Adversarial Knapsacks

Morad Elsaify, Ben Ewing, Usa Kerdnunvong, Neil Pruthi

December 11, 2018

1 Introduction

Suppose we have a game in which each agent faces a knapsack problem. Each agent chooses items to place in her knapsack, and then goes out into the world. The value of each item is a *random variable*. We might imagine, for instance, an agent choosing *bearspray* for her knapsack, and then going out into a world in which there are many bears, and bearspray is highly coveted. An agent wins the game if the sum of the values of the items in her knapsack is greater than that of any other agent.

1.1 Premise of the Game

1.2 Applications

2 Two Player Variance Independence

First, we consider the simplest specification of this problem. Let there be 2 players and 2 items. Denote the two items X_1 and X_2, and further assume that the items have identical means $\mathbb{E}[X_1] = \mathbb{E}[X_2] = \mu$ and symmetric distributions.

Definition 1 (Symmetric Distributions)**.** Let a probability distribution be symmetric if there exists a value x_0 such that its probability density function $f(\cdot)$ exhibits the following property:

$$f(x_0 - \delta) = f(x_0 + \delta) \ \ \forall \delta \in \mathbb{R}$$

Further, the median and mean of a symmetric distribution are both equal to x_0. All symmetric distributions have no skewness.

Symmetric distributions provide a good starting point for this analysis, as it abstracts away from skewness and include the most common distributions used in academic research (uniform, normal, Student's t, etc.) Focusing on these distributions lets us analyze the importance of variance (and all even moments) on each player's decision processes. The tractability resulting from assuming 2 players gives our first important result.

Theorem 1. *Let X_1 and X_2 be two continuous, independent random variables with symmetric probability density functions $f_1(\cdot)$ and $f_2(\cdot)$ and expectation $\mathbb{E}[X_1] = \mathbb{E}[X_2] = \mu$. Then, $\mathbb{P}(X_1 > X_2) = \mathbb{P}(X_2 > X_1) = 1/2$.*

Proof. WLOG, let $\mu = 0$. Let $f_{1-2}(\cdot)$ denote the probability density function of $X_1 - X_2$. $f_{1-2}(\cdot)$ is given by

$$f_{1-2}(z) = \int_{-\infty}^{\infty} f_1(z+x) f_2(x)\, dz.$$

Now, I assert that $f_{1-2}(\cdot)$ is symmetric about zero. To see this, use symmetry of $f_1(\cdot)$ and $f_2(\cdot)$:

$$f_{1-2}(z) = \int_{-\infty}^{\infty} f_1(-z-x) f_2(-x)\, dz$$

Let $y = -x$. Substituting this, we have

$$f_{1-2}(z) = \int_{-\infty}^{\infty} f_1(-z+y) f_2(y)\, dy.$$

This is simply equal to $f_{1-2}(-z)$. Thus, $f_{1-2}(z) = f_{1-2}(-z)\ \forall z$, so f_{1-2} is symmetric about zero. The mass above and below zero must be equal. As a result, we have:

$$\mathbb{P}(X_1 > X_2) = \int_{0}^{\infty} f_{1-2}(x)\, dx = 1/2$$

and

$$\mathbb{P}(X_2 > X_1) = \int_{-\infty}^{0} f_{1-2}(x)\, dx = 1/2,$$

which proves our result □

Corollary 1. *Let X_1 and X_2 be two discrete, independent random variables with symmetric probability distribution functions $f_1(\cdot)$ and $f_2(\cdot)$ and expectation $\mathbb{E}[X_1] = \mathbb{E}[X_2] = \mu$. Then, $\mathbb{P}(X_1 > X_2) = \mathbb{P}(X_2 > X_1)$.*

Theorem 1 does not perfectly correspond to Corollary 1 due to the possibility of ties in discrete random variables. As a result, while the probability of X_1 exceeding X_2 is equal to the probability of X_2 exceeding X_1, these probabilities do not necessarily equal $1/2$ due to the fact that the probability of ties is not necessarily zero.

Theorem 1 and Corollary 1 provide a powerful result for this baseline setup. The only relevant statistic for examining the setup described above is expected value. With two players, and two items with symmetric distributions, both players always prefer the item with the higher expected value. If both items have the same expected value, then both players are indifferent between the two items.

