Statistique en grande dimension et apprentissage
Data-driven Bandwidth Selection for Histogram

Dinh Hung Tu
École Normale Supérieure de Cachan
Master MVA Mathématiques Vision Apprentissage

May 23, 2016

Problem set up

From the problem set up, we are given the following equations:

$$\hat{f}_{n,h}(x) = \frac{1}{h}\sum_{k=1}^{K}\hat{p}_k \mathbb{1}_{C_k}(x) \tag{1}$$

$$\text{where } \hat{p}_k = \frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{C_k}(x) \text{ is the empirical probability} \tag{2}$$

$$p_k = \int_{C_k} f(x)dx \text{ theoretical probability} \tag{3}$$

$$\hat{J}_n(h) = \frac{2}{(n-1)h} - \frac{n+1}{(n-1)h}\sum_{k=1}^{K}\hat{p}_k^2 \tag{4}$$

$$\xi_n^{(1)}(h) = ||\hat{f}_{n,h} - f||_2^2 - ||f||_2^2 - \hat{J}_n(h) \tag{5}$$

Answers

Problem 1. Show that $\xi_n(h) = \xi_n^{(1)}(h) + \xi_n^{(2)}(h)$ where

$$\xi_n^{(1)}(h) = \frac{2n}{(n-1)h}\sum_{k=1}^{K}(\hat{p_k}^2 - \mathbb{E}[\hat{p_k}^2])$$

$$\xi_n^{(2)}(h) = \frac{2}{h}\sum_{k=1}^{K}p_k(p_k - \hat{(p_k)}$$

Proof. First term in the definition of $\xi_n(h)$

$$\begin{aligned}
||\hat{f}_{n,h} - f||_2^2 &= \int_0^1 (\hat{f}_{n,h} - f)^2 dx \\
&= \sum_{k=1}^{K} \int_{C_k} (\hat{f}_{n,h} - f)^2 dx \\
&= \sum_{k=1}^{K} \int_{C_k} \left(\frac{1}{h} \hat{p}_k - f \right)^2 dx \\
&= \sum_{k=1}^{K} \left(\frac{1}{h^2} \int_{C_k} \hat{p}_k^2 dx - \frac{2}{h} \int_{C_k} \hat{p}_k f dx + \int_{C_k} f^2 dx \right)
\end{aligned}$$

Since $\hat{p}_k$ is a constant in the interval C_k and from definition of p_k in 3, we have:

$$\begin{aligned}
||\hat{f}_{n,h} - f||_2^2 &= \sum_{k=1}^{K} \left(\frac{1}{h^2} \hat{p}_k^2 \int_{C_k} dx - \frac{2}{h} \hat{p}_k p_k \right) + \int_0^1 f^2 dx \\
&= \frac{1}{h} \sum_{k=1}^{K} \left(\hat{p}_k^2 - 2 \hat{p}_k p_k \right) + ||f||_2^2
\end{aligned}$$

From the definition in 5:

$$\begin{aligned}
\xi_n(h) &= ||\hat{f}_{n,h} - f||_2^2 - ||f||_2^2 - \hat{J}_n(h) \\
&= \frac{1}{h} \sum_{k=1}^{K} \left(\hat{p}_k^2 - 2 \hat{p}_k p_k \right) - \hat{J}_n(h) \\
&= \frac{1}{h} \sum_{k=1}^{K} \left(\hat{p}_k^2 - 2 \hat{p}_k p_k \right) - \frac{2}{(n-1)h} + \frac{n+1}{(n-1)h} \sum_{k=1}^{K} \hat{p}_k^2 \\
&= \frac{2n}{(n-1)h} \sum_{k=1}^{K} \hat{p}_k^2 - \frac{2}{h} \sum_{k=1}^{K} \hat{p}_k p_k - \frac{2}{(n-1)h}
\end{aligned}$$

Consider $n\hat{p}_k$ as a random variable which follows a binomial distribution $n\hat{p}_k \sim \mathcal{B}(n, p_k)$, we have:

$$\begin{aligned}
\mathbb{E}[\hat{p}_k] &= p_k \\
\text{Var}(\hat{p}_k) &= \frac{p_k(1 - p_k)}{n} \\
\mathbb{E}[\hat{p}_k^2] &= (\mathbb{E}[\hat{p}_k])^2 + \text{Var}(\hat{p}_k) = p_k^2 + \frac{p_k(1 - p_k)}{n} \\
&= \frac{p_k}{n} + \frac{n-1}{n} p_k^2
\end{aligned}$$

