Quantile constructions for complete lattices

With an example of measuring extremeness/outlyingness of opinions

Georg Schollmeyer

Department of Statistics, LMU Munich

Background

- There are many multivariate generalizations of the notion of quantiles
- For most generalizations a geometrical understanding of $\mathbb{R}^d$ is underlying
- Thus, especially affine equivariance is an important issue
- However, there are some situations, where a purely order-theoretic understanding of the underlying space is also reasonable (and a geometrical understanding is only partially reasonable)

Order-theoretic data analysis

Examples where an order theoretic data-analysis could be useful are:

- multivariate poverty-measurement
- analysis of commonalities and differences among voting profiles in social choice theory
- **measurement of 'extremeness' of opinions (later)**
- analysis of distribution-function-valued data (e.g., distribution of age in different households)
- idempotent descriptive analysis of Rasch-type data
- generally: statistical analysis of formal concepts in formal concept analysis (FCA, introduced by Wille in 1984)

Posets and Complete Lattices

A complete lattice $\mathbb{L} = (L, \leq)$ is a **partially** ordered set (i.e., a set L with a reflexive, transitive and antisymmetric relation $\leq$) for which every arbitrary subset $S \subseteq L$ has

- a **least upper bound** (called **supremum**, denoted by $\bigvee S$)
- and a **greatest lower bound** (called **infimum**, denoted by $\bigwedge S$).

Probability measures on complete lattices

Definition

Let $(\mathbb{L}, \mathcal{F}, m)$ be a probability space where $\mathbb{L}$ is a complete lattice. For a given $x \in \mathbb{L}$ the set

$$\downarrow x := \{y \in \mathbb{L} \mid y \leq x\}$$

is called principal ideal generated by x. Assume that all principal ideals are measurable. Then with

$$B_m : \mathcal{F} \longrightarrow \mathbb{R} : x \mapsto m(\downarrow x)$$

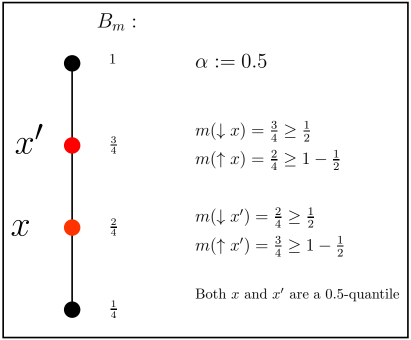
we denote the corresponding belief function.
(If not all principal ideals are measurable we can take $B_m(x) = m_*(\downarrow x)$, where m_* is the inner measure associated to m.)

Usual definition of quantiles in $(\mathbb{R}, \leq)$

Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$ be a probability space. Then every $x \in \mathbb{R}$ with

$$B_m(x) = m(\downarrow x) \geq \alpha \quad \text{and}$$
$$m(\uparrow x) \geq 1 - \alpha$$

is called an α-quantile.

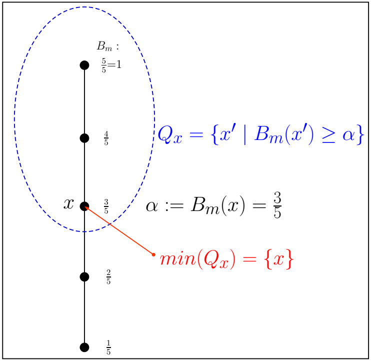


Asymmetric Definition of (lower) quantiles in $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$

Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$ be a probability space. Then every $x \in \mathbb{R}$ which is a minimal element in the set

$$Q_x := \{x' \in \mathbb{R} \mid \underbrace{m(\downarrow x')}_{=B_m(x)} \geq \underbrace{m(\downarrow x)}_{=B_m(x)=:\alpha} \}$$

is called a **lower quantile** of level $\alpha = B_m(x) = m(\downarrow x)$.



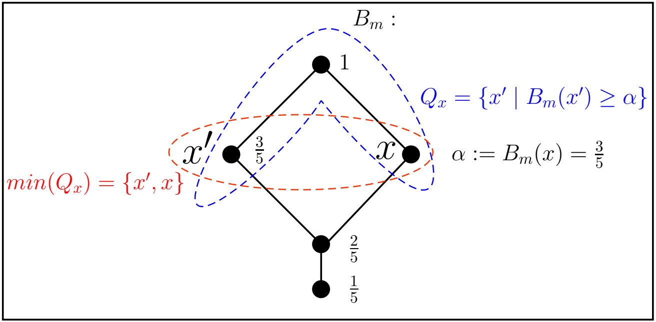
Properties of lower quantiles in $(\mathbb{R}, \mathbb{B}(\mathbb{R}), m)$

1. For every $\alpha \in \operatorname{Im}(B_m)$ there exists exactly one lower quantile of level α, denoted by q_α.
2. For every $\alpha, \alpha' \in \operatorname{Im}(B_m)$ we have

$$\alpha \leq \alpha' \Longrightarrow q_\alpha \leq q_{\alpha'}$$

and in particular, different quantiles are always comparable.

