How Likelihoodists Should (and Should Not) Respond to Borel's Paradox*

Greg Gandenberger

March 14, 2014

Note: This document is a work in progress. Please do not cite it without permission. Comments are welcome at greg@gandenberger.org.

Abstract

Borel's paradox gives rise to counterexamples to the Law of Likelihood. Those counterexamples do not strike at the heart of the likelihoodist position, but they do require refining it. One can address them without doing substantial damage to the likelihoodist position by either restricting the Law of Likelihood so that it does not apply when the likelihood ratio in question depends on an arbitrary choice among sigma fields or maintaining that evidential favoring is relative to a sigma field in such cases. Several seemingly promising alternatives to those two responses do not work, which suggests that a likelihoodist must adopt one of them to avoid refutation.

Contents

*Thanks to Adam Brodie, Kenny Easwaran, Branden Fitelson, Satish Iyengar, Michael Lew, Wayne Myrvold, Alexander Pruss, Bryan Roberts, Teddy Seidenfeld, and Nathan Van-Houdnos for discussion concerning the ideas in this paper.

1 Introduction

The Law of Likelihood says that a datum E evidentially favors a hypothesis H_1 over an incompatible[1] hypothesis H_2 if and only if the likelihood ratio $k = \Pr(E|H_1)/\Pr(E|H_2) > 1$, with k measuring the degree of favoring. This claim is problematic when either H_1 or H_2 has probability zero. When H has probability zero, $\Pr(E|H)$ is defined in Kolmogorov's theory of regular conditional distributions only relative to a sigma field in which H is embedded. One can embed a pair of hypotheses H_1 and H_2 in a sigma field that treats them differently even though the setup of the problem is symmetric with respect to those hypotheses. Doing so gives rise to instances of Borel's paradox,[2] in which probabilities conditional on probability-zero hypotheses that intuitively should be equal turn out to be unequal. In this way, one can produce counterexamples to the Law of Likelihood in which $\Pr(E|H_1)/\Pr(E|H_2) > 1$ even though the setup of the problem is symmetric with respect to H_1 and H_2.

I present a counterexample of that kind in Section 2. In Section 3, I present two possible responses to such counterexamples that I argue are sufficient to block them while doing little if any substantial damage to the likelihoodist position. In Section 4, I argue that six seemingly promising alternatives to those responses are unsatisfactory. The failure of those alternatives suggests at least weakly that accepting one of the responses I defend in Section 3 is not only sufficient but also necessary for a likelihoodist to avoid refutation.

Borel's paradox arises only for probability-zero hypotheses in the absence of a preferred sigma field on the hypothesis space. Thus, counterexamples like the one given here say nothing about applications of the Law of Likelihood to

[1]Standard formulations of the Law of Likelihood do not require that H_1 and H_2 be incompatible. I argue for this requirement in Chapter 2, Section 2.

[2]"Borel's paradox" has also been called "Bertrand's Paradox" because Joseph Bertrand seems to have been the first to discuss it (**?**). Émile Borel responds to Bertrand's discussion in his (**?**, 100–4). It is also sometimes called "the Borel–Kolmogorov paradox" because Kolmogorov gives the canonical response to it in his (**?**, 50–1)

hypotheses with strictly positive probability, nor do they preclude the possibility of a sophisticated version of the Law of Likelihood that yields intuitively correct results when applied to probability-zero hypotheses. For these reasons, they do not strike at the heart of the likelihoodist position. However, they do call for refining that position.

2 A counterexample to the Law of Likelihood based on Borel's paradox

In the following example, standard mathematical techniques yield $\Pr(E|H_1)/\Pr(E|H_2) > 1$, but it is intuitively clear that the datum E evidentially neutral between H_1 and H_2. Thus, it is a counterexample to the Law of Likelihood.

> **Example 1.** Consider a unit sphere equipped with an arbitrary system of latitudes and longitudes. Let E be the datum that a point P randomly selected from a uniform distribution on the surface of the sphere lies within one degree latitude and one degree longitude of the intersection of the equator and the prime meridian. Let H_1 be the hypothesis that P lies on the "prime meridional circle," i.e. union of the prime meridian and the line of 180° longitude, omitting the points at which that circle intersects the equator. Let H_2 be the hypothesis that P lies on the equator, omitting the same two points.

By stipulation, the system of latitudes and longitudes in this example is arbitrary. Moreover, the datum E is symmetric with respect to the equator and the prime meridional circle. Thus, it is intuitively clear that E is evidentially neutral between H_1 and H_2. But standard mathematical techniques yield the likelihood ratio $\Pr(E|H_1)/\Pr(E|H_2) = 1.57$, which according to the Law of Likelihood indicates that E favors H_1 over H_2. Thus, the Law of Likelihood as standardly formulated is false.

I derive the anomalous likelihood ratio in Appendix A. Roughly speaking, it arises from the fact that lines of longitude are farther apart at the equator than they are at the poles, while lines of latitude are equally spaced all the way around the sphere. Thus, the claim that E favors H_1 over H_2 would be defensible if for some $\epsilon > 0$, H_1 said that P lies with ϵ radians of the prime meridional circle and H_2 said that P lies with ϵ radians of the equator. It is not defensible because those hypotheses say that P lies *exactly* on their respective great circles.

Characterizing the great circle to which H_1 refers as a meridional circle and the great circle to which H_2 refers as a zonal circle and using standard mathematical techniques to compute $\Pr(E|H_1)$ and $\Pr(E|H_2)$ is tantamount to treating $\Pr(E|H_1)$ and $\Pr(E|H_2)$ as realized values of regular conditional distributions relative to a sigma field that is not symmetric with respect to H_1 and H_2.

One might think that it would be possible to avoid the counterexample by requiring that $\Pr(E|H_1)$ and $\Pr(E|H_2)$ be computed relative to a sigma field that treats them symmetrically. I discuss this response and other seemingly promising but unsuccessful responses in Section 4. Before that, I discuss in Section 3 two responses to the counterexample that I argue are successful.

3 Two successful responses to the counterexample

Likelihoodists can obviously avoid the counterexample presented here by restricting the Law of Likelihood to hypotheses with strictly positive probability. However, I argue in Subsection 3.1 that this response is stronger than necessary.

A weaker response that suffices is to restrict the Law of Likelihood so that it does not apply when the likelihood ratio in question depends on an arbitrary choice among sigma fields. The primary worry that this response faces is that it might rule out the application of the Law of Likelihood to cases of genuine scientific interest. However, I argue in Subsection 3.1 that it does not in fact do so.

Another possible response to the counterexample is to claim that the seemingly anomalous likelihood ratio it generates is a correct measure of the degree to which E favors H_1 over H_2 *relative to the sigma field with respect to which it is computed*; more generally, evidential favoring is relative to the sigma field chosen when the likelihood ratio in question depends on an arbitrary choice among sigma fields. The primary worry this response faces is that it is not clear what the Law of Likelihood is supposed to do for us if we adopt it in cases in which the likelihood ratio in question depends on an arbitrary choice among sigma fields. We do not seem to have an informal notion of "evidential favoring relative to a sigma field" for it to explicate. In Subsection 3.2, I argue that this worry is not fatal and give some admittedly weak reasons to prefer relativizing the Law of Likelihood to restricting it.

