Learning with Misspecified Models*

Dániel Csaba† Bálint Szőke‡

November 10, 2018

Abstract

We consider Bayesian learning about a stable environment when the learner's entertained probability distributions (likelihoods) are all misspecified. We evaluate likelihoods according to the long-run average payoff of the policy function they induce. We then show that, generically, the value that the Bayesian learner attains in the long run is lower than what would be achievable with her misspecified set of likelihoods. We introduce two kinds of indifference curves over the learner's set: one based on the likelihoods' induced long-run average payoff, and another capturing their statistical similarity. In case of misspecification, we show that misalignment of these curves can lead the Bayesian learner to focus on payoff-irrelevant features of the environment. On the other hand, under correct specification this misalignment has no bite. We provide conditions under which it is feasible to construct an exponential family that allows the learner to implement the best attainable policy in the long-run irrespective of misspecification. We demonstrate applications of the introduced concepts through examples.

*We thank Timothy Christensen for helpful discussions and comments.

†Department of Economics, New York University, E-mail: daniel.csaba@nyu.edu.

‡Department of Economics, New York University, E-mail: balint.szoke@nyu.edu.

1 Introduction

We consider a Bayesian decision maker (DM) learning about an exogenous stochastic process with the aim of implementing a policy function that maximizes her long-run average payoff. As a point of reference, we start with a set of distributions (likelihoods) $\mathcal{P}$, such that the unknown data generating process (DGP) belongs to $\mathcal{P}$, and define a^* as the DM's best policy relative to this set. Learning about the DGP entails assigning nonzero prior probabilities to a *subset* of $\mathcal{P}$ and using signals to infer which one of these likelihoods is most probable. If the entertained set contains the data generating process, i.e., the learner's set is correctly specified, and satisfies certain regularity conditions, Bayesian learning has the desirable property that it converges to the DGP as more and more data arrives. This property is important for the decision maker, because it implies that she can implement a^* in the long run. However, since $\mathcal{P}$ can be highly complex, the assumption of correct specification is often implausible. It is therefore of interest to see if and when a learner with a misspecified set could implement a^* in the long run.

We show that for Bayesian learning with a misspecified set of likelihoods the answer is typically negative. What's more, misspecification can cause the learner to implement a policy function that is suboptimal even *relative to the entertained set.* We call the implemented policy suboptimal if there exist other likelihoods in the DM's set that could induce policies with higher long-run average payoff. We provide conditions under which such outcome is avoidable, and what's more the implementation of a^* is guaranteed in the long run. The key idea is to tailor the misspecified set to preferences. We gain intuition about this process by introducing indifference curves over the decision maker's set that link likelihoods with the same level of induced long-run average payoff. We contrast the arising "geometry" with another one that characterizes statistical similarity of the different likelihoods. When the decision maker's set is correctly specified, the alignment of the two geometries is irrelevant for determining the long term consequences of learning. However, in case of misspecification, inconsistency among these curves makes the Bayesian learner to focus on "wrong" features of the DGP, which could lead to the implementation of a suboptimal policy in the long-run.

We demonstrate our introduced concepts through examples of well-known economic problems. After diagnosing the source of suboptimality, we pose conditions under which implementing the optimal policy a^* in the long-run is feasible even if all considered likelihoods are wrong. In particular, we recommend to construct an *exponential family* using the decision maker's "welfare-relevant moments" as sufficient statistics. These moments are derived from the underlying decision problem and as such they are determined by preference as well as pertinent features of the environment. By using our exponential family as the set of entertained likelihoods, the decision maker ensure that she learns about the "right" features of the DGP.

We follow the framework of statistical decision theory proposed by Wald [1950]. The primitives are: (i) a payoff function and (ii) a set of entertained likelihoods and a prior. Despite having a fully-embraced prior to make forward-looking decisions, ideally, the DM would want to maximize her payoff function *under the DGP*. That said, we deviate from the subjectivist Bayesian view by assuming that the DM treats her likelihoods as instruments to choose "good" policies, rather than as indisputable part of preferences or as providing psychological value or utility by themselves. This assumption is necessary for any reasonable form of assessment of models that are misspecified. Should we take the subjectivist view, the DM would always do her best according to the wrong model her decisions are based on.

In order to avoid this circular argument, we follow Blume et al. [2018] and take the perspective of an "outside observer". We confine ourselves to the limiting probability distributions emerging from Bayesian learning in ergodic settings. This allows us to compare potentially misspecified priors based on the long-run average payoffs that their limit point-induced policy functions generate. We focus on Bayesian learning, but our findings are applicable to any consistent learning rule such that the entertained hypotheses are representable with probability distributions of the observable variables and the probability of each hypothesis gets updated as new information arrives.[1]

Our results illustrate that the complexity of the entertained set should be guided by the DM's preferences as opposed to the statistical complexity of the environment. Even if the DGP is highly complex, a simple misspecified set can implement the optimal policy if it is targeted at the appropriate features of the environment. Our condition can thus be viewed as a recipe for model building.

The rest of the paper is structured as follows. Section 2 introduces the key concepts and notation. In section 3, we demonstrate how misspecification can lead to suboptimal long-run behavior through a series of examples. Section 4 presents our recommendation that can be used to resolve the source of suboptimality in certain cases. The related literature is summarized in section 5. Section 6 concludes.

