

MATH 559 - ASSIGNMENT 3 - SOLUTIONS

1. The *Trinomial*(n, θ_1, θ_2) distribution is a bivariate distribution with pmf

$$p_{Y_1, Y_2}(y_1, y_2; \theta_1, \theta_2) = \frac{n!}{y_1! y_2! (n - y_1 - y_2)!} \theta_1^{y_1} \theta_2^{y_2} (1 - \theta_1 - \theta_2)^{n - y_1 - y_2} \quad 0 \leq y_1, y_2, y_1 + y_2 \leq n,$$

where $n \geq 1$ is a fixed integer, and parameters (θ_1, θ_2) are parameters with parameter space

$$\Theta = \{(\theta_1, \theta_2) : 0 < \theta_1, \theta_2, \theta_1 + \theta_2 < 1\}.$$

- (a) Find the posterior for (θ_1, θ_2) under the conjugate *Dirichlet*($\alpha_1, \alpha_2, \alpha_3$) prior, with pdf

$$\pi_0(\theta_1, \theta_2) = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} \theta_1^{\alpha_1-1} \theta_2^{\alpha_2-1} (1 - \theta_1 - \theta_2)^{\alpha_3-1}$$

with support Θ , where $\alpha_1, \alpha_2, \alpha_3 > 0$ are hyperparameters.

SOLUTION: We have that, on the support of the prior,

$$\begin{aligned} \pi_n(\theta_1, \theta_2) &\propto \{\theta_1^{y_1} \theta_2^{y_2} (1 - \theta_1 - \theta_2)^{n - y_1 - y_2}\} \times \{\theta_1^{\alpha_1-1} \theta_2^{\alpha_2-1} (1 - \theta_1 - \theta_2)^{\alpha_3-1}\} \\ &= \theta_1^{y_1 + \alpha_1 - 1} \theta_2^{y_2 + \alpha_2 - 1} (1 - \theta_1 - \theta_2)^{n - y_1 - y_2 + \alpha_3 - 1} \end{aligned}$$

and so we may deduce that

$$\pi_n(\theta_1, \theta_2) \equiv \text{Dirichlet}(y_1 + \alpha_1, y_2 + \alpha_2, n - y_1 - y_2 + \alpha_3)$$

3 MARKS

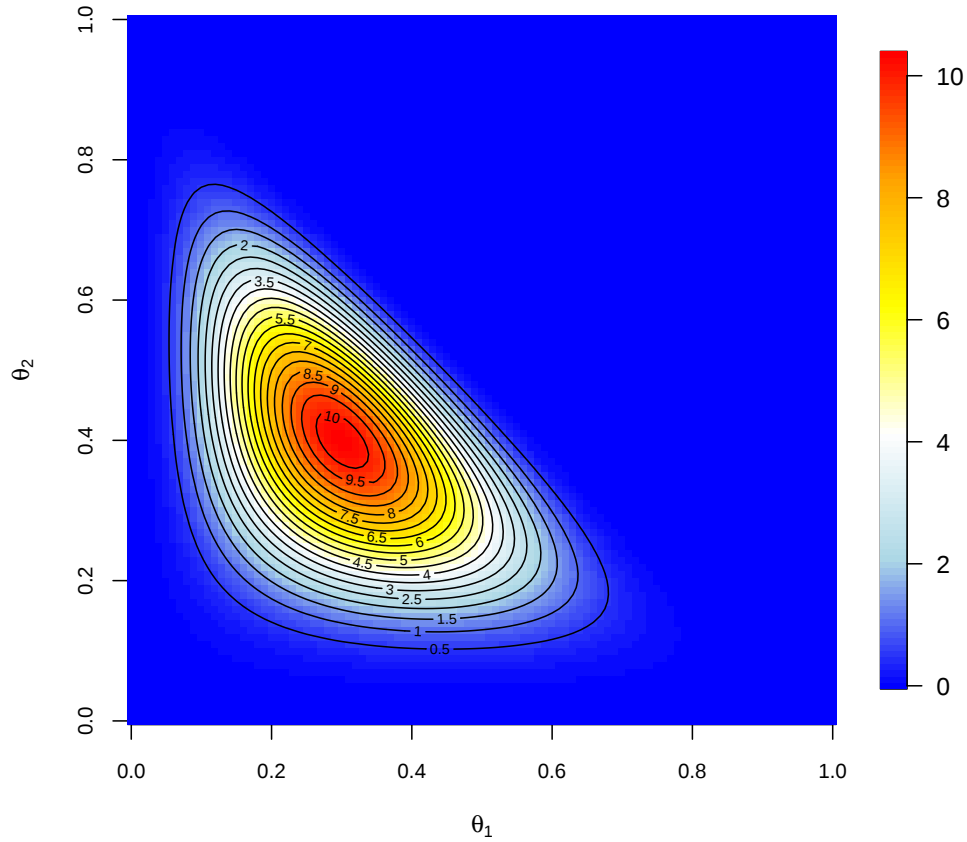
- (b) Plot the joint posterior if $\alpha_1 = \alpha_2 = \alpha_3 = 1$, and $n = 10, y_1 = 3, y_2 = 4$.

