Title: Tacking by Conjunction, Genuine Confirmation and Bayesian Convergence

Author: Gerhard Schurz (DCLPS, HHU)

Abstract:

Tacking by conjunction is a well-known problem for Bayesian confirmation theory. In the first section of the paper we point out disadvantages of orthodox Bayesian solution proposals to this problem and develop an alternative solution based on a strengthened concept of probabilistic confirmation, called genuine confirmation. In the second section we illustrate the application of the concept of genuine confirmation to Goodman-type counter-inductive generalizations and to post-facto speculations. In the final section we demonstrate that genuine confirmation is a necessary condition for Bayesian convergence to certainty based on the accumulation of conditionally independent pieces of evidence.

1. From Tacking by Conjunction To Genuine Confirmation

Tacking by conjunction is a deep problem of orthodox Bayesian confirmation theory. It is based on the insight that to each hypothesis H that is confirmed by a piece of evidence E one can 'tack' an irrelevant hypothesis X so that $H \wedge X$ is also confirmed by

E, in the Bayesian sense of "confirmation" as *probability-raising*, i.e. P(H|E) > P(H) ("P" for "probability"). To illustrate, according to the orthodox account each piece of evidence that confirms Newtonian mechanics also confirms the conjunction of Newtonian mechanics and creationism, although creationism is irrelevant to both Newtonian mechanics and the given evidence. This does not accord well with the pre-theoretic notion of confirmation that Bayesians purport to explicate.

Particularly counterintuitive is the *special* case of tacking by conjunction in which the irrelevant hypothesis is directly tacked to the evidence. Thus E confirms E∧X for every arbitrary hypothesis X, provided only that E and E∧X are *P-contingent*, where a proposition is called "P-contingent" if its probability is different from 0 and 1. For example, "snow is white" confirms "snow is white and creationism". Author (2014) calls this type of 'confirmation' "pseudo-confirmation". The probabilistic fact underlying pseudo-confirmation is simple (Proof in appendix A1):

Theorem 1 (Fact underlying pseudo-confirmation):

Assume H and E are P-contingent. Then E confirms H iff P(E|H) > P(E). Subcase: E |== H. Special case: H = E∧X.

Recent years have seen an increasing interest in the tacking by conjunction problem. Existing Bayesian solution proposals try to soften the negative impact of this result by showing that although H∧X is confirmed by E, it is so only to a lower degree (cf.

Fitelson 2002; Hawthorne and Fitelson 2004, and Crupi and Tentori 2010 who extended the focus to cases where H is disconfirmed by the evidence). Although these solution proposals provide important insights to the Bayesian confirmation model, they suffer from two drawbacks:

(1.) In application to the special case of the tacking problem in which X is directly tacked to E one would intuitively expect the tacked-on hypothesis "E∧X" to not be confirmed at all, but it counts as confirmed according to 'diminished confirmation' proposals.

(2.) These proposals are measure-sensitive in the sense that the 'diminished confirmation' claim holds only for some of the prominent Bayesian confirmation measures, but is violated for others (cf. co-author and author 2019).

One can easily see, however, that E increases the probability of E∧X only because E is a content element of E∧X and increases its own probability to 1 ($P(E|E) = 1$), while E does not increase the probability of the content element X that *logically transcend*s E, which means by definition that X is not entailed by E. More generally speaking, E does not need to raise the probability of the E-transcending content elements of a hypothesis H, in order to confirm H in the Bayesian sense. Gemes and Earman (Earman 1992, 98n5) have called this type of pseudo-confirmation "confirmation by (mere) *content-cutting*". To avoid this problem one ought to require that the confirmation takes place in those content elements of the hypothesis that are not logically contained in the evidence. Thus, in order for E to count as genuine confirmation of E∧X, E has to confirm X. This is the idea of genuine confirmation developed in author (2014a) and co-author and author (2019).

