SETS OF PROBABILITY MEASURES AND CONVEX COMBINATION SPACES

Miriam Alonso de la Fuente
Pedro Terán

Department of Statistics and Operations Research and Mathematics Didactics
University of Oviedo

Thursday, 13th

Universidad de Oviedo

Statistical Methods with Imprecise Random Elements + COmparison of DIstributions of Random Elements

- Random sets and fuzzy random variables.
- Statistical analysis of fuzzy and interval-valued data.
- Stochastic ordering of random elements.

Universidad de Oviedo

Convergence in distribution of fuzzy random variables.

- Skorokhod representation theorem (Alonso de la Fuente and Terán (2022)).
- Vitali convergence theorem (Alonso de la Fuente and Terán (2022, 2023)).
- Dominated convergence theorem (Alonso de la Fuente and Terán (2022, 2023)).
- Continuous mapping theorem (Alonso de la Fuente and Terán (2022)).

And these results can be extended to more general spaces, such as...

Universidad de Oviedo

Definition (Terán and Molchanov (2006))

Let $(\mathbb{E}, d)$ be a metric space with a *convex combination operation* $[\cdot,\cdot]$ which for any $n \geq 2$ numbers $\lambda_1, \ldots \lambda_n > 0$ satisfying $\sum_{i=1}^{n} \lambda_i = 1$ and any $v_1, \ldots, v_n \in \mathbb{E}$ this operation produces an element of $\mathbb{E}$, denoted $[\lambda_i, v_i]_{i=1}^{n}$ or $[\lambda_1, v_1; \ldots; \lambda_n, v_n]$. We will say that $\mathbb{E}$ is a *convex combination space* if the following axioms are satisfied:

(CC1) (Commutativity) For every permutation σ of $\{1, \ldots, n\}$, $[\lambda_i, v_i]_{i=1}^{n} = [\lambda_{\sigma(i)}, v_{\sigma(i)}]_{i=1}^{n}$,

(CC2) (Associativity) $[\lambda_i, v_i]_{i=1}^{n+2} = [\lambda_1, v_1; \ldots, \lambda_n, v_n; \lambda_{n+1} + \lambda_{n+2}, [\frac{\lambda_{n+j}}{\lambda_{n+1}+\lambda_{n+2}}; v_{n+j}]_{j=1}^{2}]$;

(CC3) (Continuity) If $u, v \in \mathbb{E}$ and $\lambda^{(k)} \to \lambda \in (0,1)$, then $[\lambda^{(k)}, u; 1-\lambda^{(k)}, v] \to [\lambda, u; 1-\lambda, v]$;

(CC4) (Negative curvature) For all $u_1, u_2, v_1, v_2 \in \mathbb{E}$ and $\lambda \in (0,1)$,

$$d([\lambda, u_1; 1-\lambda, u_2], [\lambda, v_1; 1-\lambda, v_2]) \leq \lambda d(u_1, v_1) + (1-\lambda) d(u_2, v_2);$$

(CC5) (Convexification) For each $v \in \mathbb{E}$, there exists $\lim_{n\to\infty}[n^{-1}, v]_{i=1}^{n}$, which will be denoted by $\mathbf{K}_{\mathbb{E}}(v)$.

Universidad de Oviedo

- Banach spaces (Terán and Molchanov (2006)).
- Cumulative distribution functions (Terán and Molchanov (2006)).
- Compact convex subsets of $\mathbb{R}^d$ with the Hausdorff (Terán and Molchanov (2006)) and the Bartels-Pallaschke metric (Alonso de la Fuente and Terán (2023)).

Universidad de Oviedo

Denote by $W_1(\mathbb{R})$ the space of probability measures in $\mathbb{R}$ with finite expectation. The L^1-Wasserstein metric in $W_1(\mathbb{R})$ is defined by

$$w_1(P,Q) = \inf_{\mathcal{L}(X)=P,\mathcal{L}(Y)=Q} \|X-Y\|_1 = \inf_{\mathcal{L}(X)=P,\mathcal{L}(Y)=Q} E[|X-Y|].$$

Definition (Li and Lin (2017))

