Bayesian inverse problems I: Probability and Conditioning

Jonas Latz

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

MANCHESTER 1824
The University of Manchester

Outline

Introduction: quantify uncertainties in inverse problems

Probabilities represent uncertainties

Learning through conditioning

Outlook

Uncertainties in inverse problems

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.

As we know: the problem of identifying f^* is ill-posed

- basically no chance to get the 'true' f^*
- classical approach: obtain a good estimate f^δ that converges to f^*, if the noiselevel $\delta \downarrow 0$.
- Questions: How good is the estimate f^δ? What do we actually know? Is there something we don't know? How do we quantify the remaining uncertainties?

Uncertainties in inverse problems

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.

As we know: the problem of identifying f^* is ill-posed

- basically no chance to get the 'true' f^*
- classical approach: obtain a good estimate f^δ that converges to f^*, if the noiselevel $\delta \downarrow 0$.
- Questions: How good is the estimate f^δ? What do we actually know? Is there something we don't know? How do we quantify the remaining uncertainties?

Quantify uncertainties in inverse problems

- Questions: How good is the estimate f^δ? What do we actually know? Is there something we don't know? How do we quantify the remaining uncertainties?

Use the **Bayesian approach to inverse problems**.

Idea:

1. Model the unknown parameter as a random variable F following a prior distribution $\mu_{\text{prior}} := \mathbb{P}(F \in \cdot)$ to represent its uncertainty before seeing the data.
2. Observe that $\mathcal{A}(F) + E = g$.
3. Learn what you have observed by deriving/approximating the posterior distribution

$$\mu_{\text{post}} := \mathbb{P}(F \in \cdot | \mathcal{A}(F) + E = g).$$

$\rightarrow$ ...but that was a bit quick!

Quantify uncertainties in inverse problems

- Questions: How good is the estimate f^δ? What do we actually know? Is there something we don't know? How do we quantify the remaining uncertainties?

Use the **Bayesian approach to inverse problems**.

Idea:

1. Model the unknown parameter as a random variable F following a prior distribution $\mu_{\text{prior}} := \mathbb{P}(F \in \cdot)$ to represent its uncertainty before seeing the data.
2. Observe that $\mathcal{A}(F) + E = g$.
3. Learn what you have observed by deriving/approximating the posterior distribution

$$\mu_{\text{post}} := \mathbb{P}(F \in \cdot | \mathcal{A}(F) + E = g).$$

$\rightarrow$ ...but that was a bit quick!

Quantify uncertainties in inverse problems

- Questions: How good is the estimate f^δ? What do we actually know? Is there something we don't know? How do we quantify the remaining uncertainties?

Use the **Bayesian approach to inverse problems**.

Idea:

1. Model the unknown parameter as a random variable F following a prior distribution $\mu_{\text{prior}} := \mathbb{P}(F \in \cdot)$ to represent its uncertainty before seeing the data.
2. Observe that $\mathcal{A}(F) + E = g$.
3. Learn what you have observed by deriving/approximating the posterior distribution

$$\mu_{\text{post}} := \mathbb{P}(F \in \cdot | \mathcal{A}(F) + E = g).$$

$\rightarrow$ ...but that was a bit quick!

Outline

Literature

SEMINAL WORKS.

Kaipio + Somersalo 2005: Statistical and Computational Inverse Problems, Springer.

Stuart 2010: Inverse problems: A Bayesian perspective, Acta Numer. 19:451-559.

Tarantola 2005: Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM.

INTRODUCTORY WORKS.

Allmaras + al. 2013: Estimating Parameters in Physical Models through Bayesian Inversion: A Complete Example, SIAM Rev. 55(1):149-167.

L. + Sprungk 2022: Solving inverse problems with Bayes' theorem, Snapshots Mod. Math. Oberwolfach.

PHILOSOPHICAL MATERIAL.

Bayes 1763: An Essay towards solving a Problem in the Doctrine of Chance, Philos. Trans. R. Soc. 53:370-418.

Cox 1946: Probability, Frequency and Reasonable Expectation, Am. J. Phys. 14(1):1-13.

