6. Density Estimation

1. **Cross Validation**
2. **Histogram**
3. **Kernel Density Estimation**
4. **Local Polynomials**
5. **Higher Dimensions**
6. **Mixture Models**
7. **Converting Density Estimation Into Regression**

6.1 Cross Validation

Suppose we observe $X_1, \ldots, X_n$ from an unknown density f. Our goal is to estimate f nonparametrically. Finding the best estimator $\widehat{f}_n$ in some sense is equivalent to finding the optimal smoothing parameter h.

How to measure the performance of $\widehat{f}_n$?

- Risk/Integrated Mean Square Error(IMSE)
$$R(\widehat{f}_n, f) = \mathsf{E}\left(L(\widehat{f}_n, f)\right)$$
where $L(\widehat{f}_n, f) = \int (\widehat{f}_n(x) - f(x))^2 dx$.
- One can find an optimal estimator that minimizes the risk function:
$$\widehat{f}_n^* = \underset{\widehat{f}_n}{\operatorname{argmin}}\, R(\widehat{f}_n, f),$$
but the risk function is *unknown*!

Estimation of the Risk function

- Use *leave-one-out cross validation* to estimate the risk function.
- One can express the loss function as a function of the smoothing parameter h
$$\begin{aligned} L(h) &= \int (\widehat{f}_n(x) - f(x))^2 dx \\ &= \underbrace{\int (\widehat{f}_n(x))^2 dx - 2\int \widehat{f}_n(x) f(x) dx}_{J(h)} + \int f^2(x) dx \end{aligned}$$
- Cross-validation estimator of the risk function (up to constant)
$$\widehat{J}(h) = \left(\int \widehat{f}_n(x) dx\right)^2 - \frac{2}{n}\sum_{i=1}^{n} \widehat{f}_{(-i)}(X_i)$$
where $\widehat{f}_{(-i)}$ is the density estimator obtained after removing $i^t h$ observation

Finding an optimal estimator $\widehat{f}_n$

- $E(\widehat{J}(h)) = E(J(h)) = R(h) + c$
- The optimal smoothing parameter
$$h^* = \underset{h}{\operatorname{argmin}}\, J(h)$$
- Nonparametric estimator $\widehat{f}$ can be expressed as a function of h and the best estimator $\widehat{f}_n^*$ can be obtained by Plug-in the optimal smoothing parameter
$$\widehat{f}^* = \widehat{f}_{h^*}$$

6.2 Histogram

Histogram Estimator

- Without loss of generality, we assume that the support of f is [0,1]. Divide the support into m equally sized bins

$$B_1 = \left[0, \frac{1}{m}\right), B_2 = \left[\frac{1}{m}, \frac{2}{m}\right), \ldots, B_m = \left[\frac{m-1}{m}, 1\right]$$

- Let $h = \frac{1}{m}$, $p_j = \int_{B_j} f(x)dx$ and $Y_j = \sum_{i=1}^n I(X_i \in B_j)$
- The histogram estimator is defined by

$$\widehat{f}_n(x) = \sum_{j=1}^{n} \frac{\widehat{p}_j}{h} I(x \in B_j)$$

where $\widehat{p}_j = \frac{Y_j}{n}$.

Theorem 1. *Suppose that f' is absolutely continuous and $\int (f')^2 du < \infty$. Then*

$$R(\widehat{f}_n, f) = \frac{h^2}{12}\int (f')^2 du + \frac{1}{nh} + o(h^2) + o\left(\frac{1}{n}\right).$$

The optimal bandwidth is

$$h^* = \frac{1}{n^{1/3}}\left(\frac{6}{\int (f'(u))^2 du}\right)^{1/3}.$$

With the optimal binwidth,

$$R(\widehat{f}_n, f) \approx \frac{C}{n^{2/3}}$$

where $C = \left(\frac{3}{4}\right)^{2/3} \left(\int (f'(u))^2 du\right)^{1/3}$.

