Department of Mathematics and Statistics
WinBUGS and OpenBUGS applications
Exam 9.3.2011

(5 questions, 30 points total. You may write in English/Finnish)

1. Explain the meaning of the following terms (1 point each):
(A) Improper prior, (B) Parent node, (C) Posterior predictive distribution,
(D) Full conditional density, (E) Zero's trick, (F) Thinning.

(A) A prior 'distribution' with an infinite integral, so that it cannot be normalized to make a proper distribution. (Posterior might still be proper, but not necessarily).
(B) A node in a DAG that has direct descendant(s) which is/are either (1) stochastic and conditionally dependent on the values of parent node(s), or (2) deterministic functions of the parent node(s).
(C) Predictive distribution that is conditional on observed data, not conditional on unknown parameters. This is obtained by 'integrating out' the unknown parameter: $P(X \mid \text{data}) = \int_\Theta P(X \mid \theta)P(\theta \mid \text{data})\mathbf{d}\theta$ where $P(X \mid \theta)$ is the parametric model of X and $P(\theta \mid \text{data})$ is the posterior of θ.
(D) This is needed in Gibbs-sampling algorithms. Assume the full distribution is d-dimensional $\pi(x_1, \ldots, x_d)$. The full conditional, say for x_1, is $\pi(x_1 \mid x_2, \ldots, x_d)$, and likewise for all other components.
(E) Method of expressing the probability of an event (or logical statement) when this has no direct representation among the probability distributions available in WinBUGS. To model such probability (or probability density) $\pi(X \mid \theta)$, it is written in the form of $\exp(-(-\log(\pi(X \mid \theta))))$ which correspond to Poisson probability of 'zero' with parameter $\lambda = -\log(\pi(X \mid \theta))$. An arbitrary constant may be needed, to be added to this to ensure that the Poisson parameter is positive. This corresponds to multiplication of $\pi(X \mid \theta)$ by a constant so that it does not alter the likelihood contribution for θ.
(F) Taking only every kth sample from the MCMC output, to be used as draws from the target distribution. Thinning reduces autocorrelations so that we could get nearly independent Monte Carlo samples.
2. (3+3 points):
(A) Explain the advantages and disadvantages of using conjugate priors in Bayesian models. (B) Explain the stepwise variable selection method (in regression models) based on indicators and DIC.

(A) Some advantages are: the posterior then becomes a known standard distribution for which means, modes, medians, variances, quantiles may be easily available. Also, the mean can be interpreted as a weighted sum of data average and 'prior data average'. The conjugate prior is clearly equivalent to some amount of 'prior data'. No MCMC needed for sampling. Disadvantage: closed form solutions exist only for a limited number of problems, so that these can be too naive approaches (both for representing distribution of observations $\pi(X \mid \theta)$ and prior $\pi(\theta)$), but nevertheless a good starting point.
(B) Every explanatory variable is multiplied by an indicator so that any combination of variables can be expressed as a vector of 0/1-indicators. We can start e.g. with a full model with all variables included. Then, candidate models are constructed by changing each of the indicator variables at a time, creating p candidate models if we have p variables which can either be included or excluded. DIC is computed for each of the models, and the model with lowest DIC is selected as the new 'current' model. Again, candidate models are constructed similarly and DIC computed for all models, and the

best model selected at each step. After some number of steps, we find the current model to have the lowest DIC compared to all candidate models, and the procedure stops. This model may not be the best of all 2^p models.

3. (6 points):
Explain the models behind the following BUGS implementations:

```
model{
for(i in 1:5){ # groups
 m[i] <- inprod(beta[],X[i,1:6])
for(j in 1:4){ # measurements
y[i,j] ~ dnorm(mu[i,j],tau)
mu[i,j] <- inprod(beta[1:6],X[i,1:6])
            # X given in data list
}}
tau ~ dgamma(0.001,0.001); s <- pow(tau,-0.5)
for(i in 1:6){beta[i] ~ dnorm(0,0.001)}
}
 # data:
 list(y=structure(.Data=c(
 -0.34,  1.34, 3.25, 0.22,
  0.89, -0.07, 1.08, 4.56,
  4.12,  1.90, 2.45, 0.27,
  4.95,  2.44, 2.34, 4.62,
  3.62,  9.84, 6.14, 4.63),.Dim=c(5,4)),
 # Alternative 1:
X=structure(.Data=c(0,1,0,0,0,0,
                    0,0,1,0,0,0,
                    0,0,0,1,0,0,
                    0,0,0,0,1,0,
                    0,0,0,0,0,1),.Dim=c(5,6)))
# Alternative 2:
 X=structure(.Data=c(1,-1,-1,-1,-1,-1,
                     1,0,1,0,0,0,
                     1,0,0,1,0,0,
                     1,0,0,0,1,0,
                     1,0,0,0,0,1),.Dim=c(5,6)))
# Alternative 3:
 X=structure(.Data=c(1,0,0,0,0,0,
                     1,0,1,0,0,0,
                     1,0,0,1,0,0,
                     1,0,0,0,1,0,
                     1,0,0,0,0,1),.Dim=c(5,6)))
```

