Identifiability

In statistical and cognitive modeling, particularly when we deal with complex or custom models, a crucial concept is *identifiability*. A model is considered *non-identifiable* if there exist different combinations of the model's parameters that always produce exactly the same predictions. In the context of likelihood-based models, this implies that distinct parameter sets yield identical likelihood values for any potential observed data. As a result no amount of observed data can distinguish between these parameter constellations. This makes it impossible to infer parameters from data – the data do not provide enough information to uniquely determine the true underlying parameters.

There are nuances to different kinds of non-identifiability, but certainly the most problematic type is **structural non-identifiability**, which is not due to insufficient data or poor experimental design, but is inherent in the mathematical structure of the model's equations themselves.

Fitting a non-identifiable model using MCMC methods typically results in several observable difficulties:

- Poor MCMC convergence, characterized by high values of the $\hat{R}$ convergence statistic (much greater than 1) and low Effective Sample Sizes (ESS).
- Posterior distributions that look unstable, that drift, or take on unusual shapes.
- Strong correlations among the parameters affected by the non-identifiability in the posterior samples.

Those three symptoms are visible – but the fourth symptom is the problem: unidentified models are not suitable for statistical inference. You cannot estimate their parameters and you cannot draw conclusions about them. Avoiding or addressing non-identifiability is a necessary step for valid model-based inference.

Illustrative Example: A Hierarchical Signal Detection Theory Model

Consider a hierarchical Signal Detection Theory (SDT) model as an example. Consider a simple SDT model that predicts counts of Hits (H_{ip}) and False Alarms (F_{ip}) for person p in condition i.

The data are modeled via a **likelihood** derived from Binomial distributions, where the success probabilities are determined by SDT parameters: sensitivity (d'_{ip}) and criterion (c_p), along with trial counts (n_s, n_n). Jointly, these equations (which we have covered many times) can be used to compute the likelihood of the data given the parameters. We can write this abstractly as:

$$(H_{ip}, F_{ip}) \sim SDT\left(d'_{ip}, c_p, n_s, n_n\right)$$

Within a **hierarchical signal detection theory (HSDT) model**, the parameters d'_{ip} and c_p have some sort of structure. Let's suppose a very reasonable HSDT model, in which d'_{ip} is defined as the sum of a person-specific intercept α_p and a condition-specific effect γ_i:

$$d'_{ip} = \alpha_p + \gamma_i$$

The condition effect γ_i is specified linearly based on a condition-level predictor:

$$\gamma_i = \zeta_0 + \zeta_1 \cdot \text{predictor}_i \qquad \text{(Condition effect on } d'\text{)}$$

ζ_0 functions as the intercept for the condition effect and ζ_1 is the slope (i.e., effect size) of the condition effect.

The person intercept α_p is drawn from a group-level distribution to allow pooling of information across people:

$$\alpha_p \sim \text{Normal}(\mu_\alpha, \sigma_\alpha^2) \qquad \text{(Person sensitivity intercept)}$$

The person criterion c_p is also modeled hierarchically (but less relevant here):

$$c_p \sim \text{Normal}(\mu_c, \sigma_c^2) \qquad \text{(Person criterion)}$$

The identifiability issue in this formulation is concentrated within the specification of d'_{ip}.

Analysis of Non-Identifiability

The source of the non-identifiability in this model is revealed by substituting the definition of γ_i into the equation for d'_{ip}:

$$\begin{aligned} d'_{ip} &= \alpha_p + \gamma_i \\ &= \alpha_p + (\zeta_0 + \zeta_1 \cdot \text{predictor}_i) \\ &= (\alpha_p + \zeta_0) + \zeta_1 \cdot \text{predictor}_i \end{aligned}$$

The parameter d'_{ip} depends on the sum of the person intercept α_p and the condition effect intercept ζ_0, in addition to the term involving ζ_1 and the predictor. A jargon-y way of describing this setup is that d'_{ip} has two intercepts (or rather, two parameters that each act as an intercept).

Since the model likelihood depends on d'_{ip} and c_p, it is determined by the value of the sum $(\alpha_p + \zeta_0)$ and the term $\zeta_1 \cdot \text{predictor}_i$.

We can now consider two distinct parameter sets. Let one set contain (α_p, ζ_0) and another contain $(\tilde{\alpha}_p, \tilde{\zeta}_0)$, where $\tilde{\alpha}_p = \alpha_p - C$ and $\tilde{\zeta}_0 = \zeta_0 + C$ for an arbitrary constant C. The sum for the second set is:

$$\tilde{\alpha}_p + \tilde{\zeta}_0 = (\alpha_p - C) + (\zeta_0 + C) = \alpha_p + \zeta_0$$

As is hopefully obvious, any constant C could be added to ζ_0 and subtracted from α_p without altering their sum. Consequently, the calculated value of d'_{ip} remains unchanged. As d'_{ip} (and an identifiable c_p) determines the likelihood, these distinct parameter combinations yield identical likelihood distributions, always. In other words, the individual values of α_p and ζ_0 are arbitrary (even though their sum is not).

