THE SIMPLE ANALYTICS OF DECISION THEORY

1. The Setup

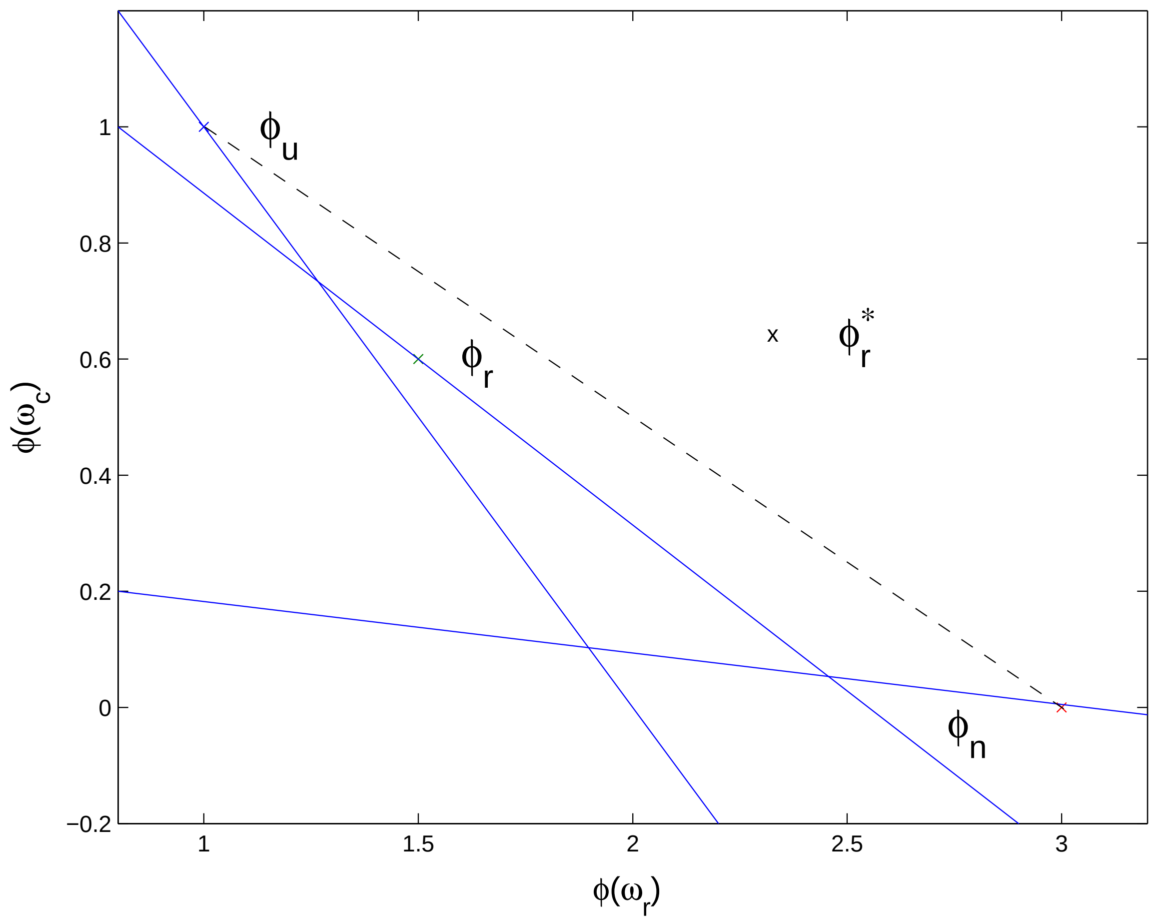
- A space $\Omega = \{\omega_1, \ldots, \omega_n\}$ of possible "states of the world".
- A set Φ of possible actions, or decisions, $\phi \in \Phi$.
- Each ϕ implies a pattern of "losses" across the different values of ω_i, and since these implied losses are the only aspects of ϕ we will be concerned with, we will think of Φ as a set of functions that map Ω into the real line.

2. Example

ω_r :	rain
ω_c :	clear
ϕ_u :	carry umbrella
ϕ_r :	carry plastic raincoat
ϕ_n :	carry no raingear

	ω_r	ω_c
ϕ_u	1	1
ϕ_r	1.5	.6
ϕ_n	3	0

3. Actions/decisions as points in $\mathbb{R}^n$



4. Ranking Rules

$\phi_1 \succ \phi_0 \Leftrightarrow$

$$(\forall \omega_i \in \Omega)\big(\phi_1(\omega_i) \leq \phi_0(\omega_i)\big) \text{ and } (\exists \omega_i \in \Omega)\big(\phi_1(\omega_i) < \phi_0(\omega_i)\big)$$

- ϕ is *admissible* in Φ iff $(\nexists \phi^* \in \Phi)\phi^* \succ \phi$
- ϕ_0 is *Bayesian* in Φ iff there is a probability P on Ω s.t.

