INFINITE LOTTERIES, PERFECTLY THIN DARTS AND INFINITESIMALS

ALEXANDER R. PRUSS

Abstract. One of the problems that Bayesian regularity, the thesis that all contingent propositions should be given probabilities strictly between zero and one, faces is the possibility of random processes that randomly and uniformly choose a number between zero and one. According to classical probability theory, the probability that such a process picks a particular number in the range is zero, but of course any number in the range can indeed be picked. There is a solution to this particular problem on the books: a measure that assigns the same *infinitesimal* probability to each number between zero and one. I will show that such a measure, while mathematically interesting, is pathological for use in confirmation theory, for the same reason that a measure that assigns an infinitesimal probability to each possible outcome in a countably infinite lottery is pathological. The pathology is that one can force someone to assign a probability within an infinitesimal of one to an unlikely event.

1. Introduction

Bayesian regularity, the thesis that all contingent propositions should be given probabilities strictly between zero and one, is difficult to sustain. Classical probability theory assigns the extreme probabilities zero or one to a number of contingent propositions. Some of these contingent propositions are even quite important to Bayesians. Thus, famously Savage's convergence theorem (Savage, 1954, 52–54) shows that under some assumptions the probability that posterior probabilities for what is in fact a truth will converge to certainty is one. Yet the proposition that the posterior probabilities will thus converge is contingent.

The best bet for Bayesian regularists has been to hope that an extension of the probability calculus that allows for infinitesimal probabilities could save regularity. Unfortunately, recent results give little comfort to such hope. Williamson (2007) has shown that the probability that a fair coin will come up heads on each of infinitely many throws must be taken to be zero and not just infinitesimal (though see Weintraub, 2008, for a riposte), and Pruss (manuscript) proved that no matter how much one expands the range of values of the probability calculus, if the sample space is large enough regularity will fail. There is, however, one area where the

Forthcoming in *Thought*.

infinitesimal strategy seems to have had an early success, and this is in Bernstein and Wattenberg's (1969) demonstration that you can produce a finitely additive measure that assigns the same non-zero infinitesimal probability to each point of the interval $[0, 1] = \{x : 0 \leq x \leq 1\}$. A two-dimensional version of the Bernstein and Wattenberg measure could, for instance, provide a regular finitely additive probability measure for the outcomes of the throw of a dart with a perfectly defined (either perfectly sharp or perfectly symmetric) tip.

This paper removes the hope for the infinitesimalist strategy that Bernstein and Wattenberg gave it. This will be done by adapting an example of Dubins (1975) to show that the Bernstein and Wattenberg measure, and any measure like it, is pathological. But first I will show that any measure that assigns an infinitesimal number to every outcome of a countably infinite lottery will be pathological. The pathology in question should be particularly galling to Bayesians: the measure allows for setups where there is a falsehood p, to which one antecedently assigned a low probability, and which is such that every outcome of an experiment makes p come out to have a posterior probability only infinitesimally less than certain. This is, of course, a severe case of failure of conglomerability, where conglomerability is the property of a probability distribution P whereby $P(A) < x$ but $P(A|T) > x$ for every member T of a partition of the probability space.[1] I will then show how a similar pathology makes pathological any measure that assigns an equal infinitesimal probability to every number in $[0, 1]$.

Except when I am explicitly considering classical Lebesgue measure, I will only assume that the probability measures are finitely additive. Defining countable additivity for measures that take infinitesimal values would be conceptually problematic.

2. Infinite lotteries

2.1. **Pathological confirmation.** Suppose it is possible to actually have a lottery that assigns (positive) infinitesimal probabilities (not necessarily the same ones) to all positive integers n (for one mathematical construction, see Benci, Horsten and Wenmackers, 2011). A positive infinitesimal is an ι such that for every real number $x > 0$ we have $0 < \iota < x$. This is equivalent to the condition that for every (finite) natural number n we have $0 < \iota \leq 1/n$.

