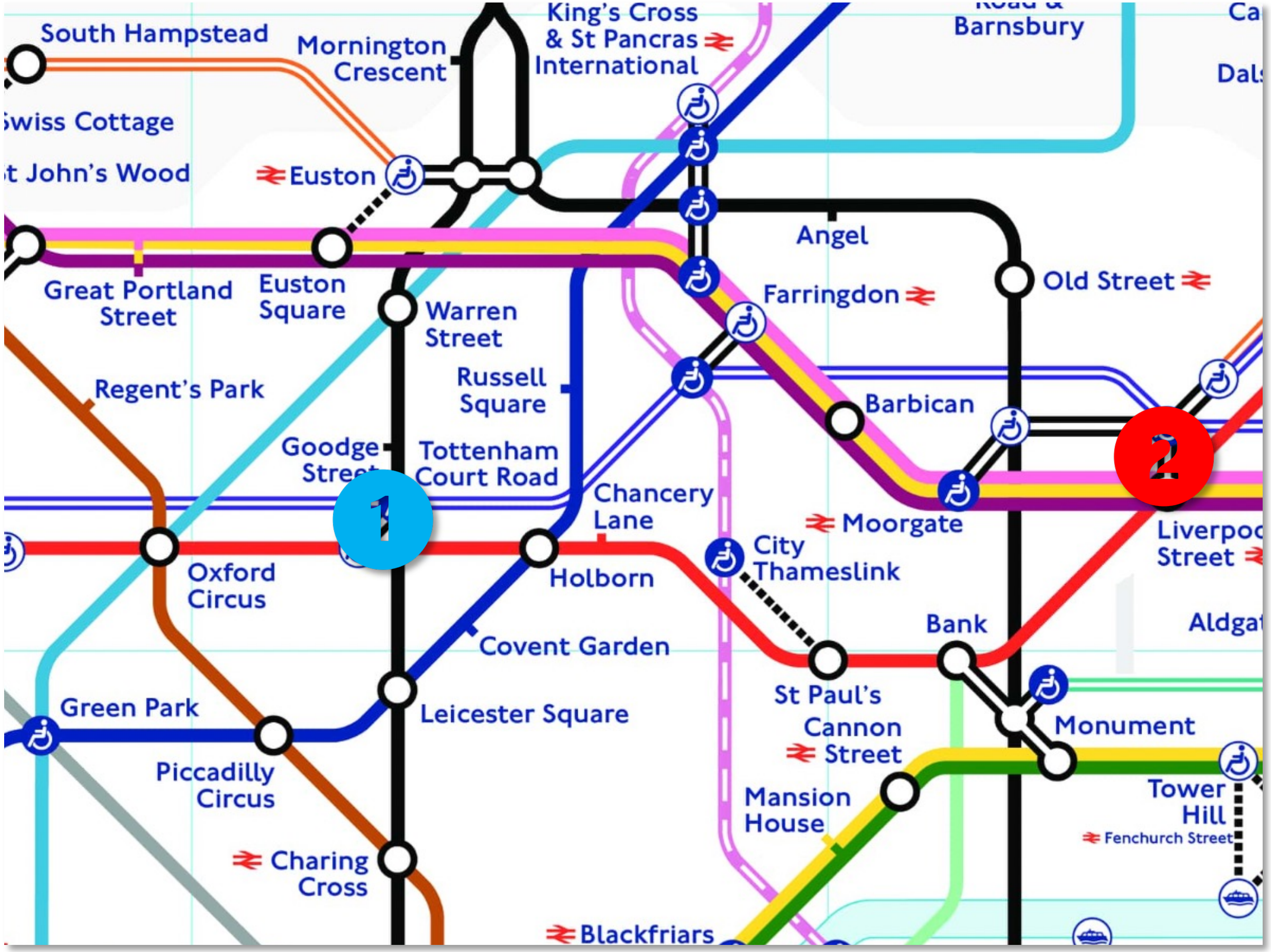
Bayesian Inference

Noor Sajid
Wellcome Centre for Human Neuroimaging, UCL

Special thanks to Peter Ziedman, Chris Mathys, Jean Daunizeau, Jérémie Mattout and Karl Friston for previous versions of this talk



Which route would you choose?

Overview

Brief primer on ill-posed problems

Random variables and PDFs

Axioms of probability

Bayes rule and approximate inference

Demo: variational Bayes in SPM

Overview

Brief primer on ill-posed problems

Random variables and PDFs

Axioms of probability

Bayes rule and approximate inference

Demo: variational Bayes in SPM

Primer

Formally, define the probability of an event, e:

$$p(e) = \frac{\sum e}{\sum s}, \in [0,1]$$

where s the sample space of all possible outcomes.

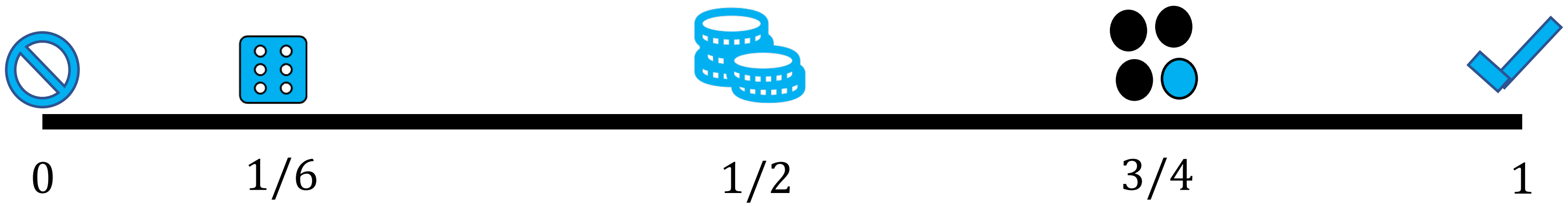
Holds under the assumption that each outcome in the sample space is equally likely

Primer

Formally, define the probability of an event, e:

$$p(e) = \frac{\sum e}{\sum s}, \in [0,1]$$

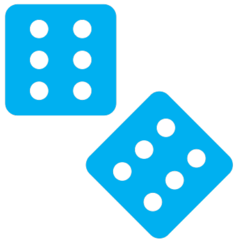
where s the sample space of all possible outcomes.



Primer

Formally, define the probability of an event, e:

$$p(e) = \frac{\sum e}{\sum s}, \in [0,1]$$

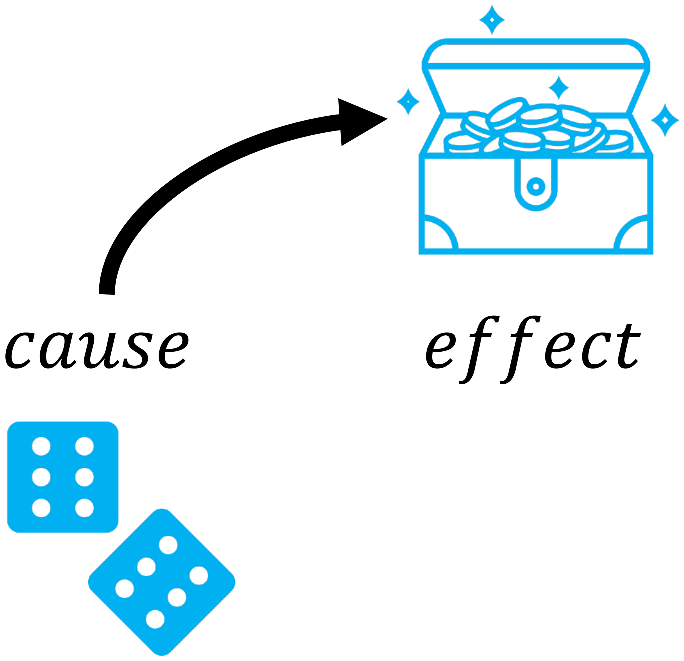


Fair a six–sided die, the sample space:

