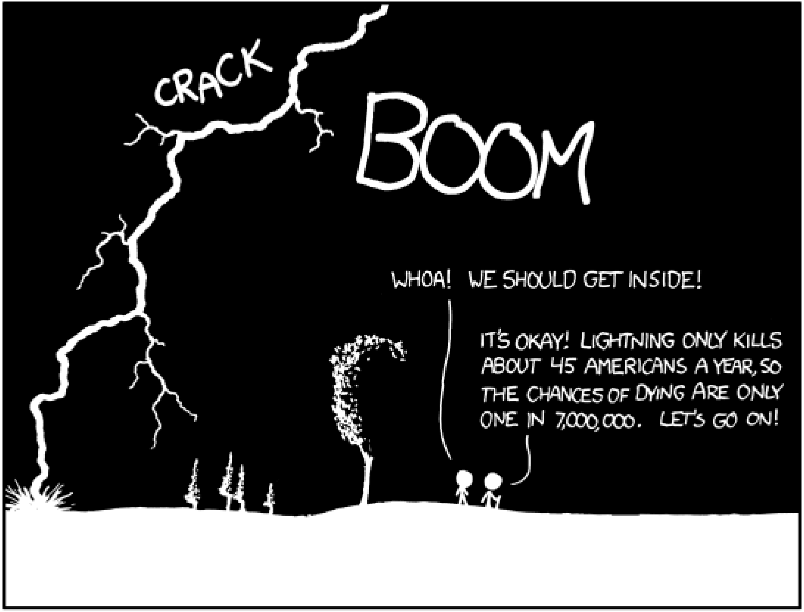
Bayesian inverse problems II: Bayes' Theorem and Well-posedness

Jonas Latz

Department of Mathematics, University of Manchester
www.latzplacian.org

The University of Manchester

Outline



THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT STATISTIC IS ONE IN SIX.

What has happened so far...

Consider the inverse problem of the form

$$\mathcal{A}(f^*) + e = g,$$

where $f^* \in X$ is the unknown parameter/signal/image, $g \in Y$ is observed data, e is noise, and $\mathcal{A} : X \to Y$ is a linear or non-linear operator.

- we model f^* and e as random variables $F \sim \mu_{\text{prior}}$ and $E \sim \mu_{\text{noise}}$, respectively, to especially represent our reasonable expectation about f^*
- Bayesian idea: obtain the reasonable expectation that incorporates the observation, given by the so-called posterior:

$$\mathbb{P}(F \in \cdot | \mathcal{A}(F) + E = g)$$

- Representing conditional distributions is a bit tough, conditional densities are nice though!
- We have generalised the setting a bit, by defining a more general problem:
  - $F : \Omega \to X$, $G : \Omega \to Y$ are (dependent) random variables.
  - Given that we observe $G = g$, find F.

What has happened so far...

Consider the inverse problem of the form

$$\mathcal{A}(f^*) + e = g,$$

where $f^* \in X$ is the unknown parameter/signal/image, $g \in Y$ is observed data, e is noise, and $\mathcal{A} : X \to Y$ is a linear or non-linear operator.

- we model f^* and e as random variables $F \sim \mu_{\text{prior}}$ and $E \sim \mu_{\text{noise}}$, respectively, to especially represent our reasonable expectation about f^*
- Bayesian idea: obtain the reasonable expectation that incorporates the observation, given by the so-called posterior:

$$\mathbb{P}(F \in \cdot | \mathcal{A}(F) + E = g)$$

- Representing conditional distributions is a bit tough, conditional densities are nice though!
- We have generalised the setting a bit, by defining a more general problem:
  - $F : \Omega \to X$, $G : \Omega \to Y$ are (dependent) random variables.
  - Given that we observe $G = g$, find F.

