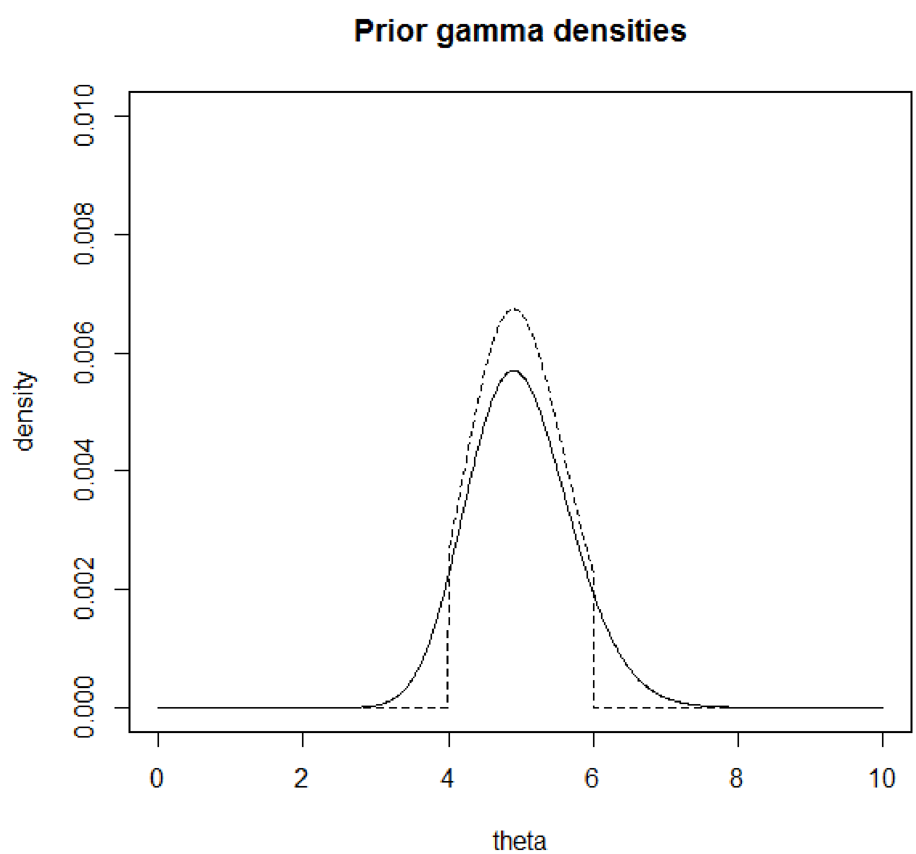
STAT601 Lab 2: Bayesian Computation & Truncated Distributions

