The Tale of Galton's Mean: The Influence of Experts

David M. Levy
Center for Study of Public Choice
George Mason University
DavidMLevy@Gmail.com

Sandra J. Peart
Dean, Jepson School of Leadership Studies
University of Richmond
SPeart@Richmond.edu

4 October 2007

This paper examines the question of expertise and trust in expertise by way of an important historical example, the statistical discussion surrounding measures of central tendency. The broad question we explore, is whom to trust in situations when answers to a problem differ: the well educated expert, or the collective wisdom of the uneducated many? In statistical terms, one might paraphrase this as a question about whether to trust a model (even in the face of contrary evidence) or the aggregated claims of the populous. This, of course, is a debate that touches on hierarchy and egalitarianism. The claim for hierarchy is often grounded on specialization: the expert, being educated, is a better judge of the right answer than others. The egalitarian position counters that, if many people have come to the conclusion that an answer is correct, there must be something to their claim.

Francis Galton entered into this long-standing debate over the wisdom of the educated few versus the collective wisdom of the uneducated many. He did so, with major contributions to both sides of the nineteenth century discussions of expertise and hierarchy in human affairs. First, as is well known, Galton co-founded the "science" of eugenics in which the educated "few" – biologists, economists and sociologists – were to decide which qualities were best for breeding an improved human race (positive eugenics) and which were fit for deprivation and neglect (negative eugenics). In this, he favoured the expertise associated with scientific knowledge, and he foresaw that experts could pick winning evolutionary strategies for the human race.

Notwithstanding this position, however, Galton also revealed an egalitarian side: in a situation in which expertise might not exist or might be difficult to detect (that of guessing the weight of an ox at a county fair), Galton put forward a solution that entailed choosing the median guess. He justified his procedure on the grounds that the proposed solution would place a bound on the influence of any one guess since the guesses of "cranks" are difficult to distinguish from those of experts. In a series of papers that were close to a century ahead of his time, Galton used his statistical insight to explain how a voting mechanism might yield a reasonable estimate of the unknown parameter.

Our examination begins with Galton's egalitarian contribution on voting as estimation as a way to bound influence. But this is only the beginning of our story. Even more interesting, perhaps, than Galton's contribution itself, is the "retelling" that has occurred since. Though Galton defended the median as the estimator for the ox weight, the tale of Galton's median was changed soon after. In addition to the median location of central tendency, Pearson suggested the mean. Along the way in the historical retelling, embellishments were added so that Galton's original procedure was entirely misconstrued. Our story recounts the misconstruing. As we do so, a key question is whether the tale was changed deliberately (falsified) or whether, not knowing the truth, the retold (and different) tale was passed on unwittingly.

The question of expertise enters into this because, if experts pass along false information (wittingly or unwittingly), they become part of a process by which errors are diffused. If experts are trusted (and if they trust the experts whose work they cite), mistakes are diffused and remain uncorrected. If, by contrast, expert results are not accepted automatically, then their results are more often subject to the scrutiny associated with replication. Knowing this, an expert whose work might be scrutinized will check the work carefully. An expert who expects deference, acceptance of results without checking, will check the work less carefully. Paradoxically, such trust increases the probability of mistakes being made or passed along. Thus, we expect that the errors made by experts will tend to be systematic because they will be repeated by those who trust them.

Galton's Median

100 years ago, the pages of *Nature* were graced by a pair of contributions in which Galton imagined a majority-rule democracy as a problem of estimation. First, he established that the sample *median* was the appropriate estimator for a majority-rule decision process (Galton 1907a); second, he tested the properties of this vision by examining the distribution of guesses at an ox judging contest (1907b). To quote from Galton's *Memories of my life*

> So I procured a loan of the cards after the ceremony was past, and worked them out in a memoir published in *Nature* It appeared that in this instance the *vox populi* was correct to within 1 per cent, of the real value; it was 1207 pounds instead of 1198 pounds (1909, p. 281)

By plotting the actual guesses against a hypothetical normal distribution, Galton found striking evidence of non-normality. This supported his statistical intuition that the mean would be a poor measure of central tendency or "interpretation of their collective view":

> 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 affect 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. (1909, p. 282) [Galton's emphasis][1]

Galton's follower on eugenics and other topics, Karl Pearson, discussed the second paper in the series at length.

---

[1] In light of recent work in statistical politics (Levy 1989, Bassett & Persky 1999) the pair of Galton papers were judged worth republishing with some commentary (Levy & Peart 2002 and Peart & Levy 2005, pp. 273-82).

