Recent Developments in the Cross-Entropy Method

Dirk P. Kroese

Department of Mathematics, The University of Queensland, Australia

Contents

CE for Estimation and Optimization

The CE method can be used to solve the following types of problems:

1. Estimation:

$$\text{Estimate } \ell = \mathbb{E}[H(\mathbf{X})],$$

where $\mathbf{X}$ is a random vector/process taking values in some set $\mathscr{X}$ and H is function on $\mathscr{X}$.

In particular, the estimation of *rare event* probabilities:
$\ell = \mathbb{P}(S(\mathbf{X}) \geq \gamma)$, where S is another function on $\mathscr{X}$.

2. Optimisation:

$$\text{Determine } \max_{\boldsymbol{x} \in \mathscr{X}} \; S(\boldsymbol{x}),$$

where S is function on $\mathscr{X}$.

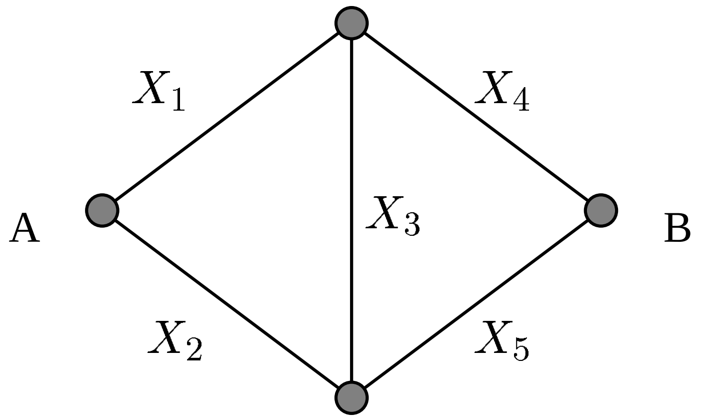
General Procedure

- Generate of a sample of random data (trajectories, vectors, etc.) according to a specified random mechanism.
- Update the parameters of the random mechanism, on the basis of the data, in order to produce a “better” sample in the next iteration.

SOME APPLICATIONS OF CE

Stochastic Shortest Path

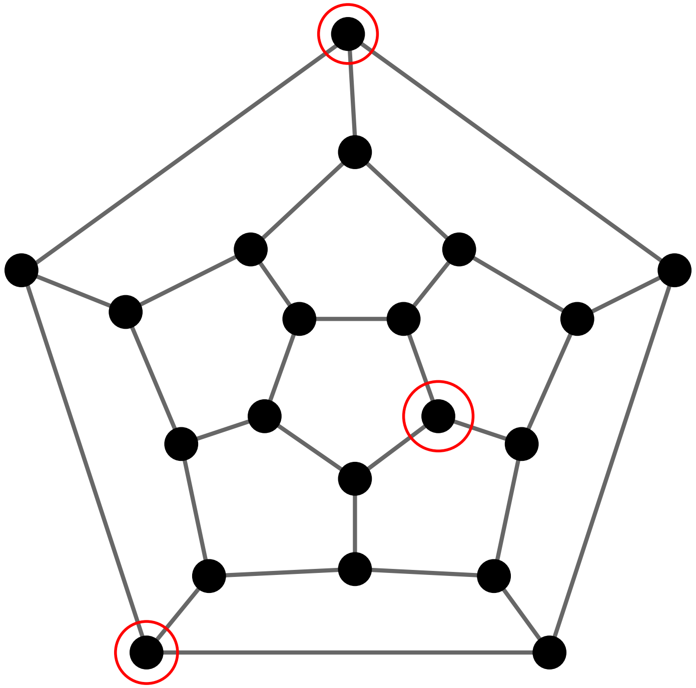
In this graph the random weights $X_1, \ldots, X_5$ are independent and exponentially distributed with means $0.25, 0.4, 0.1, 0.3, 0.2$.



Estimate the probability that the length of the shortest path from A to B is greater than or equal to 2

Network Reliability

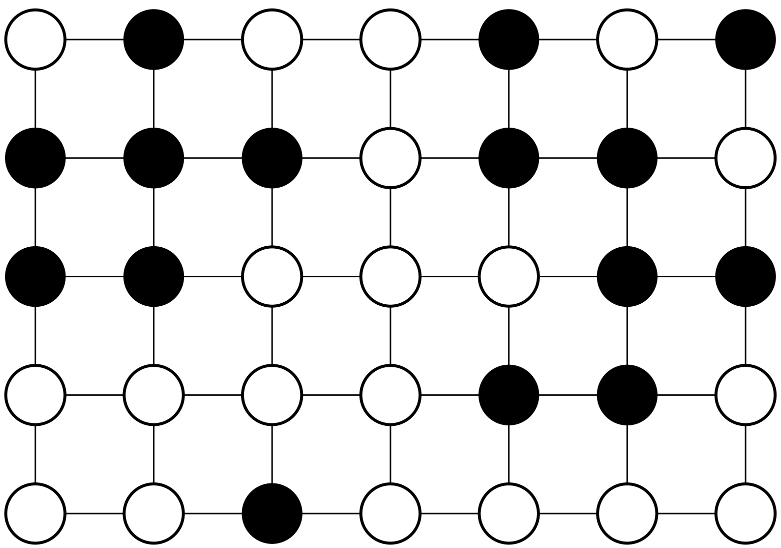
In this network all links have a probability of 0.01 of failing (independently).



Estimate the probability that the terminal nodes are not connected.

Ising Model

This is an example of a configuration $\mathbf{x}$ in the Ising model.



Energy of configuration:

$$H(\mathbf{x}) = \sum_{i<j} \psi_{ij}\, I_{\{x_i = x_j\}}$$

for some known $\{\psi_{ij}\}$.

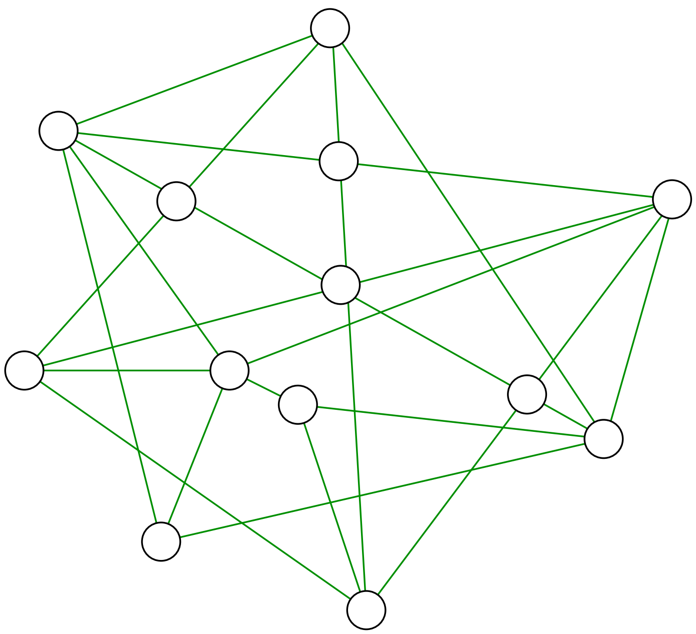
The probability of $\mathbf{x}$ occurring is

$$f(\mathbf{x}) = \mathrm{e}^{H(\mathbf{x})}/Z \ .$$

What is the normalisation constant (partition function) Z?

