The Middle Opinion. USA 2020.

Chapter 14a (post-election draft, 6th November)[1]

The Majority Problem and Central Optimality

Note: written on 11th December 2019

Overview

The majority criterion is commonly used in social situations to decide which option to choose. However in certain circumstances the majority option is not the best option and it is this that we refer to as majority suboptimality. Note that here the notion of 'best' depends on a specific definition of value.

In certain circumstances, a central option is the unique optimum and so every non-central option is suboptimal – in particular this applies to any non-central option favoured by a majority. In other words, majorities which do not embrace the central option – what might be referred to as one-sided majorities - are suboptimal.

Possible conceptions of the centre are: mode, median, mean and midpoint. A possible criterion for optimality is maximum value and minimum value deviation/polarisation.

First though we note Peter Emerson's work on the majority problem and also the ways in which the majority problem has manifested itself in the Brexit process.

The distribution of political values is a foundational concept. The notion here of a centre implies the distribution of values in a political space. The Borda Count and Condorcet criterion make no reference to a political space and are thus more general criteria with more general direct applicability. My hypothesis is that often a political space can be identified and thus make a notion of centre relevant.

Peter Emerson and the majority problem

Peter Emerson has long argued against the use of the majority criterion. He has seen the majority problem in many countries. Only now with Brexit do I fully appreciate

[1] Gordon Burt: gordonjburt@gmail.com. Website: https://sites.google.com/site/gordonburtmathsocsci/all.

October 22, 2020

the majority problem. Peter and I have agreed and disagreed over many years but now, throughout this Brexit year, repeatedly I have thought, “yes, this is a core idea which I recall Peter enunciating many years ago”. Thank you, Peter.

Emerson, Peter. *Majority Voting as a Catalyst of Populism. Preferential Decision-making for an Inclusive Democracy.* Springer: 2019.

This timely book presents a critique of binary majority rule and provides insights into why, in many instances, the outcome of a two-option ballot does not accurately reflect the will of the people. Based on the author's first-hand experience, majority-voting is argued to be a catalyst of populism and its divisive outcomes have prompted countless disputes throughout Europe and Asia. In like manner, simple majority rule is seen as a cause of conflict in war zones, and of dysfunction in so-called stable democracies. In order to safeguard democracy, an all-party power-sharing approach is proposed, which would make populism less attractive to voters and governments alike. In geographically arranged chapters, well-tested alternative voting procedures (e. g. non-majoritarian Modified Borda Count) are presented in case studies of Northern Ireland, Central Europe, the Balkans, the Caucasus, Russia, China, North Korea and Mongolia.

Brexit: the majority problems

Throughout the Brexit process, majorities have not behaved the way people have wanted them to. Was David Cameron perhaps over-confident, seeing himself as the winner of majorities: “if I can win the Scottish referendum then I can win the EU referendum”? There was a discrepancy between the majority in the opinion polls and the majority in the referendum itself. The referendum produced only a narrow majority allowing the losers to challenge the validity and hope for a change. In fact discrepancy and narrowness have continued throughout the period since the referendum. The referendum majority was for a non-specific option, allowing argument as to what specific option enjoyed a majority. In the past year there has been the failure of Mrs May’s deal to gain a majority in parliament – on three separate occasions. The indicative votes in parliament were taken to mean that there was no majority for anything.

Now at last Mr Johnson’s deal has won a majority in parliament for a second reading (but not a majority for the government’s proposed rushed timetable). It might seem then that Mr Johnson has succeeded where Mrs May had failed. An alternative view sees both occasions as being characterised by ‘contradictory majorities’: Mrs May’s deal was defeated by a contradictory majority, a coalition of two opposing extremes; the second reading of the bill for Mr Johnson’s deal was approved by a contradictory majority, a coalition of those who genuinely approved the deal and those who wanted to proceed to the second reading in order to replace it entirely or to amend it in major ways.

A rather different majority problem is addressed in what follows: the possibility that a majority option is suboptimal.

Note that suboptimality is not the only problem with the majority criterion. In that there are many possible majorities, the criterion is indeterminate. In that there may be no majorities, the criterion is indecisive.

Majority suboptimality … central optimality

Here we are concerned with a special case. The set of options is assumed to form an option space and value is defined in terms of distance in the option space. The majority criterion is assumed to select the majority option, namely the mode of the opinion distribution in option space.

