

# On Prior-Data Conflict in Predictive Bernoulli Inferences

**Gero Walter, Thomas Augustin**

Department of Statistics  
Ludwig-Maximilians-Universität München (LMU)  
{gero.walter; thomas}@stat.uni-muenchen.de

**Frank P.A. Coolen**

Department of Mathematics  
Durham University  
frank.coolen@durham.ac.uk

## Abstract

By its capability to deal with the multidimensional nature of uncertainty, imprecise probability provides a powerful methodology to sensibly handle prior-data conflict in Bayesian inference. When there is strong conflict between sample observations and prior knowledge the posterior model should be more imprecise than in the situation of mutual agreement or compatibility. Focusing presentation on the prototypical example of Bernoulli trials, we discuss the ability of different approaches to deal with prior-data conflict.

We study a generalized Bayesian setting, including Walley's Imprecise Beta-Binomial model and his extension to handle prior data conflict (called pdc-IBBM here). We investigate alternative shapes of prior parameter sets, chosen in a way that shows improved behaviour in the case of prior-data conflict and their influence on the posterior predictive distribution. Thereafter we present a new approach, consisting of an imprecise weighting of two originally separate inferences, one of which is based on an informative imprecise prior whereas the other one is based on an uninformative imprecise prior. This approach deals with prior-data conflict in a fascinating way.

**Keywords.** Bayesian inference; generalized iLUCK-models; imprecise Beta-Binomial model; imprecise weighting; predictive inference; prior-data conflict.

## 1 Introduction

Imprecise probability has shown to be a powerful methodology to cope with the multidimensional nature of uncertainty [8, 2]. Imprecision allows the quality of information, on which probability statements are based, to be modeled. Well supported knowledge is expressed by comparatively precise models, while highly imprecise (or even vacuous) models reflect scarce (or no) knowledge on probabilities. This flexible, multidimensional perspective on uncertainty

modeling has intensively been utilized in generalized Bayesian inference to overcome the criticism of the arbitrariness of the choice of single prior distributions in traditional Bayesian inference. In addition, only imprecise probability models react reliably to the presence of prior-data conflict, i.e. situations where “the prior [places] its mass primarily on distributions in the sampling model for which the observed data is surprising” [9, p. 894]. Lower and upper probabilities allow a specific reaction to prior-data conflict and offer reasonable inferences if the analyst wishes to stick to his prior assumptions: starting with the same level of ambiguity in the prior specification, wide posterior intervals can reflect conflict between prior and data, while no prior-data conflict will lead to narrow intervals. Ideally the model could provide an extra ‘bonus’ of precision if prior assumptions are very strongly supported by the data. Such a model would have the advantage of (relatively) precise answers when the data confirm prior assumptions, while still rendering more cautionary answers in the case of prior-data conflict, thus leading to cautious inferences if, and only if, caution is needed.

Although Walley [18, p. 6] explicitly emphasizes this possibility to express prior-data conflict as one of the main motivations for imprecise probability, it has received surprisingly little attention. Rare exceptions include two short sections in [18, p. 6 and Ch. 5.4] and [14, 7, 23]. The popular IDM [19, 3] and its generalization to exponential families [15] do not reflect prior-data conflict. [21] used the basic ideas of [18, Ch. 5.4] to extend the approach of [15] to models that show sensitivity to prior-data conflict.

In this paper a deeper investigation of the issue of prior-data conflict is undertaken, focusing on the prototypic special case of predictive inference in Bernoulli trials: We are interested in the posterior predictive probability for the event that a future Bernoulli random quantity will have the value 1, also called a ‘success’. This event is not explicitly included in the nota-

tion, i.e. we simply denote its lower and upper probabilities by  $\underline{P}$  and  $\bar{P}$ , respectively. This future Bernoulli random quantity is assumed to be exchangeable with the Bernoulli random quantities whose observations are summarized in the data, consisting of the number  $n$  of observations and the number  $s$  of these that are successes. In our analysis of this model, we will often consider  $s$  as a real-valued observation in  $[0, n]$ , keeping in mind that in reality it can only take on integer values, but the continuous representation is convenient for our discussions, in particular in our predictive probability plots (PPP), where for given  $n$ ,  $\underline{P}$  and  $\bar{P}$  are discussed as functions of  $s$ .

Section 2.1 describes a general framework for generalized Bayesian inference in this setting. The method presented in [18, Ch. 5.4.3], called ‘pdc-IBBM’ in this paper, is considered in detail in Section 2.2 and we show that its reaction to prior-data conflict can be improved by suitable modifications of the underlying imprecise priors. A basic proposal along these lines is discussed in Section 2.3 with further alternatives sketched in Section 2.4. Section 3 addresses the problem of prior-data conflict from a completely different angle. There we combine two originally separate inferences, one based on an informative imprecise prior and one on an uninformative imprecise prior, by an imprecise weighting scheme. The paper concludes with a brief comparison of the different approaches.

## 2 Imprecise Beta-Binomial Models

### 2.1 The Framework

The traditional Bayesian approach for our basic problem is the Beta-Binomial model, which expresses prior beliefs about the probability  $p$  of observing a ‘success’ by a Beta distribution. With<sup>1</sup>  $f(p) \propto p^{n^{(0)}y^{(0)}-1}(1-p)^{n^{(0)}(1-y^{(0)})-1}$ ,  $y^{(0)} = E[p]$  can be interpreted as prior guess of  $p$ , while  $n^{(0)}$  governs the concentration of probability mass around  $y^{(0)}$ , also known as ‘pseudo counts’ or ‘prior strength’.<sup>2</sup> These denominations are due to the role of  $n^{(0)}$  in the update step: With  $s$  successes in  $n$  draws observed, the posterior parameters are<sup>3</sup>

$$n^{(n)} = n^{(0)} + n, \quad y^{(n)} = \frac{n^{(0)}y^{(0)} + s}{n^{(0)} + n}. \quad (1)$$

Thus  $y^{(n)}$  is a weighted average of the prior parameter  $y^{(0)}$  and the sample proportion  $s/n$ , and potential prior data conflict is simply averaged out.

<sup>1</sup>Our notation relates to [18]’s as  $n^{(0)} \leftrightarrow s_0$ ,  $y^{(0)} \leftrightarrow t_0$ .

<sup>2</sup> $(0)$  denotes prior parameters;  $(n)$  posterior parameters.

<sup>3</sup>The model is prototypic for conjugate Bayesian analysis in canonical exponential families, for which updating of the parameters  $n^{(0)}$  and  $y^{(0)}$  can be written as (1).

Overcoming the dogma of precision, formulating generalized Bayes updating in this setting is straightforward. By Walley’s Generalized Bayes Rule [18, Ch. 6] the imprecise prior  $\mathcal{M}^{(0)}$ , described by convex sets of precise prior distributions, is updated to the imprecise posterior  $\mathcal{M}^{(n)}$  obtained by updating  $\mathcal{M}^{(0)}$  elementwise. In particular, the convenient conjugate analysis used above can be extended: One specifies a prior parameter set  $\Pi^{(0)}$  of  $(n^{(0)}, y^{(0)})$  values and takes as imprecise prior the set  $\mathcal{M}^{(0)}$  consisting of all convex mixtures of Beta priors with  $(n^{(0)}, y^{(0)}) \in \Pi^{(0)}$ . In this sense, the set of Beta priors corresponding to  $\Pi^{(0)}$  gives the set of extreme points for the actual convex set of priors  $\mathcal{M}^{(0)}$ . Updating  $\mathcal{M}^{(0)}$  with the Generalized Bayes’ Rule results in the convex set  $\mathcal{M}^{(n)}$  of posterior distributions that conveniently can be obtained by taking the convex hull of the set of Beta posteriors, which in turn are defined by the set of updated parameters  $\Pi^{(n)} = \{(n^{(n)}, y^{(n)}) \mid (n^{(0)}, y^{(0)}) \in \Pi^{(0)}\}$ . This relationship between the sets  $\Pi^{(0)}$  and  $\Pi^{(n)}$  and the sets  $\mathcal{M}^{(0)}$  and  $\mathcal{M}^{(n)}$  will allow us to discuss different models  $\mathcal{M}^{(0)}$  and  $\mathcal{M}^{(n)}$  by depicting the corresponding parameter sets  $\Pi^{(0)}$  and  $\Pi^{(n)}$ . When interpreting our results, care will be needed with respect to convexity. Although  $\mathcal{M}^{(0)}$  and  $\mathcal{M}^{(n)}$  are convex, the parameter sets  $\Pi^{(0)}$  and  $\Pi^{(n)}$  generating them need not necessarily be so. Indeed, convexity of the parameter set is not necessarily preserved in the update step: Convexity of  $\Pi^{(0)}$  does not imply convexity of  $\Pi^{(n)}$ .

