Public Choice **113:** 357–365, 2002.

Galton's two papers on voting as robust estimation *

DAVID M. LEVY & SANDRA PEART
[1]*Center for Study of Public Choice, George Mason University, Fairfax VA 22031; e-mail: DavidMLevy@aol.com;* [2]*Department of Economics, Baldwin-Wallace College, Berea OH 44017, U.S.A.; e-mail: speart@bw.edu*

Accepted 1 June 2001

Abstract. The relationship between voting and robust estimation was discussed by Francis Galton in 1907. His two papers in *Nature* are discussed and reprinted.

1. Introduction

Why has it taken so long for theorists of politics to see the relationship between voting and estimation?[1] This good question was asked by Bassett and Persky (1999: 299) when they took the concept of robustness from statistical theory and applied it usefully to voting procedures.[2] Their problem is even more complicated than they propose. The relationship between voting and robust estimation was explained carefully in two 1907 contributions to *Nature* by Galton.

It is easy to believe that important contributions can be overlooked when their author is outside the research community or they are published in obscure journals.[3] But two 1907 articles by Galton in *Nature*?[4] How is it possible for a contribution to be more centrally located in a discussion? Karl Pearson who describes them carefully suggests that Galton published them in *Nature* for "immediate attention ... at the cost of later oblivion."[5] Is the difficulty here that Galton is simply sixty years ahead of the statistical literature?[6] He evidently uses a technique – an influence curve – which not become commonplace until the 1970s to make his case against the sample mean and in favor of the sample median.[7]

* We have benefitted from conversation and correspondence with Gib Bassett, Roger Congleton, Stephen Stigler and thank them and James Buchanan for their enthusiastic support. Nicola Tynan found several errors.

2. Why a median?

Galton considers a group deciding upon an amount where instead of *preferences*, each member is supposed to possess an *estimate* of the true amount. What estimate is appropriate for the group? Surely not the sample mean:

> That conclusion is clearly *not* the *average* of all the estimates, which would give a voting power to "cranks" in proportion to their crankiness. One absurdly large or small estimate would leave a greater impress on the result than one of reasonable amount, and the more an estimate diverges from the bulk of the rest, the more influence would it exert.[8]

In modern terms, even though the sample mean has ideal properties at normality because it has an unbounded influence curve – the influence of "cranks" is "in proportion to their crankiness" – it is far too dangerous to employ for serious purposes, i.e., those questions involving real money. Consider on the contrary the nice properties of the sample median. Majority rule is the median estimate:

> I wish to point out that the estimate to which least objection can be raised is the *middlemost* estimate, the number of votes that it is too high being exactly balanced by the number of votes that it is too low. Every other estimate is condemned by a majority of voters as being either too high or too low, the middlemost alone escaping this condemnation.

In the next contribution Galton introduces a device which Stigler (1977) would make famous: judging the properties of an estimator by how it works with real data. While Stigler worked within an austere context – estimators' ability to recover the true parameters of the physical world from experimental data – Galton used a rather homey example – how the median guess of 787 contestants paying a 6d entrance fee recovered the weight of an ox. The median guess was 0.8% high. Galton finds the distribution of guesses strikingly abnormal.

In his later *Memories* (1908: 281) this founder of eugenics admits to an egalitarian conclusion arising from these papers:

> The result seems more creditable to the trustworthiness of a democratic judgment than might have been expected. But the proportion of the voters who were practised in judging weights undoubtedly surpassed that of the voters in ordinary elections who are versed in politics.[9]

> I endeavoured in the memoirs just mentioned, to show the appropriateness of utilising the *Median* vote in Councils and in Juries, whenever they have to consider money questions. Each juryman has his own view of

> what the sum should be. I will suppose each of them to be written down. The best interpretation of their collective view is to my mind *certainly not* the average, because the wider the deviation of an individual member from the average, of the rest, the more largely would it effect the result. In short, unwisdom is given greater weight than wisdom. In all cases in which one vote is supposed to have one value, the median value *must* be the truest representative of the whole, because any other value would be negatived if put to the vote.

3. "Immediate attention"

Galton was 85 when he wrote these papers. Suppose that Pearson was right about the consequence of publishing in *Nature* and further supposing that Galton himself recognized the consequence, what would require immediate attention? First, the title "Vox Populi" recalls Galton's Carlylean doctrine of the foolishness of majority rule democracy. This is Galton from 1872:

> I propose, in these pages, to discuss a curious and apparently anomalous group of base moral instincts and intellectual deficiencies, to trace their analogies in the world of brutes, and to examine the conditions, through which they have been evolved. I speak of the slavish aptitudes, from the leaders of men, and the heroes and the prophets, are exempt, but which are irrepressible elements in the disposition of average men. I refer to the natural tendency of the vast majority of our race to shrink from the responsibility of standing and acting alone, to their exaltation of the *vox populi*, even when they know it to be the utterance of a mob of nobodies, into the *vox Dei*, to their willing servitude to tradition, authority and custom. Quoted in Pearson (1924: 72).

