Università di Pisa

Dipartimento di Matematica
Corso di Laurea Magistrale in Matematica

Tesi di Laurea Magistrale

Exploring the Optimal Reinsurance Problem: Convex Linearization, Optimal Transport and Case Studies

Relatori:
Prof. *Dario Trevisan*
Prof. *Hansjörg Albrecher*

Candidato:
David Vencato

Controrelatore:
Prof. *Mario Maurelli*

Anno Accademico 2023-2024

Contents

Introduction

Finding the optimal reinsurance contract is a classical problem in actuarial risk theory, which can be summarized as follows. An insurer holds a portfolio of n risks, represented by a random vector X defined on a probability space $(\Omega, \mathcal{F}, P)$ taking values in $\mathbb{R}^n_+$ and distributed according to a fixed probability measure μ. A reinsurer must decide what position to take relative to the insurer by selecting a portfolio represented by a random vector R such that $0 \leq R_i \leq X_i$ P-a.s. for all $i = 1, \ldots, n$. Indeed, the reinsurer cannot assume a negative risk nor compensate more than the insurer's exposure. Consequently, a reinsurance contract can be viewed as a probability measure $\eta \in \mathscr{P}(\mathbb{R}^n_+ \times \mathbb{R}^n_+)$ such that the marginal distribution on the first component coincides with the fixed probability μ, and η is concentrated on $\mathcal{A}_R := \{(x, y) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ \mid 0 \leq y_i \leq x_i \text{ for every } i = 1, \ldots, n\}$. Denoting the space of reinsurance contracts by $\mathcal{M}$, we endow it with the weak topology. In this framework, a reinsurance contract is not necessarily deterministic, i.e. $\eta = (\mathrm{id}, f)_{\#}\mu$ for some measurable function $f : \mathbb{R}^n_+ \to \mathbb{R}^n_+$. Instead, we allow for the possibility that the reinsurer, even after observing the realization x of the vector X, may decide what to do using an additional random mechanism. Given a functional $\mathcal{P}$ that quantifies the risk associated with each contract and a subset $\mathcal{S} \subseteq \mathcal{M}$ representing specific constraints agreed upon by the two parties, the problem of finding the optimal reinsurance contract reduces to finding $\eta^* \in \mathcal{S}$ such that:

$$\mathcal{P}(\eta^*) = \min_{\eta \in \mathcal{S}} \mathcal{P}(\eta). \tag{1}$$

The first result in Chapter 2 is to show that $\mathcal{M}$ is compact. Therefore, under the assumptions that $\mathcal{S}$ is closed and $\mathcal{P}$ is lower semi-continuous, the problem (1) always admits a solution.

Section 2.2 develops assuming that $\mathcal{P}$ is lower semi-continuous and $\mathcal{S} = \{\eta \in \mathcal{M} \mid \mathcal{G}(\eta) \leq 0\}$, where $\mathcal{G} = (g_1, \ldots, g_m) : \mathcal{M} \to \mathbb{R}^m$ is a lower semi-continuous function. We use the notation $d\mathcal{P}(\eta, \theta - \eta)$ to denote the directional derivative of $\mathcal{P}$ at η in the direction $\theta - \eta$

for all $\eta, \theta \in \mathcal{M}$ (in fact, we will only require the existence of the right-hand limit of the difference quotient). Similarly, we define $d\mathcal{G}(\eta, \theta - \eta)$. Given an optimal contract η^*, we aim to identify a convex set $\mathcal{C}$ containing η^* such that the maps $\eta \mapsto d\mathcal{P}(\eta^*, \eta - \eta^*)$ and $\eta \mapsto d\mathcal{G}(\eta^*, \eta - \eta^*)$, when restricted to $\mathcal{C}$, are linear over convex combinations. This allows us to characterize the measure η^*: as a consequence of the Hahn-Banach separation theorem, there must exist $r^* \in \mathbb{R}^+$ and $\lambda^* \in \mathbb{R}^m_+$ such that $\lambda^* \cdot \mathcal{G}(\eta^*) = 0$ and

$$r^* d\mathcal{P}(\eta^*; \eta - \eta^*) + \lambda^* \cdot d\mathcal{G}(\eta^*; \eta - \eta^*) \geq 0 \tag{2}$$

for every $\eta \in \mathcal{C}$.

Now, the goal is to transform this result on measures into a result concerning the support points of η^*. To this end, given $(x, y) \in \mathrm{Supp}(\eta^*)$, $t \in \mathbb{R}^n$, and $\varepsilon \geq 0$, we define specific perturbations $\eta_{x,y,t,\varepsilon}$ of the measure η^* such that

$$\eta_{x,y,t,\varepsilon}(A) := \eta^*(A) - \eta^*(A \cap B_\varepsilon(x, y)) + \eta^*((A - (0, t)) \cap B_\varepsilon(x, y))$$

for any measurable set $A \subseteq \mathbb{R}^n_+ \times \mathbb{R}^n_+$. For $-y < t < x - y$ and sufficiently small ε, we show that the perturbations belong to $\mathcal{M}$, suggesting that these new measures are valid candidates to be evaluated in (2). Indeed, under the additional assumption that $d\mathcal{P}(\eta^*, \cdot)$ and $d\mathcal{G}(\eta^*, \cdot)$ are integral operators on $\mathcal{C} - \eta^*$ with continuous kernels p_{η^*} and g_{η^*} respectively, for a point (x, y) in the support of η^*, the following must hold:

$$r^* p_{\eta^*}(x, y) + \lambda^* \cdot g_{\eta^*}(x, y) = \min_{t \in I} \left(r^* p_{\eta^*}(x, t) + \lambda^* \cdot g_{\eta^*}(x, t) \right) \tag{3}$$

where $I := \overline{J}$ and $J := \{t \in (0, x) \mid 0 \text{ is a limit point of } \{\varepsilon \in [0, \delta) \mid \eta_{x,y,t-y,\varepsilon} \in \mathcal{C}\}\}$.

In the same setting described above, a special case arises when $\mathcal{C} = \mathcal{M}$. In this scenario, the result obtained is much stronger than (3) and is proven directly using the property of analytic subsets of the product of Polish spaces, which guarantees the existence of universally measurable sections. Specifically, it is shown that the measure η^* is concentrated on the points (x, y) such that

$$y \in \arg\min_{t \in [0, x]} \left(r^* p_{\eta^*}(x, t) + \lambda^* \cdot g_{\eta^*}(x, t) \right).$$

The chapter continues with Section 2.3, where the goal is to view the reinsurance problem as an optimal transport problem. Therefore, we return to the initial assumptions on $\mathcal{P}$ and $\mathcal{S}$, which are only needed to ensure the existence of an optimal contract η^*. We assume that $\mathcal{P}(\eta^*) \leq \mathcal{P}((1-t)\eta^* + t\eta)$ for every $\eta \in \mathcal{S}$, $0 \leq t \leq 1$, and, moreover, that $d\mathcal{P}(\eta, \cdot)$ is an integral operator on $\mathcal{C} - \eta$ with a measurable kernel p_η for every $\eta \in \mathcal{S}$. In this framework, it is proven that

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} q_{\eta^*}(x, y)\, \eta^*(dx, dy) = \min_{\nu \in \pi_2(\mathcal{S})} \mathcal{C}(\mu, \nu), \tag{4}$$

where the function $q_{\eta^*}(x,y) := p_{\eta^*}(x,y)$ if $(x,y) \in \mathcal{A}_R$ and $q_{\eta^*} := +\infty$ otherwise, $\pi_2(\mathcal{S}) := \{(\pi_2)_\# \eta : \eta \in \mathcal{S}\}$ and

$$\mathcal{C}(\mu,\nu) := \min_{\eta \in \mathcal{S} \cap \Pi(\mu,\nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} q_{\eta^*}(x,y)\, \eta(dx,dy). \tag{5}$$

The minimization problem in (5) is an optimal transport problem constrained by $\mathcal{S}$. Therefore, we would like to get rid of the constraint in order to use results from optimal transport and solve (5). Thus, we consider the one-dimensional case which, in the Setting 2.3.4, allows us to show that the optimal contract must be deterministic of the form $(\mathrm{id}, g_\nu)_\# \mu$, where g_ν is the monotone map between μ and the probability ν that minimizes the right-hand side of (4). All the necessary theory to construct the monotone map and demonstrate its main properties is presented in Chapter 1.

Finally, Chapter 3 presents five case studies where the approaches described in Chapter 2 are applied to find the optimal reinsurance contract. In the first example, the objective is to minimize the expected value of the loss function

$$\mathcal{L} := \sum_{i=1}^n \big(X_i - R_i + (1+\beta_i)\mathbb{E}[R_i]\big),$$

assuming $0 < \beta_1 < \cdots < \beta_n$, under the constraint $\mathrm{Var}\big(\sum_{i=1}^n (X_i - R_i)\big) \le c$, for a constant $0 < c < \mathrm{Var}\big(\sum_{i=1}^n X_i\big)$. In this case, we can choose $\mathcal{C} = \mathcal{M}$, and under the assumption that μ is absolutely continuous with respect to the Lebesgue measure (it will be sufficient that μ does not concentrate too much mass on the singleton $\{0\}$), we are able to show that the optimal contract is deterministic and derive its general form depending on two parameters.

The second case maintains the same assumptions as the first but replaces the variance constraint with one based on a Value at Risk (VaR) measure (which will be properly defined). In this scenario, it is no longer possible to select $\mathcal{C} = \mathcal{M}$. Instead, given $v^* := \widehat{\mathrm{VaR}}_\alpha(\eta^*)$, we restrict $\mathcal{C}$ to contracts $\eta \in \mathcal{M}$ that belong to the same VaR_α-level set of η^*. By doing so, we eliminate g_{η^*} in (3). At this point, the idea is to divide $\mathcal{A}_R$ into a partition of four sets:

$$V_1 := \left\{ (x,y) \in \mathcal{A}_R \mid \sum_{i=1}^n (x_i - y_i) < v^*, \sum_{i=1}^n x_i \le v^* \right\};$$

$$V_2 := \left\{ (x,y) \in \mathcal{A}_R \mid \sum_{i=1}^n (x_i - y_i) < v^*, \sum_{i=1}^n x_i > v^* \right\};$$

$$D_2 := \left\{ (x,y) \in \mathcal{A}_R \mid \sum_{i=1}^n (x_i - y_i) > v^* \right\};$$

$$D_3 := \left\{ (x,y) \in \mathcal{A}_R \mid \sum_{i=1}^n (x_i - y_i) = v^* \right\}.$$

So, we use the result (3) to show the following steps: if $(x,y) \in \text{supp}(\eta^*) \cap (V_1 \cup D_2)$, then $y = 0$; in case $(x,y) \in (V_2 \cup (D_3 - \mathbb{R}^n_+ \times \{0\})) \cap \text{supp}(\eta^*)$, it holds that $(x,y) \in D_3$; for $(x,y) \in (D_3 - \mathbb{R}^n_+ \times \{0\}) \cap \text{supp}(\eta^*)$, we find that $y = y^*(x)$, where $y^*(x)$ will be the minimizer of p_{η^*} in a specific domain. Finally, it is shown that, in the regions of space where, for a given $(x,y) \in \text{supp}(\eta^*)$, both $y = 0$ and $y = y^*(x)$ are theoretically possible, they cannot both hold simultaneously, which ensures that the optimal solution is deterministic. We also describe the optimal contract as the parameter v^* varies.

The third and fourth examples, instead, demonstrate how the approach based on optimal transport theory can be effective when the constraints take a more general form. Specifically, in these two cases, the constraints respectively fix the expected value and the variance of the second marginal of any contract η. The results in Section 2.3 allows us to reformulate the reinsurance problems into an optimization problems that can be efficiently solved using analytical tools, enabling us to determine the complete form of the optimal contract.

Finally, the last case study involves a change of perspective: the second marginal of a contract η represents $X - R$ rather than R. The functional to minimize is the variance of the sum of the reinsured amounts, while the constraints fix all the marginals of the probability $(\pi_2)_\#\eta$ for each reinsurance contract η. Thus, this is a multi-marginal optimal transport problem, which we address in the specific case $n = 2$, having fixed the distributions of X_1 and X_2, and through a discretization process of the random variables. This approach allows us to reduce the problem to a linear optimization problem, which is solved using a specific Python library. The results show that introducing an additional degree of randomness is necessary to achieve optimality.

Notation

The following symbols are commonly used throughout this work:

- $\mathbb{N}$: the set of natural numbers.
- $\mathbb{R}^n$: the n-dimensional Euclidean space.
- $\overline{\mathbb{R}^n}$: $\mathbb{R}^n$ extended with $\pm\infty$.
- $\mathbb{R}^n_+$: the set $\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_i \geq 0 \text{ for all } i \in \{1, \dots, n\}\}$.
- $\mathcal{A}_R$: the set $\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid 0 \leq y_i \leq x_i,\ i = 1, \dots, n\}$.
- $[x, y]$: the closed box $[x_1, y_1] \times \cdots \times [x_n, y_n]$ with $x, y \in \mathbb{R}^n$.
- $x \leq y$: the partial ordering in $\mathbb{R}^n$ (total for $n = 1$) such that $x \leq y$ if and only if $x_i \leq y_i$ for every $\in \{1, \dots, n\}$.
- $x \cdot y$: the Euclidean scalar product in $\mathbb{R}^n$.
- id: the identity function.
- π_i: the projection of a product space onto the i-th component.
- $\mathbb{1}_A$: the indicator function of a set A.
- $\mathscr{B}(X)$: the Borel sets of the Polish space X.
- $\mathscr{P}(X)$: the set of probability measures on $\mathscr{B}(X)$.
- $T_\# \mu$: the image measure of μ under T.
- F_μ: the cumulative distribution function (CDF) of μ.

- $F^{[-1]}$: the pseudo-inverse of a nondecreasing and right-continuous function F.
- $\mathrm{spt}(\gamma)$: the support of the measure γ.
- $\nu \prec_1 \mu$: the first-order stochastic dominance, i.e. $F_\mu(x) \le F_\nu(x)$ for every $x \in \mathbb{R}^n$.
- $\Pi(\mu,\nu)$: the set of transport coupling between μ and ν.
- $\mu \otimes \nu$: the product measure of μ and ν.
- $\overline{\theta}$: is the integral $\int_{\mathbb{R}} x\, \theta(dx)$ where $\theta \in \mathscr{P}(\mathbb{R})$.
- $\mathscr{L}^n$: the n-dimensional Lebesgue measure.
- $\mu \llcorner A$: the measure μ restricted to the set A, i.e. $\mathbb{1}_A \cdot \mu$.
- $C_b(X)$: the space of continuous and bounded functions from X to $\mathbb{R}$.
- $\gamma_n \rightharpoonup \gamma$: the sequence of probabilities $\{\gamma_n\}_{n\in\mathbb{N}}$ converges to γ in the weak topology.
- $\mathcal{M}$: the space of reinsurance treaties.
- $\mathcal{P}$: the functional from $\mathcal{M}$ to $\mathbb{R}$ representing the risk carried by the treaties.
- $\mathcal{S}$: a subset of $\mathcal{M}$ representing the constraints.
- $\mathcal{I}$: the real vector space of signed measures on $\mathbb{R}^n_+ \times \mathbb{R}^n_+$.
- $df(u;h)$: the directional derivative whenever it is well defined.
- ∇: the gradient operator.
- S_k: the set of permutations of the numbers $\{1,\dots,k\}$.
- $\mathbb{E}[X]$: the expected value of the random variable X.
- $[X]$: the expected value of the random variable X.
- $\mathrm{Var}(X)$: the variance of the random variable X.
- $\mathrm{VaR}_\alpha(X)$: the Value at Risk measure of probability α of the random variable X.
- f_X: the continuous density function (or probability mass function) of the continuous (or discrete) random variable X.
- F_X: the cumulative distribution function of the random variable X.
- $\otimes$: Kronecker product between two matrices.
- $\mathbb{I}_M$: the identity matrix of dimension M.
- 1_M: the column vector of dimension M with all entries equal to one.

Chapter 1

Preliminaries on Optimal Transport

In this chapter, we introduce some fundamental notions of optimal transport. Our main goal is to prove the existence of the so-called monotone map (Theorem 1.3.6) and to highlight its role within the theory of one-dimensional optimal transport, as it provides the solution to a specific case of Monge problem (Theorem 1.3.8). So, the following discussion is a compilation of selected results from [7], [9], and [11], tailored to provide precisely the necessary background for a comprehensive and self-contained presentation.

1.1 Monge and Kantorovich problems

Definition 1.1.1: *Let $T : X \to Y$ be a Borel function and $\mu \in \mathscr{P}(X)$. The push-forward of μ by T or the image of μ under T is the probability measure over Y defined as:*

$$T_{\#}\mu(A) := \mu(T^{-1}(A))$$

for every $A \in \mathscr{B}(Y)$.

Definition 1.1.2: *Let X and Y be two Polish spaces, $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$, and $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ be a cost function. The Monge problem is:*

$$\inf_{T:X\to Y} \left\{ \int_X c(x, T(x))\, d\mu(x) : T_{\#}\mu = \nu \right\}. \tag{1.1}$$

Any Borel map T that attains the infimum in (1.1) *is called an optimal transport map.*

Definition 1.1.3: *Let X and Y be two Polish spaces, $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$, and $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ be a cost function. The Kantorovich problem is:*

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int_{X \times Y} c(x,y)\, d\gamma(x,y) \tag{1.2}$$

where $\Pi(\mu,\nu) = \{\gamma \in \mathscr{P}(X \times Y) : (\pi_1)_\# \gamma = \mu$ and $(\pi_2)_\# \gamma = \nu\}$ is the set of transport couplings between μ and ν. Any coupling γ that attains the infimum in (1.2) *is called an optimal coupling.*

Remark 1.1.4: In the definition of Kantorovich, the mass has greater "freedom of movement" compared to the Monge formulation. Specifically, in the Monge problem, a particle x is transported directly to a single point $T(x)$. In contrast, the Kantorovich formulation allows for particles to split and be transported to multiple locations.

Remark 1.1.5: If a solution of (1.2) is of the form $(\mathrm{id}, T)_\# \mu$, then it is called an optimal Monge coupling, and the map T is an optimal transport map for (1.1).

From now on, we consider $\Pi(\mu,\nu)$ endowed with respect to the weak topology, hence:

$$\gamma_n \rightharpoonup \gamma \iff \int_{X \times Y} g\, d\gamma_n \to \int_{X \times Y} g\, d\gamma, \quad \forall g \in C_b(X \times Y).$$

First of all, let us recall Prokhorov's theorem, which will be useful for proving the compactness of $\Pi(\mu,\nu)$.

Theorem 1.1.6: *([6], pp. 59-60). Let X be a Polish space. A family $\mathscr{A} \subset \mathscr{P}(Z)$ of probabilty measures on Z is relatively compact with respect to the weak topology if and only if it is tight, i.e. for every $\varepsilon > 0$ there exists $K_\varepsilon \subset Z$ compact such that*

$$\sup_{\gamma \in \mathscr{A}} \gamma(Z \setminus K_\varepsilon) \leq \varepsilon.$$

Proposition 1.1.7: *([11], p. 4). The set $\Pi(\mu,\nu)$ is compact.*

Proof. First of all, let us show that $\Pi(\mu,\nu)$ is relatively compact using Theorem 1.1.6 with $\mathscr{A} = \Pi(\mu,\nu)$ and $Z = X \times Y$. We fix $\varepsilon > 0$. Since $\{\mu\}$ and $\{\nu\}$ are relatively compact respectively in $\mathscr{P}(X)$ and $\mathscr{P}(Y)$, we can again apply Prokhorov's theorem to find $K_1 \subset X$ and $K_2 \subset Y$ such that $\mu(X \setminus K_1) \leq \frac{\varepsilon}{2}$ and $\nu(Y \setminus K_2) \leq \frac{\varepsilon}{2}$. Now, $K_1 \times K_2$ is compact in $X \times Y$ and for every $\gamma \in \Pi(\mu,\nu)$ we have:

$$\gamma(X \times Y \setminus K_1 \times K_2) \leq \gamma((X \setminus K_1) \times Y) + \gamma(X \times (Y \setminus K_2)) = \mu(X \setminus K_1) + \nu(Y \setminus K_2) = \varepsilon.$$

It remains to show that $\Pi(\mu,\nu)$ is closed: let $\{\gamma_n\}_{n\in\mathbb{N}} \subset \Pi(\mu,\nu)$ be such that $\gamma_n \rightharpoonup \gamma$, we need to verify that $\gamma \in \Pi(\mu,\nu)$. Using the commutativity between the pushforward and the limit in weak convergence, we obtain:

$$(\pi_1)_\#\gamma = (\pi_1)_\# \left(\lim_{n\to\infty} \gamma_n\right) = \lim_{n\to\infty} ((\pi_1)_\#\gamma_n) = \lim_{n\to\infty} \mu = \mu$$

and similarly for $(\pi_2)_\#\gamma$. □

Theorem 1.1.8: *([11], p. 4). Assume that the cost function c is lower semi-continuous and bounded from below. Then there exists an optimal coupling for the Kantorovich problem* (1.2)*.*

Proof. We only need to verify that the functional $J : \Pi(\mu,\nu) \to \mathbb{R} \cup \{+\infty\}$ such that $J(\gamma) := \int_{X\times Y} c(x,y)\, d\gamma(x,y)$ is lower semi-continuous with respect to the weak convergence of probabilities; the thesis then follows directly from Weierstrass's theorem. Consider a sequence of continuous and bounded function $\{c_n\}_{n\in\mathbb{N}}$ converging increasingly to c and set $J_n(\gamma) := \int_{X\times Y} c_n(x,y)\, d\gamma(x,y)$. By monotone convergence, we have $J(\gamma) = \sup_{n\in\mathbb{N}} J_n(\gamma)$. Also, every J_n is continuous with respect to the weak convergence, hence J is lower semi-continuous as it is a supremum of continuous functionals. □

1.2 Optimality and c-CM set

Definition 1.2.1: *Let X and Y be two Polish spaces and $c : X \times Y \to \mathbb{R}\cup\{+\infty\}$. We say that $\Gamma \subset X \times Y$ is c-cyclically monotone (abbreviated as c-CM) if for every $k \in \mathbb{N}$, every permutation $\sigma \in S_k$ and every finite family of points $(x_1,y_1),\dots,(x_k,y_k) \in \Gamma$ it holds:*

$$\sum_{i=1}^{k} c(x_i,y_i) \le \sum_{i=1}^{k} c(x_i,y_{\sigma(i)})$$

Theorem 1.2.2: *([7], p. 30). Suppose γ is an optimal coupling for* (1.2) *with cost function c continuous. Then* $\mathrm{spt}(\gamma)$ *is a c-CM set.*

Proof. Suppose by contradiction that there exist $k \in \mathbb{N}$, a permutation $\sigma \in S_k$, and $(x_1,y_1),\dots,(x_k,y_k) \in \mathrm{spt}(\gamma)$ such that

$$\sum_{i=1}^{k} c(x_i,y_i) > \sum_{i=1}^{k} c(x_i,y_{\sigma(i)}).$$

Fix $\varepsilon < \frac{1}{2k}\left(\sum_{i=1}^{k} c(x_i,y_i) - \sum_{i=1}^{k} c(x_i,y_{\sigma(i)})\right)$; given the continuity of c for each $i \in \{1,\dots,k\}$, there exist two positive reals r_i and s_i such that for every $(x,y) \in B_{r_i}(x_i) \times B_{r_i}(y_i)$, we have $c(x,y) > c(x_i,y_i) - \varepsilon$, and for every $(x,y) \in B_{s_i}(x_i) \times B_{s_i}(y_{\sigma(i)})$, we have

$c(x,y) < c(x_i, y_{\sigma(i)}) + \varepsilon$. We define $r := \min\{r_1, s_1, \dots, r_k, s_k\}$.
Now, we set $V_i := B_r(x_i) \times B_r(y_i)$ and $W_i := B_r(x_i) \times B_r(y_{\sigma(i)})$; since $(x_i, y_i) \in \operatorname{spt}(\gamma)$, it follows that $\gamma(V_i) > 0$ for every i; thus, the measure $\gamma_i := (\gamma \llcorner V_i) / \gamma(V_i)$ is well-defined, and consequently, $\mu_i := (\pi_1)_\# \gamma_i$ and $\nu_i := (\pi_2)_\# \gamma_i$. Furthermore, for every i consider a measure $\tilde{\gamma}_i \in \Pi(\mu_i, \nu_{\sigma(i)})$ (for example $\tilde{\gamma}_i = \mu_i \otimes \nu_{\sigma(i)}$). Let $\varepsilon_0 < \frac{1}{k} \min_i \gamma(V_i)$, we define the measure:

$$\tilde{\gamma} := \gamma - \varepsilon_0 \sum_{i=1}^k \gamma_i + \varepsilon_0 \sum_{i=1}^k \tilde{\gamma}_i.$$

We will find a contradiction by showing that $\tilde{\gamma} \in \Pi(\mu, \nu)$ and

$$\int_{\mathbb{R}\times\mathbb{R}} c(x,y)\, d\tilde{\gamma}(x,y) < \int_{\mathbb{R}\times\mathbb{R}} c(x,y)\, d\gamma(x,y),$$

which implies that γ cannot be optimal. Let us proceed systematically:

- *$\tilde{\gamma}$ is a positive measure*: for every i we have

$$\varepsilon_0 \gamma_i = \frac{\varepsilon_0}{\gamma(V_i)} \gamma \llcorner V_i < \frac{1}{k} (\gamma \llcorner V_i) \le \frac{1}{k} \gamma,$$

  therefore $\varepsilon_0 \sum_{i=1}^k \gamma_i < \gamma$ and it is sufficient to ensure the positivity of $\tilde{\gamma}$.

- *$\tilde{\gamma}$ is a probability*: since γ_i is a probability measure concentrated on V_i and $\tilde{\gamma}_i$ is concentrated on W_i, it follows immediately that $\tilde{\gamma}(X \times Y) = 1$.

- *$\tilde{\gamma}$ is a coupling from μ to ν:* let us compute the marginals of $\tilde{\gamma}$:

$$(\pi_1)_\# \tilde{\gamma} = (\pi_1)_\# \gamma - \varepsilon_0 \sum_{i=1}^k (\pi_1)_\# \gamma_i + \varepsilon_0 \sum_{i=1}^k (\pi_1)_\# \tilde{\gamma}_i = \mu - \varepsilon_0 \sum_{i=1}^k \mu_i + \varepsilon_0 \sum_{i=1}^k \mu_i = \mu;$$

$$(\pi_2)_\# \tilde{\gamma} = (\pi_2)_\# \gamma - \varepsilon_0 \sum_{i=1}^k (\pi_2)_\# \gamma_i + \varepsilon_0 \sum_{i=1}^k (\pi_2)_\# \tilde{\gamma}_i = \nu - \varepsilon_0 \sum_{i=1}^k \nu_i + \varepsilon_0 \sum_{i=1}^k \nu_{\sigma(i)} = \nu.$$

- *Computation of the integral:*

$$\begin{aligned}
\int_{\mathbb{R}\times\mathbb{R}} c(x,y)\, d\gamma(x,y) - \int_{\mathbb{R}\times\mathbb{R}} c(x,y)\, d\tilde{\gamma}(x,y) &= \varepsilon_0 \sum_{i=1}^k \int_{V_i} c(x,y)\, d\gamma_i(x,y) \\
&\quad - \varepsilon_0 \sum_{i=1}^k \int_{W_i} c(x,y)\, d\tilde{\gamma}_i(x,y) \\
&\ge \varepsilon_0 \sum_{i=1}^k (c(x_i, y_i) - \varepsilon) - \varepsilon_0 \sum_{i=1}^k (c(x_i, y_{\sigma(i)}) + \varepsilon) \\
&= \varepsilon_0 \left(\sum_{i=1}^k c(x_i, y_i) - \sum_{i=1}^k c(x_i, y_{\sigma(i)}) - 2k\varepsilon \right) > 0.
\end{aligned}$$

$\square$

1.3 One-Dimensional Case

Definition 1.3.1: *Given a nondecreasing and right-continuous function $F : \mathbb{R} \to [0,1]$, its pseudo-inverse is the function $F^{[-1]} : [0,1] \to \overline{\mathbb{R}}$ such that*

$$F^{[-1]}(x) := \inf\{t \in \mathbb{R} \mid F(t) \geq x\}.$$

Remark 1.3.2: It follows immediately from the definition and the right continuity of F that:

$$F^{[-1]}(x) \leq a \iff F(a) \geq x. \tag{1.3}$$

Proposition 1.3.3: *([7], p. 60). If $\mu \in \mathscr{P}(\mathbb{R})$, then $(F_\mu^{[-1]})_\#(\mathscr{L}^1 \llcorner [0,1]) = \mu$. Moreover, if $\mu, \nu \in \mathscr{P}(\mathbb{R})$ and we set $\eta := (F_\mu^{[-1]}, F_\nu^{[-1]})_\#(\mathscr{L}^1 \llcorner [0,1])$, then $\eta \in \Pi(\mu,\nu)$ and $\eta((-\infty,a] \times (-\infty,b]) = \min\{F_\mu(a), F_\nu(b)\}$.*

Proof. Since the cumulative distribution function characterizes a measure, to verify the first part of the statement, it is sufficient to check that the CDF's of the image measure and of μ coincide. Indeed, using (1.3) it holds that:

$$\begin{aligned} F_{(F_\mu^{[-1]})_\#(\mathscr{L}^1 \llcorner [0,1])}(a) &= \mathscr{L}^1(\{x \in [0,1] \mid F_\mu^{[-1]}(x) \leq a\}) = \mathscr{L}^1(\{x \in [0,1] \mid F_\mu(a) \geq x\}) \\ &= \mathscr{L}^1([0, F_\mu(a)]) = F_\mu(a). \end{aligned}$$

This first part also proves that $\eta \in \Pi(\mu,\nu)$. To conclude, let us compute similarly to before:

$$\begin{aligned} \eta((-\infty,a] \times (-\infty,b]) &= \mathscr{L}^1(\{x \in [0,1] \mid F_\mu^{[-1]}(x) \leq a,\ F_\nu^{[-1]}(x) \leq b\}) \\ &= \mathscr{L}^1(\{x \in [0,1] \mid F_\mu^{[-1]}(a) \geq x,\ F_\nu^{[-1]}(b) \geq x\}) = \min\{F_\mu(a), F_\nu(b)\}. \end{aligned}$$

□

Definition 1.3.4: *We set $\gamma_{\text{mon}} := (F_\mu^{[-1]}, F_\nu^{[-1]})_\#(\mathscr{L}^1 \llcorner [0,1])$ and we call it the comonotone transport coupling between μ and ν. It is also known as the Brenier map.*

Now, let us try to define a nondecreasing transport map between μ and ν when μ is atomless. However, we first need a technical lemma.

