Statistical Analysis

What do we need?

Outline

- Introduction
  - Frequentist vs Bayesian
  - What others do: ADMX
  - What others do: HAYSTAC
- What I think we need
  - Basic pipeline
  - Building Blocks

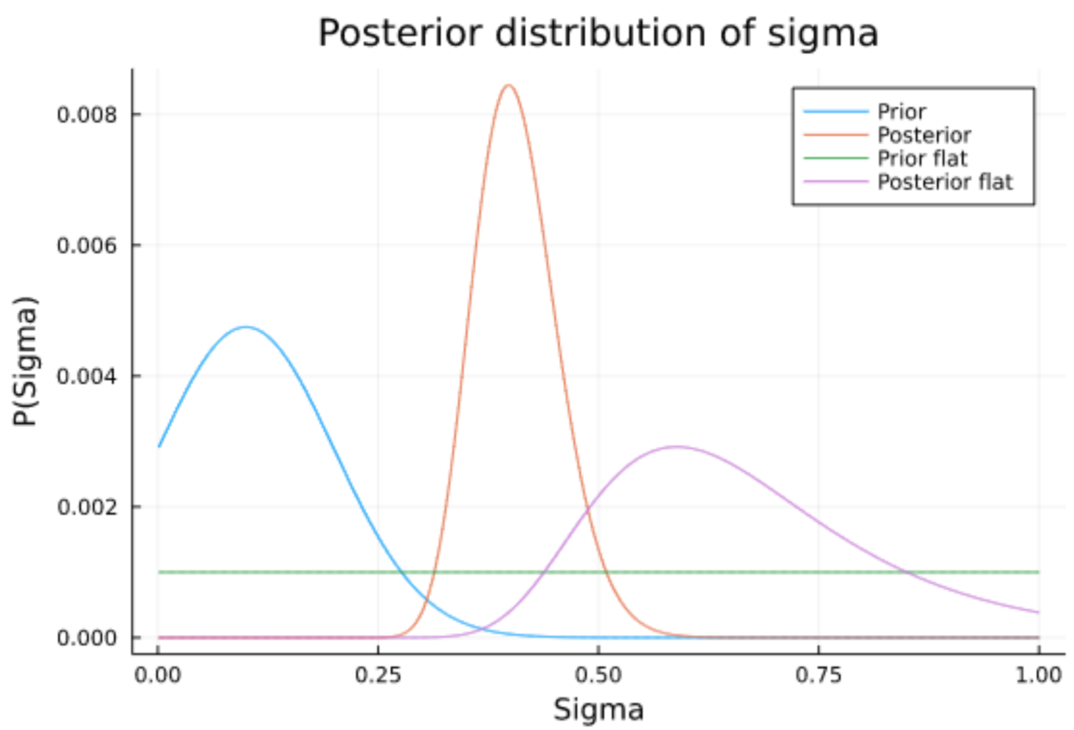

BAYESIAN VS FREQUENTIST

Introduction: Freq v Bayes

- Your friend is throwing one dart at a dartboard. How far away from the center is the dart going to land P(r, θ) = ?? Where θ are some parameters.
- He throws infinitely many times → P(r) known exactly
- He throws 10 times:

```
r = [0.02, 0.05, 0.11, 0.13,
     0.25, 0.38, 0.51, 0.67,
     0.91, 1.30]
```

- P(r, θ|data) = ?



