Problem Set 1 - Proposed Solutions

Taric Latif Padovani

October 17, 2025

Problem 1. **(a)**

Since $K(\cdot)$ is a (symmetric second-order) kernel, it satisfies $K(u) \geq 0$ for all u and $\int_{-\infty}^{\infty} K(u)\,du = 1$. Thus, for any fixed x,

$$\hat{f}_n(x) \;=\; \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) \;\geq\; 0,$$

because it is a positive scalar multiple of a sum of nonnegative terms.

Next, compute the integral:

$$\int_{-\infty}^{\infty} \hat{f}_n(x)\,dx = \frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{\infty} K\left(\frac{X_i - x}{h}\right)\,dx.$$

For each i, perform the change of variables $u = \frac{X_i - x}{h}$, so $x = X_i - hu$ and $dx = -h\,du$. The limits remain $\pm\infty$, hence

$$\int_{-\infty}^{\infty} K\left(\frac{X_i - x}{h}\right)\,dx = \int_{\infty}^{-\infty} K(u)\,(-h)\,du = h\int_{-\infty}^{\infty} K(u)\,du = h.$$

Therefore,

$$\int_{-\infty}^{\infty} \hat{f}_n(x)\,dx = \frac{1}{nh}\sum_{i=1}^{n} h = \frac{1}{n}\sum_{i=1}^{n} 1 = 1.$$

Hence, $\hat{f}_n(x) \geq 0$ for all x and $\int_{-\infty}^{\infty} \hat{f}_n(x)\,dx = 1$ for all n.

(b)

Using i.i.d. and linearity of expectation,

$$\mathbb{E}[\hat{f}_n(x)] = \frac{1}{h}\,\mathbb{E}\left[K\left(\frac{X_1 - x}{h}\right)\right] = \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{z - x}{h}\right) f(z)\,dz.$$

With the change of variables $u = (z - x)/h$ (so $z = x + hu$, $dz = h\,du$),

$$\mathbb{E}[\hat{f}_n(x)] = \int_{-\infty}^{\infty} K(u)\,f(x + hu)\,du =$$

$$= f(x) + \int_{-\infty}^{\infty} K(u)\,(f(x + hu) - f(x))\,du$$

where the final equality uses $\int_{-\infty}^{\infty} K(u)du = 1$.

We want to show that the second term is $o(1)$. Fix $\varepsilon > 0$. Since $f(x)$ is continuous in some neighborhood $\mathcal{N}$ there exists a $\delta > 0$ such that $|\nu| \leq \delta$ implies $|f(x+\nu) - f(x)| \leq \varepsilon$. Set $h \leq \delta/a$. Then $|u| \leq a$ implies $|hu| \leq \delta$ and $|f(x+hu) - f(x)| \leq \varepsilon$. Then,

$$\begin{aligned}\left|\mathbb{E}\left[\hat{f}(x) - f(x)\right]\right| &= \left|\int_{-a}^{a} K(u)\left(f(x+hu) - f(x)\right)du\right| \\ &\leq \int_{-a}^{a} K(u)\left|f(x+hu) - f(x)\right|du \\ &\leq \varepsilon \int_{-a}^{a} K(u)\,du \\ &= \varepsilon.\end{aligned}$$

Since ε is arbitrary, this shows that $\left|\mathbb{E}\left[\hat{f}(x) - f(x)\right]\right| = o(1)$ as $h \to 0$, as claimed.

Therefore:

$$\mathbb{E}[\hat{f}_n(x)] = f(x) + o(1)$$

(e) (I will answer this first, the reason will be clear)

