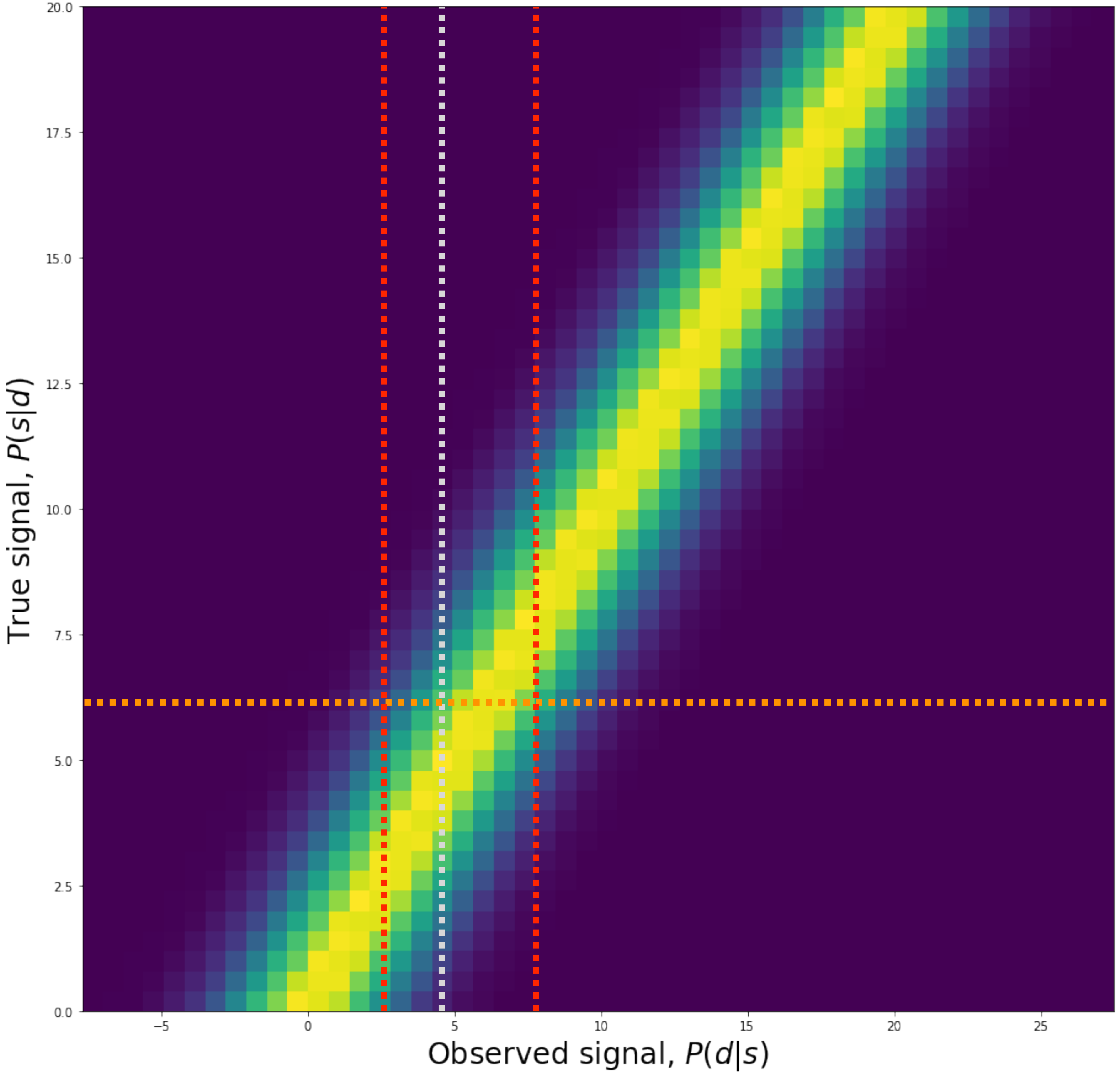
Cautionary examples of statistical errors

Miguel F. Morales
Bryna Hazelton

Review: how to calculate a confidence interval/ upper limit

Simulated observations

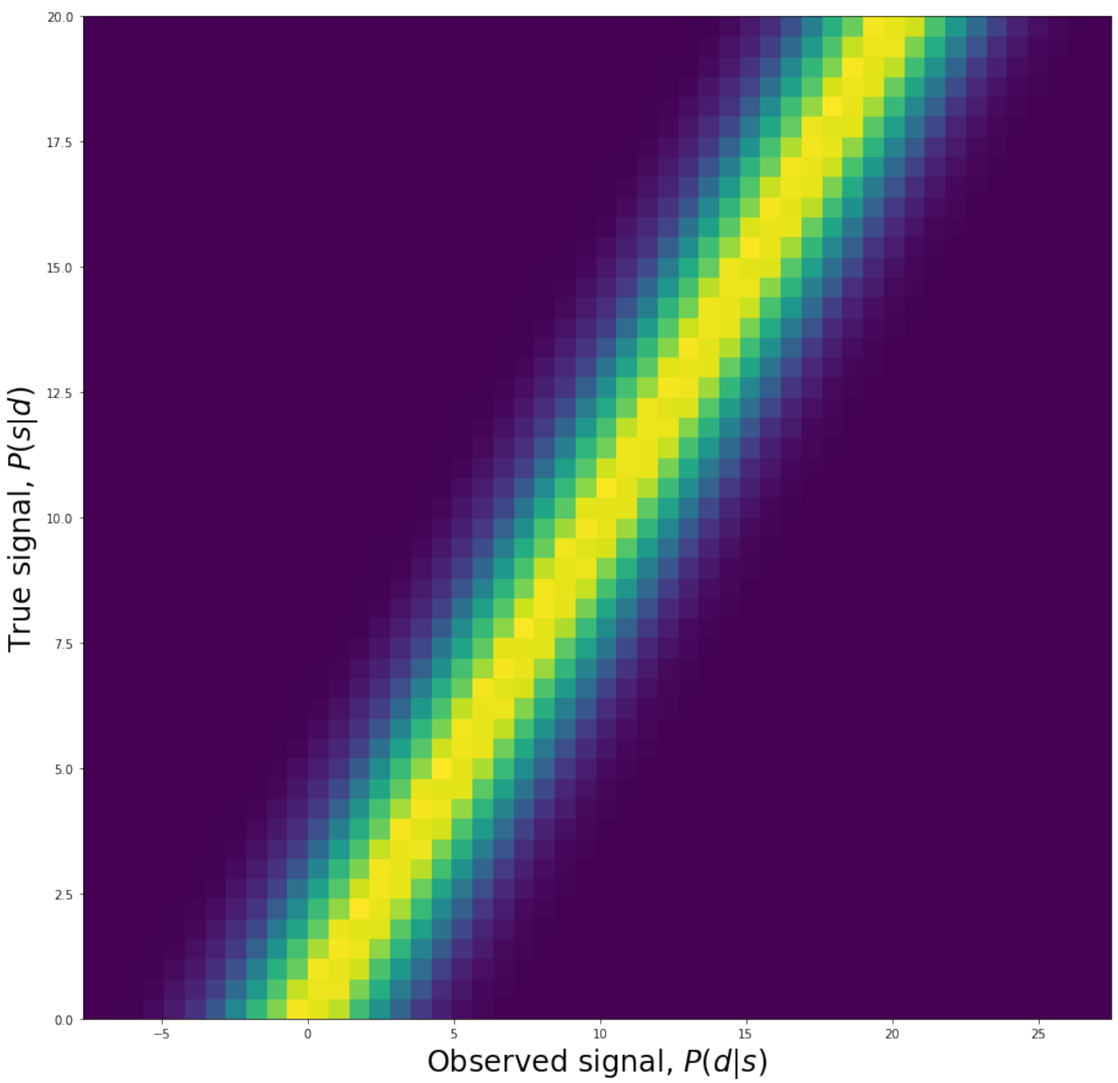


How to calculate

- Simulate: determine $P(d\,|\,s)$ for every signal value

- Signal injection: take signal free data and inject a face signal to see what it looks like; repeat for every signal value

