BDA HW1

Ziyue Wang

1/27/2021

```r
set.seed(0)
```

```r
n_batch <- 100
# With high probablity the total number of patients that...
# ...arrive before 4pm will be less than 100.

# If more than 100 we have a loop to continue adding patients as written down below.

arrive_gap <- rexp(n=n_batch, rate=1/10)
meet_time <- runif(n=n_batch, 5, 20)

while (sum(arrive_gap) <= (16-9) * 60) {
    print("Somehow we have an unbelievable number of patients before 4pm")
    arrive_time <- c(arrive_gap, rexp(n=n_batch, rate=1/10))
    meet_time <- c(meet_time, runif(n=n_batch, 5, 20))
}
```

How many patients visited our office?

```r
last_patient <- tail(which(cumsum(arrive_gap) <= (16-9) * 60), n=1)
last_patient
```

```
## [1] 43
```

```r
if (length(last_patient) == 0) {
    print("Nobody comes before 4pm, what happened?")
}
```

Continue Simulating for patient waiting information

```r
arrive_time <- cumsum(arrive_gap)

doctor_schedule <- data.frame(matrix(0, last_patient+1, 3))
colnames(doctor_schedule) <- c('A', 'B', 'C')

wait_time <- c()

for (i in 1:last_patient) {
    hello_doctor <- doctor_schedule[i, ]
    next_avaliable <- which.min(hello_doctor)

    if (hello_doctor[next_avaliable] <= arrive_time[i]) {
```

```r
        ### There is at least one doctor avaliable for the ith patient on arrival. ###

        doctor_schedule[i+1, ] <- doctor_schedule[i, ]
        # Keep the current schedule of other two doctors for next patient.
        doctor_schedule[i+1, next_avaliable] <- arrive_time[i] + meet_time[i]
        # Change the meeting doctor's schedule.

    } else {
        ### No doctor is avaliable at this time. Let's wait. ###

        wait_time <- c(wait_time, hello_doctor[next_avaliable] - arrive_time[i])
        doctor_schedule[i+1, ] <- doctor_schedule[i, ]
        doctor_schedule[i+1, next_avaliable] <-
          doctor_schedule[i, next_avaliable] + meet_time[i]
        # Notice that the way we calculate next avaliable time for this doctor...
        # ...is different comparing to the above chunk.

    }

    ### Poor doctors, they couldn't even take a break. ###
}
```

How many had to wait?

```r
how_many_had_to_wait <- length(wait_time)
how_many_had_to_wait
```

```
## [1] 4
```

What was their average wait?

```r
what_was_their_average_wait <- mean(as.numeric(wait_time))
if (is.na(what_was_their_average_wait)) what_was_their_average_wait <- 0
# in case no one ever waited.
what_was_their_average_wait # in minutes
```

```
## [1] 3.490962
```

When did the office close?

```r
when_did_the_office_close <- format(as.POSIXct((max(doctor_schedule) + 9*60) * 60,
                                        origin = "1970-01-01", tz = "UTC"),
                                        "%H:%M")
if (when_did_the_office_close <= '16:00') when_did_the_office_close <- '16:00'
# In case all patients are treated before 4pm (and...
# ...the next potential patient will arrives after 4pm).

# It depends on how you understand '... closes when the last patient is through with the doctor'.
# The question is how do you know this is the last patient before 4pm. You don't.
# So I think it's reasonable to wait until 4pm to close.
# This definition will certainly change the distribution of 'close time' as a random variable.

when_did_the_office_close
```

```
## [1] "16:09"
```

