Math76 Summer 2020

Introduction to Bayesian Computation

Lecture 22:
Posterior Predictive Checks and Basic Hierarchical Modeling
12 August 2020

Typical Workflow

1. Find data y.
    - e.g., Height of a baseball after being hit.
2. Model data using a likelihood function $f(y|x)$.
    - e.g., Physics model of baseball trajectory with additive noise.
3. Develop prior distribution $f(x)$.
    - e.g., Typical range of initial velocities and positions.
4. Sample posterior $f(x|y) \propto f(y|x)f(x)$.
    - e.g., run MCMC
5. Use samples to answer questions

$$\mathbb{P}[x \in A] \approx \frac{1}{N} \sum_{k=1}^{N} I\left[x^{(k)} \in A\right]$$

The Problem

Analysis is entirely dependent on posterior predictive distribution providing a "good" or "consistent" representation of the data and apriori information.

↳ Our assumptions are valid

↳ $f(y|x)$ is a "good" statistical representation of the process that generated the data

↳ prior isn't adversely biasing the posterior

Y := RV representing the "observable" quantity

Y_{obs} := Vector of observed values

Y_{pred} := RV representing model predictions after observing Y_{obs}

$$\left.\begin{aligned} f(y) &= \int f(y,x)\,dx \\ &= \int f(y|x)\,\underline{f(x)}\,dx \end{aligned}\right\} \text{prior predictive density}$$

↳ evidence is $f(y = Y_{obs}) = f(Y_{obs})$

$$f(Y_{pred} \mid Y_{obs}) = \int f(Y_{pred} \mid x)\,\underline{f(x \mid Y_{obs})}\,dx \; \Big\} \text{ posterior predictive density}$$

Example: Regression

$$y = Vx + \varepsilon \qquad x \sim N(\mu_0, \Sigma_0)$$
$$\varepsilon \sim N(0, \Sigma_\varepsilon)$$

$$f(y) = ?$$

$$y \sim N\left(V\mu_0,\ V\Sigma_0 V^T + \Sigma_\varepsilon\right)$$ prior predictive distribution

$$y_{pred} = V x_{post} + \varepsilon$$
$$x_{post} \sim f(x \mid y_{obs}) = N\left(\mu_{post},\ \Sigma_{post}\right)$$

$$y_{pred} \sim N\left(V\mu_{post},\ V\Sigma_{post}V^T + \Sigma_\varepsilon\right)$$

How does $f(y_{pred})$ compare to my observations y_{obs}?

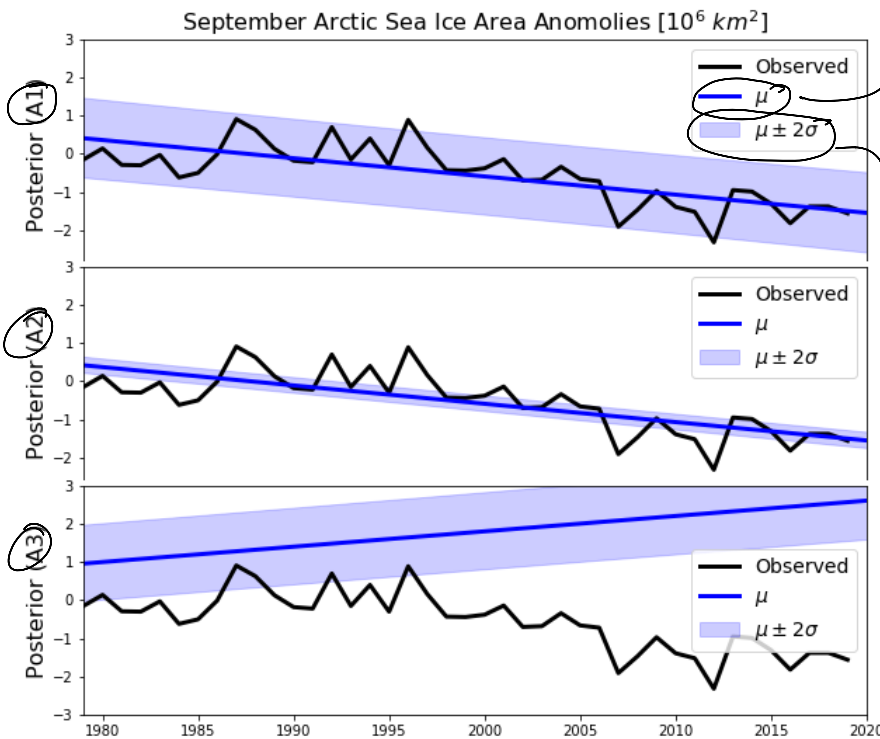
↳ Is the posterior predictive "consistent" w/ the data?