SOLUTION: Using the code provided (any plotting method is allowed):

```
#Dirichlet pdf
Dir.post<-function(th1,th2,y1v,y2v,nv,a1v=1,a2v=1,a3v=1){

  if(th1+th2 >= 1){
    return(0.0)
  }else{
    c1<-gamma(nv+a1v+a2v+a3v)
    c2<-(gamma(a1v+y1v)*gamma(a2v+y2v)*gamma(nv-y1v-y2v+a3v))
    dval<-c1*exp(y1v*log(th1)+y2v*log(th2)+(nv-y1v-y2v)*log(1-th1-th2))/c2
  }
  return(dval)
}
f <- Vectorize(Dir.post,vectorize.args=c("th1","th2"))
th1v<-seq(0.0,1,by=0.01)
th2v<-seq(0.0,1,by=0.01)
y1<-3;y2<-4;n<-10
dmat<-outer(th1v,th2v,f,y1v=y1,y2v=y2,nv=n)

library(fields,quietly=TRUE)
par(pty='s',mar=c(4,3,2,2))
colfunc <- colorRampPalette(c("blue","lightblue","white","yellow","orange","red"))
image.plot(th1v,th2v,dmat,col=colfunc(100),
           xlab=expression(theta[1]),ylab=expression(theta[2]),cex.axis=0.8)
contour(th1v,th2v,dmat,add=T,nlevels=20)
```



4 MARKS

(c) Find the marginal posterior for

$$\phi = \frac{\theta_1}{\theta_1 + \theta_2}$$

under the prior in (a).

SOLUTION: We define the bivariate 1-1 transformation,

$$\phi = \frac{\theta_1}{\theta_1 + \theta_2} \quad \psi = \theta_1 + \theta_2$$

so that for the inverse transformations we have $\theta_1 = \psi\phi$, $\theta_2 = \psi(1 - \phi)$. Note that in the new parameterization, we have that the support of the new prior/posterior will be

$$\{(\phi, \psi) : 0 < \phi < 1, 0 < \psi < 1\} \equiv (0, 1) \times (0, 1).$$

The Jacobian of the transformation is

$$\begin{vmatrix} \psi & -\psi \\ \phi & 1 - \phi \end{vmatrix} = \psi$$

and therefore the posterior for the new parameters is

$$\pi_n^*(\phi, \psi) = \pi_n(\psi\phi, \psi(1 - \phi))\psi \quad (\phi, \psi) \in (0, 1) \times (0, 1)$$

which we compute as

$$\begin{aligned} \pi_n^*(\phi, \psi) &\propto (\psi\phi)^{y_1 + \alpha_1 - 1} (\psi(1 - \phi))^{y_2 + \alpha_2 - 1} (1 - \psi\phi - \psi(1 - \phi))^{n - y_1 - y_2 + \alpha_3 - 1} \psi \\ &= \{\phi^{y_1 + \alpha_1 - 1} (1 - \phi)^{y_2 + \alpha_2 - 1}\} \{\psi^{y_1 + y_2 + \alpha_1 + \alpha_2 - 2} (1 - \psi)^{n - y_1 - y_2 + \alpha_3}\} \end{aligned}$$

Therefore we can deduce that the marginal posterior for ϕ is $Beta(y_1 + \alpha_1, y_2 + \alpha_2)$, as the joint posterior factorizes into the product of the marginal for ϕ and the marginal for ψ . 5 MARKS

2. The Gibbs posterior for iid data drawn from distribution F_0 using prior $\pi_0^\dagger(\theta)$ is formed by computing the density

$$\pi_n^\dagger(\theta) = \frac{\exp\left\{-\eta \sum_{i=1}^n \ell(y_i, \theta)\right\} \pi_0^\dagger(\theta)}{\int \exp\left\{-\eta \sum_{i=1}^n \ell(y_i, t)\right\} \pi_0^\dagger(t) dt}$$

defined when the denominator is finite, where $\ell(y, \theta)$ is a non-negative function from $\mathcal{Y} \times \Theta$ to $\mathbb{R}^+$, and η is a fixed positive constant. The true value of the parameter, θ_0 , is defined by

$$\theta_0 = \arg \min_t \int \ell(y, t) dF_0(y)$$

that is, it is the loss-minimizing value of the parameter.

- (a) Suppose that $\ell(y, \theta)$ is at least three times differentiable with respect to θ for almost all y (that is, the set of y values for which the function is NOT differential contains probability equal to zero under F_0) at each $\theta \in \Theta$. Suppose that θ_0 lies in an open subset of Θ .

Describe the behaviour of $\pi_n^\dagger(\theta)$ as $n \rightarrow \infty$.

