Westfälische
Wilhelms-Universität
Münster

The Bayesian Approach to Inverse Problems: Theoretical Aspects

Invited Talk at the University of Cambridge, UK

Felix Lucka

13.11.2012

Outline

General Setting

Consider a typical inverse problem given by

$$\mathcal{K}(u^\dagger) = f^{\dagger,\infty},$$

where:

- $u^\dagger \in \mathcal{U}$ is to be recovered.
- $\mathcal{K} : \mathcal{U} \longrightarrow \mathcal{X}$ describes the underlying physical process.
- $f^{\dagger,\infty} \in \mathcal{X}$ is the infinite dimensional, noise-free data.
- $f^\dagger \in \mathbb{R}^m, f := P(f^{\dagger,\infty}) := K(u^\dagger)$ is the finite dimensional, noise-free data.
- $f \in \mathbb{R}^m$ is the actually observed, finite dimensional, noisy data.

Inverse Problems from an Applied Mathematics Perspective

1. Use techniques developed for analyzing, solving and simulating the underlying PDEs (*functional analysis, spectral theory, Sobolev spaces, distribution theory, variational calculus,...*).

2. Analyze $\mathcal{K}$ w.r.t. $\mathcal{U}$ and $\mathcal{X}$ to identify the structure of its deficiency:
   - Singular systems?
   - Existence and uniqueness of solutions?
   - Stability w.r.t. to deterministic perturbations $f^{\dagger,\infty} + \delta\eta$.

3. Approximate ill-posed problem by a well-posed problem in a reasonable, controlled manner. For instance, by variational regularization:

$$u_\lambda^\delta = \underset{u \in \mathcal{F} \subset \mathcal{U}}{\operatorname{argmin}} \left\{ \mathcal{D}(f^{\dagger,\infty}, u) + \lambda \cdot \mathcal{J}(u) \right\}, \quad \text{e.g.} \quad \mathcal{D}(f^{\dagger,\infty}, u) = \|f^{\dagger,\infty} - \mathcal{K}(u)\|_2^2$$

4. Analysis of regularized solution over optimality conditions: Convergence rates? Source conditions? Exact recovery of features?

Some Preliminaries in Notation and Probability

"Distribution" refers to probability distributions.

Various statistical terms are used loosely, hopefully without causing confusion.

Instead of thinking in terms of probability $\mathbb{P}(X = x)$ of the outcome of random variable X, we will often think in terms of the self-information $I(x) = -\log(\mathbb{P}(X = x))$:

- $I(x)$ measures the information content or surprise associated with the outcome "$X = x$".
- $I(x)$ is positive and additive for independent random variables X and Y:

$$\begin{aligned} I(x, y) &= -\log(\mathbb{P}(X = x, Y = y)) = -\log(\mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)) \\ &= -\log(\mathbb{P}(X = x)) - \log(\mathbb{P}(Y = y)) = I(x) + I(y) \end{aligned}$$

- In the spirit of statistical physics, we will also speak of $I(x)$ as the energy

A distribution $p(x) = \exp(-I(x))$ is log-concave, if its energy $I(x)$ is convex.

Inverse Problems from a General Statistical Perspective

Accounting for the stochastic nature of the noise (or modeling it that way):

- Inverse problem becomes a special instance of statistical inference.
- Introduction of complete framework is beyond the scope of this talk.
- u is a *model* of reality.
- Forward modeling, discretization and data contamination are summed up in a *forward mapping* $u \mapsto p_{like}(f|u)$ which links u to a likelihood distribution for $f \in \mathcal{X}$.

Statistical Forward Modeling

Model the data generation process to construct the likelihood distribution.

Examples:

- Additive Gaussian noise: $F = K(u) + E, \quad E \sim \mathcal{N}(0, \Sigma_\varepsilon)$

$$\Longrightarrow p_{like}(f|u) = \mathcal{N}(f; K(u), \Sigma_\varepsilon) \propto \exp\left(-\frac{1}{2}\|\Sigma_\varepsilon^{-1/2}(f - K(u))\|_2^2\right)$$

$\Longrightarrow$ Signal and noise are independent.

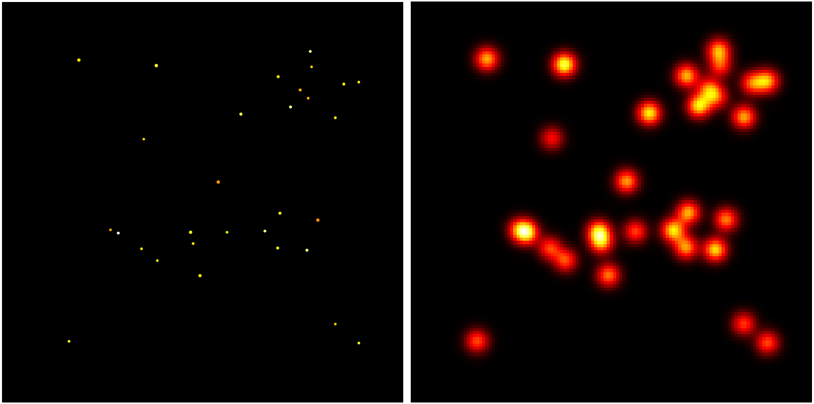
- Poisson (counting) noise: $F \sim \mathrm{Pois}(K(u) + \eta), \quad K(u), \eta \geqslant 0$

$$\Longrightarrow p_{like}(f|u) = \prod_{i=1}^{m} \frac{(K(u)+\eta)_i^{f_i}}{f_i!} \exp\left(-(K(u)+\eta)_i\right)$$

$$\propto \exp\left(-|K(u)+\eta|_1 + f^T \log(K(u)+\eta)\right)$$

$\Longrightarrow$ Signal and noise are not independent.

