Welcome to

MaxEnt 2006

Twenty sixth International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering

July 8-13, 2006, CNRS, Paris, France

Thanks to

- Edwin T. Jaynes International Center for Bayesian Methods and Maximum Entropy
- Centre national de la recherche scientifique (CNRS), France
- Université de Paris-sud 11, Orsay, France
- Laboratoire des signaux et systèmes (L2S), France
- Ministère de la Défense, Direction de la recherche et développement, France
- International Society for Bayesian Analysis

Thanks to

Local organizers:

Olivier Féron, Mahieddine Ichir,

Adel Mohammadpour,

Nadia Bali, Sofia Fékih-Salem, Isabelle Smith,

Guy Demoment,

Jean-François Giovannelli & Thomas Rodet

and also

Our photographe: Mahine Mohammad-Djafari

A few points:

- Please check your registration statut, receipts, participation at conference dinner, and for those who registration fees have been weaved, please sign your registration forms.
- Today, as well as the other days, the lunches are offert.
- Conference dinner will be tuseday July 11, at 19h30.
- From tomorrow, the conference is located at CNRS. Please bring an ID.

- We are going to request each participants to Peer-Review one or two papers
- Please do this review as soon as possible during the conference
- The authors have to account for the reviewers comments for preparing their final papers.
- Final camera-ready papers are due on September 15, 2006 (Firm deadline).
- It is really unfortunate that some of the peoples could not come due to the visa problems.

Maximum Entropy and Bayesian inference: Where do we stand and where do we go ?

Ali MOHAMMAD-DJAFARI

Laboratoire des Signaux et Systèmes

UMR 08506 du CNRS-Supélec-UPS

Supélec, Plateau de Moulon

91192 Gif-sur-Yvette, FRANCE.

djafari@lss.supelec.fr

http://djafari.free.fr

http://www.lss.supelec.fr

Contents

- Definitions: Information, Entropy, Relative entropy, different types of data
- Assigning probabilities: Maximum Entropy Principle (MEP)
- Updating probabilities: Minimizing the Relative Entropy (Kullbak-Leibler) (MKL)
- Link between ME and Maximum Likelihood (ML)
- Link between MKL and Bayesian approach
- Assigning priors (Conjugate, Reference and Jeffreys priors)
- Computing posteriors
- Multivariate extentions
- Case of inverse problems

Introduction

Quantity of interest:	$X \in \{\omega_1, \cdots, \omega_n\}$
Probabilities:	$\boldsymbol{p} = \{p_1, \cdots, p_n\}$
Information quantities:	$\boldsymbol{I} = \{I_1, \cdots, I_n\}, \quad I_j = \ln \frac{1}{p_j} = -\ln p_j$
Entropy:	$H(\boldsymbol{p}) = \mathrm{E}\{I_j\} = -\sum_{j=1}^n p_j \ln p_j$
Relative Entropy (Kullbak-Leibler):	$KL(\boldsymbol{p}:\boldsymbol{q}) = \sum_{j=1}^n p_j \ln p_j/q_j$
Prior probabilities:	$\boldsymbol{q} = \{q_1, \cdots, q_n\}$
Data type 1: Expected values:	$d_k = \mathrm{E}\{\phi_k(X)\} = \sum_{j=1}^n p_j \phi_k(\omega_j), \quad k = 1, \cdots, K$
Data type 2: N direct samples:	$\boldsymbol{x} = \{x_1, \cdots, x_N\}$
Data type 3: N indirect samples:	$\boldsymbol{y} = \{y_1, \cdots, y_N\}$ with $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$
Data type 4:	
N indirect noisy samples:	$\boldsymbol{y} = \{y_1, \cdots, y_N\}$ with $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon}$

Assigning probabilities

: Given a set of data type 1: $d_k = \mathrm{E}\{\phi_k(X)\} = \sum_{j=1}^{n} p_j\, \phi_k(\omega_j),\ k = 1, \cdots, K$ assign the probabilities $\boldsymbol{p} = \{p_1, \cdots, p_n\}$

: Infinite number of possible solutions.
Maximum Entropy Principle (MEP):
Between all the possible solutions choose the one with maximum entropy

maximize $H(\boldsymbol{p}) = -\sum_j p_j\, \ln p_j$ s.t. $\sum_j p_j\, \phi_k(\omega_j) = d_k, \quad k = 1, \cdots, K$

