# A hyperbolic sum rule for probability: solving the recursive “Chicken & Egg” problem (Response to Reviewers, round 2)

M.C.Parker, C.Jeynes, 4th July 2024

# REVIEWER REPORTS (25th September 2023; 2nd round)

## Reviewer #1:

***2nd Review of A hyperbolic sum rule for probability: solving the recursive “Chicken & Egg” problem***

Let me thank the authors for their revision. Their line of thinking has become more clear to me. I still argue below that it is wrong, and therefore **I can not recommend publication**.

I am relieved to see that the authors now accept the validity what they call the Conventional Sum Rule (CSR) of probability theory for any “non-recursive probability”. This means we can still benefit from the tremendous amount of technological achievements conventional probability theory enabled. For “recursive probabilities” (or maybe more precisely, the recursive interdependence of the arguments of the probability) they claim that their Hyperbolic Sum Rule (HSR) holds. In their view, it seems, there are two valid sum rules, CSR and HSR, and it is a property of the probability (or their arguments), namely being “recursive” or not, that decides which of those has to be applied.

Unfortunately, the authors avoid to give any mathematical rigorous definition of “recursive probabilities” that would allow to define the set of recursive probabilities, but only one example. Such a definition would be essential, as otherwise it is unclear when which formula has to be applied. I will argue below that **the set of recursive probabilities is empty** by showing that their show case example for a recursive probability belongs into the realm of conventional probabilities for which the CSR holds. When the best example is not working, it is reasonable to assume that other attempts to construct a recursive probability wont work either unless proven otherwise. At least a clear mathematical definition of this set and a working example need to be provided. This is currently not the case

Furthermore, I still claim that **the derivation of the HSR consists of incorrect steps**. The revised version of the first “derivation” helped me to spot at least one clear misconception. To my further relief I saw that the second referee is quiet competent in the details of the Cox theorem derivation and leave it to them to show the incorrect step in the HSR derivation based on that approach (I believe it is the ad-hoc replacement of Cox’s *Sm*(x) = 1 − x*m* for the probability of a negated statement with something else, which violates Cox’s desiderata).

There are a number of statements in the manuscript I would argue about (e.g. the statement “the HSR encompasses a wider domain of physical application that includes the possibility of (mutual) recursion between phenomena, whereas the CSR a priori excludes the possibility of recursion.”: HSR can not hold for only potential recursive probabilities, as a potential recursive probability can turn out to be not recursive and then the CSR holds, which is not a limiting case of the HSR). However, given that there are still very mayor concerns, I do not address these lesser pressing issues here.

**The set of recursive probabilities is empty**: I only can guess what “recursive” could mean to the authors from examining two vague data points:

First, the manuscript contains the statement that “We first note that the HSR is well-behaved in that it yields the same results as the CSR in the limit of non-recursion (when the events A and B are mutually exclusive)”. This statement seems to hint (but, yes, does not formally imply) that any set of mutually non-exclusive events are recursive and therefore not subject to the CSR. This would obliviously be wrong. The example from my first report of A and B being the first and second toss of a fair coin showing head, respectively, are mutually non-exclusive events, for which the authors agreed that the CSR and not the HSR applies. Thus, this statement is therefore at least misleading and needs to be made clearer and – more important – it does not provide the urgently needed definition of “recursion”.

Second, there is only the demonstrator for recursive probabilities by the example in which the events A and B correspond to Alice and Bob being elected in two nearby countries, respectively. There certainly can be an interdependence of the two events as there easily could be confounding events that influenced both, for example Alice and Bob having made public statements about each other prior to the elections.

The CSR for probabilities is correctly stated as

P(A + B) = P(A) + P(B) − P(A|B)P(B),

but then it is argued that P(A|B) = P(A) as we do not have any information that lets us assume something else. However, the authors argue strongly that information is physical and not subjective, so our lack of information is irrelevant her, as the physical world should determine what P(A|B) is and not our knowledge about it, and this physical probability needs to be entered into the formula. It expresses exactly the interdependence of A and B. (To make this point clearer: If I do not know the physical masses of electrons and a protons, I am still not allowed to assume them to be the same.) Only if probabilities would be subjective, lack of knowledge about interdependence could (but needs not: If being aware of the possibility of influences, all possible influences have to be taken into account and marginalized. This can lead to P(A|B) = P(A) if the considered possibilities of influences are somehow perfectly balanced, but needs not necessarily) result in P(A|B) = P(A), but the very premise of the paper is that probabilities are physical.

