MSMB-Chapter4-Mixture Models

Aleeza Gerstein

2019-10-20

Mixture Models

Finite mixtures

```
coinflips = (runif(10000) > 0.5)
table(coinflips)
```

```
## coinflips
## FALSE  TRUE
##  4937  5063
```

```
oneFlip = function(fl, mean1 = 1, mean2 = 3, sd1 = 0.5, sd2 = 0.5) {
  if (fl) {
    rnorm(1, mean1, sd1)
  } else {
    rnorm(1, mean2, sd2)
  }
}
fairmix = vapply(coinflips, oneFlip, numeric(1))
ggplot(tibble(value = fairmix), aes(x = value)) +
  geom_histogram(fill = "purple", binwidth = 0.1)
```

Question 4.1 How can you use R's vectorized syntax to remove the 'vapply' loop and generate the 'fairmix' vector more efficiently? Question 4.2 Using your improved code, perform one million flips and make a histogram with 500 bins

```
sd <- 0.25
coinflips_mil = (runif(1000000) > 0.5)
mean_mil <- ifelse(coinflips_mil, 1, 3)
values <- rnorm(length(coinflips_mil), mean_mil, 0.25)
ggplot(tibble(value = values), aes(x = value)) +
  geom_histogram(fill = "purple", bins = 500)
```

