Posterior previsions for the parameter of a binomial model via natural extension of a finite number of judgments

Vincent Fortin
Hydro-Québec's research institute (IREQ), Varennes, Québec
vinfort@ireq.ca

Éric Parent
French Institute of Forestry, Agricultural and Environmental Engineering (ENGREF), Paris
parent@engref.fr

Bernard Bobée
Hydro-Québec/NSERC/Alcan Industrial Chair in Statistical Hydrology at INRS-Eau, Sainte-Foy, Québec, Canada
bernard_bobee@inrs-eau.uquebec.ca

Abstract

In this paper, we investigate the use of assessments of conditional previsions for modeling prior information on the parameter of a binomial model as a way of obtaining non-vacuous posterior previsions via natural extension. More specifically, we argue that a useful method for obtaining an imprecise prevision for the parameter θ of a binomial model, given a sample of size n showing r successes, is to assess imprecise previsions for θ which are conditional on samples having sizes larger than n. Inferences obtained using this approach are compared to Walley's proposal for learning from a bag of marbles.

Keywords: conditional lower previsions, prior information modeling, natural extension, generalized Bayes rule, binomial model

1 Introduction

Even if the theory of coherent lower previsions (CLP, *Walley* [1991]) is quite appealing for decision-making under uncertainty, there are still few published applications of this theory, in part because inferences obtained using natural extension tend to be excessively imprecise from the point of view of the practitioner. As was pointed out by *Walley* [1996b], this occurs when the assumptions and prior judgments are themselves excessively imprecise. However, one of the appealing features of CLP theory is the possibility of making inferences and taking rational decisions based on few other assumptions than coherence. If one still wants to base inferences on natural extension, the challenge therefore resides in developing, for various statistical models, strategies for encoding prior information efficiently, in the sense that few assessments of imprecise previsions be necessary to obtain useful results.

The purpose of this paper is to propose and illustrate a simple method for using natural extension to obtain imprecise, but still useful, coherent posterior previsions for the parameter of a binomial model from a small number of prior judgments. It aims at showing the advantages of CLP theory, and in particular of natural extension, for the type of problems encountered by the practitioner who has to make inferences, and eventually take action, based on small samples.

Various inference frameworks that have been proposed for the binomial model are reviewed and discussed in Section 2. In Section 3, we propose a different approach, based on a finite number of assessments of the conditional expectation of the parameter of the binomial model for sample sizes larger than the one at hand. These ideas are illustrated in Section 4, and compared to the approach for learning from a bag of marbles proposed by *Walley* [1996a]. Section 5 contains a general discussion and some conclusions.

2 Inference from Binomial Data

Let (X_n) be a sequence of i.i.d. Bernoulli random variable that can only assume the values 1 (a success) or 0 (a failure). Let $\theta=P(X_n=1)$ be the probability of observing a success. It can be shown that the probability of observing r successes in n trials has a binomial distribution $f(r,n;\theta)=n!/[r!(n-r)!]\cdot\theta^r\cdot(1-\theta)^{n-r}$. In an inference setting, one is interested in obtaining information on the parameter θ from a finite observed sequence $\mathbf{x}=(x_n)$ containing $r=\sum x_n$ successes.

The Bayesian approach to this problem is to summarize all prior information on this parameter by a prior distribution $b(\theta)$ and to update it, using Bayes' rule, to obtain a posterior distribution $b(\theta|r,n)\propto f(r,n;\theta)\cdot b(\theta)$ for the parameter of interest. The difficulty lies with the elicitation of the prior distribution, especially when there is limited prior and sampling information. In this case, different *reference* or *non-informative* priors have been proposed based on different criteria (see *Bernardo and Smith* [1994] for a Bayesian perspective on this problem). Non-informative priors proposed in the literature have two things in common: they are relatively smooth and symmetrical, thus leading to inferences that are dominated by the effect of the sample, even if the prior expectation of θ is precisely ½ in all cases.

The limitations of Bayesian theory for dealing with near-ignorance are treated in the literature either (1) by introducing new theories for modeling uncertainty (including theories of imprecise probabilities), or (2) by arguing that for practical purposes these limitations can be overcome by a proper sensitivity analysis (see for example *Berger* [1985]). Given these results, and with limited comprehension of the fundamental limitations of Bayesian theory, the practitioner will likely only check the sensitivity of inferences obtained through Bayes' rule by trying out different priors. In some cases, this leads to results that are fundamentally wrong.

