Sets of Priors
Reflecting Prior-Data Conflict and Agreement

Gero Walter[1], Frank Coolen[2]

[1]Eindhoven University of Technology, Eindhoven, NL
[2]Durham University, Durham, UK

g.m.walter@tue.nl

IPMU Eindhoven 2016-06-21

Bayesian Inference

expert info $+$ data $\rightarrow$ 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

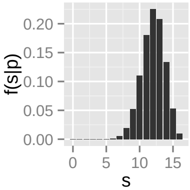
Bayesian Inference

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

Binomial distribution

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



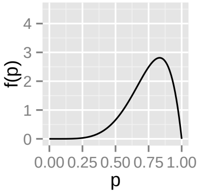
Bayesian Inference

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

► Bayes' Rule

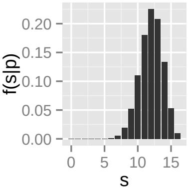
Beta prior

$$p \sim \mathrm{Beta}(\alpha^{(0)}, \beta^{(0)})$$



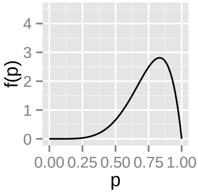
Binomial distribution

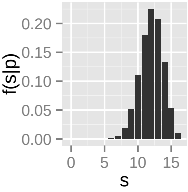
$$s \mid p \sim \mathrm{Binomial}(n, p)$$

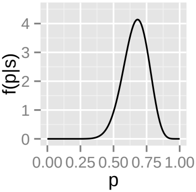


Bayesian Inference

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

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

- 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

What if expert information and data tell different stories?

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

Prior-Data Conflict

What if expert information and data tell different stories?

- 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

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

Prior-Data Conflict

What if expert information and data tell different stories?

- 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

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

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

Prior-Data Conflict

What if expert information and data tell different stories?

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

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

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

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

Prior-Data Conflict

What if expert information and data tell different stories?

- 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 $\hat{p}$

Prior-Data Conflict

What if expert information and data tell different stories?

- 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 $\hat{p}$

Prior-Data Conflict

What if expert information and data tell different stories?

- 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 $\hat{p}$!

Prior-Data Conflict

What if expert information and data tell different stories?

- 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

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

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

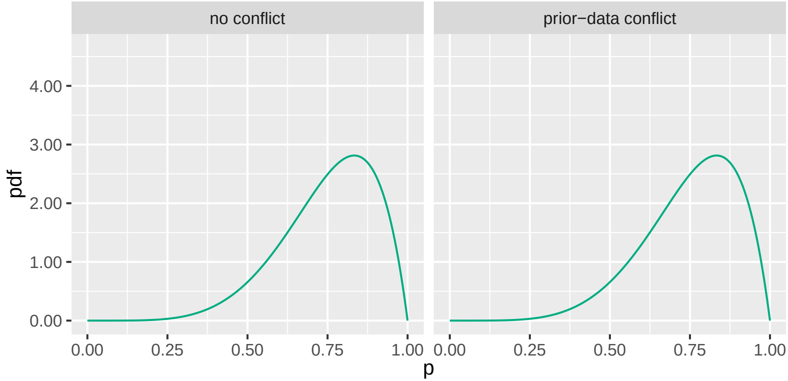
ML estimator $\hat{p}$

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

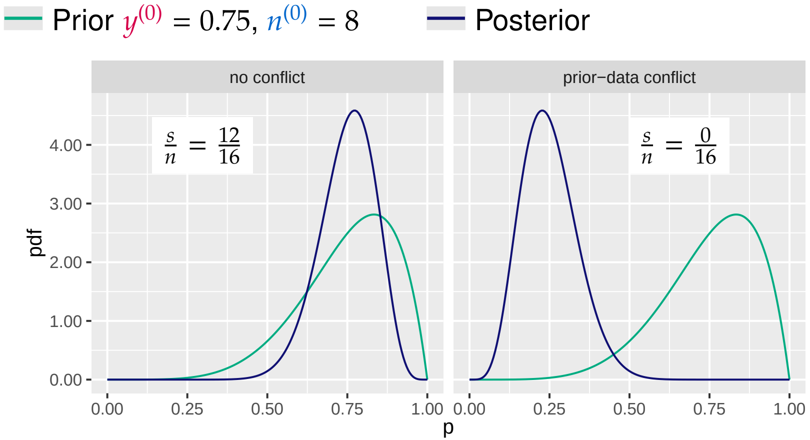
$$\mathrm{Var}[p \mid s] = \frac{y^{(n)}(1 - y^{(n)})}{n^{(n)} + 1}$$ decreases with n!

