Imprecise Probability: General Ideas and Statistical Approaches

Gero Walter

Eindhoven University of Technology, Eindhoven, NL

g.m.walter@tue.nl

Technische Universiteit Eindhoven University of Technology

Thanks to Matthias Troffaes for introductory slides.

Hannover 2016-12-07

Requirements for an Uncertainty Model

Operational

How can uncertainty be reliably

- measured?
- communicated?

Requirements for an Uncertainty Model

Operational

How can uncertainty be reliably

- measured?
- communicated?

Inference

How can we use our uncertainty model for

- statistical reasoning?
- decision making?

Uncertainty via Probability

Definition

An event is a statement that may, or may not, hold
—typically, something that may happen in the future.

Notation: $A, B, C, \dots$

Examples

- tomorrow, it will rain
- in the next year, at most 3 components will fail

Uncertainty via Probability

Definition

An event is a statement that may, or may not, hold
—typically, something that may happen in the future.

Notation: A, B, C, $\ldots$

Examples

- tomorrow, it will rain
- in the next year, at most 3 components will fail

how to express our uncertainty regarding events?

Probability: Definition

Definition

The probability of an event is a number between 0 and 1.

Notation: $P(A)$, $P(B)$, $P(C)$, ...

Examples

- for A = 'tomorrow, it will rain'
  my probability $P(A)$ is 0.2
- for B = 'in the next year, at most 3 components will fail'
  my probability $P(B)$ is 0.0173

Probability: Definition

Definition

The probability of an event is a number between 0 and 1.

Notation: $P(A)$, $P(B)$, $P(C)$, $\ldots$

Examples

- for A = 'tomorrow, it will rain'
  my probability $P(A)$ is 0.2
- for B = 'in the next year, at most 3 components will fail'
  my probability $P(B)$ is 0.0173

what do these numbers actually mean?
how would you measure them?

Probability: Interpretation

Interpretation: Trivial Cases

$P(A) = 0 \iff A$ is practically impossible — logically?

$P(A) = 1 \iff A$ is practically certain

Interpretation: Trivial Cases

$P(A) = 0 \iff A$ is practically impossible logically?

$P(A) = 1 \iff A$ is practically certain

what about values between 0 and 1, such as $P(A) = 0.2$?

Probability: Interpretation

Interpretation: Trivial Cases

$P(A) = 0 \iff A$ is practically impossible　　　logically?

$P(A) = 1 \iff A$ is practically certain

what about values between 0 and 1, such as $P(A) = 0.2$?

Interpretation: General Case

- it's (like) a frequency
- it's a degree of belief (► betting rate)
- it's something else

Probability: Frequency Interpretation

$P(A) = 0.2$ means:

- in 1 out of 5 times, it rains tomorrow

Probability: Frequency Interpretation

$P(A) = 0.2$ means:

- in 1 out of 5 times, it rains tomorrow
  ??? (tomorrow is not repeatable!)

Probability: Frequency Interpretation

$P(A) = 0.2$ means:

- in 1 out of 5 times, it rains tomorrow
  ??? (tomorrow is not repeatable!)
- on a 'day like this', in 1 out of 5 times, it rains the next day

Probability: Frequency Interpretation

$P(A) = 0.2$ means:

- in 1 out of 5 times, it rains tomorrow
  ??? (tomorrow is not repeatable!)
- on a 'day like this', in 1 out of 5 times, it rains the next day

Frequency Interpretation

+ intuitive, easy to understand
– needs reference class, only for repeatable events
– needs plenty of data, or strong symmetry assumptions
! aleatory

Probability: Betting Interpretation

$P(A) = 0.2$ means:

- I would *now* pay at most €0.2
  if *tomorrow* I am paid €1 in case it rains

Probability: Betting Interpretation

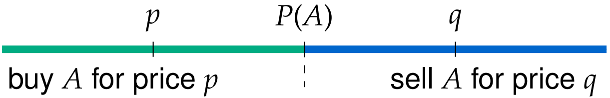
$P(A) = 0.2$ means:

- I would *now* pay at most €0.2
  if *tomorrow* I am paid €1 in case it rains
- I would *tomorrow* pay €1 in case it rains
  if I am *now* paid at least €0.2

Probability: Betting Interpretation

$P(A) = 0.2$ means:

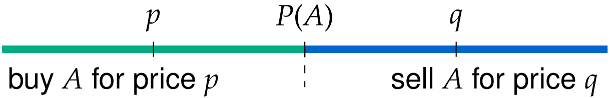
- I would *now* pay at most €0.2
  if *tomorrow* I am paid €1 in case it rains
- I would *tomorrow* pay €1 in case it rains
  if I am *now* paid at least €0.2



Probability: Betting Interpretation

$P(A) = 0.2$ means:

- I would *now* pay at most €0.2
  if *tomorrow* I am paid €1 in case it rains
- I would *tomorrow* pay €1 in case it rains
  if I am *now* paid at least €0.2



Betting Interpretation (degree of belief)

+ no reference class, works also for one-shot events

– needs plenty of elicitation or plenty of data

! epistemic

Dealing With Severe Uncertainty

in case of partial elicitation and/or sparse data
it may be hard to specify an exact probability
but you may still confidently bound your probability

this becomes more and more relevant
as problems become larger and larger

Bounding Methods

Confidence intervals (Frequentist Statistics)

- – choice of confidence level α?
- – p-value fallacy (Gigerenzer, Krauss, and Vitouch 2004)
  a.k.a. prosecutor's fallacy
- + no prior needed, only likelihood

Bounding Methods

Confidence intervals (Frequentist Statistics)

– choice of confidence level α?
– p-value fallacy (Gigerenzer, Krauss, and Vitouch 2004)
  a.k.a. prosecutor's fallacy
+ no prior needed, only likelihood

Credible intervals (Bayesian Statistics)

– choice of credible level α?
– choice of prior?
– dealing with prior ignorance and prior-data conflict?
+ no p-value fallacy

Bounding Methods

Confidence intervals (Frequentist Statistics)

– choice of confidence level α?
– p-value fallacy (Gigerenzer, Krauss, and Vitouch 2004)
  a.k.a. prosecutor's fallacy
+ no prior needed, only likelihood

Credible intervals (Bayesian Statistics)

– choice of credible level α?
– choice of prior?
– dealing with prior ignorance and prior-data conflict?
+ no p-value fallacy

Interval probability (bounding probabilities directly)

– choice of prior bounds?
+ no confidence / credible level issues
+ no prior ignorance issues
+ no p-value fallacy

Lower and Upper Probability: Definition

Definition

The lower and upper probability of an event
are numbers between 0 and 1.

Notation: $\underline{P}(A)$, $\overline{P}(A)$, ...

Examples

- for A = 'tomorrow, it will rain'
  my lower probability $\underline{P}(A)$ is 0.1
  my upper probability $\overline{P}(A)$ is 0.4

Lower and Upper Probability: Definition

Definition

The lower and upper probability of an event
are numbers between 0 and 1.

Notation: $\underline{P}(A)$, $\overline{P}(A)$, ...

Examples

- for A = 'tomorrow, it will rain'
  my lower probability $\underline{P}(A)$ is 0.1
  my upper probability $\overline{P}(A)$ is 0.4

what do these numbers actually mean?
how would you measure them?

$\underline{P}$ and $\overline{P}$: Betting Interpretation

$\underline{P}(A) = 0.1$ and $\overline{P}(A) = 0.4$ means:

- I would *now* pay at most €0.1
  if *tomorrow* I am paid €1 in case it rains

$\underline{P}$ and $\overline{P}$: Betting Interpretation

$\underline{P}(A) = 0.1$ and $\overline{P}(A) = 0.4$ means:

- I would *now* pay at most €0.1
  if *tomorrow* I am paid €1 in case it rains
- I would *tomorrow* pay €1 in case it rains
  if I am *now* paid at least €0.4

$\underline{P}$ and $\overline{P}$: Betting Interpretation

$\underline{P}(A) = 0.1$ and $\overline{P}(A) = 0.4$ means:

- I would *now* pay at most €0.1
  if *tomorrow* I am paid €1 in case it rains
- I would *tomorrow* pay €1 in case it rains
  if I am *now* paid at least €0.4

$\underline{P}$ and $\overline{P}$: Betting Interpretation

$\underline{P}(A) = 0.1$ and $\overline{P}(A) = 0.4$ means:

- I would *now* pay at most €0.1
  if *tomorrow* I am paid €1 in case it rains
- I would *tomorrow* pay €1 in case it rains
  if I am *now* paid at least €0.4

Betting Interpretation

- \+ no reference class, works also for one-shot events
- \+ works with partial elicitation and / or sparse data
- ! epistemic

Events: Formal Definition

Definition

The possibility space Ω is
the set of all possible outcomes of the problem at hand.

Example

interested in reliability of a system with 5 components
e.g. number of components that fail in the next year
$\Omega = \{0, 1, 2, 3, 4, 5\}$

Events: Formal Definition

Definition

The possibility space Ω is
the set of all possible outcomes of the problem at hand.

