1

NPS(1)

- Parametric statistics X - r.v. , $X_1 \dots X_n$ - i.i.d

$p(x) \in \{p(x, \vec{\theta})\}$, e.g for $X \sim N$ $\vec{\theta} = (\mu, \sigma^2)$

Non-parametric ($\approx$ distr.-free) $p(x)$ - contin. diff.

- Parametric regression $y_i = \langle \vec{a}, \vec{X_i} \rangle + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$

↑ linear assumption

Non-parametric : $y_i = f(\vec{X_i}) + \varepsilon_i$, f - cont. func.

$\mathbb{E}(\varepsilon_i) = 0$

$f(\vec{x_i}) = g(\langle \vec{a}, x_i \rangle)$, $g_i : \mathbb{R} \to \mathbb{R}$

↑ semi-parametric assumption

- Parametric ass. for stat. test:

$N(\mu, \sigma^2) \to$ t-test

Lemma $X_1, \dots X_n$ - i.i.d. $N(\mu, \sigma^2)$

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad \hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu}_n)^2$$

$$\frac{n \hat{\sigma}_n^2}{\sigma^2} \sim \chi^2_{n-1}, \text{ where } \hat{\sigma}_n \perp\!\!\!\perp \hat{\mu}_n \; ⚝$$

$$\left\{ \begin{array}{l} \dfrac{n \frac{1}{n} \sum (X_i - \mu)^2}{\sigma^2} = \\ = \sum \underbrace{\dfrac{X_i - \mu}{\sigma}}_{\sim N(0,1)} \sim \chi^2_n \end{array} \right.$$

Col. $\hat{\mu}_n \sim N(\mu, \sigma^2/n)$ $\quad \dfrac{\hat{\mu}_n - \mu}{\hat{\sigma}_n} \cdot \sqrt{n-1} \sim t_{n-1}$

$$⚝ \left\{ \begin{array}{l} \dfrac{\xi_0 \cdot \sqrt{n}}{\sqrt{\xi_1^2 + \dots + \xi_n^2}} \sim t_n \\ \xi_1^2 + \dots \xi_n^2 \sim \chi^2_n \end{array} \right.$$

① Prob. density estimation

Abs. cont. dist. $F(x) = P\{X \le x\} = \int_{-\infty}^{\infty} p(u)\,du$ PDF

Estimation of F

EDF $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \le x\}$

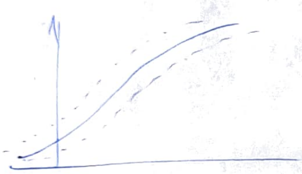
$\mathbb{E}(\hat{F}_n(x)) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\mathbb{1}\{X_i \le x\}] = F(x) \Rightarrow \hat{F}_n(x)$ - unbiased

Glivenko - Cantelli Thm $\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow[n\to\infty]{} 0$

Dvoretzky - Kiefer - Wolfowitz inequality

Massart ('90) $\mathbb{P}\{\sqrt{n} \sup_x |\hat{F}_n(x) - F(x)| > y\} \le \underbrace{2e^{-2y^2}}_{\alpha'}, \forall y \in \mathbb{R}$

$y = \sqrt{-\frac{1}{2} \log \frac{\alpha}{2}}$



$\mathbb{P}\{F \in \underbrace{(\hat{F}_n(x) - \frac{y}{\sqrt{n}}, \hat{F}_n(x) + \frac{y}{\sqrt{n}})}_{}\} > 1 - \alpha'$

conf. band of width $\sim 1/\sqrt{n}$

$\hat{p}_n(x) \neq \hat{F}_n'(x)$ - impossible to estimate

$P\{a \le X < b\} = \int_a^b p(u)\,du$

$\mathbb{P}\{x - h/2 \le X < x + h/2\}, h \to 0$

$= \int_{x-h/2}^{x+h/2} p(u)\,du = p(x) \cdot h \Rightarrow p(x) = \frac{\mathbb{P}\{\cdot\}}{h}$

$\hat{p}_n(x) = \frac{\#\{i : x - h/2 \le X_i < x + h/2\}}{nh}$

<u>Histogram</u> $c_1, \dots, c_m$ - collection of centering points

$\hat{f}_n(x) := \frac{1}{nh} \#\{i : X_i \in B_j\}$ if $x \in B_j$

$\forall j = \overline{1,m}$: $\underset{\text{bins}}{B_j} \ni c_j$; $B_i \cap B_j = \emptyset$, $i \neq j$

$B_1 \cup \dots \cup B_m = [A, B]$

(e.g.) $[A, B] = [0, 1]$

$B_j = [\frac{j-1}{m} ; \frac{j}{m})$, $j = \overline{1, m-1}$

$B_m = [\frac{m-1}{m}, 1]$

$E\hat{f}_h(x) \approx f(x)$

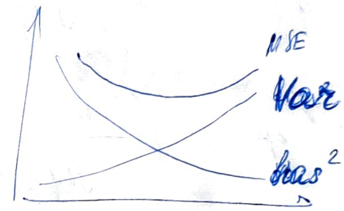
$\hat{f}_h(x) = \frac{1}{nh} \cdot \sum_{i=1}^{n} \mathbb{1}\{X_i \in B_j\}, \ \forall x \in B_j$

$E\hat{f}_i(x) = \frac{1}{nh} \sum_{i=1}^{n} \mathbb{E}\underbrace{[\mathbb{1}\{X_i \in B_j\}]}_{P\{X_i \in B_j\}} = \frac{1}{nh} \sum_{i=1}^{n} P\{X_j \in B_j\} =$

$= \frac{1}{nh} \sum^{n} \mathbb{P}\{c_j - 1/2 < x_i < c_j + 1/2\} = p(c_j)$

- MSE $\quad \mathbb{E}[(\hat{f}_n(x) - f(x))^2] \quad \Rightarrow$ bias-variance trade-off

MSE $(\hat{f}_n(x))$



MISE (mean integrated sq. er.)

$MISE(\hat{f}) = \int_R \mathbb{E}(\hat{f}_n(x) - f(x))^2 dx$

by Fub. thm

Bias - Variance decomposition

$MSE(\hat{f}_h(x)) = E[(\hat{f}_h(x) - E\hat{f}_h(x)) + (E\hat{f}_h(x) - f(x))^2] = \dots =$

$= bias^2 + Var$

ex. $X_1 \dots X_n \sim p$ $\quad \Rightarrow MSE(h(0)) = \left(\frac{h^2}{24}\right)^2 (1 + \bar{o}(1)) +$

$+ \frac{p(0)}{nh}(1 + \bar{o}(1))$

$p(0)$

$\hat{p}_h(0) = \frac{\#\{i : X_i \in [-h/2, h/2]\}}{nh} = \frac{\sum \mathbb{1}\{X_i \in [-h/2, h/2]\}}{nh}$

$E\hat{p}_h(0) = \frac{n \cdot p\{X_i \in [-h/2, h/2]\}}{nh} = \frac{p_h}{h}$

$Var\,\hat{p}_h(0) = \frac{n\,p_h(1-p_h)}{n^2h^2} = \frac{p_h(1-p_h)}{nh^2}$

$p_h = p\{-h/2 \le x_i < h/2\} = \int_{-h/2}^{h/2} p(u)\,du = \int_{-h/2}^{h/2} (p(0) + p'(0)u + p''(0)u^2 + o(u^2))\,du$

$= h \cdot p(0) + \frac{1}{2} \cdot 2 \cdot \frac{1}{3}(h/2) \dots$

$E\hat{p}_h(0) = p(0) + h^2/24 + \bar{o}(h^2)$

$Var\,\hat{p}_h(0) = 1/nh^2(h\,p(0) + \bar{o}(h)) = \frac{p(0)}{nh} \cdot (1 + \bar{o}(1))$

Thm f on $[0;1]$ – pdf

$[0,1] = [0, \frac{1}{n}) \cup (\frac{1}{n} ; \frac{2}{n}) \cup \ldots [\frac{n-1}{m}, 1]$

$\int_0^1 (f'(u))^2 du < \infty$

Then $MISE(\hat{f}_n) = \int \mathbb{E}(\hat{f}(x) - f(x))^2 \, dx =$

$$= \underbrace{\left(\frac{h^2}{12} \int_0^1 (f'(u))^2 du + \frac{1}{nh}\right)}_{g(h)} (1 + \bar{o}(1))$$

$$g'(h) = \frac{h}{6} \int (f'(u))^2 du - \frac{1}{nh^2}$$

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

Scott's rule $N(0, \sigma^2)$; $f(u) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$

$$\int_0^1 (f'(u))^2 du = \frac{1}{4\sqrt{\pi}\,\sigma^3} \Rightarrow h_{opt} = \frac{1}{n^{1/3}} (24\sqrt{\pi})^{1/3} \sigma = \frac{3{,}5 \cdot \sigma}{n^{1/3}}$$

$X_1, \ldots, X_n \Rightarrow \hat{\sigma}$

by Scott's rule

Freedman – Diaconis rule $3{,}5\,\sigma \approx 2 \cdot IQR$

$$\int \tilde{K}^2(x)\,dx = \int \sigma_k^2 \left(k^*(\sigma_k x)\right)^2 dx = \sigma_k \int (k^*(y))^2 dy$$

$$F(\tilde{K}) = \left(\sigma_k \int (k^*(y))^2 dy\right)^{4/5} = (\sigma_k^2)^{2/5} \left(\int (k^*(y))^2 dy\right)^{4/5} = F(k^*)$$

2) $\underset{\substack{k\text{-even} \\ \int K(x)dx = 1 \\ \int x^2 K(x) dx = 1 \\ a,b>0}}{\operatorname{argmin}} \left[\int k^2(x)\,dx \right] = \operatorname{argmin} \Big[\int K^2(x)dx + 2b \underbrace{\int x^2 K(x)dx}_{=1} - a^2 \underbrace{\int K(x)\,dx}_{=1} \Big]$

$$= \underset{K}{\operatorname{argmin}} \left[\int \left(K(x) + b(x^2 - a^2)\right)^2 dx - b^2 \int (x^2 - a^2)^2 dx \right]$$

$$= \underset{K}{\operatorname{argmin}} \int \left(K(x) + b(x^2 - a^2)\right)^2 dx$$

$= p(x) \underbrace{\int K(u)\, du}_{1} - h\, p'(x) \underbrace{\int u K(u)\, du}_{0} \quad \ldots$

$\langle \ldots \rangle$

$AMISE \xrightarrow[h]{} \min$

$= G(x)$

$$G'(x) = -\frac{1}{nh^2} \int K^2(u)\, du + h^3 \left(\int x^2 K(x)\, dx\right)^2 \int (p''(u))^2\, du = 0$$

$$h_{opt} = \left(\frac{\int K^2(u)\, du}{n\left(\int x^2 K(x) dx\right)^2 \cdot \int (p''(u))^2 du}\right)^{1/5}.$$

$$MISE\,(\hat{p}_n(x)) = \frac{5}{4} n^{-4/5} \left(\int x^2 K(x)\, dx\right)^{2/5} \left(\int K^2(x) dx\right)^{4/5} \cdot \left(\int (p''(x))^2\, dx\right)^{1/5}$$

? Which $K(x)$ min MISE?

K - even, $\int K(x)\, d(x) = 1$

$$\underset{\substack{K\text{-even}}}{\arg\min} \underbrace{\left(\int x^2 K(x) dx\right)^{2/5} \cdot \left(\int K^2(x) dx\right)^{4/5}}_{\mathcal{F}(K)} = K^*$$

$\int K(x)\, dx = 1$

1) It is enough to consider kernels satisfying $\int x^2 K(x) dx = 1$

In fact, if $\int x^2 K^*(x)\, dx = \sigma_K^2 \neq 1$,

then we can consider $\tilde{K}(x) = \sigma_K K^*(\sigma_K x)$

K^* - even $\Rightarrow \tilde{K}$ - even

$\int K^*(x)\, dx = 1 \Rightarrow \int \tilde{K}(x)\, dx = \int \sigma_K K^*(\sigma_K x)\, dx = 1$

$\int x^2 \tilde{K}(x) dx = \int x^2 \sigma_K^2 K^*(\sigma_K x)\, dx = \frac{1}{\sigma_K^2} \int y^2 K^*(y)\, dy = 1$

Density estimation:

- Histogram – piece-wise constant function

$X_1, \dots X_n$, $p(x)$ - pdf

$$\hat{p}_n(x) = \frac{1}{nh} \# \{ i : X_i \in B_j \} \quad \text{if } x \in B_j$$

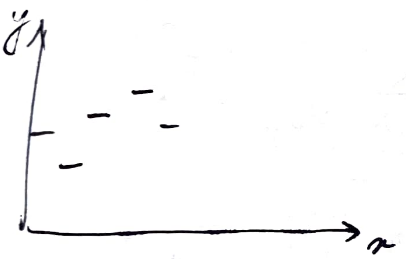
B_j - bins

$c_1, \dots c_n$ – central points

$B_j = [c_j - h/2, \; c_j + h/2)$, $j = 1 \dots M-1$

$[c_j - h/2, \; c_j + h/2]$, $j = M$

$[A, B] = B_1 \cup \dots \cup B_n$



$MISE(\hat{p}_n(x)) = \int_{\mathbb{R}} MSE(\hat{p}_n(x)) \, dx$ (integration of "point" errors)

