $\mathbb{E}[|X_1|^2] < +\infty$. We set $m := \mathbb{E}[X_1]$. For any $t \geq 0$:

$$\mathbb{E}[R_t] = u + (p - \lambda m)t, \quad t \geq 0.$$

Hence $\mathbb{E}[R_t] > 0$ for any $t \geq 0$ if and only if $\eta := \frac{p}{\lambda m} - 1 > 0$.

Definition 2.2.3. *The coefficient η is called the safety loading of the insurer and the condition $\eta > 0$ is called the net profit condition.*

It is clear that choosing p such that $\mathbb{E}[R_t] > 0$ for any t seems to be a reasonable minimal requirement for giving any form of control of the risk. We can now turn to a more precise notion of risk control.

2.2.3 Ruin probability

The notion of ruin is central in Insurance and corresponds to a regulation requirement : the firm must design its contracts so that ruin is avoided. In what follows we work with the surplus process

$$R_t = u + pt - \sum_{i=1}^{N_t} X_i, \quad t \geq 0$$

that we see as a mapping of the initial capital u. To avoid cumbersome notations we do not explicitly make use of u in the notation and simply write R.

Definition 2.2.4 (Ruin)**.** *The random time $\tau(u) := \inf\{t > 0; R_t \leq 0\}$ is called the ruin time. Of course we make use of the notation that $\inf \emptyset := +\infty$. We set the ruin probability the quantity :*

$$\Psi(u) := \mathbb{P}[\tau(u) < +\infty] = \mathbb{P}\left[\inf_{t>0} R_t < 0\right].$$

We also define the probability of solvency as $\psi(u) := 1 - \Psi(u)$.

The second order analysis we perform corresponds to a study of the ruin probability. We start with the following result :

Proposition 2.2.5.

$$\lim_{t\to+\infty} \frac{1}{t}\sum_{i=1}^{N_t} X_i = \lambda m, \ \mathbb{P} - a.s.$$

and

1. *if $\eta < 0$, then $\Psi(u) = 1$, for all $u > 0$*
2. *if $\eta > 0$, then there exists $u_0 > 0$ such that $\Psi(u) < 1$, for all $u > u_0$.*

Proof. We have using Wald's identity that

$$\mathbb{E}\left[\left|\frac{1}{N_t}\sum_{i=1}^{N_t} X_i - m\right|^2\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\left|\frac{1}{N_t}\sum_{i=1}^{N_t}(X_i - m)\right|^2\right] \\
&= \mathbb{E}\left[\frac{1}{N_t^2}\sum_{i=1}^{N_t}(X_i - m)^2\right] + 2\mathbb{E}\left[\frac{1}{N_t^2}\sum_{1=i<j\leq N_t}(X_i - m)(X_j - m)\right] \\
&= \mathbb{E}\left[|X - m|^2\right]\mathbb{E}\left[\frac{1}{N_t}\right] \to_{t\to+\infty} 0.
\end{aligned}
$$

As $\lim_{t\to+\infty}\frac{N_t}{t} = \lambda$, $\mathbb{P}$-a.s. we have that $\lim_{t\to+\infty}\sum_{i=1}^{N_t} X_i = \lambda m,\ \mathbb{P}-$ a.s.. Recall that $\eta = \frac{p}{\lambda m} - 1$ so

$$\eta > 0 \Leftrightarrow p > \lambda m; \quad \text{and} \quad \eta < 0 \Leftrightarrow p < \lambda m.$$

Letting $S_t := u - R_t = -pt + \sum_{i=1}^{N_t} X_i$ we have that

$$\Psi(u) = \mathbb{P}\left[\inf_{t>0}\{pt - \sum_{i=1}^{N_t} X_i\} < -u\right] = \mathbb{P}\left[-\sup_{t>0} S_t < -u\right] = \mathbb{P}\left[\sup_{t>0} S_t > u\right].$$

Assume $\eta < 0$, then $\frac{S_t}{t} \to_{t\to+\infty} -p + \lambda m > 0$, $\mathbb{P}$-a.s. meaning that $S_t \to_{t\to+\infty} +\infty$, $\mathbb{P}$-a.s. and so that $\Psi(u) = 1$ for any $u > 0$. If $\eta > 0$, then $\frac{S_t}{t} \to_{t\to+\infty} -p + \lambda m = -\infty$, $\mathbb{P}$-a.s. meaning that $\sup_{t>0} S_t < +\infty$, $\mathbb{P}$-a.s. and so there exists $u_0 > 0$ such that $\Psi(u) < 1$ for any $u > u_0$ as

$$\sup_{t>0} S_t = +\infty, \mathbb{P} - a.s. \Leftrightarrow \mathbb{P}\left[\sup_{t>0} S_t < u\right] = 1,\ \forall u > 0.$$

□

So what can we get out of this result ? Clearly as we have already observed it seems clear that choosing p such that $\eta > 0$ is the least to ask. From now we will assume that this net profit condition is satisfied. Then, in that case we get that there exists a level of initial capital u_0 from which $\Psi(u) < 1$. That does not sound really great but maybe it is. Until now we only focused on the choice of parameter p but this is here that the choice of u comes into play. Assume that p satisfies the net profit condition, then one can study the mapping $u \mapsto \Psi(u)$. Under specific conditions, we will prove that (see TD-TP) that for some choice of distribution of the claims one gets :

$$\Psi(u) \leq e^{Cu},\ u \geq u_0;\ C > 0$$

for some constant C that may be approximated and which depends on the law of X_1. This results is extremely interesting as it shows that the ruin probability decreases exponentially fast with u.

Chapter 3

Monte-Carlo methods for actuarial sciences in a nutshell

We have seen that one of the main object of interest is the notion of ruin probability. We will work in TD-TP on some theoretical treatment of it but we may also be interested to investigate the numerical simulation of the risk process or of related quantities. Most of the material presented in this chapter might be known to the reader, but if not we decide to cover the basic elements so that this chapter is self-contained.

3.1 A detour on the random simulation of a random variable

Throughout this section, X will denote a real-valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 3.1.1 (Cumulative distribution)**.** *We name Cumulative Distribution Function (Cdf) of X the mapping F_X*

$$F_X(x) := \mathbb{P}[X \leq x], \; x \in \mathbb{R}.$$

Proposition 3.1.2. *F_X is non-decreasing, left-limited, right-continuous, valued in $[0,1]$ and*

$$F_X(x-) = \mathbb{P}[X < x], \quad \text{and} \quad \Delta_x F_X := F_X(x) - F_X(x-) = \mathbb{P}[X = x]; \; \forall x \in \mathbb{R}$$

so F_X has a discontinuity at a point x if and only if $\mathbb{P}[X = x] > 0$ (the distribution of X is said to have atoms).

We will give more precisions on the notion of quantile functions in the next chapter (see for instance Definition 4.3.2Quantileprop.4.3.2) but we recall below the so-called left quantile :

Definition 3.1.3. *For $\lambda \in [0,1]$, we set*

$$q_X^-(\lambda) = \inf\{z \in \mathbb{R}, \; \mathbb{P}[X \leq z] \geq \lambda\}.$$

Proposition 3.1.4. *The map q_X^- is non-decreasing, left-continuous, right-limited and*

1. $q_X^-(F_X(\lambda)) \geq \lambda, \; \forall \lambda \in [0,1]$

2. $[q_X^-(\lambda) \leq x] \Leftrightarrow [\lambda \leq F_X(x)],\ \forall \lambda \in [0,1],\ \forall x \in \mathbb{R}$

3. $[F_X(q_X^-(\lambda)) = \lambda] \Leftrightarrow [F_X$ *is continuous at* $q_X^-(\lambda)],\ \forall \lambda \in [0,1].$

With these properties at hand we can now come to the main notion for the simulation of a r.v. X.

