Bayesian Inference and Non-Bayesian Prediction and Choice: Foundations and an Application to Entry Games with Multiple Equilibria

Larry G. Epstein   Kyoungwon Seo

A regulator must choose a policy for a number of markets. The consequences of the policy depend on firm behavior. The regulator has data on firm behavior in related markets. She wants to learn from these data and then choose a policy

We want to model her inference and choice

First we need to describe her theory of the environment

It is similar to that assumed in the literature on partial identification & (entry) games: Bresnahan-Reis (1991), Berry (1992), Tamer (2003), Ardillas-Lopez-Tamer (2008), Ciliberto-Tamer (2009), Bajari-Hong-Ryan (2010)

ENTRY GAMES (Jovanovic 1989)

- Two firms $j = 1, 2$; many markets/experiments $i = 1, 2, ...$

	out	in
out	$0, 0$	$0, -\epsilon_{i2}$
in	$-\epsilon_{i1}, 0$	$\eta^{1/2} - \epsilon_{i1}, \eta^{1/2} - \epsilon_{i2}$

- Structural parameter: $\eta \in (0, 1]$
  $\epsilon_{i,j}$ realizations are known to players but unknown to analyst
  For each i, $(\epsilon_{i1}, \epsilon_{i2})$ is uniformly distributed on $[0, 1]^2$, iid across markets

- Policy-maker: observe some outcomes, estimate η and then choose policy

$\{(0,0),(1,1)\}$	if $0 \leq \epsilon_{i1}, \epsilon_{i2} \leq \eta^{1/2}$
$\{(0,0)\}$	otherwise

Theory: **pure strategy** Nash equilibrium (no theory of selection)
$S = \{(0,0),(1,1)\}$ the set of possible outcomes in each market

Two consequences of multiplicity:

- Inference: Complicated because a given sample can be interpreted in different ways. May not be able to identify η even with infinite data

- Prediction and choice: Ignorance of the selection mechanism implies a set of likelihoods even given knowledge of the parameter η
  - probability interval $\Pr((1,1)) \in [0, \eta]$

Noteworthy Features (typical in the entry game literature)

1. Markets are ex ante indistinguishable, but may differ (selection)
- can generalize to allow markets to have different observable characteristics

2. Two kinds of uncertainty: risk (ϵ's) and ambiguity (selection)

3. Two kinds of factors driving experiments: (i) common factors/parameters (here η) and (ii) idiosyncratic factors that vary across experiments in way that is <u>not understood</u> (selection)
Latter suggests *heterogeneity and correlation* of an *unknown* form
and also difficulty of learning about selection - *as though sampling from a sequence of different Ellsberg urns*

Seeking to make choices that are robust to these known unknowns

We describe a unified axiomatic model of choice and inference for a decision-maker with above concerns

due to having an *incomplete theory* of her environment

Modeling the policy-maker rather than the econometrician per se

Choice drives inference. Preference is primitive

We generalize the exchangeable Bayesian model
(Savage (1954), Anscombe-Aumann (1963) and de Finetti (1937)) of
decision-making under uncertainty in a setting with *repeated experiments*

Inference: justification for Bayesian methods - Moon & Schorfheide (2012)

Choice: based on *belief function utility*, a special case of both
CEU (Schmeidler, 1989) and MEU (Gilboa-Schmeidler, 1989))

Partial Identification & Entry Games

Above entry game is representative of a range of models used in applied IO, where take seriously that economic theories are typically **incomplete** and that auxiliary assumptions made for convenience should be avoided (Manski, Tamer)

What does this literature on partial identification provide?