Given such a lottery, we can then set up the following system by adapting an example of Dubins (1975; see also Kadane, Schervish and Seidenfeld, 1996) to the case of infinitesimals. First I throw a fair die without showing you its outcome. If the die lands showing a number other than six, I independently choose a positive integer according to our lottery. Call this

[1] For an excellent and highly accessible discussion of conglomerability, see Arntzenius, Elga and Hawthorne (2004).

process for choosing the positive integer "Process A". If the die shows six, I independently choose a random positive integer in such a way that the probability of choosing n is equal to 2^{-n}. Call this "Process B". I then announce to you the integer I picked, but not the process I picked it by. Suppose you know with certainty (i.e., probability one) that I am following this procedure and what number I announced.

Say the number I announced is n. Let H be the hypothesis that the die comes up with six. The prior probability is $P(H) = 1/6$. Let us calculate your posterior probability. Let X_n be the event of my announcing n. Then $P(X_n|H) = 2^{-n}$. On the other hand, $P(X_n|\sim H)$ is some infinitesimal ι. By Bayes' Theorem,

$$\begin{aligned} P(H|X_n) &= \frac{P(X_n|H)P(H)}{P(X_n|H)P(H) + P(X_n|\sim H)P(\sim H)} \\ &= \frac{2^{-n}\cdot(1/6)}{2^{-n}\cdot(1/6)+\iota\cdot(5/6)} \\ &= \frac{1}{1+2^n\cdot 5\iota}. \end{aligned}$$

Then $0 < 1 - P(H|X_n) = 2^n \cdot 5\iota/(1 + 2^n \cdot 5\iota) \leq 2^n \cdot \iota$, and so $P(H|X_n)$ differs by an infinitesimal from 1 as an infinitesimal multiplied by $2^n \cdot 5$ is still an infinitesimal. And this is true no matter which positive integer I announced as n.

Thus, simply by performing another experiment after the die throw, in a way that depends on the die throw, and honestly telling you the result the experiment produced, I can make you be only infinitesimally short of certainty that the die showed a six, no matter how the die actually landed. Thus, in five out of six experiments, you will end up being all but certain of a falsehood that you initially assigned low probability to.

This provides a way in which an agent you know for sure to be perfectly truthful and omniscient about the relevant domains can mislead you into assigning probability infinitesimally close to 1 to any falsehood that you assign a probability non-infinitesimally greater than 0 to, without relying on any mistaken beliefs on your part. For instance, let us suppose that you assign probability 10^{-12} to phlogiston theory. The omniscient and perfectly truthful agent tells you that if phlogiston theory is false, she will run Process A, and if phlogiston theory is true, she will run Process B. She announces to you the number chosen, and using a very similar Bayes' Theorem calculation to the above, you then assign a probability within an infinitesimal of 1 to phlogiston theory.

2.2. **Error.** The die and lottery example assumed that there was no perceptual error and that you know for sure the procedure that I was following. This assumption simplifies the calculations, but for greater realism, we should take into account the possibility that small chances of error might infinitely overshadow an infinitesimal.

There are two potential sources of error. One is my misreading the outcome of the die throw and triggering the wrong process, and the second is your getting wrong the outcome of the process, due to an error on my or on your part. We can capture the two sources of error as follows. Let A (respectively, B) be the event of Process A and not Process B (respectively, Process B and not Process A) being activated. This time, let X_n be the event that either Process A or Process B produce result n. Let X_n^* be the event that you took it to be the case that Process A or Process B produced n (this can go wrong either through my announcing incorrectly or your misapprehension). Let H be the die throw resulting in a six. There are three ways to have a relevant error: we could have a six but not B (first type of error, when there is a six), we could have a non-six but not A (first type of error, where there is no six), or we could have one of the processes produce result a result that you do not take to have occurred (second type of error). Thus, if E is the event of a relevant error occurring,

$$E = (\sim B \,\&\, H) \vee (\sim A \,\&\sim H) \vee \exists k (X_k^* \,\&\sim X_k). \tag{1}$$

I will assume $P(E) \leq 1/100$. We will also suppose that $P(X_n|A \,\&\sim H)$ is infinitesimal for all n, while $P(X_n|B \,\&\, H) = 2^{-n}$.

Our pathological conclusion in the errorless case was that no matter what you observe, you end up believing H to a degree infinitesimally short of certainty. Our new pathological conclusion will be that no matter what number you take to have been produced by Process A or Process B, you end up believing $H \vee E$ to a degree within an infinitesimal of certainty, assuming you begin with $P(H \vee E)$ being non-infinitesimal and positive. This is pathological since we know that $H \vee E$ has at most a $1/6 + 1/100 \approx 18\%$ chance of being true. Thus most of the time you'll end up infinitesimally short of certainty about a falsehood, and unless you are lucky enough to get other evidence that involves likelihoods that are infinitesimally close to extreme values, you will never get out of this error by Bayesian means.