Galton's "Mean" in the Retelling

In neither of these contributions, did Galton report the sample mean. He worked with the median as an estimate of central location and the interquartile range as the estimate of scale. In his commentary, Pearson questioned whether Galton fit the most precise normal curve to the data. In opposition to Galton, Pearson advocated methods of moments, and the sample mean and standard deviation, as estimates for the normal's unknown parameters. Pearson commented on Galton's practice and suggested that the method of moments provided a better estimate from which to construct a normal than Galton's quantiles:

> Unfortunately, the percentile method of tabulation does not permit of very ready determination of the mean and standard deviation and so of getting the best normal distribution. But I find after some labour: mean 1197, standard deviation 61.895 .... the study of popular judgments and their value is an important matter and Galton rightly chose this material to illustrate it. The result, he concludes, is more creditable to the trustworthiness of a democratic judgment than might be expected, and this is more than confirmed, if the material be dealt with by the "average" method, *not* the "middlemost" judgment, the result then being only 1 lb. in 1198 out. (1924, pp. 404-05)

So, in his commentary, Pearson is clear that his average is a better estimate of location than Galton's median. The existence of Pearson's mean was cited in work circulating since 2002 (Levy & Peart, 2004, p. 17; Peart & Levy 2005, p. 99).

Pearson's retelling of the ox judging tale apparently served as a starting point for the 2004 popular account of the modern economics of information aggregation, James Surowieki's *Wisdom of Crowds*. Here is Surowieki's retelling of the tale:

> Galton borrowed the tickets from the organizers and ran a series of statistical tests on them Galton arranged the guesses (which totaled 787 in all, after he had to discard thirteen because they were illegible) in order from highest to lowest and graphed them to see if they would form a bell curve. Then, among other things, he *added all the contestants' estimates, and calculated the mean* of the group's guesses. That number represented, you could say, the collective wisdom of the Plymouth crowd. If the crowd were a single person, that was how much it would have guessed the ox weighed. (2004, p. 3)
>
> The crowd has guessed that the ox, after it had been slaughtered and dressed would weigh *1,197* pounds. After it had been slaughtered and dressed, the ox weighted 1,1198. (2004, p. 3) [emphasis added]

It will now be apparent that this description of what Galton did with the guesses misrepresents what Galton actually did. Galton was clear that the distribution of guesses was *not* normal, writing that "The abnormality of the distribution of the estimates now becomes manifest, ..." (Galton 1907b, p. 451). Surowieki has replaced Galton's statement with the claim that Galton "graphed them [the guesses] to see if they would form a bell curve" – allowing the remaining possibility that the guesses might be normal. Galton's principled *opposition* to the mean as the voice of the people, which Pearson *supplemented* by the use of the mean, is now described as Galton's *use* of the mean. Finally, the reported estimate of the *vox populi* has been changed from 1207 to 1197.

Replicating Galton's Estimation

However the new estimate of location came to be part of Surowieki's account, whoever read Pearson's reworking of Galton's data did not know what to make of what Pearson did not publish. Modern economists have been criticized severely for assuming that distributions are normal without warrant. Particularly edifying in this respect is the criticism offered by Mandelbrot and Hudson

(2004), who take great pains to explain the importance of the kurtosis of a distribution in terms of how far we might reasonably expect observations to fall from center. They document how dramatic the movements of real world asset prices are by reporting the kurtosis of many different stock markets. They explain to their readers that "kurtosis" is taken from the Greek word for shape.

According to the *Oxford English Dictionary*, "kurtosis" was coined by Pearson himself in 1905. Pearson developed a large sample test for normality on the basis of excess kurtosis (Pearson 1905) but he did not report the kurtosis of Galton's ox data.

We need not speculate about the distribution of Galton's data since they have been carefully preserved in the Galton Papers at University College London, (Galton Papers 179/B). Moreover, the work sheets for Pearson's mean and standard deviation are also in the folder. With these in hand, we can recompute the sample median and compare our results with those of Galton.

In our recomputations (using Shazam, White 20006) the sample median (the 394th observation) is 1208, whereas the *Nature* paper reported 1207 (the 393rd observation). Checking to find where the difference occurred reveals that Galton *in fact* computed the median as 1208 (Figure 1) and the manuscript he sent to *Nature* has a median of 1208 (Figure 2). When the article was published in *Nature*, however, it reported a median of 1207.