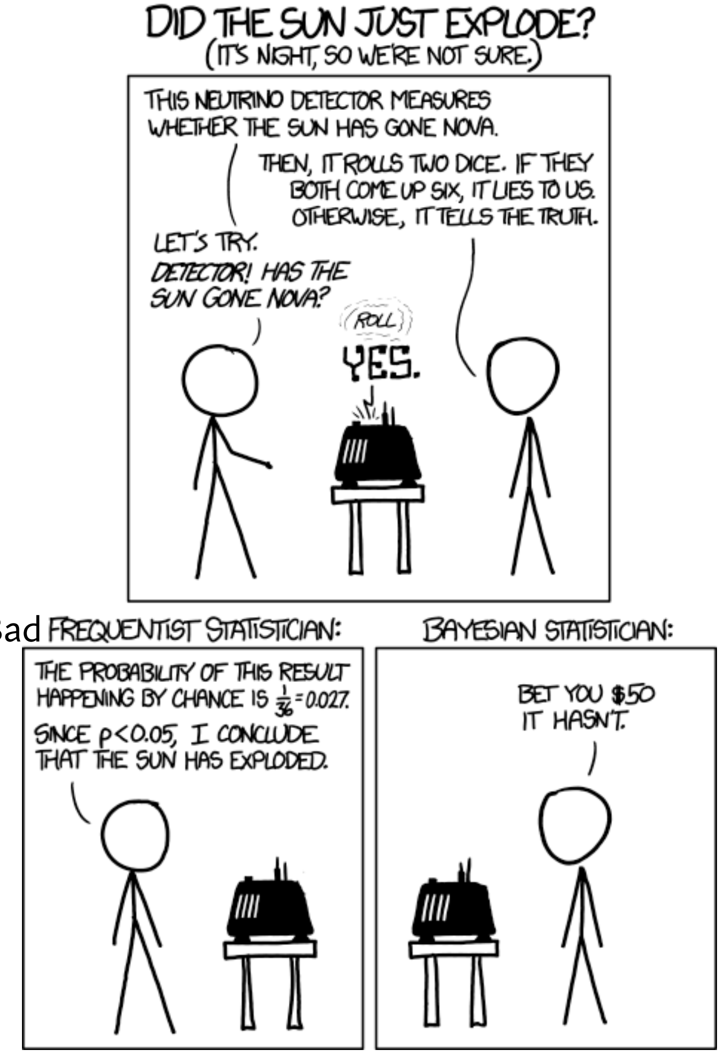
Introduction: Freq v Bayes

- Frequentist:
  - The parameter is **not** a random variable
  - Find e.g. maximum of P(data|θ)
  - Confidence interval “If I repeat the measurement 100 times, I will be right 95 times” (but usually I know nothing about the 99 measurements I did not make)
- Bayesianist:
  - The parameter is a random variable
  - P(θ|data) = P(data|θ) × P(θ) / P(data)
  - Credible interval “95% of the probability density of θ after looking at the data that I have is in this interval” (but if I did the measurement 100 times it is possible that the true value lies in this interval only in 10% of cases)

Introduction: Model selection

- Frequentist:
  - Data ”is” random variable, gaussian distribution around the correct model
  - Compare models using a test statistics e.g. by calculating reduced χ^2 taking degrees of freedom into account
  - Test statistic needs to have known distribution, so e.g. difference in reduced χ^2 can be converted to p value
- Bayesian:
  - Compare Bayes Factor for two models/ hypothesis M_1 and M_2: $\frac{P(\text{data}|M_1)}{P(\text{data}|M_2)}$. This is not (always) equal to frequentist likelihood ratio, because $P(\text{data}|M_1) = \int P(\text{data}|\theta_1, M_1)P(\theta_1|M_1)d\theta_1 \neq \max_{\theta_1} P(\text{data}|\theta_1 M_1)$.
  - Posterior odds: Bayes Factor $\times \frac{P(M_1)}{P(M_2)}$. Problem: What is P(“axion at this frequency”)? (Allen had some ideas)
- Be careful with degrees of freedom!

