Snapshots of modern mathematics
from Oberwolfach

№ 6/2022

Solving inverse problems with Bayes' theorem

Jonas Latz[1] • Björn Sprungk[2]

The goal of inverse problems is to find an unknown parameter based on noisy data. Such problems appear in a wide range of applications including geophysics, medicine, and chemistry. One method of solving them is known as the Bayesian approach. In this approach, the unknown parameter is modelled as a random variable to reflect its uncertain value. Bayes' theorem is applied to update our knowledge given new information from noisy data.

1 Inverse problems

We begin with a general problem description. Let $\mathbb{Z}$ be the set of all integers, that is, $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$. Let $f : \mathbb{Z} \to \mathbb{Z}$ be a function, and let $d \in \mathbb{Z}$ be some integer. Moreover, we assume that there is a value $t \in \mathbb{Z}$ such that

$$f(t) = d. \tag{1}$$

The task in *inverse problems* is to find $t \in \mathbb{Z}$ using the function f and the value d.

[1] JL is supported by Technical University of Munich (TUM) and DFG through the International Graduate School of Science and Engineering at TUM.

[2] BS is grateful for the support by the DFG project 389483880 and by the DFG-funded RTG 2088 "Discovering structure in complex data: statistics meets optimization and inverse problems".

1.1 Interpretation

Finding t is similar to root-finding problems: given some equation, such as $x^2 - 2 = 0$, find the unknown value that satisfies the equation. Here, we will try to understand the problem a bit differently.

Imagine a mathematical model that can be used to represent some general system of interest, such as the motion of planets. We would like to tune the parameters of the general model to study a specific instance, such as the motion of the dwarf planet Pluto. To this end, we need to find the correct parameter t of the general model that represents the specific behaviour of Pluto.

We assume that t satisfies Equation (1), where d and f are given. Here d is understood as a dataset. For example, d may be records of positions of Pluto as seen from Earth. The function f is an "observation operator" and represents the observations within the model. Hence, we have observed a dataset d and we aim to use this to find the correct parameter t for our model. For some u, $f(u)$ predicts which data we would observe, if u were the true parameter. We aim to find a parameter t which approximates u. Of course, the movement of Pluto does not really fit in the setting where f, d, and t are discrete. We will, however, stick to this setting, as it simplifies the following discussion considerably.

Next we consider an example with three different cases.

Example 1. *Let $f(u) = u^2$, for $u \in \mathbb{Z}$. We consider the following cases:*

1. *If $d = 0$, we know that $t = 0$.*
2. *If $d = -1$, we cannot find any $t \in \mathbb{Z}$ satisfying $f(t) = d$.*
3. *If $d = 1$, the true value t can be either 1 or -1.*

The inverse problems and their solutions are illustrated in Figure 1. In the second and third example we cannot identify the true parameter t. The issue of having no or many true parameters is very typical of inverse problems. We discuss this issue in the next section.

1.2 Issues in inverse problems

In inverse problems, we often face the problem that there is not a unique parameter satisfying $f(t) = d$. As we saw in the previous example, there may be zero solutions or more than one solution.

When the data d is subject to observational noise, there may be no parameters which satisfy the equation. In this case, the data is actually given by

$$f(t) + n = d, \tag{2}$$

where n is noise. Noise can arise, for example, from the inaccuracy of the instruments used for measurements. Since we do not know the value of the

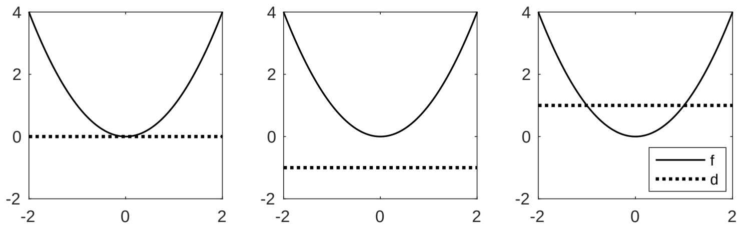

Figure 1: Plots of the function f and the dataset d in the first case (left figure), the second case (central figure), and the third case (right figure) in Example 1. The intersections of the solid line (f) and the dotted line (d) show the solutions of the inverse problem. We see that we can have one solution (first case), no solution (second case), or two solutions (third case).

noise, we now actually need to identify both t and n. In this case, however, we end up with an infinite number of potential solutions.

Proposition 1. *There are infinitely many combinations of $t, n \in \mathbb{Z}$ satisfying Equation* (2)*.*

Proof. Let $t \in \mathbb{Z}$ be chosen arbitrarily, and let $n = d - f(t)$. Then, the tuple (t, n) satisfies Equation (2). Since there are infinitely many ways to choose t, Equation (2) has infinitely many solutions. □

When ignoring noise by considering Equation (1) instead of Equation (2), we may run into an inverse problem with no solution. When considering noise, as in Equation (2), we obtain an inverse problem with an infinite number of solutions.

