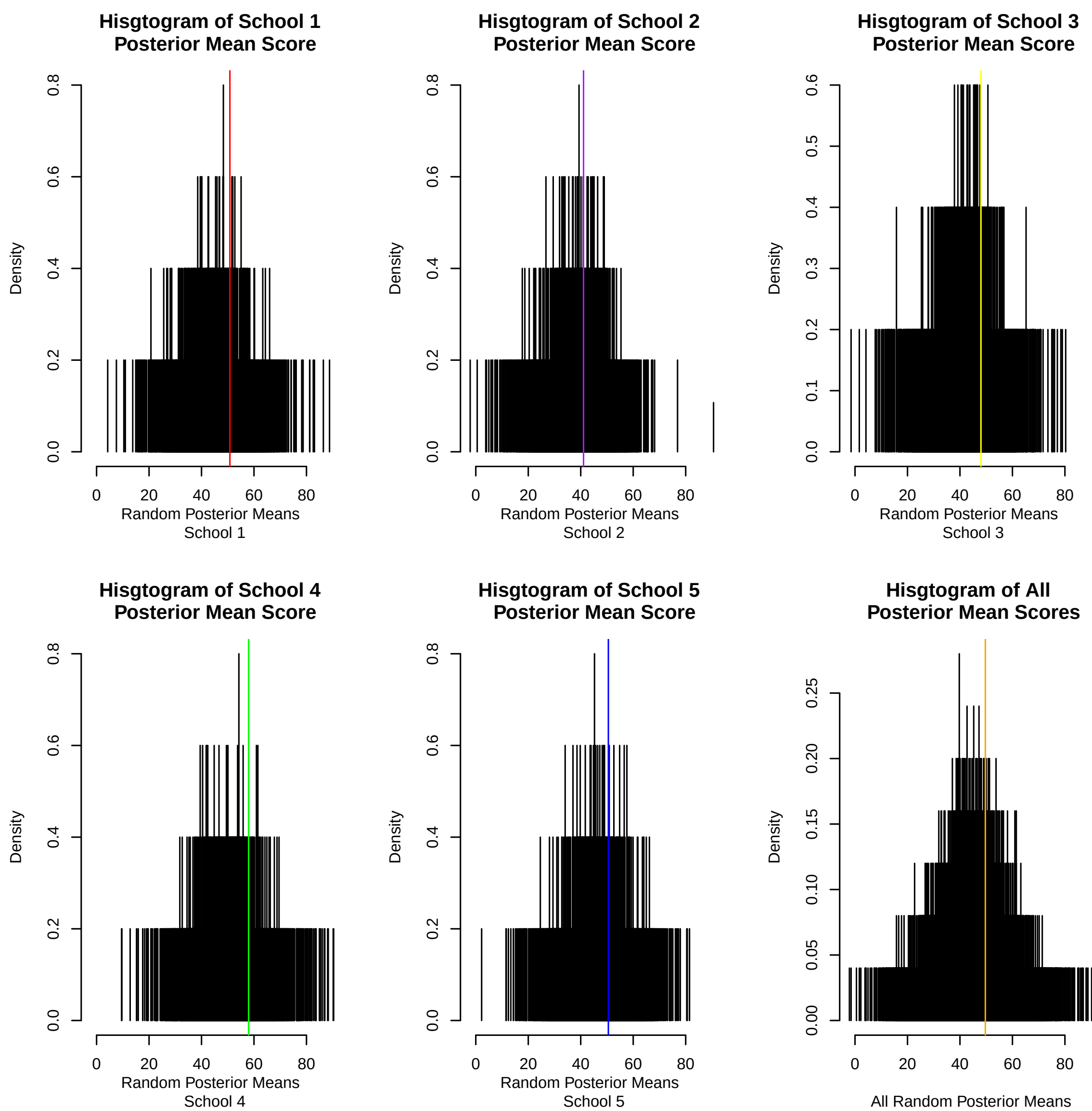
1. *Stein's Paradox: Do not do.*

2. *Derivation, see attached.*

3. *(30 points) Generate data from a two-dimensional hierarchical normal model (e.g. students within a school). Write code for a Gibbs sampler and convince me that your code returns the correct answer.*

   2D means two levels. The plots below will hopefully convince you my code worked because the true values are all within the posterior distributions and I used a weakly informative prior, so the priors should not affect the posteriors as much as the simulated data, which were simulated from the true process.



