SSC386D / M394C: Monte Carlo Statistical Methods — Midterm 1

Solutions are due **We, March 11** as harcopy, or as pdf per email (please use subject “MCMC Midterm”) by Fr, March 13, 5pm in my mailbox (8th floor RLM, or stop by my office).

Open book, open notes, internet research (wikipedia or whatever you care) is fine. Just no teamwork please, and help from anyone else outside class.

You should plan around 5 hours for working through the problems (w/o implementation). Actual implementation (item (f)) might take longer. However, there is **no formal time limit**.

Likelihood (sampling model):

Let x_i, $i = 1, \ldots, n$, denote a set of multinomial random variables, with outcomes $x_i \in \{1, \ldots, N\}$. Let $y = (y_1, \ldots, y_N)$ denote a contingency table summarizing the multinomial experiment, i.e., y_j is the frequency of outcome j. Let $\pi_j = Pr(x_i = j)$ denote the unknown probabilty of observing outcome j, $\sum \pi_j = 1.0$. We write

$$y \sim Mn(n; \pi_1, \ldots, \pi_N)$$

Prior:

Due to the nature of the experiment we believe *a priori* that some of the π_j are much larger than others. We refer to this subset of outcomes with much larger probability as the “prevalent” outcomes A_1, and to the set of not so likley outcomes as “rare” outcomes A_0. We will formally define A_0 and A_1 below.

Parametrization: To define a prior probability model for $\theta = (\pi_1, \ldots, \pi_N)$ it is convenient to introduce latent indicators z_i, and a change of variables for the π_j:

$$\pi_j = \begin{cases} \pi^* q_j & \text{if } z_j = 1 \\ (1 - \pi^*)\, r_j & \text{if } z_j = 0, \end{cases} \tag{1}$$

with $\sum q_j = 1$ and $\sum r_j = 1$ (let $q_j = 0$ if $z_j = 0$, and $r_j = 0$ if $z_j = 1$). In words, the latent variable z_j is an indicator for outcome j being a prevalent outcome, i.e., $A_1 = \{j : \; z_j = 1\}$ and $A_0 = \{j : \; z_j = 0\}$. The probability of observing some prevalent outcome is π^*, with π^* close to 1.0; q_j is the probability of outcome j given that we observe a prevalent outcome, and r_j is the probability of j given a rare outcome. For later reference we define

$$M_1 = \#A_1 \text{ and } M_0 = N - M_1$$

as the number of prevalent and rare outcomes, respectively.

Prior Probability Model $p(z, \pi^*, q, r)$: We assume

$$Pr(z_j = 1) = \rho, \tag{2}$$

a beta prior on the total probability of prevalent outcomes:

$$\pi^* \sim Be(a^*, b^*), \tag{3}$$

and a Dirichlet prior for the partitioning of π^* into the cell probabilities π_j, $j \in A_1$. Let $\tilde{q}_h$, $h = 1, \ldots, M_1$ denote the non-zero weights q_j.

$$(\tilde{q}_1, \ldots, \tilde{q}_{M_1}) \sim Dir(M_1; a_1, \ldots, a_1). \tag{4}$$

The use of equal Dirichlet parameters a_1 reflects the fact that prior believes about $(q_1, \ldots, q_{M_1})$ are *a priori* exchangeable. Similarly for r_j, $j \in A_0$.

$$(\tilde{r}_1, \ldots, \tilde{r}_{M_0}) \sim Dir(M_0; a_0, \ldots, a_0). \tag{5}$$

Hyperparameters:

The hyperparameters ρ, a^*, b^*, a_1, a_0 are fixed. Use, for example,

$$\rho = 0.1, a^* = 9, b^* = 1, a_1 = 0.1 \text{ and } a_0 = 10.$$

Data

posted on the canvas page as *sage.dta*, together with the pdf file of this exam. The file gives the observed values $x_i \in \{0, N\}$, $i = 1, \ldots, n$, with $n = 700$ and $N = 77$.

