~~Everything~~ Some of What You Always Wanted to Know About the APT*

(Via Some Abandoned IG D.K Resolution 22 Comments)

Joshua E. Hill, PhD

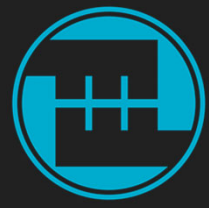
CMUF Entropy WG
20240319-4

* But Were Afraid to Ask

Preface: [Draft IG D.K Resolution 22] Comments

Initially, this presentation was a set of comments on the IG D.K Draft Resolution 22; these were essentially:

- It isn't clear why this requirement should be restricted to wide raw data.
- The "half-entropy" failure mode isn't completely specified.
- The required statistical power is too large (34-71 times the required statistical power for developer-defined health tests).
- The recommended "mapping" solution tends to make the APT worse, not better.

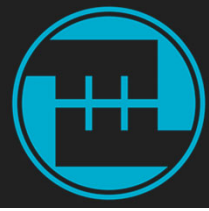


Preface: [Draft IG D.K Resolution 22] Comments

Initially, this presentation was a set of comments on the IG D.K Draft Resolution 22; these were essentially:

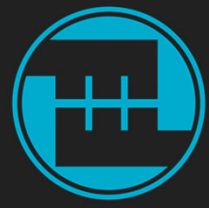
- It isn't clear why this requirement should be restricted to wide raw data.
- The "half-entropy" failure mode isn't completely specified.
- The required statistical power is too large (34-71 times the required statistical power for developer-defined health tests).
- The recommended "mapping" solution tends to make the APT worse, not better.

… but, you know, spread out over 60 slides with graphs and stuff…



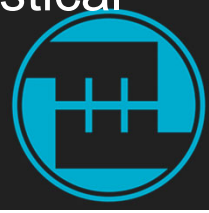
Preface: [Draft IG D.K Resolution 22] Comments

NIST has recently indicated that they are substantially changing the text, so you get the fun technical drippings!



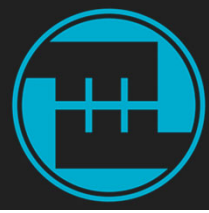
Part 0: Statistical Tests

- Statistical tests are traditionally analyzed using their *false positive rate* (α, here the probability that the test incorrectly detects a failure when the entropy source is working normally.)
- From a security perspective, we're more interested in the *statistical power* $(1 - \beta)$ of the test (here the probability that the test correctly detects a failure when the entropy source is in a particular failure mode).
- This is done because there are many ways for a source to fail (so the statistical power is failure-mode specific), but there is only one α.
- Commonly (but not always), increasing α results in an increase in the statistical power and decreasing α results in a decrease in the statistical power.



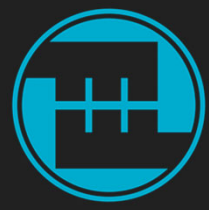
Part 0: The APT

1. A = next()
2. B = 1.
3. For i = 1 to W–1
    a) If (A == next()) B=B+1
    b) If (B ≥ C) signal a failure
4. 4. Go to Step 1.



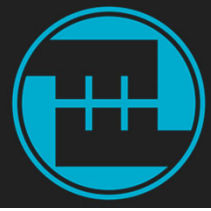
Part 1: Important APT Characteristics

- The APT is categorical, so how data is encoded (e.g., the raw data width) isn't relevant.
- Commonly, the distribution parameter that establishes false positive rate for a selected cutoff is the probability of the most likely symbol, $p_{\max}$ ($p_{\max} \approx 2^{-H}$ for an IID noise source). Here, we discuss this probability in terms of the "apparent entropy":
  - $H_{\text{apparent}} = -\log_2 p_{\max}$
  - $H \leq H_{\text{apparent}}$ (in all sources because of the MCV estimator).



Part 0: The APT Failure Mode

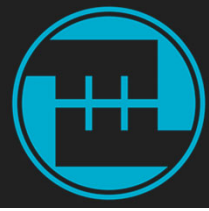
- The APT is intended to (eventually) detect the most likely symbol becoming dramatically more common than expected.
  - This situation naturally arises, even for **some** non-IID sources.
- The APT is unlikely to detect any failure mode that does not make the most likely symbol more common.