2 General framework and notation

The environment is described by a probability space $(\Omega, \mathcal{F}, P)$ such that a strictly stationary and ergodic observable state vector X takes values in the measurable space $(\mathcal{X}, \Xi)$ with distribution P. Alternative descriptions of the environment can be obtained by replacing P with some other distribution. Loosely, we use $\mathcal{P}$ to denote the set of distributions over $\mathcal{X}$ that are deemed plausible for the specific problem at hand.

The decision maker chooses a *policy function*, $a\colon \mathcal{X} \mapsto \mathcal{C}$, that assigns a particular action from some choice set $\mathcal{C}$ to every realization x of the state vector X. Let $\mathcal{A}$ denote the collection of all possible policy functions. The payoffs are described by the period *utility function*, $u\colon \mathcal{C} \times \mathcal{X} \mapsto \mathbb{R}$, and possibly depend on the value of the states. To state the decision maker's objective, we introduce a functional, $\mathsf{U}\colon \mathcal{A} \times \mathcal{P} \mapsto \mathbb{R}$, defined as,

$$\mathsf{U}(a, Q) := \int_{\mathcal{X}} u(a(x), x)\mathrm{d}Q(x). \tag{1}$$

U is the *expected payoff* induced by policy function $a \in \mathcal{A}$ under distribution $Q \in \mathcal{P}$. Ideally, the decision maker would want a policy function $a^* \in \mathcal{A}$ that maximizes $\mathsf{U}(\cdot, P)$—the long-run average payoff—providing her with the highest attainable value, $\overline{\mathsf{U}} := \mathsf{U}\left(a^*, P\right)$.

However, P is unknown, so the decision maker has to solve an alternative problem with objective $\mathsf{U}(\cdot, Q)$ in which P is replaced with some approximating distribution Q. To this end, she entertains a set of *likelihoods* $\mathcal{M}$, i.e., a family of strictly stationary and ergodic probability distributions $Q_\theta \in \mathcal{P}$ each indexed by a finite

[1] This includes both Bayesian learning and anticipated utility learning accompanied with some frequentist procedures. On the other hand, for the sake of clean exposition, we do not consider 'active' learning, that is, the implemented policy function does not affect the information the DM observes or the mechanism that she wants to learn about.

parameter vector θ,

$$\mathcal{M} := \{Q_\theta \ : \ \theta \in \Theta\}, \qquad \text{where} \quad \Theta \subseteq \mathbb{R}^p. \tag{2}$$

We are interested in situations, when there is no guarantee that $P \in \mathcal{M}$, i.e., that $\mathcal{M}$ is *correctly specified*. We call the set $\mathcal{M}$ *misspecified* if $P \notin \mathcal{M}$.

Initialized with some prior distribution over $\mathcal{M}$, Bayes' rule induces a sequence of posteriors that summarize the decision maker's best guesses for P at every point in time after the available data is taken into account. It is well known that under certain regularity conditions [Shalizi, 2009], these posteriors will eventually concentrate on the likelihoods in $\mathcal{M}$ that minimize the Kullback-Leibler (KL) divergence from the data generating mechanism,[2]

$$\Theta_{\mathrm{KL}} := \arg\min_{\theta \in \Theta} \ D_{\mathrm{KL}}\left(P \parallel Q_\theta\right). \tag{3}$$

In other words, after observing an infinite sequence of signals, Bayes' rule will suggest using a distribution from Θ_{KL} as the best approximation of P. However, while the elements of Θ_{KL} has interesting information theoretic interpretations, it is not obvious whether they provide the highest long-run average payoff in $\mathcal{M}$.

We formulate this question by defining a function that quantifies the relative performance of alternative likelihoods in terms of the long-run average payoff of the action they induce. For fixed P, this function will give a "performance measure" of Q analogous to D_{KL}. A crucial component of this object is the *best response function*, that is a mapping, $b\colon \mathcal{P} \mapsto \mathcal{A}$, defined as,[3]

$$b(Q) := \arg\max_{a \in \mathcal{A}} \ \mathsf{U}(a, Q). \tag{4}$$

Clearly, $b(P) = a^*$. Combining b with the expected payoff function, U, yields $D_{\mathsf{U}}\colon \mathcal{P} \times \mathcal{P} \mapsto \mathbb{R}_+$, defined as,

$$D_{\mathsf{U}}(P \parallel Q) := \mathsf{U}\left(b(P), P\right) - \mathsf{U}\left(b(Q), P\right) = \overline{\mathsf{U}} - \mathsf{U}\left(b(Q), P\right). \tag{5}$$

For fixed P, both D_{KL} and D_{U} attain their global minima (zero) at $Q = P$. Moreover, the likelihoods that yield the highest long-run average payoff in $\mathcal{M}$ can be characterized by,

$$\Theta_{\mathsf{U}} := \arg\min_{\theta \in \Theta} \ D_{\mathsf{U}}\left(P \parallel Q_\theta\right). \tag{6}$$

Our main question is whether the "minimal optimality condition" of $\Theta_{\mathrm{KL}} \subseteq \Theta_{\mathsf{U}}$ is satisfied. Can the Bayesian learning rule guarantee that in the long run the decision maker will get the highest value that is attainable with her set $\mathcal{M}$? We show that the answer depends on the relationship between D_{KL} and D_{U}. If the decision maker's set is correctly specified, the answer is clearly affirmative, because the two measures attain their global minima at the same point. On the other hand, in the more relevant misspecified case, we argue that the answer is *generically* in the negative, because in general D_{KL} and D_{U} are inconsistent with each other.

