

Probabilistic Conditionals Are *Almost* Monotonic

Matthew Johnson and Rohit Parikh *

April 16, 2007

Abstract

We examine some old and new paradoxes of probability, give a rough account of probabilistic conditionals, and prove some new results about non-monotonicity in probabilistic conditionals. It is well known that such conditionals are not monotonic – a conditional which is true can become false when additional hypotheses are added. We show that nonetheless, the conditionals are *usually* monotonic, in both finite and countably infinite universes, or roughly speaking that we do not actually have to worry about non-monotonicity in practice.

1 Probability & Paradox

Einstein famously said that *God does not play dice*. This attitude of his led to a certain estrangement with Quantum Mechanics and even with Physics during the last years of his life, otherwise marked by a close friendship with Kurt Gödel.

Einstein aside, probability is a puzzling phenomenon, for what exactly does it mean? One supposedly unproblematic way to define it is via frequencies. To say that a coin is fair, i.e., that the probability of heads is .5, can then be interpreted as, *If we toss the coin many times, then approximately half of the results will be heads*. But how many is many? Of course, if we toss the coin a hundred times, we probably will not get exactly fifty heads, and even if we toss it a thousand times, the proportion of heads can diverge quite a bit from .5. This is not *likely*, but explaining that word *likely* looks like it involves probability.

Moreover, if Jack wants to insure a car, the frequency definition is no good. The insurance company needs to know the probability that *he, Jack* will have an accident. And clearly there is no frequency available for the insurance company to resort to. Jack only wants to insure his car *once*. And even if he wants to insure the car a second time, he will be a year

*{mjohnson@cs.,rparikh}@gc.cuny.edu, City University of New York. Research supported in part by a grant from PSC-CUNY grants program. Versions of this paper were given in colloquia at Boston University, CUNY Graduate Center, and IHPST in Paris. A preliminary version of this paper appeared at LFCS 2007 [16].

older, and presumably more experienced. The probability of his having an accident, however we define it, will not be the same the second time.

An alternative approach is that due to Ramsey, de Finetti, Savage, etc. Ramsey defines probability subjectively, in terms of the bets which an agent is willing to accept [19]. Your subjective probability of heads is .5 if you are willing to accept a double or nothing bet *in favour* of heads, and also a double or nothing bet *against* heads. But what bets you want to accept is up to you. There is nothing intrinsically objective about your bet-acceptance attitudes.

There *are* standards of rationality which will apply. If you think in October 2004 that the probability that Bush will win is .5 and also that the probability that Kerry will win is .6, then bets can be placed in such a way that no matter what happens you will lose money. For an opponent can bet \$20 at even odds against Bush winning, and also \$16 at 3:2 odds against Kerry winning. Then if Bush wins, you gain \$20 from the first bet but lose \$24 on the second bet. If Kerry wins, you gain \$16 from the second bet and lose \$20 from the first, and finally if neither wins, you lose both bets. You lose in all three cases.

In other words, a *Dutch book* can be made against you, a term which already occurs with de Finetti [8] who says that he is puzzled by the name – the Dutch are even more puzzled! It is of course well known that an agent against whom Dutch book cannot be made has a subjective probability which satisfies the Kolmogorov axioms like $0 \leq p(E) \leq 1$ and $p(E \vee F) = p(E) + p(F)$ when $E \wedge F$ is null.¹

But even if two people are both rational in the sense that Dutch book cannot be made against either of them, it need not be the case that they both have the same subjective probability. Thus the question *What is the actual probability $p(E)$ of the event E ?* cannot be answered. The only solace we have is that if two agents start by assigning positive (but different) probabilities to the same events, and they have the same experiences, and they both revise using Bayes' law, then their probabilities will converge in the long run.

So the notion of probability does have some foundational problems. But I want to put these problems aside, and talk about some paradoxes, both old and recent, and conclude with a new result about probabilistic conditionals.

1.1 The Saint Petersburg Paradox

“The St. Petersburg game is played by flipping a fair coin until it comes up tails, and the total number of flips, n , determines the prize, which equals 2^n . Thus if the coin comes up tails the first time, the prize (in dollars) is $2^1 = 2$, and the game ends. If the coin comes up heads the first time, it is flipped again. If it comes up tails the second time, the prize is $2^2 = 4$, and the game ends. If it comes up heads the second time, it is flipped again. And

¹Isaac Levi [13] considers *families* of probabilities, but we shall not go into that here.

so on. There are an infinite number of possible consequences (runs of heads followed by one tail) possible. The probability of a consequence of n flips ($P(n)$) is $1/2^n$, and the expected payoff of each consequence is the prize times its probability” [21].

How much should you pay to participate in the St. Petersburg game? It is easily calculated that the expected payoff is infinite. You will win \$2 with probability .5, \$4 with probability .25, \$8 with probability .125, etc, adding up in all to ∞ . Therefore *any* amount whatsoever is acceptable as payment to enter the game. But most people rebel at the thought.

One possible solution to this puzzle is to use utilities rather than payoffs. If we say that the utility value of a financial payoff is *logarithmic* in the payoff, then the expected utility of the St. Petersburg game will be finite and it would be a mistake to pay an amount whose utility would exceed the expected utility of the game. This was essentially the suggestion of Daniel Bernoulli.

But Bernoulli’s suggestion would not prevent other paradoxes, for if the payoff were 2^{2n} with n = the number of heads before a tail, then the payoff from each outcome multiplied by the probability of that outcome would again be exponential in dollars, and hence again linear in *utility*. This would again give an infinite expected utility to the game. See also Paul Weirich [22].

1.2 The Sleeping Beauty

But let us go on to our second, much more modern paradox due to Adam Elga [7], based on an earlier paper of Piccione and Rubinstein [18]. *Sleeping Beauty* (SB) is put to sleep on a Sunday using some drug, and after that a fair coin is tossed. If the coin comes up heads, she is woken on Monday and asked the question Q (to be described later). If the coin comes up tails, she is woken up on Monday, asked the question Q , and then *put back* to sleep, woken up again on Tuesday and again asked the question Q . Sleeping Beauty knows this procedure and that the coin is fair. But when she is woken, she does not know what day of the week it is and whether she was woken before or not. Of course she does know that the day is either Monday or Tuesday, but she does not know which.

The question Q is, *What is the probability that the coin landed heads?*

One answer is that it is .5. The coin is fair and SB knows this. Moreover, she knew all along that she would be woken up. The fact that she is woken up and asked Q is not surprising information which might change the probability. So the answer is .5. *Or is it?*

Suppose it is .5, and so she accepts a double or nothing bet on heads each time that Q is asked. Over a hundred trials, there will be fifty heads and fifty tails, roughly. When the coin lands heads, she will *win* one dollar (say) and when it lands tails, she will *lose* two dollars, one for each time she is asked Q . So she will end up losing \$50 net. This is not compatible

with a .5 probability, and the right probability would be 1/3.

Which is the right answer? And is there a right answer? Various people have written on this and let us just refer you to recent papers by Halpern, and by Bradley and Leitgeb [11, 4]. Halpern claims that the problem here is *asynchrony*, an issue that arises often in distributed computing. Bradley and Leitgeb claim that the connection between betting and probability is only valid under certain presumptions, which fail in this case.

1.3 Probabilistic Conditionals

How should we interpret a conditional like, “If John comes to the party, then so will Mary,” i.e., of the form *If A then B*? The standard interpretation used in mathematics is to treat it as equivalent to $A \rightarrow B$ i.e., as $\neg A \vee B$ (see [20, 17]).

But often, this does not fit our intuition. The following example due to Dorothy Edgington is instructive. Clearly if God does not exist, then he cannot answer our prayers. Consider the statement *S*: *If God does not exist, then it is not the case that if I pray, my prayers will be answered*. Many who disagree about the existence of God will tend to accept *S*. There are two *ifs* in *S*. Can they both be expressed as material conditionals? So suppose we symbolize *S* as $\neg G \rightarrow \neg(P \rightarrow A)$, interpreting both implications, the main one on the outside, and the subsidiary one inside the parentheses, as material implications. Also suppose I don’t pray. Then *P* is false and $P \rightarrow A$ is true. Hence $\neg(P \rightarrow A)$ is false. But then for *S* to be true, $\neg G$ must be false and hence *G* must be true. *I can prove the existence of God simply by not praying!*

Even those who believe in God will find this argument fishy and will look more kindly on other ways, *beside the material conditional*, of interpreting the indicative conditional.²

One suggestion, associated with Ernest Adams and Edgington herself [1, 6], is to treat the conditional probabilistically. Thus asserting *If A then B* is tantamount to saying that the probability of *B* given *A* is high. Someone who says, “If John comes to the party, Mary will come too,” is asserting that $p(M|J)$ is high, perhaps more than .9.

