J. Math. Anal. Appl. **157**(1991), 211–236

Certainty Equivalents and Information Measures: Duality and Extremal Principles *

Aharon Ben-Tal[†] Adi Ben-Israel[‡] Marc Teboulle[§]

February 1, 1988
Revised March 1, 1989

Abstract

Given a convex function $\phi : \mathbb{R}_+ \to \mathbb{R}$, the **Csiszár ϕ-divergence** (Csiszár (1978)) is a function $I_\phi : \mathbb{R}^n_+ \times \mathbb{R}^n_+ \to \mathbb{R}$,

$$I_\phi(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^{n} q_i \, \phi(\frac{p_i}{q_i})$$

I_ϕ is a generalized measure of entropy, Aczél (1984), whose distance-like properties make it useful in stochastic optimization (Ben-Tal and Teboulle (1987)) and other applications (Teboulle (1988)).

We establish relations between I_ϕ and three **certainty equivalents**, the **expected utility** (EU) (Von-Neumann and Morgenstern (1947)), the **recourse certainty equivalent** (RCE) (Ben-Tal and Ben-Israel (1991)) and **Yaari's certainty equivalent** (YCE) (Yaari (1987)). These relations provide a duality framework for **economics of uncertainty** and **entropy**, giving new interpretations and extremal principles for I_ϕ, EU, RCE and YCE.

Key words: Decision-making under uncertainty. Certainty equivalents. Entropy. Csiszár's ϕ-divergence. Conjugate functions. Duality. Insurance.

*Supported by the National Science Foundation Grant ECS-8604354.

[†]Faculty of Industrial Engineering and Management, Technion-Israel Institute of Technology, Haifa 32000, Israel.

[‡]RUTCOR-Rutgers Center for Operations Research and Department of Mathematics, Rutgers University, New Brunswick, NJ 08903.

[§]Department of Mathematics and Statistics, University of Maryland, Baltimore County Campus, Baltimore, MD 21228. The research of this author was supported by AFORS Grant 0218-88 and NSF grant ECS-8802239

1 Introduction

Decision making under uncertainty presupposes the ability to rank random variables (RV's) by associating with a RV X a constant CE(X), the **certainty equivalent** (CE) of X. The simplest such CE is the **expected value** of X, $E\,X$. In **expected utility** (EU) theory, the CE is $E\,u(X)$, where $u(\cdot)$ is the decision maker's **utility** function [1].

Expected utility, as a paradigm for decisions under uncertainty, has been criticized on many grounds, see e.g. Machina (1987), and references therein. Recently various alternatives to EU have been suggested in the literature, including the the **Yaari certainty equivalent** (YCE), and the authors' **recourse certainty equivalent** (RCE) (Ben-Tal and Ben-Israel (1988), Ben-Tal and Teboulle (1986), discussed below.

Outside and independently of the economics of uncertainty, the need to deal with uncertainty in other fields gave rise to different methods and tools. In information theory the main tools are **entropy** (Shannon (1948)), **relative entropy** (Kullback and Leibler (1951)) and the more general notion of **divergence** (DIV for short), Csiszár (1978), a cardinal measure of "distance" between RV's.

In this paper we establish relations between three objects used in the study of uncertainty, namely E (expectation), CE (certainty equivalent) and DIV (divergence). Each member of the triple {E, CE, DIV} is the optimal value of an extremum problem involving (the sum or difference of) the other two. These extremal principles have concrete economic interpretations. They allow a duality theory for the models of economics under uncertainty, new interpretations of entropy and a unified framework for the above CE's.

[1]A more natural CE in EU theory is $u^{-1}\,E\,u(X)$, given in the units of X and not in "utils"

We turn now to specifics. Let the RV X, assuming values

$$x_1, x_2, \ldots, x_n$$

with probabilities

$$\text{prob}\ \{X = x_i\} = p_i, \quad i = 1, \ldots, n$$

be denoted by

$$X = [\mathbf{x}, \mathbf{p}] \tag{1.1}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{p} = (p_1, \ldots, p_n)$. The three CE's mentioned above are:

1. The **expected utility** (EU), given a **utility** u,

$$Eu(X) := \sum_{i=1}^{n} p_i\, u(x_i). \tag{1.2}$$

2. The **recourse certainty equivalent** (RCE) (Ben-Tal and Teboulle (1986), (1987) and Ben-Tal and Ben-Israel (1991))

$$S_u(X) := \sup_z \{z + \sum_{i=1}^{n} p_i\, u(x_i - z)\}, \tag{1.3}$$

given a **value risk function** u. [2].

The (RCE) was introduced by Ben-Tal and Teboulle (1986) ,(1987) under the name New Certainty Equivalent. For an interpretation, see Section 4.2, and for further details and applications see Ben-Tal and Ben-Israel (1988). 3. **Yaari's certainty equivalent** (YCE) (Yaari (1987)),

$$Y_f([\mathbf{x}, \mathbf{p}]) := \sum_{i=1}^{n} f(\bar{F}_i)\, \Delta x_i \tag{1.4}$$

given a function $f : [0, 1] \to \mathbb{R}$. Here

$$\bar{F}_i := \sum_{j=i}^{n} p_j$$

and

$$\Delta x_i := x_i - x_{i-1},\ i = 1, \ldots, n$$

For interpretations of Yaari's Certainty Equivalent see Yaari (1987) and Röell (1987).

Next comes the divergence measure. Let ϕ be a real valued function defined and convex on the nonnegative real line.

[2]The use of the same letter u for **utility** in (1.2) and **value risk function** in (1.3) facilitates the comparison between $Eu(X)$ and $S_u(X)$, see Theorems 4.1 and 5.1 below.

The **Csiszár ϕ-divergence**, Csiszár (1978), is a function $I_\phi : \mathbb{R}^n_+ \times \mathbb{R}^n_+ \to \mathbb{R}$,

$$I_\phi(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^{n} q_i\, \phi(\frac{p_i}{q_i}) \tag{1.5}$$

having certain properties of a "distance" between $\mathbf{p}$ and $\mathbf{q}$, but I_ϕ is not a metric: The triangle inequality does not hold, and in general, for $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n_{++}$,

$$I_\phi(\mathbf{p}, \mathbf{q}) \neq I_\phi(\mathbf{q}, \mathbf{p}),$$ [3]

However, for any $\phi : \mathbb{R}_+ \to \mathbb{R}$ which is convex and **normalized** i.e.

$$\phi(1) = 0 \tag{1.6}$$

the **adjoint** function $\phi^\diamond$, defined by,

$$\phi^\diamond(t) = t\, \phi(\frac{1}{t}), \quad \forall t \in \mathbb{R}_+ \tag{1.7}$$

is convex, normalized, and satisfies

$$I_\phi(\mathbf{p}, \mathbf{q}) = I_{\phi^\diamond}(\mathbf{q}, \mathbf{p}), \quad \forall \mathbf{p}, \mathbf{q} \in \mathbb{R}^n_{++} \tag{1.8}$$

The needed properties of the ϕ-divergence, and the adjoint function, are proved in Section 2. We also require some results from convex analysis, collected in Section 3.

Sections 4, 5 and 6 contain the main results, which are briefly described below.

Let $\mathbb{P}^n$ denote the **n-dimensional probability vectors**,

$$\mathbb{P}^n := \{\mathbf{p} = (p_i) : \mathbf{p} \in \mathbb{R}^n_+, \sum_{i=1}^{n} p_i = 1\}, \tag{1.9}$$

and let $\mathbb{P}^n_{++}$ denote the **positive** probability vectors,

$$\mathbb{P}^n_{++} := \{\mathbf{p} \in \mathbb{P}^n : p_i > 0,\ \forall i\}. \tag{1.10}$$

Also let $u : \mathbb{R} \to \mathbb{R}$ be closed, concave, and strictly increasing, and let

$$\phi = -u_*, \tag{1.11}$$

where u_* is the **concave conjugate** of u, Rockafellar (1970), see Section 3 for definition. Then for any RV $X = [\mathbf{x}, \mathbf{p}]$, the RCE S_u is,

$$S_u([\mathbf{x}, \mathbf{p}]) = \inf_{\mathbf{q} \in \mathbb{P}^n} \{I_\phi(\mathbf{q}, \mathbf{p}) + \sum_{i=1}^{n} q_i\, x_i\} \tag{1.12}$$

[3]Thus I_ϕ is a **directed divergence** from $\mathbf{p}$ to $\mathbf{q}$, but we call it **divergence** for short.

