Problem Set 1: Computing dynamic discrete choice models

Rachel Anderson

March 1, 2019

1 Setup

Let $\{d_{it}, x_{it}\}_{i=1,\ldots,N,\ t=1\ldots,T}$ denote the decision (d) and state (x) associated with bus i in period t. Notably, $d_{it} = 1$ if the decisionmaker replaces the bus engine in period t, and $x_{it} \in X = \{0, 1, \ldots, 20\}$ refers to bus mileage. In addition to the observables d_{it}, x_{it}, and the identity of the bus, there is an unobservable state variable $s_i \in \{1, 2\}$ which is the type of bus i. Type 1 is the type that is cheaper to replace. The goal of this analysis is to estimate the planner's willingness to pay to replace the expensive buses with the inexpensive type of bus.

2 Likelihood

The likelihood contribution of the data associated with a single bus $\{x_{it}, d_{it}\}_{t=1}^T$ is

$$\mathcal{L}(x_{i1}, \ldots, x_{iT}, d_{i1}, \ldots, d_{iT}|x_0, d_0, \theta) = \prod_{t=1}^{T} P(d_{it}|x_{it}, \theta)F(x_{it}|x_{it-1}, d_{it-1})$$

Taking logs,

$$\ell_i(\theta) \equiv \ell(x_{i1}, \ldots, x_{iT}, d_{i1}, \ldots, d_{iT}|x_0, d_0, \theta) = \sum_{t=1}^{T} \log F(x_{it} \mid x_{it-1}, d_{it-1}; \theta_F) + \sum_{t=1}^{T} \log P(d_{it} \mid x_{it}; \theta)$$

We are given $\theta_F = (\Gamma_{nr}, \Gamma_r)$ so that $F(x_{it} \mid x_{it-1}, d_{it-1}, \theta_F)$ does not depend on the unknown $\theta = (p_1, b, c)$; hence, the first term drops out of the likelihood.

$$\ell_i(\theta) = \sum_{t=1}^{T} \log P(d_{it} \mid x_{it}; \theta)$$

Using $p_1 = P(s_i = 1)$, the marginal probability that bus i is Type 1, we use the law of total probability to write:

$$P(d_{it} \mid x_{it}, \theta) = p_1 P(d_{it}|x_{it}, s_i = 1) + (1 - p_1)P(d_{it} \mid x_{it}, s_i = 2)$$

where

$$\begin{aligned}
p(d_{it} = 2|x_{it}, s_i, \theta) &= p(u_{i,t,r} > u_{i,t,nr} \mid x_{it}, s_i, \theta) \\
&= p(v(2, x_{it}, s_i) - v(1, x_{it}, s_i) \geq \epsilon_{i,t,nr} - \epsilon_{i,t,r}) \\
&= F_\epsilon\left(v(2, x_{it}, s_i) - v(1, x_{it}, s_i)\right) \\
&= \frac{\exp(v(2, x_{it}, s_i) - v(1, x_{it}, s_i))}{1 + \exp(v(2, x_{it}, s_i) - v(1, x_{it}, s_i))}
\end{aligned}$$

since the difference of two Gumbel i.i.d random variables follows a logistic distribution. Similarly,

$$Pr(d_{it} = 1|x_{it}, s_i, \theta) = 1 - p(d_{it} = 2|x_{it}, s_i, \theta)$$

Combining all of these derivations, we write the log likelihood contribution of a single bus i:

$$\sum_{t=1}^{T} \log \Big\{ p_1 \Big(Pr(d_{it} = 2\ |x_{it}, s_i = 1, \theta)^{\mathbb{1}(d_{it}=2)} (1 - Pr(d_{it} = 2|x_{it}, s_i = 1, \theta)^{1-\mathbb{1}(d_{it}=2)} \Big)$$
$$+ (1 - p_1) \Big(Pr(d_{it} = 2\ |x_{it}, s_i = 2, \theta)^{\mathbb{1}(d_{it}=2)} (1 - Pr(d_{it} = 2|x_{it}, s_i = 2, \theta)^{1-\mathbb{1}(d_{it}=2)} \Big) \Big\}$$