Part 0: APT Cutoffs

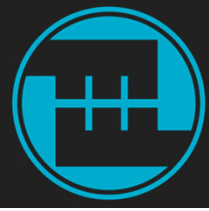
In the SP 800-90B Section 4.4.2 APT cutoff procedure:

- There are relevant corrections ([HJ 2019, Comment 10b]), but they do not change the essential results.
- The SP 800-90B Section 4.4.2 analysis approach:
  - Makes an underlying IID assumption.
  - Bounds the false positive rate (α) using the assumption that the APT reference symbol (A) is the most likely symbol.
- The actual false positive rate and statistical power associated with a particular APT cutoff selection are distribution-dependent and can be estimated via large-scale simulation.



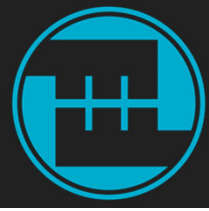
Part 0: APT Cutoffs

- Most noise sources aren't IID, so if the APT cutoff is established using the assessed entropy, then the targeted false positive rate isn't likely attained using this procedure. In this case:
  - The actual false positive rate is likely to be much smaller than intended.
  - The APT is only likely to detect failures that result in the apparent entropy being much less than the assessed entropy.
- Very large cutoffs (near the window size, associated with very low entropy assessments) and very small cutoffs (near 0, associated with very high entropy assessments), may result in a lower than desired statistical power.



Part 1: Important APT Characteristics

- For low apparent entropies, the number of symbols that can be output can't have much impact, as the most probable symbol is very likely to occur.
- For higher apparent entropies, the residual probability must be associated with some symbols, so the observed APT false positive / statistical power are distribution-dependent.
- For very high apparent entropies, the number of symbols again doesn't have much of an impact (as the distribution is necessarily close to uniform).

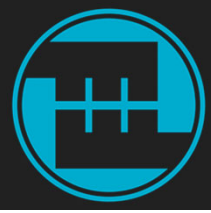


Statistical Power: The Issue

- [Draft IG D.K Resolution 22] requires a statistical power of 50%.
- The per-window statistical power, γ, required by SP 800-90B Section 4.5 Criterion (b) is a function of the window size.

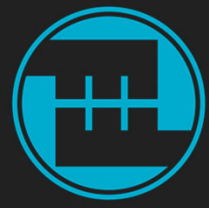
$$1 - \beta \geq 1 - \left(\frac{1}{2}\right)^{1/\left\lfloor \frac{50{,}000}{W} \right\rfloor}$$

Window Size (W)	Minimum Per-Window Statistical Power $(1-\beta)$
512	0.7%
1024	1.4%



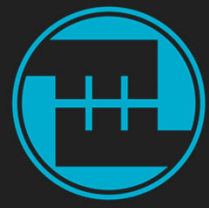
Mapping and the APT: The Issue

- Performing an APT on (many-to-one) mapped data tends to obscure shifts in the probability distribution of the unmapped data.
- If we use the same APT cutoff in the mapped-data APT, then the mapped-data APT necessarily has the same or better statistical power, but also likely has a larger false positive rate.
- We are instructed by SP 800-90B Section 4.4.2 to select an APT cutoff based on the targeted false positive rate, so presumably after mapping we follow the same procedure. Such a (larger) mapped-data APT cutoff will result in a similar false positive rate for the mapped-data APT but may also result in a dramatically worse statistical power.



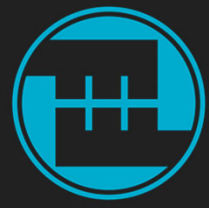
A Relevant Aside on JEnt

- NIST is particularly interested in how the Jitter Entropy Library (JEnt) APT performs, as internally JEnt uses wide (64-bit) raw data.
- In practice, the delta format used may require a large integer to represent it (particularly with the idiomatic JEnt 3.0.1 and earlier delta format), but the number of symbols present is smaller than the raw data width suggests.
- The way that the symbols are encoded isn't relevant to the APT behavior (the APT is categorical).