α_p and ζ_0 are said to be confounded additively. At the group level, this manifests as an inability to distinguish between the mean person sensitivity μ_α and the condition intercept ζ_0. The model structure permits trade-offs between these parameters while maintaining the same likelihood. This constitutes *structural non-identifiability.*

Common MCMC Symptoms

MCMC diagnostics can help us identify non-identifiability. When a non-identified model is fit using MCMC methods, several diagnostic indicators should raise red flags:

- **Poor Convergence** for the parameters involved in the confounding (e.g., μ_α and ζ_0). This is shown by high $\hat{R}$ values and low Effective Sample Sizes (ESS).

- **Wandering Trace Plots**. The MCMC chains for the confounded parameters may exhibit poor mixing and fail to stabilize, often appearing to explore a ‘ridge’ of high likelihood.

- **Strong Posterior Correlation**. Visual analysis of the joint posterior distribution (or pairwise plots) of the confounded parameters may show a dependency. In this additive case, a negative linear correlation between μ_α and ζ_0 is expected, but in general the relationship need not be linear, and it may even involve many parameters. (In fact, in the example scenario, every $alpha_p$ is also involved.)

- **Sampling Difficulties**. Numerical stability issues may arise, potentially leading to sampler warnings such as divergences.

These observations collectively signal the presence of non-identifiability.

PyMC Implementation of the Non-Identified Model

The following PyMC code snippet illustrates the parameter definitions that lead to the non-identifiability in the hierarchical SDT model. (Find the full code in the GitHub repository, `0-introduction/src/sdt/sdt_identifiability.py`.)

```python
# Uninformative priors for Sensitivity (d') parameters
mu_alpha = pm.Normal("mu_alpha", mu=0.0, sigma=1.0e9) # << Part 1
sigma_alpha = pm.HalfNormal("sigma_alpha", sigma=1.0e9)
zeta0 = pm.Normal("zeta0", mu=0.0, sigma=1.0e9)       # << Part 2
zeta1 = pm.Normal("zeta1", mu=0.0, sigma=1.0e9)

# Person sensitivity intercepts (drawn from mu_alpha)
alpha_p = pm.Normal("alpha_p", mu=mu_alpha, sigma=sigma_alpha)

# Calculate gamma_i (Condition effect on d')
# (Assuming 'cond_pred_data' is available elsewhere)
gamma_i = zeta0 + zeta1 * cond_pred_data

# Calculate d' using alpha_p and gamma_i
d_ip = pm.Deterministic("d_ip", alpha_p[...] + gamma_i) # << Part 3

# ... SDT likelihood calculations using d_ip and c_p ...
```

Listing 1: PyMC code snippet for the non-identified SDT model parameters.

The inclusion of ‘mu_alpha‘ (Part 1, mean of person intercepts) and ‘zeta0‘ (Part 2, intercept of condition effects), combined with their additive contribution to ‘d_ip‘ (Part 3 via ‘alpha_p‘ and ‘gamma_i‘), introduces a structural redundancy.

Diagnosis Method 1: Trace Plots

Evaluation of MCMC trace plots provides a visual diagnostic for convergence and mixing issues. Using ‘arviz.plot_trace()‘, focused on parameters such as ‘mu_alpha‘ and ‘zeta0‘, is standard practice.

When reviewing trace plots, pay attention to:

- Indications of poor mixing among the distinct chains for ‘mu_alpha‘ and ‘zeta0‘ – the chains should look like they are drawing samples from a stationary distribution and not wandering up and down. A well-mixed trace plot is sometimes described as a ‘fat, hairy caterpillar’.
- Evidence of chains exploring separate regions or failing to converge to a common stationary distribution. All chains should converge to the same region of the parameter space.
- Cross-check visual findings with high $\hat{R}$ values reported in the MCMC summary.

Figure 1 presents trace plots from a fit of the non-identified model. Here we see the expected poor mixing and wandering for the confounded parameters. By contrast, Figure 2 shows well-behaved traces from an identified version of the model.

MCMC Output Summary

Quantitative assessment of MCMC performance is provided by the summary table. The summary for the non-identified model looks like this:

```
    --- Non-Identified SDT Model Summary ---
             mean       sd   hdi_3%  hdi_97%  mcse_mean  mcse_sd  ess_bulk  ess_tail  r_hat
mu_alpha   40.779  268.443 -304.985  570.827    128.271   70.199       5.0      17.0   2.70
zeta0     -38.809  268.503 -569.032  307.068    128.298   70.209       5.0      17.0   2.70
zeta1      -0.397    0.028   -0.446   -0.350      0.008    0.001      15.0     180.0   1.22
mu_c        0.048    0.058   -0.020    0.173      0.027    0.014       5.0      26.0   2.35
Sampling 4 chains for 1_500 tune and 1_000 draw iterations (6_000 + 4_000 draws total) took 379s.
There were 1000 divergences after tuning. Increase `target_accept` or reparameterize.
```

The high $\hat{R}$ values (2.70) and extremely low ESS (5.0) for ‘mu_alpha‘ and ‘zeta0‘ indicate poor convergence. Seeing 1000 divergences is also a red flag (ideally, there are none).