▷Estimation & inference - frequentist approach: focus on asymptotics

Our presumption: *One of the main motivations for empirical work in economics is to evaluate policies.* One important purpose of this is making decisions ... (Tamer, 2009)

How does frequentist literature feed into choice?
With infinite data, have 'identified set' and might use multiple-priors
With finite samples?? Arguably inference and choice are not separable and choice drives inference

$\triangleright$Bayesian statistical methods: Moon-Schorfheide (2010), Liao-Jiang (2010)
Bayesian inference, but what about choice? SEU cannot capture noted concerns

$\triangleright$Choice treated by Manski (2011, for example) and Kasy (2011)
Not axiomatic, and much different models

Statistical decisions modeled by Stoye (2012), Kitagawa (2012), Menzel (2011)
- 'robust inference' and connections drawn to Gilboa-Schmeidler
- payoffs depend on true parameter and not on outcomes of experiments (unlike in every economic choice problem)
- reduced form? we focus on the underlying economic decision problem

A ‘normative’ model

We view our model as offering a *prescription* or *recommendation* for choice and inference, based on clearly stated general principles that a policy-maker would be able to accept or reject. Thus normative, with understanding that:

- Prescriptive while respecting DM’s limitations

- STP, Independence Axiom and probabilistic sophistication are not always compelling (Gilboa, Postelwaite & Schmeidler, 2012).
  Ellsberg behavior is a normative critique of probabilistic sophistication

- There are conflicting principles: one can’t have everything

Not everyone would buy it, but the alternatives are not great

OUTLINE

1. The exchangeable Bayesian model (de Finetti)

2. Belief functions

3. Foundations

4. Representation: exchangeable belief function utility

5. Entry game again

PRIMITIVES

- $\Omega = S^{\infty} = S_1 \times ... \times S_i \times ...$, ($S$ finite)
  In our case, the ordering of markets is arbitrary, not temporal

- $\mathcal{F}$ is set of all (simple) acts, $f : \Omega \longrightarrow [0, 1]$

- Given preference (binary relation) $\succeq$ on $\mathcal{F}$

- Outcomes are utils (as though risk neutral) & in *probability units*

  $f(\omega) = p$ means $f(\omega) \sim (\overline{c}, p; \underline{c}, 1 - p)$
  think of "$f(\omega) = u(c(\omega))$", where $u(\overline{c}) = 1$ and $u(\underline{c}) = 0$

The Benchmark - Exchangeable SEU

CHOICE is SEU (Savage/Anscombe-Aumann/de Finetti):
$\succeq$ has utility function $U : \mathcal{F} \longrightarrow \boldsymbol{R}$,

$$U(f) = \int_{\Omega} f(\omega)\, dP,$$

where the Bayesian (predictive) prior $P \in \Delta(\Omega)$ is *exchangeable*, that is,

$$P(\cdot) = \int_{\Delta(S)} \ell^{\infty}(\cdot)\, d\mu(\ell)$$

INFERENCE is by application of Bayes' Rule (Dynamic Consistency)

BUT: Imposes certainty that experiments are conditionally identical and independent; in entry game, iid selection probability

Made behavioral below. We mean "Bayesian behaves as if indifferent to uncertainty about heterogeneity and correlation"

We try to retain the attractive elements of the Bayesian model
- simple intuitive axioms, a simple representation and updating rule -
while accommodating a concern with differences between experiments
(poorly understood selection)

BELIEF FUNCTIONS

X (compact metric); for example, $X = S$ or $X = S^{\infty}$

$\nu : \Sigma_X \to [0,1]$, any non-additive set function constructed from $\left(\widehat{X}, m, \Gamma\right)$:

$$(\widehat{X}, m) \overset{\Gamma}{\rightsquigarrow} (X, \nu) \quad \overset{f}{\longrightarrow} \; Z$$

$$\nu(A) \equiv m\left(\left\{\widehat{x} \in \widehat{X} : \Gamma(\widehat{x}) \subset A\right\}\right)$$

Set of (predictive) priors:

$$\begin{aligned} core(\nu) &\equiv \{P \in \Delta(X) : P(\cdot) \geq \nu(\cdot)\} \\ &= \int_{\widehat{X}} \Delta(\Gamma(\widehat{x}))\, dm(\widehat{x})\} \end{aligned}$$

FACT: For a binary state space $\{B, N\}$, every belief function ν corresponds to a probability interval $[\nu(B), 1-\nu(N)]$, and conversely

BELIEF FUNCTION UTILITY: For any state space X,

$$\begin{aligned} U(f) &= U_\nu(f) = \int_X f d\nu \quad \text{(Choquet integral)} \\ &= \min_{P \in core(\nu)} \int_X f dP \quad \text{(multiple-priors)} \end{aligned}$$

Why Belief Functions in the Entry Game?

Each market corresponds to a binary experiment: $S = \{B, N\}$

For each η, the *equilbrium correspondence* is

$$\Gamma_\eta : \{(\epsilon_{i1}, \epsilon_{i2})\} = [0,1]^2 \rightsquigarrow \{B, N\}$$

Γ_η and uniform distribution induce a belief function θ_η on S,

$$\theta_\eta \longleftrightarrow [0, \eta]$$

Equilibrium *sequence* correspondence is

$$\Gamma_\eta^\infty : \widehat{\Omega} = [0,1]^2 \times [0,1]^2 \times \ldots \rightsquigarrow \{B, N\}^\infty = \Omega$$

Using iid assumption for $(\epsilon_i)_{i=1}^\infty$, we obtain the *iid product* $(\theta_\eta)^\infty$, a belief function on $\Omega = S^\infty$

More on IID products

For $\theta = \theta_\eta$, $\Gamma = \Gamma_\eta$,

$$core\left(\theta^{\infty}\right) = \int_{\left([0,1]^2\right)^{\infty}} \Delta\left(\Gamma\left(\epsilon_1\right) \times \Gamma\left(\epsilon_2\right) \times ...\right) dm^{\infty}\left(\epsilon_1, \epsilon_2, ...\right)$$

All forms of correlation and heterogeneity are admitted
In particular,

$$core\left(\theta^{\infty}\right) \supset core\left(\theta\right) \otimes core\left(\theta\right) \otimes ...$$

But the product satisfies

Product property: $\theta^{\infty}\left(A \times B \times S^{\infty}\right) = \theta^{\infty}\left(A \times S^{\infty}\right)\ \theta^{\infty}\left(B \times S^{\infty}\right)$

FOUNDATIONS: Axioms for $\{\succeq_{n,s^n}\}$ on $\mathcal{F}$. Preferences at each node

BELIEF FUNCTION UTILITY: $\succeq_{n,s^n}$ is represented by

$$U_{n,s^n}(f) = \int_\Omega f d\nu_{n,s^n}$$

$\pi \in \Pi$: set of all finite permutations of $\mathbb{N}$
permuted act $(\pi f)(s_1, s_2, ...) = f\left(s_{\pi(1)}, s_{\pi(2)}, ...\right)$

SYMMETRY: $f \sim_{n,s^n} \pi f$ no reason to distinguish between markets
Symmetry for every history: even after sample $B_1, N_1, B_2, N_2, ..., B_k, N_k$.
Order has no significance in cross-section
Sample might suggest a distinguishing characteristic of markets previously overlooked - we rule out such changes of paradigm

WHY WEAKEN INDEPENDENCE?

Bayesian model adds Independence Axiom (randomization never valuable)
Write simply $\succeq$ for the generic conditional preference
Entry game, $B \prec N$; say $B \sim .8N$
Aversion to ambiguity about:

- Single experiment (Ellsberg): $\frac{1}{2}B_1 + \frac{1}{2}(.8N_1) \succ .8N_1$
- Heterogeneity: $\frac{1}{2}B_1 + \frac{1}{2}N_1 \succ \frac{1}{2}B_1 + \frac{1}{2}N_2$
- Correlation/patterns:

$$\begin{aligned} & \tfrac{1}{2}\{B_1B_2, N_1N_2\} + \tfrac{1}{2}\{B_1B_3, N_1N_3\} \\ \succ\ & \{B_1B_2, N_1N_2\} \sim \{B_1B_3, N_1N_3\} \end{aligned}$$