4. *Show that* $p(H_o|Data) = (1 + \frac{1-\pi_o}{\pi_o} BF^{-1})$.

p$(H_o|Data) = p(H_o|\underset{\sim}{x}) = (1 + \frac{1-\pi_o}{\pi_o} BF^{-1})^{-1}$

$= (1 + \frac{(1-\pi_o)p(H_1|\underset{\sim}{x})p(H_o)}{\pi_o P(H_o|\underset{\sim}{x})p(H_1)})^{-1}$

$= (1 + \frac{(p(H_1)p(H_1|\underset{\sim}{x})p(H_o)}{p(H_o)p(H_o|\underset{\sim}{x})p(H_1)})^{-1}$

$\rightarrow$

$p(H_o|\underset{\sim}{x}) = p(H_o|\underset{\sim}{x})[p(H_o\underset{\sim}{|}x) + p(H_1\underset{\sim}{|}x)]^{-1}$

$\rightarrow$

$1 = 1$ since $[p(H_o\underset{\sim}{|}x) + p(H_1\underset{\sim}{|}x)] = 1$

5. *(10 points) Summarize the paper you have selected for the final project. What about this paper/method to you still not understand (be specific!)?*

Diggle, P., Lophaven, S. (2006). Bayesian geostatistical design. Scandinavian Journal of Statistics, 33(1), 53-64.

Given a pre-existing sampling design, future studies may require fewer or more sampling points. Criteria for such called retrospective and prospective designs, respectively, reward accuracy in spatial prediction while accounting for parameter estimation. The authors' Bayesian design criteria approach, the spatial average of VarS(x)—y, and then compared to other developed criteria, "classical". Comparisons between whether ν^2 is known or unknown minimally affected the criteria. Under a prospective lattice design, close pairs is favored to in-fill added designs, but changes in locations of added points within the respective designs resulted in small changes in the criteria.

Still not understood:

- *p. 54 line 2:* What is meant by "proper allowance" exactly?
- *p. 56 equation (1):* Why do they have to integrate out θ to make predictions? Is $[s|Y, \theta]$ not sufficient?
- In general, I don't understand the exact model fit to point reference data. Also, the authors say they assumed an exponential correlation function on p. 57, and I don't know how that was done, in a classical or bayesian framework.
- *bottom of p. 57:* In general, what about classical design criteria results in "well separated" monitoring sites compared to the Bayesian approach?
- *p. 61:* What are the implications of the locations of the added points not affecting the criteria much as the way the added points entered? Also, why did the authors do the 5 replicates?
- **Other terms/notation:**
  - kringing

– stoichasitc process

– Notation p. 56, how does [S|Y] differ from [T|Y], T = target, so is this like popuation (target) and sample (observed)?

– Method in Diggle et al. (2003) to generate direct Monte Carlo samples

– Where does the noise-to-signal variance ratio come in?

– Eutrophication

– Exact definition of diffuse prior, I assume it's flat but not improper, but will look this up.

