Inverse Theory - a modern method of data analysis

Lecture 9

W. Dębski

15.06.2021

Probabilistic point of view

inverse problem

$\Updownarrow$

inference process

$\Updownarrow$

joining available information

Joining information according to Tarantola

Two distributions describing different pieces of information about the same object

1. $p(x)$
2. $q(x)$

$$\mathbf{p} \wedge \mathbf{q}(\mathbf{x}) = \frac{\mathbf{p}(\mathbf{x})\,\mathbf{q}(\mathbf{x})}{\mu(\mathbf{x})}$$

$\mu(x)$ - non-informative probability

A posteriori pdf

$$\sigma(\mathbf{m},\mathbf{d}) = \frac{p(\mathbf{m},\mathbf{d})\, q(\mathbf{m},\mathbf{d})\, f(\mathbf{m},\mathbf{d})}{\mu^2(\mathbf{m},\mathbf{d})}$$

Marginal*A posteriori* distribution

$$\sigma_m(m) = \int_D \sigma(m,d) dD$$

$$\sigma_d(d) = \int_M \sigma(m,d) dM$$

A posteriori pdf

$$\sigma_m(m) = f(m) \cdot L(m, d^{obs})$$

$$L(\mathbf{m}, \mathbf{d}^{obs}) = \int_D p(\mathbf{d}, \mathbf{d}^{obs}) \frac{q(\mathbf{m}, \mathbf{d})}{\mu(\mathbf{m}, \mathbf{d})} \mathrm{d}\mathbf{d}$$

Probabilistic solution - features

The solution: *a posteriori* probability distribution

- always exists
- fully nonlinear
- include all uncertainties
- possible full error analysis
- physical well define meaning (and role) of *a priori* term
- requires methods of exploring $\sigma(\mathbf{m})$
- non-parametric inverse problems?

Exploring *a posteriori* probability

- searching for maximum of $\sigma(\mathbf{m})$
- calculate point estimators
- marginal distributions
- sampling $\sigma(\mathbf{m})$

Exploring *a posteriori* probability: $\mathbf{m}^{ml}$ solution

Basic characteristic of $\sigma(\mathbf{m})$:

location of the (global) maximum
- the most likelihood $\mathbf{m}^{ml}$ value

$$\mathbf{m}^{ml} : \quad \sigma(\mathbf{m}) = \max$$

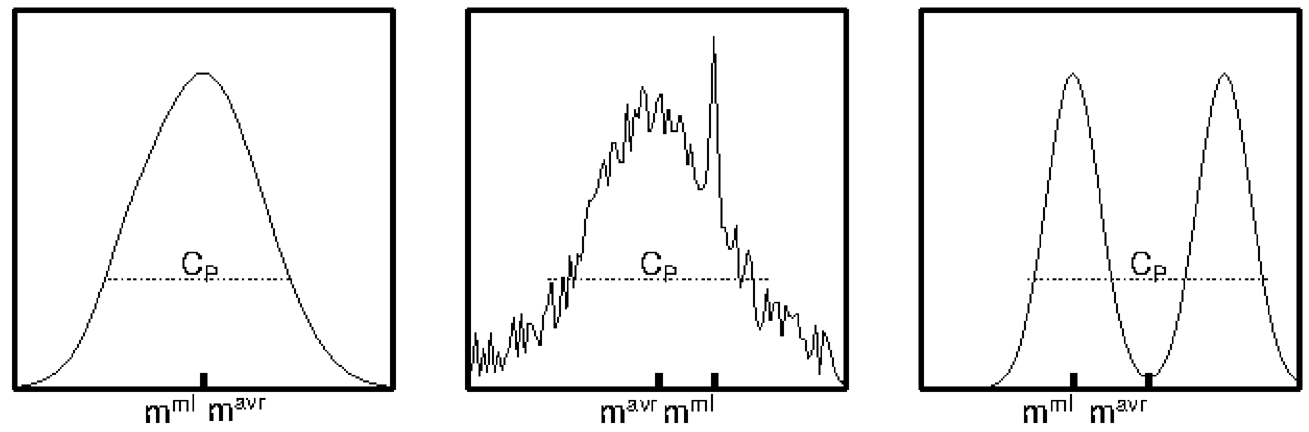
Problem reduced to optimization approach

Point estimators

Characterization of $\sigma(\mathbf{m})$ by its moments:

- average value: $\mathbf{m}^{avr} = \int_M \mathbf{m}\, \sigma(\mathbf{m}) \mathrm{d}\mathbf{m}$
- covariance $C_{ij} = \int_M (m_i - m_i^{avr})(m_j - m_j^{avr})\, \sigma(\mathbf{m}) \mathrm{d}\mathbf{m}$
- higher order moments

Point estimators



Marginal *a posteriori* distribution

- 1D marginals

$$\sigma_i(m_i) = \int_{\mathbf{m} \neq m_i} \sigma(\mathbf{m}) \, \mathrm{d}\mathbf{m}$$

- 2D marginals

$$\sigma_{ij}(m_i, m_j) = \int_{\mathbf{m} \neq m_i, m_j} \sigma(\mathbf{m}) \, \mathrm{d}\mathbf{m}$$

- higher dimension marginals

Require efficient methods of calculation multi-dimensional integrals

Inversion efficiency - Shanon's measure

$$\mathbf{m}^{apr} \Longrightarrow \mathbf{m}^{pos}$$

Information gain

$$I^{ef} = \int_M \sigma^{pos}(\mathbf{m}) \log\left(\frac{\sigma^{pos}(\mathbf{m})}{\sigma^{apr}(\mathbf{m})}\right) \mathrm{d}\mathbf{m}$$

