Bayesian Kriging - Part I

Applied Spatial Statistics

Bayesian statistics

- Most introductory statistics courses use classical/frequentist methods
- The main ideas are sampling distributions, confidence intervals, p-values, etc.
- Bayesian statistics is a different way to do statistics
- These new ideas can apply to any statistical analysis, including spatial analyses
- In this lecture we'll introduce very basic Bayesian concepts and apply them to a geostatistical analysis

Advantages of Bayesian Kriging

- A key advantage of Bayesian method for spatial statistics is in uncertainty quantification
- For example, so far we computed standard errors for $\hat{\beta}$ and predictions $\hat{Y}_0$ using a "plug-in" estimator of the spatial covariance parameters, $\hat{\theta}$
- Bayesian methods allow us to account for uncertainty in $\hat{\theta}$
- Bayesian computational methods are also useful for advanced models such as non-Gaussian data
- We can also naturally bring in prior knowledge, which is helpful for parameters that are hard to estimate like the spatial range

Disadvantages of Bayesian Kriging

- Bayesian methods are generally slower than MLE
- You have to specify prior distributions (we'll discuss) which is somewhat subjective
- In some fields, Bayesian methods are less popular than MLE

The Bayesian approach

- Simple example: Let $Y \in \{0, 1, ..., n\}$ be the number of successes in n independent trials
- For example, n patients are given a vaccine and Y achieve immunity
- The probability that a given patient achieves immunity is $\theta \in [0, 1]$
- This implies the model $Y|\theta \sim \text{Binomial}(n, \theta)$
- Our goal is to estimate the parameter θ

The Bayesian approach

- As in classical statistics, Bayesians view the parameter θ as fixed and unknown
- However, we express our uncertainty about it using probability distributions
- The distribution before observing the data is the **prior distribution**
- Example: Prob($\theta > 0.5$) $= 0.6$.
- Probability statements like this are intuitive (to me at least)
- This is subjective in that people may have different priors

The Bayesian approach

- Our uncertainty about θ is changed (hopefully reduced) after observing the data
- The **Likelihood function** is the distribution of the observed data given the parameters
- This is the same likelihood function used in MLE
- Therefore, when the prior information is weak, Bayesian and maximum likelihood estimates are similar
- Even in this case, the interpretations are different

The Bayesian approach

- The uncertainty distribution of θ after observing the data is the **posterior distribution**
- **Bayes theorem** provides the rule for updating the prior

$$p(\theta|Y) = \frac{f(Y|\theta)\pi(\theta)}{m(Y)}$$

- In words: Posterior $\propto$ Likelihood$\cdot$prior
- A key difference between Bayesian and frequentist statistics is that all inference is conditional on the single data set we observed Y

Back to the example

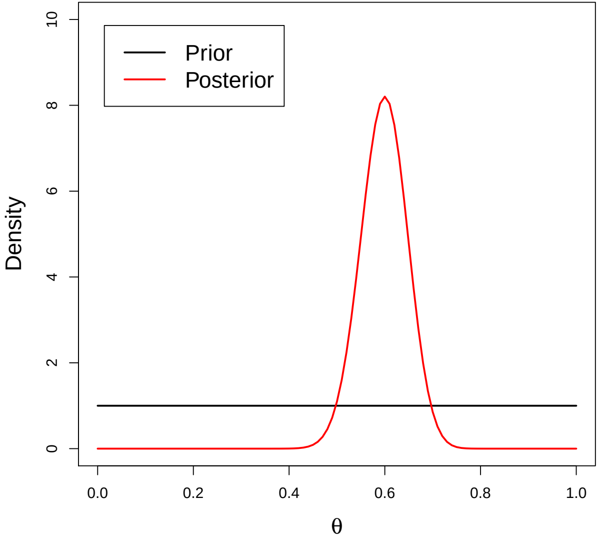
- Say we observed $Y = 60$ successes in $n = 100$ trials
- The parameter $\theta \in [0, 1]$ is the true probability of success
- In most cases we would select a prior that puts probability on all values between 0 and 1
- If we have no relevant prior information we might use the prior
$$\theta \sim \text{Uniform}(0, 1)$$
so that all values between 0 and 1 are equally likely
- This is an example of an **uninformative prior**

