

Constriction for Sets of Probabilities

Michele Caprio and Teddy Seidenfeld

PRECISE Center, Dept. of Computer and Information Science, University of Pennsylvania
Depts. of Philosophy and Statistics, Carnegie Mellon University

ISIPTA 2020

11-14 JULY, OVIEDO (ASTURIAS) - SPAIN



Universidad de Oviedo

Motivation

- **Bayes' rule** $P(A | E) = \frac{P(E|A)P(A)}{P(E)}$ is arguably the most common updating rule for subjective beliefs
- Its IP version, called **generalized Bayes' rule** is given by

$$\underline{P}(A | E) = \inf_{P \in \mathcal{P}} \frac{P(E | A)P(A)}{P(E)} \quad \text{and} \quad \bar{P}(A | E) = \sup_{P \in \mathcal{P}} \frac{P(E | A)P(A)}{P(E)}$$

Motivation

- **Bayes' rule** $P(A | E) = \frac{P(E|A)P(A)}{P(E)}$ is arguably the most common updating rule for subjective beliefs
- Its IP version, called **generalized Bayes' rule** is given by

$$\underline{P}(A | E) = \inf_{P \in \mathcal{P}} \frac{P(E | A)P(A)}{P(E)} \quad \text{and} \quad \bar{P}(A | E) = \sup_{P \in \mathcal{P}} \frac{P(E | A)P(A)}{P(E)}$$

- “When we observe new evidence in the form of a partition \mathcal{E} , can we have that $[\underline{P}(A | E), \bar{P}(A | E)] \subset [\underline{P}(A), \bar{P}(A)]$, for all $E \in \mathcal{E}$?”

Motivation

- Not necessarily! For example, the inclusion **does not hold**
 - if \mathcal{E} is a **finite partition**

¹except for Density Ratio neighborhoods

Motivation

- Not necessarily! For example, the inclusion **does not hold**
 - if \mathcal{E} is a **finite partition**
 - if the IP interval is **closed** and (precise) probabilities are **countably additive**

¹except for Density Ratio neighborhoods

Motivation

- Not necessarily! For example, the inclusion **does not hold**
 - if \mathcal{E} is a **finite partition**
 - if the IP interval is **closed** and (precise) probabilities are **countably additive**
- With generalized Bayes' updating
 - Dilation – i.e. the opposite inclusion – is commonplace for IP neighborhood models¹
 - Constriction is possible only in specialized settings

¹except for Density Ratio neighborhoods

Motivation

- Not necessarily! For example, the inclusion **does not hold**
 - if \mathcal{E} is a **finite partition**
 - if the IP interval is **closed** and (precise) probabilities are **countably additive**
- With generalized Bayes' updating
 - Dilation – i.e. the opposite inclusion – is commonplace for IP neighborhood models¹
 - Constriction is possible only in specialized settings
- “That is odd and fascinating! Can you tell us more about those settings?”
- That's the question we explore in this paper

¹except for Density Ratio neighborhoods

Three cases that allow for constriction

- No new evidence is collected

Three cases that allow for constriction

- No new evidence is collected
- Convex pooling of different opinions

Three cases that allow for constriction

- No new evidence is collected
- Convex pooling of different opinions
- Non-Bayesian updating

No new evidence

- If a generic procedure generates an IP set \mathcal{P} and prescribes a way of selecting one element P^* from $\text{Conv}(\mathcal{P})$, we give sufficient conditions for constriction

- If a generic procedure generates an IP set \mathcal{P} and prescribes a way of selecting one element P^* from $\text{Conv}(\mathcal{P})$, we give sufficient conditions for constriction
- Examples
 - de Finetti's Fundamental Theorem of Probability (de Finetti, 1974, Section 3.10)
 - Halmos' Extension (Halmos, 1950, Exercise 48.4), (Billingsley, 1995, Section 4.13)
 - Jaynes' MaxEnt Kesavan (2008)
 - Generalized Fiducial Inference Hannig et al. (2016)

- DeGroot consensus model [DeGroot \(1974\)](#)
- k individuals, each having their own subjective probability distribution
- For agent i , the opinions of all the other $k - 1$ agents represent new evidence

- DeGroot consensus model [DeGroot \(1974\)](#)
- k individuals, each having their own subjective probability distribution
- For agent i , the opinions of all the other $k - 1$ agents represent new evidence
- Instead of conditioning on those, agent i pools their own opinion with that of the other agents
- Repeating this process for all agent i , the group reaches (asymptotically) an agreement on a common subjective probability distribution

- We give conditions for constricting when the agent endorses
 - Intentional forgetting [Golding and MacLeod \(1998\)](#)

Non-Bayesian updating

- We give conditions for constricting when the agent endorses
 - Intentional forgetting [Golding and MacLeod \(1998\)](#)
 - Levi-neutrality [Levi \(2009\)](#)

