INSS 2026 Stats Presentation: How To Work As A Group And Basic NLL

July 10th, 2026

INSS Stat Group VI, Santa Barbara, CA

Jinyoung Kim, Ali Kurmus, Dario Pullia, Clara Saia

Working as group:

A **Turk**, A **Korean** and **two Italians** walk into a bar…

you complete the joke

But on a 5 sigma level its related to stats

Approach to the project:

- *Silent reading* of the problems for the first 20 minutes.
- Deciding what kind of a presentation we want:

  A. Do all of the questions by splitting

  B. Focus on each question and have detailed answers even if don't have time to present all.

Maximizing Success Likelihood (Minimizing failure log likelihood)

Utility / "Learning Impact" Model

$$\mathcal{U}(n) = \frac{D(n)^{\alpha}\ I(n)^{\beta}\ A(n)^{\gamma}\ C}{n^{\delta}}$$

where

- **n** = number of questions we tackle (out of 5)
- **D** = depth of discussion per question
- **I** = individual understanding gained
- **A** = audience learning during presentation
- **C** = collaboration efficiency (team synergy)
- $\alpha, \beta, \gamma, \delta > 0$ are weights reflecting our priorities

Key assumption (fixed total time T and opportunity):

When we spread over more questions, the depth, understanding and audience impact per question are diluted.

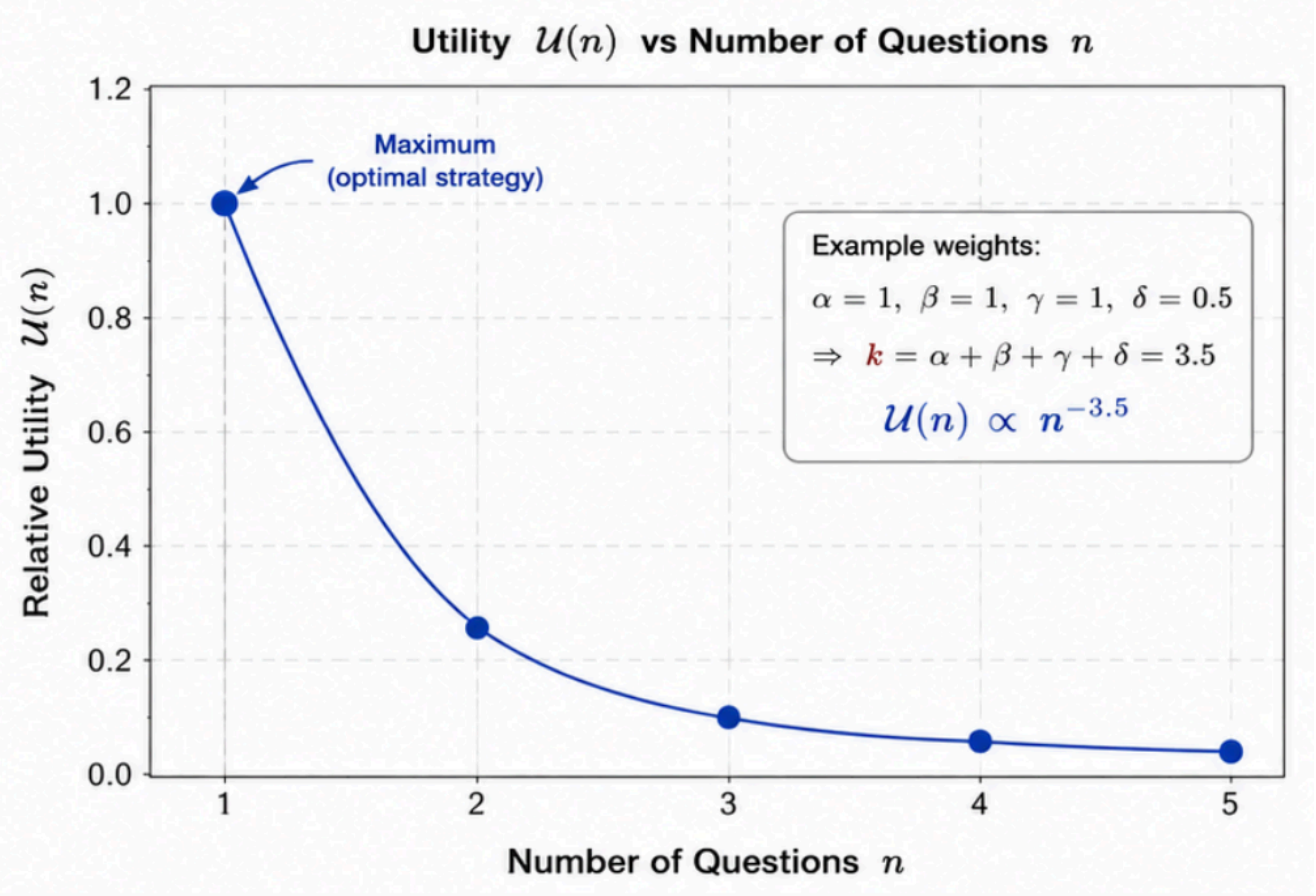
We model this as

$$D(n) \propto n^{-1}, \quad I(n) \propto n^{-1}, \quad A(n) \propto n^{-1}$$

Therefore,

$$\mathcal{U}(n) \propto \frac{C}{n^{\alpha+\beta+\gamma+\delta}} \equiv \frac{C}{n^{k}} \qquad (k > 0)$$

which is maximized for the smallest possible n.