For example, in an otherwise excellent paper on an application of Dempster-Shafer theory of evidence to a hydrological engineering problem, *Caselton and Luo* [1992] illustrate Bayesian decision theory by considering the problem of designing a highway drainage culvert in a simplified setting where only two design flow values are considered, Q_1 and Q_2, where $Q_1<Q_2$. They then express the expected utility of each decision Q_i as a linear function of the probability of exceedance $\theta_i = P(Q>Q_i)$, modeling their near-ignorance on the values of θ_i by non-informative priors. They then show that the decision is actually not sensitive to the prior, even in the case where there is no sampling information, and that the smaller culvert should be built.

But by using only symmetrical priors, they assumed in all cases that $E(\theta_i)=½$, which implies that $E[P(Q>Q_1)]=E[P(Q>Q_2)]$ and that consequently $E[P(Q_1<Q\leq Q_2)]=0$. This unreasonable hypothesis amounts to supposing that, prior to any observation, the decision-maker (DM) feels that flood flows between Q_1 and Q_2 are impossible, which then explains why a Bayesian decision analysis concludes that, under a wide range of priors, there is no reason to pay for additional protection against floods of this magnitude. As this example shows, the limitations of Bayesian decision analysis for dealing with near-ignorance have important implications for the practitioner, and cannot be avoided by a routine sensitivity analysis.

Walley [1996a] proposed an imprecise Dirichlet-multinomial model (IDM) that is appropriate to characterize near-ignorance and which can be used to obtain coherent posterior previsions once a random sample is observed. In the particular case of the binomial model, this model amounts to using for inference the lower envelope of the precise previsions generated by the family of beta priors $\{b(\theta;s,t);\ t\in[0,1]\}$, where s is a positive real number specified by the DM and $b(\theta;s,t)$ is the p.d.f. of the beta distribution:

$$b(\theta) \propto \theta^{st-1}(1-\theta)^{s-st-1} \qquad [1]$$

We will refer to this model as the imprecise beta-binomial (IBB) model. It can be shown that the IBB model generates vacuous prior previsions for θ, i.e. $\underline{P}(\theta)=0$ and $\bar{P}(\theta)=1$, as well as the following non-vacuous posterior previsions:

$$\begin{aligned} \underline{P}(\theta \mid r,n) &= \frac{r}{n+s} \\ \bar{P}(\theta \mid r,n) &= \frac{r+s}{n+s} \end{aligned} \qquad [2]$$

where $\underline{P}(\theta|r,n)$ and $\bar{P}(\theta|r,n)$ correspond respectively to the lower and upper previsions of the gamble θ conditional on the observation of r successes in n Bernoulli trials. The parameter s can therefore be interpreted as the weight of the prior information, relative to the size of the random sample available for inference. In practice, it can be derived from the assessment of a single conditional prevision. Note that the lower and upper expected values obtained via Dempster's rule for the binomial model correspond to the particular case s=1 (see *Dempster* [1968] and *Dempster and Kong* [1987]).

This is a tractable model, which represents adequately near-ignorance and leads to non-vacuous coherent posterior previsions. However, it still is a parametric model: previsions are completely determined by the value of the parameter s: by choosing a value of s, the DM specifies buying and selling prices for the gamble θ, for every possible value of r and n. In the next section we discuss how inferences can be obtained by assessing conditional previsions for only a few values of r and n.

3 Extending a Finite Set of Judgments

One way of loosening the strong parametric hypothesis of the IBB model is to assess only a finite number of conditional previsions for θ. If a DM may not be comfortable with the idea of setting prices for all conditional gambles, he may still be ready to set prices for some conditional gambles. In this section, we discuss the evaluation, via natural extension, of the imprecise probability of observing r successes in n trials, given a finite number of assessments of coherent conditional imprecise prevision. But first, let us define what is meant by natural extension from a finite set of judgments.

3.1 Natural Extension

This section summarizes material that is found in *Walley* [1991] and *Walley* [1996b] on natural extension from a finite set of judgments.

Consider a finite set $\mathcal{D}$ of k judgments which can be translated into constraints on conditional lower previsions, i.e. $\mathcal{D}=\{\underline{P}(X_j\,|B_j)\geq\mu_j, j=1...k\}$ where X_j are gambles (i.e. real-valued bounded functions) defined on a possibility space Θ and B_j are subsets of Θ. Note that constraints on conditional upper previsions can be defined in terms of lower previsions since $\bar{P}(X \mid B)=\mu \Leftrightarrow \underline{P}(-X \mid B)=-\mu$ by definition of an upper prevision. *Walley* [1996b] defines the natural extension of $\underline{P}$ to any conditional lower prevision (eq. 1, p. 15):

