Milne's Argument for the $\log$-ratio measure*

Franz Huber†‡

---

penultimate version: please cite the paper in *Philosophy of Science*

* First received: December 2006, last revised version received: December 2007.

† Franz Huber, Formal Epistemology Research Group, Zukunftskolleg and Department of Philosophy, University of Konstanz, P.O. Box X906, 78457 Konstanz, Germany. E-mail: franz.huber@uni-konstanz.de

‡ I am grateful to Jiji Zhang for pointing out an error in a previous version of this paper, and to Branden Fitelson, Chris Hitchcock, and two anonymous referees for helpful comments and suggestions.

My research was supported by the Ahmanson Foundation and the German Research Foundation through its Emmy Noether Program.

Abstract

This paper shows that a slight variation of the argument in Milne (1996) yields the log-likelihood ratio l rather than the log-ratio measure r as "the one true measure of confirmation".

1 Introduction

In his (1996) Peter Milne shows that

$$r\,(H, E, B) = \log\left[\Pr\,(H \mid E \cap B) \,/\, \Pr\,(H \mid B)\right]$$

is "the one true measure of confirmation" in the sense that r is the one and only function satisfying the following five constraints on measures of confirmation C.

1. $C\,(H, E, B) \overset{>}{\underset{<}{=}} 0$ iff $\Pr\,(H \mid E \cap B) \overset{>}{\underset{<}{=}} \Pr\,(H \mid B)$.

2. $C\,(H, E, B)$ is a function of the values $\Pr\,(X \mid B)$ and $\Pr\,(Y \mid Z \cap B)$ assume on the at most 16 truth-functional combinations X, Y, Z of E and H.

3a. If $\Pr\,(E \mid H \cap B) < \Pr\,(F \mid H \cap B)$ and $\Pr\,(E \mid B) = \Pr\,(F \mid B)$, then $C\,(H, E, B) \leq C\,(H, F, B)$.

3b. If $\Pr\,(E \mid H \cap B) = \Pr\,(F \mid H \cap B)$ and $\Pr\,(E \mid B) < \Pr\,(F \mid B)$, then $C\,(H, E, B) \geq C\,(H, F, B)$.

4a. $C\,(H, E \cap F, B) - C\,(H, E \cap G, B)$ is determined by $C\,(H, E, B)$ and the difference $C\,(H, F, E \cap B) - C\,(H, G, E \cap B)$.

4.b If $C\,(H, E \cap F, B) = 0$, then $C\,(H, E, B) + C\,(H, F, E \cap B) = 0$.

5. If $\Pr\,(E \mid H \cap B) = \Pr\,(E \mid T \cap B)$, then $C\,(H, E, B) = C\,(T, E, B)$.

Among these (1), (3), and (5) concern the relation between confirmation and probability, while (2) and (4) concern confirmation alone. I will only be concerned with the former.

(1) is logically equivalent to

1+. $C\,(H, E, B) \overset{>}{\underset{<}{=}} 0$ iff $\Pr\,(E \mid H \cap B) \overset{>}{\underset{<}{=}} \Pr\,(E \mid B)$.

This makes clear that (1), (3), and (5) say what happens to confirmation $C\,(H, E, B)$ if various relations between the likelihood of hypothesis H on evidence E and background information B, $\Pr\,(E \mid H \cap B)$, and the prior of E given B, $\Pr\,(E \mid B)$, obtain.

(1+) is logically equivalent to

1*. $C(H,E,B) \underset{<}{\overset{>}{=}} 0$ iff $\Pr(E \mid H \cap B) \underset{<}{\overset{>}{=}} \Pr\left(E \mid \overline{H} \cap B\right)$.

Similarly, (3b) is logically equivalent to

3b*. If $\Pr(E \mid H \cap B) = \Pr(F \mid H \cap B)$ and $\Pr\left(E \mid \overline{H} \cap B\right) < \Pr\left(F \mid \overline{H} \cap B\right)$, then $C(H,E,B) \geq C(H,F,B)$.

While (1+) and (3b) focus on relations between likelihoods and priors, (1*) and (3b*) say the same thing by focusing on relations between likelihoods and what, following Fitelson (2007), we call *catch-alls*, $\Pr\left(E \mid \overline{H} \cap B\right)$. Let us see where this shift in focus takes us.

