In Case I am Wrong: Ambiguity Aversion Preference Concerned about Model Mis-specification

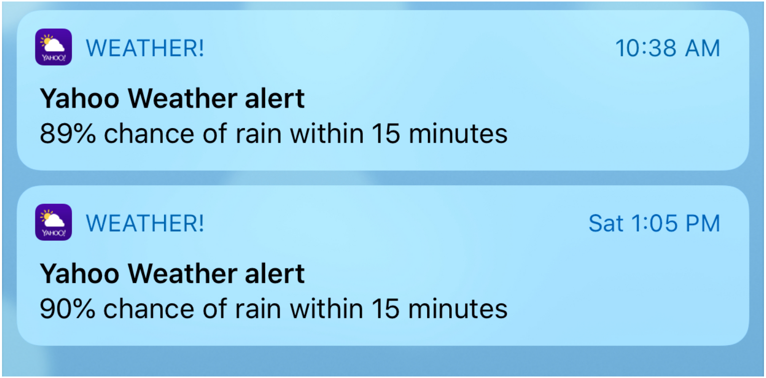
A discussion of [Hansen and Sargent, 2001] and [Strzalecki, 2011]

Tao Wang

Decision Making under Risk and Uncertainty Seminar, 2018

Outline

While it is raining outside, Yahoo Weather tells me this...



- What if it says: 50% chance of raining in next 15 minutes?

Unification of Language

If you are a statistician, you never say you are wrong, you just call it a model mis-specification.

Decision Making with Non-unqiue Priors

- You need a thought experiment to select a prior among multiples to "complete" your preference and to make the decision.
- You imagine playing the game with a malevolent nature who tries the best to go against your will.
- The way to select the prior from multiple priors may not be unique.
- You usually have some clues about the true probability. You may take this into account when evaluating different priors.
- We study such a utility representation.

From Unique to Non-unique Priors

- **Subjective Expected Utility(EU)**

$$V(f) = \int_{s \in S} u(f(s)) d\pi$$

- **Maxmin Expected Utility(MEU)**

$$V(f) = \min_{\pi \in \Delta(S)} \{ \int_{s \in S} u(f(s)) d\pi \}$$

- **Multiplier Utility(MP)**

$$V(f) = \min_{\pi \in \Delta(S)} \{ \int_{s \in S} u(f(s)) d\pi + \theta H(\pi || \bar{\pi}) \}$$

- $\pi \in \Delta(S)$, prior probability distributions over state spaces.
- $f \in F : S \to X$ is the act.
- $\bar{\pi}$: the best guess of the true probability.
- $\theta \in [0, +\infty)$: the degree of confidence.
- $H(\pi || \bar{\pi})$: difference of the prior from the best guess.

Interpretation: Concern of Model Mis-specification

$$V(f) = \min_{\pi \in \Delta(S)} \{ \int_{s \in S} u(f(s)) d\pi + \theta H(\pi || \bar{\pi}) \}$$

- Let us not worry about the definition of $H(\pi||\bar{\pi})$ for now.
- $\theta \to +\infty$, fully trust the best guess. Do not discriminate prirors at all. $\theta H(\pi||\bar{\pi})$ are equal $\forall \pi \in \Delta(S)$. Simply Expected Utility with $\bar{\pi}$ if it is unique prior.
- $\theta = 0$, does not trust your guess at all, thus simply ignores it. Do not discriminate all priors, either. Back to MEU.
- $0 < \theta << \infty$, concerned about model mis-specification.
- If all priors give the same expected utility, then choose the prior that is closest to the best guess.

Some Intuition: Back to the Thought Experiment

$$V(f) = \min_{\pi \in \Delta(S)} \{ \int_{s \in S} u(f(s)) d\pi + \theta H(\pi || \bar{\pi}) \}$$

- Imagining playing the game with a malevolent nature who tries the best to go against your will.
- MEU: the malevolent nature tries to lower your expected utility.
- MP: the nature also perturbs your beliefs.