$$\underline{E}(X \mid B)=\sup\{\mu:(\exists\boldsymbol{\lambda}\geq 0) \text{ s.t. } B(X-\mu)\geq S(\boldsymbol{\lambda})\} \quad [3]$$

where $\lambda=(\lambda_1,\lambda_2,...,\lambda_k)^{\mathrm{T}}$, $\lambda\geq 0$ means $\max(\lambda_j)\geq 0$, $S(\boldsymbol{\lambda})=\sum_{j=1}^{k}\lambda_j B_j(X_j-\mu_j)$, and finally $X\geq Y$ means $X(\theta)\geq Y(\theta)$ for every $\theta\in\Theta$. Note that on the left-hand side of this equation, B denotes a subset B of the possibility space Θ, whereas on the right-hand side of the same equation, $B(\theta)$ corresponds to the indicator function of the same subset. To simplify the notation, a set and its indicator function will always be denoted by the same symbol.

According to *Walley* [1996b], p. 16, when the initial judgments avoid sure loss, the natural extensions $\underline{E}$ are the minimal coherent lower previsions that satisfy the initial constraints, meaning that $\underline{E}(X_j\,|B_j)=\mu_j$ and that any other coherent lower prevision $\underline{P}^*$ that satisfy the initial constraints must dominate $\underline{E}$, i.e. $\underline{P}^*(X|B)\geq\underline{E}(X|B)$ for all gambles X and events B. For unconditional lower previsions, eq. [3] is equivalent to [*Walley*, 1991, lemma 3.1.3(a)]:

$$\underline{E}(X)=\sup_{\boldsymbol{\lambda}\geq 0}\left\{\inf_{\theta\in\Theta}\left[X-S(\boldsymbol{\lambda})\right]\right\} \quad [4]$$

In practice, it is easier to solve eq. [4] than eq. [3]. The generalized Bayes rule, presented in the next section makes it possible to derive conditional previsions from unconditional previsions.

3.2 The Generalized Bayes Rule

When $\underline{E}(B)>0$, $\underline{E}(X|B)$ can also be obtained as the supremum of μ such that $\underline{E}(B(X-\mu))\geq 0$:

$$\underline{E}(X \mid B)=\sup\left\{\mu:\underline{E}\left[B\cdot\left(X-\mu\right)\right]\geq 0\right\} \quad [5]$$

Proof. Since $\underline{E}$ is coherent, by definition $\underline{E}(X+Y)\geq\underline{E}(X)+\underline{E}(Y)$ and $\underline{E}(\delta X)=\delta\underline{E}(X)$ for every $\delta>0$. Choose μ such that $\underline{E}(B(X-\mu))\geq 0$. It follows that for every $\delta>0$, $\underline{E}(B(X-\mu+\delta))\geq\underline{E}(B(X-\mu+\delta))+\delta\underline{E}(B)$. By hypothesis $\underline{E}(B)>0$, therefore, $\underline{E}(B(X-\nu))>0$, where $\nu=\mu-\delta$. By definition of $\underline{E}(B(X-\nu))$ (eq. [3]), this implies that $\sup\{\eta:(\exists\lambda\geq 0) \text{ s.t. } B(X-\nu)-\eta\geq S(\lambda)\}>0$, which in turn implies that $(\exists\lambda\geq 0)$ s.t. $B(X-\nu)\geq S(\lambda)$. Since $\underline{E}(X|B)=\sup\{\mu:(\exists\lambda\geq 0) \text{ s.t. } B(X-\mu)\geq S(\lambda)\}$, it follows that $\underline{E}(X|B)\geq\nu=\mu-\delta$. Hence, for any μ such that $\underline{E}(B(X-\mu))\geq 0$, $\underline{E}(X|B)\geq\mu-\delta$ for every $\delta>0$, which is equivalent to $\underline{E}(X|B)\geq\sup\{\mu: \underline{E}[B(X-\mu)]\geq 0\}$.

For the reverse inequality, consider that $B(X-\mu)\geq S(\lambda)$ is equivalent to $\inf[B(X-\mu)-S(\lambda)]\geq 0$. Consequently, $\underline{E}(X|B)=\sup\{\mu:(\exists\lambda\geq 0) \text{ s.t. } \inf[B(X-\mu)-S(\lambda)]\geq 0\}$. Also, $(\exists\lambda\geq 0)$ s.t. $\inf[B(X-\mu)-S(\lambda)]\geq 0$ implies that $\underline{E}[B(X-\mu)]=\sup\{\inf[B(X-\mu)-S(\lambda)]\geq 0\}\geq 0$. Hence, $\{\mu:(\exists\lambda\geq 0) \text{ s.t. } \inf[B(X-\mu)-S(\lambda)]\geq 0\}\subseteq\{\mu:\underline{E}[B(X-\mu)]\geq 0\}$ and $\underline{E}(X|B)\leq\sup\{\mu:\underline{E}[B(X-\mu)]\geq 0\}$.♦

