STAT 824 hw 01

Hoeffding's inequality, KS test, Brownian bridge, kernel density estimation, Hölder smoothness, higher order kernels, CV for KDE bandwidth selection

1. Let B be a Brownian motion and B_0 be a Brownian bridge.

   (a) Show $\text{Var} \int_0^1 B(t)dt = 1/3$.

   Solution: Note that $\mathbb{E} \int_0^1 B(t)dt = 0$ (regarding $B(t)$ as a sum of Normals with zero mean and regarding also the integral as a sum, pass the expectation operator through the sums). Then write

$$
\begin{aligned}
\text{Var} \int_0^1 B(t)dt &= \mathbb{E} \int_0^1 \int_0^1 B(t)B(s)dtds \\
&= \int_0^1 \int_0^1 \mathbb{E}B(t)B(s)dtds \\
&= \int_0^1 \int_0^1 \text{Cov}(B(t), B(s))dtds \\
&= \int_0^1 \int_0^1 \min(s,t)dtds \\
&= \int_0^1 \int_0^1 [s\mathbf{1}(s < t) + t\mathbf{1}(t < s)]dtds \\
&= 2\int_0^1 \int_0^t tdt \\
&= \int_0^1 t^2 dt \\
&= 1/3.
\end{aligned}
$$

   We can see $\text{Cov}(B(t), B(s)) = \min(s,t)$ by noting that for $Z_1, Z_2, \dots \overset{\text{ind}}{\sim} \text{Normal}(0,1)$ we have

$$
\lim_{n\to\infty} \text{Cov}\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} Z_i, \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} Z_i\Big) = \min(s,t).
$$

   (b) Show $\text{Var} \int_0^1 B_0(t)dt = 1/12$.

Solution: We have

$$\begin{aligned}
\operatorname{Var}\int_0^1 B_0(t)dt &= \operatorname{Var}\int_0^1 (B(t) - tB(1))dt \\
&= \operatorname{Var}\int_0^1 B(t)dt + \operatorname{Var}\int_0^1 tB(1)dt - 2\operatorname{Cov}\Big(\int_0^1 B(t)dt, \int_0^1 tB(1)dt\Big) \\
&= 1/3 + \operatorname{Var}((1/2)B(1)) - 2\operatorname{Cov}\Big(\int_0^1 B(t)dt, (1/2)B(1)\Big) \\
&= 1/3 + 1/4\operatorname{Var}(B(1)) - \int_0^1 \operatorname{Cov}(B(t), B(1))dt \\
&= 1/3 + 1/4 - 1/2 \\
&= 1/12.
\end{aligned}$$

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

Solution: Noting that $X_i - p \in [-p, 1-p]$ for each $i = 1, \dots, n$, we can use Hoeffding's inequality to write

$$\begin{aligned}
P(\bar{X}_n - p > \varepsilon) &= P\Big(n^{-1}\sum_{i=1}^n (X_i - p) > \varepsilon\Big) \\
&= P\Big(\sum_{i=1}^n (X_i - p) > n\varepsilon\Big) \\
&\leq \exp\Big(-\frac{2n^2\varepsilon^2}{n}\Big) \\
&= e^{-2n\varepsilon^2}.
\end{aligned}$$

Likewise $P(\bar{X}_n - p < -\varepsilon) = P(-(\bar{X}_n - p) > \varepsilon) \leq e^{-2n\varepsilon^2}$, so

$$P(|\bar{X}_n - p| > \varepsilon) = P(\bar{X}_n - p < -\varepsilon) + P(\bar{X}_n - p > \varepsilon) \leq 2e^{-2n\varepsilon^2}.$$

(b) Let $Y_1, \dots, Y_n$ be iid with continuous cdf F_Y. Show that $P(|\hat{F}_n(y) - F(y)| \leq \varepsilon) \geq 1 - 2e^{-2n\varepsilon^2}$ for each $y \in \mathbb{R}$, where $\hat{F}_n(y) = n^{-1}\sum_{i=1}^n \mathbf{1}(Y_i \leq y)$.

Solution: Let $p = F_Y(y)$ and $X_i = \mathbf{1}(Y_i < y)$. Then $X_i \sim$ Bernoulli(p) and $\hat{F}_n(y) - F_Y(y) = \bar{X}_n - p$. Therefore $P(|\hat{F}_n(y) - F_Y(y)| > \varepsilon) \leq 2e^{-2n\varepsilon^2}$ from the first part.

(c) Explain how what you proved in part (b) is different from the DKW inequality.

