Statistical and Computational Inverse Problems with Applications

Part 5: Bayesian inverse problems

Aku Seppänen

Inverse Problems Group
Department of Applied Physics
University of Eastern Finland
Kuopio, Finland

Jyväskylä Summer School
August 11-13, 2014

Bayesian inverse problems

- Model for observations G

$$G = h(F, N)$$

  where G = observations, F = unknown (of interest), N = observation noise.
- In Bayesian framework, G, F and N are modeled as random variables.
- We denote realizations of G, F and N by g, f and n, respectively.

- The random variables G and F have a joint probability density $\pi(f, g)$
- **The prior density** $\pi(f)$ expresses what we know about the unknown prior to the measurements.

$$\pi_{\mathrm{pr}}(f) = \int_{\mathbb{R}_M} \pi(f, g)\mathrm{d}g$$

- If we would know the value of the unknown, that is, $F = f$, the conditional probability density of G given this information, would be

$$\pi(g|f) = \frac{\pi(f, g)}{\pi_{\mathrm{pr}}(f)}, \quad \text{if } \pi_{\mathrm{pr}}(f) \neq 0$$

The conditional probability of G is called **the likelihood function**, because it expresses the likelihood of different measurement outcomes with $F = f$ given.

- Assume finally that the measurement data $G = g_{\text{obs}}$ is given. The conditional probability distribution

$$\pi(f|g_{\text{obs}}) = \frac{\pi(f, g_{\text{obs}})}{\pi(g_{\text{obs}})} = \frac{\pi(g_{\text{obs}}|f)\pi_{\text{pr}}(f)}{\pi(g_{\text{obs}})}$$

  is called the **posterior distribution** of F. This distribution expresses what we know about F after the realized observation $G = g_{\text{obs}}$.
- In the Bayesian framework, the inverse problem is expressed in the following way: Given the data $G = g_{\text{obs}}$, find the conditional probability distribution $\pi(f|g_{\text{obs}})$ of the variable F.

- Bayes' theorem of inverse problems:

$$\pi_{\text{post}}(f) = \pi(f|g_{\text{obs}}) = \frac{\pi(g_{\text{obs}}|f)\pi_{\text{pr}}(f)}{\pi(g_{\text{obs}})}$$

- The posterior density $\pi_{\text{post}}(f)$ is the full solution of the Bayesian inverse problem.
- Note: $\pi(g_{\text{obs}})$ does not depend on F; acts as a normalizing constant – usually unimportant.
- In summary, solving an inverse problem may be broken into three subtasks:
    1. Based on all the prior information of the unknown F, find a prior probability density $\pi_{\text{pr}}(f)$ that reflects judiciously this prior information.
    2. Find the likelihood function $\pi(g_{\text{obs}}|f)$ that describes the interrelation between the observation and the unknown.
    3. Develop methods to explore the posterior probability density.

Point estimates

- Computing the **maximum a posteriori estimate** f_{MAP}:

$$f_{\mathrm{MAP}} = \arg\max_{f} \pi(f|g_{\mathrm{obs}})$$

  is an optimization problem.

- Computing the **conditional mean estimate** f_{CM}:

$$f_{\mathrm{CM}} = \mathbb{E}\{f|g_{\mathrm{obs}}\} = \int f\pi(f|g_{\mathrm{obs}})\mathrm{d}f$$

  is an integration problem (usually MCMC).

- Note: the maximum likelihood (ML) estimate f_{ML} is a non-Bayesian estimator

$$f_{\mathrm{ML}} = \arg\max_{f} \pi(g_{\mathrm{obs}}|f)$$

- f_{ML} answers the question "Which value of the unknown is most likely to produce the measured data y?". In ill-posed inverse problems, f_{ML} is quite useless: It often corresponds to solving the classical inverse problem without regularization.

Spread estimates

- The conditional (posterior) covariance

$$\Gamma_{f|g_{\mathrm{obs}}} = \int (f - f_{\mathrm{CM}})(f - f_{\mathrm{CM}})^{\mathrm{T}} \pi(f|g_{\mathrm{obs}}) \mathrm{d}f$$

- Integration problem (usually MCMC).
- Also other spread estimates are used: credibility limits, etc.

