Probability Revision, The Uniformity Rule, and the Chan- Darwiche Metric

Carl G. Wagner[1]

Abstract The author has proposed a rule of probability revision dictating that identical learning be reflected in identical ratios of new to old odds. Following this rule ensures that the final result of a sequence of probability revisions is undisturbed by an alteration in the temporal order of the learning prompting these revisions. There is also a close connection between this rule and an intriguing metric on probability measures introduced by Chan and Darwiche.

Keywords: Bayes factor, Chan-Darwiche metric, probability revision.

1 The Commutativity Principle

Consider the following belief revision schema, representing two possible sequential revisions of the probability measure p:

$$p \longrightarrow q \longrightarrow r \qquad \text{and} \qquad p \longrightarrow s \longrightarrow t. \tag{1}$$

Suppose that the revisions of p to q, and of s to t, are prompted by identical learning, and that the revisions of q to r, and of p to s, are prompted by identical learning. It is then widely held that it ought to be the case that $r = t$. As van Fraassen (1989) puts it, two persons who undergo identical learning experiences on the same day, but in a different order, ought to agree in the evening if they had exactly the same opinions in the morning. Call this the *Commutativity Principle.*

A simple rule of probability revision ensures that the Commutativity Principle is satisfied. This *Uniformity Rule,* occurring in particular cases in Wagner (1997, 1999, 2001, 2002), and given general formulation in Wagner

Department of Mathematics, The University of Tennessee, Knoxville, TN 37996-1300, U.S.A. wagner@math.utk.edu, Telephone: 865.974.4305. Fax: 865.974.6567.

(2003), dictates that identical learning be reflected in identical ratios of new to old odds, also known as *Bayes factors.* This note explores the connection between the Uniformity Rule and an intriguing metric on probability measures introduced by Chan and Darwiche (2002). The upshot is that revisions of two different probability measures based on identical learning, when effected by the Uniformity Rule, move us the same Chan-Darwiche distance from the priors in question.

2 Terminology and Notation

A sigma algebra $\mathbf{A}$ of subsets of Ω is *purely atomic* if the family $\mathbf{A}^*$ of atomic events in $\mathbf{A}$ is countable, and constitutes a partition of Ω. Every finite algebra is purely atomic, whatever the cardinality of Ω, and if Ω is countable, then every sigma algebra on Ω is purely atomic (Renyi 1970, Theorems 1.6.1, 1.6.2). If q is a revision of probability measure p, and A and B are events, then the *probability factor* (or *relevance quotient*) $\Pi_{q,p}(A)$ is the ratio

$$\Pi_{q,p}(A) := \frac{q(A)}{p(A)} \tag{1}$$

of new to old probabilities, and the *Bayes factor* $\beta_{q,p}(A : B)$ is to ratio

$$\beta_{q,p}(A : B) := \frac{\frac{q(A)}{q(B)}}{\frac{p(A)}{p(B)}} \tag{2}$$

of new to old odds. When $q(.) = p(.|E)$, then (2) is simply the likelihood ratio $\frac{p(E|A)}{p(E|B)}$. More generally,

$$\beta_{q,p}(A : B) = \frac{\Pi_{q,p}(A)}{\Pi_{q,p}(B)}, \tag{3}$$

a simple, but useful, identity.

In what follows we assume for simplicity that all probability measures are *strictly coherent,* i.e., that all nonempty events have positive probability. With the addition of certain technical conditions, however, Theorem 1 below holds for arbitrary probabilities.

3 Bayes Factors and Commutativity

The following theorem demonstrates that the Commutativity Principle is satisfied for purely atomic sigma algebras when identical learning is represented by identical Bayes factors at the level of atomic events.

Theorem 1. *Suppose that the probabilities in the revision schema*

$$\begin{array}{ccc} p & \longrightarrow & q \\ & & \downarrow \\ \downarrow & & r \\ & & \\ s & \longrightarrow & t \end{array}$$

are defined on a purely atomic sigma algebra $\mathbf{A}$*, with* $\mathbf{A}^*$ *denoting the set of atomic events in* $\mathbf{A}$*. If the Bayes factor identities*

$$\beta_{q,p}(A:B) = \beta_{t,s}(A:B), \quad for \quad all \quad A, B \in \mathbf{A}^*, \quad and \tag{1}$$

$$\beta_{r,q}(A:B) = \beta_{s,p}(A:B), \quad for \quad all \quad A, B \in \mathbf{A}^*. \tag{2}$$

hold, then $r = t$*. Indeed, for all* $A \in \mathbf{A}^*$*, we have the explicit formula*

$$r(A) = t(A) = \frac{\left[\frac{q(A)s(A)}{p(A)}\right]}{\sum_{B \in \mathbf{A}^*} \frac{q(B)s(B)}{p(B)}} \tag{3}$$

