ST4234: Bayesian Statistics

Tutorial 2 Solution, AY 19/20

Solutions

1. Let Y denote the number of re-offenders out of 43 individuals released from incarceration. $Y \sim \text{Binomial}(43, \theta)$. In this study, $Y = 15$ is observed.

   (a) Using a Beta(2,8) prior for θ, the posterior $p(\theta|y)$ is $\text{Beta}(2 + 15, 8 + 43 - 15) = \text{Beta}(17, 36)$. The posterior mean is $\frac{17}{17+36} = 0.3208$. The posterior mode is $\frac{16}{17+36-2} = 0.3137$. The posterior standard deviation is

   $$\sqrt{\frac{17 \cdot 36}{(17 + 36)^2(17 + 36 + 1)}} = 0.0635.$$

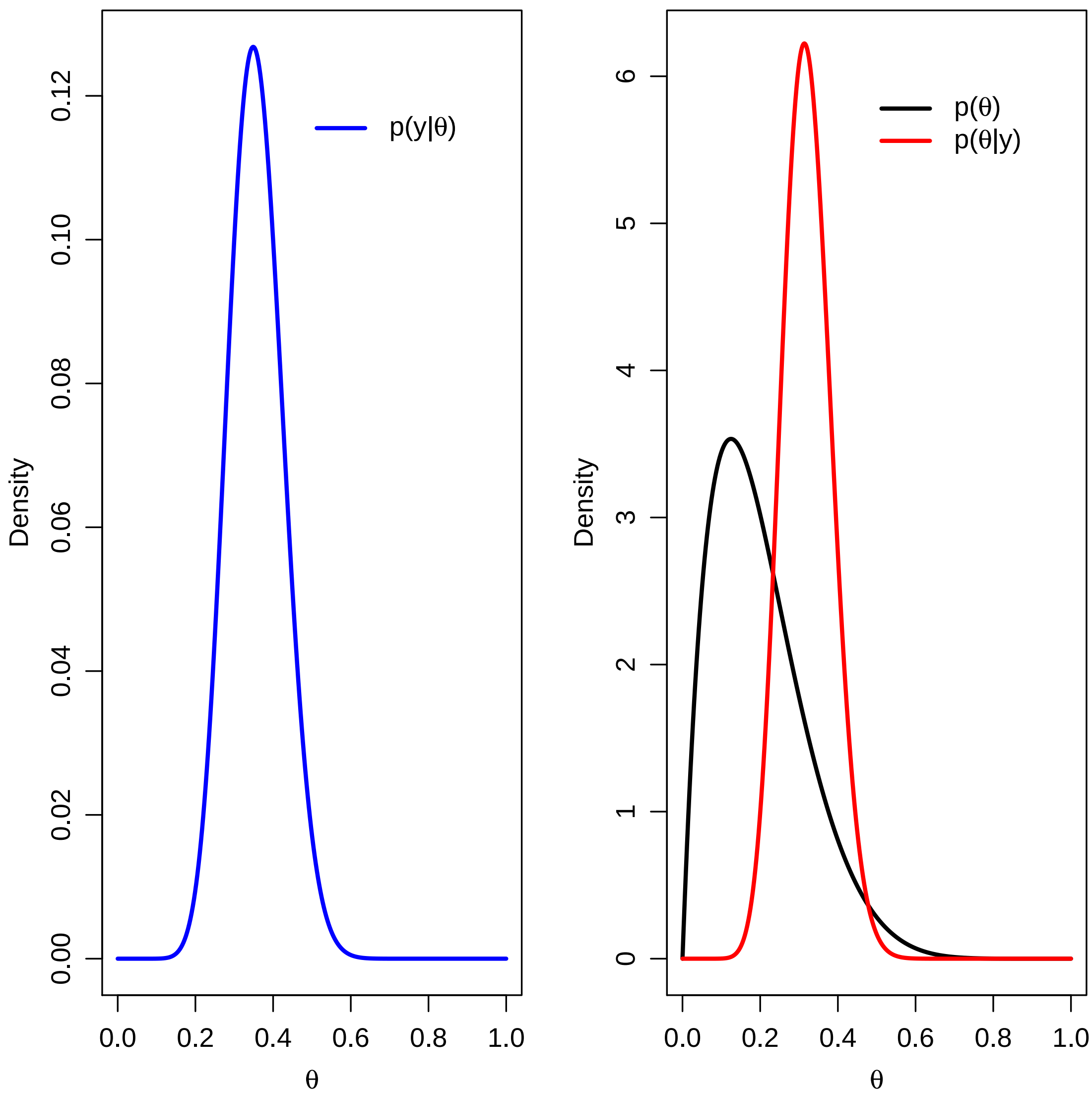
   A 95% quantile-based confidence interval is [0.2033, 0.4510].

```
qbeta(c(0.025,0.975), 17, 36)

## [1] 0.2032978 0.4510240

theta <- seq(from=0,to=1,by=0.001)
y <- 15;  n <- 43
a0 <- 2;  b0 <- 8
a <- a0 + y;  b <- b0 + n - y
likelihood <- dbinom(y, n, theta)
par(mar=c(3.5,3.5,1,1))
par(mgp=c(2.1,0.8,0))
par(mfrow=c(1,2))
plot(theta, likelihood, type="l", lwd=2.5, col="blue",
     xlab=expression(theta), ylab="Density")
legend(0.45,0.12,legend=expression(paste("p(y|", theta, ")")),
       lty=1, lwd=2.5, cex=1, col="blue", bty="n")
plot(theta, dbeta(theta,a0,b0), type="l", lwd=2.5,
```

```
      xlab=expression(theta), ylab="Density", ylim=c(0,6.2))
points(theta, dbeta(theta,a,b), type="l", lwd=2.5, col="red")
legend(0.45,6,legend = c(expression(paste("p(", theta, ")")),
       expression(paste("p(", theta, "|y)"))), lty=1,
       lwd=2.5, cex=1, col=c("black","red"), bty="n")
```