Lemma 1. *Let $X_1, \ldots, X_n$ be symmetric random variables with means $\mu_1, \ldots, \mu_n$. Then, $X = \sum_{i=1}^{n} X_i$ is symmetric with mean $\mu = \sum_{i=1}^{n} \mu_i$.*

Proof. In the proof of Theorem 1, set $X_2 = -X_i$ for any i. (Note that if X_i is symmetric, then so too is $-X_i$.) The random variable $X_1 - X_2 = X_1 + X_i$ is therefore symmetric and has mean $\mu_1 + \mu_i$. Repeat this process for all $i = 2, \ldots, n$ to obtain the result. □

Lemma 1 can be used to further generalize this result in which there are m items with symmetric distributions but potentially different means. For any allocation of the m items such that the knapsacks of both players have the same mean, higher moments do not affect the probability of winning.

Theorem 2 (Two-Player Variance Independence)**.** *Let $X_1, \ldots, X_m$ be independent random variables with symmetric probability distributions and means $\mu_1, \ldots, \mu_m$. For any two knapsacks given by the set of indices $K_1, K_2 \subseteq \{1, \ldots, m\}$ such that $\sum_{i \in K_1} \mu_i = \sum_{i \in K_2} \mu_i$,*

$$\mathbb{P}\left(\sum_{i \in K_1} X_i > \sum_{i \in K_2} X_i\right) = \mathbb{P}\left(\sum_{i \in K_2} X_i > \sum_{i \in K_1} X_i\right).$$

In other words, any two combinations of symmetric random variables with the same mean have the same probability of exceeding the other.

Proof. Combine Lemma 1 and Theorem 1. □

The theory in this section provides several directions to consider. The first is the potential importance of variance in settings in which there are more than 2 players. This gives rise to a "Mean-Variance Tradeoff," which we discuss in the next section. Second, we can generate useful insights on the importance of skew using asymmetric distributions, which we discuss in Section 4.

3 Mean-Variance Tradeoff

In the previous section, we established that in a two player setting in which items are independent and have symmetric distributions, the only relevant statistic is expected value. Players care only about the expected value of the items, and no other properties of the distribution, including variance. However, the case in which (1) there are more than two players, and (2) items have potentially different means, variance can become an important consideration.

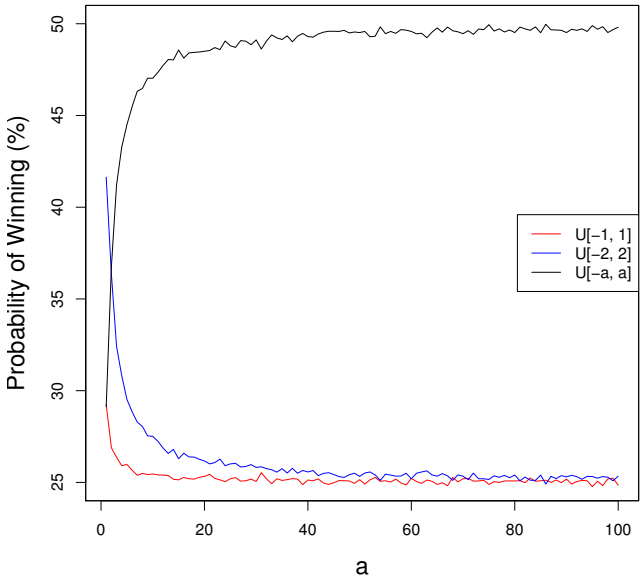
To see this, consider the following illustrative example. There are three items, $X_1 \sim U[-1, 1]$, $X_2 \sim U[-2, 2]$, and $X_3 \sim U[-100, 100]$. It is clear that $\mathbb{E}[X_1] = \mathbb{E}[X_2] = \mathbb{E}[X_3] = 0$ and $\mathbb{V}[X_3] > \mathbb{V}[X_2] > \mathbb{V}[X_1]$. The probability that X_3 is the largest value is bounded below by:

$$\mathbb{P}(X_3 \geq \max\{X_1, X_2\}) \geq \mathbb{P}(X_3 \geq 2) = 0.49.$$

In other words, X_3 wins at least 49% of the time. This is due to the fact that X_3 has a high variance, so it can have realizations that are much higher than are possible for X_1 and X_2. Thus, with more than 2 players, variance becomes an important consideration.