To conclude, the CSR is perfectly well able to handle the mutual dependence of this as “recursive” labeled case, so that it can fairly be assumed to do so in any other case one might come up. If I am not shown a single example that can not be dealt with CSR, but requires HSR, I continue to claim that the set of “recursive” probabilities is empty and HSR therefore meaningless. It would have been the obligation of the authors to give such a case.

***The derivation of the HSR consists of incorrect steps:*** I will refer to the equation numbers of the revised manuscript. Eq. 2d is correct according to the CSR and the product rule. Eq. 4a would be correct if the introduced brackets are a split of the unity, [f(. . .)]A + [g(. . .)]B := a f(. . .) + b g(. . .), with a + b = 1, as in the case of Eq. 4a f = g and thus [f(. . .)]A + [g(. . .)]B := (a + b) f(. . .) = f(. . .) which recovers Eq. 2d. And indeed the text indicates that the “prior priors” a = P(”A is prior to B”) and a = P(”B is prior to A”) are two probabilities for (possibly) mutual exclusive and exhaustive possibilities, that need add up to one. I do not want to enter the discussion of what ”A is prior to B” could mean (I think there is a confusion between the concepts “prior knowledge” and “prior cause” happening here, but that is not relevant for my point, and I accept for the moment the paper’s premise that probabilities are physical, which might have caused this confusion). So far so good.

However, in the calculation of the paper these two factors are set to a = P(A|BC) and b = P(B|AC) in Eq. 4b implying P(”A is prior to B”) ≡ P(A|BC) and P(”B is prior to A”) ≡ P(B|AC). Not only is no reason provided why this should hold (for nearly arbitrary A, B, C), it also violates a + b = 1. Thus, the step from Eq. 2b to 4b is not a mathematical allowed transformation, but a modification. Thus we are not talking about a derivation any more.

I agree (now) that the step from 4b to 4c is an allowed transformation, but the step to 5b is another (in my eyes unjustified) modification.

Modifying the laws of conventional probabilities for “recursive” probabilities would be fine, if it could be shown that they are dealing with situations conventional probabilities are not already able to handle properly. This demonstration is lacking and my previous arguments about the emptiness of the set of recursive probabilities hint towards it being impossible.

To conclude, I repeat the final statement from my review of the first version: As I see no way how to fix this severe inconsistency even with a mayor revision **I can not recommend the paper to stay in review**.

## Reviewer #2:

I agree with eq.(2b). I take A, B, C to be binary propositions, i.e. statements that are either true or false (although it might not be obvious which). I agree with (2c), which is the product rule of probability. Therefore I agree with (2d), which puts (2b) and (2c) together. But the reasoning at the bottom of p8 and top of p9 is incoherent and must be wrong, because it leads from (2d) which is correct to (4b) which is wrong, as I shall now show.

Each of the square brackets in (4b) is equal to p(AB|C), so that (4b) can be written

p(A+B|C) = p(A|C) + p(B|C) - p(AB|C) {p(B|AC) + p(A|BC)} (\*)

But (2b) reads

p(A+B|C) = p(A|C) + p(B|C) - p(AB|C)

which is only the same as (\*) if the expression in curly brackets in (\*), p(B|AC)+p(A|BC), is equal to 1. This is not generally the case; indeed if A=B then p(B|AC)+p(A|BC)=1+1=2, and the LHS of (4b) then simplifies (by Boolean algebra) to p(A|C) whereas the RHS simplifies to zero. So equation (4b) is wrong. And so is everything downstream of it.

The authors should pay much closer attention to the arbitrary functions that appear in the analysis of Cox (who derives the product rule before the sum rule) and Knuth (the authors' refs [24 and [29], who derives the sum rule before the product rule). In these works, these functions are absorbed in a redefinition of probability. The hyperbolic functions that appear in the present paper are merely one specific choice of these functions. There is nothing special about them.

I am not impressed that the paper, which is ultimately about probability, throws in relativity (including black holes - general relativity as well as special), quantum mechanics, thermodynamics and even complex time (at the bottom of p11). Probability is about making inference systematic. Did we really need to know about quantum theory and relativity in order to improve probabilistic inference in learning about the past from archaeological traces?

This paper remains unsuitable for publication.

# AUTHOR RESPONSE

Previously we thanked the Editor for selecting excellent, highly specialist and perceptive Reviewers, and we also thanked the Reviewers themselves for considering our work so carefully, commenting that “*it is not often that one gets such specific and helpful reviewing*”. These Reviews of our Revision are equally impressive and helpful.

