Belief function robustness in estimation

Alessio Benavoli

Abstract We consider the case in which the available knowledge does not allow to specify a precise probabilistic model for the prior and/or likelihood in statistical estimation. We assume that this imprecision can be represented by belief functions. Thus, we exploit the mathematical structure of belief functions and their equivalent representation in terms of closed convex sets of probability measures to derive robust posterior inferences.

1 Introduction

Lower and Upper probabilities induced from multivalued mappings were introduced by Dempster [1]. Shafer [2] called them belief and plausibility functions. Associated with a belief function there is a closed convex set of probability measures of which the belief function is a lower bound [1, 3, 4]. On the other hand, the lower bound of a convex set of probability measures is not necessarily a belief function, e.g., [3, Sec. 5.13.4]. Wasserman [5, 6] has shown that the mathematical structure of belief functions makes them suitable for generating classes of prior distributions to be used in robust Bayesian inference. In particular, in case the prior is expressed via a belief function and the likelihood is a precise probability measures, he has derived a closed form solution for the upper and lower bounds of the posterior probability content of a measurable subset of the parameter space (even in case of infinite spaces). In this paper, we extend this work in three directions. First, we compute upper and lower bounds of the posterior expectations for any bounded scalar function g of interest in statistical estimation. Second, we consider the case in which also the likelihood model (not only the prior) may be expressed via belief functions. By using the formalism of Walley's theory of coherent lower previsions [3], we provide closed form solutions for the lower and upper expectations of g. Third, we show the application of this model to several cases of practical interest.

2 Belief function

In this section we revise some properties of belief functions. Let $\mathscr{X}$ be a Polish space (e.g., Euclidean space) with Borel σ-algebra $\mathscr{B}(\mathscr{X})$ and let $\mathscr{Z}$ be a convex, compact, metrizable subset of a locally convex topological vector space with Borel σ-algebra $\mathscr{B}(\mathscr{Z})$ [5]. Let P_Z be a probability measure on $(\mathscr{Z}, \mathscr{B}(\mathscr{Z}))$ and let Γ be a map taking points in $\mathscr{Z}$ to nonempty, closed subsets of $\mathscr{X}$.[1] For each $A \subseteq \mathscr{X}$, define the belief and plausibility function as [1, 5]:

Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA),
Galleria 1, CH-6928 Manno (Lugano), Switzerland, e-mail: alessio@idsia.ch

[1] Natural conditions, such as upper or lower semi-continuity, may be imposed on Γ to guarantee measurability [5]

$$
\begin{aligned}
\underline{P}(A) = Bel(A) &= P_Z(\{z_i \in \mathscr{Z} : \Gamma(z_i) \subset A\}), \\
\overline{P}(A) = \ \ Pl(A) &= P_Z(\{z_i \in \mathscr{Z} : \Gamma(z_i) \cap A \neq \emptyset\}).
\end{aligned} \tag{1}
$$

The fourtuple $(\mathscr{Z}, \mathscr{B}(\mathscr{Z}), P_Z, \Gamma)$ is called a source for Bel. Bel and $P1$ are related by $Bel(A) = 1 - P1(A^C)$, where A^c is the complement of A. An intuitive explanation [5] of Bel and $P1$ is as follows. Draw z randomly according to P_Z. Then $Bel(A)$ is the probability that the random set $\Gamma(z)$ is contained in A and $P1(A)$ is the probability that the random set $\Gamma(z)$ hits A [7]. Here, a simple example [3, Sec. 5.13.3] that explains the construction of belief functions through multivalued mappings.

