Learning from ambiguous urns*

Massimo Marinacci

Dipartimento di Statistica e Matematica Applicata, Università di Torino,
Piazza Arbarello 8, 10122 Torino, Italy
(e-mail: massimo@econ.unito.it)

Received: November 1999; revised version: July 2001

Abstract We provide conditions under which ambiguity fades away in sampling with replacement from the same “ambiguous” urn.

1 Introduction and Setting

Consider a decision maker (DM) who has to make a decision based on the drawings of an urn of known composition. The confidence he has in his decisions will depend on the quality of the information available on the balls’ proportions, the more he knows, the more he will feel confident.

Suppose the DM can sample with replacement from this urn before making a decision. Regardless of how poor is his a priori information about the balls’ proportions, it is natural to expect that eventually, as the number of observations increases, he will become closer and closer to learn the true balls’ proportion and become more and more confident in his decisions.

The purpose of this note is to prove that this intuition is correct. It is mostly a methodological contribution on the limiting behavior of this type of decisions, mathematically all results are simple. Marinacci (1999), a related – though mathematically more complicated – paper on limit laws for non-additive probabilities, provides a discussion of learning under ambiguity and considers cases where learning might not occur over time in the presence of ambiguity. We refer the interested reader to that paper for details.

* We wish to thank Paolo Ghirardato, Marco Scarsini, an anonymous referee, and especially Larry Epstein for helpful comments.

The financial support of MIUR is gratefully acknowledged.
We consider a predictive parametric setting,¹ where $\Theta = \{\theta_0, \ldots, \theta_M\}$ is a finite set of parameters and $\{P_\theta : \theta \in \Theta\}$ a family of countably additive probability measures – the likelihood functions – defined on a $\sigma$-algebra $\Sigma$ of subsets of an observation space $\Omega$. Let $P_\theta^\infty$ be the infinite product measure on $(\Omega^\infty, \Sigma^\infty)$ that makes the coordinate random variables independent with common distribution $P_\theta$. The a priori information on the parameters is summarized by one or more prior additive probabilities $\mu : 2^\Theta \to [0, 1]$.

Let $\mathcal{X}$ be a space of consequences (endowed with its power set), on which decision makers (DMs) have bounded utility functions $u : \mathcal{X} \to \mathbb{R}$. An act $f : \Omega \to \mathcal{X}$ is a $\Sigma$-measurable map that associates a consequence to each observation in $\Omega$. As it will become clear later, we are considering one-step-ahead acts (see Section 1.2).

In this model, the DMs know that the process is governed by a probability in the family $\{P_\theta : \theta \in \Theta\}$, but they may not know which is the right one. Depending on the DMs’ a priori information, we can distinguish three different scenarios, that we illustrate using the following simple bet. Suppose there is an urn containing 100 balls, each of which is either white or red. Suppose that the DM chooses a color, and draws a ball from the urn. If the ball that is drawn matches the DM’s selected color, the DM gets $100. Otherwise the DM gets nothing.

The consequence space is $\mathcal{X} = \{0, 100\}$, while the observation space $\Omega$ consists of two elements: $x_R = \text{"A ball drawn turns out to be red,"}$ and $x_W = \text{"A ball drawn is white."}$ The two possible acts are: $f_r = \text{"Choose color } R,\text{"}$ and $f_w = \text{"Choose color } W,\text{"}$ They are formally described in the following table:

<table>
<thead>
<tr>
<th></th>
<th>$x_R$</th>
<th>$x_W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_r$</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>$f_w$</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

Finally, $\theta$ is the number of, say, white balls, so that $\Theta = \{0, 1, \ldots, 100\}$ and $P_\theta(x_w) = \theta/100$ for each $\theta \in \Theta$.

1.1 Scenario 1: Risky Urns

Suppose the DM knows the exact proportion of white and black balls, say 50 white and 50 black. In terms of the model, this means that the DM knows the true parameter $\theta_0$ – in this case $\theta_0 = 50$. The prior probability $\mu$ is degenerated on $\theta_0$, i.e., $\mu = \delta_{\theta_0}$.²

This is a situation of risk, and we assume that the DM’s choice criterion is:

$$\int_{\Omega} u(f(x)) \, dP_{\theta_0},$$

for each act $f$.

---

¹ See, e.g., Bernardo and Smith (1994) and Fortini, Ladelli, and Regazzini (2000).