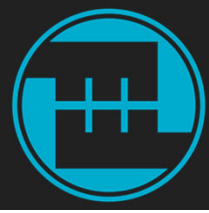
A Relevant Aside on JEnt

- JEnt 3.1.0 and later sets the APT cutoff based on the `osr`.
  - $H_{\text{submitter}} = \frac{1}{\texttt{osr}}$.
  - The target false positive rate is $\alpha = 2^{-30}$.
  - In earlier versions of JEnt, this cutoff is effectively fixed at 326 (erroneously off-by-one from the SP 800-90B Section 4.4.2 value for $H = 1.0$ for $\alpha = 2^{-30}$) or the APT is absent.



A Relevant Aside on JEnt

- For modern versions, this behavior is consistent with the SP 800-90B Section 4.4.2 procedure, but it is important to note that $\mathrm{H} \leq H_{\mathrm{submitter}} \ll H_{\mathrm{apparent}}$.
  - The actual false positive rate is dramatically less than the targeted false positive rate of $\alpha = 2^{-30}$.
  - The impact on the statistical power is dependent on the identified failure mode and the raw data distribution.

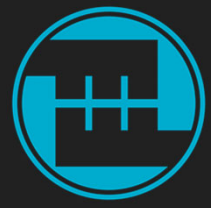


JEnt: An Example

Using JEnt 3.4.1 on a modern-ish Intel platform (Cascade Lake).

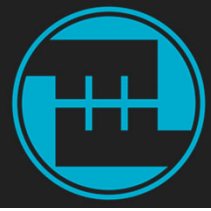
- A data set of 1 billion raw samples has about 25k distinct symbols (fewer than 2^{15}).
- By default, the `osr` is set to 3 on this platform, so here $\mathrm{H} = H_{\text{submitter}} = \frac{1}{3}$.

Entropy Estimate Type	Min Entropy	APT Cutoff for $\alpha = 2^{-30}$
Assessed Entropy	0.333333	459
Apparent Entropy	7.68651	18
8-Bit Mapped Apparent Entropy	7.41904	19
4-Bit Mapped Apparent Entropy	3.99920	71



A Relevant Aside on JEnt

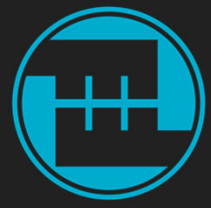
- Given a large sample of raw data from the noise source in a failure mode, we can estimate the maximum cutoff that attains a targeted failure rate.
- The distribution underlying the particular failure mode being examined is relevant.
- For this example, we simulated these failure modes by repeatedly replacing randomly selected data samples in the unmapped data set with the MLS until the desired rate was attained.
- Any translation then occurs on the resulting full-width raw data.



JEnt: An Example

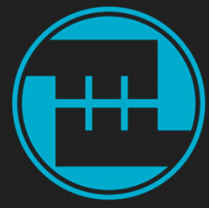
- What cutoff is required to detect various half-entropy modes at the targeted statistical powers? (Cutoff bounds in red are not met by the selected cutoffs.)

	Failure Mode Apparent Entropy	Maximum APT Cutoff for $1-\beta = 0.7\%$	Maximum APT Cutoff for $1-\beta = 50\%$
Unmapped Half Assessed	0.166667	473	455
Mapped to 8 bits Half Assessed	0.166492	473	455
Mapped to 4 bits Half Assessed	0.156459	473	455
Unmapped Half Apparent	3.84326	44	2
Mapped to 8 bits Half Apparent	3.82424	44	2
Mapped to 4 bits Half Apparent	3.01632	44	2



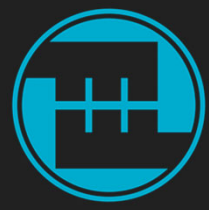
JEnt APT: You Get What's on the Label

- In most cases, the APT did well when detecting the failure mode it was configured to detect.
  - Setting the cutoffs using the apparent entropy is different than setting them using the assessed entropy.
- If you configure the APT based on the assessed entropy (commonly less than half the apparent entropy), then this APT isn't likely to detect a half-apparent-entropy failure mode.



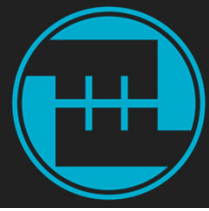
JEnt: A Conclusion

- For these distributions, mapping from 22k symbols to 256 symbols didn't have a substantial impact on the statistical power of the test for a fixed APT cutoff, though it did require a slightly larger cutoff to attain the same targeted false positive rate (and using this larger cutoff would reduce the statistical power).
- Mapping to less than the apparent entropy caused problems in the half-apparent-entropy failure modes.
- There was no observed benefit to mapping to a narrower raw data width.



