# When does ambiguity fade away?

Filippo Massari* Jonathan Newton

Bocconi University Kyoto University

October 20, 2019

**Abstract**

Is long-run ambiguity a possible outcome of the multiple prior Bayesian learning model? If the prior support is finite, long-run ambiguity is known to be a possible outcome only if the learning problem is misspecified (Marinacci and Massari, 2019). Conversely, here we show that, under natural assumptions, ambiguity fades away on most paths if the prior support is naturally rich.

*Keywords*: Ambiguity, learning, robust statistical decisions, misspecified learning.

*JEL Classification*: D81, D83, C11

## 1 Introduction

Researchers have considered the implications of ambiguity for many economic phenomena. Examples include trade (Kajii and Ui, 2006), portfolio selection (Garlappi et al., 2006), risk pricing (Augustin and Izhakian, 2019), savings behavior (Hansen et al., 1999), job search (Nishimura and Ozaki, 2004) and the possibility of speculative bubbles (Werner, 2019).[1] Given the salience of ambiguity in economic and financial research, it is natural to wonder about how persistent it is. In the current paper we focus on the multiple prior model of ambiguity and consider conditions under which ambiguity fades away in the long run as a consequence of learning.

When a Bayesian decision-maker's set of priors comprises a finite set of iid models that includes the true model, Marinacci (2002) shows that ambiguity fades away over time as the decision-maker learns the true model. Marinacci and Massari (2019) drop the iid assumption and allow the problem to be misspecified so that it is impossible for the decision-maker to learn the true model. Nevertheless, they can still provide tight conditions under which ambiguity

*Corresponding author. E-mail: *massari3141@gmail.com*.

[1]The reader is referred to the survey article by Gilboa and Marinacci (2016) for more examples.

fades away. However, many applications, including all those mentioned above, feature decision-makers with sets of priors that are not finite, but are instead compact sets with positive Lebesgue measure on some parameter space. It is this latter setup that we study in the current paper, showing that for the exponential family of models, ambiguity fades away on most sequences. This result extends the fundamental results of (Doob, 1949; Berk, 1966; Clarke and Barron, 1990) from the single prior to the multiple prior setting.

The conditions for our result are relatively weak. It applies to any sequence of observations such that a unique maximum likelihood estimate exists at any given date sufficiently far along the sequence. This holds, for example, if the prior support is convex, in which case the concavity of the log-likelihood function implies a unique maximum likelihood estimate. Over time, all the posteriors concentrate on a shrinking neighborhood of this estimate and ambiguity fades away. Notably, the result holds even if the maximum likelihood estimate does not converge to a limit: all priors eventually concentrate around the estimate, even if the estimate itself changes over time.

From an applied perspective, this result suggests that ambiguity should not be a concern for a decision-maker who makes a large number of decisions and whose payoff gets averaged over time. This is because the strong law of large numbers allows us to average payoffs irrespective of the priors. Of course, the impact of ambiguity fading away will differ across models. For example **(I)** Kajii and Ui (2006) give necessary and sufficient conditions under which trade can take place under ambiguity. Trade that does take place in these conditions will be unaffected by ambiguity fading away, but additional opportunities for trade may arise.[2] **(II)** Werner (2019) shows that speculative trading bubbles can arise when market participants have common but ambiguous beliefs. Consequently, if ambiguity fades away, then another explanation for long-run speculative trade is required. **(III)** Garlappi et al. (2006) consider mean-variance portfolio selection with an ambiguous parameter. If ambiguity fades away, then the model eventually returns to the classical mean-variance model (Markowitz, 1952; Sharpe, 1970).[3]

When does our result not apply? Firstly, it does not apply to situations in which the decision-maker needs to make an important one-off decision such as buying a house or health insurance. In this case, there is no long term learning as there is no long term. Another possibility is that ambiguity may persist for some exogenous reason. For example, there may be aspects

[2] In the model of Kajii and Ui (2006), trade between two players is possible if and only if their sets of priors do not overlap. It is easy to see that if their sets of priors do not overlap under ambiguity, then the players will differ in their beliefs after ambiguity has faded away. Conversely, even if their sets of priors overlap under ambiguity, it is possible that the players will differ in their beliefs after ambiguity has faded away.

