MATH524 – Spring 2025

Problem Set: Week 2

1. (**Bias-variance trade-off**) Suppose f is a well-defined PDF and we want to estimate $f(0)$. Let $h > 0$ be a small positive number.

   (a) Show that $f(0)$ can be estimated by $\hat{f}_n(0) = X/nh$, where X is the number of observations in an interval of length h that contains 0.
   Hint: Start with estimating the probability $\mathbb{P}(-\frac{h}{2} < X < \frac{h}{2})$. **Solution**: Starting with the hint,

   $$p_h = \mathbb{P}(-\frac{h}{2} < X < \frac{h}{2}) = \int_{-h/2}^{h/2} f(x)\,\mathrm{d}x \approx hf(0).$$

   Thus, $f(0) \approx p_h/h$. Let X be the number of observations in $(-h/2, h/2)$. Then $X \sim \mathrm{Binom}(n, p_h)$. We know that p_h can be estimated by $\hat{p}_h = X/n$. Thus,

   $$\hat{f}_n(0) \approx \frac{X}{nh}$$

   (b) Show that the bias of this estimator takes the form Ah^2 and determine the exact value of A. **Solution**: Again, using the fact that $X \sim \mathrm{Binom}(n, p_h)$, $\mathbb{E}[X] = np_h$. Then, by second-order taylor expansion around 0,

   $$f(x) \approx f(0) + xf'(0) + \frac{x^2}{2}f''(0).$$

   Then,

   $$p_h = \int_{-h/2}^{h/2} f(x)\,\mathrm{d}x \approx \int_{-h/2}^{h/2} \left(f(0) + xf'(0) + \frac{x^2}{2}f''(0).\right)\,\mathrm{d}x = hf(0) + \frac{f''(0)h^3}{24}.$$

   Combining this with the result from part (a),

   $$\mathbb{E}[\hat{f}_n(0)] = \frac{\mathbb{E}[X]}{nh} = \frac{p_h}{h} \approx f(0) + \frac{f''(0)h^2}{24}.$$

   This gives that the bias is

   $$\mathbb{E}[(\hat{f}_n(0) - f(0))] \approx \frac{f''(0)h^2}{24}.$$

   So, $A = f''(0)/24$.

   (c) Show that the variance is of order $(nh)^{-1}$. **Solution**: We start by again using the fact that X is binomially distributed and so have variance $np_h(1-p_h)$. Therefore,

   $$\mathrm{Var}(\hat{f}_n(0)) = \frac{\mathrm{Var}(X)}{n^2h^2} = \frac{p_h(1-p_h)}{nh^2} \approx \frac{p_h}{nh^2}$$

   which follows from the fact that h is small and therefore $1 - p_h \approx 1$. Then,

   $$\mathrm{Var}(\hat{f}_n(0)) \approx \frac{hf(0) + \frac{f''(0)h^3}{24}}{nh^2} = \frac{f(0)}{nh} + \frac{f''(0)h}{24n} \approx \frac{f(0)}{nh}.$$

(d) Sketch the MSE of $\hat{f}_n$ and interpret how the bias and variance change with h. **Solution**: The sketch should resemble a u-curve. This is the classical bias-variance picture that shows how larger values of h reduce bias but increase variance. The MSE optimal bandwidth is then typically chosen to balance the bias and variance, somewhere near the minimum of the MSE curve.

2. (**Scheffé's theorem**) Let (f_n) be a sequence of densities and f be another density such that $f_n \to f$ almost everywhere.

Show that

$$\int_{-\infty}^{\infty} |f_n(x) - f(x)| \, \mathrm{d}x \to 0.$$

Hint: You may choose to prove this result by first considering the integral $g_n = f - f_n$ separately over $\{x : g_n(x) > 0\}$ and $\{x : g_n(x) \leq 0\}$ and using DCT.

Solution: As

$$\begin{aligned}\int_{-\infty}^{\infty} |f_n(x) - f(x)| \, \mathrm{d}x &= \int_{-\infty}^{\infty} g_n(x)\mathbf{1}_{\{g_n(x)\geq 0\}} \, \mathrm{d}x + \int_{-\infty}^{\infty} \{-g_n(x)\}\mathbf{1}_{\{g_n(x)\leq 0\}} \, \mathrm{d}x \\ &= 2\int_{-\infty}^{\infty} g_n(x)\mathbf{1}_{\{g_n(x)\geq 0\}} \, \mathrm{d}x,\end{aligned}$$

since $\int_{-\infty}^{\infty} g_n(x) \, \mathrm{d}x = 0$.