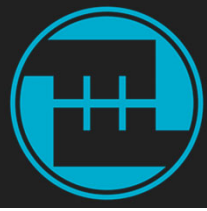
Mapping and the APT: Bounds

- To capture “best case” view of the APT, we’ll examine an IID source.
- This noise source’s output follows the near-uniform distribution from [HD 2012], and which produces all possible 16-bit output symbols with an assessed 10 bits of min entropy.
- SP 800-90B Section 4.4.2 directs us to choose the window size $W = 512$ and an APT cutoff between 8 ($\alpha = 2^{-20}$) and 12 ($\alpha = 2^{-40}$).
- In the SP 800-90B Section 4.5 Criterion (b) half-entropy failure mode, our statistical power is over 99% for $\alpha = 2^{-20}$ and over 92% for $\alpha = 2^{-40}$ using the unmapped data.



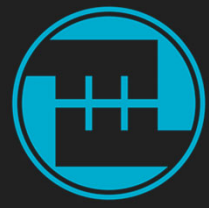
Mapping and the APT: An Example

- If we map the raw data by truncating it to 4 bits, then the entropy for the mapped data would be about 3.98 bits. For a 512-symbol window ($W = 512$) the APT cutoffs are then 63 ($\alpha = 2^{-20}$) to 79 ($\alpha = 2^{-40}$).
- The SP 800-90B Section 4.5 Criterion (b) half-entropy failure mode produces a mapped entropy of about 3.45 bits.
  - This half-entropy failure mode doesn't halve the mapped entropy!
  - The dramatically more probable pre-mapped symbol is now less obvious after mapping.
- Our statistical power is now about 1.5% for the $\alpha = 2^{-20}$ cutoff and essentially 0 for the $\alpha = 2^{-40}$ cutoff.

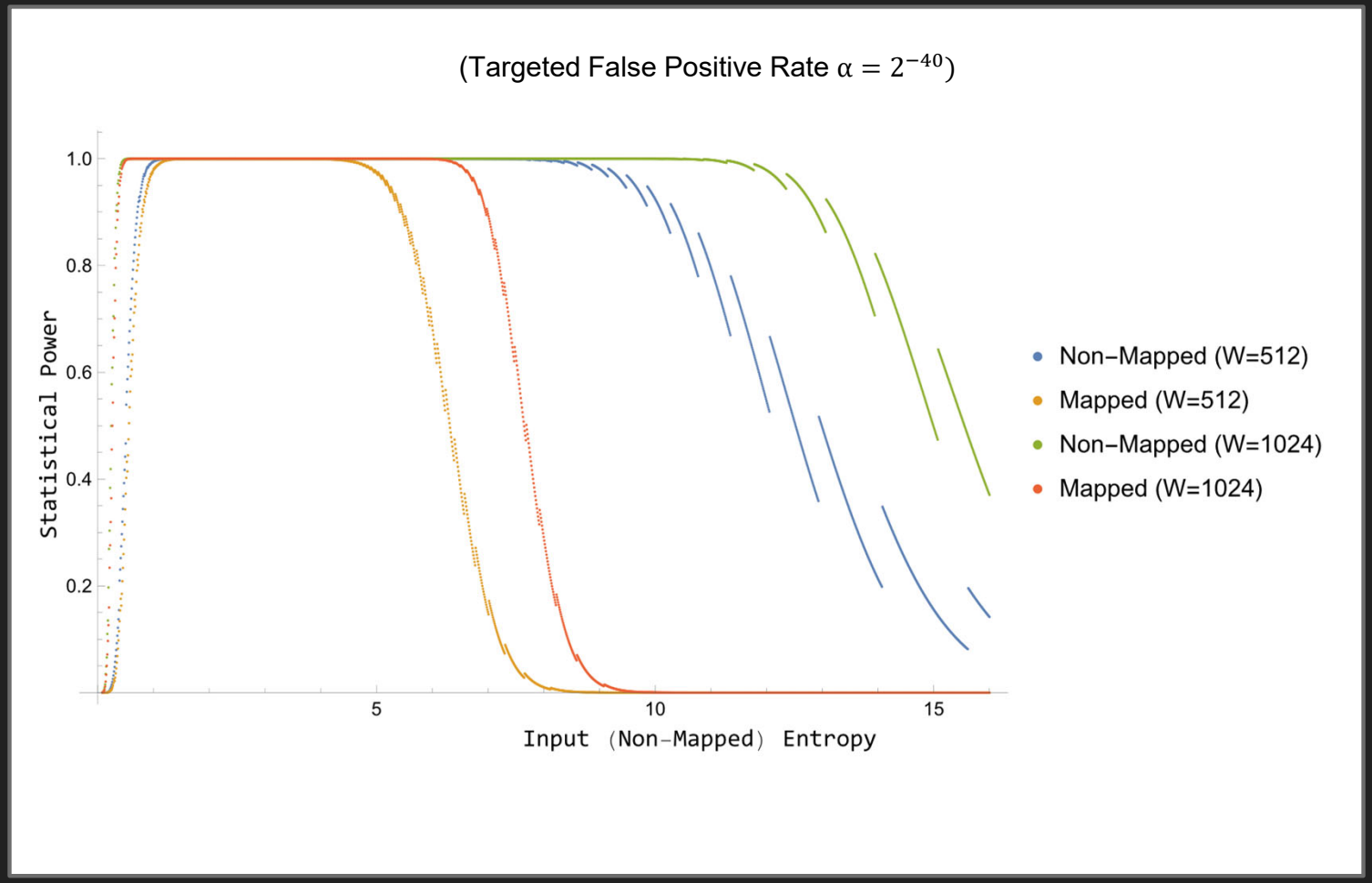


Mapping and the APT: Example Post-Mortem

- In this example, the behavior of the APT using mapped data is dramatically worse than for the APT using unmapped data.
- This occurs across many distributions, so long as there are many values that have a reasonable chance of being output.



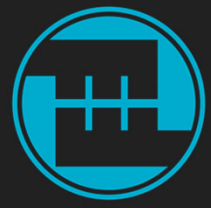
Mapping and the APT: Thousands of Examples



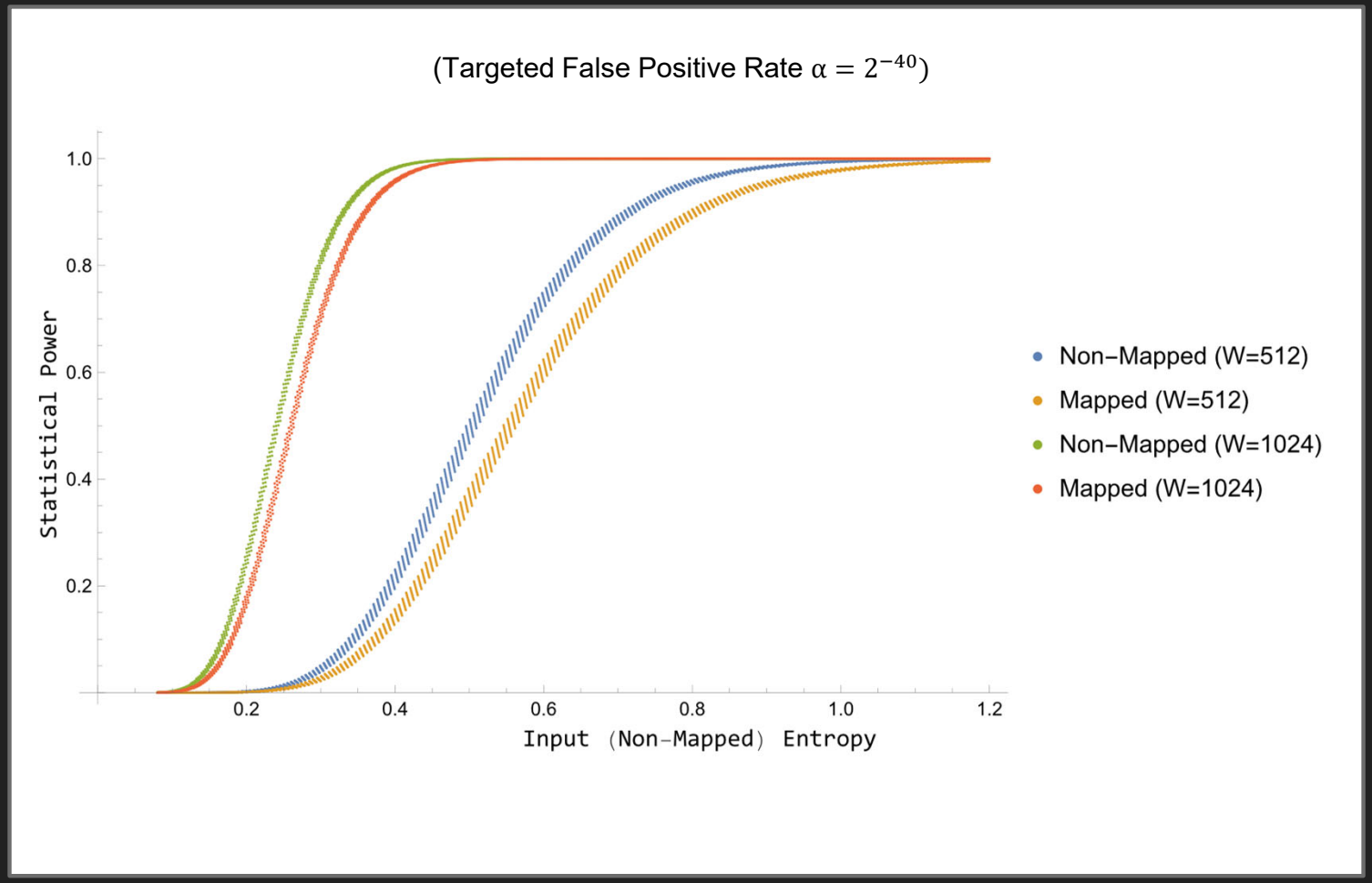
Mapping and the APT: Kilo-Post-Mortem

In these examples:

- For a fixed window size the statistical power for the mapped data is often worse and never better than for the unmapped data.

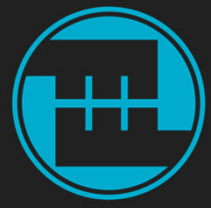


Mapping and the APT: Enhance!



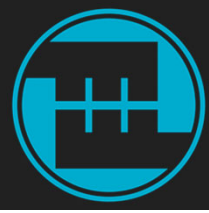
Mapping and the APT: Enhanced-Post-Mortem

- For the low apparent entropy issue (one of the “APT cutoff is too large” issues), increasing the window size helps more than mapping hurts.
- This is because all the other symbols become unlikely when the most likely symbol is very likely to occur.