Example

interested in reliability of a system with 5 components
e.g. number of components that fail in the next year
$\Omega = \{0, 1, 2, 3, 4, 5\}$

Definition

An event is a subset of Ω. Notation: A, B, C, $\ldots$

Example

in the next year, at most 3 components will fail
would be represented by the event $A = \{0, 1, 2, 3\}$

$\underline{P}$ and $\overline{P}$: Formal Definition

Definition

A lower probability $\underline{P}$
maps *every* event $A \subseteq \Omega$ to a real number $\underline{P}(A)$.

The upper probability $\overline{P}$ is simply defined as
$\overline{P}(A) = 1 - \underline{P}(A^c)$, for all $A \subseteq \Omega$

$\underline{P}$ and $\overline{P}$: Formal Definition

Definition

A lower probability $\underline{P}$
maps *every* event $A \subseteq \Omega$ to a real number $\underline{P}(A)$.

The upper probability $\overline{P}$ is simply defined as
$\overline{P}(A) = 1 - \underline{P}(A^c)$, for all $A \subseteq \Omega$

- A^c = complement (or negation) of A = all elements *not* in A

$\underline{P}$ and $\overline{P}$: Formal Definition

Definition

A lower probability $\underline{P}$
maps *every* event $A \subseteq \Omega$ to a real number $\underline{P}(A)$.

The upper probability $\overline{P}$ is simply defined as
$\overline{P}(A) = 1 - \underline{P}(A^c)$, for all $A \subseteq \Omega$

- A^c = complement (or negation) of A = all elements *not* in A
- the identity $\overline{P}(A) = 1 - \underline{P}(A^c)$ is implied by the betting interpretation

$\underline{P}$ and $\overline{P}$: Formal Definition

Definition

A lower probability $\underline{P}$
maps *every* event $A \subseteq \Omega$ to a real number $\underline{P}(A)$.

The upper probability $\overline{P}$ is simply defined as
$\overline{P}(A) = 1 - \underline{P}(A^c)$, for all $A \subseteq \Omega$

- A^c = complement (or negation) of A = all elements *not* in A
- the identity $\overline{P}(A) = 1 - \underline{P}(A^c)$ is implied by the betting interpretation
- *every* event $\leftrightarrow$ sparse data?
  we can always set $\underline{P}(A) = 0$ and $\underline{P}(A^c) = 0$ ($\iff$ $\overline{P}(A) = 1$)!

$\underline{P}$ and $\overline{P}$: Formal Definition

Definition

A lower probability $\underline{P}$
maps *every* event $A \subseteq \Omega$ to a real number $\underline{P}(A)$.

The upper probability $\overline{P}$ is simply defined as
$\overline{P}(A) = 1 - \underline{P}(A^c)$, for all $A \subseteq \Omega$

- A^c = complement (or negation) of A = all elements *not* in A
- the identity $\overline{P}(A) = 1 - \underline{P}(A^c)$ is implied by the betting interpretation
- *every* event $\leftrightarrow$ sparse data?
  we can always set $\underline{P}(A) = 0$ and $\underline{P}(A^c) = 0$ ($\iff$ $\overline{P}(A) = 1$)!
- $\underline{P}$ specification for related events may allow to raise $\underline{P}(A)$ (correcting for consistency)

$\underline{P}$ and $\overline{P}$: Credal Sets

Definition

A probability measure P
maps *every* event $A \subseteq \Omega$ to a number $P(A)$ in $[0, 1]$ and satisfies

- $P(\emptyset) = 0$,
- $P(\Omega) = 1$, and
- $P(A) = \sum_{\omega \in A} P(\{\omega\})$.

$\underline{P}$ and $\overline{P}$: Credal Sets

Definition

A probability measure P
maps *every* event $A \subseteq \Omega$ to a number $P(A)$ in $[0, 1]$ and satisfies

- $P(\emptyset) = 0$,
- $P(\Omega) = 1$, and
- $P(A) = \sum_{\omega \in A} P(\{\omega\})$.

Definition

The credal set $\mathcal{M}$ of $\underline{P}$
is the set of all probability measures P for which
$\underline{P}(A) \leq P(A) \leq \overline{P}(A)$ for all $A \subseteq \Omega$.

$\underline{P}$ and $\overline{P}$: Credal Sets

Definition

A probability measure P
maps *every* event $A \subseteq \Omega$ to a number $P(A)$ in $[0, 1]$ and satisfies

- $P(\emptyset) = 0$,
- $P(\Omega) = 1$, and
- $P(A) = \sum_{\omega \in A} P(\{\omega\})$.

Definition

The credal set $\mathcal{M}$ of $\underline{P}$
is the set of all probability measures P for which
$\underline{P}(A) \leq P(A) \leq \overline{P}(A)$ for all $A \subseteq \Omega$.

Sensitivity Interpretation of $\underline{P}$ and $\overline{P}$

One of the probability measures P in the credal set $\mathcal{M}$
is the correct one, *but we do not know which*.

$\underline{P}$ and $\overline{P}$: Credal Sets

Definition

A probability measure P
maps *every* event $A \subseteq \Omega$ to a number $P(A)$ in $[0, 1]$ and satisfies

- $P(\emptyset) = 0$,
- $P(\Omega) = 1$, and
- $P(A) = \sum_{\omega \in A} P(\{\omega\})$.

Definition

The credal set $\mathcal{M}$ of $\underline{P}$
is the set of all probability measures P for which
$\underline{P}(A) \leq P(A) \leq \overline{P}(A)$ for all $A \subseteq \Omega$.

Sensitivity Interpretation of $\underline{P}$ and $\overline{P}$

One of the probability measures P in the credal set $\mathcal{M}$
is the correct one, *but we do not know which*.

crucial: no distribution over $\mathcal{M}$ assumed!

Sensitivity Interpretation for Credal Sets

Uncertainty about probability statements

smaller credal set = more precise probability statements

Lottery A

Number of winning tickets:
exactly known as 5 out of 100

► $P(\text{win}) = 5/100$

Lottery B

Number of winning tickets:
not exactly known, supposedly
between 1 and 7 out of 100

► $P(\text{win}) = [1/100, 7/100]$

Uncertainty about probability statements

smaller credal set = more precise probability statements

Lottery A

Number of winning tickets:
exactly known as 5 out of 100

▸ $P(\text{win}) = 5/100$

Lottery B

Number of winning tickets:
not exactly known, supposedly
between 1 and 7 out of 100

▸ $P(\text{win}) = [1/100, 7/100]$

- Separate uncertainty *within the model* (probability statements) from uncertainty *about the model* (how certain about statements)

Uncertainty about probability statements

smaller credal set = more precise probability statements

Lottery A
Number of winning tickets:
exactly known as 5 out of 100
▶ $P(\text{win}) = 5/100$

Lottery B
Number of winning tickets:
not exactly known, supposedly
between 1 and 7 out of 100
▶ $P(\text{win}) = [1/100, 7/100]$

- Separate uncertainty *within the model* (probability statements) from uncertainty *about the model* (how certain about statements)
- Systematic sensitivity analysis, robust Bayesian approach

Bayesian Inference

expert info + data → complete picture

Bayesian Inference

expert info	+	data	→	complete picture
prior distribution	+	sample distribution	→	posterior distribution
$f(p)$	$\times$	$f(s \mid p)$	$\propto$	$f(p \mid s)$

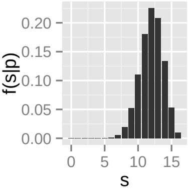
► Bayes' Rule

expert info	+	data	→	complete picture
prior distribution	+	sample distribution	→	posterior distribution
$f(p)$	$\times$	$f(s \mid p)$	$\propto$	$f(p \mid s)$

► Bayes' Rule

Binomial
distribution

$$s \mid p \sim \text{Binomial}(n, p)$$

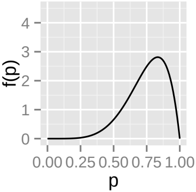


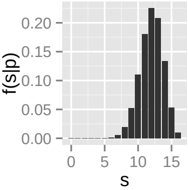
Bayesian Inference

expert info	+	data	→	complete picture
prior distribution	+	sample distribution	→	posterior distribution
$f(p)$	$\times$	$f(s \mid p)$	$\propto$	$f(p \mid s)$

▶ Bayes' Rule

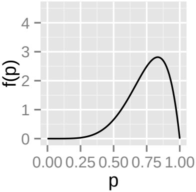
Beta prior	Binomial distribution
$p \sim \text{Beta}(\alpha^{(0)}, \beta^{(0)})$	$s \mid p \sim \text{Binomial}(n, p)$