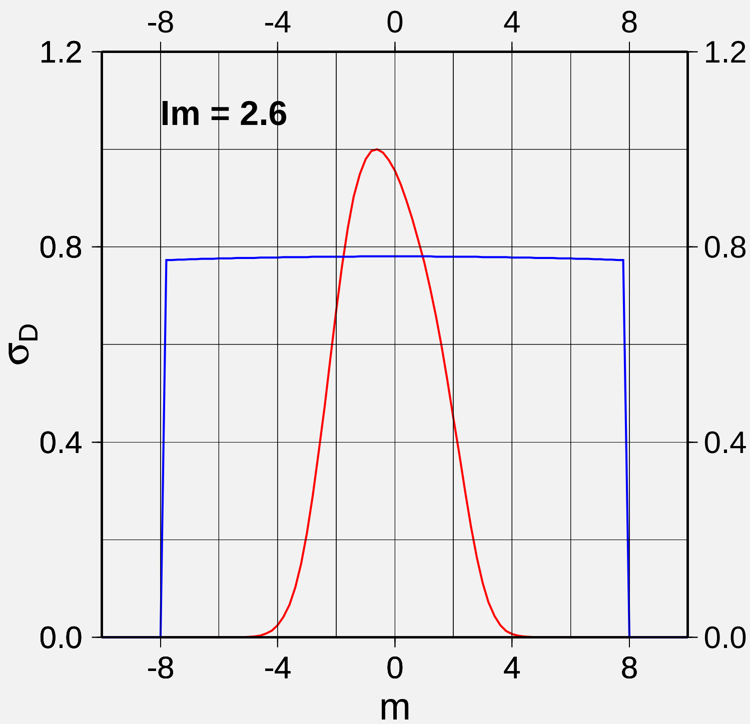
Non-parametric inversion - comparing alternative solutions

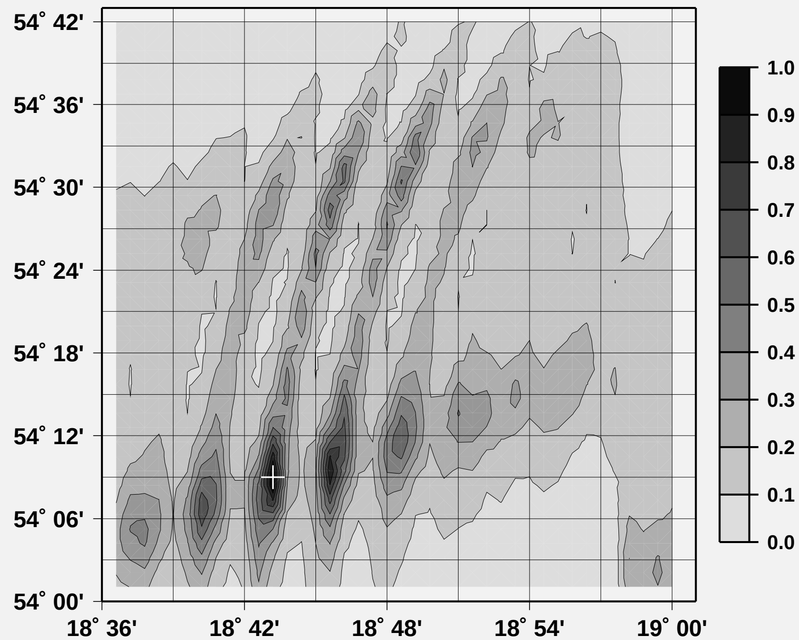
$$I^{12} = \int_M \sigma^{1}(\mathbf{m}) \log\left(\frac{\sigma^{1}(\mathbf{m})}{\sigma^{2}(\mathbf{m})}\right) \mathrm{d}\mathbf{m}$$

Markov Chain Monte Carlo sampler

Regular grid sampling

$N_m = 1$ $N_m = 2$

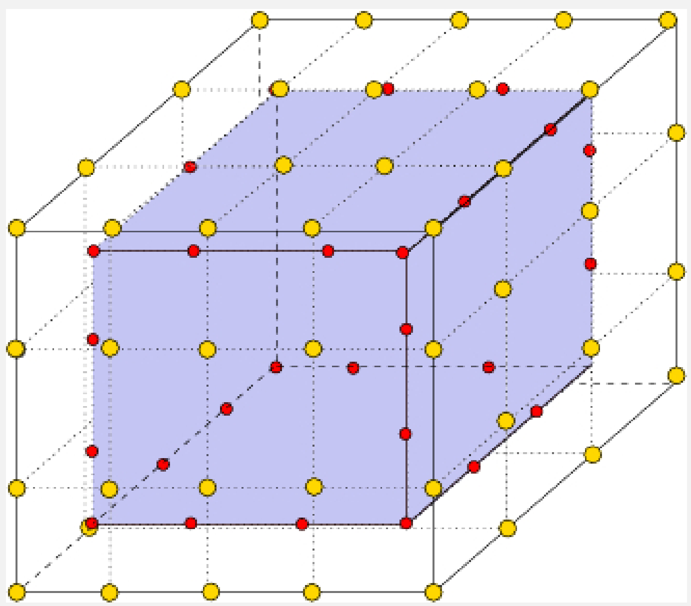




Regular grid sampling - curse of dimensionality problem

Nonuniform sampling:

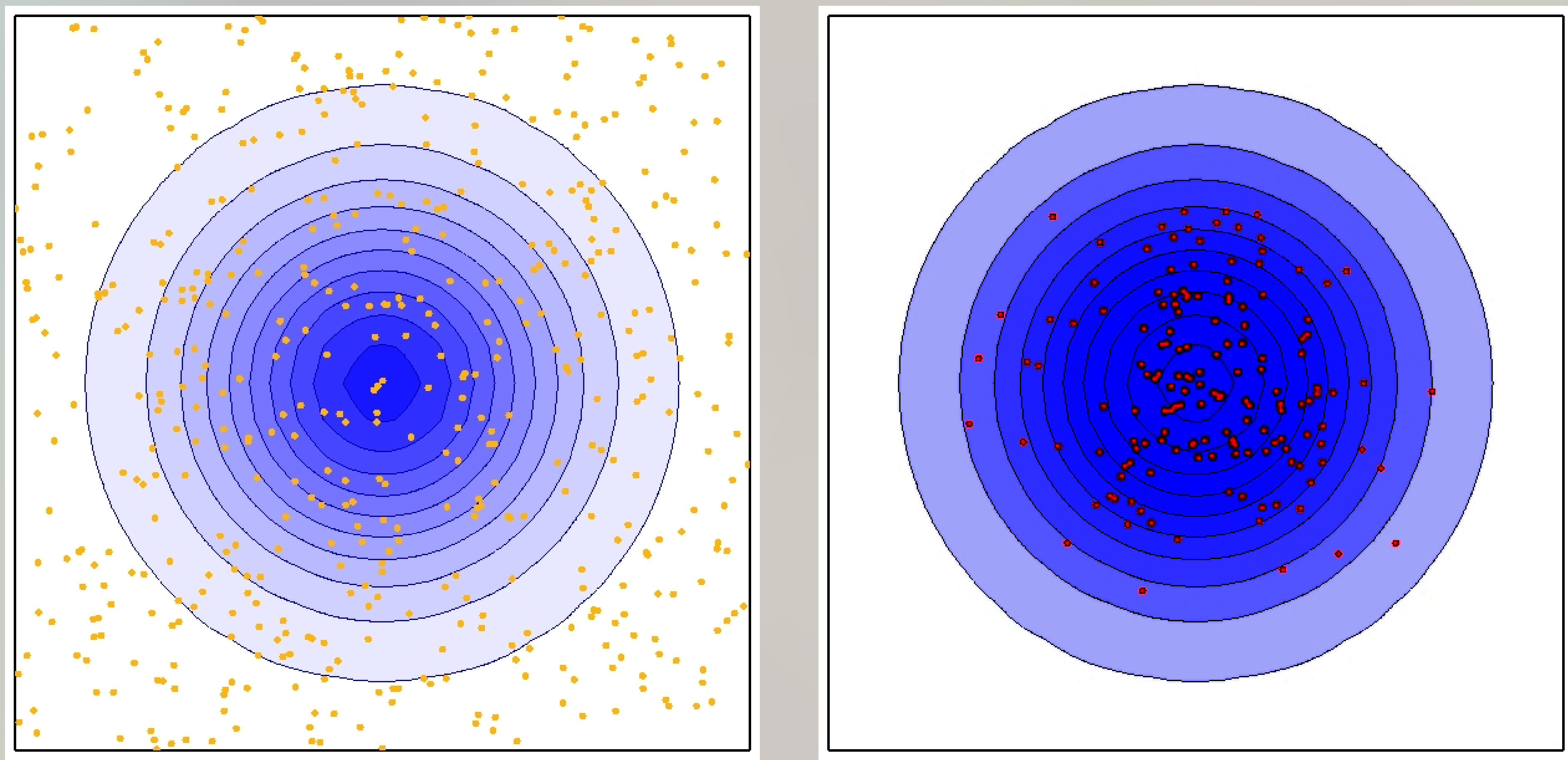
$$\frac{N_V}{N} = \left(\frac{p-2}{p}\right)^N \overset{p \gg 2}{\approx} e^{-2N/p} \xrightarrow{N} 0$$



Regular grid sampling - curse of dimensionality problem

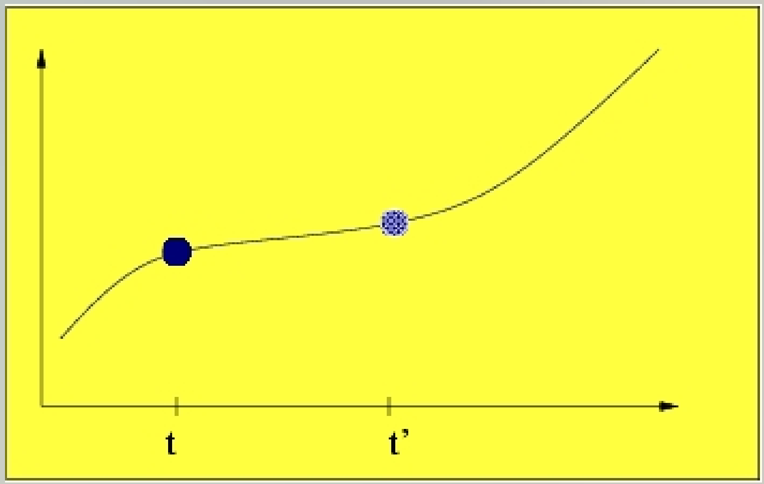
N/p	10	100	1000
2	0.64	0.96	0.996
3	0.51	0.94	0.994
5	0.33	0.90	0.990
10	0.10	0.82	0.980
100	$2.0\,10^{-10}$	0.13	0.818
1000	$1.2\,10^{-97}$	$1.6\,10^{-11}$	0.135
10000	—	$1.8\,10^{-88}$	$2\,10^{-9}$