Throughout, we are interested in the posterior predictive probability  $[\underline{P}, \bar{P}]$  for the event that a future draw is a success. In the Beta-Bernoulli model, this probability is equal to  $y^{(n)}$ , and we get<sup>4</sup>

$$\underline{P} = \underline{y}^{(n)} := \min_{\Pi^{(n)}} y^{(n)} = \min_{\Pi^{(0)}} \frac{n^{(0)}y^{(0)} + s}{n^{(0)} + n}, \quad (2)$$

$$\bar{P} = \bar{y}^{(n)} := \max_{\Pi^{(n)}} y^{(n)} = \max_{\Pi^{(0)}} \frac{n^{(0)}y^{(0)} + s}{n^{(0)} + n}. \quad (3)$$

### 2.2 Walley’s pdc-IBBM

Special imprecise probability models are now obtained by specific choices of  $\Pi^{(0)}$ . If one fixes  $n^{(0)}$  and varies  $y^{(0)}$  in an interval  $[\underline{y}^{(0)}, \bar{y}^{(0)}]$ , Walley’s [18, Ch. 5.3] model with learning parameter  $n^{(0)}$  is obtained, which typically is used in its near-ignorance form  $[\underline{y}^{(0)}, \bar{y}^{(0)}] \rightarrow (0, 1)$ , denoted as the imprecise Beta (Binomial/Bernoulli) model (IBBM)<sup>5</sup>, which is a special case of the popular Imprecise Dirichlet (Multinomial) Model [19, 20]. Unfortunately, in this basic form with fixed  $n^{(0)}$  the model is insensitive to prior-

<sup>4</sup>[15, 21, 22] use the prototypical character of (1) underlying (2) and (3) to generalize this inference to models based on canonical exponential families.

<sup>5</sup>We use ‘IBBM’ also for the model with prior information.

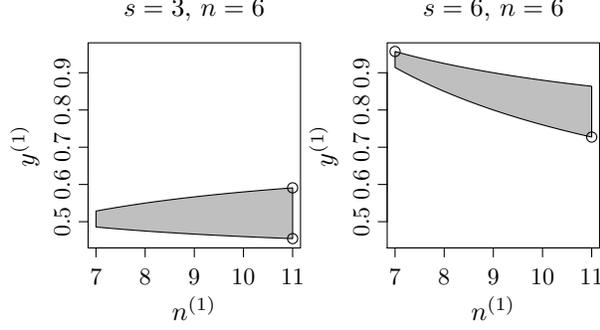


Figure 1: Posterior parameter sets  $\Pi^{(n)}$  for rectangular  $\Pi^{(0)}$ . Left: *spotlight* shape; right: *banana* shape.

data conflict [21, p. 263]. Walley [18, Ch. 5.4] therefore generalized this model by additionally varying  $n^{(0)}$ . In his extended model, called pdc-IBBM in this paper, the set of priors is defined via the set of prior parameters  $\Pi^{(0)} = [\underline{n}^{(0)}, \bar{n}^{(0)}] \times [\underline{y}^{(0)}, \bar{y}^{(0)}]$ , being a two-dimensional interval, or a rectangle set. Studying inference in this model, it is important to note that the set of posterior parameters  $\Pi^{(n)}$  is not rectangular anymore. The resulting shapes are illustrated in Figure 1: For the prior set  $\Pi^{(0)} = [1, 5] \times [0.4, 0.7]$ —thus assuming a priori the fraction of successes to be between 40% and 70% and rating these assumptions with at least 1 and at most 5 pseudo observations—the resulting posterior parameter sets  $\Pi^{(n)}$  are shown for data consisting of 3 successes in 6 draws (left) and with all 6 draws successes (right). We call the left shape *spotlight*, and the right shape *banana*. In both graphs, the elements of  $\Pi^{(n)}$  yielding  $\underline{y}^{(n)}$  and  $\bar{y}^{(n)}$ , and thus  $\underline{P}$  and  $\bar{P}$ , are marked with a circle.

The transition point between the *spotlight* and the *banana* shape in Figure 1 is the case when  $\frac{s}{n} = \bar{y}^{(0)}$ . Then  $\bar{y}^{(n)}$ , being a weighted average of  $\bar{y}^{(0)}$  and  $\frac{s}{n}$ , is attained for all  $n^{(0)} \in [\underline{n}^{(0)}, \bar{n}^{(0)}]$ , and the top border of  $\Pi^{(n)}$  in the graphical representation of Figure 1 is constant. Likewise,  $\underline{y}^{(n)}$  is constant if  $\frac{s}{n} = \underline{y}^{(0)}$ . Therefore, (2) and (3) can be subsumed as

$$\underline{P} = \begin{cases} \frac{\bar{n}^{(0)}\underline{y}^{(0)}+s}{\bar{n}^{(0)}+n} & \text{if } s \geq n \cdot \underline{y}^{(0)} =: S_1 \\ \frac{\underline{n}^{(0)}\underline{y}^{(0)}+s}{\underline{n}^{(0)}+n} & \text{if } s \leq n \cdot \underline{y}^{(0)} =: S_1 \end{cases},$$

$$\bar{P} = \begin{cases} \frac{\bar{n}^{(0)}\bar{y}^{(0)}+s}{\bar{n}^{(0)}+n} & \text{if } s \leq n \cdot \bar{y}^{(0)} =: S_2 \\ \frac{\underline{n}^{(0)}\bar{y}^{(0)}+s}{\underline{n}^{(0)}+n} & \text{if } s \geq n \cdot \bar{y}^{(0)} =: S_2 \end{cases}.$$

The interval  $[S_1, S_2]$  gives the range of expected successes  $[n \cdot \underline{y}^{(0)}, n \cdot \bar{y}^{(0)}]$  and will be called ‘Total Prior-Data Agreement’ interval, or TPDA. For  $s$  in the TPDA, we are ‘spot on’:  $\underline{y}^{(n)}$  and  $\bar{y}^{(n)}$  are attained

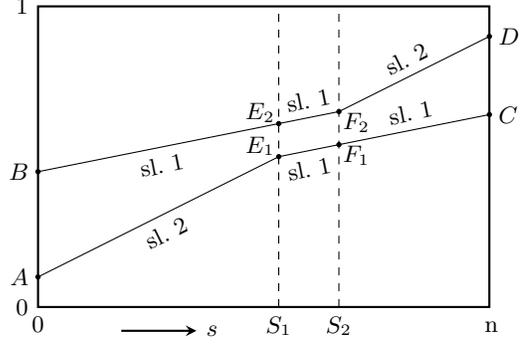


Figure 2:  $\underline{P}$  and  $\bar{P}$  for models in Sections 2.2 and 2.3.

for  $\bar{n}^{(0)}$  and  $\Pi^{(n)}$  has the *spotlight* shape. But if the observed number of successes is outside TPDA,  $\Pi^{(n)}$  goes *bananas* and either  $\underline{P}$  or  $\bar{P}$  is calculated with  $\underline{n}^{(0)}$ .

