Hierarchical Models

Lecture 5 Scribing

Ammar Raṣid

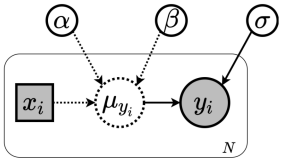
School of Electrical Engineering and Computer Science
University of Ottawa

March 8, 2022

Overview

1. Bayesian Regression
2. Hierarchical vs. Pooled Bayesian Regression
3. Robustness Analysis & Model Checking

Bayesian Regression



- Assumes prior over regression parameters $\theta = \{\alpha, \beta_1, ...\beta_K, \sigma\}$.
- Noise often assumed to be a zero-mean Gaussian $\epsilon \sim \mathcal{N}(0, \sigma)$.
- The goal is to learn the posterior $p(\theta|D)$.
- $\hat{y}$ is sampled given k-dimensional features X, and the distributions of parameters θ.
- X can be transformed by any function ϕ (e.g. polynomial, logarithmic, ...etc.).

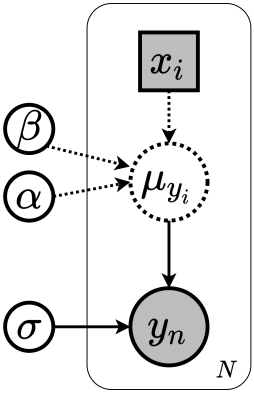
$$p(\hat{y}|X, \theta) \sim \mathcal{N}\left(\alpha + \sum_{k=1}^{K} \beta_i \phi(x_k), \epsilon\right)$$

Pooled vs. Separate vs. Hierarchical

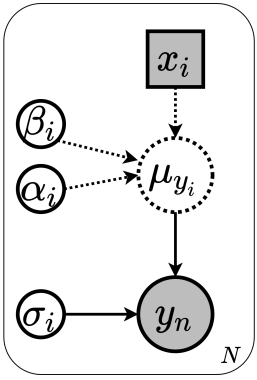
Considering observations of the same attributes from J groups:

- **Pooled** models assume all observations have the same θ, ignoring their group info.
- **Separate** models assume personalized θs for each observation.
- **Hierarchical** models assume different θs for observations of each group, while all θs are sampled from a common distribution.

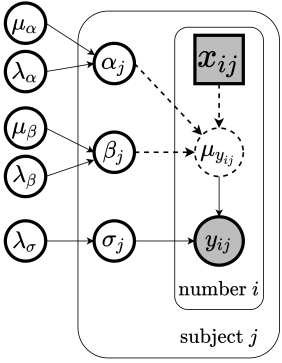
Pooled vs. Separate vs. Hierarchical



(a) Pooled



(b) Separate



(c) Hierarchical

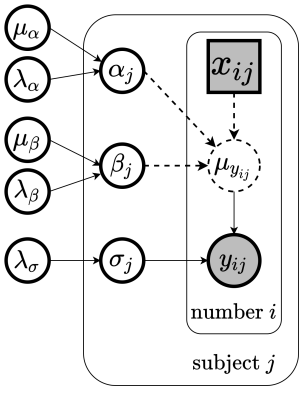
Example: Predicting BP by PAT

- Blood pressure is normally measured using cuff based methods. New contactless methods for blood pressure measurements rely on measuring time difference between two biomedical waveforms. This time difference is called pulse transit time or pulse arrival time (PAT).

$$BP = \alpha + \beta * \ln(PAT) + \epsilon$$

- The dataset contains multiple PAT and BP readings from 10 subjects.

Hierarchical Bayesian Regression



Common Priors

$$\mu_\alpha \sim \mathcal{N}(0, 10)$$
$$\lambda_\alpha \sim \mathit{Gamma}(0.01, 0.01)$$
$$\mu_\beta \sim \mathcal{N}(0, 10)$$
$$\lambda_\beta \sim \mathit{Gamma}(0.01, 0.01)$$
$$\lambda_\sigma \sim \mathit{Gamma}(0.01, 0.01)$$

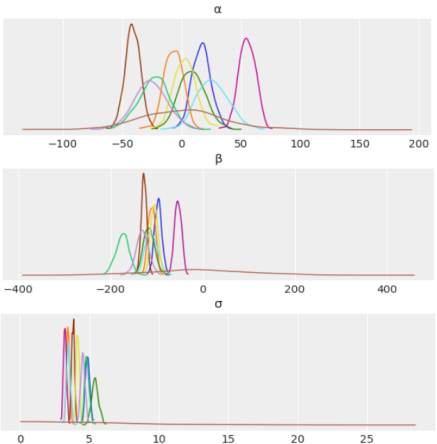
Posterior

$$y_{ij} \sim \mathcal{N}(\alpha_j + \ln(x_{ij})\beta_j, \sigma)$$

Group Priors

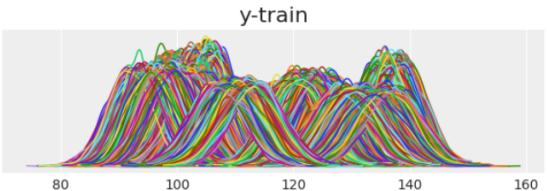
$$\alpha_j \sim \mathcal{N}(\mu_\alpha, 1/\sqrt{\lambda_\alpha})$$
$$\beta_j \sim \mathcal{N}(\mu_\beta, 1/\sqrt{\lambda_\beta})$$
$$\sigma_j \sim |\mathcal{N}(0, 1/\sqrt{\lambda_\sigma})|$$

