Discussion on Choosing Priors

March 5, 2019

Priori, Priori, Give Me the Formioli

- Assume we have i.i.d. samples $\mathbf{x} = [x[0], x[1], \cdot, x[N-1]]$ and have calculated our likelihood function $p(\mathbf{x}|\theta)$.
- We want to use some Bayesian methods, so we need to pick a prior distribution on the parameters θ.
- For example, we are predicting average temperature based on past weather station measurements using the posterior predictive distribution.

Case 1: No Prior Information

- When we have no information about our parameter, we typically use Non-Informative Priors.
- The easiest example is the Improper Uniform Prior $\pi(\theta) = 1 \; \forall\theta$.
  - This means that the posterior is just the likelihood, and we can employ all the frequentist tools we've learned in class.
- Another example is Jeffrey's Prior, $\pi(\theta) = \sqrt{\det(\mathbf{I}(\theta))}$
  - From the definition of Fisher Information, we are essentially using the likelihood function twice.
- In the temperature example with an uninformative prior our predictions will follow the trend of the previous days, e.g. a moving average.

Non-Informative Prior

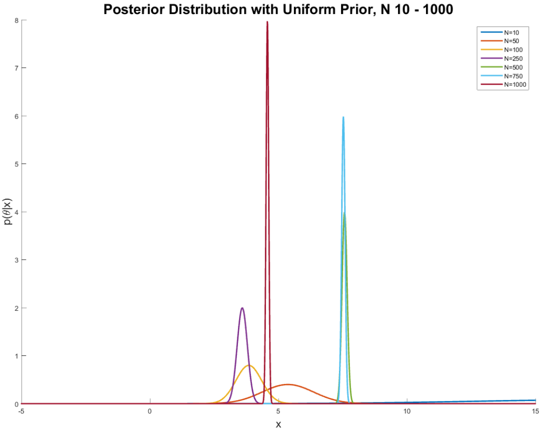

Figure 1: Posterior Distributions for different values of N, attempting to estimate the mean of a normal distribution with an uninformative prior

Case 2: Wealky Informative Priors

- Using the uninformative prior, and given some wild temperature swings, we could easily predict something like -20° or 140°.
- We would like to constrain the range of temperatures into the physically plausible range.
- Proper Uniform Prior: $\pi(\theta) = \mathbb{I}_{[\text{low},\text{high}]}(\theta)$ constrains θ to the record low and high temperatures.
- High variance Prior: Consider $\pi(\theta) \sim N(\mu_\theta, 50)$, where μ_θ is the yearly average temperature.
- Having high variance makes an almost uniform prior but most temperatures will be in a physically plausible range.

Weakly Informative Prior

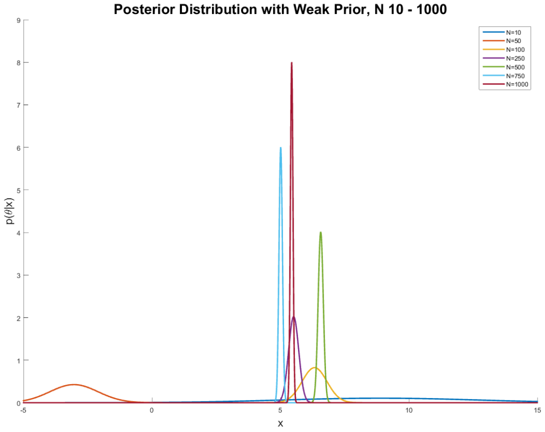

Figure 2: Posterior Distributions for different values of N, attempting to estimate the mean of a normal distribution with a weakly informative prior

Case 3: Strongly Informative Priors

- In some cases, for example calibrating a gps or sensor, we have very good information about the parameters.
- The most common Strongly Informative Prior is $\pi(\theta) \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$, where:
  - μ_θ is the "value given on the box"
  - σ_θ^2 is a very small value.
- What is the most informative prior

Strong Prior

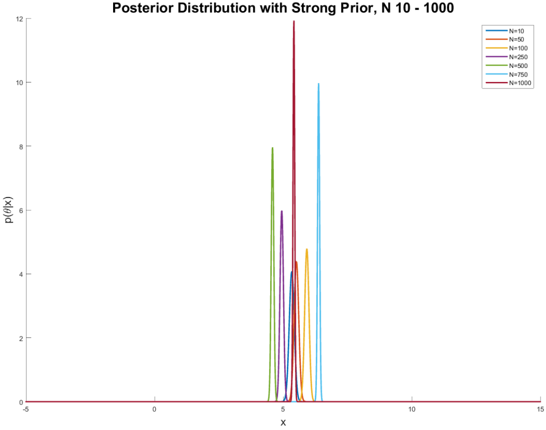

Figure 3: Posterior Distributions for different values of N, attempting to estimate the mean of a normal distribution with a strong prior

Prior vs. Likelihood: Who Wins?

- Consider the exponential family case where

$$p(\mathbf{x}|\theta) = h(\mathbf{x}) \exp(\eta(\theta)^T \mathbf{T}(\mathbf{x}) - NB(\theta))$$

with a conjugate prior:

$$\pi(\theta) = h(\nu) \exp(\eta(\theta)\nu - MB(\theta))$$

- Then the posterior will be proportional to:

$$\exp(\eta(\theta)(T(\mathbf{x}) + \nu) - (N + M)B(\theta))$$

- If N is larger than M, then $T(\mathbf{x} >> \nu)$ most of the time and the posterior can be approximated by the likelihood.