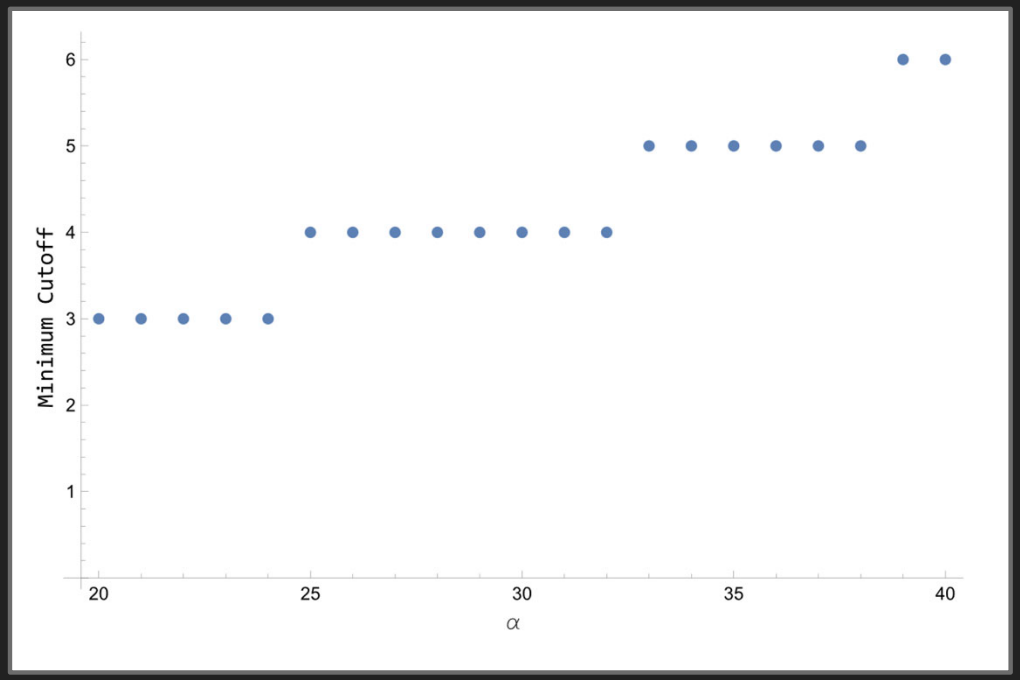


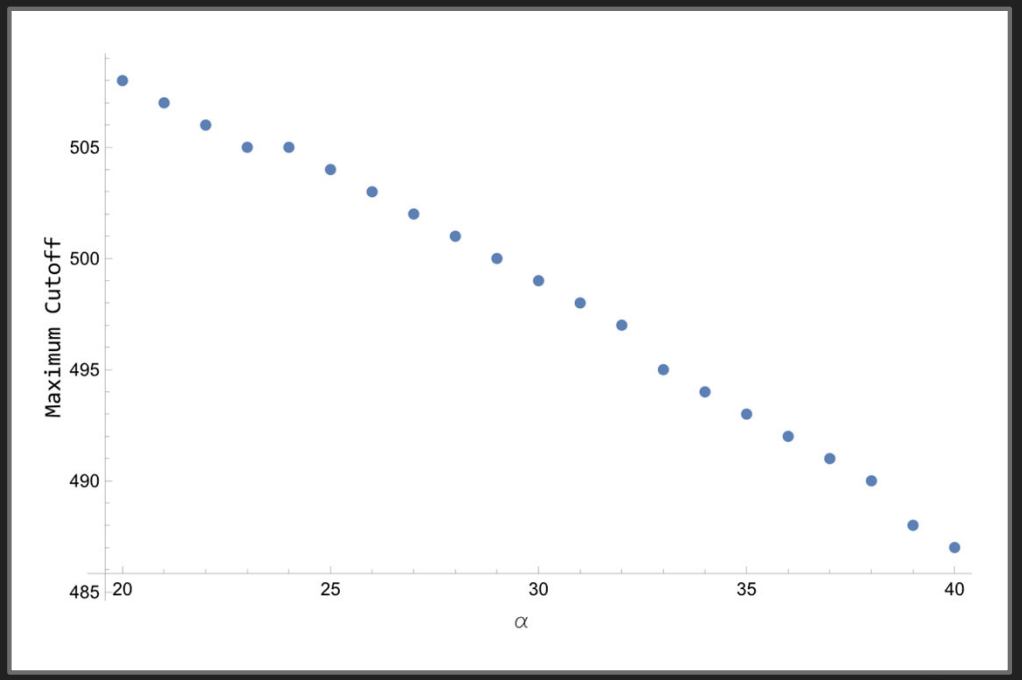
Mapping and the APT: Meta-Post-Mortem

- These examples support the notion that mapping isn't a good general solution to a low statistical power.
- The problem we see here is a lower than desired statistical power when the APT cutoffs are too small or too large.
- In all the tested situations, increasing the window size helps both the "APT cutoff is too large" and the "APT cutoff is too small" issues.
  - The efficacy of the APT tends to increase as the window size increases.
- To satisfy the SP 800-90B Section 4.5 Criterion (b) requirements using the corrected SP 800-90B Section 4.4.2 cutoff procedure and $W = 512$, cutoffs need to be in a range that depends on α.



Mapping and the APT: Meta-Post-Mortem





References

- [HD 2012] Patrick Hagerty and Tom Draper. *Entropy Bounds and Statistical Tests*. 2012.
- [HJ 2019] Joshua E. Hill and Benjamin Jackson. *NIST Special Publication 800-90B Comments*. Version 1.9, December 2019.
- [Draft IG D.K] NIST. *D.K Interpretation of SP 800-90B Requirements*. February 6, 2024.