Recap and outlook

Introducing Bayes' Theorem

- Bayes' Theorem will give us a strategy to obtain the posterior density
- More than anything, it is an inversion formula, it let's us flip a conditional density:

$$\pi_{F|G=g} \overset{\leftarrow}{\rightarrow} \pi_{G|F=f}$$

- so, we first need to access $\pi_{G|F=f}$
  - we refer to $L(g|f) := \pi_{G|F=f}(g)$ as (data) likelihood of g given f
  - it describes how well the parameter f describes the data g (or, how likely is the data f if the chosen parameter is g?)
  - the observed data is a sample from the data generating distribution

$$g \sim L(\cdot|f^*).$$

Introducing Bayes' Theorem

- Bayes' Theorem will give us a strategy to obtain the posterior density
- More than anything, it is an inversion formula, it let's us flip a conditional density:

$$\pi_{F|G=g} \leftrightarrows \pi_{G|F=f}$$

- so, we first need to access $\pi_{G|F=f}$
  - we refer to $L(g|f) := \pi_{G|F=f}(g)$ as (data) likelihood of g given f
  - it describes how well the parameter f describes the data g (or, how likely is the data f if the chosen parameter is g?)
  - the observed data is a sample from the data generating distribution

$$g \sim L(\cdot|f^*).$$

Introducing Bayes' Theorem

- Bayes' Theorem will give us a strategy to obtain the posterior density
- More than anything, it is an inversion formula, it let's us flip a conditional density:

$$\pi_{F|G=g} \overset{\leftarrow}{\rightarrow} \pi_{G|F=f}$$

- so, we first need to access $\pi_{G|F=f}$
  - we refer to $L(g|f) := \pi_{G|F=f}(g)$ as (data) likelihood of g given f
  - it describes how well the parameter f describes the data g (or, how likely is the data f if the chosen parameter is g?)
  - the observed data is a sample from the data generating distribution

$$g \sim L(\cdot|f^*).$$

Introducing Bayes' Theorem

- Bayes' Theorem will give us a strategy to obtain the posterior density
- More than anything, it is an inversion formula, it let's us flip a conditional density:

$$\pi_{F|G=g} \overset{\leftarrow}{\rightarrow} \pi_{G|F=f}$$

- so, we first need to access $\pi_{G|F=f}$
  - we refer to $L(g|f) := \pi_{G|F=f}(g)$ as (data) likelihood of g given f
  - it describes how well the parameter f describes the data g (or, how likely is the data f if the chosen parameter is g?)
  - the observed data is a sample from the data generating distribution

$$g \sim L(\cdot|f^*).$$

Introducing Bayes' Theorem

- Bayes' Theorem will give us a strategy to obtain the posterior density
- More than anything, it is an inversion formula, it let's us flip a conditional density:

$$\pi_{F|G=g} \overset{\leftarrow}{\rightarrow} \pi_{G|F=f}$$

- so, we first need to access $\pi_{G|F=f}$
  - we refer to $L(g|f) := \pi_{G|F=f}(g)$ as (data) likelihood of g given f
  - it describes how well the parameter f describes the data g (or, how likely is the data f if the chosen parameter is g?)
  - the observed data is a sample from the data generating distribution

$$g \sim L(\cdot|f^*).$$

Likelihoods in inverse problems

In our example with additive noise,

$$\mathcal{A}(F) + E = g,$$

with $Y := \mathbb{R}^d$ and $E \sim \mathrm{N}(0, \Gamma)$, we have

$$L(f|g) \propto \exp\left(-\frac{1}{2}\|\Gamma^{-1/2}(\mathcal{A}(f) - g)\|^2\right)$$

Observations:

- Considering this likelihood, we retrieve the classical least-squares formulation of an inverse problem when applying maximum likelihood estimation
- in statistics, we call a statistical model parametric if $X \subseteq \mathbb{R}^n$; our theory works for fairly general 'parameters' $f \in X$ that I will continue to call so.
- If Y is infinite-dimensional the definition of a likelihood is a bit tougher, as there is no Lebesgue measure. If applicable (non-trivial!), the Cameron–Martin Theorem can be used to obtain a likelihood

Bayes' Theorem.

Having understood likelihoods, we can now move on to Bayes' Theorem:

Bayes' Theorem.

