# Lab week 12 MAST90125: Bayesian Statistical learning

## Making inference in noisy case.

Management at a 24 hour healthline are interested in phone call duration. The available data was the standardised length of phone calls, and which hour ($t = 4, \ldots, 22$) the phone call was initiated. The following model was assumed,

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

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

$$k(x, x') = \sigma_K^2 e^{-l \times \sin[(x-x')\pi/24]^2},$$

with $\sigma_K^2$ fixed to 1.21, and $l$ fixed to 0.5.

The researchers are interested in making predictions of phone call duration, $\tilde{\boldsymbol{\mu}}(t)$ for hours $t = 0, \ldots, 23$. As an initial step, they assumed $\sigma^2$ was 0.49.

### Instructions for lab

Download `call.csv` from LMS.

```
call<-read.csv('./calldata.csv',header=TRUE)
y<-call$length
t<-call$hour
n<-length(y)
```

- Based on the information provided, determine the joint distribution of data $\mathbf{y}$ and predictions $\tilde{\boldsymbol{\mu}}(t)$.
- Determine the distribution of $\tilde{\boldsymbol{\mu}}(t)$ conditional on $\mathbf{y}$, $\sigma^2$, $\sigma_K^2$ and $l$.
- Plot the predictions with 90 % and 95 % credible intervals along with the observed data. For this problem, would the Highest posterior density (HPD) and central credible intervals be different? Comment, in a Bayesian language, on the behaviour of predictions where no data was observed.

*Solution:*

If $y_i|\mu(t) \sim \mathcal{N}(\mu(t), \sigma^2)$ and $p(\mu(t)) = \mathcal{N}(0, k(t,t))$, then the joint distribution is

$$\begin{pmatrix} \mathbf{y} \\ \tilde{\boldsymbol{\mu}}(\mathbf{t}) \end{pmatrix} = \mathcal{N}\left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{k}(\mathbf{t},\mathbf{t}) + \sigma^2 \mathbf{I}_n & \mathbf{k}(\mathbf{t},\tilde{\mathbf{t}}) \\ \mathbf{k}(\tilde{\mathbf{t}},\mathbf{t}) & \mathbf{k}(\tilde{\mathbf{t}},\tilde{\mathbf{t}}) \end{pmatrix} \right)$$

The conditional distribution can be determined by arranging the kernel:

$$e^{-0.5\begin{pmatrix} \mathbf{y}' & \tilde{\boldsymbol{\mu}}(\mathbf{t})' \end{pmatrix} \begin{pmatrix} \mathbf{k}(\mathbf{t},\mathbf{t}) + \sigma^2 \mathbf{I}_n & \mathbf{k}(\mathbf{t},\tilde{\mathbf{t}}) \\ \mathbf{k}(\tilde{\mathbf{t}},\mathbf{t}) & \mathbf{k}(\tilde{\mathbf{t}},\tilde{\mathbf{t}}) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{y} \\ \tilde{\boldsymbol{\mu}}(\mathbf{t}) \end{pmatrix}}$$

To do this, the block matrix inversion formula will be useful.

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{pmatrix}$$

Applying the block matrix inverse result, we get

$$\begin{aligned} e^{-0.5\left(\mathbf{y}' \quad \tilde{\boldsymbol{\mu}}(\tilde{\mathbf{t}})'\right)\begin{pmatrix} \mathbf{k}(\mathbf{t},\mathbf{t})+\sigma^2\mathbf{I}_n & \mathbf{k}(\mathbf{t},\tilde{\mathbf{t}}) \\ \mathbf{k}(\tilde{\mathbf{t}},\mathbf{t}) & \mathbf{k}(\tilde{\mathbf{t}},\tilde{\mathbf{t}}) \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{y} \\ \tilde{\boldsymbol{\mu}}(\mathbf{t}) \end{pmatrix}} &= e^{-0.5\left(\mathbf{y}' \quad \tilde{\boldsymbol{\mu}}(\mathbf{t})'\right)\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{D} \end{pmatrix}\begin{pmatrix} \mathbf{y} \\ \tilde{\boldsymbol{\mu}}(\tilde{\mathbf{t}}) \end{pmatrix}} \\ &= e^{-0.5(\mathbf{y}'\mathbf{A}\mathbf{y}+\mathbf{y}'\mathbf{B}\tilde{\boldsymbol{\mu}}(\tilde{\mathbf{t}})+\tilde{\boldsymbol{\mu}}(\tilde{\mathbf{t}})'\mathbf{B}'\mathbf{y}+\tilde{\boldsymbol{\mu}}(\tilde{\mathbf{t}})'\mathbf{D}\tilde{\boldsymbol{\mu}}(\tilde{\mathbf{t}}))} \end{aligned}$$