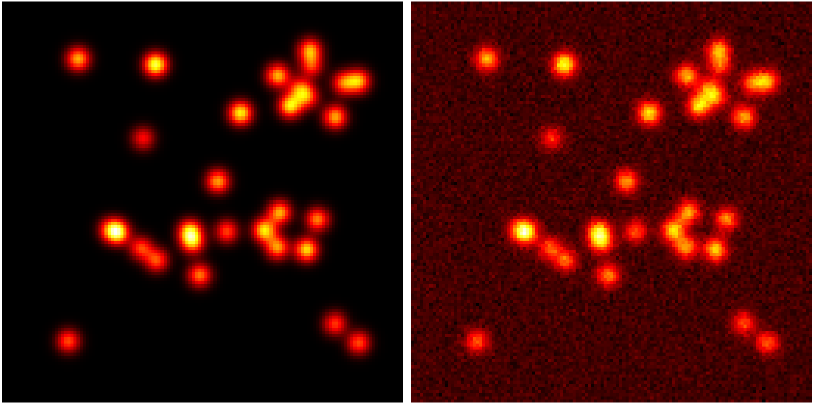
Statistical Forward Modeling



(a) $u^\dagger$ (b) $K(u^\dagger)$

Figure: An image deblurring example

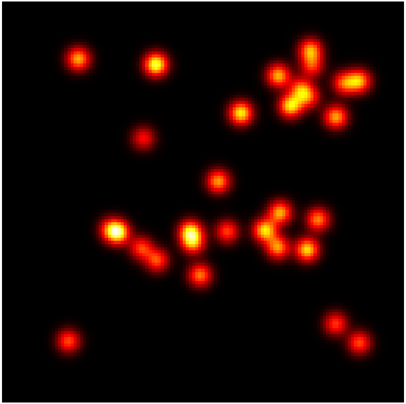
Statistical Forward Modeling



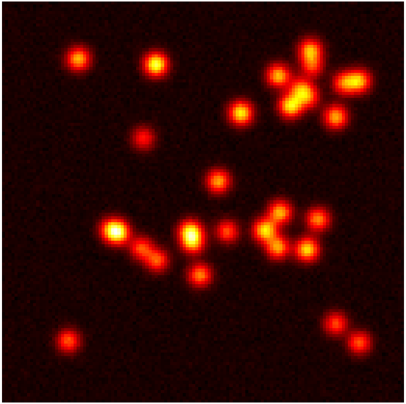
(a) $K(u^\dagger)$ (b) f, $\sigma_{rel} = 3\%$

Figure: Gaussian noise

Statistical Forward Modeling



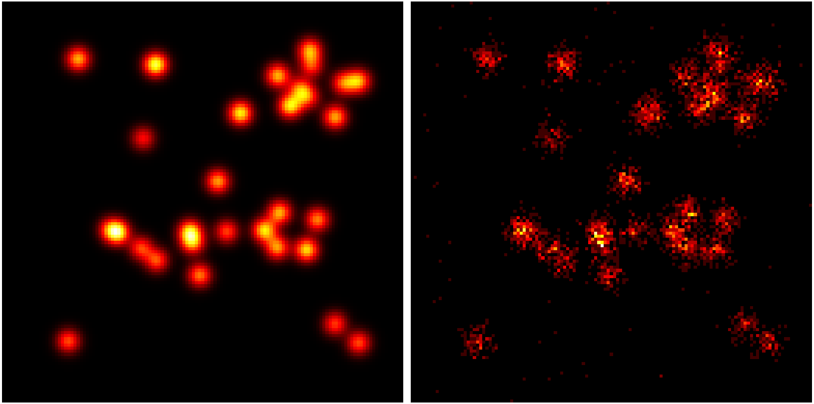
(a) $K(u^\dagger)$



(b) f, $\sigma_{rel} = 1\%$

Figure: Gaussian noise

Statistical Forward Modeling

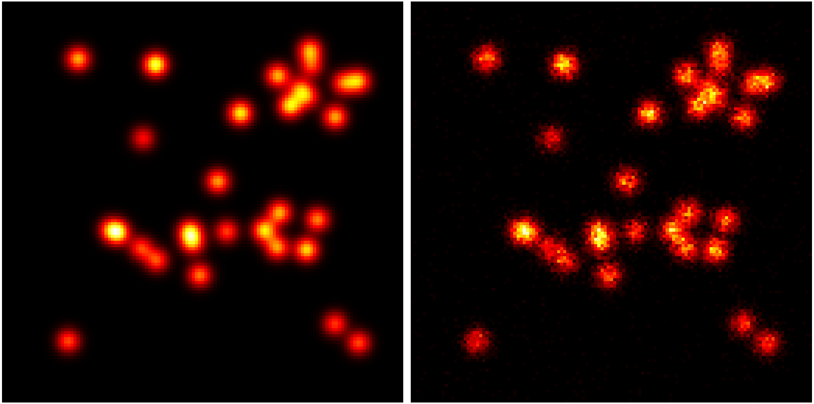


(a) $K(u^\dagger)$

(b) f, short counting time

Figure: Poisson noise, 0.1% background

Statistical Forward Modeling



(a) $K(u^\dagger)$ (b) f, long counting time

Figure: Poisson noise, 0.1% background

Inverse Methods in Statistical Inference

- Inverse methods are estimators, i.e., mapping from $\mathcal{X}$ to $\mathcal{U}$.
- As a consequence, they are random variables.
- Statistical decision theory developed concepts to classify and validate estimators.
- Many of these concepts are closely related to concepts of deterministic regularization.

Naive Maximum Likelihood Estimation

A popular estimator is the maximum likelihood estimator (MLE):

$$\hat{u}_{\text{ML}} := \operatorname*{argmax}_{u \in \mathcal{U}} \; p_{like}(f|u)$$

Doomed to fail for typical ill-posed inverse problems:

Gaussian noise:

$$\begin{aligned}\hat{u}_{\text{ML}} &= \operatorname*{argmax}_{u \in \mathcal{U}} \left\{ \exp\left(-\frac{1}{2} \|\Sigma_\varepsilon^{-1/2}(f - K(u))\|_2^2 \right) \right\} \\ &= \operatorname*{argmin}_{u \in \mathcal{U}} \left\{ \frac{1}{2} \|\Sigma_\varepsilon^{-1/2}(f - K(u))\|_2^2 \right\} \\ &\iff u \quad \text{solves} \quad K(u) = f\end{aligned}$$

i,e, every least-squares solution is an MLE, choose, e.g., the u with smallest L_2 norm $\rightsquigarrow$ minimum norm solution (MNS).

