2.4 McDiarmid inequality in action

If g has the bounded differences property

if $\left|g(x_1, \dots x_{i-1}, x_i, \dots x_n) - g(x_1 \dots x_{i-1}, x_i', \dots x_n)\right| \le c_i$

and if $x_1, \dots x_n$ are independent then

$$P\{g(\underbrace{x_1, \dots, x_n}_{x^n}) - E[g(x^n)] \ge t\} \le e^{-\frac{2t^2}{\sum_i c_i^2}}$$

Example $P(X_i \in [a_i, b_i]) = 1$ all i

$$g(x_1, x_i, x_n) = x_1 + \dots + x_n \qquad x_i \in [a_i, b_i]$$

Has b.d.p. with $c_i = b_i - a_i$

so $P\{S_n - E[S_n] \ge t\} \le e^{-\frac{2t^2}{\sum_i c_i^2}}$

$$S_n = g(x^n) = x_1 + \dots + x_n.$$

Empirical probability distⁿ

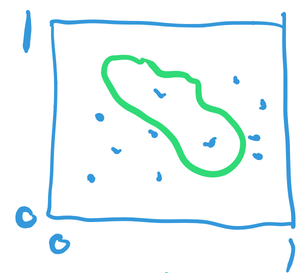
$\mathcal{X}$ = some space (or set)

P be a prob. distⁿ for $\mathcal{X}$.

Let $X_1, \dots, X_n$ indendent, distⁿ P.

If $A \subset \mathcal{X}$

$E[1_{\{X_i \in A\}}] = P\{X_i \in A\}$
$= P(A)$



$$P_n(A) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \in A\}}$$

$$P_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$$

2/11

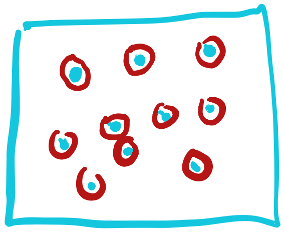
$n\left(\frac{1}{n}\right)^2 = \frac{1}{n}$

For fixed A:

$$P\{|P_n(A) - P(A)| \geq \varepsilon\} \overset{?}{\leq} e^{-2\varepsilon^2 n}$$

what's

$$P\{\underbrace{\sup_A |P_n(A) - P(A)|}_{\downarrow \text{ with probability } 1} \geq \varepsilon\} = 1 \qquad \varepsilon < 1$$



$P \sim$ uniform distn

$$P_n(A) = 1$$

$$P(A) \le n\pi\varepsilon^2 = \frac{\pi}{n} \to 0$$

$$\varepsilon = \frac{1}{n}$$

How about $g(x_1, \dots, x_n)$ - b.d.p. with $c_i = \frac{1}{n}$ for all

$$P\Big\{ \overbrace{\sup_{C \in \mathcal{C}} |P_n(C) - P(C)|}^{g(x_1,\dots,x_n)} \ge \varepsilon \Big\}$$

$\mathcal{C}$ some class of sets

e.g. axis aligned rectangles.

- consider moving x_i to x_i'

$P_n(C)$ could change by

$\pm \frac{1}{n}$

∴ McDiarmid

$$P\Big\{ \Big| \sup_{C \in \mathcal{C}} |P_n(C) - P(C)| - E\Big[\sup_{C \in \mathcal{C}} |P_n(C) - P(C)|\Big] \Big| \ge \varepsilon \Big\}$$

$$\le 2e^{-2n\varepsilon^2}$$

Estimating expectations.

