Geometric reasons for normalising variance to aggregate preferences

Owen Cotton-Barratt

January 23, 2013

Abstract

We investigate the classical social choice question of how to aggregate utility functions to make a decision. Following Harsanyi's theorem, if the method is Pareto it consists of taking sums of normalised functions. We explore formal approaches to normalising.

We present a geometric view of the space of utility functions. We explain how this picture makes the variance of a function the most natural notion of size, and hence normalising the variance the most obvious aggregation rule.

We define the expected utility of voting, and explain that it is a generalisation of voting power to the case with more than two outcomes. We show that as the number of preferences aggregated becomes large (and with mild extra assumptions), normalising variance is the only method which gives equal utility of voting to all sets of preferences.

We explore the distinction between tactical and strategic voting, where the latter captures the way people will vote not holding any expectations about how others will vote. With the same assumptions as before we prove that variance normalisation is the only weighted sum method which is immune to strategic voting.

1 Introduction

In this paper we are concerned with the classic problem of how to aggregate preferences. We will use some standard assumptions to frame this problem formally. Within this framework, the question arises how best to normalise utility functions to make them scale-comparable (see Section 2).

It has often been assumed that the only sensible formal method available here is to normalise the utility functions so that they have equal ranges: a minimum of 0 and a maximum of 1, for example. We will challenge this assumption, and present results which suggest that instead normalising the utility functions so that they have equal variances (or equivalently equal standard deviations) is the correct approach.

By constructing the space of utility functions, and considering it as a geometric object, we find very natural normalisation rules (Section 3). Different geometries on the space give rise to different normalisations.

The familiar Euclidean geometry tells us to normalise the utility functions so that they have equal variances. The chessboard geometry (where the distance between two points is the maximum of the differences in the coordinates) recovers the range normalisation rule. And the taxicab geometry (where the distance between two points is the sum of the differences of the coordinates) gives rise to a third normalisation, which equalises the average difference between the value of the utility function and the median of that function.

We will then make the additional assumptions that our normalisations depend only on the function being normalised, that the number of preferences being aggregated is large compared to the number of options available, and that the aggregated preferences will be used to select an outcome (Section 4). These assumptions are justified at greater length in Appendix A. We use these assumptions to compute the expected utility of voting – that is, of having preferences counted in the aggregation. This is a natural generalisation of the notion of voting power (well established in the theory of simple voting games, where there are only two outcomes). In Theorem 14 we show that the only normalisation method which is fair in the sense that the expected utility of voting is the same for everyone is the variance normalisation.

Finally we will consider the implications of tactical voting for different normalisation methods (Section 5). We say an aggregation method based on a normalisation is susceptible to strategic voting if it is possible to improve the expected utility of voting by misstating your preferences, even without knowing anything about how others will vote. In Theorem 19 we prove that with the same assumptions as before variance normalisation is not susceptible to strategic voting, and moreover that it is the only normalisation method with this property.

2 Establishing problem

We will consider the following general problem:

Problem 1 Given differing preferences about the best course of action, how should one decide?

Versions of this question arise in different fields, and some have attracted considerable study. The two most well-known cases are:

1. (Social Welfare) Given preferences representing what is good for the individuals comprising a society, how should a decision be chosen to maximise societal good?

2. (Voting) Given a set of individuals who will be asked to state their preferences, how should this information be used to reach a decision?

These two have differences – for instance in the voting scenario there is an extra complication in that stated preferences may differ from actual preferences – but the structures of the problems are similar. In each case there are several sets of preferences, and we are trying to combine them in a sensible way to give aggregate preferences.

We restrict ourselves to scenarios where there is an isolated decision to make (it is not embedded in a longer chain of decisions). Moreover we are primarily interested in cases with more than two outcomes, as the two-outcome scenario is simpler and more studied. We will explain how some of our ideas generalise well-known techniques in the two-outcome case.

2.1 Assumptions

We will consider scenarios with a finite set of alternatives $O = \{o_1, \ldots, o_n\}$, a finite set of individuals I, and for each $i \in I$ some preferences over the alternatives.[1]

We restrict our consideration to preferences which satisfy the von Neumann–Morgenstern axioms for rationality. These are seemingly fairly weak and obvious criteria about preferences over lotteries of alternatives; we will call preferences satisfying these *rational*. Von Neumann and Morgenstern showed

[1] Though we limit ourselves here to the case with finitely many alternatives, with some choices of measure this restriction may be dropped.

that they are equivalent to what appears to be a much stronger condition: they proved that *any* rational set of preferences can be represented by a utility function [7]:

Definition 1 *A* utility function u *is a function* $O \to \mathbb{R}$ *(which extends linearly to lotteries over* O*). It represents a collection of preferences if it assigns higher values to preferred outcomes.*

If utility functions u, u' differ by a positive affine transformation (*i.e.* $\exists c > 0, k \in \mathbb{R}$ such that $cu + k = u'$) then they represent the same preferences. Moreover if they represent the same preferences then they only differ by such a positive affine transformation.

Definition 2 *The* utility class $[u]$ *of a utility function* u *is the equivalence class of* u *under the relation which sets as equivalent all utility functions representing the same preferences. We denote by* U *the space of utility functions on* O*, and by* $\mathcal{U}$ *the space of utility classes on* O*.*

Thus we assume that for each $i \in I$, the preferences of i are captured by a utility function $u_i \in U$ (really the preferences correspond to an element $[u_i] \in \mathcal{U}$, but it is convenient to represent the utility class by a utility function).

We are looking for methods which somehow combine these utility functions to give an aggregate set of preferences. If such a set of aggregate preferences is rational, then again (by the von Neumann–Morgenstern theorem) it will be expressible by a utility function.