For comparison, the summary from the identified model below indicates successful convergence, with $\hat{R}$ values near 1.0 and substantially higher ESS for all parameters, including the newly named ‘mu_alpha_intercept‘.

```
--- Identified SDT Model Summary ---
               mean       sd   hdi_3%  hdi_97%  mcse_mean  mcse_sd  ess_bulk  ess_tail  r_hat
mu_alpha_int  1.936    0.149    1.659    2.221      0.008    0.004     380.0     598.0   1.01
zeta1        -0.398    0.033   -0.463   -0.341      0.001    0.000    3558.0    3080.0   1.00
mu_c          0.072    0.062   -0.046    0.186      0.003    0.002     426.0     735.0   1.01
Sampling 4 chains for 1_500 tune and 1_000 draw iterations (6_000 + 4_000 draws total) took 12s.
```

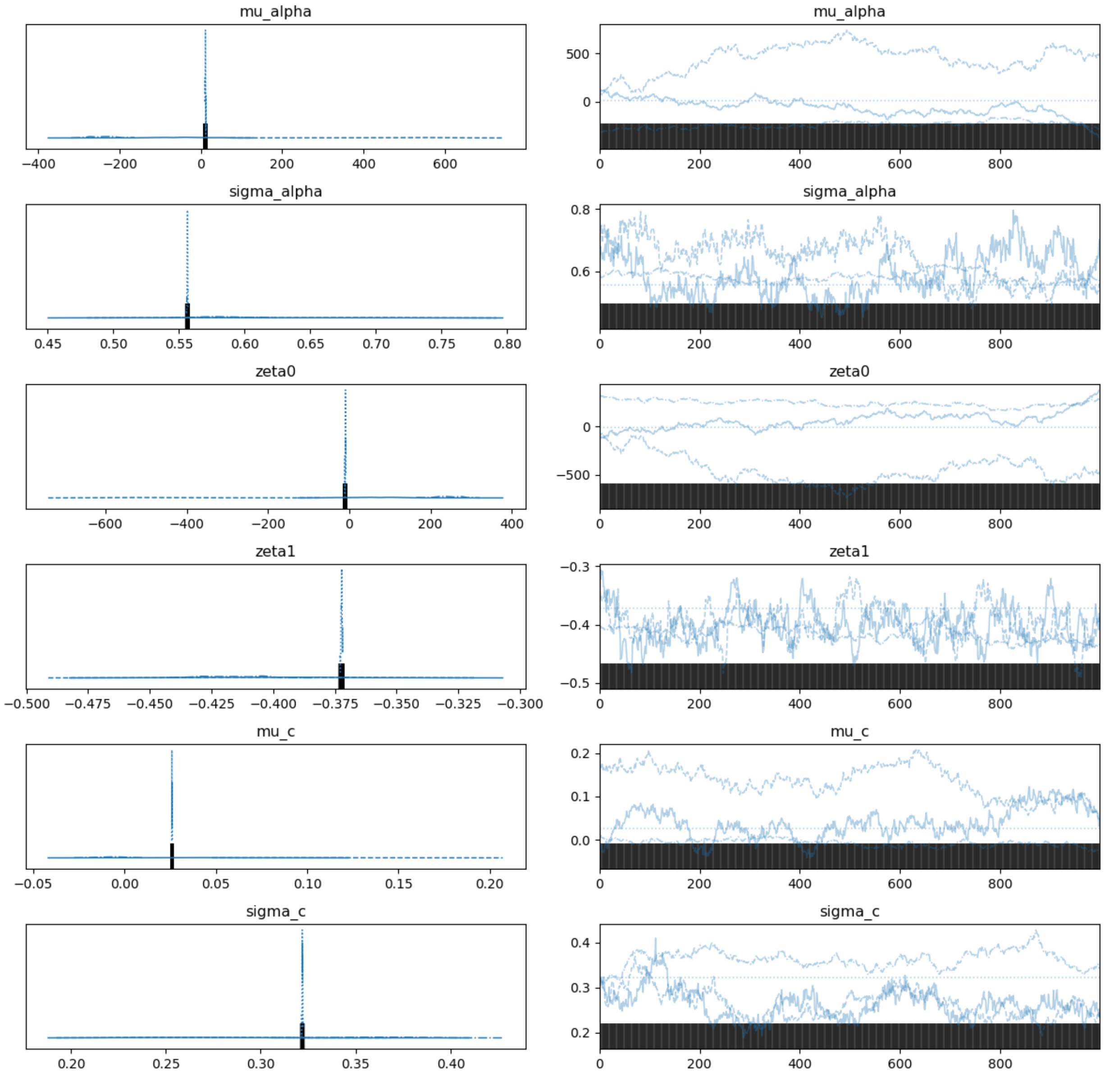

Figure 1: Trace plots for key parameters from the non-identified SDT model. Poor mixing and wandering chains are μ_α and ζ_0's lot.