Substitute this result into the following sum:

$$\begin{aligned}
\xi_n^{(1)}(h) &= \frac{2n}{(n-1)h}\sum_{k=1}^{K}(\hat{p}_k^2 - \mathbb{E}[\hat{p}_k^2]) \\
&= \frac{2n}{(n-1)h}\sum_{k=1}^{K}\left(\hat{p}_k^2 - \frac{p_k}{n} - \frac{n-1}{n}p_k^2\right)
\end{aligned} \tag{6}$$

$$\begin{aligned}
\xi_n^{(2)}(h) &= \frac{2}{h}\sum_{k=1}^{K}p_k(p_k - \hat{p}_k) = \frac{2}{h}\sum_{k=1}^{K}p_k(p_k - \hat{p}_k) \\
\xi_n^{(1)}(h) + \xi_n^{(2)}(h) &= \frac{2n}{(n-1)h}\left(\sum_{k=1}^{K}\hat{p}_k^2 - \frac{1}{n}\underbrace{\sum_{k=1}^{K}\hat{p}_k}_{=1}\right)\underbrace{-\frac{2}{h}\sum_{k=1}^{K}p_k^2 + \frac{2}{h}\sum_{k=1}^{K}p_k(p_k - \hat{p}_k)}_{=0} \\
&= \frac{2n}{(n-1)h}\left(\sum_{k=1}^{K}\hat{p}_k^2 - \frac{1}{n}\right) - \frac{2}{h}\sum_{k=1}^{K}p_k\hat{p}_k
\end{aligned} \tag{7}$$

Results above show that $\xi_n(h) = \xi_n^{(1)}(h) + \xi_n^{(2)}(h)$ □

Problem 2. (a) Let us set

$$Z_i = \frac{2}{h}\sum_{k=1}^{K}p_k(p_k - \mathbb{1}_{C_k}(X_i)) \text{ i=1,...,n}$$

Check that $\xi_n^{(2)}(h) = \frac{1}{n}\sum_{i=1}^{n}Z_i$

Proof.

$$\begin{aligned}
\xi_n^{(2)}(h) &= \frac{1}{n}\sum_{i=1}^{n}Z_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{2}{h}\sum_{k=1}^{K}p_k(p_k - \mathbb{1}_{C_k}(X_i)) \\
&= \frac{2}{h}\sum_{k=1}^{K}\frac{1}{n}\sum_{i=1}^{n}p_k(p_k - \mathbb{1}_{C_k}(X_i)) \\
&= \frac{2}{h}\sum_{k=1}^{K}p_k(p_k - \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{C_k}(X_i)) \\
&= \frac{2}{h}\sum_{k=1}^{K}p_k(p_k - \hat{p}_k)
\end{aligned}$$

Since X_i is iid, Z_i is also iid. □

Problem 2. (b) Show that the variance of Z_i is upper-bounded by

$$\frac{4}{h^2}\sum_{k=1}^{K} p_k^3(1-p_k)$$

Proof.

$$\begin{aligned}
Z_i &= \frac{2}{h}\sum_{k=1}^{K} p_k(p_k - \mathbb{1}_{C_k}(X_i)) \\
&= \underbrace{\frac{2}{h}\sum_{k=1}^{K} p_k^2}_{\text{const}} - \frac{2}{h}\sum_{k=1}^{K} p_k \mathbb{1}_{C_k}(X_i) \\
\text{Let denote } g(X_i) &= \sum_{k=1}^{K} p_k \mathbb{1}_{C_k}(X_i), \text{ by the properties of variance:} \\
\mathrm{Var}(Z_i) &= \frac{4}{h^2}\mathrm{Var}(g) \\
\mathbb{E}(g) &= \sum_{k=1}^{K} p_k \times \Pr(X_j \in C_k) = \sum_{k=1}^{K} p_k^2 \\
\mathbb{E}(g^2) &= \sum_{k=1}^{K} p_k^2 \times \Pr(X_j \in C_k) = \sum_{k=1}^{K} p_k^3 \\
\mathrm{Var}(g) &= \mathbb{E}(g^2) - [\mathbb{E}(g)]^2 = \sum_{k=1}^{K} p_k^3 - \left(\sum_{k=1}^{K} p_k^2\right)^2 \\
&\leq \sum_{k=1}^{K} p_k^3 - \sum_{k=1}^{K} p_k^4 \text{ (by Cauchy-Schwarz)} \\
&= \sum_{k=1}^{K} p_k^3(1-p_k) \\
\mathrm{Var}(Z_i) &= \frac{4}{h^2}\mathrm{Var}(g) = \frac{4}{h^2}\sum_{k=1}^{K} p_k^3(1-p_k)
\end{aligned}$$

