HW3

Sam Olson

Q1

Following the solution or instructor's comments on Part 1, take the MCMC output files provided and approximate the posterior distribution of the weight of fish at length 15 inches for each of the 75 lakes. Also produce 90% credible intervals for the posterior mean weight. Order the posterior means small to large, and graph the posterior means along with line segments representing credible intervals with lake index on the vertical axis. An example for a hypothetical example with 5 situations (will be our lakes) is shown in Figure 1. Your plot will have 75 intervals going up the vertical axis (and will be unlikely to extend to 5 on the horizontal axis). Note that you will have to re-index the lakes upon ordering the values. In real life we would want to retain a mapping from that re-indexing to the lake number of the original data so that IDNR could identify which lake is which.

Answer

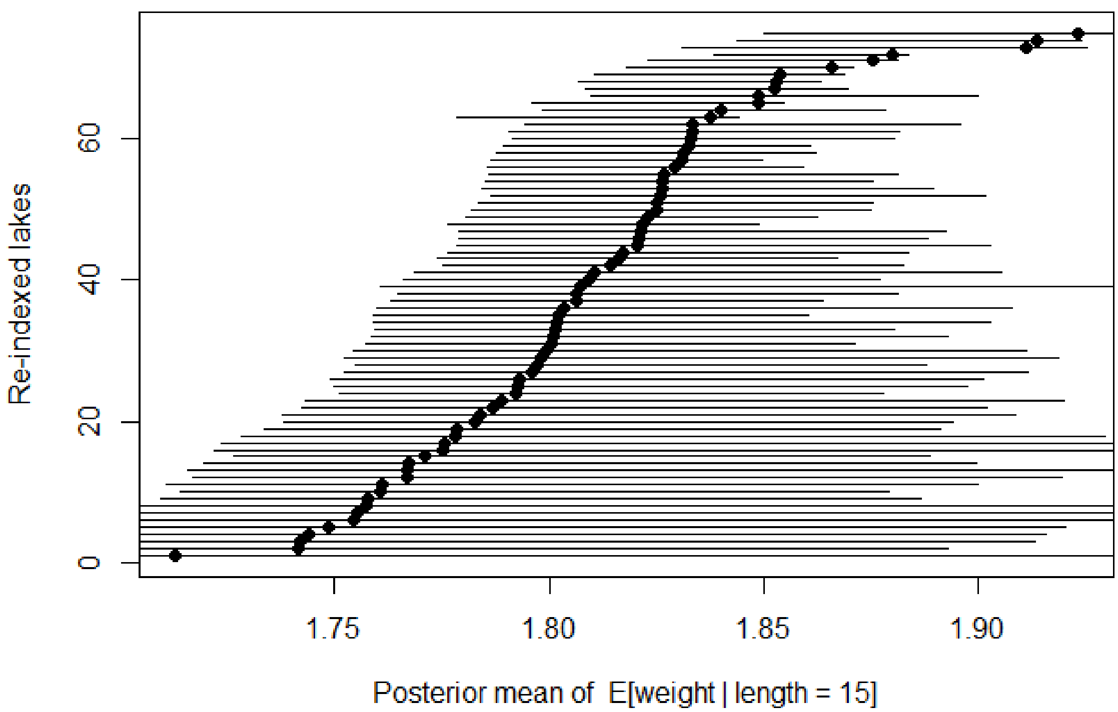

Figure 1: Posterior mean expected weight of a 15-inch largemouth bass by lake, ordered from smallest to largest. Horizontal line segments show 90% credible intervals.

Q2

Also take the MCMC output files and approximate, for each lake, the posterior mean of the probability that a 15 inch large mouth bass exceeds 1.83 pounds. Order these values and plot them on the horizontal axis against the re-indexed lakes (1 through 75) on the vertical axis. This plot will look quite a bit like an empirical distribution function.