Proposition 3.1.5. *1.* $q_X^-(U) \overset{\mathcal{L}}{=} X, \quad U \sim \mathcal{U}([0,1])$

2. If F_X *is continuous then* $F(X) \sim \mathcal{U}([0,1])$.

Proof. 1. We set $Y := q_X^-(U), \quad U \sim \mathcal{U}([0,1]$ and compute the cdf of Y. We have for $x \in \mathbb{R}$

$$\begin{aligned} F_Y(x) &= \mathbb{P}[Y \leq x] \\ &= \mathbb{P}[q_X^-(U) \leq x] \\ &= \int_0^1 \mathbf{1}_{q_X^-(u) \leq x} du \\ &= \int_0^1 \mathbf{1}_{u \leq F_X(x)} du \\ &= F_X(x). \end{aligned}$$

2. As $X = \overset{\mathcal{L}}{=} q_X^-(U)$ and since F_X is continuous $F_X(X) \overset{\mathcal{L}}{=} F_X(q_X^-(U)) \overset{\mathcal{L}}{=} U$.

□

The takeaway message of the previous proposition is that it reduces the simulation of samples of X to the one of uniform random variables knowing the quantile function q_X^-. This is at the core of the random simulators in any programmation language. First of all it is not easy to generate a uniform random variable because it is not clear how to generate a random number. In fact it is not (we use the terminology of pseudo-random generators) as they all follow a deterministic algorithm based on a numerical sequence involving a so-called seed which corresponds to the initial value of the numerical sequences. Second, q_X^- may not be attainable, like for a Gaussian random variable.

3.2 Foundations of the Monte-Carlo simulation

We have seen that we may interested in computing an expectation or a probability. This might not be theoretically attainable and thus calls for a numerical proxy. One of the most popular stochastic method for that purpose has been introduced in the middle of the 20th century in the context of the Manhattan project and is referred to the *Monte-Carlo method* after the physicist Nicholas Metropolis. We describe the basics of this method in this section.

Once again, throughout this section, X will denote a real-valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}[|X|^2] < +\infty$. We set

$$m := \mathbb{E}[X]; \quad \sigma := \sqrt{\mathbb{E}[X^2] - \mathbb{E}[X]^2}.$$

Our aim is to provide an algorithm that approximate the value m. To this end we come back to (basic) Statistics. We set

Definition 3.2.1. *Let $n \geq 1$.*

1. *A n-sample of X is a family of random variables $(X_1, \dots, X_n)$ where the random variables are defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and are iid (independent and identically distributed) with common law X.*
2. *Given $(X_1, \dots, X_n)$ a n-sample of X we name S_n (later on refereed as the empirical estimate of m) the quantity :*

$$S_n := \frac{1}{n} \sum_{i=1}^{n} X_i.$$

As its name suggests, S_n provides an estimator of m. This result follows from the Law of Large Numbers; we recall below a version of the Strong Law of Large Numbers :

Proposition 3.2.2. *S_n provides a strongly consistent estimator of m that is*

$$\lim_{n \to +\infty} S_n = m, \ \mathbb{P} - a.s..$$

Proposition 3.2.3. 1. *$\mathbb{E}[S_n] = m$ for all $n \geq 1$, that is S_n is an unbiased estimate of m.*

2. *$v(S_n) := \mathbb{E}[|S_n|^2] - \mathbb{E}[S_n]^2 = \frac{\sigma^2}{n}$ for all $n \geq 1$.*

Proof. We only focus on the variance calculation.

$$\begin{aligned} v(S_n) &= \mathbb{E}[(S_n - m)^2] \\ &= \mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - m)\right|^2\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}[|X_i - m|^2] + 2\frac{1}{n^2}\sum_{1 \leq i < j \leq n}\mathbb{E}[(X_i - m)(X_j - m)] \\ &= \frac{\sigma^2}{n} + 0 \end{aligned}$$

since the variables $X_1, \dots, X_n$ are iid. □

So we have an estimator of m but how we do not need the precision of this estimation. These information are given by the Central Limit Theorem.

Proposition 3.2.4. 1. *Let $Z \sim \mathcal{N}(0,1)$, we have that*

$$\frac{\sqrt{n}(S_n - m)}{\sigma} \xrightarrow[n \to +\infty]{\mathcal{L}} Z.$$

In particular for any $a < b$,

$$\lim_{n \to +\infty} \mathbb{P}\left[\frac{\sqrt{n}(S_n - m)}{\sigma} \in [a,b]\right] = \mathbb{P}[a \leq Z \leq b] = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

2. *For $\alpha \in (0,1)$ we set $z_{1-\frac{\alpha}{2}}$ the $\frac{\alpha}{2}$-quantile of z (that is $\mathbb{P}[Z > z_{1-\frac{\alpha}{2}}] = \frac{\alpha}{2}$). We have*

$$\mathbb{P}\left[S_n - \frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}} \leq m \leq S_n + \frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}\right] = 1 - \alpha,$$

in other words $IC_{1-\alpha} := \left[S_n - \frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}; S_n + \frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}\right]$ is a Confidence Intervalle for m with level $1 - \alpha$.

Discussion on the precision of the Monte-Carlo approximation

We denote by $IC_{1-\alpha}|$ the length of the confidence interval and note that $|IC_{1-\alpha}| = \frac{2\sigma z_{1-\frac{\alpha}{2}}}{\sqrt{n}}$; this provides the precision of the Monte-Carlo approximation. If the error e is of form $e := 10^{-p}$ for some $p \geq 1$ then one needs to choose $n = O(2p)$ so that $|IC_{1-\alpha}| \leq e$. This gives the classical rate of convergence of the Monte-Carlo method its precision behaves as $n^{-1/2}$. However this is some sort of first ordre approximation as σ^2 (the variance of the random variable X) comes into play and appears in the precision of the approximation. It is thus important to have an information on it. It might even be unknown. In that case one has to adapt his/her strategy to provide at the same time an estimation of σ. We have

Proposition 3.2.5. *Let $n \geq 2$ and $(X_1, \ldots, X_n)$ a n-sample of X. We set :*

$$\hat{\sigma}_n^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - S_n)^2.$$

We have that for $Z \sim \mathcal{N}(0,1)$

$$\frac{\sqrt{n}(S_n - m)}{\hat{\sigma}_n^2} \xrightarrow[n \to +\infty]{\mathcal{L}} Z$$

and $IC_{1-\alpha,\hat{\sigma}_n} := \left[S_n - \frac{z_{1-\frac{\alpha}{2}} \hat{\sigma}_n}{\sqrt{n}}; S_n + \frac{z_{1-\frac{\alpha}{2}} \hat{\sigma}_n}{\sqrt{n}} \right]$ *is a Confidence Intervalle for m with level $1-\alpha$.*

Proof. This follows from Slutsky's theorem and the convergences :

$$\frac{\sqrt{n}(S_n - m)}{\sigma} \xrightarrow[n \to +\infty]{\mathcal{L}} Z,$$

$$\frac{\hat{\sigma}_n}{\hat{\sigma}_n} \to_{n \to +\infty} 1, \quad \text{in } \mathbb{P} \text{ probability}.$$

□

3.2.1 What about rare events ?

Definition 3.2.6. *Let $A \in \mathcal{F}$ be an event. We say that A is a rare event if $\mathbb{P}[A] \leq 10^{-4}$.*

Of course the threshold of 10^{-4} is somehow arbitrary, but the takeaway message is that it is of interest in Engineering to control the reliability of a device, or system or process. Depending on the application one has to deliver a process together with a reliability analysis whose level will of course depend on the application. The purpose of this section is to point out that when $\mathbb{P}[A]$ is small it may lead to some numerical issue.