Why Called Multiplier Preference?

- MP can be turned into a constrained problem with MEU with a corresponding lagrangian multiplier. [Hansen and Sargent, 2001]
- MP problem

$$max_{f \in F}\{min_{\pi \in \Delta(S)} \int_S u(f(s))d(\pi) + \theta H(\pi||\bar{\pi})\}$$

- Constrained MEU problem

$$max_{f \in F}\{min_{\pi \in \Delta(S)} \int_S u(f(s))d(\pi)\} \quad \text{s.t} \quad H(\pi||\bar{\pi}) \leq \eta$$

MP is a Special Case of Variational Preference(VP)

$$V(f) = min_{\pi \in \Delta(S)}\{\int u(f(s))d\pi + \underbrace{c(\pi)}_{\theta H(\pi||\bar{\pi}) \text{ in MP}}\}$$

- $\pi \in \Delta(S)$, prior probability distributions over state spaces.
- $f \in F : S \to X$ is the act.
- $\bar{\pi}$: the best guess as a reference probability.
- $c : \Delta(S) \to [0, +\infty)$, a convex cost or penalty function. How comfortable you are with different priors.

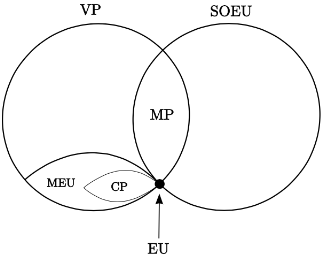
MP is also a Special Case of Second-Order Expected Utility(SOEU)

$$V(f) = \int_S \underbrace{\phi}_{\phi=\phi_\theta \text{ in MP}} (u(f(s)))d\bar{\pi}$$

- ϕ is some increasing and concave tansformation function.
- $\phi_\theta(u) = \begin{cases} -exp(-\frac{u}{\theta}) & \theta < \infty \\ u & \theta = \infty \end{cases}$
- Higher θ, higher risk aversion.
- $\pi \in \Delta(S)$, prior probability distributions over state spaces.
- $f \in F : S \to X$ is the act.
- $\bar{\pi}$: the best guess as a reference probability.

Mapping MP to Theoretical Landscape

- **VP**: variational preferences;
- **MP**: multiplier preferences;
- **SOEU**: second-order expected utility preferences;
- **EU**: expected utility preferences;
- **MEU**: maxmin expected utility preferences;
- **CP**: constraint preferences.



Setup: Anscombe-Aumann's Frameowork

- $\Delta(S) = (S, \Sigma)$ where S is state space, Σ is the sigma-algebra measure over S. $\Delta^c(S) \subset \Delta(S)$ is set of countably additive probability measures.
- $f : S \to \Delta(X)$, an act is a mapping from state space S to objective lotteries $\Delta(X)$, where X is the outcome space. F is the set of acts f.
- $\Delta(X)$: set of constant acts, namely the outcome induced by the act does not depend on state.
- $\succcurlyeq \in FxF$, binary relation over acts.

Axioms

- **A.1. Weak Order.** Binary relation $\succcurlyeq$ is transitory and complete.
- **A.2. Weak Certainty Independence.** $\forall f, g \in F, h, h' \in \Delta(X)$, constant acts$and \alpha \in (0,1), \alpha f + (1-\alpha)h \succcurlyeq \alpha g + (1-\alpha)h \Rightarrow \alpha f + (1-\alpha)h' \succcurlyeq \alpha g + (1-\alpha)h'$.
- **A.3. Continuity.** $\forall g, f, h \in F$, the sets $\{\alpha \in [0,1] | \alpha g + (1-\alpha)f \succcurlyeq h\}$ and $\{\alpha \in [0,1] | n \succcurlyeq \alpha g + (1-\alpha)f\}$ are closed.
- **A.4. Monotonicity.** If for $f, g \in F, f(s) \succcurlyeq g(s) \forall s \in S$, then $f \succcurlyeq g$.
- **A.5. Ambiguity Aversion.** If $f, g \in F$ and $\alpha \in (0,1)$, then $f \sim g \Rightarrow \alpha f + (1-\alpha)g \succcurlyeq f$.

