

Week 3: Intro to Bayes

24/01/26

Happiness

Question 1

Consider the happiness example from the lecture, with 118 out of 129 women indicating they are happy. We are interested in estimating θ , which is the (true) proportion of women who are happy. Calculate the MLE estimate $\hat{\theta}$ and 95% confidence interval.

```
# theta hat is sample mean
theta_hat <- 118/129

theta_hat
```

```
[1] 0.9147287
```

```
# mean +- 1.96*se
CI_95 <- c(theta_hat - 1.96 * sqrt((theta_hat * (1-theta_hat))/129), theta_hat + 1.96 * sqrt((theta_hat * (1-theta_hat))/129))

CI_95
```

```
[1] 0.8665329 0.9629244
```

Question 2

Assume a Beta(1,1) prior on θ . Calculate the posterior mean for $\hat{\theta}$ and 95% credible interval.

```
# mean of beta(118 + 1 = 119, 129 - 118 + 1 = 12)
post_theta_hat <- 119/(119 + 12)

post_theta_hat
```

```
[1] 0.9083969
```

```
# credible interval using quantiles qbeta
cred_int <- qbeta(c(0.025,0.975), 119,12)

cred_int
```

```
[1] 0.8536434 0.9513891
```

Question 3

Now assume a $\text{Beta}(10,10)$ prior on θ . What is the interpretation of this prior? Are we assuming we know more, less or the same amount of information as the prior used in Question 2?

The interpretation of this prior is that we are assuming our data has 10 successes (or 10 happy women) to 10 failures (or 10 unhappy women). We are now assuming that we know more information than the prior using in Q2 as in Q2 we assumed 1 success to 1 fail.

Jane Austen

We are going to estimate the mean and standard deviation of average words per line in Jane Austen books, using grid approximation. First, load in the data, calculate the average words and look at the histogram.

```
library(tidyverse)
library(tidytext)
library(janeaustenr)
library(extraDistr)

original_books <- austen_books() |>
  group_by(book) |>
  mutate(linenumber = row_number(),
         chapter = cumsum(str_detect(text,
```

```

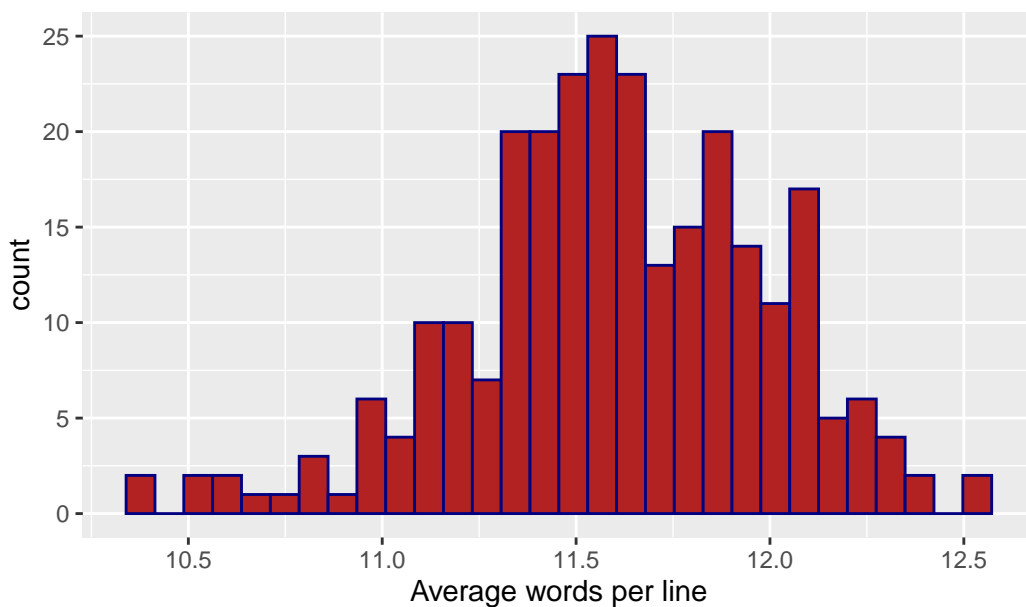
    regex("^chapter [\\divxlc]",
          ignore_case = TRUE)))) |> ungroup()

mean_words <- original_books |>
  filter(chapter>0) |>
  unnest_tokens(word, text) |>
  group_by(book, linenumber, chapter) |>
  tally() |>
  group_by(book, chapter) |>
  summarize(mean_words_per_line = mean(n))

mean_words |>
  ggplot(aes(mean_words_per_line)) +
  geom_histogram(fill = "firebrick", color = "navy")+
  labs(x = "Average words per line", title = "Distribution of average words per line across a