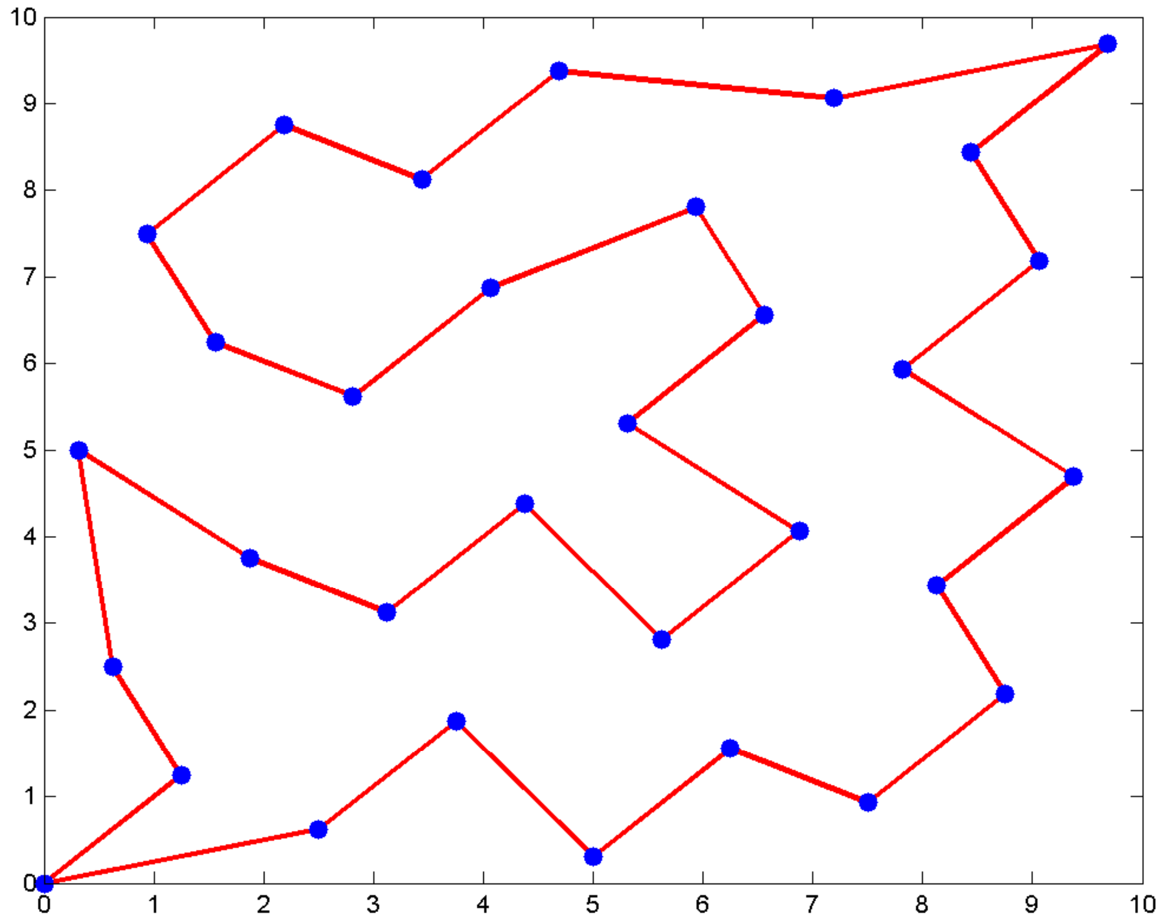
Max-cut Problem

We wish to colour the nodes white and black.



How should we colour the nodes so that the total number of links between the two groups is maximized?

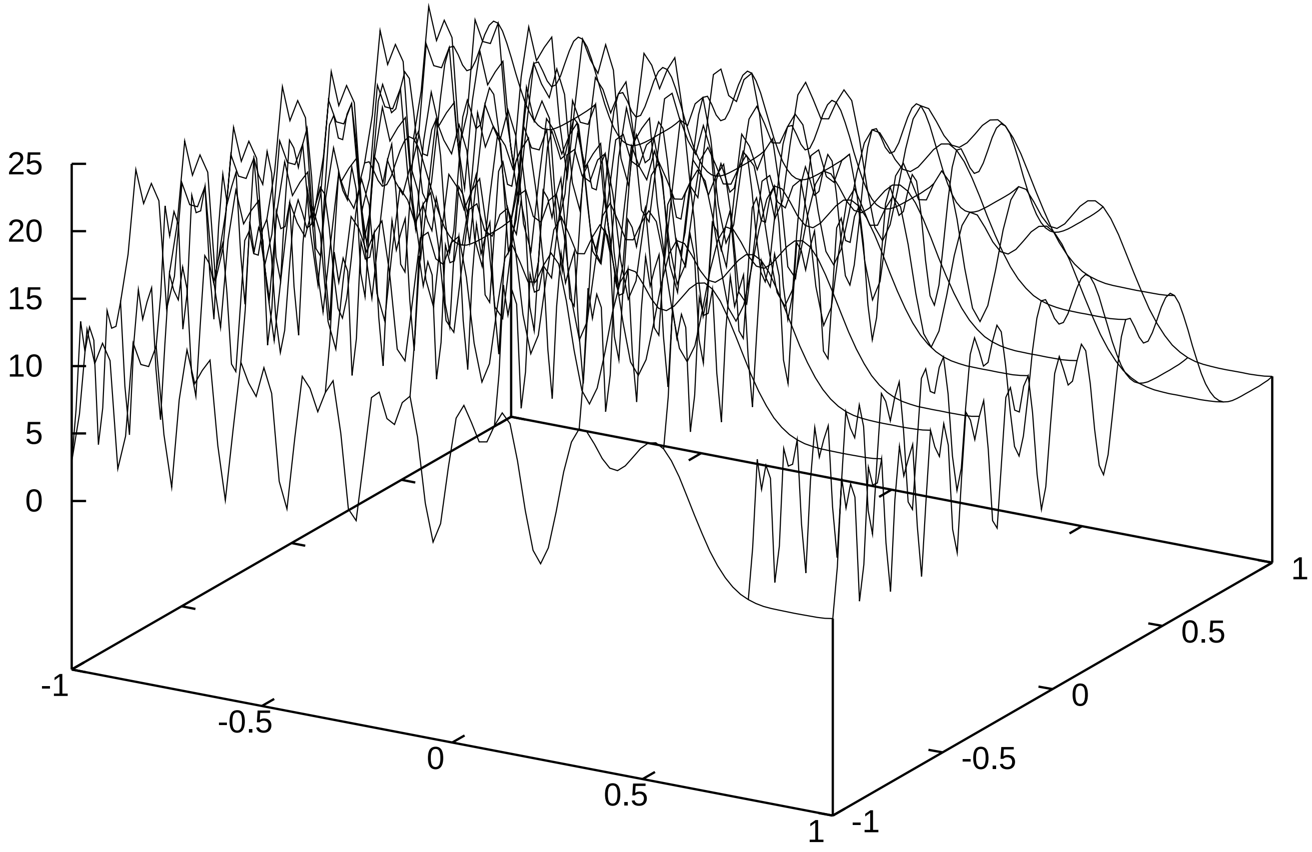
Traveling Salesman Problem



What is the shortest cycle through all the points?

