MCMC M1 : A

MD

Write the joint probability model.

$$p(\pi^*, q_1, \dots, q_N, r_1, \dots, r_N, z_1, \dots, z_N, y_1, \dots, y_N) = \text{"Joint"}$$

"Joint" distribution:

Prior:

$$Pr(z_j = 1) = p \implies Pr(z_1, \dots, z_N) = p^{M_1}(1-p)^{M_0}$$

$$Pr(\pi^* | a^*, b^*) = \frac{1}{B(a^*, b^*)} \cdot (\pi^*)^{a^*-1}(1-\pi^*)^{b^*-1}$$

$$Pr(q_1, \dots, q_N | z_j, a_1) = \prod_{j=1}^{N} \left[1_{z_j}(1) \cdot \frac{q_j^{a_1 - 1}}{B(a_1)} + 1_{z_j}(0) \cdot 1 \right]$$

Cases where $z_j = 1$, gets Dirichlet prob

Cases where $z_j = 0$, $q_j = 0$ with prob $= 1$

$$Pr(r_1, \dots, r_N | z_j, a_0) = \prod_{j=1}^{N} \left[1_{z_j}(1) \cdot 1 + 1_{z_j}(0) \cdot \frac{r_j^{a_0 - 1}}{B(a_0)} \right]$$

Cases where $z_j = 1$, $r_j = 0$ w/ prob $= 1$

Cases where $z_j = 0$, gets Dirichlet prob.

For each j:

$$Pr(\pi_j | z_j, \pi^*, q_j, r_j) = 1_{z_j}(1) \cdot \pi^* q_j + 1_{z_j}(0) \cdot (1-\pi^*) r_j$$

Likelihood; for each y:

$$P(y | n, \pi_1, \dots, \pi_N) = \prod_{j=1}^{N} \pi_j^{y_j}$$

"Joint": $P(y | n, \pi_1, \dots, \pi_N) \cdot \prod_{j=1}^{N} \left[Pr(\pi_j | z_j, \pi^*, q_j, r_j) \right] \cdot Pr(r_1, \dots, r_N | z_j, a_0) \cdot Pr(q_1, \dots, q_N | z_j, a_1) \cdot Pr(\pi^* | a^*, b^*) \cdot Pr(z_1, \dots, z_n)$

$$\text{Joint} = \overbrace{\prod_{j=1}^{N} \pi_j^{y_j}}^{P(y \mid \pi_{1,\dots,N})} \times \underbrace{\prod_{j=1}^{N} \left[\pi^* \cdot q_j \cdot I(z_j=1) + (1-\pi^*) \cdot r_j \cdot I(z_j=0) \right]}_{P(\underline{\pi} \mid \underline{z}, \pi^*, \underline{q}, \underline{r})}$$

$$\times \underbrace{\prod_{j=1}^{N} \left[1 \cdot I(z_j=1) + \frac{r_j^{a_0-1}}{B(a_0)} \cdot I(z_j=0) \right]}_{P(\underline{r} \mid \underline{z}, a_0)}$$

$$\times \underbrace{\prod_{j=1}^{N} \left[\frac{q_j^{a_1-1}}{B(a_1)} \cdot I(z_j=1) + 1 \cdot I(z_j=0) \right]}_{P(\underline{q} \mid \underline{z}, a_1)}$$

$$\times \underbrace{\frac{1}{B(a^*, b^*)} (\pi^*)^{a^*-1} (1-\pi^*)^{b^*-1}}_{P(\pi^* \mid a^*, b^*)}$$

$$\times \underbrace{\rho^{\sum_{j=1}^{N} I(z_j=1)} \cdot (1-\rho)^{\sum_{j=1}^{N} I(z_j=0)}}_{P(\underline{z} \mid \rho)}$$

$$= \prod_{j=1}^{N} \pi_j^{y_j} \times \left[\prod_{j \in A_1} q_j \right] \cdot (\pi^*)^{M_1} \cdot \left[\prod_{j \in A_0} r_j \right] \cdot (1-\pi^*)^{M_0}$$

