What Conditional Probability Must (Almost) Be

Kenny Easwaran

Probability theory first reached its modern form in the work of Kolmogorov.[1] Along with his famous axioms for unconditional probability, he gave a formula for calculating conditional probabilities. "If $P(A) > 0$, then the quotient $P_A(B) = \frac{P(AB)}{P(A)}$ is defined to be the *conditional probability* of the event B under the condition A."[2] (Throughout this paper, I will use the more standard notation $P(B|A)$ in place of Kolmogorov's $P_A(B)$.) However, since this equation gives no answer for conditional probabilities when the antecedent has probability 0, several philosophers have given different axiomatizations, taking conditional probabilities as basic and defining unconditional probabilities in terms of them.[3] In his recent paper [4], Alan Hájek points out that conditional probability is in fact a pre-theoretic notion, and thus can't be taken to be a purely technical. Thus, each of these proposed sets of axioms is an analysis of the notion, and not a definition, despite Kolmogorov's use of the word "defined". Hájek then goes on to argue that Kolmogorov's analysis is insufficient, and that we must therefore adopt something like Popper's axioms instead, taking conditional probability to be basic and analyzing unconditional probability in terms of it.

However, I will argue that there is *no* analysis of conditional probability that could be correct while assigning a value to every pair of events, as Popper requires.[4] This argument will rely on a sort of "reflection principle"[5] stating that if B is the event that exactly one of some pairwise mutually impossible events E_α occurs, then $P(A|B) \geq \min\{P(A|E_\alpha)\}$. In addition, I will show that there is a function that satisfies the standard axioms as well as this principle, and that this function was discussed by Kolmogorov already in his foundational work. Any function that could claim to represent conditional probability must almost equal this function. However, this function will take three arguments instead of the standard two for conditional probability. Also, [11] points out that in at least some probability spaces, any such function must violate certain intuitive constraints on a probability function. Thus, conditional probability must often be taken to be defined merely relatively, not absolutely as Hájek

[1] [6], p. 2

[2] [6], p. 6

[3] See [8], appendix (v*), [10] and [12]. [9] also gives an axiomatization taking conditional probabilities to be basic, but he does so to solve a different problem, and his axiomatization still faces the zero divisor problem.

[4] "Whenever there is a probability $p(b, a)$ - *i.e.* a probability of b given a - then there is always a probability $p(a, b)$ also." [8], p. 326

[5] See van Fraassen, "Ulysses and the Sirens"

wants. In addition, in some cases it may not be able to be defined at all! At any rate, whenever it exists, it must be (almost) equal to a function given by Kolmogorov himself, allowing conditional probability to be analyzed in terms of unconditional.

1 Hájek's Argument

Hájek starts by proving what he calls the "Four Horn Theorem":

> Any probability assignment defined on an uncountable algebra on an uncountable set either 1. assigns zero probability to uncountably many propositions; or 2. assigns infinitesimal probability to uncountably many propositions; or 3. assigns no probability whatsoever to uncountably many propositions; or 4. assigns vague probability to uncountably many propositions.[6]

In any of the cases forced to exist by this theorem, Kolmogorov's ratio analysis forces $P(A|B)$ to be undefined. However, Hájek gives examples of each of these cases in which there is a clear, intuitive value for what the conditional probability should be. Thus, he concludes the ratio analysis is incorrect, because it fails to account for the full extension of the conditional probability function.

He then goes on to note the following:

> The examples of vague and undefined probabilities suggest that the problem with the ratio analysis is not that it is a *ratio* analysis, as opposed to some other function of unconditional probabilities. The problem lies in the very attempt to analyze conditional probabilities in terms of unconditional probabilities at all. It seems that any other putative analysis that treated unconditional probability as more basic than conditional probability would meet a similar fate - as Kolmogorov's elaboration (RV) did.[7]

However, what he calls "Kolmogorov's elaboration" is the function that I will end up supporting. While the ratio analysis leaves out some conditional probabilities that should be defined, I will argue that the accounts Hájek prefers taking conditional probability as basic will include some conditional probabilities that shouldn't be defined. While this elaboration will still have some trouble dealing with cases of vague or undefined probabilities, most accounts of how to deal with such unconditional probabilities will lead to a natural generalization of this elaboration. Thus, I contend that this elaboration is what conditional probability must (almost) be.

