Bristol-Groningen Summer School 2015

Social epistemology from Condorcet to Aumann

⋆

Jan-Willem Romeijn
University of Groningen

Contents

1 The wisdom of crowds

1.1 The Condorcet formula

A jury of n members is trying Jack for murder. A number n_1 vote that Jack is guilty, H^1, while the remaining n_0 members vote that he is innocent, H^0. Jurors are characterised by

$$P(V_i^j | H^j \cap V_{i'}^{j'}) = P(V_i^j | H^j) = q_j > 1/2.$$

If H^j is in fact true, the event that jury member i votes for H^j, denoted V_i^j, has some fixed chance q_j, the competence. We assume that the competences are greater than one half.

Wise juries

We can now introduce Condorcet's jury theorem. Say that H^1 is true. For an ever larger jury size n, consider the relative frequency of voters in favour,

$$f_1 = \frac{n_1}{n} = 1 - f_0.$$

By the law of large numbers, the fraction f_1 tends to q_1 with probability 1. Because $q_1 > 1/2$, we have:

> *Condorcet's jury theorem (1785)*
> On the assumption of H^1, the probability of a correct majority vote $\Delta = n_1 - n_0 > 0$ tends to 1 in the limit.

Bayes meets Condorcet

Rather than calculating the probability of a majority of votes given the truth of H^j, we might ask for the probability of the hypothesis H^j given some majority of votes:

$$P(H^1|V_{n\Delta}) = \frac{q_1^{n_1}(1-q_1)^{n_0}P(H^1)}{P(V_{n\Delta})}.$$

Here $V_{n\Delta}$ is the vote of the entire jury. With this we can derive a Bayesian version of Condorcet's theorem:

> *Inverse of Condorcet's theorem*
>
> Let the jury size n go to infinity and assume a fixed relative majority larger than ½, then the probability of H^1 will tend to 1.

Condorcet formula

Under the idealising assumptions that

- the priors of H^0 and H^1 are equal, $P(H^1) = P(H^0)$, and that
- the competences of jury members on H^0 and H^1 are equal, $q_0 = q_1 = q$,

We can now derive the following posterior odds:

$$\frac{P(H^1|V_{n\Delta})}{P(H^0|V_{n\Delta})} = \left(\frac{q}{1-q}\right)^{\Delta}.$$

It depends only on the absolute margin of the jury vote, and not on the number of jurors.

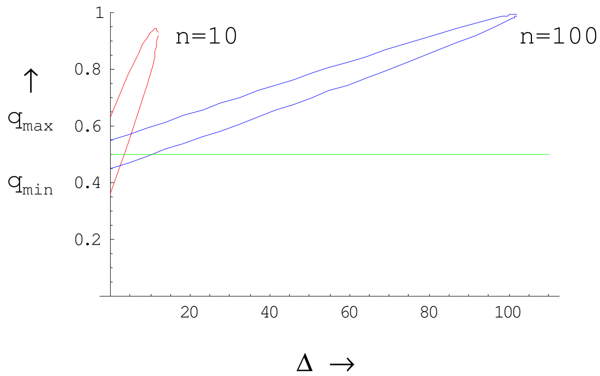
1.2 Counterintuitive consequences

List (2004) emphasises the significance of the absolute margin for jury votes. But the sole dependence on Δ is rather puzzling. Which of the following two juries do you prefer?

	Small jury	Large jury
Number of jurors	10	100
Number in favour (n_1)	10	56
Number against (n_0)	0	44
Absolute margin (Δ)	10	12

A classical statistical analysis

A 95% confidence interval for juror competence q shows that the votes are a freak accident, or otherwise that the jurors from the smaller jury are more competent.



Learning competence

The vote of the jury somehow reveals the competence of the jury, and this can be used in choosing between jury verdicts. Here we disregard a number of related viewpoints:

- Without the assumption of symmetric competence, the posterior odds do depend on the jury size. But they do not do so in the relevant way.
- A unanimous vote may well result from mindless groupthink rather than high competence. But here we assume jurors to be independent.
- As Bovens and Hartmann (2004) argue, the coherence of jurors might indicate the veracity of the jury verdict.

