Dynamics of Information and Optimal Control of Reproduction in Evolutionary Systems

Roman V. Belavkin

Middlesex University

February 16, 2011

Introduction

- Optimisation of dynamic information systems (e.g. strategic games, machine learning, evolutionary systems).

Introduction

- Optimisation of dynamic information systems (e.g. strategic games, machine learning, evolutionary systems).
- Three kinds of problems (Tsypkin, 1973):

Introduction

- Optimisation of dynamic information systems (e.g. strategic games, machine learning, evolutionary systems).
- Three kinds of problems (Tsypkin, 1973):

  Deterministic : maximise $f(\omega)$.

Introduction

- Optimisation of dynamic information systems (e.g. strategic games, machine learning, evolutionary systems).
- Three kinds of problems (Tsypkin, 1973):

  Deterministic : maximise $f(\omega)$.

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

Introduction

- Optimisation of dynamic information systems (e.g. strategic games, machine learning, evolutionary systems).
- Three kinds of problems (Tsypkin, 1973):

  Deterministic : maximise $f(\omega)$.

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

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

Introduction

- Optimisation of dynamic information systems (e.g. strategic games, machine learning, evolutionary systems).
- Three kinds of problems (Tsypkin, 1973):

  Deterministic : maximise $f(\omega)$.

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

  Adaptive : estimate $p(\omega)$ and maximise $\mathbb{E}_p\{f(\omega)\}$.
- Information as a bounded resource.

Introduction

- Optimisation of dynamic information systems (e.g. strategic games, machine learning, evolutionary systems).
- Three kinds of problems (Tsypkin, 1973):

  Deterministic : maximise $f(\omega)$.
  Stochastic : maximise $\mathbb{E}_p\{f(\omega)\} = \sum f(\omega)\,p(\omega)$.
  Adaptive : estimate $p(\omega)$ and maximise $\mathbb{E}_p\{f(\omega)\}$.

- Information as a bounded resource.
- Learning, evolving and adaptive systems:

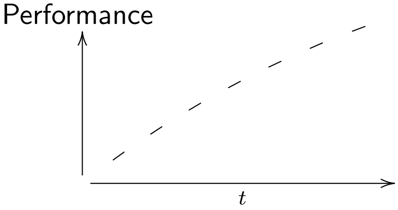
Introduction

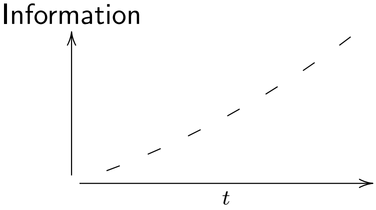
- Optimisation of dynamic information systems (e.g. strategic games, machine learning, evolutionary systems).
- Three kinds of problems (Tsypkin, 1973):

  Deterministic : maximise $f(\omega)$.

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

  Adaptive : estimate $p(\omega)$ and maximise $\mathbb{E}_p\{f(\omega)\}$.
- Information as a bounded resource.
- Learning, evolving and adaptive systems:





Representation in Paired Spaces

- Ω — a set of observable states (measurable, locally compact).

Representation in Paired Spaces

- Ω — a set of observable states (measurable, locally compact).
- 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 set of observable states (measurable, locally compact).
- 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 set of observable states (measurable, locally compact).
- 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 set of observable states (measurable, locally compact).
- 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 set of observable states (measurable, locally compact).
- 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* $\|y\|_1 = |\langle 1, y\rangle|$

Statistical Manifold

$$\mathcal{P}(\Omega) := \{y \in Y : y \geq 0, \|y\|_1 = 1\}$$

- Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$

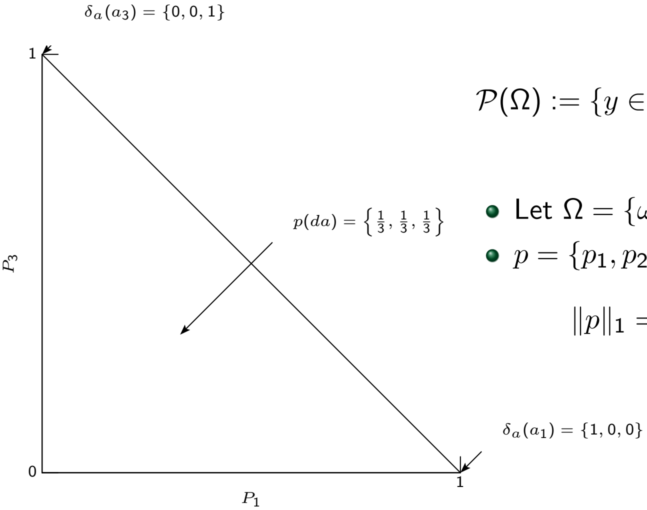
Statistical Manifold

$$\mathcal{P}(\Omega) := \{y \in Y : y \geq 0, \|y\|_1 = 1\}$$

- Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$
- $p = \{p_1, p_2, p_3\} \in \mathbb{R}^3$, $p_i \geq 0$:

$$\|p\|_1 = p_1 + p_2 + p_3 = 1$$

Statistical Manifold

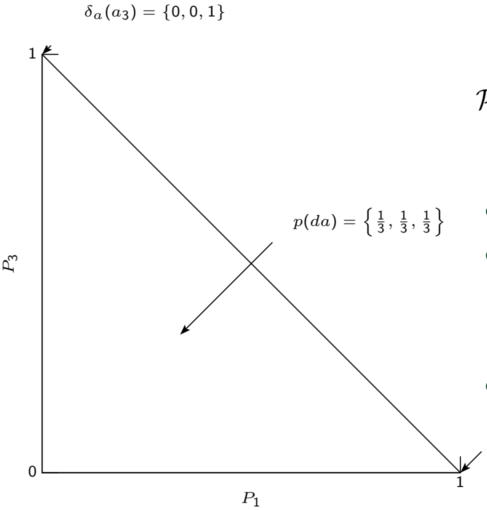


$$\mathcal{P}(\Omega) := \{y \in Y : y \geq 0, \|y\|_1 = 1\}$$

- Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$
- $p = \{p_1, p_2, p_3\} \in \mathbb{R}^3$, $p_i \geq 0$:

$$\|p\|_1 = p_1 + p_2 + p_3 = 1$$

Statistical Manifold



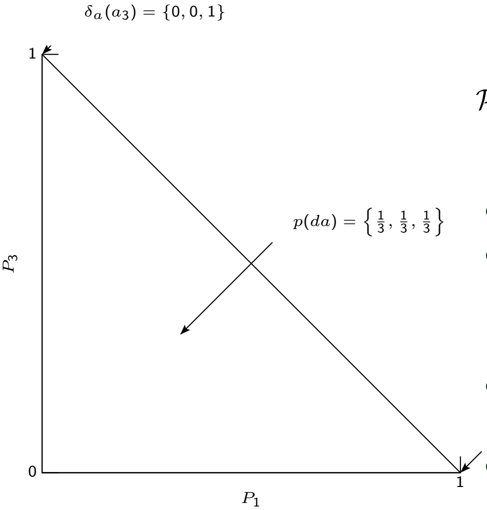
$$\mathcal{P}(\Omega) := \{y \in Y : y \geq 0, \|y\|_1 = 1\}$$

- Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$
- $p = \{p_1, p_2, p_3\} \in \mathbb{R}^3$, $p_i \geq 0$:

$$\|p\|_1 = p_1 + p_2 + p_3 = 1$$

- Vertexes of n-simplex are $m = n + 1$ Dirac δ-measures.

Statistical Manifold



$$\mathcal{P}(\Omega) := \{y \in Y : y \geq 0, \|y\|_1 = 1\}$$

- Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$
- $p = \{p_1, p_2, p_3\} \in \mathbb{R}^3$, $p_i \geq 0$:

$$\|p\|_1 = p_1 + p_2 + p_3 = 1$$

- Vertexes of n-simplex are $m = n + 1$ Dirac δ-measures.
- Choquet simplex if Ω is infinite.

Measures of Performance and Information

Expected utility

Let $(\Omega, \mathcal{R}, p)$ be a probability space, and $x : \Omega \to \mathbb{R}$ a utility. Then

$$\langle x, p \rangle = \mathbb{E}_p\{x\}$$

Measures of Performance and Information

Expected utility

Let $(\Omega, \mathcal{R}, p)$ be a probability space, and $x : \Omega \to \mathbb{R}$ a utility. Then

$$\langle x, p \rangle = \mathbb{E}_p\{x\}$$

Information resource

A closed functional $F : Y \to \mathbb{R} \cup \{\infty\}$ with $F(p)|_{\mathcal{P}} = I(p, q)$, where $I : \mathcal{P} \times \mathcal{P} \to \mathbb{R}_+ \cup \{\infty\}$ is an information distance. For example

$$F_{KL}(p) := I_{KL}(p, q) = \mathbb{E}_p\{\ln p - \ln q\}$$

Measures of Performance and Information

Expected utility

Let $(\Omega, \mathcal{R}, p)$ be a probability space, and $x : \Omega \to \mathbb{R}$ a utility. Then

$$\langle x, p \rangle = \mathbb{E}_p\{x\}$$

Information resource

A closed functional $F : Y \to \mathbb{R} \cup \{\infty\}$ with $F(p)|_{\mathcal{P}} = I(p, q)$, where $I : \mathcal{P} \times \mathcal{P} \to \mathbb{R}_+ \cup \{\infty\}$ is an information distance. For example