Likelihood models, additive noise case

- Additive noise model

$$G = h(F) + N$$

  where F and N are mutually independent.
- $\pi(n)$ is assumed to be known.
- When $F = f$ is fixed $\Rightarrow$ $g = \underbrace{h(f)}_{\text{fixed}} + n$

  $\Rightarrow$ G is distributed as N; only shifted by $h(f)$
  $\Rightarrow \pi(g|f) = \pi_{\text{noise}}(g - h(f))$
- Then posterior

$$\pi(f|g_{\text{obs}}) \propto \pi(f)\pi(g_{\text{obs}}|f) = \pi_{\text{pr}}(f)\pi_{\text{noise}}(g_{\text{obs}} - h(f))$$

Additive Gaussian noise

- Gaussian noise $N \sim \mathcal{N}(0, \Gamma_n)$

$$\pi_{\mathrm{noise}}(n) \propto \exp(-\frac{1}{2} n^{\mathrm{T}} \Gamma_n^{-1} n)$$

$$\pi(g_{\mathrm{obs}} | f) \propto \exp\left(-\frac{1}{2}(g_{\mathrm{obs}} - h(f))^{\mathrm{T}} \Gamma_n^{-1} (g_{\mathrm{obs}} - h(f))\right)$$

- If further, a Gibbs-type prior

$$\pi(f) \propto \exp(-A(f))$$

- Then,

$$\begin{aligned} \pi(f | g_{\mathrm{obs}}) &\propto \pi(g_{\mathrm{obs}} | f) \pi(f) \\ &= \exp\left(-\frac{1}{2}(g_{\mathrm{obs}} - h(f))^{\mathrm{T}} \Gamma_n^{-1} (g_{\mathrm{obs}} - h(f)) - A(f)\right) \end{aligned}$$

- in this case, the MAP estimate is of the form

$$\begin{aligned} f_{\mathrm{MAP}} &= \arg\max_f \pi(f|g_{\mathrm{obs}}) \\ &\propto \arg\min_f \left\{ \frac{1}{2}(g_{\mathrm{obs}} - h(f))^{\mathrm{T}} \Gamma_n^{-1} (g_{\mathrm{obs}} - h(f)) + A(f) \right\} \\ &= \arg\min_f \left\{ \frac{1}{2}(g_{\mathrm{obs}} - h(f))^{\mathrm{T}} L_n^{\mathrm{T}} L_n (g_{\mathrm{obs}} - h(f)) + A(f) \right\} \\ &= \arg\min_f \left\{ \frac{1}{2} \| L_n (g_{\mathrm{obs}} - h(f)) \|^2 + A(f) \right\} \end{aligned}$$

where $L_n^{\mathrm{T}} L_n = \Gamma_n^{-1}$.

- If Gaussian prior model $F \sim \mathcal{N}(\eta_f, \Gamma_f)$

$$\pi_{\mathrm{pr}}(f) \propto \exp\left(-\frac{1}{2}(f-\eta_f)^{\mathrm{T}}\Gamma_f^{-1}(f-\eta_f)\right)$$

- Then

$$\begin{aligned} f_{\mathrm{MAP}} &= \arg\max_f \pi(f|g_{\mathrm{obs}}) \\ &\propto \arg\min_f \left\{\frac{1}{2}(g_{\mathrm{obs}}-h(f))^{\mathrm{T}}\Gamma_n^{-1}(g_{\mathrm{obs}}-h(f)) \right. \\ &\qquad \left. +\frac{1}{2}(f-\eta_f)^{\mathrm{T}}\Gamma_f^{-1}(f-\eta_f)\right\} \\ &= \arg\min_f \left\{\frac{1}{2}\|L_n(g_{\mathrm{obs}}-h(f))\|^2 + \|L_f(f-\eta_f)\|^2\right\} \end{aligned}$$

which is equivalent to Generalized Tikhonov regularization with choices $\alpha L_\alpha = L_f,\ L_f^{\mathrm{T}} L_f = \Gamma_f^{-1}$ and $f_* = \eta_f = \mathbb{E}(f)$.

- Similarly, the maximum likelihood estimate

$$
\begin{aligned}
f_{\mathrm{ML}} &= \arg\max_f \pi(g_{\mathrm{obs}}|f) \\
&= \arg\min_f \left\{ \frac{1}{2} \| L_n(g_{\mathrm{obs}} - h(f)) \|^2 \right\}
\end{aligned}
$$

  which is equivalent to Gauss-Markov estimate
- Non-regularized, does not work with ill-posed inverse problems.