Equation [5] is known as the generalized Bayes rule (GBR). Natural extension and GBR are mainly justified in *Walley* [1991] by considerations of coherence between two-stage gambles $\underline{P}(\cdot|\mathbf{B})$ defined on different partitions $\mathbf{B}$ of Θ (see section 8.1.4 of *Walley* [1991]). Here, we do not require that the sets B_j make up one or more partitions of Θ. However, it is possible to define for each B_j the partition $\mathbf{B}_j=\{B_j,B_j^c\}$ and make the additional judgment that $\underline{P}(X_j|B_j^c)\geq\inf\{X_j: B_j^c>0\}$. As this is an implicit requirement for separate coherence of $\underline{P}(\cdot|\mathbf{B}_j)$, this would however lead exactly to the same inferences.

Proof. Adding any judgment of the type $\underline{P}(X_0|B_0)\geq\inf\{X_0: B_0>0\}$ increases $S(\lambda)$ to $S(\lambda_0,\lambda)=S(\lambda)+\lambda_0 B_0[X_0-\inf\{X_0: B_0>0\}]$. Since $\lambda_0\geq 0$, $B_0\geq 0$ and $X_0\geq\inf\{X_0: B_0>0\}$ whenever $B_0>0$, it follows that $S(\lambda)\leq S(\lambda_0,\lambda)$. Consider the values of μ over which the supremum is taken in eq. [3]. This set does not change when $S(\lambda)$ is replaced by $S(\lambda_0,\lambda)$. Indeed, if μ belongs to this set, i.e. $B(X-\mu)$ dominates $S(\lambda)$ for some $\lambda\geq 0$, then $B(X-\mu)$ also dominates $S(0,\lambda)$ since $S(0,\lambda)=S(\lambda)$. On the contrary, if for every $\lambda\geq 0$ there exists $\theta\in\Theta$ such that $B(\theta)(X(\theta)-\mu)<S(\lambda)(\theta)$, then, for the same value of θ, $B(\theta)(X(\theta)-\mu)<S(\lambda_0,\lambda)(\theta)$ for every $\lambda_0>0$, since $S(\lambda)\leq S(\lambda_0,\lambda)$ for every $\lambda_0>0$.♦

3.3 Inferences Based on a Precise Sampling Model

In an inference setting, one does not observe a subset of the parameter space, but an observation $x\in\Omega$ related to θ by a sampling model. We consider in this section the case where the sampling model $f(x;\theta)$ is precise and the sampling space is discrete, so that $f(x;\theta)$ corresponds to the probability of observing $\{x\}$ given θ. Define $P(X|\Theta)$ as the price one is willing to pay to buy or sell the two-stage gamble in which one learns which θ has occurred before stating his price for the gamble X. This is a function of θ related to $f(x;\theta)$ by $P(X|\Theta)(\theta)=E^{f(x;\theta)}(X)$, which can be regarded as a gamble defined on Θ. In particular,

$P(\{x\}|\Theta) = f(x;\theta)$ is the likelihood function of the sample x. Considerations of coherence over the product-space $\Theta\times\Omega$ suggest that [*Walley*, 1991, section 8.3]:

$$\underline{E}(X) = \underline{E}\left[P(X \mid \Theta)\right] \qquad [6]$$

Natural extension to conditional lower previsions can still be obtained via the GBR by replacing B by $\{x\}$: $\underline{E}(X|x)=\sup\{\mu: \underline{E}[\{x\}(X-\mu)]\geq 0\}$ [*Walley*, 1991, section 8.4]. According to eq. [6], when X is constant over Ω, $\underline{E}[\{x\}\cdot(X-\mu)]=\underline{E}[P(\{x\}\cdot(X-\mu)|\Theta)]=\underline{E}[P(\{x\}|\Theta)\cdot(X-\mu)]= \underline{E}[B\cdot(X-\mu)]$, where $B(\theta)=P(\{x\}|\Theta)$ is the likelihood function. Therefore, inferences about θ depend only on the data through the likelihood function, and equations [3] to [5] still hold when the indicator functions B and B_j are replaced a discrete likelihood function.