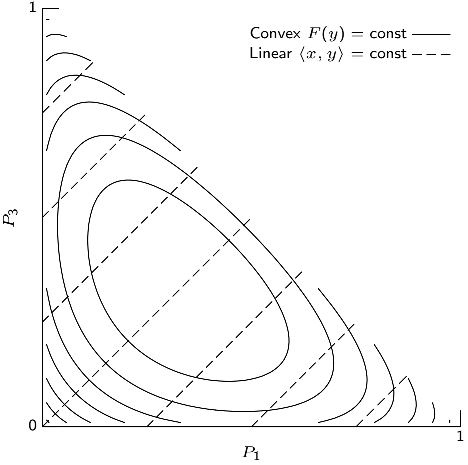
$$F_{KL}(p) := I_{KL}(p, q) = \mathbb{E}_p\{\ln p - \ln q\}$$

Remark

Other information distances are

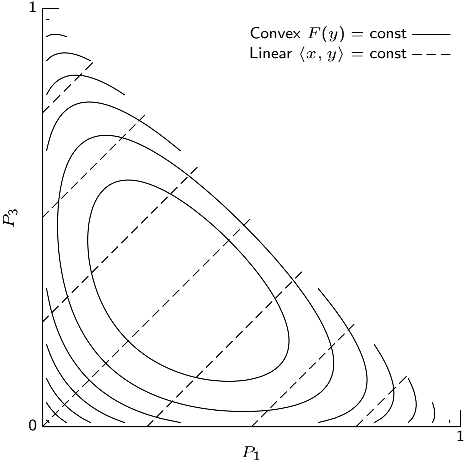
$$I_V(p, q) = \|p - q\||_1 \,, \quad I_F(p, q) = 2 \arccos\langle 1, p^{1/2} q^{1/2} \rangle$$

Information Dynamics



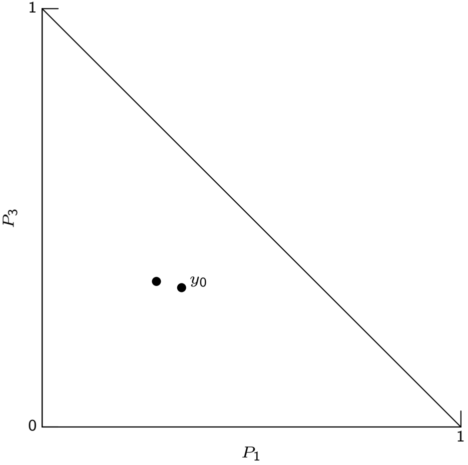
- $\mathbb{E}_p\{x\} = \langle x, p\rangle$ is linear.

Information Dynamics



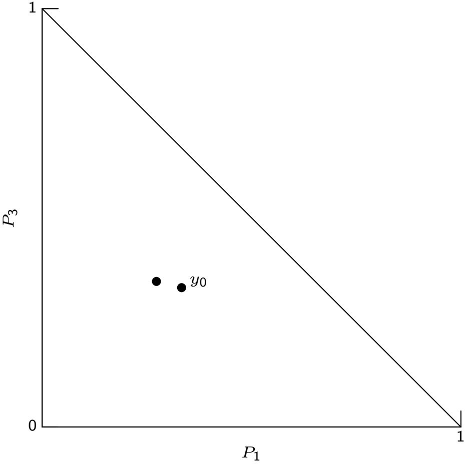
- $\mathbb{E}_p\{x\} = \langle x, p\rangle$ is linear.
- $\mathbb{E}_p\{\ln p - \ln q\}$ is convex.

Information Dynamics



- $\mathbb{E}_p\{x\} = \langle x, p\rangle$ is linear.
- $\mathbb{E}_p\{\ln p - \ln q\}$ is convex.
- $y \in \mathcal{P}$ are states and $\{y(t) \in \mathcal{P} : t \geq 0\}$ is an information trajectory.

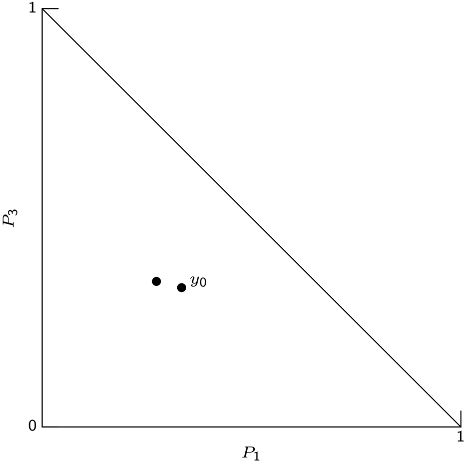
Information Dynamics



- $\mathbb{E}_p\{x\} = \langle x, p\rangle$ is linear.
- $\mathbb{E}_p\{\ln p - \ln q\}$ is convex.
- $y \in \mathcal{P}$ are states and $\{y(t) \in \mathcal{P} : t \geq 0\}$ is an information trajectory.
- Replicator dynamics $\dot{p} := dp(t)/dt$:

$$\dot{p} = [x - \mathbb{E}_p\{x\}]\, p$$

Information Dynamics



- $\mathbb{E}_p\{x\} = \langle x, p\rangle$ is linear.
- $\mathbb{E}_p\{\ln p - \ln q\}$ is convex.
- $y \in \mathcal{P}$ are states and $\{y(t) \in \mathcal{P} : t \geq 0\}$ is an information trajectory.
- Replicator dynamics $\dot{p} := dp(t)/dt$:

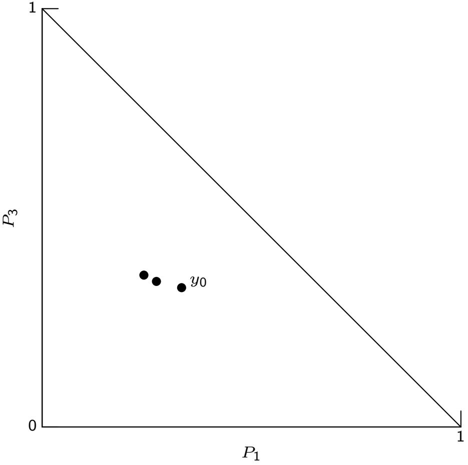
$$\dot{p} = [x - \mathbb{E}_p\{x\}]\, p$$

- Integral solution

$$p_t = e^{tx - \Psi_x(t)} p_0\,,\ p_0 = q$$

where $\Psi_x(t) := \ln\langle 1, e^{tx} p_0\rangle$.

Information Dynamics



- $\mathbb{E}_p\{x\} = \langle x, p\rangle$ is linear.
- $\mathbb{E}_p\{\ln p - \ln q\}$ is convex.
- $y \in \mathcal{P}$ are states and $\{y(t) \in \mathcal{P} : t \geq 0\}$ is an information trajectory.
- Replicator dynamics $\dot{p} := dp(t)/dt$:

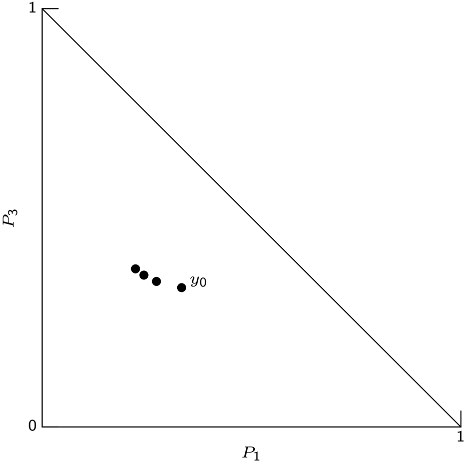
$$\dot{p} = [x - \mathbb{E}_p\{x\}]\, p$$

- Integral solution

$$p_t = e^{tx - \Psi_x(t)} p_0\,,\ \ p_0 = q$$

where $\Psi_x(t) := \ln\langle 1, e^{tx} p_0\rangle$.

Information Dynamics



- $\mathbb{E}_p\{x\} = \langle x, p\rangle$ is linear.
- $\mathbb{E}_p\{\ln p - \ln q\}$ is convex.
- $y \in \mathcal{P}$ are states and $\{y(t) \in \mathcal{P} : t \geq 0\}$ is an information trajectory.
- Replicator dynamics $\dot{p} := dp(t)/dt$:

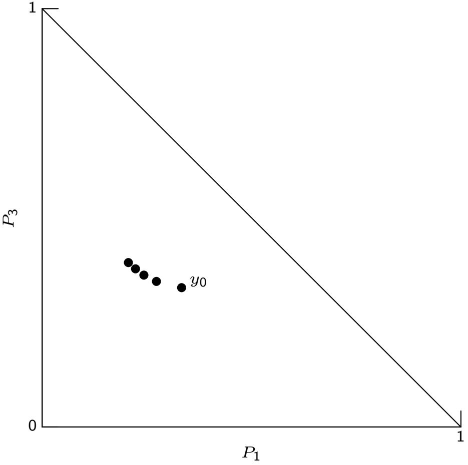
$$\dot{p} = [x - \mathbb{E}_p\{x\}]\, p$$

- Integral solution

$$p_t = e^{tx - \Psi_x(t)} p_0\,,\ \ p_0 = q$$

where $\Psi_x(t) := \ln\langle 1, e^{tx} p_0\rangle$.

