Lab 4: Normal-Gamma

Olivier Binette

28/08/2020

Agenda

1. Some notes about the labs
2. Quick review of the Normal-Gamma distribution
3. Walkthrough of Lab 4 Tasks 1-3

- I've included code to help you with Tasks 4-5

4. Questions / Office Hours

Some notes about the labs

Some notes about the labs

Each Friday at the Lab:

- Review background material as needed.
- We go through Tasks 1-3 of the week's lab.
- Questions and Office Hours regarding the labs or homeworks.

Feel free to:

- Send questions / requests in advance!
- Let us know if you have requests for something to be reviewed or talked about.

Also:

- Feel free to email me at any time during the week with questions or for any reason.
- olivier.binette@gmail.com
- You can also post on Piazza.

Some notes about the labs

Each Friday at the Lab:

- Review background material as needed.
- We go through Tasks 1-3 of the week's lab.
- Questions and Office Hours regarding the labs or homeworks.

Feel free to:

- Send questions / requests in advance!
- Let us know if you have requests for something to be reviewed or talked about.

Also:

- Feel free to email me at any time during the week with questions or for any reason.
- olivier.binette@gmail.com
- You can also post on Piazza.

Some notes about the labs

Each Friday at the Lab:

- Review background material as needed.
- We go through Tasks 1-3 of the week's lab.
- Questions and Office Hours regarding the labs or homeworks.

Feel free to:

- Send questions / requests in advance!
- Let us know if you have requests for something to be reviewed or talked about.

Also:

- Feel free to email me at any time during the week with questions or for any reason.
- olivier.binette@gmail.com
- You can also post on Piazza.

Some notes about the labs

Active learning is important:

- Take notes as we go along
- Reflect on the material and question yourself as we go along
- Interrupt to ask questions

We're trying to build a bit more interactivity into the labs

- But we also want to be quick and efficient
- Feedback welcome here!

Some notes about the labs

Active learning is important:

- Take notes as we go along
- Reflect on the material and question yourself as we go along
- Interrupt to ask questions

We're trying to build a bit more interactivity into the labs

- But we also want to be quick and efficient
- Feedback welcome here!

Some notes about the labs

Active learning is important:

- Take notes as we go along
- Reflect on the material and question yourself as we go along
- Interrupt to ask questions

We're trying to build a bit more interactivity into the labs

- But we also want to be quick and efficient
- Feedback welcome here!

Some notes about the labs

Active learning is important:

- Take notes as we go along
- Reflect on the material and question yourself as we go along
- Interrupt to ask questions

We're trying to build a bit more interactivity into the labs

- But we also want to be quick and efficient
- Feedback welcome here!

Some notes about the labs

Active learning is important:

- Take notes as we go along
- Reflect on the material and question yourself as we go along
- Interrupt to ask questions

We're trying to build a bit more interactivity into the labs

- But we also want to be quick and efficient
- Feedback welcome here!

Quick review of the Normal-Gamma distribution

Quick review of Normal-Gamma model

Context

We have data $X_i \sim^{i.i.d.} N(\mu, 1/\lambda)$ and we want a prior on (μ, λ).

Definition (Normal-Gamma)

If

$$\begin{aligned} \mu \mid \lambda &\sim N(m, (c\lambda)^{-1}) \\ \lambda &\sim \text{Gamma}(a, b), \end{aligned}$$

then we say that $(\mu, \lambda) \sim \text{NormalGamma}(m, c, a, b)$.

Quick review of Normal-Gamma model

Context

We have data $X_i \sim^{i.i.d.} N(\mu, 1/\lambda)$ and we want a prior on (μ, λ).

Definition (Normal-Gamma)

If

$$\begin{aligned} \mu \mid \lambda &\sim N(m, (c\lambda)^{-1}) \\ \lambda &\sim \text{Gamma}(a, b), \end{aligned}$$

then we say that $(\mu, \lambda) \sim \text{NormalGamma}(m, c, a, b)$.

Quick review of Normal-Gamma model

Interpretation

- m: Prior mean
- c: Prior "sample size" (for how many data points would you be willing to trade your prior?)
- a/b is the expected *precision parameter* λ of the data X_i.
- $\sqrt{a}/b$ is the standard deviation of λ and represents your uncertainty about λ.

