

# KDE Stuff

## Average of KDEs

Is the average of several KDEs the same as the KDE of the whole data set? Suppose we have  $\mathcal{X}_1 = X_{11}, \dots, X_{1N}$  and  $\mathcal{X}_2 = X_{21}, \dots, X_{2N}$ . The individual estimates of the densities for  $\mathcal{X}_1$  and  $\mathcal{X}_2$  are

$$\hat{f}^{(1)}(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - X_{1j}}{h}\right)$$

and

$$\hat{f}^{(2)}(x) = \frac{1}{Nh} \sum_{j=1}^N K\left(\frac{x - X_{2j}}{h}\right),$$

respectively. The KDE of  $\mathcal{X}_1$  and  $\mathcal{X}_2$  together would be

$$\hat{f}(x) = \frac{1}{2Nh} \sum_{i=1}^2 \sum_{j=1}^N K\left(\frac{x - X_{ij}}{h}\right),$$

If we take the average of the two estimators, then we have

$$\begin{aligned} \hat{f}^{(1,2)}(x) &= \frac{1}{2}(\hat{f}^{(1)}(x) + \hat{f}^{(2)}(x)) \\ &= \frac{1}{2} \left[ \frac{1}{Nh} \left( \sum_{j=1}^N K\left(\frac{x - X_{1j}}{h}\right) + \sum_{j=1}^N K\left(\frac{x - X_{2j}}{h}\right) \right) \right] \\ &= \hat{f}(x). \end{aligned} \tag{1}$$

Am I missing something? Let's check an example.

```

set.seed(9283)
df <- dplyr::tibble(u = runif(100),
                    xc1 = rnorm(100, m = 0, s = 1),
                    xc2 = rnorm(100, m = 3, s = 1),
                    x1 = ifelse(u <= .3, xc1, xc2),
                    x2 = rnorm(100, m = 0, s = 1))
d1 <- density(df$x1, bw = .5, from = -3.5, to = 6.5)
d2 <- density(df$x2, bw = .5, from = -3.5, to = 6.5)
d3 <- density(c(df$x1, df$x2), bw = .5, from = -3.5, to = 6.5)
df_dens <- dplyr::tibble(x = c(d1$x, d2$x, d3$x, d1$x),
                        y = c(d1$y, d2$y, d3$y, (d2$y + d1$y)/2),
                        density = rep(c("Mixture", "N(3, 1)", "Combined", "Average"),
                                     each = 512))

```

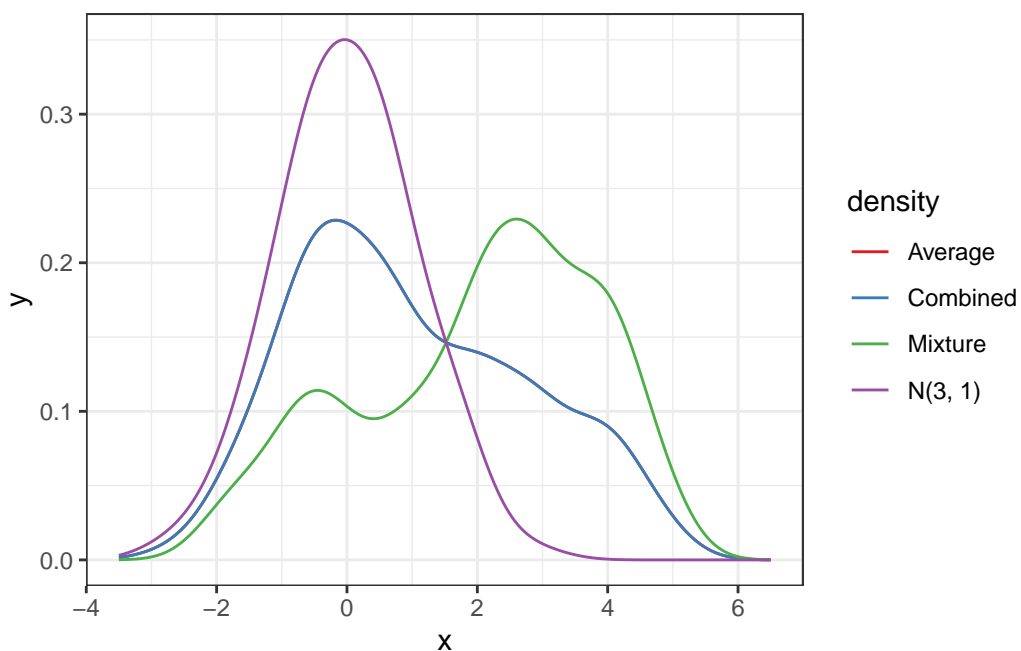


Figure 1: Densities of individual points.

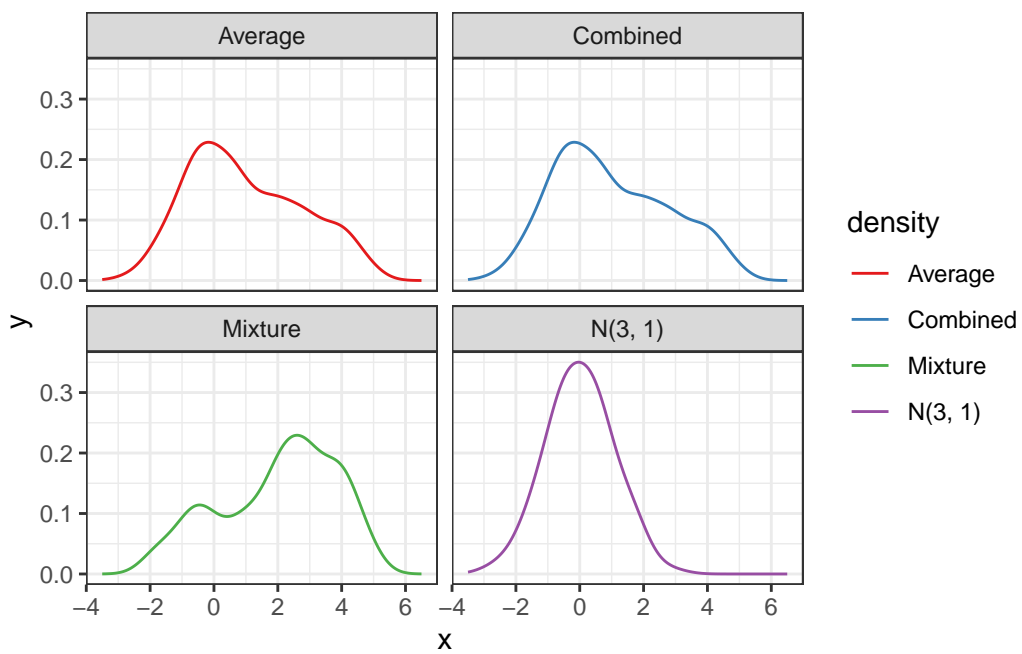


Figure 2: Densities of individual points (facetted).