Let $\mathcal{P},\mathcal{Q}$ be sets of probability measures. Then the *generalized Wasserstein metric* between $\mathcal{P}$ and $\mathcal{Q}$ is

$$\mathcal{W}_1(\mathcal{P},\mathcal{Q}) = \max\{\sup_{P\in\mathcal{P}} \inf_{Q\in\mathcal{Q}} w_1(P,Q), \sup_{Q\in\mathcal{Q}} \inf_{P\in\mathcal{P}} w_1(P,Q)\}.$$

Universidad de Oviedo

Let $\mathcal{P}$ be a set of probability measures and $\varphi : \Omega \to \mathbb{R}$ a continuous function. A *sublinear expectation* is defined as

$$\mathbb{E}^{\mathcal{P}}[\varphi] = \sup_{\mu \in \mathcal{P}} E_\mu[\varphi].$$

Definition

We denote by $\mathcal{P}_1(\mathbb{R})$ the set of all sets of probability measures in the real line such that

- $\mathcal{P}$ is weakly compact
- For an arbitrary point $r \in \mathbb{R}$,

$$\lim_{K \to \infty} \mathbb{E}^{\mathcal{P}}[d(r,\cdot) I_{\{x \in \mathbb{R} : d(r,x) \geq K\}}] = 0.$$

Universidad de Oviedo

Definition (Li and Lin (2017))

We denote by $\mathcal{P}_1^c(\mathbb{R})$ the set of all sets of probability measures in the real line such that

- $\mathcal{P}$ is weakly compact and convex
- For an arbitrary point $r \in \mathbb{R}$,

$$\lim_{K\to\infty} \mathbb{E}^{\mathcal{P}}[d(r,\cdot)I_{x\in\mathbb{R}:d(r,x)\geq K}] = 0.$$

Universidad de Oviedo

1. $(W_1(\mathbb{R}), w_1)$ is a convex combination space.
2. $\mathcal{P}_1(\mathbb{R}) = \mathcal{K}(W_1(\mathbb{R}))$.
3. $(\mathcal{P}_1(\mathbb{R}), \mathcal{W}_1)$ is a convex combination space.
4. $(\mathcal{P}_1^c(\mathbb{R}), \mathcal{W}_1)$ is a convex combination space.

Universidad de Oviedo

Steps 1 & 2

Theorem

$(W_1(\mathbb{R}), w_1)$ *is a convex combination space with the convex combination operation*

$$[\lambda_i, P_i] = \mathcal{L}(\sum_{i=1}^{n} \lambda_i X_i),$$

where $\mathcal{L}(X_i) = P_i$ *and* X_i *are independent. The convexification operator is*

$$\mathbf{K}_{W_1(\mathbb{R})}(P) = \delta_{b(P)}.$$

Proposition

$\mathcal{P}_1(\mathbb{R}) = \mathcal{K}(W_1(\mathbb{R}))$.

Universidad de Oviedo

Theorem

$(\mathcal{P}_1(\mathbb{R}), \mathcal{W}_1)$ is a convex combination space with the convex combination operation

$$[\lambda_i, \mathcal{P}_i] = \{\mathcal{L}(\sum_{i=1}^{n} \lambda_i X_i) : \mathcal{L}(X_i) \in \mathcal{P}_i, X_i \text{ independent}\}.$$

The convexification operator is

$$\mathbf{K}_{\mathcal{K}(W_1(\mathbb{R}))} = \overline{co} \circ \mathbf{K}_{W_1(\mathbb{R})}.$$

Theorem

$(\mathcal{P}_1^c(\mathbb{R}), \mathcal{W}_1)$ is a convex combination space with the operations inherited from $\mathcal{P}_1(\mathbb{R})$.

Universidad de Oviedo

Convexification property

Given $\mathcal{P} \in \mathcal{P}_1(\mathbb{R})$ we define

- $\underline{b}(\mathcal{P}) = \inf_{P \in \mathcal{P}} b(P) = \inf_{\mathcal{L}(X) \in \mathcal{P}} E[X]$ (lower barycenter),
- $\overline{b}(\mathcal{P}) = \sup_{P \in \mathcal{P}} b(P) = \sup_{\mathcal{L}(X) \in \mathcal{P}} E[X]$ (upper barycenter).