Proof. For any $x, u \in B_j$,

$$f(u) = f(x) + (u-x)f'(x) + \frac{(u-x)^2}{2} f''(\widetilde{x}).$$

for some $\widetilde{x}$ between x and u. Hence,

$$\begin{aligned} p_j &= \int_{B_j} f(u)du \\ &= \int_{B_j} \left(f(x) + (u-x)f'(x) + \frac{(u-x)^2}{2} f''(\widetilde{x})\right) du \\ &= f(x)h + hf'(x)\left(h\left(j - \frac{1}{2}\right) - x\right) + O(h^3). \end{aligned}$$

Therefore, the bias of $\widehat{f}_n$ is

$$\begin{aligned} b(x) &\equiv \mathsf{E}\left(\widehat{f}_n(x)\right) - f(x) = \frac{1}{h}\mathsf{E}\left(\widehat{p}_j\right) - f(x) \\ &= \frac{p_j}{h} - f(x) \\ &= \frac{1}{h}\left(f(x)h + hf'(x)\left(h\left(j-\frac{1}{2}\right) - x\right) + O(h^3)\right) - f(x) \\ &= f'(x)\left(h\left(j - \frac{1}{2}\right) - x\right) + O(h^2). \end{aligned}$$

By the mean value theorem, for some $\widetilde{x}_j \in B_j$,

$$\begin{aligned} \int_{B_j} b^2(x)dx &= \int_{B_j} (f'(x))^2 \left(h\left(j-\frac{1}{2}\right) - x\right)^2 dx + O(h^4) \\ &= (f'(\widetilde{x}_j))^2 \int_{B_j} \left(h\left(j-\frac{1}{2}\right) - x\right)^2 dx + O(h^4) \\ &= (f'(\widetilde{x}_j))^2 \frac{h^3}{12} + O(h^4). \end{aligned}$$

Hence

$$\begin{aligned}\int_0^1 b^2(x)dx &= \sum_{j=1}^m \int_{B_j} b^2(x)dx + O(h^3)\\ &= \sum_{j=1}^m (f''(\widetilde{x}_j))^2 \frac{h^3}{12} x + O(h^3)\\ &= \frac{h^2}{12}\int_0^1 (f''(x))^2 dx + o(h^2).\end{aligned}$$

For the variance of $\widehat{f}_n$:

$$v(x) \equiv \mathsf{Var}\left(\widehat{f}_n(x)\right) = \frac{1}{h^2}\mathsf{Var}\left(\widehat{p}_j\right) = \frac{p_j(1-p_j)}{nh^2}.$$

By the mean value theorem, for some $x_j \in B_j$,

$$p_j = \int_{B_j} f(x)dx - hf(x_j).$$

Therefore,

$$\begin{aligned}\int_0^1 v(x)dx &= \sum_{j=1}^m \int_{B_j} v(x)dx = \sum_{j=1}^m \int_{B_j} \frac{p_j(1-p_j)}{nh^2}dx\\ &= \frac{1}{nh^2}\sum_{j=1}^m \int_{B_j} p_j - \frac{1}{nh^2}\sum_{j=1}^m \int_{B_j} p_j^2\\ &= \frac{1}{nh} - \frac{1}{nh^2}\sum_{j=1}^m p_j^2 = \frac{1}{nh} - \frac{1}{nh}\sum_{j=1} h^2 f^2(x_j)\\ &= \frac{1}{nh} - \frac{1}{nh}\left(\int_0^1 f^2(x)dx + o(1)\right) = \frac{1}{nh} + o\left(\frac{1}{n}\right)\end{aligned}$$

□

Theorem 2. *The cross-validation estimator of risk for the histogram is*

$$\widehat{J}(h) = \frac{2}{h(n-1)} - \frac{(n+1)}{h(n-1)}\sum_{j=1}^m \widehat{p}_j^2.$$

Theorem 3. *Let $m = m(n)$ be the number of bins in the histogram $\widehat{f}_n$. Assume that $m(n) \to \infty$ and $m(n) = \log n/n \to 0$ as $n \to \infty$. Define*

$$l_n(x) = \left(max\left\{\sqrt{\widehat{f}_n(x)} - c, 0\right\}\right)^2, \quad u_n(x) = \left(\sqrt{\widehat{f}_n(x)} - c, 0\right)^2$$

where $c = \frac{z_{\alpha/(2m)}}{2}\sqrt{\frac{m}{n}}$. Then

$$\Pr\left(l_n(x) \le \mathsf{E}\left(\widehat{f}_n\right) \le u_n(x)\forall x\right) \ge 1-\alpha.$$

6.3 Kernel Density Estimation

- Given a kernel K and a positive number h, called the **bandwidth**, the **kernel density estimator** is:

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{i=1}^n \frac{1}{h}K\left(\frac{x-X_i}{h}\right).$$

