Optimal measures and transition kernels

Roman V. Belavkin

Middlesex University

February 11, 2011

Introduction

- Optimisation of dynamic information systems (e.g. learning or evolutionary systems).

Introduction

- Optimisation of dynamic information systems (e.g. learning or evolutionary systems).
- Three kinds of problems

Introduction

- Optimisation of dynamic information systems (e.g. learning or evolutionary systems).
- Three kinds of problems

  Deterministic : maximise $f(x)$.

Introduction

- Optimisation of dynamic information systems (e.g. learning or evolutionary systems).
- Three kinds of problems

  Deterministic : maximise $f(x)$.

  Stochastic : maximise $\mathbb{E}_p\{f(x)\} = \sum f(x)\, p(x)$.

Introduction

- Optimisation of dynamic information systems (e.g. learning or evolutionary systems).
- Three kinds of problems

  Deterministic : maximise $f(x)$.

  Stochastic : maximise $\mathbb{E}_p\{f(x)\} = \sum f(x)\,p(x)$.

  Adaptive : estimate $p(x)$ and maximise $\mathbb{E}_p\{f(x)\}$.

Introduction

- Optimisation of dynamic information systems (e.g. learning or evolutionary systems).
- Three kinds of problems

  Deterministic : maximise $f(x)$.

  Stochastic : maximise $\mathbb{E}_p\{f(x)\} = \sum f(x)\,p(x)$.

  Adaptive : estimate $p(x)$ and maximise $\mathbb{E}_p\{f(x)\}$.
- Learning, evolving and adaptive systems:

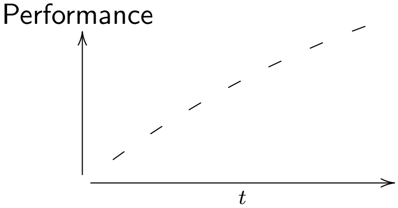
Introduction

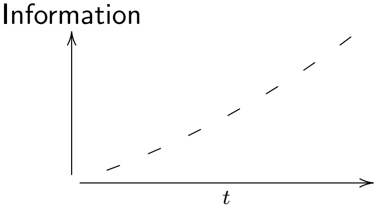
- Optimisation of dynamic information systems (e.g. learning or evolutionary systems).
- Three kinds of problems

  Deterministic : maximise $f(x)$.

  Stochastic : maximise $\mathbb{E}_p\{f(x)\} = \sum f(x)\,p(x)$.

  Adaptive : estimate $p(x)$ and maximise $\mathbb{E}_p\{f(x)\}$.
- Learning, evolving and adaptive systems:





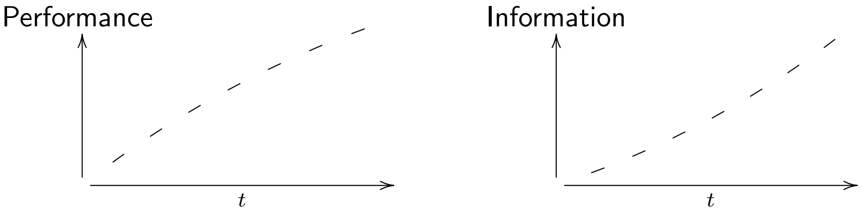
Introduction

- Optimisation of dynamic information systems (e.g. learning or evolutionary systems).
- Three kinds of problems

  Deterministic : maximise $f(x)$.

  Stochastic : maximise $\mathbb{E}_p\{f(x)\} = \sum f(x)\,p(x)$.

  Adaptive : estimate $p(x)$ and maximise $\mathbb{E}_p\{f(x)\}$.
- Learning, evolving and adaptive systems:



- In practice, stochastic and non-deterministic algorithms are generally more successful (e.g. simulated annealing, genetic algorithms).

Representation in Paired Spaces

- Ω — a measurable set, a locally compact space.

Representation in Paired Spaces

- Ω — a measurable set, a locally compact space.
- X, Y — linear spaces in duality via $\langle \cdot, \cdot \rangle : X \times Y \to \mathbb{R}$:

$$\langle x, y \rangle := \sum_{\Omega} x(\omega)\, y(\omega)\,, \quad \langle x, y \rangle := \int_{\Omega} x(\omega)\, dy(\omega)\,, \quad \langle x, y \rangle := \mathrm{tr}\,\{xy\}$$

Representation in Paired Spaces

- Ω — a measurable set, a locally compact space.
- X, Y — linear spaces in duality via $\langle\cdot,\cdot\rangle : X \times Y \to \mathbb{R}$:

$$\langle x, y\rangle := \sum_{\Omega} x(\omega)\, y(\omega)\,, \quad \langle x, y\rangle := \int_{\Omega} x(\omega)\, dy(\omega)\,, \quad \langle x, y\rangle := \mathrm{tr}\,\{xy\}$$

Observables and Measures

- $X := C_c(\Omega, \mathbb{R})$ space of continuous function with compact support.

Representation in Paired Spaces

- Ω — a measurable set, a locally compact space.
- X, Y — linear spaces in duality via $\langle\cdot,\cdot\rangle : X \times Y \to \mathbb{R}$:

$$\langle x, y\rangle := \sum_{\Omega} x(\omega)\, y(\omega)\,, \quad \langle x, y\rangle := \int_{\Omega} x(\omega)\, dy(\omega)\,, \quad \langle x, y\rangle := \mathrm{tr}\,\{xy\}$$

Observables and Measures

- $X := C_c(\Omega, \mathbb{R})$ space of continuous function with compact support.
- $Y := \mathcal{M}(\Omega)$ space of Radon measures on Ω.

Representation in Paired Spaces

- Ω — a measurable set, a locally compact space.
- X, Y — linear spaces in duality via $\langle \cdot, \cdot \rangle : X \times Y \to \mathbb{R}$:

$$\langle x, y \rangle := \sum_{\Omega} x(\omega)\, y(\omega)\,, \quad \langle x, y \rangle := \int_{\Omega} x(\omega)\, dy(\omega)\,, \quad \langle x, y \rangle := \mathrm{tr}\,\{xy\}$$

Observables and Measures

- $X := C_c(\Omega, \mathbb{R})$ space of continuous function with compact support.
- $Y := \mathcal{M}(\Omega)$ space of Radon measures on Ω.

Remark

- *X is an ordered, commutative, linear $*$-algebra with* $\|x\|_\infty = \sup x(\omega)$

Representation in Paired Spaces

- Ω — a measurable set, a locally compact space.
- X, Y — linear spaces in duality via $\langle\cdot,\cdot\rangle : X \times Y \to \mathbb{R}$:

$$\langle x, y\rangle := \sum_{\Omega} x(\omega)\, y(\omega)\,, \quad \langle x, y\rangle := \int_{\Omega} x(\omega)\, dy(\omega)\,, \quad \langle x, y\rangle := \mathrm{tr}\,\{xy\}$$

Observables and Measures

- $X := C_c(\Omega, \mathbb{R})$ space of continuous function with compact support.
- $Y := \mathcal{M}(\Omega)$ space of Radon measures on Ω.

Remark

- *X is an ordered, commutative, linear $*$-algebra with* $\|x\|_\infty = \sup x(\omega)$
- *Y is a module over $X \subset Y$ with* $\|y\|_1 = |\langle 1, y\rangle|$

Exponential Family

- Let $\{y_\beta\}_x$ be a family of

$$y_\beta := e^{\beta x}\, y_0 \,, \quad y_0 > 0$$

Exponential Family

- Let $\{y_\beta\}_x$ be a family of

$$y_\beta := e^{\beta x}\, y_0\,, \quad y_0 > 0$$

- Normalisation $p = y/\|y\|_1$ for $\|y\|_1 = |\langle 1, y\rangle|$ gives

$$p_\beta = e^{\beta x - \Psi(\beta)}\, y_0$$

where $\Psi(\beta) := \ln\langle 1, e^{\beta x}\, y_0\rangle$

Exponential Family

- Let $\{y_\beta\}_x$ be a family of

$$y_\beta := e^{\beta x}\, y_0\,, \quad y_0 > 0$$

- Normalisation $p = y/\|y\|_1$ for $\|y\|_1 = |\langle 1, y\rangle|$ gives

$$p_\beta = e^{\beta x - \Psi(\beta)}\, y_0$$

where $\Psi(\beta) := \ln\langle 1, e^{\beta x}\, y_0\rangle$

- Observe that y_β for $\beta > 0$ are mutually absolute continuous

$$y_1(E) = 0 \quad \Longleftrightarrow \quad y_2(E) = 0 \qquad \forall\, E \in \mathcal{R}(\Omega)$$

Exponential Family

- Let $\{y_\beta\}_x$ be a family of

$$y_\beta := e^{\beta x}\, y_0\,, \quad y_0 > 0$$

- Normalisation $p = y/\|y\|_1$ for $\|y\|_1 = |\langle 1, y\rangle|$ gives

$$p_\beta = e^{\beta x - \Psi(\beta)}\, y_0$$

  where $\Psi(\beta) := \ln\langle 1, e^{\beta x}\, y_0\rangle$

- Observe that y_β for $\beta > 0$ are mutually absolute continuous

$$y_1(E) = 0 \quad \Longleftrightarrow \quad y_2(E) = 0 \qquad \forall\, E \in \mathcal{R}(\Omega)$$

- A number of properties: Rao-Cramer, maximum entropy.

Composite Systems and Transition Kernels

- Let $\Omega = A \times B$, $(A, \mathcal{A})$, $(B, \mathcal{B})$.

Composite Systems and Transition Kernels

- Let $\Omega = A \times B$, $(A, \mathcal{A})$, $(B, \mathcal{B})$.
- Conditional probability $p(A_i \mid b)$:

$$p(A_i \mid B_j) = \frac{p(A_i \cap B_j)}{P(B_j)}\,, \qquad p(B_j) > 0$$

Composite Systems and Transition Kernels

- Let $\Omega = A \times B$, $(A, \mathcal{A})$, $(B, \mathcal{B})$.
- Conditional probability $p(A_i \mid b)$:

$$p(A_i \mid B_j) = \frac{p(A_i \cap B_j)}{P(B_j)}\,, \qquad p(B_j) > 0$$

- $p(A_i \mid B_k)$ is a *transition kernel*, if it is measurable w.r.t. $\mathcal{B}$, $\forall\, A_i \in \mathcal{A}$.

Composite Systems and Transition Kernels

- Let $\Omega = A \times B$, $(A, \mathcal{A})$, $(B, \mathcal{B})$.
- Conditional probability $p(A_i \mid b)$:

$$p(A_i \mid B_j) = \frac{p(A_i \cap B_j)}{P(B_j)}\,, \qquad p(B_j) > 0$$

- $p(A_i \mid B_k)$ is a *transition kernel*, if it is measurable w.r.t. $\mathcal{B}$, $\forall\, A_i \in \mathcal{A}$.
- A function $a = f(b)$ corresponds to deterministic kernel

$$p(A_i \mid b) = \delta_{f(b)}(A_i) = \begin{cases} 1 & \text{if } f(b) \in A_i \\ 0 & \text{otherwise} \end{cases}$$

Composite Systems and Transition Kernels

- Let $\Omega = A \times B$, $(A, \mathcal{A})$, $(B, \mathcal{B})$.
- Conditional probability $p(A_i \mid b)$:

$$p(A_i \mid B_j) = \frac{p(A_i \cap B_j)}{P(B_j)}\,, \qquad p(B_j) > 0$$

- $p(A_i \mid B_k)$ is a *transition kernel*, if it is measurable w.r.t. $\mathcal{B}$, $\forall\, A_i \in \mathcal{A}$.
- A function $a = f(b)$ corresponds to deterministic kernel

$$p(A_i \mid b) = \delta_{f(b)}(A_i) = \begin{cases} 1 & \text{if } f(b) \in A_i \\ 0 & \text{otherwise} \end{cases}$$

- Observe that

$$dp_\beta(a \mid b) = e^{\beta x(a,b) - \Psi(\beta,b)}\, dp_0(a) \to \delta_{f(b)}(a) \quad \text{as } \beta \to \infty$$

Optimisation

- Let $(\Omega, \lesssim)$ be a preference relation (total pre-order).