[3] Garlappi et al. (2006); Hansen et al. (1999) belong to a special class of ambiguous models known as $\varepsilon$-*contamination* models (see, e.g. Berger, 2013), in which the set of priors consists of all models within some distance $\varepsilon$ of an estimated model. Such models satisfy our condition of a positive Lebesgue measure of models in the support of the decision-maker.

of the problem that cannot be learned or that the decision-maker refuses to learn for some reason. Clearly, ambiguity can then persist with respect to these aspects (see, e.g. Epstein and Schneider, 2007).

## 2 Probabilities

We consider a family of models $M = \{P_\theta : \theta \in \Theta\}$ parametrized by a positive Lebegue measure parameter set $\Theta \subset \mathbb{R}^n$, defined on a $\sigma$-algebra $\Sigma^\infty$ of subsets of $X^\infty$ with representative element $x^\infty = x_1, x_2, ...$, where $X^\infty := \times^\infty X$ is the infinite Cartesian product of a state space $X$ with representative element $x$ and $\sigma$-algebra $\Sigma$. With a slight abuse of notation, we use $P_\theta(x^t)$ to denote the probability that model $P_\theta$ attaches to the cylinder with base $x^t$ $(i.e., Cyl(x^t) := \{x_1, ..., x_t, X_{t+1}, X_{t+2}, ...\})$, as well as the likelihood that model $P_\theta$ attaches to the partial sequence $(x_1, ..., x_t)$. The prior information about the parameters is summarized by prior distributions $\mu \in \Delta\Theta$. The set of prior distributions is $\mathcal{C}$. For any prior distribution $\mu \in \mathcal{C}$ the joint distribution of the parameters and the observations is $P_\mu \in \Delta(\Theta \times X^\infty)$, defined by, for all sets $A \subseteq \Theta$ and all cylinders $x^t$,

$$P_\mu(A \times x^t) := \int_A P_\theta(x^t) d\mu.$$

We denote by $\mu(.|x^t) \in \Delta\Theta$ the usual posterior given the observations $x^t$,[4] while $P_\mu(.|x^t) \in \Delta(\Theta \times X)$ is the one-step-ahead predictive distribution of $x_{t+1}$, given observations $x^t$. By definition, for all $A \subseteq \Theta$ we have

$$P_\mu(A \times x_{t+1}|x^t) := \int_A P_\theta(x_{t+1}|x^t) d\mu(.|x^t) := \int_A P_\theta(x_{t+1}|x^t) \frac{P_\theta(x^t)d\mu}{\int_\Theta P_\theta(x^t)d\mu}.$$

## 3 Decisions

Let $C$ be the space of consequences on which the decision-maker has a bounded utility function $u : C \to \mathbb{R}$. We consider one-step-ahead acts, i.e., $\Sigma-$measurable maps $f : X \to C$ that associates a consequence to each observation in $X$. The decision criterion adopted by the decision-maker depends on the quality of his prior information. In evaluating an act when the decision-maker has multiple priors, the decision-maker has to consider in the first period, for each act $f$, the set

$$\left\{ \int_X u(f(x)) dP_\mu(x|\emptyset) : \mu \in \mathcal{C} \right\}.$$

[4]We rule out the possibility of observing an event which is impossible according to all models in $M$.

Subsequently, as the decision-maker incorporates past realizations using Bayes' rule, the decision-maker has to consider, for each act $f$, the set:

$$\left\{ \int_X u(f(x)) dP_\mu(x|x^t) : \mu \in \mathcal{C} \right\}.$$

Possible summaries of this set are the infimum and supremum:

$$\sup_{\mu \in \mathcal{C}} \int_X u(f(x)) dP_\mu(x|x^t) \quad ; \quad \inf_{\mu \in \mathcal{C}} \int_X u(f(x)) dP_\mu(x|x^t).$$

# 4 Long-run ambiguity

As in (Marinacci, 2002; Marinacci and Massari, 2019), we consider the difference between the decision-maker 's expected utility under the most advantageous prior and under the least advantageous prior in $\mathcal{C}$ to be a measure of the ambiguity that a decision-maker perceives in evaluating an act $f$. A tight sufficient condition for this difference to be zero is that the posteriors calculated from all priors in $\mathcal{C}$ eventually coincide. That is, ambiguity fades away if all priors eventually converge to the same (possibly incorrect) model.