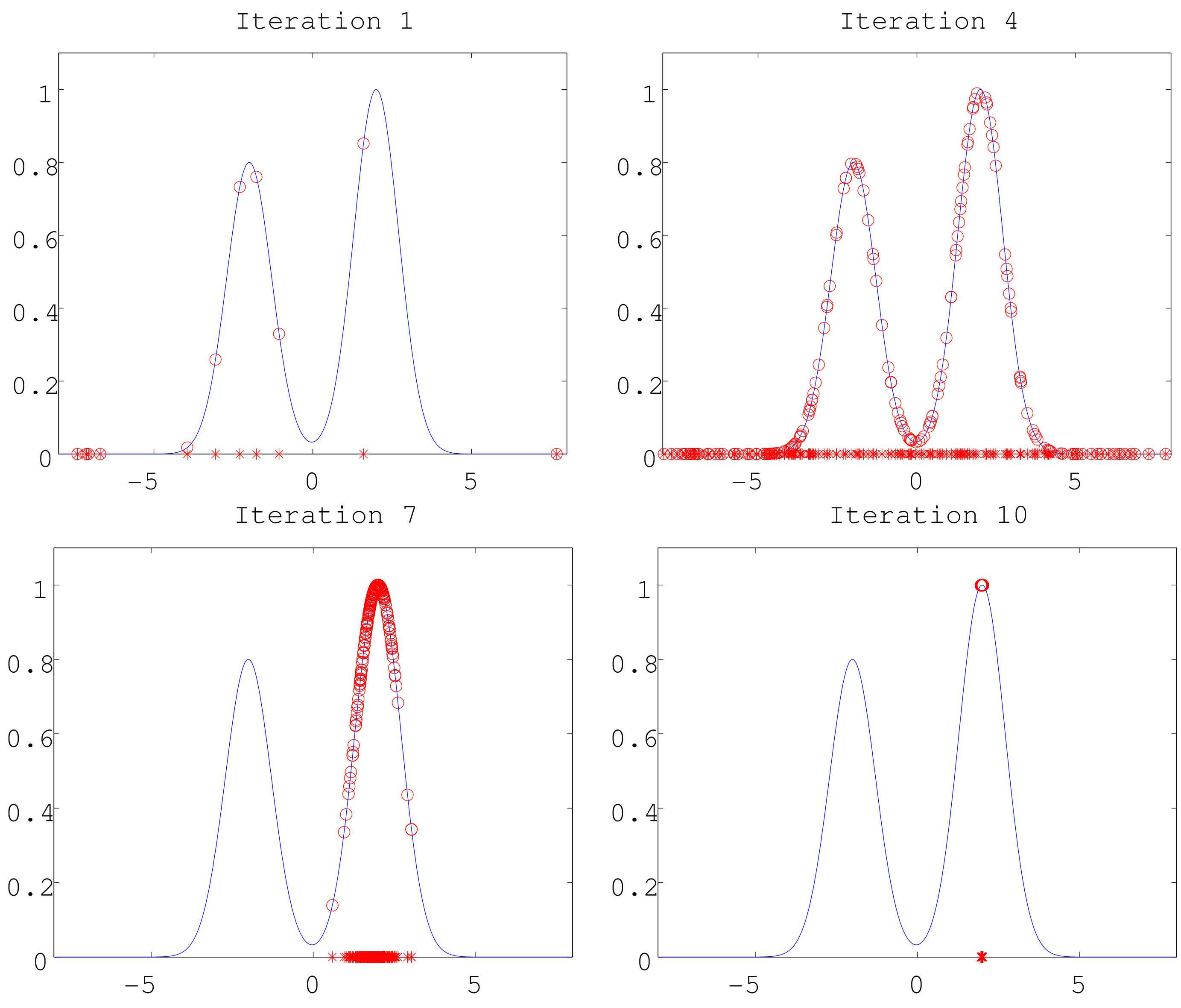
A Multi-extremal function

This is the trigonometric function.



What is its global maximum, and where is it attained?

Continuous Optimization (continued)



Sequence Alignment

A sequence alignment is an arrangement of two sequences $1, \ldots, n_1$ and $1, \ldots, n_2$ into two stacked rows, possibly including "spaces" (two opposite spaces not allowed).

$$\begin{cases} 1 & 2 & - & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & - & 2 & 3 & 4 & - & - & 5 & 6 & - & - \end{cases}$$

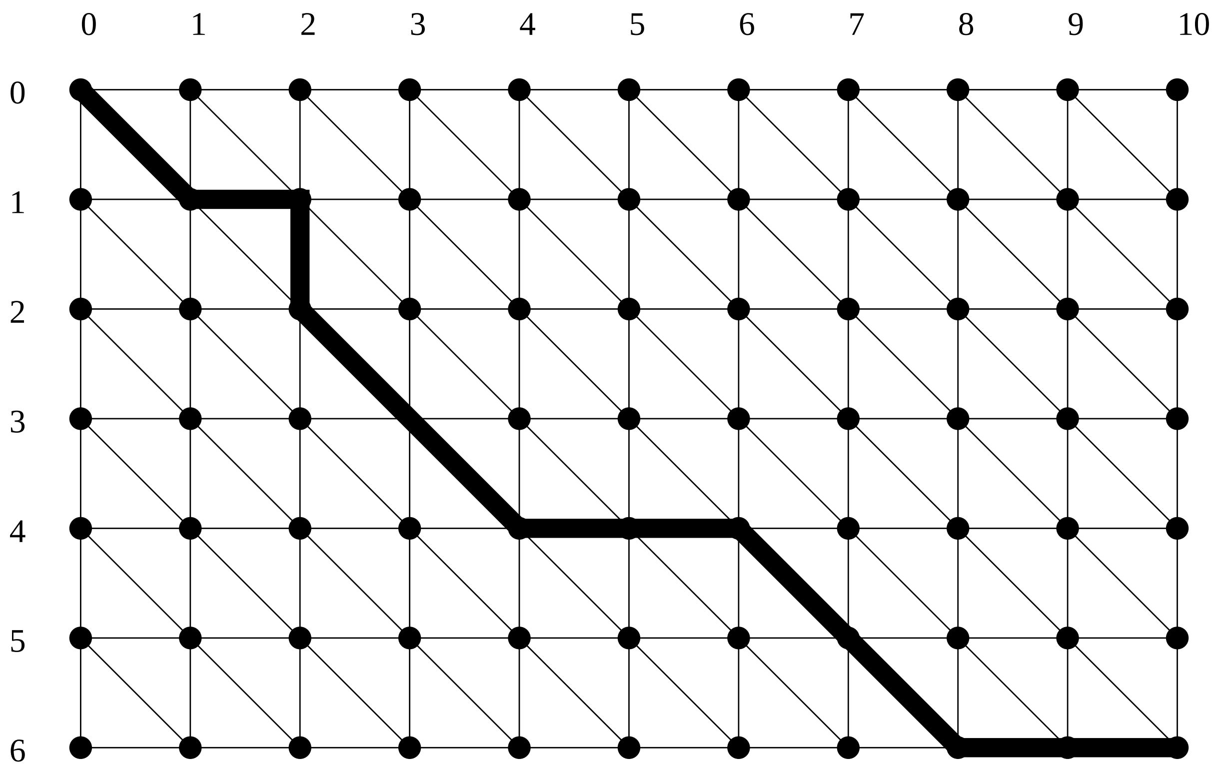
The two sequences of numbers could be associated with the positions of characters in a DNA or protein sequence, e.g.,

```
AGTGCAGATA        1 2 3 4 5 6 7 8 9 10
ACTG--GA--        1 2 3 4 - - 5 6 - -
```

What is the "best" alignment?

Alignment graph

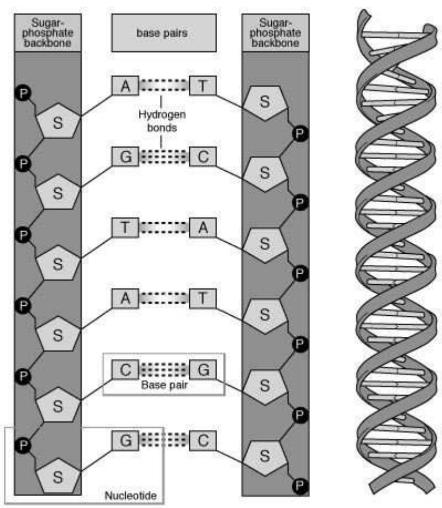
Each alignment can be characterised as a path through a directed graph.



Find the path with the smallest score.

GC Content

The GC content of a portion of DNA is the proportion of GC pairs that it contains.

