Prior-Data Conflict in Generalised Bayesian Inference

Gero Walter

Department of Statistics
Ludwig-Maximilians-Universität München (LMU)

September 16th, 2013

Outline

1. Bayesian inference & prior-data conflict
2. Generalised Bayesian inference with sets of priors
   (joint work with Thomas Augustin)
3. Common-cause failure modeling
   (joint work with Matthias Troffaes and Dana Kelly)

Bayesian Inference & Prior-Data Conflict

The Bayesian approach to statistical inference

prior $p(\vartheta)$ + likelihood $f(\boldsymbol{x} \mid \vartheta)$ $\longrightarrow$ posterior $p(\vartheta \mid \boldsymbol{x})$

All inferences are based on the posterior (e.g., point estimate, ...)

Bayesian Inference & Prior-Data Conflict

The Bayesian approach to statistical inference

prior $p(\vartheta)$ + likelihood $f(\boldsymbol{x} \mid \vartheta)$ $\longrightarrow$ posterior $p(\vartheta \mid \boldsymbol{x})$

All inferences are based on the posterior (e.g., point estimate, ...)

Assigning a certain prior distribution on ϑ

$=$ defining a conglomerate of probability statements (on ϑ).

Bayesian Inference & Prior-Data Conflict

The Bayesian approach to statistical inference

prior $p(\vartheta)$ + likelihood $f(\boldsymbol{x} \mid \vartheta)$ $\longrightarrow$ posterior $p(\vartheta \mid \boldsymbol{x})$

All inferences are based on the posterior (e.g., point estimate, ...)

Assigning a certain prior distribution on ϑ

$=$ defining a conglomerate of probability statements (on ϑ).

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 & Moshonov, 2006)
- there are not enough data to overrule the prior

Prior-Data Conflict: Basic Example

- Bernoulli observations: 0/1 observations (team wins no/yes)

Prior-Data Conflict: Basic Example

- Bernoulli observations: 0/1 observations (team wins no/yes)
- given: a set of observations and strong prior information

Prior-Data Conflict: Basic Example

- Bernoulli observations: 0/1 observations (team wins no/yes)
- given: a set of observations and strong prior information
- we are, e.g., interested in (predictive) probability P that team wins in the next match

Prior-Data Conflict: Basic Example

- Bernoulli observations: 0/1 observations (team wins no/yes)
- given: a set of observations and strong prior information
- we are, e.g., interested in (predictive) probability P that team wins in the next match

Beta-Binomial Model

$$
\begin{array}{r|lcl}
\text{data}: & s \mid \theta & \sim & \mathrm{Binom}(n, \theta) \\
\text{conjugate prior}: & \theta \mid n^{(0)}, y^{(0)} & \sim & \mathrm{Beta}(n^{(0)}, y^{(0)}) \\
\hline
\text{posterior}: & \theta \mid n^{(n)}, y^{(n)} & \sim & \mathrm{Beta}(n^{(n)}, y^{(n)})
\end{array}
$$

where s = number of wins in the n matches observed

Beta-Binomial Model

Beta-Binomial Model

$$
\begin{array}{r|lcl}
\text{data :} & s \mid \theta & \sim & \mathrm{Binom}(n, \theta) \\
\text{conjugate prior:} & \theta \mid n^{(0)}, y^{(0)} & \sim & \mathrm{Beta}(n^{(0)}, y^{(0)}) \\
\hline
\text{posterior:} & \theta \mid n^{(n)}, y^{(n)} & \sim & \mathrm{Beta}(n^{(n)}, y^{(n)})
\end{array}
$$

Beta-Binomial Model

Beta-Binomial Model

$$
\begin{array}{r|lcl}
\text{data}: & s \mid \theta & \sim & \mathrm{Binom}(n, \theta) \\
\text{conjugate prior}: & \theta \mid n^{(0)}, y^{(0)} & \sim & \mathrm{Beta}(n^{(0)}, y^{(0)}) \\
\hline
\text{posterior}: & \theta \mid n^{(n)}, y^{(n)} & \sim & \mathrm{Beta}(n^{(n)}, y^{(n)})
\end{array}
$$