**Definition 1.** *Ambiguity fades away at path $x^\infty \in X^\infty$ if,*

$$\lim_{t \to \infty} \left[ \sup_{\mu', \mu'' \in \mathcal{C}} \int_X \left| dP_{\mu''}(x|x^t) - dP_{\mu'}(x|x^t) \right| \right] = 0 \tag{1}$$

*where, for each $t > 0, x^t$ indicates the first $t$ realizations of path $x^\infty$.*

Notably, Definition 1 makes no reference to the true model, which in any practical learning situation is not known by the decision-maker .

# 5 Main result

In this section, we identify conditions that guarantee that ambiguity fades away in the long-run when $\Theta$ has positive Lebesgue measure. These regularity conditions are borrowed from Grünwald (2007) conditions for the BIC approximation (Schwarz (1978), Clarke and Barron (1990)), to which we add a compactness assumption on the set of priors $\mathcal{C}$ to ensure convergence.

**Definition 2.** *The learning problem is **regular** if*

***A1:*** *The set of priors, $\mathcal{C}$, is compact;*

***A2:*** *$\mathcal{M}$ is a member of the exponential family with finitely many parameters[5];*

[5] E.g., Multinomial, Gaussian, Poisson, with either iid or Markov (K) structure with finitely many lags.

***A3:*** *each prior in* $\mathcal{C}$ *is defined on the same subset* $\Theta_0$ *of* $\Theta$*. Further,*

- *the interior of* $\Theta_0$ *is nonempty;*
- *the closure of* $\Theta_0$ *is a compact subset of the interior of* $\Theta$*;*

***A4:*** *each prior in* $\mathcal{C}$ *is differentiable, strictly bounded above and away from 0;*

***A5:*** *the realized sequence belongs to* $\hat{S}(\mathcal{M}, \Theta_0)$*: the set of sequences such that for all large* $t$*, the maximum likelihood estimator* $\hat{\theta}(x_{1:t})$ *exists, is unique and is in* $\Theta_0$*.*

**Theorem 1.**
*If the learning problem is regular, ambiguity fades away on* $\hat{S}(\mathcal{M}, \Theta_0)$*.*

*Proof.* See Appendix. □

Theorem 1 is similar in spirit and generalizes Blackwell and Dubins (1962)'s and Kalai and Lehrer (1994)'s results on merging. It is more general because it holds in a larger set of sequences. Instead of postulating the existence of a true distribution and deriving almost sure results, we prove convergence to the same predictive distributions in all paths in which the sequence of maximum likelihood parameters belongs to a well-behaved subset of $\Theta$. Importantly, a regular learning problem needs not be well-specified. Ambiguity fades away on $\hat{S}(\mathcal{M}, \Theta_0)$, a set of sequences that depends on $\mathcal{M}$ and $\Theta$, not on the true model. $\hat{S}(\mathcal{M}, \Theta_0)$ is very large. If $\Theta_0$ is a connected set, $\hat{S}(\mathcal{M}, \Theta_0)$ contains every sequence in which $\hat{\theta}(x_{1:t})$ converges to a single point in $\Theta_0$ *and* all sequences in which $\hat{\theta}(x_{1:t})$ does not converge, but it remains in $\Theta_0$ for all $t$ large.

Furthermore, if the decision-maker is interested in learning parameters of the multinomial model and priors are defined on the whole simplex, ambiguity fades away in all sequences.

**Corollary 1.** *If the learning problem is regular,* $\mathcal{M}$ *is the multinomial family (with arbitrary but finite Markov correlation structure) and* $\Theta$ *is the full simplex (cartesian product of the simplex if not i.i.d.), ambiguity fades away in every sequence.*

*Proof.* If the maximum likelihood parameter is in the interior of $\Theta$ for most periods, Theorem 1 holds. If it converges to the boundary of $\Theta$, the technique of Xie and Barron (2000) guarantees convergence of the priors. □

For example, there are no sequences on which a Bayesian decision-maker who believes he is facing i.i.d. realizations from an ambiguous coin, and whose prior information of the bias of the coin is captured by a family of Beta priors with parameters in two strictly positive, compact intervals $[a, b], [c, d]$ on $(0, 1)$ (i.e., $\mathcal{C} = \{Beta(\alpha, \beta),\ \alpha \in [a, b], \beta \in [c, d]\}$) suffer long-run ambiguity.