Variance-Bias Decomposition for Statistical Inverse Problems

How does ill-posedness manifest in the statistical framework? Variance-bias decomposition of mean squared error:

$$\mathbb{E}_{u^\dagger}[\|\hat{u}_{\mathrm{ML}} - u^\dagger\|_2^2] = \underbrace{\mathbb{E}_{u^\dagger}[\|\hat{u}_{\mathrm{ML}} - \mathbb{E}_{u^\dagger}[\hat{u}_{\mathrm{ML}}]\|_2^2]}_{\text{Variance}} + \|\underbrace{\mathbb{E}_{u^\dagger}[\hat{u}_{\mathrm{ML}}] - u^\dagger}_{\text{Bias}}\|_2^2$$

- Bias is called approximation error in deterministic theory.
- MLE has no bias but extremely large variance due to instability.
- Regularization lowers variance but introduces bias (usually shrinkage towards some set $\mathcal{D} \subset \mathcal{U}$ $\rightsquigarrow$ shrinkage estimators).

$\Longrightarrow$ well-known trade-off between a stable and a good reconstruction.

Frequentist vs. Bayesian Approach to Statistical Inverse Problems

Frequentist approach: Use statistical decision theory to derive and analyze estimators.

- u is unknown but deterministic.
- Keywords: Ridge or robust regression, penalized maximum likelihood, LASSO, shrinkage estimators.

Bayesian approach: Rethink the concept of probability to incorporate a-priori information on the solution.

- u is a random variable.

Caution: "Statistical inverse problems" $\neq$ "Bayesian inverse problems"!

Evans and Stark, 2002.
Inverse Problems as Statistics.

Outline

Leading Guidelines of Bayesian Inference

Account for high uncertainty of the inverse problem by a probabilistic formulation:

1. All variables are modeled as random variables.

2. This randomness is not a property of the objects to be recovered, but reflects our lack of information about them.

3. Available information about them is explicitly encoded in their probability distributions.

4. The solution of the inverse problem is a posterior probability distribution over the unknowns.

Leading Guidelines of Bayesian Inference

Account for high uncertainty of the inverse problem by a probabilistic formulation:

1. All variables are modeled as random variables.
   $\Longrightarrow$ *Not only f but also u (and $\mathcal{K}$) become random variables*.
2. This randomness is not a property of the objects to be recovered, but reflects our lack of information about them.
3. Available information about them is explicitly encoded in their probability distributions.
4. The solution of the inverse problem is a posterior probability distribution over the unknowns.

Leading Guidelines of Bayesian Inference

Account for high uncertainty of the inverse problem by a probabilistic formulation:

1. All variables are modeled as random variables.
   $\Longrightarrow$ *Not only f but also u (and $\mathcal{K}$) become random variables.*
2. This randomness is not a property of the objects to be recovered, but reflects our lack of information about them.
   $\Longrightarrow$ *We don't introduce randomness, its apparently already there.*
3. Available information about them is explicitly encoded in their probability distributions.
4. The solution of the inverse problem is a posterior probability distribution over the unknowns.

Leading Guidelines of Bayesian Inference

Account for high uncertainty of the inverse problem by a probabilistic formulation:

1. All variables are modeled as random variables.
   $\Longrightarrow$ *Not only f but also u (and $\mathcal{K}$) become random variables.*
2. This randomness is not a property of the objects to be recovered, but reflects our lack of information about them.
   $\Longrightarrow$ *We don't introduce randomness, its apparently already there.*
3. Available information about them is explicitly encoded in their probability distributions.
   $\Longrightarrow$ *"Bayesian modeling";* central point in the Bayesian approach.
4. The solution of the inverse problem is a posterior probability distribution over the unknowns.

Leading Guidelines of Bayesian Inference

Account for high uncertainty of the inverse problem by a probabilistic formulation:

1. All variables are modeled as random variables.
   $\Longrightarrow$ *Not only f but also u (and $\mathcal{K}$) become random variables.*
2. This randomness is not a property of the objects to be recovered, but reflects our lack of information about them.
   $\Longrightarrow$ *We don't introduce randomness, its apparently already there.*
3. Available information about them is explicitly encoded in their probability distributions.
   $\Longrightarrow$ *"Bayesian modeling";* central point in the Bayesian approach.
4. The solution of the inverse problem is a posterior probability distribution over the unknowns.
   $\Longrightarrow$ *by Bayes rule!*

Leading Guidelines of Bayesian Inference

Account for high uncertainty of the inverse problem by a probabilistic formulation:

1. All variables are modeled as random variables.
   $\Longrightarrow$ *Not only f but also u (and $\mathcal{K}$) become random variables.*
2. This randomness is not a property of the objects to be recovered, but reflects our lack of information about them.
   $\Longrightarrow$ *We don't introduce randomness, its apparently already there.*
3. Available information about them is explicitly encoded in their probability distributions.
   $\Longrightarrow$ *"Bayesian modeling";* central point in the Bayesian approach.
4. The solution of the inverse problem is a posterior probability distribution over the unknowns.
   $\Longrightarrow$ *by Bayes rule!*

Bayesian estimation: Infer information of interest from the posterior.

Bayesian Modeling

Central step in the Bayesian approach:

- u becomes a random variable.
- Its distribution $p_{prior}(s)$ (prior) reflects a-priori assumptions/knowledge.
- Task of the prior: Render the estimation problem well-posed.

What information to incorporate into a prior $p_{pr}(u)$, ideally?

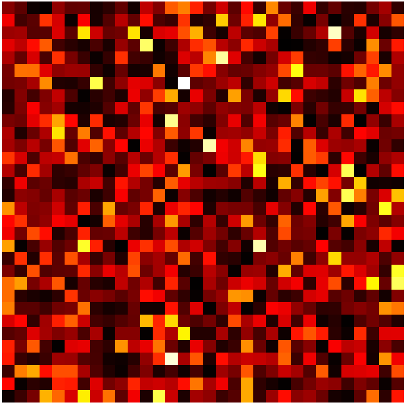
- Representation that captures the most distinct features of u:
  Basis, frames, dictionaries?
- Statistics of u w.r.t. this representation: Moments, decay characteristics (sparsity)?
- "Natural" information.
- Empirical information.
- Known geometrical/contextual features.

Bayesian Modeling

What else to think of?

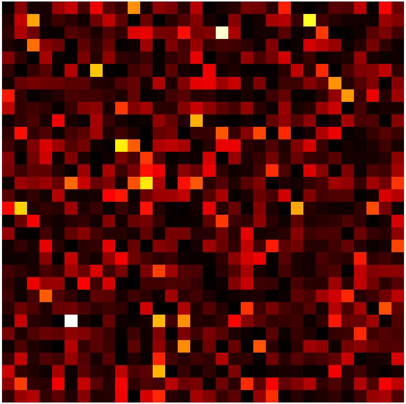
- Prior needs to be sufficiently informative.
- How to choose a prior that introduces *only* the wanted information without unwanted additional information? $\Longrightarrow$ *Maximum entropy probability distributions*.
- Computational convenience: *Exponential families*, *log-concave priors*.

