Exercises V (these are computational practices, rather than mathematical problems. The goal is to work through the examples and understand each step and how it's done with R or BUGS)

1. Assume the true disease incidence varies over a number of geographical regions i, e.g. $i = 1, \ldots, 100$. To make your own data, generate these true values λ_i from Gamma$(5, 1000)$. (in R: `la <- numeric(); for(i in 1:100){la[i]<-rgamma(1,5,1000)}`). Then, for each generated λ_i, generate the actual disease count X_i for each region from Poisson$(N_i\lambda_i)$, assuming the exposure (population count) is $N_i = i \times 10$. (in R: use `x[i]<-rpois(1,N[i]*la[i])`). Now, you know the 'true values' of λ_i, and you have 'observed data'. Compute the observed incidences $\hat{\lambda}_i = X_i/N_i$ and plot them as a function of population size. Observe how these behave when N_i is small. Finally, Compute in OpenBUGS the posterior distribution of all λ_i assuming the prior $\lambda_i \sim$ Gamma(α, β), $\alpha \sim$ Exp(0.01), $\beta \sim$ Exp(0.01). (This makes a hierarchical model). Compute posterior medians and 95% CIs and compare them with the observed incidences, $\hat{\lambda}_i$, as a function of population size. (After running the simulation, and after having sample of λ_i this can be done by clicking `Inference`, then `Comparison` to open 'Comparison tool', and setting 'lambda' as 'node', 'laobs' as 'other', and 'N' as 'axis', then click `model fit`). BUGS code below:

```
model{
for(i in 1:100){
N[i] <- i*10
x[i] ~ dpois(par[i]); par[i] <- lambda[i]*N[i]
lambda[i] ~ dgamma(a,b)
laobs[i] <- x[i]/N[i]
}
a ~ dexp(0.01); b~dexp(0.01)
}
list(x=c(your generated data))
```

2. Assume there are 10 types of bacteria. In a sample of 5 detected bacteria, there were 3 of type I, 1 of type III, and 1 of type VII. The sample could be modeled as multinomial sample $\pi(X_1, \ldots, X_{10}) =$ Multinom$(N, p_1, \ldots, p_{10})$ with $N = 5$, (so $\sum_i X_i = 5$). Parameters p_i represent the percentages of each bacteria type in a large population, (so $\sum_i p_i = 1$). What is the posterior distribution of the percentages p_i based on this small sample, assuming the prior is 'uninformative' Dirichlet$(1, \ldots, 1)$. Is the data or the prior more influential on the result? Repeat the analysis with a similar Dirichlet-prior with 'prior sample size' of one. What is the problem with using the improper prior Dirichlet$(0, \ldots, 0)$? (Check the marginal posterior distribution of p_i for some bacteria type that was not observed). Finally, run the model with the two proper priors in OpenBUGS.

```
model{
p[1:10] ~ ddirich(a[])
x[1:10] ~ dmulti(p[],5)
}
list(x=c(3,0,1,0,0,0,1,0,0,0),a=c(1,1,1,1,1,1,1,1,1,1))
```

3. Assume normal model $\pi(X_i) = \mathrm{N}(\mu, \sigma^2)$ with both parameters unknown. The goal is to compute posterior distribution $\pi(\mu, \sigma^2 \mid X_1, \ldots, X_n)$. The prior distribution is assumed to be chosen independently for both parameters: $\pi(\mu) \propto 1$, and $\pi(\sigma^2) \propto 1/\sigma^2$. (Both are improper priors). The

full conditional distributions for μ and $\tau = 1/\sigma^2$ are then $\pi(\mu \mid \sigma^2, X_1, \ldots, X_n) = \mathrm{N}(\bar{X}, \sigma^2/n)$ and $\pi(\tau \mid \mu, X_1, \ldots, X_n) = \mathrm{Gamma}(n/2, 0.5\sum_{i=1}^{n}(X_i - \mu)^2)$. Construct a Gibbs sampler in R and simulate the posterior distribution. The data are (actually generated from $\mathrm{N}(0,1)$):

```
x<-c(0.34,  0.87, -0.80,  0.65,  1.34,  0.91, -0.71,  0.30, -1.08,  0.73, -1.23,  1.52,
2.53,  0.90,  0.06,  0.64,  0.58, -0.45, -2.40, -1.58)

n <- length(x)
mu <- numeric();tau<-numeric()
tau[1] <- 1; sig2[1]<-1/tau[1]
mu[1] <- 0
for(i in 2:3000){
mu[i] <- rnorm(1,mean(x),(1/tau[i-1])/n)
tau[i] <- rgamma(1,n/2,0.5*sum((x-mu[i])^2)); sig2[i]<-1/tau[i]
}
```

4. Do the same modeling as above in OpenBUGS by writing the priors and the conditional distribution for the data points. In BUGS, you need to specify proper prior distributions. To make them as 'uninformative' as possible, you might choose `mu` $\sim$ `dunif(-100,100)` or `mu` $\sim$ `dnorm(0,0.001)`. Also, knowing that the improper prior $\pi(\sigma^2) = 1/\sigma^2$ is the same as to have $\pi(\log(\sigma)) \propto 1$, you can set uniform prior for $\log(\sigma)$ as `logs` $\sim$ `dunif(-100,100)` and then define `sigma2 <- exp(logs)*exp(logs)`. To plot scatter plot in BUGS, click `Inference` and within that `Correlations` to open 'Correlation Tool' where you can input 'mu' and 'sigma2' and select 'scatter'.

```
model{
mu ~ dunif(-100,100)
sigma2 <- exp(logs)*exp(logs); tau <- 1/sigma2
logs ~ dunif(-100,100)
for(i in 1:20){
x[i] ~ dnorm(mu,tau)
}
}
```

5. A sample of $n = 50$ individuals are tested for some disease, $x = 7$ of them were found to be positive. The diagnostic test has sensitivity q (That is $q = P(\text{test is '+'} \mid \text{truly disease})$). The population prevalence of the disease is p. The goal is to estimate the population prevalence based on the sample results. Background information is that the test sensitivity is known to be about 85 $\pm$5%. This could be roughly described by a Beta-prior with mean 0.85 and variance $(0.05/1.96)^2$ (using simple normal approximation of 95% interval $\pm 1.96\sigma$). Calculate the parameters for this Beta-prior and write OpenBUGS model to compute the posterior disribution of p, q. The prior for p can be $\mathrm{U}(0,1)$. The model for observations is Binomial$(50, p * q)$. Plot the marginal distribution of p and the 2D-scatterplot of p, q. Check what the result would be if the prior of sensitivity was uninformative $\mathrm{U}(0,1)$.

```
model{
aa <- -m*(m*m-m+var)/var; bb<-(m*m-m+var)*(m-1)/var
```

```
m <- 0.85; var <- pow(0.05/1.96,2)
q ~ dbeta(aa,bb)
p ~ dunif(0,1)
x ~ dbin(pr,50); pr<-p*q
x <- 7
}
```