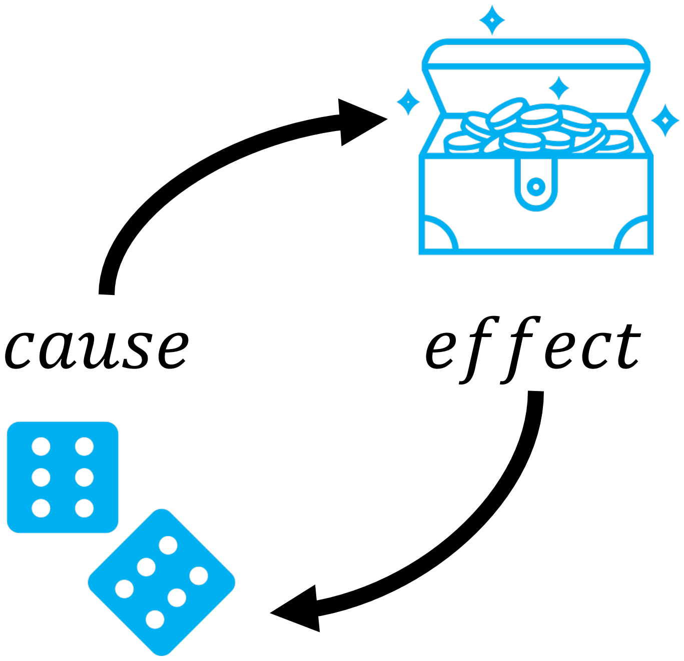
$$\text{s} \in \{1, 2, 3, 4, 5, 6\}\ i.e., \sum s = 6$$

if $e \in \{2,4,6\}\ i.e., \sum e = 3 \Rightarrow p(e) = \frac{3}{6} = 0.5$

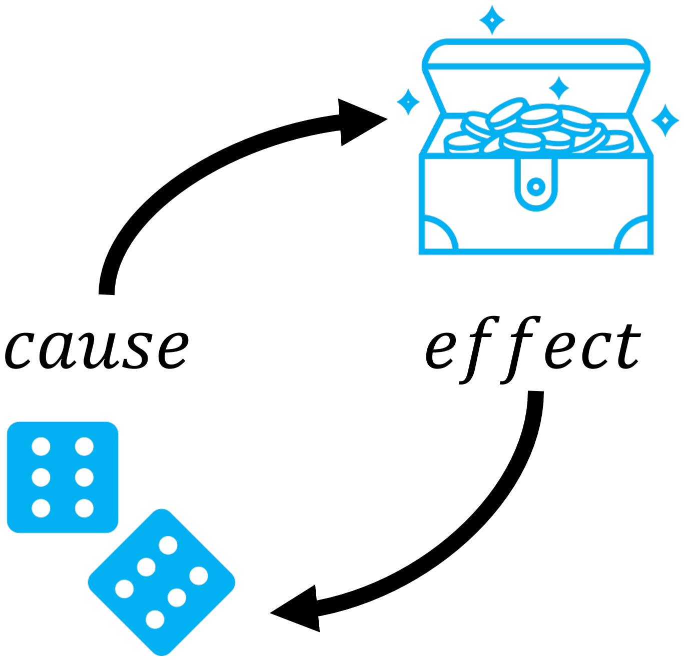
Ill-posed problem



Ill-posed problem



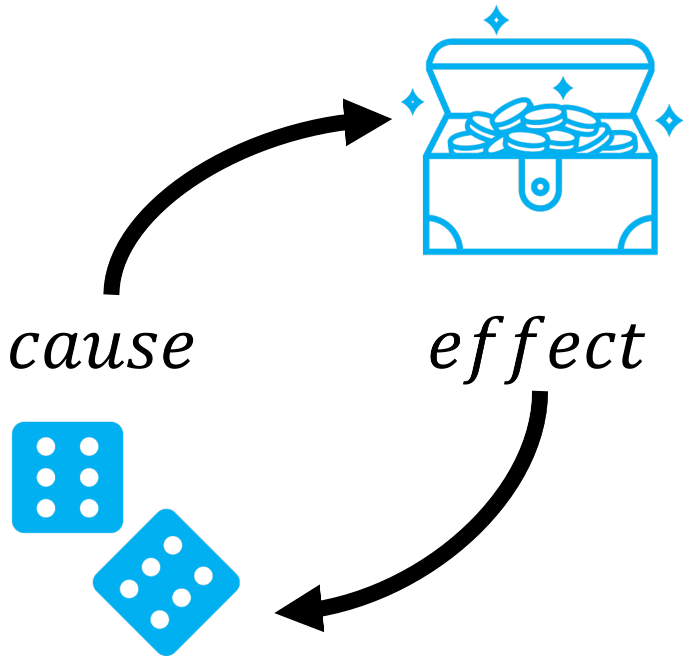
Ill-posed problem



Forward: $p(effect|cause) \in [0,1]$

Inverse: $p(cause|effect) \in [0,1]$

Ill-posed problem



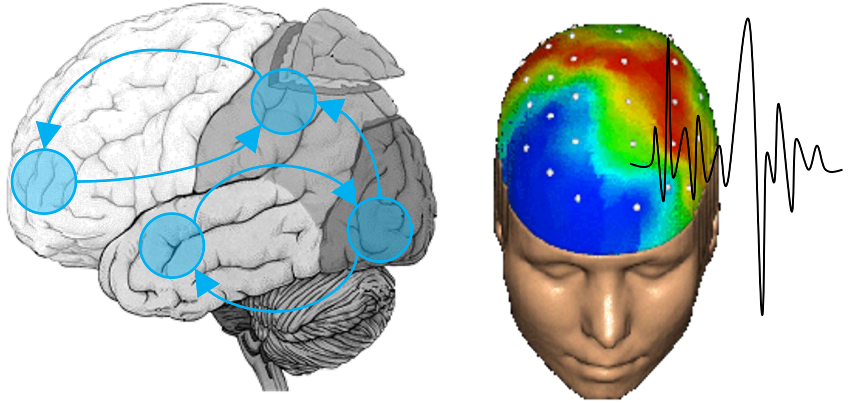
$$p(cause|effect) = \frac{p(effect|cause)p(cause)}{p(effect)}$$

Forward: $p(effect|cause) \in [0,1]$

Inverse: $p(cause|effect) \in [0,1]$

Example of ill-posed problem*

$recordings, y$

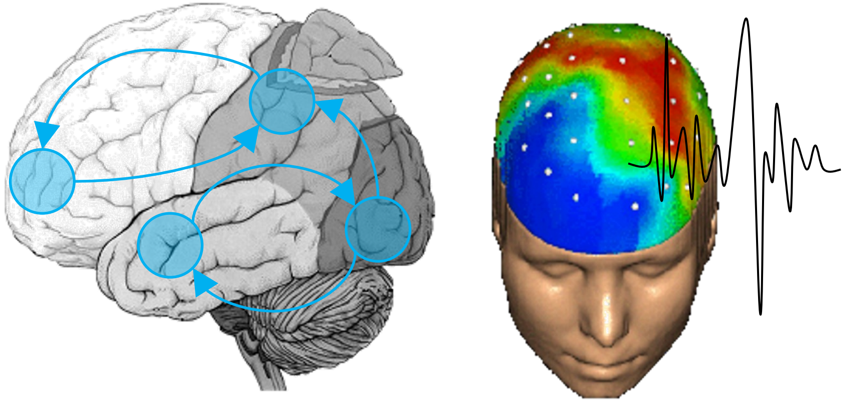


$neural\ source:\ \theta$

*Example from Peter Zeidman and Chris Mathys

Example of ill-posed problem*

$recordings, \mathrm{y}$



$neural\ source\!: \ \theta$

$$\theta = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \mathrm{y} = \begin{bmatrix} 0.2 & \cdots & 0.9 \\ \vdots & \ddots & \vdots \\ 0.4 & \cdots & 0.1 \end{bmatrix}$$