: Lagrangian $\mathcal{L} = -\sum_{j=1}^{n} p_j\, \ln p_j + \sum_{k=0}^{K} \lambda_k \left(\sum_{j=1}^{n} p_j\, \phi_k(\omega_j) - d_k\right)$

Stationnary point: $\begin{cases} \frac{\partial \mathcal{L}}{\partial p_j} = 0 \longrightarrow p_j = \frac{1}{Z(\boldsymbol{\lambda})} \exp\left[-\sum_{k=1}^{K} \lambda_k \phi_k(\omega_j)\right] \\ \frac{\partial \mathcal{L}}{\partial \lambda_k} = 0 \longrightarrow -\frac{\partial \ln Z(\boldsymbol{\lambda})}{\partial \lambda_k} = d_k \longrightarrow \lambda^* \end{cases}$

ME solution: $p_j = \dfrac{1}{Z(\boldsymbol{\lambda}^*)} \exp\left[-\sum_{k=1}^{K} \lambda_k^* \phi_k(\omega_j)\right] = \exp\left[-\lambda_0 - \sum_{k=1}^{K} \lambda_k^* \phi_k(\omega_j)\right]$

where $Z(\boldsymbol{\lambda}) = \exp[\lambda_0] = \sum_{j=1}^{n} \exp\left[-\sum_{k=1}^{K} \lambda_k \phi_k(\omega_j)\right]$

Some properties of ME solution

$$
\begin{aligned}
-\frac{\partial \ln Z(\boldsymbol{\lambda})}{\partial \lambda_k} &= -\frac{\partial \lambda_0(\boldsymbol{\lambda})}{\partial \lambda_k} = \mathrm{E}\left\{\phi_k(X)\right\} \\
-\frac{\partial \ln Z(\boldsymbol{\lambda})}{\partial \lambda_k \partial \lambda_l} &= -\frac{\partial \lambda_0(\boldsymbol{\lambda})}{\partial \lambda_k \partial \lambda_l} = \mathrm{E}\left\{\phi_k(X)\phi_l(X)\right\} \\
H &= \lambda_0 + \sum_k \lambda_k \mathrm{E}\left\{\phi_k(X)\right\} \\
H_{\max} &= \lambda_0 + \sum_k \lambda_k d_k
\end{aligned}
$$

Updating probabilities

: Given the prior probabilities $\boldsymbol{q}$ and a set of data type 1:
$d_k = \mathrm{E}\{\phi_k(X)\} = \sum_{j=1}^{n} p_j\, \phi_k(\omega_j),\ k = 1, \cdots, K$, update $\boldsymbol{q}$ to $\boldsymbol{p}$

: Minimum Kullbak-Leibler principle (MKLP):

$$\text{minimize } K(\boldsymbol{p} : \boldsymbol{q}) = \sum_{j=1}^{n} p_j \ln p_j/q_j \quad \text{s.t.} \quad \sum_{j=1}^{n} p_j\, \phi_k(\omega_j) = d_k,\ k = 1, \cdots, K$$

MKL solution:

$$p_j = \frac{q_j}{Z(\boldsymbol{\lambda})} \exp\left[-\sum_{k=1}^{K} \lambda_k \phi_k(\omega_j)\right]$$

$$Z(\boldsymbol{\lambda}) = \sum_{j} q_j \exp\left[-\sum_{k=1}^{K} \lambda_k \phi_k(\omega_j)\right]$$

Assigning a probability law (Continuous case)

X a continuous quantity with probability density function $p(x)$

: Given a set of data type 1: $d_k = \mathrm{E}\{\phi_k(X)\} = \int p(x)\,\phi_k(x)\,\mathrm{d}x \; k = 1, \cdots, K$, assign the pdf $p(x)$

: Maximum Entropy Principle (MEP):

$$\begin{aligned} &\text{maximize} \quad H(\boldsymbol{p}) = -\int p(x)\,\ln p(x)\,\mathrm{d}x \\ &\quad \text{s.t.} \quad \int p(x)\phi_k(x)\,\mathrm{d}x = d_k, \quad k = 1, \cdots, K \end{aligned}$$

: Lagrangian $\mathcal{L} = -\int p(x)\,\ln p(x)\,\mathrm{d}x + \sum_{k=0}^{K} \lambda_k \left(\int p(x)\,\phi_k(x)\,\mathrm{d}x - d_k\right)$

