ST790 — Fall 2022
Imprecise-Probabilistic Foundations of Statistics

Ryan Martin
North Carolina State University
www4.stat.ncsu.edu/~rmartin

Week 02b

This lecture

- More on random sets & capacities
- Higher-order monotonicity
- Choquet's theorem
- Finite $\mathbb{X}$ case
- Maxitivity
- ...

Random sets

- For a random set $\mathcal{X}$ on $\mathbb{X}$, the two (dual) capacities are

$$\begin{aligned} \underline{\Pi}(A) &= \mathsf{P}(\mathcal{X} \subseteq A) \\ \overline{\Pi}(A) &= \mathsf{P}(\mathcal{X} \cap A \neq \varnothing), \quad A \subseteq \mathbb{X} \end{aligned}$$

- Easy to see that these are capacities according to our previous definition; in particular, *monotonicity* is clear
- Moreover, $\underline{\Pi}$ is *2-monotone*,[1] i.e.,

$$\underline{\Pi}(A \cup B) \geq \underline{\Pi}(A) + \underline{\Pi}(B) - \underline{\Pi}(A \cap B),$$

  and $\overline{\Pi}$ is *2-alternating* (inequality reversed)
- Capacities associated with random sets satisfy more:
  - $\underline{\Pi}$ is ∞-*monotone* and $\overline{\Pi}$ is ∞-*alternating*
  - $\overline{\Pi}$ is upper semicontinuous

[1] Homework!

Properties

- $\psi : 2^{\mathbb{X}} \to [0,1]$ is K-monotone if, for all $A_1, \dots, A_K$,

$$\psi\Big(\bigcup_{k=1}^{K} A_k\Big) \geq \sum_{\varnothing \neq I \subseteq \{1,\dots,K\}} (-1)^{|I|+1}\, \psi\Big(\bigcap_{i \in I} A_i\Big)$$

- K-alternating if the reverse inequality holds
- ∞-monotone/alternating if K-monotone/alternating for all K
- Note: probabilities are ∞-monotone and ∞-alternating[2]
- Proof that $\overline{\Pi}$ is ∞-alternating:

$$\begin{aligned}
\overline{\Pi}\Big(\bigcup_{k=1}^{K} A_k\Big) &= \mathrm{P}\Big\{\bigcup_{k=1}^{K} (\mathcal{X} \cap A_k) \neq \varnothing\Big\} \\
&= \mathrm{P}(\mathcal{X} \cap A_1 \neq \varnothing \text{ or } \cdots \text{ or } \mathcal{X} \cap A_K \neq \varnothing) \\
&= \cdots
\end{aligned}$$

[2] Recall the inclusion–exclusion formula...

Properties, cont.

- $\psi : 2^{\mathbb{X}} \to [0,1]$ is *upper semicontinuous*[3] (USC) if

$$K_n \downarrow K_\infty \implies \psi(K_n) \downarrow \psi(K_\infty), \quad n \to \infty \quad \text{(compact } K\text{'s)}$$

- Countably additive probabilities are upper semicontinuous
- $\overline{\Pi}$ is upper semicontinuous
  - $E_n = \{\omega : \mathcal{X}(\omega) \cap K_n \neq \varnothing\}$
  - $K_n \downarrow K_\infty$ implies $E_n \downarrow E_\infty$
  - upper semicontinuity of P implies $\mathsf{P}(E_n) \downarrow \mathsf{P}(E_\infty)$

Choquet's theorem.

ψ is ∞-alternating and USC iff there exists a random set $\mathcal{X}$ such that $\psi(\cdot) = \mathsf{P}(\mathcal{X} \cap \cdot \neq \varnothing)$

[3] For technical reasons, it's enough to focus only on compacts

Properties, cont.

