Hierarchical Bayesian Analysis

LSST Discovery Fellowship Program Day 4

Greg Gilbert | LSST Discovery Workshop | 23 May 2025

Modeling choices

Physical
What processes do you include?
What approximations do you make?

Statistical
Are data i.i.d.?
Is there correlated noise?
Do you account for data collection?

Model specification
Parameterization
Priors
Convergence criteria

Sampler
Grid search
Maximum likelihood
Markov Chain Monte Carlo
Nested Sampling

Bayes Theorem

$$P(\theta \mid D) = \frac{P(D \mid \theta)\, P(\theta)}{P(D)}$$

Likelihood

Prior

Posterior

Evidence

Hierarchical Bayesian Modeling self-consistently **modifies the prior**

Hierarchical Bayes Theorem

Parameters

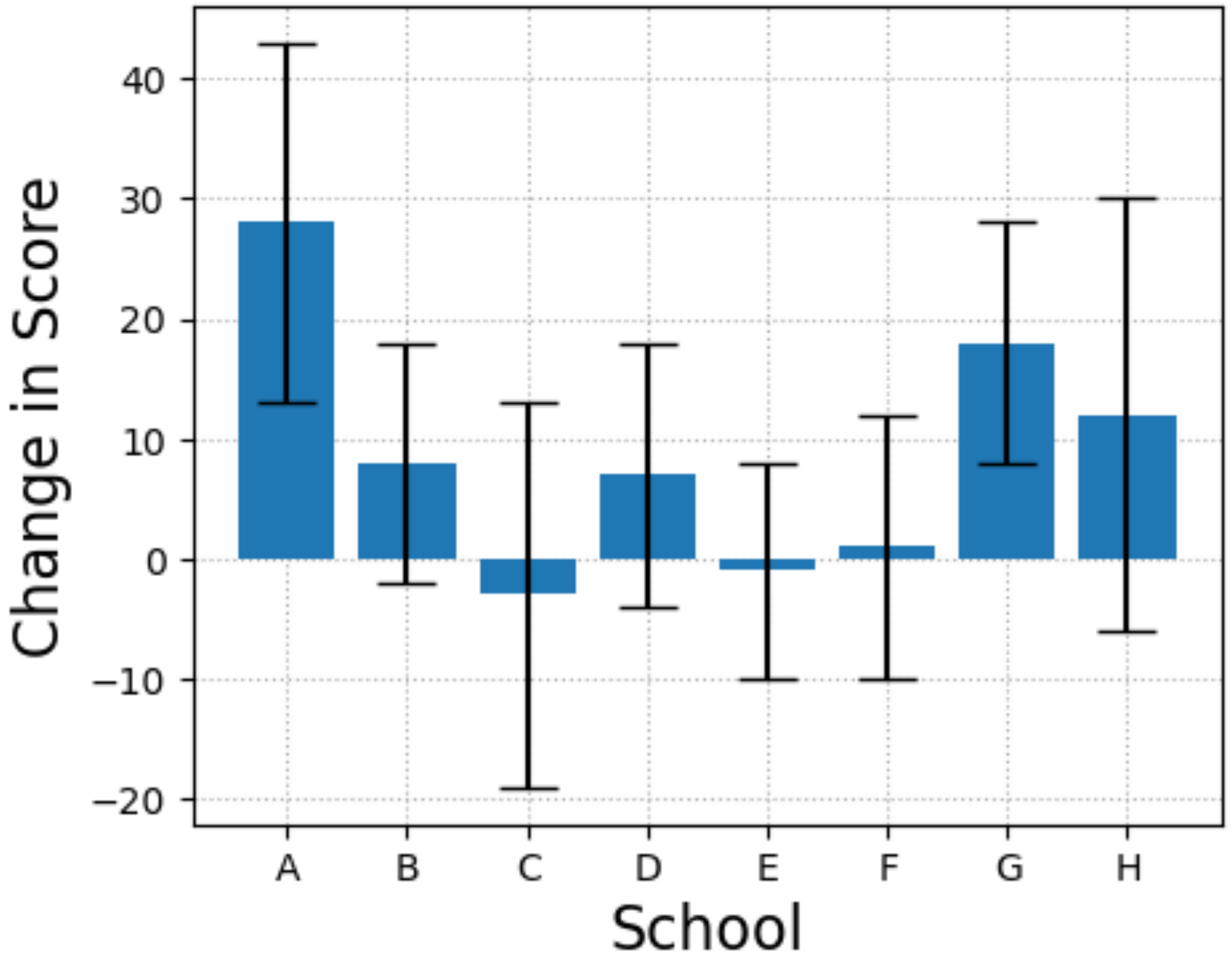
Hyper-parameters

$$P(\theta, \alpha \mid D) = \frac{P(D \mid \theta)P(\theta \mid \alpha)P(\alpha)}{P(D)}$$

The prior can be thought of as the **population-level distribution**

The Eight Schools Problem

The set-up: students from eight schools have participated in a test-prep program. The mean score improvement ΔS and uncertainty on the mean σ_μ for each school are recorded.



$$\mu = [28, 8, -3, 7, -1, 1, 18, 12]$$
$$\sigma_\mu = [15, 10, 16, 11, 9, 11, 10, 18]$$

Question: can the measured effect size for School A (28 pts) be attributed to the test-prep program?

