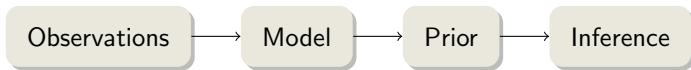# Prior elicitation for variance parameters in Bayesian hierarchical models

Ingeborg Gullikstad Hem
Norwegian University of Science and Technology
Department of Mathematical Sciences

June 28th 2022

We consider *latent Gaussian models* where the linear predictor is Gaussian conditional on the model parameters

```
Observations  →  Model  →  Prior  →  Inference
```
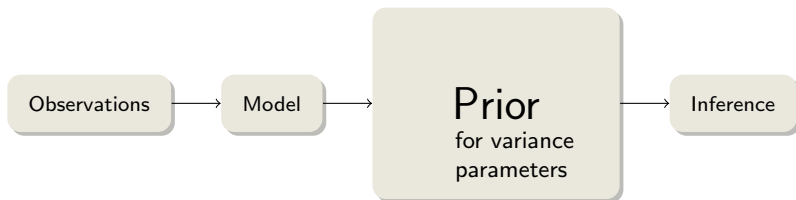
We consider *latent Gaussian models* where the linear predictor is Gaussian conditional on the model parameters
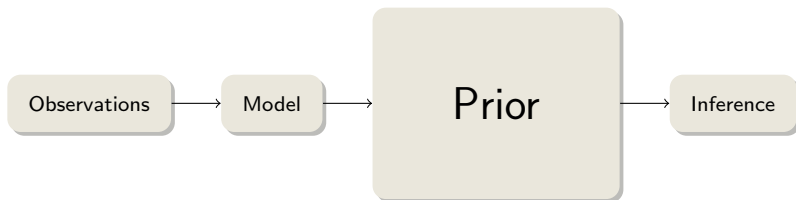
We consider *latent Gaussian models* where the linear predictor is Gaussian conditional on the model parameters

Note: We do not focus on fixed effects, and assign wide Gaussian priors

Outline

- Motivating example: Wheat breeding
- Standard prior approach
- New approach: Hierarchical variance decomposition
- New method: Hierarchical decomposition prior framework
- New tool: R package `makemyprior`
- Neonatal mortality
- Discussion and summary

Motivating example: Wheat breeding

- **Purpose:** More grain yield (weight of grain from each strain)
- **How:** Find which strains of wheat are genetically best
- **Challenge:** Separate environmental (residual) and genetic variation

- Observations: Grain yield y_i (phenotype) and genetic markers
- Model: $y_i = \mu + \underbrace{a_i + d_i + x_i}_{g_i} + \varepsilon_i, \ i = 1, \dots, 100$
  - ▶ $\mu \sim \mathcal{N}(0, \sigma = 1000)$ (intercept)
  - ▶ $* \sim \mathcal{N}(\mathbf{0}, \sigma_*^2 \mathbf{V}_*)$, $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$
  - ▶ Genetic = additive + nonadditive (= dominance + epistasis)
- Prior: *Use information from a geneticist*
- Inference: Stan through `rstan`
- See Hem *et al.* (2021)[1] for details

[1]Ingeborg Gullikstad Hem et al. "Robust Modeling of Additive and Nonadditive Variation With Intuitive Inclusion of Expert Knowledge". In: *Genetics* 217.3 (2021). iyab002. DOI: 10.1093/genetics/iyab002.

Information from the geneticist

1. The broad-sense heritability ($\frac{\sigma_g^2}{\sigma_g^2 + \sigma_\varepsilon^2}$) is around 25%
2. Additive (**a**), dominance (**d**) and epistasis (**x**) account for most of, some of and a little of the genetic variance (σ_g^2)

Information from the geneticist

1. The broad-sense heritability ($\frac{\sigma_g^2}{\sigma_g^2 + \sigma_\varepsilon^2}$) is around 25%
2. Additive (a), dominance (d) and epistasis (x) account for most of, some of and a little of the genetic variance (σ_g^2)
   85%   **10%**   **5%**

