Hierarchical Bayesian Analysis - Hierarchical logit and models of bounded rationality

Alina Ferecatu

Rotterdam School of Management,
Erasmus University

EMAC

May 31st, 2018

Overview of models

- A choice model of buying decisions given price and promotions

  Hierarchical binary logit example

- A bounded rationality model

  Adaptive learning example

Hierarchical binary logit example

A choice model of buying decisions given price and promotions:

- 500 consumers with 10 purchase occasions each

$$y_{ijt} \sim \mathcal{B}(p_{ijt})$$

$$logit(p_{ijt}) \sim \mathcal{N}(x_{ijt}\beta_{ij})$$

$$\beta_i \sim MultiNormal(z_i\delta, \Sigma)$$

$$\delta \sim \mathcal{N}(\delta_0, \sigma)$$

Non-conjugate prior

$$\Sigma = diag(\tau)\,\Omega\,diag(\tau)$$

$$\tau \sim gamma(a, b)$$

$$\Omega \sim LKJcorr(\nu)$$

```
data {
int<lower=1> nvar; // number of parameters in the logit regressio
int<lower=0> N; // number of observations
int<lower=1> nind; // number of individuals
int<lower=0,upper=1> y[N];
int<lower=1,upper=nind> ind[N]; // indicator for individuals
row_vector[nvar] x[N];
}

parameters {
vector[nvar] delta;
vector<lower=0>[nvar] tau;
vector[nvar] beta[nind];
corr_matrix[nvar] Omega; // Vbeta - prior correlation
}

model {
to_vector(delta) ~ normal(0, 5);
to_vector(tau) ~ gamma(2, 0.5);
Omega ~ lkj_corr(2);

for (h in 1:nind)
beta[h]~multi_normal(delta, quad_form_diag(Omega, tau));

for (n in 1:N)
y[n] ~ bernoulli_logit(x[n] * beta[ind[n]]);
}
```

Noncentered (Re)Parameterization - the "Matt Trick"

Consider a model case with a diagonal variance-covariance matrix

- Assume our intercept model: $\beta_i \sim \mathcal{N}(\delta, \tau)$
- We can decompose that into: $\mathcal{N}(\delta, \sigma) \stackrel{d}{=} \delta + \tau\mathcal{N}(0, 1)$
- The trick applies to other distributions in the location-scale family
- The transformation:
  1. declare α_i in the parameters block and β_i in the transformed parameters block
  2. draw $\alpha_i \sim \mathcal{N}(0, 1)$ & $\tau \sim gamma(a, b)$
  3. compute $\beta_i = \delta + \tau\alpha_i$

Noncentered (Re)Parameterization - the multivariate case

- Assume our intercept model: $\beta_{\mathbf{i}} \sim MultiNormal(\delta, \Sigma)$
  - If Σ_{kk} is small, then β_{ik} needs to fall into a small range, NUTS needs a small step size
  - If Σ_{kk} is large, then β_{ik} can fall into a wide range, NUTS needs a large step size/ lots of small steps
- The transformation:
  1. declare α_i in the parameters block and β_i in the transformed parameters block
  2. draw $\alpha_i \sim \mathcal{N}(0, 1)$ & $\tau \sim gamma(a, b)$
  3. compute $\beta_{\mathbf{i}} = \delta + \tau \mathbf{L} \alpha_{\mathbf{i}} \sim \mathcal{N}(\delta, \tau^2 \mathbf{L}\mathbf{L}^T)$
  4. where $\tau \mathbf{L}$ is the Cholesky factor of $\Sigma = \tau^2 \mathbf{L}\mathbf{L}^T$, and τ is the standard deviation of the errors.

Noncentered (Re)Parameterization - the implementation

$$y_{ijt} \sim \mathcal{B}(p_{ijt})$$

$$logit(p_{ijt}) \sim \mathcal{N}(x_{ijt}\beta_{ij})$$

$$\beta_{\mathbf{i}} = \delta + \tau \mathbf{L}\alpha_{\mathbf{i}}$$

$$\alpha_i \sim \mathcal{N}(0, 1)$$

$$\delta \sim \mathcal{N}(\delta_0, \sigma)$$

Non-conjugate prior

$$\tau \sim gamma(a, b)$$

$$\Omega \sim LKJcorr(\nu)$$

```
data {
.....
}

parameters {
matrix[nvar, nind] alpha; // nvar*H parameter matrix
row_vector[nvar] delta;
vector<lower=0>[nvar] tau;
cholesky_factor_corr[nvar] L_Omega;
}


transformed parameters{
row_vector[nvar] beta[nind];
matrix[nind,nvar] Vbeta_reparametrized;
Vbeta_reparametrized = (diag_pre_multiply(tau, L_Omega)*alpha)';

for (h in 1:nind)
beta[h]=delta+Vbeta_reparametrized[h];
}

model {
......
}
```