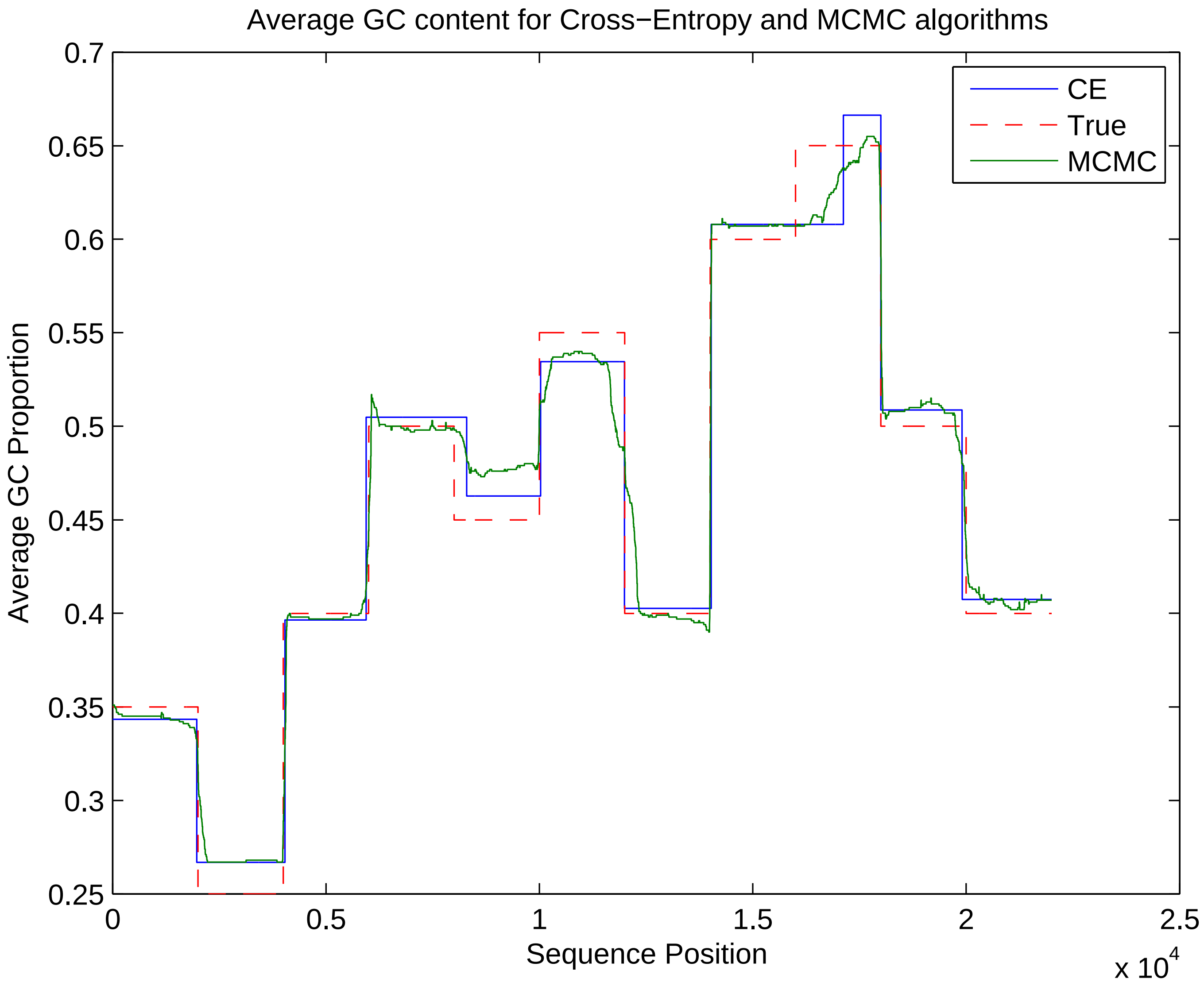


Sharp changes in GC content can be observed in the human and other genomes.

```
1                              25                             50
tgagatttatatagttgataaagcta ctccctacccatccccgcctcatctag
01010000000001001000001100 1011100111001111110100100
```

This can be viewed as a Bayesian multiple change point problem. The objective is to maximize the posterior pdf.

CE Solution



Average GC content for Cross−Entropy and MCMC algorithms
Average GC Proportion
Sequence Position
x 10^4
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0
0.5
1
1.5
2
2.5
CE
True
MCMC

Counting Representations

Let $A = \{1, 2, 3, 4, 5\}$, $A_1 = \{1, 2, 5\}$, $A_2 = \{1, 3, 4\}$, $A_3 = \{2, 3, 5\}$, $A_4 = \{1, 2, 3, 4\}$ and $A_5 = \{1, 3, 5\}$.

Corresponding matrix: $$\begin{pmatrix} 1 & \boxed{1} & 0 & 0 & 1 \\ 1 & 0 & 0 & \boxed{1} & 0 \\ 0 & 1 & 1 & 0 & \boxed{1} \\ 1 & 1 & \boxed{1} & 1 & 0 \\ \boxed{1} & 0 & 1 & 0 & 1 \end{pmatrix}.$$

A distinct representative is a vector $(x_1, \ldots, x_n)$ such that $x_i \in A_i$ for all i, and all are distinct. For example, $(2, 4, 5, 3, 1)$.

How many distinct representatives are there?

THEORETICAL DEVELOPMENT OF CE

Importance Sampling

Consider the estimation of

$$\ell = \mathbb{E}_f[H(\mathbf{X})] = \int H(\mathbf{x})\, f(\mathbf{x})\, d\mathbf{x}\ .$$

An important example of $H(\mathbf{X})$ is the indicator function

$$H(\mathbf{X}) = I_{\{S(\mathbf{x}) \geq \gamma\}} = \begin{cases} 1 & \text{if } S(\mathbf{X}) \geq \gamma \\ 0 & \text{otherwise}\ . \end{cases}$$

The likelihood ratio or importance sampling estimator of ℓ is

$$\hat{\ell} = \frac{1}{N} \sum_{i=1}^{N} H(\mathbf{X}_i)\, \frac{f(\mathbf{X}_i)}{g(\mathbf{X}_i)},$$

where $\mathbf{X}_1, \ldots, \mathbf{X}_N \sim g$.

Parametric VM Method

The best importance sampling density g (for $H \geq 0$) is

$$g^*(\mathbf{x}) := \frac{H(\mathbf{x})\, f(\mathbf{x})}{\ell} \; .$$

If instead we choose $g(\mathbf{x}) = f(\mathbf{x}; \mathbf{v})$ in the same parametric family as $f(\mathbf{x}) = f(\mathbf{x}; \mathbf{u})$, then the optimal parameter follows from the parametric variance minimization program:

$$\min_{\mathbf{v}} \mathbb{E}_{\mathbf{w}} \left[H^2(\mathbf{X}) W(\mathbf{X}; \mathbf{u}, \mathbf{v}) W(\mathbf{X}; \mathbf{u}, \mathbf{w}) \right] ,$$

which can be estimated via the stochastic counterpart approach.

Cross-Entropy (CE) Method

In the Cross-Entropy method we choose $g = f(\cdot; \mathbf{v})$ such that the **Kullback-Leibler** or **cross-entropy** distance between g^* and $f(\cdot; \mathbf{v})$ is minimal.

This is equivalent to solving

$$\max_{\mathbf{v}} \mathbb{E}_{\mathbf{w}} \left[H(\mathbf{X}) \, W(\mathbf{X}, \mathbf{w}, \mathbf{u}) \, \ln f(\mathbf{X}; \mathbf{v}) \right].$$

We may *estimate* the optimal solution $\mathbf{v}^*$ by solving the following stochastic counterpart:

$$\max_{\mathbf{v}} \frac{1}{N} \sum_{i=1}^{N} H(\mathbf{X}_i) \, W(\mathbf{X}_i; \mathbf{u}, \mathbf{w}) \, \ln f(\mathbf{X}_i; \mathbf{v}) \; ,$$

where $\mathbf{X}_1, \ldots, \mathbf{X}_N$ is a random sample from $f(\cdot; \mathbf{w})$.

Further Insights

- For rare-event estimation problems, where $H(\mathbf{X}) = I_{\{S(\mathbf{X}) \geq \gamma\}}$, the original CE and VM programs do not work (most indicators will be zero: $\hat{v} = 0/0$).
- Instead use a two-stage algorithm:
  1. Find γ_1 such that $\mathbb{P}_{\mathbf{u}}(S(\mathbf{X}) \geq \gamma_1) \geq \rho$, for rarity parameter ρ, which is typically chosen $0.1 \leq \rho \leq 0.01$.
  2. Determine the optimal $\mathbf{v}_1$ corresponding to γ_1.
  3. Iterate until $\gamma_T > \gamma$.
- By not specifying γ this procedure can also be used to maximise $S(\mathbf{x})$.

Challenges & Possible Remedies

- Degeneracy of the likelihood ratio, leads to poor estimates of $\mathbf{v}^*$ for high-dimensional problems. Remedies:
  - Screening, Parameter-free methods.
- For estimation problems the "shrinking" to a degenerate distribution is too rapid. Remedies:
  - Smoothing, injection.
- For counting problems the sampling distribution must be close to g^*. Parameterization does not always work. Remedies:
  - MCE, GCE, Parameter-free methods.

TWO RECENT APPLICATIONS

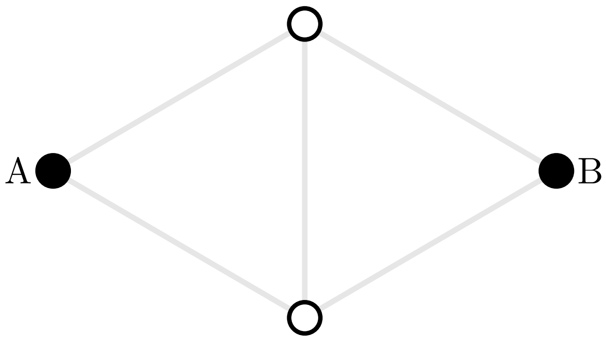
Network Planning

Problem Description:

Network Planning

Problem Description:

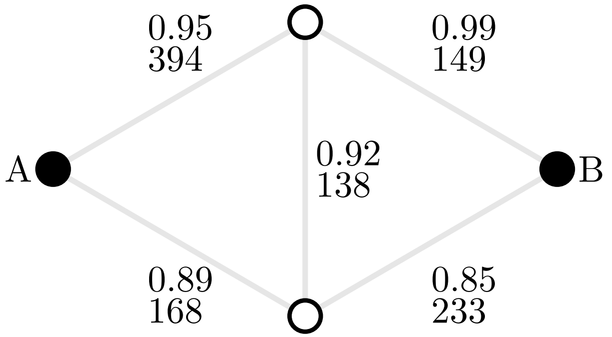
- We wish to purchase links $\in \{1, 2, \ldots\}$ to design a network, subject to a fixed budget.