(b) Using a Beta(8,2) prior for θ, the posterior $p(\theta|y)$ is $\text{Beta}(8+15, 2+43-15) = \text{Beta}(23, 30)$. The posterior mean is $\frac{23}{23+30} = 0.4340$. The posterior mode is $\frac{23-1}{23+30-2} =$

0.4314. The posterior standard deviation is

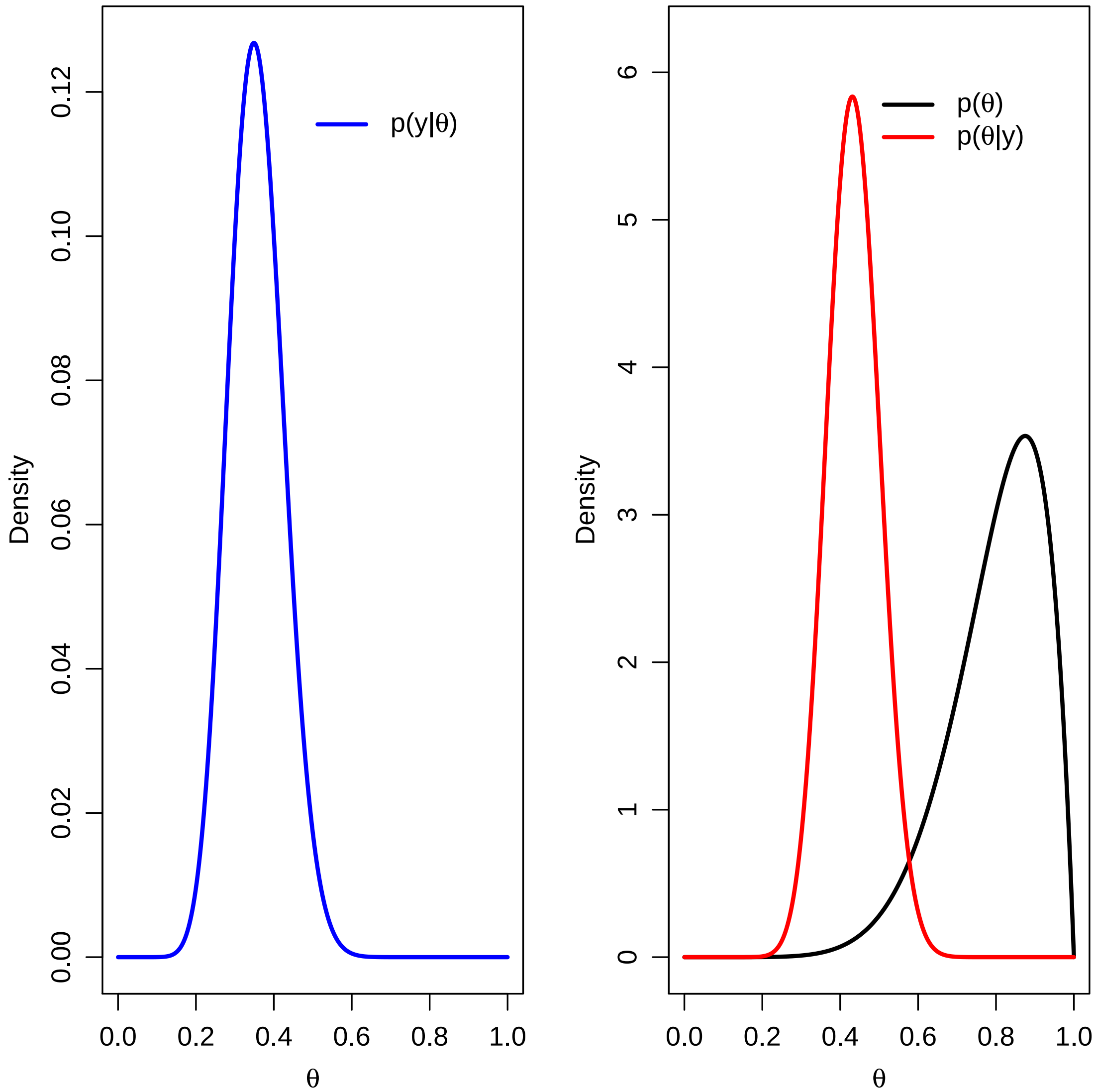
$$\sqrt{\frac{23 \cdot 30}{(23+30)^2(23+30+1)}} = 0.0674.$$

A 95% quantile-based confidence interval is [0.3047, 0.5680].

```
qbeta(c(0.025,0.975), 23, 30)

## [1] 0.3046956 0.5679528

theta <- seq(from=0,to=1,by=0.001)
y <- 15;  n <- 43
a0 <- 8;  b0 <- 2
a <- a0 + y;  b <- b0 + n - y
likelihood <- dbinom(y, n, theta)
par(mar=c(3.5,3.5,1,1))
par(mgp=c(2.1,0.8,0))
par(mfrow=c(1,2))
plot(theta, likelihood, type="l", lwd=2.5, col="blue",
     xlab=expression(theta), ylab="Density")
legend(0.45,0.12,legend=expression(paste("p(y|", theta, ")")),
       lty=1, lwd=2.5, cex=1, col="blue", bty="n")
plot(theta, dbeta(theta,a0,b0), type="l", lwd=2.5,
     xlab=expression(theta), ylab="Density", ylim=c(0,6.2))
points(theta, dbeta(theta,a,b), type="l", lwd=2.5, col="red")
legend(0.45,6,legend = c(expression(paste("p(", theta, ")")),
       expression(paste("p(", theta, "|y)"))), lty=1,
       lwd=2.5, cex=1, col=c("black","red"), bty="n")
```