To prove that $P(H \vee E|X_n^*)$ is within an infinitesimal of 1, we need only show that $P(X_n^*|H \vee E)$ is positive and non-infinitesimal while $P(X_n^*|\sim(H \vee E))$ is at most infinitesimal (i.e., zero or infinitesimal) for every n, in light of the following lemma where we let a be the event $H \vee E$ and b be the event X_n^* and we use the observation that if β is positive and non-infinitesimal and γ is at most infinitesimal, then γ/β is infinitesimal.

Lemma 1. *If $P(a)$ is positive and non-infinitesimal, $P(b|a)$ is positive, and $P(b|\sim a)/P(b|a)$ is at most infinitesimal, then $1 - P(a|b)$ is at most infinitesimal.*

Proof. Let $\alpha = P(a)$, $\beta = P(b|a)$ and $\gamma = P(b|\sim a)$. Then by Bayes, $P(a|b) = \alpha\beta/(\alpha\beta + (1-\alpha)\gamma) = 1 - (1-\alpha)\gamma/(\alpha\beta + (1-\alpha)\gamma)$. But the latter fraction is no greater than $(1-\alpha)\gamma/(\alpha\beta)$ while γ/β is zero or infinitesimal and α is positive and non-infinitesimal, so that $(1-\alpha)\gamma/(\alpha\beta)$ is at most infinitesimal. $\square$

The arguments that $P(X_n^*|H \vee E)$ is positive and non-infinitesimal while $P(X_n^*|\sim(H \vee E))$ is at most infinitesimal are left as an exercise to the reader.

Now, it is pathological enough that you would need to assign a probability infinitesimally close to 1 to $H \vee E$, no matter which X_n^* occurs. But what probability should you assign to H then? This depends on further assumptions. Say that two events, A and B, are *not highly correlated* provided that $P(A \,\&\, B) \leq 10P(A)P(B)$. You could, then, have a case where it is reasonable to assume that E is not highly correlated with X_n^* for any n. We have seen that $P(H \vee E|X_n^*) \geq 1-\iota$ for a positive infinitesimal ι for all n. So

$1-\iota \leq P(H \vee E \mid X_n^*) \leq P(H \mid X_n^*) + P(E \mid X_n^*) \leq P(H \mid X_n^*) + 10P(E)$, due to the lack of high correlation between E and X_n^*. But $P(E) \leq 1/100$, so $P(H \mid X_n^*) \geq 9/10 - \iota$. In other words, no matter what turns out, you should be fairly confident (with credence greater than 0.89) that the die showed a six, and we can increase the confidence level here simply by decreasing the probability $P(E)$ of relevant error.

Using a similar line of reasoning, we can relax the idealizing assumptions in the phlogiston case. If you have an agent capable of running infinite lotteries of the right sort and whom you are sufficiently confident to be sufficiently reliable (and in such a way that there is no high correlation between reports and relevant errors), the agent can convince you, that phlogiston theory is likely to be true, simply by doing her honest best to announce the result of Process A if phlogiston theory is false and of Process B if it's true. Thus the damage done by the non-conglomerable probability distribution in Process A cannot be confined to propositions about lottery outcomes. Once probability distributions with infinitesimal probabilities for all outcomes are permitted, the damage can easily spread to other domains. This makes problematic Arntzenius, Elga and Hawthorne's (2004) recommendation that we simply accept non-conglomerable probability distributions where they come up.

2.3. **An honest con game.** There is another pathological result about infinite lotteries, paralleling the observation (Kadane, Schervish and Seidenfeld, 1996, 1230–1231) that if you have merely finite additivity in a setting without infinitesimals, it can be rational to pay to refuse information. Good's Theorem (Good, 1967) rules this out for classical probabilities. A corollary of our case is that Good's Theorem fails for infinite lotteries with infinitesimal probabilities.