$MSE(\hat{p}_n(x)) = E[(\hat{p}_n(x) - p(x))^2]$

e.g. $\hat{p}_n(0) = \frac{1}{nh} \# \{ i : X_i \in [-h/2, h/2] \}$

$MSE(\hat{p}_n(0)) = Bias^2(\hat{p}_n(0)) + Var(\hat{p}_n(0))$

$Bias(\hat{p}_n(0)) = \frac{p(0) h^2}{24} \cancel{(1+o(1))} \to 0$, $h \to 0$ (as. unbiased)

$Var(\hat{p}_n(0)) = \frac{p(0)}{nh} (1 + o(1)) \to 0$, $nh \to \infty$

$\Leftrightarrow \hat{p}_n(0) - E\hat{p}_n(0) \xrightarrow{M.S.S} 0 \Rightarrow \hat{p}_n(0) - E\hat{p}_n(0) \xrightarrow{P} 0$

① $h \to 0$ $\mathbb{E} p_h(0) \to p(0)$ $\Rightarrow$ $\mathbb{E} p_h(0) \xrightarrow{p} p(0)$

$\Rightarrow$ $\hat{p}_h(0) \xrightarrow{p} p(0)$

- consistency in one point (more powerful, comes from BVD)

Thm if $\sup p(0) = [0, 1]$

$\oplus$ $p \in C^2$, $p \in D^2$, $\int (p'(u))^2 du < \infty$,

$$MISE(\hat{p}_h) = \Big[\underbrace{\frac{h^2}{12}\int (p'(u))^2 du + \frac{1}{nh}}_{\text{AMISE (asympt.)}}\Big](1 + o(1))$$

$$h^* = \arg\min AMISE(\hat{p}_h) = \frac{1}{n^{1/3}}\left(\frac{6}{\int (p'(u)) du}\right)$$

$$\left.\begin{array}{l} Bias^2 \sim h^2 \\ Var^2 \sim \frac{1}{nh} \end{array}\right\} \Rightarrow \quad h^2 = \frac{1}{nh}$$

$$h = n^{-1/3}$$

1) $$MISE = \big(\underbrace{f_1(h) \cdot h^{k_1} + f_2(h)/h^{k_2}}_{G(h)}\big)(1 + o(h))$$

$$G'(h) = f_1(h) \cdot k_1 h^{k_1 - 1} + f_2(h) \cdot (-k_2) \cdot \frac{1}{h^{k_2+1}} = 0$$

$$h_{opt} = \left(\frac{f_2(h)}{f_1(h)} \cdot \frac{k_1}{k_2}\right)^{\frac{1}{k_1+k_2}}$$

2) $f_1(h) \cdot h^{k_1} = f_2(h) \cdot h^{k_2} \Rightarrow h = \left(\frac{f_2(h)}{f_1(h)}\right)^{\frac{1}{k_1+k_2}}$

correct (~~diff~~ differs by multiplicative constant, though order is the same, but easier to compute, ① ▷ $MISE(\hat{p}_n) = \int_0^1 Bias^2(\hat{p}_n(x))\,dx + \int_0^1 Var(\hat{p}_n(x))\,dx$

$$Bias\, \hat{p}_n(x) = E\hat{p}_n(x) - p(x) \underset{x\in B_j}{=} E\Big[\frac{1}{nh}\sum_{i=1}^n \mathbb{1}\{X_i \in B_j\}\Big] - p(x)$$

$$= \frac{1}{nh}\sum_{i=1}^n P\{X_i \in B_j\} - p(X) = \frac{1}{h}p_j - p(x),$$

where $p_j = P\{X_1 \in B_j\}$

$$Var(\hat{p}_n(x)) = Var\Big(\frac{1}{nh}\sum_{i=1}^n \mathbb{1}\{X_i \in B_j\}\Big) = \frac{1}{n^2h^2}\sum_{i=1}^n Var\,\mathbb{1}\{X_i\in B_j\} =$$

$$= \frac{1}{nh^2}\, \underbrace{E\,\mathbb{1}\{X_i \in B_j\}}_{p_j} - \underbrace{E\,\mathbb{1}\{X_1 \in B_j\}}_{p_j}{}^2 = \frac{1}{nh^2}(p_j - p_j^2)$$

② $$\int_0^1 Var\,\hat{p}_n(x)\,dx = \sum_{j=1}^N \int_{B_j} Var\,\hat{p}_n(x)\,dx =$$

$$= \sum_{j=1}^N \int_{B_j} \frac{1}{nh}(p_j - p_j^2)\,dx = \sum_{j=1}^N \frac{1}{nh}(p_j - p_j^2)$$

$$= \frac{1}{nh}\cdot \underbrace{\sum_{j=1}^N p_j}_{=1} - \frac{1}{nh}\sum_{j=1}^N p_j^2 =$$

$$P\{X_j \in B_j\} = \int_{B_j} p(u)\,du \overset{\text{by mean value thm}}{=} p(\bar{x}_j)\cdot h$$

$$= \frac{1}{nh} - \frac{1}{n\cancel{h}}\underbrace{\sum_{j=1}^N p^2(\bar{x}_j)\, h^{\cancel{2}}}_{\text{Rieman integral, } h\to 0} = \frac{1}{nh}(1 + \bar{o}(1))$$

$\begin{cases} n \to \infty \\ h \to 0 \end{cases}$ $(h(n))$

$nh \to \infty$

$\int_0^1 p^2(x)\,dx + o(1)$

$\Downarrow$

constant $+\, o(1)$

③ $\int_0^1 \mathrm{Bias}^2(\hat{p}_n(x))\,dx = \int_0^1 \left(\frac{1}{h} p_j - p(x)\right)^2 dx \ = \ ✪$

$p_j = \mathbb{P}\{X_i \in B_j\} = \int_{B_j} p(u)\,du \underset{\text{exp.}}{\overset{\text{Taylor}}{=}}$

$\underset{x \in B_j}{=} \int_{B_j} \left(p(x) + (u-x)\,p'(x) + \frac{(u-x)^2}{2} p''(\underset{B_j}{\underset{\cap}{\tilde{x}}})\right) du$

$= \underbrace{\int_{B_j} p(x)\,du}_{= p(x)\cdot h} + \int_{B_j} (u-x)\,p'(x)\,du + \underbrace{\int_{B_j} \frac{(u-x)^2}{2} \cdot p''(\tilde{x})\,du}_{\leq h^2/2 \cdot \max_{x \in B_j} p''(x) \cdot h} =$

$= p'(x) \int_{(j-1)h}^{jh} (u-x)\,du$

$= p'(x) \cdot h\left(-\frac{1}{2}h + jh - x\right)$

$(B_j = [(j-1)h, j\cdot h))$

bounded

$\Rightarrow$

$= \bar{\bar{o}}(h^2)$

$= p(x)\cdot h + p'(x)\cdot h\left(-\frac{1}{2}h + jh - x\right) + \bar{\bar{o}}(h^2) \Rightarrow$ to ✪

✪ $= \sum_{j=1} \int_{B_j} \left(\frac{1}{h} p_j - p(x)\right)^2 dx = \sum_{j=1}^{h} \int_{B_j} \Big(p'(x) \cdot \left(-\frac{1}{2}h + jh - x\right)$

$+ \bar{o}(h)\Big)^2 dx =$

$= \sum_{j=1}^{m} \int_{B_j} (p'(x))^2 \left(h\left(j - \frac{1}{2}\right) - x\right)^2 dx + \bar{\bar{o}}(h^3) = \int (p'(x))^2\,dx \cdot \frac{h^2}{12}\left(1 + \bar{\bar{o}}(1)\right)$

$= \sum_{j=1}^{h} (p'(\underset{B_j}{\underset{\cap}{\bar{x}_j}}))^2 \underbrace{\int_{B_j} \left(h\left(j - \frac{1}{2}\right) - x\right)^2 dx}_{1/12\, h^3}$

□

$= \Big(\underbrace{\sum_{j=1}^{m} (p'(\bar{x}_j))^2 h}_{\xrightarrow[h \to 0]{} \int (p'(x))^2 dx}\Big)\, h^2/12$

Kernel density estimation

Kernel $k : \mathbb{R} \to \mathbb{R}_+$

$\int_{\mathbb{R}} k(x)\,dx = 1$ ($\Rightarrow$ k - density function)

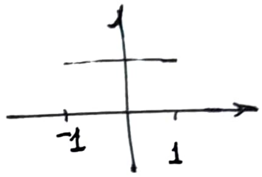
Assumption

1) k - even function 2) $\int_{\mathbb{R}} x^2 k(x)\,dx < \infty$

e.g.:

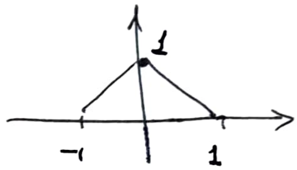
1) naive kernel (boxcar kernel

$$K(x) = \tfrac{1}{2}\, \mathbb{1}\{|x| < 1\}$$



2) triangular kernel

$$K(x) = (1 - |x|)\, \mathbb{1}\{|x| < 1\}$$

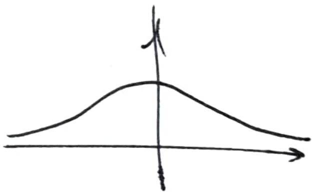


3) Epanichnicov kernel

$$K(x) = \tfrac{3}{4}(1 - x^2)\, \mathbb{1}\{|x| < 1\}$$

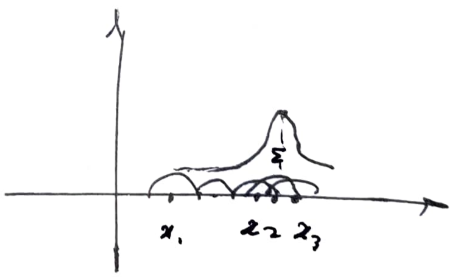
4) Gaussian Kernel

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$



$X_1, \dots, X_n \sim p(x)$

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



$\operatorname{supp} K = [-1, 1] \Rightarrow K\left(\frac{x - X_i}{h}\right) \neq 0,\ \forall x_i \in [X_i - L, X_i + L]$

e.g. naive kernel

$$\hat{p}_n(x) = \frac{1}{2nh} \sum_{i=1}^{n} \mathbb{1}\{|x - X| < h\}$$

doesn't matter

1) MISE

2) h

3) K

NPS(3)

Thm $p \in D^2$ (Tsybakov p. 19[3])

$\int (p''(u))^2 du < \infty$

$$MISE(\hat{p}_n) = \int E[(\hat{p}_n(x) - p(x))^2] dx =$$

$$= \Big[\underbrace{\frac{1}{nh}\int K^2(u) du}_{Var} + \underbrace{\frac{h^4}{4}\left(\int x^2 K(x) dx\right)^2 \int (p''(u))^2 du}_{Bias}$$

$$I_2 = \frac{1}{nh^2} \int \left(\int K\left(\frac{x-y}{h}\right) p(y)\, dy \right)^2 dx \ \textcircled{\leq}$$

$$\sqrt{K\left(\frac{x-y}{h}\right)} \cdot \sqrt{K\left(\frac{x-y}{h}\right)}\, p(y)$$

$$\int f(x)\, g(x)\, dx < \sqrt{\int f^2(x) dx} \cdot \sqrt{\int g^2(x) dx}$$

$$\leq \frac{1}{nh} \int \left[\int K\left(\frac{x-y}{h}\right) dy \int K\left(\frac{x-y}{h}\right) p^2(y)\, dy \right] dx \ =$$

$$\int K\left(\frac{x-y}{h}\right) dy = \left\{ dy = -h\, du \right\} = h \int K(u)\, du = h$$

$$\int K\left(\frac{x-y}{h}\right) p^2(y)\, dy = h \int K(u)\, p^2(x-uh)\, du$$

$$\leq \frac{1}{h} \int \int K(u) p^2(x-uh)\, du\, dx = \frac{1}{h} \int K(u) \left[\int_{-\infty}^{+\infty} p^2(x-uh)\, dx \right] du$$

$$\int_{-\infty}^{\infty} p^2(x)\, dx$$

$$= \frac{1}{h} \int p^2(x)\, dx = \bar{o}\, (I_1) \quad \int \operatorname{Var} \hat{p}_n(x) = I_1 (1 + \bar{o}(1))$$

$$\int \operatorname{Bias}^2(\hat{p}(x))\, dx$$

$$\operatorname{Bias}(\hat{p}(x)) = E\, \hat{p}_n(x) - p(x) = E\left(\frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)\right) - p(x) =$$

$$= \frac{1}{h}\, E\, K\left(\frac{x-x_i}{h}\right) - p(x) = \frac{1}{h} \int K\Big(\underbrace{\frac{x-y}{h}}_{u}\Big)\, p(y)\, dy - p(x) =$$

$$= \int K(u)\, \underbrace{p(x-uh)}\, du$$

$$p(x) - uhp'(x) + \frac{1}{2} u h^2 p''(x) + \ldots$$

$$\begin{cases} \int\limits_{|x|\le a} b(a^2 - x^2)\,dx = 1 \\ b\int\limits_{|x|\le a} x^2 (a^2 - x^2)\,dx = 1 \end{cases}$$

$$K^* = \begin{cases} -b(x^2 - a),\ |x| < a \\ 0,\ |x| \ge a \end{cases}$$

$$K^*(x) = \begin{cases} \frac{3}{4\sqrt{5}}\left(1 - x^2/5\right),\ |x| < \sqrt{5} \\ 0,\ |x| > \sqrt{5} \end{cases}$$ – Epanichnikov kernel

$K(x) = 3/4\,(1 - x^2)$, $|x| < 1 \Rightarrow K^*(x) = \sigma_n K(x\sigma_n)$

$\sigma_n = 1/\sqrt{5}$

$$\underset{K}{\operatorname{argmin}} \left(\int K^2(x)\,dx\right)^{4/5} \left(\underbrace{\int x^2 K(x)\,dx}_{F(x)}\right)^{2/5} = K^*$$