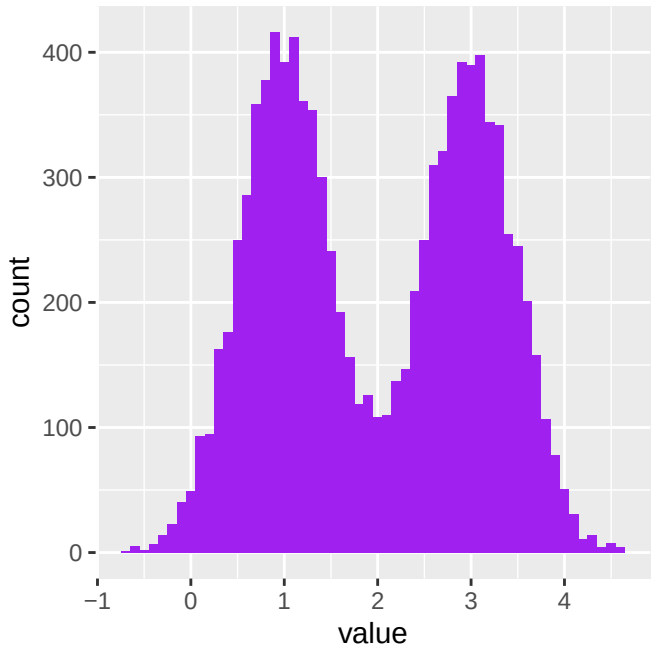
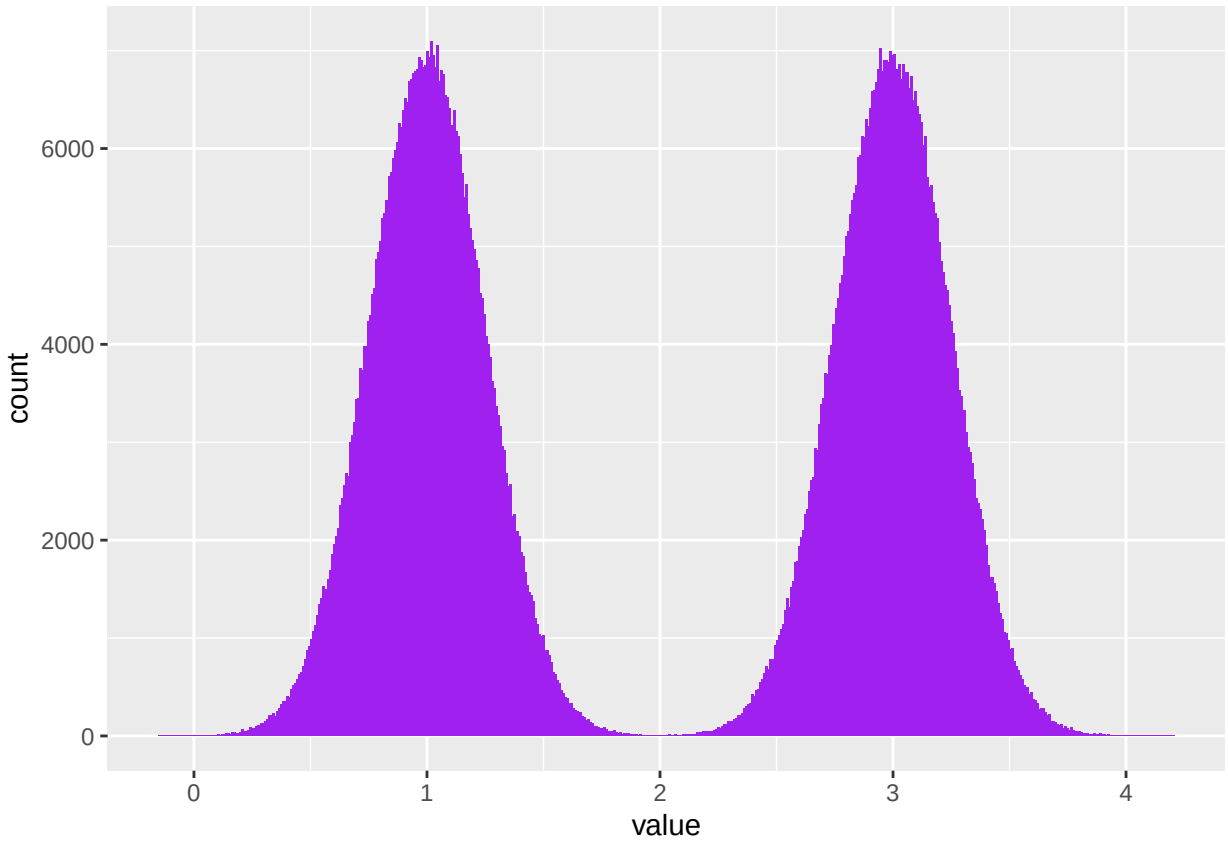
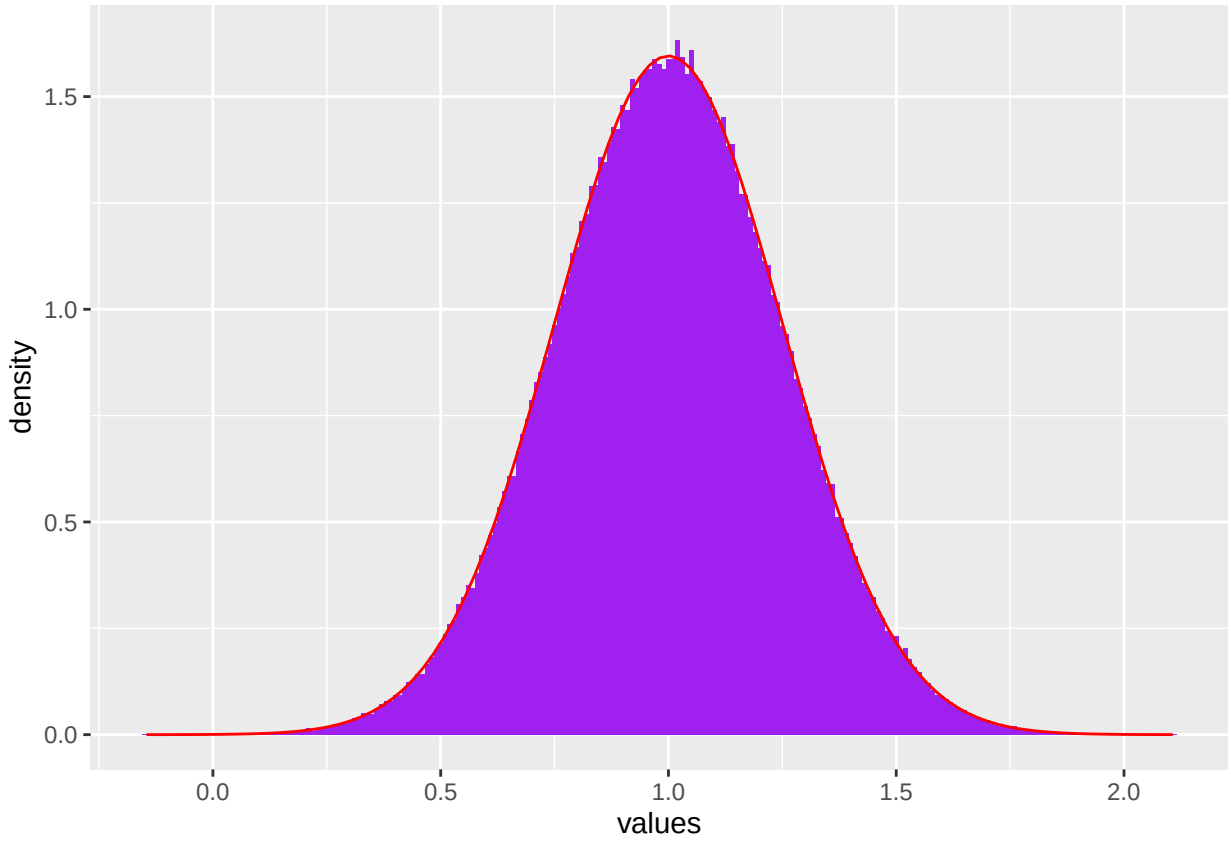


Figure 1: Histogram of 10,000 random draws from a fair mixture of two normals. The left hand part of the histogram is dominated by numbers generated from (A), on the right from (B).



