Probit regression
Example 6.2.1 from Press (2003)

• Occurrence or nonoccurrence of infection following birth by Ceasarian section.

• Response variable: $y = 1$ if caesarian birth resulted in an infection; $y = 0$ otherwise.

• Explanatory variables

$x_1 = 1$: caesarian was nonplanned
$x_2 = 1$: risk factors were present at birth
$x_3 = 1$: antibiotics were given as a prophylaxis

• Model: For $i = 1, \ldots, n = 251$

$$
\begin{aligned}
y_i | \theta_i &\sim Bern(\theta_i) \\
\theta_i &= \Phi\left(\beta_0 + \sum_{j=1}^{3} \beta_j x_{ij}\right)
\end{aligned}
$$

• Prior: $\beta \sim N(0, 5I_4)$.

- Random walk Metropolis-Hastings

$$\beta^* \sim N(\beta^{(j-1)}, \tau^2 V)$$

where $\tau^2 = 1$ is the tuning parameter and V is obtained by inverting the negative of the Hessian matrix (the matrix of second derivatives) of the log-likelihood function evaluate at

$$\widehat{\beta} = (-1.09, 0.61, 1.20, -1.90)$$

- MCMC setup

$M_0 = 1000$, $M = 1000$ and step= 20.

- Posterior summary

```
-------------------------------------------
beta      Mean    s.d.    Lower    Upper
-------------------------------------------

beta1   -1.080   0.213   -1.514   -0.685
beta2    0.603   0.238    0.134    1.069
beta3    1.177   0.256    0.708    1.689
beta4   -1.897   0.262   -2.403   -1.400
-------------------------------------------
```

- Predictive densities

$p(data|model)$ is computed, by MC integration, based a sample of size 10000 from the prior distributions. Then, assuming that $Pr(model)$ is uniform across models,

