Problem 1:

This problem continues with the HCB pollution data from Assignments 6 and 7.

Surface	Bottom
3.74	5.44
4.61	6.88
4.00	5.37
4.67	5.44
4.87	5.03
5.12	6.48
4.52	3.89
5.29	5.85
5.74	6.85
5.48	7.16

As with Assignments 6 and 7, assume the observations are normally distributed with unknown depth-specific means Θ_s and Θ_b and precisions P_s and P_b. Assume that experts have provided the following prior information based on previous studies.

- The unknown means Θ_s and Θ_b are independent and normally distributed with mean μ and standard deviation τ. The unknown precisions P_s and P_b are independent of Θ_s and Θ_b and have gamma distributions with shape α and scale β.
- Experts specified a 95% prior credible interval of [3, 9] for Θ_s and Θ_b. A good fit to this credible interval is obtained by setting the prior mean to μ =6 and the prior standard deviation to τ=1.5.
- A 95% prior credible interval of [0.75, 2.0] is given for the unknown standard deviations Σ_s and Σ_b. This translates to a credible interval of [0.25, 1.8] for $P_s = \Sigma_s^{-2}$ and $P_b = \Sigma_b^{-2}$. A good fit to this credible interval is obtained by setting the prior shape to α = 4.5 and the prior scale to β = 0.19.

Find the following conditional distributions.

- The conditional distribution for Θ_s given the other parameters and the observations.
- The conditional distribution for Θ_b given the other parameters and the observations.
- The conditional distribution for P_s given the other parameters and the observations.
- The conditional distribution for P_b given the other parameters and the observations.

Solution:

We do not have a conjugate prior for this problem, so we cannot calculate the posterior distribution in closed form. But this is a semi-conjugate prior, so we can calculate the conditional distributions we need for Gibbs sampling.

Surface:

Conditional on P_s , the posterior distribution for Θ_s is independent of Θ_b and P_b, and can be found by normal-normal conjugate updating:

- The observations are normal with mean Θ_s and standard deviation $\Sigma_s = 1/\sqrt{P_s}$.
- The prior distribution for Θ_s is normal distribution with mean μ =6 and standard deviation τ=1.5.

Therefore, the posterior distribution of Θ_s is normal with mean μ_s* and standard deviation τ_s*, where:

$$\mu_s^* = \frac{\frac{\mu}{\tau^2} + P_s n_s \bar{x}_s}{\frac{1}{\tau^2} + P_s n_s} = \frac{2.67 + 48.04 P_s}{0.44 + 10 P_s} \quad \text{and} \quad \tau_s^* = \left(\frac{1}{\tau^2} + P_s n_s\right)^{-1/2} = (0.44 + 10 P_s)^{-1/2}$$

Conditional on Θ_s, the posterior distribution for P_s is independent of Θ_b and P_b, and can be found by the equations given in Module 6 for updating the distribution of the precision conditional on the mean.

The posterior distribution of P_s given the surface observations and Θ_s is Gamma with shape α_s* and scale β_s*, where

$\alpha_s^* = \alpha + ½\, n_s = 4.5 + 5 = 9.5$, and

$$\beta_s^* = \left(1/0.19 + \tfrac{1}{2}\Sigma_{i=1}^{10}(x_{is} - \Theta_s)^2\right)^{-1} = \left(\frac{1}{0.19} + 0.5\left(234.7 + 2 \times 48.04\Theta_s + \Theta_s^2\right)\right)^{-1}$$

Bottom:

Conditional on P_b , the posterior distribution for Θ_b is independent of Θ_s and P_s, and can be found by normal-normal conjugate updating:

- The observations are normal with mean Θ_b and standard deviation $\Sigma_b = 1/\sqrt{P_b}$.
- The prior distribution for Θ_b is normal distribution with mean μ =6 and standard deviation τ=1.5.

Therefore, the posterior distribution of Θ_b is normal with mean μ_s* and standard deviation τ_s*, where:

$$\mu_b^* = \frac{\frac{\mu}{\tau^2} + P_b n_b \bar{x}_b}{\frac{1}{\tau^2} + P_b n_b} = \frac{2.67 + 58.39 P_b}{0.44 + 10 P_b} \quad \text{and} \quad \tau^* = \left(\frac{1}{\tau^2} + P_b n_b\right)^{-1/2} = (0.44 + 10 P_b)^{-1/2}$$

Conditional on Θ_b, the posterior distribution for P_b is independent of Θ_s and P_s, and can be found by the equations given in Module 6 for updating the distribution of the precision conditional on the mean.