□

Problem 2. (c) Deduce from previous questions an upper bound of the form

$$\mathrm{Var}[\xi_n^{(2)}(h)] \leq \frac{C||f||_\infty^2}{n}$$

Proof. ..Incomplete.. □

Problem 3. (a) Let us denote $T = \{(i,j) \in \{1,...,n\}^2 \text{ s.t. } i < j\}$ and introduce the random variables

$$\mathcal{U} = \sum_{k=1}^{K} \left(\mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j) - p_k^2\right), \text{ if } t = (i,j)$$

Show that $\mathbb{E}(\mathcal{U}_t) = 0$ for every t and that

$$\xi_n^{(1)}(h) = \frac{4}{nh(n-1)} \sum_{t \in T} \mathcal{U}_t$$

Proof. Let denote $\mathcal{U}_{t,k}$ as follows

$$\mathcal{U}_{t,k} = \mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j) - p_k^2$$
$$\mathcal{U}_t = \sum_{k=1}^{K} \mathcal{U}_{t,k}$$

As $\mathbb{1}_{C_k}(X_i) \in \{1,0\}$ we have $\mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j) = \begin{cases} 1, & \text{if } \mathbb{1}_{C_k}(X_i) = 1 \text{ and } \mathbb{1}_{C_k}(X_j) = 1 \\ 0, & \text{otherwise} \end{cases}$

Therefore, $\mathcal{U}_{t,k}$ can take one of its two values $\mathcal{U}_{t,k}^{(1)}, \mathcal{U}_{t,k}^{(0)}$ with their respective probability. In specific:

$$\mathcal{U}_{t,k} = \mathcal{U}_{t,k}^{(1)} = 1 \times 1 - p_k^2 = 1 - p_k^2 \text{ with probability } \overset{(1)}{\underset{t,k}{\Pr}}$$
$$\overset{(1)}{\underset{t,k}{\Pr}} = \Pr(X_i \in C_k) \times \Pr(X_j \in C_k) = p_k^2$$
$$\mathcal{U}_{t,k} = \mathcal{U}_{t,k}^{(0)} = 0 - p_k^2 = -p_k^2 \text{ with probability } \overset{(0)}{\underset{t,k}{\Pr}}$$
$$\overset{(0)}{\underset{t,k}{\Pr}} = 1 - \overset{(1)}{\underset{t,k}{\Pr}} = 1 - p_k^2$$
$$\mathbb{E}[\mathcal{U}] = \sum_{k=1}^{K} \mathbb{E}[\mathcal{U}_{t,k}] = \sum_{k=1}^{K} \left(\mathcal{U}_{t,k}^{(1)} \times \overset{(1)}{\underset{t,k}{\Pr}} + \mathcal{U}_{t,k}^{(0)} \times \overset{(0)}{\underset{t,k}{\Pr}}\right)$$
$$= \sum_{k=1}^{K} \left((1 - p_k^2) \times p_k^2 + (-p_k^2) \times (1 - p_k^2)\right) = 0$$

For the second equality, note that size of set T is $\frac{n(n-1)}{2}$

$$
\begin{aligned}
\sum_{t\in T}\mathcal{U}_t &= \sum_{k=1}^{K}\sum_{t\in T}\left(\mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j) - p_k^2\right) \\
&= \sum_{k=1}^{K}\left[\sum_{t\in T}\left(\mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j)\right) - \sum_{t\in T} p_k^2\right] \\
&= \sum_{k=1}^{K}\left[\sum_{t\in T}\left(\mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j)\right) - \frac{n(n-1)}{2} p_k^2\right]
\end{aligned}
$$

By the definition of $\hat{p}_k$ in 2, we have the total number of sample X_i falling within bin C_k.

