Los Angeles County Vote Counting Errors: Theory, Conclusions, Fixes

For almost a decade, Judy Alter and Ken Aaron have struggled to collect the data from the 1% manual tally of ballots counted by the Microcomputer Tally System (MTS) of Los Angeles County. They gave me the data for four elections, June and November 2010 and 2012. Brian Dolan did an analysis of the data for 2006. What struck me most was that year-after-year, a few of the precincts – five out of 50 – seemed much more poorly counted than the rest. This essay charts my efforts to rigorously conclude that these outliers are *not* quirks of a random process. The outliers occur because five of the machines are performing much more poorly than the rest. This type of conclusion is not so easy to justify, and you will need to wade through some fairly detailed math to follow the reasoning. However, my conclusions can be stated in concrete terms without reference to the math.

Here are the five conclusions:

(a) 5 out of the 50 machines used in LA County are counting with "off-the-wall" inaccuracy.

(b) The Registrar's office knows which machines count which precincts. That information is required for the procedure by which precincts are assigned for manual tally; we get almost, but not quite, exactly one precinct from each machine. There has been an extra "wrinkle", in that, *for a given precinct*, the ballots cast on Election Day (referred to as Election Night, EN) and the vote by mail ballots (VBM) do not seem to have always been counted on the same machine. In fact, in 2010 and in June 2012, the 5 worst-counted precincts were *not* the same for EN as for VBM, but in November 2012 the 5 worst-counted precincts *were* the same. This indicates that quality control of the machines is possible.

(c) If the contest between two candidates is tighter than the *standard deviation of the error rate*, the vote count should be reevaluated. The standard deviation of the error rate can be substantially reduced by fixing or eliminating the 5 worst machines.

(d) We observe no evidence for changing the vote count to favor one candidate or measure over another. That is, we see almost no over-counts, only dropped counts, and the count-dropping is roughly the same for all candidates and measures. However, the Registrar refuses to inform us which precinct is assigned to which machine, and the geographic/ethnographic identities of the precinct assignments are unknown to us. If this lack of transparency is deliberate, it suggests that the Registrar is capable of skewing the election count by handicapping particular districts with bad machines.

(e) For localized contests, such as the county committees of the political parties, where there are a lot of contestants and few people voting for them, it is desirable to do a hand count because random errors even at a modest rate of 0.5% may affect the outcome.

In the rest of this essay, I try to lay out some of the theory and reasoning that underlie these conclusions. My numerical results for the empirical error rates are similar to Brian Dolan's work for

2006, though our computer programs and emphasis are different. The voting error rate statistics seem to have changed little over the past 6 years.

The recurring and basic question is one that you must have encountered before: How do you extrapolate from a sample of available data to the whole population? The "Central Limit Theorem" says, in effect, that the probability distribution of a random variable *z* follows the Gaussian or normal distribution when the sample size, *N*, is very large. The probability distribution for a sample of smaller size is usually not Gaussian; the shape of the distribution is sensitive to the model for fluctuations in the value of *z*. Different systems have different values of *N* for which the asymptotic Gaussian approximation becomes valid. For non-Gaussian distributions, aberrations from the observed mean are *more probable* than with a Gaussian. That is, deviations from the mean or mode that are highly improbable for a Gaussian distribution are not so improbable for others. Another issue that comes up is this: Suppose some variable, *z*, has a distribution *P*(*z*), but you don't measure *z*, you measure some function of *z*, i.e. *y*(*z*). It turns out that

$$P(y) = P(z)\frac{dz}{dy} .$$

Therefore, even if *P*(*z*) is Gaussian, *P*(*y*) may not be Gaussian.

Most people take the Gaussian bell curve, or *normal* distribution, with its mean μ and standard deviation σ (or variance, σ^2) as God-given. The mean and variance of a collection of data are pretty intuitive to calculate. The mean is the average value of *z*, and the variance is the average square of the difference between *z* and the mean. (There is a fudge factor, but ignore that here.) But the mean and variance of a sample are usually not the same as the mean and variance of the distribution. Only if *N* is "sufficiently large" does the sample's mean equal the distribution's mean, and the sample's variance equal the distribution's variance.