Definition 3 *A* choice method *is a function* $f : \mathcal{U}^m \to \mathcal{U}$*, where* $m = |I|$*. It is said to be* anonymous *if it is invariant under permutation of the labels of the individuals, and* neutral *if it is invariant under permutation of the labels of the outcomes.*

We will at times abuse notation and refer to a choice method as a function from U^m, or to U, or both. This is always with the understanding that a choice method from U^m factors through the quotient $U^m \to \mathcal{U}^m$, and if it takes value $u^* \in U$ this is just a convenient way of specifying the point $[u^*] \in \mathcal{U}$.

2.2 The case for weighted sums

A choice method $(u_1, \ldots, u_m) \mapsto u^*$ is said to be *Pareto* if for any two lotteries A, B, if $\forall i \in I, u_i(A) \geq u_i(B)$ and $\exists i \in I : u_i(A) \gneq u_i(B)$, then $u^*(A) \gneq u^*(B)$. In other words, if nobody prefers B to A and someone prefers A to B, then collectively the group prefers A to B.

There are different ways one could interpret the social choice problem. We will be concerned with the following version:

Question 4 Social Choice *Given a set I of individuals with utility functions $u_i (i \in I)$ over an option set O, what is the fairest Pareto choice method?*

This still isn't a purely technical question, since the word *fairest* is doing some work. Later in this paper we will consider two possible interpretations of this, and see that they lead to the same answer.

Before we develop the idea of fairness, we consider the other restriction; that the method be Pareto. This seems like a very obvious requirement, and a weak one. A remarkable theorem of Harsanyi places strong restrictions on the type of method which can be Pareto:

Theorem 5 (Harsanyi's aggregation theorem) *If u^* is a Pareto preference aggregation which is rational, then there exist weights $\lambda_i \in [0, \infty)$ such that $[u*] = [\sum_i \lambda_i u_i]$.*

This theorem was originally proved in [5], but there have been various proofs since then, some of which are quite geometric in nature (*e.g.* [2]).

So if we want to make rational Pareto aggregates of rational preferences we are restricted to methods which are equivalent to taking weighted sums. Since choice methods are really functions on $\mathcal{U}^m$, this must also involve a choice $\bar{u}_i$ of *normalised representatives* of each utility class $[u_i]$ as well the weighting factor λ_i; but so long as they are non-zero we can simply absorb the λ_i into the choice of $\bar{u}_i$.

Corollary 6 *Any rational Pareto choice method f may be decomposed as a collection $\{f_i\}_{i \in I}$ of* normalisation functions $\mathcal{U}^m \to U$ *such that $f = [\sum f_i]$ and $[f_i([u_1], \ldots, [u_m])] = [u_i]$.*

Given this result, one route to understanding rational Pareto choice methods is to explore possible normalisation methods. This is the course the rest of this paper will take. We will look for natural normalisation methods, as these might lead to natural aggregations. We will also investigate how properties of normalisations interact with properties of the resulting choice methods.

We make one more definition.

Definition 7 *The normalisation functions f_i are* intrinsic *if f_i factors through projection to the i^{th} factor of $\mathcal{U}^m$;* i.e. *they depend only on the utility function being normalised, not on the utility functions of other individuals.*

All of the normalisation methods we mention will be intrinsic.

At one point later we will have to restrict our consideration to just intrinsic functions, and some of our results follow this restriction. However this is not so obviously correct as our assumptions that the choice methods be rational and Pareto, for example. The best defence I can offer for the assumption is that the concept that each set of preferences has a natural weight is appealing. I also observe that (as far as I am aware) no author has proposed a normalisation method which is not intrinsic. Otherwise this assumption may have to be treated as a matter of convenience, to facilitate our analysis, and our conclusions limited to its domain.

2.3 Existing approaches

For simplicity we henceforth assume that each u_i is non-constant; otherwise it expresses no preferences, and all normalisations lead to the same utility class for the weighted sum (including the 'zero normalisation', which is equivalent to ignoring it).

Despite the fact that the assumptions we have made so far are relatively weak, the results well-known, and their conclusions clear, there has been relatively little exploration of the merits of different normalisation methods. However, a few methods have been suggested. One is particularly widely known:

Example 8 Range normalisation *is the intrinsic normalisation where $\bar{u}_i$ is the transformed version of u_i such that its highest value is* 1 *and its lowest value is* 0 *(so the* range *of the utility function has been normalised to be* $[0, 1]$. *This has enjoyed widespread use in economics.*

Range normalisation certainly has some nice properties. It is neutral, anonymous, and intrinsic.

Indeed Hausman has argued that if there is any natural normalisation is must be range normalisation [6], but his arguments are really in favour of normalisations which equalise the size of the utility functions, and he seems to assume that range is the only sensible measure of size in this context. We will see in Section 3 that there are some others at least as natural, and in Section 4 we provide theorems which give reasons to prefer to normalise variance in particular.

Example 9 *It has also been suggested, for example by Sen [9], to normalise utility functions such that the difference between the average utility and the worst utility is a constant (1, say).*

We will see in Section 3.3 that the range normalisation does capture a very reasonable notion of size, which Sen's alternative does not. However it is not the only normalisation to capture size, and it does not seem the most natural.

3 Geometric motivation: the space of utility functions

We have seen that rational preferences may be expressed by a unique choice of utility class. The problem is that they do not come with canonical scales, which would allow comparisons between different utility functions. We are seeking to impose a scale given only the information of the utility class, and matters of scale and size can often be most clearly understood geometrically. In this section we explore this idea.