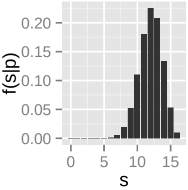


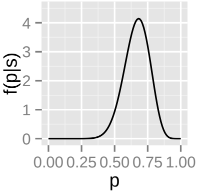


Bayesian Inference

expert info	+	data	→	complete picture
prior distribution	+	sample distribution	→	posterior distribution
$f(p)$	$\times$	$f(s \mid p)$	$\propto$	$f(p \mid s)$
				► Bayes' Rule
Beta prior		Binomial distribution		Beta posterior
				► conjugacy
$p \sim \text{Beta}(\alpha^{(0)}, \beta^{(0)})$		$s \mid p \sim \text{Binomial}(n, p)$		$p \mid s \sim \text{Beta}(\alpha^{(n)}, \beta^{(n)})$







Bayesian inference with sets of priors

set of priors $\mathcal{M}^{(0)} \rightarrow$ sets of posteriors $\mathcal{M}^{(n)}$
by updating element by element

Bayesian Inference with Credal Sets

Bayesian inference with sets of priors

set of priors $\mathcal{M}^{(0)} \rightarrow$ sets of posteriors $\mathcal{M}^{(n)}$
by updating element by element

the Generalized Bayes Rule (GBR, Walley 1991)
ensures *coherence* (a consistency property)

Bayesian Inference with Credal Sets

Bayesian inference with sets of priors

set of priors $\mathcal{M}^{(0)} \to$ sets of posteriors $\mathcal{M}^{(n)}$
by updating element by element

the Generalized Bayes Rule (GBR, Walley 1991)
ensures *coherence* (a consistency property)

Bounds for inferences (point estimate, . . .) by min / max over $\mathcal{M}^{(n)}$

Bayesian Inference with Credal Sets

Bayesian inference with sets of priors

set of priors $\mathcal{M}^{(0)} \rightarrow$ sets of posteriors $\mathcal{M}^{(n)}$
by updating element by element

the Generalized Bayes Rule (GBR, Walley 1991)
ensures *coherence* (a consistency property)

Bounds for inferences (point estimate, ...) by min / max over $\mathcal{M}^{(n)}$

this offers advantages over usual Bayesian inference in case of

- prior ignorance / weakly informative priors

Bayesian Inference with Credal Sets

Bayesian inference with sets of priors

set of priors $\mathcal{M}^{(0)} \to$ sets of posteriors $\mathcal{M}^{(n)}$
by updating element by element

the Generalized Bayes Rule (GBR, Walley 1991)
ensures *coherence* (a consistency property)

Bounds for inferences (point estimate, ...) by min / max over $\mathcal{M}^{(n)}$

this offers advantages over usual Bayesian inference in case of

- prior ignorance / weakly informative priors
- prior-data conflict

‘Non-informative’ Priors

How to construct a prior if we do not have a lot of information?

‘Non-informative’ Priors

How to construct a prior if we do not have a lot of information?

Laplace: Principle of Indifference

Use the uniform distribution.

‘Non-informative’ Priors

How to construct a prior if we do not have a lot of information?

Laplace: Principle of Indifference

Use the uniform distribution.

Obvious issue: this depends on the parametrisation!

‘Non-informative’ Priors

How to construct a prior if we do not have a lot of information?

Laplace: Principle of Indifference

Use the uniform distribution.

Obvious issue: this depends on the parametrisation!

Example

An object of 1kg has uncertain volume V between 1ℓ and 2ℓ.

‘Non-informative’ Priors

How to construct a prior if we do not have a lot of information?

Laplace: Principle of Indifference

Use the uniform distribution.

Obvious issue: this depends on the parametrisation!

Example

An object of 1kg has uncertain volume V between 1ℓ and 2ℓ.

- Uniform distribution over volume $V \implies E(V) = 1.5\ell$.

‘Non-informative’ Priors

How to construct a prior if we do not have a lot of information?

Laplace: Principle of Indifference

Use the uniform distribution.

Obvious issue: this depends on the parametrisation!

Example

An object of 1kg has uncertain volume V between 1ℓ and 2ℓ.

- Uniform distribution over volume $V \implies E(V) = 1.5\ell$.
- Uniform distribution over density $\rho = 1/V \implies$
  $E(V) = E(1/\rho) = \int_{0.5}^{1} 2/\rho \, \mathrm{d}\rho = 2(\ln 1 - \ln 0.5) = 1.39\ell$

‘Non-informative’ Priors

How to construct a prior if we do not have a lot of information?

Laplace: Principle of Indifference

Use the uniform distribution.

Obvious issue: this depends on the parametrisation!

Example

An object of 1kg has uncertain volume V between 1ℓ and 2ℓ.

- Uniform distribution over volume $V \implies E(V) = 1.5\ell$.
- Uniform distribution over density $\rho = 1/V \implies$ $E(V) = E(1/\rho) = \int_{0.5}^{1} 2/\rho \, \mathrm{d}\rho = 2(\ln 1 - \ln 0.5) = 1.39\ell$

The uniform distribution does not really model prior ignorance!

‘Non-informative’ Priors

How to construct a prior if we do not have a lot of information?

Laplace: Principle of Indifference

Use the uniform distribution.

Obvious issue: this depends on the parametrisation!

Example

An object of 1kg has uncertain volume V between 1ℓ and 2ℓ.

- Uniform distribution over volume $V \implies E(V) = 1.5\ell$.
- Uniform distribution over density $\rho = 1/V \implies$
  $E(V) = E(1/\rho) = \int_{0.5}^{1} 2/\rho \, \mathrm{d}\rho = 2(\ln 1 - \ln 0.5) = 1.39\ell$

The uniform distribution does not really model prior ignorance!
(Jeffreys prior is transformation-invariant,
but depends on the sample space and can break decision making!)

Prior Ignorance via Sets of Probabilities

How to construct a prior if we do not have a lot of information?

Prior Ignorance via Sets of Probabilities

How to construct a prior if we do not have a lot of information?

Boole: Probability Bounding

Use the set of all probability distributions (vacuous model).

Results no longer depend on parametrisation!

Prior Ignorance via Sets of Probabilities

How to construct a prior if we do not have a lot of information?

Boole: Probability Bounding

Use the set of all probability distributions (vacuous model).

Results no longer depend on parametrisation!

Example

An object of 1kg has uncertain volume V between 1ℓ and 2ℓ.

Prior Ignorance via Sets of Probabilities

How to construct a prior if we do not have a lot of information?

Boole: Probability Bounding

Use the set of all probability distributions (vacuous model).

Results no longer depend on parametrisation!

Example

An object of 1kg has uncertain volume V between 1ℓ and 2ℓ.

- Set of all distributions over volume $V \implies E(V) \in [1, 2]$.

Prior Ignorance via Sets of Probabilities

How to construct a prior if we do not have a lot of information?

Boole: Probability Bounding

Use the set of all probability distributions (vacuous model).

Results no longer depend on parametrisation!

Example

An object of 1kg has uncertain volume V between 1ℓ and 2ℓ.

- Set of all distributions over volume $V \implies E(V) \in [1, 2]$.
- Set of all distributions over density $\rho = 1/V$
  $$\implies E(V) = E(1/\rho) \in [1, 2]$$

Prior Ignorance via Sets of Probabilities

Theorem

The set of posterior distributions resulting from a vacuous set of prior distributions is again vacuous, regardless of the likelihood.

We can never learn anything when starting from a vacuous set of priors!

Prior Ignorance via Sets of Probabilities

Theorem

The set of posterior distributions resulting from a vacuous set of prior distributions is again vacuous, regardless of the likelihood.

We can never learn anything when starting from a vacuous set of priors!

Solution: Near-Vacuous Sets of Priors

Only insist that the prior predictive, or other classes of inferences, are vacuous.

This can be done using sets of conjugate priors (Walley 1996; Benavoli and Zaffalon 2012; 2015).

