Draft: Appeal fees and insurance proposal

September 27, 2018

We analyze fees that should be paid in Kleros between the parties to (binary) disputes. Typically, we will have two parties which we denote by Alice and Bob. Unless indicated otherwise, we consider the situation where Alice has one the most recent ruling of the dispute.

Let x be the value that must be paid in arbitration fees for the next round. A first observation is that each of Alice and Bob must pay at least x in appeal fees if the appeal fees are to be refunded to the winning party. Moreover, any scheme where an insurer pays arbitration fees for an honest parties in which successful insurers are paid out of the fees returned to the winner and (are not paid out of the value at stake in the dispute), the winning party must in fact receive more than what she put in for arbitration fees.

Denote:

- s_A - the additional "stake" Alice must pay in case of an appeal to not lose the case, namely Alice must pay a total of $x + s_A$
- s_B - the "stake" Bob must pay in case of appeal to not lose the case, namely Bob must pay a total of $x + s_B$.

1 Structure of proposed insurance mechanism

Remember that in order to appeal, Bob must pay $x + s_B$ and then Alice must pay $x + s_A$ to not forfeit the appeal. We detail a sort of crowd-sourced insurance mechanism that can cover these fees:

- The party (Alice or Bob) might be given the opportunity to pay these fees directly/themselves; if a party refuses to do so and/or a fixed period of time elapses where she hasn't, then the fees can be "adopted/crowd-funded" as follows:
  - Any user can pay some percentage of the required fees x. Denote $s_r = \frac{\text{contribution of } \mathcal{USR}_r}{x+s_{\text{party}}}$.
  - If less than $x + s_{\text{party}}$ is raised, everyone is refunded and that party's fees are not paid. (In Bob's case the dispute is not appealed; if Bob's fees are paid and Alice's are not, Alice loses the dispute.)

– Once $x + s_{\text{party}}$ is raised, the contract stops accepting additional contributions.
– If the other party ultimately wins the dispute (possibly after additional appeals), the adopters/crowdfunders lose their contributions.
– If the crowdfunded side of the dispute ultimately wins, each contributor $\mathcal{USR}_r$ receives back the contribution she paid towards $x + s_{\text{party}}$ and a corresponding portion of the losing party's stake given by $s_{\text{losing party}} \cdot s_r$.

2 Analysis assuming rational insurers whose total resources match that of any attack

Suppose that rational insurers exist who are willing to fund appeal fees when they estimate

$$E[\text{return on insurance}] > 0.$$

Note that if multiple parties contribute to the fees, the return will be divided proportionally, but this does not change whether this expected value is positive or negative, so without loss of generality we assume that there exists a single insurer with a given prior belief on Alice and Bob's respective winning chances, Isaac.

Remark 1. *In practice, if there are actually multiple insurers who are willing to pay a party's entire arbitration fee and are capable of making this decision quickly after the window to pay the fee has opened (for example, because they have already analyzed the case in advance), then it is possible that the insurers will get into a gas war among each other to have their insurance transaction included. This is not likely to be a worse problem than, for example, multiple parties rushing to challenge an image in the Doge pilot, so we will consider this effect to be negligible for the current work. However, as a sort of instance of the honest unity problem we may study these dynamics in future work.*

Suppose that rational actors (Isaac) evaluate Alice's chances of eventually winning at p_A. Then

$$E[\text{return on insurance}] = p_A(s_B) + (1 - p_A)(-x - s_A) \geq 0 \Leftrightarrow p_A \geq \frac{x + s_A}{x + s_A + s_B}.$$

Denote this threshold by

$$t_A = \frac{x + s_A}{x + s_A + s_B}. \tag{1}$$

Then Isaac will only finance Alice's appeal if he estimates that $p_A \geq t_A$.

