STAT4CI3/6CI3 Computational Methods for Inference

Assignment 4 Solutions

R code for this solution in a plain text file is also available separately

---

Q. 1 **a)** The joint density of the data conditional on the rates and the observation times is

$$f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\lambda}, \boldsymbol{t}) = \prod_{i=1}^{n} \mathrm{P}(X_i = x_i \mid \lambda_i, t_i) = \prod_{i=1}^{n} \frac{\mathrm{e}^{-\lambda_i t_i}(\lambda_i t_i)^{x_i}}{x_i!}$$

The joint prior for $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ is

$$\pi_{\boldsymbol{\lambda}}(\boldsymbol{\lambda} \mid \alpha, \beta) = \prod_{i=1}^{n} \pi_{\lambda_i}(\lambda_i \mid \alpha, \beta) = \prod_{i=1}^{n} \left(\frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda_i^{\alpha-1} \, \mathrm{e}^{-\lambda_i/\beta} \right)$$

And we are given that the prior for β is $\pi_\beta(\beta) \propto \beta^{-1}$.
Hence the joint posterior for the parameter vector $\boldsymbol{\theta} = (\lambda_1, \ldots, \lambda_n, \beta)$ is

$$\begin{aligned}
\pi(\boldsymbol{\theta} \mid \boldsymbol{x}, \boldsymbol{t}) &\propto f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\lambda}, \boldsymbol{t}) \pi_{\boldsymbol{\lambda}}(\boldsymbol{\lambda} \mid \alpha, \beta) \pi(\beta) \\
&= \prod_{i=1}^{n} \frac{\mathrm{e}^{-\lambda_i t_i}(\lambda_i t_i)^{x_i}}{x_i!} \prod_{i=1}^{n} \left\{ \frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda_i^{\alpha-1} \, \mathrm{e}^{-\lambda_i/\beta} \right\} \frac{1}{\beta} \\
&\propto \prod_{i=1}^{n} \left\{ \mathrm{e}^{-\lambda_i t_i} \, \lambda_i^{x_i} \right\} \prod_{i=1}^{n} \left\{ \frac{1}{\beta^\alpha} \lambda_i^{\alpha-1} \, \mathrm{e}^{-\lambda_i/\beta} \right\} \frac{1}{\beta} \\
&\propto \prod_{i=1}^{n} \left\{ \lambda_i^{x_i+\alpha-1} \, \mathrm{e}^{-(t_i+1/\beta)\lambda_i} \right\} \frac{1}{\beta^{n\alpha+1}}
\end{aligned}$$

[5/4 marks]

b) The conditional posterior distribution for $\boldsymbol{\lambda}$ given the data and β can be found by just taking any parts of the joint posterior that depend on a λ_i. Hence we get

$$\begin{aligned}
\pi_{\boldsymbol{\lambda}} &\propto \prod_{i=1}^{n} \left\{ \lambda_i^{x_i+\alpha-1} \, \mathrm{e}^{-(t_i+1/\beta)\lambda_i} \right\} \\
&\propto \prod_{i=1}^{n} \left\{ \frac{1}{\Gamma(x_i+\alpha})\left(\frac{\beta}{\beta t_i + 1} \right)^{x_i+\alpha} \lambda_i^{x_i+\alpha-1} \, \mathrm{e}^{-\lambda_i/\left(\frac{\beta}{\beta t_i+1}\right)} \right\} \\
&= \prod_{i=1}^{n} \pi_{\lambda_i}(\lambda_i \mid \beta, \boldsymbol{x}, \boldsymbol{t})
\end{aligned}$$

Since the joint conditional posterior can be written as a product of univariate conditional posterior distributions we have that $\lambda_1, \ldots, \lambda_n$ are independent and that

$$\pi_{\lambda_i}(\lambda_i \mid \beta, \boldsymbol{x}, \boldsymbol{t}) \;=\; \frac{1}{\Gamma(x_i+\alpha})\left(\frac{\beta}{\beta t_i+1}\right)^{x_i+\alpha} \lambda_i^{x_i+\alpha-1}\, \mathrm{e}^{-\lambda_i/\left(\frac{\beta}{\beta t_i+1}\right)}$$

Hence we have that

$$\lambda_i \mid \beta, \boldsymbol{x}, \boldsymbol{t} \;\sim\; \text{Gamma}\left(x_i+\alpha, \beta/(\beta t_i+1)\right)$$

