Computational Statistics

Dec 01, 2021

Let $\hat{f}$ be the kernel density estimator :

Thm Let $R(x) = \mathbb{E}\left(\left(f(x) - \hat{f}(x) \right)^2 \right)$ be the risk at x. Then

$$R(x) = \frac{1}{4} \sigma_k^4 h^4 f''(x)^2 + \frac{f(x)}{nh} \int K^2(x)\, dx + O\left(\frac{1}{n}\right) + O(h^6).$$

Proof : $\mathbb{E}(\hat{f}(x)) = \mathbb{E}\left(\frac{1}{n} \sum_i \frac{1}{h} K\left(\frac{x-x_i}{h}\right)\right)$

$$= \mathbb{E}\left(\frac{1}{h} K\left(\frac{x-x_i}{h}\right)\right)$$

$$= \frac{1}{h} \int K\left(\frac{x-t}{h}\right) f(t)\, dt$$

$$= \int K(u) f(x-hu)\, du$$, expand f around $hu = 0$.

$$= \int K(u) \left(f(x) - \cancel{hu f'(x)} + \frac{h^2u^2}{2} f''(x) + \dots \right) du$$

$$= f(x) + \frac{1}{2} h^2 f''(x) \underbrace{\int u^2 K(u)\, du}_{\sigma_k^2} + \dots$$

assuming K is even.

$\Rightarrow$ bias $= \mathbb{E}\left(\hat{f}(x) - f(x) \right) = \frac{1}{2} h^2 \sigma_k^2 f''(x) + O(h^4)$

Similarly, $\mathrm{var}(\hat{f}(x)) = \dfrac{f(x)}{nh} \int K^2(u)\, du. + O\left(\frac{1}{n}\right).$

$\Rightarrow$ R = bias^2 + variance

Then for the optimal bandwidth,

$$\text{soln. } \frac{dR}{dh} = 0 \quad \Rightarrow \quad h \sim \frac{\sigma_K^2}{n^{1/5}}$$

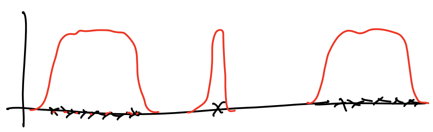
$$\Rightarrow \quad R \sim \mathcal{O}\left(\frac{1}{n^{4/5}}\right)$$

vs for the histogram : $R \sim \mathcal{O}\left(\frac{1}{n^{2/3}}\right)$

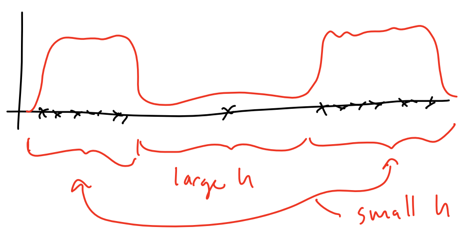
<u>Note</u> Assuming only that $\int (f'')^2 < \infty$, the rate of $\frac{1}{n^{4/5}}$ is <u>the best</u> that can be obtained

<u>Adaptive Methods</u> Choose h locally depending on clustering of data and other considerations.

<u>Ex:</u>



uniformly small h



<u>Multivariate Version</u>

Same mathematical idea as in the one-dimensional case.

One option is to use what is known as a product kernel:

$$K_h^d \underbrace{(x_1, \ldots, x_d)}_{\vec{x}} = \prod_{j=1}^{d} \frac{1}{h_j} K\left(\frac{x_j - x_j'}{h_j}\right)$$

K_h^d must satisfy kernel properties.

then $\hat{f}(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n} K_h^d(\vec{x}-\vec{x}_i)$

$= \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\frac{1}{h_j}K\left(\frac{x_j - x_{ij}}{h_j}\right)$

Risk can be estimated in the same way using multivariable Taylor series for f.