Quick review of Normal-Gamma model

Standard "non-informative" choice

- $m = 0$ (bias your estimate of μ towards 0).
- $c = 1$ (but bias it just a little bit; your prior is worth the same as only 1 data point).
- a/b: this you have to choose and should be on the scale of your data. E.g. if talking about people's heights, you might expect a standard deviation of around 30cm. So put $a/b = (30cm)^{-1}$.
- $\sqrt{a}/b$: You're not really sure what value of a/b you should really have picked (you could have gone for $(20cm)^{-1}$ or $(40cm)^{-1}$). Better to make it big than too small. As $\sqrt{a}/b \to \infty$, your prior becomes *non-informative*.

Quick review of Normal-Gamma model

Standard "non-informative" choice

- $m = 0$ (bias your estimate of μ towards 0).
- $c = 1$ (but bias it just a little bit; your prior is worth the same as only 1 data point).
- a/b: this you have to choose and should be on the scale of your data. E.g. if talking about people's heights, you might expect a standard deviation of around 30cm. So put $a/b = (30cm)^{-1}$.
- $\sqrt{a}/b$: You're not really sure what value of a/b you should really have picked (you could have gone for $(20cm)^{-1}$ or $(40cm)^{-1}$). Better to make it big than too small. As $\sqrt{a}/b \to \infty$, your prior becomes *non-informative*.

Quick review of Normal-Gamma model

Standard “non-informative” choice

- $m = 0$ (bias your estimate of μ towards 0).
- $c = 1$ (but bias it just a little bit; your prior is worth the same as only 1 data point).
- a/b: this you have to choose and should be on the scale of your data. E.g. if talking about people's heights, you might expect a standard deviation of around 30cm. So put $a/b = (30cm)^{-1}$.
- $\sqrt{a}/b$: You're not really sure what value of a/b you should really have picked (you could have gone for $(20cm)^{-1}$ or $(40cm)^{-1}$). Better to make it big than too small. As $\sqrt{a}/b \to \infty$, your prior becomes *non-informative*.

Quick review of Normal-Gamma model

Standard "non-informative" choice

- $m = 0$ (bias your estimate of μ towards 0).
- $c = 1$ (but bias it just a little bit; your prior is worth the same as only 1 data point).
- a/b: this you have to choose and should be on the scale of your data. E.g. if talking about people's heights, you might expect a standard deviation of around 30cm. So put $a/b = (30cm)^{-1}$.
- $\sqrt{a}/b$: You're not really sure what value of a/b you should really have picked (you could have gone for $(20cm)^{-1}$ or $(40cm)^{-1}$). Better to make it big than too small. As $\sqrt{a}/b \to \infty$, your prior becomes *non-informative*.

Quick review of Normal-Gamma model

Posterior distribution

If

$$X_1, X_2, \ldots, X_n \sim^{ind} N(\mu, 1/\lambda),$$
$$(\mu, \lambda) \sim \text{NormalGamma}(m, c, a, b),$$

then

$$(\mu, \lambda) \mid \{X_i\}_{i=1}^n \sim \text{NormalGamma}(M, C, A, B)$$

with

$$M = \frac{cm + n\overline{X}}{c+n},$$
$$C = c + n,$$
$$A = a + n/2,$$
$$B = b + \tfrac{1}{2}\left(cm^2 - CM^2 + \sum X_i^2\right).$$

Quick review of Normal-Gamma model

Posterior distribution

If

$$X_1, X_2, \ldots, X_n \sim^{ind} N(\mu, 1/\lambda),$$
$$(\mu, \lambda) \sim \text{NormalGamma}(m, c, a, b),$$

then

$$(\mu, \lambda) \mid \{X_i\}_{i=1}^{n} \sim \text{NormalGamma}(M, C, A, B)$$

with

$$M = \frac{cm + n\overline{X}}{c+n},$$
$$C = c + n,$$
$$A = a + n/2,$$
$$B = b + \tfrac{1}{2}\left(cm^2 - CM^2 + \sum X_i^2\right).$$

Quick review of Normal-Gamma model

Posterior distribution

If

$$X_1, X_2, \ldots, X_n \sim^{ind} N(\mu, 1/\lambda),$$
$$(\mu, \lambda) \sim \text{NormalGamma}(m, c, a, b),$$

then

$$(\mu, \lambda) \mid \{X_i\}_{i=1}^n \sim \text{NormalGamma}(M, C, A, B)$$

with

$$M = \frac{cm + n\overline{X}}{c + n},$$
$$C = c + n,$$
$$A = a + n/2,$$
$$B = b + \tfrac{1}{2}\left(cm^2 - CM^2 + \sum X_i^2\right).$$