where $\mathbf{D} = (\mathbf{k}(\tilde{\mathbf{t}},\tilde{\mathbf{t}}) - \mathbf{k}(\tilde{\mathbf{t}},\mathbf{t})(\mathbf{k}(\mathbf{t},\mathbf{t})+\sigma^2\mathbf{I}_n)^{-1}\mathbf{k}(\mathbf{t},\tilde{\mathbf{t}}))^{-1}$, $\mathbf{B} = -(\mathbf{K}(\mathbf{t},\mathbf{t}) + \sigma^2\mathbf{I}_n)^{-1}\mathbf{k}(\mathbf{t},\tilde{\mathbf{t}})\mathbf{D}$ and $\mathbf{A} = (\mathbf{k}(\mathbf{t},\mathbf{t})+\sigma^2\mathbf{I}_n)^{-1}(\mathbf{I}_n - \mathbf{k}(\mathbf{t},\tilde{\mathbf{t}})\mathbf{B}')$.

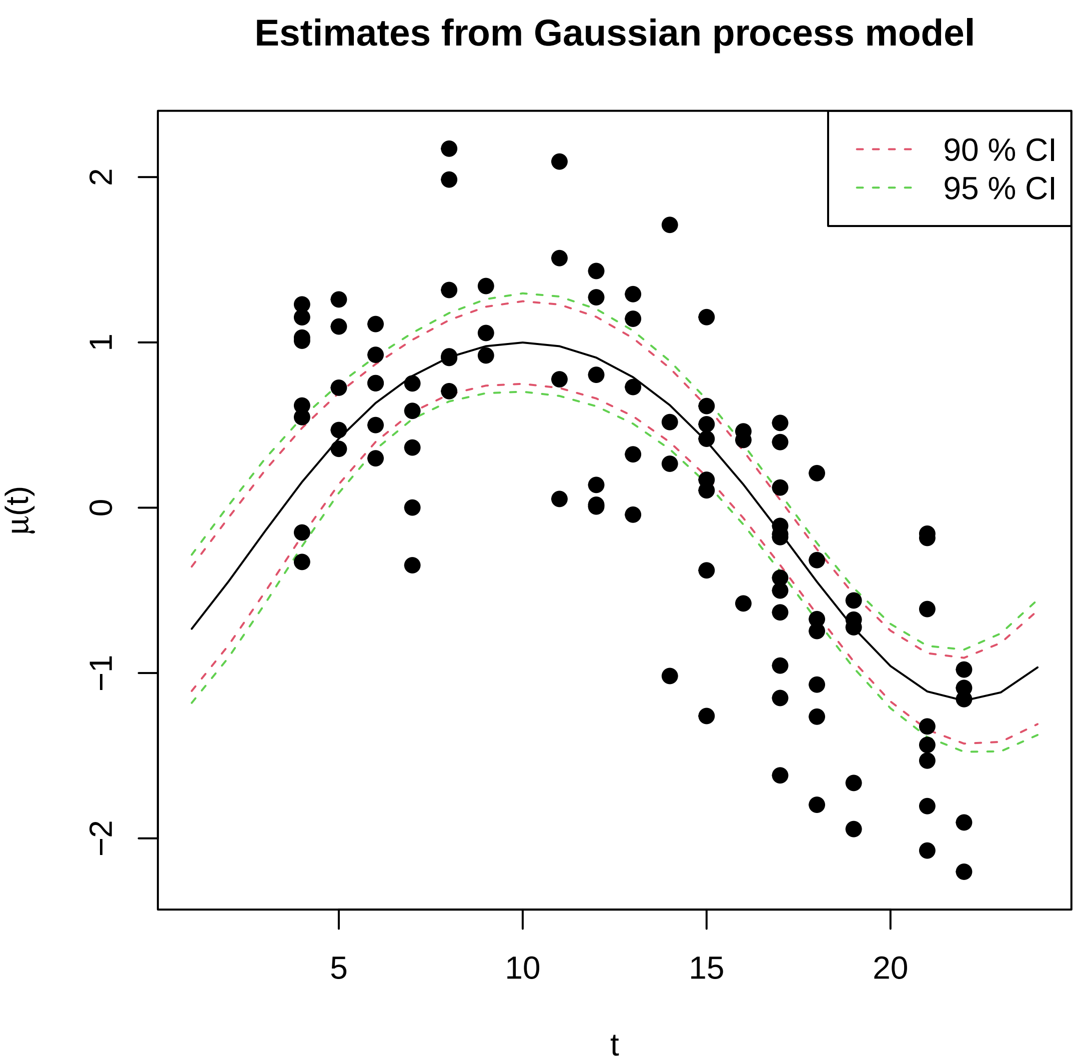
From this, we can determine that $\boldsymbol{\mu}(\tilde{\mathbf{t}})|\mathbf{y},\mathbf{t},\sigma_K^2,l,\sigma^2$ is normally distributed with mean $-\mathbf{D}^{-1}\mathbf{B}'\mathbf{y}$ and variance-covariance matrix $\mathbf{D}^{-1}$, or in the original notation,

