Homework 1
Due Friday Feb 1 3:00 pm

1. Let $X_1, \ldots, X_n \sim P$ where $X_i \in [0,1]$ and P has density p. Let $\widehat{p}$ be the histogram estimator using m bins. Let $h = 1/m$. Recall that the L_2 error is $\int (\widehat{p}(x) - p(x))^2 = \int \widehat{p}^2(x)dx - 2\int \widehat{p}(x)p(x)dx + \int p^2(x)dx$. As usual, we may ignore the last term so we define the loss to be

$$L(h) = \int \widehat{p}^2(x)dx - 2\int \widehat{p}(x)p(x)dx.$$

(a) Suppose we used the direct estimator of the loss, namely, we replace the intergal with the average to get

$$\widehat{L}(h) = \int \widehat{p}^2(x)dx - \frac{2}{n}\sum_i \widehat{p}(X_i).$$

Show that this fails in the sense that it is minimized by taking $h = 0$.

(b) Recall that the leave-one-out estimator of the risk is

$$\widehat{L}(h) = \int \widehat{p}^2(x)dx - \frac{2}{n}\sum_i \widehat{p}_{(-i)}(X_i).$$

Show that

$$\widehat{L}(h) = \frac{2}{(n-1)h} - \frac{n+1}{n^2(n-1)h}\sum_j Z_j^2$$

where Z_j is the number of observations in bin j.

2. Let $\widehat{p}_h$ be the kernel density estimator (in one dimension) with bandwidth $h = h_n$. Let $s_n^2(x) = \mathrm{var}(\widehat{p}_h(x))$.

(a) Show that, under appropriate conditions,

$$\frac{\widehat{p}_h(x) - p_h(x)}{s_n(x)} \rightsquigarrow N(0,1)$$

where $p_h(x) = \mathbb{E}[\widehat{p}_h(x)]$.

Hint: Recall that the Lyapunov central limit theorem says the following: Suppose that $Y_1, Y_2, \ldots$ are independent. Let $\mu_i = \mathbb{E}[Y_i]$ and $\sigma_i^2 = \mathrm{Var}(Y_i)$. Let $s_n^2 = \sum_{i=1}^n \sigma_i^2$. If

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}[|Y_i - \mu_i|^{2+\delta}] = 0$$

for some $\delta > 0$. Then $s_n^{-1}\sum_i (Y_i - \mu_i) \rightsquigarrow N(0,1)$.

(b) Assume that the smoothness is $\beta = 2$. Suppose that the bandwidth h_n is chosen optimally. Show that

$$\frac{\widehat{p}_h(x) - p(x)}{s_n(x)} \rightsquigarrow N(b(x), 1)$$

for some constant $b(x)$ which is, in general, not 0.

3. Let $X_1, \ldots, X_n \sim P$ where $X_i \in [0,1]$. Assume that P has density p which has a bounded continuous derivative. Let $\widehat{p}_h(x)$ be the kernel density estimator. Show that, in general, the bias is of order $O(h)$ at the boundary. That is, show that $\mathbb{E}[\widehat{p}_h(0)] - p(0) = Ch$ for some $C > 0$.

4. Let p be a density on the real line. Assume that p is m-times continuosly differentiable and that $\int |p^{(m)}|^2 < \infty$. Let K be a higher order kernel. This means that $\int K(y)dy = 1$, $\int y^j K(y)dy = 0$ for $1 \leq j \leq m-1$, $\int |y|^m K(y)dy < \infty$ and $\int K^2(y)dy < \infty$. Show that the kernel estimator with bandwidth h satisfies

$$\mathbb{E} \int (\widehat{p}(x) - p(x))^2 dx \leq C \left(\frac{1}{nh} + h^{2m} \right)$$

for some $C > 0$. What is the optimal bandwidth and what is the corresponding rate of convergence (using this bandwidth)?

5. Let $X_1, \ldots, X_n \sim P$ where $X_i \in [0,1]$ and P has density p. Let $\phi_1, \phi_2, \ldots$ be an orthonormal basis for $L_2[0,1]$. Hence $\int_0^1 \phi_j^2(x)dx = 1$ for all j and $\int_0^1 \phi_j(x)\phi_k(x)dx = 0$ for $j \neq k$. Assume that the basis is uniformly bounded i.e. $\sup_j \sup_{0 \leq x \leq 1} |\phi_j(x)| \leq C < \infty$. We may expand p as $p(x) = \sum_{j=1}^{\infty} \beta_j \phi_j(x)$ where $\beta_j = \int \phi_j(x)p(x)dx$. Define

$$\widehat{p}(x) = \sum_{j=1}^{k} \widehat{\beta}_j \phi_j(x)$$

where $\widehat{\beta}_j = (1/n) \sum_{i=1}^{n} \phi_j(X_i)$.

(a) Show that the risk is bounded by

$$\frac{ck}{n} + \sum_{j=k+1}^{\infty} \beta_j^2$$

for some constant $c > 0$.

(b) Define the Sobolev ellipsoid $E(m, L)$ of order m as the set of densities of the form $p(x) = \sum_{j=1}^{\infty} \beta_j \phi_j(x)$ where $\sum_{j=1}^{\infty} \beta_j^2 j^{2m} < L^2$. Show that the risk for any density in $E(m, L)$ is bounded by $c[(k/n) + (1/k)^{2m}]$. Using this bound, find the optimal value of k and find the corresponding risk.

6. Recall that the total variation distance between two distributions P and Q is $\mathrm{TV}(P,Q) = \sup_A |P(A) - Q(A)|$. In some sense, this would be the ideal loss function to use for density estimation. We only use L_2 because it is easier to deal with. Here you will explore some properties of TV.

(a) Suppose that P and Q have densities p and q. Show that

$$\mathrm{TV}(P,Q) = (1/2) \int |p(x) - q(x)| dx.$$

(b) Let T be any mapping. Let X and Y be random variables. Then

$$\sup_A |P(T(X) \in A) - P(T(Y) \in A)| \leq \sup_A |P(X \in A) - P(Y \in A)|.$$

(c) Let K be a kernel. Recall that the convolution of a density p with K is $(p \star K)(x) = \int p(z)K(x-z)dz$. Show that

$$\int |p \star K - q \star K| \leq \int |K| \int |p - q|.$$

Hence, smoothing reduces L_1 distance.

(d) Let p be a density on $\mathbb{R}$ and let p_n be a sequence of densities. Suppose that $\int (p - p_n)^2 \to 0$. Show that $\int |p - p_n| \to 0$.

(e) Let $\widehat{p}$ be a histogram on $\mathbb{R}$ with binwidth h. Under some regularity conditions it can be shown that

$$\mathbb{E} \int |\widehat{p} - p| \approx \frac{\sqrt{2}}{\pi n h} \int \sqrt{p} + \frac{1}{4} h \int |p'|.$$

Hence, this risk can be unbounded if $\int \sqrt{p} = \infty$. A density is said to have a regularly varying tail of order r if $\lim_{x\to\infty} p(tx)/p(x) = t^r$ for all $t > 0$ and $\lim_{x\to-\infty} p(tx)/p(x) = t^r$ for all $t > 0$. Suppose that p has a regularly varying tail of order r with $r < -2$. Show that the risk bound above is bounded.