Example: Imprecise Beta Model (IBM)

- Bernoulli observations: 0/1 observations (failure/success)

Example: Imprecise Beta Model (IBM)

- Bernoulli observations: 0/1 observations (failure/success)
- given: s successes in n i.i.d. trials and strong prior information

Example: Imprecise Beta Model (IBM)

- Bernoulli observations: 0/1 observations (failure/success)
- given: s successes in n i.i.d. trials and strong prior information
- we are, e.g., interested in probability for success in next trial

Example: Imprecise Beta Model (IBM)

- Bernoulli observations: 0/1 observations (failure/success)
- given: s successes in n i.i.d. trials and strong prior information
- we are, e.g., interested in probability for success in next trial

Beta-Binomial Model

data :	$s \mid p$	$\sim$	$\text{Binomial}(n, p)$
conjugate prior:	$p \mid \alpha^{(0)}, \beta^{(0)}$	$\sim$	$\text{Beta}(\alpha^{(0)}, \beta^{(0)})$
posterior:	$p \mid \alpha^{(n)}, \beta^{(n)}$	$\sim$	$\text{Beta}(\alpha^{(n)}, \beta^{(n)})$

Example: Imprecise Beta Model (IBM)

- Bernoulli observations: 0/1 observations (failure/success)
- given: s successes in n i.i.d. trials and strong prior information
- we are, e.g., interested in probability for success in next trial

Beta-Binomial Model

data :	$s \mid p$	$\sim$	$\text{Binomial}(n, p)$	
conjugate prior:	$p \mid \alpha^{(0)}, \beta^{(0)}$	$\sim$	~~$\text{Beta}(\alpha^{(0)}, \beta^{(0)})$~~	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$p \mid \alpha^{(n)}, \beta^{(n)}$	$\sim$	~~$\text{Beta}(\alpha^{(n)}, \beta^{(n)})$~~	$\text{Beta}(n^{(n)}, y^{(n)})$

Example: Imprecise Beta Model (IBM)

- Bernoulli observations: 0/1 observations (failure/success)
- given: s successes in n i.i.d. trials and strong prior information
- we are, e.g., interested in probability for success in next trial

Beta-Binomial Model

data :	$s \mid p$	$\sim$	$\text{Binomial}(n, p)$	
conjugate prior:	$p \mid \alpha^{(0)}, \beta^{(0)}$	$\sim$	~~$\text{Beta}(\alpha^{(0)}, \beta^{(0)})$~~	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$p \mid \alpha^{(n)}, \beta^{(n)}$	$\sim$	~~$\text{Beta}(\alpha^{(n)}, \beta^{(n)})$~~	$\text{Beta}(n^{(n)}, y^{(n)})$

Vary hyperparameters $(n^{(0)}, y^{(0)})$ in a set ▸ set of priors $\mathcal{M}^{(0)}$

Example: Imprecise Beta Model (IBM)

- Bernoulli observations: 0/1 observations (failure/success)
- given: s successes in n i.i.d. trials and strong prior information
- we are, e.g., interested in probability for success in next trial

Beta-Binomial Model

data :	$s \mid p$	$\sim$	$\text{Binomial}(n, p)$	
conjugate prior:	$p \mid \alpha^{(0)}, \beta^{(0)}$	$\sim$	~~$\text{Beta}(\alpha^{(0)}, \beta^{(0)})$~~	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$p \mid \alpha^{(n)}, \beta^{(n)}$	$\sim$	~~$\text{Beta}(\alpha^{(n)}, \beta^{(n)})$~~	$\text{Beta}(n^{(n)}, y^{(n)})$

Vary hyperparameters $(n^{(0)}, y^{(0)})$ in a set ► set of priors $\mathcal{M}^{(0)}$

Set of posteriors $\mathcal{M}^{(n)}$ via $= \left\{(n^{(n)}, y^{(n)}) : (n^{(0)}, y^{(0)}) \in \quad \right\}$

Bounds for inferences (point estimate, …) by min/max over .

Reparametrisation of the Beta Distribution

- reparametrisation helps to understand the parameter update:

$$n^{(0)} = \alpha^{(0)} + \beta^{(0)}, \qquad y^{(0)} = \frac{\alpha^{(0)}}{\alpha^{(0)} + \beta^{(0)}}, \quad \text{which are updated as}$$

$$n^{(n)} = n^{(0)} + n, \qquad y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}$$

Reparametrisation of the Beta Distribution

- reparametrisation helps to understand the parameter update:

$$n^{(0)} = \alpha^{(0)} + \beta^{(0)}\,, \qquad y^{(0)} = \frac{\alpha^{(0)}}{\alpha^{(0)} + \beta^{(0)}}\,, \qquad \text{which are updated as}$$

$$n^{(n)} = n^{(0)} + n\,, \qquad y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n}\, y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}$$

$$y^{(0)} = \mathrm{E}[p]$$

Reparametrisation of the Beta Distribution

- reparametrisation helps to understand the parameter update:

$$n^{(0)} = \alpha^{(0)} + \beta^{(0)}\,, \qquad y^{(0)} = \frac{\alpha^{(0)}}{\alpha^{(0)} + \beta^{(0)}}\,, \qquad \text{which are updated as}$$

$$n^{(n)} = n^{(0)} + n\,, \qquad y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n}\, y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}$$

$$y^{(0)} = \mathrm{E}[p] \qquad y^{(n)} = \mathrm{E}[p \mid s]$$

Reparametrisation of the Beta Distribution

- reparametrisation helps to understand the parameter update:

$$n^{(0)} = \alpha^{(0)} + \beta^{(0)}, \qquad y^{(0)} = \frac{\alpha^{(0)}}{\alpha^{(0)} + \beta^{(0)}},$$

which are updated as

$$n^{(n)} = n^{(0)} + n, \qquad y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}$$

$y^{(0)} = \mathrm{E}[p]$

$y^{(n)} = \mathrm{E}[p \mid s]$

ML estimator $\hat{p}$

Reparametrisation of the Beta Distribution

- reparametrisation helps to understand the parameter update:

$$n^{(0)} = \alpha^{(0)} + \beta^{(0)}, \qquad y^{(0)} = \frac{\alpha^{(0)}}{\alpha^{(0)} + \beta^{(0)}},$$

which are updated as

$$n^{(n)} = n^{(0)} + n, \qquad y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}$$

$n^{(0)}$ = pseudocounts

$y^{(0)} = \mathrm{E}[p]$

$y^{(n)} = \mathrm{E}[p \mid s]$

ML estimator $\hat{p}$

Reparametrisation of the Beta Distribution

- reparametrisation helps to understand the parameter update:

$$n^{(0)} = \alpha^{(0)} + \beta^{(0)}\,, \qquad y^{(0)} = \frac{\alpha^{(0)}}{\alpha^{(0)} + \beta^{(0)}}\,, \qquad \text{which are updated as}$$

$$n^{(n)} = n^{(0)} + n\,, \qquad y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n}\, y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}$$

$n^{(0)}$ = pseudocounts

$y^{(0)} = \mathrm{E}[p]$

$y^{(n)} = \mathrm{E}[p \mid s]$

ML estimator $\hat{p}$

$\mathrm{E}[p \mid s] = y^{(n)}$ is a weighted average of $\mathrm{E}[p]$ and $\hat{p}$!

Reparametrisation of the Beta Distribution

- reparametrisation helps to understand the parameter update:

$$n^{(0)} = \alpha^{(0)} + \beta^{(0)}, \qquad y^{(0)} = \frac{\alpha^{(0)}}{\alpha^{(0)} + \beta^{(0)}},$$ which are updated as

$$n^{(n)} = n^{(0)} + n, \qquad y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}$$

$n^{(0)}$ = pseudocounts

$y^{(0)} = \mathrm{E}[p]$

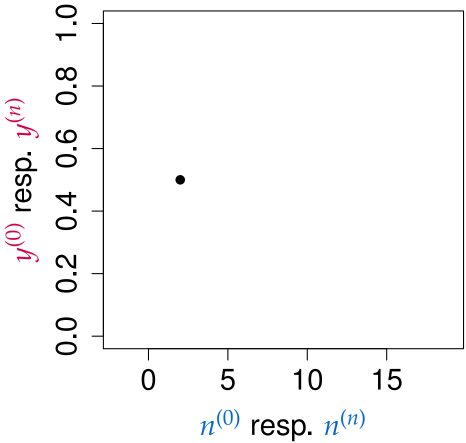
$y^{(n)} = \mathrm{E}[p \mid s]$

ML estimator $\hat{p}$

$\mathrm{E}[p \mid s] = y^{(n)}$ is a weighted average of $\mathrm{E}[p]$ and $\hat{p}$!

$\mathrm{Var}[p \mid s] = \frac{y^{(n)}(1 - y^{(n)})}{n^{(n)} + 1}$ decreases with n!

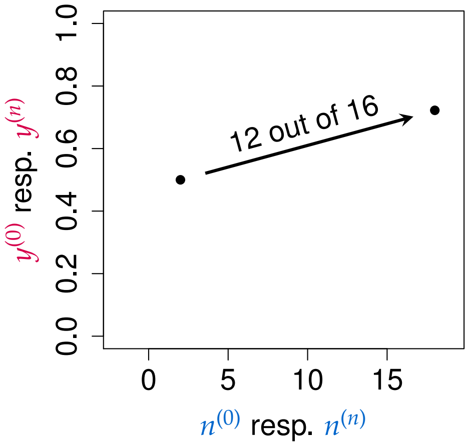
Beta-Binomial Model (BBM)



single prior (uniform)

prior $n^{(0)} = 2$, $y^{(0)} = 0.5$

data $s/n = 12/16 = 0.75$



single prior (uniform)