Bayesian Modeling: Simple Examples



Gaussian (L2) prior:

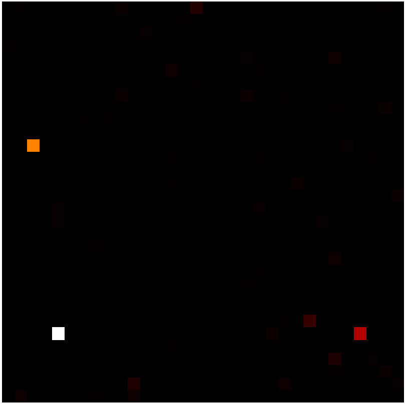
$$p_{pr}(u) \propto \exp\left(-\frac{1}{2}\|u\|_2^2\right)$$



Laplace (L1) prior:

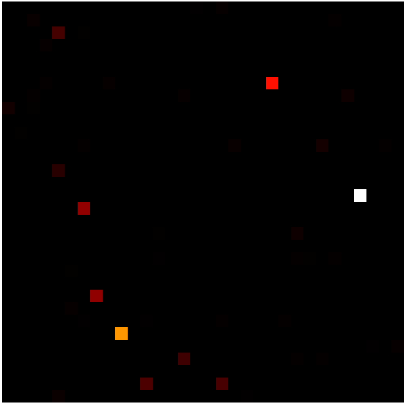
$$p_{pr}(u) \propto \exp\left(-|u|_1\right)$$

Bayesian Modeling: Simple Examples



Cauchy prior:

$$p_{pr}(u) \propto \prod_i \frac{1}{1+u_i^2}$$



Lévy prior:

$$p_{pr}(u) \propto \prod_i \frac{1}{|u_i|^{3/2}} \exp\left(-\frac{1}{2\,|u_i|}\right)$$

Bayesian Modeling: Simple Examples

Use prior on the increments of a function ($u_i := u(t_i)$):

$$p_{pr}(u) \propto \prod_i p^{inc}(|u_{i+1} - u_i|)$$

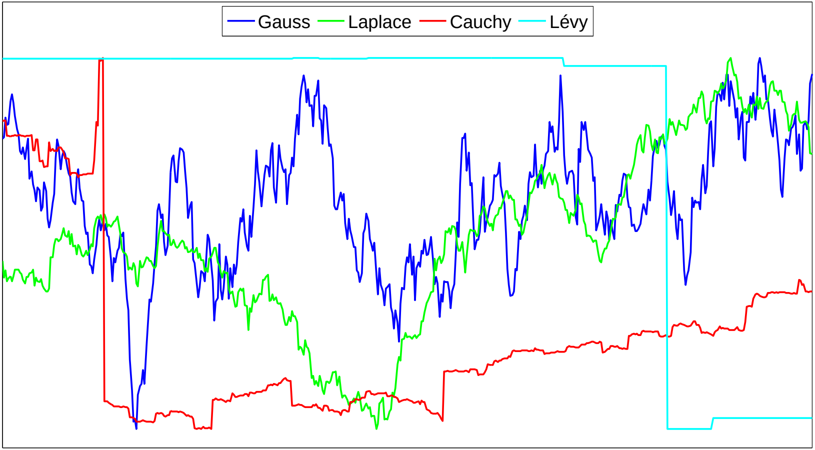
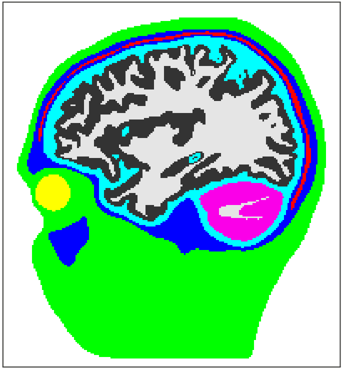


Figure: Different increment priors for $n = 500$ (rescaled and re-centered).

Bayesian Modeling: An Advanced Example

Automatic head tissue segmentation of MRI-images:

T1-weighted MR image



Voxel-based segmentation

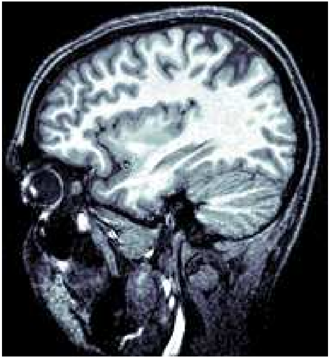
Not an inverse problem in the very strict sense, but a good example for Bayesian modeling.

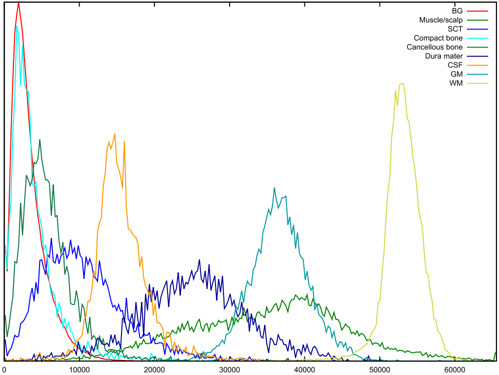
Work by Benjamin Lanfer, BESA GmbH; PhD project with Martin Burger and Carsten Wolters.

Bayesian Modeling: An Advanced Example

Challenges:

- Noise, artifacts & inhomogeneities.
- Strong differences in gray values and image quality across MR machines and sequences.
- Tissue regions with differing gray value characteristics.
- Partial volume effects.
- High dimensional but unstructured and discrete.