R Code

4.
```
Y.school.mathscore<-dget("https://www.stat.washington.edu/~pdhoff/Book/Data/data/Y.school.mathscore")

Y <- Y.school.mathscore
#head(Y)

### say there are 5 schools with 12 classes in each school,
### say elementary schools in a town the size of bozeman

Z <- NULL
Z$school <- c(rep(1,12), rep(2,12), rep(3,12), rep(4,12), rep(5,12))

### there are two classes of each k-5 grades, lets say all teachers within a
### grade level produce the same quality of students

### to make things interesting, set school means to first five of the hoff data set

true.thetaj <- by(Y[,2], as.factor(as.character(Y[,1])), mean)[1:5]

### simplify things, each school has the same variability, use the variation from all schools in hoff

true.sigma2 <- var(Y[,2])

### make class data from school data
set.seed(53206)
a1 <- rnorm(12,true.thetaj[1], sd = sqrt(true.sigma2))
a2 <- rnorm(12,true.thetaj[2], sd = sqrt(true.sigma2))
a3 <- rnorm(12,true.thetaj[3], sd = sqrt(true.sigma2))
a4 <- rnorm(12,true.thetaj[4], sd = sqrt(true.sigma2))
a5 <- rnorm(12,true.thetaj[5], sd = sqrt(true.sigma2))

Z$mathscore <- c(a1,a2,a3,a4,a5)

Y <- data.frame(Z)

### weakly informative priors

#prior on sigma2
nu.0<-1
sigmasq.0<-100

#prior on tau2, uses same sigmasq.0 from prior on sigma2, but here it is called tausq.0
eta.0<-1
tausq.0<-100

#prior on mu = mean of group math scores, don't have any intuition about these
#prior mean is 50, set prior variance to 144 say

mu.0<-0
gammasq.0<-144
###

### starting values

m <- 5 # number of schools

#find the mean, variance, and sample size of each school
n<-sv<-ybar<-rep(NA,m)
for(j in 1:m)
{
  #mean by group
ybar[j]<-mean(Y[Y[,1]==j,2])
   #var by group
sv[j]<-var(Y[Y[,1]==j,2])
   #n by group
```

```
n[j]<-sum(Y[,1] ==j)
}
theta<-ybar
#sigma2 is the average group variance
sigma2<-mean(sv)
mu<-mean(theta)
#tau2 is the variance of the group means
tau2<-var(theta)
###

### setup MCMC
set.seed(5326)
S<-5000
THETA<-matrix( nrow=S,ncol=m)
MST<-matrix( nrow=S,ncol=3)
###

### MCMC algorithm
for(s in 1:S)
{

# sample new values of the thetas
for(j in 1:m)
{
vtheta<-1/(n[j]/sigma2+1/tau2)

etheta<-vtheta*(ybar[j]*n[j]/sigma2+mu/tau2)
theta[j]<-rnorm(1,etheta,sqrt(vtheta))
}
#}
#sample new value of sigma2
nun<-nu.0+sum(n)
ss<-nu.0*sigmasq.0;

ss1 <- ss + sum((Y[1:12,2] - theta[1])^2)
ss2 <- ss + sum((Y[13:24,2] - theta[2])^2)
ss3 <- ss + sum((Y[25:36,2] - theta[3])^2)
ss4 <- ss + sum((Y[37:48,2] - theta[4])^2)
ss5 <- ss + sum((Y[49:60,2] - theta[5])^2)

ssk <- c(ss1,ss2,ss3,ss4,ss5)

#don't understand how the code below works
for(j in 1:m){
ss<-ss+sum((Y[,2]-theta[j])^2)
}

sigma2<-1/rgamma(1,nun/2,ss/2)

#sample a new value of mu
vmu<- 1/(m/tau2+1/gammasq.0)
emu<- vmu*(m*mean(theta)/tau2 + mu.0/gammasq.0)
mu<-rnorm(1,emu,sqrt(vmu))

# sample a new value of tau2
etam<-eta.0+m
ss<- eta.0*tausq.0 + sum( (theta-mu)^2 )
tau2<-1/rgamma(1,etam/2,ss/2)

#store results
#for(s in 1:S) {
THETA[s,]<-theta
MST[s,]<-c(mu,sigma2,tau2)
}

par(mfrow=c(2,3))
hist(THETA[,1], breaks = c(min(THETA) - 0.001, seq(min(THETA), max(THETA),by = 0.001),
                          max(THETA) + 0.001),
     main = "Hisgtogram of School 1 \n Posterior Mean Score", xlab =
       "Random Posterior Means \n School 1")
abline(v=true.thetaj[1], col = "red")

hist(THETA[,2], breaks = c(min(THETA) - 0.001, seq(min(THETA), max(THETA),by = 0.001),
                          max(THETA) + 0.001),
     main = "Hisgtogram of School 2 \n Posterior Mean Score", xlab =
       "Random Posterior Means \n School 2")
abline(v=true.thetaj[2], col = "purple")

hist(THETA[,3], breaks = c(min(THETA) - 0.001, seq(min(THETA), max(THETA),by = 0.001),
                          max(THETA) + 0.001),
     main = "Hisgtogram of School 3 \n Posterior Mean Score", xlab =
       "Random Posterior Means \n School 3")
 abline(v=true.thetaj[3], col = "yellow")

 hist(THETA[,4], breaks = c(min(THETA) - 0.001, seq(min(THETA), max(THETA),by = 0.001),
                          max(THETA) + 0.001),
     main = "Hisgtogram of School 4 \n Posterior Mean Score", xlab =
       "Random Posterior Means \n School 4")

abline(v=true.thetaj[4], col = "green")

hist(THETA[,5], breaks = c(min(THETA) - 0.001, seq(min(THETA), max(THETA),by = 0.001),
                          max(THETA) + 0.001),
     main = "Hisgtogram of School 5 \n Posterior Mean Score", xlab = "Random Posterior Means \n School 5")
 abline(v=true.thetaj[5], col = "blue")

 hist(THETA, breaks = c(min(THETA) - 0.001, seq(min(THETA), max(THETA),by = 0.001),
```

```
                             max(THETA) + 0.001),
     main = "Hisgtogram of All \n Posterior Mean Scores", xlab = "All Random Posterior Means")
abline(v=mean(true.thetaj), col = "orange")
```