The following questions will require you to make use of several functions within the `nclbayes` package. As discussed in class, there are a few options for accessing these functions.

Option 1: Installing the nclbayes package

Do not copy-and-paste this code into R – you will have to type it in by hand very carefully!

```
install.packages("nclbayes", repos="http://R-Forge.R-project.org")
library(nclbayes)
```

Option 2: Accessing the functions from a webfile

Do not copy-and-paste this code into R – you will have to type it in by hand very carefully!

```
source("https://www.mas.ncl.ac.uk/~nlf8/teaching/mas3902/nclbayes.txt")
```

Option 3: Accessing the functions from Canvas

- In Canvas, click *Modules* and then *Main course files*
- Scroll down until you see a link for *R code from "nclbayes" package*
- Either download this file, saving it somewhere convenient, and then copy-and-paste all of its content into your RStudio session...
- ...Or just copy-and-paste directly into your RStudio session

Questions to try

The increase in protein synthesis observed in athletes taking part in a specific muscle-building training programme is assumed Normally distributed with mean μ and standard deviation σ grams per day.

After extensive discussions with sports scientists, a Normal-Gamma prior distribution is assigned to $(\mu, \tau)^T$; specifically,

$$\begin{pmatrix} \mu \\ \tau \end{pmatrix} \sim NGa(2, 0.25, 3, 0.2),$$

where $\tau = 1/\sigma^2$.

1. *[No R needed in this question]* Using the material from Chapter 2 of the lecture notes, state fully the marginal prior distributions for μ, τ and σ. For example,

$$\mu \sim t_6(2, 0.267).$$

3.3	2.3	2.2	3.1	2.2	2.2	2.8	3.0	2.1	3.3
2.0	3.7	1.9	2.2	2.4	2.8	1.9	2.2	2.2	2.6

Table 1: Increase in protein synthesis, in grams per day, for 20 athletes.

2. The scientists have collected data from a random sample of 20 participants who underwent this training program – their increase in protein synthesis, in grams per day, is shown in Table 1 above.

   (a) Enter these data into R; for example:

```
protein = c(3.3, 2.3, 2.2, 3.1, ..., 2.6)
```

   (b) Use R to find

$$\bar{x} = \frac{1}{20}\sum_{i=1}^{20} x_i \quad \text{and} \quad s^2 = \frac{1}{20}\sum_{i=1}^{20}(x_i - \bar{x})^2.$$

   [Hint: Recall from lectures that $s^2 = \frac{n-1}{n}s_u^2$, where s_u^2 is the variance obtained using the `var` *command in R, and n is the sample size.]*

3. Find the posterior distribution for $(\mu, \tau)^T$, using results from Chapter 2 of the lecture notes. At this point you might want to use R like a big calculator – for example:

```
b = 2; c = 0.25; g = 3; h = 0.2
n = length(protein)
xbar = mean(protein)
B = (b*c + n*xbar)/(c + n)
C = c + n
G = ...
H = ...
```

4. *[No R needed in this question]* Using the material from Chapter 2 of the lecture notes, state fully the marginal posterior distributions for μ, τ and σ. For example,

$$\mu|\boldsymbol{x} \sim t_{26}(2.514, 0.011).$$

5. Produce plots of the prior and posterior (superimposed on the same graph) for each of μ, τ and σ. For example:

```
#Partition the plotting space for one plot for each parameter
par(mfrow = c(1, 3))
x.mu = seq(0, 4, 0.001)
mu.prior = dgt(x.mu, 2*g, b, h/(g*c))
mu.posterior = dgt(x.mu, 2*G, B, H/(G*C))
plot(x.mu, mu.prior, type="l", xlab="mu", ylab="Density", lty=2, ylim=
    c(0,max(mu.posterior)))
lines(x.mu, mu.posterior, type="l", lwd=2)
```

   [Hint: You may need to experiment a little with the plotting ranges (e.g. the `xlim` *and* `ylim` *arguments), and the sequence of values over which you compute the densities (e.g.* `x.mu` *above), in order to get nice looking plots.]*

6. Complete the table below showing the 95% highest density intervals for each of μ, τ and σ, before and after seeing the data in Table 1. For example, for μ:

```
> (c(qgt(0.025, 2*g, b, h/(g*c)), qgt(0.975, 2*g, b, h/(g*c))))
[1] 0.7364202 3.2635798
> (c(qgt(0.025, 2*G, B, H/(G*C)), qgt(0.975, 2*G, B, H/(G*C))))
[1] 2.299727 2.727433
```

	Prior 95% HDI	Posterior 95% HDI
μ	$\sim t_6(2, 0.267) \longrightarrow (0.736, 3.264)$	$\sim t_{26}(2.514, 0.011) \longrightarrow (2.300, 2.727)$
τ		
σ		

7. Use the `NGacontour` command to produce plots showing the joint prior density for $(\mu, \tau)^T$. For example:

```
par(mfrow=c(1,1))
mu = seq(0.5, 3.5, len=1000); tau = seq(0, 40, len=1000)
NGacontour(mu, tau, b, c, g, h, lty=2)
```

8. On the plot you produced in question 7, superimpose the contour plot for the joint posterior density. Colour these contours red so they're distinguishable from those for the prior.

9. Now use the `NGacontour` command to produce a plot showing the 90%, 95% and 99% confidence regions for the joint prior and posterior distributions. Again, colour the contours red for the posterior distribution.

10. Using your answers to questions 6–9, compare and contrast your marginal and joint prior and posterior densities.

11. Find the prior and posterior probability that the mean increase in protein synthesis is at least 2.5 grams per day. *[Hint: Try using the `pgt` command.]*

12. Find the 95% posterior predictive highest density interval.