Axioms, continued..

- **A.6. Non-triviality.** $f \succcurlyeq g$ for some $f, g \in F$.
- **A.8. Weak Monotone Continuity.** For $f, g \in F$ and $h \in \Delta(X)$, $\{E_n\}_{n\geq 1} \in \Sigma$ with $E_1 \supseteq E_2 \ldots$, and $\cap_{n\geq 1} E_n = \emptyset$, then $f \succcurlyeq g$ implies there exists $n_0 \geq 1$ such that $hE_{n_0} f \succ g$, where $hE_{n_0} f = h(s) \forall s \in E_{n_0}$ and $hE_{n_0} f = f(s) \forall s \notin E_{n_0}$.
- P.2. Sure-Thing Principle. For all $E \in \Sigma$, $f, g, h, h' \in F$, if $f_E h \succcurlyeq g_E h$, then $f_E h' \succcurlyeq g_E h'$.

Representation Theorem [Strzalecki, 2011]

Theorem

Axioms A.1-A.6, A.8 and P.2 are satisfied if and only if $\succcurlyeq$ has a MP representation. Two tuples $(u, \theta, \bar{\pi})$ and $(u', \theta', \bar{\pi}')$ represent the same preference if and only if $\bar{\pi}$ and $\bar{\pi}'$ are identical. Also, u and θ are unique up to affine transformation.

1. u and θ jointly determined as they need to be in the same scale.
2. Sure-Thing Principle is added to aximization of VP by [Maccheroni et al., 2006]. There, the representation needs to pin down a convex, lower semicontinuous, and grounded function $c(\pi)$.

Essence of the Proof

- MP is a the intersection of VP and SOEU.
  - Only a special form of $H(\pi||\bar{\pi})$ can be a *SOEU*. (We will discuss in detail next).
  - Only a special form of ϕ_θ can be VP.

Reinterepretation of Ellsberg Paradox

- Objective Urn A: 50 red, 50 blue.
- Subjective Urn B: X red, 100-X blue.
- A denotes bet in A for blue. Same for others.
- Bet prize 100 or 0.
- **Observed Preference**: $A \sim A \succ B \sim B$
- **EU**: $U(A) = V(A) = V(B) = V(B) = \frac{1}{2}u(100) + \frac{1}{2}u(0)$.
- **MEU:** $V(A) = V(A) = \frac{1}{2}(u(100) + u(0)) > u(0) = V(B) = V(B)$.
- **MP:** $V(A) = V(A) = \phi_\theta(\frac{1}{2}u(100) + \frac{1}{2}u(0)) > V(B) = V(B) = \frac{1}{2}\phi_\theta(u(100)) + \frac{1}{2}\phi_\theta(u(0)) \quad \forall\theta < +\infty$

Elicitation of the MP Preference

- **Revealed Preference Approach**
  - u can be solved by looking certainty equivalence for objective urn under an Ellsberg's experiment.
  - θ can be solved by compare certainty equivalence for the objective urn and subjective urn.
- **Hypothetical Questions in Surveys**
  - Ask about subjective beliefs.
  - And the respondent's confidence about her answers.

Measure of Divergence n Probability Distributions: Relative Entropy or Kullback−Leibler divergence

- If π and $\bar{\pi}$ are measures in the same state space,

$$H(\pi||\bar{\pi}) = \begin{cases} \int_S log(\frac{d\bar{\pi}}{d\pi})d\bar{\pi} & \pi \in \Delta^c(\bar{\pi}) \\ +\infty & \text{otherwise} \end{cases}$$

