Stat 591 Notes – Logistic regression and Metropolis–Hastings example

Ryan Martin (rgmartin@uic.edu)

November 2, 2013

Introduction

(This part taken from Example 1.13 in Robert and Casella, 2004.) In 1986, the space shuttle Challenger exploded during takeoff, killing the seven astronauts aboard. The explosion was a result of an O-ring failure, a splitting of a rubber ring that seals parts of the ship together. The accident was believed to be caused by the unusually cold weather (31°F) at the time of the launch, as there is reason to believe that the O-ring failure probabilities increase as the temperature decreases.

Data presented in the R code section below (see, also, p. 15 in Robert and Casella, 2004) gives temperature at flight time (in °F) and failure/success (0/1) of the O-ring for the flight at the given temperature. The temperatures range from 53°F to 81°F, with the majority of O-ring failures at the lower temperatures. Notice that the temperature at the Challenger launch was significantly lower those studied in the given data.

Logistic regression model

For a binary response, Y, with a predictor variable, X, logistic regression is a standard model. Specifically, Y, given $X = x$ is modeled as a Bernoulli random variable, with success probability $p(x)$, where $p(x)$ satisfies

$$\log \frac{p(x)}{1-p(x)} = \alpha + \beta x \iff p(x) = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}.$$

Here $\theta = (\alpha, \beta)$ is the unknown parameter. The likelihood function is

$$L(\alpha,\beta) \propto \prod_{i=1}^n \Bigl(\frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x_i}}\Bigr)^{y_i} \Bigl(\frac{1}{1+e^{\alpha+\beta x_i}}\Bigr)^{1-y_i},$$

and this can be maximized to produce the maximum likelihood estimators $\hat\theta = (\hat\alpha, \hat\beta)$. These, along with standard error estimates and significance tests (based on asymptotic normality) are obtained using the `glm` function in R, and the results are:

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  15.0429     7.3786   2.039   0.0415 *
x            -0.2322     0.1082  -2.145   0.0320 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Here we see that both α and β are significant at the 0.05 level. A plot of the estimated regression function is given in Figure 1. This confirms the hunch that the O-ring failure probability is decreasing in temperature. In particular, by extrapolation, we can speculate that the probability of O-ring failure at 31°F is almost certain, a fact which was apparently unknown at the time of the Challenger launch.

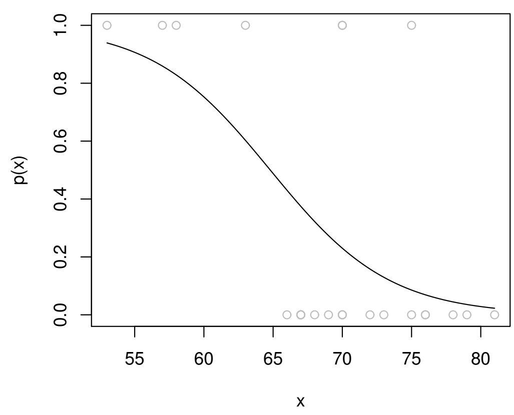


Figure 1: Plot of the estimated function $p(x)$ based on maximum likelihood.

Bayesian analysis

Following Robert and Casella, 2004, Example 7.11, we consider a prior for $\theta = (\alpha, \beta)$ of the following form:

$$\pi(\alpha, \beta) = \pi_\alpha(\alpha \mid b)\pi_\beta(\beta) = \frac{1}{b} e^{\alpha} e^{-e^{\alpha}/b},$$

where b is a hyper parameter to be specified. This is just an exponential prior (with scale b) for e^α and a flat prior for β; Robert and Casella claim this prior leads to a proper posterior. For the hyper parameter b, they propose a data-dependent choice, i.e., they pick b so that the prior mean for α equals the MLE $\hat{\alpha}$. A little bit of work reveals that this is achieved by setting

$$\hat{b} = e^{\hat{\alpha}+\gamma}, \quad \text{where } \gamma = 0.577216 \text{ is "Euler's constant."}$$

The motivation with this choice is to have the prior be centered somewhere near the center of the likelihood, so that the prior has minimal effect. Since the prior involves a hyper parameter estimated from data, this is really more like an empirical Bayes analysis, but let's not worry about this.