Information from the geneticist

1. Total variation (y) is 25% genetic ($a + d + x$) and 75% environmental (ε)
2. Genetic variation is 85% additive (a) and 15% nonadditive ($d + x$)
3. Nonadditive variation is 67% dominance (d) and 33% epistasis (x)
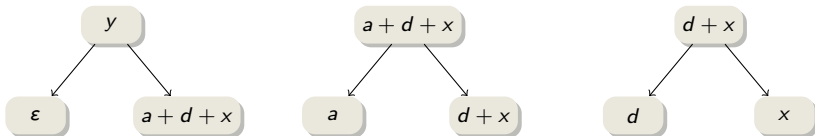
Standard prior approach

1. Total variation (y) is 25% genetic ($a + d + x$) and 75% environmental (ε)
2. Genetic variation is 85% additive (a) and 15% nonadditive ($d + x$)
3. Nonadditive variation is 67% dominance (d) and 33% epistasis (x)

- Four variances: σ_ε^2, σ_a^2, σ_d^2, σ_x^2
- Independent variance priors (inverse gamma, half-Cauchy, penalized complexity prior[2], ...)
- Prior info is not about the variances directly
- No information about magnitude

---

[2]Daniel Simpson et al. "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors". In: *Statistical Science* 32.1 (2017), pp. 1–28.
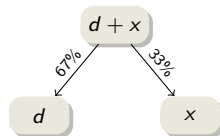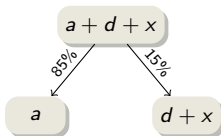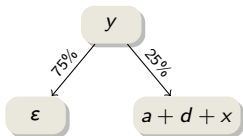
Hierarchical variance decomposition

1. Total variation (y) is 25% genetic ($a + d + x$) and 75% environmental (ε)
2. Genetic variation is 85% additive (a) and 15% nonadditive ($d + x$)
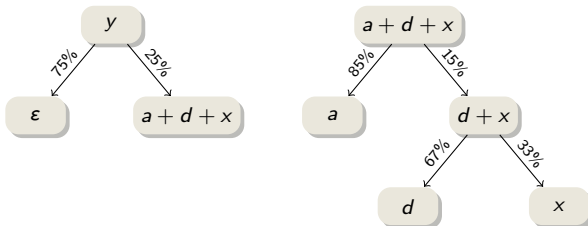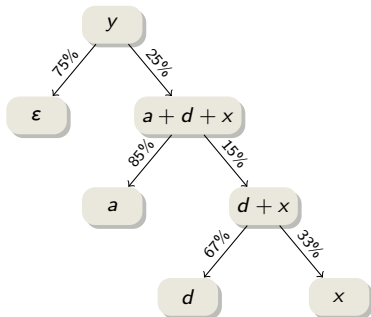3. Nonadditive variation is 67% dominance (d) and 33% epistasis (x)

Hierarchical variance decomposition

1. Total variation (y) is 25% genetic ($a + d + x$) and 75% environmental (ε)
2. Genetic variation is 85% additive (a) and 15% nonadditive ($d + x$)
3. Nonadditive variation is 67% dominance (d) and 33% epistasis (x)

Hierarchical variance decomposition

1. Total variation (y) is 25% genetic ($a + d + x$) and 75% environmental (ε)
2. Genetic variation is 85% additive (a) and 15% nonadditive ($d + x$)
3. Nonadditive variation is 67% dominance (d) and 33% epistasis (x)

Hierarchical variance decomposition

1. Total variation (y) is 25% genetic ($a + d + x$) and 75% environmental (ε)
2. Genetic variation is 85% additive (a) and 15% nonadditive ($d + x$)
3. Nonadditive variation is 67% dominance (d) and 33% epistasis (x)

Hierarchical decomposition prior framework[3]

- Create joint prior for variance parameters from a prior tree
- Prior tree is based on the model structure and prior (expert) knowledge
- Assigning prior to the total latent variance (sum of variances)
  - ▶ Empirical data variation does not depend on model
  - ▶ Straight forward to add and remove model components
  - ▶ Any variance prior is possible, Jeffreys' prior useful for Gaussian data
- Assigning priors to splits in tree (representing variance partitions)
  - ▶ Ignorance through symmetric Dirichlet priors
  - ▶ Shrinkage through PC priors
- Keep fixed effects outside the HD prior framework
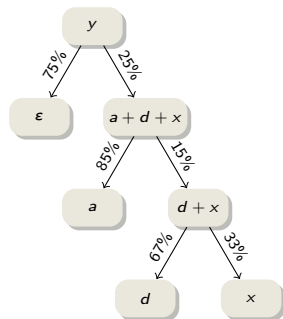
[3]Geir-Arne Fuglstad et al. "Intuitive Joint Priors for Variance Parameters". In: *Bayesian Analysis* 15.4 (2020), pp. 1109–1137.

