Measures everywhere

Applications

Sergei Zuyev

University of Strathclyde, Glasgow, U.K.

Applications already considered

- Estimation of mixture distribution
- Generalisations of Kiefer-Wolfowitz theorem in optimal design
- Russo’s Formula and Gamma-type results in stochastic geometry
- Numeric integration of functions and Approximation of convex bodies

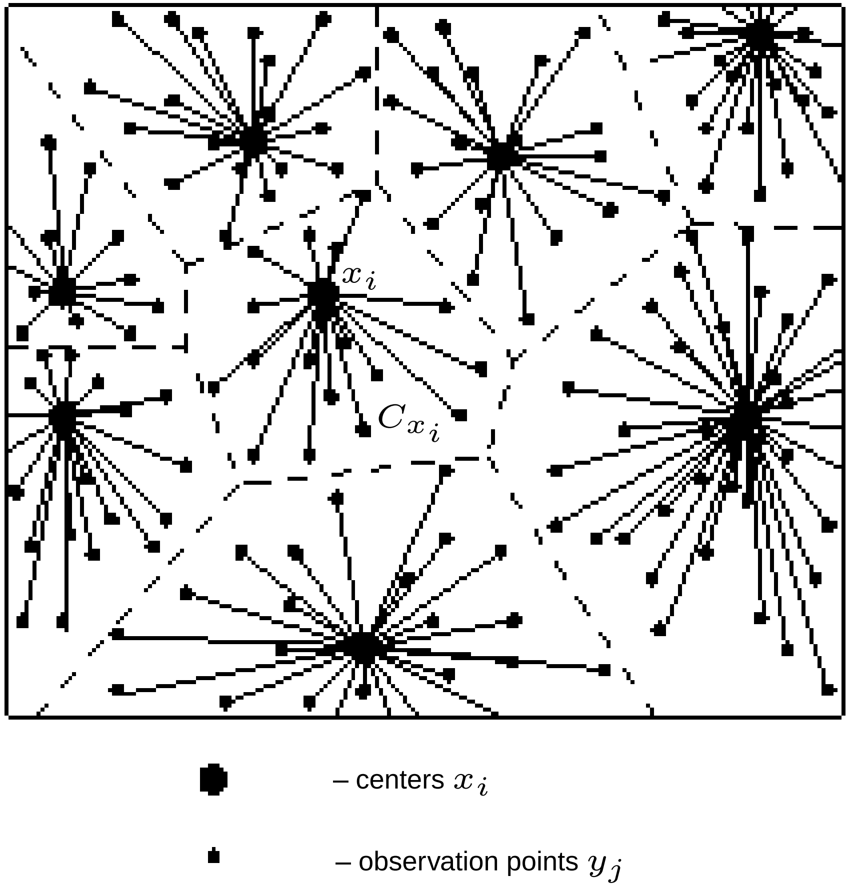
Clustering

Data points $y_j \in X$,

$1 \leq j \leq n$.

Ward criterion: find clusters' centres: $\mathbf{x} = (x_1, \ldots, x_k)$ minimising

$$\sum_{x_i} \sum_{y_j \sim C_{x_i}(\mathbf{x})} \rho^2(x_i, y_j)$$



Note: the objective function is non-convex w.r.t. $x_1, \ldots, x_k$ (k-means)!

Poissonisation

Cluster centres: Poisson process Π_μ with intensity μ with $\mu(X) = k$

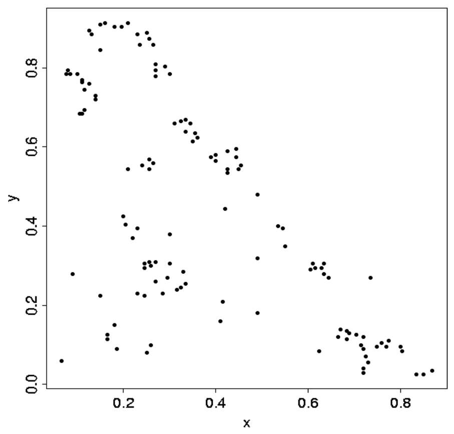
$$\mathbf{E}_\mu \left[\sum_{x_i \in \Pi_\mu} \sum_{y_j \sim C_{x_i}(\Pi_\mu)} \rho^2(x_i, y_j) \right] = \sum_{y_j} \mathbf{E}_\mu [\rho^2(y_j, \Pi_\mu)]$$