Stationnary point:
$$\begin{cases} \frac{\partial \mathcal{L}}{\partial p(x)} = 0 \longrightarrow p(x) = \frac{1}{Z(\boldsymbol{\lambda})} \exp\left[-\sum_{k=1}^{K} \lambda_k \phi_k(x)\right] \\ \frac{\partial \mathcal{L}}{\partial \lambda_k} = 0 \longrightarrow -\frac{\partial \ln Z(\boldsymbol{\lambda})}{\partial \lambda_k} = d_k \longrightarrow \lambda^* \end{cases}$$

$$\text{ME solution:} \quad p(x) = \frac{1}{Z(\boldsymbol{\lambda}^*)} \exp\left[-\sum_{k=1}^{K} \lambda_k^* \phi_k(x)\right]$$

Updating probability laws (Continuous case)

: Given the prior probabilities $q(x)$ and a set of data type 1:
$d_k = \mathrm{E}\{\phi_k(X)\} = \int p(x)\,\phi_k(x)\,\mathrm{d}x,\ k = 1, \cdots, K$, update $q(x)$ to $p(x)$.

: Minimum Kullbak-Leibler principle (MKLP):

$$\text{minimize}\quad K(p:q) = \int p(x)\,\ln[p(x)/q(x)]\,\mathrm{d}x$$

$$\text{s.t.}\quad \int p(x)\,\phi_k(x)\,\mathrm{d}x = d_k, \quad k = 1, \cdots, K$$

MKL solution:
$$p(x) = \frac{q(x)}{Z(\boldsymbol{\lambda})}\exp\left[-\sum_{k=1}^{K}\lambda_k\phi_k(x)\right]$$

$$Z(\boldsymbol{\lambda}) = \int q(x)\exp\left[-\sum_{k=1}^{K}\lambda_k\phi_k(x)\right]\,\mathrm{d}x$$

Simple examples

x.	x	$q(x)$	$\mathrm{E}\{\phi_k(x)\} = d_k$	$p(x)$
	$x > 0$	cte	$\mathrm{E}\{X\} = d_1 = \mu$	$\frac{1}{z(\lambda)} \exp[-\lambda x]$
	$x > 0$	cte	$\mathrm{E}\{X\} = d_1$ $\mathrm{E}\{\ln X\} = d_2$	$\frac{1}{z(\boldsymbol{\lambda})} \exp[-\lambda_1 x - \lambda_2 \ln x]$ $= \mathcal{G}(\lambda_2 + 1, \lambda_1)$
	$x > 0$	$\mathcal{G}(\alpha_0, \beta_0)$	$\mathrm{E}\{X\} = d_1 = \beta/\alpha$ $\mathrm{E}\{\ln X\} = d_2 = \alpha$	$\mathcal{G}(\alpha, \beta)$
	$x \in R$	cte	$\mathrm{E}\{X\} = d_1 = \mu_0$ $\mathrm{E}\{(X-\mu)^2\} = d_2 = \sigma_0^2$	$\mathcal{N}(\mu_0, \sigma_0^2)$
	$x \in R$	$\mathcal{N}(\mu_0, \sigma_0^2)$	$\mathrm{E}\{X\} = d_1 = \mu_1$ $\mathrm{E}\{(X-\mu)^2\} = d_2 = \sigma_1^2$	$\mathcal{N}(\mu, \sigma^2)$, $\mu = \sigma^{-2}(\mu_0/\sigma_0^2 + \mu_1/\sigma_1^2)$, $\sigma^{-2} = [1/\sigma_0^2 + 1/\sigma_1^2)]$

Link between MEP and Maximum Likelihood (ML)

Quantity of interest: X, a continuous random variable

Data type 1: $d_k = \mathrm{E}\{\phi_k(X)\} = \int p(x)\,\phi_k(x)\,\mathrm{d}x, \quad k = 1, \cdots, K$

ME solution: $p(x;\boldsymbol{\lambda}) = \frac{1}{Z(\boldsymbol{\lambda})} \exp\left[-\sum_{k=1}^{K} \lambda_k \phi_k(x)\right]$

λ solution of: $-\frac{\partial \ln Z(\boldsymbol{\lambda})}{\partial \lambda_k} = d_k, \quad k = 1, \cdots, K$

Data type 2: N direct samples: $\boldsymbol{x} = \{x_1, \cdots, x_N\}$

Choose a param. family: $p(x;\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp\left[-\sum_{k=1}^{K} \theta_k \phi_k(x)\right]$