² As usual, $\delta_\theta$ is the Dirac measure on $\theta \in \Theta$, i.e., $\delta_\theta(A) = 1$ if and only if $\theta \in A$ for all $A \subseteq \Theta$. 
1.2 Scenario 2: Unambiguously Uncertain Urns

Suppose the DM does not know the proportion of white and black balls, but he has enough a priori information to form a sharply defined unique prior \( \mu : 2^\Theta \to [0, 1] \). For instance, the DM may consider all parameters as equally likely, and so \( \mu(\theta) = 1/|\Theta| \). Or, in contrast, the DM may think that \( \theta = 50 \) is the more likely parameter and have \( \mu(\theta) \) proportional to \( (\theta/100)(1 - \theta/100) \) for each \( \theta \in \Theta \).

This is a situation of unambiguous uncertainty, and we assume that the DM’s choice criterion before making any observation is:

\[
\int_\Theta \left[ \int_\Omega u(f(x)) \, dP_\theta \right] \, d\mu, \quad (2)
\]

for each act \( f \).

Unlike the risky case, now the DM may want to gather some observations in order to learn some information about the parameter and then update his choice criterion in light of the improved information. For example, in the betting example the DM may have gathered some observations about previous drawings (with replacement) from the same urn before having to bet on the next drawing.

For any prior probability \( \mu \), let \( P_\mu : 2^\Theta \otimes \Sigma^\infty \to [0, 1] \) be the joint distribution of the parameter and the observations defined by

\[
P_\mu(A \times B) = \int_A P_\theta^\infty(B) \, d\mu
\]

for all \( A \subseteq \Theta \) and \( B \in \Sigma^\infty \). We denote by \( \mu(\cdot \mid X_1, \ldots, X_n) : 2^\Theta \to [0, 1] \) the usual posterior given the observations \( X_1, \ldots, X_n \), while \( P_{\mu,n} : \Sigma \to [0, 1] \) is the one-step-ahead predictive distribution given the observations \( X_1, \ldots, X_n \) defined by

\[
P_{\mu,n}(B) = \int_\Theta P_\theta(B) \, d\mu(\cdot \mid X_1, \ldots, X_n),
\]

for each \( B \in \Sigma \).

After some observations have been made, the choice criterion (2) becomes:

\[
\int_\Theta \left[ \int_\Omega u(f(x)) \, dP_\theta \right] \, d\mu(\cdot \mid X_1, \ldots, X_n), \quad (3)
\]

for each act \( f \). Equivalently, we can write Eq. (3) in terms of the predictive distribution as:

\[
\int_\Omega u(f(x)) \, dP_{\mu,n}. \quad (4)
\]

As Eqs. (3) and (4) clarify, we are considering one-step-ahead acts, whose payoffs only depend on the observation that the DM is about to collect in the next drawing. Even though this is a class of somewhat simple-minded acts, we expect that our results hold for more sophisticated acts whose payoffs may depend on more general streams of future observations than
just the one-step-ahead. Such acts, as well as a general dynamic decision-theoretic setting, can be found described in the recent paper of Epstein and Schneider (2001), to which we refer the interested reader (their purpose is to axiomatize the multiple priors model in a general dynamic setting).

1.3 Scenario 3: Ambiguous Urns

Suppose the DM neither knows the proportion of white and black balls nor has enough \textit{a priori} information to form a sharply defined unique prior $\mu : 2^\Theta \to [0, 1]$. But, suppose that there is a set $\mathcal{C}$ of prior probabilities compatible with his limited \textit{a priori} information. For example, when $\Theta \subseteq \mathbb{R}$, the DM may only know that $\theta \geq \alpha$ and $\theta \leq \beta$, and so

$$\mathcal{C} = \{\mu : \mu ([\alpha, \beta]) = 1\}.$$

In evaluating acts in this scenario, the DM has to take into account the following set of criteria:

$$\left\{ \int_\Theta \left( \int_\Omega u(f(x)) \, dP_\theta \right) \, d\mu : \mu \in \mathcal{C} \right\},$$

for each act $f$. A possible summary of this set of criteria is the \textit{infimum}, that is,\footnote{\textit{Infimum}.}

$$\inf_{\mu \in \mathcal{C}} \int_\Theta \left( \int_\Omega u(f(x)) \, dP_\theta \right) \, d\mu. \quad (5)$$