1.3 A model using unknown competences

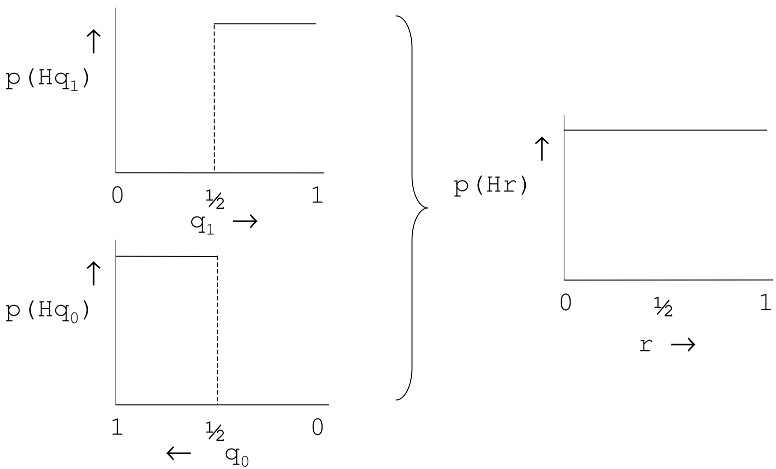
We split the hypotheses H^0 and H^1 up into H_{q_0} and H_{q_1} respectively. The hypotheses H^j each consist of a range of statistical hypotheses:

$$P(H^0) = \int_{1/2}^{1} P(H_{q_0})\, dq_0 \qquad P(H^1) = \int_{1/2}^{1} P(H_{q_1})\, dq_1\,.$$

The hypotheses H_{q_j} concern the competences $P(V_j^i | H_{q_j} \cap V_{j'}^{i'}) = q_j$. We assume that the prior is equal and uniform over $(1/2, 1)$, for both q_0 and q_1.

Transforming the problem

We can turn this into a well-known statistical problem by a suitable translation of the parameters q_j to a single $r \in [0, 1]$.



Posterior for the hypothesis

We model the impact of the jury vote on the probability assignments over q_0 and q_1 by modelling its impact on the probability assignment over r. The posterior over H_r is a Beta distribution,

$$P(H_r|V_{n\Delta}) = \frac{(n+1)!}{n_0!n_1!} r^{n_1}(1-r)^{n_0}.$$

The posterior probability of the hypotheses H^0 is:

$$P(H^0|V_{n\Delta}) = \frac{(n+1)!}{n_0!n_1!} \int_0^{1/2} r^{n_1}(1-r)^{n_0}\, dr$$

1.4 Analytic and numerical results

We retain an important consequence of the Condorcet formula. On the assumption that $\Delta = n_1 - n_0 > 0$, we have

$$P(H^1|V_{n\Delta}) > P(H^0|V_{n\Delta}).$$

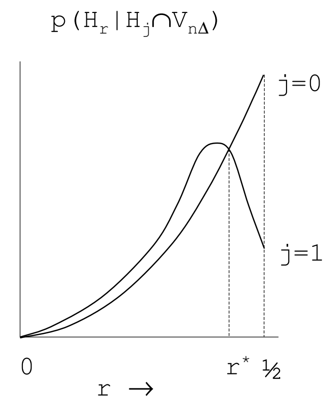
But we can also show that

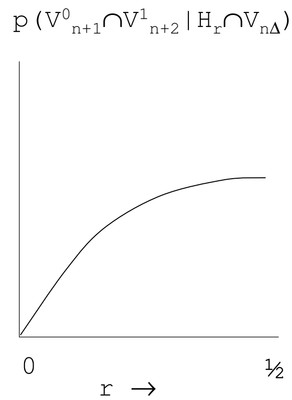
$$\frac{P(H^1|V_{n+2,\Delta})}{P(H^0|V_{n+2,\Delta})} < \frac{P(H^1|V_{n\Delta})}{P(H^0|V_{n\Delta})}.$$

This repairs the counterintuitive choice between the two juries.