Example 1. Suppose that our information on $\mathscr{X}$ is a report from an unreliable witness that the event $B \subset \mathscr{X}$ has occurred. We might consider two possible explanations: either the witness really observed B, or he observed nothing at all. These hypotheses are represented by z_1 and z_2, with multivalued mapping $\Gamma(z_1) = B$ and $\Gamma(z_2) = \mathscr{X}$. If we assess the probability $P_Z(z_1) = p$ and $P_Z(z_2) = 1 - p$, this corresponds to the belief function $Bel(A) = p$ if $A \supseteq B$ and $A \neq \mathscr{X}$; $Bel(A) = 1$ if $A = \mathscr{X}$ and zero otherwise. □

This lack of knowledge expresses via a belief function can equivalently be represented through a set of probability measures, i.e., the set of all probabilities on X that are compatible with the bounds Bel and Pl [1]:

$$
\mathscr{P}_X = \{P_X : \ Bel(A) \leq P_X(A) \leq Pl(A) \ \textit{for any} \ A \subseteq \mathscr{X}\}. \tag{2}
$$

For this reason, Bel is also called lower probability $\underline{P}$ (and Pl upper probability $\overline{P}$), since it is the lower (upper) envelope of a set of probability measures. Thus, associated to each belief function, there is a closed convex set of probability measures of which a belief function is a lower bound but, on the other hand, the lower bound $\underline{P}$ of a closed convex set of probability measures is not necessarily a belief function [3]. To be a belief function, the lower probability $\underline{P}$ has to satisfy the property of ∞-monotonicity. There are many closed convex sets of distributions that are used in practical applications that are not belief functions. By restricting closed convex sets of distributions to be belief functions one looses in generality but gains in tractability. In fact, because of the ∞-monotonicity property, belief functions satisfy several nice properties. Besides tractability, belief functions are also a useful source of closed convex set of probabilities. For instance, the multivalued mapping mechanism can be used to define belief functions also in the case the set $\mathscr{X}$ is continuous.

Example 2. Consider the case $\mathscr{X} = \mathscr{Z} = \mathbb{R}$ and thus $\mathscr{B}(\mathscr{Z})$ and $\mathscr{B}(\mathscr{X})$ coincide with the standard Borel σ-algebra in $\mathbb{R}$. Since $\mathscr{X} = \mathscr{Z}$, we are considering a map from $\mathscr{X}$ to itself and, thus, for simplicity we can denote z with x. Assume that $p(x)$ is the probability density w.r.t. the Lebesgue measure on $\mathbb{R}$ associated to P_Z (assuming it exists) and consider the case $p(x) = U_{[a,b]}(x)$, i.e., the uniform density on the interval $[a,b]$. Consider then the multivalued mapping $\Gamma(x) = [x-c, x+c]$ with $c > 0$ which maps each point x in the interval $[x-c, x+c]$. This originates the following lower/upper probabilities for the interval $[r,s]$ with $r < s$:

$$\begin{aligned}\underline{P}([r,s]) &= \int\limits_{x\in[a,b]} I_{\{x:\ [x-c,x+c]\subset[r,s]\}}(u)\tfrac{1}{b-a}du,\\ \overline{P}([r,s]) &= \int\limits_{x\in[a,b]} I_{\{x:\ [x-c,x+c]\cap[r,s]\neq\emptyset\}}(u)\tfrac{1}{b-a}du,\end{aligned} \tag{3}$$

where $I_{\{A\}}$, defined by $I_{\{A\}}(x) = 1$ if $x \in A$ and $I_{\{A\}}(x) = 0$ if $x \notin A$ is called the indicator of A. Notice that the inclusion $[x-c, x+c] \subset [r,s]$ holds for all $x \in [r+c, s-c]$, while the condition $[x-c,x+c]\cap[r,s] \neq \emptyset$ is satisfied by all $x \in [r-c, s+c]$. By setting $[r,s] = (-\infty, x]$, one can compute the lower/upper cumulate distribution function:

$$\underline{P}((-\infty,x]) = \begin{cases} 0 & x < a+c,\\ \frac{x-a-c}{b-a} & a+c \leq x < b+c,\\ 1 & x \geq b+c,\end{cases} \quad \overline{P}((-\infty,x]) = \begin{cases} 0 & x < a-c,\\ \frac{x-a+c}{b-a} & a-c \leq x < b-c,\\ 1 & x \geq b-c.\end{cases} \tag{4}$$