Let π_{prior} be the density of the prior with respect to our reference measure ν_X. If $L(g|\cdot) \in \mathbf{L}^1(X, \mu_{\mathrm{prior}}; \mathbb{R})$ and $L(g|\cdot) > 0$ for $g \in Y$ $\mathbb{P}(G \in \cdot)$-a.s., we have

$$Z(g) := \int_X L(g|f) \mathrm{d}\mu_{\mathrm{prior}}(f) \in (0, \infty)$$

and μ_{post} has a density with respect to ν_X, which is given by

$$\pi_{\mathrm{post}} = \frac{L(g|\cdot)\pi_{\mathrm{prior}}}{Z(g)} \qquad (\nu_X\text{-a.e.}),$$

for $g \in Y$, $\mathbb{P}(G \in \cdot)$-a.s..

Bayes' Theorem.

Sketch of a proof.

1. $Z(g)$ is an integral of a function that is positive everywhere. Thus, $Z(g) > 0$.
2. We can write
$$\pi_{\text{post}} = \pi_{F|G=g} = \frac{\pi_{F,G}(\cdot, g)}{\pi_G(g)},$$
if $\pi_G(g) > 0$. From the definition of the converse
$$L(\cdot|f) = \pi_{G|F=f} = \frac{\pi_{F,G}(f, \cdot)}{\pi_F(f)}, \text{ where } \pi_F = \pi_{\text{prior}}$$
and some algebra we obtain
$$\pi_{F,G}(f, g) = L(g|f)\pi_{\text{prior}}(f) \text{ and, by the first step: } \pi_G(g) = Z(g) > 0.$$
Thus, Bayes' Theorem is true. □

Bayes' Theorem.

Sketch of a proof.

1. $Z(g)$ is an integral of a function that is positive everywhere. Thus, $Z(g) > 0$.
2. We can write

$$\pi_{\text{post}} = \pi_{F|G=g} = \frac{\pi_{F,G}(\cdot, g)}{\pi_G(g)},$$

if $\pi_G(g) > 0$. From the definition of the converse

$$L(\cdot|f) = \pi_{G|F=f} = \frac{\pi_{F,G}(f, \cdot)}{\pi_F(f)}, \text{ where } \pi_F = \pi_{\text{prior}}$$

and some algebra we obtain

$$\pi_{F,G}(f, g) = L(g|f)\pi_{\text{prior}}(f) \text{ and, by the first step: } \pi_G(g) = Z(g) > 0.$$

Thus, Bayes' Theorem is true. □

Bayes' Theorem.

Sketch of a proof.

1. $Z(g)$ is an integral of a function that is positive everywhere. Thus, $Z(g) > 0$.
2. We can write

$$\pi_{\text{post}} = \pi_{F|G=g} = \frac{\pi_{F,G}(\cdot, g)}{\pi_G(g)},$$

if $\pi_G(g) > 0$. From the definition of the converse

$$L(\cdot|f) = \pi_{G|F=f} = \frac{\pi_{F,G}(f, \cdot)}{\pi_F(f)}, \text{ where } \pi_F = \pi_{\text{prior}}$$

and some algebra we obtain

$$\pi_{F,G}(f,g) = L(g|f)\pi_{\text{prior}}(f) \text{ and, by the first step: } \pi_G(g) = Z(g) > 0.$$

Thus, Bayes' Theorem is true. $\square$

Bayes' Theorem.

Sketch of a proof.

1. $Z(g)$ is an integral of a function that is positive everywhere. Thus, $Z(g) > 0$.
2. We can write

$$\pi_{\text{post}} = \pi_{F|G=g} = \frac{\pi_{F,G}(\cdot, g)}{\pi_G(g)},$$

if $\pi_G(g) > 0$. From the definition of the converse

$$L(\cdot|f) = \pi_{G|F=f} = \frac{\pi_{F,G}(f, \cdot)}{\pi_F(f)}, \text{ where } \pi_F = \pi_{\text{prior}}$$

and some algebra we obtain

$$\pi_{F,G}(f, g) = L(g|f)\pi_{\text{prior}}(f) \text{ and, by the first step: } \pi_G(g) = Z(g) > 0.$$

Thus, Bayes' Theorem is true. □

Bayes' Theorem.

Sketch of a proof.

