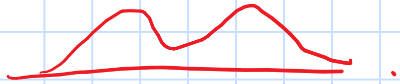
Recap



Our goal is to estimate the density of a distribution $p(x)$

$\rightarrow P = \int_R p(x')\,dx$

$\tilde{}$ ↳ smoothed version of $p(x)$

$\boxed{\star \text{ Goal } p_n(x) \rightarrow p(x)}$

→ $\boxed{K : \# \text{ of samples that fall in } R}$

$\mathbb{E}[k] = nP(x)$ x: point in R

n is the total number of sample points

$\boxed{V : \text{Volume enclosing } R}$

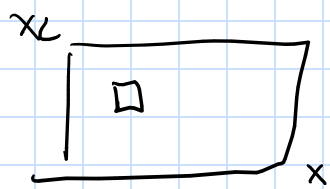
$\rightarrow p_n(x) = \dfrac{K/n}{V_n}$

$\boxed{\begin{array}{l} n \rightarrow \infty \\ V \rightarrow 0 \end{array}}$

To have Pn(x) converge to P(x) there are 3 conditions:

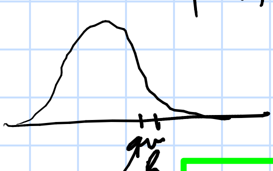
$\boxed{\#1}$ This condition assumes that the space averaged $\boxed{P/V}$ will converge to $p(x)$ provided that the $\boxed{\text{regions shrink uniformly}}$

$$\lim_{n \to \infty} V_n = 0$$



$\boxed{\#2}$ This condition $\boxed{\text{only makes sense for } p(x) = 0}$, assures us that frequency ratio will $\boxed{\text{converge in probability to } P}$

$$\lim_{n \to \infty} k_n = \infty$$



$\boxed{\#3}$ $\boxed{\text{This condition is needed for any type of convergence}}$

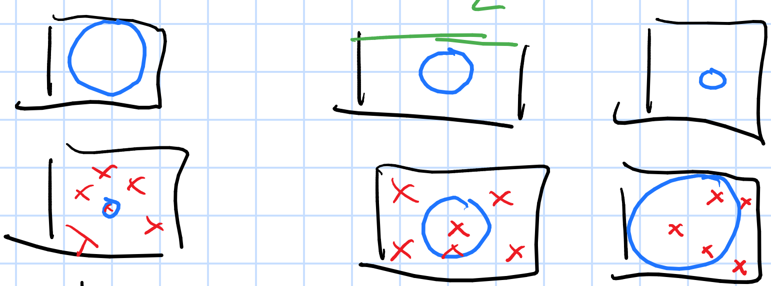
$$\lim_{n = \infty} k_n/n = 0$$

The condition have to be met for any convergance

$n=1$ $V_1 = 1$ $n=4$ $V_4 = \frac{1}{2}$ $n=9$ $V_4 = \frac{1}{3}$

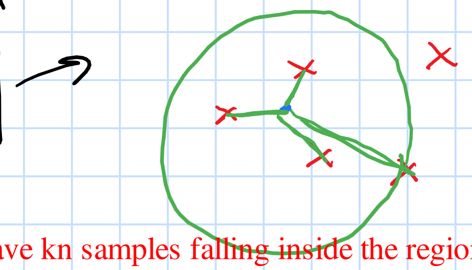
$V_n = \frac{1}{\sqrt{n}}$

$k_n = \sqrt{n}$



$k_1 = 1$ $k_4 = 2$ $k_9 = 3$

focus on fixing a region and continously shrinking it.



grown the region until we have kn samples falling inside the region

Density Estimation

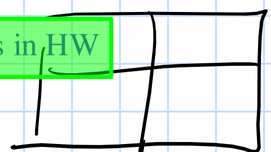
① Shrink an initial region by specifying the volume V_n as a function of n [Parzan window]

② Specify k_n as a function of n [nearest neighbor methods]

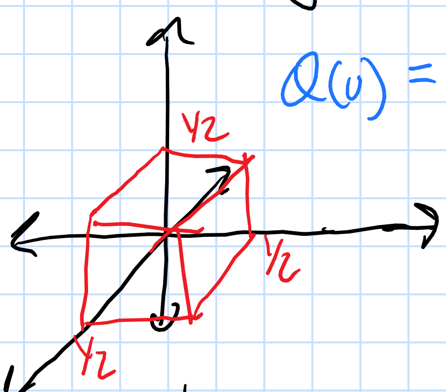
With enough number of sample both methods would give the same results.