Modeling Options

Independent: Each school is analyzed separately

Pooled: All schools are analyzed in one group

Hierarchical: The relationships between groups are considered

Independent

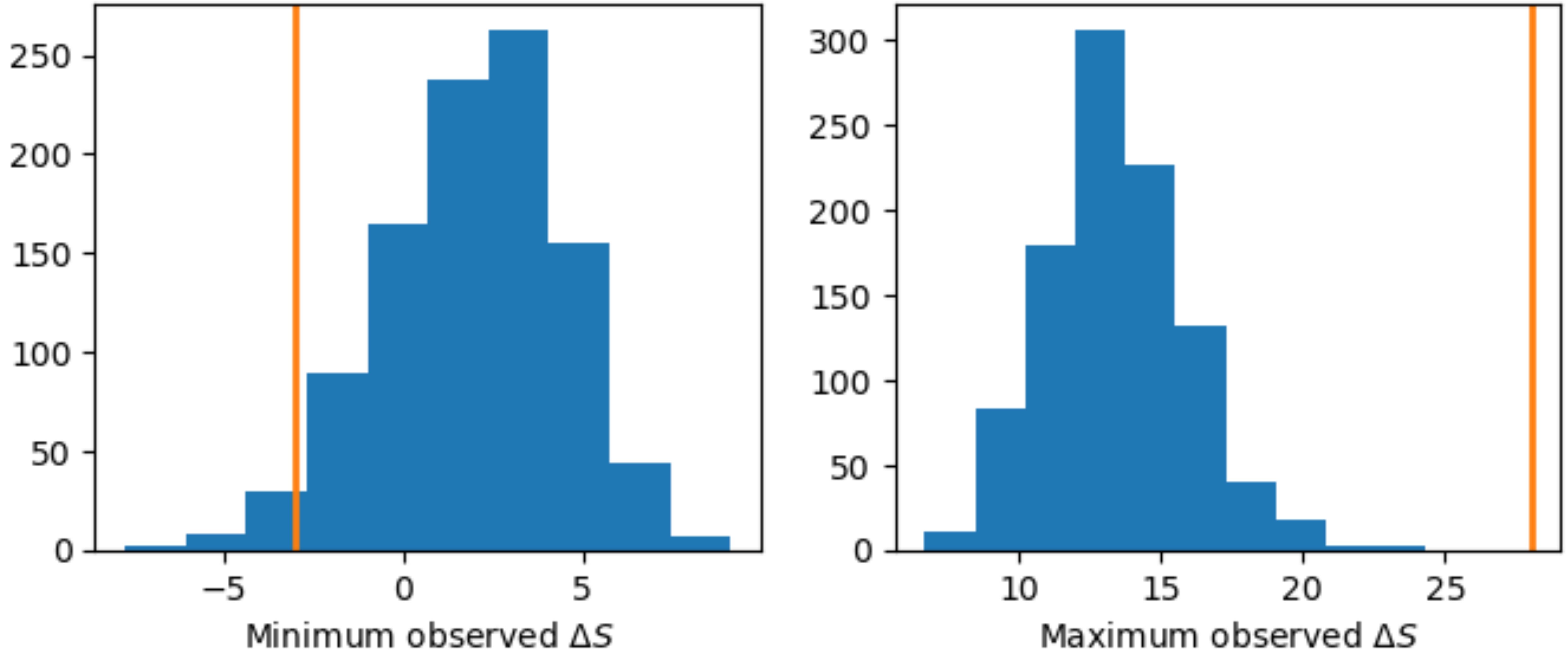
$$\mu = [28, 8, -3, 7, -1, 1, 18, 12]$$
$$\sigma_{\mu} = [15, 10, 16, 11, 9, 11, 10, 18]$$

Pooled

$$\Delta S = 7.7 \pm 4.1$$

The Eight Schools Problem

Question: can the measured effect size for School A (28 pts) be attributed to the test-prep program?



