Relations between physical and mathematical statistics

Armen Allahverdyan
(Yerevan Physics Institute)

-- Maximum aposteriori estimation and maximum-likelihood

-- Hidden Markov models

-- $T=0$ 1d Ising model

-- outline

Math. statistics: data recovery from noise
precision, suffciency, algorithmic procedures,
machine learning, data mining

Phys. statistics: macroscopic matter (heat capacity, conductivity,...)
conservation laws, weak-coupling,
entropy and temperature

Math. and phys. statistics: introducing probabilistic description
greatly simplifies things

Direct relation between math. and phys. statistics

Maximum aposteriori estimation (MAP)

$\mathbf{x} = (x_1, \ldots, x_N)$ internal

$\mathbf{p}(\mathbf{x})$

$\mathbf{y} = (y_1, \ldots, y_N)$ observed

$\mathbf{p}(\mathbf{y}|\mathbf{x}) = \prod_{k=1}^{N} p(y_k|x_k)$

noise

Question: estimate **x** given **y**

MAP: maximize over **x**
<u>(probabilities are known)</u>

$$\mathbf{p}(\mathbf{x}|\mathbf{y}) = \frac{\mathbf{p}(\mathbf{y}|\mathbf{x})\mathbf{p}(\mathbf{x})}{\mathbf{p}(\mathbf{y})}$$

weak noise

$$\mathbf{p}(\mathbf{y}|\mathbf{x})\mathbf{p}(\mathbf{x}) \simeq \delta(\mathbf{x}-\mathbf{y}) = \prod_{k=1}^{N} \delta(x_k - y_k)$$

MAP: minimize over **x**

$$-\ln[\,\mathbf{p}(\mathbf{y}|\mathbf{x})\mathbf{p}(\mathbf{x})\,] \equiv H(\mathbf{y},\mathbf{x}).$$

MAP --> maximum likelihood

$$\mathbf{p}(\mathbf{x}) \propto \mathrm{const}$$

Minimum description length principle: the best hypothesis **x** for data **y** minimizes the *complexity* of data encoded via **x** + the *complexity* of the hypothesis **x**

Structural hierarchy of random processes

independent process

$$p(x_1, x_2, ..., x_n) = p(x_1)p(x_2)...p(x_n)$$

Markov process:
one-step memory

$$p(x_n \mid x_{n-1}, x_{n-2}, x_{n-3}...,) = p(x_n \mid x_{n-1})$$

Hidden Markov process:
noisy Markov process

No finite-memory processes anymore

Simplest hidden Markov process

$$p(\mathbf{x}) = \prod_{k=2}^{N} p(x_k|x_{k-1})p(x_1), \qquad x_k = \pm 1$$

$$p(1|1) = p(-1|-1) = 1-q,$$

$$p(1|-1) = p(-1|1) = q$$

$$p(\mathbf{y}|\mathbf{x}) = \prod_{k=1}^{N} \pi(y_k|x_k), \qquad y_k = \pm 1$$

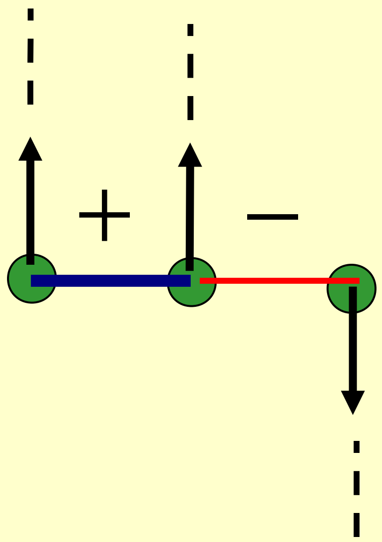
Symmetric noise

$$\pi(-1|1) = \pi(1|-1) = \epsilon$$

$$\pi(1|1) = \pi(-1|-1) = 1-\epsilon,$$

y : hidden Markov process

Hamiltonian of 1d Ising model in random fields



ferromagnetic +

anti-ferromagnetic −

$$H(\mathbf{y},\mathbf{x}) = -J\sum_{k=1}^{N} x_k x_{k+1} - h\sum_{k=1}^{N} y_k x_k$$

frozen disorder

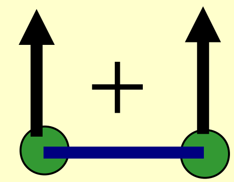
$$J = \frac{1}{2}\ln\left[\frac{1-q}{q}\right] \qquad h = \frac{1}{2}\ln\left[\frac{1-\epsilon}{\epsilon}\right]$$