Monte Carlo sampling

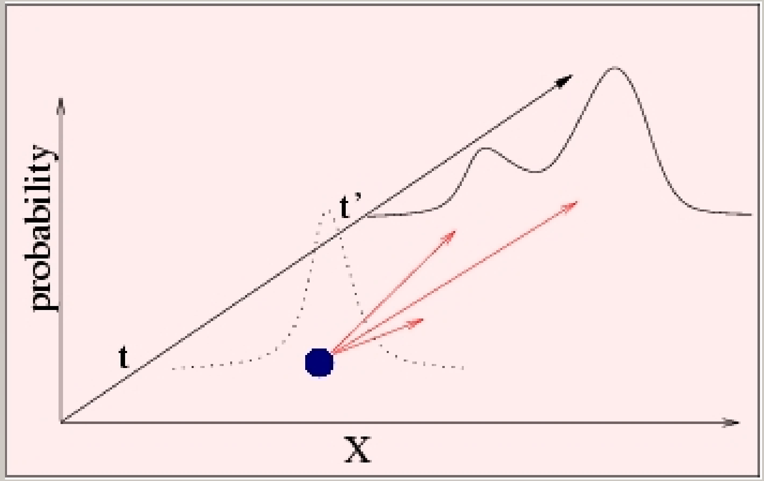


Stochastic processes (with short memory)

deterministic



stochastic



$$X(t+dt) = X(t) + f(X(t),t)dt \quad P(x',t') = \sum_x P(x,t) K(t',x';t,x)$$

Markow chains

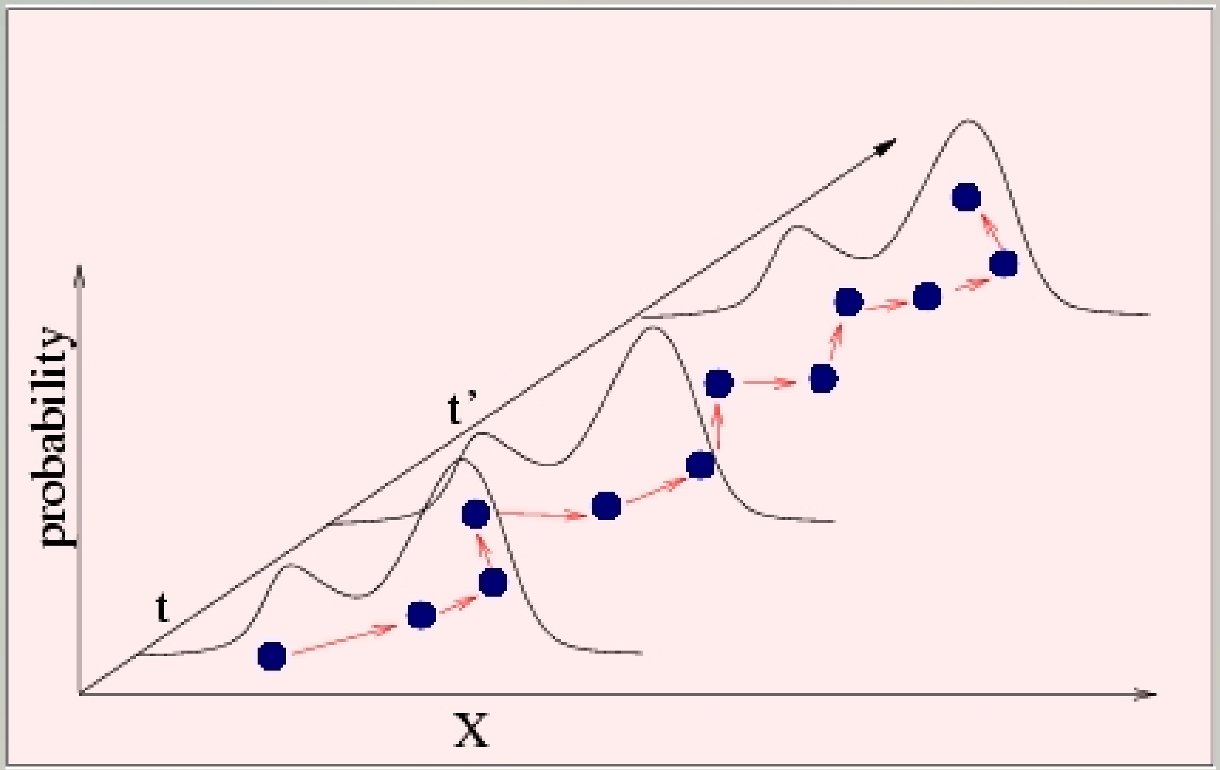
- discrete time: $t_1, t_2, \cdots$
- short memory $K(t_i, x; t_{i-1}, x_{i-1})$
- ergodic process:
  - stationary $K(t, x; t', x') = K(x, x')$
  - irreducible
  - aperiodic
- Stationary probability distribution: $\pi(x') = \pi(x)$