The code implements a one-way ANOVA model for estimating group effects for categorical groups. The first alternative uses parametrization without intercept, whereas the second alternative uses sum-

to-zero constraint and the third alternative uses corner constraints for parameter identifiability.

4. (6 points):
Write a BUGS model code for the following model

$$\pi(\alpha_1,\ldots,\alpha_5,\mu,\sigma \mid Y,C) \propto \prod_{i=1}^{m}\pi(Y_i \mid \theta_i)\prod_{i=1}^{n}P(Y_{i+m} > C_i \mid \theta_{i+m})\prod_{i=1}^{K}\pi(\alpha_i \mid \mu,\sigma)\pi(\mu)\pi(\sigma)$$

where Y are exponentially distributed 'event times' with means $1/\theta_i$ using log-link: $\log(1/\theta_i) = \alpha_k$ when individual i belongs to group k. (Every individual belongs to some group $k \in \{1,...,K\}$). Group level parameters α_k have normal prior $\mathrm{N}(\mu,\sigma^2)$, and μ has hyper prior $\mathrm{N}(0,1000)$ and σ has hyper prior $\mathrm{U}(0,1000)$. Data are of the form: $Y = c(.,...)$ a vector of length $m+n$ (n missing values) and $C = c(.,...)$ a vector of length n (censoring points for missing Y), and $g = c(.,...)$ a vector of length $m+n$ containing group numbers $k \in \{1,...,K\}$ for all individuals.

```
model{
for(i in 1:m){
  Y[i]~dexp(theta[i])
  theta[i] <- exp(-alpha[g[i]])
  }
for(i in 1:n){
  Y[i+m] ~ dexp(theta[i+m])I(C[i],)
  theta[i+m] <- exp(-alpha[g[i+m]])
  }
for(i in 1:K){alpha[i]~dnorm(mu,tau)}
mu ~ dnorm(0,0.001); sigma ~ dunif(0,1000)
tau <- 1/(sigma*sigma)
}
# example initials:
list(Y=c(NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,
9,9,9),mu=0,sigma=1)
# example data:
list(m=17,n=3,K=5,Y=c(14,8.5,1,7,3,13.3,8.8,6,3,2,17.9,8.9,2,3,
1,13.7,8.4,NA,NA,NA),
C=c(3,1,2),g=c(1,2,3,4,5,1,2,3,4,5,1,2,3,4,5,1,2,3,4,5))
```

5. (6 points):
Find and explain the errors in the following BUGS model code.

```
model{
N <- 5   # this should be given in the data
for(i in 1:N){
y[i] ~ dbinom(n[i],p[i])
# should be 'dbin' and parameters (p[i],n[i])
p[i] <- logit(inprod(alpha[],x[i,]))
```

```
# should be logit(p[i]) <- inprod(alpha[],x[i,]))
}
for(i in 1:4){alpha[i]~dnorm(0,100^2)
# should be '1:3' although alpha[4] does not interfere.
# should have dnorm(0,0.0001) with low precision,
# also 100^2 is syntax error anyway.
# missing '}' in the loop.
z ~ dbern(step(mean(p)-0.5)) #P(average(p) >0.5)
# should be z <- step(mean(p[])-0.5) for
# calculating this probability.
# (Actually, should account for the fact that step(0)=1
# and the required ">0.5" but since p is continuous,
# this does not matter).
# also mean(p) is syntax error without "p[]"
}
# data:
list(y=c(0,20,10,50,50),n=c(50,10,40,70,60),
# inconsistent data where n[2] < y[2]
x=structure(.Data=c(1.2,5.1,1.4,
              3.3,4.1,1.5,
              3.2,2.6,3.1,
              2.7,1.3,4.3,
              2.2,1.2,5.4),.Dim=c(5,3) ))
```