The space U of utility functions on $\{o_1, \ldots, o_n\}$ is in a natural one-to-one correspondence with $\mathbb{R}^n$, where u corresponds to $(u(1_1), \ldots, u(o_n))$. We will use this correspondence to inform our intuitions about what makes normalisations natural. The purpose of normalising is to be fair by making all of the functions the same *size* before adding, and notions of size and distance are much closer to the surface in $\mathbb{R}^n$.

First we will give a picture of the utility classes in this space.

Definition 10 *A utility function is* nihilistic *if it assigns the same value to every outcome. The* line of nihilism, *L, is the line* $\{(\lambda, \ldots, \lambda) : \lambda \in \mathbb{R}\} \subset \mathbb{R}^n$ *corresponding to nihilistic functions.*

The line of nihilism is a special utility class; the one corresponding to the belief that all outcomes are equally good. If $u \in U$ and $l \in L$ then $[u] = [u+l]$, so elements of L function as a kind of zero in the space. So while our normalisation should pick out a normalised representative of each utility class, this normalised representative need only be defined up to addition of nihilistic functions. That is, instead of a normalised point in U we may look for a normalised line K parallel to L.

Two utility functions u, u' lie in the same utility class if there exist $c > 0, k \in \mathbb{R}$ such that $\forall o \in O, u'(o) = cu(o) + k$. If they are not nihilistic, this happens precisely when u' lies in the plane spanned by u and l (where l is the nihilistic function assigning value 1 to everything), and u and u' lie on the same side of L in this plane. Thus the non-nihilistic utility classes correspond to half-planes in $\mathbb{R}^n$, each with boundary L. The other half of the half-plane associated with $[u]$ is associated with $[-u]$ (see for example Figure 1).

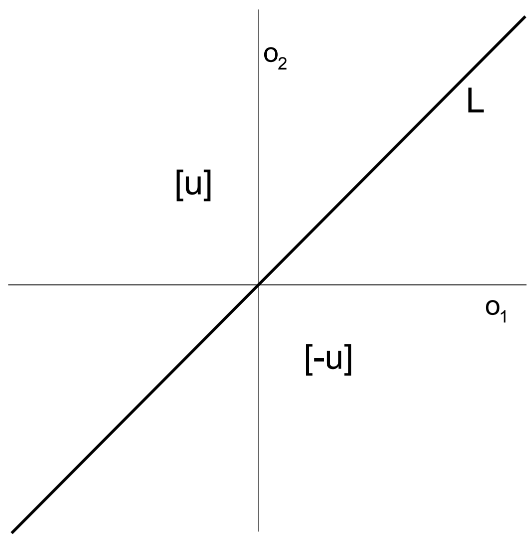

Figure 1: A picture of U when $n = 2$. In this case there are just two non-nihilistic utility classes corresponding to the half-planes on either side of L.

There is a reasonable sense in which the size of a utility function is its distance from nihilism. We will use obvious notions of distance on U to capture this size, and hence normalisation methods (since the idea of normalisation is that all normalised theories should have the same size).

There is more than one metric which we might consider on U, coming from its $\mathbb{R}^n$ structure. The Euclidean (or l^2) metric is the most standard, and the most homogeneous. Admittedly it is not immediately obvious that we should seek a homogeneous metric, as there are some distinguished directions in U corresponding to changing opinion of one outcome while leaving other preferences fixed. We will briefly investigate two others: the l^1 metric defines the distance between two points to be the sum of the differences in each coordinate; and the l^∞ metric defines it to be the greatest coordinate difference.

For ease of thought and to aid diagrams, we will for the time being work in the hyperplane $H = \{(x_1, \ldots, x_n) \in U : \sum_{i=1}^n x_i = 0$ which is perpendicular (in the Euclidean metric) to L. We imbue H not with the subspace metric, however, but with the quotient metric where the distance between two points $h, h' \in H$ is equal to the distance between the lines $h+L$ and $h'+L$ in U which correspond to the normalised utility classes of h and h'. Each normalised utility class K has a single representative in H, and L intersects H at the origin. Each non-nihilistic utility class corresponds to a ray in H with its end at L.

3.1 Euclidean Normalisation

With the Euclidean metric, the shortest path from any point to the line of nihilism will run perpendicular to the line. Given a point $p = (p_1, \ldots, p_n) \in U$, we may replace it by $\bar{p} = (p_1 - m, \ldots, p_n - m)$, where m is the mean of $p_1, \ldots, p_n$. Since $\bar{p}$ is just a translation of p parallel to the line of nihilism, it will lie at the same distance from that line. Moreover $\bar{p}$ lies in the plane through the origin $\{(x_1, \ldots, x_n) : \sum x_i = 0\}$ which is perpendicular to the line of nihilism, so the closest point to $\bar{p}$ on L is the origin. The distance between these two points is $(\sum(p_i - m)^2)^{1/2}$, which is nothing other than the standard deviation $\sigma(\bar{p}) = \sigma(p)$.

Thus to normalise the Euclidean distance from nihilism one normalises standard deviation, or, equivalently, the variance σ^2.