Information Dynamics



- $\mathbb{E}_p\{x\} = \langle x, p\rangle$ is linear.
- $\mathbb{E}_p\{\ln p - \ln q\}$ is convex.
- $y \in \mathcal{P}$ are states and $\{y(t) \in \mathcal{P} : t \geq 0\}$ is an information trajectory.
- Replicator dynamics $\dot{p} := dp(t)/dt$:

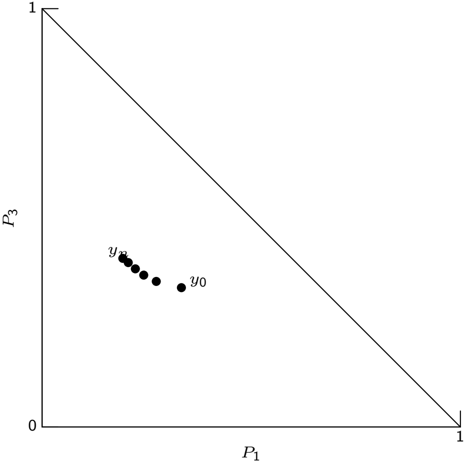
$$\dot{p} = [x - \mathbb{E}_p\{x\}]\, p$$

- Integral solution

$$p_t = e^{tx - \Psi_x(t)} p_0\,,\ p_0 = q$$

where $\Psi_x(t) := \ln\langle 1, e^{tx} p_0\rangle$.

Information Dynamics



- $\mathbb{E}_p\{x\} = \langle x, p\rangle$ is linear.
- $\mathbb{E}_p\{\ln p - \ln q\}$ is convex.
- $y \in \mathcal{P}$ are states and $\{y(t) \in \mathcal{P} : t \geq 0\}$ is an information trajectory.
- Replicator dynamics $\dot{p} := dp(t)/dt$:

$$\dot{p} = [x - \mathbb{E}_p\{x\}]\, p$$

- Integral solution

$$p_t = e^{tx - \Psi_x(t)} p_0\,,\ \ p_0 = q$$

where $\Psi_x(t) := \ln\langle 1, e^{tx} p_0\rangle$.

Utility of Information

- If $x \in X$ is fitness (utility), then the value of event p relative to q is

$$\langle x, p - q \rangle = \mathbb{E}_p\{x\} - \mathbb{E}_q\{x\}$$

Utility of Information

- If $x \in X$ is fitness (utility), then the value of event p relative to q is

$$\langle x, p - q \rangle = \mathbb{E}_p\{x\} - \mathbb{E}_q\{x\}$$

Definition (Utility of information)

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

Utility of Information

- If $x \in X$ is fitness (utility), then the value of event p relative to q is

$$\langle x, p - q \rangle = \mathbb{E}_p\{x\} - \mathbb{E}_q\{x\}$$

Definition (Utility of information)

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

- Stratonovich (1965) defined $U_x(I)$ for Shannon information (i.e. $F(y) = \mathbb{E}_p\{\ln p - \ln q\}$, where p is joint and q product of marginals).

Utility of Information

- If $x \in X$ is fitness (utility), then the value of event p relative to q is

$$\langle x, p - q \rangle = \mathbb{E}_p\{x\} - \mathbb{E}_q\{x\}$$

Definition (Utility of information)

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

- Stratonovich (1965) defined $U_x(I)$ for Shannon information (i.e. $F(y) = \mathbb{E}_p\{\ln p - \ln q\}$, where p is joint and q product of marginals).
- Related problem (inverse of $U_x(I)$):

$$I_x(U) := \inf\{F(y) : \langle x, y \rangle \geq U\}$$

Utility of Information

- If $x \in X$ is fitness (utility), then the value of event p relative to q is

$$\langle x, p - q \rangle = \mathbb{E}_p\{x\} - \mathbb{E}_q\{x\}$$

Definition (Utility of information)

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

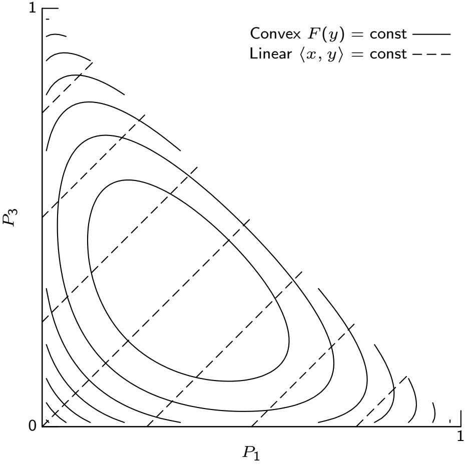
- Stratonovich (1965) defined $U_x(I)$ for Shannon information (i.e. $F(y) = \mathbb{E}_p\{\ln p - \ln q\}$, where p is joint and q product of marginals).
- Related problem (inverse of $U_x(I)$):

$$I_x(U) := \inf\{F(y) : \langle x, y \rangle \geq U\}$$

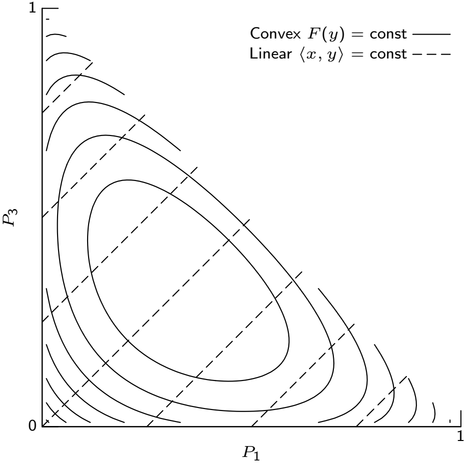
Question

Which transformations (controls) of the system lead to $U_x(I)$ on $\mathcal{P}(\Omega)$?

Optimal Information Dynamics



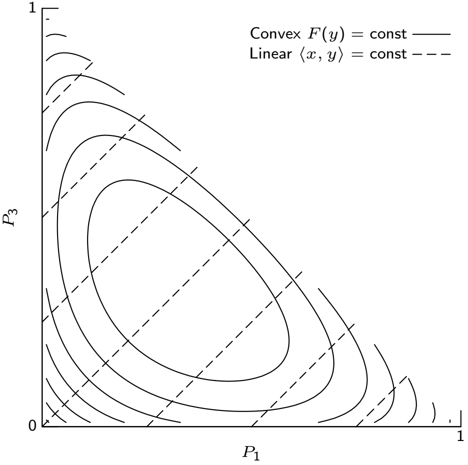
Optimal Information Dynamics



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

$$\begin{aligned} \partial\langle x, y_\beta\rangle &\in \beta^{-1}\partial F(y_\beta) \\ F(y_\beta) &= I \\ \beta^{-1} &\in \partial U_x(I) \end{aligned}$$

Optimal Information Dynamics

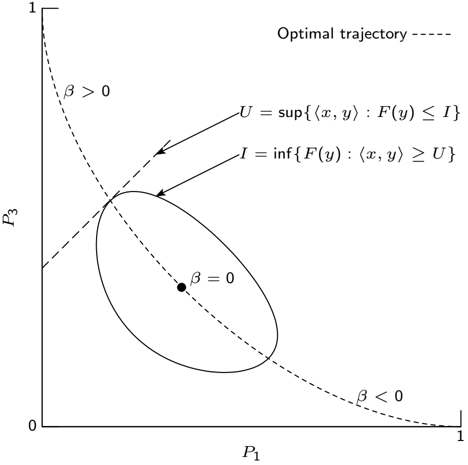


- The gradient of information must coincide with the gradient of the expected utility:

$$\begin{aligned} \partial\langle x, y_\beta\rangle &\in \beta^{-1}\partial F(y_\beta) \\ F(y_\beta) &= I \\ \beta^{-1} &\in \partial U_x(I) \end{aligned}$$

- Optimal solutions (Belavkin, 2010):

Optimal Information Dynamics



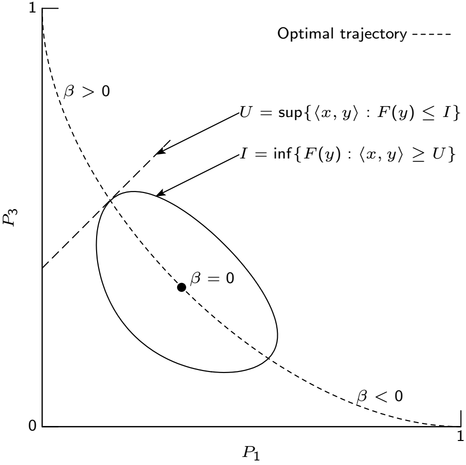
- The gradient of information must coincide with the gradient of the expected utility:

$$\begin{aligned} \partial\langle x, y_\beta\rangle &\in \beta^{-1}\partial F(y_\beta) \\ F(y_\beta) &= I \\ \beta^{-1} &\in \partial U_x(I) \end{aligned}$$

- Optimal solutions (Belavkin, 2010):

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

Optimal Information Dynamics



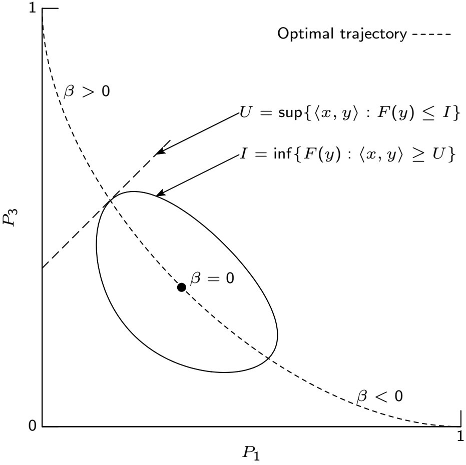
- The gradient of information must coincide with the gradient of the expected utility:

$$\begin{aligned} \partial\langle x, y_\beta\rangle &\in \beta^{-1}\partial F(y_\beta) \\ F(y_\beta) &= I \\ \beta^{-1} &\in \partial U_x(I) \end{aligned}$$