Regarding (3a) Milne (1996, 21) states that it "corrresponds more or less to the claim[...] that, other things being equal, a theory is better confirmed by evidence the more likely the theory makes the evidence." More than one thing can be equal, though. Often not all of them can be equal simultaneously. According to (3a) the prior of the evidence is held fixed: $\Pr(E \mid B)$ is equal to $\Pr(F \mid B)$.

Consider the catch-all counterpart

3a*. If $\Pr(E \mid H \cap B) < \Pr(F \mid H \cap B)$ and $\Pr\left(E \mid \overline{H} \cap B\right) = \Pr\left(F \mid \overline{H} \cap B\right)$, then $C(H,E,B) \leq C(H,F,B)$,

According to (3a*) the catch-all, the likelihood of $\overline{H}$ on the evidence, is held fixed: $\Pr\left(E \mid \overline{H} \cap B\right)$ is equal to $\Pr\left(F \mid \overline{H} \cap B\right)$. Given that the theory makes the one evidence more likely than the other, i.e. $\Pr(E \mid H \cap B) < \Pr(F \mid H \cap B)$, not both of these other things can be equal.

Regarding (5) Milne (1996, 22) says that it "is a weak consequence of the Likelihood Principle":

> In comparing the evidential bearing (relative to background knowledge B) of E on the hypotheses H and T we need consider only $\Pr(E \mid H \cap B)$ and $\Pr(E \mid T \cap B)$. (Milne 1996, 22)

Note that, in the presence of (1-4), (5) is equivalent to the otherwise stronger

5+. If $\Pr(E \mid H \cap B) = \Pr(F \mid T \cap B)$ and $\Pr(E \mid B) = \Pr(F \mid B)$, then $C(H,E,B) = C(T,F,B)$.

This is so because $r(H,E,B)$ satisfies (5+).

Here is the catch-all counterpart of (5+):

5*. If $\Pr(E \mid H \cap B) = \Pr(F \mid T \cap B)$ and $\Pr\left(E \mid \overline{H} \cap B\right) = \Pr\left(F \mid \overline{T} \cap B\right)$, then $C(H, E, B) = C(T, F, B)$.

Let us rename (2) and (4) by (2*) and (4*), respectively. Then things can be put as follows. In the presence of (2) and (4), the conjunction of (1), (3), and (5) says that $C(H, E, B)$ is a function of the likelihood of H on E, $\Pr(E \mid H \cap B)$, and the prior of E, $\Pr(E \mid B)$ – increasing with the former, and decreasing with the latter.

In the presence of (2*) and (4*), the conjunction of (1*), (3*), and (5*) says that $C(H, E, B)$ is a function of the likelihood of H on E, $\Pr(E \mid H \cap B)$, and the catch-all, i.e. the likelihood of $\overline{H}$ on E, $\Pr\left(E \mid \overline{H} \cap B\right)$ – increasing with the former, and decreasing with the latter.

2 Catch-alls or Priors?

A variation of Milne's proof (presented in Appendix 1) shows that

$$l(H, E, B) = \log\left[\Pr(E \mid H \cap B) / \Pr\left(E \mid \overline{H} \cap B\right)\right]$$

is another true measure of confirmation in the sense that l is the one and only function satisfying (1*-5*).

As Fitelson (2001, 29) observes, l satisfies (1-4). It is worth noting that r satisfies (1*-4*). So the difference between r and l lies in (5) versus (5*): l does not satisfy (5), and r does not satisfy (5*).

Thus r and l agree that confirmation depends on the likelihood of H on E, $\Pr(E \mid H \cap B)$, and one other factor. They also agree on how to compare the likelihood of H on E to the other factor, viz. by taking logarithms of ratios. What they disagree about is the other factor the likelihoods of H on E should be compared to: r says the other factor is the prior of the evidence E, $\Pr(E \mid B)$, while l says it is the catch-all, i.e. the likelihood of $\overline{H}$ on the evidence E, $\Pr\left(E \mid \overline{H} \cap B\right)$.