The notion of genuine confirmation is based on the notion of a *content element*. A definition of this notion for predicate languages has been given in co-author and author (2017, def. 4.2) and Author 2014b, def. 3.12-2) as follows (where propositional variables count as 0-placed predicates):

Definition 1: C is a content element of (hypothesis) H iff (i) H logically entails C (H |== C), (ii) no predicate in C is replaceable by an arbitrary new predicate with the same place number, salva validitate of H |== C, and (iii) C is elementary in the sense that C is not L(ogically) equivalent with a conjunction $C_1 \wedge C_2$ of conjuncts both of which are *shorter* than C.

The shortness criterion is related to the well-known concept of *minimal description length* in machine learning (Grünwald 2000); it is relativized to an underlying language with ¬, ∧, ∨,∃ and ∀ as primitive logical symbols, assuming that defined symbols are eliminated by their definitions. In propositional logic an equivalent version of this definition has been given in terms of shortest clauses (co-author and author 2017, def. 4.1; 2019, def. 3). Note that (p∨q)∧(p∨¬q) is not an admissible conjunctive decomposition of p, which avoids the Popper-Miller (1983) objection to inductive confirmation, which runs as follows: every hypothesis H is logically equivalent to the conjunction (H∨E)∧(H∨¬E). But H∨E is entailed by E and H∨¬E is provably disconfirmed by E, so "inductive" confirmation is impossible. But neither ($H \lor E$) nor ($H \lor \neg E$) are content elements of H.

Other technical definitions of content elements are possible − examples are Friedman's (1974) "independently acceptable elements", Gemes' (1994) "content parts" and Fine's (2017) "verifiers". The technical details don't matter as long as the core idea is captured, namely the decomposition of a hypothesis into a set of smallest content elements that are not further conjunctively decomposable in relevant ways and whose conjunction is L-equivalent to the original hypothesis.

The notion of genuine confirmation (GC) has been explicated by co-author and author (2019) in three versions: qualitative full GC, qualitative partial GC and quantitative GC:

Definition 2: Assume E does not entail H.[1] Then:

1.1 Qualitative full GC: E fully genuinely confirms H iff (i) $P(X|E) > P(X)$ holds for all E-transcending content elements X of H.

1.2 Qualitative partial GC: E partially genuinely confirms H iff $P(X|E) \geq P(X)$ holds for all and $P(X|E) > P(X)$ holds for some E-transcending content elements X of H.

1.3 Quantitative GC: The degree of genuine confirmation that E provides for H is the

---

[1] We leave it open whether one wants to count logical entailment (E |== H) as a case of 'genuine confirmation' or not. In this case, H has no E-transcending content elements.

sum of the confirmation degrees, conf(E,H), over all E-transcending content elements X of H, divided by their number (where "conf(E,H)" is one of the standard Bayesian confirmation measures, e.g., the difference measure).

Note that although the notion of genuine confirmation (in particular that of genuine full confirmation) strengthens ordinary Bayesian confirmation considerably, it is spelled out within the ordinary Bayesian framework.

2. Applications of Genuine Confirmation

In co-author and author (2019) it is shown that the so-defined measure has a number of attractive features. For example, it can solve problem of measure sensitivity. Moreover, qualitative partial GC implies positive quantitative GC; thus the qualitative and the quantitative notions of GC are in coherence. In this paper we elaborate some attractive features of qualitative confirmation.

Partial (qualitative) genuine confirmation is sufficient to rule out the special case of tacking by conjunction in which the irrelevant hypothesis X is directly tacked on the evidence. This includes an important subcase, namely the problem of Bayesian pseudo-confirmation of Goodman-type *counter-inductive generalizations*. Let E be the evidence that all observed emeralds have been green, H_1* the hypothesis that all unobserved emeralds will be green and H_2* the hypothesis that call unobserved emeralds will be red. Then the inductive generalization H_1 is L-equivalent with $E \wedge H_1$* and the counter-inductive generalization H_2 is L-equivalent with $E \wedge H_2^*$. Now, following from theorem 1, E confirms both H_1 and H_2 in the pseudo-sense. However, E's confirmation of H_2 is not a genuine one, because E does not confirm H_2's E-transcending content element H_2^*. Moreover, note that E will only confirm the E-transcending inductive projection H_1^* of E, and thus genuinely confirm H_1, if the underlying probability function P satisfies certain additional *inductive* principles, such as de Finetti's exchangeability (invariance of P under permutation of individual constants) and regularity ($P(S) \neq 0, 1$ for every analytically contingent S).

