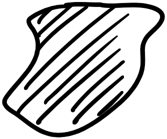
What is Uncertainty Quantification?

Philipp Wacker

February 2022

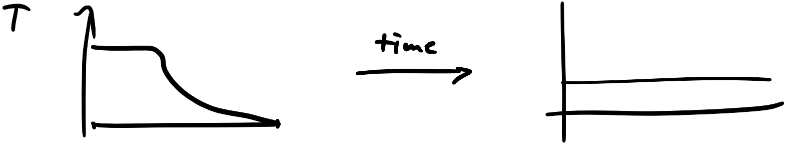
Calculation: Obtaining quantifiable information about objects

- **Given** a geometrical object, what is its volume/diameter/mean curvature?



Calculation: Obtaining quantifiable information about objects

- **Given** a geometrical object, what is its volume/diameter/mean curvature?
- **Given** an initial heat distribution, what is an object's temperature evolution?

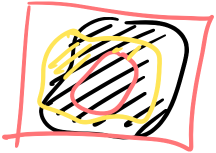


Calculation: Obtaining quantifiable information about objects

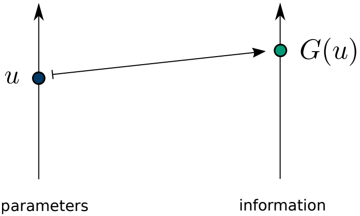
- **Given** a geometrical object, what is its volume/diameter/mean curvature?
- **Given** an initial heat distribution, what is an object's temperature evolution?
- On manufacturing a product: **Given** parameters like demand, price elasticity, production cost etc; What is the optimal amount produced and what is the best price?

params $\longrightarrow$ info.

Calculation: Obtaining quantifiable information about objects



- **Given** a geometrical object, what is its volume/diameter/mean curvature?
- **Given** an initial heat distribution, what is an object's temperature evolution?
- On manufacturing a product: **Given** parameters like demand, price elasticity, production cost etc; What is the optimal amount produced and what is the best price?

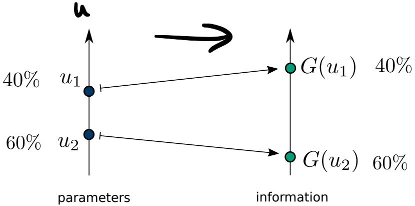


Propagation of uncertainty

Given only **a rough idea** of your parameters, what is the "best guess" information?

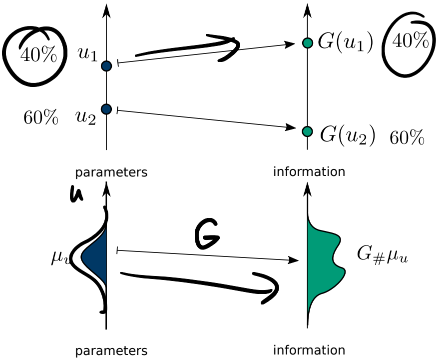
Propagation of uncertainty

Given only **a rough idea** of your parameters, what is the "best guess" information?



Propagation of uncertainty

Given only **a rough idea** of your parameters, what is the "best guess" information?

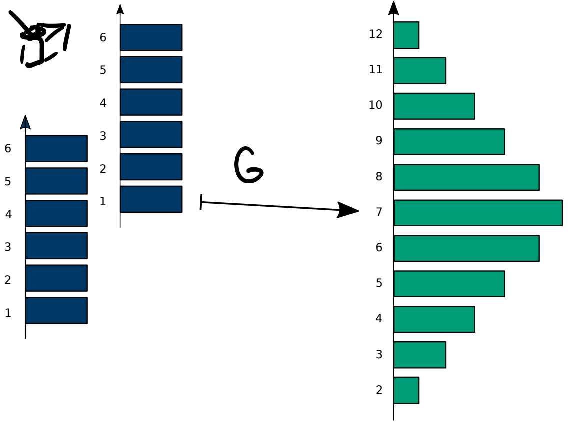


Push-forward of measures:

$$G_\#\mu(A) := \mu(G^{-1}(A))$$

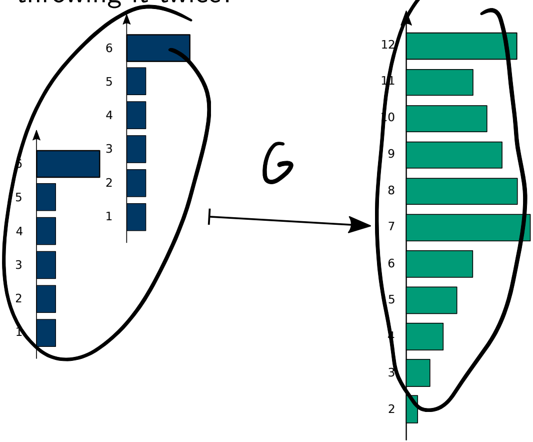
Example

Take a usual 6-sided die. What is the "best guess" sum of throwing it twice?



