1. Let $X_1, \ldots, X_n$ be iid from a distribution F with density f. The likelihood function for f is

$$\mathcal{L}_n(f) = \prod_{i=1}^{n} f(X_i).$$

If the model is set $\mathcal{F}$ of all probability density functions, what is the maximum likelihood estimator of f?

The MLE of f is given by

$$\hat{f}_{MLE} = \underset{\hat{f} \in \mathcal{F}}{\operatorname{argmax}} \mathcal{L}_n(\hat{f})$$

2. Prove equation (4.10). $$R(f(x), \hat{f}_n(x)) = \mathsf{bias}_x^2 + \mathbb{V}_x \tag{4.10}$$

Proof:

$$R(f(x), \hat{f}_n(x)) = \mathbb{E}(f(x) - \hat{f}_n(x))^2$$

$$= \mathbb{E}(f^2(x)) - 2\mathbb{E}(f(x) \cdot \hat{f}_n(x)) + \mathbb{E}(\hat{f}_n^2(x))$$

$$= f^2(x) - 2f(x)\mathbb{E}(\hat{f}_n(x)) + \mathbb{E}(\hat{f}_n^2(x))$$

$$= f^2(x) - 2f(x)\mathbb{E}(\hat{f}_n(x)) + \mathbb{E}(\hat{f}_n^2(x)) - \mathbb{E}^2(\hat{f}_n(x)) + \mathbb{E}^2(\hat{f}_n(x))$$

$$= (\mathbb{E}(\hat{f}_n(x)) - f(x))^2 + \mathbb{E}(\hat{f}_n^2(x)) - \mathbb{E}^2(\hat{f}_n(x))$$

$$= Bia^2 + Var(\hat{f}_n(x))$$

$\square$

3. Let $X_1, \ldots, X_n$ be an IID sample from a $N(\theta, 1)$ distribution with density $f_\theta(x) = (2\pi)^{-1/2} e^{-(x-\theta)^2/2}$. Consider the density estimator $\widehat{f}(x) = f_{\widehat{\theta}}(x)$ where $\widehat{\theta} = \overline{X}_n$ is the sample mean. Find the risk of $\widehat{f}$.

$$\hat{\theta} \sim N(\theta, \tfrac{1}{n})$$

$$g_{\hat{\theta}}(t) = (2\pi/n)^{-\frac{1}{2}} \exp\{-n(t-\theta)^2/2\}$$

Fix any $x \in \mathbb{R}$

$$\mathbb{E}\hat{f}(x) = (2\pi)^{-\frac{1}{2}} \mathbb{E}\left[\exp\{-\tfrac{1}{2}(x-\hat{X})^2\}\right]$$

$$= (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{+\infty} \exp\{-(x-t)^2/2\} \cdot (2\pi/n)^{-\frac{1}{2}} \exp\{-n(t-\theta)^2/2\} dt$$

$$= \frac{\sqrt{n}}{2\pi} \int_{-\infty}^{+\infty} \exp\{-\tfrac{1}{2}[(x-t)^2 + n(t-\theta)^2]\} dt.$$

$$= \frac{\sqrt{n}}{2\pi} \int_{-\infty}^{+\infty} \exp\{-\tfrac{1}{2}[t^2 - 2xt + x^2 + nt^2 - 2n\theta t + n\theta^2]\} dt.$$

$$= \frac{\sqrt{n}}{2\pi} \int_{-\infty}^{+\infty} \exp\{-\tfrac{1}{2}(n+1)[t^2 - 2\frac{(x+n\theta)}{n+1}t + \frac{n\theta^2+x^2}{n+1}]\} dt$$

$$= \frac{\sqrt{n}}{2\pi} \int_{-\infty}^{+\infty} \exp\{-\tfrac{1}{2}(n+1)[t^2 - 2\frac{(x+n\theta)}{n+1}t + \frac{n\theta^2+x^2}{n+1} + \left(\frac{x+n\theta}{n+1}\right)^2$$

$$- \left(\frac{x+n\theta}{n+1}\right)^2]\} dt$$

$$= \frac{\sqrt{n}}{2\pi} \int_{-\infty}^{+\infty} \exp\{-\frac{n+1}{2}(t - \frac{x+n\theta}{n+1})^2\} \cdot \exp\{-\frac{n(x-\theta)^2}{2(n+1)}\} dt$$

