# Bayesian Inference and Prior-Data Conflict

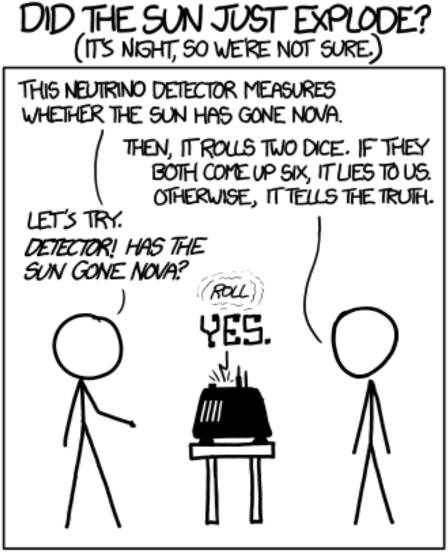
Gero Walter

Eindhoven University of Technology, Eindhoven, NL

g.m.walter@tue.nl

2015-12-15

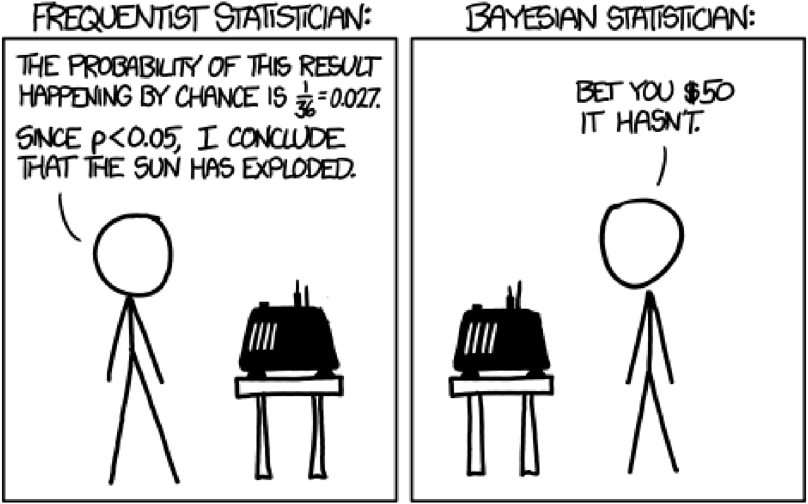
# Bayesian Inference




https://xkcd.com/1132

# Bayesian Inference






https://xkcd.com/1132

# Bayesian Inference



expert info + data → complete picture

# Bayesian Inference



| | | | | |
|---|---|---|---|---|
| expert info | + | data | $\rightarrow$ | complete picture |
| prior distribution | + | sample distribution | $\rightarrow$ | posterior distribution |
| $f(\theta)$ | $\times$ | $f(\boldsymbol{x} \mid \theta)$ | $\propto$ | $p(\theta \mid \boldsymbol{x})$ |

► Bayes' Rule

# Bayesian Inference

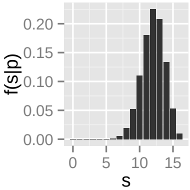


| | | | | |
|---|---|---|---|---|
| expert info | + | data | → | complete picture |
| prior distribution | + | sample distribution | → | posterior distribution |
| $f(\theta)$ | $\times$ | $f(\boldsymbol{x} \mid \theta)$ | $\propto$ | $p(\theta \mid \boldsymbol{x})$ |

► Bayes' Rule

Binomial distribution

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



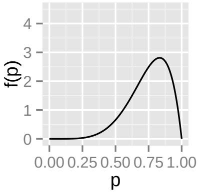
# Bayesian Inference

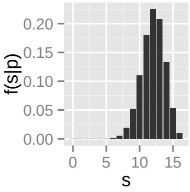


| | | | | |
|---|---|---|---|---|
| expert info | + | data | $\rightarrow$ | complete picture |
| prior distribution | + | sample distribution | $\rightarrow$ | posterior distribution |
| $f(\theta)$ | $\times$ | $f(\boldsymbol{x} \mid \theta)$ | $\propto$ | $p(\theta \mid \boldsymbol{x})$ |

► Bayes' Rule

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

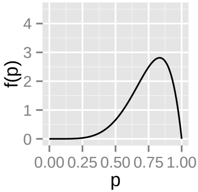


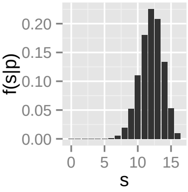



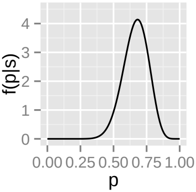
# Bayesian Inference



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




Technische Universiteit Eindhoven University of Technology

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

- conjugate prior makes learning about parameter tractable, just update hyperparameters: $\alpha^{(0)} \rightarrow \alpha^{(n)}$, $\beta^{(0)} \rightarrow \beta^{(n)}$
- closed form for some inferences: $\text{E}[p \mid s] = \frac{\alpha^{(n)}}{\alpha^{(n)}+\beta^{(n)}}$