More generally, consider a function $\Gamma : 2^\mathbb{R} \to \mathbb{R}$ such that $\Gamma(A) \in cl(A)$ for all $A \subseteq \mathbb{R}$, which gives the following general criterion:

$$\Gamma \left( \left\{ \int_\Theta \left( \int_\Omega u(f(x)) \, dP_\theta \right) \, d\mu : \mu \in \mathcal{C} \right\} \right).$$

As in the unambiguous uncertainty case, the DM may also want to gather some observations from previous drawings with replacement from the same urn. In this case, we have to consider

$$\mathcal{C}(X_1, \ldots, X_n) = \{\mu(\cdot | X_1, \ldots, X_n) : \mu \in \mathcal{C}\},$$

the set of all posteriors obtained by updating the elements of $\mathcal{C}$. The choice criterion becomes:

$$\Gamma_n \left( \left\{ \int_\Theta \left( \int_\Omega u(f(x)) \, dP_\theta \right) \, d\mu : \mu \in \mathcal{C}(X_1, \ldots, X_n) \right\} \right). \quad (6)$$

Equivalently, in predictive terms we have\footnote{\textit{Equivalently, in predictive terms}.}

$$\Gamma_n \left( \left\{ \int_\Omega u(f(x)) \, dP_{\mu,n} : \mu \in \mathcal{C} \right\} \right). \quad (7)$$
We index \( \Gamma \) with \( n \) so that the DM may have different choice rules at different drawings, under the only requirement that \( \Gamma_n (A) \in \text{cl}(A) \) for all \( A \subseteq \mathbb{R} \) and all \( n \geq 1 \).

If, for example, the DM uses the "infimum rule" (5) at all drawings, then Eq. (7) becomes

\[
\inf_{\mu \in \mathcal{C}} \int_{\Omega} u(f(x)) \, dP_{\mu,n},
\]

which is the functional form often implicitly used when dealing with the multiple priors models axiomatized by Gilboa and Schmeidler (1989).

A natural measure of the ambiguity that the DM perceives in evaluating an act \( f \) is given by the difference

\[
\sup_{\mu \in \mathcal{C}(X_1, \ldots, X_n)} \int_{\Theta} \left[ \int_{\Omega} u(f(x)) \, dP_{\theta} \right] d\mu - \inf_{\mu \in \mathcal{C}(X_1, \ldots, X_n)} \int_{\Theta} \left[ \int_{\Omega} u(f(x)) \, dP_{\theta} \right] d\mu
= \sup_{\mu \in \mathcal{C}} \int_{\Omega} u(f(x)) \, dP_{\mu,n} - \inf_{\mu \in \mathcal{C}} \int_{\Omega} u(f(x)) \, dP_{\mu,n},
\]

which we denote by \( \psi(f) \). It is easy to check that we get back to unambiguous uncertainty if and only if \( \psi(f) = 0 \) for all acts \( f \).

2 Learning with Ambiguous Urns

In this paper we focus on the ambiguous case. In particular, consider an ambiguous urn and DMs that can bet on this urn after gathering some observations about previous drawings (with replacement) from the same urn. We want to find out conditions under which, as evidence builds up and posteriors are computed, the ambiguity perceived by the DMs disappears.

Let \( \mathcal{M} \) be the set of all probabilities defined on the power set of \( \Theta \). For any two subsets \( C_1 \) and \( C_2 \) of \( \mathcal{M} \), set:

\[
\rho(C_1, C_2) = \sup_{(\mu_1, \mu_2) \in C_1 \times C_2} \left( \max_{A \subseteq \Theta} |\mu_1(A) - \mu_2(A)| \right).
\]

**Definition 1** A sequence \( \{C_n\}_{n \geq 1} \subseteq \mathcal{M} \) converges uniformly to the singleton \( \{\delta_0\} \), denoted \( C_n \overset{L}{\to} \delta_0 \), if \( \rho(C_n, \{\delta_0\}) \to 0 \).