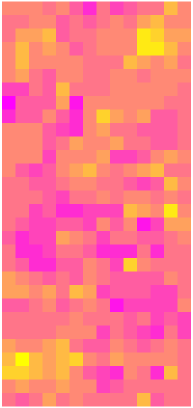




Bayesian Modeling: An Advanced Example

Idea: Use Markov random field prior that incorporates

1. Contextual information based on anatomical knowledge.
2. Geometric, atlas-based (empirical) information.



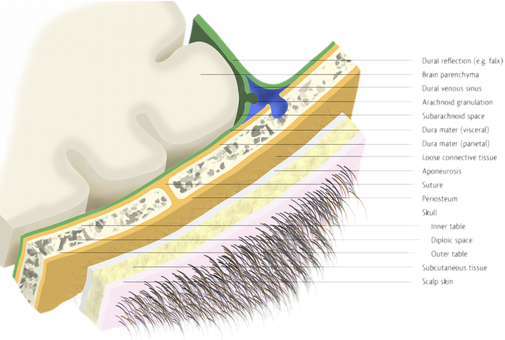
Markov random field (MRF) prior:

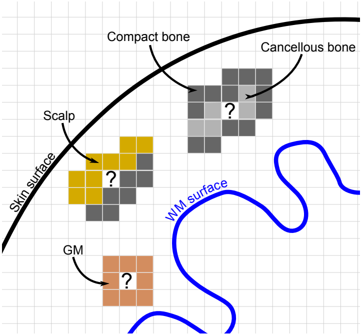
- Popular in image processing and computer vision.
- Example: Discrete total variation (TV) prior, Ising prior.
- Components of u are arranged in a neighborhood system
- Conditional distribution of u_i only depends on the neighboring components:

$$p_{prior}(x_i | x_{[-i]}) = p_{prior}(x_i | x_k, k \in \mathcal{N}_i)$$

Bayesian Modeling: An Advanced Example

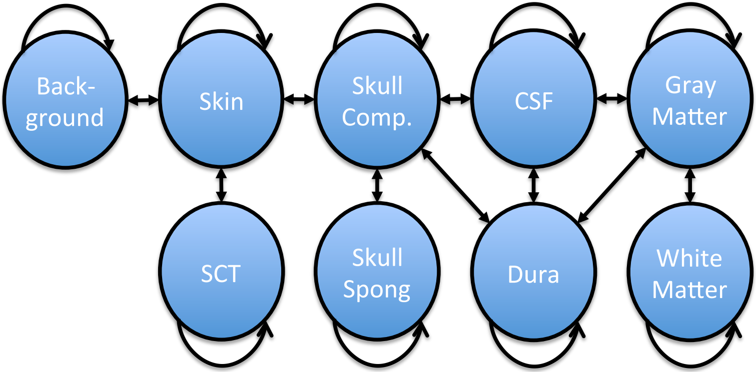
Contextual anatomical knowledge: Layered and nested structure of tissues.





Bayesian Modeling: An Advanced Example

Contextual anatomical knowledge is encoded in graph and implemented in prior

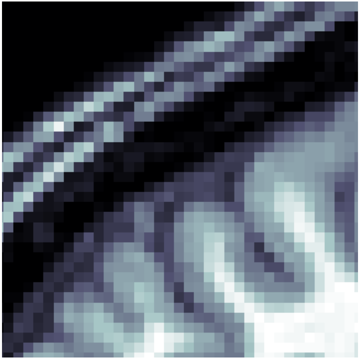


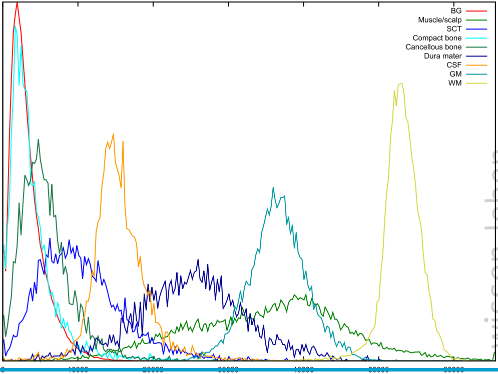
Bayesian Modeling: An Advanced Example

Problem: Several tissue gray values overlap

Qualitative geometric knowledge:

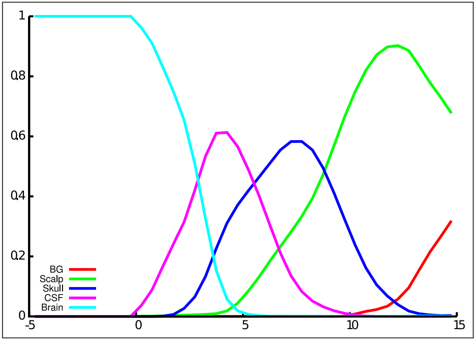
Voxels close to the skin surface are more likely to be skin whereas voxels close to the white matter surface are more likely to be gray matter.

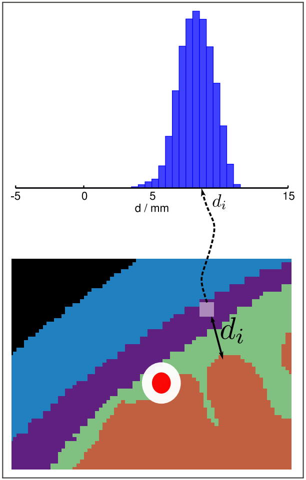




Bayesian Modeling: An Advanced Example

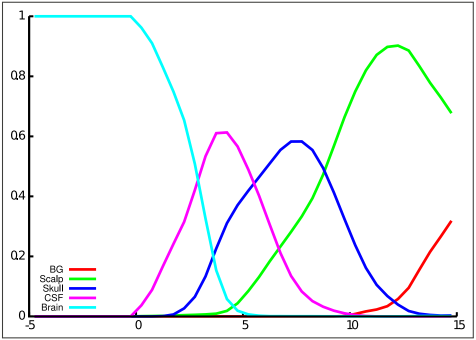
Turn qualitative into quantitative information by deriving empirical histograms from manually segmented images.

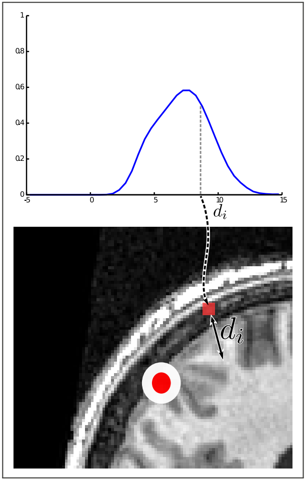




Bayesian Modeling: An Advanced Example

Turn qualitative into quantitative information by deriving empirical histograms from manually segmented images.





More Bayesian Modeling

More recent and fancy examples in the upcoming talks!

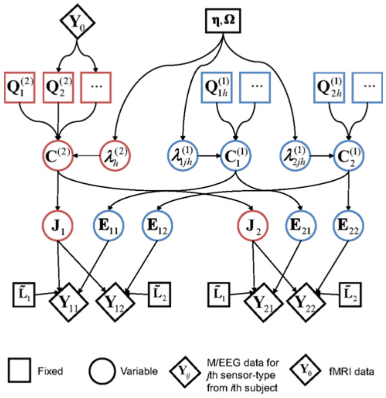
e.g., how and why people draw images like on the right.