Such situations are typical for inverse problems. In fact, inverse problems are usually "ill-posed problems". This simply means that they are not "well-posed" in the sense described by Jacques Salomon Hadamard (1865–1963): a *well-posed problem* admits a unique solution which depends continuously on the data [6].

In order to solve (ill-posed) inverse problems in practice, various approaches have been developed and discussed in the literature. The most popular of these is probably the regularised least squares approach. To find the parameter t satisfying Equation (1) using the least squares approach, one computes

$$\min_{t \in \mathbb{Z}} \frac{1}{2}(f(t) - d)^2. \tag{3}$$

This method was developed in the early 19th century by Carl Friedrich Gauß (1777–1855) and Adrien-Marie Legendre (1752–1833) independently. It was successfully applied by Gauß to determine the motion of the dwarf planet Ceres [12].

The least squares approach can overcome the problem that Equation (1) has no solutions due to noise in the data. However, there might still be multiple solutions to Equation (3). To this end, we can add another term in Equation (3) which encodes additional information or preferences about the sought-after t. For instance, if multiple minimizers in Equation (3) exist, we may prefer the one closest to zero. In that case, we seek

$$\min_{t \in \mathbb{Z}} \frac{1}{2}(f(t) - d)^2 + \frac{\alpha}{2} t^2, \tag{4}$$

where $t \mapsto \frac{1}{2}t^2$ denotes the regularising (or penalty) functional, and $\alpha > 0$ is a parameter steering the influence of the regularisation.

In recent years, the Bayesian approach to inverse problems has gained popularity. In this approach, regularisation and additional information to inverse problems is introduced in a probabilistic way. We will explain and discuss the Bayesian approach in this snapshot.

2 Conditional probability and Bayesian statistics

2.1 Probability and knowledge

We say that U is a *random variable* taking values in $\mathbb{Z}$ if the value of $U \in \mathbb{Z}$ is unknown and determined by the outcome of a random experiment, such as tossing a coin. We describe the value of U by its probability distribution. The probability that the value of U is in a set A, where $A \subseteq \mathbb{Z}$, is denoted $\mathrm{P}(U \in A)$.

What does this mean? We follow the "Bayesian" or "subjective interpretation" of probabilities. In his original essay, Thomas Bayes (1702–1761) [2] gave the following definition: "The *probability of any event* is the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon it's happening." Let us translate this into modern terms. Let $A \subseteq \mathbb{Z}$ be some event. (In probability theory, an *event* is some subset of the possible outcomes.) We anticipate that we obtain €1 if the event A occurs. Hence, the "value of the thing expected upon it's happening" is €1. The probability is "the value at which an expectation depending on the happening of the event ought to be computed". We interpret this now as the price that we are paying in a fair game to win €1 if A occurs.

Example 2. *Somebody tosses a coin. We win €1 if the coin shows head. Since we expect the coin to be fair, we bet €0.50. After we hand over the money, the coin is tossed, and we see whether we get €1 or not. Now, assume somebody places a coin in a box and asks us whether it shows head or not. If we guess correctly, we win €1. Again, we would bet €0.50, but this time the game is* *entirely based on our limited knowledge or uncertainty about the coin rather than actual randomness.*

How do we know that €0.50 is a fair price to win €1 in the first game discussed in Example 2? Probabilities of random experiments are typically defined using their *frequentist interpretation*: the probability $P(A)$ of an event A is the value obtained in the ratio

$$\frac{\text{number of occurences of event } A \text{ within } n \text{ independent experiments}}{n}$$

when n goes to infinity. If we flip a coin infinitely often, we will see head half of the time. We note that this interpretation is indeed useful to describe the outcome of a random and repeatable experiment.

However, this interpretation fails if we consider uncertainty about non-repeatable events, say, the outcome of a particular football match. The Bayesian interpretation, on the other hand, is capable of interpreting probabilities for any kind of event, including repeatable random experiments where it coincides with the frequentist interpretation. Thus, we stick to the Bayesian approach in the following.

We consider the value of the variable U to be uncertain and use $\mathrm{P}(U \in \cdot)$ to describe our knowledge concerning U. The function $\mathrm{P}(U \in \cdot)$ is called the *probability distribution* of U.

2.2 Learning by conditioning

If we model knowledge concerning a variable by a probability measure, we also need a way to model the process of learning. We do this using conditional probability[3]. Let $B \subseteq \mathbb{Z}$, with $\mathrm{P}(B) > 0$. Then, the probability that $U \in A$ occurs given that we know already that B occurs, is given by

$$\mathrm{P}(U \in A \mid B) = \frac{\mathrm{P}(U \in A \cap B)}{\mathrm{P}(B)}. \tag{5}$$

The function $\mathrm{P}(U \in A \mid B)$ is called the *conditional probability distribution of A given B.*[4] Learning that the event B occurs is represented by the map

$$\mathrm{P}(U \in \cdot) \longmapsto \mathrm{P}(U \in \cdot \mid B).$$

Example 3. *We shuffle a complete poker deck. Then, we draw a card and check its suit, that is, whether it is ♢,♡, ♠, or ♣. If it is ♣, we obtain €1.*

[3] You can refer to [10] for an interactive and visual introduction to conditional proabability.

