Sets of canonical parameters in imprecise Bayesian inference with conjugate priors

Gero Walter

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

November 1st, 2011

Introduction

- Bernoulli observations: 0/1 observations (success yes/no, head/tails when tossing a coin)

Introduction

- Bernoulli observations: 0/1 observations
  (success yes/no, head/tails when tossing a coin)
- given: a set of observations (12 out of 16 tosses were heads)

Introduction

- Bernoulli observations: 0/1 observations
  (success yes/no, head/tails when tossing a coin)
- given: a set of observations (12 out of 16 tosses were heads)
- additional to observations, we have strong prior information
  (we are convinced that $P(\text{heads})$ should be around 0.75)

Introduction

- Bernoulli observations: 0/1 observations
  (success yes/no, head/tails when tossing a coin)
- given: a set of observations (12 out of 16 tosses were heads)
- additional to observations, we have strong prior information
  (we are convinced that P(heads) should be around 0.75)
- interested in certain inferences,
  e.g. probability P that the next observation is a head.

Introduction

- Bernoulli observations: 0/1 observations (success yes/no, head/tails when tossing a coin)
- given: a set of observations (12 out of 16 tosses were heads)
- additional to observations, we have strong prior information (we are convinced that P(heads) should be around 0.75)
- interested in certain inferences, e.g. probability P that the next observation is a head.
- **prior-data conflict:** if P(heads) for the coin is actually very different from our prior guess (i.e., prior information and data are in conflict), this should show up in the inferences (probability P and, e.g., confidence intervals)

Beta-Bernoulli/Binomial Model (BBM)

- one standard statistical model for this situation:
  Beta-Bernoulli/Binomial Model
- here in parameterization used, e.g., by Walley:

Beta-Bernoulli/Binomial Model (BBM)

- one standard statistical model for this situation: Beta-Bernoulli/Binomial Model
- here in parameterization used, e.g., by Walley:

Data :	s	$\sim$	$\text{Binom}(p, n)$
conjugate prior:	p	$\sim$	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$p \mid s$	$\sim$	$\text{Beta}(n^{(n)}, y^{(n)})$

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

Beta-Bernoulli/Binomial Model (BBM)

- one standard statistical model for this situation: Beta-Bernoulli/Binomial Model
- here in parameterization used, e.g., by Walley:

Data :	s	$\sim$	$\text{Binom}(p, n)$
conjugate prior:	p	$\sim$	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$p \mid s$	$\sim$	$\text{Beta}(n^{(n)}, y^{(n)})$

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

Beta-Bernoulli/Binomial Model (BBM)

- one standard statistical model for this situation: Beta-Bernoulli/Binomial Model
- here in parameterization used, e.g., by Walley:

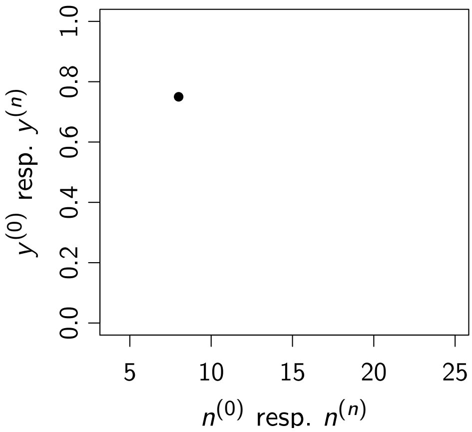
Data :	s	$\sim$	$\text{Binom}(p, n)$
conjugate prior:	p	$\sim$	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$p \mid s$	$\sim$	$\text{Beta}(n^{(n)}, y^{(n)})$

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

$$\text{Var}(p \mid s) = \frac{y^{(n)}(1 - y^{(n)})}{n^{(n)} + 1}$$

➡ no reaction to prior-data conflict!

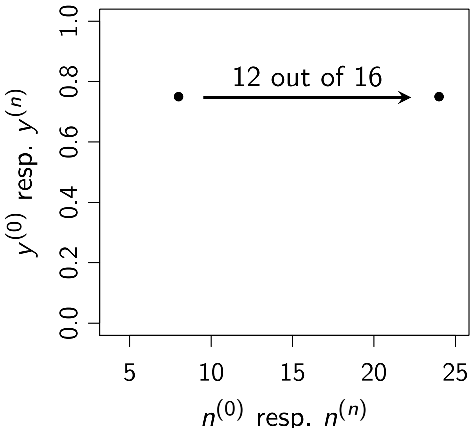
Beta-Bernoulli/Binomial Model (BBM)



no conflict:

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

Beta-Bernoulli/Binomial Model (BBM)

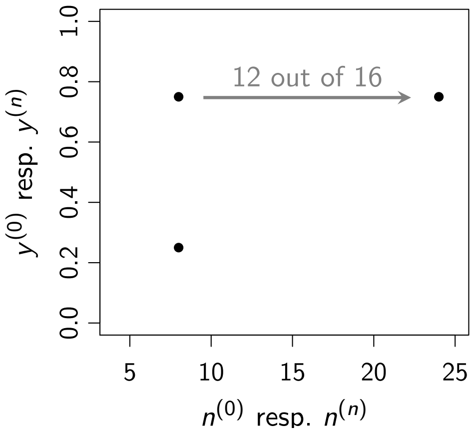


no conflict:

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

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

Beta-Bernoulli/Binomial Model (BBM)



no conflict:

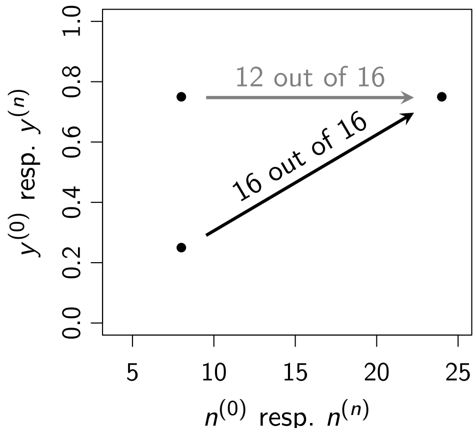
prior $n^{(0)} = 8$, $y^{(0)} = 0.75$
data $s/n = 12/16 = 0.75$

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

prior-data conflict:

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

Beta-Bernoulli/Binomial Model (BBM)



no conflict:

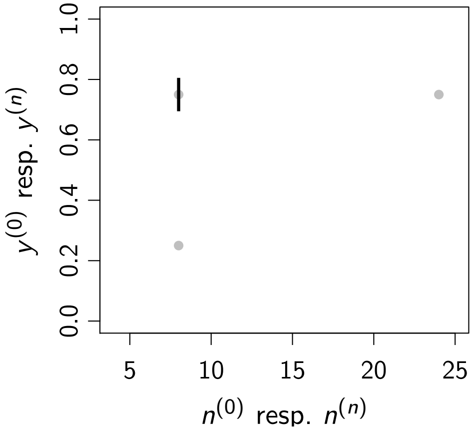
prior $n^{(0)} = 8$, $y^{(0)} = 0.75$
data $s/n = 12/16 = 0.75$

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

prior-data conflict:

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

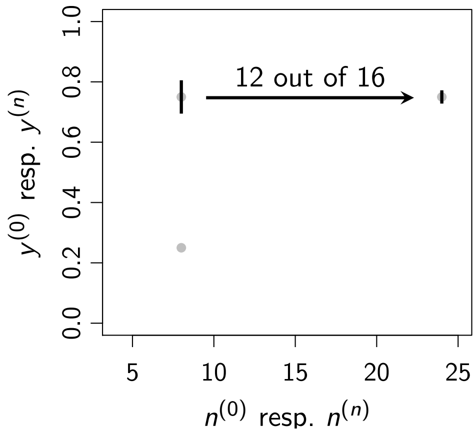
Imprecise BBM (IBBM) $\hat{=}$ IDM with prior information



no conflict:

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

Imprecise BBM (IBBM) $\hat{=}$ IDM with prior information

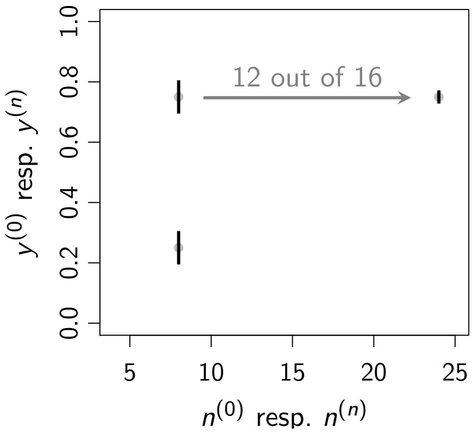


no conflict:

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 (IBBM) $\hat{=}$ IDM with prior information



no conflict:

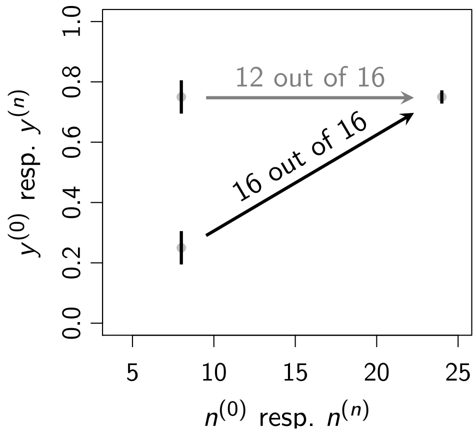
prior $n^{(0)} = 8,\ y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

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

prior-data conflict:

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

Imprecise BBM (IBBM) $\hat{=}$ IDM with prior information



no conflict:

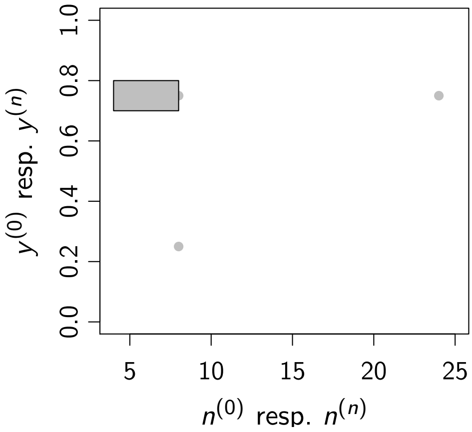
prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

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

prior-data conflict:

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

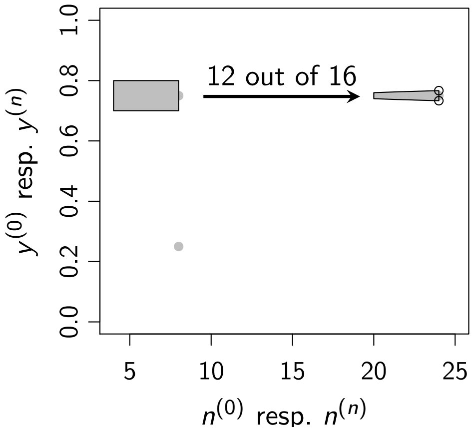
pdc-Imprecise BBM (pdc-IBBM): Walley 1991, Ch.5.4.3



no conflict:

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

pdc-Imprecise BBM (pdc-IBBM): Walley 1991, Ch.5.4.3

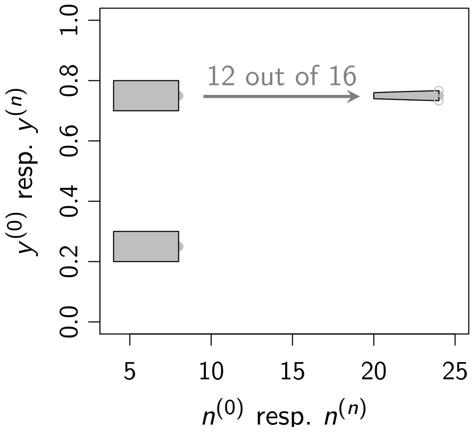


no conflict:

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

"spotlight" shape

pdc-Imprecise BBM (pdc-IBBM): Walley 1991, Ch.5.4.3



no conflict:

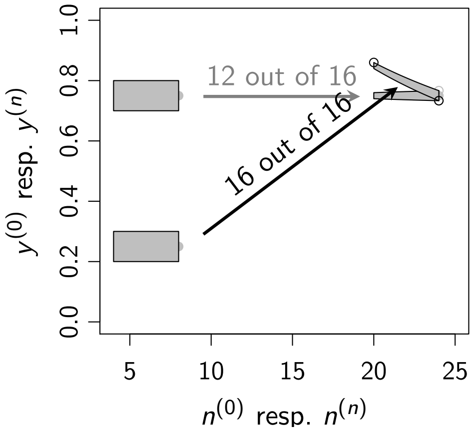
prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

"spotlight" shape

prior-data conflict:

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

pdc-Imprecise BBM (pdc-IBBM): Walley 1991, Ch.5.4.3



no conflict:

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

▼

"spotlight" shape

prior-data conflict:

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

▼

"banana" shape

pdc-IBBM – Some Characteristics and Open Ends

- reaction to prior-data conflict due to different 'updating speeds' depending on $n^{(0)}$

pdc-IBBM – Some Characteristics and Open Ends

- reaction to prior-data conflict due to different 'updating speeds' depending on $n^{(0)}$
- can be generalized to conjugate family distributions (Walter & Augustin 2009) (same weighted average structure for $y^{(n)}$)

pdc-IBBM – Some Characteristics and Open Ends

- reaction to prior-data conflict due to different 'updating speeds' depending on $n^{(0)}$
- can be generalized to conjugate family distributions (Walter & Augustin 2009) (same weighted average structure for $y^{(n)}$)
- posterior parameter sets not rectangular (but can be described by their four 'corners')

pdc-IBBM – Some Characteristics and Open Ends

- reaction to prior-data conflict due to different ‘updating speeds’ depending on $n^{(0)}$
- can be generalized to conjugate family distributions (Walter & Augustin 2009) (same weighted average structure for $y^{(n)}$)
- posterior parameter sets not rectangular
  (but can be described by their four ‘corners’)
- pdc-IBBM goes “bananas” irrespective of n or $n^{(0)}$
  i.e. also for small sample sizes and weak prior information

pdc-IBBM – Some Characteristics and Open Ends

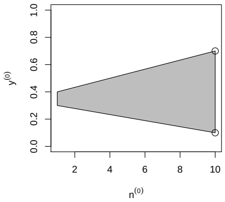
- reaction to prior-data conflict due to different 'updating speeds' depending on $n^{(0)}$
- can be generalized to conjugate family distributions (Walter & Augustin 2009) (same weighted average structure for $y^{(n)}$)
- posterior parameter sets not rectangular
  (but can be described by their four 'corners')
- pdc-IBBM goes "bananas" irrespective of n or $n^{(0)}$
  i.e. also for small sample sizes and weak prior information
- open ends:
  - why rectangles? why same $y^{(0)}$ interval at $\underline{n}^{(0)}$ and $\overline{n}^{(0)}$?
  - 'conjugate' set description possible? (invariant under updating)
  - what could be an easy elicitation procedure for parameter sets?

Allow different $y^{(0)}$ intervals at $\underline{n}^{(0)}$ and $\overline{n}^{(0)}$:

"snout left" (shorter $y^{(0)}$ interval at $\underline{n}^{(0)}$)

Encoding the same amount of prior information for different prior strengths $n^{(0)}$:

- for low $n^{(0)}$, can be precise with $y^{(0)}$
- for high $n^{(0)}$, must be cautious with $y^{(0)}$

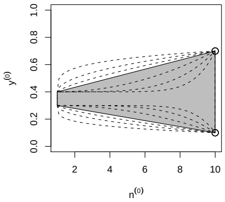


Allow different $y^{(0)}$ intervals at $\underline{n}^{(0)}$ and $\overline{n}^{(0)}$:

"snout left" (shorter $y^{(0)}$ interval at $\underline{n}^{(0)}$)

Encoding the same amount of prior information for different prior strengths $n^{(0)}$:

- for low $n^{(0)}$, can be precise with $y^{(0)}$
- for high $n^{(0)}$, must be cautious with $y^{(0)}$

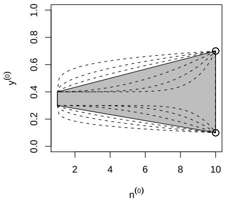


Allow different $y^{(0)}$ intervals at $\underline{n}^{(0)}$ and $\overline{n}^{(0)}$:

"snout left" (shorter $y^{(0)}$ interval at $\underline{n}^{(0)}$)

Encoding the same amount of prior information for different prior strengths $n^{(0)}$:

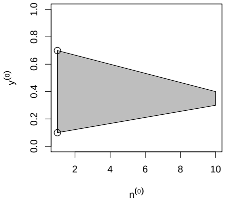
- for low $n^{(0)}$, can be precise with $y^{(0)}$
- for high $n^{(0)}$, must be cautious with $y^{(0)}$



"snout right" (shorter $y^{(0)}$ interval at $\overline{n}^{(0)}$)

Imprecision like confidence interval length for different pseudocounts $n^{(0)}$:

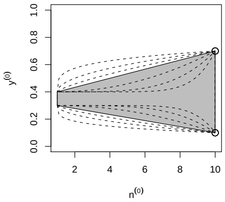
- for few $n^{(0)}$, imprecise interval for $y^{(0)}$
- for many $n^{(0)}$, precise interval for $y^{(0)}$



Allow different $y^{(0)}$ intervals at $\underline{n}^{(0)}$ and $\overline{n}^{(0)}$:

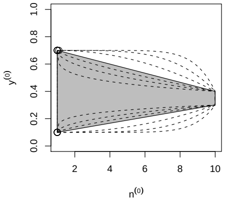
“snout left” (shorter $y^{(0)}$ interval at $\underline{n}^{(0)}$)
Encoding the same amount of prior information for different prior strengths $n^{(0)}$:

- for low $n^{(0)}$, can be precise with $y^{(0)}$
- for high $n^{(0)}$, must be cautious with $y^{(0)}$

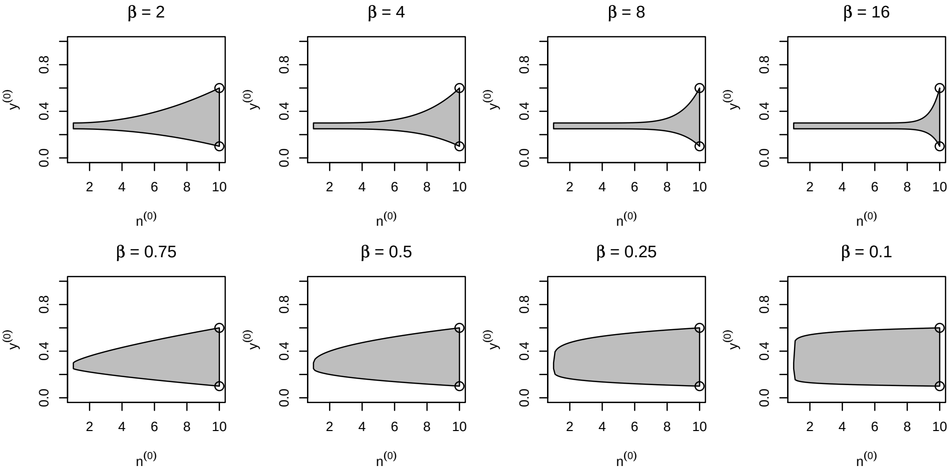


“snout right” (shorter $y^{(0)}$ interval at $\overline{n}^{(0)}$)
Imprecision like confidence interval length for different pseudocounts $n^{(0)}$:

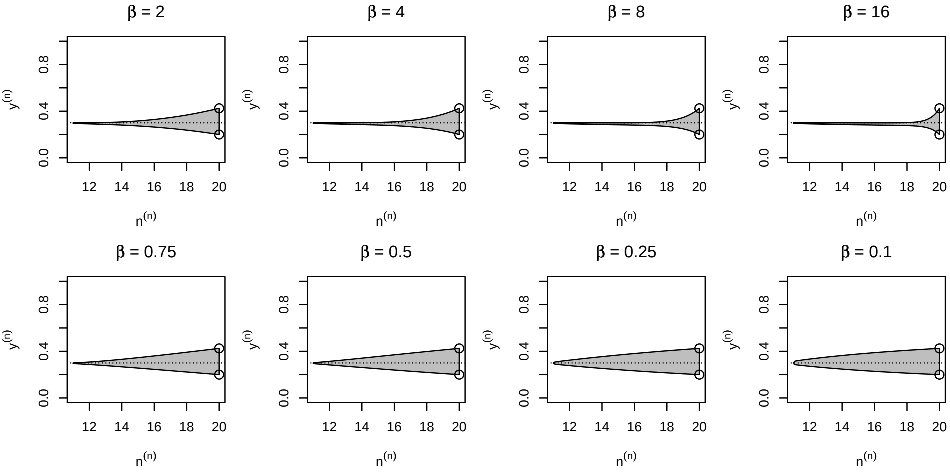
- for few $n^{(0)}$, imprecise interval for $y^{(0)}$
- for many $n^{(0)}$, precise interval for $y^{(0)}$