To summarize, the predictive probability plot (PPP), displaying  $\underline{P}$  and  $\bar{P}$  for  $s \in [0, n]$ , is given in Figure 2. For the pdc-IBBM, the specific values are

$$A = \frac{\underline{n}^{(0)}\underline{y}^{(0)}}{\underline{n}^{(0)}+n} \quad C = \frac{\bar{n}^{(0)}\underline{y}^{(0)}+n}{\bar{n}^{(0)}+n}$$

$$B = \frac{\bar{n}^{(0)}\bar{y}^{(0)}}{\bar{n}^{(0)}+n} \quad D = \frac{\underline{n}^{(0)}\bar{y}^{(0)}+n}{\underline{n}^{(0)}+n}$$

$$\text{sl. 1} = \frac{1}{\bar{n}^{(0)}+n} \quad E_1 = \underline{y}^{(0)} \quad E_2 = \frac{\bar{n}^{(0)}\bar{y}^{(0)}+n\underline{y}^{(0)}}{\bar{n}^{(0)}+n}$$

$$\text{sl. 2} = \frac{1}{\underline{n}^{(0)}+n} \quad F_2 = \bar{y}^{(0)} \quad F_1 = \frac{\bar{n}^{(0)}\underline{y}^{(0)}+n\bar{y}^{(0)}}{\bar{n}^{(0)}+n}.$$

As noted by [18, p. 224], the posterior predictive imprecision  $\Delta = \bar{P} - \underline{P}$  can be calculated as

$$\Delta = \frac{\bar{n}^{(0)}(\bar{y}^{(0)} - \underline{y}^{(0)})}{\bar{n}^{(0)}+n} + \frac{\bar{n}^{(0)} - \underline{n}^{(0)}}{(\underline{n}^{(0)}+n)(\bar{n}^{(0)}+n)} \Delta(s, \Pi^{(0)}),$$

where  $\Delta(s, \Pi^{(0)}) = \inf\{|s - ny^{(0)}| : y^{(0)} \in [\underline{y}^{(0)}, \bar{y}^{(0)}]\}$  is the distance of  $s$  to the TPDA. If  $\Delta(s, \Pi^{(0)}) \neq 0$ , we have an effect of additional imprecision as desired, increasing linearly in  $s$ , because  $\Pi^{(n)}$  is going *bananas*. However, when considering the fraction of observed successes instead of  $s$ , the onset of this additional imprecision immediately if  $\frac{s}{n} \notin [\underline{y}^{(0)}, \bar{y}^{(0)}]$  seems very abrupt. Moreover, and even more severe, it happens irrespective of the number of trials  $n$ . When updating successively, this means that all single Bernoulli observations, being either 0 or 1, have to be treated as if being in conflict (except if  $\bar{y}^{(0)} = 1$  and  $s = n$  or if  $\underline{y}^{(0)} = 0$  and  $s = 0$ ). Furthermore, regarding  $s/n = 7/10$  as an instance of prior-data conflict when  $\bar{y}^{(0)} = 0.6$  had been assumed seems somewhat picky. To explore possibilities to amend this behaviour, alternative approaches are explored next.

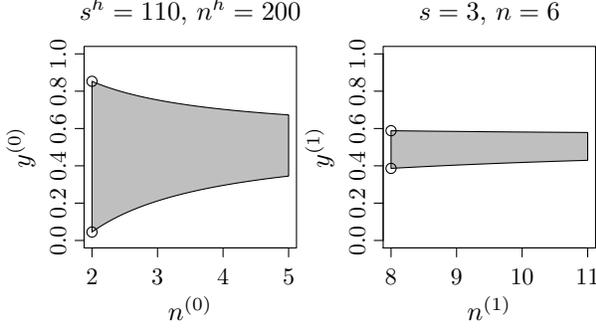


Figure 3:  $\Pi^{(0)}$  and  $\Pi^{(n)}$  for the *anteater* shape.

### 2.3 Anteater Shape Prior Sets

Choosing a two-dimensional interval  $\Pi^{(0)}$  seems logical but the resulting inference is not fully satisfactory in case of prior data conflict. Recall that  $\Pi^{(0)}$  is used to produce  $\mathcal{M}^{(0)}$ , which then is processed by the Generalized Bayes rule. Any shape can be chosen for  $\Pi^{(0)}$ , including the composure of single pairs  $(n^{(0)}, y^{(0)})$ . In this section we investigate an alternative shape, with  $y^{(0)}$  a function of  $n^{(0)}$ , aiming at a more advanced behaviour in the case of prior-data conflict. To elicit  $\Pi^{(0)}$ , one could consider a thought experiment<sup>6</sup>: Given the hypothetical observation of  $s^h$  successes in  $n^h$  trials, which values should  $\underline{P}$  and  $\bar{P}$  take? In other words, what would one like to learn from data  $s^h/n^h$  in accordance with prior beliefs? As a simple approach, we can define  $\Pi^{(0)}$  such that  $\underline{P} = \underline{c}$  and  $\bar{P} = \bar{c}$  are constants in  $n^{(n)} = n^{(0)} + n^h$ . Then, the lower and upper bounds for  $y^{(0)}$  must be

$$\begin{aligned} \underline{y}^{(0)}(n^{(0)}) &= ((n^h + n^{(0)})\underline{c} - s^h)/n^{(0)}, \\ \bar{y}^{(0)}(n^{(0)}) &= ((n^h + n^{(0)})\bar{c} - s^h)/n^{(0)}, \end{aligned} \quad (4)$$

for  $n^{(0)}$  in an interval  $[\underline{n}^{(0)}, \bar{n}^{(0)}]$  derived by the range  $[\underline{n}^{(n)}, \bar{n}^{(n)}]$  one wishes to attain for  $\underline{P}$  and  $\bar{P}$  given the  $n^h$  hypothetical observations.<sup>7</sup> The resulting shape of  $\Pi^{(0)}$  is as in Figure 3 (left) and called *anteater* shape. Rewriting (4),  $\Pi^{(0)}$  is now defined as

$$\left\{ (n^{(0)}, y^{(0)}) \mid n^{(0)} \in [\underline{n}^{(0)}, \bar{n}^{(0)}], \right. \\ \left. y^{(0)}(n^{(0)}) \in \left[ \underline{c} - \frac{n^h}{n^{(0)}} \left( \frac{s^h}{n^h} - \underline{c} \right), \bar{c} + \frac{n^h}{n^{(0)}} \left( \bar{c} - \frac{s^h}{n^h} \right) \right] \right\}.$$

With the reasonable choice of  $\underline{c}$  and  $\bar{c}$  such that  $\underline{c} \leq s^h/n^h \leq \bar{c}$ ,  $\Pi^{(0)}$  can be interpreted as follows: The range of  $y^{(0)}$  protrudes over  $[\underline{c}, \bar{c}]$  on either side far enough to ensure  $\underline{P} = \underline{c}$  and  $\bar{P} = \bar{c}$  if updated with  $s = s^h$  for  $n = n^h$ , the amount of protrusion decreasing in  $n^{(0)}$  as the movement of  $y^{(0)}(n^{(0)})$  towards

<sup>6</sup>AKA ‘pre-posterior’ analysis in the Bayesian literature.

<sup>7</sup>For the rest of the paper, we tacitly assume that  $n^h, s^h, n^{(0)}$  and  $\underline{c}/\bar{c}$  are chosen such that  $y^{(0)} \geq 0$  resp.  $\bar{y}^{(0)} \leq 1$  to generate Beta distributions as priors.

$s^h/n^h$  is slower for larger values of  $n^{(0)}$ . As there is a considerable difference in behaviour if  $n > n^h$  or  $n < n^h$ , these two cases are discussed separately.