This model can be used to account for lack of information on the support of the uniform distribution. We are eliciting a support of length $b-a$ but we are not completely sure about its extremes. □

This approach can be extended to any PDF $p(x)$ (e.g., see [5] for the Gaussian case). Assume $\mathscr{X} = \mathscr{Z} = \mathbb{R}$, we discuss two other models (the first is discussed in [5]) generated by multivalued mappings.

ε-contamination: $p(x) = (1-\varepsilon)\pi'(x) + \varepsilon\delta_{\{z_0\}}(x)$, $\Gamma(x) = x$ if $x \neq z_0$ and $\Gamma(x) = \mathbb{R}$ if $x = z_0$, where $\delta_{\{z_0\}}$ is a Dirac's delta on z_0 and π' is PDF such that $\pi'(z_0) = 0$ and $\pi' = \pi$ if $x \neq z_0$, then:

$$\underline{P}(A) = (1-\varepsilon)\int_A \pi(x)dx, \;\; \overline{P}(A) = (1-\varepsilon)\int_A \pi(x)dx + \varepsilon.$$

When $\varepsilon = 1$, we have a vacuous model $\underline{P}(A) = 0$ and $\overline{P}(A) = 1$.
heavy-tail: $p(x) = \mathscr{N}(x;0,1)$, $\Gamma(x) = [x, 1/x)$ if $x \in [0,1)$ and $\Gamma(x) = x$ if $x \geq 1$ (symmetric for the negative axis). Consider $A = (-\infty, w]$ with $w > 1$, we can then compute the lower upper distribution of X:

$$\underline{P}((-\infty,w]) = \tfrac{1}{2} + \int\limits_1^w \mathscr{N}(x;0,1)dx + \int\limits_{1/w}^1 \mathscr{N}(w;0,1)dx = \tfrac{1}{2}\left(\text{erf}\left(\tfrac{w}{\sqrt{2}}\right) - \text{erf}\left(\tfrac{1}{\sqrt{2}w}\right)\right),$$

where $\text{erf}(w) = \frac{2}{\sqrt{\pi}}\int_0^w e^{-t^2}dt$, while

$$\overline{P}((-\infty,w]) = \tfrac{1}{2} + \int\limits_0^w \mathscr{N}(x;0,1)dx = \tfrac{1}{2}\left(1 + \text{erf}\left(\tfrac{w}{\sqrt{2}}\right)\right).$$

By differentiating $\underline{P}((-\infty,w])$ w.r.t. w, one gets:

$$\tfrac{d}{dw}\underline{P}((-\infty,w]) = \tfrac{1}{2}\left(\sqrt{\tfrac{2}{\pi}}e^{-\frac{w^2}{2}} + \frac{\sqrt{\frac{2}{\pi}}e^{-\frac{1}{2w^2}}}{w^2}\right).$$

Observe that the derivative goes to zero as $1/w^2$ and, thus, it has the same tail behaviour of the Cauchy density $\gamma/[\pi((w-w_0)^2+\gamma^2)]$. This belief function can be used for instance for robustness to outliers when employed as likelihood model or as a sort of weak-informative prior (when employed as prior model).