Repeat 100 times

- Write above chunks of code into a function and repeat the simulation for 100 times.
- Comments are deleted. Please see above chunks for code explaination.

```
Hurtado_Health_Center <- function(n_rep) {
    result <- data.frame(matrix(, n_rep, 4))
    for (n in 1:n_rep) {
        n_batch <- 100
        arrive_gap <- rexp(n=n_batch, rate=1/10)
        meet_time <- runif(n=n_batch, 5, 20)

        while (sum(arrive_gap) <= (16-9) * 60) {
            print("Somehow we have an unbelievable number of patients before 4pm")
            arrive_time <- c(arrive_gap, rexp(n=n_batch, rate=1/10))
            meet_time <- c(meet_time, runif(n=n_batch, 5, 20))
        }

        last_patient <- tail(which(cumsum(arrive_gap) <= (16-9) * 60), n=1)
        result[n, 1] <- last_patient

        if (length(last_patient) == 0) {
            print("Nobody comes before 4pm, what happened?")
        }

        arrive_time <- cumsum(arrive_gap)

        doctor_schedule <- data.frame(matrix(0, last_patient+1, 3))
        colnames(doctor_schedule) <- c('A', 'B', 'C')

        wait_time <- c()

        for (i in 1:last_patient) {
            hello_doctor <- doctor_schedule[i, ]
            next_avaliable <- which.min(hello_doctor)

            if (hello_doctor[next_avaliable] <= arrive_time[i]) {

                doctor_schedule[i+1, ] <- doctor_schedule[i, ]
                doctor_schedule[i+1, next_avaliable] <- arrive_time[i] + meet_time[i]

            } else {

                wait_time <- c(wait_time, hello_doctor[next_avaliable] - arrive_time[i])
                doctor_schedule[i+1, ] <- doctor_schedule[i, ]
                doctor_schedule[i+1, next_avaliable] <-
                  doctor_schedule[i, next_avaliable] + meet_time[i]

            }
        }

        how_many_had_to_wait <- length(wait_time)
```

```
        result[n, 2] <- how_many_had_to_wait

        what_was_their_average_wait <- mean(as.numeric(wait_time))
        if (is.na(what_was_their_average_wait)) what_was_their_average_wait <- 0
        result[n, 3] <- what_was_their_average_wait

        when_did_the_office_close <- max(doctor_schedule)
        if (when_did_the_office_close <= 420) when_did_the_office_close <- 420
        result[n, 4] <- when_did_the_office_close
    }
    return(result)
}
```

Take a glance at simulation results

```
res <- Hurtado_Health_Center(100)
colnames(res) <- c('p_number', 'wait_count', 'wait_time_avg', 'close')
head(res)
```

```
##   p_number wait_count wait_time_avg    close
## 1       42          2      2.049634 430.4233
## 2       47          6      2.563495 434.8551
## 3       49          7      2.035736 426.2373
## 4       37          3      4.354265 426.7133
## 5       31          3      8.758707 420.4448
## 6       32          2      3.406471 420.0000
```

```
tail(res)
```

```
##     p_number wait_count wait_time_avg    close
## 95        46          3      2.796880 426.8117
## 96        52          7      6.390701 432.3017
## 97        40          2      3.625204 435.1705
## 98        48          7      4.815937 437.7422
## 99        44         17      6.033319 420.0000
## 100       40          2      3.087213 422.2469
```