Proof of inequality

Given the likelihoods $r(1-r)$ for H_r, the marginal likelihood of the hypothesis H^0 is larger because most of the mass lies close to $r = ½$.





Limiting behaviour

For constant Δ, we find the asymptotic behaviour

$$\lim_{n\to\infty} P(H^0|V_{n\Delta}) = \frac{1}{2}.$$

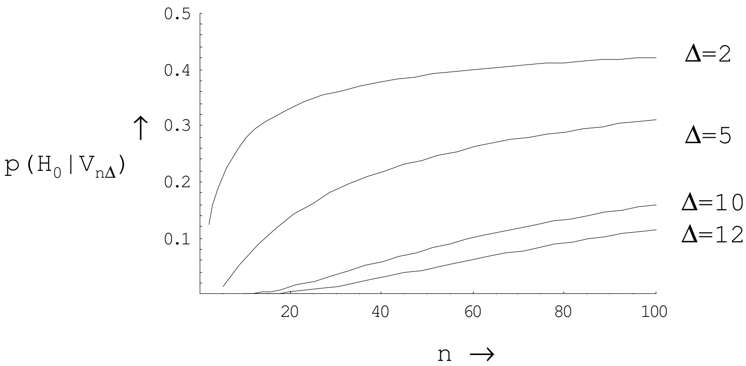
For constant fractional majority, $f = \Delta/n > 0$, we have

$$\lim_{n\to\infty} P(H^0|V_{n,nf}) = 0.$$

In fact the increase in Δ need not be linear in n. It is enough if Δ increases more quickly than $\sqrt{n}$.

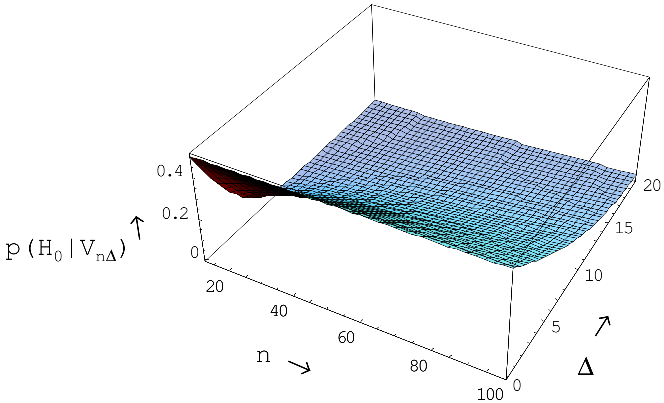
Dependence on jury size

These results are in accordance with the aforementioned intuitions on the relation between jury votes and the hypothesis voted over.



Dependence on jury size and margin

For fixed jury size *n*, the probability of H^0 decreases with increasing majority size Δ. And for fixed Δ and increasing *n*, the probability of H^0 increases towards ½.



1.5 Conclusions

In the model with competence learning we have:

- The probability that the jury majority verdict is incorrect is monotonically increasing in the jury size n, if the absolute margin Δ remains constant.

- The probability that the jury majority verdict is incorrect tends to one-half as n tends to infinity, if Δ remains constant in this limit.

- The probability that the jury majority verdict is incorrect tends to zero as n tends to infinity, if the fractional majority, $f = \Delta/n$, tends to a nonzero constant in this limit.

Important consequences

For the discussion on voting rules, two consequences of this must be given extra emphasis.

- The exclusive dependence on the absolute margin seems to be an artefact of idealising assumptions, and not something inherent to real jury verdicts.
- Both the normal Condorcet jury theorem and the converse Condorcet jury theorem for posterior odds remain valid in the new model.

Hence, against List (2004), we insist on the significance of the relative margin.

Further research

Some suggestions on how to develop the results of the present paper:

- It is important for the practical applicability of Condorcet-style results to relax the assumption on the independence of the jurors (Bovens and Hartmann 2004).
- An entirely different line of research concerns the possible variation of competences within the jury (Dietrich 2008).
- Much can be gained by applying the present insights to the discussion over the coherence measures proposed in Bovens and Hartmann (2004).