For ruling out all sorts of tacking by conjunction, full (qualitative) genuine confirmation is needed. A further important application of full GC is the elimination of the pseudo-confirmation of *post-facto speculations*. By this we mean the confirmation of hypotheses that contain *theoretical* concepts or, more generally, *latent variables* that are not present in the evidence. By postulating sufficiently many latent variables and suitable principles connecting them with the observed variables, one can explain any observation whatsoever. For example, the fact that grass is green (E) pseudo-confirms the hypothesis (H) that "God wanted that grass is green and whatever God wants, happens". Here "God's wishes" figure as the latent variable. Author (2014a) suggests to understand the pseudo-confirmation of post-facto speculations based on Worrall's (2016) concept of *use-novel evidence*. Worrall's account starts from the observation that the values of the latent variables of a general type of hypothesis are *fitted* towards the evidence. Author (2014a) argues that the unfitted hypothesis H_{unfit} should be understood as a content element of the fitted hypothesis H_{fit}, which is obtained as the existential quantification over the possible values of the latent variables. If H_{unfit} is so general that it can be fitted to every evidence, then H_{unfit} cannot be said to be confirmed merely by the fact that H_{unfit} was fitted to a particular evidence E_1, leading to H_{fit} (although by theorem 1 H_{fit}'s probability has increased, $P(H_{fit}|E_1) > P(H_{fit})$). For example, in the case of the "God-has-wanted-it" hypothesis, H_{unfit} would be the hypothesis "∃X(God wants X and whatever God wants, happens)". According to our account, this hypotheses cannot be genuinely confirmed by theological post-facto explanations of events. This follows straightforwardly from $P(E_1|H_{unfit}) = P(\neg E_1|H_{unfit})$, which holds because H_{unfit} can be fitted to any evidence whatsoever. Only if the fitted hypotheses is confirmed by a second *use-novel* piece of evidence E_2, i.e. one to which H_{unfit} has not been fitted and which H_{fit} could have predicted, then H_{unfit} can be said to be confirmed via the confirmation of H_{fit} by E_1 and E_2. For obviously it is not possible to fit H_{unfit} to a given evidence E_1 and then to confirm the so-obtained H_{fit} by any other evidence E_2 whatsoever. In this way, the concept of genuine confirmation provides a probabilistic justification of Worrall's criterion of use novelty. As a side remark we mention that the use-novelty criterion is by no means a purely philosophical invention, but is employed in a famous computational learning method, namely cross validation (Shalev-Shwartz and Ben-David 2014, sec. 11.2).

When we argued above that the probability of an E-transcending content element of H is or is not raised conditional on an evidence E that raises H's probability, we frequently argued by considerations of intuition. Probability theory itself does not tell us the value of P(E|C). Based on the considerations above we suggest the following rationality criteria for the spread of the evidence-induced probability increase from a hypothesis H to its E-transcending content elements.:

Necessary criteria for spread of probability increase:

If H increases E's probability, then the resulting probability increase of H by E spreads from H to an E-transcending content element C of H ($P(C|E) > P(C)$) *only if:*

(1.) C is necessary within H to make E probable, i.e., there exists no conjunction H* of content elements of H that makes E at least equally probable ($P(E|H^*) \geq P(E|H)$) but does not entail C, and

(2.) it is not the case that C is an existential quantification, $C = \exists x H(x)$, and H results from a parameter-adjustment of x in H(x) towards the evidence E, such that an equally good fitting of H(x) would have been possible for every possible alternative evidence E'.

In the next section be explain a particular important application of the concept of genuine confirmation: it is a precondition for an important form of Bayesian convergence.