Now $g_n(x)\mathbf{1}_{\{g_n(x)\geq 0\}} \leq \max\{f(x) - f_n(x), 0\} \leq f(x)$, which is integrable, and $g_n \to 0$ almost everywhere by assumption. Thus

$$\int_{-\infty}^{\infty} |f_n(x) - f(x)| \, \mathrm{d}x = 2\int_{-\infty}^{\infty} g_n(x)\mathbf{1}_{\{g_n(x)\geq 0\}} \, \mathrm{d}x \to 0,$$

by the dominated convergence theorem.

3. (**Properties of the local polynomial estimator**) Prove the following two properties of the local polynomial estimator:

(a) $\mathbb{E}[\hat{r}_n(x)] = \sum l_i(x)r(x_i)$ **Solution**: We know that

$$\hat{r}_n(x) = \sum_j^p \hat{\beta}_j x_j = x^\top \hat{\beta} = l(x)^\top Y$$

where $\hat{\beta} = (X^\top X)^{-1}X^\top Y$ and $l(x)^\top = x^\top (X^\top X)^{-1}X^\top$. Taking expectation on both sides,

$$\mathbb{E}[\hat{r}_n(x)] = \mathbb{E}[l(x)^\top Y] = l(x)^\top r(x) = \sum_i l_i(x)r(x_i),$$

by the linearity of expectation and mean-zero noise.

(b) $\mathrm{Var}(\hat{r}_n(x)) = \sigma^2\|l(x)\|^2$ **Solution**:

We already know that $\hat{r}_n(x) = \sum_i l_i(x)Y_i$. The variance follows directly,

$$\mathrm{Var}(\hat{r}_n(x)) = \sigma^2 \sum_i l_i^2(x) = \sigma^2\|l(x)\|^2.$$

4. (**Bias of local polynomial estimators**) Recall that for the regression model

$$y_i = r(x_i) + \varepsilon_i,$$

with $\mathbb{E}[\varepsilon_i|x_i] = 0$ and $\mathbb{E}[\varepsilon_i^2|x_i] = \sigma^2$, the Nadaraya-Watson estimator is a local constant estimator of the form

$$\hat{r} = \frac{\sum_i K\left(\frac{x-x_i}{h}\right) y_i}{\sum_j K\left(\frac{x-x_j}{h}\right)}.$$

For this question we assume K is a second-order kernel. You may assume sufficient smoothness of any functions necessary in evaluating expressions.

(a) Show that the regression function can be written as

$$r(x) = \frac{\int y f(x,y)\,\mathrm{d}y}{f(x)}$$

where $f(x)$ is the marginal density of x_i. **Solution**: This formulation follows directly by defining

$$f(x) = \int f(x,y)\,\mathrm{d}y$$

and

$$f(x,y) = \int K_h(x_i - x)K_h(y_i - y)\,\mathrm{d}y.$$

Suppose that we have some fixed x for which we want to estimate the model. Note that we can equivalently write the regression model as

$$y_i = r(x) + (r(x_i) - r(x)) + \varepsilon_i.$$

(b) Using this reformulation of the regression model, show that we can write the regression function estimator at x as

$$\hat{r}(x) = r(x) + \frac{\hat{m}_1(x)}{\hat{f}(x)} + \frac{\hat{m}_2(x)}{\hat{f}(x)}$$

where $\hat{m}_1$ and $\hat{m}_2$ are functions of r, K and $\{x_i\}$. **Solution**: Taking the reformulation and multiplying by the kernel function and summing over the data on both sides of the equation,

$$\begin{aligned}
&\frac{1}{nh}\sum_i K(\frac{x_i - x}{h})y_i \\
&= \frac{1}{nh}\sum_i K(\frac{x_i - x}{h})r(x) + \frac{1}{nh}\sum_i K(\frac{x_i - x}{h})(r(x_i) - r(x)) + \frac{1}{nh}\sum_i K(\frac{x_i - x}{h})\varepsilon_i \\
&= \hat{f}(x)r(x) + \hat{m}_1(x) + \hat{m}_2(x).
\end{aligned}$$

Dividing by $\hat{f}(x)$ on both sides, we get the desired result.