Quantiles and Histogram

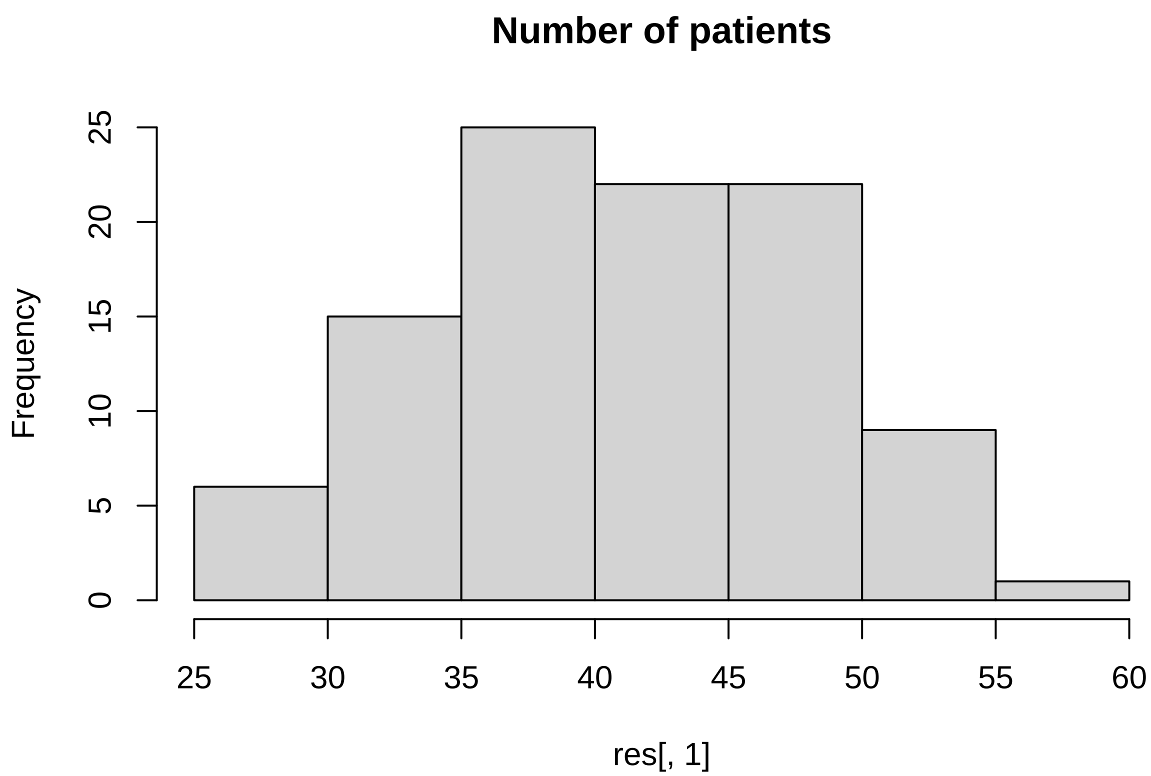
```
print(paste('Meidan for patients # is', quantile(res[,1], probs=.5)))
```

```
## [1] "Meidan for patients # is 41"
```

```
print(paste('50% interval for patients # is', quantile(res[,1], probs=.25),
            'to', quantile(res[,1], probs=.75)))
```

```
## [1] "50% interval for patients # is 37 to 47"
```

```
hist(res[,1], main = 'Number of patients')
```