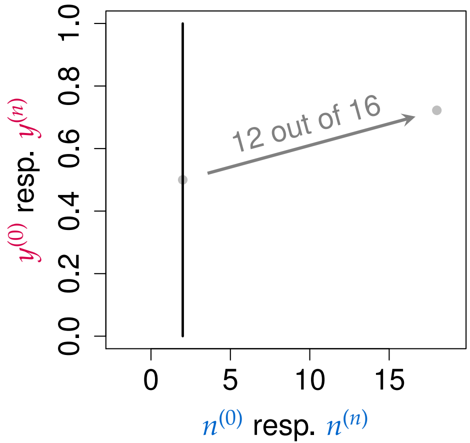
prior $n^{(0)} = 2$, $y^{(0)} = 0.5$

data $s/n = 12/16 = 0.75$

▼

$n^{(n)} = 18$, $y^{(n)} = 0.72$

Beta-Binomial Model (BBM)



single prior (uniform)

prior $n^{(0)} = 2$, $y^{(0)} = 0.5$
data $s/n = 12/16 = 0.75$

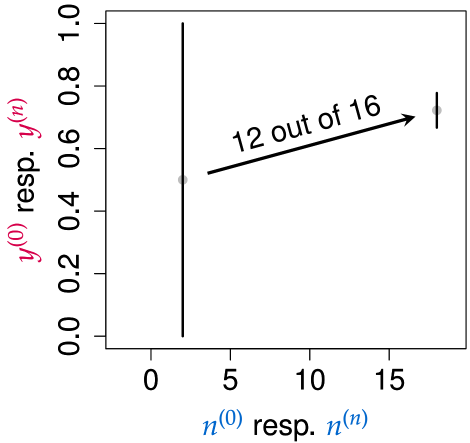
▼

$n^{(n)} = 18$, $y^{(n)} = 0.72$

near-vacuous set of priors

prior $n^{(0)} = 2$, $y^{(0)} \in (0, 1)$
data $s/n = 12/16 = 0.75$

Beta-Binomial Model (BBM)



single prior (uniform)

prior $n^{(0)} = 2$, $y^{(0)} = 0.5$
data $s/n = 12/16 = 0.75$

▼

$n^{(n)} = 18$, $y^{(n)} = 0.72$

near-vacuous set of priors

prior $n^{(0)} = 2$, $y^{(0)} \in (0, 1)$
data $s/n = 12/16 = 0.75$

▼

$n^{(n)} = 18$, $y^{(n)} \in (0.67, 0.77)$

Prior-Data Conflict

What if expert information and data tell different stories?

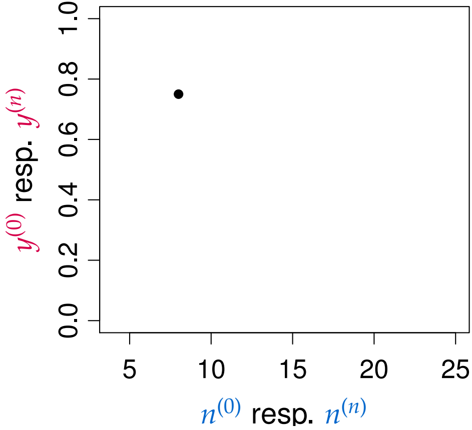
Prior-Data Conflict

What if expert information and data tell different stories?

Prior-Data Conflict

- *informative prior beliefs* and *trusted data* (sampling model correct, no outliers, etc.) are in conflict
- "[...] the prior [places] its mass primarily on distributions in the sampling model for which the observed data is surprising" (Evans and Moshonov 2006)
- there are not enough data to overrule the prior

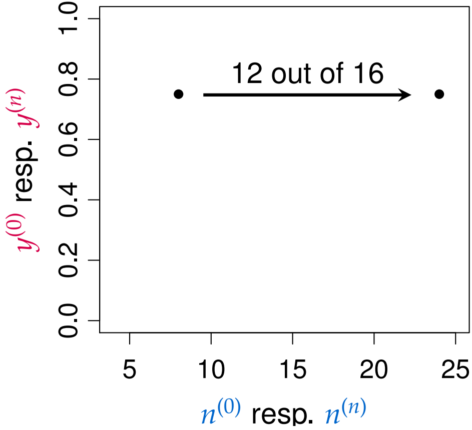
Beta-Binomial Model (BBM)



no conflict:

prior $n^{(0)} = 8$, $y^{(0)} = 0.75$
data $s/n = 12/16 = 0.75$

Beta-Binomial Model (BBM)

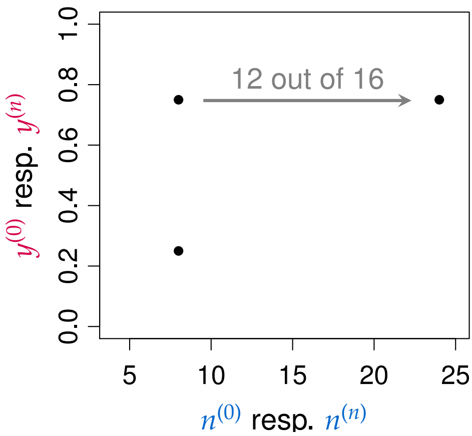


no conflict:

prior $n^{(0)} = 8$, $y^{(0)} = 0.75$
data $s/n = 12/16 = 0.75$

$n^{(n)} = 24$, $y^{(n)} = 0.75$

Beta-Binomial Model (BBM)



no conflict:

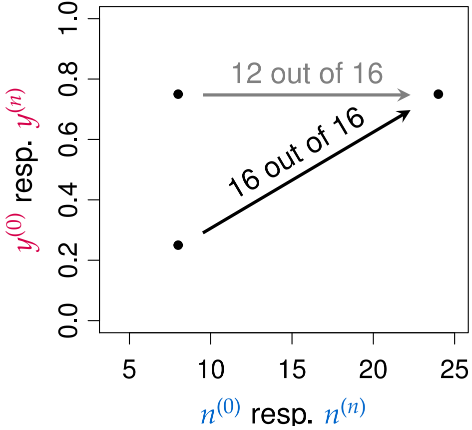
prior $n^{(0)} = 8$, $y^{(0)} = 0.75$
data $s/n = 12/16 = 0.75$

$n^{(n)} = 24$, $y^{(n)} = 0.75$

prior-data conflict:

prior $n^{(0)} = 8$, $y^{(0)} = 0.25$
data $s/n = 16/16 = 1$

Beta-Binomial Model (BBM)



no conflict:

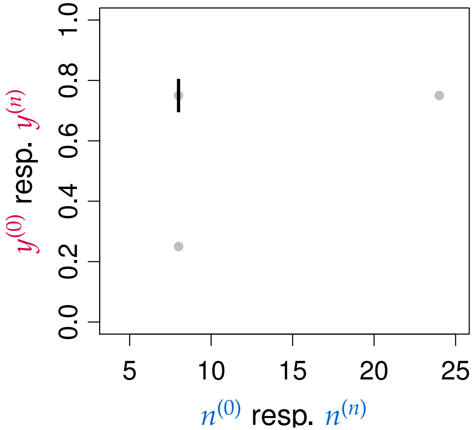
prior $n^{(0)} = 8$, $y^{(0)} = 0.75$
data $s/n = 12/16 = 0.75$

$n^{(n)} = 24$, $y^{(n)} = 0.75$

prior-data conflict:

prior $n^{(0)} = 8$, $y^{(0)} = 0.25$
data $s/n = 16/16 = 1$

IDM (Walley 1996); Quaeghebeur and de Cooman (2005)

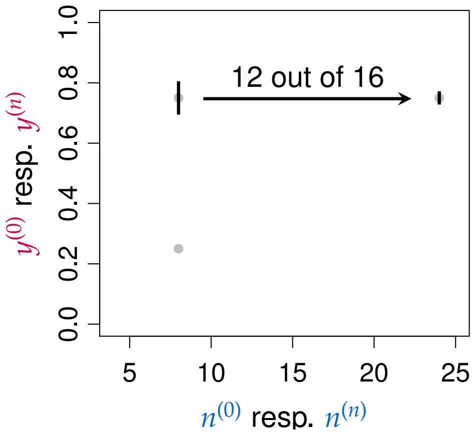


no conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

Imprecise BBM

IDM (Walley 1996); Quaeghebeur and de Cooman (2005)



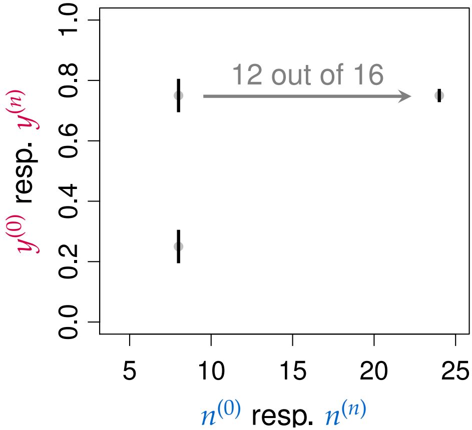
no conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