Suppose we play a coin toss guessing game. After I have tossed the coin, I am certainly free to offer you additional information and a chance to change your guess. It seems there is nothing dishonest about this—one can always offer someone with whom one has a contract an option to change the contract—as long as I honestly communicate to you the additional information. But if we can have infinite lotteries with infinitesimal probabilities of outcomes, then I can bilk you out of your money by honestly communicating additional information.

For suppose I offer you a game where you get \$20 if you guess the coin correctly and you pay me \$10 if you guess wrong. Of course, you accept. I toss the coin. If you guessed wrong, 99% of the time I just reveal the coin and collect my \$10. But 1% of the time when you guessed wrong, and always when you guessed right, I do the following. If you guessed right, I independently perform Process A, the infinite lottery on $\{1, 2, ...\}$ with infinitesimal probabilities for each outcome. If you guessed wrong, I independently perform process B, where each positive integer n has probability 2^{-n}. I then announce

(a) all of the above procedures, including the fact that I offer such a chance to change the guess in 100% of the cases where you guessed right and 1% of the cases where you guessed wrong,
(b) the positive integer resulting from my process,

but of course I do not announce whether I performed Process A or B. And then I ask you if you want to change your guess.

For simplicity, suppose you accept everything I say with certainty.

What should you do? When you learn of my procedures, without learning of the positive integer, and you find out that I let you change the guess in 1% of the cases where you guessed wrong and in all the cases where you guessed right, it seems a no-brainer—you should stand pat, since the fact that I am letting you change the guess is strong evidence that your guess was correct. But then I announce the positive integer coming from Process A or B. And now you need to change. Let R be the event of your initial guess being right. Let X_n be the event of my Process A or B resulting in n. Then $P(X_n|R)$ is infinitesimal, since if you were right, I would do the infinite lottery with infinitesimal outcomes, while $P(X_n|\sim R) = 2^{-n}$. Applying Lemma 1, we conclude that your posterior for $\sim R$ will be within an infinitesimal of one, and hence your posterior for R will be at most infinitesimal. Consequently, you should change your guess.

Let us suppose you do what you should do. Then this is a pretty good game for me. In 99% of the cases where you initially guessed wrong, I get \$10. In 1% of the cases where you initially guessed wrong, you will change your guess and get \$20. And in all the cases where you initially guessed right, you will change your guess and I will get \$10. My expected payoff from the game is $(0.5 \cdot 0.99 + 0.5)(\$10) - (0.5 \cdot 0.01)(\$20) = \$9.85$. My risk is low: I win in 99.5% of the games I play, and in the few that I lose, I don't lose much. It's a great con game, and it seems quite honest, since I'm only giving you true information and giving you a chance to change your guess. Isn't this an absurd conclusion?

I also leave it as an exercise to the reader to check that as soon as you hear (a), it becomes rational for you to *pay me* \$29 to shut up and not go on to announce my result. And so if I make my announcement of (a) *before* I engage in the lotteries, I don't even have to bother with running the lotteries—the credible threat of them is good enough to get you to pay me \$29 to stop.

Any probability measure that allows for such things is pathological. Therefore, any probability measure that assigns infinitesimal probabilities in a countably infinite lottery is pathological.

3. Uniform measures on $[0, 1]$

As we saw, any measure that assigns infinitesimal values to the outcomes of a *countably* infinite lottery is pathological. But what about a measure that assigns infinitesimal values to the outcomes of a process with values on a continuous interval like the Bernstein and Wattenberg (BW) measure on $[0, 1]$? Here, I will confine myself to measures like the BW one that assign *the same* infinitesimal, ι, to each real number in $[0, 1]$. I will also assume that the measure assigns a probability to every element of an algebra of sets (a collection of sets closed under complementation and finite union) that contains every countable subset of $[0, 1]$.

It turns out that such a measure is also pathological, if only for the reason that it can be used to generate something like an infinite lottery. The trick will be to partition $[0, 1]$ into a collection (of course uncountable) of countably infinite sets. Then a number X from $[0, 1]$ is picked according to the BW-measure, and it is announced which member, T, of the partition the number falls into. Given this information, the exact value of X in effect provides a countably infinite lottery, with infinitesimal probabilities, over the members of T.

