The Geometry of Imprecise Inference

Miķelis Bickis
University of Saskatchewan, Saskatoon, Saskatchewan, Canada

Manifold of probability measures

Let $\mathcal{Y}$ be an observation space and P_0 a probability measure on some σ-algebra of events defined on $\mathcal{Y}$. Let P_1 be another probability measure having the same null sets. Then the log likelihood ratio for distinguishing these two measures can be written as

$$\log \frac{dP_1}{dP_0} = v - I(P_0|P_1) \tag{1}$$

where

$$E_0(v) = \int v \, dP_0 = 0$$

and

$$I(P_0|P_1) = \int \log \frac{dP_0}{dP_1} \, dP_0$$

is the Kullback-Leibler information from P_0 to P_1.
From this we can write

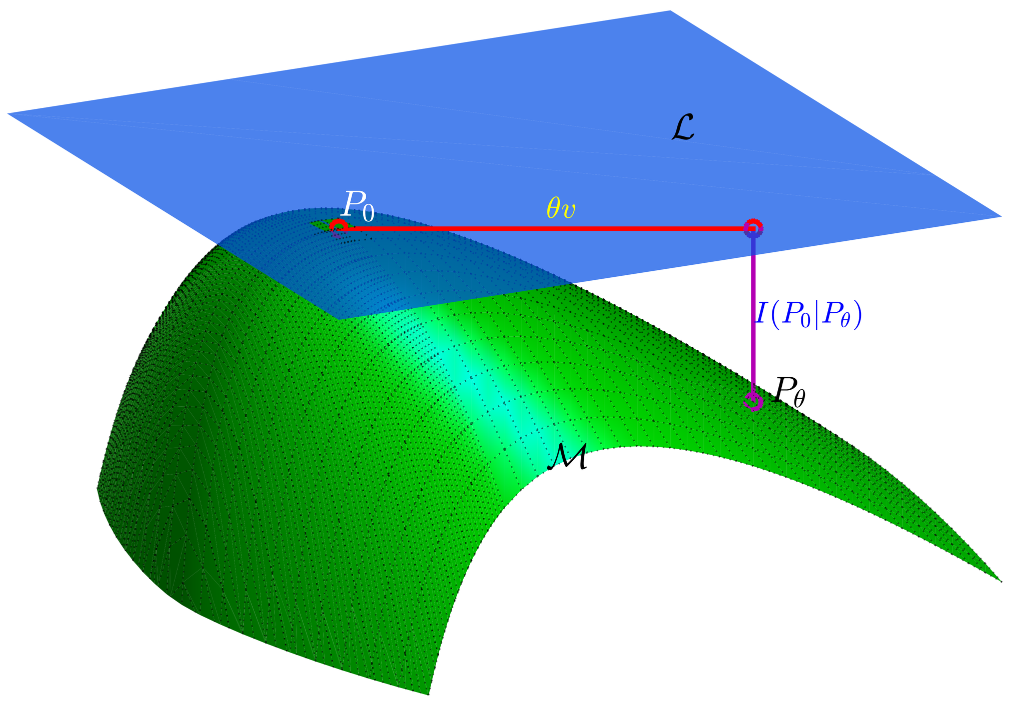
$$P_1(A) = \int \mathbf{1}_A \, e^{v - I(P_0|P_1)} \, dP_0,$$

and introducing a scalar parameter θ we can define a one-dimensional exponential family $\mathcal{P} = \{P_\theta\}$ by

$$P_\theta(A) = \int \mathbf{1}_A \, e^{\theta v - I(P_0|P_\theta)} \, dP_0, \tag{2}$$

provided that

$$\int e^{\theta v} \, dP_0 < \infty.$$



More generally, given a list $\mathbf{v} = (v_1, \ldots, v_k)$ of linearly independent random variables with

$$E_0(v_i) = 0, \quad i = 1, \ldots, k \tag{3}$$

we can define a k-dimensional exponential family

$$P_{\boldsymbol{\theta}}(A) = \int \mathbf{1}_A \, e^{\boldsymbol{\theta}^\top \mathbf{v} - I(P_0|P_{\boldsymbol{\theta}})} \, dP_0, \tag{4}$$

where

$$\boldsymbol{\theta} \in \Theta = \{\boldsymbol{\theta} \in \mathbb{R}^k : I(P_0|P_{\boldsymbol{\theta}}) < \infty\}.$$

Θ will be a convex set in $\mathbb{R}^k$. The set of probability measures thus defines a k-dimensional manifold

$$\mathcal{M} = \{(-I(P_0|P_{\boldsymbol{\theta}}), \theta_1, \ldots, \theta_k) : I(P_0|P_{\boldsymbol{\theta}}) < \infty\}$$

embedded in $\mathbb{R}^{k+1}$. This manifold can be projected one-to-one onto $\mathcal{L}$, its tangent space at P_0. The set of probability measures can then be represented uniquely by vectors in this tangent space, or by the parameters in Θ.