As a general response to the Reviewers, it has become clear to us that where previously they highlighted major areas of confusion (leading us to completely rewrite the MS), this time there are only issues of misunderstanding due to imperfect exposition which we have endeavoured to make clearer. Of course, these remain our responsibility (and we have again thoroughly revised the work for a new submission to *JPhysA*). We specifically address all the Reviewers’ points below:

## Reviewer #1:

We thank Reviewer#1, again, for their exceptionally detailed and thoughtful review. They claim that a number of fatal flaws remain in our work: we will refute these claims, while of course acknowledging that our text should have been clearer so that such suspicions were not aroused in the first place.

Reviewer#1 says that we do not give a mathematically rigorous definition of ‘recursive probabilities’, however, recursion is clearly present when the probabilities *p*(*A*|*B*) and *p*(*B*|*A*) are both non-zero at the same time. That is to say, for recursion we require *p*(*A*|*B*)>0 and *p*(*B*|*A*)>0, with both probabilities being valid independent of any temporal (or causal or logical) constraints. This is a sufficient mathematical definition for recursive probabilities which we now provide appropriately in the text. Of course, Bayes theorem also in general requires both these two probabilities to be non-zero, but the choice of priority (associated with each side of Bayes’ theorem) assumes a causal (temporal) priority as appropriate to the given context, such that an additional temporal (or causal or logical) constraint therefore also exists.

Reviewer#1 claims that we have only a single “showcase” example (the running of elections in two separate countries) for the HSR; but of course, our Eq.1c also shows that the probability of reflection for an optical etalon obeys the same HSR sum rule equation. And, of course, such a physical device exists, is correctly described by the HSR, and is recursive (multiply reflective!) such that Reviewer#1’s assertion that “the set of recursive probabilities is empty” is clearly wrong! But it is true that we did not sufficiently highlight this.

Reviewer#1 provides an important discussion requiring careful consideration when they state that the “*HSR cannot hold for only potential recursive probabilities, as a potential recursive probability can turn out to be not recursive and then the CSR holds, which is not a limiting case of the HSR*.” When there are quantitative or qualitative unknowns in any formulation of the probability of a phenomenon there exists the implication of ‘potential’ constraints or contributions affecting the calculation of the probability. The *Principle of Indifference* (**PI**) recognises the presence of these unknowns (potential and actual), and generates a probabilistic analysis that accounts for them without introducing bias; so that Jaynes was able to formulate the principle of *Maximum Entropy* (**MaxEnt**) as a quantitative approach that avoids the inclusion (inadvertent or otherwise) of unjustified or biassed “information” into the probabilistic description of any phenomenon. There always exists an infinity of ‘potential’ influences on any system: *Question*, how do we appropriately account for *all* of them? *Answer*, employing a MaxEnt solution alongside the PI (in the appropriate domain of applicability). Thus for Reviewer#1 to suggest that the HSR doesn’t deal correctly with the issue of *potential* recursion is wrong, since the HSR is proven MaxEnt (see Appendix C) within its domain of applicability (which includes recursion); in contrast to the CSR, which is also MaxEnt, but within its narrower domain of applicability (which explicitly excludes recursion).

Reviewer#1 also claims that the CSR is not a limiting case of the HSR. However, Eqs.8 make it plain that the CSR can be thought of as a “truncated case of the HSR”; that is, the CSR is applicable for the situation when there is specific mutual dependence (with a specified or strongly implied logical/temporal constraint) between the event probabilities: that is, if we *possess specific a priori knowledge* about the mutual dependence of these probabilities. This can be seen more clearly in the factorisation (roots) of the self-reciprocal function S=(1-*x*)/(1+*x*)=(1-*x*)(1-*x*+*x*2-*x*3+…) when only the first root (1-x) is taken (as per Cox’s and Jaynes’ analysis for the CSR where *S*=(1-*x*): see Eqs.8 in the text) and all the other roots (corresponding to dependencies) are neglected. Thus the CSR can indeed be thought of as the limiting case for the HSR when all these higher-order dependencies are ignored *a priori* (due to specified information). This was originally discussed in the vicinity of Eq.5b. But our text was far from clear and has been revised, particularly in the discussion after Eq.7.

The final point of Reviewer#1 relating to the scenario of elections in two countries includes the charge that our treatment requires *influences* to be *somehow perfectly balanced.* It appears that Reviewer#1 doesn’t properly understand the power of PI/MaxEnt to take into account all possible extraneous influences – they don’t need to be “*somehow perfectly balanced*”! The PI/MaxEnt principles are properly deployed to ensure that the infinitude of qualitatively and quantitively unknown potential influences on the phenomenon are neutrally taken into account when calculating probabilities, avoiding the inclusion (inadvertent or otherwise) of unjustified information and/or bias into the probabilistic description.