Jaynes 2003: Probability Theory: The Logic of Science, Cambridge Univ. Press.

Kolmogorov 1933: Grundbegriffe der Wahrscheinlichkeitsrechnung, Springer.

Schwarz 2018: No interpretation of probability, Erkenntnis 83:1195-1212.

Richard von Mises 1928: Wahrscheinlichkeit Statistik und Wahrheit, Springer. (English version avaialble: Probability, Statistics, and Truth).

WELL-POSEDNESS OF BAYESIAN INVERSE PROBLEMS

Dashti + Stuart 2017: The Bayesian Approach to Inverse Problems, Handbook of Uncert. Quantif., pp. 311-428.

L. 2020: On the Well-posedness of Bayesian Inverse Problems, SIAM J. Uncert. Quantif. 8(1):451-482.

L. 2023: Bayesian Inverse Problems are Usually Well-posedness, SIAM Rev. 65(3):831-865.

Sprungk 2020: On the local Lipschitz stability of Bayesian inverse problems, Inverse Probl. 36(5):055015.

COMPUTATIONS.

Cotter + al. 2013: MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster, Stat. Sci. 28(3):242-446.

Robert + Casella 2004: Monte Carlo Statistical Methods, Springer.

BACKGROUND PROBABILITY THEORY.

Ash + Doleans-Dade 1999: Probability and Measure Theory, Acad. Press.

STATISTICAL CONSIDERATIONS.

Nickl 2023: Bayesian non-linear statistical inverse problems, EMS.

Outline:

Introduction: quantify uncertainties in inverse problems

Probabilities represent uncertainties

Learning through conditioning

Outlook

Probabilities represent uncertainties?!

What actually is a probability?

- Several different schools of thought in probability

> *'If minor differences are counted, the number of schools seems to be somewhere between two and the number of authors, and probably nearer the latter number.'* [Cox 1946].

> *'[...]none of the usual interpretations of probability provide an adequate interpretation of probabilistic theories in science.'* [Schwarz 2018].

- Let's talk about:
  - frequentist probability → a measure of randomness
  - Bayesian probability /'reasonal expectation' → a measure of knowledge/uncertainty

Probabilities represent uncertainties?!

What actually is a probability?

- Several different schools of thought in probability

  > *'If minor differences are counted, the number of schools seems to be somewhere between two and the number of authors, and probably nearer the latter number.'* [Cox 1946].

  > *'[...]none of the usual interpretations of probability provide an adequate interpretation of probabilistic theories in science.'* [Schwarz 2018].

- Let's talk about:
  - frequentist probability $\rightarrow$ a measure of randomness
  - Bayesian probability /'reasonal expectation' $\rightarrow$ a measure of knowledge/uncertainty

Probabilities represent uncertainties?!

What actually is a probability?

- Several different schools of thought in probability

  > *'If minor differences are counted, the number of schools seems to be somewhere between two and the number of authors, and probably nearer the latter number.'* [Cox 1946].

  > *'[...]none of the usual interpretations of probability provide an adequate interpretation of probabilistic theories in science.'* [Schwarz 2018].

- Let's talk about:
  - frequentist probability $\rightarrow$ a measure of randomness
  - Bayesian probability /'reasonal expectation' $\rightarrow$ a measure of knowledge/uncertainty

What is a probability? Frequentist answer.

'If we repeat the complex $\mathcal{S}$ n times, where n is large, and let m be the number of occurrences of the event A, we can basically be sure, that $P(A)$ is not very different fom m/n' *[Kolmogorov 1933]*.[1]

- Let A be some event in a random experiment. Repeat the experiment n times independently and define

$$P_{\mathrm{freq}}(A) := \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} \mathbf{1}(A \text{ occured in experiment } i)$$

- Cox calls this structure "ensemble"

[1] *Original:* 'Man kann praktisch sicher sein, daß, wenn man den Komplex der Bedingungen $\mathcal{S}$ eine große Anzahl von n Malen wiederholt und dabei durch m die Anzahl der Fälle bezeichnet, bei denen das Ereignis A stattgefunden hat, das Verhältnis m/n sich von $P(A)$ wenig unterscheidet'.