[2] The KL divergence is $D_{\mathrm{KL}}(P \parallel Q) = \int \log \frac{p(x)}{q(x)} \mathrm{d}P(x) = \int \log p(x) \mathrm{d}P(x) - \int \log q(x) \mathrm{d}P(x)$. Note that in principle Θ_{KL} is a set, however, we shall only consider cases when it is a singleton. To emphasize this, we will use the notation θ_{KL}.

[3] Considering only learning rules that converge to a single likelihood asymptotically, we can determine the best responding functions without having to consider mixture distributions.

3 Illustrative examples

Consider an income fluctuation problem with a risk averse agent who can borrow or lend at a constant rate of interest, r, subject to the period budget constraint (and the corresponding transversality condition),

$$c_t + w_t = (1+r)w_{t-1} + x_t,$$

where c_t denotes consumption, x_t is the agent's uncertain labor income, and w_t is her financial wealth at the end of period t. For simplicity, we impose $\beta(1+r) = 1$, where β is the agent's discount factor. Suppose that the reference set $\mathcal{P}$ includes distributions according to which every period labor income is drawn *i.i.d.* from some unknown distribution $P \in \mathcal{P}$.

To illustrate the effect of preferences on determining the relevant features of labor income, we derive best response functions for two well-known utility specifications,[4]

(i) quadratic preferences with $u(c) = -\frac{1}{2}(c-\bar{c})^2$ and $\bar{c} > 0$, that imply:

$$b_{\mathrm{q}}(Q) = rw + \mathbb{E}_Q\left[X\right]; \tag{7}$$

(ii) constant absolute risk averse (CARA) preferences with $u(c) = -\eta^{-1}\exp(-\eta c)$ and $\eta > 0$, that imply:

$$b_{\mathrm{c}}(Q) = rw + \mathbb{E}_Q\left[X\right] - \frac{1}{r\eta}\log\mathbb{E}_Q\left[\exp\left(-\frac{r\eta}{1+r}X\right)\right]. \tag{8}$$

Evidently, in both cases, the specific Q affects the agent's best action through the implied expected labor income. In fact, as the well-known *certainty equivalence* property of linear-quadratic control problems suggests, with quadratic preferences the first moment $m_{\mathrm{q}}(Q) := \mathbb{E}_Q[X]$ is the only feature of the income process that matters for the agent's decisions.

On the other hand, with CARA preferences precautionary motives are also present, so the agent's optimal action depends on higher moments as well. Nonetheless, the effect of higher moments are conveniently summarized by the last term of (8), so similar to the quadratic case, we can define a vector,

$$m_{\mathrm{c}}(Q) := \left(\mathbb{E}_Q\left[X\right], \quad \mathbb{E}_Q\left[\exp\left(-\frac{r\eta}{1+r}X\right)\right]\right)',$$

that captures all welfare-relevant features of X. The fact that in both cases, we are able to summarize the relevant features of X with a vector of moments permits an intuitive characterization of D_{U}.

Fact 1. *Let $i \in \{q, c\}$. If two distributions Q, Q' are such that $m_i(Q) = m_i(Q')$, they induce the same policy functions, thus $b_i(Q) = b_i(Q')$ and $D_{\mathsf{U}}(P \parallel Q) = D_{\mathsf{U}}(P \parallel Q')$.*

As for the entertained set of likelihoods, assume first that the agent attempts to learn about P using

[4] See the seminal papers by Hall [1978] and Caballero [1990] for the quadratic and CARA specifications, respectively.

likelihoods that describe X as being *i.i.d.* log-normal parameterized by $\theta = (\mu, \sigma^2)$,

$$\mathcal{M} = \left\{ \ \log\mathcal{N}\left(\mu, \sigma^2\right) \ : \ (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+ \ \right\}. \tag{9}$$

During the learning process, Q is a mixture model arising from combining the distributions in $\mathcal{M}$ using posterior probabilities. However, since the agent is assumed to settle on a single likelihood asymptotically, for our purposes, it is sufficient to focus only on the individual parametric distributions of $\mathcal{M}$. In particular, Fact 1 implies that if there is a $Q_\theta \in \mathcal{M}$ such that $m(Q_\theta) = m(P)$, then the agent is able to implement a^* and attain the maximum possible long-run average payoff, $D_{\mathsf{U}}(P \parallel Q_\theta) = 0$, even if all of her likelihoods are wrong. Indeed, because the moments in m enter the policy functions in an additive manner, there are in principle many likelihoods in $\mathcal{M}$ that can satisfy this property.

In particular, when the payoff function is quadratic, the D_{U}-minimizing likelihoods are characterized by,

$$\Theta^{\mathrm{q}}_{\mathsf{U}} = \left\{ \ \left(\mu, \sigma^2\right) \ : \ \exp\left(\mu + \frac{\sigma^2}{2}\right) = \mathbb{E}_P[X] \ \right\}.$$

As for CARA preferences, notice that the extra moment that the agent cares about is the Laplace transform of the perceived distribution of labor income evaluated at $\eta r/(1+r)$ and use $\mathcal{L}(Q)$ to denote this object. Although there is no closed form solution for the Laplace transform of the lognormal distribution, good approximations exist that we can use to compute $\mathcal{L}(Q)$ for the likelihoods in $\mathcal{M}$. For the sake of transparency, we use $\mathcal{L}(\mu, \sigma^2)$ to denote these values. That said, the D_{U}-minimizing likelihoods with CARA utility are,

$$\Theta^{\mathrm{c}}_{\mathsf{U}} = \left\{ \ \left(\mu, \sigma^2\right) \ : \ \exp\left(\mu + \frac{\sigma^2}{2}\right) - \frac{1}{\eta r}\log\mathcal{L}\left(\mu, \sigma^2\right) = \mathbb{E}_P[X] - \frac{1}{\eta r}\log\mathcal{L}(P)\right\}.$$