Lemma 1.3.5: *([7], p. 61). If $\mu \in \mathscr{P}(\mathbb{R})$ is atomless, then $(F_\mu)_\#\mu = \mathscr{L}^1 \llcorner [0,1]$ and for every $l \in [0,1]$, the set $\{x \mid F_\mu(x) = l\}$ is μ-negligible.*

Proof. Since μ is atomless, the function F_μ is continuous. In particular, for every $a \in (0,1)$ the set $\{x \mid F_\mu(x) \leq a\}$ is a closed set of the form $(-\infty, x_a]$ such that $F_\mu(x_a) = a$. Thanks to this, for every $a \in (0,1)$, we obtain:

$$F_{(F_\mu)_\#\mu}(a) = \mu(\{x \mid F_\mu(x) \leq a\}) = F_\mu(x_a) = a = \mathscr{L}^1 \llcorner [0,1]([0,a]) = F_{\mathscr{L}^1 \llcorner [0,1]}(a).$$

Hence, together with $F_{(F_\mu)_\#\mu}(0) = F_{\mathscr{L}^1 \llcorner [0,1]}(0) = 0$ and $F_{(F_\mu)_\#\mu}(1) = F_{\mathscr{L}^1 \llcorner [0,1]}(1) = 1$, we have $(F_\mu)_\#\mu = \mathscr{L}^1 \llcorner [0,1]$.
If, for the sake of contradiction, there exists $l \in [0,1]$ such that $\mu(\{x \mid F_\mu(x) = l\}) > 0$, we would have:

$$0 < \mu(\{x \mid F_\mu(x) = l\}) = (F_\mu)_\#\mu(\{l\}) = \mathscr{L}^1 \llcorner [0,1](l) = 0.$$

$\square$

Theorem 1.3.6: *([7], p. 61). Let $\mu, \nu \in \mathscr{P}(\mathbb{R})$ and suppose that μ is atomless. Then, there exists a unique nondecreasing map $T_{\text{mon}} : \mathbb{R} \to \mathbb{R}$ such that $(T_{\text{mon}})_\#\mu = \nu$.*

Proof. We define:

$$T_{\text{mon}}(x) := F_\nu^{[-1]}(F_\mu(x)).$$

Notice that this quantity is well defined if and only if $F_\mu(x) \in (0,1)$; since $\{x \mid F_\mu(x) = 0\}$ and $\{x \mid F_\mu(x) = 1\}$ are negligible thanks to Lemma 1.3.5, the map T_{mon} is well-defined μ-a.e. It follows directly from the definition of the pseudo-inverse that T_{mon} is nondecreasing. Furthermore, using Proposition 1.3.3 and Lemma 1.3.5, we obtain:

$$(T_{\text{mon}})_\#\mu = (F_\nu^{[-1]} \circ F_\mu)_\#\mu = (F_\nu^{[-1]})_\#(F_\mu)_\#\mu = (F_\nu^{[-1]})_\#(\mathscr{L}^1 \llcorner [0,1]) = \nu.$$

Now, we have to prove the uniqueness. Consider a nondecreasing function T such that $T_\#\mu = \nu$. Since the monotonicity of T, it holds $(-\infty, x] \subset T^{-1}((-\infty, T(x)])$ and so:

$$F_\mu(x) = \mu((-\infty, x]) \leq \mu(T^{-1}((-\infty, T(x)])) = \nu((-\infty, T(x)]) = F_\nu(T(x)).$$

This implies that $T(x) \geq F_\nu^{[-1]}(F_\mu(x))$. Suppose that the inequality is strict, by definition there exists an $\varepsilon_0 > 0$ such that for every $\varepsilon \in (0, \varepsilon_0)$ it holds $F_\nu(T(x) - \varepsilon) \geq F_\mu(x)$. Furthermore, for every $\varepsilon > 0$ it is valid that $T^{-1}((-\infty, T(x) - \varepsilon)) \subset (-\infty, x)$ and so from the same previous chain we get $F_\nu(T(x) - \varepsilon) \leq F_\mu(x)$. Hence, for every x where $T(x) > F_\nu^{[-1]}(F_\mu(x))$, x is in an interval where F_ν is constant. We know that the intervals where F_ν is constant is a countable quantity and let $\{l_i\}_{i \in \mathbb{N}}$ be the values of F_ν on these intervals. From the above, it follows that $\{x \mid T(x) > F_\nu^{[-1]}(F_\mu(x))\} \subset \bigcup_{i \in \mathbb{N}} \{x \mid F_\mu(x) = l_i\}$, and thus $T(x) = F_\nu^{[-1]}(F_\mu(x))$ μ-a.e. $\square$

Now let us see an important property of γ_{mon} and T_{mon} which will be fundamental for finding the solution to the optimal transport problem for a significant class of cost functions.

Lemma 1.3.7: *([7], p. 62). Let $\mu, \nu \in \mathscr{P}(\mathbb{R})$ and $\gamma \in \Pi(\mu, \nu)$. If γ satisfies the property:*

$$\forall (x_1, y_1), (x_2, y_2) \in \operatorname{spt}(\gamma),\ x_1 < x_2 \implies y_1 \leq y_2, \tag{1.4}$$

then $\gamma = \gamma_{\text{mon}}$*. In particular, there is a unique* $\gamma \in \Pi(\mu, \nu)$ *satisfying* (1.4)*. Furthermore, if* μ *is atomless, we have* $\gamma = (\text{id}, T_{\text{mon}})_{\#}\mu$*.*

Proof. Every measure η on $\mathbb{R}^2$ is characterized by the values $\eta((-\infty, a] \times (-\infty, b])$ with $(a, b) \in \mathbb{R}^2$. Thus, thanks to Proposition 1.3.3 we have to verify that:

$$\gamma((-\infty, a] \times (-\infty, b]) = \min\{F_\mu(a), F_\nu(b)\}$$

for every $(a, b) \in \mathbb{R}^2$. Let us define the sets $A := \{(-\infty, a] \times (b, +\infty)\}$ and $B := \{(a, +\infty) \times (-\infty, b]\}$. In particular, $\gamma(A)$ and $\gamma(B)$ cannot both be strictly positive at the same time; otherwise, it would contradict the assumption in (1.4). Hence, we get the following equality:

$$\gamma((-\infty, a] \times (-\infty, b]) = \min\{\gamma(((-\infty, a] \times (-\infty, b]) \cup A), \gamma(((-\infty, a] \times (-\infty, b]) \cup B)\}.$$

However, for the two terms in the argument of the minimum, we have:

$$\gamma(((-\infty, a] \times (-\infty, b]) \cup A) = \gamma((-\infty, a] \times \mathbb{R}) = \gamma(\pi_1^{-1}((-\infty, a])) = \mu((-\infty, a]) = F_\mu(a)$$

and

$$\gamma(((-\infty, a] \times (-\infty, b]) \cup B) = \gamma(\mathbb{R} \times (-\infty, b]) = \gamma(\pi_2^{-1}((-\infty, b])) = \nu((-\infty, b]) = F_\nu(b).$$

This concludes the first part of the proof. Now suppose μ is atomless. First, we want to define a function $T(x)$ μ-a.e. such that $\gamma = (\text{id}, T)_{\#}\mu$. It is clear that we have:

$$\begin{aligned}\mathbb{R} =& \{x \mid \text{card}(\{y \mid (x, y) \in \text{spt}(\gamma)\}) = 0\} \cup \{x \mid \text{card}(\{y \mid (x, y) \in \text{spt}(\gamma)\}) = 1\} \\ &\cup \{x \mid \text{card}(\{y \mid (x, y) \in \text{spt}(\gamma)\}) > 1\}.\end{aligned}$$

We will now analyze the three cases:

- $\{x \mid \text{card}(\{y \mid (x, y) \in \text{spt}(\gamma)\}) = 0\}$: this set is μ-negligible, and thus we do not need to do anything;
- $\{x \mid \text{card}(\{y \mid (x, y) \in \text{spt}(\gamma)\}) = 1\}$: we define $T(x) = y$ where y is the only element such that $(x, y) \in \text{spt}(\gamma)$;
- $\{x \mid \text{card}(\{y \mid (x, y) \in \text{spt}(\gamma)\}) > 1\}$: we fix x in this set and we define I_x as the minimal interval I such that $\text{spt}(\gamma) \cap (\{x\} \times \mathbb{R}) \subset \{x\} \times I$. In particular, the map that sends x to the interior of I_x is a bijection. Indeed, the map is well-defined since the interior of each I_x is non-empty because I_x is not a singleton. It is injective because if $x_1 \neq x_2$, then the interiors of I_{x_1} and I_{x_2} are disjoint, as a direct consequence of (1.4). But a family of pairwise disjoint open intervals in $\mathbb{R}$ is countable, and therefore the cardinality of $\{x \mid \text{card}(\{y \mid (x, y) \in \text{spt}(\gamma)\}) > 1\}$ is countable as well. Since μ is atomless, this set is μ-negligible.

Hence, we have defined μ-a.e. the map we were looking for. Using (1.4), it follows that the map is non-decreasing, and thus, by Theorem 1.3.6, we must have $T = T_{\text{mon}}$. $\square$

Theorem 1.3.8: *([7], p. 63). Let $h : \mathbb{R} \to \mathbb{R}$ be a strictly convex and bounded from below function and $\mu, \nu \in \mathscr{P}(\mathbb{R})$. Suppose that* (1.2) *has a finite value and the cost function is of the form $c(x, y) = h(y - x)$ (or $h(x - y)$). Then, the Kantorovich problem has a unique solution which is given by γ_{mon}. Furthermore, if μ is atomless, this optimal coupling is induced by the Monge map T_{mon}.*

Proof. We can use Theorem 1.1.8 to ensure the existence of an optimal γ. Thanks to Lemma 1.3.7, we only need to verify that property (1.4) holds.
Let $(x_1, y_1), (x_2, y_2) \in \operatorname{spt}(\gamma)$ such that $x_1 < x_2$, and assume by contradiction that $y_1 > y_2$. Define $a := y_1 - x_1$, $b := y_2 - x_2$, and $\delta := x_2 - x_1 > 0$. By Theorem 1.2.2, we know that the support of an optimal γ is a c-CM set Γ when c is continuous. In particular:

$$h(y_1 - x_1) + h(y_2 - x_2) \leq h(y_2 - x_1) + h(y_1 - x_2),$$

that is

$$h(a) + h(b) \leq h(b + \delta) + h(a - \delta).$$

By definition, $b > a$, while $b + \delta$ and $a - \delta$ lie in the segment $[b, a]$. In particular, it is easy to veirfy that:

$$b + \delta = (1 - t)b + ta, \quad a - \delta = tb + (1 - t)a, \quad \text{for } t = \frac{\delta}{a - b} \in (0, 1).$$

Finally, the strict convexity of the function h gives us a contradiction:

$$h(a) + h(b) \leq h(b + \delta) + h(a - \delta) < (1 - t)h(b) + th(a) + th(b) + (1 - t)h(a) = h(a) + h(b).$$

$\square$

Chapter 2

The Reinsurance Problem

In this chapter, we address the problem of optimal reinsurance through two different approaches, both presented in [10]. For a broader perspective on the topic, we refer to the monograph [8].

We begin with Section 2.1, where we define the space of admissible reinsurance treaties $\mathcal{M}$ and formulate the corresponding constrained optimization problem (2.1). We then establish an existence result for this problem in Proposition 2.1.5. In Section 2.2, by applying a linearization process over convex sets, we characterize the optimal reinsurance contract (Proposition 2.2.2) and identify the points in its support (Proposition 2.2.5). Finally, in Section 2.3, we highlight the duality that emerges when interpreting the reinsurance problem as an optimal transport problem. The main result of this section is Proposition 2.3.6, which ensures that, in the one-dimensional case and under conditions covering a wide variety of problems, the optimal contract is deterministic.

Regarding the notation used in this chapter, we consider $\mathbb{R}^n$ endowed with the partial ordering $\leq$ such that $x \leq y$ if and only if $x_i \leq y_i$ for all $i = 1, \ldots, n$. Also, the notation $x \cdot y$ indicates the Euclidean scalar product in $\mathbb{R}^n$. In this context, for every $x, y \in \mathbb{R}^n$, the set $[x, y]$ represents the closed box $[x_1, y_1] \times \cdots \times [x_n, y_n]$ and we define the set $\mathbb{R}^n_+ := \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_i \geq 0 \text{ for all } i \in \{1, \ldots, n\}\}$.

2.1 Formulation and Existence

Given a probability space $(\Omega, \mathcal{F}, P)$, let $X = (X_1, \ldots, X_n)$ be a nonnegative random vector representing a portfolio consisting of n risks associated with one or more insurers, intended for partial reinsurance. We assume that each X_i has a finite first moment. The distribution of X_i is denoted by $\mu_i := (X_i)_\# P$, while the joint distribution of X is given by $\mu := X_\# P$.

Definition 2.1.1: *A random reinsurance treaty for the portfolio X with joint distribution μ is a probability measure $\eta \in \mathscr{P}(\mathbb{R}^n_+ \times \mathbb{R}^n_+)$ such that:*

(i) the first marginal of η is equal to μ, i.e. $(\pi_1)_\# \eta = \mu$;

(ii) the measure η is concentrated on the set $\mathcal{A}_R := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid 0 \le y_i \le x_i,\ i = 1, \dots, n\}$, i.e. $\eta(\mathcal{A}_R) = 1$.

We denote by $\mathcal{M}$ the space of reinsurance treaties and endow it with the weak topology.

Remark 2.1.2: We recall that for probability measures it also holds:

$$\eta_n \rightharpoonup \eta \iff \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} g \, d\eta_n \to \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} g \, d\eta, \quad \forall g \in C_c(\mathbb{R}^n_+ \times \mathbb{R}^n_+).$$

Remark 2.1.3: Let R denote the vector with distribution $(\pi_2)_\# \eta$, which represents the portfolio of risks held by the reinsurer. Condition (ii) in Definition 2.1.1 ensures that $R, X - R \in \mathbb{R}^n_+$, meaning that the reinsurer will not assume a negative amount of risk, but cannot cover more than the portion of the original loss borne by the insurer.
The definition helps us explain, also from a practical point of view, how contracts are decided. Suppose that $X = x$ occurs; we can use η to calculate the conditional distribution of R given $X = x$ and thus choose a value $R = r$. A special case takes place when the support of η is a subset of the graph of a function $f : \mathbb{R}^n \to \mathbb{R}^n$; in this case, we evidently refer to non-random treaties since $r = f(x)$.

Definition 2.1.4: *Consider a functional $\mathcal{P} : \mathcal{M} \to \mathbb{R}$ that represents the risk carried by the treaty, and a subset $\mathcal{S} \subseteq \mathcal{M}$ (which may represent specific requirements or conditions). Then, the optimal reinsurance problem is to find $\eta^* \in \mathcal{S}$ such that*

$$\mathcal{P}(\eta^*) = \min_{\eta \in \mathcal{S}} \mathcal{P}(\eta). \tag{2.1}$$

Any η^* satisfying this condition is called an optimal reinsurance contract.

Proposition 2.1.5: *([5], pp. 449-450). If $\mathcal{P}$ is lower semi-continuous and $\mathcal{S}$ is closed, then the optimal reinsurance problem* (2.1) *always has a solution.*

Proof. The crux of the proof lies in showing that $\mathcal{M}$ is compact. Indeed, $\mathcal{S}$ would be a closed subset of a compact set, and thus itself compact. At this point, we can directly apply the Weierstrass theorem, which guarantees the existence of an optimal reinsurance contract.

First, it is well known that $\mathcal{M}$ with the weak topology is metrizable; therefore, showing its sequential compactness is equivalent to proving its compactness. Let us fix the classical notation $\langle \eta, g\rangle := \int_{\mathbb{R}_+^n\times\mathbb{R}_+^n} g\, d\eta$ for all $\eta\in\mathcal{M}$ and for all $g\in C_c(\mathbb{R}_+^n\times\mathbb{R}_+^n)$. Let $\{\eta_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}$ and $\{g_n\}_{n\in\mathbb{N}}$ a countable dense subset of $C_c(\mathbb{R}_+^n\times\mathbb{R}_+^n)$. Now, for every $g\in C_c(\mathbb{R}_+^n\times\mathbb{R}_+^n)$, we have that:

$$|\langle \eta, g\rangle| \le \int_{\mathbb{R}_+^n\times\mathbb{R}_+^n} |g|\, d\eta \le \|g\|_\infty \cdot \eta(\mathbb{R}_+^n\times\mathbb{R}_+^n) = \|g\|_\infty < +\infty.$$

Then, the sequence of real numbers $\{\langle \eta_n, g_1\rangle\}_{n\in\mathbb{N}}$ is bounded and thus admits a convergent subsequence $\left\{\left\langle \eta_{n_k^1}, g_1\right\rangle\right\}_{k\in\mathbb{N}}$. Similarly, for each $i\in\mathbb{N}$, we can define inductively a subsequence of $\left\{\eta_{n_k^i}\right\}_{k\in\mathbb{N}}$, which we denote by $\left\{\eta_{n_k^{i+1}}\right\}_{k\in\mathbb{N}}$, such that $\left\{\left\langle \eta_{n_k^{i+1}}, g_{i+1}\right\rangle\right\}_{k\in\mathbb{N}}$ converges. Hence, by applying a diagonal argument, we obtain the sequence $\left\{\eta_{n_k^k}\right\}_{k\in\mathbb{N}}$ that is a subsequence of $\{\eta_n\}_{n\in\mathbb{N}}$ such that $\left\{\left\langle \eta_{n_k^k}, g_i\right\rangle\right\}_{k\in\mathbb{N}}$ converges for all $i\in\mathbb{N}$. The next step is to show that the sequence $\left\{\left\langle \eta_{n_k^k}, g\right\rangle\right\}_{k\in\mathbb{N}}$ is Cauchy for all $g\in C_c(\mathbb{R}_+^n\times\mathbb{R}_+^n)$. Fix $\varepsilon>0$. Let $j\in\mathbb{N}$ be an index such that $\|g-g_j\|_\infty < \frac{\varepsilon}{4}$. Since every convergent sequence of real numbers is Cauchy, there exists M such that for all $m,k>M$ we have $\left|\left\langle \eta_{n_m^m}, g_j\right\rangle - \left\langle \eta_{n_k^k}, g_j\right\rangle\right| < \frac{\varepsilon}{2}$. So, we obtain:

$$\begin{aligned}\left|\left\langle \eta_{n_m^m}, g\right\rangle - \left\langle \eta_{n_k^k}, g\right\rangle\right| &\le \left|\left\langle \eta_{n_m^m}, g_j\right\rangle - \left\langle \eta_{n_k^k}, g_j\right\rangle\right| + \left|\left\langle \eta_{n_m^m}, g-g_j\right\rangle\right| + \left|\left\langle \eta_{n_k^k}, g-g_j\right\rangle\right| \\ &\le \left|\left\langle \eta_{n_m^m}, g_i\right\rangle - \left\langle \eta_{n_k^k}, g_i\right\rangle\right| + 2\|g-g_i\|_\infty < \frac{\varepsilon}{2} + 2\cdot\frac{\varepsilon}{4} = \varepsilon.\end{aligned}$$

For the completeness of $\mathbb{R}$, the sequence $\left\{\left\langle \eta_{n_k^k}, g\right\rangle\right\}_{k\in\mathbb{N}}$ is convergent, and thus the functional $g\mapsto \lim_{k\to\infty}\left\langle \eta_{n_k^k}, g\right\rangle$ is well defined. It follows directly from the definition that the functional is positive, linear, and bounded, hence continuous. We can therefore apply the Riesz representation theorem, which yields a positive measure η such that:

$$\langle \eta, g\rangle = \lim_{k\to\infty}\left\langle \eta_{n_k^k}, g\right\rangle, \quad \forall g\in C_c(\mathbb{R}_+^n\times\mathbb{R}_+^n).$$

By definition, $\eta_{n_k^k}\rightharpoonup\eta$ so, in order to conclude, we just need to check that $\eta\in\mathcal{M}$. Using the commutativity between the pushforward and the limit in weak convergence, we obtain:

$$(\pi_1)_\#\eta = (\pi_1)_\#\Big(\lim_{k\to\infty}\eta_{n_k^k}\Big) = \lim_{k\to\infty}\Big((\pi_1)_\#\eta_{n_k^k}\Big) = \lim_{k\to\infty}\mu = \mu.$$

Furthermore, since $\mathcal{A}_R$ is a closed set, the weak convergence implies $\limsup_{k\to\infty}\eta_{n_k^k}(\mathcal{A}_R)\le\eta(\mathcal{A}_R)$, hence $\eta(\mathcal{A}_R)=1$. □

From now on, we will always assume that $\mathcal{P}$ is lower semi-continuous and $\mathcal{S}$ is closed, ensuring the existence of an optimal treaty.

2.2 Linearization and Optimal Treaties

Definition 2.2.1: *Let Ω_c be a convex subset of a vector space U and let V be a normed space. Given a function $f : \Omega_c \to V$, we define its directional derivative at $u \in \Omega_c$ in the direction $h \in U$ as:*

$$df(u; h) = \lim_{t \to 0^+} \frac{f(u + th) - f(u)}{t},$$

provided the expression is well-defined and the limit exists.

We shall see how we can give meaning to the definition in our case:

- we set $U = \mathcal{I}$, where $\mathcal{I}$ is the set of signed measures on $\mathbb{R}^n_+ \times \mathbb{R}^n_+$ endowed with the usual real vector space structure:

  $$(\mu_1 + \mu_2)(A) = \mu_1(A) + \mu_2(A) \qquad \text{and} \qquad (c\mu)(A) = c \cdot \mu(A)$$

  for every $A \in \mathscr{B}(\mathbb{R}^n_+ \times \mathbb{R}^n_+)$ and for every $c \in \mathbb{R}$. In this way, it follows directly from the definitions that a convex combination of two probability measures concentrated on $\mathcal{A}_R$ is itself a probability measure concentrated on $\mathcal{A}_R$, and by using the fact that the pushforward of measures is a linear operator, we obtain for all $\eta_1, \eta_2 \in \mathcal{M}$:

  $$(\pi_1)_\# \left(t\eta_1 + (1 - t)\eta_2\right) = t(\pi_1)_\#\eta_1 + (1 - t)(\pi_1)_\#\eta_2 = t\mu + (1 - t)\mu = \mu.$$

  Hence, we can pick $\mathcal{M}$ as the convex subset of the vector space $\mathcal{I}$.

- We choose $V = \mathbb{R}^m$ with the Euclidean norm.

- Let us consider f as the functional $\mathcal{P}$. In general, if $\eta, \theta \in \mathcal{M}$, then $\eta + t\theta$ is not a probability measure for $t > 0$, and thus the notation $d\mathcal{P}(\eta; \theta)$ loses its meaning. However, due to the convexity of $\mathcal{M}$, it makes sense to write $(1 - t)\eta + t\theta = \eta + t(\theta - \eta)$ for $0 \leq t \leq 1$, and therefore the notation $d\mathcal{P}(\eta; \theta - \eta)$ is also meaningful. Remembering that η and θ are two vectors, then $\theta - \eta$ lies in the affine space centered at η and so $d\mathcal{P}(\eta; \theta - \eta)$ captures the infinitesimal change in $\mathcal{P}$ as we move from η towards θ.

Having clarified these aspects, we can state and prove the following proposition.

Proposition 2.2.2: *([10], p. 10). Assume that the functional $\mathcal{P}$ is lower semicontinuous. Let the set of constraints $\mathcal{S}$ be defined as $\mathcal{S} = \{\eta \in \mathcal{M} \mid \mathcal{G}(\eta) \leq 0\}$ for a lower semicontinuous function $\mathcal{G} = (g_1, ..., g_m) : \mathcal{M} \to \mathbb{R}^m$, and let η^* represent an optimal reinsurance contract. Consider the set $\mathcal{D} \subset \mathcal{M}$ given by*

$$\mathcal{D} = \{\eta \in \mathcal{M} \mid d\mathcal{P}(\eta^*; \eta - \eta^*) \text{ and } d\mathcal{G}(\eta^*; \eta - \eta^*) \text{ exist}\}.$$

Assume there exists a subset $\mathcal{C}$ *of* $\mathcal{D}$ *such that* $\eta^* \in \mathcal{C}$*,* $\mathcal{C}$ *is convex, and for any* $\eta_1, \eta_2 \in \mathcal{C}$*, it holds that*

$$d\mathcal{P}\left(\eta^*; (1-t)\eta_1 + t\eta_2 - \eta^*\right) = (1-t)d\mathcal{P}(\eta^*; \eta_1 - \eta^*) + td\mathcal{P}(\eta^*; \eta_2 - \eta^*) \tag{2.2}$$

for $0 \leq t \leq 1$*, with a similar statement for* $d\mathcal{G}$*.*
Under these conditions, there exist $r^* \in \mathbb{R}_+$ *and* $\lambda^* \in \mathbb{R}^m_+$ *such that* $\lambda^* \cdot \mathcal{G}(\eta^*) = 0$ *and*

$$r^* d\mathcal{P}(\eta^*; \eta - \eta^*) + \lambda^* \cdot d\mathcal{G}(\eta^*; \eta - \eta^*) \geq 0 \tag{2.3}$$

for every $\eta \in \mathcal{C}$*. Also, if* $\mathcal{G}$ *is constant on* $\mathcal{C}$ *or if there exists* $\eta \in \mathcal{C}$ *such that* $\mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta - \eta^*) < 0$*, then* r^* *in* (2.3) *is positive.*

Proof. First of all, we want to use Proposition 2.1.5 to ensure the existence of η^*. This can be done because the condition $g_i \leq 0$ defines a closed set since g_i is lower semicontinuous for all $i = 1, \ldots, m$.
Now, we define the sets:

$$A = \{(r, \lambda) \in \mathbb{R} \times \mathbb{R}^m \mid r \geq d\mathcal{P}(\eta^*; \eta - \eta^*),\ \lambda \geq \mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta - \eta^*) \text{ for some } \eta \in \mathcal{C}\};$$
$$B = \{(r, \lambda) \in \mathbb{R} \times \mathbb{R}^m \mid r < 0,\ \lambda < 0\}.$$

Suppose for contradiction that sets A and B intersect at a point, meaning there exist $r < 0$, $\lambda < 0$, and $\eta \in \mathcal{C}$ such that:

$$d\mathcal{P}(\eta^*; \eta - \eta^*) \leq r < 0$$

and

$$\mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta - \eta^*) \leq \lambda < 0.$$

We find a contradiction by showing that η^* is not optimal, as $\eta^* + s(\eta - \eta^*)$ represents a better reinsurance treaty for some $0 < s < 1$. First, we note that $d\mathcal{P}(\eta^*; \eta^* - \eta^*) = 0$ and so $\eta^* + s(\eta - \eta^*) \neq \eta^*$ for all $s \neq 0$. Now, using the definition of directional derivative and contradiction hypothesis there exists a real number $x_1 < 0$ such that:

$$\lim_{t \to 0^+} \frac{\mathcal{P}(\eta^* + t(\eta - \eta^*)) - \mathcal{P}(\eta^*)}{t} = x_1,$$

so by definition of limit we can find a $0 < \delta_1 < 1$ such that for all $t \in (0, \delta_1)$ it holds:

$$\frac{\mathcal{P}(\eta^* + t(\eta - \eta^*)) - \mathcal{P}(\eta^*)}{t} \in (2x_1, 0). \tag{2.4}$$

With the same argument, for all $i = 1, \ldots m$, we can find a real number $x_2^i < 0$ and $0 < \delta_2^i < 1$ such that for all $t \in (0, \delta_2^i)$:

$$g_i(\eta^*) + \frac{g_i(\eta^* + t(\eta - \eta^*)) - g_i(\eta^*)}{t} \in (2x_2^i, 0). \tag{2.5}$$

Since $\eta^* \in \mathcal{S}$, we have $g_i(\eta^*) \leq 0$, and thus $g_i(\eta^*)(t-1) \geq 0$ for all $t \in (0,1)$; so (2.5) implies:

$$g_i(\eta^* + t(\eta - \eta^*)) \leq 0. \tag{2.6}$$

Now, let $s := \min\{\delta_1, \delta_2^1, \dots, \delta_2^m\}$, we obtain from (2.4) that $\mathcal{P}(\eta^* + s(\eta - \eta^*)) < \mathcal{P}(\eta^*)$ and from (2.6) that $\mathcal{G}(\eta^* + s(\eta - \eta^*)) \leq 0$. So, η^* can not be an optimal reinsurance treaty.
Furthermore, we want to show that A is convex. Consider two points $(r_1, \lambda_1), (r_2, \lambda_2) \in A$, for any $t \in [0,1]$ we need to verify that for any $t \in [0,1]$ the convex combination $(1-t)(r_1, \lambda_1) + t(r_2, \lambda_2)$ is also in A. By definition of A, there exist $\eta_1, \eta_2 \in \mathcal{C}$ such that:

$$r_1 \geq d\mathcal{P}(\eta^*; \eta_1 - \eta^*), \quad \lambda_1 \geq \mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta_1 - \eta^*),$$

$$r_2 \geq d\mathcal{P}(\eta^*; \eta_2 - \eta^*), \quad \lambda_2 \geq \mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta_2 - \eta^*).$$

Let $\eta_t := (1-t)\eta_1 + t\eta_2$. Since $\mathcal{C}$ is convex $\eta_t \in \mathcal{C}$, so using the previous relations and (2.2) we obtain:

$$(1-t)r_1 + tr_2 \geq (1-t)d\mathcal{P}(\eta^*; \eta_1 - \eta^*) + td\mathcal{P}(\eta^*; \eta_2 - \eta^*) = d\mathcal{P}(\eta^*; \eta_t - \eta^*);$$

$$(1-t)\lambda_1 + t\lambda_2 \geq (1-t)(\mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta_1 - \eta^*)) + t(\mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta_2 - \eta^*)) = \mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta_t - \eta^*).$$

Moreover, it is evident that A is non-empty (using η^* we obtain $(0, \mathcal{G}(\eta^*)) \in A$) and that B is a non-empty, open, convex set. We can therefore apply the Hahn-Banach separation theorem: there exists a non-vanishing vector $(r^*, \lambda^*) \in \mathbb{R} \times \mathbb{R}^m$ and a scalar $\alpha \in \mathbb{R}$ such that the hyperplane $(r^*, \lambda^*) \cdot x = \alpha$ satisfies the separation conditions

$$r^* r + \lambda^* \cdot \lambda \geq \alpha \quad \forall (r, \lambda) \in A,$$

$$r^* r + \lambda^* \cdot \lambda < \alpha \quad \forall (r, \lambda) \in B.$$

Let us notice that $\alpha = 0$: the point $(0,0) \in A$, in fact $\eta^* \in C$, $\mathcal{P}(\eta^*; \eta^* - \eta^*) = 0$, $\mathcal{G}(\eta^*; \eta^* - \eta^*) = 0$ and $\mathcal{G}(\eta^*) \leq 0$. Hence, $\alpha \leq 0$. On the other hand, by definition of B, $\alpha > -\varepsilon$ for all $\varepsilon > 0$ and so $\alpha \geq 0$.
We have obtained the inequality $r^* r + \lambda^* \cdot \lambda < 0$ for every $r < 0$ and $\lambda < 0$ but this can be true if and only if $r^* \in \mathbb{R}^+$ and $\lambda^* \in \mathbb{R}^m_+$. Furthermore, by definition of A, for every $\eta \in \mathcal{C}$ the point $(d\mathcal{P}(\eta^*; \eta - \eta^*), \mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta - \eta^*)) \in A$, hence:

$$r^* d\mathcal{P}(\eta^*; \eta - \eta^*) + \lambda^* \cdot (\mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta - \eta^*)) \geq 0. \tag{2.7}$$

In particular, by choosing $\eta = \eta^*$, we obtain $\lambda^* \cdot \mathcal{G}(\eta^*) \geq 0$. However, from the conditions $\mathcal{G}(\eta^*) \leq 0$ and $\lambda^* \geq 0$, it also follows that $\lambda^* \cdot \mathcal{G}(\eta^*) \leq 0$. Therefore, we conclude that $\lambda^* \cdot \mathcal{G}(\eta^*) = 0$, which proves (2.3) replacing it in (2.7).
Let us now prove the second part of the statement.

