On Ken Rahn's Statistical Analysis of the Neutron Activation Data in the JFK Assassination[1]

A. Snyder

Sun May 27 20:49:57 PDT 2001

Ken Rahn's statistical analysis of the neutron activation analysis (NAA) of the bullets and bullet fragments in the JFK assassination case is flawed: it leads him to an incorrect conclusion about the definitiveness of the NAA evidence. A simple perusal of Guinn's measurements of samples from the different bullets (see appendix A) will show you that Guinn's claim that accidental matches were "extremely unlikely or very improbable" is incorrect. Elaborate statistical arguments obscure this basic fact, but let's have a look at them anyway.

The problems with Rahn's paper include the following:

→ He introduces and knocks down the "straw man" of 5 separate bullets being responsible for the 5 evidentially important samples.

  - As nobody is seriously proposing this hypothesis, introducing it only confuses the issue.

[1] www.kenrahn.com/JFK/Scientific_topics/NAA/NAA_and_assassination_II/Key_problem

- By assuming a probability of one (apart from a irrelevant combinatoric factor) for "genuine" match hypotheses, Rahn fails to make a real comparison of the competing hypotheses.

  - In any complex event such as an assassination, the probability of what actually happened is bound to be small. To test different hypotheses you have to compare the ratio of the probability of the observed events being produced by one hypothesis to that of them being produced by the other. If one neglects the possibility of one of the hypotheses failing, there is no comparison.

- Rahn's "key" probability $P_{tightness}$ is misleading. This is especially so since he assumes "probabilities of unity" for "fragments considered genuine," but focusing on "tightness" can lead to mistaken conclusions even when $P_{tightness}$ is computed correctly for both hypotheses. Because it is insensitive to the separation being too big for genuine matches, $P_{tightness}$ is a poor choice for discriminating between genuine and accidental matches.

- In extracting the probability ($P_{5-0}^{overall} = 0.9967$ in Rahn's notation) that his preferred hypothesis is correct he implicitly assumes *a priori* probabilities for the various scenarios. These probabilities are not known, so such a calculation is meaningless.

  - This is the same error made by the HSCA acoustic panel when they claimed there was a 95% chance there was a shooter on the grassy knoll. All they could sensibly claim was that there was 5% chance the signal they saw was the result of noise which one has to suspect would not have sounded very convincing to a congressman.

Only two issues:

Rahn calculates probabilities for a lot of hypotheses that nobody is proposing. This introduces a lot of confusing combinatorics that are irrelevant to the issues of interest. The only issues relevant to the case for or against conspiracy are:

⟶ Does the "tightness" of the grouping of the three "limo" fragments rule out the possibility of more than one bullet contributing to this group?

⟶ Does the apparent match between CE399 (the "magic" bullet) and the Connally wrist fragment establish that the fragment came from CE399?

Each of these issues can be considered separately.

The fact that five independent fragment are unlikely to have produced the 3/2 grouping seen in the evidence is irrelevant to these issues.

An assumption:

For purposes of this note, I'm going to use Rahn's assumption that the probability of bullet antimony content is uniformly distributed between 0 and $1200ppm$. I don't agree with his assertion that it doesn't matter; deviations from flat could easily increase the probability of accidental matches by a factor of 2 or more. This might not matter for the 5 independent bullet hypothesis, since 2^5 times a very small number is still small, but for the hypotheses of interest it matters. However, as it is easy to work with, I'm going to use it anyway for illustrative purposes.

With this assumption, the probability that a random bullet has an antimony fraction between x and $x + dx$ is given by $P_{sb}(x)dx$ where:

$$P_{Sb}(x) = \frac{1}{1200} \tag{1}$$

This function is called the probability density function or PDF for short.

CE399 and friend:

The key ingredient in Rahn's calculation is P_{tight} – the probability that two measurements of antimony content would be as close or closer together than they are. For two different bullets he estimates

$$P_{tight} = \frac{2 \times \Delta x}{1200} \tag{2}$$

where Δx is the difference in antimony content between the two measurements under consideration.

