Problem Set 4

Patrick Altmeyer

08 March, 2021

Problem 13

Problem 14

Problem 15

The function below generates data with the desired distribution:

```
sim_data <- function(n = 100, d = 5, a = 0.5) {
    y <- 2 * (rbinom(n, 1, 0.5) - 0.5)  # generate binary outcome
    X <- matrix(rnorm(n * d), n) + matrix(kronecker(y,
        matrix(c(a, rep(0, d - 1)))), n, byrow = T)
    return(list(X = X, y = y))
}
```

Figure 1 provides a quick sanity check: it plots the class-conditional densities of $\mathbf{X}$ where $d = 5$, $n = 10^5$ and $a = 0.5$. The data looks normally distributed and the vectors of class conditional empirical means rounded to the nearest decimal are

$$\bar{\mathbf{X}}_{\mathbf{y}=1} = \begin{bmatrix} 0.5 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \bar{\mathbf{X}}_{\mathbf{y}=-1} = \begin{bmatrix} -0.5 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

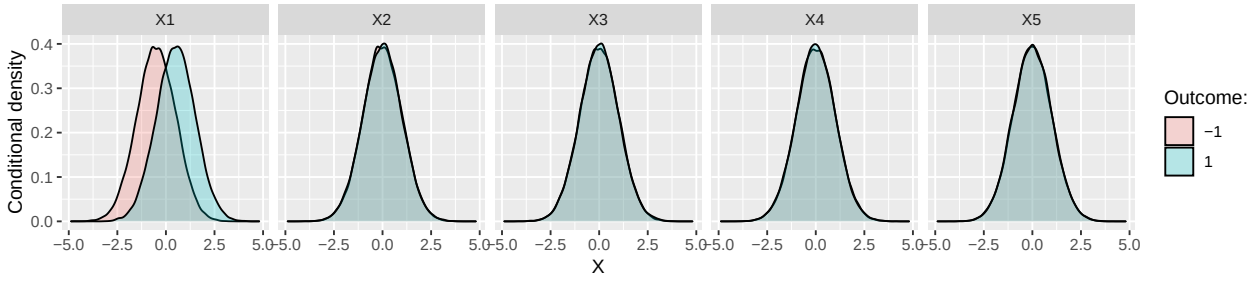

Figure 1: Class-conditional empirical densities.