Choice of prior distribution of the variance components

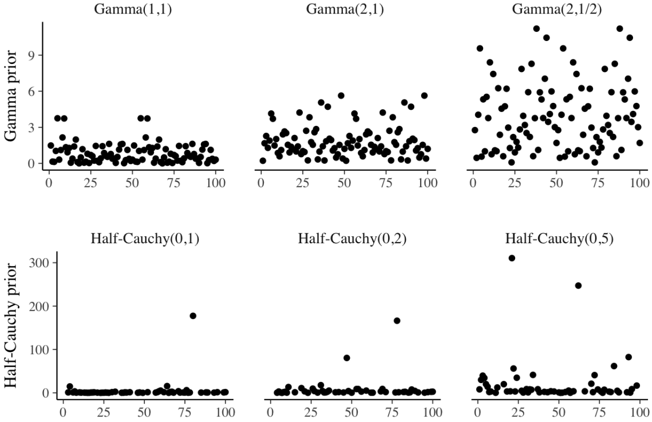

Figure 1: Choice of prior for the variance τ

Traceplots of model parameters

The noncentered reparametrization helps tremendously

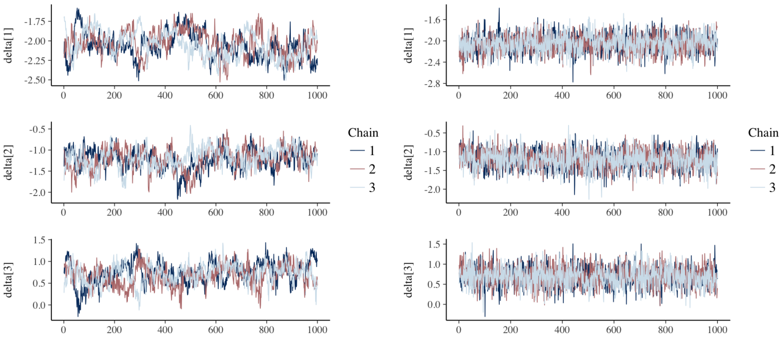

Figure 2: Traceplot of δ parameters (after burnin) under the centered *(left)* vs. the noncentered reparametrization *(right)*, using package *bayesplot*

Summary statistics

Effective sample size and convergence properties

```
$summary
                mean     se_mean        sd       2.5%      97.5% n_eff     Rhat
delta[1] -2.0464282 0.003151008 0.1725878 -2.3917852 -1.7207339  3000 1.001876
delta[2] -1.2073192 0.004632004 0.2537053 -1.6928554 -0.7322237  3000 1.002893
delta[3]  0.6915687 0.004494100 0.2461520  0.2134696  1.1791871  3000 1.001500
```

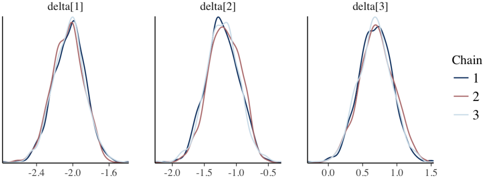

Figure 3: Density plot of δ parameters (after burnin) under the noncentered reparametrization, using package *bayesplot*

Individual level parameters

Most parameters are within the 95% highest density intervals

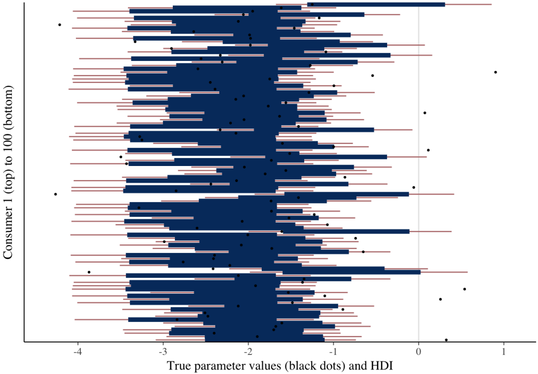

Figure 4: True values (black dots) and the 80% and 95% highest density intervals for the intercept, for the first 100 consumers

Model checking

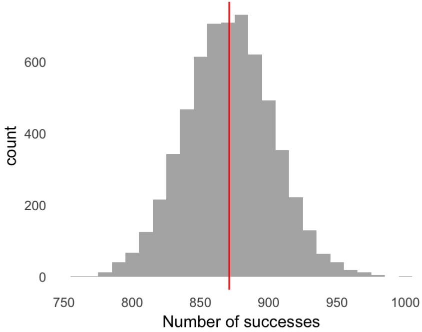

Figure 5: Number of successes: posterior replications vs. true value

Model checking

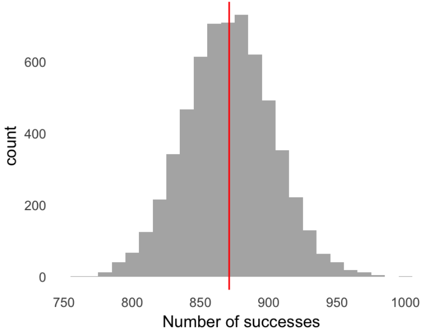

Figure 5: Number of successes: posterior replications vs. true value

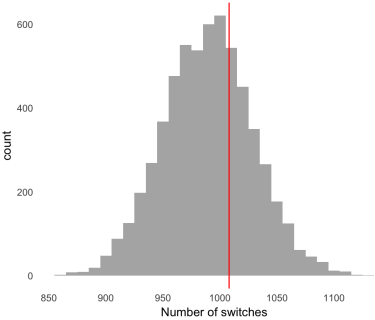

Figure 6: Switches between buying/ not buying: posterior replications vs. true value

Compute hit rates and MSEs based on posterior replications

Model comparison

Likelihood-based measures: *Leave-one-out cross-validation*

Table 1: Model comparison based on LOO-CV, using package *loo*

	Variance model		*Full covariance model*		*NCP model*	
	Estimate	SE	Estimate	SE	Estimate	SE
elpd_loo	-1874.3	43.1	-1871.9	43.2	-1871.2	43
p_loo	398.6	12.5	364.4	11.8	363.7	11.8
looic	3748.6	86.2	3743.7	86.4	3742.4	86.5

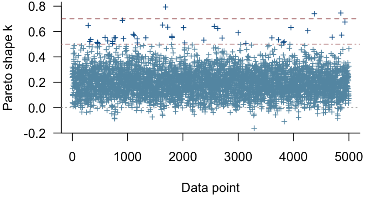

Figure 7: PSIS diagnostic plot, using package *loo*


Thank
you