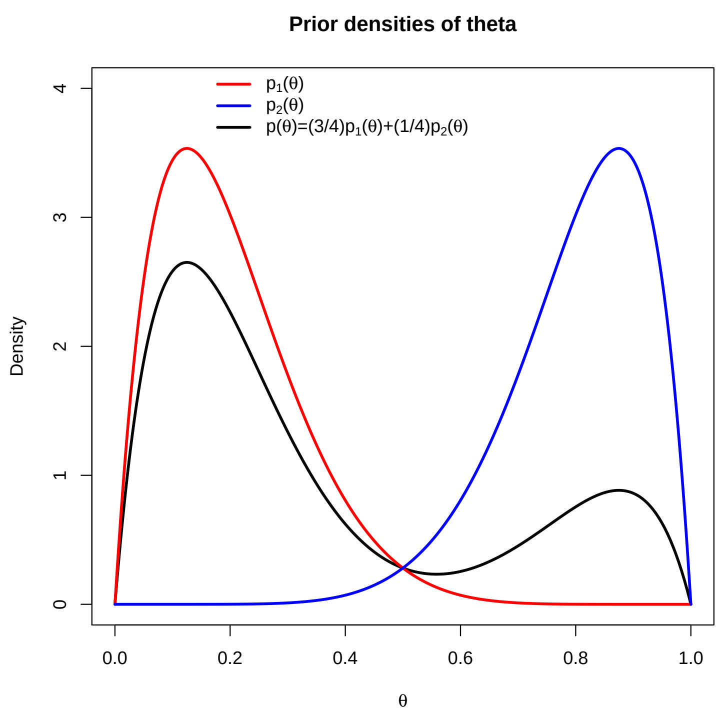
(c) This prior density can be rewritten as the following mixture density

$$
\begin{aligned}
p(\theta) &= \frac{3}{4}\frac{\Gamma(10)}{\Gamma(2)\Gamma(8)}\theta(1-\theta)^7 + \frac{1}{4}\cdot\frac{\Gamma(10)}{\Gamma(8)\Gamma(2)}\theta^7(1-\theta)] \\
&= \frac{3}{4}p_1(\theta) + \frac{1}{4}p_2(\theta),
\end{aligned}
$$

where $p_1(\theta)$ is the density of $\text{Beta}(2,8)$ and $p_2(\theta)$ is the density of $\text{Beta}(8,2)$.

The plots of the two prior in (a), (b), and (c) are plotted below in the same plot.

```
beta_mixture <- function(theta){
        (3/4)*dbeta(theta,2,8) + (1/4)*dbeta(theta,8,2)
}
theta <- seq(from=0, to=1, by=0.001)
plot(theta, beta_mixture(theta), type="l", lwd=2.5,
     ylim=c(0,4), xlab=expression(theta), ylab="Density",
         main="Prior densities of theta")
points(theta, dbeta(theta,2,8), type="l", lwd=2.5, col="red")
points(theta, dbeta(theta,8,2), type="l", lwd=2.5, col="blue")
legend(0.15, 4.2, legend = c(expression(paste("p"[1], "(", theta, ")")),
       expression(paste("p"[2], "(", theta, ")")),
           expression(paste("p(",theta,")=(3/4)p"[1],"(",theta,")",
           "+(1/4)p"[2],"(",theta,")"))),
           lty=1, lwd=2.5, cex=1, bty="n", col=c("red","blue","black"))
```



This prior may represents a prior opinion that 75% of the individuals released from incarceration has an average probability of 0.2 (with a mode of 0.125) of relapse into criminal behavior while the remaining 25% has a higher average probability of 0.8 (with a mode of 0.875) of relapse. Alternatively, this prior may represents a prior opinion that for 75% of the individuals released from incarceration, 2 out of 10 will relapse into criminal behavior while for the remaining 25%, 8 out of 10 will relapse into criminal behavior.