Hierarchical Posterior: θ



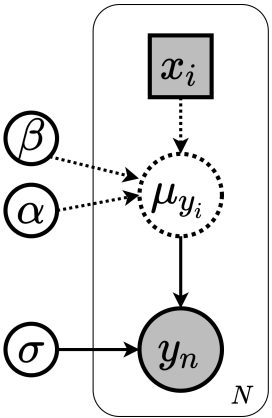
Posterior: $P(\theta|D)$

- The hierarchical model has learnt different αs, βs, and σs for different subjects.



Posterior Predictive: $P(y|X, \theta)$

Pooled Bayesian Regression



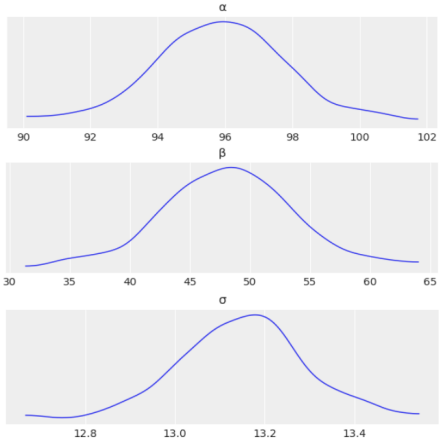
Priors

$$\alpha \sim \mathcal{N}(0, 10)$$
$$\beta \sim \mathcal{N}(0, 10)$$
$$\sigma \sim |\mathcal{N}(0, 10)|$$

Posterior

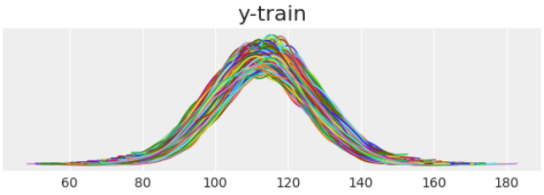
$$y_{ij} \sim \mathcal{N}(\alpha + \ln(x_{ij})\beta, \sigma)$$

Pooled Posterior: θ



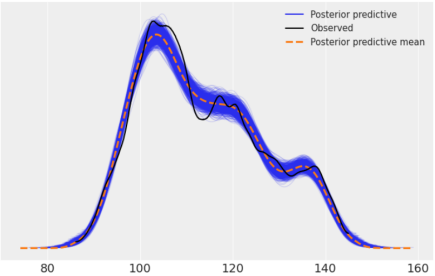
Posterior: $P(\theta|D)$

- The pooled model is too constrained to learn different αs, βs, and σs, for different subjects.

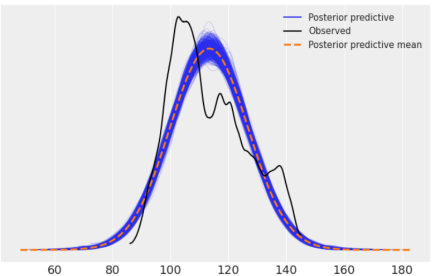


Posterior Predictive: $P(y|X,\theta)$

Hierarchical vs Pooled: Posterior Predictive Check (PPC)



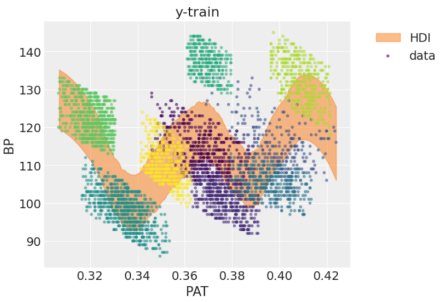
(a) Hierarchical Model's PPC



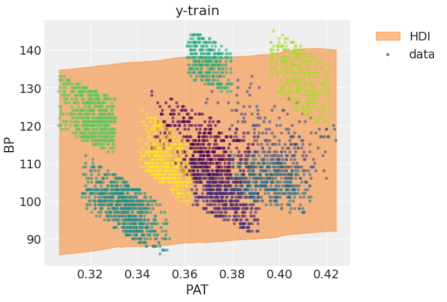
(b) Pooled model's PPC

- The hierarchical model has fit the data noticeably better as it learnt different Gaussians for different subjects.
- The pooled model learnt a single normal distribution around the data mean.