1. $Z(g)$ is an integral of a function that is positive everywhere. Thus, $Z(g) > 0$.
2. We can write

$$\pi_{\mathrm{post}} = \pi_{F|G=g} = \frac{\pi_{F,G}(\cdot, g)}{\pi_G(g)},$$

if $\pi_G(g) > 0$. From the definition of the converse

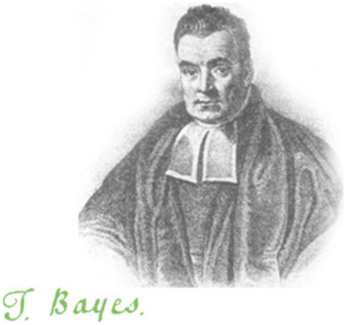
$$L(\cdot|f) = \pi_{G|F=f} = \frac{\pi_{F,G}(f, \cdot)}{\pi_F(f)}, \text{ where } \pi_F = \pi_{\mathrm{prior}}$$

and some algebra we obtain

$$\pi_{F,G}(f, g) = L(g|f)\pi_{\mathrm{prior}}(f) \text{ and, by the first step: } \pi_G(g) = Z(g) > 0.$$

Thus, Bayes' Theorem is true. □

Bayes'(?) theorem



ce qui est le principe énoncé ci-dessus, lorsque toutes les causes sont *à priori* également possibles. Si cela n'est pas, en nommant p la probabilité *à priori* de la cause que nous venons de considérer; on aura $E = Hp$; et en suivant le raisonnement précédent, on trouvera

$$P = \frac{Hp}{S.Hp};$$

ce qui donne les probabilités des diverses causes, lorsqu'elles ne sont pas toutes, également possibles *à priori*.

Pour appliquer le principe précédent à un exemple, supposons qu'une urne renferme trois boules dont chacune ne puisse être que,

Figure: Left: Maybe Bayes. Right: Certainly not Bayes [[Laplace 1812]].

Bayesian inverse problem

Bayesian inverse problem: Given likelihood $L : Y \times X \to \mathbb{R}$ measurable and μ_{prior} that satisfy Bayes' theorem.

$$\text{Find: } \mu_{\text{post}} \text{ with } \nu_X\text{-density } \pi_{\text{post}} = \frac{L(g|\cdot)\pi_{\text{prior}}}{Z(g)} \qquad (\nu_X\text{-a.e.}) \qquad \text{(BIP)}$$

for any data set $g \in Y$.

. . . .

- Is (BIP) well-posed?

Bayesian inverse problem

Bayesian inverse problem: Given likelihood $L : Y \times X \to \mathbb{R}$ measurable and μ_{prior} that satisfy Bayes' theorem.

$$\text{Find: } \mu_{\mathrm{post}} \text{ with } \nu_X\text{-density } \pi_{\mathrm{post}} = \frac{L(g|\cdot)\pi_{\mathrm{prior}}}{Z(g)} \qquad (\nu_X\text{-a.e.}) \qquad \text{(BIP)}$$

for any data set $g \in Y$.

. . . .

- Is (BIP) well-posed?

Outline

When is a (BIP) well-posed?

[Dashti, Stuart 2017: **The Bayesian Approach to Inverse Problems.** In Ghanem, Higdon, Owhadi (editors): Handbook of Uncertainty Quantification, 311–428.] [Engel, Hafemeyer, Münch, Schaden 2019: **An application of sparse measure valued Bayesian inversion to acoustic sound source identification.** Inverse Probl. 35(7)] [Ernst, Sprungk, Starkloff 2015: **Analysis of the Ensemble and Polynomial Chaos Kalman Filters in Bayesian Inverse Problems.** SIAM/ASA JUQ 3(1):823–851.] [Dolera, Mainini 2019: **On uniform continuity of posterior distributions.** Stat Probabil Lett (in Press)] [Hosseini 2017: **Well-Posed Bayesian Inverse Problems with Infinitely Divisible and Heavy-Tailed Prior Measures.** SIAM/ASA JUQ 5(1):1024–1060.] [Hosseini, Nigam 2017: **Well-Posed Bayesian Inverse Problems: Priors with Exponential Tails.** SIAM/ASA JUQ 5(1):436–465,] [Iglesias, Lin, Stuart 2014: **Well-posed Bayesian geometric inverse problems arising in subsurface flow.** Inverse Probl., 30(11):114001.] [Iglesias, Lu, Stuart 2016: **A Bayesian level set method for geometric inverse problems.** Interfaces and Free Boundaries 18:181–217.] [Kahle, Lam, JL, Ullmann 2019: **Bayesian parameter identification in Cahn–Hilliard models for biological growth**. SIAM/ASA JUQ 7(2):526-552.] [JL 2020: **On the Well-posedness of Bayesian Inverse Problems,**. SIAM/ASA JUQ 8(1):451-482.] [JL 2023: **Bayesian Inverse Problems are Usually Well-posedness**. SIAM Rev. 65(3):831-865.] [JL, Eisenberger, Ullmann 2019: **Fast Sampling of parameterised Gaussian random fields**. Comput. Methods in Appl. Mech. Engrg. 348:978-1012.] [Sprungk 2019: **On the Local Lipschitz Robustness of Bayesian Inverse Problems.** arXiv:1906.07120] [Stuart 2010: **Inverse problems: a Bayesian perspective.** Acta Numerica 19:451-559.] [Sullivan 2017: **Well-posed Bayesian inverse problems and heavy-tailed stable quasi-Banach space priors.** Inverse Probl. Imaging 11(5):857–874.],...