Introduction: Freq v Bayes



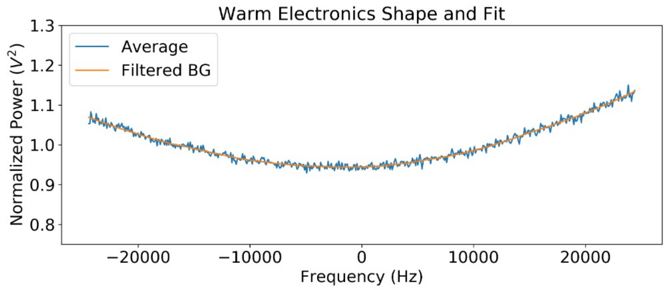
Bad

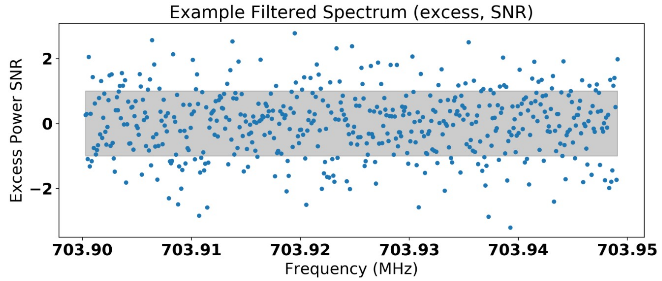
- Easy reading resources:
  - http://jakevdp.github.io/blog/2015/08/07/frequentism-and-bayesianism-5-model-selection/
  - https://stats.stackexchange.com/questions/2272/whats-the-difference-between-a-confidence-interval-and-a-credible-interval

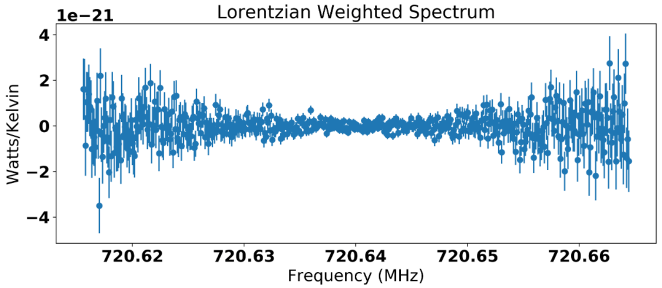
ADMX Run 1b

https://arxiv.org/abs/2010.06183

- Baseline removal via SG filter
- Convolution with 6th order Pade filter for some more background removal
- Divide by mean and subtract 1 → white noise (+ axion?)
- Lorentz weighting to account for signal power drop away from cavity resonance
- Various manual cuts throwing away many raw spectra
- Construct combined spectrum merging measurements bin by bin accounting for differences in signal power and noise temperature
- Filter with Axion Lineshape from Maxwell Boltzmann velocity distribution
- Optimal Weighting process from HAYSTAC