[5/4 marks]

c) Similarly when looking for the conditional posterior of β given the values of $\lambda_1, \ldots, \lambda_n$ and the data we can just take the parts of the joint posterior density which depend on β so we get

$$\pi_\beta(\beta \mid \boldsymbol{\lambda}, \boldsymbol{x}, \boldsymbol{t}) \;\propto\; \prod_{i=1}^{n}\left\{\mathrm{e}^{-\lambda_i/\beta}\right\}\frac{1}{\beta^{n\alpha+1}} \;=\; \frac{1}{\beta^{n\alpha+1}}\,\mathrm{e}^{-\sum\lambda_i/\beta}$$

Now suppose that $\tau = 1/\beta$ then the conditional posterior of τ given the rates and the data is

$$\begin{aligned}\pi_\tau(\tau \mid \boldsymbol{\lambda}, \boldsymbol{x}, \boldsymbol{t}) &= \pi_\beta \tau^{-1} \mid \boldsymbol{\lambda}, \boldsymbol{x}, \boldsymbol{t})\left|-\tau^{-2}\right| \\ &\propto \tau^{n\alpha+1}\,\mathrm{e}^{-\tau\sum\lambda_i}\,\tau^{-2} \\ &\propto \frac{(\sum\lambda_i)^{n\alpha}}{\Gamma(n\alpha)}\tau^{n\alpha-1}\,\mathrm{e}^{-\tau\sum\lambda_i}\end{aligned}$$

Hence we see that

$$\tau \mid \boldsymbol{\lambda}, \boldsymbol{x}, \boldsymbol{t} \;\sim\; \text{gamma}\left(n\alpha, \left(\sum\lambda_i\right)^{-1}\right)$$

and so we have

$$\beta \mid \boldsymbol{\lambda}, \boldsymbol{x}, \boldsymbol{t} \;\sim\; \text{inverse-gamma}\left(n\alpha, \left(\sum\lambda_i\right)^{-1}\right)$$

[5/4 marks]

d) To run a Gibbs sampler we need to get an initial value for the chain. It is simplest to initialize beta but we cannot generate from the prior distribution of beta. There are a few reasonable approaches such as

1. Generate $\log\beta$ from a uniform on some finite range and then take exponents to get a β.

2. Generate β from an invers-gamma with parameters $n\alpha$ and $\left(\sum\hat{\lambda}_i\right)^{-1}$ where $\hat{\lambda}_i = x_i/t_i$ are the mle's of the the λ_i from the Poisson model. This is equivalent to initializing the λ_i to the fixed value $\hat{\lambda}_i$.

3. Since $\mathrm{E}(\lambda_i) = \alpha\beta$ we can use a similar strategy to the above to initialize β to $\beta_0 = \alpha/\overline{\lambda}$ which is the mean of the distribution generated from in option 2.

Provided that the chain mixes well and we use some reasonable burn-in period, it should not matter what value we use to initialize the chain.
Here is my code to run this chain.

```
gibbs.pumps <- function(N, x, t, alpha, beta) {
  # A function to run a Gibbs Sampler for the
  # Pumps problem.
  #
  # The arguments are the total chain length,
  # the data x and t and the hyperparameter alpha
  # I will also allow inputting an initial value
  # for beta if desired.

  # Check to ensure valid data x and t
  if (any((x<0)|(x%%1!=0)))
    stop("Invalid data in x")
  if (any (t<0)) stop("Invalid Observation Times")
  n <- length(x)
  if (length(t)!=n)
    stop("x and t must be the same length")

  # Initial value for beta using option 3 in
  # the solution if no value is given
  if (missing(beta))
    beta <- alpha/mean(x/t)
  chain <- matrix(NA, ncol=n+1, nrow=N)
  for (i in 1:N){
    lambda <- rgamma(n, x+alpha, scale=1/(t+1/beta))
    beta <- 1/rgamma(1, n*alpha, scale=1/sum(lambda))
    chain[i,] <- c(lambda, beta)
  }
  chain
}

pumps <- data.frame(Fails=c(5, 1, 5, 14, 3, 19, 1, 1, 4, 22),
                    Times=c(94.32, 15.72, 62.88, 125.76, 5.24, 31.44,
                            1.05, 1.05, 2.1, 10.48))
set.seed(20190408)
pumps.out <- gibbs.pumps(15000, pumps$Fails, pumps$Times, 1.8)
```