$$\times \left[\prod_{j \in A_0} \frac{r_j^{a_0-1}}{B(a_0)} \right]$$

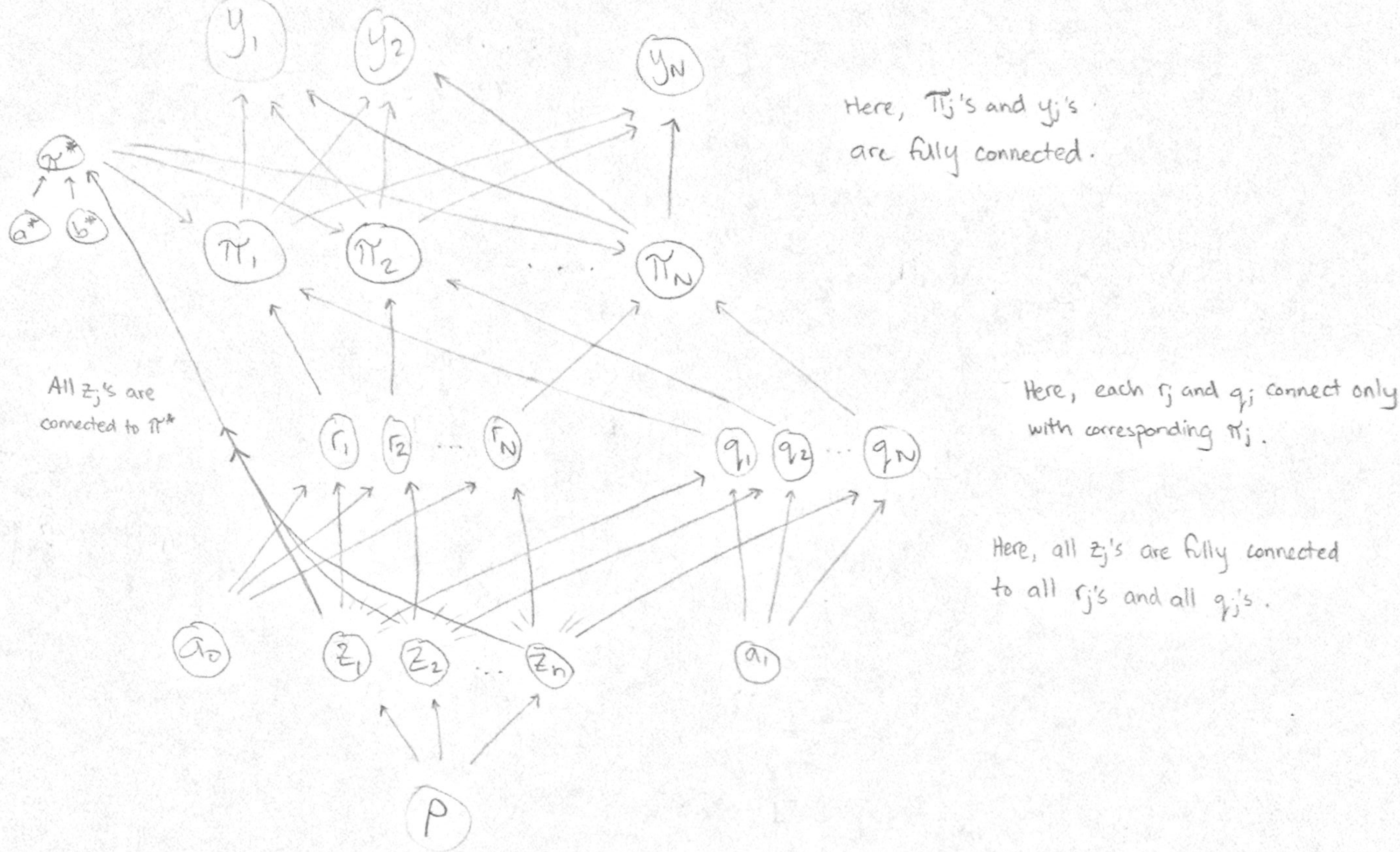
$$\times \left[\prod_{j \in A_1} \frac{q_j^{a_1-1}}{B(a_1)} \right]$$

$$\times \frac{1}{B(a^*, b^*)} (\pi^*)^{a^*-1} (1-\pi^*)^{b^*-1}$$

$$\times \rho^{M_1} (1-\rho)^{M_0}$$

MCMC M1: B

Graphical Model



Edges pointing to a node i represent dependencies of i on those parents in i's complete conditional posterior.