If  $n > n^h$ , the PPP graph in Figure 2 holds again, now with the values

$$\begin{aligned} A &= \frac{\underline{c}(n^{(0)} + n^h) - s^h}{n^{(0)} + n} & S_1 &= s^h + \underline{c}(n - n^h) & E_1 &= \underline{c} \\ B &= \frac{\bar{c}(n^{(0)} + n^h) - s^h}{n^{(0)} + n} & S_2 &= s^h + \bar{c}(n - n^h) & F_2 &= \bar{c} \\ C &= \frac{\underline{c}(\bar{n}^{(0)} + n^h) - s^h + n}{\bar{n}^{(0)} + n} & \text{sl. 1} &= 1/(\bar{n}^{(0)} + n) \\ D &= \frac{\bar{c}(n^{(0)} + n^h) - s^h + n}{n^{(0)} + n} & \text{sl. 2} &= 1/(n^{(0)} + n) \end{aligned}$$

$$\begin{aligned} E_2 &= \underline{c} + \frac{\bar{n}^{(0)} + n^h}{\bar{n}^{(0)} + n} (\bar{c} - \underline{c}) = \bar{c} - \frac{n - n^h}{\bar{n}^{(0)} + n} (\bar{c} - \underline{c}) \\ F_1 &= \bar{c} - \frac{\bar{n}^{(0)} + n^h}{\bar{n}^{(0)} + n} (\bar{c} - \underline{c}) = \underline{c} + \frac{n - n^h}{\bar{n}^{(0)} + n} (\bar{c} - \underline{c}). \end{aligned}$$

As for the pdc-IBBM, the TPDA boundaries  $S_1$  and  $S_2$  mark the transition points where either  $\underline{y}^{(n)}$  or  $\bar{y}^{(n)}$  are constant in  $n^{(0)}$ . We now have

$$\frac{S_1}{n} = \underline{c} + \frac{n^h}{n} \left( \frac{s^h}{n^h} - \underline{c} \right), \quad \frac{S_2}{n} = \bar{c} - \frac{n^h}{n} \left( \bar{c} - \frac{s^h}{n^h} \right),$$

so this TPDA is a subset of  $[\underline{c}, \bar{c}]$ . The *anteater* shape is, for  $n > n^h$ , even more strict than the pdc-IBBM, as, e.g.,  $\underline{y}^{(0)}(\bar{n}^{(0)}) = \underline{c} - \frac{n^h}{\bar{n}^{(0)}} \left( \frac{s^h}{n^h} - \underline{c} \right) < \frac{S_1}{n}$ .

The situation for  $n < n^h$  is illustrated in Figure 4, where  $A, B, C, D, E_1, F_2$  and slopes 1 and 2 are the same as for  $n > n^h$ , but

$$\begin{aligned} E_2 &= \underline{c} + \frac{\underline{n}^{(0)} + n^h}{\underline{n}^{(0)} + n} (\bar{c} - \underline{c}) = \bar{c} + \frac{n^h - n}{\underline{n}^{(0)} + n} (\bar{c} - \underline{c}), \\ F_1 &= \bar{c} - \frac{\underline{n}^{(0)} + n^h}{\underline{n}^{(0)} + n} (\bar{c} - \underline{c}) = \underline{c} - \frac{n^h - n}{\underline{n}^{(0)} + n} (\bar{c} - \underline{c}). \end{aligned}$$

Note that now  $S_2 < S_1$ , so the TPDA is  $[S_2, S_1]$ . In this interval,  $\underline{P}$  and  $\bar{P}$  are now calculated with  $\underline{n}^{(0)}$ ; for  $s \notin [S_2, S_1]$  the same situation as for  $n > n^h$  applies, with the bound nearer to  $s/n$  calculated with  $\underline{n}^{(0)}$  and the other with  $\bar{n}^{(0)}$ .

The upper transition point  $S_1$  can now be between  $\bar{y}^{(0)}(\underline{n}^{(0)})$  and  $\bar{y}^{(0)}(\bar{n}^{(0)})$ , and having  $S_1$  decreasing in  $n$  now makes sense: the smaller  $n$ , the larger  $S_1$ , i.e. the more tolerant is the *anteater* set. The switch over  $S_1$  (with  $s/n$  increasing) is illustrated in the three graphs in Figures 3 (right) and 5 (left, right): First,  $\Pi^{(0)}$  from Figure 3 (left) is updated with  $s/n = 3/6 < S_1/n$ , leading again to an *anteater* shape, and so we get  $\underline{P}$  and  $\bar{P}$  from the elements of  $\Pi^{(n)}$  at  $\underline{n}^{(n)}$ , as marked with circles. Second, the transition point is reached for  $s = S_1 = 4.27$ , and now  $\underline{P}$  is attained for any  $n^{(n)} \in [\underline{n}^{(n)}, \bar{n}^{(n)}]$ , as emphasized by the arrow. Third, as soon as  $s$  exceeds  $S_1$  (in the graph:

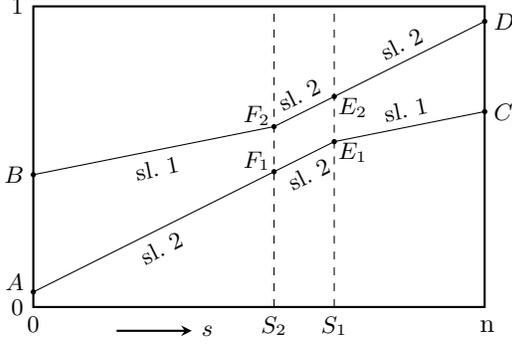


Figure 4:  $\underline{P}$  and  $\bar{P}$  for the *anteater* shape if  $n < n^h$ .

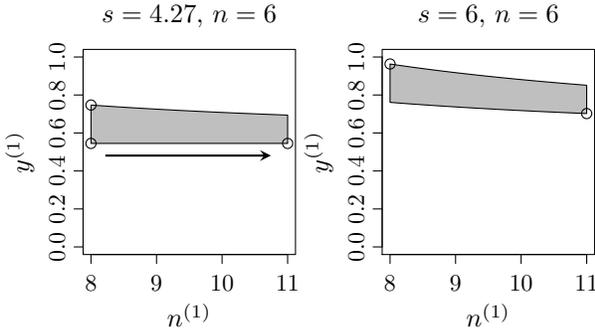


Figure 5: Posterior parameter sets  $\Pi^{(n)}$  for *anteater* prior sets  $\Pi^{(0)}$ . Left: the transition point where  $\underline{y}^{(n)}$  is attained for all  $n^{(n)}$ , right: the *banana* shape.

$s/n = 6/6$ , it holds that  $\underline{y}^{(n)}(\underline{n}^{(n)}) > \underline{y}^{(n)}(\bar{n}^{(n)})$ , and  $\underline{P}$  is now attained at  $\bar{n}^{(n)}$ . As for the pdc-IBBM, for  $s$  outside the TPDA  $\Pi^{(n)}$  goes *bananas*, leading to additional imprecision. The imprecision  $\Delta = \bar{P} - \underline{P}$  if  $n < n^h$  is

$$\Delta = \frac{\underline{n}^{(0)} + n^h}{\underline{n}^{(0)} + n} (\bar{c} - \underline{c}) + \frac{\bar{n}^{(0)} - \underline{n}^{(0)}}{(\underline{n}^{(0)} + n)(\bar{n}^{(0)} + n)} \Delta(s, n, \mathbf{c}),$$

where  $\Delta(s, n, \mathbf{c}) = n|c^* - \frac{s}{n}| - n^h|c^* - \frac{s^h}{n^h}|$  and  $c^* = \arg \max_{c \in [\underline{c}, \bar{c}]} |\frac{s}{n} - c|$  is the boundary of  $[\underline{c}, \bar{c}]$  with the largest distance to  $s/n$ . For  $s \in [S_2, S_1]$ ,  $\Delta(s, n, \mathbf{c}) = 0$ , giving a similar structure as for the pdc-IBBM except that  $\Delta(s, n, \mathbf{c})$  does not directly give the distance of  $s/n$  to  $\Pi^{(0)}$  but is based on  $[\underline{c}, \bar{c}]$ . The imprecision increases again linearly with  $s$ , but now also with  $n$ . The distance of  $s/n$  to the opposite bound of  $[\underline{c}, \bar{c}]$  (weighted with  $n$ ) is discounted by the distance of  $s^h/n^h$  to the same bound (weighted with  $n^h$ ). In essence,  $\Delta(s, n, \mathbf{c})$  is thus a reweighted distance of  $s/n$  to  $s^h/n^h$ . The more dissimilar these fractions are, the larger the posterior predictive imprecision is.

For  $n = n^h$ ,  $S_1 = S_2 = s^h$  so the TPDA is reduced to a single point. In this case, the *anteater* shape

$n > n^h$	$s < S_1$ banana	$s \in [S_1, S_2]$ spotlight	$s > S_2$ banana
$n = n^h$	$s < s^h$ banana	$s = s^h$ rectangular	$s > s^h$ banana
$n < n^h$	$s < S_2$ banana	$s \in [S_2, S_1]$ anteater	$s > S_1$ banana

Table 1: Shapes of  $\Pi^{(n)}$  if  $\Pi^{(0)}$  has the *anteater* shape.

can be considered as an equilibrium point, with any  $s \neq s^h$  leading to increased posterior imprecision. In this case, the weights in  $\Delta(s, n, \mathbf{c})$  coincide, and so the posterior imprecision depends directly on  $|s - s^h|$ .

For  $n > n^h$  the transition behaviour is as for the pdc-IBBM: As long as  $s \in [S_1, S_2]$ ,  $\Pi^{(n)}$  has the *spotlight* shape, where both  $\underline{P}$  and  $\bar{P}$  are calculated with  $\bar{n}^{(n)}$ ;  $\Delta$  for  $s \in [S_1, S_2]$  is thus calculated with  $\bar{n}^{(n)}$  as well. If, e.g.,  $s > S_2$ ,  $\bar{P}$  is attained with  $\underline{n}^{(n)}$ , and  $\Delta(s, n, \mathbf{c})$  gives directly the distance of  $s/n$  to  $s^h/n^h$ , the part of which is inside  $[\underline{c}, \bar{c}]$  is weighted with  $n$ , and the remainder with  $n^h$ . Table 1 provides an overview of the possible shapes of  $\Pi^{(n)}$ .

## 2.4 Intermediate Résumé

Despite the (partly) different behaviour inside the TPDA, both pdc-IBBM and the *anteater* shape display only two different slopes in their PPPs (Figures 2 and 4), with either  $\underline{n}^{(n)}$  or  $\bar{n}^{(n)}$  used to calculate  $\underline{P}$  and  $\bar{P}$ . It is possible to have shapes such that for some  $s$  other values from  $[\underline{n}^{(n)}, \bar{n}^{(n)}]$  are used. As a toy example, consider  $\Pi^{(0)} = \{(1, 0.4), (3, 0.6), (5, 0.4)\}$ , so consisting only of three parameter combinations  $(n^{(0)}, y^{(0)})$ .  $\bar{P}$  is then derived as  $\bar{y}^{(n)} = \max\{\frac{0.4+s}{1+n}, \frac{1.8+s}{3+n}, \frac{2+s}{5+n}\}$ , leading to

$$\bar{y}^{(n)} = \begin{cases} \frac{0.4+s}{1+n} & \text{if } s > 0.7n + 0.3 \\ \frac{1.8+s}{3+n} & \text{if } 0.1n - 1.5 < s < 0.7n + 0.3 \\ \frac{2+s}{5+n} & \text{if } s < 0.1n - 1.5 \end{cases} .$$

So, in a PPP we would observe the three different slopes  $1/(1+n)$ ,  $1/(3+n)$  and  $1/(5+n)$  depending on the value of  $s$ . Our conjecture is therefore that with carefully tailored sets  $\Pi^{(0)}$ , an arbitrary number of slopes is possible, and so even smooth curvatures. Using a thought experiment as for the *anteater* shape,  $\Pi^{(0)}$  shapes can be derived to fit any required behaviour. Another approach for constructing a  $\Pi^{(0)}$  that is more tolerant with respect to prior-data conflict could be as follows: As the onset of additional imprecision in the pdc-IBBM is caused by the fact that  $\bar{y}^{(n)}(\underline{n}^{(n)}) > \bar{y}^{(n)}(\bar{n}^{(n)})$  as soon as  $s/n > \bar{y}^{(0)}$ , we could define the  $y^{(0)}$  interval at  $\underline{n}^{(0)}$  to be narrower than the  $y^{(0)}$  interval at  $\bar{n}^{(0)}$ , so that the *banana* shape results only when  $s/n$  exceeds  $\bar{y}^{(0)}(\bar{n}^{(0)})$

far enough. Having a narrower  $y^{(0)}$  interval at  $\underline{n}^{(0)}$  than at  $\bar{n}^{(0)}$  could also make sense from an elicitation point of view: We might be able to give quite a precise  $y^{(0)}$  interval for a low prior strength  $\underline{n}^{(0)}$ , whereas for a high prior strength  $\bar{n}^{(0)}$  we must be more cautious with our elicitation of  $y^{(0)}$ , i.e. giving a wider interval. The rectangular shape for  $\Pi^{(0)}$  as discussed in Section 2.2 seems thus somewhat peculiar. One could also argue that if one has substantial prior information but acknowledges that this information may be wrong, one should not reduce the weight of the prior  $n^{(0)}$  on the posterior while keeping the same informative interval of values of  $y^{(0)}$ .

Generally, the actual shape of a set  $\Pi^{(0)}$  influences the inferences, but for a specific inference only a few aspects of the set are relevant. So, while a detailed shape of a prior set may be very difficult to elicit, it may not even be that relevant for a specific inference. A further general issue seems unavoidable in the generalized Bayesian setting as developed here, namely the dual role of  $n^{(0)}$ . On the one hand,  $n^{(0)}$  governs the weighting of prior information  $y^{(0)}$  with respect to the data  $s/n$ , as mentioned in Section 2.1: The larger  $n^{(0)}$ , the more  $\underline{P}$  and  $\bar{P}$  are dominated by  $\underline{y}^{(0)}$  and  $\bar{y}^{(0)}$ . On the other hand,  $n^{(0)}$  governs also the degree of posterior imprecision: the larger  $n^{(0)}$ , the larger c.p.  $\Delta$ . A larger  $n^{(0)}$  thus leads to more imprecise posterior inferences, although a high weight on the supplied prior information should boost the trust in posterior inferences if  $s$  in the TPDA, i.e. the prior information turned out to be appropriate. In the next section, we thus develop a different approach separating these two roles: Now, two separate models for predictive inference, each resulting in different precision as governed by  $n^{(0)}$ , are combined with an imprecise weight  $\alpha$  taking the role of regulating prior-data agreement.

### 3 Weighted Inference

We propose a variation of the Beta-Binomial model that is attractive for prior-data conflict and has small yet fascinating differences with the models in Sections 2.2 and 2.3. We present a basic version of the model in Section 3.1, followed by an extended version in Section 3.2. Opportunities to generalize the model are mentioned in Section 3.3.

#### 3.1 The Basic Model

The idea for the proposed model is to combine the inferences based on two models, each part of an imprecise Bayesian inferential framework using sets of prior distributions, although the inferences can also result from alternative inferential methods. The combination is not achieved by combining the two sets of

prior distributions into a single set, but by combining the posterior predictive *inferences* by imprecise weighted averaging. When the weights assigned to the two models can vary over the whole range  $[0, 1]$  we actually return to imprecise Bayesian inference with a prior set, as considered in this subsection. In Section 3.2 we restrict the values of the model weights. The basic model turns out to be relevant from many perspectives, in particular to highlight similarities and differences with the methods presented in Sections 2.2 and 2.3, and it is a suitable starting point for more general models. These aspects will be discussed in Subsection 3.3.

We consider the combination of the imprecise posterior predictive probabilities  $[\underline{P}^i, \bar{P}^i]$  and  $[\underline{P}^u, \bar{P}^u]$  for the event that the next observation is a success with

$$\underline{P}^i = \frac{s^i + s}{n^i + n + 1} \quad \text{and} \quad \bar{P}^i = \frac{s^i + s + 1}{n^i + n + 1}, \quad (5)$$

$$\underline{P}^u = \frac{s}{n + 1} \quad \text{and} \quad \bar{P}^u = \frac{s + 1}{n + 1}. \quad (6)$$

The superscript  $i$  indicates ‘informative’, in the sense that these lower and upper probabilities relate to an ‘informative’ prior distribution reflecting prior beliefs of similar value as  $s^i$  successes in  $n^i$  observations. The superscript  $u$  indicates ‘uninformative’, which can be interpreted as absence of prior beliefs. These lower and upper probabilities can for example result from Walley’s IBBM, with  $\underline{P}^i$  and  $\bar{P}^i$  based on the prior set with  $n^{(0)} = n^i + 1$  and  $y^{(0)} \in \left[ \frac{s^i}{n^i + 1}, \frac{s^i + 1}{n^i + 1} \right]$ , and  $\underline{P}^u$  and  $\bar{P}^u$  on the prior set with  $n^{(0)} = 1$  and  $y^{(0)} \in [0, 1]$ . There are other methods for imprecise statistical inference that lead to these same lower and upper probabilities, including Nonparametric Predictive Inference for Bernoulli quantities [4]<sup>8</sup>, where the  $s^i$  and  $n^i$  would only be included if they were actual observations, for example resulting from a second data set that one may wish to include in the ‘informative’ model but not in the ‘uninformative’ model.

The proposed method combines these lower and upper predictive probabilities by imprecise weighted averaging. Let  $\alpha \in [0, 1]$ , we define

$$\underline{P}_\alpha = \alpha \underline{P}^i + (1 - \alpha) \underline{P}^u, \quad \bar{P}_\alpha = \alpha \bar{P}^i + (1 - \alpha) \bar{P}^u, \quad (7)$$

and as lower and upper predictive probabilities for the event that the next Bernoulli random quantity is a success<sup>9</sup>

$$\underline{P} = \min_{\alpha \in [0, 1]} \underline{P}_\alpha \quad \text{and} \quad \bar{P} = \max_{\alpha \in [0, 1]} \bar{P}_\alpha.$$

<sup>8</sup>See also [www.npi-statistics.com](http://www.npi-statistics.com).

<sup>9</sup>While in (2) and (3), prior and sample information are imprecisely weighted, here informative and uninformative models are combined.

Allowing  $\alpha$  to take on any value in  $[0, 1]$  reduces this method to the IBBM with a single prior set, as discussed in Section 2, with the prior set simply generated by the union of the two prior sets for the ‘informative’ and the ‘uninformative’ models as described above. For all  $s$  these minimum and maximum values are obtained at either  $\alpha = 0$  or  $\alpha = 1$ . With switch points  $S_1 = (n+1)\frac{s^i}{n^i} - 1$  and  $S_2 = (n+1)\frac{s^i}{n^i}$ , they are equal to

$$\underline{P} = \begin{cases} \underline{P}^u = \frac{s}{n+1} & \text{if } s \leq S_2 \\ \underline{P}^i = \frac{s^i+s}{n^i+n+1} & \text{if } s \geq S_2, \end{cases}$$

$$\bar{P} = \begin{cases} \bar{P}^i = \frac{s^i+s+1}{n^i+n+1} & \text{if } s \leq S_1 \\ \bar{P}^u = \frac{s+1}{n+1} & \text{if } s \geq S_1. \end{cases}$$

The PPP graph for this model is displayed in Figure 6. The upper probability for  $s = S_1$  and the lower probability for  $s = S_2$  are both equal to  $\frac{s^i}{n^i}$ . The TPDA contains only a single possible value of  $s$  (except if  $S_1$  and  $S_2$  are integer), namely the one that is nearest to  $\frac{s^i}{n^i}$ . The specific values for this basic case are

$$A = 0 \quad B = \frac{s^i + 1}{n^i + n + 1} \quad C = \frac{s^i + n}{n^i + n + 1}$$

$$D = 1 \quad E = \frac{s^i}{n^i} - \frac{1}{n+1} \quad F = \frac{s^i}{n^i} + \frac{1}{n+1}$$

$$\text{sl. 1} = \frac{1}{n^i + n + 1} \quad \text{sl. 2} = \frac{1}{n+1}.$$

If  $s$  is in the TPDA it reflects optimal agreement of the ‘prior data’ ( $n^i, s^i$ ) and the (really observed) data ( $n, s$ ), so it may be a surprise that both the lower and upper probabilities in this case correspond to  $\alpha = 0$ , so they are fully determined by the ‘uninformative’ part of the model. This is an important aspect, it will be discussed in more detail and compared to the methods of Section 2 in Subsection 3.3. For  $s$  in the TPDA both  $\underline{P}$  and  $\bar{P}$  increase with slope  $\frac{1}{n+1}$  and  $\Delta = \frac{1}{n+1}$ .

Figure 6, with the specific values for this basic case given above, illustrates what happens for values of  $s$  outside this TPDA. Moving away from the TPDA in either direction, the imprecision increases as was also the case in the models in Section 2. For  $s$  decreasing towards 0, this is effectively due to the smaller slope of the upper probability, while for  $s$  increasing towards 1 it is due to the smaller slope of the lower probability. For  $s \in [0, S_1]$ , the imprecision is  $\Delta = \frac{s^i+1}{n^i+n+1} - \frac{sn^i}{(n^i+n+1)(n+1)}$ . For  $s \in [S_2, n]$  the imprecision is  $\Delta = \frac{1}{n+1} - \frac{s^i}{n^i+n+1} + \frac{sn^i}{(n^i+n+1)(n+1)}$ . For the two extreme possible cases of prior data conflict, with either  $s^i = n^i$  and  $s = 0$  or  $s^i = 0$  and  $s = n$ , the imprecision is  $\Delta = \frac{n^i+1}{n^i+n+1}$ . For this combined model with  $\alpha \in [0, 1]$ , we have  $\underline{P} \leq \frac{s}{n} \leq \bar{P}$  for all  $s$ , which is attractive from the perspective of objective inference.

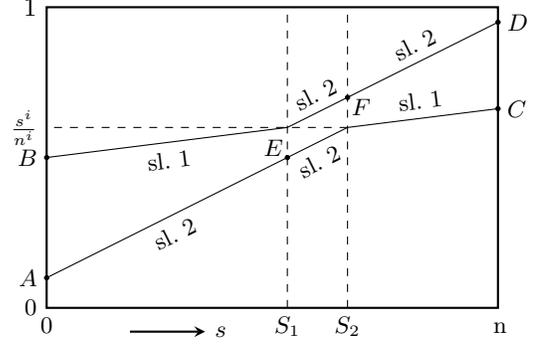


Figure 6:  $\underline{P}$  and  $\bar{P}$  for the weighted inference model.

### 3.2 The Extended Model

We extend the basic model from Subsection 3.1, perhaps remarkably by reducing the interval for the weighting variable  $\alpha$ . We assume that  $\alpha \in [\alpha_l, \alpha_r]$  with  $0 \leq \alpha_l \leq \alpha_r \leq 1$ . We consider this an extended version of the basic model as there are two more parameters that provide increased modelling flexibility. It is important to remark that, with such a restricted interval for the values of  $\alpha$ , this weighted model is no longer identical to an IBBM with a single set of prior distributions. One motivation for this extended model is that the basic model seemed very cautious by not using the informative prior part if  $s$  is in the TPDA. For  $\alpha_l > 0$ , the informative part of the model influences the inferences for all values of  $s$ , including the one in the TPDA. As a consequence of taking  $\alpha_l > 0$ , however, the line segment  $(s, \frac{s}{n})$  with  $s \in [0, n]$  will not always be in between the lower and upper probabilities anymore, specifically not at, and close to,  $s = 0$  and  $s = n$ , as follows from the results presented below.

The lower and upper probabilities resulting from the two models that are combined by taking an imprecise weighted average are again as given by formulae (5)-(6), with the weighted averages  $\underline{P}_\alpha$  and  $\bar{P}_\alpha$ , for any  $\alpha \in [\alpha_l, \alpha_r]$ , again given by (7). This leads to the lower and upper probabilities for the combined inference

$$\underline{P} = \min_{\alpha \in [\alpha_l, \alpha_r]} \underline{P}_\alpha \quad \text{and} \quad \bar{P} = \max_{\alpha \in [\alpha_l, \alpha_r]} \bar{P}_\alpha.$$

The lower and upper probabilities have, as function of  $s$ , the generic forms presented in Figure 6, with  $[S_1, S_2] = \left[ (n+1)\frac{s^i}{n^i} - 1, (n+1)\frac{s^i}{n^i} \right]$  as in Section 3.1. The specific values for Figure 6 are

$$A = \frac{\alpha_l s^i}{n^i + n + 1} \quad B = \frac{1}{n+1} + \frac{\alpha_r [s^i (n+1) - n^i]}{(n^i + n + 1)(n+1)}$$

$$D = 1 - \frac{\alpha_l (n^i - s^i)}{n^i + n + 1} \quad C = \frac{n}{n+1} - \frac{\alpha_r [(n^i - s^i)(n+1) - n^i]}{(n^i + n + 1)(n+1)}$$

$$\begin{aligned} \text{sl. 1} &= \frac{n^i+n+1-\alpha_r n_i}{(n^i+n+1)(n+1)} & E &= \frac{s^i}{n^i} - \frac{1}{n+1} \left[ 1 - \frac{\alpha_l n^i}{n^i+n+1} \right] \\ \text{sl. 2} &= \frac{n^i+n+1-\alpha_l n_i}{(n^i+n+1)(n+1)} & F &= \frac{s^i}{n^i} + \frac{1}{n+1} \left[ 1 - \frac{\alpha_l n^i}{n^i+n+1} \right]. \end{aligned}$$

The increase in imprecision when  $s$  moves away from the TPDA can again be considered as caused by the informative part of the model, which is logical as the uninformative part of the model cannot exhibit prior-data conflict.

The possibility to choose values for  $\alpha_l$  and  $\alpha_r$  provides substantially more modelling flexibility compared to the basic model presented in Section 3.1. One may, for example, wish to enable inferences solely based on the informative part of the model, hence choose  $\alpha_r = 1$ , but ensure that this part has influence on the inferences in all situations, with equal influence to the uninformative part in case of TPDA. This latter aspect can be realized by choosing  $\alpha_l = 0.5$ . When compared to the situation in Section 3.1, this choice moves, in Figure 6,  $A$  and  $D$  away from 0 and 1, respectively, but does not affect  $B$  and  $C$ . It also brings  $E$  and  $F$  a bit closer to the corresponding upper and lower probabilities, respectively, hence reducing imprecision in the TPDA.

### 3.3 Weighted Inference Model Properties

The basic model presented in Section 3.1 fits in the Bayesian framework, but its use of prior information is different to the usual way in Bayesian statistics. The lower and upper probabilities are mainly driven by the uninformative part, which e.g. implies that  $\underline{P} \leq \frac{s}{n} \leq \overline{P}$  for all values of  $s$ . While in (imprecise, generalized) Bayesian statistics any part of the model that uses an informative prior can be regarded as adding information to the data, the informative part of the basic model leads to more careful inferences when there is prior-data conflict. Figure 6 shows that, for the basic case of Section 3.1, the points  $A$  and  $D$  are based only on the uninformative part of the model, but the points  $B$  and  $C$  are based on the informative part of the model.

Prior-data conflict can be of different strength, one would expect to only talk about ‘conflict’ if consideration is required, hence the information in the prior and in the data should be sufficiently strong. The proposed method in Section 3.1 takes as starting point inference that is fully based on the data, it uses the informative prior part of the model to widen the interval of lower and upper probabilities in the direction of the value  $\frac{s^i}{n^i}$ . For example, if one observed  $s = 0$ , the upper probability of a success at the next observation is equal to  $\frac{s^i+1}{n^i+n+1}$ , which reflects inclusion of the information in the prior set for the informative part of the model that is most supportive for this

event, equivalent to  $s^i + 1$  successes in  $n^i + 1$  observations. As such, the effect of the prior information is to weaken the inferences by increasing imprecision in case of prior-data conflict.

One possible way in which to view this weighted inference model is as resulting from a multiple expert or information source problem, where one wishes to combine the inferences resulting individually from each source. The basic model of Section 3.1 leads to the most conservative inference such that no individual model or expert disagrees, while the restriction on weights provides a guaranteed minimum level for the individual contributions to the combined inference.

It should be emphasized that the weighted inference model has wide applicability. The key idea is to combine, by imprecise weighting, the actual inferences resulting from multiple models, and as such there is much scope for the use and further development of this approach. The individual models could even be models such as those described in Sections 2.2 and 2.3, although that would lead to more complications. If the individual models are coherent lower and upper probabilities, i.e. provide separately coherent inferences, then the combined inference via weighted averaging and taking the lower and upper envelopes is also separately coherent<sup>10</sup>.

In applications, it is often important to determine a sample size (or more general design issues) before data are collected. If one uses a model that can react to prior-data conflict, this is likely to lead to a larger data requirement. One very cautious approach is to choose  $n$  such that the maximum possible resulting imprecision does not exceed a chosen threshold. In the models presented in this paper, this maximum imprecision will always occur for either  $s = 0$  or  $s = n$ , whichever is further away from the TPDA. In such cases, a preliminary study has shown an attractive feature if one can actually sample sequentially. If some data are obtained with success proportion close to  $s^i/n^i$ , the total data requirement (including these first observations) to ensure that the resulting maximum imprecision cannot exceed the same threshold level is substantially less than had been the case before any data were available. This would be in line with intuition, and further research into this and related aspects is ongoing, including of course the further data need in case first sampled data is in conflict with  $(n^i, s^i)$ , and the behaviour of the models of Section 2 in such cases.

The weighted inference method combines the inferences based on two models, and can be generalized to allow more than two models and different inferential methods. It is also possible to allow more imprecise

<sup>10</sup>This follows from e.g. [18, 2.6.3f]

sion in each of the models that are combined, leading to more parameters in the overall model that can be used to control the behaviour of the inferences. Similar post-inference combination via weighted averaging, but with precise weights, has been presented in the frequentist statistics literature [11, 13], where the weights are actually determined based on the data and a chosen optimality criterion for the combined inference. In Bayesian statistics, estimation or prediction inferences based on different models can be similarly combined using Bayes factors [12], which are based on both the data (via the likelihood function) and prior weightings for the different models. In our approach, we do not use the data or prior beliefs about the models to derive precise weights for the models, instead we cautiously base our combined lower and upper predictive probabilities on those of the individual models with a range of possible weights. This range is set by the analyst and does not explicitly take the data or prior beliefs into account, but it provides flexibility with regard to the relative importance given to the individual models.

## 4 Insights and Challenges

We have discussed two different classes of inferential methods to handle prior-data conflict in the Bernoulli case. These can be generalized to the multinomial case corresponding to the IDM. It also seems possible to extend the approaches to continuous sampling models like the normal or the gamma distribution, by utilizing the fact that the basic form of the updating of  $n^{(0)}$  and  $y^{(0)}$  in (1) underlying (2) and (3) is valid for arbitrary canonical exponential families [15, 21]. Further insight into the weighting method may also be provided by comparing it to Generalized Bayesian analysis based on sets of conjugate priors consisting of nontrivial mixtures of two Beta distributions. There, however, the posterior mixture parameter depends on the other parameters. For a deeper understanding of prior-data conflict it may also be helpful to extend our methods to coarse data, in an analogous way to [17] and [16], and to look at other model classes of prior distributions, most notably at contamination neighbourhoods. Of particular interest here may be to combine both types of prior models, considering contamination neighbourhoods of our exponential family based-models with sets of parameters, as developed in the Neyman-Pearson setting by [1, Section 5].

The models presented here address prior-data conflict in different ways, either by fully utilizing the prior information in a way that is close to the traditional Bayesian method, where this information is added to data information, or by not including them initially as in Section 3. All these models show the desired in-

crease of imprecision in case of prior-data conflict. It may be of interest to derive methods that explicitly respond to (perhaps surprisingly) strong prior-data agreement. One possibility to achieve this with the methods presented here is to consider the TPDA as this situation of strong agreement in which one wants imprecision reduced further than compared to an ‘expected’ situation, and to choose the prior set (Section 2) or the two inferential models (Section 3) in such a way to create this effect. This raises interesting questions for elicitation, but both approaches provide opportunities for this and we consider it as an important topic for further study.

Far beyond further extensions one has, from the foundational point of view, to be aware that there are many ways in which people might react to prior-data conflict, and we may perhaps at best hope to catch some of these in a specific model and inferential method. This is especially important when the conflict is very strong, and indeed has to be considered as full contradiction of modeling assumptions and data, which may lead to a revision of the whole system of background knowledge in the light of surprising observations, as Hampel argues.<sup>11</sup> In this context applying the weighting approach to the NPI-based model for categorical data [6] may provide some interesting opportunities, as it explicitly allows to consider not yet observed and even undefined categories [5].

There is another intriguing way in which one may react to prior-data conflict, namely by considering the combined information to be of less value than either the real data themselves or than both information sources. Strong prior beliefs about a high success rate could be strongly contradicted by data, as such leading to severe doubt about what is actually going on. The increase of imprecision in case of prior-data conflict in the methods presented in this paper might be interpreted as reflecting this, but there may be other opportunities to model such an effect. It may be possible to link these methods to some popular approaches in frequentist statistics, where some robustness can be achieved or where variability of inferences can be studied by round robin deletion of some of the real observations. This idea may open up interesting research challenges for imprecise probability models, where the extent of data reduction could perhaps be related to the level of prior-data conflict. Of course, such approaches would only be of use in situations with substantial amounts of real data, but as mentioned before these are typically the situations where prior-data conflict is most likely to be of sufficient relevance to take its modelling seriously. As

<sup>11</sup>See in particular the discussion of the structure and role of background knowledge in [10].

(imprecise, generalized) Bayesian methods all work essentially by *adding* information to the real data, it is unlikely that such new methods can be developed within the Bayesian framework, although there may be opportunities if one restricts the inferences to situations where one has at least a pre-determined number of observations to ensure that posterior inferences are proper. For example, one could consider allowing the prior strength parameter  $n^{(0)}$  in the IBBM to take on negative values, opening up a rich field for research and discussions.

## Acknowledgements

We thank the referees for very helpful comments.

## References

- [1] T. Augustin. Neyman-Pearson testing under interval probability by globally least favorable pairs – Reviewing Huber-Strassen theory and extending it to general interval probability. *Journal of Statistical Planning and Inference*, 105(1):149–173, 2002.
- [2] T. Augustin, F.P.A. Coolen, S. Moral, and M.C.M. Troffaes, editors. *ISIPTA '09: Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*. SIPTA, 2009.
- [3] J.-M. Bernard. Special Issue on the Imprecise Dirichlet Model. *International Journal of Approximate Reasoning*, 50:201–268, 2009.
- [4] F. P. A. Coolen. Low structure imprecise predictive inference for Bayes' problem. *Statistics & Probability Letters*, 36:349–357, 1998.
- [5] F.P.A. Coolen and T. Augustin. Learning from multinomial data: a nonparametric predictive alternative to the Imprecise Dirichlet Model. In F.G. Cozman, R. Nau, and T. Seidenfeld, editors, *ISIPTA '05: Proc. of the Fourth International Symposium on Imprecise Probabilities and their Applications*, pages 125–135, 2005.
- [6] F.P.A. Coolen and T. Augustin. A nonparametric predictive alternative to the Imprecise Dirichlet Model: the case of a known number of categories. *International Journal of Approximate Reasoning*, 50:217–230, 2009.
- [7] Frank P. A. Coolen. On Bernoulli experiments with imprecise prior probabilities. *The Statistician*, 43:155–167, 1994.
- [8] G. de Cooman, J. Vejnarová, and M. Zaffalon, editors. *ISIPTA '07: Proceedings of the Fifth International Symposium on Imprecise Probabilities and Their Applications*. SIPTA, 2007.
- [9] M. Evans and H. Moshonov. Checking for prior-data conflict. *Bayesian Analysis*, 1:893–914, 2006.
- [10] F. Hampel. How can we get new knowledge? In T. Augustin, F.P.A. Coolen, S. Moral, and M.C.M. Troffaes, editors, *ISIPTA '09: Proc. of the Sixth International Symposium on Imprecise Probabilities: Theories and Applications*, pages 219–227, 2009.
- [11] N.L. Hjort and G. Claeskens. Frequentist model average estimators. *Journal of the American Statistical Association*, 98:879–899, 2003.
- [12] R.E. Kass and A.E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90:773–795, 1995.
- [13] N.T. Longford. An alternative to model selection in ordinary regression. *Statistics and Computing*, 13:67–80, 2003.
- [14] L. P. Pericchi and P. Walley. Robust Bayesian credible intervals and prior ignorance. *International Statistical Review*, 58:1–23, 1991.
- [15] E. Quaeghebeur and G. de Cooman. Imprecise probability models for inference in exponential families. In F.G. Cozman, R. Nau, and T. Seidenfeld, editors, *ISIPTA '05. Proc. of the Fourth International Symposium on Imprecise Probabilities and Their Applications*, pages 287–296, 2005.
- [16] M.C.M. Troffaes and F.P.A. Coolen. Applying the imprecise Dirichlet model in cases with partial observations and dependencies in failure data. *International Journal of Approximate Reasoning*, 50(2):257–268, 2009.
- [17] L.V. Utkin and T. Augustin. Decision making under imperfect measurement using the imprecise Dirichlet model. *International Journal of Approximate Reasoning*, 44(3):322–338, 2007.
- [18] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.
- [19] P. Walley. Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B*, 58:3–57, 1996.
- [20] P. Walley and J.-M. Bernard. Imprecise probabilistic prediction for categorical data. Technical Report CAF-9901, Paris 8, 1999.
- [21] G. Walter and T. Augustin. Imprecision and prior-data conflict in generalized Bayesian inference. *Journal of Statistical Theory and Practice*, 3:255–271, 2009.
- [22] G. Walter, T. Augustin, and A. Peters. Linear regression analysis under sets of conjugate priors. In G. de Cooman, J. Vejnarová, and M. Zaffalon, editors, *ISIPTA '07: Proc. of the Fifth International Symposium on Imprecise Probabilities: Theories and Applications*, pages 445–455, 2007.
- [23] K.M. Whitcomb. Quasi-Bayesian analysis using imprecise probability assessments and the generalized Bayes' rule. *Theory and Decision*, 58:209–238, 2005.