To be more precise, let $\mathcal{T}$ be any partition of $[0, 1]$ into countably infinite sets, such that every member T of $\mathcal{T}$ can be written as $\{T_1, T_2, ...\}$ for distinct elements T_i. (In the Appendix, an explicit construction is given, without using the Axiom of Choice.) For any number $x \in [0, 1]$, there is a unique member T of $\mathcal{T}$, and a unique positive integer number n such that $x = T_n$. Denote this T and n as $T(x)$ and $n(x)$, respectively. Then in our notation $x = (T(x))_{n(x)}$.

Now, suppose that I throw a fair die and independently select the number X in $[0, 1]$ as above. I then inform you of the identity of the set $T(X)$ (perhaps by engaging in a supertask where I tell you all its members, in the order $(T(X))_1, (T(X))_2, ...$). Next, if the die shows something other than six, I let $n = n(X)$, and if it shows six, I independently of X select a positive integer number n where the probability of n equals 2^{-n}. I then announce n.

Let H be the event of the die showing six. Let $E_{T,n}$ be the event of $T(X)$ being the set T and of the chosen number being n. This event is measurable (see the Appendix).

Now, if H happens, then we picked a positive integer n with probability 2^{-n} independently of X. Thus:

$$P(E_{T,n}|H) = 2^{-n}P(T(X) = T)$$

for any $T \in \mathcal{T}$ and positive integer n. Now, observe that the event $T(X) = T$ holds if and only if X is one of $T_1, T_2, ...$ Thus by additivity $P(T(X) = T) \geq P(X = T_1) + ... + P(X = T_k) = k\iota$, for any finite k. Hence

$$P(E_{T,n}|H) \geq 2^{-n}k\iota, \tag{2}$$

for all positive k.

On the other hand, if H does not happen, and if $T \in \mathcal{T}$, then $E_{T,n}$ happens if and only if $T(X) = T$ and $n(X) = n$, i.e., if and only if $X = T_n$. Thus:

$$P(E_{T,n}|\sim H) = P(X = T_n|\sim H) = P(X = T_n) = \iota, \tag{3}$$

where the second equality used independence of X and H.

It follows from (2) and (3) that:

$$\frac{P(E_{T,n}|\sim H)}{P(E_{T,n}|H)} \leq \frac{\iota}{2^{-n} \cdot k\iota} = \frac{2^n}{k}$$

for any positive integer k. The only way this can happen is if the left hand side of the inequality is zero or infinitesimal. By Lemma 1, it follows that $P(H|E_{T,n})$ is within an infinitesimal of one.

So, if it is possible to pick a uniformly distributed number in $[0, 1]$ with infinitesimal probabilities for individual outcomes, then I can rationally force you to have a credence within an infinitesimal of one that the die shows six, simply by performing an experiment and honestly informing you of its result.

Classical Lebesgue measure prevents this pathology. We can still define the events $E_{T,n}$ as above. But Lebesgue measure assigns probability zero to $X = T_n$ and also to $T(X) = T$, the latter having probability zero by countable additivity since it occurs when and only when $X \in \{T_1, T_2, ...\}$. Consequently $P(E_{T,n}|H)$ and $P(E_{T,n}|\sim H)$ are both zero, and Bayes' Theorem cannot be applied in the way we applied it in Lemma 1.

Assuming $T \in \mathcal{T}$, it is very tempting to say that $X \in T$ (which is equivalent to $T(X) = T$) is "infinitely more likely" than $X = T_n$ in the Lebesgue case. This temptation should be resisted, since if we succumb, then we have to say that $E_{T,n}$ is infinitely more likely on H than on $\sim H$, and hence that every result we get in the above setup strongly confirms H over $\sim H$.

4. Conclusions

The one apparent success in using infinitesimals to handle what classical probability theory considers to be zero probability events was the production of a uniform measure on $[0,1]$ that assigns an infinitesimal to each point. Such a measure is mathematically coherent and well-defined. However, this mathematical success is of no help to confirmation theorists, since any such measure will be pathological for very much the same reason that a measure for countably infinite lotteries that assigns infinitesimal probability to each ticket is pathological. Namely, it allows one to have an event whose credence—and, one may add, objective chance—is significantly less than one half and a setup where no matter what results, one will end up within an infinitesimal of certainty that the event occurred.

