

# Assignment 3 Question 2 MAST90125: Steven Maharaj 695281

**Due: Friday 25 October 2019**

There are places in this assignment where R code will be required. Therefore set the random seed so assignment is reproducible.

```
set.seed(695281) #Please change random seed to your student id number.  
library(mvtnorm)
```

## Question Two: (13 marks)

Researchers are interested in phone call duration. The available data consisted of 100 standardised phone call durations, and which hour ( $t = 1, \dots, 10$ ) the call was initiated. The researchers assumed the following model,

$$\begin{aligned}p(y_i|\mu(t)) &= \mathcal{N}(\mu(t), \sigma^2) \\p(\mu(t)) &= \mathcal{N}(0, k(t, t))\end{aligned}$$

such that the covariance function  $k(x, x')$  is squared exponential,

$$k(x, x') = \sigma_K^2 e^{-l \times (x - x')^2},$$

with  $\sigma_K^2$  fixed to 1.85 and  $l$  fixed to 0.05. The data can be downloaded from LMS as `call.csv`.

- a) The researchers are interested in making predictions of phone call duration,  $\tilde{\mu}(t)$  for hours  $t = 1, \dots, 12$ . As an initial step, they assumed  $\sigma^2$  was 0.95. Assume the data  $\mathbf{y}$  is conditioned on a single realisation of the Gaussian process prior. Do the following:
- Based on the information provided, determine the joint distribution of data  $\mathbf{y}$  and predictions  $\tilde{\mu}(t)$ .
  - Determine the distribution of  $\tilde{\mu}(t)$  conditional on  $\mathbf{y}$ ,  $\sigma^2$ ,  $\sigma_K^2$  and  $l$ . Show working
  - Plot the predictions with 95 % and 99 % credible intervals along with the observed data. Comment, in a Bayesian language, on the behaviour of predictions where no data was observed.

Answer: The joint distribution of the data  $\mathbf{y}$  and predictions is  $\mu(\tilde{\mathbf{x}})$  is For lecture it was shown that

$$p\left(\begin{array}{c} \mathbf{y} \\ \mu(\tilde{\mathbf{x}}) \end{array}\right) = \mathcal{N}\left(\left(\begin{array}{c} \mathbf{m}(\mathbf{x}) \\ \mathbf{m}(\tilde{\mathbf{x}}) \end{array}\right), \left(\begin{array}{cc} \mathbf{k}(\mathbf{x}, \mathbf{x}) + \Sigma & \mathbf{k}(\mathbf{x}, \tilde{\mathbf{x}}) \\ \mathbf{k}(\tilde{\mathbf{x}}, \mathbf{x}) & \mathbf{k}(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}) \end{array}\right)\right)$$

For this question joint distribution of the data  $\mathbf{y}$  and predictions is  $\mu(\tilde{\mathbf{t}})$  is

For lecture it was shown that

$$p\left(\begin{array}{c} \mathbf{y} \\ \mu(\tilde{\mathbf{t}}) \end{array}\right) = \mathcal{N}\left(\left(\begin{array}{c} \mathbf{0} \\ \mathbf{0} \end{array}\right), \left(\begin{array}{cc} \mathbf{k}(\mathbf{t}, \mathbf{t}) + \Sigma & \mathbf{k}(\mathbf{t}, \tilde{\mathbf{t}}) \\ \mathbf{k}(\tilde{\mathbf{t}}, \mathbf{t}) & \mathbf{k}(\tilde{\mathbf{t}}, \tilde{\mathbf{t}}) \end{array}\right)\right)$$

where  $k(t_i, \tilde{t}_j) = \sigma_K^2 e^{-l \times (t_i - \tilde{t}_j)^2}$  and  $\Sigma = \sigma^2 \mathbf{I}$

Now we will find  $\mu(\tilde{\mathbf{t}})$  conditional on  $\mathbf{y}, \sigma^2, \sigma_K^2, l$ .

For lectures, the noiseless case had the follow distribution.

$$\mu(\tilde{\mathbf{t}}) | \cdot \sim \mathcal{N}(\mathbf{m}(\tilde{\mathbf{t}}) + \mathbf{k}(\tilde{\mathbf{t}}, \mathbf{t}) \mathbf{k}(\mathbf{t}, \mathbf{t})^{-1} (\mathbf{y} - \mathbf{m}(\mathbf{t})), \mathbf{k}(\tilde{\mathbf{t}}, \tilde{\mathbf{t}}) - \mathbf{k}(\tilde{\mathbf{t}}, \mathbf{t}) \mathbf{k}(\mathbf{t}, \mathbf{t})^{-1} \mathbf{k}(\mathbf{t}, \tilde{\mathbf{t}}))$$

for this question  $\mathbf{m}(\tilde{\mathbf{t}}) = \mathbf{0}$ , we must add  $\Sigma$  to the first element of covariance block matrix. In other words  $\mathbf{k}(\mathbf{t}, \mathbf{t}) \rightarrow \mathbf{k}(\mathbf{t}, \mathbf{t}) + \Sigma$ . Then we have that following