Remark: The constraint (3) ensures that the vector space $\mathcal{L}$ is tangent to the manifold. It is not essential for constructing a correspondence between vectors and probability measures.

Dual manifold of priors and posteriors

A prior distribution Π_0 can be considered equivalently as a probability measure on $\mathcal{M}$, $\mathcal{L}$, or Θ. We will consider it defined on $\mathcal{L}$. Then the construction analogous to (1) requires functions defined on $\mathcal{L}$. In fact, for any observation $y \in \mathcal{Y}$, the *evaluation functional* which maps

$$v \mapsto v(y)$$

is a linear function on $\mathcal{L}$ as is the mapping

$$v \mapsto v(y) - \int w(y) \, d\Pi_0(w),$$

the latter also satisfying (3). Denote by π_0 the density of Π_0. Then for a family defined by (4) Bayes' rule will give the posterior density

$$\pi_y(v) = \frac{\pi_0(\boldsymbol{\theta}) \exp(v(y) - I(P_0|P_v))}{\int \exp(w(y) - I(P_0|P_w)) \, d\Pi_0(w)}.$$

This can be written in the equivalent form:

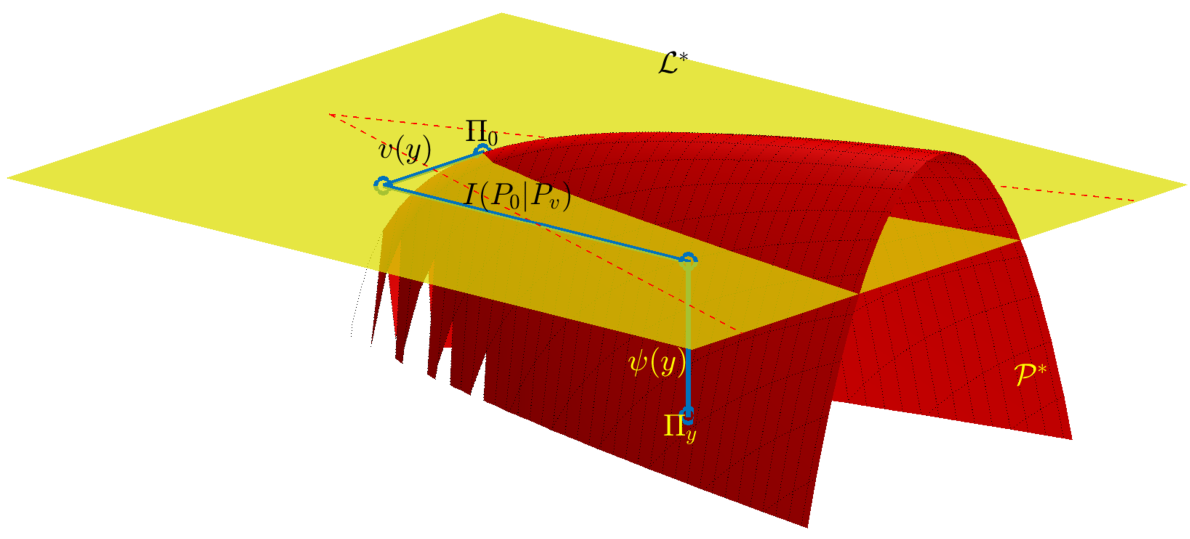
$$\log \frac{d\Pi_y}{d\Pi_0}(v) = v(y) - I(P_0|P_v) - \psi(y) \tag{5}$$

where

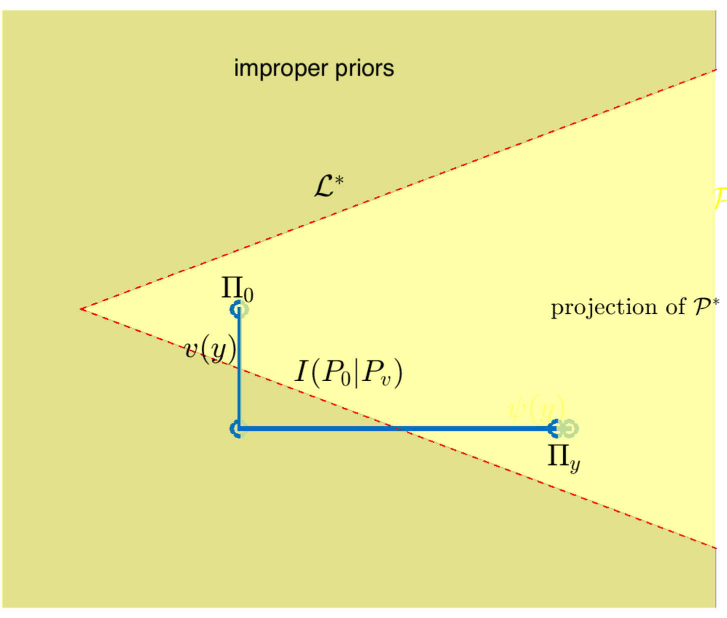
$$\psi(y) = \log \int \exp\left(v(y) - I(P_0|P_v)\right) \, d\Pi_0(v).$$

This is analogous to (1) in which the posterior distribution can be expressed as a shift from the prior by the evaluation functional as well as the function $v \mapsto -I(P_0|P_v)$.
Note that

- the first two terms in (5) do not depend on the prior,
- the second term does not depend on the observation y,
- the third term depends on y but not on the vector v.