Prior-Data Conflict

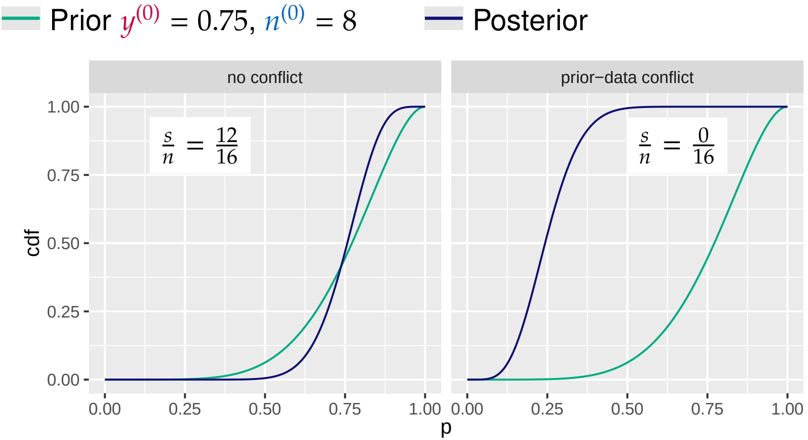
Prior $y^{(0)} = 0.75$, $n^{(0)} = 8$



Prior-Data Conflict

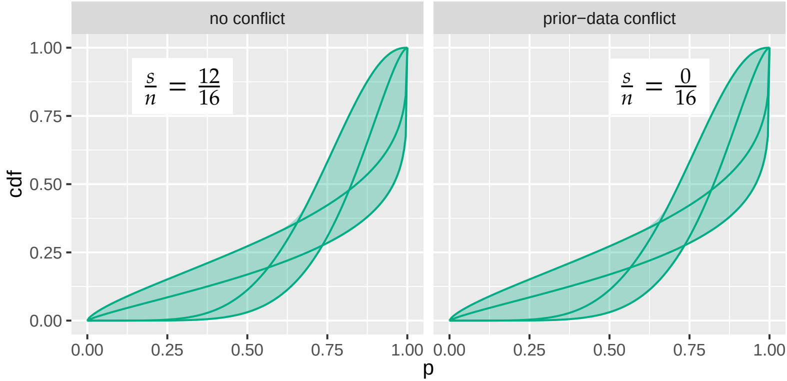


Prior-Data Conflict

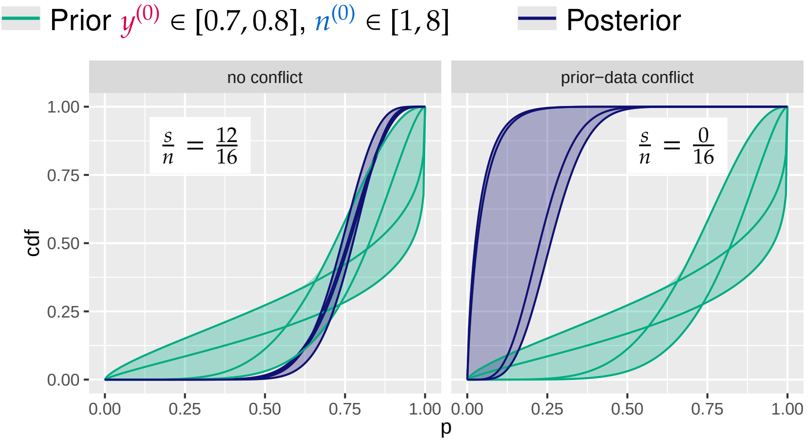


Sets of Priors

