Forecasting using belief functions: an application to marketing econometrics

Orakanya Kanjanatarakul[a], Songsak Sriboonchitta[b], Thierry Denœux[c]

[a]*Faculty of Management Sciences, Chiang Mai Rajabhat University, Thailand*
[b]*Faculty of Economics, Chiang Mai University, Thailand*
[c]*UMR 7253 Heudiasyc, Université de Technologie de Compiègne, CNRS, France*

Abstract

A method is proposed to quantify uncertainty on statistical forecasts using the formalism of belief functions. The approach is based on two steps. In the estimation step, a belief function on the parameter space is constructed from the normalized likelihood given the observed data. In the prediction step, the variable Y to be forecasted is written as a function of the parameter θ and an auxiliary random variable Z with known distribution not depending on the parameter, a model initially proposed by Dempster for statistical inference. Propagating beliefs about θ and Z through this model yields a predictive belief function on Y. The method is demonstrated on the problem of forecasting innovation diffusion using the Bass model, yielding a belief function on the number of adopters of an innovation in some future time period, based on past adoption data.

Keywords: Prediction, Dempster-Shafer Theory, Evidence Theory, Bass model, Innovation diffusion, Sales forecasting, Statistical inference, Likelihood, Uncertainty.

1. Introduction

Forecasting may be defined as the task of making statements about events that have not yet been observed. As such statements can usually not be guaranteed to be true, handling uncertainty is a critical issue in any forecasting task. Forecasting methods can be classified as statistical, causal or judgmental, depending on the kind of information used (respectively, data, causal relations or expert opinions). Whatever the approach used, a forecast cannot be trusted unless it is accompanied by some measure of uncertainty. Most of the time, forecast uncertainty is described by subjective probabilities or prediction intervals.

Recently, new formal frameworks for handling uncertainty have emerged and have become increasingly used in various application areas. One such framework is the Dempster-Shafer theory of belief functions [6], [33], [35]. In this approach, a piece of evidence about some question of interest is represented by a belief function, which is mathematically equivalent to a random set [29]. Independent pieces of evidence are then combined using an operation called Dempster's rule to obtain a unique belief function quantifying our state of knowledge about the question of interest. Since probability measures are special belief functions, and Bayesian conditioning can be seen as a special case of Dempster's rule, Dempster-Shafer theory is formally an extension of Bayesian probability theory. In particular, both approaches yield the same conclusions from the same initial information. However, the theory of belief function has greater expressive power and it can be argued to yield more sensible results in the presence of deep uncertainty. The reader is referred to, e.g., [34], [37] for detailed discussions on the comparison between belief function and probabilistic reasoning. Recent examples of applications of Dempster-Shafer theory can be found, e.g., in Refs. [27], [25], [24], [3], [11], among others.

Although the theory of belief functions has gained increasing popularity in the last few years, applications to forecasting have been, until now, very limited. The purpose of this paper is to demonstrate the application of belief function theory to statistical forecasting problems, with emphasis on situations where data are scarce and, consequently, uncertainty is high and needs to be quantified.

As an important application area, we will consider marketing econometrics and, more specifically, the forecasting of innovation diffusion. This has been a topic of considerable practical and academic interest in the last fifty years [28]. Typically, when a new product is launched, sale forecasts can only be based on little data and uncertainty has to be quantified to avoid making wrong business decisions based on unreliable forecasts [21], [20], [22]. The approach described in this paper uses the Bass model for innovation diffusion [2] together with past sales data to quantify the uncertainty on future sales using the formalism of belief functions. The forecasting method exemplified here can be applied to any forecasting problem when a statistical model can be postulated and some historical data is available.

The rest of this paper is organized as follows. The notion of likelihood-based belief function will first be recalled in Section 2 and the forecasting problem will be addressed in Section 3. These notions will then be applied to innovation diffusion in Section 4. Finally, the relationship with existing approaches will be discussed in Section 5 and Section 6 will conclude the paper.

2. Likelihood-based belief function

Basic knowledge of the theory of belief functions will be assumed throughout this paper. A complete exposition in the finite case can be found in Shafer's book [33] and the relation with random sets is explained in [30]. The reader is referred to [3] for a quick introduction on those aspects of this theory needed for statistical inference. In this section, the definition of a belief function from the likelihood function and its justification as proposed in [10] will first recalled. The forecasting problem will then be addressed in the next section.

Let $X \in \mathbb{X}$ denote the observable data, $\theta \in \Theta$ the parameter of interest and $f_\theta(x)$ the probability mass or density function describing the data-generating mechanism. Statistical inference consists in making meaningful statements about θ after observing the outcome x of the random experiment. This problem has been addressed in the belief function framework by many authors, starting to Dempster's seminal work [5], [6], [7], [9]. In contrast with Dempster's approach relying on an auxiliary variable (see Section 3 below), Shafer proposed, on intuitive grounds, a more direct approach in which a belief function Bel_x^Θ on Θ is built from the likelihood function. This approach was further elaborated by Wasserman [39] and discussed by Aickin [1], among others. It was recently justified by Denœux in [10], from the following three basic principles:

Likelihood principle. This principle states that all the relevant information from the random experiment is contained in the likelihood function, defined by $L_x(\theta) = \alpha f_\theta(x)$ for all $\theta \in \Theta$, where α is any positive multiplicative constant [15], [14]. This principle was shown by Birnbaum [4] to result from the principles of sufficiency and conditionality, which are of immediate intuitive appeal. This principle entails that Bel_x^Θ should be defined only from the likelihood function.

Compatibility with Bayesian inference. This principle states that, if a Bayesian prior $\pi(\theta)$ is available, combining it with Bel_x^Θ using Dempster's rule [33] should yield the Bayesian posterior. It follows from this principle that the contour function $pl_x(\theta)$ associated to Bel_x^Θ should be proportional to the likelihood function:

$$pl_x(\theta) \propto L_x(\theta). \quad (1)$$

Least Commitment Principle. According to this principle, when several belief functions are compatible with some constraints, we should select the least committed one according to some informational ordering [13, 36].

In [10], it was shown that the least committed belief function verifying (1) according to the commonality ordering [12] is the consonant belief function Bel_x^Θ whose contour function is the relative likelihood function:

$$pl_x(\theta) = \frac{L_x(\theta)}{\sup_{\theta' \in \Theta} L_x(\theta')}. \tag{2}$$

This belief function is called the likelihood-based belief function on Θ induced by x. The corresponding plausibility function can be computed from pl_x as:

$$Pl_x^\Theta(A) = \sup_{\theta \in A} pl_x(\theta), \tag{3}$$

for all $A \subseteq \Theta$. The focal sets of Bel_x^Θ are the levels sets of $pl_x(\theta)$ defined as follows:

$$\Gamma_x(\omega) = \{\theta \in \Theta | pl_x(\theta) \geq \omega\}, \tag{4}$$

for $\omega \in [0,1]$. These sets may be called *plausibility regions* and can be interpreted as sets of parameter values whose plausibility is greater than some threshold ω. The belief function Bel_x^Θ is equivalent to the random set induced by the Lebesgue measure λ on $[0,1]$ and the multi-valued mapping Γ_x from $[0,1]$ to 2^Θ [30]. In particular, the following equalities hold:

$$Bel_x^\Theta(A) = \lambda(\{\omega \in [0,1] | \Gamma_x(\omega) \subseteq A\}) \tag{5a}$$

$$Pl_x^\Theta(A) = \lambda(\{\omega \in [0,1] | \Gamma_x(\omega) \cap A \neq \emptyset\}), \tag{5b}$$

for all $A \subseteq \Theta$ such that the above expressions are well-defined.

We can also remark that the maximum likelihood estimate (MLE) of θ can be interpreted as the value of θ with the highest plausibility, and likelihood regions as defined by Edwards [14], among others, are identical to plausibility regions.

Example 1. *As an example, assume that we observe a random variable X having a binomial distribution:*

$$f_\theta(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}. \tag{6}$$

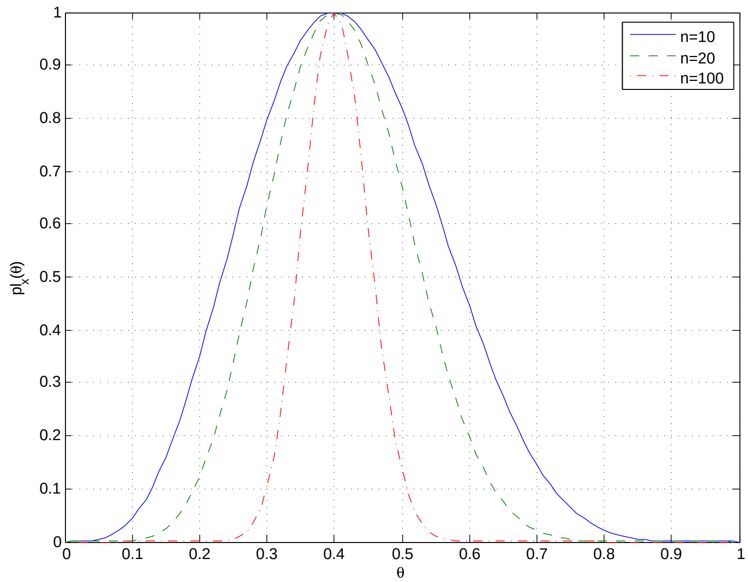


Figure 1: Contour functions (normalized likelihood functions) for the binomial distribution with $\hat{\theta} = 0.4$ and $n \in \{10, 20, 100\}$.