$$\mu(\tilde{\mathbf{t}}) | \mathbf{y}, \sigma^2, \sigma_K^2, l \sim \mathcal{N} \left( \mathbf{k}(\tilde{\mathbf{t}}, \mathbf{t}) [\mathbf{k}(\mathbf{t}, \mathbf{t}) + \Sigma]^{-1} \mathbf{y}, \mathbf{k}(\tilde{\mathbf{t}}, \tilde{\mathbf{t}}) - \mathbf{k}(\tilde{\mathbf{t}}, \mathbf{t}) [\mathbf{k}(\mathbf{t}, \mathbf{t}) + \Sigma]^{-1} \mathbf{k}(\mathbf{t}, \tilde{\mathbf{t}}) \right)$$

```
df <- read.csv("call.csv")

mu.fun<-function(y,t,l,sigma2,sigmak2,n){

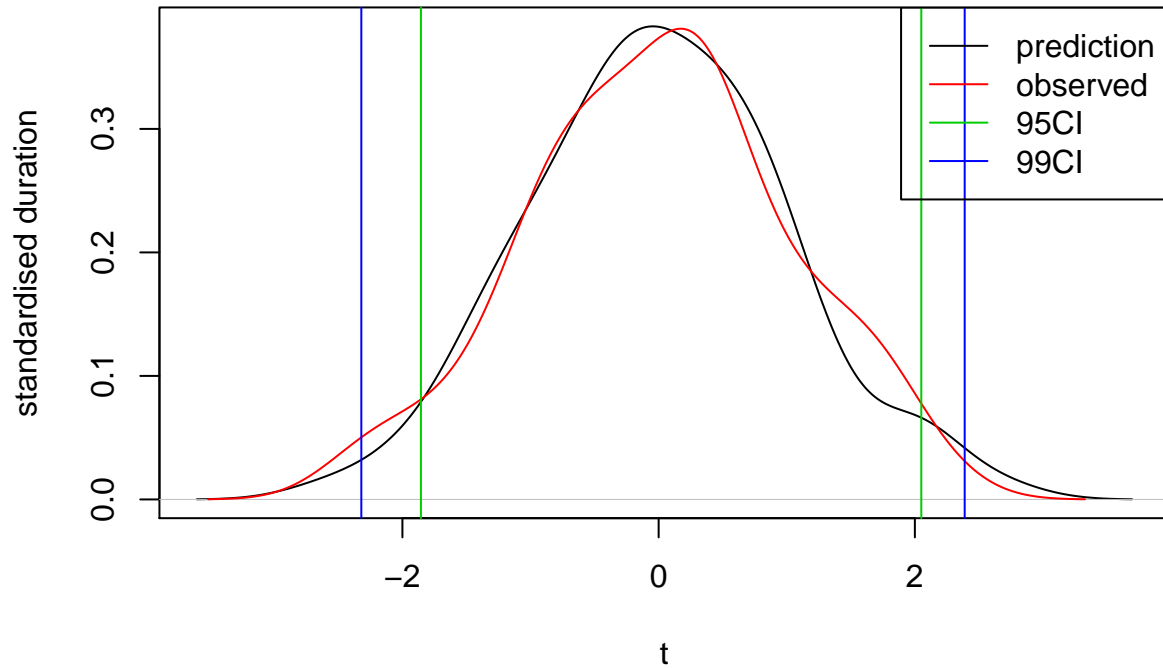
  np<-length(t) #number of location to evaluate Gaussian process.
  mT <- matrix(df$hour,np,np)
  mT_hat<-matrix(t,np,np)
  Sigma <- sigma2*diag(np)
  # kt_t<- sigmak2*exp(-l*(mT-t(mT))) + Sigma
  kt_t<- Sigma
  kt_th<- sigmak2*exp(-l*(mT-mT_hat)^2)
  kth_th<- solve(Sigma)
  result<-rmvnorm(n,mean=t(kt_th)%*% solve(kt_t)%*%y,sigma= kth_th - ( t(kt_th)%*% solve(kt_t)%*%kt_th
  return(result)
}

#An example of generating function with $n=5$.
t<-sort(sample(1:12,100,replace = TRUE)) #generate 100 points for gaussian process to be evaluated at.
test<-mu.fun(y=df$duration,t=t,l=0.05,sigma2=0.95,sigmak2=1.85,n=1)#plotting result

## Warning in rmvnorm(n, mean = t(kt_th) %*% solve(kt_t) %*% y, sigma = kth_th
## - : sigma is numerically not positive semidefinite

test_std <- (test-mean(test))/sd(test)
plot(density(test_std),main='realisations of Gaussian process', xlab="t",ylab="standardised duration")
den <- density(df$duration)
lines(density(df$duration),col=2)
abline(v=quantile(test_std,c(0.025,0.975)),col=3)
abline(v=quantile(test_std,c(0.005,0.995)),col=4)
legend("topright",legend = c("prediction","observed","95CI","99CI"),col = 1:4,lty = 1)
```

## realisations of Gaussian process



We see that on the right side of the distribution the upper limits are further from zero than lower limits. This is because there is more uncertainty at times we did not observe ( $t = 11, 12$ ).

- b) Was the inference performed in part a) fully Bayesian? If not, how would you make the analysis fully Bayesian, noting any particular difficulties that arise by assuming a Gaussian process prior. Your answer should not exceed one page of writing.

Answer:

The inference from part a) was not fully Bayesian as we fixed  $\sigma^2, \sigma_K^2, l$ . In order to make the inference fully Bayesian one should assign priors to  $\sigma^2, \sigma_K^2, l$ .

A difficulty of using a Gaussian prior is that  $\mathbf{K}(t, t)$  will not be full rank for  $t_i = t_j$  for every element this would mean  $\mathbf{K}(t, t)$  is not invertible.