Example

Take a usual 6-sided die. What is the "best guess" sum of throwing it twice?

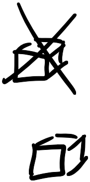


Forward

Now we know how to propagate uncertainty forward. Can we go the other way?

Forward

Now we know how to propagate uncertainty forward. Can we go the other way?
Given the sum of two dice is 10. What is the probability that the first die showed “2” / showed “5”?



$1/3$

$\Sigma = 10$

Forward

Now we know how to propagate uncertainty forward. Can we go the other way?
Given the sum of two dice is 10. What is the probability that the first die showed "2" / showed "5"?
Bayes' law allows inversion of probabilistic relations:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B)}$$

$$\text{posterior} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}$$

where $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

Bayes' Law: Common Sense – Amplified

Unknown parameter $u \sim \rho_0$. We are **given a measurement** y such that

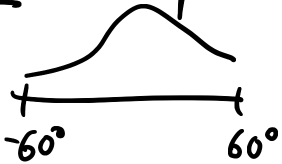
$$y = G(u) + \epsilon$$

Bayesian inversion

and $\boxed{\epsilon \sim \rho_\epsilon}$ is measurement noise. We assume that we know the densities ρ_0 and ρ_ϵ, but not u and ϵ.

1. $\rho_\varepsilon \sim N(0, \sigma^2)$ noise known statistically

2. ρ_0 prior on parameter space



Bayes' Law: Common Sense – Amplified

Unknown parameter $u \sim \rho_0$. We are **given a measurement** y such that

$$y = G(u) + \epsilon$$

and $\epsilon \sim \rho_\epsilon$ is measurement noise. We assume that we know the densities ρ_0 and ρ_ϵ, but not u and ϵ.
What is the correct "backward propagation" of this information y?
What is "u **given** y"?

Bayes' Law: Common Sense – Amplified

Unknown parameter $u \sim \rho_0$. We are **given a measurement** y such that

$$y = G(u) + \epsilon$$

and $\epsilon \sim \rho_\epsilon$ is measurement noise. We assume that we know the densities ρ_0 and ρ_ϵ, but not u and ϵ.
What is the correct "backward propagation" of this information y?
What is "u **given** y"?
Does not work:

$$u = G^{-1}(y)$$

- $\epsilon \neq 0$

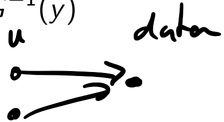
Bayes' Law: Common Sense – Amplified

Unknown parameter $u \sim \rho_0$. We are **given a measurement** y such that

$$y = G(u) + \epsilon$$

and $\epsilon \sim \rho_\epsilon$ is measurement noise. We assume that we know the densities ρ_0 and ρ_ϵ, but not u and ϵ.
What is the correct "backward propagation" of this information y?
What is "u **given** y"?
Does not work:

$$u = G^{-1}(y)$$



- $\epsilon \neq 0$
- G not one-to-one

Bayes' Law: Common Sense – Amplified

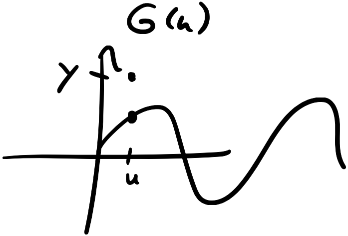
Unknown parameter $u \sim \rho_0$. We are **given a measurement** y such that