2.1 Upper and lower expectation

The previous section has discussed several belief functions generated through a multivalued mappings. We have also seen that a belief function can equivalently be interpreted as a lower probability model defined on the subsets of $\mathscr{X}$ and also as a lower expectation model defined on the indicator functions over the subsets of $\mathscr{X}$, i.e., $\underline{E}(I_{\{A\}}) = \underline{P}(A)$. Assume that we know the functional $\underline{P}(A) = \underline{E}(I_{\{A\}})$ for any subset A of $\mathscr{X}$ how can we extend this lower probability model to compute $\underline{E}(g)$ for any bounded real-valued function of interest g. It can be shown that

$$\underline{E}(g) = \inf_{P_X \in \mathscr{P}_X} \int g(x) P_X(dx), \;\; \overline{E}(g) = \sup_{P_X \in \mathscr{P}_X} \int g(x) P_X(dx). \tag{5}$$

Thus, the interpretation of belief functions as closed convex set of probability measures allows to computer lower and upper expectations for any bounded real valued function. Since belief function are multivalued mapping, it has been proved in [5] that (5) is equal to:

$$\underline{E}(g) = \int g_*(z) P_Z(dx), \;\; \overline{E}(g) = \int g^*(z) P_Z(dz), \tag{6}$$

where $g_*(z) = \inf_{x\in\Gamma(z)} g(x)$ and $g^*(z) = \sup_{x\in\Gamma(z)} g(x)$. This fact has important implications for computation because it reduces the problem of calculating extrema over the set of probability measures $\mathscr{P}_X$ to that of finding extrema of g over subsets of $\mathscr{X}$ followed by a single integral over Z.

Example 3. Consider for instance the ε-contamination model discussed in the previous section, then

$$\begin{aligned} &\underline{E}(g) = \int g_*(z) P_Z(dz) = \int dz \left[(1-\varepsilon)\pi'(z) + \varepsilon\delta_{\{z_0\}}(z)\right] \inf_{x\in\Gamma(z)} g(x), \\ &= \int_{\mathscr{X}-\{z_0\}} (1-\varepsilon)\pi'(z) g(z) dz + \varepsilon \inf_{x\in\mathbb{R}} g(x) = \int (1-\varepsilon)\pi(z) g(z) dz + \varepsilon \inf_{x\in\mathbb{R}} g(x). \end{aligned} \tag{7}$$

In case $\pi(z) = \mathscr{N}(z; x_0, \sigma_0^2)$ and in the case the vacuous part is restricted to $[-a,a]$ with $a>0$, one gets $\underline{E}(g) = \int (1-\varepsilon) g(z) \mathscr{N}(z; x_0, \sigma_0^2) dz + \varepsilon \inf_{x\in[-a,a]} g(x)$. □

2.2 Statistical inference

Assume that $\mathscr{X} \subseteq \mathbb{R}$. Consider a likelihood model $p(y|x)$, where Y denotes the observation variable taking values from a sample space $(\mathscr{Y}, \mathscr{B}(\mathscr{Y}))$ and $x \in \mathscr{X}$. Assume that the prior information over X is expressed through a belief function or, equivalently, through the closed convex set of probability measures associated to the belief function, how can we compute the lower/upper posterior expectation of a bounded real-valued function g given the observation $\tilde{y}$?

Theorem 1. *Assume that $p(y|x)$ is $\mathscr{B}(\mathscr{Y})\times\mathscr{B}(\mathscr{X})$-measurable and bounded. Assume that the value $\tilde{y}$ of Y is observed and that $\underline{E}_X(E_Y(\delta_{\{\tilde{y}\}}|X)) = \underline{E}\big(p(\tilde{y}|x)\big) > 0$, where $\delta_{\{\tilde{y}\}}$ is a degenerate limiting measure (e.g., Dirac's delta) on $\mathscr{Y}$. The lower posterior expectation $\underline{E}(g|\tilde{y})$ is the unique solution μ of the following equation:*

$$\underline{E}_X\left(E_Y\left((g-\mu)\delta_{\{\tilde{y}\}}|X\right)\right) = 0. \quad \square \tag{8}$$

This equation is called Generalized Bayes rule (GBR) [3, Ch. 6].