$$\pi(x) = \int_{x'} \pi(x') K(x; x') dx'$$

Stationary Markov Chain

$$p(x_i) = \sum_k p(x_k)\, K(x_i; x_k)$$

$$p(x) = \int_{x'} \pi(x') K(x; x') dx'$$

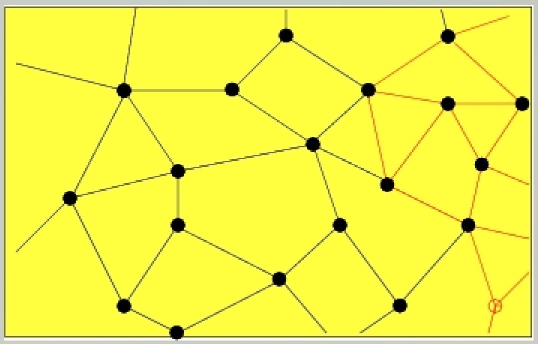


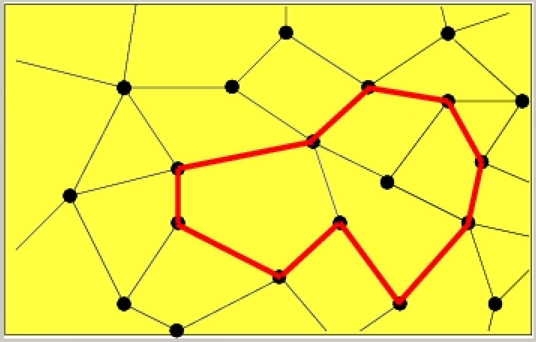
Markov chain - example

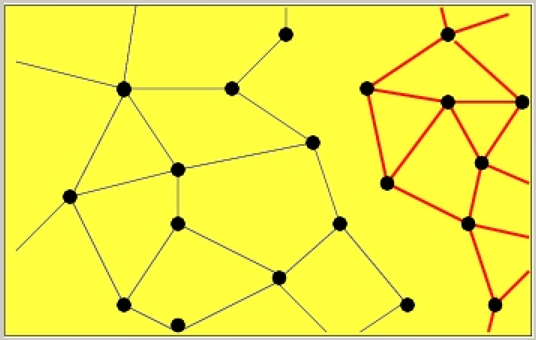
random walk

$$x_i \Longrightarrow x_{i+1} = x : \quad p_{x_i \rightarrow x_{i+1}} = K(x_{i+1}|x_i)$$