A corresponding result, in terms of the directed divergence from $\mathbf{p}$ to $\mathbf{q}$ is

$$S_u([\mathbf{x},\mathbf{p}]) = \inf_{\mathbf{q}\in\mathbb{P}^n} \{I_{\phi^\diamond}(\mathbf{p},\mathbf{q}) + \sum_{i=1}^n q_i\, x_i\} \tag{1.13}$$

where the adjoint $\phi^\diamond$ is related to u through the **convex conjugate** of its inverse,

$$\phi^\diamond = (u^{-1})^*, \tag{1.14}$$

see Theorem 4.1. An analogous result for the EU is

$$Eu([\mathbf{x},\mathbf{p}]) = \inf_{\mathbf{q}\in\mathbb{R}^n_+} \{I_\phi(\mathbf{q},\mathbf{p}) + \sum_{i=1}^n q_i\, x_i\}, \tag{1.15}$$

see Theorem 5.1. A comparison of (1.12) and (1.15) shows a surprising similarity between EU and RCE, in spite of the marked difference in their implications for economic behavior under uncertainty (see Ben-Tal and Ben-Israel (1991)).

Yaari's certainty equivalent (1.4), see § 6, is simi larly described, by a "dual" extremal principle involving expectation and divergence. Let $f : [0,1] \to [0,1]$ be strictly increasing, closed and convex. Then

$$Y_f([\mathbf{x},\mathbf{p}]) = \sup_{\mathbf{\Delta y}\in\mathbb{R}^n_+} \{\sum_{i=1}^n p_i\, y_i - I_\phi(\mathbf{\Delta y},\mathbf{\Delta x})\}\ ^4 \tag{1.16}$$

where $\phi = f^*$, the convex conjugate of f.

The converse results, expressing I_ϕ in terms of RCE and EU, are given in Theorems 4.3 and 5.2. First the RCE,

$$I_\phi(\mathbf{q},\mathbf{p}) = \sup_{x\in\mathbb{R}^n} \{S_u([\mathbf{x},\mathbf{p}]) - \sum q_i\, x_i\} \tag{1.17}$$

and then the EU,

$$I_\phi(\mathbf{q},\mathbf{p}) = \sup_{\mathbf{x}\in\mathbb{R}^n} \{E_u([\mathbf{x},\mathbf{p}]) - \sum_{i=1}^n q_i\, x_i\} \tag{1.18}$$

These results give I_ϕ as the **optimal value** of an **insurance plan**, for an RCE maximizer and an EU maximizer, respectively. In fact, I_ϕ has another representation in terms of RCE, as the optimal insurance plan under budget constraints, see Theorem 4.2.

[4] Note the analogy to (1.12) and (1.13), which can also be written in terms of $\mathbf{\Delta P}$ and $\mathbf{\Delta Q}$ instead of $\mathbf{p}$ and $\mathbf{q}$, where $\mathbf{P}$ and $\mathbf{Q}$ are the **cumulative probability vectors**, $\mathbf{P} = (P_i = \sum_{j=1}^i p_i)$ and $\mathbf{Q} = (Q_i = \sum_{j=1}^i q_i)$.

A "dual" extremal principle, giving the ϕ-divergence between difference vectors $\mathbf{\Delta y}$ and $\mathbf{\Delta x}$ in terms of Yaari's certainty equivalent, appears in Theorem 6.2:

$$I_\phi(\mathbf{\Delta y},\mathbf{\Delta x}) = \sup_{\mathbf{p}\in\mathbb{R}^n_+} \{\sum_i y_i\, p_i - Y_f([\mathbf{x},\mathbf{p}])\} \tag{1.19}$$

where $f(t) = \phi^*(t)$.

It thus appears that Csiszár's ϕ-divergence, whose appeal has so far been restricted to information theory and statistics, is a fundamental concept in economics of uncertainty, providing a unified framework for the certainty equivalents considered here, and perhaps others. For the convenience of the reader we collect below a list of acronysms frequently used in the sequel:

RV :	Random Variable(s)
EU :	Expected Utility
CE :	Certainty Equivalent
DIV :	Divergence
RCE :	Recourse Certainty Equivalent
YCE :	Yaari Certainty Equivalent

2 The ϕ-divergence functional

Given a convex function $\phi : \mathbb{R}_+ \to \mathbb{R}$, the **$\phi$-divergence functional**

$$I_\phi(\mathbf{p},\mathbf{q}) := \sum_{i=1}^n q_i\, \phi(\frac{p_i}{q_i}) \tag{2.1}$$

was introduced in Csiszár (1978) as a generalized measure of information, a "distance function" on the set of probability distributions $\mathbb{P}^n$. The restriction here to discrete distributions is only for convenience, similar results hold for general distributions.

The basic properties of I_ϕ are proved here under minimal assumptions.

As in Csiszár (1967) we interpret undefined expressions by

$$\begin{aligned} \phi(0) &= \lim_{t\to 0^+} \phi(t) \\ 0\phi(\frac{0}{0}) &= 0 \\ 0\,\phi(\frac{a}{0}) &= \lim_{\epsilon\to 0} \epsilon\,\phi(\frac{a}{\epsilon}) \\ &= a \lim_{t\to\infty} \frac{\phi(t)}{t}, \quad a > 0. \end{aligned}$$

The following results (Lemmas 2.1, 2.2 and Corollary 2.1) were essentially given by Csiszár and Korner (1982). We include the proofs for completeness.

Lemma 2.1 (Joint convexity) *If $\phi : \mathbb{R}_+ \to \mathbb{R}$ is convex, then $I_\phi(\mathbf{p}, \mathbf{q})$ is jointly convex in $\mathbf{p}$ and $\mathbf{q}$.*

Proof. Let $f(x, y) = y\,\phi(\frac{x}{y})$. It suffices to show that $f(x, y)$ is jointly convex in x and y, i.e. for any (x, y), $(u, v) \in \mathbb{R}^2_+$, $0 < \lambda < 1$,

$$(\lambda y + \bar\lambda v)\,\phi(\frac{\lambda x + \bar\lambda u}{\lambda y + \bar\lambda v}) \le \lambda\, y\, \phi(\frac{x}{y}) + \bar\lambda\, v\, \phi(\frac{u}{v}) \tag{2.2}$$

where $\bar\lambda := 1 - \lambda$. By the convexity of ϕ,

$$\begin{aligned} \phi(\frac{\lambda x + \bar\lambda u}{\lambda y + \bar\lambda v}) &= \phi(\frac{\lambda y}{\lambda y + \bar\lambda v}\frac{x}{y} + \frac{\bar\lambda v}{\lambda y + \bar\lambda v}\frac{u}{v}) \\ &\le \frac{\lambda}{\lambda y + \bar\lambda v}\, y\,\phi(\frac{x}{y}) + \frac{\bar\lambda}{\lambda y + \bar\lambda v}\, v\,\phi(\frac{u}{v}) \end{aligned}$$

which, when multiplied by $\lambda y + \bar\lambda v$, gives (2.2). □

Lemma 2.2 *Let $\phi : \mathbb{R}_+ \to \mathbb{R}$ be convex. Then for any $\mathbf{p}$, $\mathbf{q} \in \mathbb{R}^n_+$,*

$$I_\phi(\mathbf{p}, \mathbf{q}) \ge (\sum_{i=1}^n q_i)\,\phi\left(\frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n q_i}\right) \tag{2.3}$$

If ϕ is strictly convex, equality holds in (2.3) iff

$$\frac{p_1}{q_1} = \frac{p_2}{q_2} = \cdots = \frac{p_n}{q_n} \tag{2.4}$$

Proof. By Jensen's inequality, for $\lambda = (\lambda_i) \in \mathbb{P}^n$,

$$\sum_{i=1}^n \lambda_i\,\phi(x_i) \ge \phi(\sum_{i=1}^n \lambda_i\, x_i) \tag{2.5}$$

with equality, for strictly convex ϕ, iff $x_1 = x_2 = \cdots = x_n$.