The normal distribution is

$$P(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(z-\mu)^2}{2\sigma^2}} .$$ The *Gaussian* or *normal* distribution

The error function, erf, describes how much of the area under the Gaussian distribution for *z* lies between –*w* and +*w*:

$$\tfrac{1}{2}\,\mathrm{erf}\left(\tfrac{w}{\sigma\sqrt{2}}\right) = \int_0^w \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{t^2}{2\sigma^2}}\,dt .$$ The *error function* in relation to the Gaussian

In the above definition, the variable *t* corresponds to $z-\mu$, so that *t*=0 corresponds to $z=\mu$. A *t*-value outside of the range between –*w* and +*w* is an *outlier* if the total (integrated) probability of finding these extremes is small. For a Gaussian, the total probability that *z* deviates from μ by more than 2σ is five percent. The total probability that *z* deviates from μ by more than 4σ is one hundredth of one per cent, i.e. 1 in 10 thousand. So if you observe outliers differing from the mean of your sample by three or four standard deviations, you tend to think that the outliers came from a different population than the rest of the samples.

Let me repeat: If you observe data that are "off-the-wall" relative to the mean of your sample, you tend to infer that the aberrant data belong to a different distribution. But, your conclusion about such outliers may not be valid, because the distribution of the variable in question is not Gaussian, or the sample may be too small.

So what does this have to do with the issue of vote counting accuracy? Our evaluation of the accuracy of vote counting in Los Angeles County is based on comparing the manual and machine counted tally for one percent of the ~5,000 precincts. Almost all the machine inaccuracies that we observe are due to the failure to record ballot marks; the hand count is assumed to be perfect. Thus the *error rate* corresponds to the number of missed marks, or undercounts, divided by the total number of marks in the tally. The *average* or *global error rate* is the total number of missed marks for all contests, divided by the total number of marks for all contests. There are corresponding formulas for the error rate for a particular contest, or for a particular voting precinct, or even for a particular contest *in* a particular voting precinct: Again, it is the number of missed marks divided by the number of marks involved in that tally – contest, precinct, or contest within a precinct.

But when the jurisdiction gets smaller, we have to be careful. Obviously, the loss of one or two votes could have a significant effect on inferences about the outcome, especially considering that we are sampling only 1% of the precincts. The model for these error rate phenomena is the "Poisson distributed process". Poisson formulas are applicable to phenomena that have a fairly low intrinsic rate of occurrence. For example, year-in-year-out, the global error rate in these vote-counts has been around 0.7%. That is, 7 marks are missed out of 1000. But the error rate in *subsets* of the data, such as in individual precincts, sometimes appears to be several times higher than the global average (and sometimes lower, even zero), and one needs to ask whether this is due to random fluctuations or to a higher intrinsic error rate for those precincts.

The mathematical behavior of Poisson distributions is rather complicated, but well defined. The Poisson probability distribution is a function of *two variables in addition to the characteristic rate parameter*, e.g. 0.7%, or 0.007. In this context, the first variable is the "sample size", x, which corresponds to the total number of marks in the sample, e.g. a precinct, and the second variable is the number of "hits", c, in the sample, e.g. the number of dropped marks, so that c/x corresponds to the "error rate for the precinct", or "hit rate".

If you had a fictitiously large number of precincts all with the same size, x, the *distribution* of c/x would look like Figure 1, if x were 150:

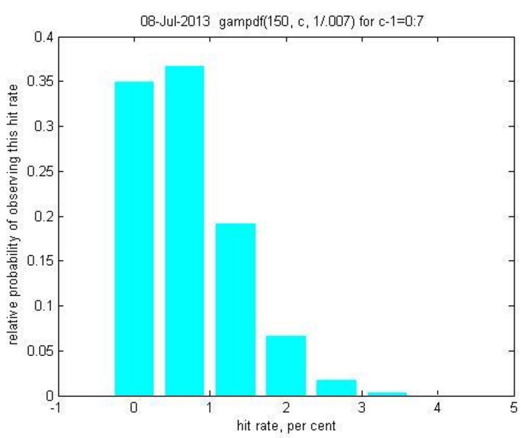

Figure 1: Display of a gamma/Poisson probability distribution for constant x=150 and variable integer count, c. The probability distribution is plotted as a bar graph because c takes on only integer values. The hit rate is given in percent, 100 times c/x.