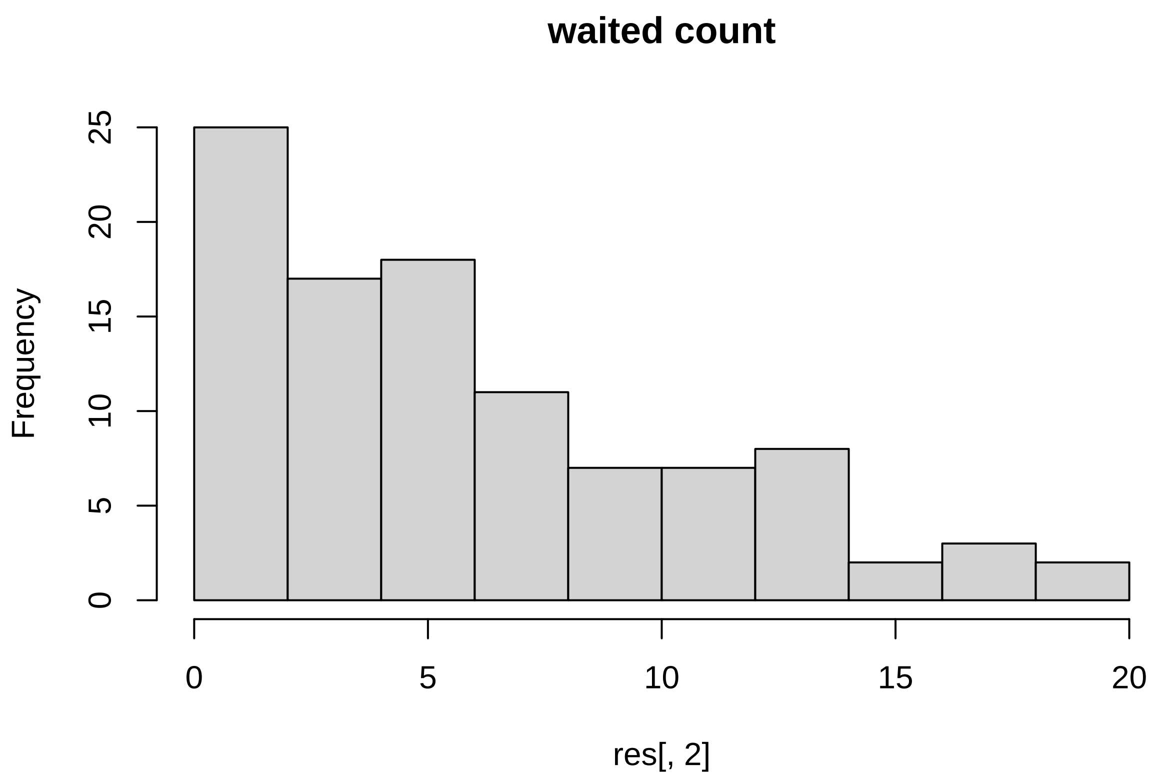

```
print(paste('Meidan for waited count is', quantile(res[,2], probs=.5)))
```

```
## [1] "Meidan for waited count is 5"
```

```
print(paste('50% interval for waited count is', quantile(res[,2], probs=.25),
            'to', quantile(res[,2], probs=.75)))
```

```
## [1] "50% interval for waited count is 2.75 to 10"
```

```
hist(res[,2],  main = 'waited count')
```



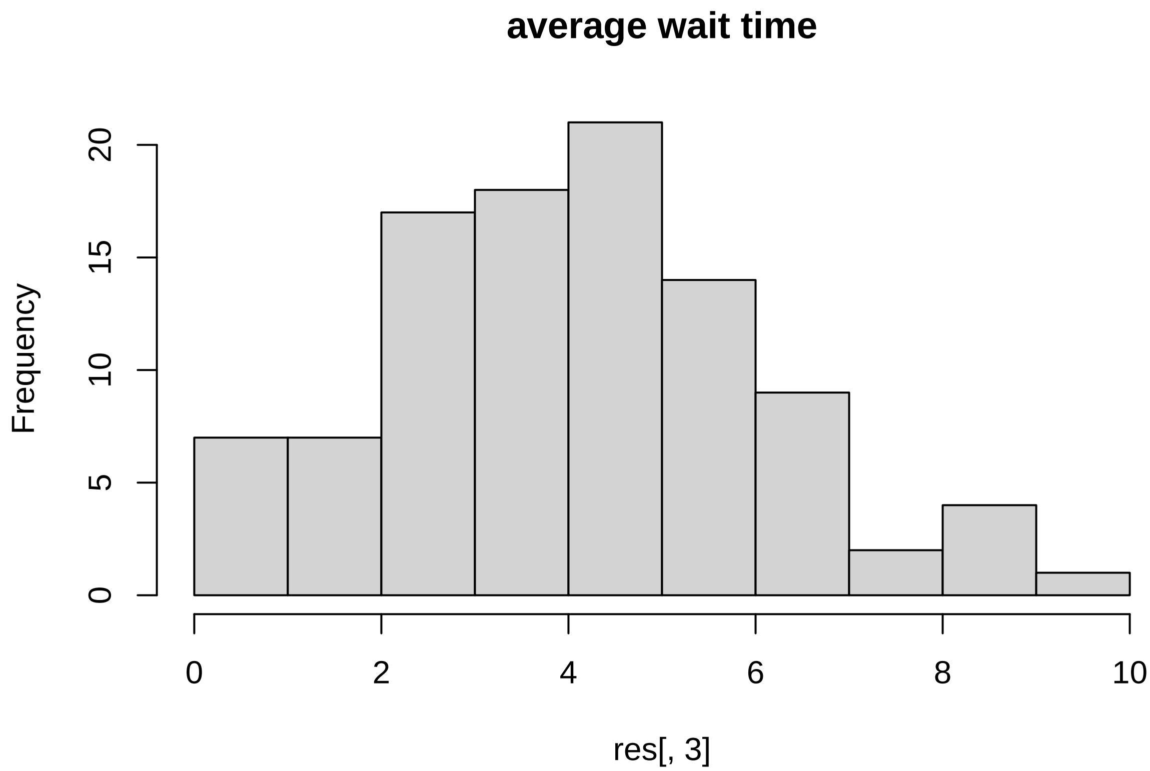

```
print(paste('Meidan for average wait time is', quantile(res[,3], probs=.5)))
```

```
## [1] "Meidan for average wait time is 4.12498239125524"
```

```
print(paste('50% interval for average wait time is', quantile(res[,3], probs=.25),
            'to', quantile(res[,3], probs=.75)))
```

```
## [1] "50% interval for average wait time is 2.56338285421338 to 5.32078694792142"
```

```
hist(res[,3], main = 'average wait time')
```