$$= \frac{\sqrt{n}}{2\pi} \cdot \exp\{-\frac{n(x-\theta)^2}{2(n+1)}\} \cdot \sqrt{\frac{2\pi}{n+1}} = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{n+1}} \exp\{-\frac{n(x-\theta)^2}{2(n+1)}\}$$

$$\mathbb{E}\hat{f}^2(x) = \frac{1}{2\pi} \mathbb{E}\left[\exp\{-(x-\bar{X})^2\}\right]$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-(x-t)^2} (2\pi/n)^{-\frac{1}{2}} \exp\{-n(t-\theta)^2/2\} dt$$

$$= \frac{\sqrt{n}}{2\pi\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left\{-\frac{n+2}{2}\left(t-\frac{2x+n\theta}{n+2}\right)^2\right\} \cdot \exp\left\{-\frac{n(x-\theta)^2}{n+2}\right\}$$

$$= \frac{\sqrt{n}}{2\pi\sqrt{2\pi}} \cdot \sqrt{\frac{2\pi}{n+2}} \cdot \exp\left\{-\frac{n(x-\theta)^2}{n+2}\right\}$$

$$= \frac{1}{2\pi} \cdot \sqrt{\frac{n}{n+2}} \cdot \exp\left\{-\frac{n(x-\theta)^2}{n+2}\right\}$$

$$MSE = E(\hat{f}(x)-f(x))^2 = E[\hat{f}^2(x)] - 2f_\theta(x)E[\hat{f}(x)] + f_\theta^2(x)$$

$$= \frac{1}{2\pi} \cdot \sqrt{\frac{n}{n+2}} \cdot \exp\left\{-\frac{n(x-\theta)^2}{n+2}\right\} - \frac{1}{\pi}\sqrt{\frac{n}{n+1}} \exp\left\{\frac{-(2n+1)(x-\theta)^2}{2(n+1)}\right\}$$

$$+ \frac{1}{2\pi}\exp\{-(x-\theta)^2\}.$$

4. Recall that the Kullback–Leibler distance between two densities f and g is $D(f,g) = \int f(x)\log\big(f(x)/g(x)\big)dx$. Consider a one-dimensional parametric model $\{f_\theta(x) : \theta \in \mathbb{R}\}$. Establish an approximation relationship between L_2 loss for the parameter θ and Kullback–Leibler loss. Specifically, show that $D(f_\theta, f_\psi) \approx (\theta-\psi)^2 I(\theta)/2$ where θ is the true value, ψ is close to θ and $I(\theta)$ denotes the Fisher information.

$$D(f_\theta, f_\psi) = \int f_\theta(x) \cdot \log(f_\theta(x)/f_\psi(x))\, dx$$

$$= \int f_\theta(x) \cdot [\log f_\theta(x) - \log f_\psi(x)]\, dx$$

By Taylor expansion at θ

$$\log f_\theta(x) - \log f_\psi(x) = (\psi-\theta) \cdot \frac{1}{f_\theta(x)} \cdot \frac{\partial f_\theta(x)}{\partial\theta}$$
$$+ \frac{(\theta-\psi)^2}{2}\left(-\frac{1}{f_\theta^2(x)} \cdot \frac{\partial f_\theta(x)}{\partial\theta} + \frac{1}{f_\theta(x)} \cdot \frac{\partial^2 f_\theta(x)}{\partial\theta^2}\right).$$
$$+ o(\theta-\psi)^2$$

$$D(f_\theta, f_\psi) \simeq (\psi-\theta)\, E_\theta\left[\frac{\partial}{\partial\theta}\ln f_\theta(x)\right] - \frac{(\theta-\psi)^2}{2} \cdot E_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_\theta(X)\right]$$

