Laplace approximation notes

Oswaldo Gressani

December 12, 2022

Idea behind the Laplace approximation

Let $\mathcal{D}$ denote the observed data and θ a scalar parameter of interest. Using Bayes' rule, we can write the posterior density as:

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})},$$

and by omitting the normalizing constant $p(\mathcal{D}) = \int p(\mathcal{D}|\theta)p(\theta)d\theta$, usually called *the evidence* or *marginal likelihood*, we can write $p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$. The ideas of Laplace can be used to approximate a complex posterior by computing a second-order Taylor expansion of the log-posterior around its mode $\hat{\theta}$:

$$\begin{aligned}\log p(\theta|\mathcal{D}) \quad &\approx \quad \log p(\hat{\theta}|\mathcal{D}) + \frac{\partial \log p(\theta|\mathcal{D})}{\partial \theta}\bigg|_{\theta=\hat{\theta}} (\theta - \hat{\theta}) \\ &\quad + \frac{1}{2}\, \frac{\partial^2 \log p(\theta|\mathcal{D})}{\partial \theta^2}\bigg|_{\theta=\hat{\theta}} (\theta - \hat{\theta})^2.\end{aligned}$$

As the first derivative of the log-posterior evaluated at the mode is equal to zero, it can be discarded:

$$\log p(\theta|\mathcal{D}) \approx \log p(\hat{\theta}|\mathcal{D}) - \frac{\tau}{2}(\theta - \hat{\theta})^2, \tag{1}$$

where $\tau = -(\partial^2 \log p(\theta|\mathcal{D})/\partial\theta^2)|_{\theta=\hat{\theta}}$. Recall that the logarithm of a Gaussian density for θ with mean μ and variance σ^2 is given by:

$$C - \frac{1}{2\sigma^2}(\theta - \mu)^2, \tag{2}$$

where C is a normalization constant ensuring that the Gaussian density integrates to one. From (2), one recognizes that (1) is the Laplace approximation to $p(\theta|\mathcal{D})$ with mean $\mu = \hat{\theta}$ and variance equal to the inverse of the negative of the curvature of the posterior at the mode, i.e. $\sigma^2 = \tau^{-1}$.

A simple illustration

Consider the example of Robert (1996), where the density to be approximated is given by:

$$p(\theta|\theta_0) \propto \frac{\exp(-\theta^2/2)}{(1+(\theta-\theta_0)^2)^\nu}, \ \theta \in \mathbb{R}, \nu > 0,$$

Let us define $g(\theta|\theta_0) := g(\theta) = \log p(\theta|\theta_0)$. We thus have:

$$g(\theta) = \frac{-\theta^2}{2} - \nu \log\left(1+(\theta-\theta_0)^2\right).$$

$$\frac{\partial g(\theta)}{\partial \theta} = -\theta - \frac{2\nu(\theta-\theta_0)}{1+(\theta-\theta_0)^2}.$$

$$\frac{\partial^2 g(\theta)}{\partial \theta^2} = -\left(1 + \frac{2\nu(1+(\theta-\theta_0)^2) - 4\nu(\theta-\theta_0)^2}{(1+(\theta-\theta_0)^2)^2}\right).$$

It follows that the Laplace approximation to the target distribution is:

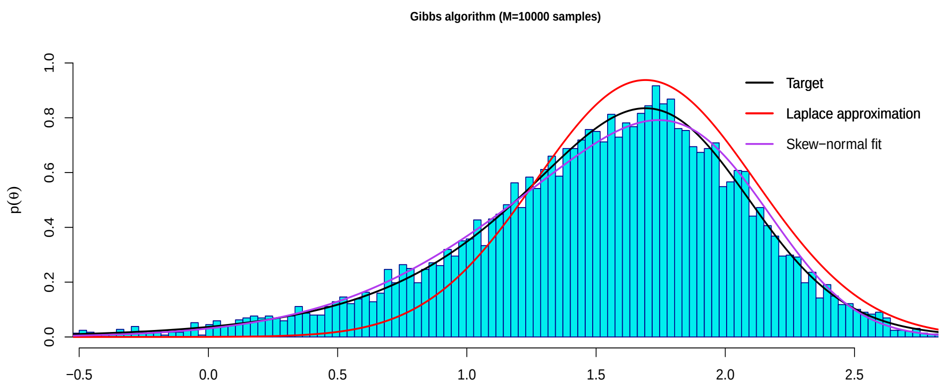
$$\widetilde{p}_G(\theta|\theta_0) = \mathcal{N}\left(\widehat{\theta}, \left(-\frac{\partial^2 g(\theta)}{\partial \theta^2}\right)^{-1}\Bigg|_{\theta=\widehat{\theta}}\right),$$

