

Comparing Non-informative Priors for Estimation and Prediction in Spatial Models

Vigre Semester Report

by: Regina Wu

Advisor: Cari Kaufman

January 31, 2010

1 Introduction

Gaussian random fields with specified mean and covariance functions are commonly used to model spatial data, but it is often a challenge to decide which non-informative prior to choose for the unknown mean and covariance parameters. In Berger, de Oliveira, and Sansó (2001), the Jeffreys-rule prior, independence Jeffreys prior, and reference prior are presented, with the reference prior recommended for its proper posterior and good frequentist coverage probabilities in parameter estimation. We examine the behaviors and characteristics of these prior distributions and work toward proposing a simpler yet comparable non-informative prior.

2 Behavior of Non-informative Priors

We first examine priors having an exponential covariance function and a null X-matrix, which conveniently simplifies the reference prior, independence Jeffreys prior, and Jeffreys-rule prior to be mathematically identical. We examine the prior's behavior in response to increasing the number of equally-spaced samples in both a fixed and increasing domain. For easier comparison, we plot the ratio of the prior to the prior's value at $\theta = 1$ rather than the unnormalized prior itself. From Figure 1, we see that increasing the sampling size has no affect on the shape of the prior ratios under an increasing domain, but under a fixed domain it causes more weight to be shifted toward smaller theta values. The idea that the prior may be dependent upon sampling locations is further examined in Figure 2 which compares the shape of the prior ratios under four types of sampling locations bounded in $[0,1]$. Results show that sets having locations closer to one another, such as the one chosen to be clustered in the center of $[0,1]$ tend to place more weight on smaller theta values.

An important characteristic of the prior presented in Berger et al., 2001 is that the prior approaches $\frac{1}{\theta}$ as theta gets larger. Figure 3 shows the convergence in both the fixed and increasing domains. This motivates finding a simple approximation to the computationally intensive priors currently proposed. However, a major barrier to using $\frac{1}{\theta}$ as an approximation is how to deal with values near $\theta = 0$ as $\frac{1}{\theta}$ shoots upward towards infinity.

3 Simulation Study

To compare the performance of the reference prior and Jeffreys-rule prior, we replicated and extended the simulation study in Berger et al., 2001 which compared the frequentist coverage of equal-tailed Bayesian credible intervals for the range parameter θ and their expected log lengths. Both priors produced intervals with reasonable coverage when only an intercept

Figure 1: Behavior of Priors with Increasing Sampling Size n under Fixed Domain (left) and Increasing Domain (right) (Bold Line: $n = 100$, Dashed Line: $n = 200$, Dotted Line: $n = 300$)

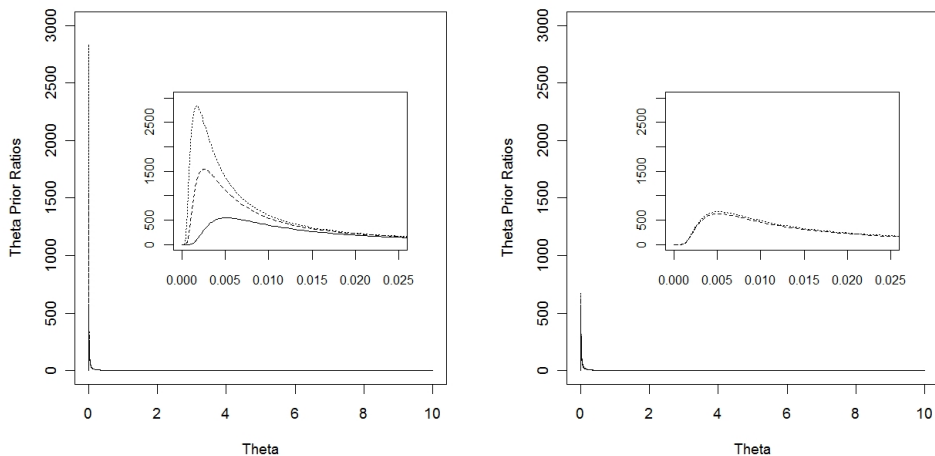


Figure 2: Effect of Varying Sampling Locations on Prior (Dashed Line= systematically spaced locations, Solid Line= random locations, Dotted Line= locations taken near the endpoints of $[0,1]$, Dotted-Dashed Line= locations clustered in the center of $[0,1]$)

