1.- $D = \{X_1, \dots, X_n\}$

a) Notice that for a given resample, the ith observation can appear more than once.
Let Y_j = Number of times X_j appears in the resample, then $Y_j \sim Bin(k, 1/n)$
It follows

$$\mathbb{P}(X_j \text{ appears in the resample}) = \mathbb{P}(Y_j > 0) = 1 - \mathbb{P}(Y_j = 0) = 1 - (1 - 1/n)^k$$

b) Let Z_i a random variable defined as $Z_i = \begin{cases} 1 & X_i \text{ appears in the resample} \\ 0 & X_i \text{ doesn't appear in the resample} \end{cases}$
It follows

$$\mathbb{E}[Z_i] = 1 \left[\mathbb{P}(Z_i = 1)\right] = \mathbb{P}(Y_i > 0) = 1 - \left(1 - \frac{1}{n}\right)^k$$

Now, consider $Z = \sum_{i=1}^n Z_i$, then Z counts the number of unique observations present in the resample. The expected number of observations is then

$$\mathbb{E}[Z] = \mathbb{E}\left[\sum_{i=1}^n Z_i\right] = \sum_{i=1}^n \mathbb{E}[Z_i] = \sum_{i=1}^n \left[1 - \left(1 - \frac{1}{n}\right)^k\right] = n\left[1 - \left(1 - \frac{1}{n}\right)^k\right]$$

We can summarize this as

$$\underbrace{n\left[1 - \left(1 - \frac{1}{n}\right)^k\right]}_{\text{Expected number of observations}} \qquad \underbrace{1 - \left(1 - \frac{1}{n}\right)^k}_{\text{Expected proportion}}$$

c) If $k = n$ And $n \to \infty$ then

$$\lim_{n \to \infty} 1 - \left(1 - \frac{1}{n}\right)^n = 1 - e^{-1} \approx 0.632$$

d) The 0.632 rule states that for a sample with size N, the average number of distinct observations in each bootstrap sample is about 0.632(N). 0.632 was the proportion of unique observations we derived on part (c)

3.- For linear smoothers we derived in class

$$df(\hat{y}) = \frac{1}{\sigma^2} tr(Cov(WY, Y)) = \frac{1}{\sigma^2} tr(W\, Cov(Y,Y)) = \frac{1}{\sigma^2} tr(W \sigma^2 I_n) = tr(W) = \sum_{i=1}^{n} w(X_i, X_i)$$

For k-nearest neighbors we can define

$$w(X_i, X) = \begin{cases} 1/k & X_i \text{ is one of the } k \text{ nearest neighbors of } X \\ 0 & \text{o.w.} \end{cases}$$

It follows

$$df(\hat{y}) = \sum_{i=1}^{n} w(X_i, X_i) = \sum_{i=1}^{n} \frac{1}{k} = \frac{n}{k}$$ $\therefore$ k-nearest neighbors haves $\frac{n}{k}$ degrees of freedom

4.- Here $\hat{Y} = \hat{f}(X) = \bar{Y}$, it follows

$$df(\hat{Y}) = \frac{1}{\sigma^2} \sum_{i=1}^{n} Cov(\hat{Y}_i, Y_i) = \frac{1}{\sigma^2} \sum_{i=1}^{n} Cov\left(\frac{\sum_{j=1}^{n} Y_j}{n}, Y_i\right)$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{n} \frac{1}{n} Cov\left(\sum_{j=1}^{n} Y_j, Y_i\right) = \frac{1}{\sigma^2} \sum_{i=1}^{n} \frac{1}{n} Cov(Y_i, Y_i)$$ Obs. Y_i, Y_j uncorrelated when $i \neq j$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{n} \frac{\sigma^2}{n} = 1$$

5.- Consider the following. Suppose we have a sample $D = \{(X_1, Y_1), \dots, (X_{11}, Y_{11})\}$ where $X_i = i$ and $Y_i \in \mathbb{R}$, $i \in \{1, \dots, 11\}$.