- Optimal solutions (Belavkin, 2010):

$$y_\beta = e^{\beta x}\, y_0$$

Optimal Information Dynamics



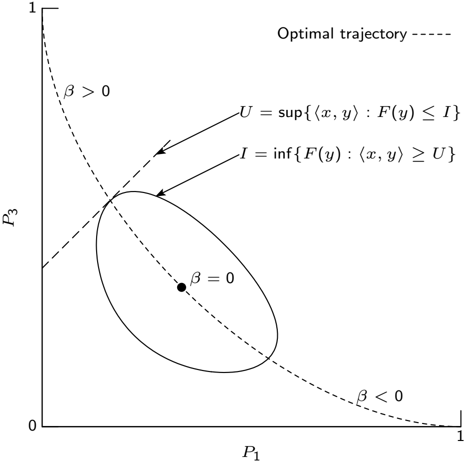
- The gradient of information must coincide with the gradient of the expected utility:

$$\begin{aligned} \partial\langle x, y_\beta\rangle &\in \beta^{-1}\partial F(y_\beta) \\ F(y_\beta) &= I \\ \beta^{-1} &\in \partial U_x(I) \end{aligned}$$

- Optimal solutions (Belavkin, 2010):

$$p_\beta = e^{\beta x - \Psi_x(\beta)}\, p_0$$

Optimal Information Dynamics



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

$$\begin{aligned} \partial\langle x, y_\beta\rangle &\in \beta^{-1}\partial F(y_\beta) \\ F(y_\beta) &= I \\ \beta^{-1} &\in \partial U_x(I) \end{aligned}$$

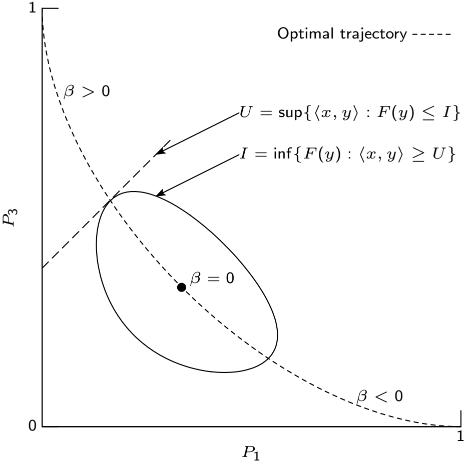
- Optimal solutions (Belavkin, 2010):

$$p_\beta = e^{\beta\, x - \Psi_x(\beta)}\, p_0$$

- Replicator dynamics in β:

$$\dot{p}_\beta = [x - \mathbb{E}_p\{x\}]\, p_\beta$$

Optimal Information Dynamics



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

$$\begin{aligned} \partial\langle x, y_\beta\rangle &\in \beta^{-1}\partial F(y_\beta) \\ F(y_\beta) &= I \\ \beta^{-1} &\in \partial U_x(I) \end{aligned}$$

- Optimal solutions (Belavkin, 2010):

$$p_\beta = e^{\beta\, x - \Psi_x(\beta)}\, p_0$$

- Replicator dynamics in β:

$$\dot{p}_\beta = [x - \mathbb{E}_p\{x\}]\, p_\beta$$

Question

Time or Information?

References122

Evolution as an Information Dynamic System

- EPSRC Sandpit '*Math of Life*' (July, 2009):

Evolution as an Information Dynamic System

- EPSRC Sandpit '*Math of Life*' (July, 2009):

- Three year project

Evolution as an Information Dynamic System

- EPSRC Sandpit '*Math of Life*' (July, 2009):

- Three year project

  Middlesex University : Roman Belavkin

Evolution as an Information Dynamic System

- EPSRC Sandpit '*Math of Life*' (July, 2009):

- Three year project

  Middlesex University : Roman Belavkin
  University of Warwick : John Aston

Evolution as an Information Dynamic System

- EPSRC Sandpit '*Math of Life*' (July, 2009):

- Three year project

  Middlesex University : Roman Belavkin
  University of Warwick : John Aston
  University of Keele : Alastair Channon

Evolution as an Information Dynamic System

- EPSRC Sandpit '*Math of Life*' (July, 2009):

- Three year project

  Middlesex University : Roman Belavkin
  University of Warwick : John Aston
  University of Keele : Alastair Channon
  University of Manchester : Chris Knight

Evolution as an Information Dynamic System

- EPSRC Sandpit '*Math of Life*' (July, 2009):

- Three year project

  Middlesex University : Roman Belavkin
  University of Warwick : John Aston
  University of Keele : Alastair Channon
  University of Manchester : Chris Knight

- Total value $\pounds 519,648$

Summary

- Evolution in biology occurs via death and birth of individuals.

Summary

- Evolution in biology occurs via death and birth of individuals.
- Reproduction involves selection, mutation and recombination.

Summary

- Evolution in biology occurs via death and birth of individuals.
- Reproduction involves selection, mutation and recombination.
- Is there a theory of optimal control of reproduction?

Summary

- Evolution in biology occurs via death and birth of individuals.
- Reproduction involves selection, mutation and recombination.
- Is there a theory of optimal control of reproduction?
- Meta-heuristics in combinatorial optimisation (Fogarty, 1989; Bäck, 1993; Vafaee et al., 2010).

Summary

- Evolution in biology occurs via death and birth of individuals.
- Reproduction involves selection, mutation and recombination.
- Is there a theory of optimal control of reproduction?
- Meta-heuristics in combinatorial optimisation (Fogarty, 1989; Bäck, 1993; Vafaee et al., 2010).
- Do biological organisms control reproduction?

Summary

- Evolution in biology occurs via death and birth of individuals.
- Reproduction involves selection, mutation and recombination.
- Is there a theory of optimal control of reproduction?
- Meta-heuristics in combinatorial optimisation (Fogarty, 1989; Bäck, 1993; Vafaee et al., 2010).
- Do biological organisms control reproduction?

Problem

To describe evolution and the dynamics of variables involved in genetic reproduction via information theoretic approach.

Summary

- Evolution in biology occurs via death and birth of individuals.
- Reproduction involves selection, mutation and recombination.
- Is there a theory of optimal control of reproduction?
- Meta-heuristics in combinatorial optimisation (Fogarty, 1989; Bäck, 1993; Vafaee et al., 2010).
- Do biological organisms control reproduction?

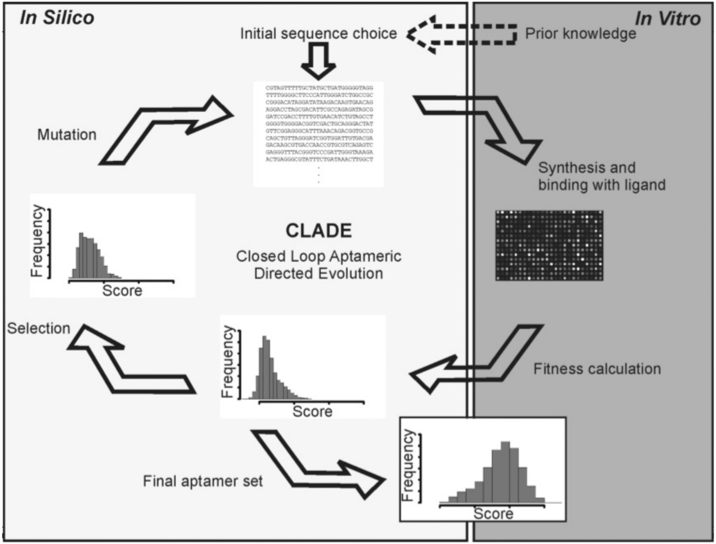
Problem

To describe evolution and the dynamics of variables involved in genetic reproduction via information theoretic approach.

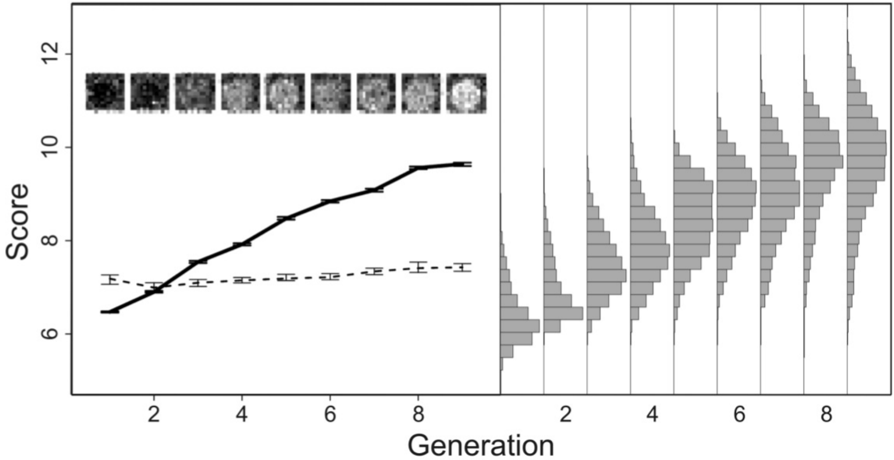
Problem

To use the theory for predictions and models, experiments for its verification, as well as inform and extend the theory

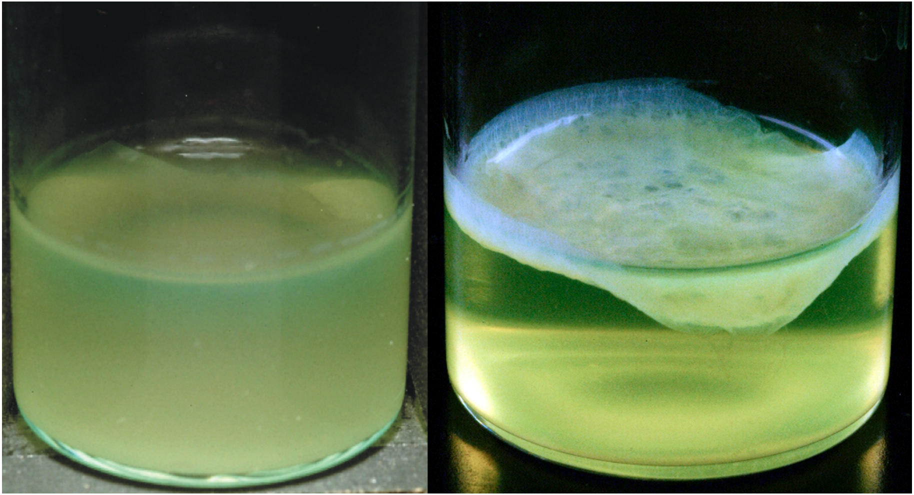
In Vitro, In Silico (C. Knight (2009))



Evolution *In Vitro*



Evolution *In Vivo*



Individual Organisms

Alphabet : $\alpha = \{G, A, T, C\}$ or $\{0, 1\}$

Individual Organisms

Alphabet : $\alpha = \{G, A, T, C\}$ or $\{0, 1\}$

Genotype : A sequence $\omega = `AATTCGC \ldots'$, $\omega \in \Omega := \alpha^n$.