- The choice of kernel K is not crucial but the choice of bandwidth h is important.
- We assume that K satisfies

$$\int K(x)dx = 1, \quad \int xK(x)dx = 0 \text{ and } \sigma_K^2 \equiv \int x^2K(x) > 0.$$

Theorem 4. *Let $R_x = \mathsf{E}\,(f(x) - \widehat{f}(x))^2$ be the risk at point x and let $R = \int R_x dx$ denote the integrated risk. Assume that f'' is absolutely continuous and that $int(f'''(x))^2 dx < \infty$. Also, assume that K satisfies*

$$\int K(x)dx = 1, \quad \int xK(x)dx = 0 \text{ and } \sigma_K^2 \equiv \int x^2 K(x) > 0.$$

Then,

$$\begin{aligned} R_x &= \frac{1}{4}\sigma_K^4 h_n^4 (f''(x))^2 + \frac{f(x)\int K^2(x)dx}{nh_n} + O(n^{-1}) + O(h_n^6), \\ R &= \frac{1}{4}\sigma_K^4 h_n^4 \int (f''(x))^2 dx + \frac{\int K^2(x)dx}{nh_n} + O(n^{-1}) + O(h_n^6). \end{aligned}$$

Proof. Write $K_h(x, X) = \frac{1}{h}K\left(\frac{x-X}{h}\right)$ and $\widehat{f}_n(x) = \frac{1}{n}\sum_{i=1}^n K_h(x, X_i)$.

$$\begin{aligned} \mathsf{E}\,(\widehat{f}_n(x)) &= \mathsf{E}\,(K_h(x, X)) \\ &= \int \frac{1}{h}K\left(\frac{x-t}{h}\right) f(t)dt \\ &= \int K(u) f(x - hu)du \\ &= \int K(u)\left(f(x) - huf'(x) + \frac{h^2u^2}{2}f''(x) + \cdots\right) du \\ &= f(x) + \frac{1}{2}h^2 f''(x) \int u^2 K(u)du. \end{aligned}$$

Hence

$$\text{bias}(\widehat{f}_n(x)) = \frac{1}{2}\sigma_K^2 h_n^2 f''(x) + O(h^4).$$

Similarly,

$$\mathsf{Var}\,(\widehat{f}_n(x)) = \frac{f(x)\int K^2(x)dx}{nh_n} + O(n^{-1}).$$

□

Bandwidth selection

- The optimal smoothing bandwidth is

$$h^* = \left(\frac{c_2}{c_1^2 A(f) n}\right)^{1/5}$$

  where $c_1 = \int x^2 K(x)dx, c_2 = \int (K(x))^2 dx$ and $A(f) = \int (f''(x))^2 dx$.

- The only unknown quantity in h^* is $A(f) = \int (f''(x))^2 dx$
- $A(f) = \int (f''(x))^2 dx = \int (f^{(4)}(x)) f(x) dx = \mathsf{E}\,(f^{(4)})$ where $f^{(r)}$ denote the r^{th} derivative of f

The Normal reference rule

- Assume that f is Normal, one can compute the h^* with the Gaussian kernel.

$$h^* = 1.06\sigma n^{-1/5}$$

- σ is estimated by $\widehat{\sigma} = \min\{s, \text{IQR}/1.34\}$ where s is the sample standard deviation and IQR is the interquartile range.
- The selected bandwidth is:

$$h_n = \frac{1.06\widehat{\sigma}}{n^{1/5}}.$$

Plug-in method

Let $\psi_r = \mathsf{E}\left(f^{(r)}\right)$. To estimate ψ_4 by using the Kernel method, one need to choose the optimal bandwidth which is a functional of ψ_6.

1. Estimate ψ_8 with the bandwidth chosen the normal reference rule.
2. Estimate ψ_6 with the bandwidth depending on $\widehat{\psi}_8$
3. Estimate ψ_4 with the bandwidth depending on $\widehat{\psi}_6$
4. The selected bandwidth is

$$h^* = \left(\frac{c_2}{c_1^2 \widehat{\psi}_4 n}\right)^{1/5}$$

Cross-validation

- Cross-validation score function:

$$\widehat{J}(h) = \int \widehat{f}^2(x)dx - \frac{2}{n}\sum_{i=1}^{n} \widehat{f}_{-i}(X_i)$$