# 6 Appendix

In this appendix, $h(x) = o(g(x))$, abbreviates $\lim_{x\to\infty} \frac{h(x)}{g(x)} = 0$;
$\hat{\theta}_t := \hat{\theta}(x_{1:t})$ is the maximum likelihood parameter on the partial history $x_{1:t}$;
$D\left(P(.|\hat{\theta}_t)||P(.|\theta)\right) := E_{P(.|\hat{\theta}_t)} \ln \frac{P(x|\hat{\theta}_t)}{P(x|\theta)}$ is the K-L divergence between $P(.|\hat{\theta}_t)$ and $P(.|\theta)$.

**Proof of Theorem 1**

*Proof.* $\mathcal{C}$ compact $\Rightarrow arg\max_{\mu\in\mathcal{C}} \int_X u(f(x))dP_\mu(.|x_{1:t})$ and $arg\min_{\mu\in\mathcal{C}} \int_X u(f(x))dP_\mu(.|x_{1:t})$ exist.

Thus, it suffices to show that if the learning problem is regular, then $\forall x_{1:\infty} \in \hat{S}(\mathcal{M},\Theta_0)$ and $\forall g,h \in \mathcal{C}$, $\lim_{t\to\infty} \int_X |dP_g(x|x_{1:t}) - dP_h(x|x_{1:t})| = 0$.

$$0 \le \lim_{t\to\infty} \int_X |dP_g(x|x_{1:t}) - dP_h(x|x_{1:t})| := \lim_{t\to\infty} \int_X \left| \frac{P_g(x_{1:t+1})}{P_g(x_{1:t})} - \frac{P_h(x_{1:t+1})}{P_h(x_{1:t})} \right| dx$$

$$\underbrace{=^a}_{\substack{\text{by} \\ \text{Lemma 1}}} \int_X \lim_{t\to\infty} \left| \int_{\Theta_0} P(x|\theta) \left( \frac{e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} g(\theta)}{\int_{\Theta_0} e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} g(\theta)\ d\theta} \frac{P(x_{1:t}|\hat{\theta}_t)}{P(x_{1:t}|\hat{\theta}_t)} + \right.\right.$$

$$\left.\left. - \frac{e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} h(\theta)}{\int_{\Theta_0} e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} h(\theta)\ d\theta} \frac{P(x_{1:t}|\hat{\theta}_t)}{P(x_{1:t}|\hat{\theta}_t)} \right) d\theta \right| dx$$

$$=^{b,c} \int_X \lim_{t\to\infty} \left| \int_{B_t} P(x|\theta) \left( \frac{e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} g(\theta)}{\int_{B_{t-1}} e^{-tD\left(P(.|\hat{\theta}_{t-1})||P(.|\theta)\right)} g(\theta)\ d\theta dx} - \frac{e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} h(\theta)}{\int_{B_{t-1}} e^{-tD\left(P(.|\hat{\theta}_{t-1})||P(.|\theta)\right)} h(\theta)\ d\theta} \right) d\theta \right| dx$$

$$\le^d \int_X \lim_{t\to\infty} P^+(x) \max \left\{ \left| L^+ \frac{\frac{\sqrt{2\pi} g_t^+}{\sqrt{t I_t^-}}}{\frac{\sqrt{2\pi} g_{t-1}^-}{\sqrt{(t-1) I_{t-1}^+}}} - L^- \frac{\frac{\sqrt{2\pi} h_t^-}{\sqrt{t I_t^+}}}{\frac{\sqrt{2\pi} h_{t-1}^+)}{\sqrt{(t-1) I_{t-1}^-}}} \right| ; \left| L^- \frac{\frac{\sqrt{2\pi} g_t^-}{\sqrt{t I_t^+}}}{\frac{\sqrt{2\pi} g_{t-1}^+}{\sqrt{(t-1) I_{t-1}^-}}} - L^+ \frac{\frac{\sqrt{2\pi} h_t^+}{\sqrt{t I_t^-}}}{\frac{\sqrt{2\pi} h_{t-1}^-}{\sqrt{(t-1) I_{t-1}^+}}} \right| \right\} dx$$

$$=^e 0.$$

a) We can exchange the order of limit and integration by the Lebesgue dominated convergence theorem.