2 Conditioning, pooling, and voting

2.1 Opinion pooling and Bayesian updates

Raquel and Quassim are both pondering over the proposition A. Raquel's belief is $P_R(A) = r$, Quassim's is $P_Q(A) = q$. Pooling determines that

$$r' = wq + (1 - w)r.$$

The parameter $w \in [0, 1]$ determines how much Raquel moves towards Quassim. It measures the trust of Raquel in Quassim.

Conditionalizing on opinions

A different model for Raquel's accommodating Quassim's opinion employs Bayesian conditionalization on Quassim's opinion:

$$r' = P_R(A|\ulcorner q \urcorner \cap \ulcorner r \urcorner) = P_R(A|\ulcorner r \urcorner)\frac{P_R(\ulcorner q \urcorner|A \cap \ulcorner r \urcorner)}{P_R(\ulcorner q \urcorner|\ulcorner r \urcorner)}.$$

Here Quassim's belief that $P_Q(A) = q$ is denoted by $\ulcorner q \urcorner$, and similarly Raquel's belief is denoted $\ulcorner r \urcorner$. They are categorical events concerning probabilistic opinions.

Aumann, Bayes, Condorcet

We can provide a Bayesian model of opinion pooling. The formal link may be employed to relate pooling to other Bayesian models of epistemic interaction.

- First we focus on the interpretation of the trust parameter w, elaborating its relation to the so-called truth-conduciveness of jurors from Condorcet's theorem.

- In the next lecture we relate the representation of pooling as updating to consensus formation and disagreement among peers, in particular the agreement theorem of Aumann.

2.2 Pooling as updating

Genest and Schervish (1985) establish that we can always find likelihoods $P_R(\ulcorner q \urcorner | A \cap \ulcorner r \urcorner)$ such that, after conditionalizing on $\ulcorner q \urcorner$, Raquel's belief in A equals the result of pooling.

Corollary of Genest and Schervish (1985)
Let $P_Q(A) = q$, $P_R(A) = P_R(A|\ulcorner r \urcorner) = r$, and let $P'_R(A) = r' = wq + (1 - w)r$ be the result of linear pooling. We choose

$$P_R(\ulcorner q \urcorner | A \cap \ulcorner r \urcorner) = g(q, r)\left(1 - w + \frac{w}{r} q\right),$$

$$P_R(\ulcorner q \urcorner | \neg A \cap \ulcorner r \urcorner) = g(q, r)\left(1 + \frac{r}{1-r} w - \frac{w}{1-r} q\right).$$

Pooling as updating (continued)

Here $g(q,r) = P_R(\ulcorner q \urcorner | \ulcorner r \urcorner)$ is such that

$$\int_0^1 g(q,r)dq = 1,$$

$$\int_0^1 qg(q,r)dq = r.$$

Then the Bayesian update on $\ulcorner q \urcorner$ is identical to the update by linear pooling, $P_R(A | \ulcorner q \urcorner \cap \ulcorner r \urcorner) = r' = P'_R(A)$.

It is intuitive that the trust parameter w shows up as the skewness of the likelihood function.

Raquel's expectations about Quassim

The interpretation of the constraints on $P(\ulcorner q \urcorner | \ulcorner r \urcorner)$ is rather natural: Raquel's distribution for Quassim's opinion q is centred on r.

- Raquel might use a peaked Beta-distribution to express that she thinks Quassim will think much like herself.
- She can use a U-shaped Beta-distribution to express the idea that Quassim will be opinionated.

In all of this it is assumed that the events $\ulcorner q \urcorner$ form a partition. We can drop constraints by presuming Quassim might return a blank.

2.3 Truth-conducive voting

In Condorcet's setting, jurors are asked to vote for or against a proposition A, denoted by V and $\neg V$ respectively. It is standardly assumed that the jurors are competent:

$$c_A = P(V|A) > \frac{1}{2} > P(V|\neg A),$$

$$c_{\neg A} = P(\neg V|\neg A) > \frac{1}{2} > P(V|\neg A).$$

Under this assumption we can derive that for ever larger juries, the majority vote is ever more probable to be correct.