3.2 Normalising using l^1

With the l^1 metric we consider the distance from p to the point $\underline{x} = (x, \ldots, x) \in L$ as a function of x; this will be equal to $\sum |p_i - x|$. This is a continuous function, and the derivative of this with respect to x is defined except at values p_i, and equal to the number of points in the tuple $(p_1, \ldots, p_n)$ which lie above x minus the number of points in the tuple which lie below x. It

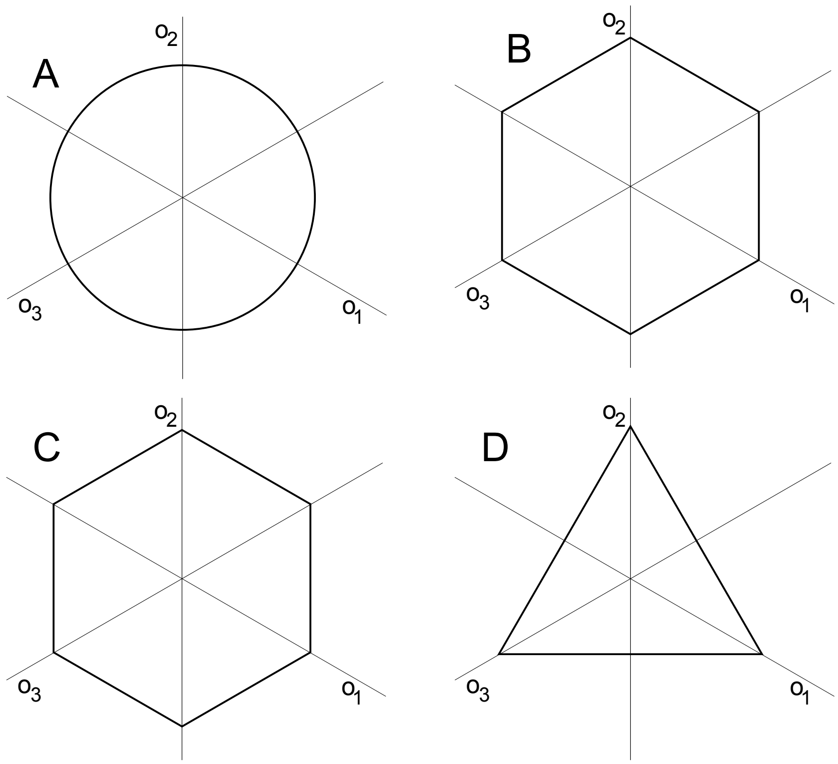

Figure 2: Four instances of H when $n = 3$; the 3 lines meeting at the centre are the projections of the axes of U onto H. The shapes show the normalised representatives (one in each ray from the origin) coming from different methods. In A we have the unit circle coming from the Euclidean (l^2) metric on U; B and C have unit hexagons coming from the l^1 and l^∞ metrics respectively; and D has the triangle which would be normalised to under the rule mentioned in Example 9. The hexagons in B and C are the same because the metrics on H induced by l^1 and l^∞ coincide up to a multiplicative constant when $n \leq 3$. They diverge in higher dimensions. When $n = 4$, H is 3-dimensional, and the corresponding figures would be a sphere in A, a cube in B, a rhombic dodecahedron in C, and a tetrahedron in D.

is easily seen that this takes its minimum value when the same number of points among $(p_1, \ldots, p_n)$ lie above x as lie below.

Thus a closest point on the line of nihilism lies at $\underline{x}$, where x is the median value of the p_i, at distance equal to the sum of coordinate distances to the median. Normalising with respect to the l^1 metric corresponds to normalising the average distance to the median.

3.3 Normalising using l^∞

With the l^∞ metric the distance from p to $\underline{x}$ will be $\max_{\mathrm{p_i}} |\mathrm{p_i} - \mathrm{x}| = \max\{|\max(\mathrm{p_i}) - \mathrm{x}|, |\min(\mathrm{p_i}) - \mathrm{x}|\}$. This is obviously minimised when $x = (\max(\mathrm{p_i}) + \min(\mathrm{p_i}))/2$, with distance $(\max(\mathrm{p_i}) - \min(\mathrm{p_i}))/2$. Thus normalising with respect to this recovers the traditional range normalisation.

3.4 Comparison of metrics

We remarked that since our space does have distinguished axes, it is not automatic that we should seek a homogeneous metric. The l^∞ metric gives a strange notion of distance here, though. It implies that two utility functions agreeing precisely except on one pair of outcomes, which are transposed, are as far apart from each other as if one of them is further disturbed so that they disagree everywhere, but nowhere by more than the transposed pair.

Intuitively, the disturbance feels like it should move them further apart. We are therefore inclined to prefer the l^1 and the Euclidean distances in this space.

3.4.1 Canonical forms

Although from an aggregation perspective it is irrelevant *what* the respective quantities are normalised to, or *which* representative of the normalised utility class is chosen to be summed, in practice it can be helpful to have a canonical choice. Range normalisation has just such a default, with the unique choice of representative being that such that the worst option has utility 0 and the best option has utility 1.

It would be useful to establish such defaults for the other normalisation methods here. Fortunately there are some obvious choices.

In the remainder of this paper we will give additional reasons for preferring the variance normalisation, and we will not return to the normalisation given by the l^1 metric beyond this brief comment: it seems natural to map the median value to 0 and have the mean distance from the median be 1.

For variance normalisation, we suggest that it is most canonical to normalise to a variance of 1 and choose a representative with a mean of 0. This is, for example, the form of the standard normal distribution. It has the nice property that the 0 point is meaningful, representing the value of the uniform lottery over the outcomes, so everything with a positive value is preferred to a (uniform) random choice.

4 Voting utility and the large electorate assumption

In Section 3 we explored geometric interpretations of size. These led us to conclude that a few normalisation methods, in particular normalising variance, are particularly natural. This was suggestive, but as we could not find strong reasons to prefer any one geometry, it was not conclusive. In this Section and Section 5 we present some simple desirable properties for normalisation rules and show that under mild assumptions normalising the variance is the *only* rule which has these properties.