Hierarchical vs Pooled: Highest Density Interval (HDI) Training



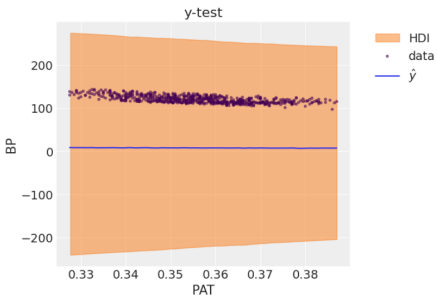
(a) Hierarchical Model's HDI



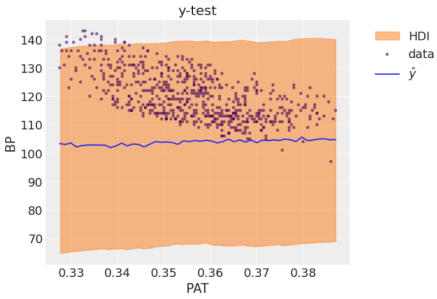
(b) Pooled model's HDI

- The hierarchical model reported a noticeably much narrower HDI than the pooled model on the training set.
- The HDI of the pooled model shows a high level of uncertainty due to the underfitting of the data.

Hierarchical vs Pooled: Highest Density Interval (HDI) Testing



(a) Hierarchical Model's HDI

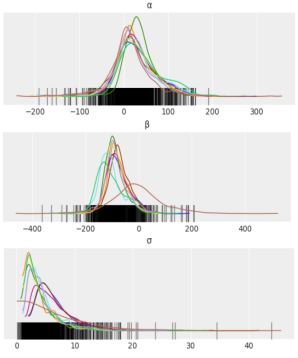


(b) Pooled model's HDI

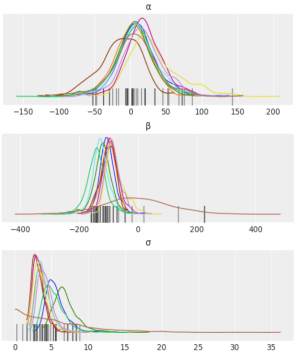
- The hierarchical model's HDI on the test subject is impractically wide, with a mean $\hat{y}$ close to zero due to the flattened parameters.
- Whereas the pooled model wide HDI seemed much narrower compared to the hierarchical model's HDI.

Robustness Analysis: Hierarchical $P(\theta|D)$

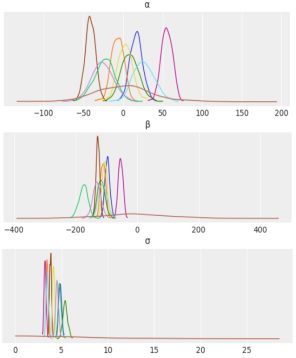
As the number of data points per subject increases, The hierarchical model can distinguish between subjects more clearly, while the parameters corresponding to the test subject (i.e. the brown curve) flatten/die out.



(a) 3 dp / subject



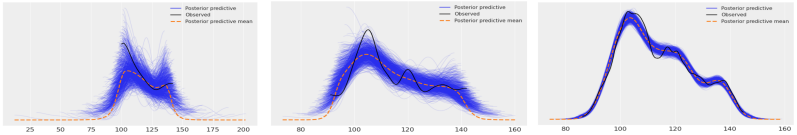
(b) 10 dp / subject



(c) All dp / subject

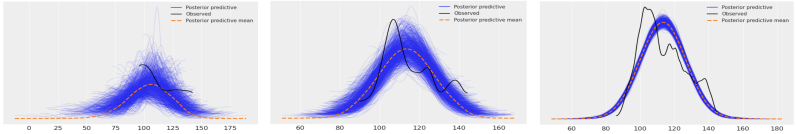
Robustness Analysis: PPC

Hierarchical Model



(a) 3 dp / subject (b) 10 dp / subject (c) All dp / subject

Pooled Model



(a) 3 dp / subject (b) 10 dp / subject (c) All dp / subject

As the number of data points per subject increase, the posterior predictive tend to have lower variance.

Model Check: WAIC & LOO

The hierarchical model yields lower error on the training data consistently, and has higher supportive evidence (i.e. lower WAIC and LOO). However, since the test subject has not been seen in the training data, The hierarchical model yields higher error as the support for training subjects increases, and the support for testing subjects fades in comparison.

dp/subject	Model	WAIC	LOO	Train MAE	Test MAE
3	Hierarchical	161.85	163.28	2.12	99.72
	Pooled	236.06	236.08	13.63	16.61
10	Hierarchical	523.19	524.97	3.14	109.26
	Pooled	725.65	725.66	11.04	8.65
All	Hierarchical	23,564.9	23,564.9	3.47	112.63
	Pooled	33,320.2	33,320.2	11.26	8.62

Table: Comparing WAIC & LOO of Hierarchical and Pooled Models