As an illustration consider A an event and let $p := \mathbb{P}[A]$. Our aim is to use the MC scheme for estimating p; indeed $p = \mathbb{E}[\mathbf{1}_A]$. To this end, let ≥ 1 and $(\mathit{1}_{A_1}, \ldots, \mathit{1}_{A_n})$ a n-sample of the random variable $X := \mathbf{1}_A$. We name $\hat{p}_n$ the empirical mean estimateur S_n that is :

$$\hat{p}_n := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{A_i}.$$

It is clear that $n\hat{p}_n \sim \mathcal{B}(n,p)$ and thus

$$\mathbb{E}[\hat{p}_n] = p; \quad v(\hat{p}_n) = \frac{p(1-p)}{n}.$$

One can build a (theoretical) confidence interval of level $1-\alpha$ ($\alpha \in (0,1)$) for p whose length is given by $2z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = 2z_{1-\frac{\alpha}{2}}\frac{p(1-p)}{\sqrt{n}}$. We use the word "theoretical" as obviously we are in the case where $\sigma = \sqrt{p(1-p)}$ is unknown, our goal is just to point out that even at the theoretical level the size of the confidence interval is equal up to a universal constant to $\sqrt{\frac{p(1-p)}{n}}$. We can thus define (once again the theoretical) the relative error $e(n) := \frac{\sqrt{\frac{p(1-p)}{n}}}{p} = \sqrt{\frac{(1-p)}{np}}$. For instance

p	10^{-1}	10^{-4}	10^{-7}
$n;\ e(n) = 10\%$	10^{-3}	10^{-6}	10^{-9}
$n;\ e(n) = 1\%$	10^{-5}	10^{-8}	10^{-11}

and if one takes into account the constant $2z_{1-\frac{\alpha}{2}}$ it amounts to something like 4 for $\alpha = 0.05$. So for this choice of α, and a relative error of 10%, n has to amount to 410^9 to estimate $p = 10^{-7}$.

To simplify all these considerations, let us just mention that statistically speaking in average $\sum_{i=1}^n \mathbf{1}_{A_i} = 1$ for $n = O(p^{-1})$ and thus $\hat{p}_n = \frac{1}{n} = O(p)$, hence for $p = 10^{-7}$ this is of the order of the numerical error. To illustrate this phenomenon we suggest the following exercise.

Exercise 1 : Let $X \sim \mathcal{N}(0,1)$. Using a numerical approximation we have an estimation of $p := \mathbb{P}[X > 5] \equiv 2.8710^{-7}$, it is thus a rare event. For $n \geq 2$, we consider $(X_1, \ldots, X_n)$ a n-sample of X.

1. Recall the empirical mean estimator for p (we name it $\hat{p}_n$).
2. Let $\alpha = 0.10$. We approximate $z_{1\frac{\alpha}{2}} = z_{0.95}$ by 1.96. Simulate 50 outcomes of the confidence interval $IC_{1-\alpha}$ either by assuming that the variance is known and by using the estimation $p \equiv 2.8710^{-7}$ or by estimating it. Give the proportion of these confidence intervals which contained the approximate value of $p \equiv 2.8710^{-7}$ for different values of n : 10^5, 10^6, 10^7.

3.3 The importance sampling MC method

Nota : The importance sampling method translates in French as "échantillonnage d'importance".

3.3.1 Introduction and motivation

We have seen that when working with rare events the MC method fails to give an appropriate proxy. The reason lies in the fact that for an event of probability of ordre 10^5 over 10^5

simulations of copies 1_{A_i} statistically only one will be equal to 1 or put it differently over 10^5 simulations, statistically

$$\hat{p}_n \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{A_i} = \frac{1}{n} = 10^{-5}.$$

In the example of the previous exercise $A = \{X > 5\}$ with $X \sim \mathcal{N}(0,1)$. It is clear that one simulates outcomes of X with probability $1-\alpha$ the values simulated lies in $[-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}]$:

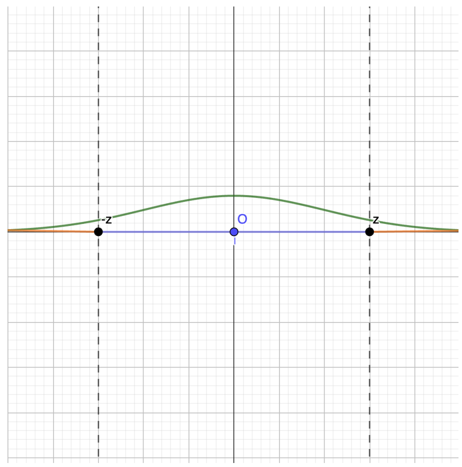

Figure 3.1: Quantile (designed with GeoGebra)

The values around 5 we would like to simulated belong to the area where the probability of outcome is very small. One would need to shift the distribution so that the region around 5 belongs to the zone between the quantiles as follows :

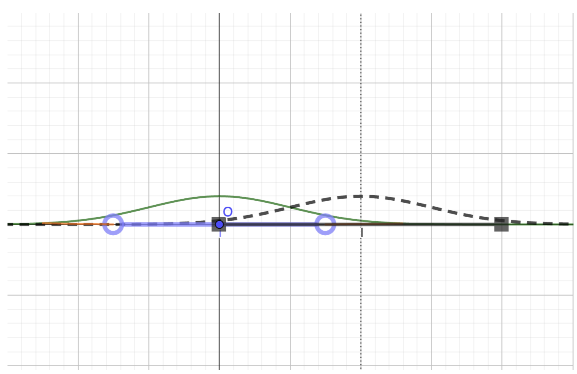

Figure 3.2: Shifted distribution (designed with GeoGebra)

This method is at the heart of the importance sampling method.

3.3.2 Principle of the importance sampling method

Through this section we consider $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space on which a random variable X (with $\mathbb{E}[|X|^2] < +\infty$) and on which for any $n \geq 1$, a n-sample $(X_1, \ldots, X_n)$ can be defined. We also consider $h : \mathbb{R}\ \mathbb{R}_+^*$ in $L^1(\mathbb{R}, dx)$. We assume in addition that X has a density f

on $\mathbb{R}$ that is[1] there exists $f : \mathbb{R} \to \mathbb{R}_+$ with $\int_{\mathbb{R}} |f(x)|^2 dx < +\infty$ and $\int_{\mathbb{R}} f(x)dx = 1$ and $F_x(x) = \int_{-\infty}^{x} f(t)dt$. Our objective is to numerically approximate the quantity m :

$$m := \mathbb{E}_f[h(X)] = \int_{\mathbb{R}} h(x)f(x)dx,$$

note that to be completely precise we make use of the notation $\mathbb{E}_f$ this means that the random variable X has density f w.r.t. the Lebesgue measure. To numerically approximate m we may of course make use of the classical MC algorithm. More precisely, we recall the estimator $\hat{m}_n^{(1)}$

$$\hat{m}_n^{(1,f)} := \frac{1}{n}\sum_{i=1}^{n} h(X_i),$$

where the X_i are sampled according to f. We have seen the limitation of this approach. Consider then $g : \mathbb{R} \to \mathbb{R}_+^*$ (in particular $\{f > 0\} \subset \{g > 0\}$). We have

$$\begin{aligned} m &= \mathbb{E}_f[h(X)] \\ &= \int_{\mathbb{R}} h(x)f(x)dx \\ &= \int_{\mathbb{R}} \frac{h(x)f(x)}{g(x)} g(x)dx \\ &= \mathbb{E}_g[\tilde{h}(X)], \end{aligned}$$

where $\tilde{h}(x) := \frac{h(x)f(x)}{g(x)}$ and $\mathbb{E}_g[\cdot]$ means that X has density g with respect to the Lebesgue measure. Hence we can apply the MC method to this new formulation and consider the MC estimator of m, $\hat{m}_n^{(2)}$ as

$$\hat{m}_n^{(2,g)} := \frac{1}{n}\sum_{i=1}^{n} \tilde{h}(X_i),$$

where the X_i are sampled according to g. What can we say about these two estimators ?