3.1 Response 1: Restrict the Law of Likelihood so that it does not apply when the likelihood ratio in question depends on an arbitrary choice among sigma fields

Restricting the Law of Likelihood so that it does not apply to probability-zero hypotheses is of course sufficient to block all counterexamples arising from Borel's paradox. However, it is stronger than we need because Borel's paradox does not arise for all pairs of probability-zero hypotheses.

In particular, Borel's paradox does not arise for *simple statistical hypotheses*, where I define H as a simple statistical hypothesis with respect to datum E if and only if there is a constant a such that $\Pr(E|H) = a$ in the absence of any contingent background knowledge.[3][4] The following example illustrates this point:

> **Example 2.** Suppose one were to draw a number r at random from a uniform distribution on the unit interval and then flip a coin with the corresponding bias r for heads. Consider the datum E' that the coin lands heads as evidence in relation to the hypotheses $H^* : r = .25$ and $H^\dagger : r = .5$.

H^* and $H^\dagger$ are simple statistical hypotheses with respect to E': each assigns a definite probability to E' "all by itself," i.e. without reference to any contingent background knowledge. Thus, the Law of Likelihood yields the intuitively correct verdict that E' weakly favors $H^\dagger$ over H^* without any problem.

In addition, Borel's paradox does not arise for probability-zero *composite* statistical hypotheses when the likelihood function does not depend on an arbitrary choice of sigma field, where a composite statistical hypothesis with respect to E is a disjunction of simple statistical hypotheses with respect to E. Consider, for instance, the following variant on Example 2:

> **Example 3.** Suppose one were to draw two numbers r and q independently from a uniform distribution on the unit interval, flip a coin with

[3]The proposition that H is a simple statistical hypothesis with respect to E in this sense is not equivalent to the proposition that $\Pr(E|H)$ is a tautological probability in the sense of (**?**, 141). $\Pr(E|H)$ is a tautological probability in this sense if and only if there is a constant a such that $\Pr(E|H, K) = a$ for any K such that $H \cap K \neq \varnothing$. Consider the hypothesis H that some random variable X has a standard normal distribution and the datum E that $X \in (-2, 2)$. H is a simple statistical hypothesis with respect to E in my sense, as it should be ($\Pr(E|H) = .95$ in the absence of any contingent background knowledge). But $\Pr(E|H)$ is not a tautological probability. For instance, $\Pr(E|H, K)$ is .95 if K is a tautology but is .68 if $K \equiv X \in (-1, 1)$.

[4]One might object that H_1 and H_2 in Example 1 count as simple statistical hypotheses with respect to E on my definition when they are conjoined with the information I that P is uniformly distributed over the surface of the sphere. But $\Pr(E|H_1\&I)$ and $Pr(E|H_2\&I)$ have definite values only relative to a sigma field, so I have classified $H_1\&I$ and $H_2\&I$ as simple statistical hypotheses only relative to a sigma field, which is fine.

the corresponding bias r for heads, and report heads with probability q if the coin land heads and report nothing otherwise. Consider the datum E'' that the coin is reported to land heads as evidence in relation to the hypotheses $H^{**} : r = .25$ and $H^{\ddagger} : r = .5$.

H^{**} and $H^{\ddagger}$ are probability-zero composite hypotheses with respect to E'': H^{**} is a disjunction of simple statistical hypotheses each of which ascribes probability $.25q$ to E'' for some unknown $0 \leq q \leq 1$, and $H^{\ddagger}$ is a disjunction of simple statistical hypotheses each of which ascribes probability $.5q$ to E'' for the same unknown q. Yet the Law of Likelihood yields the intuitively correct verdict that E'' weakly favors $H^{\ddagger}$ over H^{**} without any difficulty: $\Pr(E''|H^{\ddagger})/\Pr(E''|H^{**}) = .5q/.25q = 2$.

The Law of Likelihood is generally taken to apply to composite statistical hypotheses if and only if a prior probability distribution over the hypothesis space is available (leaving aside admittedly *ad hoc* generalizations of the Law (**?**, Ch. 7)), but Example 1 indicates that the availability of a prior probability distribution is not enough when one or both of the hypotheses have prior probability zero.[5] Borel's paradox arises in such cases when the relevant likelihood ratio does depend on an arbitrary choice of sigma field, as it does in Example 1.[6]

There is a third class of hypotheses that we might call *substantive* that are not straightforwardly identified with even disjunctions of simple statistical hypotheses.[7] High-level scientific theories such as the theory of intelligent design

[5]When $H_c = \cup_i H_i$, where the H_i are simple statistical hypotheses, $\Pr(E|H_c)$ is given by $\int \Pr(E|H_i)f(H_i|H_c)\,dP = [\Pr(E|H_i)/\Pr(H_c)]\int f(H_i)dP$ when $\Pr(H_i) > 0$. This formula breaks down when $\Pr(H_i) = 0$ because in that case $\int f(H_i)dP = 0$ as well.

[6]H_1 and H_2 are composite statistical hypotheses because each of them is an uncountable disjunction of hypotheses according to which the randomly selected point P is a particular point on the associated great circle. Each of those disjuncts is a simple statistical hypothesis with respect to E because it assigns probability 1 to E if the associated point lies within one degree of the intersection of the equator and the prime meridian and assigns probability 0 to E otherwise.

[7]One might object that I have classified what are intuitively conjunctions of statistical and

and the theory of evolution typically fall into this category. I would like to allow the application of the Law of Likelihood to such hypotheses insofar as it is possible to pin down the relevant likelihood functions. For instance, it seems reasonable to say that the fossil record favors the theory of evolution over young-earth creationism on the grounds that the features of the fossil record that we observe are more or less what we would expect if the theory of evolution is true but are quite different from what we would expect if young-earth creationism were true, unless we are willing to give substantial credence to the idea that God, if he exists, would give us seemingly misleading evidence in order to test our faith. Pinning down likelihood ratios for substantive hypotheses may require supplying background knowledge, auxiliary assumptions, and/or subjective judgments, which limits their usefulness but need not make them completely useless in all cases.

The process of using the Law of Likelihood in a loose way to make qualitative judgments about substantive hypotheses is certainly not affected by a choice of sigma fields even if the substantive hypothesis in question is plausibly regarded as having prior probability zero. Thus, Borel's paradox is not a difficulty for substantive hypotheses.

I see no need to restrict the Law of Likelihood to hypotheses of positive probability when the relevant likelihood ratio has a well-defined value that does not depend on an arbitrary choice of sigma field. Counterexamples based on Borel's paradox certainly do not indicate such a need.