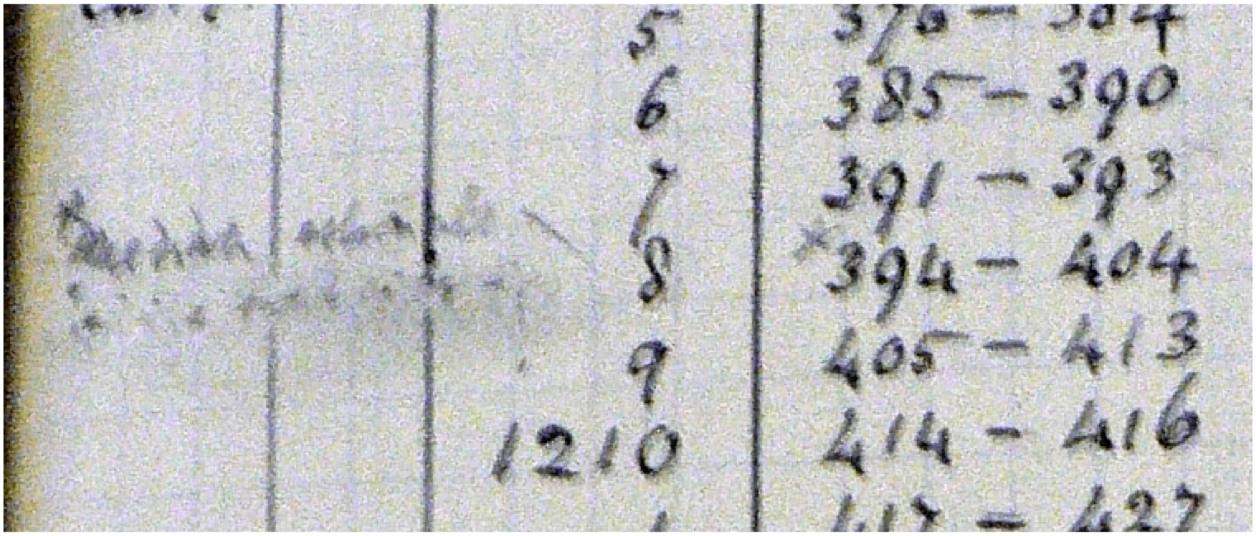

Figure 1. Galton's Work Sheet

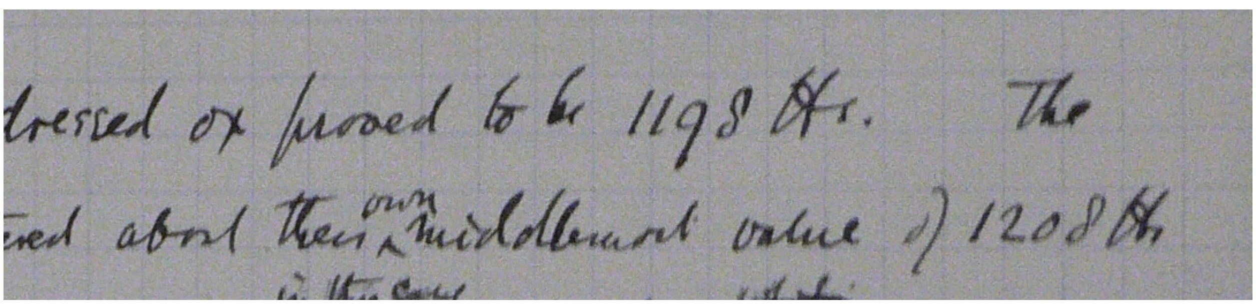

Figure 2. Manuscript for *Nature*

Our computations of quantiles, using *Shazam*10's regression quantile procedure, sometimes yield an observation higher than Galton's report, e.g., he reports the 5th quantile as 1074 (the 39th observation) and we compute 1078 (the 40th observation). Our mean is close to that of Pearson, at 1196.7, while our standard deviation of 73.6 is considerably different from his. The excess kurtosis is computed as 3.04 with a standard deviation of 0.17. The skew is -0.41 with a standard deviation of

0.09. (These are computed as residual diagnostics for a least squares regression.) Normality is decisively rejected by Pearson's tests. Powerful modern tests are equally decisive.

The extraordinary non-normality, higher than what is reported for the price movements of stocks on the Nasdaq (Mandelbrot and Hudson 2004, p. 97), strongly urges against the use of the sample mean. Alternatively, if we use a device unavailable to Galton and make the decision on which estimator has the smaller bootstrap standard error (Léger & Romano, 1990), we find with 10k replications the bootstrap standard error of the sample mean to be 2.6 and that of the sample median 2.2. Galton's choice seems warranted using a late 20th century criterion.