Network Planning

Problem Description:

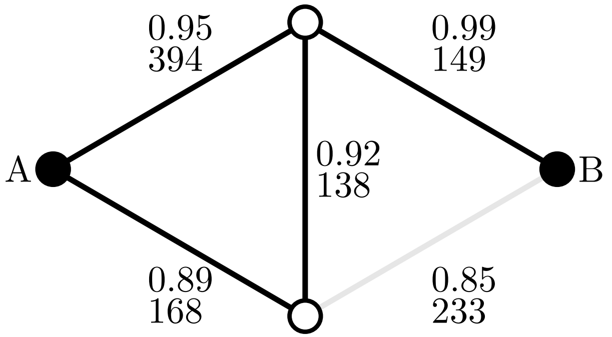
- We wish to purchase links $\in \{1, 2, \ldots\}$ to design a network, subject to a fixed budget.
- Each link i has a *cost* c_i and *reliability* p_i.



Network Planning

Problem Description:

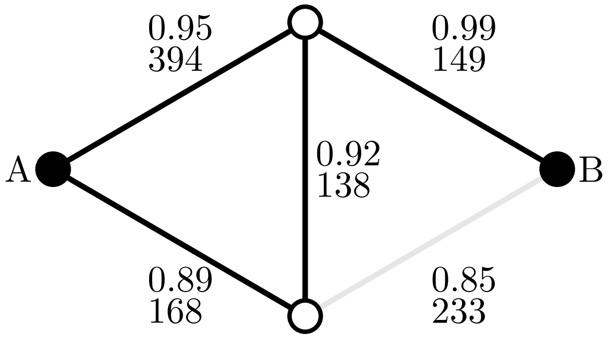
- We wish to purchase links $\in \{1, 2, \ldots\}$ to design a network, subject to a fixed budget.
- Each link i has a *cost* c_i and *reliability* p_i.
- Certain nodes called *terminal* nodes in the graph (i.e. in this case, nodes A and B) must be connected.



Network Planning

Problem Description:

- We wish to purchase links $\in \{1, 2, \ldots\}$ to design a network, subject to a fixed budget.
- Each link i has a *cost* c_i and *reliability* p_i.
- Certain nodes called *terminal* nodes in the graph (i.e. in this case, nodes A and B) must be connected.



Objective: Determine which links to purchase in order to maximize the system reliability, i.e., the probability that the terminal nodes are connected.

Mathematical Program

Let $x_i = 1$ if link i is purchased, and let $x_i = 0$ otherwise. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be the purchase vector, and let $r(\mathbf{x})$ be the reliability of the network. Define $\bar{r}(\mathbf{x}) = 1 - r(\mathbf{x})$ as the unreliability.

The objective is to Determine

$$\max_{\mathbf{x}} r(\mathbf{x}) \quad \text{or} \quad \min_{\mathbf{x}} \overline{r}(\mathbf{x})$$

subject to

$$\sum_{i=1}^{n} x_i c_i \leq C_{\max} \ ,$$

where $C_{\max}$ is the total budget.

Difficulties

- For large networks the exact calculation of the system (un)reliability is very difficult, and hence estimation of the network reliability via simulation becomes a viable option.
- However, for small link unreliabilities (typical) CMC becomes infeasible.
- We use permutation Monte Carlo to estimate the (small) unreliabilities instead.
- The network planning problem becomes a *simulation-based* (or *noisy*) optimisation problem.

Permutation Monte Carlo

- Observe a *dynamic* network – each link i has an exponential repair time with rate $\lambda_i = -\ln(1 - p_i)$.
- Assume that all links are failed at $t = 0$ and that all repair times are independent of each other.
- Let $\mathbf{Y}(t)$ be the state of links at time t. Then, $(\mathbf{Y}(t))$ is a Markov process with state space $\{0, 1\}^n$.
- Define Π as the *order* in which the links become operational.

Reliability Estimation: PMC

At $t = 1$, the probability that link e is operational is p_e. Therefore

$$r = \mathbb{E}[\varphi(\mathbf{Y}(1))]$$

where φ is the structure function.

Reliability Estimation: PMC

At $t = 1$, the probability that link e is operational is p_e.

By conditioning on Π,

$$r = \mathbb{E}[\mathbb{E}[\varphi(\mathbf{Y}(1)) \mid \Pi]]$$

Reliability Estimation: PMC

At $t = 1$, the probability that link e is operational is p_e.

By conditioning on Π,

$$r = \mathbb{E}[\underbrace{\mathbb{E}[\varphi(\mathbf{Y}(1)) \mid \Pi]}_{G(\Pi)}]$$

Reliability Estimation: PMC

At $t = 1$, the probability that link e is operational is p_e.

By conditioning on Π,

$$r = \mathbb{E}[\underbrace{\mathbb{E}[\varphi(\mathbf{Y}(1)) \mid \Pi]}_{G(\Pi)}]$$

The network reliability can be rewritten as

$$r = \mathbb{E}[G(\Pi)].$$

Reliability Estimation: PMC

At $t = 1$, the probability that link e is operational is p_e.

By conditioning on Π,

$$r = \mathbb{E}[\underbrace{\mathbb{E}[\varphi(\mathbf{Y}(1)) \mid \Pi]}_{G(\Pi)}]$$

The network reliability can be rewritten as

$$r = \mathbb{E}[G(\Pi)].$$

It is computed using the convolution.

- indirect approach - transform technique.
- direct approach - matrix exponentiation.

Reliability Estimation: PMC

At $t = 1$, the probability that link e is operational is p_e.

By conditioning on Π,

$$r = \mathbb{E}[\underbrace{\mathbb{E}[\varphi(\mathbf{Y}(1)) \mid \Pi]}_{G(\Pi)}]$$

The network reliability can be rewritten as

$$r = \mathbb{E}[G(\Pi)].$$

An unbiased estimator $\widehat{r}$ of r is computed via

$$\widehat{r} = \frac{1}{N} \sum_{i=1}^{N} G(\Pi_{(i)})$$

where $\Pi_{(1)}, \ldots, \Pi_{(N)}$ are independent identically distributed random permutations.

CE Approach

The CE method consists of two steps which are iterated:

- generate random purchase vectors $\mathbf{X}_1, \ldots, \mathbf{X}_N$ according to the following mechanism, parameterized by a vector $\mathbf{a}$:
  1. Generate a uniform random permutation $(e_1, \ldots, e_n)$. Set $k = 1$.
  2. Calculate $C = c_{e_k} + \sum_{i=1}^{k-1} X_{e_i} c_{e_i}$.
  3. If $C \leq C_{\max}$, draw $X_{e_k} \sim \mathsf{Ber}(a_{e_k})$. Otherwise set $X_{e_k} = 0$.
  4. If $\sum_{i=1}^{k} X_{e_i} c_{e_i} > C_{\max}$ then stop; otherwise set $k = k + 1$ and reiterate from step 2.
- update the $\mathbf{a}$ as the mean of the elite samples.

Main CE Algorithm

1 Initialise $\widehat{\mathbf{a}}_0$. Set $t = 1$ (iteration counter).

2 Generate a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_N$ using $\mathbf{a} = \widehat{\mathbf{a}}_{t-1}$. Let $\hat{r}_{(1)}, \ldots, \hat{r}_{(N)}$ be the order statistics of the estimates $\hat{r}(\mathbf{X}_1), \ldots, \hat{r}(\mathbf{X}_N)$. Let $\widehat{\gamma}_t = \hat{r}_{(\lceil (1-\rho)N \rceil)}$ be the worst reliability of the elite samples.

3 Update $\widehat{\mathbf{a}}_t$ as the mean of the elite samples.

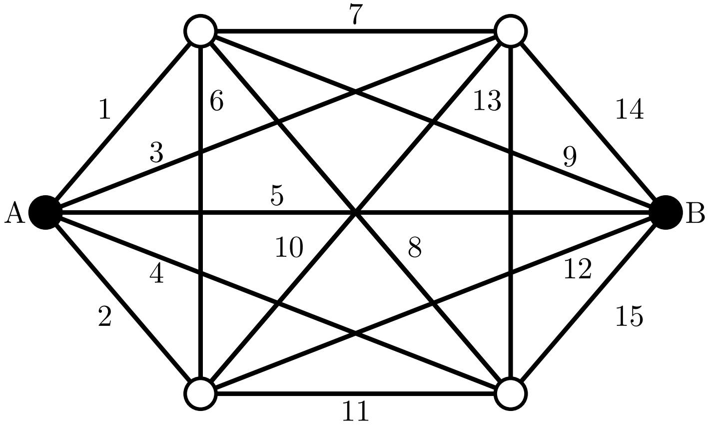
4 Stop if

$$\max(\min(\widehat{\mathbf{a}}_t, 1 - \widehat{\mathbf{a}}_t)) \leq \beta$$

for some small fixed β; otherwise set $t = t + 1$ and reiterate from step 2.