$\in$ range(G)

$$y = \boxed{G(u)} + \epsilon$$

and $\epsilon \sim \rho_\epsilon$ is measurement noise. We assume that we know the densities ρ_0 and ρ_ϵ, but not u and ϵ.
What is the correct "backward propagation" of this information y?
What is "u **given** y"?
Does not work:

$$u = G^{-1}(y)$$



- $\epsilon \neq 0$
- G not one-to-one
- $y \notin \text{range}(G)$

Bayes' Law: Common Sense – Amplified

Unknown parameter $u \sim \rho_0$. We are **given a measurement** y such that

$$y = G(u) + \epsilon$$

and $\epsilon \sim \rho_\epsilon$ is measurement noise. We assume that we know the densities ρ_0 and ρ_ϵ, but not u and ϵ.
What is the correct "backward propagation" of this information y?
What is "u **given** y"?
Does not work:

$$u = G^{-1}(y)$$

- $\epsilon \neq 0$
- G not one-to-one
- $y \notin \text{range}(G)$
- G^{-1} numerically unstable

typical: G compact

Backward via Bayes' law

Let $u \sim \rho_0$ and $\epsilon \sim \rho_\epsilon$.

$$y = G(u) + \epsilon$$

Backward via Bayes' law

Let $u \sim \rho_0$ and $\epsilon \sim \rho_\epsilon$.

data, unknown, unknown, but irrelevant

$$y = G(u) + \epsilon \Rightarrow \mathbb{P}(y|u)$$

Then $u|y \sim \rho_y$ where

$$\rho_y(\underline{u}) = \frac{\rho_\epsilon(y - G(\underline{u})) \cdot \rho_0(\underline{u})}{Z} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}$$

or formally

$$\mathbb{P}(u|y) = \frac{\mathbb{P}(y|u) \cdot \mathbb{P}(u)}{\mathbb{P}(y)}.$$

distr. of u given data y

Backward via Bayes' law

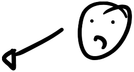
Let $u \sim \rho_0$ and $\epsilon \sim \rho_\epsilon$.

$$y = G(u) + \epsilon$$

Then $u|y \sim \rho_y$ where

$$\rho_y(\underline{u}) = \frac{\rho_\epsilon(y - G(\underline{u})) \cdot \rho_0(\underline{u})}{Z} = \frac{\text{likelihood} \ \cdot \ \text{prior}}{\text{evidence}}$$

or formally



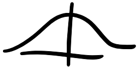
$$\mathbb{P}(u|y) = \frac{\mathbb{P}(y|u) \cdot \mathbb{P}(u)}{\mathbb{P}(y)}.$$

This can be made rigorous

- for singular/continuous distributions and
- in very general infinite-dimensional spaces.

Example

data param



Let $y = \exp(-u) + \epsilon$ with $u \sim N(0, 5^2)$ and $\underline{\epsilon \sim N(0, 1)}$.
Assume that $y = -1.0$.

$\varepsilon \in [-2, 2]$

prior

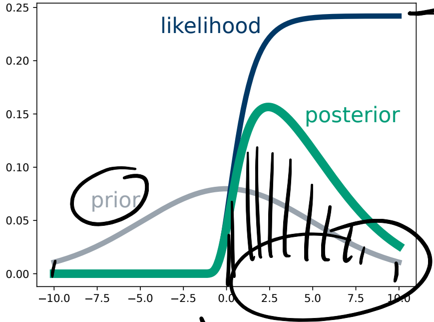
$0 \notin \text{range}(\exp(-\cdot))$

Example

prior

-1.0

Let $y = \exp(-u) + \epsilon$ with $u \sim N(0, 5^2)$ and $\epsilon \sim N(0, 1)$.
Assume that $y = -1.0$.



maximum likelihood approach

$\longrightarrow u$

forward

e^{-u}

Bayes' Law: Common Sense – Amplified

Bayes' Law is the mathematically correct way of merging

prior knowledge and (possibly faulty) data

into

posterior knowledge.

Bayes' Law: Common Sense – Amplified

Bayes' Law is the mathematically correct way of merging

prior knowledge   and   (possibly faulty) data

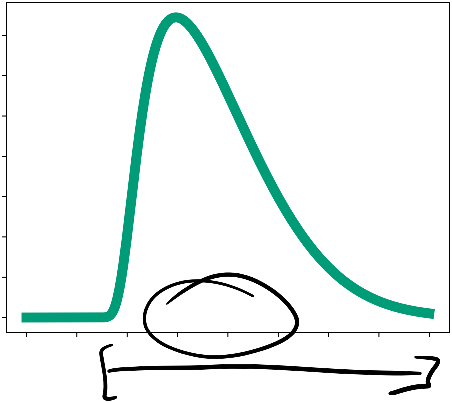
into

posterior knowledge.