E.g. After conditioning on all other variables, q is still dependent on a_1, so receives an edge from a_1 to q.

MCMC M1: C   Conditional Posterior Distributions   MD

Given the joint, which is proportional (to a constant) to the posterior, each conditional posterior can be found by regarding all other variables as constant.

$p(z_i|\ldots)$: Probability terms for π_j, r_j, q_j, and z_j are all functions of z.

$$P(z|\ldots) \propto P(\underline{\pi}|\underline{z},\pi^*,q,r)\cdot P(\underline{r}|\underline{z},a_0)\cdot P(q|\underline{z},a_1)\cdot P(\underline{z}|\rho)$$

SEE BELOW *

$$\propto \left[\prod_{j\in A_1} q_j\right](\pi^*)^{M_1}\cdot\left[\prod_{j\in A_0} r_j\right]\cdot(1-\pi^*)^{M_0}\cdot\left[\prod_{j\in A_0} r_j^{a_0-1}\right]\cdot\left[\prod_{j\in A_1} q_j^{a_1-1}\right]\cdot\rho^{M_1}(1-\rho)^{M_0}$$

$$\propto \left[\prod_{j\in A_1} q_j^{a_1}\right]\cdot\left[\prod_{j\in A_0} r_j^{a_0}\right]\cdot(\pi^*)^{M_1}(1-\pi^*)^{M_0}\cdot\rho^{M_1}(1-\rho)^{M_0}$$

$\sim Dir(M_1; a_1+1,\ldots,a_1+1)$
$\cdot Dir(M_0; a_0+1,\ldots,a_0+1)$
$\cdot Beta(M_1+1, M_0+1)$
$\cdot Beta(M_1+1, M_0+1)$

$P(\pi^*|\ldots)$: Only the π_j and π^* probability distributions contain π^*.

$$P(\pi^*|\ldots) \propto \prod_{j=1}^{N}\left[\delta(z_j=1)\cdot\pi^* q_j + \delta(z_j=0)\cdot(1-\pi^*)\cdot r_j\right]\cdot(\pi^*)^{a^*-1}(1-\pi^*)^{b^*-1}$$

$$\propto \left[\prod_{j\in A_1} \pi^* q_j\right]\cdot\left[\prod_{j\in A_0} (1-\pi^*)\cdot r_j\right]\cdot(\pi^*)^{a^*-1}(1-\pi^*)^{b^*-1}$$

Recall $|A_1| = M_1$, $|A_0| = M_0$

$$\propto \left[\prod_{j\in A_1} q_j\right]\left[\prod_{j\in A_0} r_j\right](\pi^*)^{M_1+a^*-1}\cdot(1-\pi^*)^{M_0+b^*-1}$$

$$\sim Beta(M_1+a^*, M_0+b^*)$$

$$* \; P(z_i|\ldots) \propto I(z_i=1)\left[q_i^{a_1}\cdot\pi^*\cdot\rho\right] + I(z_i=0)\cdot\left[r_i^{a_0}\cdot(1-\pi^*)(1-\rho)\right]$$

$P(q|\ldots)$: Only the π_j an $q_1, \ldots, q_N$ terms contain q.

$$P(q|\ldots) \propto \left[\prod_{j \in A_1} q_j\right] \cdot \left[\prod_{j \in A_1} \frac{q_j^{a_1 - 1}}{B(a_1)}\right]$$

$$\propto \prod_{j \in A_1} q_j \cdot q_j^{a_1 - 1} = \prod_{j \in A_1} q_j^{a_1} \sim Dir\left(M_1; a_1+1, \ldots, a_1+1\right)$$

$P(r|\ldots)$: Only the π_j and $r_1, \ldots, r_N$ terms contain r.

$$P(r|\ldots) \propto \left[\prod_{j \in A_0} r_j\right] \left[\prod_{j \in A_0} \frac{r_j^{a_0 - 1}}{B(a_0)}\right]$$

$$\propto \prod_{j \in A_0} r_j \cdot r_j^{a_0 - 1} = \prod_{j \in A_0} r_j^{a_0} \sim Dir\left(M_0; a_0+1, \ldots, a_0+1\right)$$

MCMC M1 : D

Find posterior $p(z \mid \pi^*, y)$, marginalizing w.r.t. q and r.