Answer

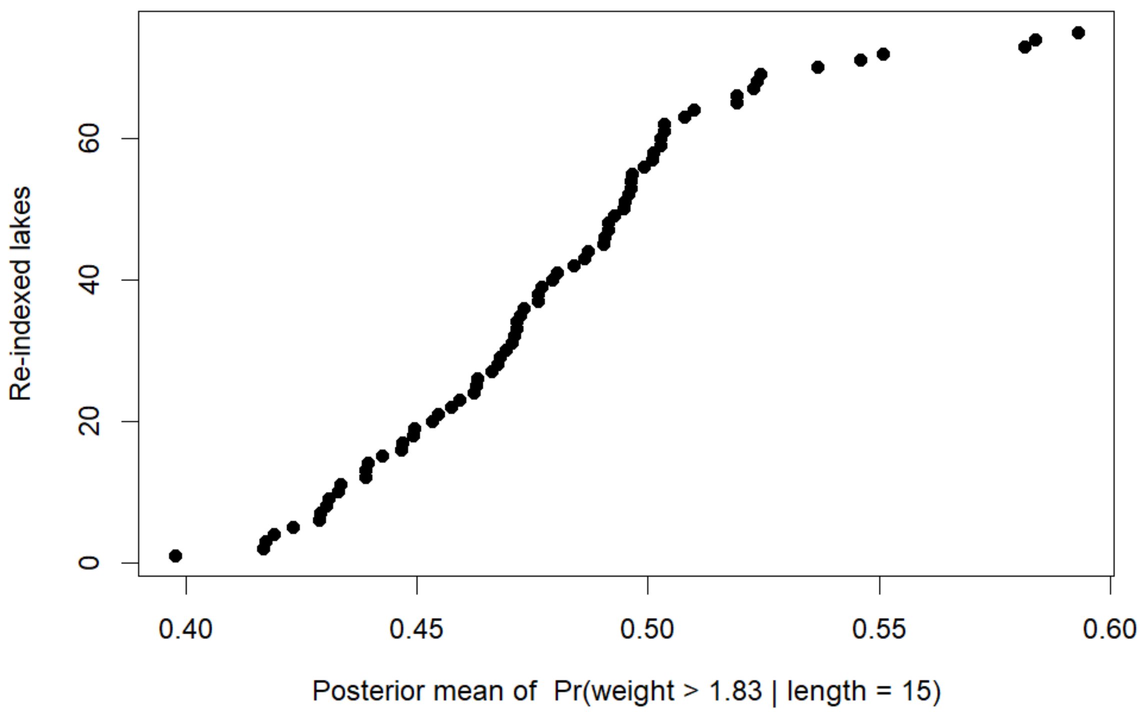

Figure 2: Posterior mean probability that a 15-inch largemouth bass exceeds 1.83 pounds, by lake. Lakes are ordered from smallest to largest.

Appendix

```r
# Read-in data
dat <- read.table("LMBdat_for601.txt", header = TRUE)
m1 <- read.table("LMBMCMC1.txt", header = TRUE)
m2 <- as.matrix(read.table("LMBMCMC2.txt", header = TRUE))
# Objects for later
n_lakes <- 75
lake_levels <- sort(unique(dat$lknum))
S <- nrow(m2)
# Q1: length value
x0 <- 15
# Q2: weight value
w0 <- 1.83

# Posterior draws (Source: LMBMCMC2.txt)
beta_s <- m2[, 1:n_lakes, drop = FALSE]
sig2_s <- m2[, (n_lakes + 1):(2*n_lakes), drop = FALSE]
alpha_s <- m2[, 2*n_lakes + 1]

################################################
# Q1: Posterior mean weight | length 15, by lake
################################################
# Note: From background/model setup
# mu[i] = beta_i * x_i^alpha
# mu15[s,i] = beta_i[s] * 15^(alpha[s])
mu15 <- beta_s * (x0 ^ alpha_s)

# posterior mean
mu15_mean <- colMeans(mu15)
# 90% credible interval for posterior mean
mu15_ci90 <- t(apply(X = mu15, MARGIN = 2, FUN = quantile, probs = c(0.05, 0.95)))

# organize statistics
q1_tbl <- data.frame(
  lake_id = 1:n_lakes,
  lknum = lake_levels,
  post_mean_mu15 = mu15_mean,
  lo90 = mu15_ci90[, 1],
  hi90 = mu15_ci90[, 2]
)

# order small to large
# this also re-indexes the lakes
q1_tbl_ord <- q1_tbl[order(q1_tbl$post_mean_mu15), ]
q1_tbl_ord$rank <- seq_len(nrow(q1_tbl_ord))

# plot
par(mar = c(4, 6, 2, 2))
plot(x = q1_tbl_ord$post_mean_mu15,
     y = q1_tbl_ord$rank,
     xlab = expression(paste("Posterior mean of  E[weight | length = ", 15, "]")),
     ylab = "Re-indexed lakes",
```

```r
     pch = 19)
segments(x0 = q1_tbl_ord$lo90,
         y0 = q1_tbl_ord$rank,
         x1 = q1_tbl_ord$hi90,
         y1 = q1_tbl_ord$rank)

################################################
# Q2: Posterior mean > 1.83 | length 15, by lake
################################################
# Note: Based on model formulation
# Under theta = 1
# We have normally distributed errors with sd = sqrt(sig2_i) * mu15
sd15 <- sqrt(sig2_s) * mu15
# standardize
z <- (w0 - mu15) / sd15
p_exceed_draws <- 1 - pnorm(z)
# aggregate by lake
p_exceed_mean <- colMeans(p_exceed_draws)

q2_tbl <- data.frame(
  lake_id = 1:n_lakes,
  lknum = lake_levels,
  post_mean_prob_exceed = p_exceed_mean
)

# order small to large (again)
# this also re-indexes the lakes (again)
q2_tbl_ord <- q2_tbl[order(q2_tbl$post_mean_prob_exceed), ]
q2_tbl_ord$rank <- seq_len(nrow(q2_tbl_ord))

# plot
par(mar = c(4, 6, 2, 2))
plot(x = q2_tbl_ord$post_mean_prob_exceed,
     y = q2_tbl_ord$rank,
     xlab = expression(paste("Posterior mean of  Pr(weight > ", 1.83, " | length = 15)")),
     ylab = "Re-indexed lakes",
     pch = 19)
```