STAT 200B 2019 Week14

Soyeon Ahn

1 Density estimation

$X_1, \ldots, X_n$: iid with probability density funciton f Nonexistence of MLE

$$\max \left\{ \prod f(X_i) : f \geq 0, \int f = 1 \right\} = \infty$$

Note that no element in the class $\left\{ f \geq 0, \int f = 1 \right\}$ attains the maximum.

1.1 Histogram

- B_j: j-th bin
- b_j: width of the B_j. Assume an equal binwidth, $h = \frac{1}{m}$

$$\begin{aligned} &\text{(Area)} = P(X_1 \in B_j) \overset{(i)}{\approx} \text{(bin frequency)}/n \\ &\text{(Area)} \overset{(ii)}{\approx} f(x) b_j \end{aligned}$$

For small b_j, the approximation (i) is bad. For large b_j, the approximation (ii) is bad.

Let v_j denote the number of observations in bin B_j, and define

$$\begin{aligned} \hat{p}_j &= v_j/n \\ p_j &= \int_{B_j} f(u) du \end{aligned}$$

Note that $\hat{p}_j$ is the plug-in estimator of p_j, with $E[\hat{p}_j] = p_j$ and $V[\hat{p}_j] = p_j(1 - p_j)/n$, since v_j follows $Bin(n, p_j)$.

Define the histogram estimator of the density f to be

$$\begin{aligned} \hat{f}_n(x) &= \begin{cases} \hat{p}_1/h & x \in B_1 \\ \hat{p}_2/h & x \in B_2 \\ \vdots & \\ \hat{p}_m/h & x \in B_m \end{cases} \\ &= \sum_{j=1}^{m} \frac{\hat{p}_j}{h} I\{x \in B_j\} \end{aligned}$$

Note that $E(\hat{f}_n(x)) = \frac{p_j}{h} = \frac{1}{h}\int_{B_j} f(u)du = \frac{1}{h}\int_{(j-1)h}^{jh} f(u)du$ and $f(u) - f(x) \approx (u-x)f^{'}(x)$.

[Simple version]

$$\begin{aligned} bias = E(\hat{f}_n(x)) - f(x) &= \frac{1}{h}\int_{(j-1)h}^{jh} f(u)du - f(x) \\ &= \frac{1}{h}\int_{(j-1)h}^{jh} \Big(f(u) - f(x)\Big)du \\ &\approx \frac{1}{h}\int_{(j-1)h}^{jh} (u-x)f^{'}(x)du \\ &= \frac{1}{h}\Big((j-\frac{1}{2})h - x\Big)hf^{'}(x) \\ &= \Big((j-\frac{1}{2})h - x\Big)f^{'}(x) \end{aligned}$$

$$\begin{aligned} V(\hat{f}_n(x)) &= \frac{1}{nh^2}p_j(1-p_j) \approx \frac{1}{nh^2}p_j \\ &= \frac{1}{nh^2}\int_{(j-1)h}^{jh} f(u)du \\ &= \frac{1}{nh^2}\int_{(j-1)h}^{jh} f(x) + (u-x)f^{'}(x)du \\ &= \frac{1}{nh^2}\Big(f(x)h + \Big((j-\frac{1}{2})h - x\Big)hf^{'}(x)\Big) \\ &= \frac{1}{nh}\Big(f(x) + \Big((j-\frac{1}{2})h - x\Big)f^{'}(x)\Big) \\ &\approx \frac{1}{nh}f(x) \end{aligned}$$

Note that the histgoram estimator is consistent when $h \to 0$, and $nh \to \infty$. Small h: large variance but small bias. small h: large bias but small variance.

[Advanced version]

$$\begin{aligned} bias &= \frac{1}{h}\int_{(j-1)h}^{jh} \Big(f(u) - f(x)\Big)du \\ &= \frac{1}{h}\int_{(j-1)h}^{jh} \left((u-x)f^{'}(x) + (u-x)\int_0^1 f^{'}(x + z(u-x)) - f^{'}(x)dz\right)du \\ &= \frac{1}{h}\Big((j-\frac{1}{2})h - x\Big)hf^{'}(x) + \int_0^1\int_{(j-1)h}^{jh} \frac{(u-x)}{h}\{f^{'}(x+z(u-x)) - f^{'}(x)\}dudz \\ &= \Big((j-\frac{1}{2})h - x\Big)f^{'}(x) + \text{second term} = \Big((j-\frac{1}{2})h - x\Big)f^{'}(x) + o(h) \end{aligned}$$

$$
\begin{aligned}
\text{second term} &\le \int_0^1 \int_{(j-1)h}^{jh} \Big|\frac{(u-x)}{h}\Big|\Big|f^{'}(x+z(u-x)) - f^{'}(x)\Big| dudz \\
&\le \int_0^1 \int_{(j-1)h}^{jh} \sup_{|y|<h} |f^{'}(x+y) - f^{'}(x)| dudz \\
&= h \sup |f^{'}(x+y) - f^{'}(x)| = o(h)
\end{aligned}
$$

$$
\begin{aligned}
V(\hat{f}_n(x)) &= \frac{1}{nh^2} \int_{(j-1)h}^{jh} f(u)du \\
&= \frac{1}{nh}\frac{1}{h} \int_{(j-1)h}^{jh} f(u)du \\
&= \frac{1}{nh}(f(x) + o(1)) \\
&= \frac{1}{nh} f(x) + o(\frac{1}{nh})
\end{aligned}
$$

1.2 Mean integrated sqaured error

$$
\begin{aligned}
MISE(\hat{f}_n) = E[L(f, \hat{f}_n)] &= E\int \{f(x) - \hat{f}_n(x)\}^2 dx = \int MSE(\hat{f}_n(x))dx \\
&= \frac{1}{nh} + \frac{1}{12}h^2 \int f^{'2} + o\Big(\frac{1}{nh} + h^2\Big)
\end{aligned}
$$

$$
\begin{aligned}
\int bias^2 dx &= \sum_j \int_{B_j} \Big((j-\frac{1}{2})h - x\Big)^2 f^{'}\Big((j-\frac{1}{2})h\Big)^2 dx + o(h^2) \\
&= \sum_j f^{'}\Big((j-\frac{1}{2})h\Big)^2 \int_{B_j} \Big((j-\frac{1}{2})h - x\Big)^2 dx + o(h^2) \\
&= \sum_j f^{'}\Big((j-\frac{1}{2})h\Big)^2 \Big(2\frac{1}{3}\Big(\frac{h}{2}\Big)^3\Big) + o(h^2) \\
&= \frac{1}{12}h^2 + o(h^2)
\end{aligned}
$$

$$
\int var\{\hat{f}_n(x)\}dx \approx \frac{1}{nh} + o(\frac{1}{nh})
$$

Note that $h^* = \frac{1}{n^{1/3}}\left(\frac{6}{\int [f'(u)]^2 du}\right)^{1/3}$ minimizes the MISE, and MISE $\sim n^{-2/3}$. In parametric approach with $N(\theta, 1)$, for example, MISE of $N(\bar{X}, 1) \sim n^{-1}$