<u>Curse of Dimensionality</u>

If we want the risk $R \sim 0.1$ at $\vec{x}=0$ for $f \sim \text{Normal}(0,1) \in \mathbb{R}^d$, using the optimal bandwidth then $n \sim c^d$:

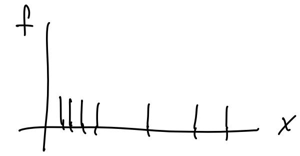
d	n
1	4
2	19
4	223
8	43,700
9	187,000
10	842,000

Resampling Techniques

First: Empirical distribution function

Let $x_1, \dots, x_n$ be samples from some unknown distribution F.
Then $\hat{f}_n$, $\hat{F}_n$ are the empirical PDF and CDF:

$$\hat{f}_n(x) = \begin{cases} 1/n & , \text{ if } x = x_i \quad (\text{assume } x_i \neq x_j \text{ if } i \neq j) \\ 0 & , \text{ otherwise} \end{cases}$$

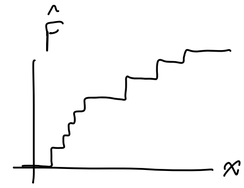


And then

$$\hat{F}(x) = \int_{-\infty}^{x} \hat{f}(x)\, dx$$

$$= \int_{-\infty}^{x} d\hat{F}(x)$$

$$= \frac{m}{n} \quad \text{where } m = \#\ x_i \leq x.$$



Ex: If $\hat{X} \sim \hat{f}$, then $E(\hat{X}) = \int x \, d\hat{F}(x) = \sum_{i=1}^{n} x_i \frac{1}{n}$

The Jackknife

Goal: Partition observations in order to estimate properties of an estimator.

Suppose we have a random sample $y_1, \dots, y_n$, and we compute a statistic $T = T(y_1, \dots, y_n)$ to estimate θ, a parameter. (E.g. $T = \frac{1}{n} \sum y_i$, $\approx E(Y)) = \theta$.

1. Partition data set into $\underline{r}$ groups, each of size $\underline{k}$ $(n = kr)$.

2. Remove j^{th} group and re-compute the estimator from the remaining samples: call it $T_{(-j)}$

   E.g. $T_{(-j)} = \frac{1}{n-k} \sum y_{i(j)}$

T and $T_{(-j)}$ will have similar properties: bias, etc.

<u>Another</u> estimate for θ is the average of the $T_{(-j)}$'s:

$$\hat{\theta} = \overline{T_{(\cdot)}} = \frac{1}{r}\sum_{j=1}^{r} T_{(-j)}.$$

Why do this? The original T gives only one value, and therefore $\mathrm{Var}(T)$ can be hard to estimate. By subsampling, we are generating approximate samples and can compute sample variances.

<u>Consider</u>: the weighted difference:

$$T_j^* = rT - (r-1)T_{(-j)}$$

pseudovalues

The "jackknifed" T is then

$$J(T) = \frac{1}{r}\sum_{j=1}^{r} T_j^*$$

$$= \overline{T}^*$$

Often it is optimal to take $k=1$ and $r=n$ (similar to leave-one-out cross validation)

<u>Variance Estimation</u>

The T_j^*'s are <u>not</u> independent, but if we treat them independently we can use $\mathrm{Var}(J(T))$ as an estimate for $\mathrm{var}(T)$:

$$\hat{\mathrm{var}}(J(T)) = \underbrace{\frac{1}{r(r-1)}\sum_{j=1}\left(T_j^* - J(T)\right)^2}_{\text{sample variance}}.$$

Bias Correction

Recall the bias of an estimator is

$$\text{bias}(T) = E(T) - \theta \qquad (*)$$

Assume this can be written as

$$= \sum_{q=1}^{\infty} \frac{a_q}{n^q} \qquad \text{(constant bias can be easily fixed)}$$

$$= \frac{a_1}{n} + \frac{a_2}{n^2} + \cdots$$

What is the bias of $J(T)$ if $(*)$ is true? (assume $k=1$)

$$\begin{aligned}
\text{Bias}(J(T)) &= E(J(T)) - \theta \\
&= n\left(E(T) - \theta\right) - \frac{n-1}{n}\sum_{j=1}^{n} E\left(T_{(-j)} - \theta\right) \\
&= n\sum_{1}^{\infty}\frac{a_q}{n^q} - (n-1)\sum_{1}^{\infty}\frac{a_q}{(n-1)^q} \\
&= a_2\left(\frac{1}{n} - \frac{1}{n-1}\right) + a_3\left(\frac{1}{n^2} - \frac{1}{(n-1)^2}\right) + \cdots \\
&= -a_2\left(\frac{1}{n(n-1)}\right) + \cdots \quad = O\left(\frac{1}{n^2}\right) \qquad \text{A reduction!}
\end{aligned}$$

Bootstrap (Ch 13 in Gentle)

Goal: Estimate standard errors and confidence sets for statistics.