$$P = \mathrm{E}[\theta \mid n^{(n)}, y^{(n)}]$$

Beta-Binomial Model

Beta-Binomial Model

$$
\begin{array}{r|lcl}
\text{data}: & s \mid \theta & \sim & \mathrm{Binom}(n, \theta) \\
\text{conjugate prior}: & \theta \mid n^{(0)}, y^{(0)} & \sim & \mathrm{Beta}(n^{(0)}, y^{(0)}) \\
\hline
\text{posterior}: & \theta \mid n^{(n)}, y^{(n)} & \sim & \mathrm{Beta}(n^{(n)}, y^{(n)})
\end{array}
$$

$$
P = \mathrm{E}[\theta \mid n^{(n)}, y^{(n)}] = y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}
$$

Beta-Binomial Model

Beta-Binomial Model

$$
\begin{array}{r|lcl}
\text{data}: & s \mid \theta & \sim & \mathrm{Binom}(n, \theta) \\
\text{conjugate prior}: & \theta \mid n^{(0)}, y^{(0)} & \sim & \mathrm{Beta}(n^{(0)}, y^{(0)}) \\
\hline
\text{posterior}: & \theta \mid n^{(n)}, y^{(n)} & \sim & \mathrm{Beta}(n^{(n)}, y^{(n)})
\end{array}
$$

$$P = \mathrm{E}[\theta \mid n^{(n)}, y^{(n)}] = y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}$$

$$n^{(n)} = n^{(0)} + n$$

Beta-Binomial Model

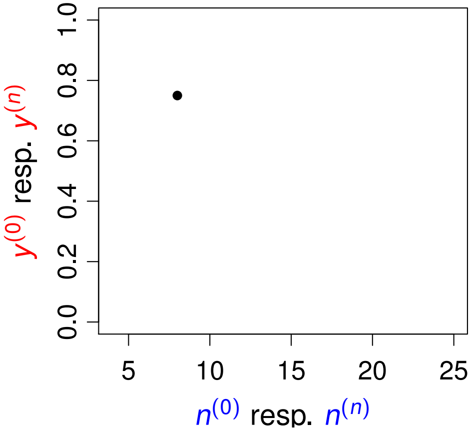
Beta-Binomial Model

$$\begin{array}{r|lcl} \text{data}: & s \mid \theta & \sim & \text{Binom}(n, \theta) \\ \text{conjugate prior}: & \theta \mid n^{(0)}, y^{(0)} & \sim & \text{Beta}(n^{(0)}, y^{(0)}) \\ \hline \text{posterior}: & \theta \mid n^{(n)}, y^{(n)} & \sim & \text{Beta}(n^{(n)}, y^{(n)}) \end{array}$$

$$P = \mathrm{E}[\theta \mid n^{(n)}, y^{(n)}] = y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}$$

$$n^{(n)} = n^{(0)} + n \qquad \mathrm{Var}(\theta \mid n^{(n)}, y^{(n)}) = \frac{y^{(n)}(1 - y^{(n)})}{n^{(n)} + 1}$$

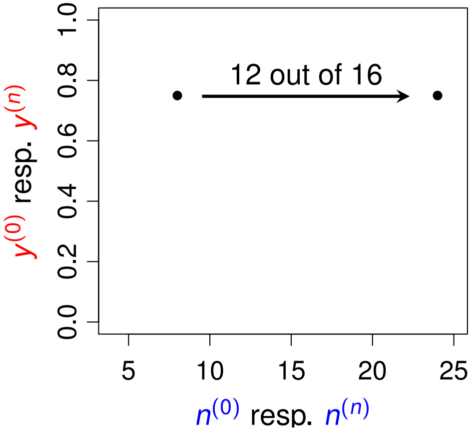
Beta-Binomial Model (BBM)



no conflict:

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

Beta-Binomial Model (BBM)

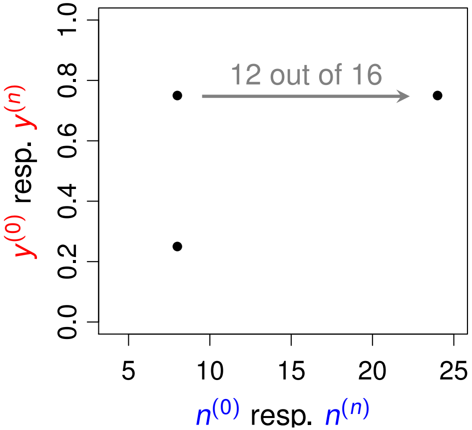


no conflict:

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:

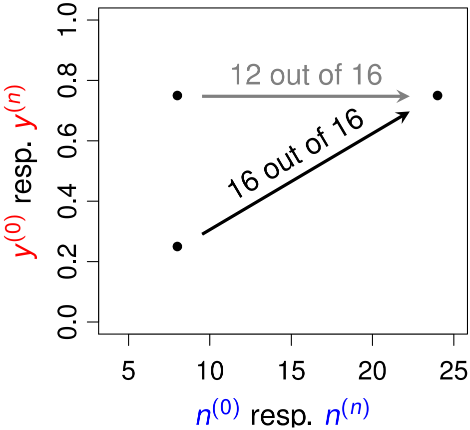
prior $n^{(0)} = 8$, $y^{(0)} = 0.75$
data $s/n = 12/16 = 0.75$

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

prior-data conflict:

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

Beta-Binomial Model (BBM)



no conflict:

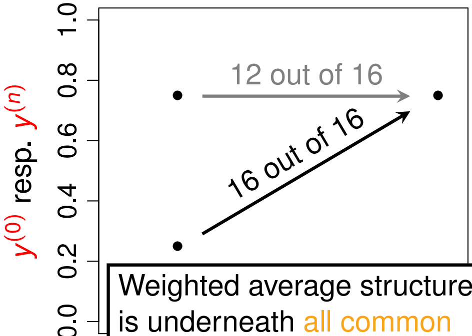
prior $n^{(0)} = 8$, $y^{(0)} = 0.75$
data $s/n = 12/16 = 0.75$

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

prior-data conflict:

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

Beta-Binomial Model (BBM)



Weighted average structure is underneath all common conjugate priors for exponential family sampling distributions!

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 $s/n = 16/16 = 1$

Why Generalise Bayesian Inference?

Bayesian theory lacks the ability to specifiy the degree of uncertainty in probability statements encoded in a (prior, posterior) distribution.

Why Generalise Bayesian Inference?

Bayesian theory lacks the ability to specifiy the degree of uncertainty in probability statements encoded in a (prior, posterior) distribution.

Variance or stretch of a distribution for describing uncertainty?

Why Generalise Bayesian Inference?

Bayesian theory lacks the ability to specifiy the degree of uncertainty in probability statements encoded in a (prior, posterior) distribution.

Variance or stretch of a distribution for describing uncertainty?

→ Does not work in the case of prior-data conflict:
In conjugate updating, the posterior variance does not depend on the degree of prior-data conflict in most cases.

→ How to express the precision of a probability statement?

Imprecision

Add imprecision as new model dimension:
Sets of priors model uncertainty in probability statements

Imprecision

Add imprecision as new model dimension:
Sets of priors model uncertainty in probability statements

Interpretation

smaller sets $\longleftrightarrow$ more precise probability statements

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

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

Bayesian Inference with Sets of Priors

Standard Bayesian inference procedure

prior + likelihood $\longrightarrow$ posterior

using Bayes' Rule

All inferences are based on the posterior

Bayesian Inference with Sets of Priors

Standard Bayesian inference procedure

prior + likelihood ➡ posterior

using Bayes' Rule

All inferences are based on the posterior

Generalised Bayesian inference procedure

set of priors + likelihood ➡ set of posteriors

Coherence (consistency of inferences) ensured by using

Generalised Bayes' Rule (GBR, Walley 1991)

= element-wise application of Bayes' Rule

All inferences are based on the set of posteriors

Bayesian Inference with Sets of Priors

Standard Bayesian inference procedure

prior + likelihood ➡ posterior

using Bayes' Rule

All inferences are based on the posterior

Generalised Bayesian inference procedure

set of priors + likelihood ➡ set of posteriors

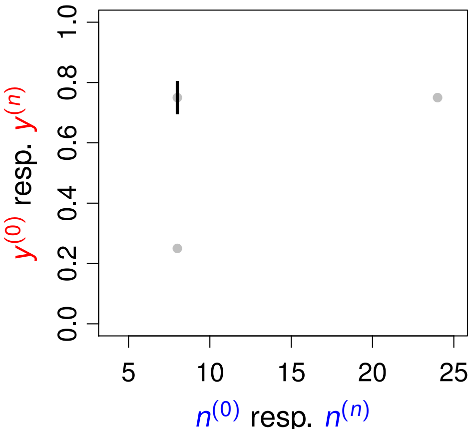
Coherence (consistency of inferences) ensured by using *Generalised Bayes' Rule* (GBR, Walley 1991)