Optimisation

- Let $(\Omega, \lesssim)$ be a preference relation (total pre-order).
- Let $x : \Omega \to \mathbb{R}$ be a utility representation:

$$a \lesssim b \quad \iff \quad x(a) \leq x(b)$$

Optimisation

- Let $(\Omega, \lesssim)$ be a preference relation (total pre-order).
- Let $x : \Omega \to \mathbb{R}$ be a utility representation:

$$a \lesssim b \quad \iff \quad x(a) \leq x(b)$$

- Let $(\Omega, \mathcal{R}, p)$ be a probability space, and $\mathbb{E}_p\{x\} := \langle x, p \rangle$.

Optimisation

- Let $(\Omega, \lesssim)$ be a preference relation (total pre-order).
- Let $x : \Omega \to \mathbb{R}$ be a utility representation:

$$a \lesssim b \quad \iff \quad x(a) \leq x(b)$$

- Let $(\Omega, \mathcal{R}, p)$ be a probability space, and $\mathbb{E}_p\{x\} := \langle x, p \rangle$.
- Let $\mathcal{P}(\Omega) := \{y : y \geq 0\,,\ \|y\|_1 = 1\}$ be a statistical manifold

Optimisation

- Let $(\Omega, \lesssim)$ be a preference relation (total pre-order).
- Let $x : \Omega \to \mathbb{R}$ be a utility representation:

$$a \lesssim b \quad \Longleftrightarrow \quad x(a) \leq x(b)$$

- Let $(\Omega, \mathcal{R}, p)$ be a probability space, and $\mathbb{E}_p\{x\} := \langle x, p \rangle$.
- Let $\mathcal{P}(\Omega) := \{y : y \geq 0\,,\ \|y\|_1 = 1\}$ be a statistical manifold
- Then, for all p, $q \in \mathcal{P}(\Omega)$

$$q \lesssim p \quad \Longleftrightarrow \quad \mathbb{E}_q\{x\} \leq \mathbb{E}_p\{x\}$$

Optimisation

- Let $(\Omega, \lesssim)$ be a preference relation (total pre-order).
- Let $x : \Omega \to \mathbb{R}$ be a utility representation:

$$a \lesssim b \quad \iff \quad x(a) \leq x(b)$$

- Let $(\Omega, \mathcal{R}, p)$ be a probability space, and $\mathbb{E}_p\{x\} := \langle x, p \rangle$.
- Let $\mathcal{P}(\Omega) := \{y : y \geq 0\,,\ \|y\|_1 = 1\}$ be a statistical manifold
- Then, for all p, $q \in \mathcal{P}(\Omega)$

$$q \lesssim p \quad \iff \quad \mathbb{E}_q\{x\} \leq \mathbb{E}_p\{x\}$$

- We can extend $(\mathcal{P}, \lesssim)$ to $(Y, \lesssim)$ by $\{y : \langle x, y \rangle \geq 0\}$:

$$z \lesssim y \quad \iff \quad \langle x, z \rangle \leq \langle x, y \rangle$$

Optimisation

- Let $(\Omega, \lesssim)$ be a preference relation (total pre-order).
- Let $x : \Omega \to \mathbb{R}$ be a utility representation:

$$a \lesssim b \quad \Longleftrightarrow \quad x(a) \leq x(b)$$

- Let $(\Omega, \mathcal{R}, p)$ be a probability space, and $\mathbb{E}_p\{x\} := \langle x, p \rangle$.
- Let $\mathcal{P}(\Omega) := \{y : y \geq 0\,,\ \|y\|_1 = 1\}$ be a statistical manifold
- Then, for all p, $q \in \mathcal{P}(\Omega)$

$$q \lesssim p \quad \Longleftrightarrow \quad \mathbb{E}_q\{x\} \leq \mathbb{E}_p\{x\}$$

- We can extend $(\mathcal{P}, \lesssim)$ to $(Y, \lesssim)$ by $\{y : \langle x, y \rangle \geq 0\}$:

$$z \lesssim y \quad \Longleftrightarrow \quad \langle x, z \rangle \leq \langle x, y \rangle$$

Remark

$\mathbb{E}_p\{x\}$ ($\langle x, \cdot \rangle$) is a unique functional such that $\lesssim$ on $\mathcal{P}(\Omega)$ (Y) is compatible with the linear structure of Y and Archimedean.

Information

- Information distance $I : \mathcal{P} \times \mathcal{P} \to \mathbb{R}_+ \cup \{\infty\}$ (e.g. distance in variation, Fisher information metric).

Information

- Information distance $I : \mathcal{P} \times \mathcal{P} \to \mathbb{R}_+ \cup \{\infty\}$ (e.g. distance in variation, Fisher information metric).
- Kullback-Leibler divergence (Kullback, 1959):

$$I_{KL} := \mathbb{E}_p\{\ln p - \ln q\} = \left\langle \ln \frac{p}{q}, p \right\rangle$$

Information

- Information distance $I : \mathcal{P} \times \mathcal{P} \to \mathbb{R}_+ \cup \{\infty\}$ (e.g. distance in variation, Fisher information metric).
- Kullback-Leibler divergence (Kullback, 1959):

$$I_{KL} := \mathbb{E}_p\{\ln p - \ln q\} = \left\langle \ln \frac{p}{q}, p \right\rangle$$

Definition (Information resource)

A closed functional $F : Y \to \mathbb{R} \cup \{\infty\}$ such that $F(y)|_{\mathcal{P}} = I(y, y_0)$.

Information

- Information distance $I : \mathcal{P} \times \mathcal{P} \to \mathbb{R}_+ \cup \{\infty\}$ (e.g. distance in variation, Fisher information metric).
- Kullback-Leibler divergence (Kullback, 1959):

$$I_{KL} := \mathbb{E}_p\{\ln p - \ln q\} = \left\langle \ln \frac{p}{q}, p \right\rangle$$

Definition (Information resource)

A closed functional $F : Y \to \mathbb{R} \cup \{\infty\}$ such that $F(y)|_{\mathcal{P}} = I(y, y_0)$.

Example (Relative information)

$$F_{KL}(y) := \begin{cases} \left\langle \ln \frac{y}{y_0}, y \right\rangle - \langle 1, y - y_0 \rangle & \text{if } y > 0 \text{ and } y_0 > 0 \\ \langle 1, y_0 \rangle & \text{if } y = 0 \text{ and } y_0 > 0 \\ \infty & \text{otherwise} \end{cases}$$

Utility (Value) of Information

- Consider the problem (Shannon, 1948; Jaynes, 1957; Stratonovich, 1965):

$$\sup\{\mathbb{E}_p\{x\} : I_{KL}(p.q) \leq \lambda\}$$

Utility (Value) of Information

- Consider the problem (Shannon, 1948; Jaynes, 1957; Stratonovich, 1965):

$$\sup\{\mathbb{E}_p\{x\} : I_{KL}(p.q) \leq \lambda\}$$

- Related problem:

$$\inf\{I_{KL}(p,q) : \mathbb{E}_p\{x\} \geq \upsilon\}$$

Utility (Value) of Information

- Consider the problem (Shannon, 1948; Jaynes, 1957; Stratonovich, 1965):

$$\sup\{\mathbb{E}_p\{x\} : I_{KL}(p.q) \leq \lambda\}$$

- Related problem:

$$\inf\{I_{KL}(p,q) : \mathbb{E}_p\{x\} \geq \upsilon\}$$

- Generalisation: given $x \in X$ let $\overline{x} : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$:

$$\overline{x}(\lambda) := \sup\{\langle x, y\rangle : F(y) \leq \lambda\}$$

with $\overline{x}(\lambda) := -\infty$ if $\lambda < \inf F$, and $\exists\, x : \overline{x}(\infty) = \infty$.

Utility (Value) of Information

- Consider the problem (Shannon, 1948; Jaynes, 1957; Stratonovich, 1965):

$$\sup\{\mathbb{E}_p\{x\} : I_{KL}(p.q) \leq \lambda\}$$

- Related problem:

$$\inf\{I_{KL}(p,q) : \mathbb{E}_p\{x\} \geq \upsilon\}$$

- Generalisation: given $x \in X$ let $\overline{x} : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$:

$$\overline{x}(\lambda) := \sup\{\langle x, y\rangle : F(y) \leq \lambda\}$$

  with $\overline{x}(\lambda) := -\infty$ if $\lambda < \inf F$, and $\exists\, x : \overline{x}(\infty) = \infty$.
- Function $\overline{x}(\lambda)$ has inverse

$$\overline{x}^{-1}(\upsilon) := \inf\{F(y) : \langle x, y\rangle \geq \upsilon\}$$

Unbounded Information

- $\mathcal{P}(\Omega)$ is a compact, convex set in Y.

Unbounded Information

- $\mathcal{P}(\Omega)$ is a compact, convex set in Y.
- Let $\operatorname{ext}\mathcal{P} \subseteq \partial\mathcal{P}$ be the set of extreme points $\delta \in \mathcal{P}$

Unbounded Information

- $\mathcal{P}(\Omega)$ is a compact, convex set in Y.
- Let $\operatorname{ext}\mathcal{P} \subseteq \partial\mathcal{P}$ be the set of extreme points $\delta \in \mathcal{P}$

Proposition

$\forall\, x \in X$, $\exists\, \delta_x \in \operatorname{ext}\mathcal{P}$ *such that*

$$\langle x, \delta_x \rangle = \sup\{\langle x, p \rangle : p \in \mathcal{P}\}$$

Unbounded Information

- $\mathcal{P}(\Omega)$ is a compact, convex set in Y.
- Let $\operatorname{ext}\mathcal{P} \subseteq \partial\mathcal{P}$ be the set of extreme points $\delta \in \mathcal{P}$

Proposition

$\forall\, x \in X$, $\exists\, \delta_x \in \operatorname{ext}\mathcal{P}$ *such that*

$$\langle x, \delta_x \rangle = \sup\{\langle x, p\rangle : p \in \mathcal{P}\}$$

Proof.

$$\exists\, \Delta \subseteq \operatorname{ext}\mathcal{P} \qquad : \qquad \langle x, \delta_x \rangle = \sup\{\langle x, p\rangle : p \in \mathcal{P}\}\,,\ \forall\, \delta_x \in \Delta$$

Unbounded Information

- $\mathcal{P}(\Omega)$ is a compact, convex set in Y.
- Let $\operatorname{ext}\mathcal{P} \subseteq \partial\mathcal{P}$ be the set of extreme points $\delta \in \mathcal{P}$

Proposition

$\forall\, x \in X$, $\exists\, \delta_x \in \operatorname{ext}\mathcal{P}$ *such that*

$$\langle x, \delta_x\rangle = \sup\{\langle x, p\rangle : p \in \mathcal{P}\}$$

Proof.

$$\begin{aligned} \exists\, \Delta \subseteq \operatorname{ext}\mathcal{P} \quad &: \quad \langle x, \delta_x\rangle = \sup\{\langle x, p\rangle : p \in \mathcal{P}\}\,,\ \forall\, \delta_x \in \Delta \\ &\iff \quad \langle x, \bar{p}\rangle = \sup\{\langle x, p\rangle : p \in \mathcal{P}\}\,,\ \forall\, \bar{p} \in \operatorname{cl}\operatorname{co}\Delta \end{aligned}$$

