RECOVERING A PRIOR FROM A POSTERIOR: SOME PARAMETERIZATIONS

OF JEFFREY CONDITIONING

(running head: Recovering a Prior)

Carl G. Wagner

Department of Mathematics

The University of Tennessee

Knoxville, TN 37996-1320

USA

Telephone: 865-974-2461

Fax: 865-974-6576

Email: cwagner@tennessee.edu

RECOVERING A PRIOR FROM A POSTERIOR

*Abstract.* Given someone’s fully specified posterior probability distribution *q* and information about the revision method that they employed to produce *q*, what can you infer about their prior probabilistic commitments? This question furnishes an entrée into a thoroughgoing discussion of a class of parameterizations of Jeffrey conditioning in which the parameters furnish information above and beyond that incorporated in . Our analysis highlights the ubiquity of Bayes factors (ratios of new to old odds) in the study of probability revision.

**1. Introduction.** In what follows,  is a set of possible states of the world and **A** is an algebra of subsets of . ***P*** denotes the set of all finitely additive probability measures (henceforth, simply termed *probabilities*) on **A**. If *q* is a revision of the probability *p*, and *A* is an event, the *relevance quotient* (Carnap’s term)  is equal to the ratio . If *A* and *B* are events, the *Bayes factor*  is the ratio of the posterior odds on *A* against *B* to the prior such odds, i.e.,   When *q* comes from *p* by conditioning on the event *E*, then  is just the *likelihood ratio*  Bayes factors and relevance quotients are clearly related by the formula .

Probabilistic epistemology is typically a forward-looking enterprise. Given the prior *p*, new evidence, and a judgement that a particular method of revision is warranted, applying that method leads in general to a unique posterior *q*. In this note, we consider the inverse problem: Given someone’s fully specified posterior *q* and information about the revision method that they employed to produce *q*, what can you infer about their prior probabilistic commitments? Unsurprisingly, the answer dependson the extent to which you are apprised of revision parameters (if any) that are not already implicit in *q*. Suppose, for example, you are informed that the probability *q* has come from some *p* by conditioning on the event *E*, or by Jeffrey conditioning on the partition **E**. As will be familiar to many readers, this constrains the admissible priors  in the former case to those satisfying  , for all **A**, and in the latter to those satisfying  , for all **A** and  (Jeffrey 1992, 117-119). As shown in sections 2 and 3, these *rigidity* constraints still allow infinitely many admissible priors. In section 4, however, we show that for a broad class of alternative parameterizations of Jeffrey conditioning each posterior *q* determines a unique admissible prior, as a result of the fact that such parameterizations furnish information over and above that included in *q*. In section 5 we examine three such parameterizations, and note that two of these, due to Field (1978), and to Jeffrey and Hendrickson (1988/89) employ parameters that, in an important sense, efface all traces of the relevant prior.

**2. Revision by conditioning.** Suppose that you are given the fully specified probability *q****P*** , and you are told that *q* was derived from some probability *p* by conditioning on the event *E***A**. Unless *E* (in which case *p* is necessarily identical to *q*), there are infinitely many priors *p* such that *q*(), as shown by the following theorem.

THEOREM 1. Suppose that *q****P***, *E***A**, and ** A probability *p****P*** satisfies the equation *q*(*A*) for all **A** if and only if there exists a number  and a probability ***P***, with  such that

(1)



*Proof.*  Sufficiency. It is straightforward to show that *p*, as defined by (1), is a probability on **A**, with  In particular, if  then  Since  

and it follows from (1) that, for all**A**,



Note that since  there always exists an ***P*** with  For any  we can simply take *r* to be the *point mass at*  defined by (*i*)  if , and (*ii*)  if  Necessity. Suppose that *q*(*A*) for all **A**. If  then  and we may set  and  any probability such that  If , then



and so we may set  and  

**3. Revision by Jeffrey conditioning.** To avoid the discussion of distracting (and unenlightening) special cases, we restrict consideration, in this section and the next, to the set ***P*****+** of so-called *strictly coherent* probabilities on **A**, i.e., to the set of all such that  whenever  Suppose that **E**  is a partition of  consisting of members of **A**, where  You are given the fully specified probability  and you are told that  was derived from some (unidentified) probability  by Jeffrey conditioning (henceforth, JC) on the partition **E**, i.e., that

(2)  for all **A**,

where, of course, . Here again, there are infinitely many probabilities  satisfying formula (2).

THEOREM 2. Given the fully specified probability ***P*** +,  comes from the probability  ***P*** + on the partition **E** by JC if and only if there exists a sequence  of positive real numbers, with , such that, for all **A**,

(3) .

*Proof.* Sufficiency. It is straightforward to show that ***P*** +. Moreover, (3) implies that

(4) .

So, by the total probability law (Evans, et al, 2004), along with (4),

(5) =.



Necessity. Just as (3) implies (4), formula (2) implies that . In this case, *p* is assumed to be fully specified. So, with  we have

. 