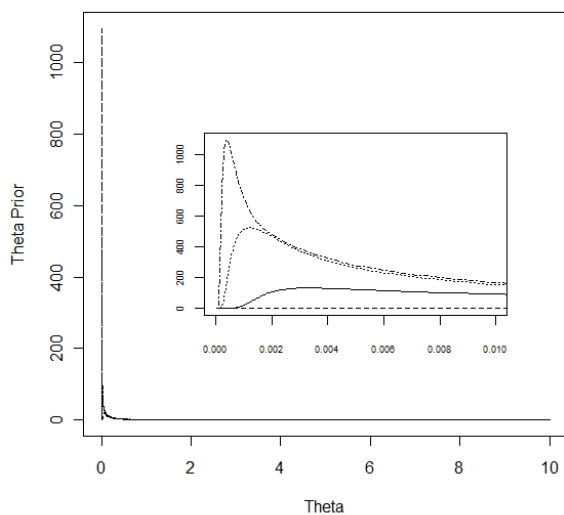
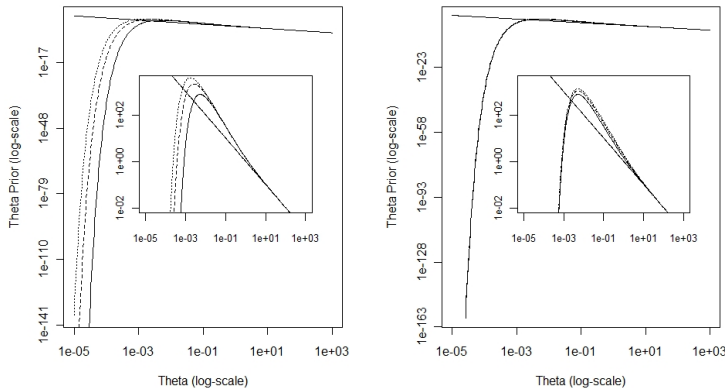


Figure 3: *Convergence of Priors to $\frac{1}{\theta}$ on Fixed Domain (left) and Increasing Domain (right)*



term was included in the mean of the Gaussian process, but the coverage of the Jeffreys-rule prior was very poor when additional terms were added, particularly when the spatial correlation was also strong. The authors attributed this poor performance to the change in the marginal posterior of θ as a result of the increase in degrees of freedom in the posterior for σ . We extend this simulation study by comparing mean squared errors in addition to coverage rates, and posterior predictions of the Gaussian process in addition to inference about θ .

As in Berger et al., 2001, our simulated spatial data are modeled under 6 different isotropic Gaussian random fields $Z(\cdot)$. Each model is a combination of a mean function $E\{Z(s)\}$ and covariance function $C(d)$, where $E\{Z(s)\}$ is either the constant value 0.15 ($p=1$) or the function $0.15 - .65x - .1y + 9x^2 - xy + 1.2y^2$ ($p=6$), and the covariance function $C(d) = .12exp\{-d/\theta\}$ is exponential with θ being either .2, .5, or 1.0.

From each of the six Gaussian fields, we simulate 3000 sets of samples at equally-spaced locations in the union of a training set $S_t = \{(x, y) | x, y \in \{0, 0.25, 0.5, 0.75, 1\}\}$ and a validation set $S_v = \{(x, y) | x, y \in \{.125, 0.375, 0.625, 0.875\}\}$.

For each of the 3000 sample replications z_j in our training set, we generate 1000 random

walk Metropolis Hastings θ samples under both the reference prior and Jeffreys-rule prior. Candidates are sampled from a normal distribution, and the first 200 values from each set of posterior samples are removed for burn-in.

To see how well the range parameter θ is estimated under the two priors, we compute the mean squared error (MSE) between the posterior mean $\bar{\theta}_j$ of each sample and the true θ value of either .2, .5, or 1.0. We also calculate the frequentist coverage probabilities of the true parameter θ under both the reference prior and Jeffreys-rule prior by computing equal tailed 95% credible intervals ($L(\theta_j)$, $U(\theta_j)$), from which we calculate the coverage probability $\hat{p}_\theta = \frac{1}{3000} \sum_{i=1}^{3000} I_{L(\theta_j) < \theta < U(\theta_j)}$ with the associated standard error estimate of $\sqrt{\frac{\hat{p}_\theta(1-\hat{p}_\theta)}{3000}}$. The MSE's are shown in Table 1 and the frequentist coverage probabilities in Table 2.