Running 1000 bootstrap trials suggests NO

Eight Schools: Hierarchical Model

$$\alpha_\mu \sim \text{Normal}(\mu, \sigma)$$
$$\alpha_\sigma \sim \text{Half-Cauchy}(\beta)$$

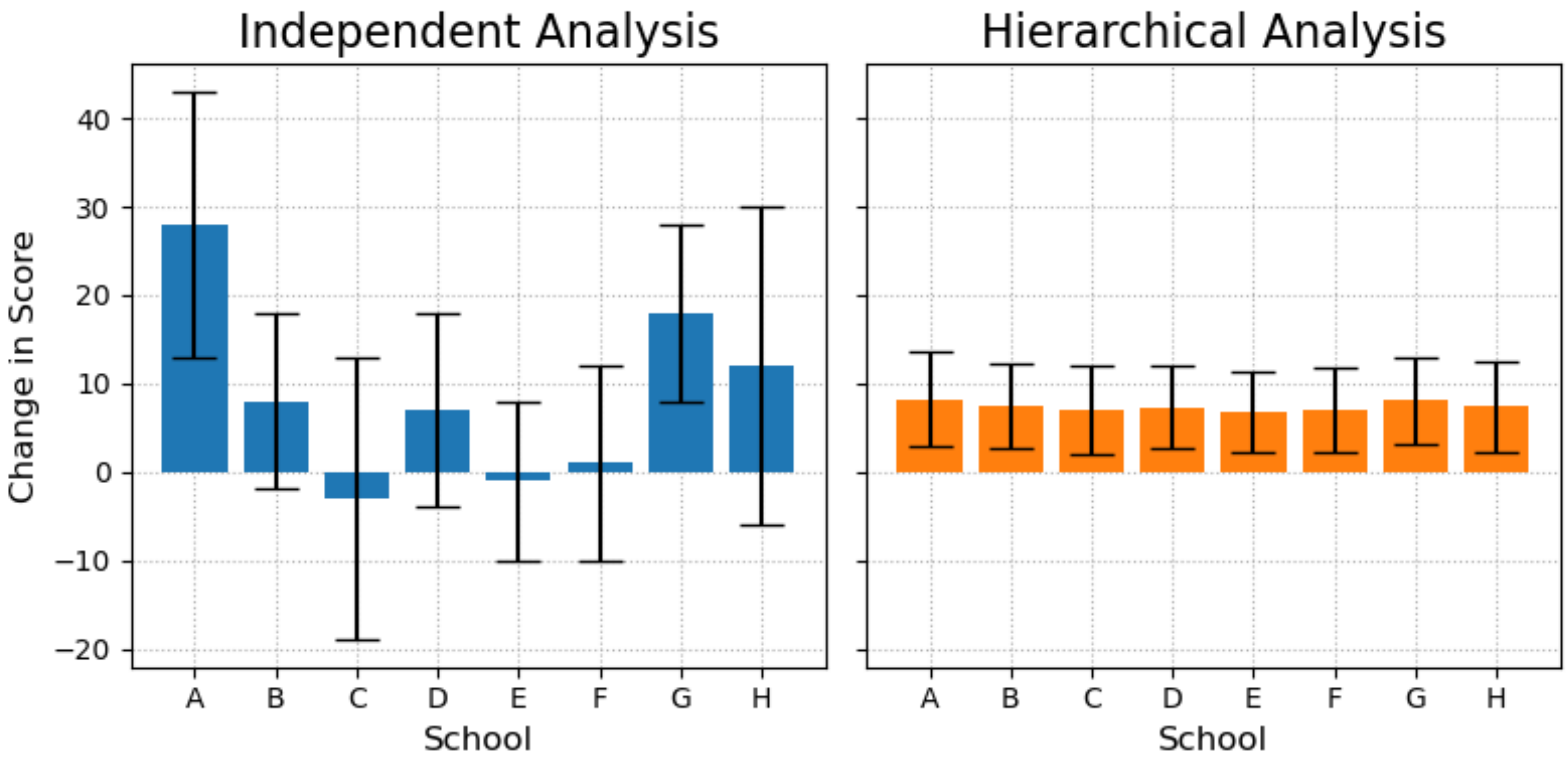
The hyper-parameters $\alpha \equiv \{\alpha_\mu, \alpha_\sigma\}$ describe the population distribution (i.e. the prior)