Figure 10 from: Henson et al. (2011) in *Frontiers in Human Neuroscience*, 5:76.

Even more of nice examples for various inverse problem scenarios:

- Kaipio and Somersalo, 2005. *Statistical and Computational Inverse Problems*

Multisubject fusion model



Source and sensor space

Step 4.: The Posterior

Merge information before the measurement (prior) with the information by the measurement (likelihood) by Bayes rule:

$$p_{post}(u|f) = \frac{p_{like}(f|u)p_{prior}(u)}{p(f)}$$

- Conditional distribution of u given f is called posterior distribution.
- Represents all information on u after observing f.
- Complete solution to the inverse problem in Bayesian Inference
- $p(f) = \int p_{like}(f|u)p_{prior}(u)\mathrm{d}u$ is called model evidence. Neglected in normal Bayesian inference but central in advanced topics like model comparison / selection / averaging.

Bayesian Inference

Bayesian inference is the process of exploiting the information contained in the posterior.

- Point estimates: Collapse all information into a single point.
- Credible regions estimates: Search for sets that bound the unknowns with certain probability.
- Extreme value probabilities: Given the data, what is the probability that a feature $g(u)$ exceeds some critical value?.
- Conditional covariance estimates: Spatial distribution of variance and dependencies.
- Histogram estimates: Analyze single component distributions.

Outline

Bayesian Inference

The most simple Bayesian inference technique is to derive a point estimate for u from the posterior:

1. Maximum a-posteriori-estimate (MAP):

$$\hat{u}_{\text{MAP}} := \underset{u \in \mathbb{R}^n}{\operatorname{argmax}} \; p_{post}(u|f)$$

   Practically: High-dimensional optimization problem.

2. Conditional mean-estimate (CM):

$$\hat{u}_{\text{CM}} := \mathbb{E}\,[u|f] = \int_{\mathbb{R}^n} u \; p_{post}(u|f) \mathrm{d}u$$

   Practically: High-dimensional integration problem.

Connections to Variational Regularization & Ridge Regression

Consider

$$p_{like}(f|u) \propto \exp\left(-\,\mathcal{D}(f, K(u))\right), \qquad \mathcal{D}(f, K(u)) : \text{negative log-likelihood}$$
$$p_{prior}(u) \propto \exp\left(-\,\lambda \mathcal{J}(u)\right), \qquad \text{"Gibbs prior"}$$

then

$$p_{post}(u|f) \propto \exp\left(-\,\left(\mathcal{D}(f, K(u)) + \lambda \mathcal{J}(u)\right)\right)$$

and

$$\hat{u}_{\mathrm{MAP}} = \underset{u \in \mathcal{U}}{\operatorname{argmin}} \left\{ \mathcal{D}(f, K(u)) + \lambda\, \mathcal{J}(u) \right\}$$

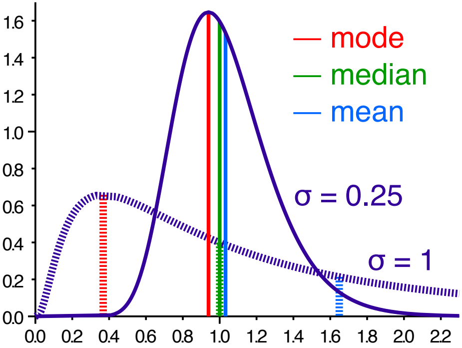
Relation Between MAP and CM Estimates

Immediate Question: Relation between MAP and CM estimate.

In the following, I will summarize.

- The "classical" view on this problem which favors the CM over MAP estimate.
- Observations and discussions which don't support to the classical view.
- Recent work by Martin Burger.

Relation Between MAP and CM Estimates

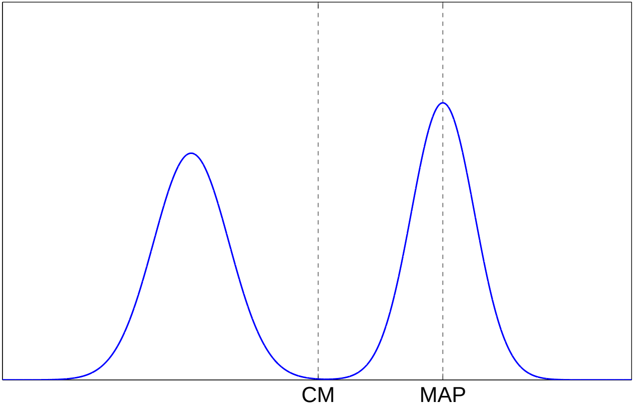


source: Wikimedia Commons

- CM estimate is the mean of the posterior
- MAP estimate the (highest) mode of the posterior.

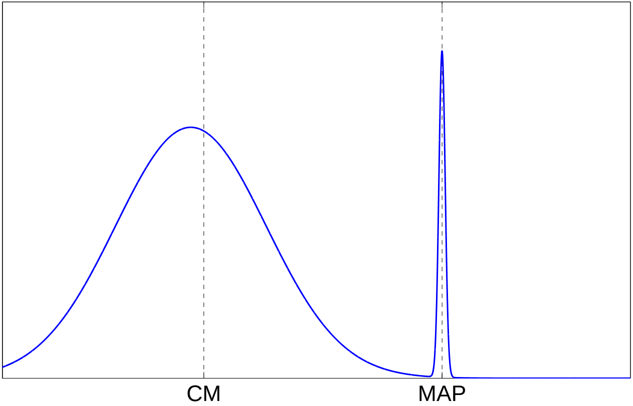
Relation Between MAP and CM Estimates: Classical View

Then people start drawing pictures from hypothetical distributions to show that none is better in general.



Relation Between MAP and CM Estimates: Classical View

Then people start drawing pictures from hypothetical distributions to show there is no general answer to the question which of them is better.



Relation Between MAP and CM Estimates: Classical View

A theoretical argument "decides" the conflict: The Bayes cost formalism.

- An estimator is a random variable, as it relies on f and u.
- How does it perform on average? Which estimator is "best"?
- $\rightsquigarrow$ Define a cost function $\Psi(u, \hat{u}(f))$.
- Bayes cost is the expected cost:

$$BC(\hat{u}) = \iint \Psi(u, \hat{u}(f))\, p_{like}(f|u)\, \mathrm{d}f\, p_{prior}(u)\, \mathrm{d}u$$

- Bayes estimator $\hat{u}_{BC}$ for given Ψ minimizes Bayes cost.