- Let's take a look at the finite case:
  - $\mathbb{X}$ is a finite set
  - domain $2^{\mathbb{X}}$ of $\mathcal{X}$ is finite
- This simplifies description of the random set's dist'n
- That is, all we need is the *mass function* of $\mathcal{X}$,

$$f(A) = \mathsf{P}(\mathcal{X} = A), \quad A \in 2^{\mathbb{X}}$$

- Then the capacities can be evaluated as

$$\underline{\Pi}(A) = \mathsf{P}(\mathcal{X} \subseteq A) = \sum_{B \in 2^{\mathbb{X}} : B \subseteq A} f(B)$$

$$\overline{\Pi}(A) = \mathsf{P}(\mathcal{X} \cap A \neq \varnothing) = \sum_{B \in 2^{\mathbb{X}} : B \cap A = \varnothing} f(B)$$

Example

Let $\mathbb{X} = \{a, b, c\}$ and define a random set $\mathcal{X}$ with mass function:

A	$\emptyset$	$\{a\}$	$\{b\}$	$\{c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$	$\mathbb{X}$
$f(A)$	0.0	0.1	0.1	0.2	0.3	0.0	0.2	0.1

Get the distribution function by summing, e.g.,

$$\begin{aligned} F(\{a, c\}) &= \mathrm{P}(\mathcal{X} \subseteq \{a, c\}) \\ &= \sum_{B : B \subseteq \{a,c\}} f(B) \\ &= f(\emptyset) + f(\{a\}) + f(\{c\}) + f(\{a, c\}) \\ &= 0.3 \end{aligned}$$

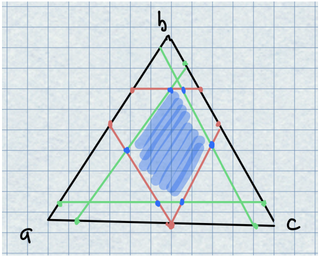
Example, cont.

Repeating this for all the subsets gives the distribution function:

A	$\varnothing$	$\{a\}$	$\{b\}$	$\{c\}$	$\{a,b\}$	$\{a,c\}$	$\{b,c\}$	$\mathbb{X}$
$F(A)$	0.0	0.1	0.1	0.2	0.5	0.3	0.5	1.0

Moreover, can calculate the lower/upper prob's assigned to each element and visualize the corresponding credal set

A	$\underline{\Pi}(A)$	$\overline{\Pi}(A)$
$\{a\}$	0.1	0.5
$\{b\}$	0.1	0.7
$\{c\}$	0.2	0.5



Properties, cont.

- Like with random variables, if we know, e.g., $\mathsf{P}(\mathcal{X} \subseteq A)$ for all A, then we should be able to recover the mass function
- i.e., 1–1 correspondence between mass and dist'n functions
- Connection is via *Möbius inversion*

$$f(A) = \sum_{B \in 2^{\mathbb{X}} : B \subseteq A} (-1)^{|A \cap B^c|} \mathsf{P}(\mathcal{X} \subseteq B)$$

- Very general theory behind this (e.g., Nguyen Sec. 4.2)[4]
- *Intuition.* For a discrete bivariate random variable, how to get the joint mass function from the joint CDF?

[4] I'll walk you through a simple, combinatorial proof in HW

Example, cont.

- Check that we can recover mass function from the distribution function using the Möbius formula
- For example,

$$\begin{aligned}\sum_{B:B\subseteq\{a,c\}} (-1)^{2-|B|}F(B) &= F(\varnothing) - F(\{a\}) - F(\{c\}) + F(\{a,c\}) \\ &= 0.0 - 0.1 - 0.2 + 0.3 \\ &= 0 \\ &= f(\{a,c\}) \quad \checkmark\end{aligned}$$

- Similarly for other subsets...

Maxitivity

- A functional $\psi : 2^{\mathbb{X}} \to [0,1]$ is called *maxitive* if

$$\psi(A \cup B) = \max\{\psi(A), \psi(B)\}, \quad \text{all } A, B$$

- Extends to finite (and arbitrary) unions
- Common measures of dimension/complexity[5] are maxitive
- Maxitivity implies:
  - monotonicity[6] (easy)
  - 2-alternating (pretty easy)
  - ∞-alternating (not as easy...)
- Choquet: ψ maxitive + USC $\implies \psi(\cdot) = \mathsf{P}(\mathcal{X} \cap \cdot \neq \varnothing)$
- Turns out maxitivity forces a very special structure...