Use complete conditional from (C) to compute this posterior :

$$P(z \mid \pi^*, y) = \iint P(z \mid \pi^*, y, q, r)\, P(q \mid \pi^*, y)\, P(r \mid \pi^*, y)\, dq\, dr$$

Marginalizing out from the complete conditional is integrating the enlarged distribution with respect to q and r. This is equivalent to ignoring terms without q and r.

$$\propto \left[\prod_{j \in A_1} q_j^{a_1}\right]\left[\prod_{j \in A_0} r_j^{a_0}\right] \underbrace{(\pi^*)^{M_1}(1-\pi^*)^{M_0} \cdot \rho^{M_1}(1-\rho)^{M_0}}_{\text{ignore}}$$

$$\propto \mathrm{Dir}(M_1; a_1+1, \ldots) \cdot \mathrm{Dir}(M_0; a_0+1, \ldots)$$

For MCMC II, used instead $P(z_i \mid \pi^*, y) = \delta(z_i=1)\left[q_i^{a_1}\, \pi^* \cdot \rho\right] + \delta(z_i=0)\left[r_i^{a_0}(1-\pi^*)(1-\rho)\right]$

as derived from part (C). [Bottom of page].

MCMC M1: E

Violation of Irreducibility

A Gibbs sampler should be able to reach any configuration of parameters in a finite number of steps, and with positive probability. This is called irreducibility.

Irreducibility is violated if there exists any $\underbrace{\text{configuration of parameters}}_{\equiv \text{ point in the subspace}}$ that cannot be reached ($Pr=0$) in a finite number of steps.

If steps are in the order z_i^{t+1}, $\pi^{*(t+1)}$, q^{t+1}, r^{t+1} suppose that in the current q, $q_i > 0$ and $r_i = 0$, because $z_i = 1$. Now suppose that in the 1st update, that of z_i^{t+1}, z_i^{t+1} becomes $= 0$. The current setting for the second update is:

$$z_i^{(t+1)} = 0, \quad q_i^{(t)} = a : a > 0, \quad r_i^{(t)} = 0$$

The 2nd update will then proceed. $P(\pi^* | \ldots) \propto \prod_{j=1}^{N} \left[\pi^* q_j \cdot I(z_j=1) + (1-\pi^*) r_j \cdot I(z_j=0) \right]$

The ith term of this product will be $\pi^* \cdot q_i \, I(z_i=1) + (1-\pi^*) r_i \cdot I(z_i=0)$.

Since $z_i^{(t+1)} = 0$, the first term is zero because the indicator is zero, and the second term is zero because $r_i = 0$.

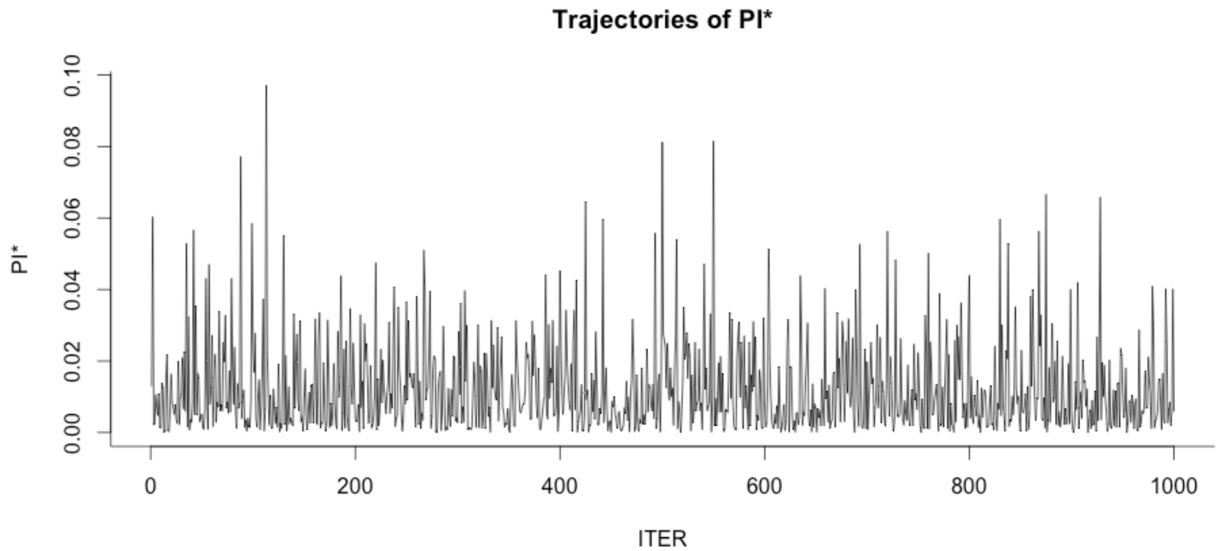
This entire product, $P(\pi^* | \ldots)$, equals zero, making some configurations impossible to reach.