The posterior distribution of P_b given the surface observations and Θ_b is Gamma with shape α_b* and scale β_b*, where

$\alpha_b^* = \alpha + ½\, n_b = 4.5 + 5 = 9.5$, and

$$\beta_b^* = \left(1/0.19 + \tfrac{1}{2}\Sigma_{i=1}^{10}(x_{ib} - \Theta_b)^2\right)^{-1} == \left(\frac{1}{0.19} + 0.5\left(350.2 + 2 \times 58.39\Theta_b + \Theta_b^2\right)\right)^{-1}$$

***Problem 2*:**

Using the distributions you found in Part 1, draw 10,000 Gibbs samples of $(\Theta_s, \Theta_b, P_s, P_b)$. Estimate 90% credible intervals for Θ_s, Θ_b, $\Sigma_s = P_s^{-1/2}$, $\Sigma_b = P_b^{-1/2}$, and Θ_b - Θ_s.

***Solution*:**
This problem can be solved by modifying the R code provided for the reaction times example. We can use either the hand-designed Gibbs sampler or the JAGS model. R code is provided for both.

If we are hand-coding the Gibbs sampler, it works just like the examples we have seen, except that we have to repeat the process for both the surface and the bottom parameters and observations. The R code provided with this solution does this.

If we are using JAGS, we can define a single JAGS model and call it twice, once for surface and once for bottom. To use JAGS, we need to install it using the procedure described in M6.4, page 5.

To use JAGS, the first step is to define a model. I defined the following JAGS model and saved it in a file called `wolfriver.model.jags`. We have to remember that JAGS uses mean/precision for normal distributions and shape/rate for gamma distributions, and does not allow other paramterizations. Also JAGS does not allow arithmetic in arguments to a density function, so we need to define a new variable for precision and rate, and use those in the density functions.

```
model {
   for(i in 1:n) {
          x[i]~dnorm(theta,rho)  # mean theta precision rho
   }
   prec = 1/1.5^2           # JAGS uses precision in dnorm
   theta~dnorm(6,prec)      # mean 6, stdv 1.5
   rate = 1/0.19            # JAGS uses rate in dgamma
   rho~dgamma(4.5,rate)     # shape 4.5, scale 0.19
}
```

Then I called this model twice, once for surface and once for bottom. The model file uses the variables x and n, so before calling it I set x to the data and n to the number of observations. The code for doing this is:

```
## Fit surface posterior distribution
n=length(wolf.surface)  # number of observations
x=wolf.surface          # observations
surface.fit <- jags(data=list("x", "n"),
                           inits=function(){list("theta"=c(mean(x)),
                              "rho"=c(1/var(x)))},
                           parameters.to.save = c("theta", "rho"),
                           n.chains=1, n.iter=10000, n.burnin=0,n.thin=1,
                           model.file="wolfriver.model.jags")


## Fit bottom posterior distribution
n=length(wolf.bottom)  # number of observations
x=wolf.bottom          # observations
bottom.fit <- jags(data=list("x", "n"),
                    inits=function(){list("theta"=c(mean(x)),
                           "rho"=c(1/var(x)))},
                    parameters.to.save = c("theta", "rho"),
                    n.chains=1, n.iter=10000, n.burnin=0,n.thin=1,
                    model.file="wolfriver.model.jags")
```

To find 90% credible intervals for the parameters, we need to extract the chains from the JAGS fit objects. For this, we use the `MCMCchains` function in the MCMCvis package:

```
surface.chains=as.data.frame(MCMCchains(surface.fit))   # extract the chains
to data frame
quantile(surface.chains$theta,c(0.05,0.95))
quantile(1/sqrt(surface.chains$rho),c(0.05,0.95))
bottom.chains=as.data.frame(MCMCchains(bottom.fit))  # extract the chains to
data frame
quantile(bottom.chains$theta,c(0.05,0.95))
quantile(1/sqrt(bottom.chains$rho),c(0.05,0.95))
quantile(bottom.chains$theta - surface.chains$theta,c(0.05,0.95))
```

