Homework 7

Shao-Ting Chiu (UIN:433002162)

10/30/22

Table of contents

Description

- Course: STAT638, 2022 Fall

  Read Chapter 7 in Hoff. Then, do the following exercises: 7.1, 7.3, 7.4.

  Problem 7.1 considers the standard/joint Jeffreys prior (as opposed to the independent Jeffreys prior considered on the lecture slides). You may find the following hints useful:

- You can write $y_i - \theta$ as $(y_i - \bar{y}) + (\bar{y} - \theta)$ and expand the quadratic form in the exponent in the multivariate normal likelihood accordingly.
- $\sum_i b_i^T A c = c^T A(\Sigma_i b_i)$
- Brute-force integration can sometimes be avoided if the integrand is proportional to a known density (e.g., multivariate normal), as any density integrates to 1 and the normalizing constant is known for known densities. For 7.3, note that the `rWishart()` function in R returns a three-dimensional array, so we have to index the array as [„1] to get to the actual matrix located within the array.

Computational Environment Setup

Third-party libraries

```julia
using Pkg
Pkg.activate("hw7")
using Distributions
using DataFrames
using Turing
using Plots
using DelimitedFiles
using LinearAlgebra
using Statistics
using Turing
```

```
  Activating project at `~/Documents/GitHub/STAT638_Applied-Bayes-Methods/hw/hw7`
```

Version

```julia
Pkg.status()
VERSION
```

```
Status `~/Documents/GitHub/STAT638_Applied-Bayes-Methods/hw/hw7/Project.toml`
  [a93c6f00] DataFrames v1.4.2
  [31c24e10] Distributions v0.25.76
  [91a5bcdd] Plots v1.35.5
```

```
  [fce5fe82] Turing v0.21.12
  [8bb1440f] DelimitedFiles
```

```
v"1.8.2"
```

Problem 7.1

Jeffrey's prior: For the multivariate normal model, Jeffreys' rule for generating a prior distribution on (θ, Σ) gives $p_J(\theta, \Sigma) \propto |\Sigma|^{-(p+2)/2}$.

(a)

Explain why the function p_J cannot actually be a probability density for (θ, Σ).

The density is independent of θ. The integration can be infinity and beyond 1.

(b)

Let $p_J(\theta, \Sigma|y_1, \dots, y_n)$ be the probability density that is proportional to $p_J(\theta, \Sigma) \times p(y_1, \dots, y_n|\theta, \Sigma)$. Obtain the form of $p_J(\theta, \Sigma|y_1, \dots, y_n)$, $p_J(\theta|\Sigma, y_1, \dots, y_n)$ and $p_J(\Sigma|y_1, \dots, y_n)$.

$$
\begin{aligned}
p_J(\theta, \Sigma|y_1, \dots, y_n) &\propto p_J(\theta, \Sigma) \times p(y_1, \dots, y_n|\theta, \Sigma) && (1)\\
&\propto |\Sigma|^{-\frac{p+2}{2}} \times \left[|\Sigma|^{-\frac{n}{2}} \exp\left(-tr(S_\theta \Sigma^{-1})\right)\big/ 2\right] && (2)\\
&\propto |\Sigma|^{-\frac{p+n+2}{2}} \exp\left(-tr(S_\theta \Sigma^{-1})/2\right) && (3)
\end{aligned}
$$

$$
\begin{aligned}
p_J(\theta|\Sigma, y_1, \dots, y_n) &\propto \exp\left[-\sum_{i=1}^{n}(y_i - \theta)^T \Sigma^{-1}(y_i - \theta)/2\right] && (4)\\
&\propto \exp\left[-n(\bar{y} - \theta)^T \Sigma^{-1}(\bar{y} - \theta)/2\right] && (5)\\
&\propto Normal(\theta; \bar{y}, \frac{\Sigma}{n}) && (6)
\end{aligned}
$$

$$
\begin{aligned}
p_J(\Sigma|y_1, \dots, y_n, \theta) &\propto |\Sigma|^{-\frac{p+n+2}{2}} \exp\left(-tr(S_\theta \Sigma^{-1})/2\right) && (7)\\
&\propto inverse-Wishart(\Sigma; n+1, S_\theta^{-1}) && (8)
\end{aligned}
$$