[4] Intuitively, $P(U \in A \mid B)$ indicates how likely it is that U is in A, assuming that the event B holds.

Otherwise, we get no money. Since the suits are uniformly distributed among the deck, we know that

$$\mathrm{P}(U = \diamondsuit) = \mathrm{P}(U = \heartsuit) = \mathrm{P}(U = \spadesuit) = \mathrm{P}(U = \clubsuit) = 0.25, \tag{6}$$

which we have again computed with the frequentist interpretation of probabilities. Hence, we would pay €0.25 to participate in the game above. However, if somebody checks the card and tells us whether it is red or black, we would decide differently. If the person tells us that the suit is red, we obtain the following conditional probability:

$$\begin{aligned}\mathrm{P}(U = \clubsuit \mid \textit{suit is red}) &= \mathrm{P}(U = \clubsuit \mid U \in \{\diamondsuit, \heartsuit\}) \\ &= \frac{\mathrm{P}(U \in \{\diamondsuit, \heartsuit\} \cap \{\clubsuit\})}{\mathrm{P}(U \in \{\diamondsuit, \heartsuit\})} \\ &= \frac{\mathrm{P}(U \in \emptyset)}{\mathrm{P}(U \in \{\diamondsuit, \heartsuit\})} = \frac{0}{0.5} = 0.\end{aligned} \tag{7}$$

Hence, we will not spend any money on a game in which we cannot win anything. Moving on to the case where the suit is black.

$$\begin{aligned}\mathrm{P}(U = \clubsuit \mid \textit{suit is black}) &= \mathrm{P}(U = \clubsuit \mid U \in \{\spadesuit, \clubsuit\}) \\ &= \frac{\mathrm{P}(U \in \{\spadesuit, \clubsuit\} \cap \{\clubsuit\})}{\mathrm{P}(U \in \{\spadesuit, \clubsuit\})} \\ &= \frac{\mathrm{P}(U = \clubsuit)}{\mathrm{P}(U \in \{\spadesuit, \clubsuit\})} = \frac{0.25}{0.5} = 0.5.\end{aligned} \tag{8}$$

Hence, we know that there is a 50% chance for the card to show $\clubsuit$*. Now, we would spend €0.50 to play the game mentioned above.*

2.3 Bayes' theorem

In practice, we typically cannot compute the conditional probability using the formula in Equation (5), since we cannot access the joint probability in the numerator of the term on the right-hand side. However, we sometimes can access the *inverted conditional probability* $\mathrm{P}(B \mid U \in A)$, which is the probability of B given that the value of U is in A. In this case, we can use *Bayes' theorem* to go from $\mathrm{P}(B \mid U \in A)$ to $\mathrm{P}(U \in A \mid B)$.

Theorem 1 (Bayes)**.** *Let* $A, B \subseteq \mathbb{Z}$ *and* $\mathrm{P}(B) > 0$*. Then,*

$$\mathrm{P}(U \in A \mid B) = \frac{\mathrm{P}(B \mid U \in A) \cdot \mathrm{P}(U \in A)}{\mathrm{P}(B)}.$$

Proof. Let us first assume that $\mathrm{P}(U \in A) > 0$. Then, according to Equation (5), we have

$$\mathrm{P}(B \mid U \in A) \cdot \mathrm{P}(U \in A) = \mathrm{P}(\{U \in A\} \cap B),$$

where we exchange $\{U \in A\}$ and B. Hence, we have

$$\frac{\mathrm{P}(B \mid U \in A) \cdot \mathrm{P}(U \in A)}{\mathrm{P}(B)} = \frac{\mathrm{P}(\{U \in A\} \cap B)}{\mathrm{P}(B)} = \mathrm{P}(U \in A \mid B).$$

Let us now assume that $\mathrm{P}(U \in A) = 0$. Then, we have

$$\mathrm{P}(\{U \in A\} \cap B) \leq \mathrm{P}(\{U \in A\}) = 0,$$

by the monotonicity of probabilities. According to Equation (5), we thus have $\mathrm{P}(U \in A \mid B) = 0$. Hence,

$$\mathrm{P}(U \in A \mid B) = 0 = \frac{\mathrm{P}(B \mid U \in A) \cdot 0}{\mathrm{P}(B)} = \frac{\mathrm{P}(B \mid U \in A)\mathrm{P}(U \in A)}{\mathrm{P}(B)},$$

which again proves our claim. □

We can now use the result of Theorem 1 to compute the conditional probabilities in Example 3 with the playing cards in a different way.