Quick review of Normal-Gamma model

Posterior distribution

If

$$\begin{aligned} X_1, X_2, \ldots, X_n &\stackrel{ind}{\sim} N(\mu, 1/\lambda), \\ (\mu, \lambda) &\sim \mathrm{NormalGamma}(m, c, a, b), \end{aligned}$$

then

$$(\mu, \lambda) \mid \{X_i\}_{i=1}^n \sim \mathrm{NormalGamma}(M, C, A, B)$$

with

$$\begin{aligned} M &= \frac{cm + n\overline{X}}{c + n}, \\ C &= c + n, \\ A &= a + n/2, \\ B &= b + \tfrac{1}{2}\left(cm^2 - CM^2 + \sum X_i^2\right). \end{aligned}$$

Quick review of Normal-Gamma model

Posterior distribution

If

$$X_1, X_2, \ldots, X_n \sim^{ind} N(\mu, 1/\lambda),$$
$$(\mu, \lambda) \sim \text{NormalGamma}(m, c, a, b),$$

then

$$(\mu, \lambda) \mid \{X_i\}_{i=1}^n \sim \text{NormalGamma}(M, C, A, B)$$

with

$$M = \frac{cm + n\overline{X}}{c + n},$$
$$C = c + n,$$
$$A = a + n/2,$$
$$B = b + \tfrac{1}{2}\left(cm^2 - CM^2 + \sum X_i^2\right).$$

Quick review of Normal-Gamma model

Posterior distribution

If

$$X_1, X_2, \ldots, X_n \sim^{ind} N(\mu, 1/\lambda),$$
$$(\mu, \lambda) \sim \text{NormalGamma}(m, c, a, b),$$

then

$$(\mu, \lambda) \mid \{X_i\}_{i=1}^n \sim \text{NormalGamma}(M, C, A, B)$$

with

$$\begin{aligned}
M &= \frac{cm + n\overline{X}}{c + n}, \\
C &= c + n, \\
A &= a + n/2, \\
B &= b + \tfrac{1}{2}\left(cm^2 - CM^2 + \sum X_i^2\right).
\end{aligned}$$

Quick review of Normal-Gamma model

Let's code this up

```r
rNormalGamma <- function(k, m, c, a, b) {
  lambda = rgamma(k, a, b)
  mu = rnorm(k, m, sqrt(1/(c*lambda)))
  return(rbind(mu, lambda))
}

rNormalGamma(3, 0, 1, 1, 1)
```

```
##             [,1]      [,2]      [,3]
## mu     0.8476881 0.2924268 0.4561089
## lambda 1.3114119 1.6198819 4.3739960
```

Quick review of Normal-Gamma model

Let's code this up

```r
normalGammaParams.post <- function(m, c, a, b, X) {
  n = length(X)
  M = (m*c + sum(X))/(c + n)
  C = c + n
  A = a + n/2
  B = b + (c*m^2 - C*M^2 + sum(X^2))/2
  return(list(m=M, c=C, a=A, b=B))
}

X = rnorm(100, 0, 1)
normalGammaParams.post(0, 1, 1, 1, X)
```

```
## $m
## [1] -0.2039534
##
## $c
## [1] 101
##
## $a
```

Quick review of Normal-Gamma model

Let's code this up

```r
rNormalGamma.post <- function(k, m, c, a, b, X) {
  params = append(list(k=k),
                  normalGammaParams.post(m, c, a, b, X))
  do.call(rNormalGamma, params)
}

params.post = rNormalGamma.post(500, 0, 1, 1, 1, X)
rowMeans(params.post)
```

```
##         mu     lambda
## -0.2064241  1.1244954
```

Lab 4

Lab 4

Do teacher's expectations influence student achievement?

- Students had an IQ test at the begining and end of a year; the data is the difference in IQ score.
- 20% of the students were randomly chosen; their teacher was told they were "spurters" (high performers)

```
spurters = c(18, 40, 15, 17, 20, 44, 38)
controls = c(-4, 0, -19, 24, 19, 10, 5, 10,
             29, 13, -9, -8, 20, -1, 12, 21,
             -7, 14, 13, 20, 11, 16, 15, 27,
             23, 36, -33, 34, 13, 11, -19, 21,
             6, 25, 30,22, -28, 15, 26, -1, -2,
             43, 23, 22, 25, 16, 10, 29)
```

Lab 4

Do teacher's expectations influence student achievement?