From the last item, we have:

$$\mathbb{E}[\hat{f}_n(x)] = \int_{-\infty}^{\infty} K(u)\, f(x+hu)\, du$$

By the mean-value theorem

$$\begin{aligned}f(x+hu) &= f(x) + f'(x)hu + \tfrac{1}{2}f''(x+hu^*)h^2u^2 \\ &= f(x) + f'(x)hu + \tfrac{1}{2}f''(x)h^2u^2 + \tfrac{1}{2}\big(f''(x+hu^*) - f''(x)\big)h^2u^2\end{aligned}$$

where u^* lies between 0 and u. Substituting and using $\int_{-\infty}^{\infty} K(u)u\,du = 0$ and $\int_{-\infty}^{\infty} K(u)u^2\,du = 1$ we find

$$\mathbb{E}\left[\hat{f}(x)\right] \;=\; f(x) + \tfrac{1}{2}f''(x)h^2 + h^2R(h),$$

where

$$R(h) = \tfrac{1}{2}\int_{-\infty}^{\infty}\big(f''(x+hu^*) - f''(x)\big)u^2K(u)\,du.$$

It remains to show that $R(h) = o(1)$ as $h \to 0$. Fix $\varepsilon > 0$. Since $f''(x)$ is continuous in some neighborhood $\mathcal{N}$ there exists a $\delta > 0$ such that $|v| \leq \delta$ implies $|f''(x+v) - f''(x)| \leq \varepsilon$. Set $h \leq \delta/a$. Then $|u| \leq a$ implies $|hu^*| \leq |hu| \leq \delta$ and $|f''(x+hu^*) - f''(x)| \leq \varepsilon$. Then

$$|R(h)| \;\leq\; \tfrac{1}{2}\int_{-\infty}^{\infty}\left|f''(x+hu^*) - f''(x)\right|u^2K(u)\,du \;\leq\; \tfrac{\varepsilon}{2}.$$

Since ε is arbitrary this shows that $R(h) = o(1)$. This completes the proof.

(c)

$$V(\hat{f}_{x_0}) = Var\left(\frac{1}{nh}\sum_i K(\frac{x_i - x_0}{h})\right) = \frac{1}{n^2}Var(\sum_i \frac{1}{h}K(\frac{x_i - x_0}{h}))$$

x_i are i.i.d., therefore

$$... = \frac{1}{n^2}\ n\ Var(\frac{1}{h}K(\frac{x_i - x_0}{h})) = \frac{1}{n}\ Var\left(\frac{1}{h}K(\frac{x_i - x_0}{h})\right)$$

$$... = \frac{1}{n}\ E(\left[\frac{1}{h}K(\frac{x_i - x_0}{h})\right]^2) - \frac{1}{n}\left[\underbrace{E(\frac{1}{h}K(\frac{x_i - x_0}{h}))}_{=E(\hat{f}(x_0))}\right]^2$$

Now consider the first term:

$$E(\left[\frac{1}{h}K(\frac{x_i - x_0}{h})\right]^2) = \int \left[\frac{1}{h}K(\frac{t - x_0}{h})\right]^2 f(t)dt$$

Substitute $z := \frac{t-x_0}{h}$, $\frac{\partial z}{\partial t} = \frac{1}{h}$

$$\int \left[\frac{1}{h}K(\frac{t - x_0}{h})\right]^2 f(t)dt = \int \frac{1}{h}K(z)^2 f(zh + x_0)dz$$

Now Taylor with a first-order expansion $f(zh + x_0) = f(x_0) + f'(x_0)(zh + x_0 - x_0) + O(h^2)$

$$= \int \frac{1}{h}K(z)^2 \left(f(x_0) + f'(x_0)zh + \underbrace{\frac{1}{2}f''(\hat{x})z^2h^2}_{O(h^2)} \right) dz$$

where $\hat{x} \in (x_0, x_0 + zh)$ (I am using the Mean Value Theorem). Break this up to get:

$$E(\left[\frac{1}{h}K(\frac{x_i - x_0}{h})\right]^2) = \frac{1}{h}f(x_0)\int K(z)^2dz + f'(x_0)\int zK(z)^2dz + O(h)$$

Notice that $\frac{1}{h}\frac{1}{2}f''(\hat{x})z^2h^2 = \frac{1}{2}f''(\hat{x})z^2h \to O(h)$.