ADMX Run 1b ctd

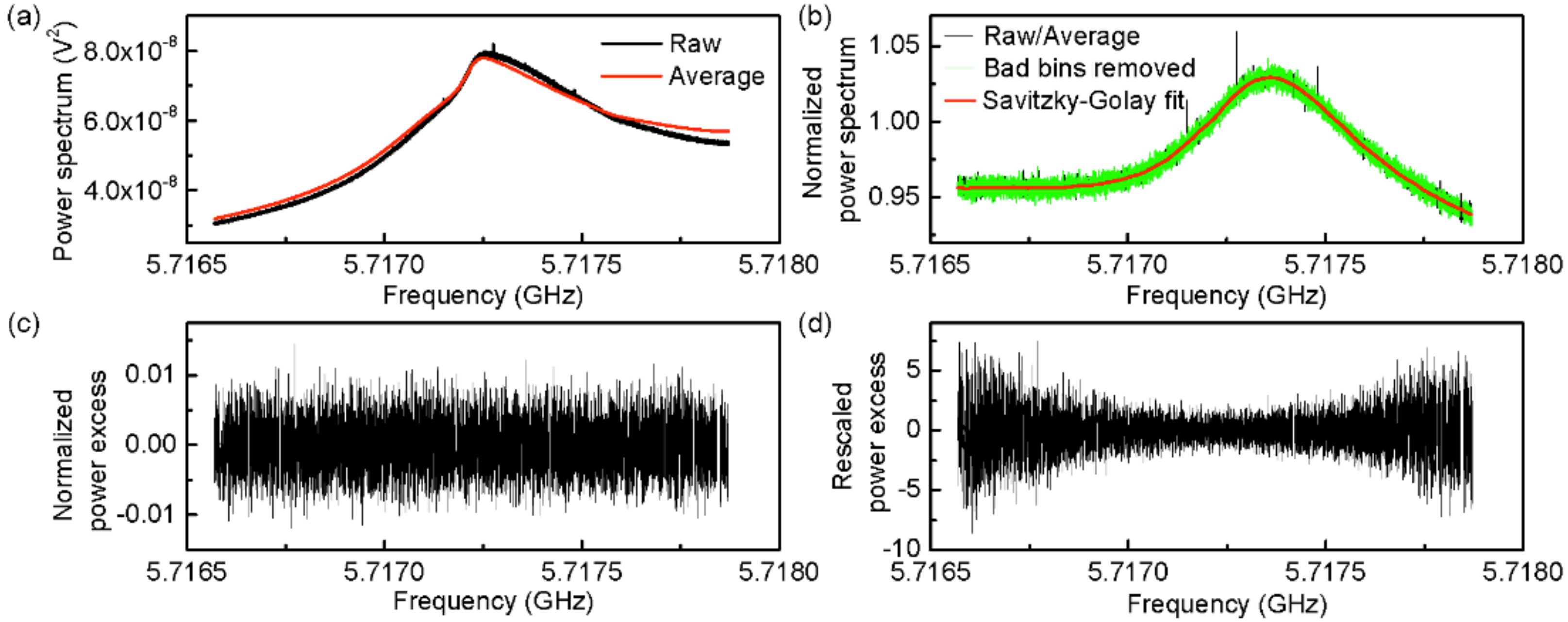
https://arxiv.org/abs/2010.06183

- 1773 evenly-spaced, DFSZ2-like synthetic software axions, 20 / MHz
- Thereof 1684 found (~95%) with average of ~81.8% of input power → effect from signal processing
- Hardware synthetics (not present in Run 1a) → not relevant right now
- Rescan procedure → not relevant right now
- Axion results in $\mu = g_{a\gamma}^2 \eta$ with SNR η, no axion results in $\mu = 0$.
- Limit = μ for which 90% confidence limit that measurement does not contain axion “CDF for truncated normal distribution” → no idea what they were actually doing here
- Resolution too fine for plot → combine 200 bins using 100 samples from each bin (assume Normal with measured power as mean and measured uncertainty as std), clipping all below 0, combining 2k samples, using 90% CDF

HAYSTAC

https://arxiv.org/abs/1706.08388

- Various cut using auxilliary data throwing away few spectra
- Subtract average baseline obtained using SG fit
- Remove bins with IF interference
- Use SG fit to remove residual baseline* (divide, subtract 1)
- Multiply freq dependent noise Temp and signal Power



*SG filter doesn't allow for gaps

HAYSTAC ctd

https://arxiv.org/abs/1706.08388

- Bins in processed spectra follow normal distribution with known mean, std (in absence of axions, but correlations from SG filter possible)
- Use central limit theorem to investigate influence of correlations
- Combine bins with same freq using Maximum Likelihood weights* (seems only relevant if signal power and noise temp differ between raw spectra)
- Combine bins with adjacent freqs according to axion lineshape using ML weights (raw bin width 100 Hz, axion linewidth 5 kHz)
- Define rescan threshold as tradeoff between number of rescans vs. excludable SNR
- Quantify signal attenuation from SG fit (basically did that already)
- Injected 10 hardware axions