Unbounded Information

- $\mathcal{P}(\Omega)$ is a compact, convex set in Y.
- Let $\operatorname{ext}\mathcal{P} \subseteq \partial\mathcal{P}$ be the set of extreme points $\delta \in \mathcal{P}$

Proposition

$\forall\, x \in X$, $\exists\, \delta_x \in \operatorname{ext}\mathcal{P}$ *such that*

$$\langle x, \delta_x\rangle = \sup\{\langle x, p\rangle : p \in \mathcal{P}\}$$

Proof.

$$\begin{aligned}
\exists\, \Delta \subseteq \operatorname{ext}\mathcal{P} \quad &: \quad \langle x, \delta_x\rangle = \sup\{\langle x, p\rangle : p \in \mathcal{P}\}\,,\ \forall\, \delta_x \in \Delta \\
&\iff \quad \langle x, \bar{p}\rangle = \sup\{\langle x, p\rangle : p \in \mathcal{P}\}\,,\ \forall\, \bar{p} \in \operatorname{cl}\operatorname{co}\Delta \\
\Delta = \emptyset \quad &\iff \quad \langle x, \bar{p}\rangle > \langle x, p\rangle\,,\ \forall\, p \in \operatorname{cl}\operatorname{co}\operatorname{ext}\mathcal{P}
\end{aligned}$$

Unbounded Information

- $\mathcal{P}(\Omega)$ is a compact, convex set in Y.
- Let $\operatorname{ext}\mathcal{P} \subseteq \partial\mathcal{P}$ be the set of extreme points $\delta \in \mathcal{P}$

Proposition

$\forall\, x \in X$, $\exists\, \delta_x \in \operatorname{ext}\mathcal{P}$ *such that*

$$\langle x, \delta_x\rangle = \sup\{\langle x, p\rangle : p \in \mathcal{P}\}$$

Proof.

$$\begin{aligned}
\exists\, \Delta \subseteq \operatorname{ext}\mathcal{P} \quad &: \quad \langle x, \delta_x\rangle = \sup\{\langle x, p\rangle : p \in \mathcal{P}\}\,,\ \forall\, \delta_x \in \Delta \\
&\iff \quad \langle x, \bar{p}\rangle = \sup\{\langle x, p\rangle : p \in \mathcal{P}\}\,,\ \forall\, \bar{p} \in \operatorname{cl}\operatorname{co}\Delta \\
\Delta = \emptyset \quad &\iff \quad \langle x, \bar{p}\rangle > \langle x, p\rangle\,,\ \forall\, p \in \operatorname{cl}\operatorname{co}\operatorname{ext}\mathcal{P} \\
&\implies \quad \bar{p} \notin \mathcal{P} \qquad (\operatorname{cl}\operatorname{co}\operatorname{ext}\mathcal{P} = \mathcal{P})
\end{aligned}$$

Necessary and Sufficient Optimality Conditions

Proposition

Element y_β *solves* $\overline{x}(\lambda) := \sup\{\langle x, y\rangle : F(y) \leq \lambda\}$ *iff:*

$$y_\beta \in \partial F^*(\beta x)\,, \quad F(y_\beta) = \lambda\,, \quad \beta^{-1} \in \partial \overline{x}(\lambda)\,, \quad \beta^{-1} > 0$$

Necessary and Sufficient Optimality Conditions

Proposition

Element y_β *solves* $\overline{x}(\lambda) := \sup\{\langle x, y\rangle : F(y) \leq \lambda\}$ *iff:*

$$y_\beta \in \partial F^*(\beta x)\,, \quad F(y_\beta) = \lambda\,, \quad \beta^{-1} \in \partial\overline{x}(\lambda)\,, \quad \beta^{-1} > 0$$

Proof.

1. The Lagrangian function is

$$K(y, \beta^{-1}) = \langle x, y\rangle + \beta^{-1}[\lambda - F^{**}(y)]$$

Necessary and Sufficient Optimality Conditions

Proposition

Element y_β *solves* $\overline{x}(\lambda) := \sup\{\langle x, y\rangle : F(y) \leq \lambda\}$ *iff:*

$$y_\beta \in \partial F^*(\beta x)\,, \quad F(y_\beta) = \lambda\,, \quad \beta^{-1} \in \partial\overline{x}(\lambda)\,, \quad \beta^{-1} > 0$$

Proof.

1. The Lagrangian function is

$$K(y, \beta^{-1}) = \langle x, y\rangle + \beta^{-1}[\lambda - F^{**}(y)]$$

2. Zero in $\partial K(y, \beta^{-1})$ gives the following Kuhn-Tucker conditions:

$$\beta x \in \partial F^{**}(y_\beta)\,, \quad F^{**}(y_\beta) = F(y_\beta) = \lambda\,, \quad \beta^{-1} \in \partial\overline{x}(\lambda)$$

Necessary and Sufficient Optimality Conditions

Proposition

Element y_β *solves* $\overline{x}(\lambda) := \sup\{\langle x, y\rangle : F(y) \leq \lambda\}$ *iff:*

$$y_\beta \in \partial F^*(\beta x)\,, \quad F(y_\beta) = \lambda\,, \quad \beta^{-1} \in \partial\overline{x}(\lambda)\,, \quad \beta^{-1} > 0$$

Proof.

1. The Lagrangian function is

$$K(y, \beta^{-1}) = \langle x, y\rangle + \beta^{-1}[\lambda - F^{**}(y)]$$

2. Zero in $\partial K(y, \beta^{-1})$ gives the following Kuhn-Tucker conditions:

$$\beta x \in \partial F^{**}(y_\beta)\,, \quad F^{**}(y_\beta) = F(y_\beta) = \lambda\,, \quad \beta^{-1} \in \partial\overline{x}(\lambda)$$

3. Sufficient by convexity of F^{**}.

Necessary and Sufficient Optimality Conditions

Proposition

Element y_β *solves* $\overline{x}(\lambda) := \sup\{\langle x, y\rangle : F(y) \leq \lambda\}$ *iff:*

$$y_\beta \in \partial F^*(\beta x)\,, \quad F(y_\beta) = \lambda\,, \quad \beta^{-1} \in \partial\overline{x}(\lambda)\,, \quad \beta^{-1} > 0$$

Example

- For $F_{KL}(y)$

$$y_\beta = e^{\beta\, x}\, y_0 \quad \Longrightarrow \quad p_\beta = e^{\beta\, x - \Psi(\beta)}\, y_0$$

Necessary and Sufficient Optimality Conditions

Proposition

Element y_β *solves* $\overline{x}(\lambda) := \sup\{\langle x, y\rangle : F(y) \leq \lambda\}$ *iff:*

$$y_\beta \in \partial F^*(\beta x)\,, \quad F(y_\beta) = \lambda\,, \quad \beta^{-1} \in \partial\overline{x}(\lambda)\,, \quad \beta^{-1} > 0$$

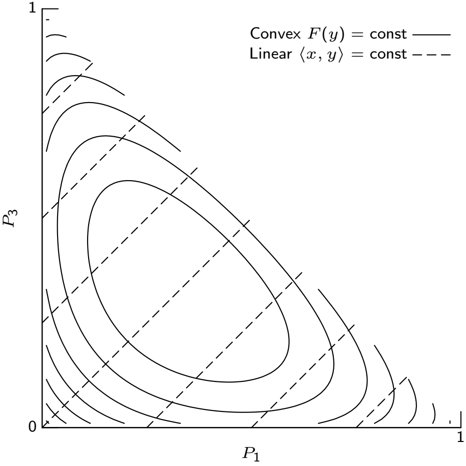
Example

- For $F_{KL}(y)$

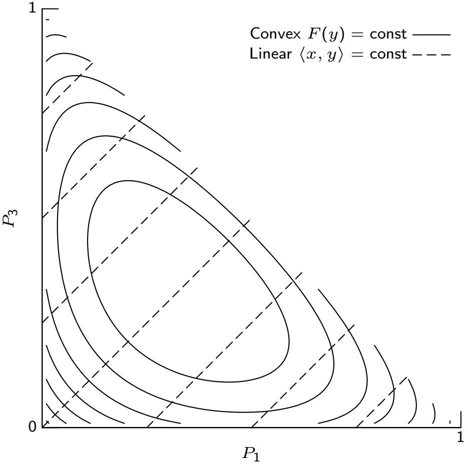
$$y_\beta = e^{\beta\, x}\, y_0 \quad \Longrightarrow \quad p_\beta = e^{\beta\, x - \Psi(\beta)}\, y_0$$

- If $x(a,b) = -\frac{1}{2}(a-b)^2$, a,$b \in \mathbb{R}^1$, p_β is Gaussian with $\sigma^2 = \beta^{-1}$.

Geometric Interpretation



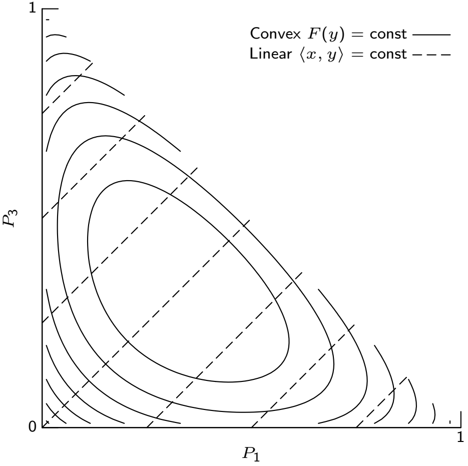
Geometric Interpretation



- The gradient of information must coincide with the gradient of the expected utility:

$$\partial\langle x, y_\beta\rangle \in \beta^{-1}\partial F(y_\beta)$$

Geometric Interpretation



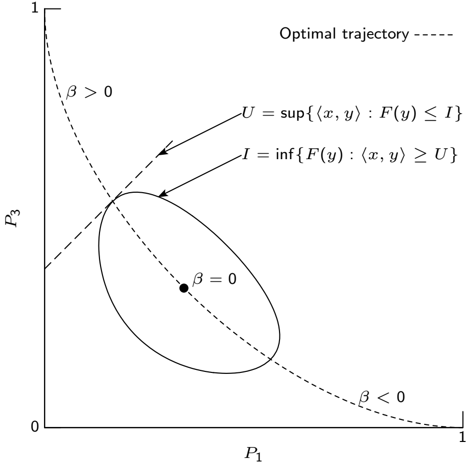
- The gradient of information must coincide with the gradient of the expected utility:

$$\partial\langle x, y_\beta\rangle \in \beta^{-1}\partial F(y_\beta)$$

- Optimal solutions are derived from the above eikonal inclusion

$$y_\beta \in \partial F^*(\beta x)$$

Geometric Interpretation



- The gradient of information must coincide with the gradient of the expected utility:

$$\partial\langle x, y_\beta\rangle \in \beta^{-1}\partial F(y_\beta)$$

- Optimal solutions are derived from the above eikonal inclusion

$$y_\beta \in \partial F^*(\beta x)$$

- If F is strictly convex, then y_β is unique.

Optimal Value Function $\overline{x}(\lambda)$

Proposition

Function $\overline{x}(\lambda) := \sup\{\langle x, y\rangle : F(y) \leq \lambda\}$ *is concave and strictly increasing for* $\lambda \in (\inf F, F(\delta_x))$.

Optimal Value Function $\overline{x}(\lambda)$

Proposition

Function $\overline{x}(\lambda) := \sup\{\langle x, y\rangle : F(y) \leq \lambda\}$ *is concave and strictly increasing for* $\lambda \in (\inf F, F(\delta_x))$.

Proof.

Let y_{β_1}, y_{β_2} be solutions for $\lambda_1 \leq \lambda_2$.

$$\langle x_1 - x_2, y_1 - y_2\rangle \geq 0\,, \qquad \forall\, y_i \in \partial F^*(x_i)$$

Optimal Value Function $\overline{x}(\lambda)$

Proposition

Function $\overline{x}(\lambda) := \sup\{\langle x, y\rangle : F(y) \leq \lambda\}$ *is concave and strictly increasing for* $\lambda \in (\inf F, F(\delta_x))$.