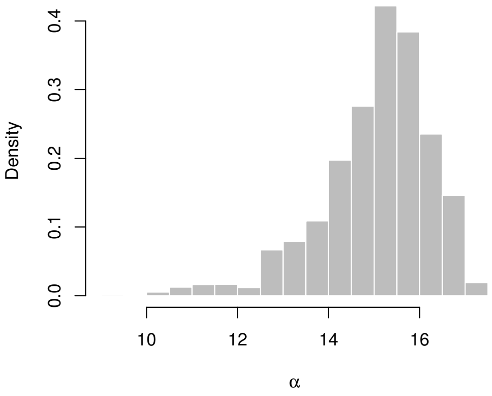
To implement the Metropolis–Hastings method to sample from the posterior, we shall consider an independence proposal distribution with density

$$g(\alpha, \beta) = \pi_\alpha(\alpha \mid \hat{b})\varphi(\beta),$$

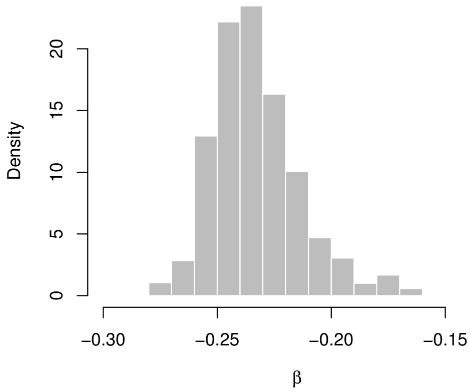
where π_α is as before, and $\varphi(\beta)$ is a normal density with mean $\hat{\beta} = -0.2322$ and standard deviation $\sigma_\beta = 0.1082$, from the MLEs above. Since both the prior and proposal have the same π_α terms, these will cancel out in the Metropolis–Hastings acceptance probability calculation. That is, for a given $\theta = (\alpha, \beta)$ and a proposed $\theta' = (\alpha', \beta')$, one must compute the acceptance probability

$$\min\Big\{\frac{L(\alpha', \beta')}{L(\alpha, \beta)} \frac{\varphi(\beta)}{\varphi(\beta')}, 1\Big\}.$$

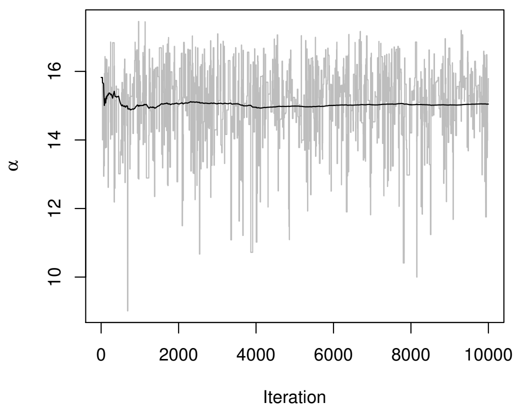
R code is given below to implement this Metropolis–Hastings sampler and summarize the results.



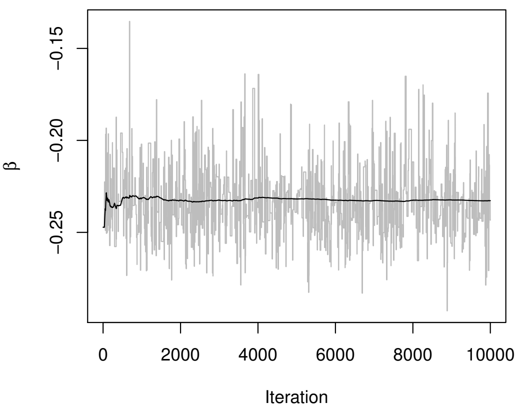
(a) Marginal posterior for α



(b) Marginal posterior for β



(c) Trace plot for α

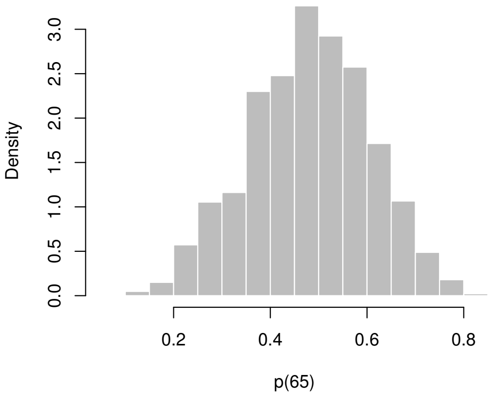


(d) Trace plot for β

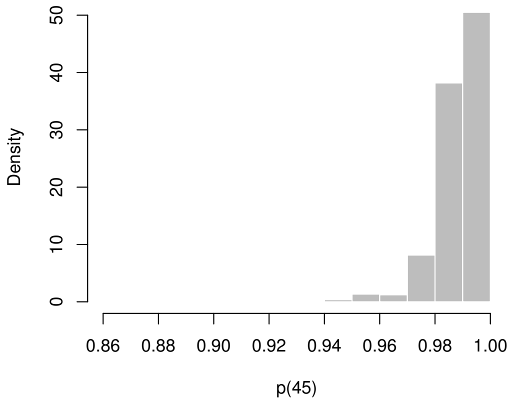
Figure 2: Marginal posteriors (top row) and trace plots with mean path (bottom row).