This proves that there exist cases where points in the parameter space have zero probability, contradicting irreducibility. ∎

MCMC M1.F

(a) Trajectories of simulated values of pistar(t) against iteration / convergence diagnostics used.

The trajectories are spastic (not smooth), and appear to jump up and down.

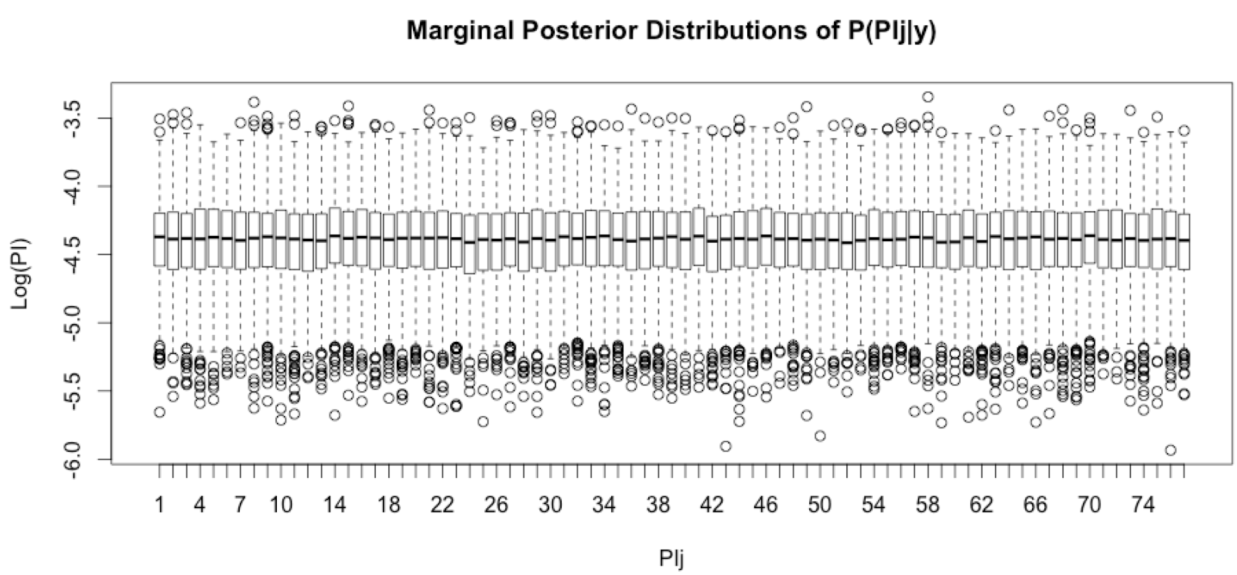


(b) Convergence diagnostics used.

As part of convergence diagnostics, one checks the mixing of the variable over many iterations (here, there is decent mixing with `n=1000`), the stationarity of the values (here, the values hover around one value, that of ~`0.01`), and whether the beginning portion of the chain appears highly dependent and moving in a noticeable direction (here, it seems there is no significant burn-in required).