Proof.

Let y_{β_1}, y_{β_2} be solutions for $\lambda_1 \leq \lambda_2$.

$$\begin{aligned} & \langle x_1 - x_2, y_1 - y_2\rangle \geq 0\,, & \forall\, y_i \in \partial F^*(x_i) \\ \Longrightarrow\quad & (\beta_2 - \beta_1)\langle x, y_{\beta_2} - y_{\beta_1}\rangle \geq 0\,, & (y_\beta \in \partial F^*(\beta x)) \end{aligned}$$

Optimal Value Function $\overline{x}(\lambda)$

Proposition

Function $\overline{x}(\lambda) := \sup\{\langle x, y\rangle : F(y) \leq \lambda\}$ *is concave and strictly increasing for* $\lambda \in (\inf F, F(\delta_x))$.

Proof.

Let y_{β_1}, y_{β_2} be solutions for $\lambda_1 \leq \lambda_2$.

$$\begin{aligned}
& \langle x_1 - x_2, y_1 - y_2\rangle \geq 0\,, && \forall\, y_i \in \partial F^*(x_i) \\
\implies\ & (\beta_2 - \beta_1)\langle x, y_{\beta_2} - y_{\beta_1}\rangle \geq 0\,, && (y_\beta \in \partial F^*(\beta x)) \\
\implies\ & \beta_1 \leq \beta_2\,, && (\langle x, y_{\beta_2} - y_{\beta_1}\rangle \geq 0)
\end{aligned}$$

Optimal Value Function $\overline{x}(\lambda)$

Proposition

Function $\overline{x}(\lambda) := \sup\{\langle x, y\rangle : F(y) \leq \lambda\}$ *is concave and strictly increasing for* $\lambda \in (\inf F, F(\delta_x))$.

Proof.

Let y_{β_1}, y_{β_2} be solutions for $\lambda_1 \leq \lambda_2$.

$$\begin{aligned}
& \langle x_1 - x_2, y_1 - y_2\rangle \geq 0\,, && \forall\, y_i \in \partial F^*(x_i) \\
\Longrightarrow\quad & (\beta_2 - \beta_1)\langle x, y_{\beta_2} - y_{\beta_1}\rangle \geq 0\,, && (y_\beta \in \partial F^*(\beta x)) \\
\Longrightarrow\quad & \beta_1 \leq \beta_2\,, && (\langle x, y_{\beta_2} - y_{\beta_1}\rangle \geq 0) \\
\Longrightarrow\quad & \beta^{-1} \in \partial\overline{x}(\lambda) \text{ is antitone} &&
\end{aligned}$$

Optimal Value Function $\overline{x}(\lambda)$

Proposition

Function $\overline{x}(\lambda) := \sup\{\langle x, y\rangle : F(y) \leq \lambda\}$ *is concave and strictly increasing for* $\lambda \in (\inf F, F(\delta_x))$.

Proof.

Let y_{β_1}, y_{β_2} be solutions for $\lambda_1 \leq \lambda_2$.

$$
\begin{aligned}
& \langle x_1 - x_2, y_1 - y_2\rangle \geq 0\,, && \forall\, y_i \in \partial F^*(x_i) \\
\Longrightarrow\quad & (\beta_2 - \beta_1)\langle x, y_{\beta_2} - y_{\beta_1}\rangle \geq 0\,, && (y_\beta \in \partial F^*(\beta x)) \\
\Longrightarrow\quad & \beta_1 \leq \beta_2\,, && (\langle x, y_{\beta_2} - y_{\beta_1}\rangle \geq 0) \\
\Longrightarrow\quad & \beta^{-1} \in \partial\overline{x}(\lambda) \text{ is antitone} && \\
\Longrightarrow\quad & \overline{x}(\lambda) \text{ is concave} &&
\end{aligned}
$$

Optimal Value Function $\overline{x}(\lambda)$

Proposition

Function $\overline{x}(\lambda) := \sup\{\langle x, y\rangle : F(y) \leq \lambda\}$ *is concave and strictly increasing for* $\lambda \in (\inf F, F(\delta_x))$.

Proof.

Let y_{β_1}, y_{β_2} be solutions for $\lambda_1 \leq \lambda_2$.

$$
\begin{aligned}
& \langle x_1 - x_2, y_1 - y_2\rangle \geq 0\,, && \forall\, y_i \in \partial F^*(x_i) \\
\implies\; & (\beta_2 - \beta_1)\langle x, y_{\beta_2} - y_{\beta_1}\rangle \geq 0\,, && (y_\beta \in \partial F^*(\beta x)) \\
\implies\; & \beta_1 \leq \beta_2\,, && (\langle x, y_{\beta_2} - y_{\beta_1}\rangle \geq 0) \\
\implies\; & \beta^{-1} \in \partial\overline{x}(\lambda) \text{ is antitone} \\
\implies\; & \overline{x}(\lambda) \text{ is concave} \\
\implies\; & \overline{x}(\lambda) \text{ is strictly increasing if } \lambda \in (\inf F, F(\delta_x))
\end{aligned}
$$

□

Zero Solution

Proposition

Let $y_\beta \geq 0$ *for all* $\lambda = F(y)$. *Then* $y_\beta = 0$ *implies*

$$x = 0 \quad \text{or} \quad \inf F = F(0) \quad \text{or} \quad F(\delta_x) = F(0)$$

Zero Solution

Proposition

Let $y_\beta \geq 0$ *for all* $\lambda = F(y)$. *Then* $y_\beta = 0$ *implies*

$$x = 0 \quad \textit{or} \quad \inf F = F(0) \quad \textit{or} \quad F(\delta_x) = F(0)$$

Proof.

Assume $x \neq 0$ and

$$\inf F < F(0) < F(\delta_x)$$

Zero Solution

Proposition

Let $y_\beta \geq 0$ *for all* $\lambda = F(y)$*. Then* $y_\beta = 0$ *implies*

$$x = 0 \quad \textit{or} \quad \inf F = F(0) \quad \textit{or} \quad F(\delta_x) = F(0)$$

Proof.

Assume $x \neq 0$ and

$$\begin{aligned} & \inf F < F(0) < F(\delta_x) \\ \implies & \inf F < F(y_1) < F(0) < F(y_2) < F(\delta_x) \quad (F \text{ is closed}) \end{aligned}$$

Zero Solution

Proposition

Let $y_\beta \geq 0$ *for all* $\lambda = F(y)$*. Then* $y_\beta = 0$ *implies*

$$x = 0 \quad \textit{or} \quad \inf F = F(0) \quad \textit{or} \quad F(\delta_x) = F(0)$$

Proof.

Assume $x \neq 0$ and

$$\begin{aligned} & \inf F < F(0) < F(\delta_x) \\ \implies & \inf F < F(y_1) < F(0) < F(y_2) < F(\delta_x) \quad (F \text{ is closed}) \\ \implies & \langle x, y_1 \rangle < 0 < \langle x, y_2 \rangle \quad (\overline{x}(\lambda) \text{ str.inc.}) \end{aligned}$$

Zero Solution

Proposition

Let $y_\beta \geq 0$ *for all* $\lambda = F(y)$*. Then* $y_\beta = 0$ *implies*

$$x = 0 \quad \textit{or} \quad \inf F = F(0) \quad \textit{or} \quad F(\delta_x) = F(0)$$

Proof.

Assume $x \neq 0$ and

$$\begin{array}{rcl} & \inf F < F(0) < F(\delta_x) & \\ \implies & \inf F < F(y_1) < F(0) < F(y_2) < F(\delta_x) & (F \text{ is closed}) \\ \implies & \langle x, y_1 \rangle < 0 < \langle x, y_2 \rangle & (\overline{x}(\lambda) \text{ str.inc.}) \\ \implies & \langle x_+ - x_-, y_1 \rangle < 0 < \langle x_+ - x_-, y_2 \rangle & (x_+, x_- \in X_+ \text{ is reproducin} \end{array}$$

Zero Solution

Proposition

Let $y_\beta \geq 0$ *for all* $\lambda = F(y)$*. Then* $y_\beta = 0$ *implies*

$$x = 0 \quad \textit{or} \quad \inf F = F(0) \quad \textit{or} \quad F(\delta_x) = F(0)$$

Proof.

Assume $x \neq 0$ and

$$\begin{array}{rcl}
 & \inf F < F(0) < F(\delta_x) & \\
\Longrightarrow & \inf F < F(y_1) < F(0) < F(y_2) < F(\delta_x) & (F \text{ is closed}) \\
\Longrightarrow & \langle x, y_1 \rangle < 0 < \langle x, y_2 \rangle & (\overline{x}(\lambda) \text{ str.inc.}) \\
\Longrightarrow & \langle x_+ - x_-, y_1 \rangle < 0 < \langle x_+ - x_-, y_2 \rangle & (x_+, x_- \in X_+ \text{ is reproducin} \\
\Longrightarrow & x_+ < x_- \quad \text{and} \quad x_+ > x_- & (y_\beta \geq 0)
\end{array}$$

Zero Solution

Proposition

Let $y_\beta \geq 0$ *for all* $\lambda = F(y)$*. Then* $y_\beta = 0$ *implies*

$$x = 0 \quad \textit{or} \quad \inf F = F(0) \quad \textit{or} \quad F(\delta_x) = F(0)$$

Proof.

Assume $x \neq 0$ and

$$\begin{array}{lcl}
 & \inf F < F(0) < F(\delta_x) & \\
\Longrightarrow & \inf F < F(y_1) < F(0) < F(y_2) < F(\delta_x) & (F \text{ is closed}) \\
\Longrightarrow & \langle x, y_1 \rangle < 0 < \langle x, y_2 \rangle & (\overline{x}(\lambda) \text{ str.inc.}) \\
\Longrightarrow & \langle x_+ - x_-, y_1 \rangle < 0 < \langle x_+ - x_-, y_2 \rangle & (x_+, x_- \in X_+ \text{ is reproducin} \\
\Longrightarrow & x_+ < x_- \quad \text{and} \quad x_+ > x_- & (y_\beta \geq 0) \\
\Longrightarrow & x = 0 &
\end{array}$$

Subsets of Measure Zero

- σ-algebra $\mathcal{R}(\Omega) \iff$ linear algebra (space) $X := C_c(\Omega)$

Subsets of Measure Zero

- σ-algebra $\mathcal{R}(\Omega) \iff$ linear algebra (space) $X := C_c(\Omega)$
- Sub-algebra $\mathcal{R}(E) \subset \mathcal{R}(\Omega) \iff$ sub-algebra (sub-space) $M \subset X$

Subsets of Measure Zero

- σ-algebra $\mathcal{R}(\Omega) \iff$ linear algebra (space) $X := C_c(\Omega)$
- Sub-algebra $\mathcal{R}(E) \subset \mathcal{R}(\Omega) \iff$ sub-algebra (sub-space) $M \subset X$
- Measure $y(E) = 0$ on $E \subset \Omega \iff y(M) = \langle x, y \rangle = 0$, $y \in M^\perp$

$$M^\perp := \{y \in Y : \langle x, y \rangle = 0 \,, \forall\, x \in M\}$$

Subsets of Measure Zero

- σ-algebra $\mathcal{R}(\Omega) \iff$ linear algebra (space) $X := C_c(\Omega)$
- Sub-algebra $\mathcal{R}(E) \subset \mathcal{R}(\Omega) \iff$ sub-algebra (sub-space) $M \subset X$
- Measure $y(E) = 0$ on $E \subset \Omega \iff y(M) = \langle x, y\rangle = 0$, $y \in M^\perp$

$$M^\perp := \{y \in Y : \langle x, y\rangle = 0\,,\ \forall\, x \in M\}$$

- Dual of $M \subset X$ is factor space $M^* = Y/M^\perp$ of equivalence classes

$$[y] \in Y/M^\perp\,, \qquad [0] = M^\perp$$

Subsets of Measure Zero

- σ-algebra $\mathcal{R}(\Omega) \iff$ linear algebra (space) $X := C_c(\Omega)$
- Sub-algebra $\mathcal{R}(E) \subset \mathcal{R}(\Omega) \iff$ sub-algebra (sub-space) $M \subset X$
- Measure $y(E) = 0$ on $E \subset \Omega \iff y(M) = \langle x, y \rangle = 0$, $y \in M^\perp$

$$M^\perp := \{y \in Y : \langle x, y \rangle = 0\,,\ \forall\, x \in M\}$$

- Dual of $M \subset X$ is factor space $M^* = Y/M^\perp$ of equivalence classes

$$[y] \in Y/M^\perp\,, \qquad [0] = M^\perp$$

- Let $P_M : X \to M \subset X$ be a continuous linear projection.