3. From Genuine Confirmation to Bayesian Convergence

An important part of Bayesian epistemology are convergence theorems. According to them the conditional probability of a hypotheses can be driven to near certainty, if many confirming and mutually conditionally independent pieces of evidence for this hypotheses are accumulated (Earman 1992, 141ff.). Most versions of Bayesian convergence theorems have been formulated for hypotheses not containing latent variables, typically hypotheses that are obtainable from the evidence by enumerative induction. For example, it has been shown that if P is countably additive, then $\lim_{n\to\infty}P(p(Fx)=r \mid (E_1,\ldots,E_n)) = 1$, where each E_i is Fa_i or $\neg Fa_i$ and F's frequency limit in the sequence $(E_1,\ldots,E_n)$ is r (this is a consequence of the theorem of Gaifman and Snir 1982). More important, however, is convergence theorem for hypotheses containing latent variables. A well-known convergence theorem for this case is the following (proof in appendix A2):

Theorem 2 - convergence to certainty:

If a P-contingent hypothesis H satisfies the following conditions

(a) H is confirmed by each of the P-contingent pieces of evidence $E_1,\ldots,E_n$ (i.e., $P(E_i|H) > P(E_i)$ for all $i\in\{1,\ldots,n\}$),

(b) the pieces of evidences are mutually independent conditional on H, i.e., $P(E_i|H\wedge E_1\wedge\ldots\wedge E_{i-1}) = P(E_i|H)$ for all $i \in\{1,\ldots,n\}$ (and some ordering of the E_i's),

(c) and they are also mutually independent conditional on $\neg$H,

then $\lim_{n\to\infty}P(H|E_1\wedge\ldots\wedge E_n) = 1$.

Convergence to certainty in spite of a small prior probability is the ideal case of scientific confirmation. The confirmation of Darwinian evolution theory by multiple pieces of evidence constitutes an example. Theorem 2 is a reformulation of the *Condorcet jury theorem*, with the agreeing reports of the independent witnesses being equated with the independent evidences (Bovens and Hartmann 2003; List 2004). Surprisingly, however, a necessary condition for convergence to certainty is full genuine confirmation. The existence of only one E-transcending content element of H, call if C, that is not confirmed by any one of the evidences E_i, is sufficient to prevent convergence to certainty. Since C's probability is not raised by any of the E_i it holds that $P(C|E_1\wedge\ldots\wedge E_n) = P(C)$. But $P(C|E_1\wedge\ldots\wedge E_n) = P(C)$ is an upper bound of $P(H|E_1\wedge\ldots\wedge E_n)$, since H entails C. Thus $P(H|E_1\wedge\ldots\wedge E_n)$ is forced to stay below P(C), which is small, and cannot approach certainty.

Theorem 3 – failure of convergence to certainty:

If a hypotheses H satisfies conditions (a) and (b) of theorem 2, but contains a content element C that is not confirmed by any of the evidences E_i, then

(i) $\lim_{n\to\infty}P(H|E_1\wedge\ldots\wedge E_n) \leq P(C)$, and

(ii) condition (c) of theorem 2 fails.

Note that if case of theorem 3(i) obtains and H starts from a low prior, then H's probability is still increasing conditional on the accumulating pieces evidence, however, it does not converge to 1, but to P(C) (from below).

In conclusion, genuine confirmation is a precondition for the sustainable confirmation of hypotheses that are allowed to contain latent variables. While the proof of theorem (i) is obvious from the arguments above, it is prima facie puzzling how this result squares with theorem 2. It turns out that entailment of an irrelevant content elements undermines the independence of the pieces of evidence conditional on the negation of the hypothesis, which is the content of theorem 3(ii). Theorem 3(ii) points towards a general limitation of the convergence theorem 3; because of its importance we state the proof right here in the text (not in the appendix). For whenever the negation of the hypotheses, $\neg H$, can be decomposed into a partition of finer hypotheses that convey different probabilities to the evidence, then the independence of the pieces of evidence conditional on $\neg H$ fails. For example, assume $\neg H$ splits into two disjoint hypotheses H_2, H_3 such that $P(E_i|H_2)$ is much larger than $P(E_i|H_3)$ (for all i), although $(E_i|H_2 \vee H_3) = P(E_i|\neg H) < P(E_i)$, which follows from $P(E_i) < P(E_i|H)$ and the P-contingency of E_i and H. Then $P(E_j|\neg H \wedge E_i) > P(E_j|\neg H)$ will hold, because the fact that E_i obtained makes it more probable that H_2 and not H_3 obtained, which in turn makes E_j more probable.