In the first case, we assume $\mathcal{G}$ is constant on $\mathcal{C}$, and we proceed with a general reasoning. For every $\eta \in \mathcal{C}$ we have $d\mathcal{G}(\eta^*, \eta - \eta^*) = 0$ so we just need to prove:

$$r^* d\mathcal{P}(\eta^*; \eta - \eta^*) \geq 0 \tag{2.8}$$

for some $r^* > 0$. Since $\eta^* \in \mathcal{C}$ and $\mathcal{C}$ is convex it holds for all $t \in [0,1]$:

$$\mathcal{G}((1-t)\eta^* + t\eta) = \mathcal{G}(\eta^*) \leq 0$$

meaning that $(1-t)\eta^* + t\eta \in \mathcal{S}$. Hence, from the optimality of η^*, it must be that $\mathcal{P}(\eta^*) \leq \mathcal{P}((1-t)\eta^* + t\eta)$ for all $t \in [0,1]$. Thus, by definition

$$d\mathcal{P}(\eta^*; \eta - \eta^*) = \lim_{t \to 0^+} \frac{\mathcal{P}((1-t)\eta^* + t\eta) - \mathcal{P}(\eta^*)}{t} \geq 0$$

for every $\eta \in \mathcal{C}$. So (2.8) holds for all $r^* \geq 0$.
In the second case, for hypothesis there exists $\eta \in \mathcal{C}$ such that $\mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta - \eta^*) < 0$. Referring to the proof of (2.3), suppose by contradiction $r^* = 0$. Since (r^*, λ^*) is non-vanishing it must be that $\lambda^* \cdot (\mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta - \eta^*)) < 0$ which contradicts (2.7). □

The result we just obtained concerns measures; now we would like to transform it into a result that pertains to the points in the support of η^*. The first natural thought is to use the Dirac measures $\delta_{(x,y)}$. However, since we will frequently require that μ is absolutely continuous with respect to the Lebesgue measure, there is no hope that $\delta_{(x,y)} \in \mathcal{M}$ for some $(x,y) \in \mathcal{A}$. Therefore, we need to define different measures that we can use in (2.3). The next two lemmas go in this direction and are preparatory to our goal.

Lemma 2.2.3: *Let $(x,y) \in \mathcal{A}_R$ with $(x,y) \neq (0,0)$, and let $t \in \mathbb{R}^n$. We set:*

$$\delta = \begin{cases} \min\left\{d\left((x,y), \partial\mathcal{A}_R\right), d\left((x,y), \partial\mathcal{A}_R - (0,t)\right)\right\} & \text{if } 0 < y < x, \\ d\left((x,y), \partial\mathcal{A}_R - (0,t)\right) & \text{if } (x,y) \in \partial\mathcal{A}_R, \end{cases}$$

where in a vector space V, the notation $W - v$ refers to the translation of the subspace W by the vector v.
Furthermore, let $\varepsilon \in \mathbb{R}$ and suppose $-y < t < x - y$ and $0 \leq \varepsilon < \delta$. Then, we have $\delta > 0$ and the measures $\eta_{x,y,t,\varepsilon}$, defined as

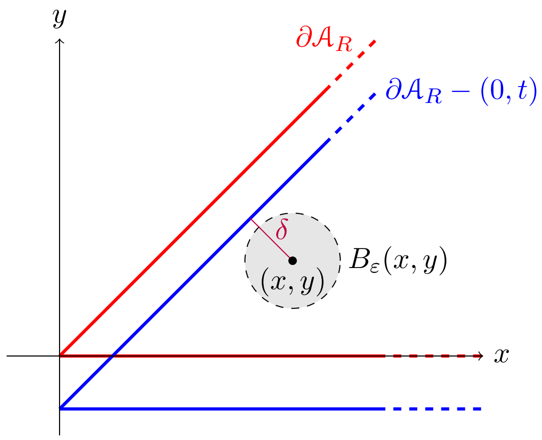
$$\eta_{x,y,t,\varepsilon}(A) = \eta^*(A) - \eta^*(A \cap B_\varepsilon(x,y)) + \eta^*((A - (0,t)) \cap B_\varepsilon(x,y))$$

for any measurable set $A \subseteq \mathbb{R}^n_+ \times \mathbb{R}^n_+$, belong to the space $\mathcal{M}$.

Proof. Let us start with the case $0 < y < x$ and recall that, for a closed set C, we have $d(C,x) = 0$ if and only if $x \in C$. Since $0 < y < x$, it follows that $(x,y) \notin \partial\mathcal{A}_R$. Moreover,

$(x, y) \in \mathcal{A}_R - (0, t)$ if and only if $(x, y + t) \in \mathcal{A}_R$, which is equivalent to $-y \leq t \leq x - y$. Hence $(x, y) \notin \partial\mathcal{A}_R - (0, t)$ and $\delta > 0$.
From the definition of δ, we have that $B_\varepsilon(x, y) \subset \mathcal{A}_R \cap (\mathcal{A}_R - (0, t)) \subset \mathbb{R}^n_+ \times \mathbb{R}^n_+$. Thus, the measure is well-defined. It is evidently positive, and furthermore, the following hold:

- $\eta_{x,y,t,\varepsilon}(\mathcal{A}_R) = \eta^*(\mathcal{A}_R) - \eta^*(B_\varepsilon(x, y)) + \eta^*(B_\varepsilon(x, y)) = \eta^*(\mathcal{A}_R) = 1$;
- $(\pi_1)_\# \eta_{x,y,t,\varepsilon}(X) = \eta_{x,y,t,\varepsilon}(X \times \mathbb{R}^n_+) = \eta^*(X \times \mathbb{R}^n_+) - \eta^*((X \times \mathbb{R}^n_+) \cap B_\varepsilon(x, y))$
$+ \eta^*((X \times (\mathbb{R}^n_+ - t)) \cap B_\varepsilon(x, y)) = (\pi_1)_\# \eta^*(X) = \mu(X)$
for all $X \in \mathscr{B}(\mathbb{R}^n_+)$, where we have used that $(\mathbb{R}^n_+ - t) \cap \pi_2(B_\varepsilon(x, y)) = \mathbb{R}^n_+ \cap \pi_2(B_\varepsilon(x, y))$ since $\pi_2(B_\varepsilon(x, y)) \subset (\mathbb{R}^n_+ \cap (\mathbb{R}^n_+ - t))$.



Now, let us consider the case where $(x, y) \in \partial\mathcal{A}_R$. With the same argument as before, we have that $\delta > 0$. We observe that, in the case where $y = 0$, the set $(A - (0, t)) \cap B_\varepsilon(x, y)$ is not necessarily contained in $\mathbb{R}^n_+ \times \mathbb{R}^n_+$; however, with an abuse of notation, we can always think of probability measures on $\mathbb{R}^n_+ \times \mathbb{R}^n_+$ extended over all $\mathbb{R}^n \times \mathbb{R}^n$ by assigning them zero outside the first set. Hence, we have:

- $\eta_{x,y,t,\varepsilon}(\mathcal{A}_R) = \eta^*(\mathcal{A}_R) - \eta^*(B_\varepsilon(x, y) \cap (\mathbb{R}^n_+ \times \mathbb{R}^n_+)) + \eta^*(B_\varepsilon(x, y) \cap (\mathbb{R}^n_+ \times \mathbb{R}^n_+)) = \eta^*(\mathcal{A}_R) = 1$
where we have used that $\eta^*(\mathcal{A}_R \cap B_\varepsilon(x, y)) = \eta^*(B_\varepsilon(x, y) \cap (\mathbb{R}^n_+ \times \mathbb{R}^n_+))$ and $B_\varepsilon(x, y) \subset (\mathcal{A}_R - (0, t))$.
- $(\pi_1)_\# \eta_{x,y,t,\varepsilon}(X) = \eta_{x,y,t,\varepsilon}(X \times \mathbb{R}^n_+) = \eta^*(X \times \mathbb{R}^n_+) - \eta^*((X \times \mathbb{R}^n_+) \cap B_\varepsilon(x, y))$
$+ \eta^*((X \times (\mathbb{R}^n_+ - t)) \cap B_\varepsilon(x, y) \cap (\mathbb{R}^n_+ \times \mathbb{R}^n_+)) = (\pi_1)_\# \eta^*(X) = \mu(X)$
for all $X \in \mathscr{B}(\mathbb{R}^n_+)$, since $\pi_2(B_\varepsilon(x, y)) \subset (\mathbb{R}^n_+ - t)$.

□

Lemma 2.2.4: *Let $(x, y) \in Supp(\eta^*)$ with $(x, y) \neq (0, 0)$. Then, for every $-y < t < x-y$ and for any continuous function $f : \mathbb{R}^n_+ \times \mathbb{R}^n_+ \to \mathbb{R}^n$, it holds:*

$$\lim_{\varepsilon \to 0} \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} f \, d\eta_\varepsilon = f(x, y+t) - f(x, y), \quad \eta_\varepsilon := \frac{\eta_{x,y,t,\varepsilon} - \eta^*}{\eta^*(B_\varepsilon(x, y))}. \tag{2.9}$$

Proof. By definition of $\eta_{x,y,t,\varepsilon}$, for every measurable set $A \subseteq \mathbb{R}^n_+ \times \mathbb{R}^n_+$ we can write $\eta_\varepsilon(A) = \theta_\varepsilon(A) - \pi_\varepsilon(A)$ with

$$\theta_\varepsilon(A) := \frac{\eta^* \left((A - (0, t)) \cap B_\varepsilon(x, y) \right)}{\eta^* \left(B_\varepsilon(x, y) \right)} \quad \text{and} \quad \pi_\varepsilon(A) := \frac{\eta^* \left(A \cap B_\varepsilon(x, y) \right)}{\eta^* \left(B_\varepsilon(x, y) \right)},$$

that are well-defined since $(x, y) \in \mathrm{Supp}(\eta^*)$ hence $\eta^* \left(B_\varepsilon(x, y) \right) > 0$. Now, we notice that $\theta_\varepsilon = \phi_\# \pi_\varepsilon$ where $\phi : \mathbb{R}^n_+ \times \mathbb{R}^n_+ \to \mathbb{R}^n_+ \times \mathbb{R}^n$ such that $\phi(u, v) = (u, v+t)$. Hence, we obtain:

$$\begin{aligned}
\int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} f(u, v) \, d\eta_\varepsilon(u, v) &= \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} f(u, v) \, d\theta_\varepsilon(u, v) - \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} f(u, v) \, d\pi_\varepsilon(u, v) \\
&= \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} f(u, v) \, d(\phi_\# \pi_\varepsilon)(u, v) - \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} f(u, v) \, d\pi_\varepsilon(u, v) \\
&= \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} f(u, v+t) \, d\pi_\varepsilon(u, v) - \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} f(u, v) \, d\pi_\varepsilon(u, v) \\
&= \frac{1}{\eta^* \left(B_\varepsilon(x, y) \right)} \left(\int_{B_\varepsilon(x,y)} f(u, v+t) \, d\eta^*(u, v) - \int_{B_\varepsilon(x,y)} f(u, v) \, d\eta^*(u, v) \right).
\end{aligned}$$

At this point, it only remains to apply the Lebesgue differentiation theorem to terms, obtaining:

$$\lim_{\varepsilon \to 0} \frac{1}{\eta^* \left(B_\varepsilon(x, y) \right)} \left(\int_{B_\varepsilon(x,y)} f(u, v+t) \, d\eta^*(u, v) \right) = f(x, y+t);$$

$$\lim_{\varepsilon \to 0} \frac{1}{\eta^* \left(B_\varepsilon(x, y) \right)} \left(\int_{B_\varepsilon(x,y)} f(u, v) \, d\eta^*(u, v) \right) = f(x, y).$$

□

Proposition 2.2.5: *([10], p. 11). Let us remain in the same setup of Proposition 2.2.2. Assume that $d\mathcal{P}$ and $d\mathcal{G}$ are integral operators on $\mathcal{C} - \eta^*$ with continuous kernels, i.e. there exist continuous functions $p_{\eta^*} : \mathbb{R}^n_+ \times \mathbb{R}^n_+ \to \mathbb{R}$ and $g_{\eta^*} : \mathbb{R}^n_+ \times \mathbb{R}^n_+ \to \mathbb{R}^m$ such that for every $\eta \in \mathcal{C} - \eta^*$ hold*

$$d\mathcal{P}(\eta^*, \eta) = \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} p_{\eta^*}(u, v) \, \eta(du, dv) \quad \textit{and} \quad d\mathcal{G}(\eta^*, \eta) = \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} g_{\eta^*}(u, v) \, \eta(du, dv).$$

Then, for every $(x, y) \in Supp(\eta^*) \setminus (0, 0)$, it holds:

$$r^* p_{\eta^*}(x, y) + \lambda^* \cdot g_{\eta^*}(x, y) = \min_{t \in I} \left(r^* p_{\eta^*}(x, t) + \lambda^* \cdot g_{\eta^*}(x, t) \right)$$

where $I := \overline{J}$, $J := \{t \in (0, x) \mid 0$ is a limit point of $\{\varepsilon \in [0, \delta) \mid \eta_{x,y,t-y,\varepsilon} \in \mathcal{C}\}\}$ and $y \in J$.

Proof. Since $\eta^* \in \mathcal{C}$, we have $\{\varepsilon \in [0,\delta) \mid \eta_{x,y,0,\varepsilon} \in \mathcal{C}\} = [0,\delta)$ and so $y \in I$. Now, choosing $t \in J$, for every $\varepsilon \in [0,\delta)$ such that $\eta_{x,y,t-y,\varepsilon} \in \mathcal{C}$, we can rewrite (2.3) as:

$$r^* \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} p_{\eta^*}(u,v)\,(\eta_{x,y,t-y,\varepsilon} - \eta^*)(du,dv) + \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} g_{\eta^*}(u,v)\,(\eta_{x,y,t-y,\varepsilon} - \eta^*)(du,dv) \geq 0.$$

So, using linearity:

$$\int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} (r^* p_{\eta^*}(u,v) + g_{\eta^*}(u,v))\; d\left(\frac{\eta_{x,y,t-y,\varepsilon} - \eta^*}{\eta^*(B_\varepsilon(x,y))}\right) \geq 0.$$

At this point, it is sufficient to use the Lemma 2.2.4 where we use $t-y$ instead of y (the hypotheses coincide since $-y < t-y < x-y$ if and only if $0 < t < x$). So, we obtain:

$$r^*\,(p_{\eta^*}(x,t) - p_{\eta^*}(x,y)) + \lambda^* \cdot (g_{\eta^*}(x,t) - g_{\eta^*}(x,y)) \geq 0.$$

Hence, we have proved the statement for J. To conclude, we only need to note that for every element $t \in I$, there exists a sequence $\{t_n\}_{n\in\mathbb{N}} \subset J$ such that $t_n \to t$. Given the continuity of p_{η^*} and g_{η^*}, the result follows by taking the limit as $n \to \infty$. $\square$

Remark 2.2.6: Let us note that under the assumptions of the Proposition 2.2.5, we must know that the point $(x,y) \in \mathrm{Supp}(\eta^*)$, which is, however, the contract we are trying to determine. Therefore, what has been said should be understood as providing necessary conditions for solving the problem, when we are able to determine the sets I. Also, the functions p_{η^*} and g_{η^*} depends on η^*, so we are trying to solve a minimization problem that depends on the optimal contract we have chosen. Nevertheless, in most cases, they will depend on η^* only on a finite number of parameters that can be considered fixed at the outset.

Now, let us see what happens when the set $\mathcal{C}$ becomes the whole of $\mathcal{M}$. But first, we recall a result concerning the existence of sections for analytic sets.

Proposition 2.2.7: *([4], p. 199). Every analytic subset A of the product of Polish spaces X and Y admits a section s that is universally measurable, meaning it is measurable with respect to the completion of the Borel σ-algebra relative to any probability measure.*

Proposition 2.2.8: *([10], p. 14). We remain in the setting of the Proposition 2.2.5 with the addition of two assumptions: $\mathcal{C} = \mathcal{D} = \mathcal{M}$, and the function $m : \mathbb{R}^n_+ \to \mathbb{R}$ defined by*

$$m(x) = \inf_{y\in[0,x]} (r^* p_{\eta^*}(x,y) + \lambda^* \cdot g_{\eta^*}(x,y))$$

is measurable. Then, if we set $K := \{(x,y) \in \mathcal{A}_R \mid y \in \arg\min_{t\in[0,x]} (r^* p_{\eta^*}(x,t) + \lambda^* \cdot g_{\eta^*}(x,t))\}$, we have $\eta^*(K) = 1$.

Proof. During the proof, in order to avoid cumbersome notation, we will use m to denote both the function from the statement and the function $\tilde{m} : \mathbb{R}^n_+ \times \mathbb{R}^n_+ \to \mathbb{R}$ given by $\tilde{m}(x, y) = m(x)$. We define the function $h : \mathbb{R}^n_+ \times \mathbb{R}^n_+ \to \mathbb{R}$ such that $h(x, y) = r^* p_{\eta^*}(x, y) + \lambda^* \cdot g_{\eta^*}(x, y)$. Thus, we can rewrite the set K as:

$$K = \{(x, y) \in \mathcal{A}_R \mid h(x, y) = m(x, y)\}.$$

Given the measurability of m, the set K is measurable, and the thesis is well-defined. Since $\eta^*(\mathcal{A}_R) = 1$, we have to show that $h(x, y) = m(x, y)$ η^*-a.e. By definition, $h(x, y) \geq m(x, y)$, so we just need the inequality:

$$\int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} h(x, y)\eta^*(dx, dy) \leq \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} m(x, y)\eta^*(dx, dy).$$

By choosing K as the analytic set, the Proposition 2.2.7 gives us that there exists a function $s : \mathbb{R}^n_+ \to \mathbb{R}^n_+ \times \mathbb{R}^n_+$ such that $s(x) \in K$, $\pi_1 \circ s = \mathrm{id}_{\mathbb{R}^n_+}$, and s is measurable with respect to the μ-completion of $\mathbb{R}^n_+$. This last property allows us to well-define the pushforward $\gamma := s_\# \mu$. Not only that, this probability is in $\mathcal{M}$; in fact:

- $(\pi_1)_\# \gamma = (\pi_1)_\# s_\# \mu = (\pi_1 \circ s)_\# \mu = (\mathrm{id}_{\mathbb{R}^n_+})_\# \mu = \mu$;
- $\eta(\mathcal{A}_R) = s_\# \mu(\mathcal{A}_R) = \mu(s^{-1}(\mathcal{A}_R)) = \mu(\mathbb{R}^n_+) = 1$.

Since $\mathcal{M} = \mathcal{C}$, we can use Proposition 2.2.2 with γ:

$$\begin{aligned}
0 &\leq r^* d\mathcal{P}(\eta^*; \gamma - \eta^*) + \lambda^* \cdot d\mathcal{G}(\eta^*; \gamma - \eta^*) \\
&= r^* \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} p_{\eta^*}(x, y)(\gamma - \eta^*)(dx, dy) + \lambda^* \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} g_{\eta^*}(x, y)(\gamma - \eta^*)(dx, dy) \\
&= \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} h(x, y)(\gamma - \eta^*)(dx, dy) \\
&= \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} h(x, y)\gamma(dx, dy) - \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} h(x, y)\eta^*(dx, dy).
\end{aligned}$$

Finally, using the definition of γ and the fact that $s(x) \in K$, we obtain:

$$\int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} h(x, y)\gamma(dx, dy) = \int_{\mathbb{R}^n_+} h \circ s(x)\mu(dx) = \int_{\mathbb{R}^n_+} m(x)\mu(dx) = \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} m(x, y)\eta^*(dx, dy),$$

and this completes the proof. □

Corollary 2.2.9: *([10], p. 16). We use the same assumptions and notations as in the Proposition 2.2.8. Suppose $\hat{\eta} \in \mathcal{M}$ such that $d\mathcal{P}$ and $d\mathcal{G}$ are integral operators on $\mathcal{C} - \hat{\eta}$ as well with $p_{\eta^*} = p_{\hat{\eta}}$ and $g_{\eta^*} = g_{\hat{\eta}}$. If $\hat{\eta}(K) = 1$, then the analogue of* (2.3) *holds with $\hat{\eta}$ instead of η^*:*

$$r^* d\mathcal{P}(\hat{\eta}; \eta - \hat{\eta}) + \lambda^* \cdot d\mathcal{G}(\hat{\eta}; \eta - \hat{\eta}) \geq 0.$$

Proof. Let us show that the following equality holds:

$$r^* d\mathcal{P}(\hat{\eta}; \eta - \hat{\eta}) + \lambda^* \cdot d\mathcal{G}(\hat{\eta}; \eta - \hat{\eta}) = r^* d\mathcal{P}(\eta^*; \eta - \eta^*) + \lambda^* \cdot d\mathcal{G}(\eta^*; \eta - \eta^*),$$

which, by expanding the terms on both sides, is equivalent to verify that:

$$\int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} h(x, y)\, \hat{\eta}(dx, dy) = \int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} h(x, y)\, \eta^*(dx, dy).$$

As noted in the proof of Proposition 2.2.8, $\hat{\eta}(K) = 1$ is equivalent to have $h(x, y) = m(x, y)$ $\hat{\eta}$-a.e.. Hence, we obtain:

$$\begin{aligned}
\int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} h(x, y)\, \hat{\eta}(dx, dy) &= \int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} m(x, y)\, \hat{\eta}(dx, dy) \\
&= \int_{\mathbb{R}_+^n} m(x)\, \mu(dx) \\
&= \int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} m(x, y)\, \eta^*(dx, dy) \\
&= \int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} h(x, y)\, \eta^*(dx, dy)
\end{aligned}$$

where the last equality is due to Proposition 2.2.8. □

Corollary 2.2.10: *([10], p. 15). In the same setting of Proposition 2.2.8, we suppose that the functional $\mathcal{P}$ is given by:*

$$\mathcal{P}(\eta) = \int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} p(x, y)\, \eta(dx, dy),$$

where $p : \mathbb{R}_+^n \times \mathbb{R}_+^n \to \mathbb{R}$ is a continuous function and the function $\mathcal{G} = 0$. We define the set

$$H := \left\{ (x, y) \in \mathcal{A}_R \mid y \in \arg\min_{t \in [0,x]} p(x, t) \right\}.$$

Then, a measure $\eta \in \mathcal{M}$ is an optimal reinsurance contract if and only if $\eta(H) = 1$.

Proof. Proof of ($\Rightarrow$). By definition we have for all $\gamma \in \mathcal{M}$:

$$\begin{aligned}
&d\mathcal{P}(\eta; \gamma - \eta) = \lim_{t \to 0^+} \frac{\mathcal{P}(\eta + t(\gamma - \eta)) - \mathcal{P}(\eta)}{t} = \int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} p(x, y)\, (\gamma - \eta)(dx, dy); \\
&d\mathcal{G}(\eta; \gamma - \eta) = 0.
\end{aligned}$$

Hence, we can choose $p_\eta = p$ and $g_\eta = 0$. In this way, the set H is exactly the set K in Proposition 2.2.8, so $\eta(H) = 1$.

Proof of ($\Leftarrow$). Let us take η^* optimal reinsurance contract. Now, h defined in the proof of

Proposition 2.2.8 is $h(x, y) = r^* p(x, y)$ and using the same chain of equality in the proof of Corollary 2.2.10 we have:

$$\int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} r^* p(x, y)\, \eta(dx, dy) = \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} r^* p(x, y)\, \eta^*(dx, dy),$$

which gives us the desired equality $\mathcal{P}(\eta) = \mathcal{P}(\eta^*)$. □

2.3 Optimal Transport Approach

In this section, we will analyze what was previously discussed from the perspective of optimal transport. In our framework, as in the previous section, μ represents the distribution of the fixed portfolio, while ν represents the distribution for the risk carried by the reinsurer, which can vary and we have to choose it in an optimal way. In contrast to before, however, we do not impose restrictions on the set $\mathcal{S}$, which may be formed by an infinite number of constraints, including bilateral ones. At this point, we state two assumptions:

Assumption 2.3.1: *If $\eta^* \in \mathcal{S}$ is an optimal reinsurance contract, then for every $\eta \in \mathcal{S}$ and $0 \leq t \leq 1$, it holds that:*

$$\mathcal{P}(\eta^*) \leq \mathcal{P}((1-t)\eta^* + t\eta).$$

Assumption 2.3.2: *For every $\eta \in \mathcal{S}$, the directional derivative $d\mathcal{P}(\eta; \cdot)$ is a well-defined map from $\mathcal{S} - \eta$ to $\mathbb{R}$ and can be expressed as an integral operator, meaning there exists a measurable function $p_\eta : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ such that:*

$$d\mathcal{P}(\eta; \theta - \eta) = \int_{\mathbb{R}^n \times \mathbb{R}^n} p_\eta(x, y)(\theta - \eta)(dx, dy)$$

for every $\theta \in \mathcal{S}$.

Keeping in mind that our objective is a linearization of the problem, as in the previous section, the second assumption aligns with this direction, while the first seems somewhat unusual. It would have been more natural to impose some geometric condition on the structure of $\mathcal{S}$, such as its convexity. However, by not doing so, we are able to remain more general while still obtaining relevant results.

Now, we would like to establish that an optimal reinsurance contract satisfies conditions that can be framed as an optimal transport problem; the next proposition goes in this direction:

Proposition 2.3.3: *([10], p. 17). Under Assumptions 2.3.1 and 2.3.2, and given an optimal reinsurance contract η^*, it holds that:*

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} q_{\eta^*}(x, y)\eta^*(dx, dy) = \min_{\nu \in \pi_2(\mathcal{S})} \mathcal{C}(\mu, \nu), \tag{2.10}$$

where the function $q_{\eta^*} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *coincides with* p_{η^*} *on* $\mathcal{A}_R$ *and takes the value* $+\infty$ *otherwise,* $\pi_2(\mathcal{S}) := \{(\pi_2)_\#\eta : \eta \in \mathcal{S}\}$ *and*

$$\mathcal{C}(\mu, \nu) := \min_{\eta \in \mathcal{S} \cap \Pi(\mu,\nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} q_{\eta^*}(x, y)\eta(dx, dy). \tag{2.11}$$

Proof. Assumptions 2.3.1 and 2.3.2 imply:

$$\begin{aligned} 0 \le \lim_{t \to 0^+} \frac{\mathcal{P}(\eta^* + t(\eta - \eta^*)) - \mathcal{P}(\eta^*)}{t} = d\mathcal{P}(\eta^*, \eta - \eta^*) = &\int_{\mathbb{R}^n \times \mathbb{R}^n} p_{\eta^*}(x, y)\eta(dx, dy) \\ &- \int_{\mathbb{R}^n \times \mathbb{R}^n} p_{\eta^*}(x, y)\eta^*(dx, dy) \end{aligned}$$

for every $\eta \in \mathcal{S}$ and η^* an optimal contract. So, we have:

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} p_{\eta^*}(x, y)\eta^*(dx, dy) = \min_{\eta \in \mathcal{S}} \int_{\mathbb{R}^n \times \mathbb{R}^n} p_{\eta^*}(x, y)\eta(dx, dy). \tag{2.12}$$

Since all the contracts involved have support on $\mathcal{A}_R$, we can replace p^*_η with q^*_η, obtaining

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} q_{\eta^*}(x, y)\eta^*(dx, dy) = \min_{\eta \in \mathcal{S}} \int_{\mathbb{R}^n \times \mathbb{R}^n} q_{\eta^*}(x, y)\eta(dx, dy). \tag{2.13}$$

The last step immediately follows decomposing the minimum in (2.13), keeping in mind our desired outcome and that ν is the distribution that can vary. □

The equation (2.10) represents the (constrained) optimal problem we were seeking. It is important to emphasize that the property we have found is a necessary condition of being an optimal reinsurance contract but is not completely characterizing. For example, we are not taking into account that we need to minimize the functional $\mathcal{P}$. Also, we are facing an issue similar to that in the previous section: q_{η^*} depends on η^*, and thus we are confronted with an optimization problem that depends on the chosen optimal reinsurance treaty. The positive aspect, however, is that in most cases, the cost function will depend on η^* only on a finite number of parameters that can be considered fixed at the outset. Furthermore, it is useful to make the following consideration: if for every $\nu \in \pi_2(\mathcal{S})$ there exists a unique solution of (2.11), and this solution is an optimal Monge coupling, then η^* will also be deterministic. In fact, by definition, η^* solves the problem $\mathcal{C}(\mu, \nu^*)$ with $\nu^* := (\pi_2)_\#\eta^*$. Therefore, by the uniqueness of the solution, we would have $\eta^* = (\mathrm{id}, \gamma^*)_\#\mu$ for some function γ^*. Let us now explore how to ensure the existence of this last property.