Individual Organisms

Alphabet : $\alpha = \{G, A, T, C\}$ or $\{0, 1\}$

Genotype : A sequence $\omega =$ '$AATTCGC\ldots$', $\omega \in \Omega := \alpha^n$.

Metric : We can define $d : \Omega \times \Omega \to \mathbb{R}_+$, such as the Hamming metric: $d(a, b) := |\{i : a_i \neq b_i\}|$.

Individual Organisms

Alphabet : $\alpha = \{G, A, T, C\}$ or $\{0, 1\}$

Genotype : A sequence $\omega = `AATTCGC\ldots'$, $\omega \in \Omega := \alpha^n$.

Metric : We can define $d : \Omega \times \Omega \to \mathbb{R}_+$, such as the Hamming metric: $d(a, b) := |\{i : a_i \neq b_i\}|$.

Environment : imposes a preference relation (total pre-order) $(\Omega, \lesssim)$.

Individual Organisms

Alphabet : $\alpha = \{G, A, T, C\}$ or $\{0, 1\}$

Genotype : A sequence $\omega = `AATTCGC\ldots'$, $\omega \in \Omega := \alpha^n$.

Metric : We can define $d : \Omega \times \Omega \to \mathbb{R}_+$, such as the Hamming metric: $d(a, b) := |\{i : a_i \neq b_i\}|$.

Environment : imposes a preference relation (total pre-order) $(\Omega, \lesssim)$.

Fitness : is a pre-order embedding $f : (\Omega, \lesssim) \to (\mathbb{R}, \leq)$:

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

Individual Organisms

Alphabet : $\alpha = \{G, A, T, C\}$ or $\{0, 1\}$

Genotype : A sequence $\omega = `AATTCGC\ldots'$, $\omega \in \Omega := \alpha^n$.

Metric : We can define $d : \Omega \times \Omega \to \mathbb{R}_+$, such as the Hamming metric: $d(a, b) := |\{i : a_i \neq b_i\}|$.

Environment : imposes a preference relation (total pre-order) $(\Omega, \lesssim)$.

Fitness : is a pre-order embedding $f : (\Omega, \lesssim) \to (\mathbb{R}, \leq)$:

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

Definition (Monotonic landscape)

f is locally monotonic (isomorphic) relative to a metric d, if there exist $S(\top, l) := \{\omega : d(\top, \omega) < l\}$, $\top = \vee\Omega$ such that $\forall\, a, b \in S(\top, l)$:

$$-d(\top, a) \leq -d(\top, b) \quad \Longrightarrow (\iff) \quad f(a) \leq f(b)$$

Optimisation of Population Dynamics

Population of size n is a multiset of $\omega \in \Omega$.

Optimisation of Population Dynamics

Population of size n is a multiset of $\omega \in \Omega$.

Measure of each ω in a population is $y \in \mathbb{R}^{|\Omega|}$, $y \geq 0$.

Optimisation of Population Dynamics

Population of size n is a multiset of $\omega \in \Omega$.

Measure of each ω in a population is $y \in \mathbb{R}^{|\Omega|}$, $y \geq 0$.

Frequency is $p = y/\|y\|_1$ (i.e. $\|p\|_1 = 1$ and $p \geq 0$).

Optimisation of Population Dynamics

Population of size n is a multiset of $\omega \in \Omega$.

Measure of each ω in a population is $y \in \mathbb{R}^{|\Omega|}$, $y \geq 0$.

Frequency is $p = y/\|y\|_1$ (i.e. $\|p\|_1 = 1$ and $p \geq 0$).

Expected fitness of a population

$$\mathbb{E}_p\{f\} := \sum_{\omega \in \Omega} f(\omega)\, p(\omega)$$

Optimisation of Population Dynamics

Population of size n is a multiset of $\omega \in \Omega$.

Measure of each ω in a population is $y \in \mathbb{R}^{|\Omega|}$, $y \geq 0$.

Frequency is $p = y/\|y\|_1$ (i.e. $\|p\|_1 = 1$ and $p \geq 0$).

Expected fitness of a population

$$\mathbb{E}_p\{f\} := \sum_{\omega \in \Omega} f(\omega)\, p(\omega)$$

Dynamics of populations

$$\dot{p} = [f - \mathbb{E}_p\{f\}]\, p$$

Optimisation of Population Dynamics

Population of size n is a multiset of $\omega \in \Omega$.

Measure of each ω in a population is $y \in \mathbb{R}^{|\Omega|}$, $y \geq 0$.

Frequency is $p = y/\|y\|_1$ (i.e. $\|p\|_1 = 1$ and $p \geq 0$).

Expected fitness of a population

$$\mathbb{E}_p\{f\} := \sum_{\omega \in \Omega} f(\omega)\, p(\omega)$$

Dynamics of populations

$$\dot{p} = [f - \mathbb{E}_p\{f\}]\, p$$

Question

How to control parameters to achieve optimal evolution?

Control of Mutation

- $\Omega = \{\alpha\}^n$ is a metric space (e.g. with Hamming metric d).

Control of Mutation

- $\Omega = \{\alpha\}^n$ is a metric space (e.g. with Hamming metric d).
- Mutation of a by radius r is a transfer into a sphere:

$$S(a, r) := \{b \in \Omega : d(a, b) = r\}$$

Control of Mutation

- $\Omega = \{\alpha\}^n$ is a metric space (e.g. with Hamming metric d).
- Mutation of a by radius r is a transfer into a sphere:

$$S(a,r) := \{b \in \Omega : d(a,b) = r\}$$

- If θ is the mutation rate, then

$$P_\theta(b \in S(a,r)) = \binom{n}{r}\theta^r(1-\theta)^{n-r}$$

Control of Mutation

- $\Omega = \{\alpha\}^n$ is a metric space (e.g. with Hamming metric d).
- Mutation of a by radius r is a transfer into a sphere:

$$S(a, r) := \{b \in \Omega : d(a, b) = r\}$$

- If θ is the mutation rate, then

$$P_\theta(b \in S(a, r)) = \binom{n}{r}\theta^r(1-\theta)^{n-r}$$

- If $a \in S(\top, l)$, what is the probability that its offspring $b \in S(\top, k)$?

Control of Mutation

- $\Omega = \{\alpha\}^n$ is a metric space (e.g. with Hamming metric d).
- Mutation of a by radius r is a transfer into a sphere:

$$S(a,r) := \{b \in \Omega : d(a,b) = r\}$$

- If θ is the mutation rate, then

$$P_\theta(b \in S(a,r)) = \binom{n}{r}\theta^r(1-\theta)^{n-r}$$

- If $a \in S(\top, l)$, what is the probability that its offspring $b \in S(\top, k)$?
- Thus, $d(a,\top) = l$, $d(a,b) = r$, $d(b,\top) = k$:

$$P_\theta(b \in S(a,r) \cap S(\top,k) \mid a \in S(\top,l)) = |S(a,r) \cap S(\top,l)|_l \theta^r(1-\theta)^{n-r}$$

Control of Mutation

- $\Omega = \{\alpha\}^n$ is a metric space (e.g. with Hamming metric d).
- Mutation of a by radius r is a transfer into a sphere:

$$S(a,r) := \{b \in \Omega : d(a,b) = r\}$$

- If θ is the mutation rate, then

$$P_\theta(b \in S(a,r)) = \binom{n}{r}\theta^r(1-\theta)^{n-r}$$

- If $a \in S(\top, l)$, what is the probability that its offspring $b \in S(\top, k)$?
- Thus, $d(a,\top) = l$, $d(a,b) = r$, $d(b,\top) = k$:

$$P_\theta(b \in S(a,r) \cap S(\top,k) \mid a \in S(\top,l)) = |S(a,r) \cap S(\top,l)|_l \theta^r (1-\theta)^{n-r}$$

- where for $\alpha = 2$ (binary strings):

$$|S(a,r) \cap S(\top,k)|_l = \begin{cases} 0\,, & \text{if } l_+ \notin \mathbb{N} \cup \{0\} \\ \binom{n-l}{l_+}\binom{l}{l_-}\,, & \text{otherwise} \end{cases}$$

with $l_+ = \frac{1}{2}(r + k - l)$ and $l_- = \frac{1}{2}(l - |r-k|) = \min\{r,k\} - l_+$.