Table 1: *MSE (and Standard Errors) of θ Samples*

θ	.2	.5	1
<i>p</i> = 1			
Reference Prior	.8498 (.2229)	3.5694 (.7940)	4.6910 (.9063)
Jeffreys-rule Prior	1.0478 (.3327)	4.5504 (.9807)	6.0384 (1.0631)
<i>p</i> = 6			
Reference Prior	.1606 (.1299)	.0629 (.0066)	.4973 (.0244)*
Jeffreys-rule Prior	.1486 (.1362)	.2783 (.1323)	.8652 (.1255)

(*) indicates significantly different values where there is no overlap between values ± 2 SD's

Table 2: *Frequentist Coverage Probabilities (and Standard Errors) of Bayesian Equal-Tailed 95% Credible Intervals for θ*

θ	.2	.5	1
<i>p</i> = 1			
Reference Prior	.9763 (.0028)	.9693 (.0031)*	.9430 (.0042)*
Jeffreys-rule Prior	.9810 (.0025)	.9230 (.0049)	.8640 (.0063)
<i>p</i> = 6			
Reference Prior	1.000 (.0000)*	.9707 (.0031)*	.6307 (.0088)*
Jeffreys-rule Prior	.8317 (.0068)	.2160 (.0075)	.0663 (.0045)

The MSE values in Table 1 show that the accuracy of the posterior mean in estimating the range parameter θ tends to be slightly better for the reference prior than for the Jeffreys-

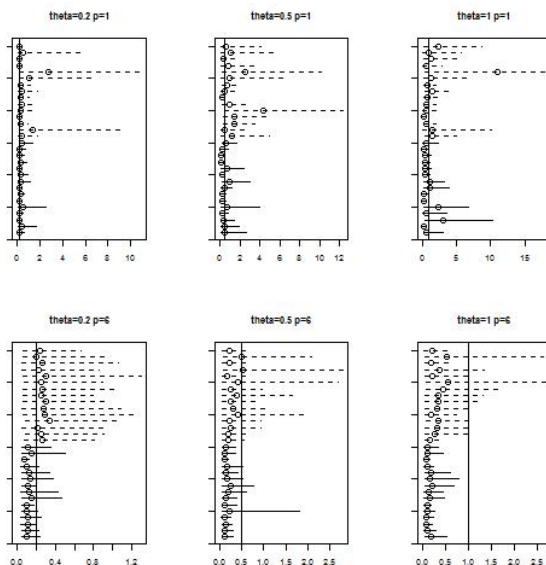
rule prior, but this difference is not significant except in the case of $p=6$ and $\theta = 1$. Although there is little significant difference in MSE values, Table 2 shows that the frequentist coverage probabilities of the 95% credible intervals are significantly different between the two priors in all cases except when $p=1$ and $\theta = .2$. The credible intervals under the reference prior consistently have closer to nominal coverage of the true θ compared to the intervals under the Jeffreys-rule prior which perform particularly poorly when $p = 6$. These results agree with those presented in Berger et al., 2001, although the coverage probabilities under the reference prior tended to be closer to the nominal .95 value, except when $p = 6$ and $\theta = 1$. However, it is expected that both priors perform poorly at $p = 6$ and $\theta = 1$, as the strong spatial correlation reduces the effective sample sizes.

To better understand the cause of the difference between the two tables, we examine the lengths of the credible intervals, computed as $\frac{1}{3000} \sum_{j=1}^{3000} (U(\theta_j) - L(\theta_j))$. Under all 12 cases, the credible intervals under the reference prior have significantly longer lengths compared to the credible intervals under the Jeffreys-rule prior.

We illustrate this in forest plots for a random subset of 15 of the 3000 iterations, shown in Figure 3. We plot both the posterior means and the endpoints of the 95% credible intervals. From the plot we can clearly see that the reference prior produces longer intervals, especially when $p=6$. Under both priors, the posterior means of each $\{\theta_j\}$ land about equally far from the true value of θ , but because the credible intervals under the reference prior are longer in length, they are more likely to capture the true θ . This explains why in Table 1 the MSE's are not significantly different between the priors in almost every case, yet the frequentist coverage probabilities are significantly higher for the reference prior.

In addition to examining the effect of prior choice on parameter estimation, it is also important to investigate its effect on prediction. For this portion of the simulation we revisit our validation set, S_v . For each sampled θ , we make predictions at each of the 16

Figure 4: Forest Plots of Posterior θ (Dashed Line=Reference Prior and Solid Line= Jeffreys-Rule Prior)



locations by sampling from a multivariate t-distribution found in Handcock and Stein, 1993. Combining these predictions, we have a set $\{z_{jk}\}$ for each location k , from which we are able to calculate a posterior mean \bar{z}_{jk} . We compute the mean squared prediction error (MSPE) by averaging over both locations and replications, shown in Table 3. As before, the frequentist coverage probabilities of Bayesian equal-tailed 95% credible intervals are also calculated for the samples taken at the 16 locations. We calculated the frequentist coverage probability by averaging indicators of $L(\{z_{jk}\}) < z'_{jk} < U(\{z_{jk}\})$ over both locations and replications. The results are shown in Table 4.