A Paradox of Expertise?

Suppose we distinguish between those who are sufficiently well-known to publish under their own name (experts) and the rest of the population whose contributions are so small as to be unrecognized and anonymous. While experts might be motivated by external goods such as money and praise, the only reward for the non-experts is the internal reward associated with careful work or getting it right. Surowieki's report of "Galton's mean" offers a simple test for how experts and ordinary people alike succeed (or not) in detecting falsehoods.

First, consider the anonymous contributors to *Wikipedia* who reported:

> (Note that although Surowiecki's description implies that Galton used the *mean*, inspection of the original paper indicates that the *median* was used.)

The Wikipedia contributor was unable to discover the source of the mean since it evidently is not in Galton's paper. No matter, the falsehood was detected. Second, how about people who write under their own name and might be expected to be careful? Surowieki's (incorrect) version is now gaining considerable currency, e.g., Sunstein (2006, p. 24), Solomon (2006), Caplan (2007, p. 8). Apparently, none of these experts has verified Surowieki's account. This is understandable. Since Surowieki is considered to be an expert in the area, there would be no perceived need to do so.

There seems to be a paradox of expertise at work here. The expectation of being careful seems to substitute for actually being careful. If it is widely believed that an expert is careful, then the expected marginal external benefit of checking the expert's result might be very small indeed. This suggests that mistakes, once made by experts, will be widely diffused and persistent.

One of Galton's defenses for the sample median as the *vox populi* was it that bounds the influence of any individual voter. Replication and checking of the work of experts may be a way to bound the influence of experts.

Bibliography

Archive

Galton Papers. 179/B. Folder 2. "FG's Working Papers on Livestock Competition." University College London

Published Works

Bassett, Gilbert Jr. and Joseph Persky. 1999. "Robust Voting." *Public Choice* 99: 299-310.

Caplan, Bryan. 2007. *The Myth of the Rational Voter: Why Democracies Choose Bad Politics.* Princeton: Princeton University Press.

Galton, Francis. 1907. "One Vote, One Value." *Nature* 75: 414.
http://galton.org/cgi-bin/searchImages/galton/search/essays/pages/galton-1907-vote-value.htm

Galton, Francis. 1907. "Vox Populi." *Nature* 75: 450-51.
http://galton.org/cgi-bin/searchImages/galton/search/essays/pages/galton-1907-vox-populi_1.htm

Galton, Francis. 1909. *Memories of My Life.* New York: Dutton.
http://galton.org/cgi-bin/searchImages/galton/search/books/memories/pages/memories_0289.htm

Léger, Christian and Joseph P. Romano. 1990. "Bootstrap Adaptive Estimation: The Trimmed-Mean Example." *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* 18 (December): 297-314

Levy, David M. 1989 "The Statistical Basis of Athenian-American Constitutional Theory." *Journal of Legal Studies* 18: 79-103.

Levy, David M. and Sandra J. Peart. 2002. "Galton's Two Papers on Voting as Robust Estimation." *Public Choice* 113: 357-65.

Levy, David M. and Sandra J. Peart. 2004. "Statistical Prejudice: From Eugenics to Immigrants." *European Journal of Political Economy* 20: 5-22.

Mandlebrot, Benoit and Richard L. Hudson. 2004. *The Misbehavior of Markets*. New York: Basic Books.

Pearson, Karl. 1905. "'Das Fehlergesetz und Seine Verallgemeinerungen Durch Fechner and Pearson.' A Rejoinder." *Biometrika* 4: 169-212.

Pearson, Karl. 1924. *The Life, Letters and Labours of Francis Galton: Researches of Middle Life.* Cambridge: Cambridge University Press. http://galton.org/cgi-bin/searchImages/galton/search/pearson/vol2/pages/vol2_0468.htm

Peart, Sandra J. and David M. Levy. 2005. *The "Vanity of the Philosopher": From Equality to Hierarchy in Classical Economics.* Ann Arbor: University of Michigan Press.

Solomon, Miriam. 2006. "Norms of Epistemic" *Diversity Episteme: A Journal of Social Epistemology* 3.1 :23-36

Sunstein, Cass. R. 2006. *Infotopia: How Many Minds Produce Knowledge*. New York: Oxford University Press

Surowiecki, James. 2005. *The Wisdom of Crowds*. New York: Anchor Books.

"The Wisdom of Crowds." *Wikipedia*. http://en.wikipedia.org/wiki/Wisdom_of_crowds

White, Kenneth J. 2006. *Shazam 10: Professional Edition*. Northwest Econometrics.