Analysis Using Absorbing Markov Chain

- We have the following Markov transition probability

$$P_\theta(k \mid l) := P_\theta(b \in S(\top, k) \mid a \in S(\top, l))$$

Analysis Using Absorbing Markov Chain

- We have the following Markov transition probability

$$P_\theta(k \mid l) := P_\theta(b \in S(\top, k) \mid a \in S(\top, l))$$

- It defines linear transformation $T_\theta : \mathcal{P} \to \mathcal{P}$ of $p_t = P(l)$:

$$p_{t+1} = T_\theta p_t = \sum_{l=0}^{n} P_\theta(k \mid l) P(l) \quad \text{or} \quad p_{t+m} = T_\theta^m p_t$$

Analysis Using Absorbing Markov Chain

- We have the following Markov transition probability

$$P_\theta(k \mid l) := P_\theta(b \in S(\top, k) \mid a \in S(\top, l))$$

- It defines linear transformation $T_\theta : \mathcal{P} \to \mathcal{P}$ of $p_t = P(l)$:

$$p_{t+1} = T_\theta p_t = \sum_{l=0}^{n} P_\theta(k \mid l) P(l) \quad \text{or} \quad p_{t+m} = T_\theta^m p_t$$

- The top element $\top \in \Omega$ can be assumed as an absorbing state.

Analysis Using Absorbing Markov Chain

- We have the following Markov transition probability

$$P_\theta(k \mid l) := P_\theta(b \in S(\top, k) \mid a \in S(\top, l))$$

- It defines linear transformation $T_\theta : \mathcal{P} \to \mathcal{P}$ of $p_t = P(l)$:

$$p_{t+1} = T_\theta p_t = \sum_{l=0}^{n} P_\theta(k \mid l) P(l) \quad \text{or} \quad p_{t+m} = T_\theta^m p_t$$

- The top element $\top \in \Omega$ can be assumed as an absorbing state.
- Using function $\theta = \theta(l)$, we can compute operator T_θ and its fundamental matrix

$$T_\theta = \begin{pmatrix} I & 0 \\ R & Q \end{pmatrix} \qquad N = (I - Q)^{-1}$$

Analysis Using Absorbing Markov Chain

- We have the following Markov transition probability

$$P_\theta(k \mid l) := P_\theta(b \in S(\top, k) \mid a \in S(\top, l))$$

- It defines linear transformation $T_\theta : \mathcal{P} \to \mathcal{P}$ of $p_t = P(l)$:

$$p_{t+1} = T_\theta p_t = \sum_{l=0}^{n} P_\theta(k \mid l) P(l) \quad \text{or} \quad p_{t+m} = T_\theta^m p_t$$

- The top element $\top \in \Omega$ can be assumed as an absorbing state.
- Using function $\theta = \theta(l)$, we can compute operator T_θ and its fundamental matrix

$$T_\theta = \begin{pmatrix} I & 0 \\ R & Q \end{pmatrix} \qquad N = (I - Q)^{-1}$$

- Given $p_0 = P(l)$, compute the expected time of convergence:

$$\mathbb{E}_P\{t\} = \sum_{l=1}^{n} t_l P(l) = \sum_{l=1}^{n} P(l) \sum_{k=1}^{n} n_{lk}$$

Optimisation of Mutation in One Step

- The expected fitness after transformation $a \mapsto b$ is

$$\mathbb{E}_\theta\{f\} = \sum_{b\in\Omega} f(b)\, P_\theta(b) = \sum_{b\in\Omega} f(b) \sum_{a\in\Omega} P_\theta(b \mid a) P(a)$$

Optimisation of Mutation in One Step

- The expected fitness after transformation $a \mapsto b$ is

$$\mathbb{E}_\theta\{f\} = \sum_{b\in\Omega} f(b)\, P_\theta(b) = \sum_{b\in\Omega} f(b) \sum_{a\in\Omega} P_\theta(b \mid a) P(a)$$

- The optimal mutation rate is given by condition

$$\frac{d\mathbb{E}_\theta\{f\}}{d\theta} = \sum_{b\in\Omega} f(b) \frac{d}{d\theta} P_\theta(b) = 0$$

Optimisation of Mutation in One Step

- The expected fitness after transformation $a \mapsto b$ is

$$\mathbb{E}_\theta\{f\} = \sum_{b\in\Omega} f(b)\, P_\theta(b) = \sum_{b\in\Omega} f(b) \sum_{a\in\Omega} P_\theta(b \mid a) P(a)$$

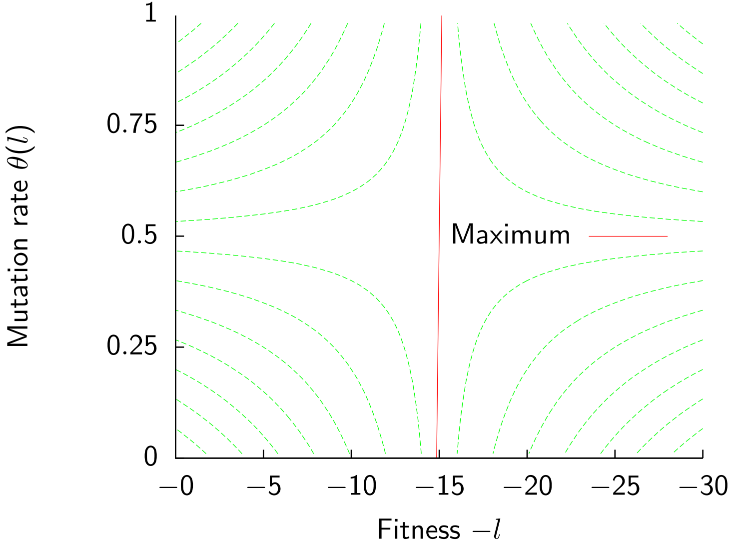
- The optimal mutation rate is given by condition

$$\frac{d\mathbb{E}_\theta\{f\}}{d\theta} = \sum_{b\in\Omega} f(b) \frac{d}{d\theta} P_\theta(b) = 0$$

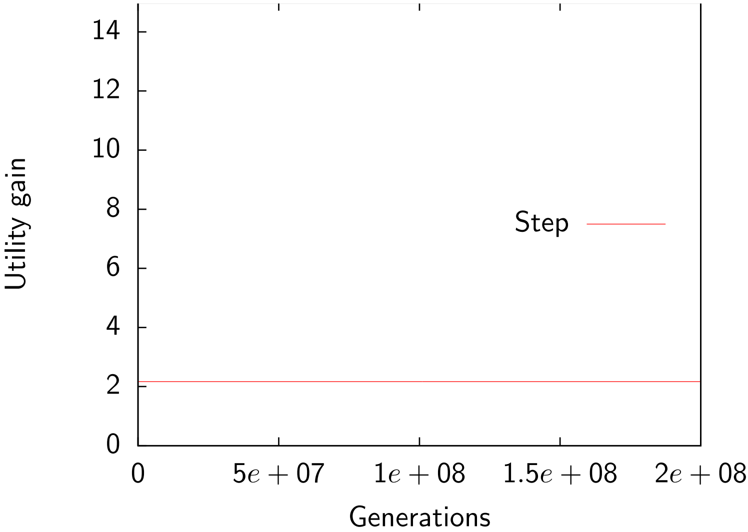
- If $f(b) = -d(\top, b)$ and $P(a) = \delta_a(\omega)$, then

$$\bar{\theta}(l) := \begin{cases} 0 & \text{if } l < n/2 \\ 1 & \text{if } l > n/2 \end{cases}$$

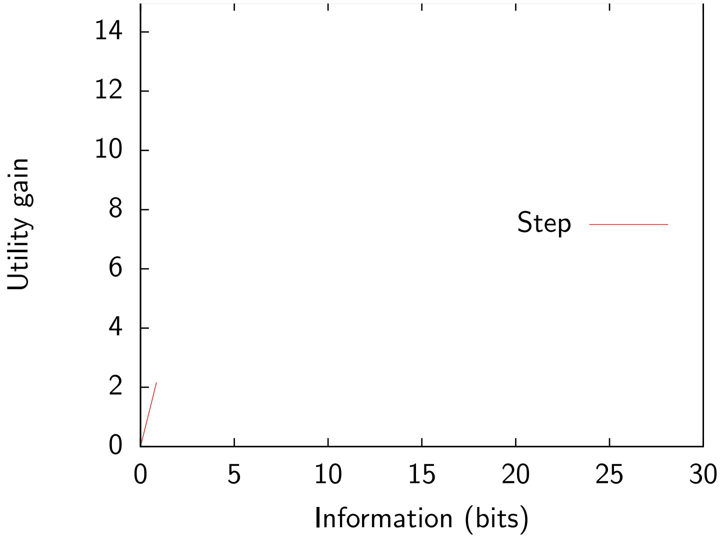
Step function



Step function



Step function



Maximising Fitness Growth

- Let $\mathbb{E}_0\{f\} = \langle f, p_0 \rangle$, and consider

$$\Delta\mathbb{E}_\theta\{f\} = \mathbb{E}_\theta\{f\} - \mathbb{E}_0\{f\}$$

Maximising Fitness Growth

- Let $\mathbb{E}_0\{f\} = \langle f, p_0 \rangle$, and consider

$$\Delta\mathbb{E}_\theta\{f\} = \mathbb{E}_\theta\{f\} - \mathbb{E}_0\{f\}$$

- Consider utility

$$u(\theta) := \begin{cases} 0 & \text{if } \Delta\mathbb{E}_\theta\{f\} \leq 0 \\ 1 & \text{if } \Delta\mathbb{E}_\theta\{f\} > 0 \end{cases}$$