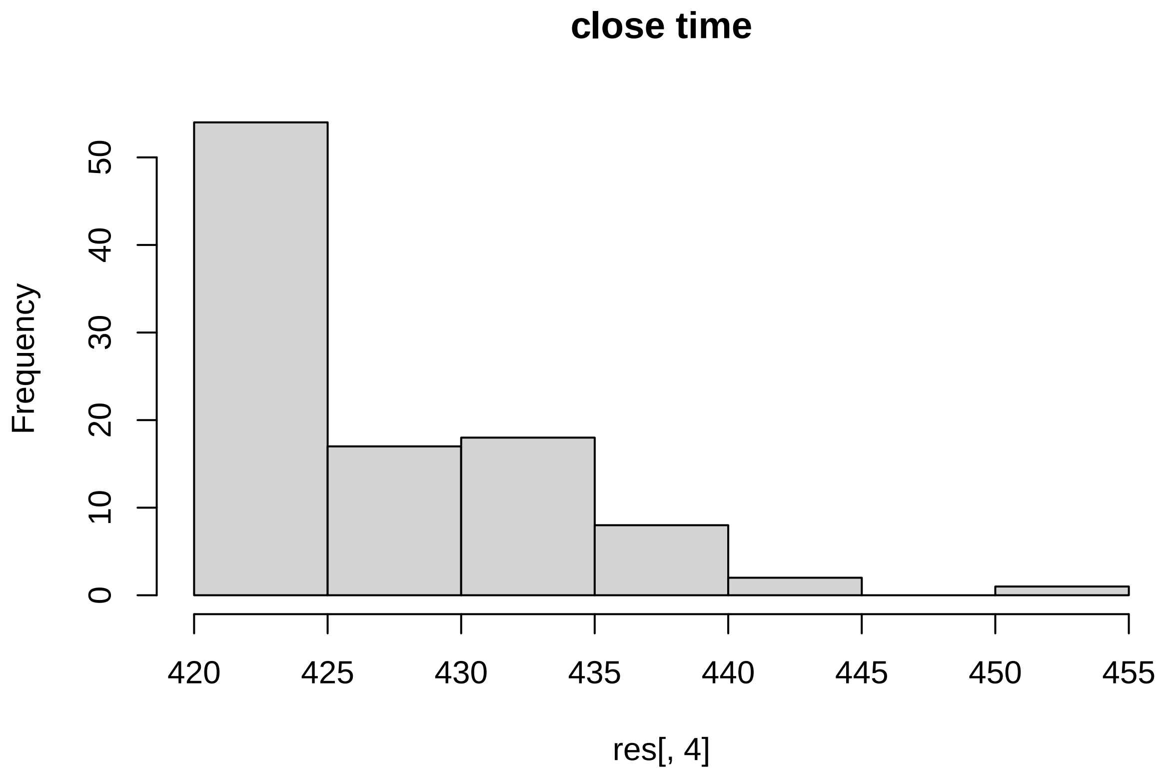

```r
print(paste('Meidan for close time (in minutes after 9am) is', quantile(res[,4], probs=.5)))
```

```
## [1] "Meidan for close time (in minutes after 9am) is 424.198327354349"
```

```r
print(paste('50% interval for close time (in minutes after 9am) is', quantile(res[,4], probs=.25),
            'to', quantile(res[,4], probs=.75)))
```

```
## [1] "50% interval for close time (in minutes after 9am) is 420 to 430.903126228668"
```

```r
hist(res[,4], main = 'close time')
```