Numerical Example

Suppose that any link in this network can be purchased.



We wish to purchase the links that yield the smallest network unreliability, using the CE method.

Numerical Example

CE Parameters:

N	ρ	β
300	0.1	0.02

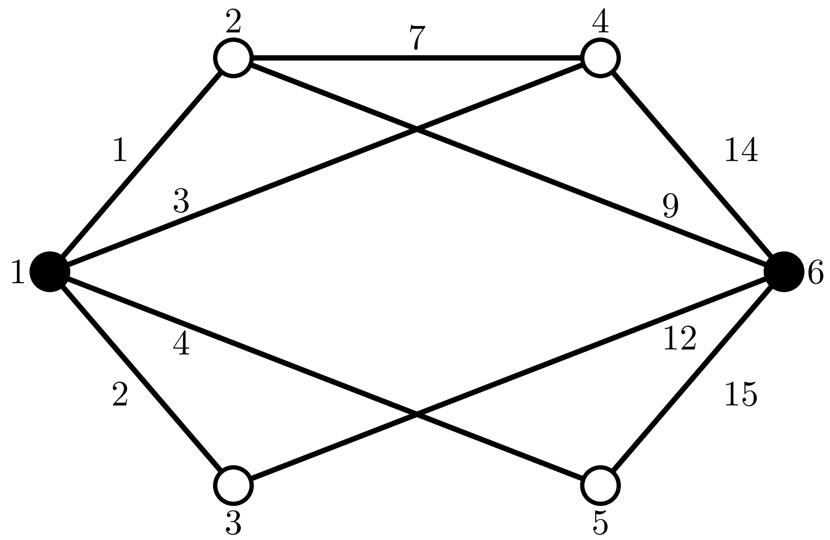
Let $C_{\max} = 1$. Link costs and reliabilities:

i	c_i	p_i	i	c_i	p_i	i	c_i	p_i
1	0.2416	0.8081	6	0.2301	0.8611	11	0.2894	0.8527
2	0.2822	0.7956	7	0.1546	0.8063	12	0.2712	0.8294
3	0.1531	0.8036	8	0.1984	0.8304	13	0.1632	0.7376
4	0.1586	0.8835	9	0.2221	0.7203	14	0.2461	0.8677
5	2.3000	0.7665	10	0.1835	0.7693	15	0.2424	0.8698

Evolution

Optimal Network

CE find the optimal network, even with noisy estimates.



The exact optimal network unreliability is 0.006584.

• Kroese, D.P., Nariai, S, Hui, K-P (2007). Network Reliability Optimization via the Cross-Entropy Method. *IEEE Trans. Rel.* **56** (2), 275–287.

Optimal Control for Epidemic Models

In the Susceptible-Infective model the proportion of infectives in a constant population follows the DE (μ is the death rate):

$$i'(t) = \beta(u(t))\, i(t)\, (1 - i(t)) - \mu\, i(t), \quad i(0) = i_0, \quad (\star)$$

We assume that the infection rate β is a function of the rate $u(t)$ at which budget is spent, e.g., $\beta(u) = \frac{\beta_{\max}}{1+u}$.

Objective : minimise the new infective cases over $[0, T]$

$$\min_u J(u) = \min_u \int_0^T \beta(u)\, i(t)\, (1 - i(t))\, dt,$$

subject to $\int_0^T u(t)dt = K,\ u(t) \geq 0$, and the evolution ($\star$).

Solving OC via CE

Suppose we want to find γ^*, assuming it exists.

$$\gamma^* = J(\mathbf{u}^*) = \min_{\mathbf{u} \in \mathcal{U}} J(\mathbf{u}),$$

where $\mathbf{u} = \left\{ u^{(r)}(t), r = 1, \ldots, R, t \in [0, T] \right\}$ where R is the number of control functions.

Solving OC via CE

Suppose we want to find γ^*, assuming it exists.

$$\gamma^* = J(\mathbf{u}^*) = \min_{\mathbf{u}\in\mathcal{U}} J(\mathbf{u}),$$

where $\mathbf{u} = \left\{ u^{(r)}(t), r = 1, \ldots, R, t \in [0, T] \right\}$ where R is the number of control functions.

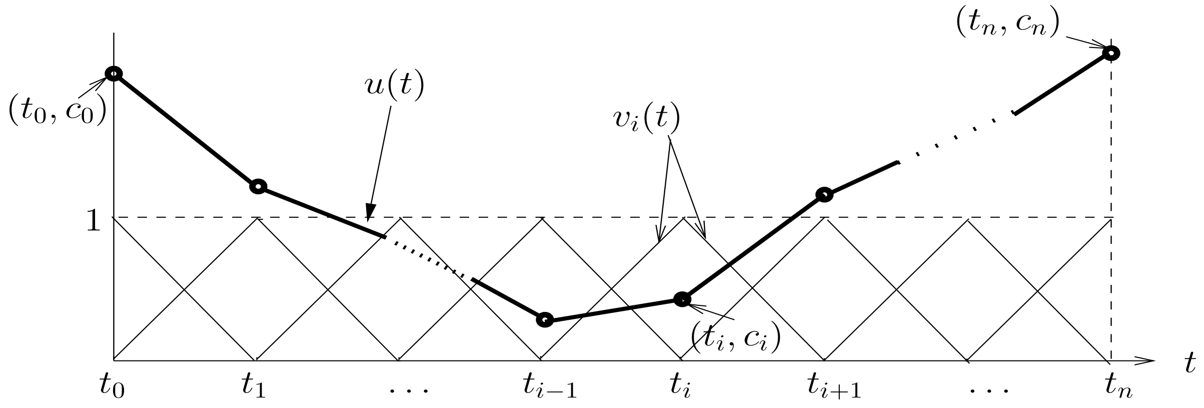
Instead of solving this functional optimization program directly, we consider a related parametric optimization program, namely,

$$\gamma_{\mathbf{c}^*} = J(\mathbf{u}_{\mathbf{c}^*}) = \min_{\mathbf{c}\in\mathcal{C}} J(\mathbf{u}_{\mathbf{c}})$$

$\mathbf{u}_{\mathbf{c}} = \{u_{\mathbf{c}}^{(r)}, r = 1, \ldots, R\}$, $\mathbf{c} = \{c_i^{(r)}, r = 1, \ldots, R;\ i = 0, \ldots, n\}$.

Interpolation via FEM

The control function $u_{\mathbf{c}}^{(j)}$ could be obtained by interpolating the set of points $\{(t_i, c_i^{(j)}), i = 0, \ldots, n\}$ using the Finite Element Method.



If the collection $\{\mathbf{u_c}, \mathbf{c} \in \mathcal{C}\}$ is chosen large enough, then $\gamma^* \approx \gamma_{\mathbf{c}^*}$ and $\mathbf{u}^* \approx \mathbf{u}_{\mathbf{c}^*}$.

CE Algorithm for Optimal Control

1. Initialize: Choose $\boldsymbol{\mu}_0 = \{\mu_{0i}^{(r)}, r = 1, \ldots, R; i = 0, \ldots, n\}$ and $\boldsymbol{\sigma}_0 = \{\sigma_{0i}^{(r)}, r = 1, \ldots, R; i = 0, \ldots, n\}$. Set $k := 1$.

2. Draw: Generate a random sample $\mathbf{C}_1, \ldots, \mathbf{C}_N \sim \mathsf{N}(\boldsymbol{\mu}_{\mathsf{k}-1}, \boldsymbol{\sigma}^2_{\mathsf{k}-1})$ with $\mathbf{C}_m = \{C_{mi}^{(r)}, r = 1, \ldots, R;\ i = 0, 1, \ldots, n\}$.

3. Evaluate: For each control vector $\mathbf{C}_m$ evaluate the objective function $J(\mathbf{u}_{\mathbf{C}_m})$, e.g., by solving the ODE system using RK techniques.

4. Select: Find the N^{elite} best performing (=elite) samples, based on the values $\{J(\mathbf{u}_{\mathbf{C}_m})\}$. Let $\mathcal{I}$ be the corresponding set of indices.