Question 4.3 Plot a histogram only of those values for which coinflips is 'TRUE'. Hint: use 'y = ..density ..' in the call to 'aes' to specify that the vertical axis shows the proportion of counts and set the binwidth to 0.01. (b) Overlay the line corresponding to $\phi(z)$

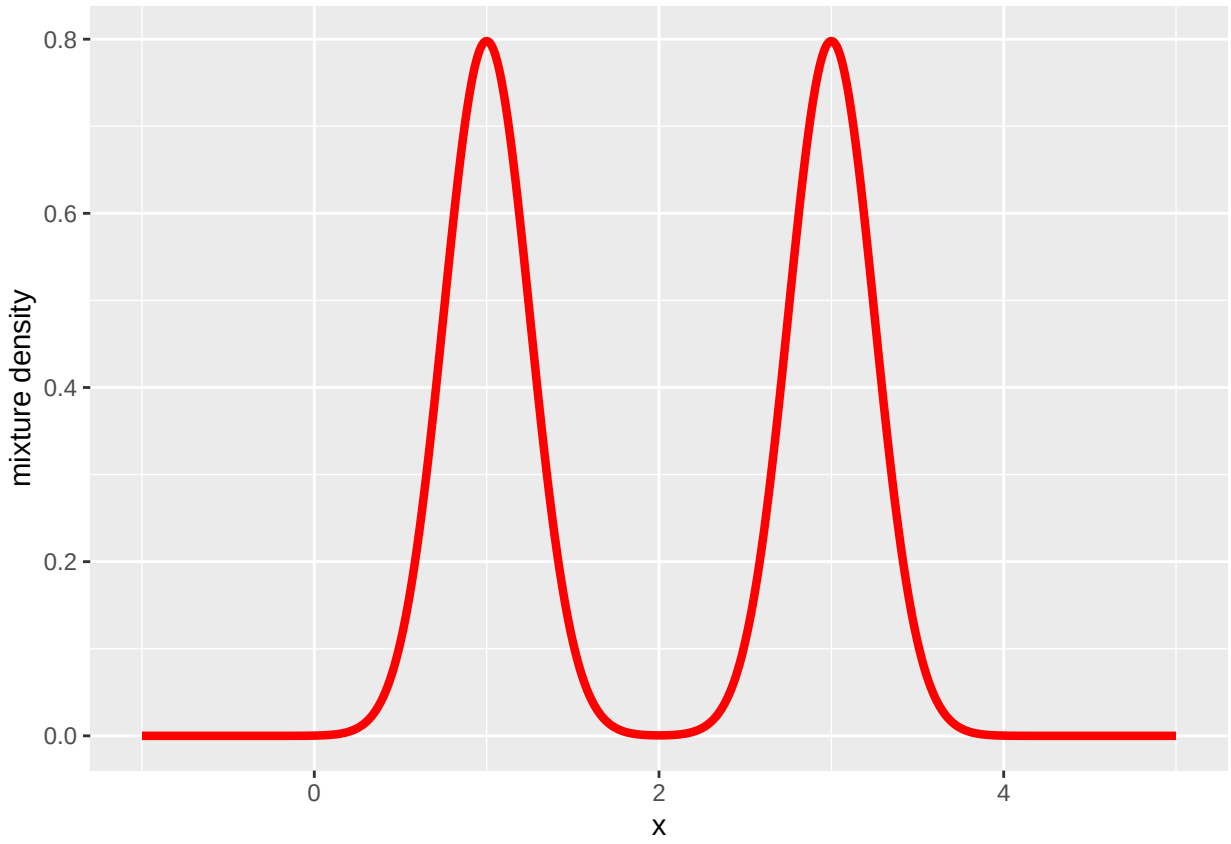
```
fair <- tibble(coinflips_mil, values)
ggplot(filter(fair, coinflips_mil), aes(values)) +
  geom_histogram(aes(y = ..density..), fill = "purple", binwidth = 0.01) +
  stat_function(fun = dnorm, args = list(mean = 1, sd = 0.25), color="red")
```



You can write the formula for the density of all values as a sum of the two densities.

```
means = c(1, 3)
sds   = c(0.25, 0.25)

fairtheory = tibble(
  x = seq(-1, 5, length.out = 1000),
  f = 0.5 * dnorm(x, mean = means[1], sd = sds[1]) +
    0.5 * dnorm(x, mean = means[2], sd = sds[2]))
ggplot(fairtheory, aes(x = x, y = f)) +
  geom_line(color = "red", size = 1.5) + ylab("mixture density")
```


0.8
0.6
0.4
0.2
0.0
mixture density
0
2
4
x