Truth-conducive jurors

We can drop the constraint on the absolute competence of the jurors and assume that jurors have a positive truth-conduciveness Δ:

$$P(V|A) - P(V|\neg A) = c_A + c_{\neg A} - 1 = \Delta > 0,$$

$$P(\neg V|\neg A) - P(\neg V|A) = \Delta > 0.$$

Under this assumption we can still derive a Bayesian version of the Condorcet theorem: votes for A increase the posterior of A, and votes against A decrease it.

2.4 Pooling as voting

We relate the Condorcet setting to opinion pooling, by making opinion pooling categorical and by using the Bayesian representation of pooling.

- The juror Quassim casts a categorical vote, but Raquel takes him to express a probabilistic opinion.
- The votes for or against A are captured by the events $\ulcorner q > r \urcorner$ and $\ulcorner q < r \urcorner$.
- Raquel accommodates the coarse-grained opinion by means of Bayesian conditioning, just as in voting.
- For this she uses the marginal likelihood of the events $\ulcorner q > r \urcorner$ and $\ulcorner q < r \urcorner$, as determined by the Bayesian equivalent of pooling.

A simplifying assumption

Raquel distinguishes between Quassim offering a degree of belief in A nearby the extremes, but within these two ranges assumes the distribution to be uniform:

$$P_R(\ulcorner q \urcorner | \ulcorner r \urcorner) = \begin{cases} l & \text{if } q < \epsilon r, \\ h & \text{if } q > 1 - \epsilon(1 - r), \\ 0 & \text{else.} \end{cases}$$

Solving for the constraints yields:

$$l = \frac{1 - r}{\epsilon r}, \qquad h = \frac{r}{\epsilon(1 - r)}.$$

Truth-conducive pooling

Using the afore-mentioned likelihoods and the uniform priors within $\ulcorner q > r \urcorner$ and $\ulcorner q < r \urcorner$, we can derive:

$$P_R(\ulcorner q > r \urcorner | A) - P_R(\ulcorner q > r \urcorner | \neg A) = w\left(1 - \frac{\epsilon}{2}\right),$$

and the same for $\neg A$. Now notice the similarity with the truth-conduciveness of jurors from Condorcet:

$$P(V|A) - P(V|\neg A) = \Delta.$$

Trust as truth-conduciveness

The formulas involving a trust parameter in pooling can be interpreted as expressing a form of truth-conduciveness. For diminishing ϵ we have

$$\Delta = w.$$

The main conclusion is that we can thus express the trust w intuitively and as internal to the model of beliefs.

Modelling choices

Some considerations on the modelling choices needed to relate trust to beliefs:

- The shape of the distribution $P_R(\ulcorner q \urcorner | \ulcorner r \urcorner)$ is not crucial. For smaller ϵ it looses import altogether.
- The translation of a categorical vote into a probabilistic opinion can be motivated by a threshold notion of full belief.
- The literature has several other interpretations of w that are less specific but fit well with the current proposal.
- The link between trust and belief may nevertheless come across as somewhat contrived.

2.5 Follow-up research

Opinion pooling has a Bayesian reconstruction, which allows us to connect pooling to voting. This may stimulate research at the intersection of pooling with other disciplines.

- The peer disagreement debate has little contact with formal interactive epistemology. The current paper may serve a constructive role in establishing contact, e.g., by illuminating the strategy of splitting-the-difference in disagreement.

- Another application concerns consensus formation. The consensus formation process of DeGroot (1974) and Lehrer and Wagner (1981) can be modelled as a dynamic approach to common knowledge (Geneakoplos and Polemarchakis 1982).

