STAT 824 hw 01
DKW inequality, KS test, Brownian bridge, kernel density estimation, Hölder smoothness, higher order kernels, CV for KDE bandwidth selection

1. Hoeffding's Inequality states that for any independent zero-mean random variables $Y_1, \ldots, Y_n$ such that $Y_i \in [a_i, b_i]$ for $i = 1, \ldots, n$, for each $\varepsilon > 0$ we have
$$P(\textstyle\sum_{i=1}^n Y_i > \varepsilon) \le e^{-t\varepsilon} \prod_{i=1}^n e^{t^2(b_i - a_i)/8} \quad \text{for all } t \in \mathbb{R}.$$

   (a) Use Hoeffding's inequality to show that for $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Bernoulli}(p)$, for every $\varepsilon > 0$, we have
$$P(|\bar{X}_n - p| > \varepsilon) \le 2e^{-2n\varepsilon^2}.$$

   > We can write
   > $$P(\bar{X}_n - p > \varepsilon) = P(n^{-1} \sum_{i=1}^n (X_i - p) > \varepsilon)$$
   > $$= P(\sum_{i=1}^n (X_i - p) > n\varepsilon), \quad \text{where } X_i - p \in [-p, 1-p]$$
   > $$\le e^{-tn\varepsilon} \prod_{i=1}^n e^{t^2(1-p-(-p))/8}$$
   > $$= e^{-n\varepsilon t + nt^2/8}$$
   >
   > for any $t \in \mathbb{R}$. The value of t which minimizes the right hand side (gives the sharpest bound) is $t = 4\varepsilon$. Plugging this value in gives
   > $$P(\bar{X}_n - p > \varepsilon) \le e^{-2n\varepsilon^2}.$$
   > Likewise $P(\bar{X}_n - p < -\varepsilon) = P(-(\bar{X}_n - p) > \varepsilon) \le e^{-2n\varepsilon^2}$, so
   > $$P(|\bar{X}_n - p| > \varepsilon) = P(\bar{X}_n - p < -\varepsilon) + P(\bar{X}_n - p > \varepsilon) \le 2e^{-2n\varepsilon^2}.$$

   (b) Prove the DKW inequality using the result in part (a).

   > Let $Y_1, \ldots, Y_n$ be iid with cdf F_Y. For any $y \in \mathbb{R}$ let $p = F_Y(y)$ and $X_i = \mathbf{1}(Y_i < y)$. Then $X_i \sim \text{Bernoulli}(p)$ and $\hat{F}_n(y) - F_Y(y) = \bar{X}_n - p$. Therefore $P(|\hat{F}_n(y) - F_Y(y)| > \varepsilon) \le 2e^{-2n\varepsilon^2}$. Since this holds for all $y \in \mathbb{R}$ the DWK inequality is proven.

2. Consider the random variables X and Y, where
$$Y \sim \text{Normal}(0, 1)$$
$$X|\delta \sim \text{Normal}(a \cdot \delta, 1 - a^2), \text{ where } \delta \in \{-1, 1\}, \text{ with } P(\delta = 1) = 1/2,$$

for some $a > 0$. Suppose random samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ are drawn.

(a) Give $\mathbb{E}X$.

We have $\mathbb{E}X = \mathbb{E}(\mathbb{E}[X|\delta]) = a\mathbb{E}\delta = 0$.

(b) Give $\operatorname{Var} X$.

We get $\operatorname{Var} X = \operatorname{Var}(\mathbb{E}[X|\delta]) + \mathbb{E}(\operatorname{Var}[X|\delta]) = \operatorname{Var}(a\delta) + \mathbb{E}(1 - a^2) = 1$, since $\operatorname{Var} \delta = 1$.

(c) Give $\mathbb{E}Y^3$ and $\mathbb{E}X^3$.

The distributions are symmetric around 0 so they will have third moment equal to zero.