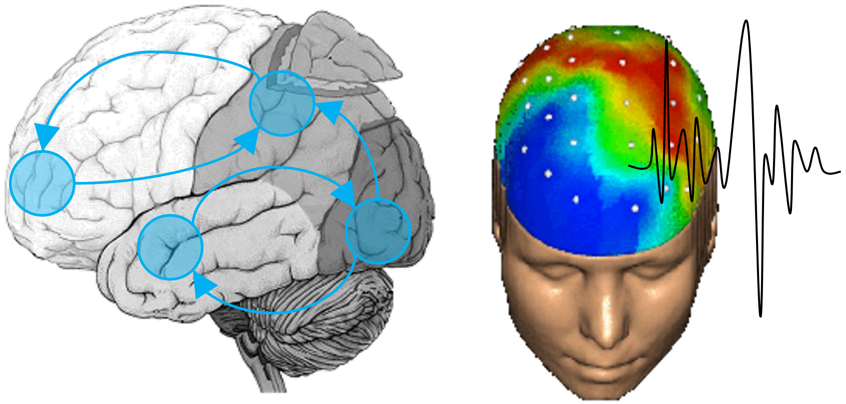
$$p(\theta = [\cdot] | y)$$

$$i.e., p(cause|effect)$$

*Example from Peter Zeidman and Chris Mathys

Example of ill-posed problem*

recordings, y



neural source: θ

$$\theta = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \mathrm{y} = \begin{bmatrix} 0.2 & \cdots & 0.9 \\ \vdots & \ddots & \vdots \\ 0.4 & \cdots & 0.1 \end{bmatrix}$$

$$p(\theta = [\cdot] | y)$$
$$i.e., p(cause|effect)$$

ill-posed problems

*Example from Peter Zeidman and Chris Mathys

Overview

Brief primer on ill-posed problems

Random variables and PDFs

Axioms of probability

Bayes rule and approximate inference

Demo: variational Bayes in SPM

Random variables

Random variable, X is a function that assigns a real number to each outcome in the sample space, Ω, of a random process:

$$X: \Omega \rightarrow \mathbb{R},$$

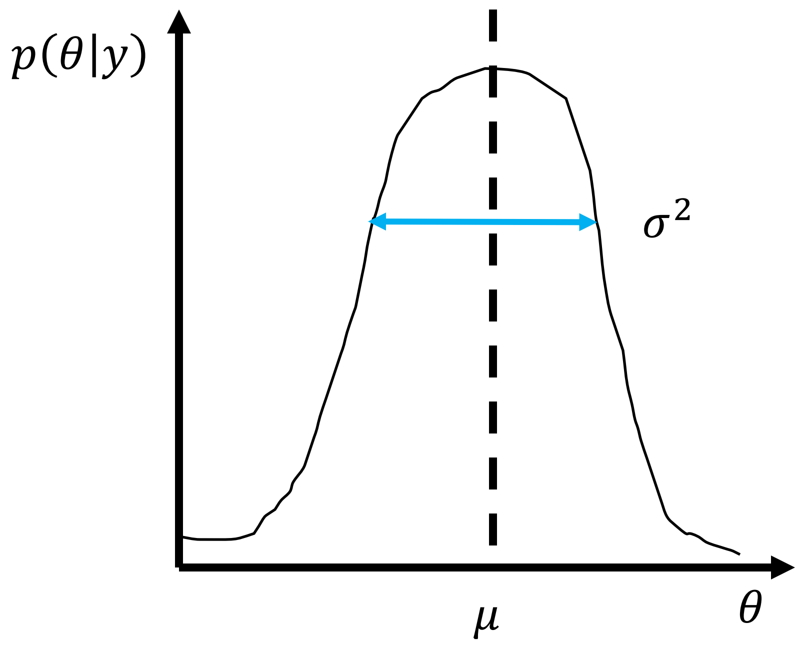
where $\mathbb{R}$ denotes the set of real numbers.

We can have two types of random variables: discrete and continuous

Example of continuous random variable

Here, we want to infer:

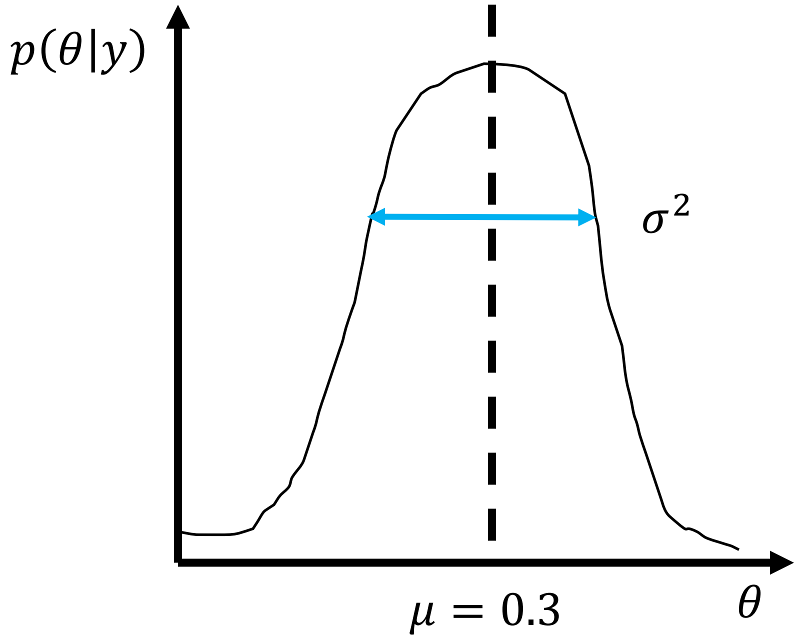
$$p(\theta|y) \sim \mathcal{N}(\mu, \sigma^2)$$