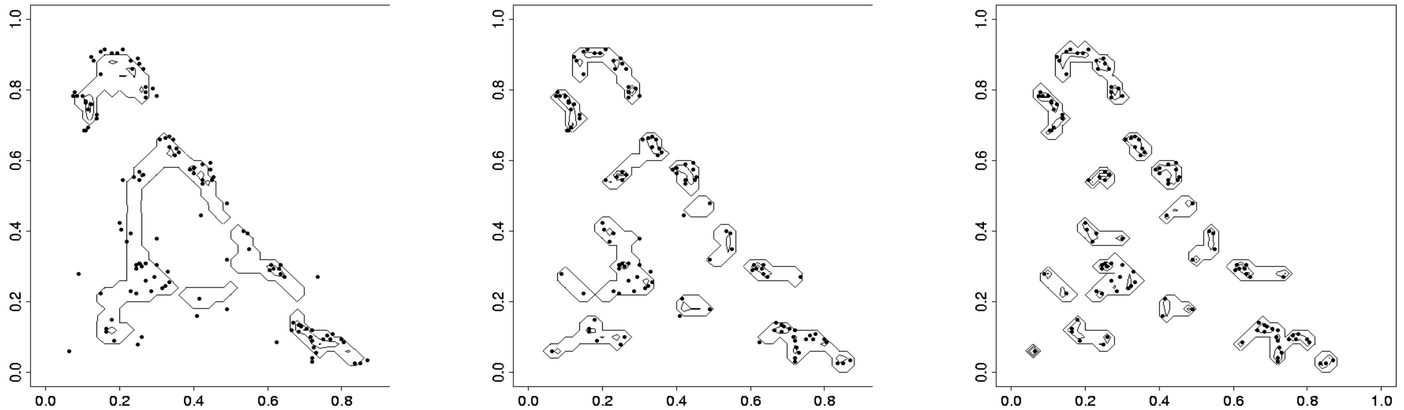
$$= \sum_{y_j} \int \exp\{-\mu(b_{\sqrt{t}}(y_j) \cap X)\} dt$$

Note: The objective function is strictly convex w.r.t. μ!

Redwood data

Data and level sets for optimal μ for total mass $n = 20, 50, 100$





Telecommunications example

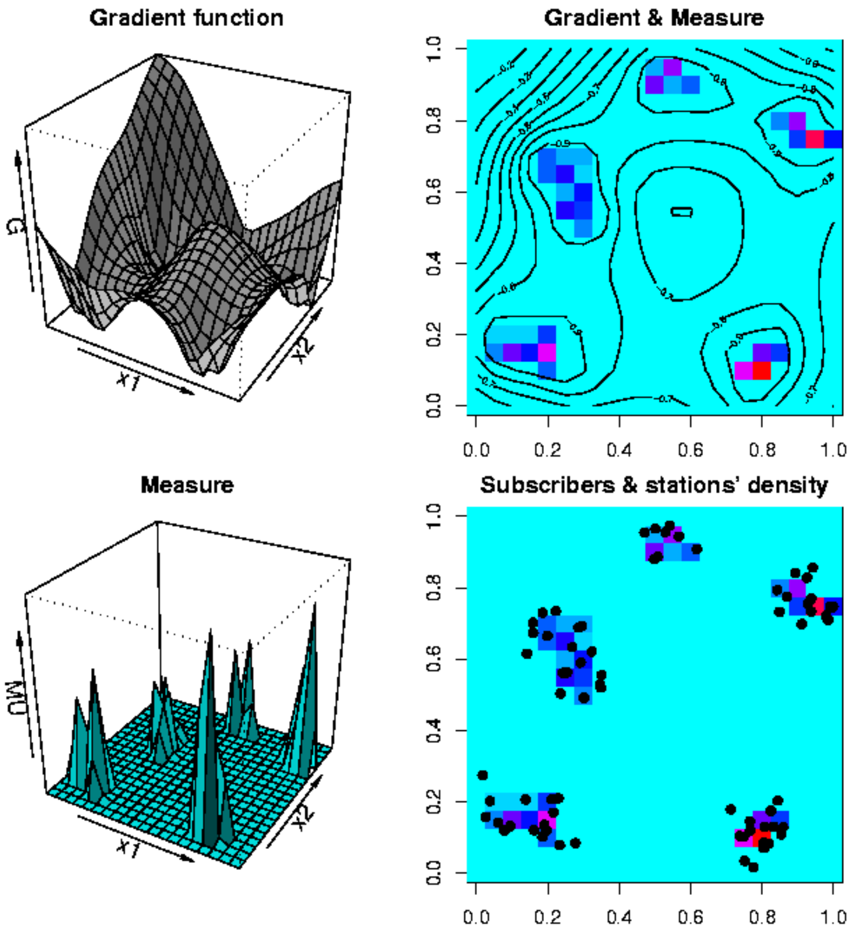
- Daughter points $\equiv$ subscribers (or demand).
- Cluster centers $\equiv$ local exchanges (stations)

Problem: Find density μ of stations minimising the average connections cost of subscribers to the stations:

$$\mathbf{E}_\mu \sum_{x_i} \sum_{y_j \sim C_{x_i}(\Pi)} \rho^\beta(x_i, y_j)\,.$$

❑ **High intensity solution:** Density of stations $p_\mu(x) \propto q(x)^{d/(d+\beta)}$, where q is the density of the demand ($d = 2$ typically)

Optimal placement of stations



Monte Carlo integration

Aim: calculate $\int_X f(y)dy$, $X \subset \mathbb{R}^d$

$$\int_X f(y)dy \approx [f(U_1) + \cdots + f(U_n)]/n = I_n \, ,$$

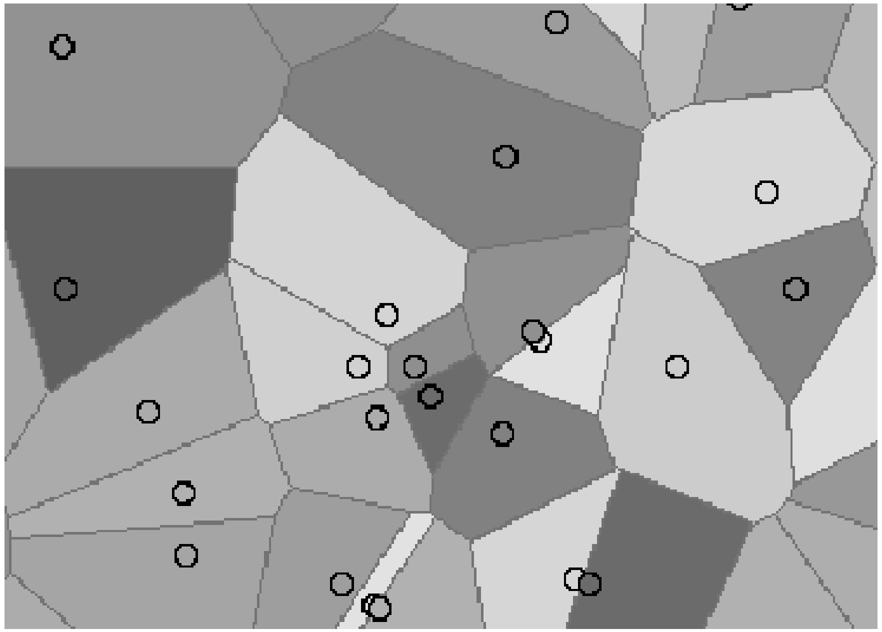
$$\mathrm{Var} I_n = \frac{1}{n}\varphi(X)$$

where $\varphi(X) = \ell(X) \int_X f(y)^2 dy - \Big[\int_X f(y)\, dy\Big]^2$

Stratification: split X into k sub-regions and

sample n/k points from every sub-region

Find optimal stratification



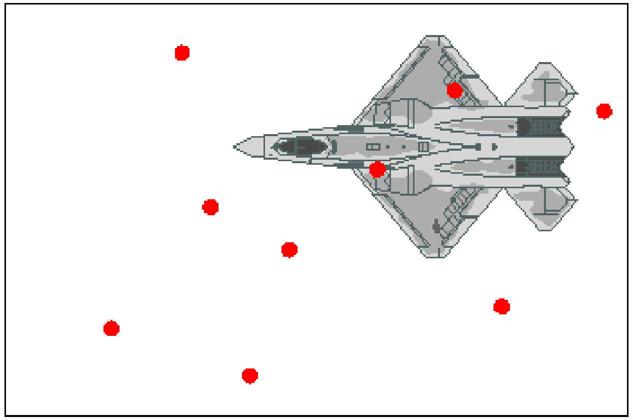
Variance (stratified case): is k/n times $F(\Pi) = \sum_{x_i \in \Pi} \varphi(C_{x_i}(\Pi))$.