3 Odds or Probabilities?

Things can be put differently still. Let $O(H \mid B)$ and $O(H \mid E \cap B)$ stand for the prior and posterior odds of H, respectively,

$$O(H \mid B) = \frac{\Pr(H \mid B)}{\Pr\left(\overline{H} \mid B\right)} \text{ and } O(H \mid E \cap B) = \frac{\Pr(H \mid E \cap B)}{\Pr\left(\overline{H} \mid E \cap B\right)}.$$

Then, as Joyce (2003, table 5) observes,

$$r(H,E,B) = \log\left[\frac{\Pr(H \mid E \cap B)}{\Pr(H \mid B)}\right] \text{ and } l(H,E,B) = \log\left[\frac{O(H \mid E \cap B)}{O(H \mid B)}\right].$$

Seen this way r and l agree that it is differences between priors and posteriors that matter for confirmation. They also agree on how to measure those differences, viz. by taking the logarithm of the ratio of posterior over prior. What they disagree about is, to speak with Joyce (2003, sct. 3), the question whether we should consider differences in "total evidence" as measured by $\Pr(H \mid E \cap B)$ and $\Pr(H \mid B)$, or differences in "net evidence" as measured by $O(H \mid E \cap B)$ and $O(H \mid B)$.

4 Conclusion

Milne 1996 presents his argument as a desideratum/explicatum argument for r as opposed to other measures of confirmation. His confirmation theoretic monism presupposes that there is one and only one true measure of confirmation. Joyce 2003, sct. 3, on the other hand, favors a confirmation theoretic pluralism according to which, among others, each of r and l "measures an important evidential relationship, but that the relationships they measure are importantly different."[1]

This pluralistic view suggests to view Milne's 1996 argument and the above variation not so much as arguments for or against one particular measure of confirmation. Rather, they can be viewed as *characterizations* that tell us, descriptively, what particular measures focus on, without telling us, prescriptively, what we *should* focus on. The latter, normative question seems to be beyond the reach of desiderata/explicata approaches, but to belong to the realm of means-ends epistemology or epistemic consequentialism (Percival 2002; Stalnaker 2002) as exemplified, for probability, by Joyce (1998), and, for confirmation, by Huber (2005).

[1] Actually Joyce 2003 considers e^r and e^l, that is, r and l without the $\log$.

Appendix 1: A Variation of Milne's (1996) Proof

The following proof is entirely due to Milne 1996, appendix 1, although all errors are, of course, mine.

(2*) entails that $C(H, E, B)$ is a function of $\Pr(E \mid H \cap B)$, $\Pr\left(E \mid \overline{H} \cap B\right)$, and $\Pr(H \mid B)$. (5*) entails that $C(H, E, B)$ is independent of $\Pr(H \mid B)$. So $C(H, E, B) = F\left(\Pr(E \mid H \cap B), \Pr\left(E \mid \overline{H} \cap B\right)\right)$ for some $F : [0,1]^2 \to \Re^*$, where $\Re^* = \Re \cup \{\pm\infty\}$.

(1*) entails that $F(x, x) = 0$ for all $x \in [0, 1]$. As

$$\begin{aligned}\Pr(E \cap F \mid H \cap B) &= \Pr(E \mid H \cap B) \cdot \Pr(F \mid E \cap H \cap B)\\ \Pr\left(E \cap F \mid \overline{H} \cap B\right) &= \Pr\left(E \mid \overline{H} \cap B\right) \cdot \Pr\left(F \mid E \cap \overline{H} \cap B\right),\end{aligned}$$

(4*) entails that there is a possibly partial $G : \Re^{*2} \to \Re^*$ such that for all $x, y, z_1, z_2, w_1, w_2 \in [0, 1]$

$$F(x \cdot z_1, y \cdot w_1) - F(x \cdot z_2, y \cdot w_2) = G(F(x, y), F(z_1, w_1) - F(z_2, w_2)). \quad (1)$$

The range of F is assumed to be a real interval.

$F(1, 1) = 0$, and so

$$F(x \cdot z, y \cdot w) - F(x, y) = G(F(x, y), F(z, w)), \quad (2)$$

which yields $G(0, u) = u$ and $G(u, 0) = 0$ for $x = y = 1$ and $z = w = 1$, respectively. Equation (2) and the previous equation give us

$$\begin{aligned}F(x \cdot z, x \cdot w) &= F(x, x) + G(F(x, x), F(z, w))\\ &= F(z, w).\end{aligned}$$

If $x/z = y/w$, then $F(x, z) = F\left(\frac{z}{w} \cdot y, \frac{x}{y} \cdot w\right)$ and $z/w = x/y$, or $F\left(\frac{w}{z} \cdot x, \frac{y}{x} \cdot z\right) = F(y, w)$ and $w/z = y/x$. Hence $F(x, z) = F(t \cdot y, t \cdot w)$ or $F(t \cdot x, t \cdot z) = F(y, w)$ for some $t \in [0, 1]$.