What is a probability? Frequentist answer.

'If we repeat the complex $\mathcal{S}$ n times, where n is large, and let m be the number of occurrences of the event A, we can basically be sure, that $P(A)$ is not very different fom m/n' [Kolmogorov 1933].[1]

- Let A be some event in a random experiment. Repeat the experiment n times independently and define

$$P_{\text{freq}}(A) := \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \mathbf{1}(A \text{ occured in experiment } i)$$

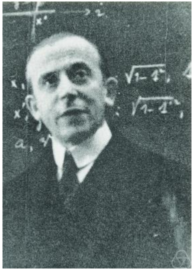
- Cox calls this structure "ensemble"

[1] *Original:* 'Man kann praktisch sicher sein, daß, wenn man den Komplex der Bedingungen $\mathcal{S}$ eine große Anzahl von n Malen wiederholt und dabei durch m die Anzahl der Fälle bezeichnet, bei denen das Ereignis A stattgefunden hat, das Verhältnis m/n sich von $P(A)$ wenig unterscheidet'

What is a probability? Frequentist answer.

- Kolmogorov was not the first one to discuss frequentist probability
- Richard von Mises, e.g., introduced a framework ("collectives") for a mathematical description of frequentist probabilities (at this stage not fully rigorous, it has been improved later)

Richard von Mises: *Wahrscheinlichkeit Statistik und Wahrheit*, Springer, 1928. (English translation version available: *Probability, Statistics, and Truth.*)



What is a probability? Bayesian answer.

'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 chance of the thing expected upon it's happening.' [Bayes 1763]

Let A be some event in a random experiment.

- Bayes: The probability $P_{\mathrm{Bayes}}(A)$ is the price we are willing to pay to win 1 if A occurs
- defines probability through reasonable expectation $\rightarrow$ randomness is not required

Coin flip examples.

Experiment.

We flip an idealised coin. $A := \{\text{coin shows head}\}$.

- Frequentist: $P_{\mathrm{freq}}(A) = \frac{1}{2}$
- Bayesian: $P_{\mathrm{Bayes}}(A) = \frac{1}{2}$

Experiment.

Someone places a coin unintentionally on a table. It is out of sight. $A := \{\text{coin shows head}\}$.

- Frequentist: $P_{\mathrm{freq}}(A)$ probably undefined; no ensemble
- Bayesian: $P_{\mathrm{Bayes}}(A) = \frac{1}{2}$

Coin flip examples.

Experiment.

We flip an idealised coin. $A := \{\text{coin shows head}\}$.

- Frequentist: $P_{\mathrm{freq}}(A) = \frac{1}{2}$
- Bayesian: $P_{\mathrm{Bayes}}(A) = \frac{1}{2}$

Experiment.

Someone places a coin unintentionally on a table. It is out of sight.
$A := \{\text{coin shows head}\}$.

- Frequentist: $P_{\mathrm{freq}}(A)$ probably undefined; no ensemble
- Bayesian: $P_{\mathrm{Bayes}}(A) = \frac{1}{2}$

Further examples by Cox

For the coin example it is rather clear when there is an ensemble

> *'there is so gradual a transition from the cases in which there is a discoverable ensemble and those in which there is none that a theory which requires a sharp distinction between them offers serious difficulties'* [Cox 1946]

- mathematical: *'finding the least number of cubes for the expression of large integers'*

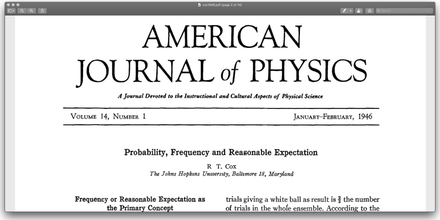
  $$\text{Given } N \in \mathbb{N} \text{ find } \min\left\{K \in \mathbb{N} : \forall n \in \mathbb{N}, n \geq N, \exists a \in \mathbb{N}_0^K : n = \sum_{k=1}^{K} a_k^3\right\} (\leq 9).$$

  (Is there an ensemble?)
- physical: finding a physical constant from measurements (is there an ensemble?)

