In *Analysis* 69 (2009): 630-35

AN ARGUMENT FOR THE LIKELIHOOD-RATIO MEASURE OF CONFIRMATION

José L. Zalabardo, UCL

In the recent literature on confirmation, there are two leading approaches to the provision of a probabilistic measure of the degree to which a hypothesis is confirmed by evidence. The first is to construe the degree to which evidence E confirms hypothesis H as a function that is directly proportional to p(H | E) and inversely proportional to p(H). I shall refer to this as the *probability approach*. The second approach construes the notion as a function that is directly proportional to the true-positive rate—the probability of the evidence if the hypothesis is true, p(E | H)—and inversely proportional to the false-positive rate—the probability of the evidence if the hypothesis is false, p(E | ~H). These reverse conditional probabilities—of the evidence on the truth or falsehood of the hypothesis—are sometimes known as likelihoods. I shall refer to the approach to confirmation that uses them as the *likelihood approach*.

For each of these approaches, there are two obvious options as to how to define the confirmation function. The first is to define it as the *difference* between the magnitude that is treated as increasing confirmation (p(H | E) or p(E | H)) and the magnitude that is treated as decreasing confirmation (p(H) or p(E | ~H)). The second is to define confirmation as the quotient of these two magnitudes.

This yields four different measures of confirmation, represented in the following table:

	Differences	*Ratios*
Probabilities	$PD(H, E) = P(H \mid E) - P(H)$[1]	$PR(H, E) = \dfrac{p(H \mid E)}{p(H)}$[2]
Likelihoods	$LD(H, E) = p(E \mid H) - p(E \mid {\sim}H)$[3]	$LR(H, E) = \dfrac{p(E \mid H)}{p(E \mid {\sim} H)}$[4]

[1] See (Earman 1992; Gillies 1986; Jeffrey 1992).
[2] See (Milne 1996). Milne argues for log(PR) as the true measure of confirmation. Log(PR) is ordinally equivalent to PR, i.e. PR(H, E) > PR(H*, E*) iff log(PR(H, E)) > log(PR(H*, E*)).
[3] LD hasn't received much support among confirmation theorists but it is the account of weak evidential support put forward by Robert Nozick (Nozick 1981).

All these measures agree on whether or not E provides some support for H. Intuitively, we want to say that E provides some support for H just in case learning that E is true would make us assign a higher probability to H, i.e. when $P(H \mid E) > P(H)$. Clearly this will be the situation if $PD(H, E) > 0$ or $PR(H, E) > 1$. And on the likelihood measures we have the same situation. Using Bayes' Theorem we can easily show that

$$p(H \mid E) > p(H) \text{ iff } p(E \mid H) > p(E \mid \sim H)$$

Hence we have that PD(H, E) > 0 iff LD(H, E) > 0, and PR(H, E) > 1 iff LR(H, E) > 1.

But while the four measures agree on whether or not E confirms H, they disagree on the degree to which E confirms H. For any two of these measures, there are hypotheses H, H* and evidence E, E* such that the two measures disagree on whether E supports H to a higher degree than E* supports H*. In other words, if we take these measures to order hypothesis-evidence pairs by degree of confirmation (the degree to which the evidence confirms the hypothesis), each of them generates a different ordering. Our choice between these accounts of confirmation should be dictated by the plausibility of the orderings they generate.

We can argue in this way that the probability-difference measure is inferior to the probability-ratio alternative. George Schlesinger (Schlesinger 1995) has provided a cogent illustration of the superiority of PR over PD. Schlesinger asks us to compare two scenarios. In the first, we consider a type of aircraft which is regarded as extremely safe, with a $1/10^9$ probability of crashing in a single flight. However, further inspection of the structure of the aircraft reveals a flaw as a result of which the probability of one of these planes crashing is actually 1/100. The second scenario concerns troops landing gliders behind enemy lines. We start from the assumption that someone taking part in one of these operations has a 26%

[4] See (Fitelson 2001; Roush 2005). Fitelson argues for log(LR), which is ordinally equivalent to LR.

chance of perishing, but one day the commander announces that due to peculiar weather conditions the risk has increased from 26% to 27%.