- Use math: Bayes' theorem



Bayes' theorem

$$P(s\,|\,d) = \frac{\overbrace{P(d\,|\,s)}^{\text{Physics Model}}\overbrace{P(s)}^{\text{Prior}}}{\underbrace{\cancel{P(d)}}_{\text{Normalization}}}$$

Bayes' theorem

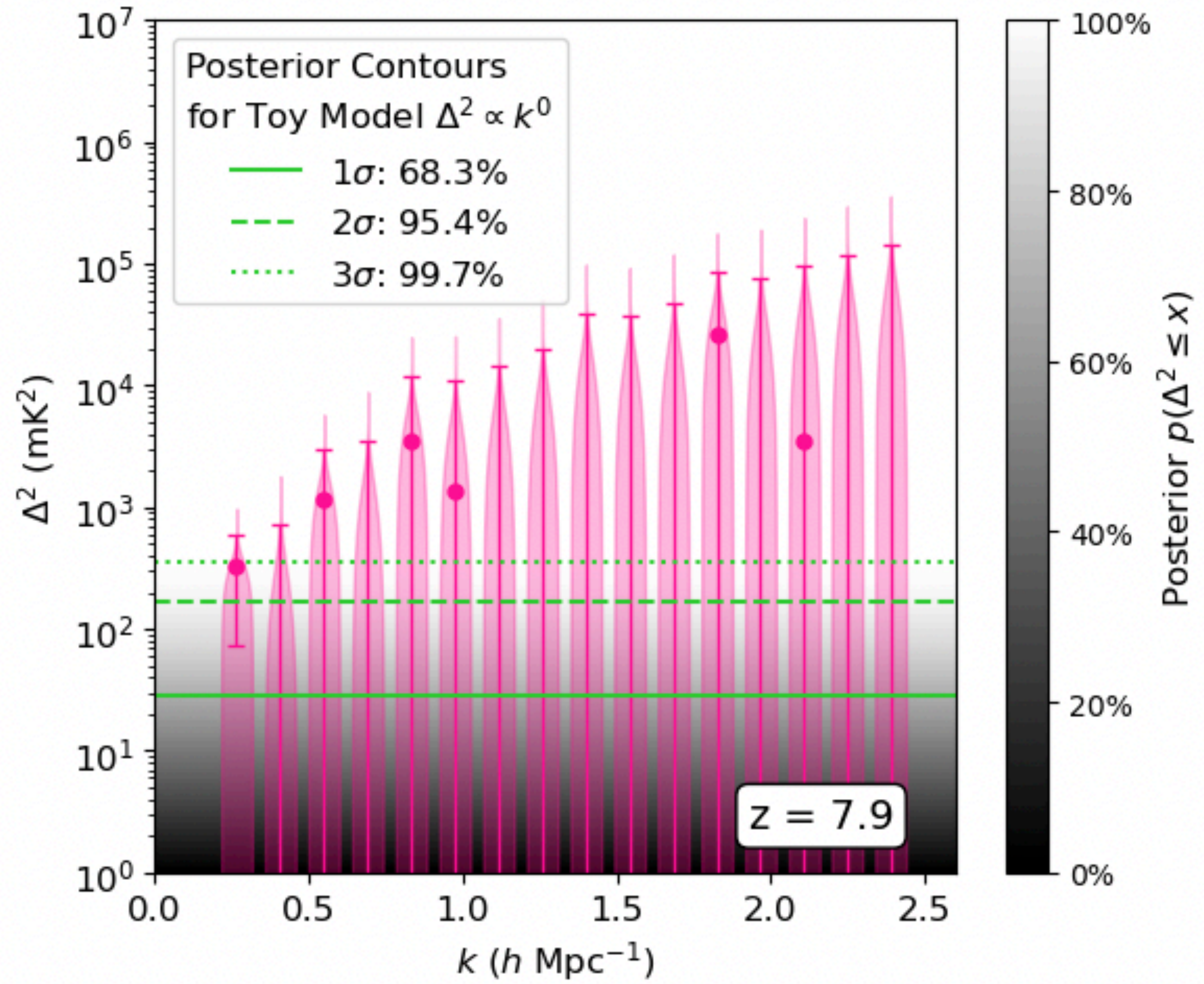
$$P(s\,|\,d) = \frac{\overbrace{P(d\,|\,s)}^{\text{Physics Model}}\overbrace{P(s)}^{\text{Prior}}}{\underbrace{\cancel{P(d)}}_{\text{Normalization}}}$$

- Always interpreting in the framework of a model
    - May be generic/data driven, e.g. a 'top hat pulse'
- Prior necessary to invert
    - Powerful when driven by knowledge
    - ***Dangerous*** when driven by assumptions

