Homework #5

Instructor: Roshan Vengazhiyil, Brani Vidakovic   *Name:* Nick Korbit*, gtID:* 903263968

Problem 1

Let x be the total blood volume of normal newborn babies in whom the cord was clamped early. We then define y as the total blood volume of normal newborn babies in whom the cord was not clamped until the placenta began to descend. We model x and y as Gamma with non-informative $(0.01, 0.01)$ priors:

$$x \sim \mathcal{G}a\left(a_x, b_x\right)$$
$$y \sim \mathcal{G}a\left(a_y, b_y\right)$$
$$a_x, b_x, a_y, b_y \sim \mathcal{G}a\left(0.001, 0.001\right)$$

Given observed 16 data points we specify an OpenBUGS model as

```
for (i in 1:n) {
        x[i] ~ dgamma(a_x, b_x)
        y[i] ~ dgamma(a_y, b_y)
}

a_x ~ dgamma(0.001, 0.001)
b_x ~ dgamma(0.001, 0.001)
a_y ~ dgamma(0.001, 0.001)
b_y ~ dgamma(0.001, 0.001)
```

We also know that the mean of Gamma is α/β, so we add calculation of means and their difference to the model:

```
mean_x <- a_x/b_x
mean_y <- a_y/b_y
diff <- mean_x - mean_y
```

We initialize values for the priors as ones:

```
list(a_x=1, b_x=1, a_y=1, b_y=1)
```

Let's now run an OpenBUGS simulation. We burn the first 10000 observation and update the model with the next 100000 samples. First, we plot densities for means and their difference:
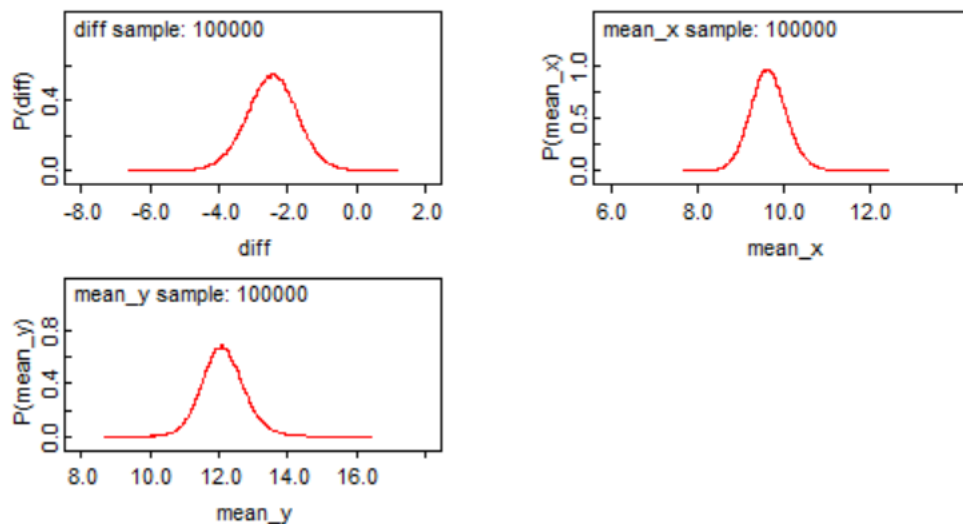
We notice that visually the difference is far from 0 and the means themselves are very different. Let's see the statistics for the observed variables.

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
a_x	35.56	12.56	0.4219	14.93	34.21	63.5	10001	100000
a_y	27.68	10.02	0.3183	11.82	26.42	50.46	10001	100000
b_x	3.689	1.312	0.04407	1.536	3.544	6.618	10001	100000
b_y	2.288	0.8355	0.02654	0.9662	2.183	4.18	10001	100000
diff	-2.465	0.7546	0.002682	-3.987	-2.454	-1.008	10001	100000
mean_x	9.66	0.4354	0.001577	8.837	9.648	10.56	10001	100000
mean_y	12.13	0.6187	0.002146	10.96	12.1	13.41	10001	100000

The 95% credible set for the difference in means is $(-3.987, -1.008)$, so that 0 is outside the set.

Note: the full OpenBUGS code is available at *babies0.odc* in the attached archive.

Problem 2

Let y be the variable we would like to predict – geographic location of wolves. We model y as

$$y_i \sim \mathcal{B}er\left(p_i\right)$$

$$logit\left(p_i\right) = \log \frac{p_i}{1 - p_i} = \alpha + \beta_{gender} x_{gender} + \beta_{x3} x_3 + \beta_{x7} x_7$$

For each of the coefficients we set a normal prior with $\mu = 0$ and $precision = 0.01$. So that the OpenBUGS model is

```
# Training
for (i in 1:n) {
        logit(p[i]) <- alpha + b.gender*gender[i] + b.x3*x3[i] + b.x7*x7[i]
        arctic[i] ~ dbern(p[i])
}

alpha ~ dnorm(0.0, 0.01)
b.gender ~ dnorm(0.0, 0.01)
b.x3 ~ dnorm(0.0, 0.01)
b.x7 ~ dnorm(0.0, 0.01)
```