- The selected bandwidth is

$$h_* = \operatorname*{argmin}_h \widehat{J}(h)$$

Theorem 5 (Stone's Theorem). *Suppose f is bounded. Let $\widehat{f}_h$ denote the kernel estimator with bandwidth h and let h_* denote the bandwidth chosen by cross-validation. Then*

$$\frac{\int \left(f(x) - \widehat{f}_{h_*}\right)^2 dx}{\inf_h \int \left(f(x) - \widehat{f}_h\right)^2 dx} \stackrel{a.s.}{\to} 1.$$

Computation

- The cross-validation score function $\widehat{J}(h)$ can be approximated by

$$\widehat{J}(h) = \frac{1}{nh^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K^*\left(\frac{X_i - X_i}{h}\right) + \frac{2}{nh}K(0) + O\left(\frac{1}{n^2}\right)$$

  where $K^*(x) = K^{(2)}(x) - 2K(x)$ and $K^{(2)}K(z-y)K(y)dy$.

- For the computation of $\widehat{f}_n$ and $\widehat{J}(h)$, use ;
  - fast Fourier transform (FFT)
  - binning strategy
- For the details, See Silverman (1981) and Wand and Jones (1995).

Optimal Convergence rate

From Theorem 4, the optimal convergence rate is $O(n^{-4/5})$ if the optimal bandwidth is used.

Theorem 6. *Let $\mathcal{F}$ be the set of all pdfs and let $f^{(m)}$ denote the m^{th} derivative of f. Define*

$$\mathcal{F}_m(c) = \left\{ f \in \mathcal{F} : \int |f^{(m)}(x)|^2 dx \le c^2 \right\}.$$

For any estimator $\widehat{f}_n$,

$$\sup_{f \in \mathcal{F}_m(c)} \mathsf{E}_f \int (\widehat{f}_n(x) - f(x))^2 \ge b n^{-2m/(2m+1)}.$$

where $b > 0$ is a universal constant that depends only on m and c.

6.4 Local Polynomials

Smoothed log-likelihood

- Recall that the nonparametric MLE of the pdf is

$$p = (p_1, \ldots, p_n) \equiv \underset{p}{\operatorname{argmin}}\, \mathcal{L}(p)$$

where

$$\mathcal{L}(p) = \sum_{i=1}^{n} \log p_i - n\left(\sum_{i=1}^{n} p_i - 1\right).$$

- A smoothed version of the log-likelihood at x (up to constant) is

$$\mathcal{L}_x(p) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \log p_i - n \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) p_i.$$

Local log-likelihood

- The kernel density estimator:

$$\widehat{f}_K(x) = \underset{p}{\operatorname{argmax}}\, \mathcal{L}_x(p).$$

- The log-likelihood function of f:

$$\mathcal{L}(f) = \sum_{i=1}^{n} \log f(X_i) - n\left(\int f(u)du - 1\right).$$

- The **local log-likelihood at target value** x is

$$\mathcal{L}_x(f) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \log f(X_i) - n \int K\left(\frac{x - u}{h}\right) f(u)du.$$

Local likelihood density estimator

- Approximate $\log f(u)$ with a polynomial

$$P_x(a, u) = \sum_{j=0}^{p} \frac{a_j}{j!}(x - u)^j.$$

- The **local polynomial log-likelihood** is

$$\mathcal{L}_x(a) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) P_x(a, X_i) - n \int K\left(\frac{u - X_i}{h}\right) e^{P_x(a,u)} du.$$

- The **local likelihood density estimator** is

$$\widehat{f}_n(x) = e^{P_x(\widehat{a}, x)} = e^{\widehat{a}_0},$$

where

$$\widehat{a} = (\widehat{a}_0, \ldots, \widehat{a}_p)^T \equiv \underset{a}{\operatorname{argmax}}\, \mathcal{L}_x(a)$$

6.5 Higher Dimensions

Suppose $\mathbf{X_i} = (X_{i1}, \ldots, X_{id})$. The multivariate kernel estimator is

$$\widehat{f}_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{|\mathbf{H}|} K(\mathbf{H^{-1}}(\mathbf{x} - \mathbf{X})).$$

We often assume a simple form of the bandwidth matrix or kernel. For example, $\mathbf{H}$ =diag$(h_1, \ldots, h_d)$. Then

$$\widehat{f}_n(x) = \frac{1}{n h_1 \cdots n_d} \sum_{i=1}^{d} \left\{ \prod_{j=1}^{d} K\left(\frac{x_j - X_{ij}}{h_j}\right) \right\}$$

- The risk function;

$$R \approx \frac{1}{4}\sigma_K^4 \left[\sum_{j=1}^{d} h_j^4 \int f_{jj}^2(x)dx + \sum_{j \neq k} h_j^2 h_k^2 \int f_{jj} f_{kk} dx \right] + \frac{(\int K^2(x)dx)^d}{nh_1 \cdots h_d}.$$

- The optimal bandwidth $h_i^* = O(n^{-1/(4+d)})$.
- In practice, one can choose the optimal bandwidth by cross-validation and often assume a simple form of the bandwidth matrix. $\mathbf{H} = h \cdot I$.
- The curse of dimensionality.