I used the phrase "fictitiously large number of precincts" deliberately. A "distribution" or "probability distribution" of something – call it *z* – is the probability of observing *z* if you did an infinite number of identical experiments. That is, the *probability distribution* of *z* is a special *histogram* of the relative number of occurrences of *z* for the case of an infinite number of samples. But the above *distribution* is not the *histogram* that you would observe with, say, 50 precincts. One simulation, using random number math gives Figure 2:

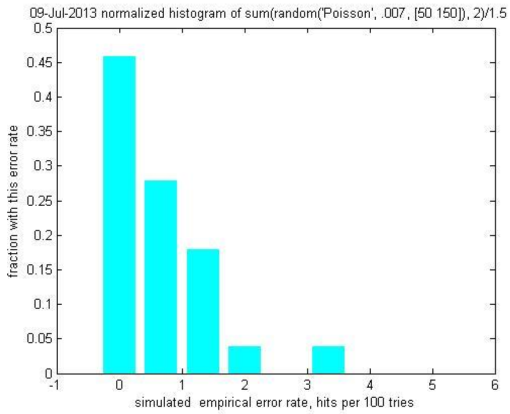

Figure 2: Simulated histogram of the error rate among 50 precincts, when there are 150 marks per precinct and the missed marks are Poisson distributed with an intrinsic rate of 0.7%.

If instead of just 50 precincts, you simulated the data for 5000 precincts of 150 marks each, you would get something like Figure 3:

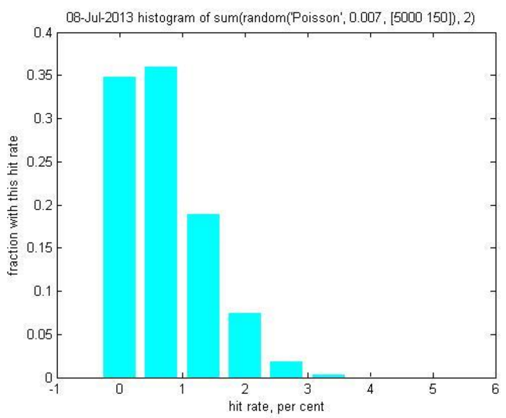

Figure 3: Simulated histogram of the error rate among 5000 precincts, when there are 150 marks per precinct and the missed marks are Poisson distributed with an intrinsic rate of 0.7%.

Figure 3 looks almost indistinguishable from the distribution for a "fictitiously large" number of precincts in Figure 1.

From this you can see that the shape of the observed histogram depends on the number of precincts in the study, but not strongly. In both cases – 50 precincts and 5000 precincts – the histogram is *not* symmetric around the global average, e.g. 0.7%; the peak is at or near zero, but there is a long tail.

The shape of the observed histogram (or the theoretical distribution) depends strongly on the value of *x*, i.e. the "number of marks" in the precincts. In our case we have the extra complication that *x* varies from one precinct to another. If *x* is really large, say 15,000, most of the precincts will have an error rate pretty close to the global average error rate, even if there are only 50 precincts in the sample. Simulations demonstrate this in Figure 4:

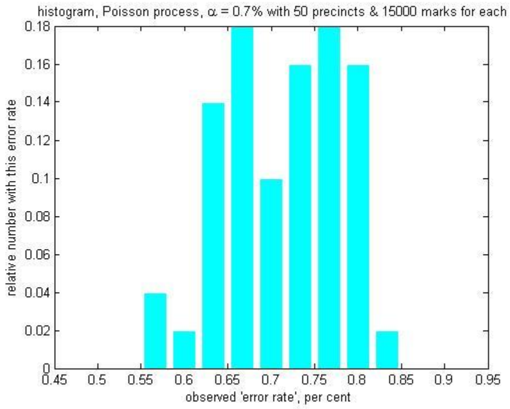

Figure 4: Simulated histogram of the error rate among 50 precincts, when there are 15,000 marks per precinct and the missed marks are Poisson distributed with an intrinsic rate of 0.7%.