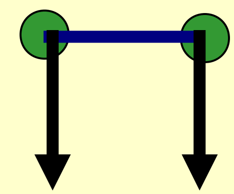
MAP-sequence --> the lowest energy state

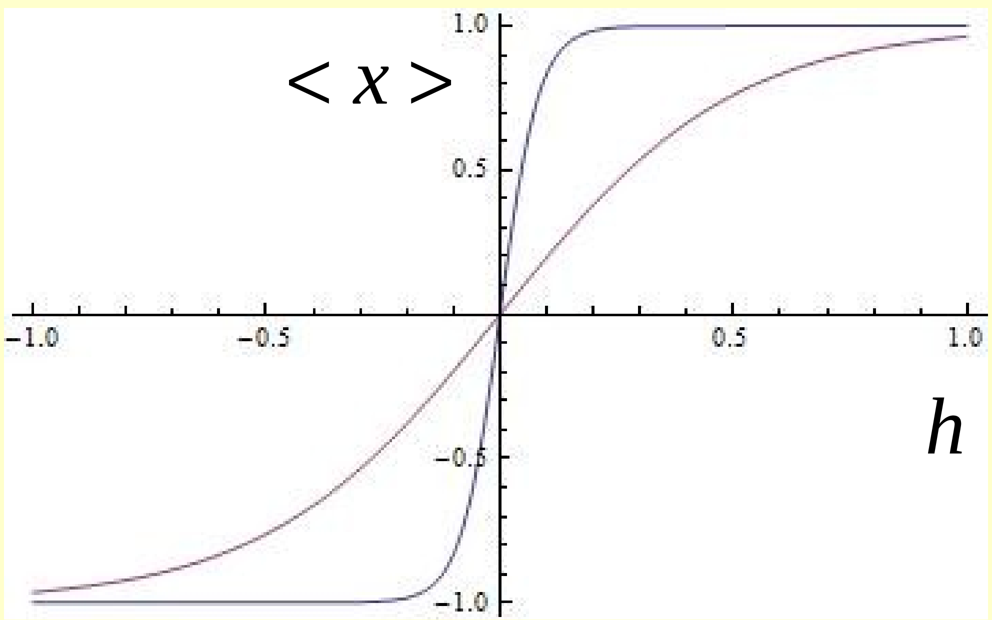
Ferromagnetic 1d Ising

$$H(\mathbf{y},\mathbf{x}) = -J\sum_{k=1}^{N} x_k x_{k+1} - h\sum_{k=1}^{N} y_k x_k$$

$$y_k = 1,\ J > 0$$







low-temperature first-order phase transition gauge transformation

$$z_i = x_i y_i, \qquad \tau_i = y_i y_{i+1}$$

$$H(\boldsymbol{\tau}, \mathbf{z}) \ = \ -J\sum_{k=1}^{N} \tau_k z_k z_{k+1} - h\sum_{k=1}^{N} z_k$$

$\hat{\mathbf{z}}(\boldsymbol{\tau})$

$$\boldsymbol{\tau} \ = (\boldsymbol{\tau}_1, \mathrm{w}_1, \boldsymbol{\tau}_2, \mathrm{w}_2, \ldots)$$

$$\hat{\mathbf{z}} = (\hat{\mathbf{z}}_1, \mathrm{w}_1, \hat{\mathbf{z}}_2, \mathrm{w}_2, \ldots)$$