b) $B_t$ is a neighbourhood of the maximum likelihood such that $diam(B_t) \to^{t\to\infty} 0$ at the appropriate rate.

c) By Lemma 2 **(i)**, $\int_{\Theta_0\setminus B_t} e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} h(\theta)\ d\theta \to 0$ exponentially fast and it can be ignored in the calculation the limit.

d) By Lemma 2 **(ii)**, with $P^+(x) = \sup_{\theta'\in B_{t-1}} P(x|\theta') < 1$, $L^+ = \frac{\sup_{\theta\in B_{t-1}} p(\sigma^{t-1}|\theta)}{\inf_{\theta\in B_{t-1}} p(\sigma^{t-1}|\theta)}$, $L^- = \frac{\inf_{\theta\in B_{t-1}} p(\sigma^{t-1}|\theta)}{\sup_{\theta\in B_{t-1}} p(\sigma^{t-1}|\theta)}$, $I_t^- = \inf_{\theta'\in B_t} I(\theta')$, $I_t^+ = \sup_{\theta'\in B_t} I(\theta')$, $g_t^- = \inf_{\theta'\in B_t} g(\theta')$, $g_t^+ = \sup_{\theta'\in B_t} g(\theta')$.

e) $g(.), h(.), I(.)$ and $L$ are differentiable strictly positive $\forall\theta \in B_t \subset \Theta_0$,[6]. By definition of $\{B_t\}_{t=1}^\infty$, $\sup_{\theta'\in B_t, \theta''\in B_{t-1}} ||\theta' - \theta''|| \to 0$, so that $L_t^+ - L_t^- \to 0$, $\frac{g_t^+}{g_t^-} \to 1$, $\frac{I_t^+}{I_t^-} \to 1$ $\frac{t}{t-1} \to 1$ uniformly and the limit follows $\forall x_{1:\infty} \in \hat{S}(\mathcal{M},\Theta_0)$.

□

[6] $g(.)$ and $h(.)$, by **A4**; $I(.)$ and $L$ because $\mathcal{M}$ is a member of the exponential family (**A2**).

**Lemma 1.** *Let $\mathcal{M}$ be a member of the exponential family parameterized by $\Theta_0$ and $g$ a function that satisfies **A3-A4**, then:*

$$\ln \int_{\Theta_0} P(x_{1:t}|\theta)g(\theta)\ d\theta = \ln \int_{\Theta_0} e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} g(\theta)\ d\theta + \ln P(x_{1:t}|\hat{\theta}_t)$$

*Proof.* Standard result (see, e..g., Grünwald, 2007, Chapter 8). □

**Lemma 2.** *Let $\mathcal{M}$ be a member of the exponential family parametrized by $\Theta_0$ and $g$ a function that satisfies **A3-A4**, then, $\forall x_{1:\infty} \in \hat{S}(\mathcal{M}, \Theta_0)$,*

$$\int_{\Theta_0} e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} g(\theta)\ d\theta = \int_{\Theta_0 \setminus B_t} e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} g(\theta)\ d\theta + \int_{B_t} e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} g(\theta)\ d\theta,$$

*and the following bounds holds uniformly when $B_t$ is a neighbourhood of the maximum likelihood such that $diam(B_t) \to^{t\to\infty} 0$ at the appropriate rate.*
***(i) First integral:*** $\exists k, a < \infty : \mathcal{I}_1 = \int_{\Theta_0 \setminus B_t} e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} g(\theta) d\theta < k e^{-at^{2\alpha}} \to 0.$
***(ii) Second integral:*** *Let $\mathcal{I}_2 = \int_{B_t} e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} g(\theta)\ d\theta, I(\theta_t)$ be the Fisher information evaluated at the maximum likelihood parameter.[7], $I_t^- = \inf_{\theta' \in B_t} I(\theta')$, $I_t^+ = \sup_{\theta' \in B_t} I(\theta')$, $g_t^- = \inf_{\theta' \in B_t} g(\theta')$, $g_t^+ = \sup_{\theta' \in B_t} g(\theta')$, then*

$$\frac{g_t^-}{\sqrt{tI_t^+ 2\pi}} \leq \mathcal{I}_2 \leq \frac{g_t^+}{\sqrt{tI_t^- 2\pi}}.$$