This is an expression we can plug back into $V(\hat{f}(x_0)) = \frac{1}{n}\ E(\left[\frac{1}{h}K(\frac{x_i - x_0}{h})\right]^2) - \frac{1}{n}\left[E(\hat{f}(x_0)\right]^2$

$$V(\hat{f}(x_0)) = \frac{1}{nh}f(x_0)\int K(z)^2dz + \frac{1}{n}f'(x_0)\int zK(z)^2dz\ +\ \frac{1}{n}O(h)$$
$$- \frac{1}{n}\left[E(\hat{f}(x_0))\right]^2$$

We know from question (f) that $E(\hat{f}(x_0)) = f(x_0) + \frac{1}{2}h^2 f''(x_0) + o(h^2)$ to plug in last term

$$V(\hat{f}(x_0)) = \frac{1}{nh}f(x_0)\int K(z)^2dz + \frac{1}{n}f'(x_0)\int zK(z)^2dz\ +\ \frac{1}{n}O(h)$$
$$- \frac{1}{n}\left[f(x_0) + \frac{1}{2}h^2 f''(x_0) + o(h^2)\right]^2$$

Notice that for $n \to \infty$ and $h \to 0$ all of these terms disappear **except** the first term. We cannot know what happens to nh if $n \to \infty$ and $h \to 0$. But we can check what happens when we assume $nh \to \infty$ (and thus $\frac{1}{nh} \to 0$)

$$V(\hat{f}(x_0)) = \frac{1}{nh} f(x_0) \int K(z)^2 dz + o\left(\frac{1}{nh}\right)$$

For example, take the second term $\frac{1}{n} f'(x_0) \int zK(z)^2 dz$

$$\frac{\frac{1}{n} f'(x_0) \int zK(z)^2 dz}{\frac{1}{nh}} = hf'(x_0) \int zK(z)^2 dz \xrightarrow{h \to 0} 0$$

Or the third term $\frac{1}{n} O(h)$, which is a product with the first factor going to zero for $n \to \infty$ and the second factor going to a constant for $h \to 0$.

(d)

Let us assume $nh \to \infty, n \to \infty, h \to 0$. Note that, form previous results, we know that $V(\hat{f}(x_0))$ and $\mathbb{E}[\hat{f}(x_0)] - f(x_0)$ will converge to zero. we thus have:

$$\lim_{nh \to \infty} \operatorname{Bias}\big(\hat{f}_{nh}(x_0)\big) = \lim_{nh \to \infty} \Big(\mathbb{E}[\hat{f}_{nh}(x_0)] - f(x_0)\Big) = 0,$$

and

$$\lim_{nh \to \infty} \operatorname{Var}\big(\hat{f}_{nh}(x_0)\big) = 0.$$

The mean squared error (MSE) admits the decomposition

$$\operatorname{MSE}\big(\hat{f}_{nh}(x_0)\big) = \mathbb{E}\Big[\big(\hat{f}_{nh}(x_0) - f(x_0)\big)^2\Big] = \operatorname{Var}\big(\hat{f}_{nh}(x_0)\big) + \Big(\operatorname{Bias}\big(\hat{f}_{nh}(x_0)\big)\Big)^2.$$

Hence, $\operatorname{MSE}\big(\hat{f}_{nh}(x_0)\big) \to 0$ as $nh \to \infty$.

For any $\varepsilon > 0$, apply Markov's inequality to the nonnegative random variable $\big(\hat{f}_{nh}(x_0) - f(x_0)\big)^2$:

$$\mathbb{P}\Big(\big|\hat{f}_{nh}(x_0) - f(x_0)\big| > \varepsilon\Big) = \mathbb{P}\Big(\big(\hat{f}_{nh}(x_0) - f(x_0)\big)^2 > \varepsilon^2\Big) \le \frac{\mathbb{E}\Big[\big(\hat{f}_{nh}(x_0) - f(x_0)\big)^2\Big]}{\varepsilon^2} = \frac{\operatorname{MSE}\big(\hat{f}_{nh}(x_0)\big)}{\varepsilon^2}.$$

Since $\operatorname{MSE}\big(\hat{f}_{nh}(x_0)\big) \to 0$, the right-hand side tends to zero, and thus

$$\lim_{nh \to \infty} \mathbb{P}\Big(\big|\hat{f}_{nh}(x_0) - f(x_0)\big| > \varepsilon\Big) = 0.$$

Therefore, $\hat{f}_{nh}(x_0) \xrightarrow{P} f(x_0)$, i.e., the estimator is consistent.