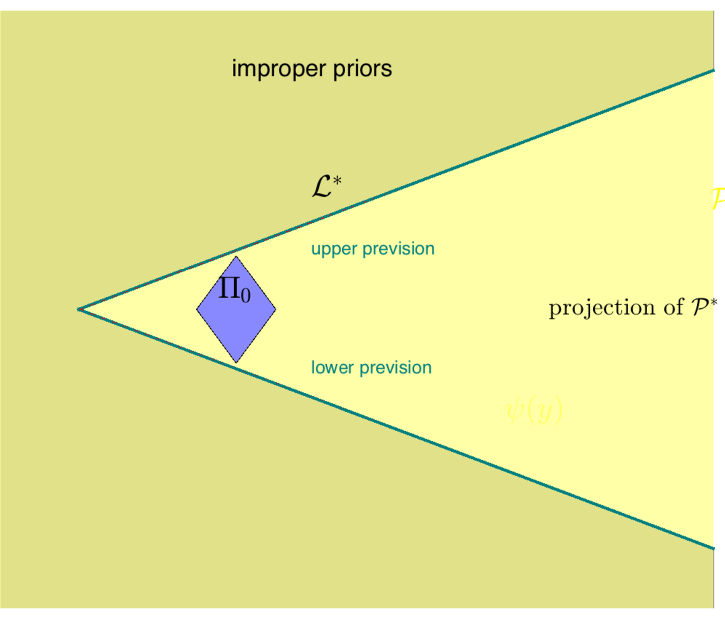
By ignoring $\psi(y)$, each possible posterior can be projected uniquely onto the linear space $\mathcal{L}^*$ spanned by the evaluation functionals and $I(P_0|P_v)$. The dimension of $\mathcal{L}^*$ is one greater than that of $\mathcal{L}$. Using the construction (2), the elements of $\mathcal{L}^*$ will then define an exponential family $\mathcal{P}^*$ containing the prior and all possible posteriors.



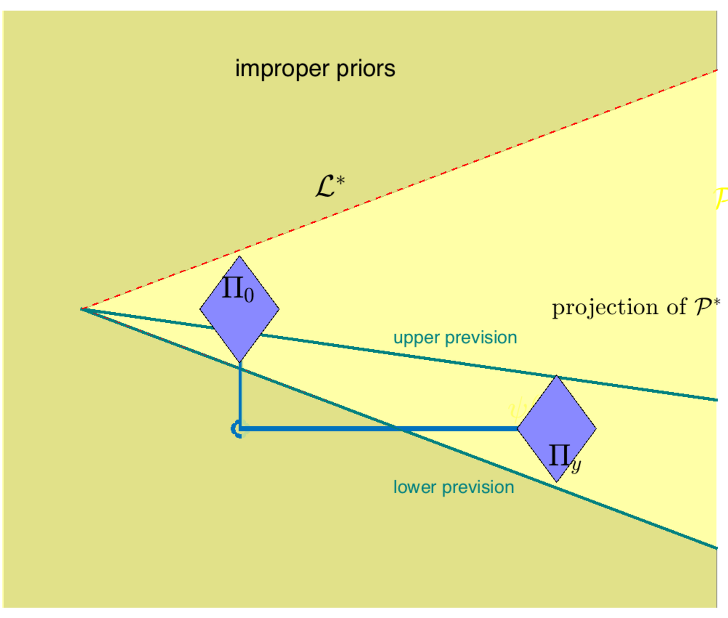
In the present construction, the space $\mathcal{L}^*$ is not tangent to the manifold $\mathcal{P}^*$. This could be achieved by subtracting from $v(y)$ and $I(P_0|P_v)$ their Π_0-expectations. Such an adjustment would change the specific correspondence between $\mathcal{L}^*$ and $\mathcal{P}^*$, but not the principle.

Imprecise updating

An imprecise prior can be represented as a (convex) set of priors. Such a set can be visualized as a set in $\mathcal{L}^*$. Since the first two terms in (5) do not depend on the prior, and the third terms is "projected away", Bayesian updating of the prior set appears as a simple translation of the entire set.



Upper and lower posterior previsions of parametric functions can be determined as suprema and infima over the translated set. If the level sets of such functions are linear in $\mathcal{L}^*$, then the extrema will occur at the extreme points of the posterior set.



Research supported by the Natural Sciences and Engineering Research Council of Canada

ISIPTA'15, Pescara, Italy, 20 July 2015