The likelihood-based belief function induced by x has the following contour function:

$$pl_x(\theta) = \frac{\theta^x (1-\theta)^{n-x}}{\hat{\theta}^x (1-\hat{\theta})^{n-x}} = \left(\frac{\theta}{\hat{\theta}}\right)^{n\hat{\theta}} \left(\frac{1-\theta}{1-\hat{\theta}}\right)^{n(1-\hat{\theta})}, \tag{7}$$

for all $\theta \in \Theta = [0,1]$, where $\hat{\theta} = x/n$ is the MLE of θ. Function $pl_x(\theta)$ is plotted in Figure 1 for $\hat{\theta} = 0.4$ and $n \in \{10, 20, 100\}$. We can see that the contour function becomes more specific as n increases.

As $pl_x(\theta)$ is unimodal and continuous, each plausibility region $\Gamma_x(\omega)$ for $\omega \in [0,1]$ is a closed interval $[U(\omega), V(\omega)]$ and Bel_x^Θ is equivalent to the a closed random interval $[U, V]$ [8]. The marginal cumulative probability distribution of U and V can be obtained as follows:

$$\begin{aligned}
F_U(u) &= \Pr(U \leq u) && \text{(8a)}\\
&= \Pr([U, V] \cap (-\infty, u] \neq \emptyset) && \text{(8b)}\\
&= Pl_x^\Theta((-\infty, u]) && \text{(8c)}\\
&= \begin{cases} pl_x(u) & \text{if } u \leq \hat{\theta} \\ 1 & \text{otherwise,} \end{cases} && \text{(8d)}
\end{aligned}$$

and

$$\begin{aligned}
F_V(v) &= \Pr(V \leq v) && \text{(9a)}\\
&= \Pr([U,V] \subseteq (-\infty, v]) && \text{(9b)}\\
&= Bel_x^{\Theta}((-\infty, v]) && \text{(9c)}\\
&= 1 - Pl_x^{\Theta}((v, +\infty)) && \text{(9d)}\\
&= \begin{cases} 0 & \textit{if } v \leq \hat{\theta} \\ 1 - pl_x(v) & \textit{otherwise.} \end{cases} && \text{(9e)}
\end{aligned}$$

3. Forecasting

As explained in the previous section, the inference problem consists in making statements about some parameter θ after observing a realization x of some random quantity $X \sim f_\theta(x)$. The *forecasting* problem is, in some sense, the inverse of the previous one: given some knowledge about θ obtained by observing x (represented here by a belief function), we wish to make statements about some random quantity $Y \in \mathbb{Y}$ whose conditional distribution $g_{x,\theta}(y)$ given $X = x$ depends on θ. For instance, in a sales forecasting problem, x may be the numbers of sales observed in the past n years, while Y may be number of sales to be realized in some future time period.

We will first propose a general solution to the forecasting problem in Section 3.1 and illustrate it with two examples in Section 3.2. Numerical coefficients summarizing a predictive belief function will then introduced in Section 3.3 and the relationship with the Bayesian approach will be discussed in Section 3.4.

3.1. Problem solution

A solution to the forecasting problem can be found using the sampling model used by Dempster [9] for inference. In this model, the data Y is expressed as a function of the parameter θ and an unobserved auxiliary variable $Z \in \mathbb{Z}$ with known probability distribution μ independent of θ:

$$Y = \varphi(\theta, Z), \tag{10}$$

where φ is defined in such a way that the distribution of Y for fixed θ is $g_{x,\theta}(y)$. When Y is continuous, a model of the form (10) can be obtained as $Y = F_{x,\theta}^{-1}(Z)$, where $F_{x,\theta}$ is the conditional cumulative distribution function (cdf) of Y given $X = x$ and Z has a continuous uniform distribution in $[0, 1]$.

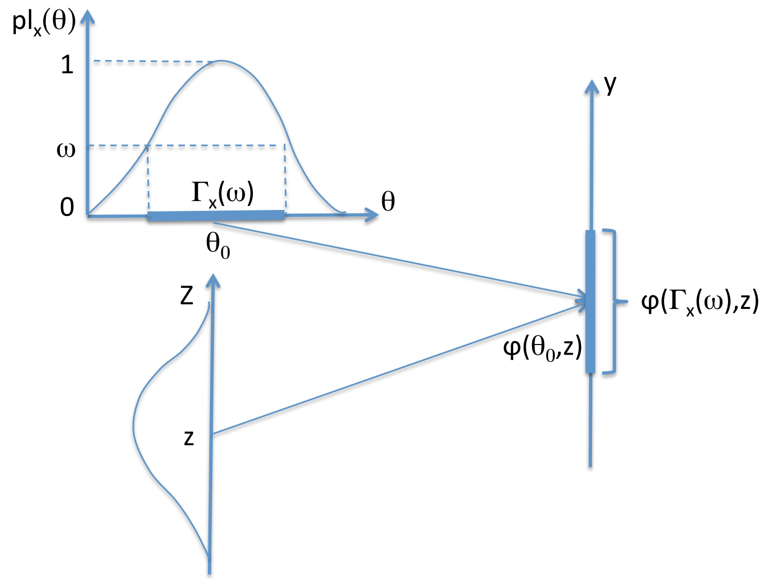


Figure 2: Random set on $Y = \varphi(\theta, Z)$ induced by the belief function on θ and the probability distribution of Z.

We note that, in the general case, both φ and μ may depend on x; however, we do not make this dependence explicit to keep the notation simple.

By composing the multi-valued mapping $\Gamma_x : [0,1] \to 2^\Theta$ with φ, we get a new multi-valued mapping Γ'_x from $[0,1] \times \mathbb{Z}$ to $2^\mathbb{Y}$ defined as follows:

$$\begin{array}{rccc} \Gamma'_x : & [0,1] \times \mathbb{Z} & \to & 2^\mathbb{Y} \\ & (\omega, z) & \to & \varphi(\Gamma_x(\omega), z). \end{array} \tag{11}$$

This definition is illustrated in Figure 2: function φ maps each pair (θ_0, z) to some value $y_0 = \varphi(\theta_0, z)$. The set $\Gamma'_x(\omega, z) = \varphi(\Gamma_x(\omega), z)$ is defined as the set of all values $\varphi(\theta_0, z)$ for θ_0 in $\Gamma_x(\omega)$.

As the distribution of Z does not depend on θ, Z and the underlying random variable ω associated with Bel_x^Θ are independent. The product measure $\lambda \otimes \mu$ on $[0,1] \times \mathbb{Z}$ and the multi-valued mapping Γ'_x thus induce *predictive* belief and plausibility functions on $\mathbb{Y}$ defined, respectively, as follows:

$$Bel_x^\mathbb{Y}(A) = (\lambda \otimes \mu)\left(\{(\omega, z) | \varphi(\Gamma_x(\omega), z) \subseteq A\}\right), \tag{12a}$$

$$Pl_x^\mathbb{Y}(A) = (\lambda \otimes \mu)\left(\{(\omega, z) | \varphi(\Gamma_x(\omega), z) \cap A \neq \emptyset\}\right), \tag{12b}$$

for all $A \subseteq \mathbb{Y}$. In particular, when $\mathbb{Y}$ is the real line, we may define the lower and upper predictive cdfs of Y as, respectively,

$$F_x^L(y) = Bel_x^{\mathbb{Y}}((-\infty, y]), \tag{13a}$$

$$F_x^U(y) = Pl_x^{\mathbb{Y}}((-\infty, y]), \tag{13b}$$

for any $y \in \mathbb{R}$.

3.2. Examples

The quantities defined in Equations (12) and (13) can sometimes be expressed analytically. Otherwise, they can be approximated by Monte Carlo simulation. Examples for these two cases are given below.