Exercise I

Exercise 1

A physics analysis with a large background B seeks the extract a signal S from and extended likelihood $S+B$ fit to a multivariate discriminant. The binned PDFs for signal and background, as well as the data distributions are shown in Fig. 1. The histograms of Fig. 1 can be downloaded in `ROOT` or `npz` format from

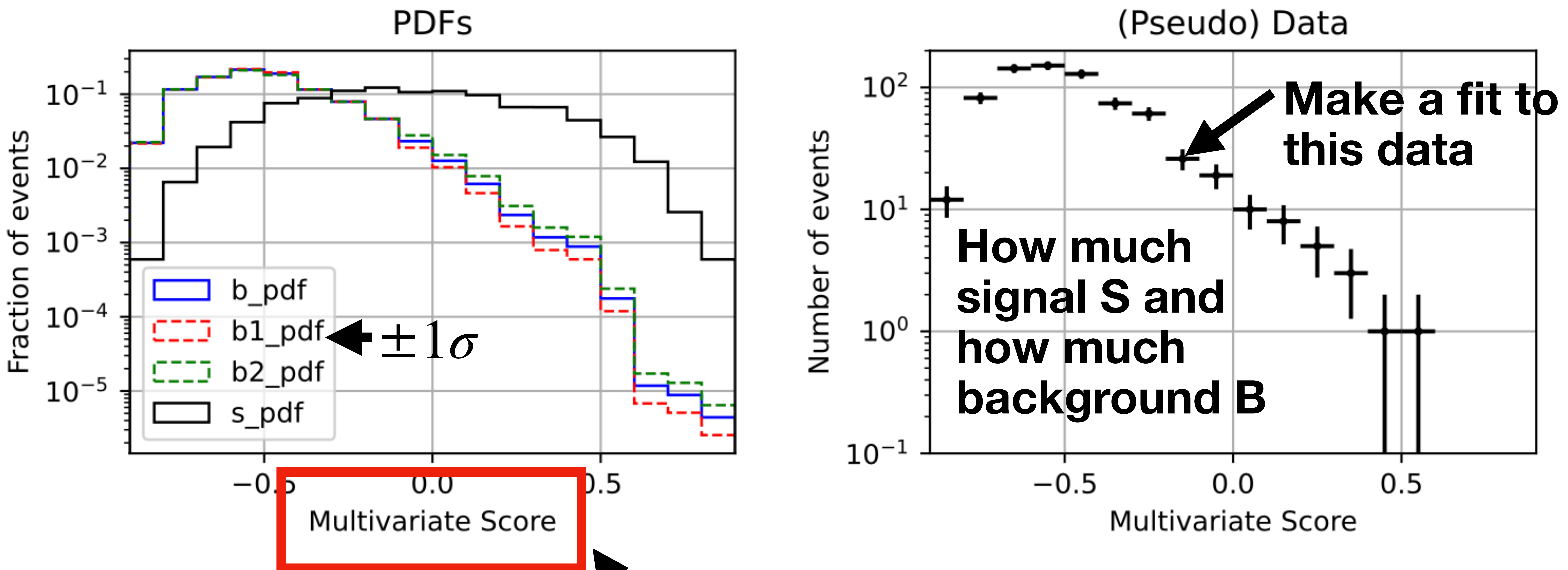


Figure 1: Left: signal and background PDFs. **b_pdf** id the background PDF, **b1_pdf** and **b2_pdf** represent the $\pm 1\sigma$ variations. Right: data)

classification between background and signal

Step 1: What is the data

- It is a **binned** data. For each i, the observed data are:
  $d_i =$ number of observed events in bin i .
- The object we fit is not one number. It is the full list of bin counts:
  $(d_1, d_2, \ldots, d_N)$.
- These are event counts, the natural statistical model is a Poisson model. If the model predicts an expected number of events μ_i in bin i
  $$P(d_i \mid \mu_i) = \frac{\mu_i^{d_i} e^{-\mu_i}}{d_i!} .$$
- The fit changes the model parameters until the predicted bin counts μ_i make the observed counts d_i as likely as possible.

Question For The Audience:

- The models show **fractions of events across bins**.