(d) Fix $n = 60$ and $m = 80$ and, for each $a \in \{1 - (1/2)^j : j = 0, 1, \ldots, 8\}$, generate 100 random samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ (the densities are pictured below) and report for each value of a the proportion of times the two-sample Kolmogorov-Smirnov test rejects the null hypotheses of equal cdfs (make a table). Do NOT use the `ks.test` function; write your own code and turn it in along with the table.
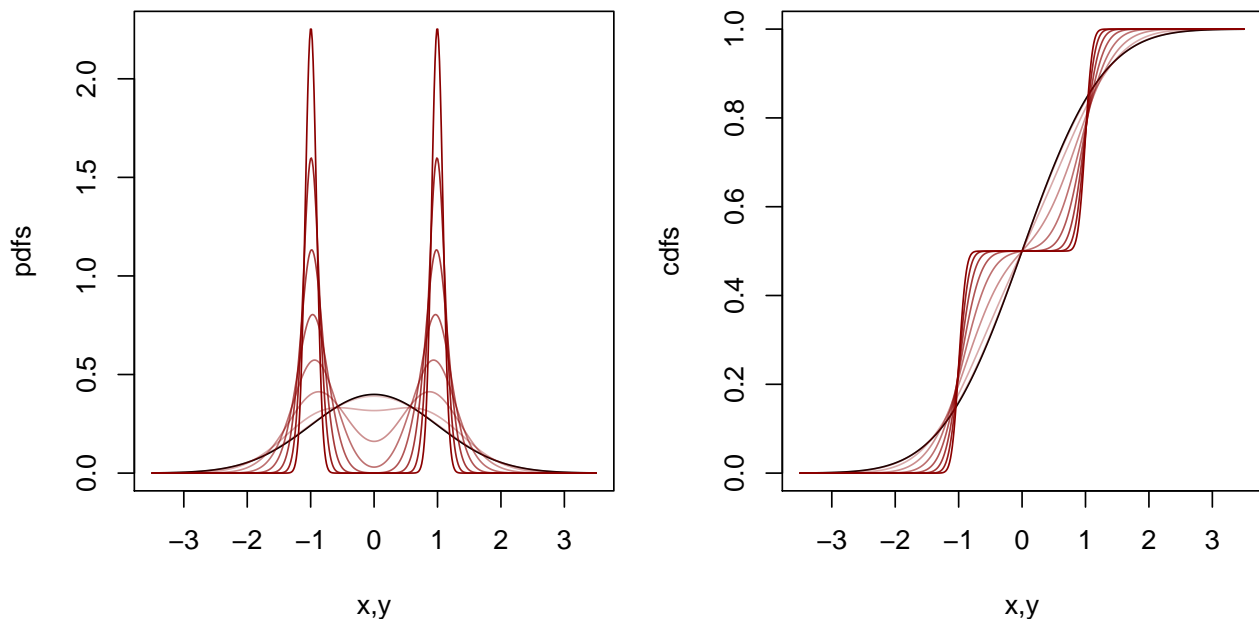
I got the table

j	0	1	2	3	4	5	6	7	8
power	0.06	0.07	0.08	0.19	0.51	0.79	0.98	1.00	1.00

(e) Generate a large number of Brownian bridges in order to get approximations to the 0.70, 0.80, 0.90, 0.95, and 0.99 quantiles of the distribution with cdf $\mathrm{KS}(x) = 1 - 2\sum_{i=1}^{\infty}(-1)^{i+1}e^{-2i^2x^2}$. Turn in a table of these values.

The code

```
N <- 5000
S <- 2000
KS <- numeric(S)
for( s in 1:S){

  B <- cumsum(c(0,rnorm(N,0,sqrt(1/N))))
  t <- c(0:N)/N
  B0 <- B - t * B[N+1]
  KS[s] <- max(abs(B0))

}

quantile(KS,c(.7,.8,.9,.95,.99))
```

gave me the quantiles

0.70	0.80	0.90	0.95	0.99
0.95	1.06	1.21	1.36	1.61

3. Given a random sample $X_1, \ldots, X_n$, find $\int_{\mathbb{R}} x \hat{f}_n(x) dx$ and $\int_{\mathbb{R}} x^2 \hat{f}_n(x) dx$ when

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right),$$

where $h > 0$ and

(a) K the standard Normal density.