$$Pr(model|data) \propto p(data|model).$$

```
----------------------------
Model          Pr(model|data)
----------------------------
x1 x2 x3               73.01
x2 x3                  26.98
x1 x3                   0.01
x1 x2                   0.00
x1                      0.00
x2                      0.00
x3                      0.00
----------------------------
```
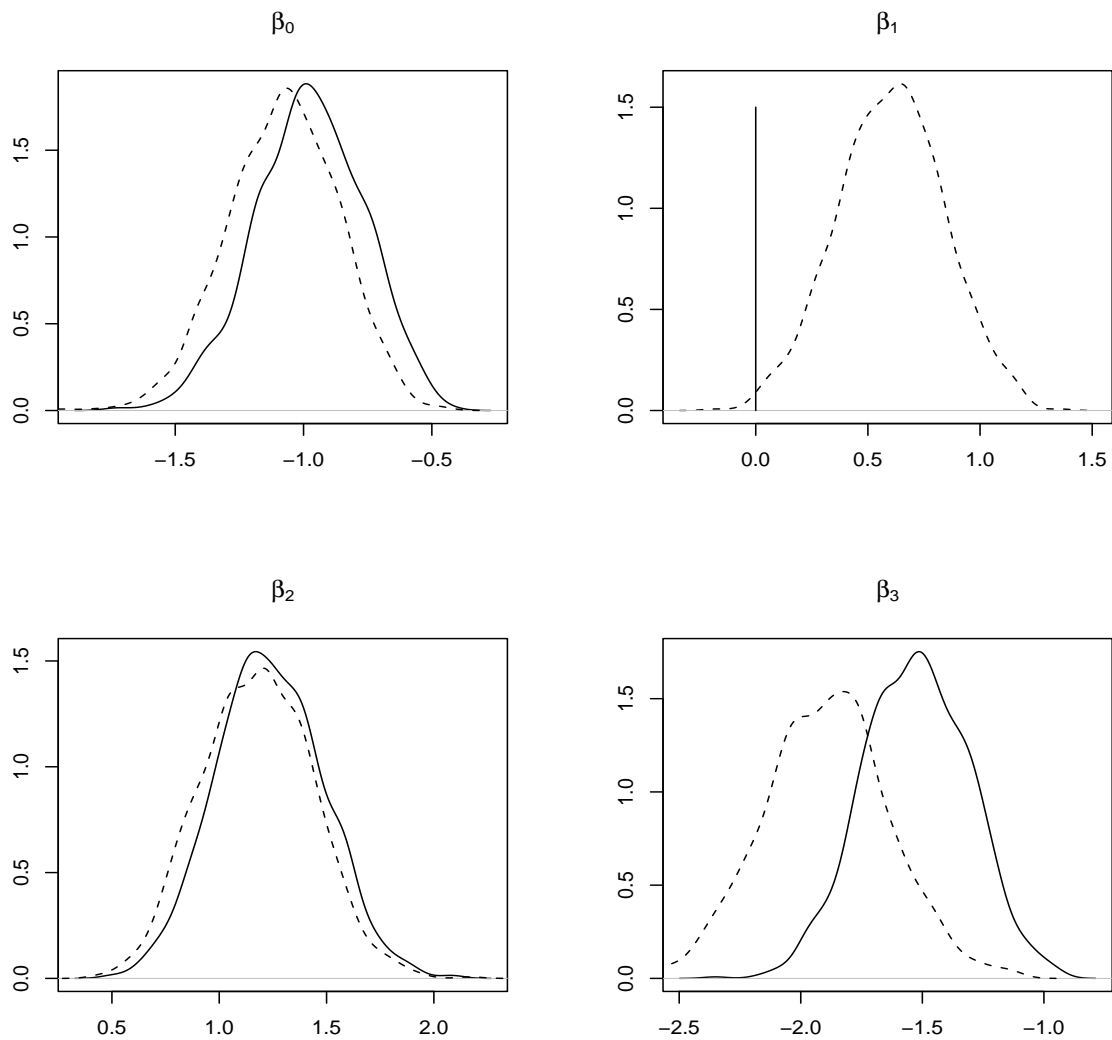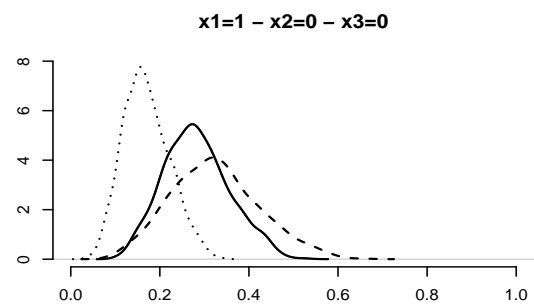
- Posterior summary (complete model)

```
------------------------------------------
beta      Mean     s.d.     Lower     Upper
------------------------------------------
beta1   -1.080    0.213    -1.514    -0.685
beta2    0.603    0.238     0.134     1.069
beta3    1.177    0.256     0.708     1.689
beta4   -1.897    0.262    -2.403    -1.400
------------------------------------------
```
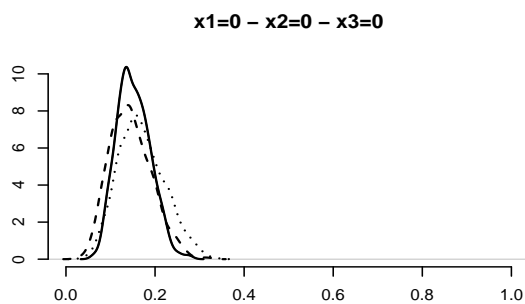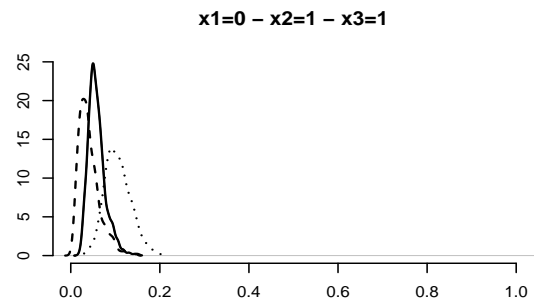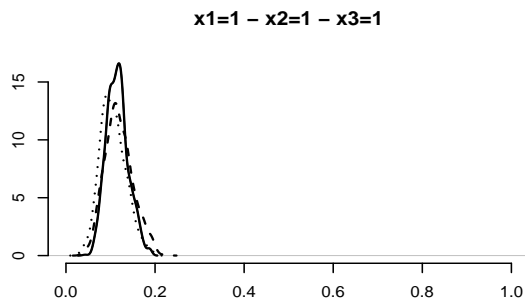
- Posterior summary (reduced model)

```
------------------------------------------
beta      Mean     s.d.     Lower     Upper
------------------------------------------
beta0   -0.977    0.211    -1.406    -0.583
beta2    1.235    0.249     0.756     1.730
beta3   -1.526    0.217    -1.956    -1.111
------------------------------------------
```