[7/6 marks]

e) Here is my code to do this and the results

```
mcmc.interval <- function(x, prob=0.95)
  quantile(x, c((1-prob)/2, 1-(1-prob)/2))
t(round(apply(pumps.out[-(1:5000),], 2, mcmc.interval),4))
         2.5%  97.5%
 [1,]  0.0278 0.1319
 [2,]  0.0278 0.3764
 [3,]  0.0400 0.1925
```

```
 [4,]  0.0700 0.1902
 [5,]  0.1807 1.2319
 [6,]  0.3723 0.8905
 [7,]  0.1214 1.8526
 [8,]  0.1235 1.9248
 [9,]  0.3851 2.3991
[10,]  1.1006 2.5601
[11,]  0.1856 0.6170
```

The first 10 of these intervals are for the individual failure rates. Looking at these we can see that the final few pumps are less reliable (higher failure rates). Three of these were only observed for a short time resulting in wide intervals but the final pump certainly seems problematic. **[3/2 marks]**

Q. 2 **a)** **(i)** For the parametric we first estimate θ by $\hat{\theta}$ and then generate samples from the Beta$(1, \hat{\theta})$ distribution.

```
q2.x <- c(0.580, 0.730, 0.986, 0.691, 0.532, 0.819, 0.382, 0.706,
          0.310, 0.791, 0.462, 0.653, 0.589, 0.664, 0.750, 0.786,
          0.590, 0.892, 0.712, 0.713)
q2.n <- length(q2.x)
q2.xbar <- mean(q2.x)
q2.theta.hat <- q2.xbar/(1-q2.xbar)

R <- 100000
set.seed(20190408)
q2.xstar.1 <- matrix(rbeta(R*q2.n, q2.theta.hat, 1), ncol=q2.n)
q2.xbar.star.1 <- rowMeans(q2.xstar.1)
q2.theta.star.1 <- q2.xbar.star.1/(1-q2.xbar.star.1)
q2.bias.1 <- mean(q2.theta.star.1)-q2.theta.hat
q2.se.1 <- sd(q2.theta.star.1)
```

Hence we get $\hat{b}_{\text{boot}}(\hat{\theta}) = 0.0831$ and $\text{se}_{\text{boot}}(\hat{\theta}) = 0.5189$. **[7/5 marks]**

(ii) For the nonparametric bootstrap we simply change from generating samples from the fitted beta distribution to sampling with replacement from the observed data.

```
set.seed(20190408)
q2.xstar.2 <- matrix(sample(q2.x, R*q2.n, replace=T), ncol=q2.n)
q2.xbar.star.2 <- rowMeans(q2.xstar.2)
q2.theta.star.2 <- q2.xbar.star.2/(1-q2.xbar.star.2)
q2.bias.2 <- mean(q2.theta.star.2)-q2.theta.hat
q2.se.2 <- sd(q2.theta.star.2)
```

Hence we get $\hat{b}_{\text{boot}}(\hat{\theta}) = 0.0345$ and $\text{se}_{\text{boot}}(\hat{\theta}) = 0.3355$ so there is a smaller estimated bias and smaller standard error if we do not assume the data comes from a beta(θ) distribution. Note that since I actually generated the original data from the beta$(1, \theta)$ distribution, it is likely that the parametric bootstrap is closer to being correct and so it is likely that the nonparametric bootstrap is underestimating both the bias and standard error in this case. **[6/5 marks]**

b) The following code will give the two normal confidence intervals.

```
q2.CI.norm.1 <- q2.theta.hat-q2.bias.1-qnorm(c(0.975,0.025))*q2.se.1
q2.CI.norm.2 <- q2.theta.hat-q2.bias.2-qnorm(c(0.975,0.025))*q2.se.2
```

This results in the intervals

Parametric:	[0.9020 2.9359]
Nonparametric:	[1.3100 2.6253]

[4/4 marks]

c) For the other bootstrap intervals we have the following code

```
q2.ahat.1 <- sort(q2.theta.star.1)[c(0.025*R, 0.975*R)]
q2.CI.basic.1 <- 2*q2.theta.hat-rev(q2.ahat.1)
q2.CI.perc.1 <- q2.ahat.1
```

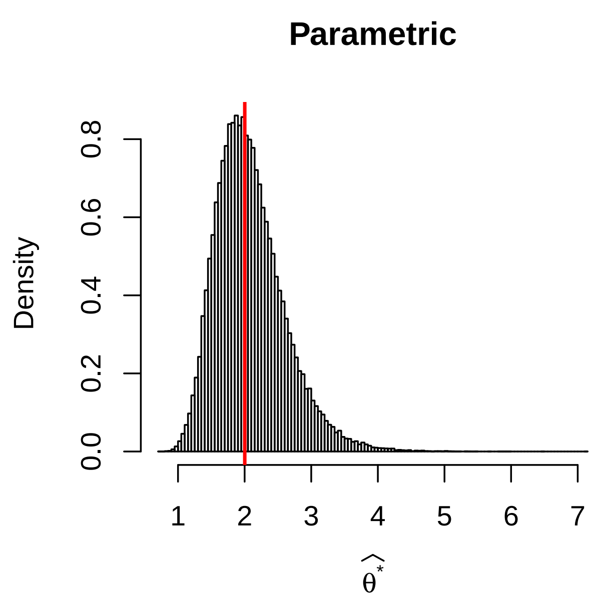
```
q2.ahat.2 <- sort(q2.theta.star.2)[c(0.025*R, 0.975*R)]
q2.CI.basic.2 <- 2*q2.theta.hat-rev(q2.ahat.2)
q2.CI.perc.2 <- q2.ahat.2
```