Assume x_j iid: $p(\boldsymbol{x};\boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{1}{Z(\boldsymbol{\theta})} \exp\left[-\sum_{k=1}^{K} \theta_k \phi_k(x_j)\right]$

Define the Likelihood: $\mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{1}{Z^n(\boldsymbol{\theta})} \exp\left[-\sum_{j=1}^{N}\sum_{k=1}^{K} \theta_k \phi_k(x_j)\right]$

ML solution: $\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{\mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta})\}$

$\widehat{\boldsymbol{\theta}}$ solution of: $-\frac{\partial \ln Z(\boldsymbol{\theta})}{\partial \theta_k} = \frac{1}{n}\sum_{j=1}^{N} \phi_k(x_j)$

Link between MKL and Bayesian approach

uantity of interest: X, a continuous random variable with prior $q(x)$

ata type 1: $d_k = \mathrm{E}\{\phi_k(X)\} = \int p(x)\,\phi_k(x)\,\mathrm{d}x, \quad k = 1, \cdots, K$

MKL solution: $p(x|\boldsymbol{\lambda}) = \frac{q(x)}{Z(\boldsymbol{\lambda})} \exp\left[-\sum_{k=1}^{K} \lambda_k \phi_k(x)\right]$

λ solution of: $-\frac{\partial \ln Z(\boldsymbol{\lambda})}{\partial \lambda_k} = d_k, \quad k = 1, \cdots, K$

ote that

$$
\begin{aligned}
p(x|\boldsymbol{\lambda}) &= \frac{q(x)}{Z(\boldsymbol{\lambda})} \exp\left[-\sum_{k=1}^{K} \lambda_k \phi_k(x)\right] \\
&\propto q(x) \quad \exp\left[-\sum_{k=1}^{K} \lambda_k \phi_k(x)\right]
\end{aligned}
$$

a posteriori $\propto$ a priori Data Likelihood

Link between MKL and Bayesian approach (continued)

Data type 2: N direct samples: $\boldsymbol{x} = \{x_1, \cdots, x_N\}$

Choose a param. family: $p(x|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp\left[-\sum_{k=1}^{K} \theta_k \phi_k(x)\right]$

Define the Likelihood: $\mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{1}{Z^n(\boldsymbol{\theta})} \exp\left[-\sum_{j=1}^{N}\sum_{k=1}^{K} \theta_k \phi_k(x_j)\right]$

Assign a prior on: $\boldsymbol{\theta}$ $\pi(\boldsymbol{\theta})$

Apply the Bayes rule: $p(\boldsymbol{\theta}|\boldsymbol{x}) \propto \pi(\boldsymbol{\theta})\ \mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta})$

a posteriori a priori Likelihood

Extract an information on $\boldsymbol{\theta}$:

the mean: $\widehat{\boldsymbol{\theta}} = \int \boldsymbol{\theta}\ p(\boldsymbol{\theta}|\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = \frac{\int \boldsymbol{\theta}\ \mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}}{\int \mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}}$

or the mode: $\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{\pi(\boldsymbol{\theta})\ \mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta})\}$

Signification of $p(\boldsymbol{x}|\widehat{\boldsymbol{\theta}})$ and link with $p(\boldsymbol{x}|\boldsymbol{\lambda})$?

Data type 1: $d_k = \mathrm{E}\{\phi_k(X)\} = \int p(x)\,\phi_k(x)\,\mathrm{d}x, \quad k = 1, \cdots, K$

$$p(x|\boldsymbol{\lambda}) \propto q(x|\boldsymbol{\lambda}_0) \exp\left[-\sum_{k=1}^{K} \lambda_k \phi_k(x)\right]$$

$$\text{a posteriori} \propto \text{a priori} \quad \text{Data Likelihood}$$

Data type 2: N direct samples: $\boldsymbol{x} = \{x_1, \cdots, x_N\}$

$$p(\boldsymbol{\theta}|\boldsymbol{x}) \propto \pi(\boldsymbol{\theta}|\boldsymbol{x}_0) \exp\left[-\sum_{j=1}^{N}\sum_{k=1}^{K} \theta_k \phi_k(x_j)\right]$$

$$\text{a posteriori} \propto \text{a priori} \quad \text{Data Likelihood}$$

few questions:

- How to assign $q(x|\boldsymbol{\lambda}_0)$ or $\pi(\boldsymbol{\theta}|\boldsymbol{x}_0)$?
- How to use $p(x|\boldsymbol{\lambda})$ or $p(\boldsymbol{\theta}|\boldsymbol{x})$?
- How to compute $\mathrm{E}\{X\}$ using $p(x|\boldsymbol{\lambda})$ or $\mathrm{E}\{\boldsymbol{\theta}\}$ using $p(\boldsymbol{\theta}|\boldsymbol{x})$?
- Any link between $q(x|\boldsymbol{\lambda}_0)$ and $\pi(\boldsymbol{\theta}|\boldsymbol{x}_0)$ or between $p(x|\boldsymbol{\lambda})$ and $p(\boldsymbol{\theta}|\boldsymbol{x})$?