Likewise, the example from Reviewer#1 “*If I do not know the physical masses of electrons and protons, I am still not allowed to assume them to be the same*” is also wrong for precisely the same reason. The PI and MaxEnt indeed require us to admit the possibility of the masses being *the same* in the absence of any other knowledge, in which case the MaxEnt/PI solution is indeed equal masses. The point here of course is that qualitative information (or lack of it) is also information, and the Bayesian treatment of probabilities takes this into account.

With respect to Reviewer#1’s derivation of the HSR, unfortunately, they make an incorrect assumption in their formalism which renders their analysis wrong. That is to say (if we understand it correctly) their quantities *f* and *g* are, respectively, the two forms on the RHS of our Eq.2d; so that they are suggesting (correctly, as we understand it) that *f*=*g*=*p*(*A* OR *B*), since they write: *f = g and thus [f(. . .)]A + [g(. . .)]B := (a + b) f(. . .) = f(. . .) which recovers Eq. 2d…*, since Reviewer#1 defines the quantities: *a = P(“A is prior to B”) and b = P(“B is prior to A”)*, which, being mutually exclusive, clearly means that *a*+*b*=1. But Reviewer#1 then conflates their definition for *a* and *b*, with the probabilities *a=P*(*A*|*BC*) and *b=P*(*B*|*AC*), which are clearly different quantities! This is not something we do: we neither invoke nor use the probabilities: *P(“A is prior to B”)* and *P(“B is prior to A”)*.

That is to say, the quantities *P*(*A*|*BC*) and *P*(*B*|*AC*) are clearly defined for complementary priorities; but they do not represent the probabilities for the priorities. Therefore (for example) *P*(*A*|*BC*) is the probability that *A* occurs given that *B* has occurred, but it does NOT equal the probability that *B* is prior to *A*! By introducing these additional (extraneous) probability terms *P(“A is prior to B”)* and *P(“B is prior to A”)* into their analysis, it is not surprising that the Reviewer#1 achieves a contradictory and erroneous result.

Clearly, Reviewer#1 has given our paper very careful thought, and equally clearly we could have been more helpful in helping readers avoid subtle errors. Although the paper does not have the fatal flaws that Reviewer#1 suspected we are grateful for their careful reading that has made us thoroughly revise our text for clarity.

## Reviewer #2:

Reviewer#2 complains that our Eq.2b is inconsistent with our Eq.4b. This appears to miss the point entirely! The one is the CSR and the other is the HSR (albeit truncated to first-order) – of course they are different! The HSR applies to different cases and is clearly distinguishable from the CSR.

More specifically, the Reviewer#2 is ignoring the consideration of the temporal (or logical) *ordering* of the propositions, which is the point of our paper and is what distinguishes between the CSR and HSR (and which is the burden of writing Eq.4b the way we have). In particular, Reviewer#2 correctly identifies that the quantity “p(B|AC)+p(A|BC)” must equal 1. But then, as a result of ignoring the temporal/logical ordering of {A,B}, Reviewer#2 gets the inevitable contradiction “A=B”! But Eqs.4 require the quantities p(B|AC) and p(A|BC) to be mutually exclusive (they cannot both be true at the same time), and therefore p(B|AC)+p(A|BC)=1; indeed by identifying that the propositions are *ordered* then the possibility A=B is excluded on purely logical grounds. The Reviewer#2 has helpfully identified a clear pedagogical misapprehension, and we have modified the text in our paper to try to overcome the potential for this particular misapprehension.

Reviewer#2 points out that previous workers (Cox & Knuth) specify arbitrary functions of which the “*hyperbolic functions that appear in the present paper*” are but “*one specific choice*”. However Reviewer#2 is incorrect to claim that “*There is nothing special about them*” because we show that they lead to a novel and powerful treatment of otherwise inaccessible cases. Neither Cox nor Knuth noticed these special properties for specific choices of their “arbitrary functions”, although in both cases they were prescient enough to notice the very existence of these functions, very helpfully for us.

Reviewer#2 complains about how we generalise the scope of the work, saying that “*Probability is about making inference systematic. Did we really need to know about quantum theory and relativity in order to improve probabilistic inference in learning about the past from archaeological traces?*” We have made a detailed technical case for saying that probability theory is ***not*** merely about “*making inference systematic*”! Of course, inference being systematic is a *sine qua non*, but we are making a larger point here, that “*probability*” should be regarded as *physical* (in some sense, which we try to make definite within the context of the theatre of spacetime). But this is a *Gestalt* switch, which perhaps means we should not be surprised that some readers will resist.