Example 2. *Let Y be a random variable with a Bernoulli distribution $\mathcal{B}(\theta)$. It can be generated using the following equation:*

$$Y = \varphi(\theta, Z) = \begin{cases} 1 & \text{if } Z \leq \theta \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

where Z has a uniform distribution in the interval $[0,1]$. Assume that Bel_x^Θ is induced by a random closed interval $\Gamma_x(\omega) = [U(\omega), V(\omega)]$ (this is the case, in particular, if the observed data X used to estimate θ have a binomial distribution, as in Example 1). We have

$$\varphi([U(\omega), V(\omega)], z) = \begin{cases} 1 & \text{if } Z \leq U(\omega) \\ 0 & \text{if } Z > V(\omega) \\ \{0,1\} & \text{otherwise.} \end{cases} \tag{15}$$

Consequently, the predictive belief function $Bel_x^{\mathbb{Y}}$ can be computed as follows:

$$Bel_x^{\mathbb{Y}}(\{1\}) = (\lambda \otimes \mu)(\{(\omega, z) | Z \leq U(\omega)\}) \tag{16a}$$

$$= \int_0^1 \mu(\{z | z \leq U(\omega)\}) f(\omega) d\omega \tag{16b}$$

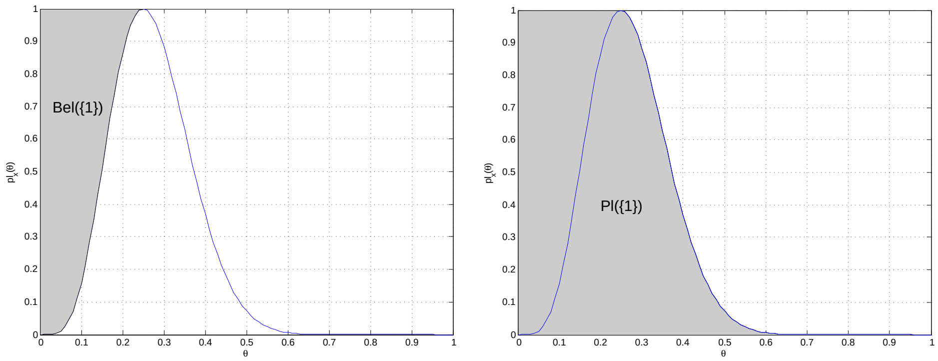
$$= \int_0^1 U(\omega) f(\omega) d\omega = \mathbb{E}(U) \tag{16c}$$

and

$$Bel_x^{\mathbb{Y}}(\{0\}) = (\lambda \otimes \mu)(\{(\omega, z) | Z > V(\omega\}) \tag{16d}$$

$$= 1 - (\lambda \otimes \mu)(\{(\omega, z) | Z \leq V(\omega)\}) \tag{16e}$$

$$= 1 - \mathbb{E}(V). \tag{16f}$$



(a) (b)

Figure 3: Predictive belief and plausibility of success for a Bernoulli trial based on the contour function $pl_x(\theta)$ on the probability of success θ.

Equivalently,

$$Pl_x^{\mathbb{Y}}(\{1\}) = 1 - Bel_x^{\mathbb{Y}}(\{0\}) = \mathbb{E}(V). \tag{17}$$

As $\Pr(U \geq 0) = \Pr(V \geq 0) = 1$*, we can write:*

$$Bel_x^{\mathbb{Y}}(\{1\}) = \int_0^{+\infty} (1 - F_U(u))du \tag{18a}$$

$$= \int_0^{\hat{\theta}} (1 - pl_x(u))du \tag{18b}$$

$$= \hat{\theta} - \int_0^{\hat{\theta}} pl_x(u)du \tag{18c}$$

and

$$Pl_x^{\mathbb{Y}}(\{1\}) = \int_0^{+\infty} (1 - F_V(v))du \tag{18d}$$

$$= \hat{\theta} + \int_{\hat{\theta}}^1 pl_x(v)dv. \tag{18e}$$

These two quantities can be represented as the areas of regions delimited by the contour function, as shown in Figure 3. The difference $Pl_x^{\mathbb{Y}}(\{1\}) - Bel_x^{\mathbb{Y}}(\{1\})$*, which is the mass* $m_x^{\mathbb{Y}}(\{0,1\})$ *assigned to ignorance, is simply the area under the contour function* pl_x*. It tends to zero as the sample size* n *tends to infinity.*

Example 3. *As a second example, let us consider the case where* $\mathbf{X} = (X_1, \ldots, X_n)$ *is an i.i.d. sample from a normal distribution* $\mathcal{N}(m, \sigma^2)$ *(referred to as "past data"), and* $Y \sim \mathcal{N}(m, \sigma^2)$ *is a not yet observed random variable ("future data") drawn independently from the same distribution. The contour function on* $\theta = (m, \sigma^2)$ *given a realization* $\mathbf{x}$ *of* $\mathbf{X}$ *is*

$$pl_{\mathbf{x}}(m, \sigma^2) = \frac{(2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - m)^2\right)}{(2\pi s^2)^{-n/2} \exp\left(-\frac{1}{2s^2}\sum_{i=1}^n (x_i - \overline{x})^2\right)} \tag{19a}$$

$$= \left(\frac{s^2}{\sigma^2}\right)^{n/2} \exp\left(\frac{n}{2} - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - m)^2\right), \tag{19b}$$

where $\overline{x}$ *and* s^2 *are, respectively, the sample mean and the sample variance. The future data* Y *can be written as*

$$Y = \varphi(\theta, Z) = m + \sigma Z, \tag{20}$$

with $Z \sim \mathcal{N}(0,1)$*. For any* (ω, z) *in* $[0,1] \times \mathbb{R}$*, the set* $\varphi(\Gamma_{\mathbf{x}}(\omega), z)$ *is the interval* $[y^L(\omega, z), y^U(\omega, z)]$ *defined by the following lower and upper bounds:*

$$y^L(\omega, z) = \min_{\{(m,\sigma^2) | pl_x(m,\sigma^2) \geq \omega\}} m + \sigma z \tag{21a}$$

$$y^U(\omega, z) = \max_{\{(m,\sigma^2) | pl_x(m,\sigma^2) \geq \omega\}} m + \sigma z, \tag{21b}$$

which can be computed using a constrained nonlinear optimization algorithm. By drawing independently N *pairs* (ω_i, z_i)*,* $i = 1, \ldots, N$*, we get* N *intervals* $[y^L(\omega_i, z_i), y^U(\omega_i, z_i)]$*. For any* $A \subset \mathbb{R}$*, the quantities* $Bel_{\mathbf{x}}^{\mathbb{Y}}(A)$ *and* $Pl_{\mathbf{x}}^{\mathbb{Y}}(A)$ *defined by (12) can be approximated by*

$$\widehat{Bel}_{\mathbf{x}}^{\mathbb{Y}}(A) = \frac{1}{N} \#\{i \in \{1, \ldots, N\} \mid [y^L(\omega_i, z_i), y^U(\omega_i, z_i)] \subseteq A\}, \tag{22a}$$

$$\widehat{Pl}_{\mathbf{x}}^{\mathbb{Y}}(A) = \frac{1}{N} \#\{i \in \{1, \ldots, N\} \mid [y^L(\omega_i, z_i), y^U(\omega_i, z_i)] \cap A \neq \emptyset\}. \tag{22b}$$

3.3. Summarizing the predictive belief function

Any of the two functions $Bel_x^{\mathbb{Y}}$ and $Pl_x^{\mathbb{Y}}$ completely describes our knowledge of Y, given the observed data x. However, to facilitate interpretation by the decision-maker, it may be useful to summarize them in the form of a small number of coefficients. Assuming Y to be a real random variable, its lower and upper expectations [30] with respect to $Bel_x^{\mathbb{Y}}$ are defined, respectively, as follows:

$$\mathbb{E}_x^L(Y) = \int \min \varphi(\Gamma_x(\omega), z)\ d\lambda(\omega) d\mu(z), \tag{23a}$$

$$\mathbb{E}_x^U(Y) = \int \max \varphi(\Gamma_x(\omega), z)\ d\lambda(\omega) d\mu(z). \tag{23b}$$

A point prediction of Y can also be obtained by plugging the MLE $\widehat{\theta}$ in (10) and taking the expectation with respect to Z:

$$\widehat{y} = \int \varphi(\widehat{\theta}, z) d\mu(z), \tag{24}$$

which is the MLE of the conditional expectation of Y given $X = x$. As $\widehat{\theta} \in \Gamma_x(\omega)$ for all $\omega \in [0,1]$, the following inequalities hold:

$$\mathbb{E}_x^L(Y) \leq \widehat{y} \leq \mathbb{E}_x^U(Y). \tag{25}$$

If Y is continuous, we may also define its lower and upper predictive quantiles at level α, for any $\alpha \in (0,1)$, as:

$$q_\alpha^L = (F_x^U)^{-1}(\alpha), \tag{26a}$$

$$q_\alpha^U = (F_x^L)^{-1}(\alpha). \tag{26b}$$

By definition, q_α^L and q_α^U are thus, respectively, the values such that

$$Pl_x^{\mathbb{Y}}((-\infty, q_\alpha^L]) = \alpha \tag{27a}$$

and

$$Bel_x^{\mathbb{Y}}((-\infty, q_\alpha^U]) = \alpha \tag{27b}$$

or, equivalently,

$$Pl_x^{\mathbb{Y}}((q_\alpha^U, +\infty)) = 1 - \alpha. \tag{27c}$$

For any $\alpha \in (0, 0.5]$, we may compute the α-quantile interval $(q_\alpha^L, q_{1-\alpha}^U]$, which has an obvious interpretation: the plausibility that Y will lie below this interval and the plausibility that Y will lie above it are both equal to α. Because of the sub-additivity of $Pl_x^{\mathbb{Y}}$, we may conclude that

$$Pl_x^{\mathbb{Y}}(\overline{(q_\alpha^L, q_{1-\alpha}^U]}) \leq 2\alpha. \tag{28a}$$

or, equivalently,

$$Bel_x^{\mathbb{Y}}((q_\alpha^L, q_{1-\alpha}^U]) \geq 1 - 2\alpha. \tag{28b}$$

The definitions of lower and upper quantiles can be extended to the case where Y is discrete by linearly interpolating the lower and upper cdfs between the discrete values of Y.