Choose

$$\lambda_i = \frac{q_i}{\sum\, q_j}, \quad x_i = \frac{p_i}{q_i}$$

then (2.5) gives

$$\frac{1}{\sum\, q_j}\left(\sum q_i\,\phi\left(\frac{p_i}{q_i}\right)\right) \ge \phi\left(\frac{\sum p_i}{\sum q_j}\right)$$

which is (2.3). □

Corollary 2.1 (Nonnegativity) *Let $\phi : \mathbb{R}_+ \to \mathbb{R}$ be convex and normalized, i.e.,*

$$\phi(1) = 0 \tag{2.6}$$

Then for any $\mathbf{p}$, $\mathbf{q} \in \mathbb{R}^n_+$ with

$$\sum_{i=1}^n p_i = \sum_{j=1}^n q_j \tag{2.7}$$

we have the inequality

$$I_\phi(\mathbf{p}, \mathbf{q}) \ge 0 \tag{2.8}$$

If ϕ is strictly convex, equality holds in (2.8) iff

$$p_i = q_i,\ \forall i. \tag{2.9}$$

Proof. By Lemma 2.2,

$$I_\phi(\mathbf{p}, \mathbf{q}) \ge (\sum q_j)\phi(1) = 0$$

with equality iff (2.4), which under (2.7) is equivalent to (2.9). □

In particular, if $\mathbf{p}, \mathbf{q}$ are probability vectors, then (2.7) is assured. Corollary 2.1 then shows, for strictly convex and normalized $\phi : \mathbb{R}_+ \to \mathbb{R}$,

$$\begin{aligned} I_\phi(\mathbf{p}, \mathbf{q}) &\ge 0 \quad \forall \mathbf{p}, \mathbf{q} \in \mathbb{P}^n, & (2.10) \\ &= 0 \quad \text{iff } \mathbf{p} = \mathbf{q} & (2.11) \end{aligned}$$

These are "distance properties". However I_ϕ is not a metric: It violates the triangle inequality, and is **assymetric**, i.e. for general $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n_+$,

$$I_\phi(\mathbf{p}, \mathbf{q}) \ne I_\phi(\mathbf{q}, \mathbf{p}) \tag{2.12}$$

However, symmetry holds for the divergence $I_\phi + I_{\phi^\diamond}$, where $\phi^\diamond$ is the adjoint of ϕ. Properties of $\phi^\diamond$ are collected in the following lemma, whose proof is omitted.

Lemma 2.3 (Properties of the Adjoint) *Let $\phi : \mathbb{R}_+ \to \mathbb{R}$ be convex, $\phi(1) = 0$, and define the* **adjoint** *of ϕ as*

$$\phi^\diamond(t) := t\,\phi(\frac{1}{t}), \quad t > 0. \tag{2.13}$$

Then:
(a) $\phi^{\diamond\diamond} = \phi$.
(b) $\phi^\diamond(\cdot)$ is convex.
(c) $\phi^\diamond(1) = 0$
(d) For any $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n_{++}$

$$I_\phi(\mathbf{p}, \mathbf{q}) = I_{\phi^\diamond}(\mathbf{q}, \mathbf{p}) \quad \square \tag{2.14}$$

In the examples below we get, for suitable choices of the kernel ϕ, some of the best known distance functions I_ϕ used in mathematical statistics (Justice (1986), Kapur (1984)), information theory (Burbea and Rao (1982), Gallager (1968), Shannon (1948)) and signal processing (Frieden (1975), Leahy and Goutis (1986)). For such ϕ we also write the adjoint $\phi^\diamond$.

Example 2.1 (Kullback-Leibler) *For*

$$\phi(t) := t \log t, \quad t > 0, \tag{2.15}$$

the ϕ-divergence is

$$I_\phi(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}, \tag{2.16}$$

the **Kullback-Leibler distance**, *Kullback (1959), Kullback and Leibler (1951). The adjoint of ϕ is*

$$\phi^\diamond = t(\frac{1}{t} \log \frac{1}{t}) = -\log t \tag{2.17}$$

In particular, for

$$\mathbf{p} = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$$

the $\phi^\diamond$-divergence $I_{\phi^\diamond}$ gives the negative **Shannon entropy** *of $\mathbf{q} = (q_1, q_2, \ldots, q_n)$, Shannon (1948), plus the constant $\log n$,*

$$\begin{aligned} I_{\phi^\diamond}(\mathbf{p}, \mathbf{q}) &= -\sum_{i=1}^{n} q_i \log \frac{1}{n q_i} \\ &= \sum_{i=1}^{n} q_i \log q_i + \log n \end{aligned} \tag{2.18}$$

Example 2.2 (Hellinger) *Let*

$$\phi(t) := (1 - \sqrt{t})^2, \quad t > 0. \tag{2.19}$$

Then I_ϕ gives the **Hellinger distance**, *Beran (1977),*

$$I_\phi(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} (\sqrt{p_i} - \sqrt{q_i})^2 \tag{2.20}$$

which is symmetric. Here ϕ is self-adjoint,

$$\phi^\diamond = \phi$$

Example 2.3 (Renyi) *For $\alpha > 1$ let,*

$$\phi(t) := t^\alpha, \quad t > 0. \tag{2.21}$$

Then

$$I_\phi(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} (p_i)^\alpha (q_i)^{1-\alpha} \tag{2.22}$$

*the α-***order entropy**, *Renyi (1961). The adjoint is*

$$\phi^\diamond(t) = t^{1-\alpha} \tag{2.23}$$

Example 2.4 *Let*

$$\phi(t) := (t-1)^2, \quad t > 0, \tag{2.24}$$

with

$$I_\phi(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} \left(\frac{p_i^2 - q_i^2}{q_i} \right) \tag{2.25}$$

the **χ^2-distance**. *The adjoint is*

$$\begin{aligned} \phi^\diamond(t) &= (\sqrt{t} - \frac{1}{\sqrt{t}})^2 \\ &= t + \frac{1}{t} - 2 \end{aligned}$$

Example 2.5 *Let*

$$\phi(t) := \mid t - 1 \mid, \quad t > 0, \tag{2.26}$$

which is self-adjoint, $\phi^\diamond = \phi$. The corrseponding divergence, called the **variation distance**, *is symmetric*

$$I_\phi(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} \mid p_i - q_i \mid \tag{2.27}$$

Example 2.6 (Indicator function) *Let $\alpha < 1 < \beta$ ($\alpha \leq 0$) and let*

$$\phi(t) := \begin{cases} 0 & \text{if } \alpha \leq t \leq \beta, \\ \infty & \text{otherwise,} \end{cases} \tag{2.28}$$

the **indicator function** *of the interval $[\alpha, \beta]$, denoted by $\delta(t \mid [\alpha, \beta])$. The adjoint $\phi^\diamond$ is likewise an indicator function,*

$$\phi^\diamond(t) = \delta(t \mid \left[\frac{1}{\beta}, \frac{1}{\alpha}\right]).$$

The corresponding divergence is

$$I_\phi(\mathbf{p}, \mathbf{q}) = \begin{cases} 0 & \text{if } \alpha \leq \frac{p_i}{q_i} \leq \beta, \quad \forall i = 1, \ldots, n \\ \infty & \text{otherwise} \end{cases} \tag{2.29}$$

3 Convex analysis

This section contains required results from convex analysis. For a function $f : \mathbb{R}^n \to \mathbb{R}$, we recall the definitions of **conjugate** functions, Rockafellar (1970):

$$\begin{aligned} f^*(y) &:= \sup_x \{\mathbf{y}\cdot\mathbf{x} - f(\mathbf{x})\} \quad (3.1) \\ &\qquad \text{the } \textbf{convex conjugate}, \\ f_*(y) &:= \inf_x \{\mathbf{y}\cdot\mathbf{x} - f(\mathbf{x})\} \quad (3.2) \\ &\qquad \text{the } \textbf{concave conjugate}\ . \end{aligned}$$

The conjugacy operation is fundamental to derive duality results. For various interpretations and applications the reader is refered to Rocakfellar (1970) and Roberts and Varberg (1973).
The following lemma, whose proof is omitted, is useful in the sequel.