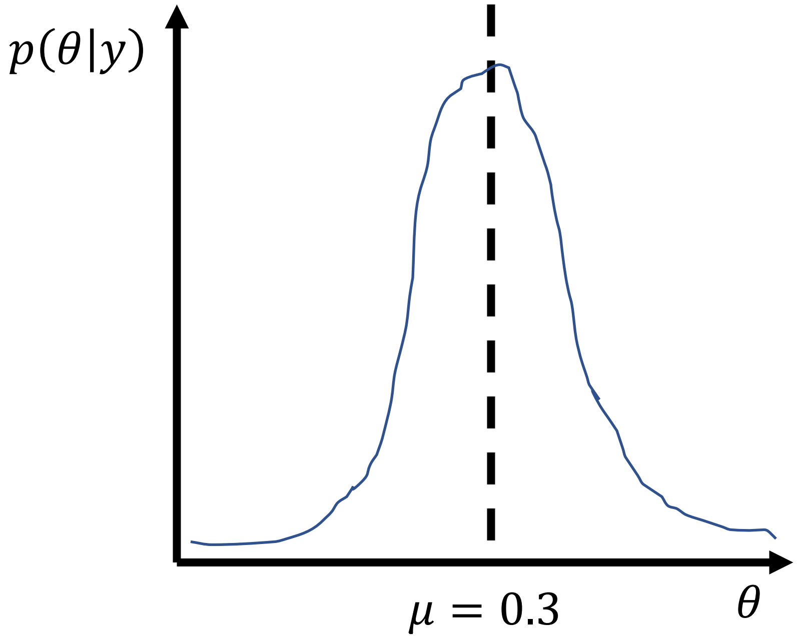


Example of continuous random variable

Here, we want to infer:

$$p(\theta|y) \sim \mathcal{N}(\mu, \sigma^2)$$

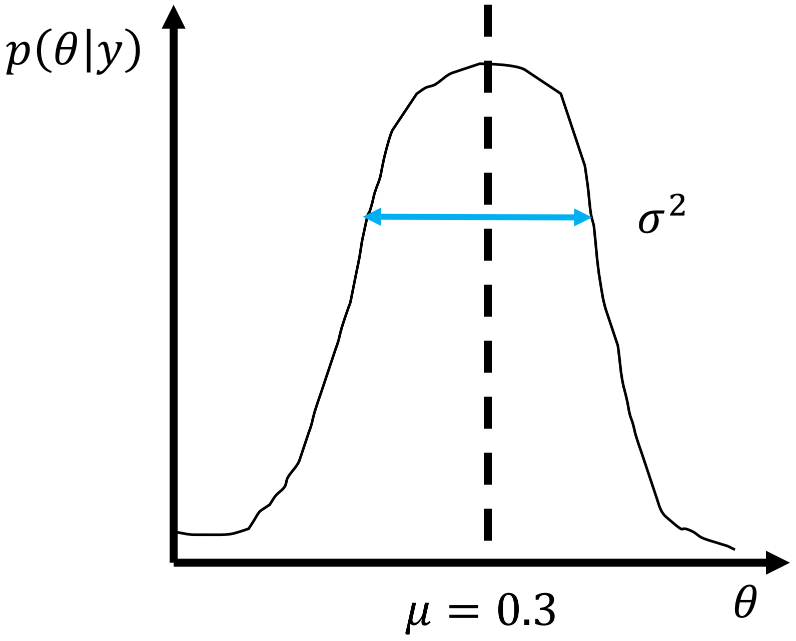


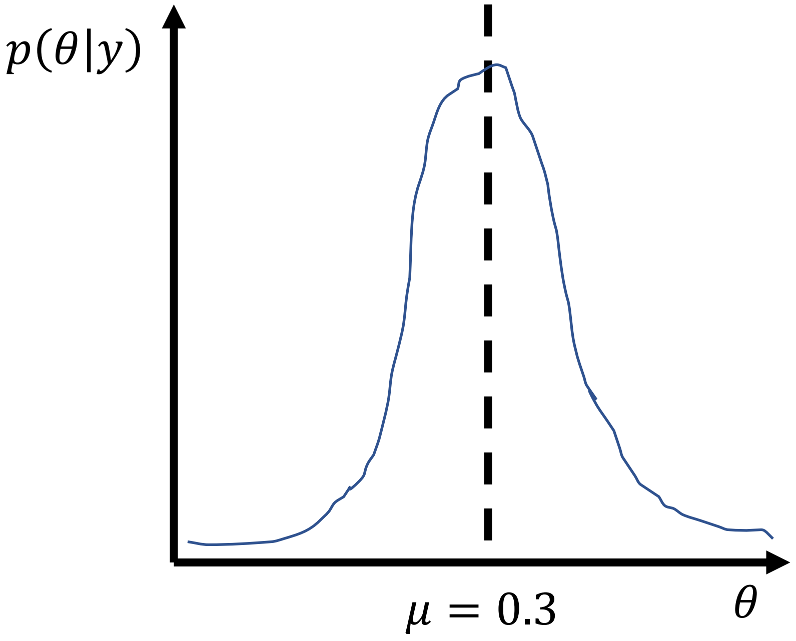


Example of continuous random variable

Here, we want to infer:

$$p(\theta|y) \sim \mathcal{N}(\mu, \sigma^2), \pi = \frac{1}{\sigma^2}$$

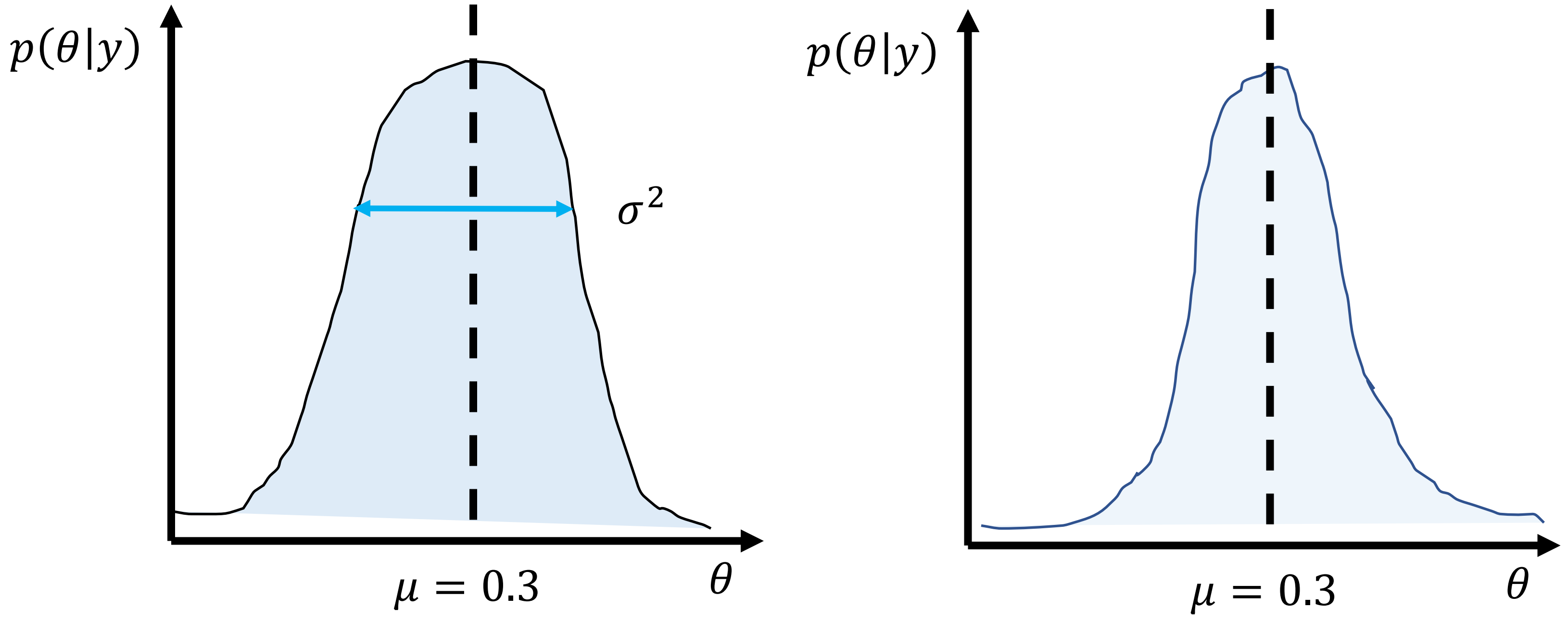




Example of continuous random variable

Here, we want to infer:

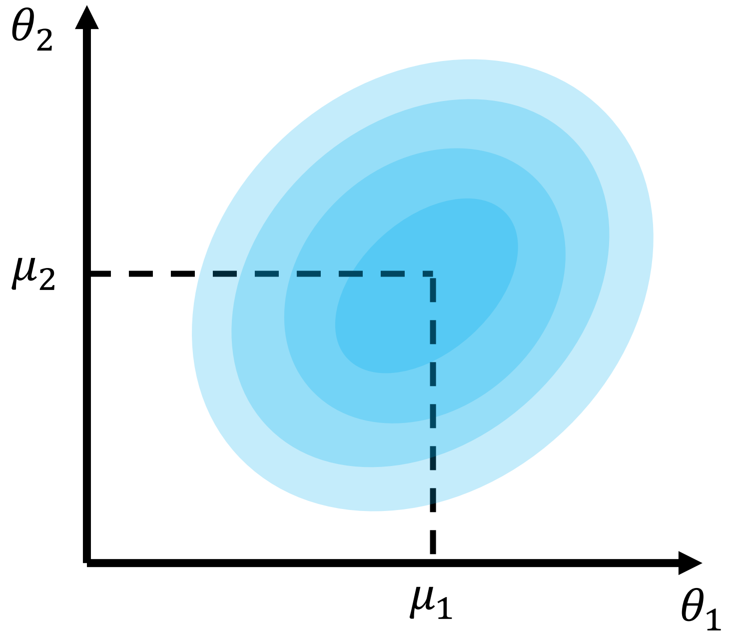
$$p(\theta|y) \sim \mathcal{N}(\mu, \sigma^2), \pi = \frac{1}{\sigma^2}$$

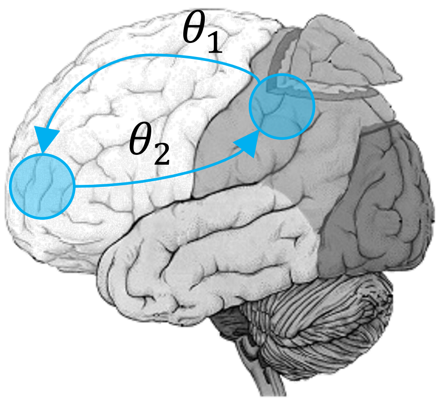


Example of continuous random variable

Here, we want to infer:

$$p(\bar{\theta}|y) \sim \mathcal{N}(\bar{\mu}, \Sigma), \quad \bar{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$





neural source: θ

Overview

Brief primer on ill-posed problems

Random variables and PDFs

Axioms of probability

Bayes rule and approximate inference

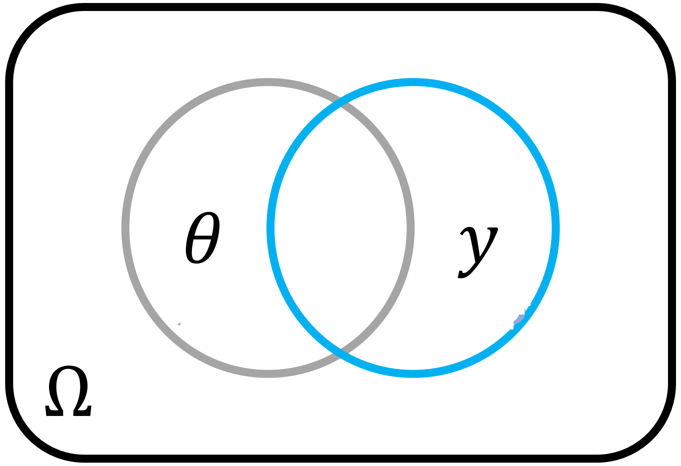
Demo: variational Bayes in SPM

Different kinds of probabilities

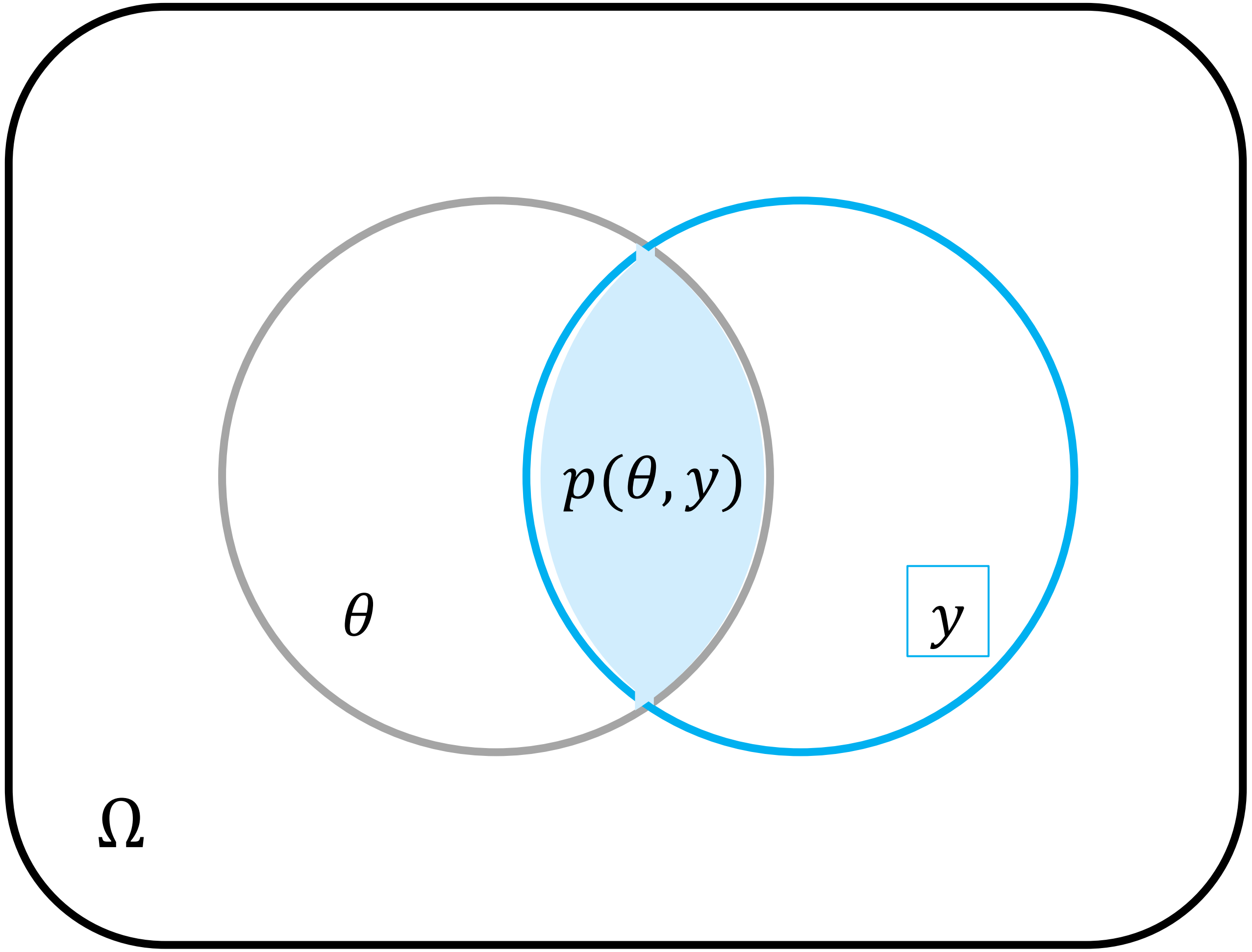
We assume a probability space Ω with subsets y and θ.