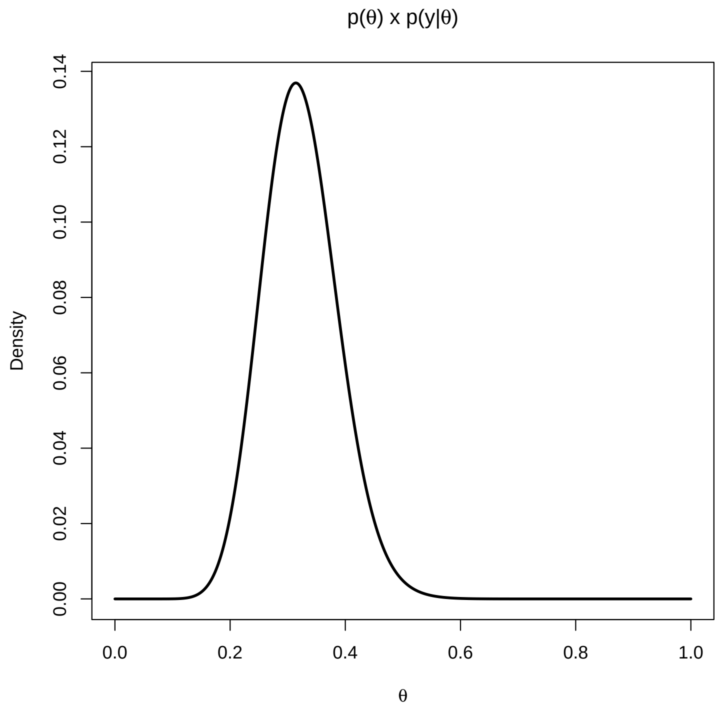
(d) (i)

$$
\begin{aligned}
p(\theta) \times p(y|\theta) &= \frac{\Gamma(10)}{4\Gamma(2)\Gamma(8)}[3\theta(1-\theta)^7 + \theta^7(1-\theta)] \cdot \binom{43}{15}\theta^{15}(1-\theta)^{28} \\
&= \frac{9!}{4(7!)}\binom{43}{15}[3\theta^{16}(1-\theta)^{35} + \theta^{22}(1-\theta)^{29}] \\
&= 18\binom{43}{15}[3\theta^{16}(1-\theta)^{35} + \theta^{22}(1-\theta)^{29}]. \qquad (1)
\end{aligned}
$$

(ii) The posterior distribution is a mixture of Beta(17,36) and Beta(23,30).

(iii) The posterior mode is 0.3141 and it lies between the modes in (a) and (b). It is much closer to the mode in (a), which is expected as the prior in (a) has a much higher weightage than the prior in (b).

```
y <- 15
n <- 43
posterior_prop <- function(theta,y,n){
        beta_mixture(theta) * dbinom(y,n,theta)
}
plot(theta, posterior_prop(theta,y=y,n=n), type="l", lwd=2.5,
        xlab=expression(theta), ylab="Density",
        main=expression(paste("p(", theta, ") x p(y|", theta, ")")))
```




```
(posterior_mode <- optimize(posterior_prop, interval=c(0,1),
 maximum=TRUE, n=n, y=y))
```

```
## $maximum
## [1] 0.3140734
##
## $objective
## [1] 0.1369141
```

(iv) Continued from (1),

$$\begin{aligned} p(\theta|y) &\propto p(\theta)p(y|\theta) \\ &\propto 3\theta^{16}(1-\theta)^{35} + \theta^{22}(1-\theta)^{29} \\ \Rightarrow p(\theta|y) &= \frac{3\theta^{16}(1-\theta)^{35} + \theta^{22}(1-\theta)^{29}}{C} \\ \int_0^1 p(\theta|y)\,\mathrm{d}\theta = 1 \Rightarrow C &= \int_0^1 3\theta^{16}(1-\theta)^{35} + \theta^{22}(1-\theta)^{29}\,\mathrm{d}\theta \end{aligned}$$

$$\Rightarrow C = 3\mathrm{B}(17,36) + \mathrm{B}(23,30),$$

where $\mathrm{B}(a,b)$ is the beta function. Therefore

$$\begin{aligned}
p(\theta|y) &= \frac{3\theta^{16}(1-\theta)^{35} + \theta^{22}(1-\theta)^{29}}{3\mathrm{B}(17,36)+\mathrm{B}(23,30)} \\
&= \frac{3\mathrm{B}(17,36)}{3\mathrm{B}(17,36)+\mathrm{B}(23,30)} \cdot \frac{\theta^{16}(1-\theta)^{35}}{\mathrm{B}(17,36)} \\
&\quad + \frac{\mathrm{B}(23,30)}{3\mathrm{B}(17,36)+\mathrm{B}(23,30)} \cdot \frac{\theta^{22}(1-\theta)^{29}}{\mathrm{B}(23,30)}.
\end{aligned}$$