(f)

Local convexity affects the sign and magnitude of the finite-sample bias through the curvature $f''(x)$. Since $R(K) > 0$ for standard symmetric kernels, the leading bias term has the same sign as $f''(x)$:

- **Convex region ($f''(x) > 0$):** The estimator is biased upward; $\mathbb{E}[\hat{f}_h(x)] > f(x)$ to leading order.

- **Concave region ($f''(x) < 0$):** The estimator is biased downward; $\mathbb{E}[\hat{f}_h(x)] < f(x)$ to leading order.
- **Near inflection/flat regions ($f''(x) \approx 0$):** Bias is small (higher-order terms may dominate).

The magnitude of this bias increases with both the local curvature $|f''(x)|$ and the bandwidth h (quadratically in h). Local slope $f'(x)$ does not enter the leading bias because symmetric kernels eliminate first-order terms. Thus, convexity matters precisely through the second derivative: more pronounced convexity/concavity yields larger positive/negative bias, respectively.

Problem 2. (a)

Assume regressors x_i are i.i.d. with density f, the true process is $y_i = m(x_i) + \varepsilon_i$, kernel estimator $\hat{m}(x_0) = \sum_i w_{ioh} y_i$ with:

$$w_{ioh} = \frac{\frac{1}{nh} K(\frac{x_i - x_0}{h})}{\frac{1}{nh} \sum_i K(\frac{x_i - x_0}{h})}.$$

To find the asymptotic distribution of the NW estimator $\hat{m}$, analyze the following expression:

$$\sqrt{nh}(\hat{m}(x_0) - m(x_0)) = \sqrt{nh} \sum_{i=1}^{n} w_{i0h}(m(x_i) - m(x_0) + \varepsilon_i)$$

Plug in definition of w_{i0h}, denominator is identical :

$$\sqrt{nh}(\hat{m}(x_0) - m(x_0)) = \sqrt{nh} \frac{\frac{1}{nh} \sum_{i=1}^{n} K(\frac{x_i - x_0}{h})(m(x_i) - m(x_0) + \varepsilon_i)}{\hat{f}(x_0)} =$$

$$= \frac{1}{\sqrt{nh}} \frac{\sum_{i=1}^{n} K(\frac{x_i - x_0}{h})(m(x_i) - m(x_0) + \varepsilon_i)}{\hat{f}(x_0)}.$$

Break up the numerator:

$$\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} K\left(\frac{x_i - x_0}{h}\right)(m(x_i) - m(x_0)) + \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} K\left(\frac{x_i - x_0}{h}\right)\varepsilon_i.$$

Now let us find the expected value of this term. Notice that the second term is zero in expectation **if** x_i and ε_i are assumed to be independent $cov(x, \varepsilon) = 0$. Focus on the first term:

$$E\left(\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} K\left(\frac{x_i - x_0}{h}\right)(m(x_i) - m(x_0))\right).$$

Remember that x_i i.i.d. :

$$= \frac{1}{\sqrt{nh}} nE\left(K\left(\frac{x - x_0}{h}\right)(m(x) - m(x_0))\right)$$

$$= \frac{\sqrt{n}}{\sqrt{h}} \int K\left(\frac{x - x_0}{h}\right)(m(x) - m(x_0))f(x)dx.$$

Substitute with $z = (x - x_0)/h$, need to expand with $\frac{h}{h}$

$$= \sqrt{nh} \int K(z)(m(x_0 + hz) - m(x_0))f(x_0 + hz)dz$$

Taylor twice! First order at $f(x_0)$, second order at $m(x_0)$:

$$= \sqrt{nh} \int K(z)(m'(x_0)zh + m''(x_0)\frac{1}{2}h^2z^2)(f(x_0) + hzf'(x_0))dz.$$