Subsets of Measure Zero

- σ-algebra $\mathcal{R}(\Omega) \iff$ linear algebra (space) $X := C_c(\Omega)$
- Sub-algebra $\mathcal{R}(E) \subset \mathcal{R}(\Omega) \iff$ sub-algebra (sub-space) $M \subset X$
- Measure $y(E) = 0$ on $E \subset \Omega \iff y(M) = \langle x, y\rangle = 0$, $y \in M^\perp$

$$M^\perp := \{y \in Y : \langle x, y\rangle = 0\,,\ \forall\, x \in M\}$$

- Dual of $M \subset X$ is factor space $M^* = Y/M^\perp$ of equivalence classes

$$[y] \in Y/M^\perp\,, \qquad [0] = M^\perp$$

- Let $P_M : X \to M \subset X$ be a continuous linear projection.
- Then the restriction of F^* to M and its dual are

$$F^*(x)|_M = F^*(P_M x) \qquad F^{**}([y]) := \inf\{F^{**}(y) : y \in [y]\}$$

Mutual Absolute Continuity

Theorem

Let $\{y_\beta\}_x \ni y_\beta \geq 0$ *for all* $\lambda = F(y)$ *and* F^* *strictly convex. Then*

Mutual Absolute Continuity

Theorem

Let $\{y_\beta\}_x \ni y_\beta \geq 0$ *for all* $\lambda = F(y)$ *and* F^* *strictly convex. Then*

1. $\exists\, \{y^\circ_\beta\}_x \subseteq \{y_\beta\}_x$ *such that* y°_β *are mutually absolutely continuous* $\forall\, \lambda \in (\inf F, F(\delta_x))$.

Mutual Absolute Continuity

Theorem

Let $\{y_\beta\}_x \ni y_\beta \geq 0$ *for all* $\lambda = F(y)$ *and* F^* *strictly convex. Then*

1. $\exists\, \{y^\circ_\beta\}_x \subseteq \{y_\beta\}_x$ *such that* y°_β *are mutually absolutely continuous* $\forall\, \lambda \in (\inf F, F(\delta_x))$.
2. y_0 *and* δ_x *(*$\inf F = F(y_0)$ *and* $\sup\{\langle x, y\rangle : y \in \mathrm{dom} F\} = \langle x, \delta_x\rangle$*) are absolutely continuous w.r.t. all* y°_β.

Mutual Absolute Continuity

Theorem

Let $\{y_\beta\}_x \ni y_\beta \geq 0$ *for all* $\lambda = F(y)$ *and* F^* *strictly convex. Then*

1. $\exists \{y_\beta^\circ\}_x \subseteq \{y_\beta\}_x$ *such that* y_β° *are mutually absolutely continuous* $\forall \lambda \in (\inf F, F(\delta_x))$.
2. y_0 *and* δ_x *(*$\inf F = F(y_0)$ *and* $\sup\{\langle x, y\rangle : y \in \mathrm{dom} F\} = \langle x, \delta_x\rangle$*) are absolutely continuous w.r.t. all* y_β°.
3. *If* F^{**} *is strictly convex, then* $\{y_\beta^\circ\}_x = \{y_\beta\}_x \setminus \{y_0, \delta_x\}$.

Mutual Absolute Continuity

Theorem

Let $\{y_\beta\}_x \ni y_\beta \geq 0$ for all $\lambda = F(y)$ and F^* strictly convex. Then

1. $\exists\, \{y^\circ_\beta\}_x \subseteq \{y_\beta\}_x$ *such that* y°_β *are mutually absolutely continuous* $\forall\, \lambda \in (\inf F, F(\delta_x))$.
2. y_0 *and* δ_x ($\inf F = F(y_0)$ *and* $\sup\{\langle x, y\rangle : y \in \mathrm{dom} F\} = \langle x, \delta_x\rangle$) *are absolutely continuous w.r.t. all* y°_β.
3. *If* F^{**} *is strictly convex, then* $\{y^\circ_\beta\}_x = \{y_\beta\}_x \setminus \{y_0, \delta_x\}$.

Proof.

- $y_\beta(M) = 0 \iff y_\beta \in [0] \in \partial F^*(\beta P_M x)$, $0 < \beta^{-1} < \infty$

Mutual Absolute Continuity

Theorem

Let $\{y_\beta\}_x \ni y_\beta \geq 0$ *for all* $\lambda = F(y)$ *and* F^* *strictly convex. Then*

1. $\exists\,\{y_\beta^\circ\}_x \subseteq \{y_\beta\}_x$ *such that* y_β° *are mutually absolutely continuous* $\forall\,\lambda \in (\inf F, F(\delta_x))$.
2. y_0 *and* δ_x *(*$\inf F = F(y_0)$ *and* $\sup\{\langle x, y\rangle : y \in \mathrm{dom}F\} = \langle x, \delta_x\rangle$*) are absolutely continuous w.r.t. all* y_β°.
3. *If* F^{**} *is strictly convex, then* $\{y_\beta^\circ\}_x = \{y_\beta\}_x \setminus \{y_0, \delta_x\}$.

Proof.

- $y_\beta(M) = 0 \iff y_\beta \in [0] \in \partial F^*(\beta P_M x)$, $0 < \beta^{-1} < \infty$
- $P_M x = 0$ or $\inf F^{**} = F^{**}([0])$ or $F^{**}([\delta_x]) = F^{**}([0])$

Mutual Absolute Continuity

Theorem

Let $\{y_\beta\}_x \ni y_\beta \geq 0$ *for all* $\lambda = F(y)$ *and* F^* *strictly convex. Then*

1. $\exists \{y^\circ_\beta\}_x \subseteq \{y_\beta\}_x$ *such that* y°_β *are mutually absolutely continuous* $\forall \lambda \in (\inf F, F(\delta_x))$.
2. y_0 *and* δ_x ($\inf F = F(y_0)$ *and* $\sup\{\langle x, y\rangle : y \in \mathrm{dom} F\} = \langle x, \delta_x\rangle$) *are absolutely continuous w.r.t. all* y°_β.
3. *If* F^{**} *is strictly convex, then* $\{y^\circ_\beta\}_x = \{y_\beta\}_x \setminus \{y_0, \delta_x\}$.

Proof.

- $y_\beta(M) = 0 \iff y_\beta \in [0] \in \partial F^*(\beta P_M x)$, $0 < \beta^{-1} < \infty$
- $P_M x = 0$ or $\inf F^{**} = F^{**}([0])$ or $F^{**}([\delta_x]) = F^{**}([0])$
- $\partial F^{**}(y) = \{x\}$ implies $\beta P_M x = 0$, $\forall \beta \in \overline{\mathbb{R}}$.

Mutual Absolute Continuity

Theorem

Let $\{y_\beta\}_x \ni y_\beta \geq 0$ *for all* $\lambda = F(y)$ *and* F^* *strictly convex. Then*

1. $\exists \{y_\beta^\circ\}_x \subseteq \{y_\beta\}_x$ *such that* y_β° *are mutually absolutely continuous* $\forall \lambda \in (\inf F, F(\delta_x))$.
2. y_0 *and* δ_x *(*$\inf F = F(y_0)$ *and* $\sup\{\langle x, y\rangle : y \in \mathrm{dom} F\} = \langle x, \delta_x\rangle$*) are absolutely continuous w.r.t. all* y_β°.
3. *If* F^{**} *is strictly convex, then* $\{y_\beta^\circ\}_x = \{y_\beta\}_x \setminus \{y_0, \delta_x\}$.

Proof.

- $y_\beta(M) = 0 \iff y_\beta \in [0] \in \partial F^*(\beta P_M x)$, $0 < \beta^{-1} < \infty$
- $P_M x = 0$ or $\inf F^{**} = F^{**}([0])$ or $F^{**}([\delta_x]) = F^{**}([0])$
- $\partial F^{**}(y) = \{x\}$ implies $\beta P_M x = 0$, $\forall \beta \in \overline{\mathbb{R}}$.
- $\forall \lambda \in [\inf F, \sup F]\ \exists y_\beta \in \{y_\beta\}_x : y_\beta(M) = 0$

Mutual Absolute Continuity

Theorem

Let $\{y_\beta\}_x \ni y_\beta \geq 0$ *for all* $\lambda = F(y)$ *and* F^* *strictly convex. Then*

1. $\exists\,\{y^\circ_\beta\}_x \subseteq \{y_\beta\}_x$ *such that* y°_β *are mutually absolutely continuous* $\forall\,\lambda \in (\inf F, F(\delta_x))$.
2. y_0 *and* δ_x *(*$\inf F = F(y_0)$ *and* $\sup\{\langle x, y\rangle : y \in \mathrm{dom} F\} = \langle x, \delta_x\rangle$*) are absolutely continuous w.r.t. all* y°_β.
3. *If* F^{**} *is strictly convex, then* $\{y^\circ_\beta\}_x = \{y_\beta\}_x \setminus \{y_0, \delta_x\}$.

Proof.

- $y_\beta(M) = 0 \iff y_\beta \in [0] \in \partial F^*(\beta P_M x)$, $0 < \beta^{-1} < \infty$
- $P_M x = 0$ or $\inf F^{**} = F^{**}([0])$ or $F^{**}([\delta_x]) = F^{**}([0])$
- $\partial F^{**}(y) = \{x\}$ implies $\beta P_M x = 0$, $\forall \beta \in \overline{\mathbb{R}}$.
- $\forall\,\lambda \in [\inf F, \sup F]\ \exists\, y_\beta \in \{y_\beta\}_x : y_\beta(M) = 0$
- Construct $\{y^\circ_\beta\}_x$: $M = \sup\{M' \subset X : \exists\, y^\circ_\beta \in \{y_\beta\}_x,\ y^\circ_\beta(M') = 0\}$

Strict Convexity of F^*

- If F^* is not strictly convex, then for some $x \in X$

$$\exists\, \beta_1 x \neq \beta_2 x \in \partial F(y_\beta)$$

Strict Convexity of F^*

- If F^* is not strictly convex, then for some $x \in X$

$$\exists\, \beta_1 x \neq \beta_2 x \in \partial F(y_\beta)$$

- If x is understood as a utility, then information F does not 'distinguish' between $\beta_1 x$ or βx_2:

$$F(y_\beta) = \lambda \quad \Longleftrightarrow \quad \beta_1 x \neq \beta_2 x$$

Strict Convexity of F^*

- If F^* is not strictly convex, then for some $x \in X$

$$\exists\, \beta_1 x \neq \beta_2 x \in \partial F(y_\beta)$$

- If x is understood as a utility, then information F does not 'distinguish' between $\beta_1 x$ or βx_2:

$$F(y_\beta) = \lambda \quad \iff \quad \beta_1 x \neq \beta_2 x$$

- Information distance $I(y, y_0) = F(y)$ is often required to satisfy the additivity axiom (Chentsov, 1972):

$$I(yz, y_0 z_0) = I(y, y_0) + I(z, z_0)$$

Strict Convexity of F^*

- If F^* is not strictly convex, then for some $x \in X$

$$\exists\, \beta_1 x \neq \beta_2 x \in \partial F(y_\beta)$$

- If x is understood as a utility, then information F does not 'distinguish' between $\beta_1 x$ or βx_2:

$$F(y_\beta) = \lambda \quad \Longleftrightarrow \quad \beta_1 x \neq \beta_2 x$$

- Information distance $I(y, y_0) = F(y)$ is often required to satisfy the additivity axiom (Chentsov, 1972):

$$I(yz, y_0 z_0) = I(y, y_0) + I(z, z_0)$$

- Hence

$$\partial F(\cdot) = \ln(\cdot) \quad \text{and} \quad \partial F^*(\cdot) = \exp(\cdot)$$

Optimal Transition Kernels

Corollary

Let $p_\beta \in \mathcal{P}(A \times B)$ *be solutions* $\forall\, \lambda = F(p)$. *Let* F^* *be strictly convex.*
Let $p_0 \in \partial F^*(0) \subset \mathrm{Int}(\mathcal{P}(A \times B))$.