An example is the first stage of the Conservative party leadership election: Boris Johnson was the mode, winning more votes than any other candidate, but was the extreme candidate in Brexit space and was not the choice of the mean voter (nor indeed of the median voter).

Adopting certain definitions and making certain assumptions it can be proved mathematically that the majority criterion is suboptimal. The argument involves bringing together two topics, the first is optimal social design and the second is statistical democracy.

Peter Emerson has long argued against the use of the majority criterion. Here we consider one particular argument against the majority criterion:

> Value declines with the distance from the ideal. Social mean value declines with the mean distance from the social mean ideal. Majorities which do not embrace the mean ideal are suboptimal. In particular this is the case for one-sided majorities.

We provide a model which measures the amount by which the majority option is suboptimal in terms of failing to maximise value and failing to minimise distance. This result is located within a framework which might be referred to as statistical democracy. Key concepts are: option set, opinion space, opinion distribution, opinion distribution parameters.

Opinion polls should be designed and media articles should be written to illuminate public opinion from a statistical democracy perspective. In particular they should enable the best option to be identified. We might refer to all this as location politics – prompting the recollection that Walter Isard was founder of both location economics and of peace science.

Optimal social design

Note: in what follows the set of options is assumed to form an option space and value is defined in a particular way.

There is a relationship between value, loss and distance. Value is maximised when mean squared distance (loss) is minimised. This occurs when the reference point is the mean of the ideal points.

Suppose an individual has an ideal point x* which has a value v* for the individual. Suppose the value of any other point x is $v(x)=v^*-k(x-x^*)^2$. (This is sometimes referred to as a quadratic loss function.) Suppose individuals have the same form of value function with the same k but with different v* and different x*. Then the mean value V(r) of some reference point r for the population is:

.	$$V(r)=V^*-kMSD(r)$$

where V* is the mean of the ideal points and MSD(r) is the mean square distance in relation to the reference point r.

The mean squared distance from the reference point r is given by:

.	$$MSD(r) = MSD(\mu) + (\mu-r)^2$$

The mean squared distance is minimised when the reference point is the mean μ of the ideal points.

Result
The mean squared distance from the reference point r occurs when $r=\mu$. In other words the mean is the reference point which minimises the mean squared distance.

Other reference points are suboptimal in the sense that they have larger mean squared distance. In particular the mode is suboptimal in this sense (except where the mode and mean coincide).

Opinion distribution

Consider a set of individuals - a population – and a set of options. Suppose each individual values the set of options in some way, either placing a value on each or having a set of preferences between the different options. Of particular interest is an individual's ideal (the most valued option) or first preference option. We refer to the individual as having a value function over the set of options and the population as having a set of value functions over the set of options. A social choice criterion is some rule which selects one of the options as the social choice, dependent on the set of value functions.

An important feature of society is the population distribution of opinion over the option set. What happens in society – and what should happen – is affected by the distribution of opinion. Parameters of the distribution can be used as criteria for selecting options …

A common method in social decision making is majority voting, selecting the option with the most votes – in statistics this is referred to as the mode of the distribution. The mode can be defined for any set of options.

A special situation is when the set of options can be thought of as forming an option space. This occurs if it is possible to define an ordering on the set of options or a distance between each pair of options.

If the set of options form an option space, then it is possible to define other parameters such as the median or mean.
It may be that other parameters such as the median or mean may be a better criterion in that they reduce spread – which corresponds to the notion of polarisation in social settings.

To the three averages already mentioned I would add the midpoint, a very crude indicator indeed. A corresponding very crude measure of spread/polarisation is the maximum distance, the distance of the most distant individual from the 'average'. The distance minimax criterion selects the option where this maximum distance is a minimum. The minimax criterion is satisfied by the midpoint, by definition. The minimax distance is somewhat akin to one aspect of Rawls' theory of justice.[2]

Distribution parameters

Here we assume an option space. We are interested in distributions over a finite one-dimensional interval. These can always be transformed into distributions over the unit interval and so this is how we shall approach the topic. Rather than the sophisticated distributions usually studied in statistical theory we shall study some rudimentary distributions that nevertheless capture certain key ideas.

Figure 1 A finite one-dimensional interval: the unit interval

0 1

We are interested in a number of the parameters of distributions, partly in their own right but also as potential criteria for choosing between options. One kind of parameter are averages or measures of central tendency. The three commonly discussed are mode, median and mean. The median is sometimes between the mode and the mean, but this is not always the case.