What we would like to prove is that if \( \theta \) is the true parameter, then the sequence \( C(X_1, \ldots, X_n) \) usually (i.e., \( P_\theta^\infty - a.e. \)) uniformly converges to the singleton \( \{\delta_0\} \). Basically, it is a Bayesian consistency problem with a twist, which is the need of uniform convergence of the posteriors in the set \( C(X_1, \ldots, X_n) \), a set that is typically uncountable. As Theorem 1 will make clear, for our purposes it is not enough that the posteriors converge one at a time, but they all have to converge at a similar rate. This is where our work differs from the literature on the consistency of Bayesian estimators (see, e.g., Diaconis and Freedman, 1986).
Definition 2 Given a pair \((C_0, \theta_0)\), with \(C_0 \subseteq \mathcal{M}\) and \(\theta_0 \in \Theta\), we say that ambiguity fades away at \((C_0, \theta_0)\) if
\[ C_0(X_1, \ldots, X_n) \overset{\mu}{\rightarrow} \delta_{\theta_0} \quad \text{in } P_{\theta_0}^\infty - \text{a.e.} \]

In words, ambiguity fades away at \((C_0, \theta_0)\) if there exists a set \(A \in \Sigma^\infty\) with \(P_{\theta_0}(A) = 1\) such that for all \(\varepsilon > 0\) and all \(x = (x_1, \ldots, x_n, \ldots) \in A\) there exists a positive integer \(N(x, \varepsilon)\) such that for all \(n \geq N(x, \varepsilon)\) and all \(\mu \in C_0\)
\[ \mu(\theta_0 \mid X_1 = x_1, \ldots, X_n = x_n) \geq 1 - \varepsilon \]
\[ \mu(\theta \mid X_1 = x_1, \ldots, X_n = x_n) \leq \varepsilon \quad \text{for all } \theta \neq \theta_0. \]

Given a set \(C \subseteq \mathcal{M}\), the (strong) support \(\text{supp} (C)\) of \(C\) is the set
\[ \{\theta \in \Theta : \mu(\theta) > 0 \text{ for all } \mu \in C\}. \]

We can now state a first convergence result.

Lemma 1 Let \(C\) be a compact subset of \(\mathcal{M}\). Then ambiguity fades away at \(\{C, \theta\}\) if and only if \(\theta \in \text{supp} (C)\).

In words, if one of the parameters in \(\text{supp} (C)\) is the true one, then the set of posteriors converges to it, while such a convergence cannot occur if the true parameter lies outside \(\text{supp} (C)\).

In light of Lemma 1, it is natural to ask when the set \(\text{supp} (C)\) is "large," for instance when \(\mu(\text{supp} (C)) = 1\) for all \(\mu \in C\).

Definition 3 We say that a set \(C \subseteq \mathcal{M}\) is consistent if any two probabilities \(\mu_1\) and \(\mu_2\) in \(C\) are mutually absolutely continuous, i.e., \(\mu_1(\theta) > 0 \iff \mu_2(\theta) > 0\) for all \(\theta \in \Theta\).

It is easy to check that \(\mu(\text{supp} (C)) = 1\) for all \(\mu \in C\) if and only if \(C\) is consistent. Therefore, the set \(C\) is large provided all prior probabilities in \(C\) agree on which sets are null.

Having established Lemma 1, we can turn our attention to our main object of interest, the limiting behavior of the set
\[ \left\{ \int_{\Theta} \left[ \int_{\Omega} u(f(x)) \, dP_{\theta} \right] \, d\mu : \mu \in C(X_1, \ldots, X_n) \right\} = \left\{ \int_{\Omega} u(f(x)) \, dP_{\mu, n} : \mu \in C \right\}. \]

We study this behavior using the functions \(\Gamma_n\) and \(\Psi\), introduced respectively in Eqs. (6) and (8).

Theorem 1 Let \(C\) be a compact subset of \(\mathcal{M}\) and suppose \(\theta_0 \in \text{supp} (C)\) is the true parameter. Then, for all acts \(f\), the following statements are true \(P_{\theta_0}^\infty\)-a.e.:

(i) \(\lim_{n \to \infty} \Psi(f) = 0\),

(iii) \( \lim_{n \to \infty} \Gamma_n \left( \left\{ \int_{\Omega} u(f(x)) \, dP_\theta \right\} \mu : \mu \in \mathcal{C}(X_1, ..., X_n) \right) \)
\[ = \int_{\Omega} u(f(x)) \, dP_\theta, \]
(iii) \( \lim_{n \to \infty} \Gamma_n \left( \left\{ \int_{\Omega} u(f(x)) \, dP_{\mu,n} : \mu \in \mathcal{C} \right\} \right) = \int_{\Omega} u(f(x)) \, dP_\theta. \)

In words, point (i) says that the ambiguity associated with each act disappears as \( n \) goes to infinity, while points (ii) and (iii) say in two equivalent ways that as observations build up, the DM is choosing more and more as if he knew the true parameter. Observations eventually "prevail", regardless of the size of the set \( \mathcal{C} \) and of the particular choice function \( \Gamma_n \) adopted at each drawing.