```

Distribution of average words per line across all chapters in all books



We will assume the likelihood of average words y , is Normal

$$y|\mu, \sigma \sim N(\mu, \sigma^2)$$

```
y <- mean_words$mean_words_per_line
```

Let's assume a Normal prior on μ and a half Normal prior on σ :

$$\mu \sim N(\mu_0, \sigma_0^2)$$

$$\sigma \sim N^+(0, 0.1)$$

Now let's set up a grid of size 50 and calculate our prior grid

```
mu0 <- 10
sigma_mu0 <- 5

# how many mu's and how many sigma's?
G<-50 ; H<-50
mean_grid<-seq(9,15,length=G)
sd_grid<-seq(0.1,0.5,length=H)

# priors
unstmean_prior <- dnorm(mean_grid, mu0, sigma_mu0)
mean_prior <- unstmean_prior/sum(unstmean_prior)
unstsd_prior <- dhnorm(sd_grid, sigma = 0.1)
sd_prior <- unstsd_prior/sum(unstsd_prior)
# joint prior
unstprior_grid<-matrix(nrow=G,ncol=H)
for(g in 1:G) {
  for(h in 1:H) {
    unstprior_grid[g,h]<- mean_prior[g]*sd_prior[h]
  }
}
prior_grid<- unstprior_grid/sum(unstprior_grid)
# sum(prior_grid) # check
```

Now calculate the posterior grid:

```
logunstpost_grid<-matrix(nrow=G,ncol=H)
for(g in 1:G) {
  for(h in 1:H) {
    logunstpost_grid[g,h]<- log(prior_grid[g,h])+ sum((dnorm(y,mean_grid[g],sd_grid[h], log = TRUE)))
  }
}
# to avoid weird computer rounding
post_grid <- exp(logunstpost_grid - max(logunstpost_grid))/
  sum(exp(logunstpost_grid - max(logunstpost_grid)))
# sum(post_grid) # check
```

Question 4

Create `mean_post` and `sd_post`, the posterior marginal distributions for μ and σ .

```
mean_post <- rowSums(post_grid)

sd_post <- colSums(post_grid)
```

Question 5

Calculate the posterior mean values for μ and σ .

```
mu_post <- sum(mean_post*mean_grid)

mu_post
```

```
[1] 11.5716
```

```
sigma_post <- sum(sd_post*sd_grid)

sigma_post
```

```
[1] 0.3815707
```