# 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

# Prior-Data Conflict: Example



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

# Prior-Data Conflict: Example



- Bernoulli observations: 0/1 observations (failure/success)
- given: a set of $n$ i.i.d. observations and strong prior information

# Prior-Data Conflict: Example



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

# Prior-Data Conflict: Example



- Bernoulli observations: 0/1 observations (failure/success)
- given: a set of $n$ i.i.d. observations 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: | $p \mid \alpha^{(n)}, \beta^{(n)}$ | $\sim$ | $\text{Beta}(\alpha^{(n)}, \beta^{(n)})$ |

where $s$ = number of successes in the $n$ observed trials

# Reparametrisation of the Beta Distribution



- reparametrisation helps to understand effect of prior-data conflict:

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

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

# Reparametrisation of the Beta Distribution



- reparametrisation helps to understand effect of prior-data conflict:

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

which are updated as

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

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

# Reparametrisation of the Beta Distribution



- reparametrisation helps to understand effect of prior-data conflict:

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

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

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

# Reparametrisation of the Beta Distribution

- reparametrisation helps to understand effect of prior-data conflict:

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

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

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

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

ML estimator $\hat{p}$

# Reparametrisation of the Beta Distribution



- reparametrisation helps to understand effect of prior-data conflict:

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

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 effect of prior-data conflict:

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

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 $\hat{p}$!**

# Reparametrisation of the Beta Distribution



- reparametrisation helps to understand effect of prior-data conflict:

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

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

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

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

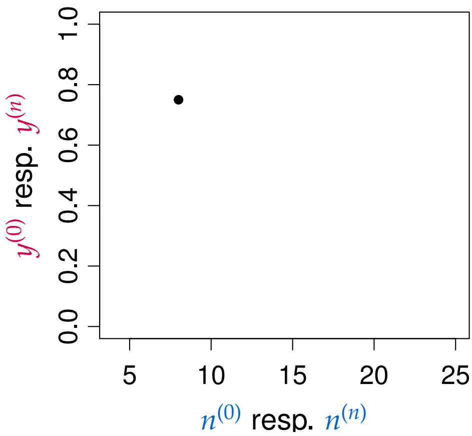
ML estimator $\hat{p}$

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

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

# Beta-Binomial Model (BBM)





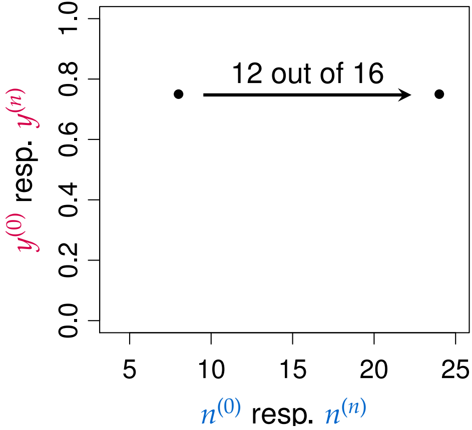

**no conflict:**

prior $n^{(0)} = 8$, $y^{(0)} = 0.75$

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

# Beta-Binomial Model (BBM)





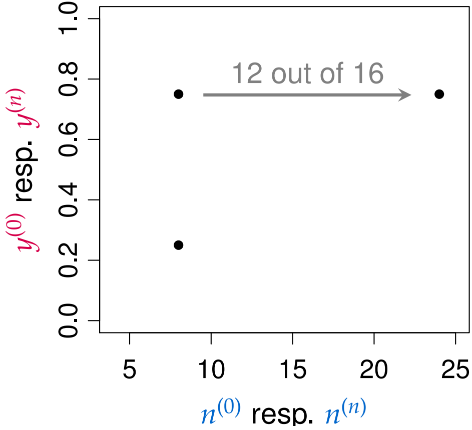

**no conflict:**

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 $s/n = 12/16 = 0.75$

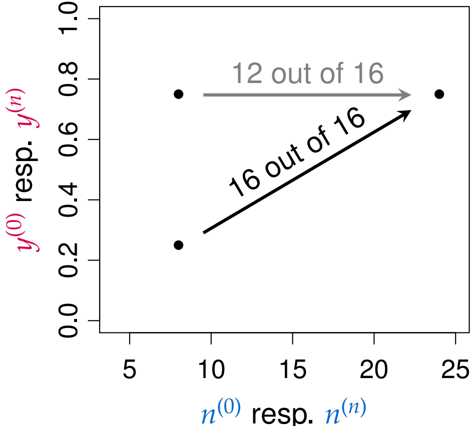
$n^{(n)} = 24$, $y^{(n)} = 0.75$

**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 $s/n = 16/16 = 1$

# Canonical Conjugate Priors



Averaging property holds *for all conjugate models* (!)