Of course, because the agent's long-run policy function is determined by her learning process, the main question is whether this process settles on one of the elements of $\Theta^{\mathrm{q}}_{\mathsf{U}}$ or $\Theta^{\mathrm{c}}_{\mathsf{U}}$. As discussed above, $\theta_{\mathrm{KL}} = (\mu_{\mathrm{KL}}, \sigma^2_{\mathrm{KL}})$ is the model in $\mathcal{M}$ that minimizes the KL divergence relative to P. Using the definition of D_{KL}, this amounts to minimizing $\mathbb{E}_P[-\log q_\theta]$, where q_θ denotes the density of Q_θ with respect to the Lebesque measure. Given that the entertained likelihoods in $\mathcal{M}$ are lognormals we obtain,

$$\mu_{\mathrm{KL}} = \mathbb{E}_P\left[\ln X\right], \qquad \sigma^2_{\mathrm{KL}} = \mathbb{E}_P\left[(\ln X)^2\right] - \mathbb{E}_P\left[\ln X\right]^2, \tag{10}$$

irrespective of the form of P.

Comparing the best response functions (7) and (8) with (10), one can immediately see a form of inconsistency: while the agent's policy function depends on the mean of X, Bayesian learning with lognormal likelihoods aims to match the mean of $\ln X$. In other words, with both utility functions, the vector m including the relevant moments is inconsistent with the set $\mathcal{M}$ that "tells" KL-divergence which moments to match.

Although the specific form of P is irrelevant for this conclusion, further insight can be gained by looking at a particular example with $P \notin \mathcal{M}$. To this end, suppose that P is such that $\ln X$ is distributed as a two-component mixture normal distribution,

$$\ln X \overset{\mathrm{iid}}{\sim} \lambda \cdot \mathcal{N}(\mu_1, \sigma_1^2) + (1-\lambda) \cdot \mathcal{N}(\mu_2, \sigma_2^2), \tag{11}$$

hence the agent's long-run behavior is determined by the likelihood $Q_{\theta_{\mathrm{KL}}} \in \mathcal{M}$ with:

$$\mu_{\mathrm{KL}} = \lambda\mu_1 + (1-\lambda)\mu_2, \qquad \sigma^2_{\mathrm{KL}} = \lambda(\sigma^2_1 + \mu^2_1) + (1-\lambda)(\sigma^2_2 + \mu^2_2) - \mu^2_{\mathrm{KL}}.$$

Notice that for a general quintuple $(\lambda, \mu_1, \sigma^2_1, \mu_2, \sigma^2_2)$, $m_{\mathrm{i}}\left(Q_{\theta_{\mathrm{KL}}}\right) \neq m_{\mathrm{i}}(P)$ for $i \in \{\mathrm{q}, \mathrm{c}\}$, so the asymptotically implemented policy function is suboptimal relative to those induced by the likelihoods in $\Theta^{\mathrm{q}}_{\mathsf{U}}$ or $\Theta^{\mathrm{c}}_{\mathsf{U}}$. The extent to which this suboptimality matters depends on the characteristics of the utility function, here captured by the risk aversion parameter, η, and the agent's environment captured by the interest rate r.