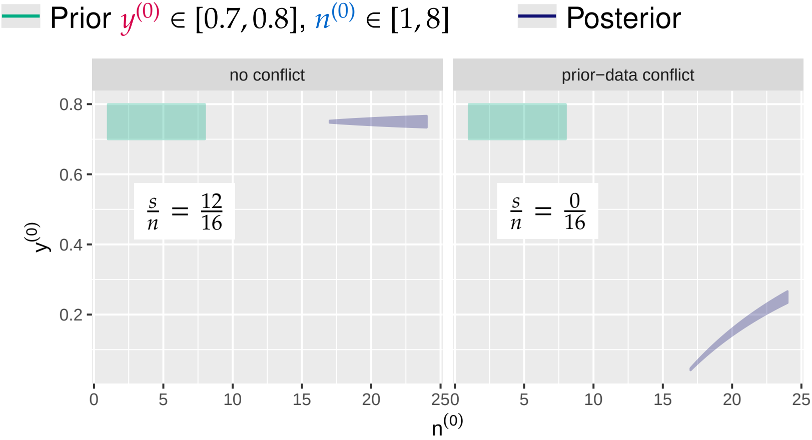
- Prior $y^{(0)} \in [0.7, 0.8]$, $n^{(0)} \in [1, 8]$



Sets of Priors



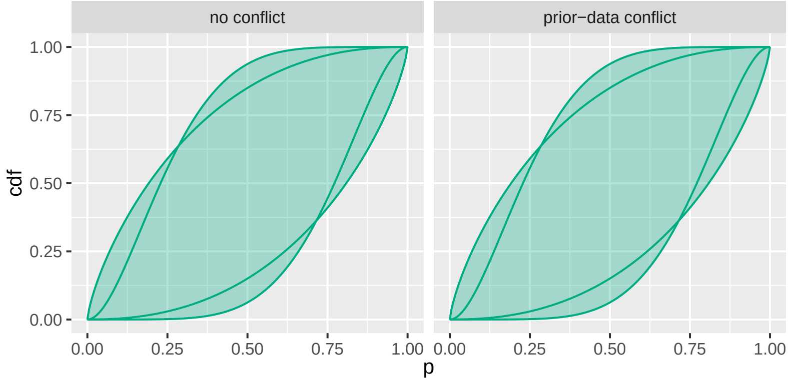
Rectangular Prior Parameter Set



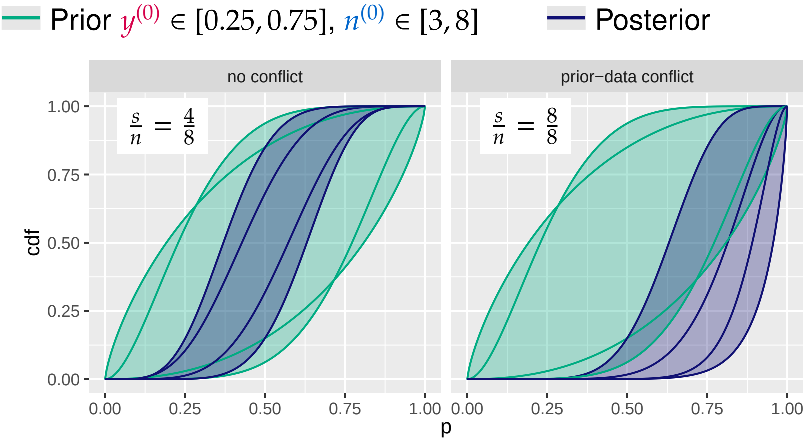
(Walter and Augustin 2009; Walter 2013)