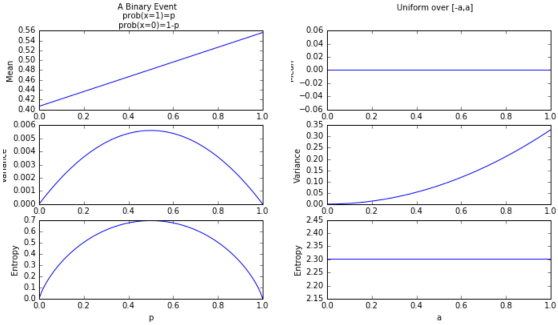
- Expected logarithmic difference between the two probabilities evaluated with the probability of $\bar{\pi}$, as it is the reference distribution.
- **Intuition**: average surprise of seeing π given the prior $\bar{\pi}$.
- **Non-negativity**: $H(\pi||\bar{\pi}) \in [0, +\infty)$, 0 if two distributions are identical. $+\infty$ if the two do not share outcome space.
- **Invariant under Transformation**: independent from the outcome. Purely a discription of probability distributions.
- **Non-symmetry**: not exactly a distance. $H(p||q) \neq H(q||p)$.
- **Convexity**: convext function of the pair of proabilities.

Entropy

- A random variable X whose probability density function is $f(x) \geq 0 \quad \forall x \in X$.
- **Entropy** for the probability distribution of X is defined as $H(X) = E_X(-log(f(x)) = -\int_{x \in X} log(f(x))f(x)$
- **Intuition**: the less likely the event of x, the more information it contains when it happens. On average, how surprised would I feel?
- **Less surprise, more certainty**.
- Relative Entropy, Conditional Entropy, Joint Entropy etc. defined correspondingly.

Entropy: Examples with Random Variables

Both entropy and variance seem to capture some notion of dispersion. But they are different.



More Intuition: Black Cats and White Cats

- You start from a prior that there are equal number of black and white cats in this world. High uncertainty.
- After you keep seeing black cats more often than white ones. $80\% v.s. 20\%$. There is surprise. It is useful to learn.
- Posterior belief shifts toward black cats. Lower entropy, lower uncertainty. New informaiton lowers subjective uncertainty.
- Learning won't happen if you see black and white equally often. Zero relative entropy. There is no new information.
- Therefore, Relative Entropy $=$Surprise$=$ Information Gain.

Entropy in Different Contexts

- Termodynamics:
  - Disorder or chaos of a system.
- Information Theory [Shannon, 1948]
  - Entropy: average information content. Units: bits for log_2; nats for *ln*.
  - Relative Entropy: information loss if a different coding system being used.
- Statistical Inference
  - Information gain from prior to posterior.
  - Negative log likelihood of the realized data with the true model.
- Machine Learning
  - information gain using one model compared to the other.
- Economics
  - subjective uncertainty. (more work can be done)

Different Ways of Characterizing the Problem

- Decision Maker: adopting a subjective belief while allowing for the possiblity of the belief being inconsistent with the true world.
- Statistician: try to make inference but with preservation as to whether the model used it correct or not.
- Engineers: robust control by taking approximating model and statistically purtubing it. Maxmin objective function. [Hansen and Sargent, 2008]

Concluding Remarks

- Unique Prior $\Rightarrow$ Multiple Priors $\Rightarrow$ Priors and Posteriors with Learning.
- Decision maker does not only makes the decision, but also LEARNS.
- Especially important in modeling real-world economic and finance decisions.
- In a dynamic decision-making this is useful and important.

Hansen, L. and Sargent, T. J. (2001).
Robust control and model uncertainty.
American Economic Review, 91(2):60–66.

Hansen, L. P. and Sargent, T. J. (2008).
Robustness.
Princeton university press.

Maccheroni, F., Marinacci, M., and Rustichini, A. (2006).
Ambiguity aversion, robustness, and the variational representation of preferences.
Econometrica, 74(6):1447–1498.

Shannon, C. E. (1948).
A mathematical theory of communication.
Bell system technical journal, 27(3):379–423.

Strzalecki, T. (2011).

Axiomatic foundations of multiplier preferences.
Econometrica, 79(1):47–73.