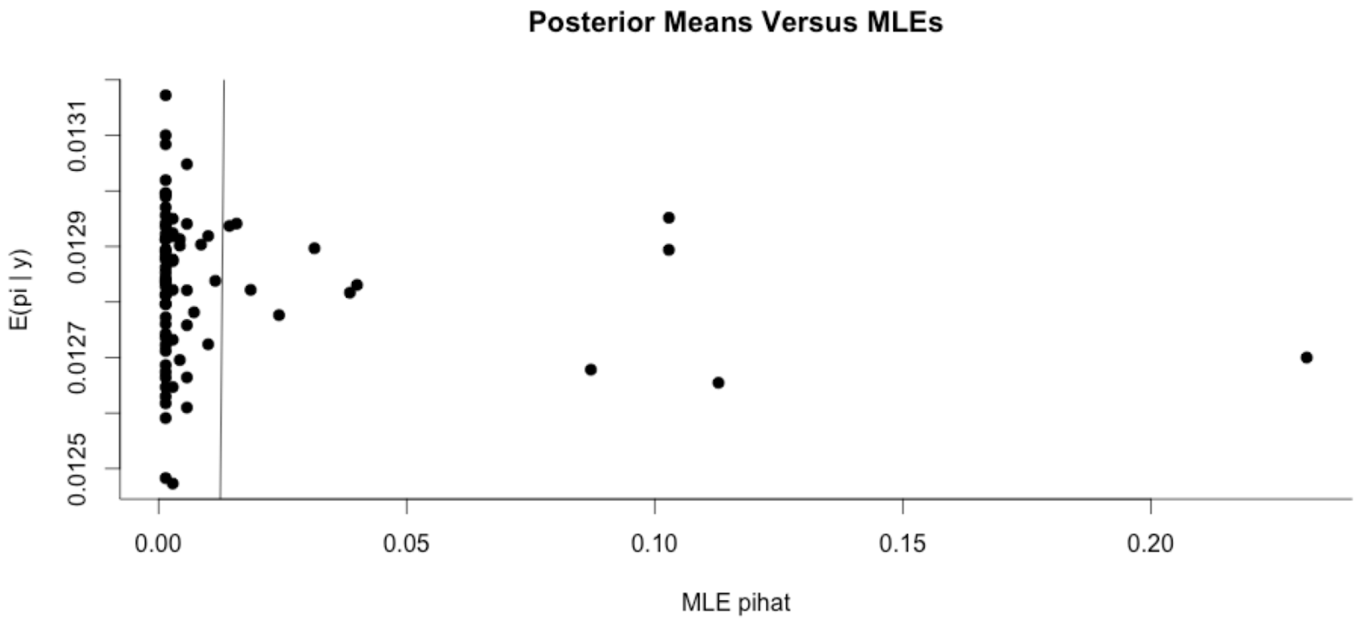
(c) Boxplot of the (marginal) posterior distributions p(πj |y).

The values for each `PIj` vary between ~`exp(-6)` to ~`exp(-3.5)`, centered at ~`exp(-4.4)` or ~`0.012`. There is a narrow interquartile range, and many lower outliers, a potentially concerning detail that requires further investigation.