3 Agreement and consensus formation

3.1 Consensus formation

Opinion pooling can be iterated. Writing $P^i_R(A) = r_i$ and $P^i_Q(A) = q_i$, we have on every round $i > 0$:

$$r_{i+1} = (1 - w_R)r_i + w_R q_i.$$

and similarly for Quassim. If the trust parameters w_R and w_Q are fixed and positive, the result is a convergent series of opinion pairs:

$$\langle r_1, q_1\rangle, \langle r_2, q_2\rangle, \ldots, \langle r_i, q_i\rangle, \ldots, \langle p, p\rangle$$

The same things apply to opinion pools with more than two agents.

Aumann's agreement

Aumann uses a Bayesian model for the interacting agents. He proves that the agreement of opinions is automatic if we assume the opinions to be common knowledge. He writes:

> "It seems to me that the Harsanyi doctrine is implicit in much of [the literature on opinion pooling]. . . The result of this paper may be considered a theoretical foundation for the reconciliation of subjective probabilities [i.e., by means of pooling]."

Surprisingly, there is no account of how the agreement theorem relates to iterated opinion pooling.

Agreement and consensus

The remainder of this paper provides a reconstruction of the approach to consensus based on Aumann's result. More precisely:

- The consensus formation can be represented as a dynamic approach to common knowledge.
- The Bayesian rendering of opinion pooling determines the requisite constraints on the common prior.

This rationalizes consensus formation via pooling, thereby revealing its conditions for applicability. Moreover, it offers a new perspective on pooling as information sharing through higher-order knowledge.

3.2 The agreement theorem

In Aumann's (1976) theorem, Raquel and Quassim are in the following epistemic situation:

- They share a space Ω of possible worlds and an initial probability assignment P over it.
- Both have their own information partition $\mathcal{R}$ and $\mathcal{Q}$, with elements R_i and Q_j. These partitions are common knowledge.
- Each of them has private information, in the form of one element from their partition, R_0 and Q_0.

Common knowledge

A central notion of the theorem is that of common knowledge, which relies on a specific conception of knowing:

> *Knowledge of knowledge of. . .*
>
> Raquel knows the proposition X iff R_0 is included in X. She knows that Quassim knows X iff all Q_j that intersect with R_0, i.e., that can be reached from R_0, are included in X. And so on.

Common knowledge of the posteriors r and q is associated with a particular set C_{rq} that is included in both $\ulcorner r \urcorner$ and $\ulcorner q \urcorner$.

A static result

Using this notion of common knowledge, Aumann proves the following.

> *Aumann's agreement theorem*
>
> If two people have the same priors, and their posteriors for an event A are common knowledge, then these posteriors are equal.

We can divide the set C_{rq} using the partitions $\mathcal{R}$ and $\mathcal{Q}$. For all R_i and Q_j overlapping with C_{rq}:

$$P(A|R_i \cap C_{rq}) = r \qquad \text{and} \qquad P(A|Q_j \cap C_{rq}) = q.$$

It does not matter in what direction we marginalize the probability for A, hence $r = q$.

Reaching agreement

Geanakoplos and Polemarchakis (1982) provide an account of how agreement may arise from an exchange of opinions between Raquel and Quassim.

- At the outset both receive private information, R_0 and Q_0, from their own information partition.
- At each round i they exchange their posteriors r_i and q_i.
- With this new information they exclude members from the information partition of the other agent.
- And they update their own opinions accordingly.

Sharing information

In the first round Raquel conditions on Quassim's opinion $\ulcorner q_1 \urcorner$ and obtains a new opinion about A:

$$P_R^2(A) = P_R^1(A|\ulcorner q_1 \urcorner) = P(A|\ulcorner q_1 \urcorner \cap R_0).$$

The set $\ulcorner q_1 \urcorner$ comprises the elements from Quassim's information partition that are consistent with the probability assignment $P_Q^1(A) = q_1$:

$$\ulcorner q_1 \urcorner = \{Q_j : P(A|Q_j) = q_1\}.$$

Quassim also deletes everything outside $\ulcorner q_1 \urcorner$ from the sets of Q_j that Raquel considers possible. Raquel and Quassim similarly update on $\ulcorner r_1 \urcorner$.