  Normalized $\sum_i s_i = 1, \qquad \sum_i b_i = 1.$

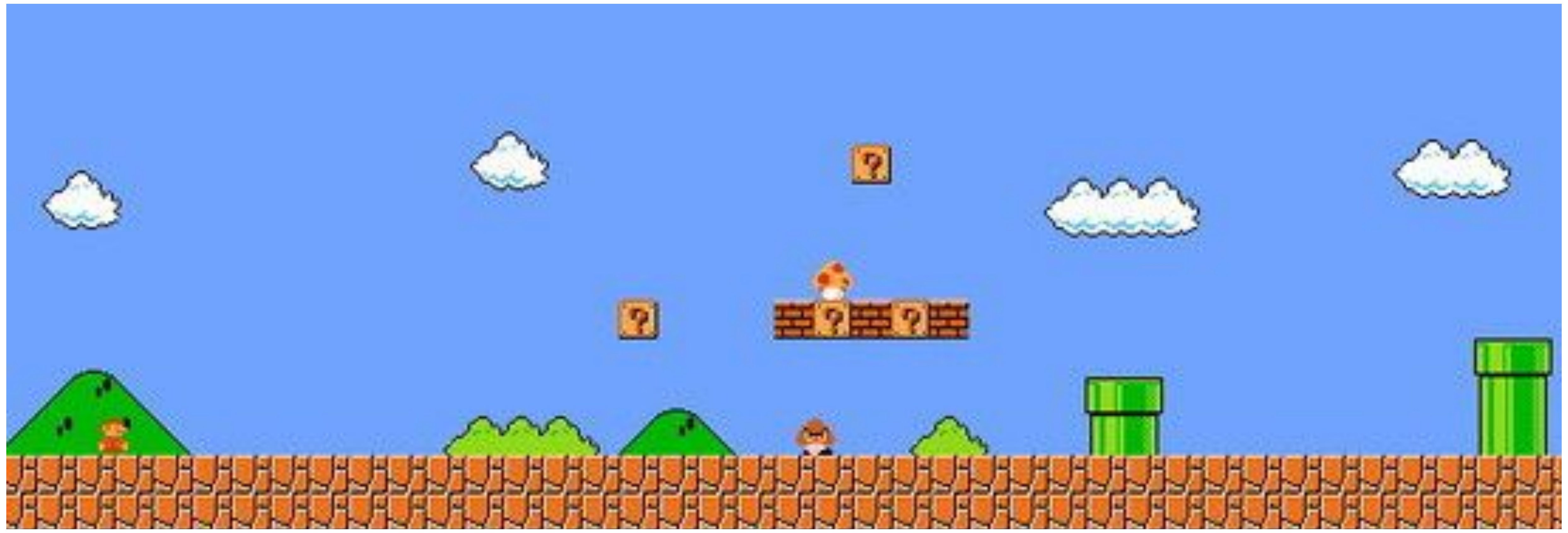
- If the total number of signal events is S, then the expected signal count in bin i is Ss_i .
- If the total number of background events is B, then the expected background count in bin i is Bb_i
- Total expected count $\mu_i(S, B) = Ss_i + Bb_i$.
- **Two questions possible:**
  1. Given that we observed exactly **N** events, what fraction of them look **signal-like** versus **background-like**?
  2. What total number of signal and background events best explains both the **shape** and the **total event count**?

Building The Likelihood

- The full binned likelihood is the product over bins: $L(S,B) = \prod_i P(d_i \mid \mu_i(S,B))$
- Equivalently: $L(S,B) = \prod_i \dfrac{\mu_i(S,B)^{d_i} e^{-\mu_i(S,B)}}{d_i!}$.
- The negative log-likelihood, dropping constants that do not depend on S or B, is NLL$(S,B) = \sum_i \left[\mu_i(S,B) - d_i \ln \mu_i(S,B)\right]$.
- **This is what we minimize!**
- We all like to see familiar things :)
- For large counts expanding around $\mu_i \approx d_i$gives approximately

  NLL $\approx$ constant $+ \dfrac{1}{2} \sum_i \dfrac{(d_i - \mu_i)^2}{\mu_i}$
- **This is the chi square!**

Now Fitting Level I: Background has no uncertainty



- The model would only need the yields S and B: $\mu_i(S, B) = Ss_i + Bb_i$.