$I(K) = \frac{F(K_{ep})}{F(K)}$ – efficiency of the ker. K

$I(K) \le 1$, $I(K_{ep}) = 1$

$I(K)$ are very close to to 1

5 methods for bandwidth selection

nrd, nrd0, ucv, bcv, SJ.

nrd: $h_{opt} = \left(\frac{\int K^2(x)\,dx}{\left(\int x^2 K(x)\,dx\right)^2 \cdot \int (p''(u))^2\,du}\right)^{1/5}$

$p(x) \sim N(0, \sigma^2)$

$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$h_{opt} = \underbrace{(4/3)^{1/5}}_{1{,}06}\ \sigma\ n^{-1/5}$

$\sigma = \min\left(\sigma_\delta, \frac{1{,}34}{IQR}\right)$

nrd0 (Silverman) $1{,}06 \leftrightarrow 0{,}9$

Cross-validation

$$\int (\hat{f}_n(x) - f(x))^2 dx = \int \hat{f}_n(x) dx - 2\int \hat{f}_n(x) f(x) dx + \int f^2(x) dx$$

$$\hat{h}_{cv} = \arg\min \left(\int \hat{f}_n^2(x) dx - 2 \underbrace{\int \hat{f}_n(x) f(x) dx}_{E\hat{f}_n(X)} \right)$$

$$= 1/n \sum_{i=1}^{n} \hat{f}_{-i}(X_i)$$

$\uparrow$ based on $X_1, \dots, X_n$

Stone's thm:

$$\frac{\inf_{h \in R_+} \int (\hat{f}_h(x) - f(x))^2 dx}{\int (\hat{f}_{n,\hat{h}_{cv}}(x) - f(x))^2 dx} \xrightarrow[n \to \infty]{} 1 \quad a.s.$$

$$\int \hat{f}_n(x) dx = \int \left(\frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \right)^2 dx =$$

$$= \frac{1}{n^2h^2} \int \sum_{i=1}^{n} \left(K\left(\frac{x - x_i}{h}\right)\right) \sum_{i=1}^{n} \left(K\left(\frac{x_i - x_j}{h}\right)\right) dx$$

$$= \frac{1}{n^2h^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \int K\left(\frac{x - x_i}{h}\right) K\left(\frac{x - x_j}{h}\right) dx =$$

$$= \frac{1}{n^2h^2} \sum_i \sum_j \int K(u) K\left(u + \frac{x_j - x_i}{h}\right) dx =$$

convolution

$$= \frac{1}{n^2h^2} \sum_i \sum_j K * K\left(\frac{X_j - X_i}{h}\right)$$

UCV:

$$h_{opt} = \arg\min \left(\frac{1}{n^2h^2} \sum\sum K*K\left(\frac{x_i - x_j}{h}\right) - \frac{2}{n(n-1)h} \sum\sum K\left(\frac{x_i - X_j}{h}\right) + \frac{2}{(n-1)h} K(0) \right)$$

BCV ($n-1 \approx n$):

$$h_{opt} = \arg\min \left(\frac{1}{n^2h^2} \sum\sum \bar{K}\left(\frac{x_i - x_j}{h}\right) + \frac{2}{(n-1)h} K(0) \right)$$

SJ (Sheather-Jones): AMISE $\int (\hat{f}''_n(x))^2 dx$ $\quad (\hat{f}''_n(x) \approx f''(x))$

Thm (Van der Vaart)

$\mathcal{P}_m = \{ p$ - probability density function

$\int (p^{(m)}(x))^2 dx < \infty \}$

Then for $\forall$ estimate $\hat{p}_n$ of p

$$\sup_{p \in \mathcal{P}_m} \mathbb{E} \int (\hat{p}_n(x) - p(x))^2 dx \geq C \cdot n^{-2m/(2m+1)}$$

Hist : $m=1 \Rightarrow n^{-2/3}$

Kernel DE : $m=2 \Rightarrow n^{-4/5}$

NPS(4)

$(\vec{X}_1, y_1), \ldots, (\vec{X}_n, y_n)$, $\vec{X}_1, \ldots \vec{X}_n \in \mathbb{R}^p$, $y_1 \ldots y_n \in \mathbb{R}$

$y_i = f(\vec{X}_i) + \varepsilon_i \qquad \mathbb{E}\,\varepsilon_i = 0$

$f: \mathbb{R}^d \to \mathbb{R}$ - regression function

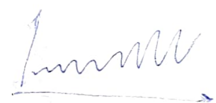
Aim: estimate f based on the data

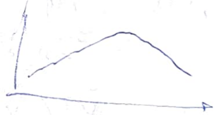
Param: $\varepsilon_i \sim N(0, \sigma^2)$

$$f(\vec{X}_i) = \beta^T \vec{X}_i$$

$$\hat{\beta} = (X^T X)^{-1} X^T \vec{y} \in \mathbb{R}^p \qquad X - Mat(n \times p)$$

e.g. fluctuations in temperature → frequency (multiples) → power





$\hat{f}_n : \sum_{i=1}^{n} (y_i - \hat{f}_n(\vec{X}_i))^2$

Cross-validation: $\sum_{i=1}^{n} (y_i - \hat{f}_{(-i)}(\vec{X}_i))^2$

$\hat{f}_{(-i)}$ - est. based on $(X_1, y_1), \ldots, (X_{i-1}, y_{i-1}), (X_{i+1}, y_{i+1}) \ldots$

↑ leave - one - out

GCV: $\hat{f}_n(\vec{X}_i) = y^T X((X^T X)^{-T})^T \cdot \vec{X} = (\vec{X}^T (X^T X)^{-1} X^T)\vec{y}$

$$\hat{f}_n : \begin{pmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{pmatrix} = \begin{pmatrix} f(X_1) \\ \vdots \\ f(X_n) \end{pmatrix} = (X(X^T X)^{-1} X^T)\vec{y} \Rightarrow \hat{\vec{y}} = H \cdot \vec{y}, \text{ where}$$

$$H = X(X^T X)^{-1} X^T$$

Def $\hat{y} = HY$ - linear smoothing method, H - smoothing matrix

$\operatorname{tr} H$ - effective degrees of freedom

$$\operatorname{tr} H = \operatorname{tr}(X(X^TX)^{-1}X^T) = \operatorname{tr}(\underbrace{(X^TX)^{-1}}_{p\times p}\underbrace{(X^TX)}_{p\times p}) = \operatorname{tr}(I_p) = p$$

Thm For linear regression:

$$y_i - \hat{f}_{(-i)}(X_i) = \frac{y_i - \hat{f}_n(X_i)}{1-h_{ii}}, \quad \text{where } H = (h_{ij})$$

Application to GCV

$$\sum_{i=1}^n (y_i - f_{(-i)}(X_i))^2 = \sum_{i=1}^n \left(\frac{y_i - \hat{f}_n(X_i)}{1-h_{ii}}\right)^2 \approx$$

$h_{11} + \dots + h_{nn} = \operatorname{tr} H \Rightarrow h_{ii} = \frac{\operatorname{tr} H}{n}$

assume that $h_{11} = \dots = h_{nn}$

$$= \frac{\sum_{i=1}^n (y_i - \hat{f}_n(x_i))^2}{(1-(\operatorname{tr}+1)/n)^2} \approx \left(1+2\frac{\operatorname{tr}+1}{n}\right)\cdot\sum_{i=1}^n (y_i - f(\bar{X}_i))^2$$

$$\frac{1}{(1-x)^2} = 1+2x$$

$$\triangleright \quad \hat{f}_{(-i)}(\bar{X}_i) = \bar{X}_i^T (X_{(-i)}^T X_{(-i)})^{-1} X_{(-i)}^T y_{(-i)}$$

$$\text{l.h.s.} = y_i - \bar{X}_i^T (X_{(-i)}^T X_{(-i)})^{-1} X_{(-i)}^T y_{(-i)} =$$

$$\underbrace{X_{(-i)}^T X_{(-i)}}_{p\times p} = \underbrace{X^T X}_{p\times p} - \underbrace{\bar{X}_i \bar{X}_i^T}_{(p\times 1)\cdot(1\times p)}$$

$\bar{X}_s = (x_{s1}, \dots x_{sp})$

$X = (x_{sk})_{\substack{s=\overline{1,n}\\ k=\overline{1,p}}}$

$$X_{(-i)}^T \vec{y}_{(-i)} = X^T y - \vec{X}_i y_i$$

$$(X_{(-n)}^T X_{(-n)})_{jh} = \sum_{s=1}^{n-1} x_{sj} \cdot x_{sh} =$$

$$\sum_{s=1}^n x_{sj} x_{sh} - x_{nj} x_{nh} =$$

$$(X^TX)_{jh} - (\bar{X}_n \bar{X}_n^T)_{jh}$$

Sherman - Morrison - Woodbury formula

$(A - B)^{-1}$ $\quad A \in \mathbb{R}^{n\times n}$ - invertible matrix

$\bar{u}, \bar{v} \in \mathbb{R}^n$

Then $$(A + \bar{u}\bar{v}^T)^{-1} = A^{-1} - \frac{A^{-1}\bar{u}\bar{v}^TA^{-1}}{1+\bar{v}^TA^{-1}\bar{u}}, \quad 1+\bar{v}^TA^{-1}\bar{u} \neq 0$$

Check: $(A+\bar{u}\bar{v}^T)\left(A^{-1}-\dfrac{A^{-1}uv^TA^{-1}}{1+v^TA^{-1}u}\right)$

$$= I_n + \bar{u}\bar{v}^TA^{-1} - \frac{u\boxed{v^TA^{-1}}}{1+v^TA^{-1}u} - \frac{uv^TA^{-1}u\boxed{v^TA^{-1}}}{1+v^TA^{-1}u}$$

$$= I_n + \cancel{uv^TA^{-1}} - \frac{u\cancel{(1+v^TA^{-1}u)}v^TA^{-1}}{\cancel{1+v^TA^{-1}u}} = I_n$$

$A = X^TX$, $\bar{u} = -\bar{x}_i$, $\bar{v} = x_i$

$$(X^TX - \bar{x}_ix_i^T) = (X^TX)^{-1} + \frac{(X^TX)^{-1}\vec{x}_i\vec{x}_i^T(X^TX)^{-1}}{1-\underbrace{\vec{x}_i^T(X^TX)^{-1}\vec{x}_i}_{h_{ii}}}$$

$$= y_i - \bar{x}_i^T\left((X^TX)^{-1} + \frac{(X^TX)^{-1}x_ix_i^T(X^TX)^{-1}}{1-h_{ii}}\right)(X^Ty - \bar{x}_iy_i) =$$

$h_{ii} = \bar{x}_i^T(X^TX)^{-1}\bar{x}_i$

$$= y_i - \bar{x}_i^T(X^TX)^{-1}X^Ty + h_{ii}y_i - \frac{h_{ii}x_i^T(X^TX)^{-1}X^Ty}{1-h_{ii}} + \frac{h_{ii}^2y_i}{1-h_{ii}}$$

$$\text{r.h.s.} = \frac{y_i - x_i^T(X^TX)^{-1}X^Ty}{1-h_{ii}} = \frac{e_i}{1-h_{ii}}$$

$$= (e_i + h_{ii}y_i)^{\cdot(1-h_{ii})} - \frac{h_{ii}(y_i - e_i)}{1-h_{ii}} + \frac{h_{ii}^2y_i}{1-h_{ii}} = \frac{e_i - e_ih_{ii} + h_{ii}y_i - h_{ii}^2y_i - h_{ii}y_i + h_{ii}e_i + h_{ii}^2y_i}{1-h_{ii}}$$

$$= \cancel{y_i - e_i - \frac{h_{ii}e_i}{1-h_{ii}} + h_{ii}y_i \cdot \frac{1-h_{ii}+h_{ii}}{1-h_{ii}}} = \frac{e_i}{1-h_{ii}} \quad \square$$

Local linear method

any "good" function is locally linear

How to determine vicinity → sup smu (super smoother)
→ LOESS

NPS(5)

Non-parametric Regression:

$(X_1, Y_1), \dots, (X_n, Y_n)$ ind.

