Inference and Optimal Experimental Design for Percolation and Random Graph Models

(based on joint work with Gavin Gibson and Stan Zachary)

Andrei Bejan

Computer Laboratory, University of Cambridge

17 February 2009

Abstract

In this talk the problems of inference and optimal design for stochastic interaction models based on random graphs are considered. By a design it is meant a (spatial) arrangement of locations of observation nodes in some (metric) space within a framework of a certain probabilistic rule of the appearance of edges.

As traditionally, such probabilistic rule may involve parameter(s) that one would like to infer on. A host of questions is put for such models. These questions include the problems of inference, optimal design, elaboration of algorithms of identifying such designs, etc.

An optimal design is the “most informative” arrangement.

Abstract

In this talk the problems of inference and optimal design for stochastic interaction models based on random graphs are considered. By a design it is meant a (spatial) arrangement of locations of observation nodes in some (metric) space within a framework of a certain probabilistic rule of the appearance of edges.

As traditionally, such probabilistic rule may involve parameter(s) that one would like to infer on. A host of questions is put for such models. These questions include the problems of inference, optimal design, elaboration of algorithms of identifying such designs, etc.

An optimal design is the “most informative” arrangement.

Abstract

In this talk the problems of inference and optimal design for stochastic interaction models based on random graphs are considered. By a design it is meant a (spatial) arrangement of locations of observation nodes in some (metric) space within a framework of a certain probabilistic rule of the appearance of edges.

As traditionally, such probabilistic rule may involve parameter(s) that one would like to infer on. A host of questions is put for such models. These questions include the problems of inference, optimal design, elaboration of algorithms of identifying such designs, etc.

An optimal design is the “most informative” arrangement.

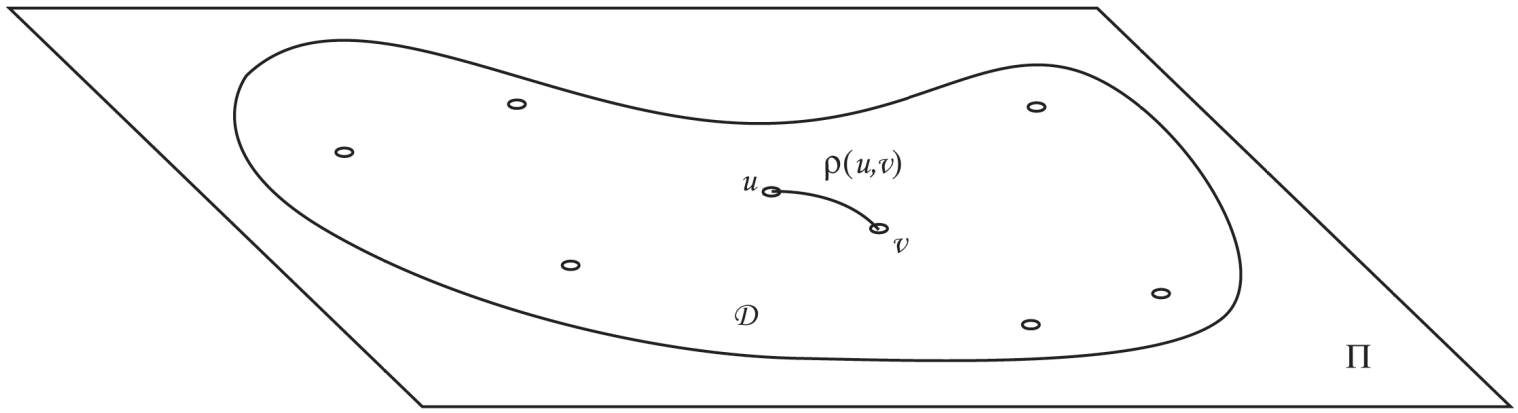


Figure: n objects are considered in the region $\mathcal{D} \subseteq \Pi$; there is a link between each pair (u, v) of them with probability $\mathbb{P}(u \leftrightarrow v; \theta) = p(\rho(u, v); \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^k$.

The function ρ is defined on $\mathcal{D}$ and may well be considered to represent a metric. If so, then (i) $p : \mathbb{R}^+ \times \Theta \to [0, 1]$; (ii) the function $p(\rho, \theta)$ is monotonically decreasing in ρ for any fixed θ.

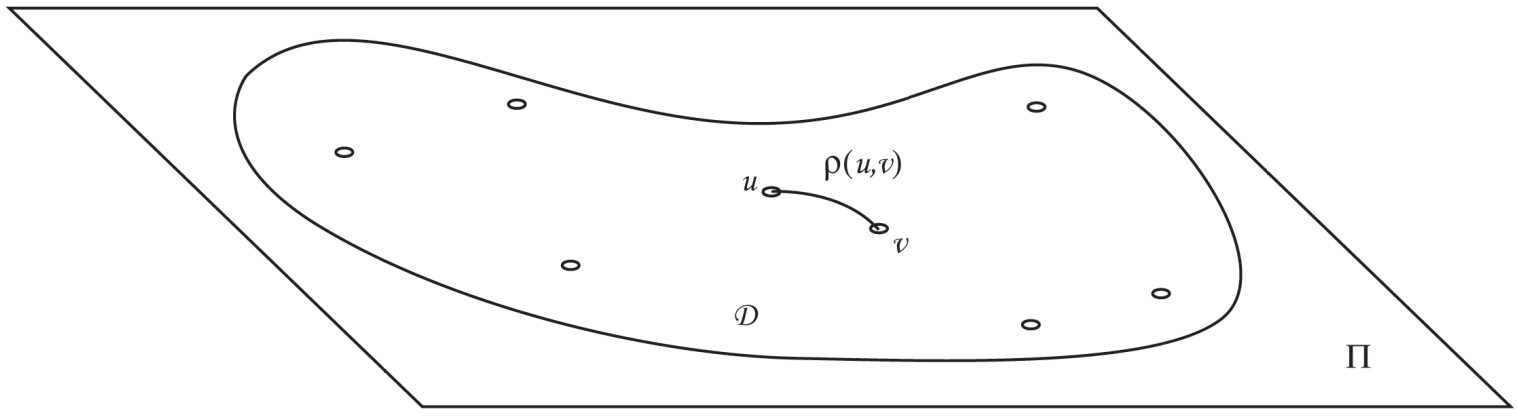


Figure: n objects are considered in the region $\mathcal{D} \subseteq \Pi$; there is a link between each pair (u, v) of them with probability $\mathbb{P}(u \leftrightarrow v; \theta) = p(\rho(u, v); \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^k$.

The function ρ is defined on $\mathcal{D}$ and may well be considered to represent a metric. If so, then (i) $p : \mathbb{R}^+ \times \Theta \to [0, 1]$; (ii) the function $p(\rho, \theta)$ is monotonically decreasing in ρ for any fixed θ.

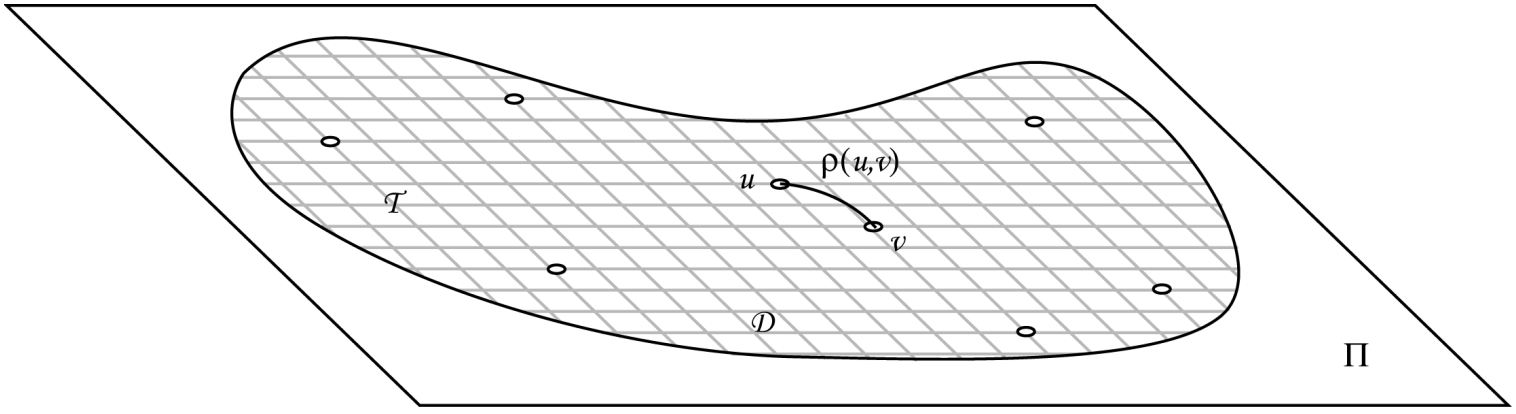


Figure: n objects are considered in the region $\mathcal{D} \subseteq \Pi$; there is a "connection" between each pair (u, v) of them with probability $\mathbb{P}(u \leftrightarrow v; \theta) = p(\rho(u, v); \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^k$.

The function ρ is defined on $\mathcal{D}$ and may well be considered to represent a metric. If so, then (i) $p : \mathbb{R}^+ \times \Theta \to [0, 1]$; (ii) the function $p(\rho, \theta)$ is monotonically decreasing in ρ for any fixed θ.