Similarly, rational actors would only finance Bob's appeal if

$$p_B(s_A) + (1 - p_B)(-x - s_B) \geq 0 \Leftrightarrow p_B \geq t_B = \frac{x + s_B}{x + s_A + s_B}.$$

Then

$$t_B - t_A = \frac{s_B - s_A}{x + s_A + s_B}. \tag{2}$$

Remark 2. *As Alice and Bob are playing a negative sum game (the payment to the jurors is consuming part of their appeal fees), it is impossible to calibrate the stakes s_A and s_B such that perfectly rational insurers that evaluate Alice and Bob's winning chances in appeal both at 50% will fund the appeals of each. Indeed, we will see that there is some range of estimations of p_A in which is profitable to finance the appeal fees of neither Alice nor Bob. In practice as the evaluations of Alice and Bob's chances will vary in the population of insurers it is possible that they might both have their fees funded.*

We consider the consequences of a few possiblities for s_A and s_B:

2.0.1 Possibility 1: $s_A = s_B$

So note that if $s_A = s_B$, $t_A = t_B$ but

$$t_A = t_B = \frac{x + s_A}{x + 2s_A}.$$

So for example,

- if $s_A = 0$, $t_A = 1$ and insurers are never incentivized to fund appeals,
- if $s_A = x$, $t_A = 2/3$ (in particular if both Alice's and Bob's chances are estimated by Isaac to between 1/3 and 2/3, he is not incentivized to fund either of them), and
- as $s_A \to \infty$ for fixed x, the threshold $t_A = t_B$ tends to 1/2.

In this setting, it may often be the case that an honest party Alice that won the previous round would not have a high enough expected return for insurers to have an incentive to fund her appeal.

2.0.2 Possibility 2: $t_A = 1/2$ (recommended)

Instead, suppose we want $t_A = 1/2$. Then, rearranging Equation 1 we must have

$$s_B = x + s_A.$$

Plugging this into Equation 2, we have

$$t_B = \frac{x}{2x + 2s_A} + 1/2.$$

Then again s_A is a parameter that could be tuned by the governance process. To illustrate several choices:

- if $s_A = 0$, then $s_B = x$ and $t_B = 1$. In this case, insurers are never incentivized to fund appeals for Bob,

- if $s_A = x$, then $s_B = 2x$ and $t_B = 3/4$, (these may be reasonable choices) and
- as $s_A \to \infty$ for fixed x, then s_B also tends to infinity and t_B tends to $1/2$.

So, in the context of taking $t_A = 1/2$ there is a basic tradeoff here between the size of the arbitration fees $x+s_A$ that we are willing to impose on Alice, the party that won the previous round (as well as the even higher fees $x + s_B = 2x + s_A$), versus the threshold probability t_B of an eventual Bob victory that would be required for insurers to be willing to fund his fees. As s_A tends to infinity for fixed x, t_B still tends to $1/2$.

We now address the question of how insurers might estimate p_A and p_B. Consider the related probability π_A - the probability that a randomly selected PNK will correspond to a juror that votes with Alice in an idealized setting where there are no attacks that influence jurors' votes and where this probability does not change from round to round (such as by new information becoming available). Similarly we denote by π_B the probability that a randomly selected PNK will correspond to a juror that votes with Bob under the same conditions.

2.1 Insurers have a perfect knowledge of π_A

If Isaac knows a priori $\pi_A \neq 1/2$ then he knows who would eventually win the dispute with probability arbitrarily close to one if the case was appealed to a sufficiently large juror pool. (Or, more realistically, in the setting where there is a maxAppel set that is very large and on which it is unviable for an attacker Eve to launch attacks to influence the last appeal round(s), Isaac knows who would win that round with high probability.) Hence, if $\pi_A > 1/2$, Isaac would know that Alice would eventually win the dispute as long as she paid whatever arbitration fees are required of her. Moreover, Isaac would know that funding each of these appeals is profitable as long as $s_A > 0$.

With similar reasoning, if Isaac is merely very confident that π_A is even slightly greater than $1/2$, then he will estimate p_A as close to one.

2.2 Insurer's estimate of π_A updates after successive appeal rounds

On the other hand, after each round of voting insurers will have learned information about π_A, so Isaac's estimate of p_A may change over time. Of course, if there is an ongoing attack, jurors' votes may not be representative of the true value of π_A. In the following discussion, we make the assumption that there are no attacks that change the jurors' voting incentives, and that in a given appeal round where there are N jurors, the number of votes that Alice receives is distributed as Binomial(N, π_A). Then Isaac will have a prior for his estimation of the (unknown) value of π_A that he can update after each round. Here we can analyze how this insurance proposal handles bank attacks - where an attacker Eve with a large budget continually appeals until the opposing party does not

fund an appeal. In future work, we think a reasonable model for handling attacks that change voter incentives (such as bribes, $p+\epsilon$ attacks, pre-revelation attacks, etc) and incorporate them into this analysis will be for Isaac to not consider any votes seen while such an attack is active. In practice, insurers may be likely to increase their estimation of π_A, concluding that Alice likely has a winning argument, if they see that Alice is being attacked.

