Assignment 3

STAT 3006, 2021/22

Due date: Apr 19, 2022 23:59, Tuesday

Question 1

(30%) There are 500 i.i.d. Poisson distributed random variables (r.v.s) with mean parameter λ. You know 35 r.v.s take values at zero; 73 r.v.s at one; 136 r.v.s at two; 95 r.v.s at three; 83 r.v.s at four; and 78 r.v.s are more than or equal to five. Based on the information, please carry out a hybrid Gibbs sampler to estimate λ.

Note: assign the prior $\pi(\lambda) \propto 1/\lambda$ to λ; $2,000$ iterations are needed, and collect posterior samples in the last $1,000$ iterations; use posterior mean to estimate λ.

Question 2

(30%) There are 1,000 samples $\{X_1, X_2, \ldots, X_{1000}\}$ in a data set `Assg3_Q2.txt`. For each sample i, $X_i = (X_{i1}, X_{i2}, X_{i3})$. You know that these samples are from three clusters. For sample i, we use Z_i to denote the cluster number to which sample i belongs. The proportion of the three clusters is denoted by π_1, π_2, π_3, where $\sum_{k=1}^{3} \pi_k = 1$. Specifically, for each sample i, $P(Z_i = k) = \pi_k$. Given $Z_i = k$, $X_{i1} \sim Bino(10, \theta_{1k})$, $X_{i2} \sim Bino(20, \theta_{2k})$, and $X_{i3} \sim Bino(30, \theta_{3k})$. Estimate parameters θ_{jk} with $j = 1, 2, 3, k = 1, 2, 3$, (π_1, π_2, π_3) and $(Z_i)_{i=1}^{1,000}$.

Note: assign $Dirichlet(2, 2, 2)$ prior to (π_1, π_2, π_3), and assign uniform prior $p(\theta_{jk}) \propto 1$ to θ_{jk}, with $j = 1, 2, 3; k = 1, 2, 3$; implement Gibbs sampler 5,000 iterations, and only samples in the last 2,000 iterations are kept; use posterior mean to estimate $\boldsymbol{\theta}$, $\boldsymbol{\pi}$ and use posterior mode to estimate $\mathbf{Z}$.

Question 3

(40%) There are four candidates in an election. The election comprises of two voting stages, morning and afternoon, and there are 100 votes for each stage. Denote y_{ij} as the vote number of Candidate j at stage i, with $i = 1, 2$ and $j = 1, 2, 3, 4$. However, an accident occurs during the vote counting such that we only know that the vote number of some candidates are larger or equal to 15 as the following table:

	Candidate 1	Candidate 2	Candidate 3	Candidate 4
Stage I	≥ 15	≥ 15	22	31
Stage II	28	≥ 15	26	≥ 15

If we assume that $\boldsymbol{Y}_i = (y_{i1}, y_{i2}, y_{i3}, y_{i4}), i = 1, 2$ satisfies the same multinomial distribution with 100 trials and event probabilities $\boldsymbol{p} = (p_1, p_2, p_3, p_4)$. Based on the available information, please carry out a hybrid Gibbs sampler to estimate $\boldsymbol{p}$.

Note: assign the Dirichlet prior $\pi(\boldsymbol{p}) \propto p_1p_2p_3p_4$ to $\boldsymbol{p}$. Regard y_{12} and y_{22} as latent variables and update y_{12} by the proposal:

1. if $15 < y_{12}^{(t-1)} < 32$, then $P(y_{12}^* = y_{12}^{(t-1)} - 1 \big| y_{12}^{(t-1)}) = P(y_{12}^* = y_{12}^{(t-1)} + 1 \big| y_{12}^{(t-1)}) = 0.5$;
2. if $y_{12}^{(t-1)} = 15$, then $P(y_{12}^* = 16 \big| y_{12}^{(t-1)}) = 1$;
3. if $y_{12}^{(t-1)} = 32$, then $P(y_{12}^* = 31 \big| y_{12}^{(t-1)}) = 1$.

Similarly, you can obtain a proposal distribution for y_{22}. 10,000 iterations are needed, and collect posterior samples in the last 5,000 iterations; use posterior mean to estimate $\boldsymbol{p}$.

Requirements

-	in the paper report	in the R code file
Q1	Detailed derivation of hybrid Gibbs sampler an estimate for λ	R code
Q2	Detailed derivation of Gibbs sampler all estimates for $\boldsymbol{\theta}, \boldsymbol{\pi}, \mathbf{Z}$ Trace plots for θ_{11}, Z_1 and π_1	R code
Q3	Detailed derivation of hybrid Gibbs sampler estimates for $\boldsymbol{p}$	R code