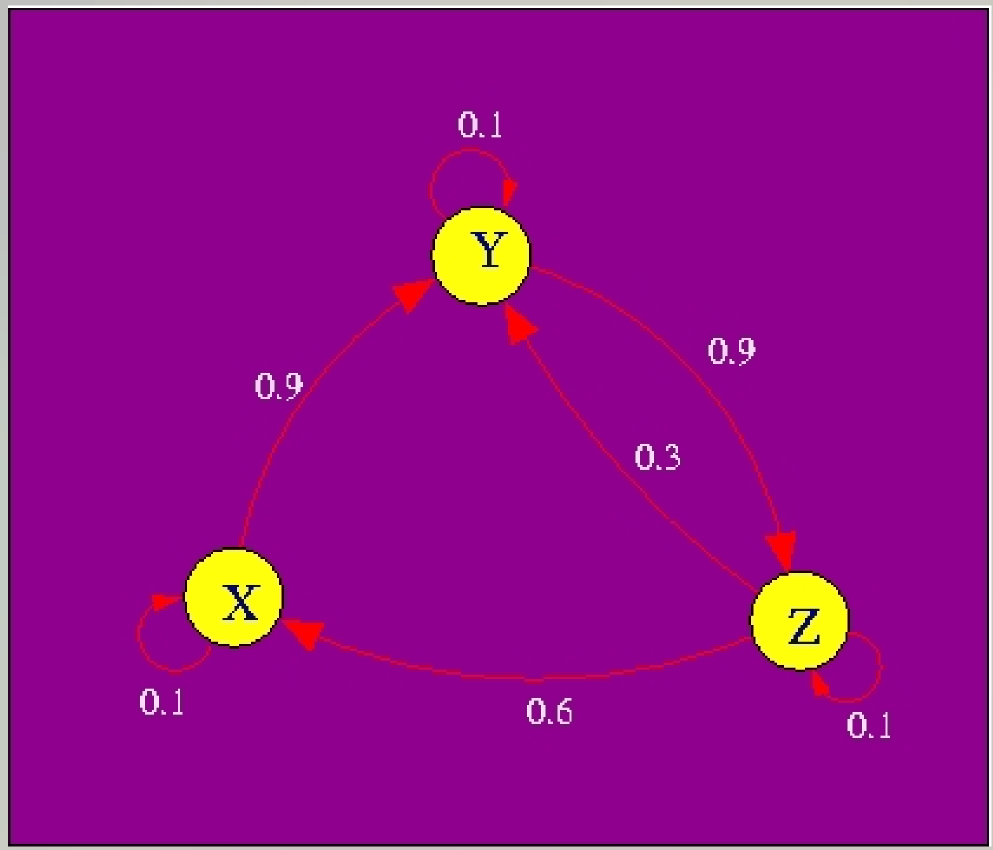
stationary aperiodic irrducible







Algorithm - example



Algorithm - example

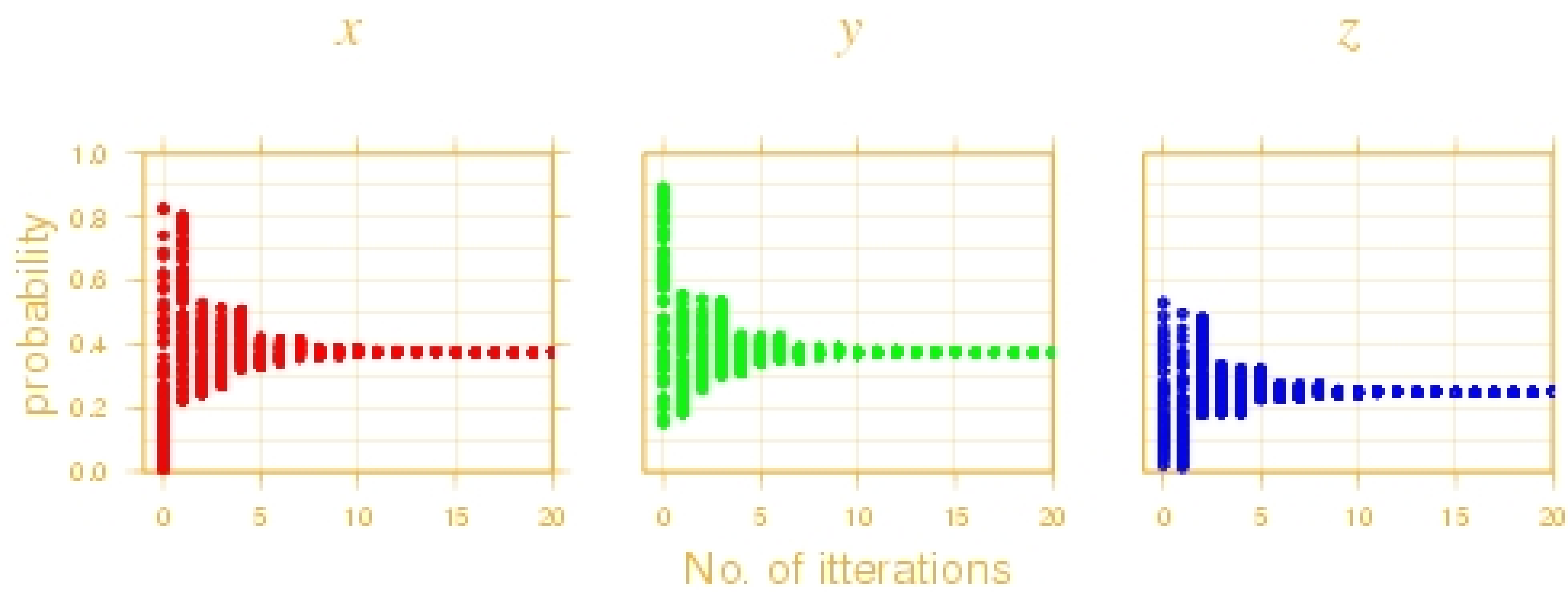
$$\mathbf{p}_i = (p_x, p_y, p_z)$$

Evolution equation

$$\mathbf{p}_{i+1} = \mathbf{p}_i \cdot \begin{pmatrix} 0.1 & 0.9 & 0 \\ 0 & 0.1 & 0.9 \\ 0.6 & 0.3 & 0.1 \end{pmatrix}$$

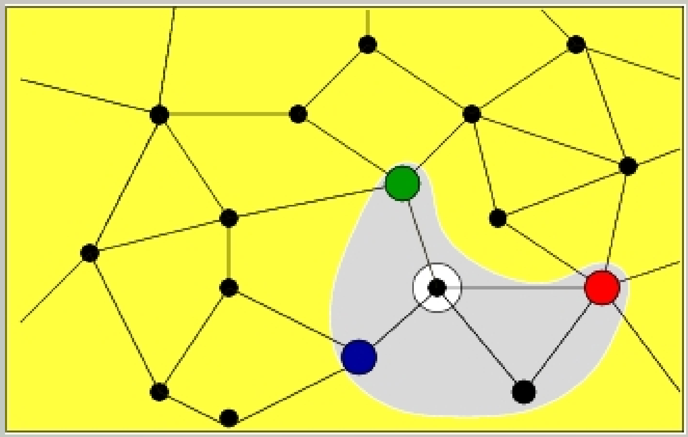
$\mathbf{p}_{i\to\infty} =???$

Algorithm - example



Defining MC

1.



2.