Non-Bayesian updating

- We give conditions for constricting when the agent endorses
 - Intentional forgetting [Golding and MacLeod \(1998\)](#)
 - Levi-neutrality [Levi \(2009\)](#)
 - Geometric updating rule [Gong and Meng \(2021\)](#)

- We give conditions for constricting when the agent endorses
 - Intentional forgetting [Golding and MacLeod \(1998\)](#)
 - Levi-neutrality [Levi \(2009\)](#)
 - Geometric updating rule [Gong and Meng \(2021\)](#)
 - Dempster's updating rule [Dempster \(1967\)](#)

- We give conditions for constricting when the agent endorses
 - Intentional forgetting [Golding and MacLeod \(1998\)](#)
 - Levi-neutrality [Levi \(2009\)](#)
 - Geometric updating rule [Gong and Meng \(2021\)](#)
 - Dempster's updating rule [Dempster \(1967\)](#)
 - Gärdenfors' generalized imaging updating rule [Gärdenfors \(1988\)](#)

COME TO THE POSTER FOR MORE!



Constriction for Sets of Probabilities
Michele Caprio & Teddy Seidenfeld
PRECISE Center, Dept. of Computer and Information Science, University of Pennsylvania
Dept. of Philosophy and Statistics, Carnegie Mellon University
caprio@seas.upenn.edu, teddy@stat.cmu.edu



1. What does it mean to constrain?

Consider the set of probability distributions representing an agent's knowledge on the selection of a sign (upper U , lower L) on a coin flip. An experimenter generates a new measurement of heads to be used to predict the choice of the probability of either the procedure are compared to those before the procedure. An experimenter's knowledge is updated (generalized). Lower credences place our constraints on the probabilities. The set of probabilities will still be well-ordered and without evidence being obtained, and we characterize these possibilities.

2. Bayes updating does not constrict

Consider a signal of interest X if a given set of probability measures $\mathcal{P} = \{P_1, \dots, P_n\}$ and denote by U and L the lower and upper probabilities associated with \mathcal{P} , respectively. Call \mathcal{P} a general probability that produces a bias assessment of heads, and denote by \mathcal{P}^* any of the lower and upper probabilities resulting from such constraints, respectively. Then, we say that constraints (generalization) is $\mathcal{P}^* \subseteq \mathcal{P}$ if and only if $P_i^* \subseteq P_i$. We say that a nearly constraint is if one of the two inequalities is strict.



3. Constricting without new evidence

Theorem 3.1 Suppose a general procedure generates a set $\mathcal{P} = \{P_1, \dots, P_n\}$ of probabilities on $\{U, L\}$ such that $\mathcal{P}^* \subseteq \mathcal{P}$, and this procedure is a way of extending the set \mathcal{P}^* to \mathcal{P} . Then, \mathcal{P}^* is a general procedure. Assume that \mathcal{P}^* is $\mathcal{P}^* \subseteq \mathcal{P}$, while \mathcal{P} is a general order and that we have that:

- $\mathcal{P}^* = \{P_1^*, \dots, P_n^*\}$ has non-trivial credences $\{P_i^*(U), P_i^*(L)\}$ for all $i \in \{1, \dots, n\}$ and $P_i^*(U) > P_i^*(L)$.
- In addition, $\mathcal{P}^* \subseteq \mathcal{P}$, for all $i \in \{1, \dots, n\}$.
- $\mathcal{P}^* \subseteq \mathcal{P}$ is a nearly constraint.

Example: In the *Philosophical Theoria of Probability 7*, Section 11.18, "Narrow" is a theorem 3.1 above (4.4), where Seidenfeld's generalization is not necessary.

4. Constricting based on common priors

Consider two agents who share a common prior, each having been independently distributed P on the common value of some parameter $\theta \in \Theta$. For agents 1 and 2, the signals of the other 1 - agents represent new evidence. Instead of conditioning on their agent's own data (which is not relevant to that of the other agent), the options are mutually representing the same information that agent 1 sends to the other agent. The signals of the prior, instead of being relevant to the prior of all agents, it requires (conditionally) an agreement on a common sufficient probability distribution. Theorem 4.1 from Seidenfeld (1992) says that every agent in a set of two (or more) agents \mathcal{P} has a common prior P .

Example: In Seidenfeld (1992), Section 4.1, "The Signal" we have that $P(U, L) = P(U, U) = P(L, U) = P(L, L) = 1/4$. This means that the list of sequences $\{U, L\}^n$ is not \mathcal{P} , if the condition in Theorem 4.1 is satisfied. Here, $\mathcal{P} = \{P_1, P_2\}$. Consider how we can build.



