ONLINE SUPPLEMENTAL MATERIAL

for

PRINCIPLES OF STATISTICAL INFERENCE: LIKELIHOOD AND THE BAYESIAN PARADIGM

STEVE C. WANG

Department of Geological and Environmental Sciences
Stanford University
450 Serra Mall, Building 320, Stanford, CA 94306

and

Department of Mathematics and Statistics
Swarthmore College
500 College Ave, Swarthmore, PA 19081
<scwang@swarthmore.edu>

Below is the code for producing the figures and working the examples. This was written in R version 2.10.1 (http://r-project.org) on an Apple Macintosh running OS X 10.6.4.

```
# Example 3 calculations
p0 <- .5
phat <- 442/779
dbinom(442,779,p0)/dbinom(442,779,phat)
teststat <- -2 * log( dbinom(442,779,p0)/dbinom(442,779,phat) )
teststat
1-pchisq(teststat, 1)
```

```
# Figure 1
par(mfrow=c(3,1), mai=c(.6,.6,.4,.2))

# graph of P(x | p), assuming p = .465
n <- 297
p <- 138/297
x <- 105:170
plot(x, dbinom(x,n,p), type="h", ylab="P(x | p)", bty="n", col=gray(.7))
points(x, dbinom(x,n,p), pch=16, cex=.5)
text(170,.044, "1", cex=1.6)

# graph of L(p), assuming x = 138
n <- 297
p <- seq(.33,.6,.001)
x <- 138
plot(p, dbinom(x,n,p), type="l", ylab="L(p)", bty="n")
text(.6,.044, "2", cex=1.6)
points(x/n, -.0012, pch=17)

# graph of log L(p), assuming x = 138
n <- 297
p <- seq(.33,.6,.001)
x <- 138
plot(p, dbinom(x,n,p, log=T), type="l", ylab="log L(p)", bty="n")
text(.6,-3.9, "3", cex=1.6)
points(x/n, -14.9, pch=17)
```

```r
# Figure 2
par(mfrow=c(1,2), mai=c(1,.6,.4,.2))

# graph of a likelihood function that is strongly concave
n <- 1000
p <- seq(.3,.7,.002)
x <- .5*n
plot(p, dbinom(x,n,p), type="l", ylab="", ylim=c(0,.027), bty="L", yaxt="n")
mtext("L(p)", side=2, line=.8)
text(.7,.026, "1", cex=1.2)

# graph of a likelihood function that is weakly concave
n <- 20
p <- seq(.3,.7,.002)
x <- .5*n
plot(p, dbinom(x,n,p), type="l", ylab="", ylim=c(0,.19), bty="L", yaxt="n")
mtext("L( p)",  side=2, line=.8)
text(.7,.18, "2", cex=1.2)
```

```r
# Figure 3
# lost wallets example
par(mfrow=c(3,1), mai=c(.7,1,.2,1))

# graph of prior on p
p <- seq(0,1,.001)
a <- 2
b <- 3
plot(p, dbeta(p,a,b), type="l", ylab="", yaxt="n", bty="L")
mtext("f(p)", side=2, line=1, cex=.65)
text(1, 1.6, "1. Prior", cex=1.2, adj=1)

# graph of likelihood
n <- 40
x <- 0
p <- seq(0,1,.001)
plot(p, dbinom(x,n,p), type="l", ylab="", yaxt="n", bty="L")
mtext("L(p)", side=2, line=1, cex=.65)
text(1, .925, "2. Likelihood", cex=1.2, adj=1)

# graph of posterior
a <- 2
b <- 3
n <- 40
x <- 0
p <- seq(0,1,.001)
plot(p, dbeta(p,a+x,b+n-x), type="l", ylab="", yaxt="n", bty="L")
mtext("f(p | x)", side=2, line=1, cex=.65)
text(1, 15.2, "3. Posterior", cex=1.2, adj=1)
```

```
# Figure 4
# examples of Beta distributions
par(mfrow=c(4,2), mai=c(.8,.6,.2,.1))
p <- seq(0,1,.001)
plot(p, dbeta(p,.8,4), type="l", ylab="", yaxt="n", bty="L")
mtext("f(p)", side=2, line=1, cex=.7)
text(.92, 9.4, "1", cex=1.4)
plot(p, dbeta(p,1,2), type="l", ylab="", yaxt="n", bty="L")
mtext("f(p)", side=2, line=1, cex=.7)
text(.92, 1.9, "2", cex=1.4)
plot(p, dbeta(p,1.8,5), type="l", ylab="", yaxt="n", bty="L")
mtext("f(p)", side=2, line=1, cex=.7)
text(.92, 2.4, "3", cex=1.4)
plot(p, dbeta(p,4,4), type="l", ylab="", yaxt="n", bty="L")
mtext("f(p)", side=2, line=1, cex=.7)
text(.92, 2.05, "4", cex=1.4)
plot(p, dbeta(p,.85,.85), type="l", ylab="", yaxt="n", bty="L")
mtext("f(p)", side=2, line=1, cex=.7)
text(.88, 1.97, "5", cex=1.4)
plot(p, dbeta(p,1,1), type="l", ylab="", yaxt="n", bty="L")
mtext("f(p)", side=2, line=1, cex=.7)
text(.92, 1.34, "6", cex=1.4)
plot(p, dbeta(p,11.5,8.2), type="l", ylab="", yaxt="n", bty="L")
mtext("f(p)", side=2, line=1, cex=.7)
text(.92, 3.33, "7", cex=1.4)
plot(p, dbeta(p,3,2), type="l", ylab="", yaxt="n", bty="L")
mtext("f(p)", side=2, line=1, cex=.7)
text(.98, 1.65, "8", cex=1.4)
```

```r
# Figure 5
# effect of sample size on posterior
par(mfcol=c(3,2), mai=c(.4,.6,.3,.2))

