AdvStDaAn, Worksheet, Week 12

Michael Lappert

06 Juni, 2022

Contents

Task 1 . . . 1
Task 2 . . . 2
Task 3 . . . 5
Question Task . . . 6

Task 1

You are working in a call center. In the past there were about 30 incomming calls per hour that have to be forwarded. The times between calls can be assumed to be exponentially distributed with rate λ.

Based on your knowledge you model a prior on λ that is Gamma distributed with mean 30 and standard deviation 10.

You recorded 120 calls in 3 hours.

Calculate a 90% credible interval for the parameter λ based on observed data. Calculate the MAP estimate for the parameter λ.

Assume you would have recorded 1200 calls in 30 hours, calculate the 90% credible interval for the parameter λ and the MAP based on these observations.

In the past: 30 calls/hour
Recorded: 120 calls in 3 hours (40 calls/hour)
prior: gamma distribution with $\mu = 30$ and $\sigma = 10$

For the gamma distribution the following equations can be used to calculate a and b:
$\mu = \frac{a}{b}$ and
$\sigma = \frac{a}{b^2}$

solve for one and insert in second equation resulting in:

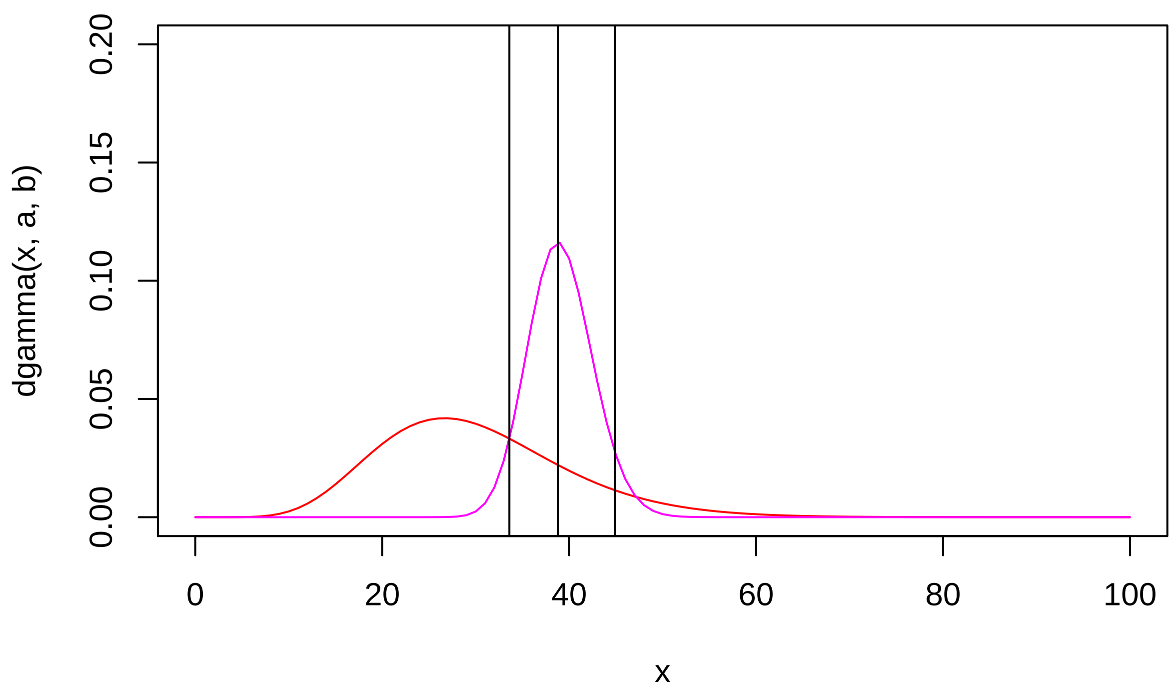
```r
a <- 9
b <- 0.3

# plot prior density
curve(dgamma(x,a,b), 0, 100, ylim = c(0, 0.2), col = 'red')
```

```r
# plot posterior density
curve(dgamma(x,a+120,b+3), 0, 100, col="magenta", add=T)
# equal tailed 90% credible interval
abline(v = qgamma(c(0.05,0.95), a+120, b+3))
qgamma(c(0.05,0.95), a+120, b+3)
```

```
## [1] 33.60715 44.91908
```

```r
# mode
mode = (a+120-1)/(b+3)
abline(v=mode)
```




```r
# more observations
# equal tailed 90% credible interval
qgamma(c(0.05,0.95),a+1200,b+30)
```

```
## [1] 38.03239 41.80711
```

```r
# mode
mode = (a+1200-1)/(b+30)
```

Task 2

The concentration of Carbon dioxide CO2 in the air is reported to be normally distributed with mean 325 ppm and standard deviation 2 (under certain conditions). You can take this as prior information.

