MATH 598: TOPICS IN STATISTICS

Bayesian modelling with the Binomial model

Suppose that a model is to be constructed under an assumption of exchangeability with the following components:

- Finite realization $y_1, \ldots, y_n$ recorded;
- $\mathcal{Y} \equiv \{0, 1\}$;
- $f_Y(y; \theta) \equiv Bernoulli(\theta)$
- $\pi_0(\theta)$ a prior density on $[0, 1]$.

We consider a data-generating scenario where the true value of the parameter is $\theta_0 = 0.2$ and consider a sample of size $n = 20$.

```
set.seed(234)
n<-20
theta0<-0.2
y<-rbinom(n,1,theta0)
s.n<-sum(y)
```

Bayesian inference for θ

The Bayesian calculation in this problem is based on the de Finetti representation

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = \int_0^1 \left\{ \prod_{i=1}^n f_Y(y_i; \theta) \right\} \pi_0(d\theta) \tag{1}$$

where

$$\prod_{i=1}^n f_Y(y_i; \theta) = \prod_{i=1}^n \theta^{y_i}(1 - \theta_i)^{1-y_i} = \theta^{s_n}(1 - \theta)^{n-s_n}$$

where

$$s_n = \sum_{i=1}^n y_i.$$

Prior distributions: The prior distribution $\pi_0(d\theta)$ can be chosen however deemed suitable.

- **Conjugate prior:** A conjugate prior for θ is a prior where the resulting posterior has the same family form as the prior. Here, if we choose $\pi_0(d\theta) \equiv Beta(\alpha, \beta)$, with corresponding density function

  $$\pi_0(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1 - \theta)^{\beta-1} \qquad 0 \leq \theta \leq 1$$

  for parameters $\alpha, \beta > 0$, then it follows that

  $$\pi_n(\theta) \propto \theta^{s_n}(1 - \theta)^{n-s_n}\theta^{\alpha-1}(1 - \theta)^{\beta-1} = \theta^{s_n+\alpha-1}(1 - \theta)^{n-s_n+\beta-1}$$

  so that $\pi_n(\theta) \equiv Beta(\alpha_n, \beta_n)$

  $$\alpha_n = s_n + \alpha \qquad\qquad \beta_n = n - s_n + \beta.$$

  The two Beta prior hyperparameters control the shape of the prior – the prior is reasonably flexible: the expectation of the distribution is

  $$\frac{\alpha}{\alpha + \beta}$$

  - $\alpha, \beta > 1$: prior is unimodal with mode at

    $$\frac{\alpha - 1}{\alpha + \beta - 2}.$$

  - $\alpha < 1$: prior density is unbounded with asymptote at $\theta = 0$;
  - $\beta < 1$: prior density is unbounded with asymptote at $\theta = 1$.

The predictive distribution for $Y_{n+1}, \ldots, Y_{n+m}$ given $Y_1 = y_1, \ldots, Y_n = y_n$ can also be easily computed under the Beta prior: for $(y_{n+1}, \ldots, y_{n+m}) \in \{0, 1\}^m$,

$$
\begin{aligned}
f_{Y_{n+1},\ldots,Y_{n+m}|Y_1,\ldots,Y_n}(y_{n+1},\ldots,y_{n+m}|y_1,\ldots,y_n) &= \int_0^1 \prod_{i=n+1}^{n+m} f_Y(y_i;\theta)\pi_n(\theta)\,d\theta \\
&= \int_0^1 \theta^{s_{n,m}}(1-\theta)^{m-s_{n,m}} \frac{\Gamma(\alpha_n+\beta_n)}{\Gamma(\alpha_n)\Gamma(\beta_n)}\theta^{\alpha_n-1}(1-\theta)^{\beta_n-1}\,d\theta \\
&= \frac{\Gamma(\alpha_n+\beta_n)}{\Gamma(\alpha_n)\Gamma(\beta_n)} \int_0^1 \theta^{s_{n,m}+\alpha_n-1}(1-\theta)^{m-s_{n,m}+\beta_n}\,d\theta \\
&= \frac{\Gamma(\alpha_n+\beta_n)}{\Gamma(\alpha_n)\Gamma(\beta_n)} \frac{\Gamma(s_{n,m}+\alpha_n)\Gamma(m-s_{n,m}+\beta_n)}{\Gamma(m+\alpha_n+\beta_n)} \\
&= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(s_n+\alpha)\Gamma(n-s_n+\beta)} \frac{\Gamma(s_{n+m}+\alpha)\Gamma(n+m-s_{n+m}+\beta)}{\Gamma(n+m+\alpha+\beta)}
\end{aligned}
$$