and the full likelihood of the sample os

$$\mathcal{L}(\theta) = \sum_{i=1}^{N} \ell_i(\theta)$$

3 Summary of results

Table 1 displays the parameter estimates and runtime (in seconds) of each algorithm. Table 2 shows the estimates of

$$\frac{V_1[0] - V_2[x]}{|c[1]|} \qquad x = 1, \ldots, 20$$

implied by the estimates of each procedure. The point estimate reported for the Bayesian DDC procedure is the posterior mean of 10,000 draws (with 3,017 accepted and 100 draw burn-in).

Table 1: Parameter estimates and runtime for various estimation procedures

Method	Time (seconds)	$\hat{\theta} = (p_1, b_1, b_2, c_1, c_2)$
NFXP	13.25	(0.433, −0.864, −1.052, −11.191, −13.797)
EM	26.59	(0.415, −0.827, −1.072, −10.845, −13.541)
2-Step EM	7.333	(0.415, −0.815, −1.072, −10.688, −13.530)
Bayesian DDC*	40.77	(0.415, −0.816, −0.79, −10.68, −14.0372)

4 Notes on Implementing Bayesian DDC

For the Metropolis-Hastings Step, I tried using Gamma and Normal distributions centered on the last accepted value of θ as my proposal densities. I also tried using a Uniform distribution or Beta(1,1) for my jump distribution for p_1. For simplicity, however, I decided on the following specification:

1. Across all draws I fix $p_1 = p_1^{MLE}$, and focus on drawing from the posterior of β.

2. I use independent normals as the proposal distribution for each β_k with the following parameters:

$$g(\beta_k'|\beta_k) \sim \mathcal{N}\left(\beta_k, \frac{1}{100}\right)$$

Table 2: Estimated willingness to pay to switch bus type by state

State	NFXP	EM	2-Step EM	Bayesian
1	0.336	0.333	0.340	0.314
2	0.609	0.591	0.619	0.566
3	0.820	0.778	0.836	0.758
4	0.977	0.912	1.000	0.898
5	1.091	1.014	1.121	1.003
6	1.169	1.092	1.205	1.088
7	1.214	1.146	1.253	1.158
8	1.236	1.176	1.276	1.214
9	1.244	1.190	1.284	1.256
10	1.247	1.196	1.287	1.283
11	1.248	1.198	1.288	1.298
12	1.249	1.199	1.289	1.305
13	1.249	1.199	1.289	1.309
14	1.249	1.199	1.289	1.311
15	1.249	1.199	1.289	1.312
16	1.249	1.199	1.289	1.312
17	1.249	1.199	1.289	1.312
18	1.249	1.199	1.289	1.312
19	1.249	1.199	1.289	1.312
20	1.249	1.199	1.289	1.312

3. I use independent priors with a similar form:

$$\pi(\beta_k) \sim \mathcal{N}\left(\beta_k^{MLE}, \frac{1}{100}\right)$$

so that the prior is always centered around the MLE. I chose this specification because otherwise the chain would wander off and never return to the MAP estimate.

4. I use the Euclidean norm for my kernel weights when updating V, so that

$$K(\theta - \theta^{*(n)}) \equiv \sum_i \sqrt{(\beta_1 - \beta_1^*)^2 + (\beta_2 - \beta_2^*)^2 + \cdots + (\beta_4 - \beta_4^*)^2}$$

in the formula

$$\hat{E}^{(m)} V(x', t, \epsilon') = \sum_{n=m-N(t)}^{m} V^{(n)}(x', t, \epsilon', \theta^{*(n)}) \frac{K(\theta - \theta^{*(n)})}{\sum_{n=m-N(t)}^{m} K(\theta - \theta^{*(n)})})$$

I use 10,000 draws and keep track of the 50 past values of $\hat{V}$.

5. I use a burn-in of 100 draws. Since I use independent jump distributions and priors, I don't expect much serial correlation in the draws. My main reason for using burn-in is so that the estimates of $\hat{V}$ are reasonably calibrated.