Minimal Empirical Density Estimation

Matthew Leonawicz

December 24, 2014

1 R code

1.1 Setup

First simulate some data and perform some rudimentary density estimation.

```
set.seed(47)
x <- c(rnorm(1000, -3, 1), rnorm(500, -1, 1), rpois(500, 10))  # Simulate multimodal distribution

n <- 20  # default for density() is n=512
den.smooth <- density(x, adjust = 1.5, n = n)  # I tend to smooth it a bit
den <- density(x, adjust = 1, n = n)  # But I store one without additional smoothing
```

1.2 Graph one

```
# win.graph(10,5) layout(matrix(1:2,1,2))

# hist(x, freq=F) lines(den.smooth, lwd=2) # my preferred smoothed density
# estimate based on x

hist(x, freq = F)
for (i in 1:1000) {
    # reproducing a sample from distribution of x based on den which I carry
    # through my code
    sample.boot <- sample(den$x, size = 1000, prob = den$y, rep = T)
    lines(density(sample.boot, adjust = 1), lwd = 1, col = "#FF000001")  # No extra smoothing with small
    # print(i)
}
# A larger bootstrap sample will pin down the distribution accurately enough
# if necessary
sample.boot <- sample(den$x, size = 10000, prob = den$y, rep = T)
lines(density(sample.boot, adjust = 1.5), lwd = 2, col = "#FF0000")  # smoothing affordable
```

1.3 Graph two

Almost the same as above.

```
# As before but adding an approx() step
hist(x, freq = F)
for (i in 1:1000) {
```

```
    ap <- approx(den$x, den$y, n = 1000)  # reintroduce interpolation before sampling
    sample.boot2 <- sample(ap$x, size = 1000, prob = ap$y, rep = T)
    lines(density(sample.boot2, adjust = 1), lwd = 1, col = "#0000FF01")
    # print(i)
}
sample.boot2 <- sample(ap$x, size = 10000, prob = ap$y, rep = T)
lines(density(sample.boot2, adjust = 1), lwd = 2, col = "#0000FF")
```