Assigning priors: Conjugate, Reference and Jeffrey's priors

ata type 2: n direct samples: $\boldsymbol{x}_n = \{x_1, \cdots, x_n\}$

$$p(\boldsymbol{\theta}|\boldsymbol{x}_n) = \frac{1}{z_n(\boldsymbol{x}_n)}\,\pi(\boldsymbol{\theta})\,f(\boldsymbol{x}_n|\boldsymbol{\theta}) \quad \text{with} \quad f(\boldsymbol{x}_n|\boldsymbol{\theta}) = \exp\left[-\sum_{j=1}^{n}\sum_{k=1}^{K}\theta_k\phi_k(x_j)\right]$$

- Conjugate priors: posterior $p(\boldsymbol{\theta}|\boldsymbol{x})$ and prior $\pi(\boldsymbol{\theta})$ in the same family

$$\text{Exponential family:} \quad \pi(\boldsymbol{\theta}|\boldsymbol{x}_0) = \frac{1}{z_0(\boldsymbol{x}_0)}\pi_0(\boldsymbol{\theta})\exp\left[-\sum_{j=1}^{n_0}\sum_{k=1}^{K}\phi_k(x_{0_j})\theta_k\right]$$

where $\boldsymbol{x}_{0n_0} = \{x_{0_1}, \cdots, x_{0_{n_0}}\}$

- Reference priors: $n_0 = 1$

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}_0) = \frac{1}{z_0(\boldsymbol{x}_0)}\pi_0(\boldsymbol{\theta})\exp\left[-\boldsymbol{\phi}^t\boldsymbol{\theta}\right] \quad \text{with } \boldsymbol{\phi} = [\phi_1(x_0), \cdots, \phi_K(x_0)]^t$$

- Jeffrey's priors:

Jeffrey's priors

$$p(\boldsymbol{\theta}|\boldsymbol{x}_n) = \frac{1}{z_n(\boldsymbol{x}_n)}\ \pi(\boldsymbol{\theta})\ f(\boldsymbol{x}_n|\boldsymbol{\theta})$$

ain in information:

$$KL[p(\boldsymbol{\theta}|\boldsymbol{x}_n) : \pi(\boldsymbol{\theta})] = \int\int \frac{1}{z_n(\boldsymbol{x}_n)}\ \pi(\boldsymbol{\theta})\ f(\boldsymbol{x}_n|\boldsymbol{\theta}) \ln \frac{f(\boldsymbol{x}_n|\boldsymbol{\theta})}{z_n(\boldsymbol{x}_n)}\, \mathrm{d}\boldsymbol{x}_n\, \mathrm{d}\boldsymbol{\theta}$$

hen $n \longrightarrow \infty$

$$\pi(\boldsymbol{\theta}) = |I_n(\boldsymbol{\theta})|^{1/2} \text{ with } I_n(\boldsymbol{\theta}) = \begin{bmatrix} \cdots & \cdots & \cdots \\ & \frac{\partial f(\boldsymbol{x}_n|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k \partial \boldsymbol{\theta}_l} & \\ \cdots & \cdots & \cdots \end{bmatrix}$$

sher's Information matrix.

Multivariate extensions

a random vector with $p(\boldsymbol{x})$. Data type 1:

$$d_k = \mathrm{E}\{\phi_k(\boldsymbol{X})\} = \int p(\boldsymbol{x})\,\phi_k(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x},\ k = 1, \cdots, K$$

$$p(\boldsymbol{x}|\boldsymbol{\lambda}) \propto q(\boldsymbol{x}|\boldsymbol{\lambda}_0) \exp\left[-\sum_{k=1}^{K}\lambda_k\phi_k(\boldsymbol{x})\right]$$