The worry for any proposal to restrict the scope of the Law of Likelihood is

a substantive hypotheses as statistical. For instance, I have classified $H : X \sim \mathcal{N}(0, 1)$ and the moon is made of green cheese as statistical with respect to $E : X \in (1, 1)$. But it does not seem highly counterintuitive to me to say that H is a statistical hypothesis with respect to E when H ascribes a definite probability to E and also provides additional "substantive" information. More importantly, classifying H as statistical with respect to E in such cases is appropriate for the purposes for which I want to distinguish among different kinds of hypotheses in the first place, namely to delineate the class of cases that create problems for the Law of Likelihood as standardly formulated.

that it might exclude cases of genuine scientific interest, thereby limiting any potential usefulness of the likelihoodist approach. I contend that restricting the scope of the Law of Likelihood to cases in which the relevant likelihood ratio does not depend on an arbitrary choice of sigma field would exclude few if any cases of genuine scientific interest.

One reason not to worry about evidential favoring for probability-zero hypotheses in general is that Bayesian updating on positive-probability data will never turn a probability-zero hypothesis into a positive-probability hypothesis. $\Pr(H|E) = \Pr(E|H)\Pr(H)/\Pr(E)$, so if $\Pr(H) = 0$ and $\Pr(E) > 0$, then $\Pr(H|E)$. Zero-probability data are generally not a concern. The would be a concern if truly continuous sample spaces were possible because every outcome in a continuous sample space has probability zero. But because real measurement techniques have finite precision and bounded range, the true sample space for an experiment is always discrete. And if an outcome occurs that is outside of the sample space for a hypothesis, then that outcome indicates either that the hypothesis is simply false or that some auxiliary hypothesis has failed.

But what about point null hypotheses? Scientists often test the hypothesis that some parameter is exactly zero, for instance. If the parameter space is continuous, such a point null hypothesis might plausibly be regarded as having prior probability zero. Thus, it seems that probability-zero hypotheses are often of genuine scientific interest.

Point null hypotheses are often simple statistical hypotheses. That is, they imply a particular probability distribution over the sample space. In such cases, the Law of Likelihood applies without any difficulties, and the proposal under discussion would not stand in the way.

For instance, suppose that a random variable X is assumed to come from a normal distribution with known variance 1 and unknown mean μ. (Of course, I am using a continuous probability distribution as a convenient idealization only.) A hypothesis of the form $\mu = \mu_0$ implies a particular probability distribution for X, so the Law of Likelihood applies to it without any difficulty.

On the other hand, suppose X were assumed to come from a normal distribution with unknown variance σ^2 and unknown mean μ. A hypothesis of the form $\mu = \mu_0$ implies only a class of probability distributions for X. There are many methods for eliminating the "nuisance parameter" σ^2. When a prior probability is available, one can eliminate σ in a principled way by integrating the probability density function with respect to that distribution. But integrating out one parameter in a two-dimensional parameter space is exactly the procedure that generates the anomalous likelihood function in Example 1.

Likelihoodists can respond to the claim that a given point null hypothesis is of genuine scientific interest in at least three ways:

1. The point null hypothesis has positive probability.

2. The point null hypothesis is an oversimplified representation of what is implicitly a range null hypothesis (i.e., that the true parameter value lies within ϵ of the value given by the point null hypothesis for some $\epsilon > 0$).

3. The scientists are doing something that is not of genuine scientific interest.

Of course, a fourth option is to concede that there are hypotheses of genuine scientific interest to which the Law of Likelihood does not apply.

Options 1 and 2 cover many of the cases in which scientists consider point null hypotheses. In cases from the history of science in which a true point null hypothesis is clearly of genuine scientific interest, we would want (at least in retrospect) to assign positive probability to that null hypothesis. Take, for instance, the Michelson-Morley experiment. That experiment was originally intended to measure the direction and velocity of the aether wind. It is commonly regarded today as having provided strong evidence that there is no aether wind. This case is somewhat complicated because many scientists might have assigned probability zero to the hypothesis that there was no aether wind prior to the Michelson-Morley experiment. However, the introduction of aether drag theories and special relativity enlarged the hypothesis space to include plausible theoretical possibilities that implied the absence of any aether drag. Thus, in retrospect we would want to say that the point null hypothesis of no aether wind did have positive probability.

To take one more example, consider the fact that no stellar parallax was detected until the nineteenth century. The hypothesis that there is no stellar parallax at all was of genuine scientific interest because it followed from standard geocentric astronomical theories but not from Copernicus's heliocentric theory. Of course, that hypothesis would not have been regarded as having probability zero at the time of Copernicus.

Option 1 thus covers at least many of the cases in which a true point null hypothesis is of genuine scientific interest. Option 2 in turn covers at least many of the cases in which scientists perform a formal test of a point null hypothesis that we know to be literally false. This practice is somewhat common in the social sciences. For instance, a recent paper in *Psychological Science* reports results from tests of various versions of the point null hypothesis that children's tendency toward altruistic behavior is the same one month before and one month after experiencing a major earthquake (**?**). This point null hypothesis is arguably highly implausible—perhaps even a probability-zero claim. It would be incredible if the traumatic experience of living through an earthquake had *precisely* zero effect on altruistic behavior. But it would not be incredible if the effect were small enough to be essentially negligible. In cases like this one involving implausible point null hypotheses, it is reasonable to regard researchers as testing the vague range null hypothesis that the departure from the point null hypothesis is negligible rather than testing the point null hypothesis itself.

Option 2 seems to be problematic from a likelihoodist perspective because likelihood ratios are not well-defined for composite hypotheses such as range null hypotheses when a prior probability distribution is not available, which is the usual case in science according to most likelihoodists. However, a likelihoodist can put bounds on likelihood ratios involving range hypothesis. For instance, he or she might be able to say that the likelihood ratio for the maximum likelihood estimate of the mean of some distribution against the hypothesis that the mean is in some interval around zero is at least k for some k.

The burden is on opponents of likelihoodism to argue that there are point null hypotheses of genuine scientific interest that are not covered by either Option 1 or Option 2. More generally, the burden is on them to argue that restricting the Law of Likelihood to cases in which the likelihood ratio does not depend on an arbitrary choice of sigma field excludes cases of genuine scientific interest. At worst, the Law of Likelihood does not apply to some (presumably unusual) kinds of cases to which we would like it to apply. This conclusion would hardly be devastating to the likelihoodist position.

3.2 Response 2: Relativize evidential favoring to the sigma field chosen when the likelihood ratio in question depends on an arbitrary choice among sigma fields

An alternative to restricting the Law of Likelihood to cases in which the likelihood ratio does not depend on an arbitrary choice of sigma field is to maintain that evidential favoring is relative to the sigma field chosen in such cases. I argue in this section that there is nothing but verbal preference to decide between these two responses.

Prima facie, relativizing the notion of evidential favoring to a sigma field seems preferable to restricting the Law of Likelihood simply because it allows the Law of Likelihood to apply to a broader range of cases. However, this apparent advantage is empty because the notion of evidential favoring relative to a sigma field lacks both practical utility and intuitive significance. What does it mean to say regarding Example 1 that E favors H_1 over H_2 to the degree 1.57 relative to the sigma field implicit in the calculation performed? Unlike in cases involving hypotheses with positive probabilities, that statement cannot be given a Bayesian interpretation in terms of the relationship between the posterior and prior odds (which are undefined).