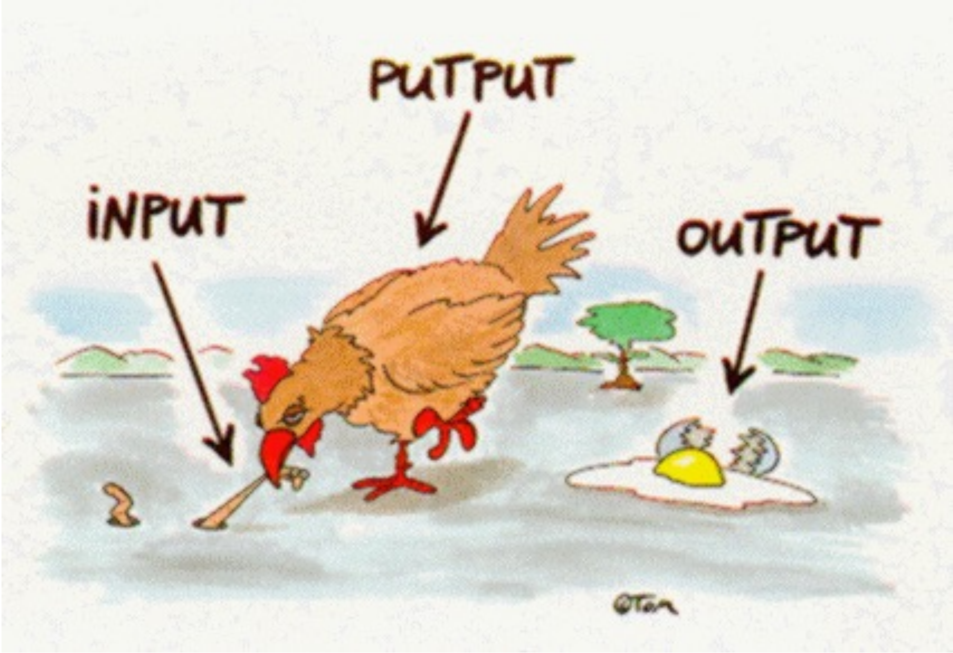
*https://towardsdatascience.com/probability-concepts-explained-maximum-likelihood-estimation-c7b4342fdbb1

Questions?

Basic (Statistics) Pipeline

- Input:
  - Raw power spectrum (uncertainty on T sys)
  - Boost factor (uncertainty)
  - magnetic field
  - disk size
  - information on data cuts
  - Amplifier Impedance for Y factor calib
  - Influence of RF response on axion signal shape
- Putput:
  - SG filtering
  - Statistical cross checks
- Output:
  - Freq dependent $|g_{a\gamma}|$ confidence limit

$$P(\omega)\mathrm{d}\omega = \frac{\rho_a}{m_a^2}\, g_{a\gamma}^2\, B_e^2\, A\, \beta(\omega)\, \frac{q_e}{\hbar}\sqrt{2/\pi}\,\frac{v(\omega)}{\sigma_v v_{\mathrm{lab}}}$$
$$\times \exp\left(-\frac{v(\omega)^2 + \ v_{\mathrm{lab}}^2}{2\sigma_v^2}\right)\sinh\left(\frac{v(\omega)v_{\mathrm{lab}}}{\sigma_v^2}\right)\mathrm{d}\omega$$



Putput

Vertical stacking

Background subtraction

Limit setting

Synthetic Axion

Gaussianity tests

Coverage test

Other tests?

Vertical Stacking

- Not entirely sure what HAYSTAC is doing, but just adding bin by bin should be fine?
- Maybe some effect from frequency drift of Booster dip position in time? T sys differences between different measurements?
- Different samplers of 2022? Probably no problem, just average!

