A Model of Consumer Learning for Service Quality and Usage

RAGHURAM IYENGAR, ASIM ANSARI, and SUNIL GUPTA

APPENDIX W: PRIORS AND FULL CONDITIONAL DISTRIBUTIONS

Markov chain Monte Carlo procedures are used for numerical Bayesian inference. The following set of full conditionals and priors for $\boldsymbol{\psi}_\omega, \boldsymbol{\Lambda}_\omega, \boldsymbol{\omega}_i, \{\lambda_j^2\}, \{\tau_{ij}^2\}, \{c_j\}, \{d_j\}, \mu, \delta^2, \{m_{it}\}$ and $\{\mathbf{EU}_{ijt}\}$ are used:

a) The full conditional for $\boldsymbol{\psi}_\omega$ is multivariate normal $N(\boldsymbol{\psi}, \mathbf{V}_\psi)$, where the precision matrix, $\mathbf{V}_\psi^{-1} = \mathbf{C}_\omega^{-1} + \sum_{i=1}^{I} \boldsymbol{\Lambda}_\omega^{-1}$ and $\boldsymbol{\psi} = \mathbf{V}_\psi[\mathbf{C}_\omega^{-1}\boldsymbol{\eta}_\omega + \sum_{i=1}^{I} \boldsymbol{\Lambda}_\omega^{-1}\boldsymbol{\omega}_i]$. I is the number of households. The prior for $\boldsymbol{\psi}_\omega$ is normal where the prior mean, $\boldsymbol{\eta}_\omega = 0$ and the prior covariance, $\mathbf{C}_\omega = \text{diag}(1000)$.

b) The full conditional for the precision matrix $\boldsymbol{\Lambda}_\omega^{-1}$ is a Wishart distribution $W(\mathbf{V}_\lambda, \varrho_\lambda)$, where $\mathbf{V}_\lambda = (\sum_{i=1}^{I}(\boldsymbol{\omega}_i - \boldsymbol{\psi}_\omega)(\boldsymbol{\omega}_i - \boldsymbol{\psi}_\omega)' + \varrho\mathbf{R})^{-1}$ and $\varrho_\lambda = \varrho + I$. The prior for $\boldsymbol{\Lambda}_\omega^{-1}$ is $W((\varrho\,\mathbf{R})^{-1}, \varrho)$ with $\varrho = N_p + 1$, where N_p is the dimension of $\boldsymbol{\omega}_i$ and $\mathbf{R}$ is identity.

c) Draws for $\boldsymbol{\omega}_i$ are obtained by using a random walk multivariate Metropolis-Hastings step. On the (m+1)th iteration, this involves generating a candidate $\boldsymbol{\omega}_i^c$ from a multivariate normal proposal density $N(\boldsymbol{\omega}_i^{(m)}, \Upsilon_i)$. The proposal density is centered on the old value of $\boldsymbol{\omega}_i^{(m)}$ from iteration m. The tuning constant is chosen to ensure rapid mixing of the chain. The generated candidate $\boldsymbol{\omega}_i^c$ is accepted with the following acceptance probability

$$\alpha(\boldsymbol{\omega}_i^{(m)}, \boldsymbol{\omega}_i^c) = \min\left\{1, \frac{L(\boldsymbol{\omega}_i^c)\varphi(\boldsymbol{\omega}_i^c \mid \boldsymbol{\psi}_\omega, \boldsymbol{\Lambda}_\omega)}{L(\boldsymbol{\omega}_i^{(m)})\varphi(\boldsymbol{\omega}_i^{(m)} \mid \boldsymbol{\psi}_\omega, \boldsymbol{\Lambda}_\omega)}\right\}.$$

where $\varphi()$ represents the normal density for the prior and L is the density for the likelihood. If the candidate is accepted, then $\boldsymbol{\omega}_i^{(m+1)} = \boldsymbol{\omega}_i^c$, otherwise, $\boldsymbol{\omega}_i^{(m+1)} = \boldsymbol{\omega}_i^{(m)}$. The parameters for the different individuals can be drawn in sequence.

d) As the utilities of the plans are assumed to be independent given $\boldsymbol{\omega}_i$, we can consider each λ_j^2 separately. We set an inverse gamma prior $IG(a, b)$. Then, the posterior is $IG(\frac{N}{2} + a, [\frac{1}{2}\sum_{i=1}^{I}\sum_{t=1}^{T_i}(EU_{ijt} - EU_{ijt}^m)^2 + b^{-1}]^{-1})$ where T_i is the number of observations of household i and N is the total number of observations. Here, EU_{ijt}^m is the mean of the expected utility function. We fixed a at 3.0 and b at 2.0.