Another useful summary of of predictive belief function is the predictive contour function $pl_x(y) = Pl_x^{\mathbb{Y}}(\{y\})$. We note that $Bel_x^{\mathbb{Y}}$ is generally not consonant, so that it cannot be completely recovered from its contour function. When function $pl_x(y)$ has a unique maximum $\widetilde{y}$, it may be taken as a point prediction of Y, with an easy interpretation as the most plausible value of Y given x. The point estimates $\widehat{y}$ and $\widetilde{y}$ are different in general, the latter being arguably more suitable as a point prediction of Y.

3.4. Relationship with the Bayesian posterior predictive distribution

To conclude this section, we can remark that the predictive belief function $Bel_x^{\mathbb{Y}}$ boils down to the Bayesian posterior predictive distribution of Y given $X = x$ when a prior probability distribution $\pi(\theta)$ is available and combined with the belief function Bel_x^{Θ} by Dempster's rule. As recalled in Section 2, the combined belief function $Bel_x^{\Theta} \oplus \pi$ is then, by construction, the posterior probability distribution $f_x(\theta)$ of θ given $X = x$ and we then have, for any measurable subset $A \subseteq \mathbb{Y}$:

$$
\begin{aligned}
Bel_x^{\mathbb{Y}}(A) &= \Pr(\varphi(\theta, Z) \in A | x) && \text{(29a)}\\
&= \int_{\Theta} \Pr(\varphi(\theta, Z) \in A | \theta, x) f_x(\theta) d\theta && \text{(29b)}\\
&= \int_{\Theta} \left(\int_A g_{x,\theta}(y) dy \right) f_x(\theta) d\theta && \text{(29c)}\\
&= \int_A \left(\int_{\Theta} g_{x,\theta}(y) f_x(\theta) d\theta \right) dy && \text{(29d)}\\
&= \int_A g_x(y) dy, && \text{(29e)}
\end{aligned}
$$

which is the posterior predictive probability that Y belongs to A, given x.

The forecasting method introduced in this paper is thus a proper generalization of the Bayesian approach. The two methods coincide when a prior probability distribution of the parameter is provided. However, this is not required in the belief function approach, making it less arbitrary than the Bayesian approach in the absence of prior knowledge about the data distribution. This important point will be further discussed in Section 5 below.

4. Application to innovation diffusion

In this section, the inference and forecasting methodology outlined in the previous section will be applied to the problem of forecasting the diffusion of innovation. In spite of the considerable amount of work on this topic since the 1960's [28], the Bass model [2] remains one of the most widely used models of innovation diffusion (see, e.g., [38], [23], [21]). It will first be presented in Section 4.1. Parameter inference and sales forecasting using this model in the belief function framework will then be addressed in Sections 4.2 and 4.3, respectively.

4.1. The Bass model

The Bass model is based on the following assumption: the probability that an initial purchase of an innovative product will be made at t, given that no purchase has yet been made, is an affine function of the number of previous buyers [2]. Formally, let $f(t)$ denote the likelihood of purchase at time t for eventual adopters, and

$$F(t) = \int_0^t f(u)du. \tag{30}$$

The likelihood of purchase at time t for eventual adopters, given that no purchase has yet been made, is assumed to be of the form

$$\frac{f(t)}{1 - F(t)} = p + qF(t), \tag{31}$$

where p is called the coefficient of innovation and q the coefficient of imitation. Using the initial value $F(0) = 0$, integration of the above equation yields

$$F(t) = \frac{1 - \exp[-(p+q)t]}{1 + (p/q)\exp[-(p+q)t]}, \tag{32}$$

which is the probability that an eventual adopter will buy the product before time t. If c denotes the probability of eventually adopting the product, the unconditional probability of adoption before time t for an individual taken at random from the population is

$$\Phi_\theta(t) = cF(t), \tag{33}$$

with $\theta = (p, q, c)$. In a sample of size M taken from the process, the expected number of eventual adopters is cM.

Bass [2] initially proposed to estimate the model parameters using an ordinary least squares method. A maximum likelihood approach was later proposed by Schmittlein and Mahajan [32]. This latter approach is described hereafter.

4.2. Parameter estimation

Typically, individual adoption times are not available, but we know the number of adopters between some time intervals. Let x_i denote the observed number of adopters in time interval $[t_{i-1}, t_i)$, for $i = 1, \ldots, T-1$, where t_0 is the initial time where the innovation was launched. The number of individuals in the sample of size M who did not adopt the product at time t_{T-1} is

$$x_T = M - \sum_{i=1}^{T-1} x_i. \tag{34}$$

We note that the sample may actually consist of the whole population of potential adopters, or a subset of that population in the case where data are collected from a survey.

The probability that an individual adopts the innovation between times t_{i-1} and t_i, for $i = 1, \ldots, T-1$, is $p_i = \Phi_\theta(t_i) - \Phi_\theta(t_{i-1})$, and the probability that an individual does not adopt the innovation before t_{T-1} is $p_T = 1 - \Phi_\theta(t_{T-1})$. Consequently, the observed data $\mathbf{x} = (x_1, \ldots, x_T)$ is a realization of a multinomial random vector $\mathbf{X}$ with probabilities $(p_1, \ldots, p_T)$ and the likelihood function is

$$L_{\mathbf{x}}(\theta) \propto \prod_{i=1}^{T} p_i^{x_i} = \left(\prod_{i=1}^{T-1} [\Phi_\theta(t_i) - \Phi_\theta(t_{i-1})]^{x_i} \right) [1 - \Phi_\theta(t_{T-1})]^{x_T}. \tag{35}$$

Explicit formulas for the MLE $\widehat{\theta}$ of θ do not exist and an iterative optimization procedure must be used. The belief function on θ is defined by the following contour function:

$$pl_{\mathbf{x}}(\theta) = \frac{L_{\mathbf{x}}(\theta)}{L_{\mathbf{x}}(\widehat{\theta})}, \tag{36}$$

and the marginal contour functions on each individual parameter are

$$pl_{\mathbf{x}}(p) = \sup_{q,c} pl_{\mathbf{x}}(\theta) \tag{37a}$$

$$pl_{\mathbf{x}}(q) = \sup_{p,c} pl_{\mathbf{x}}(\theta) \tag{37b}$$

$$pl_{\mathbf{x}}(c) = \sup_{p,q} pl_{\mathbf{x}}(\theta). \tag{37c}$$

Example 4. *To illustrate the estimation method outlined above, we considered the Ultrasound data used in [32]. These data were collected from 209 hospitals through the U.S.A., i.e.,* $M = 209$. *The hospitals were asked to*

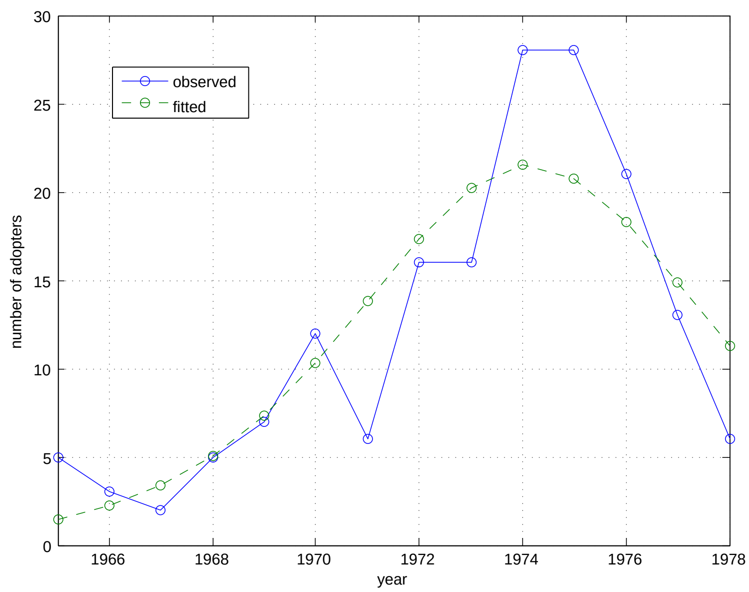


Figure 4: Actual and fitted numbers of adopters for the Ultrasound data.