This can be done iteratively: The posterior becomes the new prior and can be combined with new data.

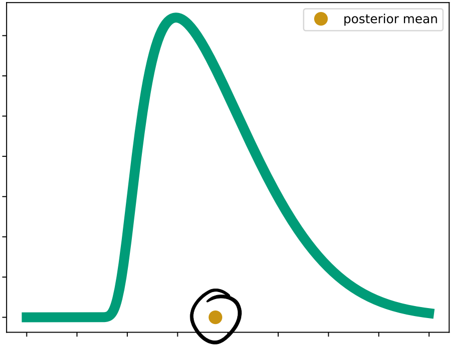
Now what?

Bayes' answer is a **measure**. How do we use it (especially in high/infinite dimensions)? $d > 2$



Now what?

Bayes' answer is a **measure**. How do we use it (especially in high/infinite dimensions)?

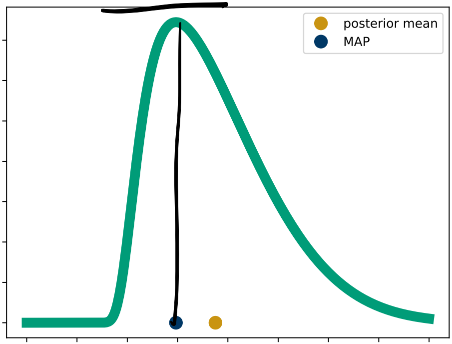


- Posterior mean.

(a parameter)

Now what?

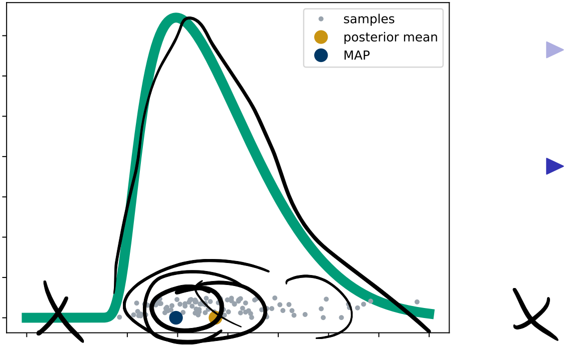
Bayes' answer is a **measure**. How do we use it (especially in high/infinite dimensions)?



- Posterior mean.
- maximum a posteriori estimator.

Now what?

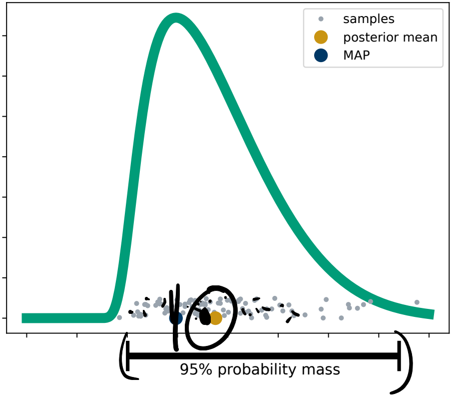
Bayes' answer is a **measure**. How do we use it (especially in high/infinite dimensions)?



- Posterior mean.
- maximum a posteriori estimator.
- samples (typical events).

Now what?

Bayes' answer is a **measure**. How do we use it (especially in high/infinite dimensions)?



- Posterior mean.
- maximum a posteriori estimator.
- samples (typical events).
- uncertainty information/ spread.

Extracting information from a (posterior) measure

- Computing the maximum a posteriori estimator (MAP) is an optimization problem

Extracting information from a (posterior) measure

- Computing the maximum a posteriori estimator (MAP) is an optimization problem
- Sampling – How do we obtain them?

Extracting information from a (posterior) measure

- Computing the maximum a posteriori estimator (MAP) is an optimization problem
- Sampling – How do we obtain them?
- Computing the posterior mean can be reduced to sampling:

$$\mathbb{E}_\mu X \approx M_N := \underbrace{\frac{1}{N}\sum_{i=1}^{N}}X_i$$

if $X_i \sim \mu$ are samples.