Solution: The DKW inequality is a much stronger result because it is a probability bound for the maximum difference between $\hat{F}_n(y)$ and $F(y)$ across all y. What we proved in part (b) is for a single fixed value of y. So in part (b) we proved

$$\sup_{y\in\mathbb{R}} P(|\hat{F}_n(y) - F(y)| \le \varepsilon) \ge 1 - 2e^{-2n\varepsilon^2},$$

whereas the DKW inequality is

$$P(\sup_{y\in\mathbb{R}} |\hat{F}_n(y) - F(y)| \le \varepsilon) \ge 1 - 2e^{-2n\varepsilon^2}.$$

The proof of the DWK inequality is much more complicated [1].

3. For any random variable $Y \in [a, b]$, show that $\operatorname{Var} Y \le (b-a)^2/4$.

Solution: For $y \in [a, b]$ we have

$$(y - (a+b)/2)^2 \le (b - (a+b)/2)^2 = (b-a)^2/4.$$

So

$$\begin{aligned}
\operatorname{Var} Y &= \operatorname{Var}(Y - (a+b)/2) \\
&= \mathbb{E}(Y - (a+b)/2)^2 - [\mathbb{E}(Y - (a+b)/2)]^2 \\
&\le \mathbb{E}(Y - (a+b)/2)^2 \\
&\le (b-a)^2/4.
\end{aligned}$$

4. Consider the random variables X and Y, where

$$\begin{aligned}
Y &\sim \text{Normal}(0, 1) \\
X|\delta &\sim \text{Normal}(a\cdot\delta, 1-a^2), \text{ where } \delta \in \{-1, 1\}, \text{ with } P(\delta = 1) = 1/2,
\end{aligned}$$

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

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

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

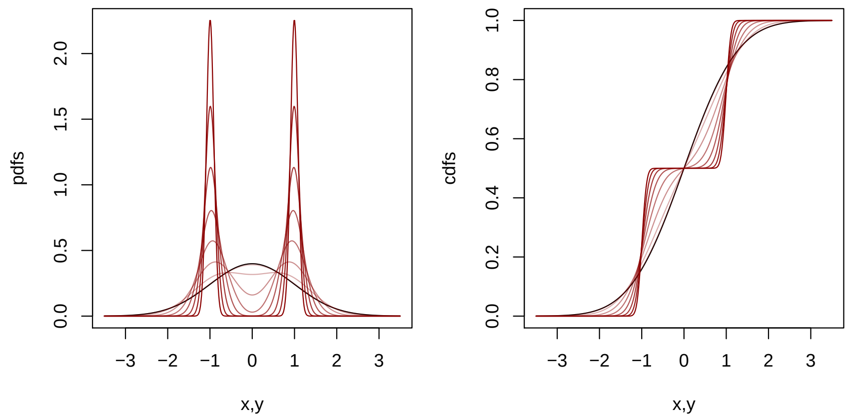
(b) Give $\operatorname{Var} X$.

Solution: 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 $\mathbb{E}X^3$.

Solution: 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 $a \in \{1 - (1/2)^j : j = 0, 1, \dots, 8\}$, generate 100 random samples $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ (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 R function `ks.test()`; write your own code in R or python and turn it in along with the table.



Solution: 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.

Solution: 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

5. 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.

Solution: 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.

Solution: 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 $\int_{\mathbb{R}} x^2\hat{f}_n(x)dx - (\int_{\mathbb{R}} x\hat{f}_n(x)dx)^2$?

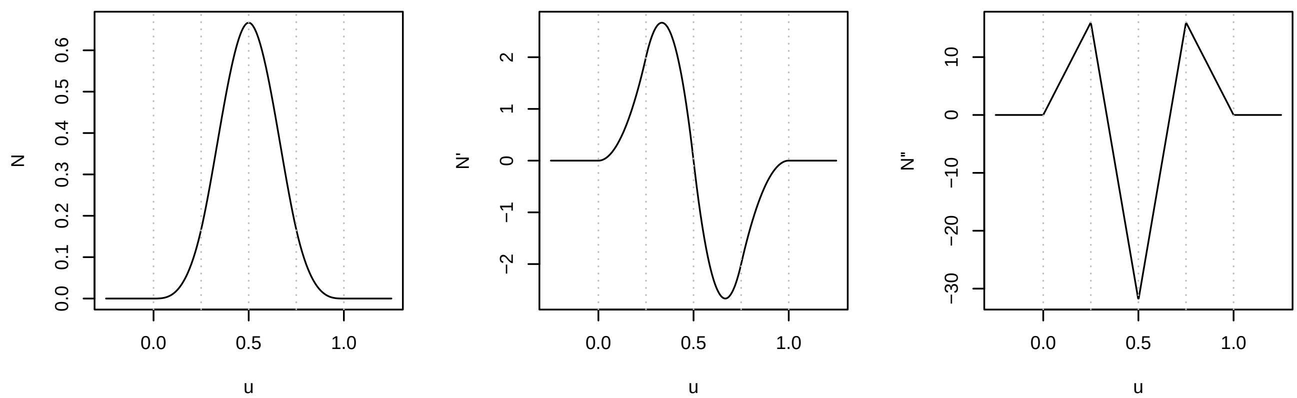
Solution: The Gaussian kernel is a kernel of order 1, and causes the variance according to $\hat{f}_n$ to increase; the kernel of order 2 preserves the variance in the sense that the variance according to $\hat{f}_n$ is equal to the empirical variance of $X_1, \ldots, X_n$.