Breakout Exercise

Below are three different posterior predictive plots.

1. Which plot is the most "consistent" with the data?
2. What might be causing the behavior in the other plots?

A1 = Assumptions 1

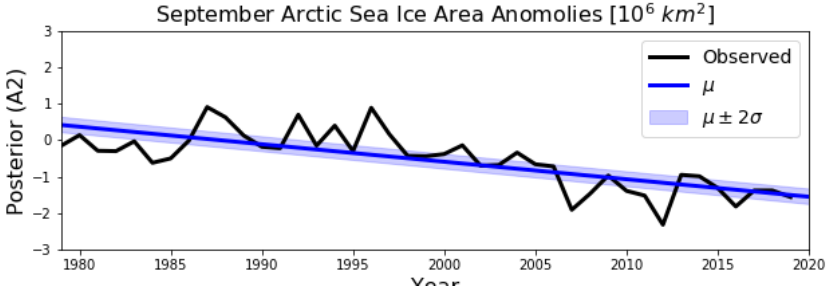


$= \sqrt{\mu_{post}}$

$= \mathbb{E}[y_{pred}]$

$\sigma^2 = Var[y_{pred}]$

Breakout Solution



Notes

- Slope seems to agree w/ data
- Variance of posterior is too small

$$\Sigma_0 - K \Sigma_0 V^T \; ??$$

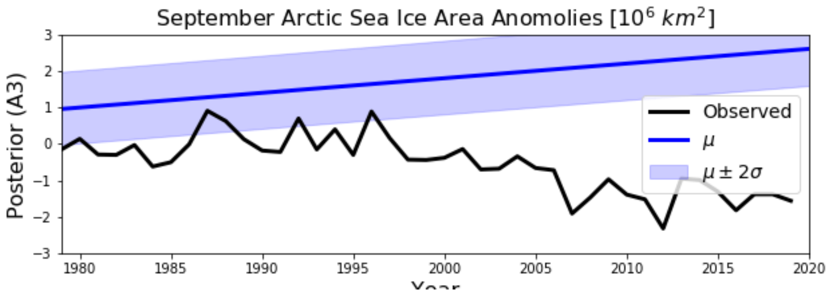
prior variance too small

Σ_ε too small

"good" result, $\sigma_\varepsilon^2 \approx (0.5)^2$

this result, $\sigma_\varepsilon^2 \approx (0.1)^2$

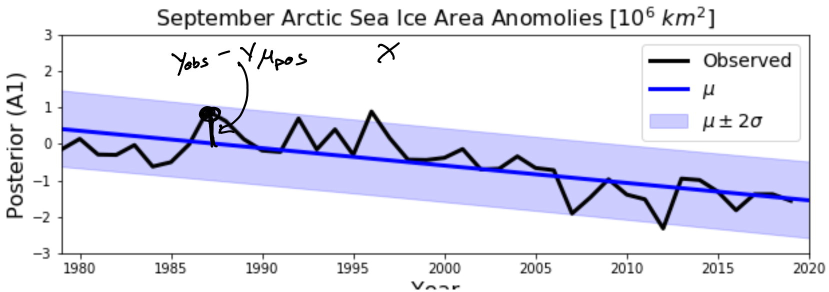
Breakout Solution



Notes

- posterior predictive variance seems reasonable
- slope is off...
  - prior mean on slope is probably positive
  - also prior variance on slope is too small

Breakout Solution



Consistent

↳ posterior predictive "cover" the observations

Posterior Predictive Variance

$$y_{pred} = V x_{post} + \varepsilon$$

error term

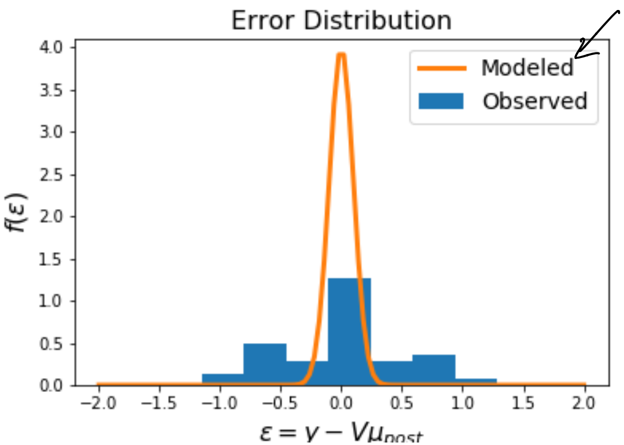
$$\mathrm{Cov}[y_{pred}] = \underbrace{V \Sigma_{post} V^T}_{\text{parameter variability}} + \underbrace{\Sigma_\varepsilon}_{\text{error term}}$$