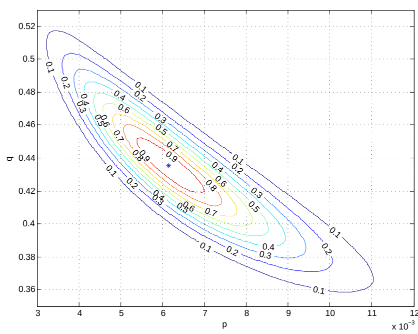
identify themselves as adopters or nonadopters of an ultrasound equipment and, if adopters, to provide the date of adoption [31].

The actual and fitted numbers of adopters are shown in Figure 4. The MLE estimates are $\widehat{p} = 0.0061$, $\widehat{q} = 0.4353$ *and* $\widehat{c} = 0.9236$. *These values are close, but not identical to those reported in [32]. Discrepancies may be due to the use of different optimization algorithms (we used the Matlab function* `fmincon`*). Figure 5 shows two-dimensional slices of the contour function, with one of the three parameters fixed to its MLE.*

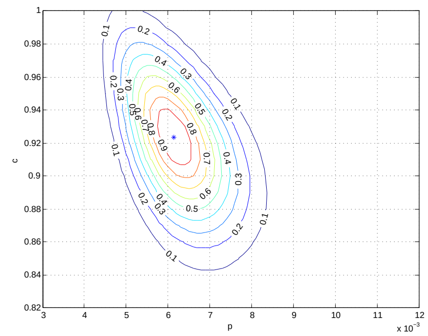
Finally, Figure 6 displays the marginal contour functions (37) for parameters p, q *and* c. *These marginal plausibilities can be used to compute plausibility intervals for each of the three parameters. These intervals are identical to likelihood intervals [14]. For instance, setting the threshold to 0.8, we get the following intervals for* p, q *and* c:

$$p \in [0.0050, 0.0075], \quad q \in [0.401, 0.469], \quad c \in [0.90, 0.96]. \tag{38}$$

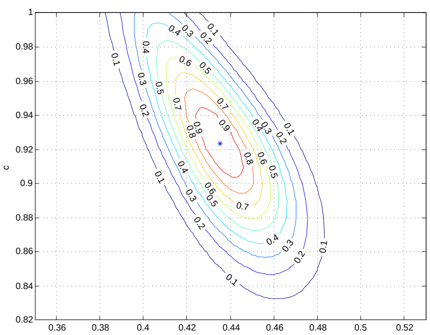
We can remark that these intervals are not based on asymptotic approximations, in contrast to the approximate confidence intervals given in [32]. The relative merits of plausibility and confidence intervals will be discussed in Section 5.



(a)

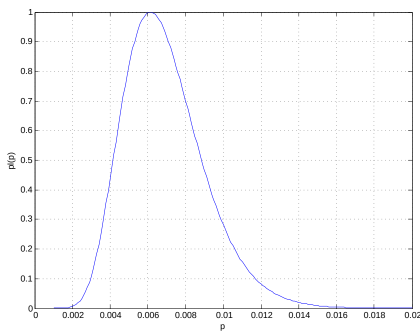


(b)

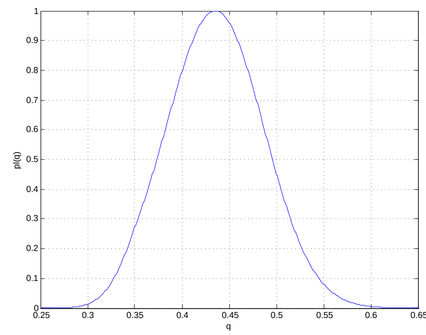


(c)

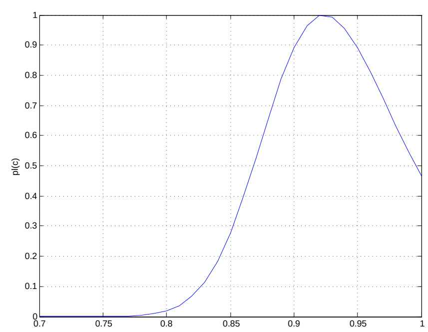
Figure 5: Contour plots of $pl_{\mathbf{x}}(\theta)$ in two-dimensional parameter subspaces with (a): $c = \widehat{c}$, (b): $q = \widehat{q}$ (b) and (c): $p = \widehat{p}$.



(a)



(b)



(c)

Figure 6: Marginal plausibilities of parameters p (a), q (b) and c (c) for the Ultrasound data.

4.3. Sales forecasting with the Bass model

Let us assume we wish to predict at time t_{T-1} the number of sales Y between times τ_1 and τ_2, with $t_{T-1} \leq \tau_1 < \tau_2$. The probability of purchase for an individual in that period, given that no purchase has been made before t_{T-1} is

$$\pi_\theta = \frac{\Phi_\theta(\tau_2) - \Phi_\theta(\tau_1)}{1 - \Phi_\theta(t_{T-1})}. \tag{39}$$

Let Q be the number of potential adopters at time t_{T-1}, assumed to be known. If M is the size of the whole population, then $Q = x_T$. Since Q individuals did not adopt the innovation before time t_{T-1}, Y has, conditionally on $\mathbf{x}$, a binomial distribution $\mathcal{B}(Q, \pi_\theta)$. It can thus be written as

$$Y = \varphi(\theta, \mathbf{Z}) = \sum_{i=1}^{Q} \mathbb{1}_{[0,\pi_\theta]}(Z_i), \tag{40}$$

where $\mathbb{1}_{[0,\pi_\theta]}$ is the indicator function of the interval $[0, \pi_\theta]$ ($\mathbb{1}_{[0,\pi_\theta]}(Z_i) = 1$ if $Z_i \leq \pi_\theta$ and $\mathbb{1}_{[0,\pi_\theta]}(Z_i) = 0$ otherwise) and $\mathbf{Z} = (Z_1, \ldots, Z_Q)$ has a uniform distribution in $[0, 1]^Q$. Individual i will adopt the innovation if $Z_i \leq \pi_\theta$, and will not adopt it if $Z_i > \pi_\theta$.

To determine $\varphi(\Gamma_x(\omega), \mathbf{z})$ for any $\omega \in [0, 1]$ and $\mathbf{z} = (z_1, \ldots, z_Q) \in [0, 1]^Q$, we may reason as follows. Let $[\pi_\theta^L(\omega), \pi_\theta^U(\omega)]$, with

$$\pi_\theta^L(\omega) = \min_{\{\theta | pl_x(\theta) \geq \omega\}} \pi_\theta, \tag{41a}$$

$$\pi_\theta^U(\omega) = \max_{\{\theta | pl_x(\theta) \geq \omega\}} \pi_\theta, \tag{41b}$$

be the range of π_θ when θ varies in $\Gamma_x(\omega)$.

Let $N_1(\omega)$, $N_2(\omega)$ and $N_2(\omega)$ denote, respectively, the number of Z_i's that fall in the intervals $[0, \pi_\theta^L(\omega))$, $[\pi_\theta^L(\omega), \pi_\theta^U(\omega))$ and $[\pi_\theta^U(\omega), 1]$:

$$N_1(\omega) = \sum_{i=1}^{Q} \mathbb{1}_{[0,\pi_\theta^L(\omega))}(Z_i), \tag{42a}$$

$$N_2(\omega) = \sum_{i=1}^{Q} \mathbb{1}_{[\pi_\theta^L(\omega),\pi_\theta^U(\omega))}(Z_i), \tag{42b}$$

$$N_3(\omega) = \sum_{i=1}^{Q} \mathbb{1}_{[\pi_\theta^U(\omega),1]}(Z_i). \tag{42c}$$

The vector $(N_1(\omega), N_2(\omega), N_3(\omega))$ has a multinomial distribution with parameters Q and

$$\mathbf{p}(\omega) = \left(\pi_\theta^L(\omega), \pi_\theta^U(\omega) - \pi_\theta^L(\omega), 1 - \pi_\theta^U(\omega)\right). \tag{43}$$

Assuming that $\pi_\theta \in [\pi_\theta^L(\omega), \pi_\theta^U(\omega)]$, we know that at least $N_1(\omega)$ individuals will adopt the product and at least $N_3(\omega)$ individuals will not adopt it. The minimum and maximum values of $Y = \varphi(\theta, \mathbf{Z})$ are thus, respectively,

$$Y^L(\omega, \mathbf{Z}) = N_1(\omega), \tag{44}$$
$$Y^U(\omega, \mathbf{Z}) = Q - N_3(\omega) = N_1(\omega) + N_2(\omega), \tag{45}$$

and the range of the number of adopters between times τ_1 and τ_2 is

$$\varphi(\Gamma_x(\omega), \mathbf{Z}) = [Y^L(\omega, \mathbf{Z}), Y^U(\omega, \mathbf{Z})]. \tag{46}$$

Given ω, $Y^L(\omega, \mathbf{Z})$ and $Y^L(\omega, \mathbf{Z})$ have binomial distributions $\mathcal{B}(Q, \pi_\theta^L(\omega))$ and $\mathcal{B}(Q, \pi_\theta^U(\omega))$, respectively. Their conditional expectations given ω are thus, respectively, $Q \cdot \pi_\theta^L(\omega)$ and $Q \cdot \pi_\theta^U(\omega)$.