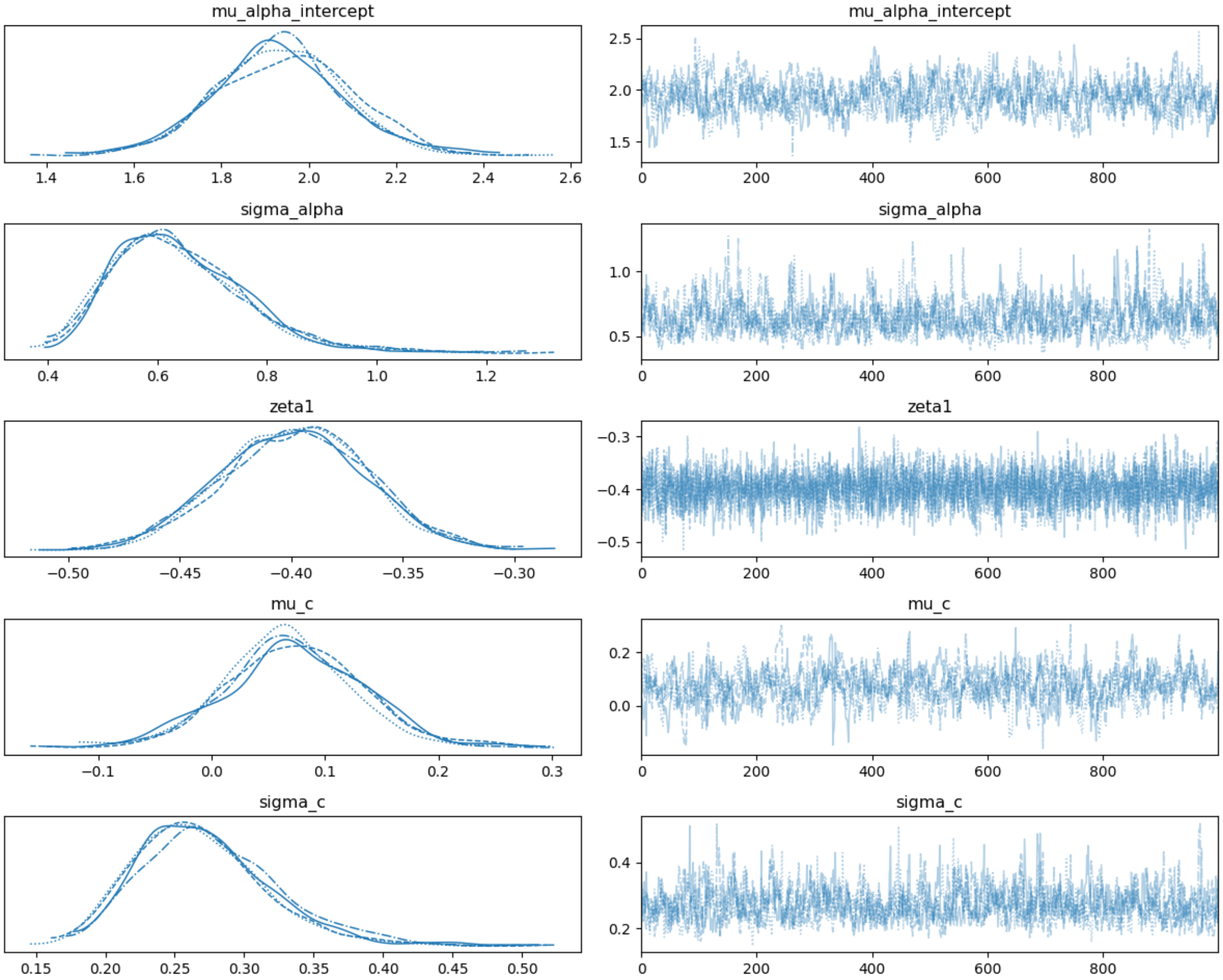

Figure 2: Trace plots for key parameters from the identified SDT model. Chains show good mixing and convergence.

Diagnosis Method 2: Pair Plots

Pair plots show the joint posterior distributions of pairs of parameters. This is a diagnostic for (possibly non-linear) correlations indicative of identifiability issues. The ‘arviz.plot_pair()‘ function, applied to ‘mu_alpha‘ and ‘zeta0‘, is useful here.

When examining pair plots for potentially confounded parameters, features to identify include:

- A distinct correlation forming a “ridge” across the scatter plot of posterior samples.
- Specifically for this additive confounding, a *negative, linear* correlation between ‘mu_alpha‘ and ‘zeta0‘.
- Marginal distributions along the diagonal that may appear unusually flat or poorly defined – they will have excess variance compared the true posterior distributions.

Figure 3 displays the pair plot for ‘mu_alpha‘ and ‘zeta0‘ from the non-identified model – a strong negative linear correlation is visible.

Strategies for Addressing Non-Identifiability

Non-identifiability must be addressed to maintain valid statistical inference. Common strategies include:

a) **Reparameterization (Preferred)**. This involves modifying the model’s parameterization to remove the inherent redundancy. Methods include:

   - Removing redundant parameters. In the example, removing ζ_0 eliminates the additive confounding.
   - Imposing sum-to-zero constraints on effects. This is an alternative method to resolve additive confounding but a little trickier to implement.
   - Modeling effects as contrasts relative to a baseline condition. This implicitly removes a redundant intercept.

   Reparameterization addresses the structural source of the issue.

b) **Informative Priors**. Applying informative priors to confounded parameters can constrain the posterior and improve MCMC convergence diagnostics. However, this does not resolve the *structural* non-identifiability. The resulting posterior distributions for the affected parameters will be heavily influenced by the chosen prior. That is not by itself a problem, but it is important not to forget that you may have made prior assumptions that drive your conclusions.

c) **Exogenous Predictors**. Adding an exogenous predictor that is not confounded with the parameters of interest can resolve the non-identifiability. In the example, replacing the mean person intercept μ_α with a function of a predictor variable (e.g., $\mu_\alpha = \zeta_1 \cdot \text{predictor}_i$) removes the confounding.