The histogram in Figure 4 is qualitatively very different from the previous ones, because there is nothing near zero per cent and nothing out at 3%. The mean error rate is 0.7%, the variance of the error rate is .017%, and the standard deviation of the error rate is 0.130%. When you first look at the above figure, you may tend to dwell on the jagged shape and the glitch at 0.7%. That occurs because we have a histogram of only 50 precincts. But for the fictitiously large number of precincts, the error rate distribution would look like the familiar bell curve:

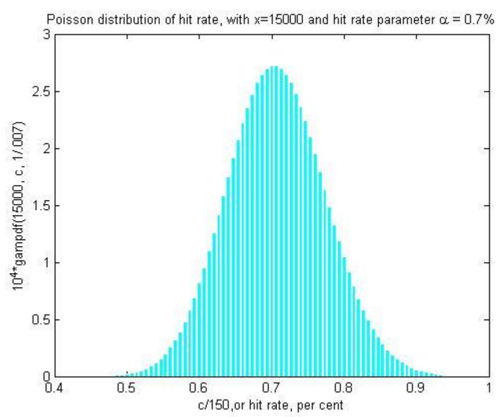

Figure 5: Display of a gamma/Poisson probability distribution for constant *x*=15,000 and variable integer count, *c*, with an intrinsic error rate, α, of 0.7%. The error rate (hit rate along the horizontal axis) is plotted in percent, and corresponds to 100 times *c*/*x*.

From these experiments, we see that if the number of marks per precinct is large, the distribution of error rates will tend to center around the global mean, even if we have only 50 precincts. (That assumes that all the precincts have the same intrinsic error rate.) On the other hand, if the number of marks per precinct is small, e.g. 150, both the empirical histogram and the idealized distribution of error rates have substantial amplitude at zero and the intrinsic rate, but a long tail; error rates several times the intrinsic rate are not improbable.

The following figure shows that with 750 marks per precinct, we can expect the behavior ascribed to a "large number of marks". That is, the histogram is almost symmetric around the peak at the nominal frequency, and near zero at zero.

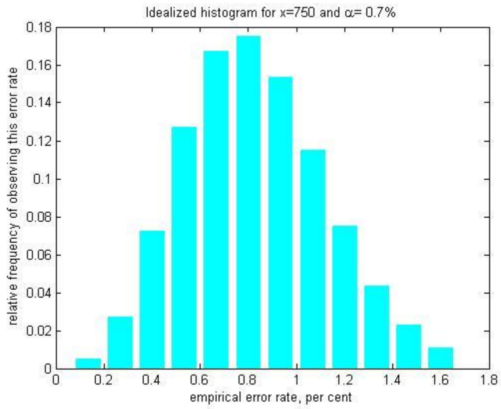

Figure 6: Display of a gamma/Poisson probability distribution of the error rate *c*/*x*, for constant *x*=750, variable integer count, *c*, with an intrinsic error rate, α, of 0.7%. The error rate (hit rate along the horizontal axis) is plotted in percent, and corresponds to 100 times *c*/*x*.

Why is the number of ballot marks per precinct important? This parameter determines the shape of the histogram we can expect. If the number of ballot marks per precinct falls below about 300, the distribution is no longer symmetric about the peak, and we can expect long tails with apparent outliers, and, at the same time, an apparent concentration near error rates of zero. Figure 7 shows the idealized histogram for *x*=300:

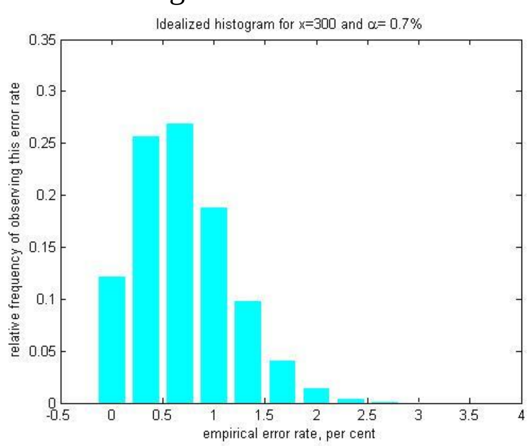

Figure 7: Display of a gamma/Poisson probability distribution of the error rate *c*/*x*, for constant *x*=300, variable integer count, *c*, with an intrinsic error rate, α, of 0.7%. The error rate (hit rate along the horizontal axis) is plotted in percent, and corresponds to 100 times *c*/*x*.