[6] [4], p. 284
[7] [4], p. 315

2 The Borel Paradox

I will concede all of the specific intuitions that Hájek uses in his paper. However, I will show that any natural generalization of these intuitions will lead to inconsistency even in one situation that he himself discusses.

> Suppose that we have a uniform probability measure over the Earth's surface (imagine it to be a perfect sphere). What is the probability that a randomly chosen point lies in the western hemisphere (W), given that it lies on the equator (E)? $1/2$, surely. But the probability that the point lies on the equator is 0, since the equator has no area. ... We could have asked a similar question for each of the uncountably many lines of latitude, with the same answer in each case.[8]

Hájek attributes this scenario to Borel, and I will follow Kolmogorov in calling this the "Borel Paradox".[9] I will use somewhat different notation from Hájek and go into more detail than Kolmogorov. In addition, I will follow Hájek and Kolmogorov in sometimes identifying an event with a certain set of worlds or outcomes of an experiment, but this will be merely for notational convenience, rather than as a metaphysical claim about events.

Consider a sphere of surface area 1, with center O, and let X be a point chosen uniformly at random from the surface. For any (measurable) region E, $P(E)$ is given by the area of E. If Y is any point on the sphere, and $0 \leq \theta_1 \leq \theta_1 \leq \pi$, then I will define the event S_{Y,θ_0,θ_1} to occur just in case $\theta_0 \leq \angle XOY \leq \theta_1$. For instance, if Y is the north pole, then $S_{Y,0,\pi/2}$ is the northern hemisphere, and $S_{Y,2\pi/3,3\pi/4}$ is the region between 30 and 45 degrees south latitude. By a simple integration, we can find that the area of this region (and thus the probability of the event) is $\frac{\cos\theta_0 - \cos\theta_1}{2}$. I will also define C_Y to be $S_{Y,\pi/2,\pi/2}$, so that if Y is either pole then C_Y is the equator, if Y is on the equator at 90 degrees west longitude then C_Y is the Greenwich meridian, and in general C_Y is some great circle on the surface of the earth. It is easy to verify that $P(C_Y) = 0$.

2.1 Two Generalizations of Hájek's Intuition

If we let Y be the point on the equator at 90 degrees west longitude, and Z be the north pole, then $W = S_{Y,0,\pi/2}$ is the western hemisphere and $E = C_Z$ is the equator. Hájek suggested that intuitively, $P(W|E) = 1/2$, even though the ratio is undefined. There are two ways to generalize this intuition to angles other than 0 and $\pi/2$. One way is to note that exactly half of the length of E lies within W, and to suppose that since the unconditional distribution is uniform on the surface of the earth, the distribution conditional on E should be uniform on the length of E. Another way is to note that the symmetry of

[8] [4], p. 289

[9] [6], pp. 50-51

the earth suggests that $P(W|E_\alpha)$ should be equal for any great circle E_α that goes through Y. Thus, W should be independent of each E_α (since every point other than Y and its opposite is in exactly one E_α), and in particular it should be independent of E. Thus, $P(W|E)$ should equal $P(W)$, which is 1/2.

However, these two intuitions come apart for most pairs of angles other than $(0, \pi/2)$. So let N be the north pole and let A be $S_{N,0,\pi/6}$, which is the collection of all points with latitude at least 60 degrees north. Let B be the great circle comprising the Greenwich meridian. Since exactly 1/6 of the length of B lies within the region A, the first intuition suggests that $P(A|B) = 1/6$. But the symmetry argument of the second intuition applies if we consider all the lines of longitude now, and they suggest that $P(A|B) = P(A) = \frac{2-\sqrt{3}}{4}$. In general, for S_{Y,θ_0,θ_1}, the first intuition gives $P(A|B) = \frac{\theta_1-\theta_0}{\pi}$ while the second gives $P(A|B) = P(A) = \frac{\cos\theta_0-\cos\theta_1}{2}$. For any of the uncountably many possible values of θ_0, at most three of the uncountably many possibilities for θ_1 will make these two values agree. Thus, almost always, at least one of these two intuitions will have to be wrong. In a sense, Hájek just happened to be lucky in choosing 0 and $\pi/2$ so that he didn't have to choose between these intuitions. This separation of intuitions is what led Kolmogorov to call this scenario the "Borel Paradox".