> We get $\int_{\mathbb{R}} x \hat{f}_n(x) dx = \bar{X}_n$ and $\int_{\mathbb{R}} x^2 \hat{f}_n(x) dx = n^{-1} \sum_{i=1}^{n} X_i^2 + h^2$.

(b) K is a kernel of order 2.

> We get $\int_{\mathbb{R}} x \hat{f}_n(x) dx = \bar{X}_n$ and $\int_{\mathbb{R}} x^2 \hat{f}_n(x) dx = n^{-1} \sum_{i=1}^{n} X_i^2$.

(c) What is the effect of these different kernels on $\int_{\mathbb{R}} x^2 \hat{f}_n(x) dx - (\int_{\mathbb{R}} x \hat{f}_n(x) dx)^2$?

> The Gaussian kernel is a kernel of order 1, and causes the variance according to $\hat{f}_n$ to increase; the kernel of order 2 preserves the variance in the sense that the variance accoring to $\hat{f}_n$ is equal to the empirical variance of $X_1, \ldots, X_n$.

4. Consider the function

$$N(x) = \begin{cases} \frac{32}{3}u^3, & 0 \le u < 1/4 \\ 32u^2 - 32u^3 - 8u + 2/3, & 1/4 \le u < 2/4 \\ 32(1-u)^2 - 32(1-u)^3 - 8(1-u) + 2/3, & 2/4 \le u < 3/4 \\ \frac{32}{3}(1-u)^3, & 3/4 \le u < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Identify β and L such that $N \in \mathcal{H}(\beta, L)$.

We find that N has two continuous derivatives, and the second derivative satisfies a Lipschitz condition with $L = 192$. So $N \in \mathcal{H}(\beta = 3, L = 192)$. For a visual, we have
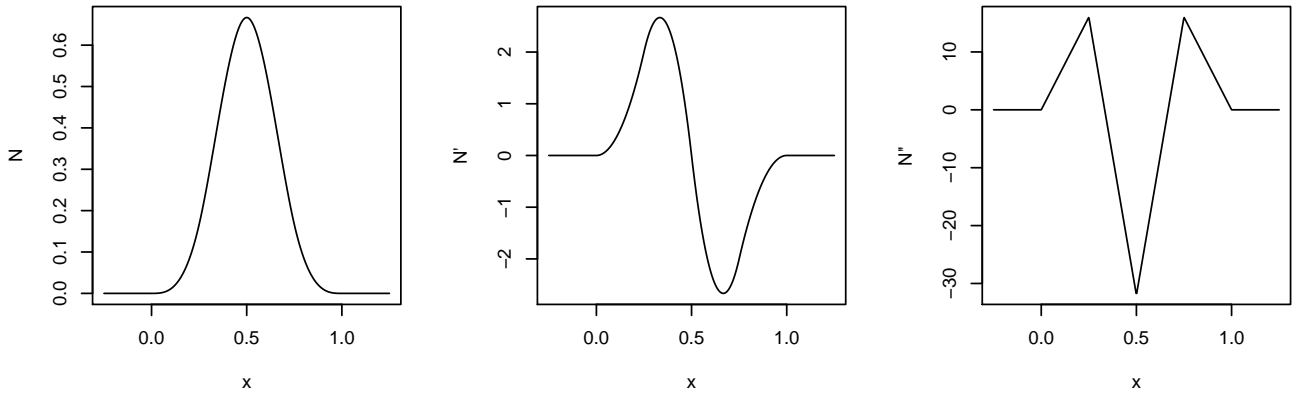
5. Let $\{\varphi_m(\cdot)\}_{m=0}^{\infty}$ represent polynomials defined by

$$\varphi_0(u) = \frac{1}{\sqrt{2}}, \quad \varphi_m(u) = \sqrt{\frac{2m+1}{2}} \frac{1}{2^m m!} \frac{d^m}{du^m} \left[(u^2 - 1)^m \right], \quad m = 1, 2, \dots$$

for $u \in [-1, 1]$. These are known as the Legendre polynomials on $[-1, 1]$; they are *orthonormal with respect to the Lebesgue measure*, which means they have the property

$$\int_{-1}^{1} \varphi_m(u)\varphi_k(u)du = \begin{cases} 1, & m = k \\ 0, & m \ne k. \end{cases}$$