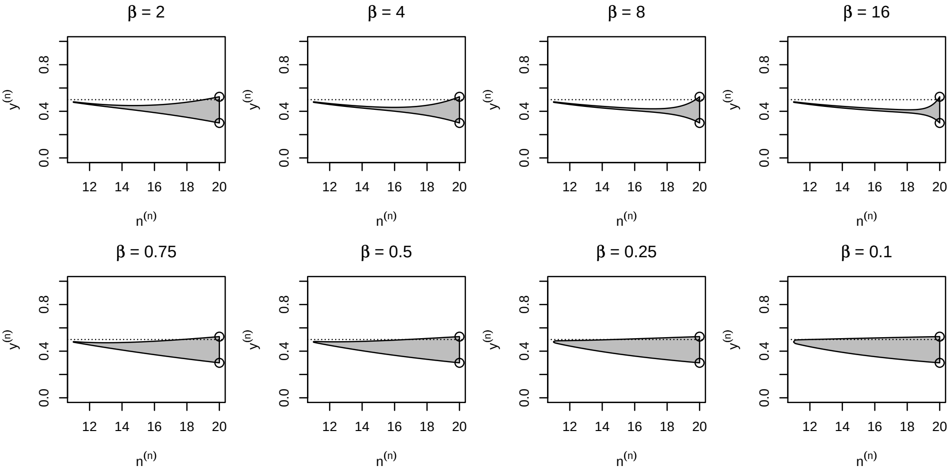
"snout left": Priors



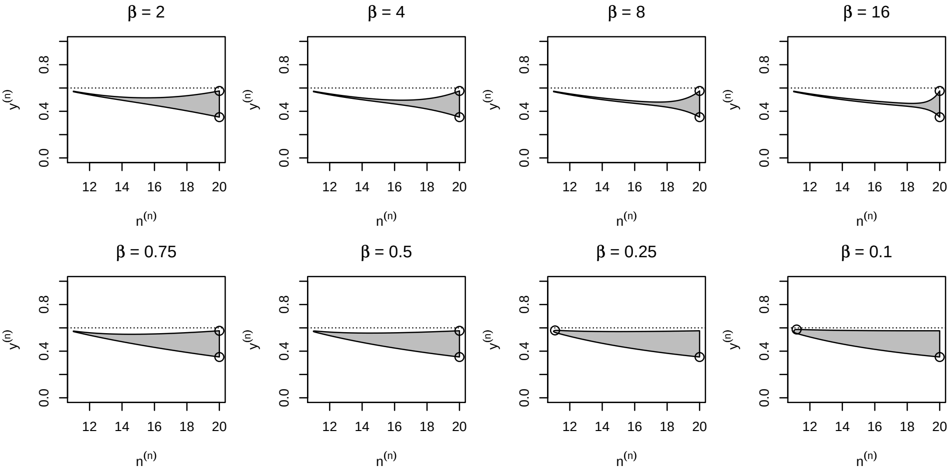
"snout left": Posteriors when $s/n = 3/10$



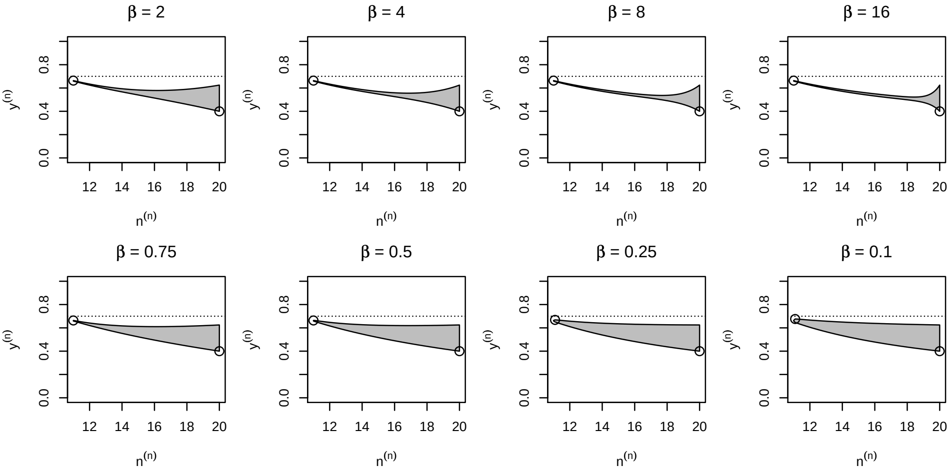
“snout left”: Posteriors when $s/n = 5/10$



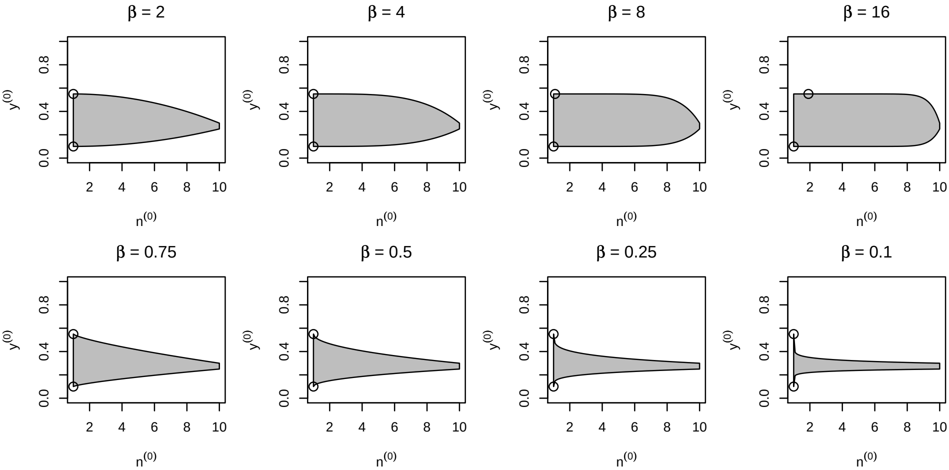
"snout left": Posteriors when $s/n = 6/10$