Example 4 (Example 3 revisited)**.** *We now aim to compute the conditional probabilities in Equation* (7) *and Equation* (8) *using the result from Theorem 1. To this end, we need the conditional probabilities* $\mathrm{P}(U \in \{\diamondsuit, \heartsuit\} \mid U = \clubsuit)$ *and* $\mathrm{P}(U \in \{\spadesuit, \clubsuit\} \mid U = \clubsuit)$*. We can easily deduce what these conditional probabilities are. If we know already that* $U = \clubsuit$*, then* $U \notin \{\diamondsuit, \heartsuit\}$ *with probability 1, and* $U \in \{\spadesuit, \clubsuit\}$ *with probability* 1*. Therefore,*

$$\mathrm{P}(U \in \{\diamondsuit, \heartsuit\} \mid U = \clubsuit) = 0 \qquad \mathrm{P}(U \in \{\spadesuit, \clubsuit\} \mid U = \clubsuit) = 1.$$

Now, we apply Theorem 1. We obtain

$$\mathrm{P}(U = \clubsuit \mid U \in \{\diamondsuit, \heartsuit\}) = \frac{\mathrm{P}(U \in \{\diamondsuit, \heartsuit\} \mid U = \clubsuit) \cdot \mathrm{P}(U = \clubsuit)}{\mathrm{P}(U \in \{\diamondsuit, \heartsuit\})} = \frac{0 \cdot 0.25}{0.5} = 0$$

and

$$\mathrm{P}(U = \clubsuit \mid U \in \{\spadesuit, \clubsuit\}) = \frac{\mathrm{P}(U \in \{\spadesuit, \clubsuit\} \mid U = \clubsuit) \cdot \mathrm{P}(U = \clubsuit)}{\mathrm{P}(U \in \{\spadesuit, \clubsuit\})} = \frac{1 \cdot 0.25}{0.5} = 0.5,$$

as in Example 3.

Let $A = \{u\} \subseteq \mathbb{Z}$. Now, conditional probabilities of the type

$$\mathrm{P}(B \mid U \in A) = \mathrm{P}(B \mid U = u)$$

frequently appear in statistics. In these expressions, an event B is observed and based on this observation some true parameter t shall be identified. A function like $u \mapsto \mathrm{P}(B \mid U = u)$ is called the *(data) likelihood*. It shows the likelihood of observing the data (that is B), given that the unknown true parameter is equal to u.

Given the likelihood, there are multiple ways to obtain a parameter estimate. A popular method is the "maximum likelihood" method. Here, one just maximises $u \mapsto \mathrm{P}(B \mid U = u)$. Hence, the parameter is estimated by the value for which the event B is particularly likely.

Example 5. *We are again given a deck of a total of 52 playing cards. We aim to find out how many of these playing cards are* ♣*. To this end, we draw 10 cards in the following fashion: we draw a card, note down whether it shows* ♣ *or not, put it back, shuffle the deck and then continue. Four of the ten cards show* ♣*. Hence, we observe the event*

$$B = \{4\clubsuit\}.$$

Let U be the number of ♣ *in the deck. Then, for each $u = 0, \ldots, 52$,*

$$\mathrm{P}(B \mid U = u) = \left(\frac{u}{52}\right)^4 \left(1 - \frac{u}{52}\right)^6,$$

and $\mathrm{P}(B \mid U = u) = 0$, if $u < 0$ or $u > 52$. The function $\mathrm{P}(B \mid U = u)$ attains its maximum at $u = 21$. Hence, 21 is the maximum likelihood estimate for the number of ♣ *in the deck. We plot the likelihood and the maximum likelihood in Figure 2.*

The Bayesian paradigm is an alternative to the maximum likelihood framework. We consider u to be uncertain and model it by the random variable U. The probability distribution of U models our knowledge concerning u prior to observing B. Therefore, $\mathrm{P}(U \in \cdot)$ is called the *prior probability distribution*. Now we observe the data and learn by conditioning. Having observed the data, our state of knowledge is given by the *posterior probability distribution* $\mathrm{P}(U \in \cdot \mid B)$.

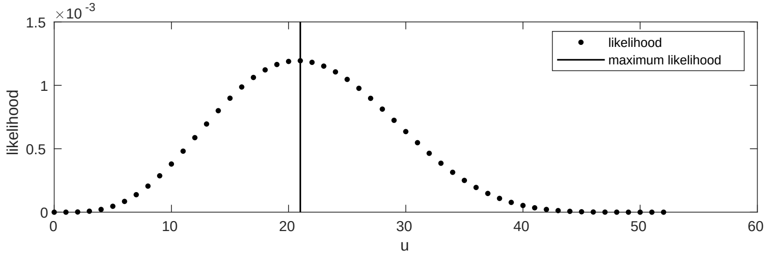

Figure 2: Plot of the likelihood and maximum likelihood in Example 5. The maximum likelihood marks the point at which the likelihood function is maximised.