2.2 Vindication of the Second Intuition

Because these intuitions seem to have led us astray, I will now argue more carefully and show that the second intuition is basically correct and that $P(A|B)$ is almost certainly $P(A)$. Let A be the region S_{Y,θ_0,θ_1}, and let $\mathcal{E}$ be the set of all great circles E_α that go through Y. Assume that there is some conditional probability $P(A|E_\alpha)$ for each $E_\alpha \in \mathcal{E}$. I will show that for almost all the E_α, $P(A|E_\alpha) = P(A)$. That is, if B is the union of the E_α such that $P(A|E_\alpha) \neq P(A)$, then $P(B) = 0$.

To do this, I will define a function h on the surface of the sphere, such that if w is any point on the surface other than the poles, I will let $h(w) = P(A|E_\alpha) - P(A)$, where E_α is the unique element of $\mathcal{E}$ containing w. If w is a pole, then I will let $h(w) = 0$. The desired result will be to show that $P(h(w) \neq 0) = 0$. To do this, I will have to assume that we accept Kolmogorov's axiom of countable additivity for probability functions. This axiom has been questioned by some probabilists. But even without using this assumption, I will be able to prove for every positive ϵ that $P(h(w) > \epsilon) = 0$ and $P(h(w) < -\epsilon) = 0$. With countable additivity, this is equivalent to $P(h(w) \neq 0) = 0$, but even without it, this seems like a strong enough constraint on h to justify saying that this means that $P(A|E_\alpha)$ is "almost" the same function as the constant function $P(A)$.

To prove this result by contradiction, I will assume that there is some positive ϵ such that $P(h(w) > \epsilon)$ is positive. (The case where $P(h(w) < -\epsilon)$ is similar.) Let B_ϵ be the event that $h(w) > \epsilon$. Because of the way h is defined, its value only depends on which great circle E_α the point w lies on. Thus, B_ϵ is the disjoint union of some collection of these circles. For each of these circles, it is clear by definition that $P(A|E_\alpha) > P(A) + \epsilon$. I will argue that therefore $P(A|B_\epsilon) \geq P(A)+\epsilon$. This is an instance of the "reflection principle" mentioned above.

Assume that $P(A|B_\epsilon) = P(A)+\epsilon-\delta$. When we are in a state of knowledge[10] that tells us that X is in B_ϵ and nothing else, this is the probability that we should assign to A. But now imagine we set up some experiment to tell us which unique E_α that composes B_ϵ the point X is in. We would then assign probability $P(A|E_\alpha)$ to the occurence of A. Because each $P(A|E_\alpha) > P(A)+\epsilon$, we see that this would result in an increase in our credence for A of at least δ, no matter how the experiment turned out. Thus, it seems that just performing the experiment without observing the result will allow us to increase our credence in A by at least δ, which is absurd. To avoid this outcome, $P(A|B_\epsilon)$ should be at least $P(A) + \epsilon$ as claimed above.

But now, since I have assumed that $P(B_\epsilon)$ is positive (for the sake of a contradiction that I will soon achieve), everyone will agree that $P(A) + \epsilon \leq P(A|B_\epsilon) = \frac{P(A\&B_\epsilon)}{P(B_\epsilon)}$, because problems with the ratio analysis arise only when the antecedent has probability zero. Multiplying through, we see that $P(A)P(B_\epsilon) + \epsilon P(B_\epsilon) \leq P(A\&B_\epsilon)$, so that $P(A)P(B_\epsilon) < P(A\&B_\epsilon)$. However, it is not hard to check that because A is rotationally symmetric around the point Y and B is composed entirely of great circles through Y, $P(A\&B_\epsilon)$ must equal $P(A)P(B_\epsilon)$, so this is a contradiction. Therefore, for any positive ϵ, $P(B_\epsilon) = P(h > \epsilon) = 0$, QED.

Since $h(w)$ measured the difference between $P(A|E_\alpha)$ and $P(A)$ for the great circle E_α containing w, this means that the conditional probability of A must be almost equal to the unconditional probability almost everywhere. This is exactly what the second intuition said.