Setting 2.3.4: *Consider the following hypothesis:*

(i) the functional $\mathcal{P}$ *is of the form*

$$\mathcal{P}(\eta) = \mathcal{P}_1(T_\#\eta) + \mathcal{P}_2((\pi_2)_\#\eta), \tag{2.14}$$

where $T : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ *such that* $T(x, y) = x - y$. *Furthermore, the functionals* $\eta \mapsto \mathcal{P}_1(T_\# \eta)$ *and* $\eta \mapsto \mathcal{P}_2((\pi_2)_\# \eta)$ *from* $\mathcal{M}$ *to* $\mathbb{R}$ *satisfy the analogues of Assumption 2.3.2, i.e. there exist two measurable functions* $h_\eta : \mathbb{R}^n \to \mathbb{R}$ *and* $k_\eta : \mathbb{R}^n \to \mathbb{R}$ *such that:*

$$d\mathcal{P}_1(T_\#\eta; T_\#\theta - T_\#\eta) = \int_{\mathbb{R}^n} h_\eta(x)(T_\#\theta - T_\#\eta)(dx),$$

$$d\mathcal{P}_2((\pi_2)_\#\eta; (\pi_2)_\#\theta - (\pi_2)_\#\eta) = \int_{\mathbb{R}^n} k_\eta(x)((\pi_2)_\#\theta - (\pi_2)_\#\eta)(dx)$$

for every $\eta, \theta \in \mathcal{S}$*;*

(ii) for every $\nu \in \pi_2(\mathcal{S})$*, it holds that* $\Pi(\mu, \nu) \cap \mathcal{M} \subset \mathcal{S}$*;*

(iii) h_{η^*} *is strictly convex and bounded from below for every optimal reinsurance contract* η^**;*

(iv) the random vector X *has a continuous distribution.*

Let us begin with a simple lemma that ensures that hypothesis (i) is not in contradiction with the assumption made earlier:

Lemma 2.3.5: *If* $\mathcal{P}$ *satisfies (i), then it also satisfies Assumption 2.3.2 with*

$$p_\eta(x, y) = h_\eta(x - y) + k_\eta(y).$$

Proof. Proceeding directly from the definition, we obtain for every $\eta, \theta \in \mathcal{S}$:

$$\begin{aligned}
d\mathcal{P}(\eta; \eta - \theta) &= \lim_{t \to 0^+} \frac{\mathcal{P}(\eta + t(\eta - \theta)) - \mathcal{P}(\eta)}{t} \\
&= \lim_{t \to 0^+} \frac{\mathcal{P}_1(T_\#(\eta + t(\eta - \theta))) + \mathcal{P}_2((\pi_2)_\#(\eta + t(\eta - \theta))) - \mathcal{P}_1(T_\#\eta) - \mathcal{P}_2((\pi_2)_\#\eta)}{t} \\
&= \lim_{t \to 0^+} \frac{\mathcal{P}_1(T_\#\eta + t(T_\#(\eta - \theta))) - \mathcal{P}_1(T_\#\eta)}{t} \\
&\quad + \lim_{t \to 0^+} \frac{\mathcal{P}_2((\pi_2)_\#\eta + t((\pi_2)_\#\eta - (\pi_2)_\#\theta)) - \mathcal{P}_2((\pi_2)_\#\eta)}{t} \\
&= d\mathcal{P}_1(T_\#\eta; T_\#\theta - T_\#\eta) + d\mathcal{P}_2((\pi_2)_\#\eta; (\pi_2)_\#\theta - (\pi_2)_\#\eta) \\
&= \int_{\mathbb{R}^n} h_\eta(x)(T_\#\theta - T_\#\eta)(dx) + \int_{\mathbb{R}^n} k_\eta(x)((\pi_2)_\#\theta - (\pi_2)_\#\eta)(dx) \\
&= \int_{\mathbb{R}^n \times \mathbb{R}^n} h_\eta(x - y)(\theta - \eta)(dx, dy) + \int_{\mathbb{R}^n \times \mathbb{R}^n} k_\eta(y)(\theta - \eta)(dx, dy) \\
&= \int_{\mathbb{R}^n \times \mathbb{R}^n} h_\eta(x - y) + k_\eta(y)(\theta - \eta)(dx, dy).
\end{aligned}$$

□

Proposition 2.3.6: *([10], p. 17). We adopt the framework established in Setting 2.3.4, Assumption 2.3.1 and we set $n = 1$. Under these conditions, for every $\nu \in \pi_2(\mathcal{S})$, there exists a unique solution of* (2.11)*, and this solution is an optimal Monge coupling. In particular, every optimal contract is deterministic.*

Proof. Let $\nu \in \pi_2(\mathcal{S})$. Since $\mathcal{S} \subset \mathcal{M}$, from hypothesis (ii) we have that $\Pi(\mu,\nu) \cap \mathcal{M} = \Pi(\mu,\nu) \cap \mathcal{S}$ and so:

$$\begin{aligned}\mathcal{C}(\mu,\nu) &:= \min_{\eta \in \mathcal{S}\cap\Pi(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} q_{\eta^*}(x,y)\eta(dx,dy) = \min_{\eta \in \mathcal{M}\cap\Pi(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} q_{\eta^*}(x,y)\eta(dx,dy) \\ &= \min_{\eta \in \mathcal{M}\cap\Pi(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} p_{\eta^*}(x,y)\eta(dx,dy) \geq \inf_{\eta \in \Pi(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} p_{\eta^*}(x,y)\eta(dx,dy) \\ &= \int_{\mathbb{R}} k_{\eta^*}(y)\nu(dy) + \inf_{\eta \in \Pi(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} h_{\eta^*}(x-y)\eta(dx,dy).\end{aligned}$$

In the last sum, we have thus obtained a constant (fixed ν) and a Kantorovich optimization problem. In particular, the following holds:

$$\begin{aligned}\inf_{\eta \in \Pi(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} h_{\eta^*}(x-y)\eta(dx,dy) &\leq \mathcal{C}(\mu,\nu) - \int_{\mathbb{R}} k_{\eta^*}(y)\nu(dy) \\ &\leq \mathcal{C}(\mu,\nu^*) - \int_{\mathbb{R}} k_{\eta^*}(y)\nu(dy) \\ &= \int_{\mathbb{R}\times\mathbb{R}} q_{\eta^*}(x,y)\eta^*(dx,dy) - \int_{\mathbb{R}} k_{\eta^*}(y)\nu(dy) < +\infty,\end{aligned}$$

where the last inequality follows from Assumption 2.3.2. Hence, the optimization problem is finite and we can use Theorem 1.3.8, obtaining that the unique solution π^*_ν is given by the Monge map $T_{\text{mon}} =: g_\nu$, i.e. $\pi^*_\nu = (\text{id}, g_\nu)_\# \mu$.
Now, we observe that $\nu \prec_1 \mu$, i.e. $F_\nu(x) \geq F_\mu(x)$ for every $x \in \mathbb{R}$. In fact, let us call γ the measure in $\mathcal{S}$ such that $(\pi_2)_\# \eta = \nu$, in particular $\eta \in \Pi(\mu,\nu) \cap \mathcal{M}$. Thus, it holds:

$$F_\mu(x) = \mu((-\infty,x]) = \gamma((-\infty,x]\times\mathbb{R}) = \gamma((-\infty,x]\times(-\infty,x]) \leq \gamma(\mathbb{R}\times(-\infty,x]) = \nu((-\infty,x]) = F_\nu(x).$$

The last step is to show that $0 \leq g_\nu(x) \leq x$ μ-a.s., so that $\pi^*_\nu \in \Pi(\mu,\nu) \cap \mathcal{M} = \Pi(\mu,\nu) \cap \mathcal{S}$, and therefore it is also a solution to $\mathcal{C}(\mu,\nu)$.
Using (1.3) we get $g_\nu(x) \leq x$ if and only if $F_\nu(x) \geq F_\mu(x)$, which is the definition of $\nu \prec_1 \mu$. Furthermore, $F_\mu(x) > 0$ if and only if $(-\infty, x]$ is not μ-negligible and since $\nu((\mathbb{R}_+)^c) = 0$ we have $g_\nu \geq 0$ μ-a.s. for $x \geq 0$. □

Remark 2.3.7: In any dimension n, the inequality from the previous proof always holds:

$$\mathcal{C}(\mu,\nu) \geq \int_{\mathbb{R}} k_{\eta^*}(y)\,\nu(dy) + \inf_{\eta \in \Pi(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} h_{\eta^*}(x-y)\,\eta(dx,dy).$$

If we assume that h is also superlinear (i.e., $\lim_{|x|\to+\infty} h(x)/|x| = +\infty$), then, thanks to the Gangbo-McCann theorem ([3], p. 63), we know that there exists a unique solution to the optimization problem in the last term given by the Monge map:

$$T(x) = x - (\nabla h_{\eta^*})^{-1} \circ \nabla\phi(x),$$

where ϕ is an h_{η^*}-concave function. The problem is that, a priori, it is not automatic that:

$$0 \leq T(x) \leq x,$$

which would guarantee the same result as Proposition 2.3.6 for any n.

For the next proposition, we adopt the notation $\pi_{i,j} := \pi_j \circ \pi_i$, where $\pi_i : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ for $i = 1, 2$ and $\pi_j : \mathbb{R}^n \to \mathbb{R}$ for $j = 1, \ldots, n$.

Proposition 2.3.8: *([10], p. 19). Suppose that $\mathcal{P}$ can be written as $\mathcal{P} = \mathcal{F} \circ \mathcal{R}$ where $\mathcal{F} : \mathscr{P}(\mathbb{R})^n \to \mathbb{R}$ is a lower semi-continuous functional and $\mathcal{R} : \mathcal{M} \to \mathscr{P}(\mathbb{R})^n$ is defined as:*

$$\mathcal{R}(\eta) = ((\pi_{2,1})_{\#}\eta, \ldots, (\pi_{2,n})_{\#}\eta)\,.$$

As for the constraints, we assume that $\mathcal{S} = \mathcal{U}^{-1}(E)$ where E is a closed set in a topological space O, and $\mathcal{U} = \mathcal{V} \circ \mathcal{R}$ with $\mathcal{V} : \mathscr{P}(\mathbb{R})^n \to O$ a continuous map on the image of $\mathcal{R}$. Let C be a copula for the distribution μ of X. Furthermore, let η^* be an optimal treaty, and we set $\nu_i^* := (\pi_{2,i})_{\#}\eta^*$ for each $i = 1, \ldots, n$. We define θ as the probability measure on $\mathbb{R}^n \times \mathbb{R}^n$ with the following cumulative distribution function:

$$F_\theta(x_1, \ldots, x_n, y_1, \ldots, y_n) = C(\min\{F_{\mu_1}(x_1), F_{\nu_1^*}(y_1)\}, \ldots, \min\{F_{\mu_n}(x_n), F_{\nu_n^*}(y_n)\}).$$

Then θ is also an optimal treaty. In particular, if each F_{μ_i} is continuous, then there exists a deterministic optimal contract, with components given by the functions:

$$R_i(x) = F_{\nu_i^*}^{[-1]} \circ F_{\mu_i}(x), \quad i = 1, \ldots, n.$$

Proof. Let us begin with the first part of the statement and show that θ is optimal:

- *the first marginal of θ is μ*: let us compute the cumulative distribution function of $(\pi_1)_{\#}\theta$

$$\begin{aligned}
F_{(\pi_1)_{\#}\theta}(x_1, \ldots, x_n) &= \theta((-\infty, x_1] \times \cdots \times (-\infty, x_n] \times \mathbb{R} \times \cdots \times \mathbb{R}) \\
&= \lim_{y_1\to+\infty,\ldots,y_n\to+\infty} \theta((-\infty, x_1] \times \cdots \times (-\infty, x_n] \times (-\infty, y_1] \times \cdots \times (-\infty, y_n]) \\
&= \lim_{y_1\to+\infty,\ldots,y_n\to+\infty} F_\theta(x_1, \ldots, x_n, y_1, \ldots, y_n) \\
&= \lim_{y_1\to+\infty,\ldots,y_n\to+\infty} C(\min\{F_{\mu_1}(x_1), F_{\nu_1^*}(y_1)\}, \ldots, \min\{F_{\mu_n}(x_n), F_{\nu_n^*}(y_n)\}) \\
&= C(\min\{F_{\mu_1}(x_1), 1\}, \ldots, \min\{F_{\mu_n}(x_n), 1\}) \\
&= C(F_{\mu_1}(x_1), \ldots, F_{\mu_n}(x_n)) = F_\mu(x_1, \ldots, x_n),
\end{aligned}$$

where the last equality is due to the Sklar's theorem.

- *θ is concentrated on $\mathcal{A}_R$*: consider a point $(x_1, \dots, x_n, y_1, \dots, y_n)$ such that there exists a strictly negative coordinate. Since μ_i and ν_i^* have support on $\mathbb{R}_+$ for every $i = 1, \dots, n$, it follows that $F_\theta(x_1, \dots, x_n, y_1, \dots, y_n) = 0$ because a copula is zero if any one of the arguments is zero. Therefore, the support of θ is contained in $\mathbb{R}^n_+ \times \mathbb{R}^n_+$.
Now, we see that $\nu_i^* \prec_1 \mu_i$ for every $i = 1, \dots, n$; indeed:

$$
\begin{aligned}
F_{\mu_i}(x) &= (\pi_{1,i})_\# \eta^*((-\infty, x]) \\
&= \eta^*(\mathbb{R} \times \dots \times \mathbb{R} \times \underbrace{(-\infty, x]}_{i\text{-th}} \times \mathbb{R} \times \dots \times \mathbb{R}) \\
&= \eta^*(\mathbb{R} \times \dots \times \mathbb{R} \times \underbrace{(-\infty, x]}_{i\text{-th}} \times \mathbb{R} \times \dots \times \mathbb{R} \times \underbrace{(-\infty, x]}_{(n+i)\text{-th}} \times \mathbb{R} \times \dots \times \mathbb{R}) \\
&\leq \eta^*(\mathbb{R} \times \dots \times \mathbb{R} \times \underbrace{(-\infty, x]}_{(n+i)\text{-th}} \times \mathbb{R} \times \dots \times \mathbb{R}) \\
&= (\pi_{2,i})_\# \eta^*((-\infty, x]) = F_{\nu_i^*}(x).
\end{aligned}
$$

Suppose by contradiction that there exists $(x_1, \dots, x_n, y_1, \dots, y_n) \in \operatorname{spt}(\theta) \cap \big((\mathbb{R}^n_+ \times \mathbb{R}^n_+) \setminus \mathcal{A}_R\big)$. Thus, there exists a coordinate i such that $y_i > x_i$. Without loss of generality, suppose $i = 1$. Hence, it holds :

$$
F_\theta(x_1, \dots, x_n, x_1, y_2, \dots, y_n) < F_\theta(x_1, \dots, x_n, y_1, y_2, \dots, y_n).
$$

On the other hand, we have:

$$
\begin{aligned}
&F_\theta(x_1, \dots, x_n, x_1, y_2 \dots, y_n) \\
&= C(\min\{F_{\mu_1}(x_1), F_{\nu_1^*}(x_1)\}, \dots, \min\{F_{\mu_n}(x_n), F_{\nu_n^*}(y_n)\}) \\
&= C(F_{\mu_1}(x_1), \min\{F_{\mu_2}(x_2), F_{\nu_2^*}(y_2)\}, \dots, \min\{F_{\mu_n}(x_n), F_{\nu_n^*}(y_n)\}) \\
&= C(\min\{F_{\mu_1}(x_1), F_{\nu_1^*}(y_1)\}, \dots, \min\{F_{\mu_n}(x_n), F_{\nu_n^*}(y_n)\}) \\
&= F_\theta(x_1, \dots, x_n, y_1, y_2 \dots, y_n),
\end{aligned}
$$

which leads to a contradiction.

- *θ satisfies the constraints*: it suffices to see that $\mathcal{R}(\theta) = \mathcal{R}(\eta^*)$, that is, for every $i = 1, \dots, n$, it must hold that $(\pi_{2,i})_\# \theta = \nu_i^*$. Similarly to the first point, we have:

$$
\begin{aligned}
F_{(\pi_{2,i})_\#\theta}(y) &= \lim_{\substack{x_1 \to +\infty, \dots, x_n \to +\infty, \\ y_1 \to +\infty, \dots, y_{n+i-1} \to +\infty, \\ y_{n+i+1} \to +\infty, \dots, y_n \to +\infty}} F_\theta(x_1, \dots, x_n, y_1, \dots, y_{n-i-1}, y, y_{n-i+1}, \dots, y_n) \\
&= C(1, \dots, 1, F_{\nu_i^*}(y), 1, \dots, 1) = F_{\nu_i^*}(y)
\end{aligned}
$$

where, for the last equality, we have used that a copula is equal to x if one argument is x and all others are one.

- *Optimality*: for the exact same reason as the previous point.

Let us now focus on the second part of the statement. Consider the deterministic contract:

$$\tilde{\theta} := (\mathrm{id}, (F_{\nu_1^*}^{[-1]} \circ F_{\mu_1}, \ldots, F_{\nu_n^*}^{[-1]} \circ F_{\mu_n}))_{\#}\mu.$$

We note that the maps $F_{\nu_i^*}^{[-1]} \circ F_{\mu_i}$ are well-defined by Theorem 1.3.6. We fix an element $(x_1, \ldots, x_n, y_1, \ldots, y_n) \in \mathbb{R}^n \times \mathbb{R}^n$. Let c_i be the real number such that $F_{\mu_i}(c_i) = F_{\nu_i^*}(y_i)$ for every $i = 1, \ldots, n$. We observe that $F_{\mu_i}(\min\{x_i, c_i\}) = \min\{F_{\mu_i}(x_i), F_{\nu_i^*}(y_i)\}$ since $x_i \leq c_i$ if and only if $F_{\mu_i}(x_i) \leq F_{\nu_i^*}(y_i)$. Thus, it holds:

$$\begin{aligned}
&F_{\tilde{\theta}}(x_1, \ldots, x_n, x_1, y_2 \ldots, y_n) \\
&= \mu(\{(z_1, \ldots, z_n) \mid z_1 \leq x_1, \ldots, z_n \leq x_n, F_{\nu_1^*}^{[-1]} \circ F_{\mu_1}(z_1) \leq y_1, \ldots, F_{\nu_n^*}^{[-1]} \circ F_{\mu_n}(z_n) \leq y_n\}) \\
&= \mu(\{(z_1, \ldots, z_n) \mid z_1 \leq x_1, \ldots, z_n \leq x_n, F_{\mu_1}(z_1) \leq F_{\nu_1^*}(y_1), \ldots, F_{\mu_n}(z_n) \leq F_{\nu_n^*}(y_n)\}) \\
&= \mu(\{(z_1, \ldots, z_n) \mid z_1 \leq x_1, \ldots, z_n \leq x_n, z_1 \leq c_1, \ldots, z_n \leq c_n\}) \\
&= F_\mu(\min\{x_1, c_1\}, \ldots, \min\{x_n, c_n\}) \\
&= C(F_{\mu_1}(\min\{x_1, c_1\}), \ldots, F_{\mu_n}(\min\{x_n, c_n\})) \\
&= C(\min\{F_{\mu_1}(x_1), F_{\nu_1^*}(y_1)\}, \ldots, \min\{F_{\mu_n}(x_n), F_{\nu_n^*}(y_n)\}).
\end{aligned}$$

Hence, we conclude by the first part of the proposition. □

Remark 2.3.9: Proposition 2.3.8 tells us that if the risk measure and the constraints depend only on the reinsured distribution, then under appropriate continuity assumptions, randomization is unnecessary. However, it is not necessarily the case that all optimal contracts are deterministic, in contrast to Proposition 2.3.6.

Chapter 3

Case Studies

In this chapter, we will explore some applications of the results seen in the previous sections, all based on Section 6 of the article [10].

The first case study in Section 3.1 was originally examined, in a slightly different setting, by De Finetti in [1], a work that is still considered a cornerstone of reinsurance theory. Proposition 2.2.8 is therefore used to derive the structure of the optimal contract. In Section 3.2, the constraint remains an inequality, but this time it is based on the Value at Risk measure. This prevents us from following the previous approach, making Proposition 2.2.5 essential. The third and fourth applications instead rely on the theory developed in Section 2.3: the constraints become equalities, and the reformulation via optimal transport proves to be effective. In the case of Section 3.3, the constraint is imposed on the expected value of R. After applying Proposition 2.3.6, we use the characterization of projections onto a closed convex set in the function space $L^2(\mathbb{R}, \mu)$ to determine the solution. In Section 3.4, the constraint is instead on the variance. The procedure follows the same steps as in the previous case, but obtaining the optimal contract structure requires the use of a Lagrangian operator. Finally, Section 3.5 presents a case where the nature of the solution differs from all previous ones. Specifically, when the constraint fixes the marginals of the distribution of $X - R$, and in the case where $n = 2$ with distributions defined on a discrete space, we discover that optimality cannot be achieved through a deterministic contract.

Regarding the notation used in this chapter, as before, let $T : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ be the map defined as $T(x, y) = x - y$. Furthermore, given $\theta \in \mathscr{P}(\mathbb{R})$, we define $\overline{\theta} := \int_{\mathbb{R}} x\, \theta(dx)$.

3.1 A Classical Case with $\mathcal{C} = \mathcal{M}$

An insurer operating at the first line has n sub-portfolios with insurance risks $X_1, \dots, X_n$. The goal is to determine a reinsurance contract $R = (R_1, \dots, R_n)$ that minimizes the total expected loss after reinsurance under a constraint on the retained aggregate variance. The premium for the i-th contract is calculated using the expected value principle with a safety loading factor β_i, meaning the premium is given by:

$$P_i = (1 + \beta_i)\mathbb{E}[R_i].$$

Here, the safety loading β_i is an additional charge included in the premium to account for uncertainties, risk margins or insurer profitability. Consequently, the total loss $\mathcal{L}$ incurred by the insurer after reinsurance can be expressed as:

$$\mathcal{L} := \sum_{i=1}^{n} \left(X_i - R_i + (1 + \beta_i)\mathbb{E}[R_i]\right).$$

Furthermore, suppose that all the β_i's are different and ordered such that $0 < \beta_1 < \dots < \beta_n$. To make an appropriate choice of $\mathcal{P}$, given η in $\mathcal{M}$, we compute:

$$\begin{aligned}
\mathbb{E}[\mathcal{L}] &= \sum_{i=1}^{n} \left(\mathbb{E}[X_i] - \mathbb{E}[R_i] + (1 + \beta_i)\mathbb{E}[R_i]\right) \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i] + \mathbb{E}\left[\sum_{i=1}^{n} \beta_i R_i\right] \\
&= \sum_{i=1}^{n} \int_{\Omega} X_i(\omega)\, dP(\omega) + \int_{\Omega} \sum_{i=1}^{n} \beta_i R_i(\omega)\, dP(\omega) \\
&= \sum_{i=1}^{n} \int_{\mathbb{R}_+} x\, \mu_i(dx) + \int_{\mathbb{R}_+^n} \sum_{i=1}^{n} \beta_i y_i\, ((\pi_2)_{\#}\eta)(dy) = \sum_{i=1}^{n} \overline{\mu_i} + \int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} \sum_{i=1}^{n} \beta_i y_i\, \eta(dx, dy).
\end{aligned}$$

The term $\sum_{i=1}^{n} \overline{\mu_i}$ depends only on μ, and since this distribution is considered fixed, we can define $\mathcal{P}$ as:

$$\mathcal{P}(\eta) := \int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} \sum_{i=1}^{n} \beta_i y_i\, \eta(dx, dy).$$

Regarding the constraints, in this example, we assume that there exists a constant c such that $0 < c < \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)$, and for which the contract must satisfy the inequality $\mathrm{Var}\left(\sum_{i=1}^{n}(X_i - R_i)\right) \le c$. Thus, with steps analogous to how we found $\mathbb{E}[\mathcal{L}]$ and recalling that a reinsurance treaty η represents the joint distribution of the random vector (X, R), we can define $\mathcal{G}$ as:

$$\mathcal{G}(\eta) := \int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} \left(\sum_{i=1}^{n}(x_i - y_i)\right)^2 - \left(\sum_{i=1} \int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} (x_i - y_i)\, \eta(dx, dy)\right)^2 \eta(dx, dy) - c.$$

Now, using the notation in Proposition 2.2.5, let us find p_{η^*} and g_{η^*} such that $\mathcal{P}$ and $\mathcal{G}$ are integral operators with continuous kernels. Fixing η^* as the optimal contract, it holds for every $\eta \in \mathcal{C}$ that:

$$
\begin{aligned}
d\mathcal{P}(\eta^*, \eta - \eta^*) &= \lim_{t \to 0^+} \frac{\int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} \sum_{i=1}^n \beta_i y_i \, (\eta^* + t(\eta - \eta^*))(dx, dy) - \int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} \sum_{i=1}^n \beta_i y_i \, \eta^*(dx, dy)}{t} \\
&= \int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} \sum_{i=1}^n \beta_i y_i \, (\eta - \eta^*)(dx, dy).
\end{aligned}
$$

Hence, we can define $p_{\eta^*} : \mathbb{R}_+^n \times \mathbb{R}_+^n \to \mathbb{R}$ such that

$$
p_{\eta^*}(x, y) := \sum_{i=1}^n \beta_i y_i.
$$

Let us now try to do the same for $\mathcal{G}$ and define $A(x, y) := \sum_{i=1}^n (x_i - y_i)$ and $B(\eta) := \sum_{i=1}^n \int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} (x_i - y_i) \, \eta(dx, dy)$. Then it follows:

$$
\begin{aligned}
d\mathcal{G}(\eta^*, \eta - \eta^*) &= \lim_{t \to 0^+} \frac{\int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} A(x, y)^2 - B(\eta^* + t(\eta - \eta^*))^2 \, (\eta^* + t(\eta - \eta^*))(dx, dy) - c}{t} \\
&\quad - \frac{\int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} A(x, y)^2 - B(\eta^*)^2 \, \eta^*(dx, dy) - c}{t} \\
&= \lim_{t \to 0^+} \frac{\int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} A(x, y)^2 \, (\eta^* + t(\eta - \eta^*))(dx, dy) - B(\eta^* + t(\eta - \eta^*))^2}{t} \\
&\quad - \frac{\int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} A(x, y)^2 \, \eta^*(dx, dy) - B(\eta^*)^2}{t} \\
&= \int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} A(x, y)^2 \, (\eta - \eta^*)(dx, dy) - 2B(\eta^*) B(\eta - \eta^*) \\
&= \int_{\mathbb{R}_+^n \times \mathbb{R}_+^n} A(x, y)^2 - 2B(\eta^*) A(x, y) \, (\eta - \eta^*)(dx, dy).
\end{aligned}
$$

Noting that $B(\eta^*) = \sum_{i=1}^n \overline{(\pi_i \circ T)_\# \eta^*}$, we define $g_{\eta^*} : \mathbb{R}_+^n \times \mathbb{R}_+^n \to \mathbb{R}$ as

$$
g_{\eta^*}(x, y) := \left(\sum_{i=1}^n (x_i - y_i) \right)^2 - 2 \left(\sum_{i=1}^n \overline{(\pi_i \circ T)_\# \eta^*} \right) \left(\sum_{i=1}^n (x_i - y_i) \right).
$$

Throughout this argument, we consider $B(\eta^*)$ as a fixed parameter that limits how much mass the probability μ can place at the point zero. Specifically, we assume that

$$
\mu(\{0\}) < \frac{c}{B(\eta^*)^2}. \tag{3.1}
$$

Given the definitions of p_{η^*} and g_{η^*}, and using the identity $\eta^* = t\eta^* + (1-t)\eta^*$, it follows immediately that (2.2) is satisfied for every $\eta_1, \eta_2 \in \mathcal{M}$. Therefore, we have $\mathcal{C} = \mathcal{M}$ and the assumptions of Proposition 2.2.8 hold. Thus, we know that there exist two numbers r^* and $\overline{\lambda^*}$, both non-negative, such that the support of η^* is contained in the set:

$$K = \{(x,y) \in \mathcal{A}_R \mid y \in \arg\min_{t\in[0,x]} \big(r^* p_{\eta^*}(x,t) + \overline{\lambda^*} g_{\eta^*}(x,t)\big)\}.$$

In particular, it holds:

$$\begin{aligned}\mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*, \eta - \eta^*) = &\int_{\mathbb{R}^n_+\times\mathbb{R}^n_+} \left(\sum_{i=1}^n x_i - y_i\right)^2 \eta(dx,dy) \\ &- 2B(\eta^*)\int_{\mathbb{R}^n_+\times\mathbb{R}^n_+} \sum_{i=1}^n (x_i - y_i)\ \eta(dx,dy) + B(\eta^*)^2 - c. \quad (3.2)\end{aligned}$$

