Lecture 15: Spatial Prediction and Model Selection

Conditional Multivariate Normal Theory

The conditional distribution, $p(\boldsymbol{Y_1}|\boldsymbol{Y_2}, \boldsymbol{\beta}, \sigma^2, \phi, \tau^2)$ is normal with:

- $E[\boldsymbol{Y_1}|\boldsymbol{Y_2}] = \boldsymbol{\mu_1} + \Omega_{12}\Omega_{22}^{-1}(\boldsymbol{Y_2} - \mu_2)$

- $Var[\boldsymbol{Y_1}|\boldsymbol{Y_2}] = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$

Posterior Predictive Distribution

The (posterior) predictive distribution $p(Y(\boldsymbol{s_0})|y)$ can be written as

$$p(Y(\boldsymbol{s_0})|y) = \int p(Y(\boldsymbol{s_0})|y, \boldsymbol{\theta})p(\boldsymbol{\theta}|y)d\boldsymbol{\theta}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma^2, \phi, \tau^2\}$.

The posterior predictive distribution gives a probabilistic forecast for the outcome of interest that does not depend on any unknown parameters.

These are the individual samples for a specific point that we have seen in STAN.

Prediction

We will use cross-validation or a test/training approach to compare predictive models.

Consider three data structures: continuous, count, and binary; how should we evaluate predictions in these situations?

Loss Functions

Loss functions penalize predictions that deviate from the true values.

For continuous or count data, squared error loss and absolute error loss are common.

With binary data, a zero-one loss is frequently used.

However, these metrics are all focused on point estimates.

If we think about outcomes distributionally, empirical coverage probability can be considered. For instance, our 95 % prediction intervals should, on average, have roughly 95 % coverage.

With interval predictions, the goal is to have a concentrated predictive distribution around the outcome.

The Continuous Rank Probability Score (CRPS) defined as

$$CRPS(F, y) = \int_{-\infty}^{\infty} \left(F(u) - 1(u \geq y)\right)^2 du,$$

where F is the CDF of the predictive distribution, is a metric that measures distance between an observed value and a distribution.

```r
crps_sample(y = 0, dat = 0)
```

```
## [1] 0
```

```r
crps_sample(y = 0, dat = 2)
```

```
## [1] 2
```

```r
crps_sample(y = 0, dat = c(2,-2))
```

```
## [1] 1
```

```r
crps_sample(y = 0, dat = c(2,0,-2))
```

```
## [1] 0.4444444
```

```r
crps_sample(y = 0, dat = c(2,0,0,0,0,-2))
```

```
## [1] 0.1111111
```

CRPS

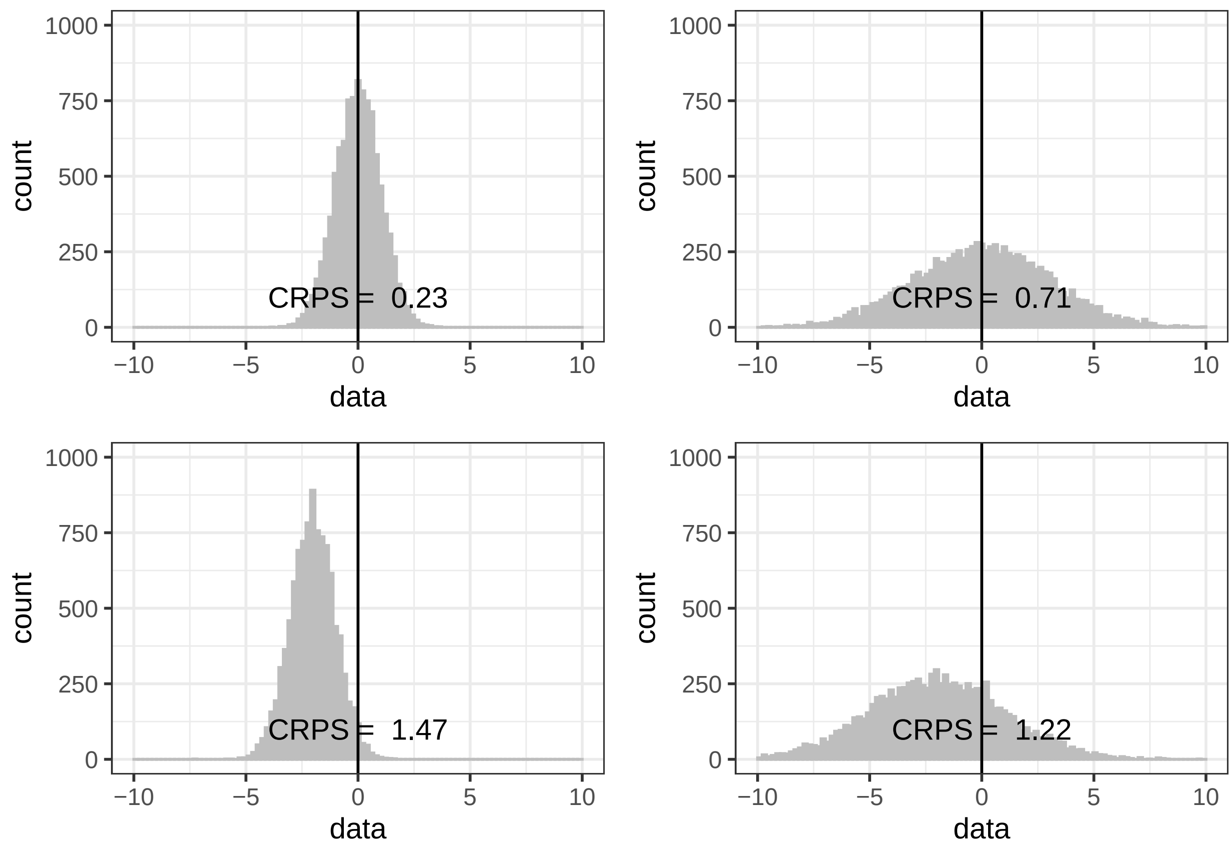
Consider four situations and sketch the predictive distribution and the resultant CRPS for each scenario. How does the MSE function in each setting?

1. Narrow predictive interval centered around outcome.

2. Wide predictive interval centered around outcome.

3. Narrow predictive interval with outcome in tail.

4. Wide predictive interval with outcome in tail.


1000
750
500
250
0
count
−10
−5
0
5
10
data
CRPS = 0.23
CRPS = 0.71
CRPS = 1.47
CRPS = 1.22