Vague Prior Information

Prior $y^{(0)} \in [0.25, 0.75]$, $n^{(0)} \in [3, 8]$

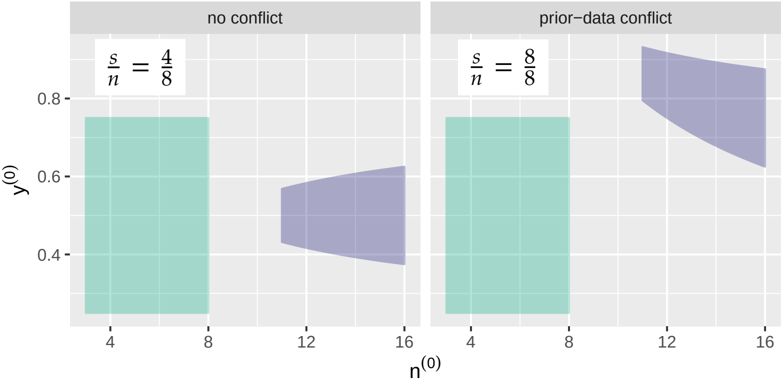


Vague Prior Information



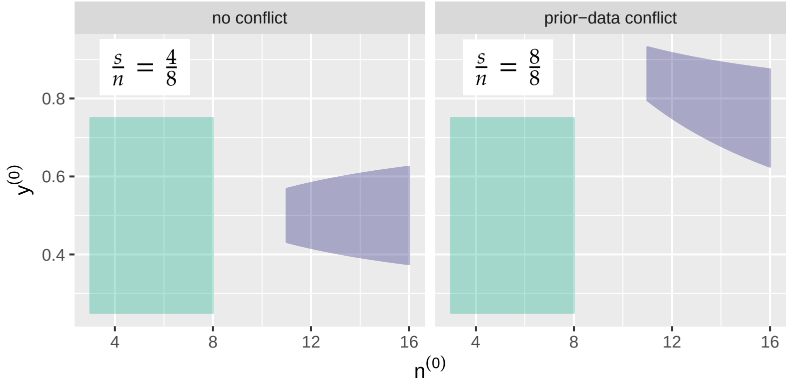
Vague Prior Information

Prior $y^{(0)} \in [0.25, 0.75]$, $n^{(0)} \in [3, 8]$ Posterior



Vague Prior Information

Prior $y^{(0)} \in [0.25, 0.75]$, $n^{(0)} \in [3, 8]$ Posterior



$$n^{(0)} \mapsto n^{(0)} + n, \qquad y^{(0)} \mapsto y^{(0)} + \frac{s - n y^{(0)}}{n^{(0)} + n}$$

New Parametrisation

Bickis (2015): use parameters $(\eta_0^{(0)}, \eta_1^{(0)})$ defined as

$$\eta_0^{(0)} = n^{(0)} - 2\,, \qquad \eta_1^{(0)} = n^{(0)}(y^{(0)} - \frac{1}{2})$$

Then the Bayesian update corresponds to

$$\eta_0^{(0)} \mapsto \eta_0^{(0)} + n\,, \qquad \eta_1^{(0)} \mapsto \eta_1^{(0)} + s - \frac{n}{2}\,.$$

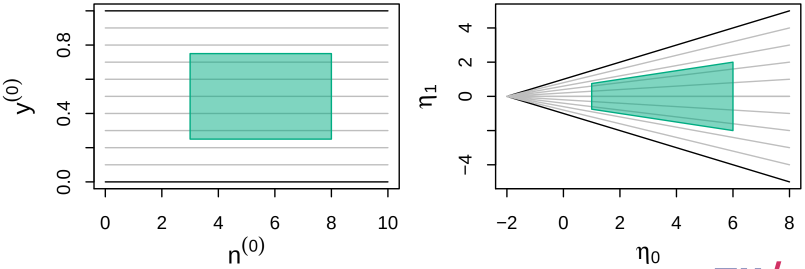
New Parametrisation

Bickis (2015): use parameters $(\eta_0^{(0)}, \eta_1^{(0)})$ defined as

$$\eta_0^{(0)} = n^{(0)} - 2\,, \qquad \eta_1^{(0)} = n^{(0)}\Big(y^{(0)} - \frac{1}{2}\Big)$$