Proposition 3.3.1. *1.* $\mathbb{E}_f[\hat{m}_n^{(1,f)}] = m = \mathbb{E}_g[\hat{m}_n^{(2,g)}]$.

2. The lenght of the confidence intervals of some level $1-\alpha$ ($\alpha \in (0,1)$) are given by

$$|IC_{1-\alpha}(\hat{m}_n^{(1,f)})| = \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\sqrt{\mathbb{E}_f[h(X)^2] - m^2}; \quad |IC_{1-\alpha}(\hat{m}_n^{(2,g)})| = \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\sqrt{\mathbb{E}_g[\tilde{h}(X)^2] - m^2}.$$

In other words the introduction of g is worth trying if

$$\int_{\mathbb{R}} |h(x)|^2 f(x)dx > \mathbb{E}_f[h(X)^2] \leq \mathbb{E}_g[\tilde{h}(X)^2 = \int_{\mathbb{R}} |\tilde{h}(x)|^2 g(x)dx.$$

Hence : can we find some g^* such that

$$\int_{\mathbb{R}} |h(x)|^2 f(x)dx > \int_{\mathbb{R}} |\tilde{h}(x)|^2 g^*(x)dx = \int_{\mathbb{R}} |\tilde{h}(x)|^2 g^*(x)dx?$$

[1]Here we deal with r.v. with a density on $\mathbb{R}$. In fact all the results presented here can be obtained for a random variable with state space $\mathbb{N}$, $\mathbb{R}$, etc equipped with a measure ν.

Note that for any g

$$\int_{\mathbb{R}} |\tilde{h}(x)|^2 g(x)dx = \int_{\mathbb{R}} \frac{|h(x)f(x)|^2}{|g(x)|^2}|g^*(x)dx = \int_{\mathbb{R}} \frac{|h(x)f(x)|^2}{g(x)}dx$$

and so we can try to minimize the previous quantity to get an optimal $g^*(x) := \frac{h(x)f(x)}{\mathbb{E}_f[h(X)]} = \frac{h(x)f(x)}{m}$ for which :

$$\begin{aligned}
&v_{g^*}\left(\tilde{h}(X)\right) \\
&= \mathbb{E}_{g^*}[|\tilde{h}(X)|^2] - m^2 \\
&= \int_{\mathbb{R}} |\tilde{h}(x)|^2 g^*(x)dx - m^2 \\
&= \int_{\mathbb{R}} \frac{|h(x)f(x)|^2}{g^*(x)}dx - m^2 \\
&= \int_{\mathbb{R}} \left|\frac{h(x)f(x)}{g^*(x)}\right|^2 g(x)dx - m^2 \\
&= \int_{\mathbb{R}} \left|\frac{h(x)f(x)}{g^*(x)}\right|^2 g(x)dx - m^2 = 0.
\end{aligned}$$

So the variance can be reduced to 0, but it involves m in the expression of g^* ? So it proves that the variance can be reduced but how to use it in practice ?

3.3.3 Illustration on the rare event probability estimation

We come back to the setting of Exercise 1What about rare events ?exo.1. Consider $X \sim \mathcal{N}(0,1)$ and $p := \mathbb{P}[X > c]$ with $c = 5$. We have seen that the usual MC estimator $hatm_n^{(1,f)}$ (with $f(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$ and $h(x) := \mathbf{1}_{x>c}$) is not appropriate to simulate p. For $\theta \in \mathbb{R}$ we set $g_\theta(x) := \frac{e^{-\frac{(x-\theta)^2}{2}}}{\sqrt{2\pi}}$. We have

$$\begin{aligned}
p &= \mathbb{E}_f[\mathbf{1}_{X>c}] \\
&= \int_{\mathbb{R}} \mathbf{1}_{x>c} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}dx \\
&= \int_{\mathbb{R}} \mathbf{1}_{x>c} \frac{e^{-\frac{x^2}{2}}}{e^{-\frac{(x-\theta)^2}{2}}} \frac{e^{-\frac{(x-\theta)^2}{2}}}{\sqrt{2\pi}}dx \\
&= \int_{\mathbb{R}} \mathbf{1}_{x>c} e^{-\theta x} e^{\frac{\theta^2}{2}} \frac{e^{-\frac{(x-\theta)^2}{2}}}{\sqrt{2\pi}}dx \\
&= \mathbb{E}_{g_\theta}[\tilde{h}(X)],
\end{aligned}$$

with $\tilde{h}(x) := \mathbf{1}_{x>c} e^{-\theta x + \frac{\theta^2}{2}}$. We can try to minimize the variance of the estimator $\hat{m}_n^{(2,g_\theta)}$ with respect to θ which amonts to minimize :

$$\vartheta(\theta) := \mathbb{E}_{g_\theta}[\tilde{h}(X)^2].$$

We have

$$\begin{aligned}
\vartheta(\theta) &= \mathbb{E}_{g_\theta}[\tilde{h}(X)^2] \\
&= \int_c^{+\infty} e^{-2\theta x+\theta^2} \frac{e^{-\frac{(x-\theta)^2}{2}}}{\sqrt{2\pi}} dx \\
&= \int_c^{+\infty} e^{-2\theta x+\theta^2} \frac{e^{-\frac{(x-\theta)^2}{2}}}{\sqrt{2\pi}} dx \\
&= \int_c^{+\infty} e^{-2\theta x+\theta^2+x\theta-\frac{\theta^2}{2}} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \\
&= \int_c^{+\infty} e^{-\theta x+\frac{\theta^2}{2}} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \\
&= \mathbb{E}_{g_0}\left[\mathbf{1}_{X>c}\, e^{-\theta X+\frac{\theta^2}{2}}\right]
\end{aligned}$$

hence one might choose $\theta > c$ and even optimally $\theta^* = c$ in that case we have a second moment

$$\mathbb{E}_{g_0}\left[\mathbf{1}_{X>c}\, e^{-cX+\frac{c^2}{2}}\right] = e^{\frac{c^2}{2}}\ \mathbb{E}_f\left[\mathbf{1}_{X>c}\, e^{-cX}\right].$$

Maybe the reader should try to compute some confidence intervals for this new estimator like in Exercise 1What about rare events ?exo.1 and compute the number of these confidence intervals that contain p.

We will also apply the importance sampling method for approximating the ruin probability in the Cramér-Lundberg model.