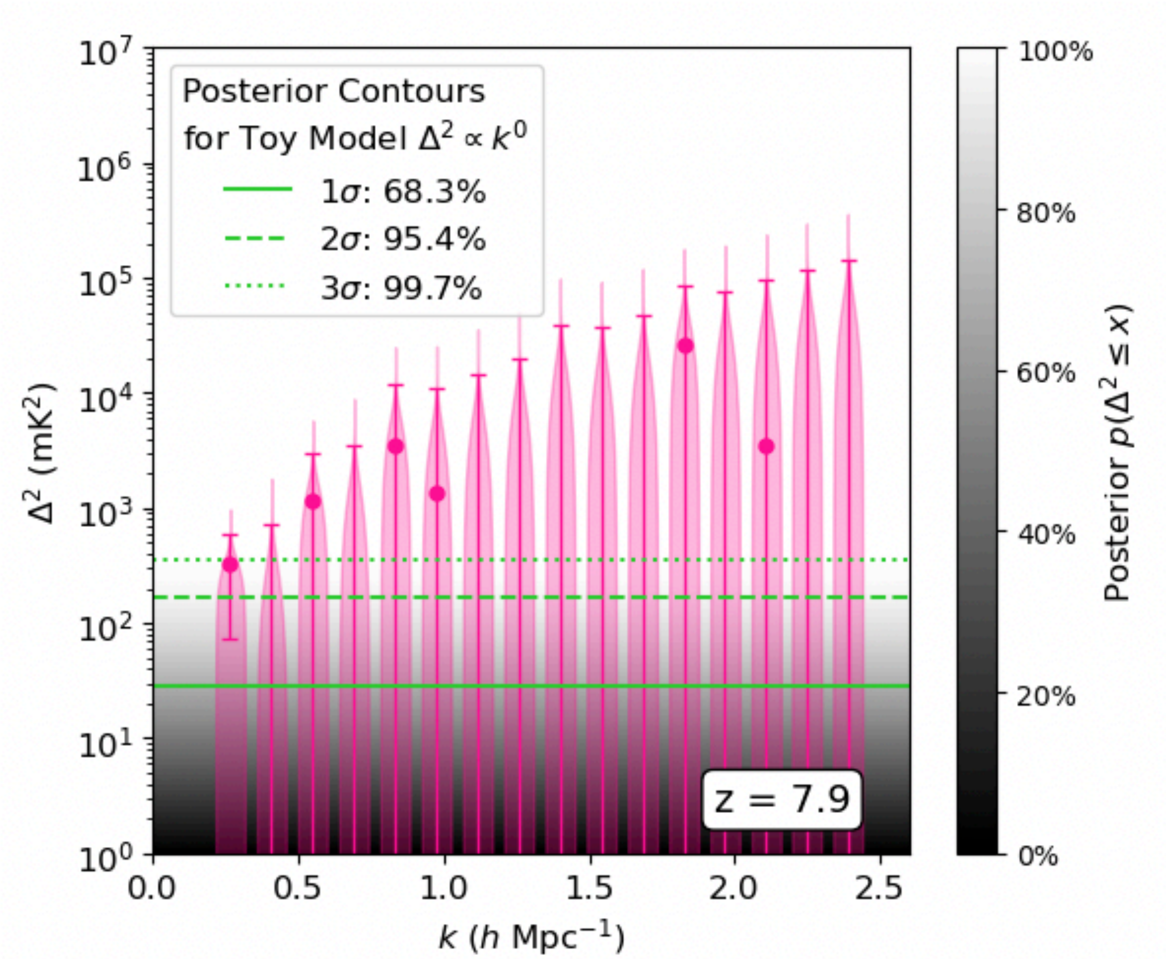
Implicit or explicit logarithmic priors

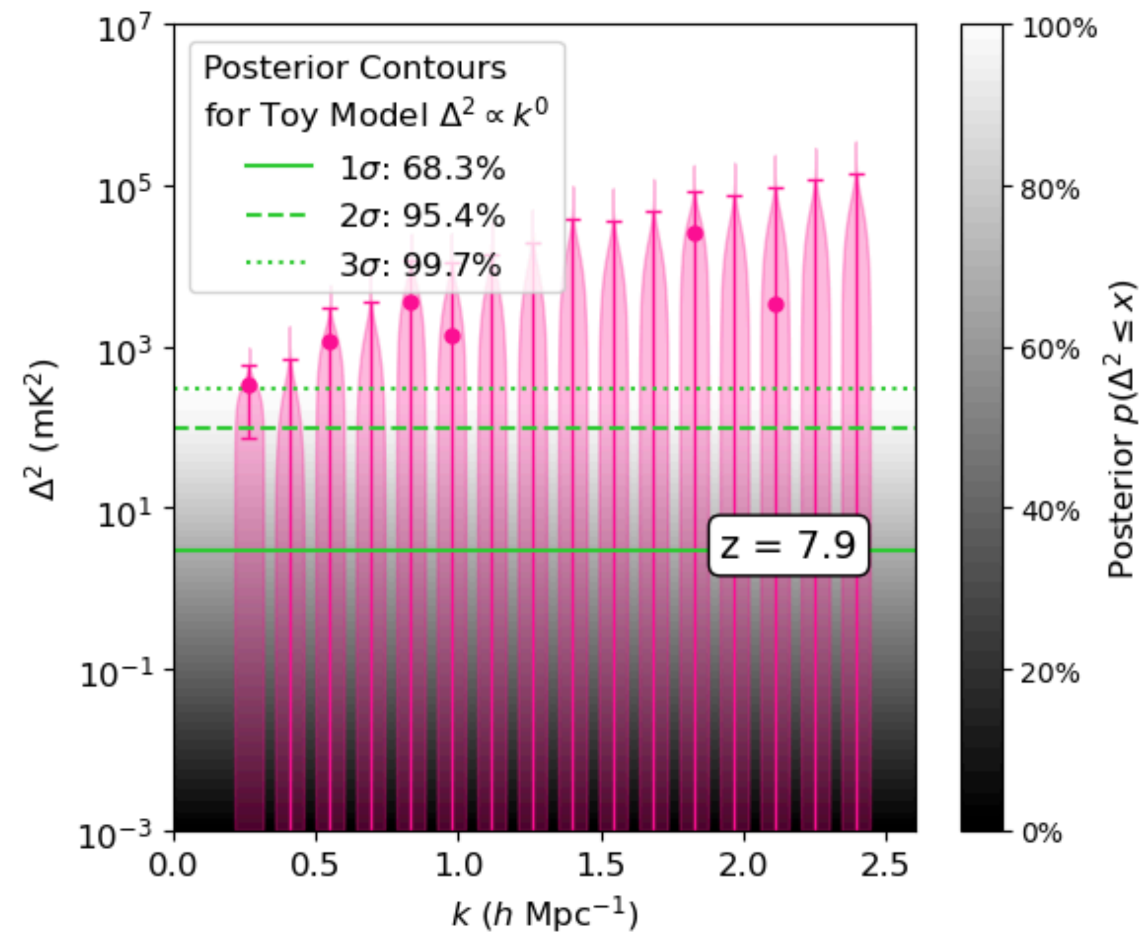
Case study

- Combined upper limit combining many limits
- Use Bayes' theorem to invert
- Logarithmic prior