Now we get 4 terms in the sum. One of them contains h^3, so we ignore it asymptotically. One term contains $\int zK(z)$ which is zero. Remaining two terms:

$$= \sqrt{nh} \left(\int K(z)h^2z^2m'(x_0)f'(x_0)dz + \int K(z)\frac{1}{2}h^2z^2m''(x_0)f(x_0))dz \right)$$

$$= \sqrt{nh}h^2(m'(x_0)f'(x_0) + \frac{1}{2}m''(x_0)f(x_0)) \left(\int z^2K(z)dz \right).$$

Define the term $B(x_0)$ as "the bias term of the kernel regression estimator":

$$B(x_0) := h^2 \left[(m'(x_0)\frac{f'(x_0)}{f(x_0)} + \frac{1}{2}m''(x_0) \right] \int z^2K(z)dz$$

such that we get:

$$E\left[\sqrt{nh}(\hat{m}(x_0) - m(x_0))\right] = \sqrt{nh}\frac{f(x_0)}{\hat{f}(x_0)}B(x_0).$$

If we can be sure that $\hat{f}(x_0) \xrightarrow{p} f(x_0)$, then we have a bias term of order $O(h^2)$ in the limit. $\hat{m}(x_0) \to m(x_0)$ requires $h \to 0$, but also need a large enough "sample size" around x_0. $\hat{m}(x_0) \xrightarrow{p} m(x_0)$ if $h \to 0$ and $nh \to \infty$ with the bias term:

$$B(x_0) = h^2 \left[m'(x_0)\frac{f'(x_0)}{f(x_0)} + \frac{1}{2}m''(x_0) \right] \int z^2K(z)dz.$$

In our case, $m'(x) = \beta$ and $m''(x) = 0$.

(b)

If $\beta > 0$, then $\operatorname{sign} B(x) = \operatorname{sign}\{f'(x)\}$. Hence

$$B(x) \begin{cases} > 0, & \text{where the density is increasing } (f'(x) > 0), \\ < 0, & \text{where the density is decreasing } (f'(x) < 0), \\ = 0, & \text{at stationary points of } f \text{ (e.g., at a mode or anti-mode).} \end{cases}$$

(c)

If $\beta < 0$, the signs reverse:

$$B(x) \begin{cases} < 0, & \text{where } f'(x) > 0, \\ > 0, & \text{where } f'(x) < 0, \\ = 0, & \text{where } f'(x) = 0. \end{cases}$$

(d)

The NW estimator averages nearby Y's with weights proportional to the local density of X around x. If f is increasing at x (more mass to the right than to the left), the weighted average uses relatively more $X_i > x$ observations. When $\beta > 0$ the regression function increases, so those right–of–x observations have larger conditional means $m(X_i)$, pulling the average upward $\Rightarrow$ positive bias. If $\beta < 0$, the same asymmetry pulls the average downward $\Rightarrow$ negative bias. When f is decreasing the argument flips, and at $f'(x) = 0$ the asymmetry vanishes, yielding no (first–order) bias.

Problem 3. **a)** The paper by Karlan and Zinman asks whether demand for microcredit in less-developed economies is truly insensitive to interest rates—a key assumption behind policies urging microfinance institutions to raise rates to reduce subsidies. Their most striking finding is that demand is in fact price-sensitive, with borrowing dropping sharply when rates exceed standard levels, and that loan size responds much more strongly to changes in loan maturity than to interest rates, highlighting the importance of liquidity constraints.

b) **Target object and estimator.** With $Y = \mathbf{1}\{\text{applied} = 1\}$ and $X =$ `offer4`, the target function is the propensity

$$p(x) \equiv \Pr(Y = 1 \mid X = x) \;=\; \mathbb{E}[Y \mid X = x].$$

The Nadaraya–Watson (local-constant) kernel estimator with bandwidth $h > 0$ and symmetric kernel K is

$$\widehat{p}_h(x) = \frac{\sum_{i=1}^n K_h(X_i - x)\, Y_i}{\sum_{i=1}^n K_h(X_i - x)}, \qquad K_h(u) = \frac{1}{h} K\Big(\frac{u}{h}\Big).$$

c) **Gaussian kernel with three bandwidths and plot.** Use the Gaussian kernel $K(u) = \phi(u) = (2\pi)^{-1/2} e^{-u^2/2}$. Let $\hat{\sigma}_X$ be the sample s.d. of X and n the sample size. Define:

$$h_{\text{S}} = 1.06\, \hat{\sigma}_X\, n^{-1/5}, \qquad h_{\text{small}} = 0.30\, h_{\text{S}}, \qquad h_{\text{large}} = 3\, h_{\text{S}}.$$