Given the prior and the likelihood, we can obtain the posterior using Theorem 1. Indeed, the posterior probability of $U = u$ for any $u \in \mathbb{Z}$ is given by

$$\underbrace{\mathrm{P}(U = u \mid B)}_{\text{posterior}} = \frac{\overbrace{\mathrm{P}(B \mid U = u)}^{\text{likelihood}}\overbrace{\mathrm{P}(U = u)}^{\text{prior}}}{\underbrace{\mathrm{P}(B)}_{\text{evidence}}}.$$

The *evidence* $\mathrm{P}(B)$ can, for instance, be computed by summing over the prior and the likelihood:

$$\mathrm{P}(B) = \sum_{u' \in \mathbb{Z}} \mathrm{P}(B \mid U = u')\mathrm{P}(U = u').$$

We now go back to our playing card example, this time with a Bayesian approach.

Example 6 (Example 5 revisited). *We want to apply a Bayesian approach to identify the number of ♣ cards in our deck. To this end, we need to model our prior assumptions. We consider the following three examples:*

(i) We do not have any information about the number of ♣ cards in the deck, other than that this number lies between 0 *and* 52*. Hence, according to the prior measure, all values within these bounds are equally likely:*

$$\mathrm{P}(U = u) = \mathrm{P}(U = u') \qquad u, u' = 0, 1, \ldots, 52.$$

Therefore, we have $\mathrm{P}(U = u) = 1/53$*, if* $u = 0, 1, ..., 52$*, and* $\mathrm{P}(U = u) = 0$*, otherwise. This prior is called* uniform *or* uninformative.

(ii) We have a certain tendency to believe there are 26 ♣, but we are still pretty unsure. Therefore, we use a symmetric "binomial"[5] *prior centred around* 26*:*

$$\mathrm{P}(U = u) = \binom{52}{u} 0.5^{52}.$$

Here, $\mathrm{P}(U = u)$ *is the probability of obtaining head* u *times when tossing a fair coin* 52 *times.*

(iii) We are very certain that the number of ♣ cards is close to 13*. Indeed, we assume that*

$$\begin{aligned} &\mathrm{P}(U = 13) = 0.3, \\ &\mathrm{P}(U = 12) = \mathrm{P}(U = 14) = 0.2, \\ &\mathrm{P}(U = 11) = \mathrm{P}(U = 15) = 0.15, \\ \textit{and} \quad &\mathrm{P}(U = u) = 0 \quad \textit{in any other case.} \end{aligned} \tag{9}$$

Hence, we exclude the possibility that $u > 15$ *or* $u < 11$*.*

We plot the priors $\mathrm{P}(U = u)$ and the posteriors $\mathrm{P}(U = u \mid B)$ in the three examples in Figure 3. Like in Example 5, the event B is that four out of the ten cards that were drawn show ♣. Given the uninformative prior in (i), the posterior measure is a scaled version of the likelihood in Example 5 and Figure 2. In (ii), with a distribution centered on 26, the prior and posterior are very similar. The prior is already very informative, focusing around a value of 26. The data corrects the prior a bit by moving it a bit to the left. In (iii), we first observe that the posterior probabilities are 0, whenever the prior probabilities are 0. Hence, we cannot overturn the prior if it is 0. Other than that, the posterior does not change the prior probability at $u = 13$. However, it increases the probability for $u > 13$ and lowers it for $u < 13$. Hence, as in (ii), we see a drag towards the maximum likelihood.

3 Bayesian inverse problems

Having introduced the paradigm of Bayesian statistics, we now aim to use it to solve inverse problems. We now model the unknown parameter u in the inverse problem setting as a random variable U. This random variable follows a prior distribution $\mathrm{P}(U \in \cdot)$. Using what we have learnt about inverse problems, we can define an observation B from a dataset d:

$$B = \{f(U) = d\},$$

[5] A binomial distribution is a probability distribution corresponding to an experiment with exactly two possible outcomes, for example, a coin toss. A good place to develop an intuition for some commonly occurring probability distributions, including binomial distribution, is [9].