Synthetic Axion

- Should have equal SNR to have good statistics for one specific SNR
- Problem: Use which Boostfactor? One set with known, one set with Boostfactors drawn from distribution?
- Throw in software synthetics with specific power (instead of gag)
- Problem: Quantifying influence of response to axion peaks requires hardware synthetics. Can we do that for 2023, 2022? 22 Transfer function width munch narrower than signal shape (see is the bump significant, Bela tries to find data), 23 vice versa (data exists for narrow gaussian peak)

T calibration

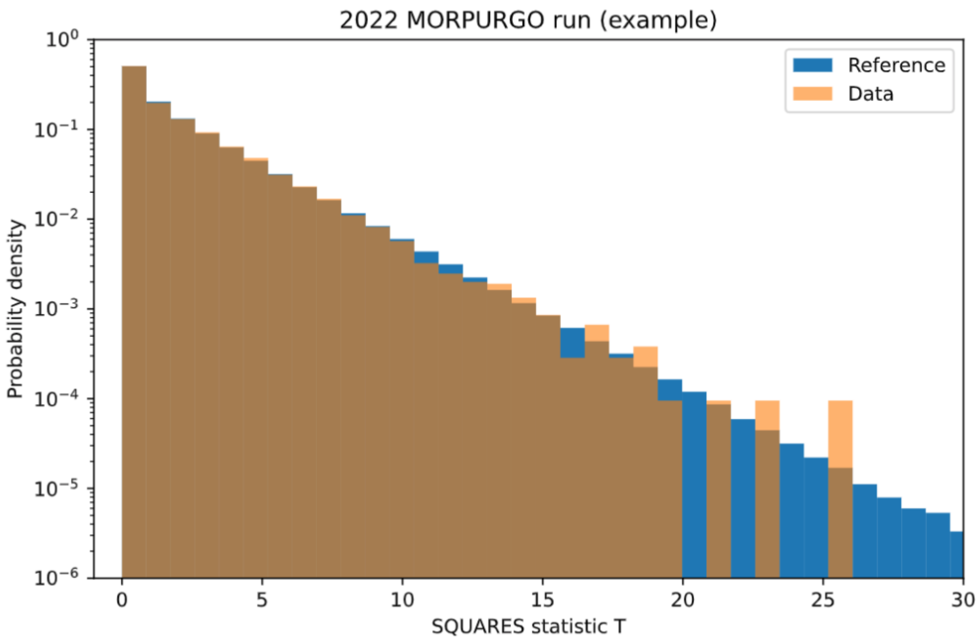
- Major issue: Investigate how to deal with Olafs non-linearity for 2022 data -> Three independent data (sim, 50Mhz, 2GHz) for T sys, decide which to trust? Take difference as uncertainty? Power calibration -> Bernardo
- Y factor method filter and quantify influence

Background subtraction

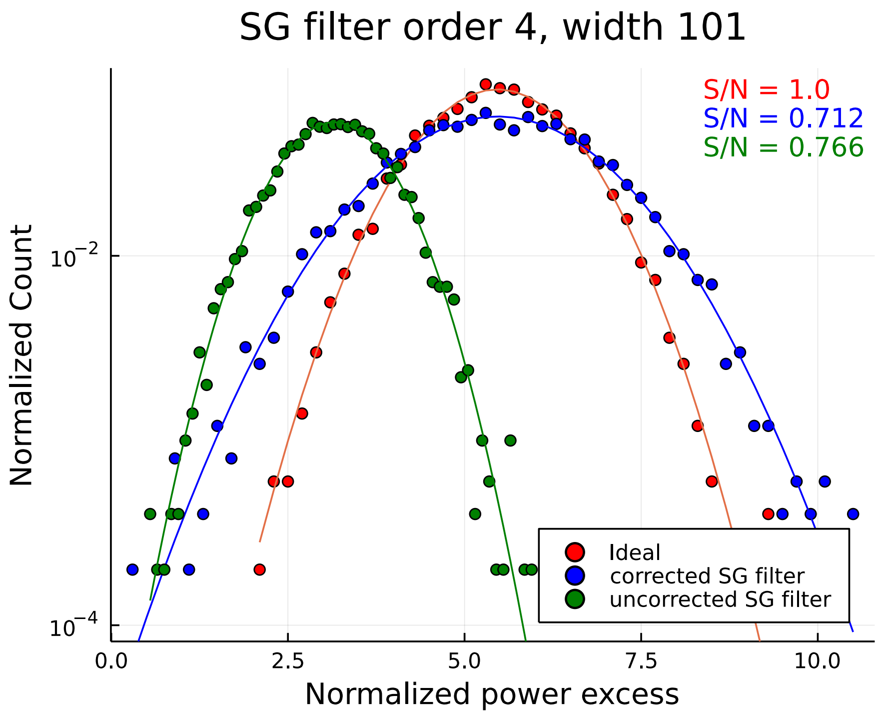
- Use SG filters. Simple. Fast.
- Demonstrate effect of different SG filters for Appendix using (HAYSTAC) Gaussianity test (basically coded up)
- Investigate possibility of parametric background fit for 23 data -> David