Chapter 4

An introduction to Risk Measures

The material presented in these notes can be found in [4, 5].

4.1 Motivation and general notations

Throughout these notes $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space and T a finite positive real number.

Definition 4.1.1 (Set of financial positions)**.** *A financial position is any bounded random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$. We set $\mathcal{X}$ the set of financial positions :*

$$\begin{aligned}\mathcal{X} &:= \{X \in L_\infty(\Omega, \mathcal{F}, \mathbb{P})\} \\ &= \left\{X : \Omega \to \mathbb{R},\ \textit{such that } X \textit{ is } \mathcal{F}-\textit{measurable and } \|X\|_\infty := \sup_{\omega\in\Omega} |X(\omega)| < +\infty\right\}.\end{aligned}$$

Hence, X represents here the payoff[1] associated to a financial position or its return[2]. Before going further we give two examples of financial positions.

Example 4.1.2. *1. Consider a financial market composed of a risky asset $S := (S_0, S_T)$ and a riskless asset $S^0 := (S^0_0, S^0_T)$ with S_t (resp. S^0_t) the value of the asset S (resp. S^0) at time t ($t = 0$ or $t = T$). Riskless asset means here that S^0 is associated to an interest rate $r \geq 0$ as follows :*

$$S^0_0 := 1, \quad S^0_T := 1 + r.$$

For the risky asset, $S_0 > 0$ is a positive number and $S_T > 0$ is a $\mathcal{F}$-measurable bounded random variable. A portfolio (which represents an investment on this market) is a pair of real numbers (x, π) with $x \geq 0$ and $\pi \in \mathbb{R}$ which represents the number of risky assets S in the portfolio, so that the wealth associated $X^{(x,\pi)} := (X^{(x,\pi)}_0, X^{(x,\pi)}_T)$ to a portfolio (x, π) is defined as :

$$X^{(x,\pi)}_0 := x, \quad X^{(x,\pi)}_T := \pi S_T + (x - \pi S_0)(1 + r).$$

[1] Fr. Gain, c'est à dire la valeur à la maturité T

[2] Fr. Rendement

Note that x is the capital initial and that π could be negative which corresponds to a short sale[3]*. Note as well that the value which is not invested on the risky asset is invested on the reiskless asset.*

$X_T^{(x,\pi)}$ for some portfolio (x,π) is an example of a financial position. In that case note that X is non-negative. Another way to look at such an investment is to consider the financial position X defined below and which is the return[4] *of the discounted value*[5] *of $X_T^{(x,\pi)}$ associated to this portfolio:*

$$X := \frac{(1+r)^{-1}X_T^{(x,\pi)} - X_0^{(x,\pi)}}{X_0^{(x,\pi)}} = \frac{\pi((1+r)S_T - S_0)}{x},$$

which is not necessarily a positive random variable.

2. *The second example is the one of a cumulated loss R_T in Insurance in the Cramér-Lundberg model (with $c = 0$) as follows :*

$$R_T = u - \sum_{i=1}^{N_T} Y_i,$$

where N_T is a Poisson random variable with parameter $\lambda T > 0$ and $(Y_i)_{i\geq 1}$ is a sequence of iid bounded random variables. In that case X can be takes as : $X = R_T \wedge p$ for p in $\mathbb{N}$ (so that X is bounded).

What we aim in these notes is to study a tool (that we will call a risk measure) which allows one to "measure" the risk associated to a given financial position. We mention that the modern theory of risk measures has been introduced in [1].

4.2 Risk measures

4.2.1 Definition

We aim at defining a tool, that is a map $\rho : \mathcal{X} \to \mathbb{R}$ which associates to any financial position X a real number $\rho(X)$ that models the risk carried out by the position X. More precisely, given a financial position X, $\rho(X)$ represents the *monetary measure of risk of the financial position X.*

Remark 4.2.1. *What does the word "monetary" mean ? To answer that question, we need to explain what the real number $\rho(X)$ stand for. In this model, $\rho(X)$ will be seen as the quantity of cash (positive or negative) such that $X + \rho(X)$ is "acceptable". In particular, assume for simplicity that "Y is acceptable if and only if $Y \geq 0$, $\mathbb{P}$-a.s.", then :*

- *$\rho(X) > 0$ means that X is risky, as one needs to add a positive amount of cash ($\rho(X)$) to X so that $X + \rho(X)$ is acceptable. Hence in that case, $\rho(X)$ represents a reserve that covers (partly) the losses of the financial position X.*

[3] Fr. : Vente à découvert

[4] Fr. : Rendement

[5] The discounted value of $X_T^{(x,\pi)}$ is defined as : $\frac{X_T^{(x,\pi)}}{1+r}$

- *$\rho(X) < 0$ means that X is riskless, as one can withdraw a positive amount of cash $(-\rho(X))$ so that $X + \rho(X)$ is acceptable. Hence in that case, $\rho(X)$ can be seen as an extra amount of cash that can be consumed.*

What is a "good" measure of risk ? We can consider some examples :

Example 4.2.2.
Set $\rho_1 : \mathcal{X} \to \mathbb{R}$ as follows :

$$\rho_1(X) := \mathbb{E}[-X].$$

Set $\rho_2 : \mathcal{X} \to \mathbb{R}$ as follows :

$$\rho_2(X) := \sup_{\omega\in\Omega} -X(\omega) = -\inf_{\omega\in\Omega} X(\omega).$$

Set $\rho_3 : \mathcal{X} \to \mathbb{R}$ as follows :

$$\rho_3(X) := \mathbb{E}^{\mathbb{Q}}[-X] = \mathbb{E}\left[-X\frac{d\mathbb{Q}}{d\mathbb{P}}\right],$$

where $\frac{d\mathbb{Q}}{d\mathbb{P}}$ is a positive $L^1(\Omega, \mathcal{F}, \mathbb{P})$-random variable with $\mathbb{E}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\right] = 1$.

One sees that there are many ways of defining a risk measure. In order to be useful, we would like that a risk measure to satisfy some axioms (which make sense from the economics point of view). As an example, if a position is less risky than another one then the associated risk should be ordered in a logical way. Another example, is the one of diversification : it is an economics principle that diversification should reduce the risk. One may like to have a risk measure that satisfies such principles (that we will call axioms).

Definition 4.2.3 (Axioms). *Let $\rho : \mathcal{X} \to \mathbb{R}$ be a map. We say that :*

1. *ρ is monotonic if*
$$\forall X, Y \in \mathcal{X},\ [X \leq Y,\ \mathbb{P}-a.s. \Rightarrow \rho(X) \geq \rho(Y)],$$
2. *ρ is translation (or cash) invariant if*
$$\forall X \in \mathcal{X},\ \forall m \in \mathbb{R}, \quad \rho(X+m) = \rho(X) - m,$$
3. *ρ is normalized if*
$$\rho(0) = 0,$$
4. *ρ is convex if*
$$\forall X, Y \in \mathcal{X}, \forall \lambda \in [0,1], \quad \rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y),$$
5. *ρ is positive homogeneous if*
$$\forall X \in \mathcal{X}, \forall \beta \geq 0, \quad \rho(\beta X) = \beta\rho(X),$$
6. *ρ is subadditive if*
$$\forall X, Y \in \mathcal{X}, \quad \rho(X+Y) \leq \rho(X) + \rho(Y).$$

We comment on the meaning of these axioms.

