Chapter 4: Solutions to Exercises

Exercise 1

Because $X_{n+1} \perp X_{1:n} \mid \theta$, we can sample from $X_{n+1}|x_{1:n}$ as follows:

1. sample $\theta \sim p(\theta|x_{1:n})$, and then

2. sample $x_{n+1} \sim p(x_{n+1}|\theta)$.

Since $p(x_{n+1}|\theta) = \mathcal{N}(x_{n+1}|\theta, \lambda^{-1})$, step 2 is equivalent to setting $x_{n+1} = \theta + z$, where z is sampled from $\mathcal{N}(0, \lambda^{-1})$, independently. Thus, the distribution of $X_{n+1}|x_{1:n}$ is the same as the distribution of $\tilde{\boldsymbol{\theta}} + Z$, where

$$\tilde{\boldsymbol{\theta}} \sim p(\theta|x_{1:n})$$
$$Z \sim \mathcal{N}(0, \lambda^{-1})$$

independently. Since we know $p(\theta|x_{1:n}) = \mathcal{N}(\theta|M, L^{-1})$, where $L = \lambda_0 + n\lambda$ and

$$M = \frac{\lambda_0\mu_0 + \lambda\sum_{i=1}^{n}x_i}{\lambda_0 + n\lambda},$$

then by the formula for the sum of independent Normals, we have that

$$\tilde{\boldsymbol{\theta}} + Z \sim \mathcal{N}(M, \, L^{-1} + \lambda^{-1}).$$

Hence,

$$p(x_{n+1}|x_{1:n}) = \mathcal{N}(x_{n+1} \mid M, \, L^{-1} + \lambda^{-1}).$$

Exercise 2

Yum!

Exercise 3

In some ways, an i.i.d. Normal model is appropriate, but in other ways, it is lacking—specifically, the Normal assumption is probably justified, but the i.i.d. assumption may not be. We could reasonably expect the distribution for any given year to be approximately Normally distributed, because annual snowfall represents the sum of the snowfall over many days, and even though there is dependence between the snowfall in successive days, the amount of dependence probably decays quickly enough that the central limit theorem takes
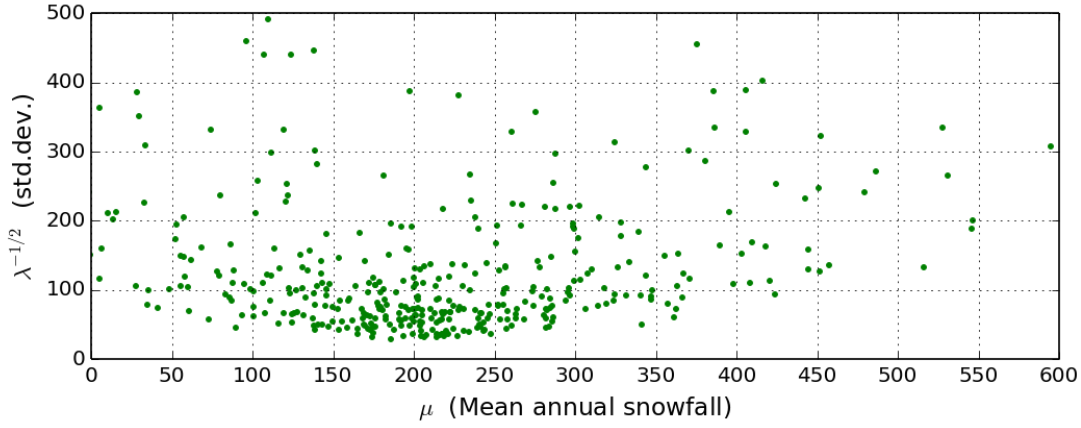
Figure 1: Scatterplot of samples from the prior for the snowfall exercise.

effect. (Note: Technically, it cannot be exactly Normal, since snowfall cannot take negative values—however, when the mean is relatively large compared to the standard deviation, the probability of negative values is exceedingly small.) One possible issue with the i.i.d. assumption is that there may be changes in climate over time, causing the amount of snow to increase or decrease (violating the identically distributed assumption). Another possible issue is that there may be dependence between successive years (violating the independence assumption).