Let $(\mathbb{L}, \mathcal{F}, m)$ be a probability space where $\mathbb{L}$ is a complete lattice (and all principal ideals are measurable). Then every $x \in \mathbb{L}$ which is a minimal element in the set $Q_x = \{x' \in \mathbb{L} \mid B_m(x') \geq B_m(x)\}$ is called a **lower quantile** of level $\alpha = B_m(x) = m(\downarrow x)$.



Quantile system for a probability measure

Definition

Let $(\mathbb{L}, \mathcal{F}, m, B_m)$ *be given. Define the system*

$$\mathfrak{Q} = \left\{ \bigwedge Q_x \mid x \in \mathbb{L} \right\}$$

as the system of lower (pre-)quantiles associated to $(\mathbb{L}, \mathcal{F}, m, B_m)$.

Construction of a qualitative and a quantitative map

Definition

Define the map Φ *as*

$$\Phi : \mathbb{L} \longrightarrow \mathbb{L} : x \mapsto \bigwedge \{q \in \mathfrak{Q} \mid q \geq x\}.$$

The mapping Φ *can be understood as a qualitative measure of location.*
(Φ *is a so-called closure operator.)*

Definition

Define the **level function** λ *as*

$$\lambda : \mathbb{L} \longrightarrow \mathbb{R} : x \mapsto B_m(\Phi(x)).$$

This mapping could be understood as a quantitative measure of location. Because (in regular cases) λ *satisfies*

$$\forall x \in \mathbb{L}, \alpha \in [0,1] : \lambda(x) \leq \alpha \iff x \leq \underbrace{\max \lambda^{-1}(\downarrow \alpha)}_{\in \mathbb{L}}$$

the location measure λ *has a qualitative representation in* $\mathbb{L}$.

Example

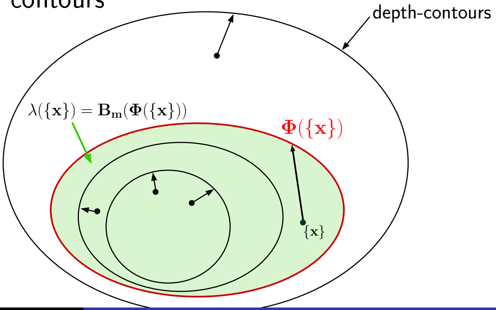
Let $\mathbb{L} :=$ be the complete lattice of all compact convex sets in $\mathbb{R}^d$ ordered by set inclusion $\subseteq$. Then

$$\bigwedge M \;\widehat{=}\; \bigcap M \quad ; \quad \bigvee M \;\widehat{=}\; \overline{co}\left(\bigcup M\right).$$

If we treat $\mathbb{R}^d$-valued random variables as one-point set-valued variables in $\mathbb{L}$ then we get essentially Tukey's half-space depth:

level function $\lambda \;\widehat{=}\;$ (a transformation of) Tukeys outlyingness function

$\mathfrak{Q} \;\widehat{=}\;$ depth contours



Example: Analysis of different attitudes w.r.t. different principles of justice

- Allbus 2014: 8 Questions about 4 different principles of justice (c.f., Liebig, S., May, M., (2009): Dimensionen sozialer Gerechtigkeit)
  1. merit principle ('Leistungsprinzip')
  2. principle of equality ('Gleichheitsprinzip')
  3. right principle ('Anrechtsprinzip')
  4. demand principle ('Bedarfsprinzip')
- e.g.,: statement S1: *'It is fair when those, who perform well at work, earn more than others'* (transl. G.S.)
- agreement measured on a scale from 1: 'full agreement' to 5: 'full disagreement',
- simplified here to 2 binary variables: agreement (yes/no), disagreement (yes/no)

person i

agreement

			1	2	...	n
merit principle	S1	fair: better performance, higher merit	x	x		
	S2	fair: get only what one has achieved through own efforts		x		x
principle of equality	S3	fair: equal living conditions	x	x		
	S4	fair: equally distributed income				
right principle	S5	fair: advantages due to origin				
	S6	fair: social superiors have better living conditions				
demand principle	S7	fair: take care of the weak	x	x		x
	S8	fair: support carers	x	x		

disagreement

			1	2	...	n
merit principle	S1	fair: better performance, higher merit				
	S2	fair: get only what one has achieved through own efforts				
principle of equality	S3	fair: equal living conditions				
	S4	fair: equally distributed income	x	x		x
right principle	S5	fair: advantages due to origin		x		
	S6	fair: social superiors have better living conditions		x		x
demand principle	S7	fair: take care of the weak				
	S8	fair: support carers				