Lemma 3.1

(a) ϕ *convex* $\Longrightarrow (-\phi)_*(x) = -\phi^*(-x),\ \forall x,$
(b) u *concave* $\Longrightarrow (-u)^*(x) = -u_*(-x),\ \forall x.$ □

Some odd properties of conjugates and inverses are collected in the following theorem. Of particular interest is (the seemingly new) part (d), giving the conjugate of the inverse function.

Theorem 3.1 *Let* $u : \mathbb{R} \to \mathbb{R}$ *be strictly increasing and concave. Then:*
(a) dom $u_* \subset \mathbb{R}_+$
(b) dom $(u^{-1})^* \subset \mathbb{R}_+$
(c) $y \in$ *dom* $(u^{-1})^* \Longleftrightarrow \frac{1}{y} \in$ *dom* u_* *, for all* $y > 0$
(d) For all $y > 0$

$$\begin{aligned} (u^{-1})^*(y) &= -y\, u_*(\frac{1}{y}) \quad (3.3) \\ &= -(u_*)^\diamond(y), \quad \textit{see (2.13).} \end{aligned}$$

Proof. (a) By Rockafellar (1970), §24,

$$\text{ri (dom } u_*) \subset \text{range } \partial u \subset \text{dom } u_*$$

(where ri denotes relative interior and ∂u is the subdifferential of u) and by the monotonicity of u,

$$\begin{aligned} \text{range } \partial u &\subset \mathbb{R}_{++} \\ \therefore\ \text{ri (dom } u_*) &\subset \mathbb{R}_{++} \\ \therefore\ \text{dom } u_* &\subset \mathbb{R}_{+} \end{aligned}$$

(b) Similarly proved.
(c) Suppose

$$\begin{aligned} \frac{1}{y} \in \text{dom } u_* &:= \{x : u_*(x) > -\infty\} \\ \text{but } y \notin \text{dom } (u^{-1})^* &:= \{t : (u^{-1})^*(t) < \infty\} \end{aligned}$$

Then for any $M > 0$ there exists $s = s(M)$ such that

$$sy - u^{-1}(s) > M$$

Let $t = u^{-1}(s)$.

$$\begin{aligned} \therefore\ u(t)y - t &> M \\ \frac{t}{y} - u(t) &< -\frac{M}{y} \quad \text{(since } y > 0) \\ \therefore\ \frac{1}{y} &\notin \text{dom } u_*, \quad \text{a contradiction.} \end{aligned}$$

The converse is proved similarly.
(d) By definition of the convex conjugate we have :

$$\begin{aligned} (u^{-1})^*(y) &= \sup_t\{t\,y - u^{-1}(t)\} \\ &= \sup_{t,\,s}\{t\,y - s : s = u^{-1}(t)\} \\ &= \sup_s\{y\,u(s) - s\} \\ &= -y \inf_s \{\frac{s}{y} - u(s)\} = -y\,u_*(\frac{1}{y}) \\ &\qquad \text{for all } y > 0\ . \quad \square \end{aligned}$$

An illustrative application of Theorem 3.1(d) is the following result, describing the inverse function as an extremal value.

Corollary 3.1 *Let* $u : \mathbb{R} \to \mathbb{R}$ *be strictly increasing and concave. Then*

$$u^{-1}(t) = \sup_{x>0} \{x\,t - (u_*)^\diamond(x)\} \quad (3.4)$$

Proof. Taking the convex conjugate of both sides of (3.3) and using (2.13) the result follows. □

Example 3.1 *Let*

$$u(t) = \log t - t, \quad t > 0.$$

The inverse u^{-1} *cannot be computed analytically, however we can calculate its conjugate* $(u^{-1})^*$*. Indeed, the concave conjugate of* u *is*

$$u_*(x) = 1 + \log\,(1 + x), \quad (x > -1)$$

and, by (3.3),

$$(u^{-1})^*(x) = x \log \frac{x}{1+x} - x, \quad (x > -1)$$

A "companion" of Theorem 3.1(a) is the following:

Lemma 3.2 *Let $\phi : \mathbb{R} \to \mathbb{R}$ be convex. Then*
(a) dom $\phi \subseteq \mathbb{R}_+ \Longrightarrow \phi^*$ is nondecreasing.
(b) dom $\phi \subseteq \mathbb{R}_{++} \Longrightarrow \phi^*$ is increasing.

Proof. (a) Let $y_2 > y_1$. Then, by the gradient inequality for the convex function ϕ^*,

$$\phi^*(y_2) - \phi^*(y_1) \geq (y_2 - y_1)\, y^*, \quad \forall y^* \in \partial\phi^*(y_1) \subset \text{range } \partial\phi^* \tag{3.5}$$

But

$$\text{ri (dom } \phi) \subset \text{range } \partial\phi^* \subset \text{dom } \phi \subseteq \mathbb{R}_+$$

For $y^* \in \partial\phi^*(y_1)$ we therefore have $y^* \geq 0$, and by (3.5),

$$\phi^*(y_2) \geq \phi^*(y_1)$$

(b) Similarly proved. □.

Remark. If in Lemma 3.2(a) ϕ is essentially smooth in $\mathbb{R}_+$ (see Rockafellar (1970)), then ϕ^* is strictly increasing.

In the following theorem we introduce a functional (3.6), needed in the sequel, but of interest in itself.

Theorem 3.2 *For $h : \mathbb{R}^n \to \mathbb{R}$ define $h_+ : \mathbb{R}^n \to \mathbb{R}$ by*

$$h_+(\mathbf{y}) := \sup_{\substack{z \in \mathbb{R} \\ \mathbf{x} \in \mathbb{R}^n}} \{z + h(\mathbf{x} - z\mathbf{e}) : \ \mathbf{y} \cdot \mathbf{x} = 0\} \tag{3.6}$$

where $\mathbf{e}^T = (1, 1, \ldots, 1)$. Then for any $\mathbf{y} \in \mathbb{R}^n$ with $\sum y_i \neq 0$,

$$h_+(\mathbf{y}) = -h_*(\frac{\mathbf{y}}{\sum y_i}) \tag{3.7}$$

Proof. Let $\mathbf{v} := \frac{\mathbf{y}}{\sum y_i}$. Then

$$-h_*(\mathbf{v}) = -\inf_{\mathbf{u} \in \mathbb{R}^n} \{\mathbf{v} \cdot \mathbf{u} - h(\mathbf{u})\} \tag{3.8}$$

For any $\mathbf{u} \in \mathbb{R}^n$ let

$$z = -\mathbf{v} \cdot \mathbf{u} \tag{3.9}$$
$$\mathbf{x} = \mathbf{u} + z\mathbf{e} \tag{3.10}$$

Then $\mathbf{v} \cdot \mathbf{x} = 0$ since $\mathbf{e} \cdot \mathbf{v} = 1$.

$$\begin{aligned} \therefore \ -h_*(\mathbf{v}) &= -\inf_{\substack{z \in \mathbb{R} \\ \mathbf{x} \in \mathbb{R}^n \\ \mathbf{v} \cdot \mathbf{x} = 0}} \{-z - h(\mathbf{x} - z\mathbf{e})\}, \\ &= \sup_{\substack{z \in \mathbb{R} \\ \mathbf{x} \in \mathbb{R}^n \\ \mathbf{v} \cdot \mathbf{x} = 0}} \{z + h(\mathbf{x} - z\mathbf{e})\}. \quad \square \end{aligned}$$

If h is strictly concave and continuously differentiable, we can get explicit expressions for the optimal z^* and $\mathbf{x}^*$ in Theorem 3.2. Indeed, by differentiating the infimand in (3.8), it follows that the optimal $\mathbf{u}^*$ satisfies

$$\begin{aligned} \mathbf{v} &= \nabla h(\mathbf{u}^*) \\ \therefore \ \mathbf{u}^* &= (\nabla h)^{-1}(\mathbf{v}) \\ &= (\nabla h)^{-1}(\frac{\mathbf{y}}{\mathbf{e} \cdot \mathbf{y}}) \end{aligned}$$

