Risk Forms: Representation, Disintegration, and Application to Partially Observable Two-Stage Systems

Darinka Dentcheva* and Andrzej Ruszczyński†

July 9, 2018

Abstract

We introduce the concept of a risk form, which is a real functional on the product of two spaces: the space of measurable functions and the space of measures on a Polish space. We prove a dual representation theorem and generalize the Kusuoka representation to this setting. For a risk form acting on a product space, we define marginal and conditional forms and we prove a disintegration formula, which represents a risk form as a composition of its marginal and conditional forms. We apply the proposed approach to two-stage stochastic programming problems with partial information and decision-dependent observation distribution.

Keywords: Risk Measures, Kusuoka Representation, Risk Disintegration, Two-Stage Stochastic Programming, Partially Observable Systems

1 Introduction

The theory of risk measures is one of the main directions of recent developments in stochastic optimization. It has found multitude of applications, far beyond the original motivation in finance. The main setting is the following: a probability space \((\Omega, \mathcal{F}, P)\) is given and a space \(\mathcal{L}^p\) of real-valued measurable functions on \(\Omega\) is defined (usually, \(\mathcal{L}^p(\Omega, \mathcal{F}, P)\) with \(p \in [1, \infty]\)). A (convex) risk measure is a convex, monotonic, and translation-equivariant functional \(\rho: \mathcal{L}^p \to \mathbb{R}\). We refer to [18], [22], [1], and [10] for initial contributions, and to [11], [29], [23], [31], [23] and the survey [3] for detailed presentation, applications, and further references.

Two key results provide variational representation of risk measures. One of them, called dual representation, can be derived from the theory of conjugate duality, as shown in [29]. Another representation is known as Kusuoka representation of law invariant coherent measures of risk [20]. It is derived from the dual representation by employing the Hardy-Littlewood-Pólya inequality (see [13]), under several assumptions about the properties of the measures of risk. In all these developments, the original probability measure \(P\) is assumed fixed, which is essential for the use of convex analysis techniques in the spaces of measurable functions.

We propose a different approach. We fix a Polish space \(\mathcal{X}\) with its Borel \(\sigma\)-algebra \(\mathcal{B}(\mathcal{X})\), but we allow arbitrary probability measures \(P\) on this space. In section [2] we introduce real-valued functionals of two arguments, \(\rho[Z, P]\), where \(Z\) is a bounded measurable function on \(\mathcal{X}\) and \(P\) is a probability measure.

---

*Stevens Institute of Technology, Department of Mathematical Sciences, Hoboken, NJ 07030, USA
†Rutgers University, Department of Management Science and Information Systems, Piscataway, NJ 08854, USA
on \((\mathcal{X}, \mathcal{B}(\mathcal{X}))\). In analogy to the bilinear form \(E[Z, P] = \int_{\mathcal{X}} Z(x) P(dx)\), we call \(\rho[Z, P]\) a risk form. Transition risk mappings, which arose in our recent research on risk-averse control \([7, 8, 27]\), are special cases of risk forms.

Under assumptions which are less restrictive than in the fixed probability measure case, we prove a generalized Kusuoka representation of risk forms in section \([3]\). The risk representation has a universal character; it remains valid for all probability measures \(P\).

The second contribution of the paper is the risk disintegration formula and its implications. In section \([4]\) we introduce the property of conditional consistency of risk forms. We prove that forms enjoying this property can be represented as compositions of two forms, which we call marginal and conditional forms. This result generalizes the decomposition of the bilinear form, resulting from the disintegration of probability measures. While our approach is related to the theory of dynamic and conditional risk measures (see \([30, 25, 26, 9, 4, 28, 2, 24, 19, 15, 5]\) and the references therein), it allows for variable probability measures and does not have any time structure associated with it; the order of conditioning may be arbitrary. These results are generalized in section \([5]\), where we consider multi-step disintegration and prove the generalized tower property of conditional risk forms, as a counterpart of the tower property of conditional expectations.

Our final contribution is the application of the risk form theory to two-stage risk-averse optimization of models with partial observation (section \([6]\)). Opposite to classical two-stage models, we assume that only partial information is available at the second stage, which allows for the update of the conditional distribution of the unobserved part. This setting was first considered in \([21]\) and \([14]\), in a special case, and with a postulated structure of the overall measure of risk. We generalize and justify the earlier contributions, by proving the equivalence of the overall risk optimization and two-stage optimization in this setting. We also allow for decision-dependent observation distribution and develop a risk-averse Bayes formula. In the risk-neutral case, stochastic programming models with endogeneous (decision-dependent) uncertainty have been discussed in \([16]\), where the probability distribution and the first stage decision are linked by a special constraint.

## 2 Risk Models with Variable Probability Measures

Consider a Polish space \(\mathcal{X}\) and its Borel \(\sigma\)-algebra \(\mathcal{B}(\mathcal{X})\). Let \(\mathcal{P}(\mathcal{X})\) be the set of probability measures on \((\mathcal{X}, \mathcal{B}(\mathcal{X}))\). The space of all real-valued bounded measurable functions on \(\mathcal{X}\) is denoted by \(\mathbb{B}(\mathcal{X})\). We use \(x\) to denote an element of \(\mathcal{X}\) and \(\delta_x\) to denote the Dirac measure concentrated at \(x\). The symbol \(1\) stands for the function in \(\mathbb{B}(\mathcal{X})\), which is constantly equal to 1. For two elements \(V, W \in \mathbb{B}(\mathcal{X})\), the notation \(V \overset{\rho}{\leq} W\) means that \(P\{V \leq \eta\} = P\{W \leq \eta\}\) for all \(\eta \in \mathbb{R}\).

A probabilistic model is a pair \([Z, P] \in \mathbb{B}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})\). We are interested in evaluating risk of probabilistic models. Our goal is to propose a universal approach to risk evaluation of a family of probabilistic models.

**Definition 2.1.** A measurable functional \(\rho : \mathbb{B}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \to \mathbb{R}\) is called a risk form.

(i) It is monotonic, if \(V \leq W\) implies \(\rho[V, P] \leq \rho[W, P]\) for all \(P \in \mathcal{P}(\mathcal{X})\);

(ii) It is normalized if \(\rho[0, P] = 0\) for all \(P \in \mathcal{P}(\mathcal{X})\);

(iii) It is translation equivariant if for all \(V \in \mathbb{B}(\mathcal{X})\), all \(a \in \mathbb{R}\), and all \(P \in \mathcal{P}(\mathcal{X})\), \(\rho[a 1 + V, P] = a + \rho[V, P]\);

(iv) It is positively homogeneous, if for all \(V \in \mathbb{B}(\mathcal{X})\), all \(\beta \in \mathbb{R}_+\), and all \(P \in \mathcal{P}(\mathcal{X})\), \(\rho[\beta V, P] = \beta \rho[V, P]\);

(v) It is law invariant if \(V \overset{\rho}{\sim} W\) implies \(\rho[V, P] = \rho[W, P]\);

(vi) It has the support property, if \(\rho\left[\mathbf{1}_{\text{supp}(P)} V, P\right] = \rho[V, P]\) for all \((V, P) \in \mathbb{B}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})\).
An example of a risk form is the expected value, which is a well-understood bilinear form:

$$\mathbb{E}[Z,P] = \int_{\mathbb{X}} Z(x) P(dx).$$

In our analysis, we are interested mainly in forms depending on one or both arguments in a nonlinear way. Our concept of law invariance is broader than that used in the literature, because it allows for the probability measure $P$ to vary. If the risk form is law invariant, then it has the support property, because $\forall x \in \mathbb{X}$.

**Lemma 2.2.** If a risk form $\rho : \mathcal{B}(\mathbb{X}) \times \mathcal{P}(\mathbb{X}) \rightarrow \mathbb{R}$ has the normalization, translation equivariance, and support properties then for every $Z \in \mathcal{B}(\mathbb{X})$ and every $x \in \mathbb{X}$

$$\rho[Z, \delta_x] = Z(x). \tag{1}$$

**Proof.** Using the support property twice, the translation property, and the normalization property, we obtain the chain of equations:

$$\rho[Z, \delta_x] = \rho[1_x Z, \delta_x] = \rho[Z(x) 1, \delta_x] = Z(x) + \rho[0, \delta_x] = Z(x).$$

This property was called state-consistency in [7].