Our goal is to find a contract η such that (3.2) is negative. If $B(\eta^*)^2 - c < 0$, then by choosing η as the full reinsurance contract, we immediately achieve the desired result. Now, suppose $B(\eta^*)^2 - c \geq 0$. For every $\varepsilon > 0$, we define $K_\varepsilon := \big\{x \in \mathbb{R}^n_+ \mid \sum_{i=1}^n x_i \geq \varepsilon\big\}$ and consider $f^\varepsilon := (f_1^\varepsilon, \dots, f_n^\varepsilon)$ defined in such a way that $\sum_{i=1}^n x_i - f_i(x_i) = \varepsilon$ if $x \in K_\varepsilon$, otherwise $f_\varepsilon(x) = x$. In this way, $0 \leq f_\varepsilon \leq \mathrm{Id}$ and the measure $\eta_\varepsilon := (\mathrm{id}, f_\varepsilon)_{\#}\mu \in \mathcal{M}$. Equation (3.2) then becomes:

$$\begin{aligned}\mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*, \eta_\varepsilon - \eta^*) = &\int_{\mathbb{R}^n_+\times\mathbb{R}^n_+} \left(\sum_{i=1}^n x_i - f_i^\varepsilon(x_i)\right)^2 \mu(dx) \\ &- 2B(\eta^*)\int_{\mathbb{R}^n_+\times\mathbb{R}^n_+} \sum_{i=1}^n (x_i - f_i^\varepsilon(x_i))\ \mu(dx) + B(\eta^*)^2 - c \\ = &\ \mu(K_\varepsilon)\varepsilon^2 - 2B(\eta^*)\mu(K_\varepsilon)\varepsilon + B(\eta^*)^2 - c.\end{aligned}$$

There exists an $\varepsilon > 0$ such that the expression in the last line is strictly less than zero if and only if

$$\mu(K_\varepsilon) > \frac{B(\eta^*)^2 - c}{B(\eta^*)^2} =: \alpha \in [0,1). \quad (3.3)$$

If, for contradiction, there exists no α such that (3.3) holds for every ε, then

$$\mu\left(x \in \mathbb{R}^n_+ \setminus \{0\} \mid \sum_{i=1}^n x_i < \varepsilon\right) > 1 - \alpha - \mu(\{0\}).$$

But as $\varepsilon \to 0^+$, the left-hand side tends to zero while the right-hand side is strictly positive by (3.1). Thus, by part of the statement of Proposition 2.2.2, we can assume $r^* > 0$. Therefore, setting $\lambda^* := \frac{\overline{\lambda^*}}{r^*}$, we can rewrite K as:

$$K = \{(x,y) \in \mathcal{A}_R \mid y \in \arg\min_{t\in[0,x]} (p_{\eta^*}(x,t) + \lambda^* g_{\eta^*}(x,t))\}.$$

Furthermore, it holds that $\lambda^* > 0$; in fact, the minimum of $p_{\eta^*}(x,\cdot)$ occurs at $y=0$, and so the contract η_0, for which no reinsurance takes place, would be optimal. However, this leads to a contradiction since:

$$\mathcal{G}(\eta_0) = \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) - c > 0.$$

Naturally, we want to compute the partial derivatives of the function $h(x,\cdot) := p_{\eta^*}(x,\cdot) + \lambda^* g_{\eta^*}(x,\cdot)$. Using the definitions of p_{η^*} and g_{η^*} we obtain:

$$\frac{\partial h(x,y)}{\partial y_k} = \beta_k - 2\lambda^*\left(\sum_{i=1}^{n}(x_i - y_i)\right) + 2\lambda^* B(\eta^*), \quad k = 1,\dots,n.$$

Setting the expressions we found equal to zero and using that $\lambda^* \neq 0$, we have:

$$\sum_{i=1}^{n}(x_i - y_i) = \frac{\beta_k}{2\lambda^*} + B(\eta^*), \quad k = 1,\dots,n.$$

The right-hand side varies with k since $\beta_i \neq \beta_j$ if $i \neq j$, while the left-hand side is constant with respect to k. Thus, no $y \in \mathbb{R}_+^n$ exists that solves the system. Hence, the minima occur on the boundary $\partial\mathcal{A}_R$, and we need to try a different approach. We define $s_X := \sum_{i=1}^n x_i$ and $s_Y := \sum_{i=1}^n y_i$ and rewrite $h(x,y)$:

$$\begin{aligned}
h(x,y) &= \sum_{i=1}^{n} \beta_i y_i + \lambda^*(s_X - s_Y)^2 - 2\lambda^* B(\eta^*)(s_X - s_Y) \\
&= \sum_{i=1}^{n}(\beta_i - \beta_n) y_i + \beta_n s_Y + \lambda^*\left(s_Y^2 + 2(B(\eta^*) - s_X)s_Y\right) + \lambda^* s_X^2 - 2\lambda^* B(\eta^*) s_X \\
&= -\sum_{i=1}^{n}(\beta_n - \beta_i) y_i + \lambda^*\left(s_Y - s_X + B(\eta^*) + \frac{\beta_n}{2\lambda^*}\right)^2 + \lambda^* s_X^2 - 2\lambda^* B(\eta^*) s_X \\
&\quad - \lambda\left(B(\eta^*) + \frac{\beta_n}{2\lambda^*} - s_X\right)^2.
\end{aligned}$$

In this way, in the last expression only the first two terms depend on y, and thus the problem reduces to finding the minima of the function $a(x,\cdot)$ defined as:

$$a(x,y) = -\sum_{i=1}^{n}(\beta_n - \beta_i) y_i + \lambda^*\left(s_Y - s_X + B(\eta^*) + \frac{\beta_n}{2\lambda^*}\right)^2.$$

Let us try to characterize these points through the following two observations:

- *let y^* be a minimizer for $a(x,\cdot)$ in $[0,x]$. If there exists $i \geq 2$ such that $y_i^* > 0$, then $y_j^* = x_j$ for every $j = 1,\dots,i-1$.*

In fact, suppose by contradiction there exists $j < i$ such that $y_j^* < x_j$. We define a new vector $\tilde{y}$:

$$\tilde{y}_k = \begin{cases} y_k^*, & \text{if } k \neq i, j, \\ y_i^* - \min\{x_j - y_j^*, y_i^*\}, & \text{if } k = i, \\ y_j^* + \min\{x_j - y_j^*, y_i^*\}, & \text{if } k = j. \end{cases}$$

By assumption, $x_j - y_j^*$ and y_i^* are strictly positive. Thus, $\tilde{y} \neq y^*$ and $\tilde{y} \in [0, x]$. Furthermore, $\sum_{k=1}^n y_k^* = \sum_{k=1}^n \tilde{y}_k$ and so we have:

$$a(x, y^*) - a(x, \tilde{y}) = (\beta_i - \beta_j) \min\{x_j - y_j^*, y_i^*\} > 0$$

where we have used that the β_k's are ordered increasingly. This is a contradiction to the minimality of y^*.

- *Let y^* be a minimizer for $a(x, \cdot)$ in $[0, x]$. If there exists j such that $y_j^* = 0$, then $y_i^* = 0$ for every $i = j, \ldots, n$.*
  Using the previous result, if for the sake of contradiction there exists $i > j$ such that $y_i^* > 0$, then $x_j = 0$. However, we can always assume $x_j \neq 0$ for every $j = 1, \ldots, n$, since otherwise y_j would be forced to be zero, and we could consider x and y as vectors in a lower dimension.

Thus, given x, the solution can be of n types:

$y^{(1)} := (y_1, 0, \cdots, 0), \cdots, y^{(j)} := (x_1, \cdots, x_{j-1}, y_j, 0, \cdots, 0), \cdots, y^{(n)} := (x_1, \cdots, x_{n-1}, y_n)$

where, for every $j = 1, \ldots, n$, $y_j^{(j)}$ minimizes in $[0, x_j]$ the mapping

$$z \mapsto -\sum_{i=1}^{j-1} (\beta_n - \beta_i) x_i - (\beta_n - \beta_j) z + \lambda^* \left(z - \sum_{i=j}^n x_i + B(\eta^*) + \frac{\beta_n}{2\lambda^*} \right)^2. \tag{3.4}$$

The function in (3.4) is an upward-facing parabola, and the vertex's z-coordinate is:

$$z_V = \sum_{i=j}^n x_i - B(\eta^*) - \frac{\beta_j}{2\lambda^*}.$$

In particular, $z_V \in [0, x_j]$ if and only if $\sum_{i=j}^n x_i \geq B(\eta^*) + \frac{\beta_j}{2\lambda^*}$ and $\sum_{i=j+1}^n x_i \leq B(\eta^*) + \frac{\beta_j}{2\lambda^*}$. Furthermore, if $z_V < 0$, then the function is increasing in the interval $[0, x_j]$, and the minimum will occur at 0. On the other hand, if $z_V > x_j$, the function is decreasing in the same interval, and the minimum will occur at x_j. We can summarize the above as follows:

$$y_j^{(j)} = \begin{cases} 0, & \text{if } \sum_{i=j}^n x_i < B(\eta^*) + \frac{\beta_j}{2\lambda^*}, \\ \sum_{i=j}^n x_i - B(\eta^*) - \frac{\beta_j}{2\lambda^*}, & \text{if } \sum_{i=j}^n x_i \geq B(\eta^*) + \frac{\beta_j}{2\lambda^*} \quad \text{and} \quad \sum_{i=j+1}^n x_i < B(\eta^*) + \frac{\beta_j}{2\lambda^*}, \\ x_j, & \text{if } \sum_{i=j+1}^n x_i \geq B(\eta^*) + \frac{\beta_j}{2\lambda^*}. \end{cases} \tag{3.5}$$

Let us define j^* as the index defined by

$$j^* = \max \left\{ j \in \{1, \ldots, n\} \mid \sum_{i=j}^{n} x_i \geq B(\eta^*) + \frac{\beta_j}{2\lambda^*} \right\}$$

or $j^* = 1$ if the set on the right hand side is empty. We want to prove that $y^{(j^*)}$ minimizes $a(x, \cdot)$. In particular, we want to show that the function $k \mapsto a(x, y^{(k)})$ is decreasing up to j^* and increasing from j^* onwards, which implies the desired result. Hence, let us analyze the two possible cases:

- $1 \leq k < j^*$: it holds

$$\sum_{i=k+1}^{n} x_i \geq \sum_{i=j^*}^{n} x_i \geq B(\eta^*) + \frac{\beta_{j^*}}{2\lambda^*} \geq B(\eta^*) + \frac{\beta_k}{2\lambda^*}, \tag{3.6}$$

  so we have that

$$y^{(k)} = (x_1, \ldots, x_{k-1}, x_k, 0, \ldots, 0).$$

  At the same time, however, $y^{(k+1)}$ minimizes the function $a(x, \cdot)$ among the vectors of the form $(x_1, \ldots, x_{k-1}, x_k, z, \ldots, 0)$ (with respect to z), and thus it must hold that $a(x, y^{(k)}) \geq a(x, y^{(k+1)})$;

- $j^* \leq k < n - 1$: we have

$$\sum_{i=k+1}^{n} x_i < B(\eta^*) + \frac{\beta_{k+1}}{2\lambda^*}, \tag{3.7}$$

  that leads to the equality:

$$y^{(k+1)} = (x_1, \ldots, x_{k-1}, x_k, 0, 0, \ldots, 0).$$

  However, we know $y^{(k)}$ minimizes the function $a(x, \cdot)$ among the vectors of the form $(x_1, \ldots, x_{k-1}, z, 0, \ldots, 0)$, hence $a(x, y^{(k)}) \leq a(x, y^{(k+1)})$.

Moreover, the uniqueness of the minimum of (3.4) implies that the increase and decrease of $k \mapsto a(x, y^{(k)})$ must be strict, and thus the minimum just found is also unique. In particular, this means that the contract is deterministic with R given by:

$$R(x) = (x_1, \ldots, x_{j^*-1}, y_{j^*}^{(j^*)}, 0, \cdots, 0).$$

Furthermore, the system (3.5) can be rewritten to obtain

$$y_{j^*}^{(j*)} = \min \left(\left(\sum_{i=j^*}^{n} x_i - \frac{\beta_{j^*}}{2\lambda^*} - B(\eta^*) \right)_+ , x_{j^*} \right).$$

Finally, using (3.6) and (3.7) we get:

$$x_k = \min\left(\left(\sum_{i=k}^{n} x_i - \frac{\beta_k}{2\lambda^*} - B(\eta^*)\right)_+, x_k\right), \qquad k = 1, \cdots, j^* - 1;$$

$$0 = \min\left(\left(\sum_{i=k}^{n} x_i - \frac{\beta_k}{2\lambda^*} - B(\eta^*)\right)_+, x_k\right), \qquad k = j^* + 1, \cdots, n.$$

This leads to the final form of R, which in components is:

$$R_k(x) = \min\left(\left(\sum_{i=k}^{n} x_i - \frac{\beta_k}{2\lambda^*} - B(\eta^*)\right)_+, x_k\right), \qquad k = 1, \cdots, n.$$

We note that we have described the optimal contract as the parameters λ^* and $B(\eta^*)$ vary, which have so far been considered fixed. In specific cases, it is possible to optimize the solution as they vary using analytical tools or numerical simulations.

3.2 The Value at Risk Measure as a Constraint

In this second case, we stay in the same setting of the previous example with respect to the functional $\mathcal{P}$, assuming X has a density. However, we change the constraint. To do so, we introduce some definitions regarding quantile risk measures.

Definition 3.2.1: *For any random variable X on a probability space $(\Omega, \mathcal{F}, P)$ and for any constant $\alpha \in (0,1)$, the Value at Risk measure at probability α is defined as*

$$\mathrm{VaR}_\alpha(X) := \min\{v \in \mathbb{R} : P(X > v) \leq \alpha\}.$$

It is indeed a minimum because the cumulative distribution function of X is right-continuous.

Definition 3.2.2: *For any $\eta \in \mathcal{M}$ we define:*

$$\widehat{\mathrm{VaR}}_\alpha(\eta) := \min\left\{v \in \mathbb{R} : \eta\left(\left\{(x,y) \mid \sum_{i=1}^{n}(x_i - y_i) > v\right\}\right) \leq \alpha\right\}.$$

Also in this case, it is a minimum because it can be rewritten in terms of the cumulative distribution function of the probability $(T_S)_\# \eta$, where $T_S(x,y) := \sum_{i=1}^{n}(x_i - y_i)$.

That said, let us fix $\alpha \in (0,1)$ and suppose there exists a constant c such that $0 < c < \mathrm{VaR}_\alpha\left(\sum_{i=1}^{n} X_i\right)$. The constraint in this case is given by:

$$\mathcal{G}(\eta) := \widehat{\mathrm{VaR}}_\alpha(\eta) - c \leq 0.$$

Due to the condition on c, we cannot have full reinsurance and nor no reinsurance contracts are optimal. In this example, it is clear that we cannot proceed as before by attempting to use $\mathcal{M}$ as the convex set $\mathcal{C}$. The nature of $\mathcal{G}$, in fact, does not allow us to treat it as an integral operator over such a large set. For this reason, we fix an optimal treaty η^* and set $v^* := \widehat{\mathrm{VaR}}_\alpha(\eta^*)$. Hence, we define $\mathcal{C}$ as follows:

$$\mathcal{C} := \left\{\eta \in \mathcal{M} \mid \widehat{\mathrm{VaR}}_\alpha(\eta) = v^*\right\}.$$

Now we have the advantage that, for every $\eta \in \mathcal{C}$, it holds that $d\mathcal{G}(\eta^*, \eta - \eta^*) = 0$, and thus we can use:

$$p_{\eta^*}(x,y) := p(x,y) := \sum_{i=1}^{n} \beta_i y_i, \qquad g_{\eta^*}(x,y) := 0.$$

However, we need to verify that $\mathcal{C}$ is convex: let $\eta_1, \eta_2 \in \mathcal{C}$ and $t \in (0,1)$. Thus, it holds:

$$t\eta_1\left(\left\{(x,y) \mid \sum_{i=1}^{n}(x_i - y_i) > v^*\right\}\right) + (1-t)\eta_2\left(\left\{(x,y) \mid \sum_{i=1}^{n}(x_i - y_i) > v^*\right\}\right) \le t\alpha + (1-t)\alpha \le \alpha.$$

Hence, $\widehat{\mathrm{VaR}}_\alpha(t\eta_1 + (1-t)\eta_2) \le v^*$. If there existed $\overline{v} < v^*$ such that $\widehat{\mathrm{VaR}}_\alpha(t\eta_1 + (1-t)\eta_2) = \overline{v}$, then at least one of $\widehat{\mathrm{VaR}}_\alpha(\eta_1)$ and $\widehat{\mathrm{VaR}}_\alpha(\eta_2)$ would have to be less than or equal to $\overline{v}$, contradicting the minimality of v^*. We are therefore in the right conditions to apply Proposition 2.2.5. Let us break down the solution into steps.

Step 1: *We set*

$$D_1 := \left\{(x,y) \in \mathcal{A}_R \mid \sum_{i=1}^{n}(x_i - y_i) < v^*\right\},$$

$$D_2 := \left\{(x,y) \in \mathcal{A}_R \mid \sum_{i=1}^{n}(x_i - y_i) > v^*\right\}.$$

For every $(x,y) \in Supp(\eta^*)$ we denote with $I(x,y)$ the set where we take the minimum in Proposition 2.2.5. If $(x,y) \in D_1$, then

$$\{z \in [0,x] \mid (x,z) \in \overline{D_1}\} \subset I(x,y).$$

If $(x,y) \in D_2$, it holds that $0 \in I(x,y)$.

Proof. Let us start with $(x,y) \in D_1 \setminus \partial\mathcal{A}_R$ and define $T := \{-y < t < x - y \mid (x, y+t) \in D_1\}$, that is, the set such that $(x,y) \in D_1 - (0,t)$. Hence, for $t \in T$, we define

$$\delta := \min\left\{d\left((x,y), \partial\mathcal{A}_R\right), d\left((x,y), \partial\mathcal{A}_R - (0,t)\right), d((x,y), \partial D_1), d((x,y), \partial D_1 - (0,t))\right\}.$$

In this way, $\delta > 0$ (since D_1 is open), and by Lemma 2.2.3, $\eta_{x,y,t,\varepsilon} \in \mathcal{M}$ for every $0 \le \varepsilon < \delta$ (the δ used here is even smaller than the one used in the lemma).

We want to prove $\widehat{\mathrm{VaR}}_\alpha(\eta_{x,y,t,\varepsilon}) = v^*$. By the definition of δ, $B_\varepsilon(x,y) \subset D_1 \cap (D_1 - (0,t))$, hence $B_\varepsilon(x,y) \cap D_2 = B_\varepsilon(x,y) \cap (D_2 - (0,t)) = \emptyset$ and so

$$\eta_{x,y,t,\varepsilon}\left(\left\{(x,y) \mid \sum_{i=1}^n (x_i - y_i) > v^*\right\}\right) = \eta_{x,y,t,\varepsilon}(D_2) = \eta^*(D_2) \le \alpha,$$

which gives us the inequality $\widehat{\mathrm{VaR}}_\alpha(\eta_{x,y,t,\varepsilon}) \le v^*$. On the other hand, let

$$\varepsilon' := \min\{d((x,y),\partial D_1), d((x,y),\partial D_1 - (0,t))\} - \varepsilon,$$

that is positive since $\varepsilon < \delta \le \min\{d((x,y),\partial D_1), d((x,y),\partial D_1 - (0,t))\}$. Furthermore, for every u such that $v^* - \varepsilon' < u < v^*$ we set

$$D_u := \left\{(x,y) \in \mathcal{A}_R \mid \sum_{i=1}^n (x_i - y_i) > u\right\}.$$

As before, it holds that $B_\varepsilon(x,y) \subset D_u^c \cap (D_u^c - (0,t))$. This is because ∂D_1 and ∂D_u^c are defined, respectively, through the hyperplanes $\sum_{i=1}^n (x_i - y_i) = v^*$ and $\sum_{i=1}^n (x_i - y_i) = u$, with $D_u^c \subset D_1$. By construction, the second hyperplane is precisely located between $B_\varepsilon(x,y)$ and the first hyperplane. Therefore, $B_\varepsilon(x,y) \subset D_u^c$ (a similar argument applies to $B_\varepsilon(x,y) \subset D_u^c - (0,t)$). Thus, we obtain:

$$\eta_{x,y,t,\varepsilon}\left(\left\{(x,y) \mid \sum_{i=1}^n (x_i - y_i) > u\right\}\right) = \eta_{x,y,t,\varepsilon}(D_u) = \eta^*(D_u) > \alpha,$$

where the last inequality follows form the minimality of v^*. Hence $\widehat{\mathrm{VaR}}_\alpha(\eta_{x,y,t,\varepsilon}) = v^*$.
We have shown that 0 is a limit point of the set $\{\varepsilon \in [0,\delta) \mid \eta_{x,y,t,\varepsilon} \in \mathcal{C}\}$ if $t \in T$. This leads to the following chain:

$$\begin{aligned} J(x,y) &:= \{z \in (0,x) \mid 0 \text{ is a limit point of } \{\varepsilon \in [0,\delta) \mid \eta_{x,y,z-y,\varepsilon} \in \mathcal{C}\}\} \\ &\supset \{z \in (0,x) \mid z - y \in T\} \\ &= \{z \in (0,x) \mid (x,z) \in D_1\}. \end{aligned}$$

We get what we wanted by recalling that $I(x,y) = \overline{J(x,y)}$.
For the case $(x,y) \in D_1 \cap \partial\mathcal{A}_R$, we can proceed similarly by defining

$$\delta = \min\{d((x,y),\partial\mathcal{A}_R \setminus \{(0,t)\}), d((x,y),\partial D_1 \setminus \{(0,t)\})\},$$

$$\varepsilon' = d((x,y),\partial D_1 \setminus \{(0,t)\}) - \varepsilon.$$

Now, let us consider the case with D_2. As before, start with $(x,y) \in D_2 \setminus \partial\mathcal{A}_R$ and define $T := \{-y < t < x - y \mid (x, y+t) \in D_2\}$. Hence, for $t \in T$, we set:

$$\delta := \min\{d((x,y),\partial\mathcal{A}_R), d((x,y),\partial\mathcal{A}_R - (0,t)), d((x,y),\partial D_2), d((x,y),\partial D_2 - (0,t))\}.$$

In the same manner of the previous case, $\delta > 0$, and by Lemma 2.2.3, $\eta_{x,y,t,\varepsilon} \in \mathcal{M}$ for every $0 \leq \varepsilon < \delta$.
We want to prove that $\widehat{\mathrm{VaR}}_\alpha(\eta_{x,y,t,\varepsilon}) = v^*$. By the definition of δ, $B_\varepsilon(x,y) \subset D_2 \cap (D_2 - (0,t))$, and so:

$$\eta_{x,y,t,\varepsilon}\left(D_2\right) = \eta^*\left(D_2\right) - \eta^*(B_\varepsilon(x,y)) + \eta^*(B_\varepsilon(x,y)) \leq \alpha,$$

which leads to the inequality $\widehat{\mathrm{VaR}}_\alpha(\eta_{x,y,t,\varepsilon}) \leq v^*$. On the other hand, suppose by contradiction that $\widehat{\mathrm{VaR}}_\alpha(\eta_{x,y,t,\varepsilon}) = \tilde{v} < v^*$. Then the set:

$$D_{\tilde{v}} := \left\{(x,y) \in \mathcal{A}_R \mid \sum_{i=1}^n (x_i - y_i) > \tilde{v}\right\}$$

is such that $D_2 \subset D_{\tilde{v}}$. In particular, $B_\varepsilon(x,y) \subset D_{\tilde{v}} \cap (D_{\tilde{v}} - (0,t))$, and thus:

$$\alpha \geq \eta_{x,y,t,\varepsilon}\left(D_{\tilde{v}}\right) = \eta^*\left(D_{\tilde{v}}\right) > \alpha.$$

Hence, $\widehat{\mathrm{VaR}}_\alpha(\eta_{x,y,t,\varepsilon}) = v^*$. Then we obtain that $J(x,y)$ satisfies:

$$J(x,y) \supset \{z \in (0,x) \mid z - y \in T\} = \{z \in (0,x) \mid (x,z) \in D_2\}.$$

This implies that $\{z \in [0,x] \mid (x,z) \in \overline{D_2}\} \subset I(x,y)$. In particular, since $(x,y) \in D_2$, we have $0 \in I(x,y)$. We can conclude for the case $(x,y) \in D_2 \cap \mathcal{A}_R$ in the same way as $D_1 \cap \mathcal{A}_R$. □

We observe that we can partition $\mathcal{A}_R$ into four disjoint sets:

$$\mathcal{A}_R = V_1 \cup V_2 \cup D_2 \cup D_3,$$

where V_1 and V_2 form a partition of D_1 and are defined as

$$V_1 := \left\{(x,y) \in \mathcal{A}_R \mid \sum_{i=1}^n (x_i - y_i) < v^*, \sum_{i=1}^n x_i \leq v^*\right\},$$

$$V_2 := \left\{(x,y) \in \mathcal{A}_R \mid \sum_{i=1}^n (x_i - y_i) < v^*, \sum_{i=1}^n x_i > v^*\right\},$$

and D_3 is

$$D_3 := \left\{(x,y) \in \mathcal{A}_R \mid \sum_{i=1}^n (x_i - y_i) = v^*\right\}.$$

Step 2: For every $(x,y) \in \mathrm{supp}(\eta^*) \cap (V_1 \cup D_2)$, then $y = 0$.

Proof. Let us start with $(x, y) \in \mathrm{Supp}(\eta^*) \cap V_1$. By Proposition 2.2.5, there exists $r^* > 0$ ($\mathcal{G}$ is constant on $\mathcal{C}$) such that:

$$y \in \operatorname{argmin}\{r^* p_{\eta^*}(x, z) \mid z \in I(x, y)\} = \operatorname{argmin}\left\{\sum_{i=1}^{n} \beta_i z_i \mid z \in I(x, y)\right\}.$$

However, since $\sum_{i=1}^n x_i \leq v^*$, from the first step we have

$$0 \in \{z \in [0, x] \mid (x, z) \in \overline{D_1}\} \subset I(x, y).$$

and so it must hold $y = 0$. Similarly, for $(x, y) \in \mathrm{Supp}(\eta^*) \cap D_2$. $\square$

To determine the behavior of the support points in V_2 and D_3, we first need the following result.

Step 3: For fixed $x \in \mathbb{R}^n_+$ and $v > 0$ such that $\sum_{i=1}^n x_i > v$, it holds that the minimum point of $z \mapsto \sum_{i=1}^n \beta_i z_i$ on

$$\left\{z \in [0, x] \mid \sum_{i=1}^{n} (x_i - z_i) \leq v\right\}$$

is unique and is given by

$$y(x, v) := \left(x_1, \ldots, x_{i_0 - 1}, \sum_{i=i_0}^{n} x_i - v, 0, \ldots, 0\right),$$

where i_0 is the index such that $\sum_{j=i_0+1}^n x_j \leq v$ and $\sum_{j=i_0}^n x_j > v$.

Proof. In the case $n = 1$, the result follows directly. Let $n \geq 2$. First, we observe that any minimum point $\tilde{y}$ must satisfy $\tilde{y}_j = x_j$ for every $j < i_0$. Otherwise, let j_0 be the first index where this does not hold or $j_0 = 1$ in case $\tilde{y}_j < x_j$ for every $j \in \{1, \ldots, n\}$. Then, there must exist an index $k > j_0$ such that $\tilde{y}_k \neq 0$, otherwise the following chain of inequalities would hold:

$$\begin{aligned}
x_{j_0} > \tilde{y}_{j_0} &= \sum_{i=1}^{j_0} \tilde{y}_i - \sum_{i=1}^{j_0 - 1} \tilde{y}_i = \sum_{i=1}^{n} \tilde{y}_i - \sum_{i=1}^{j_0 - 1} \tilde{y}_i \geq \sum_{i=1}^{n} x_i - v - \sum_{i=1}^{j_0 - 1} \tilde{y}_i \\
&= \sum_{i=j_0}^{n} x_i - v = x_{j_0} + \sum_{i=j_0+1}^{n} x_i - v \geq x_{j_0}.
\end{aligned}$$

So, let $k_0 > j_0$ be an index such that $\tilde{y}_{k_0} > 0$. Hence, we can define $\hat{y}$ such that $\hat{y}_i = \tilde{y}_i$ if $i \neq j_0, k_0$, $\hat{y}_{j_0} = \tilde{y}_{j_0} + \min\{x_{j_0} - \tilde{y}_{j_0}, \tilde{y}_{k_0}\}$ and $\hat{y}_{k_0} = \tilde{y}_{k_0} - \min\{x_{j_0} - \tilde{y}_{j_0}, \tilde{y}_{k_0}\}$. In particular, $\hat{y} \in [0, x]$, $\hat{y} \neq \tilde{y}$ and they hold:

$$\sum_{i=1}^{n} \tilde{y}_i = \sum_{i=1}^{n} \hat{y}_i;$$

$$\sum_{i=1}^{n} \beta_i \hat{y}_i - \sum_{i=1}^{n} \beta_i \tilde{y}_i = (\beta_{j_0} - \beta_{k_0}) \min\{x_{j_0} - \tilde{y}_{j_0}, \tilde{y}_{k_0}\} < 0.$$

This leads to a contradiction due to the optimality of $\tilde{y}$. Thus, we have $\tilde{y}_j = x_j$ for every $j < i_0$.