= element-wise application of Bayes' Rule

All inferences are based on the set of posteriors

Let hyperparameters $(n^{(0)}, y^{(0)})$ vary in a set ➡ set of priors

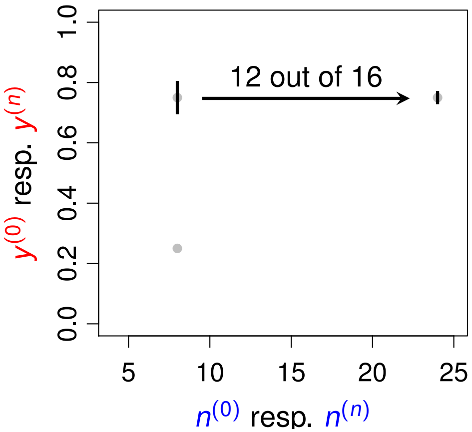
Imprecise BBM with $n^{(0)}$ fixed: IDM (Walley 1996) Quaghebeur & de Cooman (2005)



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) Quaghebeur & de Cooman (2005)

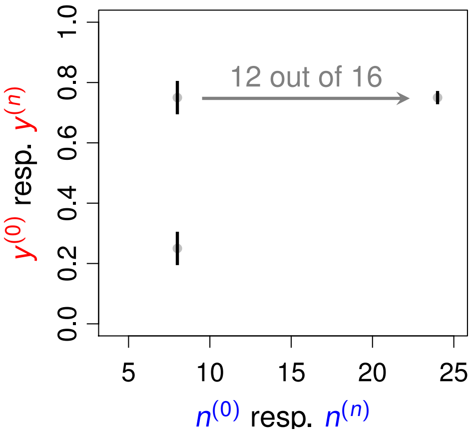


no conflict:

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) Quaghebeur & de Cooman (2005)



no conflict:

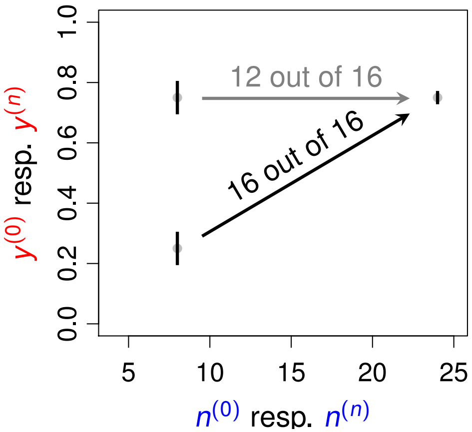
prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

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

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) Quaghebeur & de Cooman (2005)



no conflict:

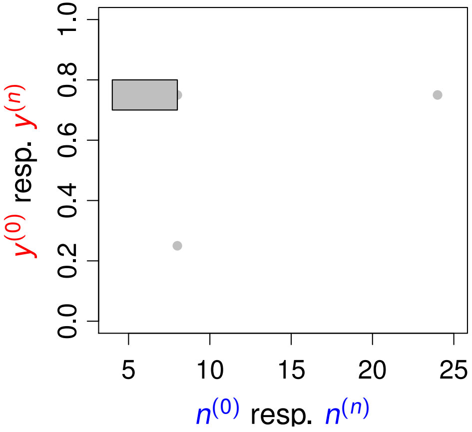
prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

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

prior-data conflict:

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

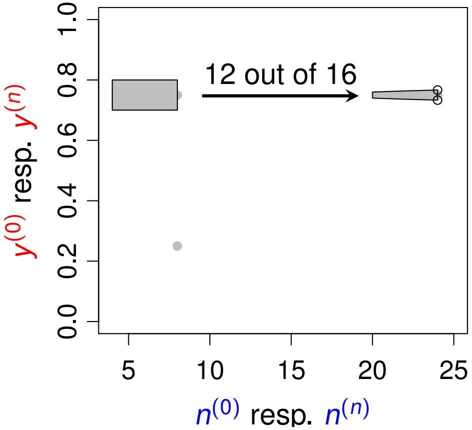
Imprecise BBM with $[\underline{n}^{(0)}, \overline{n}^{(0)}]$: Walley (1991, §5.4.3) Walter & Augustin (2009)



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 $[\underline{n}^{(0)}, \overline{n}^{(0)}]$: Walley (1991, §5.4.3) Walter & Augustin (2009)

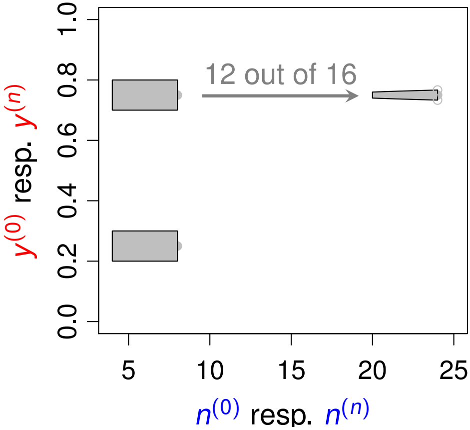


no conflict:

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 $[\underline{n}^{(0)}, \overline{n}^{(0)}]$: Walley (1991, §5.4.3) Walter & Augustin (2009)



no conflict:

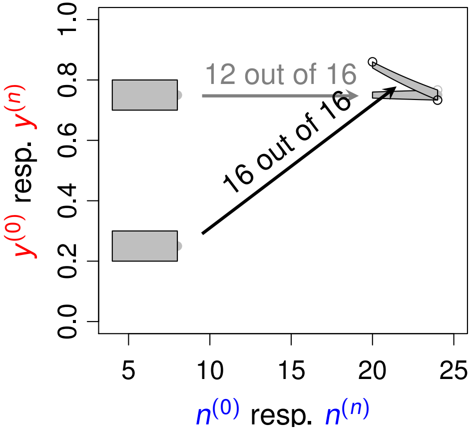
prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

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

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 $[\underline{n}^{(0)}, \overline{n}^{(0)}]$: Walley (1991, §5.4.3) Walter & Augustin (2009)



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 $s/n = 16/16 = 1$

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

Model Discussion

- Easy to handle, generally favourable inference properties, e.g.: $n \to \infty$

Model Discussion

- Easy to handle, generally favourable inference properties, e.g.:
  $n \to \infty$ ➡ $y^{(n)}$ stretch in $\to 0$

Model Discussion

- Easy to handle, generally favourable inference properties, e.g.:
  $n \to \infty$ ➡ $y^{(n)}$ stretch in $\to 0$ ➡ precise inferences

Model Discussion

- Easy to handle, generally favourable inference properties, e.g.:
  $n \to \infty$ $\Longrightarrow$ $y^{(n)}$ stretch in $\to 0$ $\Longrightarrow$ precise inferences
- Set shape is crucial modeling choice:
  trade-off between model complexity and model behaviour

Model Discussion

- Easy to handle, generally favourable inference properties, e.g.:
  $n \to \infty$ ➡ $y^{(n)}$ stretch in $\to 0$ ➡ precise inferences
- 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; Quaghebeur & 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