We specify the initial values as 1 for α and zeros for other coefficients:

```
list(alpha=1, b.gender=0, b.x3=0, b.x7=0)
```

Knowning the characteristics of the new observation we proceed with inference:

```
# Inference
logit(p_new) <- alpha + b.gender*xgender + b.x3*xx3 + b.x7*xx7

DATA
list(n=25, xx3=5.28, xx7=1.78, xgender=1)
```

Let's now run an OpenBUGS simulation. We start with burning the first 10000 observation and update the model with the next 100000 samples. First, we plot density for the new point p_{new}:
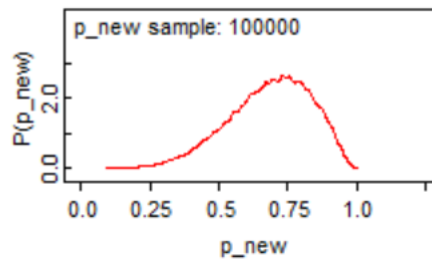
We notice that visually the prediction is more skewed towards $y = 1$ – the Arctic location. Let's investigate statistics:

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
p_new	0.6846	0.1489	7.172E-4	0.3595	0.6996	0.925	10001	100000

The mean of the posterior (our Bayes estimator) is 0.6846, so that we conclude that with probability $p = 0.6846$ our new observation point belongs to the Arctic location ($y = 1$). Notice also the relatively large 95% credible set bounds – $(0.3595, 0.925)$. So we cannot safely assume that the new point is from the Arctic.

Note: the full OpenBUGS code is available at *wolves0.odc* in the attached archive.

Problem 3

Let's define y as the number of micronuclei and x as the dose (in Gy). We model y as a Poisson regression:

$$y_i \sim \mathcal{P}oi\left(\lambda_i\right)$$
$$\log\left(\lambda_i\right) = \beta_0 + \beta_1 x_i$$

For each of the coefficients we set a normal prior with $\mu = 0$ and *precision* $= 0.01$. So that the OpenBUGS model is

```
# Training
for (i in 1:n)
{
        y[i] ~ dpois(lambda[i])
        lambda[i] <- exp(beta0 + beta1 * x[i])
}

beta0 ~ dnorm(0, 0.01)
beta1 ~ dnorm(0, 0.01)
```
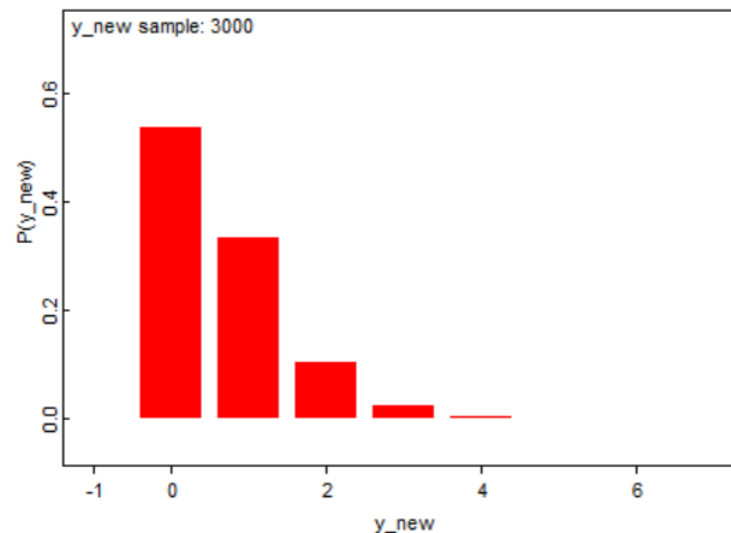
Knowning the characteristics of the new observation we proceed with inference:

```
# Inference
lambda_new <- exp(beta0 + beta1 * xdose)
y_new ~ dpois(lambda_new)
}

DATA
list(n=6000,  xdose=3.5)
```

Let's now run an OpenBUGS simulation. We start with burning the first 500 observation and update the model with the next 3000 samples. First, we plot density for the new point y_{new}:

We notice that the probability mass is skewed towards $n = 0$ and $n = 1$ for dose of 3.5 Gy. Let's investidate statistics.

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
beta0	-2.826	0.06333	0.007494	-2.951	-2.821	-2.71	501	3000
beta1	0.6748	0.01904	0.002226	0.6439	0.6738	0.7123	501	3000
lambda_new	0.6291	0.01624	7.871E-4	0.6	0.6283	0.6642	501	3000
y_new	0.6287	0.8046	0.01486	0.0	0.0	3.0	501	3000

We conclude that the average number of micronuclei for dose of 3.5 Gy is 0.6287. That's our Bayes estimator (the mean of the posterior distribution).

Note: the full OpenBUGS code is available at *micronuclei0.odc* in the attached archive.

References

[1] Engineering Biostatistics: An Introduction using MATLAB and WinBUGS. Brani Vidakovic - Wiley Series in Probability and Statistics.