*Proof.* By additivity,

$$\int_{\Theta_0} e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} g(\theta)\ d\theta = \int_{\Theta_0 \setminus B_t} e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} g(\theta)\ d\theta + \int_{B_t} e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} g(\theta)\ d\theta.$$

For ease of exposition,[8] we focus on the case in which $\mathcal{M}$ is the Bernoulli family, so that $P(.|\theta) = \theta$ and $B_t$ can be chosen as follow: $B_t = \{\theta \in [\hat{\theta}_t - t^{-\frac{1}{2}+\alpha}, \hat{\theta}_t + t^{-\frac{1}{2}+\alpha}]\}$ with $0 < \alpha < \frac{1}{2}$. To gain intuition, take $\alpha$ very small, so that $B_t$ is a neighborhood of the maximum likelihood that shrinks to 0 at a rate slightly slower than $1/\sqrt{t}$. Because $x_\infty \in \hat{S}(\mathcal{M}, \Theta_0)$, $B_t$ concentrates around $\hat{\theta}_t$. Because $g$ is continuous and strictly positive in $\Theta_0$, there is a $T : \forall t > T, B_t \subset \Theta_0$ where $\Theta_0$ is a compact subset of $\Theta$ in which $g > 0$. We always assume $t > T$.

The proof is done by performing a second-order Taylor expansion of $D(P(.|\hat{\theta}_t))||P(.|\theta))$ to bound the two integrals. $\mathcal{M}$ is a member of the exponential family; thus , $D(p^{\hat{\theta}_t}||P)$ can be well approximated in $B_t$ as follows (see Grünwald, 2007, chapter 19):

$$D\left(P(.|\hat{\theta}_t)||P(.|\theta)\right) = \frac{1}{2}\left(\hat{\theta}_t - \theta\right)^2 I(\theta^*) \tag{2}$$

for some $\theta^* \in B_t$ such that $\theta^*$ lies between $\theta$ and $\hat{\theta}_t$.

**(i) First integral:** Because $D\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)$, as a function of $\theta$, is strictly convex, has a minimum at $\theta = \hat{\theta}_t$, and is increasing in $|\theta - \hat{\theta}_t|$, the following holds:

$$0 < \int_{\Theta_0 \setminus B_t} e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} g(\theta)\ d\theta < \int_{\Theta_0 \setminus B_t} e^{-t \min_{\theta \in \Theta_0 \setminus B_t} D\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} g(\theta)\ d\theta.$$

[7] Which is positive definite because $\mathcal{M}$ is a member of the exponential family.
[8] The result generalizes to other members of the exponential family straightforwardly.

By Equation 2 and the definition of $B_t$

$$\min_{\theta \in \Theta_0 \setminus B_t} D\left(P(.|\hat{\theta}_t)||P(.|\theta)\right) \geq \frac{1}{2} t^{-1+2\alpha} \min_{\theta \in int(\Theta_0)} I(\theta)$$

so that, since $I(\theta)$ is continuous and $> 0$ for all $\theta \in \Theta_0$, and $\int_{\Theta_0 \setminus B_t} g(\theta)\, d\theta < \infty$,

$$0 < \int_{\Theta_0 \setminus B_t} e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} g(\theta)\, d\theta < \int_{\Theta_0 \setminus B_t} e^{-t\left(\frac{1}{2}t^{-1+2\alpha} \min\limits_{\theta \in int(\Theta_0)} I(\theta)\right)} g(\theta)\, d\theta < ke^{-at^{2\alpha}},$$

for $a = \frac{1}{2} \min\limits_{\theta \in int(\Theta_0)} I(\theta) > 0$ and $k = \int_{\Theta_0 \setminus B_t} g(\theta)\, d\theta < \int_{\Theta_0} g(\theta)\, d\theta < \infty$.