Figure 1 plots the probability of winning for the above example in which we set $X_3 \sim U[-a, a]$, and let $a = 1, \ldots, 100$. For small values of a, X_2 clearly performs the best; however, this performance quickly drops off. When $a \geq 20$, the probability of X_2 winning is almost indistinguishable from the probability of X_1 winning. The importance of variance is clearly evident from this figure.

Figure 1: Affect of Variance in Uniform Distributions



We quantify the importance of variance by asking the question: "How much excess variance does one require to be indifferent between that item and an item with an excess mean of x, and how does this change with the number of players?" To answer this question, we begin with a setup in which there are N players and N items (each player gets exactly 1 item). All items are independent of each other and come from the same distribution family (i.e., all distributions are uniform). $N-2$ of the items are identically distributed with mean μ and variance σ^2. 1 item has mean $\mu + x$ and variance σ^2 (the "high mean item"). The remaining item has mean μ and variance $\sigma^2 + s^2$ (the "high variance item"). We denote $x$the "excess mean," and s^2 the "excess variance."

We estimate, via simulations, a value of s that makes agents indifferent between the high mean item and the high variance item. We can think of s as a function of x and N that maps the excess mean and the number of players to the excess variance. Using this approach, we can characterize the function $s(x, N)$ for various distributions.

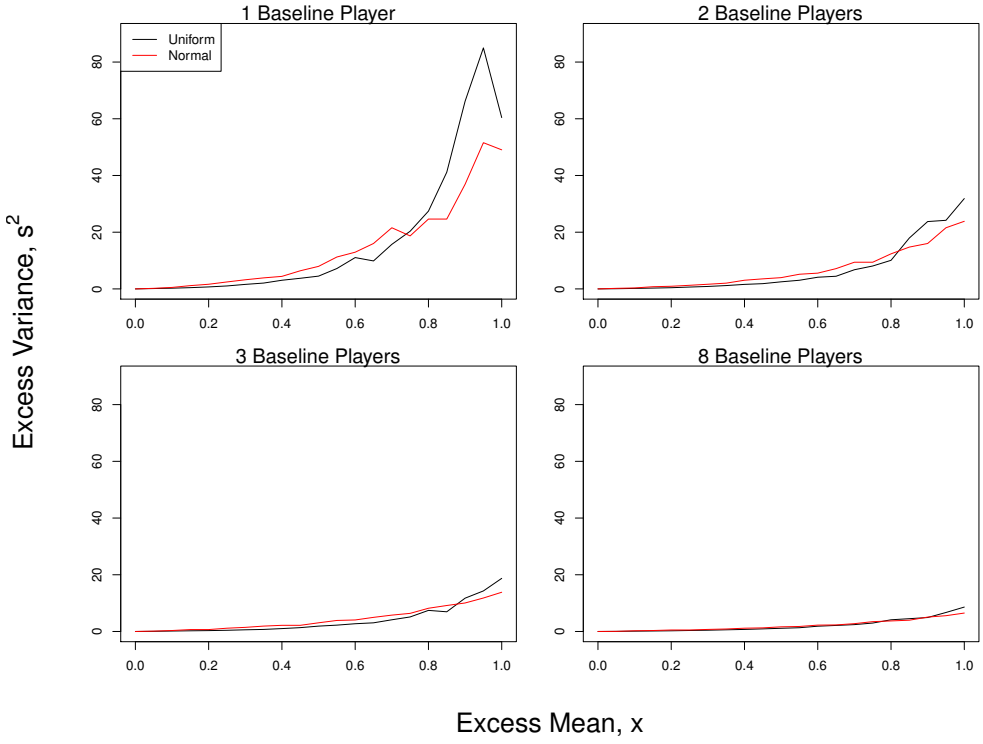
We consider two specifications in this simulation. The first features $N-2$ items $\{X_i\}_{i=1}^{N-2} \overset{iid}{\sim} U[-1, 1]$. The high mean item is given by $X_{N-1} \sim U[-1+x, 1+x]$, and the high variance item is given by $X_N \sim U[-a, a]$.[1] The second specification features $N-2$ items $\{X_i\}_{i=1}^{N-2} \overset{iid}{\sim} N(0, 1)$. The high mean item is given by $X_{N-1} \sim N(x, 1)$, and the high variance item is given by $X_N \sim N(0, 1+s)$. Without loss of generality, these two specifications completely characterize the mean-variance tradeoff for all possible parametrizations of the uniform and normal distributions in which $N-2$ items are iid.