22	48	54	96	96	94	96	22	22	102	99	33	87	96	99	96	26	28	26	117
53	116	57	28	40	28	33	96	26	33	96	98	53	102	96	33	96	952	54	96
102	33	102	33	28	85	96	33	102	85	96	25	102	33	26	43	96	2	102	87
96	94	33	96	96	87	28	28	18	54	96	102	102	96	96	96	96	85	33	85
18	33	96	84	83	85	96	57	99	89	26	96	102	96	96	33	96	28	631	748
33	85	25	85	7	96	67	33	85	120	96	22	96	85	33	35	33	96	102	54
26	96	43	22	43	28	98	948	39	33	102	54	33	102	102	85	96	96	96	96
99	48	96	96	85	96	33	96	26	28	102	96	87	33	99	96	28	2	102	22
102	87	102	26	85	57	22	94	102	75	85	28	96	117	85	33	48	28	98	28
35	28	85	33	96	54	28	487	96	503	102	85	33	54	28	96	618	33	98	96
182	96	26	96	94	112	87	96	26	28	397	87	87	99	102	96	33	22	96	57
18	94	28	85	443	28	112	85	102	85	102	87	22	636	85	2	96	96	43	75
22	85	96	96	96	96	33	33	813	854	33	96	87	85	85	96	85	85	33	84
85	102	96	348	420	748	96	22	87	33	96	96	32	96	96	28	102	85	28	54
96	33	98	28	26	35	33	96	96	85	85	102	102	87	40	96	22	33	85	96
85	85	96	85	102	33	96	393	96	28	118	96	120	26	87	85	102	85	96	85
85	87	96	102	96	28	96	85	102	85	96	96	98	26	96	33	33	28	85	118
33	42	39	33	28	26	28	96	102	369	85	102	102	96	96	96	28	117	94	28
96	96	85	85	54	28	96	22	102	102	33	85	112	96	33	87	85	28	102	96
85	30	43	102	33	33	96	960	96	33	118	99	18	28	53	98	33	54	96	96
96	98	113	33	116	22	87	85	96	96	96	26	102	96	22	96	221	71	96	57
96	57	39	96	32	57	96	87	102	85	28	28	102	102	96	94	33	33	96	85
102	54	96	96	85	85	96	43	96	33	102	33	85	102	96	534	33	57	98	28
33	33	33	26	102	96	33	102	26	87	85	33	87	630	26	94	87	958	102	85
96	28	96	96	118	102	116	33	53	102	26	102	882	22	33	102	28	118	83	96
94	96	33	116	102	96	117	96	96	33	26	22	85	85	85	54	32	22	99	723
30	26	22	98	28	85	85	96	26	33	87	28	33	96	28	161	26	85	102	54
28	96	85	102	550	18	28	96	96	96	96	85	85	96	49	28	28	33	96	102
102	85	33	71	96	33	33	116	94	102	85	102	43	689	28	102	102	26	96	366
347	96	96	7	85	96	96	26	96	102	28	40	85	96	94	28	85	96	28	96
102	28	85	87	85	85	87	96	96	28	57	28	85	32	96	96	28	85	587	96
102	85	96	54	28	102	163	87	96	26	96	28	85	96	85	102	22	96	102	87
898	102	54	87	28	33	96	39	57	846	326	96	28	94	22	96	33	33	40	54
48	96	94	96	116	118	28	939	96	96	116	96	28	87	864	102	54	33	96	102
28	33	33	96	33	85	57	26	28	22	121	96	85	102	680	28	87	85	96	102

(a)[2pts] Joint probability model. Write out the joint probability model

$$p(\pi^\star, q_1, r_1, \ldots, q_N, r_N, z_1, \ldots, z_N, y_1, \ldots, y_N).$$

(b)[2pts] Graphical Model. Show a graphical model represenation of the probability model. Use circles (or whatever shape you like) for each r.v., and connect any two random variables that are *not* conditionally indpendent given all other variables. [1]

(c)[4pts] Conditional posterior distributions. Find the complete conditional posterior distributions

1. $p(z_i | \ldots)$
2. $p(\pi^* | \ldots)$
3. $p(q | \ldots)$
4. $p(r | \ldots)$

Here $\ldots$ denotes "all other parameters and the data y."

(d)[2pts] Marginalizing w.r.t. q and r: Find the posterior distribution $p(z | \pi^*, y)$, marginalizing w.r.t. q and r.

(e)[4pts] MCMC I. Consider a Gibbs sampling scheme based on sampling from the complete conditional posterior distributions found in (c), i.e. an MCMC scheme with steps:

1. $z_i^{t+1} \sim p(z_i | \ldots)$
2. $\pi^{*(t+1)} \sim p(\pi^* | \ldots)$
3. $q^{t+1} \sim p(q | \ldots)$
4. $r^{t+1} \sim p(r | \ldots)$.

Show that MCMC I violates irreducibility (of course, don't implement MCMC I – it would not work).

[1] I realize we did not talk much about graphical models in class. Nothing fancy requird here - just a representation of conditional independence structure by an undirected graph; or a directed acyclical grap (DAG) for the model (if you know how to do this).

(f)[4pts] MCMC II. Implement a Gibbs sampling scheme based on sampling from the conditional posterior distributions found in (d), (c3), (c4) and (c2):

1. $z_i \sim p(z_i|z_{-i}, \pi^*, y)$, $i = 1, \ldots, n$
2. $q \sim p(q|\ldots)$
3. $r \sim p(r|\ldots)$.
4. $\pi^* \sim p(\pi^*|\ldots)$

Use a convergence diagnostic of your choice to determine when to stop. In your report, please discuss and show:

(a) Trajectories of simulated values $\pi^{*(t)}$ against iteration t.

(b) Convergence diagnostic used.

(c) Boxplot of the (marginal) posterior distributions $p(\pi_j|y)$. Use one figure with multiple boxplots (using, for example, the R command `boxplot(.)`).

(d) Plot posterior means $E(\pi_j|y)$ against the m.l.e.'s $\hat{\pi}_j = y_j/n$. Include the 45 degree line and discuss the shrinkage pattern you (should) see.