$$\min_{\phi \in \Phi} \left\{ E_P\phi = \sum_i P(\omega_i)\phi(\omega_i) \right\} = E_P\phi_0 \,.$$

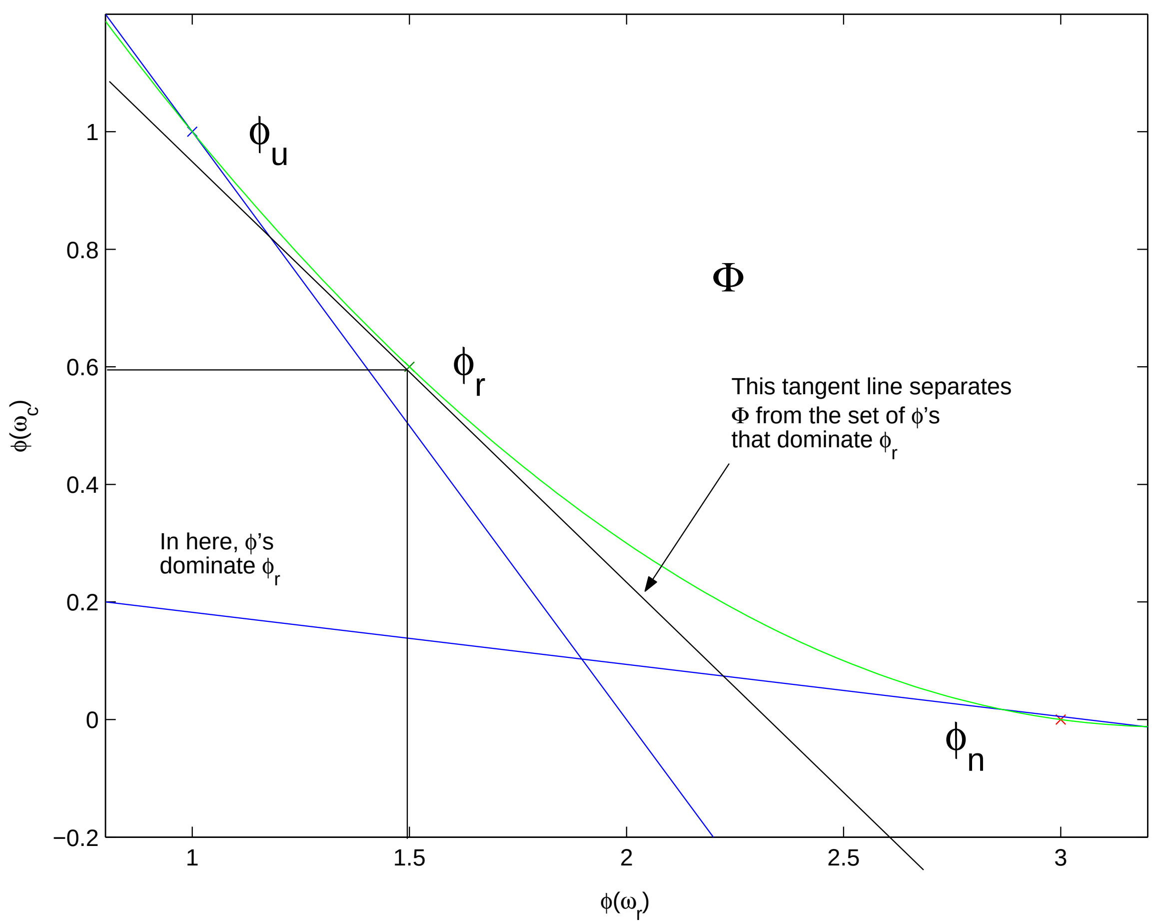
5. Complete Class Theorem

- If Φ is closed, convex, and bounded, every admissible ϕ is Bayesian and
- the admissible Bayesian ϕ's form a **minimal complete class**). I.e.,

$$(\forall\phi \text{ non-Bayesian})(\exists\phi_b \text{ Bayesian})\phi_b \succ \phi\,,$$

  and any strict subset of the Bayesian ϕ's fails to have this property.
- Like the conclusion that competitive equilibria are efficient, this is a consequence of the **separating hyperplane theorem**. Probabilities here play a role very close to that of prices in competitive general equilibrium theory.

6. Admissible rules as Bayesian



7. Remarks

- The separating hyperplane theorem only tells us that there will be linear functions separating convex sets. But probabilities, the coefficients in this linear function, have to be non-negative and sum to one.

- Looking at the graph, how do we know that the probabilities can't be negative? The summing to one is just a normalization.
- This is very much like the argument that prices should be positive and the fact that we are free to choose the numeraire for prices.