- If further, the observation model is linear

$$G = KF + N$$

- Then

$$\pi(f|g_{\mathrm{obs}}) \quad \propto \quad \arg\min_f \left\{ \frac{1}{2}(g_{\mathrm{obs}} - Kf)^{\mathrm{T}}\Gamma_n^{-1}(g_{\mathrm{obs}} - Kf) \right. \\ \left. + \frac{1}{2}(f - \eta_f)^{\mathrm{T}}\Gamma_f^{-1}(f - \eta_f) \right\}$$

is Gaussian, and it can shown, that

$$f_{\mathrm{MAP}} = f_{\mathrm{CM}} = (K^{\mathrm{T}}\Gamma_n^{-1}K + \Gamma_f^{-1})^{-1}(K^{\mathrm{T}}\Gamma_n^{-1}g_{\mathrm{obs}} + \Gamma_f^{-1}\eta_f)$$

$$\Gamma_{\mathrm{post}} = (K^{\mathrm{T}}\Gamma_n^{-1}K + \Gamma_f^{-1})^{-1}$$

- See **Matlab example 5.1**.

Computation of the integration based estimates

- Many estimators are of the form

$$\int u(f)\pi(f|g_{\mathrm{obs}})\mathrm{d}f$$

- For $f_{\mathrm{CM}},\ u(f) = f$
- For $\Gamma_{f|g},\ u(f) = (f - f_{\mathrm{CM}})(f - f_{\mathrm{CM}})^{\mathrm{T}}$
- Analytical evaluation in most cases impossible
- Traditional numerical quadratures not applicable when N is large (number of points needed unreasonably large, support of $\pi(f|g_{\mathrm{obs}})$ may not be well known) $\Rightarrow$ Monte Carlo integration.

Monte Carlo integration

- Monte Carlo integration
  1. Draw an ensemble $\{f^{(k)},\ k = 1, \ldots, Q\}$ of i.i.d samples from $\pi_{\text{post}}(f)$
  2. Estimate
$$\int u(f)\pi(f|g_{\text{obs}})\mathrm{d}f \approx \frac{1}{Q}\sum_{k=1}^{Q} u(f^{(k)})$$
- Direct sampling from $\pi_{\text{post}}(f)$ usually not possible $\Rightarrow$ Markov Chain Monte Carlo (MCMC) integration:
  1. Draw $\{f^{(k)},\ k = 1, \ldots, Q\}$ by simulating a Markov chain (with equilibrium distribution $\pi_{\text{post}}(f)$)
  2. Estimate
$$\int u(f)\pi(f|g_{\text{obs}})\mathrm{d}f \approx \frac{1}{Q}\sum_{k=1}^{Q} u(f^{(k)})$$
- Algorithms for MCMC: Metropolis-Hastings algorithm, Gibbs sampler

Metropolis-Hastings

- Generation of an ensemble $\{f^{(k)},\ k = 1, \ldots, Q\} \sim \pi_{\text{post}}(f)$ using Metropolis-Hastings algorithm:
  1. Pick an initial value $f^{(1)}$ and set $\ell = 1$
  2. Set $f = f^{(\ell)}$.
  3. Draw a candidate sample f' from proposal density
  $$f' \sim q(f, f')$$
  and compute the acceptance factor
  $$\alpha(f, f') = \min\left(1, \frac{\pi_{\text{post}}(f')q(f', f)}{\pi_{\text{post}}(f)q(f, f')}\right)$$
  4. Draw $t \in [0, 1]$ from uniform probability distribution ($t \sim \text{uni}(0, 1)$).
  5. If $\alpha(f, f') \geq t$, set $f^{(\ell+1)} = f'$, else $f^{(\ell+1)} = f$. Increment $\ell \to \ell + 1$.
  6. When $\ell = Q$ stop, else repeat from step 2.

- Great flexibility in choosing the proposal density $q(f, f')$; almost any density would do the job (eventually).
- However, the choice of $q(f, f')$ is a crucial part of successful Metropolis-Hastings MCMC; it determines the efficiency of the algorithm
- **Matlab example 5.2**
  - $f \in \mathbb{R}^2$, and posterior

$$\pi_{\text{post}}(f) \propto \exp\left\{-10(f_1^2 - f_2)^2 - (f_2 - \frac{1}{4})^4\right\}$$

  - We choose the random walk proposal distribution

$$q(f, f') \propto \exp\left(-\frac{1}{2\gamma^2}\|f' - f\|^2\right)$$

  Note: for this choice $q(f, f') = q(f', f)$.