Relation Between MAP and CM Estimates: Classical View

Main classical arguments pro CM and contra MAP estimates:

- CM is Bayes estimator for $\Psi(u, \hat{u}) = \|u - \hat{u}\|_2^2$ (MSE).
- Also the minimum variance estimator.
- The mean value is intuitive, it is the "center of mass", the known "average".
- MAP estimate is asymptotic Bayes estimator of

$$\Psi_\epsilon(u, \hat{u}) = \begin{cases} 0, & \text{if} \quad \|u - \hat{u}\|_\infty \leqslant \epsilon \\ 1 & \text{otherwise,} \end{cases}$$

  for $\epsilon \to 0$ (uniform cost). $\Longrightarrow$ It is not a proper Bayes estimator.

MAP and CM seem theoretically and computationally fundamentally different $\Longrightarrow$ one should decide.

"A real Bayesian would not use the MAP estimate".

Relation Between MAP and CM Estimates: Observations and Discussions

- Theoretical considerations could often not be validated numerically (computational challenges $\rightsquigarrow$ next talk)
- Now, this is possible for non-trivial prior distributions.
- For reasonable priors, they look quite similar. Fundamentally different?
- If a CM estimate looks good, it looks like the MAP estimate.
- MAP estimates are sparser, sharper, look and perform better,...
- Recent theoretical results:
  - Gribonval, 2011: CM are MAP estimates for different priors.
  - Lassas, Siltanen, Saksman: CM has issues with discretization invariance for TV priors (more later).
  - Louchet-Moisan, 2012: More results for TV priors.

Relation Between MAP and CM Estimates: Recent Work by M. Burger

For the case of

- Linear $\mathcal{K}$
- Additive Gaussian noise
- Log-concave prior, i.e., $p_{prior}(u) \propto \exp\left(-\lambda \mathcal{J}(u)\right)$, where $\mathcal{J}(u)$ is convex.

Martin Burger developed several ideas (unpublished) to shed new light on the issue.

He uses Bregman distances as a main tool:

$$D_{\mathcal{J}}^{q}(u, v) = \mathcal{J}(u) - \mathcal{J}(v) - \langle q, u - v \rangle, \qquad q \in \partial \mathcal{J}(v)$$

I will report some key results here.

Relation Between MAP and CM Estimates: Recent Work by M. Burger

"A real Bayesian would not use the MAP estimate as it is not a proper Bayes estimator.".

"MAP estimate is asymptotic Bayes estimator of

$$\Psi_\epsilon(u, \hat{u}) = \begin{cases} 0, & \text{if} \quad \|u - \hat{u}\|_\infty < \epsilon \\ 1 & \text{otherwise,} \end{cases}$$

for $\epsilon \to 0$. ???$\Longrightarrow$??? It is not a proper Bayes estimator."

"MAP estimator is asymptotic Bayes estimator for some degenerate Ψ"
$\not\Rightarrow$ "MAP can't be Bayes estimator for proper Ψ" !!!!

Relation Between MAP and CM Estimates: Recent Work by M. Burger

The MAP estimate is the Bayes estimator for the convex cost function

$$\Psi_{\mathrm{MAP}}(u, \hat{u}) = \|K(\hat{u} - u)\|_2^2 + \lambda D_{\mathcal{J}}(\hat{u}, u)$$

The CM estimate is the Bayes estimator for the convex cost function

$$\Psi_{\mathrm{CM}}(u, \hat{u}) = \|K(\hat{u} - u)\|_2^2 + \beta \|L(\hat{u} - u)\|_2^2$$

for a regular L and $\beta > 0$.

- CM estimate fulfills “average optimality condition”.
- Interesting characterization of difference between CM and MAP estimate by optimality conditions.
- Comparison:

$$\mathbb{E}_{(u|f)} \|L(\hat{u}_{\mathrm{CM}} - u)\|_2^2 \leqslant \mathbb{E}_{(u|f)} \|L(\hat{u}_{\mathrm{MAP}} - u)\|_2^2$$
$$\mathbb{E}_{(u|f)} D_{\mathcal{J}}(\hat{u}_{\mathrm{MAP}}, u) \leqslant \mathbb{E}_{(u|f)} D_{\mathcal{J}}(\hat{u}_{\mathrm{CM}}, u)$$

Outline

Infinite Dimensional Bayesian Inversion

Extension of the Bayesian approach to function spaces:
"First infer, then discretize"

- Analyzing the effects of discretization, approximation and measurement setup.
- New insights for practical challenges in high dimensional inversion?
- Essential for consistent multi-level / multi-grid algorithms.

Challenges:

- No analogue of Lebesgue measure on an infinite-dimensional Banach space.
- Discretization invariance: A-priori knowledge should be independent of m and n.
  $\Longrightarrow$ Which priors represent the same a-prior information for all m and n and in the limits $m, n \longrightarrow \infty$?

Infinite Dimensional Bayesian Inversion

A (very incomplete) collection of publications on this issue:

- Matti Lassas and Samuli Siltanen, 2004.
  Can one use total variation prior for edge-preserving Bayesian inversion?
- Matti Lassas, Eero Saksman and Samuli Siltanen, 2009.
  Discretization-invariant Bayesian inversion and Besov space priors
- Tapio Helin 2009.
  On infinite-dimensional hierarchical Bayesian probability models in statistical inverse problems
- Andrew Stuart, 2009.
  Inverse Problems: A Bayesian Perspective
- Masoumeh Dashti, Stephen Harris and Andrew Stuart, 2012.
  Besov priors for Bayesian inverse problems

Treating Nuisance Parameters in the Bayesian Approach

Models might contain parameters that are uncertain:

- Noise parameters
- Parameters of the forward problem, usually
  - PDE coefficients in linear problems.
  - source/boundary terms in non-linear problems.
  - sensor parameters.
- Discretization errors can also be regarded as nuisance parameters as well.

Their choice might be crucial but difficult in practical scenarios.

Bayesian paradigm:

- All parameters of the model that carry uncertainty are modeled as random variables.
- Our knowledge about their values is modeled by a prior.
- They are subject to Bayesian inference techniques as well!