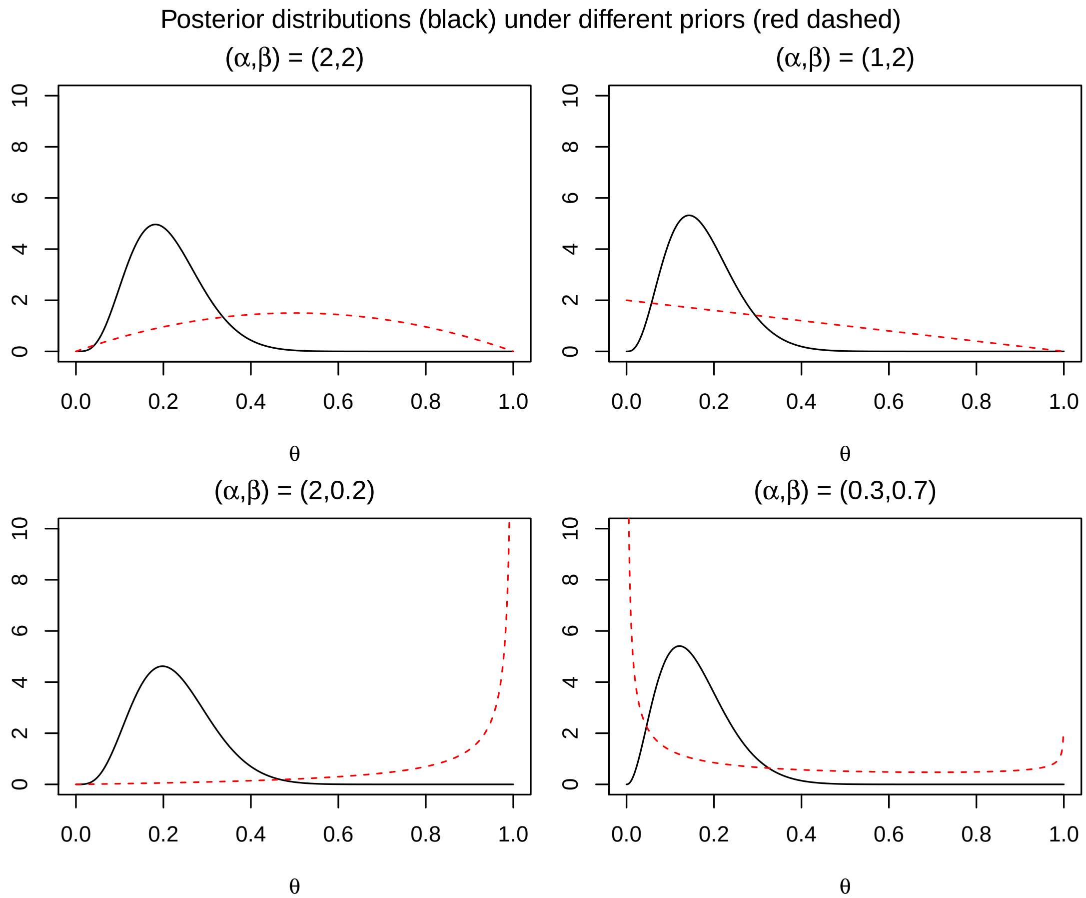
where

$$
s_{n,m} = \sum_{i=n+1}^{n+m} y_i \qquad s_{n+m} = s_n + s_{n,m} = \sum_{i=1}^{n+m} y_i.
$$