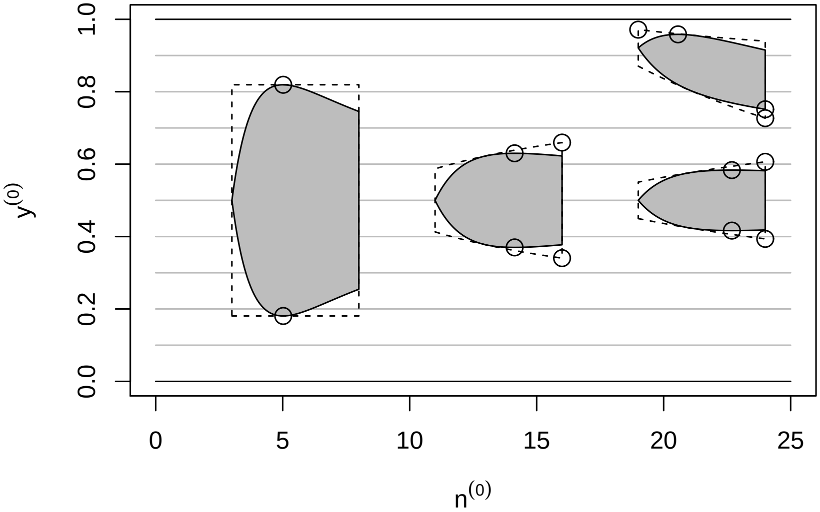
Model Discussion

- Easy to handle, generally favourable inference properties, e.g.:
  $n \to \infty$ ➡ $y^{(n)}$ stretch in $\to 0$ ➡ precise inferences
- 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; Quaghebeur & 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 & Augustin 2009):
  have non-trivial forms (banana / spotlight), but prior-data conflict sensitivity and closed form for min / max $y^{(n)}$ over
  implementation: **R** package `luck`

Model Discussion

- Easy to handle, generally favourable inference properties, e.g.:
  $n \to \infty$ ➡ $y^{(n)}$ stretch in $\to 0$ ➡ precise inferences
- 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; Quaghebeur & 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 & Augustin 2009):
  have non-trivial forms (banana / spotlight), but prior-data conflict sensitivity and closed form for min / max $y^{(n)}$ over
  implementation: **R** package `luck`
- Other set shapes are possible, but may be more difficult to handle

Parameter Set Shape for Strong Prior-Data Agreement



Common-Cause Failures

Source: Wikimedia Commons, http://commons.wikimedia.org/wiki/File:Fukushima_I_by_Digital_Globe.jpg

Common-Cause Failures

common-cause failure

simultaneous failure of several redundant components
due to a common or shared root cause (Høyland & Rausand, 1994)

Source: Wikimedia Commons, `http://commons.wikimedia.org/wiki/File:Fukushima_I_by_Digital_Globe.jpg`

Common-Cause Failures

common-cause failure

simultaneous failure of several redundant components
due to a common or shared root cause (Høyland & Rausand, 1994)

Source: `http://www.diakont.com/solutions/nuclear-energy/plant-systems/diesel-generator-control-systems/`

Source: Wikimedia Commons, `http://commons.wikimedia.org/wiki/File:Fukushima_I_by_Digital_Globe.jpg`

Conclusion

- 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

Conclusion

- 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
  - Hyperparameter set shape is important
  - Reasonable choice: *rectangular* $= [\underline{n}^{(0)}, \overline{n}^{(0)}] \times [\underline{y}^{(0)}, \overline{y}^{(0)}]$
    (Walter & Augustin 2009: *generalised iLUCK-models*, `luck`)
  - Bounds for hyperparameters 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
  - Shape for more precision in case of strong prior-data agreement is in development (joint work with Frank Coolen and Mik Bickis)

References

Evans, M. and H. Moshonov (2006). "Checking for Prior-Data Conflict". In: *Bayesian Analysis* 1, pp. 893–914.

Høyland, Arnljot and Marvin Rausand (1994). *System reliability theory: models and statistical methods*. A Wiley interscience publication. New York, NY: Wiley. ISBN: 0-471-59397-4.

Quaeghebeur, E. and G. de Cooman (2005). "Imprecise probability models for inference in exponential families". In: *ISIPTA '05*. Ed. by F. Cozman, R. Nau, and T. Seidenfeld. Manno: SIPTA, pp. 287–296.

Troffaes, Matthias, Gero Walter, and Dana Kelly (2013). *A Robust Bayesian Approach to Modelling Epistemic Uncertainty in Common-Cause Failure Models*. Preprint available at http://arxiv.org/abs/1301.0533. Accepted for publication at: Reliability Engineering & System Safety.

Walley, Peter (1991). *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall. ISBN: 0-412-28660-2.

Walley, Peter (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.

Walter, Gero and Thomas Augustin (2009). "Imprecision and Prior-data Conflict in Generalized Bayesian Inference". In: *Journal of Statistical Theory and Practice* 3, pp. 255–271.

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