8. More Remarks

- We've derived probability without any reference to repeated trials and frequencies or to a symmetry that implies a set of events must be equiprobable.
- This approach is called **subjective probability**, because it arises from consideration of the behavior of a single rational decision maker.
- Our derivation is **normative**, meaning it describes optimal behavior. It need not describe actual behavior of less than fully rational people.
- Even rational agents need only behave *as if* they were minimizing expected loss. They need not actually think in terms of probabilities, any more than businessmen need to compute marginal products in order to arrive at efficient production decisions.

9. Convexity of Φ

- This is a strong assumption. It doesn't apply in our raincoat/umbrella/no-raingear example.
- It is probably no worse than our conventional assumption of convexity of production possibility sets, though.
- Just as we explain diminishing returns by arguing that farmers use the best land first, we can explain that the most cost-effective "wetness-reduction" methods will be undertaken first (plastic bag headgear, cheap collapsible umbrellas) and that the last increments of rain protection (Gore-Tex foul weather gear, e.g.) will be very costly (and thus cause high losses if it doesn't rain).

10. Other approaches, other assumptions

- Convexity plays a big role in this derivation because we have not used any properties of the "loss" measure except that less loss is better.
- If we instead assume that a rich set of ϕ's can be ranked (even if they aren't feasible) and that losses are in a certain sense comparable across states, we can derive the existence of probabilities without convexity of Φ. This is in fact the more common approach.

11. Inference

Actions ϕ are taken in two stages: An initial stage with no knowledge of where we are in Ω, followed by a second stage after we observe some data. We express this by writing $\phi = \phi(\omega, \delta_0, \delta_1(X))$, where $\delta_0 \in \mathbb{R}$ is the initial stage choice and $\delta_1(X(\omega))$ is the second stage choice, a function of the random variable $X : \Omega \to \mathbb{R}$. This is really only a special case of our previous setup. For each real-valued choice of δ_0 and choice of the function

$\delta_1 : \mathbb{R} \to \mathbb{R}$, $\phi(\omega, \delta_0, \delta_1(X(\omega)))$ is a function mapping $\Omega \to \mathbb{R}$ as before. Φ is the set of all such functions obtainable through feasible choices of δ_0 and δ_1.

So by our previous argument we know that, assuming Φ is closed, convex and bounded, any admissible action ϕ will minimize expected loss for some probability on Ω. Note that for any $\omega_i \in \Omega$, we can write $P[\omega_i] = P[\omega_i \mid X = x] \cdot P[\{\omega_j \mid X(\omega_j) = x\}]$. This follows directly from the definition of conditional probability. So

$$
\begin{aligned}
E_P[\phi(\delta_0, \delta_1)] &= \sum_{i=1}^{n} \phi\left(\omega_i, \delta_0, \delta_1(X(\omega_i))\right) P[\omega_i] \\
&= \sum_{x \in S} \sum_{\{\omega_i \mid X(\omega_i) = x\}} \phi\left(\omega_i, \delta_0, \delta_1(x)\right) P[\omega_i \mid X = x] P[\{\omega_j \mid X(\omega_j) = x\}] \\
&= E_P\left[E_P[\phi(\delta_0, \delta_1) \mid X = x]\right].
\end{aligned}
$$

Note that this is a special case of the **law of iterated expectations**. If there are no constraints that link our choice of $\delta_1(x)$ to our choice of $\delta_1(y)$ for $x \neq y$ or to our choice of δ_0, then we can choose δ_1 separately for each possible value of $X(\omega)$. And it is clear from the formula that to minimize expected loss overall, we should always choose $\delta_1(x)$ to minimize $E[\phi(\delta_0, \delta_1(x)) \mid X = x]$.

Conclusion Optimal decision making, done at various states of knowledge, requires using an initial probability distribution, and then updating that distribution to a new conditional distribution given the data, after new information is observed. Inference is the process of updating the distribution based on observed data.

12. OTHER VIEWS OF INFERENCE

The view that inference is nothing more or less than the updating of probability distributions based on new information is the **Bayesian** view of inference. It links inference very directly to decision making.

The Bayesian view has nothing to say about where the probability distribution that applies before the data is seen might come from. There are formal rules for using data to update a distribution reflecting beliefs about uncertain prospects, but there are no formal rules about how to come up with the initial beliefs — not even any way to express the idea that we are initially "completely ignorant" or "completely unprejudiced".

Other ways of thinking about inference try to avoid this loose end by in one way or another restricting "probabilities" to apply only to "objective" types of uncertainty. This is what disconnects these approaches from decision theory.

13. A QUESTION TO CHECK YOUR UNDERSTANDING

In our rain gear example, suppose the list of feasible ϕ's (i.e. the whole Φ set) consisted of the following list of points:

	$\phi(\omega_r)$	$\phi(\omega_c)$
1	0.95	0.62
2	0.23	0.79
3	0.61	0.92
4	0.49	0.74
5	0.89	0.18
6	0.76	0.41
7	0.46	0.94
8	0.02	0.92
9	0.82	0.41
10	0.44	0.89

Which are admissible? Which are Bayes?