Wheat breeding priors

Standard:

$$\sigma_\varepsilon^2 \sim \pi(\mathrm{m} = V \cdot 0.75)$$

$$\sigma_a^2 \sim \pi(\mathrm{m} = V \cdot 0.25 \cdot 0.85)$$

$$\sigma_d^2 \sim \pi(\mathrm{m} = V \cdot 0.25 \cdot 0.15 \cdot 0.67)$$

$$\sigma_x^2 \sim \pi(\mathrm{m} = V \cdot 0.25 \cdot 0.15 \cdot 0.33)$$

Must choose some V, and update all priors when changing model

Hierarchical:

$$V_{a+d+x+\varepsilon} \sim \text{Jeffreys'}$$

$$\omega_{\frac{g}{g+\varepsilon}} \sim \pi(\mathrm{m} = 0.25)$$

$$\omega_{\frac{a}{g}} \sim \pi(\mathrm{m} = 0.85)$$

$$\omega_{\frac{d}{d+x}} \sim \pi(\mathrm{m} = 0.67)$$

# `makemyprior`[4]

- HD prior
- Compatible with `rstan` and INLA

[4]Ingeborg Gullikstad Hem, Geir-Arne Fuglstad, and Andrea Riebler. `makemyprior`: *Intuitive Construction of Joint Priors for Variance Parameters in R*. *arXiv preprint* arXiv:2105.09712v2 [stat.CO]. 2022, https://CRAN.R-project.org/package=makemyprior

Wheat breeding

```
formula <- y ~
    mc(a, model = "generic0", Cmatrix = Q_a, constr = TRUE) +
    mc(d, model = "generic0", Cmatrix = Q_d, constr = TRUE) +
    mc(x, model = "generic0", Cmatrix = Q_x, constr = TRUE)
prior <- make_prior(formula, wheat_data)

new_prior <- makemyprior_gui(prior)
res <- inference_stan(new_prior)
```
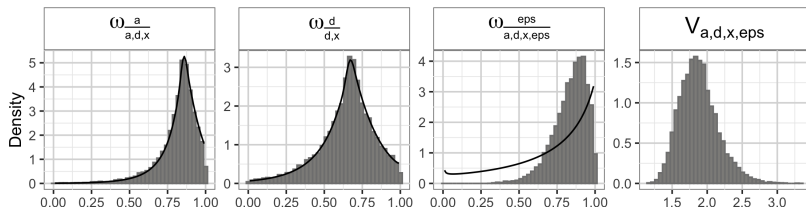

- Without the prior knowledge, ε and x are too hard to separate
- Standard approach leads to less stable inference than hierarchical approach
- Improved selection of genetically best individuals with the hierarchical approach

Neonatal mortality in Kenya

- Observations: Number of live births $b_{i,j}$ and neonatal[5] deaths $y_{i,j}$ in cluster j in county i, urbanity indicator of cluster $x_{i,j}$ (urban = 1)
- Model:

$$y_{i,j}|b_{i,j}, p_{i,j} \sim \text{Bin}(b_{i,j}, p_{i,j}), \ i = 1, \ldots, 47, \ j = 1, \ldots, m_i$$
$$\text{logit}(p_{i,j}) = \mu + x_{i,j}\beta + u_i + v_i + \nu_{i,j}$$

  ▶ $\mu, \beta \sim \mathcal{N}(0, \sigma = 1000)$
  ▶ $v_i, \nu_{i,j} \sim N(0, \sigma_*^2)$,
    u is Besag effect with variance σ_u^2
- Prior: Use spatial structure and avoid overfitting
- Inference: Stan using `rstan`
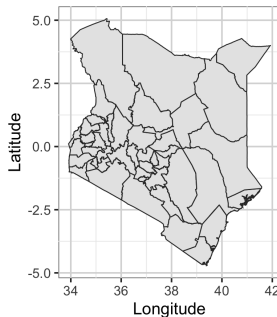
[5] First 30 days of life