$$\propto \text{a priori} \quad \text{Data Likelihood}$$

- Minimizing $KL(p:q) \longrightarrow$ Minimizing $D(\boldsymbol{\lambda};\boldsymbol{\lambda}_0)$ (Primal-Dual optimization)
- $D(\boldsymbol{\lambda};\boldsymbol{\lambda}_0)$ is a distance measure and its expression depends on $q(\boldsymbol{x}|\boldsymbol{\lambda}_0)$
- If $q(\boldsymbol{x}|\boldsymbol{\lambda}_0)$ is separable then $p(\boldsymbol{x}|\boldsymbol{\lambda})$ is also separable
- If we note by

$$\mathrm{E}_q\{\boldsymbol{X}\} = \int \boldsymbol{x}\, q(\boldsymbol{x}|\lambda_0)\,\mathrm{d}\boldsymbol{x} = \boldsymbol{x}_q \quad \text{and} \quad \mathrm{E}_p\{\boldsymbol{X}\} = \int \boldsymbol{x}\, p(\boldsymbol{x}|\lambda)\,\mathrm{d}\boldsymbol{x} = \boldsymbol{x}_p$$

then minimizing $KL(p:q) \longrightarrow$ minimizing $\Delta(\boldsymbol{x}_p : \boldsymbol{x}_q)$.

ata type 3: M indirect samples: $\boldsymbol{y} = \{y_1, \cdots, y_M\}$ where $\boldsymbol{A}$ is a $M \times N$ matrix
d $\boldsymbol{y} = \mathrm{E}\{\boldsymbol{A}\boldsymbol{X}\} = \boldsymbol{A}\mathrm{E}\{\boldsymbol{X}\}$ and the prior measure $q(\boldsymbol{x}|\lambda_0)$ and

$$\mathrm{E}_q\{\boldsymbol{X}\} = \int \boldsymbol{x}\, q(\boldsymbol{x}|\lambda_0)\, \mathrm{d}\boldsymbol{x} = \boldsymbol{x}_0$$

$$p(\boldsymbol{x}|\boldsymbol{\lambda}) \propto q(\boldsymbol{x}|\boldsymbol{\lambda}_0) \exp\left[-\sum_{k=1}^{K} \lambda_k [\boldsymbol{A}\boldsymbol{x}]_k)\right]$$

$$\mathrm{E}_p\{\boldsymbol{X}\} = \int \boldsymbol{x}\, p(\boldsymbol{x}|\lambda)\, \mathrm{d}\boldsymbol{x} = \boldsymbol{x}$$

- Minimizing $KL(p:q) \longrightarrow$ Minimizing $D(\boldsymbol{\lambda};\boldsymbol{\lambda}_0) \longrightarrow$ Minimizing $\Delta(\boldsymbol{x}:\boldsymbol{x}_0)$ subject to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$.
- $\Delta(\boldsymbol{x};\boldsymbol{x}_0)$ is a distance measure and its expression depends on the family form of $q(\boldsymbol{x}|\boldsymbol{\lambda}_0)$
- If $q(\boldsymbol{x}|\boldsymbol{\lambda}_0)$ is separable then $\Delta(\boldsymbol{x};\boldsymbol{x}_0) = \sum_{j=1}^{N} \Delta_j(x_j;x_{0j})$
- If $q(\boldsymbol{x})$ Gaussian, then $D(\boldsymbol{\lambda};\boldsymbol{\lambda}_0) = \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\|^2$ and $\Delta(\boldsymbol{x};\boldsymbol{x}_0) = \|\boldsymbol{x} - \boldsymbol{x}_0\|^2$
- If $q(\boldsymbol{x})$ Poisson measure, then $\Delta(\boldsymbol{x};\boldsymbol{x}_0) = \sum_j x_j \ln(x_j/x_{0j}) + (x_j - x_{0j})$

ata type 4: M indirect samples: $\boldsymbol{y} = \{y_1, \cdots, y_M\}$ where $\boldsymbol{A}$ is a $M \times N$ matrix
d $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon}$ and the prior probability laws:

$$p_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}|\boldsymbol{\theta}_1), \qquad p(\boldsymbol{x}|\boldsymbol{\theta}_2) = \frac{1}{Z(\boldsymbol{\theta}_2)} \exp\left[-\boldsymbol{\theta}_2^t \boldsymbol{\phi}(\boldsymbol{x})\right],$$

ase 1: $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ known.