Remarks 4.2.4. *1. $X \leq Y$, $\mathbb{P}$-a.s. means that X is riskier than Y (take for example $Y = 0$ so X only leads to losses) so the amount of cash that needs to be added to X so that it becomes admissible (for instance non-negative) is higher than the one needed for Y.*

2. *Translation invariance just means that if one adds a (deterministic) amount of cash to X, then the amount of cash needed to make it acceptable is substracted by m (if $m > 0$, the risk is reduced, the risk is increased otherwise).*

3. *$\rho(0) = 0$ is just a convention.*

4. *The convex feature is the mathematical translation of the diversification principle in Economics.*

5. *The positive homogeneous principle, means that the risk is somehow proportional but only for positive factors. Indeed, if one would ask this property to be true for any β then it would imply that ρ is linear. So we would end up with examples 1 and 3 in Examples 4.2.2prop.4.2.2 only.*

Definition 4.2.5 (Risk measure)**.** *A risk measure is a map $\rho : \mathcal{X} \to \mathbb{R}$ which is cash invariant and monotonic.*

We will always assume that a risk measure is normalised.

Definition 4.2.6. *A risk measure ρ is said to be a*

1. *convex risk measure if it is convex,*

2. *coherent risk measure if it is convex and positive homogeneous.*

Exercise 2 : For each risk measure of Examples 4.2.2prop.4.2.2, list the properties (of Definition 4.2.3Axiomsprop.4.2.3) that are satisfied.

Exercise 3 : Let ρ be a risk measure. Prove that if ρ satisfy two of the three properties below then it satisfies the third one :

- Convexity
- Positive homogeneous
- Subadditive.

Proposition 4.2.7. *Let ρ be a risk measure. Then ρ is a Lipschitz function with respect to the sup norm that is :*

$$|\rho(X) - \rho(Y)| \leq \|X - Y\|_\infty, \quad \forall X, Y \in \mathcal{X},$$

where we recall that $\|X\|_\infty = \sup_{\omega \in \Omega} |X(\omega)|$.

Proof. Let $X, Y \in \mathcal{X}$. We have that either $\rho(X) - \rho(Y) > 0$ or $\rho(X) - \rho(Y) \leq 0$ (these two cases being symmetric). Assume $\rho(X) - \rho(Y) \leq 0$. We have that :

$$X - Y \leq \|X - Y\|_\infty,\ \mathbb{P} - \text{a.s.},$$

so

$$X \leq Y + \|X - Y\|_\infty,\ \mathbb{P} - \text{a.s.}.$$

So by monotonicity and cash invariance, we have that

$$\rho(X) \geq \rho(Y + \|X - Y\|_\infty) = \rho(Y) - \|X - Y\|_\infty.$$

So

$$0 \geq \rho(X) - \rho(Y) \geq -\|X - Y\|_\infty,$$

and thus $|\rho(X) - \rho(Y)| \leq \|X - Y\|_\infty$. The case where $\rho(X) - \rho(Y) > 0$ is treated by exchanging the role of X and Y.

□

4.2.2 Acceptance set

We have seen that for a given risk measure ρ, and a financial position X, $\rho(X)$ denotes an amount of cash such that $X + \rho(X)$ is acceptable. But what does that mean ?

Definition 4.2.8 (Acceptance set)**.** *Let ρ be a risk measure. The acceptance set $\mathcal{A}_\rho$ associated to ρ is defined as :*

$$\mathcal{A}_\rho := \{X \in \mathcal{X},\ \rho(X) \leq 0\}.$$

In other words, a position X is acceptable (from the perspective of ρ) if its monetary measure of risk $\rho(X)$ is non-positive. In other words, it is not a reserve but a consumption. Why does it make sense ? Consider a risk measure ρ (recall it is normalised). By cash invariance, we have that :

$$\rho(X + \rho(X)) = \rho(X) - \rho(X) = 0.$$

So $X + \rho(X)$ belongs to $\mathcal{A}_\rho$. In addition, by monotonicity, $\rho(X)$ seems to be the minimal value which when added to X makes it acceptable. We will prove this claim in the result below.

Theorem 4.2.9. *Let ρ be a risk measure. Let $\mathcal{A}_\rho$ the acceptance set associated to ρ. We have :*

1. *$\mathcal{A}_\rho$ is non-empty, sequentially closed with respect to the sup norm.*

2. *Let X in $\mathcal{A}_\rho$ and Y in $\mathcal{X}$ such that $Y \geq X$, $\mathbb{P}$-a.s. then Y belongs to $\mathcal{A}_\rho$.*

3. *For any X in $\mathcal{X}$,*
$$\rho(X) = \inf\{m \in \mathbb{R},\ X + m \in \mathcal{A}_\rho\}.$$

4. *For any X in $\mathcal{A}_\rho$, $X + \rho(X)$ belongs to $\mathcal{A}_\rho$.*

5. *ρ is convex if and only if $\mathcal{A}_\rho$ is convex.*

6. *ρ is positive positive homegeneous if and only if $\mathcal{A}_\rho$ is cone.*

7. *ρ is coherent if and only if $\mathcal{A}_\rho$ is convex cone.*

Proof. 1. By normalisation 0 belongs to $\mathcal{A}_\rho$. In addition, the fact that $\mathcal{A}_\rho$ is closed is a consequence of Proposition 4.2.7prop.4.2.7.

2. As $Y \geq X$, $\mathbb{P}$-a.s. and $\rho(X) \leq 0$, we have that $\rho(Y) \leq \rho(X) \leq 0$, so Y belongs to $\mathcal{A}_\rho$.

3. We have :

$$\begin{aligned}\inf\{m \in \mathbb{R},\ X + m \in \mathcal{A}_\rho\} &= \inf\{m \in \mathbb{R},\ \rho(X+m) \leq 0\}\\ &= \inf\{m \in \mathbb{R},\ \rho(X) - m \leq 0\}\\ &= \inf\{m \in \mathbb{R},\ \rho(X) \leq m\}\\ &= \rho(X).\end{aligned}$$

4. Follows from the fact that for any X in $\mathcal{A}_\rho$, $\rho(X + \rho(X)) = \rho(X) - \rho(X) = 0$.

5. Assume $\mathcal{A}_\rho$ is convex. Let X, Y in $\mathcal{A}_\rho$ and λ in $[0,1]$. We have that $X + \rho(X)$ and $Y + \rho(Y)$ belong to $\mathcal{A}_\rho$. As $\mathcal{A}_\rho$ is assumed to be convex, we have that $\lambda(X + \rho(X)) + (1-\lambda)(Y + \rho(Y))$ belongs to $\mathcal{A}_\rho$. Hence :

$$0 \geq \rho(\lambda(X+\rho(X)) + (1-\lambda)(Y+\rho(Y))) = \rho(\lambda X + (1-\lambda)Y) - (\lambda\rho(X) + (1-\lambda)\rho(Y)),$$

which proves that ρ is convex.

Assume ρ is convex. Let X, Y in $\mathcal{A}_\rho$ and λ in $[0,1]$. We have :

$$\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y) \leq 0.$$

So $\lambda X + (1-\lambda)Y$ belongs to $\mathcal{A}_\rho$.

6. Left as an exercise.

7. Follows from 5. and 6..

□

Note that the acceptance set $\mathcal{A}_\rho$ completely define ρ and *vice-versa*.