Even though this is not the preferred approach, let’s briefly look at the effect of informative priors.

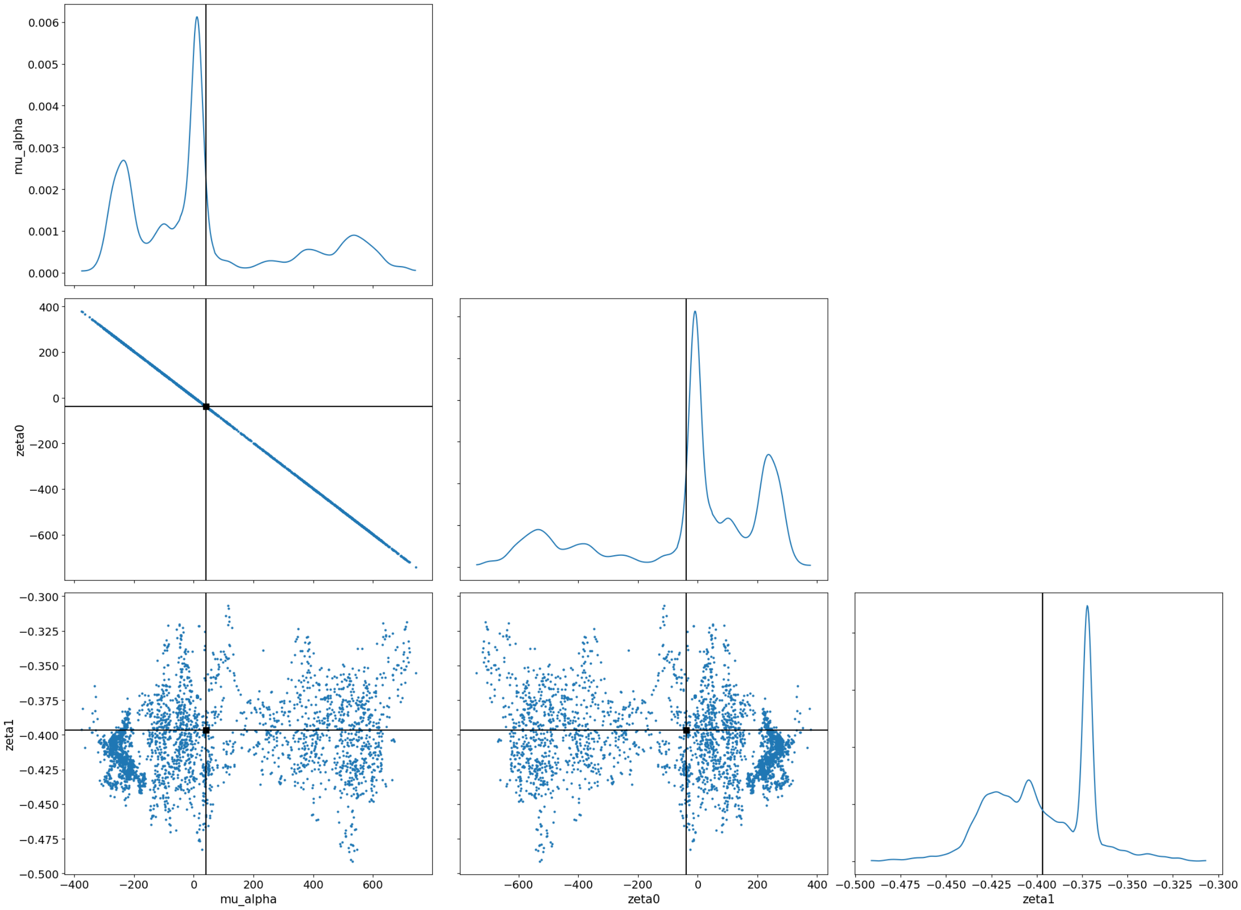

Figure 3: Pair plot for μ_α and ζ_0 from the non-identified SDT model. A strong negative linear correlation indicates confounding, and the marginal posteriors on the diagonal have strange complex shapes.

Effect of Informative Priors

Employing informative priors can lead to improved MCMC convergence statistics even in a structurally non-identified model. I sometimes call those models “classically unidentified” because they are not identified by the data, but they can be identified by the data and a Bayesian prior.

Using $\mu_\alpha \sim \text{Normal}(0, 1.5)$ (a more informative prior than before) in the non-identified SDT model yields convergence statistics shown below.

```
    --- Semi-Identified SDT Model Summary ---
    mean     sd  hdi_3%  hdi_97%  mcse_mean  mcse_sd  ess_bulk  ess_tail  r_hat
mu_alpha     1.301  0.827  -0.310    2.807      0.018    0.013    2023.0    2365.0   1.00
zeta0        0.629  0.825  -0.924    2.183      0.018    0.013    2002.0    2339.0   1.00
zeta1       -0.398  0.033  -0.459   -0.338      0.001    0.001    3915.0    2744.0   1.00
mu_c         0.067  0.064  -0.051    0.190      0.003    0.002     480.0     720.0   1.01
```

The $\hat{R}$ values are now near 1.00, and ESS values are considerably higher than in the standard non-identified fit. Trace plots (Figure 4) also appear more stable. This looks almost good!