Parzan Windows

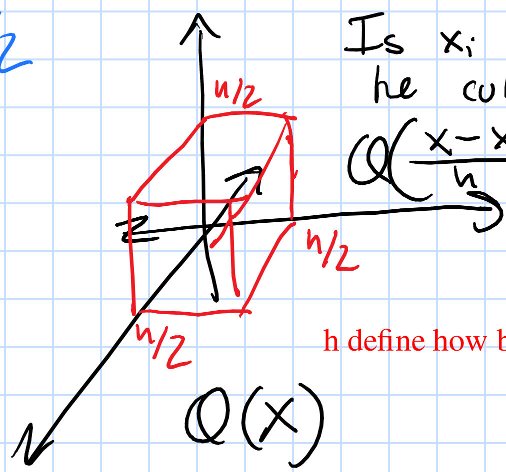
Might use this in HW



R is going to be defined as a d-dimensional hypercube



$$\varphi(u) = \begin{cases} 1 & |u_i| \leq 1/2 \\ 0 & \text{otherwise} \end{cases}$$



Is x_i inside the cube?

$$\varphi\left(\frac{x - x_i}{h}\right) = 1$$

h define how big is the cube

How many samples fall inside the cube?

$\varphi(x)$

$$\boxed{k_n = \sum_{i=1}^{n} \varphi\left(\frac{x - x_i}{h}\right)}$$

n: # number of sample

h: width of the window (cube)

x: is a point to get p(x)

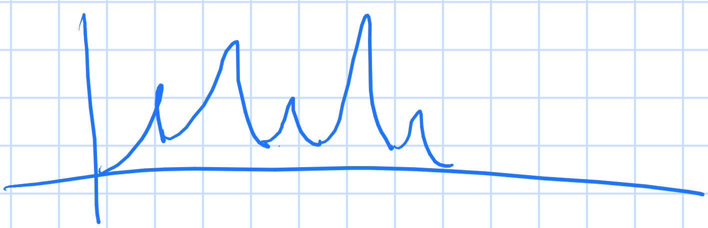
V: Volume

ϕ: Kernel (smoothing function)

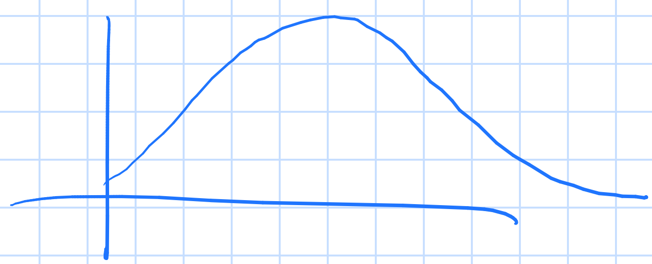
$$p_n(x) = \frac{k_n/n}{V} = \boxed{\frac{1}{nV}\sum_{i=1}^{n} \phi\left(\frac{x-x_i}{h}\right)}$$

$$\phi(x) = \exp\left(-\frac{\|x-x_i\|_2^2}{\sigma}\right)$$

Choose σ to small



Choose σ to be too large



K-NN estimators

① Select an initial volume around x to estimate $p(x)$

② Grow the window until k samples fall inside of the Region R

→ R is the region w/ the NN of x

③ $p_n(x) = \frac{k/n}{V}$

K-NNs for classification

$$p(\omega_i \mid x) = \frac{k_i}{K}$$

$\{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\}$

$\omega \in \{1, 2, 3\}$

Classify x!

$\|x - x_1\|_2^2 = d_1$

$\|x - x_2\|_2^2 = d_2$

$\vdots$

$d_{n=100}$

$d_{50}, d_{22}, d_{13}, d_{44}, d_2$