Representing reasonable expectation through probability

- In the Bayesian framework, we do not require randomness in a parameter/object to describe its state with a probability distribution
- Is probability a good way to describe knowledge/uncertainty/reasonable expecation?

  Yes. Theorem by Cox [Cox 1946] 'shows'[2] that a fairly general 'measure of reasonable expectation' on a Boolean algebra (of logical statements) has to be a probability measure (actually: probabiliity content).



AMERICAN JOURNAL of PHYSICS

A Journal Devoted to the Instructional and Cultural Aspects of Physical Science

VOLUME 14, NUMBER 1 JANUARY–FEBRUARY, 1946

Probability, Frequency and Reasonable Expectation

R. T. Cox

The Johns Hopkins University, Baltimore 18, Maryland

Frequency or Reasonable Expectation as the Primary Concept

trials giving a white ball as result is ⅗ the number of trials in the whole ensemble. According to the

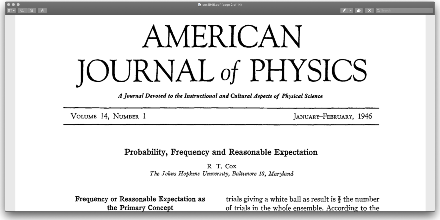
  No. There are some uncertainties we cannot really describe with probability. Try instead: random sets, fuzzy logic, possibility theory.

[2] rigorous version by Jaynes [Jaynes 2003]

Representing reasonable expectation through probability

- In the Bayesian framework, we do not require randomness in a parameter/object to describe its state with a probability distribution
- Is probability a good way to describe knowledge/uncertainty/reasonable expecation?

Yes. Theorem by Cox [Cox 1946] 'shows'[2] that a fairly general 'measure of reasonable expectation' on a Boolean algebra (of logical statements) has to be a probability measure (actually: probabiliity content).


AMERICAN JOURNAL *of* PHYSICS

A Journal Devoted to the Instructional and Cultural Aspects of Physical Science

VOLUME 14, NUMBER 1 JANUARY–FEBRUARY, 1946

Probability, Frequency and Reasonable Expectation

R. T. Cox

The Johns Hopkins University, Baltimore 18, Maryland

Frequency or Reasonable Expectation as the Primary Concept

trials giving a white ball as result is ⅗ the number of trials in the whole ensemble. According to the

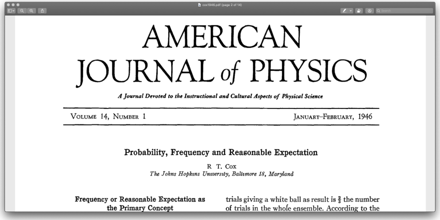
No. There are some uncertainties we cannot really describe with probability. Try instead: random sets, fuzzy logic, possibility theory.

[2] rigorous version by Jaynes [Jaynes 2003]

Representing reasonable expectation through probability

- In the Bayesian framework, we do not require randomness in a parameter/object to describe its state with a probability distribution
- Is probability a good way to describe knowledge/uncertainty/reasonable expecation?

Yes. Theorem by Cox [Cox 1946] 'shows'[2] that a fairly general 'measure of reasonable expectation' on a Boolean algebra (of logical statements) has to be a probability measure (actually: probabiliity content).



AMERICAN JOURNAL *of* PHYSICS

A Journal Devoted to the Instructional and Cultural Aspects of Physical Science

Volume 14, Number 1 — January–February, 1946

Probability, Frequency and Reasonable Expectation

R. T. Cox

The Johns Hopkins University, Baltimore 18, Maryland

Frequency or Reasonable Expectation as the Primary Concept

trials giving a white ball as result is $\frac{2}{3}$ the number of trials in the whole ensemble. According to the

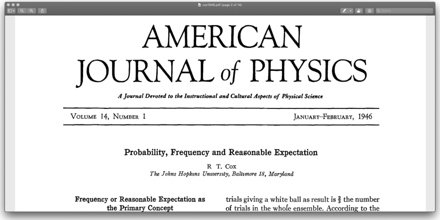
No. There are some uncertainties we cannot really describe with probability. Try instead: random sets, fuzzy logic, possibility theory.