Xue Feng

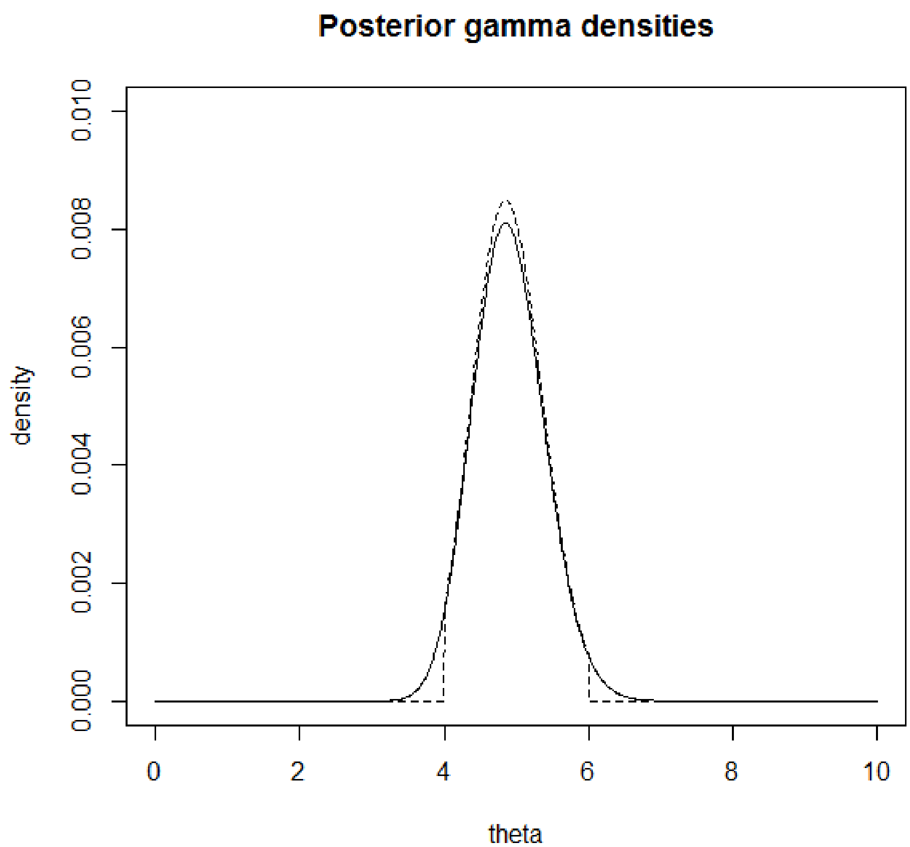
January 26, 2013

We consider 2 prior distributions for θ. Both are plotted below.

1. Gamma(shape=50, scale=0.1)
2. Gamma(shape=50, scale=0.1) truncated to [4,6]



Now we take the likelihood to be $y|\theta \sim \text{Poisson}(\theta)$. We observe independent data ($n = 10$): $y = (2,1,9,4,3,3,7,7,5,7)$. The posterior distributions are plotted below, and the 2.5th and 97.5th percentiles are reported for both distributions.



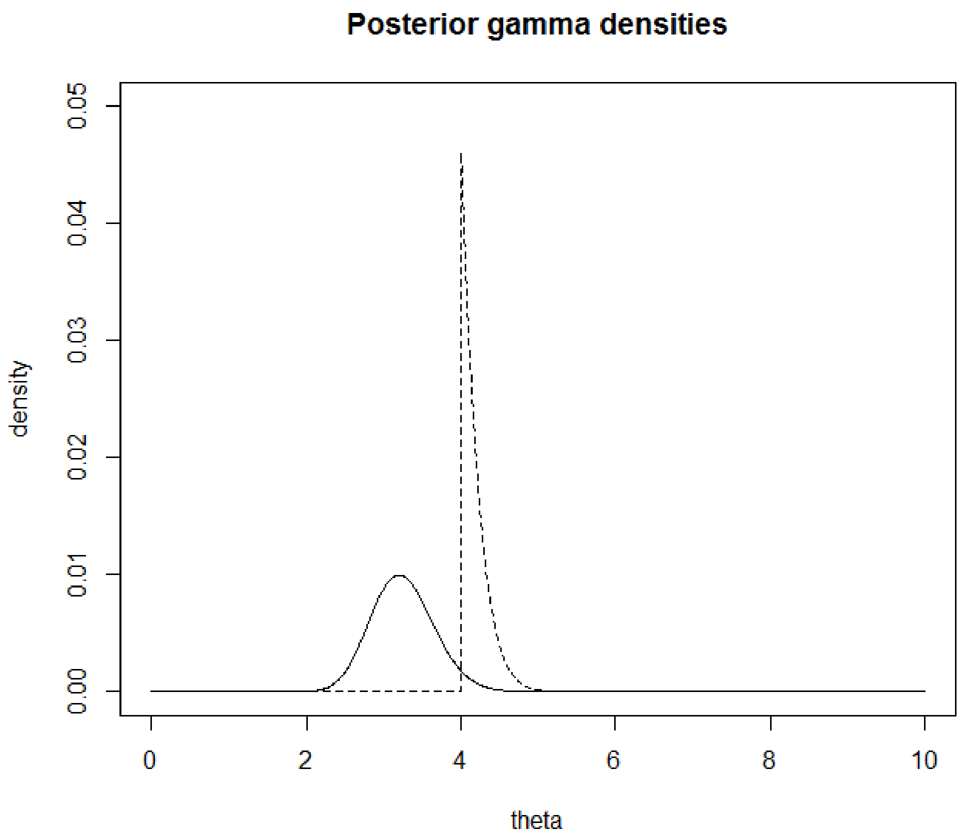
The 95% central credible intervals for the posterior distributions are:

1. Gamma(shape = $\sum y_i + 50$, scale $= \frac{1}{n+\frac{1}{0.1}}$) → **[3.978, 5.917]**
2. Gamma(shape = $\sum y_i + 50$, scale $= \frac{1}{n+\frac{1}{0.1}}$) truncated to [4,6] → **[4.124, 5.791]**

<u>Comments:</u>

The 95% central credible interval for the truncated posterior distribution is slightly narrower compared to the untruncated case because we've limited the range of possible values to [4,6]. Both posterior distributions have higher density within [4,6] compared to the prior.