You conducted an experiment and measured the concentration of Carbon dioxide CO2 at different times in the air: 310 ppm, 320 ppm, 324 ppm, 307 ppm and 329 ppm. The data are normally distributed with a known fixed standard deviaton 9.

What's an equally tailled 95% credible interval for the concentration of Carbon dioxide CO2 after seeing the data? Perform an exact Bayesian analysis and use an ABC approach to compare the results.

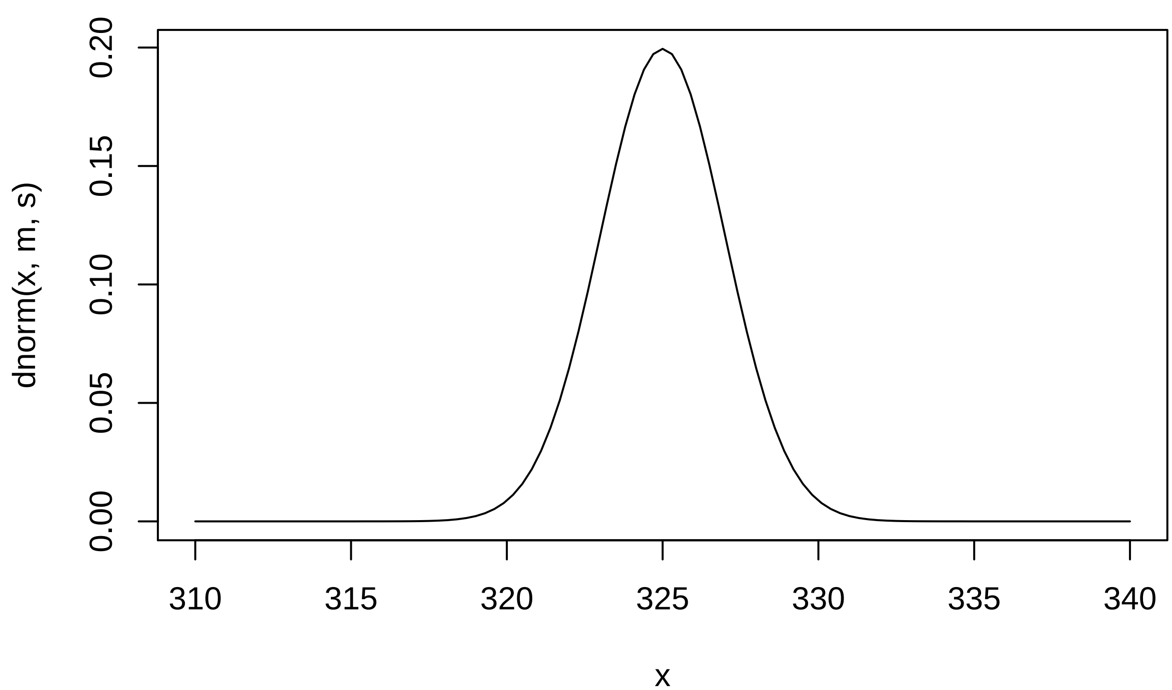
Exact Bayesian analysis:

Mean of observed carbon dioxide concentrations: mean(310, 320, 324, 307, 329) = 318 ppm

The posterior standard deviation of the expected concentration μ is $\left(\frac{1}{2^2} + \frac{5}{9^2}\right)^{-1} = 3.208.$

The posterior mean of the expected concentration μ is $\left(\frac{325}{2^2} + \frac{1590}{9^2}\right)\left(\frac{1}{2^2} + \frac{5}{9^2}\right)^{-1} = 323.6.$