From this, we can define 3 kinds of probabilities:

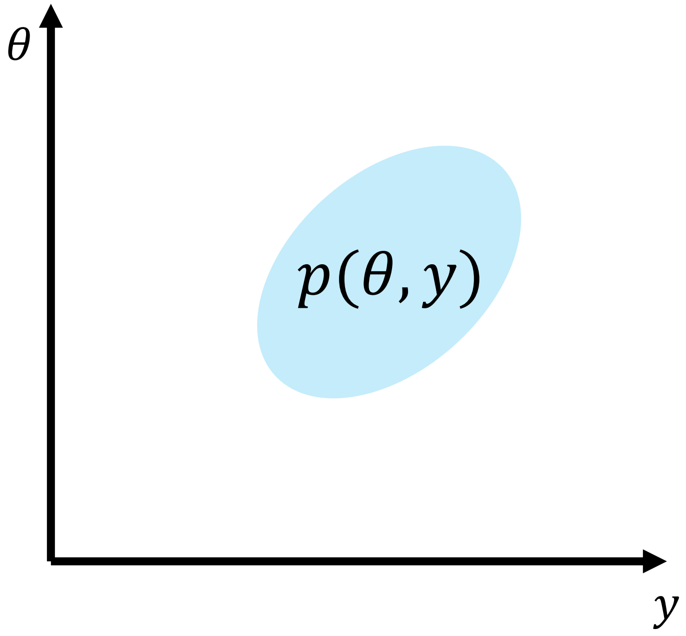
1. Joint probabilities e.g., $p(\theta, y)$



Joint probability



$$\int p(y, \theta)\, d\theta dy = 1$$

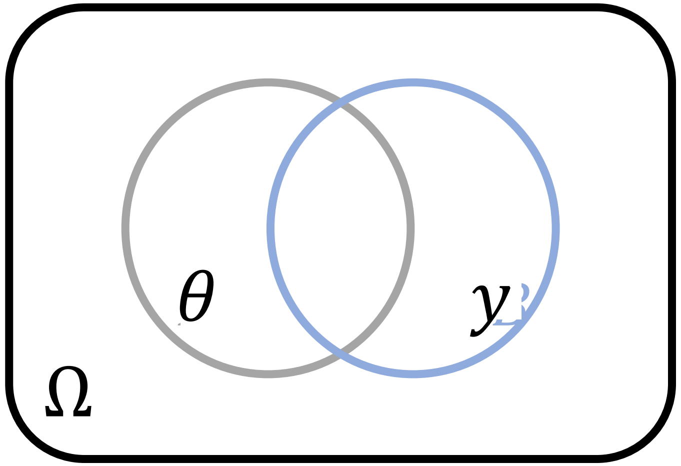


Different kinds of probabilities

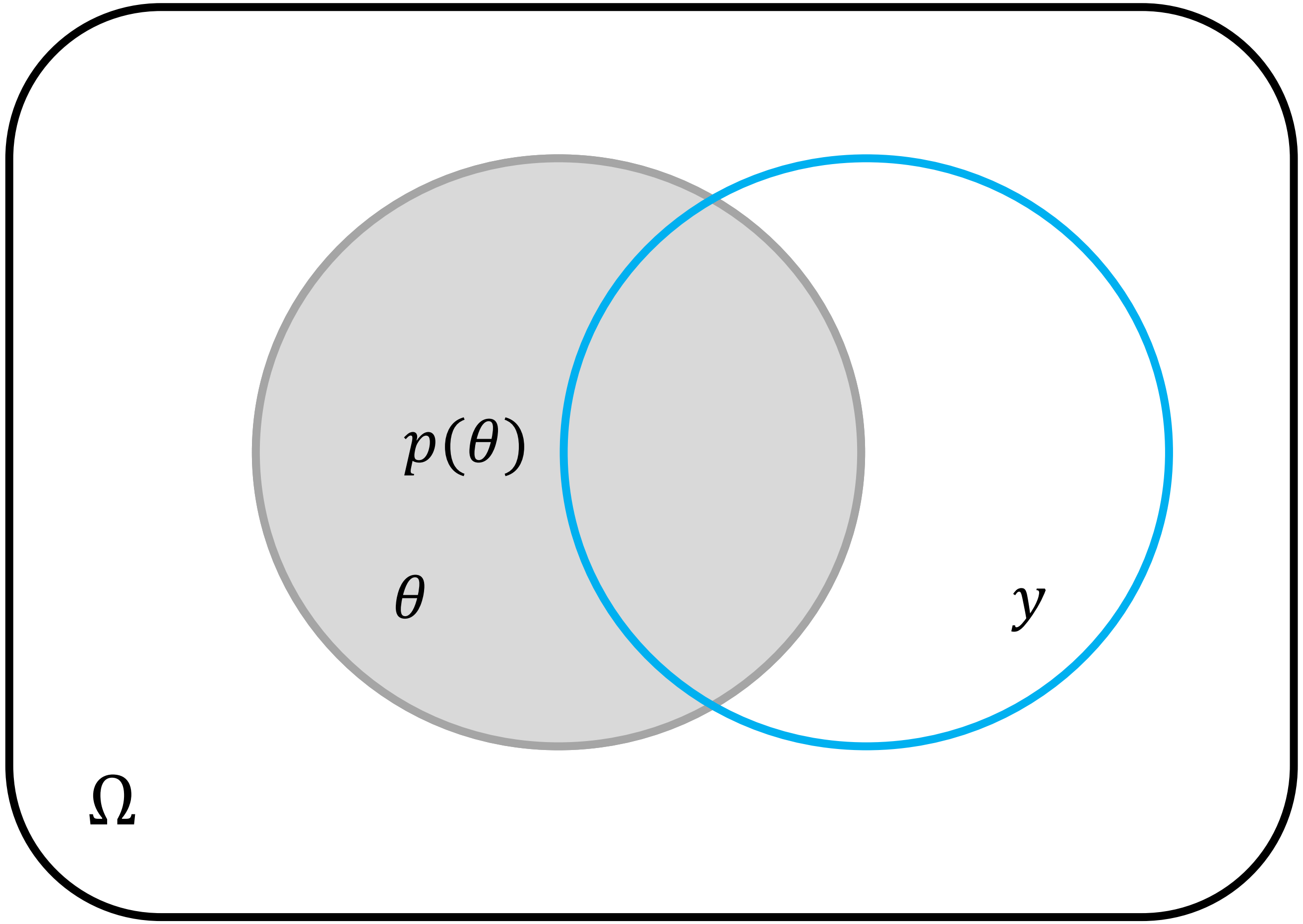
We assume a probability space Ω with subsets y and θ.