$(x_1, \ldots, x_n) = \boldsymbol{x} \overset{iid}{\sim}$ canonical exponential family

$$f(\boldsymbol{x} \mid \theta) \propto \exp\left\{\langle \psi, \tau(\boldsymbol{x}) \rangle - n b(\psi)\right\} \qquad \left[\psi \text{ transformation of } \theta\right]$$

(includes Binomial, Multinomial, Normal, Poisson, Exponential, …)

# Canonical Conjugate Priors



Averaging property holds *for all conjugate models* (!)

$(x_1, \ldots, x_n) = \boldsymbol{x} \overset{iid}{\sim}$ canonical exponential family

$$f(\boldsymbol{x} \mid \theta) \propto \exp\left\{\langle \psi, \tau(\boldsymbol{x}) \rangle - n b(\psi)\right\} \qquad \left[\psi \text{ transformation of } \theta\right]$$

(includes Binomial, Multinomial, Normal, Poisson, Exponential, ...)

- conjugate prior: $f(\psi \mid n^{(0)}, y^{(0)}) \propto \exp\left\{n^{(0)}\left[\langle \psi, y^{(0)} \rangle - b(\psi)\right]\right\}$

# Canonical Conjugate Priors



Averaging property holds *for all conjugate models* (!)

$(x_1, \ldots, x_n) = \boldsymbol{x} \overset{iid}{\sim}$ canonical exponential family

$$f(\boldsymbol{x} \mid \theta) \propto \exp\left\{\langle \psi, \tau(\boldsymbol{x}) \rangle - n b(\psi)\right\} \qquad \left[\psi \text{ transformation of } \theta\right]$$

(includes Binomial, Multinomial, Normal, Poisson, Exponential, . . . )

- conjugate prior: $f(\psi \mid n^{(0)}, y^{(0)}) \propto \exp\left\{n^{(0)}\left[\langle \psi, y^{(0)} \rangle - b(\psi)\right]\right\}$
- (conjugate) posterior: $f(\psi \mid n^{(0)}, y^{(0)}, \boldsymbol{x}) \propto \exp\left\{n^{(n)}\left[\langle \psi, y^{(n)} \rangle - b(\psi)\right]\right\}$

where $y^{(n)} = \dfrac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \dfrac{n}{n^{(0)} + n} \cdot \dfrac{\tau(\boldsymbol{x})}{n}$ and $n^{(n)} = n^{(0)} + n$

# Canonical Conjugate Priors



Averaging property holds *for all conjugate models* (!)

$(x_1, \ldots, x_n) = \boldsymbol{x} \overset{iid}{\sim}$ canonical exponential family

$$f(\boldsymbol{x} \mid \theta) \propto \exp\left\{\langle \psi, \tau(\boldsymbol{x})\rangle - n b(\psi)\right\} \qquad \left[\psi \text{ transformation of } \theta\right]$$

(includes Binomial, Multinomial, Normal, Poisson, Exponential, . . . )

- conjugate prior: $f(\psi \mid n^{(0)}, y^{(0)}) \quad \propto \exp\left\{n^{(0)}\left[\langle \psi, y^{(0)}\rangle - b(\psi)\right]\right\}$
- (conjugate) posterior: $f(\psi \mid n^{(0)}, y^{(0)}, \boldsymbol{x}) \propto \exp\left\{n^{(n)}\left[\langle \psi, y^{(n)}\rangle - b(\psi)\right]\right\}$

where $\quad y^{(n)} = \dfrac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \dfrac{n}{n^{(0)} + n} \cdot \dfrac{\tau(\boldsymbol{x})}{n}$ and $\quad n^{(n)} = n^{(0)} + n$

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

# Imprecise / Interval Probability



- Averaging property holds *for all conjugate models* (!)
  Can we mitigate this and still keep tractability?

# Imprecise / Interval Probability



- Averaging property holds *for all conjugate models* (!)
  Can we mitigate this and still keep tractability?
- Prior $f(p)$ is a collection of probability statements:

$$\int_a^b f(p)\,\mathrm{d}p = P(a \le p \le b)$$

How can we express uncertainty
about these probability statements?

# Imprecise / Interval Probability



- Averaging property holds *for all conjugate models* (!)
  Can we mitigate this and still keep tractability?
- Prior $f(p)$ is a collection of probability statements:

$$\int_a^b f(p)\,\mathrm{d}p = P(a \le p \le b)$$

How can we express uncertainty
about these probability statements?

- **Add imprecision as new modelling dimension:**
  **Sets of priors model uncertainty in probability statements and allow to better model partial or vague information on $p$.**

# Imprecise / Interval Probability



- Averaging property holds *for all conjugate models* (!)
  Can we mitigate this and still keep tractability?
- Prior $f(p)$ is a collection of probability statements:

$$\int_a^b f(p)\,\mathrm{d}p = P(a \leq p \leq b)$$

How can we express uncertainty
about these probability statements?