Essential role in our analysis will be played by the increasing convex order (second order stochastic dominance for preference of smaller outcomes).

**Definition 2.3.** A probabilistic model $[Z, P]$ is smaller than a probabilistic model $[Z', P']$ in the increasing convex order, written $[Z, P] \preceq [Z', P']$, if for all $\eta \in \mathbb{R}$

$$\int_{\mathbb{X}} [Z(x) - \eta]_+ P(dx) \leq \int_{\mathbb{X}} [Z'(x) - \eta]_+ P'(dx).$$

This concept allows us to consider risk forms consistent with the increasing order.

**Definition 2.4.** A risk form $\rho : \mathcal{B}(\mathbb{X}) \times \mathcal{P}(\mathbb{X}) \rightarrow \mathbb{R}$ is consistent with the increasing convex order, if

$$[Z, P] \preceq [Z', P'] \implies \rho[Z, P] \leq \rho[Z', P'].$$

Evidently, consistency with the increasing convex order implies monotonicity and law invariance. We call two functions $Z, V \in \mathcal{B}(\mathbb{X})$ comonotonic, if

$$(Z(x') - Z(x))(V(x') - V(x)) \geq 0, \quad \forall x, x' \in \mathbb{X}.$$  

**Definition 2.5.** A risk form $\rho : \mathcal{B}(\mathbb{X}) \times \mathcal{P}(\mathbb{X}) \rightarrow \mathbb{R}$ is comonotonically convex, if for all comonotonic functions $Z, V \in \mathcal{B}(\mathbb{X})$, all $P \in \mathcal{P}(\mathbb{X})$, and all $\lambda \in [0, 1]$,

$$\rho[\lambda Z + (1 - \lambda) \lambda V, P] \leq \lambda \rho[Z, P] + (1 - \lambda) \rho[V, P].$$

3
3 Duality and Kusuoka Representation

In this section, we generalize the Kusuoka representation of [20] to risk forms. In the extant literature, this representation is always derived under the assumption that the probability measure is fixed (see, e.g., [12, 17, 24]). We show that a more general result using variable probability measures is true.

With every stochastic model \( [Z, P] \) we associate its distribution function,
\[
F[Z, P](z) = P[Z \leq z], \quad z \in \mathbb{R},
\]
and its quantile function
\[
\Phi[Z, P](p) = \inf \{ \eta : P[Z \leq \eta] \geq p \}, \quad p \in (0, 1].
\]
The quantile functions are elements of the space \( \mathbb{Q}_b \) of bounded, nondecreasing, and left-continuous functions on \( (0, 1] \).

We make the following assumption.

**Assumption 1.** A function \( Z_0 \in \mathbb{B}(\mathcal{X}) \) and a measure \( P_0 \in \mathcal{P}(\mathcal{X}) \) exist, such that \( \Phi[Z_0, P_0](p) = p \), for all \( p \in (0, 1] \). With no loss of generality, we may assume that \( Z_0(x) > 0 \) for all \( x \in \mathcal{X} \).

In words, an atomless probabilistic model exists in \( \mathbb{B}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \).

We first adapt the general duality result of [6] for risk models on the space of quantile functions. We denote by \( \mathcal{M} \) the set of countably additive finite measures on \( \mathcal{B}(0, 1] \). For every risk form \( \rho : \mathbb{B}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \to \mathbb{R} \), we define the conjugate functional \( \rho^* : \mathcal{M} \to \mathbb{R} \cup \{+\infty\} \) as follows:
\[
\rho^*(\mu) = \sup_{[Z, P] \in \mathbb{B}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})} \left\{ \int_0^1 \Phi[Z, P](p) \mu(dp) - \rho[Z, P] \right\}. \tag{2}
\]

We recall that a total preorder \( \preceq \) on the space \( \mathbb{Q}_b \) is a binary relation, which is reflexive, transitive and complete. It is directed if it satisfies the following conditions:

(i) For any real numbers \( \alpha < \beta \), the relation \( \alpha \preceq \beta \) is true;

(ii) For every \( \Psi \in \mathbb{Q}_b \), numbers \( \alpha \) and \( \beta \) exist such that \( \alpha \preceq \Psi \preceq \beta \).

In [6], we introduced the following properties of preorders:

**Dual Translation:** For all \( \Psi_1 \) and \( \Psi_2 \) in \( \mathbb{Q}_b \) and all \( c \in \mathbb{R} \)
\[
\Psi_1 \preceq \Psi_2 \implies \Psi_1 + c \preceq \Psi_2 + c.
\]

**Dual Monotonicity:** For all \( \Psi_1 \) and \( \Psi_2 \) in \( \mathbb{Q}_b \)
\[
\Psi_1 \preceq \Psi_2 \text{ pointwise } \implies \Psi_1 \preceq \Psi_2.
\]

**Theorem 3.1.** Suppose Assumption [7] is satisfied. If a risk form \( \rho : \mathbb{B}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \to \mathbb{R} \) is normalized, translation equivariant, comonotonically convex, and consistent with the increasing convex order, then a uniquely defined closed convex set
\[
\mathcal{D}_\rho \subseteq \{ \mu \in \mathcal{M} : \mu(0, 1] \text{ is nondecreasing and convex on } (0, 1], \mu(0, 1] = 1 \}
\]
exists, such that for all \( [Z, P] \in \mathbb{B}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \)
\[
\rho[Z, P] = \sup_{\mu \in \mathcal{D}_\rho} \left\{ \int_0^1 \Phi[Z, P](p) \mu(dp) - \rho^*(\mu) \right\}. \tag{4}
\]

If, additionally, the risk form is positively homogeneneous, then \( \rho^*(\mu) = 0 \) for all \( \mu \in \mathcal{D}_\rho \).
Proof. First, we show that the risk form \( \rho[\cdot, \cdot] \) defines a functional \( R \) on the space of quantile functions \( Q_b \) on \( (0, 1) \) by the identity:

\[
\rho[Z, P] = R(\Phi[Z, P]).
\]  

(5)

Indeed, \( \rho[\cdot, \cdot] \) is law invariant, due to its consistency with the increasing convex order. Therefore, if \( \Phi[Z, P] = \Phi[Z', P'] \) then \( \rho[Z, P] = \rho[Z', P'] \). Thus the functional \( R \) is well-defined on the set of quantile functions \( \{ \Phi[Z, P] : Z \in B(\mathcal{X}), P \in \mathcal{P}(\mathcal{X}) \} \). Due to Assumption [1] for every function \( \Psi \in Q_b \), we can define \( Z(x) = \Psi(Z_0(x)), x \in \mathcal{X}, \) and \( \Psi^{-1}(z) = \sup \{ p \in (0, 1) : \Psi(p) \leq z \} \). Then, for every \( z \in \mathbb{R} \),

\[
P_0\{ x : Z(x) \leq z \} = P_0\{ x : \Psi(Z_0(x)) \leq z \} = P_0\{ x : Z_0(x) \leq \Psi^{-1}(z) \} = \Psi^{-1}(z).
\]  

(6)

Consequently, the distribution function of \( Z \) under \( P_0 \) is the inverse of \( \Psi \), and thus \( \Phi[Z, P_0] = \Psi \). This means that the domain of \( R \) is the entire space \( Q_b \).