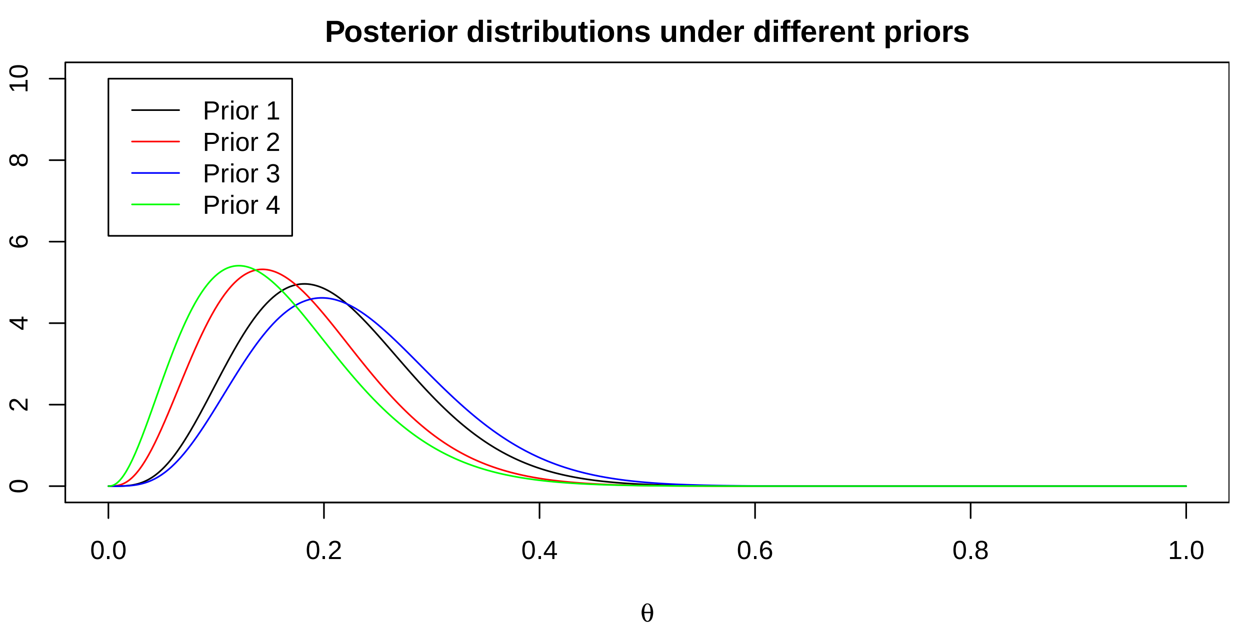
We can investigate the behaviour of the posterior for different settings of the hyperparameters:

- $(\alpha, \beta) = (2, 2)$;
- $(\alpha, \beta) = (1, 2)$;
- $(\alpha, \beta) = (3, 0.2)$;
- $(\alpha, \beta) = (0.3, 0.7)$.

```
par(oma=c(2,2,1,4),mar=c(4,3,2,0),mfrow=c(2,2))
xv<-seq(0,1,by=0.001)

al<-2;be<-2                  ##Prior 1
al.n<-s.n+al;be.n<-n-s.n+be
yv1<-dbeta(xv,al.n,be.n)
mtxt<-substitute(paste('(',alpha,',',beta,')',' = (',a,',',b,')'), list(a=al,b = be))
plot(xv,yv1,type='l',main=mtxt,ylim=range(0,10),xlab=expression(theta))
lines(xv,dbeta(xv,al,be),col='red',lty=2)

al<-1;be<-2                  ##Prior 2
al.n<-s.n+al;be.n<-n-s.n+be
yv2<-dbeta(xv,al.n,be.n)
mtxt<-substitute(paste('(',alpha,',',beta,')',' = (',a,',',b,')'), list(a=al,b = be))
plot(xv,yv2,type='l',main=mtxt,ylim=range(0,10),xlab=expression(theta))
lines(xv,dbeta(xv,al,be),col='red',lty=2)

al<-2;be<-0.2                ##Prior 3
al.n<-s.n+al;be.n<-n-s.n+be
yv3<-dbeta(xv,al.n,be.n)
mtxt<-substitute(paste('(',alpha,',',beta,')',' = (',a,',',b,')'), list(a=al,b = be))
plot(xv,yv3,type='l',main=mtxt,ylim=range(0,10),xlab=expression(theta))
lines(xv,dbeta(xv,al,be),col='red',lty=2)

al<-0.3;be<-0.7                  ##Prior 4
al.n<-s.n+al;be.n<-n-s.n+be
yv4<-dbeta(xv,al.n,be.n)
mtxt<-substitute(paste('(',alpha,',',beta,')',' = (',a,',',b,')'), list(a=al,b = be))
plot(xv,yv4,type='l',main=mtxt,ylim=range(0,10),xlab=expression(theta))
lines(xv,dbeta(xv,al,be),col='red',lty=2)
mtext(text="Posterior distributions (black) under different priors (red dashed)",
side=3,line=0,outer=TRUE)
```