parameter variability << error term

$$f(y_{pred}) \approx N(V\mu_{post}, \Sigma_\varepsilon)$$

$$\underbrace{y_{obs} - V\mu_{post}}_{\text{residual}} \approx \underbrace{\varepsilon}_{\text{model of error}} \qquad \varepsilon \sim N(0, \sigma_\varepsilon^2)$$

Another Look – A2 (Residual Plot)

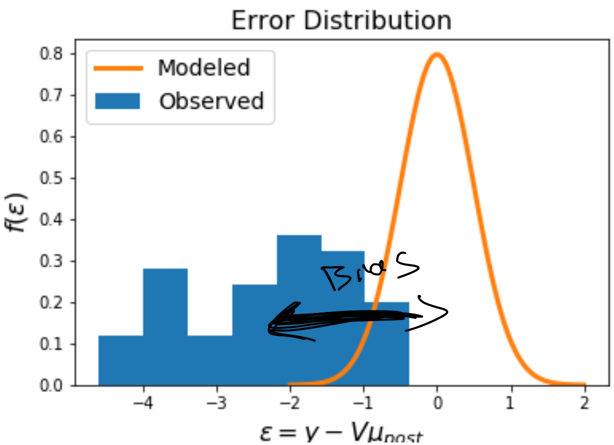


$N(0, \sigma_e^2)$

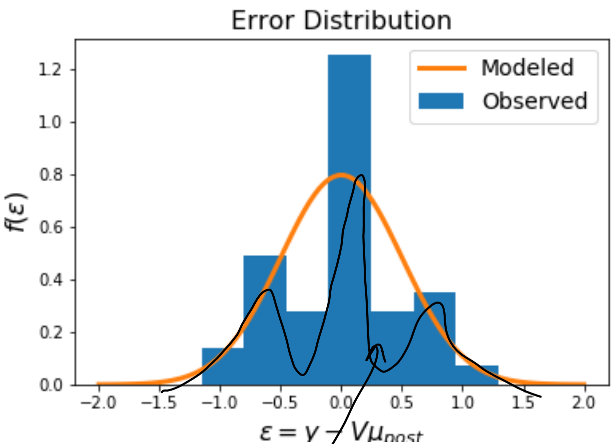
$(y_{obs,i} - V\mu_{pred})$ vs.

Another Look – A3

(Prior on slope was pos)



Another Look – A1



might indicate that a more complex distribution for ε could be necessary

Quantitative Posterior Predictive Checks

Compare posterior predictive distribution to y_{obs} via a test statistic.

$$T(y_{pred}) := \text{test statistic}$$

$$p = \mathbb{P}\text{rob}\underbrace{\left[T(y_{pred}) \geq T(y_{obs}) \right]}_{\text{Probability that } y_{pred} \text{ is "more extreme" than } y_{obs}}$$

One possible test stat:

Black "y_{obs}" line

Blue "mean line" from regression plots

$$T(y_{pred}) = \chi^2 \text{ discrepancy} = \sum_{i=1}^{N} \frac{\left(y_{obs,i} - \mathbb{E}[y_{pred,i}] \right)^2}{\text{Var}[y_{pred,i}]}$$

Hierarchical Inference

Idea: Make model more flexible by including prior and likelihood hyperparameters as additional inference targets. Write Bayes' rule over parameter <u>and</u> "hyperparameters."

If you don't really know σ_ε^2, make it RV and try to infer it's value from the data.

$$f(x, \sigma_\varepsilon^2 \mid y_{obs}) \propto f(y_{obs} \mid x, \sigma_\varepsilon^2)\, f(x, \sigma_\varepsilon^2)$$

Assume independent prior

$$\vdots$$

$$f(y_{obs} \mid x, \sigma_\varepsilon^2)\, f(x)\, f(\sigma_\varepsilon^2)$$

$$\Downarrow$$

USE MCMC to explore $f(x, \sigma_\varepsilon^2)$