Maximising Fitness Growth

- Let $\mathbb{E}_0\{f\} = \langle f, p_0 \rangle$, and consider

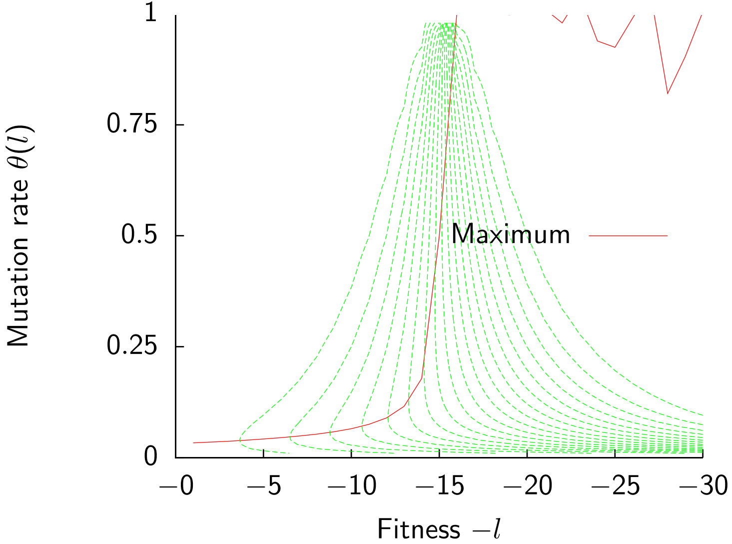
$$\Delta\mathbb{E}_\theta\{f\} = \mathbb{E}_\theta\{f\} - \mathbb{E}_0\{f\}$$

- Consider utility

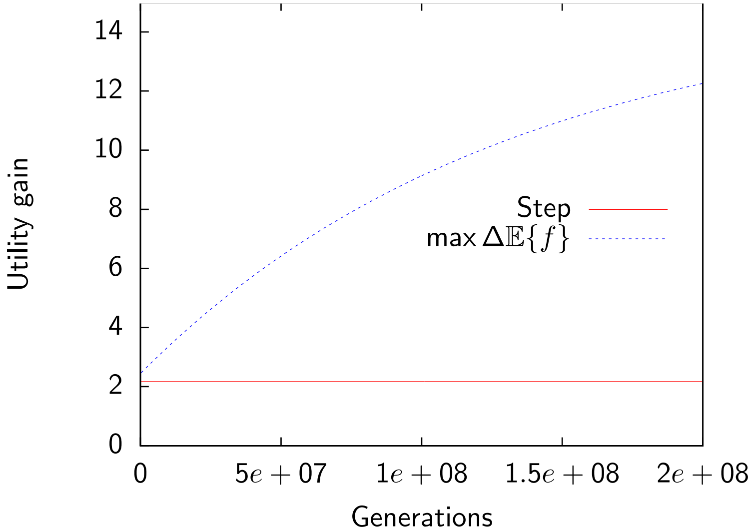
$$u(\theta) := \begin{cases} 0 & \text{if } \Delta\mathbb{E}_\theta\{f\} \leq 0 \\ 1 & \text{if } \Delta\mathbb{E}_\theta\{f\} > 0 \end{cases}$$

- The maximiser of $u(\theta)$ can be found numerically.

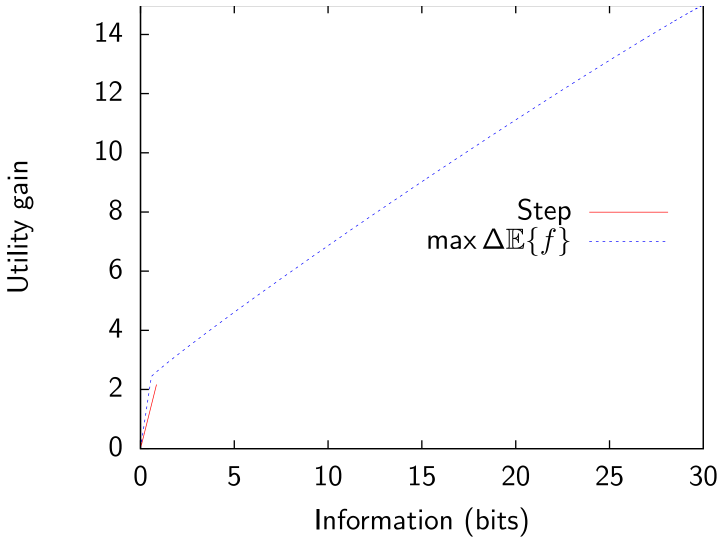
$\Delta\mathbb{E}\{f\}$ function



$\Delta\mathbb{E}\{f\}$ function



$\Delta\mathbb{E}\{f\}$ function



Needle in a Haystack

- Consider fitness

$$f(l) := \begin{cases} 1 & \text{if } l = 0 \\ 0 & \text{otherwise} \end{cases}$$

Needle in a Haystack

- Consider fitness

$$f(l) := \begin{cases} 1 & \text{if } l = 0 \\ 0 & \text{otherwise} \end{cases}$$

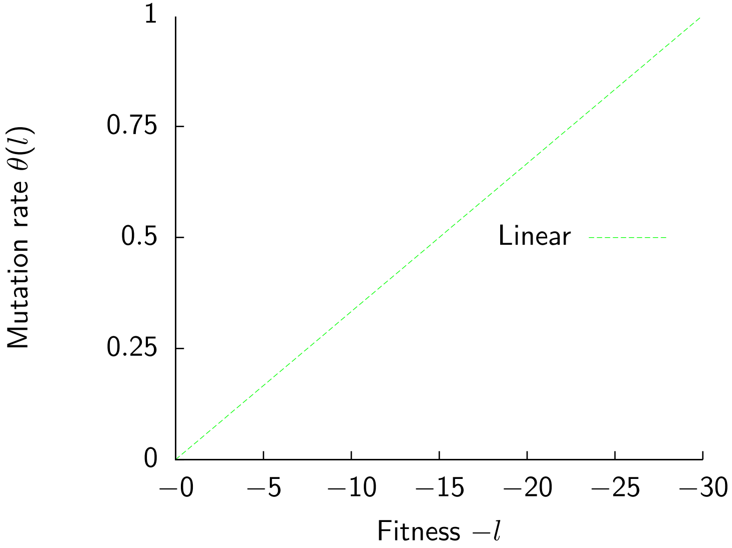
- Maximum probability of absorption

$$T_\theta = \begin{pmatrix} I & 0 \\ R & Q \end{pmatrix} \quad \text{where } r_{lk} = P_\theta(0 \mid l \neq 0) = \theta^l (1-\theta)^{n-l}$$

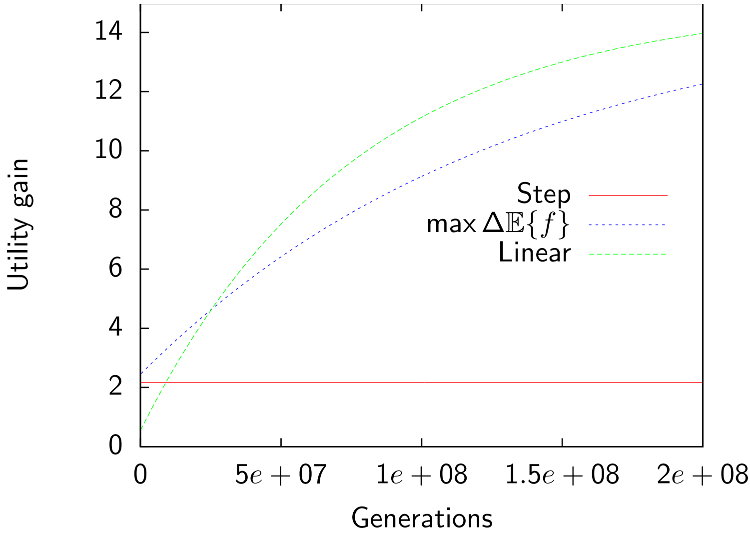
Needle in a Haystack

- Consider fitness
$$f(l) := \begin{cases} 1 & \text{if } l = 0 \\ 0 & \text{otherwise} \end{cases}$$
- Maximum probability of absorption
$$T_\theta = \begin{pmatrix} I & 0 \\ R & Q \end{pmatrix} \quad \text{where } r_{lk} = P_\theta(0 \mid l \neq 0) = \theta^l (1-\theta)^{n-l}$$
- Condition $dP_\theta(0 \mid l \neq 0)/d\theta = 0$ is satisfied for $l = n\theta$, which gives linear function:
$$\theta(l) = \frac{l}{n}$$

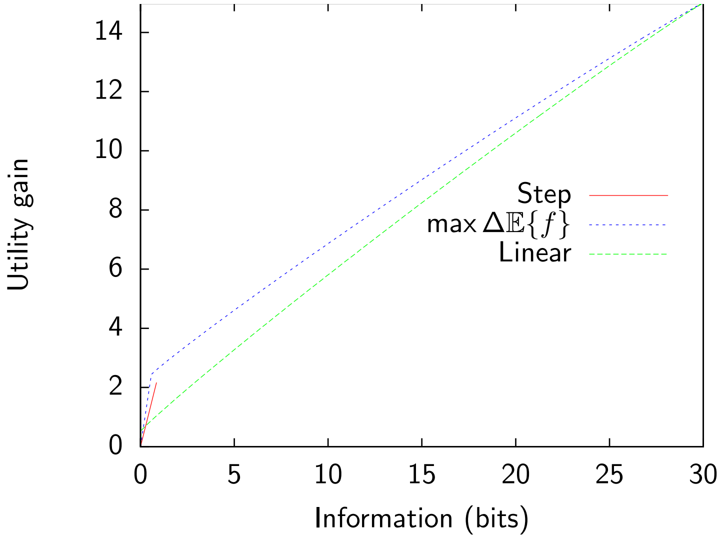
Linear function



Linear function



Linear function



Information Heuristics

- $f(\omega) = -l$ is not all information that can be used.

