Betting on Doom

Ronald L. Rivest
MIT Institute Professor
rivest@mit.edu

December 9, 2023

Abstract

We examine bets on "Doom", where both bettors die if Doom should happen. Can two parties bet rationally in this case? We argue they can.

1 Introduction

Probability theory began with the study of gambling [5]. We extend the study here by considering situations where the outcome bet upon may destroy both bettors (if "Doom" happens), but where both parties survive if Doom doesn't happen.

We call our two bettors "Alice" and "Bob"[1]

- **Alice**: an optimist, who thinks Doom is unlikely, and
- **Bob**: a pessimist, who thinks Doom is quite likely.

They wish to bet on whether Doom will happen or not.

Of course, if Doom doesn't happen, then Bob can simply pay off Alice for what he owes her, due to the bet. Alice was right, and she won the bet.

But if Doom does happen, neither Alice nor Bob will be around any longer to settle up. How can Alice pay Bob in this case? Bob was right, but how can he obtain value for being right? How can he make a bet?

2 Time

The key is to consider the "time value of money."

To make things work, Doom should not happen immediately, but rather at some point in the future.

Let's suppose that the bet then is that "Doom will not happen for at least t years" (for a given specific t, such as $t = 50$).

[1] Alice and Bob first appeared in the RSA paper [2], they have since appeared many times in other papers [4].

Alice wishes to bet that both she and Bob (or their proxies) will be alive at the end of t years, whereas Bob wishes to bet that Doom will happen within the next t years, killing everyone.

Can they formulate a payoff scheme that makes betting between them work?

3 Odds

The bet wants to have the form:

- If Doom doesn't happen within the next t years, then Bob pays Alice $P dollars at the end of the t years, but
- if Doom happens within the next t years, then Bob collects $Q from Alice at the end of t years.

However, the second possibility doesn't make sense, since neither Bob nor Alice will be around if Doom happens.

(Here P and Q are parameters agreed upon by Bob and Alice, reflecting their judgment of the odds that Doom will happen within the next t years. See Investopedia [1].)

4 Early payment

The trick is to have Alice pay Bob an appropriate amount *at the time the bet is made*, whose value today is equivalent to the value $Q will have in t years.

If Alice loses, then both parties are destroyed, but Bob got his payment early, so he is happy.

If Alice wins, and Doom does *not* happen within the next t years, then she should collect $(P+Q) from Bob at the end of t years:

- Bob repays Alice for the $Q (equivalent) he received when the bet was made, and in addition
- Bob pays Alice $P for winning the bet.

How much should Bob receive from Alice when the bet is made? The amount (call it $Q') in today's dollars should be equivalent to $Q in dollars at time t.

For example, if Alice and Bob agree on a "discount rate" δ (so that one dollar at time t is worth δ^t today), then

$$Q' = Q\delta^t \ .$$

More generally, Alice and Bob can agree on the three parameters P, Q, and Q' when they make the bet. Really, there are only the two parameters:

- Q' — how much Alice pays Bob when the bet is made, and
- $P + Q$ — how much Bob pays Alice at time t if Doom hasn't happened.

5 Discussion

Such bets may be feasible in practice. The flow of value from optimists to pessimists might be used in interesting ways.

Note that the bet doesn't incentivize Bob to make Doom more likely, assuming that Bob would rather live than win a bet and die.

Indeed, Bob may use the payment of $Q' that Alice made to him to *increase the likelihood that Alice wins the bet.*

Part of the negotiation of the parameters (P, Q, Q') of the bet would be considerations by either party as to what they might be able to do with $Q' to affect the odds of Doom happening.

One might consider such a bet as an "estimate of the likelihood of Doom occurring of $Q/(P+Q)$," to tie this all back to the foundations of probability.

This note was motivated by discussions with Martin Hellman on the risk of nuclear war and by the article [3] on the probability $p(doom)$ of AGI causing Doom.

It would not surprise me if this approach has already appeared in the literature, but I haven't seen it.

References

[1] Investopedia. *How do Odds Work in Betting?* URL: https://www.investopedia.com/articles/investing/042115/betting-basics-fractional-decimal-american-moneyline-odds.asp.

[2] R[onald] L. Rivest, A[di] Shamir, and L[eonard M.] Adleman. "A method for obtaining digital signatures and public-key cryptosystems". In: *CACM* 21.2 (Feb. 1978), pp. 120–126. ISSN: 0001-0782. DOI: 10.1145/359340.359342. URL: http://doi.acm.org/10.1145/359340.359342.

[3] Kevin Roose. *Silicon Valley Confronts a Grim New A.I. Metric.* NY Times (December 6, 2023). URL: https://www.nytimes.com/2023/12/06/business/dealbook/silicon-valley-artificial-intelligence.html.

[4] Wikipedia. *Alice and Bob.* URL: https://en.wikipedia.org/wiki/Alice_and_Bob.

[5] Wikipedia. *History of Probability.* URL: https://en.wikipedia.org/wiki/History_of_probability.