$Y_i = f(X_i) + \varepsilon_i$ $\quad \mathbb{E}\varepsilon_i = 0$

f - linear smoother

$$\begin{pmatrix} \hat{f}(x_1) \\ \dots \\ \hat{f}(x_n) \end{pmatrix} = H \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad X =$$

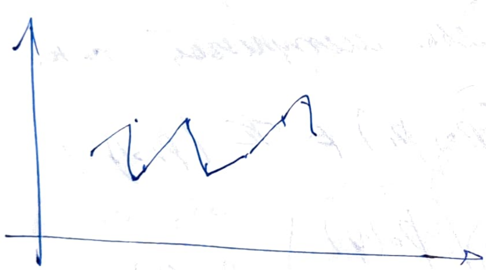
for lin. reg.:

$f(X_i) = \vec{X}_i^T \underset{\mathbb{R}^n}{\vec{\beta}}$ $\quad H = X(X^TX)^{-1}X^T$ $\quad \operatorname{tr} H = m$, $\vec{X}_1, \dots, \vec{X}_n \in \mathbb{R}^n$

smoothing matrix

for other forms

$\operatorname{tr} H$ - effective degree of freedom



$\sum_{i=1}^n (y_i - \hat{f}_n(X_i))^2 = 0$

GCV:

$$\left[\sum (y_i - \hat{f}_n(X_i))^2\right]\underbrace{\left(1 + 2\frac{\operatorname{tr} H}{n}\right)}_{?\text{, when } \operatorname{tr} H}$$

Motivation:

$H = (h_{ij})$

$$y_i - \hat{f}_{(i)}(X_i) = \frac{y_i - \hat{f}_n(X_i)}{1 - h_{ii}},$$

$$AIC_c = \log\left(\frac{1}{n}\sum(y_i - \hat{f}_n(X_i))^2\right) + 1 + \frac{2(\operatorname{tr} H + 1)}{n - \operatorname{tr} H - 2}$$

(*) $AIC_c = AIC + \frac{2k^2+2K}{n-k-1}$, $\cancel{AIC}$ $2K + n[\ln(2\pi\, RSS/n) + 1] = 2k + n \cdot \ln RSS = 2k - 2\ln L$

Model missspecification

True model $y_i = f_0(X_i) + \varepsilon_i \quad \mathbb{E}\,\varepsilon_i = 0 \quad \varepsilon_i \sim N(0, \sigma_0^2)$

$$y_i \sim N(f_0(x_i);\ \sigma_0^2)$$

$$\begin{pmatrix} y_1 \\ \cdots \\ y_n \end{pmatrix} \sim N\left(\underbrace{\begin{pmatrix} f_0(x_1) \\ \cdots \\ f_0(x_n) \end{pmatrix}}_{\vec{\mu}_0 =},\ \sigma_0^2 I_n\right)$$

Candidate model $y_i = f_1(X_i) + \varepsilon_i \quad \mathbb{E}\,\varepsilon_i = 0 \quad \varepsilon_i \sim N(0, \sigma_1^2)$

$$y_i \sim N(f_1(X_i),\ \sigma_1^2)$$

$$\begin{pmatrix} y_1 \\ \cdots \\ y_n \end{pmatrix} \sim N\left(\underbrace{\begin{pmatrix} f_1(y_1) \\ \cdots \\ f_1(x_n) \end{pmatrix}}_{\vec{\mu}_1 =},\ \sigma_1^2 I_n\right)$$

KL - divergence (discrepancy) (*) не явл. метрикой т.к.

$$K(p_0, p_1) \neq K(p_1, p_0)$$

$$K(p_0, p_1) = \mathbb{E} \ln \frac{p_0(y)}{p_1(y)} = \int \ln\left(\frac{p_0(y)}{p_1(y)}\right) \cdot p_0(y)\, dy$$

Lemma: $y \sim p_0$ Assume that $p_0(y) = 0 \Leftrightarrow p_1(y) = 0$

(p_0, p_1 have the same support)

$$K(p_0, p_1) \geq 0 \quad \forall p_0, p_1$$

$$K(p_0, p_1) = 0 \Leftrightarrow p_0(y) = p_1(y) \quad \forall y \in \mathbb{R}$$

▷ Jensen's inequality:

$\forall z$ - r.v. $\forall \varphi$ - convex:

$$E[\varphi(z)] \geq \varphi(E z)$$

$\textcircled{=} \Leftrightarrow z = const$

$\varphi(x) = -\ln x \leftarrow$ convex

$z = \frac{p_1(y)}{p_0(y)}$, $y \sim p_0$

$$E_0\left[-\ln\left(\frac{p_1(y)}{p_0(y)}\right)\right] \geq \underbrace{-\ln\left(E \frac{p_1(y)}{p_0(y)}\right)}_{= -\ln \int \frac{p_1(y)}{p_0(y)} p_0(y)\,dy = 0} \Rightarrow KL(p_0, p_1) \geq 0$$

($\uparrow$ w.r.t. $y \sim p_0$ (measure of p_0)

$\int \ln \frac{p_0(y)}{p_1(y)} \cdot p_0(y)\,dy$

+ $\varphi(x) = x^2$

$E Z^2 \geq (E Z)^2$

since

$Var\, Z \geq 0$

Assume $z = \frac{p_1(y(\omega))}{p_0(y(\omega))} = const$

$p_0(y) = 0 \Leftrightarrow p_1(y) = 0$

if $p_0(y) \neq 0$

$\exists \omega \in \Omega : y(\omega) = y$

$\Rightarrow p_1(y(\omega)) = c \cdot p_0(y(\omega)) \Rightarrow$

$\Rightarrow p_1(y) = c \cdot p_0(y) \Rightarrow$

$\Rightarrow \underbrace{\int p_1(y)\,dy}_{1} = c \cdot \underbrace{\int p_0(y)\,dy}_{=1} \Rightarrow c = 1$

$$\mathcal{K}(p_0, p_1) = \mathbb{E}_0 \ln p_0(y) - \mathbb{E}_0 \ln p_1(y)$$

w.r.t. to true model

(?) by tradition

$$\underset{p_1}{\operatorname{argmin}} \mathcal{K}(p_0, p_1) = \underset{p_1}{\operatorname{argmin}} (-2\, \mathbb{E}_0 \ln p_1(y)) \stackrel{\leq}{=}$$

AIC

lin regr.: true model $\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \sim N(\vec{\mu}_0, \sigma_0^2 I_n)$

cand. model $\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \sim N(\vec{\mu}_1, \sigma_1^2 I_n)$

$$p_1(\vec{y}) = \left(\frac{1}{\sqrt{2\pi}\,\sigma_1}\right)^n e^{-\frac{1}{2\sigma_1^2}(\vec{y} - \vec{\mu}_1)^T(\vec{y} - \vec{\mu}_1)}$$

$$\ln p_1(\vec{y}) = -n \log \sqrt{2\pi} - n \log \sigma_1 - \frac{1}{2\sigma_1^2}(\vec{y} - \vec{\mu}_1)^T(\vec{y} - \vec{\mu}_1)$$

$$\stackrel{\leq}{=} \underset{\mu_1, \sigma_1^2}{\operatorname{argmin}} \Big[\underbrace{-2n \log \sigma_1}_{= n \log \sigma_1^2} + \frac{1}{\sigma_1^2} \mathbb{E}_0 \underbrace{(\vec{y} - \vec{\mu}_1)^T(\vec{y} - \vec{\mu}_1)}_{} \Big] \stackrel{(=)}{=}$$

$\vec{y} \sim \vec{\mu}_0 + \sigma_0 \cdot \vec{\varepsilon}_0$, $\vec{\varepsilon}_0 \sim N(0, I_n)$

$$= \mathbb{E}_0 \left[[(\mu_0 - \mu_1) + \sigma_0 \bar{\varepsilon}_0)]^T \cdot [(\mu_0 - \mu_1) + \sigma_0 \bar{\varepsilon}_0)] \right]$$

$$\stackrel{(=)}{=} \underset{\mu_1, \sigma_1^2}{\operatorname{argmin}} \Big[n \log \sigma_1^2 + \frac{1}{\sigma_1^2} \Big[(\vec{\mu}_0 - \vec{\mu}_1)^T (\vec{\mu}_0 - \vec{\mu}_1) + \underbrace{2(\mu_0 - \mu_1)^T \cdot \sigma_0 \cdot \mathbb{E}\, \bar{\varepsilon}_0}_{= 0} + \sigma_0^2\, \mathbb{E}_0 \underbrace{[\varepsilon_0^T \varepsilon_0]}_{n} \Big] \Big] \stackrel{(=)}{=}$$

$$= \mathbb{E}\left[(\varepsilon_{0,1} \ldots \varepsilon_{0,n}) \begin{pmatrix} \varepsilon_{0,1} \\ \varepsilon_{0,n} \end{pmatrix} \right] \Rightarrow \mathbb{E}[\varepsilon_{01}^2 + \ldots + \varepsilon_{0n}^2] = n =$$

$\Leftrightarrow \operatorname{argmin}\left[n \log \sigma_1^2 + \frac{(\mu_0-\mu_1)^T(\mu_0-\mu_1)}{\sigma_1^2} + \frac{\sigma_0^2}{\sigma_1^2} n\right]$

we do not know $\mu_0, \sigma_0 \Rightarrow$ change the problem:

Cond. model $\div$ linear regression:

$\bar{y} = X\bar{\beta} + \bar{\varepsilon}, \quad \bar{\varepsilon} \sim N(0, \sigma_1^2 I_n)$

$\bar{\mu}_1 = X \cdot \hat{\bar{\beta}}, \quad \sigma_1^2 = \frac{1}{n}(\bar{y} - X\hat{\bar{\beta}})^T(\bar{y} - X\hat{\bar{\beta}})$

Analogue of the thm for lin. reg.:

1) $\bar{\mu}_1 \perp\!\!\!\perp \sigma_1^2$ independent (true for N)

2) $\eta_1 = \frac{n\sigma_1^2}{\sigma_0^2} \sim \chi^2_{n-m}$

3) $\eta_2 = \frac{n-m}{nm} \cdot \frac{(\mu_1-\mu_0)^T(\mu_1-\mu_0)}{\sigma_1^2} \sim F(n, n-m)$

$\Rightarrow \operatorname{argmin}\left[n \log \sigma_1^2 + \mathbb{E}\frac{(\mu_0-\mu_1)^T(\mu_0-\mu_1)}{\sigma_1^2} + \mathbb{E}\frac{\sigma_0^2}{\sigma_1^2} n\right]$

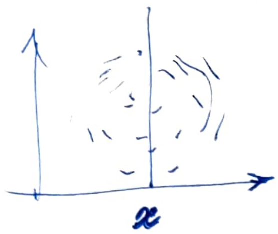
$= \operatorname{argmin}\left[\cancel{n} \log \sigma_1^2 + \frac{\cancel{n}m}{n-m}\underbrace{\mathbb{E}\,\eta_2}_{=\frac{n-m}{n-m-2}} + n^2 \cdot \underbrace{\mathbb{E}\left(\frac{1}{\eta_1}\right)}_{=\frac{1}{n-m-2}}\right]$

$= \operatorname{argmin}\left[\underbrace{\log \sigma_1^2}_{\log\left(\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2\right)} + \underbrace{\frac{m+n}{n-m-2}}_{=1+\frac{2(m+1)}{n-m-2}}\right] = \operatorname{argmin} \log\left(\frac{1}{n}\sum(\hat{y}_i - y_i)^2\right) + 1 + \frac{2(\operatorname{tr} H + 1)}{n - \operatorname{tr} H - 2}$

Kernel regression estimates

$$y_i = f(X_i) + \varepsilon_i$$

$$f(x) = \mathbb{E}[y_i \mid X_i = x]$$



Nadaraya - Watson

$$f(x) = \int_{\mathbb{R}} p_{Y|X}(y|x)\, y\, dy = \int_{\mathbb{R}} \frac{p_{(X,Y)}(x,y)}{p_X(x)}\, y\, dy$$

$$p_X(x) \leftrightarrow \hat{p}_h^X(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)$$

$$p_{(X,Y)}(x,y) \longleftrightarrow \hat{p}_h^{(X,Y)}(x,y) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) K\left(\frac{Y_i - y}{h}\right)$$

$$\hat{f}_n(x) = \int_{\mathbb{R}} \frac{1}{h} \frac{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) K\left(\frac{Y_i - y}{h}\right) y}{\sum K\left(\frac{X_i - x}{h}\right)}\, dy =$$

$$\int K\Big(\underbrace{\frac{Y_i - y}{h}}_{u}\Big)\, y\, dy = h \int K(u)(Y_i - uh)\, du =$$

$$= h\Big(Y_i \underbrace{\int K(u)\, du}_{=1} - h^2 \underbrace{\int K(u)\, u\, du}_{=0}\Big) = h\, Y_i$$

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

Comment :

1) $\sum_{i=1}^{n} (y_i-\beta)^2 K\left(\frac{x_i-x}{h}\right) \longrightarrow \min_{\beta}$

$$\frac{\partial}{\partial\beta} = -2\sum (y_i-\beta) K\left(\frac{x_i-x}{h}\right) = 0 \Rightarrow \beta = \frac{\sum K\left(\frac{x_i-x}{h}\right) y_i}{\sum K\left(\frac{x_i-x}{h}\right)}$$

Generalization : $\sum (y_i-\beta_0-\beta_1 x_i)^2 K\left(\frac{x_i-x}{h}\right)$

2) $\operatorname{tr} H$ $\quad \hat{Y} = HY \qquad \hat{y}_k = \sum_{i=1}^{n} \frac{K\left(\frac{x_i-x_k}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{x_j-x_k}{h}\right)} y_i$

$$H = \begin{pmatrix} \frac{K(0)}{\sum_j K\left(\frac{x_j-x_1}{h}\right)} & \frac{K\left(\frac{x_2-x_1}{h}\right)}{\sum_j K\left(\frac{x_j-x_1}{h}\right)} & \cdots & \frac{K\left(\frac{x_n-x_1}{h}\right)}{\sum_j K\left(\frac{x_j-x_n}{h}\right)} \\ \vdots & \frac{K(0)}{\sum_j K\left(\frac{x_j-x_2}{h}\right)} & & \vdots \\ \vdots & & \ddots & \vdots \\ \cdots & \cdots & \cdots & \frac{K(0)}{\sum_j K\left(\frac{x_j-x_n}{h}\right)} \end{pmatrix}$$

$$\operatorname{tr} H = K(0) \cdot \sum_{i=1}^{n} \frac{1}{\sum_{j=1}^{n} K\left(\frac{x_j-x_i}{h}\right)}$$

h - very small
in x_i-vicinity
we have 2 points (x_i, x_{i+1})

$$\operatorname{tr} H = K(0) \cdot \sum \frac{1}{K(0)} \qquad = \frac{\sum_{i=1}^{n} K\left(\frac{x_i-x}{h}\right) y_i}{\sum_{i=1}^{n} K\left(\frac{x_i-x}{h}\right)}$$

$$\operatorname{tr} H = K(0) \sum_{i}^{n} \frac{1}{K(0) + K\left(\frac{x_i+x_{i+1}}{2}\right)} < n$$

