23.1 Kernel density estimators

Suppose we have data $X_1, \cdots, X_n \overset{ind}{\sim} F$ with support on $\mathbb{R}$. Assume $\exists\, p$ such that $F(x) = \int_{-\infty}^{x} p(x)\,\mathrm{d}x$. We want to estimate p, but that is hard. Instead we can estimate F using empirical cdf.

$$F_n(x_0) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x_0) \tag{23.1}$$

By DKW we know

$$\mathbb{P}(\sup_{x_0} |F_n(x_0) - F(x_0)| > \epsilon) \leq 2e^{-2n\epsilon^2} \tag{23.2}$$

For sufficiently small $h > 0$ we can write an approximation

$$\begin{aligned}
p(x_0) &\approx \frac{F(x_0+h) - F(x_0-h)}{2h} && (23.3)\\
&\approx \frac{F_n(x_0+h) - F_n(x_0-h)}{2h} && (23.4)\\
&= \frac{1}{2nh}\sum_{i=1}^{n} I(x_0 - h < x_i \leq x_0 + h) && (23.5)\\
&=: \frac{1}{nh}\sum_{i=1}^{n} K_0\left(\frac{x_i - x_0}{h}\right) && (23.6)
\end{aligned}$$

where $K_0(u) = \frac{1}{2} I(-1 < u \leq 1)$.

A generalization of this estimation is

$$\widehat{p}_n(x_0) =: \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right) \tag{23.7}$$

where $K : \mathbb{R} \to \mathbb{R}$ is an integrable function satisfying $\int K(u)\,\mathrm{d}u = 1$. Such a function K is called a kernel and the parameter h is called a bandwidth of the estimator. $\widehat{p}_n(x_0)$ is called the kernel density estimator (KDE) or the Parzen-Rosenblatt estimator.

Some classical examples of kernels are the following:

1. the Gaussian kernel: $K(u) = \frac{1}{\sqrt{2\pi}} exp(-\frac{u^2}{2})$
2. the Silverman kernel: $K(u) = \frac{1}{2} exp(-\frac{|u|}{2}) \sin(\frac{|u|}{\sqrt{2}} + \frac{\pi}{4})$
3. the Epanechnikov kernel: $K(u) = \frac{3}{4}(1-u)^2 I(|u| \leq 1)$

A few note on the KDE:

1. Truncation is common.
2. If the kernel K takes only nonnegative values, then $\widehat{p}_n$ is a density function.
3. We won't require K takes only nonnegative values.
4. KDE can be generalized to higher dimension.

23.2 Mean squared error of kernel estimators at fixed x_0

Mean squared error of kernel estimators at fixed x_0 is

$$\begin{aligned} MSE(x_0) &= \mathbb{E}\big[(p(x_0) - \widehat{p}_n(x_0))^2\big] && (23.8) \\ &= b^2(x_0) + \sigma^2(x_0) && (23.9) \end{aligned}$$

where $b(x_0)$ is the bias and $\sigma^2(x_0)$ is the variance.

Variance part

Proposition 23.1. *Let $p(x) \leq p_{max} < \infty$ for and x. Suppose $\int K^2(u)\,\mathrm{d}u < \infty$. Then for $n \leq 1$,*

$$\sigma^2(x_0) \leq \frac{p_{max} \int K^2(u)\,\mathrm{d}u}{nh} \tag{23.10}$$

Proof Let

$$\eta_i(x_0) = K(\frac{X_i - x_0}{h}) - \mathbb{E}\Big[K(\frac{X_i - x_0}{h})\Big] \tag{23.11}$$

$$\begin{aligned} \mathbb{E}\big[\eta_i^2(x_0)\big] &\leq \mathbb{E}\Big[K^2(\frac{X_i - x_0}{h})\Big] && (23.12) \\ &= \int K^2(\frac{z - x_0}{h})p(z)\,\mathrm{d}z && (23.13) \\ &\leq p_{max} h \int K^2(u)\,\mathrm{d}u && (23.14) \end{aligned}$$

Thus

$$\begin{aligned} \sigma^2(x_0) &= \mathbb{E}\Big[(\frac{1}{nh}\sum_{i=1}^{n} \eta_i(x_0))^2\Big] && (23.15) \\ &= \frac{1}{nh^2}\mathbb{E}\big[\eta_i^2(x_0)\big] && (23.16) \\ &\leq \frac{C_1}{nh} && (23.17) \end{aligned}$$