The first intuition suggested that since the unconditional probability was uniformly distributed over the surface of the sphere, the conditional probability should be uniformly distributed along the length of the great circle. But this presumes that area and length are related in the way that unconditional and conditional probability should be. This presumption sounds natural at first, but I think it is a bit too fast. If the space we were considering hadn't had a uniform distribution, this presumption would clearly have been unjustified. Although the second intuition seems slightly odd, it is supported by the reflection principle I mentioned above. This reflection principle will give rise to similar constraints on conditional probabilities in arbitrary probability spaces, and thus this intuition generalizes where the first one doesn't. Therefore, I suggest that the second intuition is the correct generalization of Hájek's intuition, even though the first seems initially slightly more natural.

[10]The argument in this paragraph assumes that the probability function described here is a subjective probability function. But even if it is supposed to be an objective function, I think the principle should still hold.

3 Generalization

In the case of the Borel paradox, I argued that where A is some region with rotational symmetry around Y, for almost all of the great circles E_α through Y, it must be the case that $P(A|E_\alpha) = P(A)$. This was effectively done by finding a function g_A (in this case the constant function whose value was always $P(A)$) such that for any B that is the union of some of the E_α, we must have $P(A\&B) = \int_B g_A(w)dw$. Then letting $f(w) = P(A|E_\alpha)$ for the unique E_α containing w, I used the reflection principle to show that $P(A\&B) = \int_B f(w)dw$. Letting $h(w) = f(w) - g(w)$, I showed that h must be almost equal to 0.

The same procedure will work for arbitrary probability spaces and arbitrary partitions $\mathcal{E}$ of the space into various E_α. Assuming that conditional probabilities are always defined, I will let $f(w) = P(A|E_\alpha)$. If B is the union of finitely many such E_α, then clearly $P(A\&B) = \sum P(A|E_\alpha)P(E_\alpha) = \int_B f(w)dw$. If B is an arbitrary union of some of the E_α, the proof requires the reflection principle, and is done in the first appendix. Now, if f and g_A are *any* functions satisfying this integral equation (so that for all B that are unions of some E_α, we have $P(A\&B) = \int_B f(w)dw = \int_B g_A(w)dw$), then the function $h(w) = f(w) - g_A(w)$ must be almost everywhere zero. (If not, then we can just let B be some region of positive probability on which f and g_A differ by at least ϵ and we get a contradiction.) Thus, if g_A is a function satisfying the integral equation for events B composed of elements from some partition $\mathcal{E}$, then the conditional probability of A given events in this partition must be almost equal to the value of g_A.

If we assume countable additivity, then the Radon-Nikodym theorem of real analysis guarantees that such a function g_A always exists.[11] If such a function exists, then it is what conditional probability must almost be. Since the equality is only "almost" equality, note that if $P(E_\alpha) = 0$, then $P(A|E_\alpha)$ may differ from g_A on the set E_α. So it seems that this function doesn't specify the conditional probability for antecedents of probability zero, which is exactly when we need to use this rather than the ratio analysis. However, only a relatively small number of the E_α can vary from the constraint, so the existence of a function g_A satisfying the integral equation will settle the values of almost all conditional probabilities, though a few of these values may be incorrect.

Note that this is much better than the ratio analysis, which stayed silent for probabilities conditional on any event of probability 0, while this one gives us an answer, albeit one that might be wrong for a few of these events.

4 Problems with the Analysis

Now that I've shown that this second intuition is in general right, and is also better than the first because it generalizes to arbitrary probability spaces, I will show some potentially unappetizing consequences of adopting it. In the end, I think these problems will combine with the results I have shown so far to suggest that conditional probabilities must not be taken to be basic, and in fact

[11] [6], p. x; [2], p. x

must be undefined in many cases, even when all the unconditional probabilities have well-defined values.

4.1 A Three-Place Function

Let Y be the north pole, let Z_1 be the point on the equator at 90 degrees west longitude, and let Z_2 be the intersection of the equator and the Greenwich meridian. Let A_1 be $S_{Y,0,\pi/6} \cup S_{Y,5\pi/6,\pi}$, which is the union of the disc mentioned above with its mirror image in the south. That is, A_1 is the set of all points either further than 60 degrees north or further than 60 degrees south. Let A_2 be $S_{Z_1,\pi/3,2\pi/3}$, which is the band generated by rotating A_1 around the point Z_1. Let B be C_{Z_2} - the great circle through Y and Z_1, comprising the lines of longitude 90 degrees west and 90 degrees east.

By the symmetries mentioned above around Y, $P(A_1|B)$ is almost certainly $P(A_1) = \frac{2-\sqrt{3}}{2}$. By the symmetry around Z_1, we see that $P(A_2|B)$ is probably $P(A_2) = 1/2$. However, if B is assumed, then A_1 occurs iff A_2 does, so $P(A_1|B)$ should equal $P(A_2|B)$. If the former were larger, then simple rules of the probability calculus that every proposed set of axioms satisfy would suggest that $P(A_1 \& \neg A_2|B) > 0$, which would assign positive probability to an impossible event. To avoid this, the two conditional probabilities must be equal.

One might try to avoid the potentially dire consequences of this trilemma by noting that one or both of the $P(A_i|B)$ could diverge from the value suggested above, because each of these values has to occur *almost* everywhere, rather than everywhere. I explore this possibility in the second appendix, but note here only that this option seems to require some extra set-theoretic assumptions to pursue. In addition, the function given will be highly non-symmetric and the construction will use the axiom of choice, preventing us from knowing any particular one of its values.

Instead, I propose noticing the other assumption used in the argument much earlier - that all the stated conditional probabilities exist. Because the first intuition was discredited, and the second intuition gives two separate values here, I suggest that there must be *no* particular value for this conditional probability. Since the analyses of Popper and others require that every conditional probability exist, these analyses are wrong for the opposite reason that Kolmogorov's ratio analysis was. These accounts give a value where there is none, just as Kolmogorov's account gave no value where there was one.

However, some role for conditional probability can be saved by defining conditional probabilities only relative to a partition of the space, rather than absolutely. Recall that for any partition $\mathcal{E}$ of the space, the conditional probability f_A must satisfy the integral equation $P(A \& B) = \int_B f_A(w)dw$ for all B that are unions of some elements of the partition. Rather than letting this function give the values $P(A|E_\alpha)$ absolutely, I will relativize the function to the partition, so that it gives $P(A|E_\alpha, \mathcal{E})$, where E_α is some element of the partition $\mathcal{E}$. Thus, conditional probability is a three-place function, rather than a two-place function as we might have expected.

This move is not as bad as one might fear. Note that when $B = E_\alpha$ is some member of the partition $\mathcal{E}$ with positive unconditional probability, the integral equation requires that $P(A\&B) = \int_B f(w)dw = P(A|E_\alpha)P(E_\alpha)$. But then this just means that $P(A|E_\alpha) = \frac{P(A\&E_\alpha)}{P(E_\alpha)}$, so that the value is given by the ratio analysis, and doesn't depend on the third argument at all. Thus, we can be forgiven for not having noticed this general dependency on the partition considered, because it didn't arise in the cases normally considered, where the antecedent has positive probability.

In addition, in all the cases Hájek considers, even though $P(E_\alpha) = 0$, the value of $P(A|E_\alpha, \mathcal{E})$ is independent of the partition $\mathcal{E}$ from which E_α is drawn. He may have just been lucky, but I think that he chose these particular examples for their maximal intuitive pull. It seems plausible that the strength of the intuition is related to the lack of ambiguity in value on this account. Thus, these particular conditional probabilities can be defined absolutely, even though in general the value might be relativized to a partition of the space.

4.2 An Argument for Relativization

Regardless of the importance of these mathematical considerations, there are independent reasons one might prefer a relativized conditional probability function to an absolute one. Alfred Rényi made the following argument in an attempt to support a different point, but I think it applies here:

> In general, it makes sense to ask for the probability of an event A only if the conditions under which the event A may or may not occur are specified and the value of the probability depends essentially on these conditions. In other words, *every probability is in reality a conditional probability*. This evident fact is somewhat obscured by the practice of omitting the explicit statement of the conditions if it is clear under which conditions the probability of an event is considered.[12]

As stated, I think that Hájek doesn't believe this argument is sound, or else his paper would have been redundant. But I think that as stated, even Rényi didn't believe this argument to be sound. The "conditions under which the event A may or may not occur" sound like they should include background assumptions, like the one "that the pack is complete and well shuffled, etc."[13] But Rényi never talks about a probability space within which the event of the pack being complete and well shuffled has a measure. Rényi only ever talks about conditional probabilities, but the antecedents of these conditionals are always events *within* a probability space, rather than these background assumptions that are necessary to define the probability space to begin with. Thus, I think Rényi's argument has shown merely that every probability is in reality *relative to a model*, and not actually conditional.

[12] [9], pp. 34-35
[13] [9], p. 35

Thus, what looks like a one-place unconditional probability function is actually a function with a hidden place for a probability model to appear as an argument. Taking conditional probabilities seems to add one more place for the antecedent of the conditional, but I claim here that it actually adds one more place as well, for the partition from which the antecedent is drawn.

This extra piece of relativization is just the sort one might expect. Just as an unconditional event is always drawn from some probability model, the antecedent of a conditional probability is always drawn from some hypothetical experiment. If one learns E_α, then there was some set $\mathcal{E}$ of ways that the experiment performed could have turned out, and I claim that these possible outcomes will partition the space in just the way required for this relativized conditional probability function. This is true whether the probability function involved is objective or subjective, and whether the conditionalization is on actual knowledge or a hypothetical advance in one's knowledge.

[In the well-known Monty Hall paradox, if the contestant knows that Monty will open a random door that doesn't contain the prize, but not the one she originally chose, then the probability is 2/3 that the prize is behind door 1, given that the contestant originally chose door 2 and Monty revealed door 3. But if the contestant instead knows merely that Monty will open a random door other than the one she originally chose, then the probability is 1/2 that the prize is behind door 1, given that the contestant originally chose door 2 and Monty revealed the lack of prize behind door 3. The conditional probabilities depend on the way the particular piece of knowledge was gathered, and the relativization mentioned here is just a generalization of this fact.]

[This may also be seen as a version of the reference class problem. While an unconditional probability depends on specifying the model that provides the reference class for alternative events that could occur, the conditional probability also depends on specifying the reference class of alternatives to the antecedent.]

Thus, the relativization required is not as big a problem as one might have initially feared.

4.3 Seidenfeld's Objection

A more pressing concern is raised by [11]. In this paper, an extra desideratum for conditional probability is mentioned. In addition to the Kolmogorov axioms for unconditional probability, Seidenfeld et al require also that for any element $B \in \mathcal{E}$, the function is "proper" at B, i.e. that $P(B|B, \mathcal{E}) = 1$. They also consider $P(A|B, \mathcal{E})$ when $\mathcal{E}$ is an arbitrary "sub-σ-field" that contains B from the original probability space, rather than just a partition.

In some spaces, like the uniform distribution on the interval $[0, 1]$ with events being only the countable and co-countable sets, and $\mathcal{E}$ consisting of all the events, there is a function satisfying all my constraints but not the additional one. This would clearly be a problem if every function I endorsed had this property. However, there is a particularly natural function they describe that satisfies their additional constraint as well as all of mine, so this space is not much of a problem.

However, in other spaces, they show that for certain values of $\mathcal{E}$, any function satisfying the integral equation must be improper at some points, and in some cases must be improper almost everywhere! They claim that this is evidence that conditional probabilities should be given not by the three-place function I suggest, but rather a finitely additive two-place function described in [1]. However, they also point out that the function they recommend fails to satisfy the integral equation for some unusual events B.

I think this is just a stronger version of my claim from before that the conditional probabilities don't exist in these cases. Before, I claimed that they don't exist in any absolute sense, but merely relative to a partition. Seidenfeld et al have shown that relativized to certain partitions, any such function will have to violate other important constraints. Thus, I suggest that in these cases, even the relative conditional probabilities don't exist. Their preferred functions violate the reflection principle that I proposed; my preferred functions violate the propriety requirement they proposed. Any analysis that takes conditional probability as basic will have to violate one or both of these conditions, based on my results and those of Seidenfeld et al. Thus, there is no conditional probability in these cases.

However, these cases seem to arise only in particular spaces and when considering general sub-σ-fields rather than just partitions of the space as I consider. In other cases, I think the function Kolmogorov proposed will be a satisfactory treatment of conditional probability. And when it doesn't work, nothing can. This is what conditional probability must (almost) be, whenever it exists.

5 Appendices

5.1 Proof that conditional probabilities satisfy the integral equation

Let $\mathcal{E}$ be some partition of a probability space into disjoint events E_α. Let $f(w) = P(A|E_\alpha)$, where E_α is the unique element of the partition containing the point w. I want to show that for any B that is the union of some collection of E_α, $P(A\&B) = \int_B f(w)dw$.

