

STA 360/601 Lab 7

Michael Jauch

March 2016

Introduction

Suppose we have a collection of random variables $\{X_n\}_{n=1}^N$, each representing an observation of some process at stage $n \in \{1, \dots, N\}$. In our case, N will be the number of days in July and X_n will be the high temperature on the n^{th} day of July.

The expected high temperature and variation in high temperature doesn't change much in the month of July, so we might assume these quantities are the same for each day. We could model the temperatures as $X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$ – or in multivariate normal notation, $X \sim N(\mu \mathbf{1}_N, \sigma^2 \mathcal{I}_N)$, where $X = (X_1, \dots, X_N)^T$, $\mathbf{1}_N$ is an $N \times 1$ vector of ones, and \mathcal{I}_N is the $N \times N$ identity matrix.

However, we know that temperatures on subsequent days tend to be similar, so the independence assumption is not reasonable. Also, if we make the independence assumption, we can't use our model to predict temperatures in the near future, which we might like to do.

A simple model that allows for correlation between observations is the autoregressive process of order 1, also known as an AR(1) process. In this case, $X \sim N(\mu \mathbf{1}_N, \sigma^2 \Phi_N)$, where Φ_N is a correlation matrix defined as

$$\Phi_N = \begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^{N-1} \\ \phi & 1 & \phi & \dots & \phi^{N-2} \\ \phi^2 & \phi & 1 & \dots & \phi^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{N-1} & \phi^{N-2} & \phi^{N-3} & \dots & 1 \end{bmatrix}.$$

The parameters σ^2 and $\phi \in [-1, 1]$ control the marginal variance of each X_n and the correlation between the observations, respectively. We will assume that $\phi \in [0, 1]$, so that there is positive correlation between high temperatures on subsequent days rather than negative correlation.

Task 1

Plot the sequence of high temperatures in the file `temp_data.csv`. This file, downloaded from the Weather Underground website, contains high tempera-

tures recorded at Raleigh-Durham airport in July of 2015. (In what follows, we will ignore the fact that the data only take integer values).

Simulate from a multivariate normal $X \sim N(\mu \mathbf{1}_N, \sigma^2 \mathcal{I}_N)$ with $\mu = 89$ and $\sigma^2 = 12$ using the `mvrnorm` function in the `MASS` library. Plot X similarly to how you plotted the real data.

Compare the appearance of the plot of the true data and the plot of the simulated data.

Task 2

Fix $\mu = 87$ and $\sigma^2 = 12$ and sample/plot $X \sim N(\mu \mathbf{1}_N, \sigma^2 \Phi_N)$ for various values of $\phi \in [0, 1]$. How do the samples change? What choice of ϕ results in samples that look the most like the real data?

To do this, you may use the following function, which, given N and ϕ , constructs the correlation matrix Φ_N :

```
AR1cor <- function(phi, N){
  phi_mat <- mat.or.vec(N,N)
  for(i in 2:N){
    for(j in 1:(i-1))
      phi_mat[i,j] <- phi^(i-j)
  }
  phi_mat <- phi_mat + t(phi_mat) + diag(N)
  return(phi_mat)
}
```

Task 3

We want to do inference for the parameters μ, σ^2 and ϕ , so we use the following prior distributions:

$$\begin{aligned}\mu &\sim N(\theta, \tau) \\ \sigma^2 &\sim \text{IG}(a, b) \\ \phi &\sim \text{Unif}(0, 1)\end{aligned}$$

where θ, τ, a , and b are fixed values. Derive full conditionals for μ, σ^2 and describe (mathematically) how you would update ϕ using a Metropolis step with a uniform proposal distribution.

Task 4

Code a sampler. Run it for 10000 iterations. You might try $\theta = 85$, $\tau = 100$, $a = 11$, $b = 100$, and $\epsilon = .15$. Make trace plots to assess convergence. Give 95% credible intervals for μ , σ^2 , and ϕ .

Just for “fun” task 5 (will not be graded)

Think about you would predict high temperatures in the last week of July if you had only observed the high temperatures in the first 24 days. Try it out.