e) The full conditional for τ_{ij}^2 is an inverse gamma distribution. We assume the heterogeneity distribution for τ_{ij}^2 to be independent for each plan j. We set an inverse gamma prior $IG(c_j, d_j)$. Then, the posterior is $IG(\frac{T_{ij}}{2} + c_j, [\frac{1}{2}\sum_{t=1}^{T_{ij}}(x_{it}^{actual} - E(x_{ijt}))^2 + d_j^{-1}]^{-1})$ where T_{ij} is the number of observations of consumer i where they chose plan j, x_{it}^{actual} is the actual quantity consumed by consumer i at time t under plan j and $E(x_{ijt})$ is the expected quantity.

f) Draws for c_j are obtained by using a random walk univariate Metropolis-Hastings step. We assume $c_j \sim IG(r_c, s_c)$. Let lc denote $\log(c_j)$. For the (m+1)th iteration, we then generate a candidate $lc^{(c)}$ from a univariate normal proposal density $N(lc^{(m)}, \tau_c^2)$. The proposal density is centered on the old value of $lc^{(m)}$ from iteration m and the variance of the proposal density, τ_c^2, is set to ensure rapid mixing of the chain. The generated candidate $lc^{(c)}$ is accepted with the following acceptance probability

$$\alpha(lc^{(m)}, lc^{(c)}) = \min\left\{1, \frac{L(e^{lc^{(c)}})IG(e^{lc^{(c)}} \mid r_c, s_c)}{L(e^{lc^{(m)}})IG(e^{lc^{(m)}} \mid r_c, s_c)}\right\}. \quad (1)$$

where $IG()$ represent the inverse gamma density for the prior and L is the density for the likelihood. If the candidate is accepted, then $lc^{(m+1)} = lc^{(c)}$, otherwise, $lc^{(m+1)} = lc^{(m)}$. We set r_c to 3.0 and s_c to 2.0 for the priors. The tuning constant τ_c^2 was set to 0.008.

g) The full conditional distribution for d_j is an inverse gamma distribution. We set an inverse gamma prior $IG(r_d, s_d)$. With this prior, the posterior is $IG(r_d + Ic_j, [\sum_{i=1}^{I} \tau_{ij}^{-2} + s_d^{-1}]^{-1})$ where I is the total number of households. We set $r_d = 3.0$ and $s_d = 2.0$.

h) To obtain the time-varying parameters (states) m_{it}, $i = 1$ to I, $t = 1$ to T_i, we use the forward-filtering backward-sampling algorithm, which consists of two steps. In a forward step, the moments of the updated distribution of each state is computed using a Kalman filter approach. To initialize the forward step, we set the prior for m_{i1} as N(0,100) and fix σ_{i1} to 1.0. In the backward step, each parameter is sampled from its conditional distribution conditioned on the preceding draw.

In the forward step, the moments of each state are computed recursively. Let the posterior at time $t-1$ be $p(m_{it-1} \mid D_{it-1}) \sim N(\hat{m}_{it-1}, \hat{C}_{it-1})$, where, D_{it-1} is the information set at time-period $t-1$ for consumer i and includes the utilities up to time period $t-1$ and the other unknowns. The prior for m_{it} can be written as $p(m_{it} \mid D_{it-1}) \sim N(w_{1it}\hat{m}_{it-1} + w_{2it}\mu, w_{1it}^2\hat{C}_{it-1} + w_{2it}^2\delta^2)$. After the data for period t is observed, conditional on the expected utilities for the observations in period t ($\tilde{\mathbf{EU}}_{it}$), the posterior at time t can be written as $p(m_{it} \mid D_{it-1}, \tilde{\mathbf{EU}}_{it}) \sim N(\hat{m}_{it}, \hat{C}_{it})$ by combining the prior and the likelihood. Hence, the posterior parameters are given by $\hat{C}_{it}^{-1} = (w_{1it}^2\hat{C}_{it-1} + w_{2it}^2\delta^2)^{-1} + \Sigma_{j=1}^{J}\lambda_j^{-2}$ and $\hat{m}_{it} = \hat{C}_{it}[(w_{1it}^2\hat{C}_{it-1} + w_{2it}^2\delta^2)^{-1}(w_{1it}\hat{m}_{it-1} + w_{2it}\mu) + \Sigma_{j=1}^{J}\lambda_j^{-2}\tilde{u}_{ijt}]$. Here, $\tilde{u}_{ijt}$ refers to the adjusted expected utilities. These are the expected utilities adjusted for quantity expectation, intercepts and the deterministic main effects. These posterior moments can be computed and stored in the forward step of the algorithm.

For describing the backward step, let $\tilde{m}_i = \{m_{i1}, m_{i2}, \ldots, m_{iT_i}\}$ and let the parameters that are not in $\tilde{m}_i$ be written as $\tilde{\boldsymbol{\alpha}}$. Then we can write

$p(\tilde{m}_i \mid \tilde{\mathbf{D}}_{iT}, \boldsymbol{\alpha}) = p(m_{iT} \mid \tilde{\mathbf{D}}_{iT}, \boldsymbol{\alpha}) p(m_{iT-1} \mid m_{iT}, \tilde{\mathbf{D}}_{iT-1}, \boldsymbol{\alpha}) \ldots p(m_{i1} \mid m_{i2}, \tilde{\mathbf{D}}_{i1}, \boldsymbol{\alpha})$ which depends upon the identity

$p(m_{iT-k} \mid m_{iT-k+1}, \tilde{D}_{iT}, \boldsymbol{\alpha}) = p(m_{iT-k} \mid m_{iT-k+1}, \tilde{D}_{iT-k}, \boldsymbol{\alpha})$. Therefore, $\tilde{m}_i$ can be sampled using the following steps:

- Sample m_{iT_i} from $N(\hat{m}_{iT_i}, \hat{C}_{iT_i})$

- For $t = T_i - 1, \ldots, 1$, sample m_{it} from $p(m_{it} \mid m_{it+1}, D_{it}, \tilde{\boldsymbol{\alpha}})$,

where, $(m_{it} \mid m_{it+1}, D_{it}, \tilde{\boldsymbol{\alpha}}) \sim N(q_{it}, Q_{it})$. The precision, $Q_{it}^{-1} = \hat{C}_{it}^{-1} + \frac{w_{1it}^2}{w_{2it}^2 \delta^2}$, and the mean

$q_{it} = Q_{it}[\hat{C}_t^{-1} \hat{m}_{it} + \frac{w_{1it}(m_{it+1} - w_{2it}\mu)}{w_{2it}^2 \delta^2}]$.

i) The full conditional distribution for μ is a normal distribution. We set a normal prior, $N(\eta\ , C)$. Recall that the transition equation can be written in the following manner.

$$\begin{aligned} m_{it+1} &= w_{1it} m_{it} + w_{2it}\mu + \chi_{it}, \\ \chi_{it} &\sim N(0, w_{2it}^2 \delta^2). \end{aligned}$$

Thus, we can rewrite the above equation as follows.

$$\begin{aligned} \frac{m_{it+1} - w_{1it} m_{it}}{w_{2it}} &= \mu + \varkappa_{it}, \\ \varkappa_{it} \sim N(0, \delta^2). \end{aligned}$$

Using the above equation and combining it with the prior specification, we obtain that the posterior distribution for μ is normally distributed as well. We denote this posterior distribution by $N(\overline{m}, \overline{V})$, where $\overline{V}^{-1} = C^{-1} + \sum_{i=1}^{I} \frac{(T_i - 1)}{\delta^2}$ and $\overline{m} = \overline{V}[C^{-1}\eta + \frac{1}{\delta^2} \sum_{i=1}^{I} \sum_{t=1}^{T_i - 1} (\frac{m_{it+1} - w_{1it} m_{it}}{w_{2it}})$]. Here, T_i is the total number of observations for consumer i.

j) Draws for δ^2 are obtained by a random walk univariate Metropolis-Hastings step. We denote log(δ) by lsig and assume that lsig $\sim N(p, q)$. For the (m+1)th iteration, we then generate a candidate $lsig^c$ from a univariate normal proposal density $N(lsig^{(m)}, \tau_\delta^2)$. The proposal density is centered on the old value of $lsig^{(m)}$ from iteration m and the variance of the proposal density, τ_δ^2, is set to ensure rapid mixing of the chain. The generated candidate $lsig^c$ is accepted with the following acceptance probability

$$\alpha(lsig^{(m)}, lsig^c) = \min\left\{1, \frac{L((e^{lsig^c})^2)\varphi(lsig^c \mid p, q)}{L((e^{lsig^m})^2)\varphi(lsig^m \mid p, q)}\right\}. \quad (2)$$

where $\varphi()$ represent the normal density for the prior and L is the likelihood from the transition equation. If the candidate is accepted, then $lsig^{(m+1)} = lsig^c$, otherwise, $lsig^{(m+1)} = lsig^{(m)}$. We set p to 0 and q to 1.0. The tuning constant τ_δ^2 was set to 0.02.

k) The expected utilities $\mathbf{EU}_{ijt}$ for any given observation can be drawn in a data augmentation step.

This involves drawing each plan expected utility from a truncated conditional normal distribution where the truncation points depend upon whether the utility is for the chosen plan or not.