Proposition

Let $\mathcal{P} \in \mathcal{P}_1(\mathbb{R})$. Then

$$\mathbf{K}_{\mathcal{P}_1(\mathbb{R})}(\mathcal{P}) = \{\delta_x : x \in [\underline{b}(\mathcal{P}), \overline{b}(\mathcal{P})]\}.$$

Theorem

Let $\mathcal{P} \in \mathcal{P}_1(\mathbb{R})$. Then

$$d(\mathcal{L}(n^{-1}\sum_{i=1}^{n} X_i), \{\delta_x : x \in [\underline{b}(\mathcal{P}), \overline{b}(\mathcal{P})]\}) \to 0.$$

Universidad de Oviedo

Consequences

Theorem (Strong law of large numbers)

Let Γ be an integrable random element of $\mathcal{P}_1(\mathbb{R})$. Let $\{\Gamma_n\}_n$ be pairwise independent random elements of $\mathcal{P}_1(\mathbb{R})$ distributed as Γ. Then

$$\mathcal{W}_1([n^{-1};\Gamma_i]_{i=1}^n, E[\Gamma]) \to 0$$

almost surely.

Theorem (Dominated convergence theorem)

Let Γ_n, Γ be random elements of $\mathcal{P}_1(\mathbb{R})$ such that

$$\mathcal{W}_1(\Gamma_n, \{\delta_0\}) \leq g$$

for some $g \in L^1(\Omega, \mathcal{A}, P)$. If $\Gamma_n \to \Gamma$ weakly then

$$\mathcal{W}_1(E[\Gamma_n], E[\Gamma]) \to 0.$$

Universidad de Oviedo

Theorem (Jensen's inequality)

Let φ be a lower semicontinuous function, i.e.,

$$\mathcal{W}_1(Q_n, Q) \to 0 \Rightarrow \liminf_n \varphi(Q_n) \geq \varphi(Q),$$

and midpoint convex, i.e., such that

$$\varphi([1/2, \mathcal{P}; 1/2, Q]) \leq \frac{\varphi(\mathcal{P}) + \varphi(Q)}{2}$$

for all $\mathcal{P}, Q \in \mathcal{P}_1(\mathbb{R})$. Let Γ be an integrable random element of $\mathcal{P}_1(\mathbb{R})$ such that $E(\varphi(\Gamma)) < \infty$. Then

$$\varphi(E(\Gamma)) \leq E(\varphi(\Gamma)).$$

Universidad de Oviedo

References

- M. Alonso de la Fuente, P. Terán (2022). Some results on convergence and distributions of fuzzy random variables. Fuzzy Sets and Systems 435, 149–163.
- M. Alonso de la Fuente, P. Terán (2023). Convergence theorems for random elements in convex combination spaces. Fuzzy Sets and Systems 458, 69–93.
- P. C. Álvarez-Esteban, E. del Barrio, J. A. Cuesta-Albertos, C. Matrán (2008). Trimmed comparison of distributions. J. Amer. Statist. Assoc. 103, 697–704.
- L. Ambrosio, N. Gigli, G. Savare (2005). Gradient flows in metric spaces and in the Wasserstein space of probability measures. Birkhäuser.
- Y. Chen, Z. Lin, H. Müller. Wasserstein regression. J. Amer. Statist. Assoc., to appear.
- P. M. Esfahani, D. Kuhn (2018). Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. Math. Programming 171, 115–166.

Universidad de Oviedo

References

- O. Kallenberg (2002). Foundations of modern probability. Second edition, Springer.
- Y. Kim, B. Pass (2017). Wasserstein barycenters over Riemannian manifolds. Adv. Math. 307, 640–683.
- X. Li and Y. Lin (2017). Generalized Wasserstein distance and weak convergence of sublinear expectations. J. Theoret. Probab. 30, 581–593.
- V. M. Panaretos, Y. Zemel (2020). An invitation to Statistics in Wasserstein space. Springer.
- S. Peng (2019). Nonlinear expectations and stochastic calculus under uncertainity. Springer.

Universidad de Oviedo

References

- P. Terán (2014). Jensen's inequality for random elements in metric spaces and some applications. J. Math. Anal. Appl. 414, 756–766.
- P. Terán, I. Molchanov (2006). The law of large numbers in a metric space with a convex combination operation. J. Theoret. Probab. 9, 875–898.
- C. Villani (2003). Topics in Optimal Transport. American Mathematical Society.
- C. Villani (2008). Optimal Transport: Old and New. Springer.
- Y. Zhao, X. Zhang (2019). Calculating spatial configurational entropy of a landscape mosaic based on the Wasserstein metric. Landscape Ecol. 34, 1849–1858.

Universidad de Oviedo