We will now change our language slightly, and refer to aggregation methods as voting systems, and the sets of preferences being aggregated as belonging to voters. This is because it is convenient to compare our set-up with existing theory which uses this language. We are not at this point restricting the scope of aggregation situations we consider. In Section 5 we will focus on the extra structure present in voting scenarios where voters may misrepresent their preferences in the expectation of a better outcome.

A key notion in the study of simple voting games, which have an assembly of voters and some rule for passing motions according to who votes for them, is *voting power*. It expresses the *a priori* likelihood that an individual's vote will make a difference to the outcome [3].

We would like to measure a similar concept in the context of making decisions by taking normalised sums of utility functions. The main difference is that, since we have a utility function for each voter, we are interested in not just whether they can affect the outcome, but by how much. Certainly this is what the voters care about: preferring a fair chance of a significant improvement (in their eyes) to a slightly larger chance of a much smaller one.

Definition 11 *The* voting utility *of a voter is the expected utility gained by voting.*

This is not quite a complete definition, because to take an expectation you must have a prior over the results with your vote counted or not (and hence probably over how others will vote). However it is reasonable to assume that in most situations you have some prior, and given a prior it is a natural quantity to consider. In Section 4.1 we will explore the consequences of some assumptions about the prior.

Voting utility has been studied in the context of whether it's rational for people to make the effort to vote[8], but although it captures precisely what people care about, it doesn't seem to have been used as a measure of the fairness of different voting methods.

Of course as voting utilities are expectations of individuals' utility functions, to compare them we must be able to compare these utility functions – which in some sense is the whole problem we have set out to address! But if we assume that we have a sensible normalisation method, the purpose of the normalisations is to provide a reasonable way to make the functions comparable. We will therefore use the same normalisation to compare voting utilities.

Earlier we made an analogy with voting power, used in the case of simple voting games where there are just two outcomes. With this last assumption voting utility is actually a generalisation of voting power, since in the case with just two outcomes there are only two utility classes, depending on which outcome is preferred, as we saw in Figure 1. So the voting utility is determined entirely by the expected chance of changing the outcome; when others are assumed to vote uniformly at random this is the definition of voting power.

4.1 Large electorate assumption

Henceforth we will assume that our normalisation methods are intrinsic (see Definition 7). This is a non-trivial assumption, but it is both intuitive, and useful in facilitating analysis. The assumption is intuitive in expressing that the degree to which a vote is counted does not depend on other votes; indeed all of the previously suggested normalisation methods discussed above in Section 2.3 are intrinsic. It is useful in considering voting utility, for if the normalisation method is intrinsic then adding a single vote consists of adding a known vector (representing the normalised vote) in U to one drawn from a distribution depending on the prior beliefs about others' votes (representing the sum of the other normalised votes). If the method is not intrinsic then each of these vectors may depend on the other, potentially in some complicated manner.

As we saw, the definition of voting utility is dependent on the choice of a prior over the votes of others. Rather than specify such a prior at this stage, we reason about a class of non-pathological priors.

Since our normalisation method is intrinsic, a prior over the votes of others will lead to a prior over vectors in U, representing the sum of the other normalised votes. Essentially we will assume that the number of other voters is large compared to the number of outcomes available (so the expected size of the vector in U is large compared to the size of the vector which represents the vote we will add in), and that our prior does not have a precise idea of how the voting will go, hence has a degree of continuity.

We suggest this assumption is reasonable since if a method is correct it should at least work in the limiting case with many voters.

Assumption 12 (Large Electorate Assumption) *In considering voting utility it is enough to consider pairs of outcomes. Moreover the chance that a voter will switch the chosen outcome from b to a is proportional to the difference in the utilities of a and b in their normalised vote.*

A technical justification for this assumption is given in Appendix A. Roughly speaking, as the number of voters grows large, the chance that the outcome is close enough that a single vote can move it to any side of a tie between three or more outcomes becomes small compared to the chance that it will be close between two outcomes. This explains the first part of the assumption. The second condition, local linearity, comes from the fact that having many voters will smooth out any uncertainty.

4.2 Computing voting utility

We will now use these assumptions to compute the voting utility for a utility function.

Given a prior distribution μ over pairs of outcomes, the voting utility for a voter with utility function u and (normalised) vote v is:

$$\mathcal{VU}_u(v) = \sum_{\{a,b\}\in O^2} \mu(\{a,b\})(v(a) - v(b))(u(a) - u(b)). \tag{1}$$

Here the factor $\mu(\{a,b\})(v(a) - v(b))$ is the chance that the vote makes us switch to a from b (where a negative value is understood to mean a chance of switching to b from a), and the factor of $u(a) - u(b)$ captures the utility value of such a switch.

4.2.1 *A priori* voting utility

Voting power is an *a priori* measurement of how likely a vote is to affect an outcome, computed by assuming that all the possible ways for other people to vote are equally likely. When voters submit a utility function to be normalised there are no longer finitely many possible ballots, so we cannot weight them all equally. One should still be able to use an analogous definition by choosing an appropriate measure on $\mathcal{U}$, but to avoid having to make a choice of measure we will just note that, at least, an *a priori* method should be symmetric in its expectations between the different outcomes.

So if we want to find the *a priori* voting utility we may assume that $\mu(a, b)$ is constant ($=k$) across all pairs $a \neq b \in O$. This assumption allows us to simplify our expression of voting utility:

Lemma 13 *With the large electorate assumption, the* a priori *voting utility $\mathcal{VU}_u(v)$ for a voter with utility function u and normalised vote v is $k(\sum_{a\in O}(nv(a)u(a)) - \sum_{a\in O}(v(a)\sum_{b\in O} u(b))$, where k is a constant depending on the scenario but not on u or v.*

Proof Putting the assumption about μ into Equation 1 gives:

$$\mathcal{VU}_u(v) = \sum_{\{a,b\}\in O^2} k(v(a) - v(b))(u(a) - u(b)). \tag{2}$$

Now we'll take some algebraic steps to simplify this slightly. First, decomposing the sum over pairs into two sums over individual outcomes gives

$$\mathcal{VU}_u(v) = k/2 \sum_{a\in O} \sum_{b\in O} (v(a) - v(b))(u(a) - u(b)), \tag{3}$$

where it's irrelevant that we've introduced terms for $a = b$, since these will all vanish, and a factor of $1/2$ has been introduced to balance the fact that each pair $\{a, b\}$ is now counted twice, once in either order.

Now we expand, to get:

$$\mathcal{VU}_u(v) = k/2 \sum_{a\in O} \sum_{b\in O} (v(a)u(a) + v(b)u(b)) - (v(a)u(b) + v(b)u(a)). \tag{4}$$

Since $u(a)$ and $v(a)$ are independent of b, and likewise in reverse, this gives:

$$\mathcal{VU}_u(v) = k/2(\sum_{a\in O}(nv(a)u(a))+\sum_{b\in O}(nv(b)u(b))-\sum_{a\in O}\sum_{b\in O}(v(a)u(b)+v(b)u(a)), \tag{5}$$

where the coefficients of n arise because we are summing over a set of size n. Then since a and b are just indexing letters we can combine the first terms and reorder the later ones to get

$$\mathcal{VU}_u(v) = k/2(2\sum_{a\in O}(nv(a)u(a))-\sum_{a\in O}((v(a)\sum_{b\in O}u(b))+(u(a)\sum_{b\in O}v(b)))). \tag{6}$$

Finally, this reduces to

$$\mathcal{VU}_u(v) = k(\sum_{a\in O}(nv(a)u(a)) - \sum_{a\in O}(v(a)\sum_{b\in O}u(b)). \tag{7}$$

□

As a corollary we have the following theorem.

Theorem 14 *If an intrinsic normalisation method is used to produce a voting method and to measure the voting utility of voters, it gives everyone equal voting utility under the condition of the large electorate assumption if and only if it is variance normalisation.*

Proof We take $v = u$ in the conclusion of Lemma 13 to get: $\mathcal{VU}_u(u) = nk(\sum_{a\in O}((u(a))^2) - (\sum_{a\in O}(u(a))^2$. This expression is simply nk times the variance of u. □

It is often seen as desirable that voting systems be fair, and neutrality captures an important component of this: it means that a person's vote is counted the same regardless of the person making it. Equalising voting utility adds an extra dimension to fairness, meaning that (as best we can measure) the value of each vote to the person making it is the same, regardless of their preferences.

5 Tactical voting and strategic voting

In Section 4 we introduced the concept of voting utility and explored some of the consequences in abstract cases. We now restrict to voting scenarios with the additional complication that although we would like to maximise a weighted sum of the voter's preferences, we cannot necessarily expect the voters to truthfully report these.

Rather, we should in general expect voters to vote so as to maximise their voting utility. For a fixed voting method, if there exist scenarios where by reporting a utility function u' instead of their true utility function u some voter can lead to a strictly preferred scenario, we say that the voting method is susceptible to *tactical voting*.

Since Arrow's celebrated theorem [1] there have been a number of results about the susceptibility of different systems to tactical voting. In particular the Gibbard–Satterthwaite theorem presents a strong obstruction to escaping tactical voting altogether [4]:

Theorem 15 *If a voting system with more than two outcomes is Pareto and not susceptible to tactical voting, it is a dictatorship.*

It seems that trying to avoid tactical voting is too much, then. But there is a similar, weaker condition.

Definition 16 *A voting system is susceptible to* strategic voting *if it is possible for a voter to increase their* a priori *voting utility by falsely stating their preferences.*

Remark 17 *Some authors use* strategic voting *to refer to what we have here called tactical voting, but it is less common. The clash with our usage here is unfortunate, but we feel the distinction between 'strategic' and 'tactical' is natural and helpful enough to justify our appropriation of the term.*

A system which is immune to tactical voting is said to be *preference revealing*; immunity to strategic voting corresponds to being preference revealing under restricted conditions. Specifically, when the voter has no expectations that other voters will favour or disfavour particular outcomes.

This seems like a weak requirement, and in the context of preference aggregation obviously a desirable one. For Harsanyi's theorem tells us that if we are not maximising some weighted sum of the true utility functions we may not even manage a Pareto outcome. And if voters will vote strategically even in the case where they have no information about others, we cannot ever expect a true representation of their preferences, so will have no guarantee of Pareto outcomes.

In light of this, we might hope at least that the normalisations the geometries in Section 3 suggested as natural would be immune to strategic voting. However this is not true of all of them. Range voting is well known to be susceptible to strategic voting, and we recapitulate the argument in our terms:

Example 18 *Let v, v' be utility functions differing only on outcome o, where v takes a higher value. Then applying Lemma 13 we see that $\mathcal{VU}_u(v) > \mathcal{VU}_u(v')$ if and only if $u(o)$ is higher than the mean value of u over O.*

With the range normalisation rule, the utility function may be altered by moving the non-extremal values freely while remaining normalised (so long as they do not become new extremes). Therefore strategic range voting degenerates to the case where all outcomes which are better than the mean are ranked equally top (at 1, say), and all which are worse than the mean are ranked equally bottom.