[5] e.g., packing dimension (Nguyen's book, p. 81–86)

[6] So, ψ is a capacity (modulo boundary conditions at $\varnothing$, $\mathbb{X}$)

Example

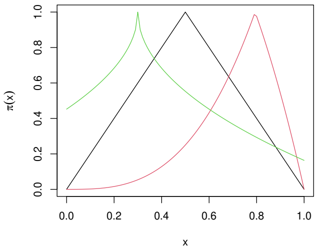
Define $\mathcal{X} = \{x \in \mathbb{X} : h(x) \geq U\}$, for an upper semicontinuous function $h : \mathbb{X} \to [0,1]$, and $U \sim \mathrm{Unif}(0,1)$.

- Recall

$$\overline{\Pi}(A) = \sup_{x \in A} h(x), \quad A \subseteq \mathbb{X}$$

- Maxitivity:

$$\begin{aligned} \overline{\Pi}(A \cup B) &= \sup_{x \in A \cup B} h(x) \\ &= \max\{\sup_{x \in A} h(x), \sup_{x \in B} h(x)\} \\ &= \max\{\overline{\Pi}(A), \overline{\Pi}(B)\} \end{aligned}$$



Hitting prob, $\mathbb{X} = [0,1]$

Maxitivity, cont.

- Fact:[7] the only maxitive functionals are of the type above, i.e.,

$$\psi(A) = \sup_{x \in A} h(x), \quad A \subseteq \mathbb{X}, \quad \text{for some } h$$

- So, for $\mathcal{X}$'s with a maxitive upper prob, the distribution is *completely determined* by its hitting probability, i.e.,

$$\mathsf{P}(\mathcal{X} \cap A \neq \varnothing) \equiv \sup_{x \in A} \underbrace{\mathsf{P}(\mathcal{X} \ni x)}_{\text{hitting prob}}, \quad A \subseteq \mathbb{X}$$

- Remarks:
  - compare to mass/density function in ordinary probability
  - recall my previous comment about sup's in p-values

[7] Molchanov's Proposition 1.16 (2005 version)

Random sets in statistical inference

- Dempster's framework[8] from the 1960s, a generalization of both fiducial & Bayes, was based on random sets
- Basic idea:
  - statistical model: $X = a(\theta, U)$
  - $U \sim \mathsf{P}_U$ is an auxiliary variable, a "random seed"
- For given (x, u), let $\mathbb{T}_x(u) = \{\theta \in \mathbb{T} : x = a(\theta, u)\}$
- Note: $u \mapsto \mathbb{T}_x(u)$ is a set-valued map[9]
- *Fiducial flip:* take the "conditional distribution" of U, given $X = x$, to be the marginal P_U
- Quantify uncertainty about θ, given $X = x$, via the dist'n of the (data-dependent) random set $\mathbb{T}_x(U)$ with $U \sim \mathsf{P}_U$

[8] Modern description in Dempster (2008 *IJAR*)

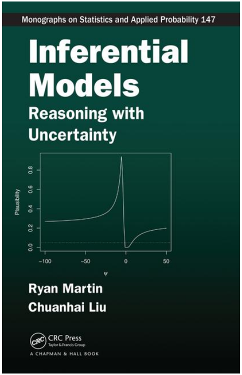
[9] Dempster and others sometimes call this a *multivalued map*

Random sets in statistical inference, cont.

- Dempster's formulation is subjective; no concern about statistical properties
- Rather than the *fiducial flip*
  - quantify uncertainty about U, given $X = x$, by a random set $\mathcal{U} \subseteq \mathbb{U}$
  - $\mathbb{T}_x(\mathcal{U}) = \bigcup_{u \in \mathcal{U}} \mathbb{T}_x(u)$
  - $\overline{\Pi}_x(A) = \mathsf{P}_{\mathcal{U}}\{\mathbb{T}_x(\mathcal{U}) \cap A \neq \varnothing\}$
- For suitable $\mathcal{U}$, inference about θ based on dist'n of $\mathbb{T}_x(\mathcal{U})$ is *valid*,

$$\sup_{\theta \in A} \mathsf{P}_{X|\theta}\{\overline{\Pi}_X(A) \leq \alpha\} \leq \alpha$$

for all $\alpha \in [0,1]$ and all $A \subseteq \mathbb{T}$



Next lecture

- Possibility theory
- Examples
- Properties
- ...