High intensity solution:

$$p_\mu(x) \propto \|\mathrm{grad} f\|^{d/(d+1)}$$

Optimal random search

(How to catch a random set using Poisson traps?)



- $Y \subset X$ is a random closed set in-dependentdnt of Π.
- **Maximise trapping probability** $\mathbf{P}_\mu\{Y \cap \Pi \neq \emptyset\}$
- $\overline{\Delta}_\mu(x) = \mathbf{E}_\mu[e^{-\mu(Y)}\, \mathbb{1}_{x\in Y}]$

Example: $X = \{0, 1, 2, \ldots\}$

$Y = \{\xi\}$ geometrically distributed random singleton: $\mathbf{P}\{\xi = \{i\}\} = pq^i$

If $\mu(\{i\}) = m_i$, then (maximisation!)

$$\overline{\Delta}_\mu(x) = e^{-m_i} pq^i \begin{cases} = u & i \in \operatorname{supp}\mu, \\ \leq u & \forall i \end{cases}$$

Thus $m_i = -\log(u/(pq^i))$ on $\operatorname{supp}\mu$ and hence $\operatorname{supp}\mu$ is finite as otherwise m_i become negative.

- ❑ E.g., if $p = q = 0.5$ and $a = \mu(X) = 1$, then

$\operatorname{supp}\mu = \{0, 1\}$ with $m_0 = 0.847, m_1 = 0.153$,

trapping probability is 0.3211

- ❑ Compare:

0.5 = trapping probability using the fixed trap at 0 (not Poisson).

But trapping probability *given* $\Pi(X) > 0$ is $0.3211/(1 - e^{-1}) = 0.509$.

Catching a random ball

$X \subseteq \mathbb{R}^d$, $Y = b_\rho(\xi)$ random ball of radius ρ at $x \in X$

ξ and ρ are independent and have continuous densities

$$\overline{\Delta}_{ap(x)\lambda}(x) \propto -p_\xi(x) b_d d^{-1}$$

$$\times \left[\frac{p_\rho(0)(d+1)\Gamma(1+1/d)}{(ap(x)b_d)^{1+1/d}} + \frac{p'_\rho(0)(d+2)\Gamma(1+2/d)}{(ap(x)b_d)^{1+2/d}} + \cdots \right]$$

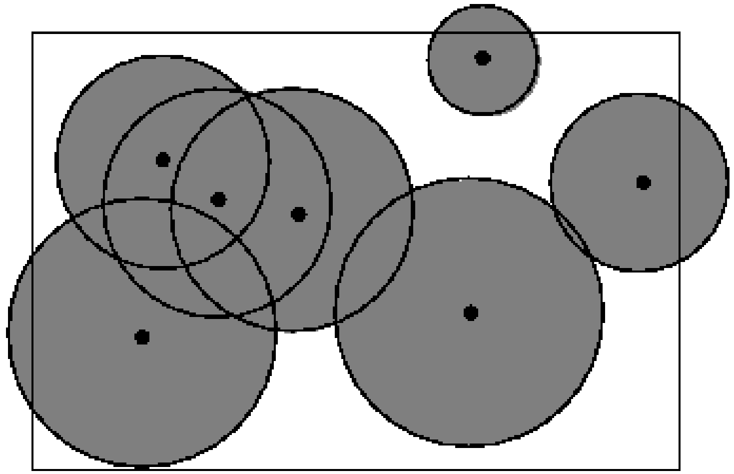
High intensity solution:

$p(x) \propto (p_\xi(x))^{d/(d+k+1)}$, where k is the first non-zero $p_\rho^{(k)}(0)$.

Design of materials

Boolean model:

$$\Xi = \bigcup_{x_i \in \Pi_\mu} (x_i + \Xi_i)$$



Π Poisson process with intensity measure μ,

Ξ_0 is a typical grain (e.g., $b_\xi(0)$).