$n^{(n)} = 24$, $y^{(n)} \in [0.73, 0.77]$

Imprecise BBM

IDM (Walley 1996); Quaeghebeur and de Cooman (2005)



no conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

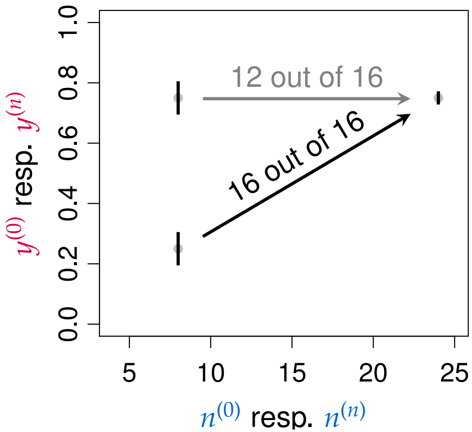
$n^{(n)} = 24$, $y^{(n)} \in [0.73, 0.77]$

prior data conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.2, 0.3]$
data $s/n = 16/16 = 1$

Imprecise BBM

IDM (Walley 1996); Quaeghebeur and de Cooman (2005)



no conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

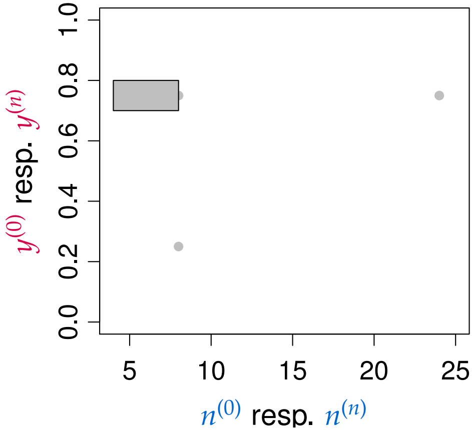
$n^{(n)} = 24$, $y^{(n)} \in [0.73, 0.77]$

prior data conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.2, 0.3]$
data $s/n = 16/16 = 1$

Imprecise BBM with $n^{(0)}$ interval

Walley (1991, §5.4.3); Walter and Augustin (2009); Walter (2013)



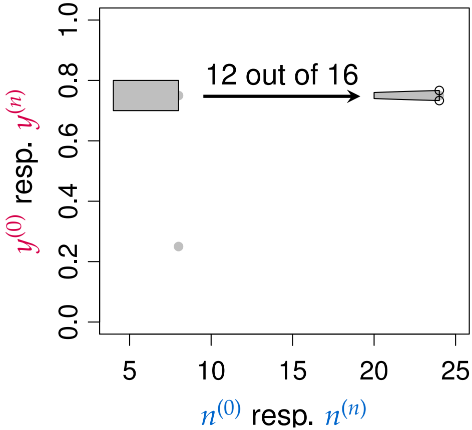
no conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

Technische Universiteit Eindhoven University of Technology

Imprecise BBM with $n^{(0)}$ interval

Walley (1991, §5.4.3); Walter and Augustin (2009); Walter (2013)



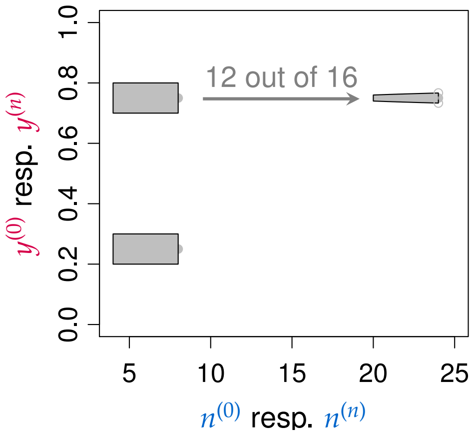
no conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

$y^{(n)} \in [0.73, 0.77]$

Imprecise BBM with $n^{(0)}$ interval

Walley (1991, §5.4.3); Walter and Augustin (2009); Walter (2013)



no conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

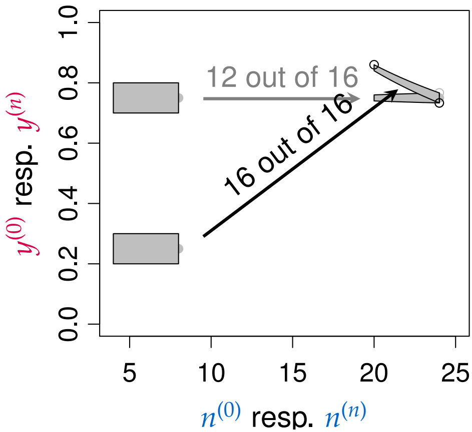
$$y^{(n)} \in [0.73, 0.77]$$

prior-data conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.2, 0.3]$
data $s/n = 16/16 = 1$

Imprecise BBM with $n^{(0)}$ interval

Walley (1991, §5.4.3); Walter and Augustin (2009); Walter (2013)



no conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

$y^{(n)} \in [0.73, 0.77]$

prior-data conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.2, 0.3]$
data $s/n = 16/16 = 1$

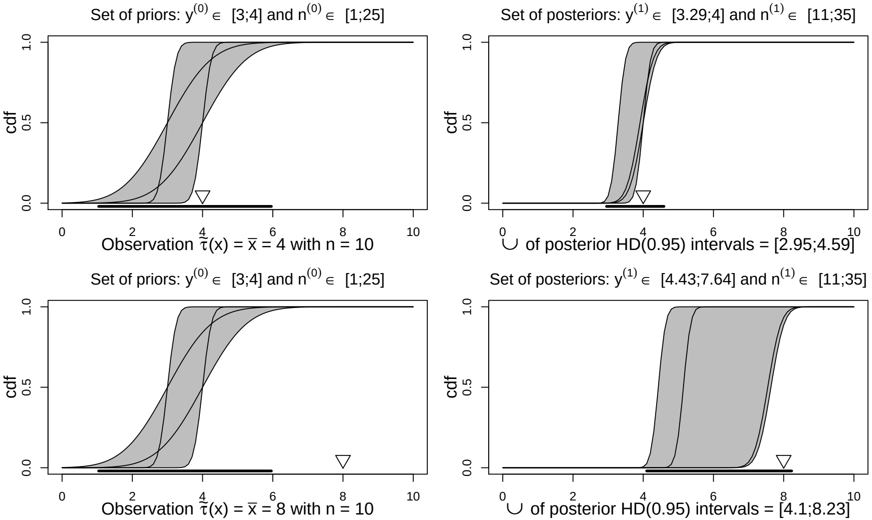
$y^{(n)} \in [0.73, 0.86]$

Example: Scaled Normal Data

Example: Scaled Normal Data

Data :	$\boldsymbol{x} \mid \mu$	$\sim$	$\mathrm{N}(\mu, 1)$
conjugate prior:	$\mu \mid n^{(0)}, y^{(0)}$	$\sim$	$\mathrm{N}(y^{(0)}, 1/n^{(0)})$
posterior:	$\mu \mid n^{(n)}, y^{(n)}$	$\sim$	$\mathrm{N}(y^{(n)}, 1/n^{(n)}) \quad (\tau(\boldsymbol{x})/n = \bar{x})$

Example: Scaled Normal Data



Canonical Exponential Families

Conjugate priors like the Beta can be constructed for sample distributions (likelihood) from:

Definition (Canonical exponential family)

$$f(x \mid \psi) = h(x) \exp\left\{\psi^T \tau(x) - b(\psi)\right\}$$

- includes multinomial, normal, Poisson, exponential, ...
- ψ generally a transformation of original parameter θ

Canonical Exponential Families

Conjugate priors like the Beta can be constructed for sample distributions (likelihood) from:

Definition (Canonical exponential family)

$$f(x \mid \psi) = h(x) \exp\left\{\psi^T \tau(x) - b(\psi)\right\}$$

- includes multinomial, normal, Poisson, exponential, ...
- ψ generally a transformation of original parameter θ

Definition (Family of conjugate priors)

A family of priors for i.i.d. sampling from the can. exp. family:

$$f(\psi \mid n^{(0)}, y^{(0)}) \propto \exp\left\{n^{(0)}\left[\psi^T y^{(0)} - b(\psi)\right]\right\}$$

with hyper-parameters $n^{(0)}$ and $y^{(0)}$.