A possible response to this objection is that while a measure of the degree to which E favors H_1 over H_2 evidentially relative to a particular sigma field has little or no intuitive significance, it takes on intuitive significance as the degree to which E favors H_1 over H_2 evidentially *simpliciter* in a context in which that sigma field is the "relevant" or "preferred" one. Thus, while Proposal 2 makes the Law of Likelihood more or less idle for probability-zero hypotheses in the absence of a preferred sigma field, it allows for a sense in which the principle nevertheless applies in such cases and is "ready and waiting," so to speak, for a preferred sigma field to be specified.

This response is not persuasive as an argument for relativizing the Law of Likelihood rather than restricting it because it concedes that the Law is idle in the cases to which it applies when relativized but not when restricted. For this reason, it seems that the choice between relativizing and restricting is inconsequential and can be decided on the basis of verbal preference alone.

For what it is worth, I can see two weak arguments for preferring relativizing the Law of Likelihood to restricting it. First, the relativization is in some sense already built into the Law of Likelihood. It is reasonable to assume that the likelihood functions to which the Law of Likelihood refers are to be understood according to Kolmogorov's theory of regular conditional distributions because that theory is the standard one taught to graduate students and presupposed in most theoretical and applied work. On that theory, probabilities conditional on probability-zero hypotheses are defined only relative to a sigma field. Thus, it is reasonable to assume that the Law of Likelihood is to be understood in such a way that it has a definite application to probability-zero hypotheses only relative to a sigma field.

This argument has a major weakness: the major expositors of the Law of Likelihood do not seem to have considered the difficulties that Borel's paradox presents. Thus, the relativization in question was built into the Law of Likelihood *accidentally*. The claim that this fact should motivate us to embrace that relativization is far from compelling.

A second argument for relativizing the notion of evidential favoring to a sigma field alleges that in the absence of a sigma field, probability-zero hypotheses are not even toy scientific theories in a proper sense.[8] Many philosophers of science have argued that scientific theories are better thought of as models than as mere sets of sentences. Models include methods for generating predictions from data. A probability-zero hypothesis does not imply definite probabilities for observations until it is embedded in a sigma field. Thus, according to our best accounts of what it is to be a scientific theory, probability-zero hypotheses in the absence of a sigma field do not qualify.[9]

This argument is weak because it assumes that the Law of Likelihood should apply only to "proper" scientific theories according to a contentious account of

[8] Thanks to Alexander Pruss for suggesting this argument in personal correspondence.

[9] One might prefer to say not that evidential favoring is relative to a sigma field, but rather that the Law of Likelihood only applies to probability-zero hypotheses *qua* elements of sigma fields. I regard these views as mere notational variants of one another. I am not thereby committed to regarding Responses 1 and 2 as mere notational variants of one another because I intend Response 1 to imply that the Law of Likelihood does not apply to H_1 and H_2 even *qua* embedded in some kind of structure on the hypothesis space.

what it requires to be such a theory.

Relativizing the notion of evidential favoring to a sigma field where the choice of sigma field makes a difference to the likelihood ratio seems to be a tenable alternative to simply denying that the Law of Likelihood applies to such cases. I prefer the relativizing approach for basically aesthetic reasons, but I do not see that anything important hangs on the choice between them.

4 Six unsuccessful alternative responses to the counterexample

I have argued that restricting the Law of Likelihood so that it does not apply to cases in which the likelihood ratio depends on an arbitrary choice of sigma field and regarding evidential favoring as relative to a sigma field are viable responses to counterexamples based on Borel's paradox. Even someone who accepts that claim might think that those responses are unnecessary or that some other response is preferable. In this section I argue that six seemingly promising alternative responses do not work, which suggests that adopting one of the two responses I endorse is necessary for saving the Law of Likelihood from refutation.

4.1 Alternative 1: The fact that each hypothesis omits two points on the relevant great circle defuses the counterexample

It might seem suspicious that H_1 and H_2 in Example 1 each omit two points from the relevant great circle. However, *the omission of those points plays no role in generating the anomalous likelihood ratio* (See Appendix A). Because those points have no length, the same likelihood ratio arises regardless of how they are handled.

I exclude the points at which the relevant great circles intersect only so that one cannot avoid the counterexample by restricting the Law of Likelihood to mutually exclusive hypotheses, as I advocate in Chapter 2.

4.2 Alternative 2: The fact that the likelihood ratio is small excuses the Law of Likelihood

The value 1.57 of the anomalous likelihood ratio is much smaller than the value of 8 that likelihoodists conventionally require in order to declare a result "fairly strong" evidence (**?**, 761). One might think that this fact somehow excuses the Law of Likelihood. But it does not in fact do so, for two reasons:

1. Regardless of what the Law of Likelihood says about the degree of evidential favoring in this case, it still implies the incorrect qualitative claim that the result favors the prime meridional circle hypothesis over the equator hypothesis.

2. One could produce an analogous but more dramatic result by using a strange coordinate system. As I explain above, the fact that likelihood ratio in this case is greater than one comes from the fact that lines of longitude are farther apart at the equator than near the poles, while lines of latitude are spaced equally all the way around the sphere. One could get a larger likelihood ratio by using a system of "pseudo-longitudes" that exaggerates this effect around the prime meridian and/or a system of "pseudo-latitudes" that are closer together near the prime meridianional circle than elsewhere around the equator.

4.3 Alternative 3: The fact that measurement techniques have finite precision defuses the counterexample

Borel himself points out that actual methods of observation do not allow you to learn that a particular point on a sphere lies on a particular great circle (**?**, 102–3). From a position on the prime meridian, for instance, you might be able to use astronomical observations and a chronometer to determine that your longitude is between $0.1''$ East and $0.1''$ West, but you would never be able to determine that your longitude is exactly 0.

This point would be relevant if $\Pr(E|H_1)$ were supposed to reflect uncertainty about the the longitude of P given a *measurement* with a high but finite degree of precision indicating that it lay on the equator, and $\Pr(E|H_2)$ were supposed to reflect uncertainty about the the latitude of P given a measurement with a high but finite degree of precision indicating that it lay on the prime meridional circle. But those probabilities are supposed to reflect uncertainty given *hypotheses*, not measurements. The fact that measurement techniques have finite precision does not preclude considering "sharp" hypotheses.

The fact that measurement techniques have finite precision may allow Bayesians to escape difficulties arising from Borel's paradox because Bayesians condition on observed evidence. But it does not help likelihoodists escape such difficulties because likelihoodists condition on hypotheses.[10]

4.4 Alternative 4: Restrict the Law of Likelihood to hypotheses that belong to a common model

One might think that the Law of Likelihood should be restricted to hypotheses that belong to a "common model" and that H_1 and H_2 in Example 1 do not

[10] Thanks to Alexander Pruss for providing this compact formulation of my response to the claim that Borel's paradox is just as much a problem for Bayesians as it is for likelihoodists.

belong to a common model.