6. Consider the function

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

Find a positive integer β and $L > 0$ such that $N(u) \in \mathcal{H}(\beta, L)$.

Solution: We find that N has two continuous derivatives, and the second derivative satisfies a Lipschitz condition with $L = 192$. So the best answer is $N(u) \in \mathcal{H}(\beta = 3, L = 192)$. But this is not all we can say: For a visual, we plot below the function $N(u)$ as well as its first two derivatives $N'(u)$ and $N''(u)$.



We have

$$N'(u) = \begin{cases} 32u^2, & 0 \leq u < 1/4 \\ 64u - 96u^2 - 8, & 1/4 \leq u < 2/4 \\ -64(1-u) + 96(1-u)^2 + 8, & 2/4 \leq u < 3/4 \\ -32(1-u)^2, & 3/4 \leq u < 1 \\ 0, & \text{otherwise} \end{cases}$$

and

$$N''(u) = \begin{cases} 96u, & 0 \leq u < 1/4 \\ 64 - 192u, & 1/4 \leq u < 2/4 \\ 64 + 96(1-u), & 2/4 \leq u < 3/4 \\ 64(1-u), & 3/4 \leq u < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Note that $N''(x)$ is not differentiable.

From here we see the following:

1. We can say $N(u)$ belongs to the $\mathcal{H}(3, L)$ class for any $L \geq 192$, since the maximum absolute slope of any line tangent to $N''(u)$ is 192. We have $|N''(u) - N''(u')| \leq 192|u - u'|$ for any $u, u' \in (0, 1)$.

2. We can also say $N(u)$ belongs to the $\mathcal{H}(2, L)$ class for any $L \geq 32$, since the maximum absolute slope of any line tangent to $N'(u)$ is $32 = \sup_{u \in [0,1]} |N''(u)| = |N''(1/2)|$. We have $|N'(u) - N'(u')| \leq 32|u - u'|$ for any $u, u' \in (0, 1)$.

3. Indeed, we can also say $N(u)$ belongs to the $\mathcal{H}(1, L)$ class (which is the same as the Lipschitz class) for any $L \geq 8/3$, since the maximum absolute slope of any line tangent to $N(u)$ is $8/3 = \sup_{u \in [0,1]} |N'(u)| = N'(1/3)$. We have $|N(u) - N(u')| \leq (8/3)|u - u'|$ for any $u, u' \in (0, 1)$.

However, when placing a function in a Hölder class, we prefer to take the maximum number of derivatives; then after fixing the number of derivatives, we prefer to take the smallest L.

7. Let $\{\varphi_m(\cdot)\}_{m=0}^{\infty}$ represent polynomials defined by

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

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

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

Proposition 1.3 of [2] gives that the function $K : \mathbb{R} \to \mathbb{R}$ given by

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

is a kernel of order ℓ.

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

Solution: We end up with the kernel corresponding to the Rosenblatt estimator; that is

$$K(u) = \frac{1}{2}\mathbf{1}(|u| \leq 1).$$

(b) Use (1) to construct a kernel of order 2.