$\frac{\partial}{\partial\theta} \ln f_\theta(x)$ is score function so has expectation 0

$$-E\left[\frac{\partial^2}{\partial\theta^2} \ln f_\theta(x)\right] = I(\theta)$$

$$\Rightarrow D(f_\theta, f_\psi) \approx \frac{(\theta-\psi)^2}{2} I(\theta)$$

1. In Example 5.24, construct the smoothing matrix L and verify that $\nu = m$.

5.24 Example (Regressogram). Suppose that $a \le x_i \le b$ $i = 1, \ldots, n$. Divide (a, b) into m equally spaced bins denoted by $B_1, B_2, \ldots, B_m$. Define $\widehat{r}_n(x)$ by

$$\widehat{r}_n(x) = \frac{1}{k_j} \sum_{i:x_i \in B_j} Y_i, \quad \text{for } x \in B_j \tag{5.25}$$

where k_j is the number of points in B_j. In other words, the estimate $\widehat{r}_n$ is a step function obtained by averaging the Y_is over each bin. This estimate is called the **regressogram**. An example is given in Figure 4.6. For $x \in B_j$ define $\ell_i(x) = 1/k_j$ if $x_i \in B_j$ and $\ell_i(x) = 0$ otherwise. Thus, $\widehat{r}_n(x) = \sum_{i=1}^n Y_i \ell_i(x)$. The vector of weights $\ell(x)$ looks like this:

$$\ell(x)^T = \left(0, 0, \ldots, 0, \frac{1}{k_j}, \ldots, \frac{1}{k_j}, 0, \ldots, 0\right).$$

To see what the smoothing matrix L looks like, suppose that $n = 9$, $m = 3$ and $k_1 = k_2 = k_3 = 3$. Then,

$$L = \frac{1}{3} \times \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

In general, it is easy to see that there are $\nu = \mathrm{tr}(L) = m$ effective degrees of freedom. The binwidth $h = (b-a)/m$ controls how smooth the estimate is. and the smoothing matrix L has the form ■

$$m \cdot h = b - a$$

$$l_j(x) = \left[0, \ldots, 0, \frac{1}{k_j}, \ldots, \frac{1}{k_j}, 0, 0, \ldots, 0\right] \quad \text{where } k_j = |B_j|$$

$$\sum_{j=1}^m k_j = n.$$

$$\Rightarrow L = \begin{bmatrix} \overbrace{\begin{matrix} \frac{1}{k_1} & \cdots & \frac{1}{k_1} \\ \vdots & & \\ \frac{1}{k_1} & & \frac{1}{k_1} \end{matrix}}^{k_1} \Big\} k_1 & & & 0 \\ & \underbrace{\begin{matrix} \frac{1}{k_2} & \cdots & \frac{1}{k_2} \\ \frac{1}{k_2} & \cdots & \frac{1}{k_2} \end{matrix}}_{k_2} \Big\} k_2 & & \\ 0 & & \ddots & \\ & & & \overbrace{\begin{matrix} k_m & \cdots & k_m \\ k_m & & k_m \end{matrix}}^{k_m} \Big\} k_m \end{bmatrix}$$

$$\nu = \mathrm{tr}(L) = \sum_{j=1}^m k_j \cdot \frac{1}{k_j} = 1 \times m = m.$$

2. Prove Theorem 5.34.

5.34 Theorem. *Let $\widehat{r}_n$ be a linear smoother. Then the leave-one-out cross-validation score $\widehat{R}(h)$ can be written as*

$$\widehat{R}(h) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i - \widehat{r}_n(x_i)}{1 - L_{ii}}\right)^2 \tag{5.35}$$

where $L_{ii} = \ell_i(x_i)$ is the i^{th} diagonal element of the smoothing matrix L.

Proof: Fix any positive real number h.

By definition. $\widehat{R}(h) = \frac{1}{n}\sum_{i=1}^{n}(Y - \widehat{r}_{-i}(x_i))^2$

where $\widehat{r}_{-i}(x_i) = \sum_{j\neq i}\left(\frac{w_j(x) Y_j}{\sum_{k\neq i} w_k(x)}\right) = \frac{\sum_{j=1}^{n} w_j(x_i) Y_j - w_i(x_i) Y_i}{\sum_{j=1}^{n} w_j(x_i) - w_i(x_i)} = \frac{\widehat{r}_n(x_i) - L_{ii} Y_i}{1 - L_{ii}}$

$\Rightarrow \widehat{R}(h) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{r}_{-i}(x_i))^2$

$= \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \frac{\widehat{r}_n(x_i) - L_{ii}Y_i}{1 - L_{ii}}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i - \widehat{r}_n(x_i)}{1 - L_{ii}}\right)^2$

$\square$