Proof.

$$0 = \underline{E}_X\left(E_Y\left((g-\mu)\delta_{\{\tilde{y}\}}|X\right)\right) = \underline{E}\left((g-\mu)p(\tilde{y}|x)\right) = \inf_{P_X\in\mathscr{P}_X}\int(g(x)-\mu)p(\tilde{y}|x)P_X(dx)$$
$$= \inf_{p\in\mathscr{P}_X}\int p(\tilde{y}|x)P(dx)\left(\frac{\int g(x)p(\tilde{y}|x)P(dx)}{\int p(\tilde{y}|x)P(dx)}-\mu\right).$$

Being $\int p(\tilde{y}|x)P_X(dx) = \underline{E}\big(p(\tilde{y}|x)\big) > 0$ by hypothesis, it follows that $\mu = \inf_{p_X\in\mathscr{P}_X}\frac{\int g(x)p(\tilde{y}|x)P_X(dx)}{\int p(\tilde{y}|x)P_X(dx)}$. Therefore, GBR is equivalent to apply Bayes rule to all probability measures in $\mathscr{P}_X$ and, then, take the infimum. The following proof has been derived by [3, Sec. 6.4.1.] replacing the indicator with a Dirac's delta to account for the fact that Y is a continuous variable. $\square$

Corollary 1. *Exploiting (6) and applying (8) to belief function, it results that the lower posterior expectation $\underline{E}(g|\tilde{y})$ is the unique solution μ of the following equation:*

$$\underline{E}\big((g-\mu)p(\tilde{y}|x)\big) = \int P_Z(dz)\inf_{x\in\Gamma(z)}(g(x)-\mu)p(\tilde{y}|x) = 0. \tag{9}$$

$\square$

For the proof, see Theorem 2. Equation (9) can be extended to the case of n i.i.d. observation, by simply replacing $p(\tilde{y}|x)$ with $\prod_{i=1}^n p(\tilde{y}_i|x)$. Assume now the case in which also the likelihood model is expressed through a belief function characterized by $(\mathscr{U},\mathscr{B}(\mathscr{U}),\Gamma(\cdot|x),P_{U|x})$ for each value of the conditional variable x.

Theorem 2. *Assume that the value $\tilde{y}$ of Y is observed, that $\underline{E}_X(\underline{E}_Y(\delta_{\{\tilde{y}\}}|X)) > 0$ and $\underline{E}_Y(\delta_{\{\tilde{y}\}}|x)$ is well defined for each $x\in\mathscr{X}$. The lower posterior expectation $\underline{E}(g|\tilde{y})$ is the unique solution μ of the following equation:*

$$\underline{E}_X\left(\underline{E}_Y\left((g-\mu)\delta_{\{\tilde{y}\}}|X\right)\right) = 0, \tag{10}$$

which for belief functions becomes:

$$0 = \int P_Z(dz)\inf_{x\in\Gamma(z)}\int P_{U|x}(du|x)\inf_{y\in\Gamma(u|x)}\delta_{\{\tilde{y}\}}(y)(g(x)-\mu). \tag{11}$$

$\square$

This is the extension of Corollary 1 to the case also the likelihood is a belief function. The proof of this theorem can be derived from the proof of [8, Th. 2] by using the expression for the lower expectation in (6).[2] The above result is very important for practical applications as shown in the next examples.

3 ε-Contamination and interval estimation

Consider an ε-contamination model for $(\mathscr{U},\mathscr{B}(\mathscr{U}),\Gamma(\cdot|x),P_{U|x})$, i.e., $P_{U|x}=(1-\varepsilon_m)\mathscr{N}(u;x,\sigma^2)+\varepsilon_m\delta_{\{u_0\}}(u|x)$ and $\Gamma(u|x)=y$ and $\Gamma(u_0|x)=\mathscr{A}_Y(x)=[x-b,x+b]$ for $b>0$. In the domain of the variable Y, this model is equivalent to: $y=x+(1-\varepsilon_m)n+\varepsilon_m v$, where n is a Gaussian noise with zero mean and variance σ^2, while v is a noise with unknown distribution. The only knowledge about v is its support $[-b,b]$ (norm bounded noise). This model can be used to account for the uncertainty in the measurement process which is due to a white noise component (n) and to the finite precision of the instrument (v), so it is very important for practical applications. Assume that also the prior over X is a ε-contamination model at the end of the Example 3. Applying (11) one gets:

$$\begin{aligned}0=&\int dz\left[(1-\varepsilon)\mathscr{N}(z;x_0,\sigma_0^2)+\varepsilon\delta_{\{z_0\}}(z)\right]\inf_{x\in\Gamma(z)}\\&\int du\left[(1-\varepsilon_m)\mathscr{N}(u;x,\sigma^2)+\varepsilon_m\delta_{\{u_0\}}(u|x)\right]\inf_{y\in\Gamma(u|x)}\delta_{\{\tilde{y}\}}(y)(g(x)-\mu).\end{aligned}\tag{12}$$

Consider the case where $\delta_{\{\tilde{y}\}}(y)$ is the limit for $|\Omega(\tilde{y})|\to 0$ of the following sequence of functions $\frac{1}{|\Omega(\tilde{y})|}I_{\{\Omega(\tilde{y})\}}$, where $\Omega(\tilde{y})$ is a ball centred at $\tilde{y}$ which does not depend on x and $|\Omega(\tilde{y})|$ is its Lebesgue volume [3, Sec. 6.10.4], [8]. Then, the previous integral equation can be rewritten as:

$$\begin{aligned}0=&\frac{1}{|\Omega(\tilde{y})|}\int dz\left[(1-\varepsilon)\mathscr{N}(z;x_0,\sigma_0^2)+\varepsilon\delta_{\{z_0\}}(z)\right]\inf_{x\in\Gamma(z)}\\&\int du\left[(1-\varepsilon_m)\mathscr{N}(u;x,\sigma^2)+\varepsilon_m\delta_{\{u_0\}}(u|x)\right]\inf_{y\in\Gamma(u|x)}I_{\{\Omega(\tilde{y})\}}(y)(g(x)-\mu).\end{aligned}\tag{13}$$

Since $|\Omega(\tilde{y})|$ is positive it can be simplified in the equation, which for $|\Omega(\tilde{y})|\to 0$ can be written as:

$$\begin{aligned}0=&\int dz\left[(1-\varepsilon)\mathscr{N}(z;x_0,\sigma_0^2)+\varepsilon\delta_{\{z_0\}}(z)\right]\inf_{x\in\Gamma(z)}\\&\left[(1-\varepsilon_m)\mathscr{N}(\tilde{y};x,\sigma^2)(g(x)-\mu)+\varepsilon_m\inf_{y\in\mathscr{A}_Y(x)}I_{\{\Omega(\tilde{y})\}}(y)(g(x)-\mu)\right]\\=&\int dz\left[(1-\varepsilon)\mathscr{N}(z;x_0,\sigma_0^2)+\varepsilon\delta_{\{z_0\}}(z)\right]\inf_{x\in\Gamma(z)}\\&\left[(1-\varepsilon_m)\mathscr{N}(\tilde{y};x,\sigma^2)(g(x)-\mu)-\varepsilon_m I_{\{x:\,\tilde{y}\in\mathscr{A}_Y(x)\}}(x)(g(x)-\mu)^-\right],\end{aligned}\tag{14}$$

where $(g(x)-\mu)^-=-\min(g(x)-\mu,0)$ is the negative part of $g-\mu$. Simplifying the other integral and exploiting that $\Gamma(z_0)=[-a,a]$, $\mathscr{A}_Y(x)=[x-b,x+b]$ and, thus, $\tilde{y}\in\mathscr{A}(x)$ implies $x\in[\tilde{y}-b,\tilde{y}+b]$, one finally gets:

[2] Observe that the proof in [8, Th. 2] has been obtained by assuming that the observation variables are discretized. Intuitively, we can see Theorem 2 as the limit of this result when the size of the discretization interval goes to zero.