CE Algorithm for Optimal Control

5. Update: for all $r = 1, \ldots, R$; $i = 0, 1, \ldots, n$, let

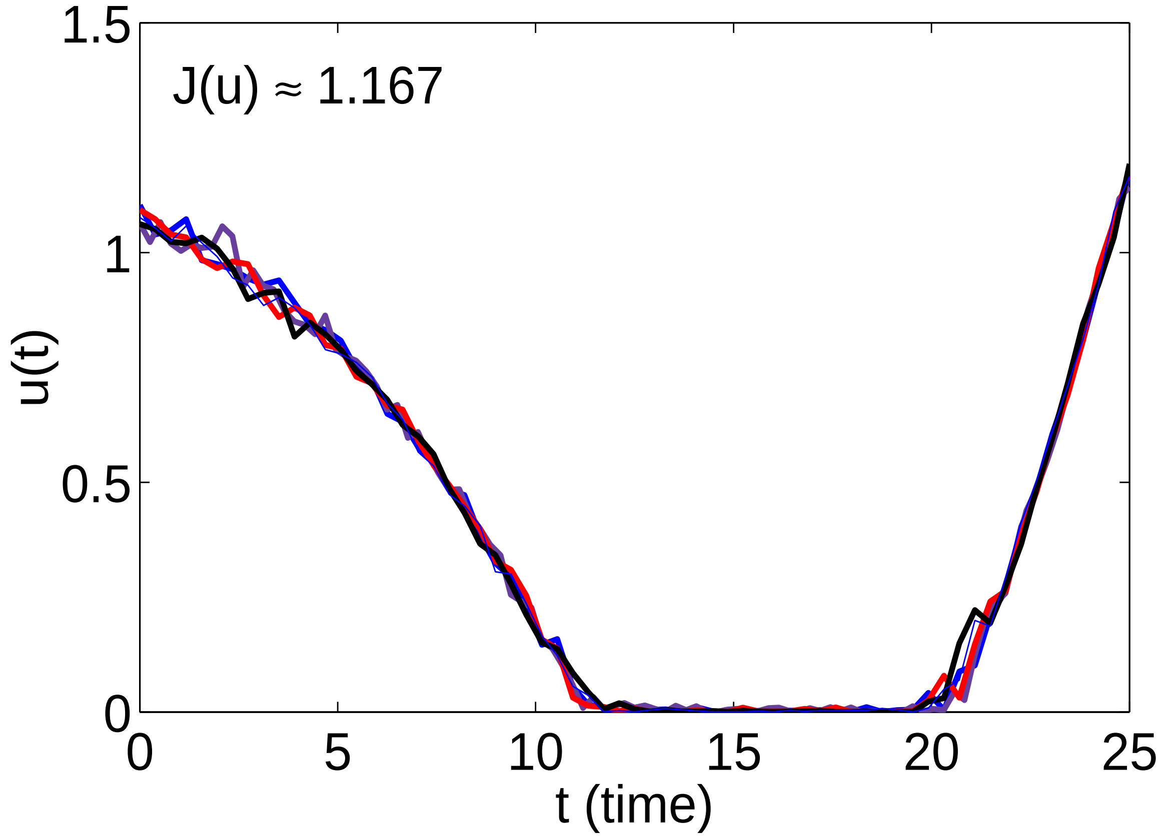
$$\widetilde{\mu}_{ik}^{(r)} := \frac{1}{N^{\text{elite}}} \sum_{m \in \mathcal{I}} C_{im}^{(r)} \quad \text{and} \quad \tilde{\sigma}_{ik}^{(r)\,2} := \frac{1}{N^{\text{elite}}} \sum_{m \in \mathcal{I}} \left(C_{im}^{(r)} - \mu_{ik}^{(r)} \right)^2 .$$

6. Smooth: For a fixed smoothing parameter $0 < \alpha \leq 1$, let

$$\hat{\boldsymbol{\mu}}_k := \alpha \widetilde{\boldsymbol{\mu}}_k + (1 - \alpha) \hat{\boldsymbol{\mu}}_{k-1}, \quad \hat{\boldsymbol{\sigma}}_k := \alpha \widetilde{\boldsymbol{\sigma}}_k + (1 - \alpha) \hat{\boldsymbol{\sigma}}_{k-1}$$

7. Stop: Repeat 2–6 until $\max_{i,r} \sigma_{ik}^{(r)} < \varepsilon$. Let L be the final iteration number. Return $\boldsymbol{\mu}_L$ as an estimate of the optimal control parameter $\mathbf{c}^*$.

CE Solution for the SI Problem



Application: HIV Control

Objective Function

$$\min_{\mathbf{u}} J(\mathbf{u}) = \min_{\mathbf{u}} \int_0^T \sum_{r=1}^R \sum_{j=1}^R \beta_{jr} \left[\frac{i_M^{(j)}}{n_M^{(j)}} s_F^{(r)} + \frac{i_F^{(j)}}{n_F^{(j)}} s_M^{(r)} \right] dt,$$

Application: HIV Control

Objective Function

$$\min_{\mathbf{u}} J(\mathbf{u}) = \min_{\mathbf{u}} \int_0^T \sum_{r=1}^R \sum_{j=1}^R \beta_{jr} \left[\frac{i_M^{(j)}}{n_M^{(j)}} s_F^{(r)} + \frac{i_F^{(j)}}{n_F^{(j)}} s_M^{(r)} \right] dt,$$

the rate at which new infectives are generated.

Application: HIV Control

subject to:

(1). integral constraints

$$\int_0^T u^{(r)}\, dt = K^{(r)},\ 0 \leq u^{(r)},\ r = 1, \ldots, R.$$

(2). state (dynamic) constraints

$$\begin{aligned}
\frac{ds_F^{(r)}}{dt} &= B_F - \sum_{j=1}^{R} \beta_{jr}\, \frac{i_M^{(j)}}{n_M^{(j)}}\ s_F^{(r)} - \mu\, s_F^{(r)}, \\
\frac{di_F^{(r)}}{dt} &= \sum_{j=1}^{R} \beta_{jr}\, \frac{i_M^{(j)}}{n_M^{(j)}}\ s_F^{(r)} - (\mu + \gamma)\, i_F^{(r)}, \\
\frac{ds_M^{(r)}}{dt} &= B_M - \sum_{j=1}^{R} \beta_{jr}\, \frac{i_F^{(j)}}{n_F^{(j)}}\ s_M^{(r)} - \mu\, s_M^{(r)}, \\
\frac{di_M^{(r)}}{dt} &= \sum_{j=1}^{R} \beta_{jr}\, \frac{i_F^{(j)}}{n_F^{(j)}}\ s_M^{(r)} - (\mu + \gamma)\, i_M^{(r)},
\end{aligned}$$

Application: HIV Control

subject to:

the rate at which budget is spent at time *t* in patch *r*.

(1). integral constraints

$$\int_0^T u^{(r)} dt = K^{(r)},\ 0 \le u^{(r)},\ r = 1, \ldots, R.$$

(2). state (dynamic) constraints

$$\begin{aligned}
\frac{ds_F^{(r)}}{dt} &= B_F - \sum_{j=1}^R \beta_{jr} \frac{i_M^{(j)}}{n_M^{(j)}}\, s_F^{(r)} - \mu\, s_F^{(r)},\\
\frac{di_F^{(r)}}{dt} &= \sum_{j=1}^R \beta_{jr} \frac{i_M^{(j)}}{n_M^{(j)}}\, s_F^{(r)} - (\mu + \gamma)\, i_F^{(r)},\\
\frac{ds_M^{(r)}}{dt} &= B_M - \sum_{j=1}^R \beta_{jr} \frac{i_F^{(j)}}{n_F^{(j)}}\, s_M^{(r)} - \mu\, s_M^{(r)},\\
\frac{di_M^{(r)}}{dt} &= \sum_{j=1}^R \beta_{jr} \frac{i_F^{(j)}}{n_F^{(j)}}\, s_M^{(r)} - (\mu + \gamma)\, i_M^{(r)},
\end{aligned}$$

Application: HIV Control

subject to:

the total budget available in patch r.