Posterior distribution

- The likelihood is $Y|\theta \sim \text{Binomial}(n, \theta)$
- The uniform prior is $\theta \sim \text{Uniform}(0, 1)$
- Then it turns out the posterior is

$$\theta|Y \sim \text{Beta}(Y + 1, n - Y + 1)$$

Bayesian learning: $Y = 60$ and $n = 100$

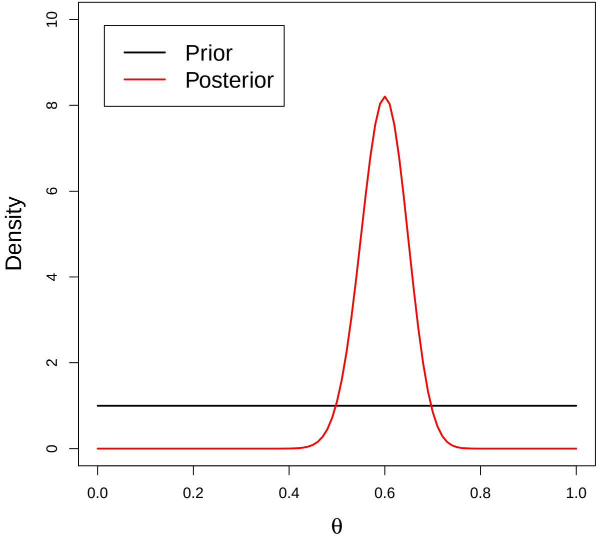


Beta prior

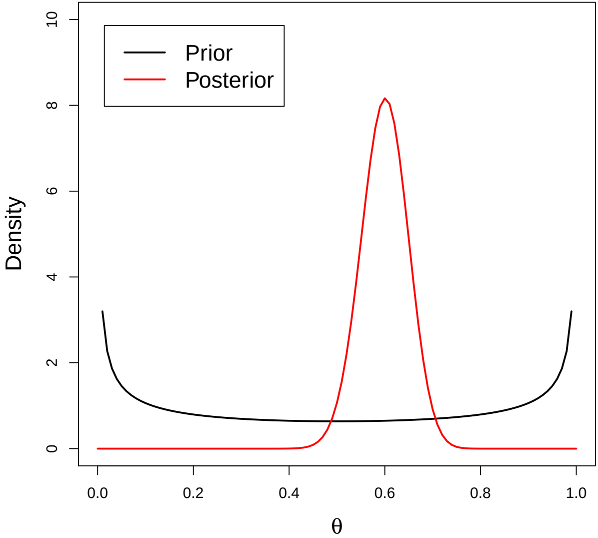
- The uniform prior represents prior ignorance
- To encode prior information we need a more general prior
- The beta distribution is a common prior for a parameter that is bounded between 0 and 1
- If $\theta \sim \text{Beta}(a, b)$ then the posterior is

$$\theta | Y \sim \text{Beta}(Y + a, n - Y + b)$$

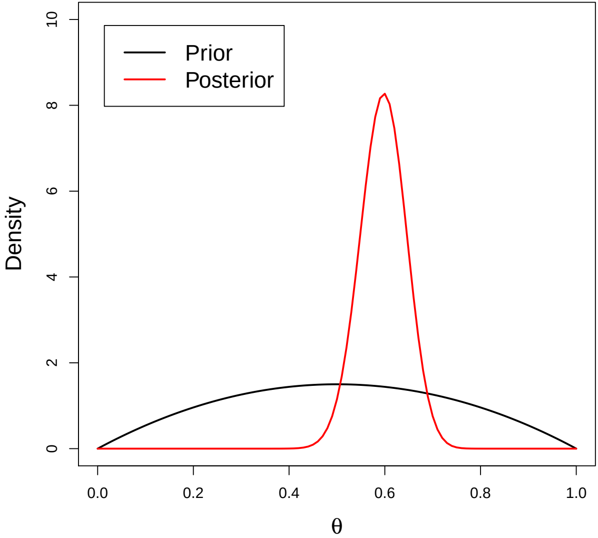
Prior 1: $\theta \sim \text{Beta}(1, 1)$



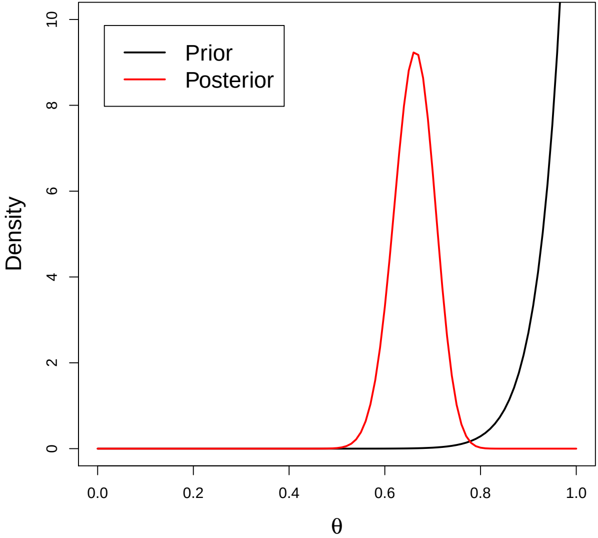
Prior 2: $\theta \sim \text{Beta}(0.5, 0.5)$



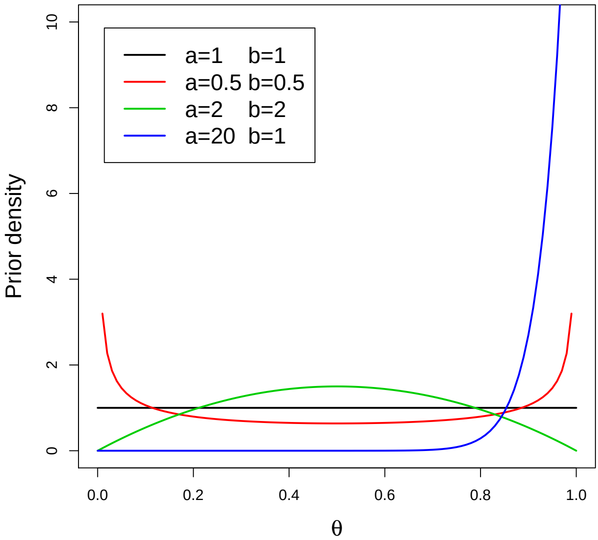
Prior 3: $\theta \sim \text{Beta}(2, 2)$



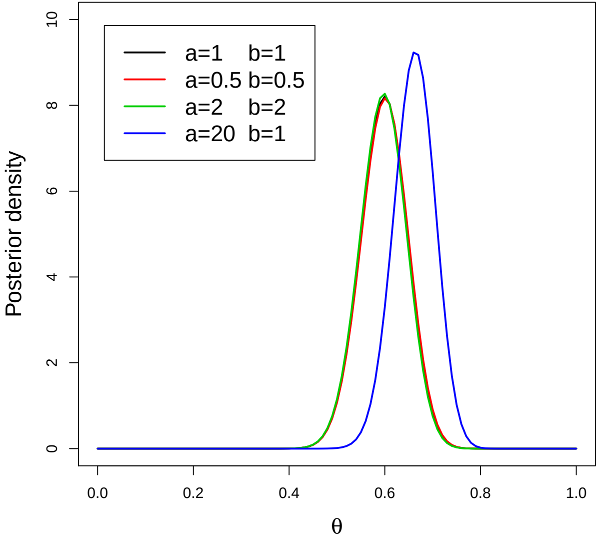
Prior 4: $\theta \sim \text{Beta}(20, 1)$



Plot of different beta priors



Plots of the corresponding posteriors



Summarizing the posterior distribution

- If there is only one parameter we can simply plot the posterior distribution
- However, if there are many parameters the posterior is hard to plot
- Instead we can extract means, standard deviations and 95% intervals to summarize the posterior
- These results can be put in a table and interpreted similar to classical statistics
- For example, if a regression coefficient's 95% posterior interval excludes zero we can say its effect is significant

Sensitivity to the prior

		Prior			Posterior		
a	b	Mean	SD	P>0.5	Mean	SD	P>0.5
1	1	0.50	0.29	0.50	0.60	0.05	0.98
0.5	0.5	0.50	0.50	0.50	0.60	0.05	0.98
2	2	0.50	0.22	0.50	0.60	0.05	0.98
20	1	0.95	0.05	1.00	0.66	0.04	1.00