4.2.3 Robust representation of risk measures

We have seen some examples of risk measures. The question we ask ourselves is : "Can we give the form of a risk measure ? " In other words, are there other risk measures than those we have cited ? For instance, the Daniell-Stone Theorem (which is related to the Riesz' representation theorem) states that for any real-valued continuous (in some sense) linear map ℓ on $\mathcal{X}$, there exists a probability measure $\mathbb{Q}$ such that : $\ell(X) = \mathbb{E}^{\mathbb{Q}}[X]$ for any X in $\mathcal{X}$. In other words, any linear map on $\mathcal{X}$ is an integral (or an expectation). So the question is : what happens if one replaces "linear" by "coherent" or "convex" for a risk measure ?

We need a notation. Let[6] :

$$\mathcal{M} := \{\mathbb{Q} \text{ probability measure on } (\Omega, \mathcal{F})\},$$

and

$$\mathcal{M}_f := \{\mathbb{Q} \text{ finitely additive probability measure on } (\Omega, \mathcal{F})\}.$$

Naturally, $\mathcal{M}_f \subset \mathcal{M}$ (but the converse is not true).

We start with the following proposition.

Proposition 4.2.10. *Let $\mathcal{Q} \subset \mathcal{M}$.*

1. *Set :*

$$\rho(X) := \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[-X], \quad \forall X \in \mathcal{X}.$$

   Then, ρ is a coherent risk measure.

2. *Let $\gamma : \mathcal{Q} \to \mathbb{R}$ such that $\sup_{\mathbb{Q}\in\mathcal{Q}} \gamma(Q) < +\infty$. Set*

$$\rho(X) := \sup_{\mathbb{Q}\in\mathcal{Q}} \left(\mathbb{E}^{\mathbb{Q}}[-X] - \gamma(\mathbb{Q})\right), \quad \forall X \in \mathcal{X}.$$

   Then, ρ is a convex risk measure.

Proof. 1. Left as an exercise.

2. For the second part, there is a trap to be avoid. Indeed, given X and Y in $\mathcal{X}$, the sup used for computing $\rho(X)$ may not be attained. But if it is, there is no reason that it is attained for the same element $\mathbb{Q}$ for $\rho(X)$ and $\rho(Y)$. So one needs to work a little bit. One way is to use the definition of supremum.

   Let X, Y in $\mathcal{X}$ and λ in $[0,1]$. Let $\mathbb{Q}$ in $\mathcal{Q}$. We have that :

$$\begin{aligned}
&\mathbb{E}^{\mathbb{Q}}[-(\lambda X + (1-\lambda)Y)] - \gamma(\mathbb{Q}) \\
&= \lambda \left(\mathbb{E}^{\mathbb{Q}}[-X] - \gamma(\mathbb{Q})\right) + (1-\lambda)\left(\mathbb{E}^{\mathbb{Q}}[-Y] - \gamma(\mathbb{Q})\right) \\
&\leq \lambda\rho(X) + (1-\lambda)\rho(Y).
\end{aligned}$$

   As the supremum is the least bigger element, we get that :

$$\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y).$$

□

What can be said on the converse statement. Unfortunately, there is a slight technical issue to be considered. More precisely, we have that :

Theorem 4.2.11 (Robust representation)**.** *Let ρ be a risk measure.*

[6] A finitely additive probability measure $\mathbb{Q}$ is a map $\mathbb{Q} : \mathcal{F} \to [0,1]$, such that $\mathbb{Q}[\Omega] = 1$ and for any $n \in \mathbb{N}^*$, and any $(A_i)_{i=1,\cdots,n} \subset \mathcal{F}$ pairwise disjoint, $\mathbb{Q}[\cup_{i=1}^n A_i] = \sum_{i=1}^n \mathbb{Q}[A_i]$.

1. *ρ is a coherent risk measure if and only if there exists a set $\mathcal{Q} \subset \mathcal{M}_f$ such that*

$$\rho(X) = \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[-X], \quad \forall X \in \mathcal{X};$$

2. *ρ is a convex risk measure if and only if there exists a set $\mathcal{Q} \subset \mathcal{M}_f$ such that*

$$\rho(X) = \sup_{\mathbb{Q}\in\mathcal{Q}} \left(\mathbb{E}^{\mathbb{Q}}[-X] - \alpha_{min}(\mathbb{Q})\right), \quad \forall X \in \mathcal{X},$$

with

$$\alpha_{min}(\mathbb{Q}) := \sup_{X\in\mathcal{A}_\rho} \mathbb{E}^{\mathbb{Q}}[-X],\ \mathbb{Q} \in \mathcal{M}.$$

4.3 Quantile functions and risk measures

4.3.1 Introduction

We have seen some examples in the previous section and the "general" form of a risk measure (coherent or convex). We aim in studying here a class of risk measures which take (*a priori*) a different form. Indeed, historically, measuring risk has been seen as estimating the probability that a financial claim leads to losses (that is the set $\{X < 0\}$ or $\{X < -a\}$, for some $a > 0$). In other words, given a level $\lambda \in (0,1)$ a position could be considered to be acceptable if :

$$\mathbb{P}\left[X < 0\right] \leq \lambda.$$

As we have seen, defining a set of acceptable claims is equivalent to define a risk measure and we define the risk measure $V@R_\lambda$ (called the value at Risk at level λ) as follows :

$$V@R_\lambda(X) := \inf\{m \in \mathbb{R},\ \mathbb{P}[m + X < 0] \leq \lambda\}, \quad X \in \mathcal{X}. \tag{4.3.1}$$

Our goal in this section is to study this (historical) risk measure and to provide related risk measures which will cope against the drawbacks of $V@R_\lambda$. To perform our analysis we will express more clearly $V@R_\lambda$ as a quantile. To this end we recall in the next section some elements on quantile functions.
Exercise 4 : Prove that $V@R_\lambda$ is a positive homogeneous risk measure.

Exercise 5 : Let $X \sim \mathcal{N}(\mu, \sigma^2)$ (note that X is not a bounded random variable so strictly speaking it is not a financial position in the way we have defined it). We have

$$V@R_\lambda(X) = -\mu + \sigma\Phi^{-1}(1-\lambda),$$

where Φ is the CDF of the $\mathcal{N}(0,1)$, that is, $\Phi(x) := \displaystyle\int_{-\infty}^{x} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy.$

$V@R_\lambda$ is one of the most used risk measure in practise (maybe due to this explicit form for Gaussian random variables, which can be extended to log-Normal random variables). However, it has several huge drawbacks.

Remark 4.3.1. *$V@R_\lambda$ is not a convex (so it does not encourage diversification). Another main drawback lies in the fact that $V@R_\lambda$ indicates the level from which one will face losses but it does not give information on the amplitude of these losses.*

4.3.2 Quantile functions and Value at Risk

Throughout this section, X denotes an element of $\mathcal{X}$ and F denotes its CDF ($F(x) := \mathbb{P}[X \leq x]$, $x \in \mathbb{R}$).