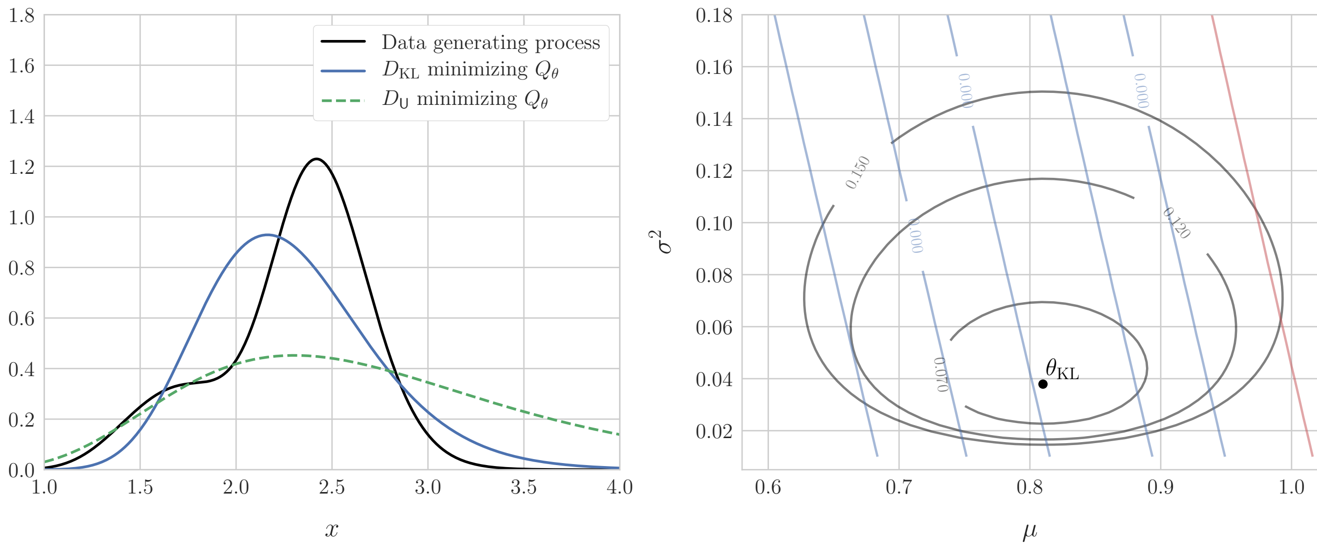

Figure 1: Left: Densities of the data generating process and best approximations within $\mathcal{M}$ according to D_{KL} and D_{U}. Right: Indifference curves with CARA preferences and the information geometry over $\mathcal{M}$. Note: The parameters are $(\lambda, \mu_1, \sigma_1, \mu_2, \sigma_2) = (0.3, 0.6, 0.2, 0.9, 0.1)$, $\beta = 0.98$, and $\eta = 2.5$. This implies that under P the mean of X is 2.29 and the standard deviation is 0.41.

For simplicity, we focus here on CARA preferences, but the case of quadratic utility is very similar. The left panel of Figure 1 depicts densities of the data generating process along with two misspecified likelihoods that are closest to it according to D_{KL} and D_{U}. While the blue distribution, used by the learner for making decisions, matches statistical aspects of the data generating process better, the green dashed likelihood induces higher long-run average payoff given that it lies in $\Theta^{\mathrm{c}}_{\mathsf{U}}$. In fact, the value generated by the green likelihood equals to that implied by the black distribution P.

To shed more light on the mechanism underlying this suboptimality, the right panel of Figure 1 depicts level curves corresponding to the projections of D_{KL} and D_{U} on $\mathcal{M}$. Following the notion of "information geometry" based on the KL divergence we name the latter "utility geometry".[5] Distributions on the ellipses have equal KL divergence relative to P. The further an ellipsis from θ_{KL} the higher the corresponding KL divergence. On the other hand, the indifference curves of D_{U} exhibit a strikingly different geometry. The red line represents $\Theta^{\mathrm{c}}_{\mathsf{U}}$, i.e., likelihoods in $\mathcal{M}$ that induce a policy function that performs identically to a^*. The blue lines correspond to likelihoods that result in policy functions which have equal performance in terms of long-run average payoff. The further we are from the red line the worse the performance.

[5]This is just a crude analogy since D_{U} is not a divergence in general.

The difference between the two "geometries" emerges from the properties of D_{KL} and D_{U}. While the level curves of D_{U} are influenced by the preference parameters, (η, β), the iso-entropies of D_{KL} depend primarily on the learner's set of models $\mathcal{M}$. As we saw, log-normal distributions imply that Bayes learning focuses on the moments $\mathbb{E}_P\left[\,\ln X\,\right]$ and $\mathbb{E}_P\left[\,(\ln X)^2\,\right]$, so to some extent, learning with this specific set, $\mathcal{M}$, directs the agent's attention on irrelevant features of the data generating process.

An alternative $\mathcal{M}$

While the lognormal family discussed above looks like a plausible choice to describe labor income, admittedly, it is more common to use normal distributions as a "default prior". To see how $\mathcal{M}$ can affect our previous conclusions, assume that the agent's set contains likelihoods that describe X as if it was drawn *i.i.d.* from a normal distribution parameterized by $\theta = \left(\mu, \sigma^2\right)$,

$$\mathcal{M} = \left\{ \; \mathcal{N}\left(\mu, \sigma^2\right) \; : \; (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+ \; \right\}. \tag{12}$$

For simplicity, let P be the same as above. The main difference from the previous analysis lies in the way KL-divergence determines θ_{KL}. With Gaussian likelihoods, D_{KL} aims to match the following moments:

$$\mu_{\text{KL}} = \mathbb{E}_P\left[\,X\,\right], \qquad\qquad \sigma^2_{\text{KL}} = \mathbb{V}_P\left[\,X\,\right].$$

Interestingly, these targets are consistent with quadratic preferences, in the sense that $m_{\text{q}}(Q_{\theta_{\text{KL}}}) = m_{\text{q}}(P)$. In this case, learning is successful even if all likelihoods are misspecified, because KL-divergence focuses on a moment that is relevant to the agent's objective. In fact, by learning about both the mean and the variance of X, the agent is wasting resources: σ^2 is irrelevant to her long-run payoffs, so it appears inefficient to include models with different variances in $\mathcal{M}$.

The situation is different with CARA utility. In this case, the best policy a^* is a function of $\mathcal{L}(P)$ that depends on all higher order moments of P not just the mean and variance, so it cannot be guaranteed that $m_{\text{q}}\left(Q_{\theta_{\text{KL}}}\right) = m_{\text{q}}(P)$. In other words, while changing the set $\mathcal{M}$ to include Gaussian likelihoods enables the agent with quadratic utility to implement a^* in the long-run, the same does not hold if preferences are CARA.

4 Misspecification-proof learning

The central message of the previous examples is that if the statistical and utility relevant aspects of $\mathcal{M}$ are not aligned, learning with misspecified likelihoods can lead to the implementation of suboptimal policies. As demonstrated by the negative examples in section 3, the problem with the DM's set is that KL divergence matches features of the environment that are irrelevant to the $\mathcal{P}$-optimal policy function, a^*. In this sense, Bayesian learning is focusing on wrong features of the data generating process.