$$0 = \int (1-\varepsilon)\mathscr{N}(x;x_0,\sigma_0^2)dx\left[(1-\varepsilon_m)\mathscr{N}(\tilde{y};x,\sigma^2)(g(x)-\mu)-\varepsilon_m I_{\{x\in[\tilde{y}-b,\tilde{y}+b]\}}(x)(g(x)-\mu)^-\right] + \varepsilon \inf_{x\in[-a,a]}\left[(1-\varepsilon_m)\mathscr{N}(\tilde{y};x,\sigma^2)(g(x)-\mu)-\varepsilon_m I_{x\in[\tilde{y}-b,\tilde{y}+b]}(x)(g(x)-\mu)^-\right].$$

Notice that in case $\varepsilon = \varepsilon_m = 0$ (no imprecision) and $g = X$, then $\mu = E(X|\tilde{y}) = (1/\sigma_0^2 + 1/\sigma^2)^{-1}(x_0/\sigma_0^2 + \tilde{y}/\sigma^2)$ that is the well known expression for the posterior mean in the Gaussian case. In the vacuous case, $\varepsilon = \varepsilon_m = 1$ (full imprecision), one gets:

$$0 = -\sup_{x\in[-a,a]} I_{\{x\in[\tilde{y}-b,\tilde{y}+b]\}}(x)(g(x)-\mu)^- = \sup_{x\in[-a,a]\cap[\tilde{y}-b,\tilde{y}+b]}(g(x)-\mu)^-. \quad (15)$$

For $g = X$ and assuming that the two intervals overlap, one has $\mu = \underline{E}(X|\tilde{y}) = \max(\tilde{y}-b,-a)$. Similarly, we can compute $\overline{E}(X|\tilde{y}) = \min(\tilde{y}+b,a)$. This is the well known updating formula in *interval estimation*, for instance in *set-membership estimation* [9, 10]. Figure 1 shows the expression for $\underline{E}(X|\tilde{y})$ in case $\varepsilon = 0$, $\varepsilon_m = 0.5$, $\tilde{y} = 1$, $x_0 = 0$, $\sigma_0^2 = \sigma^2 = 1$ and different values of b, i.e., $b \in [0,4]$. It can be observed that for $b < (1/\sigma_0^2 + 1/\sigma^2)^{-1}(x_0/\sigma_0^2 + \tilde{y}/\sigma^2) = 0.5$, the posterior mean coincides with that of the case $\varepsilon_m = 0$. The lower expectations starts to decrease when the support of the norm-bounded noise v, i.e., $[\tilde{y}-b,\tilde{y}+b]$, becomes larger than the interval $[0.5,1.5]$.

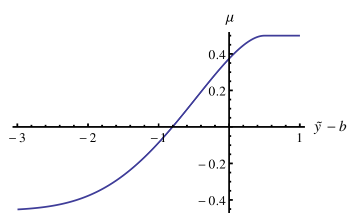

Fig. 1 Lower expectation $\mu = \underline{E}(X|\tilde{y})$ versus $\tilde{y}-b$ for $\tilde{y} = 1$ and $b \in [0,4]$.

4 Heavy-tail belief function model

Consider the model discussed at the end of Section 2 for the variable X with $p(x) = \mathscr{N}(x;0,2.19)$, i.e., variance 2.19. Assume that the measurement process is described by a normal density function $\mathscr{N}(y;x,\sigma^2)$. We can then use Corollary 1 to compute the lower posterior expectation of some function of interest g of X, i.e., $\underline{E}[g|y]$ is the unique solution μ of

$$0 = \int dx \mathscr{N}(x;0,2.19) \inf_{x'\in\Gamma(x)}(g(x')-\mu)\prod_{i=1}^n \mathscr{N}(y_i;x,\sigma^2), \quad (16)$$

where $\Gamma(x) = [x,1/x)$ for $x \in (-1,1)$ and $\Gamma(x) = x$ otherwise. The above equation can be solved numerically by discretizing X. In Table 1 (last row) we have reported the lower and upper posterior mean of X computed according to (16) in case $g = X$. For the sake of comparison we have reported also the posterior means obtained by the prior $p_1(x) = \mathscr{N}(x;0,2.19)$ (Normal distribution), denoted by $E_1(X|y)$ in the table, and $p_2(x) = \mathscr{C}(x;0,1)$ (Cauchy distribution), denoted by $E_2(X|y)$. Both these two prior distributions have prior mean equal to zero and and prior quartiles equal to ± 1.[3] From Table 1 it can be noticed that at the increasing of the prior-data

y	0	1	2	4.5	10
$E_1(X\|y)$	0	0.69	1.37	3.09	6.87
$E_2(X\|y)$	0	0.55	1.28	4.01	9.80
$\underline{E}(X\|y), \overline{E}(X\|y)$	-0.26,0.26	0.68,1.46	1.35,1.93	2.78,4.52	5.42,14.01

Table 1 Posterior mean computed for the three different prior models.

conflict (increasing of y) the Cauchy prior is more robust than the Normal prior, i.e., its posterior mean is closer to the value of the measurement. The third row in the table shows that the choice of a set of priors based on the heavy-tail belief function model further increases the robustness. Notice in fact that, for a small prior-data conflict, the interval $[\underline{E}(X|y), \overline{E}(X|y)]$ is tight and includes the posterior mean $E_1(X|y)$. However, at the increasing of the conflict, the interval enlarges highlighting the presence of a prior-data conflict, and its centre moves towards y similarly to the posterior mean of the Cauchy prior that moves towards y.

5 Conclusions

We have derived robust inferences based on classes of priors and likelihoods generated by belief functions. As future work, we intend to apply this work to practical estimation problems and to derive more closed convex sets of probability measures by using the multivalued mapping mechanism of belief functions.

Acknowledgements This work has been partially supported by the Swiss NSF grants n. 200020-121785/1.

References

1. A. P. Dempster, "Upper and lower probabilities induced by a multiple-valued mapping," *Ann. Math. Stat.*, vol. 38, pp. 325–339, 1967.
2. G. Shafer, *A mathematical theory of evidence*. Princeton University Press, 1976.
3. P. Walley, *Statistical Reasoning with Imprecise Probabilities*. New York: Chapman and Hall, 1991.
4. E. Miranda, G. de Cooman, and I. Couso, "Lower previsions induced by multi-valued mappings," *Journal of Statistical Planning and Inference*, vol. 133, no. 1, pp. 173–197, 2005.
5. L. A. Wasserman, "Prior envelopes based on belief functions," *The Annals of Statistics*, vol. 18, no. 1, pp. pp. 454–464, 1990.
6. L. A. Wasserman, "Belief functions and statistical inference," *The Canadian Journal of Statistics*, vol. 18, no. 3, pp. pp. 183–196, 1990.
7. I. Molchanov, *Theory of random sets*. Springer Verlag, 2005.
8. A. Benavoli, M. Zaffalon, and E. Miranda, "Robust filtering through coherent lower previsions," *Automatic Control, IEEE Transactions on*, vol. 56, pp. 1567 –1581, July 2011.
9. F. C. Schweppe, "Recursive state estimation: Unknown but bounded errors and system inputs," in *Adaptive Processes, Sixth Symposium on*, vol. 6, pp. 102 –107, oct. 1967.
10. D. Bertsekas and I. Rhodes, "Recursive state estimation for a set-membership description of uncertainty," *Automatic Control, IEEE Transactions on*, vol. 16, pp. 117 – 128, apr 1971.
11. J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York: Springer Series in Statistics, 1985.

[3] This example has been adapted from [11, Sec. 4.7.1.].