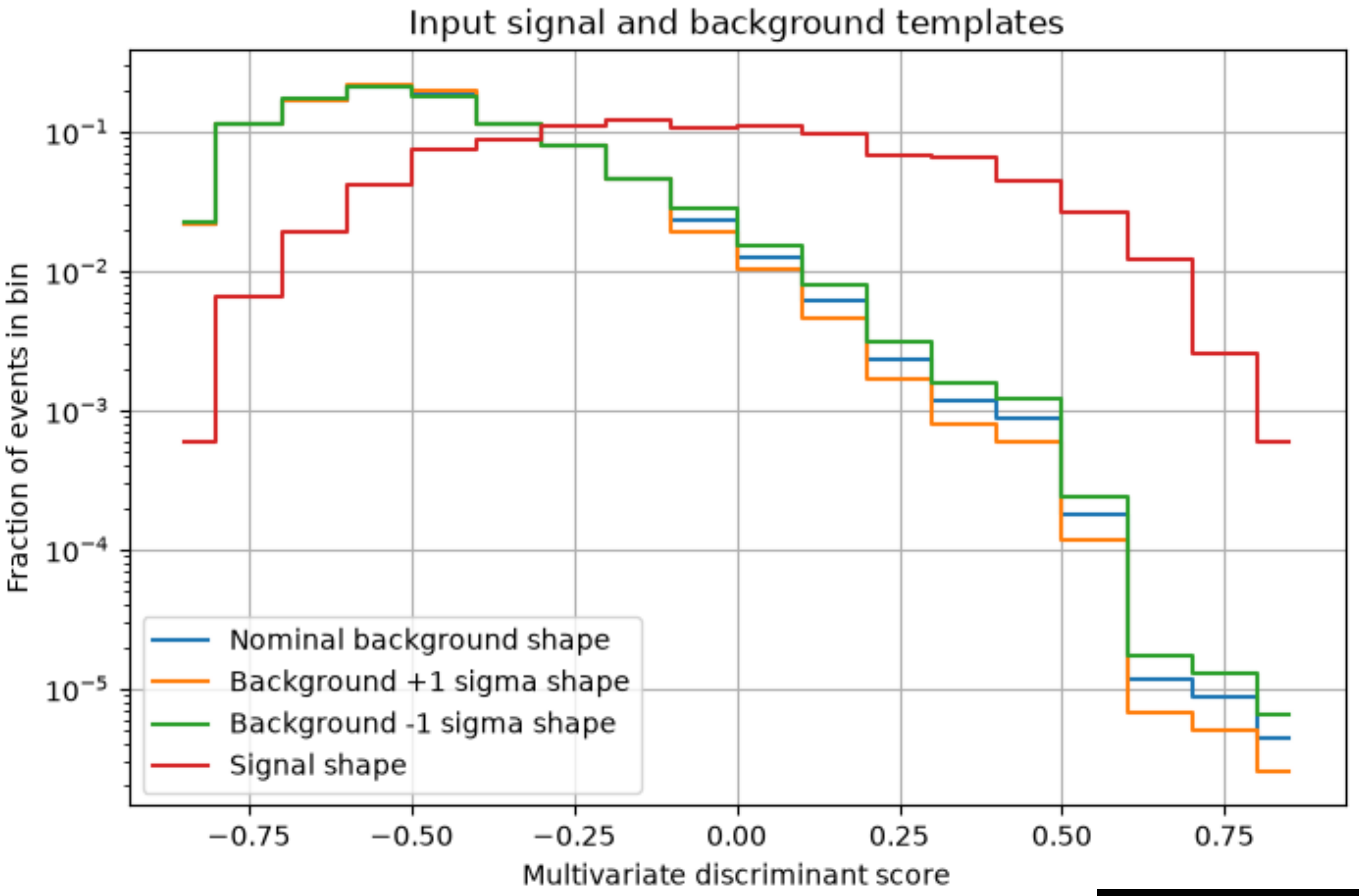
- We ask: For each possible pair (S, B), how likely is it that Poisson fluctuations around $\mu_i(S, B)$would have produced the observed counts d_i?

- The best fit is the pair $(\hat{S}, \hat{B})$ that makes this likelihood largest, or equivalently makes the NLL smallest

- Clean!

Fitting Level II: Background has known uncertainty

- We introduce a nuisance parameter α:
- $\alpha = 0$: use the nominal background shape.
- $\alpha = 1$: use the $+1\sigma$ shifted background shape.
- $\alpha = -1$: use the -1σ shifted background shape.
- Morphing idea from Stats Lecture, slide 9.