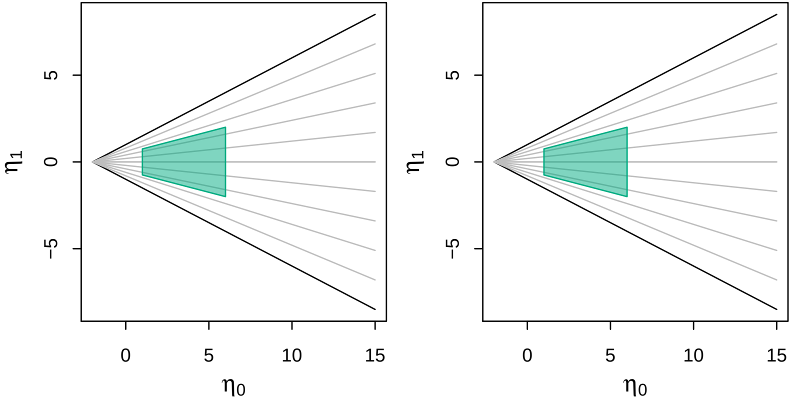
Then the Bayesian update corresponds to

$$\eta_0^{(0)} \mapsto \eta_0^{(0)} + n\,, \qquad \eta_1^{(0)} \mapsto \eta_1^{(0)} + s - \frac{n}{2}\,.$$



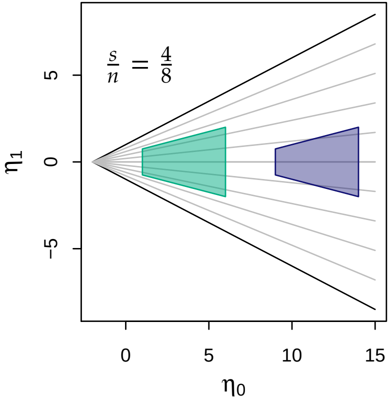
New Parametrisation: Update

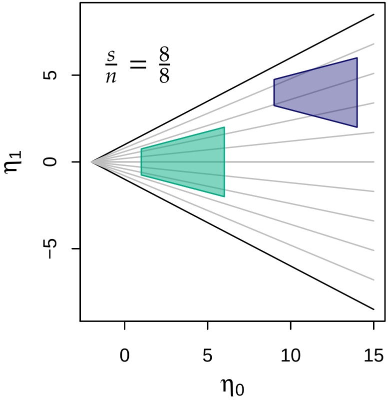
Prior $y^{(0)} \in [0.25, 0.75]$, $n^{(0)} \in [3, 8]$



New Parametrisation: Update

Prior $y^{(0)} \in [0.25, 0.75]$, $n^{(0)} \in [3, 8]$ Posterior

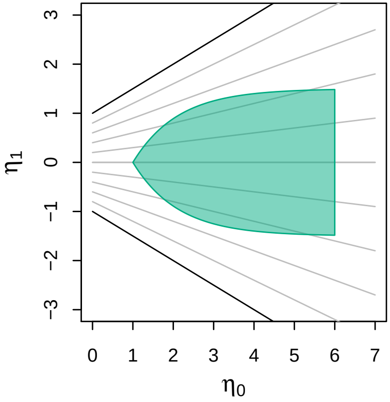


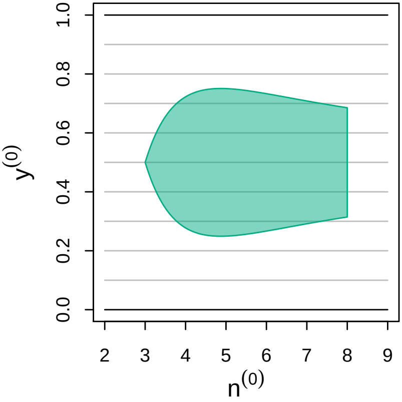


$$\eta_0^{(0)} \mapsto \eta_0^{(0)} + n, \qquad \eta_1^{(0)} \mapsto \eta_1^{(0)} + s - \frac{n}{2}$$