We verify the assumptions of Theorem 4 of [6]. We define a preference relation \( \preceq \) on \( Q_b \) by setting

\[
\Psi_1 \preceq \Psi_2 \text{ if and only if } R(\Psi_1) \leq R(\Psi_2).
\]

Clearly, the relation \( \preceq \) is a total preorder with \( R \) being its numerical representation. Since \( \rho[\cdot, \cdot] \) is normalized and translation equivariant, the identity (5) implies that \( R \) is normalized and translation equivariant as well. We observe that \( R \) is monotonic, i.e., if \( \Psi_1 \preceq \Psi_2 \) (pointwise), then \( R(\Psi_1) \leq R(\Psi_2) \). Indeed, let \( \Psi_1 \preceq \Psi_2 \) and set \( Z_1(x) = \Psi_1(Z_0(x)), Z_2(x) = \Psi_2(Z_0(x)) \) for all \( x \in \mathcal{X} \). Similar to (6),

\[
P_0(Z_1 \leq z) = \Psi_1^{-1}(z) \geq \Psi_2^{-1}(z) = P_0(Z_2 \leq z)
\]

for all \( z \in \mathbb{R} \). The last relation implies that \( [Z_1, P_0] \preceq [Z_2, P_0] \). The consistency of \( \rho \) with the increasing convex order entails

\[
R(\Psi_1) = \rho[Z_1, P_0] \leq \rho[Z_2, P_0] = R(\Psi_2),
\]

which is the desired monotonicity. The properties of \( R \) further imply that the order \( \preceq \) is directed, monotonic, and satisfies the dual translation property.

For any two comonotonic functions \( Z_1 \) and \( Z_2 \) in \( B(\mathcal{X}) \), any \( \lambda \in [0, 1] \), and any \( P \in \mathcal{P}(\mathcal{X}) \),

\[
\Phi[\lambda Z_1 + (1 - \lambda)Z_2, P] = \lambda \Phi[Z_1, P] + (1 - \lambda) \Phi[Z_2, P].
\]

The comonotonic convexity assumption implies that

\[
R(\lambda \Phi[Z_1, P] + (1 - \lambda) \Phi[Z_2, P]) = \rho[\lambda Z_1 + (1 - \lambda)Z_2, P] \\
\leq \lambda \rho[Z_1, P] + (1 - \lambda) \rho[Z_2, P] = \lambda R(\Phi[Z_1, P]) + (1 - \lambda) R(\Phi[Z_2, P]).
\]

Since any two functions \( \Psi_1, \Psi_2 \in Q_b \) can be represented as

\[
\Psi_1 = \Phi[Z_1, P_0], \text{ with } Z_1(x) = \Psi_1(Z_0(x)), x \in \mathcal{X}, \\
\Psi_2 = \Phi[Z_2, P_0], \text{ with } Z_2(x) = \Psi_2(Z_0(x)), x \in \mathcal{X},
\]

and the functions \( Z_1 \) and \( Z_2 \) are comonotonic by construction, the functional \( R \) is convex.

Consider a function \( \Psi \in Q_b \). For \( (a, b) \subset (0, 1) \) we define

\[
\Psi_{(a,b)}(p) = \begin{cases} 
\frac{1}{b-a} \int_a^b \Psi(s) \, ds & \text{if } p \in (a,b), \\
\Psi(p) & \text{otherwise.}
\end{cases}
\]  

(7)
Directly from (7) we observe that for every \( \alpha \in (0, 1] \)
\[
\int_{1-\alpha}^1 \Psi_{[a,b]}(p) \, dp \leq \int_{1-\alpha}^1 \Psi(p) \, dp.
\]

Therefore, for any \([Z, P]\) and \([V, Q]\) such that \( \Psi_{[a,b]} = \Phi[Z, P] \) and \( \Psi = \Phi[V, Q] \), we have \([Z, P] \preceq [V, Q] \). Due to the consistency of \( \rho[\cdot, \cdot] \) with the increasing convex order,
\[
R(\Psi_{[a,b]}) = \rho[Z, P] \leq \rho[V, Q] = R(\Psi).
\]

Therefore, the preorder \( \preceq \) is risk averse in the sense of [6, Def. 2]. It follows from [6, Th. 4] that a set \( \mathcal{D} \) satisfying (3) exists, such that
\[
R(\Psi) = \sup_{\mu \in \mathcal{D}} \left\{ \int_0^1 \Psi(p) \mu(dp) - R^*(\mu) \right\},
\]
with
\[
R^*(\mu) = \sup_{\Psi \in \mathcal{Q}_b} \left\{ \int_0^1 \Psi(p) \mu(dp) - R(\Psi) \right\}.
\]

Moreover, \( R^*(\mu) = 0 \) for \( \mu \in \mathcal{D} \), if \( R \) is positively homogeneous. The assertion of the theorem follows now from the substitution (5).

Theorem 3.1 allows us to derive a generalization of the celebrated Kusuoka representation of law invariant coherent measures of risk (see [20], [12], and [24, sec. 2.2.4] for an overview of relevant results).

**Definition 3.2.** The Average Value at Risk at level \( \alpha \in [0, 1] \) of a probabilistic model \([Z, P]\) is defined as follows:
\[
\text{AVaR}_\alpha[Z, P] = \begin{cases} 
\frac{1}{\alpha} \int_{1-\alpha}^1 \Phi[Z, P](p) \, dp & \text{if } \alpha \in (0, 1), \\
\Phi[Z, P](1) & \text{if } \alpha = 0, \\
\mathbb{E}[Z, P] & \text{if } \alpha = 1.
\end{cases}
\]

Then, repeating the considerations leading to [6, Cor. 1] verbatim, we obtain the following result.

**Corollary 3.3.** Suppose the conditions of Theorem 3.1 are satisfied and the risk form \( \rho[\cdot, \cdot] \) is positively homogeneous. Then a convex subset \( \Lambda_\rho \) of the set of probability measures on \([0, 1]\) exists, such that for all \([Z, P] \in \mathcal{B}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \)
\[
\rho[Z, P] = \sup_{\lambda \in \Lambda_\rho} \int_0^1 \text{AVaR}_\alpha[Z, P] \lambda(ds).
\] (8)

It is worth stressing that in the extant literature, the Kusuoka representation was derived for probabilistic models with a fixed atomless probability measure \( P \). Our approach proves the validity of the Kusuoka representation for models with arbitrary probability measure \( P \), as long as an atomless model exists. The set \( \Lambda \) is the same for all models \([Z, P]\).
4  The risk disintegration formula

Our main interest is measuring risk on product spaces. Consider two Polish spaces $X$ and $Y$ and their corresponding Borel $\sigma$-algebras $\mathcal{B}(X)$ and $\mathcal{B}(Y)$. We can disintegrate any $P \in \mathcal{P}(X \times Y)$ into its marginal $P_x \in \mathcal{P}(X)$ and a transition kernel $P_{y|x} : X \to \mathcal{P}(Y)$ as follows: $P(dx,dy) = P_x(dx)P_{y|x}(dy|x)$.

Let $\mathcal{D}(Y|X)$ be the space of all measurable mappings $Q : X \to \mathcal{P}(Y)$ (transition kernels). For any measure $\lambda \in \mathcal{P}(X)$ and any kernel $Q \in \mathcal{D}(Y|X)$, the measure $P = \lambda \circ Q$ defined as $P(dx,dy) = \lambda(dx)Q(dy|x)$ is an element of $\mathcal{P}(X \times Y)$.

Suppose the risk form $\rho : \mathcal{B}(X \times Y) \times \mathcal{P}(X \times Y) \to \mathbb{R}$ is monotonic, translation equivariant, and normalized. Then it induces a mapping $\rho_{y|x} : \mathcal{B}(X) \times \mathcal{P}(Y|X) \to \mathcal{B}(X)$ defined as follows:

$$\rho_{y|x}[Z,Q](x) = \rho[Z,\delta_x \circ Q], \quad x \in X.$$  \hspace{1cm} (9)

We call the mapping $\rho_{y|x}[\cdot,\cdot]$ the conditional risk form associated with $\rho[\cdot,\cdot]$.

To verify that those values are indeed elements of $\mathcal{B}(X)$ let $c \in \mathbb{R}$ be such that $Z \leq c \mathbf{1}$. Then, by monotonicity, translation equivariance and normalization,

$$\rho[Z,\delta_x \circ Q] \leq \rho[c \mathbf{1},\delta_x \circ Q] = c.$$

The lower bound is similar, and thus the function $\rho_{y|x}[Z,Q]$ is bounded. It is measurable, as a composition of measurable mappings.