Suppose that the insurer's prior distribution for π_A is given by $\text{Beta}(a_0, b_0)$. So his expected value for π_A is $\frac{a}{a+b}$ and his certainty in this believe scales with the size of $a+b$. Consequently, if $a+b$ is very large, Isaac's estimate of π_A will change little with new evidence.

In the following results, we drop the assumption that Alice won the previous dispute and the results apply equally to Bob.

Proposition 1. *If $\pi_A > 1/2$, then the probability that Alice loses the ith appeal round is at most*

$$\frac{\pi_A}{2(2^{i+2}-1)\left(\pi_A - \frac{1}{2}\right)^2}.$$

Proof. Note that there are $2^{i+2}-1$ total votes in the ith appeal round.

By [1][p.151], if X is distributed as $\text{Binomial}(N, \pi)$, then if $r \leq N\pi$ then

$$\text{Prob}(X \leq r) \leq \frac{(N-r)\pi}{(N\pi - r)^2}.$$

Denote by X the number of votes for Alice in the ith round. As we assume that $\pi_A > 1/2$, in our case this inequality translates into

$$\begin{aligned}\text{Prob}\left(X \leq \frac{2^{i+2}-1}{2}\right) &\leq \frac{\left(2^{i+2}-1-\frac{2^{i+2}-1}{2}\right)\pi_A}{\left((2^{i+2}-1)\pi_A - \frac{2^{i+2}-1}{2}\right)^2} \\ &= \frac{\pi_A}{2\left[2^{i+2}-1\right]\left(\pi_A - \frac{1}{2}\right)^2}.\end{aligned}$$

□

Proposition 2. *Suppose that $\pi_A > 1/2$ and that Alice's required arbitration fees are always paid. Then the total probability that Alice wins every arbitration round from the $R+1$st round onward is lower bounded by*

$$\left(\max\left\{1 - \frac{\pi_A}{2^{R+3}\left(\pi_A - \frac{1}{2}\right)^2}, 0\right\}\right)^2$$

Proof. Suppose that

$$R \geq \log_2\left(\frac{\pi_A}{2^3\left(\pi_A - \frac{1}{2}\right)^2}\right).$$

Indeed, if this is not the case then the statement holds trivially.

Note that the jurors drawn from one appeal round to the next are drawn independently from the fixed distribution Binomial $\left(2^{i+2}-1, \pi_A\right)$; hence the results of these rounds are independent.

Then

$$\begin{aligned} &P(\text{Alice wins every round from } R+1 \text{ onward}) \\ &\geq \prod_{i=R+1}^{\infty} P\left(\begin{array}{c}\text{proportion of votes in} \\ \text{ith round} > 1/2\end{array}\right) \\ &\geq \prod_{i=R+1}^{\infty}\left[1-\frac{\pi_A}{2[2^{(i+2)}-1]\left(\pi_A-\frac{1}{2}\right)^2}\right], \end{aligned}$$

where due to our assumption on R all of the terms in the product are positive and we are applying the bound from Proposition 1.

However, note that this product converges (to a non-zero value); in fact in general

$$\prod_{i=N}^{\infty}\left(1-\frac{c}{2^i}\right) \geq \left(1-\frac{c}{2^N}\right)^2,$$

when $N \geq \log_2(c)$.

To see this note that

$$\log_2\left(\prod_{i=N}^{\infty}\left(1-\frac{c}{2^i}\right)\right) = \sum_{i=N}^{\infty}\log_2\left(1-\frac{c}{2^i}\right) \geq \sum_{i=N}^{\infty} 2^{N-i}\log_2(1-c2^{-N})$$

where we have used the fact that $2^i \geq 2^N$ for all the i we consider so

$$2^{-N}\log_2(1-c2^{-i}) \geq 2^{-i}\log_2(1-c2^{-N}).$$

Then

$$\log_2\left(\prod_{i=N}^{\infty}\left(1-\frac{c}{2^i}\right)\right) \geq 2\log_2(1-c2^{-N})$$

establishing the claim.

