

Great! Here's a **Bayesian assignment with a completely different context**, still emphasizing key learning objectives like mixed effects modeling, Bayesian inference, bias estimation, and use of priors — but this time in the **context of online product reviews and ratings**.

Bayes Assignment 7 of 2025: Online Product Ratings Analysis

Dr. Sean van der Merwe
Instructions

Context:

You are a data scientist at a company managing a large e-commerce platform. Each product on the platform is rated by customers who purchased it. However, not all customers leave reviews, and those who do may differ in generosity, severity, or reliability. Furthermore, products belong to categories (e.g., electronics, clothing), and some categories receive more polarised ratings than others.

Your task is to use **Bayesian hierarchical modeling** to estimate the **true quality of products** while accounting for **customer bias**, **product category effects**, and **missingness in the rating data**.

Tasks:

- 1. What are the sources of residual (error) variability in this problem?**
Consider factors like customer mood, delivery issues, personal preferences, etc.
- 2. If every customer rated every product they bought and was perfectly objective,** would averaging the scores be fair? What assumptions would this rely on, and what hidden assumptions may still be problematic?
- 3. Summarize your assigned dataset.**
Explore:
 - Missingness patterns (e.g., which products have few ratings?)
 - Skewness or multimodality in rating distributions
 - Differences between product categories
- 4. Reshape the data into long form for modeling.**
Each row should include: `product_id`, `customer_id`, `rating`, `product_category`, and any available covariates (e.g., `verified_purchase`, `review_length`).
- 5. Specify which effects are fixed and which are random.**
 - Should product quality be modeled as a random effect?
 - What about customer bias, or category-specific shifts?
 - Justify your choices.
- 6. Fit a Bayesian mixed effects model using vague priors.**
Model the rating as a function of product quality, customer bias, and category effect. Assume a simple error structure for residuals.

7. **Produce model-based estimates for each product's true quality.**

Include:

- Posterior means or medians
- 95% credible intervals
- Prediction intervals for future ratings

8. **Estimate customer-level biases.**

Who are the harshest and kindest raters? Who is most consistent?

9. **Now bring in external product information (e.g., expert reviews or brand reputation scores).**

Use this to construct **subjective priors** on product quality. Fit an updated model and discuss:

- How did the posterior estimates shift?
- Are the intervals narrower or wider?
- Is this approach fair across product categories?

10. **Suppose some reviews are suspected to be fake or influenced (e.g., bots or paid ratings).**

Propose a strategy to incorporate **credibility scores** or **review trustworthiness indicators** into the model. How would you identify and downweight suspicious ratings?

11. **Ensure auditability and reproducibility of your work:**

- Create two timestamped backups before submission
- Use GitHub for version control
- Submit a public repository link to earn bonus marks

Would you like this version packaged with a simulated dataset, data dictionary, or code starter pack in R or Python? Let me know your preference!