GCV:

$$\underbrace{\sum(\hat{y}_i - y_i)^2}_{\downarrow\, 0} \underbrace{\left(1 + \frac{2 \operatorname{tr} H}{h}\right)}_{\uparrow}$$

penalty $\dfrac{K(x_2 - x_i)}{\sum_k K\left(\frac{x_i - x_k}{h}\right)}$

1) kernels

2) $h \nearrow$ GCV, $h \searrow$ AIC

3) opt. problem

$$\sum (y_i - \beta)^2 K\left(\frac{x_i - x}{h}\right)$$

$$\sum (y_i - \beta_0 - \beta_1 x_i)^2 K\left(\frac{x_i - x}{h}\right)$$

1

Projection estimate (wavelets)

$$\mathcal{L}^2([A,B]) = \{f : \int_A^B f^2(x)\,dx < \infty\}$$

$$\langle f, g \rangle = \int_A^B f(x)\,g(x)\,dx$$

$$\|f\| = (\langle f, f \rangle)^{1/2} = \left(\int_A^B f^2(x)\,dx\right)^{1/2}$$

Def $\{\varphi_1, \varphi_2, \dots\}$ - sequence is orthonormal

if $\langle \varphi_i, \varphi_j \rangle = \begin{cases} 0 & , i \neq j \\ 1 & , i = j \end{cases}$

Seq. is complet if: ~~the~~ ~~follow~~

$$\langle f, \varphi_j \rangle = 0 \quad \forall j \Rightarrow f \equiv 0$$

Def Basis is an orthonormal and complete seq

Ex 1 Legendre polynomials

$$\sqrt{\frac{2j+1}{2}}\,\frac{1}{j!\,2^j}\left((x^2-1)^j\right)^{(j)}$$

$j = 0, 1, 2, \dots$

basis in $\mathcal{L}^2([-1, 1])$

$j=0$ $\sqrt{1/2}$ $\langle \sqrt{1/2}, \sqrt{1/2} \rangle = \int_{-1}^{1} (\sqrt{1/2})^2\,dx = 1$

$j=1$ $\sqrt{3/2} \cdot 1/2\,(x^2-1)' = \sqrt{3/2}\,x$

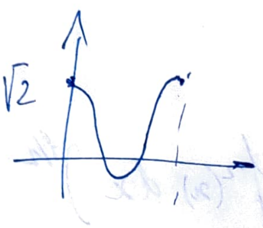
..

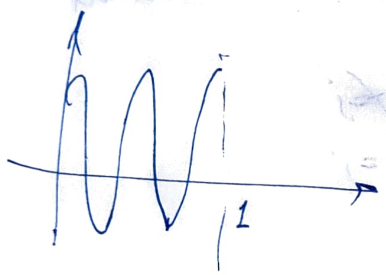
$j=n \Rightarrow P(n)$

ex 2) trigonometric basis $R^2([0,1])$

$\varphi_0(x) = 1$

$\varphi_{2k-1} = \sqrt{2}\cos(2\pi k x)$ $\qquad$ $\varphi_{2k}(x) = \sqrt{2}\sin(2\pi k x)$





NPS(6)

Projetion Estimation

$\mathcal{L}^2[A,B] = \{f : \int_A^B f^2(x)dx < \infty\}$

Scalar product : $\langle f, g\rangle = \int_A^B f(x)g(x)dx$

norm : $\|f\| = \sqrt{\langle f, f\rangle} = \sqrt{\int_A^B f^2(x)dx}$

Basis – a orthonormal and complete collection of function $\varphi_1, \varphi_2 \dots$

$\langle \varphi_i, \varphi_j\rangle = 0$, $i \neq j$ $\qquad$ $f : \langle \varphi_i, f\rangle = 0 \; \forall i$

$\|\varphi_i\| = 1, \forall i$ $\qquad$ $\Rightarrow f = 0$

<u>Thm</u> Let $\varphi_1, \varphi_2 \dots$ be a basis in $\mathcal{L}^2([A,B])$

Then $\forall f \in \mathcal{L}^2([A,B])$

$f(x) = \sum_{j=1}^{\infty} \theta_j \cdot \varphi_j(x)$, where $\theta_j = \int_A^B f(u)\varphi_j(u)du$

Parseval's identity $\|f\|^2 = \sum_{j=1}^{\infty} \theta_j^2$

replacement inf. sum $\to$ fin.

J (resolution)

- <u>PDE</u> : $X_1, \dots, X_n \sim p$

Assume that $p \in \mathcal{L}^2([A,B]) \Rightarrow p(x) = \sum_{j=1}^{\infty} \theta_j \varphi_j(x)$

$\theta_j = \int_A^B p(u)\varphi_j(u)du = \underset{p}{E}\,\varphi_j(X)$

$\Downarrow$

$\hat{\theta}_j = \sum_{i=1}^{n} \varphi_j(X_i)$

$\hat{p}_n(x) = \frac{1}{n}\sum_{j=1}^{J}\sum_{i=1}^{n} \varphi_j(X_i)\varphi_j(x)$

- Regression estimator $(X_1, Y_1) \dots (X_n, Y_n)$

$Y_i = f(X_i) + \varepsilon_i \qquad \mathbb{E}\varepsilon_i = 0$

$f \in \mathcal{L}^2([A,B]) : f(x) = \sum_{j=1}^{\infty} \theta_j \varphi_j(x)$, $\infty \leftrightarrow J$

$\theta_j = \int_A^B f(u)\varphi_j(u)du \approx \sum_{i=1}^{n} f(X_i)\varphi_j(X_i) \cdot \frac{B-A}{n}$, $f(X_i) \to Y_i$

$X_1 \dots X_n$ – approx. equidistant

$\Delta X \approx \frac{B-A}{n}$

$\Rightarrow \hat{\theta}_j = \frac{B-A}{n}\sum_{i=1}^{n}\varphi_j(X_i)Y_i$

$\hat{f}_n(x) = \frac{B-A}{n}\sum^{J}\left(\sum_{i=1}^{n}\varphi_j(X_i)Y_i\right)\varphi_j(x)$

Assumptions:

1. f belong to the Sobolev space $\quad [A,B] = [0,1]$

$$W(\beta, L) = \{ f : [0,1] \to \mathbb{R},\ f \in \mathcal{L}^2([0,1]),$$
$\beta > 0,\ L > 0$
$$\sum_{j=1}^{\infty} a_j^2 \theta_j^2 \le L^2 \}$$

$$a_j = \begin{cases} j^\beta & \text{for even } j \\ (j-1)^\beta & \text{for odd } j \end{cases}$$

<u>Prop.</u> $f \in W(\beta, L) \Rightarrow \int_0^1 (f^{(\beta)}(x))^2 dx < \infty$

<u>Thm:</u> $\lim_{n\to\infty} \sup_{f \in W(\beta,L)} E\left[n^{\frac{2\beta}{2\beta+1}} \| \hat{f}_{n,J} - f \|^2 \right] \le C$, prov. that $J = \lfloor \alpha \cdot n^{\frac{1}{2\beta+1}} \rfloor$ for some $\alpha > 0$

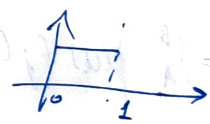
$$\hat{f}_{n,J}(x) = \sum_{j=1}^{J} \hat{\theta}_j \varphi_j(x)$$

where $\varphi_j(x)$ – trigonometric basis

* by Van der Vaart's Thm
$$\sup_{p \in \{..\}} E\left[n^{\frac{2\beta}{2\beta+1}} \| \hat{p}_n - p \| \right] \le c, \ \forall \hat{p}_n$$

Wavelet basis $[A,B] = [0,1]$

Haar father wavelet $\quad \phi(x) = 1 \quad \forall x \in [0,1]$



Haar mother wavelet $\quad \psi(x) = \begin{cases} 1 & x \in (1/2, 1] \\ -1 & x \in [0, 1/2) \end{cases}$

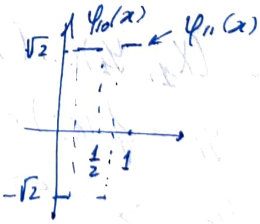
$$\psi_{jk}(x) = 2^{j/2} \psi(2^j x - k)$$

$$j = 1, 2, \ldots, ? \ldots \quad k = 0, \ldots, (2^j - 1)$$

$j = 1$: $\psi_{10}(x) = \sqrt{2} \cdot \psi(2x)$

$\psi_{11}(x) = \sqrt{2}\, \psi(2x - 1)$

$= \psi_{10}(x - \frac{1}{2})$

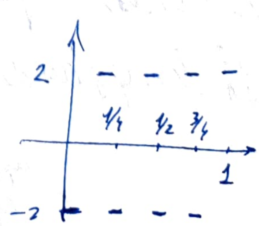


$j = 2$: $\psi_{20}(x) = 2\, \psi(4x)$

$\psi_{21}(x) = 2\, \psi(4x - 1) = \psi_{20}(x - 1/4)$

$\psi_{22}(x) = \psi_{20}(x - 1/2)$

$\psi_{23}(x) = \psi_{20}(x - 3/4)$



Prop. (ϕ, ψ, ψ_{jk}) is a basis in $\mathcal{L}^2([0,1])$

Resolution $(\phi, \varphi, \varphi_{jk}, \ j=1\dots J \quad k=0,\dots,(2^j-1))$

$$1+1+\sum_{j=1}^{J} 2^j = 1+\sum_{j=0}^{J} 2^j = 1+\frac{2^{J+1}-1}{2-1} = 2^{J+1}$$

$$\theta_{jk} = \int_0^1 \varphi_{jk}(u)\, f(u)\, du$$

$\Downarrow$

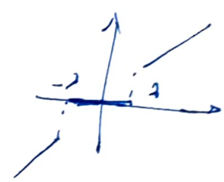
$\hat\theta_{jk} = \frac{1}{n}\sum_{i=1}^{n} \varphi_{jk}(x_i)\, y_i \approx 0$ ← sparsity (a lot of coef. ≈ 0)

$\Downarrow$

thresholding

1) Hard thresholding (discont.)

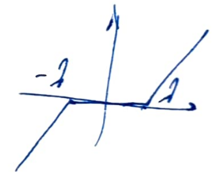
$$\hat\theta_{jk} \to \hat\theta^*_{jk} = \begin{cases} 0, & |\theta_{jk}| \le \lambda \\ \theta_{jk}, & |\theta_{jk}| > \lambda \end{cases}$$



2) Soft thresholding (continuous)

$$\hat\theta_{jk} \to \hat\theta^*_{jk} = \operatorname{sign}(\hat\theta_{jk}) \cdot (|\hat\theta_{jk}| - \lambda)_+ =$$

$$= \begin{cases} \hat\theta_{jk} - \lambda, & \hat\theta_{jk} > \lambda \\ \hat\theta_{jk} + \lambda, & \hat\theta_{jk} < -\lambda \\ 0, & |\hat\theta_{jk}| \le \lambda \end{cases}$$



Thm $\theta_{jk} = 0, \ \forall j, \forall k$, Assume $n = 2^J$

$y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$ i.i.d.

soft thresholding with $\lambda = \sigma\sqrt{\frac{2\log(n)}{n}}$

then $P\{\hat\theta^*_{jk} = 0, \forall j,k\} \xrightarrow[n\to\infty]{} 1$

▷ $\hat\theta_{jk} = \frac{1}{n}\sum_{i=1}^{n} \varphi_{jk}(x_i)\, y_i \sim N\left(\frac{1}{n}\sum_{i=1}^{n} \varphi_{jk}(x_i) f(x_i), \ \frac{\sigma^2}{n^2}\sum_{i=1}^{n} \varphi^2_{jk}(x_i)\right)$

$(y_i \sim N(f(x_i), \sigma^2))$

$$\varphi^2_{jk}(x) = 2^j$$

$$\operatorname{Var} \hat\theta_{jk} = \frac{\sigma^2}{n} \cdot \frac{1}{n} \cdot 2^j \le \frac{\sigma^2}{n}$$

$P\{\hat\theta^*_{jk} = 0 \ \forall j,k\} = P\{|\hat\theta_{jk}| < \lambda, \forall j,k\}$

$P\{\hat\theta^*_{jk} \ne 0$ for some $j,k\} = P\{|\hat\theta_{jk}| > \lambda\} = P\left\{\frac{\sqrt{n}}{\sigma}|\hat\theta_{jk}| > \frac{\sqrt{n}}{\sigma}\lambda\right\}$

by Mill's inequality:

$Z \sim N(0,1) \Rightarrow P\{|Z| > x\} \le \sqrt{\frac{2}{\pi}}\frac{1}{x} e^{-x^2/2}$

(I)

$$\hat{f}_n(x) = \sum_{j=1}^{J} \left(\frac{1}{n} \sum_{i=1}^{n} \varphi_j(X_i) Y_i \right) \varphi_j(x)$$

Legendre polynomials on $[-1, 1]$:

$$\psi_j(x) = \frac{1}{2^j j!} \sqrt{\frac{2j+1}{2}} \left[(x^2-1)^j \right]^{(j)}$$