Therefore, by (3.9) and (3.10), the optimal z^* and $\mathbf{x}^*$ are

$$\begin{aligned} z^* &= -\mathbf{v}^T\mathbf{u}^* = -\frac{\mathbf{y}^T}{\mathbf{e} \cdot \mathbf{y}} (\nabla h)^{-1} (\frac{\mathbf{y}}{\mathbf{e} \cdot \mathbf{y}}) \\ \mathbf{x}^* &= \mathbf{u}^* + z^*\mathbf{e} = \left(I - \frac{\mathbf{e}\mathbf{y}^T}{\mathbf{e}^T\mathbf{y}}\right) (\nabla h)^{-1} (\frac{\mathbf{y}}{\mathbf{e} \cdot \mathbf{y}}) \\ &= P_{\mathbf{e}^\perp} (\nabla h)^{-1} (\frac{\mathbf{y}}{\mathbf{e} \cdot \mathbf{y}}) \end{aligned}$$

where $P_{\mathbf{e}^\perp}$ is the orthogonal projection perpendicular to $\mathbf{e}$. Similarly,

$$-z^*\mathbf{e} = \frac{\mathbf{e}\mathbf{y}^T}{\mathbf{e}^T\mathbf{y}} (\nabla h)^{-1} (\frac{\mathbf{y}}{\mathbf{e} \cdot \mathbf{y}}) = P_{\mathbf{e}} (\nabla h)^{-1} (\frac{\mathbf{y}}{\mathbf{e} \cdot \mathbf{y}}) \tag{3.11}$$

where $P_{\mathbf{e}}$ is the perpendicular projection along $\mathbf{e}$.

4 The recourse certainty equivalent

In this section we relate the **recourse certainty equivalent** (RCE), see (1.3), and the **Csiszár ϕ-divergence**, (2.1).

Given a strictly increasing and concave $u : \mathbb{R} \to \mathbb{R}$, the function ϕ defined by

$$\phi = -u_* \tag{4.1}$$

is convex, and by Theorem 3.1(a), $\phi : \mathbb{R}_+ \to \mathbb{R}$. For such u, ϕ, an extremal principle for the RCE S_u in terms of I_ϕ is given in Theorem 4.1, and interpreted in Section 4.2.

Conversely, for a convex $\phi : \mathbb{R}_+ \to \mathbb{R}$, the function u defined by

$$u(t) := -\phi^*(-t) \qquad (4.2)$$

is, under certain assumptions on ϕ, concave. Theorem 4.2 then gives an extremal priciple expressing I_ϕ in terms of S_u. The economic interpretation, in Section 4.4, is that I_ϕ is the optimal value of an insurance plan, for an RCE maximizer.

4.1 An extremal principle for the RCE

Theorem 4.1 *Let $u : \mathbb{R} \to \mathbb{R}$ be strictly increasing, closed and concave, and let $\phi = -u_*$ and $\phi^\diamond = (u^{-1})^*$, see Theorem 3.1(d). Then for any RV $X = [\mathbf{x}, \mathbf{p}]$,*

$$S_u([\mathbf{x},\mathbf{p}]) = \inf_{\mathbf{q}\in\mathbb{P}^n} \{I_\phi(\mathbf{q},\mathbf{p}) + \sum_{i=1}^n q_i\, x_i\} \qquad (4.3)$$

$$= \inf_{\mathbf{q}\in\mathbb{P}^n} \{I_{\phi^\diamond}(\mathbf{p},\mathbf{q}) + \sum_{i=1}^n q_i\, x_i\} \qquad (4.4)$$

Proof. In the problem

$$\inf_{\substack{\sum q_i = 1 \\ q_i \geq 0, \forall i}} \{I_\phi(\mathbf{q},\mathbf{p}) + \sum_{i=1}^n q_i\, x_i\}$$

the objective is convex in $\mathbf{q}$ and the constraints are linear. Therefore the optimal value is equal to the optimal value of the Lagrangian dual problem,

$$\sup_\lambda \inf_{\mathbf{q}\geq\mathbf{0}} \{I_\phi(\mathbf{q},\mathbf{p}) + \sum_{i=1}^n q_i\, x_i + \lambda(1 - \sum q_i)\} =$$

$$= \sup_\lambda \inf_{\mathbf{q}\geq\mathbf{0}} \{I_\phi(\mathbf{q},\mathbf{p}) + \sum_{i=1}^n q_i\,(x_i - \lambda) + \lambda\}$$

$$= \sup_\lambda \{\lambda + \sum_i \inf_{q_i\geq 0}\{q_i\,(x_i-\lambda) + \sum_i p_i \phi(\frac{q_i}{p_i})\}\}$$

$$= \sup_\lambda \{\lambda + \sum_i p_i \inf_{q_i\geq 0}\{\frac{q_i}{p_i}\,(x_i-\lambda) - u_*(\frac{q_i}{p_i})\}\}$$

$$= \sup_\lambda \{\lambda + \sum_i p_i \inf_{t\geq 0}\{t\,(x_i-\lambda) - u_*(t)\}\}$$

$$= \sup_\lambda \{\lambda + \sum_i p_i\, u_{**}(x_i-\lambda)\}$$

(since $\operatorname{dom} u_* \subset \mathbb{R}_+$, Theorem 3.1(a))

$$= \sup_\lambda\{\lambda + Eu(X-\lambda)\}$$

($u = u_{**}$ since u is closed concave)

$$= S_u([\mathbf{x},\mathbf{p}]).$$

The equality (4.4) follows from (4.3) and (2.14). □

An infinite dimensional version of this theorem was proved in Ben-Tal and teboul le (1987).

4.2 A duality interpretation

We recall from Ben-Tal and Ben-Israel (1991) that, for a RV X, the RCE $S_u(X)$ is the optimal value of the problem

$$\sup\ \{z :\ \text{“}z \leq X\text{”}\} \qquad \text{(P)}$$

where optimality is in the sense of **recourse** or **two stage optimization** (Dantzig (1955), Dantzig and Madansky (1961)), using the **value risk function** u to account for the stochastic constraint

$$\text{“}z \leq X\text{”}.$$

We interpret the RHS of (4.3) as the following dual of (P),

$$\inf\ \{z :\ \text{“}z \geq X\text{”}\} \qquad \text{(D)}$$

where the stochastic constraint

$$\text{“}z \geq X\text{”} \qquad (4.5)$$

is “enforced” by using a **stochastic penalty function**, Ben-Tal (1985).

Given the RV

$$X = [\mathbf{x},\mathbf{p}]$$

and z, define the subset of probability vectors

$$R\{z\} := \{\mathbf{q} \in \mathbb{P}^n :\ z \geq \sum_i q_i\, x_i\}, \qquad (4.6)$$

representing the set of RV's

$$\{Z = [\mathbf{x},\mathbf{q}] :\ \mathbf{q} \in R\{z\}\}$$

with the same support as X, which satisfy (4.5) “in the mean”.

We then interpret the problem (D) as

$$\inf\ \{z + P(z)\} \qquad \text{(D1)}$$

where $P(z)$ is the **penalty**

$$P(z) := \text{“dist”}\,(\mathbf{p}, R\{z\}) = \inf_{\mathbf{q}\in R\{z\}} \text{“dist”}\,(\mathbf{q},\mathbf{p}) \qquad (4.7)$$

and “dist” is the “distance” induced by the ϕ-divergence,

$$P(z) = \inf_{\mathbf{q}\in R\{z\}} I_\phi(\mathbf{q},\mathbf{p}) \qquad (4.8)$$

Therefore, the problem (D1) is

$$\inf_{\substack{z \\ \mathbf{q}\in R\{z\}}} \{z + I_\phi(\mathbf{q},\mathbf{p})\} =$$

$$= \inf_{\mathbf{q}\in\mathbb{P}^n} \inf_{z\geq\sum q_i\,x_i} \{z + I_\phi(\mathbf{q},\mathbf{p})\}$$

$$= \inf_{\mathbf{q}\in\mathbb{P}^n} \{I_\phi(\mathbf{q},\mathbf{p}) + \sum q_i\,x_i\}$$

which is the RHS of (4.3).