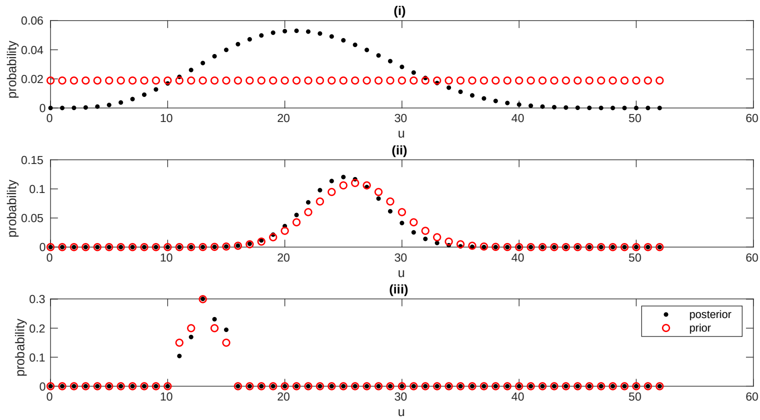

Figure 3: Plots of the priors and posteriors in Example 6. In (i), the prior is uniform (uninformative), and the posterior is identical to the likelihood. In (ii), the prior favours a value close to 26, while the posterior only barely corrects this prior knowledge. In (iii), the prior only allows for a very small number of values, and the posterior is strongly influenced by this prior choice and only suggests higher probabilities for the values $\{14, 15\}$.

if the observations are noise-free, and

$$B = \{f(U) + N = d\},$$

if the observations are noisy. In the noisy case, the noise is now also given by a random variable N. In either of these cases, we aim to compute the posterior distribution of U given B, that is,

$$\mathrm{P}(U \in \cdot \mid f(U) = d), \quad \text{or} \quad \mathrm{P}(U \in \cdot \mid f(U) + N = d).$$

This posterior distribution is considered the solution of the Bayesian inverse problem. Hence, we have a unique solution in terms of a probability distribution. Note that it can typically also be shown that the posterior depends continuously on the data, and therefore, Bayesian inverse problems are well-posed [11]. To compute the posterior distributions, we need to know the associated likelihoods. We will derive these in the following section.

3.1 Likelihoods in inverse problems

We start with the noise-free case. We will derive the likelihood of observing the dataset d given that the true data is u, for some $u \in \mathbb{Z}$,

$$\mathrm{P}(f(U) = d \mid U = u).$$

Note that we fix the random variable $U = u$. Therefore, the event $\{f(U) = d\}$ is fully deterministic – we can say for sure whether it occurs or not. It can only occur if $u \in \mathbb{Z}$ is chosen such that $f(u) = d$. Therefore, we obtain

$$\mathrm{P}(f(U) = d \mid U = u) = \begin{cases} 1, & \text{if } f(u) = d, \\ 0, & \text{otherwise,} \end{cases}$$

as the likelihood.

In the noisy case, we first need to define the probability distribution of the noise N. We denote the probability of the event $\{N = n\}$, for $n \in \mathbb{Z}$, by

$$P_N(n) := \mathrm{P}(N = n).$$

We now again fix $U = u$, for $u \in \mathbb{Z}$, and obtain

$$\begin{aligned} \mathrm{P}(f(U) + N = d \mid U = u) &= \mathrm{P}(f(u) + N = d) \\ &= \mathrm{P}(N = d - f(u)) \\ &= P_N(d - f(u)) \end{aligned}$$

as the likelihood. Hence, if we know the noise model P_N, we can immediately deduce the likelihood for the inverse problem.

3.2 Revisiting the quadratic examples

We now revisit Example 1 and compute Bayesian solutions to these inverse problems. We start with the noise-free case.

Example 7 (Example 1 revisited)**.** *We consider the inverse problem $f(u) = u^2$, with three cases: $d = -1$, $d = 0$, and $d = 1$. As a prior for U, we choose two examples:*

(i) A uniform prior on $\{-3, -2, \ldots, 3\}$, that is,

$$\mathrm{P}(U = u) = \begin{cases} 1/7, & \textit{if } -3 \le u \le 3, \\ 0, & \textit{otherwise} \end{cases} \qquad (u \in \mathbb{Z}).$$

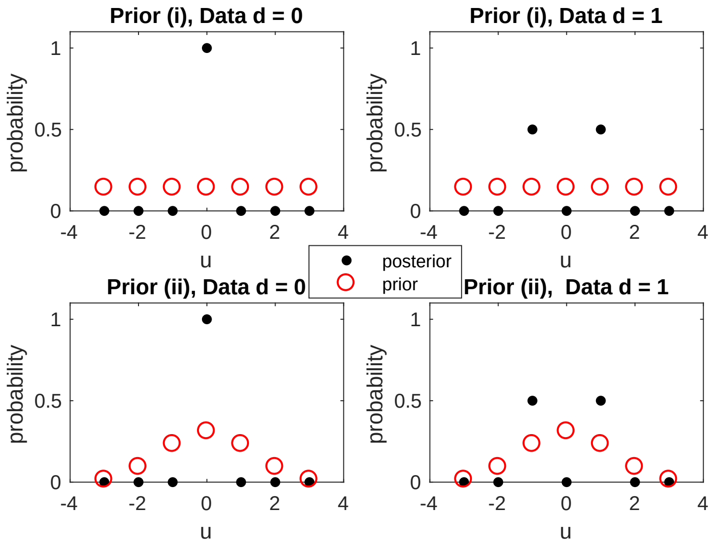

Figure 4: Plots of the priors and posteriors in Example 7. In the top row, the prior is uniform, while in the bottom row the prior is binomial. In this problem with noise-free, highly informative data, the prior has no influence.