Higher-order knowledge

The process is then iterated on the resulting smaller space. Notice that Raquel and Quassim update on events that sit ever higher up in a hierarchy of knowledge,

$$\ulcorner q_j \urcorner = \bigcup \left\{ Q_k : P(A|Q_k \cap \ulcorner r_{j-1} \urcorner) = q_j \right\} \cap \ulcorner q_{j-1} \urcorner \cap \ulcorner r_{j-1} \urcorner,$$

and similar for r_i. A more explicit representation of this can be given in so-called Harsanyi type space but for present purposes we may collapse type space onto the information partitions.

3.3 Consensus as agreement

Iterated pooling and reaching agreement manifest in similar ways:

$$\langle r_1, q_1\rangle, \langle r_2, q_2\rangle, \ldots, \langle r_i, q_i\rangle, \ldots, \langle p, p\rangle$$

We can determine the prior over $\mathcal{R} \vee \mathcal{Q}$ such that approaching agreement fits the iterated pooling process. The constraints are

$$P(A|Q_k \cap \ulcorner r_{j-1} \urcorner) = q_j$$

for all j and for all k such that $Q_k \cap \ulcorner q_{j-1} \urcorner$ non-empty. A similar constraint set must hold for r_i.

Main result

Using Genest and Schervish (1985), this set of constraints can be imposed coherently onto the partitioned space Ω.

- The constraints leave some space for variation, especially in the distributions that agents choose for the opinions of other agents.
- Alternative pooling operations will lead to different constraints, e.g., on what opinions we expect others to have.
- It is not part of the model that agents can derive the values of trust parameters from the revealed opinions.
- The assignments within $\ulcorner q_j \urcorner^C$ are determined by the pooling that would happen if Quassim had revealed something other than q_j.

Multiple agents

The result generalizes to any number of agents, whether they learn the other opinions sequentially or all at the same time. Quassim and Simone, say, can be viewed as constituting a single agent:

$$r_{i+1} = \big(1-(w_Q+w_S)\big)r_i + (w_Q+w_S)\left(\frac{w_Q q_i + w_S s_i}{w_Q+w_S}\right).$$

However the model for multiple agents is organized, the likelihoods will encode specific assumptions about dependencies among the agents:

$$P(\ulcorner q_i \urcorner \cap \ulcorner s_i \urcorner | A \cap \ulcorner r_i \urcorner) \neq P(\ulcorner q_i \urcorner | A \cap \ulcorner r_i \urcorner) \times P(\ulcorner s_i \urcorner | A \cap \ulcorner r_i \urcorner).$$

3.4 Discussion

The result constitutes a bridge between two models of epistemic interaction. It seems natural to transport things over the bridge.

- Can we view pooling as a kind of information sharing? There seem to be conceptual differences between consensus and agreement.
- How can we interpret the common priors? They specify the conditions under which the epistemic shortcut of pooling is warranted.
- In addition, we may look for a taxonomy of consensus formations and consensus failures in terms of the common prior, or lack thereof.

A new angle on pooling

The standard view is that in the context of agreement, the exchange of opinions constitutes implicit information sharing, whereas in pooling the exchange amounts to a series of concessions.

- Consider the evidence presented by the opinions: after the initial round, the information is in opinion *changes*, i.e., in responses of the agents to each other.
- In the agreement context, the opinions refer to ever higher orders of knowledge. We may reinterpret the exchange in the context of pooling along similar lines.

The model of consensus can be viewed afresh, in terms of the conception of opinion as evidence stemming from the agreement context.

Pooling as shortcut

Following Genest and Schervish (1985), the agreement version of consensus by iterated pooling provides insights into the implicit assumptions of pooling.

- Pooling operations ignore many aspects of the distributions in the Bayesian model. We can motivate a consensus formation process by determining what aspects are relevant.
- Pooling entails specific dependencies among the opinions of agents. The Bayesian model offers insights into these dependencies, and this might help to motivate them.

Thank you

The slides of this talk will be available at http://www.philos.rug.nl/ romeyn. Papers are available upon request. For comments and questions, email j.w.romeijn@rug.nl.