$$
\sum_{i=1}^{n}\mathbb{1}_{C_k}(X_i) = n\hat{p}_k
$$

Number of pairs $X_i \neq X_j$ that are both in the bin C_k.

$$
\begin{aligned}
\sum_{t\in T}\mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j) &= \sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j) \\
&= \frac{1}{2}n\hat{p}_k(n\hat{p}_k - 1)
\end{aligned}
$$

Combining all of the above results gives us

$$
\begin{aligned}
\frac{4}{nh(n-1)}\sum_{t\in T}\mathcal{U}_t &= \frac{4}{nh(n-1)}\sum_{k=1}^{K}\left[\frac{1}{2}n\hat{p}_k(n\hat{p}_k - 1) - \frac{n(n-1)}{2}p_k^2\right] \\
&= \frac{2n}{h(n-1)}\sum_{k=1}^{K}\left[\hat{p}_k^2 - \frac{p_k}{n} - \frac{n-1}{n}p_k^2\right]
\end{aligned}
$$

Comparing with $\xi_n^{(1)}(h)$ at 6

$$
\xi_n^{(1)}(h) = \frac{4}{nh(n-1)}\sum_{t\in T}\mathcal{U}_t
$$

$\square$

Problem 3. (b) What is the value of the covariance between $\mathcal{U}_t$ and $\mathcal{U}_{t'}$

Proof. If the set (i, j, i', j') are all different, variables $\mathcal{U}_t$,$\mathcal{U}_{t'}$ are independent random variables. By the property of covariance we have:

$$
\begin{aligned}
\mathrm{Cov}(\mathcal{U}_t, \mathcal{U}_{t'}) &= \mathbb{E}[\mathcal{U}_t \times \mathcal{U}_{t'}] - \underbrace{\mathbb{E}[\mathcal{U}_t]\mathbb{E}[\mathcal{U}_{t'}]}_{=0} \\
&= \mathbb{E}[\mathcal{U}_t \times \mathcal{U}_{t'}] = \mathbb{E}[\mathcal{U}_t] \times \mathbb{E}[\mathcal{U}_{t'}] = 0 \times 0 = 0
\end{aligned}
$$

$\square$

Problem 3. (c) Check that for every $t = (i, j) \in T$:

$$\mathbb{E}\left[\left(\sum_{k=1}^{K} \mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j)\right)^2\right] = \sum_{k=1}^{K} \mathbb{E}\left[(\mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j))^2\right] = \sum_{k=1}^{K} p_k^2$$

Proof. To prove the second equality, we make use of indicator function property in exactly similar deduction as in Problem 2(a).

$\mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j) = \begin{cases} 1, & \text{if both } \mathbb{1}_{C_k}(X_i) = 1 \text{ and } \mathbb{1}_{C_k}(X_j) = 1 \\ 0, & \text{otherwise} \end{cases}$

Hence,

$$\mathbb{E}\left[(\mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j))^2\right] = 1 \times p_k^2 + 0 \times (1 - p_k^2) = p_k^2$$

To prove the first equality, we have:

$\left(\sum_{k=1}^{K} \mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j)\right)^2 = \begin{cases} 1^2 = 1, & \text{if both } X_i, X_j \in \text{the same bin } C_k, \forall k \in \{1, ..., K\} \\ 0^2 = 0, & \text{otherwise when } 1 \times 0 = 0 \end{cases}$

$$\text{Prob}(X_i, X_j \in C_k; \forall k \in \{1, ..., K\}) = \sum_{k=1}^{K} \text{Prob}(X_i, X_j \in C_k) = \sum_{k=1}^{K} p_k^2$$

In conclusion,

$$\mathbb{E}\left[\left(\sum_{k=1}^{K} \mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j)\right)^2\right] = \sum_{k=1}^{K} p_k^2$$

□

Problem 3. (d) Let us define the set

$$\mathcal{W} = \{(i, j, i', j') \in T \times T \text{ s.t. } (i, j) \neq (i', j') \text{ and } (i, j) \cap (i', j') \neq \emptyset \tag{8}$$

Check the cardinality of $\mathcal{W}$.