**4. A class of alternative parameterizations of JC.** Let , **A** , and **E**  be as above, and let ***P*** +. Let  be *any sequence of positive real numbers whatsoever*, and consider revising the prior *p* to the posterior *q* by the formula

(6)  for all  **A**.

It is straightforward to check that the set function *q* is indeed a probability on **A**. Initial appearances notwithstanding, formula (6) furnishes no new and exotic method of probability revision, for if , then . So (6) may be equivalently expressed as

(7) , with  as defined above,

whence  obviously comes from *p* by JC on **E**. Alternatively, one can of course verify that

(8) 

The family of revisions by JC on **E** given by formulas of type (6) are not merely a sub-family of the family of all JC revisions on **E** given by the classical formulas of type (2), but coterminous with the latter family. There are a number of ways to demonstrate this, which requires showing that each formula of type (2) can be equivalently expressed as a formula of type (6). This demonstration is deferred until Section 5 below, in order to continue our exploration of the general class of formulas (6) without interruption.

We begin by connecting the mysterious quantities  with more salient parameters.

THEOREM 3. For all 

(9) .

*Proof.*  By formula (6),

. 

Given the fully specified posterior *q*, the partition **E**, and the parameters , there is a *unique prior p* satisfying the revision formula (6).

THEOREM 4. Formula (6) holds for probabilities *p*, *q*  ***P*** + if and only if

(10) 

*Proof.* Sufficiency. It is straightforward to confirm that (10) implies (6). Necessity. From (6) and its consequence (9),

  , and so

(11)  , whence,

(12) 

Summing each side of (12) from  to  yields

 , whence,

(13)  

Substituting the right-hand side of (13) for  in (12) yields

(14)  ,

which establishes (10) when . But by (8) and (14),

  

 . 

As suggested in Section 1 above, what accounts for the “invertibility” of formulas of type (6) is that we are given information over and above that incorporated in *q*. In the following section we examine three special cases of formula (6).

**5. Three parameterizations of JC of type (6).** We now complete the demonstration, promised earlier, that the classes of probability revisions given by formulas of type (6) are coterminous with those of type (2). As the three parameterizations of JC that accomplish this are of independent mathematical and philosophical interest, we present a fairly detailed discussion of each of them.

*1. Relevance quotient parameterization.* The simplest way to recast formula (2) in the form given by (6) is

(15) , where .

There is a certain artificiality in formula (15), since the numerator . But it falls squarely under the rubric furnished by (6), and so all earlier results about parameterizations of the form (6), in particular Theorem 4, hold for (15).

What (15) fails to do is to exhibit  as a function of the prior and an *input parameter* that expresses what is learned *from new evidence alone*.1 For the factors  incorporate information about the prior *p*. As a simple illustration of this assertion, note that from  we may deduce that .2 The following example elaborates this observation. Suppose that **E** is a partition of , with **A**, and we envision revising *p* to *q* by JC on **E**, with , , and  Here,  and  Suppose now that we are given *only* these values of  and . We can then deduce information about the prior , namely, that *p* cannot be among those probabilities  satisfying , since  Of course, we can normalize these summands by dividing each by  441/437. Setting =21/38 and , it follows that , and so the latter values can form the basis of a revision of  to  by JC on **E**. Now, however,  and  On the other hand,  This is no coincidence since, (i) normalized relevance quotients are basically alter egos of Bayes factors (Wagner 2002, 271), and (ii) as will be seen below, Bayes factors are ubiquitous in representions of what is learned from new evidence alone.

We next turn to two parameterizations of JC based on genuine input factors.

*2. Field’s parameterization.* Hartry Field (1978) succeeded in deriving from (2) the intriguing formula

(16) , where :  the geometric mean  3

whence,

(17) 

Field used the parameterization (17) to establish, for the first time, sufficient conditions for successive JC revisions on different finite partitions to commute. 4

It is a straightforward exercise to show that the Field parameter product  In fact this property characterizes formulas of type (17) among those of type (6).

THEOREM 4. Suppose that  is a sequence of positive real numbers, with . For *every* ***P*** +, let  be defined on the partition **E**  by  . Since , the values  provide the foundation of a revision by JC of  to on **E**. Moreover, , so that  is precisely the parameter labeled  in formula (17).

*Proof.*  

*Remark 1.*  Suppose that  is revised to  by JC on **E**, but you are only apprised of the values of the Field parameters  associated with this revision. It follows from Theorem 4 that you can infer nothing whatsoever about , since, unlike relevance quotient parameters, the parameters  are compatible , in the obvious sense, with every probability in ***P*** +.

*2. The Jeffrey-Hendrickson parameterization.*  Jeffrey and Hendrikson (1988/89) derived from (2) the somewhat simpler formula

(18)  where  .

Note that it is always the case that  In fact this property characterizes formulas of type (18) among those of type (6).