Such probabilistic conditionals received a blow from results of David Lewis [14] who showed that (on pain of triviality) such conditionals cannot be interpreted as *propositions*. In other words, there cannot be a proposition (a set of possible worlds) *C* such that $p(C) = p(B|A)$ for all probability measures *P*.

²This particular variety of conditional is called the *indicative conditional*, to distinguish it from so called subjunctives or counterfactuals.

2 Near-monotonicity: finite setting

But let us leave that worry aside and ask about the *logic* of probabilistic conditionals interpreted not as an implication (i.e., as a connective) but instead as a consequence relation \sim . This avoids the Lewis' problem because we are *not saying* that *If A then B* is a proposition. Let us say that we accept $A \sim B$ if $p(B|A)$ is high, where the \sim represents the indicative conditional interpreted probabilistically.

Arlo-Costa and Parikh [2] looked at probabilistic conditionals in the context of cores,³ a notion investigated by Bas van Fraassen, and they looked at some conditions on non-monotonic relations considered by Dov Gabbay, and by Kraus, Lehmann and Magidor [9, 12].

Various rules of inference apply to such consequence relations. Thus from $A \sim B$ and $B \models C$ we can derive $A \sim C$, where \models represents the classical consequence. This rule is (RW) or right weakening and is sound. So is the rule (AND) which derives $A \sim B \wedge C$ from $A \sim B$ and $A \sim C$ (provided we make some sacrifice in probability⁴). But a rule which does *not* hold is monotonicity (M), or *strengthening the antecedent*. This would be the rule, *from $A \sim B$ and $C \models A$ we should be able to derive $C \sim A$* . In particular we would like to be able to derive $(A \wedge X) \sim B$ from $A \sim B$.

$$(M) \quad \frac{A \sim B}{A \wedge X \sim B}$$

Alas, the rule (M) is known not to be sound. The probability of B given A may be high, and the probability of B given A and X may be low. For instance, if our domain is integers upto 100, then the probability that n is odd given that it is prime is quite high. But the probability that it is odd given that it is prime *and* less than 4 is only .5.

A well known example involves birds. Given that Tweety is a bird there is a high probability that it flies. But given that Tweety is a bird and a penguin, the probability drops to 0.

What we show below is that the rule (M) is *mostly* sound. That is that provided that A is large enough, for *most* X , the conclusion continues to hold.

Probabilistic conditionals are *mostly* monotonic. And this is good news, for clearly, while accepting that the monotonicity condition does not hold *universally*, we do want it to hold *usually*.

For consider birds. If the dictum, "Birds fly," could be destroyed at the drop of a hat, it would be useless. We could not conclude that female birds fly, that blue birds fly, or that the bird sitting on the window sill is likely to fly. It is almost always the case that when we know that some creature is a bird, we also have *some* additional information X . And

³[2] interpret "high" as 1.

⁴If $p(B|A) > .95$ and $p(C|A) > .95$, then $p(B \wedge C|A) > .9$

usually, we do not drop the dictum “Birds fly,” when we have some additional information. Thus it must be the case that the dictum is reasonably sturdy. Information like, “It is a penguin,” is *unusual*. It is this sturdiness that we will prove below.

2.1 A concrete example

We are representing propositions as sets of possible worlds. One of us has already objected to this identification [15]. But the representation is commonly accepted and a result which uses it ought to have a relevance. The following theorem is stated rather loosely, but will be followed up by a more precisely stated theorem.

Theorem 2.1 *Suppose that W is a finite space, with all points equally likely. Suppose that A, B are sufficiently large subsets of W and $p(B|A) \geq .95$. Then for most randomly selected $X \subseteq W$, $p(B \cap X|A \cap X) = p(B|A \cap X) > .948$.*

Of course the set theoretic operation \cap corresponds to the logical operation \wedge . In terms of \sim it means that if you know $A \sim B$ and want to know if $A \wedge X \sim B$, the answer will be, “Most likely.”

We have used the numbers .95, .948, and the uniform distribution on W for convenience, but of course the result holds more generally, as will be evident from the proof. What “sufficiently large” B means will be made more explicit below. The technique of proof requires the central limit theorem and the binomial distribution’s Gaussian approximation. It is quite accessible.