Optimal Transition Kernels

Corollary

Let $p_\beta \in \mathcal{P}(A \times B)$ *be solutions* $\forall\, \lambda = F(p)$*. Let* F^* *be strictly convex.*
Let $p_0 \in \partial F^*(0) \subset \mathrm{Int}(\mathcal{P}(A \times B))$*.*
Then p_β *is deterministic if and only if* $\lambda \geq F(\delta_x)$ *or* $\langle x, p_\beta \rangle = \langle x, \delta_x \rangle$*.*

Optimal Transition Kernels

Corollary

Let $p_\beta \in \mathcal{P}(A \times B)$ *be solutions* $\forall\, \lambda = F(p)$. *Let* F^* *be strictly convex.*
Let $p_0 \in \partial F^*(0) \subset \mathrm{Int}(\mathcal{P}(A \times B))$.
Then p_β *is deterministic if and only if* $\lambda \geq F(\delta_x)$ *or* $\langle x, p_\beta \rangle = \langle x, \delta_x \rangle$.

Proof.

$(\Rightarrow)$ Assume $p_\beta \iff F(p_\beta) < F(\delta_x)$ $(\langle x, p_\beta \rangle < \langle x, \delta_x \rangle)$ is deterministic:

$$p_\beta(f(B_j) \mid B_j) = 1 \text{ and } p_\beta(A \setminus f(B_j) \mid B_j) = 0$$

Optimal Transition Kernels

Corollary

Let $p_\beta \in \mathcal{P}(A \times B)$ *be solutions* $\forall\, \lambda = F(p)$. *Let* F^* *be strictly convex.*
Let $p_0 \in \partial F^*(0) \subset \mathrm{Int}(\mathcal{P}(A \times B))$.
Then p_β *is deterministic if and only if* $\lambda \geq F(\delta_x)$ *or* $\langle x, p_\beta \rangle = \langle x, \delta_x \rangle$.

Proof.

$(\Rightarrow)$ Assume $p_\beta \iff F(p_\beta) < F(\delta_x)$ $(\langle x, p_\beta \rangle < \langle x, \delta_x \rangle)$ is deterministic:

$$\begin{aligned} & p_\beta(f(B_j) \mid B_j) = 1 \text{ and } p_\beta(A \setminus f(B_j) \mid B_j) = 0 \\ \Longrightarrow \quad & p_\beta(A \setminus f(B_j), B_j) = 0\,, \quad p_\beta \notin \mathrm{Int}(\mathcal{P}(A \times B)) \end{aligned}$$

Optimal Transition Kernels

Corollary

Let $p_\beta \in \mathcal{P}(A \times B)$ *be solutions* $\forall\, \lambda = F(p)$. *Let* F^* *be strictly convex.*
Let $p_0 \in \partial F^*(0) \subset \mathrm{Int}(\mathcal{P}(A \times B))$.
Then p_β *is deterministic if and only if* $\lambda \geq F(\delta_x)$ *or* $\langle x, p_\beta \rangle = \langle x, \delta_x \rangle$.

Proof.

$(\Rightarrow)$ Assume $p_\beta \iff F(p_\beta) < F(\delta_x)$ $(\langle x, p_\beta \rangle < \langle x, \delta_x \rangle)$ is deterministic:

$$
\begin{aligned}
& p_\beta(f(B_j) \mid B_j) = 1 \text{ and } p_\beta(A \setminus f(B_j) \mid B_j) = 0 \\
\implies \quad & p_\beta(A \setminus f(B_j), B_j) = 0\,, \quad p_\beta \notin \mathrm{Int}(\mathcal{P}(A \times B)) \\
\implies \quad & p_\beta \notin \partial F^*(0)\,, \quad F(p_\beta) = \lambda \in (\inf F, F(\delta_x))
\end{aligned}
$$

Optimal Transition Kernels

Corollary

Let $p_\beta \in \mathcal{P}(A \times B)$ *be solutions* $\forall\, \lambda = F(p)$. *Let* F^* *be strictly convex.*
Let $p_0 \in \partial F^*(0) \subset \mathrm{Int}(\mathcal{P}(A \times B))$.
Then p_β *is deterministic if and only if* $\lambda \geq F(\delta_x)$ *or* $\langle x, p_\beta \rangle = \langle x, \delta_x \rangle$.

Proof.

$(\Rightarrow)$ Assume $p_\beta \iff F(p_\beta) < F(\delta_x)$ $(\langle x, p_\beta \rangle < \langle x, \delta_x \rangle)$ is deterministic:

$$
\begin{aligned}
& p_\beta(f(B_j) \mid B_j) = 1 \text{ and } p_\beta(A \setminus f(B_j) \mid B_j) = 0 \\
\Longrightarrow \quad & p_\beta(A \setminus f(B_j), B_j) = 0\,, \quad p_\beta \notin \mathrm{Int}(\mathcal{P}(A \times B)) \\
\Longrightarrow \quad & p_\beta \notin \partial F^*(0)\,, \quad F(p_\beta) = \lambda \in (\inf F, F(\delta_x)) \\
\Longrightarrow \quad & p_\beta(A \setminus f(B_j), B_j) = 0\,, \ \forall\, \lambda \in [\inf F, \sup F]
\end{aligned}
$$

Optimal Transition Kernels

Corollary

Let $p_\beta \in \mathcal{P}(A \times B)$ *be solutions* $\forall\, \lambda = F(p)$. *Let* F^* *be strictly convex.*
Let $p_0 \in \partial F^*(0) \subset \mathrm{Int}(\mathcal{P}(A \times B))$.
Then p_β *is deterministic if and only if* $\lambda \geq F(\delta_x)$ *or* $\langle x, p_\beta \rangle = \langle x, \delta_x \rangle$.

Proof.

$(\Rightarrow)$ Assume $p_\beta \iff F(p_\beta) < F(\delta_x)$ $(\langle x, p_\beta\rangle < \langle x, \delta_x\rangle)$ is deterministic:

$$
\begin{aligned}
& p_\beta(f(B_j) \mid B_j) = 1 \text{ and } p_\beta(A \setminus f(B_j) \mid B_j) = 0 \\
\Longrightarrow \quad & p_\beta(A \setminus f(B_j), B_j) = 0\,, \quad p_\beta \notin \mathrm{Int}(\mathcal{P}(A \times B)) \\
\Longrightarrow \quad & p_\beta \notin \partial F^*(0)\,, \quad F(p_\beta) = \lambda \in (\inf F, F(\delta_x)) \\
\Longrightarrow \quad & p_\beta(A \setminus f(B_j), B_j) = 0\,, \ \forall\, \lambda \in [\inf F, \sup F] \\
\Longrightarrow \quad & \exists\, p_0(A \setminus f(B_j), B_j) = 0 \text{ and } p_0 \notin \mathrm{Int}(\mathcal{P}(A \times B))
\end{aligned}
$$

Optimal Transition Kernels

Corollary

Let $p_\beta \in \mathcal{P}(A \times B)$ be solutions $\forall\, \lambda = F(p)$. Let F^* be strictly convex.
Let $p_0 \in \partial F^*(0) \subset \mathrm{Int}(\mathcal{P}(A \times B))$.
Then p_β is deterministic if and only if $\lambda \geq F(\delta_x)$ or $\langle x, p_\beta \rangle = \langle x, \delta_x \rangle$.

Proof.

($\Rightarrow$) Assume $p_\beta \iff F(p_\beta) < F(\delta_x)$ ($\langle x, p_\beta \rangle < \langle x, \delta_x \rangle$) is deterministic:

$$
\begin{aligned}
& p_\beta(f(B_j) \mid B_j) = 1 \text{ and } p_\beta(A \setminus f(B_j) \mid B_j) = 0 \\
\Longrightarrow\quad & p_\beta(A \setminus f(B_j), B_j) = 0\,, \quad p_\beta \notin \mathrm{Int}(\mathcal{P}(A \times B)) \\
\Longrightarrow\quad & p_\beta \notin \partial F^*(0)\,, \quad F(p_\beta) = \lambda \in (\inf F, F(\delta_x)) \\
\Longrightarrow\quad & p_\beta(A \setminus f(B_j), B_j) = 0\,,\ \forall\, \lambda \in [\inf F, \sup F] \\
\Longrightarrow\quad & \exists\, p_0(A \setminus f(B_j), B_j) = 0 \text{ and } p_0 \notin \mathrm{Int}(\mathcal{P}(A \times B))
\end{aligned}
$$

($\Leftarrow$) obvious.

Strict Inequalities

Corollary

Let $p_\beta \in \mathcal{P}(A \times B)$ *be solutions* $\forall\, \lambda = F(p)$*. Let* F^* *be strictly convex.*
Let $p_0 \in \partial F^*(0) \subset \mathrm{Int}(\mathcal{P}(A \times B))$*.*

Strict Inequalities

Corollary

Let $p_\beta \in \mathcal{P}(A \times B)$ *be solutions* $\forall\, \lambda = F(p)$. *Let* F^* *be strictly convex.*
Let $p_0 \in \partial F^*(0) \subset \mathrm{Int}(\mathcal{P}(A \times B))$.
Then for all deterministic $p_f \in \mathcal{P}(A \times B)$:

Strict Inequalities

Corollary

Let $p_\beta \in \mathcal{P}(A \times B)$ *be solutions* $\forall\,\lambda = F(p)$. *Let* F^* *be strictly convex.*
Let $p_0 \in \partial F^*(0) \subset \mathrm{Int}(\mathcal{P}(A \times B))$.
Then for all deterministic $p_f \in \mathcal{P}(A \times B)$*:*

$$\langle x, p_f \rangle < \langle x, p_\beta \rangle \qquad F(p_f) = F(p_\beta) \in (\inf F, F(\delta_x))$$

Strict Inequalities

Corollary

Let $p_\beta \in \mathcal{P}(A \times B)$ *be solutions* $\forall\, \lambda = F(p)$. *Let* F^* *be strictly convex.*
Let $p_0 \in \partial F^*(0) \subset \mathrm{Int}(\mathcal{P}(A \times B))$.
Then for all deterministic $p_f \in \mathcal{P}(A \times B)$:

$$\begin{aligned} \langle x, p_f \rangle < \langle x, p_\beta \rangle &\qquad F(p_f) = F(p_\beta) \in (\inf F, F(\delta_x)) \\ F(p_f) > F(p_\beta) &\qquad \langle x, p_f \rangle = \langle x, p_\beta \rangle \in (\overline{\upsilon}_0, \langle x, \delta_x \rangle) \end{aligned}$$

Strict Inequalities

Corollary

Let $p_\beta \in \mathcal{P}(A \times B)$ *be solutions* $\forall\, \lambda = F(p)$*. Let* F^* *be strictly convex.*
Let $p_0 \in \partial F^*(0) \subset \mathrm{Int}(\mathcal{P}(A \times B))$*.*
Then for all deterministic $p_f \in \mathcal{P}(A \times B)$*:*

$$\begin{array}{rl} \langle x, p_f\rangle < \langle x, p_\beta\rangle & \quad F(p_f) = F(p_\beta) \in (\inf F, F(\delta_x)) \\ F(p_f) > F(p_\beta) & \quad \langle x, p_f\rangle = \langle x, p_\beta\rangle \in (\overline{\upsilon}_0, \langle x, \delta_x\rangle) \end{array}$$

Proof.