3.4 A Set of Judgments for the Binomial Model

Let $\mathcal{D}=\{\underline{P}(a_j\cdot\theta\,|r_j, N_j)\geq a_j\cdot\mu_j, j=1...k\}$ denote a finite set of judgments where $a_j\in\{-1,1\}$, r_j, and N_j are nonnegative natural numbers such that $r_j\leq N_j$, and μ_j is a real number specified by the DM. When a_j=1, μ_j corresponds to a lower bound for the lower prevision $\underline{P}(\theta\,|r_j, N_j)$. When a_j=-1, μ_j corresponds instead to an upper bound for the upper prevision $\overline{P}(\theta \mid r_j, N_j)$ since $\overline{P}(\theta \mid r_j, N_j) \leq \mu_j \Leftrightarrow \underline{P}(-\theta \mid r_j, N_j) \geq -\mu_j$.

To simplify numerical comparisons with the IBB model in the next section and also insure the coherence of $\mathcal{D}$, we will suppose that there exists a real number s>0 such that eq. [7] holds for j=1...k:

$$\mu_j = \begin{cases} \dfrac{r_j}{N_j+s} & \text{if} \quad a_j = 1 \\[2ex] \dfrac{r_j+s}{N_j+s} & \text{if} \quad a_j = -1 \end{cases} \qquad [7]$$

In this case, it can be seen that the IBB model is consistent with the judgments in $\mathcal{D}$, in the sense that, for the same value of s, every prior in the set $\{b(\theta;s,t);\, t\in[0,1]\}$ generate inferences that do not violate any constraint defined in $\mathcal{D}$. This is readily seen by comparing [7] and [2]. Since natural extension leads to minimal coherent lower previsions, the inferences obtained by applying the GBR will be less precise or equivalent to those obtained with the IBB model.

In practice, we believe that a DM would be more inclined to assess conditional previsions for larger sample sizes N_j. For example, there should be in most practical cases a sample size for which the DM will be inclined to accept the observed frequency as his bet on the probability of success. This is well illustrated by the popularity of the bootstrap [*Efron*, 1979] for statistical simulation purposes, even for moderate sample sizes [see for example *Friedman and Friedman*, 1995]. Indeed, from a Bayesian point of view, resampling with replacement from an observed random sample (x_n) amounts to assessing that the predictive distribution of a single observation is given by the empirical distribution of the sample [*Rubin*, 1981]. This is fortunate, as we will show that by assessing only a few conditional previsions for larger sample sizes, one can obtain non-vacuous posterior previsions from an available smaller sample.

Setting $B_j=f(r_j,N_j;\theta)$, $X=Z$ and $B=f(r,n;\theta)$ in equations [4] and [5], we obtain the following expressions for the natural extension of $\mathcal{D}$ and the GBR:

$$\underline{E}(Z) = \sup_{\boldsymbol{\lambda}\geq 0}\left\{\inf_{\theta\in[0,1]} h\left(\theta,\boldsymbol{\lambda},Z,\mathcal{D}\right)\right\} \qquad [8]$$

$$h\left(\theta,\boldsymbol{\lambda},Z,\mathcal{D}\right) = Z(\theta) - \sum_{j=1}^{k}\lambda_j\cdot f(r_j,N_j;\theta)\cdot a_j\cdot\left(\theta-\mu_j\right)$$

$$\underline{E}(X \mid r,n) = \sup\left\{\mu : \underline{E}\left[\,f(r,n;\theta)\cdot\left(X-\mu\right)\right]\geq 0\right\} \qquad [9]$$

3.5 Conditioning on events of probability zero

Prior to using the GBR for computing $\underline{E}(X|r,n)$, it must be checked that $\underline{E}(f(r,n;\theta)]>0$. Otherwise, $\underline{E}(X|r,n)$ is vacuous, i.e. $\underline{E}(X|r,n) = \inf\{X\text{: } f(r,n;\theta)>0] = \inf X$ (see section 6.10 of *Walley* [1991]).

3.6 Solving the GBR Numerically

Solving the GBR numerically can be a challenge, especially when lower previsions cannot be easily computed [*Cozman*, 1999b]. In the case where the gamble $X(\theta)$ is a polynomial, $Z(\theta)=f(r,n;\theta)\cdot(X(\theta)-\mu)$ and $h(\theta,\lambda,Z,\mathcal{D})$ are also polynomials in θ. Computing $\inf h(\theta,\lambda,Z,\mathcal{D})$ is then generally straightforward. Evaluating the supremum is more difficult because the partial derivatives with respect to each λ_j do not exist everywhere, and in particular they often do not exist at the supremum. Techniques developed for solving Lagrangian functions, which exhibit similar features, can be applied [see ch.8 of *Lasdon*, 1970]. We have used with success the tangential approximation, first proposed by *Geoffrion* [1968].