Exercise 4

Based on my prior knowledge before looking at the data, my subjective prior was:

- $m = 200$ inches (I thought the mean snowfall for each city would be around 15–18 feet)

- $c = 1$ (I was not certain about this choice of m, so I gave it a weight of 1 prior "sample")

- $a = 1/2$ (I was uncertain about the precision, so I chose a weight of 1 prior "sample")

- $b = 80^2 a$ (I thought the standard deviation of the annual snowfall would be around 80 inches)

I visualized this prior by making a scatterplot of samples from it, see Figure 1. The y-axis is in terms of the standard deviation $\lambda^{-1/2}$, since this is more interpretable than precision. Note that this prior puts non-negligible mass on negative values of the mean snowfall, which is not particularly realistic—it might be better to truncate the prior to make all of the values positive, however, this would probably make very little difference to the posterior.

Now, let's visualize the data in a histogram. See Figure 2. From visual inspection, it looks like Valdez has a slight edge—let's see what the posterior says. Using my prior, the
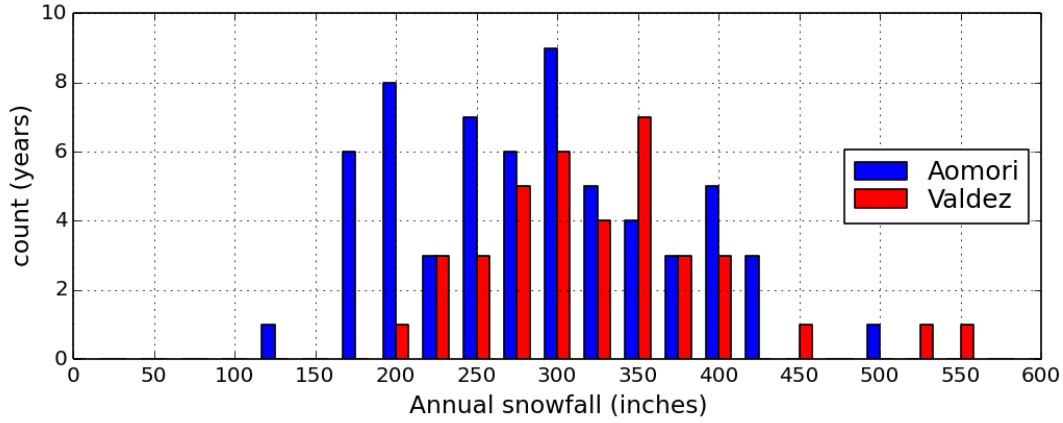
2

Figure 2: Histograms of the annual snowfall for Aomori and Valdez.

posterior parameters are:

for Aomori:
$$M = 285.5$$
$$C = 62$$
$$A = 31$$
$$B = 204034$$
for Valdez:
$$M = 322.2$$
$$C = 39$$
$$A = 19.5$$
$$B = 115204$$

See Figure 3 for a scatterplot of samples from each of the posteriors. Visually, the posterior does indeed seem to favor Valdez as having higher mean snowfall. Using 10^6 samples from each posterior, I approximate the posterior probability that Valdez has higher mean snowfall than Aomori to be:

$$\mathbb{P}(\boldsymbol{\mu}_V > \boldsymbol{\mu}_A \mid x_{1:n_A}, y_{1:n_V}) \approx 0.987$$

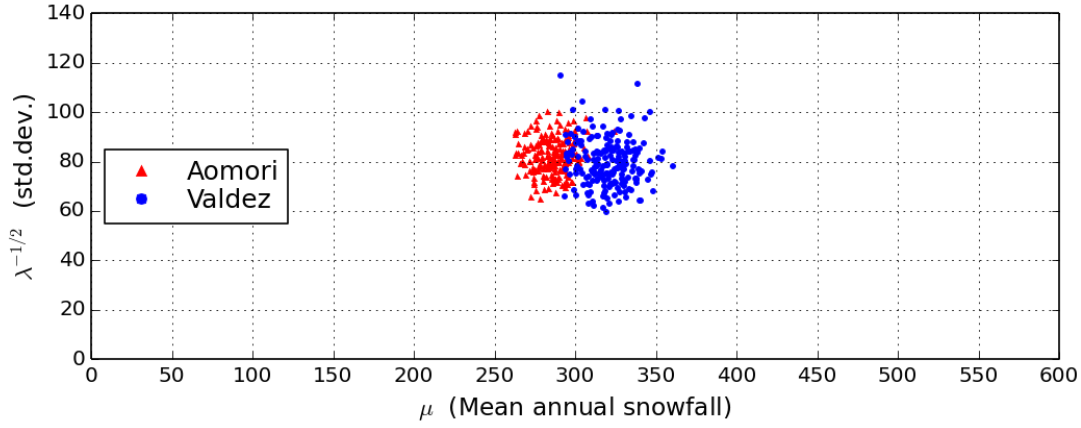where $x_{1:n_A}$ is the Aomori data and $y_{1:n_V}$ is the Valdez data.

Figure 3: Scatterplot of samples from the posteriors for Aomori and Valdez.

Exercise 5

I tried the following three different settings of hyperparameters, varying m, b, and a separately:

$$m = 500, c = 1, a = 1/2, b = 80^2 a$$
$$m = 200, c = 1, a = 1/2, b = 20^2 a$$
$$m = 200, c = 1, a = 10/2, b = 80^2 a$$

This gave the following three values for the posterior probability that Valdez has higher mean snowfall than Aomori:

$$0.989$$
$$0.987$$
$$0.987$$

This suggests that the result does not depend that strongly on the choice of prior. (A more in-depth sensitivity analysis would be needed to verify this.) See Figure 4.

There is enough data that we would have make the prior extremely different in order to get a significantly different result. This can be quantified more precisely by looking at how the posterior parameters M, C, A, B depend on m, c, a, b, the sample size, and the summary statistics of the data.
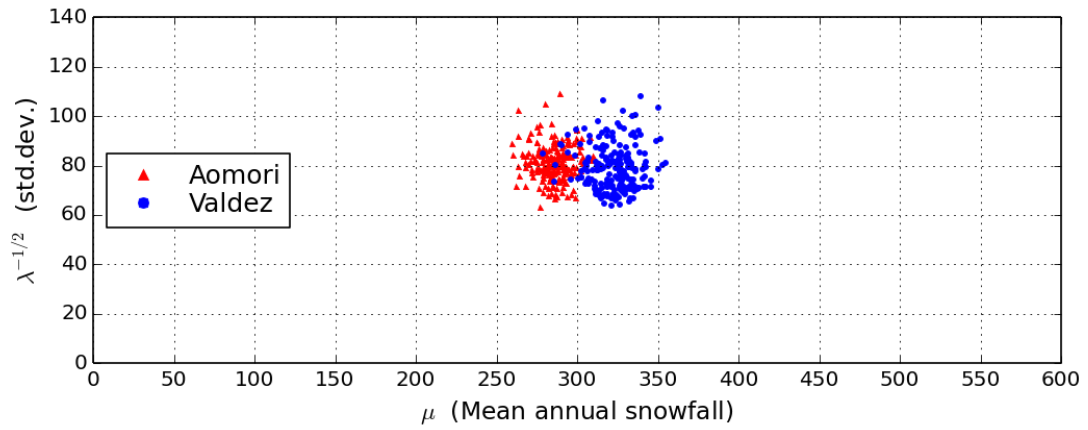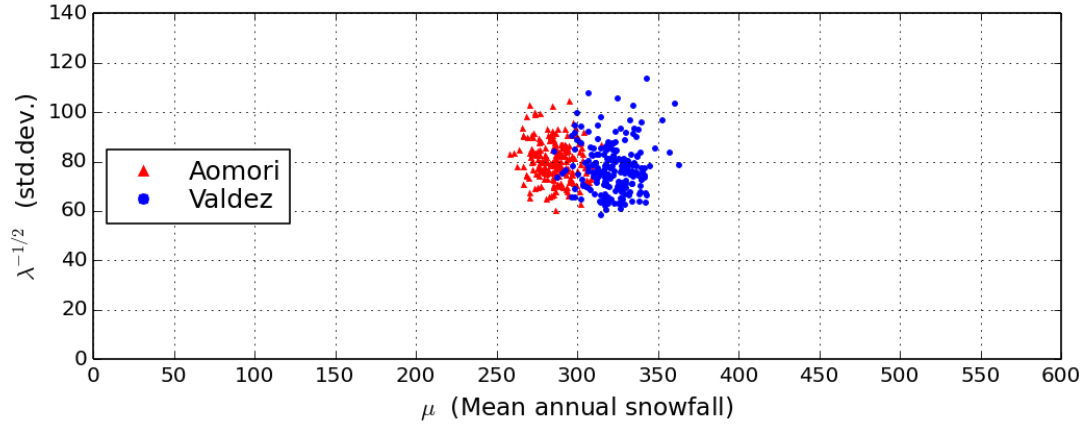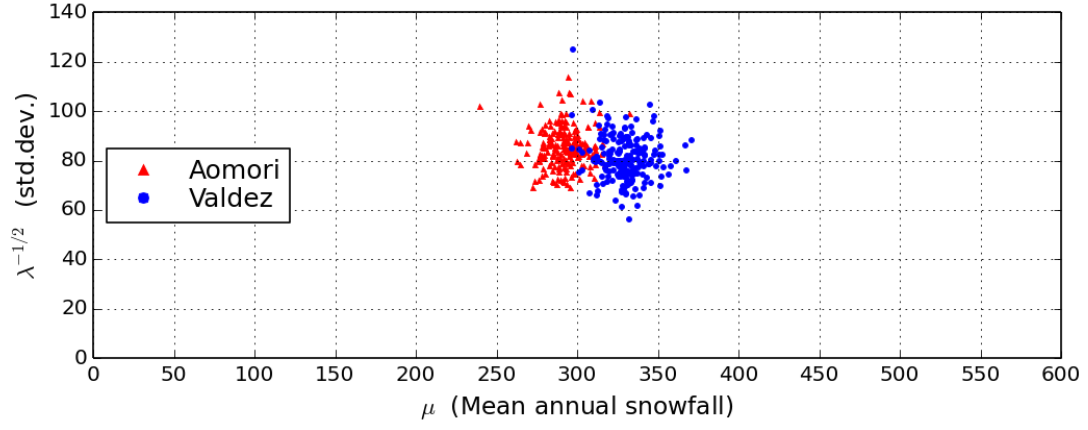
Figure 4: Scatterplots of samples from the posteriors for Aomori and Valdez, for the three different choices of hyperparameters.