Plots

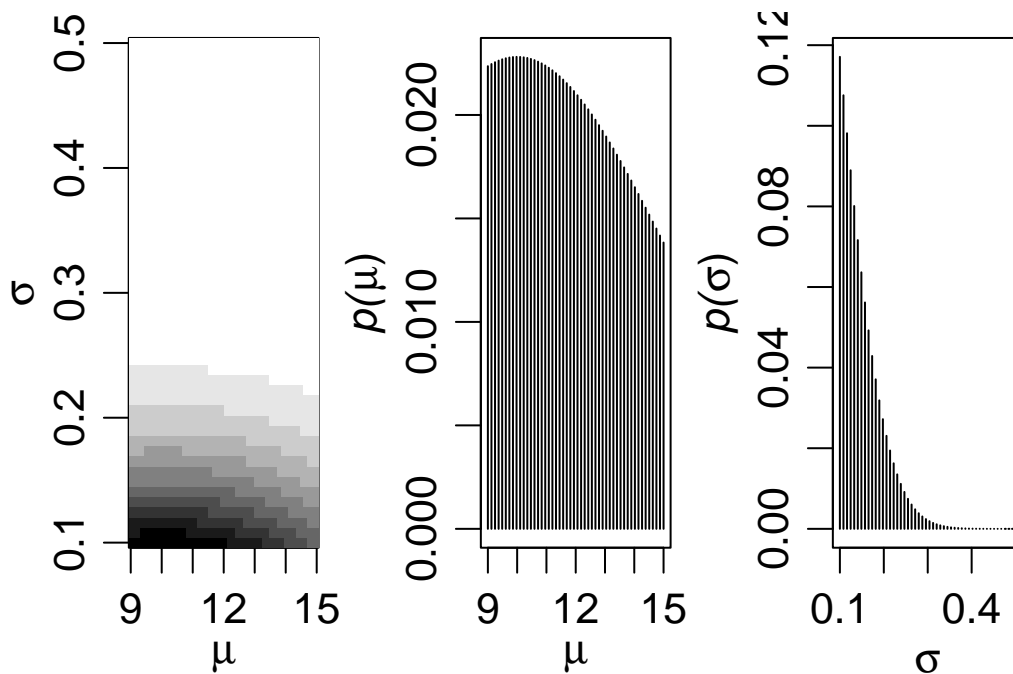
Now let's plot the prior distributions

```
# plots of discrete priors

par(mfrow=c(1,3),mar=c(2.75,2.75,.5,.5),mgp=c(1.70,.70,0), cex = 1.3)
image(mean_grid, sd_grid, prior_grid, col=gray( (10:0)/10 ),
      xlab=expression(mu), ylab=expression(sigma) )

plot(mean_grid,mean_prior,type="h",xlab=expression(mu),
      ylab=expression( paste(italic("p("),
                             mu,")",sep="")), ylim = range(0, mean_prior))

plot(sd_grid,sd_prior, type="h",xlab=expression(sigma),
      ylab=expression( paste(italic("p("),
                             sigma,")",sep="")), ylim = range(0, sd_prior))
```



Question 6

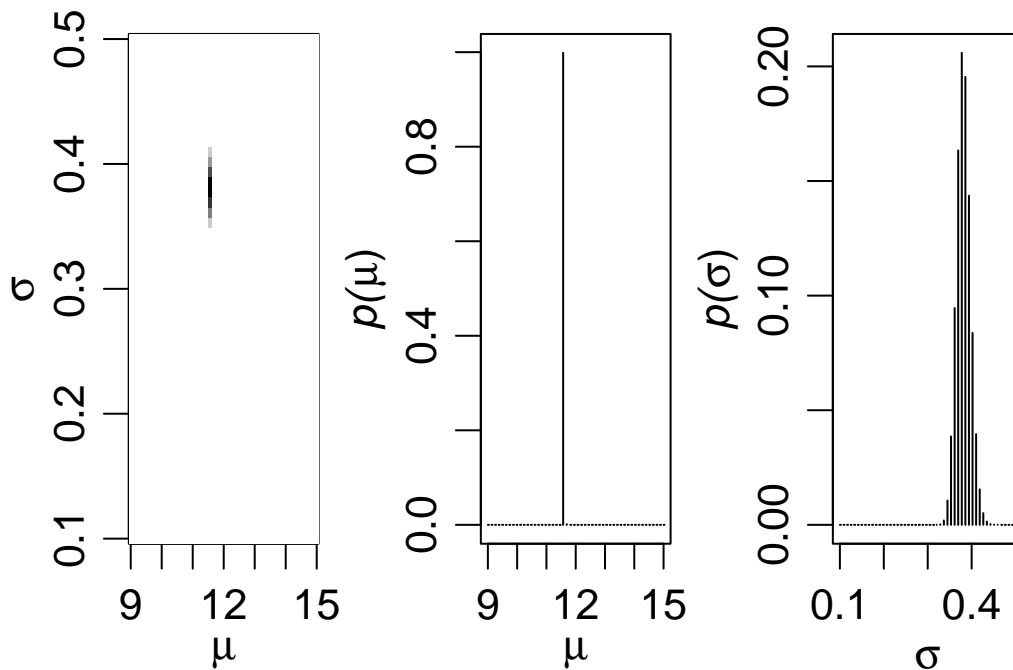
Create a plot as above but for the posterior distributions

```
# plots of discrete posteriors

par(mfrow=c(1,3),mar=c(2.75,2.75,.5,.5),mgp=c(1.70,.70,0), cex = 1.3)
image(mean_grid, sd_grid, post_grid, col=gray( (10:0)/10 ),
      xlab=expression(mu), ylab=expression(sigma) )

plot(mean_grid,mean_post,type="h",xlab=expression(mu),
     ylab=expression( paste(italic("p("),
                           mu,")",sep="")), ylim = range(0, mean_post))

plot(sd_grid,sd_post, type="h",xlab=expression(sigma),
     ylab=expression( paste(italic("p("),
                           sigma,")",sep="")), ylim = range(0, sd_post))
```