SOLUTION: Under these conditions, the loss function can be approximated in the same way as a regular log density (or log likelihood) function, that is, using a quadratic Taylor expansion. Therefore, provided the support of the prior includes θ_0 , the posterior will concentrate at θ_0 as $n \rightarrow \infty$. Under these conditions, when n is large,

$$\frac{1}{n} \sum_{i=1}^n \ell(y_i, \theta)$$

is minimized at $\theta = \theta_0$ with probability tending to 1, by the result from lectures. In addition, we may construct a Normal approximation to $\pi_n^\dagger$ by using the quadratic expansion.

Note that θ_0 may not be uniquely defined in general (for example if F_0 is a discrete distribution), in which case these results hold for one of the true loss minimizers.

4 MARKS

- (b) If $\mathcal{Y} \equiv \Theta \equiv \mathbb{R}$ and $\ell(y, \theta) = |y - \theta|$ show that the Gibbs posterior is equivalent to a standard Bayesian posterior under a particular parametric assumption.

SOLUTION: In this case

$$\exp\left\{-\eta \sum_{i=1}^n \ell(y_i, \theta)\right\} = \exp\left\{-\eta \sum_{i=1}^n |y_i - \theta|\right\}$$

suggesting that for this to be a standard Bayesian posterior, we would need the density

$$f_Y(y; \theta) \propto \exp\{-\eta|y - \theta|\}.$$

But this is a valid pdf on $\mathbb{R}$ that takes the form

$$f_Y(y; \theta) = \frac{\eta}{2} \exp\{-\eta|y - \theta|\} \quad y \in \mathbb{R}$$

which is known as the Laplace or Double Exponential distribution with location parameter θ . Here η is treated as a **known** scale parameter. 2 MARKS

- (c) If, in fact $F_0(y)$ is an *Exponential*(1) distribution, describe the behaviour of $\pi_n^\dagger(\theta)$ as $n \rightarrow \infty$ for the loss function in (b).

SOLUTION: In this case, we have from results proven in lectures that θ_0 is the **median** of F_0 , which for this distribution is the value $\log 2 = 0.69310$. Therefore the Gibbs posterior concentrates at this value as $n \rightarrow \infty$, provided this value lies in the support of $\pi_0^\dagger$. 2 MARKS

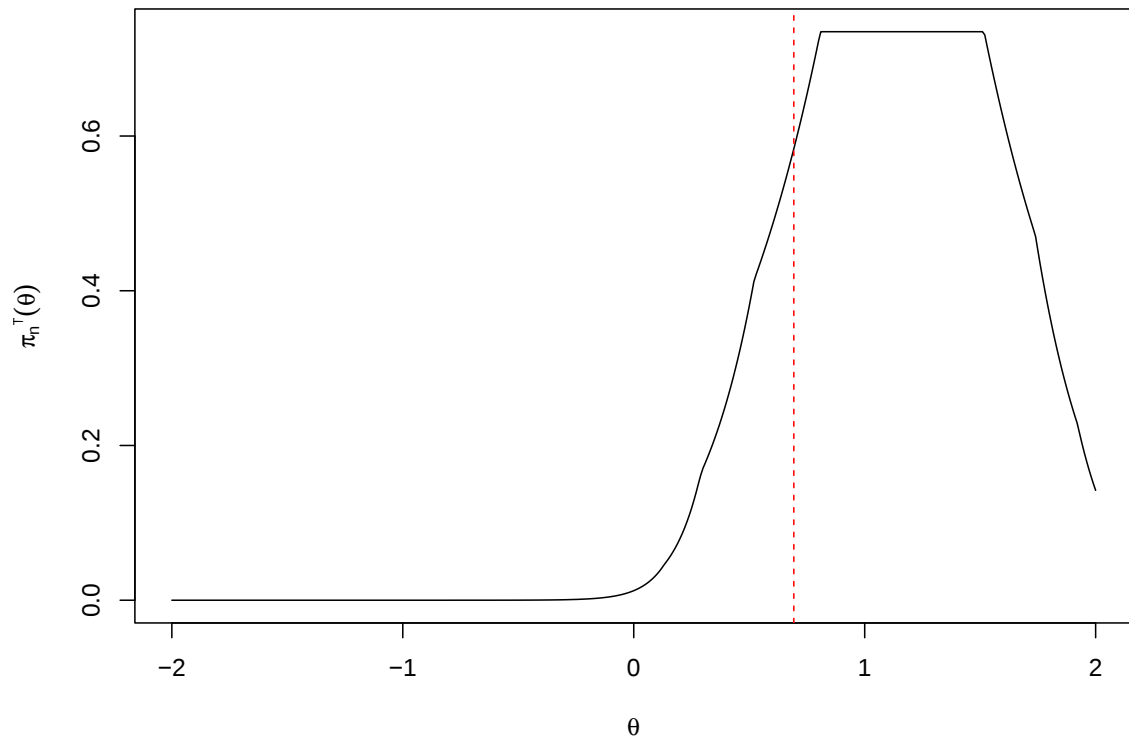
Illustration of this result (not required in solution): we may use (for convenience) the Jeffreys prior for θ , which is constant on $\mathbb{R}$ and therefore improper, but yields a valid posterior; we may also choose $\eta = 1$ for illustration.