For samples from the same bullet Rahn assumes $P_{tight} = 1$. This leads to absurd results. E.g., if $\Delta x = 0$, $P_{tight} = 0$ for the two bullet hypothesis and $P_{tight} = 1$ for the single bullet hypothesis. Case closed, right?

Wrong! P_{tight} for the single bullet hypothesis can't be 1 regardless of Δx. There is measurement error, so that for $\Delta x = 0$, P_{tight} is 0 for the single bullet hypothesis too.

By assuming $P^{single}_{tight} = 1$ Rahn is effectively ignoring the comparison of competing hypothesis that is central to testing them. The preferred hypothesis is given a free ride; it can't lose.

As an example of how Rahn's methodology leads to absurd results consider CE573 and CE141:

$\Delta x = 2$ for these two samples implies $P^{two}_{tight} = 2 \times 2/1200 = 3.4 \times 10^{-3}$. This would seem to establish that the bullet found in General Walker's wall and the unfired bullet found in the Carcano rifle left on the 6th floor are the same bullet.

If we assume the resolution on the antimony measurement is $30ppm$,[2] then $P^{single}_{tight} \approx 0.05$ for the single bullet hypothesis – not 1.0. In this case the ratio of the two bullet to single bullet tightness probability is $P^{two}_{tight}/P^{single}_{tight} \approx 0.068$ – small but not off the wall[3]!

Even when P_{tight} is computed *correctly* for the single bullet hypothesis, instead of just being set to 1.0, it can lead to incorrect conclusions in some circumstances.

Consider the following example: let's assume[4] for the sake of argument that the quoted measurement errors on CE399 and CE842 (see appendix B) can be used to construct the PDF needed. In this case the PDF for the hypothesis that CE842 is a fragment from CE399 is:

$$P_{single}(\Delta x) = \frac{1}{\sigma_{\Delta x}\sqrt{2\pi}} \times e^{-\frac{(\Delta x)^2}{2\sigma^2_{\Delta x}}} \qquad (3)$$

where resolution $\sigma_{\Delta x}$ on $\Delta x = x_{CE842} - x_{CE399}$ is $\sigma_{\Delta x} = \sqrt{\sigma^2_{CE399} + \sigma^2_{CE842}} \approx 11.4ppm$ (I denote errors are by the Greek letter σ).

---

[2] If one only uses the measurement error, CE399 and CE842 as well as the three limo fragments don't match very well. As will be disscused below, a larger error can be justified due to sampling vatiations.

[3] The odds of such a relatively low probability ratio appearing for some pair of bullets among the evidentiary samples is in fact not all that small as there ≈ 15 possible pairings.

[4] I am NOT claiming this is correct. Guinn's repeat measurements on bullets clearly show a sampling error must be included too.

The probability that this Gaussian distribution would produce a separation Δx less than $35ppm$ is indeed near one ($P_{tight}^{single} \approx 0.997$). So with $P_{tight}^{two} = 2 \times 35/1200 \approx 0.058$, the two bullet hypothesis appears to be ≈ 17 times less likely to have produced the observed separation then the single bullet hypothesis.

However, Rahn's focus on "tightness" leaves out the fact that the probability that these two fragments are as far apart as they are is quite small (≈ 0.003) – resulting in the misleading impression that the single bullet hypothesis is preferred.

A more reasonable approach is to use a "likelihood" ratio to compare our hypotheses. Likelihood comes with a nice theorem[5] that a selection based on the ratio of likelihoods will produce the minimum number of false positives (accidental matches) for a given level of false negatives (failure of true matches). More importantly the ratio of likelihoods will properly weight the degree to which samples are "tight" or "far apart." It is not biased in favor of one or the other.

The likelihood is defined as the value of the PDF for a given hypothesis evaluated at the measured value of the variable in question. Only ratios of likelihoods are meaningful.

For the two bullet hypothesis the likelihood is just $L^{two} = 1/1200$ for the simple uniform distribution model we are using. For this example, the likelihood of single bullet hypothesis using the PDF given by eqn. (3) above is $L^{single} = P_{single}(35) \approx 3 \times 10^{-4}$.