From this, we can define 3 kinds of probabilities:

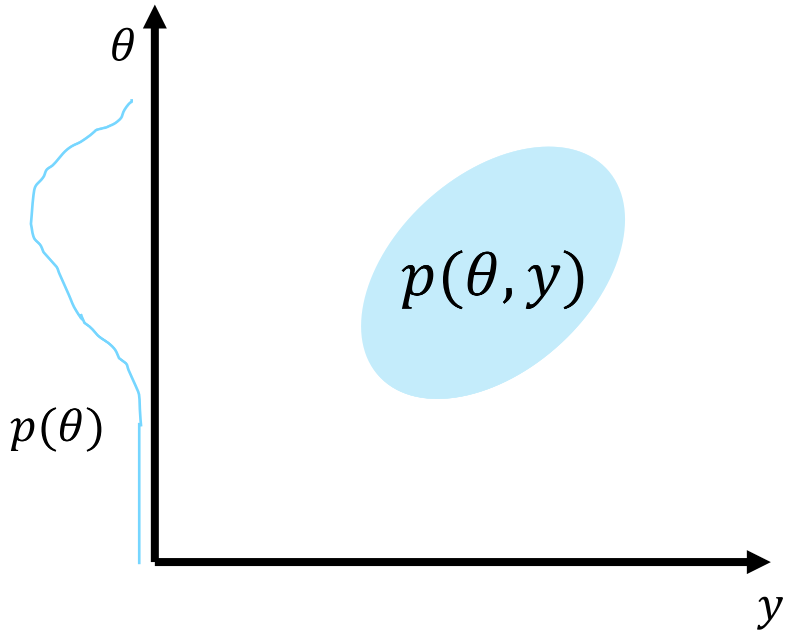
1. Joint probabilities e.g., $p(\theta, y)$
2. Marginal probabilities e.g., $p(\theta)$



Marginal probability



$$p(\theta) = \int p(y, \theta) dy$$



Example for discrete random variables

- Let A be the statement 'the sun is shining'
- Let B be the statement 'it is raining'
- $\bar{A}$ negates A, $\bar{B}$ negates B

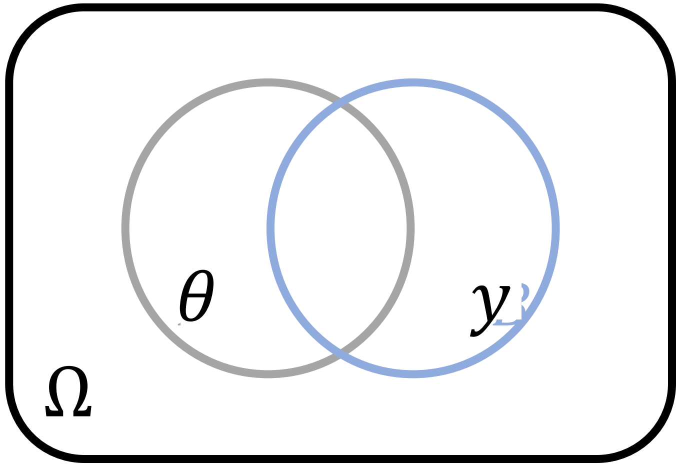
	B	$\bar{B}$	Marginal probabilities
A	0.1	0.5	
$\bar{A}$	0.2	0.2	
Marginal probabilities			Sum of all probabilities $\sum p(\cdot,\cdot) = 1$

Different kinds of probabilities

We assume a probability space Ω with subsets y and θ.

From this, we can define 3 kinds of probabilities:

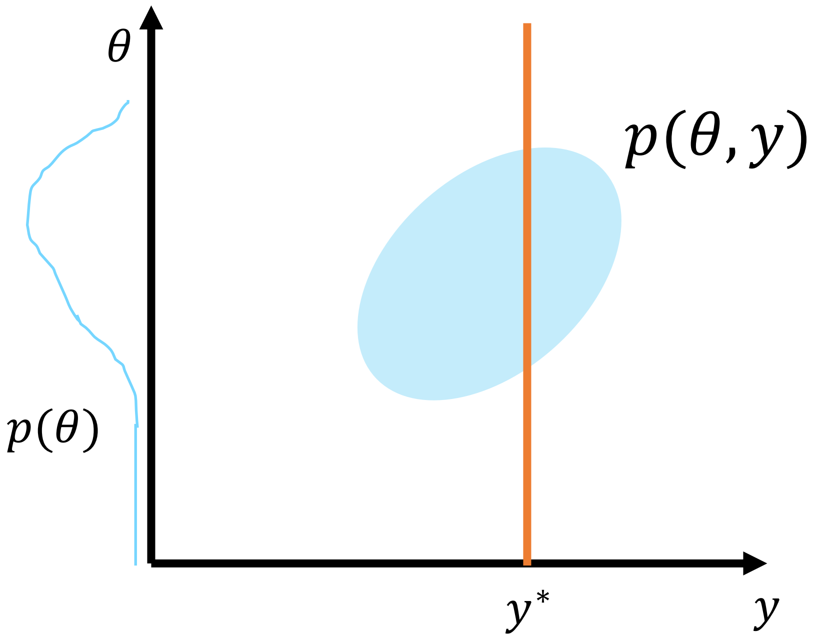
- Joint probabilities e.g., $p(\theta, y)$
- Marginal probabilities e.g., $p(y)$
- Conditional probabilities e.g., $p(\theta|y)$

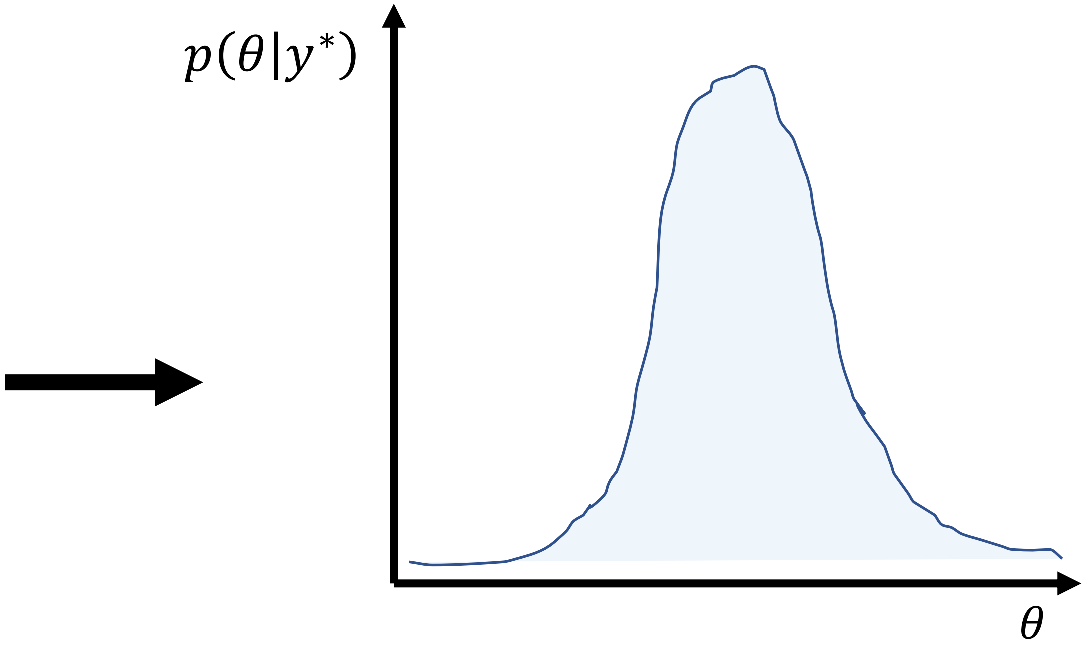


Conditional probability

$$p(\theta|y) = \int p(y,\theta)dy$$

$$p(\theta,y) = p(y|\theta)p(\theta) = p(\theta|y)p(y)$$





Example for discrete random variables

- Let A be the statement 'the sun is shining'
- Let B be the statement 'it is raining'
- $\bar{A}$ negates A, $\bar{B}$ negates B

What is the probability that the sun is shining given that it is not raining?