Posterior densities: complete model (solid line) and reduced model (dashed line).
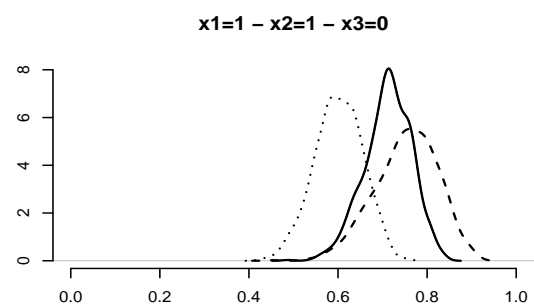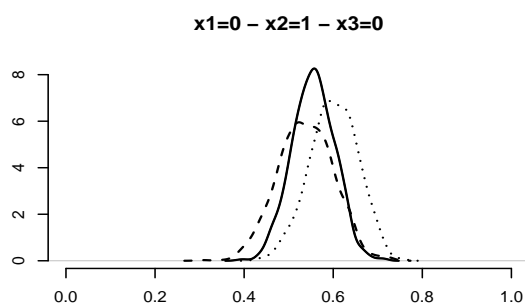
Posterior predictive densities: $Pr(y = 1|x)$, for several configuration of x. Complete model (dashed line), reduced model (dotted line) and Bayesian model averaging (solid line).

x1=1 – caesarian nonplanned
x2=1 – risk factors present birth
x3=1 – antibiotics given

	x1=1 x2=1 x3=1
	x1=0 x2=1 x3=1
	x1=0 x2=0 x3=0
	x1=1 x2=0 x3=0
	x1=0 x2=1 x3=0
	x1=1 x2=1 x3=0
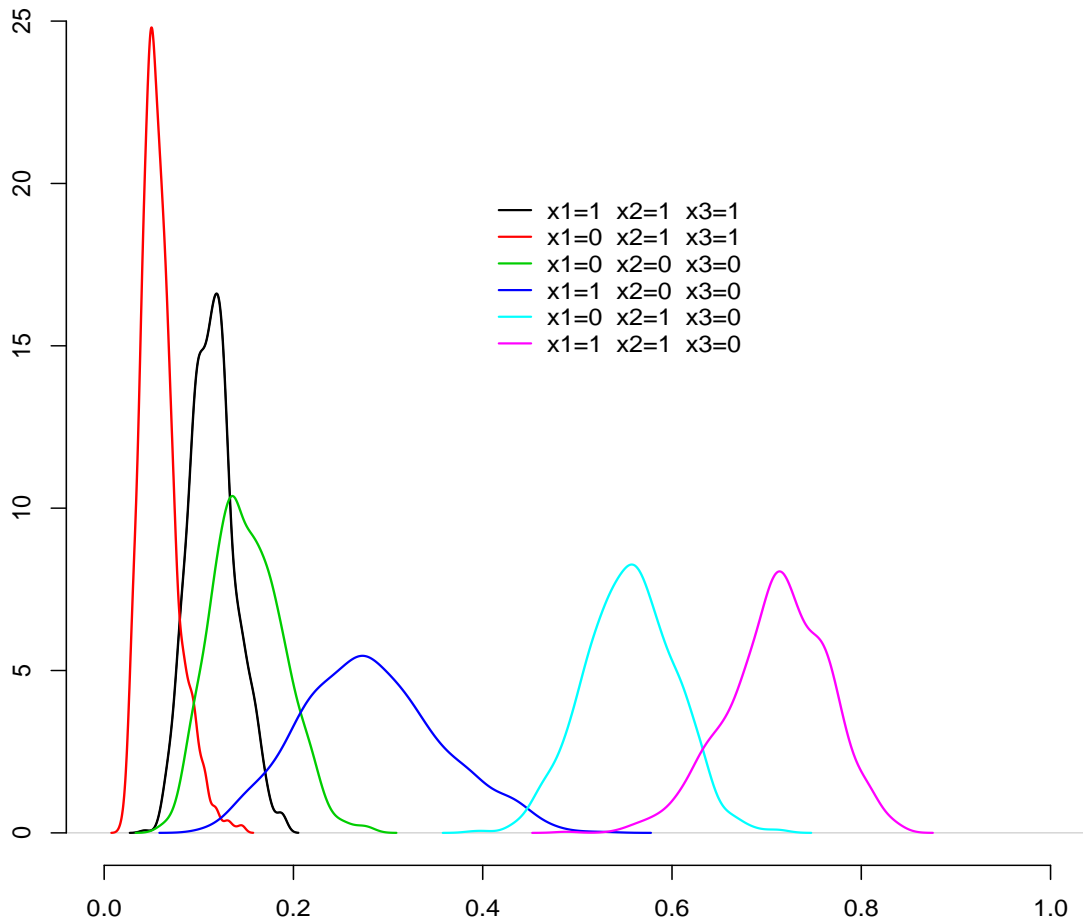
Bayesian model averaging: posterior predictive densities, $Pr(y = 1|x)$, for several configuration of x.