$[A, B] = [4, 25]$ (from cars)

$$\varphi_j(x) = \sqrt{\frac{2}{B-A}} \, \psi_j\left(2\frac{x-A}{B-A} - 1 \right) \qquad x \in [A, B]$$

$$\int_A^B \varphi_{j_1}(x) \varphi_{j_2}(x) \, dx = \frac{2}{B-A} \int_A^B \psi_{j_1}\left(2\frac{x-A}{B-A} - 1\right) \psi_{j_2}\left(\frac{x-A}{B-A} - 1\right) dx =$$

$$= \int_{-1}^{1} \psi_{j_1}(y) \psi_{j_2}(y) \, dy = \begin{cases} 0 & j_1 \neq j_2 \\ 1 & j_1 = j_2 \end{cases}$$

$$\Rightarrow \hat{f}_n(x) = 2 \cdot \sum_{j=1}^{J} \left(\frac{1}{n} \sum_{j=1}^{n} \psi_j\left(2\frac{X_i - A}{B-A} - 1\right) Y_i \right) \psi_j\left(2\frac{x-A}{B-A} - 1\right)$$

(II) Doppler function

$$f(x) = \sqrt{x(1-x)} \cdot \sin\left(\frac{2.1\pi}{x+0.05}\right) \quad ; \ x \in [0,1]$$

Wavelet basis $(\phi, \varphi, \varphi_{jk})$

$$f(x) = \underset{m[1]}{\theta_\phi \phi(x)} + \theta_\varphi \varphi(x) + \sum_{j=1}^{J} \underbrace{\sum_{k=0}^{2^j-1} \theta_{jk} \varphi_{jk}(x)}_{j=1}$$

NPS (7)

Non-parametric Statistical Tests

Testing independence hypothesis

(X, Y) Pearson correlation coefficient $\rho = \dfrac{E[(X-EX)(Y-EY)]}{\sqrt{\operatorname{Var} X}\sqrt{\operatorname{Var} Y}}$

$$\hat{\rho} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2} \cdot \sqrt{\sum (Y_i - \bar{Y})^2}}$$

$H_0 : \rho = 0$? const is N dist? * hard to check (analogue S-W, K-S or assumed)

$H_1 : \rho \neq 0$ $\Leftrightarrow$ $\forall a, b \in \mathbb{R}$ $aX + bY \sim N$ $\sim \{\frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x-\mu}{\sigma}\right)^2}, \sigma > 0, \sigma = 0$

Thm: (X, Y) - norm. dist, then

$$\frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}} \sim t_{n-2}$$

e.g. canned tuna 9 types

$\{1, \ldots, 6\} \leftarrow$ marks $\sim \operatorname{Unif}(1, 6)$

$\rho = 0 \underset{\Leftarrow}{\not\Rightarrow} X \perp\!\!\!\perp Y$

Kendall's tau (more effective)

$(X_1, Y_1), \ldots, (X_n, Y_n)$ $X_1, \ldots, X_n$ - ind. $\sim F_X$

$Y_1, \ldots, Y_n$ - ind $\sim F_Y(x)$

F_X, F_Y are cont.

$H_0 : F_{(X,Y)}(x, y) = F_X(x) \cdot F_Y(y)$

$$\tau = 2\mathbb{P}\{\underbrace{(X_2 - X_1)(Y_2 - Y_1) > 0}\} - 1$$

$X_2 > X_1$, $Y_2 > Y_1$

$X_2 < X_1$, $Y_2 < Y_1$

1) $X_1 \perp\!\!\!\perp Y_1$, $X_2 \perp\!\!\!\perp Y_2$ $\Rightarrow \tau = 0$

$\tau = 2(P\{X_2 > X_1, Y_2 > Y_2\} +$

$P\{X_2 < X_1, Y_2 < Y_1\}) - 1$ $\leftarrow$ for cont.

$= 2(\underset{=1/2}{P\{X_2 > X_1\}} \underset{=1/2}{P\{Y_2 > Y_1\}} +$

$\underset{1/2}{P\{X_2 < X_1\}} \underset{1/2}{P\{Y_2 < Y_1\}}) - 1 = 0$

2) $\tau \in [-1, 1]$

3) $\tau(X, Y) = \tau(Y, X) = \tau(-X, -Y)$;

$\tau(-X, Y) = -\tau(X, Y)$

4) $\tau = \mathbb{P}\{(X_2 - X_1)(Y_2 - Y_1) > 0\} -$

$(1 - P\{(X_2 - X_1)(Y_2 - Y_1) > 0\}) =$

$= P\{(X_2 - X_1)(Y_2 - Y_1) > 0\} -$

$P\{(X_2 - X_1)(Y_2 - Y_1) < 0\}$

$$\hat{\tau} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Q((X_i, Y_i), (X_j, Y_j))$$

$$Q((a,b),(c,d)) = \begin{cases} 1, & (c-a)(d-b) > 0 \\ -1, & (c-a)(d-b) < 0 \\ 0, & a = c,\ b = d \end{cases}$$

ties in the data (repeated values)

$X_i = X_j$ or $Y_i = Y_j$ for some $i \neq j$

$(X_i - X_j)(Y_i - Y_j) > 0 \Rightarrow (X_i, Y_i)$ and (X_j, Y_j) are concordant (concord.) $\Rightarrow K'$ pairs

— " — < 0 — " — discordant $\Rightarrow K''$ pairs

$$K' + K'' = C_n^2 = \frac{n(n-1)}{2}$$

$\begin{pmatrix} (1,1) \\ (2,2) \\ (3,3) \end{pmatrix}$

(e.g.) $n = 3$ $\quad (1, Y_1), (2, Y_2), (3, Y_3)$

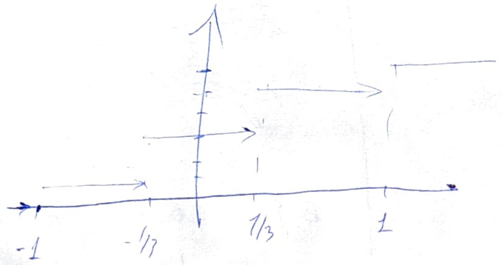
where $Y_1, Y_2, Y_3 = 1$ or 2 or 3

P	Y_1, Y_2, Y_3	$\Sigma\Sigma$
1/6	1, 3, 2	1
1/6	1, 2, 3	3
1/6	3, 1, 2	−1 (−1+1−1)
1/6	3, 2, 1	−3
1/6	2, 1, 3	1
1/6	2, 3, 1	−1

$$\hat{\tau} = \frac{1}{3} \cdot \sum_{i=1}^{2} \sum_{j=1+i}^{3} Q((X_i, Y_i), (X_j, Y_j))$$

$\forall\ j > i \quad X_j > X_i$

$$\tau = \begin{cases} 1 & , 1/6 \\ 1/3 & , 1/3 \\ -1/3 & , 1/3 \\ -1 & , 1/6 \end{cases}$$



$$E\hat{\tau} = \frac{2}{n(n-1)} E\left[\sum_{i=1}^{n-1}\sum_{j=1+i}^{n} Q((X_i, Y_i), (X_j, Y_j))\right] =$$

$$= \frac{2}{n(n-1)} \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E\left[\mathbb{1}\{(X_j - X_i)(Y_j - Y_i) > 0\} - \mathbb{1}\{(X_j - X_i)(Y_j - Y_i) < 0\}\right] =$$

$$= \frac{2}{n(n-1)} \underbrace{\sum\sum}_{C_n^2} \underbrace{\left[P\{\ldots > 0\} - P\{\ldots < 0\}\right]}_{\tau} = \tau \quad \text{(unbiasedness)}$$

$$\mathrm{Var}\,\hat{\tau} = \left(\frac{2}{n(n-1)}\right)^2 n(n-1)\left[\frac{1}{2}(1-\tau^2) + 4(n-2)\left(\delta - \left(\frac{\tau+1}{2}\right)^2\right)\right],$$

$\delta = P\{(X_2 - X_1)(Y_2 - Y_1) > 0$ and $(X_3 - X_1)(Y_3 - Y_1) > 0\} =$

+	+	+	+
+	+	−	−
−	−	+	+
−	−	−	−

$= P\{X_2 > X_1, Y_2 > Y_1, X_3 > X_1, Y_3 > Y_1\} +$
$+ P\{X_2 > X_1, Y_2 > Y_1, X_3 < X_1, Y_3 < Y_1\} +$
$+ P\{X_2 < X_1, Y_2 < Y_1, X_3 > X_1, Y_3 > Y_1\} +$
$+ P\{X_2 < X_1, Y_2 < Y_1, X_3 < X_1, Y_3 < Y_1\}$ $\ominus$

$= P\{X_1 < \min(X_2, X_3), Y_1 < \min(Y_2, Y_3)\} +$

$+ P\{X_3 < X_1 < X_2, Y_3 < Y_1 < Y_2\} +$

$+ P\{X_2 < X_1 < X_3, Y_2 < Y_1 < Y_3\} +$

$+ P\{X_1 > \max(X_2, X_3), Y_1 > \max(Y_2, Y_3)\}$

If $X \perp\!\!\!\perp Y \Rightarrow \delta = \frac{1}{3}\cdot\frac{1}{3} + \frac{1}{6}\cdot\frac{1}{6} + \frac{1}{6}\cdot\frac{1}{6} + \frac{1}{3}\cdot\frac{1}{3} = \frac{5}{18}$

$$\Rightarrow \operatorname{Var}\hat{\tau} = \frac{4}{n(n-1)}\left[\frac{1}{2}(1-0) + 4(n-2)\left(\frac{5}{18} - \left(\frac{0+1}{2}\right)^2\right)\right]$$

$$= \frac{4}{n(n-1)}\left[\frac{1}{2} + \frac{n-2}{9}\right] = \frac{4(2n+5)}{9n(n-1)}$$

Thm: if $X \perp\!\!\!\perp Y \Rightarrow \frac{\hat{\tau}}{\sqrt{\operatorname{Var}\hat{\tau}}} \xrightarrow[n\to\infty]{} N(0,1)$ large sample approx

$$\operatorname{Var}\hat{\tau} = \frac{4(2n+5)}{9n(n-1)}$$

$H_0: X \perp\!\!\!\perp Y$
$H_1: \tau \neq 0$ $\Rightarrow$ 1) if the exact distr. of $\hat{\tau}$ is known — kendall's distribution

$|\hat{\tau}| > k_{\alpha/2}$ $\leftarrow P\{\hat{\tau} > k_{\alpha/2}\} = \alpha/2$

2) otherwise

$$\left|\frac{\hat{\tau}}{\sqrt{\operatorname{Var}\hat{\tau}}}\right| > z_{\alpha/2}$$

Thm: (X, Y) - normal distr.

$$\tau = \frac{2}{\pi}\arcsin(\rho)$$

$$\boxed{\text{Spearman's } \hat{\rho}_S = \frac{\sum (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum (R_i - \bar{R})^2 \sum (S_i - \bar{S})^2}}}$$

e.g. $\{(7,1;\ 2,6),\ (7,8;\ 2,7),\ (7,5;\ 3)\}$

R_i	S_i
1	2
3	1
2	3

$(X_1, Y_1) \ldots (X_n, Y_n)$

$X_1, \ldots, X_n$ $\quad 1 \to X_{(1)} < \ldots < n \to X_{(n)}$ $\quad R_i = \operatorname{rank}(X_i)$

$\quad 1 \to Y_{(1)} < \ldots < n \to Y_{(n)}$ $\quad S_i = \operatorname{rank}(Y_i)$

Thm (Proof)

(X_1, Y_1), (X_2, Y_2)

$\tau = 2\mathbb{P}\{(X_2 - X_1)(Y_2 - Y_1) > 0\} - 1$

$U = \dfrac{X_1 - X_2}{\sqrt{2}\,\sigma_x}$ $\quad V = \dfrac{Y_1 - Y_2}{\sqrt{2}\,\sigma_y}$ $\quad \tau = 2\mathbb{P}\{UV > 0\} - 1$

$\mathbb{E}U = \mathbb{E}V = 0$ $\quad \operatorname{Var} U = \operatorname{Var} V = 1$ $\quad U, V \sim N(0,1)$

$\dfrac{2\operatorname{Var} X_1}{2\sigma_x^2}$

$\operatorname{cov}(U, V) = E[UV] = $

$= \dfrac{E[(X_1 \overset{-\mathbb{E}X_1 + \mathbb{E}X_2}{-} X_2)(Y_1 \overset{-\mathbb{E}Y_1 + \mathbb{E}Y_2}{-} Y_2)]}{2\sigma_x \sigma_y} =$

$= \dfrac{E[(X_1 - EX_1)(Y_1 - EY_1)] + E[(X_2 - E(X_2))(Y_2 - E(Y_2))]}{2\sigma_x \sigma_y} = \dfrac{2\sigma_x \sigma_y \rho}{2\sigma_x \sigma_y} = \rho$

$\frac{1}{2} - (\frac{1}{2} - P\{U < 0, V < 0\})$

$\frac{1}{2} - P\{U > 0, V < 0\}$

$\tau = 2\mathbb{P}\{U, V > 0\} - 1 = 2(\mathbb{P}\{U > 0, V > 0\} +$

$+ \underbrace{P\{U < 0, V < 0\}}_{\Phi(0,0)} - 1 = 4\Phi(0,0) - 1$

$\begin{cases} U, V \sim N(0,1) \\ \operatorname{cov}(U, V) = \rho \\ \tau = 2\mathbb{P}\{UV > 0\} - 1 \end{cases}$