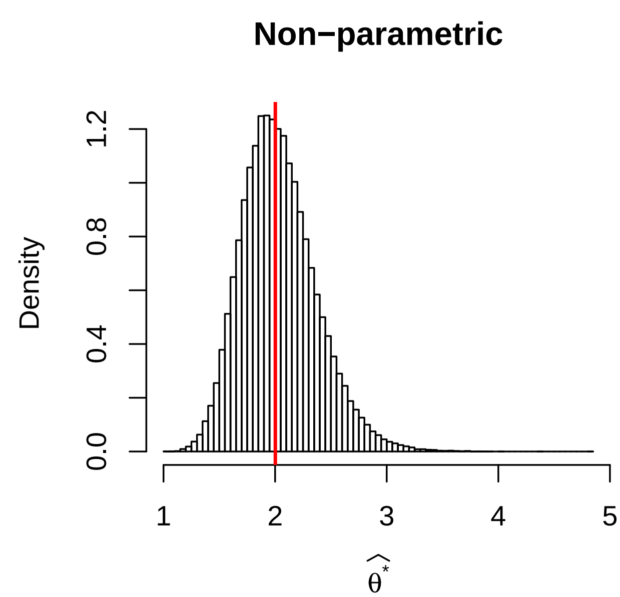
This results in the intervals

Basic	Parametric:	[0.7093 2.7268]
Basic	Nonparametric:	[1.2242 2.5345]
Percentile	Parametric:	[1.2774 3.2949]
Percentile	Nonparametric:	[1.4697 2.7800]

We see that these bootstrap intervals are much wider than the normal bootstrap intervals above and that the parametric intervals are wider than the nonparametric intervals. There is much less of a difference between the basic and percentile intervals for the nonparametric bootstrap than there is for the parametric bootstrap for which there is a very marked difference between the two intervals with the basic being much closer to the normal bootstrap interval. **[8/6 marks]**

Although it was not asked we can examine the distributions of the bootstrap replicates under the two sampling schemes. The following are the two histograms. Clearly we see that the distribution of $\hat{\theta}^*$ is much more variable and skewed when sampling from a fitted beta distribution than when sampling from the empirical distribution.





Q. 3 **a)** A non-parametric bootstrap sample consists of sampling with replacement from the observed n data points. Since ordering of the random sample does not matter, we can think of a bootstrap sample as being a collection of frequencies $(f_1, \ldots, f_n)$ such that $\sum f_i = n$ and each f_i corresponds to the number of times that x_i appears in the sample. Thus the number of bootstrap samples is the same as the number of ways that we can find n non-negative integers which sum up to n.

Now suppose that we have n balls in a line. Consider a fixed stick at the start of the row of balls and another fixed stick at the end of the the row of balls. The number of ways that we can find n non-negative integers which sum to n is then the same as the number of ways that we can partition the n balls into n sets without changing the ordering of the balls. To get n sets we need $n-1$ boundaries. If we define f_i to be the number of balls between stick $i-1$ and stick i (allowing these two sticks to be side-by-side with no balls between them also which corresponds to $f_i = 0$) then it is clear that the number of distinct partitions of the n balls is the number of ways that we can insert the $n-1$ sticks into the sequence of n balls. We can therefore think of having $2n-1$ slots each of which can contain a ball or a stick and we need to select $n-1$ of these to contain the sticks. From elementary combinatorics, we know that the number of distinct ways we can do this is

$$m_n = \binom{2n-1}{n-1}.$$

From the above argument, this must also be the number of ways that we can find n non-negative integers which sum to n and so is the number of possible distinct bootstrap samples for the non-parametric bootstrap with *iid* data. **[5/4 marks]**

b) We know that the number of distinct bootstrap samples is $m_3 = 10$ but it is important to note that they are **not** equally likely since the number of orderings that give rise to each sample may be different. The distinct bootstrap samples and the consequent bootstrap distribution of $\overline{X}^*$ are given in the table on the next page.