Problem 7.3

Australian crab data: The files `bluecrab.dat` and `orangecrab.dat` contain measurements of body depth (Y_1) and rear width (Y_2), in millimeters, made on 50 male crabs from each of two species, blud and orange. We will model these data using a bivariate normal distribution.

```
dblue = readdlm("data/bluecrab.dat")
doran = readdlm("data/orangecrab.dat");
```

(a)

For each of the two species, obtain posterior distributions of the population mean θ and covariance matrix Σ as follows: Using the semiconjugate prior distributions for θ and Σ, set μ_0 equal to the sample mean of the data, Λ_0 and S_0 equal to the sample covariance matrix and $\nu_0 = 4$. Obtain 10000 posterior samples of θ and Σ. Note that this prior distribution lossely centers the parameters around empirical estimates based on the observed data (and is very similar to the unit information prior described in the previous exericise). It cannot be consitered as our true prior distribution, as it was derived from the observed data. However, it can roughly considered as the prior distribution of someone with weak but unbiased information.

- $p(\theta) = \exp\left[-\frac{1}{2}\theta^T A_0 \theta + \theta^T b_0\right] = multivariate-normal(\mu_0, \Lambda_0)$
  - $A_0 = \Lambda_0^{-1}$
  - $b_0 = \Lambda_0^{-1}\mu_0$
- $p(\Sigma) = inverse-Whishart(\nu_0, S_0^{-1})$

```
S = 10000
function sampling(crab)
    n, p = size(crab)
      = transpose(mean(crab, dims=1))
    Λ  = S = cov(crab)
      = 4
     s = zeros(S, p)
    Σs = zeros(S, p, p);

    # Gibbs sampling
    for s in 1:S
        # update
```

```
        Λ  = inv(inv(Λ ) + n*inv(S ))
          = Λ  * (inv(Λ )*   + n*inv(S )* )
         = rand(MvNormal( vec( ), Λ ))

        # update Σ
        res = crab .- reshape( , 1, p)
        S   = transpose(res) * res
        S  = S  + S
        Σ = rand(InverseWishart(  + n, S ))
        # Store data
         s[s,:] =
        Σs[s,:, :] = Σ
    end
    return  s, Σs
end


 bs, Σbs = sampling(dblue)
 os, Σos = sampling(doran);
```

(b)