Evaluate $\widehat{p}_h(x)$ on a grid $x \in [\text{p1}(X), \text{p99}(X)]$ and plot the three curves together.

d) **Epanechnikov kernel, same bandwidths.** Use $K(u) = \frac{3}{4}(1 - u^2)\mathbf{1}\{|u| \le 1\}$ with the same three h's. Compute and overlay the three Epanechnikov curves.

e) **Qualitative comparison.** For fixed h, Epanechnikov (compact support) places zero weight beyond $|u| > 1$, so estimates are more local and often less influenced by outliers/boundaries; Gaussian (unbounded) uses small but nonzero weights far away, yielding slightly smoother tails. Across bandwidths: h_{small} captures more local variation (higher variance, lower bias);

h_{large} is very smooth (lower variance, higher bias). Since Y is binary, all curves lie in $[0, 1]$ and trace the propensity to apply as a function of the offered rate; they should slope downward if lower rates increase applications.

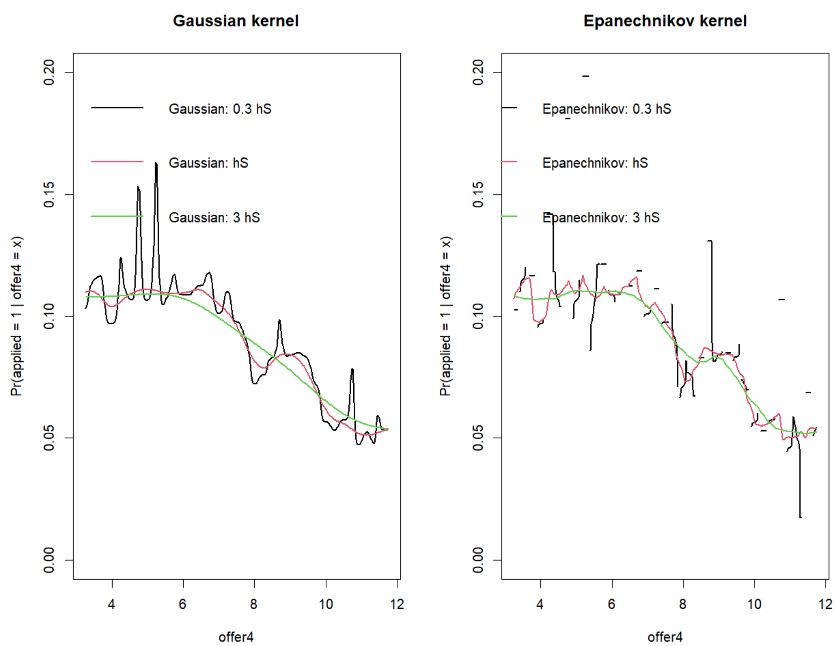
Reproducible code (R)

The following self-contained R code computes the requested estimators and produces the plots. It expects a data frame `df` with columns `applied` (0/1) and `offer4`.

Listing 1: Kernel propensity estimation and plots

```
library(here)
library(readr)

csv_path <- here("problem_sets", "PS1", "karlan.csv")

# Read data
df = read.csv(csv_path)
Y <- df$applied
X <- df$offer4
n <- length(X)
sx <- sd(X)

# Bandwidths

# Bandwidths
hS     <- 1.06 * sx * n^(-1/5)
hSmall <- 0.30 * hS
hLarge <- 3.00 * hS
Hs     <- c(hSmall, hS, hLarge)

# Grids for evaluation
qx <- quantile(X, c(0.01, 0.99), na.rm = TRUE)
xg <- seq(qx[1], qx[2], length.out = 400L)

# Kernels
k_gauss <- function(u) dnorm(u)
k_epan  <- function(u) { w <- pmax(0, 1 - u^2); 0.75 * w }

# Generic NW estimator returning a vector over xg
nw <- function(xgrid, X, Y, h, kfun){
  sapply(xgrid, function(x){
    w <- sum(kfun((X - x)/h))
    w_i <- kfun((X - x)/h)
    if(all(w == 0)) return(NA_real_)
    sum(w_i * Y)/w
  })
```

```
}

# Compute estimates
out_g  <- sapply(Hs, function(h) nw(xg, X, Y, h, k_gauss))
out_e  <- sapply(Hs, function(h) nw(xg, X, Y, h, k_epan))
colnames(out_g) <- c("Gaussian:␣0.3␣hS", "Gaussian:␣hS", "Gaussian:␣3␣hS")
colnames(out_e) <- c("Epanechnikov:␣0.3␣hS", "Epanechnikov:␣hS", "
    Epanechnikov:␣3␣hS")

# Plot helper
plot_many <- function(x, mat, main){
  matplot(x, mat, type = "l", lty = 1, lwd = 2,
          xlab = "offer4", ylab = "Pr(applied␣=␣1␣|␣offer4␣=␣x)",
          main = main, ylim = c(0,0.2))
  legend("topright", legend = colnames(mat), col = 1:ncol(mat),
         lty = 1, lwd = 2, bty = "n")
}

par(mfrow = c(1,2))
plot_many(xg, out_g, "Gaussian␣kernel")
plot_many(xg, out_e, "Epanechnikov␣kernel")
```



Problem 4. Let $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \text{Unif}[0, \theta]$ and let $X_{(1)} \le \cdots \le X_{(n)}$ denote the order statistics. Consider

$$R_n \;=\; n\big(X_{(n)} - \theta\big).$$

(a) $R_n \le 0$**.** Since $X_{(n)} = \max_i X_i \le \theta$ almost surely for Uniform$[0, \theta]$ data, we have $X_{(n)} - \theta \le 0$ a.s., hence $R_n \le 0$ a.s.

(b) Limit cdf of R_n**.** For $x \in \mathbb{R}$,

$$J_n(x) \;:=\; \mathbb{P}(R_n \le x) = \mathbb{P}\Big(X_{(n)} \le \theta + \frac{x}{n}\Big).$$

Using the cdf of the maximum of n i.i.d. Uniform$[0, \theta]$ variables, $\mathbb{P}(X_{(n)} \le t) = \big(t/\theta\big)^n$ for $0 \le t \le \theta$, we obtain the exact finite-n form

$$J_n(x) = \begin{cases} 0, & x < -n\theta, \\ \Big(1 + \dfrac{x}{n\theta}\Big)^n, & -n\theta \le x \le 0, \\ 1, & x \ge 0. \end{cases}$$

Hence, for fixed $x \le 0$,

$$J_n(x) = \Big(1 + \frac{x}{n\theta}\Big)^n \xrightarrow[n\to\infty]{} \exp\Big(\frac{x}{\theta}\Big) \qquad \text{(use } (1 + \tfrac{r}{n})^n \to e^{\,r}\text{)}.$$

For $x > 0$, $J_n(x) = 1$ for all n. Therefore $J_n(\,\cdot\,) \Rightarrow J(\,\cdot\,)$, where

$$J(x) = \mathbb{P}(-\theta X \le x) = \begin{cases} e^{x/\theta}, & x \le 0, \\ 1, & x > 0, \end{cases} \qquad \text{with } X \sim \text{Exp}(1).$$

(Indeed, $-\theta X$ has support $(-\infty, 0]$ and $\mathbb{P}(-\theta X \le x) = \mathbb{P}(X \ge -x/\theta) = e^{x/\theta}$ for $x \le 0$.)

(c) Fix n and let P_n assign mass $1/n$ to n distinct support points in $[0, \theta]$. Write $\theta_n \equiv \theta(P_n)$ for the largest support point and let $X_{(n),n}$ be the maximum of an i.i.d. sample of size n from P_n.