**(ii) Second integral:** by Equation 2,

$$\mathcal{I}_2 = \int_{B_t} e^{-tD\left(P(.|\hat{\theta}_t)||P(.|\theta)\right)} g(\theta)\, d\theta = \int_{B_t} e^{-\frac{t}{2}(\hat{\theta}_t - \theta)^2 I(\theta')} g(\theta)\, d\theta$$

where $\theta'$ depends on $\theta$. Using the notation defined above, we get

$$g_t^- \int_{B_t} e^{-\frac{t}{2}(\hat{\theta}_t - \theta)^2 I_t^+} di \leq \mathcal{I}_2 \leq g_t^+ \int_{B_t} e^{-\frac{t}{2}(\hat{\theta}_t - \theta)^2 I_t^-} di.$$

Performing the substitutions $z = (\hat{\theta}_t - \theta)\sqrt{tI_t^+}$ on the left integral and $z = (\hat{\theta}_t - \theta)\sqrt{tI_t^-}$ on the right integral, we get

$$\frac{g_t^-}{\sqrt{tI_t^+}} \int_{|z| < t^\alpha \sqrt{I_t^-}} e^{-\frac{1}{2}z^2} dz \leq \mathcal{I}_2 \leq \frac{g_t^+}{\sqrt{tI_t^-}} \int_{|z| < t^\alpha \sqrt{I_t^+}} e^{-\frac{1}{2}z^2} dz,$$

and recognize these integrals as standard Gaussian.
Because, as $t \to \infty$, $I_t^- \to I(\hat{\theta}_t)$ and $I_t^+ \to I(\hat{\theta}_t)$, the domain of integration tends to infinity for both integrals, so that they both converge to $\sqrt{2\pi}$. This approximation holds uniformly for all $x_{1:\infty} \in \hat{S}(\mathcal{M}, \Theta_0)$ because $i)$ the bound on $\mathcal{I}_1$ does not depend on $x_{1:t}$, and $ii)$ convergence of $\mathcal{I}_2$ is uniform because $g(\theta)$ and $I(\theta)$ are differentiable, positive functions of $\theta$ over the compact closure of $\Theta_0$. □

# References


Augustin, P. and Izhakian, Y. Y. (2019). Ambiguity, volatility, and credit risk. *Review of Financial Studies (forthcoming).*

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis.* Springer Science & Business Media.

Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *The Annals of Mathematical Statistics*, 37(1):51–58.

Blackwell, D. and Dubins, L. (1962). Merging of opinions with increasing information. *The Annals of Mathematical Statistics*, 33(3):882–886.

Clarke, B. S. and Barron, A. R. (1990). Information-theoretic asymptotics of bayes methods. *Information Theory, IEEE Transactions on*, 36(3):453–471.

Doob, J. L. (1949). *Application of the theory of martingales.* Colloques Internationaux du Centre National de la Recherche Scientifique Paris.

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

Garlappi, L., Uppal, R., and Wang, T. (2006). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies*, 20(1):41–81.

Gilboa, I. and Marinacci, M. (2016). Ambiguity and the bayesian paradigm. In *Readings in formal epistemology*, pages 385–439. Springer.

Grünwald, P. D. (2007). *The minimum description length principle*. MIT press.

Hansen, L. P., Sargent, T. J., and Tallarini Jr, T. D. (1999). Robust permanent income and pricing. *Review of Economic studies*, pages 873–907.

Kajii, A. and Ui, T. (2006). Agreeable bets with multiple priors. *Journal of Economic Theory*, 128(1):299–305.

Kalai, E. and Lehrer, E. (1994). Weak and strong merging of opinions. *Journal of Mathematical Economics*, 23(1):73–86.

Marinacci, M. (2002). Learning from ambiguous urns. *Statistical Papers*, 43(1):143–151.

Marinacci, M. and Massari, F. (2019). Learning from ambiguous and misspecified models. *Journal of Mathematical Economics*, 84:144–149.

Markowitz, H. (1952). Portfolio selection. *The journal of finance*, 7(1):77–91.

Nishimura, K. G. and Ozaki, H. (2004). Search and knightian uncertainty. *Journal of Economic Theory*, 119(2):299–333.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Sharpe, W. F. (1970). *Portfolio theory and capital markets*, volume 217. McGraw-Hill New York.

Werner, J. (2019). Speculative trade under ambiguity. Technical report, mimeo.

Xie, Q. and Barron, A. R. (2000). Asymptotic minimax regret for data compression, gambling, and prediction. *IEEE Transactions on Information Theory*, 46(2):431–445.