- **Add imprecision as new modelling dimension:**
  **Sets of priors model uncertainty in probability statements and allow to better model partial or vague information on $p$.**
- Separate uncertainty *within the model* (probability statements) from uncertainty *about the model* (how certain about statements)

# Imprecise / Interval Probability



- Averaging property holds *for all conjugate models* (!)
  Can we mitigate this and still keep tractability?
- Prior $f(p)$ is a collection of probability statements:

$$\int_a^b f(p)\,\mathrm{d}p = P(a \leq p \leq b)$$

How can we express uncertainty
about these probability statements?

- **Add imprecision as new modelling dimension:**
  **Sets of priors model uncertainty in probability statements and allow to better model partial or vague information on $p$.**
- Separate uncertainty *within the model* (probability statements) from uncertainty *about the model* (how certain about statements)
- Can also be seen as systematic sensitivity analysis or robust Bayesian approach.

# Sets of Prior Distributions



## Uncertainty about probability statements

smaller sets = 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

- $P(\text{win}) = [1/100, 7/100]$

# Sets of Prior Distributions



## Uncertainty about probability statements

smaller sets = 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

► $P(\text{win}) = [1/100, 7/100]$

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

# Sets of Prior Distributions



## Uncertainty about probability statements

smaller sets = 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
▶ $P(\text{win}) = [1/100, 7/100]$

Let hyperparameters $(n^{(0)}, y^{(0)})$ vary in a set ▶ set of priors $\mathcal{M}^{(0)}$

Sets of priors → sets of posteriors by updating element by element: the Generalized Bayes Rule (GBR Walley 1991) ensures *coherence* (a consistency property)

## Sets of Prior Distributions



### Uncertainty about probability statements

smaller sets = 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
► $P(\text{win}) = [1/100, 7/100]$

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

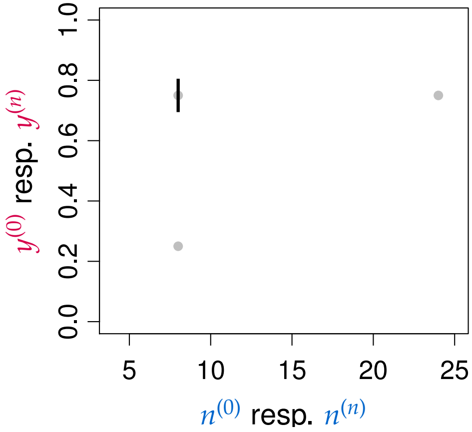
Sets of priors → sets of posteriors by updating element by element: the Generalized Bayes Rule (GBR Walley 1991) ensures *coherence* (a consistency property)

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

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

# Imprecise BBM with $n^{(0)}$ fixed

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



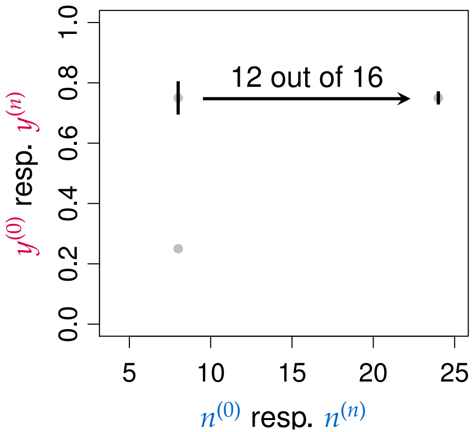

**no conflict:**

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

# Imprecise BBM with $n^{(0)}$ fixed



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



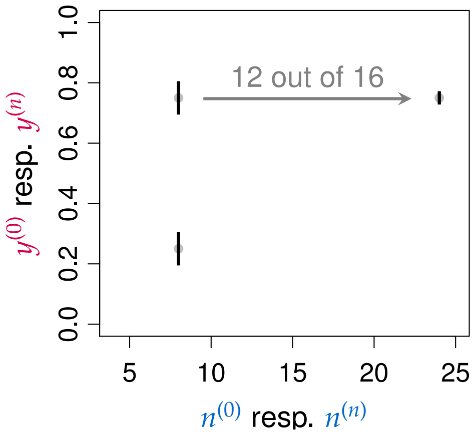

**no conflict:**

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 with $n^{(0)}$ fixed

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




**no conflict:**

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

▼

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

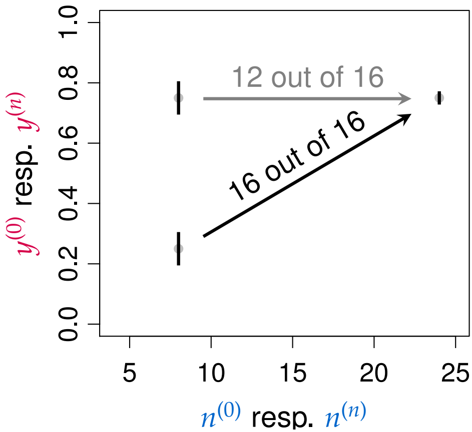
**prior data conflict:**

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

# Imprecise BBM with $n^{(0)}$ fixed



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




**no conflict:**

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

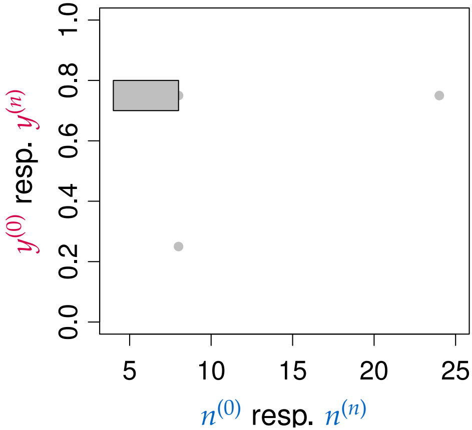
$n^{(n)} = 24$, $y^{(n)} \in [0.73, 0.77]$