6.6 Mixture Models

The Normal Mixture Model

$$f(x, \theta) = \sum_{j=1}^{K} p_j \phi(x|\mu_j, \sigma_j^2),$$

where $\phi_j(x|\mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}}$.

- Define $\theta = (p_1, \ldots, p_K, \mu_1, \ldots, \mu_K, \sigma_1^2, \ldots, \sigma_K^2)$
- Given K, how to estimate θ?
  $\rightarrow$ EM algorithm

EM algorithm

For $K = 2$,

- Define latent variables Z_i:

$$Z_i = \begin{cases} 1 & : \quad \text{if } Y_i \text{ is from group 1} \\ 0 & : \quad \text{if } Y_i \text{ is from group 2} \end{cases}$$

- The log-likelihood function is

$$\begin{aligned} \mathcal{L}(y, z, \theta) &= \sum_{i:z_i=0} \log[p\phi(y_i|\mu_1, \sigma_1^2)] + \sum_{i:z_i=1} \log[(1-p)\phi(y_i|\mu_2, \sigma_2^2)] \\ &= \sum_{i=1}^{n} \log[p\phi(y_i|\mu_1, \sigma_1^2)] + (1 - z_i) \log[(1-p)\phi(y_i|\mu_2, \sigma_2^2)] \end{aligned}$$

- [E-step] Given θ, compute $\mathsf{E}(Z_i|Y, \theta)$.

$$E(Z_i|Y = y_i, \theta) \equiv w_i = \frac{\widehat{p}\phi(y_i)}{\widehat{\phi}(y) + (1 - \widehat{p})\phi(y)}$$

- [M-step]

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmax}}\, \widetilde{\mathcal{L}}(\theta),$$

  where

$$\widetilde{\mathcal{L}}(\theta) = \sum_{i=1}^{n} (w_i \log \widehat{p}\phi(y_i) + (1 - w_i) \log(1 - p)\phi(y_i)).$$

  Then

$$\begin{aligned} \widehat{p} &= \sum_{i=1}^{n} \frac{w_i}{n}, \quad \widehat{\mu}_1 = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}, \quad \widehat{\mu}_2 = \frac{\sum_{i=1}^n (1 - w_i) y_i}{\sum_{i=1}^n (1 - w_i}, \\ \widehat{\sigma}_1^2 &= \frac{\sum_{i=1}^n w_i (y_i - \widehat{\mu}_1)^2}{\sum_{i=1}^n w_i}, \quad \widehat{\sigma}_2^2 = \frac{\sum_{i=1}^n (1 - w_i)(y_i - \widehat{\mu}_1)^2}{\sum_{i=1}^n (1 - w_i)}. \end{aligned}$$

How to estimate the number of components k?

- AIC (Akaike Information Criterion): find most predictive model

$$\text{AIC} = \mathcal{L} - q$$

  where $\mathcal{L}$ is the loglikelihood and p is the number of parameters.

- BIC (Bayesian Information Criterion): find the true model with high probability

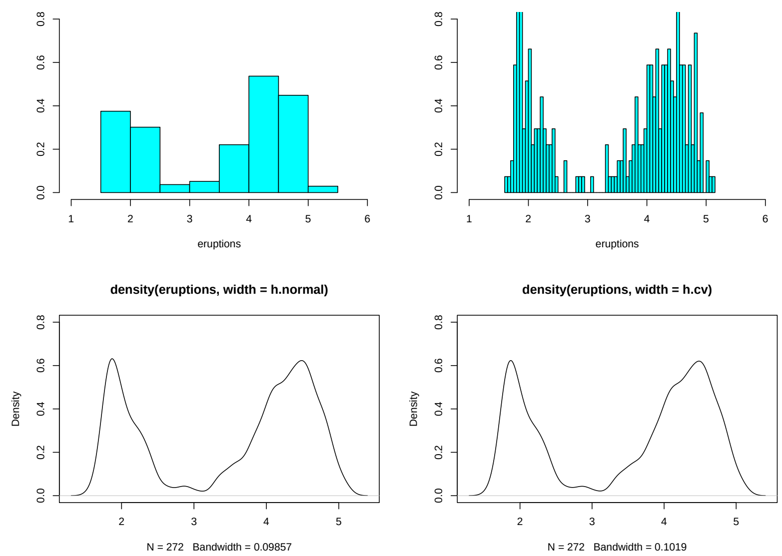
$$\text{BIC} = \mathcal{L} - \frac{q \log n}{2}$$