Hence, the weights of the mixture distribution in (ii) are $w_1 = \frac{3\mathrm{B}(17,36)}{3\mathrm{B}(17,36)+\mathrm{B}(23,30)}$ and $w_2 = \frac{\mathrm{B}(23,30)}{3\mathrm{B}(17,36)+\mathrm{B}(23,30)}$ corresponding to the two posteriors Beta(17,36) and Beta(23,30) respectively. These two weights sum to one and are proportional to the product of the respective prior weights and prior predictive distributions (recall that prior predictive distribution is defined as $\int p(y|\theta)p(\theta)\,\mathrm{d}\theta$).

(e) The posterior mean is given by

$$\begin{aligned}
\mathrm{E}(\theta|y) &= \int \theta p(\theta|y)\,\mathrm{d}\theta \\
&= w_1 \int_0^1 \theta \frac{\theta^{16}(1-\theta)^{35}}{\mathrm{B}(17,36)}\,\mathrm{d}\theta + w_2 \int_0^1 \theta \frac{\theta^{22}(1-\theta)^{29}}{\mathrm{B}(23,30)}\,\mathrm{d}\theta \\
&= w_1 \frac{17}{17+36} + w_2 \frac{23}{23+30} = \frac{51\mathrm{B}(17,36)+23\mathrm{B}(23,30)}{53[3\mathrm{B}(17,36)+\mathrm{B}(23,30)]}.
\end{aligned}$$

The posterior mean is a weighted average of the means in (a) and (b) with the weights given by w_1 and w_2 respectively.

2. (a)

$$\begin{aligned}
\mathrm{E}(y) &= \int_0^\infty y p(y)\,\mathrm{d}y \\
&= \int_0^\infty \frac{2}{\Gamma(a)} \theta^{2a} y^{2a} \exp(-\theta^2 y^2)\,\mathrm{d}y \\
&= \frac{\theta^{2a}\Gamma(a+1/2)}{\Gamma(a)\theta^{2a+1/2}} \underbrace{\int_0^\infty \frac{2}{\Gamma(a+1/2)} \theta^{2(a+1/2)} y^{2(a+1/2)-1} \exp(-\theta^2 y^2)\,\mathrm{d}y}_{=1}
\end{aligned}$$

$$= \frac{\Gamma(a+1/2)}{\theta\Gamma(a)}.$$

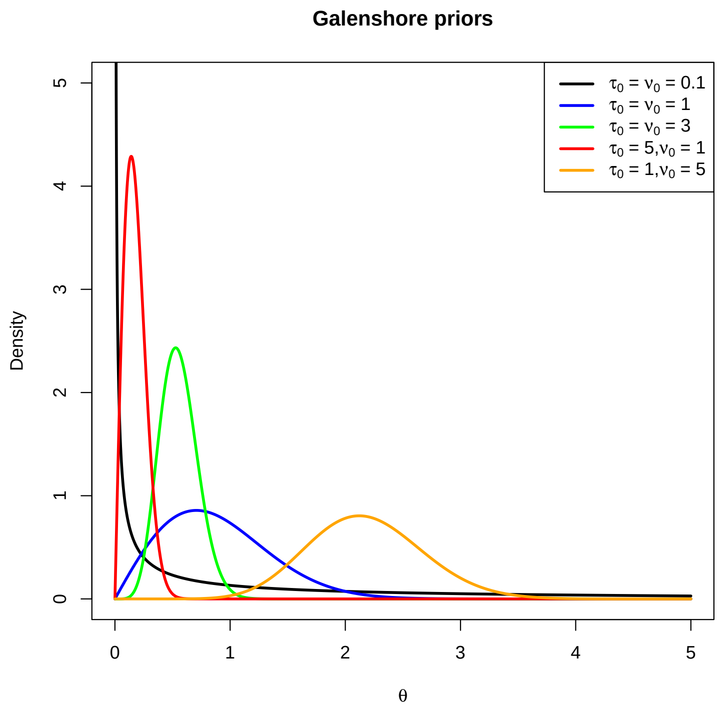
The integrand in the third line is the density of Galenshore$(2a+1/2, \theta)$.