Excepting the case where one outcome has exactly the mean utility (in which case the voting utility will be the same regardless of where it is placed), this means that strategic voting recovers the system known as approval voting.

It turns out that the normalisation suggested by the l^1 metric is also susceptible to strategic voting. A strategic voter will rank equally all of the options except the one they find furthest from the mean utility, which will be correspondingly above or below. The proof is straightforward and we leave it as an exercise for the interested reader. If everyone is assumed to vote strategically, this recovers the system where each person gets one vote and they choose a candidate and give either $+1$ or -1 to their score.

It is not just bad luck that these two methods are susceptible to strategic voting. Our next theorem shows that our third geometric notion of size, variance normalisation, is the only intrinsic normalisation method which is immune to strategic voting.

Theorem 19 *Under the large electorate assumption, a weighted sum voting method is immune to strategic voting if and only if for each individual i there is a constant k_i such that all normalised utility classes for that individual have variance k_i. In particular, an anonymous weighted sum voting method is immune to strategic voting if and only if there is some constant k such that for all individuals all normalised utility classes have variance k*

Proof First we will prove that if such k_i exist the method is immune to strategic voting. For if an individual has utility function u and declares a normalised function v, we know from Section 4.2 that their voting utility will be:

$$\mathcal{VU}_u(v) = k(\sum_{a \in O}(nv(a)u(a)) - \sum_{a \in O}(v(a)\sum_{b \in O}u(b)). \tag{8}$$

Without loss of generality we may assume that u and v are representatives of their respective utility classes lying in H, the hyperplane in U of functions with zero mean. Then the last term disappears, and we have:

$$\mathcal{VU}_u(v) = kn(\sum_{a \in O}(v(a)u(a)), \tag{9}$$

which is kn times the standard inner product of u and v. Now the Cauchy-Schwartz inequality tells us that if v has fixed norm in the space in question (which corresponds to fixed variance), then this quantity is maximised when v is in the same direction as u.

Now let us assume we have a method which is immune to strategic voting. Let f be the normalisation function. Then for any u, v it must be that $f(u).u \geq f(v).u$ (equivalently $f(u).f(u) \geq f(v).f(u)$), or it would be strategically better to declare v when holding view u. It turns out that this condition is strong enough that we can apply this condition locally and promote it to a global one.

We will use the characterisation of the inner product $u.v = |u||v|\cos\theta$, where θ is the angle between u and v. If u and v are any two vectors, then the angle between them, measured in radians, is no more than π. So for any $\epsilon > 0$ we can find a sequence of vectors $u = u_0, u_1, \ldots, u_{n-1}, u_n = v$ such that the angle between u_i and u_{i+1} is no greater than ϵ, and $n \leq (\pi/\epsilon) + 1$ (for instance by projecting u and v to the unit sphere, and placing the u_i as evenly spaced points along the geodesic arc joining u and v in the sphere).

Since we are immune to strategic voting, for every i we have $f(u_i).f(u_i) \geq f(u_{i+1}).f(u_i) \Rightarrow |f(u_i)| \geq |f(u_{i+1})|\cos\epsilon$ (as cos is a decreasing function on $[0,\pi]$). By combining these inequalities we have $|f(u_0)| \geq |f(u_n)|(\cos\epsilon)^n$. Now we use the well known inequality $\cos\theta \geq 1 - \theta^2$ (this is a good approximation when θ is small), to get

$$|f(u)| \geq |f(v)|(1-\epsilon^2)^n. \tag{10}$$

When ϵ is small, $(1-\epsilon^2)^n \geq 1 - n\epsilon^2$ (by the binomial expansion, for example). But since $n \leq (\pi/\epsilon) + 1$, we have $n\epsilon^2 \leq \pi\epsilon + \epsilon^2$. Putting this together we have:

$$|f(u)| \geq |f(v)|(1 - \pi\epsilon - \epsilon^2), \tag{11}$$

but since this holds for every positive ϵ, by letting it tend towards zero we can see that $|f(u)| \geq |f(v)|$. By a symmetric argument, $|f(v)| \geq |f(u)|$, and hence $\forall u, v, |f(u)| = |f(v)|$. Since the modulus is just equal to the variance, we can conclude that every (intrinsic) normalisation method which is immune to strategic voting normalises variance. □

Of course this immunity to strategic voting is somewhat restricted in its application, since in many scenarios voters do have some idea of how others will vote.

Remark 20 *If a voter has a probability prior ν over O expressing how likely they think each candidate is to be at the top among other voters, and that the relative placements of any two candidates will be independent of others, then a variation of the above proof tells us that an intrinsic normalisation method will reveal that voter's preferences if and only if it normalises the variance of their utility function calculated with the outcomes weighted by ν.*

6 Conclusions

Exploring the idea that normalisation is supposed to capture size, we found that normalising the variance of a utility function corresponds to our most familiar notion of size. We proceeded to use the concept of voting utility to show that if there is any natural and fair intrinsic normalisation method, it must be variance normalisation. Finally we showed that in a voting context (with many voters), unlike all other intrinsic normalisation methods, variance normalisation does not invite strategic voting.

Taken together we believe that these results provide a compelling case that when a formal intrinsic normalisation method is desired, it is correct to normalise the variance (at least if this is practical; the widely used range normalisation benefits from simplicity).

One might ask what happens if we drop the assumption of a large number of voters, or the assumption that the method is intrinsic. Either of these would complicate the analysis.

In the first case, dropping the assumption of a large electorate, there cannot simply be some other ultimately better normalisation method, for it would have to converge to variance normalisation as the number of voters increased, and if it is intrinsic it cannot depend on the number of voters so must already be variance normalisation.