In this case all 10 distinct samples result in distinct values of $\overline{X}^*$ but that is not necessarily the case.

f_1	f_2	f_3	$\boldsymbol{X}^*$	number of orderings	$\mathrm{P}(\boldsymbol{X}^*)$	$\overline{x}^*$
3	0	0	$\{2,2,2\}$	1	$\frac{1}{27}$	2
2	1	0	$\{2,2,3\}$	3	$\frac{1}{9}$	$\frac{7}{3}$
2	0	1	$\{2,2,7\}$	3	$\frac{1}{9}$	$\frac{11}{3}$
1	2	0	$\{2,3,3\}$	3	$\frac{1}{9}$	$\frac{8}{3}$
1	1	1	$\{2,3,7\}$	6	$\frac{2}{9}$	4
1	0	2	$\{2,7,7\}$	3	$\frac{1}{9}$	$\frac{16}{3}$
0	3	0	$\{3,3,3\}$	1	$\frac{1}{27}$	3
0	2	1	$\{3,3,7\}$	3	$\frac{1}{9}$	$\frac{13}{3}$
0	1	2	$\{3,7,7\}$	3	$\frac{1}{9}$	$\frac{17}{3}$
0	0	3	$\{7,7,7\}$	1	$\frac{1}{27}$	7

[8/6 marks]

Now from the definition of the bootstrap bias and variance in my notes we have

$$\begin{aligned}
b_{\text{boot}}(\overline{X}) &= \mathrm{E}(\overline{X}^* \mid \hat{F}) - \overline{x} \\
&= \sum_{\overline{x}^*} \overline{x}^* \mathrm{P}(\overline{x}^*) - \overline{x} \\
&= \frac{2}{27} + \frac{7}{27} + \frac{11}{27} + \frac{8}{27} + \frac{8}{9} + \frac{16}{27} + \frac{1}{9} + \frac{13}{27} + \frac{17}{27} + \frac{7}{27} - 4 \\
&= 0
\end{aligned}$$

[3/2 marks]

$$
\begin{aligned}
v_{\text{boot}}(\overline{X}) &= \operatorname{Var}(\overline{X}^* \mid \hat{F}) \\
&= \mathrm{E}(\overline{X}^{*2} \mid \hat{F}) - 16 \\
&= \sum_{\overline{x}^*} \overline{x}^{*2} \mathrm{P}(\overline{x}^*) - \left(\sum_{\overline{x}^*} \overline{x}^* \mathrm{P}(\overline{x}^*) \right)^2 \\
&= \frac{4}{27} + \frac{49}{81} + \frac{121}{81} + \frac{64}{81} + \frac{32}{9} + \frac{256}{81} + \frac{1}{3} + \frac{169}{81} + \frac{289}{81} + \frac{49}{27} - 4^2 \\
&= \frac{1422}{81} - 16 \\
&= \frac{14}{9}
\end{aligned}
$$

[3/2 marks]

c) We know that sampling from the empirical distribution function means that a bootstrap sample $X_1^*, \ldots, X_n^*$ are *iid* with probability mass function

$$
\mathrm{P}\left(X^* = x\right) = \frac{1}{n} \qquad x \in \mathcal{X} = \{x_1, \ldots, x_n\}
$$

Hence we have that

$$
\mathrm{E}\left(X^*\right) = \sum_{x \in \mathcal{X}} x \mathrm{P}\left(X^* = x\right) = \sum_{i=1}^{n} x_i \times \frac{1}{n} = \overline{x}
$$

Similarly we get

$$
\mathrm{E}\left((X^*)^2\right) = \sum_{x \in \mathcal{X}} x^2 \mathrm{P}\left(X^* = x\right) = \sum_{i=1}^{n} x_i^2 \times \frac{1}{n}
$$

and so

$$
\begin{aligned}
\operatorname{Var}\left(X^*\right) = \mathrm{E}\left((X^*)^2\right) - \left(\mathrm{E}\left(X^*\right)\right)^2 &= \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{x}^2 \\
&= \frac{1}{n} \left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2 \right) \\
&= \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2
\end{aligned}
$$

[3/3 marks]

Now the bootstrap replicate of the sample mean is

$$\overline{X}^* = \frac{1}{n}\sum_{i=1}^{n} X_i^*$$

and standard results for sample means tell us that

$$\begin{aligned}
\mathrm{E}\left(\overline{X}^*\right) &= \mathrm{E}\left(X^*\right) &&= \overline{x} \\
\mathrm{Var}\left(\overline{X}^*\right) &= \frac{\mathrm{Var}\left(X^*\right)}{n} &&= \frac{1}{n^2}\sum_{i=1}^{n}(x_i - \overline{x})^2
\end{aligned}$$