(b) The Galenshore density $p(y|\theta) = \frac{2}{\Gamma(a)} y^{2a-1}\theta^{2a}\exp(-\theta^2 y^2)$ belongs to the exponential family. As a function of θ, $p(y|\theta) \propto \theta^{2a}\exp(-\theta^2 y^2)$. A conjugate prior is given by

$$p(\theta) \propto \theta^{\nu}\exp(-\theta^2\tau).$$

Let $\tau = \tau_0^2$ and $\nu = 2\nu_0 - 1$. Then $p(\theta) \propto \theta^{2\nu_0-1}\exp(-\theta^2\tau_0^2)$, for $\theta > 0$, which is the Galenshore(ν_0, τ_0) distribution. This shows that Galenshore(ν_0, τ_0) is a conjugate prior. A few members from this class of conjugate priors are plotted in the figure below.

```r
theta <- seq(from=0,to=5, by=0.001)
dprior <- function(theta,tau0,nu0){
2*tau0^(2*nu0)/gamma(nu0)*theta^(2*nu0-1)*exp(-tau0^2*theta^2)
}
plot(theta, dprior(theta,0.1,0.1), type="l",lwd=2.5,
xlab=expression(theta), ylab="Density",
main="Galenshore priors",ylim=c(0,5))
points(theta, dprior(theta,1,1), type="l",lwd=2.5,col="blue")
points(theta, dprior(theta,3,3), type="l",lwd=2.5,col="green")
points(theta, dprior(theta,5,1), type="l",lwd=2.5,col="red")
points(theta, dprior(theta,1,5), type="l",lwd=2.5,col="orange")
legend("topright",
legend=c(expression(paste(tau[0]," = ",nu[0]," = 0.1")),
expression(paste(tau[0]," = ",nu[0]," = 1")),
expression(paste(tau[0]," = ",nu[0]," = 3")),
expression(paste(tau[0]," = 5,",nu[0]," = 1")),
expression(paste(tau[0]," = 1,",nu[0]," = 5"))),
col=c("black","blue","green","red","orange"),
lty=1,lwd=2.5)
```



(c) Let $\mathbf{y} = (y_1, \ldots, y_n)$.

$$\begin{aligned} p(\mathbf{y}|\theta) = \prod_{i=1}^{n} p(y_i|\theta) &= \prod_{i=1}^{n} \frac{2}{\Gamma(a)} y_i^{2a-1} \theta^{2a} \exp(-\theta^2 y_i^2) \\ &= \left(\frac{2}{\Gamma(a)}\right)^n \left(\prod_{i=1}^{n} y_i\right)^{2a-1} \theta^{2na} \exp\left(-\theta^2 \sum_{i=1}^{n} y_i^2\right). \end{aligned}$$

From Fisher-Neyman Factorization Theorem, a sufficient statistic for θ is $t = \sum_{i=1}^{n} y_i^2$.

(d)

$$\begin{aligned} p(\theta|\mathbf{y}) \propto p(\theta)p(\mathbf{y}|\theta) &= \theta^{2\nu_0 - 1} \exp(-\tau_0^2 \theta^2) \cdot \theta^{2na} \exp\left(-\theta^2 \sum_{i=1}^{n} y_i^2\right) \\ &= \theta^{2(\nu_0 + na) - 1} \exp\left\{-\left(\tau_0^2 + \sum_{i=1}^{n} y_i^2\right) .\theta^2\right\} \end{aligned}$$

Therefore the posterior distribution is the Galenshore(ν_n, τ_n) distribution where $\nu_n = \nu_0 + na$ and $\tau_n = \sqrt{\tau_0^2 + \sum_{i=1}^n y_i^2}$.