Inspired by the positive example in section 3, we show now that sometimes misspecification-proof Bayesian learning is feasible. As we saw before, Bayes rule designates KL-divergence as an implicit loss function for learning. While taking this loss function as given, we can choose $\mathcal{M}$ so that D_{KL} focuses on pertinent features of the environment. To this end, the selection of entertained likelihoods should be based on the payoff-relevant moments of the data generating process. This insight can be generalized to a broader context and serve as a guideline for specifying $\mathcal{M}$ in situations when misspecification is a major concern.

We give conditions on $\mathcal{M}$ such that the KL divergence minimizing model induces the $\mathcal{P}$-optimal policy function even if the entertained set of likelihoods is misspecified.

Assumption 1 (Moment-dependent policy function)**.**
For a given utility function, u, suppose that the $\mathcal{P}$-optimal policy function can be expressed as a function of finitely many moments of the data generating process. That is, for any $Q \in \mathcal{P}$, the best response policy function, $\hat{a} = b(Q)$, can be implicitly defined as,

$$G_u\left(\hat{a}, \mathbb{E}_Q\left[T_u(X)\right]\right) = \mathbf{0}, \tag{13}$$

where $T_u\colon \mathcal{X} \mapsto \mathbb{R}^d$ defines moments of Q, and $G_u\colon \mathcal{A} \times \mathbb{R}^d \mapsto \mathbb{R}^k$ specifies the dependence of the optimal policy function on these moments. Note that both of these functions may depend on the DM's preferences, u.

Assumption 1 implies that the data generating process does not enter in the optimal policy function other than through the moments, $m(P) = \mathbb{E}_P\left[T_u(X)\right]$. As a result, T_u defines the sufficient statistic that is needed to implement the $\mathcal{P}$-optimal policy function.

Claim 1 (Misspecification-proof model)**.**
If Assumption 1 is satisfied and the set of entertained likelihoods forms an exponential family[6] *with,*

$$\mathcal{M} = \left\{q_\theta(x) = h(x)\exp\left\{\theta \cdot T_u(x) - A(\theta)\right\} : \theta \in \Theta \subseteq \mathbb{R}^d\right\}, \tag{14}$$

for some $h\colon \mathcal{X} \mapsto \mathbb{R}_+$, where $A(\theta)$ is the cumulant function, Bayesian learning converges to the model inducing the $\mathcal{P}$-optimal policy function a^* irrespective of the actual data generating mechanism, $P \in \mathcal{P}$; that is, irrespective of misspecification.

The model being misspecification-proof follows from the properties of the exponential family.[7] The KL-divergence minimizing distribution within $\mathcal{M}$ is characterized by,

$$\mathbb{E}_{\theta_{\mathrm{KL}}}[T_u(X)] = \mathbb{E}_P[T_u(X)]. \tag{15}$$

As a result, the Bayesian learner implements the following policy function in the limit, $\hat{a}_{\mathrm{KL}} = b\left(Q_{\theta_{\mathrm{KL}}}\right)$, such that,

$$G_u\left(\hat{a}_{\mathrm{KL}}, \mathbb{E}_{\theta_{\mathrm{KL}}}[T_u(X)]\right) = \mathbf{0}. \tag{16}$$

Given equation (15) the KL-divergence minimizing likelihood exactly matches the payoff-relevant moments

[6] For ease of notation we define the exponential family through density functions. q_θ is the density of Q_θ with respect to the Lebesgue measure. We use the canonical parametrization.

[7] See any standard reference on mathematical statistics, e.g. Shao [2003].

of the data generating process, and by assumption 1 the $\mathcal{P}$-optimal policy function is implemented,

$$G_u\left(\hat{a}_{\mathrm{KL}}, \mathbb{E}_P[T_u(X)]\right) = \mathbf{0}. \tag{17}$$

The moral of claim 1 is as follows. Even if the environment described by the underlying data generating process is complex in the statistical sense—i.e. infinite dimensional—the relevant complexity of the learner's problem is defined through her objective reflecting the properties of u. If the purpose of learning is to aide decisions, the learner should select her likelihoods based on their ability to capture features of the environment that matter for making good decisions. In this sense, we treat the DM's set $\mathcal{M}$ as an instrument for making decisions rather than an indisputable part of preferences. Given that the environment's statistical complexity can seriously limit the ability of learning, we advocate for calibrating $\mathcal{M}$ not to the intricacies of the "true" environment but to features pertinent to making good decisions.

Example revisited

To see how to apply the above concepts, we revisit the income fluctuation problem of section 3. Recall that we identified a vector of payoff-relevant moments, $m_i(Q)$, for both the quadratic ($i = \mathrm{q}$) and CARA ($i = \mathrm{c}$) utility specifications. In terms of the introduced notation, these vectors can be rewritten as,

$$\begin{aligned} T_{\mathrm{q}}(x) &\doteq x & &\Rightarrow & m_{\mathrm{q}}(Q) &= \mathbb{E}_Q\left[T_{\mathrm{q}}(X)\right]; \\ T_{\mathrm{c}}(x) &\doteq \left(x,\ \exp\left(-\frac{r\eta}{1+r}x\right)\right)' & &\Rightarrow & m_{\mathrm{c}}(Q) &= \mathbb{E}_Q\left[T_{\mathrm{c}}(X)\right]. \end{aligned}$$

Evidently, in both cases, Assumption 1 is satisfied, so misspecification-proof Bayes learning is feasible. With quadratic utility the recommended set of (misspecified) likelihoods is,

$$\mathcal{M}_{\mathrm{q}} = \Big\{ \exp\left(\theta \cdot x - A_{\mathrm{q}}(\theta)\right) \ : \ \theta \in \Theta \subseteq \mathbb{R} \Big\},$$

which is consistent with our previous finding that Gaussian distributions with unknown mean and known variance can lead to the implementation of a^* in the long-run.