Proof. Since $(i, j) \cap (i', j') \neq \emptyset$ there are three unique elements in the set (i, j, i', j'). Cardinality of set $T = \{(i, j) \in \{1, ..., n\}^2 \text{ s.t. } i < j\}$ is well known to be $\frac{n(n-1)}{2}$. For the other pair (i', j') we only need to find a 3-rd value of index that are different from the index of the pair i, j. There are $(n - 2)$ options for this choice. And this 3-rd value can be paired-up with one of the two indexes from the pair i, j. In total, the number of elements in set $\mathcal{W}$ are $\frac{n(n-1)}{2} \times (n - 2) \times 2 = n(n - 1)(n - 2) \leq n(n - 1)^2$. □

Problem 3. (e) Show that for every pair $(t, t') \in \mathcal{W}$, it holds that

$$\mathbb{E}[\mathcal{U}_t\mathcal{U}_{t'}] \leq \sum_{k=1}^{K} p_k^3 \leq h^2||f||_\infty^2$$

Proof.

$$
\begin{aligned}
\mathcal{U} &= \sum_{k=1}^{K} \left(\mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j) - p_k^2\right) \\
&= \sum_{k=1}^{K} \mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j) - \sum_{k=1}^{K} p_k^2 \\
\mathcal{U}_t\mathcal{U}_{t'} &= \left[\sum_{k=1}^{K} \mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j) - \sum_{k=1}^{K} p_k^2\right] \times \left[\sum_{k=1}^{K} \mathbb{1}_{C_k}(X_{i'})\mathbb{1}_{C_k}(X_{j'}) - \sum_{k=1}^{K} p_k^2\right] \\
&= \sum_{k=1}^{K} \mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j) \times \sum_{k=1}^{K} \mathbb{1}_{C_k}(X_{i'})\mathbb{1}_{C_k}(X_{j'}) \\
&\dots - \left(\sum_{k=1}^{K} p_k^2\right)\left(\sum_{k=1}^{K} \mathbb{1}_{C_k}(X_i)\mathbb{1}_{C_k}(X_j) + \sum_{k=1}^{K} \mathbb{1}_{C_k}(X_{i'})\mathbb{1}_{C_k}(X_{j'})\right) + \underbrace{\left(\sum_{k=1}^{K} p_k^2\right)^2}_{\text{constant}}
\end{aligned}
$$

The first term in the sum takes value 1 when all of the 4 indicator functions are in the same bin. Since there is one common index among $\{(i,j),(i',j')\}$, this probability is $\sum_{k=1}^{K} p_k^3$. Otherwise as this is the product, if any of the 4 indicators function are 0 the term becomes 0. This is the complement case of the case previously, so the probability of this is $1-\sum_{k=1}^{K} p_k^3$. In addition, from previous discussion we have $\mathbb{E}\left(\sum_{k=1}^{K} \mathbb{1}_{C_k}(X_{i'})\mathbb{1}_{C_k}(X_{j'})\right) = \sum_{k=1}^{K} p_k^2$.

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{U}_t\mathcal{U}_{t'}\right] &= \sum_{k=1}^{K} p_k^3 - \left(\sum_{k=1}^{K} p_k^2\right)\left(\sum_{k=1}^{K} p_k^2 + \sum_{k=1}^{K} p_k^2\right) + \left(\sum_{k=1}^{K} p_k^2\right)^2 \\
&= \sum_{k=1}^{K} p_k^3 - \sum_{k=1}^{K} p_k^2 \\
&\leq \sum_{k=1}^{K} p_k^3 \leq h^2 ||f||_\infty^2
\end{aligned}
$$

□

Problem 3. (f) Show that

$$
\mathrm{Var}[\xi_n^{(1)}(h)] = \frac{16}{h^2 n^2 (n-1)^2} \left(\sum_{(t,t')\in\mathcal{W}} \mathbb{E}[\mathcal{U}_t, \mathcal{U}_{t'}] + \sum_{t\in T} \mathbb{E}[\mathcal{U}_t^2] \right). \tag{9}
$$

Deduce from the previous questions that

$$
\mathrm{Var}[\xi_n^{(1)}(h)] \leq \frac{16||f||_\infty^2}{n} + \frac{8||f||_\infty}{hn(n-1)}
$$