The likelihood ratio is then $\frac{L^{two}}{L^{single}} \approx 2.7$ which says that the probability that the two bullet hypothesis would produce $\Delta x = 35$ is almost three

[5] F. James, "Determinng the statistical significance of experimental results," Lectures at 1980 CERN School of Computing, Vranona, Attiki, Greece, September 1980.

times bigger than the probability that the single bullet hypothesis would produce it – rather than 17 times smaller.

However, besides not being an overwhelmingly large ratio, it is not (see footnote 4) a realistic calculation. Because of Guinn's sparse and less accurate measurements on repeated samples from the same bullet we are forced to include a sampling error in addition to the quoted measurement error – this will increase L^{single} substantially, so that two bullet hypothesis will again be disfavored – though not by a lot.

How to construct the PDF for the single bullet hypothesis is not at all clear. The repeated measurement are sparse: 4 samples from each of only 3 different bullets – I've included a copy of these results as an appendix C.

These measurements suffer from two problems: 1) the measurements were done with *an order of magnitude* worse resolution ($40-90ppm$) than those made on the evidentiary samples ($4-9ppm$); 2) there are outliers that indicate the bullets are not very homogeneous. The 2nd problem means there is fairly substantial probability of a false negative ($\approx 2/12$ taking the most obvious cases) just from the outliers. The 1st problem means that we don't really know how wide the Δx distribution would be for the "major" component of the bullet in the better measured evidentiary samples. Excluding the outliers 6002A and 6003A the distribution of the remaining measurements is consistent with the quoted measurement errors, but given the prevalence of outliers it's impossible to exclude a contribution from sample variation. A sampling error $\approx 30ppm$ level or so (that would make a CE399/CD842 match reasonable) would still be consistent with the variation observed after excluding the obvious outliers[6].

To proceed we need to make some reasonable assumption for the PDF. I

[6] Sampling error refers to a variation in antimony concentration between different samples taken from the same bullet – a variation which would still be there even if the measurement error was zero.

propose to use the following:

$$P_{single}(\Delta x) = \frac{f_{outlier}}{1200} + \frac{1 - f_{outlier}}{\sigma\sqrt{2\pi}} \times e^{-\frac{(\Delta x)^2}{2\sigma^2}} \tag{4}$$

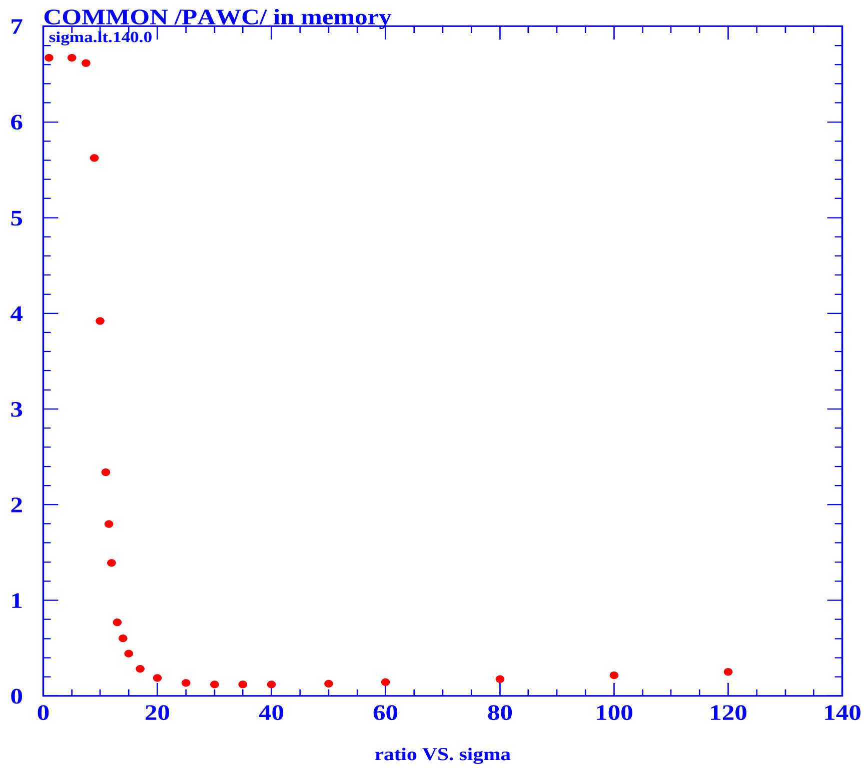
with $f_{outlier} = 0.15$ and $\sigma = 30ppm$.