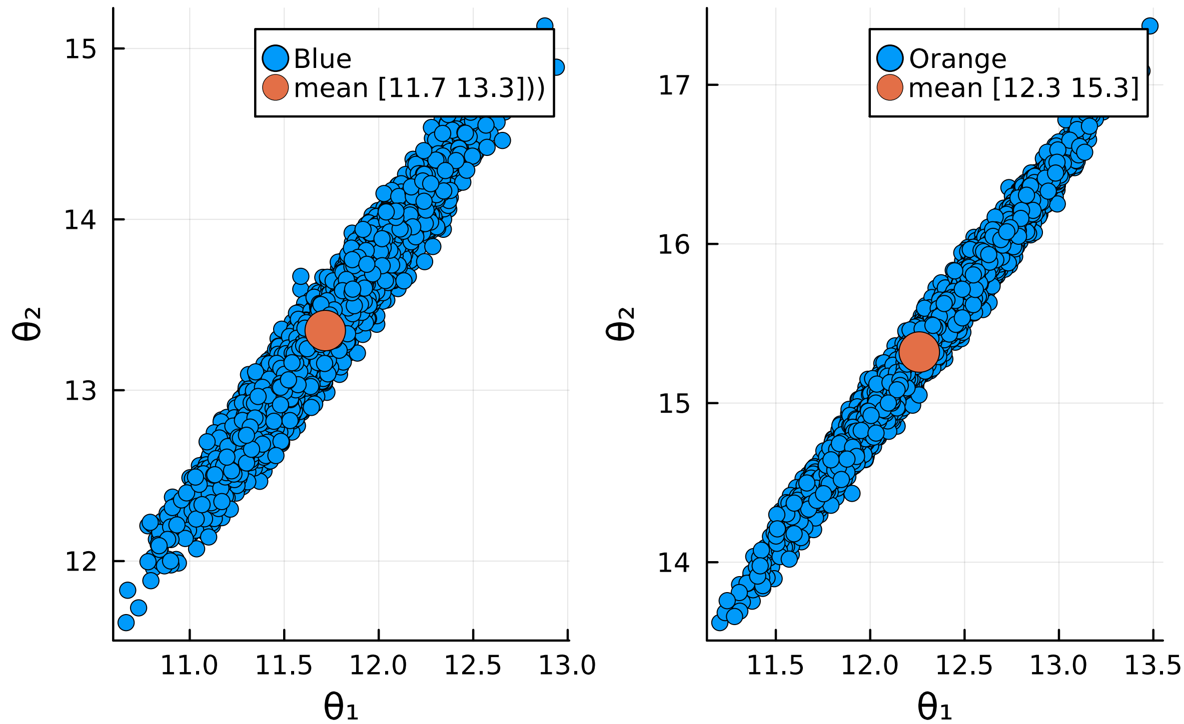
Plot values of $\theta = (\theta_1, \theta_2)'$ for each group and compare. Describe any size differences between the two groups.

```
plot
pb = scatter( bs[:,1],  bs[:,2], label="Blue")
po = scatter( os[:,1],  os[:,2], label="Orange")

 b = mean( bs, dims=1)
 o = mean( os, dims=1)

scatter!(pb, [ b[1]], [ b[2]], label="mean $(round.( b; digits = 1))))", markersize = 10)
scatter!(po, [ o[1]], [ o[2]], label="mean $(round.( o; digits = 1))", markersize = 10)

plot(pb, po, layout = (1, 2), xlabel=" ", ylabel=" ")
```




```
mean( os[:,1] .> bs[:,1])
```

```
0.8992
```

```
mean( os[:,2] .> bs[:,2])
```

```
0.9984
```

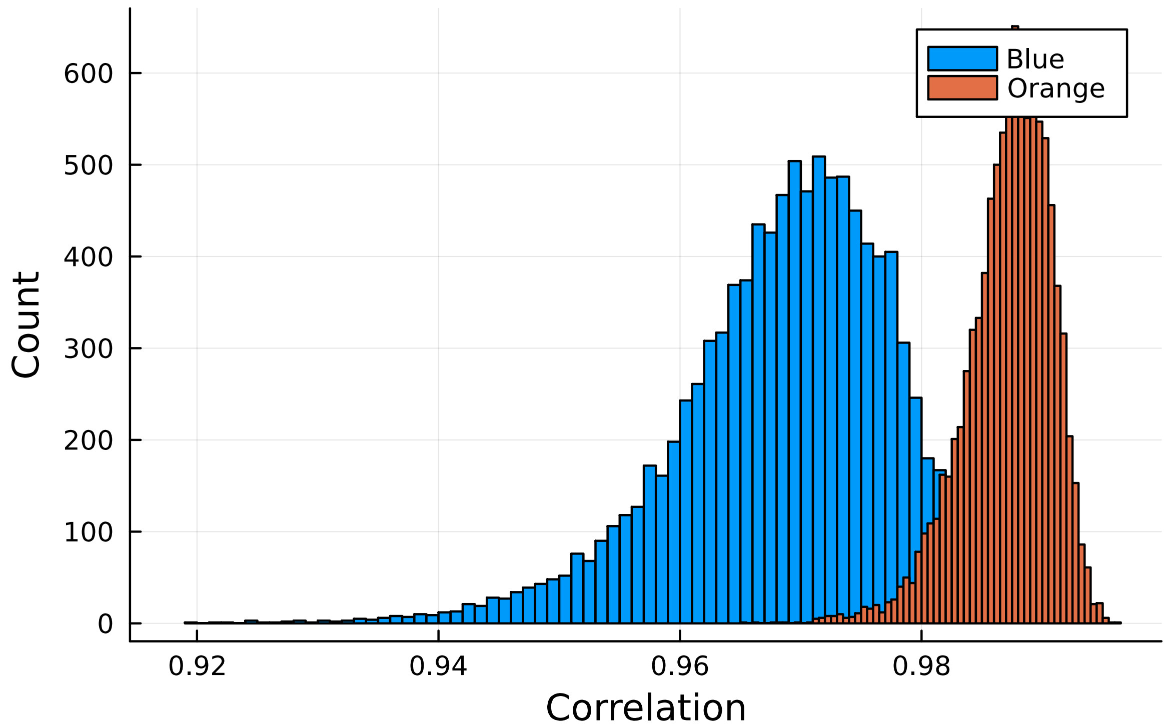
(c)