4.3 Two extremal principles for the ϕ-divergence

Theorem 4.2 *Let $\phi : \mathbb{R}_+ \to \mathbb{R}$ be convex, with dom $\phi \subset \mathbb{R}_{++}$. Define*

$$\begin{aligned} u(t) &:= -\phi^*(-t) && (4.9)\\ &= -\sup_x \{(-t)x - \phi(x)\} \\ &= \inf_x \{tx + \phi(x)\} && (4.10) \end{aligned}$$

Then for any $\mathbf{p},\mathbf{q} \in \mathbb{P}^n$,

$$I_\phi(\mathbf{q},\mathbf{p}) = \sup_{\substack{\mathbf{x}\in\mathbb{R}^n \\ \mathbf{q}\cdot\mathbf{x}=0}} S_u([\mathbf{x},\mathbf{p}]) \qquad (4.11)$$

Furthermore, if dom $\phi = \emptyset$, define the function v by

$$v(x) := (\phi^*)^{-1}(x), \qquad (4.12)$$

see Lemma 3.2. Then for any $\mathbf{p},\mathbf{q} \in \mathbb{P}^n$,

$$I_\phi(\mathbf{p},\mathbf{q}) = \sup_{\substack{\mathbf{x}\in\mathbb{R}^n \\ \mathbf{q}\cdot\mathbf{x}=0}} S_v([\mathbf{x},\mathbf{p}]) \qquad (4.13)$$

Proof. The RHS of (4.11) is

$$\sup_{\mathbf{q}\cdot\mathbf{x}=0} \sup_z \{z + \sum_i p_i\,u(x_i - z)\}$$

Define

$$h(x) := \sum_i p_i\,u(x_i).$$

Then,

$$\begin{aligned} \text{RHS of (4.11)} &= h_+(\mathbf{q}) \ \text{ see (3.6)} \\ &= -h_*(\mathbf{q}), \ \text{ by Theorem 3.2,} \\ &= -\inf_{\mathbf{x}\in\mathbb{R}^n} \{\sum_i q_i\,x_i - \sum_i p_i\,u(x_i)\} \\ &= -\sum_i \inf_{x_i} \{q_i\,x_i - p_i\,u(x_i)\} \\ &= -\sum_i p_i \inf_{x_i} \{\frac{q_i}{p_i}\,x_i - u(x_i)\} \\ &= -\sum_i p_i\,u_*(\frac{q_i}{p_i}) \\ &= -\sum_i p_i\,\phi(\frac{q_i}{p_i}) \ , \\ &\quad \text{by (4.9) and Lemma 3.1} \\ &= I_\phi(\mathbf{q},\mathbf{p}) \end{aligned}$$

Theorem 4.3 *Let $\phi : \mathbb{R}_+ \to \mathbb{R}$ be convex, and let u be defined by (4.2). Then for all $\mathbf{p},\mathbf{q} \in \mathbb{P}^n$,*

$$I_\phi(\mathbf{q},\mathbf{p}) = \sup_{x\in\mathbb{R}^n} \{S_u([\mathbf{x},\mathbf{p}]) - \sum q_i x_i\} \qquad (4.14)$$

Proof. For $g : \mathbb{R}^n \to \mathbb{R}$ define $\hat{g} : \mathbb{R}^n \to \mathbb{R}$ by

$$\hat{g}(\mathbf{y}) := \sup_{x\in\mathbb{R}^n,\ z\in\mathbb{R}} \{z + g(\mathbf{x} - z\mathbf{e}) - \mathbf{y}\cdot\mathbf{x}\}$$

where $\mathbf{e}^T = (1,\ldots,1)$. Then,

$$\begin{aligned} \hat{g}(\mathbf{y}) &= \sup_{x,z,w} \{z + g(\mathbf{w}) - \mathbf{y}\cdot\mathbf{x} : \ \ \mathbf{x} - z\mathbf{e} = \mathbf{w}\} \\ &= \sup_{z,w} \{z + g(\mathbf{w}) - \mathbf{y}\cdot(\mathbf{w} + z\mathbf{e})\} \\ &= \sup_w \{g(\mathbf{w}) - \mathbf{y}\cdot\mathbf{w}\} + \sup_z \{1 - \mathbf{y}\cdot\mathbf{e}\}z \\ &= -g_*(\mathbf{y}) \ \text{ if } \ \mathbf{y}\cdot\mathbf{e} = 1. \end{aligned}$$

Define

$$g(x) := \sum_i p_i\,u(x_i)$$

Then,

$$\begin{aligned} \text{RHS of (4.14)} &= -g_*(\mathbf{q}) \ (\text{since } \sum_i q_i = 1) \\ &= -\inf \{\sum_i q_i\,x_i - p_i\,u(x_i)\} \\ &= -\sum_i p_i \inf \{\frac{q_i}{p_i}\,x_i - u(x_i)\} \\ &= -\sum_i p_i\,u_*(\frac{q_i}{p_i}) \end{aligned}$$

$$= \sum_i p_i \, \phi(\frac{q_i}{p_i}) \;\; \text{(by (4.9) and Lemma 3.1)}$$
$$= I_\phi(\mathbf{q}, \mathbf{p}) \;\square$$

4.4 The ϕ-divergence and optimal insurance plans

We recall here the insurance model of Ben-Tal and Ben-Israel (1991), §9. The elements of this model are:

n **states of nature**
$\mathbf{p}$ $= (p_1, \dots, p_n)$ their **probabilities**
$\bar{q}_i$ = **premium** for 1$ coverage in state i, $\bar{q}_i > 0$
$\bar{B}$ = **insurance budget**
q_i $= \bar{q}_i / \sum_{j=1}^n \bar{q}_j$ = **normalized premium**
B $= \bar{B} / \sum_{j=1}^n \bar{q}_j$ = **normalized budget**
$\bar{x}_i$ = **income** in state i
$\bar{\mathbf{x}}$ $= (\bar{x}_1, \dots, \bar{x}_n)$ the **decision variable**

The insurance problem, for an RCE maximizer, is

$$\begin{array}{ll} \max & S_u([\bar{\mathbf{x}}, \mathbf{p}]) - \sum_i q_i \, \bar{x}_i \\ \text{s.t.} & \sum_i q_i \, \bar{x}_i = B, \text{ the } \textbf{budget constraint} \end{array} \tag{INS}$$

Changing the decision variables from $\bar{x}_i$ to,

$$x_i := \bar{x}_i - B \tag{4.15}$$

we can rewrite (INS) as

$$\begin{array}{ll} \max & S_u([\mathbf{x} + B\mathbf{e}, \mathbf{p}]) - B \\ \text{s.t.} & \sum_i q_i \, x_i = 0, \text{ the } \textbf{normalized budget constraint} \end{array}$$

which, by the shift additivity of S_u, Ben-Tal and Ben-Israel (1991), Theorem 2.1(a), reduces to

$$\begin{array}{ll} \max & S_u([\mathbf{x}, \mathbf{p}]) \\ \text{s.t.} & \sum_i q_i \, x_i = 0 \end{array}$$

the RHS of (4.11).

By Theorem 4.2 the ϕ-divergence functional,

$$I_\phi(\mathbf{q}, \mathbf{p}), \quad \mathbf{p}, \mathbf{q} \in \mathbb{P}^n$$

can be interpreted as the optimal value (for an RCE maximizer) of an insurance plan, subject to the budget constraint $\sum_i q_i \, x_i = 0$. Here $\mathbf{p}$ is the underlying probability vector, and $\mathbf{q}$ is the vector of normalized premiums. Alternatively, by Theorem 4.3, the ϕ-divergence can be interpreted as the (unconstrained) optimal value (for an RCE maximizer) of insurance coverage minus insurance costs (compare with Theorem 5.2).

5 Expected utility

In Section 4 we established relations between the RCE S_u and the ϕ-divergence, where u and ϕ are related by (4.2) or (4.1). Very similar relations hold between the EU and the ϕ-divergence. This is unexpected, in view of the great differences between the RCE and EU, see Ben-Tal and Ben-Israel (1991).