I asked to change lower bound from 1 to 1e-3

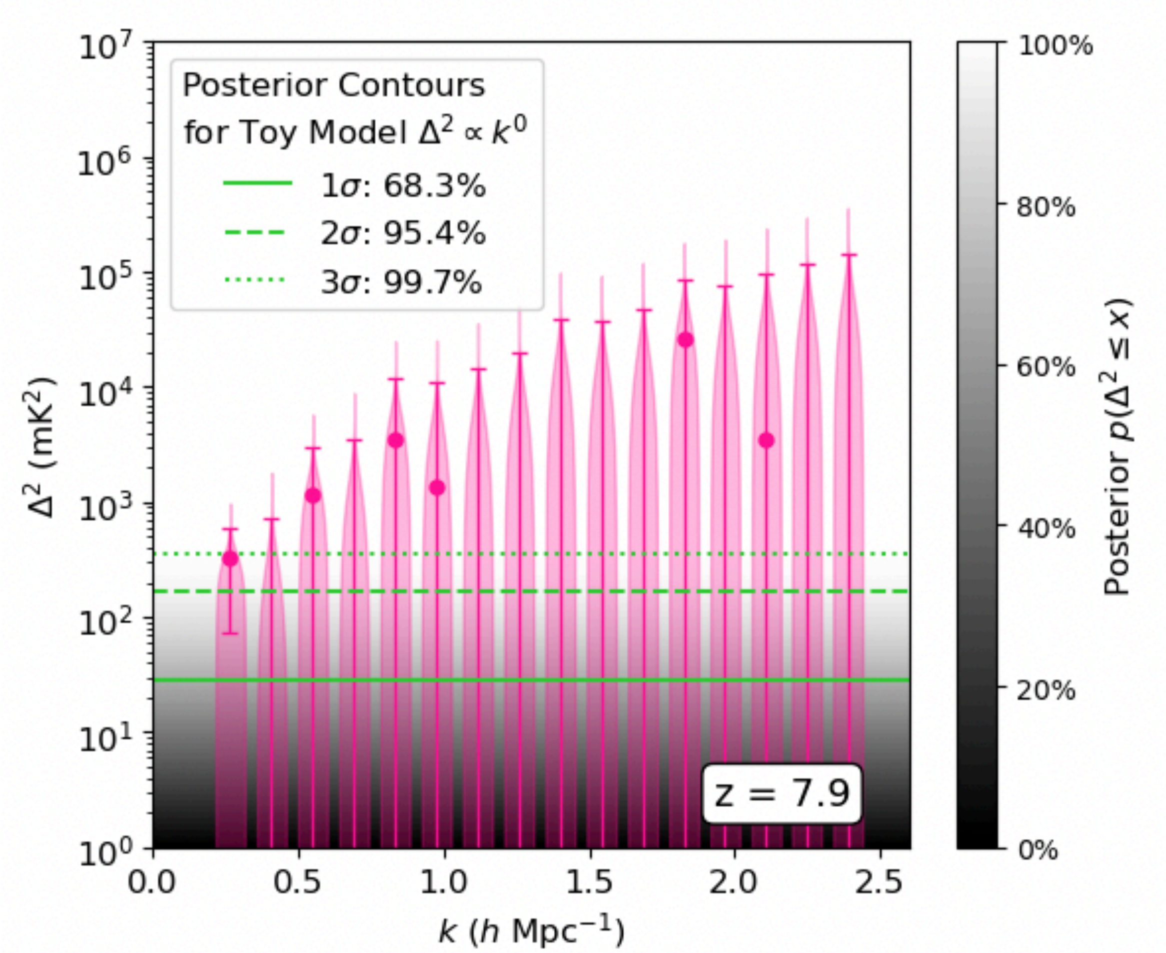




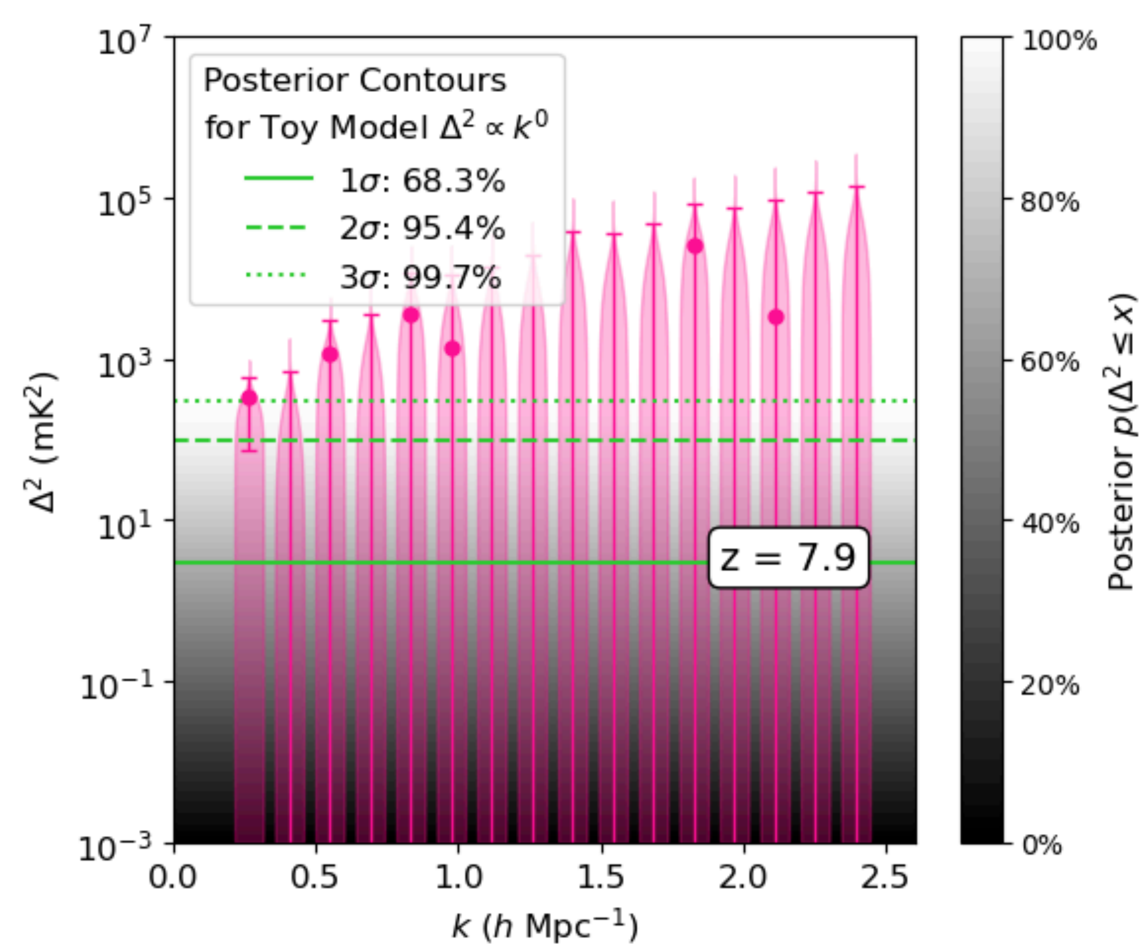
2σ upper limit: $16.7\ \mathrm{mK}^2$

2σ upper limit: $9.7\ \mathrm{mK}^2$

Moving ***lower*** prior bound by <1 mK² changed ***upper*** limit by 7 mK²!