To these three averages I would add the midpoint, a very crude indicator indeed. As long as 0 and 1 have non-zero probability density then the midpoint is 0.5 for every distribution.

As well as measures of central tendency there are measures of spread, such as mean deviation from the mean/median and standard deviation. A very crude measure of spread is the distance of the most distant individual from the 'average'.

[2] "Social and economic inequalities are to be arranged so that they are **(a)** to the greatest benefit of the least advantaged members of society, consistent with the just savings principle."
Rawls, John (1971). *A theory of justice*. p. 266. ISBN 0674000781. OCLC 41266156
https://en.wikipedia.org/wiki/A_Theory_of_Justice

The uniform distribution

Consider the uniform distribution over the interval [0,1]. For x in the interval, the probability density function is $f(x)=1$.

The median, mean and midpoint are all equal to 0.5.

For every x, $f(x)=1$ and so every point x is a mode.

Consider now measures of spread from some reference point r. Turning to measures of spread, the mean deviation from the median is 0.25.

The distance between any point x and some reference point r is $|(x-r)|$. The maximum distance from any reference point r is $\max\{r,1-r\}$. The minimax distance criterion selects the reference point $r=0.5$. The maximum distance of points from the reference point is $d=0.5$ which occurs for points $x=0$ and $x=1$.

Figure 2 The midpoint 0.5 minimises the maximum distance, d=0.5

.   :<u>   d=0.5   :   d=0.5   :</u>
.   0   0.5   1

Symmetric distributions

Consider symmetric distributions on the interval [0,1]. The median, mean and midpoint are all equal to 0.5. If the distribution is unimodal then the mode is also at 0.5.

Many measures of spread are symmetric in relation to the average. These measures of spread have a minimum at $x=0.5$.

Unimodal … triangular distributions

An important class of distributions are unimodal distributions. Here we focus on a simple type of unimodal distributions, namely the triangular distributions.

Result 1
Consider a triangular distribution over the interval [0,1] with peak at a. The midpoint is 0.5; the mode is a; the median is $\sqrt{(a/2)}$ (if $a\geq 0.5$) and $1-\sqrt{((1-a)/2)}$ (if $a\leq 0.5$); and the mean is $1/3+a/3$.

Consider the set of all triangular distributions. Each distribution has a peak at a. For each distribution, in other words for each peak a, we can identify the midpoint, the mean, the median and the mode. Figure 1 below shows what the midpoints, the means, the medians and the modes are for different distributions, different peaks a.

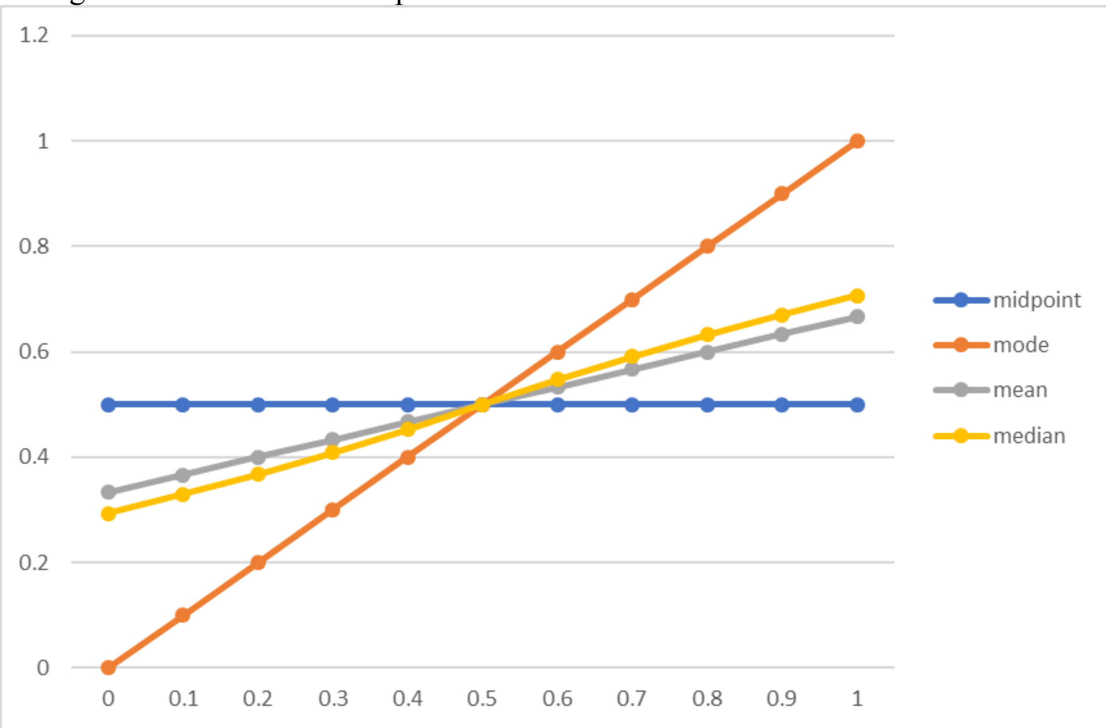
The midpoint is the same for all distributions. The mode varies a large amount depending on the distribution. The median varies much less depending on the distribution. Similarly for the mean which is close to the median. Hence:

Result 2

Considering the set of triangular distributions, the range of the mean is 1/3=0.33 to 2/3=0.67; and the range of the median is 0.3 to 0.7 approx. Both these ranges are less than the range of the mode which is 0 to 1. The midpoint is constant at 0.5.

Figure 3 Different triangular distributions with different peaks a; midpoint, mean, median and mode for different values of peak a

average for different values of peak a



. value of peak a

Proof of Result 1

Proof for the mean:

The triangle consists of two parts: the left-hand triangle with its peak at a; and the right-hand triangle with its peak at a. The bases are a and b=(1-a) respectively. So the areas are a and b respectively.

The means are 2a/3 and 1-2b/3, respectively.

The overall mean is the weighted sum of the means for the two component right-angled triangles, the weights being the areas.

$=a(2a/3)$ and $b(1-2b/3)=b+(2/3)(a^2-b^2)=b+(2/3)(a-b)=(1-a)+(2/3)(2a-1)=1/3+a/3$.

Proof of "The means are 2a/3 and 1-2b/3, respectively."

Mean=E(x)=xf(x)…

… for left triangle $=x(h/a)x=(h/a)[x^3]_0^a=(2/a^2)(a^3/3)=2a/3$

Proof for the median:

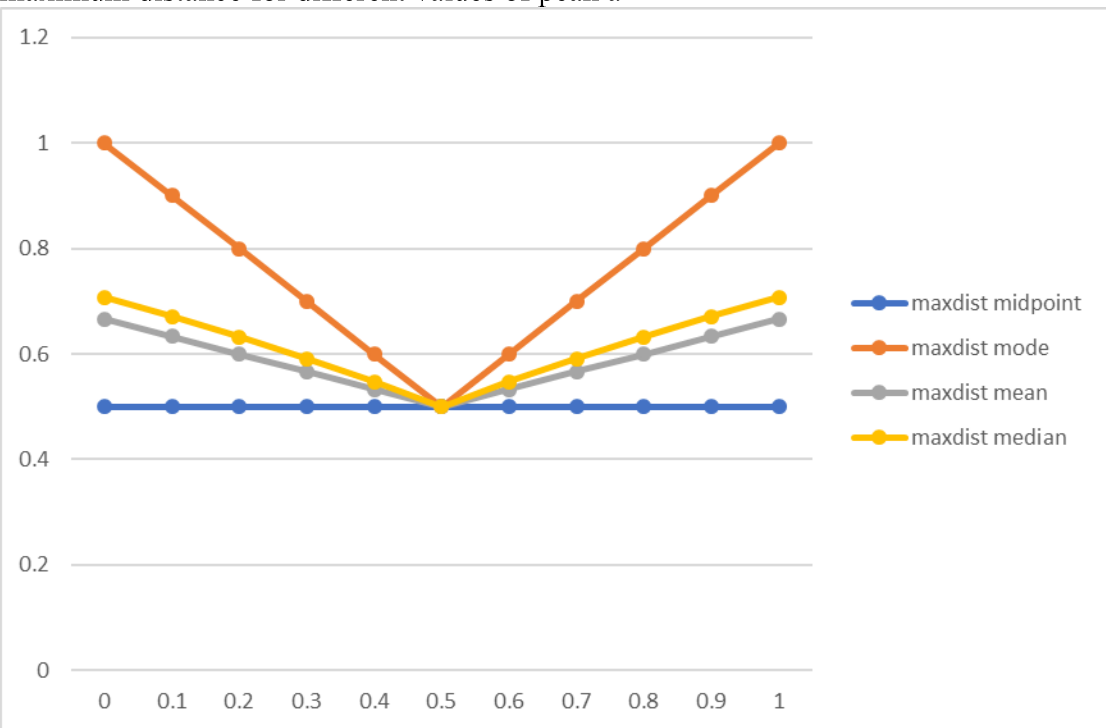
Consider a triangular distribution over the interval [0,1]. The peak is at x=a where 0≤a≤1. Let b=(1-a).