Hence from the definition of the bootstrap bias and variance given in my notes we have

$$\begin{aligned}
b_{\text{boot}}(\overline{X}) &= \mathrm{E}\left(\overline{X}^*\right) - \overline{x} &&= 0 \\
v_{\text{boot}}(\overline{X}) &= \mathrm{Var}\left(\overline{X}^*\right) &&= \frac{1}{n^2}\sum_{i=1}^{n}(x_i - \overline{x})^2
\end{aligned}$$

[3/2 marks]

Q. 4 This is a one-sample problem with a sample of n bivariate vectors. We get the observed value from the sample.

```
library(boot)
q3.n <- nrow(cd4)
q3.r <- cor(cd4$baseline, cd4$oneyear)
```

a) For the parametric bootstrap suppose that the data vectors are of the form $\boldsymbol{X}_i = (X_{i,0}, X_{i,1})^t$ where $X_{i,0}$ is the baseline cd4 count (divided by 100) for patient i and $X_{i,1}$ is the cd4 count after a year of medication. We assume that this is an observation from the bivariate normal distribution with parameters

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_0^2 & \sigma_{0,1} \\ \sigma_{0,1} & \sigma_1^2 \end{pmatrix}$$

We need to estimate these parameters. We can do so using maximum likelihood or standard unbiased estimation (they are the same for the means and only differ by a factor of $(n-1)/n$ for the variances). I will do the latter in my solution but accepted either in yours.

```
mu.hat <- colMeans(cd4)
Sigma.hat <- var(cd4) # cov(cd4) gives the same result!
```

This results in

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix} 3.288 \\ 4.093 \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 0.6571 & 0.6800 \\ 0.6800 & 1.3455 \end{pmatrix}$$

[4/3 marks]

We now simulate from the bivariate normal distribution with these parameter values using the `mvrnorm` function written by Brian Ripley (there is also a function `rmvnorm` in the `mvtnorm` package written by Friedrich Leisch and Fabian Scheipl, either is fine for this question) and for each simulated dataset we calculate the correlation coefficient.

```
library(MASS)
set.seed(20190408)
R <- 100000
q4.r.star.norm <- rep(NA, R)
for (j in 1:R) {
   cd4.star <- mvrnorm(q4.n, mu.hat, Sigma.hat)
   q4.r.star.norm[j] <- cor(cd4.star[,1], cd4.star[,2])
}
```

Finally we calculate the bootstrap bias estimate, standard error and hence the bootstrap normal interval for the true correlation coefficient ρ.

```
q4.b.boot.norm <- mean(q4.r.star.norm)-q4.r
q4.se.boot.norm <- sd(q4.r.star.norm)
q4.CI.boot.norm <- q4.r-q4.b.boot.norm-qnorm(c(0.975,0.025))*q4.se.boot.norm
```

This gives us

$$\hat{b}_{\text{boot}}(r) = -0.0092 \qquad \text{se}_{\text{boot}}(r) = 0.1172$$

and the 95% confidence interval $\big(0.5025,\ 0.9621\big)$. **[6/5 marks]**

b) To use the Fisher's transformation we simply need to define the transformation and apply it to all of our bootstrap replicates r^*. We then define the inverse of this transformation and apply it to the endpoints of the interval for ψ.

The inverse of the transformation is

$$\psi = \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right) \quad \Longleftrightarrow \quad \rho = \frac{e^{2\psi}-1}{e^{2\psi}+1}$$ **[3/2 marks]**

```
fisher <- function(r) 0.5*log((1+r)/(1-r))
fisher.inv <- function(p) (exp(2*p)-1)/(exp(2*p)+1)

q4.psi.hat <- fisher(q4.r)
q4.psi.star.norm <- fisher(q4.r.star.norm)
q4.b.boot.psi.norm <- mean(q4.psi.star.norm)-q4.psi.hat
q4.se.boot.psi.norm <- sd(q4.psi.star.norm)
q4.CI.boot.psi.norm <- q4.psi.hat-q4.b.boot.psi.norm-
                          qnorm(c(0.975,0.025))*q4.se.boot.psi.norm
q4.CI.boot.rho.norm <- fisher.inv(q4.CI.boot.psi.norm)
```