Question 7

Now repeat the exercise but with a much more spread out prior on μ . What happens to the posterior mean of μ ? What happens to the posterior sd of μ ?

```
# change mu0 and increase spread

mu0 <- 15
sigma_mu0 <- 10

# how many mu's and how many sigma's?
G<-100 ; H<-100
mean_grid<-seq(5,25,length=G)
sd_grid<-seq(0.1,0.5,length=H)

# priors
unstmean_prior <- dnorm(mean_grid, mu0, sigma_mu0)
mean_prior <- unstmean_prior/sum(unstmean_prior)
unstsd_prior <- dhnorm(sd_grid, sigma = 0.1)
sd_prior <- unstsd_prior/sum(unstsd_prior)
# joint prior
unstprior_grid<-matrix(nrow=G,ncol=H)
```

```

for(g in 1:G) {
  for(h in 1:H) {
    unstprior_grid[g,h]<- mean_prior[g]*sd_prior[h]
  }
}
prior_grid<- unstprior_grid/sum(unstprior_grid)
# sum(prior_grid) # check

logunstpost_grid<-matrix(nrow=G,ncol=H)
for(g in 1:G) {
  for(h in 1:H) {
    logunstpost_grid[g,h]<- log(prior_grid[g,h])+ sum((dnorm(y,mean_grid[g],sd_grid[h], log =
  )
}
}
# to avoid weird computer rounding
post_grid <- exp(logunstpost_grid - max(logunstpost_grid))/
  sum(exp(logunstpost_grid - max(logunstpost_grid)))

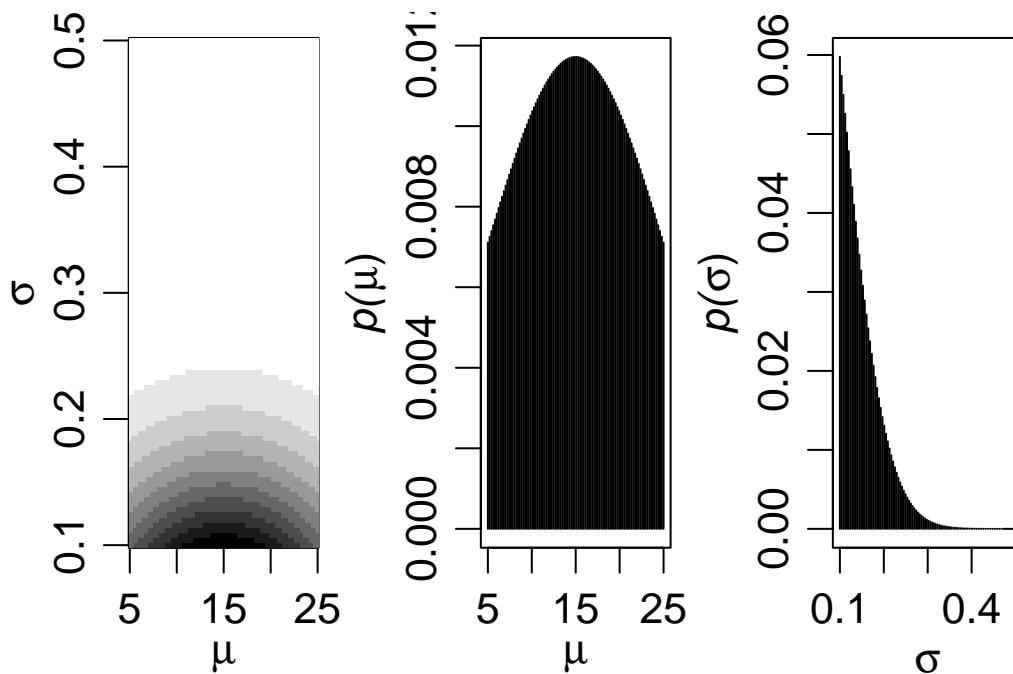
# plots of discrete priors

par(mfrow=c(1,3),mar=c(2.75,2.75,.5,.5),mgp=c(1.70,.70,0), cex = 1.3)
image(mean_grid, sd_grid, prior_grid, col=gray( (10:0)/10 ),
      xlab=expression(mu), ylab=expression(sigma) )

plot(mean_grid,mean_prior,type="h",xlab=expression(mu),
      ylab=expression( paste(italic("p("),
                             mu,")",sep="")), ylim = range(0, mean_prior))

plot(sd_grid,sd_prior, type="h",xlab=expression(sigma),
      ylab=expression( paste(italic("p("),
                             sigma,")",sep="")), ylim = range(0, sd_prior))