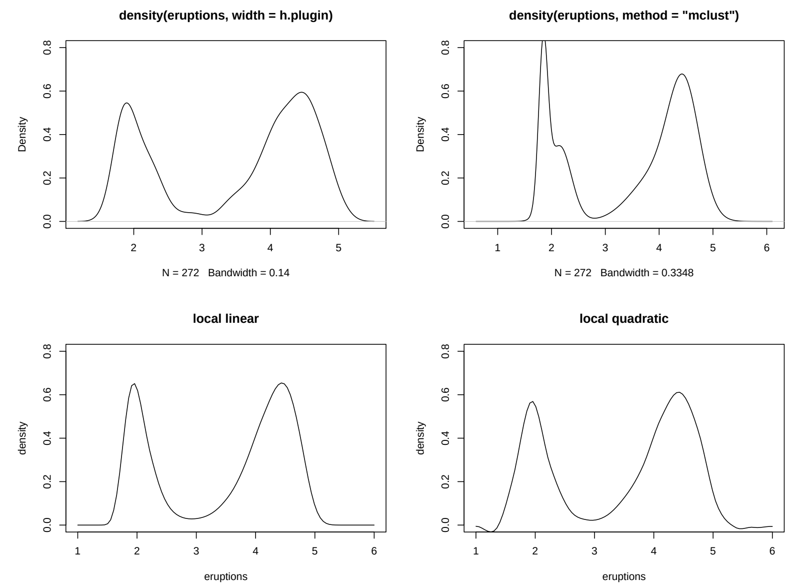
- The number of clusters is not always equal to the number of component
- Gaussian is sensitive to outliers → Replace the normal density with t-distribution density.
- If $\sigma_j^2 = \sigma^2 > 0$ is fixed and $K \to n$, then MLE of the mixture model approaches the kernel estimate where $p_j = \frac{1}{n}$ and $\widehat{\mu}_j = x_j$.


```
source("ch6.r")
library(MASS)
library(sm)
library(locfit)
library(mclust)

### Read data
faithful<-read.table("faithful.dat",header=T)
eruptions <-faithful$eruptions
n<-length(eruptions)

### Select the optimal number of bins by CV
hist.h  <- cv.hist.fun(eruptions)$mbest

### Select optimal bandwidth of kernel estimators by normal
### reference rule/Cross-validation/plug-in

sigma.hat <-min(sd(eruptions), IQR(eruptions)/1.34)
h.normal <-1.06*sigma.hat/n^(0.2)
```



```
h.cv   <- ucv(eruptions)
h.plugin <-width.SJ(eruptions)

f1<-density(eruptions,width=h.normal) # normal reference
f2<-density(eruptions,width=h.cv) # cross-validation
f3<-density(eruptions,width=h.plugin) # plugin
f4<-density(eruptions,method="mclust") # finite mixture model
f5 <-locfit(~eruptions, alpha=c(0.1,0.8),flim=c(1,6))
f6 <-locfit(~eruptions, alpha=c(0.1,0.6),flim=c(1,6), link="ident")

postscript("density.ps")
par(mfrow=c(2,2))
truehist(eruptions, xlim=c(1,6), ymax=0.8) # histogram
truehist(eruptions, nbins=hist.h, xlim=c(1,6), ymax=0.8)
plot(f1,ylim=c(0,0.8));plot(f2,ylim=c(0,0.8))
plot(f3,ylim=c(0,0.8));plot(f4,ylim=c(0,0.8))
plot(f5,ylim=c(0,0.8),main="local linear")
plot(f6,ylim=c(0,0.8),main="local quadratic")
dev.off()
```





6.7 Converting Density Estimation Into Regression

Description of Methodette

1. Suppose that $X_1, \ldots, X_n$ are from a density f on $[0,1]$
2. Create k equal width bins where $k \approx n/10$
3. Define

$$Y_j = \sqrt{\frac{k}{n}} \times \sqrt{N_j + \frac{1}{4}},$$

where $N_j = \sum_{i=1}^{n} I(X_i \in B_j)$ and B_j is the j^{th} bin.