Suppose $f : X \to [0, \underline{B}]$

Estimate $\mathbb{E}_P[f(X)]$ by $\mathbb{E}_{P_n}[f(X)] = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$

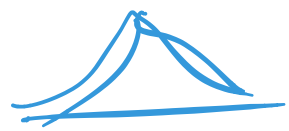
$$\mathbb{P}\left\{ \left| \underbrace{\sup_{f \in \mathcal{F}} \left| \mathbb{E}[f] - \frac{1}{n}\sum_{i=1}^{n} f(X_i) \right|}_{\text{B.D.P. } c_i = \frac{B}{n}} - \mathbb{E}[\quad] \right| \geq \varepsilon \right\}$$

$$\leq 2e^{-\frac{2n\varepsilon^2}{B^2}}$$

Density estimation

$X_1, \dots, X_n$ have pdf f over $\mathbb{R}^d$

e.g. $f(x) = \frac{1}{\sqrt{2\pi}^d} e^{-\|x\|^2/2}$

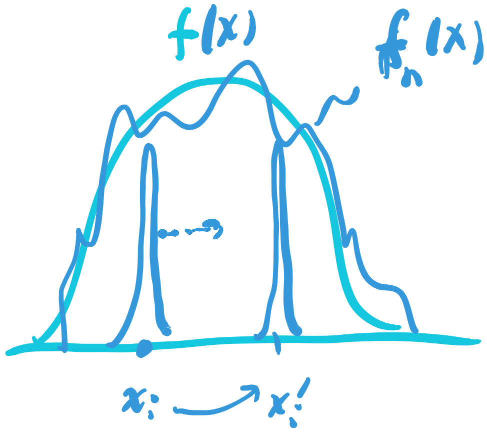


Suppose $\int_{\mathbb{R}^d} K(x)\,dx = 1$, $K \geq 0$ $h > 0$

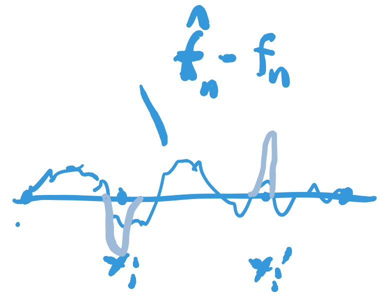
Estimator $\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$

Note $\int_{\mathbb{R}^d} \frac{1}{h} K\left(\frac{x}{h}\right) dx = 1$

$$\|f - \hat{f}_n\|_1 = \underbrace{\int_{\mathbb{R}^d} |f(x) - \hat{f}_n(x)|\,dx}_{\text{random}}$$



$$\overbrace{\int_{\mathbb{R}^d} |f - \hat{f}_n| \, dx}^{\text{Has B.D.P. } c_i \equiv 2/n}$$



$$P\left\{ \Big| \|f - \hat{f}_n\|_1 - E[\|f - \hat{f}_n\|_1] \Big| \geq \varepsilon \right\} \leq 2e^{-n\varepsilon^2/2}$$

Problem set 1 #3

Consider functions $f:[2n] \to \{0,1\}$
$\{1,2,\dots,2n\}$ $n=100$

examples

parity function $f(12)=0$
$f(13)=1$
$f(25)=1$
$f(30)=0$

prime numbers $f(2)=1$ $f(13)=1$ $f(8)=0$
$f(179)=1$ $f(12)=0$

threshold function $\mathbb{1}_{(x \geq 178)}$ $f(1)=0$ $f(196)=1$
$f(13)=0$ $f(180)=1$
$f(163)=0$ ...

example

select $f(i)=B_i$ with $B_i \sim \mathrm{Ber}(1/2)$ (fair coin flips)

$X_1=3,\ X_2=14,\ X_3=2\ \ X_4=1\ \cdots\ X_{200}=169$

$f(3)=3$ $f(14)=1$ $f(169)=0$

Samples are $\overbrace{(X_1, B_{X_1})}^{Z_1}, \cdots \overbrace{(X_n, B_{X_n})}^{Z_n}$

$\hat{f} = A((X_1,B_1)\cdots,(X_n,B_n))$

Claim 8 For an

(a) show that for any A and any $k \in [2n]$ that

$$P\{\hat{f}(k) \neq B_k\} \geq 0.25 = 1/4$$

V = fraction your estimator gets correct.

$E[V] \leq 3/4 = .75$

$P\{V \geq .9\} \leq \frac{.75}{.9} = 5/6$

So $P\{V < .9\} \geq 1/6$

B