$$\Delta S_i \sim \text{Normal}(\alpha_\mu, \alpha_\sigma)$$

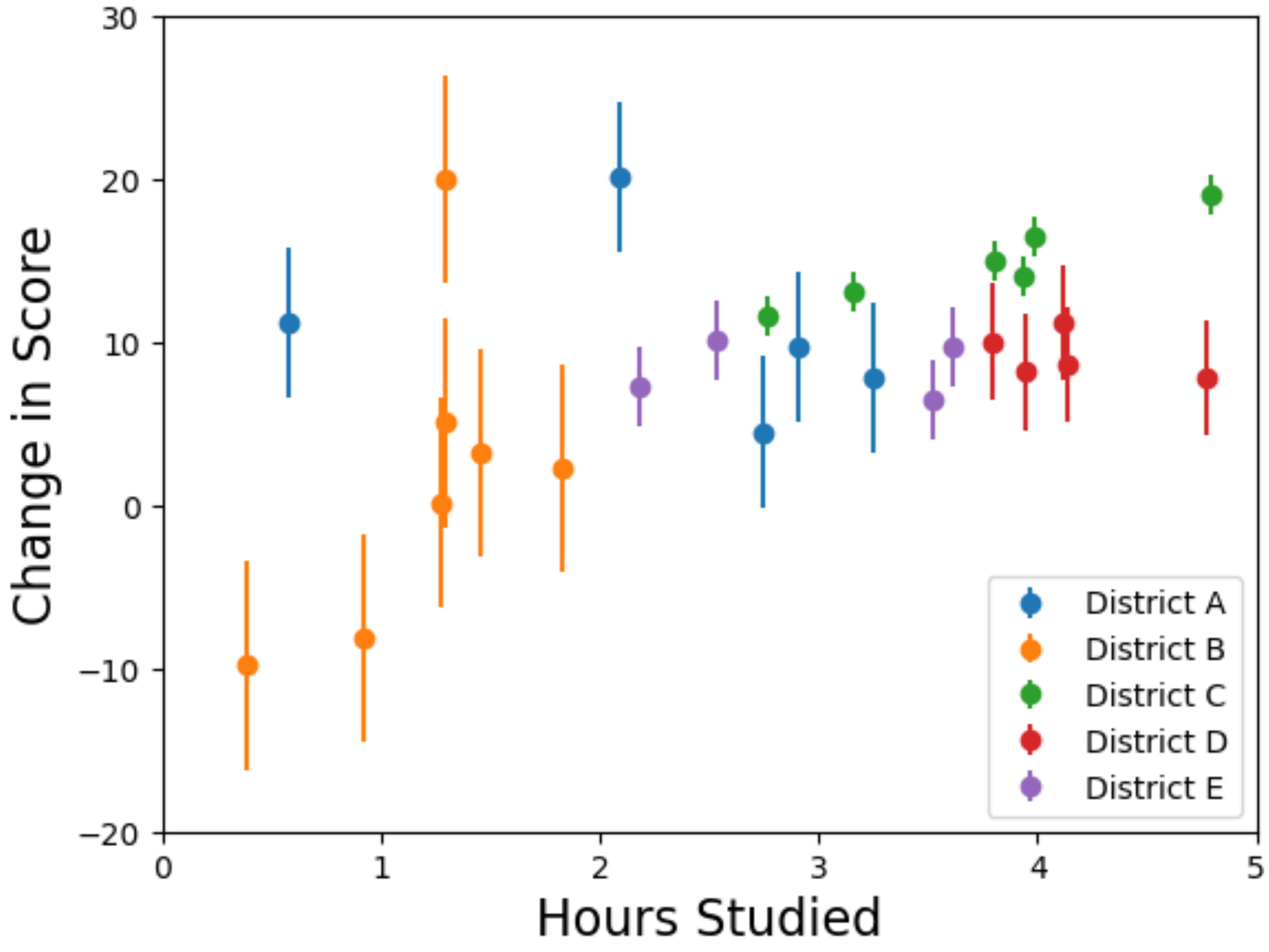
Each school's $\theta_i \equiv \{\Delta S\}_i$ is drawn from a Gaussian described by α

We will simultaneously and self-consistently infer the population hyper-parameters α and the individual member values θ