- Students had an IQ test at the begining and end of a year; the data is the difference in IQ score.
- 20% of the students were randomly chosen; their teacher was told they were "spurters" (high performers)

```
spurters = c(18, 40, 15, 17, 20, 44, 38)
controls = c(-4, 0, -19, 24, 19, 10, 5, 10,
             29, 13, -9, -8, 20, -1, 12, 21,
             -7, 14, 13, 20, 11, 16, 15, 27,
             23, 36, -33, 34, 13, 11, -19, 21,
             6, 25, 30,22, -28, 15, 26, -1, -2,
             43, 23, 22, 25, 16, 10, 29)
```

Lab 4

Do teacher's expectations influence student achievement?

- Students had an IQ test at the begining and end of a year; the data is the difference in IQ score.
- 20% of the students were randomly chosen; their teacher was told they were "spurters" (high performers)

```
spurters = c(18, 40, 15, 17, 20, 44, 38)
controls = c(-4, 0, -19, 24, 19, 10, 5, 10,
             29, 13, -9, -8, 20, -1, 12, 21,
             -7, 14, 13, 20, 11, 16, 15, 27,
             23, 36, -33, 34, 13, 11, -19, 21,
             6, 25, 30,22, -28, 15, 26, -1, -2,
             43, 23, 22, 25, 16, 10, 29)
```

Lab 4

Do teacher's expectations influence student achievement?

- Students had an IQ test at the begining and end of a year; the data is the difference in IQ score.
- 20% of the students were randomly chosen; their teacher was told they were "spurters" (high performers)

```
spurters = c(18, 40, 15, 17, 20, 44, 38)
controls = c(-4, 0, -19, 24, 19, 10, 5, 10,
             29, 13, -9, -8, 20, -1, 12, 21,
             -7, 14, 13, 20, 11, 16, 15, 27,
             23, 36, -33, 34, 13, 11, -19, 21,
             6, 25, 30,22, -28, 15, 26, -1, -2,
             43, 23, 22, 25, 16, 10, 29)
```

Lab 4

Task 1: Plot histograms for the change in IQ score for the two groups. Report your findings.

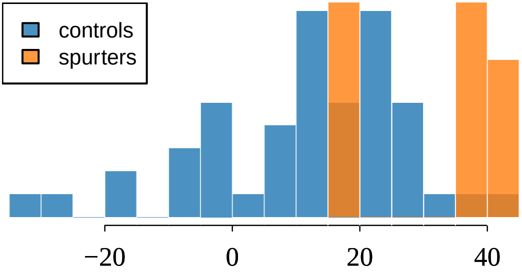
```
source(url("https://gist.githubusercontent.com/OlivierBinette/b7

hist(controls, xlim=range(c(spurters, controls)),
     col=adjustcolor(cmap.seaborn(1), alpha.f=0.8))
hist(spurters, col=adjustcolor(cmap.seaborn(2), alpha.f=0.8),
     add=TRUE)

legend("topleft", legend=c("controls", "spurters"),
       fill=adjustcolor(cmap.seaborn(c(1,2)), alpha.f=0.8),
       cex=0.7)
```

Lab 4

Task 1: Plot histograms for the change in IQ score for the two groups. Report your findings.



Lab 4

Task 2: How strongly does this data support the hypothesis that the teachers expectations caused the spurters to perform better than their classmates?

Let's use a normal model:

$$X_1, \ldots, X_{n_s} \mid \mu_s, \lambda_s^{-1} \overset{iid}{\sim} \text{Normal}(\mu_S, \lambda_S^{-1})$$
$$Y_1, \ldots, Y_{n_C} \mid \mu_c, \lambda_c^{-1} \overset{iid}{\sim} \text{Normal}(\mu_C, \lambda_C^{-1}).$$

We are interested in the difference between the means—in particular, is $\mu_S > \mu_C$?

We can answer this by computing the posterior probability that $\mu_S > \mu_C$:

$$\Pr(\mu_S > \mu_C \mid x_{1:n_S}, y_{1:n_C}).$$

Lab 4

Task 2: How strongly does this data support the hypothesis that the teachers expectations caused the spurters to perform better than their classmates?

Let's use a normal model:

$$X_1, \ldots, X_{n_s} \mid \mu_s, \lambda_s^{-1} \overset{iid}{\sim} \text{Normal}(\mu_S, \lambda_S^{-1})$$
$$Y_1, \ldots, Y_{n_C} \mid \mu_c, \lambda_c^{-1} \overset{iid}{\sim} \text{Normal}(\mu_C, \lambda_C^{-1}).$$

We are interested in the difference between the means—in particular, is $\mu_S > \mu_C$?

