On Prior-Data Conflict in Predictive Bernoulli Inferences

Gero Walter, Thomas Augustin

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

Frank P.A. Coolen

Department of Mathematics
Durham University

July 28th, 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 $P(\text{heads})$ should be around 0.75)
- interested in probability P that the next observation is a head.
  (predictive probability!)

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 probability P that the next observation is a head. (predictive probability!)
- **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 predictive 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)

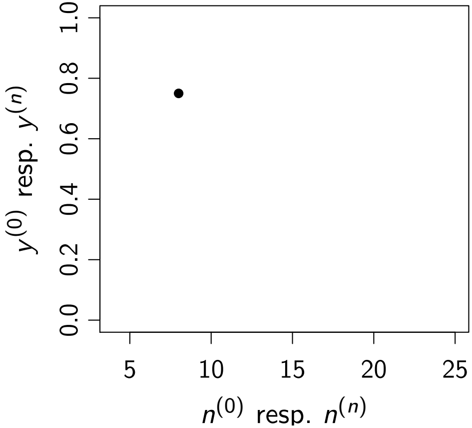
- one standard statistical model for this situation: Beta-Bernoulli/Binomial Model
- here in parameterization used, e.g., by Walley:

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!

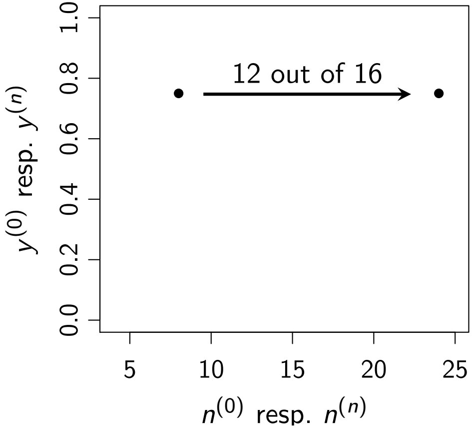
Beta-Bernoulli/Binomial Model (BBM)



no conflict:

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

Beta-Bernoulli/Binomial Model (BBM)

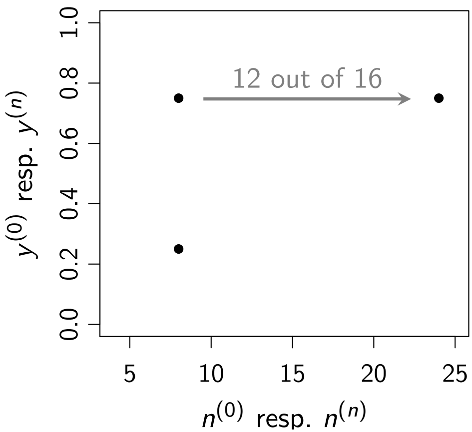


no conflict:

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:

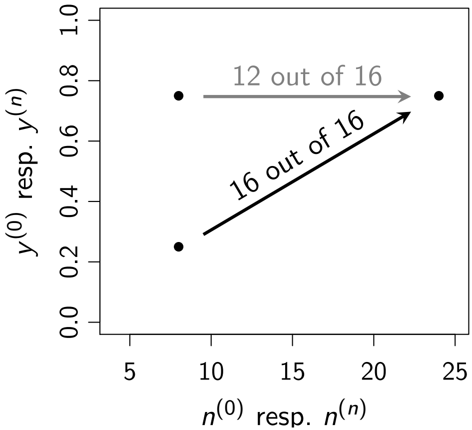
prior $n^{(0)} = 8$, $y^{(0)} = 0.75$
data $s/n = 12/16 = 0.75$

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

prior-data conflict:

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

Beta-Bernoulli/Binomial Model (BBM)



no conflict:

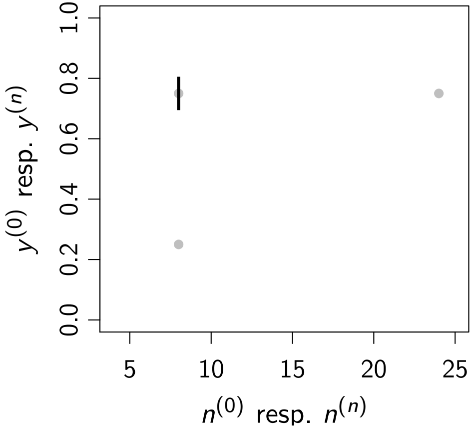
prior $n^{(0)} = 8$, $y^{(0)} = 0.75$
data $s/n = 12/16 = 0.75$

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

prior-data conflict:

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

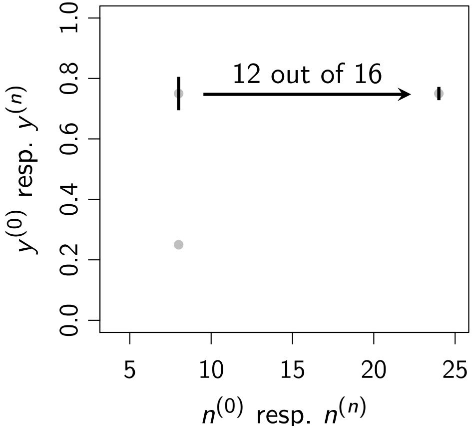
Imprecise BBM (IBBM) $\hat{=}$ IDM with prior information



no conflict:

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

Imprecise BBM (IBBM) ≙ IDM with prior information



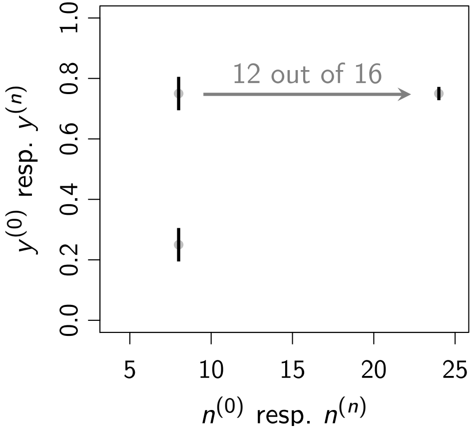
no conflict:

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:

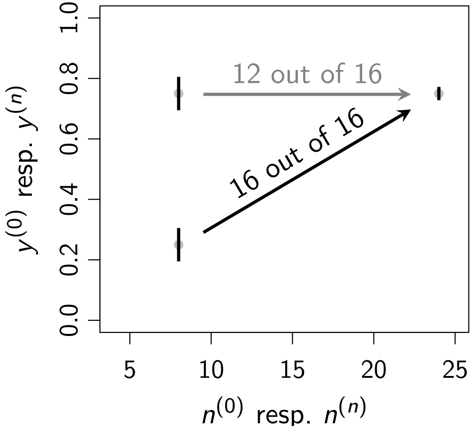
prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

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

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:

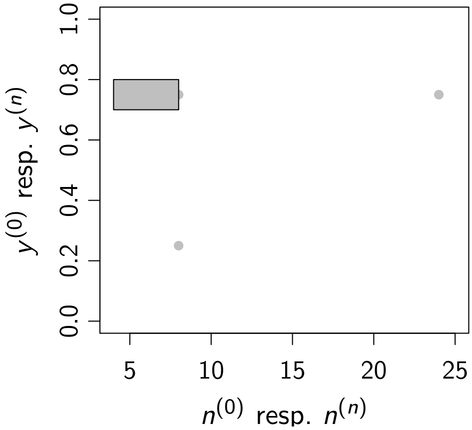
prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

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

prior-data conflict:

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

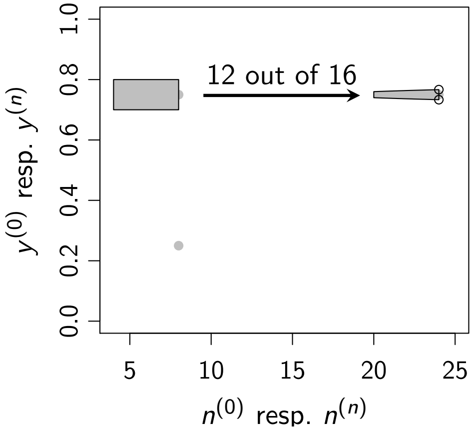
pdc-Imprecise BBM (pdc-IBBM): Walley 1991, Ch.5.4.3



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

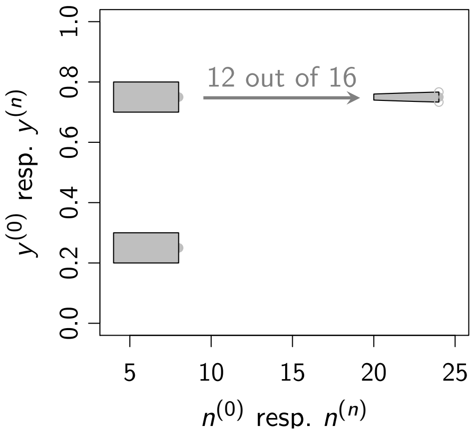


no conflict:

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 $s/n = 12/16 = 0.75$

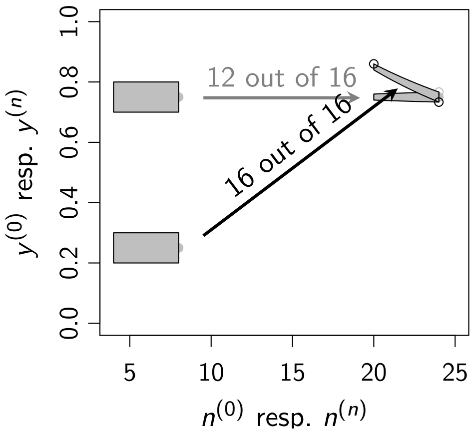
▼

"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 – Issues and alternatives

- two-dimensional interval seems natural, but generally any shape possible

pdc-IBBM – Issues and alternatives

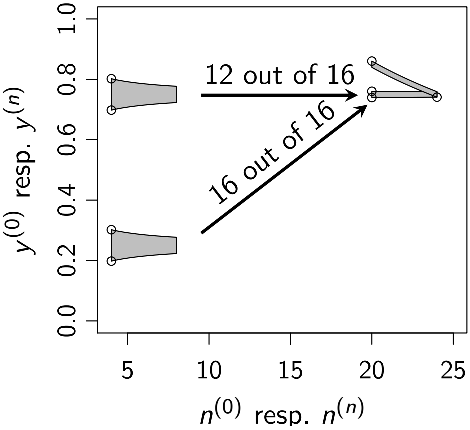
- two-dimensional interval seems natural, but generally any shape possible
- pdc-IBBM goes “bananas” irrespective of n or $n^{(0)}$

pdc-IBBM – Issues and alternatives

- two-dimensional interval seems natural, but generally any shape possible
- pdc-IBBM goes “bananas” irrespective of n or $n^{(0)}$
- a more tolerant shape?
  ↪ the “anteater” shape:

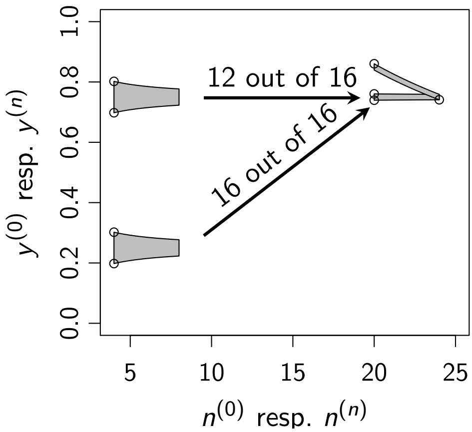
pdc-IBBM – Issues and alternatives

- two-dimensional interval seems natural, but generally any shape possible
- pdc-IBBM goes "bananas" irrespective of n or $n^{(0)}$
- a more tolerant shape?
  - ↳ the "anteater" shape:



pdc-IBBM – Issues and alternatives

- two-dimensional interval seems natural, but generally any shape possible
- pdc-IBBM goes "bananas" irrespective of n or $n^{(0)}$
- a more tolerant shape?
  ➥ the "anteater" shape:
- "anteater" is difficult to elicit
  ➥ other approach: combine predictive inferences