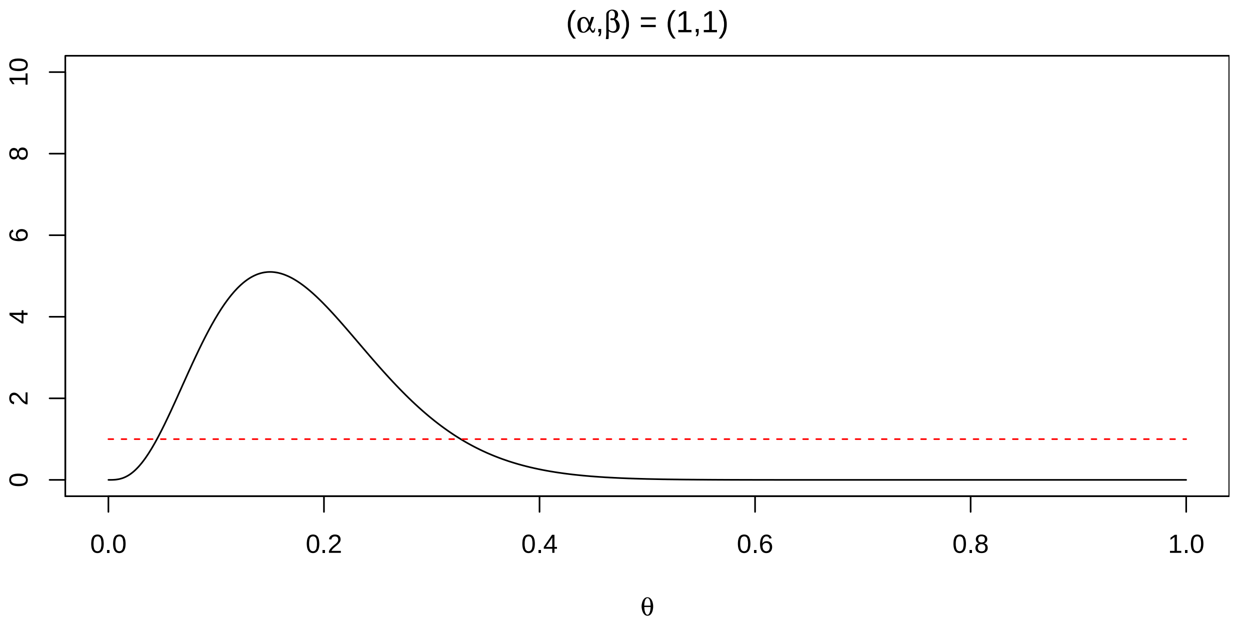



```r
par(mar=c(4,3,2,0),mfrow=c(1,1))
plot(xv,yv1,type='l',main="Posterior distributions under different priors",
ylim=range(0,10),xlab=expression(theta))
lines(xv,yv2,col='red');lines(xv,yv3,col='blue');lines(xv,yv4,col='green')
legend(0,10.0,c('Prior 1','Prior 2','Prior 3','Prior 4'),lty=1,
        col=c('black','red','blue','green'))
```



A special case of the Beta prior is when $\alpha = \beta = 1$; this corresponds to the $Uniform(0,1)$ distribution.

```
par(mar=c(4,3,2,0),mfrow=c(1,1))
xv<-seq(0,1,by=0.001)

al<-1;be<-1                    ##Prior 0
al.n<-s.n+al;be.n<-n-s.n+be
yv0<-dbeta(xv,al.n,be.n)
mtxt<-substitute(paste('(',alpha,',',beta,')',' = (',a,',',b,')'), list(a=al,b = be))
plot(xv,yv0,type='l',main=mtxt,ylim=range(0,10),xlab=expression(theta))
lines(xv,dbeta(xv,al,be),col='red',lty=2)
```



It is sometimes proposed that this prior represents 'prior ignorance' of the true value of θ. However, a Uniform prior on θ is not Uniform on transformations of θ. Consider for example a re-parameterization to $\phi = \theta^2$; this is a monotonic transform on the interval $[0,1]$, and we can compute by standard transformation methods that

$$\pi_0(\phi) = \frac{1}{2\sqrt{\phi}} \qquad 0 \le \phi \le 1.$$

that is $\pi_0(\phi) \equiv Beta(1/2, 1)$, which yields a different posterior distribution.