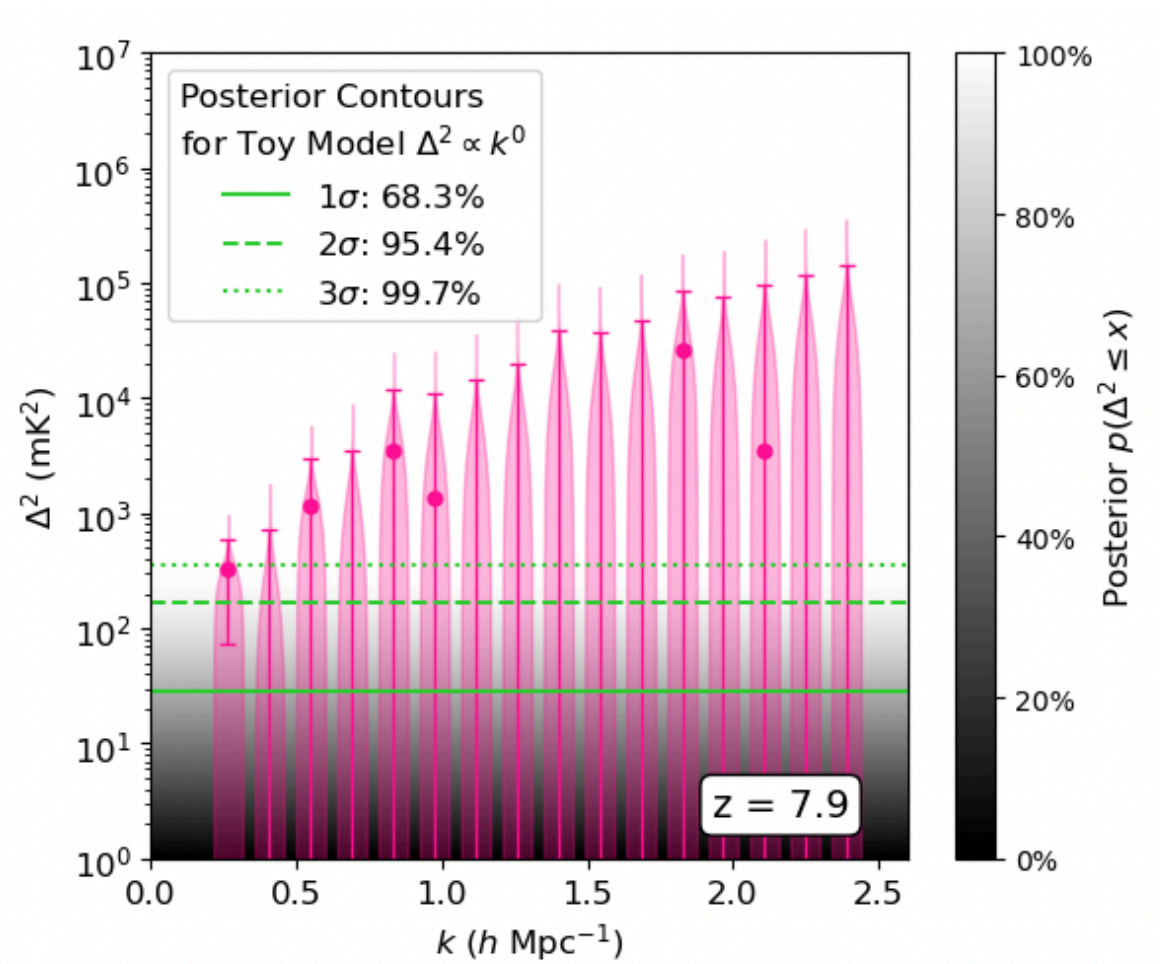


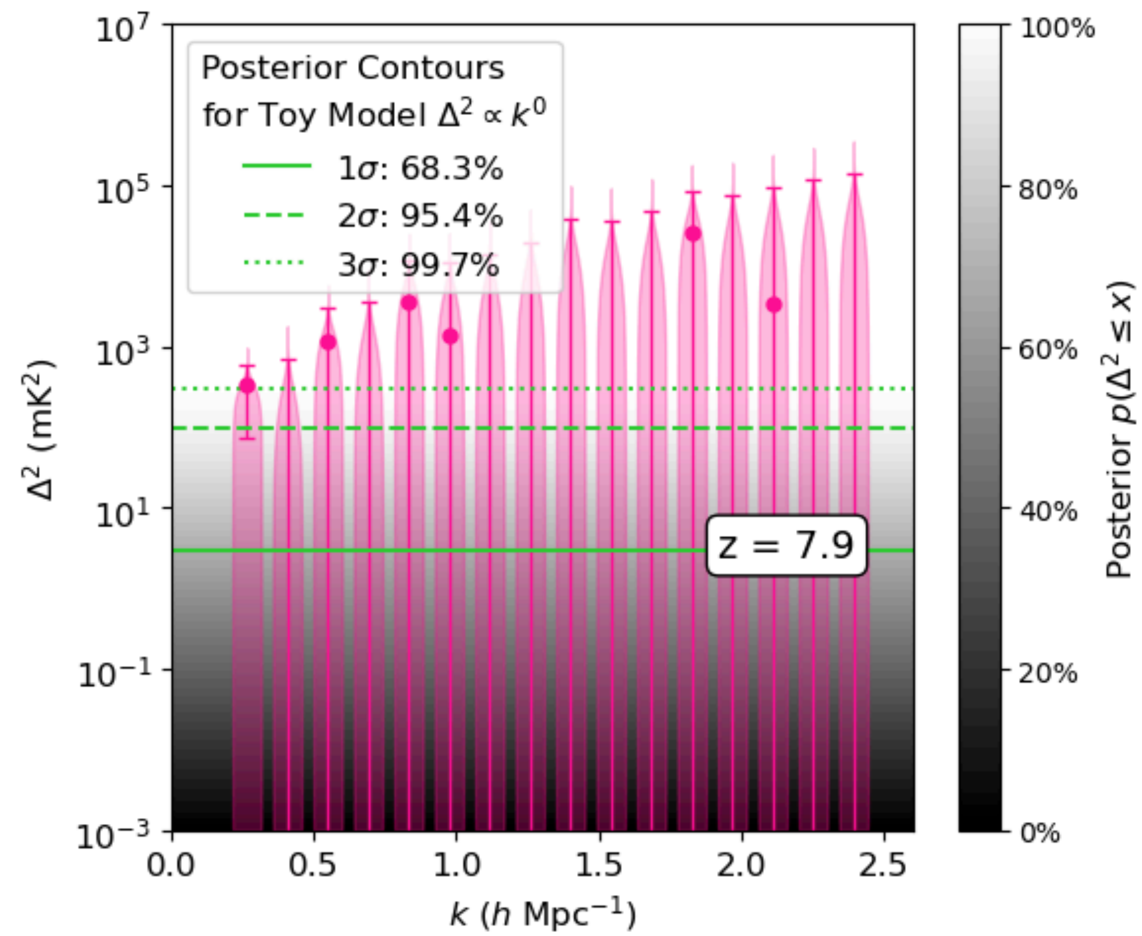
2σ upper limit: $16.7\ \mathrm{mK}^2$



2σ upper limit: $9.7\ \mathrm{mK}^2$

Which was right?





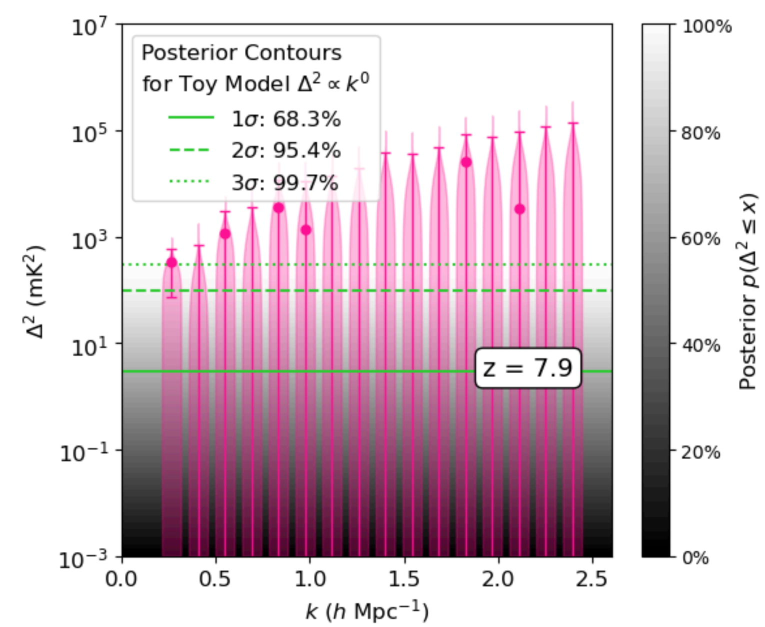
2σ upper limit: $16.7\ \mathrm{mK}^2$

2σ upper limit: $9.7\ \mathrm{mK}^2$

Neither!