As Schlesinger argues, the degree to which the inspection of the aircraft confirms the hypothesis of a plane crash is intuitively much higher than the degree to which the unusual weather conditions confirm the hypothesis of a glider mission resulting in death. This intuition is preserved by the probability ratio account. If H is the hypothesis of a plane crash and E is the inspection of the aircraft, PR(H, E) is 1/100 divided by $1/10^9$, i.e. 10^7. On the other hand, if H* is the hypothesis of a glider mission ending in tragedy, and E* is the unusual weather, PR(H*, E*) is 27/100 divided by 26/100, i.e. 1.038. PR(H, E) is much higher than PR(H*, E*). The probability-difference measure, by contrast, does not preserve the intuition. PD(H, E) is $1/100 – 1/10^9$, i.e. 1.00999… and PD(H*, E*) is 27/100 – 26/100, i.e. 1.01. PD(H, E) is slightly smaller than PD(H*, E*). I regard this as a powerful argument for preferring PR over PD as a measure of confirmation.

In the case of the likelihood approach, the superiority of the ratio measure over the difference alternative can be conclusively established. The problem for the likelihood-difference measure concerns the intuitive thought that the degree to which different bodies of data confirm a given hypothesis is directly proportional to the conditional probability of the hypothesis on the data. I.e. E confirms H to a greater degree than E* does just in case p(H | E) is greater than p(H | E*). Satisfaction of this requirement is a plausible adequacy condition for an account of confirmation. This condition is satisfied by the likelihood-ratio measure, as well as by both probability measures. However, as Fitelson has shown (Fitelson 2001), the likelihood-difference measure violates this condition.[5] I regard this as a conclusive reason for rejecting LD as a measure of confirmation.

---

[5] As Fitelson shows, the same shortcoming is present in another popular account of confirmation: p(H | E) – p(H | ~E). This account has been defended by (Joyce 1999; Christiensen 1999)

This leaves us with two alternatives: the probability ratio and the likelihood ratio. I am going to argue for the superiority of LR over PR. My strategy will consist in presenting two intuitively plausible adequacy conditions for a theory of confirmation and arguing that one of these conditions would be satisfied if we adopted LR as our measure of confirmation, but not if we adopted PR.

My adequacy conditions will relate to the importance for the degree to which some evidence confirms a hypothesis of the ratio of false positives. We can see the point first by considering the degree to which two pieces of evidence confirm a single hypothesis. Consider the degree to which a diagnosis of asthma is supported by two standard symptoms: wheezing and a dry cough. Both symptoms have a very high ratio of true positives: most people with asthma wheeze and most people with asthma have a dry cough.[6] Let's assume that the true-positive ratio is identical in each case. However, with respect to false positives, the two symptoms rate very differently. Very few people who don't have asthma wheeze, whereas quite a few people who don't have asthma have a dry cough. Hence wheezing and a dry cough have the same ratio of true positives, while wheezing has a significantly lower ratio of false positives than a dry cough does.

I want to suggest that a plausible theory of confirmation should yield the result that the features of the example that we have described suffice for concluding that wheezing confirms a diagnosis of asthma to a higher degree than a dry cough does. In general, a plausible theory of confirmation should take same true-positive ratio with a lower false-positive ratio as a sufficient condition for a higher degree of confirmation. This yields my first adequacy condition for a theory of confirmation:

(1) If $p(E \mid H) = p(E^* \mid H)$ and $p(E \mid {\sim}H) < p(E^* \mid {\sim}H)$, then E confirms H to a higher degree than E* does.

[6] I'm grateful to Jonathan Grigg and Sabine Kleinert for the medical examples.

Clearly treating LR as our measure of confirmation would satisfy (1), since if p(E | H) = p(E* | H) and p(E | ~H) < p(E* | ~H), LR(H, E) and LR(H, E*) have the same numerator, but LR(H, E) has a smaller denominator than LR(H, E*) does.

However, this gives no advantage to LR over PR, as taking the latter as our measure of confirmation would also satisfy (1). To see this, notice that Bayes' Theorem enables us to reformulate PR in the following way:

$$\text{PR(H, E)} = \frac{p(E \mid H)}{p(E)}.$$

Now, on this formulation, PR(H, E) and PR(H, E*) have the same numerator. Furthermore, the mixing principle tells us that p(E) = p(H) · p(E | H) + p(~H) · p(E | ~H). Hence, if p(E | H) = p(E* | H) and p(E | ~H) < p(E* | ~H), it follows that p(E) < p(E*). Therefore, on our reformulation, PR(H, E) has a smaller denominator than PR(H, E*) does. Since they have the same numerator, PR(H, E) is greater than PR(H, E*), as required by (1).

I want to argue next that intuition sanctions the same verdict when we are comparing the degree to which two pieces of evidence confirm different hypotheses. Compare now the degree to which wheezing confirms a diagnosis of asthma with the degree to which weight loss confirms a diagnosis of lung cancer. Most lung cancer patients lose weight. Hence, as evidence for lung cancer, weight loss has a very high true-positive ratio, which we can assume to be identical to the true-positive ratio of wheezing or a dry cough as evidence for asthma. However, lots of people who don't have lung cancer also lose weight. Hence, as evidence for lung cancer, weight loss has quite a high ratio of false positives, which we can assume to be identical to the false-positive ratio of a dry cough as evidence for asthma.