WEAK ORTHOGONAL INDEPENDENCE

Randomization is a matter of indifference sometimes

f and g are *orthogonal* if they depend on different experiments

$$- - - - \underset{J}{\overset{g}{-}} - - - - \bullet - - - - - \underset{I}{\overset{f',\, f}{-}} - - - - - \bullet - - - - -$$

WOI: For all suitably orthogonal acts, and $0 < \alpha < 1$,

$$f' \succeq f \quad \Longleftrightarrow \alpha f' + (1 - \alpha)\, g \succeq \alpha f + (1 - \alpha)\, g$$

"orthogonal acts do not hedge one another"

e.g. $\frac{1}{2}B_3 + \frac{1}{2}N_1 \sim \frac{1}{2}B_3 + \frac{1}{2}N_2$
WOI is consistent with the three examples contradicting Independence

INTUITION for "orthogonal acts do not hedge one another"

(1) No ambiguity about common factor (η)

(2) Selection is in some sense "stochastically independent" across markets
Rules out certainty that selection is identical in all markets, where expect

$$\frac{1}{2}B_1 + \frac{1}{2}(.8N_2) \succ B_1 \sim .8N_2$$

Bottom Line: WOI is a simple behavioral condition that a decision-maker can understand and accept/reject

CONSEQUENTIALISM: $f' \sim_{n,s^n} f$ if $f'(s^n, \cdot) = f(s^n, \cdot)$

Unrealized parts of the tree do not matter

COMMUTATIVITY: $\succeq_{n,\pi s^n} = \succeq_{n,s^n}$

The order of past observations does not matter

There is no natural ordering of cross-sectional data

WEAK DYNAMIC CONSISTENCY: For any $n \geq 1$, sample s^n and acts f', f over $S_{n+1} \times S_{n+2} \times ...$,

$$f' \succeq_{n,s^n} f \text{ for all } s^n \implies f' \succeq f$$

and, if in addition $f' \succ_{n,s^n} f$ for some s^n, then $f' \succ f$

Dynamic Consistency restricted to cases where

- observe outcomes in some markets and then bet on outcomes in others
- sample and then choose (update once)

How normative if permit violation of DC? can't have everything (ES, 2011): Consequentialism, DC, and Symmetry (for ex ante preference alone) are inconsistent with each of the three canonical violations of Independence described above

THEOREM: The above axioms are satisfied if and only if:

(i) **Choice:** For every (n, s^n), there exists a (unique Borel) probability measure $\mu_n(\cdot \mid s^n)$ on $Bel(S)$ such that $\succeq_{n,s^n}$ has a belief function utility with

$$\nu_n(\cdot) = \int_{Bel(S)} \theta^\infty(\cdot)\, d\mu_n(\theta)$$

(ii) **Inference:** There exists a likelihood function $L(\cdot \mid \theta)$ that is exchangeable for each θ, and s.t. $\{\mu_n\}$ is obtained via Bayes' Rule from μ_0 and L

Entry Game Again $(S = \{B, N\})$

Choice (ex ante): Need to specify μ_0 on $Bel(\{B, N\})$

Parameter $\eta \longleftrightarrow [0, \eta] \longleftrightarrow \theta_\eta$

Thus μ_0 is just a prior on the parameter $\eta \in (0, 1]$

$\nu_0(\cdot) = \int_{[0,1]} (\theta_\eta)^\infty(\cdot)\, d\mu_0(\eta)$ defines utility $U_0(\cdot)$
Each market described by the same (unknown) probability interval

Inference: How to update? $\Pr(B \mid \eta) \in [0, \eta]$, multiple likelihoods
Take an Average - *ANY* exchangeable $L(\cdot \mid \eta) = \int_{\Delta(\{B,N\})} q^{\infty} d\lambda_{\eta}(q)$
$L(\cdot \mid \eta)$ is not iid; uncertain interpretation (Acemoglu et al, 2009), Moon-Schorfheide (2012)

Surprise? Moon-Schorfheide inference is consistent with a model of choice that incorporates concern with poor understanding of selection

Bayesian machinery applies to describe process of posteriors $\{\mu_n\}$

If each λ_{η} is uniform on $[0, \eta]$, obtain intuitive results in the limit as $n \longrightarrow \infty$ along sample s^n

Learns about η but not about selection

Are preferences needed?