# sample size of n = 4
n <- 4
x <- 0
a <- 2
b <- 3
p <- seq(0,1,.001)

# graph of prior on p
plot(p, dbeta(p,a,b), type="l", ylab="", yaxt="n", bty="L")
mtext("f(p)", side=2, line=1, cex=.7)
text(.99, 1.7, "1", cex=1.5)

# graph of likelihood
plot(p, dbinom(x,n,p), type="l", ylab="", yaxt="n", bty="L")
mtext("L(p)", side=2, line=1, cex=.7)
text(.98, .95, "2", cex=1.5)

# graph of posterior
plot(p, dbeta(p,a+x,b+n-x), type="l", ylab="", yaxt="n", bty="L")
mtext("f(p | x)", side=2, line=1, cex=.7)
text(.98, 3.0, "3", cex=1.5)

# sample size of n = 100
n <- 100
x <- 0
a <- 2
b <- 3
p <- seq(0,1,.001)

# graph of prior on p
plot(p, dbeta(p,a,b), type="l", ylab="", yaxt="n", bty="L")
mtext("f(p)", side=2, line=1, cex=.7)
text(.99, 1.7, "4", cex=1.5)

# graph of likelihood
plot(p, dbinom(x,n,p), type="l", ylab="", yaxt="n", bty="L")
mtext("L(p)", side=2, line=1, cex=.7)
text(.98, .96, "5", cex=1.5)

# graph of posterior
plot(p, dbeta(p,a+x,b+n-x), type="l", ylab="", yaxt="n", bty="L")
mtext("f(p | x)", side=2, line=1, cex=.7)
text(.98, 36.5, "6", cex=1.5)
```

```r
# Figure 6: Bayesian credible interval on stratigraphic range
# Seymour Is ammonite data from Macellari 1986 -- Grossouvrites gemmatus;
par(mfrow=c(3,1), mai=c(.8,1.2,.5,1)*.6)
x <- c(1150,1140,1132.5,1110,1100,1090,1080,1067.5,
       1055,1037.5,1027.5,1015,1000)
x <- x - 1000    # rescale data ; use only finds above 1000 m
n <- length(x)
theta <- seq(1000,1200,.1)-1000
THETA <- expression(theta)      # for typesetting

# graph of prior
prior <- function(theta)  {  dnorm(theta,159.5,8)  }
plot(theta+1000, prior(theta), type="l", xlab="", ylab="", bty="L",
xaxt="n",yaxt="n")
text(1002, .045, "1. Prior", adj=0, cex=1.2)
mtext(seq(1000,1200,50), at=seq(1000,1200,50), side=1, line=.5, cex=.6)
mtext(THETA, side=1, line=2, cex=.7)
mtext(expression(paste("f(",theta,")")), side=2, line=1, cex=.7)

# graph of likelihood
lik <- function(theta,x)  {  1/theta^n * (theta>=max(x))  }
      # nonzero only if theta >= max(x)
plot(theta+1000, lik(theta,x), type="l", ylab= "",xlab="",bty="L",
xaxt="n",yaxt="n")
points(x+1000, rep(0,n), pch=16)    # add points for locations of fossil finds
text(1002, 4.8*10^(-29), "2. Likelihood", adj=0, cex=1.2)
mtext(seq(1000,1200,50), at=seq(1000,1200,50), side=1, line=.5, cex=.6)
mtext(THETA, side=1, line=2, cex=.7)
mtext(expression(paste("L",theta,")")), side=2, line=1, cex=.7)

# graph of posterior
posterior <- function(theta,x)  {  prior(theta) * lik(theta,x)  }
plot(theta+1000, posterior(theta,x),
type="l",ylab="",xlab="",bty="L",xaxt="n",yaxt="n")
points(x+1000, rep(0,n), pch=16)    # add points for locations of fossil finds
text(1002, 1.3*10^(-30), "3. Posterior", adj=0, cex=1.2)
mtext(seq(1000,1200,50), at=seq(1000,1200,50), side=1, line=.5, cex=.6)
mtext(THETA, side=1, line=2, cex=.7)
mtext(expression(paste("f(",theta," | x)")), side=2, line=1, cex=.7)

# calculate posterior mean
# this approximates integral{ theta * f(theta|x) dtheta }
sum((theta*posterior(theta,x))[-1]) / sum(posterior(theta,x)[-1])
  # need to omit the first value, which is NA

# shade in 90% credible interval
totalarea <- integrate(posterior, max(x), 250,x)$value
.9 * totalarea                  # want integral to equal this area = 1.8276e-29
upperlim <- 166.147                                     # found by trial and error
integrate(posterior, max(x), upperlim,x)   # verify that this equals 1.8276e-29
lines(c(upperlim,upperlim)+1000, c(0,posterior(upperlim,x)))
xvals <- (theta+1000)[theta<=upperlim]
xvals <- c(xvals, max(xvals))             # add last point to close the polygon
yvals <- posterior(theta,x)[theta<=upperlim]
yvals <- c(yvals,0)                       # add last point to close the polygon
polygon(xvals, yvals, col=gray(.8), border=NA)                     # add shading
points(max(x)+1000,0,cex=.8)       # replot in case it is hidden by the polygon
```