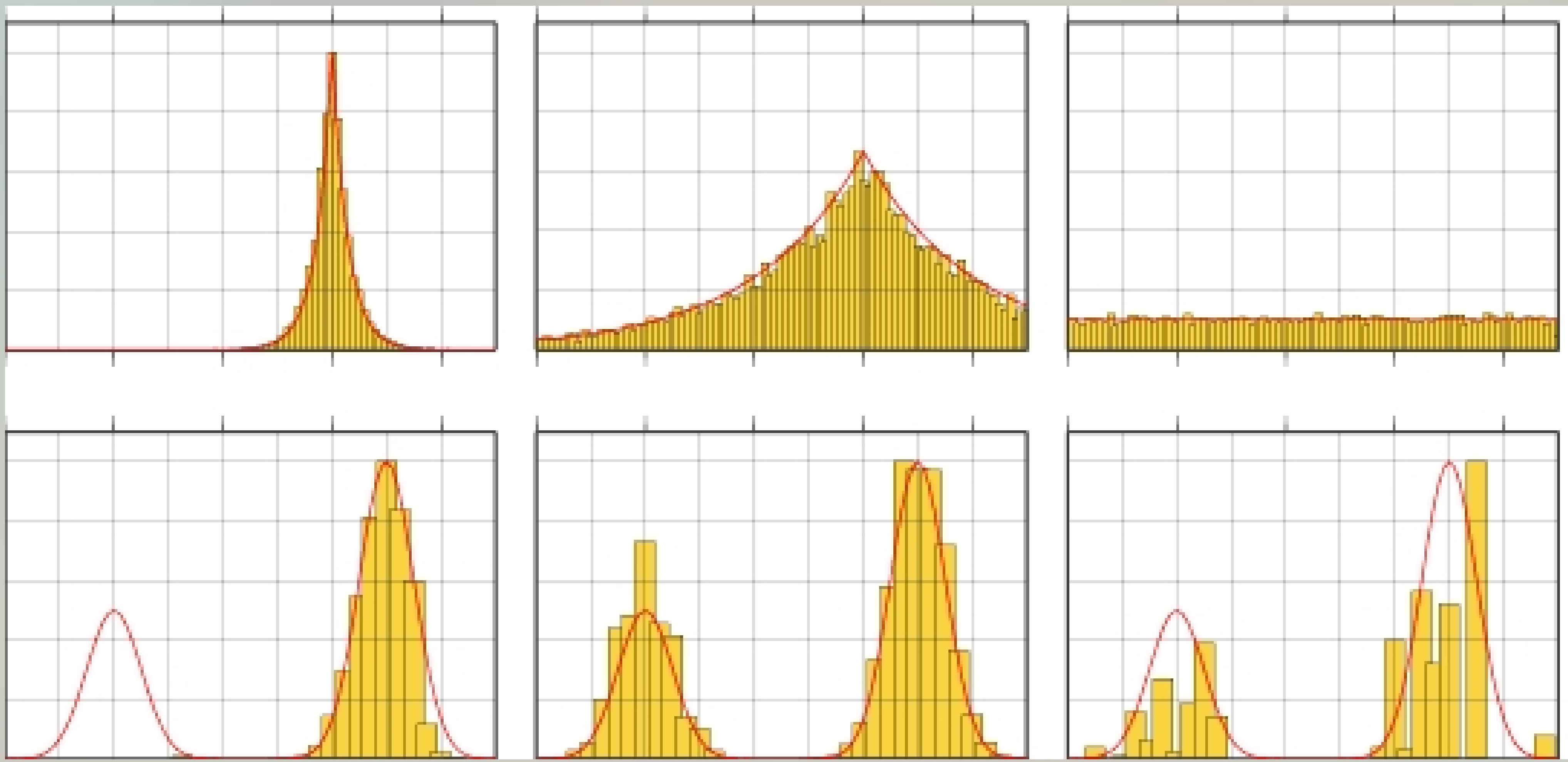
$$K = \begin{pmatrix} K_{11} & \cdots & K_{1N} \\ K_{21} & \cdots & K_{2N} \\ \vdots & \vdots & \vdots \\ K_{M1} & \cdots & K_{MN} \end{pmatrix}$$

Metropolisa-Hastings (MH) algorithm

$$x_{i+1} = x_i + \delta x_i$$

$$K(x_i; x_{i+1}) = \min\left\{1, \frac{p(x_{i+1})}{p(x_i)}\right\}$$

MH algorithm MH - limitations



MH - pseudo-code

- Utwórz rozkład "Boltzmanowski" $p(\mathbf{m},T) = \exp(-S(\mathbf{m})/T)$

  - ★ generate test sample $\mathbf{m}^\beta$: $\mathbf{m}^\beta = \mathbf{m}^\alpha + \delta\mathbf{m}$
  - ★ evaluate it $\mathbf{m}^\beta$: $p_\beta = p(\mathbf{m}^\beta, T_k)$
  - ★ create a new one $\mathbf{m}^{\alpha+1}$
    - ➡ accept $\mathbf{m}^\beta$ with probability $p = min\,(1, p_\beta/p_\alpha)$

$$\mathbf{m}^{\alpha+1} = \mathbf{m}^\beta$$

    - ➡ if $\mathbf{m}^\beta$ rejected duplicate

$$\mathbf{m}^{\alpha+1} = \mathbf{m}^\alpha$$

- Powtarzaj aż do osiągnięcia końcowej temperatury

Probability distribution - do we know them?