Note that the integral used here is the Lebesgue integral, which is defined as the supremum of the sums $\sum x_i P(h(w) = x_i)$ over functions h that are 0 outside B, everywhere bounded above by f, taking on only finitely many distinct values x_i. To make the notation clearer, for such a function h, I will define $\int h(w)dw = \sum x_i P(h(w) = x_i)$. It is clear that if h' is some function 0 outside B, everywhere bounded *below* by f, taking on only finitely many distinct values x'_i, then $\int_B f(w)dw \leq \int h'(w)dw$, because h' is an upper bound for every h that is considered in calculating the value of the integral.

Thus, if for every n I can find h_n and h'_n that are 0 outside B and such that everywhere in B, $h_n \leq f \leq h'_n$ and $\int h(w)dw \leq P(A\&B) \leq \int h'(w)dw$ and $\int h'(w)dw - \int h(w)dw \leq 1/n$, then I will have proven the integral equation. This is because the integral of h' is an upper bound for the integral of f and the integral of h is a lower bound, and since n is arbitrary, they can be made arbitrarily close.

So now let $h(w)$ and $h'(w)$ both be 0 for $w \notin B$ and let $h(w) = \max\{\frac{k}{n} : \frac{k}{n} \leq f(w)\}$ and $h'(w) = h(w) + 1/n$ for $w \in B$. Because $f(w)$ was defined in terms of which element E_α of the partition contained w, we see that f is constant on the E_α, so h and h' are too. Thus, the set B_k where $h(w) = \frac{k}{n}$ (which is also where $h'(w) = \frac{k+1}{n}$) is a union of some of the E_α. In particular, it is the union of the E_α where $\frac{k}{n} \leq P(A|E_\alpha) < \frac{k+1}{n}$.

Thus, by the reflection principle, we see that $\frac{k}{n} \leq P(A|B_k) \leq \frac{k+1}{n}$. Multiplying through by $P(B_k)$, we get $\frac{k}{n}P(B_k) \leq P(A\&B_k) \leq \frac{k+1}{n}P(B_k)$. Summing over all k from 0 to n (these are the only relevant values, because f was bounded between 0 and 1), we get $\int h(w)dw \leq P(A\&B) \leq \int h'(w)dw = \int h(w)dw + P(B)/n$. But this is just the inequality we wanted earlier, since $P(B) \leq 1$.

Thus, the argument goes through as desired.

5.2 Exploration of an attempt to define conditional probabilities in a non-relativized way

In this appendix I will show how Kolmogorov's extended analysis of conditional probabilities can be used to define every conditional probability in the Borel paradox absolutely, rather than just relative to a particular axis. However, this non-relativized conditional probability function will be highly non-unique, and specifying any particular such function will require a well-ordering of the reals. In addition, this proof requires controversial set-theoretic principles beyond just the Axiom of Choice. Because of these problems, I will endorse the relativized function of the main body of the text, rather than this absolute function, which may not even exist for many spaces.

Recall that if we consider a particular axis, then relative to the partition of the sphere into the great circles through this particular axis, the conditional probability of any set given any great circle is $(\cos\theta_0 - \cos\theta_1)/4$ if it intersects the great circle in the interval from θ_0 to θ_1 radians away. It isn't necessary that every great circle have this property for all its conditional probabilities, just that for any axis, the set of great circles that do should together form a set of measure 1 (by the result of the previous appendix). Since these values depend greatly on the point from which the angles are measured, each great circle can only give all the "correct" values for at most one axis it goes through. Thus, in order to define these conditional probabilities absolutely, I will associate each great circle with a single axis it goes through in such a way that the set of great circles associated with any particular axis covers a region of measure 1 on the surface of the sphere.

To do this, I will assume the Continuum Hypothesis (that any set of real numbers with cardinality strictly less than that of the set of all real numbers is in fact countable)[14]. There are just as many possible axes for a sphere as

[14] This proof will in fact go through using Martin's Axiom, which is weaker than the Con-

there are real numbers. So using the Axiom of Choice, let us well-order them in the shortest order-type possible, so that (by the Continuum Hypothesis), every axis has only countably many predecessors in the ordering. Then associate each great circle with the axis on it that comes earliest in the well-ordering. Thus, each great circle will be associated with a unique axis it goes through, so we can define the probability of any set conditional on this great circle to be the value it should have relative to this particular axis.