[2] rigorous version by Jaynes [Jaynes 2003]

Representing reasonable expectation through probability

- In the Bayesian framework, we do not require randomness in a parameter/object to describe its state with a probability distribution
- Is probability a good way to describe knowledge/uncertainty/reasonable expecation?

  Yes. Theorem by Cox [Cox 1946] 'shows'[2] that a fairly general 'measure of reasonable expectation' on a Boolean algebra (of logical statements) has to be a probability measure (actually: probabiliity content).



AMERICAN JOURNAL of PHYSICS

A Journal Devoted to the Instructional and Cultural Aspects of Physical Science

VOLUME 14, NUMBER 1 JANUARY–FEBRUARY, 1946

Probability, Frequency and Reasonable Expectation

R. T. Cox

The Johns Hopkins University, Baltimore 18, Maryland

Frequency or Reasonable Expectation as the Primary Concept

trials giving a white ball as result is ⅗ the number of trials in the whole ensemble. According to the

  No. There are some uncertainties we cannot really describe with probability. Try instead: random sets, fuzzy logic, possibility theory.

[2] rigorous version by Jaynes [Jaynes 2003]

So in our inverse problem

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

we neither know f^* (parameter we care about), nor e (noise we don't care about). So we now model f^* and e as random variables

$$F : \Omega \to X, \qquad E : \Omega \to Y$$

on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We define the probability distributions

- $\mu_{\text{prior}} := \mathbb{P}(F \in \cdot)$ and call it prior (distribution); it describes our reasonable expectation regarding f^* before observing the data.
- $\mu_{\text{noise}} := \mathbb{P}(E \in \cdot)$; it describes randomness or reasonable expectation of the measurement error.

How do we now inform μ_{prior} about the data?

So in our inverse problem

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

we neither know f^* (parameter we care about), nor e (noise we don't care about). So we now model f^* and e as random variables

$$F : \Omega \to X, \qquad E : \Omega \to Y$$

on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We define the probability distributions

- $\mu_{\text{prior}} := \mathbb{P}(F \in \cdot)$ and call it prior (distribution); it describes our reasonable expectation regarding f^* before observing the data.
- $\mu_{\text{noise}} := \mathbb{P}(E \in \cdot)$; it describes randomness or reasonable expectation of the measurement error.

How do we now inform μ_{prior} about the data?

Learning through conditioning

Introduction: quantify uncertainties in inverse problems

Probabilities represent uncertainties

Learning through conditioning

Outlook

How do we inform μ_{prior} about the data?

When studying random experiments, we can adjust probability distributions when receiving partial information by conditioning.

Experiment.

We roll an idealised 6-sided die. Let $A := \{\text{die shows } 6\}$. Then, $\mathbb{P}(A) = 1/6$. Now, we replace the die by one that will only ever land on...

- ...even numbers, in which case we have $\mathbb{P}(A|\text{die shows only even number}) = 1/3$.
- ...odd numbers, in which case we have $\mathbb{P}(A|\text{die shows only odd number}) = 0$.

Similarly, we can learn information when using probability to represent reasonable expectation.

How do we inform μ_{prior} about the data?

When studying random experiments, we can adjust probability distributions when receiving partial information by conditioning.

Experiment.

We roll an idealised 6-sided die. Let $A := \{\text{die shows } 6\}$. Then, $\mathbb{P}(A) = 1/6$. Now, we replace the die by one that will only ever land on...

- ...even numbers, in which case we have $\mathbb{P}(A|\text{die shows only even number}) = 1/3$.
- ...odd numbers, in which case we have $\mathbb{P}(A|\text{die shows only odd number}) = 0$.

Similarly, we can learn information when using probability to represent reasonable expectation.

How do we inform μ_{prior} about the data?

When studying random experiments, we can adjust probability distributions when receiving partial information by conditioning.

Experiment.