$$\begin{aligned}
p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}_1) &= p_{\boldsymbol{\epsilon}}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}|\boldsymbol{\theta}_1) \\
p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &\propto p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}_1)\, p(\boldsymbol{x}|\boldsymbol{\theta}_2) \\
p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}) &= p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}_1)\, p(\boldsymbol{x}|\boldsymbol{\theta}_2)/p(\boldsymbol{y}|\boldsymbol{\theta}), \quad \boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
p(\boldsymbol{y}|\boldsymbol{\theta}) &= \int p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}_1)\, p(\boldsymbol{x}|\boldsymbol{\theta}_2)\, \mathrm{d}\boldsymbol{x}
\end{aligned}$$

fering on $\boldsymbol{x}$:

Mode	$\widehat{\boldsymbol{x}}(\boldsymbol{\theta})$	$= \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})\}$	needs optimization
Mean	$\widehat{\boldsymbol{x}}(\boldsymbol{\theta})$	$= \int \boldsymbol{x}\, p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}$	needs integation
		$= \frac{\int \boldsymbol{x}\, p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}) p(\boldsymbol{x}|\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}}{\int p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}) p(\boldsymbol{x}|\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}}$	
Sampling	$\boldsymbol{x}$	$\sim p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$	Monté Carlo techniques

Case 2: $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ unknown:

$$p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}_1) = p_{\boldsymbol{\epsilon}}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}|\boldsymbol{\theta}_1)$$

$$p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}_1)\, p(\boldsymbol{x}|\boldsymbol{\theta}_2)\, \pi(\boldsymbol{\theta}), \quad \boldsymbol{\theta}=(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$$

Infering on $\boldsymbol{x}$:

$$p(\boldsymbol{x}|\boldsymbol{y}) = \int p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{\theta}$$

$$\text{Mode } \widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x}|\boldsymbol{y})\}$$

$$\text{Mean } \widehat{\boldsymbol{x}} = \int \boldsymbol{x}\, p(\boldsymbol{x}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{x} = \int\int \boldsymbol{x}\, p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{\theta}$$

Infering on $\boldsymbol{\theta}$:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \int p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}$$

$$\text{Mode } \widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{p(\boldsymbol{\theta}|\boldsymbol{y})\}$$

$$\text{Mean } \widehat{\boldsymbol{\theta}} = \int \boldsymbol{\theta}\, p(\boldsymbol{\theta}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{\theta} = \int\int \boldsymbol{\theta}\, p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{\theta}$$

Infering on $(\boldsymbol{x},\boldsymbol{\theta})$: $(\boldsymbol{x},\boldsymbol{\theta}) \sim p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})$

Joint MAP $(\widehat{\boldsymbol{x}},\widehat{\boldsymbol{\theta}}) = \arg\max_{\boldsymbol{x},\boldsymbol{\theta}} \{p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})\}$

Joint sampling $\boldsymbol{x} \sim p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y}), \quad \boldsymbol{\theta} \sim p(\theta|\boldsymbol{y})$

Gibbs sampling $\boldsymbol{x} \sim p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y}), \quad \boldsymbol{\theta} \sim p(\theta|\boldsymbol{x},\boldsymbol{y})$

Computational aspects

$$
\begin{aligned}
\boldsymbol{y} &= \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon} \\
p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}_1) &= p_{\boldsymbol{\epsilon}}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}|\boldsymbol{\theta}_1) \\
p(\boldsymbol{x}|\boldsymbol{\theta}_2) &= \frac{1}{Z(\boldsymbol{\theta}_2)} \exp\left[-\boldsymbol{\theta}_2^t \boldsymbol{\phi}(\boldsymbol{x})\right] \\
p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}_1)\, p(\boldsymbol{x}|\boldsymbol{\theta}_2)\, \pi(\boldsymbol{\theta}), \qquad \boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)
\end{aligned}
$$

timation of $\boldsymbol{x}, \boldsymbol{\theta}$:

- Mode: $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{x}, \boldsymbol{\theta})} \{p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})\}$, needs optimization
- Mean: $\begin{cases} \mathrm{E}\{\boldsymbol{x}\} = \int \boldsymbol{x}\, p(\boldsymbol{x}|\boldsymbol{y})\, \mathrm{d}\boldsymbol{x} \\ \mathrm{E}\{\boldsymbol{\theta}\} = \int \boldsymbol{\theta}\, p(\boldsymbol{\theta}|\boldsymbol{y})\, \mathrm{d}\boldsymbol{\theta} \end{cases}$, needs integrations
- Sampling: $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{\theta}}) \sim p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$, needs sampling techniques