Extracting information from a (posterior) measure

- Computing the maximum a posteriori estimator (MAP) is an optimization problem
- Sampling – How do we obtain them?
- Computing the posterior mean can be reduced to sampling:

$$\mathbb{E}_\mu X \approx M_N := \frac{1}{N}\sum_{i=1}^{N} X_i$$

  if $X_i \sim \mu$ are samples.
- Computing uncertainty information can also be done via sampling:

Emp. covariance

$$\mathbb{E}_\mu[X - \mathbb{E}_\mu X]^2 \approx \frac{1}{N}\sum_{i=1}^{N}(X_i - M_n)^2$$

One slightly less trivial example

Assume data from an (unknown) quadratic function $f(x) = a + b \cdot x + c \cdot x^2$ via

param. $u = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$

pointwise (noisy) eval. $y_i = f(x_i) + \epsilon_i$

at positions x_i with measurement noise $\epsilon_i \sim N(0, \sigma^2)$ where σ is known.

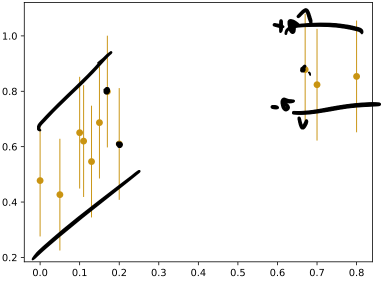
One slightly less trivial example

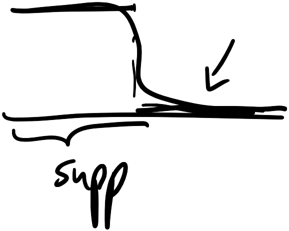
Assume data from an (unknown) quadratic function $f(x) = a + b \cdot x + c \cdot x^2$ via

$(x, \varepsilon) \sim \mu_{x,\varepsilon}$

$$y_i = f(x_i) + \epsilon_i$$

at positions x_i with measurement noise $\epsilon_i \sim N(0, \sigma^2)$ where σ is known



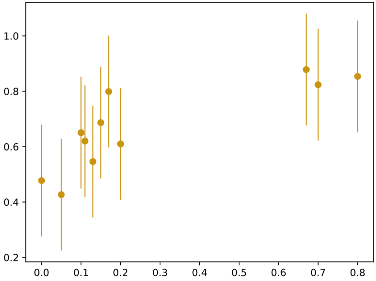


One slightly less trivial example

Assume data from an (unknown) quadratic function $f(x) = a + b \cdot x + c \cdot x^2$ via

$$y_i = f(x_i) + \epsilon_i$$

at positions x_i with measurement noise $\epsilon_i \sim N(0, \sigma^2)$ where σ is known

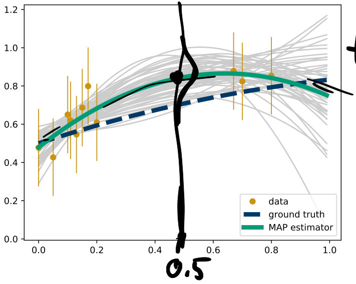


What is the "correct" set of parameters (a, b, c) that generated the data?

$\rightarrow$ measure on $\mathbb{R}^3$ (parameter space)

(a_k, b_k, c_k)

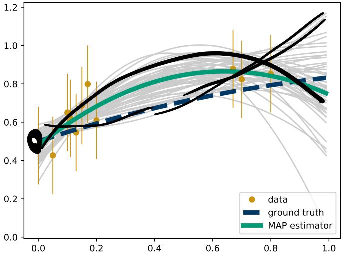
Example (cont'd)



Data space

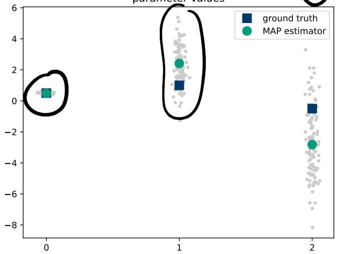
$f \quad a_{MAP}, b_{MAP}, c_{MAP}$

Example (cont'd)



Data space

$a + bx + cx^2$



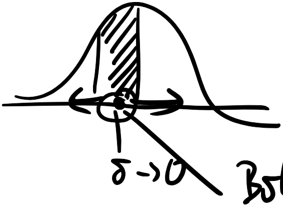
MAP

argmax posterior density

$\int_X x \, d\mu(x)$

$\int_X \ldots dx$

Radon-Nikodym density w.r.t. Lebesgue



$\delta \to 0$

$\{x : d(x,z) < \delta\} = B_\delta(z)$