We roll an idealised 6-sided die. Let $A := \{\text{die shows } 6\}$. Then, $\mathbb{P}(A) = 1/6$. Now, we replace the die by one that will only ever land on...

- ...even numbers, in which case we have $\mathbb{P}(A|\text{die shows only even number}) = 1/3$.
- ...odd numbers, in which case we have $\mathbb{P}(A|\text{die shows only odd number}) = 0$.

Similarly, we can learn information when using probability to represent reasonable expectation.

Learning from data in inverse problem

Observing the data g means knowing that the event

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

has occurred. So to update our reasonable expectation, we need to condition with respect to this event:

$$\mu_{\text{prior}} = \mathbb{P}(F \in \cdot) \longrightarrow \mathbb{P}(F \in \cdot | \mathcal{A}(F) + E = g) =: \mu_{\text{post}}.$$

We refer to the outcome as posterior (distribution) of F (given $\mathcal{A}(F) + E = g$).

Questions:

- What actually is a conditional distribution? How can we represent it?
- How do we get from prior to posterior?

Learning from data in inverse problem

Observing the data g means knowing that the event

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

has occurred. So to update our reasonable expectation, we need to condition with respect to this event:

$$\mu_{\text{prior}} = \mathbb{P}(F \in \cdot) \longrightarrow \mathbb{P}(F \in \cdot | \mathcal{A}(F) + E = g) =: \mu_{\text{post}}.$$

We refer to the outcome as posterior (distribution) of F (given $\mathcal{A}(F) + E = g$).

Questions:

- What actually is a conditional distribution? How can we represent it?
- How do we get from prior to posterior?

Some disclaimers...

In the following,

- parameter space X is a Polish/Radon space, data space Y is a separable Banach space (e.g., $X, Y \in \{\mathbb{Z}^n, \mathbb{R}^n, \mathcal{L}^2\}$)
- both these spaces are associated with Borel-σ-algebras $\mathcal{B}X$ and $\mathcal{B}Y$
- the prior μ_{prior} has a density π_{prior} with respect to the reference measure ν_X on $(X, \mathcal{B}X)$, i.e.
$$\mu_{\mathrm{prior}}(A) = \int_A \pi_{\mathrm{prior}}(x)\nu_X(\mathrm{d}x) \qquad (A \in \mathcal{B}X)$$
- We generalise the inverse problem a bit by allowing the data observation to be more general: we have dependent random variables $F : \Omega \to X$, $G : \Omega \to Y$, observe $\{G = g\}$, and aim to find
$$\mu_{\mathrm{post}} = \mathbb{P}(F \in \cdot | G = g).$$

Conditional distribution

We can define the conditional distribution

$$\mathbb{P}(F \in A | G = g) := \frac{\mathbb{P}(\{F \in A\} \cap \{G = g\})}{\mathbb{P}(G = g)} \qquad (A \in \mathcal{B}X)$$

if $\mathbb{P}(G = g) > 0$.

- elementary definition of conditional probability distributions
- only valid if $\mathbb{P}(G = g) > 0$, which is often not the case:
  - the noise distribution often is a continuous distribution (say, Gaussian), in which case $\mathbb{P}(G = g) = 0$
  - it is valid with, e.g., Poissonian noise, but we still need a more general set-up

Conditional distribution

We can define the conditional distribution

$$\mathbb{P}(F \in A | G = g) := \frac{\mathbb{P}(\{F \in A\} \cap \{G = g\})}{\mathbb{P}(G = g)} \qquad (A \in \mathcal{B}X)$$

if $\mathbb{P}(G = g) > 0$.

- elementary definition of conditional probability distributions
- only valid if $\mathbb{P}(G = g) > 0$, which is often not the case:
  - the noise distribution often is a continuous distribution (say, Gaussian), in which case $\mathbb{P}(G = g) = 0$
  - it is valid with, e.g., Poissonian noise, but we still need a more general set-up

Conditional distribution

We can define the conditional distribution

$$\mathbb{P}(F \in A | G = g) := \frac{\mathbb{P}(\{F \in A\} \cap \{G = g\})}{\mathbb{P}(G = g)} \qquad (A \in \mathcal{B}X)$$

if $\mathbb{P}(G = g) > 0$.