From this we get the 95% interval for ρ to be $\big(0.4000,\ 0.8771\big)$ which is wider than the interval calculated on the original scale and is shifted towards 0. **[4/3 marks]**

c) For the non-parametric bootstrap we replace the calls to `mvrnorm` with resampling rows of the cd4 data with replacement. The easiest way to do this is to resample the row numbers from the set of possible row numbers $\{1, \ldots, n\}$.

```
set.seed(20190408)
inds <- matrix(sample(1:q4.n, R*q4.n, replace=TRUE), ncol=q4.n)
q4.r.star.np <- rep(NA, R)
for (j in 1:R) {
   i <- inds[j,]
   q4.r.star.np[j] <- cor(cd4$baseline[i], cd4$oneyear[i])
}
q4.b.boot.np <- mean(q4.r.star.np)-q4.r
q4.se.boot.np <- sd(q4.r.star.np)
q4.CI.boot.np <- q4.r-q4.b.boot.np-qnorm(c(0.975,0.025))*q4.se.boot.np
```

I get $\hat{b}_{\text{boot}}(r) = -0.0061$, $\text{se}_{\text{boot}}(r) = 0.0914$ and 95% confidence interval $\big(0.5501,\ 0.9084\big)$. **[5/4 marks]**

Using the transformed scale we have

```
q4.psi.star.np <- fisher(q4.r.star.np)
q4.b.boot.psi.np <- mean(q4.psi.star.np)-q4.psi.hat
q4.se.boot.psi.np <- sd(q4.psi.star.np)
q4.CI.boot.psi.np <- q4.psi.hat-q4.b.boot.psi.np-
                        qnorm(c(0.975,0.025))*q4.se.boot.psi.np
q4.CI.boot.rho.np <- fisher.inv(q4.CI.boot.psi.np)
```

which gives us the interval for ρ to be $\big(0.4873,\ 0.8542\big)$.

Comparing with the results obtained using the normal model we see that the nonparametric bias and standard error estimate are both smaller (in absolute value) and the intervals are much narrower. **[3/3 marks]**

Q. 5 Required for STATS 6CI3 Students Only

a) First we note that the joint posterior distribution of μ and σ^2 is given by

$$\begin{aligned}
\pi(\mu, \sigma^2 \mid \boldsymbol{x}) &\propto L(\mu, \sigma^2; \boldsymbol{x})\pi(\mu, \sigma^2) \\
&\propto \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\} \exp\left\{-\frac{n(\mu-\overline{x})^2}{2\sigma^2}\right\}
\end{aligned}$$

where $s^2 = (n-1)^{-1}\sum(x_i - \overline{x})^2$ is the sample variance. **[2 marks]**

The conditional posterior distribution of μ for a given σ^2 will be proportional to the joint posterior of μ and σ^2 evaluated at the given value of σ^2. Since the value of σ^2 is known we can omit parts which depend only on σ^2 so we see that

$$\pi(\mu \mid \sigma^2, \boldsymbol{x}) \propto \pi(\mu, \sigma^2 \mid \boldsymbol{x}) \propto \exp\left\{-\frac{n(\mu-\overline{x})^2}{2\sigma^2}\right\}$$

We now recognize this as the kernel of a normal pdf with mean $\overline{x}$ and variance σ^2/n. Thus we have

$$(\mu \mid \sigma^2 = v, \boldsymbol{x}) \;\sim\; \text{Normal}(\overline{x}, v/n)$$

[3 marks]

b) To get the marginal posterior distribution of μ we need to integrate the joint posterior distribution with respect to σ^2. We only know that distribution up to a constant but that is okay.

To complete the integration easily we recall from class that the inverse gamma pdf is given by

$$f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{-(\alpha+1)} \exp\left\{-\frac{1}{\beta y}\right\}$$

and since this integrates to 1 we have that

$$\int_0^\infty y^{-(\alpha+1)} \exp\left\{-\frac{1}{\beta y}\right\} dy = \Gamma(\alpha)\beta^\alpha$$

Thus the marginal posterior distribution for μ is

$$\begin{aligned}
\pi_\mu(\mu \mid \boldsymbol{x}) &\propto \int_0^\infty \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{(n-1)s^2 + n(\mu-\overline{x})^2}{2\sigma^2}\right\} d\sigma^2 \\
&= \Gamma\left(\frac{n}{2}\right)\left(\frac{2}{(n-1)s^2 + n(\mu-\overline{x})^2}\right)^{n/2} \\
&\propto \left(\frac{2}{(n-1)s^2 + n(\mu-\overline{x})^2}\right)^{n/2} \\
&\propto \left((n-1)s^2 + n(\mu-\overline{x})^2\right)^{-n/2} \qquad \mu \in \mathbb{R}
\end{aligned}$$