$$\pi_n^\dagger(\theta) = \frac{\exp\left\{-\sum_{i=1}^n |y_i - \theta|\right\}}{\int_{-\infty}^{\infty} \exp\left\{-\sum_{i=1}^n |y_i - t|\right\} dt}$$

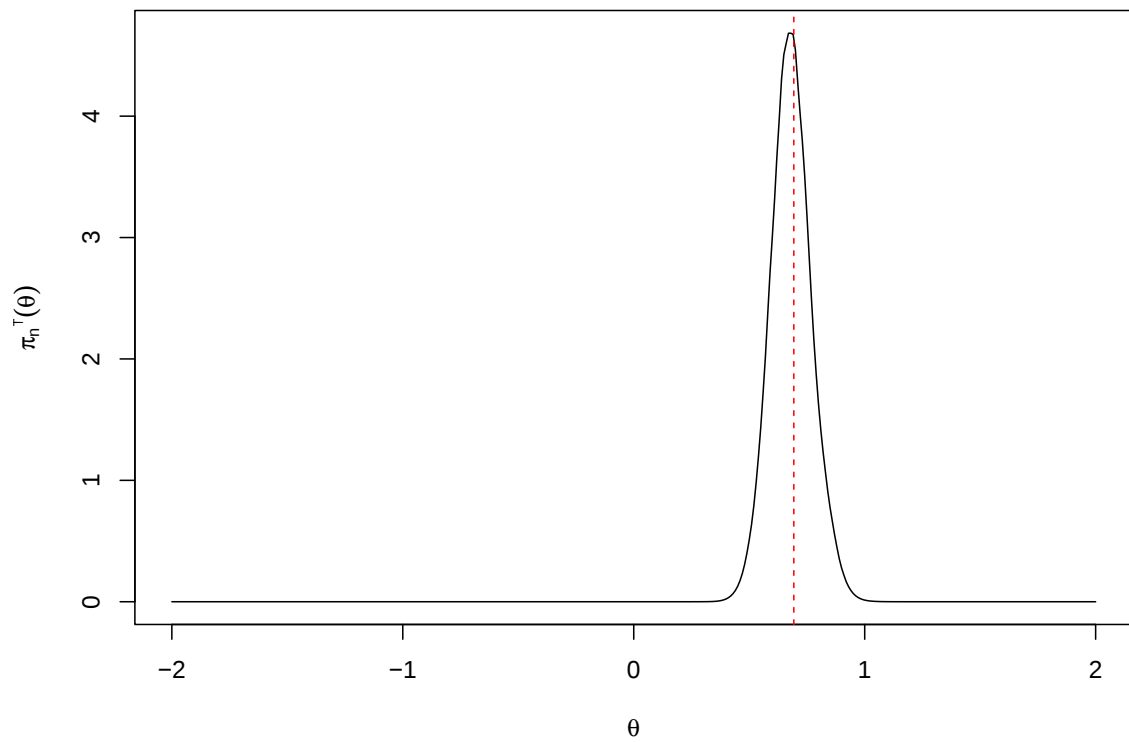
```
Gpost<-function(thv,yv,ev=1){
  return(exp(-ev*sum(abs(yv-thv))))
}
GpostI<-function(thv,yv,ev=1){
  Iv<-thv
  for(i in 1:length(thv)){
    Iv[i]<-exp(-ev*sum(abs(yv-thv[i])))
  }
  return(Iv)
}
set.seed(1101)
yl<-expression({{pi[n]}^"\u2020"}(theta))

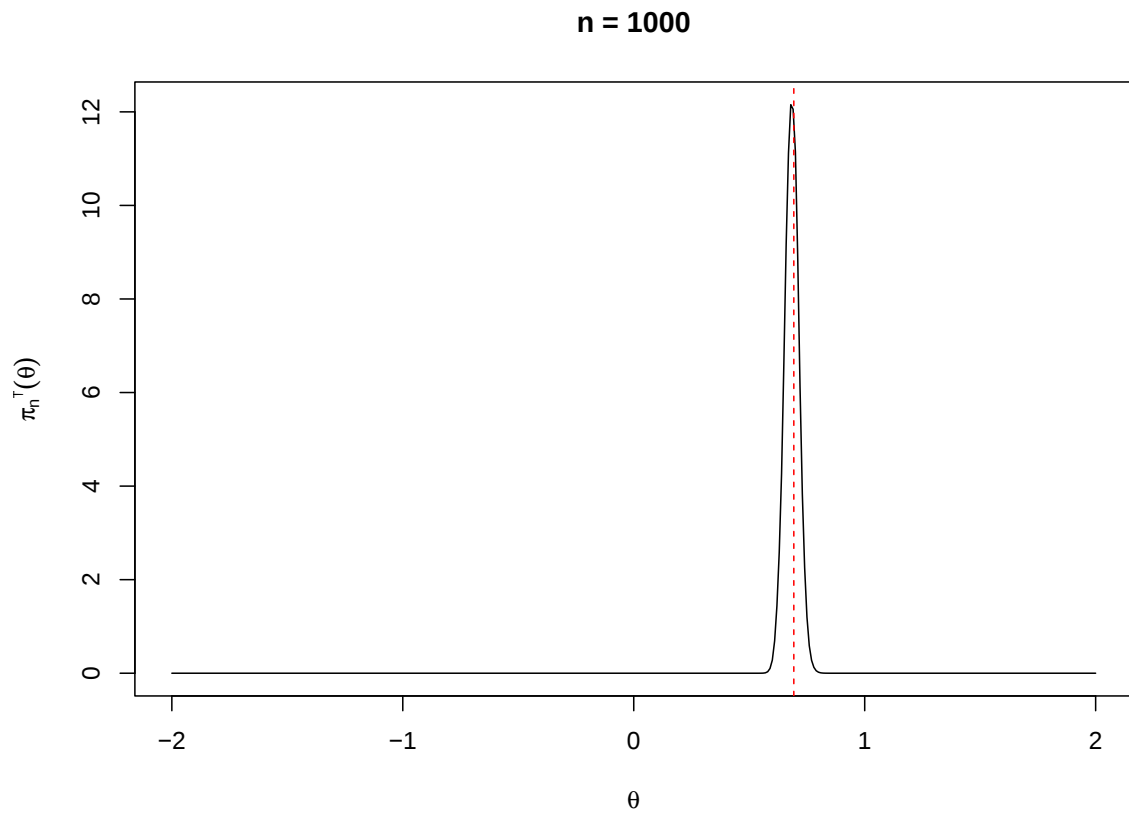
for(n in c(10,100,1000)){
  Y<-rexp(n)
  th<-seq(-2,2,by=0.01)
  Gy<-th*0
  Ival<-integrate(GpostI,lower=-Inf,upper=Inf,yv=Y)
  Ival
  for(j in 1:length(th)){
    Gy[j]<-Gpost(th[j],Y)
  }
  plot(th,Gy/Ival$value,type='l',ylab=yl,xlab=expression(theta))
  title(paste('n =',n))
  abline(v=log(2),col='red',lty=2)
}
```