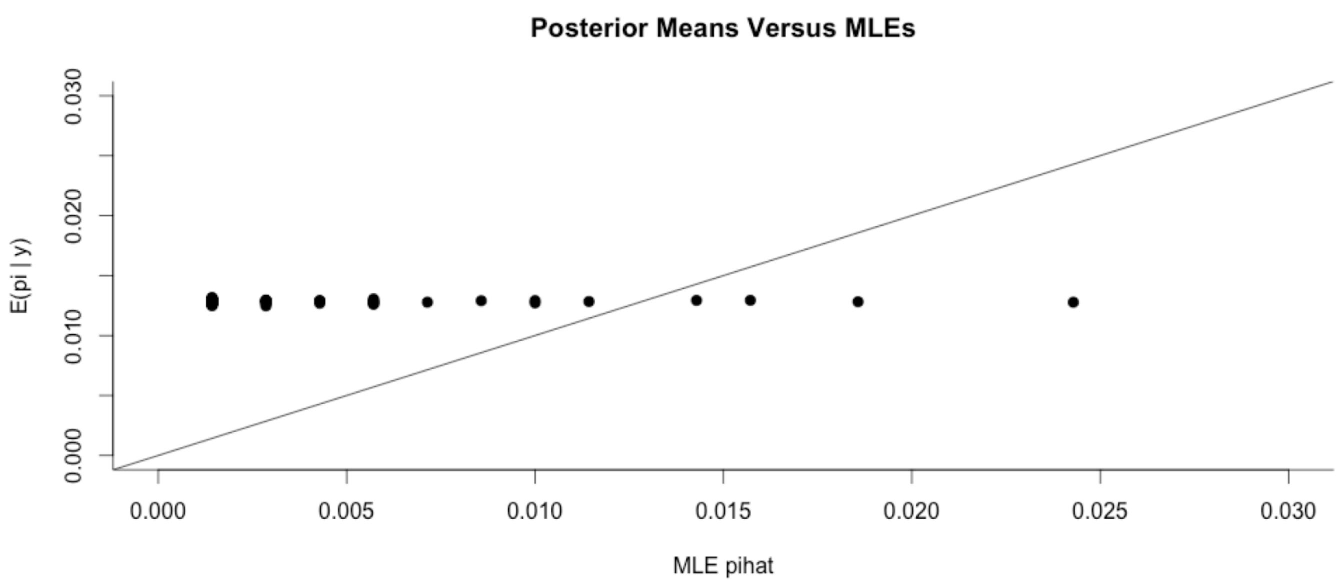


(d) Posterior means, MLEs, and shrinkage.

Here the line drawn is the 45 degree line, y=x.



Here, the graph is zoomed into the left corner, and compared against the 45 degree, y=x line. The y-axis of the upper graph shows that the posterior mean values were very concentrated in the range of about 0.0125 and 0.0130. On the equally dimensioned axes of the lower graph, it is evident that the slope of the posterior means is significantly lower than the 45 degree y=x line. This indicates a large amount of shrinkage occurred, where the model “borrowed” information from the prior mean, due to small bucket sizes in the data likelihood.



CODE

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

## DATA
setwd("~/Google Drive/2. SPRING 2015/MCMC - Prof Mueller/M1")
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); pistar=pi*, r=(r~[1],..r~[M0])
##   q=(q~[1],..q~[M1])
## You can 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) # Set of indices where z=0 / z=1.
  M0 <- length(A0);      M1 <- length(A1) # Num of indices where z=0 / z=1.
  Y0 <- sum(y[A0]);      Y1 <- sum(y[A1]) # Num of X's with z=0 / z=1.
  pistar <- sum(Y1)/n
  ## q and r: 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,pistar=pistar,q=q,r=r))
}

## main function for MCMC
gibbs <- function(n.iter, verbose) {
  TH <- NULL # initialize - will save pi*,z here
             ##             for each iteration
  PI <- NULL # similar - will save (pi1,.., piN) here
  th <-  init()
  pistar <- th$pistar        # initialize pistar = pi*
  z <- th$z            # initialize z

  # Vector of probabilities for when z=1, and zeros otherwise.
  q <- rep(0, N)
  q.probs <- th$q
  q[which(z==1)] <- q.probs

  # Vector of probabilities for when z=0, and zeros otherwise.
  r <- rep(0, N)
  r.probs <- th$r
  r[which(z==0)] <- r.probs

  for (i in 1:n.iter) { # loop over iterations
    z <- sample.z(pistar,z, q, r)       # 1. z ~ p(z | pistar, y)
```

```
    q <- sample.q(pistar,z, q)       # 2. q ~ p(q | pistar,z,y)
    r <- sample.r(pistar,z, r)       # 3. r ~ p(r | pistar,z,y)
    pistar <- sample.pistar(z)    # 4. pi
    if (verbose > 0){
      if (i %% 10 == 0) {
        # print short summary, for every 10th iteration
        print(paste("Iteration: ", i))
        print(round(z, 2))
        print(round(q, 2))
        print(round(r, 2))
        print(round(pistar, 2))
      }
    }
    ## save iteration
    TH <- rbind(TH, c(pistar,z))

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

ex <- function() {
  ## RUN these lines to get the plots
  n.iter <- 1000; verbose=0
  gbs <- gibbs(n.iter, verbose)
  ## 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 of Z
  plot(its, TH[,1],xlab="ITER",ylab="PI*",bty="l",type="l",
       main="Trajectories of PI*")