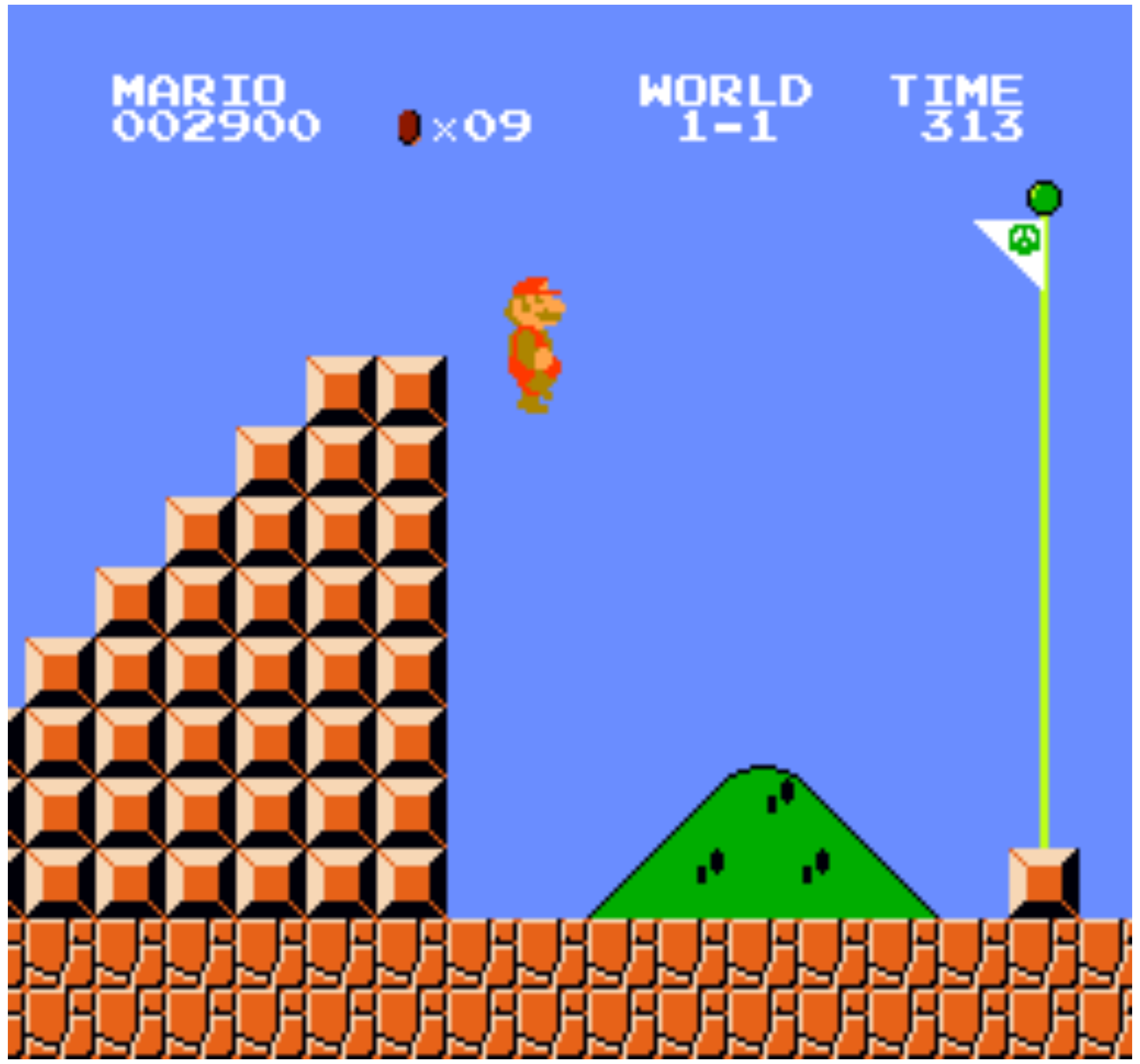


- The model becomes $\mu_i(S, B, \alpha) = Ss_i + Bb_i(\alpha)$.
- The likelihood becomes $L(S, B, \alpha) = \left[\prod_i \frac{\mu_i(S,B)^{d_i} e^{-\mu_i(S,B)}}{d_i!}\right] \cdot \boxed{e^{-\frac{\alpha^2}{2}}.}$
- The constrained NLL is therefore $\mathrm{NLL}(S, B, \alpha) = \sum_i \left[\mu_i - d_i \ln \mu_i\right] + \frac{\alpha^2}{2}$.