close time
Frequency
0
10
20
30
40
50
420
425
430
435
440
445
450
455
res[, 4]

BDA HW1

Ziyue Wang

Jan. 27, 2021

14. (a)

Eq. 2.9 & 2.10: $p(\theta|y)$

$$\propto \exp\left\{-\frac{1}{2}\left[\frac{(y-\theta)^2}{\sigma^2}+\frac{(\theta-\mu_0)^2}{\tau_0^2}\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\frac{(\sigma^2+\tau_0^2)\theta^2-2(y\tau_0^2+\mu_0\sigma^2)\theta+y^2\tau_0^2+\mu_0^2\sigma^2}{\sigma^2\tau_0^2}\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\cdot\frac{\sigma^2+\tau_0^2}{\sigma^2\tau_0^2}\cdot\left[\theta^2-2\,\frac{y\tau_0^4\sigma^2+\mu_0\sigma^4\tau_0^2}{\sigma^2+\tau_0^2}\,\theta+C\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\frac{\sigma^2+\tau_0^2}{\sigma^2\tau_0^2}\left[\left(\theta-\frac{y\tau_0^4\sigma^2+\mu_0\sigma^4\tau_0^2}{\sigma^2+\tau_0^2}\right)^2+C-\left(\frac{y\tau_0^4\sigma^2+\mu_0\sigma^4\tau_0^2}{\sigma^2+\tau_0^2}\right)^2\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\frac{\sigma^2+\tau_0^2}{\sigma^2\tau_0^2}\left(\theta-\frac{y\tau_0^4\sigma^2+\mu_0\sigma^4\tau_0^2}{\sigma^2+\tau_0^2}\right)^2\right\}\cdot\exp\{-C'\}.$$

where C, C' are constant w.r.t. θ,

drop $\exp(-C')$ and define: $\tau_1^2=\dfrac{\sigma^2\tau_0^2}{\sigma^2+\tau_0^2}=\dfrac{1}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}}$, $\mu_1=\dfrac{y\tau_0^4\sigma^2+\mu_0\sigma^4\tau_0^2}{\sigma^2+\tau_0^2}=\dfrac{\frac{1}{\tau_0^2}\mu_0+\frac{1}{\sigma^2}y}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}}$

we conclude that

$$p(\theta|y)\propto\exp\left(-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right).$$

Eq. 2.11 & 2.12:

$$p(\theta|y)\propto\exp\left\{-\frac{1}{2}\left[\frac{\theta^2-2\bar{y}\theta+C}{\sigma^2/n}+\frac{1}{\tau_0^2}(\theta-\mu_0)^2\right]\right\}$$

where $C=\frac{1}{n}\sum y_i^2$ is constant w.r.t. θ.