To be a probability distribution the area under the triangle must be 1. So the height h of the peak at a is $\frac{1}{2}(1)(h)=h/2=1$. So $h=2$.
The distribution is specified by:
$f(x)=ux$, $0\leq x\leq a$
$f(x)=v(1-x)$, $a\leq x\leq 1$
$f(a)=2$
So $u=2/a$; and $v=2/(1-a)=2/b$.
The median is at the point that divides the area/probability 50%:50%.
For $x\leq a$ the area up to point x is $p=\frac{1}{2}x(2x/a)=x^2/a$. For $p=1/2$, $x=\sqrt{(a/2)}$. The median is $\sqrt{(a/2)}$.
[$a=1/2$ median is $\frac{1}{2}$; $a=1$ median is $\sqrt{(1/2)}$, approximately 0.7].
For $x\leq a$, the median is $1-\sqrt{(b/2)}=1-\sqrt{(1-a)/2}$
[$b=a=1/2$ median is $\frac{1}{2}$; $b=1$, $a=0$ median is $1-\sqrt{(1/2)}$, approximately 0.3].

A very simple measure of spread is the maximum distance from the average. Figure 2 shows that this measure is lower in the middle, least for x=0.5 and greatest at the two extremes. The maximum distance from the midpoint is the same for all distributions. The maximum distance from the mode varies greatly for different distributions. The maximum distance from the median varies moderately for different distributions. Similarly for the mean. See Figure.

Figure 4 Different triangular distributions with different peaks a; maxdistance from midpoint, mean, median and mode for different values of peak a

maximum distance for different values of peak a



. value of peak a

We now consider in more detail the case of the mean. The measure of spread usually associated with the mean μ is the variance σ^2. This defined as follows, with E as the expectation:

$\sigma^2 = E(x-\mu)^2 = E(x^2) - \mu^2$

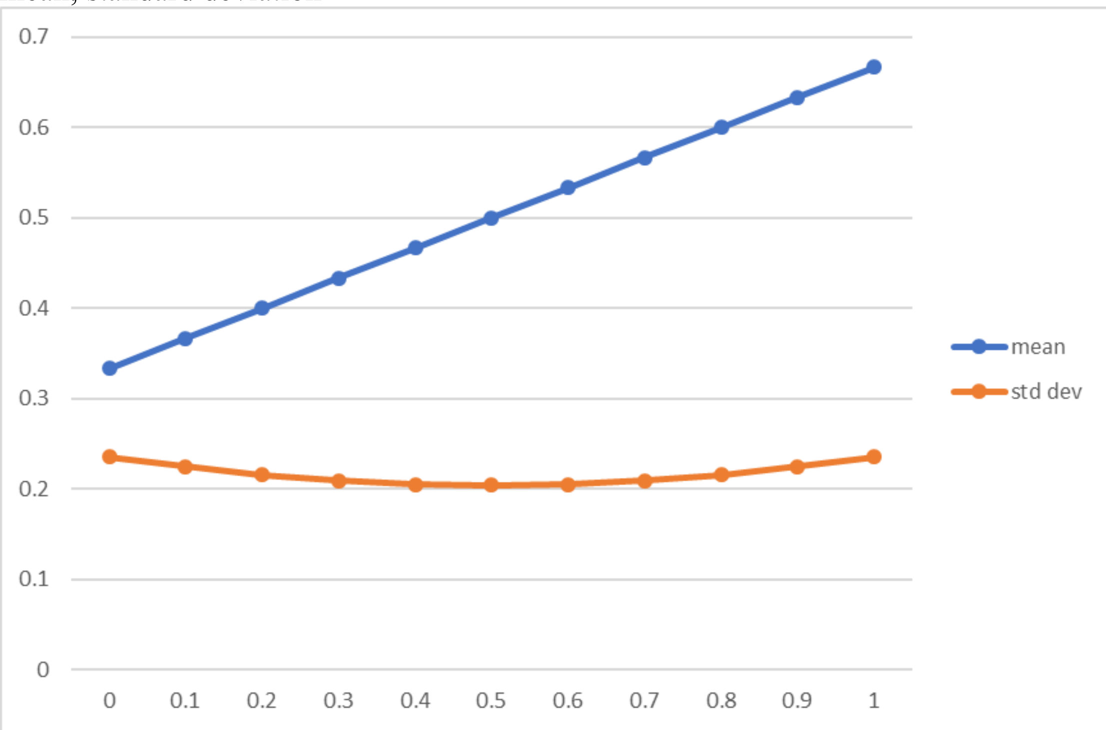
Result

For a triangular distribution with peak a,

$\sigma^2 = (1+a+a^2)/6 - \mu^2$ where $\mu = (1+a)/3$.