Based on inequalities (equalities $\iff$ $p_\beta \in \partial F^*(\beta x)$):

$$\begin{array}{rl} \langle x, y\rangle \leq F^*(x) + F(y)\,, & \quad F(y) \geq \langle x, y\rangle - F^*(x) \\ \beta\langle x, p_\beta\rangle = F^*(\beta x) + F(p_\beta)\,, & \quad F(p_\beta) = \beta\langle x, p_\beta\rangle - F^*(\beta x) \end{array}$$

$\square$

Example: Motivation

- Construct an example, where $\exists\,\lambda \in (\inf F, F(\delta_x))$ or $\upsilon \in (\overline{\upsilon}_0, \langle x, \delta_x\rangle)$ and for any deterministic p_f

$$\begin{aligned} \mathbb{E}_{p_f}\{x\} &= -\infty \qquad \text{if } I(p_f, p_0) \leq \lambda \\ I(p_f, p_0) &= \infty \qquad \text{if } \mathbb{E}_{p_f}\{x\} \geq \upsilon \end{aligned}$$

Example: Motivation

- Construct an example, where $\exists\,\lambda \in (\inf F, F(\delta_x))$ or $\upsilon \in (\overline{\upsilon}_0, \langle x, \delta_x\rangle)$ and for any deterministic p_f

$$\begin{aligned} \mathbb{E}_{p_f}\{x\} &= -\infty \qquad \text{if } I(p_f, p_0) \leq \lambda \\ I(p_f, p_0) &= \infty \qquad \text{if } \mathbb{E}_{p_f}\{x\} \geq \upsilon \end{aligned}$$

- Strict inequalities $\langle x, p_f\rangle < \langle x, p_\beta\rangle$, $F(p_f) > F(p_\beta)$ of Corollary 6 would imply that for non-deterministic solution p_β

$$\begin{aligned} \mathbb{E}_{p_\beta}\{x\} &> -\infty \qquad \text{and } I(p_\beta, p_0) \leq \lambda \\ I(p_\beta, p_0) &< \infty \qquad \text{and } \mathbb{E}_{p_\beta}\{x\} \geq \upsilon \end{aligned}$$

Communication Channel

- $p \in \mathcal{P}(A \times B)$, where $p = p(A_i \cap B_j) = p(A_i \mid B_j)\, p(B_j)$.

Communication Channel

- $p \in \mathcal{P}(A \times B)$, where $p = p(A_i \cap B_j) = p(A_i \mid B_j)\, p(B_j)$.
- No information is communicated if $p(A_i \mid B_j) = p(A_i)$.

Communication Channel

- $p \in \mathcal{P}(A \times B)$, where $p = p(A_i \cap B_j) = p(A_i \mid B_j)\, p(B_j)$.
- No information is communicated if $p(A_i \mid B_j) = p(A_i)$.
- The maximum of information is communicated by deterministic $p_f(A_i \mid b) = \delta_{f(b)}(A_i)$, where f is an injective function.

Communication Channel

- $p \in \mathcal{P}(A \times B)$, where $p = p(A_i \cap B_j) = p(A_i \mid B_j)\, p(B_j)$.
- No information is communicated if $p(A_i \mid B_j) = p(A_i)$.
- The maximum of information is communicated by deterministic $p_f(A_i \mid b) = \delta_{f(b)}(A_i)$, where f is an injective function.
- If $|f(B)| < |B|$ (non-injective), then $a = f(b)$ communicates less information than an injective function.

Communication Channel

- $p \in \mathcal{P}(A \times B)$, where $p = p(A_i \cap B_j) = p(A_i \mid B_j)\, p(B_j)$.
- No information is communicated if $p(A_i \mid B_j) = p(A_i)$.
- The maximum of information is communicated by deterministic $p_f(A_i \mid b) = \delta_{f(b)}(A_i)$, where f is an injective function.
- If $|f(B)| < |B|$ (non-injective), then $a = f(b)$ communicates less information than an injective function.
- Constant functions $|f(B)| = 1$ communicate no information.

Communication Channel

- $p \in \mathcal{P}(A \times B)$, where $p = p(A_i \cap B_j) = p(A_i \mid B_j)\, p(B_j)$.
- No information is communicated if $p(A_i \mid B_j) = p(A_i)$.
- The maximum of information is communicated by deterministic $p_f(A_i \mid b) = \delta_{f(b)}(A_i)$, where f is an injective function.
- If $|f(B)| < |B|$ (non-injective), then $a = f(b)$ communicates less information than an injective function.
- Constant functions $|f(B)| = 1$ communicate no information.

Definition (Shannon Information)

$F_S(p) = I_{KL}(p, q)$, where $p = p(A_i \cap B_j)$ and $q = p(A_i)\, p(B_j)$

Communication Channel

- $p \in \mathcal{P}(A \times B)$, where $p = p(A_i \cap B_j) = p(A_i \mid B_j)\, p(B_j)$.
- No information is communicated if $p(A_i \mid B_j) = p(A_i)$.
- The maximum of information is communicated by deterministic $p_f(A_i \mid b) = \delta_{f(b)}(A_i)$, where f is an injective function.
- If $|f(B)| < |B|$ (non-injective), then $a = f(b)$ communicates less information than an injective function.
- Constant functions $|f(B)| = 1$ communicate no information.

Definition (Shannon Information)

$F_S(p) = I_{KL}(p, q)$, where $p = p(A_i \cap B_j)$ and $q = p(A_i)\, p(B_j)$

$$\begin{aligned} F_S(p) &:= \int_{A\times B} \ln\left[\frac{dp(a,b)}{dp(a)\, dp(b)}\right] dp(a,b) \\ &= \int_B dp(b) \int_A \ln\left[\frac{dp(a \mid b)}{dp(a)}\right] dp(a \mid b) \\ &= H(a) - H(a \mid b) \end{aligned}$$

Communication Channel

- $p \in \mathcal{P}(A \times B)$, where $p = p(A_i \cap B_j) = p(A_i \mid B_j)\, p(B_j)$.
- No information is communicated if $p(A_i \mid B_j) = p(A_i)$.
- The maximum of information is communicated by deterministic $p_f(A_i \mid b) = \delta_{f(b)}(A_i)$, where f is an injective function.
- If $|f(B)| < |B|$ (non-injective), then $a = f(b)$ communicates less information than an injective function.
- Constant functions $|f(B)| = 1$ communicate no information.

Definition (Shannon Information)

$F_S(p) = I_{KL}(p, q)$, where $p = p(A_i \cap B_j)$ and $q = p(A_i)\, p(B_j)$

$$\begin{aligned} F_S(p_\beta) &= \int_A dp(a) \int_B \ln \frac{dp(b \mid a)}{dp(b)}\, dp(b \mid a) \\ &= \int_A dp(a) \int_B \left\{ \beta\, x(a,b) - \ln \int_B e^{\beta\, x(a,b)}\, db - \ln \frac{dp(b)}{db} \right\} dp(b \mid a) \\ &= \beta\, \mathbb{E}_{p_\beta}\{x\} - \Psi_0(\beta) + H\{p(b)\} \end{aligned}$$

Communication Channel

- $p \in \mathcal{P}(A \times B)$, where $p = p(A_i \cap B_j) = p(A_i \mid B_j)\, p(B_j)$.
- No information is communicated if $p(A_i \mid B_j) = p(A_i)$.
- The maximum of information is communicated by deterministic $p_f(A_i \mid b) = \delta_{f(b)}(A_i)$, where f is an injective function.
- If $|f(B)| < |B|$ (non-injective), then $a = f(b)$ communicates less information than an injective function.
- Constant functions $|f(B)| = 1$ communicate no information.

Definition (Shannon Information)

$F_S(p) = I_{KL}(p, q)$, where $p = p(A_i \cap B_j)$ and $q = p(A_i)\, p(B_j)$

$$\begin{aligned} F_S(p_f) &= \int_B dp(b) \int_A \ln \frac{\delta(f(b) - b)}{dp(a)}\, \delta(f(b) - b) \\ &= -\int_B dp(b) \ln\Big(dp(f(b))\Big) = -\int_B dp(b) \ln\left(\frac{|B|}{|f(B)|}\, dp(b)\right) \\ &= \ln|f(B)| - \ln|B| + H\{p(b)\} \end{aligned}$$

Optimal Mean-Squared Communication

- Let $a \in A$, $b \in (B, \mathcal{B}, p)$ and utility $x : A \times B \to \mathbb{R}$.

Optimal Mean-Squared Communication

- Let $a \in A$, $b \in (B, \mathcal{B}, p)$ and utility $x : A \times B \to \mathbb{R}$.
- Find $p(A_i \mid b)$ that maximises

$$\mathbb{E}_p\{x\} = \int_A \int_B x(a,b)\, dp(a,b) = \int_B dp(b) \int_A x(a,b)\, dp(a \mid b)$$

subject to $F(p(A \cap B)) \leq \lambda$.

Optimal Mean-Squared Communication

- Let $a \in A$, $b \in (B, \mathcal{B}, p)$ and utility $x : A \times B \to \mathbb{R}$.
- Find $p(A_i \mid b)$ that maximises

$$\mathbb{E}_p\{x\} = \int_A \int_B x(a,b)\, dp(a,b) = \int_B dp(b) \int_A x(a,b)\, dp(a \mid b)$$

  subject to $F(p(A \cap B)) \leq \lambda$.
- Assume $A = B = \mathbb{R}$ and $x : A \times B \to \mathbb{R}$ is

$$x(a,b) = -\frac{1}{2}(a-b)^2$$

Optimal Constant Function

- Let $dp(a \mid b) = dp(a)$ (no information $\lambda = 0$).

Optimal Constant Function

- Let $dp(a \mid b) = dp(a)$ (no information $\lambda = 0$).
- Then $p_f \iff f(b) = \text{const}$ (i.e. $|f(B)| = 1$).