Now, take $y(x,v)$ as in the statement, by definition of i_0 $y^v \in [0,x]$, and satisfy the unilateral constraint (for the point $(x, y_i^v(x))$, the equality $\sum_{i=1}^n (x_i - y(x,v)) = v$ holds). We also have:

$$\sum_{i=1}^{n} \beta_i y_i^v(x) = \sum_{i=1}^{i_0-1} \beta_i x_i + \beta_{i_0} \left(\sum_{i=i_0}^{n} x_i - v \right) \leq \sum_{i=1}^{i_0-1} \beta_i \tilde{y}_i + \beta_{i_0} \sum_{i=i_0}^{n} \tilde{y}_i \leq \sum_{i=1}^{n} \beta_i \tilde{y}_i,$$

where the last equality becomes strict if there exists $i > i_0$ such that $\tilde{y}_i \neq 0$. It remains to check the index i_0: from the inequality of the constraint, we obtain $\tilde{y}_{i_0} \geq \tilde{y}_{i_0}^v$, but if the strict inequality held, it would contradict the minimality assumption of $\tilde{y}$. Thus, it must be $\tilde{y} = y(x,v)$. □

We will use the notation $y^*(x) := y(x, v^*)$.

Step 4: If $(x,y) \in (V_2 \cup (D_3 - \mathbb{R}_+^n \times \{0\})) \cap \operatorname{supp}(\eta^*)$, then $(x,y) \in D_3$.

Proof. By definition, $\sum_{i=1}^n x_i > v^*$. Suppose, for the sake of contradiction, that $(x,y) \in V_2 \subset D_1$. Then, we would have

$$y^*(x) \in \left\{ z \in [0,x] \mid \sum_{i=1}^{n} (x_i - z_i) = v^* \right\} \subset \left\{ z \in [0,x] \mid (x,z) \in \overline{D_1} \right\} \subset I(x,y).$$

Thus, on one hand, since y minimizes $z \mapsto \sum_{i=1}^n \beta_i z_i$ in $I(x,y)$, it must hold that

$$\sum_{i=1}^{n} \beta_i y_i \leq \sum_{i=1}^{n} \beta_i y_i^*(x).$$

On the other hand, by the previous step, $y^*(x)$ is the unique minimizer of $z \mapsto \sum_{i=1}^n \beta_i z_i$ over

$$\left\{ z \in [0,x] \mid \sum_{i=1}^{n} (x_i - z_i) \leq v^* \right\}.$$

Hence,

$$\sum_{i=1}^{n} \beta_i y_i > \sum_{i=1}^{n} \beta_i y_i^*(x),$$

where the strict inequality arises from the fact that y and $y^*(x)$ cannot be the same by the definition of the latter. □

At this point, we aim to understand which points must belong to the support of the measure in D_3.

Step 5: For every $(\overline{x}, \overline{y}) \in (D_3 - \mathbb{R}^n_+ \times \{0\}) \cap \operatorname{supp}(\eta^*)$, *it holds that* $\overline{y} = y^*(\overline{x})$.

Proof. Let us argue by contradiction and suppose that $\overline{y} \neq y^*(\overline{x})$. The idea is the following: we construct a better competitor for η^* by moving the points of a suitable neighborhood of $(\overline{x}, \overline{y})$ to "less costly" points with the same first component (so as not to change the projection onto the first component of the measure), along planes parallel to $\sum_{i=1}^n (x_i - y_i) = v^*$ (thus preserving the VaR_α). Let V be a sufficiently small neighborhood of $(\overline{x}, \overline{y})$ in $\mathcal{A}_R$ such that $(x, y) \neq (x, y(x, \sum_{i=1}^n x_i - y_i))$ for every $(x, y) \in V$ (this exists due to the continuity of the minimum in Step 3 with respect to the parameters v and x). We define the function $L : \mathbb{R}^n_+ \times \mathbb{R}^n_+ \longrightarrow \mathbb{R}^n_+ \times \mathbb{R}^n_+$ as follows:

$$L(x, y) = \begin{cases} (x, y\left(x, \sum_{i=1}^n x_i - y_i\right)) & \text{if } (x, y) \in V, \\ (x, y), & \text{otherwise.} \end{cases}$$

Let $\tilde{\eta} := L_\# \eta^*$, then $\tilde{\eta}(\mathcal{A}_R) = 1$ and $(\pi_1)_\# \tilde{\eta} = (\pi_1 \circ L)_\# \tilde{\eta} = (\pi_1)_\# \tilde{\eta} = \mu$, so $\tilde{\eta} \in \mathcal{M}$. As defined in Step 1, for every $u > 0$

$$D_u := \left\{ (x, y) \in \mathcal{A}_R \mid \sum_{i=1}^n (x_i - y_i) > u \right\},$$

we have $L^{-1}(D_u) = D_u$, and thus $\mathrm{VaR}_\alpha(\tilde{\eta}) = \mathrm{VaR}_\alpha(\eta^*) = v^*$. Finally, by Step 3, $p(x, y(x, \sum_{i=1}^n x_i - y_i)) < p(x, y)$ for every $(x, y) \in V$, and since V has positive measure, it must hold that:

$$\begin{aligned} \mathcal{P}(\tilde{\eta}) &= \int_{\mathcal{A}_R \setminus V} p(x, y)\, \tilde{\eta}(dx, dy) + \int_V p(x, y)\, \tilde{\eta}(dx, dy) \\ &= \int_{\mathcal{A}_R \setminus V} p \circ L(x, y)\, \eta^*(dx, dy) + \int_V p \circ L(x, y)\, \eta^*(dx, dy) \\ &= \int_{\mathcal{A}_R \setminus V} p(x, y)\, \eta^*(dx, dy) + \int_V p\left(x, y\left(x, \sum_{i=1}^n x_i - y_i \right)\right) \eta^*(dx, dy) \\ &< \int_{\mathcal{A}_R \setminus V} p(x, y)\, \eta^*(dx, dy) + \int_V p(x, y)\, \eta^*(dx, dy) = \mathcal{P}(\eta^*). \end{aligned}$$

□

We have reached the point where, given $(x, y) \in \operatorname{Supp}(\eta^*)$, if $\sum_{i=1}^n x_i \leq v^*$, it can be $(x, y) \in V_1$ or $(x, y) \in D_3 \cap (\mathbb{R}^n_+ \times \{0\})$. In both cases, the only possibility is $y = 0$. If $\sum_{i=1}^n x_i > v^*$, we have two possibilities: either $y = y^*(x)$ (if $(x, y) \in (D_3 - \mathbb{R}^n_+ \times \{0\}) \cup V_2$) or $y = 0$ (if $(x, y) \in D_2$); in the last step we want to show that they cannot be both true.

Step 6: If we denote $k := \mu(\{x \in \mathbb{R}^n_+ \mid \sum_{i=1}^n x_i \leq v^*\})$, then there exist $d > 0$ such that $\mu(E) = 1 - \alpha - k$ with

$$E := \left\{ x \in \mathbb{R}^n_+ \mid \sum_{i=1}^n x_i > v^*,\, p(x, y^*(x)) \leq d \right\}.$$

Furthermore, for every $(x, y) \in \operatorname{Supp}(\eta^*)$ with $\sum_{i=1}^n x_i > v^*$, it holds that $y = y^*(x)$ if $x \in E$ or $y = 0$ if $x \notin E$.

Proof. Let us first verify that d is well-defined. First of all, $1 - \alpha - k > 0$, because if by contradiction $k \geq 1 - \alpha$, then the no reinsurance contract would be optimal. Indeed, we would be forced to assign all the remaining probability to the points $(x, 0)$. Otherwise, due to the absolute continuity of μ, we would find a $v < v^*$ such that

$$\eta^* \left(\left\{ (x, y) \in \mathcal{A}_R \mid \sum_{i=1}^n (x_i - y_i) > v \right\} \right) \leq \alpha,$$

contradicting the constraint $\mathrm{VaR}_\alpha(\eta^*) = v^*$. Furthermore, if we consider E as a function of d, the map $d \mapsto \mu(E_d)$ is continuous. Indeed, for every $d > 0$, we define I_d as the set:

$$I_d := \left\{ x \in \mathbb{R}^n_+ \mid \sum_{i=1}^n x_i > v^*,\, p(x, y^*(x)) = d \right\}.$$

By the definition of $p(x, y^*(x))$, the set I_d is negligible with respect to the Lebesgue measure, and hence it is also negligible with respect to μ due to the absolute continuity of the latter. Now, consider an increasing sequence $\{d_n\}_{n=1}^{+\infty}$ such that $d_n \to d$ as $n \to +\infty$. Clearly, we have $E_{d_i} \subset E_{d_{i+1}}$ for every i, and thus we obtain:

$$\lim_{n \to +\infty} \mu(E_{d_n}) = \mu \left(\bigcup_{i=1}^{\infty} E_{d_i} \right) = \mu(E_d) - \mu(I_d) = \mu(E_d).$$

On the other hand, if we take a decreasing sequence $\{d_n\}_{n=1}^{+\infty}$ such that $d_n \to d$ as $n \to +\infty$, then $E_{d_{i+1}} \subset E_{d_i}$ for every i, and hence we have:

$$\lim_{n \to +\infty} \mu(E_{d_n}) = \mu \left(\bigcap_{i=1}^{\infty} E_{d_n} \right) = \mu(E_d).$$

Since $\lim_{d \to 0^+} \mu(E_d) = 0$ and $\lim_{d \to +\infty} \mu(E_d) = 1$, there exists a d as stated in the claim. Note that $d > 0$, otherwise it would hold $k = 1 - \alpha$.
From $\mathrm{VaR}_\alpha(\eta^*) = v^*$, we get:

$$\eta^* \left(\left\{ (x, 0) \in \mathcal{A}_R \mid \sum_{i=1}^n x_i > v^* \right\} \right) \leq \alpha, \quad \eta^* \left(\left\{ (x, y^*(x)) \in \mathcal{A}_R \mid \sum_{i=1}^n x_i > v^* \right\} \right) \geq 1 - \alpha - k. \tag{3.8}$$

Let us show that the optimal contract η^* must satisfy the equalities in (3.8). Define $A_{y^*(x)} := \{(x, y^*(x)) \in \mathcal{A}_R \mid \sum_{i=1}^n x_i > v^*\}$, and assume by contradiction that $\eta^*(A_{y^*(x)}) > 1-k-\alpha$. Let $(\overline{x}, y^*(\overline{x})) \in \mathrm{Supp}(\eta^*)$ and U be a neighborhood of $\overline{x}$ in $\{x \in \mathbb{R}^n_+ \mid \sum_{i=1}^n x_i > v^*\}$ such that $\mu(U) < \eta^*(A_{y^*(x)}) - (1-k-\alpha)$ (such a choice is always possible due to the absolute continuity of μ). Consider the measure $\overline{\eta}$ obtained as the push-forward of η^* defined by the map that sends $(x, y^*(x)) \mapsto (x, 0)$ if $x \in U$, and leaves all other points unchanged. For analogous reasons to the previous step, $\overline{\eta} \in \mathcal{M}$ and $\mathcal{P}(\overline{\eta}) < \mathcal{P}(\eta^*)$. Regarding the constraint:

$$\begin{aligned}\overline{\eta}(A_{y^*(x)}) &= \eta^*(A_{y^*(x)}) - \eta^*(\{(x, y^*(x)) \mid x \in U\}) \\ &\geq \eta^*(A_{y^*(x)}) - (\eta^*(A_{y^*(x)}) - (1-k-\alpha)) = 1-k-\alpha,\end{aligned}$$

where the inequality follows from the fact that

$$\mu(U) = \eta^*(\{(x,0) \mid x \in U\}) + \eta^*(\{(x, y^*(x)) \mid x \in U\}).$$

Thus, $\mathrm{VaR}_\alpha(\overline{\eta}) \leq v^*$. Equality holds because for every $v < v^*$ we have

$$\overline{\eta}\left(\left\{(x,0) \in \mathcal{A}_R \mid \sum_{i=1}^n x_i < v\right\}\right) = \eta^*\left(\left\{(x,0) \in \mathcal{A}_R \mid \sum_{i=1}^n x_i < v\right\}\right) = \mu\left(\left\{x \in \mathbb{R}^n_+ \mid \sum_{i=1}^n x_i < v\right\}\right),$$

hence if $\mathrm{VaR}_\alpha(\overline{\eta}) < v^*$, then $\mathrm{VaR}_\alpha(\eta^*)$ would also be strictly less than v^*. Therefore, $\overline{\eta}$ is a better competitor than η^*, and for this reason, the equalities in (3.8) must hold.
At this point, let us define the sets:

$$\begin{aligned}A_{0,>} &:= \left\{(x,0) \in \mathcal{A}_R \mid \sum_{i=1}^n x_i > v^*,\, p(x, y^*(x)) > d\right\}, \\ A_{0,\leq} &:= \left\{(x,0) \in \mathcal{A}_R \mid \sum_{i=1}^n x_i > v^*,\, p(x, y^*(x)) \leq d\right\}, \\ A_{y^*(x),>} &:= \left\{(x, y^*(x)) \in \mathcal{A}_R \mid \sum_{i=1}^n x_i > v^*,\, p(x, y^*(x)) > d\right\}, \\ A_{y^*(x),\leq} &:= \left\{(x, y^*(x)) \in \mathcal{A}_R \mid \sum_{i=1}^n x_i > v^*,\, p(x, y^*(x)) \leq d\right\}.\end{aligned}$$

Rewriting what we just obtained, we have:

$$\eta^*(A_{0,>}) + \eta^*(A_{0,\leq}) = \alpha, \quad \eta^*(A_{y^*(x),>}) + \eta^*(A_{y^*(x),\leq}) = 1-k-\alpha.$$

At the same time, it holds

$$\eta^*(A_{0,>}) + \eta^*(A_{y^*(x),>}) = \mu\left(\left\{x \in \mathbb{R}^n_+ \mid \sum_{i=1}^n x_i > v^*,\, p(x, y^*(x)) > d\right\}\right) = \alpha.$$

This leads to the equality $\eta^*(A_{0,\le}) = \eta^*(A_{y^*(x),>})$. If $\eta^*(A_{0,\le}) = \eta^*(A_{y^*(x),>}) = 0$, so η^* is the deterministic contract that we have described in the statement. Let us assume by contradiction that $\eta^*(A_{0,\le}) = \eta^*(A_{y^*(x),>}) > 0$. We thus construct the deterministic contract starting from η^*, i.e. $\eta_{\text{det}} := L_\# \eta^*$ with

$$L(x,y) = \begin{cases} (x, y^*(x)) & \text{if } (x,y) \in A_{0,\le}, \\ (x,0) & \text{if } (x,y) \in A_{y^*(x),>}, \\ (x,y) & \text{otherwise.} \end{cases}$$

Clearly, $\eta_{\text{det}} \in \mathcal{M}$. Also, we have:

$$\begin{aligned} \eta_{\text{det}}\left(\left\{(x,y) \in \mathcal{A}_R \mid \sum_{i=1}^n (x_i - y_i) > v^*\right\}\right) &= \eta_{\text{det}}\left(\left\{(x,0) \in \mathcal{A}_R \mid \sum_{i=1}^n x_i > v^*\right\}\right) \\ &= \eta^*\left(L^{-1}\left(\left\{(x,0) \in \mathcal{A}_R \mid \sum_{i=1}^n x_i > v^*\right\}\right)\right) \\ &= \eta^*(A_{0,>} \cup A_{y^*(x),>}) \\ &= \eta^*(A_{0,>}) + \eta^*(A_{y^*(x),>}) = \alpha. \end{aligned}$$

As for $\overline{\eta}$ seen earlier, this is sufficient to guarantee that $\mathrm{VaR}_\alpha(\eta_{\text{det}}) = v^*$. Finally, it holds that

$$\begin{aligned} \mathcal{P}(\eta^*) - \mathcal{P}(\eta_{\text{det}}) &= \int_{\mathcal{A}_R} p(x,y)\, d\eta^* - \int_{\mathcal{A}_R} p \circ L(x,y)\, d\eta^* \\ &= \int_{(A_{y^*(x),>}) \cup (A_{y^*(x),\le})} p(x,y)\, d\eta^* - \int_{(A_{0,\le}) \cup (A_{y^*(x),\le})} p(x,y)\, d\eta^* \\ &= \int_{A_{y^*(x),>}} p(x,y)\, d\eta^* - \int_{A_{0,\le}} p(x,y)\, d\eta^* \\ &> \eta^*(A_{y^*(x),>})\, d - \eta^*(A_{0,\le})\, d = 0. \end{aligned}$$

□

Thus, to summarize, even in this case the optimal contract is deterministic and is defined as:

$$R(x) := \left(x_1, \ldots, x_{i_0 - 1}, \sum_{i=i_0}^n x_i - v^*, 0, \ldots, 0\right),$$

if $\sum_{i=1}^{i_0-1} \beta_i x_i + \beta_{i_0} \sum_{i=i_0}^n x_i - \beta_{i_0} v^* \le d$, $\sum_{j=i_0+1}^n x_j \le v^*$ and $\sum_{j=i_0}^n x_j > v^*$; otherwise, $R(x) = 0$.
As in the previous case study, in specific cases, optimizing as v^* varies becomes analytically solvable.

3.3 An Optimal Transport Case with an Expected Value Constraint

Let us now analyze a first example where we will use the perspective of optimal transport. Let $n = 1$ and let X have a continuous distribution with finite variation. We then define

$$\mathcal{P}(\eta) := \widehat{\mathrm{Var}}(T_{\#}\eta) := \int_{\mathbb{R}} x^2\, T_{\#}\eta(dx) - \left(\int_{\mathbb{R}} x\, T_{\#}\eta(dx)\right)^2$$

and $\mathcal{S} := \{\eta \in \mathcal{M} \mid \int_{\mathbb{R}} y\,(\pi_2)_{\#}\eta(dy) = c\}$ with $0 \le c \le \mathbb{E}[X]$. First, we observe that $\mathcal{S}$ is closed, which guarantees the existence of an optimal treaty. Let us consider a sequence $\{\eta_k\}_{k\in\mathbb{N}} \subset \mathcal{S}$ that converges weakly to $\tilde{\eta} \in \mathcal{M}$. Due to the finiteness of the variance of X, there exists a constant $K < +\infty$ such that the following holds:

$$K > \int_{\mathbb{R}} x^2\,\mu(dx) = \int_{\mathbb{R}\times\mathbb{R}} x^2\,\tilde{\eta}(dx,dy) \ge \int_{\mathbb{R}\times\mathbb{R}} y^2\,\tilde{\eta}(dx,dy) = \int_{\mathbb{R}\times\mathbb{R}} y^2\,(\pi_2)_{\#}\tilde{\eta}(dy).$$

From the inequality $y \le 1 + y^2$, we have:

$$\int_{\mathbb{R}} y\,(\pi_2)_{\#}\tilde{\eta}(dy) \le 1 + K.$$

This result also holds for every η_k. Using an approximating sequence f_n of the identity, with f_n continuous and bounded, we can apply the dominated convergence theorem and the definition of weak convergence to obtain:

$$\begin{aligned}\int_{\mathbb{R}} y\,(\pi_2)_{\#}\tilde{\eta}(dy) &= \lim_{n\to+\infty}\int_{\mathbb{R}} f_n(y)\,(\pi_2)_{\#}\tilde{\eta}(dy) = \lim_{n\to+\infty}\lim_{k\to+\infty}\int_{\mathbb{R}} f_n(y)\,(\pi_2)_{\#}\eta_k(dy)\\ &= \lim_{k\to+\infty}\int_{\mathbb{R}} y\,(\pi_2)_{\#}\eta_k(dy) = \lim_{k\to+\infty} c = c.\end{aligned}$$

Thus, the set $\mathcal{S}$ is closed. Furthermore, this set is clearly convex, and thus Assumption 2.3.1 is automatically satisfied. Also, by definition, $\Pi(\mu,\nu)\cap\mathcal{M} \subset \mathcal{S}$ for every $\nu \in \pi_2(\mathcal{S})$. Rewriting the functional $\mathcal{P}$ as

$$\mathcal{P}(\eta) = \int_{\mathbb{R}\times\mathbb{R}} (x-y)^2\,\eta(dx,dy) - \left(\int_{\mathbb{R}\times\mathbb{R}} (x-y)\,\eta(dx,dy)\right)^2 = \int_{\mathbb{R}\times\mathbb{R}} (x-y)^2\,\eta(dx,dy) - (\overline{\mu}-c)^2,$$

then it is evident that the assumptions in Setting 2.3.4 are satisfied, and thus we can use Proposition 2.3.6. Hence, it follows:

$$\begin{aligned}
\min_{\eta\in\mathcal{S}} \mathcal{P}(\eta) &= \min_{\eta\in\mathcal{S}} \int_{\mathbb{R}\times\mathbb{R}} (x-y)^2\,\eta(dx,dy) - (\overline{\mu}-c)^2 \\
&= \min_{\eta\in\mathcal{S}} \int_{\mathbb{R}\times\mathbb{R}} (x-y)^2\,\eta(dx,dy) \\
&= \min_{\nu\in\pi_2(\mathcal{S})} \int_{\mathbb{R}\times\mathbb{R}} (x-y)^2\,(\mathrm{id}, g_\nu)_{\#}\mu(dx,dy) \\
&= \min_{\nu\in\pi_2(\mathcal{S})} \int_{\mathbb{R}} (x-g_\nu(x))^2\,\mu(dx),
\end{aligned}$$

where we recall that g_ν is defined as

$$g_\nu(x) := F_\nu^{[-1]} \circ F_\mu(x).$$

Thus, the solution to the optimization problem will be the function $g_{\nu^*} \in L^2(\mathbb{R},\mu)$ that minimizes the distance to the identity among the functions $h \in L^2(\mathbb{R},\mu)$ that are non-decreasing, satisfy $0 \le h(x) \le x$ for every $x \ge 0$, and fulfill the integral constraint $\int_{\mathbb{R}} h(x)\mu(dx) = c$. These last two conditions form a set C that is convex and closed. Therefore, we know that the solution to our problem is given by the projection of the identity, denoted Pr(id), which is characterized by the property:

$$\langle \mathrm{id} - \mathrm{Pr(id)}, h - \mathrm{Pr(id)}\rangle_{L^2(\mathbb{R},\mu)} \le 0 \quad \text{for every} \quad h \in C.$$

Thus, we choose $a \ge 0$ such that

$$\int_{\mathbb{R}} (x-a)_+\,\mu(dx) = c,$$

which is always possible given the initial conditions on c. Now, we verify that the function $g(x) = (x-a)_+$ satisfies the previous property and is therefore Pr(id) and, since it is non-decreasing, it is also the solution to our problem. In fact, we get:

$$\begin{aligned}
\langle \mathrm{id} - g, h - g\rangle_{L^2(\mathbb{R},\mu)} &= \int_{\mathbb{R}} (x-(x-a)_+)(h(x)-(x-a)_+)\,\mu(dx) \\
&= \int_{x<a} x\,h(x)\,\mu(dx) + \int_{x\ge a} a\,(h(x)-(x-a))\,\mu(dx) \\
&= \int_{x<a} xh(x)\,\mu(dx) + a\int_{x\ge a} h(x)\,\mu(dx) - a\int_{x\ge a} (x-a)\,\mu(dx) \\
&= \int_{x<a} x\,h(x)\,\mu(dx) + ac - a\int_{x<a} h(x)\,\mu(dx) - a\int_{\mathbb{R}} (x-a)_+\,\mu(dx) \\
&= \int_{x<a} (x-a)h(x)\,\mu(dx) + ac - ac \le 0.
\end{aligned}$$

Hence, the reinsurance contract is deterministic and is given by a stop-loss contract:

$$R(x) = (x - a)_+.$$

Numerical Example

Suppose that X follows a shifted Pareto distribution with a density given by

$$f_X(x) = 324(x + 3)^{-5}, \quad x \geq 0.$$

In particular, $\mathbb{E}[X] = 1$ and $\mathrm{Var}(X) = 2$, so by choosing $c = 0.5$, the initial hypotheses are satisfied. We impose the condition to find a:

$$\begin{aligned} 0.5 = \mathbb{E}[(X - a)_+] &= \int_a^{+\infty} 324(x - a)(x + 3)^{-5}\, dx \\ &= 324 \int_{a+3}^{+\infty} t^{-4} - (a + 3)t^{-5}\, dt = 27(a + 3)^{-3}. \end{aligned}$$

Thus, we obtain $a \approx 0.77976$. In general, the random variable $R = (X - a)_+$ will concentrate a mass equal to $F_X(a)$ at the point $R = 0$, while for $r > 0$ it must hold that

$$F_R(r) = F_X(r + a).$$

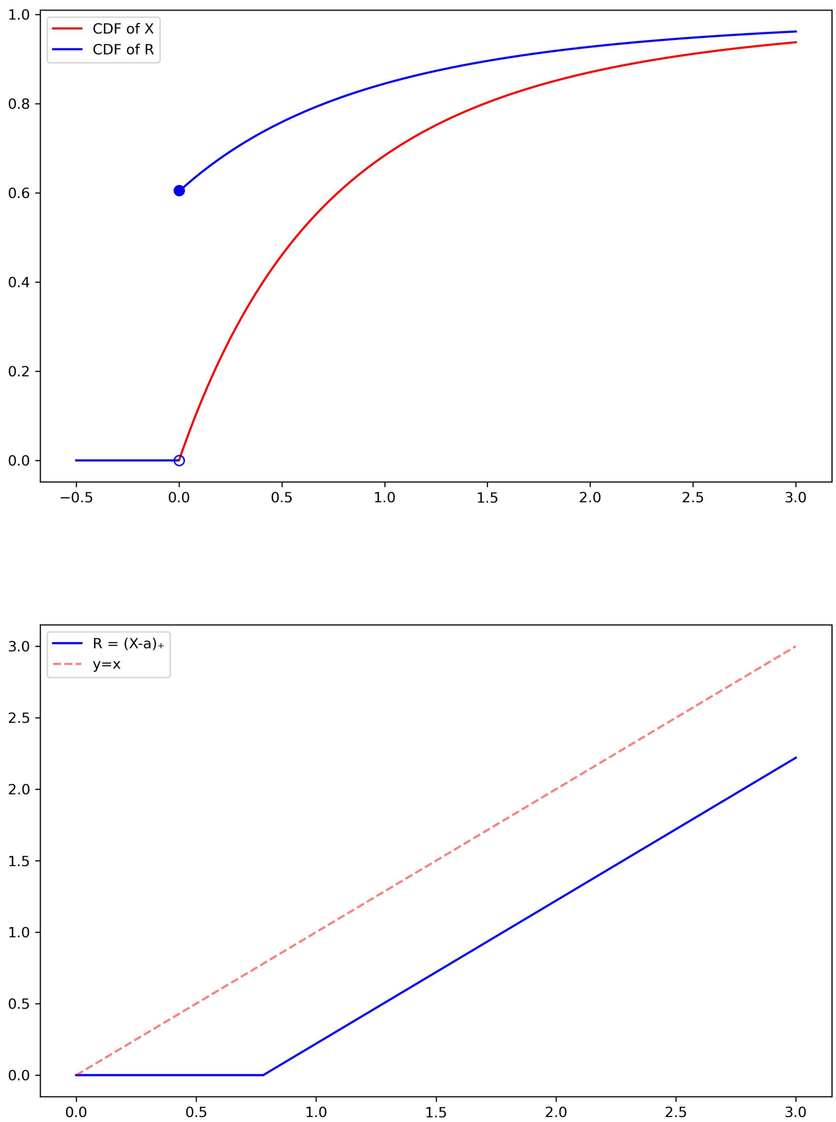


Figure 3.1: Cumulative distribution functions of X and R (top) and the (X, R) graph (bottom).