(g)[2pts] MCMC III. Consider the following MCMC scheme. We describe the MCMC by constructive definition of the transition probability, i.e., $p(\theta^{t+1}|\theta^t)$:

1. Metropolis-Hastings step to change z_i, q, r, $i = 1, \ldots, N$.
   - Generate a proposal (z_i', q', r'): [2]

   $$Pr(z_i' = 1) = 0.5,\ q' \sim p(q|z', \ldots, y),\ \text{and } r' \sim p(r|z', \ldots, y),$$

   where z' is the currently imputed z, with z_i replaced by z_i', i.e., $z' = (z_1, \ldots, z_{i-1}, z_i', z_{i+1}, \ldots, z_N)$.

   - Compute an appropriate acceptance probability α.
   - Set

   $$(z_i, q, r) = \begin{cases} (z_i', q', r') & \text{with prob } \alpha \\ (z_i, q, r) & \text{with prob } 1 - \alpha \end{cases}$$

2. $\pi^* \sim p(\pi^*|\ldots)$

Find the correct expression for the acceptance probability α (in step 1.).

No need to implement MCMC III.

(z) Mystery: Can anyone guess what the application behind this probability model could be.

[2] We use q' to mark proposals, since we already used the traditional $\tilde{q}$ for the non-zero q_i, before.

R Code

Here is some fragment R code (also on canvas), to get you started..

```
require(gtools)  # I found a Dir r.v. generator in this package
## if needed install it with
## install.packages("gtools")

## DATA
x <- scan("sage.dta")                # raw data
y <- table(x)                        # counts
N <- length(y)
names(y) <- 1:N
n <- length(x)

## HYPERPARS:
rho <- 0.1
as <- 0.9
bs <- 0.1
a1 <- .1
a0 <- 10

## initialize
## this function creates a list with
##   z=(z1,.. zN); pis=pi*, r=(r~[1],..r~[M0])
##   q=(q~[1],..q~[M1])
## Youc an use it to initialize the state of the MC
init <- function()
   { # initialize parameters
     ## z
     z <- ifelse(y<10,0,1)
     ## pi*: empirical frquency
     A0 <- which(z==0);     A1 <- which(z==1)
     M0 <- length(A0);      M1 <- length(A1)
     Y0 <- sum(y[A0]);      Y1 <- sum(y[A1])
     pis <- sum(Y1)/n
    ## r and q: empirical fequencies
    q <- y[A1]/Y1  # this is the q~ of the text
    r <- y[A0]/Y0  # this is the r~ of the text
    return(th=list(z=z,pis=pis,r=r,q=q))
  }

## main function for MCMC
gibbs <- function(n.iter=100, verbose=0)
  {
    TH <- NULL # initialize - will save pi*,z here
              ##             for each iteration
    PI <- NULL # similar - will save (pi1,.., piN) here
    th <-  init()
    pis <- th$pis        # initialize pis = pi*
    z <- th$z            # initialize z
    for(it in 1:n.iter){ # loop over iterations
      z <- sample.z(pis, z)   # 1. z ~ p(z | pis, y)
      q <- sample.q(pis,z)    # 2. q ~ p(q | pis,z,y)
      r <- sample.r(pis,z)    # 3. r ~ p(r | pis,z,y)
      pis <- sample.pis(z)    # 4. pi
      if (verbose > 0){
        if (it %% 10 ==0)       # print short summary
          prt.summary(z,q,r,pis)
      }
      ## save iteration
      TH <- rbind(TH, c(pis,z))
```

```
      pi <- rep(0,N)
      pi[z==1] <- pis*q
      pi[z==0] <- (1-pis)*r
      PI <- rbind(PI, pi)
    }
    return(list(TH=TH, PI=PI))
  }

## run the MCMC :-)
ex <- function()
  { ## RUN these lines to get the plots
    n.iter <- 500
    gbs <- gibbs(n.iter)
    ## assume gbs returns a list with elements
    ## TH = (niter x p) matrix with each row being the
    ##      state (pi, z)
    ## PI = (niter x 1) vector with pi
    TH <- gbs$TH
    PI <- gbs$PI
    its <- 1:n.iter

    ## trajectory plot
    plot(its, TH[,1],xlab="ITER",ylab="PI*",bty="l",type="l")

    ## boxplot
    boxplot(log(PI))

    ## plotting posterior means vs. mle's
    pibar <- apply(PI,2,mean) # posterior means
    pihat <- as.numeric(y)/n
    plot(pihat, pibar, type="p",
         pch=19, bty="l",xlab="MLE pihat", ylab="E(pi | y)")
    abline(0,1)

    ## same thing, zoom in to left lower corner
    plot(pihat, pibar, type="p", xlim=c(0,0.03), ylim=c(0,0.03),
         pch=19, bty="l",xlab="MLE pihat", ylab="E(pi | y)")
    abline(0,1)
  }

#######################################################
## aux functions
```

... ah the rest you have to do :-). Write functions `sample.z()`, `sample.q()`, `sample.r()` and `sample.pis()` to carry out one step of the Gibbs sampler.