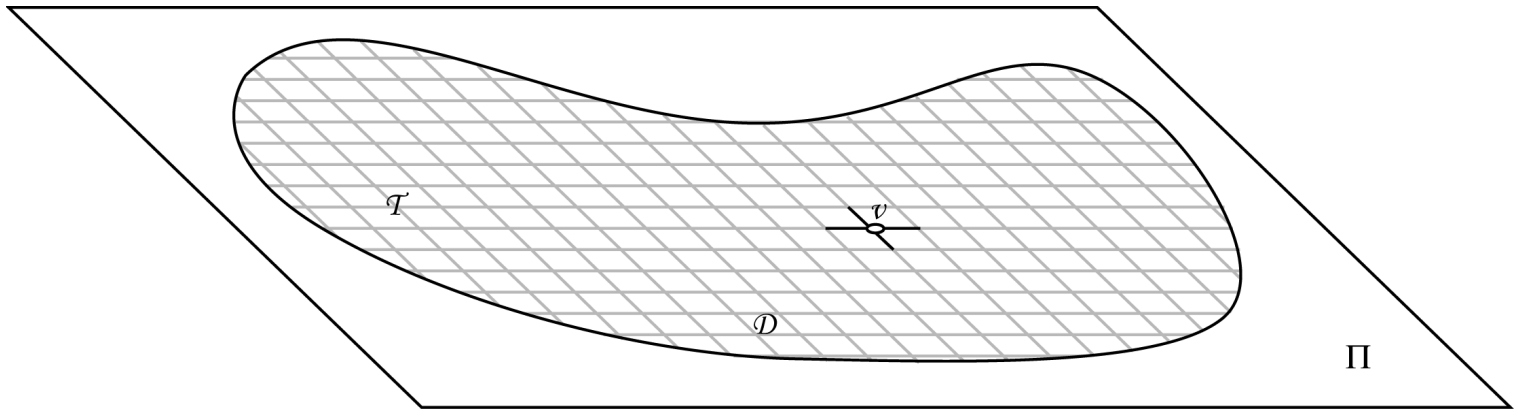


Figure: n objects are considered in the region $\mathcal{D} \subseteq \Pi$; there is a "connection" between each pair (u, v) of them with probability $\mathbb{P}(u \leftrightarrow v; \theta) = p(\rho(u, v); \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^k$.

The function ρ is defined on $\mathcal{D}$ and may well be considered to represent a metric. If so, then (i) $p : \mathbb{R}^+ \times \Theta \to [0, 1]$; (ii) the function $p(\rho, \theta)$ is monotonically decreasing in ρ for any fixed θ.

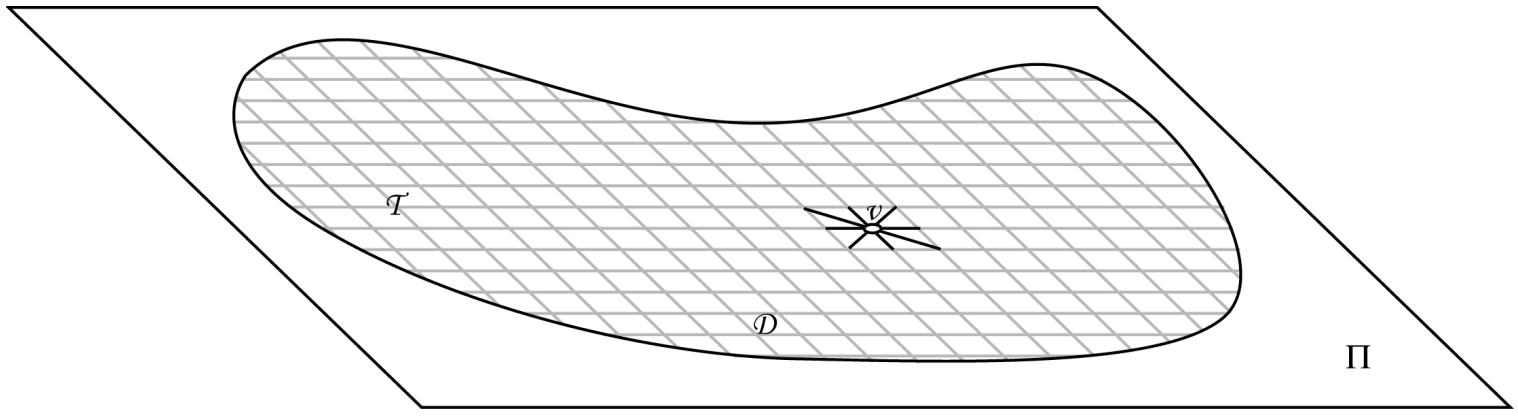


Figure: n objects are considered in the region $\mathcal{D} \subseteq \Pi$; there is a "connection" between each pair (u, v) of them with probability $\mathbb{P}(u \leftrightarrow v; \theta) = p(\rho(u, v); \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^k$.

The function ρ is defined on $\mathcal{D}$ and may well be considered to represent a metric. If so, then (i) $p : \mathbb{R}^+ \times \Theta \to [0, 1]$; (ii) the function $p(\rho, \theta)$ is monotonically decreasing in ρ for any fixed θ.

Names for $p(\rho, \theta)$:

- edge probability function
- connectivity kernel
- edge probability profile

Example

Functions

$$p(\rho, \theta) = e^{-\theta\rho},\ \theta \in \Theta \equiv [0, \infty), \tag{1}$$

$$p(\rho, \theta) = (1 + \theta_1\rho)^{-\theta_2},\ \theta \in \Theta \equiv \mathbb{R}^+ \times \mathbb{R}^+ \tag{2}$$

$$p(\rho, \theta) = (1 + \theta\rho^2)^{-1},\ \theta \in \Theta \equiv \mathbb{R}^+ \tag{3}$$

satisfy the above conditions and can be considered as connectivity kernels. These decay laws are called **exponential**, **power** and **Cauchy**, correspondingly.

Names for $p(\rho, \theta)$:

- edge probability function
- connectivity kernel
- edge probability profile

Example

Functions

$$p(\rho, \theta) = e^{-\theta\rho},\ \theta \in \Theta \equiv [0, \infty), \tag{1}$$

$$p(\rho, \theta) = (1 + \theta_1\rho)^{-\theta_2},\ \theta \in \Theta \equiv \mathbb{R}^+ \times \mathbb{R}^+ \tag{2}$$

$$p(\rho, \theta) = (1 + \theta\rho^2)^{-1},\ \theta \in \Theta \equiv \mathbb{R}^+ \tag{3}$$

satisfy the above conditions and can be considered as connectivity kernels. These decay laws are called **exponential**, **power** and **Cauchy**, correspondingly.

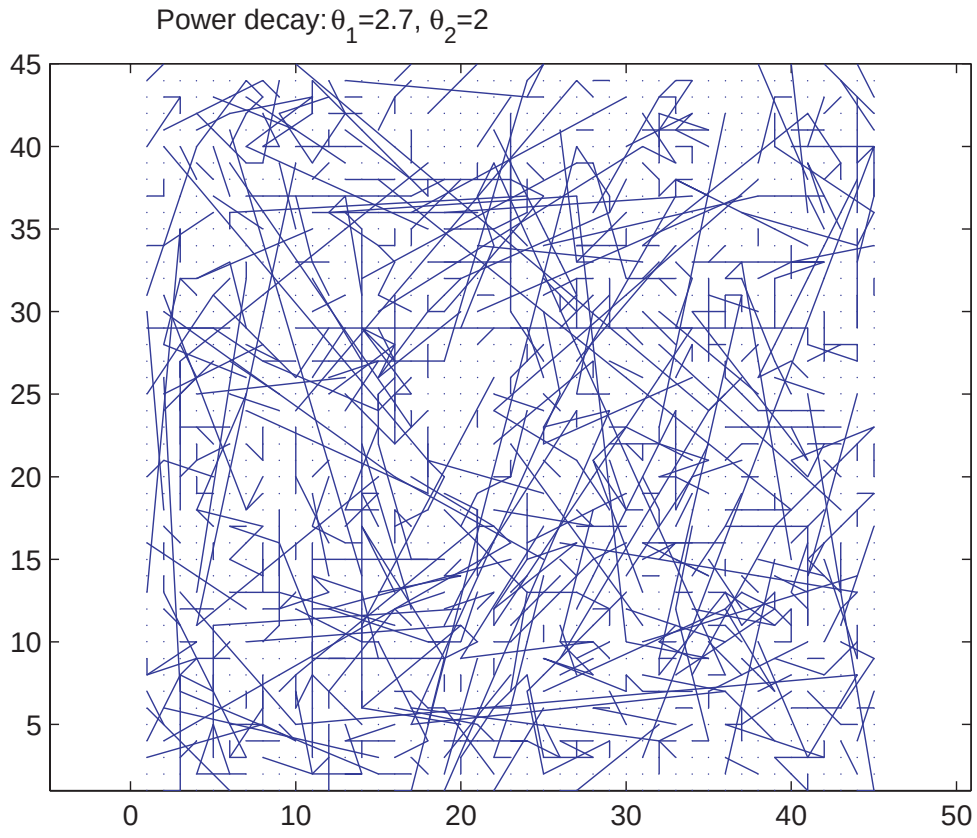


Figure: Realisation of a random graph on a square 45×45 grid governed by the power law connectivity kernel $p(\rho, \theta) = (1 + \theta_1 \rho)^{-\theta_2}$ (a latent process for smaller graphs within this region).

Model

A family of parametrised connectivity kernels

$$\mathcal{K} = \{p(\rho, \theta)\}_{\theta \in \Theta},$$

where θ specifies a probability edge function.

Statistical interest

After observing a random graph on n nodes, formation of which is based on $\mathcal{K}$ (assumption!), we want to be able to infer on θ.

Design question

Being able to choose n nodes within a set of potential nodes to observe a random graph on them governed by $\mathcal{K}$, to design an optimal arrangement of these n nodes, so that the observation is “the most informative” on θ.

Model

A family of parametrised connectivity kernels

$$\mathcal{K} = \{p(\rho, \theta)\}_{\theta \in \Theta},$$

where θ specifies a probability edge function.

Statistical interest

After observing a random graph on n nodes, formation of which is based on $\mathcal{K}$ (assumption!), we want to be able to infer on θ.

Design question

Being able to choose n nodes within a set of potential nodes to observe a random graph on them governed by $\mathcal{K}$, to design an optimal arrangement of these n nodes, so that the observation is “the most informative” on θ.

Model

A family of parametrised connectivity kernels

$$\mathcal{K} = \{p(\rho, \theta)\}_{\theta \in \Theta},$$

where θ specifies a probability edge function.