When is a (BIP) well-posed?

How do we define well-posedness?

Definition. [JL; 2020, 2023]

Given a metric space of probability measures (P, d). (BIP) is (P, d)-well-posed, if

(1) $\mu_{\mathrm{post}} \in \mathrm{P}$ exists

(2) μ_{post} is unique

(3) μ_{post} is stable, i.e.

$$\forall g \in Y, \varepsilon > 0 \ \exists \delta > 0 : d(\mu_{\mathrm{post}}(\cdot|g)), \mu_{\mathrm{post}}(\cdot|g')) \leq \varepsilon \quad (\|g - g'\|_Y \leq \delta)$$

(continuity in (P, d))

Thus, we need to choose an appropriate metric space (P, d) on which we measure continuity.

(P, d)?

[JL; 2020, 2023]

$(\mathrm{Prob}(X), d_{\mathrm{TV}})$-well-posed $\quad :\Leftrightarrow$: total variation well-posed

$$d_{\mathrm{TV}}(\mu, \mu') = \sup_{A \in \mathcal{B}X} |\mu(A) - \mu'(A)|, \quad \mu, \mu' \in \mathrm{Prob}(X) := \{\mu : \mu \text{ is probability distribution on } (X, \mathcal{B}X)\}$$

$(\mathrm{Prob}(X, \mu_{\mathrm{prior}}), \mathrm{d}_{\mathrm{Hel}})$-well-posed $\quad :\Leftrightarrow$: Hellinger well-posed

$(\mathrm{Prob}(X), d_{\mathrm{Prok}})$-well-posed $\quad :\Leftrightarrow$: weakly well-posed

(P, d)?

[JL; 2020, 2023]

$(\mathrm{Prob}(X), d_{\mathrm{TV}})$-well-posed $\quad :\Leftrightarrow$: total variation well-posed

$(\mathrm{Prob}(X, \mu_{\mathrm{prior}}), \mathrm{d}_{\mathrm{Hel}})$-well-posed $\quad :\Leftrightarrow$: Hellinger well-posed

$$\mathrm{d}_{\mathrm{Hel}}(\mu, \mu') = \sqrt{\frac{1}{2} \int \left(\sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\mu_{\mathrm{prior}}}} - \sqrt{\frac{\mathrm{d}\mu'}{\mathrm{d}\mu_{\mathrm{prior}}}} \right)^2 \mathrm{d}\mu_{\mathrm{prior}}},$$

$$\mu, \mu' \in \mathrm{Prob}(X, \mu_{\mathrm{prior}}) := \{\mu \in \mathrm{Prob}(X) : \mu \ll \mu_{\mathrm{prior}}\}$$

$(\mathrm{Prob}(X), d_{\mathrm{Prok}})$-well-posed $\quad :\Leftrightarrow$: weakly well-posed

(P, d)?