Since each circle is associated with the earliest axis on it in the well-ordering, the only way a particular great circle through an axis A can be unassociated with A is if it contains an axis that comes earlier than A in the well-ordering. But as mentioned above, there are only countably many axes earlier than A, and any pair of axes have exactly one great circle going through both of them, so there are at most countably many great circles through A that aren't associated with A. By countable additivity, the great circles not associated with A have measure 0, so the ones associated with axis A have measure 1. Since this is true for any axis A, the association of great circles with axes has the property required above, QED.

This result is highly counterintuitive. It says that although any given great circle gives the "wrong" probabilities for almost every axis on it, we can make sure that for any axis, almost every great circle through it gives the "right" probabilities. I would argue that this highlights some counterintuitive aspects of the Continuum Hypothesis (or Martin's Axiom) when combined with the Axiom of Choice.

Using this association of circles with axes, we can define all the conditional probabilities discussed above absolutely, rather than relatively, in a way that preserves all the integrals needed. However, it is not clear that it is possible to extend this definition to probabilities conditional on other sets of measure zero (say, lines of latitude rather than longitude). It is also unclear whether these probabilities will satisfy the appropriate integral equations relative to other partitions of the space not considered here.

More importantly, the conditional probabilities so defined are highly non-uniform, and depend on the well-ordering of the set of axes. Such a well-ordering can only be given in a highly non-constructive way, using the Axiom of Choice, and there are many more such well-orderings than there are real numbers. Each well-ordering gives a different set of conditional probabilities, so it is particularly hard to justify any set of these as the "correct" set of conditional probabilities for this example.[15]

---

tinuum Hypothesis. Both are known to be consistent with standard ZFC set theory. Using Martin's Axiom, every cardinal smaller than the number of real numbers behaves like $\aleph_0$, and in particular, the union of κ many sets of measure 0 is itself a set of measure 0 under standard Lebesgue measure, assuming that $\kappa < 2^{\aleph_0}$. For more information on these axioms, see [7] p. 51: "Unlike the basic axioms of ZFC, MA does not pretend to be an 'intuitively evident' principle, and in fact at first sight it seems strange and ill-motivated."

[15] Compare the arguments against assigning infinitesimal values to probabilities on [4], p. 292, where he points out that there is no way to specify a particular infinitesimal, since their identity depends on a particular set constructed using the Axiom of Choice.

6 Acknowledgements

I would like to thank audiences at the Berkeley Student Logic Colloquium and the Formal Epistemology Workshop in Austin for many insightful questions and comments. In particular, Peter Gerdes and Teddy Seidenfeld shaped several of my points, and Peter Vranas pointed out that the earlier version of the proof in the first appendix was unclear. I would also like to thank Branden Fitelson and Alan Hájek for criticisms, comments, and encouragement throughout the entire process.

References

[1] Dubins, Lester, 1975. "Finitely Additive Conditional Probabilities, Conglomerability, and Disintegrations," *The Annals of Probability* **3**: 89-99.

[2] Durrett, Richard, 2005. *Probability: Theory and Examples,* Belmont, CA: Brooks/Cole.

[3] Feller, William, 1968. *An Introduction to Probability Theory and Its Applications,* New York: John Wiley & Sons.

[4] Hájek, Alan, 2003. "What Conditional Probability Could Not Be," *Synthese* **137**: 273-323.

[5] Jech, Thomas, 2003. *Set Theory: The Third Millenium Edition, Revised and Expanded,* New York: Springer-Verlag.

[6] Kolmogorov, A. N., 1950. *Foundations of the Theory of Probability,* New York: Chelsea.

[7] Kunen, Kenneth, 1980. *Set Theory: an introduction to independence proofs,* New York: North Holland.

[8] Popper, Karl, 1959. *The Logic of Scientific Discovery,* New York: Harper & Row, 1968.

[9] Rényi, Alfred, 1970. *Foundations of Probability,* San Francisco: Holden-Day.

[10] Roeper, P. and Leblanc, H., 1999. *Probability Theory and Probability Logic,* Toronto: University of Toronto.

[11] Seidenfeld, Teddy; Schervish, Mark; and Kadane, Joseph, 2001. "Improper Regular Conditional Distributions," *The Annals of Probability* **29**: 1612-1624.

[12] van Fraassen, Bas, 1995. "Fine-Grained Opinion, Probability, and the Logic of Full Belief," *Journal of Philosophical Logic* **24**: 349-377.