Theorem 1 therefore formalizes the simple intuition that eventually ambiguity disappears, at least when sampling with replacement from the same urn.

3 Proofs

3.1 Lemma 1

The "only if" part is obvious. For, if the true parameter \( \theta_0 \) is not in \( \text{supp}(\mathcal{C}) \), there is \( \tilde{\mu} \in \mathcal{C} \) such that \( \tilde{\mu}(\theta_0) = 0 \). Hence, \( \rho(\tilde{\mu}, \delta_{\theta_0}) = 1 \), and so \( \rho(\mathcal{C}, \{\delta_{\theta_0}\}) = 1 \).

We now prove the "if" part. Since \( \mathcal{C} \) is compact, \( \min_{\mu \in \mathcal{C}} \mu(\theta) \) exists for all \( \theta \in \Theta \). In particular, \( \min_{\mu \in \mathcal{C}} \mu(\theta) > 0 \) for all \( \theta \in \text{supp}(\mathcal{C}) \). Since \( \text{supp}(\mathcal{C}) \) is finite, there exists \( \epsilon > 0 \) small enough such that

\[ \frac{\epsilon}{|\text{supp}(\mathcal{C})|} \leq \min_{\mu \in \mathcal{C}} \mu(\theta) \text{ for all } \theta \in \text{supp}(\mathcal{C}). \]

Let \( m \) be the additive probability on \( \Theta \) such that
\[ m(\theta) = \begin{cases} \frac{1}{|\text{supp}(\mathcal{C})|} & \text{for all } \theta \in \text{supp}(\mathcal{C}) \smallsetminus \{\theta_0\}, \\ 0 & \text{else}. \end{cases} \]

Let \( \nu_\varepsilon \) be the monotone set function on \( \Theta \) such that \( \nu_\varepsilon(A) = \varepsilon m(A) \) for all \( A \subseteq \Theta \) and \( \nu_\varepsilon(\Theta) = 1 \). Then \( \nu_\varepsilon \) is a convex capacity, and its core \( \text{core}(\nu_\varepsilon) \) contains the set \( \mathcal{C} \). By a result of Shapley (1971), there exists a finite set \( \mathcal{E} \) of extreme points of the core of \( \nu \). Hence, for every \( \mu \in \mathcal{C} \) there exist two probabilities \( \mu_1 \) and \( \mu_2 \) in \( \mathcal{E} \) and a constant \( \alpha \in [0,1] \) such that \( \mu(A) = \alpha \mu_1(A) + (1-\alpha)\mu_2(A) \) for all \( A \subseteq \Theta \).

Since \( \nu_\varepsilon(\theta) > 0 \) if \( \theta \in \text{supp}(\mathcal{C}) \), by a well known result on Bayesian consistent estimators due to Doob (1948), for every \( \mu \in \text{core}(\nu_\varepsilon) \) and every \( A \subseteq \Theta \) it holds
\[ |\mu(A \mid X_1, ..., X_n) - \delta_\theta(A)| \to 0 \quad P_\theta^\infty \text{ - a.e.} \]
whenever \( \theta \in \text{supp}(\mathcal{C}) \). Let \( A_{\mu,\theta} \in S_{\infty}^\infty \) be the set where such a convergence occurs. Since \( \mathcal{E} \) is finite, and \( P_\theta^\infty(A_{\mu,\theta}) = 1 \) for all \( \mu \in \mathcal{E} \), it holds that
\[ P_\theta^\infty \left( \bigcap_{\mu \in \mathcal{E}} A_{\mu,\theta} \right) = 1. \]
Set $A_{\varepsilon, \delta} = \bigcap_{\mu \in \mathcal{E}} A_{\mu, \delta}$. Fix $\varepsilon > 0$, $\hat{\delta} \in \text{supp}(\mathcal{C})$, and $x = (x_1, \ldots, x_n, \ldots) \in A_{\varepsilon, \hat{\delta}}$. As $\mathcal{E}$ is finite, there exists a positive integer $N(\varepsilon)$ such that for all $\mu \in \mathcal{E}$ it holds

$$\rho \left( \mu \left( \cdot \mid X_1 = x_1, \ldots, X_n = x_n \right), \delta_{\hat{\theta}} \right) \leq \varepsilon \quad \text{for all } n \geq N(\varepsilon). \quad (10)$$