Optimal Constant Function

- Let $dp(a \mid b) = dp(a)$ (no information $\lambda = 0$).
- Then $p_f \iff f(b) = \text{const}$ (i.e. $|f(B)| = 1$).
- Optimal $\bar{a} = f(b)$ can be found from

$$\nabla_a \mathbb{E}_p\{x\} = \nabla_a \int_B x(a, b)\, dp(b) = 0$$

Optimal Constant Function

- Let $dp(a \mid b) = dp(a)$ (no information $\lambda = 0$).
- Then $p_f \iff f(b) = \text{const}$ (i.e. $|f(B)| = 1$).
- Optimal $\bar{a} = f(b)$ can be found from

$$\nabla_a \mathbb{E}_p\{x\} = \nabla_a \int_B x(a,b)\, dp(b) = 0$$

- For $x(a,b) = -\frac{1}{2}(a-b)^2$ this gives $\bar{a} = \mathbb{E}_p\{b\}$ and

$$\mathbb{E}_{p_f}\{x\} = -\frac{1}{2}\int_B (a-b)^2\, dp(b) \le -\frac{1}{2}\mathrm{Var}\{b\}$$

Optimal Constant Function

- Let $dp(a \mid b) = dp(a)$ (no information $\lambda = 0$).
- Then $p_f \iff f(b) = \text{const}$ (i.e. $|f(B)| = 1$).
- Optimal $\bar{a} = f(b)$ can be found from

$$\nabla_a \mathbb{E}_p\{x\} = \nabla_a \int_B x(a, b)\, dp(b) = 0$$

- For $x(a, b) = -\frac{1}{2}(a - b)^2$ this gives $\bar{a} = \mathbb{E}_p\{b\}$ and

$$\mathbb{E}_{p_f}\{x\} = -\frac{1}{2} \int_B (a - b)^2\, dp(b) \leq -\frac{1}{2}\mathrm{Var}\{b\}$$

- If $dp(b) = [\pi(b^2 + 1)]^{-1}\, db$ (Cauchy distribution), then

$$\mathbb{E}_{p_f}\{x\} \leq -\infty$$

Optimal Deterministic Kernel

- Let $dp(a \mid b) \neq dp(a)$ (i.e. information $\lambda > 0$).

Optimal Deterministic Kernel

- Let $dp(a \mid b) \neq dp(a)$ (i.e. information $\lambda > 0$).
- Then $p_f \iff f(b) = a$ non-constant (e.g. $|f(B)| = 2$).

Optimal Deterministic Kernel

- Let $dp(a \mid b) \neq dp(a)$ (i.e. information $\lambda > 0$).
- Then $p_f \iff f(b) = a$ non-constant (e.g. $|f(B)| = 2$).
- Optimal $\bar{a} = f(b)$ can be found by maximising

$$\mathbb{E}_{p_f}\{x\} = -\frac{1}{2} \sum_{i=1}^{|f(B)|} \int_{B_i} (\bar{a}_i - b)^2 \, dp(b)$$

Optimal Deterministic Kernel

- Let $dp(a \mid b) \neq dp(a)$ (i.e. information $\lambda > 0$).
- Then $p_f \iff f(b) = a$ non-constant (e.g. $|f(B)| = 2$).
- Optimal $\bar{a} = f(b)$ can be found by maximising

$$\mathbb{E}_{p_f}\{x\} = -\frac{1}{2} \sum_{i=1}^{|f(B)|} \int_{B_i} (\bar{a}_i - b)^2 \, dp(b)$$

- If $|f(B)| \in \mathbb{N}$ and $dp(b) = [\pi(b^2 + 1)]^{-1} \, db$ (Cauchy distribution), then

$$\mathbb{E}_{p_f}\{x\} \leq -\infty$$

Optimal Deterministic Kernel

- Let $dp(a \mid b) \neq dp(a)$ (i.e. information $\lambda > 0$).
- Then $p_f \iff f(b) = a$ non-constant (e.g. $|f(B)| = 2$).
- Optimal $\bar{a} = f(b)$ can be found by maximising

$$\mathbb{E}_{p_f}\{x\} = -\frac{1}{2}\sum_{i=1}^{|f(B)|}\int_{B_i}(\bar{a}_i - b)^2\,dp(b)$$

- If $|f(B)| \in \mathbb{N}$ and $dp(b) = [\pi(b^2+1)]^{-1}\,db$ (Cauchy distribution), then

$$\mathbb{E}_{p_f}\{x\} \leq -\infty$$

- If $|f(B)| = \infty$, then $a = f(b)$ communicates infinite information.

Optimal Non-Deterministic Kernel

- Take Shannon information $F_S(p) = I_{KL}(dp(a,b), dp(a)\, dp(b))$.

Optimal Non-Deterministic Kernel

- Take Shannon information $F_S(p) = I_{KL}(dp(a,b), dp(a)\, dp(b))$.
- Then $dp_\beta(a \mid b)$ is exponential, and for $x(a,b) = -\frac{1}{2}(a-b)^2$

$$dp_\beta(a \mid b) = \frac{e^{\beta\, x(a,b)}\, da}{\int_A e^{\beta\, x(a,b)}\, da} = \frac{1}{\sqrt{2\pi\beta^{-1}}}\, e^{-\frac{1}{2}\beta(a-b)^2}\, da$$

Optimal Non-Deterministic Kernel

- Take Shannon information $F_S(p) = I_{KL}(dp(a,b), dp(a)\,dp(b))$.
- Then $dp_\beta(a \mid b)$ is exponential, and for $x(a,b) = -\frac{1}{2}(a-b)^2$

$$dp_\beta(a \mid b) = \frac{e^{\beta\, x(a,b)}\, da}{\int_A e^{\beta\, x(a,b)}\, da} = \frac{1}{\sqrt{2\pi\beta^{-1}}}\, e^{-\frac{1}{2}\beta(a-b)^2}\, da$$

- The expected utility is

$$\begin{aligned}
\mathbb{E}_{p_\beta}\{x\} &= -\frac{1}{2}\frac{1}{\sqrt{2\pi\beta^{-1}}}\int_B dp(b)\int_{-\infty}^{\infty}(a-b)^2\, e^{-\frac{1}{2}\beta(a-b)^2}\, da \\
&= -\frac{1}{2}\frac{\sqrt{2\pi\beta^{-3}}}{\sqrt{2\pi\beta^{-1}}}\int_B dp(b) = -\frac{1}{2}\beta^{-1}
\end{aligned}$$

Optimal Non-Deterministic Kernel

- Take Shannon information $F_S(p) = I_{KL}(dp(a,b), dp(a)\, dp(b))$.
- Then $dp_\beta(a \mid b)$ is exponential, and for $x(a,b) = -\frac{1}{2}(a-b)^2$

$$dp_\beta(a \mid b) = \frac{e^{\beta\, x(a,b)}\, da}{\int_A e^{\beta\, x(a,b)}\, da} = \frac{1}{\sqrt{2\pi\beta^{-1}}}\, e^{-\frac{1}{2}\beta(a-b)^2}\, da$$

- The expected utility is

$$\begin{aligned}\mathbb{E}_{p_\beta}\{x\} &= -\frac{1}{2}\frac{1}{\sqrt{2\pi\beta^{-1}}}\int_B dp(b)\int_{-\infty}^{\infty}(a-b)^2\, e^{-\frac{1}{2}\beta(a-b)^2}\, da \\ &= -\frac{1}{2}\frac{\sqrt{2\pi\beta^{-3}}}{\sqrt{2\pi\beta^{-1}}}\int_B dp(b) = -\frac{1}{2}\beta^{-1}\end{aligned}$$

- Lagrange multiplier β^{-1} is related to $F_S(p) = \lambda$

$$\beta = 2\pi e^{1-2[H\{p(b)\}-\lambda]}$$

Optimal Non-Deterministic Kernel

- Take Shannon information $F_S(p) = I_{KL}(dp(a,b), dp(a)\, dp(b))$.
- Then $dp_\beta(a \mid b)$ is exponential, and for $x(a,b) = -\frac{1}{2}(a-b)^2$

$$dp_\beta(a \mid b) = \frac{e^{\beta\, x(a,b)}\, da}{\int_A e^{\beta\, x(a,b)}\, da} = \frac{1}{\sqrt{2\pi\beta^{-1}}}\, e^{-\frac{1}{2}\beta(a-b)^2}\, da$$

- The expected utility is

$$\begin{aligned} \mathbb{E}_{p_\beta}\{x\} &= -\frac{1}{2}\frac{1}{\sqrt{2\pi\beta^{-1}}}\int_B dp(b) \int_{-\infty}^{\infty} (a-b)^2\, e^{-\frac{1}{2}\beta(a-b)^2}\, da \\ &= -\frac{1}{2}\frac{\sqrt{2\pi\beta^{-3}}}{\sqrt{2\pi\beta^{-1}}}\int_B dp(b) = -\frac{1}{2}\beta^{-1} \end{aligned}$$

- Lagrange multiplier β^{-1} is related to $F_S(p) = \lambda$

$$\beta = 2\pi e^{1-2[H\{p(b)\}-\lambda]}$$

- $\mathbb{E}_{p_\beta}\{x\} > -\infty$ if $H\{p(b)\} - \lambda < \infty$ or $\beta > 0$.

Another Example

- If $A = B = \mathbb{N}$, then any information constraint $\lambda < \sup F$ on $a = f(b)$ implies $|f(B)| < \infty$.

Another Example

- If $A = B = \mathbb{N}$, then any information constraint $\lambda < \sup F$ on $a = f(b)$ implies $|f(B)| < \infty$.
- Take $x(a, b) = -\frac{1}{2}(a - b)^2$ and $p(b)$ in the form

$$p(b) = \frac{1}{b^3 \zeta(3)} = \frac{1}{b^3 \sum_{b \in \mathbb{N}} b^{-3}}$$

Another Example

- If $A = B = \mathbb{N}$, then any information constraint $\lambda < \sup F$ on $a = f(b)$ implies $|f(B)| < \infty$.
- Take $x(a, b) = -\frac{1}{2}(a - b)^2$ and $p(b)$ in the form

$$p(b) = \frac{1}{b^3 \zeta(3)} = \frac{1}{b^3 \sum_{b \in \mathbb{N}} b^{-3}}$$

- Then $\mathbb{E}_{p_f}\{x\} = -\infty$ for all $F(p_f) < \sup F$.

Conclusions

- Defined optimal measures in optimisation problems with constraint on an abstract information resource $F(p) \leq \lambda$.

Conclusions

- Defined optimal measures in optimisation problems with constraint on an abstract information resource $F(p) \leq \lambda$.
- There are families of mutually absolutely continuous optimal measures, if the dual F^* of information is strictly convex.

Conclusions

- Defined optimal measures in optimisation problems with constraint on an abstract information resource $F(p) \leq \lambda$.
- There are families of mutually absolutely continuous optimal measures, if the dual F^* of information is strictly convex.
- Strict convexity of F^* is natural in the context of optimisation.

Conclusions

- Defined optimal measures in optimisation problems with constraint on an abstract information resource $F(p) \leq \lambda$.
- There are families of mutually absolutely continuous optimal measures, if the dual F^* of information is strictly convex.
- Strict convexity of F^* is natural in the context of optimisation.
- Deterministic transition kernels are strictly sub-optimal if information is bounded.

Conclusions

- Defined optimal measures in optimisation problems with constraint on an abstract information resource $F(p) \leq \lambda$.
- There are families of mutually absolutely continuous optimal measures, if the dual F^* of information is strictly convex.
- Strict convexity of F^* is natural in the context of optimisation.
- Deterministic transition kernels are strictly sub-optimal if information is bounded.
- There exist examples when $\mathbb{E}_p\{x\} = -\infty$ for any deterministic kernel satisfying information constraint.

Conclusions

- Defined optimal measures in optimisation problems with constraint on an abstract information resource $F(p) \leq \lambda$.
- There are families of mutually absolutely continuous optimal measures, if the dual F^* of information is strictly convex.
- Strict convexity of F^* is natural in the context of optimisation.
- Deterministic transition kernels are strictly sub-optimal if information is bounded.
- There exist examples when $\mathbb{E}_p\{x\} = -\infty$ for any deterministic kernel satisfying information constraint.
- No contradiction with theory of optimal statistical decisions.

Chentsov, N. N. (1972). *Statistical decision rules and optimal inference*. Moscow, U.S.S.R.: Nauka. (In Russian, English translation: Providence, RI: AMS, 1982)

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, *106, 108*, 620–630, 171–190.

Kullback, S. (1959). *Information theory and statistics*. John Wiley and Sons.

Shannon, C. E. (1948, July and October). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423 and 623–656.

Stratonovich, R. L. (1965). On value of information. *Izvestiya of USSR Academy of Sciences, Technical Cybernetics*, *5*, 3–12. (In Russian)