Treating Nuisance Parameters in the Bayesian Approach

Example: i.i.d. Gaussian measurement noise with unknown variance σ^2

Choose the prior on σ^2, e.g., the conjugate prior, i.e., the inverse gamma distribution:

$$p_{prior}(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\alpha-1} \exp\left(-\frac{\beta}{\sigma^2}\right)$$

with *shape* and *scale* parameters α and β to incorporate a-priori information (by calibration).

Joint posterior over u and σ^2 given f :

$$p_{post}(u, \sigma^2 | f) \propto p_{like}(f | u, \sigma^2) p_{prior}(\sigma^2) p_{prior}(u)$$

- Marginalization: Integrate out the uncertainty in σ^2 (form of generalized error propagation)
- Joint inference: To improve subsequent measurements, to inform subsequent inversions.

Practically, done in the same way.

Treating Nuisance Parameters: Approximation Error Modeling

A recent Bayesian technique to model errors w.r.t. the forward mapping K:

- Discretization / model reduction $K \longrightarrow K_N$
- Linearization in non-linear problems $K(u) \longrightarrow Ku$
- Unknown parameters: $K^{\gamma^\dagger} \longrightarrow K^\gamma$

$$\begin{aligned} f &= K^{\gamma^\dagger}(u) + \varepsilon \\ &= K^{\gamma^\dagger}(u) + K_N^\gamma u_N - K_N^\gamma u_N + \varepsilon \\ &= K_N^\gamma u_N + (K^{\gamma^\dagger}(u) - K_N^\gamma u_N) + \varepsilon \\ &= K_N^\gamma u_N + \underbrace{(K^{\gamma^\dagger}(u) - K_N^\gamma P_N u)}_{:=\eta} + \varepsilon \end{aligned}$$

$\rightsquigarrow$ Approximation error $\eta = \eta(u, \gamma^\dagger)$: Exact value unknown. But...
...what about its statistics?

Treating Nuisance Parameters: Approximation Error Modeling

Bayesian approach: We have $p_{prior}(u)$, $p_{prior}(\gamma)$, can we

- infer marginalized posterior on u_N: complete error model.
- Enhanced error model: Assume independence and Gaussian shape of η: $\eta \sim \mathcal{N}(\mu, \Sigma_\eta)$

  $$\mu = \mathbb{E}_{(u,\gamma)}[K^{\gamma^\dagger}(u) - K_N^\gamma P_N u]; \quad \Sigma_\eta = \mathbb{C}\mathrm{ov}_{(u,\gamma)}[K^{\gamma^\dagger}(u) - K_N^\gamma P_{n,N} u]$$

  Now modifying the measurement noise statistics:

  $$\bar{\varepsilon} := \eta + \varepsilon \sim \mathcal{N}(\mu, \Sigma_{\bar{\varepsilon}} := \Sigma_\varepsilon + \Sigma_\eta)$$

  Practical computation might involve Monte Carlo integration ($\rightsquigarrow$ next talk)
- Promising first results for electrical impedance tomography (EIT) and optical tomography.

Treating Nuisance Parameters: Approximation Error Modeling

Starting literature on this issue:

Jari P. Kaipio and Erkki Somersalo, 2005.
Statistical and Computational Inverse Problems.

Jari P. Kaipio and Erkki Somersalo, 2007.
Statistical inverse problems: Discretization, model reduction and inverse crimes.

Antti Nissinen, 2011 (PhD thesis).
Modelling Errors in Electrical Impedance Tomography.

Ville Kolehmainen, Tanja Tarvainen, Simon R. Arridge, Jari P. Kaipio, 2011.

Marginalization of uninteresting parameters in inverse problems - application to diffuse optical tomography

Outline

Take Home Messages

- The deterministic, statistical and Bayesian perspective on inverse problems are closely related.
- Every interpretation is more or less convenient and intuitive to deal with certain techniques / sub-problems.

Take Home Messages

- The deterministic, statistical and Bayesian perspective on inverse problems are closely related.
- Every interpretation is more or less convenient and intuitive to deal with certain techniques / sub-problems.
- The Bayesian approach
  - emphasizes on quantifying uncertainty and information on all stages of the problem by the use of probability distributions.
  - offers convenient ways for incorporating complex a-priori information.
  - offers more interesting inference strategies than simple point estimates.
  - offers principled ways to account for additional uncertainties appearing in practical applications.

Take Home Messages

- The deterministic, statistical and Bayesian perspective on inverse problems are closely related.
- Every interpretation is more or less convenient and intuitive to deal with certain techniques / sub-problems.
- The Bayesian approach
  - emphasizes on quantifying uncertainty and information on all stages of the problem by the use of probability distributions.
  - offers convenient ways for incorporating complex a-priori information.
  - offers more interesting inference strategies than simple point estimates.
  - offers principled ways to account for additional uncertainties appearing in practical applications.
- MAP estimates are proper Bayesian estimates, one does not have to be ashamed to use them.

Take Home Messages

- The deterministic, statistical and Bayesian perspective on inverse problems are closely related.
- Every interpretation is more or less convenient and intuitive to deal with certain techniques / sub-problems.
- The Bayesian approach
  - emphasizes on quantifying uncertainty and information on all stages of the problem by the use of probability distributions.
  - offers convenient ways for incorporating complex a-priori information.
  - offers more interesting inference strategies than simple point estimates.
  - offers principled ways to account for additional uncertainties appearing in practical applications.
- MAP estimates are proper Bayesian estimates, one does not have to be ashamed to use them.
- "Bayesian" refers to the interpretation behind this approach, not to using Bayes rule.

Concluding Advice

The Bayesian approach might be interesting for you if...

- you're interested in more than point estimates: Collapsing the posterior information into a single point was not worth the effort.
- you look for convenient ways to incorporate "non-standard" a-priori information.
- you are happy about every bit of information you can get:
  - Strongly non-linear problems
  - Little structure in the problems
  - Severely ill-posed problems
- exact bookkeeping and propagation of uncertainty is of interest.
- you're facing practical difficulties like nuisance parameters.

Thank you for your attention!

Jari Kaipio and Erkki Somersalo.
Statistical and Computational Inverse Problems,
Volume 160 of Applied Mathematical Sciences. Springer New York, 2005.

Andrew Gelman, John B. Carlin, Hal S. Stern and Donald B. Rubin
Bayesian Data Analysis
Volume 19 of Acta Numerica, 451-559 . Cambridge University Press,, 2010.

Andrew Stuart.
Inverse Problems: A Bayesian Perspective
Texts in Statistical Science, CRC Press, 2003 (2nd).