$\begin{pmatrix} U \\ V \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$

Box - Müller $\quad \xi_1 \perp\!\!\!\perp \xi_2 \sim \operatorname{Unif}(0,1)$

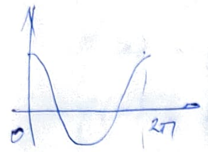
$\begin{cases} V_1 = \sqrt{-2\ln(\xi_1)} \cdot \cos(2\pi \xi_2) \\ V_2 = \sqrt{-2\ln(\xi_1)} \cdot \sin(2\pi \xi_2) \end{cases}$ $\quad V_1 \perp\!\!\!\perp V_2 \sim N(0,1)$

$\begin{cases} W_1 = V_1 \\ W_2 = \rho V_1 + \sqrt{1-\rho^2}\, V_2 \end{cases}$ $\quad W_1, W_2 \sim N(0,1)$, $\operatorname{corr}(W_1, W_2) = \rho$

$\begin{cases} W_1 = \sqrt{-2\ln(\xi_1)} \cdot \cos(2\pi \xi_2) \\ W_2 = \sqrt{-2\ln(\xi_1)} \cdot (\cos(2\pi \xi_2)\rho + \sin(2\pi \xi_2)\sqrt{1-\rho^2}) \end{cases}$ $\quad \varphi : \sin\varphi = \rho$, $\varphi = \arcsin\rho \in [-\frac{\pi}{2}, \frac{\pi}{2}]$

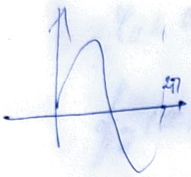
$= \sqrt{-2\ln(\xi_1)} \cdot \sin(2\pi \xi_2 + \varphi)$

$$P\{W_1<0,\ W_2<0\} = P\{\cos(2\pi\xi_2)<0;\ \sin(2\pi\xi_2+\varphi)<0\}$$



$$= \mathbb{P}\{\pi/2 < 2\pi\xi_2 \le 3\pi/2,\ \pi-\varphi \le 2\pi\xi_2 \le 2\pi-\varphi\} =$$

$$= \mathbb{P}\{\pi-\varphi \le 2\pi\xi_2 \le 3\pi/2\} =$$



$$= \mathbb{P}\{1/2-\varphi/2\pi \le \xi_2 \le 3/4\} = 3/4 - 1/2 + \varphi/2\pi = \frac{1}{4} + \varphi/2\pi$$

$\sim$ Unif

$$\tau = 4\Phi(0,0) - 1 = 4(1/4 + \varphi/2\pi) - 1 = \frac{2}{\pi}\varphi = \frac{2}{\pi}\arcsin\varphi$$

Theoretical Part

T1. $f(x) = x^p$, $x \in [0,1]$, $p \in \mathbb{N}$

$\phi(x) = 1$

$\psi(x) = \begin{cases} 1, & x \in [1/2, 1] \\ -1, & x \in [0, 1/2) \end{cases}$

$\psi_{jk}(x) = 2^{j/2} \psi(2^j x - k)$ $\quad j = 1, 2, \dots, \quad k = 0, \dots, 2^j - 1$

$$f_J(x) = \alpha \phi(x) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \beta_{jk} \psi_{jk}(x)$$

$$\alpha = \int_0^1 f(x) \cdot \phi(x)\, dx = \int_0^1 x^p \cdot 1\, dx = \frac{1}{p+1}$$

$$\beta_{jk} = \int_0^1 f(x) \psi_{jk}(x)\, dx = \int_0^1 x^p 2^{j/2} \psi(2^j x - k)\, dx =$$

$$= 2^{j/2} \left(\int_{(k+1/2)/2^j}^{(k+1)/2^j} x^p dx - \int_{k/2^j}^{(k+1/2)/2^j} x^p dx \right) = \frac{2^{j/2}}{p+1} \left(\left(\frac{k+1}{2^j}\right)^{p+1} - 2\left(\frac{k+1/2}{2^j}\right)^{p+1} + \left(\frac{k}{2^j}\right)^{p+1} \right)$$

$$f_J(x) = \frac{1}{p+1} \cdot \mathbb{1}\{x \in [0,1]\} + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \left(\frac{2^{j/2}}{p+1} \left(\left(\frac{k+1}{2^j}\right)^{p+1} - 2\left(\frac{k+1/2}{2^j}\right)^{p+1} + \left(\frac{k}{2^j}\right)^{p+1} \right) \cdot \right.$$

$$\left. \cdot \left(2^{j/2} \psi(2^j x - k) \right) \right) = \frac{1}{p+1} \mathbb{1}\{x \in [0,1]\} + \frac{1}{p+1} \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} 2^{-jp} \left((k+1)^{p+1} - 2(k+1/2)^{p+1} + k^{p+1} \right) \cdot$$

$$\cdot \psi(2^j x - k) = \frac{1}{p+1} \mathbb{1}\{x \in [0,1]\} + \frac{1}{p+1} \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} 2^{-jp} \left((k+1)^{p+1} - 2(k+1/2)^{p+1} + k^{p+1} \right) \cdot$$

$$\left(\mathbb{1}\left\{x \in \left[\frac{k+1/2}{2^j}, \frac{k+1}{2^j}\right]\right\} - \mathbb{1}\left\{x \in \left[\frac{k}{2^j}, \frac{k+1/2}{2^j}\right)\right\} \right)$$

T2. $f(x)$, $x \in [a, b]$, $B_1 \dots B_m$ – bins,

$$\hat{f}_n(x) = \frac{1}{k_j} \sum_{i: x_i \in B_j} Y_i, \quad \forall x \in B_j$$

$$\hat{f}_n(x) = \sum_{i=1}^{n} Y_i\, l_i(x), \quad \text{where } l_i(x) = \left(0, \dots, 0, \frac{1}{k_j}, \dots, \frac{1}{k_j}, 0, \dots, 0\right)$$

$$L = \begin{bmatrix} 1/k_1 & \cdots & 1/k_1 & & & \\ \vdots & & \vdots & & 0 & \\ 1/k_1 & \cdots & 1/k_1 & & & \\ & & & \ddots & & \\ & 0 & & & 1/k_m & \cdots & 1/k_m \\ & & & & \vdots & & \\ & & & & 1/k_m & \cdots & 1/k_m \end{bmatrix}$$

$$\nu = tr(L) = k_1 \cdot \frac{1}{k_1} + \dots + k_m \cdot \frac{1}{k_m} = m \leftarrow \text{number of eff degrees of freedom}$$

T3. $K(x) = (1-|x|)\,\mathbb{1}\{|x| \leq 1\}$

$(x_1, y_1) \dots (x_6, y_6)$, $x_i = i \quad \forall i = \overline{1,6}$

(i) $h = 1/2$

$$v = \operatorname{tr} H = \underset{1}{\underset{\parallel}{K(0)}} \cdot \sum_{i=1}^{6} \frac{1}{\sum_{j=1}^{6} K\left(\frac{x_j - x_i}{h}\right)} = 6$$ - eff degrees of freedom

$$H = \begin{bmatrix} 1 & & & & & \\ & 1 & & & 0 & \\ & & 1 & & & \\ & & & 1 & & \\ & 0 & & & 1 & \\ & & & & & 1 \end{bmatrix}$$

(ii) $h = 3/2$

$$v = \operatorname{tr} H = \underset{1}{\underset{\parallel}{K(0)}} \cdot \sum_{i=6}^{6} \frac{1}{\sum_{j=1}^{6} K\left(\frac{x_j - x_i}{h}\right)} = 2 \cdot \frac{3}{4} + 4 \cdot \frac{3}{5} = 3{,}9$$

$$H = \begin{bmatrix} 3/4 & 1/4 & 0 & 0 & 0 & 0 \\ 1/5 & 3/5 & 1/5 & 0 & 0 & 0 \\ 0 & 1/5 & 3/5 & 1/5 & 0 & 0 \\ 0 & 0 & 1/5 & 3/5 & 1/5 & 0 \\ 0 & 0 & 0 & 1/5 & 3/5 & 1/5 \\ 0 & 0 & 0 & 0 & 1/4 & 3/4 \end{bmatrix}$$

T4. $y_i = (x_i + 1)^2 + \varepsilon_i$, $i = \overline{1,n}$ $\quad \mathbb{E}\varepsilon_i = 0 \quad \operatorname{Var} \varepsilon_i = \sigma^2 \quad \forall i$

$X_1, \dots X_n \sim U[0,1]$

$$h_* = \left(\frac{1}{n}\right)^{1/5} \cdot \left(\frac{\sigma^2(x) \cdot \int K^2(x)dx \int f^{-1}(x)dx}{\int x^2 K^2(x) dx \int \left(r''(x) + 2r'(x)\frac{f'(x)}{f(x)}\right)^2 dx} \right)^{1/5}$$

$r(x) = (x+1)^2 \quad r'(x) = 2(x+1) \quad r''(x) = 2$

$f(x) = \mathbb{1}\{0 \leq x \leq 1\} \quad f'(x) = 0$ $\quad \int_{\mathbb{R}} (r''(x))^2 \mathbb{1}\{x \in [0,1]\} dx = \int_0^1 2^2 dx = 4$

$\int_{\mathbb{R}} f^{-1}(x)\,dx = \int_0^1 1\,dx = 1$

$\int K^2(x)\,dx = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \cdot \int_{-\infty}^{\infty} e^{-x^2} dx = \frac{1}{2\sqrt{\pi}}$

$\int_{\mathbb{R}} K^2(x) \cdot x^2 dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-x^2/2} dx = 1$

$$h_* = \left(\frac{\sigma^2 \cdot 1/2\sqrt{\pi}}{4n}\right)^{1/5} = \left(\frac{\sigma^2}{8n\sqrt{\pi}}\right)^{1/5}$$

NPS(8)

Spearman's correlation coefficient

$(X_1, Y_1) \ldots (X_n, Y_n)$

$\underbrace{X_{(1)}}_{1} < \ldots < X_{(n)}$

$$\hat{\rho}_s = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2}\sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}}$$

$R_i = \text{rank}(X_i)$

$S_i = \text{rank}(Y_i)$

$(R_1, S_1), \ldots, (R_n, S_n)$

$$\bar{R} = \frac{1}{n}\sum_{i=1}^n R_i = \frac{1}{n}\sum_{i=1}^n i = \frac{n(n+1)}{2n} = \frac{n+1}{2}$$

$$\sum_{i=1}^n (R_i - \bar{R})^2 = \sum_{i=1}^n R_i^2 - 2\bar{R}\underbrace{\sum_{i=1}^n R_i}_{n\bar{R}} + \bar{R}^2 n =$$

$$= \sum_{i=1}^n R_i^2 - n\bar{R}^2 = \underbrace{\sum_{i=1}^n R_i^2}_{\sum i^2} - n\frac{(n+1)^2}{4} =$$

$$S = \sum_{i=1}^n i^2 \; : \; (1+i)^3 = 1 + 3i + 3i^2 + i^3 \Big| \left(\sum_{i=1}^n\right)$$

$$\sum_{i=1}^n (1+i)^3 = n + 3\sum i + 3\underbrace{\sum i^2}_{S} + \sum_{i=1}^n i^3$$

$$(1+n)^3 = n + \frac{3}{2}\cdot n(n+1) + 3S + 1$$

$1 + 3n + 3n^2 + n^3$

$$S = \frac{n}{3}\left(\underline{3 + 3n + n^2 - 1 - \frac{3}{2}n - \frac{3}{2}}\right)$$

$$= n^2 + \frac{3}{2}n + \frac{1}{2} = \frac{1}{2}(n+1)(2n+1)$$

$$S = \frac{n(n+1)(2n+1)}{6}$$

$$= \sum_{i=1}^n (R_i - \bar{R})^2 = \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4} =$$

$$= \frac{n(n+1)}{2}\left(\frac{2n+1}{3} - \frac{n+1}{2}\right) = \frac{n(n+1)(n-1)}{12} = \frac{n(n^2-1)}{12}$$

$$\hat{\rho}_s = \frac{\sum\left(R_i - \frac{n+1}{2}\right)\left(S_i - \frac{n+1}{2}\right)}{\frac{n(n^2-1)}{12}} = \frac{\sum R_i S_i - \frac{n+1}{2}\left(\overset{n(n+1)}{\sum R_i + \sum S_i}\right) + \frac{n\cdot(n+1)^2}{4}}{\frac{n(n^2-1)}{12}} =$$

$$= \boxed{\frac{12\sum R_i S_i}{n(n^2-1)} - 3\frac{n+1}{n-1}} \quad (*)$$

$$\hat\rho_s = 1 - \frac{6 \sum_{i=1}^{n} (R_i - S_i)^2}{n(n^2-1)}$$

Kendall's tau $n \le 50$

Spearman's rho $n \le 22$

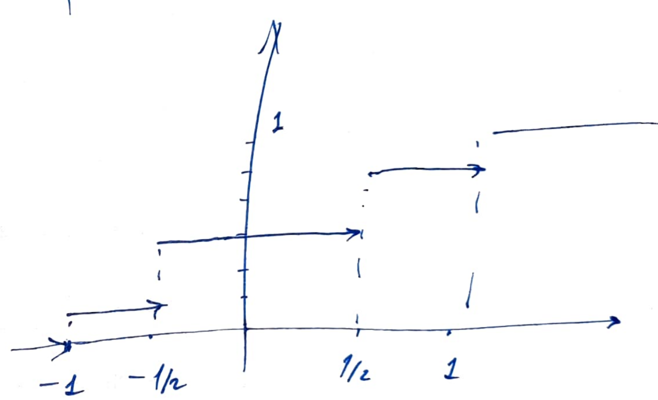
$n = 3$

$R_i = i$

$$\hat\rho_s = 1 - \frac{\sum (R_i - S_i)^2}{4}$$

(S_1, S_2, S_3)	$\sum (R_i - S_i)^2$	$\hat\rho_s$	if $X \perp\!\!\!\perp Y \Rightarrow (R_1 \dots R_n) \perp\!\!\!\perp (S_1, \dots S_n)$
1 2 3	0	1	1/6
1 3 2	2	1/2	1/6
2 1 3	2	1/2	1/6
2 3 1	6	−1/2	1/6
3 1 2	6	−1/2	1/6
3 2 1	8	−1	1/6

$$\Rightarrow \hat\rho_s = \begin{cases} 1, & 1/6 \\ 1/2, & 1/3 \\ -1/2, & 1/3 \\ -1, & 1/6 \end{cases}$$



for $n > 22$ large-sample approx.:

$$\frac{\hat\rho_s - \mathbb{E}\hat\rho_s}{\sqrt{\operatorname{Var}\hat\rho_s}} \xrightarrow[n\to\infty]{d} N(0,1) \quad // \text{ if } X \perp\!\!\!\perp Y$$