**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)



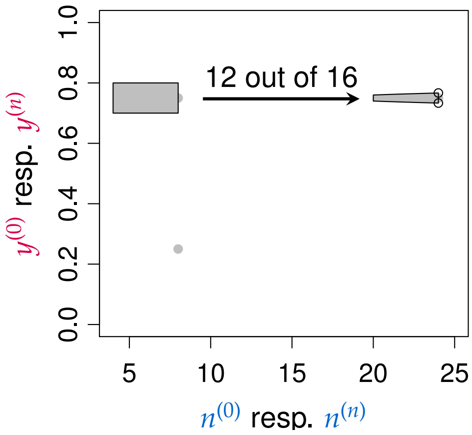

**no conflict:**

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

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



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




**no conflict:**

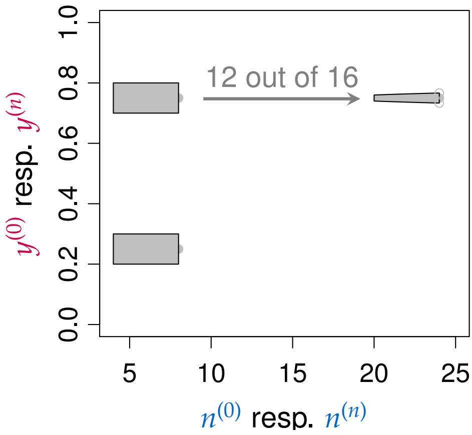
prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

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

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

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

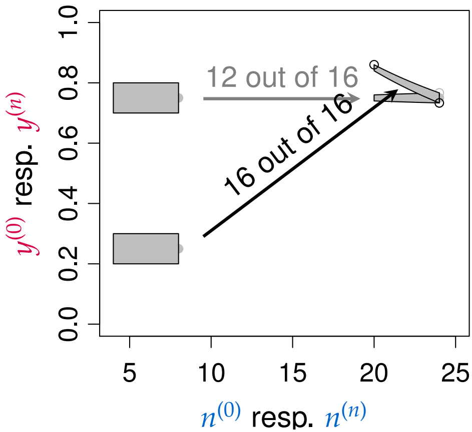
**prior-data conflict:**

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 $s/n = 12/16 = 0.75$

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

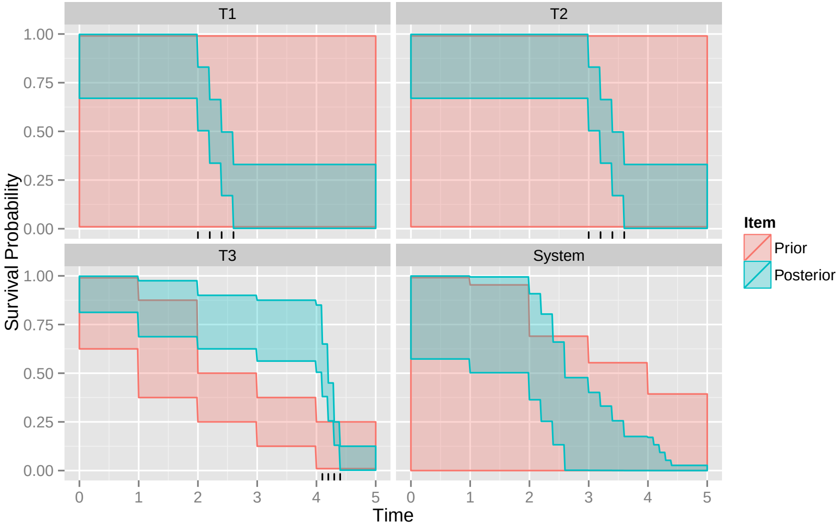
**prior-data conflict:**

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

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

# Sets of Nonparametric Survival Functions






(joint work with Louis Aslett and Frank Coolen)

# Example: Scaled Normal Data

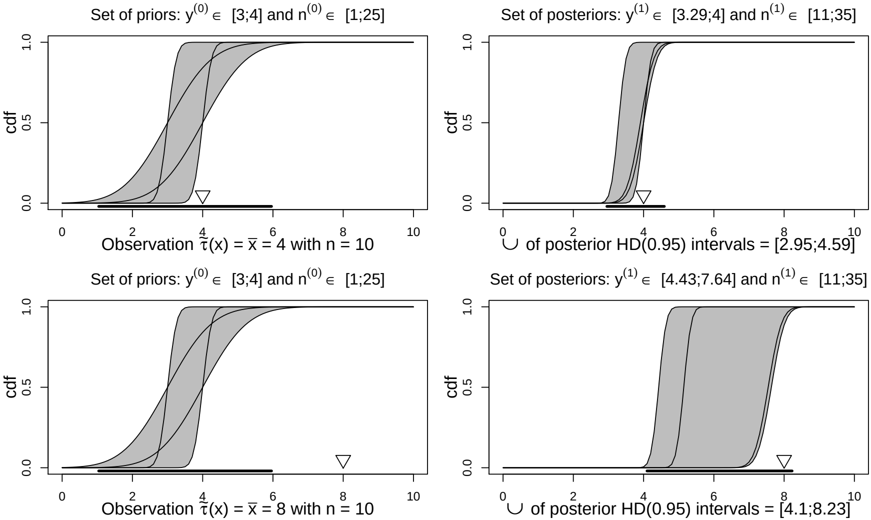


## Example: Scaled Normal Data

| | | | |
|---|---|---|---|
| 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





# General Model Properties



Good inference properties (cf. other models based on sets of priors)

- $n \to \infty$

# General Model Properties



Good inference properties (cf. other models based on sets of priors)

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

# General Model Properties



Good inference properties (cf. other models based on sets of priors)

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

# General Model Properties



Good inference properties (cf. other models based on sets of priors)

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

# General Model Properties



Good inference properties (cf. other models based on sets of priors)

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

# General Model Properties



Good inference properties (cf. other models based on sets of priors)

- $n \to \infty$ ► $y^{(n)}$ stretch in $\to 0$ ► precise inferences
- larger $n^{(0)}$ ► larger ► more vague 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 (cf. other models based on sets of priors)

- $n \to \infty$ ► $y^{(n)}$ stretch in $\to 0$ ► precise inferences
- larger $n^{(0)}$ ► larger ► more vague 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 (cf. other models based on sets of priors)

- $n \to \infty$ ► $y^{(n)}$ stretch in $\to 0$ ► precise inferences
- larger $n^{(0)}$ ► larger ► more vague 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 (cf. other models based on sets of priors)

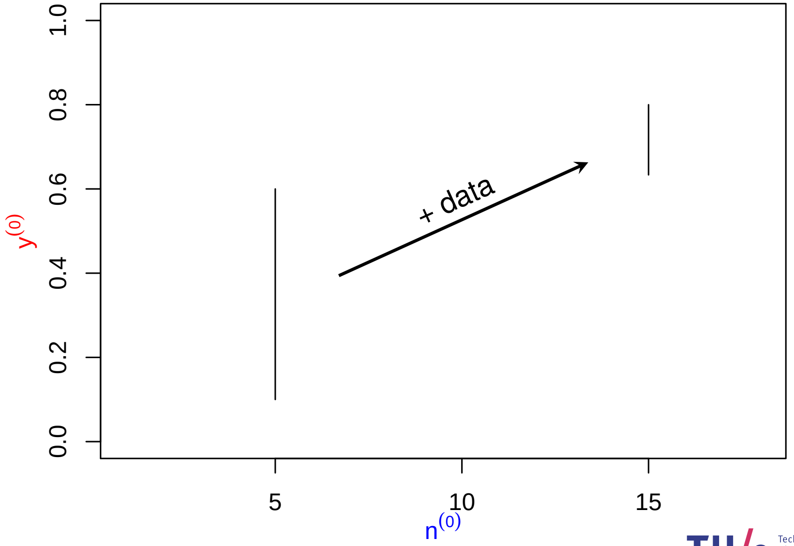
- $n \to \infty$ ▶ $y^{(n)}$ stretch in $\to 0$ ▶ precise inferences
- larger $n^{(0)}$ ▶ larger ▶ more vague inferences
- larger range of $y^{(0)}$ in ▶ larger range of $y^{(n)}$ in ▶ more vague inferences

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

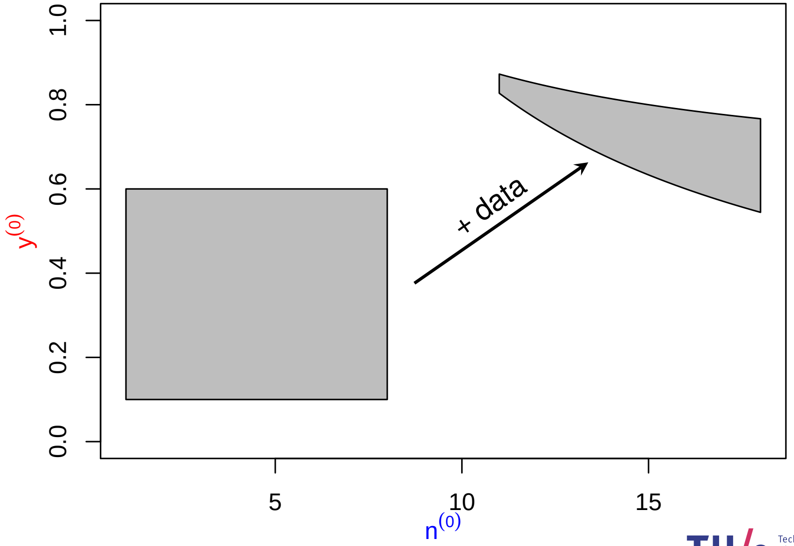
# Hyperparameter Set Shapes





# Hyperparameter Set Shapes



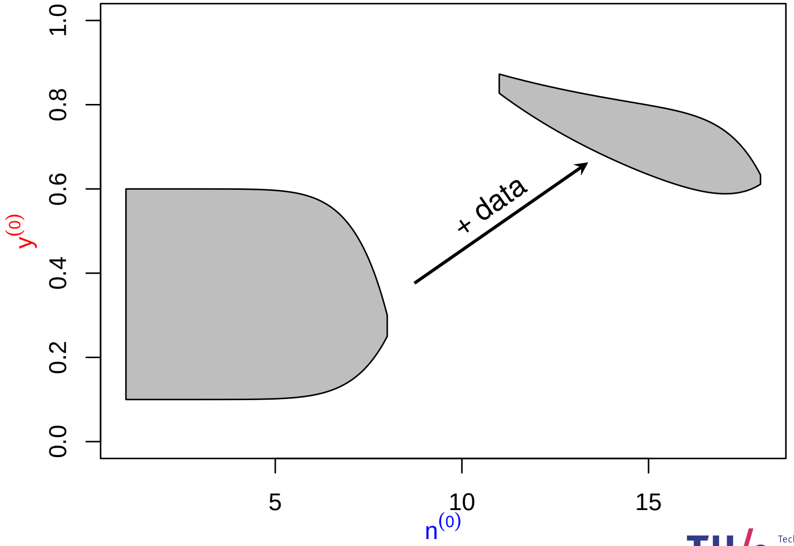





Technische Universiteit Eindhoven University of Technology

# 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 .
  For other inferences, **R** package `luck` implements optimisation over via box-constraint optimisation over

# Hyperparameter Set Shapes

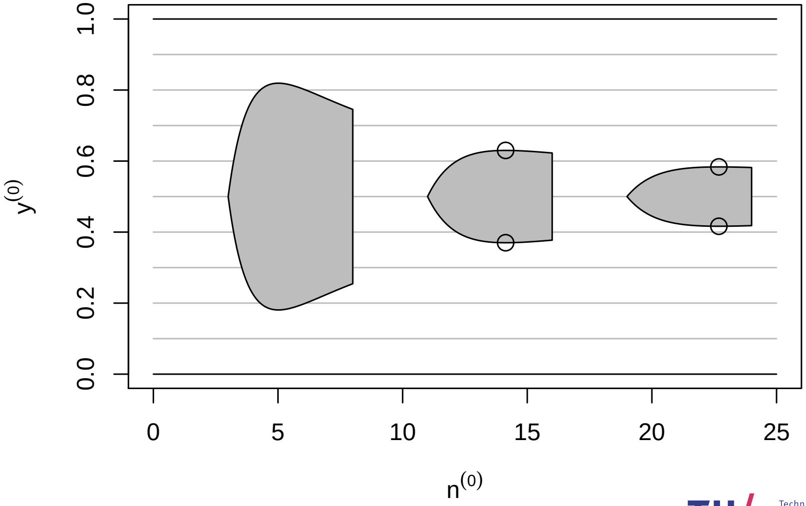


- Set shape is crucial modeling choice:
  trade-off between model complexity and model behaviour
- $= n^{(0)} \times [\underline{y}^{(0)}, \overline{y}^{(0)}]$ (Walley 1996; Quaeghebeur and de Cooman 2005):
  $= n^{(n)} \times [\underline{y}^{(n)}, \overline{y}^{(n)}]$ ► optimise over $[\underline{y}^{(n)}, \overline{y}^{(n)}]$ 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 .
  For other inferences, **R** package `luck` implements optimisation over via box-constraint optimisation over
- Other set shapes possible, but may be more difficult to handle

# Hyperparameter Set Shapes



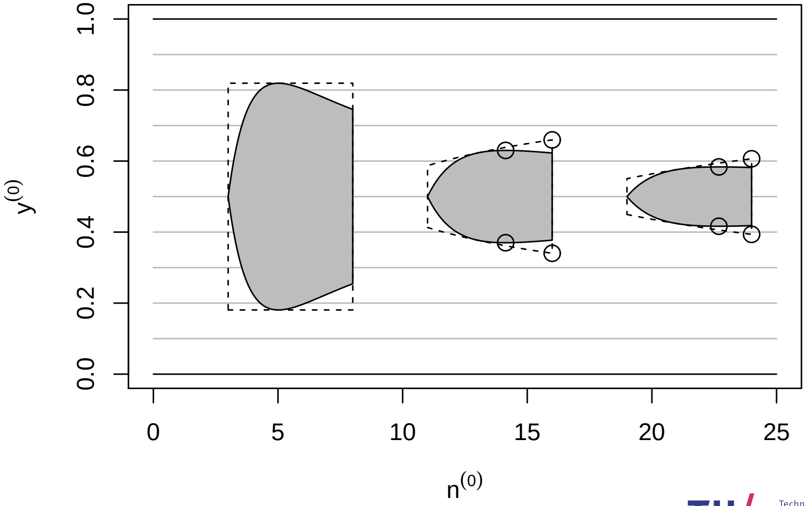
Parameter set shape for strong prior-data agreement (Walter 2013, A.2)