Thus it reduces to the same format as in Eq. 2.9 & 2.10 with σ^2 replaced by $\frac{n}{\sigma^2}$ and y replaced by $\bar{y}$ and the result follows.

(b) We've show that with prior $N(\mu_0, \tau_0^2)$

$$\theta | y_1 \sim N(\theta | \mu_1, \tau_1^2)$$

(with exchangeability)

where $\mu_1 = \dfrac{\frac{1}{\tau_0^2}\cdot\mu_0 + \frac{1}{\sigma^2} y_1}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}}$ and $\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}$.

Similarly, with prior $N(\mu_1, \tau_1^2)$,

$$\theta | y_2 \sim N(\theta | \mu_2, \tau_2^2)$$

where $\mu_2 = \dfrac{\frac{1}{\tau_1^2}\mu_1 + \frac{1}{\sigma^2} y_2}{\frac{1}{\tau_1^2} + \frac{1}{\sigma^2}}$ and $\frac{1}{\tau_2^2} = \frac{1}{\tau_1^2} + \frac{1}{\sigma^2}$.

or equivalently, $\frac{1}{\tau_2^2} = \frac{1}{\tau_1^2} + \frac{1}{\sigma^2} = \left(\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}\right) + \frac{1}{\sigma^2}$,

$$\mu_2 = \frac{\left(\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}\right)\mu_1 + \frac{1}{\sigma^2} y_2}{\left(\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}\right) + \frac{1}{\sigma^2}}$$

$$= \frac{\left(\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}\right)\cdot \dfrac{\frac{1}{\tau_0^2}\cdot\mu_0 + \frac{1}{\sigma^2} y_1}{\tau_0^2 + \frac{1}{\sigma^2}} + \frac{1}{\sigma^2} y_2}{\frac{1}{\tau_0^2} + \frac{2}{\sigma^2}}$$

$$= \frac{\frac{1}{\tau_0^2}\cdot\mu_0 + \frac{1}{\sigma^2}(y_1 + y_2)}{\frac{1}{\tau_0^2} + \frac{2}{\sigma^2}}$$

By induction we have

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \quad \text{and} \quad \frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}.$$

2.15. $E[Z^m(1-Z)^n]$

$$:= \frac{1}{B(\alpha,\beta)} \cdot \int Z^m (1-Z)^n \cdot Z^{\alpha-1} \cdot (1-Z)^{\beta-1} dZ$$

$$= \frac{B(\alpha+m, \beta+n)}{B(\alpha,\beta)} \cdot \int \frac{1}{B(\alpha+m, \beta+n)} \cdot Z^{\alpha+m-1} (1-Z)^{\beta+n-1} dZ$$

$$= \frac{B(\alpha+m, \beta+n)}{B(\alpha,\beta)} \cdot 1$$

where $B(a,b) := \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

let $m=1$, $n=0$, we have

$$EZ = E[Z^1(1-Z)^0]$$

$$= \frac{B(\alpha+1, \beta)}{B(\alpha,\beta)}$$

$$= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+1)}$$

$$= \frac{\alpha}{\alpha+\beta}$$

let $m=2$, $n=0$. we have

$$EZ^2 = E[Z^2(1-Z)^0]$$

$$= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+2)}$$

$$= \frac{(\alpha+1)\alpha}{(\alpha+\beta)(\alpha+\beta+1)}$$

Thus, $\mathrm{Var}(Z) = EZ^2 - (EZ)^2$

$$= \frac{(\alpha+1)\alpha}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{\alpha^2}{(\alpha+\beta)^2}$$

$$= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \qquad (\alpha>0,\ \beta>0)$$

2.19: (a) (support of y, θ, ϕ omitted)

$$p(\theta|y) \propto p(\theta)\,p(y|\theta)$$

$$\propto \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \theta^{\alpha-1} e^{-\beta\theta} \cdot \theta e^{-\theta y}.$$

$$\propto \theta^{(\alpha+1)-1} e^{-\theta(\beta+y)}$$

Thus $\theta|y \sim \mathrm{Gamma}(\alpha+1, \beta+y)$

(b) Let $\theta \sim \mathrm{Gamma}(\alpha, \beta)$ and $\phi := \frac{1}{\theta}$.