3.4 An Optimal Transport Example with a Variance Constraint

Compared to Section 3.3, we use the same framework changing only the set $\mathcal{S}$, defining it as

$$\mathcal{S} := \left\{ \eta \in \mathcal{M} \mid \widehat{\mathrm{Var}}((\pi_2)_\# \eta) = c \right\}$$

where c is a constant such that $0 \le c \le \mathrm{Var}(X)$. Now, the set of constraints is not convex, so it is no longer straightforward to say that Assumption 2.3.1 is satisfied. It is enough to observe that $\mathcal{P}$ is concave. Indeed, if it were so, for every η^* optimal, $\eta \in \mathcal{S}$, and $t \in [0,1]$, we would have:

$$\mathcal{P}((1-t)\eta^* + t\eta) \ge (1-t)\mathcal{P}(\eta^*) + t\mathcal{P}(\eta) = \mathcal{P}(\eta^*) + t(\mathcal{P}(\eta) - \mathcal{P}(\eta^*)) \ge \mathcal{P}(\eta^*).$$

Hence, fixed η_1, η_2 in $\mathcal{M}$ and $t \in [0,1]$, it holds:

$$\begin{aligned}
\mathcal{P}((1-t)\eta_1 + t\eta_2) &= (1-t)\int_{\mathbb{R}\times\mathbb{R}} (x-y)^2 d\eta_1 + t\int_{\mathbb{R}\times\mathbb{R}} (x-y)^2 d\eta_2 \\
&\quad - \left((1-t)\int_{\mathbb{R}\times\mathbb{R}} (x-y) d\eta_1 + t\int_{\mathbb{R}\times\mathbb{R}} (x-y) d\eta_2 \right)^2 \\
&\ge (1-t)\int_{\mathbb{R}\times\mathbb{R}} (x-y)^2 d\eta_1 + t\int_{\mathbb{R}\times\mathbb{R}} (x-y)^2 d\eta_2 \\
&\quad - \left((1-t)\left(\int_{\mathbb{R}\times\mathbb{R}} (x-y) d\eta_1\right)^2 + t\left(\int_{\mathbb{R}\times\mathbb{R}} (x-y) d\eta_2\right)^2 \right) \\
&= (1-t)\mathcal{P}(\eta_1) + t\mathcal{P}(\eta_2).
\end{aligned}$$

Furthermore, note that setting the function $p_\eta := (x-y)^2 - 2(x-y)\overline{T_\# \eta}$ ensures we are within the assumptions of Setting 2.3.4, in fact:

$$\begin{aligned}
d\mathcal{P}(\eta, \theta - \eta) &= \int_{\mathbb{R}\times\mathbb{R}} (x-y)^2 \, d(\theta - \eta) - 2\int_{\mathbb{R}\times\mathbb{R}} (x-y) \, d\eta \int_{\mathbb{R}\times\mathbb{R}} (x-y) \, d(\theta - \eta) \\
&= \int_{\mathbb{R}\times\mathbb{R}} (x-y)^2 - 2(x-y)\overline{T_\# \eta} \, d(\theta - \eta).
\end{aligned}$$

We can therefore apply Proposition 2.3.6, according to which the optimal coupling will be of the form $(\mathrm{Id}, g_\nu)_\# \mu$, with $g_\nu(x) := F_\nu^{[-1]} \circ F_\mu(x)$. The problem then becomes minimizing

$$\mathcal{F}(\nu) = \int_{\mathbb{R}} (x - g_\nu(x))^2 \, \mu(dx) - \left(\int_{\mathbb{R}} (x - g_\nu(x)) \, \mu(dx) \right)^2$$

with respect to $\nu \in \pi_2(\mathcal{S})$. By performing the change of variable $y = F_\mu(x)$ and using the fact that $F_\mu^{[-1]}(F_\mu(x)) = x$ μ-a.s., we can rewrite $\mathcal{F}$ as:

$$\mathcal{F}(\nu) = \int_0^1 (F_\mu^{[-1]}(x) - F_\nu^{[-1]}(x))^2 \, dx - \left(\int_0^1 (F_\mu^{[-1]}(x) - F_\nu^{[-1]}(x)) \, dx \right)^2.$$

The constraints then become $0 \le F_\nu^{[-1]} \le F_\mu^{[-1]}$ (to ensure that $(\mathrm{id}, g_\nu)_\# \mu \in \mathcal{M}$) and

$$c = \int_{\mathbb{R}} x^2 \, (g_\nu)_\#\mu(dx) - \left(\int_{\mathbb{R}} x \, (g_\nu)_\#\mu(dx) \right)^2 = \int_{\mathbb{R}} g_\nu(x)^2 \, \mu(dx) - \left(\int_{\mathbb{R}} g_\nu(x) \, \mu(dx) \right)^2 \tag{3.9}$$

$$= \int_0^1 F_\nu^{[-1]}(x)^2 \, dx - \left(\int_0^1 F_\nu^{[-1]}(x) \, dx \right)^2 . \tag{3.10}$$

For the same reason observed in the previous example, using the finiteness of $\mathrm{Var}(X)$, we know that $F_\nu^{[-1]} \in L^2([0,1])$. Therefore, we can view our optimization problem over functions $f \in L^2([0,1])$ satisfying (3.10) such that $0 \le f \le F_\mu^{[-1]}$ and f non-decreasing. For now, we do not consider the last two constraints and we write the Lagrangian operator $\mathcal{L} : L^2([0,1]) \times \mathbb{R} \to \mathbb{R}$ associated with $\mathcal{F}(f)$ and the bilateral constraint:

$$\begin{aligned} \mathcal{L}(f, \lambda) := &\int_0^1 (F_\mu^{[-1]}(x) - f(x))^2 \, dx - \left(\int_0^1 (F_\mu^{[-1]}(x) - f(x)) \, dx \right)^2 \\ &+ \lambda \left(\int_0^1 f(x)^2 \, dx - \left(\int_0^1 f(x) \, dx \right)^2 - c \right) . \end{aligned}$$

We can therefore compute the functional derivative with respect to a function $h \in L^2([0,1])$, given by:

$$\begin{aligned} \mathcal{L}'(f, \lambda; h) = &-2 \int_0^1 (F_\mu^{[-1]}(x) - f(x)) h(x) \, dx + 2 \int_0^1 (F_\mu^{[-1]}(x) - f(x)) \, dx \int_0^1 h(x) \, dx \\ &+ 2\lambda \int_0^1 f(x) h(x) \, dx - 2\lambda \int_0^1 f(x) \, dx \int_0^1 h(x) \, dx \\ = &-2 \int_0^1 (F_\mu^{[-1]}(x) - f(x)) h(x) \, dx + 2 \int_0^1 h(x) \int_0^1 (F_\mu^{[-1]}(y) - f(y)) \, dy \, dx \\ &+ 2\lambda \int_0^1 f(x) h(x) \, dx - 2\lambda \int_0^1 h(x) \int_0^1 f(y) \, dy \, dx \\ = &-2 \int_0^1 h(x) \left(F_\mu^{[-1]}(x) - f(x) - \int_0^1 (F_\mu^{[-1]}(y) - f(y)) \, dy - \lambda f(x) + \lambda \int_0^1 f(y) \, dy \right) dx \\ = &-2 \int_0^1 h(x) \left(F_\mu^{[-1]}(x) - f(x)(\lambda + 1) - \int_0^1 (F_\mu^{[-1]}(y) - f(y)(\lambda + 1)) \, dy \right) dx. \end{aligned}$$

We note that, for $\mathcal{L}'(f, \lambda; h) = 0$ to hold for every $h \in L^2([0,1])$, it must be:

$$F_\mu^{[-1]}(x) - f(x)(\lambda + 1) - \int_0^1 (F_\mu^{[-1]}(y) - f(y)(\lambda + 1)) \, dy = 0$$

for every $x \in [0,1]$. Hence, there exists a costant $a \in \mathbb{R}$ such that:

$$F_\mu^{[-1]} - (\lambda + 1) f = a.$$

Thus, rewriting the bilateral constraint with the f just found, we have:

$$\begin{aligned}
c &= \int_0^1 \left(\frac{F_\mu^{[-1]}(x) - a}{\lambda + 1} \right)^2 dx - \left(\int_0^1 \frac{F_\mu^{[-1]}(x) - a}{\lambda + 1}\, dx \right)^2 \\
&= \frac{1}{(\lambda+1)^2} \left(\int_0^1 F_\mu^{[-1]}(x)^2 \, dx - \left(\int_0^1 F_\mu^{[-1]}(x)\, dx \right)^2 \right) \\
&= \frac{1}{(\lambda+1)^2} \left(\int_{\mathbb{R}} x^2 \, \mu(dx) - \left(\int_{\mathbb{R}} x \, \mu(dx) \right)^2 \right) = \frac{\mathrm{Var}(X)}{(\lambda+1)^2}.
\end{aligned}$$

We choose $\lambda := \sqrt{\frac{\mathrm{Var}(X)}{c}} - 1$, thus the optimal f has to be of the form:

$$f = \frac{F_\mu^{[-1]} - a}{\sqrt{\frac{\mathrm{Var}(X)}{c}}}.$$

To ensure that $0 \leq f \leq F_\mu^{[-1]}$, given that $\frac{\mathrm{Var}(X)}{c} \geq 1$ by assumption, it is sufficient to choose a such that $0 \leq a \leq F_\mu^{[-1]}(0)$. With these choices, the optimal contract is deterministic with R given by:

$$R(x) = \frac{x - a}{\sqrt{\frac{\mathrm{Var}(X)}{c}}}.$$

Note that choosing $a = 0$ imposes no restriction: $a \neq 0$ simply represents a deterministic payment difference, and thus zero variance, from the reinsurer to the insurer. Practically speaking, this translates into a recalibration of the constant that determines $\mathcal{S}$, which leads to a deterministic exchange of the same amount between the two parties, and therefore serves no purpose. The optimal contract is thus a quota-share contract, where a fixed percentage is agreed upon in advance and reimbursed by the reinsurer.

Numerical Example

Let X follow a lognormal distribution given by:

$$f_X(x) := \frac{1}{x\sqrt{2\pi \log(3)}} \exp\left(-\frac{(\log(\sqrt{3}x))^2}{2\log(3)} \right), \quad x > 0.$$

The variance of X is thus two, and by choosing $c = 0.5$, we obtain that the optimal contract is

$$R = \frac{X}{2}.$$

In the case of a quota-share contract, if X admits a density, then R also admits a density, and in this case it is given by:

$$f_R(r) = 2 f_X(2r), \quad r > 0.$$

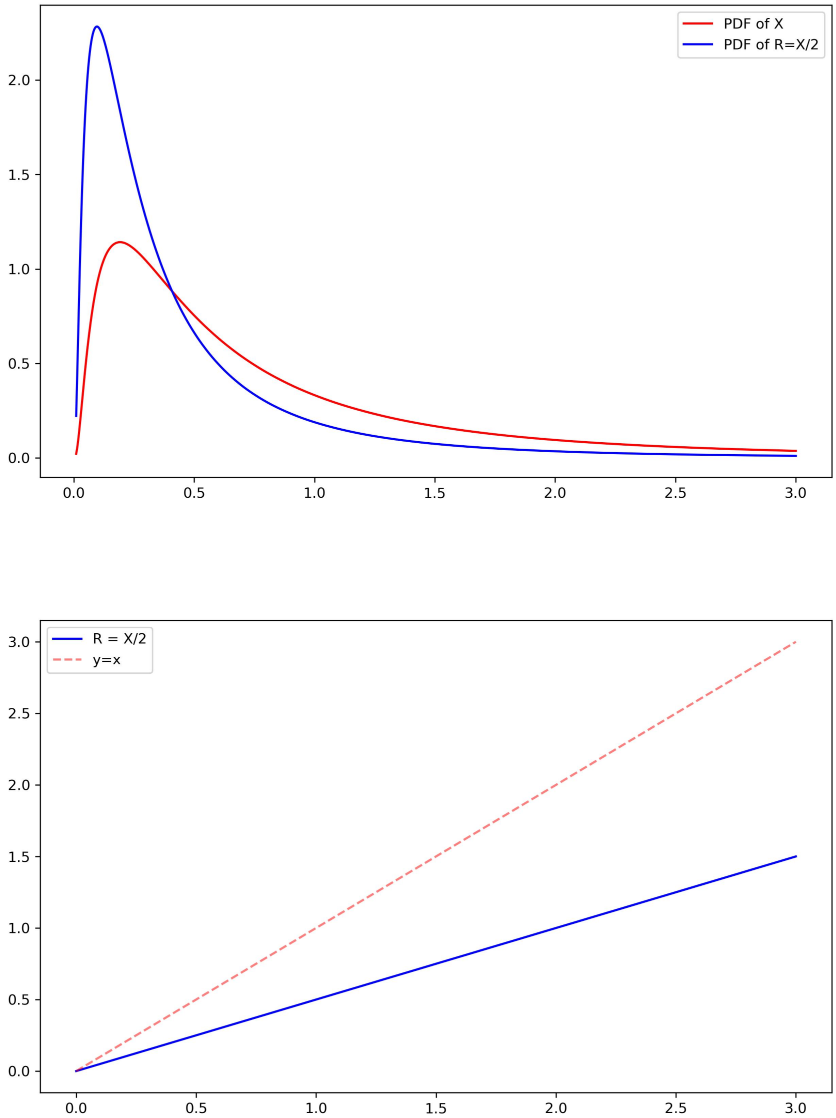


Figure 3.2: Probability density functions of X and R (top) and the (X, R) graph (bottom).

3.5 The Constraint Defining a Multidimensional Optimal Transport Problem: A Discretized Approach

In this example, we change perspective: a reinsurance treaty will represent the joint distribution of the random vector $(X, X - R)$ instead of the vector (X, R). Therefore, we fix $\nu_k \in \mathscr{P}(\mathbb{R}_+)$ for each $k = 1, \ldots, n$, and for every lower semi-continuous functional $\mathcal{P}$, we define

$$\mathcal{S} := \{\eta \in \mathcal{M} \mid (\pi_{2,k})_{\#}\eta = \nu_k,\ k = 1, \dots, n\}.$$

The notation introduced at the beginning is useful for interpreting this set of constraints from a practical point of view. In fact, we can view it as n insurers, each fixing a distribution ν_k, $k = 1, \dots, n$, that they wish to have after the reinsurance. The reinsurer must satisfy these requirements while trying to minimize the functional $\mathcal{P}$. Then, we deal with a multi-marginal optimal transport problem, which we will not attempt to solve in general but only in one particular case.
We take $n = 2$, μ absolutely continuous with finite second moment, and $\mathcal{P}$ given by the variance of the sum of the reinsured amounts. Hence, for every $\eta \in \mathcal{S}$, we obtain:

$$\begin{aligned}\mathcal{P}(\eta) &= \mathrm{Var}(R_1 + R_2) = \mathrm{Var}(X_1 - (X_1 - R_1) + X_2 - (X_2 - R_2)) \\ &= \int_{\mathbb{R}^4_+} (x_1 - y_1 + x_2 - y_2)^2\, \eta(dx_1, dx_2, dy_1, dy_2) - (\overline{\mu_1} - \overline{\nu_1} + \overline{\mu_1} - \overline{\nu_2}).\end{aligned}$$

Thus, since the second term on the right-hand side is fixed, the optimal transport problem becomes:

$$\min_{\eta \in \mathcal{S}} \mathcal{Q}(\eta) \quad \text{with} \quad \mathcal{Q}(\eta) := \int_{\mathbb{R}^4_+} (x_1 - y_1 + x_2 - y_2)^2\, \eta(dx_1, dx_2, dy_1, dy_2).$$

Since X_1 and X_2 have finite variance, the minimization problem is finite; however, solving it for an arbitrary μ is challenging, and for this reason, we fix a specific numerical setting. Suppose X_1 and X_2 are independent, with X_1 has a lognormal distribution and X_2 following a shifted Pareto distribution. The p.d.f.'s of the two variables are given by:

$$f_{X_1}(x) := \frac{1}{x\sqrt{2\pi \log(3)}} \exp\left(-\frac{(\log(\sqrt{3}x))^2}{2\log(3)}\right), \quad x > 0;$$
$$f_{X_2}(x) := 324(x+3)^{-5}, \quad x \geq 0.$$

With this choice of parameters, it is easy to verify that $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 1$ and $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = 2$. At this point, in order to linearize the problem, we use a discretization process for X_1 and X_2, called "binning".

Binning of a random variable

Suppose we have a random variable X with a density f_X with respect to the Lebesgue measure, defined on $[0, +\infty)$ or $(0, +\infty)$ and strictly positive on $(0, +\infty)$. Let $q \in (0, 1)$ and set $u = F_X^{-1}(q)$ (note that in this case $F_X^{[-1]}(q) = F_X^{-1}(q)$ for $q \in (0, 1)$). We then consider the interval $[0, u]$ and, for a fixed $N \in \mathbb{N} \setminus \{0\}$, we divide it into N subintervals I_k, $k = 1, \dots, N$, all of equal length, namely:

$$I_k := \left[\frac{(k-1)u}{N}, \frac{ku}{N}\right], \quad k = 1, \dots, N.$$

What we do is define a discrete random variable $\tilde{X}$ by moving all the probability of the interval I_k to its right endpoint (except for the point u, which, in order to properly define $\tilde{X}$, must also capture the probability that we are unable to capture with the interval $[0, u]$). Using the fact that $F_X(0) = 0$, we can compactly define the probability mass function of $\tilde{X}$ as follows:

$$f_{\tilde{X}}(x) := P[\tilde{X} = x] := \begin{cases} F_X\left(\frac{ku}{N}\right) - F_X\left(\frac{(k-1)u}{N}\right), & \text{if } x = \frac{ku}{N},\ k \in \{1, \dots, N-1\}, \\ 1 - F_X\left(\frac{(N-1)u}{N}\right), & \text{if } x = u. \end{cases} \tag{3.11}$$

Clearly, as N increases and as q approaches one, the function $f_{\tilde{X}}$ approximates f_X more and more accurately.

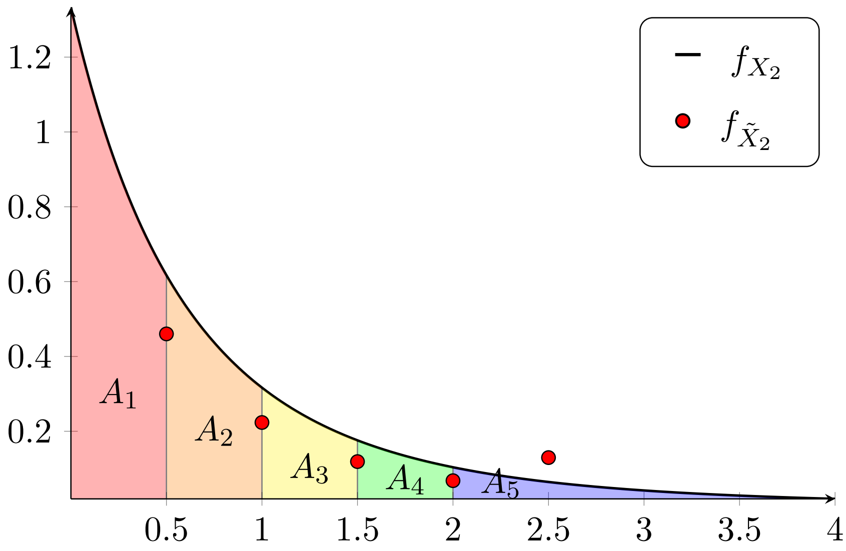


Figure 3.3: Binning of X_2 using parameters $u = 2.5$ and $N = 5$. For every $k = 1, \dots, 4$, A_k represents the area under the graph of f_{X_2} over the intervals $[0.5(k-1), 0.5k]$, while A_5 is the area under the curve over the interval $[2, +\infty)$. The red points, which represent $f_{\tilde{X}_2}$, therefore have coordinates $(0.5k, A_k)$ for each $k = 1, \dots, 5$.

At this point, we consider the binning of X_1 and X_2 where we fix the same $q \in (0, 1)$ and $N \in \mathbb{N} \setminus \{0\}$ for both. In particular, $\tilde{X}_1$ will be defined using $u_1 = F_{X_1}^{-1}(q)$ and $\tilde{X}_2$ using $u_2 = F_{X_2}^{-1}(q)$; we denote by $\tilde{\mu}$ the joint distribution of $(\tilde{X}_1, \tilde{X}_2)$. Let $\tilde{Y}_1$ and $\tilde{Y}_2$ be random variables such that

$$\tilde{Y}_1 \overset{d}{=} 0.5\tilde{X}_1 \quad \text{and} \quad \tilde{Y}_2 \overset{d}{=} \min(\tilde{X}_2, 0.5) + 0.25(\tilde{X}_2 - 0.95)_+,$$

where the notation $\overset{d}{=}$ is used to indicate equality in distribution. From a practical perspective, $\tilde{Y}_1$ could represent, for example, a quota-share contract. In contrast, recalling that $\tilde{Y}_2$ represents the amount remaining for the insurer after reinsurance, we are considering a type of contract where the reinsurer does not become involved as long as the amount to be reinsured is below 0.5 ($\tilde{Y}_2 = \tilde{X}_2$). When $0.5 \leq \tilde{X}_2 \leq 0.95$, the reinsurer reimburses the excess such that the insurer always pays 0.5 ($\tilde{Y}_2 = 0.5$). If $\tilde{X}_2 > 0.95$, then $\tilde{Y}_2 = 0.5 + 0.25(\tilde{X}_2 - 0.95)$. As defined, the values that $\tilde{Y}_1$ can take are always N, while we assume that $\tilde{Y}_2$ can take M distinct values. Therefore, we can consider all the random variables involved to be defined on the space $\Omega' := \{1, \dots, N\}^3 \times \{1, \dots, M\}$, such that for example, $\tilde{Y}_1(i, j, k, l) := \tilde{Y}_1(k)$ where $\tilde{Y}_1(k)$ is the k-th value taken by $\tilde{Y}_1$ in increasing order, and $\tilde{\nu}_1(k) := P[\tilde{Y}_1 = \tilde{Y}_1(k)]$. Similarly, for all the other random variables. Now, the optimization problem becomes finding a vector $A \in \mathbb{R}_+^{N \times N \times N \times M}$ such that

$$\begin{aligned} \sum_{\substack{k=1,\dots,N \\ l=1,\dots,M}} A_{i,j,k,l} = \tilde{\mu}(i,j), \quad (i,j) \in \{1, \dots, N\}^2, \\ \sum_{\substack{i,j=1,\dots,N \\ l=1,\dots,M}} A_{i,j,k,l} = \tilde{\nu}_1(k), \quad k \in \{1, \dots, N\}, \\ \sum_{i,j,k=1,\dots,N} A_{i,j,k,l} = \tilde{\nu}_2(l), \quad l \in \{1, \dots, M\}, \\ A_{i,j,k,l} = 0 \quad \text{if} \quad \tilde{Y}_1(k) > \tilde{X}_1(i) \quad \text{or} \quad \tilde{Y}_2(l) > \tilde{X}_2(j), \end{aligned} \tag{3.12}$$

and A minimizes the expression

$$\sum_{\substack{i,j,k=1,\dots,N \\ l=1,\dots,M}} (\tilde{X}_1(i) - \tilde{Y}_1(k) + \tilde{X}_2(j) - \tilde{Y}_2(l))^2 A_{i,j,k,l} \tag{3.13}$$

among all vectors that satisfy (3.12). To write this linear optimization problem more compactly, we start by viewing A as a vector $\hat{a} \in \mathbb{R}_+^{N^3 M}$, mapping the entry (i, j, k, l) of A to the entry $i + N(j-1) + N^2(k-1) + N^3(l-1)$ of $\hat{a}$. Similarly, we view the cost as a vector $\hat{c} \in \mathbb{R}_+^{N^3 M}$, whose entry at position $i + N(j-1) + N^2(k-1) + N^3(l-1)$ is the value $(\tilde{X}_1(i) - \tilde{Y}_1(k) + \tilde{X}_2(j) - \tilde{Y}_2(l))^2$. We denote by 1_N the column vector of dimension N with all entries equal to one, $\mathbb{I}_N$ the identity matrix of dimension N, and $\otimes$ the Kronecker product. For completeness, we recall that given a matrix $D = [d_{ij}]$ of size $m \times n$ and a matrix E of size $p \times q$, the matrix $D \otimes E$ has size $mp \times nq$ and is defined as

$$D \otimes E := \begin{bmatrix} d_{11}E & \cdots & d_{1n}E \\ \vdots & & \vdots \\ d_{m1}E & \cdots & d_{mn}E \end{bmatrix}.$$

Since $\otimes$ is a tensor product, it is, in particular, associative, and the following definition of the $(N^2+N+M)\times N^3M$ matrix $\hat{B}$ makes sense:

$$\hat{B} := \begin{bmatrix} 1_{NM}^{\intercal} \otimes \mathbb{I}_{N^2} \\ 1_{M}^{\intercal} \otimes \mathbb{I}_{N} \otimes 1_{N^2}^{\intercal} \\ \mathbb{I}_{M} \otimes 1_{N^3}^{\intercal} \end{bmatrix}.$$

We define the vector $\theta \in \mathbb{R}^{N^2+N+M}$, which has in its first N^2 components the "flattened" $\tilde{\mu}$ (i.e. the value $\tilde{\mu}(i,j)$ corresponds to the $i+N(j-1)$ entry of θ), in the next N components the marginal $\tilde{\nu}_1$, and in the last M components the marginal $\tilde{\nu}_2$. First of all, we claim that A satisfies the three equalities in (3.12) if and only if $\hat{B}\hat{a}=\theta$. Let us start by calculating the entries $(\hat{B}\hat{a})_t$ for $t \le N^2$. There exist unique $i,j \in \{1,\dots,N\}$ such that $t=i+N(j-1)$. Denoting by D_t the t-th row of a matrix D, we observe that:

$$\left(\left[1_{NM}^{\intercal} \otimes \mathbb{I}_{N^2}\right]_t\right)_u = \begin{cases} 1, & \text{if } u = t+sN^2,\ s \in \{0,\dots,NM-1\}, \\ 0, & \text{otherwise.} \end{cases}$$

From this, we obtain that:

$$\begin{aligned} (\hat{B}\hat{a})_{i+N(j-1)} &= \left[1_{NM}^{\intercal} \otimes \mathbb{I}_{N^2}\right]_{i+N(j-1)} \hat{a} \\ &= \sum_{s=0}^{NM-1} \hat{a}_{i+N(j-1)+sN^2} \\ &= \sum_{\substack{k=1,\dots,N \\ l=1,\dots,M}} \hat{a}_{i+N(j-1)+N^2(k-1)+N^3(l-1)} = \sum_{\substack{k=1,\dots,N \\ l=1,\dots,M}} A_{i,j,k,l}, \end{aligned}$$

where the penultimate equality follows from the fact that $\{sN^2 \mid s=0,\dots,NM-1\} = \{N^2(k-1)+N^3(l-1) \mid k=1,\dots,N,\ l=1,\dots,M\}$. Let us now consider $N^2 < t \le N^2+N$. Then, there exists a unique $k \in \{1,\dots,N\}$ such that $t=N^2+k$. We define the set

$$U_k := \{N^2(sN+k-1)+r \mid s \in \{0,\dots,M-1\},\ r \in \{1,\dots,N^2\}\}$$

and note that

$$\left(\left[1_{M}^{\intercal} \otimes \mathbb{I}_{N} \otimes 1_{N^2}^{\intercal}\right]_k\right)_u = \begin{cases} 1, & \text{if } u \in U_k, \\ 0, & \text{otherwise.} \end{cases}$$

Moreover, rewriting

$$N^2(sN+k-1)+r = N^3s+N^2(k-1)+r,$$

it is evident that the set U_k coincides with

$$\{i+N(j-1)+N^2(k-1)+N^3(l-1) \mid i,j \in \{1,\dots,N\},\ l \in \{1,\dots,M\}\}.$$

Hence, this leads to the following equalities:

$$\begin{aligned}(\hat{B}\hat{a})_{N^2+k} &= \left[1_M^\intercal \otimes \mathbb{I}_N \otimes 1_{N^2}^\intercal\right]_k \hat{a}\\ &= \sum_{\substack{s=0,\dots,M-1\\ r=1,\dots,N^2}} \hat{a}_{N^2(sN+k-1)+r}\\ &= \sum_{\substack{i,j=1,\dots,N\\ l=1,\dots,M}} \hat{a}_{i+N(j-1)+N^2(k-1)+N^3(l-1)} = \sum_{\substack{i,j=1,\dots,N\\ l=1,\dots,M}} A_{i,j,k,l}.\end{aligned}$$

Finally, we consider $t > N^2 + N$. Let $l \in \{1, \dots, M\}$ be such that $t = N^2 + N + l$, then

$$\left(\left[\mathbb{I}_M \otimes 1_{N^3}^\intercal\right]_l\right)_u = \begin{cases} 1, & \text{if } u = N^3(l-1) + s,\ s \in \{1, \dots, N^3\},\\ 0, & \text{otherwise.}\end{cases}$$

Thus, it holds:

$$\begin{aligned}(\hat{B}\hat{a})_{N^2+N+l} &= \left[\mathbb{I}_M \otimes 1_{N^3}^\intercal\right]_l \hat{a}\\ &= \sum_{s=1,\dots,N^3} \hat{a}_{N^3(l-1)+s}\\ &= \sum_{i,j,k=1,\dots,N} \hat{a}_{i+N(j-1)+N^2(k-1)+N^3(l-1)} = \sum_{i,j,k=1,\dots,N} A_{i,j,k,l}.\end{aligned}$$

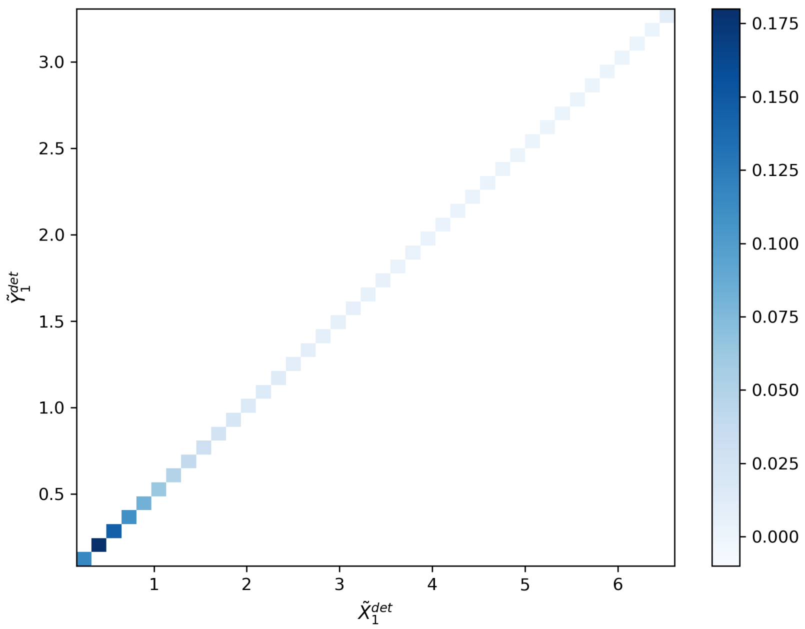
We have verified the claim. At this point, to include the final condition of (3.12), it suffices to remove from $\hat{a}$ and $\hat{c}$ the entries $i + N(j-1) + N^2(k-1) + N^3(l-1)$ for which $\tilde{Y}_1(k) > \tilde{X}_1(i)$ or $\tilde{Y}_2(l) > \tilde{X}_2(j)$ and, at the same time, we remove from $\hat{B}$ the columns at positions $i + N(j-1) + N^2(k-1) + N^3(l-1)$ for which i, j, k, l satisfy the same property. Thus, we obtain the vectors a and c, and the matrix B, respectively. Denoting by K the columns that remain after this operation, the linear optimization problem becomes:

$$\min\{c^\intercal a \mid a \in \mathbb{R}_+^K,\ Ba = \theta\}. \tag{3.14}$$

We set $N = 40$ and $q = 0.99$, and we use the `linprog` function from the `scipy.optimize` module to solve the linear optimization problem defined in (3.14) (see Appendix A). We denote by η_{det} the deterministic contract obtained using the push-forward map $(x_1, x_2) \mapsto (x_1, x_2, 0.5x_1, \min(x_2, 0.5) + 0.25(x_2 - 0.95)_+)$, while η_{opt} represents the optimal contract that solves (3.14). Furthermore, we set $(\tilde{X}_1^{\text{det}}, \tilde{X}_2^{\text{det}}, \tilde{Y}_1^{\text{det}}, \tilde{Y}_2^{\text{det}})$ and $(\tilde{X}_1^{\text{opt}}, \tilde{X}_2^{\text{opt}}, \tilde{Y}_1^{\text{opt}}, \tilde{Y}_2^{\text{opt}})$ the random vectors distributed according to η_{det} and η_{opt}, respectively. Recalling that $R = X - Y$, we obtain the following results:

$$\operatorname{Var}(\tilde{R}_1^{\text{det}} + \tilde{R}_2^{\text{det}}) = 1.05314, \qquad \operatorname{Var}(\tilde{R}_1^{\text{opt}} + \tilde{R}_2^{\text{opt}}) = 0.82568.$$

Thus, we observe a 21.60% improvement and proceed to analyze how this was achieved. The first reason can be inferred from Figures 3.4 and 3.5: by introducing randomness in the contract η_{opt}, $\tilde{Y}_1^{\text{opt}}$ and $\tilde{Y}_2^{\text{opt}}$ have a positive probability of being close to $\tilde{X}_1^{\text{opt}}$ and $\tilde{X}_2^{\text{opt}}$, respectively, which implies a reduction in the values of the reinsured amounts. Furthermore, while $\tilde{R}_1^{\text{det}}$ and $\tilde{R}_2^{\text{det}}$ are independent, the contract η_{opt} introduces different dependency relationships among the various marginals. In fact, Figure 3.6 reveals a negative dependence structure between $\tilde{R}_1^{\text{opt}}$ and $\tilde{R}_2^{\text{opt}}$, where a significant mass is concentrated near the axes. Similarly, from Figure 3.7, we observe that the mass of the joint distribution of $(\tilde{X}_1^{\text{opt}}, \tilde{Y}_2^{\text{opt}})$ is concentrated in the bottom-left corner. This can be explained as follows: on one hand, the new degree of freedom allows positive probability for points with high ordinate values; on the other hand, it must satisfy the constraints $\tilde{Y}_2^{\text{opt}} \leq \tilde{X}_2^{\text{opt}}$ and $\tilde{Y}_2^{\text{opt}} = \tilde{X}_2^{\text{opt}}$ when $\tilde{X}_2^{\text{opt}} \leq 0.5$. A similar reasoning applies to the joint distribution of $(\tilde{X}_2^{\text{opt}}, \tilde{Y}_1^{\text{opt}})$ and the constraint $\tilde{Y}_1^{\text{opt}} \leq \tilde{X}_1^{\text{opt}}$. Finally, from Figure 3.8, we can observe that $\tilde{Y}_1^{\text{opt}}$ and $\tilde{Y}_2^{\text{opt}}$ are not always concentrated at a single point given the realization of the vector $(\tilde{X}_1^{\text{opt}}, \tilde{X}_2^{\text{opt}})$. This indicates that the optimal contract utilizes randomness external to $(\tilde{X}_1^{\text{opt}}, \tilde{X}_2^{\text{opt}})$, which allows it to reduce the variance.