Proof. The intersection of set of indices $\{(i,j),(i',j')\}$ can fall into one of the following three cases: no intersection, one common index and both indices. For the case of no intersection, the two variable $\mathcal{U}_t, \mathcal{U}_{t'}$ are independent and the expectation of their product is 0, as proved in Problem 3(b). The expectation in the second and third cases are shown in Problem 3(e) and Problem 3(c) accordingly. From the result in Problem 3(a), we have.

$$
\begin{aligned}
\mathrm{Var}[\xi_n^{(1)}(h)] &\leq \left(\frac{4}{nh(n-1)}\right)^2 \sum_{t\in T} \mathrm{Var}(\mathcal{U}_t) \\
\sum_{t\in T} \mathrm{Var}(\mathcal{U}_t) &= \sum_{t\in T} \mathbb{E}\left[\left(\mathcal{U}_t - \underbrace{\mathbb{E}[U_t]}_{=0}\right)^2\right] = \sum_{t\in T} \mathbb{E}[\mathcal{U}_t^2] \\
&= \sum_{(t,t')\in\mathcal{W}} \mathbb{E}[\mathcal{U}_t, \mathcal{U}_{t'}] + \sum_{t\in T} \mathbb{E}[\mathcal{U}_t^2] + \underbrace{\sum_{t\neq t'} \mathbb{E}[\mathcal{U}_t, \mathcal{U}_{t'}]}_{=0 \text{ (independent)}} \\
&\leq \sum_{(t,t')\in\mathcal{W}} h^2 ||f||_\infty^2 + \sum_{t\in T} \sum_{k=1}^{K} p_k^2 \\
&\leq \sum_{(t,t')\in\mathcal{W}} h^2 ||f||_\infty^2 + \sum_{t\in T} h||f||_\infty \\
&= \mathrm{Card}(W) h^2 ||f||_\infty^2 + \mathrm{Card}(T) h ||f||_\infty \text{ (Cardinality)} \\
&\leq n(n-1)^2 h^2 ||f||_\infty^2 + \frac{n(n-1)}{2} h ||f||_\infty \\
\mathrm{Var}[\xi_n^{(1)}(h)] &\leq \frac{16}{n} ||f||_\infty^2 + \frac{8}{hn(n-1)} h ||f||_\infty
\end{aligned}
$$

□

Problem 4. Show that $||f||_\infty \geq 1$. Explain why it is reasonable to assume that $h > 1/n$ and show that

$$V_n(h)^{1/2} = V_n[\xi_n(h)]^{1/2} \leq \frac{9||f||_\infty}{\sqrt{n}}$$

Partial proof. Since f is the density of P with respect to Lebesgue measure, its integration in the domain must sum up to 1.

$$1 = \int_0^1 f(x)dx \leq \int_0^1 ||f||_\infty dx = ||f||_\infty \int_0^1 dx = ||f||_\infty x|_0^1 = ||f||_\infty$$

Number of bins is $n_{bins} = \frac{1}{h}$. If $n_{bins} > n$ there must be empty bins in the histogram and that is not useful. Therefore, it is reasonable to assume that none of the bins are not empty, or $h > \frac{1}{n}$ or $\frac{1}{hn} \leq 1$.

$$\begin{aligned}
V_n(h) &= V_n[\xi_n(h)] = V_n[\xi_n^{(1)}(h) + \xi_n^{(2)}(h)] \\
&= V_n[\xi_n^{(1)}(h)] + V_n[\xi_n^{(2)}(h)] \text{ (variance property)} \\
&\leq \frac{16||f||_\infty^2}{n} + \frac{8||f||_\infty}{hn(n-1)} + \frac{C||f||_\infty^2}{n} \\
&\leq \frac{(16+C)||f||_\infty^2}{n} + \frac{8||f||_\infty}{n-1} \\
&\leq \frac{(16+C)||f||_\infty^2}{n} + \frac{8||f||_\infty^2}{n(1-1/n)} \text{ (since } ||f||_\infty^2 \geq 1)
\end{aligned}$$

Assume that $n >> 1$ that $\frac{1}{n} \approx 0$

$$\begin{aligned}
V_n(h) = V_n[\xi_n(h)] &\leq \frac{(24+C)||f||_\infty^2}{n} \\
V_n(h)^{\frac{1}{2}} = V_n[\xi_n(h)]^{\frac{1}{2}} &\leq \frac{\sqrt{24+C}||f||_\infty}{\sqrt{n}}
\end{aligned}$$

□