Statistical interest

After observing a random graph on n nodes, formation of which is based on $\mathcal{K}$ (assumption!), we want to be able to infer on θ.

Design question

Being able to choose n nodes within a set of potential nodes to observe a random graph on them governed by $\mathcal{K}$, to design an optimal arrangement of these n nodes, so that the observation is “the most informative” on θ.

Different processes from the following fields give rise to the described model:

- chemical kinetics (e.g. Firth and Hinde (1997), Atkinson and Donev (1992))
  first order reaction: $A \to B$, $\mu(t;\theta) = \exp(-\theta t)$
  two consecutive first order reactions: $A \xrightarrow{\theta_1} B \xrightarrow{\theta_2} C$, $\eta(t,\theta_1,\theta_2) = \frac{\theta_1}{\theta_1-\theta_2}\{e^{-\theta_2 t} - e^{-\theta_1 t}\}$, provided that $\theta_1 > \theta_2 > 0$.
- radio networks (random mobile graphs introduced by Tyrakowski and Palka (2005) for analysis of distributed algorithms requiring synchronous communication in radio networks)
- disease evolution on networks (e.g. Read and Keeling (2003))

Different processes from the following fields give rise to the described model:

- chemical kinetics (e.g. Firth and Hinde (1997), Atkinson and Donev (1992))
  first order reaction: $A \to B$, $\mu(t;\theta) = \exp(-\theta t)$
  two consecutive first order reactions: $A \xrightarrow{\theta_1} B \xrightarrow{\theta_2} C$, $\eta(t,\theta_1,\theta_2) = \frac{\theta_1}{\theta_1-\theta_2}\{e^{-\theta_2 t} - e^{-\theta_1 t}\}$, provided that $\theta_1 > \theta_2 > 0$.
- radio networks (random mobile graphs introduced by Tyrakowski and Palka (2005) for analysis of distributed algorithms requiring synchronous communication in radio networks)
- disease evolution on networks (e.g. Read and Keeling (2003))

Different processes from the following fields give rise to the described model:

- chemical kinetics (e.g. Firth and Hinde (1997), Atkinson and Donev (1992))
  first order reaction: $A \to B$, $\mu(t;\theta) = \exp(-\theta t)$
  two consecutive first order reactions: $A \xrightarrow{\theta_1} B \xrightarrow{\theta_2} C$, $\eta(t,\theta_1,\theta_2) = \frac{\theta_1}{\theta_1-\theta_2}\{e^{-\theta_2 t} - e^{-\theta_1 t}\}$, provided that $\theta_1 > \theta_2 > 0$.
- radio networks (random mobile graphs introduced by Tyrakowski and Palka (2005) for analysis of distributed algorithms requiring synchronous communication in radio networks)
- disease evolution on networks (e.g. Read and Keeling (2003))

Different processes from the following fields give rise to the described model:

- chemical kinetics (e.g. Firth and Hinde (1997), Atkinson and Donev (1992))
  first order reaction: $A \rightarrow B$, $\mu(t;\theta) = \exp(-\theta t)$
  two consecutive first order reactions: $A \xrightarrow{\theta_1} B \xrightarrow{\theta_2} C$, $\eta(t,\theta_1,\theta_2) = \frac{\theta_1}{\theta_1-\theta_2}\{e^{-\theta_2 t} - e^{-\theta_1 t}\}$, provided that $\theta_1 > \theta_2 > 0$.
- radio networks (random mobile graphs introduced by Tyrakowski and Palka (2005) for analysis of distributed algorithms requiring synchronous communication in radio networks)
- disease evolution on networks (e.g. Read and Keeling (2003))

Different processes from the following fields give rise to the described model:

- chemical kinetics (e.g. Firth and Hinde (1997), Atkinson and Donev (1992))
  first order reaction: $A \rightarrow B$, $\mu(t;\theta) = \exp(-\theta t)$
  two consecutive first order reactions: $A \xrightarrow{\theta_1} B \xrightarrow{\theta_2} C$, $\eta(t,\theta_1,\theta_2) = \frac{\theta_1}{\theta_1-\theta_2}\{e^{-\theta_2 t} - e^{-\theta_1 t}\}$, provided that $\theta_1 > \theta_2 > 0$.
- radio networks (random mobile graphs introduced by Tyrakowski and Palka (2005) for analysis of distributed algorithms requiring synchronous communication in radio networks)
- disease evolution on networks (e.g. Read and Keeling (2003))

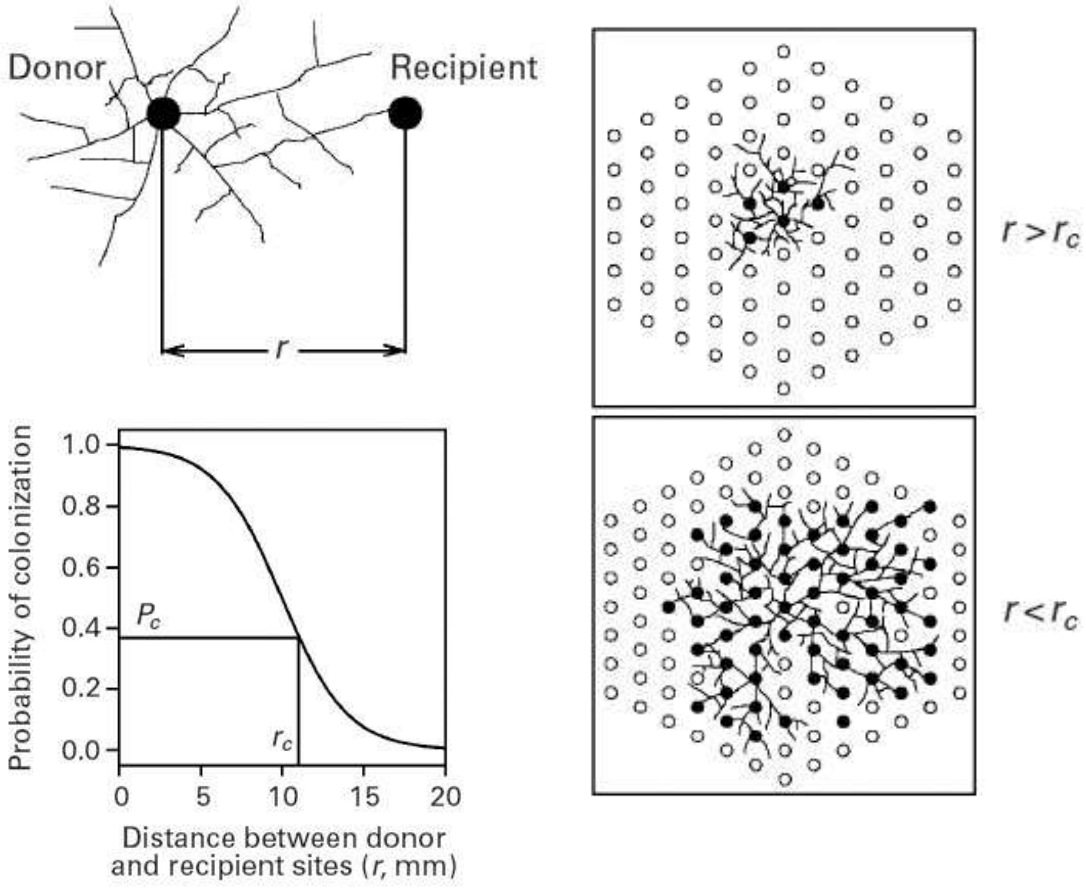


Figure: The growth of mycelial colonies as a percolation process (e.g., D.J. Bailey, W. Otten, C. Gilligan (2000) Saprotrophic invasion by the soil borne fungal plant pathogen Rhizoctonia solani and pecolation thresholds).

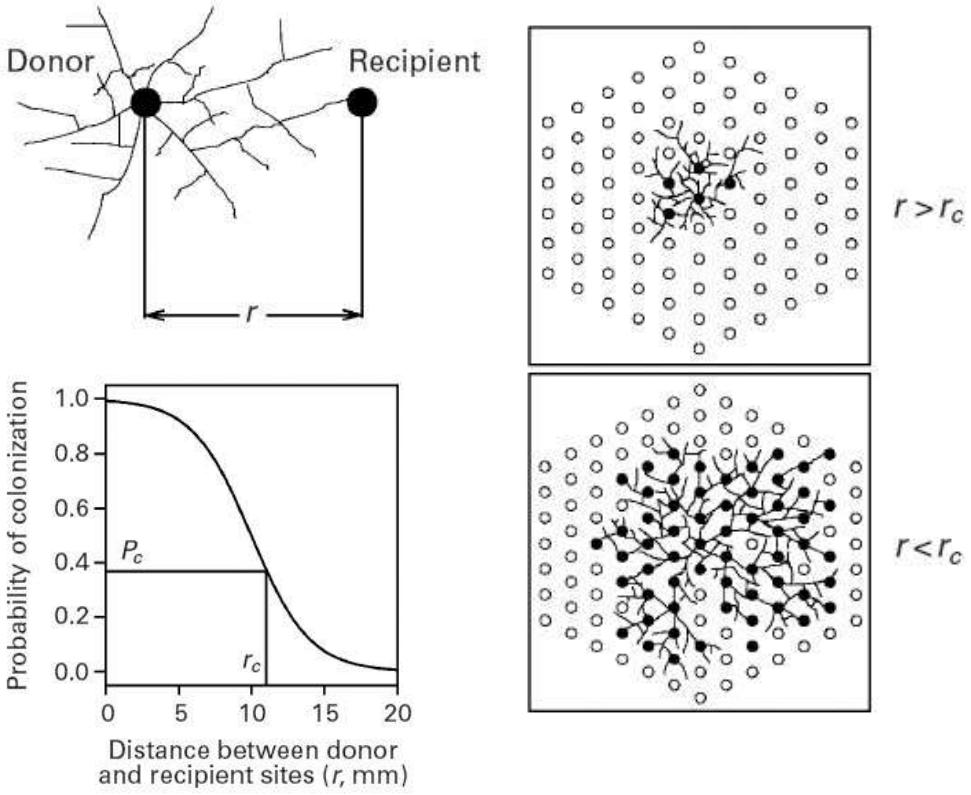


Figure: The edge probability profile may be “combined” from simpler kernels: e.g., Bailey and Gilligan (1997) studying the progress of disease in a population of radish plants exposed to primary infection by *R. solani* in the presense/absence of *T. viride* used the following form for the probability of infection: $p(\rho, \theta) = (\theta_1 + \theta_2 \rho) e^{-\theta_3 \rho}$.