However, the pair plot for ‘mu_alpha‘ and ‘zeta0‘ (Figure 5) continues to show the underlying negative correlation ridge, even if it is a little jittered now. While the prior aided convergence, the structural confounding remains.

Implementation of the Structural Fix: Reparameterization

The best approach to resolve the structural non-identifiability in this SDT model is reparameterization by removing the condition effect intercept ζ_0.

This adjustment necessitates a change in parameter interpretation. The parameter previously known as μ_α now represents the average d' when the predictor variable is zero. This parameter is renamed ‘mu_alpha_intercept‘ to reflect this. The parameter ζ_1 continues to represent the rate of change in d' with respect to the predictor, now relative to this new baseline.

The modified PyMC code for the identified model is as follows:

```python
# Priors for Sensitivity (d') parameters
mu_alpha_int = pm.Normal("mu_alpha_int", mu=0.0, sigma=1.0e9) # Was mu_alpha
sigma_alpha = pm.HalfNormal("sigma_alpha", sigma=1.0e9)
# zeta0 = pm.Normal("zeta0", mu=0.0, sigma=1.0e9) # <<< REMOVED
zeta1 = pm.Normal("zeta1", mu=0.0, sigma=1.0e9)

# Person sensitivity intercepts (drawn from mu_alpha_intercept)
alpha_p = pm.Normal("alpha_p", mu=mu_alpha_int, sigma=sigma_alpha)

# Condition Effects on Sensitivity (NO intercept)
gamma_i = pm.Deterministic("gamma_i", zeta1 * cond_pred_data)

# Calculate d' using alpha_p and gamma_i
d_ip = pm.Deterministic("d_ip", alpha_p[...] + gamma_i)

# SDT likelihood calculations are structurally identical
```

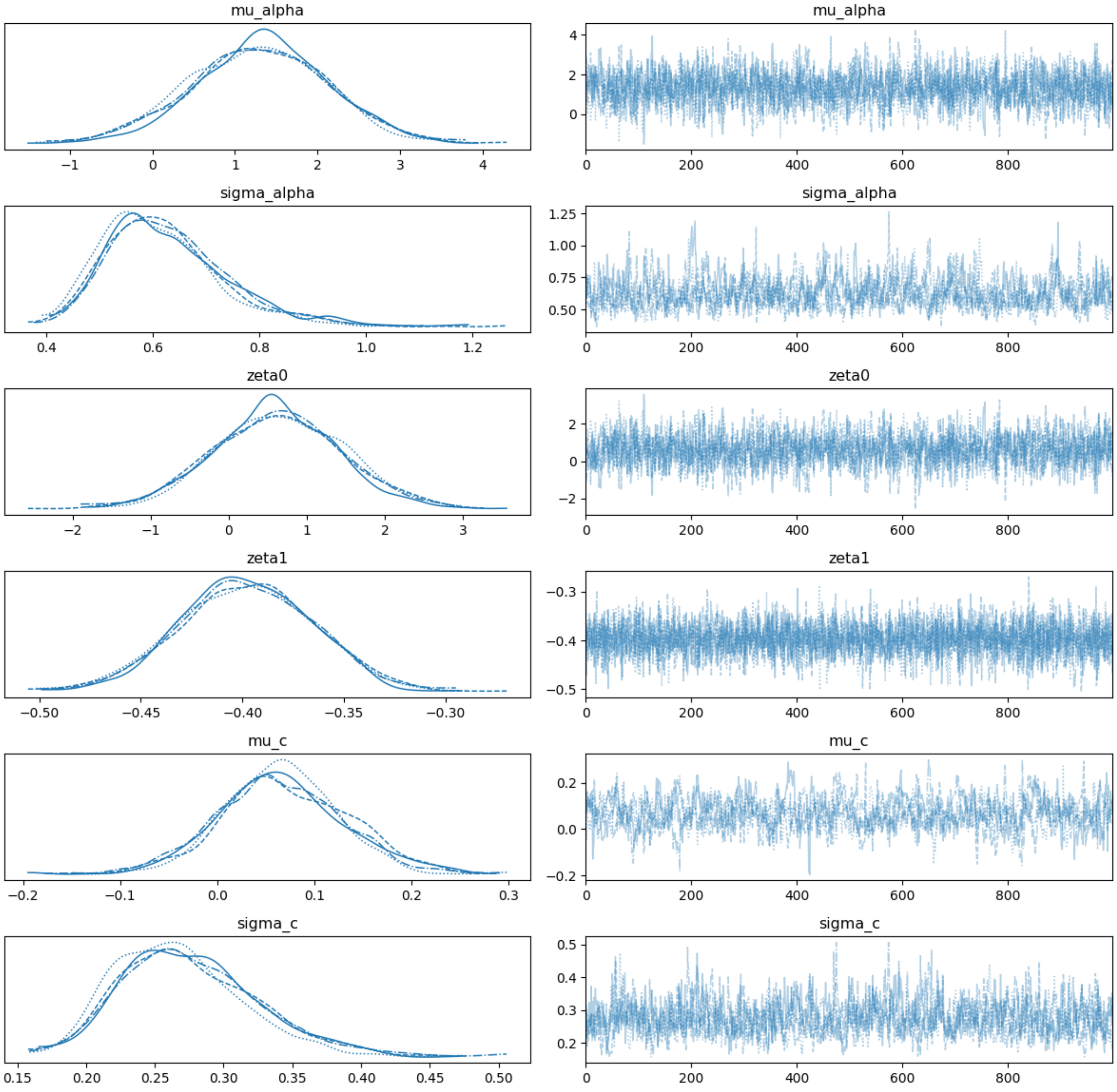

Figure 4: Trace plots from the model with an informative prior on μ_α. Apparent convergence based on traces.