- **Prior on log odds parameter:** It is possible to place a prior on any transformed version of θ, so consider

$$\phi = \log\left(\frac{\theta}{1-\theta}\right).$$

This is the log odds parameterization, as $\theta/(1-\theta)$ is the odds parameter. In the new parameterization, ϕ is a parameter taking values on $\mathbb{R}$. Note that the transformation is 1-1, and the inverse transform is defined by

$$\theta = \frac{e^\phi}{1+e^\phi}.$$

which allows us to write the likelihood in terms of ϕ. For ϕ, a $Normal(\eta, \tau^2)$ prior may be considered. In the new parameterization, therefore, the posterior distribution takes the form

$$\pi_n(\phi) = c\frac{e^{\phi s_n}}{(1+e^\phi)^n} \exp\left\{-\frac{1}{2\tau^2}(\phi-\eta)^2\right\}$$

where

$$\frac{1}{c} = \int_{-\infty}^{\infty} \frac{e^{t s_n}}{(1+e^t)^n} \exp\left\{-\frac{1}{2\tau^2}(t-\eta)^2\right\}\, dt$$

as other constants cancel. An equivalent formulation in the original parameterization can be written

$$\pi_n(\theta) = c\theta^{s_n-1}(1-\theta)^{n-s_n-1}\exp\left\{-\frac{1}{2\tau^2}\left(\log\left(\frac{\theta}{1-\theta}\right)-\eta\right)^2\right\}$$

where the constant can be re-expressed

$$\frac{1}{c} = \int_0^1 u^{s_n-1}(1-u)^{n-s_n-1}\exp\left\{-\frac{1}{2\tau^2}\left(\log\left(\frac{u}{1-u}\right)-\eta\right)^2\right\} du$$

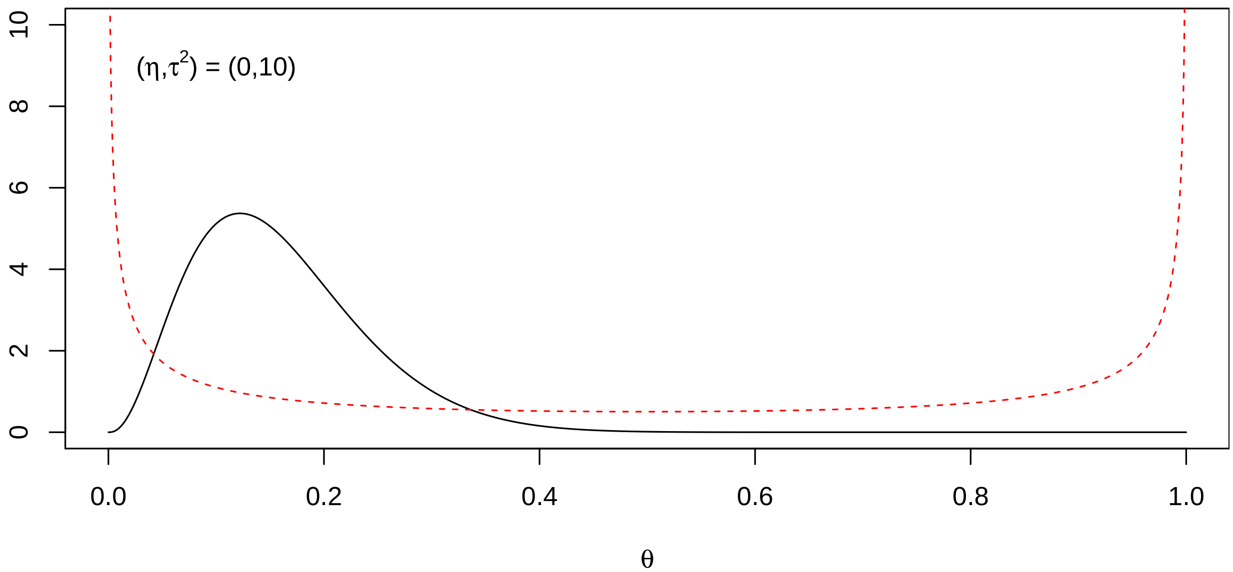
as the Jacobian of the transform is

$$\frac{dt}{du} = \frac{d}{du}\left\{\log\left(\frac{u}{1-u}\right)\right\} = \frac{1}{u(1-u)}$$

