Example 3. In this example, assume that you want to estimate $\Phi(\mu^\dagger) = \mu^\dagger[a, 1]$ for some $a \in (0, 1)$, where $\mu^\dagger$ is an unknown distribution on the interval $[0, 1]$, based on the observation of n data points $d_1, \ldots, d_n$ up to resolution δ (i.e. we observe $d_i \in B_\delta(x_i)$ with $x_i \in [0, 1]$ for $i = 1, \ldots, n$). Our purpose is to examine the sensitivity of the Bayesian answer to this problem with respect to the choice of a particular prior. Consider the model class $\mathcal{A} := \mathcal{M}([0, 1])$ consisting of the full set of probability measures on the unit interval and the class of priors

$$\Pi := \left\{\pi \in \mathcal{M}(\mathcal{A}) \,\middle|\, \mathbb{E}_{\mu\sim\pi}\big[\mathbb{E}_\mu[X]\big] = m\right\},$$

corresponding to the assumption that $\mu^\dagger$ is the realization of a random measure on $[0, 1]$ whose mean is on average m with $0 < m < a$. As in the previous example, the finite codimensional class of priors Π leads to brittleness in the sense that, although, the least upper bound on prior values is

$$\mathcal{U}(\Pi) = \frac{m}{a} < 1, \tag{1}$$

for $\delta \ll 1/n$, the least upper bound on posterior values is equal to the deterministic supremum of the quantity of interest (over $\mathcal{A}$), i.e.

$$\mathcal{U}(\Pi|B_\delta^n) = 1. \tag{2}$$

Since worst priors are obtained by selecting priors for which the probability of observing the data $\mu^n[B_\delta^n]$ is arbitrarily close to zero except when $\Phi(\mu)$ is close to its deterministic supremum, it is natural to ask if this brittleness can be avoided by adding a uniform constraint on the probability of observing the data in the model class. To investigate this question, let us introduce $\alpha \geq 1$ and a probability measure μ_0 on $[0, 1]$ with strictly positive Lebesgue density (with a prototypical example being that μ_0 is itself the uniform measure on $[0, 1]$), and consider the new model class

$$\mathcal{A}(\alpha) := \left\{\mu \in \mathcal{M}[0, 1] \,\middle|\, \frac{1}{\alpha}\mu_0^n[B_\delta^n] \leq \mu^n[B_\delta^n] \leq \alpha\mu_0^n[B_\delta^n]\right\}, \tag{3}$$

and the new class of priors

$$\Pi(\alpha) := \left\{\pi \in \mathcal{M}(\mathcal{A}(\alpha)) \,\middle|\, \mathbb{E}_{\mu\sim\pi}\big[\mathbb{E}_\mu[X]\big] = m\right\}. \tag{4}$$

Note that, for the model class $\mathcal{A}(\alpha)$, the probability of observing the data is uniformly bounded below by $\frac{1}{\alpha}\mu_0^n[B_\delta^n]$ and above by $\alpha\mu_0^n[B_\delta^n]$.

Then, the calculus developed in http://arxiv.org/abs/1304.6772 allows to us compute the least upper bound on posterior values to be

$$\lim_{\delta\to 0}\mathcal{U}\big(\Pi(\alpha)|B_\delta^n\big) = \frac{1}{1 + \frac{1}{\alpha^2}\frac{a-m}{m}} = \frac{m}{\frac{a}{\alpha^2} + m(1 - \frac{1}{\alpha^2})}. \tag{5}$$

Therefore, for $\alpha = 1$, when the probability of observing the data is uniform in the model class, prior values are equal to posterior values (i.e. $= \frac{m}{a}$), and the method is robust but learning is impossible. Moreover, when α slightly deviates from 1, then the right hand side of (5) *quickly* moves from m/a towards 1 as α increases. As a numerical application observe that for $a = \frac{3}{4}$ and $m = \frac{a}{2} = \frac{3}{8}$, we have $\lim_{\delta \to 0} \mathcal{U}\big(\Pi(\alpha)\big) = \frac{1}{2}$ and

$$\lim_{\delta \to 0} \mathcal{U}\big(\Pi(\alpha)|B_\delta^n\big) = \frac{1}{1 + \frac{1}{\alpha^2}} \,.$$

Therefore, for $\alpha = 2$, we have (irrespective of the number of data points)

$$\lim_{\delta \to 0} \mathcal{U}\big(\Pi(2)|B_\delta^n\big) = 0.8,$$

and for $\alpha = 10$, we have (irrespective of the number of data points)

$$\lim_{\delta \to 0} \mathcal{U}\big(\Pi(10)|B_\delta^n\big) \approx 0.99.$$

Moreover, if α is derived by assuming the probability of each data point to be controlled up to some tolerance γ, i.e. if the model class $\mathcal{A}(\alpha)$ is replaced by

$$\mathcal{A}_\gamma := \left\{ \mu \in \mathcal{M}[0,1] \,\middle|\, \frac{1}{\gamma}\mu_0[B_\delta(x_i)] \leq \mu[B_\delta(x_i)] \leq \gamma \mu_0[B_\delta(x_i)]\,, \text{ for } i = 1, \dots, n \right\} \tag{6}$$

for some $\gamma > 1$ and the prior class $\Pi(\gamma)$ defined in the same way as for $\Pi(\alpha)$, then it can be shown that

$$\lim_{\delta \to 0} \mathcal{U}(\Pi(\gamma)|B_\delta^n) = \frac{1}{1 + \frac{1}{\gamma^{2n}}},$$

which exponentially converges towards 1 as the number n of data points goes to infinity.

In conclusion, the effects of a uniform constraint on the probability of the data under finite information in the model class show that learning ability comes at the price of loss in stability in the following sense: when $\alpha = 1$, the data is equiprobable under all measures in the model class, posterior values are equal to prior values, and the method is robust but learning is not possible. As α deviates from one, the learning ability increases as robustness decreases, and when α is large, learning is possible but the method is brittle.