Proposition 1.3 of [1] gives that the function $K : \mathbb{R} \to \mathbb{R}$ given by

$$K(u) = \sum_{m=0}^{\ell} \varphi_m(0)\varphi_m(u)\mathbf{1}(|u| \le 1) \tag{1}$$

is a kernel of order ℓ.

(a) Use (1) to construct a kernel of order 1.

> We end up with the kernel corresponding to the Rosenblatt estimator; that is
> $$K(u) = \frac{1}{2}\mathbf{1}(|u| \le 1).$$

(b) Use (1) to construct a kernel of order 2.
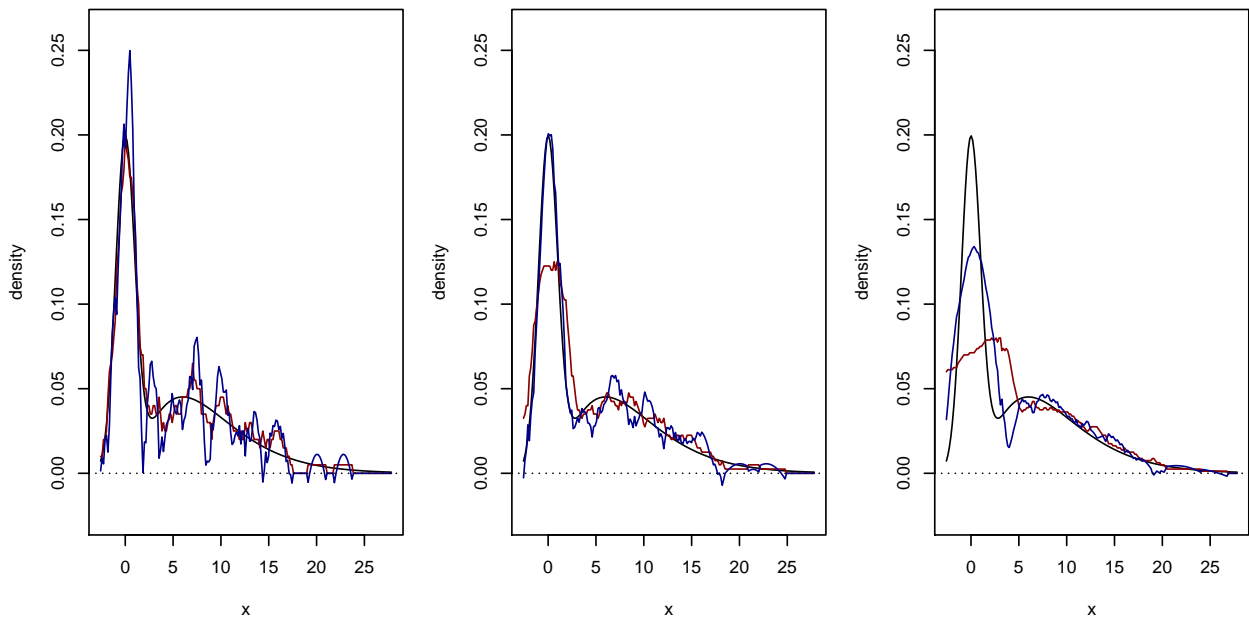
> We have
> $$\varphi_1(u) = \sqrt{\frac{3}{4}} \cdot u, \quad \varphi_2 = \sqrt{\frac{5}{2}} \cdot \frac{3u^2 - 1}{2}$$
> We obtain the kernel
> $$K(u) = \left(\frac{9}{8} - \frac{15}{8}u^2\right)\mathbf{1}(|u| \le 1).$$

(c) Generate some data and make a plot of the KDE (you must code your own KDE—no using built-in functions) based on these two kernels. Include in the plot the true density from which the data were generated (it is up to you how you generate the data. Be creative!). Turn in your code along with the plots—R Markdown is great for this!!!