This response to the counterexample faces many difficulties. First it's not clear what it means for two hypotheses to belong to a common model. In statistics jargon, "model" typically refers to a triple of the form $\{\mathcal{X}, \Theta, \mathbf{P}\}$, where $\mathcal{X}$ is a sample space, Θ is an index set, and $\mathbf{P}$ is a set of probability distributions indexed by the elements of Θ. H_1 and H_2 do not belong to such models at all because they are not probability distributions for P. Instead, they are claims about P that give rise to probability distributions when conditioned upon, but only relative to a sigma field.

One could restrict the Law of Likelihood to probability distributions themselves, as opposed to claims that merely give rise to probability distributions when conditioned upon. However, that restriction would be a costly overreaction. It would be costly because it would prevent one from applying the Law of Likelihood to substantive hypotheses, including high-level scientific theories. It would clearly be an overreaction because it would prevent one from applying the Law of Likelihood to examples such as the following in which its use is entirely unproblematic.

Example 4. Consider a set of playing cards that consists of the four aces, the kings of spades, and the king of clubs. Those cards are shuffled, one is drawn at random, and its suit is reported.

To what degree does the information that the selected card is a spade favor the hypothesis that it is a king over the hypothesis that it is an ace? The Law of Likelihood answers this question easily: $\Pr(\text{spade}|\text{king})/\Pr(\text{spade}|\text{ace}) = (1/2)/(1/4) = 2$. This answer is intuitively reasonable and has a nice interpretation as the ratio of the posterior odds to the prior odds under Bayesian updating. It is clearly unnecessary to prevent one from applying the Law of Likelihood to this example by restricting it to probability distributions themselves, as opposed to claims that merely give rise to probability distributions when conditioned upon.

It is not clear what it could mean to say that two hypotheses that merely give rise to probability distributions when conditioned upon belong to different models except that the probability distributions to which they give rise when conditioned upon belong to different models, which is not the case for H_1 and H_2. Thus, it is not clear how the proposal to restrict the Law of Likelihood to hypotheses that belong to a common model is supposed to exclude H_1 and H_2. Perhaps there is a different notion of "model" than the standard statistical notion that would make this proposal well-motivated and allow it to exclude H_1 and H_2. Advocates of the proposal owe us an account of that notion.

There is no general difficulty in applying the Law of Likelihood to pairs of hypotheses that would generally be identified as belonging to different parametric models. The hypothesis that some random variable has the standard normal distribution can be tested against the hypothesis that it has the standard Cauchy distribution, for instance. One can if one likes write down a probability density function for a model that includes both of those hypotheses. In general, if hypothesis H_1 says that some random vector X has probability density function $f_1(X)$ and another hypothesis H_2 says that X has probability density function $f_2(X)$, then those hypotheses belong to a common model according to which $f(X) = I(b = 1)f_1(X) + I(b = 2)f_2(X)$, where I is an indicator function and $b = i$ if and only if hypothesis H_i is true.

If we had a clear account of what it means to belong to a common model according to which H_1 and H_2 do not qualify, we would still need to consider why we should believe that the Law of Likelihood only applies to hypotheses from a common model. One motivation for holding this view is that R.A. **?**, 310 and A.W.F. **?**, 9, among others, characterize likelihood functions as defined only up to constants of proportionality. If those constants are necessary and each one is specific to the model in which the relevant hypothesis is embedded, then the claim that the Law of Likelihood applies only to hypotheses that belong to a common model is well-motivated. For if a pair of hypotheses H and H' belong to different models, then their likelihood ratio on some datum A is not given by $\Pr(A|H)/\Pr(A|H^*)$, but rather by $c \cdot \Pr(A|H)/c^* \cdot \Pr(A|H^*)$ for some pair of constants c and c^* that do not stand in any definite relation to one another.

However, I can see no argument beyond an appeal to authority for the claim that likelihood functions must be regarded as being defined only up to a model-specific constant of proportionality. It might be convenient to multiply a likelihood function by a constant so that, for instance, the maximum of the function is always 1. But the claim that normalizing the likelihood function in this way is useful does not require defining "likelihood function" in such a way that a normalized likelihood function is itself a likelihood function. Similarly, the fact that the Likelihood Principle says that two experimental outcomes are evidentially equivalent with respect to a set of hypotheses if their likelihood functions over those hypotheses are proportional does not require calling proportional likelihood functions the same likelihood function. It only requires giving those likelihood functions the same evidential interpretation. And the Law of Likelihood does so without any restrictions: if $\Pr(E_1|H) = c \cdot \Pr(E_2|H)$ for some constant c and all H in some partition, then the likelihood ratios $\Pr(E_1|H_1)/\Pr(E_1|H_2)$ and $\Pr(E_2|H_1)/\Pr(E_1|H_2)$ are equal for all H_1 and H_2 in that partition.

If restricting the Law of Likelihood to hypotheses from a common model is not required by the nature of likelihood functions, then perhaps it is required by examples such as the following.[11]

Example 5. Let E be a report that a certain coin with unknown bias

[11] This example is due to Michael Lew (personal correspondence).

p for heads landed heads h times in n tosses ($n \geq h$). Let H_1 be the hypothesis that $p = p_0$ for some $0 < p_0 < 1$ and that the number of tosses was fixed in advance. Let H_2 be the hypothesis that $p = p_0$ and that the number of heads was fixed in advance.

It might seem that E contains no information about whether the number of heads or the number of tosses was fixed in advance. But applying the Law of Likelihood to to this example yields the conclusion that E favors H_1 over H_2 to the degree n/h for any p_0. For

$$\frac{\Pr(E|H_1)}{\Pr(E|H_2)} = \frac{\binom{n}{h}p^h(1-p)^{n-h}}{\binom{n-1}{h-1}p^h(1-p)^{n-h}} = \frac{n}{h}.$$

It also yields the conclusion that E favors the claim that the number of tosses was fixed at n in advance over the claim that the number of heads was fixed at h in advance to the degree n/h if we assume that evidential favoring obeys the following attractive constraint (an analogue of conglomerability): if E favors $A\&C$ over $B\&C$ for all C in some partition, then it also favors A over B.[12] (Let A be the hypothesis that n is fixed, B the hypothesis that h is fixed, and C the set of hypotheses of the form $p = p_0$ for $0 < p_0 < 1$.) Thus, this example seems to speak against the claim that the Law of Likelihood applies to hypotheses from different models.

However, Example 5 is not as compelling a counterexample to the Law of Likelihood across models as it might seem. The Law of Likelihood does not yield the conclusion that E favors the hypothesis H_1^* that the number of heads was fixed in advance over the hypothesis H_2^* that the number of tosses was fixed in advance. Instead, it says that E favors the hypothesis H_1 that the number

[12]This claim can be proven if there is a probability distribution over the relevant partition—regardless of what that distributions is—but it seems eminently plausible even when that partition is a hypothesis space over which many likelihoodists would deny that there can be a meaningful probability distribution.

of heads was fixed at h over the hypothesis H_2 that the number of tosses was fixed at n, where n and h are the number of tosses and the number of heads reported, respectively.