Minimise the expected uncovered volume (convex function!):

$$\psi(\mu) = \mathbf{E}\,\mathrm{Vol}(X \setminus \Xi) = \int_X e^{-\,\mathbf{E}\,\mu(x-\Xi_0)} dx \mapsto \min$$

Gradient:

$$d(x,\mu) = -\,\mathbf{E}\left[\int_{\Xi_0+x} e^{-\,\mathbf{E}\,\mu(y-\Xi_0)} dy\right]$$

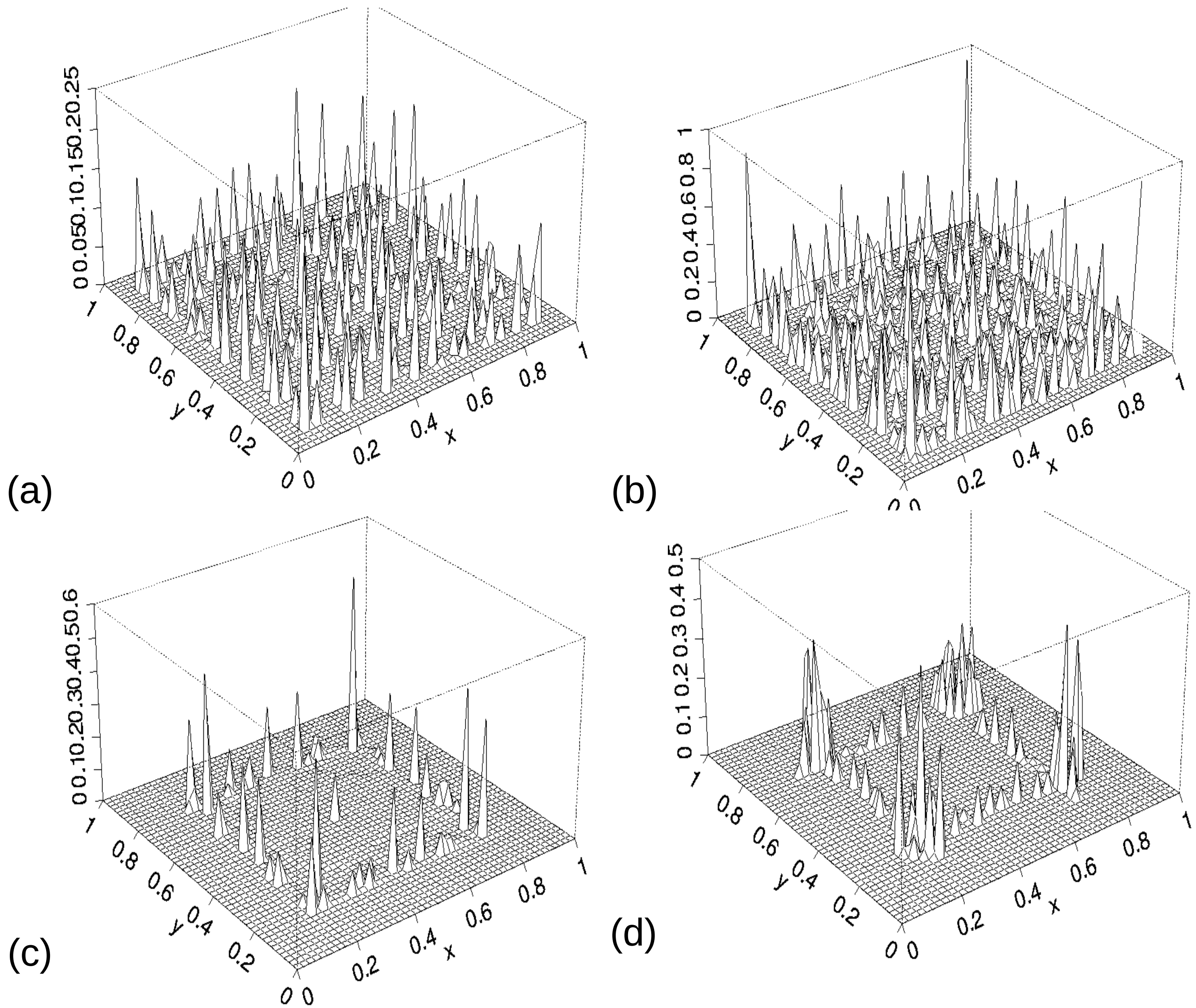
Measures maximising the expected covered area in $X = [0,1]^2$ with the fixed total mass a. The typical grain is a ball of radius r.

(a) $a = 10$, $r = 0.1$;

(b) $a = 50$, $r = 0.1$;

(c) $a = 10$, r is exponentially distributed with mean 0.1;

(d) $a = 10$, $r = 0.3$.


(a)
(b)
(c)
(d)
x
y

Other quantities

- Weighted volume $\mathbf{E}\,\Theta(\Xi\cap X)$

- Predetermined volume $\mathbf{E}\,\mathrm{Vol}(\Xi\cap X)=v$

Solve $f(\mu)=v$ or, equivalently, minimise

$$f_v(\mu)=(f(\mu)-v)^2$$

- Entropy of the uncovered phase

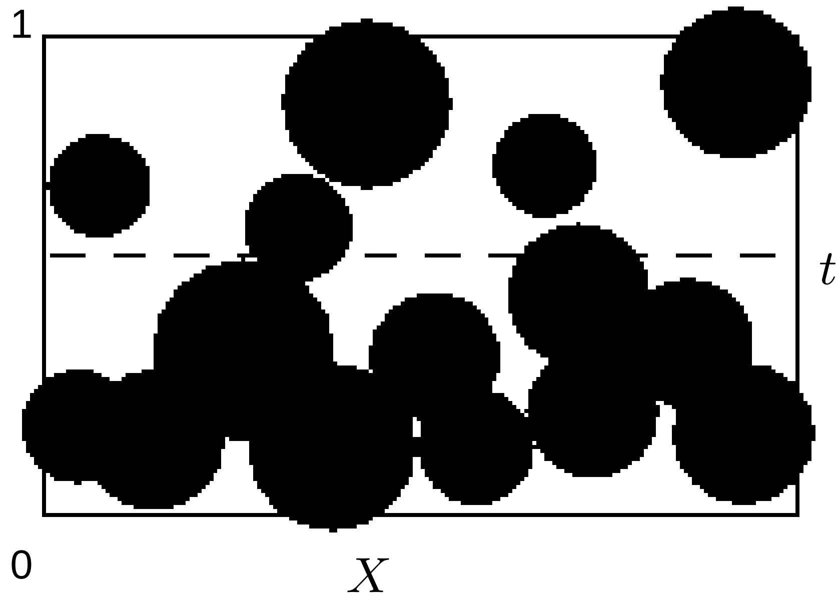
$$-\int_X (1-r(x))\log(1-r(x))dx$$

$$=\int_X e^{-\mathbf{E}\,\mu(x-\Xi_0)}\,\mathbf{E}\,\mu(x-\Xi_0)dx$$

where $r(x)=\mathbf{P}\{x\in\Xi\}$.

Functionally graded materials (FGM)

Ξ is a Boolean model in $X \times [0,1]$. The last coordinate (height) is used for grading



Expected uncovered volume at height t is called density profile

$$q(t,\mu) = \mathbf{E}_\mu \operatorname{Vol}(\Xi^c \cap (X \times \{t\}))$$
$$= \int_X \exp\{-\mathbf{E}\,\mu((x,t) - \Xi_0)\} dx\,.$$