- elementary definition of conditional probability distributions
- only valid if $\mathbb{P}(G = g) > 0$, which is often not the case:
  - the noise distribution often is a continuous distribution (say, Gaussian), in which case $\mathbb{P}(G = g) = 0$
  - it is valid with, e.g., Poissonian noise, but we still need a more general set-up

A more general definition of a conditional distribution

Definition. The (regular) conditional distribution of F given $G = \cdot$ is given by the Markov kernel

$$M : \mathcal{B}X \times Y \to [0,1]$$

that satisfies

$$\mathbb{P}(F \in A, G \in B) = \int_B M(A|g)\mathbb{P}(G \in \mathrm{d}g) \quad (= \mathbb{E}[\mathbf{1}[G \in B]M(A|G)]).$$

We write $\mathbb{P}(F \in A|G = g) := M(A|g)$, for $A \in \mathcal{B}X$ and $g \in Y$.

Theorem. $\mathbb{P}(F \in \cdot|G = \cdot)$ exists and is $\mathbb{P}(G \in \cdot)$-almost surely unique.

A more general definition of a conditional distribution

Definition. The (regular) conditional distribution of F given $G = \cdot$ is given by the Markov kernel

$$M : \mathcal{B}X \times Y \to [0, 1]$$

that satisfies

$$\mathbb{P}(F \in A, G \in B) = \int_B M(A|g)\mathbb{P}(G \in \mathrm{d}g) \quad (= \mathbb{E}[\mathbf{1}[G \in B]M(A|G)]).$$

We write $\mathbb{P}(F \in A|G = g) := M(A|g)$, for $A \in \mathcal{B}X$ and $g \in Y$.

Theorem. $\mathbb{P}(F \in \cdot|G = \cdot)$ exists and is $\mathbb{P}(G \in \cdot)$-almost surely unique.

Conditional distribution

Idea: Given two random variables F, G, the conditional distribution of F given $G = g$ allows us to represent their dependency in a two-step procedure:

1. Sample $G' \sim \mathbb{P}(G \in \cdot)$,
2. Sample $F' \sim \mathbb{P}(F \in \cdot | G = G')$.

Then, $(F', G') \stackrel{d}{=} (F, G)$. (← precise interpretation of the definition)

$\Rightarrow \mathbb{P}(F \in \cdot | G = g)$ explains how F needs to behave, if G is already known such that the joint distribution (F, G) keeps being correct.

Conditional distribution

Idea: Given two random variables F, G, the conditional distribution of F given $G = g$ allows us to represent their dependency in a two-step procedure:

1. Sample $G' \sim \mathbb{P}(G \in \cdot)$,
2. Sample $F' \sim \mathbb{P}(F \in \cdot | G = G')$.

Then, $(F', G') \overset{d}{=} (F, G)$. ($\leftarrow$ precise interpretation of the definition)

$\Rightarrow \mathbb{P}(F \in \cdot | G = g)$ explains how F needs to behave, if G is already known such that the joint distribution (F, G) keeps being correct.

Conditional densities

The definition of the conditional distribution is not constructive. To actually construct a conditional distribution, we can use densities.

Assumption. Let ν_X and ν_Y be reference measures on X and Y, respectively. We assume that the joint distribution $\mathbb{P}((F, G)) \in \cdot)$ has a (joint) density $\pi_{F,G}$ with respect to $\nu_X \otimes \nu_Y$, i.e.

$$\mathbb{P}((F, G)) \in A \times B) = \int_A \int_B \pi_{F,G}(f, g)\nu_Y(\mathrm{d}g)\nu_X(\mathrm{d}f).$$

Moreover, we define the marginal densities

- of F: $\pi_F = \int_Y \pi_{F,G}(\cdot, g)\nu_Y(\mathrm{d}g)$ and
- of G: $\pi_G = \int_X \pi_{F,G}(f, \cdot)\nu_X(\mathrm{d}f)$

Conditional densities

The definition of the conditional distribution is not constructive. To actually construct a conditional distribution, we can use densities.