From each covariance matrix obtained from the Gibbs sampler, obtain the corresponding correlation coefficient. From these values, plot posterior densities of the correlations ρ_{blue} and ρ_{orange} for the two groups. Evaluate differences between the two species by comparing these posterior distributions. In particular, obtain an approximation to $Pr(\rho_{\text{blue}} < \rho_{\text{orange}} | y_{\text{blue}}, y_{\text{orange}})$. What do the results suggest about differences between the two populations?

```julia
correlation(covmat) = covmat[1,2] / sqrt(covmat[1,1] * covmat[2,2])

corrbs = [correlation(Σbs[i,:,:]) for i in 1:S]
corros = [correlation(Σos[i,:,:]) for i in 1:S];

h = histogram(corrbs, label="Blue", xlabel="Correlation")
histogram!(h, corros, label="Orange", ylabel="Count")
```



$Pr(\rho_{\text{blue}} < \rho_{\text{orange}} | y_{\text{blue}}, y_{\text{orange}})$ is

```julia
mean(corrbs .< corros)
```

```
0.9895
```

Problem 7.4

Marriage data: The file `agehw.dat` contains data on the ages of 100 married couples sampled from the U.S. population.

```
dagew = readdlm("data/agehw.dat")[2:end, :]
size(dagew)
```

```
(100, 2)
```

(a)

Before you look at the data, use your own knowledge to formulate a semiconjugate prior distribution for $\theta = (\theta_h, \theta_w)^T$ and Σ, where θ_h, θ_w are mean husband and wife ages, and Σ is the covariance matrix.

- $\mu_0 = (50, 50)^T$
- prior corrleation: 0.7, variance 13
  - $0.7 = \frac{\sigma_{1,2}}{169}$
  - $\sigma_{1,2} = 118.3$
- $\Lambda = \begin{bmatrix} 169 & 118.3 \\ 118.3 & 169 \end{bmatrix}$
- Set $S_0^{-1} = \Lambda_0$
- $\nu_0 = p + 2 = 4$

```
n, p = size(dagew);

  = ones(p,1) .* transpose(mean(dagew, dims=1))
Λ  = S  = [ 169 118.3 ; 118.3 169]
  = p + 2;
```

(b)