If the risk form $\rho$ has the support property, then for each $x \in X$ the value of $\rho_{y|x}$ depends only on the function $Z(x,\cdot) \in \mathcal{B}(Y)$ and the measure $Q(x) \in \mathcal{P}(Y)$. We can, therefore, define the functionals $\rho_{y|x} : \mathcal{B}(Y) \times \mathcal{P}(Y) \to \mathbb{R}, x \in X$, as follows:

$$\rho_{y|x}[Z(x,\cdot),Q(x)] = \rho_{y|x}[Z,Q](x), \quad x \in X.$$  \hspace{1cm} (10)

We call them conditional risk forms associated with $\rho[\cdot,\cdot]$. Observe that any function from $\mathcal{B}(Y)$ and any measure from $\mathcal{P}(Y)$ may feature as arguments of $\rho_{y|x}[\cdot,\cdot]$.

From now on we always assume that the risk forms in question have the support property. The inequalities “$\leq$” between functions are always understood point-wise.

**Lemma 4.1.** If the risk form $\rho[\cdot,\cdot]$ is monotonic (normalized, translation equivariant), then, for every $x \in X$, the conditional risk form $\rho_{y|x}$ is monotonic (normalized, translation equivariant).

**Proof.** All the properties follow directly from the equation

$$\rho_{y|x}[Z(x,\cdot),Q(x)] = \rho[Z,\delta_x \circ Q],$$

which defines the conditional risk form. \hfill \Box

**Definition 4.2.** A risk form $\rho : \mathcal{B}(X \times Y) \times \mathcal{P}(X \times Y) \to \mathbb{R}$ is conditionally consistent if for all $Z,Z' \in \mathcal{B}(X \times Y)$ and all $Q,Q' \in \mathcal{D}(Y|X)$ the inequality

$$\rho_{y|x}[Z,Q] \leq \rho_{y|x}[Z',Q'],$$

implies that

$$\rho[Z,\lambda \circ Q] \leq \rho[Z',\lambda \circ Q'], \quad \forall \lambda \in \mathcal{P}(X).$$

7
The following result is the foundation of our further considerations.

**Theorem 4.3.** Suppose a risk form \( \rho : \mathcal{B}(\mathcal{X} \times \mathcal{Y}) \times \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \to \mathbb{R} \) is monotonic, normalized, translation equivariant, has the support property, and is conditionally consistent. Then a risk form \( \rho_\mathcal{X} : \mathcal{B}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \to \mathbb{R} \) exists, such that for all \( [Z, P] \in \mathcal{B}(\mathcal{X} \times \mathcal{Y}) \times \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \) the following formula is true:

\[
\rho[Z, P] = \rho_\mathcal{X}[\rho_{\mathcal{X} \mid \mathcal{X}}[Z, P_{\mathcal{X} \mid \mathcal{X}}], P_\mathcal{X}].
\]  

(11)

The risk form \( \rho_\mathcal{X} \) is uniquely defined by the equation

\[
\rho_\mathcal{X}[f, P_\mathcal{X}] = \rho[f, P], \quad \text{with} \quad f(x, y) \equiv f(x).
\]  

(12)

It is monotonic, normalized, translation equivariant, and has the support property.

**Proof.** Let us verify (11). Suppose \([Z, P]\) and \([Z', P']\) are such that

\[
\rho_{\mathcal{X} \mid \mathcal{X}}[Z, P_{\mathcal{X} \mid \mathcal{X}}] = \rho_{\mathcal{X} \mid \mathcal{X}}[Z', P'_{\mathcal{X} \mid \mathcal{X}}].
\]  

Then it follows from Definition 4.2 that \( \rho[Z, \lambda \circ P_{\mathcal{X} \mid \mathcal{X}}] = \rho[Z', \lambda \circ P'_{\mathcal{X} \mid \mathcal{X}}] \) for all \( \lambda \in \mathcal{P}(\mathcal{X}) \). If, additionally, the marginal measures \( P_\mathcal{X} \) and \( P'_\mathcal{X} \) are identical, by setting \( \lambda = P_\mathcal{X} = P'_\mathcal{X} \) we conclude that

\[
\rho[Z, P] = \rho[Z, P_\mathcal{X} \circ P_{\mathcal{X} \mid \mathcal{X}}] = \rho[Z', P'_\mathcal{X} \circ P'_{\mathcal{X} \mid \mathcal{X}}] = \rho[Z', P'].
\]

It follows that the value of \( \rho[Z, P] \) is fully determined by the value of the conditional risk operator \( \rho_{\mathcal{X} \mid \mathcal{X}}[Z, P_{\mathcal{X} \mid \mathcal{X}}] \) and the marginal measure \( P_\mathcal{X} \). Therefore, the disintegration formula (11) is true.

It remains to verify the properties of \( \rho_\mathcal{X} \). Set \( Z(x, y) = f(x, y) \equiv f(x) \) in (11). Then, by the support, translation equivariance, and normalization properties of \( \rho[\cdot, \cdot] \), for every \( x \in \mathcal{X} \) we obtain

\[
\rho_{\mathcal{X} \mid \mathcal{X}}[f, P_{\mathcal{X} \mid \mathcal{X}}](x) = \rho[f(x) \mathbb{1}, \delta_x \circ P_{\mathcal{X} \mid \mathcal{X}}] = f(x).
\]

Combining with (11) we observe that the identity (12) is true. All the postulated properties of \( \rho_\mathcal{X}[\cdot, \cdot] \) follow from the corresponding properties of \( \rho[\cdot, \cdot] \).

We call the identity (11) the **risk disintegration formula.** It represents the risk form \( \rho[\cdot, \cdot] \) by its marginal risk form \( \rho_\mathcal{X}[\cdot, \cdot] \) and its conditional risk operator \( \rho_{\mathcal{X} \mid \mathcal{X}}[\cdot, \cdot] \).

### 5 Composite disintegration

We now generalize our results to the product of multiple Polish spaces \( \mathcal{X}_j, j = 1, \ldots, n \), with their corresponding Borel \( \sigma \)-algebras \( \mathcal{B}(\mathcal{X}_j) \), \( j = 1, \ldots, n \). For a nonempty set of indices \( J \subseteq \{1, \ldots, n\} \), we write

\[
\mathcal{X}_J = \bigtimes_{j \in J} \mathcal{X}_j, \quad \text{and} \quad J^c = \{1, \ldots, n\} \setminus J.
\]

Let \( P \) be a probability measure on \( \mathcal{X} = \bigtimes_{j=1}^n \mathcal{X}_j \). For every \( J \) such that \( J^c \neq \emptyset \), we can disintegrate \( P \) into its marginal \( P_{\mathcal{X}_J} \in \mathcal{P}(\mathcal{X}_J) \) and a transition kernel \( P_{\mathcal{X}_J \mid \mathcal{X}_J} : \mathcal{X}_J \to \mathcal{P}(\mathcal{X}_J^c) \) as follows:

\[
P(dx_J, dx_{J^c}) = P_{\mathcal{X}_J}(dx_J) P_{\mathcal{X}_J \mid \mathcal{X}_J}(dx_{J^c} | x_J).
\]
For the case \( J = \emptyset \), trivially \( P(dx_J) = P(dx) \). We denote the set of transition kernels from \( \mathcal{X}_J \) to \( \mathcal{P}(\mathcal{X}_J) \) by \( \mathcal{D}_{\mathcal{X}_J|\mathcal{X}_J} \).

