1 Turnout: logit and probit models for binary data

Nagler (1991) analyzes voter turnout with individual level data from the 1984 Current Population Survey. Covariates include each survey respondent's age and a categorical measure of their level of education. In addition, measures of each respondent's political context are available: the number of days registration closes before the election, whether or not a gubernatorial election took place in the respondent's state, and whether the respondent lives in the South. This re-analysis uses a random 3,000 observation subset of Nagler's original 98,000 observation data set.

A probit model is simply

$$
\begin{aligned}
y_i &\sim \text{Bern}(p_i) \\
\Phi^{-1}(p_i) &= \mathbf{x}_i\boldsymbol{\beta}
\end{aligned}
$$

while a logit model is

$$
\begin{aligned}
y_i &\sim \text{Bern}(p_i) \\
\ln\left(\frac{p_i}{1 - p_i}\right) &= \mathbf{x}_i\boldsymbol{\beta}
\end{aligned}
$$

Estimating the binary response model is easily accomplished using the Gibbs sampler in WinBUGS:

```
model{
  for (i in 1:N){                       ## loop over observations
    y[i] ~ dbern(p[i]);                 ## binary outcome
    logit(p[i]) <- ystar[i];            ## logit link
    ystar[i] <- beta[1]                 ## regression structure for covariates
              + beta[2]*educ[i]
              + beta[3]*(educ[i]*educ[i])
              + beta[4]*age[i]
              + beta[5]*(age[i]*age[i])
              + beta[6]*south[i]
              + beta[7]*govelec[i]
              + beta[8]*closing[i]
              + beta[9]*(closing[i]*educ[i])
              + beta[10]*(educ[i]*educ[i]*closing[i]);

    llh[i] <- y[i]*log(p[i]) + (1-y[i])*log(1-p[i]);   # llh contributions
  }

  sumllh <- sum(llh[]);       # sum of log-likelihood contributions

  ## priors
  beta[1:10] ~ dmnorm(mu[] , B[ , ]) ;     # diffuse multivariate Normal prior
                                           # see data file
}
```

Starting values were drawn from an OLS regression of the observed binary dependent variable on the covariates, and a diffuse multivariate Normal prior employed for the $\boldsymbol{\beta}$. Note also the calculation of the likelihood, demonstrating that any function of the parameters may be calculated and stored at each iteration of the Gibbs sampler; the co-sponsor example highlights this further.

1.1 Probit model by truncated normal sampling

In an early demonstration of the Gibbs sampler, Albert and Chib (1993) showed how truncated Normal sampling could be used to implement the Gibbs sampler for a probit model for binary responses. The intuition behind this approach is to see that the probit model is simply a regression where the dependent variable y_i^* is observed only in terms of its sign: i.e.,

$$y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \sim N(0,1)\ \forall\, i = 1,\ldots,n, \tag{1}$$

and

$$\begin{aligned} y_i = 0 &\Rightarrow y_i^* < 0, \\ y_i = 1 &\Rightarrow y_i^* \geq 0. \end{aligned}$$

Setting the censoring threshold at zero is arbitrary --- any non-zero threshold will be offset by a corresponding shift in the intercept. Also, the restriction that y_i^* has its scale parameter σ^2 set to 1.0 is an arbitrary identification restriction, to overcome the scale invariance problem common to many latent variable models. To see this, consider multiplying both sides of equation (1) by any $c > 0$; $\theta = (\boldsymbol{\beta}, 1)$ fits the data as well as $\theta^* = (c\boldsymbol{\beta}, c^2), \forall\, c > 0$. Setting σ^2 to any positive constant will pin down the latent variable (and hence $\boldsymbol{\beta}$), and 1.0 is simply chosen for convenience.

The Gibbs sampler for the probit model is defined by the following set of conditional distributions:

$$\begin{aligned} y_i^*|(y_i = 0, \mathbf{x}_i, \boldsymbol{\beta}) &\sim N(\mathbf{x}_i\boldsymbol{\beta}, 1)\, I(y_i^* < 0) \quad \text{(trunc. Normal)} \quad (2) \\ y_i^*|(y_i = 1, \mathbf{x}_i, \boldsymbol{\beta}) &\sim N(\mathbf{x}_i\boldsymbol{\beta}, 1)\, I(y_i^* \geq 0) \quad \text{(trunc. Normal)} \quad (3) \\ \boldsymbol{\beta}|\mathbf{y}^*, \mathbf{X}, \mathbf{y} &\sim N(\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{B}}), \quad (4) \end{aligned}$$

where

$$\begin{aligned} \tilde{\boldsymbol{\beta}} &= (\mathbf{B}^{-1}_{\text{prior}} + \mathbf{X}'\mathbf{X})^{-1}(\mathbf{B}^{-1}_{\text{prior}}\,\boldsymbol{\beta}_{\text{prior}} + \mathbf{X}'\mathbf{y}^*) \\ \tilde{\mathbf{B}} &= (\mathbf{B}^{-1}_{\text{prior}} + \mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