$$P(s\,|\,d) = \frac{\overbrace{P(d\,|\,s)}^{\text{Physics Model}}\,\overbrace{P(s)}^{\text{Prior}}}{\underbrace{\cancel{P(d)}}_{\text{Normalization}}}$$

Log prior in linear space

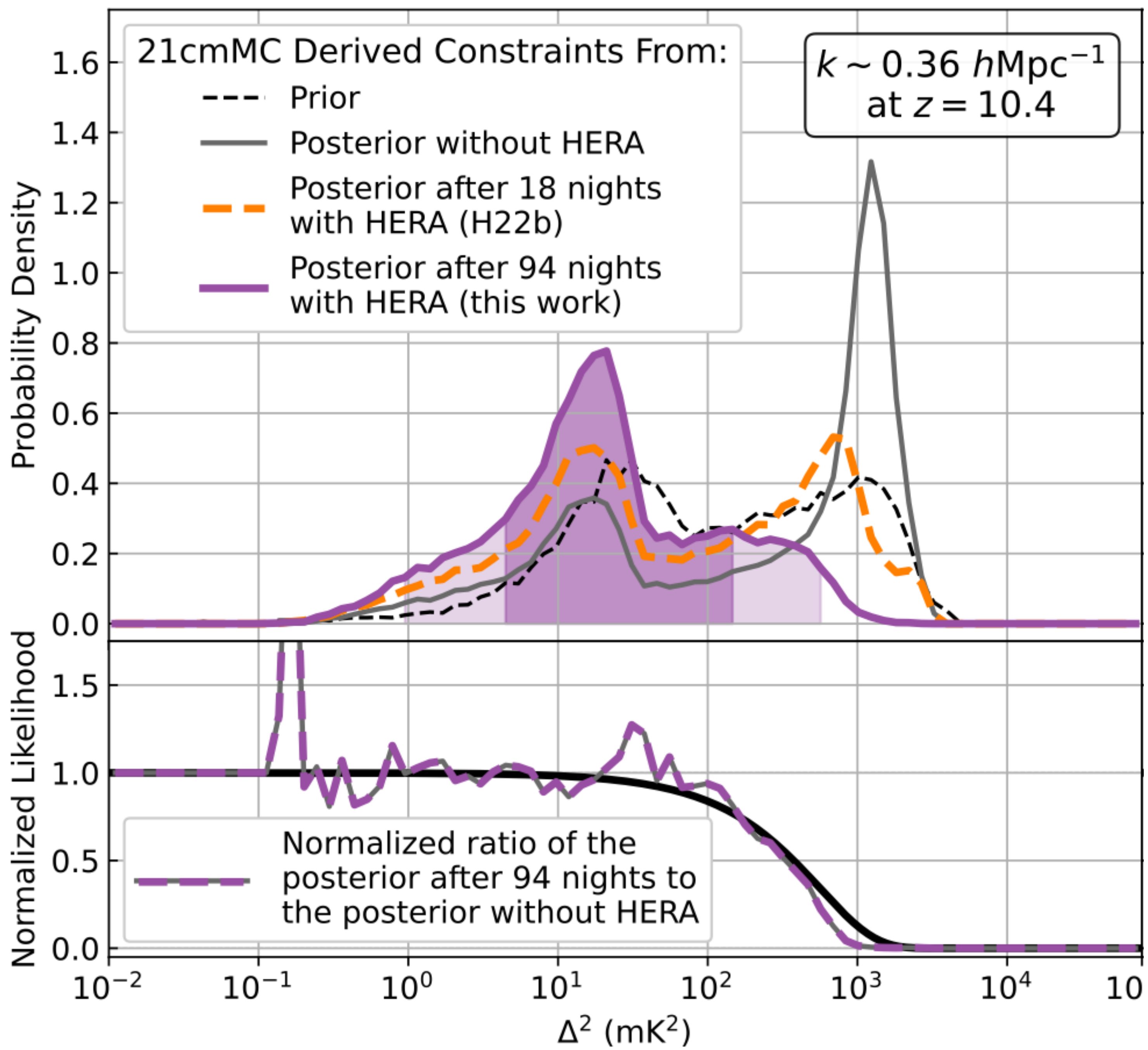


Logarithmic priors are shockingly dangerous

- A lower value must be chosen so $\int P(s) = 1$
- Upper and lower values tightly linked by normalization requirement
- Most common bad statistical mistake I see
- Rule of thumb: never use logarithmic priors

Asking the wrong question

Misusing Bayes' theorem



3 questions

- In what region does the brightest 5% of the models live?
  - $\int_{?}^{\infty} P(d\,|\,s)P(s)\,ds$

Theory landscape question

- Given only our physics model, what are the brightest 5% of the models consistent with our data?
  - $\int_{?}^{\infty} P(d\,|\,s)(\text{uniform prior})\,ds$

