Exercises III

1. Estimating genotypes in a family. Assume the parents have three possible genotypes (AA,Aa,aa). The prior probability of each allele, A or a, is 0.5 and in the prior the alleles are assumed independent. What is the prior probability of each genotype for parents? Write conditional probabilities of the genotypes of the children in a table, for each genotype combination of the parents. Draw this model as a DAG where the nodes of the DAG represent variables denoting parent genotypes and the genotypes for two children. Write the same DAG but including explicitly a probability parameter $\theta = (\theta_1, \theta_2, \theta_3)$ denoting the conditional probabilities for child genotypes. (Parameter θ depends deterministically on parental genotypes, i.e. θ is a deterministic node in BUGS syntax, whereas genotypes themselves are stochastic nodes).

2. Open either WinBUGS or OpenBUGS. (BUGS should be found in computer class, or you can download and install at home from e.g. http://www.openbugs.net/). Find out in BUGS user manual or in its help menu the list of `Distributions` and find the syntax for '`Categorical`' distribution. Also find out the list of `Functions` and find '`equals`' there. Explain each line of the following BUGS code, and how it corresponds to a DAG, and what posterior distribution is obtained for what variables, when the child-variables are assigned values from observed data. Compute in BUGS the posterior probability P(mother is of type AA | both children were AA). *Advice: steps of running BUGS models are explained in the longer course text p.34-35. (Note: probability of an event is the expected value of the (binary) indicator variable for that event, and this is approximated by Monte Carlo sample mean from the corresponding distribution of that variable.* `Sample Monitor Tool -> stats`*).* Optional further exercises: you can try adding more children to the family with their genotypes, and see how this information changes the probabilities for parents. Also, you can experiment the effect of prior probabilities of alleles by changing them. For this problem you can in principle calculate exact posterior probabilities by 'paper and pencil' too.

```
model{
mother ~ dcat(p0[1:3]);  father ~ dcat(p0[1:3])
motherAA <- equals(mother,1)
for(i in 1:2){
child[i] ~ dcat(p[mother,father,1:3])
}
p0[1]<-1/4; p0[2]<-1/2; p0[3]<-1/4

p[1,1,1] <- 1; p[2,2,1] <- 1/4; p[3,3,1] <- 0;
p[1,1,2] <- 0; p[2,2,2] <- 1/2; p[3,3,2] <- 0;
p[1,1,3] <- 0; p[2,2,3] <- 1/4; p[3,3,3] <- 1;
p[1,2,1] <- 1/2; p[2,1,1] <- 1/2;
p[1,2,2] <- 1/2; p[2,1,2] <- 1/2;
p[1,2,3] <- 0; p[2,1,3] <- 0;
p[1,3,1] <- 0; p[3,1,1] <- 0;
p[1,3,2] <- 1; p[3,1,2] <- 1;
p[1,3,3] <- 0; p[3,1,3] <- 0;
p[2,3,1] <- 0; p[3,2,1] <- 0;
p[2,3,2] <- 1/2; p[3,2,2] <- 1/2;
```

```
 p[2,3,3] <- 1/2; p[3,2,3] <- 1/2;
 }
# data given in this list:
list(child=c(1,1))
```

3. Assume the exponential model $\pi(x_i \mid \theta) = \text{Exponential}(x_i \mid \theta) = \theta \exp(-\theta x_i)$ for life times $x = (40.6, 12.5, 1.0, 16.0, 16.1, 12.2, 3.3, 19.6, 2.1, 0.5)$,
and prior $\pi(\theta) = \text{Gamma}(0.001, 0.001)$. Write the BUGS code with full likelihood coded in the form

$$L(\theta; \text{full data}) = \pi(x_1, \ldots, x_{10} \mid \theta) = \prod_{i=1}^{10} \text{Exponential}(x_i \mid \theta)$$

and alternatively in the form exploiting sufficient statistics $S = \sum x_i = 123.9$:

$$L(\theta; \text{full data}) = \pi(S \mid \theta) = \text{Gamma}(S \mid 10, \theta)$$

The data list could then be `list(S=123.9,x=c(40.6,12.5,1.0,16.0,16.1,12.2,3.3,19.6,2.1,0.5))`. Run the two models and check you get the same posterior distribution for θ in both ways. (The BUGS syntax for gamma and exponential distributions are `dgamma(,)` and `dexp()`). *As a theory background: according to probability theory, if $X_i \sim$ Exp(θ), it is the same as $X_i \sim$ Gamma$(1, \theta)$, and moreover: $\sum_{i=1}^{n} X_i \sim$ Gamma(n, θ).*

4. A machine fails if **one or both** of the events, A and B occur. The event probabilities are unknown $P(A) = p_1$ and $P(B) = p_2$ and without further knowledge of them, we choose Uniform(0,1) prior distribution for both (independently). In a sample of 700 machines, 33 failed. Write a BUGS model to estimate p_1 and p_2 from posterior distribution and explain your results discussing parameter identifiability. Then assume you had expert opinion which says that about 2 in 1000 machines fail due to event B (this could be interpreted to be equivalent to 'prior data' of sample size 1000). Include this expert knowledge in the analysis as a (beta)prior and recalculate. In both cases, plot 2D scatter plots from the posterior distribution of p_1, p_2 using `Inference -> Correlation Tool` from the BUGS menu. Explain why the 2D posterior takes the shape it does.