Now assume that H is a hypothesis that has an irrelevant content element C, $H = H_1 \wedge C$, where $P(E_i|H_1) > P(E_i)$ and C is irrelevant for E_i both unconditionally and conditionally on H_1. In this case the negation $\neg(H_1 \wedge C)$ splits into the finer partition $\neg H_1 \wedge C$, $\neg H_1 \wedge \neg C$ and $H_1 \wedge C$. While $P(E_i|\neg H_1 \wedge \pm C) < P(E_i)$ holds for both $\pm C = C$ and

$\pm C = \neg C$, the third element of the partition behaves differently, namely $P(E_i|H_1 \wedge \neg C) > P(E_i|H_1)$, and this destroys the independence of the evidence conditional on $\neg(H_1 \wedge C)$.

Fortunately there is a generalized version of theorem 3 that is relativized to a given possibly large partition of hypotheses that are assumed to be sufficiently strong to guarantee mutual conditional independence of the pieces of evidence (proof in appendix A3):

Theorem 4 - generalized convergence to certainty:

Assume a P-contingent hypothesis H_1 belongs to a partition of hypotheses $\{H_1,\ldots,H_m\}$ satisfying the following conditions:

(a) every piece of evidence *favors* H_1 over every other hypothesis by at least δ (for some $\delta>0$), i.e., $P(E_i|H_1) \geq P(E_i|H_r)+\delta$ for all $r>1$ and $i\in\{1,\ldots,n\}$, and

(b) the pieces of evidences are mutually independent conditional on every H_k ($k\in\{1,\ldots,m\}$), i.e., $P(E_i|H_k \wedge E_1 \wedge \ldots \wedge E_{i-1}) = P(E_i|H)$ for all i ($i\in\{1,\ldots,n\}$),

then (i) $P(H_1|E_1 \wedge \ldots \wedge E_n) \geq \dfrac{h}{h + (1-h)\cdot(1-\delta)^n}$, and

(ii) $\lim_{n\to\infty} P(H_1|E_1 \wedge \ldots \wedge E_n) = 1$.

If we apply theorem 4 to hypotheses that are conjunctions of several content elements, $H = H_1 \wedge \ldots \wedge H_k$, then the smallest partition of competing hypotheses that has to be checked in regard to conditional independence of the pieces of evidence is the partition $\{\pm H_1 \wedge \ldots \wedge \pm H_k : \pm H_i \in \{H_i, \neg H_i\}, 1 \le i \le k\}$, which contains 2^k elements.

Appendix: Proof of theorems:

A1. Proof of theorem 1:

This is well-known: Assuming H and E are P-contingent, then

$P(H|E) = P(H) \cdot P(E|H)/P(E)$, and $P(E|H) \cdot P(H)/P(E) > P(H)$ iff $P(E|H) > P(E)$. Q.E.D.

A2. Proof of theorem 2:

Theorem 2 follows from theorem 4 by substituting $\{H, \neg H\}$ for $\{H_1, \ldots, H_m\}$. Note that for P-contingent E and H, $P(E|H) > P(E)$ entails $P(E) > P(E|\neg H)$, which follows from the fact that $P(E) = P(E|H) \cdot P(H) + P(E|\neg H) \cdot P(\neg H)$. Thus there exists a δ such that $P(E|H) \ge P(E|\neg H) + \delta$, which is the assumption of theorem 3. Q.E.D.