Definition of $\leq$

For a person p define the corresponding vector $x \in \mathbb{L} := \{0,1\}^{16}$ with $x_i = 1 \iff$ person p has a cross in column i and thus agrees (disagrees) to the corresponding statement.
Define the relation $\leq$ as

$$p \leq q \iff \forall i \in \{1, \ldots, 16\} : p_i \leq q_i.$$

Thus person p is lower than or equal to person q if person q has at least a cross in every row in which person p has a cross:

agreement

			Quantiles 0,3 %	0,49 %	0,85 %	1,76 %	"mean opinion" 13,3 %	"median opinion" 49,45%	77%	91,85%
merit principle	S1	fair: better performance, higher merit		x	x	x	x	x	x	x
	S2	fair: get only what one has achieved through own efforts				x	x	x	x	x
principle of equality	S3	fair: equal living conditions						x	x	x
	S4	fair: equally distributed income							x	x
right principle	S5	fair: advantages due to origin								
	S6	fair: social superiors have better living conditions							x	x
demand principle	S7	fair: take care of the weak	x	x	x	x	x	x	x	x
	S8	fair: support carers	x	x	x	x	x	x	x	x

disagreement

			0,3 %	0,49 %	0,85 %	1,76 %	13,3 %	49,45%	77%	91,85%
merit principle	S1	fair: better performance, higher merit								x
	S2	fair: get only what one has achieved through own efforts								
principle of equality	S3	fair: equal living conditions							x	x
	S4	fair: equally distributed income					x	x	x	x
right principle	S5	fair: advantages due to origin			x	x	x	x	x	x
	S6	fair: social superiors have better living conditions					x	x	x	x
demand principle	S7	fair: take care of the weak								
	S8	fair: support carers								

- The higher the quantile, the more diverse is the subpopulation that lies below this quantile and is described by this quantile.
- For a given opinion profile $x \in \mathbb{L}$ we have: The higher the smallest quantile $\Phi(x)$ that is still above x the more disperse is the most specific subpopulation that contains x and is summarized by this quantile $\Phi(x)$.
- Thus higher $\Phi(x)$ or higher values of $\lambda(x) = B_m(\Phi(x))$ indicate a 'more outlying' opinion x.
- Thus we have some method of 'measuring' the outlyingness of opinions.

'Proof of concept'

- There is an (small, but statistically significant) association between the level λ and for example the political self evaluation of people on a left-right scale
- For this example, the approach could also be understood as some ordinal approach for measuring a latent concept (here: extremeness/outlyingness of opinions)
- The proposed approach has the advantage of being insensitive to adding redundant items (this would not be the case e.g., for a Rasch-type modeling that assumes locally stochastic independent items)

And finally:

- There are many other situations where ordinal data analysis with the introduced concepts are useful:
  - multivariate poverty-measurement
  - analysis of commonalities and differences among voting profiles in social choice theory
  - analysis of distribution-function-valued data (e.g., distribution of age in different households)
  - idempotent descriptive analysis of Rasch-type data
  - generally: statistical analysis of formal concepts in **formal concept analysis** (FCA, introduced by Rudolf Wille in 1984)

- Stefan Liebig and Meike May.
  Dimensionen sozialer Gerechtigkeit.
  Aus Politik und Zeitgeschichte, 59(47), 2009.

- GESIS - Leibniz - Institut fur Sozialwissenschaften.
  Allbus compact (2015): Allgemeine Bevölkerungsumfrage der Sozialwissenschaften.
  GESIS Datenarchiv, Köln. ZA5241 Datenfile Version 1.1.0,.

- Bernhard Ganter and Rudolf Wille.
  Formale Begriffsanalyse: Mathematische Grundlagen.
  Springer-Verlag, 2013.

- John W Tukey.
  Mathematics and the picturing of data.
  In *Proceedings of the international congress of mathematicians*, volume 2, pages 523–531, 1975.