Definition 4.3.2 (Quantile). *Let λ in $(0,1)$. A λ-quantile of X is an real number $q_X(\lambda)$ such that :*

$$\mathbb{P}[X \leq q_X(\lambda)] \geq \lambda \quad \text{and} \quad \mathbb{P}[X < q_X(\lambda)] \leq \lambda.$$

Proposition 4.3.3. *Let λ in $(0,1)$. The set of λ-quantile of X is the interval $[q_X^-(\lambda), q_X^+(\lambda)]$ with :*

$$q_X^-(\lambda) := \sup\{z \in \mathbb{R},\ \mathbb{P}[X < z] < \lambda\} \quad q_X^+(\lambda) := \inf\{z \in \mathbb{R},\ \mathbb{P}[X \leq z] > \lambda\}.$$

Remarks 4.3.4. *Note that :*

1. $$q_X^-(\lambda) = \inf\{z \in \mathbb{R},\ \mathbb{P}[X \leq z] \geq \lambda\}, \quad q_X^+(\lambda) := \sup\{z \in \mathbb{R},\ \mathbb{P}[X < z] \leq \lambda\}.$$

2. *If F is increasing*[7] *("strictement croissante"), then $q_X^-(\lambda) = q_X^+(\lambda)$ and we recover the uniqueness of the quantile.*

3. *If F admits intervals of constancy (that is : there exists $[a,b]$ such that $F(x) = F(a)$ for any x in $[a,b]$) then $a = q_X^-(F(a))$ and $b = q_X^+(F(a))(= q_X^+(F(b)))$. This fact is annoying as for instance this happens for random variables : $X = N_T \sim \mathcal{P}(\lambda T)$ or for $X = R_T$ (as in 2. of Example 4.1.2prop.4.1.2).*

As mentioned in the previous section, $V@R_\lambda$ amounts to a quantile as proved below.

Proposition 4.3.5. *Let $\lambda \in (0,1)$. We have that :*

$$V@R_\lambda(X) = -q_X^+(\lambda) = q_{-X}^-(1-\lambda), \quad \forall X \in \mathcal{X}.$$

Proof. Let $\lambda \in (0,1)$ and X in $\mathcal{X}$. We have by definition that $V@R_\lambda(X) = \inf\{m \in \mathbb{R},\ \mathbb{P}[m + X < 0] \leq \lambda\}$. Hence

$$\begin{aligned}
V@R_\lambda(X) &= \inf\{m \in \mathbb{R},\ \mathbb{P}[m + X < 0] \leq \lambda\} \\
&= \inf\{m \in \mathbb{R},\ \mathbb{P}[X < -m] \leq \lambda\} \\
&= -\sup\{m \in \mathbb{R},\ \mathbb{P}[X < m] \leq \lambda\} \\
&= -q_X^+(\lambda) \\
&= \inf\{m \in \mathbb{R},\ 1 - \mathbb{P}[X \geq -m] \leq \lambda\} \\
&= \inf\{m \in \mathbb{R},\ \mathbb{P}[X \geq -m] \geq 1 - \lambda\} \\
&= \inf\{m \in \mathbb{R},\ \mathbb{P}[-X \leq m] \geq 1 - \lambda\} \\
&= q_{-X}^-(1-\lambda).
\end{aligned}$$

□

[7] even if it has jumps

4.3.3 Generalisations of $V@R_\lambda$

In this section, we do not provide proofs of the main results as they are rather technical.

According to Remark 4.3.1prop.4.3.1, $V@R_\lambda$ is not a convex risk measure (so it is not a coherent risk measure). In addition, it does not give information on the amount of losses. We introduce below two risk measures who address these drawbacks.

Definition 4.3.6 (Average value at risk)**.** *The Average Value at Risk at level $\lambda \in (0,1)$ ($AV@R_\lambda$) is defined as :*

$$AV@R_\lambda(X) := \frac{1}{\lambda}\int_0^\lambda V@R_\gamma(X)d\gamma, \quad X \in \mathcal{X}. \tag{4.3.2}$$

Proposition 4.3.7. *$AV@R_\lambda$ is a coherent risk measure (in particular it is convex).*

Remark 4.3.8. *Sometimes $AV@R_\lambda$ is called Conditional Value at Risk or Expected shortfall but as we will see later this is misleading.*

Note that :

$$AV@R_\lambda(X) = -\frac{1}{\lambda}\int_0^\lambda q_X^+(\gamma)d\gamma, \quad X \in \mathcal{X},$$

and formally

$$AV@R_1(X) = -\int_0^1 q_X^+(\gamma)d\gamma = \mathbb{E}[-X], \quad X \in \mathcal{X},$$
$$AV@R_0(X) = -\inf_{\omega\in\Omega} X(\omega), \quad X \in \mathcal{X}.$$

We have seen that any coherent risk measure can be represented as in Theorem 4.2.11Robust representationprop.4.2.11. In fact we have that :

Proposition 4.3.9. *Let λ in $(0,1)$. We have that :*

$$AV@R_\lambda(X) = \sup_{\mathbb{Q}\in\mathcal{Q}_\lambda} \mathbb{E}^{\mathbb{Q}}[-X], \quad \forall X \in \mathcal{X},$$

where $\mathcal{Q}_\lambda := \left\{\mathbb{Q} << \mathbb{P},\ \sup_{\omega\in\Omega}\left|\frac{d\mathbb{Q}}{d\mathbb{P}}(\omega)\right| \leq \frac{1}{\lambda}\right\}$.

Proposition 4.3.10 (Lemma 4.46 in [4])**.** *Let λ in $(0,1)$. We have that :*

$$AV@R_\lambda(X) = \lambda^{-1}\mathbb{E}[(q_X^+ - X)_+] - q_X^+, \quad \forall X \in \mathcal{X},$$

where $x_+ := \max(x,0)$.

We now define two other risk measures which are related to $AV@R_\lambda$.

Definition 4.3.11 (WCE_λ)**.** *We set for λ in $(0,1)$, WCE_λ (Worst Conditional Expectation et level λ) the map*

$$WCE_\lambda(X) := \sup_{A\in\mathcal{F},\,\mathbb{P}[A]>\lambda} \mathbb{E}[-X|A], \quad X \in \mathcal{X}.$$

Proposition 4.3.12. *For any λ in $(0,1)$, WCE_λ is a coherent risk measure.*

We now come back to Remark 4.3.8prop.4.3.8. We have that :

Proposition 4.3.13. *Let λ in $(0,1)$. For any X in $\mathcal{X}$, we have that :*

1.

$$\begin{aligned} AV@R_\lambda(X) &\geq WCE_\lambda(X) \\ &\geq \mathbb{E}[-X| - X \geq V@R_\lambda(X)] \\ &\geq V@R_\lambda(X). \end{aligned}$$

2. *If in addition $\mathbb{P}[X \leq q_X^+(X)] = \lambda$, then*

$$AV@R_\lambda(X) = WCE_\lambda(X) = \mathbb{E}[-X| - X \geq V@R_\lambda(X)].$$

The second item of the previous proposition, is quite important. Indeed, the property $\mathbb{P}[X \leq q_X^+(X)] = \lambda$ is true (for any λ in $(0,1)$) when the CDF of X does not admit levels of constancy. This is the case for instance for Gaussian random variables. This has lead to a misuse of the terminology (see Remark 4.3.8prop.4.3.8) but in Actuarial sciences making such a difference is quite important, as for instance for $X = R_T$ (see 2. of Example 4.1.2prop.4.1.2) it is not true for any λ that $\mathbb{P}[X \leq q_X^+(X)] = \lambda$ (as an exercice consider for instance a Poisson random variable). These quantities are interesting as they account for the expected losses beyond $V@R_\lambda$.

Bibliography

[1] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Math. Finance*, 9(3):203–228, 1999.

[2] S. Asmussen. *Ruin Theory*. John Wiley & Sons, Ltd, 2010.

[3] C. Dutang. *Cours d'actuariat 1 et 2*. Université Paris-Dauphine, 2020.

[4] H. Föllmer and A. Schied. *Stochastic finance*. Walter de Gruyter & Co., Berlin, extended edition, 2011. An introduction in discrete time.

[5] N. Gantert. Lecture notes : Actuarial risk theory summer term 2013, 2013.