Where do our data lie between these two conditions of "large" and "small" number of marks per precinct? The answer is not entirely clear cut. November elections have turnout between 50% and 85%; LA County turnout was 66% in November 2012 (http://www.lavote.net/GENERAL/PDFS/PRESS_RELEASES/11072012-092030.pdf), whereas the June elections tend to have a turnout near 20%. Furthermore, the EN and VBM ballots are counted separately; I inferred that except for November 2012, the same machine did not count both the EN and the VBM ballots for each precinct. So we *had* to treat the two batches separately. The EN ballots comprise about two thirds of the number of ballots cast, so the number of marks in the VBM sample is about half the number of marks in the EN sample. Additionally, the size of the precincts is not uniform. The precincts of the EN data of 2010 and 2012 all/each have a number of ballot marks > 750. We are not so lucky with the VBM data: The average number of VBM ballot marks per precinct in June 2010 was 900, but of the five worst-counted precincts, three had about 350 ballot marks. However, the error rates for those three worst-counted precincts were *massive*: 30%, 25% and 25%, respectively – completely off the chart – even for the ideal histogram for 300 ballot marks. With an average error rate of 0.7%, the cumulative probability of an error rate of 25% or more is *zero*. I see no way to ascribe such large error rates to the random fluctuations of a Poisson process with an intrinsic error rate parameter of 0.7%. I also analyzed these data in a Bayesian formalism, which inverts the problem: given the observed *empirical* error rates, what is the probability of a particular intrinsic error rate? The conclusions are the same: A few machines are performing very poorly, and do not have the error rate characteristic of the rest.

We should have the intellectual honesty to ask how these error rates – dropped counts – might affect the *outcome* of the election. It is intuitive to suppose that if all the machines *randomly* drop counts with the same frequency, then the election outcome will be unchanged. The outcome will be the same as if the turnout were *uniformly* that much lower. That intuition is slightly incorrect. Because the process is random, not every contestant (either a candidate for political office or 'yes' or 'no' on a proposition) will have its tally reduced by the same fractional amount. The percentage uncertainty in the outcome depends on the *standard deviation of the error rate*, not the error rate.

For a true Poisson distribution the mean and mode (peak of the distribution) are the same as the nominal frequency. In the examples of this essay, the variance of the error rate distribution depends on the parameter '*x*'. (The standard deviation is the square root of the variance.) I argue that *x* is sufficiently high that the variance should be the same magnitude as the mean error rate, *or less*, and we could not observe these error rate histograms unless some of the machines had intrinsic error rates substantially higher than the rest. Another type of analysis, termed *Bayesian*, uses calculus to infer what the true error rate might be, given the empirical error rate. This is called the *posterior distribution* for the error rate. If the data all come from a process with the same error rate, the posterior distribution should match the histogram, and with a large enough sample size (the parameter *x* in the figures above), the variance in the error rate becomes very small. I do not want to go into the details here, but again there is no way to match the posterior distribution to the observed histograms using a single error rate. In the VBM data, particularly for 2010, we saw high variance of the error rate – several percent. So given the heterogeneity of the machines' error rates, it is fair to say that the standard deviation of the error rates is equal to or greater than the error rates. The election outcome will be uncertain if the percentage point difference between two contestants is smaller than the empirical standard deviation of the error rates. The standard deviation of the error rates could be greatly reduced by repairing the bad machines and performing quality control in real time on Election Day and Election Night.

The author welcomes readers of this essay to request write-ups of the actual data, or even source materials. The egregious errors of June 2010 are still posted online:
http://protectcaliforniaballots.org/RandD/2010%20June%20Audit%20Analysis.pdf
The analysis was performed programming in Matlab; the raw data are in Excel spreadsheets.

Antonie K. Churg, Ph. D.
achurg@socal.rr.com
(310) 539-6506