[JL; 2020, 2023]

$(\mathrm{Prob}(X), d_{\mathrm{TV}})$-well-posed $\quad :\Leftrightarrow$ total variation well-posed

$(\mathrm{Prob}(X, \mu_{\mathrm{prior}}), \mathrm{d}_{\mathrm{Hel}})$-well-posed $\quad :\Leftrightarrow$ Hellinger well-posed

$(\mathrm{Prob}(X), d_{\mathrm{Prok}})$-well-posed $\quad :\Leftrightarrow$ weakly well-posed

d_{Prok} metrises weak convergence, i.e. $d_{\mathrm{Prok}}(\mu, \mu') \to 0$

$$\Leftrightarrow \forall Q \text{ continuous, bounded: } \int_X Q(f)\mu(\mathrm{d}f) \to \int_X Q(f)\mu'(\mathrm{d}f) \qquad \mu, \mu' \in \mathrm{Prob}(X)$$

Main result

Theorem. [JL; 2020, 2023]

Let $\mu_{\text{prior}} \in \mathrm{Prob}(X)$. BIP is weakly, Hellinger, and total variation well-posed, if (W1)-(W4) are satisfied for $f \in X$ (μ_{prior}-a.s.) and $g \in Y$.

(W1) $L(\cdot|f)$ is a strictly positive probability density function,

(W2) $L(g|\cdot) \in \mathbf{L}^1(X, \mu_{\text{prior}})$,

(W3) $k \in \mathbf{L}^1(X, \mu_{\text{prior}})$ exists such that $L(g'|\cdot) \leq k$ for all $g' \in Y$.

(W4) $L(\cdot|f)$ is continuous.

- we can also show continuity in Wasserstein(p) distance and the Kullback–Leibler divergence, but need additional assumptions

Discussion

[JL; 2020, 2023]

(W1) $L(\cdot|f)$ is a strictly positive probability density function,

- $L(\cdot|f)$ is typically by definition of the inverse problem a probability density function
- $L > 0$: no data-parameter combinations are impossible

(W2) $L(g|\cdot) \in \mathbf{L}^1(X, \mu_{\text{prior}})$,

- Implied by (W3)

(W3) $g \in \mathbf{L}^1(X, \mu_{\text{prior}})$ exists such that $L(g'|\cdot) \leq k$ for all $g' \in Y$.

- Often satisfied by choosing g constant

(W4) $L(\cdot|f)$ is continuous.

- Simple continuity statement, in g, not f

Discussion

Assume, we have a finite number of observations, i.e. $g \in Y := \mathbb{R}^{N_{\mathrm{obs}}}$, to identify the parameter $f^* \in X$, with

$$g := \mathcal{A}(f^*) + e,$$

where the noise $E \sim \mathrm{N}(0, \Gamma)$, with $\Gamma > 0$, and $\mathcal{A} : X \to Y$ is some measurable function

. . . .

Corollary. [JL; 2020, 2023]

The Bayesian approach to the inverse problem above with any prior $\mu_{\mathrm{prior}} \in \mathrm{Prob}(X)$ is weakly, Hellinger, and total variation well-posed.

Discussion

Assume, we have a finite number of observations, i.e. $g \in Y := \mathbb{R}^{N_{\text{obs}}}$, to identify the parameter $f^* \in X$, with

$$g := \mathcal{A}(f^*) + e,$$

where the noise $E \sim \mathrm{N}(0, \Gamma)$, with $\Gamma > 0$, and $\mathcal{A} : X \to Y$ is some measurable function

. . . .

Corollary. [JL; 2020, 2023]

The Bayesian approach to the inverse problem above with any prior $\mu_{\text{prior}} \in \mathrm{Prob}(X)$ is weakly, Hellinger, and total variation well-posed.