Information Heuristics

- $f(\omega) = -l$ is not all information that can be used.
- Given $q = P(l)$, random information is

$$I_{KL}(\delta_{kl}, q) = -\log P(l)$$

Information Heuristics

- $f(\omega) = -l$ is not all information that can be used.
- Given $q = P(l)$, random information is

$$I_{KL}(\delta_{kl}, q) = -\log P(l)$$

- The expected random information is entropy:

$$\mathbb{E}_q\{I_{KL}(\delta_{kl}, q)\} = -\sum_l [\log P(l)]\, P(l)$$

Information Heuristics

- $f(\omega) = -l$ is not all information that can be used.
- Given $q = P(l)$, random information is

$$I_{KL}(\delta_{kl}, q) = -\log P(l)$$

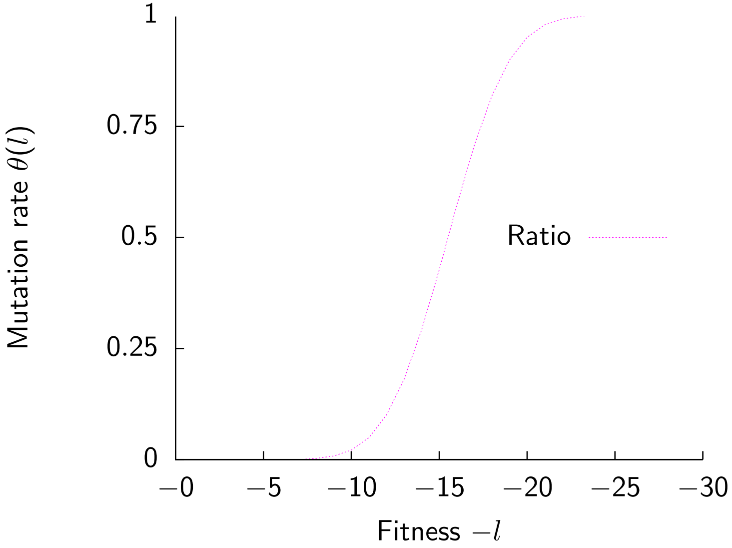
- The expected random information is entropy:

$$\mathbb{E}_q\{I_{KL}(\delta_{kl}, q)\} = -\sum_l [\log P(l)]\, P(l)$$

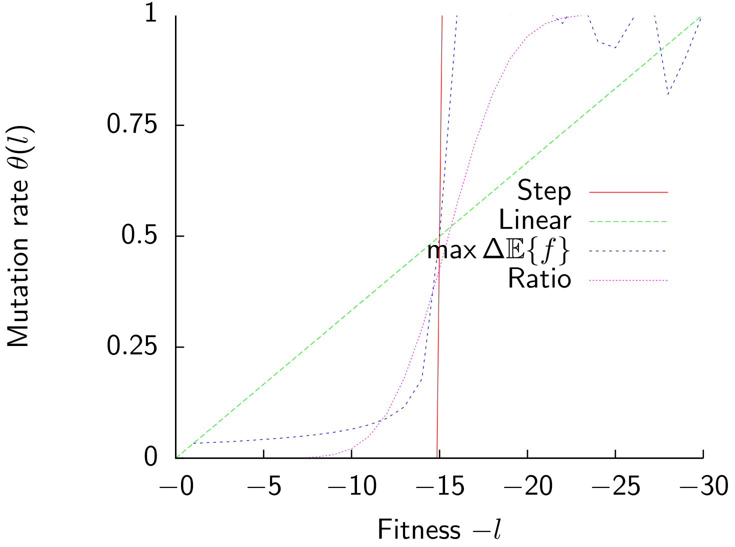
- The probability of 'improvement' is given by

$$P(k < l) = \sum_{k=0}^{l-1} P(k) = 2^{-n} \sum_{k=0}^{l-1} \binom{n}{k}$$

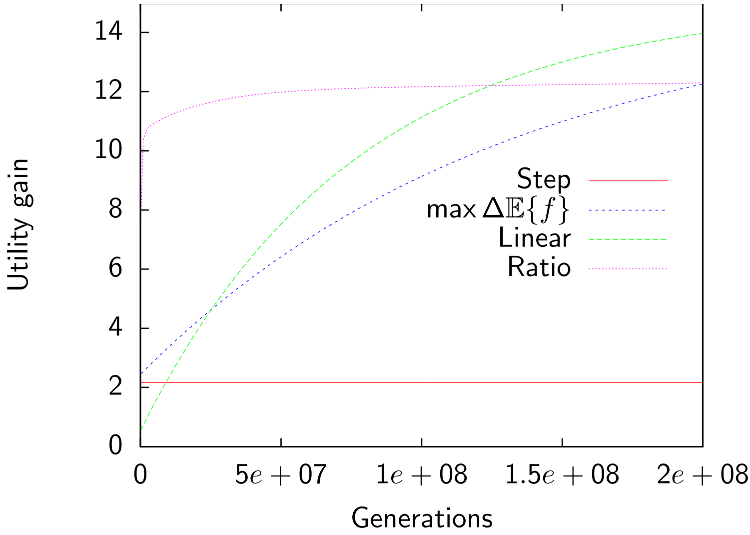
‘Informed’ Mutation function



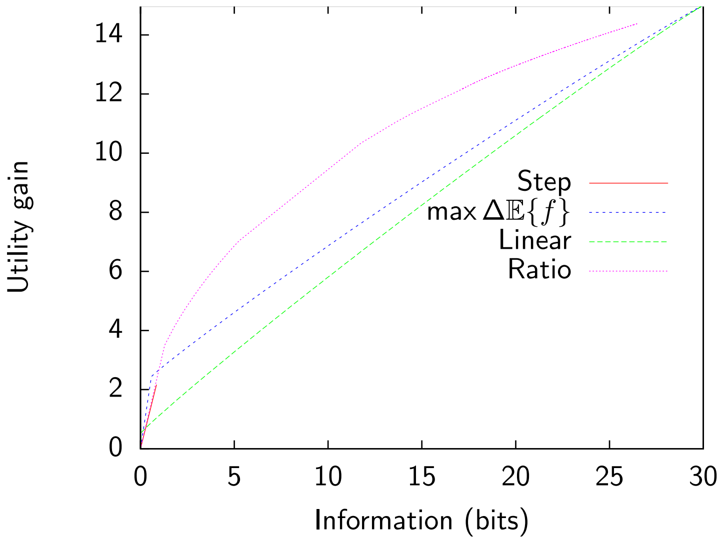
Comparing Mutation Rate Functions



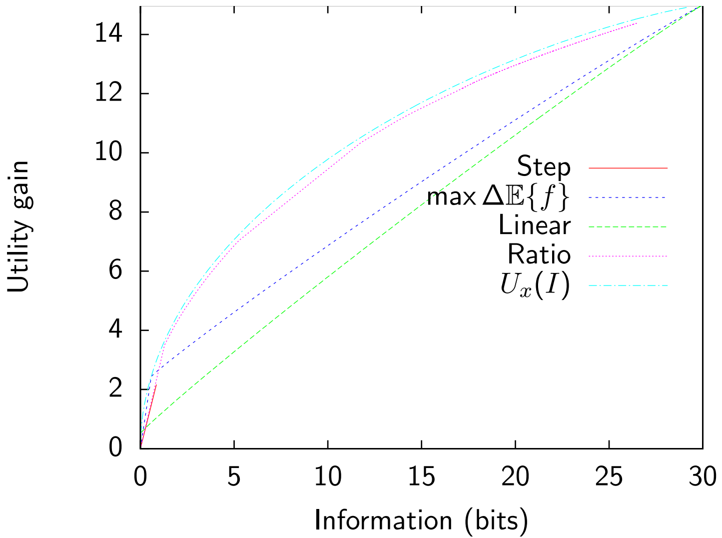
'Informed' Mutation function



‘Informed’ Mutation function



Theoretical $U_x(I)$



Conclusions and Questions

- Optimisation in time and information may not be the same.

Conclusions and Questions

- Optimisation in time and information may not be the same.
- Optimisation in information dynamics can give a different and potentially more effective methods (e.g. non-stationary, adaptive).

Conclusions and Questions

- Optimisation in time and information may not be the same.
- Optimisation in information dynamics can give a different and potentially more effective methods (e.g. non-stationary, adaptive).
- There is a problem of relating transformations of the system to information dynamics.

Bäck, T. (1993). Optimal mutation rates in genetic search. In S. Forrest (Ed.), *Proceedings of the 5th International Conference on Genetic Algorithms* (pp. 2–8). Morgan Kaufmann.

Belavkin, R. V. (2010). Information trajectory of optimal learning. In M. J. Hirsch, P. M. Pardalos, & R. Murphey (Eds.), *Dynamics of information systems: Theory and applications* (Vol. 40). Springer.

Fogarty, T. C. (1989). Varying the probability of mutation in the genetic algorithm. In J. D. Schaffer (Ed.), *Proceedings of the 3rd International Conference on Genetic Algorithms* (pp. 104–109). Morgan Kaufmann.

Knight, C. G., Platt, M., Rowe, W., Wedge, D. C., Khan, F., Day, P. J., et al. (2009). Array-based evolution of dna aptamers allows modelling of an explicit sequence-fitness landscape. *Nucleic Acids Res*, *37*(1), e6. (1362-4962 (Electronic) Journal Article Research Support, Non-U.S. Gov't)

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

Tsypkin, Y. Z. (1973). *Foundations of the theory of learning systems*.

Academic Press Inc.

Vafaee, F., Turán, G., & Nelson, P. C. (2010). Optimizing genetic operator rates using a markov chain model of genetic algorithms. In M. Pelikan & J. Branke (Eds.), (pp. 721–728). ACM.