The belief and plausibilities that Y will be less than, or equal to y are

$$Bel_x^{\mathbb{Y}}([0, y]) = \int_0^1 F_{Q, \pi_\theta^L(\omega)}(y) d\omega \tag{47a}$$
$$Pl_x^{\mathbb{Y}}([0, y]) = \int_0^1 F_{Q, \pi_\theta^U(\omega)}(y) d\omega, \tag{47b}$$

where $F_{Q,p}$ denotes the cdf of the binomial distribution $\mathcal{B}(Q, p)$. The contour function of Y is

$$pl_x(y) = \int_0^1 \Pr(y \in [Y^L(\omega, \mathbf{Z}), Y^U(\omega, \mathbf{Z})]) d\omega \tag{48a}$$
$$= \int_0^1 \left(1 - \Pr(Y^L(\omega, \mathbf{Z}) > y) - \Pr(Y^U(\omega, \mathbf{Z}) < y)\right) d\omega \tag{48b}$$
$$= \int_0^1 \left(F_{Q, \pi_\theta^L(\omega)}(y) - F_{Q, \pi_\theta^U(\omega)}(y - 1)\right) d\omega. \tag{48c}$$

The integrals in (47)-(48) can be approximated by Monte Carlo simulation. For instance, if $\omega_1, \ldots, \omega_N$ are N numbers drawn at random from the uniform distribution in $[0,1]$, we have

$$Bel_x^{\mathbb{Y}}([0,y]) \approx \frac{1}{N}\sum_{i=1}^{N} F_{Q,\pi_\theta^L(\omega_i)}(y), \tag{49a}$$

$$Pl_x^{\mathbb{Y}}([0,y]) \approx \frac{1}{N}\sum_{i=1}^{N} F_{Q,\pi_\theta^U(\omega_i)}(y), \tag{49b}$$

$$pl_x(y) \approx \frac{1}{N}\sum_{i=1}^{N} \left(F_{Q,\pi_\theta^L(\omega_i)}(y) - F_{Q,\pi_\theta^U(\omega_i)}(y-1)\right). \tag{49c}$$

Similarly, the lower and upper expectations of Y with respect to $Bel_x^{\mathbb{Y}}$ can be approximated, respectively, by

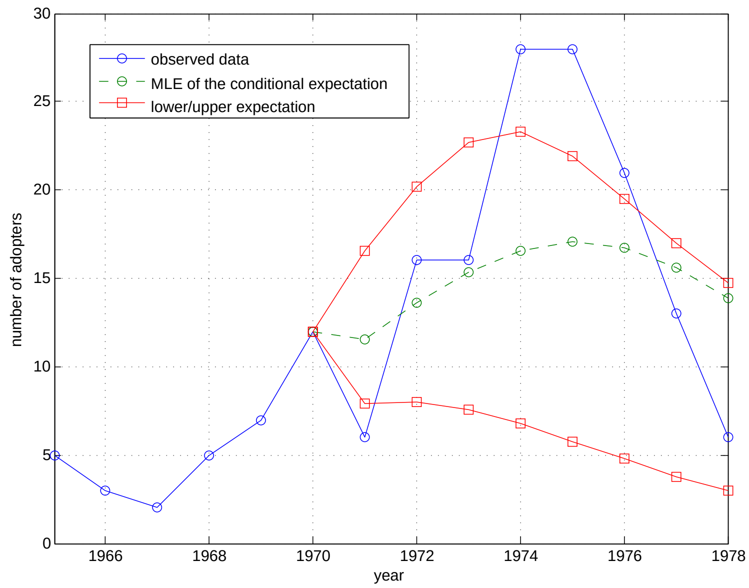
$$\overline{Y}^L = \frac{Q}{N}\sum_{i=1}^{N} \pi_\theta^L(\omega_i) \tag{50a}$$

and

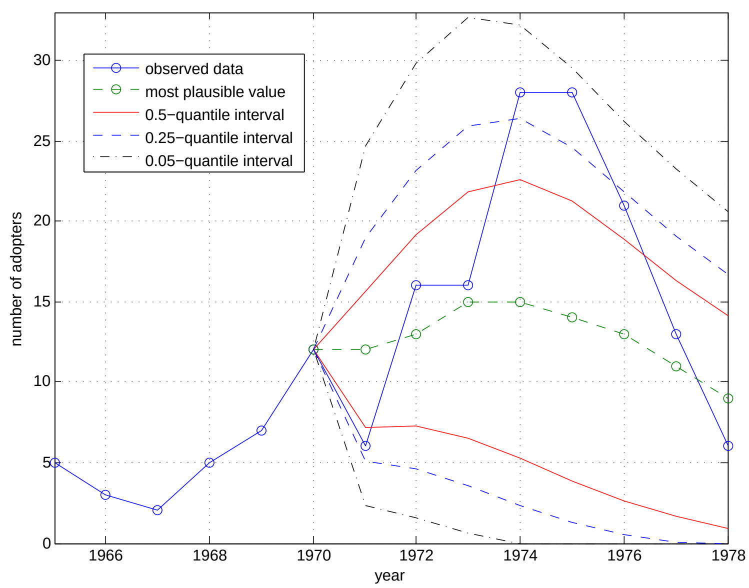
$$\overline{Y}^U = \frac{Q}{N}\sum_{i=1}^{N} \pi_\theta^U(\omega_i). \tag{50b}$$

Example 5. *The approach described above was applied to the Ultrasound data already used in Example 4. We assumed that $Q = x_T$. Figures 7 and 8 show, respectively the predictions made in 1970 and 1974 for the number of adopters in the periods 1971-1978 and 1975-1978. The estimated conditional expectation $\widehat{y}$ as well as the lower and upper expectations are shown in Figures 7(a) and 8(a). The most plausible predictions $\widetilde{y}$ are displayed in Figures 7(b) and 8(b), together with α-quantile intervals with $\alpha \in \{0.05, 0.25, 0.5\}$.*

The forecasts made in 1970 are based on little data, which is reflected by large upper-lower c and quantile intervals (Figures 7). We can see that the numbers of adopters in 1974 and 1975 is quite severely underestimated both by $\widehat{y}$ and $\widetilde{y}$, which may be due to the partial inadequacy of the model; However, the time of the peak is correctly predicted, and the observed data are contained in the 0.05-quantile intervals. As expected, the upper-lower expectation and quantile intervals are narrower for the predictions made in 1974, which are based on more historical data (Figures 8).