4 Numerical Results

In this section we illustrate the proposed approach, and compare it to the IBB model. As a first example, consider the case where the DM makes only two assessments: $\underline{P}(\theta \mid 3,5) \geq 1/2$ and $\overline{P}(\theta \mid 3,5) \leq 2/3$.

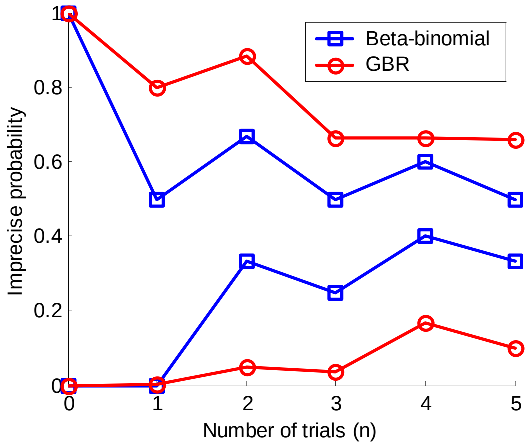

Figure 1: Posterior previsions for θ, given $r=\lfloor n/2 \rfloor$ successes in n trials, based on the judgments $\underline{P}(\theta \mid 3,5) \geq 1/2$ and $\overline{P}(\theta \mid 3,5) \leq 2/3$

This information can be used to derive, via natural extension, useful posterior previsions for smaller values of n, or it can be used to estimate the parameter s of the IBB model. Solving for s in eq. [2] leads to s=1. Figure 1 compares the imprecise probability of success obtained with both methods, given a sample of size $n \leq 5$ in which $r=\lfloor n/2 \rfloor$ successes have been observed.

As expected, the IBB model gives more precise previsions. However, natural extension of these two judgments leads to useful, non-vacuous previsions. The results can be used to assess the impacts of using the more tractable beta-binomial model, when the parameter s has been estimated from only two judgments of conditional previsions.

Compared to the IBB model, Figure 1 shows that natural extension from a finite number of judgments has one important drawback: the imprecision of the posterior probability of θ does not necessarily decrease with the sample size. Indeed, the difference between $\overline{P}(\theta \mid r,n)$ and $\underline{P}(\theta \mid r,n)$ is larger for n=5 than for n=4. This cannot happen with the IBB model. Indeed, using eq. [2], $\overline{P}(\theta \mid r,n) - \underline{P}(\theta \mid r,n) = s/(n+s)$ for the IBB model. Consequently, the imprecision on the posterior prevision of θ is strictly decreasing with the sample size. In fact, coherent lower previsions in general are subject to *dilation*: it is often possible to choose a gamble X and a partition $\mathbf{B}=\{B_1,B_2,\ldots,B_k\}$ for which the posterior previsions are more imprecise than the prior previsions, whatever B_j occurs (see *Wasserman and Seidenfeld* [1996] for an illustration of this phenomenon for the IDM model).

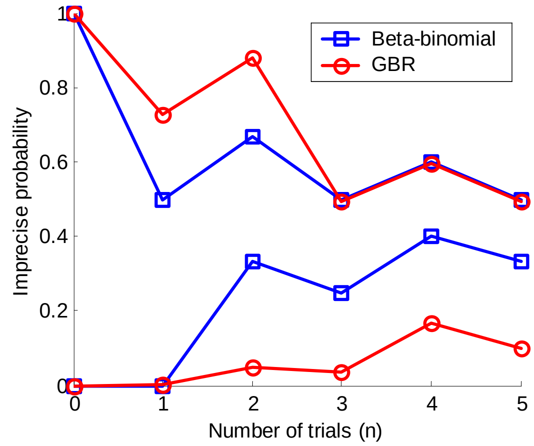

Figure 2: Posterior previsions for θ, given $r=\lfloor n/2 \rfloor$ successes in n trials, based on the judgments $\{\underline{P}(\theta \mid r,5) \geq 1/2; 3 \leq r \leq 5\}$ and $\{\overline{P}(\theta \mid r,5) \leq 2/3; r \leq 3\}$

In practice, the fact that the imprecision of posterior previsions increases with the sample size is generally an indication that the constraints used to derive the previsions via natural extension should be revisited. It is often possible for the DM to add some constraints reflecting prior beliefs that he had not taken the time to assess previously.