THEOREM 5. Suppose that  is a sequence of positive real numbers, with  For every ***P*** +, let  be defined on the partition **E**  by  . Since , the values  provide the foundation of a revision by JC of  to on **E**. Moreover, , so that  is precisely the parameter labeled  in formula (11).

*Proof.* Straightforward. 

*Remark 2. .*  Suppose that  is revised to  by JC on **E**, but you are only apprised of the values of the Jeffrey-Hendrickson parameters  associated with this revision. It follows from Theorem 5 that you can infer nothing whatsoever about , since those parameters are compatible, in the obvious sense, with every probability in ***P*** +.

As shown in Wagner (2002), formula (11) can be generalized to the case of JC revisions on any countable measurable partition (where **A** is now a sigma algebra of subsets of  and probabilities are countably additive). Using this parameterization Wagner showed that Field’s sufficiency conditions for commutativity of successive JC revisions also hold for countable partitions **E** and **F**, and proved under mild regularity conditions that these conditions are also necessary for commutativity.5

**6. Discussion.**

As noted earlier, the reason that one can “invert” a formula of type (6) is that the parameters furnish information over and above that provided by the fully specified posterior Notably, these parameters all prominently feature Bayes factors, either directly (as in the case of the Field and Jeffrey-Hendrickson parameterizations), or indirectly (as in the relevance factor parameterization, and, in the general case, as given by formula (12)).

As will be familiar to many readers, Bayes factors have a distinguished pedigree as measures of what is learned from new evidence alone, with prior probabilities expunged. Indeed, the number  was termed the *weight of evidence* by I. J. Good (1950). According to Good (1979), Alan Turing advocated using weights of evidence to represent the gain or loss of the probability of one hypothesis versus another resulting from the receipt of additional evidence. Such weights were routinely used in the code-breaking work at Bletchley Park, where Good and Turing were colleagues during WW II (Jeffrey, 2004, pp. 31-32). It is hoped that the present note succeeds in further underscoring the centrality of Bayes factors in the representation and analysis of probability revisions.

**Notes**

1. The term *input parameter* is due to Field (1978), who references the Jeffrey-Carnap correspondence (Jeffrey 1975, 363), in which Carnap poses the problem of identifying a number which “is to indicate the subjective certainty of the sentence on the basis of the observational experience.” As Field points out, Carnap seems to regard such numbers (Carnap’s “b’s”) as a probabilities, and attributes Carnap’s lack of progress on resolving this problem to this way of thinking. Jeffrey, in response to Carnap, asserts that the b’s are properly thought of as “high level theoretical terms” to be derived from (rather than used to construct) posterior probabilities.

2. More generally, if, for a given event *E* and unknown probabilities *p* and *q*, we are informed that , we can deduce that , which is informative whenever  By contrast, if we are informed that , this is compatible with  taking any value in the open interval . For setting  yields (i)  and (ii) 

3. Field actually expressed  as , where , and interpreted  in purely physicalist terms as expressing the direct and immediate effect of a given stimulus. However, Garber (1980) noted that if , repeated exposure to that stimulus would then drive the value of  toward 1. Unfortunately, Garber’s counter-example led philosophers to ignore Field’s beautiful parameterization of JC, and its application to the commutativity problem for successive JC revisions on different finite partitions. As argued in Wagner (2002), Field’s analysis can easily be divested of its (inessential) physicalist gloss.

4. Consider two successive updating schemes, , and , where (i) *q* comes from *p* by JC on **E ** , and *r* from *q* by JC on **F ** , and (ii)  comes from *p* by JC on **F**, and  from  by JC on **E**. It is a corollary of Field’s parameterization that the Bayes factor identities (*i*) , , and (*ii*) , , imply that 

5. Notably, Wagner’s proof that these conditions are necessary for commutativity relies only on the rigidity property of JC revisions.

**References**

Evans, M. and Rosenthal, J. (2004). *Probability and Statistics*. (New York: Freeman)

Field, H. (1978). A note on Jeffrey conditionalization. *Philosophy of Science*, 45, 361-367

Garber, D. (1980). Field and Jeffrey conditionalization. *Philosophy of Science*,47, 142-145

Good, I.J. (1950). *Probability and the Weighing of Evidence*. (London: Charles Griffin)

Good, I.J. (1979). A.M. Turing’s statistical work in World War II. *Studies in the History of Probability and Statistics XXVII*, *Biometrica* 66: 393-396

Jeffrey, R.C. (1975). Carnap’s empiricism, in Maxwell and Anderson (eds.), *Minnesota Studies in the Philosophy of Science*, vol. VI. (Minneapolis: University of Minnesota Press)

Jeffrey, R.C. and Hendrickson, M. (1988/89). Probabilizing pathology, *Proceedings of the Aristotelian Society* 89 (Part 3): 211-225

Jeffrey, R.C. (1992). *Probability and the Art of Judgment*.(Cambridge: Cambridge University Press)

Wagner, C. (2002). Probability kinematics and commutativity, *Philosophy of Science*,69, 266-278