Let $\mu \in \mathcal{C}$. We know there exist a finite set $\{\mu_k\}_{k \in K}$ in $\mathcal{E}$ and coefficients $\{\alpha_k\}_{k \in K}$, with $\alpha_k \in [0, 1]$ and $\sum_{k=1}^{K} \alpha_k = 1$, such that for all $A \subseteq \Theta$ we can write:

$$\mu(A \mid X_1, \ldots, X_n) = \sum_{k=1}^{K} \alpha_k \left[ \mu_k(A \mid X_1, \ldots, X_n) - \delta_{\hat{\theta}}(A) \right]$$

By (10), for all $n \geq N(\varepsilon)$ we then have:

$$\rho \left( \mu \left( \cdot \mid X_1 = x_1, \ldots, X_n = x_n \right), \delta_{\hat{\theta}} \right) \leq \varepsilon \sum_{k=1}^{K} \alpha_k \frac{\int_{\Theta_0} P_{\hat{\theta}}(X_1, \ldots, X_n) d\mu_k}{\int_{\Theta_0} P_{\hat{\theta}}(X_1, \ldots, X_n) d\mu} = \varepsilon$$

for all $\mu \in \mathcal{C}$. Hence,

$$\rho \left( \mathcal{C}(X_1, \ldots, X_n), \delta_{\hat{\theta}} \right) \leq \varepsilon \quad \text{for all } n \geq N(\varepsilon).$$

By (9), since $x \in A_{\varepsilon, \hat{\delta}}$, we conclude that

$$\rho \left( \mathcal{C}(X_1, \ldots, X_n), \{\delta_{\hat{\theta}}\} \right) \to 0 \quad P_{\hat{\theta}}^{\infty} - \text{a.e.},$$

as wanted. □

### 3.2 Theorem 1

Let $h : \Theta \to \mathbb{R}$ be a function defined by $h(\theta) = \int_{\Omega} u(f(x)) dP_{\theta}$ for all $\theta \in \Theta$. Since $u$ is bounded, without loss of generality, assume $0 \leq h(\theta) \leq M$. Given $\varepsilon > 0$, by Lemma 1 there exists $N(\varepsilon)$ such that for all $n \geq N(\varepsilon)$ and each $\mu \in \mathcal{C}(X_1, \ldots, X_n)$ it holds that

$$\left| \int_{\Theta} h(\theta) d\mu - \int_{\Theta} h(\theta) d\delta_{\theta_0} \right| \leq \int_{0}^{M} |\mu(\{h(\theta) \geq t\}) - \delta_{\theta_0}(\{h(\theta) \geq t\})| dt \leq M \rho \left( \mathcal{C}(X_1, \ldots, X_n), \{\delta_{\theta_0}\} \right) \leq M \varepsilon,$$
so that
\[
\inf_{\mu \in \mathcal{C}(x_1, \ldots, x_n)} \int_{\Theta} \left[ \int_{\Omega} u(f(x)) \, dP_{\theta} \right] - \int_{\Omega} u(f(x)) \, dP_{\theta_0} \, d\mu
\]
\[
= \inf_{\mu \in \mathcal{C}(x_1, \ldots, x_n)} \int h(\theta) \, d\mu - \int h(\theta) \, d\delta_{\theta_0} \leq M\varepsilon.
\]
A similar argument holds for the sup, and so point (i) follows easily. As to (ii), it is enough to notice that, for each \( n \geq 1 \),
\[
\inf_{\mu \in \mathcal{C}(x_1, \ldots, x_n)} \int_{\Theta} \left[ \int_{\Omega} u(f(x)) \, dP_{\theta} \right]
\leq \Gamma_n \left( \left\{ \int_{\Theta} \left[ \int_{\Omega} u(f(x)) \, dP_{\theta_0} \right] \, d\mu : \mu \in \mathcal{C}(X_1, \ldots, X_n) \right\} \right)
\leq \sup_{\mu \in \mathcal{C}(x_1, \ldots, x_n)} \int_{\Theta} \left[ \int_{\Omega} u(f(x)) \, dP_{\theta} \right].
\]
This completes the proof of the theorem since (ii) and (iii) are clearly equivalent. \( \square \)

References