A toy example

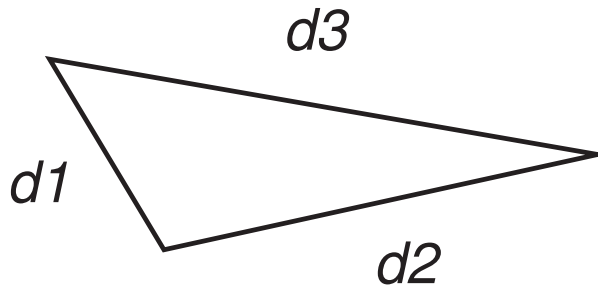


Figure: What are the optimal lengths d_1, d_2, and d_3?

The Fisher information function is additive:

$$I(\theta; d_1, d_2, d_3) = \sum_{i=1}^{3} d_i^2 \frac{e^{-\theta d_i}}{1 - e^{-\theta d_i}}$$

with the triangle inequality imposed on d_1, d_2, and d_3.

A toy example

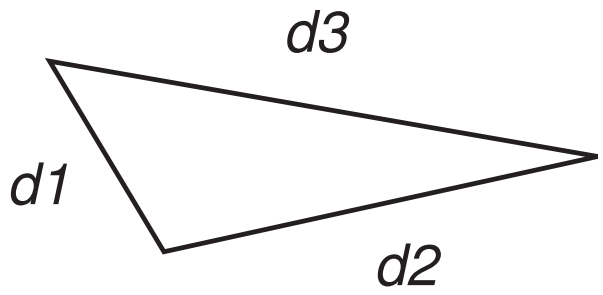


Figure: What are the optimal lengths d_1, d_2, and d_3?

The Fisher information function is additive:

$$I(\theta; d_1, d_2, d_3) = \sum_{i=1}^{3} d_i^2 \frac{e^{-\theta d_i}}{1 - e^{-\theta d_i}}$$

with the triangle inequality imposed on d_1, d_2, and d_3.

A toy example

The Fisher information function is additive:

$$I(\theta; d_1, d_2, d_3) = \sum_{i=1}^{3} d_i^2 \frac{e^{-\theta d_i}}{1 - e^{-\theta d_i}}$$

with the triangle inequality imposed on d_1, d_2, and d_3.

However,

$$d_i^2 \frac{e^{-\theta d_i}}{1 - e^{-\theta d_i}} \leq I(\rho_{max}),$$

and this upper bound is attainable for each of the three summands, hence:

$$d_1 = d_2 = d_3 = \rho_{max} \approx 1.6\theta^{-1}.$$

A toy example

The Fisher information function is additive:

$$I(\theta; d_1, d_2, d_3) = \sum_{i=1}^{3} d_i^2 \frac{e^{-\theta d_i}}{1 - e^{-\theta d_i}}$$

with the triangle inequality imposed on d_1, d_2, and d_3.

However,

$$d_i^2 \frac{e^{-\theta d_i}}{1 - e^{-\theta d_i}} \leq I(\rho_{max}),$$

and this upper bound is attainable for each of the three summands, hence:

$$d_1 = d_2 = d_3 = \rho_{max} \approx 1.6\theta^{-1}.$$

A toy example

The Fisher information function is additive:

$$I(\theta; d_1, d_2, d_3) = \sum_{i=1}^{3} d_i^2 \frac{e^{-\theta d_i}}{1 - e^{-\theta d_i}}$$

with the triangle inequality imposed on d_1, d_2, and d_3.

However,

$$d_i^2 \frac{e^{-\theta d_i}}{1 - e^{-\theta d_i}} \leq I(\rho_{max}),$$

and this upper bound is attainable for each of the three summands, hence:

$$d_1 = d_2 = d_3 = \rho_{max} \approx 1.6\theta^{-1}.$$

So called locally D-optimal designs require a good knowledge of the true parameter values.

Cochran (1973): *"... and I promise to design the best experiment for estimating θ"*.

So called locally D-optimal designs require a good knowledge of the true parameter values.
Cochran (1973): *"... and I promise to design the best experiment for estimating θ"*.

So called locally D-optimal designs require a good knowledge of the true parameter values.
Cochran (1973): *" You tell me the value of θ and I promise to design the best experiment for estimating θ"*.

So called locally D-optimal designs require a good knowledge of the true parameter values.
Cochran (1973): *" You tell me the value of θ and I promise to design the best experiment for estimating θ"* or, in other words: "if we happen to perform the experiment with ρ being very different from $\rho_{max}(\theta)$ then we loose a considerable amount of information".

Proper prior -> experiment -> posterior

How to formulate the optimality in designing an experiment in these settings?

Solution

How to obtain the solution to the optimal design problem?

“Equilateral” designs?

Is equilateral design (equidistant or regular arrangement) still optimal under the new formulation of the problem?

Proper prior -> experiment -> posterior

How to formulate the optimality in designing an experiment in these settings?

Solution

How to obtain the solution to the optimal design problem?

“Equilateral” designs?

Is equilateral design (equidistant or regular arrangement) still optimal under the new formulation of the problem?

Proper prior -> experiment -> posterior

How to formulate the optimality in designing an experiment in these settings?

Solution

How to obtain the solution to the optimal design problem?

“Equilateral” designs?

Is equilateral design (equidistant or regular arrangement) still optimal under the new formulation of the problem?

Employing a utility!

Let $\mathbf{y} = \{e(x_i, x_j) = 0 \text{ or } 1\} \in Y$ represent observations, $\mathbf{d} \in \mathcal{D}$—the design vector, and $\theta \in \Theta$—the model parameter. Through using a utility function $u(\mathbf{d}, \theta, \mathbf{y})$ we could represent the purpose and the value of the experiment after observing its outcome. Choices include:

- the negative squared error loss: $u = -\{\theta - \mathbb{E}[\theta|\mathbf{y}, d]\}^2$
- $u(\mathbf{d}, \theta, \mathbf{y}) = \sum_{1 \le i < j \le n} y_{i,j}$ = the total number of present edges
- precision in the Bayesian sense—the inverse of the posterior variance: $u = [\mathbb{V}(\theta|\mathbf{y}, \mathbf{d})]^{-1}$
- discrimination between the prior and the posterior: Kullback-Leibler divergence (Lyndley information measure, Shannon or differential entropy...)

Employing a utility!

Let $\mathbf{y} = \{e(x_i, x_j) = 0 \text{ or } 1\} \in Y$ represent observations, $\mathbf{d} \in \mathcal{D}$—the design vector, and $\theta \in \Theta$—the model parameter. Through using a utility function $u(\mathbf{d}, \theta, \mathbf{y})$ we could represent the purpose and the value of the experiment after observing its outcome. Choices include:

- the negative squared error loss: $u = -\{\theta - \mathbb{E}[\theta|\mathbf{y}, d]\}^2$
- $u(\mathbf{d}, \theta, \mathbf{y}) = \sum_{1 \le i < j \le n} y_{i,j}$ = the total number of present edges
- precision in the Bayesian sense—the inverse of the posterior variance: $u = [\mathbb{V}(\theta|\mathbf{y}, \mathbf{d})]^{-1}$
- discrimination between the prior and the posterior: Kullback-Leibler divergence (Lyndley information measure, Shannon or differential entropy...)

Employing a utility!

Let $\mathbf{y} = \{e(x_i, x_j) = 0 \text{ or } 1\} \in Y$ represent observations, $\mathbf{d} \in \mathcal{D}$—the design vector, and $\theta \in \Theta$—the model parameter. Through using a utility function $u(\mathbf{d}, \theta, \mathbf{y})$ we could represent the purpose and the value of the experiment after observing its outcome. Choices include:

- the negative squared error loss: $u = -\{\theta - \mathbb{E}[\theta|\mathbf{y}, d]\}^2$
- $u(\mathbf{d}, \theta, \mathbf{y}) = \sum_{1 \le i < j \le n} y_{i,j}$ = the total number of present edges
- precision in the Bayesian sense—the inverse of the posterior variance: $u = [\mathbb{V}(\theta|\mathbf{y}, \mathbf{d})]^{-1}$
- discrimination between the prior and the posterior: Kullback-Leibler divergence (Lyndley information measure, Shannon or differential entropy...)

Employing a utility!

Let $\mathbf{y} = \{e(x_i, x_j) = 0 \text{ or } 1\} \in Y$ represent observations, $\mathbf{d} \in \mathcal{D}$—the design vector, and $\theta \in \Theta$—the model parameter. Through using a utility function $u(\mathbf{d}, \theta, \mathbf{y})$ we could represent the purpose and the value of the experiment after observing its outcome. Choices include:

- the negative squared error loss: $u = -\{\theta - \mathbb{E}[\theta|\mathbf{y}, d]\}^2$
- $u(\mathbf{d}, \theta, \mathbf{y}) = \sum_{1 \le i < j \le n} y_{i,j}$ = the total number of present edges
- precision in the Bayesian sense—the inverse of the posterior variance: $u = [\mathbb{V}(\theta|\mathbf{y}, \mathbf{d})]^{-1}$
- discrimination between the prior and the posterior: Kullback-Leibler divergence (Lyndley information measure, Shannon or differential entropy...)

Employing a utility!