	B	$\bar{B}$	Marginal probabilities
A	0.1	0.5	0.6
$\bar{A}$	0.2	0.2	0.4
Marginal probabilities	0.3	0.7	Sum of all probabilities $\sum p(\cdot,\cdot) = 1$

Axioms of probability

1. $\int p(y, \theta)\, d\theta dy = 1$ *(Normalisation)*

2. $p(\theta) = \int p(y, \theta) dy$ *(Marginalisation – the sum rule)*

3. $p(\theta, y) = p(y|\theta)p(\theta)$

   $= p(\theta|y)p(y)$ *(Conditioning – the product rule)*

Overview

Brief primer on ill-posed problems

Random variables and PDFs

Axioms of probability

Bayes' rule and approximate inference

Demo: variational Bayes in SPM

Bayes' rule

Product rule states that:

$$p(\theta|y)p(y) = p(y|\theta)p(\theta)$$

Next, we rearrange:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

Apply the sum and product rules:

$$p(y) = \int p(y,\theta)d\theta = \int p(y|\theta)p(\theta)d\theta$$

Bayes' rule

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

Overview

Brief primer on ill-posed problems

Random variables and PDFs

Axioms of probability

Bayes rule and approximate inference

Demo: variational Bayes in SPM

Demo

General linear model formulation:

$$\underset{[20 \cdot 1]}{\bar{y}} = \underset{[20 \cdot 2]}{x}\, \underset{[2 \cdot 1]}{\bar{\beta}} + \underset{[20 \cdot 1]}{\bar{\epsilon}}$$

$$\begin{bmatrix} 1 & 18 \\ 1 & 32 \\ 1 & \dots \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

where, $\overline{\epsilon_i} \sim \mathcal{N}(0, \sigma^2)$

Demo

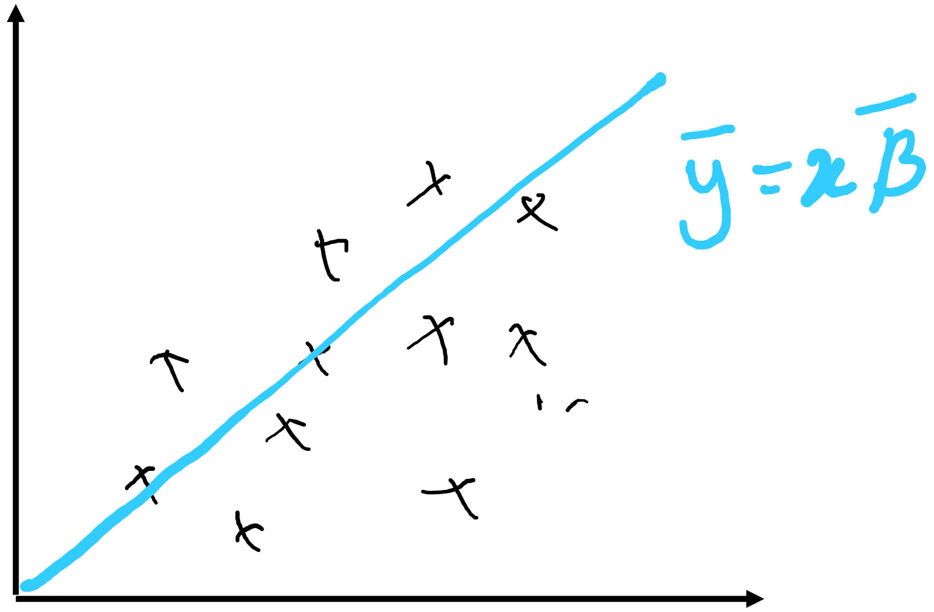
General linear model formulation:

$$\overset{[100\cdot 1]}{\bar{y}} = \overset{}{x\,\bar{\beta}} + \overset{[100\cdot 1]}{\bar{\epsilon}}$$

$$[100\cdot 2][2\cdot 1]$$

$$\begin{bmatrix} 1 & 18 \\ 1 & 32 \\ 1 & \dots \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

where, $\bar{\epsilon_i} \sim \mathcal{N}(0, \sigma^2)$



Demo

Likelihood:

$$p(y|\bar{\beta}) \sim \mathcal{N}(x\bar{\beta}, \Sigma) \qquad \Sigma = \begin{bmatrix} \sigma^2 & & \\ & \sigma^2 & \\ & & \dots \end{bmatrix}$$

Prediction is given by $x\bar{\beta}$; uncertainty can be thought of as following a normal distribution over that prediction.

Demo

Likelihood:

$$p(y|\bar{\beta}) \sim \mathcal{N}(x\bar{\beta}, \Sigma) \qquad \Sigma = \begin{bmatrix} \sigma^2 & & \\ & \sigma^2 & \\ & & \sigma^2 \end{bmatrix}$$

Prior:

$$p(\bar{\beta}) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

Prediction is given by $x\bar{\beta}$; uncertainty can be thought of as following a normal distribution over that prediction.

Demo

Posterior: $$p\left(\bar{\beta}\middle|y\right) = \frac{p\left(y\middle|\bar{\beta}\right)p(\bar{\beta})}{p(y)}$$

where, $$p(y) = \int p\left(y\middle|\bar{\beta}\right)p\left(\bar{\beta}\right)d\beta$$

Variational inference

- Calculate the evidence by marginalising out the parameters from the joint density:

$$p(y) = \int p(y|\bar{\beta})p(\bar{\beta})d\beta$$

- The evidence integral is **not** available in closed form + computing this requires variational inference.
- We introduce a variational density q that can be integrated: $q(\bar{\beta}) \approx p(\bar{\beta}|y)$
- We now make a move from $p(y) \rightarrow log\, p(y)$ to make the computations easier.

Deriving the free energy

Assumptions: $p(\bar{\beta}|y) \neq 0$ and $q(\bar{\beta}) \neq 0$

$$\log p(y) = \log p(y) + \int log \frac{p(\bar{\beta}|y)}{p(\bar{\beta}|y)} d\beta$$

$$= \int q(\bar{\beta}) \log p(y)\, d\beta + \int q(\bar{\beta}) log \frac{p(\bar{\beta}|y)}{p(\bar{\beta}|y)} d\beta$$

$$= \int q(\bar{\beta}) log \frac{p(\bar{\beta},y)}{p(\bar{\beta}|y)} d\beta$$

$$= \int q(\bar{\beta}) log \frac{p(\bar{\beta},y)}{q(\bar{\beta})} d\beta + \int q(\bar{\beta}) log \frac{q(\bar{\beta})}{p(\bar{\beta}|y)} d\beta$$

Free energy **KL-divergence**

divergence is 0 iff $q(\cdot) = p(\cdot\,|y)$

Demo

Posterior: $$p(\bar{\beta}|y) = \frac{p(y|\bar{\beta})p(\bar{\beta})}{p(y)}$$

where, $$p(y) = \int p(y|\bar{\beta})p(\bar{\beta})d\beta$$

$$p(\bar{\beta}|y) \sim \mathcal{N}(\mu, \Sigma) \text{ and } \log p(y) \approx F$$

Matlab demo

Overview

Brief primer on ill-posed problems

Random variables and PDFs

Axioms of probability

Bayes rule and approximate inference

Demo: variational Bayes in SPM

Probability theory is nothing but common sense reduced to calculation.
— Pierre-Simon Laplace, 1819