n = 10

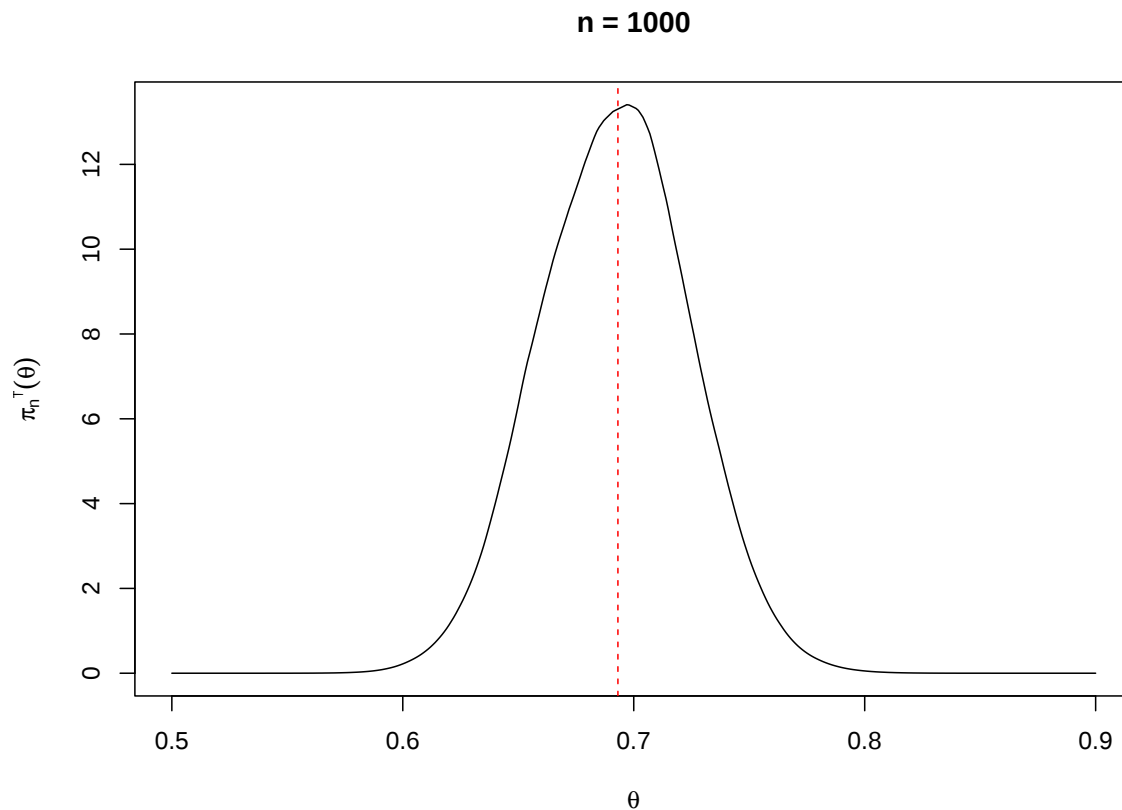


n = 100





```
th<-seq(0.5,0.9,by=0.001)
set.seed(3)
n<-1000
Y<-rexp(n)
Gy<-th*0
Ival<-integrate(GpostI,lower=-Inf,upper=Inf,yv=Y)
for(j in 1:length(th)){
  Gy[j]<-Gpost(th[j],Y)
}
plot(th,Gy/Ival$value,type='l',ylab=y1,xlab=expression(theta))
title(paste('n =',n))
abline(v=log(2),col='red',lty=2)
```



Addendum: The Gibbs posterior based on the absolute error loss can be approximated