We can answer this by computing the posterior probability that $\mu_S > \mu_C$:

$$\Pr(\mu_S > \mu_C \mid x_{1:n_S}, y_{1:n_C}).$$

Lab 4

Task 2: How strongly does this data support the hypothesis that the teachers expectations caused the spurters to perform better than their classmates?

Let's use a normal model:

$$X_1, \dots, X_{n_s} \mid \mu_s, \lambda_s^{-1} \overset{iid}{\sim} \text{Normal}(\mu_S, \lambda_S^{-1})$$
$$Y_1, \dots, Y_{n_C} \mid \mu_c, \lambda_c^{-1} \overset{iid}{\sim} \text{Normal}(\mu_C, \lambda_C^{-1}).$$

We are interested in the difference between the means—in particular, is $\mu_S > \mu_C$?

We can answer this by computing the posterior probability that $\mu_S > \mu_C$:

$$\Pr(\mu_S > \mu_C \mid x_{1:n_S}, y_{1:n_C}).$$

Lab 4

Task 2: How strongly does this data support the hypothesis that the teachers expectations caused the spurters to perform better than their classmates?

Let's use a normal model:

$$X_1, \ldots, X_{n_s} \mid \mu_s, \lambda_s^{-1} \stackrel{iid}{\sim} \text{Normal}(\mu_S, \lambda_S^{-1})$$
$$Y_1, \ldots, Y_{n_C} \mid \mu_c, \lambda_c^{-1} \stackrel{iid}{\sim} \text{Normal}(\mu_C, \lambda_C^{-1}).$$

We are interested in the difference between the means—in particular, is $\mu_S > \mu_C$?

We can answer this by computing the posterior probability that $\mu_S > \mu_C$:

$$\Pr(\mu_S > \mu_C \mid x_{1:n_S}, y_{1:n_C}).$$

Lab 4

Task 2: How strongly does this data support the hypothesis that the teachers expectations caused the spurters to perform better than their classmates?

Let's assume independent Normal-Gamma priors:

$$\text{spurters: } (\mu_S, \lambda_S) \sim \text{NormalGamma}(m, c, a, b)$$
$$\text{controls: } (\mu_C, \lambda_C) \sim \text{NormalGamma}(m, c, a, b)$$

Lab 4

Task 2: How strongly does this data support the hypothesis that the teachers expectations caused the spurters to perform better than their classmates?

Subjective choice:

- $m = 0$ (Don't know whether students will improve or not, on average.)
- $c = 1$ (Unsure about how big the mean change will be—prior certainty in our choice of m assessed to be equivalent to one datapoint.)
- $a/b = 1/25$
- $\sqrt{a}/b = 1$, thus $a = 0.0016$, $b = 0.04$.

```
m = 0
c = 1
a = 0.0016
b = 0.04
```

Lab 4

Task 2: How strongly does this data support the hypothesis that the teachers expectations caused the spurters to perform better than their classmates?

Now let's sample from the posterior distributions.

```
k = 5000
spurters.postParams.s =
  rNormalGamma.post(k, m, c, a, b, spurters)
controls.postParams.s =
  rNormalGamma.post(k, m, c, a, b, controls)
```

Lab 4

Task 2: How strongly does this data support the hypothesis that the teachers expectations caused the spurters to perform better than their classmates?

Using the Monte-Carlo approximation

$$\Pr(\mu_S > \mu_C \mid x_{1:n_S}, y_{1:n_C}) \approx \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\left(\mu_S^{(i)} > \mu_C^{(i)}\right),$$

we find

```
mean(spurters.postParams.s["mu",] >
       controls.postParams.s["mu",])
```

```
## [1] 0.963
```

Lab 4

Task 3: Provide a scatterplot of samples from the posterior distributions for the two groups. What are your conclusions?

```
plot(spurters.postParams.s["mu",],
     spurters.postParams.s["lambda",]^(-1/2),
     col=cmap.seaborn(2), alpha=0.2,
     xlab="mu", ylab="standard deviation",
     ylim=c(0,40), xlim=c(-50, 50))
points(controls.postParams.s["mu",],
      controls.postParams.s["lambda",]^(-1/2),
      col=cmap.seaborn(1), alpha=0.2)

legend("topleft", legend=c("controls", "spurters"),
       col=cmap.seaborn(c(1,2)), cex=0.7, lty=1)
```

Lab 4

Task 3: Provide a scatterplot of samples from the posterior distributions for the two groups. What are your conclusions?