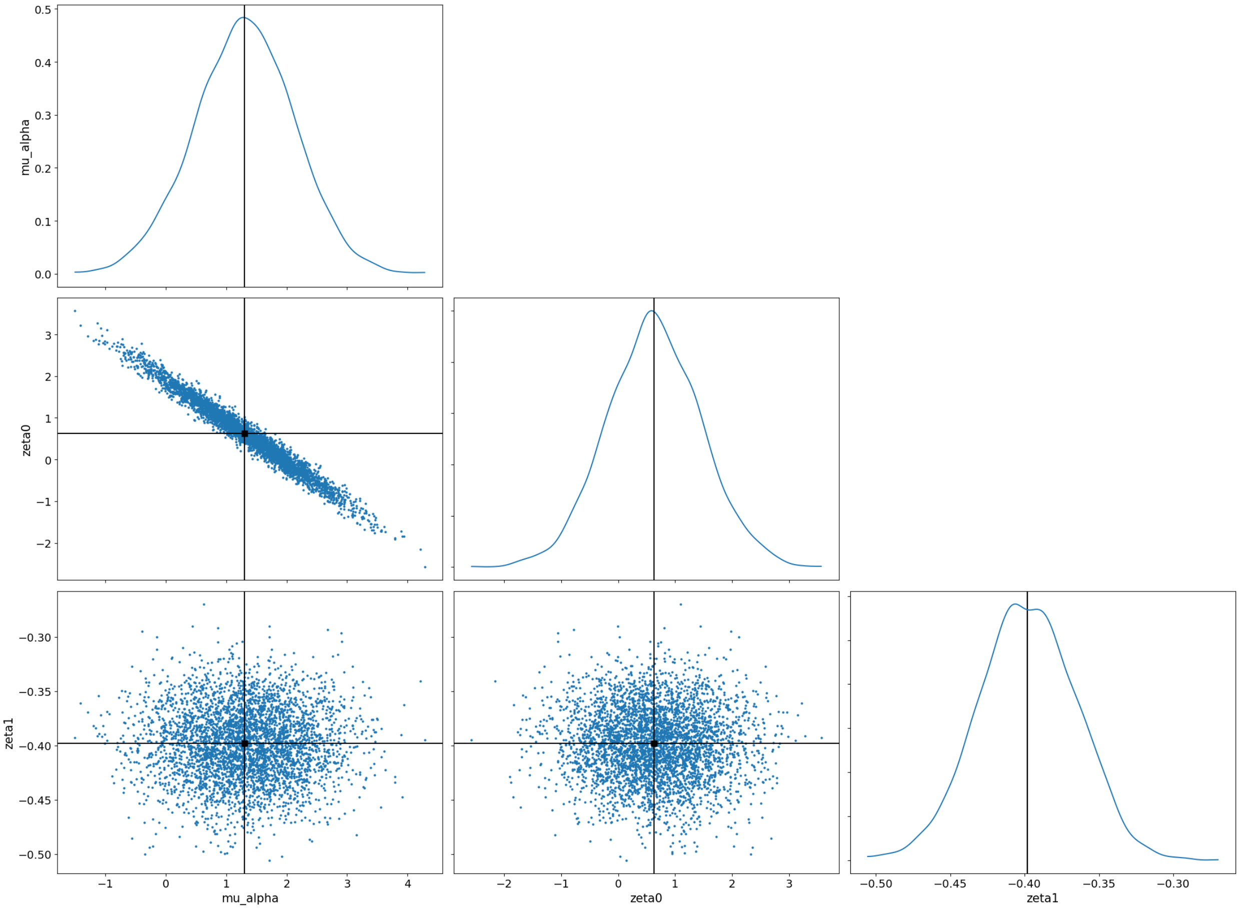

Figure 5: Pair plot for μ_α and ζ_0 from the model with informative prior. The negative correlation ridge persists.

```
# ...
```

Listing 2: PyMC code snippet for the identified SDT model parameters (removing ζ_0).

By removing the ‘zeta0‘ term and defining ‘gamma_i‘ solely based on ‘zeta1‘ and the predictor, the additive confounding is eliminated, resulting in a structurally identified model.

Results Following Reparameterization

If we now fit the reparameterized, identified model using MCMC, we see that the convergence and diagnostics are improved. (Table repeated from before.)

```
--- Identified SDT Model Summary ---
                mean     sd  hdi_3%  hdi_97%  mcse_mean  mcse_sd  ess_bulk  ess_tail  r_hat
mu_alpha_int   1.936  0.149   1.659    2.221      0.008    0.004     380.0     598.0   1.01
zeta1         -0.398  0.033  -0.463   -0.341      0.001    0.000    3558.0    3080.0   1.00
mu_c           0.072  0.062  -0.046    0.186      0.003    0.002     426.0     735.0   1.01
Sampling 4 chains for 1_500 tune and 1_000 draw iterations (6_000 + 4_000 draws total) took 12s.
```

We see:

- Good convergence ($\hat{R} \approx 1$) for all parameters.
- High Effective Sample Sizes (ESS).
- Well-mixed, stable trace plots (in Figure 2).

