Lab 6 Cokriging

Due October 23, 2014

In this exercise, we will use gstat package to conduct cokriging interpolation. We will use the dataset meuse dataset which is provided by R. The variable "lead" will be used as the target variable.

1. Dataset preparation

```
library(gstat)          #load the library
library(lattice)

?meuse                  #information on meuse
data(meuse)             #load the data
str(meuse)              #structure of the data

?meuse.grid             #Prediction grid
data(meuse.grid)
str(meuse.grid)

hist(meuse$lead, breaks=12)
hist(log10(meuse$lead), breaks=12)
```

Considering the severe sknewness with the original lead data, we will work with the log-transformed target variable. Given that many of the industrial processes result in lead pollution also produce zinc. We will use zinc as the secondary variable or co-variate for the cokriging.

```
hist(meuse$zinc, breaks=12)
hist(log10(meuse$zinc), breaks=12)
```

We will also work with the transformed covariate.

To explore the effect of co-kriging, we will create a situation where the target variable is under-sampled compared to the secondary variable.

```
#Extract a subset of the target variable
meuse.pb<-meuse[seq(1, length(meuse$lead), by=3), c("x", "y", "lead", "zinc")]
str(meuse.pb)

rownames(meuse.pb)          #the order of the extracted data
```

Attach the transformed lead data "ltpb" and the transformed zinc data "ltzn" to meuse.pb

```
meuse.pb<-cbind(meuse.pb, ltpb=log10(meuse.pb$lead), ltzn=log10(meuse.pb$zinc))
```

```
str(meuse.pb)

#Create a data frame of the lead observations at the extra points that have not been included
in the subsample
meuse.extra <- meuse[setdiff(rownames(meuse), rownames(meuse.pb)), c("x", "y", "lead")]
meuse.extra <- cbind(meuse.extra, ltpb = log10(meuse.extra$lead))
meuse.extra
```

Now we need to convert the data frame to explicitly spatial dataset.

```
class(meuse)          #class inquiry

coordinates(meuse)<- ~x+y        #convert data frame to "SpatialPointsDataFrame"
coordinates(meuse.pb) <- ~ x + y
coordinates(meuse.extra) <- ~ x + y
coordinates(meuse.grid) <- ~ x + y
class(meuse)

summary(meuse.pb)

bubble (meuse.pb, zcol= "ltpb", main= "log10 lead distribution")
```

2. Ordinary kriging

This part will focus on ordinary kriging on the lead variable

```
#variogram construction
v.ltpb<-variogram(ltpb~1, data=meuse.pb, cutoff=1800, width=200)
plot(v.ltpb)
m.ltpb<-vgm(psill=0.08, model="Sph", range =800, nugget=0.03)
plot(v.ltpb, pl=T, model=m.ltpb)
m.ltpb.f <- fit.variogram(v.ltpb, m.ltpb)
plot(v.ltpb, pl=T, model=m.ltpb.f)

#conduct ordinary kriging
k.o <- krige(ltpb ~1, locations=meuse.pb, newdata=meuse.grid, model=m.ltpb.f)
# summary statistics
summary(k.o)


#validation: prediction at the extra points
k <- krige(ltpb ~ 1, meuse.pb, meuse.extra, m.ltpb.f)
# compute and summarize validation errors
summary(k)
diff <- k$var1.pred - meuse.extra$ltpb
summary(diff)
sqrt(sum(diff^2)/length(diff))          # RMSE (precision)
```

```
sum(diff)/length(diff)          # mean error (bias)
median(diff)                    # median error
```

Assignment #1

1.Conduct a cross validation here to assess the accuracy of ordinary kriging.
2.Discuss the cross validation results compared with the validation results at extra points above?

3. Cokriging

Assess the relationship between the lead variable and the zinc variable

```
#correlation
cor(meuse.pb$ltzn, meuse.pb$ltpb)
plot(meuse.pb$ltzn, meuse.pb$ltpb)

#extract the co-variate
meuse.co<-as.data.frame(meuse)[, c("x", "y", "zinc")]
meuse.co <- cbind(meuse.co, ltzn = log10(meuse.co$zinc))
coordinates(meuse.co) <- ~ x + y

#variogram
v.ltzn <- variogram(ltzn ~ 1, meuse.co, cutoff=1800)
plot(v.ltzn, pl=T)
m.ltzn <- vgm(psill=.11, model="Sph", range=1000, nugget=.02)   #by visual inspection
m.ltzn.f <- fit.variogram(v.ltzn, m.ltzn)                       #by least square fit
plot(v.ltzn, pl=T, model=m.ltzn.f)
```

Model the coregionalization: create a gstat structure containing sets of data: subsample for lead and full sample for zinc

```
g1 <- gstat(NULL, id = "ltpb", form = ltpb ~ 1, data = meuse.pb)
g1 <- gstat(g1, id = "ltzn", form = ltzn ~ 1, data = meuse.co)

#variogram and cross-variogram
v.cross <- variogram(g1)
plot(v.cross, pl=T)

#Fit cross-variogram model
g1 <- gstat(g1, id = "ltpb", model = m.ltpb.f, fill.all=T)
g1 <- fit.lmc(v.cross, g1)
plot(variogram(g1), model=g1$model)

#Prediction at the grids
k.c1 <- predict.gstat(g1, meuse.grid)
summary(k.c1$ltpb.pred); summary(k.c1$ltpb.var)
```

Validation

```
#Interpolate at extra points
k <- predict.gstat(g1, meuse.extra)

diff <- k$ltpb.pred - meuse.extra$ltpb
summary(diff)
sqrt(sum(diff^2)/length(diff))          # RMS error (precision)
sum(diff)/length(diff)                  # mean error (bias)
median(diff)                            # median error
```

Cross-validation

```
cv.c2 <- gstat.cv(g1, nfold=52)
summary(cv.c2$residual)
sqrt(mean(cv.c2$residual^2))
mean(cv.c2$residual)
mean(cv.c2$residual^2/cv.c2$ltpb.var)
median(cv.c2$residual)
```

Assignment #2

1. How cokriging validation and the cross-validation results by cokriging compare with the those by kriging?
2. Plot the prediction and error images of k.c1 and k.o. Discuss the difference.

Assignment #3

Use “om” in the meuse dataset instead as the covariate and re-run the cokriging. Compare the prediction accuracy with that using the zinc as co-variate. Discuss why.