[1]Note that we can also express the distribution of X_N in terms of excess variance, s, as: $X_N \sim U\left[-\sqrt{3s^2+1}, \sqrt{3s^2+1}\right]$. We opt for the simpler specification of $U[-a, a]$, but the specification in terms of s highlights the relationship between a change in the bounds of the uniform distribution and variance.

Figure 2 plots $s(x, N)$ for $N = 3, 4, 5, 10$ and $x \in [0, 1]$[2] using the two specifications considered above. There seems to be an exponential relationship between excess mean and excess variance, although the variance becomes flatter with more players. For the three-player uniform specification, an excess mean of 0.2 requires an excess variance of 0.7, but an excess mean of 1 requires an excess variance of over 60.[3]

The mean-variance tradeoff becomes increasingly flat as the number of baseline players increases. This is due to the fact that excess mean is increasingly less valuable as the number of players increase. This can be seen most clearly in the uniform specification. As the number of players increase, the maximum value of the baseline players' items approaches the upper bound of the distribution (in this case, 1). Thus, a marginal increase in the excess mean is likely to have less of an effect on the win probability.[4] Excess variance, on the other hand, is much less dependent on the number of baseline players. By increasing the range (i.e., the support of the distribution), the high variance item has a higher probability of achieving a value that is above the support of any of the other items, which results in a victory with certainty.

Figure 2: Mean-Variance Tradeoff for Symmetric Distributions



Excess mean becomes increasingly less valuable as the number of players increases, while excess variance

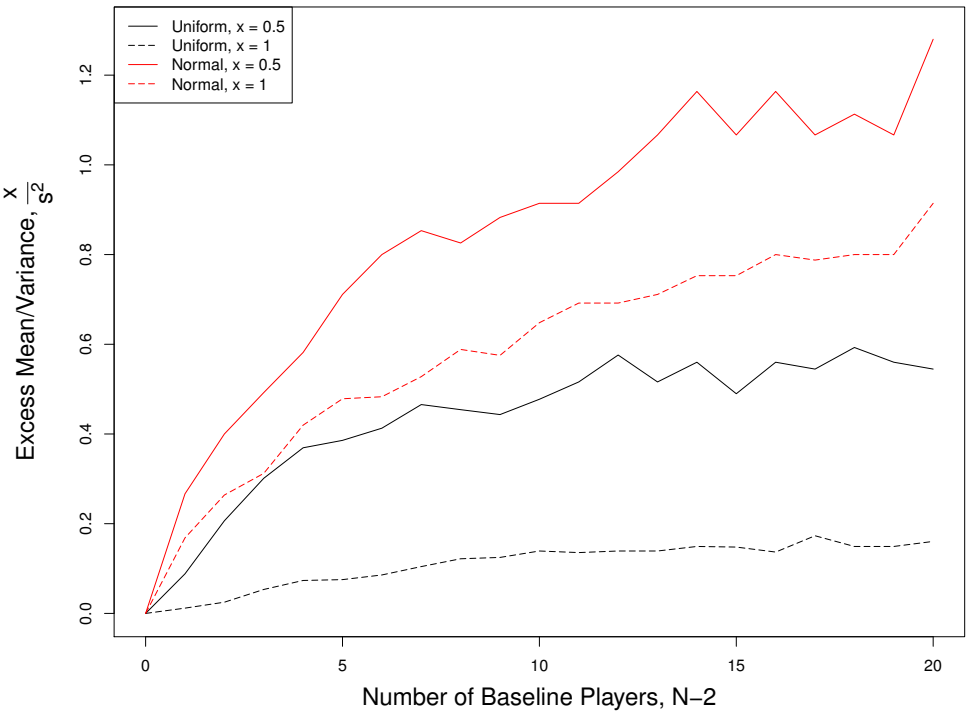
[2]Note that for the uniform specification, x can only take values in $[0, 2)$. For an excess mean above 2, the $N-2$ items that are distributed $U[-1, 1]$ will never have the maximum value, so this reduces to a two player, two item setup. As shown in Section 2, no amount of variance can compensate for the difference in expected value, so $s = \infty$. In principle, for the normal specification, x can take on any positive value, since normal distributions feature an unbounded support, so no value of x reduces to a two player setup. Nevertheless, the value of s approaches infinity as x increases.