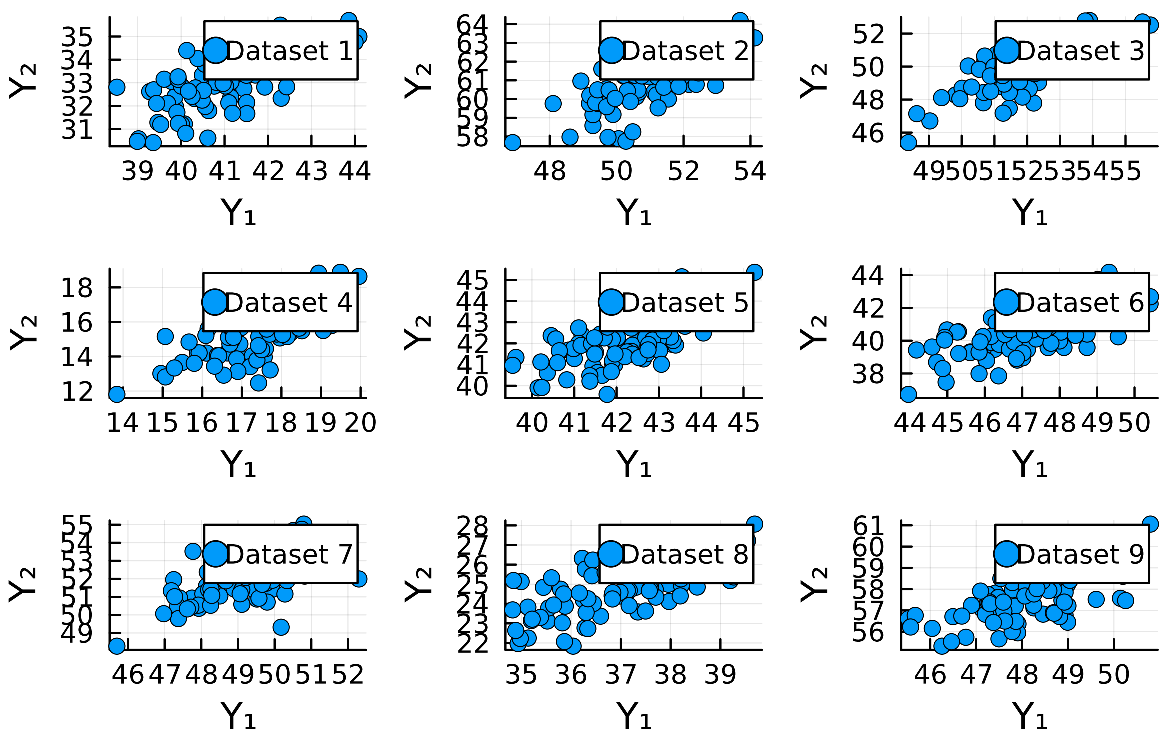
Generate a *prior predictive dataset* of size $n = 100$, by sampling (θ, Σ) from your prior distribution and then simulating $Y_1, \ldots, Y_n \sim i.i.d.$ multivariate normal (θ, Σ). Generate several such datasets, make bivariate scatterplots for each dataset, and make sure they roughly represent your prior beliefs about what such a dataset would actually look like. If your prior predictive datasets do not confirm to your beliefs, go back to part (a) and formulate a new prior. Report the prior that you eventually decide upon, and provide scatterplots for at least three prior predictive datasets.

```
N = 100
S = 9

Ypreds = zeros(S, p, N)
for i in 1:S
      = rand(MvNormal( vec( ), Λ ))
    Σ = rand(InverseWishart(  + n, S ))
    Ypreds[i,:, :] = rand(MvNormal( , Σ), N)
end

pvers = [plot() for i in 1:S]
for i in 1:S
    scatter!(pvers[i], Ypreds[i, 1, :], Ypreds[i, 2, :], label="Dataset $i")
end

plot(pvers..., layout = (3, 3), xlabel="Y ", ylabel="Y ")
```



(c)

Using your prior distribution and the 100 values in the dataset, obtain an MCMC approximation to $p(\theta, \Sigma | y_1, \dots, y_{100})$. Plot the joint posterior distribution of θ_h and θ_w, and also the marginal posterior density of the correlation between Y_h and Y_w, the ages of a husband and wife. Obtain 95% posterior confidence intervals for θ_h, θ_w and the correlation coefficient.

```
S = 10000

function mcmc(data,  , Λ , S ,  )
    y  = mean(data, dims=1)
    n, p = size(data)

     s = zeros(S, p)
    Σs = zeros(S, p, p)
    Σ = cov(data)
    for i in 1:S
        #update
        Λ  = inv(inv(Λ ) + n * inv(Σ))
        Λ [1,2] = Λ [2,1]

          = Λ  * (inv(Λ )*   + n*inv(Σ) * transpose(y ))


          = rand(MvNormal( vec( ), Λ ))

        #update Σ
        res = data .- reshape( , 1, p)
        S   = transpose(res) * res
        S  = S  + S
        Σ = rand(InverseWishart(  + n, S ))
        # Store data
         s[i,:] = 
        Σs[i,:, :] = Σ
    end

    return  s, Σs
end

 s, Σs = mcmc(dagew,  , Λ , S ,  );
corrs = [correlation(Σs[i,:,:]) for i in 1:S];
```