Appendix: The construction of $\mathcal{T}$ and measurability

First, we need to construct the partition $\mathcal{T}$, and, second, we need to show that the event $E_{T,n}$ is measurable.

To construct the partition, let $T^1 = \{0, 1, 1/2, 1/4, 1/8, ...\}$ and then for any x such that $1/2 < x < 1$, let $T^x = \{x, x/2, x/4, x/8, ...\}$. Then let $\mathcal{T} = \{T^x : 1/2 < x \leq 1\}$. We need to show that the members of $\mathcal{T}$ partition $[0,1]$, i.e., that every member y of $[0,1]$ is in exactly one of the sets T^x. To see this, suppose that $y \in [0,1]$. If $y = 0$, then $y \in T^x$ if and only if $x = 1$. So suppose that $y > 0$. Let n be the largest natural number such that $2^n y \in [0,1]$. Observe that $2^n y > 1/2$, since otherwise $2^{n+1} y$ would also be in $[0,1]$. I now claim that $y \in T^x$ if and only if $x = 2^n y$. If we can prove this, we will have shown that y is in exactly one of the T^x, and our proof that $\mathcal{T}$ partitions $[0,1]$ will be complete. Suppose first that $x = 2^n y$. Then $y = 2^{-n} x$, and since T^x contains each of $x, x/2, x/4, x/8, ...$, it must contain y as well. Conversely, suppose that $y \in T^x$ where $1/2 < x \leq 1$. Since $y \neq 0$, it follows by the construction of T^x that y must be of the form $2^{-m} x$ for some natural number m. Then $x = 2^m y$ and $1/2 < x \leq 1$. Since n is the largest natural number such that $2^n y \in [0,1]$, it follows that $m \leq n$. Suppose $m < n$ so that $2 \leq 2^{n-m}$. Then $1 \leq 2^{n-m} \cdot (1/2) < 2^{n-m} x = 2^n y \leq 1$ and so, absurdly, $1 < 1$. So m cannot be less than n, and so $m = n$. Thus, $x = 2^n y$, and that completes our proof that $\mathcal{T}$ is a partition of $[0,1]$.

It remains to show that $E_{T,n}$ is measurable for $T \in \mathcal{T}$. Observe that $E_{T,n}$ is equivalent to $(X = T_n \,\&\, H) \vee (X \in T \,\&\, N = n)$, where N is an independent random variable representing the result of picking a random number among $1, 2, 3, ...$ with respective probabilities $1/2, 1/4, 1/8, ...$ The event $X = T_n$ is measurable, since we assume that singletons are measurable under our measure on $[0,1]$ and the event $X \in T$ is measurable, since T is a countable set and countable subsets of $[0,1]$ are measurable. It follows that $E_{T,n}$ is measurable.[2]

[2]I would like to thank two anonymous readers for a number of very helpful suggestions.

References

[1] Arntzenius, F., Elga, A. and Hawthorne, J. 2004. "Bayesianism, infinite decisions, and binding", *Mind*, 113, 251–283.

[2] Benci, V., Horsten, L. and Wenmackers, S. 2011. "Non-archimedean probability." Preprint: arXiv:1106.1524v1.

[3] Bernstein, A.R. and Wattenberg, F. 1969. "Non-standard measure theory." In: W. A. J. Luxemberg, ed., *Applications of model theory of algebra, analysis, and probability.* Holt, Rinehart and Winston, New York.

[4] Dubins, L.E. 1975. "Finitely additive conditional probabilities, conglomerability, and disintegration", *Annals of Probability*, 3, 89–99.

[5] Good, I.J. 1967. "On the principle of total evidence", *British Journal for the Philosophy of Science*, 17, 319–321.

[6] Kadane, J.B., Schervish, M.J., and Seidenfeld, T. 1996. "Reasoning to a foregone conclusion", *Journal of the American Statistical Association*, 91, 1228–1235.

[7] Pruss, A.R. Manuscript. "Probability, regularity and cardinality."

[8] Savage, L.J. 1954. *The foundations of statistics*, 2nd edn. Wiley, New York.

[9] Williamson, T. 2007. "How probable is an infinite sequence of heads?", *Analysis*, 67, pp. 173–180.

[10] Weintraub, R. 2008. "How probable is an infinite sequence of heads? A reply to Williamson", *Analysis*, 68, pp. 247-250.