This difference is important, because it would seem to be problematic if the Law of Likelihood said that the data from an experiment would favor some non-random hypothesis over another non-random hypothesis regardless of the outcome of that experiment. But H_1 is a random hypothesis if H_2 is true, and H_2 is a random hypothesis if H_1 is true. The fact that you can after the fact formulate some hypothesis according to which the number of heads was fixed and some hypothesis according to which the number of tosses was fixed such that the data reported favor the former over the latter is not obviously troubling.

Even assuming the analogue of conglomerability stated above, the Law of Likelihood does not say whether E favors the H_1^* over H_2^* or vice versa without a prior probability distributions over h on the assumption that n is fixed and a prior probability distributions over n on the assumption that h is fixed. Given such distributions, what the Law of Likelihood says about whether E favors the H_1^* over H_2^* or vice versa depends on h and n, as it should.

It might seem wrong to say that any (n, h) pair favors the corresponding H_1 over the corresponding H_2, but I find this claim acceptable upon reflection. To evoke clearer intuitions, consider an extreme cases such as $n = 1000$ and $h = 1$. There are 1000 ways this result could have arisen if $n = 1000$ was fixed in advance: the head could have occurred in the first toss, the second toss, ..., or the thousandth toss. But there is only one way it could have arisen if $h = 1$ was fixed in advance: the head must have occurred on the thousandth toss. Considering this fact makes it seem rather plausible to me that the datum $n = 1000, h = 1$ does in fact favor the hypothesis that $n = 1000$ was fixed over the hypothesis that $h = 1$ was fixed to the degree $n/h = 1000$.

4.5 Alternative 5: Appeal to Symmetry

Perhaps applying the Law of Likelihood to H_1 and H_2 is inappropriate not because they belong to different models, but because the sigma field in which they are implicitly embedded treats them differently. It introduces an asymmetry that is not present in the setup of the problem. Unfortunately, the plausible suggestion that we avoid counterexamples such as Example 1 by requiring the the sigma fields relative to which we condition to get the likelihood ratios to which the Law of Likelihood refers respect the symmetries that are present in the setup of the problem leads to disaster, as I show next.

An obvious place to look for such a restriction is to symmetry considerations. A coordinate system that has the great circle in H_1 as an equator and the great circle in H_2 as a meridional circle, for instance, introduces an asymmetry that is not present in the setup of the problem. It seems right to say that conditional probabilities that reflect that asymmetry are inappropriate and should not be used.

Unfortunately, this response will not work. Avoiding easy variants on the counterexample requires adopting the following restriction on the conditional probabilities that appear in the Law of Likelihood when it is applied to that example:[13]

> **Rotation Preservation.** For any hypothesis H according to which P lies in a region that is invariant under rotations of the sphere about some axis, it is required that $\Pr(H|G)$ have a common value a for any hypothesis G according to which P lies on a great circle that passes through the poles of that axis (possibly excluding the endpoints of the axis).

The problem with adopting Rotation Preservation is that it leads to a vio-

[13] Rotation Preservation is a special case of Easwaran's symmetry principle (**?**, 81).

lation of the essentially universally accepted[14] principle of Finite Additivity in the presence of the following additional assumptions (**?**, 83–4):

> **Weak Conglomerability.** If there is a constant a such that $\Pr(A|E) = a$ for all $E \in \mathcal{E}$, where $\mathcal{E}$ is a partition, then it is required that $\Pr(A) = a$.
>
> **Conjunction Identity (my term):** If $A\&E = B\&E$, then it is required that $\Pr(A|E) = \Pr(B|E)$.

See Appendix B for a proof of this claim.

Denying Finite Additivity is not a viable option. Denying Weak Conglomerability means giving up the theory of regular conditional distributions in favor of some other approach such as the theory of coherent conditional distributions, which would cause the appeal to symmetry to collapse into the approach discussed in Subsection 4.6 below. The only remotely plausible way to deny Conjunction Identity is to maintain that it holds only relative to some structure on the hypothesis space such as a sigma field, which would cause the appeal to symmetry to collapse into the approach discussed in Subsection 3.2 below. Thus, the appeal to symmetry is not an adequate alternative to other remedies I discuss elsewhere in this paper.

4.6 Alternative 6: Appeal to an alternative theory of conditional probability

The standard theory of conditional probability for probability-zero conditioning propositions is Kolmogorov's theory of regular conditional distributions (**?**).

[14] I write "essentially universally accepted" because some formal theories that resemble probability theory violate additivity, such as Dempster-Shafer theory and John Norton's approach to representing ignorance (**?**). Additivity is appropriate in probability theory, in which degrees of belief are assumed to span belief and disbelief, as Norton puts it (**?**, 406). It is not appropriate in alternative theories in which degrees of belief are assumed to span something like knowledge and ignorance.

That theory gives rise to Borel's paradox because it makes probability conditional on a probability-zero conditioning hypothesis relative to the sigma field in which that hypothesis is embedded. Treating the great circle in H_1 as an equator and the great circle in H_2 as a meridional circle in the manner of Appendix A is in the Kolmogorov theory tantamount to embedding them in a particular sigma field that treats them differently.

There is an alternative theory of conditional probability for probability-zero conditioning propositions called the theory of coherent conditional distributions (**??**). That theory allows one to condition on events themselves without referring to an embedding sigma field. It requires only that conditional probabilities conform to the following requirements. Requirements 1-3 are conditional counterparts to Kolmogorov's axioms of unconditional probability, and Requirement 4 is a doubly conditional counterpart to the formula $\Pr(A\&B) = \Pr(A|B)\Pr(B)$:

1. Propriety: $\Pr(C|C) = 1$
2. Finite Additivity: $\Pr(A\&B|C) = \Pr(A|C) + \Pr(B|C)$ if $A\&B = \varnothing$
3. Non-Negativity: $\Pr(A|C) \geq 0$
4. Nesting (my term): $Pr(A\&B|C) = P(A|B\&C)P(B|C)$ if $B\&C =\neq \varnothing$

The theory of coherent conditional probabilities faces some serious objections, including the fact that it violates *conglomerability.* Conglomerability is the requirement that unconditional probabilities lie within the range of corresponding conditional probabilities. The claim that probability distributions should be conglomerable is highly intuitive, at least in simple finite examples.

Whether failures of conglomerability and other apparent anomalies in the theory of coherent conditional distributions are devastating to the theory or not is a topic of ongoing debate. We need not concern ourselves with that debate because regardless of how it turns out, appealing to the theory of coherent conditional distributions does not provide a response to the counterexample that allows one to apply the Law of Likelihood to probability-zero hypotheses in a way that a likelihoodist would find satisfactory.