5. Constricting based on non-Bayesian updating

If the agent's update rule to event A , given evidence E , makes sense, then we say it is correct to "bet" β and to agree to receive to the update rule in order to have the sign of the update. The agent's update rule is correct if the update rule is correct. Definition 5.1 Let $\mathcal{P} = \{P_1, \dots, P_n\}$ be a set of probabilities on $\{U, L\}$ and let $\mathcal{P}^* = \{P_1^*, \dots, P_n^*\}$ be a set of probabilities on $\{U, L\}$ such that $\mathcal{P}^* \subseteq \mathcal{P}$. The same situation occurs if we require that the condition in generalized Bayes rule holds for all $i \in \{1, \dots, n\}$ and $P_i^* \subseteq P_i$.

Definition 5.1 Let $\mathcal{P} = \{P_1, \dots, P_n\}$ be a belief function. Then the LPs and UPs associated in Seidenfeld (1992) with \mathcal{P} are not correct, $\mathcal{P}^* \subseteq \mathcal{P}$ and that $\mathcal{P}^* \subseteq \mathcal{P}$ if and only if $P_i^*(U) > P_i^*(L)$ and $P_i^*(U) > P_i^*(L)$ for all $i \in \{1, \dots, n\}$. In the belief function, assume that \mathcal{P} is a belief function having mean function μ . Consider a function $f: \mathcal{P} \rightarrow \mathcal{P}$ having constants $\mu(U) = \mu(L) = 1/2$ and $\mu(U) = \mu(L) = 1/2$ for all $i \in \{1, \dots, n\}$. For all $i \in \{1, \dots, n\}$, then the belief function $P_i^*(U) = P_i^*(L)$ obtained according to Seidenfeld's generalized updating rule, and $\mathcal{P}^* \subseteq \mathcal{P}$ is a belief function. Then $\mathcal{P}^* \subseteq \mathcal{P}$ is a belief function.

$$\mu^*(U) = \sum_{i=1}^n P_i^*(U) \mu(U) = \sum_{i=1}^n P_i^*(L) \mu(L) = \mu(L)$$

$$\mu^*(L) = \sum_{i=1}^n P_i^*(L) \mu(L) = \sum_{i=1}^n P_i^*(U) \mu(U) = \mu(U)$$

Theorem 5.2 Let $\mathcal{P} = \{P_1, \dots, P_n\}$ be the partition associated with the partition of the space of events. Assume that \mathcal{P} is a belief function such that $\mathcal{P}^* \subseteq \mathcal{P}$, and consider any event $A \in \mathcal{P}$. Then, if \mathcal{P}^* is a belief function, then $\mathcal{P}^* \subseteq \mathcal{P}$ is a belief function. Similarly, if \mathcal{P}^* is a belief function, then $\mathcal{P}^* \subseteq \mathcal{P}$ is a belief function.

Theorem 6 Assume that \mathcal{P} is a belief function. Let \mathcal{P}^* be the collection of probabilities on $\{U, L\}$ and assume that \mathcal{P}^* is a belief function having mean function μ such that $\mu(U) = \mu(L) = 1/2$. Consider any event $A \in \mathcal{P}$. We have that $\mathcal{P}^* \subseteq \mathcal{P}$ if and only if $\mathcal{P}^* \subseteq \mathcal{P}$ is a belief function. Similarly, if \mathcal{P}^* is a belief function, then $\mathcal{P}^* \subseteq \mathcal{P}$ is a belief function.

6. Conclusion

Do not change your belief function more than is necessary.

A generalization of Bayesian Inference



PRECISE Center, Dept. of Computer and Information Science, University of Pennsylvania
Dept. of Philosophy and Statistics, Carnegie Mellon University



References I

- Patrick Billingsley. *Probability and Measure*. New York, NY : Wiley, second edition, 1995.
- Bruno de Finetti. *Theory of Probability*, volume 1. New York : Wiley, 1974.
- Morris H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.
- Arthur Pentland Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38(2): 325–339, 1967.
- Jonathan M. Golding and Colin M. MacLeod. *Intentional forgetting: Interdisciplinary approaches*. Mahwah, NJ : Lawrence Erlbaum Associates, 1998.
- Ruobin Gong and Xiao-Li Meng. Judicious judgment meets unsettling updating: dilation, sure loss, and Simpson's paradox. *Statistical Science*, 36(2):169–190, 2021.

References II

- Peter Gärdenfors. *Knowledge in flux: Modeling the dynamics of epistemic states*. Boston, MA : The MIT press, 1988.
- Paul R. Halmos. *Measure Theory*. Graduate Texts in Mathematics. New York, NY : Springer, 1950.
- Jan Hannig, Hari Iyer, Randy C. S. Lai, and Thomas C. M. Lee. Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association*, 111(515):1346–1361, 2016.
- Hiremagalur K. Kesavan. Jaynes' maximum entropy principle. In Christodoulos A. Floudas and Panos M. Pardalos, editors, *Encyclopedia of Optimization*. Boston, MA : Springer, 2008.
- Isaac Levi. Why indeterminate probability is rational. *Journal of Applied Logic*, 7(4):364–376, 2009.