A3. Proof of theorem 4:

We abbreviate $P(E_i|H_1)$ as p_i and write $\Sigma\{x_1, \ldots, x_n\}$ and $\Pi\{x_1, \ldots, x_n\}$ for the sum and the product of the numbers $x_1, \ldots, x_n$, respectively. We calculate as follows. By Bayes' theorem:

$$P(H_1|E_1 \wedge \ldots \wedge E_n) = P(E_1 \wedge \ldots \wedge E_n|H_1) \cdot P(H_1)/\Sigma\{P(E_1 \wedge \ldots \wedge E_n|H_r) \cdot P(H_r) : 1 < r \le m\}.$$

Since $P(E_1 \wedge \ldots \wedge E_n|H_r) = \Pi\{P(E_i|H_r \wedge E_1 \wedge \ldots \wedge E_{i-1}) : 1 \le i \le n\}$ and condition (b) of theorem 4 we continue:

$$= \frac{h \cdot \Pi\{p_i : 1 \le i \le n\}}{h \cdot \Pi\{p_i : 1 \le i \le n\} + \Sigma\{P(H_r) \cdot \Pi\{P(E_i|H_r) : 1 \le i \le n\} : 1 \le r \le m, r > 1\}}$$

$$\ge \frac{h \cdot \Pi\{p_i : 1 \le i \le n\}}{h \cdot \Pi\{p_i : 1 \le i \le n\} + \Sigma\{P(H_r) : 1 \le r \le m, r > 1\} \cdot \Pi\{(p_i - \delta) : 1 \le i \le n\}} \quad \text{(from condition (a))}$$

$$= \frac{h \cdot \Pi\{p_i : 1 \le i \le n\}}{h \cdot \Pi\{p_i : 1 \le i \le n\} + (1-h) \cdot \Pi\{(p_i - \delta) : 1 \le i \le n\}} = \frac{1}{1 + \frac{1-h}{h} \cdot \frac{\Pi\{(p_i - \delta) : 1 \le i \le n\}}{\Pi\{p_i : 1 \le i \le n\}}}.$$

Because of $\frac{\Pi\{(p_i - \delta) : 1 \le i \le n\}}{\Pi\{p_i : 1 \le i \le n\}} \le (1-\delta)^n$ we obtain the claim of theorem 4 (i), which entails theorem 4 (ii) because of $\lim_{n\to\infty}(1-\delta)^n = 0$. Q.E.D.

References (blinded)

Author (2014a). in *Studies in History and Philosophy of Science*.

Author (2014b). [Book]

Bovens, L., and Hartmann, S. (2003). *Bayesian Epistemology*, Oxford University Press, Oxford.

Co-author and author (2019). in *Studia Logica*.

Co-author and author (2019). in *British Journal for the Philosophy of Science*.

Crupi, V. and Tentori, K. (2010). "Irrelevant Conjunction: Statement and Solution of a New Paradox", *Philosophy of Science*, 77, pp. 1–13.

Earman, J. (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*, Cambridge, MA: MIT Press.

Fitelson, B. (2002). "Putting the Irrelevance Back into the Problem of Irrelevant Conjunction", *Philosophy of Science*, 69, 611–22.

Gaifman, H., and Snir, M. (1982). "Probabilities Over Rich Languages", *Journal of Symbolic Logic*, 47, 495-548.

Grünwald, P. (2000). "Model Selection Based on Minimal Description Length", *Journal of Mathematical Psychology*, 44, 133-152.

Hawthorne, J. and Fitelson, B. (2004). "Discussion: Resolving Irrelevant Conjunction with Probabilistic Independence", *Philosophy of Science*, 71, 505–14.

List, C. (2004). "On the Significance of the Absolute Margin", *British Journal for the Philosophy of Science*, 55, pp. 521–44.

Popper, K., and Miller, D. (1983). "A Poof of the Impossibility of Inductive Probability", *Nature*, 302, 687-688.

Shalev-Shwartz, S. and S. Ben-David. (2014). *Understanding Machine Learning. From Theory to Algorithms*, Cambridge University Press, New York.