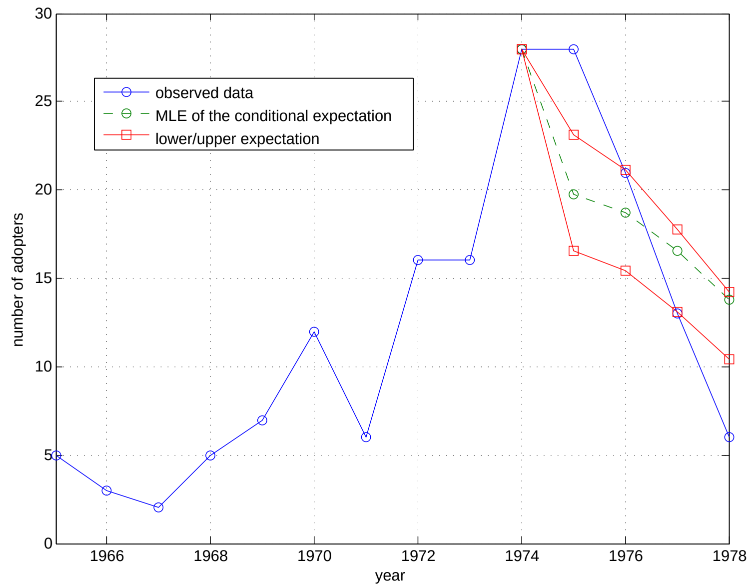


(a)

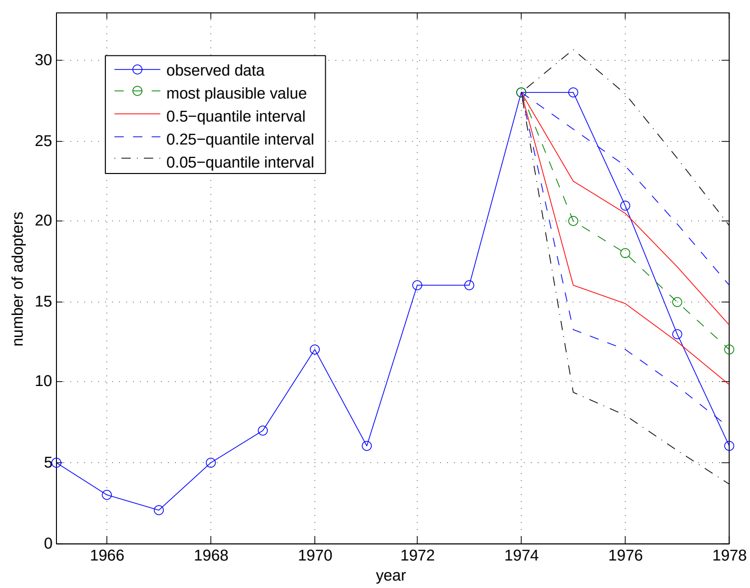


(b)

Figure 7: Predictions made in 1970 for the number of adopters in the period 1971-1978. (a): $\widehat{y}$ and lower-upper expectation intervals; (b): $\widetilde{y}$ and α-quantile intervals with $\alpha \in \{0.05, 0.25, 0.5\}$.



(a)



(b)

Figure 8: Predictions made in 1974 for the number of adopters in the period 1975-1978. (a): $\widehat{y}$ and lower-upper expectation intervals; (b): $\widetilde{y}$ and α-quantile intervals with $\alpha \in \{0.05, 0.25, 0.5\}$.

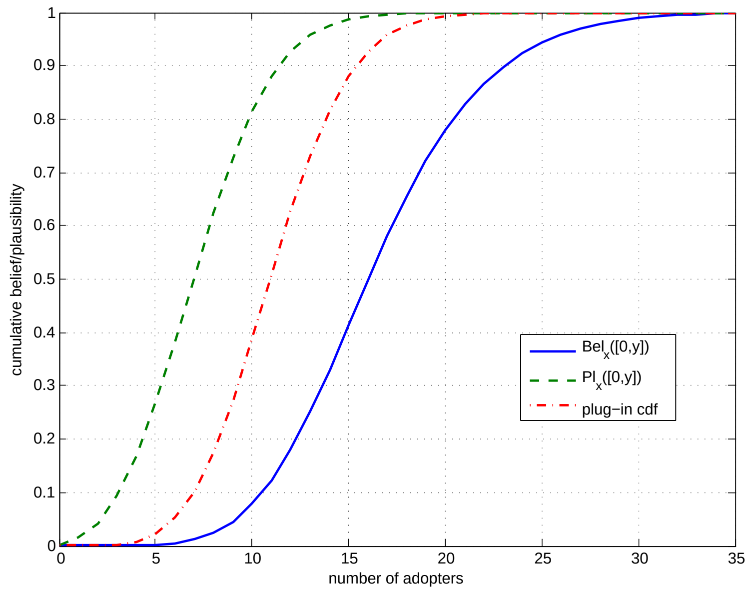


Figure 9: Lower and upper cdfs for the number of adopters in 1971, forecasted in 1970, and plug-in conditional cdf $F_{x,\widehat{\theta}}(y)$ of Y given x.

Figures 9 and 10 show additional information about the predictive belief function $Bel_x^{\mathbb{Y}}$ *concerning the forecasted number of adopters for 1971. Figure 9 displays the lower and upper cdfs* F_x^L *and* F_x^U*, together with the plug-in conditional cdf* $F_{x,\widehat{\theta}}(y)$ *of* Y *given* x*. Figure 10 displays the plausibilities* $Pl_x^{\mathbb{Y}}([y-r, y+r])$ *as functions of* y*, for* $r = 0, 1, 2, 3, 4, 5$*. This representation makes it possible to find the smallest intervals with given plausibility value. For instance, the smallest intervals with plausibility at least equal to 0.9 are* 10 ± 2*,* 11 ± 2*,* 12 ± 2 *and* 13 ± 2*.*

5. Discussion

The method just described makes it possible to quantify the uncertainty pertaining to statistical forecasts, within the formalism of Dempster-Shafer theory. This method is quite general and can be applied to any statistical model. When applied to the Bass model, it allows us to compute a belief function on the number of adopters of an innovative product, which can be summarized in different forms, such as lower/upper cumulative distribution functions, contour functions, lower/upper expectations and lower/upper quantiles. This representation as a belief function is an alternative to con-

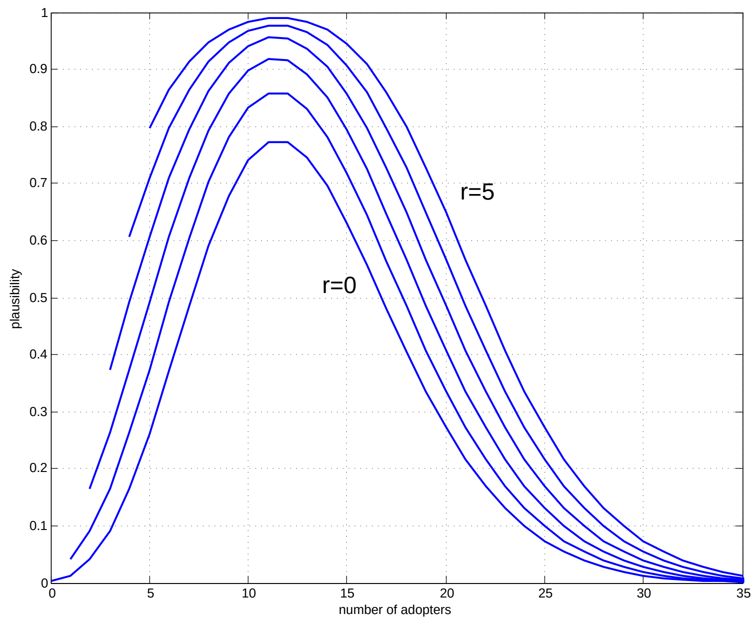


Figure 10: Plausibilities $Pl_x^{\mathbb{Y}}([y-r, y+r])$ as functions of y, from $r=0$ (lower curve) to $r=5$ (upper curve), for the number of adopters in 1971, forecasted in 1970.

fidence intervals and Bayesian predictive distributions. In this section, we discuss the advantages of the belief function approach as compared to previous models.

As shown in Sections 2 and 3.4, the methods of inference and forecasting described in this paper generalize Bayesian procedures (see, e.g., [26] for a Bayesian treatment of product adoption forecasting using the Bass model). Specifically, when a Bayesian prior is available, the belief function Bel_x^{Θ} on the parameter given the observations is, by construction, the Bayesian posterior, and the predictive belief function $Bel_x^{\mathbb{Y}}$ becomes the Bayesian predictive posterior distribution of future observations. Consequently, the belief function approach is not at odds with the Bayesian approach, but it is more general. In particular, it does not require the statistician to arbitrarily provide a prior probability distribution when prior knowledge does not exist. One might object that a uniform distribution adequately represents ignorance. However, a uniform distribution is not invariant with respect to nonlinear transformations of the parameters. For instance, Schmittlein and Mahajan [32] replace parameters p and q by $a = q/p$ and $b = p + q$. It is clear that a uniform distribution on, say, p and q, induces a nonuniform distribution on a and b, and the information contained in such a distribution is completely arbitrary. In contrast, the belief function formalism allows us to incorporate not only probabilistic prior information, but also weaker forms of prior knowledge up to complete ignorance.