$$p(\boldsymbol{\mu}(\tilde{\mathbf{t}})|\mathbf{y},\mathbf{t},\sigma_K^2,l,\sigma^2) = \mathcal{N}(\mathbf{k}(\tilde{\mathbf{t}},\mathbf{t})(\mathbf{k}(\mathbf{t},\mathbf{t})+\sigma^2\mathbf{I}_n)^{-1}\mathbf{y}, \mathbf{k}(\tilde{\mathbf{t}},\tilde{\mathbf{t}}) - \mathbf{k}(\tilde{\mathbf{t}},\mathbf{t})(\mathbf{k}(\mathbf{t},\mathbf{t})+\sigma^2\mathbf{I}_n)^{-1}\mathbf{k}(\mathbf{t},\tilde{\mathbf{t}})).$$

As we know all parameters of this distribution, we can directly determine the posterior and construct the plots without using sampling techniques.

```
call<-read.csv('calldata.csv',header=TRUE)

y<-call$length
t<-call$hour
n<-length(y)
t.all<-c(call$hour,0:23)  #times of predictions and observations
#Construct K.
np<-length(t.all)
mT<-matrix(t.all,np,np)
Kall<- 1.21*exp(-0.5*sin( (mT-t(mT))*pi/24)^2 )

yvarinv<-solve(Kall[1:n,1:n]+0.49*diag(n))
p.mean <- Kall[(n+1):np,1:n]%*%yvarinv%*%y
p.var  <- Kall[(n+1):np,(n+1):np] - Kall[(n+1):np,1:n]%*%yvarinv%*%Kall[1:n,(n+1):np]
mean.fun<-function(x){a<-1.21*exp(-0.5*sin((x-t)*pi/24)^2)%*%yvarinv%*%y;return(a) }
   #Here x is a scalar.
cov.fun<- function(x){ #Here x is a scalar
  av<-exp(-0.5*sin((x-t)*pi/24)^2)
  a<-1.21 -(1.21^2)*t(av)%*%yvarinv%*%av
  return(a) }

#Lower and upper limits of credible intervals
LL90<-qnorm(0.05,mean=p.mean,sd=sqrt(diag(p.var)))
LL95<-qnorm(0.025,mean=p.mean,sd=sqrt(diag(p.var)))
UL90<-qnorm(0.95,mean=p.mean,sd=sqrt(diag(p.var)))
UL95<-qnorm(0.975,mean=p.mean,sd=sqrt(diag(p.var)))

#Constructing plots.
ylims=c(min(c(LL95,y))-0.05,max(c(UL95,y))+0.05 )
plot(p.mean,type='l',ylim=ylims,xlab='t',ylab=expression(paste( mu,'(t)',sep='') ),
main='Estimates from Gaussian process model')
```

```
lines(LL90,col=2,lty=2)
lines(UL90,col=2,lty=2)
lines(LL95,col=3,lty=2)
lines(UL95,col=3,lty=2)
points(t,y,pch=19)
legend('topright',legend=c('90 % CI', '95 % CI'),col=2:3,lty=2)
```




```
#as curves via vectorization of mean.fun, cov.fun and other functions.
#The predictions at unobserved times produced by this vectorization are
#regarded as independent of each other, which is a problem.
mf <-Vectorize(mean.fun)
LL90<-function(x){a<-qnorm(0.05,mean=mean.fun(x), sd= sqrt(cov.fun(x)));return(a)}
UL90<-function(x){a<-qnorm(0.95,mean=mean.fun(x), sd= sqrt(cov.fun(x)));return(a)}
LL95<-function(x){a<-qnorm(0.025,mean=mean.fun(x), sd= sqrt(cov.fun(x)));return(a)}
UL95<-function(x){a<-qnorm(0.975,mean=mean.fun(x), sd= sqrt(cov.fun(x)));return(a)}
LL90v <-Vectorize(LL90)
```

```
UL90v <-Vectorize(UL90)
LL95v <-Vectorize(LL95)
UL95v <-Vectorize(UL95)

curve(mf,xlim=c(0,23),ylim=ylims,xlab='t',ylab=expression(paste( mu,'(t)',sep='') ),
      main='Estimates from Gaussian process model')
curve(LL90v, add=TRUE,col=2,lty=2)
curve(UL90v, add=TRUE,col=2,lty=2)
curve(LL95v, add=TRUE,col=3,lty=2)
curve(UL95v, add=TRUE,col=3,lty=2)
points(t,y,pch=19)
legend('topright',legend=c('90 % CI', '95 % CI'),col=2:3,lty=2)
```