domain-wall structure

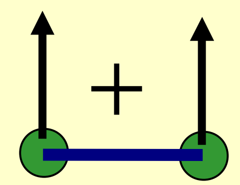
weak noise

$$h > 2J$$

$$\hat{z}_i = 1 \text{ for all } i$$

estimates = observations

$J < h < 2J,$



any number of + bonds define a wall

odd number of -- bonds

non-unique outcome

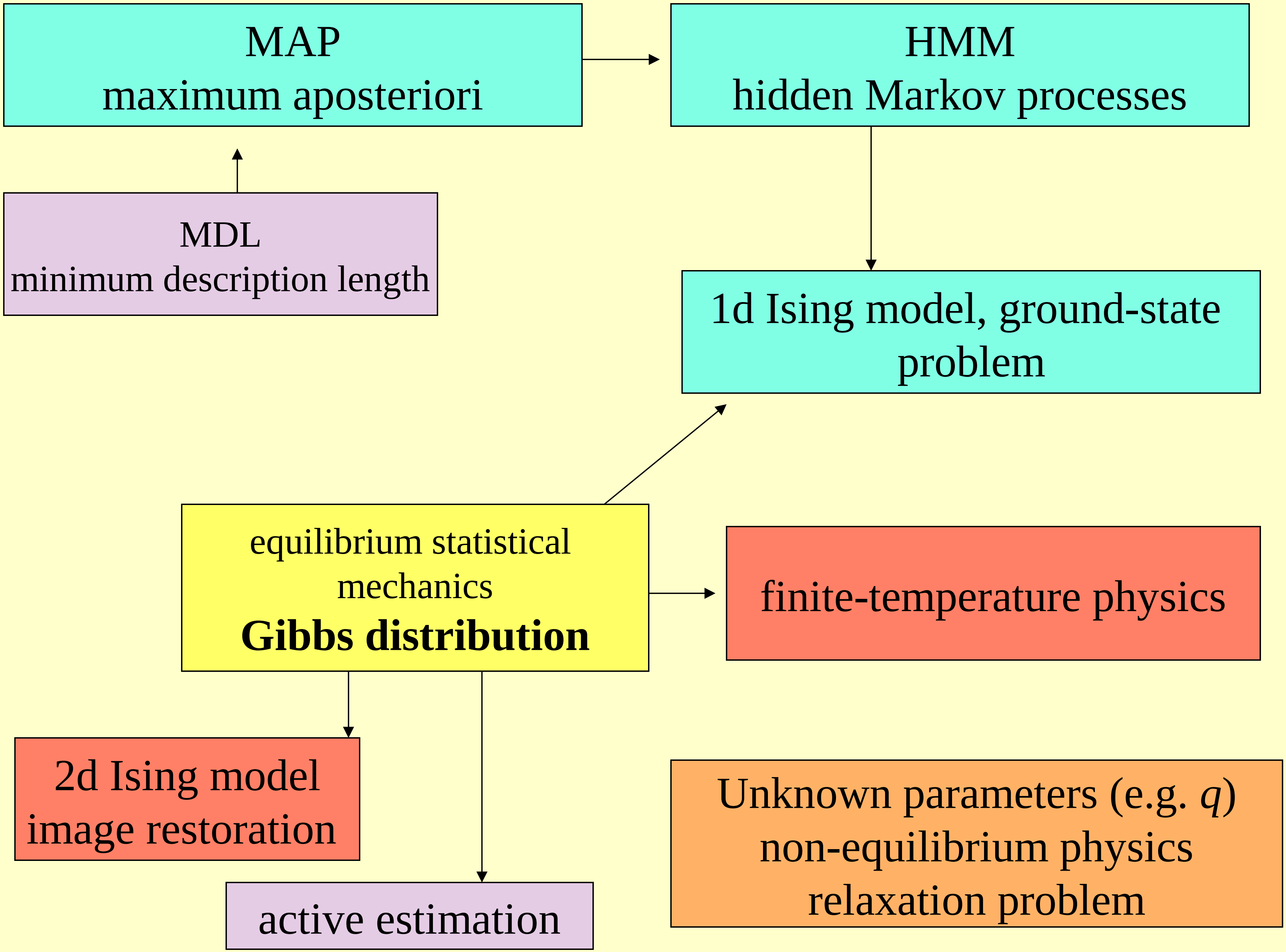
$$+ \uparrow - \uparrow - \downarrow - \uparrow +,$$
$$+ \uparrow - \downarrow - \uparrow - \uparrow + .$$

anti-ferromagnetic clusters

even number of -- bonds

unique outcome

$$+ \uparrow - \downarrow - \uparrow +$$


MAP
maximum aposteriori
HMM
hidden Markov processes
MDL
minimum description length
1d Ising model, ground-state problem
equilibrium statistical mechanics
Gibbs distribution
finite-temperature physics
2d Ising model
image restoration
Unknown parameters (e.g. *q*)
non-equilibrium physics
relaxation problem
active estimation

Law of large numbers
(self-averaging)

$$H(\mathbf{y},\mathbf{x}) \rightarrow \sum_{\mathbf{y}} \mathbf{p}(\mathbf{y}) H(\mathbf{y},\mathbf{x})$$