Outline: Given $X_1, \ldots, X_n \sim F$, statistic $T = g(X_1, \ldots, X_n)$, want $Var_F(T)$.

$\quad$ ← dependence on unknown distribution F.

Ex: $T = \bar{X}$

$$Var_F = \frac{\sigma^2}{n} \quad \text{if } var X_i = \sigma^2 = \int (x-\mu)\, dF(x)$$

function of F.

The idea of the bootstrap:

① Estimate $\mathrm{Var}_F(T)$ with $\mathrm{Var}_{\hat{F}}(T)$.

② Use simulation to approximate $\mathrm{Var}_{\hat{F}}(T)$.

In this example, step 2 is not needed because

$$\mathrm{Var}_{\hat{F}}(T) = \frac{\hat{\sigma}^2}{n} \quad \text{where} \quad \hat{\sigma}^2 = \frac{1}{n}\sum (X_i - \bar{X})^2 .$$

What is simulation? Drawing samples from some distribution, and computing averages.

<u>Ex</u>: Draw $Y_1, \dots, Y_m$ from a distribution G, by the law of large numbers

$$\bar{Y} = \frac{1}{m}\sum_{j=1}^{m} Y_j \xrightarrow{\mathbb{P}} \mathbb{E}(Y) = \int y\, dG(y) \quad \text{as } m \to \infty .$$

Choosing m large enough means that $\bar{Y} \approx \mathbb{E}(Y)$, use this as an estimate for $\mathbb{E}(Y)$.

Also if h is some function with $\int h(y)\, dy < \infty$, then

$$\frac{1}{m}\sum h(Y_j) \xrightarrow{\mathbb{P}} \mathbb{E}(h(Y)) = \int h(y)\, dG(y).$$

<u>Ex</u>:

$$\frac{1}{m}\sum (Y_i - \bar{Y})^2 = \frac{1}{m}\sum Y_i^2 - \left(\frac{1}{m}\sum Y_i\right)^2$$

$$\xrightarrow{\mathbb{P}} \int y^2\, dG(y) - \left(\int y\, dG(y)\right)$$

$$= \mathrm{Var}(Y).$$

Bootstrap Variance Estimate

If we have data X_i, but F is unknown, then estimate F with $\hat{F}$, and draw from $\hat{F}$.

$\Rightarrow$ Draw $X_1^*, \ldots, X_n^*$ from $X_1, \ldots, X_n$ with replacement.

$\Rightarrow$ Compute $T^* = g(X_1^*, \ldots, X_n^*)$

Note Some of the X_i^* will be duplicates.

Variance Algorithm

DO $i = 1, \ldots, m$

Draw $X_1^*, \ldots, X_n^*$ from $\hat{F}$

Compute $T_i^* = g(X_1^*, \ldots, X_n^*)$

COMPUTE $v_{boot} = \frac{1}{m} \sum_{j=1}^{m} (T_j^* - \bar{T}^*)^2$

bootstrap estimate of the variance $\Rightarrow$ $\hat{se} = \sqrt{v_{boot}}$

We can use exactly the same algorithm to estimate the variance of median, mode, or any other integrable statistic. $\int g < \infty$.

Bootstrap Confidence Intervals

Method 1 If T is approximately normal, e.g. an MLE, the T^* is also approximately normal (and so is v_{boot})

$\Rightarrow$ a $1-\alpha$ confidence interval for T is:

$$CI = \left(T - z_{\alpha/2}\, \hat{se}_{boot} \,,\; T + z_{\alpha/2}\, \hat{se}_{boot} \right)$$

actual T from data

$\hat{se}_{boot} = \sqrt{v_{boot}}$

Method 3 Percentile Intervals (obvious idea)

Generate $T_1^*, \ldots, T_m^*$ using simulation,

and let $T^*_{\alpha/2}$ be the $\alpha/2$ percentile from $T_1^*, \ldots, T_m^*$

$$\Rightarrow \quad CI = \left(T^*_{\alpha/2} \,,\; T^*_{1-\alpha/2} \right)$$

Requires some justification, see appendix.