Discussion

- Corollary can be generalised to infinite-dimensional data spaces and other noise models
  - infinite-dimensional case requires boundedness assumptions on $\mathcal{A}$
  - continuity and boundedness of the noise density function
- Assumptions (W1)-(W4) only imply continuity of the posterior, not more than that. Other works, [Stuart 2010], [Dashti+Stuart 2017], [Sprungk 2020],... show Lipschitz continuity and also other stability criteria under considerably stronger assumptions
- Well-posedness does not imply that the chosen model is sensible
  - additionally, we should look at the small noise/large data limit [Nickl 2023]
  - those are similar to the $\delta \downarrow 0$ in the variational approach to inverse problems and give results like

$$\mu_{\text{post}} \to \delta(\cdot - f^*) \qquad \text{('posterior consistency')}$$

or, even better,

$$\mu_{\text{post}} \approx \mathrm{N}(f^*, N_{\text{obs}}^{-1}\mathcal{I}^{-1}) \qquad \text{('Bernstein – von Mises Theorem').}$$

Discussion

- Corollary can be generalised to infinite-dimensional data spaces and other noise models
  - infinite-dimensional case requires boundedness assumptions on $\mathcal{A}$
  - continuity and boundedness of the noise density function
- Assumptions (W1)-(W4) only imply continuity of the posterior, not more than that. Other works, [Stuart 2010], [Dashti+Stuart 2017], [Sprungk 2020],... show Lipschitz continuity and also other stability criteria under considerably stronger assumptions
- Well-posedness does not imply that the chosen model is sensible
  - additionally, we should look at the small noise/large data limit [Nickl 2023]
  - those are similar to the $\delta \downarrow 0$ in the variational approach to inverse problems and give results like

$$\mu_{\text{post}} \to \delta(\cdot - f^*) \qquad \text{('posterior consistency')}$$

or, even better,

$$\mu_{\text{post}} \approx \mathrm{N}(f^*, N_{\text{obs}}^{-1}\mathcal{I}^{-1}) \qquad \text{('Bernstein -- von Mises Theorem').}$$

Discussion

- Corollary can be generalised to infinite-dimensional data spaces and other noise models
  - infinite-dimensional case requires boundedness assumptions on $\mathcal{A}$
  - continuity and boundedness of the noise density function
- Assumptions (W1)-(W4) only imply continuity of the posterior, not more than that. Other works, [Stuart 2010], [Dashti+Stuart 2017], [Sprungk 2020],... show Lipschitz continuity and also other stability criteria under considerably stronger assumptions
- Well-posedness does not imply that the chosen model is sensible
  - additionally, we should look at the small noise/large data limit [Nickl 2023]
  - those are similar to the $\delta \downarrow 0$ in the variational approach to inverse problems and give results like

$$\mu_{\text{post}} \to \delta(\cdot - f^*) \qquad \text{('posterior consistency')}$$

or, even better,

$$\mu_{\text{post}} \approx \mathrm{N}(f^*, N_{\text{obs}}^{-1}\mathcal{I}^{-1}) \qquad \text{('Bernstein – von Mises Theorem').}$$

Outline

Point estimators

It is nice to have a posterior measure, in practice, however

- that is usually computationally complicated
- not always practical

Obvious question: Can we find some point estimator (e.g. a value in X) that uses the Bayesian framework?

- posterior mean
- maximum-a-posteriori (MAP) estimator

Point estimators

It is nice to have a posterior measure, in practice, however

- that is usually computationally complicated
- not always practical

Obvious question: Can we find some point estimator (e.g. a value in X) that uses the Bayesian framework?

- posterior mean
- maximum-a-posteriori (MAP) estimator

Posterior mean

Given that it exists, we can use the mean of the posterior as a point estimator

$$f_{\mathrm{Bayes}}(g) = \int_X f \, \mu_{\mathrm{post}}(\mathrm{d}f) = \mathbb{E}_{\mathrm{post}}[F],$$

which is also called 'Bayes estimator'.

Posterior mean

Why is that a good idea? (Assume μ_{post} has finite second moment)

- The Bayes estimator is the best function in $\mathcal{L}^2(Y;X)$ to describe the mapping from G to F, i.e. it solves

$$f_{\text{Bayes}} = \operatorname{argmin}_{k \in \mathcal{L}^2(Y;X)} \int \|f - k(g)\|_X^2 \mathbb{P}((F,G) \in (\mathrm{d}f, \mathrm{d}g))$$

- Hellinger well-posedness will imply well-posedness of the posterior mean with the same modulus of continuity!

Why is using bad idea?