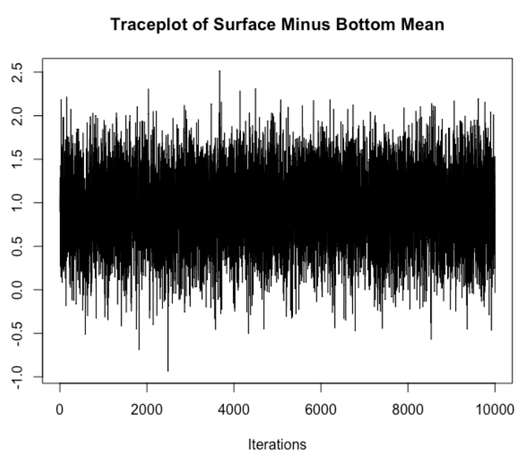
The resulting intervals will be slightly different every run, but for the run I did, they were:

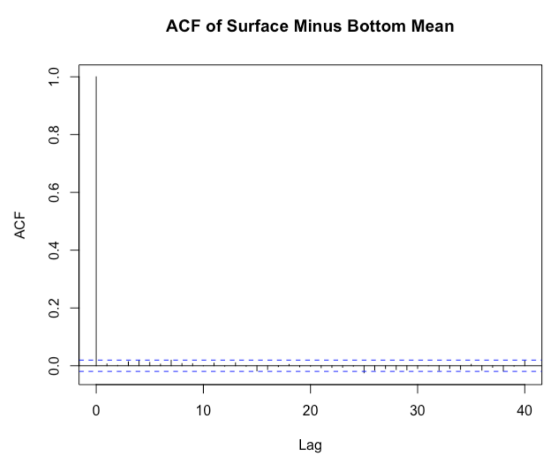
- For Θ_s: [4.37, 5.32]
- For Θ_b: [5.30, 6.40]
- For $P_s^{-1/2}$: [0.70, 1.22]
- For $P_b^{-1/2}$: [0.82, 1.44]
- For Θ_b - Θ_s: [0.29, 1.74]

***Problem 3*:**
Do a traceplot of Θ_b - Θ_s. Find the autocorrelation function of Θ_b - Θ_s and the effective sample size for your Monte Carlo sample for Θ_b - Θ_s.

***Solution*:**
To use the traceplot fuction, we need to convert the JAGS fit object, or the samples from the custom Gibbs sampler, to an MCMC object. We do this with `as.mcmc`. Commands for traceplot and acf are:

```
traceplot(as.mcmc(bottom.chains$theta - surface.chains$theta),
          main="Traceplot of Surface Minus Bottom Mean")
acf(bottom.chains$theta - surface.chains$theta,
    main="ACF of Surface Minus Bottom Mean")
```





The effective sample size was 10,000. I found it using the effectiveSize function in R.

***Problem 4*:**
Comment on your results. Compare with Assignment 6.

***Solution*:**
The MCMC diagnostics show very little autocorrelation, a full 10,000 effective sample size, and a traceplot that shows convergence. Therefore, we can be confident in the conclusions we draw from our sample.

We can directly compare the credible intervals for Θ_s and Θ_b with the exact intervals from Assignment M5:

	90% Interval for Θ_s	90% Interval for Θ_b
Gibbs sampling (6B)	[4.37, 5.32]	[5.30, 6.40]
Direct MC (5B)	[4.44, 5.16]	[5.26, 6.43]
Exact (5A)	[4.44, 5.17]	[5.25, 6.43]

The intervals from Assignment M5 are almost identical. The intervals from this assignment are also similar to the results of M5. They are somewhat different because of the influence of the prior distribution. The Assignment 6 interval for Θ_b is completely contained within the intervals from Assignment 5. The interval is narrower because the samples are centered around the prior mean of 6, thus confirming our prior estimate. The Assignment 6 interval for Θ_s is pulled toward the prior mean of 6, and is wider than the interval for Assignment 5 because the precision is pulled down by the difference between the sample mean and the prior mean.

In this assignment, we were asked for credible intervals on the standard deviations, whereas in M5 we found intervals for the precision. We can find intervals for $P_s^{-1/2}$ and $P_b^{-1/2}$ using R commands `quantile(surface.chains$rho,c(0.05,0.95))` and `quantile(bottom.chains$rho,c(0.05,0.95))`:

- For P_s: [0.676, 2.060]
- For P_b: [0.484, 1.474]

	90% Interval for P_s	90% Interval for P_b
Gibbs sampling (6B)	[0.676, 2.06]	[0.484, 1.47]
Direct MC (5B)	[0.930, 4.67]	[0.355, 1.83]
Exact (5A)	[0.927, 4.72]	[0.359, 1.83]

The intervals for P_s and P_b are narrower for Assignment M6 than for Assignment M5, reflecting the added information from the prior distribution.