Design of FGM

Assume: $\mu = dx \times \nu(dt)$ is homogeneous on X and $\Xi_0 = B_\xi(0)$

Aim: Given $h(t)$, design FGM (measure μ) such that

$$q(t,\mu) = h(t), \quad t \in [0,1] \quad \text{or}$$

$$\psi(\mu) = \int_0^1 (q(t,\mu) - h(t))^2 \nu(dt) \mapsto \min$$

$$q(t,\mu) = \exp\left\{ -\int_0^1 g(s,t)\nu(ds) \right\}$$

$$g(s,t) = b_d \,\mathbf{E}\left[\max(\xi^2 - (s-t)^2, 0)^{d/2}\right]$$

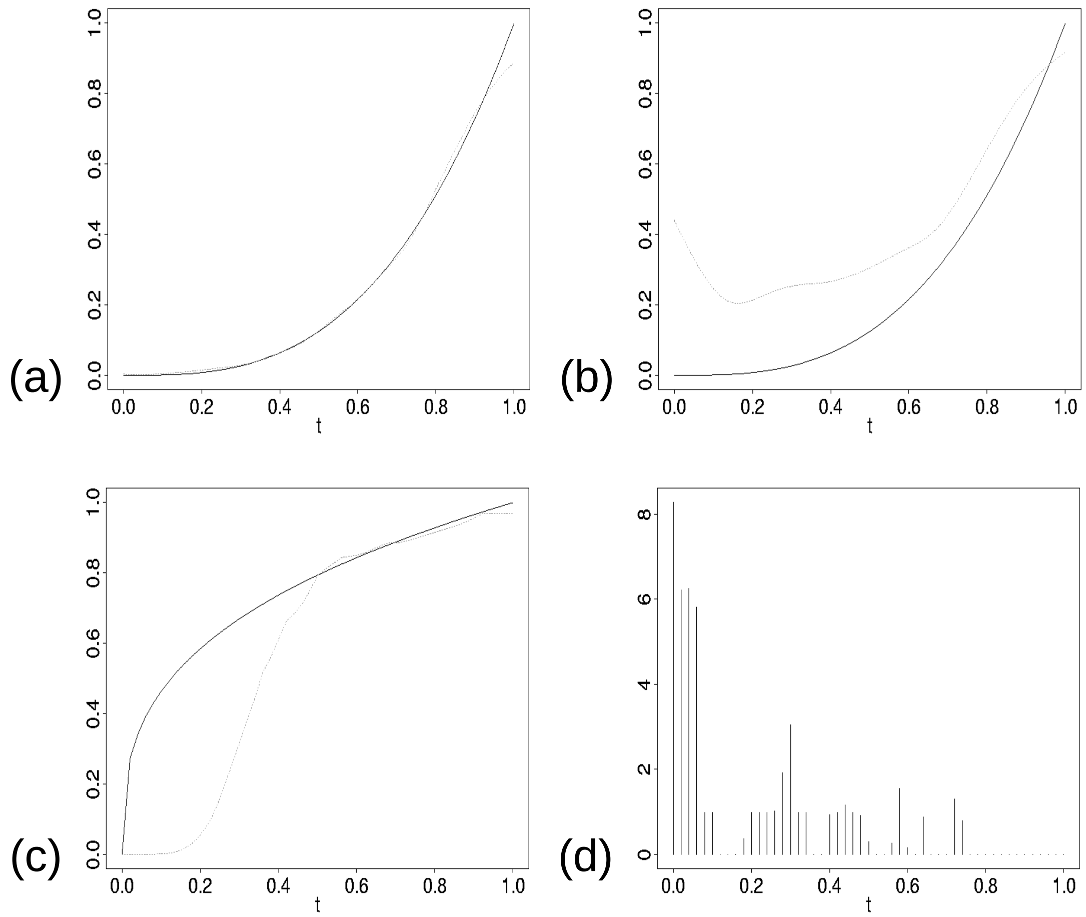
Target density profiles h (solid lines) and the calculated density profiles $q(t, \Lambda)$ (dashed lines) for optimal measures with a total mass 50.

(a) $d = 1$, $h(t) = t^3$;

(b) $d = 2$, $h(t) = t^3$;

(c) $d = 1$, $h(t) = t^{1/3}$;

(d) The optimal ν for the case (a).


(a)
(b)
(c)
(d)
t

Some subjects non-covered in the course

- Infinite mass measures
- Second order necessary conditions for $\inf$
- Specific constraints, e. g. class of absolutely continuous measures
- P-design measures
- Sequential Gamma-type results
- Hitting properties of stopping sets
- Projected gradient descent
- Other applications

References

- I. Molchanov and S. Zuyev. Variational analysis of functionals of a Poisson process. *Math. Oper. Research*, **25**, 485–508 (2000)
- I. Molchanov and S. Zuyev. Steepest descent algorithms in space of measures. Statistics and Computing, **12**, 2002, 115–123.
- I. Molchanov, S.N. Chiu, and S. Zuyev. Design of materials with given structural properties. *Phys. Rev. E*, **62**, 2000, 4544–4552
- M. van Lieshout, I. Molchanov and S. Zuyev. Clustering methods based on variational analysis in the space of measures. Biometrika, **88**, 2001, 1021–1033.

❑ http://www.stams.strath.ac.uk/~sergei