(ii) A symmetric binomial prior on the same set, centered at 0*, that is,*

$$\mathrm{P}(U = u) = \binom{6}{u+3} \cdot 0.5^6 \qquad (u \in \mathbb{Z}).$$

To compute the posterior measure, we use Bayes' formula,

$$\underbrace{\mathrm{P}(U = u \mid f(U) = d)}_{\text{posterior}} = \frac{\overbrace{\mathrm{P}(f(U) = d \mid U = u)}^{\text{likelihood}}\,\overbrace{\mathrm{P}(U = u)}^{\text{prior}}}{\underbrace{\mathrm{P}(f(U) = d)}_{\text{evidence}}}.$$

First, consider the case where $d = -1$. The posterior measure, as computed with Bayes' formula, does not exist since the evidence is everywhere equal to 0.

We plot priors and posteriors for $d = 0$ and $d = 1$ in Figure 4. For each dataset, the posterior measure does not depend on the prior. In the case $d = 0$, we obtain

$$\begin{aligned} P(U = 0 \mid f(U) = d) &= 1, \qquad \text{and} \\ P(U = u \mid f(U) = d) &= 0 \qquad (u \neq 0). \end{aligned}$$

Hence, we know for sure that the true value of t is 1. In the case where $d = 1$, we obtain

$$\begin{aligned} P(U = 0 \mid f(U) = -1) &= 1/2, \\ P(U = 0 \mid f(U) = 1) &= 1/2, \qquad \text{and} \\ P(U = u \mid f(U) = d) &= 0 \qquad (u \neq -1, 1). \end{aligned}$$

Hence, we still do not know whether $t = -1$ or $t = 1$, but we can exclude all other possibilities. We are still uncertain concerning t. However, as opposed to multiple solutions in Example 1, the posterior distribution is a unique solution to the Bayesian inverse problem. The posterior distribution quantifies precisely our remaining uncertainty concerning t.

Next, we consider the noisy case. We use the same priors as in Example 7 and two different noise models.

Example 8. *We consider the inverse problem $f(u) = u^2 + n$ for three cases: $d = -1$, $d = 0$, and $d = 1$. As a prior for U, we consider both the uniform prior (i) and the symmetric binomial prior (ii) from Example 7. Moreover, we consider the following noise models, that is, probability distributions of N:*

(a) A symmetric binomial noise on $\{-4, -3, \ldots, 4\}$, that is,

$$P_N(n) = \mathrm{P}(N = n) = \binom{8}{n+4} \cdot 0.5^8 \qquad (n \in \mathbb{Z}).$$

(b) A symmetric binomial noise on $\{-1, 0, 1\}$, that is,

$$P_N(n) = \mathrm{P}(N = n) = \binom{2}{n+1} \cdot 0.5^2 \qquad (n \in \mathbb{Z}).$$

We plot the noise distributions in Figure 5. There we see that in (a) the noise is spread more widely than in (b). That means, we expect a larger observational noise in (a) and more accurate data in (b).

We show the posterior distributions from Example 8 in Figure 6. First note that in this case, we give a positive posterior probability to a much bigger range of potential parameter values. Since we are uncertain about the observational noise, this is what we should expect. Also, we can nicely observe the influence of the noise distribution. The larger observational noise (a) gives flatter posterior

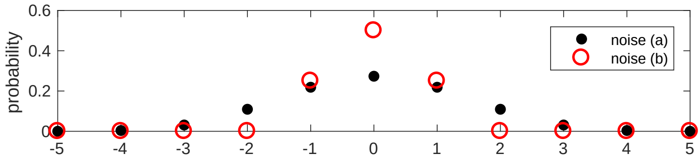

Figure 5: Plots of the noise distributions in Example 8. Noise in (a) is spread more widely than in (b), which means that we expect the data to be more inaccurate in (a) than in (b).

distributions, as opposed to the smaller observational noise (b) which gives more peaks in the posterior distributions. With a flatter posterior distribution, we anticipate a higher variation in the parameter, that is, the parameter still contains a lot of uncertainty. This is consistent with the data being very noisy. When there are more peaks in the posterior distribution, we anticipate hardly any variation in the parameter, that is, we are pretty certain about the parameter. This is consistent with the data being not very noisy. In contrast to the noise-free example, we observe an influence of the prior probability distribution.