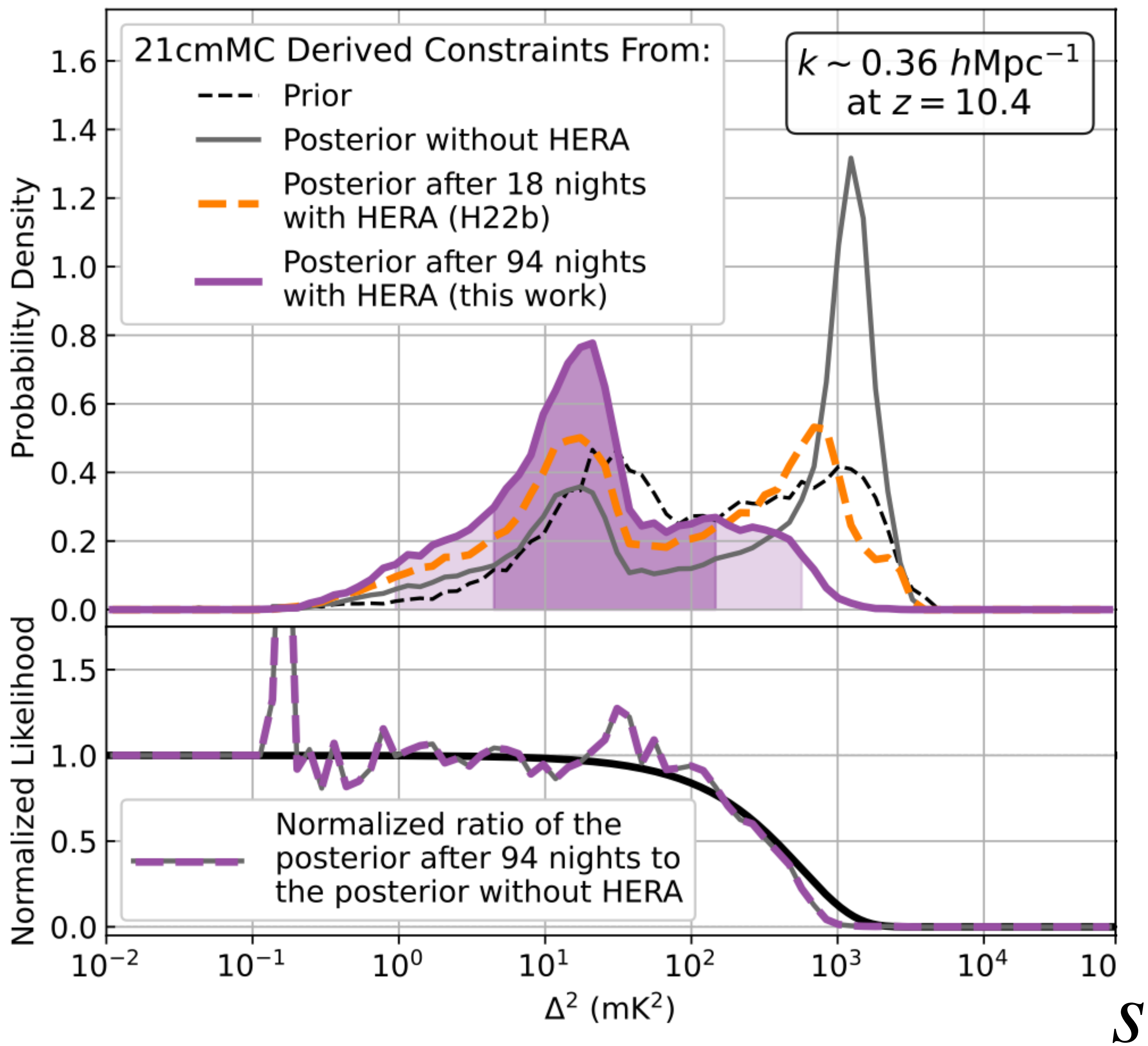
Observer's question v1

- If we repeated our experiment, at what signal level would we have detection 95% of the time?
  - $\int_{d}^{\infty} P(d\,|\,s_?)\,\partial d$

Observer's question v2

Using Bayes' theorem

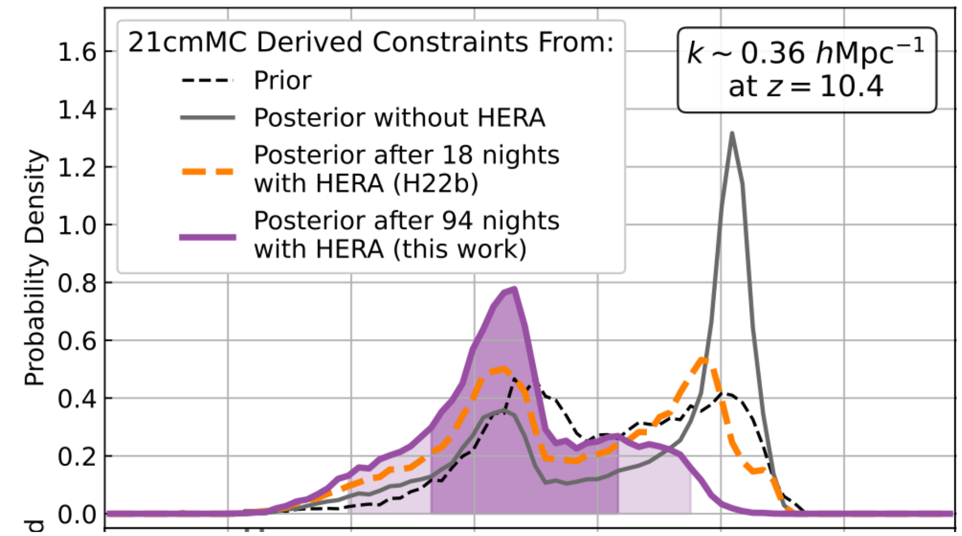
$$P(s \mid d) = P(d \mid s)P(s)$$



$$P(s\,|\,d) = P(d\,|\,s)P(s)$$

3 questions

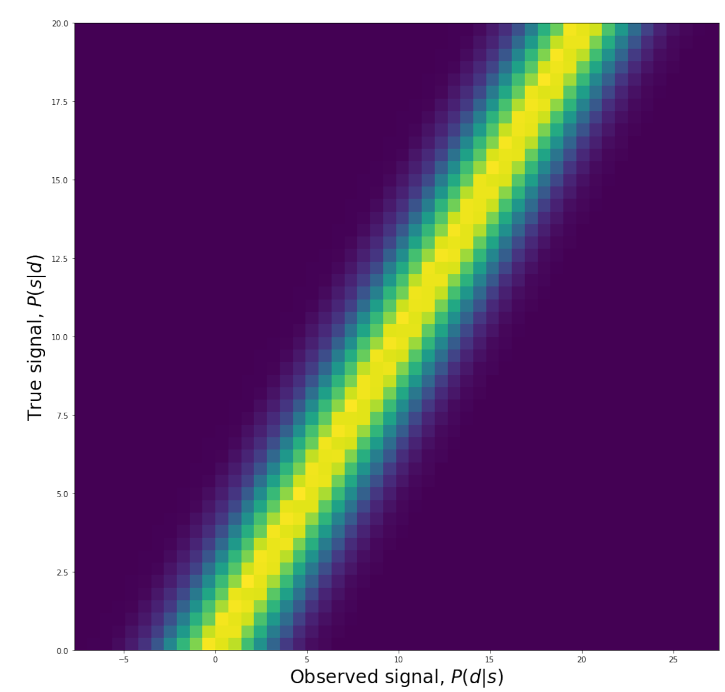
- In what region does the brightest 5% of the models live?
  - $\int_{?}^{\infty} P(d\,|\,s)P(s)\,ds$



- Given only our physics model, what are the brightest 5% of the models consistent with our data?
  - $\int_{?}^{\infty} P(d\,|\,s)(\text{uniform prior})\,ds$