The spatial prediction portion of this simulation produced flipped results from those of parameter estimation. As before, the MSPE under the reference prior is slightly better, although in general the difference is not significant as shown in Table 3. However, Table 4 shows that the frequentist coverage of predictions under the Jeffreys-rule prior are all significantly higher than those of the reference prior except when $p = 1$ and $\theta = 1$. We note that

Table 3: *MSPE (and Standard Errors) of Predictions*

θ	.2	.5	1
$p = 1$			
Reference Prior	.0747 (.0005)	.0330 (.0002)	.0167 (.0001)
Jeffreys-rule Prior	.0750 (.0005)	.0332 (.0002)	.0168 (.0001)
$p = 6$			
Reference Prior	.0742 (.0005)	.0332 (.0002)*	.0169 (.0001)*
Jeffreys-rule Prior	.0758 (.0005)	.0346 (.0002)	.0178 (.0001)

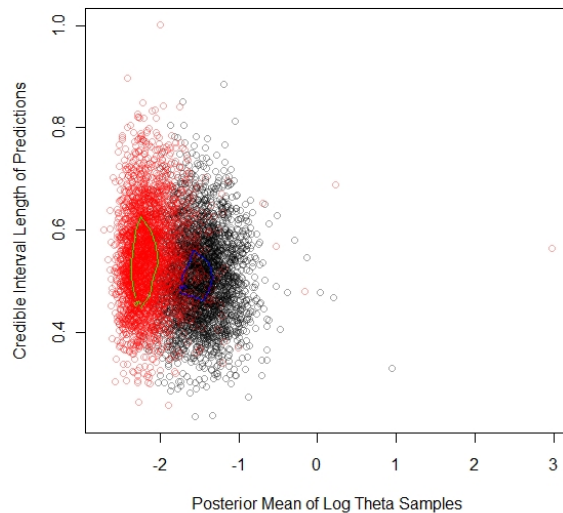
Table 4: *Frequentist Coverage Probabilities (and Standard Errors) of Bayesian Equal-Tailed 95% Credible Intervals for Predictions*

θ	.2	.5	1
$p = 1$			
Reference Prior	.8037 (.0018)*	.9525 (.0010)*	.9568 (.0010)
Jeffreys-rule Prior	.9444 (.0010)	.9571 (.0009)	.9602 (.0009)
$p = 6$			
Reference Prior	.9191 (.0012)*	.9329 (.0011)*	.9368 (.0011)*
Jeffreys-rule Prior	.9313 (.0012)	.9433 (.0011)	.9473 (.0010)

these differences are very minor, and may again be attributed to the length of the credible intervals which are significantly longer under the Jeffreys-rule prior in all cases. From these results, it seems the reasons for preferring the reference prior to the Jeffreys-rule prior are restricted to parameter estimation, and not prediction.

The relationship between parameter estimation and predictive power is compared in Figure 5. The posterior means of the log theta samples generated under the $p = 6$ and $\theta = 1$ case are plotted against the credible interval lengths of the predictions for the first of 16 total locations in the validation set. Though both priors produce biased theta samples, the reference prior is less biased and has log theta samples closer to the accurate value of zero. The slight negative trend reflects how the increase in variation as theta values get smaller results in longer credible interval lengths for predictions. This may explain why even though the Jeffreys-rule prior tends to have more biased parameter estimations, it still has higher coverage probabilities for prediction.

Figure 5: *Plot of Posterior Mean of Log Theta against Credible Interval Length for Predictions (Black=Reference prior, Red=Jeffreys-rule Prior)*



4 Further Research

Thus far, the focus has been placed on priors which assume the exponential covariance function, a special case of the Matérn covariance function when $\nu = 1$. A next step for the simulation study is to examine how the choice of the smoothness parameter ν can affect parameter estimation and predictive power of the priors. Specifically, we will vary ν in the Matérn covariance function over additional Metropolis Hastings samplings. The first will over-estimate the actual $\nu = 1$ value and the second will assume a ν value found through a Gaussian quadrature approximation. If the latter method can produce comparable results to those produced under the correct $\nu = 1$ value, it will present a systematic and useful method of choosing the smoothness parameter. Finally, we wish to further examine the behavior of the priors under the more general Matérn covariance function in hopes of finding a computationally simpler prior with a proper posterior that can produce comparable results

to the currently proposed reference prior.

References

James O. Berger, Victor De Oliveira, and Bruno Sansó. Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 96(456):1361–1374, 2001.

Mark S. Handcock and Micheal L. Stein. A Bayesian analysis of kriging. *Technometrics*, 35(4):403–410, 1993.