Exactly as in section \( 4 \), a monotonic, translation equivariant, and normalized risk form \( \rho : \mathcal{B}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \to \mathbb{R} \) on the product space induces a family of conditional risk operators \( \rho_{\mathcal{X}_J|\mathcal{X}_J} : \mathcal{B}(\mathcal{X}) \times \mathcal{D}_{\mathcal{X}_J|\mathcal{X}_J} \to \mathcal{B}(\mathcal{X}_J) \), as follows:

\[
\rho_{\mathcal{X}_J|\mathcal{X}_J}[Z, Q](x_J) = \rho[Z, \delta_x \circ Q], \quad x_J \in \mathcal{X}_J.
\]

If the risk form \( \rho \) has the support property, then for each \( x_J \in \mathcal{X}_J \) the value of (12) depends only on the function \( Z(x_J, \cdot) \in \mathcal{B}(\mathcal{X}_J) \) and the measure \( Q(x_J) \in \mathcal{P}(\mathcal{X}_J) \). As in section \( 4 \), we define the conditional risk forms functionals \( \rho_{\mathcal{X}_J|\mathcal{X}_J} : \mathcal{B}(\mathcal{X}_J) \times \mathcal{P}(\mathcal{X}_J) \to \mathbb{R}, \) as follows:

\[
\rho_{\mathcal{X}_J|\mathcal{X}_J}[Z(x_J, \cdot), Q(x_J)] = \rho_{\mathcal{X}_J|\mathcal{X}_J}[Z, Q](x_J), \quad x_J \in \mathcal{X}_J.
\]

**Definition 5.1.** A risk form \( \rho : \mathcal{B}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \to \mathbb{R} \) is strongly conditionally consistent if for all nonempty index sets \( J \subseteq \{1, \ldots, n\} \), all \( Z, Z' \in \mathcal{B}(\mathcal{X}) \) and all \( Q, Q' \in \mathcal{D}_{\mathcal{X}_J|\mathcal{X}_J} \) the inequality

\[
\rho_{\mathcal{X}_J|\mathcal{X}_J}[Z, Q] \leq \rho_{\mathcal{X}_J|\mathcal{X}_J}[Z', Q'],
\]

implies that

\[
\rho[Z, \lambda \circ Q] \leq \rho[Z', \lambda \circ Q'], \quad \forall \lambda \in \mathcal{P}(\mathcal{X}_J).
\]

Note that this property is stronger than that of Definition 4.1 even for the case of two spaces only, because both spaces may feature as \( \mathcal{X}_J \). It also implies monotonicity.

**Lemma 5.2.** Suppose the risk form \( \rho \) has the normalization, translation, and support properties. If it is strongly conditionally consistent then it is monotonic.

**Proof.** Setting \( J = \{1, \ldots, n\} \) in Definition 5.1 and using Lemma 2.2, we conclude that \( Z(x) \leq Z'(x) \) for all \( x \in \mathcal{X} \) implies that \( \rho[Z, \lambda] \leq \rho[Z', \lambda] \) for all \( \lambda \in \mathcal{P}(\mathcal{X}) \).

The following corollary results directly from Theorem 4.3 with the additional observation that monotonicity is guaranteed by Lemma 5.2.

**Corollary 5.3.** If a risk form \( \rho : \mathcal{B}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \to \mathbb{R} \) is normalized, translation equivariant, and strongly conditionally consistent then for every nonempty \( J \subset \{1, \ldots, n\} \) a risk form \( \rho_{\mathcal{X}_J} : \mathcal{B}(\mathcal{X}_J) \times \mathcal{P}(\mathcal{X}_J) \to \mathbb{R} \) exists, such that for all \( [Z, P] \in \mathcal{B}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \) the following formula holds:

\[
\rho[Z, P] = \rho_{\mathcal{X}_J}[\rho_{\mathcal{X}_J|\mathcal{X}_J}[Z, P]\rho_{\mathcal{X}_J|\mathcal{X}_J], P_{\mathcal{X}_J}],
\]

where the marginal risk form \( \rho_{\mathcal{X}_J} \) is uniquely defined by the equation (12) with \( (\mathcal{X}_J, \mathcal{X}_J) \) replacing \( (\mathcal{X}, \mathcal{Y}) \). It is monotonic, normalized, translation equivariant, and has the support property.

A question arises what is the relation between the marginal and conditional risk forms for different sets \( J \).

**Theorem 5.4.** If a risk form \( \rho : \mathcal{B}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \to \mathbb{R} \) is normalized, translation equivariant, and strongly conditionally consistent, then for every \( J \subset \{1, \ldots, n\} \) the following statements are true:

(i) For all \( x_J \in \mathcal{X}_J \) the conditional risk forms \( \rho_{\mathcal{X}_J|\mathcal{X}_J} \) are monotonic, normalized, translation equivariant, and strongly conditionally consistent;

(ii) For all \( K \subset J^c \) and all \( x_{J\cup K} \in \mathcal{X}_{J\cup K} \) we have \( \rho_{\mathcal{X}_J|\mathcal{X}_J} = \rho_{\mathcal{X}_{J\cup K}|\mathcal{X}_{J\cup K}} \).
(iii) The marginal risk form $\rho_{x_J}$ is monotonic, normalized, translation equivariant, and strongly conditionally consistent;
(iv) For all $K \subset J$ and all $x_K \in \mathcal{X}_K$ we have $\rho_{x_K} = (\rho_{x_J})_{x_K}$.

**Proof.** The monotonicity, normalization, and translation equivariance of the marginal and conditional forms follow from the corresponding properties of $\rho$ via formula (15) by considering special classes of functions in $\mathbb{B} (\mathcal{X})$: the functions that depend only on $x_J$ (for the marginal), and the functions that depend only on $x_J^c$ (for the conditionals). The proof is identical to the the last part of proof of Theorem 4.3.

It remains to prove strong conditional consistency and the tower formulae (ii) and (iv). For a fixed $x_J \in \mathcal{X}_J$, we shall verify Definition 5.1 for the conditional risk form $\rho_{x_J^C | x_J}$. Let $K \subset J^c$, and let $Z, Z' \in \mathbb{B} (\mathcal{X}_{J^c})$ and $Q, Q' \in \mathcal{D}_{\mathcal{X}_{J^c} | \mathcal{X}_K}$. Suppose

$$
\rho_{x_J^C | x_J} [Z, \delta_{x_K} \circ Q] \leq \rho_{x_J^C | x_J} [Z', \delta_{x_K} \circ Q'], \quad \forall x_K \in \mathcal{X}_K.
$$

(16)

We can formally extend the functions $Z$ and $Z'$ to the entire domain $\mathcal{X}$ by setting $\check{Z}(x_J, x_{J^c}) = Z(x_J)$ and $\check{Z}'(x_J, x_{J^c}) = Z'(x_J)$. We can also define the kernels $\check{Q}$ and $\check{Q}'$ in $\mathcal{D}_{\mathcal{X}_{J^c} | \mathcal{X}_{J \cup K}}$ by setting $\check{Q}(x_J, x_K) = Q(x_K)$ and $\check{Q}'(x_J, x_K) = Q'(x_K)$. Then

$$
\rho_{x_J^C | x_J} [Z, \delta_{x_K} \circ Q] = \rho [\check{Z}, \delta_{x_{J \cup K}} \circ \check{Q}],
$$

A similar equation is true for $Z'$ and $Q'$. Then (16) can be written as follows:

$$
\rho [\check{Z}, \delta_{x_J} \circ \check{Q}] \leq \rho [\check{Z}', \delta_{x_J} \circ \check{Q}].
$$

By the strong conditional consistency of $\rho$,

$$
\rho [\check{Z}, \psi \circ \check{Q}] \leq \rho [\check{Z}', \psi \circ \check{Q}'], \quad \forall \psi \in \mathcal{P} (\mathcal{X}_J).
$$

Let $\lambda \in \mathcal{P} (\mathcal{X}_K)$. By setting $\psi = \delta_{x_J} \circ \lambda$ in the last displayed inequality, and using the fact that $\check{Z}, \check{Z}', \check{Q}$, and $\check{Q}'$ do not depend on $x_J$, we conclude that

$$
\rho_{x_J | x_J} [Z, \lambda \circ Q] \leq \rho_{x_J | x_J} [Z', \lambda \circ Q'].
$$

This proves the strong conditional consistency of the conditional risk form $\rho_{x_J | x_J}$.