Proof. The identity (1) is equivalent to

$$\frac{t(A)q(B)s(B)}{p(B)} = \frac{q(A)s(A)t(B)}{p(A)}, \text{ for all } A, B \in \mathbf{A}^*. \tag{4}$$

Fixing A in (4), and summing over all $B \in \mathbf{A}^*$ then yields (3) for $t(A)$, since

$$\sum_{B \in \mathbf{A}^*} t(B) = 1. \tag{5}$$

The proof of (3) for $r(A)$ follows from (2) in exactly analogous fashion. ■

Remark 3.1. If p, q, r, s and t are well-defined and in place and (1) and (2) hold, then, necessarily, the sum in the denominator of the right-hand side of (3) converges. If only p, q, and s are in place at the outset and the aforementioned sum converges, then (3) defines probabilities r and t satisfying (1) and (2). So (3) furnishes a recipe for constructing a probability measure r that would be the appropriate revision of q if, in the probabilistic state q, one were to undergo learning identical to that which prompted the revision of p to s. Similarly, (3) furnishes a recipe for constructing a probability measure t that would be the appropriate revision of s if, in the probabilistic state s, one were to undergo learning identical to that which prompted the revision of p to q. However, it is easy to construct examples where the sum in the denominator of (3) fails to converge. Then there exists no probability measure t satisfying (1) and no probability r satisfying (2). Thus from the perspective of the Uniformity Rule, it is impossible in the conceptual state reflected in s (respectively, q) to experience learning identical to that which prompted the revision of p to q (respectively, of p to s).

4 The Chan-Darwiche Metric

When Ω is finite the Uniformity Rule has intriguing connections with a metric on probability measures introduced by Chan and Darwiche (2002). Assume for simplicity that all probabilities are strictly coherent [1], and defined on all subsets of Ω. Define the *Chan-Darwiche distance* $CD(p,q)$ by

$$CD(p,q) := \log(R) - \log(r), \tag{1}$$

where

$$R := \max_{\omega\in\Omega} \frac{q(\omega)}{p(\omega)} \quad \text{and} \quad r := \min_{\omega\in\Omega} \frac{q(\omega)}{p(\omega)}. \tag{2}$$

It is straightforward to show that CD is a *metric* on the set of all strictly coherent probability measure on the power set of Ω, i.e., that

$$CD(p,q) \geqslant 0, \quad \text{with} \quad CD(p,q) = 0 \quad \text{iff} \quad p = q. \tag{3}$$

$$CD(p,q) = CD(q,p), \text{ and} \tag{4}$$

$$CD(p,q) \leqslant CD(p,p') + CD(p',q). \tag{5}$$

$CD(p,q)$ yields uniform bounds on the Bayes factors $\beta_{q,p}(A:B)$:

Theorem 2. *For all nonempty events $A, B \in 2^{\Omega}$,*

$$\exp(-CD(p,q)) \leqslant \beta_{q,p}(A,B) \leqslant \exp(CD(p,q)). \tag{6}$$

Proof. Suppose that max $\frac{q(\omega)}{p(\omega)}$ and min $\frac{q(\omega)}{p(\omega)}$ are attained, respectively, at $\omega = \omega_2$ and $\omega = \omega_1$. Then

[1] On the set all probability measures on the power set of Ω, CD is, strictly speaking, no longer a metric, since it can take the extended real number ∞ as a value. Indeed, with the stipulation that $\frac{0}{0} = 1$, $CD(p,q) < \infty$ iff p and q have exactly the same support, i.e., iff $\{\omega \in \Omega : p(\omega) > 0\} = \{\omega \in \Omega : q(\omega) > 0\}$

$$\frac{q(\omega_1)p(\omega)}{p(\omega_1)} \leqslant q(\omega) \leqslant \frac{q(\omega_2)p(\omega)}{p(\omega_2)}. \tag{7}$$

Summing (7) over all $\omega \in A$, and over all $\omega \in B$ yields

$$\frac{q(\omega_1)}{p(\omega_1)} \leqslant \frac{q(A)}{p(A)}, \qquad \frac{q(B)}{p(B)} \leqslant \frac{q(\omega_2)}{p(\omega_2)} \tag{8}$$

whence,

$$\frac{\left[\frac{q(\omega_1)}{p(\omega_1)}\right]}{\left[\frac{q(\omega_2)}{p(\omega_2)}\right]} \leqslant \frac{\Pi_{q,p}(A)}{\Pi_{q,p}(B)} \leqslant \frac{\left[\frac{q(\omega_2)}{p(\omega_2)}\right]}{\left[\frac{q(\omega_1)}{p(\omega_1)}\right]} \tag{9}$$

which is equivalent to (6) by (3) of section 2 above, (1), and (2). ■

Remark 4.1 Note that the bounds in (6) are sharp, the upper bound being attained when $A = \{\omega_2\}$ and $B = \{\omega_1\}$, and the lower bound when $A = \{\omega_1\}$ and $B = \{\omega_2\}$.