by a Normal distribution, but some care is needed in constructing the approximation. The function

$$\ell(y, t) = |y - t|$$

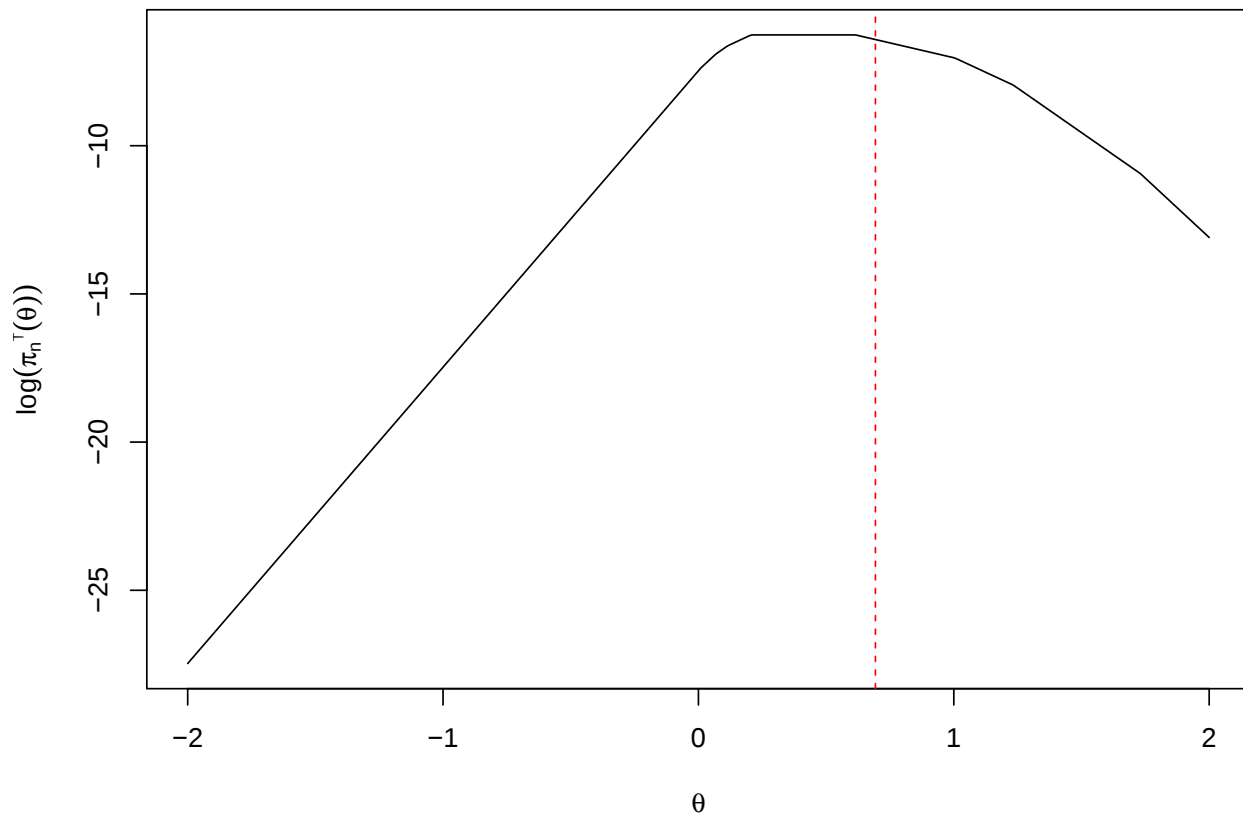
is not differentiable wrt t at $t = y$, but is differentiable everywhere else. The log-posterior

$$\log \pi_n^\dagger(\theta) = - \sum_{i=1}^n |y_i - \theta| + \text{const.}$$

is a piecewise linear function that is well approximated by a quadratic when n is large.