- You usually need to approximate μ_{post} to obtain f_{Bayes}...
- It often has a smoothing effect that is

Posterior mean

Why is that a good idea? (Assume μ_{post} has finite second moment)

- The Bayes estimator is the best function in $\mathcal{L}^2(Y;X)$ to describe the mapping from G to F, i.e. it solves

$$f_{\mathrm{Bayes}} = \mathrm{argmin}_{k \in \mathcal{L}^2(Y;X)} \int \|f - k(g)\|_X^2 \mathbb{P}((F,G) \in (\mathrm{d}f, \mathrm{d}g))$$

- Hellinger well-posedness will imply well-posedness of the posterior mean with the same modulus of continuity!

Why is using bad idea?

- You usually need to approximate μ_{post} to obtain f_{Bayes}...
- It often has a smoothing effect that is

Posterior mean

Why is that a good idea? (Assume μ_{post} has finite second moment)

- The Bayes estimator is the best function in $\mathcal{L}^2(Y;X)$ to describe the mapping from G to F, i.e. it solves

$$f_{\mathrm{Bayes}} = \mathrm{argmin}_{k \in \mathcal{L}^2(Y;X)} \int \|f - k(g)\|_X^2 \mathbb{P}((F,G) \in (\mathrm{d}f, \mathrm{d}g))$$

- Hellinger well-posedness will imply well-posedness of the posterior mean with the same modulus of continuity!

Why is using bad idea?

- You usually need to approximate μ_{post} to obtain f_{Bayes}...
- It often has a smoothing effect that is

Maximum-a-posteriori (MAP)

Idea: Maximum likelihoods (=the maximum of the likelihood) are good estimators, how about the maximum of the posterior density?

Let's assume for a moment that $X := \mathbb{R}^n$ and our prior/posterior has a Lebesgue density. Then, finding the maximum of π_{post} is identical to finding:

$$f_{\text{MAP}} \in \operatorname{argmin}_{f \in X} - \log \pi_{\text{post}}(f|g) = \operatorname{argmin}_{f \in X} - \log L(g|f) - \log \pi_{\text{prior}}(f) + \log Z(g)$$
$$= \operatorname{argmin}_{f \in X} \underbrace{- \log L(g|f)}_{\text{data misfit}} \underbrace{- \log \pi_{\text{prior}}(f)}_{\text{regulariser}},$$

which is exactly equivalent to a regularised variational problem.

Maximum-a-posteriori (MAP)

Idea: Maximum likelihoods (=the maximum of the likelihood) are good estimators, how about the maximum of the posterior density?

Let's assume for a moment that $X := \mathbb{R}^n$ and our prior/posterior has a Lebesgue density. Then, finding the maximum of π_{post} is identical to finding:

$$\begin{aligned} f_{\text{MAP}} \in \operatorname{argmin}_{f \in X} - \log \pi_{\text{post}}(f|g) &= \operatorname{argmin}_{f \in X} - \log L(g|f) - \log \pi_{\text{prior}}(f) + \log Z(g) \\ &= \operatorname{argmin}_{f \in X} \underbrace{- \log L(g|f)}_{\text{data misfit}} - \underbrace{\log \pi_{\text{prior}}(f)}_{\text{regulariser}}, \end{aligned}$$

which is exactly equivalent to a regularised variational problem.

Maximum-a-posteriori (MAP)

This equivalence of MAP and regularised variational approach is considerably more complicated if X is infinite-dimensional

- there is not really a natural reference density, so that way of defining it doesn't work.
- people define the MAP via the Onsager-Machlup functional
  - Essentially, finding the point around which the posterior assigns maximum probability
- the equivalence to variational problems is still mostly correct [Helin+Burger 2015]

Maximum-a-posteriori (MAP)

Why is using f_{MAP} a good idea?

- Getting the MAP is only an optimisation problem! Very cheap!
- Properties tend to be better for pure reconstruction

Why is using f_{MAP} a bad idea?

- Well-posedness not automatic!

Recap and outlook

Recap

Bayes' Theorem

Well-posedness of Bayesian inverse problems

Bayesian point estimators

Outlook

Recap and outlook

Recap.

- Bayes' formula allows us to obtain the posterior from prior and likelihood
- Bayesian inverse problems are usually well-posed

Outlook.

- The posterior is a conditional distribution. How can we get the its density?
- What is a good prior?
- Which properties does the posterior have?