# Hyperparameter Set Shapes



Parameter set shape for strong prior-data agreement (Walter 2013, A.2)



# Summary



- Conjugate priors are a convenient tool for Bayesian inference but there are some pitfalls
  - Hyperparameters $n^{(0)}, y^{(0)}$ are easy to interpret and elicit
  - Averaging property makes calculations simple, but leads to inadequate model behaviour in case of prior-data conflict

# Summary



- Conjugate priors are a convenient tool for Bayesian inference but there are some pitfalls
  - Hyperparameters $n^{(0)}, y^{(0)}$ are easy to interpret and elicit
  - Averaging property makes calculations simple, but leads to inadequate model behaviour in case of prior-data conflict
- Sets of conjugate priors maintain advantages & mitigate issues
  - Sets of posteriors adequately reflect vague prior information, the amount of data, and prior-data conflict
  - Hyperparameter set shape is important
  - Reasonable choice: *rectangular* $= [\underline{n}^{(0)}, \overline{n}^{(0)}] \times [\underline{y}^{(0)}, \overline{y}^{(0)}]$: *"generalised iLUCK-models"* (Walter and Augustin 2009; Walter 2013), **R** package `luck` (Walter and Krautenbacher 2013)
  - Bounds for prior hyperparameters $(n^{(0)}, y^{(0)})$ are easy to interpret and elicit
  - Additional imprecison in case of prior-data conflict leads to cautious inferences if, and only if, caution is needed

# References I



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.

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.

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

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.

# References II



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

- Neighbourhood models
  - set of distributions 'close to' a central distribution $P_0$
  - common in robust Bayesian approaches
  - example: $\varepsilon$-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 $f(\vartheta)$:

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

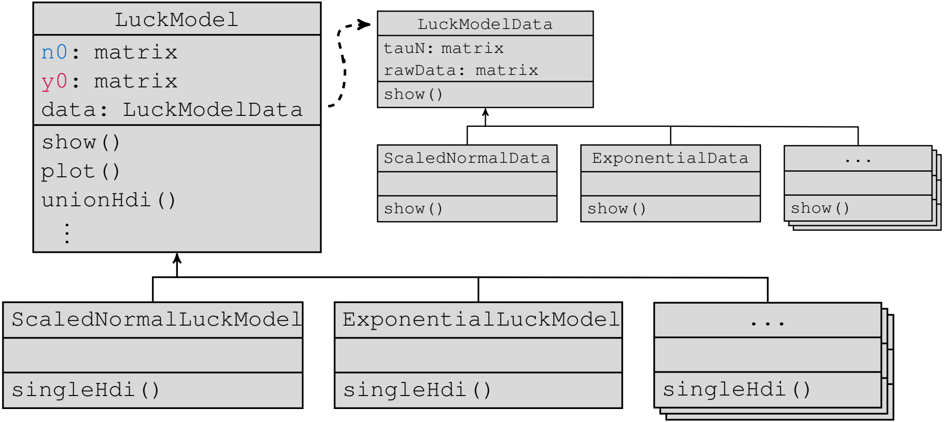
  - posterior set is bounded by updated $l(\theta)$ and $u(\theta)$
  - $u(\theta)/l(\theta)$ is constant under updating
    - ► size of the set does not decrease with $n$
    - ► too vague posterior inferences

- `S4` implementation of the general canonical prior parameter structure with rectangular sets $= [\underline{n}^{(0)}, \overline{n}^{(0)}] \times [\underline{y}^{(0)}, \overline{y}^{(0)}]$
- lean subclasses for concrete sample distributions (currently implemented: scaled normal, exponential)
- available on R-Forge:

```
install.packages("luck",repos="http://R-Forge.R-project.org")
```

or

```
install.packages("http://download.r-forge.r-project.org/src/contrib/luck_0.9.tar.gz",repos=NULL,type="source")
```


LuckModel
n0: matrix
y0: matrix
data: LuckModelData
show()
plot()
unionHdi()
⋮
LuckModelData
tauN: matrix
rawData: matrix
show()
ScaledNormalData
show()
ExponentialData
show()
...
show()
ScaledNormalLuckModel
singleHdi()
ExponentialLuckModel
singleHdi()
...
singleHdi()

# Strong Prior-Data Agreement Modelling