I want to suggest that intuition dictates that when we compare the degree to which wheezing supports a diagnosis of asthma with the degree to which weight loss supports a diagnosis of lung cancer, we should draw the same conclusion as when we compared wheezing and a dry cough as evidence for asthma. Wheezing has the same ratio of true positives with respect to asthma as weight loss does with respect to lung cancer, but the former has a lower false-positive ratio than the latter does. Hence a plausible account of confirmation should treat wheezing as confirming the asthma hypothesis to a higher degree than weight loss confirms the lung cancer hypothesis. In general, a plausible theory of confirmation should take same true-positive ratio with a lower false-positive ratio as a sufficient condition for a higher degree of confirmation, even when we are dealing with different hypotheses. This yields my second adequacy condition for a theory of confirmation:

(2) If $p(E \mid H) = p(E^* \mid H^*)$ and $p(E \mid {\sim}H) < p(E^* \mid {\sim}H^*)$, then E confirms H to a higher degree than E* confirms H*.

Clearly, the same reasoning that we gave for (1) establishes that treating LR as our measure of confirmation would satisfy (2). Notice, however, that LR might not be the only measure of confirmation that satisfies (2). (2) says nothing about the orderings that the confirmation function should yield when the true-positive ratios are different. The question that we need to ask is whether treating PR as our measure of confirmation would also satisfy (2). I am going to argue that this question should be answered in the negative.

Notice that, by our reformulation of PR, to show this it would suffice to establish that $p(E) > p(E^*)$ is compatible with $p(E \mid H) = p(E^* \mid H^*)$ and $p(E \mid {\sim}H) < p(E^* \mid {\sim}H^*)$. In these cases, (2) would dictate that E confirms H to a higher degree than E* confirms H*, but treating PR as our measure of confirmation would yield the opposite result. I am going to argue that this is the situation.

Suppose that E, as evidence for H, and E*, as evidence for H*, have the same high ratio of true positives, say 95%, but that E has a lower ratio of false positives (2.5%) than E* does (20%). Hence we have:

$$p(E \mid H) = p(E^* \mid H^*) = 19/20$$

$$p(E \mid {\sim}H) = 1/40$$

$$p(E^* \mid {\sim}H^*) = 1/5$$

Is this compatible with $p(E) > p(E^*)$? Clearly, if we assumed that $p(H) = p(H^*)$, the situation would be the same as in the one-hypothesis scenario. The mixing principle would entail that $p(E) < p(E^*)$.

However once we remove the assumption that $p(H) = p(H^*)$, the result no longer follows. Suppose, for example, that $p(H) = 1/3$ and $p(H^*) = 1/20$. Then the mixing principle tells us that $p(E) = 1/3$, while $p(E^*) = 0.2375$. Using our reformulation of PR, we have that $PR(H, E) = 2.85$ and $PR(H^*, E^*) = 4$. Hence in this case, although $p(E \mid H) = p(E^* \mid H^*)$ and $p(E \mid {\sim}H) < p(E^* \mid {\sim}H^*)$, taking PR as our measure of confirmation would yield the result that E confirms H to a lesser degree than E* confirms H*, in violation of our adequacy condition.

REFERENCES

Christiensen, David. 1999. Measuring Confirmation. *Journal of Philosophy* 96:437-461.

Earman, John. 1992. *Bayes or Bust?: A Critical Examination of Bayesian Confirmation Theory*. Cambridge, Massachusetts: MIT Press.

Fitelson, Branden. 2001. A Bayesian Account of Independent Evidence with Applications. *Philosophy of Science* 68:123-140.

———. 2001. Studies in Bayesian Confirmation Theory. PhD Dissertation, University of Wisconsin, Madison.

Gillies, Donald. 1986. In Defense of the Popper-Miller Argument. *Philosophy of Science* 53:110-113.

Jeffrey, Richard. 1992. *Probability and the Art of Judgment*. Cambridge: Cambridge University Press.

Joyce, James. 1999. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

Milne, Peter. 1996. log[p(h/eb)/p(h/b)] is the One True Measure of Confirmation. *Philosophy of Science* 63:21-26.

Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge, Massachusetts: Harvard University Press.

Roush, Sherrilyn. 2005. *Tracking Truth*. Oxford: Oxford University Press.
Schlesinger, George N. 1995. Measuring Degrees of Confirmation. *Analysis* 55:208-212.