where $\widehat{\theta}$ is obtained numerically via a root finding technique. The code below compares the Laplace approximation, the skew-normal approximation and the Gibbs sampler as candidates to approximate the target density $p(\theta|\theta_0)$.

```
#-------------------------------------------------------------------------------
# Laplace approximation, skew-normal fit and
# Gibbs sampling by completion
# Copyright, Oswaldo Gressani. All rights reserved.
#------------------------------------------------------------------------------

theta0 <- 2
nu <- 3

#--- Target distribution to sample from
ptarget <- function(theta){
val <- exp(-0.5 * theta^2)/((1+(theta-theta0)^2)^nu)
return(val)
}
```

```
#--- Plot target
x <- seq(-0.55,3, length = 1000)
dx <- x[2]-x[1]
target_val <- sapply(x, ptarget)
cnorm <- 1/(sum(target_val * dx))
target_val <- cnorm * target_val
#sum(target_val * dx)
plot(x, target_val, type="l", xlab=expression(theta),
ylab="Posterior", lwd = 2)

#-- Gibbs sampler
theta_sample <- c()
eta_sample <- c()
theta_init <- 1
M <- 10000

for(m in 1:M){
eta_sample[m] <- rgamma(n = 1, shape = nu,
rate = 0.5 * (1+(theta_init-theta0)^2))
theta_sample[m] <- rnorm(n = 1,
mean = theta0 * (eta_sample[m]/(1+eta_sample[m])),
sd = sqrt(1/(1+eta_sample[m])))
theta_init <- theta_sample[m]
}

hist(theta_sample,
breaks=seq(min(theta_sample),max(theta_sample),length.out = 200),
freq = FALSE,
xlim = c(-0.4,2.7),
ylim=c(0,1),
ylab=expression(p(theta)),
xlab=expression(theta),
col="cyan2",
border = "darkblue",
main="Gibbs algorithm (M=10000 samples)",
cex.main=0.8)
lines(x, target_val, type="l", lwd = 2)


#-------------------------------------------------------------------------------
#                              Laplace approximation
#-------------------------------------------------------------------------------

logptarget <- function(theta){
val <- -(theta^2) * 0.5 - nu * log(1+(theta-theta0)^2)
return(val)
```

```
}

gradtar <- function(theta){
val <- (-1) * theta - (2 * nu * (theta-theta0))/(1+(theta-theta0)^2)
return(val)
}

hesstar <- function(theta){
val <- (-1) - (2 * nu * (1+(theta-theta0)^2) - 4 * nu * (theta-theta0)^2)/
((1+(theta-theta0)^2)^2)
return(val)
}

# Find mode of LA with uniroot
Laplace_mode <- stats::uniroot(gradtar, interval = c(-5,5))$root

# Variance of LA
Laplace_var <- ((-1) * hesstar(Laplace_mode))^(-1)

lines(x, dnorm(x, mean = Laplace_mode, sd = sqrt(Laplace_var)), col = "red",
lwd = 2)
legend("topright", c("Target", "Laplace approximation"), col = c("black","red"),
lty = c(1,1), lwd = c(2,2), bty = "n")

#--------------------------------------------------------------------------------
#                              Skew-normal fit
#--------------------------------------------------------------------------------

# Skew-normal fit
library("blapsr")

# Extract x an y coordinates from the target
xcoord <- x
ycoord <- target_val

# Computation of the skew-normal fit
snfit <- blapsr::snmatch(xcoord, ycoord)
lines(snfit$xgrid, snfit$snfit, type = "l", col = "darkorchid2", lwd = 2)
legend("topright", c("Target", "Laplace approximation", "Skew-normal fit"),
col = c("black","red", "darkorchid 2"),
lty = c(1,1, 1), lwd = c(2,2, 2), bty = "n")
```


Gibbs algorithm (M=10000 samples)
p(θ)
0.0
0.2
0.4
0.6
0.8
1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
Target
Laplace approximation
Skew−normal fit