Solution: We have

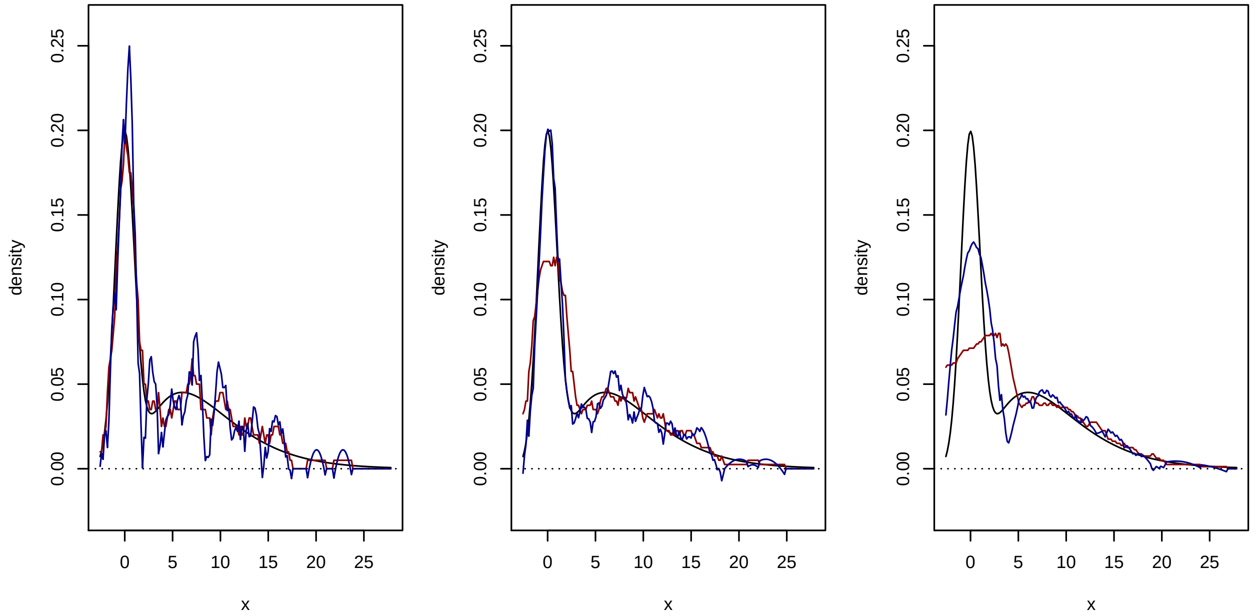
$$\varphi_1(u) = \sqrt{\frac{3}{4}} \cdot u, \quad \varphi_2 = \sqrt{\frac{5}{2}} \cdot \frac{3u^2-1}{2}$$

We obtain the kernel

$$K(u) = \left(\frac{9}{8} - \frac{15}{8}u^2\right)\mathbf{1}(|u| \leq 1).$$

(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.

Solution: My plots look like this (I considered three different bandwidths. The blue line is the 2nd order kernel.):



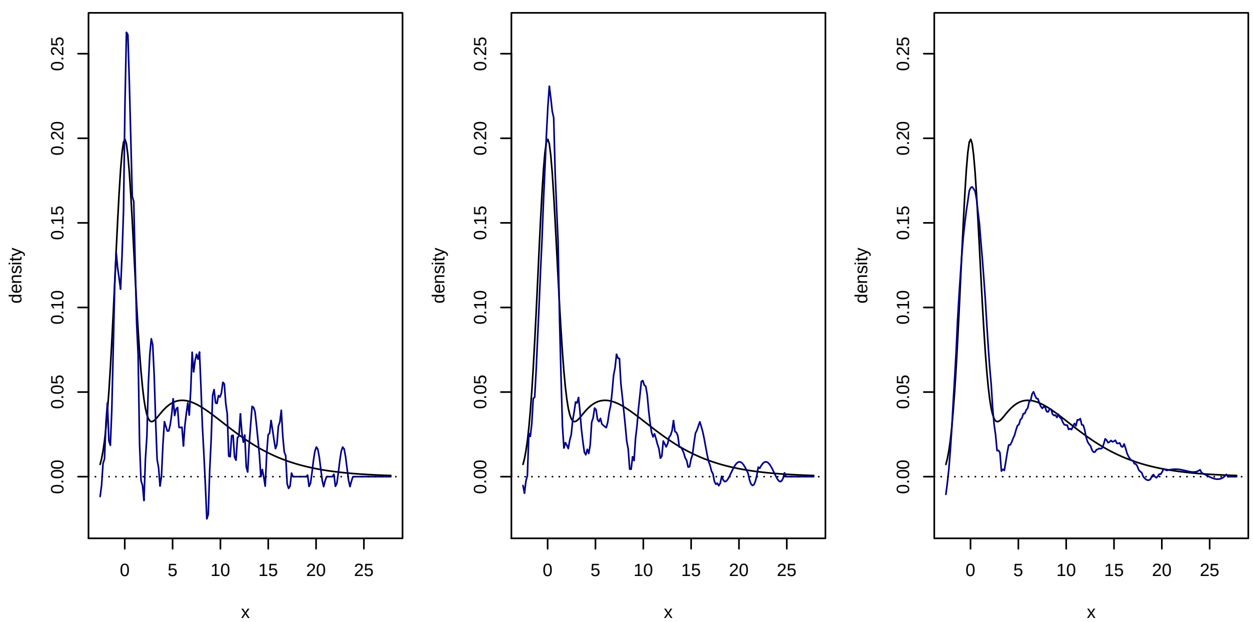
The kernel of order 4 constructed in this way happens to be

$$K(u) = \left(\frac{225}{128} - \frac{1050}{128}u^2 + \frac{945}{128}u^4\right)\mathbf{1}(|u| \leq 1),$$

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:



8. Write an (elegant) R or python 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 or python 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] Pascal Massart. The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *The annals of Probability*, pages 1269–1283, 1990.

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