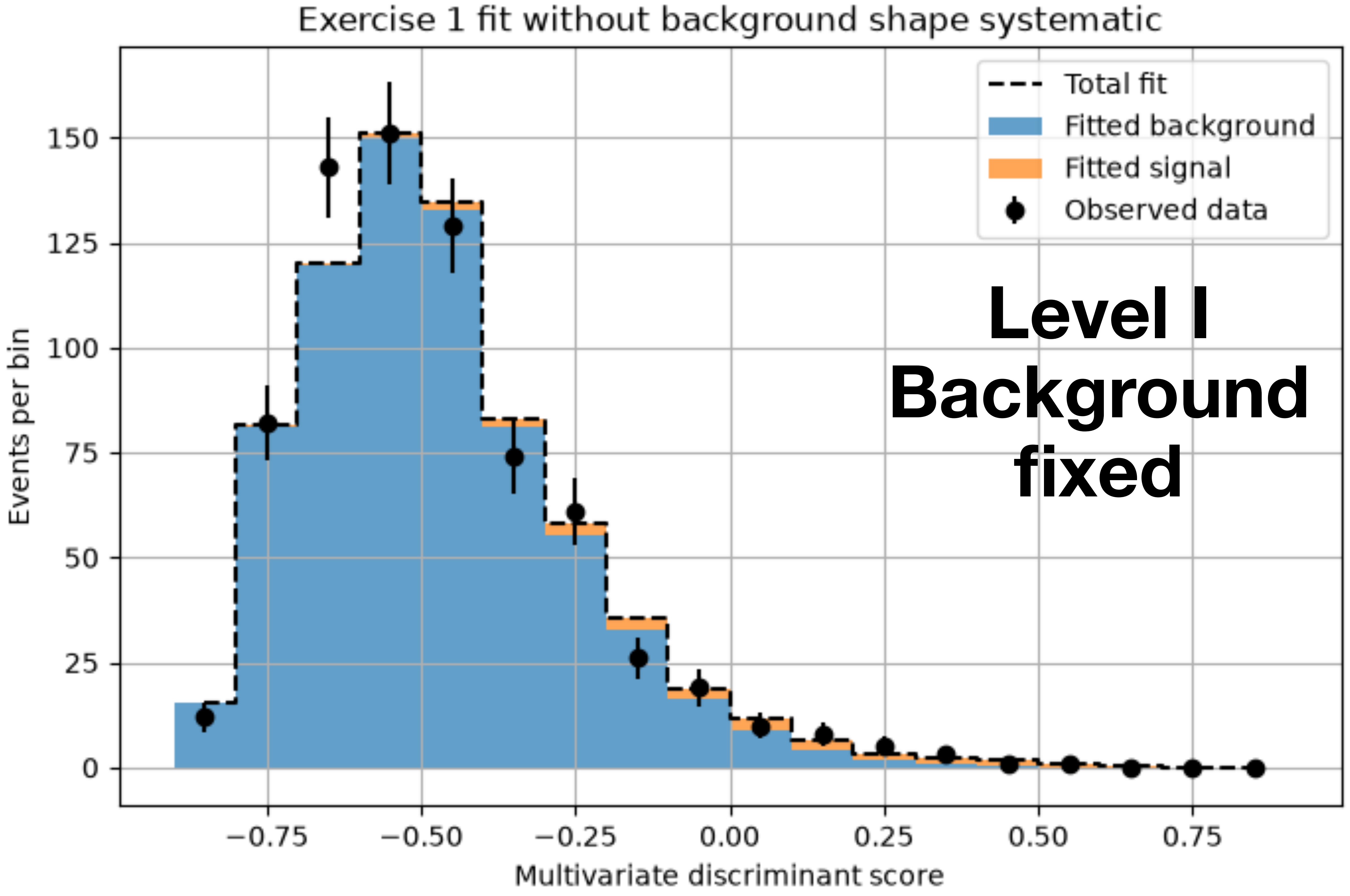
Questions For Thought Before Results

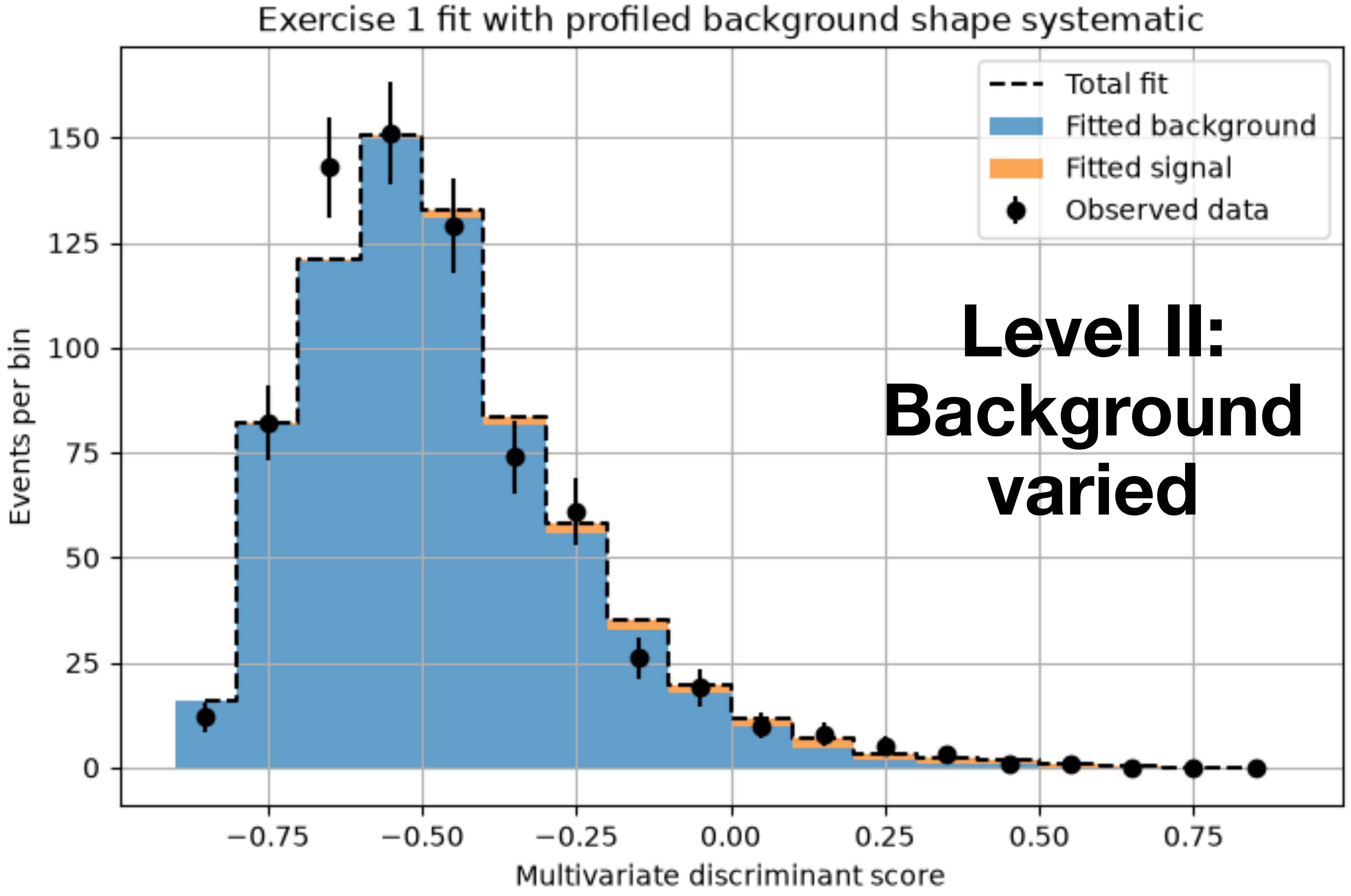
- What if we did not know 1 sigma uncertainty exactly?
    - Possibly a bayesian MCMC where we model the background as a fixed parameter plus a gaussian.
    - Both the mean and standard deviation of the gaussian would be parameters themselves who would be distributed as a gaussian.
    - But what if the parameters of the gaussian are not gaussian distributed? What if standard deviations are not symmetric?