Assumption. Let ν_X and ν_Y be reference measures on X and Y, respectively. We assume that the joint distribution $\mathbb{P}((F,G)) \in \cdot)$ has a (joint) density $\pi_{F,G}$ with respect to $\nu_X \otimes \nu_Y$, i.e.

$$\mathbb{P}((F,G)) \in A \times B) = \int_A \int_B \pi_{F,G}(f,g)\nu_Y(\mathrm{d}g)\nu_X(\mathrm{d}f).$$

Moreover, we define the marginal densities

- of F: $\pi_F = \int_Y \pi_{F,G}(\cdot, g)\nu_Y(\mathrm{d}g)$ and
- of G: $\pi_G = \int_X \pi_{F,G}(f, \cdot)\nu_X(\mathrm{d}f)$

Conditional densities

Theorem. Under the previously mentioned assumptions, $\mathbb{P}(F \in \cdot | G = g)$ has a density with respect to ν_X, given by

$$\pi_{F|G=g} = \begin{cases} \frac{\pi_{F,G}(\cdot,g)}{\pi_G(g)}, & \text{if } \pi_G(g) > 0, \\ 0, & \text{otherwise} \end{cases} \qquad (g \in Y, \mathbb{P}(G \in \cdot)\text{-a.s.}, f \in X, \nu_X\text{-a.e.}).$$

The same is true for $\mathbb{P}(G \in \cdot | F = f)$, which has the density $\pi_{G|F=f}$ with respect to ν_Y:

$$\pi_{G|F=f} = \begin{cases} \frac{\pi_{F,G}(f,\cdot)}{\pi_F(f)}, & \text{if } \pi_F(f) > 0, \\ 0, & \text{otherwise} \end{cases} \qquad (f \in X, \mathbb{P}(F \in \cdot)\text{-a.s.}, g \in Y, \nu_Y\text{-a.e.}).$$

Exercise. The elementary definition of the conditional probability is a special case in the theorem given above.

Conditional densities

Theorem. Under the previously mentioned assumptions, $\mathbb{P}(F \in \cdot | G = g)$ has a density with respect to ν_X, given by

$$\pi_{F|G=g} = \begin{cases} \frac{\pi_{F,G}(\cdot, g)}{\pi_G(g)}, & \text{if } \pi_G(g) > 0, \\ 0, & \text{otherwise} \end{cases} \qquad (g \in Y, \mathbb{P}(G \in \cdot)\text{-a.s.}, f \in X, \nu_X\text{-a.e.}).$$

The same is true for $\mathbb{P}(G \in \cdot | F = f)$, which has the density $\pi_{G|F=f}$ with respect to ν_Y:

$$\pi_{G|F=f} = \begin{cases} \frac{\pi_{F,G}(f, \cdot)}{\pi_F(f)}, & \text{if } \pi_F(f) > 0, \\ 0, & \text{otherwise} \end{cases} \qquad (f \in X, \mathbb{P}(F \in \cdot)\text{-a.s.}, g \in Y, \nu_Y\text{-a.e.}).$$

Exercise. The elementary definition of the conditional probability is a special case in the theorem given above.

Recap and outlook

Introduction: quantify uncertainties in inverse problems

Probabilities represent uncertainties

Learning through conditioning

Outlook

Recap and outlook

Recap.

- Inverse problem solutions may contain (hidden) uncertainties that we may need to quantify.
- Probabilities are a great way to describe our reasonable expectation regarding uncertain objects; conditioning allows us to learn in this framework.
- Conditional distributions are much more complicated than we thought, conditional densities allow us to represent them easily, though.

Outlook.

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

Recap and outlook

Recap.

- Inverse problem solutions may contain (hidden) uncertainties that we may need to quantify.
- Probabilities are a great way to describe our reasonable expectation regarding uncertain objects; conditioning allows us to learn in this framework.
- Conditional distributions are much more complicated than we thought, conditional densities allow us to represent them easily, though.

Outlook.

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