If we have a different set of data $y = (2,1,0,4,1,1,0,1,1,4)$ then the posterior distributions will look different.



The 95% central credible intervals for the posterior distributions are:

1. Gamma(shape = $\sum y_i + 50$, scale = $\frac{1}{n+\frac{1}{0.1}}$) → **[2.508, 4.086]**
2. Gamma(shape = $\sum y_i + 50$, scale = $\frac{1}{n+\frac{1}{0.1}}$) truncated to [4,6] → **[4.005, 4.647]**

<u>Comments:</u>

In this case, the narrowness of the 95% interval may not be an advantage of the truncated gamma distribution. According to the untruncated posterior, the true value for θ is much more likely to lie outside of the [4,6] interval, toward the lower end. However, since any probability is reduced to zero outside of the [4,6] interval for the truncated posterior, most of the density is concentrated on one side of the interval closest to where the true value may lie. This is not indicative of where the true value may be, but rather where within the truncated interval is likely to be closest to the true value. Thus, a truncated distribution should only be used when we are fairly confident of the range of the true value, such as in the first case.

```
# Plotting 2 priors
# 1. Gamma(shape=50, scale=0.1)
# 2. Gamma(shape=50, scale=0.1) truncated to [4,6]
```

```r
a = 50
b = 0.1
x <- seq(0, 10, length=1000)

# Defining required function
trunc.Gamma = function(x, a, b, lb=4, ub=6){
  y = dgamma(x, shape=a, scale=b) * as.numeric(x>lb & x<ub)/
    (pgamma(ub, shape=a, scale=b)-pgamma(lb, shape=a, scale=b))
  return(y)
}

trunc.QGamma = function(p, a, b, lb=4, ub=6){
  pTrunc = (pgamma(ub, shape=a, scale=b)-pgamma(lb, shape=a, scale=b))*p +
    pgamma(lb, shape=a, scale=b)
  q = qgamma(pTrunc, shape=a, scale=b)
  return(q)
}

trunc.Gamma.Post = function(x, a, b, aPost, bPost, lb=4, ub=6){
  y = dgamma(x, shape=aPost, scale=bPost) * as.numeric(x>lb & x<ub)/
    (pgamma(ub, shape=a, scale=b)-pgamma(lb, shape=a, scale=b))
  return(y)
}

# Plotting the priors
pd1 <- dgamma(x, shape=a, scale=b); pd1 <- pd1/sum(pd1)
pd2 <- trunc.Gamma(x, a, b); pd2 <- pd2/sum(pd2)

plot.new()
plot(x, pd1, type='l', xlab='theta', ylab='density', ylim=c(0,0.01))
lines(x, pd2, lty=2)
title('Prior gamma densities')

# Plotting the posteriors
n=10
#y <- c(2,1,9,4,3,3,7,7,5,7)
y <- c(2,1,0,4,1,1,0,1,1,4)

# posterior for the gamma distribution
aPost = sum(y)+a
bPost = 1/(n+1/b) # need to convert the rate to scale
pd1Post <- dgamma(x, shape=aPost, scale=bPost)
pd1Post <- pd1Post/sum(pd1Post)

# posterior for the truncated gamma distribution
pd2Post <- trunc.Gamma.Post(x,a,b,aPost, bPost)
pd2Post <- pd2Post/sum(pd2Post)

plot.new()
plot(x, pd1Post, type='l', xlab='theta', ylab='density', ylim=c(0,0.05))
lines(x, pd2Post, lty=2)
title('Posterior gamma densities')

# finding the 95% percentile intervals
trunc.QGamma(c(0.025, 0.975), aPost, bPost) # for the truncated posterior
qgamma(c(0.025, 0.975), shape=aPost, scale=bPost) # for the regular posterior
```