- Main difficulty: $p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$ is not, in general, separable in $\boldsymbol{x}, \boldsymbol{\theta}$ neither in components of $\boldsymbol{x}$ nor in components of $\boldsymbol{\theta}$ $\longrightarrow$ Separable Approximations

Variational Bayes

$$p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y}) \simeq q_1(\boldsymbol{x}|\boldsymbol{y})\; q_2(\boldsymbol{\theta}|\boldsymbol{y})$$

here $q_1(\boldsymbol{x}|\boldsymbol{y})$ and $q_2(\boldsymbol{\theta}|\boldsymbol{y})$ are such that $KL\left(q_1\; q_2 : p\right)$ be minimized.

$$\begin{aligned}
KL\left(q_1\; q_2 : p\right) &= \int\int q_1(\boldsymbol{x}|\boldsymbol{y})\; q_2(\boldsymbol{\theta}|\boldsymbol{y}) \ln \frac{q_1(\boldsymbol{x}|\boldsymbol{y}) q_2(\boldsymbol{\theta}|\boldsymbol{y})}{p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})}\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{\theta} \\
&= cte + \int q_1(\boldsymbol{x}|\boldsymbol{y}) \left(\int q_2(\boldsymbol{\theta}|\boldsymbol{y}) \ln \frac{q_1(\boldsymbol{x}|\boldsymbol{y}) q_2(\boldsymbol{\theta}|\boldsymbol{y})}{p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})}\, \mathrm{d}\boldsymbol{\theta} \right) \mathrm{d}\boldsymbol{x} \\
&= cte + \int q_2(\boldsymbol{\theta}|\boldsymbol{y}) \left(\int q_1(\boldsymbol{x}|\boldsymbol{y}) \ln \frac{q_1(\boldsymbol{x}|\boldsymbol{y}) q_2(\boldsymbol{\theta}|\boldsymbol{y})}{p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})}\, \mathrm{d}\boldsymbol{x} \right) \mathrm{d}\boldsymbol{\theta}
\end{aligned}$$

- $KL\left(q_1\; q_2 : p\right)$ is a convex function of q_1 and q_2. This optimization can be done iteratively

$$\begin{aligned}
\widehat{q_1} &= \arg\min_{q_1} \left\{ KL\left(q_1\; \widehat{q_2} : p\right) \right\} \\
\widehat{q_2} &= \arg\min_{q_2} \left\{ KL\left(\widehat{q_1}\; q_2 : p\right) \right\}
\end{aligned}$$

Where do we have to go ?

- Forward modeling and assigning a probability laws to the errors:

$$\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{x}) + \boldsymbol{\epsilon} \longrightarrow \mathcal{L}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}_1) = q_{\boldsymbol{\epsilon}}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}|\boldsymbol{\theta}_1) = p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}_1)$$

  Link with physics $\longrightarrow$ Likelihood

- Modeling unknown quantities $\boldsymbol{x}$ and assigning probability laws:

  Simple models: $p(\boldsymbol{x}|\boldsymbol{\theta}_1)$

  Models with hidden variables: $p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\theta}_2)$, $p(\boldsymbol{z}|\boldsymbol{\theta}_3)$

- Assigning prior laws to the hyperparameters $p(\boldsymbol{\theta})$:
  Conjugate, Reference, Entropic, Jeffreys priors

- Obtaining expressions of the posterior laws

$$\begin{aligned} p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}_1)\, p(\boldsymbol{x}|\boldsymbol{\theta}_2)\, \pi(\boldsymbol{\theta}), & \boldsymbol{\theta} &= (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\ p(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}_1)\, p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\theta}_2)\, \pi(\boldsymbol{z}|\boldsymbol{\theta}_3)\, \pi(\boldsymbol{\theta}), & \boldsymbol{\theta} &= (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3) \end{aligned}$$

- Using posterior laws to give practical solutions:
  Joint Modes, Means, Marginal modes or means, integration of nuisance parameters, ...
  - Computing modes needs huge dimentional multivariate optimization
  - Computing means needs huge dimentional multivariate integration
  - Sampling is a good tool for exploring the whole probability density and compute approximate means. However, sampling from a non-separable multivariate probability law is not so easy.
- Finding appropriate approximations to do fast computations:
  Laplace approximation, Separable approximation, Variational and Mean Field approximations
- Evaluating the performances of the obtained algorithmes
- Evaluating the remaining uncertainties

Thanks

and now starts

More deep presentations

on those aspects