This is not going to be real accurate, but it will have to do. It will at least serve to illustrate the *order of magnitude* that can be expected.

For CE399/CD842 with $\Delta x = 35$, we have $L^{two} = 1/1200 = 8.3 \times 10^{-4}$ (just as before) and $L^{single} = P_{single}(35) \approx 5.8 \times 10^{-3}$. The likelihood ratio is $\frac{L^{two}}{L^{single}} \approx 0.14$ or about $7/1$ against the two bullet theory – not exactly overwhelming. Keep in mind, too, that this is a very uncertain estimate, e.g., non-uniformity of the $P_{Sb}(x)$ distribution could easily mean L^{two} is twice as big.

Since the correct value to use for the sampling error is not known, it is interesting to see how our result depends on the error assumed. Figure 1 on the next page shows the likelihood ratio $\frac{L^{two}}{L^{single}}$ for $\Delta x = 35$ plotted vs. σ.

Figure 1: Likelihood ratio vs. assumed error for $\Delta x = 35$.



The measurement resolution lies at $\sigma \sim 11ppm$. The error I use to account for sampling error is $\sigma = 30ppm$. The flat part at very low σ in figure 1 corresponds to the flat 15% outlier component which limits how well the hypotheses can be distinguished even with perfect measurements. The rise at large σ is due the fact that it becomes impossible to distinguish the hypotheses as the resolution σ approaches the size of entire antimony content range.

There is no value of σ that's would allow us do discriminate strongly between the two bullet and single bullet hypotheses.

The 3 limo fragments:

We can apply the same likelihood ratio test as above to the three limo fragments, if we stick to comparing the hypothesis that one of the fragments is a "ringer"[7] and the other two "genuine" to the hypothesis that all three are "genuine." Each possible "ringer" can be considered in turn. The hypothesis that all three fragments are from different bullets is not very interesting, so I don't bother with it.

Given as part of the hypothesis that two of the fragments are from the same bullet, we can use their average as a single measurement of the same bullet. In principle, this average might be considered to have better resolution than a single measurement, but given the imponderables of sample variation ignoring that possibility is no worse than the approximation we're already using. I'll just stick with the same $30ppm$ resolution used above along with the same 15% outlier fraction.

The three measurements are CE567 with 602 ± 4, CE843 at 621 ± 4 and CE840 at 642 ± 3 (the average of the two measurements made on this sample).

Case I: CE840 is the "ringer"

$$\Delta x = 30.5 \Rightarrow \frac{L^{two}}{L^{single}} = \frac{8.3 \times 10^{-4}}{6.8 \times 10^{-3}} \approx 0.12 \approx 1/8 \tag{5}$$

[7] By "ringer" I mean only a different bullet; no implication that a bullet fragement was "planted" or anything like that is intended.

Case II: CE843 is the "ringer"

$$\Delta x = 1 \Rightarrow \frac{L^{two}}{L^{single}} = \frac{8.3 \times 10^{-4}}{1.1 \times 10^{-2}} \approx 0.072 \approx 1/14 \tag{6}$$

Case III: CE567 is the "ringer"

$$\Delta x = 29.5 \Rightarrow \frac{L^{two}}{L^{single}} = \frac{8.3 \times 10^{-4}}{7.1 \times 10^{-3}} \approx 0.12 \approx 1/8 \tag{7}$$

None of these are strong enough to rule out the two bullet hypothesis especially given that these estimates are highly uncertain; the actual likelihood ratios might be bigger or smaller.