Figure 5 The mean and standard deviation σ for different distributions with different peaks a

mean; standard deviation



. value of peak a

We may think of the variance as the mean squared distance from the mean. Consider some other reference point r, rather than the mean. What is the mean squared distance from the reference point r?

Result

The mean squared distance from the reference point r is $\sigma(r)^2=\sigma^2+(\mu-r)^2$.

Proof:

$\sigma(r)^2 = E(x-r)^2 = E(x-r+\mu-\mu)^2$

$= E(x-\mu)^2 + 2(\mu-r)E(x-\mu) + E(\mu-r)^2$

Result

The mean squared distance from the reference point r occurs when $r=\mu$. In other words the mean is the reference point which minimises the mean squared distance.

Other reference points are suboptimal in the sense that they have larger mean squared distance. In particular the mode is suboptimal in this sense.

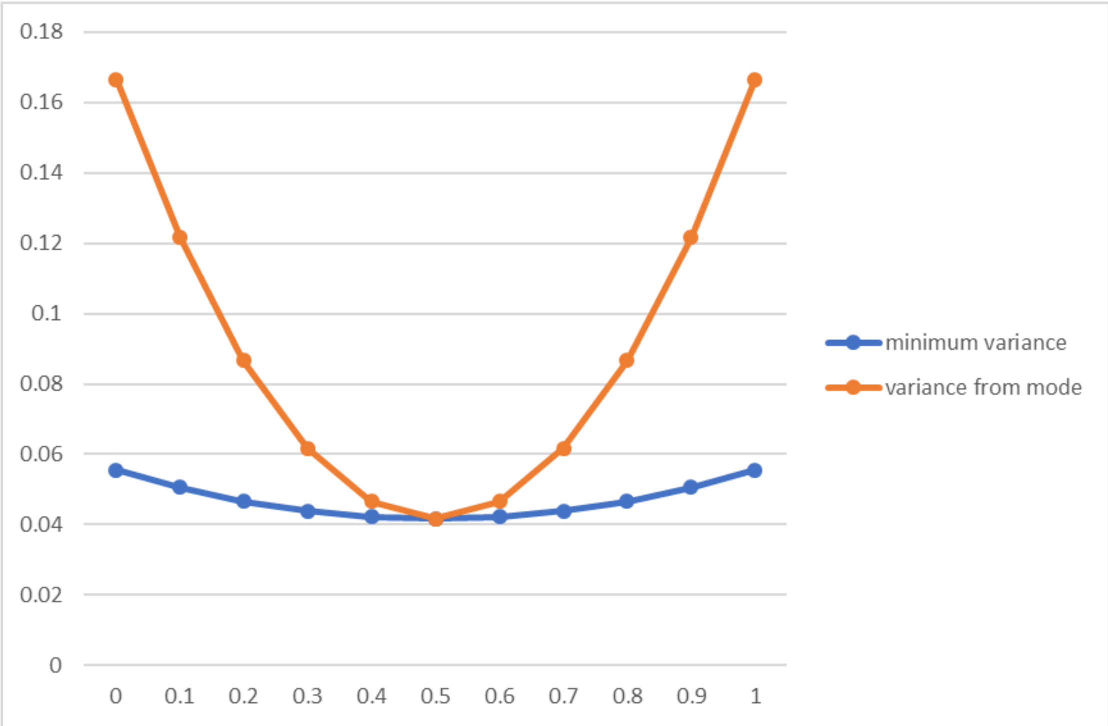
Result

The mean squared distance from the reference point r is $\sigma(r)^2=\sigma^2+(\mu-r)^2$.

Adopting certain definitions and making certain assumptions it can be proved mathematically that the majority criterion is suboptimal.

Figure 6 The mean and standard deviation σ for different distributions with different peaks a

variance



. value of peak a