  ## boxplot
  boxplot(log(PI), main="Marginal Posterior Distributions of P(PIj|y)",
          xlab="PIj", ylab="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)",
       main="Posterior Means Versus MLEs")
  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)",
       main="Posterior Means Versus MLEs")
  abline(0,1)

}
```

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

sample.z <- function(pistar, z, q, r) {
  for (i in 1:length(z)) {
    if (z[i]==1) {
      pr <- q[i]^a1*pistar*rho
    } else if (z[i]==0) {
      pr <- r[i]^a0*(1-pistar)*(1-rho)
    }
    z[i] <- ifelse(runif(1)<pr, 1, 0)
  }
  return (z)
}

sample.q <- function(pistar, z, q) {
  M1 <- length(which(z==1))
  if (M1>0) {
    new.q <- rdirichlet(1, rep(a1+1, M1))
    q[which(z==1)] <- new.q
  } else if (M1==0) {
    q <- rep(0, N)
  }
  return (q)
}

sample.r <- function(pistar, z, r) {
  M0 <- length(which(z==0))
  if (M0>0) {
    new.r <- rdirichlet(1, rep(a0+1, M0))
    r[which(z==0)] <- new.r
  } else if (M0==0) {
    r <- rep(0, N)
  }
  return (r)
}

sample.pistar <- function(z) {
  M0 <- length(which(z==0))
  M1 <- length(which(z==1))
  new.pistar <- rbeta(1, M1+as, M0+bs)
  return (new.pistar)
}
```

MCMC M1: G

MH Acceptance Probability

MD

Acceptance probability is the ratio of the posterior probability under the new parameters and the posterior probability under the current parameters. Each probability is then weighted by the inverse probability of arriving in that state from the previous one.

Let $\theta = (z_i, q, r)$ be the current values, $\theta' = (z_i', q', r')$ be the proposed values, $P(\cdot)$ be the posterior distribution, and $T(\theta|\theta')$ be the transition probability of moving from θ' to θ.

$$\alpha = \frac{P(\theta'|y)}{P(\theta|y)} \cdot \frac{T(\theta|\theta')}{T(\theta'|\theta)}$$

$$T(\theta|\theta') = \frac{1}{T(\theta'|\theta)}$$

$$T(\theta'|\theta) = \frac{T(z_i'|z,q,r)}{T(z|z_i',q,r)} \cdot \frac{T(q'|q,z,r)}{T(q|q',z,r)} \cdot \frac{T(r'|r,q,z)}{T(r|r',q,z)}$$

$$= \left[\frac{P(z_i')}{P(z_i)} \cdot \frac{T(z|z')}{T(z'|z)}\right] \cdot \left[\frac{P(q')}{P(q)} \cdot \frac{T(q|q')}{T(q'|q)}\right] \cdot \left[\frac{P(r')}{P(r)} \cdot \frac{T(r|r')}{T(r'|r)}\right]$$

$$= 1.$$

Note:
$T(z_i|z') = P(z_i)$
$T(z'|z) = P(z_i') = 0.5$
$T(q|q') = P(q)$
$T(q'|q) = P(q') = q'$
$T(r|r') = P(r)$
$T(r'|r) = P(r') = r'$

Proposing from full conditions, the MH acceptance probability will be 1. This makes it equivalent to the Gibbs sampler. ∎

MCMC M1.Z

This model could be applied to studies of market share and consumer choice. I envision the following application of each parameter.

Y - Number of people who bought cars of brands 1,...,N
Pi1,...,PiN - Probabilities of a person buying a car from each brand.
Z - Indicator of whether someone is a “mainstream” brand buyer, or an “eccentric” brand buyer.
Q - Market shares among “mainstream” brands.
R - Market shares among “eccentric” brands.
A1 - Set of “mainstream” brands.
A0 - Set of “eccentric” brands.
M1 - Number of “mainstream” brands.
M0 - Number of “eccentric” brands.
Pistar - Market type scalar, that amplifies market share of “mainstream” brands (the Qs), and dampens market share of “eccentric” brands (the Rs), especially given the right-weighted Beta(0.1, 0.9).