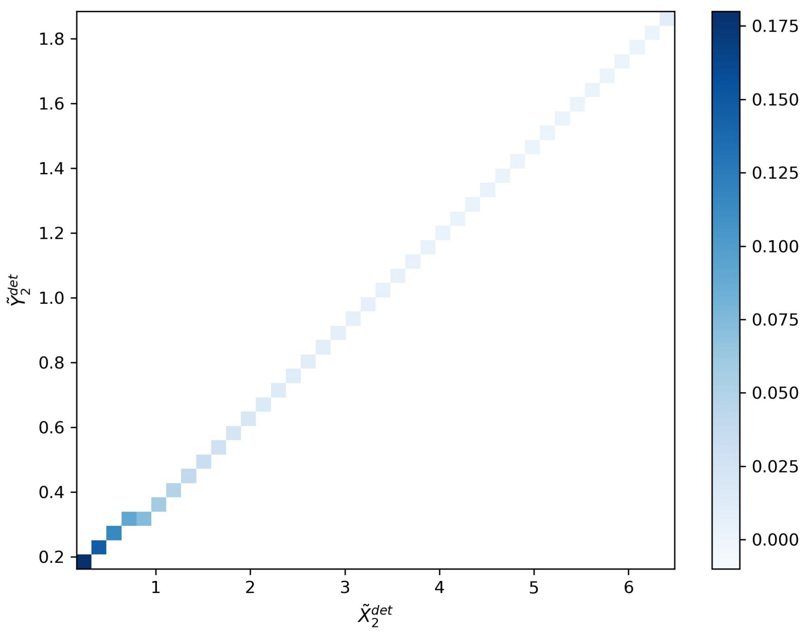


Figure 3.4: Probability mass function for the joint distributions of $(\tilde{X}_1^{\text{det}}, \tilde{Y}_1^{\text{det}})$ and $(\tilde{X}_2^{\text{det}}, \tilde{Y}_2^{\text{det}})$.

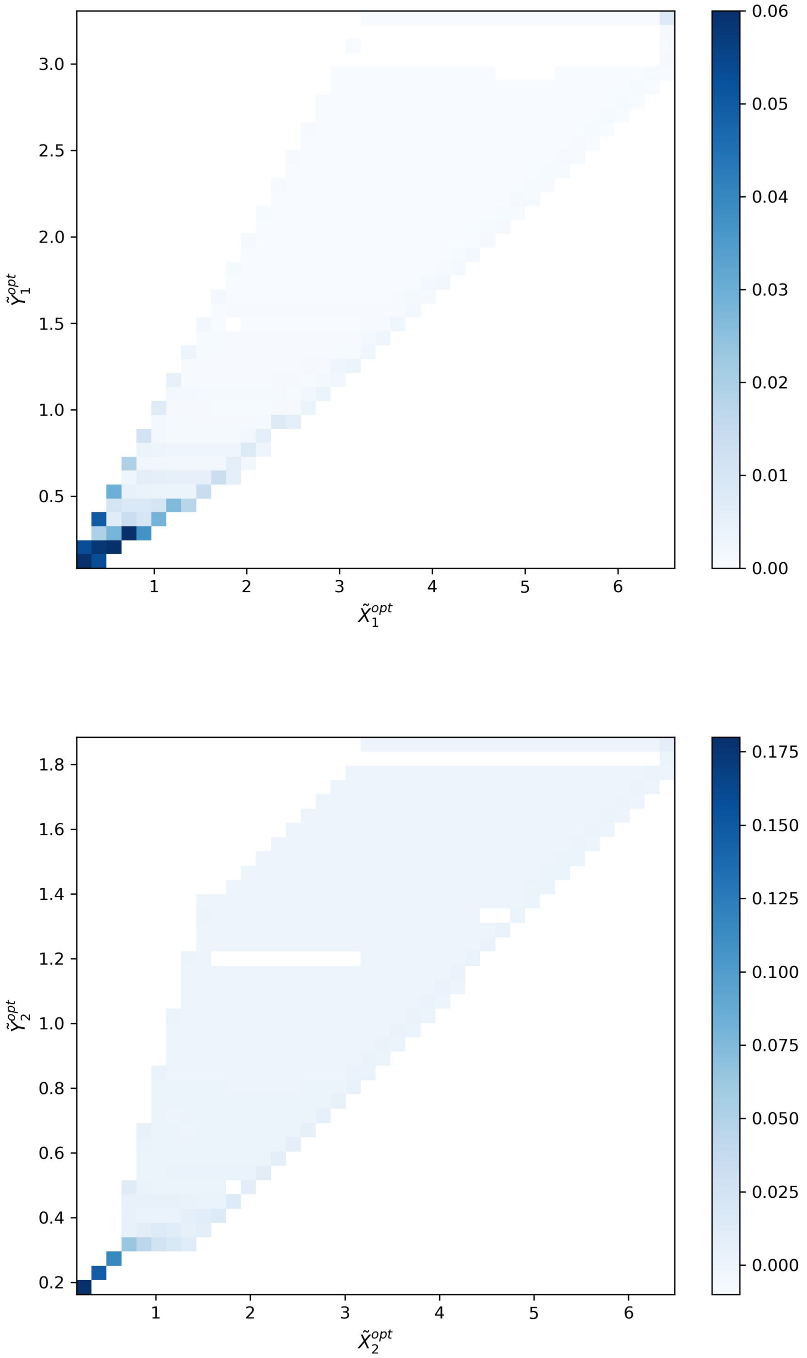


Figure 3.5: Probability mass function for the joint distributions of $(\tilde{X}_1^{\text{opt}}, \tilde{Y}_1^{\text{opt}})$ and $(\tilde{X}_2^{\text{opt}}, \tilde{Y}_2^{\text{opt}})$.

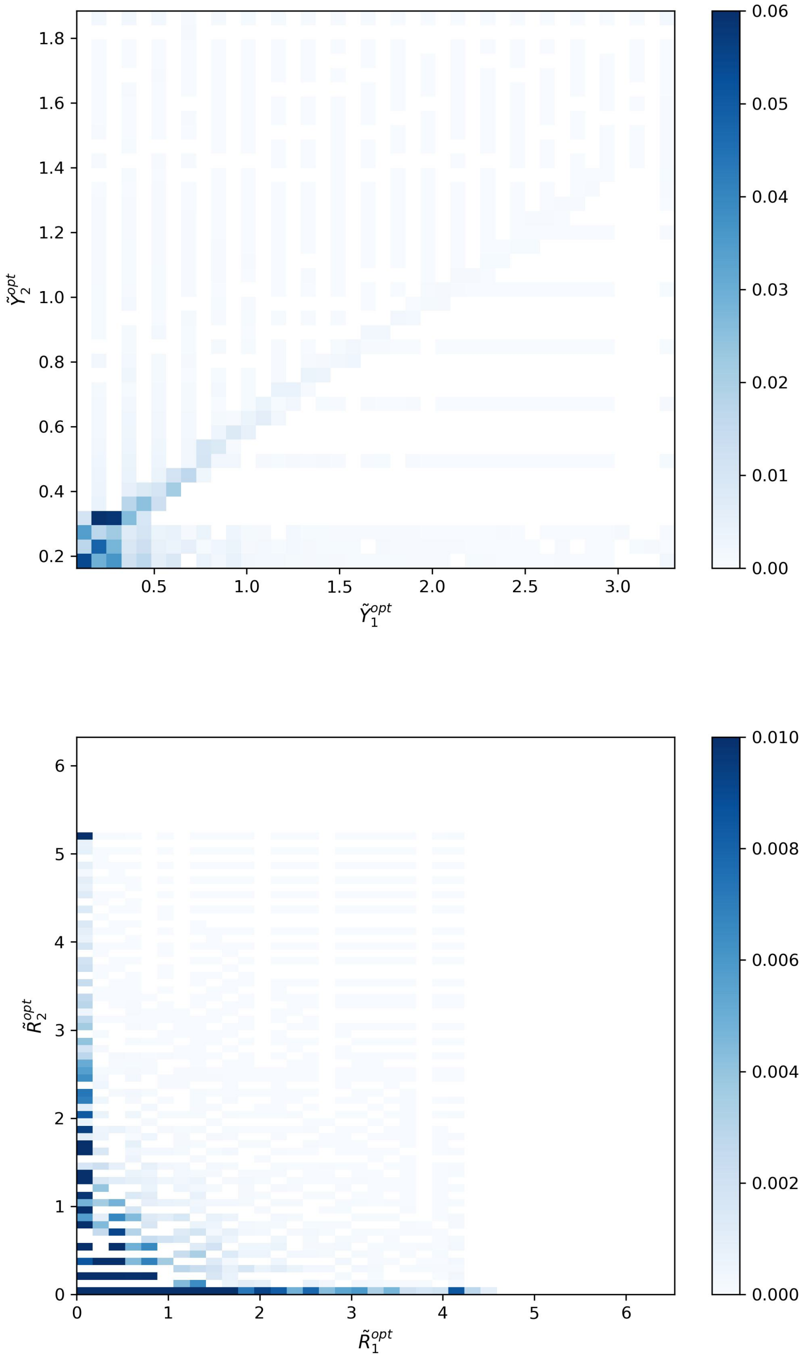


Figure 3.6: Probability mass function for the joint distributions of $(\tilde{Y}_1^{\text{opt}}, \tilde{Y}_2^{\text{opt}})$ and $(\tilde{R}_2^{\text{opt}}, \tilde{R}_1^{\text{opt}})$.

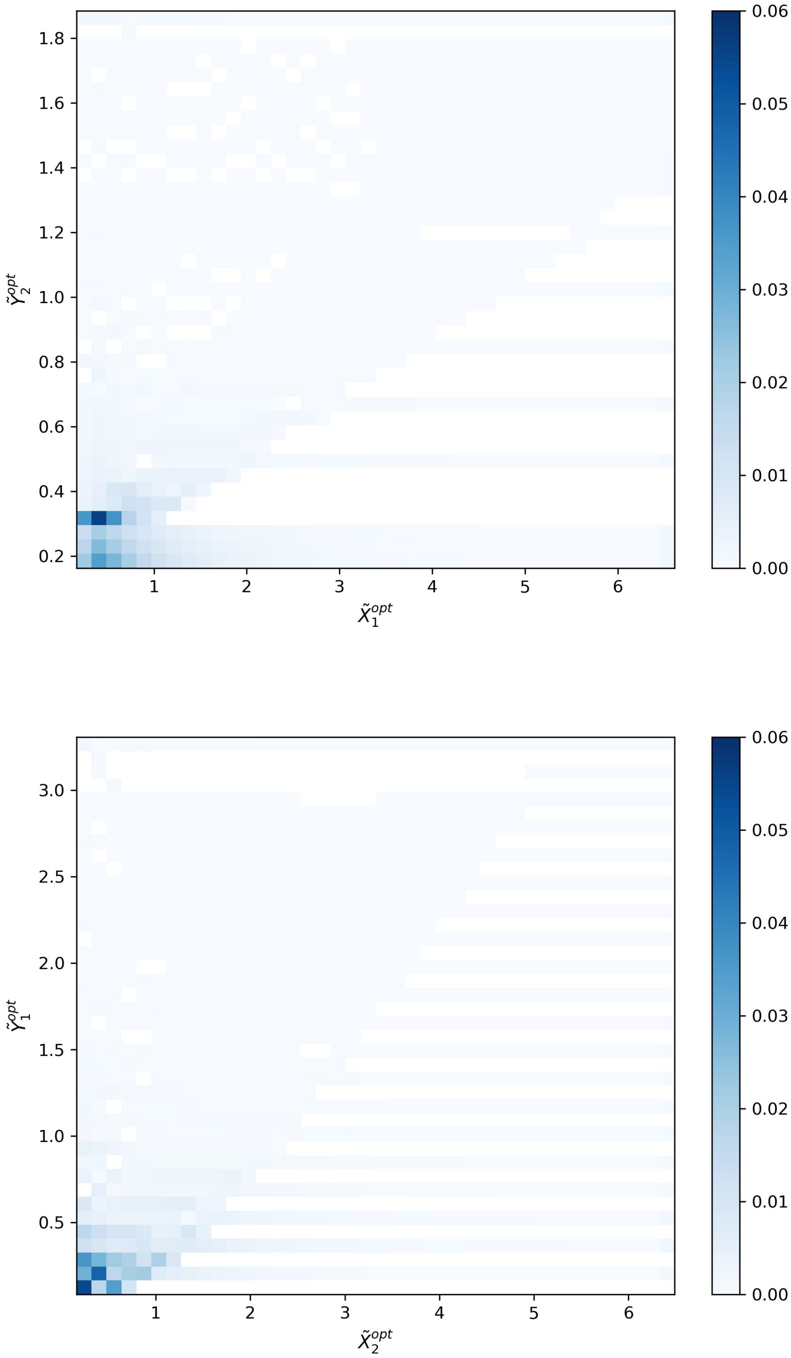
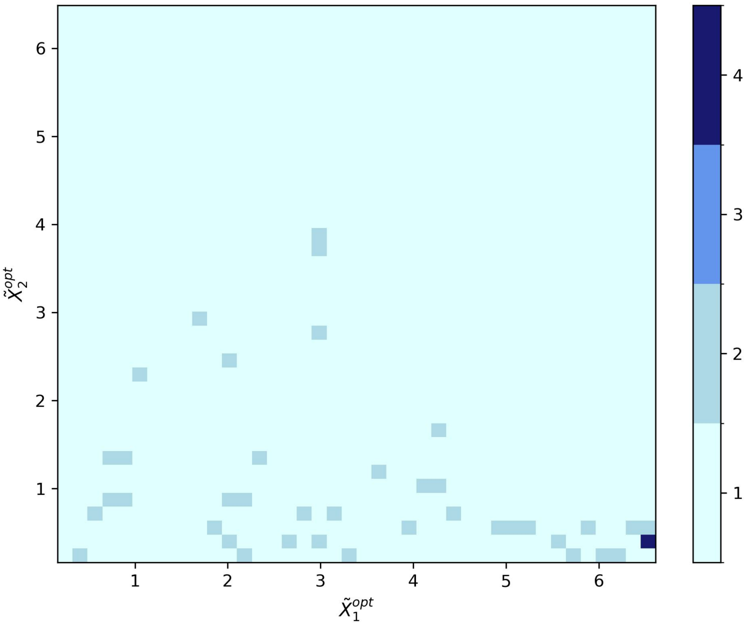


Figure 3.7: Probability mass function for the joint distributions of $(\tilde{X}_1^{\text{opt}}, \tilde{Y}_2^{\text{opt}})$ and $(\tilde{X}_2^{\text{opt}}, \tilde{Y}_1^{\text{opt}})$.



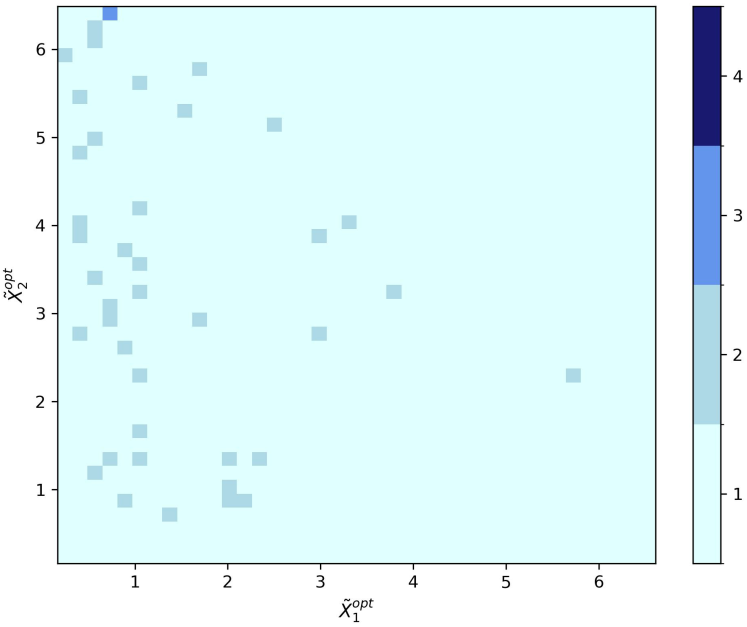


Figure 3.8: Cardinalities of the supports of the conditional distributions of $\tilde{Y}_1^{\text{opt}}$ (top) and $\tilde{Y}_2^{\text{opt}}$ (bottom) given $(\tilde{X}_1^{\text{opt}}, \tilde{X}_2^{\text{opt}})$.

Chapter 4

Conclusions and future directions

The techniques for solving the optimal reinsurance problem described in this work appear to perform well across a wide range of problems. For this reason, some of the results presented might be generalized in certain ways. Thus, we outline some possible directions for future research.

- The case study in Section 3.2 suggests how Proposition 2.2.5 works well with the VaR_α measure. Therefore, it would be possible to increase the dimension of the constraint, i.e. define $\mathcal{G} = (g_1, ..., g_m) : \mathcal{M} \to \mathbb{R}^m$ such that for every $i = 1, \dots, m$

  $$g_i(\eta) := \widehat{\mathrm{VaR}}_{\alpha_i}(\eta) - c_i$$

  with $0 \leq c_1 \leq \dots \leq c_m$ and $0 < \alpha_m < \dots < \alpha_1 < 1$. Similarly to the base case, a sensible choice seems to define

  $$\mathcal{C} := \left\{ \eta \in \mathcal{D} \mid \widehat{\mathrm{VaR}}_{\alpha_i}(\eta) = v_i^* \right\}$$

  where $v_i^* := \widehat{\mathrm{VaR}}_{\alpha_i}(\eta^*)$ for an optimal contract η^*.

- An important class of risk measures are convex risk measures. These have the particularity of possessing a dual representation (see [2]), which seems to be linked with the existence of the directional derivative. Hence, on one hand, the convex linearization approach appears to be effective. On the other hand, since we often deal with strictly convex cost functions in this context, the optimal transport perspective seems to fit well too. The choice between the two, besides being determined by the dimension on which we operate, will largely depend on the type of constraints that we face.

- As pointed out in Remark 2.3.7, there is a technical difficulty in generalizing the proof of Proposition 2.3.6 to an arbitrary n. However, this does not preclude the possibility that, for a certain class of functionals, one might be able to show that

$$0 \leq (\nabla h_{\eta^*})^{-1} \circ \nabla\phi(x) \leq \mathrm{Id}$$

where ϕ is the function defining $T(x)$ in the remark, thus guaranteeing the existence of a deterministic optimal contract.

Appendix A

Python Implementation for Section 3.5

Below, we provide the implementation of the Python code that allows us to solve the linear optimization problem (3.14). We will use the notation introduced in Section 3.5.
First, we import the necessary packages:

```
import numpy as np
from scipy.sparse import kron, identity, vstack
from scipy.stats import rv_continuous
from scipy.stats import lognorm
from scipy.optimize import linprog
```

We define the function `bin_random_variable` that allows us to perform the binning of a random variable X under the assumptions of Section 3.5. As can be seen from the following code, the function takes as arguments the cumulative distribution function of X, the quantile q, and the number N of bins, and returns the discrete random variable $\tilde{X}$ with the respective probabilities for the values it assumes.

```
# Define a function for binning a random variable
def bin_random_variable(f_X_cdf, q, N):

    # Compute the inverse CDF (quantile function) at q
    u = f_X_cdf.ppf(q)

    # Define the bin endpoints
    bin_edges = np.linspace(0, u, N + 1)

    # Initialize the discrete random variable support and probabilities
```

```
    discrete_x = []
    probabilities = []

    for k in range(1, N):
        # Right endpoint of the interval I_k
        x_k = bin_edges[k]

        # Compute the probability mass for the interval I_k
        p_k = f_X_cdf.cdf(bin_edges[k]) - f_X_cdf.cdf(bin_edges[k - 1])

        # Append the right endpoint and its probability
        discrete_x.append(x_k)
        probabilities.append(p_k)

    # Handle the last interval including the point u
    discrete_x.append(u)
    probabilities.append(1 - f_X_cdf.cdf(bin_edges[N - 1]))

    return discrete_x, probabilities
```

Then, we perform the binning of X_1 and X_2, using the parameters $q = 0.99$ and $N = 40$ for both. Hence, we obtain $\tilde{X}_1$ and $\tilde{X}_2$.

```
# Binning of X_1

# Parameters for the log-normal distribution
sigma = np.sqrt(np.log(3))
mu = -np.log(np.sqrt(3))

# Instantiate the log-normal distribution
lognormal_dist = lognorm(s=sigma, scale=np.exp(mu))

# Parameters
q = 0.99  # Quantile
N = 40    # Number of bins

# Perform binning
tilde_X_1, probabilities_tilde_X_1 = bin_random_variable(lognormal_dist, q, N)
```

```
# Binning of X_2

class CustomDistribution(rv_continuous):
    def _pdf(self, x):
        return 324 * (x + 3)**-5 if x >= 0 else 0
```

```

    def _cdf(self, x):
        if x >= 0:
            return 1 - (3 / (x + 3))**4
        return 0

# Instantiate the custom distribution
X_2_dist = CustomDistribution(a=0, name="custom_density")

# Parameters
q = 0.99  # Quantile
N = 40    # Number of bins

# Perform binning
tilde_X_2, probabilities_tilde_X_2 = bin_random_variable(X_2_dist, q, N)
```

At this stage, we define the random variables $\tilde{Y}_1$ and $\tilde{Y}_2$. For the second random variable, we also create the lists `tilde_Y_2_no_cumulative` and `original_probabilities`, which are only needed to reproduce the plots in Section 3.5. However, they are not essential for the purposes of this appendix.

```
def compute_quota_share_X(discrete, probabilities):
    discrete_quota_share = [0.5 * x for x in discrete]
    probabilities_quota_share = probabilities
    return discrete_quota_share, probabilities_quota_share

def compute_stop_loss_X(discrete, probabilities):

    probability_map = {}
    transformed_X = []
    original_probabilities = []

    for x, p in zip(discrete, probabilities):
        y=min(x, 0.5) + 0.25 * max(x - 0.95, 0)

        transformed_X.append(y)
        original_probabilities.append(p)

        if y in probability_map:
            probability_map[y] += p
        else:
            probability_map[y] = p

    discrete_stop_loss = list(probability_map.keys())
    probabilities_stop_loss = list(probability_map.values())
```

```python

    return (discrete_stop_loss, probabilities_stop_loss), (transformed_X,
    ↪ original_probabilities)

tilde_Y_1, probabilities_tilde_Y_1 = compute_quota_share_X(tilde_X_1,
↪ probabilities_tilde_X_1)
(aggregated, non_aggregated) = compute_stop_loss_X(tilde_X_2, probabilities_tilde_X_2)
tilde_Y_2, probabilities_tilde_Y_2 = aggregated

tilde_Y_2_no_cumulative, original_probabilities = non_aggregated
```

We define the vectors c and θ, as well as the matrix B. Let us notice that we initialize an auxiliary list, `col_indices`, which is used to store the columns (entries) to be retained in $\hat{B}$ $(\hat{c})$ in order to obtain B (c).

```python
N = len(tilde_X_1)
M = len(tilde_Y_2)

hat_c_size = N**3 * M

hat_c = np.zeros(hat_c_size)

col_indices = []

for l in range(1, M+1):
    for k in range(1, N+1):
        for j in range(1, N+1):
            for i in range(1, N+1):
                index = i + N * (j-1) + N**2 * (k-1) + N**3 * (l-1) -1
                hat_c[index] = (tilde_X_1[i-1] - tilde_Y_1[k-1] + tilde_X_2[j-1] -
                ↪ tilde_Y_2[l-1])**2

                if tilde_Y_1[k-1] <= tilde_X_1[i-1] and tilde_Y_2[l-1] <= tilde_X_2[j-1]:
                    col_indices.append(index)

hat_c = np.array(hat_c)

c=hat_c[col_indices]
```

```python
def initialize_theta(N, M, probabilities_tilde_X_1, probabilities_tilde_X_2,
↪ probabilities_tilde_Y_1, probabilities_tilde_Y_2):

    theta_size = N**2 + N + M

```

```
    theta = np.zeros(theta_size)

    for i in range(1, N+1):
        for j in range(1, N+1):
            index = i + (j-1) * N -1
            theta[index] = probabilities_tilde_X_1[i-1] * probabilities_tilde_X_2[j-1]

    theta[N**2:N**2 + N] =  probabilities_tilde_Y_1

    theta[N**2 + N:] = probabilities_tilde_Y_2

    return theta

theta = initialize_theta(N, M, probabilities_tilde_X_1, probabilities_tilde_X_2,
↪ probabilities_tilde_Y_1, probabilities_tilde_Y_2)
```

```
B1 = kron(np.ones((N * M, 1)).T, identity(N**2, format="csr"))
B2 = kron(kron(np.ones((M, 1)).T, identity(N, format="csr")), np.ones((N**2, 1)).T)
B3 = kron(identity(M, format="csr"), np.ones((N**3, 1)).T)

hat_B = vstack([B1, B2, B3], format="csr")

B = hat_B[:, col_indices]
```

At this point, we have all the necessary components to define the function that solves the problem (3.14):

```
def solve_linear_optimization(c, B, theta):
    """
    Solves the linear optimization problem:
    Minimize: c^T a
    subject to: Ba = theta
                a >= 0
    """
    # Number of variables (dimension of a)
    num_vars = B.shape[1]

    # Define the problem for linprog
    # Minimization of c^T a
    # Equality constraints: Ba = theta
    # Non-negativity constraints: a >= 0
    result = linprog(
        c,                  # Objective function
        A_eq=B,             # Equality constraints matrix
```

```
        b_eq=theta,         # Equality constraints vector
        bounds=(0, None),   # Non-negativity constraints
        method="highs"      # Advanced numerical method for large problems
    )

    if result.success:
        print("Optimization completed successfully.")
        print("Minimum value of the objective function:", result.fun)
        return result.x  # Returns the optimal solution a
    else:
        print("Optimization error:", result.message)
        return None

# Solve the problem
opt_result = solve_linear_optimization(c, B, theta)
```

Bibliography

[1] Bruno De Finetti. *Il problema dei pieni.* Istituto Italiano degli Attuari, 1940.

[2] R. Tyrrell Rockafellar. *Conjugate Duality and Optimization.* Vol. 16. Regional Conference Series in Applied Mathematics. Philadelphia: SIAM, 1974.

[3] Wilfrid Gangbo and Robert J McCann. “The geometry of optimal transportation”. In: (1996).

[4] Sashi Mohan Srivastava. *A course on Borel sets.* Vol. 180. Springer Science & Business Media, 2008.

[5] Manuel Guerra and Maria L Centeno. “Are quantile risk measures suitable for risk-transfer decisions?” In: *Insurance: Mathematics and Economics* 50.3 (2012), pp. 446–461.

[6] Patrick Billingsley. *Convergence of probability measures.* John Wiley & Sons, 2013.

[7] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians.* Springer International Publishing AG Switzerland, 2015.

[8] Hansjoerg Albrecher, Jan Beirlant, and Jozef L. Teugels. *Reinsurance: Actuarial and Statistical Aspects.* Wiley Series in Probability and Statistics. Hoboken, NJ: John Wiley & Sons, Inc., 2017.

[9] Luigi Ambrosio, Elia Brué, and Daniele Semola. *Lectures on Optimal Transport.* Springer, 2021.

[10] Beatrice Acciaio, Hansjörg Albrecher, and Brandon García Flores. *Optimal reinsurance from an optimal transport perspective.* 2024. arXiv: 2312.06811 [math.OC]. URL: https://arxiv.org/abs/2312.06811.

[11] Julio Backhoff-Veraguas and Martin Huesmann. “Stochastic Mass Transfer”. In: (n.d.).