Neonatal mortality

- Prefer coarser over finer effects (explain data at coarser level if possible)
  $\implies u + v > \nu$
- Prefer unstructured over structured effects (reduce risk of estimating spurious effects) $\implies v > u$
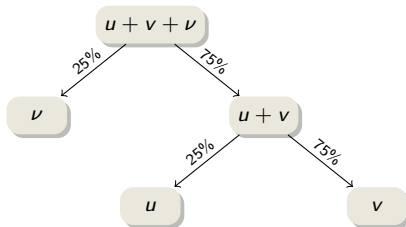
Neonatal mortality

- Prefer coarser over finer effects (explain data at coarser level if possible)
  $\implies u + v > \nu$
- Prefer unstructured over structured effects (reduce risk of estimating spurious effects) $\implies v > u$
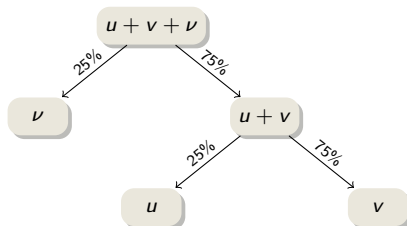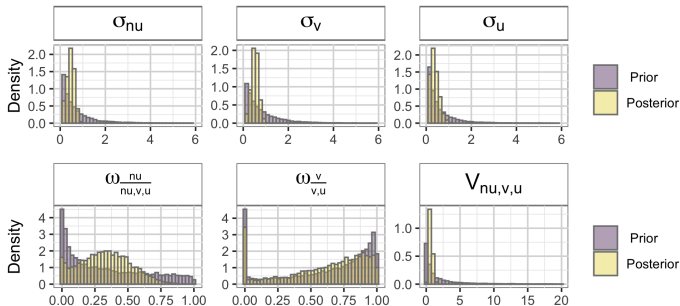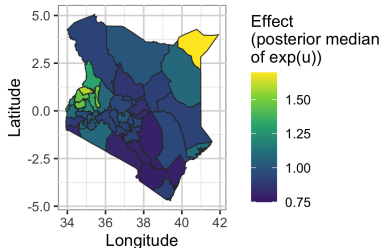

- PC priors for variance proportions with shrinkage to (u+v) and (v)
- PC prior for total variance, such that $[0.1, 10]$ is a 90% credible interval for $\exp(u_i + v_i + \nu_{i,j})$: PC(3.35, 0.05)

Neonatal mortality

```
formula <- y ~ urban + mc(nu) + mc(v) +
    mc(u, model = "besag", graph = graph_path,
       scale.model = TRUE)
prior <- make_prior(formula, neonatal_data,
    family = "binomial")

new_prior <- makemyprior_gui(prior)
res <- inference_stan(new_prior)
```

Discussion and summary

- Non-variance parameters (such as a range) can be assigned independent priors
- Idea not restricted to PC and Dirichlet priors
- Can combine several prior trees
- Taking the model structure into account

- Visualization eases communication
- Prior knowledge can be incorporated at the level it applies to
- One, but not the only, way to construct priors
- Simple to explore options with `makemyprior`

- All priors are choices, *make good ones*

Prior elicitation for variance parameters in Bayesian hierarchical models

ingeborg.hem@ntnu.no

References I

📄 Fuglstad, Geir-Arne et al. "Intuitive Joint Priors for Variance Parameters". In: *Bayesian Analysis* 15.4 (2020), pp. 1109–1137.

📄 Hem, Ingeborg Gullikstad, Geir-Arne Fuglstad, and Andrea Riebler. `makemyprior`: *Intuitive Construction of Joint Priors for Variance Parameters in R*. arXiv preprint arXiv:2105.09712v2 [stat.CO]. 2022.

📄 Hem, Ingeborg Gullikstad et al. "Robust Modeling of Additive and Nonadditive Variation With Intuitive Inclusion of Expert Knowledge". In: *Genetics* 217.3 (2021). iyab002. DOI: 10.1093/genetics/iyab002.

📄 Simpson, Daniel et al. "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors". In: *Statistical Science* 32.1 (2017), pp. 1–28.