A Source code in Julia language

Exercise 2

```julia
module Snow

using PyPlot, Distributions
draw_now() = (pause(0.001); get_current_fig_manager()[:window][:raise_]())

# Aomori
x = [188.6, 244.9, 255.9, 329.1, 244.5, 167.7, 298.4, 274.0, 241.3, 288.2,
    208.3, 311.4, 273.2, 395.3, 353.5, 365.7, 420.5, 303.1, 183.9, 229.9,
    359.1, 355.5, 294.5, 423.6, 339.8, 210.2, 318.5, 320.1, 366.5, 305.9,
    434.3, 382.3, 497.2, 319.3, 398.0, 183.9, 201.6, 240.6, 209.4, 174.4,
    279.5, 278.7, 301.6, 196.9, 224.0, 406.7, 300.4, 404.3, 284.3, 312.6,
    203.9, 410.6, 233.1, 131.9, 167.7, 174.8, 205.1, 251.6, 299.6, 274.4,
    248.0]
# Valdez
y = [351.0, 379.3, 196.1, 312.3, 301.4, 240.6, 257.6, 304.5, 296.0, 338.8,
    299.9, 384.7, 353.5, 312.8, 550.7, 327.1, 515.8, 343.4, 341.6, 396.9,
    267.3, 230.6, 277.4, 341.0, 377.0, 391.3, 337.0, 250.4, 353.7, 307.7,
    237.5, 275.2, 271.4, 266.5, 318.7, 215.5, 438.3, 404.6]

function NormalGamma_posterior(x,m,c,a,b)
    mx = mean(x)
    vx = mean((x-mx).^2)
    n = length(x)
    C = c + n
    A = a + n/2
    M = (c*m + n*mx)/C
    B = b + 0.5*n*vx + 0.5*(c*n/C)*(mx-m)^2
    return M,C,A,B,mx,vx
end

function NormalGamma_sample(n,m,c,a,b)
    lambda = rand(Gamma(a,1/b),n)
    mu = rand(Normal(0,1),n)./sqrt(c*lambda) + m
    return (mu,lambda)
end

for (rep,(m,c,a,s)) in enumerate([(200,1,1/2,80),
                                  (500,1,1/2,80),
                                  (200,1,1/2,20),
                                  (200,1,10/2,80)])
    b = s^2*a
    Mx,Cx,Ax,Bx,mx,vx = NormalGamma_posterior(x,m,c,a,b)
    My,Cy,Ay,By,my,vy = NormalGamma_posterior(y,m,c,a,b)
    sx,sy = sqrt(vx),sqrt(vy)

    lower,upper = -12.5,612.5
```

```
# Histograms
figure(1,figsize=(10,4)); clf(); hold(true); grid(true)
edges = lower:25:upper
edges,cx = hist(x,edges)
edges,cy = hist(y,edges)
bins = midpoints(edges)
bar(bins-8,cx,8,color="b",label="Aomori")
bar(bins,cy,8,color="r",label="Valdez")
xlabel("Annual snowfall (inches)",fontsize=14)
ylabel("count (years)",fontsize=14)
subplots_adjust(bottom = 0.2)
xticks(bins[1:2:end])
xlim(0,600)
ylim(0,10)