Since $\frac{1}{x}$ is 1-to-1 transformation of x, $(\frac{1}{x})^{-1} = x$,

we have

$$P_\phi(\phi) = \left|\frac{d\theta}{d(1/\theta)}\right| P_\theta(\theta)$$

$$= \left|-\frac{1}{(1/\theta)^2}\right| \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \theta^{\alpha-1} e^{-\beta\theta}$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \theta^{\alpha+1} \cdot e^{-\beta\theta}$$

$$:= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \phi^{-\alpha-1} e^{-\frac{\beta}{\phi}}$$

#.

Also, we can show $\phi | y \sim$ inv-Gamma

$$p(\phi|y) \propto p(\phi)\, p(y|\phi)$$

$$\propto \frac{\beta^{\alpha}}{\Gamma(\alpha)} \phi^{-\alpha-1} e^{-\frac{\beta}{\phi}} \cdot \frac{1}{\phi} \cdot e^{-\frac{y}{\phi}}$$

$$\propto \phi^{-\alpha-2} e^{-\frac{\beta+y}{\phi}}$$

Thus, $\phi | y \sim$ inv-Gamma $(\alpha+1, \beta+y)$.

(c) let $\theta \sim$ Gamma (α, β).

$SD(\theta) = \sqrt{\frac{\alpha}{\beta^2}}$, $E(\theta) = \frac{\alpha}{\beta}$

Thus, coeff. of variation $= \frac{SD(\theta)}{E(\theta)} = \frac{1}{\sqrt{\alpha}} = \frac{1}{2}$

This implies $\alpha = 4$. (β unknown).

Given size n sample $y = (y_1, \dots, y_n)$, the posterior is

$$p(\theta|y) \propto \prod_{i=1}^{n} (\theta e^{-\theta y_i}) \cdot \theta^{\alpha-1} e^{-\beta\theta}$$

$$\propto \theta^{\alpha+n-1} e^{-\theta(\beta + \sum_{i=1}^{n} y_i)}$$

Thus, $\theta | y \sim$ Gamma $(\alpha+n, \beta + \sum_{i=1}^{n} y_i)$

and the coeff. of variation $= \frac{1}{\sqrt{\alpha+n}} = \frac{1}{10}$.

This implies $n = 100 - \alpha = 96$.

(d) Similarly, for $\phi \sim$ inv-Gamma (α, β).

the coeff. of variation $= \frac{SD(\phi)}{E(\phi)} = \sqrt{\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}} \Big/ \frac{\beta}{\alpha-1}$ (for $\alpha > 2$)

$$= \sqrt{\frac{1}{\alpha-2}} = \frac{1}{2}$$

This implies $\alpha = 6$.

The posterior of ϕ is

$$p(\phi|y) \propto \prod_{i=1}^{n} \left(\frac{1}{\phi} e^{-\frac{y_i}{\phi}} \right) \cdot \phi^{-\alpha-1} e^{-\frac{\beta}{\phi}}$$

$$\propto \phi^{-\alpha-n-1} e^{-\frac{\beta+\Sigma y_i}{\phi}}$$

Thus, $\phi|y \sim \text{inv-Gamma}(\alpha+n,\ \beta+\sum_{i=1}^{n} y_i)$

The coeff. of variation $= \sqrt{\frac{1}{\alpha+n-2}} = \frac{1}{10}$

This implies $n = 100+2-\alpha = 96$.

The answer does not change.