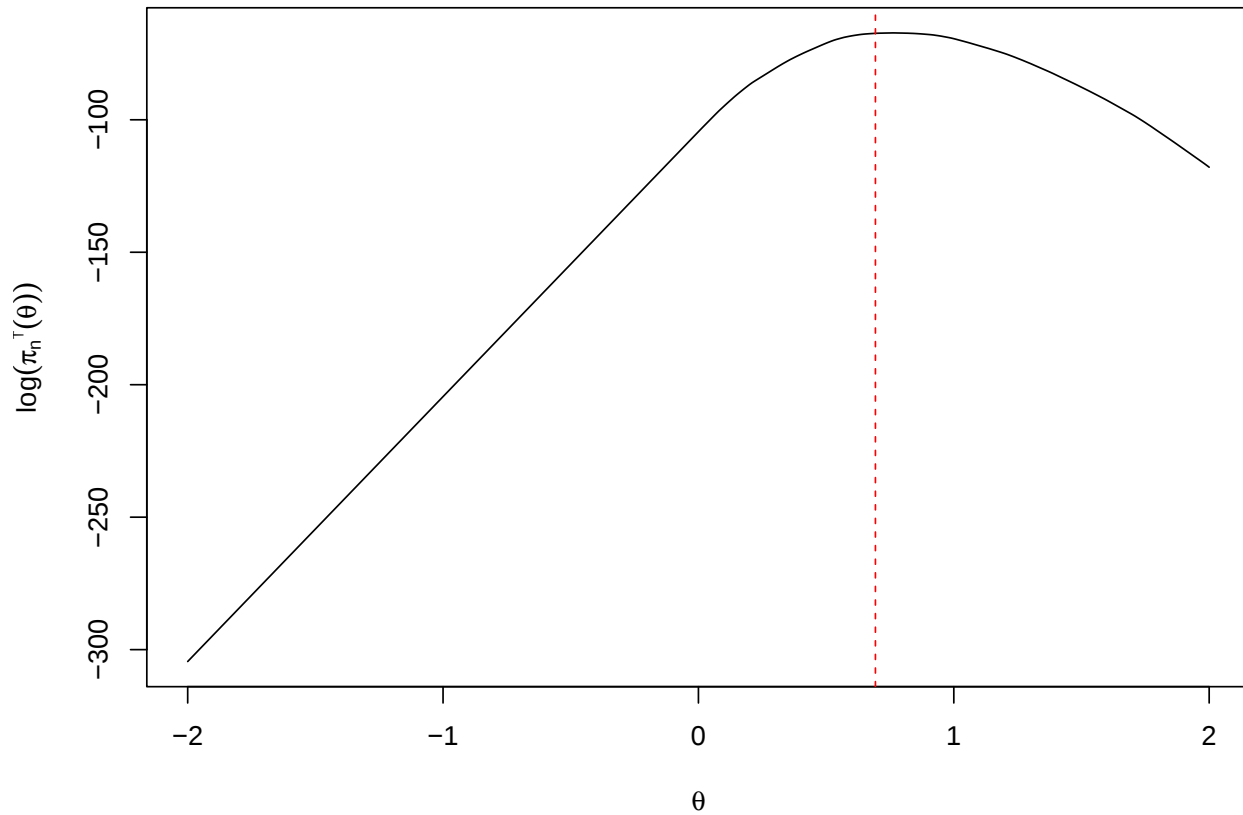
```
th<-seq(-2,2,by=0.01)
set.seed(3)
n<-10
Y<-rexp(n)
lGy<-th*0
for(j in 1:length(th)){
  lGy[j]<-log(Gpost(th[j],Y))
}
yll<-expression(log({{pi[n]}^"\u2020"(theta)}))
plot(th,lGy,type='l',ylab=yll,xlab=expression(theta))
title(paste('n =',n))
abline(v=log(2),col='red',lty=2)
```