Canonical Exponential Families

Theorem (Conjugacy)

Posterior is of the same form as the prior:

$$f(\psi \mid n^{(0)}, y^{(0)}, \boldsymbol{x}) \propto \exp\left\{n^{(n)}\left[\psi^T y^{(n)} - b(\psi)\right]\right\}$$

where

$$\boldsymbol{x} = (x_1, \ldots, x_n) \qquad \tau(\boldsymbol{x}) = \sum_{i=1}^{n} \tau(x_i)$$

$$n^{(n)} = n^{(0)} + n \qquad y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{\tau(\boldsymbol{x})}{n}$$

- $y^{(0)}$ = prior expectation of $\tau(\boldsymbol{x})/n$
- $n^{(0)}$ determines spread and learning speed

General Model Properties

Good inference properties

- $n \rightarrow \infty$

General Model Properties

Good inference properties

- $n \to \infty$ ► $y^{(n)}$ stretch in $\to 0$

General Model Properties

Good inference properties

- $n \to \infty$ ► $y^{(n)}$ stretch in $\to 0$ ► precise inferences

General Model Properties

Good inference properties

- $n \to \infty$ ► $y^{(n)}$ stretch in $\to 0$ ► precise inferences
- larger range of $y^{(0)}$ in ► larger range of $y^{(n)}$ in ► more vague inferences

General Model Properties

Good inference properties

- $n \to \infty$ ► $y^{(n)}$ stretch in $\to 0$ ► precise inferences
- larger range of $y^{(0)}$ in ► larger range of $y^{(n)}$ in ► more vague inferences

Model very easy to handle:

- Hyperparameter set defines set of priors $\mathcal{M}^{(0)}$

Technische Universiteit Eindhoven University of Technology

General Model Properties

Good inference properties

- $n \to \infty$ ► $y^{(n)}$ stretch in $\to 0$ ► precise inferences
- larger range of $y^{(0)}$ in ► larger range of $y^{(n)}$ in ► more vague inferences

Model very easy to handle:

- Hyperparameter set defines set of priors $\mathcal{M}^{(0)}$
- Hyperparameter set defines set of posteriors $\mathcal{M}^{(n)}$

General Model Properties

Good inference properties

- $n \to \infty$ ► $y^{(n)}$ stretch in $\to 0$ ► precise inferences
- larger range of $y^{(0)}$ in ► larger range of $y^{(n)}$ in
  ► more vague inferences

Model very easy to handle:

- Hyperparameter set defines set of priors $\mathcal{M}^{(0)}$
- Hyperparameter set defines set of posteriors $\mathcal{M}^{(n)}$
- $\to$ is easy: $n^{(n)} = n^{(0)} + n$, $y^{(n)} = \frac{n^{(0)}}{n^{(0)}+n} y^{(0)} + \frac{n}{n^{(0)}+n} \cdot \frac{\tau(\boldsymbol{x})}{n}$

General Model Properties

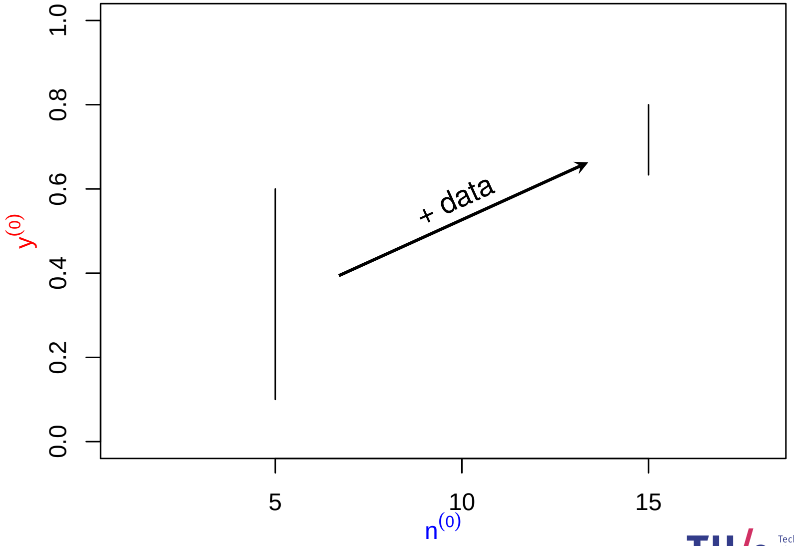
Good inference properties

- $n \to \infty$ ▶ $y^{(n)}$ stretch in $\to 0$ ▶ precise inferences
- larger range of $y^{(0)}$ in ▶ larger range of $y^{(n)}$ in ▶ more vague inferences

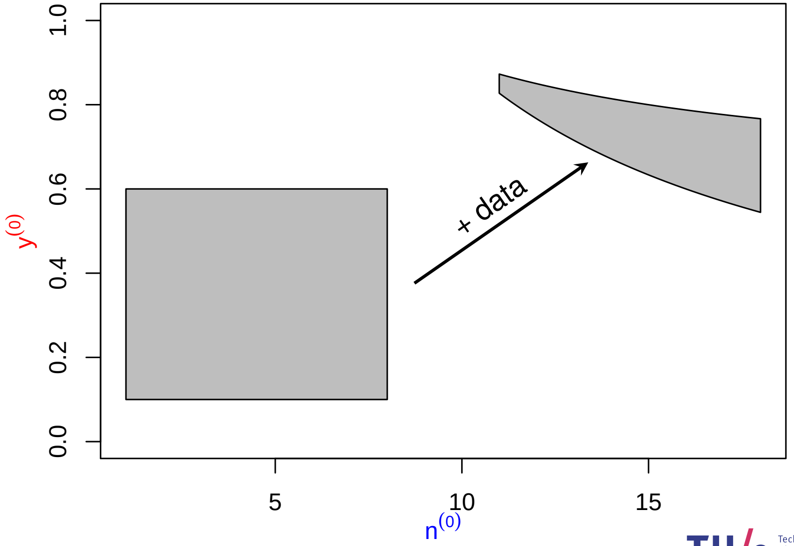
Model very easy to handle:

- Hyperparameter set defines set of priors $\mathcal{M}^{(0)}$
- Hyperparameter set defines set of posteriors $\mathcal{M}^{(n)}$
- $\to$ is easy: $n^{(n)} = n^{(0)} + n$, $y^{(n)} = \frac{n^{(0)}}{n^{(0)}+n} y^{(0)} + \frac{n}{n^{(0)}+n} \cdot \frac{\tau(\boldsymbol{x})}{n}$
- Often, optimising over $(n^{(n)}, y^{(n)}) \in$ is also easy: closed form solution for $y^{(n)}$ = posterior 'guess' for $\frac{\tau(\boldsymbol{x})}{n}$ (think: $\bar{x}$) when has 'nice' shape

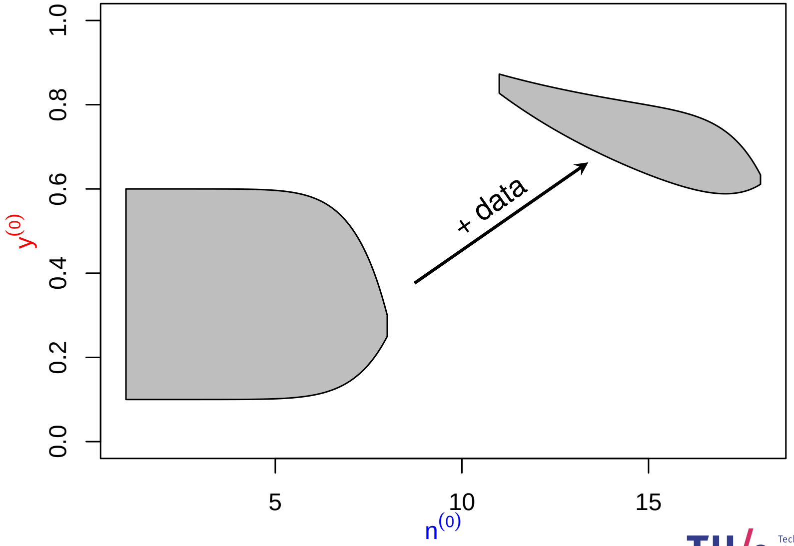
Hyperparameter Set Shapes



Hyperparameter Set Shapes



Hyperparameter Set Shapes



Hyperparameter Set Shapes

- Set shape is crucial modeling choice:
  trade-off between model complexity and model behaviour

Hyperparameter Set Shapes

- Set shape is crucial modeling choice:
  trade-off between model complexity and model behaviour
- $= n^{(0)} \times [\underline{y}^{(0)}, \overline{y}^{(0)}]$ (Walley 1996; Quaeghebeur and de Cooman 2005):
  $= n^{(n)} \times [\underline{y}^{(n)}, \overline{y}^{(n)}]$ ► optimise over $[\underline{y}^{(n)}, \overline{y}^{(n)}]$ only,
  but no prior-data conflict sensitivity

Hyperparameter Set Shapes

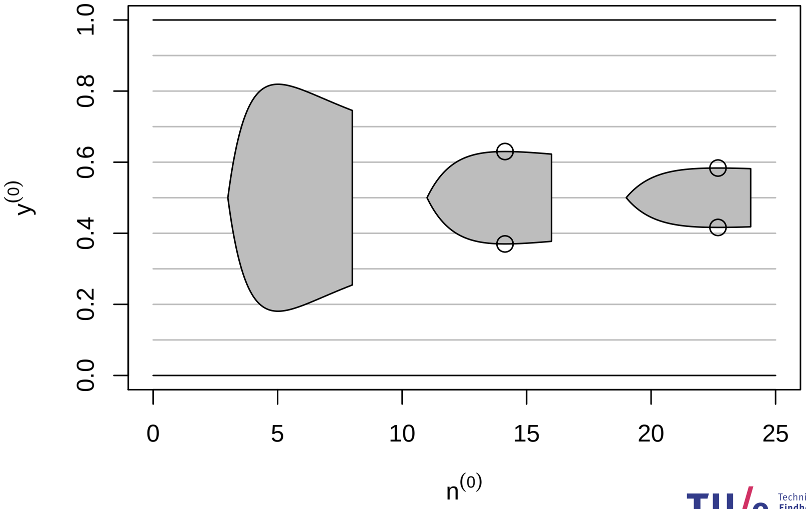
- Set shape is crucial modeling choice:
  trade-off between model complexity and model behaviour
- $= n^{(0)} \times [\underline{y}^{(0)}, \overline{y}^{(0)}]$ (Walley 1996; Quaeghebeur and de Cooman 2005):
  $= n^{(n)} \times [\underline{y}^{(n)}, \overline{y}^{(n)}]$ ► optimise over $[\underline{y}^{(n)}, \overline{y}^{(n)}]$ only,
  but no prior-data conflict sensitivity
- $= [\underline{n}^{(0)}, \overline{n}^{(0)}] \times [\underline{y}^{(0)}, \overline{y}^{(0)}]$ (Walley 1991; Walter and Augustin 2009):
  have non-trivial forms (banana / spotlight), but prior-data conflict sensitivity and closed form for $\min / \max y^{(n)}$ over
  implemented as R package `luck` (Walter and Krautenbacher 2013)