For any $\varepsilon > 0$,

$$\Pr\big(n(X_{(n),n} - \theta_n\big) \le -\varepsilon\big) = \Pr\Big(X_{(n),n} \le \theta_n - \frac{\varepsilon}{n}\Big) \le \Pr\big(X_{(n),n} < \theta_n\big)\,.$$

The last event occurs iff *none* of the n observations equals the top support point θ_n. Since $\Pr(X_i = \theta_n) = 1/n$ and the draws are independent,

$$\Pr\big(X_{(n),n} < \theta_n\big) = \Big(1 - \frac{1}{n}\Big)^n.$$

Hence, for every $\varepsilon > 0$ and all n,

$$\Pr\big(n\big(X_{(n),n} - \theta_n\big) \le -\varepsilon\big) \le \Big(1 - \frac{1}{n}\Big)^n \xrightarrow[n\to\infty]{} e^{-1}.$$

Therefore

$$\limsup_{n\to\infty} \Pr\Big(n\big(X_{(n),n} - \theta(P_n)\big) \le -\varepsilon\Big) \le e^{-1} \qquad \forall\, \varepsilon > 0.$$

(d) Let P be Unif$[0,\theta]$ and for each n let P_n put mass $1/n$ on n distinct points in $[0,\theta]$ with top point $\theta(P_n)$. From the previous item, for every $\varepsilon > 0$,

$$\Pr_{P_n}\Big(n\big(X_{(n),n} - \theta(P_n)\big) \le -\varepsilon\Big) \;=\; \Pr_{P_n}\Big(X_{(n),n} \le \theta(P_n) - \tfrac{\varepsilon}{n}\Big) \;\le\; \Big(1 - \tfrac{1}{n}\Big)^n \xrightarrow[n\to\infty]{} e^{-1}. \tag{1}$$

Let $J_n(x, P_n)$ denote the cdf of the *bootstrapped* root when the true measure is P_n, and let $J(x, P)$ be the target limit under P:

$$J(x,P) = \Pr(-\theta X \le x) = \begin{cases} e^{x/\theta}, & x \le 0, \\ 1, & x > 0, \end{cases} \qquad X \sim \mathrm{Exp}(1).$$

Pick any $\varepsilon \in (0,\theta)$. Then $J(-\varepsilon, P) = e^{-\varepsilon/\theta}$ and $e^{-\varepsilon/\theta} > e^{-1}$. By (1),

$$\limsup_{n\to\infty} J_n(-\varepsilon, P_n) = \limsup_{n\to\infty} \Pr_{P_n}\Big(n\big(X_{(n),n} - \theta(P_n)\big) \le -\varepsilon\Big) \le e^{-1} \;<\; e^{-\varepsilon/\theta} = J(-\varepsilon, P).$$

Hence $J_n(\cdot, P_n)$ does *not* converge to $J(\cdot, P)$; there is a fixed point ($x = -\varepsilon$) where the limit is strictly smaller than the target.

(e) Let P_n be the empirical distribution based on an i.i.d. sample from P. Conditionally on the data, P_n puts mass $1/n$ on n observed points and has maximum equal to θ_n. A bootstrap sample $X_1^*, \ldots, X_n^* \overset{\text{i.i.d.}}{\sim} P_n$ therefore satisfies

$$\Pr\Big(X_{(n)}^* < \theta_n \mid \text{data}\Big) = \Big(1 - \tfrac{1}{n}\Big)^n \to e^{-1},$$

so for any $\varepsilon > 0$,

$$\Pr\Big(n\Big(X_{(n)}^* - \theta_n\Big) \le -\varepsilon \,\Big|\, \text{data}\Big) \le \Big(1 - \tfrac{1}{n}\Big)^n \to e^{-1}.$$

But the true limit cdf at $x = -\varepsilon$ equals $e^{-\varepsilon/\theta} > e^{-1}$ (for $\varepsilon < \theta$). Thus the conditional bootstrap cdf fails to converge to the correct limit. Intuitively, the empirical distribution cannot place mass *above* the sample maximum, so the bootstrap cannot reproduce the correct left tail of the centered-and-scaled maximum; the maximum functional is not smooth, and the bootstrap is inconsistent in this problem.