Assume without loss of generality that $F(x, z) = F(t \cdot y, t \cdot w) = F(y, w)$ for $t \in [0, 1]$. Then $C(H, E, B) = F(x, z) = F(y, w)$ with $x/z = y/w$, and so

$$C(H, E, B) = H\left(\Pr(E \mid H \cap B) / \Pr\left(E \mid \overline{H} \cap B\right)\right) \text{ for some } H : \Re_{\geq 0} \to \Re^*.$$

For $z_2 = w_2 = 1$ Equation (1) entails

$$H(x \cdot y) = H(x) + G(H(x), H(y)) = H(y) + G(H(y), H(x)). \quad (3)$$

This and Equation (1) give us

$$\begin{aligned} G\left(H\left(x\right), H\left(y\right)\right) - G\left(H\left(x\right), H\left(z\right)\right) &= H\left(x \cdot y\right) - H\left(x \cdot z\right) &\quad (4)\\ &= G\left(H\left(x\right), H\left(y\right) - H\left(z\right)\right), &\quad (5) \end{aligned}$$

which yields

$$G\left(t, u+v\right) = G\left(t, u\right) + G\left(t, v\right).$$

For integers m, n and $u \cdot m/n$ in the range of F so that $(t, u \cdot m/n)$ is in the domain of G, we thus have $G\left(t, u \cdot m/n\right) = \frac{m}{n} \cdot G\left(t, u\right)$. (3a*) entails that $G\left(t, u\right) \leq G\left(t, v\right)$ if $u \leq v$. So for all reals r with $u \cdot r$ in the range of F so that $(t, u \cdot r)$ is in the domain of G, $G\left(t, u \cdot r\right) = r \cdot G\left(t, u\right)$. Hence $G\left(t, u\right) = u \cdot g\left(t\right)$ for some $g : \Re^{*} \rightarrow \Re_{\geq 0}$ (at this point Milne refers to Aczél 1966, 31-34).

Equation (3) entails

$$H\left(x \cdot y\right) - H\left(x \cdot z\right) = H\left(y\right) - H\left(z\right) + G\left(H\left(y\right), H\left(x\right)\right) - G\left(H\left(z\right), H\left(x\right)\right)$$

and so Equation (5) gives us

$$g\left(H\left(x\right)\right) \cdot \left(H\left(y\right) - H\left(z\right)\right) = H\left(y\right) - H\left(z\right) + H\left(x\right) \cdot \left(g\left(H\left(y\right)\right) - g\left(H\left(z\right)\right)\right). \quad (6)$$

(1*) entails $H\left(1\right) = 0$ and that H is not constant, which implies that $g\left(0\right) = 1$. For $H\left(x\right) \neq 0$ Equation (6) entails

$$g\left(H\left(y\right)\right) - g\left(H\left(z\right)\right) = \left(g\left(H\left(x\right)\right) - 1\right) \cdot \left(H\left(y\right) - H\left(z\right)\right) / H\left(x\right)$$

The left-hand side is independent of x, and so

$$g\left(H\left(x\right) - 1\right) / H\left(x\right) = k$$

for some constant $k \in \Re^{*}$.

From Equation (3) we have

$$\begin{aligned} H\left(x \cdot y\right) &= H\left(x\right) + G\left(H\left(x\right), H\left(y\right)\right)\\ &= H\left(x\right) + H\left(y\right) \cdot g\left(H\left(x\right)\right)\\ &= H\left(x\right) + H\left(y\right) \cdot \left(H\left(x\right) \cdot k + 1\right)\\ &= H\left(x\right) + H\left(y\right) + k \cdot H\left(x\right) \cdot H\left(y\right). \end{aligned}$$

(4b*) entails that $H\left(x\right) + H\left(y\right) = 0$ if $H\left(x \cdot y\right) = 0$. $k = 0$, since it is possible that $H\left(x \cdot y\right) = 0$ while $H\left(x\right) \neq 0$ and $H\left(y\right) \neq 0$ (it suffices to consider a case where E is positively relevant for H, F is negatively relevant for H, and $E \cap F$ is independent of H in the sense of some $\Pr$ – note that this argument would be problematic if the underlying probability space were fixed). Hence