[3 marks]

Now let us make the transformation from μ to the new random variable

$$T = \frac{\sqrt{n}(\theta - \overline{x})}{s} \;\Rightarrow\; \mu = \frac{sT}{\sqrt{n}} + \overline{x}$$

[*Recall that in Bayesian inference $\overline{x}$ and s^2 are considered to be fixed known quantities, so this is a simple linear transformation.*]
Thus we have

$$\begin{aligned}
\pi_T(t \mid \boldsymbol{x}) &= \pi_\mu\left(\frac{st}{\sqrt{n}} + \overline{x}\right)\frac{s}{\sqrt{n}} \\
&\propto \left((n-1)s^2 + n\left(\frac{st}{\sqrt{n}} + \overline{x} - \overline{x}\right)^2\right)^{-n/2} \\
&= \left((n-1)s^2 + s^2t^2\right)^{-n/2} \\
&= (n-1)^{-n/2}s^{-n}\left(1 + \frac{t^2}{n-1}\right)^{-n/2} \\
&\propto \left(1 + \frac{t^2}{n-1}\right)^{-n/2} \qquad t \in \mathbb{R}
\end{aligned}$$

We now recognize this as being proportional to the density of a Student's t distribution with $n-1$ degrees of freedom. (See, for example Page 625 of *Statistical Inference* by Casella & Berger or Page 211 of *Mathematical Statistics* by Hogg, McKean & Craig for the definition of t_ν density)
Hence we have that

$$T \mid \boldsymbol{x} \;=\; \left.\frac{\sqrt{n}(\mu - \overline{x})}{s}\right|\overline{x}, s \;\sim\; t_{n-1}$$

[3 marks]

c) Similarly, for the marginal posterior distribution for σ^2 we integrate over μ and get

$$\begin{aligned}
\pi_{\sigma^2}(\sigma^2 \mid \boldsymbol{x}) &= \int_{-\infty}^{\infty} \pi(\mu, \sigma^2 \mid \boldsymbol{x})\, d\mu \\
&\propto \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{n(\mu - \overline{x})^2}{2\sigma^2}\right\} d\mu \\
&= \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\} \sqrt{\frac{2\pi\sigma^2}{n}} \\
&\propto \left(\frac{1}{\sigma^2}\right)^{(n+1)/2} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\} \qquad \sigma^2 > 0
\end{aligned}$$

Comparing this to the density of the inverse gamma distribution we see that the marginal posterior distribution of σ^2 is such that

$$\sigma^2 \mid \boldsymbol{x} \;\sim\; \text{Inverse Gamma}\left(\alpha = \frac{n-1}{2}, \beta = \frac{2}{(n-1)s^2}\right)$$

[4 marks]

Now let

$$W = \frac{(n-1)s^2}{\sigma^2} \quad \Rightarrow \quad \sigma^2 = \frac{(n-1)s^2}{W}$$

Hence we have that

$$\begin{aligned}
\pi_W(w \mid \boldsymbol{x}) &= \pi_{\sigma^2}\left(\frac{(n-1)s^2}{w}\right)\left|-\frac{(n-1)s^2}{w^2}\right| \\
&= \frac{1}{\Gamma\left(\frac{n-1}{2}\right)\left(\frac{2}{(n-1)s^2}\right)^{(n-1)/2}}\left(\frac{w}{(n-1)s^2}\right)^{\frac{n-1}{2}+1}\exp\left\{-\frac{w}{2}\right\}\left(\frac{(n-1)s^2}{w^2}\right) \\
&= \frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{(n-1)/2}}w^{\frac{n-1}{2}-1}\exp\left\{-\frac{w}{2}\right\} \qquad w > 0
\end{aligned}$$

We see that this is the pdf of a gamma density with $\alpha = (n-1)/2$ and $\beta = 2$ which is the density of the χ^2_{n-1} distribution and so we have

$$W \mid \boldsymbol{x} = \left.\frac{(n-1)s^2}{\sigma^2}\right| s^2 \sim \chi^2_{n-1}.$$

[2 marks]

d) The results given above are for fixed $\overline{x}$ and s given by the dataset actually seen. They give the posterior distributions of the parameters μ and σ^2 for this single dataset. The parameters μ and σ^2 are considered the random variables in this Bayesian analysis.

In the usual frequentist results μ and σ^2 are considered as fixed but unknown values and the distributions are for the random variables $\overline{X}$ and S^2 derived from repeated samples of size n from the underlying normal(μ, σ^2) distribution. **[3 marks]**