2.2, 2.3 Kernel Density Estimation

Saturday, February 5, 2022 2:13 PM

2.2 Kernel Density Estimation

Alternative moving histogram:

$$\hat{f}_N(x;h) = \frac{1}{2nh}\sum_{i=1}^n 1_{\{x-h < X_i < x+h\}}$$

$$= \frac{1}{nh}\sum_{i=1}^n \frac{1}{2} 1_{\{-1 < \frac{x-X_i}{h} < 1\}}$$

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

- note that K defn w/ uniform density on $(-1,1)$

Defn: Kernel Density Estimator (kde):

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

notation: $K_h(z) := \frac{1}{h}K\left(\frac{z}{h}\right)$

$$\Rightarrow \hat{f}(x;h) = \frac{1}{n}\sum_{i=1}^n K_h(x-X_i)$$

normal kernel

Say $K_h(x-X_i) = \phi_h(x-X_i)$, $K \sim N(X_i, h^2)$

- bandwidth $h \sim \sigma$ of normal w/ mean X_i
- kde inherits smoothness props of kernel
  - ↳ $\phi(z)$ is infinitely differentiable

Defn: second-order moment of K:

$$\mu_2(K) := \int z^2 K(z)\,dz$$

After some derivation, see that:
unnormalized $\tilde{K}$ w/ h has
associated scaled bandwidth $\tilde{h}$
for unnormalized K

2.3 Asymptotic Properties

Baseline assumptions

1) density f sq integrable, 2x diff and 2nd deriv sq integrable
2) K is symmetric & bounded pdf w/ finite 2nd moment & sq integrable
3) $h = h_n$ => deterministic seq of pos scalars s.t. when $n \to \infty$, $h \to 0$, and $nh \to \infty$

Recall:

1) $\ldots$ $R(f) := \int f(x)^2 dx$

Recall:

1) sq int: $R(f) := \int f(x)^2 dx$
2) conv: $(f * g)(x) := \int f(x-y) g(y) dy = (g * f)(x)$

Note: binomial trick no longer applies b/c not piecewise constant

$$\mathbb{E}[\hat{f}(x;h)] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[K_h(x - X_i)]$$
$$= \int K_h(x-y) f(y) dy$$
$$= (K_h * f)(x)$$

$$\mathbb{V}ar[\hat{f}(x;h)] = \frac{1}{n}\left((K_h^2 * f)(x) - (K_h * f)^2(x)\right)$$

exact! but hard to interpret

Proofs omitted here, but results for bias/variance of the kde @ x

$$\text{Bias}[\hat{f}(x;h)] = \frac{1}{2}\mu_2(K) f''(x) h^2 + o(h^2)$$
$$\mathbb{V}ar[\hat{f}(x;h)] = \frac{R(K)}{nh} f(x) + o(1/nh)$$

asymptotic