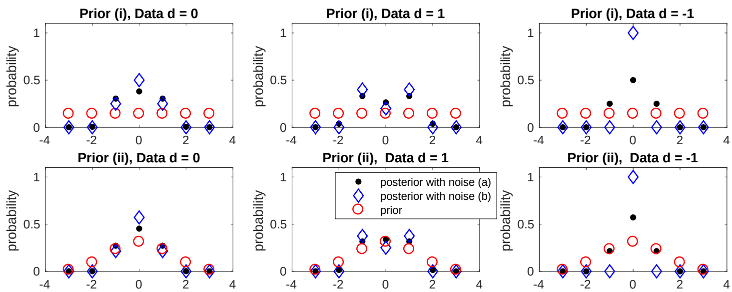

Figure 6: Plots of the priors and posteriors in Example 8. The more accurate data/smaller noise in (b) leads to a posterior that is more concentrated around the true values. In (a), the posterior is flatter. In both cases, the data is fairly accurate. Thus, the influence of the prior is small.

Consider, for instance, the case where $d = 1$, and where the noise distribution is given by (a). With the uniform prior, we give the highest posterior probability to $\{U = -1\}$ and $\{U = 1\}$, while with the centered binomial prior, we give the highest posterior probability to $\{U = 0\}$. This is due to having the prior knowledge that $\{U = 0\}$ is very likely.

4 Further reading

A comprehensive introduction to inverse problems and the regularisational approach is given by Engl, Hanke, and Neubauer [4]. A classic book on probability theory including the perspective of subjective probability is Jaynes [7]. In Whittle [14], the approach of defining probability via expectations is presented in detail. For the Bayesian approach to inverse problems, we refer to the now standard works by Kaipio and Sommersalo [8] and by Stuart [11]. Further reading on Bayesian analysis and statistics is provided by, for example, Berger [3] and Ghosh, Delampady, and Samanta [5]. Finally, some introductory tutorials on applying the Bayesian approach to inverse problems are, for instance, [1] and [13].

References

[1] M. Allmaras, W. Bangerth, J. Linhart, J. Polanco, F. Wang, K. Wang, J. Webster, and S. Zedler, *Estimating parameters in physical models through Bayesian inversion: a complete example*, SIAM Review **55** (2013), no. 1, 149–167.

[2] T. Bayes, *An essay towards solving a problem in the doctrine of chances*, Philosophical Transactions of the Royal Society **53** (1763), 370–418.

[3] J. Berger, *Statistical decision theory and Bayesian analysis*, 2nd ed., Springer, 1985.

[4] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, Springer, 2000.

[5] J. K. Ghosh, M. Delampady, and T. Samanta, *An introduction to Bayesian analysis – theory and methods*, Springer, 2006.

[6] J. Hadamard, *Sur les problèmes aux dérivés partielles et leur signification physique*, Princeton University Bulletin **13** (1902), 49–52.

[7] E. T. Jaynes, *Probability theory: the logic of science*, Cambridge University Press, 2003.

[8] J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*, Springer, 2005.

[9] Kunin, D., *Probability distributions, Seeing theory*, https://seeing-theory.brown.edu/probability-distributions/index.html, Accessed: 15-05-2022.

[10] Powell, V., *Conditional probability, Setosa.io*, https://setosa.io/conditional, Accessed: 15-05-2022.

[11] A. M. Stuart, *Inverse problems: a Bayesian perspective*, Acta Numerica **19** (2010), 451–559.

[12] D. Teets and K. Whitehead, *The discovery of Ceres: how Gauss became famous*, Mathematics Magazine **72** (1999), no. 2, 83–93.

[13] T. J. Ulrych, M. D. Sacchi, and A. Woodbury, *A Bayes tour of inversion: a tutorial*, Geophysics **66** (2001), 55–69.

[14] P. Whittle, *Probability via expectation*, 4th ed., Springer, 2000.

Jonas Latz *is an assistant professor at the School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, and the Maxwell Institute for Mathematical Sciences.*

Björn Sprungk *is a junior professor and head of the research group "Uncertainty Quantification" at Technische Universität Bergakademie Freiberg.*

Mathematical subjects
Numerics and Scientific Computing, Probability Theory and Statistics

Connections to other fields
Chemistry and Earth Science, Computer Science, Engineering and Technology, Finance, Humanities and Social Sciences, Life Science, Physics

DOI
10.14760/SNAP-2022-006-EN

Snapshots of modern mathematics from Oberwolfach provide exciting insights into current mathematical research. They are written by participants in the scientific program of the Mathematisches Forschungsinstitut Oberwolfach (MFO). The snapshot project is designed to promote the understanding and appreciation of modern mathematics and mathematical research in the interested public worldwide. All snapshots are published in cooperation with the IMAGINARY platform and can be found on www.imaginary.org/snapshots and on www.mfo.de/snapshots.

ISSN 2626-1995

Mathematisches Forschungsinstitut Oberwolfach gGmbH
Schwarzwaldstr. 9 – 11
77709 Oberwolfach
Germany

Director
Gerhard Huisken

Mathematisches Forschungsinstitut Oberwolfach

Member of Leibniz Association

IMAGINARY open mathematics