Actual Results For Level I and II Fitting



Actual Results For Level I and II Fitting





- $\hat{S}$ = 23.506 events; $\hat{B}$ = 701.508 events
- NLL_{min} = -2515.384

- $\hat{S}$ = 20.013 events; $\hat{B}$ = 705.000 events
- NLL_{min} = $\boxed{\text{-2515.537}}$ **Better**

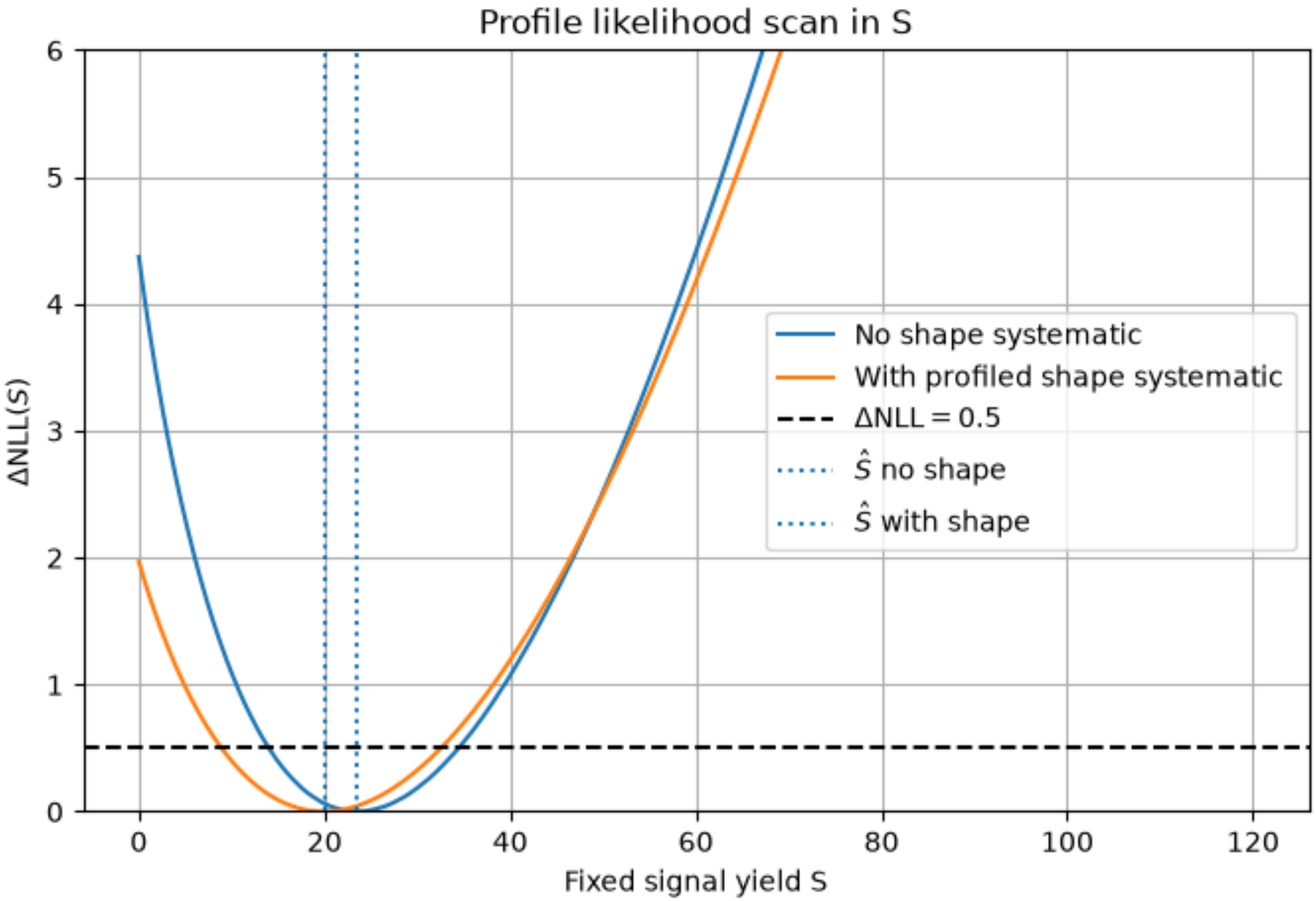
Wait!! Where are the uncertainties of these values????

Profile Likelihood Scan Uncertainties and ΔNLL $= 0.5$

- The full fit gives one best-fit value $\hat{S}$. To get the uncertainty on S, we scan possible fixed values of S.
- If I force the signal yield to be this value, what background normalization B and background-shape shift α make the fit as good as possible
- For each fixed S, we refit the nuisance parameters:
  - Without shape systematic: refit B.
  - With shape systematic: refit B and α.
- This gives the profile $\text{NLL}_{\text{profile}}(S) = \min_{B,\alpha} \text{NLL}(S, B, \alpha)$
- Then we plot $\Delta\text{NLL}(S) = \text{NLL}_{\text{profile}}(S) - \text{NLL}_{\min}$.
- The approximate one-standard-deviation interval for one parameter is where $\Delta\text{NLL} = 0.5$

Final Results

- Without shape systematic: $S = 23.51^{+10.94}_{-9.50}$ events.
- With shape systematic: $S = 20.01^{+12.52}_{-11.11}$ events.