It's worth mentioning that these test only address the issue of whether two bullets could have contributed to the three samples we have. There could have been other bullets from which no fragments were recovered – a hypothesis Rahn does not even consider. Even if the likelihood ratios above were large enough to convince most everyone that one bullet contributed all three samples, this would not rule out the possibility that another bullet struck President Kennedy or Gov. Connally.

A priori probabilities

Classical statistics only allows us to calculate conditional probabilities. Things such as, if hypothesis A is true the probability that I would observe x is $P(x; A)$, can be defined. If we don't know the *a priori* (before the measurement of x) probability, we have no means of turning the conditional probability into a absolute probability (the probability that A is true) – something we could bet on, i.e., betting odds.

Bayes' theorem[8] provides a way to calculate absolute probabilities if we know the *a priori* probabilities. E.g., suppose there are two hypotheses A & B with a priori probabilities P_A^{before} and P_B^{before}. In this simple case, with only two hypotheses, $P_A^{before} + P_B^{before} = 1$.

Further suppose that we perform a measurement of parameter X that yields a value x and that the probability of getting X between x and $x + dx$ is $P(x; A)dx$, if A is true, and is $P(x; B)dx$, if B is true. Then Bayes' theorem tells us that the probability of A being true after including the information obtained from the measurement is

$$P_A(x) = \frac{P_A^{before} \times r(x)}{P_A^{before} \times r(x) + P_B^{before}} \quad (8)$$

where $r = \frac{P(x;A)}{P(x;B)}$ is the likelihood ratio. Plug in your measured x and you get absolute probabilities – you can bet on 'em!

Perhaps a more concrete example will make this clearer. Consider a rare cancer that afflicts 1 in a million people and kills everybody who gets

[8] Thomas Bayes, "An Essay Towards Solving a Problem in the Doctrine of Chances", Philosophical Transactions *53*, 370 (1763); Reprint Biometricka *45*, 296 (1958)

it. Suppose there is a cure, but it has side-effects and kills 1 of every thousand people who use it. Clearly, it would make no sense to use this cure prophylactically on everybody; the cure would kill more people than the disease.

However, somebody invents a test; it's not, of course, perfect. It detects the cancer in 90% of the people who have it (a 10% false negative rate)[9] and has a false positive rate of only 1%. The likelihood ratio between the sick and the not sick hypothesis is 90 if the test is positive ($X = true$). Sounds great, but is it good enough?

Suppose you take the test and test negative. The probability you have the disease is reduced (using Bayes' theorem) to $\approx 10^{-7}$ – a lot smaller. However, the probability of getting this disease was already very small compared to other risk you face, so you've not learned much.

Suppose you take the test and test positive. The probability you have the disease is now $\approx 9 \times 10^{-5}$. Almost 1/10000. Scary. Unfortunately, the cure is still about 10 times more likely to kill you than the disease. You might be happier not knowing.

If the likelihood ratio or conditional probabilities are used naively without taking into account the 1/1000000 *a priori* probability, disastrously wrong decisions could be made. When the *a priori* probabilities are known, they can be used to calculate real probabilities that can be used to make important decision such as whether to risk a dangerous treatment or whether a test is worth administering widely.

However, in a unique case like the JFK assassination we have no way of knowing the *a priori* probabilities. We can't, e.g., count the occurrences of "magic" bullets in a large number of similar cases. Given this situation there is no way to calculate the probability a given hypothesis is true as

[9] X is a discrete variable in this case with only the values $true$ and $false$, but the principle involved is still the same.

Rahn attempts to do. All such attempts involve an implicit assumption about the *a priori* probabilities and these are simply not known.

The best we can achieve is to calculate the conditional probabilities and compare their ratios for our competing hypotheses. Lacking quantitative information on the *a priori* probability of each hypothesis, we are stuck with our "gut" feelings about their plausibility or implausibility. If the likelihood ratio is large enough, it will convince most people, but if it is $\sim 10/1$ (either way), only those already convinced will be convinced.