Theorem 5.1 *Let $u : \mathbb{R} \to \mathbb{R}$ be strictly increasing, closed and concave, and let $\phi = -u_*$. Then, for any RV $X = [\mathbf{x}, \mathbf{p}]$ with $\mathbf{p} \in \mathbb{P}^n_{++}$,*

$$Eu([\mathbf{x}, \mathbf{p}]) = \inf_{\mathbf{q} \in \mathbb{R}^n_+} \{ I_\phi(\mathbf{q}, \mathbf{p}) + \sum_{i=1}^n q_i \, x_i \}, \tag{5.1}$$

Proof. The RHS of (5.1) is

$$\begin{aligned} \inf_{\mathbf{q} \in \mathbb{R}^n_+} \{ \sum p_i \, \phi(\frac{q_i}{p_i}) + \sum q_i \, x_i \} &= \\ &= \sum_i p_i \inf_{q_i \geq 0} \{ x_i \, (\frac{q_i}{p_i}) + \phi(\frac{q_i}{p_i}) \} \\ &= \sum_i p_i \inf_{t \geq 0} \{ x_i \, t - u_*(t) \} \\ &= \sum_i p_i \, u_{**}(x_i), \text{ since } \operatorname{dom} u_* \subset \mathbb{R}_+ \\ &= \sum_i p_i \, u(x_i) = Eu(X). \;\square \end{aligned}$$

The representation of EU given in Theorem 5.1 can be interpreted in a similar fashion to the interpretaion of RCE given in Section 4.2. The only difference is that the set $R\{z\}$ is now defined by

$$R\{z\} := \{ \mathbf{q} \in \mathbb{R}^n_+ : \; z \geq \sum_i q_i \, x_i \},$$

instead of (4.6). Theorem 5.1 is illustrated, for the logarithmic utility, as follows.

Example 5.1 (logarithmic utility) *Let*

$$u(x) = \log(1 + x) \tag{5.2}$$

Then, by (1.11) and (3.2),

$$\begin{aligned} \phi(t) &= -u_*(t) \\ &= -\inf_{x > -1} \{ tx - \log(1 + x) \} \\ &= t - \log t - 1 \end{aligned} \tag{5.3}$$

For this ϕ, and any $\mathbf{p}, \mathbf{q} \in I\!R^n_+$,

$$\begin{aligned} I_\phi(\mathbf{p},\mathbf{q}) &= \sum_{i=1}^n q_i\, \phi\,(\frac{p_i}{q_i}) \\ &= \sum_{i=1}^n q_i \left[\frac{p_i}{q_i} - \log\,(\frac{p_i}{q_i}) - 1\right] \\ &= \sum_{i=1}^n p_i - \sum_{i=1}^n q_i - \sum_{i=1}^n q_i \log\,(\frac{p_i}{q_i}) \end{aligned} \tag{5.4}$$

In Theorem 5.1, $\mathbf{p} \in I\!P^n_{++}$, so by (5.4),

$$I_\phi(\mathbf{q},\mathbf{p}) = \sum_{i=1}^n q_i - 1 - \sum_{i=1}^n p_i \log\,(\frac{q_i}{p_i}) \tag{5.5}$$

Therefore,

$$\text{RHS of (5.1)} = \inf_{\mathbf{q}\in I\!R^n_+} \{-1+\sum_{i=1}^n q_i - \sum_{i=1}^n p_i\, \log\,(\frac{q_i}{p_i}) + \sum_{i=1}^n q_i x_i\} \tag{5.6}$$

By differentiating we find the minimizing $\mathbf{q}$

$$q_i = \frac{p_i}{1+x_i}$$

which, when substituted in (5.6), gives

$$\begin{aligned} \text{RHS of (5.1)} &= \sum_{i=1}^n p_i\, \log(1+x_i) \\ &= Eu([\mathbf{x},\mathbf{p}]) \end{aligned}$$

Theorem 5.2 *Let $\phi : I\!R_+ \to I\!R$ be convex, and let u be defined by (4.2). Then for all $\mathbf{p} \in I\!P^n$, $\mathbf{q} \in I\!R^n_+$,*

$$I_\phi(\mathbf{q},\mathbf{p}) = \sup_{\mathbf{x}\in I\!R^n} \{\sum_{i=1}^n p_i\, u(x_i) - \sum_{i=1}^n q_i\, x_i\} \tag{5.7}$$

Proof. The RHS of (5.7) is

$$\begin{aligned} &\sum_i p_i \sup_{x_i} \{u(x_i) - \frac{q_i}{p_i}\, x_i\} = \\ &= \sum_i p_i \sup_{x_i} \{-\phi^*(-x_i) - \frac{q_i}{p_i}\, x_i\} \\ &= \sum_i p_i \sup_t \{\frac{q_i}{p_i}\, t - \phi^*(t)\} \\ &= \sum_i p_i\, \phi^{**}(\frac{q_i}{p_i}) = I_\phi(\mathbf{q},\mathbf{p}) \ \Box \end{aligned}$$

Example 5.2 *We illustrate Theorem 5.2 for the logarithmic utility (5.2) of Example 5.1. For that u,*

$$\text{RHS of (5.7)} = \sup_{\mathbf{x}\in I\!R^n} \{\sum_{i=1}^n p_i\, \log(1+x_i) - \sum_{i=1}^n q_i x_i\} \tag{5.8}$$

The maximizing $\mathbf{x}$ is

$$x_i = \frac{p_i}{q_i} - 1$$

which, substituted in (5.8), gives

$$\begin{aligned} \text{RHS of (5.7)} &= \sum_{i=1}^n p_i\, \log\,(\frac{p_i}{q_i}) - \sum_{i=1}^n p_i + \sum_{i=1}^n q_i \\ &= I_\phi(\mathbf{q},\mathbf{p}), \ \text{by (5.5), since } \mathbf{p} \in I\!P^n \end{aligned}$$

In Subsection 4.4 we interpreted the ϕ-divergence $I_\phi(\mathbf{p},\mathbf{q})$ as the optimal RCE value of an insurance plan given the probabilities $\mathbf{p}$ and the normalized premiums $\mathbf{q}$. We can similarly interpret Theorem 5.2 in terms of optimal insurance plans, where the objective is to maximize expected utility. Let $\mathbf{x}, \mathbf{p}$ and $\mathbf{q}$ be as above. The insurance problem, for an EU maximizer, is

$$\sup_{\mathbf{x}\in I\!R^n} \{ \overbrace{\sum p_i\, u(x_i)}^{\text{expected utility}} - \overbrace{\sum q_i\, x_i}^{\text{cost}} \} \tag{5.9}$$

which is the RHS of (5.7).

Remark. The ϕ-divergence DIV was given above in terms of a certainty equivalent (RCE or EU) and the expectation E, see Theorems 4.3 and 5.2. Analogously, these CE's were given in terms of DIV and E, Theorems 4.1 and 5.1. These extremal principles are symmetric in E, CE and DIV in the sense that each can be expressed in terms of the other two. For example, we can express E in terms of RCE and DIV as follows,

$$E\,([\mathbf{x},\mathbf{q}]) = \sup_{\mathbf{p}\in \mathbb{R}^n_+} \{\sum_i p_i\, u(x_i) - I_\phi(\mathbf{q},\mathbf{p})\}$$

6 Yaari's certainty equivalent

Consider a random variable $X = [\mathbf{x},\mathbf{p}]$. The cummulative distribution is:

$$F_i = \sum_{j=0}^i p_j, \quad i = 0,\ldots,n. \tag{6.1}$$

where $p_0 := 0$ and the corresponding decummulative distribution is

$$\bar{F}_{i+1} := \sum_{j=i+1}^{n} p_j = 1 - F_i, \quad i = 0, \dots, n \qquad (6.2)$$

where $\bar{F}_{n+1} := 0$. Let $f : [0,1] \to \mathbb{R}$ be a convex function. For the given vector $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ define

$$\Delta x_i := x_i - x_{i-1}, \; i = 1, \dots, n$$

where $x_0 := 0$, and denote $\mathbf{\Delta x} = (\Delta x_1, \dots, \Delta x_n)^T$. The certainty eqivalent of Yaari, Yaari (1987), is then defined as follows

$$Y_f([\mathbf{x}, \mathbf{p}]) := \sum_{i=1}^{n} f(\bar{F}_i)\, \Delta x_i \qquad (6.3)$$

Without loss of generality we assume that

$$x_1 < x_2 < \cdots < x_n, \quad y_1 < y_2 < \cdots < y_n$$

Consider the ϕ-divergence between $\mathbf{\Delta x}$ and $\mathbf{\Delta y}$:

$$I_\phi(\mathbf{\Delta y}, \mathbf{\Delta x}) := \sum_{i=1}^{n} \Delta x_i\, \phi(\frac{\Delta y_i}{\Delta x_i}) \qquad (6.4)$$

¿From Corollary 2.1 it follows (for $\phi(1) = 0$) that

$$I_\phi(\mathbf{\Delta y}, \mathbf{\Delta x}) \geq 0$$

with equality if and only if $\Delta x_i = \Delta y_i$, $i = 1, \dots n$, i.e. if and only if $x_i = y_i$ for all $i = 1, \dots, n$.

