Noise Decomposition: Analysis of Variance For Binary Responses

1 Methods

1.1 Model Specification

Here we provide technical details on the method for estimating the contributions to the total response variation of signal and noise. We model the diagnostic decision of expert i on case j at occasion t as a binary response y_{ijt} arising from a Bernoulli distribution. The probability of a positive diagnosis (epilepsy) is given by a logistic function applied to a linear predictor:

$$y_{ijt} \sim \text{Bernoulli}(\pi(z_{ijt}))$$

with

$$\pi(z_{ijt}) = \frac{1}{1 + \exp(-z_{ijt})}$$

The linear predictor is specified as

$$z_{ijt} = c_j + \ell_i + p_{ij} + o_{ijt} + \bar{y},$$

where:

- c_j is the "signal" for case j,
- ℓ_i is "level noise" for expert i,
- p_{ij} is the idiosyncratic "pattern noise" capturing how expert i differentially weighs features of case j,
- o_{ijt} is the "occasion noise" for expert i on case j at occasion t, and
- $\bar{y}$ is a fixed offset (i.e., an overall intercept).

Each latent component is modeled as an independent Gaussian random variable:

$$\begin{aligned} c_j &\sim \mathcal{N}(0, \sigma_c^2), \\ \ell_i &\sim \mathcal{N}(0, \sigma_l^2), \\ p_{ij} &\sim \mathcal{N}(0, \sigma_p^2), \\ o_{ijt} &\sim \mathcal{N}(0, \sigma_o^2). \end{aligned}$$

To improve MCMC sampling (see below) in this hierarchical model, we adopt a non-centered parameterization. For example, the case effect is reparameterized as

$$z_{c,j} \sim \mathcal{N}(0, 1) \quad \text{and} \quad c_j = \sigma_c z_{c,j}.$$

Analogous non-centered formulations are used for ℓ_i, p_{ij}, and o_{ijt}.

1.2 Full Joint Distribution

Let $\phi(x; 0, \sigma^2)$ denote the density of a $\mathcal{N}(0, \sigma^2)$ random variable:

$$\phi(x; 0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{x^2}{2\sigma^2}\Big).$$

Then, the full joint density of the observed data $\{y_{ijt}\}$ and the latent variables $\{c_j\}$, $\{\ell_i\}$, $\{p_{ij}\}$, and $\{o_{ijt}\}$, given the parameters σ_c, σ_l, σ_p, and σ_o, is

$$\begin{aligned} &P\big(\{y_{ijt}\}, \{c_j\}, \{\ell_i\}, \{p_{ij}\}, \{o_{ijt}\} \mid \sigma_c, \sigma_l, \sigma_p, \sigma_o\big) = \\ &\quad \prod_{j=1}^{J} \phi(c_j; 0, \sigma_c^2) \prod_{i=1}^{I} \phi(\ell_i; 0, \sigma_l^2) \\ &\quad \times \prod_{i=1}^{I}\prod_{j=1}^{J} \phi(p_{ij}; 0, \sigma_p^2) \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{T} \phi(o_{ijt}; 0, \sigma_o^2) \\ &\quad \times \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{T} \Big\{ \pi\big(c_j + \ell_i + p_{ij} + o_{ijt} + \bar{y}\big)^{y_{ijt}} \\ &\qquad \times \Big[1 - \pi\big(c_j + \ell_i + p_{ij} + o_{ijt} + \bar{y}\big)\Big]^{1 - y_{ijt}} \Big\}. \end{aligned}$$

Because of the nonlinearity of the logistic function, the marginal likelihood (obtained by integrating out the latent variables) is analytically intractable.

1.3 Parameter Estimation via MCMC

We adopt a fully Bayesian approach by placing priors on the variance parameters σ_c, σ_l, σ_p, and σ_o (in our case, half-normal priors). Our target is the joint posterior distribution

$$P\big(\sigma_c, \sigma_l, \sigma_p, \sigma_o, \{c_j\}, \{\ell_i\}, \{p_{ij}\}, \{o_{ijt}\} \mid \{y_{ijt}\}\big),$$

which is proportional to the product of the full joint density in Equation (1) and the priors.

Because direct integration is infeasible, we use Markov chain Monte Carlo (MCMC) to sample from the posterior. Specifically, we employ the No-U-Turn Sampler (NUTS) as implemented via NumPyro in the PyMC framework. Our sampling configuration is as follows:

- 4 parallel chains,
- 2,000 tuning iterations per chain,
- 2,000 post-tuning draws per chain,
- A target acceptance rate of 0.98.

These settings are chosen to ensure adequate exploration of the high-dimensional posterior and to address potential issues related to the funnel-shaped geometry common in hierarchical models.

1.4 Variance Decomposition

After obtaining the posterior distributions, we compute the posterior means of the standard deviation parameters. To assess the relative contribution of each source of variability, we square these estimates to obtain variances. The total variance on the latent scale is given by

$$\text{Total Variance} = \sigma_c^2 + \sigma_l^2 + \sigma_p^2 + \sigma_o^2.$$

The percentage of total variance explained by each component is computed as

$$\%\ \text{Component} = \frac{\sigma^2_{\text{component}}}{\sigma_c^2 + \sigma_l^2 + \sigma_p^2 + \sigma_o^2} \times 100.$$

This decomposition allows us to compare the relative magnitude of the case signal with the various forms of noise (level, pattern, occasion) in the diagnostic judgments.