New Parameter Set Shape: Boatshape

Prior $y^{(0)} \in [0.25, 0.75]$, $n^{(0)} \in [3, 8]$, $y_c = 0.5$

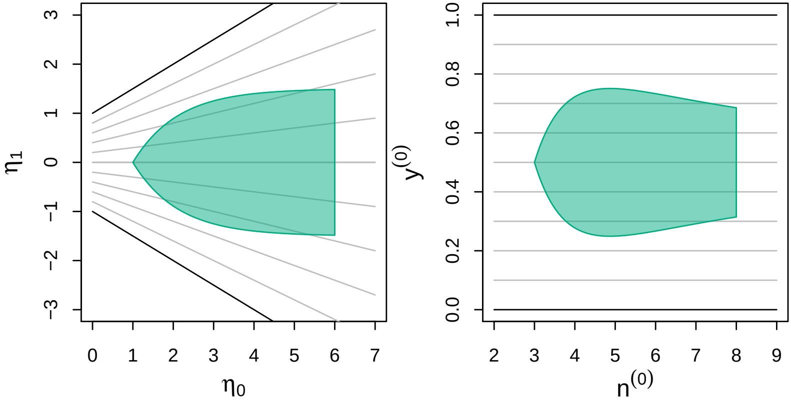




Technische Universiteit Eindhoven University of Technology

New Parameter Set Shape: Boatshape

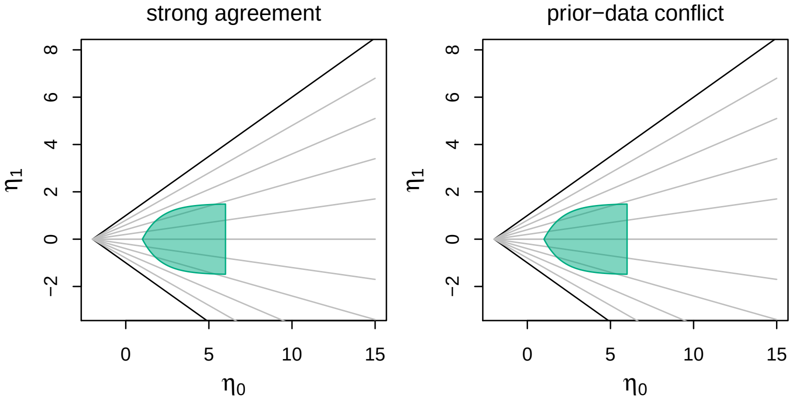
- Prior $y^{(0)} \in [0.25, 0.75]$, $n^{(0)} \in [3, 8]$, $y_c = 0.5$



$$\overline{c}^{(0)}(\eta_0) = a\left(1 - e^{-b(\eta_0 - \underline{\eta}_0)}\right) \text{ and } \underline{c}^{(0)}(\eta_0) = -a\left(1 - e^{-b(\eta_0 - \underline{\eta}_0)}\right)$$