Theorem 6.1 *Let $f : [0,1] \to [0,1]$ be strictly increasing, closed and convex, and let $f(t) = \phi^*(t)$. Then for any RV $X = [\mathbf{x}, \mathbf{p}]$*

$$Y_f([\mathbf{x}, \mathbf{p}]) = \sup_{\mathbf{\Delta y} \in \mathbb{R}^n_+} \{\sum_{i=1}^{n} p_i\, y_i - I_\phi(\mathbf{\Delta y}, \mathbf{\Delta x})\} \qquad (6.5)$$

Proof. Consider a RV $Y = [\mathbf{y}, \mathbf{p}]$. From (6.1) and (6.2) it follows that $p_i = \bar{F}_i - \bar{F}_{i+1}$ and thus

$$\sum_i p_i\, y_i = \sum_i \bar{F}_i\, \Delta y_i$$

Now,

$$\begin{aligned}
\text{RHS of (6.5)} &= \sup_{\mathbf{\Delta y} \geq \mathbf{0}} \{\sum_i \bar{F}_i\, \Delta y_i - \sum_i \Delta x_i\, \phi(\frac{\Delta y_i}{\Delta x_i})\} \\
&= \sum_i \Delta x_i \sup_{\Delta y_i \geq 0} \{\bar{F}_i \frac{\Delta y_i}{\Delta x_i} - \phi(\frac{\Delta y_i}{\Delta x_i})\} \\
&= \sum_i \Delta x_i \sup_{t \in \text{dom } \phi} \{\bar{F}_i\, t - \phi(t)\} \\
&\qquad (\text{since } \mathbb{R}_+ \supseteq \text{dom } \phi = \text{dom } f^*) \\
&= \sum_i \Delta x_i\, \phi^*(\bar{F}_i) \\
&\qquad (\text{since dom } \phi \subseteq \mathbb{R}_+) \\
&= \sum_i \Delta x_i\, f(\bar{F}_i) \\
&= Y_f([\mathbf{x}, \mathbf{p}]) \quad \Box
\end{aligned}$$

Theorem 6.2 *Let $\phi : \mathbb{R}_+ \to \mathbb{R}$ be convex and strictly increasing, and let $f(t) = \phi^*(t)$. Then for all $\mathbf{\Delta x}, \mathbf{\Delta y} \in \mathbb{R}^n_+$,*

$$I_\phi(\mathbf{\Delta y}, \mathbf{\Delta x}) = \sup_{\mathbf{p} \in \mathbb{R}^n_+} \{\sum_i y_i\, p_i - Y_f([\mathbf{x}, \mathbf{p}])\} \qquad (6.6)$$

Proof

$$\begin{aligned}
\text{RHS of (6.6)} &= \sup_{0 \leq \bar{F}_i \leq 1} \{\sum_i \Delta y_i\, \bar{F}_i - \sum_i \Delta x_i\, f(\bar{F}_i)\} \\
&= \sum_i \Delta x_i \sup_{0 \leq \bar{F}_i \leq 1} \{\frac{\Delta y_i}{\Delta x_i}\, \bar{F}_i - f(\bar{F}_i)\} \\
&= \sum_i \Delta x_i\, f^*(\frac{\Delta y_i}{\Delta x_i}) \\
&= \sum_i \Delta x_i\, \phi(\frac{\Delta y_i}{\Delta x_i}) \\
&= I_\phi(\mathbf{\Delta y}, \mathbf{\Delta x}) \quad \Box
\end{aligned}$$

References

Aczél, J. (1984) : "Measuring information beyond communication theory - why some generalized information measures may be useful, others not," *Aequationes Mathematicae*, **27**, 1-19.

Ben-Tal, A. (1985) : "The entropic penalty approach to stochastic programming," *Math. Oper. Res.*, **10**, 263-279.

Ben-Tal, A. and A. Ben-Israel (1991) : "A recourse certainty equivalent for decisions under uncertainty," *Annals of Oper. Res.*, **30**, 3–44.

Ben-Tal, A. and M. Teboulle (1986) : "Expected utility, penalty functions, and duality in stochastic nonlinear programming," *Management Sci.*, **32**, 1445-1466.

Ben-Tal, A. and M. Teboulle (1987) : "Penalty functions and duality in stochastic programming via ϕ-divergence functionals," *Math. Oper. Res.*, **12**, 224-240.

Beran, R. (1977) : "Minimum Hellinger distance estimates for parametric models," *Ann. Statist.*, **5**, 445-463.

Burbea, J. and C.R. Rao (1982) : "On the convexity of some divergence measures based on entropy functions," *IEEE Trans. on Information Th.*, **IT28**, 489-495.

Csiszár, I. (1978) : "Information measures: A critical survey," *Trans. 7th Prague Conf. on Info. Th., Statist., Decis. Funct., Random Processes and 8th European Meeting of Statist.*, volume B, pp. 73-86, Academia, Prague.

Csiszár, I. (1967) : " Information-Type measures of difference of probabilty distributions and indirect observations," *Studia Sci. Math. Hungar.*, **2**, 299-318.

Csiszár, I. and J. Körner (1981) : *Information Theory: Coding Theorems for Discrete Memoryless systems*, Academic Press, New-York.

Dantzig, G.B. (1955) : "Linear programming under uncertainty," *Manag. Sci.*, **1**, 197-206.

Dantzig, G.B. and A. Madansky (1961) : "On the solution of two-stage linear programs under uncertainty," *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 165-176, University of California, Berkeley.

Frieden, B.R. (1975) : "Image enhancement and restoration," pp. 177-248 in *Picture Processing and Digital Filtering* (T.S. Huang, Editor), Springer-Verlag, Berlin.

Gallager, R.G. (1968) : *Information Theory and Reliable Communication*, J. Wiley, New York.

Justice, J.H. (Editor), (1986) : *Maximum Entropy and Bayesian Methods in Applied Statistics*, Cambridge University Press, Cambridge.

Kapur, J.N. (1984) : "On the roles of maximum-entropy and minimum discrimination information principles in Statistics," *Technical address at the 38th Annual Conference of the Indian Society of Agricultural Statistics*, pp. 1-44.

Kullback, S. (1959) : *Information Theory and Statistics*, J. Wiley, New York.

Kullback, S. and R.A. Leibler (1951) : "On information and sufficiency," *Annals Math. Statist.*, **22**, 79-86.

Leahy, R.M. and C.E. Goutis (1986) : "An optimal technique for constraint-based image restoration and reconstruction," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **ASSP-34**, 1629-1642.

Machina, M.J. (1987) : "Choice under uncertainty: Problems solved and unsolved," *Econ. Perspectives*, **1**, 121-154.

Renyi, A. (1961) : "On measures of entropy and information," pp. 547-561 in *Proc. Fourth Berkley Sympos. Math. Statist. Prob.*, Vol. 1, Univ. of California Press, Berkley.

Roberts A. W. and D. E. Varberg (1973) : *Convex Functions*, Academic Press New York and London.

Rockafellar, R.T. (1970) : *Convex Analysis*, Princeton University Press, Princeton.

Röell, A. (1987) : "Risk aversion in Quiggin and Yaari's rank-order model of choice under uncertainty", *The Economic J.*, **97**, 143-159.

Shannon, C.E. (1948) : "A mathematical theory of communication," *Bell Syst. Tech. J.*, **27**, 379-423 and 623-656.

Teboulle, M. (1988) : "On ϕ-divergence and its applications," to appear in *Proceedings of the Conference in Honor of A. Charnes' 70th Birthday*, Austin, Texas.

Von Neumann, J. and O. Morgenstern (1947) : *Theory of Games and Economic Behavior*, Princeton University Press, Princeton.

Yaari, M.E. (1987) : "The dual theory of choice under risk," *Econometrica*, **55**, 95-115.