Let $\mathbf{y} = \{e(x_i, x_j) = 0 \text{ or } 1\} \in Y$ represent observations, $\mathbf{d} \in \mathcal{D}$—the design vector, and $\theta \in \Theta$—the model parameter. Through using a utility function $u(\mathbf{d}, \theta, \mathbf{y})$ we could represent the purpose and the value of the experiment after observing its outcome. Choices include:

- the negative squared error loss: $u = -\{\theta - \mathbb{E}[\theta|\mathbf{y}, d]\}^2$
- $u(\mathbf{d}, \theta, \mathbf{y}) = \sum_{1 \le i < j \le n} y_{i,j}$ = the total number of present edges
- precision in the Bayesian sense—the inverse of the posterior variance: $u = [\mathbb{V}(\theta|\mathbf{y}, \mathbf{d})]^{-1}$
- discrimination between the prior and the posterior: Kullback-Leibler divergence (Lyndley information measure, Shannon or differential entropy...)

Problem formulation

The design has to be chosen before observing the experiment and one needs to maximise the expectation of the utility function with respect to θ and $\mathbf{y}$. Thus,

$$\mathbf{d}^* = \arg\max_{d \in \mathcal{D}} U(\mathbf{d}),$$

where

$$U(\mathbf{d}) = \int_{\Theta \times Y} u(\mathbf{d}, \theta, \mathbf{y}) p_d(\theta, \mathbf{y}) d\theta d\mathbf{y} = \int_{\Theta \times Y} u(\mathbf{d}, \theta, \mathbf{y}) p_d(\mathbf{y}|\theta) \pi(\theta) d\theta d\mathbf{y}.$$

Here

$$p_{\mathbf{d}}(\mathbf{y}|\theta) = \prod_{1 \le i < j \le n} \left[p(\rho(x_i, x_j); \theta)\right]^{e(x_i, x_j)} \left[1 - p(\rho(x_i, x_j); \theta)\right]^{1 - e(x_i, x_j)}.$$

Problem formulation

The design has to be chosen before observing the experiment and one needs to maximise the expectation of the utility function with respect to θ and $\mathbf{y}$. Thus,

$$\mathbf{d}^* = \arg\max_{d \in \mathcal{D}} U(\mathbf{d}),$$

where

$$U(\mathbf{d}) = \int_{\Theta \times Y} u(\mathbf{d}, \theta, \mathbf{y}) p_d(\theta, \mathbf{y}) d\theta d\mathbf{y} = \int_{\Theta \times Y} u(\mathbf{d}, \theta, \mathbf{y}) p_d(\mathbf{y}|\theta) \pi(\theta) d\theta d\mathbf{y}.$$

Here

$$p_{\mathbf{d}}(\mathbf{y}|\theta) = \prod_{1 \le i < j \le n} \left[p(\rho(x_i, x_j); \theta)\right]^{e(x_i, x_j)} \left[1 - p(\rho(x_i, x_j); \theta)\right]^{1 - e(x_i, x_j)}.$$

Problem formulation

The design has to be chosen before observing the experiment and one needs to maximise the expectation of the utility function with respect to θ and $\mathbf{y}$. Thus,

$$\mathbf{d}^* = \arg\max_{d \in \mathcal{D}} U(\mathbf{d}),$$

where

$$U(\mathbf{d}) = \int_{\Theta \times Y} u(\mathbf{d}, \theta, \mathbf{y}) p_d(\theta, \mathbf{y}) d\theta d\mathbf{y} = \int_{\Theta \times Y} u(\mathbf{d}, \theta, \mathbf{y}) p_d(\mathbf{y}|\theta) \pi(\theta) d\theta d\mathbf{y}.$$

Here

$$p_{\mathbf{d}}(\mathbf{y}|\theta) = \prod_{1 \le i < j \le n} \left[p(\rho(x_i, x_j); \theta)\right]^{e(x_i, x_j)} \left[1 - p(\rho(x_i, x_j); \theta)\right]^{1 - e(x_i, x_j)}.$$

Lyndley information measure and Kullback-Leibler divergence

Kullback-Leibler divergence between posterior and prior

$$D_{KL}(\pi(\theta|\mathbf{y}) \parallel \pi(\theta)) = \int_{\Theta} \log\left(\frac{\pi(\theta|\mathbf{y},\mathbf{d})}{\pi(\theta)}\right) p(\theta|\mathbf{y},\mathbf{d})\mathrm{d}\theta.$$

Averaging over Y gives the Lyndley information measure

$$\mathbb{E}_{\mathbf{y}}\left[D_{KL}(\pi(\theta|\mathbf{y}) \parallel \pi(\theta))\right] = \ldots = \mathbb{E}_{\mathbf{y},\theta}\left[\log\left(\frac{\pi(\theta|\mathbf{y},\mathbf{d})}{\pi(\theta)}\right)\right].$$

Lyndley information measure and Kullback-Leibler divergence

Kullback-Leibler divergence between posterior and prior

$$D_{KL}(\pi(\theta|\mathbf{y}) \parallel \pi(\theta)) = \int_{\Theta} \log\left(\frac{\pi(\theta|\mathbf{y},\mathbf{d})}{\pi(\theta)}\right) p(\theta|\mathbf{y},\mathbf{d})\mathrm{d}\theta.$$

Averaging over Y gives the Lyndley information measure

$$\mathbb{E}_{\mathbf{y}}\left[D_{KL}(\pi(\theta|\mathbf{y}) \parallel \pi(\theta))\right] = \ldots = \mathbb{E}_{\mathbf{y},\theta}\left[\log\left(\frac{\pi(\theta|\mathbf{y},\mathbf{d})}{\pi(\theta)}\right)\right].$$

Kullback-Leibler divergence: two views

Progressive experimental design

There is just an experimenter A who is in possession of $\pi(\theta)$ and is designing an experiment $\mathbf{d}$ to update his or her prior knowledge to $\pi(\theta|\mathbf{y},\mathbf{d})$in the most informative way:

$$U(\mathbf{d}) = \int_{\Theta}\int_{Y} \log \frac{\pi(\theta|\mathbf{y},\mathbf{d})}{\pi(\theta)} p(\mathbf{y}|\theta,\mathbf{d})\pi(\theta)\mathrm{d}\theta\mathrm{d}\mathbf{y} \longrightarrow \max_{\mathbf{d}\in\mathcal{D}}.$$

Using Kullback-Leibler divergence: two scenarios

Instructive experimental design

In contrast to the progressive design scenario, in the instructive case there is an experimenter A, holding a prior $\pi(\theta)$, and a better informed trainer B whose knowledge about the model parameter is summarised in a distribution $\pi^*(\theta)$. The purpose of B here is to design an experiment for A in order to increase his/her knowledge ($\pi(\theta) \longrightarrow \pi(\theta \mid y)$ using the existing superior (A's) knowledge $\pi^*(\theta)$:

$$U(\mathbf{d}) = \int_Y D_{\mathrm{KL}}\{\pi(\theta \mid y, d) \parallel \pi(\theta)\} p^*(y)\, \mathrm{d}y \longrightarrow \max_{\mathbf{d}\in\mathcal{D}},$$

where $p^*(y) = \int_\Theta p(\mathbf{y}|\theta, \mathbf{d})\pi^*(\theta)\mathrm{d}\theta$.

Maximisation of the expected utility

Müller (1999) discusses simulation based methods for optimal design, where the expected utility $U(d)$ is evaluated by Monte-Carlo simulation:

$$\hat{U}(d) = \frac{1}{N} \sum_{i=1}^{N} u(d, \theta_i, y_i),$$

where $\theta_i \sim \pi(\theta)$, and $y_i \sim p_d(y|\theta)$.

Alternatives: augmented modelling, simulated annealing, genetic algorithms, etc.

Maximisation of the expected utility

Müller (1999) discusses simulation based methods for optimal design, where the expected utility $U(d)$ is evaluated by Monte-Carlo simulation:

$$\hat{U}(d) = \frac{1}{N} \sum_{i=1}^{N} u(d, \theta_i, y_i),$$

where $\theta_i \sim \pi(\theta)$, and $y_i \sim p_d(y|\theta)$.

Alternatives: augmented modelling, simulated annealing, genetic algorithms, etc.

Maximisation of the expected utility

Augmented modelling

Define an artificial distribution

$$h(d, \theta, \mathbf{y}) \propto \log u(\mathbf{d}, \theta, \mathbf{y}) \pi(\theta) p(\mathbf{y}|\theta, \mathbf{d}) \pi(\theta).$$

Under h the marginal in $\mathbf{d}$ is proportional to the expected utility $U(\mathbf{d})$. So, construct an MCMC scheme with stationary distribution $h(d, \theta, \mathbf{y})$ and take a sample of $\mathbf{d}'s$—its histogram's profile should give us information about the maximum of $U(\mathbf{d})$.