To verify (ii), consider $f \in \mathbb{B} (\mathcal{X}_{J \cup K})$, $Q \in \mathcal{D}_{\mathcal{X}_{J \cup K} | \mathcal{X}_K}$, and the natural extension $\check{f}$ of $f$ to the entire space $\mathcal{X}$, defined by $\check{f}(x_J, x_{J \cup K}) = f(x_J)$. We obtain the chain of equalities:

$$
\rho_{x_J | x_J} [f, Q] = \rho [\check{f}, \delta_{x_{J \cup K}} \circ Q] = \rho [\check{f}, \delta_{x_J} \circ \delta_{x_K} \circ Q] = \rho_{x_J^C | x_J} [\check{f}, \delta_{x_J} \circ \delta_{x_K} \circ Q] = (\rho_{x_J | x_J})_{x_K} [f, Q].
$$

Consider now the marginal risk form $\rho_{x_J}$. Let $K \subset J$, and $Q, Q' \in \mathcal{D}_{\mathcal{X}_{J \cup K} | \mathcal{X}_K}$. Suppose

$$
\rho_{x_J} [Z, \delta_{x_K} \circ Q] \leq \rho_{x_J} [Z', \delta_{x_K} \circ Q'], \quad \forall x_K \in \mathcal{X}_K.
$$

(17)

where $Z, Z' \in \mathbb{B} (\mathcal{X}_{J^c})$. We can formally extend the functions $Z$ and $Z'$ to the entire domain $\mathcal{X}$ by setting $\check{Z}(x_J, x_{J^c}) = Z(x_J)$ and $\check{Z}'(x_J, x_{J^c}) = Z'(x_J)$. We can also define the kernels $\check{Q}$ and $\check{Q}'$ in $\mathcal{D}_{\mathcal{X}_{J^c} | \mathcal{X}_K}$ by setting
From (12), we obtain the chain of equalities:

\[ \rho[\bar{Z}, \delta_{x_K} \circ \bar{Q}] \leq \rho[Z', \delta_{x_K} \circ \bar{Q}'], \quad \forall x_K \in \mathcal{X}_K. \]

By the strong conditional consistency of \( \rho \),

\[ \rho[\bar{Z}, \lambda \circ \bar{Q}] \leq \rho[Z', \lambda \circ \bar{Q}'], \quad \forall \lambda \in \mathcal{P}(\mathcal{X}_K). \]

Since \( \bar{Z} \) and \( \bar{Z}' \) do not depend on \( x_{J'} \), we conclude that

\[ \rho_{\mathcal{X}_J}[Z, \lambda \circ \bar{Q}] \leq \rho_{\mathcal{X}_J}[Z', \lambda \circ \bar{Q}'], \]

which proves the strong conditional consistency of the marginal risk form \( \rho_{\mathcal{X}_J} \).

It remains to verify the tower property (iv). For \( f \in \mathcal{B}(\mathcal{X}_K) \), we define

\[ f(x_K, x_{K'}) = f(x_K). \]

From (12), we obtain the chain of equalities:

\[ \rho_{\mathcal{X}_J}[f, P_{\mathcal{X}_J}] = \rho[f, P] = \rho_{\mathcal{X}_J}[\underline{f}, P_{\mathcal{X}_J}] = (\rho_{\mathcal{X}_J} \mathcal{X}_K)[f, (P_{\mathcal{X}_J})_{\mathcal{X}_K}] = (\rho_{\mathcal{X}_J} \mathcal{X}_K)[f, P_{\mathcal{X}_K}], \]

which is (iv). \( \square \)

6 Risk in Two-Stage Partially Observable Systems

6.1 Fixed Observation Distribution

Let us start from the following simple setting. A random vector \((X, Y)\) is distributed in the product of Polish spaces \( \mathcal{X} \times \mathcal{Y} \) according to a measure \( P \). For a bounded measurable function \( c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \), we can evaluate the risk of \( c(X, Y) \) by a risk form \( \rho[c, P] \).

However, we know that we shall be able to observe the value of \( X \). After \( X \) is observed, we might refine our risk evaluation of \( c(X, Y) \). Thus, a question arises: if a future possibility to observe \( X \) exists, what should be our present evaluation of the risk of \( c(X, Y) \), before \( X \) is observed. Note that \( Y \) is never observed.

Let us start with the problem of risk evaluation after \( X \) is observed. We can disintegrate \( P \) into its marginal \( P_{\mathcal{X}} \) on \( \mathcal{X} \) and a transition kernel \( P_{\mathcal{Y}|\mathcal{X}} \) from \( \mathcal{X} \) to \( \mathcal{P}(\mathcal{Y}) \):

\[ P(dx, dy) = P_{\mathcal{X}}(dx)P_{\mathcal{Y}|\mathcal{X}}(dy|x). \]

Suppose the risk form \( \rho \) is monotonic, normalized, has the translation property and the support property. Then the correct evaluation of the risk after \( X = x \) is observed is

\[ \rho_{\mathcal{Y}|\mathcal{X}}[c(x, \cdot), P_{\mathcal{Y}|\mathcal{X}}(x)] = \rho [c, \delta_x \circ P_{\mathcal{Y}|\mathcal{X}}]. \quad (18) \]

This is nothing else, but the conditional risk form defined in (10). As a function of \( x \), we obtain the conditional risk operator \( \rho_{\mathcal{Y}|\mathcal{X}}[c, P_{\mathcal{Y}|\mathcal{X}}] \). Now, to evaluate the overall risk, we calculate \( \rho_{\mathcal{X}}[\rho_{\mathcal{Y}|\mathcal{X}}[c, P_{\mathcal{Y}|\mathcal{X}}], P_{\mathcal{X}}] \).
We thus arrive to the following conclusion from Theorem 4.3. If the risk form \( \rho : \mathcal{B}(\mathcal{X} \times \mathcal{Y}) \times \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \to \mathbb{R} \) is normalized, translation equivariant, conditionally consistent, and has the support property, then

\[
\rho[c, P] = \rho_\mathcal{X} \left[ \rho_\mathcal{Y}\big|_\mathcal{X} \left[ c, P_\mathcal{Y}|_\mathcal{X} \right], P_\mathcal{X}\right].
\]

(19)

It follows that the two risk evaluations: without and with the perspective of inspection, are identical. The mere existence of inspection does not affect risk.

When a possibility of control exists, the situation is different. Suppose there are two Polish spaces \( \mathcal{Y}_1 \) and \( \mathcal{Y}_2 \), which we call control spaces. At stage one, a control \( u \in U_1 \subset \mathcal{Y}_1 \) is chosen, where \( U_1 \) is a subset of \( \mathcal{Y}_1 \). Then an observation of \( X \) is made. After observing the value of \( X \), we choose control \( u_2 \in U_2(X, u_1) \subset \mathcal{Y}_2 \) to minimize the risk of \( c(X, Y, u_1, u_2) \). The risk is measured by the form \( \rho[\cdot, \cdot] \). Here \( U_2 : \mathcal{X} \times \mathcal{Y}_1 \Rightarrow \mathcal{Y}_2 \) is a measurable multifunction representing the feasible set at the second stage. We shall use the symbol \( \pi(\cdot) \leq U_2(\cdot, u_1) \) to indicate that the function \( \pi \) is a measurable selection of \( U_2(\cdot, u_1) \).

We may look at this problem from two perspectives. Let us start from the functional perspective. Since \( u_2 \) can be chosen after \( X \) is observed, we may represent it as a measurable function: \( u_2 = \pi(x), x \in \mathcal{X} \). Therefore, the overall cost has the form:

\[
Z^{u_1, \pi}(x, y) = c(x, y, u_1, \pi(x)), \quad (x, y) \in \mathcal{X} \times \mathcal{Y}.
\]

(20)

The distribution of \( (X, Y) \) is \( P \). The problem takes on the form

\[
\min_{u_1, \pi} \rho \left[ Z^{u_1, \pi}, P \right]
\]

s.t. \( u_1 \in U_1 \),

\[
\pi(\cdot) \leq U_2(\cdot, u_1).
\]

(21)

We now derive a two-stage representation of this problem.