It sounds more convincing to say there is a 90% chance Oswald did (or did not) kill JFK alone, but that is propaganda not science.

Prospects

The situation could be cleaned up with more measurements to establish the properties of WCC bullets. More and more accurate measurements of repeated samples from the same bullet would allow one to characterize the single bullet PDF's more accurately. Detailed repeated measurement of CE399 would be especially useful in this regard. More measurements of a larger number of bullets would allow one to get a more accurate estimate of the antimony distribution giving us a better estimate of the two bullet likelihood.

Suppose we found that the measurement resolutions ($11.4ppm$ for the comparison of CE399 and CE842) could be used for the "major" antimony component and that outliers could be accounted for by a 15% flat component while accidental matches were enhanced by a modest factor of 1.5 due to non-uniformity. Then we would have for $\Delta x = 35ppm$:

$$\frac{L^{two}}{L^{single}} = \frac{1.2 \times 10^{-3}}{3.9 \times 10^{-4}} = 3.1 \qquad (9)$$

So in this case the two bullet hypothesis[10] is favored by a factor of ≈ 3.

However, it is unlikely that the likehihood ratio will turn out big enough or small enough to really be definitive. The NAA evidence is likely to remain as it is – **inconclusive**!

[10] The antithesis of "The Single Bullet Theory."

Appendix A: Extract from Guinn table II-A

The Trace-element composition of Mannlicher-Carcano 6.5 mm bullet leads from lots 6000, 6001, 6002 and 6003.

Lot No.	Bullet No.	Wt.	Silver (ppm)	Antimony (ppm)
6000	A	51.2	11.8 ± 0.4	173 ± 3
	B	45.6	13.5 ± 0.5	261 ± 3
6001	A	47.8	12.2 ± 0.6	158 ± 3
	B	57.9	15.3 ± 0.5	732 ± 5
	C	58.5	8.5 ± 0.4	1218 ± 7
	D	47.2	11.6 ± 0.4	161 ± 3
6002	A	51.8	9.1 ± 0.4	385 ± 4
	B	52.8	9.7 ± 0.4	949 ± 6
	C	55.3	6.0 ± 0.3	24 ± 1
	D	51.3	8.3 ± 0.6	121 ± 2
6003	A	54.3	15.9 ± 0.5	730 ± 5
	B	44.6	7.9 ± 0.4	80 ± 2
	C	44.7	8.8 ± 0.4	464 ± 5
	D	44.0	8.7 ± 0.4	240 ± 3

Appendix B: Extract from Guinn tables I and III

Measurements of evidentiary bullets and fragments

WC No.	Silver (ppm)	Antimony (ppm)
399 “Magic”	7.9 ± 1.4	833 ± 9
567 “Limo”	8.1 ± 0.6	602 ± 4
843 “Limo”	7.9 ± 0.3	621 ± 4
842 “Wrist”	9.8 ± 0.5	797 ± 7
840 “Limo”	8.6 ± 0.3	638 ± 4
2nd measure	7.9 ± 0.5	647 ± 4
573 “Walker”	20.6 ± 0.6	17 ± 2
141 “Unfired”	–	15 ± 1
2nd measure	22.4 ± 1.0	–

Appendix C: Extract from Guinn table II-C

Homogeneity measurements on four specimens from each of three individual Mannlicher-Carcano bullets

Lot	Specimen	Silver (ppm)	Antimony (ppm)
6001	C	8.5 ± 0.4	1139 ± 60
-	C1	9.5 ± 0.4	1062 ± 60
-	C2	10.1 ± 0.6	1235 ± 93
-	C3	9.2 ± 0.5	1156 ± 90
6002	A	9.9 ± 0.4	358 ± 47
-	A1	10.3 ± 0.3	983 ± 51
-	A2	9.9 ± 0.3	869 ± 47
-	A3	10.2 ± 0.5	882 ± 81
6003	A	15.9 ± 0.5	667 ± 58
-	A1	9.6 ± 0.4	395 ± 54
-	A2	8.3 ± 0.3	363 ± 39
-	A3	9.8 ± 0.4	441 ± 51