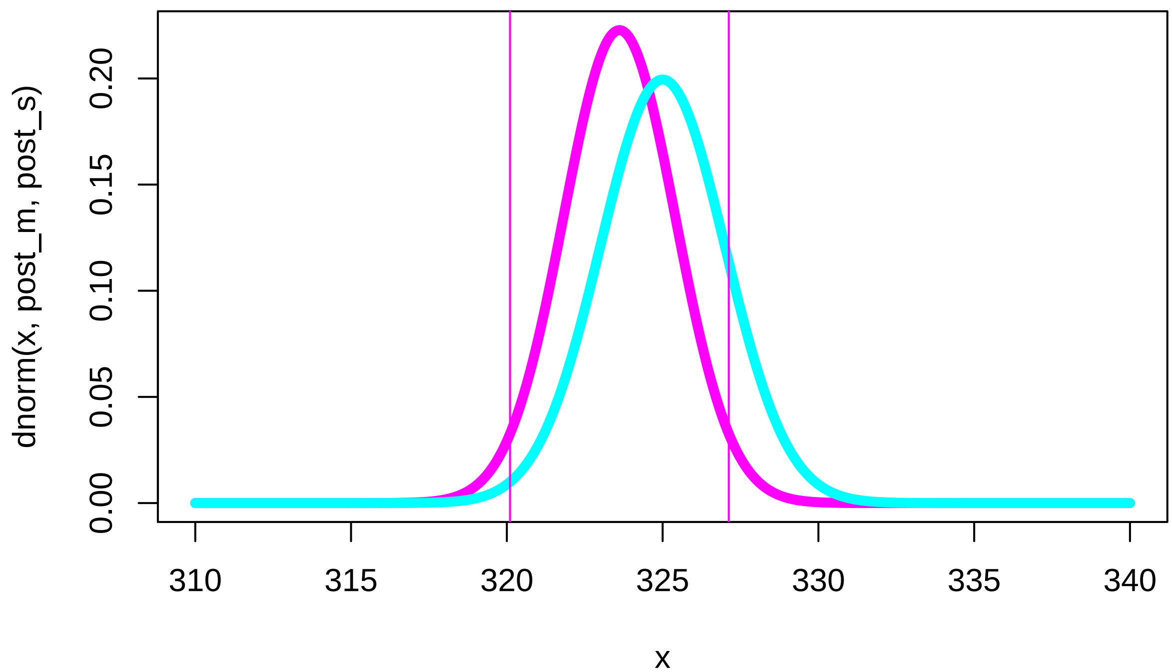
```
# given information
obs  = sum(310, 320, 324, 307, 329)
nObs = 5
sigma = 9
m = 325
s = 2
# plot prior of expected concentration of carbon dioxide
curve(dnorm(x,m,s), 310, 340)
```



```r
# calculate posterior distribution of expected concentration of carbon dioxide
post_m = (m/s^2+obs/sigma^2)/(1/s^2+nObs/sigma^2) # different written than in slides: / and no ^-1
post_m = (m/s^2+obs/sigma^2)*(1/s^2+nObs/sigma^2)^-1 # -> same as above
post_s = (1/s^2+nObs/sigma^2)^-.5
# plot posterior
curve(dnorm(x,post_m,post_s), 310, 340, col="magenta", lwd=5, n=1001)
curve(dnorm(x,m,s), 310, 340, col="cyan", lwd=5, n=1001, add=T)
qnorm(c(0.025, 0.975), post_m, post_s)
```

```
## [1] 320.1034 327.1243
```

```r
abline(v=qnorm(c(0.025,0.975), post_m, post_s), col="magenta")
```



ABC approach:

```r
# number of simulated values
nSamples = 10000
# simulate from the prior
prior = rnorm(nSamples, mean=325, sd=2)
# observed value
observed = sum(310, 320, 324, 307, 329)
observed
```

```
## [1] 1590
```

```r
# take simulated parameters from prior in data generating process and keep values
# if simulated data equals approximately observed data
keep = rep(FALSE, nSamples)
for(i in 1:nSamples)
{
  simulatedValues = rnorm(5, mean=prior[i], sd=9)
  if(abs(sum(simulatedValues)-observed)<10)
  {
      keep[i] = TRUE
  }
}
# condition on observed data
posterior = prior[keep]
# how many data are left after conditioning?
length(posterior)
```

```
## [1] 1063
```

```r
# calculate the credible interval from the sample
quantile(posterior,c(0.025,0.5,0.975))
```

```
##     2.5%      50%    97.5%
## 320.2151 323.6182 327.0033
```

Task 3

You know, that your measurements are normal distributed and centered around 1. The standard deviation is usually 0.5 ± 0.1.

Calculate the precision and propose a suitable Gamma-Prior on the precision, that approximately reflects your prior knowledge.

In an experiment you measured 2.3, 0.1, 1.2, 0.3, 1.3 and 1.6. What's the posterior distribution of the precision?