```



```
mean_post <- rowSums(post_grid)

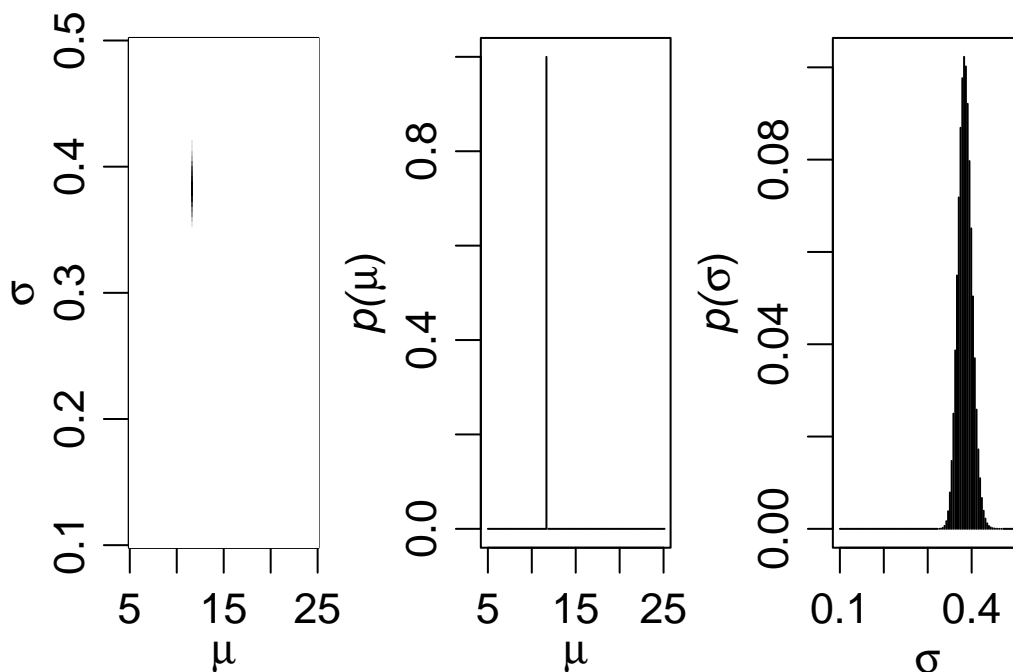
sd_post <- colSums(post_grid)

# plots of discrete posteriors

par(mfrow=c(1,3),mar=c(2.75,2.75,.5,.5),mgp=c(1.70,.70,0), cex = 1.3)
image(mean_grid, sd_grid, post_grid, col=gray( (10:0)/10 ),
      xlab=expression(mu), ylab=expression(sigma) )

plot(mean_grid,mean_post,type="h",xlab=expression(mu),
      ylab=expression( paste(italic("p("),
                             mu,")",sep="")), ylim = range(0, mean_post))

plot(sd_grid,sd_post, type="h",xlab=expression(sigma),
      ylab=expression( paste(italic("p("),
                             sigma,")",sep="")), ylim = range(0, sd_post))
```



```
mu_post <- sum(mean_post*mean_grid)
```

```
mu_post
```

```
[1] 11.66667
```

```
sigma_post <- sum(sd_post*sd_grid)
```

```
sigma_post
```

```
[1] 0.3851232
```

We see the posterior mean and sd of μ increase slightly, but not by much (mean: 11.66667 vs 11.5716, sd: 0.3851 vs 0.3816). This is because we have a large amount of data and therefore the data carries most of the weight of the estimate. Thus, despite changing the prior distribution on μ , the posterior distribution does not change much.