Summary of Short-term Research Objectives

Detian Deng

May 14, 2015

1 Model Specification

Let L be a K-dimensional Bernoulli random variable denoting the true state. Consider the general log linear model:

$$f(l;\Theta) = \exp\{\Theta_1^T l + \Theta_2^T u_2 + \ldots + \Theta_K^T u_K - A^*(\Theta)\}$$

where U_k is a $\binom{K}{k} \times 1$ vector of k-way cross-products, $k = 1, \ldots, K$, and $\Theta = (\Theta_1, \ldots, \Theta_K)$ contains the the natural parameters, which is a $(2^K - 1) \times 1$ vector.

Model restrictions, let $\tilde{l} = (l, u_2, \ldots, u_K)^T$, and $S = \sum_{j=1}^K L_j = s$ has some fixed pmf

$$\begin{aligned} \pi(s) :=& P(S = s) \\ =& \frac{1}{A(\Theta)} \sum_{\tilde{l}:S=s} \exp\{\Theta^T \tilde{l}\} \text{ , } s = 0, 1, \ldots, K \end{aligned} \tag{1}$$

$$A(\Theta) = \sum_{\tilde{l}:l\in\{0,1\}^K} \exp\{\Theta^T \tilde{l}\} \tag{2}$$

1.1 Additional Definition of Parameters

Using previous notations, with some properly defined $\boldsymbol{\gamma}$, we have:

$$\begin{aligned}P(L_i;\boldsymbol{\pi},\boldsymbol{\gamma}) =& P(L_i,S_i;\boldsymbol{\pi},\boldsymbol{\gamma})\\ =& P(L_i|S_i;\boldsymbol{\pi},\boldsymbol{\gamma})P(S_i;\boldsymbol{\pi},\boldsymbol{\gamma})\end{aligned}$$

where $P(L_i|S_i;\boldsymbol{\pi},\boldsymbol{\gamma}) = P(L_i|S_i;\boldsymbol{\gamma})$ because $1(S_i = s)$ is the sufficient statistic for π_s, furthermore S_i is the sufficient statistic for $\boldsymbol{\pi}$. Also $P(S_i;\boldsymbol{\pi},\boldsymbol{\gamma}) = \pi_{S_i}$ by definition.

Therefore, we have

$$P(L_i;\boldsymbol{\pi},\boldsymbol{\gamma}) = P(L_i|S_i;\boldsymbol{\gamma})\pi_{S_i}$$

Then we define the following parameters:

$$\begin{aligned}\gamma_{j_1,\ldots,j_s} =& P(L_{ij_1} = \ldots = L_{ij_s} = 1|S_i = s) \qquad (3)\\ \boldsymbol{\gamma} =& (\gamma_1,\gamma_2,\ldots,\gamma_{12},\ldots,\gamma_{1\ldots K})^T, \text{ where } \sum_j \gamma_j = \sum_{j\neq j'} \gamma_{jj'} = \sum_{j\neq j'\neq j''} \gamma_{jj'j''} = \ldots = \gamma_{1\ldots K} = 1\end{aligned}$$

Therefore $(\boldsymbol{\pi},\boldsymbol{\gamma})^T$ is a vector of length $2^K + K$ with degrees of freedom $2^K - 1$.

Let $J_i = \{j : L_{ij} = 1\}$. We have,

$$P(L_i;\boldsymbol{\pi},\boldsymbol{\gamma}) = \gamma_{J_i}\pi_{S_i} \qquad (4)$$

The relation between $(\boldsymbol{\pi},\boldsymbol{\gamma})$ and Θ is defined by equation (2) together with:

$$\gamma_{J_i} = \frac{\exp(\Theta^T \tilde{l}_i)}{\sum_{l:l^T 1 = S_i}\exp(\Theta^T \tilde{l})} \qquad (5)$$

with $2^K - 1 - K$ degrees of freedom.

Therefore (1), (3) and (5) together define $2^K - 1$ non-linear equations for $2^K - 1$ unknowns. If there exists a unique root for the above non-linear system, then there is a one-to-one mapping between $(\boldsymbol{\pi},\boldsymbol{\gamma})$ and Θ, which provides the re-parameterization.

1.2 Find the Re-parameterization

1.2.1 Quasi-Newton Method

Numerically solve the system defined by (1), (3) and (5). As the dimension of L grows ($K > 6$), multiple sets of starting values are needed to reach the solution. Also, solutions to high order Θ are subject to larger error.
See code GammaToTheta() in the appendix.

1.2.2 Restrict Θ to QE model

Setting all high order interaction parameter to 0, using only the equations defined by $pi_0, pi_1, \gamma_1, \ldots, \gamma_{K-1}$ and $\gamma_{11}, \ldots, \gamma_{K-2,K}$, which are in total $\frac{K(K+1)}{2}$ equations, we can solve for Θ for larger value of K.
See code GammaToTheta.QE() in the appendix.

2 Posterior Distribution

$$\begin{aligned}
P(\mu, \theta^{(2)}|L) \propto & P(L, \mu, \theta^{(2)}) \\
\propto & P(L, \mu, \theta^{(1)}, \theta^{(2)}, \pi) \\
\propto & P(L|\mu, \theta^{(1)}, \theta^{(2)}, \pi)P(\mu, \theta^{(1)}, \theta^{(2)}, \pi) \\
\propto & P(L|\theta^{(1)}, \theta^{(2)})P(\theta^{(1)}, \theta^{(2)}|\mu, \pi)P(\mu)P(\pi) \\
\propto & \mathrm{QE}(L; \theta^{(1)}, \theta^{(2)})\mathrm{UFR}(\theta^{(1)}, \theta^{(2)}|\mu, \pi)\mathrm{N}(\mathrm{logit}(\mu), \Sigma)\mathrm{tPois}(\pi)
\end{aligned}$$

where QE is the second-order log linear model.
UFR is a Multivariate distribution of $[\theta^{(1)}, \theta^{(2)}|\mu, \pi]$ subject to non-linear constrains: $M(\theta^{(1)}, \theta^{(2)}) = \mu$ and $\Pi(\theta^{(1)}, \theta^{(2)}) = \pi$, which can be sampled by a two-step procedure.
tPois is a truncated conjugate Poisson distribution defined as:

$$\begin{aligned}
\pi \sim & \mathrm{Dirichilet}(\mathrm{hist}(\vec{s})) \\
s \sim & \frac{\lambda^s}{s!}e^{-\lambda}/[1 - \sum_{s>K} \frac{\lambda^s}{s!}e^{-\lambda}]
\end{aligned}$$

2.1 On sampling $[\theta^{(1)}, \theta^{(2)}|\mu, \pi]$

2.1.1 QE model

For QE model, we have a $J_1 \times J_2$ design matrix $\tilde{L}$, where $J_1 = 2^K - 1, J_2 = \frac{1}{2}K(K+1)$. Recall that

$$A(\Theta) = \frac{1}{\pi(0)}$$

$$\pi(s) = \frac{1}{A(\Theta)} \sum_{\tilde{l}:S=s} \exp\{\Theta^T \tilde{l}\}\ ,\ s = 1, \ldots, K \tag{6}$$

$$\mu_k = \frac{1}{A(\Theta)} \sum_{\tilde{l}:l_k=1} \exp\{\Theta^T \tilde{l}\}\ ,\ k = 1, \ldots, K \tag{7}$$

Define intermediate parameter $\phi_j = \exp(\theta^T \tilde{l}_j) > 0$, $\theta = (\theta^{(1)}, \theta^{(2)}), j = 1, \ldots, J_1$ and two $K \times J_1$ sub-design matrices B, C, where $B[k, j] = 1(\tilde{l}_j^T 1 = k)$, $C[k, j] = 1(\tilde{L}[j, k] = 1)$. Thus (6) and (7) become

$$\begin{aligned}
\vec{\phi} > & 0 \\
B\vec{\phi} = & \vec{\pi}/\pi(0) \\
C\vec{\phi} = & \vec{\mu}/\pi(0)
\end{aligned}$$

Based on [1][2], we can sample $\vec{\phi}$ from Uniform distribution subject to the above linear constraints efficiently and robustly. Now we have a over-determined linear system: (J_1 equations with J_2 unknowns)

$$\tilde{L}\theta = \log \vec{\phi}$$

Then we can use Least Square method to solve for θ.

2.1.2 General model

For General model, $\tilde{L}$ is $J_1 \times J_1$, the intermediate parameters $\vec{\phi}$ and $\pi(0)$ fully specify all cell probabilities, thus the posterior distribution becomes

$$P(\mu, \vec{\phi}|L) \propto \mathrm{LL}(L; \vec{\phi})\mathrm{UFR}(\vec{\phi}|\mu, \pi)\mathrm{N}(\mathrm{logit}(\mu), \Sigma)\mathrm{tPois}(\pi)$$

Reference

1. Smith RL . Efficient Monte-Carlo Procedures for Generating Points Uniformly Dis- tributed over Bounded Regions. Operations Research, 32(6), 12961308 (1984).
2. Van den Meersche, Karel, K. E. R. Soetaert, and D. J. Van Oevelen. "xsample (): an R function for sampling linear inverse problems." Journal of Statistical Software 30 (2009).