Quick Recap

- We gave a detailed solution to a NLL fit exercise.
- We thought just talking about solutions would not be very productive. But here are they if people need it https://github.com/alikurmus/INSS_2026.
- ROOT-based solution of exercise 1: https://github.com/TFTmaster/INSS2026_stats
- Learning to work as a group was as important.
- We have other suggestions that are useful.

Best Practices For Collaboration

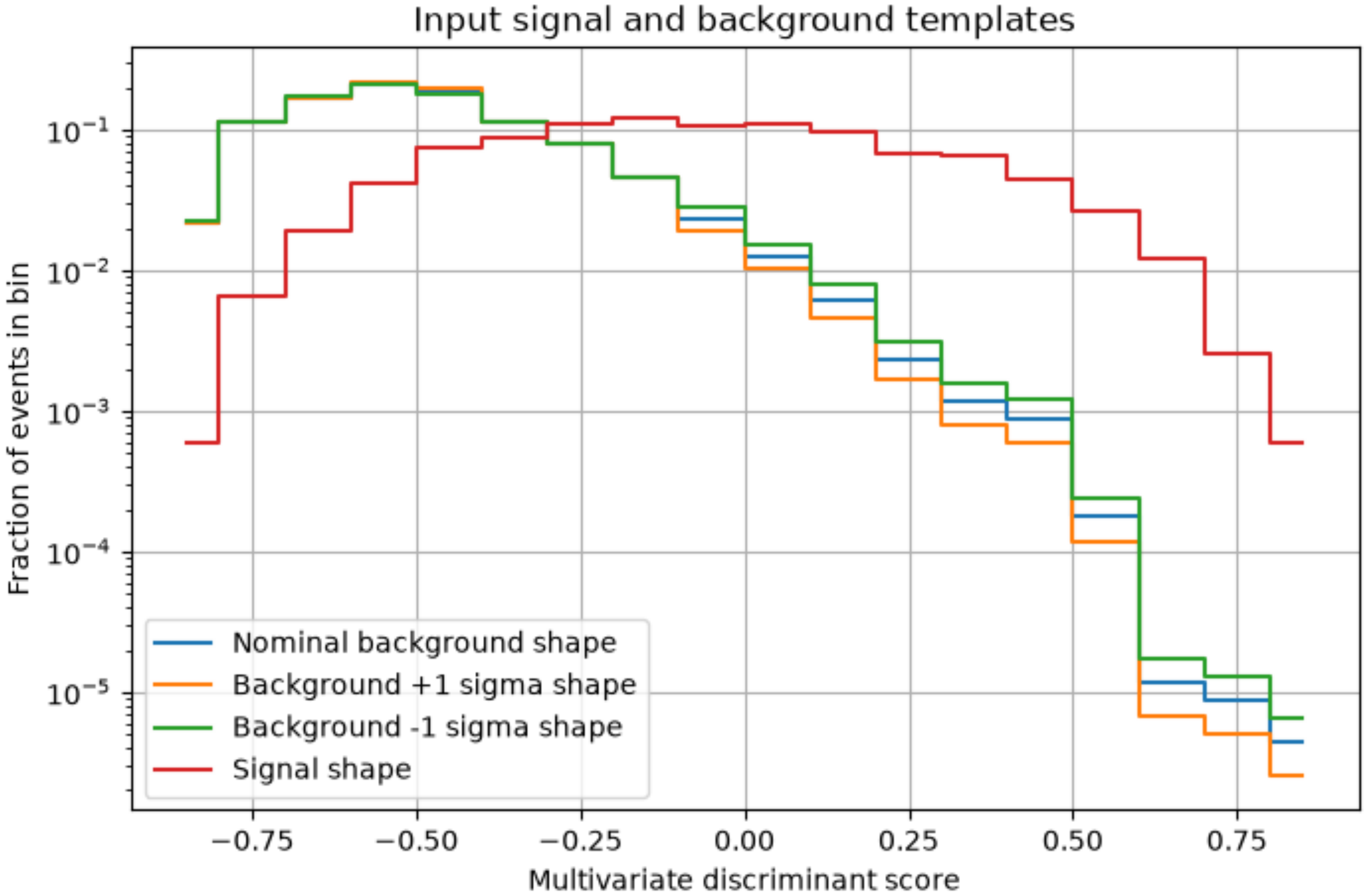
- Use GitHub/ or any version control you like.
- DO NOT COMMIT ON MAIN
- PLEASE PR
- DO NOT git add -a
- Package when possible.

Thank You For Your Time

Backup

Fitting Level II: Background has known uncertainty

- We introduce a nuisance parameter α:
- $\alpha = 0$: use the nominal background shape.
- $\alpha = 1$: use the $+1\sigma$ shifted background shape.
- $\alpha = -1$: use the -1σ shifted background shape.
- Morphing idea from Stats Lecture, slide 9.



- The model becomes $\mu_i(S, B, \alpha) = Ss_i + Bb_i(\alpha)$.
- We penalize large values with a Gaussian constraint $\frac{\alpha^2}{2} \implies \alpha \sim \mathcal{N}(0,1)$
- The constrained NLL is therefore $\text{NLL}(S, B, \alpha) = \sum_i \left[\mu_i - d_i \ln \mu_i\right] + \frac{\alpha^2}{2}$.