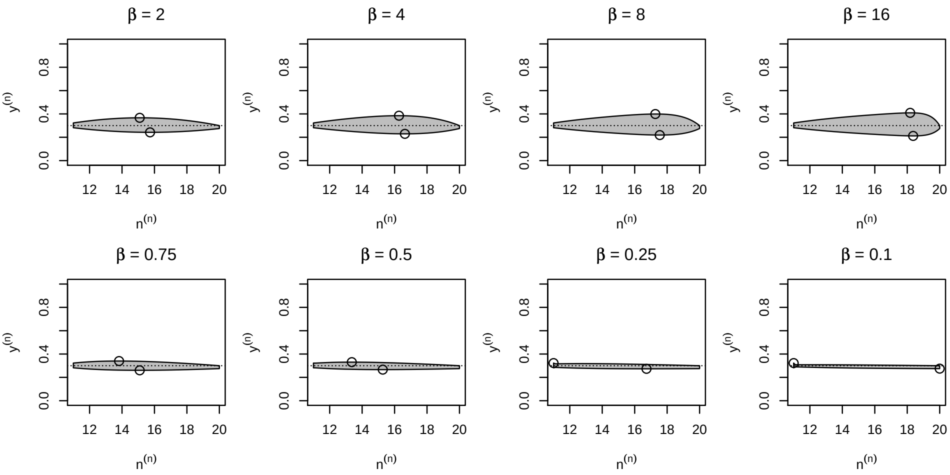
“snout left”: Posteriors when $s/n = 7/10$



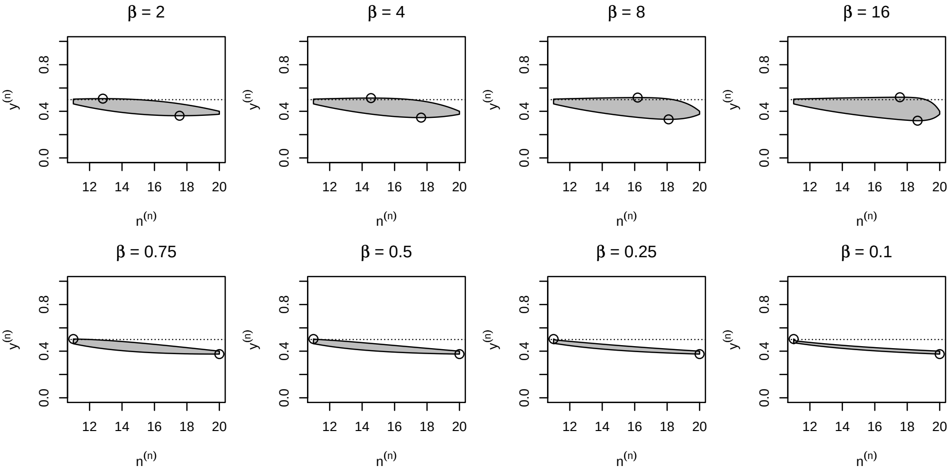
"snout right": Priors



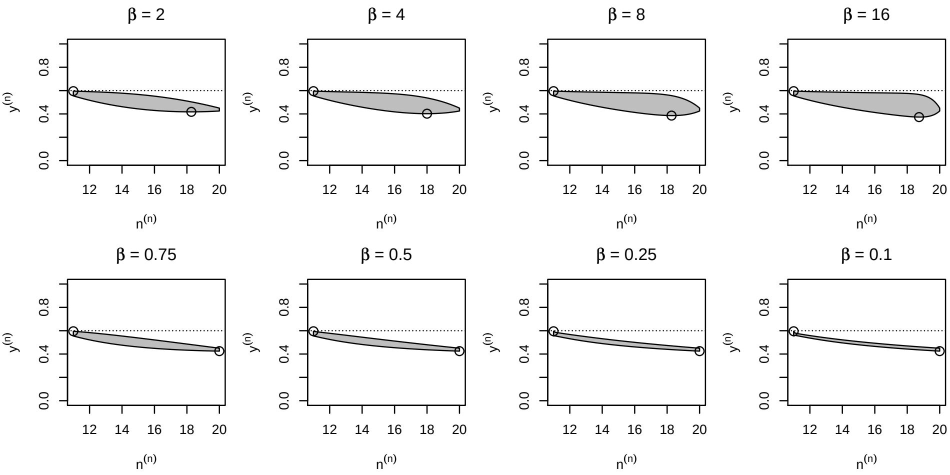
"snout right": Posteriors when $s/n = 3/10$