(c) Compute the mean and variance of $\hat{m}_2$. **Solution**: From the previous part we have

$$\hat{m}_2 = \frac{1}{nh}\sum_i K(\frac{x_i - x}{h})\varepsilon_i.$$

Then,

$$\begin{aligned}
\mathrm{Var}(\hat{m}_2) &= \mathbb{E}\left[\left(\frac{1}{nh}\sum_i K(\frac{x_i - x}{h})\varepsilon_i\right)^2\right] - \mathbb{E}[\frac{1}{nh}\sum_i K(\frac{x_i - x}{h})\varepsilon_i]^2 \\
&= \frac{\sigma^2}{nh^2}\mathbb{E}[K(\frac{x_i - x}{h})^2],
\end{aligned}$$

since by conditional independence of ε_i and x_i,

$$\mathbb{E}\left[\frac{1}{nh}\sum_i K(\frac{x_i - x}{h})\varepsilon_i\right] = 0.$$

Note that this gives us the fact that $\mathbb{E}[\hat{m}_2] = 0$. Now, all that remains is to compute the second moment of the kernel function which we can evaluate directly by change of variables.

$$\begin{aligned}\frac{1}{h}\int K(\frac{z-x}{h})^2 f(z)\,\mathrm{d}z &= \int K(u)^2 f(x+hu)\,\mathrm{d}u \\ &= g(K)f(x) + o(1),\end{aligned}$$

by taylor expansion of of f. Note that we are using $g(K) = \int K^2(u)\,\mathrm{d}u$.

Thus,

$$\mathrm{Var}(\hat{m}_2) = \frac{\sigma^2 g(K) f(x)}{nh} + o((nh)^{-1}).$$

(d) Compute the mean and variance of $\hat{m}_1$. **Solution**: We start with the mean.

$$\begin{aligned}\mathbb{E}[\hat{m}_1] &= \mathbb{E}\left[\frac{1}{nh}\sum_i K(\frac{x_i - x}{h})(r(x_i) - r(x))\right] \\ &= \frac{1}{h}\mathbb{E}\left[K(\frac{x_1 - x}{h})(r(x_1) - r(x))\right] \\ &= \frac{1}{h}\int K(\frac{z-x}{h})(r(z) - r(x))f(z)\,\mathrm{d}z \\ &= \int K(u)(r(x+hu) - r(x))f(x+hu)\,\mathrm{d}u.\end{aligned}$$

Now, simply taylor expanding both r and f,

$$\begin{aligned}&\int K(u)(r(x+hu) - r(x))f(x+hu)\,\mathrm{d}u \\ &= \int K(u)r(x+hu)f(x+hu)\,\mathrm{d}u + \int K(u)r(x)f(x+hu)\,\mathrm{d}u \\ &= \int K(u)(r(x) + hur'(x) + \frac{(hu)^2}{2}r''(x))(f(x) + huf'(x))\,\mathrm{d}u + \int K(u)r(x)(f(x) + huf'(x))\,\mathrm{d}u \\ &= \sigma_K^2 h^2[\frac{r''(x)f(x)}{2} + r'(x)f'(x)] + o(h^2).\end{aligned}$$

Similar calculations should be repeated for the variance, which results in

$$\mathrm{Var}(\hat{m}_1) = O(h/n).$$

Since the variance is of higher order, we have convergence of $\hat{m}_1$ in probability,

$$\sqrt{nh}(\hat{m}_1 - h^2\sigma_K^2 f(x)[\frac{r''(x)f(x)}{2} + r'(x)f'(x)]) \xrightarrow{p} 0.$$

(e) Invoking the CLT, compute the limiting distribution of $\hat{r}$. What is the rate of convergence? Is the estimator asymptotically biased? **Solution**: Yes, the estimator is biased since $\hat{m}_2$ does not converge to 0. Putting together all the results on $\hat{m}_1$ and $\hat{m}_2$, and using the fact that $\hat{f}(x) \xrightarrow{p} f(x)$,

$$\begin{aligned}\sqrt{nh}(\hat{r}(x) - r(x)) &= \sqrt{nh}\frac{\hat{m}_1}{\hat{f}(x)} + \frac{\sqrt{nh}\hat{m}_2}{\hat{f}(x)} \\ &\rightsquigarrow \mathcal{N}\left(h^2\sigma_K^2[\frac{r''(x)f(x)}{2} + r'(x)f'(x)], \frac{g(K)}{f(x)}\right).\end{aligned}$$

The rate of convergence is $\sqrt{nh}$.

(f) How would the parameters of the limiting distribution of $\hat{r}$ change is x were a boundary point? **Solution**: The analysis of $\hat{m}_1$ would change since the estimated derivative of $\hat{f}$ at the boundary will be biased from the data. We will not have the same convergence in probability of $\hat{f}$ and as a result the convergence of $\hat{r}$ would also have higher bias.