n = 10

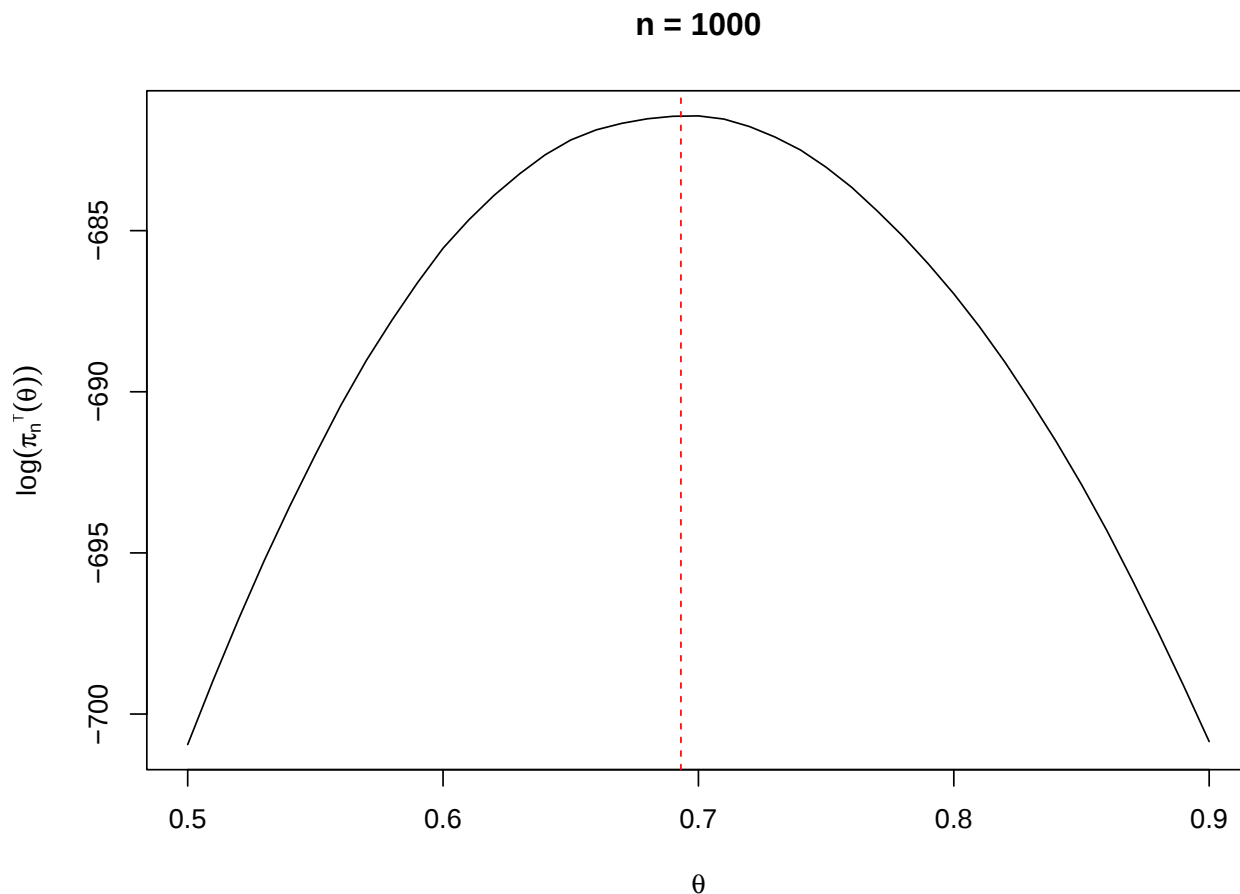


```
th<-seq(-2,2,by=0.01)
set.seed(3)
n<-100
Y<-rexp(n)
lGy<-th*0
for(j in 1:length(th)){
  lGy[j]<-log(Gpost(th[j],Y))
}
yll<-expression(log({{pi[n]}^"\u2020"}(theta)))
plot(th,lGy,type='l',ylab=yll,xlab=expression(theta))
title(paste('n =',n))
abline(v=log(2),col='red',lty=2)
```

n = 100



```
th<-seq(0.5,0.9,by=0.01)
set.seed(3)
n<-1000
Y<-rexp(n)
lGy<-th*0
for(j in 1:length(th)){
  lGy[j]<-log(Gpost(th[j],Y))
}
yll<-expression(log({{pi[n]}^"\u2020"}(theta)))
plot(th,lGy,type='l',ylab=yll,xlab=expression(theta))
title(paste('n =',n))
abline(v=log(2),col='red',lty=2)
```



However, writing $|y - \theta| = \sqrt{(y - \theta)^2}$, we have that at θ where the derivatives exist,

$$\dot{\ell}(y, \theta) = - \sum_{i=1}^n \text{sgn}|y_i - \theta| \quad \text{sgn}(x) = -\mathbb{1}_{(-\infty, 0)}(x) + \mathbb{1}_{(0, \infty)}(x)$$

$$\ddot{\ell}(y, \theta) = 0 \quad \forall \theta$$

so the second derivative cannot be used to approximate the log-posterior. Instead we may use the frequentist theory and replace the second derivative at θ by using the square of the first derivative. Here, we have

$$\sum_{i=1}^n \{\dot{\ell}(y, \theta)\}^2 = n$$

suggesting the approximation

$$\log \pi_n^\dagger(\theta) = \log \pi_n^\dagger(\hat{\theta}) - \frac{n}{2}(\theta - \hat{\theta})^2$$

where $\hat{\theta}$ is the posterior mode, so that

$$\pi_n^\dagger(\theta) \approx \text{Normal}(\hat{\theta}, 1/n).$$

```

th<-seq(0.5,0.9,by=0.01)
set.seed(3)
n<-1000
Y<-rexp(n)
lGy<-th*0
for(j in 1:length(th)){
  lGy[j]<-log(Gpost(th[j],Y))
}
yll<-expression(log({{pi[n]}^"\u2020"}(theta)))
plot(th,lGy,type='l',ylab=yll,xlab=expression(theta))
lines(th,max(lGy)-0.5*n*(th-th[which.max(lGy)])^2,col='red')
title(paste('n =',n))
abline(v=log(2),col='red',lty=2)
legend(0.785,max(lGy),c('Exact','Quadratic Appr.'),col=c('black','red'),lty=1)

```