Two dimensional optimal design problem

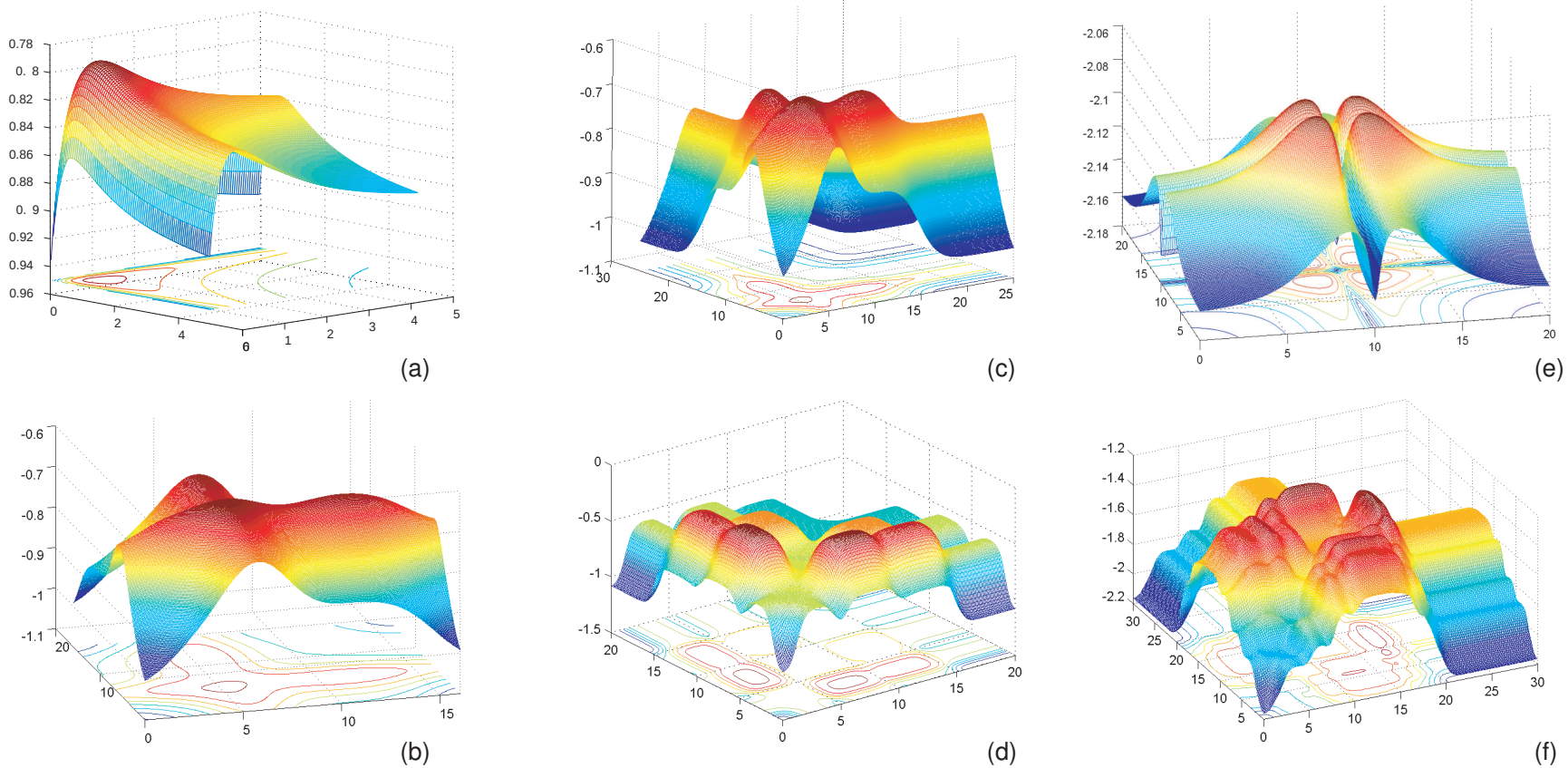


Figure: Expected utility plots for two independent random edges.

Geometric random graph case

Consider the following "step" probability edge function:

$$p(y|d,\theta) = y\mathbb{I}\{d \leq \theta\} + (1-y)\mathbb{I}\{d > \theta\}, \tag{4}$$

where y is either 0 (there is no edge) or 1 (there is an edge). Thus, the likelihood for (admissible) independent observations is either zero or one.

Theorem

The solution to the optimal design problem on a geometric random graph (4) where θ has a proper (prior) continuous increasing cdf F_θ on a non-negative support, is given by its quantiles of the order $n+1$.

Example

$$\theta \sim \mathrm{U}[0,1]$$

$$d_1 = \frac{1}{n+1}, d_2 = \frac{2}{n+1}, \ldots, d_i = \frac{i}{n+1}, \ldots, d_n = \frac{n}{n+1}.$$

Example

$$\theta \sim \mathrm{Exp}(\lambda)$$

$$d_1 = \frac{\log \frac{n+1}{n}}{\lambda}, d_2 = \frac{\log \frac{n+1}{n-1}}{\lambda}, \ldots, d_n = \frac{\log (n+1)}{\lambda}.$$

Example

$$\theta \sim \mathrm{U}[0,1]$$

$$d_1 = \frac{1}{n+1}, d_2 = \frac{2}{n+1}, \ldots, d_i = \frac{i}{n+1}, \ldots, d_n = \frac{n}{n+1}.$$

Example

$$\theta \sim \mathrm{Exp}(\lambda)$$

$$d_1 = \frac{\log \frac{n+1}{n}}{\lambda}, d_2 = \frac{\log \frac{n+1}{n-1}}{\lambda}, \ldots, d_n = \frac{\log(n+1)}{\lambda}.$$

Cook *et al.* (2008) use the derived result to provide a trivial example for showing that the sequential optimal design of replicated experiments need not be the same as the optimal design of simultaneous replicated experiments.

Reference: Cook, A. R., Gibson, G. J., Gilligan, C. A. (2008) Optimal observation times in experimental epidemic processes. *Biometrics*, **64**(3), pp. 860-868.
Web Appendices available at
`http://www.biometrics.tibs.org/datasets/070104.pdf`

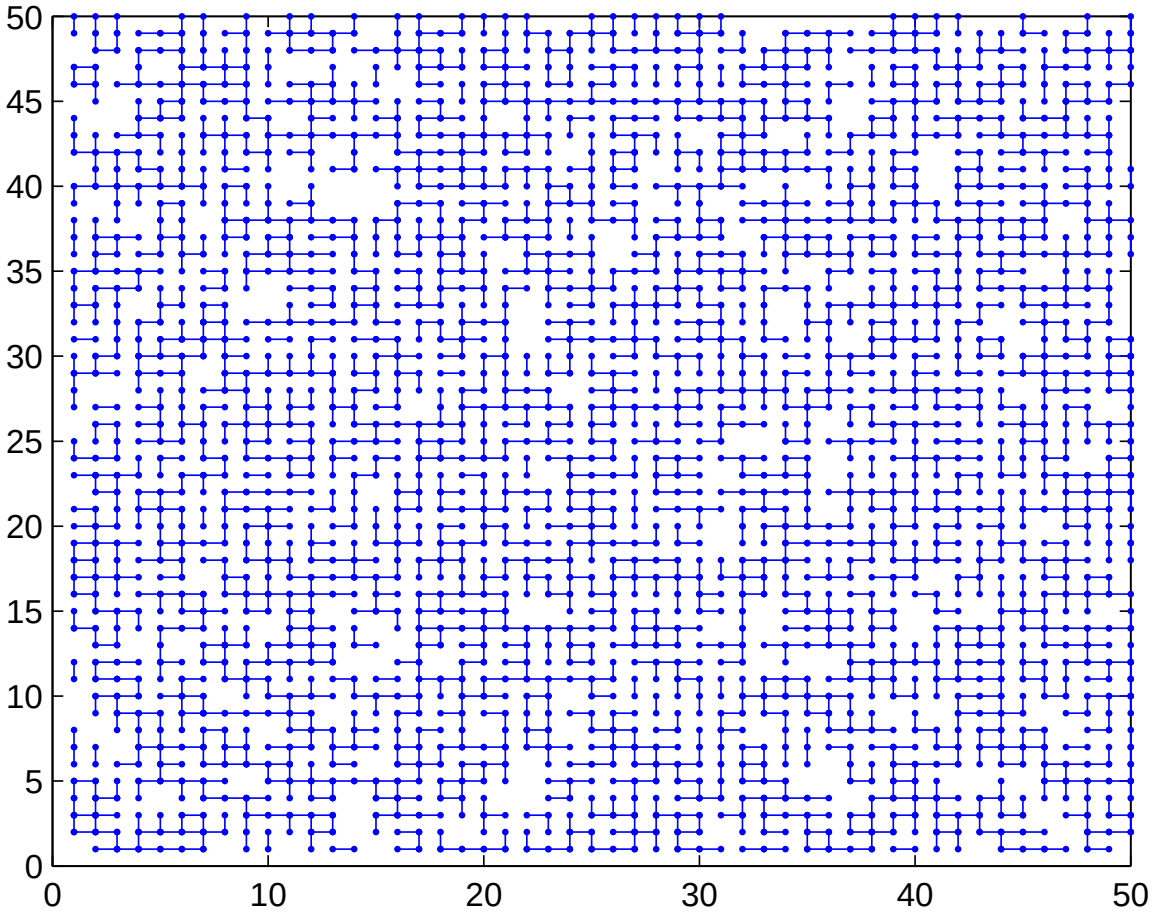


Figure: Bond percolation process in $\mathbb{Z}^2$, $p = \frac{1}{2}$.

Suppose we are given an extinct (due to natural extinction or to the boundness of the plot, or both) simple epidemic with constant life-times and some infection spread rate λ evolved in $\Pi \subseteq \mathbb{Z}^2$. In other words, suppose that we are given a set of n sites $\mathcal{C}$ from a connected component resulted from a bond percolation with $p = 1 - e^{-\lambda/4}$. The purpose is to infer on p.

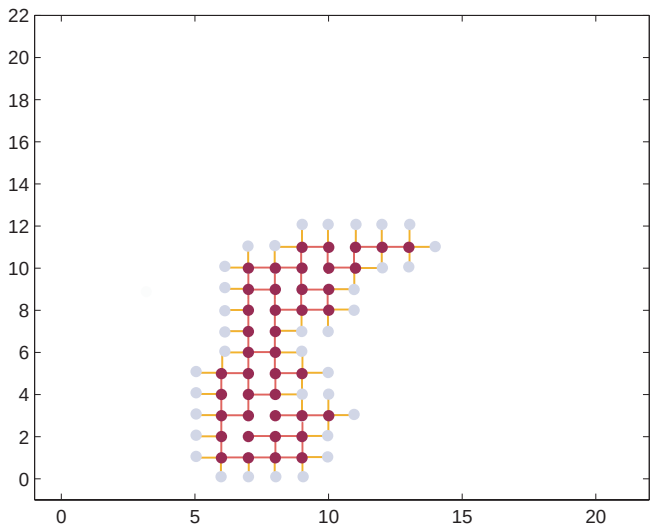


Figure: A connected component on $\mathcal{C}$ in $\mathbb{Z}^2$, $p = 0.478$. Infected sites are brown, their nearest heighbours are gray.