[3]There is a decline in the excess variance from $x = 0.95$ to $x = 1$. We do not expect the true relationship between mean and variance to exhibit such non-monotonicities. The decline shown in the figure is likely a result of our simulation results. To solve for the value of s, we rely on update rules that sequentially narrow the window of search based on which distribution has a greater probability of winning. As a result, an anomalous result can lead to the algorithm narrowing the region when in fact it should not. We hypothesize that this is the cause of the decline.

[4]For example, if the excess mean is 1, there is still only a 50% probability of the high mean item having a greater value.

is relatively unaffected. Thus, for a given excess mean x, the excess variance required to compensate for this is a decreasing function in the number of baseline players. This result is being driven in the decreasing marginal value of excess mean, rather than the marginal value of excess variance.

Figure 3: Mean-Variance Tradeoff for Varying N



We can see the relationship between the mean-variance tradeoff and the number of baseline players in Figure 3. This plots the ratio of the excess mean, x, to the excess variance, s^2, as a function of the number of baseline players. We examine the uniform and normal specifications with excess means of $x = 0.5$ and $x = 1$. We see that this ratio is concave in the number of baseline players. A lower excess mean results in a more concave function, but this is purely a result of the exponential nature of the mean-variance relationship documented in Figure 2. There is a flattening relationship due to the fact that, as mentioned above, there is diminishing marginal returns to increased excess mean. As a result, for a given excess mean, as the number of players increases, less excess variance is required to compensate for this. This gives the increasing relationship shown. The concavity of this function indicates the excess variance required decreases at a decreasing rate.

Lastly, we see that the normal distribution has a function that is consistently greater than the uniform distribution. Indeed, both normal specifications (with excess means 0.5 and 1) are both strictly greater than both uniform specifications. This indicates that for a given x and N, the uniform distribution requires a higher excess variance to compensate for this. We believe this is due primarily to the bounded support of the uniform distribution. Increasing the variance of a uniform distribution reduces the mass on any given event but increases the support. Increasing the variance of a normal distribution does not expand the support, but increases the probability of extreme values. As a result, higher excess variance is more valuable in the normal

specification than in the uniform specification, and, as such, less excess variance is required to compensate for a given excess mean.

In theory, this mean-variance tradeoff could be examined for any distribution that accepts at least two parameters, including both skewed and discrete distributions.[5] However, there are issues with both skewed and discrete distributions. For skewed distributions, altering the parametrizations not only affects the mean and variance, but also the skew. Skew, as we will see in Section 4, is an important consideration. As a result, such an analysis with skewed distributions is likely to confound the tradeoff between mean and variance with variation in skew as well. We tried this analysis for the gamma distribution, but, in our simulations, we could only obtain equivalent values for very small values of x (less than 0.3; 6% of the baseline mean). This was due to changes in the skew of the distribution. As we changed our parametrizations to have the same mean but increasing variance, the distribution became more skewed to the right, shifting most of the mass to the left. This made the resulting distribution less desirable. For a fixed mean, the variance comoved with skew, and, because of this, we were not able to solve for an indifference point in the mean-variance tradeoff.

For discrete distributions, the issue was one of computational imperfection. Because almost all discrete distributions require at least one integer-valued parameter (binomial, hypergeometric, etc.), we cannot have a continuous mapping of these parametrizations to mean and variance. As a result, there are often only one or few parametrizations that result in a given mean. Solving for an indifference between mean and variance (which requires iterating over various parametrizations until convergence) is infeasible as a result.

4 Skew

The analysis so far has focused entirely on symmetric distributions, which have no skew by definition. In this section, we briefly analyze the effect of skewness on the desirability of items. Note that, for the reasons discussed above, it is difficult to perform a tradeoff analysis between mean, variance, and skewness due to the covariation in them when there are fewer than 3 parameters (as is the case with almost all common distributions). As a result, the analysis in this section relies on general findings, rather than a quantitative evaluation.