Substituting $c = \frac{\pi_A}{2^2\left(\pi_A-\frac{1}{2}\right)^2}$ we see that

$$\prod_{i=R+1}^{\infty}\left[1-\frac{\pi_A}{2[2^{(i+2)}-1]\left(\pi_A-\frac{1}{2}\right)^2}\right] \geq \left(1-\frac{\pi_A}{2^{R+3}\left(\pi_A-\frac{1}{2}\right)^2}\right)^2.$$

□

Proposition 3. *Suppose that s_A and s_B are chosen so that the winner of the previous round is calibrated to have a threshold of 1/2. Suppose, right after*

the Rth appeal, the insurer a priori belief for the value of π_A is distributed as Beta(a, b), where $a > b$. Then the insurer will estimate

$$p_A \geq \int_{\frac{1}{2}}^{1} \left(\max \left\{ 1 - \frac{\pi_A}{2^{R+3} \left(\pi_A - \frac{1}{2} \right)^2}, 0 \right\} \right)^2 \cdot \frac{\pi_A^{a-1}(1-\pi_A)^{b-1}}{B(a,b)} d\pi_A.$$

Consequently, the insurer will be willing to pay Alice's arbitration fees if

$$\int_{\frac{1}{2}}^{1} \left(\max \left\{ 1 - \frac{\pi_A}{2^{R+3} \left(\pi_A - \frac{1}{2} \right)^2}, 0 \right\} \right)^2 \cdot \frac{\pi_A^{a-1}(1-\pi_A)^{b-1}}{B(a,b)} d\pi_A \geq t_A.$$

Here $B(a, b)$ is the beta function on a and b. Note that this bound holds for regardless of Alice won or lost the Rth round, i.e. it holds with the corresponding t_A in each case.

Proof. Isaac is trying to estimate p_A, the probability that Alice eventually wins. Based on his priors, Isaac's estimation that $\pi_A > 1/2$, so that she would win with high probability in the long run in an appeal with a sufficient number of jurors, is

$$\int_{\frac{1}{2}}^{1} \frac{\pi_A^{a-1}(1-\pi_A)^{b-1}}{B(a,b)} d\pi_A.$$

However, Alice can also lose if she has an appeal that is not funded in a subsequent round. Due to the choices of s_A and s_B, Isaac's expected return on a dollar spent on Alice's appeal fees in some later round is the same as his expected return in this round based on his **current** prior about π_A. However, the new information of the juror votes in the ensuing rounds may convince him that some future appeal fee no longer yields a sufficient return.

However, if Alice wins every following round, then as Isaac begins with a prior that $a > b$ so $\pi_A > 1/2$ has a greater than $1/2$ likelihood of being true, being updated with results where Alice receives more votes than Bob in each round, Isaac will always believe that π_A is distributed as Binomial(a', b') where $a' > b'$. Hence it will always be in Isaac's interest to fund an appeal where $t_A = 1/2$. However, as Alice wins all the remaining appeal rounds, t_A will be $1/2$ for all remaining rounds from $R + 3$ onward.

Hence, Isaac's probability that Alice will eventually win is lower bounded by his estimation of the probability that she would win all the remaining appeal rounds **if** her appeal fees are paid. By Proposition 2, this is

$$\int_{\frac{1}{2}}^{1} \left(\max \left\{ 1 - \frac{\pi_A}{2^{R+3} \left(\pi_A - \frac{1}{2} \right)^2}, 0 \right\} \right)^2 \cdot \frac{\pi_A^{a-1}(1-\pi_A)^{b-1}}{B(a,b)} d\pi_A.$$

□

Note that as all of the integrands here are bounded by 1, by the Dominated Convergence Theorem

$$\int_{\frac{1}{2}}^{1} \left(\max \left\{ 1 - \frac{\pi_A}{2^{R+3} \left(\pi_A - \frac{1}{2} \right)^2}, 0 \right\} \right)^2 \cdot \frac{\pi_A^{a-1}(1-\pi_A)^{b-1}}{B(a,b)} d\pi_A$$

$$\longrightarrow \int_{\frac{1}{2}}^{1} \frac{\pi_A^{a-1}(1-\pi_A)^{b-1}}{B(a,b)} d\pi_A$$

as $R \to \infty$. This later value is, in fact, Isaac's estimate for the probability that $\pi_A \geq 1/2$ based on his prior.

Then if Alice knows Isaac's priors, she can estimate how many appeal rounds for which she might need to pay her own appeal fees, namely how large R needs to be, to get to a point where Isaac's estimate for p_A is larger than t_A.

References

[1] Feller. *Introduction to Probability Theory and Its Applications.*