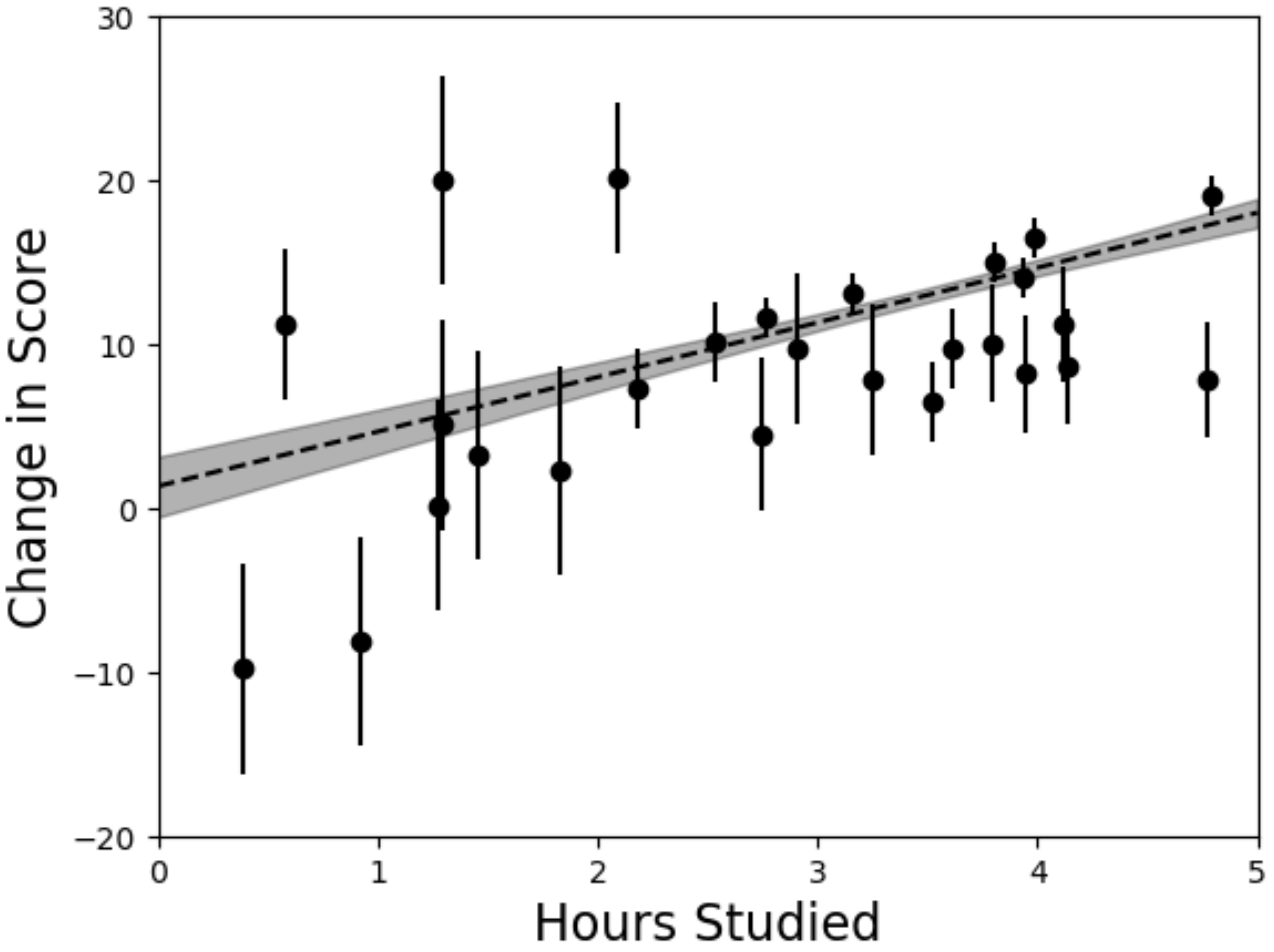
Eight Schools: Hierarchical Model



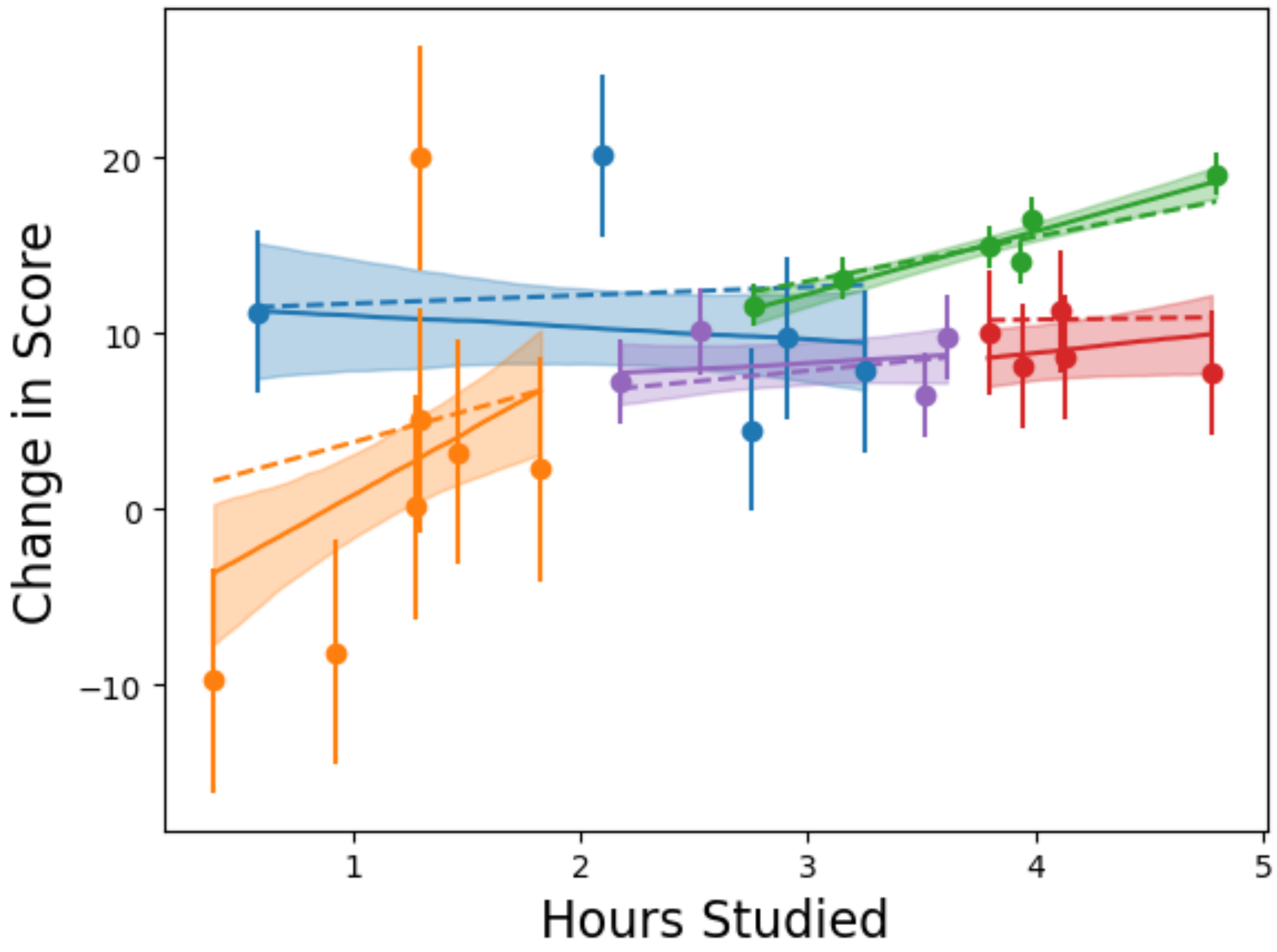
Five Districts Problem



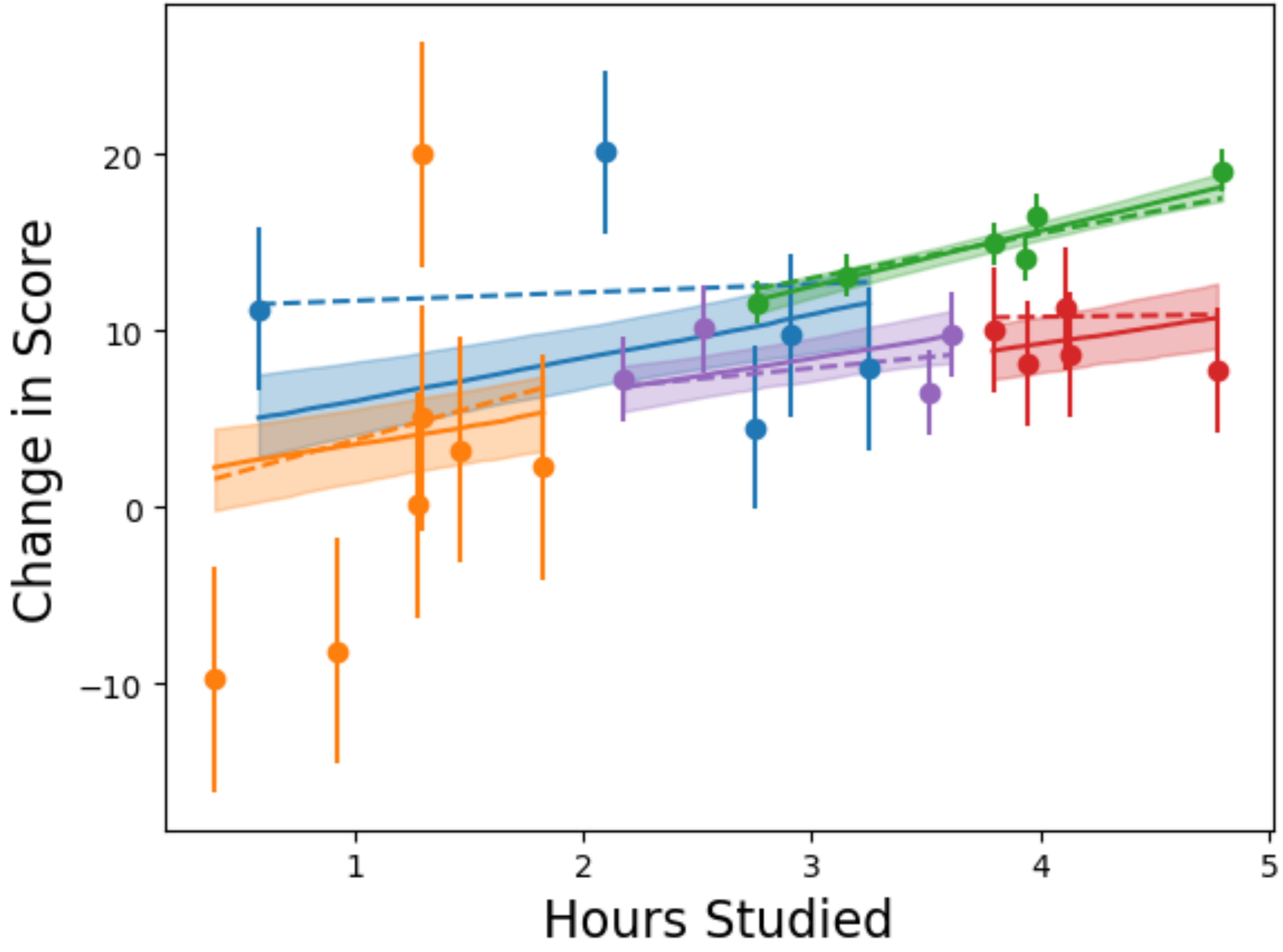
Five Districts: Pooled Model



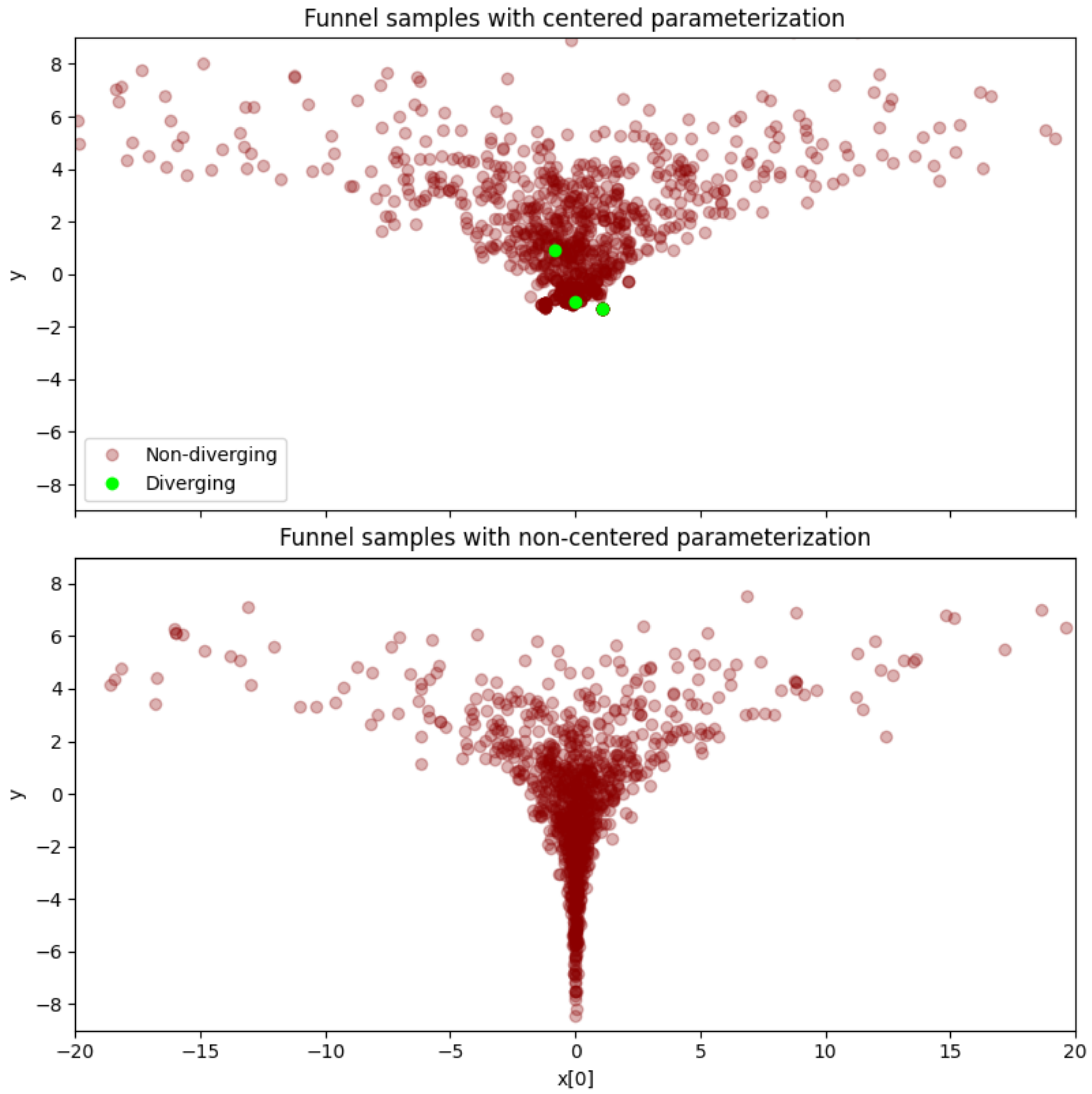
Five Districts: Independent Model



Five Districts: Hierarchical Model



Off-Centered Parameterization



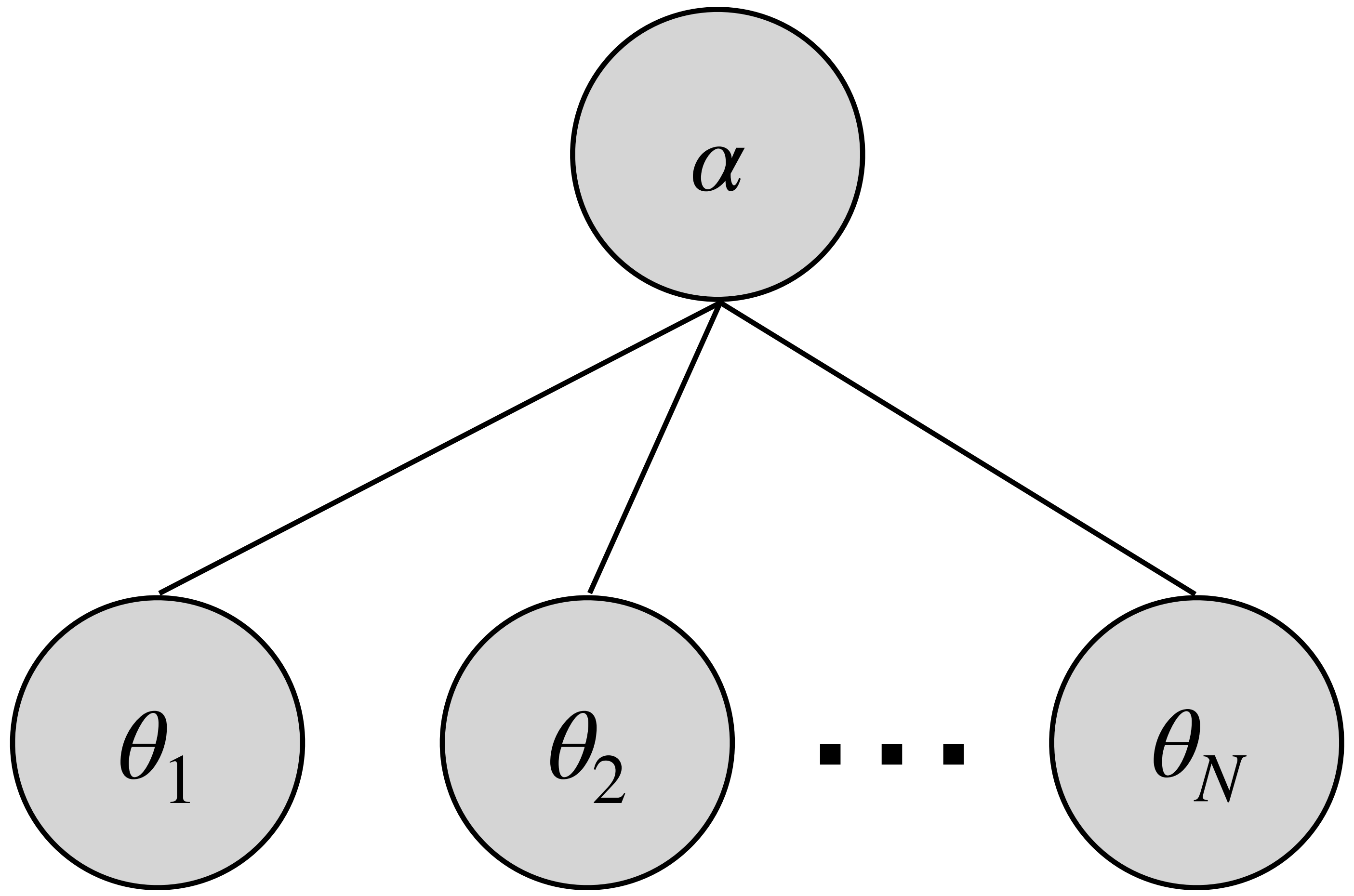
$$y \sim \text{Normal}(0,3)$$
$$x \sim \text{Normal}(0,\ e^{y})$$

$$y_{\text{off}} \sim \text{Normal}(0,1)$$