$$H\left(x \cdot y\right) = H\left(x\right) + H\left(y\right), \tag{7}$$

and so $H\left(x^{m/n}\right) = m/n \cdot H\left(x\right)$ for integers m, n. (3*) entails that $H\left(x\right) \leq H\left(y\right)$ if $x \leq y$, and so $H\left(x^{r}\right) = r \cdot H\left(x\right)$ for all $r \in \Re$. (As Milne notes, no assumptions about the domain of H need be made this time, because any number in $\Re^{*}$ can be the ratio of two probabilities – again, note that this argument would be problematic if the underlying probability space were fixed). Therefore $H\left(x\right) = c \cdot \log x$ for some constant c (at this point Milne refers to Aczél 199, 39-41) that has to be positive in view of (1*) and equals 1 by a suitable choice of the base of $\log$. Hence $C\left(H, E, B\right) = \log\left(\Pr\left(E \mid H \cap B\right) / \Pr\left(E \mid \overline{H} \cap B\right)\right)$. $\square$

Appendix 2: Fitelson's (2001) Objection

Fitelson (2001, 28) notes that "Milne's argument *implicitly* requires that the probability function Pr [...] satisfy some rather strong, unmotivated, and unintuitive constraints." In particular, "Milne's argument makes use of certain theorems [...] which force the probability function Pr (and, hence, the spaces over which the measure [of confirmation C] is defined) to satisfy various kinds of *continuity* conditions" (Fitelson 2001, 28, fn. 43). For a discussion of these Fitelson refers to Halpern (1999a, 1999b), where it is shown that Cox's (1946) theorem does not hold in finite domains.

I think it is perfectly reasonable for Milne (and proponents of the above variation of his argument) to require the domain of the measure of confirmation C to be infinite. As Halpern (1999a, sct. 5; 1999b, theorem 5) observes, one response is to say that we are not interested in a single domain in isolation, but a notion of belief or confirmation (in his or our case, respectively) that applies uniformly in all domains.

But suppose we are in fact interested in just one single field of propositions $\mathcal{A}$ over which our measure of confirmation C is defined. Suppose further $\mathcal{A}$ is finite. Even then the domain of C is uncountable, provided we assume C does not vary with the underlying probability measure Pr. That is, we only have to think of C as a mapping of probability spaces (and not propositions without probabilities) into the reals, and take its domain to be the set of all probability spaces $\langle \mathcal{A}, \mathrm{Pr} \rangle$ (for the fixed $\mathcal{A}$ from above). As far as I can tell, this assumption is implicit in all discussions of incremental confirmation. Rejecting it means to use different measures of confirmation for different probability measures on one fixed domain, rather than uniformly using the same measure of confirmation.

However, the assumption Milne (1996, 24) actually makes is that the *range* of C forms a real interval. This implies that the domain of C is uncountably infinite. As argued, the latter assumption is reasonable for Milne to make. Obviously it is another question whether the former is, too.

References

[1] Aczél, János (1966), *Lectures on Functional Equations and Their Applications*. New York, London: Academic Press.

[2] Cox, Richard T. (1946), "Probability, Frequency, and Reasonable Expectation", *American Journal of Physics* 14: 1-13.

[3] Fitelson, Branden (2001), *Studies in Bayesian Confirmation Theory*. Ph.D. Dissertation. Madison, WI: University of Wisconsin-Madison.

[4] – – – (2007), "Likelihoodism, Bayesianism, and Relational Confirmation", *Synthese* 156: 473-489.

[5] Halpern, Joseph Y. (1999a), "A Counterexample to Theorems of Cox and Fine", *Journal of AI Research* 10: 67-85.

[6] – – – (1999b), "Cox's Theorem Revisited", *Journal of AI Research* 11: 429-435.

[7] Huber, Franz (2005), "What Is the Point of Confirmation?", *Philosophy of Science* 72: 1146-1159.

[8] Joyce, James M. (1998), "A Nonpragmatic Vindication of Probabilism", *Philosophy of Science* 65: 575-603.

[9] – – – (2003), Bayes' Theorem. In E.N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*.

[10] Milne, Peter (1996), "$\log [\Pr(H \mid E \cap B) / \Pr(H \mid B)]$ Is the One True Measure of Confirmation", *Philosophy of Science* 63: 21-26.

[11] Percival, Philip (2002), "Epistemic Consequentialism", *Supplement to the Proceedings of the Aristotelian Society* 76 (1): 121-151.

[12] Stalnaker, Robert C. (2002), "Epistemic Consequentialism", *Supplement to the Proceedings of the Aristotelian Society* 76 (1): 153-168.