23.2.1 Bias part

The bias of the kernel density estimator has the form

$$b(x_0) = \mathbb{E}[\widehat{p}_n(x_0)] - p(x_0) \tag{23.18}$$
$$= \frac{1}{h}\int K(\frac{z-x_0}{h})p(z)\,\mathrm{d}z - p(x_0) \tag{23.19}$$

Definition 23.2. *Let T be an interval in $\mathbb{R}$ and let β and L be two positive numbers. The* ***Hölder class*** *$\Sigma(\beta, L)$ on T is defined as the set of $l = \lfloor \beta \rfloor$ times differentiable functions $f : T \to R$ whose derivative $f^{(l)}$ satisfies*

$$|f^{(l)}(x) - f^{(l)}(x')| \leq L|x - x'|^{\beta - l} \tag{23.20}$$

for $\forall x, x' \in T$.

Definition 23.3. *Let $l \geq 1$ be an integer. We say that $K : \mathbb{R} \to \mathbb{R}$ is* ***a kernel of order*** *l if the functions satisfy $\int K(u)\,\mathrm{d}u = 0$ and $\int u^j K(u)\,\mathrm{d}u = 0$. for $j = 1, \cdots, l$*

Proposition 23.4. *Assume that 1) $p \in \mathbb{P}(\beta, L) = \{p : p \geq 0, \int p = 1, p \in \Sigma(\beta, L)\}$, 2) K is a kernel of order $l = \lfloor \beta \rfloor$ and $\int |u|^\beta |K(u)|\,\mathrm{d}u < \infty$. Then $\forall x_0, h > 0, n \geq 1$*

$$|b(x_0)| \leq \frac{Lh^\beta}{l!}\int |u|^\beta |K(u)|\,\mathrm{d}u = C_2 h^\beta \tag{23.21}$$

Proof Let

$$u = \frac{z - x_0}{h} \tag{23.22}$$

, then $\mathrm{d}z = h\,\mathrm{d}u$ and

$$b(x_0) = \int K(u)[p(x_0 + uh) - p(x_0)]\,\mathrm{d}u \tag{23.23}$$

Since

$$p(x_0 + uh) = p(x_0) + p'(x_0)uh + \cdots + \frac{(uh)^l}{l!}p^{(l)}(x_0 + \tau uh) \tag{23.24}$$

for some $\tau \in [0, 1]$. Then

$$b(x_0) = \int K(u)\frac{(uh)^l}{l!}p^{(l)}(x_0 + \tau uh)\,\mathrm{d}u \tag{23.25}$$
$$= \int K(u)\frac{(uh)^l}{l!}[p^{(l)}(x_0 + \tau uh) - p^{(l)}(x_0)]\,\mathrm{d}u \tag{23.26}$$

Thus

$$|b(x_0)| \leq \int |K(u)|\frac{|uh|^l}{l!}|p^{(l)}(x_0 + \tau uh) - p^{(l)}(x_0)|\,\mathrm{d}u \tag{23.27}$$
$$\leq \int |K(u)|\frac{|uh|^l}{l!}L|\tau uh|^{\beta - l}\,\mathrm{d}u \tag{23.28}$$
$$\leq C_2 h^\beta \tag{23.29}$$

Theorem 23.5. *Under conditions 1) and 2) from Proposition 23.4, let $h = \alpha n^{\frac{1}{2\beta+1}}$, then for $n \geq 1$*

$$\sup_{x_0} \sup_{p \in \mathbb{P}(\beta, L)} \mathbb{E}\big[(\widehat{p}_n(x_0) - p(x_0))^2\big] \leq Cn^{-\frac{2\beta}{2\beta+1}} \tag{23.30}$$

Proof Firstly, we can show from $p \in \mathbb{P}(\beta, L)$ that $p(x_0) \le p_{max}$ for $\forall x_0 \in \mathbb{R}$.

Since

$$MSE \le \frac{C_1}{nh} + C_2 h^{\beta} \tag{23.31}$$

Let

$$h_* = (\frac{C_1}{2\beta C_2^2})^{\frac{1}{2\beta+1}} n^{-\frac{1}{2\beta+1}} \tag{23.32}$$

Then

$$MSE \le C_2^2 h_*^{2\beta} + C_1 n^{-1} h_*^{-1} = C n^{-\frac{2\beta}{2\beta+1}} \tag{23.33}$$