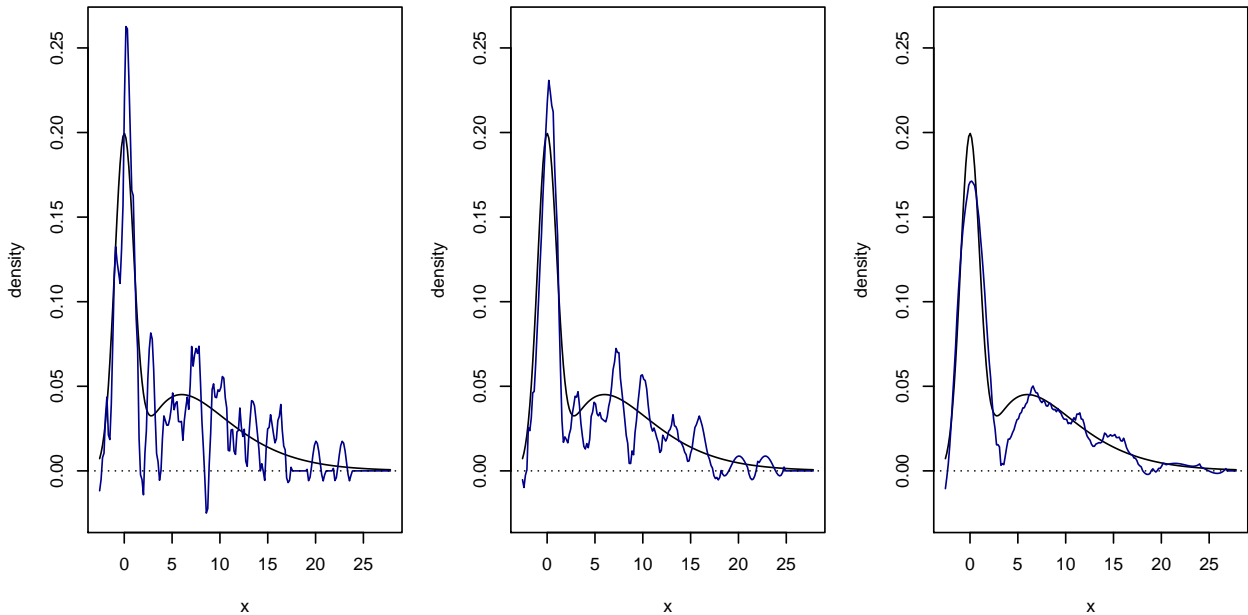> My plots look like this (I considered three different bandwidths. The blue line is the 2nd order kernel.):
>
>
>
> The kernel of order 4 constructed in this way happens to be
> $$K(u) = \left(\frac{225}{128} - \frac{1050}{128}u^2 + \frac{945}{128}u^4\right)\mathbf{1}(|u| \le 1),$$

based on
$$\varphi_3(u) = \sqrt{\frac{7}{2}} \cdot \frac{5u^2 - 3u}{2}, \quad \varphi_4(u) = \sqrt{\frac{9}{2}} \cdot \frac{35u^4 - 30u^2 + 3}{8}.$$

Here is a plot of of the KDE based on the kernel of order 4:

6. Write an (elegant) R function that chooses the bandwidth $h = \operatorname{argmin}_{h>0} CV(h)$ from a grid of candidate values. Refer to Lec 2. Make your own choice of the kernel function $K(u)$. Your function should have two arguments: x for the data values and N for the number of candidate bandwidths in the grid.

   (a) Include your R function when you turn in your hw.

   (b) Generate some data from a distribution of your choice and plot on a single set of axes

   1. the true density,
   2. the KDE under the leave-one-out CV bandwidth,
   3. and the KDE under the Sheather-Jones bandwidth.

References

[1] Alexandre B Tsybakov. *Introduction to nonparametric estimation.* Springer Science & Business Media, 2008.