**Theorem 6.1.** Let the following assumptions be satisfied:

(i) The risk form \( \rho \) is monotonic, normalized, translation equivariant, has the support property, and is conditionally consistent;

(ii) The multifunction \( U_2 \) is upper-semicontinuous and has nonempty and compact values;

(iii) The function \( c \) is uniformly bounded, measurable, and lower-semicontinuous with respect to its second argument.

Then problem (21) is equivalent to the two-stage problem:

\[
\min_{u_1 \in U_1} \rho_\mathcal{X} \left[ V(\cdot, u_1), P_\mathcal{X} \right],
\]

(22)

where \( V(\cdot, \cdot) \) is the optimal value of the second stage problem:

\[
V(x, u_1) = \min_{u_2 \in U_2(x, u_1)} \rho_\mathcal{Y}\big|_\mathcal{X} \left[ c(x, \cdot, u_1, u_2), P_\mathcal{Y}|_\mathcal{X} (x) \right], \quad x \in \mathcal{X}, \quad u_1 \in U_1.
\]

(23)

**Proof.** Since \( \rho[\cdot, \cdot] \) satisfies the assumptions of Theorem 4.3, the risk of the function \( Z^{u_1, \pi} \) can be calculated by the risk disintegration formula:

\[
\rho \left[ Z^{u_1, \pi}, P \right] = \rho_\mathcal{X} \left[ \rho_\mathcal{Y}\big|_\mathcal{X} \left[ Z^{u_1, \pi}, P_\mathcal{Y}|_\mathcal{X} \right], P_\mathcal{X} \right] = \rho_\mathcal{X} \left[ x \mapsto \rho_\mathcal{Y}\big|_\mathcal{X} \left[ c(x, \cdot, u_1, \pi(x)), P_\mathcal{Y}|_\mathcal{X} (x) \right], P_\mathcal{X} \right].
\]
Then problem (21) takes on the form:

\[
\min_{u_1, \pi} \rho_{\mathcal{X}} \left[ x \mapsto \rho_{\mathcal{Y}|\mathcal{X}} \left[ c(x, \cdot, u_1, \pi(x)), P_{\mathcal{Y}|\mathcal{X}}(x) \right], P_{\mathcal{X}} \right],
\]
subject to the same constraints. Owing to the monotonicity of the marginal risk form \( \rho_{\mathcal{X}} \), the smaller the values of the function \( x \mapsto \rho_{\mathcal{Y}|\mathcal{X}} \left[ c(x, \cdot, u_1, \pi(x)), P_{\mathcal{Y}|\mathcal{X}}(x) \right] \), the smaller the value of \( \rho_{\mathcal{X}} \). Since \( u_2 = \pi(x) \) may depend on \( x \), we may carry out the minimization with respect to \( u_2 \) inside the argument of \( \rho_{\mathcal{X}} \):

\[
\inf_{u_1 \in U_1} \inf_{\pi(x) \in U_2(x, u_1)} \rho_{\mathcal{X}} \left[ x \mapsto \rho_{\mathcal{Y}|\mathcal{X}} \left[ c(x, \cdot, u_1, \pi(x)), P_{\mathcal{Y}|\mathcal{X}}(x) \right], P_{\mathcal{X}} \right]
= \inf_{u_1 \in U_1} \rho_{\mathcal{X}} \left[ x \mapsto \inf_{u_2 \in U_2(x, u_1)} \rho_{\mathcal{Y}|\mathcal{X}} \left[ c(x, \cdot, u_1, u_2), P_{\mathcal{Y}|\mathcal{X}}(x) \right], P_{\mathcal{X}} \right].
\]

The only condition for the validity of this transformation is the measurability and boundedness of the optimal value function (23). This follows from Berge’s theorem, which can be applied due to the assumptions (ii) and (iii). In fact, they also guarantee that the optimal value function is lower semicontinuous with respect to \( u_1 \).

We conclude that problem (21) reduces to the marginal risk optimization (22). \( \square \)

Theorem 6.1 provides us with the second perspective on the problem. It has a hierarchical structure, similar to its expected-value full information counterpart: after \( X = x \) is observed, the second stage problem (23) is to minimize the conditional risk. Then, the first stage problem takes on the form of minimizing the marginal risk (22) of the second-stage optimal value. The most important conclusion is that the risk disintegration formula allows us to write the overall problem in a hierarchical structure. The extended two-stage risk-averse model, which is introduced and analyzed in (21) (see also (14)) is a special case of this problem.

### 6.2 Controlled Observation Distribution

Now we consider a more complex situation. There are still two Polish control spaces \( \mathcal{U}_1 \) and \( \mathcal{U}_2 \). However, after a control \( u_1 \in U_1 \subset \mathcal{U}_1 \) is chosen, the distribution of the observation \( X \) depends on \( u_1 \). The dependence is described by a controlled kernel \( K : \mathcal{Y} \times \mathcal{U}_1 \to \mathcal{P}(\mathcal{X}) \). After observing \( X \), we choose control \( u_2 \in U_2(X, u_1) \subset \mathcal{U}_2 \) to minimize the risk of \( c(X, Y, u_1, u_2) \). The risk is measured by the form \( \rho(\cdot, \cdot) \).

Assume the same conditions on \( U_1, U_2 \) and \( \rho \) as in the previous subsection. Let \( P_Y \) be the marginal distribution of \( Y \). After the first decision \( u_1 \) will be chosen, the joint distribution of \( (Y, X) \) will become

\[
M(u_1) = P_Y \circ K(\cdot, u_1),
\]
that is, \( M(dy, dx|u_1) = P_Y(dy)K(dx|y, u_1) \). Therefore, denoting the second stage decision by \( u_2 = \pi(x) \) (it may depend on \( x \)), our problem is to find

\[
\min_{u_1, \pi} \rho \left[ Z^{u_1, \pi}, M(u_1) \right],
\]

s.t. \( u_1 \in U_1, \pi(\cdot) \in U_2(\cdot, u_1), \) (24)

where the function \( Z^{u_1, \pi}(\cdot, \cdot) \) is given by (20).
Let us develop a two-stage version of the functional problem \((24)\). The marginal distribution of the observation result is

\[
M_\mathcal{Y}(u_1) = \int_{\mathcal{Y}} K(y, u_1) \, dP_{\mathcal{Y}}(dy)
\]

where the integral is understood in the weak sense. Since the space \(\mathcal{Y}\) is standard, the measure \(M(u_1)\) can be disintegrated into the marginal \(M_\mathcal{Y}(u_1)\) and a transition kernel \(\Gamma : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{P}(\mathcal{Y})\) as follows

\[
M(u_1) = M_\mathcal{Y}(u_1) \circ \Gamma(u_1),
\]

which reads \(M(dx, dy|u_1) = M_\mathcal{Y}(dx|u_1)\Gamma(dy|x, u_1)\). The transition kernel \(\Gamma\) is called the Bayes operator.

**Example 6.2.** Assume that the joint distribution \(M(u_1)\) of \((X,Y)\) has a density \(q(\cdot, \cdot \mid u_1)\) with respect to a finite product measure \(\mu_\mathcal{X} \otimes \mu_\mathcal{Y}\) on \(\mathcal{X} \times \mathcal{Y}\). Then the Bayes operator has the form

\[
\Gamma(A|x, u_1) = \frac{\int_A \int_{\mathcal{Y}} q(x', y' \mid u) M_\mathcal{Y}(dy) \, d\mu_\mathcal{Y}(dy')}{\int_{\mathcal{Y}} \int_{\mathcal{Y}} q(x', y' \mid u) M_\mathcal{Y}(dy) \, d\mu_\mathcal{Y}(dy')} \quad \forall A \in \mathcal{B}(\mathcal{Y}).
\]

If the formula above has a zero denominator, we can formally define \(\Gamma(x,u_1)\) to be an arbitrarily selected distribution on \(\mathcal{Y}\).