In view of (9) and the preceding remark, it is clear that $CD(p,q)$ may be equivalently defined by the formulas

$$CD(p,q) = \max_{\phi \neq A,B \subset \Omega} \log \beta_{q,p}(A:B) = \max_{\omega,\omega' \in \Omega} \log \beta_{q,p}(\{w\}:\{w'\}).^2 \tag{10}$$

The number $\log \beta_{q,p}(A:B)$ has been termed the *weight of evidence* by I.J. Good (1950). According to Good, Alan Turing was an enthusiastic advocate of using weights of evidence to measure the gain or loss of plausibility of one hypothesis vis-á-vis another as a result of the receipt of new evidence. Such weights were routinely used in the code-breaking work at Bletchley Park, where Good and Turing were colleagues during World War II (Jeffrey 2004). Indeed, Turing coined the term *ban* (after the town of Banbury, where the sheets were printed on which weights of evidence were recorded) for the unit weight of evidence, with logarithms taken to the base 10. One-tenth of a ban was termed a *deciban* (abbreviated *db*, in obvious analogy with acoustic notation). See Jeffrey (2004, pp. 32–32) and Good (1979) for further details.

Formula (10) thus provides a particularly salient formulation of the Chan-Darwiche distance, as well as an attractive and evocative unit of measurement. Moreover, there is a hand-in-glove fit between the Uniformity Rule and the Chan-Darwiche distance: If p is revised to q, and p' is revised to q', based on identical learning, and we construct q' in accord with the Uniformity Rule, then $CD(p,q) = CD(p',q')$. *So revisions based on identical learning, carried out acording to the dictates of the Uniformity Rule, move us the same CD-distance (i.e., the same number of decibans) from the priors in question.* As can be seen from the elementary example,

---

[2] Upon reading Chan-Darwiche (2002), I communicated this result to the authors, who incorporated it in Chan-Darwiche (2004).

	ω_1	ω_2		ω_1	ω_2
$p:$	$\frac{1}{2}$	$\frac{1}{2}$	$p':$	$\frac{2}{5}$	$\frac{3}{5}$
$q:$	$\frac{4}{5}$	$\frac{1}{5}$	$q':$	$\frac{8}{11}$	$\frac{3}{11}$,

where $CD(p,q) = CD(p',q') = 2\log 2$, this fails to be the case for other measure of distance, including the *Euclidean distance*

$$ED(p,q) := \left[\sum_\omega (p(\omega) - q(\omega))^2\right]^{\frac{1}{2}}, \tag{11}$$

the *variation distance*

$$\begin{aligned} V(p,q) &:= \max\{|p(A) - q(A)| : A \subset \Omega\} && (12) \\ &= \frac{1}{2}\sum_\omega |p(\omega) - q(\omega)|, && (13) \end{aligned}$$

the *Hellinger distance*

$$H(p,q) := \sum_\omega \left[\sqrt{p(\omega)} - \sqrt{q(\omega)}\right]^2, \tag{14}$$

and the *Kullback-Leibler information number*

$$KL(p,q) := \sum_\omega q(\omega)\log\left(\frac{q(\omega)}{p(\omega)}\right). \tag{15}$$

References

1. H. Chan and A. Darwiche (2002), A distance measure for bounding probabilistic belief change, *Proc.* 18^{th} *Int. Conf. on Artificial Intelligence* (AAAI-2002), Menlo Park, CA, AAAI Press, 539–545.
2. H. Chan and A. Darwiche (2004), A distance measure for bounding probabilistic belief change, *Int. J. Approx. Reas.* **38**, 149–174.
3. I. J. Good (1950), *Probability and the Weighing of Evidence,* Charles Griffin, London, chapter 6.
4. I. J. Good (1979), A. M. Turing's statistical work in World War II, *Biometrika* **66,** 393–396.
5. R. C. Jeffrey (2004), *Subjective Probability: The Real Thing,* Cambridge University Press, Cambridge.
6. A. Renyi (1970) *Foundations of Probability,* Holden-Day, San Francisco.

7. B. C. van Fraassen (1989), *Laws and Symmetry,* Clarendon Press, Oxford.
8. C. G. Wagner (1997), Old evidence and new explanation, *Philosophy of Science* **64**, 677–691.
9. C. G. Wagner (1999), Old evidence and new explanation II, *Philosophy of Science* **66**, 283–288.
10. C. G. Wagner (2001), Old evidence and new explanation III, *Philosophy of Science* **68** (Proceedings), S165–S175.
11. C. G. Wagner (2002), Probability kinematics and commutativity, *Philosophy of Science* **69**, 266-278.
12. C. G. Wagner (2003), Commuting probability revisions: the uniformity rule, *Erkenntnis* **59**, 349–364.