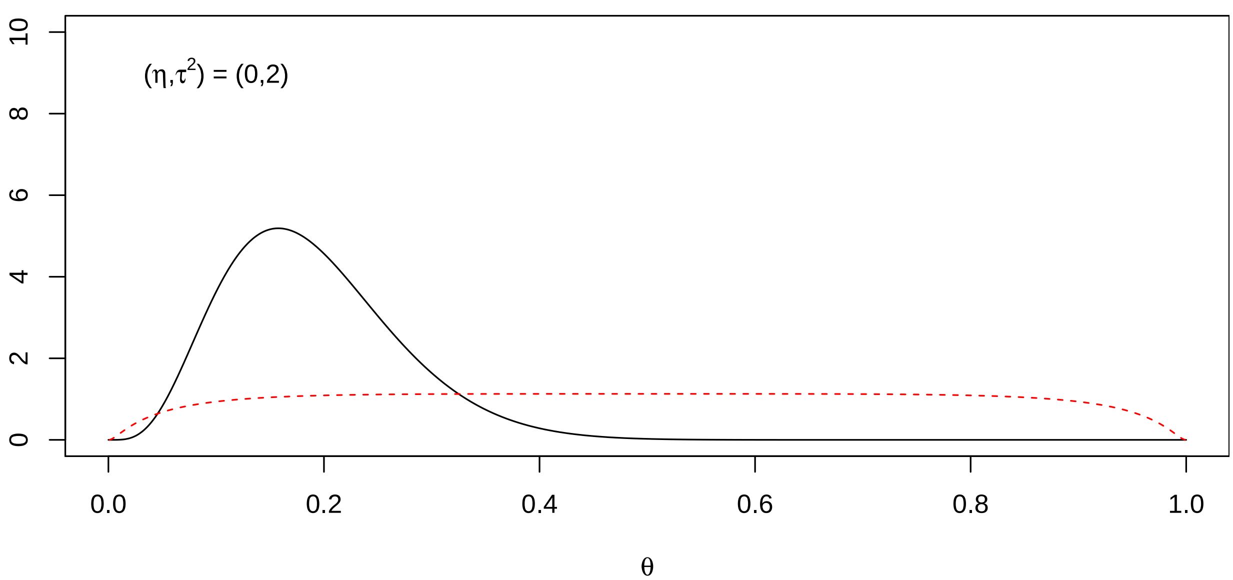
Neither $\pi_n(\theta)$ nor $\pi_n(\phi)$ can be computed analytically, as the integral defining c is not tractable. However, we can compute the integral using numerical methods.

```
#Phi parameterization
fn1<-function(x,nv,sv,ev,tv){

    val<-(exp(x*sv)/(1+exp(x))^nv)*dnorm(x,ev,sqrt(tv))
    return(val)
}
eta<-0
tau<-10
res1<-integrate(fn1,nv=n,sv=s.n,ev=eta,tv=tau,lower=-100,upper=100,subdivisions=1000)
val1<-res1$value

#Phi parameterization
fn2<-function(x,nv,sv,ev,tv){

    th<-log(x/(1-x))
    val<-x^(sv-1)*(1-x)^(nv-sv-1)*dnorm(th,ev,sqrt(tv))
    return(val)
}
res2<-integrate(fn2,nv=n,sv=s.n,ev=eta,tv=tau,lower=0,upper=1,subdivisions=1000)
val2<-res2$value

#Simpsons Rule
composite.simpson <- function(f, a, b, nv) {
  h <- (b - a) / nv
  xj <- seq.int(a, b, length.out = nv + 1)
  xj <- xj[-1]
  xj <- xj[-length(xj)]
  approx <- (h / 3) * (f(a) + 2 * sum(f(xj[seq.int(2, length(xj), 2)])) +
                          4 * sum(f(xj[seq.int(1, length(xj), 2)])) + f(b))
  return(approx)
}

fn3<-function(x){
    nv<-n;sv<-s.n;ev<-eta;tv<-tau
    ph<-log(x/(1-x))
    val<-x^(sv-1)*(1-x)^(nv-sv-1)*dnorm(ph,ev,sqrt(tv))
    return(val)
}
val3<-composite.simpson(fn3,0,1,1000)

c(val1,val2,val3)
+ [1] 0.00003418210 0.00003587907 0.00003587908
(cval<-1/val3)
+ [1] 27871.39
```

```
par(mar=c(4,3,1,0),mfrow=c(1,1))
xv<-seq(0,1,by=0.001)
eta<-0; tau<-10 ##Prior 1
cval1<-1/composite.simpson(fn3,0,1,1000)
phv<-log(xv/(1-xv))
yv0<-xv^(-1)*(1-xv)^(-1)*dnorm(phv,eta,sqrt(tau))
yv1<-cval1*fn2(xv,nv=n,sv=s.n,ev=eta,tv=tau)
mtxt<-substitute(paste('(',eta,',',tau^2,')',' = (',ev,',',tv,')'), list(ev=eta,tv = tau))
plot(xv,yv1,type='l',main='',ylim=range(0,10),xlab=expression(theta))
lines(xv,yv0,col='red',lty=2);text(0.1,9,mtxt)
```




```
eta<-0; tau<-2 ##Prior 2
cval2<-1/composite.simpson(fn3,0,1,1000)
phv<-log(xv/(1-xv))
yv0<-xv^(-1)*(1-xv)^(-1)*dnorm(phv,eta,sqrt(tau))
yv2<-cval2*fn2(xv,nv=n,sv=s.n,ev=eta,tv=tau)
mtxt<-substitute(paste('(',eta,',',tau^2,')',' = (',ev,',',tv,')'), list(ev=eta,tv = tau))
plot(xv,yv2,type='l',main='',ylim=range(0,10),xlab=expression(theta))
lines(xv,yv0,col='red',lty=2);text(0.1,9,mtxt)
```