legend(numpoints=1,loc=5,labelspacing=0.1,fontsize=15)
draw_now()
savefig("snow-histogram-$rep.png",dpi=120)

# Scatterplot of prior
N = 500
mu,lambda = NormalGamma_sample(N,m,c,a,b)
figure(2,figsize=(10,4)); clf(); hold(true)
plot(mu,1.0./sqrt(lambda),"go",markersize=3,markeredgecolor="g")
xlabel("\$\\mu\$ (Mean annual snowfall)",fontsize=14)
ylabel("\$\\lambda^{-1/2}\$ (st.dev.)",fontsize=14)
subplots_adjust(bottom = 0.2)
xticks(bins[1:2:end])d
xlim(0,maximum(bins))
ylim(0,500)
grid(true)
draw_now()
savefig("snow-prior-$rep.png",dpi=120)

# Scatterplots of posterior
N = 200
mu_x,lambda_x = NormalGamma_sample(N,Mx,Cx,Ax,Bx)
mu_y,lambda_y = NormalGamma_sample(N,My,Cy,Ay,By)
figure(3,figsize=(10,4)); clf(); hold(true)
plot(mu_x,1.0./sqrt(lambda_x),"r^",label="Aomori",markersize=3,
    markeredgecolor="r")
plot(mu_y,1.0./sqrt(lambda_y),"bo",label="Valdez",markersize=3,
    markeredgecolor="b")
xlabel("\$\\mu\$ (Mean annual snowfall)",fontsize=14)
ylabel("\$\\lambda^{-1/2}\$ (std.dev.)",fontsize=14)
subplots_adjust(bottom = 0.2)
xticks(bins[1:2:end])
xlim(0,maximum(bins))
ylim(0,140)
grid(true)
```

```
        legend(loc=6,numpoints=1,labelspacing=0.1,fontsize=15,markerscale=2)
        draw_now()
        savefig("snow-posteriors-$rep.png",dpi=120)

        # Compute posterior quantities
        N = 10^6
        mu_x,lambda_x = NormalGamma_sample(N,Mx,Cx,Ax,Bx)
        mu_y,lambda_y = NormalGamma_sample(N,My,Cy,Ay,By)

        println("==================================================")
        println("Prior parameters:")
        println("prior: m,c,a,b = $m, $c, $a, $b")
        println()
        println("Summary statistics:")
        println("n(x) = ",length(x))
        println("n(y) = ",length(y))
        println("mean(x) = ",mx)
        println("mean(y) = ",my)
        println("std(x) = ",sx)
        println("std(y) = ",sy)
        println("posterior_x: M,C,A,B = $Mx, $Cx, $Ax, $Bx")
        println("posterior_y: M,C,A,B = $My, $Cy, $Ay, $By")
        println()
        println("P(mu_x > mu_y | data) = ",mean(mu_x.>mu_y))
        println("P(mu_x < mu_y | data) = ",mean(mu_x.<mu_y))
        println()
    end

end
```