Prediction intervals constitute another way to represent forecast uncertainty. In [32], the authors give formulas to compute approximate confidence intervals on the parameters of Bass models, based on the Fisher information matrix. They do not, however, provide formulas for prediction intervals, which would be of greater interest to the decision-maker. Indeed such intervals are often difficult to obtain for complex models. Although confidence and prediction intervals have a clear frequentist interpretation in a repeated sampling context, their use to quantify the uncertainty pertaining to predictions based on a single dataset can be questioned (see, e.g., a thorough discussion in [14] on this issue). Also, it is not clear how prediction intervals could be combined with uncertain information from other sources, such as experts or marketing surveys. For instance, assume that a survey has been carried out among a sample of hospitals, asking the respondents whether they intend to buy the ultrasound equipment in the next two years. How could such information be combined with prediction intervals computed from past sales data? In contrast, the belief function formalism is sufficiently general to represent and combine any kind of information [33]. Finally, another advantage of belief functions over prediction intervals is that the former framework includes a formal decision theory. In practice, decision-makers will often use sales forecasts to plan investments such as new production lines or sale outlets. In the belief function framework, sound decision procedures can be based, e.g., on the Choquet expected utility with respect to the belief or plausibility measures (see, e.g., [17], [18], [19], [16] for axiomatic justifications of decision strategies based on nonadditive measures and belief functions). No such formal and well-justified decision procedures seem to exist for prediction intervals.

6. Conclusions

Uncertainty quantification is an important component of any forecasting methodology. The approach introduced in this paper allows us to represent forecast uncertainty in the belief function framework, based on past data and a statistical model. The method is based on two steps. In the first step, evidence on parameter θ is represented by a consonant belief function defined from the normalized likelihood function given observed data. In the second step, the quantity of interest Y is written as $\varphi(\theta, Z)$, where

Z is an auxiliary variable with known distribution. Beliefs on θ and Z are then propagated through φ, resulting in a belief function on Y. The Bayesian predictive probability distribution is recovered when a prior on θ is available, a condition that is not required here. The belief function formalism makes it possible to combine information from several sources (such as expert opinions and statistical data) and to prescribe a course of action based on a sound decision-theoretic framework.

As a concrete example to illustrate the above methodology, we considered the problem of forecasting sales of an innovative product using the Bass model. However, the method is very general and can be used with any parametric model such as, e.g., linear regression and time series models. Many issues related to the forecasting problem remain to be investigated in this framework, such as taking into account model uncertainty, combining expert judgement with statistical data or exploiting samples from similar populations using, for instance, the method proposed in [10].

Acknowledgement

This research was supported by a grant from the Office of the Higher Education Commission of Thailand. It was also carried out in the framework of the Labex MS2T, which was funded by the French Government, through the program "Investments for the future" by the National Agency for Research (reference ANR-11-IDEX-0004-02). The authors thank Sébastien Destercke for his helpful comments and suggestions.

References

[1] M. Aickin. Connecting Dempster-Shafer belief functions with likelihood-based inference. *Synthese*, 123:347–364, 2000.

[2] F. M. Bass. A new product growth for model consumer durables. *Management Science*, 15(5):215–227, 1969.

[3] N. Ben Abdallah, N. Mouhous Voyneau, and T. Denœux. Combining statistical and expert evidence using belief functions: Application to centennial sea level estimation taking into account climate change. *International Journal of Approximate Reasoning*, 55:341–354, 2014.

[4] Allan Birnbaum. On the foundations of statistical inference. *Journal of the American Statistical Association*, 57(298):269–306, 1962.

[5] A. P. Dempster. New methods for reasoning towards posterior distributions based on sample data. *Annals of Mathematical Statistics*, 37:355–374, 1966.

[6] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[7] A. P. Dempster. Upper and lower probability inferences based on a sample from a finite univariate population. *Biometrika*, 57:515–528, 1967.

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

[9] A. P. Dempster. The Dempster-Shafer calculus for statisticians. *International Journal of Approximate Reasoning*, 48(2):365–377, 2008.

[10] T. Denœux. Likelihood-based belief function: justification and some extensions to low-quality data. *International Journal of Approximate Reasoning (in press)*, 2014. http://dx.doi.org/10.1016/j.ijar.2013.06.007

[11] S. Destercke and M. Sallak. An extension of universal generating function in multi-state systems considering epistemic uncertainties. *IEEE Transactions on reliability*, 62(2):504–514, 2013.

[12] D. Dubois and H. Prade. A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets. *International Journal of General Systems*, 12(3):193–226, 1986.

[13] D. Dubois and H. Prade. The principle of minimum specificity as a basis for evidential reasoning. In B. Bouchon and R.R. Yager, editors, *Uncertainty in Knowledge-Based Systems*, volume 286 of *Lecture Notes in Computer Science*, pages 75–84. Springer Berlin Heidelberg, 1987.

[14] A. W. F. Edwards. *Likelihood (expanded edition)*. The John Hopkins University Press, Baltimore, USA, 1992.

[15] R. A. Fisher. *Statistical methods and scientific inference*. Oliver and Boyd, Edinburgh, 1956.

[16] P. Ghirardato. Coping with ignorance: unforeseen contingencies and non-additive uncertainty. *Economic theory*, 17:247–276, 2001.

[17] I. Gilboa. Expected utility with purely subjective non-additive probabilities. *Journal of Mathematical Economics*, 16:65–88, 1987.

[18] J.-Y. Jaffray. Application of linear utility theory to belief functions. In B. Bouchon, L. Saitta, and R.R. Yager, editors, *Uncertainty and Intelligent Systems*, volume 313 of *Lecture Notes in Computer Science*, pages 1–8. Springer Berlin Heidelberg, 1988.

[19] J.-Y. Jaffray and P. Wakker. Decision making with belief functions: compatibility and incompatibility with the sure-thing principle. *Journal of Risk and Uncertainty*, 8:255–271, 1994.

[20] O. Kanjanatarakul. *An Economic Analysis of the Royal Project Foundation: Sales Forecast of Feta Cheese and Impact on Thai Economy.* PhD thesis, Faculty of Economics, Chiang Mai University, 2013.

[21] O. Kanjanatarakul and K. Suriya. Comparison of sales forecasting models for an innovative agro-industrial product: Bass model versus logistic function. *The Empirical Econometrics and Quantitative Economics Letters*, 1(4):89–106, 2012.

[22] O. Kanjanatarakul and K. Suriya. Sufficient number of observations for sales forecasts with limited information: A case of an innovative agro-industrial product. *International Journal of Intelligent Technologies and Applied Statistics*, 6(2):121–134, 2013.

[23] J. Kanniainen, S. J. Mäkinen, R. Piché, and A. Chakrabarti. Forecasting the diffusion of innovation: a stochastic Bass model with log-normal and mean-reverting error process. *IEEE Transactions on Engineering Management*, 58(2):228–249, 2011.

[24] M. Kurdej, J. Moras, V. Cherfaoui, and Ph. Bonnifait. Controlling remanence in evidential grids using geodata for dynamic scene perception. *International Journal of Approximate Reasoning*, 55(1, Part 3):355 – 375, 2014.

[25] B. Lelandais, I. Gardin, L. Mouchard, P. Vera, and S. Ruan. Dealing with uncertainty and imprecision in image segmentation using belief function theory. *International Journal of Approximate Reasoning*, 55(1, Part 3):376–387, 2014.

[26] P. J. Lenk and A. G. Rao. Models from old: forecasting product adoption by hierarchical Bayes procedure. *Marketing Science*, 9(1):42–53, 1990.

[27] C. Liu and J. Xie. Large scale two sample multinomial inferences and its applications in genome-wide association studies. *International Journal of Approximate Reasoning*, 55(1, Part 3):330 – 340, 2014.

[28] N. Meade and T. Islam. Modelling and forecasting the diffusion of innovation – a 25-year review. *International Journal of Forecasting*, 22:519–545, 2006.

[29] H. T. Nguyen. On random sets and belief functions. *Journal of Mathematical Analysis and Applications*, 65:531–542, 1978.

[30] H.T. Nguyen. *An Introduction to Random Sets.* Chapman and Hall/CRC Press, Boca Raton, Florida, 2006.

[31] T. S. Robertson and Y. Wind. Organizational psychographics and innovativeness. *Journal of Consumer Research*, 7:24–31, 1980.

[32] D. C. Schmittlein and V. Mahajan. Maximum likelihood estimation for an innovation diffusion model of new product acceptance. *Marketing Science*, 1(1):57–78, 1982.

[33] G. Shafer. *A mathematical theory of evidence.* Princeton University Press, Princeton, N.J., 1976.

[34] G. Shafer. Constructive probability. *Synthese*, 48(1):1–60, 1981.

[35] G. Shafer. Perspectives in the theory and practice of belief functions. *International Journal of Approximate Reasoning*, 4:323–362, 1990.

[36] Ph. Smets. Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, 9:1–35, 1993.

[37] Ph. Smets and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66:191–243, 1994.

[38] F.-M. Tseng and Y.-C. Hu. Quadratic-interval Bass model for new product sales diffusion. *Expert Systems with Applications*, 36:8496–8502, 2009.

[39] L. A. Wasserman. Belief functions and statistical evidence. *The Canadian Journal of Statistics*, 18(3):183–196, 1990.