Gaussianity Tests

- Investigate Gaussianity of baseline (no synthetic axions) after BG subtraction (coded up using SQUARES statistics)

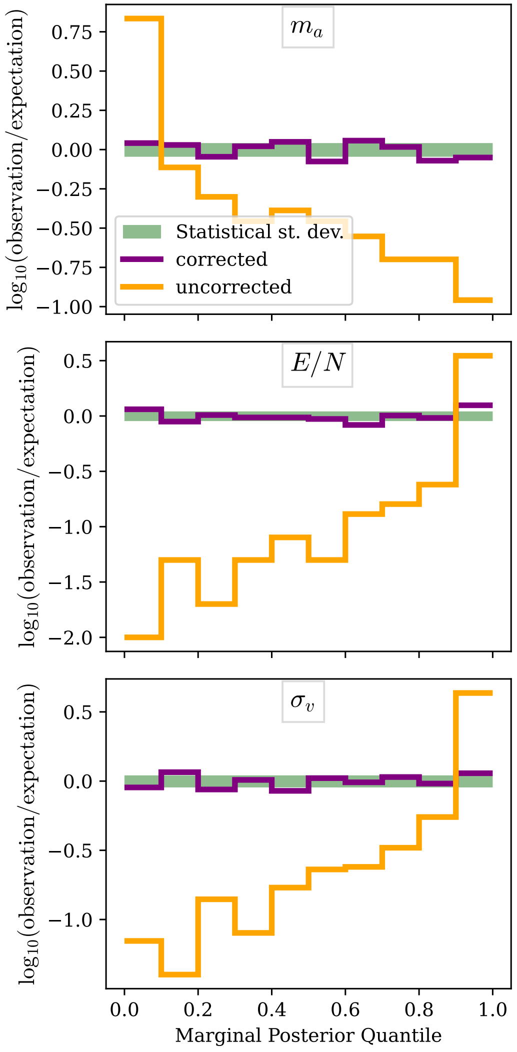


- Investigate Gaussianity of synthetic axion signal strength (coded up using HAYSTAC test)



Coverage test

- If synthetics with different parameters are used, recovery of these parameters can be investigated



Limit setting

- Quick&Dirty:
  - Define sth like a 95% confidence limit on the boostfactor
  - Do bin-by-bin Frequentist Hypothesis test between $g_{a\gamma} = 0$ and $g_{a\gamma} = x$ where we try to find x so that p-value of $g_{a\gamma} = x$ is 0.05
- Bayesian(&Dirty):
  - Need to define "prior"* distribution for boostfactor, set ρ_a fixed to be comparable to other experiments (limit has hugely different interpretation though)
  - No normalizable ALP prior in reachable $g_{a\gamma}$ range
  - Run nested sampling (already coded up), define (95% of CDF in each freq bin as limit)
  - Can use coverage test

*it depends on data so not a real prior, but for the sake of limit setting it is