If we drop the second assumption and allow non-intrinsic methods, it would be necessary to form a more complicated model before we can even speak of voting utility. It is possible that this could be done and another method with desirable properties could be found, but we are inclined to think of being intrinsic as to some degree a desirable property in itself.

In some practical situations the simple variance normalisation is not perfect. It is not independent of cloning, for example (this is easy to understand: in our workings we assumed a symmetry between the outcomes, and this is exploited in adding clones). We speculate that varying it in the direction suggested in Remark 20 might address some of these issues.

Acknowledgements

The author would like to thank Will Crouch, Toby Ord, Stuart Armstrong, James Martin, Nic Tideman, and John Quah for, variously, useful conversations and comments.

A Justifying the Large Electorate Assumption

In this appendix we'll explain why the Large Electorate Assumption will approximately hold when the number of voters is large. This is a relatively technical background section.

A normalisation function f carries the same information as the function $\bar{f} : S \to \mathbb{R}^+$ taking a utility function u in the unit sphere S in H to the length of the normalised utility function $f(u)$. We need a very mild assumption on $\bar{f}$.

Assumption 21 *There is a constant d such that the function $\bar{f}$ is bounded by d.*

Assumption 21 excludes some pathological cases from our analysis. If $\bar{f}$ is continuous then since S is compact $\bar{f}$ is automatically bounded, so this assumption is strictly weaker than continuity. Moreover if the assumption is not satisfied, there are some normalised utility functions which are arbitrarily large in favour of some outcome and against some other. If there is also symmetry between the different outcomes in the normalisation function (so the aggregation method is neutral), voting would become a game of 'name the largest number'.

Assumption 22 *Our distribution μ considers each vote to be drawn independently from an identical non-discrete distribution ν which is symmetric with respect to any permutation of the outcomes.*

Assumption 22 is a collection of conditions which express the fact that μ is supposed to be in some sense ignorant. We use it to get to Assumption 12:

Proposition 23 *Let our prior μ satisfy Assumption 22 and our normalisation function f satisfy Assumption 21. Then for a fixed number of outcomes Assumption 12 holds in the limit as the number of voters increases.*

We will prove this using the central limit theorem to get that μ converges to a normal distribution. The non-discrete condition in Assumption 22 merely rules out strong knowledge of how others will vote.

Proof Since ν is symmetric between the different outcomes, the mean will be 0 and the covariance matrix of μ will be a scalar multiple λ of the identity (where our space is the hyperplane H). By the (multivariate) central limit theorem, as the number of voters increases the sum of these votes satisfying Assumption 22 converges to a (mutivariate) normal distribution with likelihood function p, say. With m voters this normal distribution will have mean 0 and covariance matrix $m\lambda$ times the identity.

The likelihood p of a point will then be proportional to $e^{-(l^2)/2m\lambda}$, where l is the distance to the origin, as for a normal distribution in one variable.

Let $x, y \in H$ be not more than distance d apart. If $|x| = |y| + c$, then $p(x)/p(y) = e^{-(2c|y|+c^2)/2m\lambda}$ (the other terms cancel). Since $|c| \leq d$, if $|y|/m$ is small and m is large this ratio is close to 1. And as the normal distribution scales with its standard deviation, which is the square root of the variance, for sufficiently high m an arbitrarily high percentage of points y in the distribution will have arbitrarily small values of $|y|/m$.

So for large numbers of voters μ is almost everywhere almost constant at the scale of d. The two statements of Assumption 12 are a consequence of this.

From any point in H which gives a tie between three or more outcomes you can travel along the boundary between just two of them. Since the prior is close to locally constant, by travelling far enough you can see that the points within the distance d of a tie between more than two outcomes are of negligibly small measure compared to points within distance d of a tie between precisely two outcomes. While points which are close to no ties have a larger measure again, in these cases the normalised vote cannot affect the outcome, so it irrelevant for comparing voting utility.

The linearity of voting utility with the size of vote follows directly from μ being locally constant. $\square$

In fact the conditions of Assumption 22 may be partially relaxed and the results of Proposition 23 still hold. For example if the population is split into several large groups, and voters from each satisfy these conditions, the resulting distribution will still approximate normal. And the symmetry between outcomes may be dropped so long as the covariance matrix of ν is positive definite (covariance matrices are always positive semidefinite, and the requirement that it be positive definite roughly speaking excludes cases where the distribution lies in a proper subspace; this is fine in the generic case). The proofs are fairly straightforward but a little more technical, and we omit them here.

References

[1] K.J. Arrow. A difficulty in the concept of social welfare. *The Journal of Political Economy*, 58(4):328–346, 1950.

[2] K.C. Border. More on harsanyi's utilitarian cardinal welfare theorem. *Social Choice and Welfare*, 1(4):279–281, 1985.

[3] D.S. Felsenthal and M. Machover. *The measurement of voting power: theory and practice, problems and paradoxes.* Edward Elgar Cheltenham, UK, 1998.

[4] A. Gibbard. Manipulation of voting schemes: a general result. *Econometrica: Journal of the Econometric Society*, pages 587–601, 1973.

[5] J.C. Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *The Journal of Political Economy*, 63(4):309–321, 1955.

[6] D.M. Hausman. The impossibility of interpersonal utility comparisons. *Mind*, 104(415):473–490, 1995.

[7] O. Morgenstern and J. Von Neumann. Theory of games and economic behavior. 1953.

[8] W.H. Riker and P.C. Ordeshook. A theory of the calculus of voting. *The American Political Science Review*, 62(1):25–42, 1968.

[9] A.K. Sen. Collective choice and social welfare. 1970.