$$x_{\text{off}} \sim \text{Normal}(0,1)$$

$$y = y_{\text{off}} \cdot 3$$
$$x = x_{\text{off}} \cdot e^{y}$$

Image Credit: NumPyro

Hierarchical Modeling via Importance Sampling



STEP 1 Choose some broad "interim" prior $p_0(\theta)$

STEP 2 Independently fit your model to the data for the N objects to obtain K samples each, producing θ_{nk}

STEP 3 Re-weight your $N \times K$ samples by the likelihood you would have calculated under some different prior f

$$\mathscr{L} \approx \prod_{n=1}^{N} \frac{1}{K} \sum_{k=1}^{K} \frac{f(\theta; \alpha)}{p_0(\theta)}$$

updated prior

"interim" prior

This method is equivalent to what you would achieve by simultaneously fitting all N objects under the hierarchical prior

Caveats: (1) The "interim" prior must provide support over full posterior space, (2) you must generate sufficient samples

Further reading

Hierarchical Modeling

Thomas Wiecki & Danne Elbers, “The best of both worlds: hierarchical linear regression in PyMC3”
https://twiecki.io/blog/2014/03/17/bayesian-glms-3/

Thomas Wiecki, “Why hierarchical models are awesome, tricky, and Bayesian”
https://twiecki.io/blog/2017/02/08/bayesian-hierchical-non-centered/

Michael Betancourt, “Hierarchical Modeling”
https://betanalpha.github.io/assets/case_studies/hierarchical_modeling.html

Off-Centered Parameterization

NumPyro: https://num.pyro.ai/en/stable/examples/funnel.html

MC-Stan: https://mc-stan.org/docs/2_19/stan-users-guide/reparameterization-section.html

Importance Sampling

Hogg, Myers, & Bovy 2010, “Inferring the eccentricity distribution”, ApJ, 725, 2