1. Choice: We need preferences to characterize
$$\nu_n(\cdot) = \int_{Bel(S)} \theta^{\infty}(\cdot)\, d\mu_n(\theta)$$
This is NOT a result about belief functions only. Contrast with de Finetti, which can be stated as a theorem about probability measures

2. Inference: Likelihood of form
$$L(\cdot \mid \theta) = \int_{\Delta(S)} q^{\infty} d\lambda_{\theta}(q)$$
is well defined for any abstract parameter θ. But the theorem states that belief functions $\theta \in Bel(S)$ are the correct parameters in order to connect to preference. Again, preference is important

What about alternative functional forms? Along the lines of Epstein-Seo (2010, Model 1), Al Najjar-de Castro (2010), Cerreia et al (2010), Klibanoff-Mukerji-Seo (2011)

'Robust Bayesian' - robustness wrt prior beliefs

$$U^{RB}(f) = \inf_{\mu \in \mathcal{M}} \int_{\Delta(\{B,N\})} \left(\int_{\Omega} f(\omega)\, dq^{\infty}(\omega) \right) d\mu(q)$$

Why not use this functional form for choice?

Does not permit the behavioral violations of Independence expressing concern with heterogeneity and correlation

And there is no updating rule for the above model

MORE COMPLICATED ENTRY GAMES

- Profits depend also on exogenous variables (player characteristics/policy variables), $x_i = (x_{i1}, x_{i2}) \in X$

	out	in
out	$0, 0$	$0, \beta_2 x_{i2} - \epsilon_{i2}$
in	$\beta_1 x_{i1} - \epsilon_{i1}, 0$	$\beta_1 x_{i1} + \eta - \epsilon_{i1}, \beta_2 x_{i2} + \eta - \epsilon_{i2}$

Construct state space as follows: $Y =$ set of pure strategy equilibria
Uncertainty concerns outcomes for each given x, so take

$$S = Y^X$$

Each parameter (β_1, β_2, η) implies an equilibrium correspondence and "random sets" - hence belief function on $S = Y^X$

Full state: $\omega = (s_1, ..., s_i, ...)$, $s_i \in Y^X$

Act $f : \Pi_{i=1}^{\infty} Y^X \to [0, 1]$

x as a policy tool: choose x^* for market 1 - corresponds to act f^*

$$f^*(s_1, ..., s_i, ...) = u(s_1(x^*), x^*)$$

Connection to statistical decision theory

Conditional decisions solve

$$\max_{f\in\Upsilon}\int_{Bel(S)}\left[\int_{\Omega} f d\theta^{\infty}\right] d\mu_N\left(\theta \mid s^N\right)$$

Define a (feasible) *decision rule* δ as a mapping $\delta : S^N \to \Upsilon$

$\mathbb{D} = \Upsilon^{S^N}$, the set of all decision rules

Then the collection of all conditional problems, with s^N varying over all possible samples, can be reformulated as one of ex ante minimization over decision rules:

$$\min_{\delta\in\mathbb{D}}\int_{Bel(S)} r\left(\delta,\theta\right) d\mu_0\left(\theta\right) \quad \equiv \min_{\delta\in\mathbb{D}} \text{ [Bayes' risk of } \delta\text{]}$$

where

$$r\left(\delta,\theta\right) = -\boldsymbol{\Sigma}_{s^N} L\left(s^N \mid \theta\right)\left(\int_{\Omega}\delta\left(s^N\right)(\cdot)\, d\theta^{\infty}(\cdot)\right)$$

$r\left(\delta,\cdot\right)$ is the *risk function* of decision rule δ - note its special structure!