Following the removal of the confounding source, problematic correlations involving the parameter ‘mu_alpha_intercept‘ or other model parameters are gone.

Figure 6 shows the pair plot for key parameters from the identified model, including ‘mu_alpha_intercept‘. The absence of strong linear correlations shows that the structural non-identifiability has been successfully resolved.

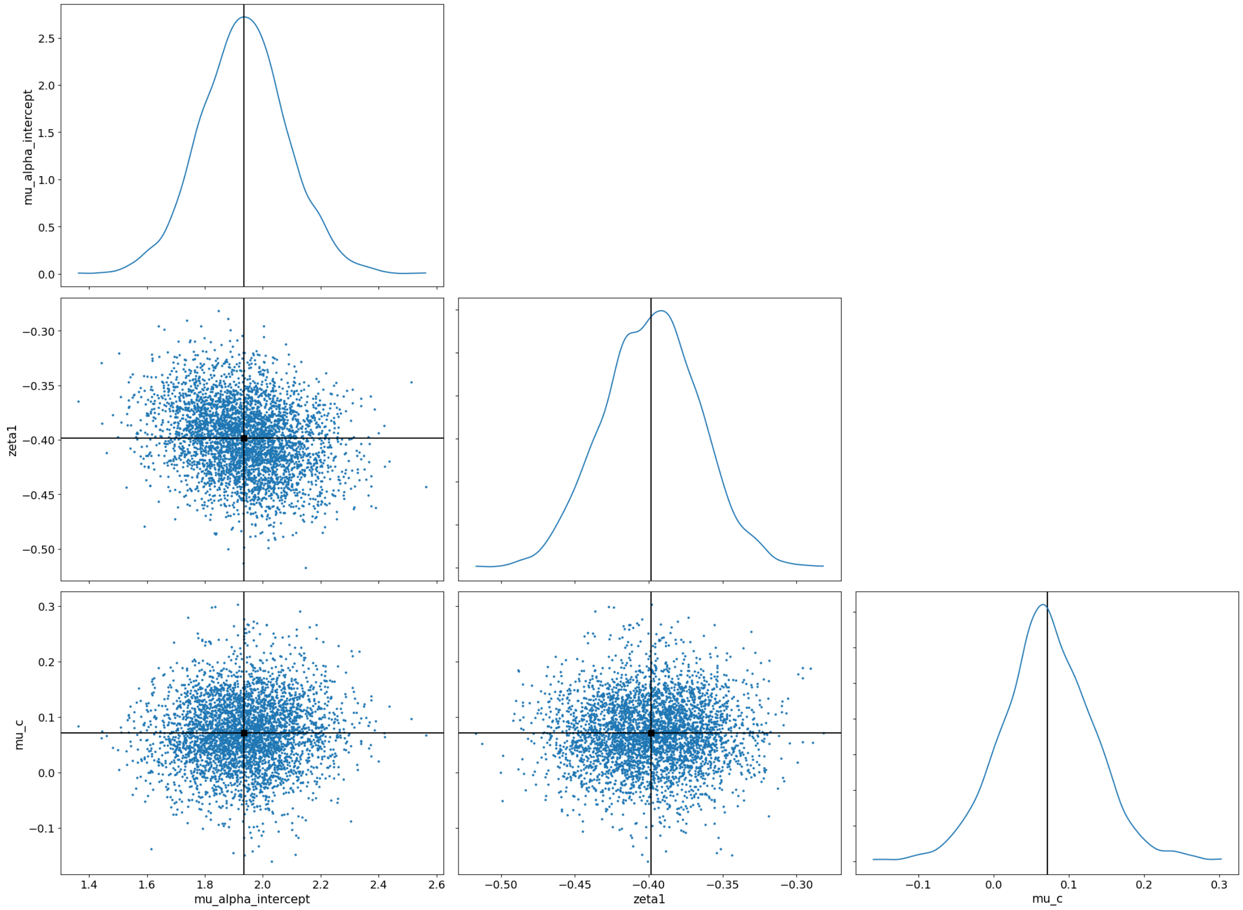

Figure 6: Pair plot for key parameters from the identified SDT model. No strong correlations are evident.

Summary Points

- Non-identifiability is a condition where distinct parameter sets produce identical data likelihoods, preventing unique parameter estimation.

- In hierarchical models, non-identifiability can arise from redundant parameters, commonly seen with additive intercepts in model components.

- The structure $d' = \text{person\_effect} + \text{condition\_effect}$ leads to non-identifiability if both effects incorporate their own intercept terms.

- Diagnosis relies on MCMC convergence statistics ($\hat{R}$, ESS) and visual checks such as trace plots and pair plots.

- The recommended solution for structural non-identifiability is reparameterization, which removes the underlying redundancy. While informative priors can sometimes assist convergence, they do not fundamentally resolve the structural issue.

- It is imperative to assess potential identifiability issues before interpreting model parameters. Models exhibiting convergence problems due to non-identifiability are unlikely to provide reliable inferences regarding the specific parameter values.