Likelihoodists intend for the Law of Likelihood to provide an objective measure of the degree to which a given datum E favors one hypothesis over another. When H_1 and H_2 are probability-zero hypotheses, the theory of coherent conditional probabilities places no constraints at all on the likelihood ratio $\Pr(E|H_1)/\Pr(E|H_2)$ that the Law of Likelihood proposes to use for this purpose. Thus, the Law of Likelihood cannot serve its intended purpose for probability-zero hypotheses if the likelihoods it contains are taken to be coherent conditional probabilities.

5 Conclusion

Borel's paradox gives rise to counterexamples to naïve versions of the Law of Likelihood. Those counterexamples seem to require either restricting the Law of Likelihood so that it does not apply in the kinds of cases in which Borel's paradox arises or relativizing the notion of evidential favoring to a sigma field in such cases. Neither of those responses are highly damaging to the basic likelihoodist position, but they do require slightly more nuanced formulations of that position than have previously been provided.

A Deriving the anomalous likelihood ratio

Let $\Theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ be a random variable that gives the latitude of P, where $\Theta = 0$ at the equator and $\Theta = \frac{\pi}{2}$ at the north pole. Let $\Phi \in [0, 2\pi)$ be a random variable that gives the longitude of P, where $\Phi = 0$ at the prime meridian. A uniform probabiity distribution over the surface has in this coordinate system the probability density function $\cos\theta/4\pi$.

We can begin to calculate $\Pr(E|H_1)$ as follows:

$$\begin{aligned}\Pr(E|H_1) &= \Pr\left(-\frac{\pi}{180} \leq \Theta, \Phi \leq \frac{\pi}{180}\middle|\Theta = 0, \Phi \neq 0, \pi\right)\\ &= \Pr\left(-\frac{\pi}{180} \leq \Phi \leq \frac{\pi}{180}\middle|\Theta = 0, \Phi \neq 0, \pi\right)\end{aligned}$$

Now, the value of this second expression is not given by the simple ratio formula $\Pr(A|B) = \Pr(A \cap B)/\Pr(B)$: the right-hand side of that formula is undefined in this case because plugging in the relevant values yields $0/0$.

A standard way around this problem is to use an analogous ratio of probability *density* functions formula to produce a conditional probability density function and then to integrate that function over the relevant domain. In this case, the relevant conditional probability density function would be

$$\begin{aligned}f_1(\Phi|\Theta = 0, \Phi \neq 0, \pi) &= \frac{f_2(\Phi, \Theta = 0|\Phi \neq 0, \pi)}{f_3(\Theta = 0|\Phi \neq 0, \pi)}\\ &= \frac{f_2(\Phi, \Theta = 0|\Phi \neq 0, \pi)}{\int_0^{\pi} f_2(\phi, \Theta = 0|\phi \neq 0, \pi)\,d\phi + \int_{\pi}^{2\pi} f_2(\phi, \Theta = 0|\phi \neq 0, \pi)\,d\phi}\\ &= \frac{\cos(0)/4\pi}{\int_0^{\pi} \cos(0)/4\pi\,d\phi + \int_{\pi}^{2\pi} \cos(0)/4\pi\,d\phi}\\ &= \frac{1}{\int_0^{\pi}\,d\phi + \int_{\pi}^{2\pi}\,d\phi}\\ &= \frac{1}{2\pi}\end{aligned}$$

Thus,

$$\begin{aligned}
\Pr(E|H_1) &= \int_{-\pi/180}^{\pi/180} \frac{1}{2\pi}\, d\phi \\
&= \frac{1}{180}
\end{aligned}$$

Notice that we would have gotten the same result if we had not conditioned on $\Phi \neq 0, \pi$. The only difference this conditioning made in our calculations was that it led us to write $f_3(\Theta = 0|\Phi \neq 0, \pi)$ as a sum of two integrals from 0 to π and from π to 2π rather than as a single integral from 0 to 2π.

We can use the same technique to calculate $\Pr(E|H_2)$:

$$\begin{aligned}
\Pr(E|H_2) &= \Pr\left(-\frac{\pi}{180} \leq \Theta, \Phi \leq \frac{\pi}{180} \middle| \Phi \in \{0 \cup \pi\}, \Theta \neq 0\right) \\
&= \frac{1}{2} \Pr\left(-\frac{\pi}{180} \leq \Theta \leq \frac{\pi}{180} \middle| \Phi \in \{0 \cup \pi\}, \Theta \neq 0\right) \qquad \text{(Symmetry)}
\end{aligned}$$

$$\begin{aligned}
f_4(\Theta|\Phi \in \{0 \cup \pi\}, \Theta \neq 0) &= \frac{f_5(\Theta, \Phi \in \{0 \cup \pi\}|\Theta \neq 0)}{f_6(\Phi \in \{0 \cup \pi\}|\Theta \neq 0)} \\
&= \frac{f_5(\Theta, \Phi \in \{0 \cup \pi\}|\Theta \neq 0)}{\int_{-\pi/2}^{0} f_5(\theta, \Phi \in \{0 \cup \pi\}|\theta \neq 0)\, d\theta + f_5(\theta, \Phi \in \{0 \cup \pi\}|\theta \neq 0)\, d\theta} \\
&= \frac{\cos\Theta/2\pi}{\int_{-\pi/2}^{0} \cos\theta/2\pi\, d\theta + \int_{0}^{\pi/2} \cos\theta/2\pi\, d\theta} \\
&= \frac{\cos\Theta}{\int_{-\pi/2}^{0} \cos\theta\, d\theta + \int_{0}^{\pi/2} \cos\theta\, d\theta} \\
&= \frac{\cos\Theta}{[\sin(0) - \sin(-\pi/2)] + [\sin(\pi/2) - \sin(0)]} \\
&= \frac{\cos\Theta}{2}
\end{aligned}$$

$$\begin{aligned}\Pr(E|H_2) &= \frac{1}{2}\int_{-\pi/180}^{\pi/180} \frac{\cos\theta}{2}\, d\theta \\ &= \frac{1}{4}[2\sin(\pi/180)] \\ &= \frac{\sin(\pi/180)}{2}\end{aligned}$$

Again, notice that conditioning on $\Theta \neq 0$ made no difference to the ultimate outcome of this calculation.

The resulting likelihood ratio is

$$\begin{aligned}\frac{\Pr(E|H_2)}{\Pr(E|H_1)} &= \frac{\sin(\pi/180)/2}{1/180} \\ &= 90\sin(\pi/180) \\ &= 1.57\end{aligned}$$

B How Rotation Preservation, Weak Conglomerability, and Conjunction Identity lead to a violation of Finite Additivity

Let A be the hypothesis that P is in the union of the circular regions of points that are within some number of degrees $0 < d < 90$ of the poles in some system of latitudes and longitudes. Let $\mathcal{G}_1$ be a "partition" of the sphere into great circles that pass through those poles. ($\mathcal{G}_1$ is not technically a partition because each of its elements contain that two poles. I deal with this issue below.) Rotation Preservation requires that $\Pr(A|G)$ be invariant under rotations of that great circle around those poles, i.e. that $\Pr(A|G) = a$ for some a and all $G \in \mathcal{G}_1$. It follows by Weak Conglomerability that $\Pr(A) = a$.