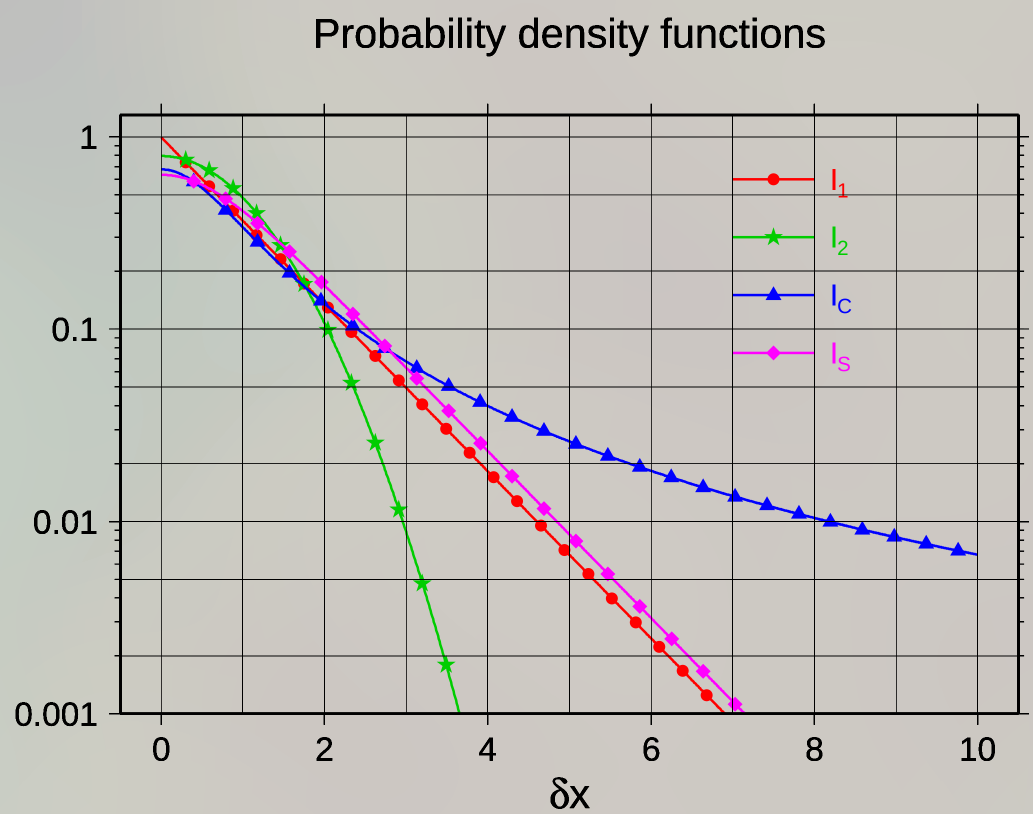
$$L(\mathbf{m},\mathbf{d}^{obs}) = \int_D p(\mathbf{d},\mathbf{d}^{obs}) \frac{q(\mathbf{m},\mathbf{d})}{\mu(\mathbf{m},\mathbf{d})} \mathrm{d}\mathbf{d}$$

While $p()$ can often be estimated (e.g. by repeating measurements) the most problematic is estimating “theoretical” uncertainties $q()$

Handicap:

$$\sigma(\delta x) \sim \exp(-||\delta x||)$$

Probability distribution - generic forms



If pdf's are unknown ...

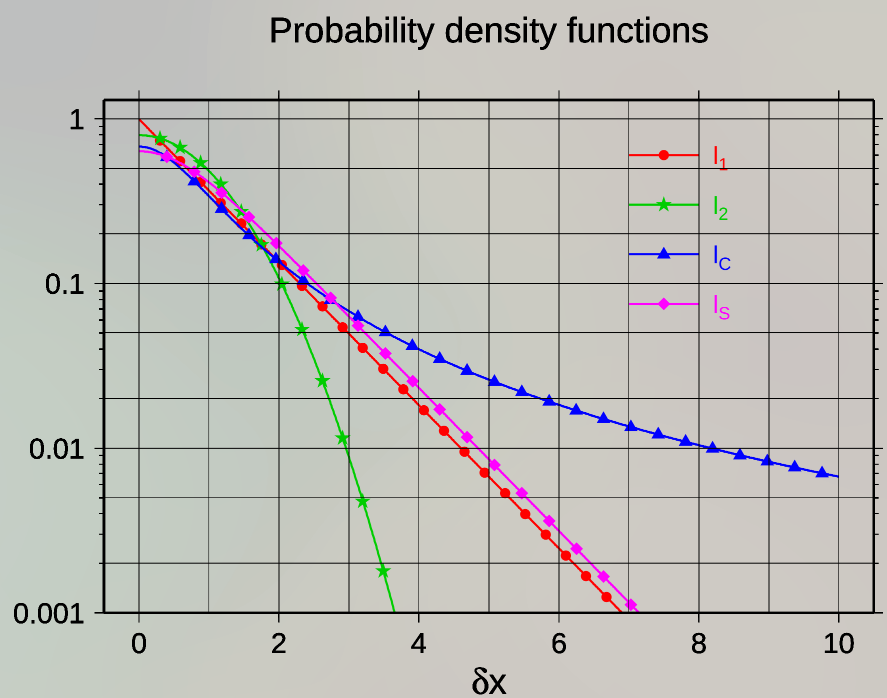
$$q(\mathbf{m},\mathbf{d}) \;= \delta(\mathbf{d}-\mathbf{G}(\mathbf{m})) \Rightarrow L(\mathbf{m}) \sim p(\mathbf{d}-\mathbf{G}(\mathbf{m}))$$

taking into account modelling errors:

$$L(\mathbf{m}) = k\exp(-||\mathbf{d}^{obs}-\mathbf{G}(\mathbf{m})||)$$

$$\sigma(\mathbf{m}) = f(\mathbf{m}) \times \exp(-||\mathbf{d}^{obs}-\mathbf{G}(\mathbf{m})||)$$

Probability distribution - generic forms



Still remaining task: how to chose the best norm ?

Using meta-information of a posteriori Bayesian solutions of the hypocenter location task for improving accuracy of location error estimation
Geophys. J. Int. Vol. 201(3), pp.1399-1408, doi: 10.1093/gji/ggv083