With the use of the Bayes operator, we can equivalently write problem \((24)\) as a two-stage problem.

**Theorem 6.3.** Let the following assumptions be satisfied:

(i) The risk form \(\rho\) is monotonic, normalized, translation equivariant, conditionally consistent, and has the support property;
(ii) The multifunction \(U_2\) is upper-semicontinuous and has nonempty and compact values;
(iii) The function \(c\) is uniformly bounded, measurable, and lower-semicontinuous with respect to its second argument.

Then problem \((24)\) is equivalent to the two-stage problem:

\[
\min_{u_1 \in U_1} \rho_\mathcal{X}[V(\cdot, u_1), P_\mathcal{X}],
\]

where \(V(\cdot, \cdot)\) is the optimal value of the second stage problem:

\[
V(x, u_1) = \min_{u_2 \in U_2(x,u_1)} \rho_{\mathcal{Y}|x}[c(x, \cdot, u_1, u_2), \Gamma(x, u_1)],
\]

\[x \in \mathcal{X}, \quad u_1 \in U_1.\]

**Proof.** With the use of the Bayes formula, we can write problem \((24)\) as follows:

\[
\min_{u_1 \in U_1} \min_{\pi(\cdot) \in U_2(\cdot, u_1)} \rho\left[Z^{u_1, \pi}, P_\mathcal{X}(u_1) \circ \Gamma(u_1)\right].
\]

Since the risk form \(\rho\) satisfies the assumptions of Theorem 4.3, we can disintegrate it to obtain the following equivalent form:

\[
\min_{u_1 \in U_1} \min_{\pi(\cdot) \in U_2(\cdot, u_1)} \rho_\mathcal{X}\left[x \mapsto \rho_{\mathcal{Y}|x}[c(x, \cdot, u_1, \pi(x)), \Gamma(x, u_1)], P_\mathcal{X}(u_1)\right].
\]

The remaining considerations are the same as in the proof of Theorem 6.1. Due to the monotonicity of the marginal risk form \(\rho_\mathcal{X}\), the smaller the values of the function \(x \mapsto \rho_{\mathcal{Y}|x}[c(x, \cdot, u_1, \pi(x)), \Gamma(x, u_1)]\), the
smaller the value of $\rho_x$. Since $u_2 = \pi(x)$ may depend on $x$, we may carry out the minimization with respect to $u_2$ inside the argument of $\rho_x$:

$$
\inf_{u_1 \in U_1} \inf_{\pi(\cdot) \in U_2(x, u_1)} \rho_x \left[ x \mapsto \rho_{\mathcal{Y}} \left[ c(x, \cdot, u_1, \pi(x)), \Gamma(x, u_1) \right], P_x(u_1) \right]
$$

$$
= \inf_{u_1 \in U_1} \rho_x \left[ x \mapsto \inf_{u_2 \in U_2(x, u_1)} \rho_{\mathcal{Y}} \left[ c(x, \cdot, u_1, u_2), \Gamma(x, u_1) \right], P_x(u_1) \right].
$$

The only difference is that the new marginal distribution and the Bayes operator are the disintegration components of the probability measure and feature in the risk disintegration formula. The “$\inf$” operation in the second stage problem can be replaced by “$\min$” because of conditions (ii) and (iii).

**References**

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

[2] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, and H. Ku. Coherent multiperiod risk adjusted values and Bellman’s principle. *Annals of Operations Research*, 152:5–22, 2007.

[3] T. R. Bielecki, I. Cialenco, and M. Pitera. A survey of time consistency of dynamic risk measures and dynamic performance measures in discrete time: $L^m$-measure perspective. *Probability, Uncertainty and Quantitative Risk*, 2(1):3, 2017.

[4] P. Cheridito, F. Delbaen, and M. Kupper. Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability*, 11:57–106, 2006.

[5] P. Cheridito and M. Kupper. Composition of time-consistent dynamic monetary risk measures in discrete time. *International Journal of Theoretical and Applied Finance*, 14(01):137–162, 2011.

[6] D. Dentcheva and A. Ruszczyński. Risk preferences on the space of quantile functions. *Mathematical Programming*, 148(1-2):181–200, 2014.

[7] D. Dentcheva and A. Ruszczyński. Risk-averse control of continuous-time Markov chains. In *2017 Proceedings of the Conference on Control and its Applications*, pages 78–85. SIAM, 2017.

[8] J. Fan and A. Ruszczyński. Risk measurement and risk-averse control of partially observable discrete-time Markov systems. *Mathematical Methods of Operations Research*, pages 1–24, 2018.

[9] H. Föllmer and I. Penner. Convex risk measures and the dynamics of their penalty functions. *Statistics & Decisions*, 24(1/2006):61–96, 2006.

[10] H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6:429–447, 2002.

[11] H. Föllmer and A. Schied. *Stochastic Finance*. Walter de Gruyter & Co., Berlin, 2011.

[12] M. Frittelli and E. Rosazza Gianin. Law invariant convex risk measures. In *Advances in mathematical economics. Volume 7, volume 7 of Adv. Math. Econ.*, pages 33–46. Springer, Tokyo, 2005.
[13] J.E. Littlewood G.H. Hardy and G. Pólya. Inequalities. Cambridge University Press, Cambridge, MA, 1934.

[14] S. Gülten and A. Ruszczyński. Two-stage portfolio optimization with higher-order conditional measures of risk. Annals of Operations Research, 229(1):409–427, 2015.

[15] A. Jobert and L. C. G. Rogers. Valuations and dynamic convex risk measures. Mathematical Finance, 18(1):1–22, 2008.

[16] T. W. Jonsbråten, R. J.-B. Wets, and D. L. Woodruff. A class of stochastic programs with decision dependent random elements. Annals of Operations Research, 82:83–106, 1998.

[17] E. Jouini, W. Schachermayer, and N. Touzi. Law invariant risk measures have the Fatou property. In Advances in mathematical economics. Vol. 9, volume 9 of Adv. Math. Econ., pages 49–71. Springer, Tokyo, 2006.

[18] M. Kijima and M. Ohnishi. Mean-risk analysis of risk aversion and wealth effects on optimal portfolios with multiple investment opportunities. Ann. Oper. Res., 45:147–163, 1993.

[19] S. Klöppel and M. Schweizer. Dynamic indifference valuation via convex risk measures. Math. Finance, 17(4):599–627, 2007.

[20] S. Kusuoka. On law-invariant coherent risk measures. In Kusuoka S. and Maruyama T., editors, Advances in Mathematical Economics, Vol. 3, pages 83–95. Springer, Tokyo, 2001.

[21] N. Miller and A. Ruszczyński. Risk-averse two-stage stochastic linear programming: Modeling and decomposition. Operations Research, 59(1):125–132, 2011.

[22] W. Ogryczak and A. Ruszczyński. From stochastic dominance to mean-risk models: Semideviations as risk measures. European Journal of Operational Research, 116(1):33–50, 1999.

[23] G.Ch. Pflug and A. Pichler. Multistage Stochastic Optimization. Springer, 2016.

[24] G.Ch. Pflug and W. Römisch. Modeling, Measuring and Managing Risk. World Scientific, Singapore, 2007.

[25] F. Riedel. Dynamic coherent risk measures. Stochastic Processes and Their Applications, 112:185–200, 2004.

[26] B. Roorda, J. M. Schumacher, and J. Engwerda. Coherent acceptability measures in multiperiod models. Mathematical Finance, 15(4):589–612, 2005.

[27] A. Ruszczyński. Risk-averse dynamic programming for Markov decision processes. Mathematical programming, 125(2):235–261, 2010.

[28] A. Ruszczyński and A. Shapiro. Conditional risk mappings. Mathematics of Operations Research, 31:544–561, 2006.

[29] A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. Mathematics of Operations Research, 31:433–542, 2006.
[30] G. Scandolo. *Risk Measures in a Dynamic Setting*. PhD thesis, Università degli Studi di Milano, Milan, Italy, 2003.

[31] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming*. SIAM, Philadelphia, Pennsylvania, 2009.