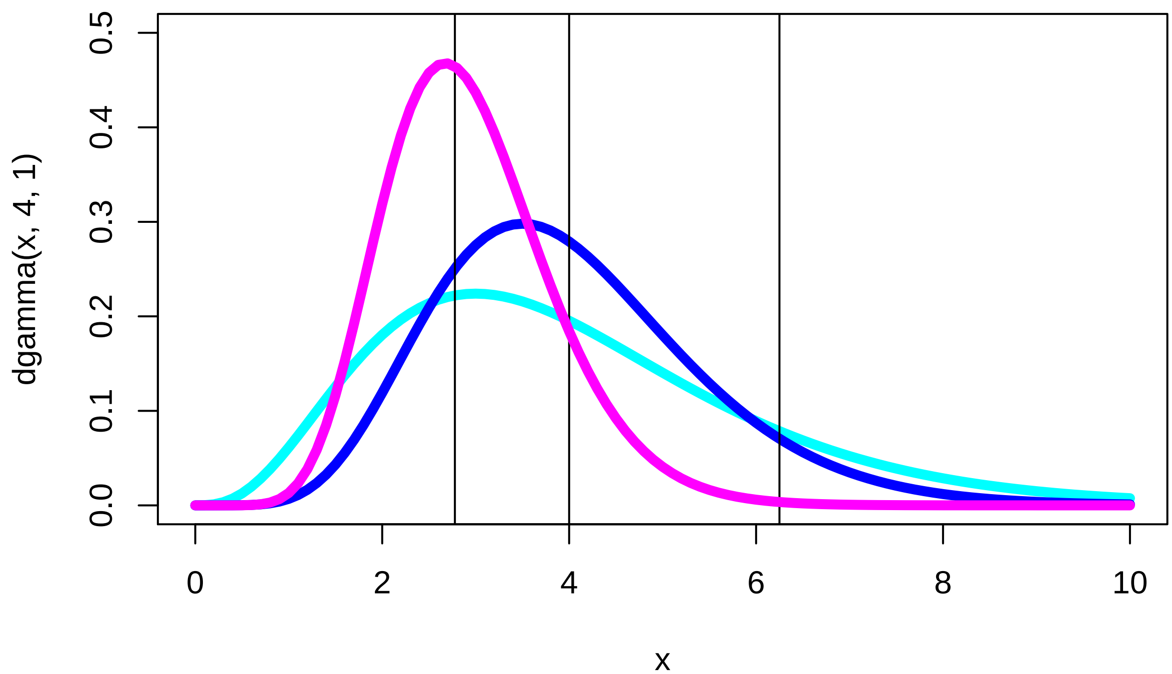
What's the posterior mean; and calculate a 90% credible interval for the precision. In what range is the standard deviation, i.e. give a 90% credible interval.

```r
# calculate a suitable Gamma prior:
# the precision (inverse of variance) lies between
1/c(0.4,0.5,0.6)^2 #mean=4 and sd is approx. 2
```

```
## [1] 6.250000 4.000000 2.777778
```

```r
# by visual inspection: Gamma(4,1) has too low variance, so try Gamma(4*2,1*2)
curve(dgamma(x,4,1), 0, 10, col="cyan", lwd=5, ylim = c(0, 0.5))
curve(dgamma(x,8,2), 0, 10, col="blue", lwd=5, add = T) #is a suitable prior with mean=4
abline(v=1/c(0.4,0.5,0.6)^2)

# posterior distribution of the precision:
obs = c(2.3, 0.1, 1.2, 0.3, 1.3, 1.6)
curve(dgamma(x, 8+6/2, 2+sum((obs-1)^2)/2), 0, 10, add=T, col="magenta", lwd=5)
```




```
# posterior mean of the precision:
(8+6/2)/(2+sum((obs-1)^2)/2)
```

```
## [1] 2.941176
```

```
# 90% credible interval for posterior precision:
posterior_precision = qgamma(c(0.05,0.95), 8+6/2, 2+sum((obs-1)^2)/2)
# 90% credible interval for posterior standard deviation:
posterior_sd = (sqrt(posterior_precision))^-1
posterior_sd
```

```
## [1] 0.7786247 0.4695636
```

```
# compare with standard deviation of observed data:
sd(c(2.3, 0.1, 1.2, 0.3, 1.3, 1.6))
```

```
## [1] 0.821381
```

Remark: The observed standard deviation on measurements is much higher than the prior on the standard deviation. This causes a shift of the prior to higher values of the posterior distribution.

Question Task

|>Q1: |»A1: