Imprecise Probabilities from Imprecise Descriptions of Real Numbers

Jonathan Lawry, Inés González-Rodríguez and Yongchuan Tang

University of Bristol, University of Cantabria and Zhejiang University

Information from Descriptions?

- Suppose x is a real value variable and LA is a set of labels describing $\Omega = \mathbb{R}$.
- If an agent learns that 'x is $high$' or perhaps 'x is $high \wedge \neg medium$', what can he/she infer about the value of x?
- Based on a prototype theory interpretation of label semantics we shall argue that such constraints naturally result in imprecise probabilities.
- The nature of this probabilistic information will depend on the available prior information about variable x.

Communicating Agents

- We consider a population of communicating agents, where each agent aims to describe elements from an underlying domain Ω, in such a way as to effectively communicate information to other agents.
- Descriptions are drawn from a finite set of labels LA, and the expressions LE generated recursively from LA using connectives $\wedge, \vee$ and $\neg$.
- For $x \in \Omega$, an agents needs to identify expressions $\theta \in LE$, which are appropriate to describe x.
- Appropriateness is governed by linguistic conventions which emerge through multiple communications between agents.

An Epistemic Model

- **Decision Problem:** Which expressions of the form 'x is θ' (for $\theta \in LE$) are assertible?
- **Epistemic Stance:** Agents assume as part of a practical decision making strategy that there is a clear dividing line between those labels which are, and those which are not appropriate to describe x.
- **Uncertainty:** Agents have only partial knowledge of the underlying linguistic conventions, obtained through the experience of their communications with other agents.
- Consequently they are uncertain as to which labels are appropriate to describe any given element of Ω.
- Label semantics introduces a calculus for such measures of appropriateness.

Mass Functions

- Agents should attempt to determine $\mathcal{D}_x$, which is the complete set of labels appropriate to describe x.
- $\mathcal{D}_x = \{medium, high\}$ means that both *medium* and *high* can be appropriately used to describe x, and no other labels are appropriate.
- $m_x : 2^{LA} \rightarrow [0, 1]$ is probability mass function on subsets of labels.
- For $F \subseteq LA$ $m_x(F)$ is the agent's subjective probability that $\mathcal{D}_x = F$.
- Assertions of the form 'x is θ' provides agents with constraints on the possible values of $\mathcal{D}_x$...

Appropriateness Measures

- For example, 'x is $high \wedge \neg medium$' provides the information that $high$ is an appropriate label to describe x, and $medium$ is not an appropriate label.
- This corresponds to the constraint $\mathcal{D}_x \in \{F \subseteq LA : high \in F, medium \notin F\}$.
- We can define a mapping $\lambda : LE \rightarrow 2^{2^{LA}}$ such that 'x is θ' provides the constraint $\mathcal{D}_x \in \lambda(\theta)$.
- For expression $\mu_\theta(x)$ is the agents subjective probability that θ is appropriate to describe x.
- This is given by $\mu_\theta(x) = \sum_{F \in \lambda(\theta)} m_x(F)$.

A Prototype Interpretation

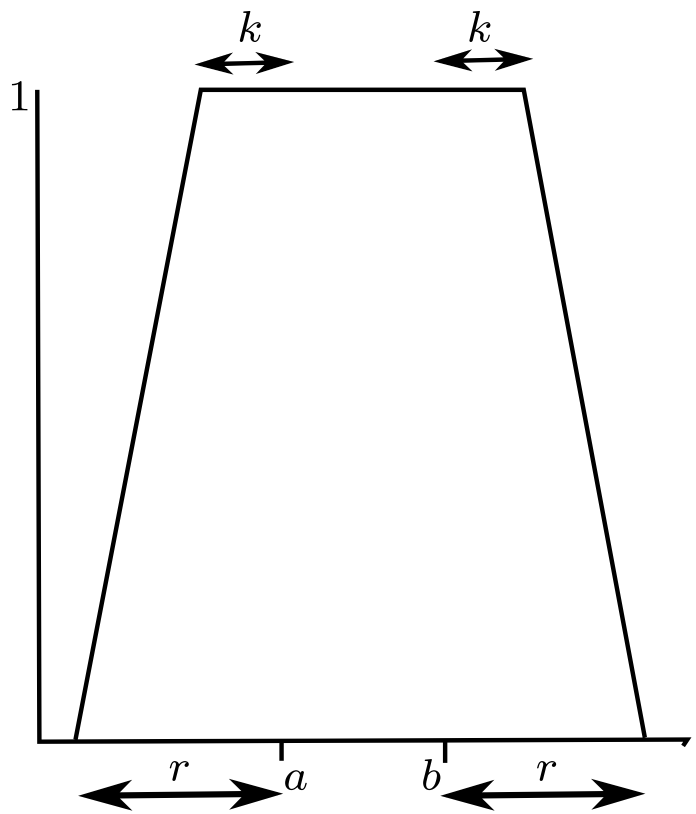
- Let $d : \Omega^2 \to [0, \infty)$ be a distance function satisfying $d(x, x) = 0$ and $d(x, y) = d(y, x)$.
- For $S, T \subseteq \Omega$ let $d(T, S) = \inf\{d(x, y) : x \in S, y \in T\}$.
- For $L_i \in \Omega$ let $P_i \subseteq \Omega$ be a set of prototypical elements for L_i.
- Let ϵ be a random variable into $[0, \infty)$ with density function δ.
- L_i is appropriate to describe x iff $d(x, P_i) \leq \epsilon$.
- $\mathcal{D}_x^\epsilon = \{L_i : d(x, P_i) \leq \epsilon\}$.
- $\forall F \subseteq \Omega \; m_x(F) = \delta(\{\epsilon : \mathcal{D}_x^\epsilon = F\})$

Neighbourhood Representation

- Let $\mathcal{N}^{\epsilon}_{L_i} = \{x \in \Omega : d(x, P_i) \leq \epsilon\}$ and then recursively so that $\forall \theta, \varphi \in \Omega$;
- $\mathcal{N}^{\epsilon}_{\theta \wedge \varphi} = \mathcal{N}^{\epsilon}_{\theta} \cap \mathcal{N}^{\epsilon}_{\varphi}$, $\mathcal{N}^{\epsilon}_{\theta \vee \varphi} = \mathcal{N}^{\epsilon}_{\theta} \cup \mathcal{N}^{\epsilon}_{\varphi}$, $\mathcal{N}^{\epsilon}_{\neg\theta} = (\mathcal{N}^{\epsilon}_{\theta})^c$.
- For θ not involving negation $\mathcal{N}^{\epsilon}_{\theta}$ is nested, otherwise not in general.
- Alternative characterisation: $\mu_\theta(x) = \delta(\{\epsilon : x \in \mathcal{N}^{\epsilon}_{\theta}\})$
- Labels represent sets of points *s*ufficiently similar to prototypes (Information Granules).
- Conditioning on information 'x is θ' corresponds to conditioning on the constraint that $x \in \mathcal{N}^{\epsilon}_{\theta}$.

Example (Uniform δ)

- Let $\Omega = \mathbb{R}$ and $d(x, y) = ||x - y||$. Let $L_i = about\ [a, b]$ where $a \leq b$, so that $P_i = [a, b]$.
- Let δ be the uniform distribution on $[k, r]$ for $0 \leq k < r$:



No Prior Knowledge about x

- Let $\Omega = \mathbb{R}$ and $d(x, y) = \|x - y\|$ and consider a set LA of number labels L_i describing $\mathbb{R}$ with prototype sets P_i each corresponding to an interval of $\mathbb{R}$.
- Given a real valued random variable x for which we know only that 'x is θ' for some $\theta \in LE$ we define upper and lower cumulative distribution functions for the probability that $x \leq y$ as follows:

$$\underline{F}(y|\theta) = \delta_\theta(\{\epsilon : \mathcal{N}_\theta^\epsilon \subseteq (-\infty, y]\}) \text{ and}$$

$$\overline{F}(y|\theta) = \delta_\theta(\{\epsilon : \mathcal{N}_\theta^\epsilon \cap (-\infty, y] \neq \emptyset\}) \text{ and where}$$

$$\delta_\theta(\epsilon) = \begin{cases} \frac{\delta(\epsilon)}{\int_{\epsilon:\mathcal{N}_\theta^\epsilon \neq \emptyset} \delta(\epsilon) d\epsilon} : \mathcal{N}_\theta^\epsilon \neq \emptyset \\ 0 : \text{otherwise} \end{cases}$$

Prior Density on x

- Now suppose we have prior information that x is distributed according to density function $p(x)$.
- Let $F(y|\mathcal{N}_\theta^\epsilon)$ denote the corresponding updated cumulative distribution given the constraint that $x \in \mathcal{N}_\theta^\epsilon$ for a fixed ϵ.
- The values of $F(y|\mathcal{N}_\theta^\epsilon)$ are uncertain given the remaining uncertainty about the value of the threshold ϵ.
- Hence, we define a second order cumulative distribution: $\forall p \in [0,1]$

$$\tilde{F}_{y,\theta}(p) = \delta_\theta(\{\epsilon : F(y|N_\theta^\epsilon) \leq p\})$$

Expected Density

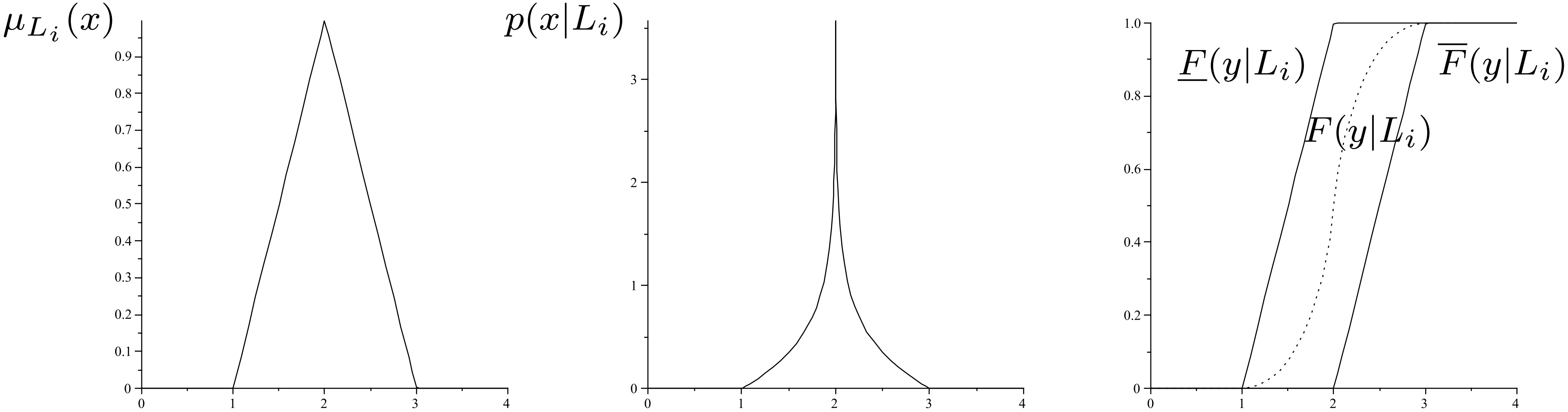
- If a precise posterior distribution is required conditional on θ, then one possibility is to take the expected value of posterior distributions given $\mathcal{N}_\theta^\epsilon$, as ϵ varies.
- Taking the expected value of $p(x|\mathcal{N}_\theta^\epsilon)$ as ϵ varies we obtain:

$$p(x|\theta) = E_{\delta_\theta}(p(x|\mathcal{N}_\theta^\epsilon))$$

- This is consistent with the upper and lower cumulative distributions:
- For $y \in \mathbb{R}$ and $\theta \in LE$, $\underline{F}(y|\theta) \leq F(y|\theta) \leq \overline{F}(y|\theta)$ where $F(y|\theta) = \int_{-\infty}^{y} p(x|\theta)dx = E_{\delta_\theta}(F(y|\mathcal{N}_\theta^\epsilon))$.

Example (Conditioning)

- Consider the number label $L_i = about\ 2$ for which $P_i = \{2\}$ and let δ be the uniform density on $[0,1]$.
- Let the prior density $p(x)$ correspond to the uniform distribution on $[0,10]$.



Appropriateness
Expected Posterior
Envelope