- **Constrained Beta prior:** Suppose that a Beta prior constrained to some sub-interval (a, b) of $[0, 1]$ is used, that is

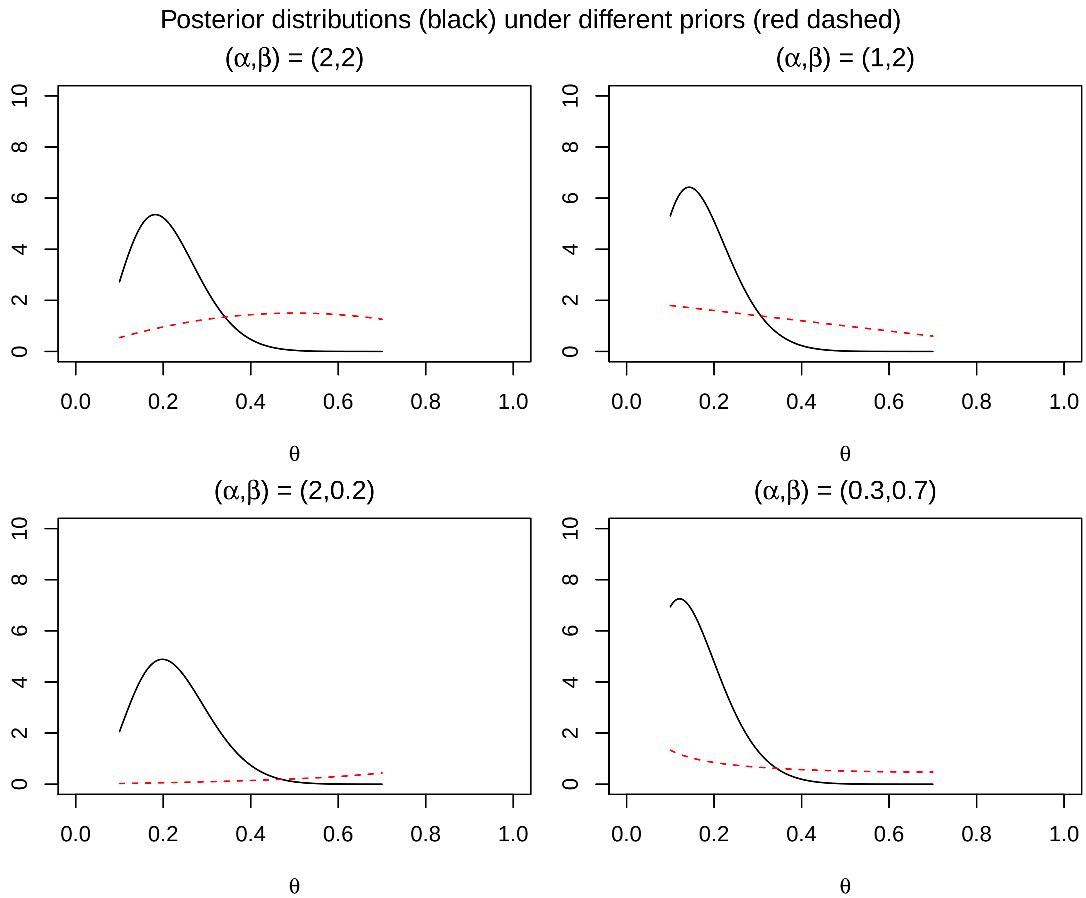
$$\pi_0(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\displaystyle\int_a^b t^{\alpha-1}(1-t)^{\beta-1}\,dt} \qquad a < \theta < b$$