Regarding CARA utility, we specify the following two-parameter family:

$$\mathcal{M}_{\mathrm{c}} = \left\{ \exp\left(\theta_1 \cdot x + \theta_2 \cdot \exp\left(-\frac{\eta r}{1+r} \cdot x\right) - A_{\mathrm{c}}(\theta_1, \theta_2)\right) \ : \ (\theta_1, \theta_2) \in \Theta \subseteq \mathbb{R}^2 \right\},$$

that renders the implied information geometry aligned with the utility geometry.

5 Related literature

Two key features of our analysis are: (i) we take an outside observer's perspective by considering standard Bayesian decision making but assessing the entertained likelihoods according to their long-run implications under some "objective reality", and (ii) we allow the set of likelihoods (prior support) to be misspecified. To the best of our knowledge, this paper is the first attempt to investigate the long-run properties of Bayesian learning by combining both of these features.

While Bayesian learning occupies a prominent place in the economics literature, most papers focus on the case of correct specification. In their classic survey, Bray and Kreps [1987] argue that this benchmark is too "sterile" and call for models that "have in place some level of inconsistency with reality". Important examples of such models are Nyarko [1991] and Fudenberg et al. [2017].[8] Similar to us, these papers analyze Bayesian learning under the assumption that the decision maker's prior is misspecified. Nevertheless, instead of assessing the usefulness of these priors as we do, these papers focus on the non-trivial dynamics of beliefs that may arise when learning is "active".

The econometrics literature paid relatively more attention on the idea of misspecification. In his seminal paper, Berk [1966] showed that when a Bayesian is learning about a parameter from a series of exchangeable signals, asymptotically, her posterior concentrates on the parameter values for which the KL-divergence of the DGP with respect to the entertained likelihoods is minimal. More recently, Shalizi [2009] arrives at the same conclusion in a much more general setting. Following the frequentist tradition, White [1996] provides a thorough analysis of maximum likelihood techniques when the model is misspecified.[9] In this case, the KL-divergence minimizing parameter, θ_{KL}, is typically called the "pseudo-true parameter". By studying large sample properties of Bayesian inference about θ_{KL}, Müller [2013] shows that one can reduce the Bayes estimator's expected loss (under the DGP) by replacing the original posterior with an artificial normal posterior centered at θ_{KL} with the "sandwich" covariance matrix. Although similar, our approach is different in the sense that our decision maker is not interested in P or θ_{KL} *per se*. Statistical closeness is important for her to the extent that it helps to make better decisions.

As for point (i), an example is Blume et al. [2018] who use an "objective" welfare criterion—similar in spirit to ours—to rank alternative market structures in the presence of belief heterogeneity (without learning). In addition, measuring the implications of learning relative to the DGP is of a similar flavor to the question of survival in financial markets analyzed by Blume and Easley [2006]. We illustrate that a learner's value function can carry invaluable information about the relative usefulness of different likelihoods when the prior is misspecified. This echoes the literature on max-min expected utility that breaks a key feature of Bayesian decision making: the separation of inference and control.[10] This literature—in contrast to the Bayesian decision rule we use—alters the manner in which optimal policies are chosen: instead of trying to maximize (4) under a single distribution, a max-min decision maker seeks policies that work "well" (not necessarily optimally) under a whole set of reasonable distributions. Attempts to marry such behavior with learning can be found in Hansen and Sargent [2007], Klibanoff et al. [2009], and Epstein and Schneider [2007].

Our recommendation in section 4 also resembles the idea of *Gibbs posteriors* advocated by Jiang and Tanner [2008] and Bissiri et al. [2016]. Instead of trying to model the DGP directly, this approach starts with some statistics of interest, θ, accompanied with a corresponding loss function, $\ell(\theta, x)$, such that θ minimizes the expected $\ell(\theta, x)$ under the DGP. It then proposes to use $\exp(-\ell(\theta, x))$ as a likelihood for Bayesian inference. In our case, the statistics θ can be viewed as our vector of payoff-relevant moments, $m(Q)$. In this sense, the

[8] See also Esponda and Pouzo [2016].

[9] Ideas similar to the justification that we give in section 4 can be spotted in various chapters of White [1996].

[10] In the Bayesian model, optimal inference about P is independent of the utility function u. See Hansen and Sargent [2018].

main difference relative to our analysis is that we do not take these moments as given, but derive them from primitives (preferences and market structure) of an economic decision problem. Similar comments apply to the so called *"focused information criterion"* developed by Claeskens and Hjort [2003]. It is a model selection tool that evaluates candidate models based on their ability to efficiently estimate a particular parameter of interest, instead of comparing their overall fit.