(1). integral constraints

$$\int_0^T u^{(r)}\, dt = K^{(r)},\ 0 \le u^{(r)},\ r = 1, \ldots, R.$$

(2). state (dynamic) constraints

$$\begin{aligned}
\frac{ds_F^{(r)}}{dt} &= B_F - \sum_{j=1}^R \beta_{jr}\, \frac{i_M^{(j)}}{n_M^{(j)}}\ s_F^{(r)} - \mu\, s_F^{(r)}, \\
\frac{di_F^{(r)}}{dt} &= \sum_{j=1}^R \beta_{jr}\, \frac{i_M^{(j)}}{n_M^{(j)}}\ s_F^{(r)} - (\mu + \gamma)\, i_F^{(r)}, \\
\frac{ds_M^{(r)}}{dt} &= B_M - \sum_{j=1}^R \beta_{jr}\, \frac{i_F^{(j)}}{n_F^{(j)}}\ s_M^{(r)} - \mu\, s_M^{(r)}, \\
\frac{di_M^{(r)}}{dt} &= \sum_{j=1}^R \beta_{jr}\, \frac{i_F^{(j)}}{n_F^{(j)}}\ s_M^{(r)} - (\mu + \gamma)\ i_M^{(r)},
\end{aligned}$$

Application: HIV Control

subject to:

the evolution of susceptibles and infectives.

(1). integral constraints

$$\int_0^T u^{(r)}\, dt = K^{(r)},\ 0 \le u^{(r)},\ r = 1, \ldots, R.$$

(2). state (dynamic) constraints

$$\frac{ds_F^{(r)}}{dt} = B_F - \sum_{j=1}^R \beta_{jr}\, \frac{i_M^{(j)}}{n_M^{(j)}}\, s_F^{(r)} - \mu\, s_F^{(r)},$$

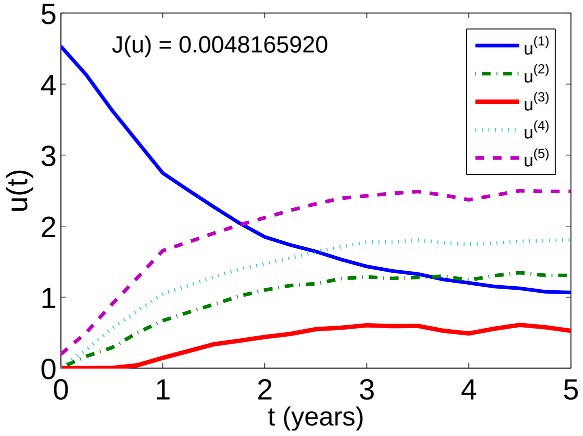
$$\frac{di_F^{(r)}}{dt} = \sum_{j=1}^R \beta_{jr}\, \frac{i_M^{(j)}}{n_M^{(j)}}\, s_F^{(r)} - (\mu + \gamma)\, i_F^{(r)},$$

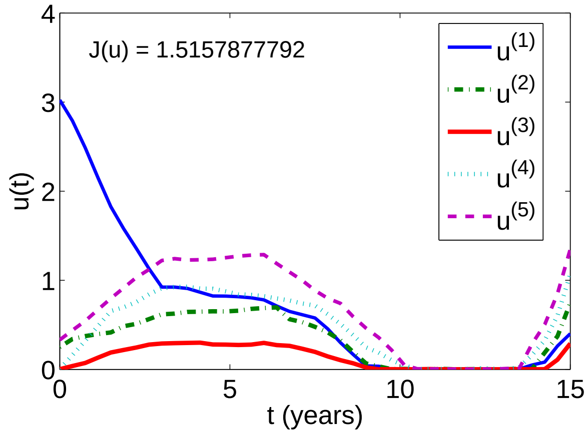
$$\frac{ds_M^{(r)}}{dt} = B_M - \sum_{j=1}^R \beta_{jr}\, \frac{i_F^{(j)}}{n_F^{(j)}}\, s_M^{(r)} - \mu\, s_M^{(r)},$$

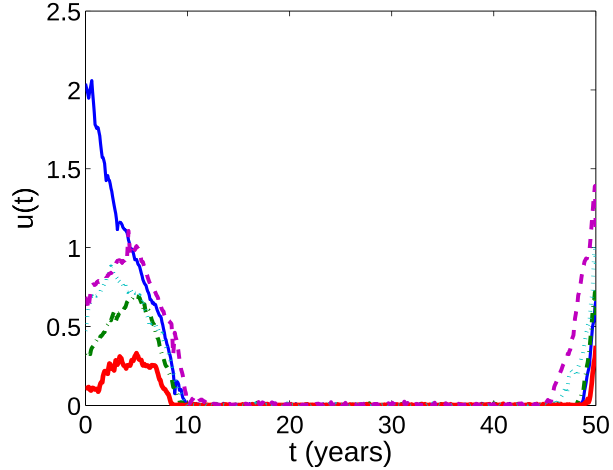
$$\frac{di_M^{(r)}}{dt} = \sum_{j=1}^R \beta_{jr}\, \frac{i_F^{(j)}}{n_F^{(j)}}\, s_M^{(r)} - (\mu + \gamma)\, i_M^{(r)},$$

Numerical Experiments

The optimal trajectory $\mathbf{u}^*$ for $T = 5, T = 15, T = 50$ years.

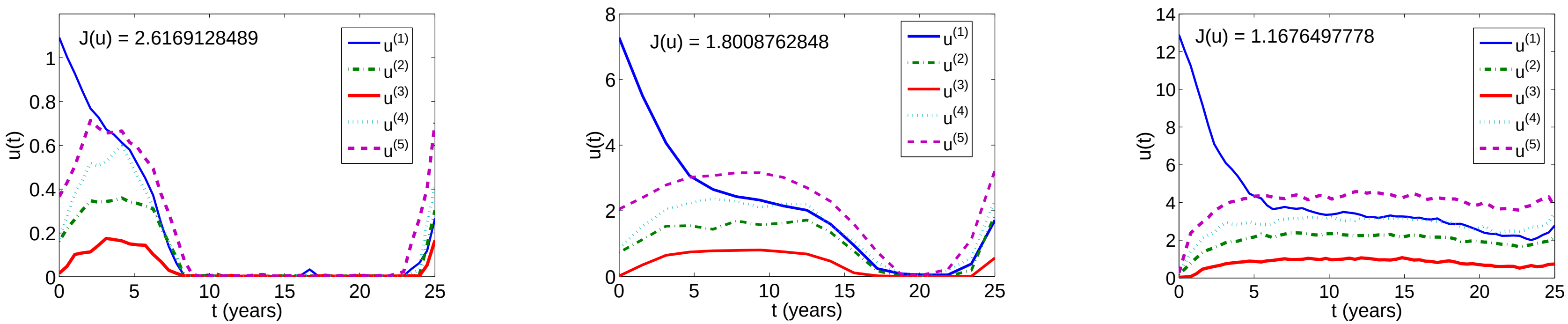






Numerical Experiments

Optimal trajectory $\mathbf{u}^*$ for $K^{(r)} \to \frac{1}{2}K^{(r)}, 5K^{(r)}, 10K^{(r)}$



• Sani, A., Kroese, D.P. (2008). Controling the number of HIV infectives in a mobile population. *Mathematical Biosciences*. Accepted for publication.

Future Research

- Parallel CE (works!)
- Overcoming likelihood ratio degeneracy
- Parameter-free methods
- Applications (lots)

Thanks

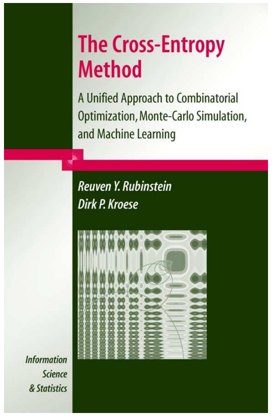
- Reuven Rubinstein
- Kin-Ping Hui, Sho Nariai, Asrul Sani, Thomas Taimre, Gareth Evans, Zdravko Botev.

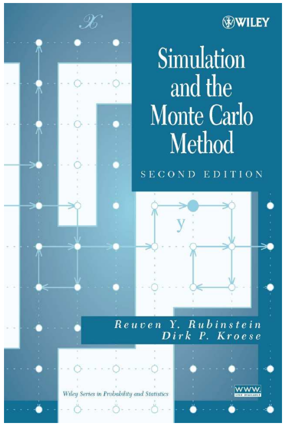
THANK YOU

See also ...

Books: R.Y. Rubinstein and D.P. Kroese.

- *The Cross-Entropy Method*, Springer-Verlag, 2004.
- *Simulation and the Monte Carlo Method*, 2nd Edition, Wiley & Sons, 2007.





The CE **home page**: `http://www.cemethod.org`

Special Issue: *Annals of Operations Research, 2009.* Monte Carlo Methods for Simulation, Optimization and Counting.