(e) Using the result in (a), $\mathrm{E}(\theta|\mathbf{y}) = \dfrac{\Gamma(\nu_n + 1/2)}{\Gamma(\nu_n)\tau_n}$, where ν_n and τ_n are given in (d).

(f) The posterior predictive distribution is

$$
\begin{aligned}
p(y_{n+1}|\mathbf{y}) &= \int_0^\infty p(y_{n+1}|\theta)p(\theta|\mathbf{y})\,\mathrm{d}\theta \\
&= \frac{4\tau_n^{2\nu_n} y_{n+1}^{2a-1}}{\Gamma(\nu_n)\Gamma(a)} \int_0^\infty \theta^{2(a+\nu_n)-1} \exp(-\theta^2(\tau_n^2 + y_{n+1}^2))\,\mathrm{d}\theta \\
&= \frac{2\tau_n^{2\nu_n} y_{n+1}^{2a-1}}{\Gamma(\nu_n)\Gamma(a)} \int_0^\infty x^{a+\nu_n-1} \exp(-x(\tau_n^2 + y_{n+1}^2))\,\mathrm{d}x \\
&\qquad (x = \theta^2,\ \mathrm{d}\theta/\mathrm{d}x = 1/(2\sqrt{x})) \\
&= \frac{2\tau_n^{2\nu_n} y_{n+1}^{2a-1}}{\Gamma(\nu_n)\Gamma(a)} \frac{\Gamma(a+\nu_n)}{(\tau_n^2 + y_{n+1}^2)^{a+\nu_n}} \\
&= \frac{2\tau_n^{2\nu_n} y_{n+1}^{2a-1}}{\mathrm{B}(a,\nu_n)(\tau_n^2 + y_{n+1}^2)^{a+\nu_n}}, \quad \text{for } y_{n+1} > 0.
\end{aligned}
$$

(This distribution does not belong to the familiar distribution families. But you can show that y_{n+1}^2 is a rescaled random variable from a F distribution.)

3. Let $\mathbf{y} = (y_1, \ldots, y_n)$. Then

$$
\begin{aligned}
p(\theta|\mathbf{y}) &\propto p(\mathbf{y}|\theta)\tilde{p}(\theta) \\
&\propto \prod_{i=1}^n [h(y_i)g(\theta)\exp\{\eta(\theta)t(y_i)\}] \sum_{k=1}^K w_k \kappa(\nu_k, \tau_k) g(\theta)^{\nu_k} \exp\{\eta(\theta)\tau_k\} \\
&\propto g(\theta)^n \exp\left\{\eta(\theta)\sum_{i=1}^n t(y_i)\right\} \sum_{k=1}^K w_k \kappa(\nu_k, \tau_k) g(\theta)^{\nu_k} \exp\{\eta(\theta)\tau_k\} \\
&\propto \sum_{k=1}^K w_k \kappa(\nu_k, \tau_k) g(\theta)^{n+\nu_k} \exp\left\{\eta(\theta)\left(\tau_k + \sum_{i=1}^n t(y_i)\right)\right\} \\
&\propto \sum_{k=1}^K \frac{w_k \kappa(\nu_k, \tau_k)}{\kappa(n+\nu_k, \tau_k + \sum_{i=1}^n t(y_i))} p\left(\theta \;\middle|\; \nu_k + n, \tau_k + \sum_{i=1}^n t(y_i)\right).
\end{aligned}
$$

$$\Rightarrow p(\theta|\mathbf{y}) = \frac{1}{C}\sum_{k=1}^{K}\frac{w_k\kappa(\nu_k,\tau_k)}{\kappa(n+\nu_k,\tau_k+\sum_{i=1}^{n}t(y_i))}p\left(\theta \;\middle|\; \nu_k+n,\tau_k+\sum_{i=1}^{n}t(y_i)\right),$$

where C is a normalization constant. Since $\int p(\theta|\mathbf{y}) = 1$,

$$C = \sum_{k=1}^{K}\frac{w_k\kappa(\nu_k,\tau_k)}{\kappa(n+\nu_k,\tau_k+\sum_{i=1}^{n}t(y_i))}.$$