Proof: We start by noting some simplifications. Since we are taking probabilities relative to A or its subsets, the points in W which are not in A play no role. So we shall assume that $A = W$. This automatically implies that $B \subseteq A$, an assumption which we could have justified independently.

We require the sets to be large, so assume for concreteness that the set $A - B$ has cardinality 10,000 and B has cardinality 190,000, in which case $p(B|A) = .95$. A random subset X of A has two parts: X_B which is simply $X \cap B$, and the remaining part X_R of X , which is $X \cap (A - B)$.

The expected size of X_B is 95,000 (half of 190,000) but it could be more or less. But by the central limit theorem [3], the standard deviation σ of the size of X_B is $.5 \times \sqrt{190,000}$, which is approximately 217.9. Thus $95,000 - 3\sigma$ is more than 94,346. It is unlikely that the actual value differs from the expected value by more than 3σ . Indeed, using standard tables, the probability that X_B has size more than 94,346 exceeds .9987. Similarly, the set X_R has expected size 5000, but the standard deviation σ' is 50. Thus with the same probability .9987, X_R has size less than 5150. Thus with probability greater than .9974, the ratio $|X_B|/|X_B \cup X_R|$ is greater than $94,346/(94,346+5,150)$ which is .9482, or very nearly .95. (The figure .9974 comes from the fact that even if both errors of .0013 (1 - .9987) were

to add up, we would still only get an error of .0026)

This means that if the sets A, B are both large, and $p(B|A) > .95$, then (when a random subset X of A is chosen) with probability greater than .9974, $p(B \cap X|A \cap X) > .9482$. \square

We can show that similar results will hold if the random set X is chosen in some other way, e.g. if we toss a die for each point of A and put a point in X only if the die shows a 6.

One could ask if the rule (M) can be called sound if it holds only for *most* X . However, note that $A \sim B$ does not say that if A is true, then B is also true 100% of the time. It only says that if A holds, then B is very likely to hold. If the rule applies 99.74% of the time, and the premise (which we accepted) only “applies” 95% of the time then it is hard to justify the premise while rejecting the rule.

2.2 The limiting case

We now state a more precise result, which actually generalizes the observation above to the case where the probability β of B relative to A is positive, but not necessarily close to 1.⁵

The intuitive idea is that we can think of the set X as a random *sample* from the space W , in which case $X \cap A$ will be a random sample from A . The expected size of $X \cap A$ is half the cardinality $|A|$ of A and its standard deviation is $.5 \times \sqrt{|A|}$. The same holds for the expected size of $X \cap B$ except for the multiplier β . Now if the actual sizes of the two sets were the *same as* their expected sizes, then we would get $p(B \cap X|A \cap X)$ to be *equal* to $p(B|A)$. Of course we cannot expect to be so lucky, for actual size can deviate from the expected size. However, as the sizes of A and B go up, the deviation matters less and less, and so the difference between $p(B \cap X|A \cap X)$ and $p(B|A)$ will tend to zero. This gives us our second result.

Theorem 2.2 *Let $\beta > 0$ be fixed, and let sets A_n, B_n increase monotonically in size with $B_n \subseteq A_n$ and $\lim_{n \rightarrow \infty} p(B_n|A_n) = \beta$. Let X_n be randomly chosen subsets of A_n and $\epsilon > 0$. Then we have*

$$\lim_{n \rightarrow \infty} Pr\left[\left| \frac{|B_n \cap X_n|}{|A_n \cap X_n|} - \beta \right| > \epsilon \right] = 0.$$

In other words, if our prior probability of B given A was $\beta > 0$, and we received additional information X , then provided A was large, we should expect the posterior probability of B given A to still be close to β , and the probability that it differs by more than ϵ goes to 0 as $|A|$ goes to infinity.

In the sections below, we will look at the case where A is not merely large but is actually infinite. Investigating that case will involve an excursion into measure theory. It will turn out that the theorem above is (roughly) a corollary of the result we prove below.

⁵We use β for this probability – a number, as we use $p(\cdot)$ for the probability *function*. We use $Pr[\cdot]$ to indicate probability for a meta-statement.

Another rule,

$$\frac{A \sim B}{\neg B \sim \neg A}$$

is not capable of a similar treatment. If our universe consists solely of the innumerable pigeons and the relatively few penguins, then “Most birds fly” will be true, but “Most non-flyers are non-birds” will be false. Indeed all non-flyers will be birds in our universe!

3 Density and Measure

Definition 3.1 *Let n be the set of natural numbers $\{0, 1, 2, \dots, n - 1\}$, in context, and for a given set X of natural numbers, let $X \cap n$ be the members of X between 0 and $n - 1$. Then let the asymptotic density of set X be defined as the limit of its relative frequency: $d(X) = \lim_{n \rightarrow \infty} \frac{1}{n} |X \cap n|$.*

Note that asymptotic density is not defined for all sets of natural numbers, since the limit may not exist, and in fact the set of subsets on which it is defined is not closed under intersection. The asymptotic density *can* be extended to obtain a true measure, using an ultrafilter. Nonetheless, we will make do with asymptotic density in this work.

A second issue to address as we move to the infinite case is what it means to choose at random a subset of an infinite set. Clearly the probability of selecting any fixed set must be 0. For concreteness, let a random subset S of N be shorthand for a set such that each S_n is a uniformly random subset of N_n for each n .

We will initially consider $p(B|A)$ when $A = N$, and then generalize to smaller sets A .

4 Near-monotonicity: infinite setting

Let $B \subseteq N$ have well defined asymptotic density, with $d(B) = \beta$. That is, we assume:

$$\lim_{n \rightarrow \infty} \frac{|B \cap n|}{n} = \beta \tag{1}$$

Let $X \subseteq N$ be chosen uniformly at random. We will now argue that we *almost always* have $p(X) = 1/2$, $p(X \wedge B) = \beta/2$, and $p(X \wedge B|X) = \beta$. Each of these claims will state that a certain property holds with probability 1, over random choices of subset X . Equivalently, the set of such choices X has measure 1. More formally, we have:

Fact 4.1 (Strong Law of Large Numbers) *Let Y_1, Y_2, \dots be a sequence of independently and identically distributed (IID) random variables with $E[Y_i] = m$, and let $S_n = \sum_{i=1}^n Y_i$. then we have $\lim_{n \rightarrow \infty} S_n/n = m$ with probability 1.*

Fact 4.2 Let X be chosen uniformly at random from N , i.e., let X_i be 0 or 1 with equal probability, for each $i > 0$. Then $Pr_X(d(X) = 1/2) = 1$.

Proof: This is simply a restatement of Borel's Normal Number Theorem [3], i.e., that with probability 1:

$$\lim_{n \rightarrow \infty} \frac{|X \cap n|}{n} = 1/2 \quad (2)$$

This also follows from the Strong Law of Large Numbers. \square

Lemma 4.3 With probability 1, we have $\lim_{n \rightarrow \infty} \frac{|X \cap B \cap n|}{|B \cap n|} = 1/2$. (Intuitively, this means $p(X|B) = 1/2$ with probability 1.)

Proof: Let $f(i)$ be the index of the i th element of set B . B will be infinite with probability 1, so assume that this is so. Let $Y_i = X_{f(i)}$. Then the random variables Y_1, Y_2, \dots are IID with $E[Y_i] = 1/2$, so by the Strong Law of Large Numbers, we have with probability 1 that $\lim_{n \rightarrow \infty} S_n/n = 1/2$. Now, what are S_n and n ?

First, by definition, $S_n = \sum_{i=1}^n X_{f(i)}$ = the number of elements among the first n members of B which are also in X . $f(i)$ is some index number, say \hat{n} . Then S_n is the number of numbers from 1 to \hat{n} that are in both B and X , i.e., $S_n = |B \cap X \cap \hat{n}|$. Second, n is the number of elements we are considering from B , i.e., the number of elements of B in the range of 1 to \hat{n} , which is $|B \cap \hat{n}|$. So we have with probability 1 that:

$$1/2 = \lim_{n \rightarrow \infty} S_n/n = \lim_{\hat{n} \rightarrow \infty} \frac{|B \cap X \cap \hat{n}|}{|B \cap \hat{n}|} \quad (3)$$

Since both n and \hat{n} go to ∞ , we can remove all the hats in the final expression. \square

We can now state our first result:

Theorem 4.4 If B has asymptotic density $d(B) = \beta$, then $\frac{d(B \wedge X)}{d(X)} = \beta$ with probability 1.

Proof: Since Equations 2 and 3 each hold with probability 1, their conjunction does as well. We assume both hold for the remainder of the proof. Equation 1 holds by assumption. Let $Eq(i)$ indicate the value of the quantities in Equation i . Then we have:

$$\begin{aligned} Eq(3) \cdot Eq(1)/Eq(2) &= \left(\lim_{n \rightarrow \infty} |B \cap X \cap n|/|B \cap n| \right) \cdot \frac{(\lim_{n \rightarrow \infty} |B \cap n|/n)}{(\lim_{n \rightarrow \infty} |X \cap n|/n)} \\ &= \lim_{n \rightarrow \infty} \frac{|B \cap X \cap n|}{|B \cap n|} \cdot \frac{|B \cap n|/n}{|X \cap n|/n} = \lim_{n \rightarrow \infty} \frac{|B \cap X \cap n|}{|X \cap n|} = \frac{d(B \wedge x)}{d(X)} \end{aligned}$$

But we also have:

$$Eq(3) \cdot Eq(1)/Eq(2) = 1/2 \cdot \beta/(1/2) = \beta$$

which yields the result. \square

Corollary 4.5 *If B has well defined asymptotic density relative to A , i.e., $\lim_{n \rightarrow \infty} \frac{|A \cap B \cap n|}{|A \cap n|} = \beta$, then $\lim_{n \rightarrow \infty} \frac{|A \cap B \cap X \cap n|}{|A \cap X \cap n|} = \beta$ with probability 1.*

Proof: We can again use the trick of renaming indices. First, *zoom in* and rewrite $\lim_{n \rightarrow \infty} |A \cap B \cap n|/|A \cap n|$ as $\lim_{n \rightarrow \infty} |B_A \cap \hat{n}|/|\hat{n}|$, where $B_A = A \cap B$ and $\hat{n} = |A \cap n|$. We can now apply the theorem to obtain $\lim_{\hat{n} \rightarrow \infty} |B_A \cap X \cap \hat{n}|/|X \cap \hat{n}| = \beta$ with probability 1. Finally, we *zoom back out*, and rewrite the expression for β as $\lim_{n \rightarrow \infty} |A \cap B \cap X \cap n|/|A \cap X \cap n|$. \square

Corollary 4.6 *For almost all B , we have $\Pr[\frac{d(X \wedge B)}{d(X)} = d(B)] = 1$ (respectively $\Pr[\frac{d(X \wedge B)}{d(X \wedge A)} = d(B|A)] = 1$). Intuitively, this means that $p(B|X) = d(B)$ (respectively $p(B|A \wedge X) = d(B|A)$) with probability 1.*

Proof: We assumed that set B has well defined asymptotic density to obtain the theorem. Note, however, that *almost all* subsets of N do have asymptotic density. This is clear from Fact 4.2: since the density is $1/2$ with probability 1, the density in particular *exists* with probability 1. \square

Corollary 4.7 (Theorem 2.2) *Let $\beta > 0$ be fixed and let sets A_n, B_n increase monotonically in size with $B_n \subseteq A_n$ and $\lim_{n \rightarrow \infty} p(B_n|A_n) = \beta$. Let X_n be randomly chosen subsets of A_n and $\epsilon > 0$. Then we have $\lim_{n \rightarrow \infty} \Pr[|\frac{|B_n \cap X_n|}{|A_n \cap X_n|} - \beta| > \epsilon] = 0$.*

Proof: Without loss of generality, assume that $A_n \subseteq A_{n+1}, B_n \subseteq B_{n+1}, X_n \subseteq X_{n+1}$ for all n . (Rename the elements of each as needed.) We can then interpret n as a bound, and the sets A_n, B_n, X_n as finite prefixes of infinite sets A, B, X_n , i.e., $A_n = A \cap n, B_n = B \cap n, X_n = X \cap n$. The assumption $\lim_{n \rightarrow \infty} p(B_n|A_n) = \beta$ means that B has well defined asymptotic density relative to A , i.e., $\lim_{n \rightarrow \infty} |A \cap B \cap n|/|A \cap n| = \beta$. Therefore Corollary 4.5 gives us that $\lim_{n \rightarrow \infty} |A \cap B \cap X \cap n|/|A \cap X \cap n| = \beta$ with probability 1, which is actually *stronger* than what we require: $\lim_{n \rightarrow \infty} \Pr[||B_n \cap X_n|/|A_n \cap X_n| - \beta| > \epsilon] = 0$. \square

5 Open Questions

- We proved the result under the assumption that B has asymptotic density, which is almost always true. The asymptotic density can be extended, using an ultrafilter, to obtain a true measure function. Can a similar result be proven for such a measure? Should it?
- Can the result be extended to an uncountable universe?
- Can the result be formulated so as to apply to a non-uniform distribution for selecting members of X , such as $1/n^2$?