Suppose we are given an extinct (due to natural extinction or to the boundness of the plot, or both) simple epidemic with constant life-times and some infection spread rate λ evolved in $\Pi \subseteq \mathbb{Z}^2$. In other words, suppose that we are given a set of n sites $\mathcal{C}$ from a connected component resulted from a bond percolation with $p = 1 - e^{-\lambda/4}$. The purpose is to infer on p.

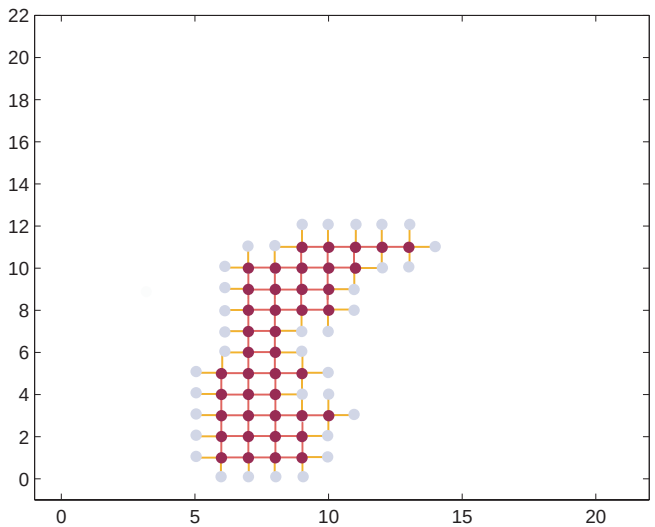
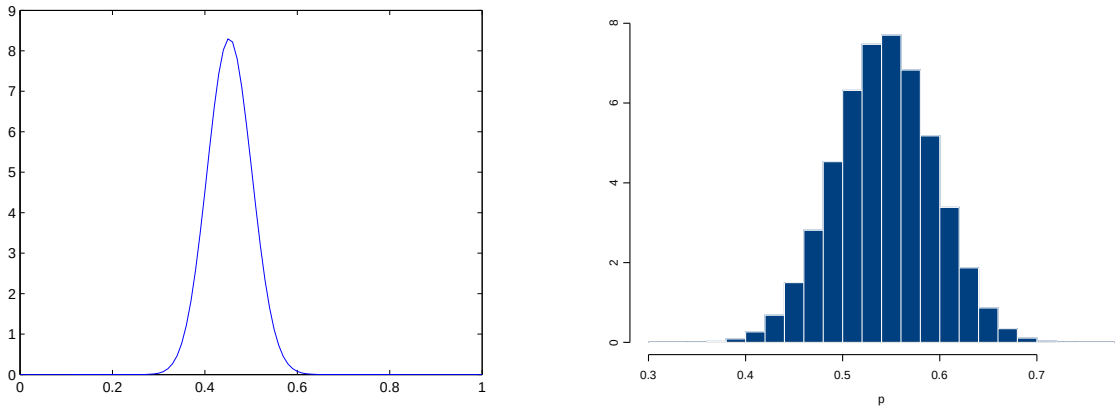


Figure: The "fully saturated" graph $S_{\mathcal{C}}$ derived from $\mathcal{C}$, $p = 0.478$. Infected sites are brown, their nearest heighbours are gray.

Algorithm 1 Sites are visible, edges are not: MCMC steps.

Require: $\mathcal{C}$, current Markov chain state $X_{2t} = (p_{2t}, G_{2t})$.
Ensure: a sample from the likelihood $f(p|\mathcal{C})$.

Gibbs step: $p_{2t+1} \sim \text{Beta}(e(G_{2t})+1, e(S_{\mathcal{C}})-e(G_{2t})+m+1)$
$X_{2t+1} := (p_{2t+1}, G_{2t})$
Metropolis-Hasting steps: choose $e_{S_G} \sim \text{Uniform}(S_G)$
if $e_{S_G} \in G_{2t}$ **then**
    if $\mathbb{I}\{G_{2t+1} \backslash \{e_{S_G}\}$ is connected$\}$ **then**
        $X_{2t+2} := (p_{2t+1}, G_{2t+1} \backslash \{e_{S_G}\})$ w.p. $\min\left(1, \frac{1-p_{2t+1}}{p_{2t+1}}\right)$
    else
        $X_{2t+2} := (p_{2t+1}, G_{2t+1})$
    end if
else
    $X_{2t+2} := (p_{2t+1}, G_{2t+1} \cup \{e_{S_G}\})$ w.p. $\min\left(1, \frac{p_{2t+1}}{1-p_{2t+1}}\right)$
end if



(a) Likelihood (b) Posterior

Figure: Likelihood for full information about the epidemic from the earlier slide and $\pi(p|\mathcal{C})$ for the case when we only know the infected sites.

The epidemic's size is known, the configuration is not

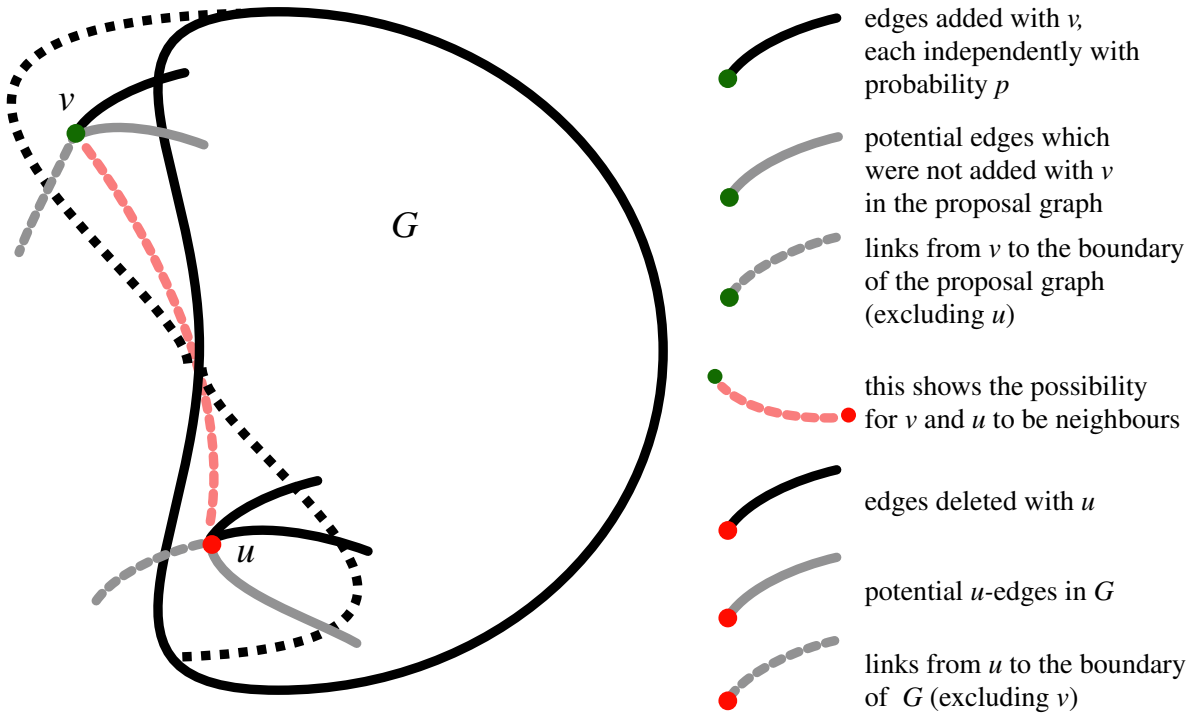


Figure: Updating connected component: graphical representation of the Metropolis-Hastings step of the MCMC algorithm

Theoretical limit results

Let $P_p(|\mathcal{C}| = n)$ be the probability that an open cluster is of size n in percolation with the edges density p. This is just the likelihood function $\mathcal{L}_n(p)$ for p:

$$\mathcal{L}_n(p) = P_p(|\mathcal{C}| = n).$$

Let $\hat{p}_n$ be the mle for p derived from $\mathcal{L}_n(p)$.

Theorem

The sequence of maximum likelihood estimators $\hat{p}_n$ for p converges to the critical probability p_c ($n \to \infty$).

Conjecture

$\pi(p|n) \propto \mathcal{L}_n(p)\pi(p) \to \delta(p - p_c)$.

Theoretical limit results

Let $P_p(|\mathcal{C}| = n)$ be the probability that an open cluster is of size n in percolation with the edges density p. This is just the likelihood function $\mathcal{L}_n(p)$ for p:

$$\mathcal{L}_n(p) = P_p(|\mathcal{C}| = n).$$

Let $\hat{p}_n$ be the mle for p derived from $\mathcal{L}_n(p)$.

Theorem

The sequence of maximum likelihood estimators $\hat{p}_n$ for p converges to the critical probability p_c ($n \to \infty$).

Conjecture

$\pi(p|n) \propto \mathcal{L}_n(p)\pi(p) \to \delta(p - p_c)$.