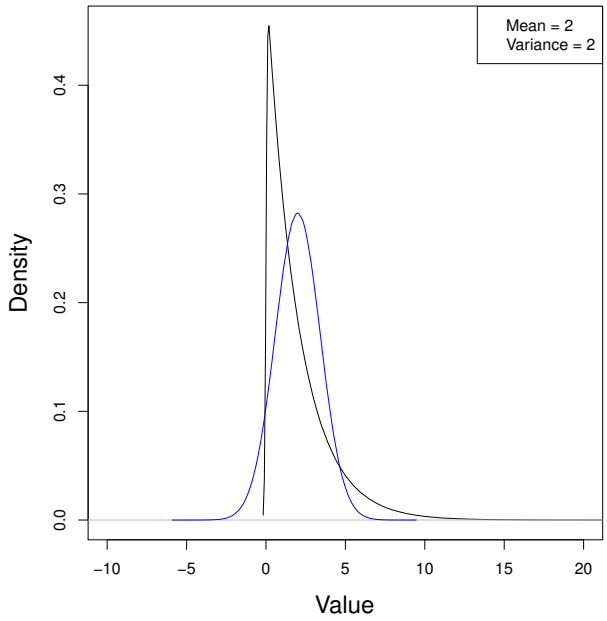
Figure 4 illustrates two distributions with different skewness, but the same mean and variance. The black line is the pdf of a gamma distribution with shape of 1 and scale of 2. The blue line is a pdf of a normal distribution with mean 2 and standard deviation $\sqrt{2}$. The gamma distribution has a skew of 2, while the normal distribution has no skew. While these distributions come from different families and exhibit different supports, this serves as a reasonable comparison to establish the importance skewness.

In our simulations, the item that is normally distributed has a value that exceeds the item that is gamma distributed 56.17% of the time. Positive skewness, therefore, does not seem to be a desirable property. This is fairly intuitive; because positive skew shifts the mass to the left of the distribution, the median is less than the mean (while, in a distribution with no skewness, the mean and median are the same). This means that the median of the gamma distribution is less than the median of the normal distribution. Indeed, the

[5] Two parameters are required, as we need to be able to vary the mean and variance independently of each other. Therefore, we need at least two degrees of freedom.

gamma distribution has a realization below its mean (2) 63.2% of the time, compared to 50% of the time for the normal distribution. Because of this, the normal distribution has a significant advantage over the gamma distribution.

Figure 4: Normal and Gamma Distributions



5 Cycles

In choosing items sequentially, it is important to understand whether cycles can exist in item comparisons. Specifically, do there exist three items X_1, X_2, and X_3 such that X_1 beats X_2 more than 50% of the time, X_2 beats X_3 more than 50% of the time, and X_3 beats X_1 more than 50% of the time? Then, a player's preferences can be expressed as $X_1 \succ X_2 \succ X_3 \succ X_1$. This intransitivity of preferences (a so-called Condorcet cycle) is generally an undesirable property, and is typically a ruled out by assumption in decision theory research.

In the adversarial knapsacks problem, cycles are possible. This is due to the fact that, even with independent random variables, the distribution of the difference of two random variables is given (in part) by the product of their probability density functions. As a result, Condorcet cycles can occur due to the interaction between these distribution functions. An example of a possible cycle is given by the following three items:

$$X_1 = 0$$

$$X_2 = \begin{cases} 1 & w.p.\ 0.51 \\ -1.04 & w.p.\ 0.49 \end{cases}$$

$$X_3 = \begin{cases} -1 & w.p.\ 0.52 \\ 1.08 & w.p.\ 0.48 \end{cases}$$

All three items have zero mean. X_2 beats X_1 51% of the time, X_1 beats X_3 52% of the time, and X_3 beats X_2 $(0.48 + (0.49 \times 0.52)) = 73.48\%$ of the time. Thus, we have $X_2 \succ X_1 \succ X_3 \succ X_2$.

Examples of Condorcet cycles that are analytically tractable tend to be "unnatural" in the sense that they are ad-hoc specifications of distributions. We are interested in whether there exists a "natural" cycle, i.e., a cycle that features items whose values follow common distributions (uniform, normal, binomial, etc.). To do so, we simulate pairwise comparisons between 22 parametrizations of 13 distributions, and search for cycles among them. We place the restriction that all distributions have the same expected value to reduce the dimensionality of the problem. We set this value to 10.