These last two expressions are simply those for the posterior mean and posterior covariance of regression parameters; the posterior mean $\tilde{\boldsymbol{\beta}}$ is the matrix-weighted average of the estimate of $\boldsymbol{\beta}$ from the data and the prior mean, where the matrix weights are the respective "precision matrices" (inverted covariance matrices) of the prior and the data. Note that with an uninformative prior $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}^*$ and $\tilde{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}$ (i.e., the posterior moments are given by simply running a regression of $\mathbf{y}^*$ on $\mathbf{X}$). Accordingly, iteration t of the Gibbs sampler consists of the following steps:

1. Sample $y_i^{*(t)}$ from respective truncated Normals in (2) and (3). This easily accomplished either by sampling from (untruncated) Normal distribution and rejecting and re-sampling draws that fail to meet the truncation constraints, although a more computationally efficient strategy is to use the "probability inverse transform" algorithm of Devroye (1986), described by Gelfand et al. (1990, 977) or Greene (1997, 179).

2. Sample $\boldsymbol{\beta}^{(t)}$ from the multivariate Normal in (4)

This scheme is easily implemented in WinBUGS, using the `I()` construct and the observed y_i to place the necessary bounds on the latent y_i^*:

Unlike the previous program, y_i is not a stochastic node, and so "automatic" imputations for any missing data on the y_i will *not* be made. We also require start values for the latent y^*; unless these are provided, WinBUGS will attempt to generate them itself, which usually takes an extremely long time. I set start values for y^* according to the following scheme: $y_i^* = 1 \iff y_i = 1, y_i^* = -1 \iff y_i = 0$ and $y_i^* = 0 \iff y_i = \text{NA}$.

Table 1 presents a comparison of the MLEs obtained using the `glm` procedure in Splus and the output of the Gibbs sampler.

The simple WinBUGS programs here should be compared with the relatively longer and more complicated Splus programs necessary to implement the Gibbs sampler for even this simple model.

References

Albert, James H. and Siddhartha Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88:669--79.

Devroye, Luc. 1986. *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.

Gelfand, Alan E., Susan E. Hills, Amy Racine-Poon and Adrian F. M. Smith. 1990. "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling." *Journal of the American Statistical Association* 85:972--985.

Greene, William H. 1997. *Econometric Analysis*. Third ed. New York: Prentice-Hall.

Nagler, Jonathan. 1991. "The Effect of Registration Laws and Education on U.S. Voter Turnout." *American Political Science Review* 85:1393--1405.

	MLE	*MCMC*
Intercept	-2.32	-2.34
	(.56)	-
	[-3.24, -1.40]	[-3.26, -1.43]
Education	.096	.11
	(.22)	-
	[-.26, .45]	[-.25, .46]
Education^2	.021	.020
	(.022)	-
	[-.015, .057]	[-.016, .056]
Age	.067	.067
	(.008)	-
	[.054, .079]	[.054, .079]
Age^2	-.00047	-.00047
	(.00008)	-
	[-.00061, -.00034]	[-.00061, -.00034]
South	-.094	-.095
	(.061)	-
	[-.19, .007]	[-.19, .006]
Gubernatorial Election	.065	.064
	(.066)	-
	[-.044, .17]	[-.046, .17]
Closing Day	-.021	-.020
	(.020)	-
	[-.053, .012]	[-.053, .012]
Education	.0063	.0061
$\times$ *Closing Day*	(.0081)	-
	[-.0071, .020]	[-.0073, .020]
Education^2	-.00061	-.00059
$\times$ *Closing Day*	(.00082)	-
	[-.0020, .00074]	[-.0020, .00076]

Table 1: **Comparison of MLEs and Gibbs sampler output, probit model of voter turnout.** Standard errors appear in parentheses for the MLEs. For the Gibbs sampler output, the mean of the last 5,000 samples is reported as the point estimate, no standard error is reported, and a 90% confidence interval is reported in brackets; the 90% confidence interval implied by the MLEs point estimate and standard error (assuming asymptotic Normality) is also reported in brackets.