The immediate context seems to be Galton's reflection upon just how eugenic policy might be made in a non-hierarchical society:

> Society would he very dull if every man resembled the highly estimable Marcus Aurelius or Adam Bede. The aim of eugenics is to represent each class or sect by its best specimens; that done, to leave them to work out their common civilization in the own way.
> A considerable list of qualities can easily be compiled that nearly everyone except "cranks" would take into account when picking out the best specimens of his class.[10]

Is that which required "immediate attention" a recantation of Galton's former views? If so it is harder to think of a more wonderful display of intellectual integrity.

Without further ado, Galton from 1907.

Notes

1. The median voter theory proposed in Downs (1957) is supposed to apply to a *distribution* of voters. The median is a parameter – the population median – not an estimate of a parameter. A search via JSTOR on Downs & median voter & Galton yield nothing.
2. The reader of Bassett and Persky (1999) might get the impression that they wrote in knowledge of Levy (1989). They did not. Their results were completely independent.
3. Stigler (1973) appeals to these facts to explain plausibly why Newcomb's and Daniell's contributions to robust estimation were neglected.
4. Galton (1908: 280–281) gives an account of these contributions in non-technical terms.
5. Pearson (1924: 400): "*The Median*. There are a number of short papers by Galton which are, perhaps, most suitably dealt with in this chapter. A good many of them appeared in the pages of *Nature*, a ready means of attracting immediate attention, but too often at the cost of later oblivion. Several of these papers concern really important points, which have, since their publication, been again and again overlooked." An 1890 *Nature* article "Dice for Statistical Experiments" is reprinted by Stigler (1999: 152–155) along with extensive commentary.
6. Stigler (1999: 151): "Francis Galton (1822–1911) was such a fertile source of statistical ideas over his long life that it should not cause surprise that he contributed to simulation as well."
7. Influence curves play a critical role in the motivation of Andrews et al. (1972).
8. What perhaps compounds the difficulties of seeing the point is that Galton has somehow obtained reputation of not deviating from the assumption of normality. Porter (1986: 139): "Galton remained one of the most loyal partisans of the error law throughout his life. Even though he was among the first to propose an alternative distribution, the so-called log-normal, in conjunction with a certain class of data, that formula involved no rejection of the conventional error law." Porter does not respond to pre-emptive counterexamples to this claim for such contexts as would justify the use of sample median, Stigler (1973: 875–876). Galton (1889: 409) uses the fact of outliers to argue against use of the range as an estimator of scale – "The difference between the extreme ends of a marshaled series is no proper measure of the variety of the men who compose it. However few may be the objects in the series, it is always possible that a giant or a dwarf, so to speak, may be included among them. The presence of either would mislead as to the range of variety likely to be found in another equally numerous sample taken from the same group." His argument in favor of the *median deviation*, instead of the mean deviation, is found on the following page. This seems clear recognition of the importance of outliers in selecting a sensible estimator of scale.
9. Might one – without the slightest disrespect – point out that there will also be more voters in elections than in ox judgment? Robin Hanson helped here in another context.
10. Galton (1904: 2). This is the only paper in JSTOR written by Galton which contains the word "crank." The search was conducted May 26, 2000.

References

Andrews, D.F, Bickel, P.J., Hanpel, F.R., Huber, P.J., Rogers, W.H. and Tukey, J.W. (1972). *Robust estimates of location*. Princeton: Princeton University Press.

Bassett, G. Jr. and Persky, R. (1999). Robust voting. *Public Choice* 99: 299–310.

Downs, A. (1957). *An economic theory of democracy*. New York: Harper & Row.

Galton, F. (1889). Address delivered at the anniversary meeting of the Anthropological Institute of Great Britain and Ireland, January 22nd, 1889. (1889). *Journal of the Anthropological Institute of Great Britain and Ireland* 18: 401–419.

Galton, F. (1904). Eugenics: Its definition, scope, and aims. *American Journal of Sociology* 10: 1–25.

Galton, F. (1907a). One vote, one value. *Nature* 75: 414.

Galton, F. (1907b). *Vox Populi*. *Nature* 75: 450–451.