Hyperparameter Set Shapes

- Set shape is crucial modeling choice:
  trade-off between model complexity and model behaviour
- $= n^{(0)} \times [\underline{y}^{(0)}, \overline{y}^{(0)}]$ (Walley 1996; Quaeghebeur and de Cooman 2005):
  $= n^{(n)} \times [\underline{y}^{(n)}, \overline{y}^{(n)}]$ ► optimise over $[\underline{y}^{(n)}, \overline{y}^{(n)}]$ only,
  but no prior-data conflict sensitivity
- $= [\underline{n}^{(0)}, \overline{n}^{(0)}] \times [\underline{y}^{(0)}, \overline{y}^{(0)}]$ (Walley 1991; Walter and Augustin 2009):
  have non-trivial forms (banana / spotlight), but prior-data conflict sensitivity and closed form for $\min / \max y^{(n)}$ over
  implemented as R package `luck` (Walter and Krautenbacher 2013)
- Other set shapes possible, but may be more difficult to handle

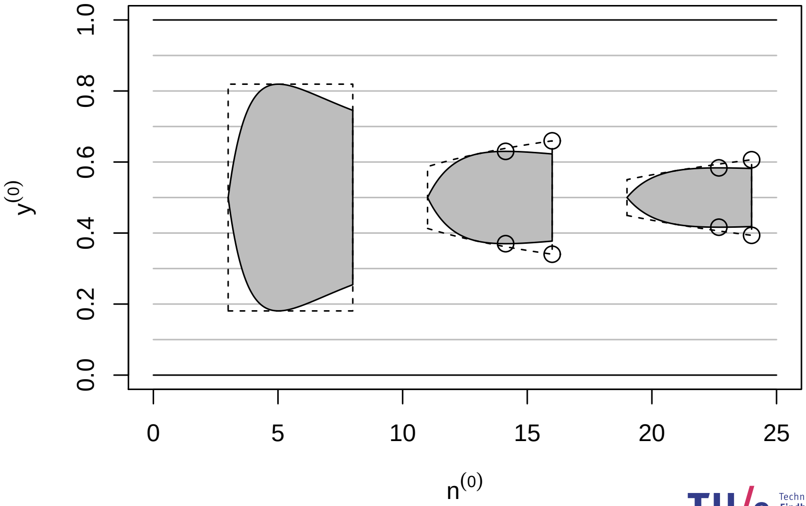
Hyperparameter Set Shapes

set shape for strong prior-data agreement (Walter and Coolen 2016)



Hyperparameter Set Shapes

set shape for strong prior-data agreement (Walter and Coolen 2016)



Robust Bayesian Analysis: Other Models

- How to define sets of priors $\mathcal{M}^{(0)}$ is a crucial modeling choice
- Sets $\mathcal{M}^{(0)}$ via parameter sets $\Pi^{(0)}$ seem to work better than other models discussed in the robust Bayes literature:
  - Neighbourhood models
    - set of distributions 'close to' a central distribution P_0
    - example: ε-contamination class: $\{P : P = (1 - \varepsilon)P_0 + \varepsilon Q, Q \in \mathcal{Q}\}$
      - ► not necessarily closed under Bayesian updating

Robust Bayesian Analysis: Other Models

- How to define sets of priors $\mathcal{M}^{(0)}$ is a crucial modeling choice
- Sets $\mathcal{M}^{(0)}$ via parameter sets $\Pi^{(0)}$ seem to work better than other models discussed in the robust Bayes literature:
  - Neighbourhood models
    - set of distributions 'close to' a central distribution P_0
    - example: ε-contamination class: $\{P : P = (1-\varepsilon)P_0 + \varepsilon Q, Q \in \mathcal{Q}\}$
      - ► not necessarily closed under Bayesian updating
  - Density ratio class / interval of measures
    - set of distributions by bounds for the density function $f(\vartheta)$:

$$\mathcal{M}^{(0)}_{l,u} = \left\{ f(\theta) : \exists c \in \mathbb{R}_{>0} : l(\theta) \leq c f(\theta) \leq u(\theta) \right\}$$

    - posterior set is bounded by updated $l(\theta)$ and $u(\theta)$
    - $u(\theta)/l(\theta)$ is constant under updating
      - ► size of the set does not decrease with n
      - ► very vague posterior inferences

Summary

- choice of prior can severely affect inferences
  even if your prior is ‘non-informative’
- solution: go imprecise
- models from canonical exponential family make this easy to do
  (Quaeghebeur and de Cooman 2005; Benavoli and Zaffalon 2012; 2015)
- allows to adequately express the quality of prior information
- close relations to robust Bayes literature
  (Berger et al. 1994; Ríos Insua and Ruggeri 2000)
- concerns uncertainty in the prior
  (uncertainty in data generating process: imprecise sampling models)
- if your prior is informative then prior-data conflict can be an issue
  (Walter and Augustin 2009; Walter 2013)

References I

Benavoli, A. and M. Zaffalon (2012). "A model of prior ignorance for inferences in the one-parameter exponential family". In: *Journal of Statistical Planning and Inference* 142, pp. 1960–1979. DOI: 10.1016/j.jspi.2012.01.023.

Benavoli, A. and M. Zaffalon (2015). "Prior near ignorance for inferences in the k-parameter exponential family". In: *Statistics* 49.5, pp. 1104–1140. DOI: 10.1080/02331888.2014.960869.

Berger et al. (1994). "An overview of robust Bayesian analysis". In: *TEST* 3, pp. 5–124.

Evans, M. and H. Moshonov (2006). "Checking for Prior-Data Conflict". In: *Bayesian Analysis* 1, pp. 893–914. URL: http://projecteuclid.org/euclid.ba/1340370946.

Gigerenzer, G., S. Krauss, and O. Vitouch (2004). "The Null Ritual – What you always wanted to know about significance testing but were afraid to ask". In: *The Sage handbook of quantitative Methodology for the social sciences*. Ed. by D. Kaplan. Thousand Oaks, CA: Sage, pp. 391–408.

Quaeghebeur, E. and G. de Cooman (2005). "Imprecise probability models for inference in exponential families". In: *ISIPTA '05. Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications*. Ed. by F. Cozman, R. Nau, and T. Seidenfeld. Manno: SIPTA, pp. 287–296. URL: http://leo.ugr.es/sipta/isipta05/proceedings/papers/s019.pdf.

Ríos Insua, D. and F. Ruggeri (2000). *Robust Bayesian Analysis*. New York: Springer.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall.

References II

Walley, P. (1996). “Inferences from multinomial data: Learning about a bag of marbles”. In: *Journal of the Royal Statistical Society, Series B* 58.1, pp. 3–34. URL: http://www.jstor.org/stable/2346164.

Walter, G. (2013). “Generalized Bayesian Inference under Prior-Data Conflict”. PhD thesis. Department of Statistics, LMU Munich. URL: http://edoc.ub.uni-muenchen.de/17059/.

Walter, G. and T. Augustin (2009). “Imprecision and Prior-data Conflict in Generalized Bayesian Inference”. In: *Journal of Statistical Theory and Practice* 3, pp. 255–271. DOI: 10.1080/15598608.2009.10411924.

Walter, G. and F. P. A. Coolen (2016). “Sets of Priors Reflecting Prior-Data Conflict and Agreement”. In: *Information Processing and Management of Uncertainty in Knowledge-Based Systems: 16th International Conference, IPMU 2016, Eindhoven, The Netherlands, June 20-24, 2016, Proceedings, Part I*. Ed. by Paulo Joao Carvalho et al. Cham: Springer International Publishing, pp. 153–164. ISBN: 978-3-319-40596-4. DOI: 10.1007/978-3-319-40596-4_14.

Walter, G. and N. Krautenbacher (2013). luck*: R package for Generalized iLUCK-models*. URL: http://luck.r-forge.r-project.org/.