Corresponding MCMC results

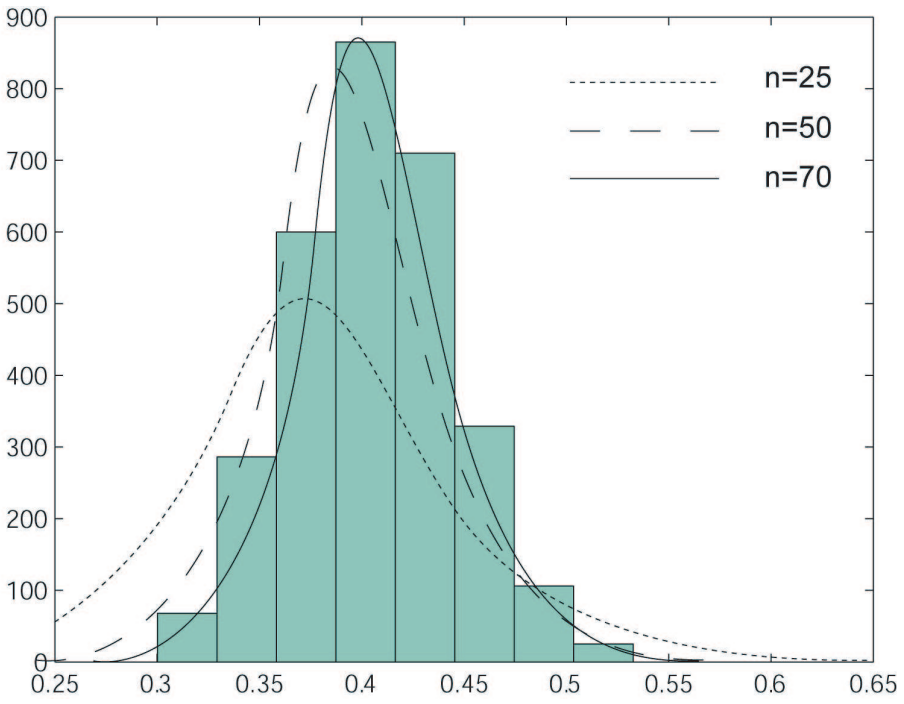


Figure: $\mathcal{L}_n(p)$ from MCMC

Optimal design for percolation process: sparsification of the grid

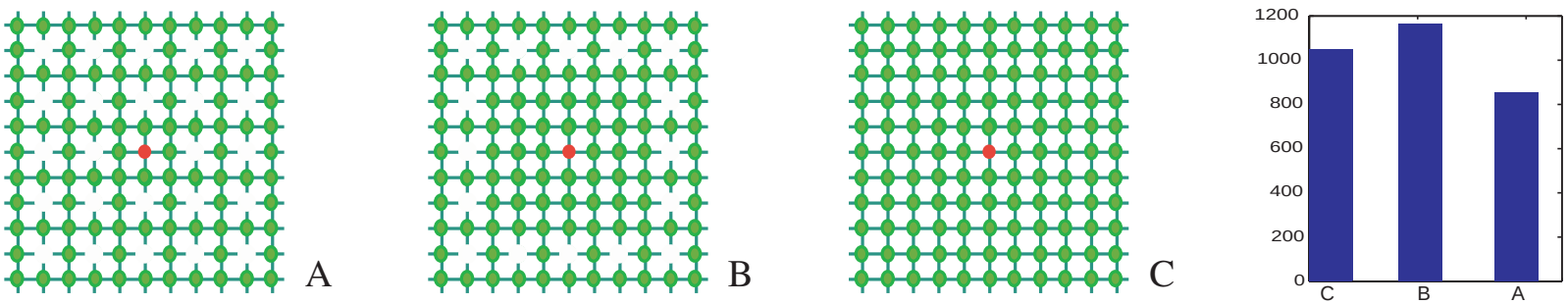


Figure: *Inner-outer* design plots A, B, and C, and marginal $h(d, p, y)$ in d, $d \in \{A, B, C\}$.

Instructive design: the fewer — the better!

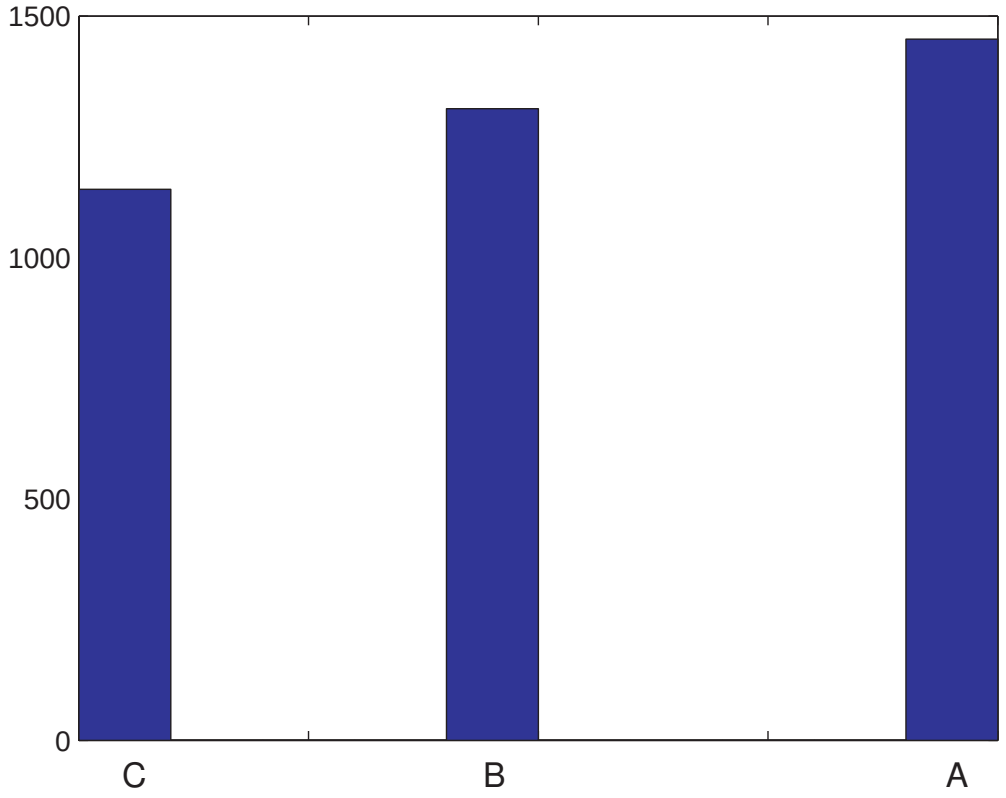


Figure: Optimal instructive design; here $\pi(p) = \mathbb{I}(p \in (0,1))$, and the true value of p was taken to be 0.9; N=6.

Rectangular, triangular, hexagonal ... grids

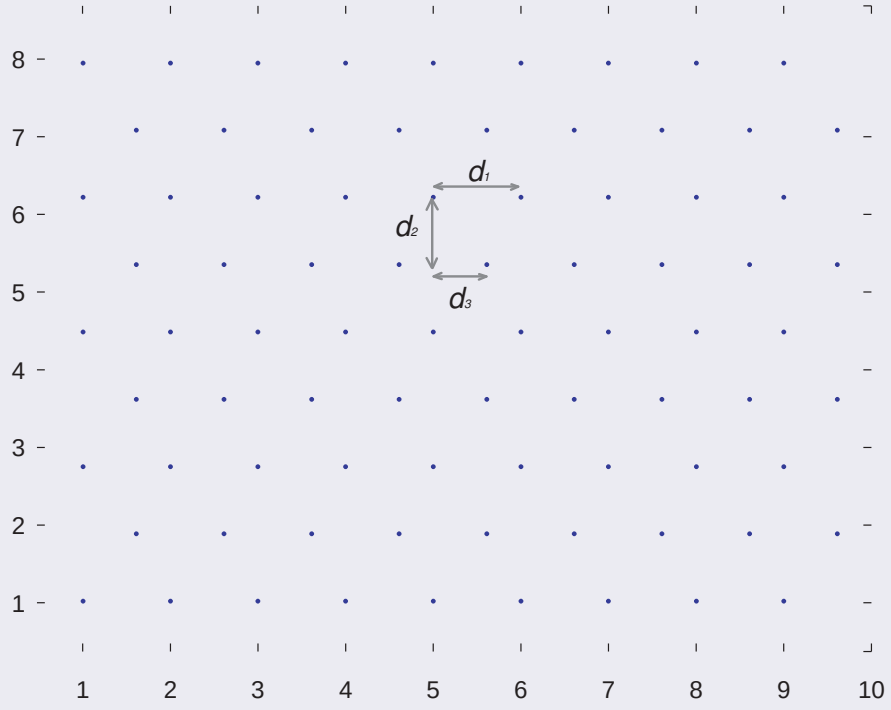


Figure: Reducing dimensionality of the design space.

Ongoing joint work with Alex Cook...

Conclusions

MCMC algorithms for inference in nearest-neighbour interaction models under incomplete observations are suggested. These algorithms can be extended to long range interaction processes and can be used for solving utility based optimal design problems via augmented probability simulation or simulated annealing.

Further work

- Optimal lattice shapes.
- Optimal arrangement of the nodes in a metric space in general settings: such problem remains to be a difficult computational task. However, grid approximation is possible and the methodology of genetic MCMC algorithms can be applied.
- Classification of the models and utility choices that give rise to unimodal expected utility surfaces: seems difficult.
- Model robustness issues (+designing experiments for checking the model assumptions?)

Bibliography

- Atkinson, A.C., Donev, A.N. (1992) *Optimum Experimental Designs*. Clarendon Press, Oxford.
- Bailey, D.J., Otten, W., Gilligan, C.A. (2000) Saprotrophic invasion by the soil-borne fungal plant pathegen *Rhizoctonia solani* and percolation thresholds. *New Phytol.* 146, 535-544.
- Bailey, D.J., Gilligan, C.A. (1997) Biological control of pathozone behaviour and disease dynamics of *Rhizoctonia solani* by *Trichoderma viride*. *New Phytol.* **136**, pp. 359-367.
- Cook, A.R., Gibson, G.J., Gilligan, C.A. (2007) Optimal observation times in experimental epidemic processes. *Biometrics*, DOI:10.1111/j.1541-0420.2007.00931.x.

- Firth, D., Hinde, J.P. (1997) Parameter neutral optimum design for non-linear models. *J.R.Statist. Soc. B*, **59**(4), pp. 799-811.
- Müller, P. (1999) Simulation-based Optimal Design. In *Bayesian Statistics 6*, J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (Eds.) Oxford University Press, pp. 459–474.
- Newman, M. (2000) The power of design. *Nature*, **405**.
- Read, J.M., Keeling, Matt, J. (2003) Disease evolution on networks: the role of contact structure. *Proc. R. Soc. Lond. B*, **270**, pp. 699–708.