Galton, F. (1908). *Memories of my life*. London: Methuen & Co.

Levy, D.M. (1989). The statistical basis of Athenian-American constitutional theory. *Journal of Legal Studies* 18: 79–103.

Pearson, K. (1924). *The life, labours and letters of Francis Galton*. Cambridge: Cambridge University Press.

Porter, T.M. (1986). *The rise of statistical thinking: 1820–1900*. Princeton: Princeton University Press.

Stigler, S.M. (1973). Simon Newcomb, Percy Daniell, and the history of robust estimation 1885–1920. *Journal of the American Statistical Association* 68: 872–879.

Stigler, S.M. (1977). Do robust estimators work with real data? *Annals of Statistics* 5: 1055–1098.

Stigler, S.M. (1999). *Statistics on the table: The history of statistical concepts and methods*. Cambridge: Harvard University Press.

One vote, one value

A certain class of problems do not as yet appear to be solved according to scientific rules, though they are of much importance and of frequent recurrence. Two examples will suffice. (1) A jury has to assess damages. (2) The council of a society has to fix on a sum of money, suitable for some particular purpose. Each voter, whether of the jury or of the council, has equal authority with each of his colleagues. How can the right conclusion be reached, considering that there may be as many different estimates as there are members? That conclusion is clearly *not* the *average* of all the estimates, which would give a voting power to "cranks" in proportion to their crankiness. One absurdly large or small estimate would leave a greater impress on the result than one of reasonable amount, and the more an estimate diverges from the bulk of the rest, the more influence would it exert. I wish to point out that the estimate to which least objection can be raised is the *middlemost* estimate, the number of votes that it is too high being exactly balanced by the number of votes that it is too low. Every other estimate is condemned by a majority of voters as being either too high or too low, the middlemost alone escaping this condemnation. The number of voters may be odd or even. If odd, there is one middlemost value: thus in 11 votes the middlemost is the 6th; in 99 votes the middlemost is the 50th. If the number of voters be even, there are two middlemost values, the mean of which must be taken; thus in 12 votes the middlemost lies between the 6th and the 7th; in 100 votes between the 50th and the 51st. Generally, in $2n$–1 votes the middlemost is the nth; in $2n$ votes it lies between the nth and the (n+1)th.

I suggest that the process for a jury on their retirement should be (1) to discuss and interchange views; (2) for each juryman to write his own independent estimate on a separate slip of paper; (3) for the foreman to arrange the slips in the order of the values written on them; (4) to take the average of the 6th and 7th as the verdict, which might be finally approved as a substantive proposition. Similarly as regards the resolutions of councils, having regard to the above ($2n$–1) and $2n$ remarks.

Francis Galton

Vox populi

In these democratic days, any investigation into the trustworthiness and peculiarities of popular judgments is of interest. The material about to be discussed refers to a small matter, but is much to the point.

A weight-judging competition was carried on at the annual show of the West of England Fat Stock and Poultry Exhibition recently held at Plymouth. A fat ox having been selected, competitors bought stamped and numbered cards, for 6*d*, each, on which to inscribe their respective names, addresses, and estimates of what the ox would weigh after it had been slaughtered and "dressed." Those who guessed most successfully received prizes. About 800 tickets were issued, which were kindly lent me for examination after they had fulfilled their immediate purpose. These afforded excellent material. The judgments were unbiassed by passion and uninfluenced by oratory and the like. The sixpenny fee deterred practical joking, and the hope of a prize and the joy of competition promoted each competitor to do his best. The competitors included butchers and farmers, some of whom were highly expert in judging the weight of cattle; others were probably guided by such information as they might pick up, and by their own fancies. The average competitor was probably as well fitted for making a just estimate of the dressed weight of the ox, as an average voter is of judging the merits of most political issues on which he votes, and the variety among the voters to judge justly was probably much the same in either case.

After weeding thirteen cards out of the collection, as being defective or illegible, there remained 787 for discussion. I arrayed [see Table 1] them in order of the magnitudes of the estimates, and converted the *cwt., quartors*, and *lbs*. in which they were made, into lbs., under which form they will be treated.

According to the democratic principle of "one vote one value," the middlemost estimate expresses the *vox populi*, every other estimate being condemned as too low or too high by a majority of the voters (for fuller explanation see "One vote, one value," *Nature*, February 28, p. 414). Now the middlemost estimate is 1207 lb., and the weight of the dressed ox proved to be 1198 lb.; so the *vox populi* was in this case 9 lb., or 0.8 per cent., of the whole weight too high. The distribution of the estimates about their middlemost value was of the usual type, so far that they clustered closely in its neighborhood and because rapidly more sparse as the distance from it increased [Diagram 1 found on p. 365]. But they were not scattered symmetrically. One quarter of them deviated more than 45 lb. above the middlemost (3.7 per cent.), and another quarter deviated more than 29 lb., below it (2.4 per cent.), therefore the range of the two middle quarters, that is, of the middlemost half, lay within those limits. It would be an equal chance that the estimate written

Table 1. Distribution of the estimates of the dressed weight of a particular living ox, made by 787 different persons.