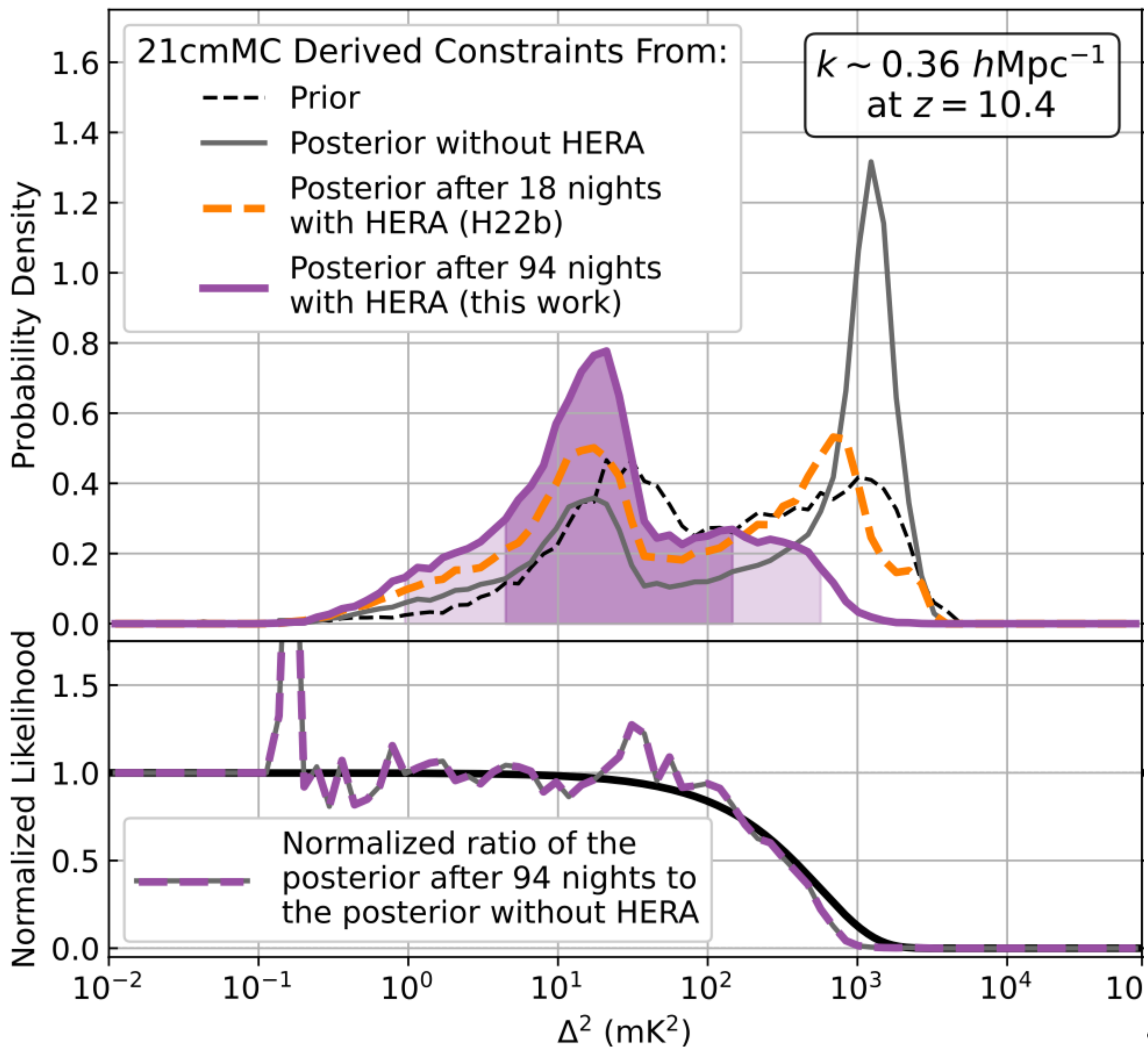
Same as above, but flat $P(s)$

- If we repeated our experiment, at what signal level would we have detection 95% of the time?
  - $\int_{d}^{\infty} P(d\,|\,s_?)\,\partial d$



Using Bayes' theorem

$$P(s \mid d) = P(d \mid s)P(s)$$

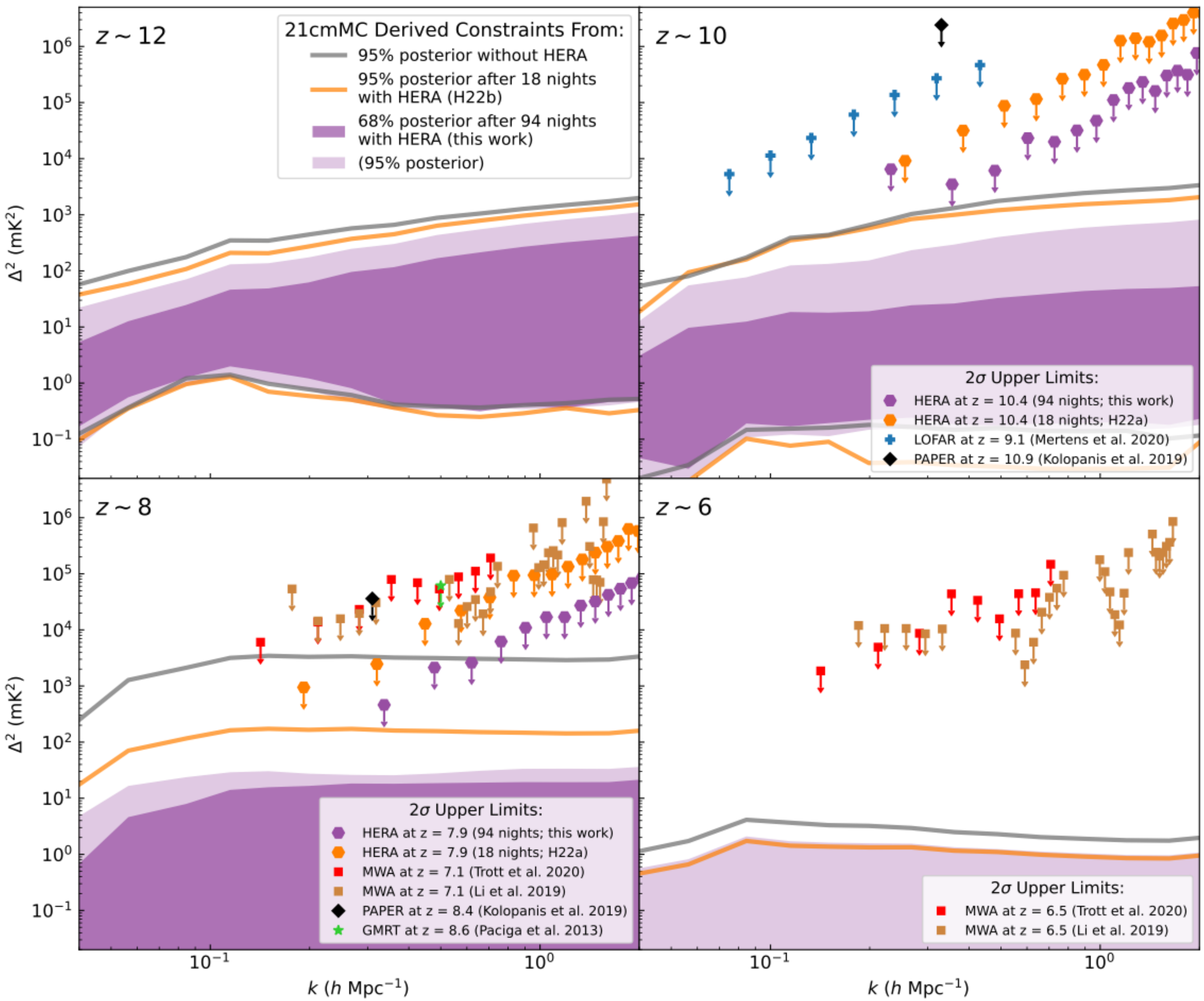


s

Mistake is not in the math

- Theory landscape is a fine question, ***for a theorist***
- Saying we have an observational upper limit (we'd see this signal 95% of the time) when what you calculated was fraction of theory landscape is ***wrong***.
  - Math answered a different question

Signature of ‘shy’ limits



Priors

$$P(s \mid d) = \frac{P(d \mid s)P(s)}{P(d)}$$

To invert you need a prior $P(s)$, key questions:

- Is your prior based on things you know? e.g. measurements, or strong theory constraints → **Use prior**

- Is your prior based on assumptions, guesses, or convenience? → **Flat prior or no prior**

Rules of thumb

- Data driven priors are fine
- You want your science conclusion to depend on your data, not on your prior → uninformative or flat priors
  - Flat priors
  - Priors that are flat over region of interest
- ‘Frequentist’ methods don’t usually run into these problems (clearer question)