In this particular case, there are some judgments that should be easy to make for the DM. Indeed, if a DM makes the judgment $\underline{P}(\theta|r_0,N_0) \geq \mu_0$, meaning that he is ready to pay at least μ_0 for the gamble θ after having observed r_0 successes in N_0 trials, he should be ready in most cases to pay at least the same amount if he observes more than r_0 successes in the same number of trials. Thus, the DM can generally replace the single constraint $\underline{P}(\theta|r_0,N_0) \geq \mu_0$ by the set of judgments $\{\underline{P}(\theta|r,N_0) \geq \mu_0;\ r_0 \leq r \leq N_0\}$. Similarly, after making the judgment $\overline{P}(\theta \mid r_0, N_0) \leq \mu_0$, a DM should generally be inclined to formulate the set of judgments $\{\overline{P}(\theta \mid r, N_0) \leq \mu_0;\ 0 \leq r \leq r_0\}$. In the previous example, by accepting this line of reasoning, the DM would add two judgments of lower previsions, and three judgments of upper previsions.

As shown by Figure 2, this leads to more precise posterior previsions (we have observed that the effect of adding these constraints becomes more important as N_0 increases). In particular, the upper prevision is now equal to the value obtained with the IBB model for $r \geq 3$. Also, the difference between the bounds of both models is generally larger for the lower prevision than for the upper prevision. Could this be caused by the fact that four judgments of upper previsions are made compared to three judgments of lower previsions?

To investigate this effect further, we computed $\underline{E}(\theta|1,2)$ and $\overline{E}(\theta\,|\,1,2)$ based on the set of judgments $\{\underline{P}(\theta|r,5)\geq r_0/6;\ r\geq r_0\}$ and $\{\overline{P}(\theta\,|\,r,5)\leq(r_0+1)/6;\ r\leq r_0\}$, for values of r_0 between 0 and 5. Note that the case studied previously corresponds to $r_0=3$, and that the total number of judgments is always equal to seven. However, the number of judgments of upper previsions increases with r_0.

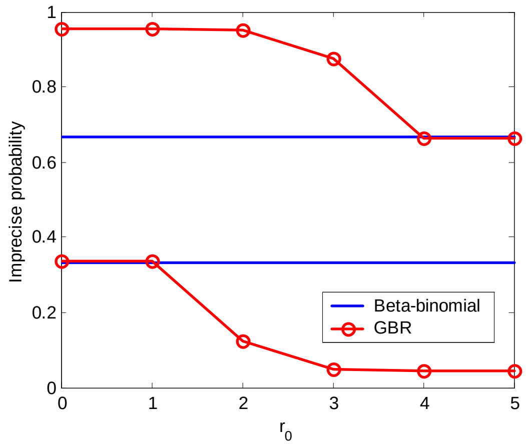

Figure 3: Posterior previsions for θ, given one success in two trials, based on the judgments $\left\{\underline{P}(\theta\,|\,r,5)\geq r_0/6;\ r\geq r_0\right\}$ and $\left\{\overline{P}(\theta\,|\,r,5)\leq(r_0+1)/6;\ r\leq r_0\right\}$

Figure 3 confirms that the lower prevision is closer to zero when the number of judgments of lower previsions decreases (i.e. with increasing r_0), and that the upper prevision is closer to one when the number of judgments of lower previsions increases (i.e. with decreasing r_0). Furthermore, it can be seen that the degree of imprecision, measured by the difference between the upper and lower prevision, increases as the ratio r_0/N_0 approaches ½. This means that an assessment of how one would react if he were to observe either only successes or only failures is somewhat more valuable that an assessment of how one would react if he were to observe an equal number of successes and failures.

5 Discussion and Conclusions

The theory of coherent lower previsions appears to be an adequate model for uncertainty in a wide range of applications [*Walley*, 2000]. One of the apparent drawbacks of the model is the fact that natural extension, which is at the heart of this theory often yields inferences that are too uninformative to be useful. In practice, there are mainly three ways of dealing with this problem: (1) use a different rule for deriving posterior previsions, which sacrifices coherence for some other less stringent consistency criteria [*Walley and de Cooman*, 1999], (2) use a set of probability measures as the basic model for uncertainty [*Cozman*, 1999a] or as a useful hypothesis for obtaining more precise results [*Walley*, 1996a], or (3) directly assess conditional probabilities. In this paper, we investigated this third approach for the binomial model.

The ability to evaluate the implications of a finite number of imprecise judgments is a significant practical advantage of CLP theory over competing models of uncertainty. However, in practice, we need to find guidelines as to which types of conditional previsions are easier to assess for the decision-maker and also lead to useful inferences.