Degree of the length of array 0°–100°	Estimates in lbs.	Centiles: Observed deviates from 1207 lbs.	Centiles: Normal p.e. = 37	Excess of observed over normal
°5	1074	–133	–90	+43
10	1109	–98	–70	+28
15	1126	–81	–57	+24
20	1148	–59	–46	+13
q_1 25	1162	–45	–37	+8
30	1174	–33	–29	+4
35	1181	–26	–21	+5
40	1188	–19	–14	+5
45	1197	–10	–7	+3
m 50	1207	0	0	0
55	1214	+7	+7	0
60	1219	+12	+14	–2
65	1225	+18	+21	–3
70	1230	+23	+29	–6
q_3 75	1236	+29	+37	–8
80	1243	+36	+46	–10
85	1254	+47	+57	–10
90	1267	+52	+70	–18
95	1293	+86	+90	–4

q_1, q_3, the first and third quartiles, stand at 25° and 75° respectively.
m, the median or middlemost value stands at 50°.
The dressed weight proved to be 1198 lbs.

on any card picked at random out of the collection lay within or without those limits. In other words, the "probable error" of a single observation may be reckoned as $\frac{1}{2}$ (45+29), or 37 lb. (3.1 per cent.). Taking this for the p.e. of the normal curve that is best adapted for comparison with the observed values, the results are obtained which appear in above table, and graphically in the diagram [see Diagram 1].

The abnormality of the distribution of the estimates now becomes manifest, and is of this kind. The competitors may be imagined to have erred *normally* in the first instance, and then to have magnified all errors that were negative and to have minified all those that were positive. The lower half of the "observed" curve agrees for a large part of its range with a normal

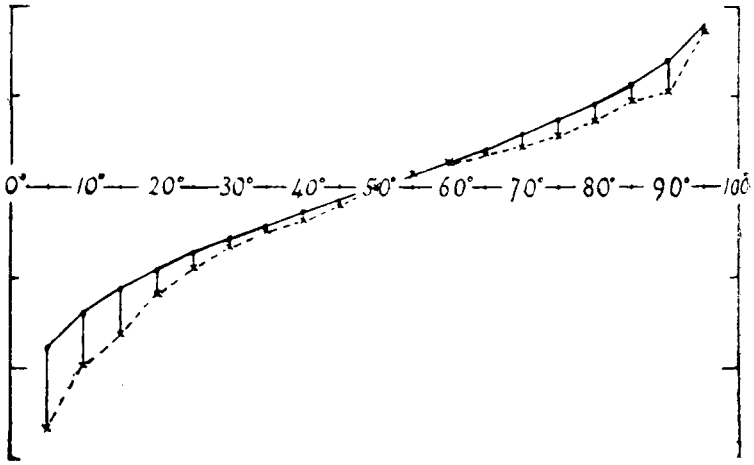

Diagram 1. From the tabular values.
The continuous line is the normal curve with p.e. = 37.
The broken line is drawn from the observations.
The lines connecting them show the differences between the observed and the normal.

curve having the p.e. = 45, and the upper half with one having its p.e. = 29. I have not sufficient knowledge of the mental methods followed by those who judge weights to offer a useful opinion as to the cause of this curious anomaly. It is partly a psychological question, in answering which the various psychophysical investigations of Fechner and others would have to be taken into account. Also the anomaly may be partly due to the use of a *small* variety of different methods, or formulae, so that the estimates are not homogeneous in that respect.

It appears then, in this particular instance, that the *vox populi* is correct to within 1 per cent. of the real value, and that the individual estimates are abnormally distributed in such a way that it is an equal chance whether one of them, selected at random, falls within or without the limits of –3.7 per cent. and +2.4 per cent. of their middlemost value.

This result is, I think, more creditable to the trustworthiness of a democratic judgment than might have been expected.

The authorities of the more important cattle shows might do service to statistics if they made a practice of preserving the sets of cards of this description, that they may obtain on future occasions, and loaned them under proper restrictions, as these have been, for statistical discussion. The fact of the cards being numbered makes it possible to ascertain whether any given set is complete.

Francis Galton