Weighted Inference / (Imprecise) Inference Fusion

Combine predictive inferences of

1. an *uninformative* model ➡ $[\underline{P}^u, \overline{P}^u]$ (near-ignorance prior)
2. an *informative* model ➡ $[\underline{P}^i, \overline{P}^i]$ (informative prior)

Weighted Inference / (Imprecise) Inference Fusion

Combine predictive inferences of

1. an *uninformative* model ➡ $[\underline{P}^u, \overline{P}^u]$ (near-ignorance prior)
2. an *informative* model ➡ $[\underline{P}^i, \overline{P}^i]$ (informative prior)

by weighing them

$$\underline{P}_\alpha = \alpha \underline{P}^i + (1-\alpha)\underline{P}^u$$

$$\overline{P}_\alpha = \alpha \overline{P}^i + (1-\alpha)\overline{P}^u$$

Weighted Inference / (Imprecise) Inference Fusion

Combine predictive inferences of

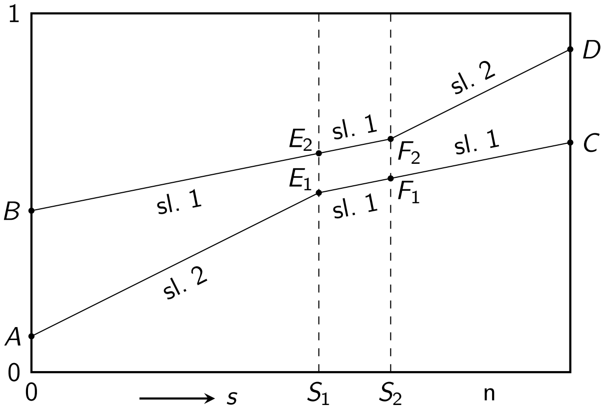
1. an *uninformative* model ➡ $[\underline{\mathrm{P}}^u, \overline{\mathrm{P}}^u]$ (near-ignorance prior)
2. an *informative* model ➡ $[\underline{\mathrm{P}}^i, \overline{\mathrm{P}}^i]$ (informative prior)

by weighing them **imprecisely**:

$$\underline{\mathrm{P}} = \min_{\alpha \in [\alpha_l, \alpha_r]} \underline{\mathrm{P}}_\alpha, \qquad \text{where} \qquad \underline{\mathrm{P}}_\alpha = \alpha \underline{\mathrm{P}}^i + (1-\alpha)\underline{\mathrm{P}}^u$$

$$\overline{\mathrm{P}} = \max_{\alpha \in [\alpha_l, \alpha_r]} \overline{\mathrm{P}}_\alpha, \qquad \text{where} \qquad \overline{\mathrm{P}}_\alpha = \alpha \overline{\mathrm{P}}^i + (1-\alpha)\overline{\mathrm{P}}^u$$

Comparison with IBBMs



The predictive probability plot (PPP) illustrates model behaviour.

Summary, Outlook

IBBMs

- sets of BBMs: vary parameters $n^{(0)}$, $y^{(0)}$
- can be generalized to any distribution from exponential family: same weighted average structure for $y^{(n)}$ (Quaghebeur & de Cooman 2005)
- generalization of pdc-IBBM: Walter, Augustin (2009)

Weighted Inference

- combine BBM inferences, not BBM models
- can be used to combine any two predictive inferences from any model, on any event of interest
- possible source for P^u / P^i: NPI (Coolen & Augustin 2009)