When the objective is to make inferences on the parameter of a statistical model based on a small random sample, we showed that a promising approach is to assess conditional previsions for larger sample sizes, both because (1) it seems possible to actually obtain these types of assessments from a decision-maker, and (2) it leads to inferences which are precise enough to be useful.

A comparison with the imprecise beta-binomial model proposed by *Walley* [1996a] showed that the IBB model can lead to much more precise inferences that what can be obtained from natural extension of the few judgments of conditional previsions used to estimate the parameter *s* of the model.

CLP theory should turn out to be useful for the practitioner, for it is not unusual in engineering, and in particular in environmental engineering, to base important decisions on limited sampling information from different sources and on practical experience [see *Bernier et al.*, 2000]. However, we feel that the practitioner will apply the theory only (1) if there are guidelines for common statistical models, (2) if robust numerical methods available and (3) if there are clear practical advantages over competing models, such as possibility theory and the robust Bayesian theory.

While the shortcomings of rules for defining conditional possibility [*Walley and de Cooman*, 1999] may convince the practitioner of using a more general model, we find that it is much more difficult to argue that CLP theory has practical advantages over robust Bayesian theory. One way of doing that might be to show with realistic examples that the tools of CLP theory (mainly natural extension and the generalized Bayes rule) can be used to obtain useful inferences from a small number of hypotheses, which the practitioner can more easily assess and defend. We are investigating further the practical merit of the ideas discussed in the present paper [*Fortin et al.*, 2001], and should present shortly an application to the hydrological engineering problem proposed by *Caselton and Luo* [1992].

Acknowledgments

We wish to thank Dr Louis Lafond for helping us solve the GBR numerically, and Charles Fortin for checking our proof that natural extension from a finite set of judgments can be computed using the GBR. The comments of an anonymous reviewer also helped improve the manuscript. This work was initiated during a stay at ENGREF, which was financed by a grant from the Government of Québec (action concertée FCAR/MEQ/MRI).

References

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., Springer-Verlag, New York.

Bernardo, J.M. et A.F.M. Smith (1994). *Bayesian Theory*, Wiley, Chichester, United Kingdom.

Bernier, J., É. Parent and J.J. Boreux (2000). *Statistique pour l'environnement: Traitement bayésien de l'incertitude*, Éditions Technique & Documentation, Paris.

Caselton, W.F. et W.B. Luo (1992). Decision making with imprecise probabilities - Dempster-Shafer theory and application. *Water Resources Research*, 28(12): 3071-3083.

Cozman, F.G. (1999a). *A Brief Introduction to the Theory of Sets of Probability Measures*, Technical report, School of Computer Science, Carnegie Mellon, Pittsburgh, PA.

Cozman, F.G. (1999b). Computing posterior upper expectations. *Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications*, Ghent, Belgium, 29 June – 2 July 1999.

Dempster, A.P. (1968). Upper and lower probabilities generated by a random closed interval. *Annals of Mathematical Statistics*, 39(3): 957-966.

Dempster, A.P. and A. Kong (1987). Comment. *Statistical Science*, 2(1): 32-36.

Efron, B. (1979). Computers and the theory of statistics: Thinking the unthinkable, *Soc. Ind. Appl. Math.*, 21: 460-480.

Fortin, V., É. Parent, J. Bernier and B. Bobée (2001). *Hydrological Information Encoding, Inference and Decision Analysis via Walley's Imprecise Probability Theory*, working paper, Hydro-Québec's research institute, Varennes.

Friedman, L.W. and H.H. Friedman (1995). Analyzing simulation output using the boostrap method, *Simulation*, 64(2): 95-100.

Geoffrion, A.M. (1968). Primal Resource – Directive Approaches for Optimizing Nonlinear Decomposable Systems, Working paper 141, The RAND Corporation.

Lasdon, L.S. (1970). *Optimization Theory for Large Systems*, MacMillan, New York.

Rubin, D.B. (1981). The Bayesian bootstrap. *The Annals of Statistics*, 9: 130-134.

Walley, P. (1991). *Statistical Reasoning with Imprecise probabilities*, Chapman and Hall, London.

Walley, P. (1996a). Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B*, 58(1): 3-57.

Walley, P. (1996b). Measures of uncertainty in expert systems. *Artificial Intelligence*, 83: 1-58.

Walley, P. (2000). Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24: 125-148.

Walley, P. and G. de Cooman (1999). Coherence of rules for defining conditional possibility. *International Journal of Approximate Reasoning*, 21(1):63-108.

Wasserman and Seidenfeld (1996). Comment. In: Inferences from multinomial data: Learning about a bag of marbles – discussion. *Journal of the Royal Statistical Society, Series B*, 58(1): 34-57.