Husband Quantiles

```
quantile( s[:,1], [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 41.80243819438264
 44.43925834268787
 47.12062893245589
```

Wife Quantiles

```
quantile( s[:,2], [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 38.44197286817298
 40.90455740795811
 43.38665068400079
```

Correlation Quantiles

```
quantile(corrs, [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 0.8593445512083464
 0.9025330899053765
 0.932486425260002
```

(d)

> Obtain 95% posterior confidence intervals for θ_h, θ_ω and the correlation coefficient using the following prior distributions:
>
> 1. Jeffrey's prior, described in Exercise 7.1;
> 2. The unit information prior, described in Exercise 7.2;
> 3. A "diffuse prior" with $\mu_0 = 0, \Lambda_0 = 10^5 \times I, S_0 = 1000 \times I$ and $v_0 = 3$.

Part I**

```
 s, Σs = mcmc(dagew,  , cov(dagew), cov(dagew), size(dagew)[1]+1);
corrs = [correlation(Σs[i,:,:]) for i in 1:S];
```

Husband Quantiles

```
quantile( s[:,1], [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 42.50765653257726
 44.42342072578426
 46.31894574583449
```

Wife Quantiles

```
quantile( s[:,2], [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 39.09343754289234
 40.89679476321716
 42.67174532022887
```

Correlation Quantiles

```
quantile(corrs, [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 0.874786310957532
 0.9037849046320076
 0.9264354343002099
```

Part II

```
 s, Σs = mcmc(dagew, transpose(mean(dagew, dims=1)), cov(dagew)/100., cov(dagew)/100., 2. +
corrs = [correlation(Σs[i,:,:]) for i in 1:S];
```

Husband Quantiles

```
quantile( s[:,1], [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 42.495025098123264
 44.41899652097021
 46.29688526783925
```

Wife Quantiles

```
quantile( s[:,2], [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 39.10629931234869
 40.88107967101249
 42.63850757744455
```

Correlation Quantiles

```
quantile(corrs, [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 0.8620945837845174
 0.9041431265569284
 0.9339649324960588
```

Part III

```
 s, Σs = mcmc(dagew, [0;0], [10^5 0; 0 10^5], [10^3 0; 0 10^3], 3);
corrs = [correlation(Σs[i,:,:]) for i in 1:S];
```

Husband Quantiles

```
quantile( s[:,1], [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 41.662362616715676
 44.387526712236955
 47.177264903268956
```

Wife Quantiles

```
quantile( s[:,2], [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 38.270618143361354
 40.86320210002676
 43.43194743045855
```

Correlation Quantiles

```
quantile(corrs, [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 0.7923898092751392
 0.8546879001574484
 0.9001162090969571
```

(e)

> Compare the confidence intervals from (d) to those obtained in (c). Discuss whether or not you think that your prior information is helpful in estimating θ and Σ, or if you think one of the alternatives in (d) is preferable. What about if the sample size were much smaller, say $n = 25$?

The prior information does not matter because the sample size is large. No matter how prior is setup, the posterior distribution is similar. However, for smaller sample size, those approaches may differ.

```
   = [50.; 50.]
Λ  = S  = [ 169 118.3 ; 118.3 169]
   = p + 2+9;

 s, Σs = mcmc(dagew,   , Λ , S ,   )
```

```
([44.84524452006495 41.22962018438789; 41.98582031644667 39.26963721685027; … ; 43.3108222288
```

Husband Quantiles

```julia
quantile( s[:,1], [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 41.936805963643764
 44.47156121792714
 47.05055100131598
```

Wife Quantiles

```julia
quantile( s[:,2], [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 38.53384171598369
 40.94720893550817
 43.33555818299781
```

Correlation Quantiles

```julia
quantile(corrs, [0.025, 0.5, 0.975])
```

```
3-element Vector{Float64}:
 0.7923898092751392
 0.8546879001574484
 0.9001162090969571
```

References