A first consideration is the marginal posterior distributions for α and β and with some assessment of the convergence of the Markov chain to the (stationary) posterior distribution. Histrograms of the posterior samples are summarized in Figure 2, panels (a) and (b). We see that these are roughly centered around the MLEs, but that the shapes are asymmetric and non-normal. The plots in Figure 2, panels (c) and (d), plot the paths of the samples as a function of the iteration. Since the curves are very jagged and quickly changing, we say that the chain is mixing well, suggesting that the sample is close to one from the true posterior. The black line displays the posterior mean as a function of iteration; both mean curves stabilize very quickly, suggesting convergence to stationarity.

For this particular problem, an interesting quantity of interest is the failure probability for a given temperature. Following Robert and Casella (2004), Example 1.13, we can consider two temperature values, namely, $x = 65°\text{F}$ and $x = 45°\text{F}$, both still considerably warmer than the day of the Challenger launch. Plots of the posterior distribution of $p(65)$ and $p(45)$ are shown in Figure 3. At the warmer temperature, the O-ring failure probability is near 0.5; however, at the cooler temperature, the failure probability is almost at 1! One can imagine that at the Challenger launch temperature, the failure



(a) $x = 65°$F



(b) $x = 45°$F

Figure 3: Posterior distributions of $p(65)$ and $p(45)$.

probability would be surely 1. Had this been known at the time, perhaps the astronauts' lives could have been saved.

```
# Generic Metropolis-Hastings function

mh <- function(x0, f, dprop, rprop, N, B) {

  x <- matrix(NA, N + B, length(x0)); fx <- rep(NA, N + B)
  x[1,] <- x0; fx[1] <- f(x0); ct <- 0
  for(i in 2:(N + B)) {

    u <- rprop(x[i-1,])
    fu <- f(u)
    r <- log(fu) + log(dprop(x[i-1,], u)) - log(fx[i-1]) - log(dprop(u, x[i-1,]))
    R <- min(exp(r), 1)
    if(runif(1) <= R) {

      ct <- ct + 1
      x[i,] <- u
      fx[i] <- fu

    } else {

      x[i,] <- x[i-1,]
      fx[i] <- fx[i-1]

    }

  }
  return(list(x=x[-(1:B),], fx=fx[-(1:B)], rate=ct / (N + B)))

}
```

```
# Data

y <- c(1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0)
x <- c(53, 57, 58, 63, 66, 67, 67, 67, 68, 69, 70, 70, 70, 70, 72, 73, 75, 75,
       76, 76, 78, 79, 81)

# Preliminary output from ML estimation

logreg.out <- glm(y ~ x, family=binomial(logit))
summary(logreg.out)
a.mle <- as.numeric(logreg.out$coefficients[1])
b.mle <- as.numeric(logreg.out$coefficients[2])
var.a.mle <- summary(logreg.out)$cov.scaled[1, 1]
var.b.mle <- summary(logreg.out)$cov.scaled[2, 2]
b.mme <- exp(a.mle + 0.577216)

# Posterior distribution

dpost <- function(theta) {

  a <- theta[1]
  b <- theta[2]
  p <- 1 - 1 / (1 + exp(a + b * x))
  lik <- exp(sum(dbinom(y, size=1, prob=p, log=TRUE)))
  dprior <- exp(a) * exp(-exp(a) / b.mme) / b.mme
  return(lik * dprior)

}

# Proposal distribution (independent proposal, so "theta0" is not used)

dprop <- function(theta, theta0) {

  a <- theta[1]
  b <- theta[2]
  pr1 <- exp(a) * exp(-exp(a) / b.mme) / b.mme
  pr2 <- dnorm(b, b.mle, sqrt(var.b.mle))
  return(pr1 * pr2)

}

rprop <- function(theta0) {

  a <- log(rexp(1, 1 / b.mme))
  b <- rnorm(1, b.mle, sqrt(var.b.mle))
  return(c(a, b))

}
```

```
# Run Metropolis-Hastings

N <- 10000
B <- 1000
mh.out <- mh(c(a.mle, b.mle), dpost, dprop, rprop, N, B)
alpha.mh <- mh.out$x[,1]
beta.mh <- mh.out$x[,2]
hist(alpha.mh, freq=FALSE, col="gray", border="white", xlab=expression(alpha))
plot(alpha.mh, type="l", col="gray", xlab="Iteration", ylab=expression(alpha))
lines(1:N, cumsum(alpha.mh) / (1:N))
hist(beta.mh, freq=FALSE, col="gray", border="white", xlab=expression(beta))
plot(beta.mh, type="l", col="gray", xlab="Iteration", ylab=expression(beta))
lines(1:N, cumsum(beta.mh)/(1:N))
p65 <- 1 - 1 / (1 + exp(alpha.mh + beta.mh * 65))
p45 <- 1 - 1 / (1 + exp(alpha.mh + beta.mh * 45))
hist(p65, freq=FALSE, col="gray", border="white", xlab="p(65)", main="")
hist(p45, freq=FALSE, col="gray", border="white", xlab="p(45)", main="")
```