Let A' be the hypothesis that P is in the band the circles just discussed sweep out when they are rotated around some axis perpendicular to the axis that passes through the poles at their centers. Let $\mathcal{G}_2$ be a "partition" of the sphere into great circles that pass through the poles of that axis. Take the unique great circle G^* that is in both $\mathcal{G}_1$ and $\mathcal{G}_2$. $A\&G^* = A'\&G^*$, so Conjunction Identity requires that $\Pr(A|G^*) = \Pr(A'|G^*)$. But by the same Rotation Preservation and Weak Conglomerability argument given above, $\Pr(A') = \Pr(A'|G^*)$. It follows that $\Pr(A) = \Pr(A')$.

But Finite Additivity implies $\Pr(A') > \Pr(A)$:

> **Finite Additivity.** If $A\&B = \varnothing$, then it is required that $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

$A' \setminus A$ has positive area, so $\Pr(A' \setminus A) > 0$. Thus, Finite Additivity requires $\Pr(A') = \Pr(A \cup (A' \setminus A)) = \Pr(A) + \Pr(A' \setminus A) > \Pr(A)$.

The fact that $\mathcal{G}_1$ and $\mathcal{G}_2$ are not technically partitions is a nuisance. Intuitively, it should not make any difference how we handle the two of the uncountably points on the surface of the sphere. But we need to handle them somehow.

?, 84, fn. 3 suggests two options. The first is to remove the two points from each element of $\mathcal{G}_1$ and $\mathcal{G}_2$ and add them to the partition. This approach requires tweaking the proof. One option is to extend Weak Conglomerability so that it applies to "not-quite-partitions" such as the modified $\mathcal{G}_1 \setminus B_1$ and $\mathcal{G}_2 \setminus B_2$, where B_1 and B_2 are the endpoints of the axes of rotation that generate $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively. Weak Conglomerability certainly seems plausible as applied to those sets, which is all that my purposes require. However, reformulating it in a general way so that it applies to those sets without generating trouble elsewhere seems to be a nontrivial task.

For that reason, I suggest the following modified proof. Let A^* be the hypothesis that P is in the union of the circular regions of points that are within some number of degrees $0 < d < 90$ of the poles in some system of latitudes and longitudes *such that* $\Pr(A|G) = 1/2$ *for some element* G *of* $\mathcal{G}_1 \setminus B_1$*, including one of those poles but not the other.* That such an A^* exists is obvious: for any G in $\mathcal{G}_1 \setminus B_1$, $\Pr(A|G)$ must go from 0 to 1 as d goes from 0 to 90, so continuity considerations dictate that $\Pr(A|G) = 1/2$ for some $0 < d < 1$. The omission of a single point (one of the poles) from A^* makes no difference to $\Pr(A^*|G)$.

Rotation Preservation requires $\Pr(A|G) = 1/2$ for all $G \in \mathcal{G}_1 \setminus B_1$. Because A^* includes one of the poles but not the other, $\Pr(A^*|B_1) = 1/2$ as well. Thus, Weak Conglomerability requires $\Pr(A^*) = 1/2$.

Let A' be the same hypothesis as before, but remove the endpoint of the axis that was removed from A and add one of the endpoints of the axes around which A was rotated to generate A'. Take the unique great circle that passes through the poles of both of those axes and remove the endpoints of those axes to get the almost-great-circle G^*. $A\&G^* = A'\&G^*$, so Conjunction Identity requires that $\Pr(A|G^*) = \Pr(A'|G^*) = 1/2$.

$\Pr(A'|G^*) = \Pr(A'|G^{**}) = 1/2$, where G^{**} is $G^* \cup B_1$. By rotation preservation, it follows that $\Pr(A'|G) = 1/2$ for all $G \in \mathcal{G}_2 \setminus B_2$. $\Pr(A'|B_2) = 1/2$ as well, so by Weak Conglomerability $\Pr(A') = 1/2$.

Thus, we have $\Pr(A) = \Pr(A')$ in violation of Finite Additivity.

The weak point of this argument is the claim that $\Pr(A|B_1) = \Pr(A|B_2) = 1/2$. This claim seems intuitively right, but trusting one's intuitions about probabilities conditional on sets of probability zero can be dangerous.

The other option Easwaran mentions is to remove the endpoints of the axes from the sphere entirely. The problem with this approach for present purposes is that it destroys the symmetry between the two great circles in the setup of the counterexample. One can restore that symmetry by removing two additional points corresponding to the endpoints of a third axis perpendicular to the other two. Better yet, remove the whole great circle perpendicular to the two great circles in that problem. Then the argument goes through essentially unmodified.

References

Bertrand, Joseph. 1889. *Calcul des probabilitiés.* Paris: Gauthier-Villars.

Borel, Émile. 1909. *Eléments de la théorie des probabilités.* Paris.

de Finetti, Bruno. 1974. *Theory of probability: a critical introductory treatment.* Probability and Statistics Series. John Wiley & Sons Australia, Limited. ISBN 9780471201410.

Dubins, Lester. 1975. "Finitely Additive Conditional Probabilities, Conglomerability and Disintegrations." *The Annals of Probability* 3:pp. 89–99. ISSN 00911798.

Easwaran, Kenny. 2008. *The Foundations of Conditional Probability.* Ph.D. thesis, University of California, Berkeley.

Edwards, A.W.F. 1972. *Likelihood. An Account of the Statistical Concept of Likelihood and Its Application to Scientific Inference. [Mit Fig. U. Tab.].* Cambridge University Press.

Fisher, Ronald A. 1922. "On the mathematical foundations of theoretical statistics." *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 222:309–368.

Good, I. J. 1949. "Comment on Statistical Inference." *Journal of the Royal Statistical Society. Series B (Methodological)* 11:140–2. ISSN 00359246.

Kolmogorov, A.N. 1956. *Foundations of the theory of probability.* AMS Chelsea Publication. Chelsea Pub. Co.

Li, Yiyuan, Li, Hong, Decety, Jean, and Lee, Kang. 2013. "Experiencing a Natural Disaster Alters Childrens Altruistic Giving." *Psychological Science* 24:1686–1695. doi:10.1177/0956797613479975.

Norton, John. 2008. "Ignorance and Indifference." *Philosophy of Science* 75:45–68. doi:10.1086/587822.

Norton, John D. 2011. "Challenges to Bayesian Confirmation Theory." In D.M. Gabbay, P.S. Bandyopadhyay, P. Thagard, and M.R. Forster (eds.), *Philosophy of Statistics*, Handbook of the Philosophy of Science, 391–439. Amsterdam: Elsevier.

Royall, Richard. 2000. "On the Probability of Observing Misleading Statistical Evidence." *Journal of the American Statistical Association* 95:pp. 760–768. ISSN 01621459.

Royall, R.M. 1997. *Statistical Evidence: A Likelihood Paradigm.* Monographs on Statistics and Applied Probability. London: Chapman & Hall.