4. Let $r(x) = \sqrt{f(x)}$ and t_j is the mid point of B_j. Then

$$Y_j \approx r(t_j) + \sigma\epsilon_j,$$

where $\epsilon_j \sim N(0,1)$ and $\sigma = \sqrt{\frac{k}{4n}}$.

5. Apply your favorite nonparametric regression method to (t_j, Y_j) to get an estimate $\widehat{r}_n$.
6. Calculate $(\widehat{r}_n(x))^2$ and normalize to be a density.

$$\widehat{f}_n(x) = \frac{(\widehat{r}^+(x))^2}{\frac{1}{k}\sum_{j=1}^{k}(\widehat{r}^+(t_j))^2}$$

where $\widehat{r}^+(x) = \max\{\widehat{r}_n(x), 0\}$.

How does it work?

- Poissonization

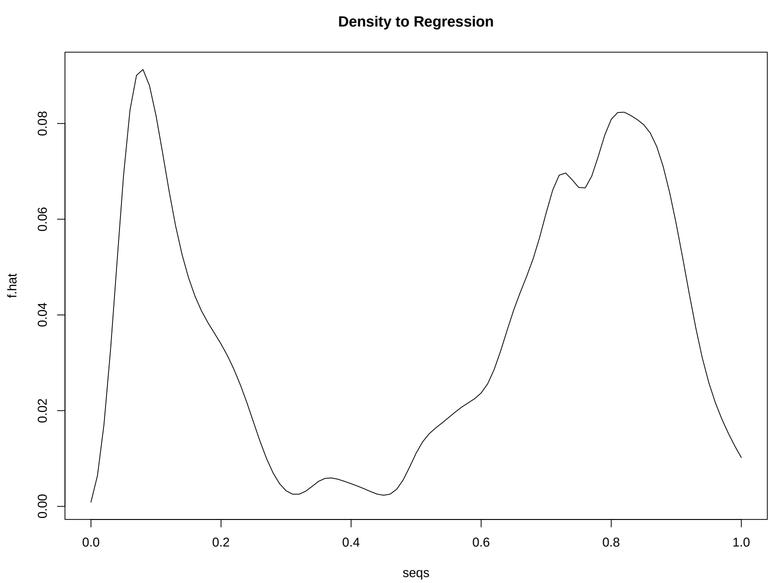
$$N_j \approx \text{Poisson}\left(n \int_{B_j} f(x)dx\right) \approx \text{Poisson}\left(\frac{nf(t_j)}{k}\right)$$

- $\mathsf{E}\,(N_j) = \mathsf{Var}\,(N_j) \approx nf(t_j)/k$.
- Variance stabilization

$$Y_j = \sqrt{\frac{k}{n}} \times \sqrt{N_j + \frac{1}{4}},$$

- $\mathsf{E}\,(Y_j) \approx \sqrt{f(t_j)}$ and $\mathsf{Var}\,(Y_j) \approx k/(4n)$


```
source("ch6.r")
library(sm)

temp <-den.to.reg(eruptions)
temp2 <-smooth.spline(temp$x,temp.y,cv=TRUE)
temp3<-temp2$y
temp3[temp3 <0] <-0
k <-length(temp3)
normal.const<-sum(temp3^2)/k
seqs <-seq(0,1,by=0.01)
temp4<-predict(temp2,seqs)$y
temp4[temp4 <0] <-0
f.hat <-(temp4)^2/normal.const*k
postscript("den_reg.ps")
plot(seqs, f.hat,main="Density to Regression")
dev.off()
```

REFERENCES

1. Brown, L.D., Cai, T. and Zhou, L. (2005). A root-unroot transform and wavelet block thresholding approach to adaptive density estimation. Unpublished manuscript.
2. Loader, C. (1999). *Local regression and likelihood.* Springer, New York.
3. McLachlan, G. and Peel, D. (2000). *Finite Mixture Models.* Wiley, New York.
4. Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* Wiley, New York.
5. Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall. London.
6. Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing.* Chapman and Hall. London.