Table 1: List of Parametrizations

#	Distributions	Parameters	Empirical Variance	Empirical Skewness
1	Uniform	$min = 0$, $max = 20$	34.09	0.007
2	Uniform	$min = 5$, $max = 15$	8.28	–0.011
3	Uniform	$min = -10$, $max = 30$	132.90	-0.017
4	Gamma	$shape = 10$, $scale = 1$	9.79	0.638
5	Gamma	$shape = 100$, $scale = 0.1$	1.00	0.209
6	Hypergeometric	$m = 20$, $n = 40$, $k = 30$	3.30	-0.025
7	Hypergeometric	$m = 40$, $n = 20$, $k = 15$	2.52	-0.122
8	Binomial	$n = 1$, $p = 0.1$	8.95	0.277
9	Binomial	$n = 15$, $p = 2/3$	3.36	-0.176
10	Poisson	$lambda = 10$	10.05	0.291
11	Geometric	$p = 1/11$	113.14	1.977
12	Negative Binomial	$size = 10/9$, $p = 0.1$	99.54	1.920
13	Negative Binomial	$size = 90$, $p = 0.9$	11.15	0.332
14	Exponential	$rate = 0.1$	103.78	1.918
15	Log-normal	$meanlog = 1.61$, $sdlog = 1.18$	297.01	7.043
16	Normal	$mean = 10$, $sd = 4$	16.07	-0.048
17	Chi-squared	$df = 10$	19.70	0.902
18	Weibull	$shape = 10$, $scale = 10.51$	1.40	-0.619
19	Weibull	$shape = 2$, $scale = 11.28$	27.73	0.622
20	Discrete Uniform	$min = 0$, $max = 20$	36.59	0.009
21	Discrete Uniform	$min = 5$, $max = 15$	10.17	-0.007
22	Discrete Uniform	$min = -10$, $max = 30$	138.45	-0.014

Table 1 presents the parametrizations and distributions considered, as well as their empirical variance and skewness. There is a wide range of variance and skewnesses considered. Variance ranges from 1 to nearly 300, while skewness ranges from -0.18 to nearly 7. These distributions provide a starting point for understanding whether Condorcet cycles can arise "naturally."

As with the example devised above, the margins in the cycle are fairly small. For example, X_2 beats X_1 51% of the time. Since our simulations estimate win probabilities with error,[6] the determination of what constitutes a "significant" win probability is essential. To calibrate this choice of the margin required for a preference to be considered significant, we rely on the theory in Section 2. Since we know for symmetric distributions, only expected values, all uniform, normal, and discrete uniform specifications should perform identically. Of the 21 pairs of these 7 distributions (3 uniform, 1 normal, 3 discrete uniform), the maximum difference between in winning probabilities is 0.64%. Thus, any margin that is greater than 0.64 should be significant at the 5% level. We thus set the critical margin to 0.64.

Using 0.64 as the critical margin, we observe a cycle between distributions 1, 18, and 6. Distribution 1 beats 18 by a margin of 0.7%, 18 beats 6 by 2.04%, and 6 beats 1 by 1%. Similar to the example above, the margins are fairly small.[7] Thus, out of the 231 pairs of distributions considered, there is a cycle between the $U\left[0,20\right]$, $Weibull\left[10,10.51\right]$, and $HypGeom\left[20,40,30\right]$.

These cycles are interesting phenomena. Not only do they violate a key axiom in standard preferences, but also form intersting game-theoretic applications. The presence of cycles essentially affects whether the first mover has an advantage. If cycles did not exist, players in a sequential game would simply choose the item that performed the best against other items. Since there are no cycles, there is a globally-preferred option, and player 2 would only be able to choose the next-best option. However, if there were cycles in all items, the first-mover would be guaranteed to lose on average. For any item player 1 chose, player 2 would be able to choose another item that performed relatively better. In the case in which there are cycles with some items, the first mover would never choose one of those items, for the same reason. This essentially reduces the option set available to the first mover.

6 Minimax and Mixtures

7 Conclusions and Future Work

[6] We use 100,000 simulations.

[7] These margins are close to the critical margin of 0.64, but our simulations do not allow for greater power in this test.