- We showed that for almost all propositions B , the result holds. Implicitly this is assuming a uniform distribution on the choice of B . What about other distributions? Can we show that the result holds for *almost all distributions* on the choice of B ?

Acknowledgement: We thank Evan Goris, Dexter Kozen, Roman Kuznets, Sergiu Hart, Ou Liu, Bud Mishra and Samer Salame for comments.

References

- [1] Ernest Adams, “Probability and the logic of conditionals,” in *Aspects of Inductive Logic*, edited by Suppes and Hintikka, (1968), North Holland, 265-316.
- [2] Horacio Arlo-Costa and Rohit Parikh “Conditional probability and defeasible inference,” *Journal of Philosophical Logic*, **34** (2005) 97-119.
- [3] Patrick Billingsley, *Probability and Measure, Third Edition*, Wiley, 1995.
- [4] Darren Bradley and Hannes Leitgeb, “When betting odds and credences come apart: more worries for the Dutch book arguments,” *Analysis*, **66.2**, 2006, 119-127.
- [5] R. Creighton Buck, “The measure theoretic approach to density”, *American Journal of Mathematics*, Vol. 68, No. 4 (Oct., 1946), pp. 560-580.
- [6] Dorothy Edgington, “On Conditionals,” *Mind*, **104**, 1995, 235-329.
- [7] Adam Elga, “Self-locating belief and the Sleeping Beauty problem,” *Analysis*, **60**, (2000) 143-147.
- [8] Bruno de Finetti, “Foresight: its logical laws, its subjective sources,” in *Studies in Subjective Probability*, ed. Kyburg and Smokler, Krieger (1980) pp. 53-118 (translation by Kyburg, original French version, 1937).
- [9] Dov Gabbay, “Theoretical foundations for nonmonotonic reasoning in expert systems,” in *Proceedings NATO Advanced Study Institute on Logics and Models of Concurrent Systems*, (ed.) K.R. Apt, (1985), 439-457.
- [10] Joseph Halpern, “Sleeping Beauty reconsidered,” *Oxford Studies in Epistemology*, edited by Gendler and Hawthorne, (2005), pp. 111-142.
- [11] James Hawthorne, “On the logic of nonmonotonic conditionals and conditional probabilities,” *Journal of Philosophical Logic*, **25** (1996) 185-218.
- [12] S. Kraus, D. Lehmann, and M. Magidor, “Nonmonotonic reasoning, preferential models and cumulative logics,” *Artificial Intelligence*, 44 (1990), 167-207.

- [13] Isaac Levi, "On indeterminate probabilities," *Journal of Philosophy*, 71 (1974), 391-418.
- [14] David Lewis, "Probabilities of conditionals and conditional probabilities," *Philosophical Review*, 85 (1976) 297-315.
- [15] R. Parikh, "Logical omniscience and common knowledge: WHAT do we know and what do WE know?," *Proceedings of the Tenth Conference on Theoretical Aspects of Rationality and Knowledge - 2005*, ed. Ron Meyden, National U. Singapore Press, pp. 62-78.
- [16] Rohit Parikh, Some puzzles about probability and probabilistic conditionals, *Symposium on Logical Foundations of Computer Science*, 2007, 449-456.
- [17] R. Parikh, review of [20], *Essays in Philosophy*, **7,1**, (2006).
- [18] M. Piccione and A. Rubinstein, "On the interpretation of decision problems with imperfect recall," *Games and Economic Behavior*, **20** (1997) 3-24.
- [19] F. P. Ramsey, "Truth and probability," (1926), reprinted in *F.P. Ramsey, Philosophical Papers*, edited by D.H.Mellor, Cambridge University Press 1990.
- [20] David Sanford, *If P then Q*, second edition, Routledge 2003.
- [21] Stanford Encyclopedia of Philosophy.
- [22] Paul Weirich, "The St. Petersburg gamble and risk," *Theory and Decision* 17, (1984) 193-202.