Consider the uniform Kernel with bandwith $h=1$, then

$$\hat{F}(X_5) = \hat{F}(5) = \frac{\sum_{i=1}^{11} K(X_i - 5) Y_i}{\sum_{i=1}^{n} K(X_i - 5)} = \frac{\frac{1}{2} Y_4 + \frac{1}{2} Y_5 + \frac{1}{2} Y_6}{\frac{1}{2} + \frac{1}{2} + \frac{1}{2}} = \frac{\frac{1}{2}(Y_4 + Y_5 + Y_6)}{\frac{3}{2}}$$

$$= \frac{1}{3}(Y_4 + Y_5 + Y_6)$$

Now, for k-nearest neighbors with $k=3$ we have

$$\hat{F}(X_5) = \frac{1}{3}(Y_4 + Y_5 + Y_6)$$

Now consider a dataset $D' = \{(X_1, Y_1), \dots (X_{11}, Y_{11})\}$ where $X_i = 2i$, $i \in \{1, \dots, 11\}$ and Y_i are the same as in D.

For the uniform Kernel with bandwith $h=1$

$$\hat{F}(X_5) = \hat{F}(5) = \frac{\sum_{i=1}^{11} K(X_i - 5) Y_i}{\sum_{i=1}^{11} K(X_i - 5)} = \frac{\frac{1}{2} Y_5}{\frac{1}{2}} = Y_5$$

For k-nearest neighbors with $k=3$

$$\hat{F}(X_5) = \frac{1}{3}(Y_4 + Y_5 + Y_6)$$

As we can see, when the points X_i were set further apart the estimate using uniform kernel changed, this didn't happen with k-nearest neighbors. This way we can see that uniform kernel and k-nearest neighbors aren't exactly the same

2.- Let P_B be the probability that any given observation in the original training sample appears in more than half of the bootstrap samples.

Let W_{ij} = Number of times the jth observation (X_j, Y_j) appears in the ith bootstrap sample

Let $Z_{ij} = \begin{cases} 1 & (X_j, Y_j) \text{ appears in the ith bootstrap sample} \\ 0 & \text{o.w.} \end{cases}$

$$\mathbb{P}(Z_{ij}=1) = \mathbb{P}(W_{ij}>0) = 1-\mathbb{P}(W_{ij}=0) = 1-\left(1-\tfrac{1}{n}\right)^n$$

Now, since the appearences in the bootstrap samples are independent, the distribution of Z_j the number of samples where (X_j, Y_j) appears is

$$Z_j = \sum_{i=1}^{B} Z_{ij}, \quad i \in \{1, \dots, B\} \quad \text{Moreover} \quad Z_j \sim Bin(B, P_n), \; P_n = 1-\left(1-\tfrac{1}{n}\right)^n$$

Computing the probability (X_j, Y_j) appears in more than half of the bootstrap samples

(I) If B is even $B = 2b$, $b \in \mathbb{N}$

$$\mathbb{P}(Z_j > b) = 1-\mathbb{P}(Z_j \leq b) = 1-\sum_{k=0}^{b}\binom{B}{k} P_n^k (1-P_n)^{B-k} = P_B$$

(II) If B is odd $B = 2b'+1$, $b' \in \mathbb{N}$ so we need $Z_j \geq b'+1$

$$\mathbb{P}(Z_j > b') = 1-\mathbb{P}(Z_j \leq b') = 1-\sum_{k=0}^{b'}\binom{B}{k} P_n^k (1-P_n)^{B-k} = P_B$$

Now, if we want all observations to appear in at least half of the bootstrap samples

(III) If B is even $B=2b$, $b \in \mathbb{N}$

$$\mathbb{P}(Z_j \geq b) = 1-\mathbb{P}(Z_j \leq b-1) = 1-\sum_{k=0}^{b-1}\binom{B}{k}(P_n)^k(1-P_n)^{B-k}$$

(iv) If B is odd $B = 2b' + 1,\ b' \in \mathbb{N}$

$$\mathbb{P}(Z_j \geq B/2) = 1 - \mathbb{P}(Z_j \leq b') = 1 - \sum_{k=0}^{b'} \binom{B}{k} (P_n)^k (1-P_n)^{B-k}$$

Now $P_B = 1 - \mathbb{P}(Z_j \leq b) \leq 1 - \mathbb{P}(Z_j \leq b-1)$

So for all observations

$$P_B^n \leq \left[1 - \mathbb{P}(Z_j \leq b-1)\right]^n \quad \therefore \text{The claim is false}$$