6 Concluding remarks

This paper shows that in a setting where misspecification is a major concern, Bayesian learning with arbitrary likelihoods can lead to outcomes that appear irrational from an "objective" point of view. Importantly, we do not mean this as a critique of Bayesian decision making. Instead, our result is meant to shed light on the advantages of viewing the decision maker's likelihoods as instruments rather than part of her preferences. In a truly unknown environment, entertaining a set which is inconsistent with the agent's payoff function is "irrational", in the sense that the decision maker would feel regret and change her mind if we exposed her to the potential consequences of her behavior.[11] Assuming that beliefs are of the same nature as attitudes toward risk or the rate of time preference makes this irrationality an "unchallangeable axiom" of behavior that we find unreasonable. In our view, there are such things as "good" and "bad" beliefs, just like there are "good" and "bad" models. Having said that, it seems sensible to impose consistency among the agent's beliefs and preferences even if learning is correctly specified.

We see two avenues for future research. First, we assumed that the decision maker fully embraces her likelihoods and uses them to derive policy functions as if they were 'subjective beliefs'. In other words, our decision maker does not acknowledge her misspecification concerns explicitly. In contrast, a growing literature pioneered by Hansen and Sargent [2008] endows agents inside economic models with the same kind of misspecification concerns as we, econometricians, face. It would be interesting to see how our results change in the presence of ambiguity aversion, that is, when policy functions are chosen in a max-min fashion. Hansen and Marinacci [2016] discuss related results and challenges. Second, for the sake of clarity, we focused on single agent problems, but it seems feasible to extend some of our insights to more general market settings as well. We leave the analysis of general equilibrium with multiple agents, learning about a common exogenous stochastic process, for future research.

[11] This definition of irrationality is motivated by Gilboa and Schmeidler [2001] and Gilboa [2009].

References

Robert H. Berk. Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Statist.*, 37(1):51–58, 02 1966.

P. G. Bissiri, C. C. Holmes, and S. G. Walker. A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5):1103–1130, 2016.

Lawrence E. Blume and David Easley. If you're so smart, why aren't you rich? belief selection in complete and incomplete markets. *Econometrica*, 74(4):929–966, 2006.

Lawrence E. Blume, Timothy Cogley, David A. Easley, Thomas J. Sargent, and Viktor Tsyrennikov. A case for incomplete markets. *Journal of Economic Theory*, 178:191 – 221, 2018. ISSN 0022-0531.

Margaret Bray and David Kreps. Rational learning and rational expectations. In G. R. Feiwel, editor, *Arrow and the Ascent of Modern Economic Theory*, pages 597–625. Palgrave Macmillan, London, 01 1987. ISBN 978-1-349-07241-5.

Ricardo J. Caballero. Consumption puzzles and precautionary savings. *Journal of Monetary Economics*, 25 (1):113–136, January 1990.

Gerda Claeskens and Nils Lid Hjort. The focused information criterion. *Journal of the American Statistical Association*, 98(464):900–945, 2003.

Larry G. Epstein and Martin Schneider. Learning under ambiguity. *The Review of Economic Studies*, 74 (4):1275–1303, 2007.

Ignacio Esponda and Demian Pouzo. Berk–nash equilibrium: A framework for modeling agents with misspecified models. *Econometrica*, 84(3):1093–1130, 2016.

Drew Fudenberg, Gleb Romanyuk, and Philipp Strack. Active learning with a misspecified prior. *Theoretical Economics*, 12(3):1155–1189, 2017.

Itzhak Gilboa. *Theory of Decision under Uncertainty.* Econometric Society Monographs. Cambridge University Press, 2009. doi: 10.1017/CBO9780511840203.

Itzhak Gilboa and David Schmeidler. *A theory of case-based decisions.* Cambridge University Press, 2001.

Robert E. Hall. Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence. *Journal of Political Economy*, 86(6):971–987, 1978.

Lars Peter Hansen and Massimo Marinacci. Ambiguity aversion and model misspecification: An economic perspective. *Statist. Sci.*, 31(4):511–515, 11 2016.

Lars Peter Hansen and Thomas J. Sargent. Recursive robust estimation and control without commitment. *Journal of Economic Theory*, 136(1):1 – 27, 2007.

Lars Peter Hansen and Thomas J. Sargent. *Robustness.* Princeton University Press, 2008.

Lars Peter Hansen and Thomas J. Sargent. *Risk, Uncertainty, and Value*. Princeton, New Jersey: Princeton University Press., 2018.

Wenxin Jiang and Martin A. Tanner. Gibbs posterior for variable selection in high-dimensional classification and data mining. *Ann. Statist.*, 36(5):2207–2231, 10 2008.

Peter Klibanoff, Massimo Marinacci, and Sujoy Mukerji. Recursive smooth ambiguity preferences. *Journal of Economic Theory*, 144(3):930 – 976, 2009.

Ulrich K. Müller. Risk of bayesian inference in misspecified models, and the sandwich covariance matrix. *Econometrica*, 81(5):1805–1849, 2013.

Yaw Nyarko. Learning in mis-specified models and the possibility of cycles. *Journal of Economic Theory*, 55(2):416 – 427, 1991.

Cosma Rohilla Shalizi. Dynamics of Bayesian updating with dependent data and misspecified models. *Electronic Journal of Statistics*, 3:1039–1074, 2009.

Jun Shao. *Mathematical Statistics*. Springer-Verlag New York Inc, 2nd edition, 2003.

Abraham Wald. *Statistical Decision Functions*. Wiley: New York, 1950.

Halbert White. *Estimation, Inference and Specification Analysis*. Number 9780521574464 in Cambridge Books. Cambridge University Press, December 1996.