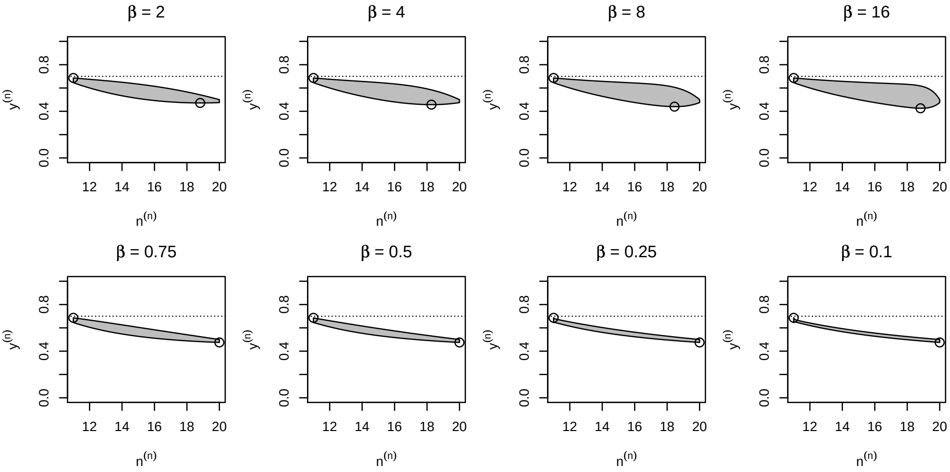
"snout right": Posteriors when $s/n = 5/10$



"snout right": Posteriors when $s/n = 6/10$



"snout right": Posteriors when $s/n = 7/10$



Shape Conjugacy?

- Such general prior shapes are difficult to elicit, do not carry over to posterior, and not all aspects of the shape might be relevant for inference(s) of interest.

Shape Conjugacy?

- Such general prior shapes are difficult to elicit, do not carry over to posterior, and not all aspects of the shape might be relevant for inference(s) of interest.
- A shape that is invariant under updating may not exist.

Shape Conjugacy?

- Such general prior shapes are difficult to elicit, do not carry over to posterior, and not all aspects of the shape might be relevant for inference(s) of interest.
- A shape that is invariant under updating may not exist.
- Sets of points or vertical lines are invariant shape *descriptions*, but may be impractical.

Shape Conjugacy?

- Such general prior shapes are difficult to elicit, do not carry over to posterior, and not all aspects of the shape might be relevant for inference(s) of interest.
- A shape that is invariant under updating may not exist.
- Sets of points or vertical lines are invariant shape *descriptions*, but may be impractical.
- Better focus on the aspects that matter for the inference(s) of interest: search for conjugate shape description specific to inferences, i.e. a shape description that is invariant only to those aspects that matter
  - Example: rectangle is generally a coarse approximation, but does not change inference on P

Elicitation

- To me still quite unclear what all those shapes mean:
  what kind of prior information requests what kind of shape?

Elicitation

- To me still quite unclear what all those shapes mean: what kind of prior information requests what kind of shape?
- Approach via elicitation: is it possible to elicit a 'conjugate' prior set from three questions on
  - $[\underline{P}, \overline{P}]$ (prior prob. for 'Head')
  - $[\underline{P}_1, \overline{P}_1]$ (posterior prob. for 'Head' given one 'Head' observed)
  - $[\underline{P}_0, \overline{P}_0]$ (posterior prob. for 'Head' given one 'Tail' observed)

Elicitation

- To me still quite unclear what all those shapes mean: what kind of prior information requests what kind of shape?
- Approach via elicitation: is it possible to elicit a 'conjugate' prior set from three questions on
  - $[\underline{P}, \overline{P}]$ (prior prob. for 'Head')
  - $[\underline{P}_1, \overline{P}_1]$ (posterior prob. for 'Head' given one 'Head' observed)
  - $[\underline{P}_0, \overline{P}_0]$ (posterior prob. for 'Head' given one 'Tail' observed)
- Current study of four-point-sets (ignoring all between $\underline{n}^{(0)}$ and $\overline{n}^{(0)}$) to understand which answers could lead to "snout left" and which to "snout right".