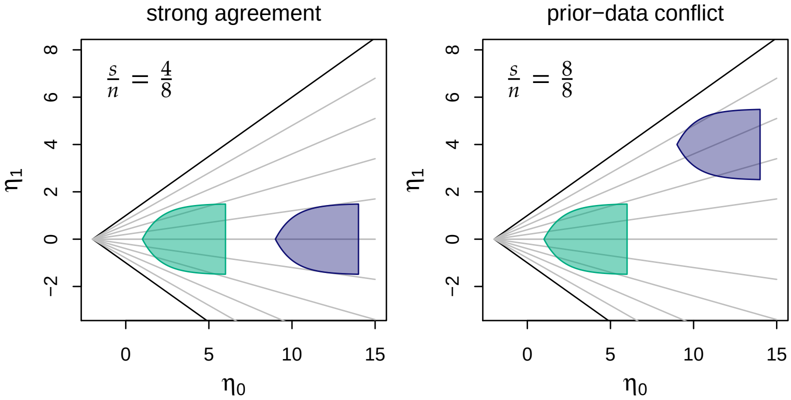
Boatshape: Update

- Prior $y^{(0)} \in [0.25, 0.75]$, $n^{(0)} \in [3, 8]$, $y_c = 0.5$

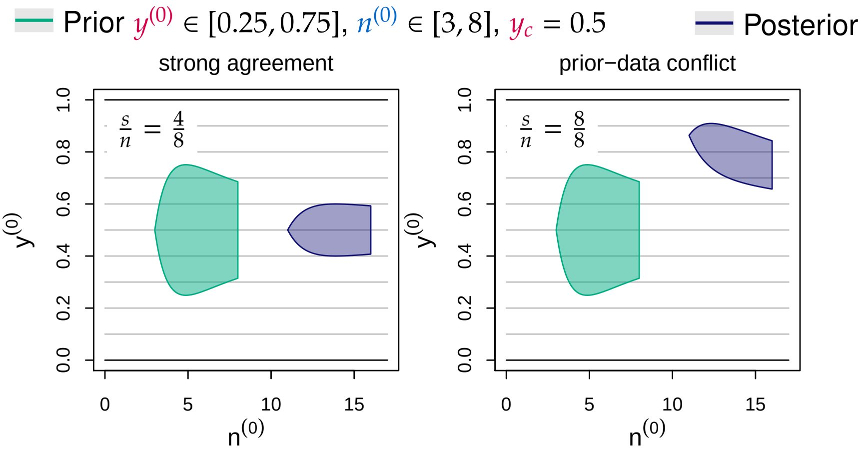


Boatshape: Update

Prior $y^{(0)} \in [0.25, 0.75]$, $n^{(0)} \in [3, 8]$, $y_c = 0.5$ Posterior

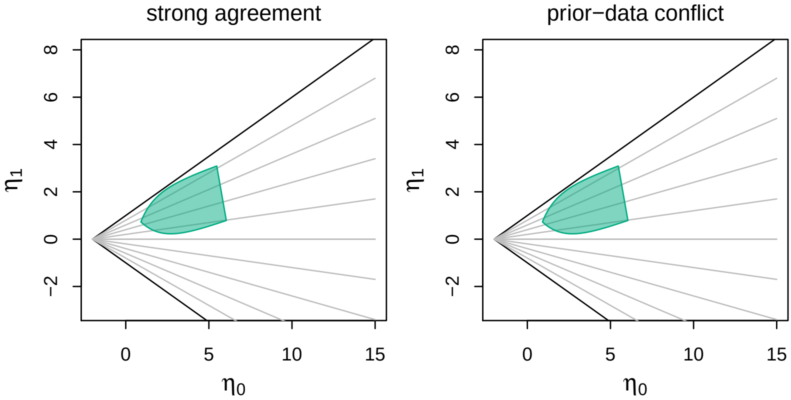


Boatshape: Update



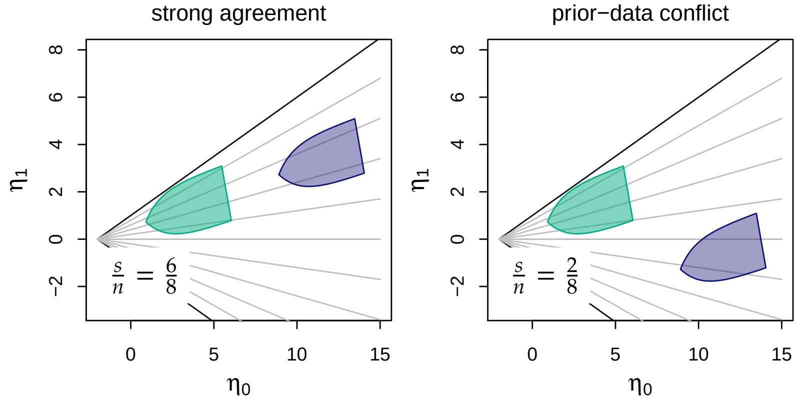
Boatshape: Update

Prior $y^{(0)} \in [0.55, 0.97]$, $n^{(0)} \in [3, 8]$, $y_c = 0.75$

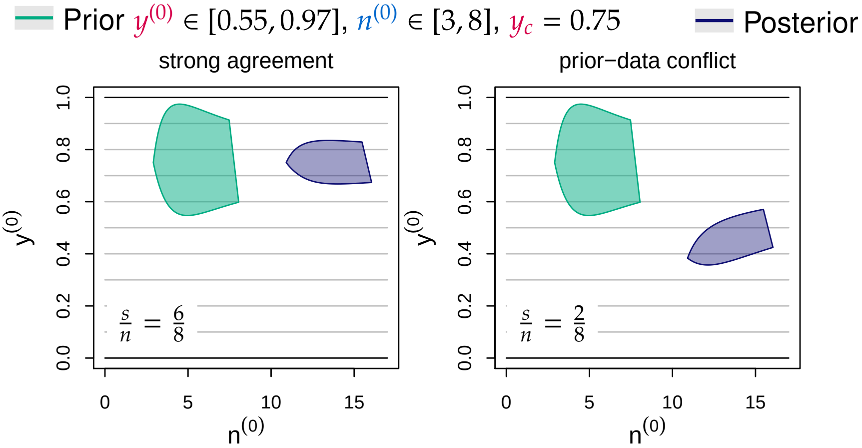


Boatshape: Update

Prior $y^{(0)} \in [0.55, 0.97]$, $n^{(0)} \in [3, 8]$, $y_c = 0.75$ Posterior

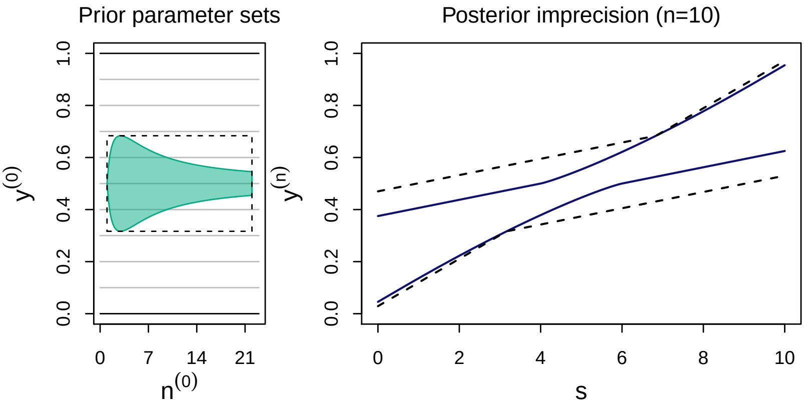


Boatshape: Update



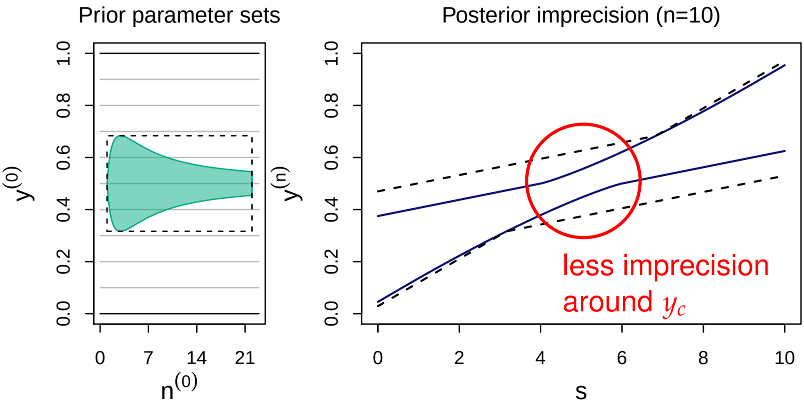
Strong Prior-Data Agreement Property

Prior $y^{(0)} \in [0.32, 0.68]$, $n^{(0)} \in [1, 22]$, $y_c = 0.5$ Posterior



Strong Prior-Data Agreement Property

Prior $y^{(0)} \in [0.32, 0.68]$, $n^{(0)} \in [1, 22]$, $y_c = 0.5$ Posterior



Summary & Outlook

Summary:

- Rectangular prior parameter sets give extra imprecision for data in conflict with the prior
- New parametrisation with purely data-dependent update shift
- Boatshape sets also give less imprecision for data exactly in line with the prior

Summary & Outlook

Summary:

- Rectangular prior parameter sets give extra imprecision for data in conflict with the prior
- New parametrisation with purely data-dependent update shift
- Boatshape sets also give less imprecision for data exactly in line with the prior

Outlook:

- Elicitation via pre-posterior analysis
- Parametrisation can be constructed for any distribution from exponential family
- Other inference properties via tailored set shape

References

Bickis, Miķelis G. (2015). “The geometry of imprecise inference”. In: *ISIPTA ’15: Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications*. Ed. by T. Augustin et al. SIPTA, pp. 47–56. URL: http://www.sipta.org/isipta15/data/paper/31.pdf.

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

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

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. DOI: 10.1080/15598608.2009.10411924.

Walter, Gero and Frank P.A. Coolen (2016). “Sets of Priors Reflecting Prior-Data Conflict and Agreement”. In: *IPMU ’16: Proceedings of the Sixteenth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Part I*. Ed. by Joao Paulo Carvalho et al. Springer, pp. 153–164.