  leading to the posterior distribution

$$\pi_n(\theta) = \frac{\theta^{\alpha_n-1}(1-\theta)^{\beta_n-1}}{\displaystyle\int_a^b t^{\alpha_n-1}(1-t)^{\beta_n-1}\,dt} \qquad a < \theta < b$$

  For example suppose $a = 0.1, b = 0.7$; we can recompute the constrained version of the posterior distributions above.

```
par(oma=c(2,2,1,4),mar=c(4,3,2,0),mfrow=c(2,2))

a<-0.1; b<-0.7
xv<-seq(a,b,by=0.001)

al<-2;be<-2                  ##Prior 1
al.n<-s.n+al;be.n<-n-s.n+be
yv1<-dbeta(xv,al.n,be.n)/(pbeta(b,al.n,be.n)-pbeta(a,al.n,be.n))
mtxt<-substitute(paste('(',alpha,',',beta,')',' = (',a,',',b,')'), list(a=al,b = be))
plot(xv,yv1,type='l',main=mtxt,xlim=range(0,1),ylim=range(0,10),xlab=expression(theta))
lines(xv,dbeta(xv,al,be),col='red',lty=2)

al<-1;be<-2                  ##Prior 2
al.n<-s.n+al;be.n<-n-s.n+be
yv2<-dbeta(xv,al.n,be.n)/(pbeta(b,al.n,be.n)-pbeta(a,al.n,be.n))
mtxt<-substitute(paste('(',alpha,',',beta,')',' = (',a,',',b,')'), list(a=al,b = be))
plot(xv,yv2,type='l',main=mtxt,xlim=range(0,1),ylim=range(0,10),xlab=expression(theta))
lines(xv,dbeta(xv,al,be),col='red',lty=2)

al<-2;be<-0.2                ##Prior 3
al.n<-s.n+al;be.n<-n-s.n+be
yv3<-dbeta(xv,al.n,be.n)/(pbeta(b,al.n,be.n)-pbeta(a,al.n,be.n))
mtxt<-substitute(paste('(',alpha,',',beta,')',' = (',a,',',b,')'), list(a=al,b = be))
plot(xv,yv3,type='l',main=mtxt,xlim=range(0,1),ylim=range(0,10),xlab=expression(theta))
lines(xv,dbeta(xv,al,be),col='red',lty=2)

al<-0.3;be<-0.7                   ##Prior 4
al.n<-s.n+al;be.n<-n-s.n+be
yv4<-dbeta(xv,al.n,be.n)/(pbeta(b,al.n,be.n)-pbeta(a,al.n,be.n))
mtxt<-substitute(paste('(',alpha,',',beta,')',' = (',a,',',b,')'), list(a=al,b = be))
plot(xv,yv4,type='l',main=mtxt,xlim=range(0,1),ylim=range(0,10),xlab=expression(theta))
lines(xv,dbeta(xv,al,be),col='red',lty=2)
mtext(text="Posterior distributions (black) under different priors (red dashed)",
side=3,line=0,outer=TRUE)
```




```r
par(mar=c(4,3,2,0),mfrow=c(1,1))
plot(xv,yv1,type='l',main="Posterior distributions under different priors",
xlim=range(0,1),ylim=range(0,10),xlab=expression(theta))
lines(xv,yv2,col='red');lines(xv,yv3,col='blue');lines(xv,yv4,col='green')
legend(0.8,10.0,c('Prior 1','Prior 2','Prior 3','Prior 4'),lty=1,
       col=c('black','red','blue','green'))
```