Thm: $$\sqrt{n-1} \cdot \hat\rho_s \xrightarrow[n\to\infty]{d} N(0,1)$$

$\mathbb{E}\hat\rho_s = 0$ // $X \perp\!\!\!\perp Y \Rightarrow (R_1 \dots R_n) \perp\!\!\!\perp (S_1, \dots S_n) \Rightarrow \mathbb{E}\hat\rho((S_1 \dots S_n), (R_1 \dots R_n)) = \rho(\dots) = 0$

$$(*) \quad E\left[\frac{12 \overbrace{\sum R_i S_i}^{=\sum i S_i}}{n(n^2-1)} - 3\frac{n+1}{n-1}\right] = \frac{12\, E(\sum i S_i)}{n(n^2-1)} - 3\frac{n+1}{n-1} =$$

$$\underbrace{\frac{12\sum i\, E\tilde{S}_i}{n(n^2-1)} - \frac{3(n+1)}{n-1}}_{\nearrow} = \frac{6\,n+1\,\sum i}{n(n^2-1)} - \frac{3(n+1)}{n-1} = \frac{3(n+1)\cancel{n}(n+1)}{\cancel{n}(n^2-1)} - \frac{3(n+1)}{n-1} = 0$$

$$\mathbb{E}\,\tilde{S}_i = \frac{1}{n}\sum i = \frac{n+1}{2}$$

$$\tilde{S}_i = \begin{bmatrix} 1 & , & 1/n \\ \vdots & & \\ n & , & 1/n \end{bmatrix} \quad n \to \infty \quad X \perp\!\!\!\perp Y$$

e.g. $\operatorname{Var} S_i = \frac{n^2-1}{12}$

$\operatorname{cov}(S_i, S_j) = -\frac{n+1}{12} \quad i \neq j$

$$\boxed{\operatorname{Var}\hat{\rho}_S = \frac{1}{n-1}}$$

Testing
- $\nearrow$ exact distr.
- $\rightarrow$ large-sample approx.
- $\searrow$ AS 89 $\leftarrow$ Edgeworth expansion

$\uparrow$ Adams

David, Kendall, Stuart 1951

$X_1, \dots, X_n$ i.i.d. $\quad \mathbb{E}X_i = 0 \quad \operatorname{Var} X_i = 1$

$$S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{d} N(0,1) \quad // \text{ CLT}$$

Edgeworth expansion:

$$\varphi_{S_n}(u) = \mathbb{E}\left[e^{iuS_n}\right] = \mathbb{E}\left[e^{i\frac{u}{\sqrt{n}}\sum_{i=1}^{n} X_i}\right] = \left[\mathbb{E}\, e^{i\frac{u}{\sqrt{n}}X_i}\right]^4 = \left(\varphi_{X_1}\left(\frac{u}{\sqrt{n}}\right)\right)^4$$

$$\varphi_{X_1}\left(\frac{u}{\sqrt{n}}\right) = \mathbb{E}\left[e^{\frac{iuX_i}{\sqrt{n}}}\right] = \mathbb{E}\left[1 + \frac{iuX_i}{\sqrt{n}} + \frac{(iu)^2 X_i^2}{2n} + \frac{(iu)^3 X_i^3}{6n^{3/2}} + \bar{o}(n^{-3/2})\right]$$

($\mathbb{E}\,\frac{iuX_i}{\sqrt{n}} = 0$, $\mathbb{E}X_i^2 = 1$)

$$= 1 + 0 + \frac{(iu)^2}{2n} + \frac{(iu)^3\gamma}{6n^{3/2}} + \bar{o}(n^{-3/2})$$

$$\varphi_{S_n}(u) = \left(1 + \frac{(iu)^2}{2n} + \frac{(iu)^3\gamma}{6n^{3/2}} + \bar{o}(n^{-3/2})\right)^n =$$

$$= \exp\left\{n \cdot \log\left(1 + \frac{(iu)^2}{2n} + \frac{(iu)^3\gamma}{6n^{3/2}} + \bar{o}(n^{-3/2})\right)\right\} = \left\{\log(1+x) = x - \frac{x^2}{2} \dots\right\}$$

$$= \exp\left\{n\left(\frac{(iu)^2}{2n} + \frac{(iu)^3\gamma}{6n^{3/2}} + \bar{o}(n^{-3/2})\right)\right\} = \exp\left\{-\frac{u^2}{2} + \frac{(iu)^3\gamma}{6n^{1/2}} + \bar{o}(n^{-1/2})\right\} =$$

$\xi = e^{-u^2/2} e^{(iu)^3 \gamma / n^{3/2}} + \dots \qquad \} \; e^x = 1 + x + \frac{x^2}{2!} \dots$

$$\stackrel{(=)}{} \; e^{-u^2/2} \left(1 + \frac{(iu)^3 \gamma}{6 n^{1/2}} + \bar{\bar{o}}(n^{-1/2})\right)$$

$$p_{S_n}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \, \Phi_{S_n}(u) \, du =$$

if p exists — inverse Fourier transform

char. function of $N(0,1)$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iux} e^{-u^2/2} du + \underbrace{\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} e^{-u^2/2} \frac{(iu)^3 \gamma}{6 n^{1/2}} du}_{-\frac{\partial^3}{\partial x} p^N(x)} + \bar{\bar{o}}(n^{-1/2}) \stackrel{(=)}{}$$

$$\| \qquad p^N(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$p^N(x) = \frac{1}{2\pi} \int e^{-iux} e^{-u^2/2} du$$

$$(p^N(x))' = \frac{1}{2\pi} \int e^{-iux} e^{-u^2/2} (-iu)$$

$$\stackrel{(=)}{} \; p^N(x) - \frac{d^3}{dx^3} p^N(x) \cdot \frac{\gamma}{6 n^{1/2}} + \bar{\bar{o}}(n^{-1/2})$$

Hermite polynomials:

$$\frac{d^k}{dx^k} p^N(x) = (-1)^k H_k(x) \, p^N(x)$$

$\llcorner$ Hermite polynom.

$$k=1: \; \frac{d}{dx} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} = -x \, p^N(x)$$

$H_1(x) = x$

$H_2(x) = x^2 - 1$

$H_3(x) = x^3 - 3x$

$$\frac{d^{k+1}}{dx^{k+1}} p^N(x) = (-1)^k \frac{d}{dx} (H_k(x) \cdot p^N(x))$$

$$= (-1)^{k+1} \cdot H_{k+1}(x) \, p^N(x)$$

$$H_{k+1}(x) \, p^N(x) = -\frac{d}{dx} (H_k(x) \, p^N(x))$$

$$\stackrel{(=)}{} \; p^N(x) + H_3 \, p^N(x) \cdot \frac{\gamma}{6 n^{1/2}} + \bar{\bar{o}}(n^{-1/2})$$

$$F_{S_n}(x) = \int_{-\infty}^{x} p_{S_n}(t) = F^N(x) - H_2(x) \, p^N(x) \frac{\gamma}{6 n^{1/2}} + \bar{\bar{o}}(n^{-1/2})$$

$$\max \left| F_{S_n}(x) - \left(F_N(x) - H_2(x)\, p^N(x) \cdot \frac{\gamma}{6 n^{1/2}} \right) \right| = \bar{\bar{o}}(n^{-1/2})$$

Bootstrap

$\theta(X) = E(X^4)$, $X_1, \dots, X_n$ - i.i.d.

$$\theta = \frac{\sum (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum (R_i - \bar{R})^2 \sum (S_i - \bar{S})^2}} \qquad (X_1, Y_1), \dots, (X_n, Y_n)$$

1) Make n ~~do~~ samples with replacement from the data
2) Compute $\hat{\theta}$ for each sample
3) repeate step 1-2 m times $\Rightarrow \hat{\theta}_1, \dots, \hat{\theta}_m$ bootstrap replicants
4) $\hat{\theta}_{(1)} < \dots < \hat{\theta}_m \Rightarrow (\hat{\theta}_{(k)}, \dots \hat{\theta}_{(m-k)})$ - $(1-\alpha)$ conf int.

$K = \lfloor m \cdot \frac{\alpha}{2} \rfloor$

$-1 \le \rho_s \le 1$

$$-1 \le 1 - \frac{6 \sum (R_n - S_n)^2}{n^3 - n} \le 1$$

$$0 \le \sum (R_n - S_n)^2 \le \frac{n^3 - n}{3}$$

$n = 3 \Rightarrow \max = 8$

$$\sum (R_n - S_n)^2 = \begin{cases} 0, & 1/6 \\ 2 & 1/3 \\ 6; & 1/3 \\ 8, & 1/6 \end{cases}$$

Test

	2	>2
ind. samples	Mann-Whitney	Kruskall-Wallis
dep. samples	Wilcoxon	Friedmans

PT	NPT
ind. samples t-test	MW U Test
rel. sampes' t-test	Wilcox. Signed Rank Test
One way Bet. subj ANOVA	Kruskal-Wallis Test
One way Repeat. subj ANOVA	Friedman Test
Pearson Correlation	Spearman Rank-order correlation

(1) Wilcoxon test

Paired replicated

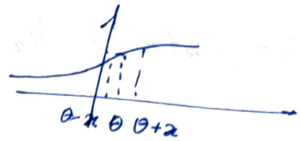
$(X_1, Y_1), \dots, (X_n, Y_n)$

$Y_1 - X_1, \dots, X_n - Y_n$

param.: $Y_i - X_i \sim N(\mu, 6^2)$ $H_0: \mu = 0$

non-param.: $\underset{F_i(x)}{\underline{Y_i - X_i}}$ is symmetric about a comma median θ

$$\boxed{F_i(\theta+x) + F_i(\theta-x) = 1}$$



If F_i

$\xi = N(\mu, \sigma^2) \Rightarrow \xi : \sigma\xi^0 + \mu \quad \sim N(0,1)$

$P\{\xi < \mu + x\} + P\{\xi < \mu - x\} =$

$= P\{\xi^0 \le \frac{x}{\sigma}\} + P\{\xi^0 \le -\frac{x}{\sigma}\}$

$= 1 - P\{\xi^0 > x/\sigma\} + P\{\xi^0 \le -x/\sigma\} = 1$

$H_0 : \theta = 0$

Wilcoxon criterion

$y_i - x_i = z_i$

$|z_{(1)}| \le \dots < |z_{(n)}|$

$$\boxed{\sum_{i=1}^{n} R_i \, \mathbb{1}\{z_i > 0\}}$$ Wilcoxon statistics

x_i	y	$y - x$	
5,2	2,1	-3.1	4
3,4	4	0,6	3
6,1	6,3	0,2	1
2,4	2,2	-0,3	2

$$SSB = \sum_{j=1}^{k} n_j (\bar{R}_j - \frac{N+1}{2})^2$$

sum of sq. between

$$SST = \sum_{j=1}^{N} \sum_{i=1}^{n_j} (R_{ij} - \frac{N+1}{2})^2$$ ← doesn't dep on data (constant)

sum of sq. total

ANOVA:

$$\frac{N-k}{k-1} \cdot \frac{SSB}{SST-SSB} \sim F_{k-1, n-k}$$

monotone w.r.t. SSB

$$= -1 + \frac{SST}{SST - SSB}$$

$$KW \xrightarrow[n_1, \dots, n_k \to \infty]{} \chi^2_{k-1}$$

NPS(8)

① Wilcoxon Test

2 dependent samples $y_i - x_i$

$(X_1, Y_1), \dots, (X_n, Y_n)$, $Z_i = Y_i - X_i$ – are jointly dep.

$H_0: \theta = 0$

$H_1: \theta \neq 0$

$\theta > 0$, $\theta < 0$

$F_i(x-\theta) + F_i(x+\theta) = 1$ $\forall x \in \mathbb{R}$

doesn't depend on i

median – θ

same for $\forall i$

F_i are cont.

(can be non-diff.)

Based on ranks

$|Z_{(1)}| < \dots < |Z_{(n)}|$

$\mathrm{rank}(|Z_i|) = R_i$

X_i	Y_i	Z_i	
2,5	1,9	−0,6	3
3,4	3,6	0,2	1
1,6	1,9	0,3	2

$$W = \sum_{i=1}^n R_i \mathbb{1}\{Z_i > 0\} = 1 + 2 = 3$$

Exact dist

$n = 3$

	W	$\theta = 0$ p
$(-, -, -)$ all negative	0	1/8
$(1, -, -)$	1	1/8
$(2, -, -)$	2	1/8
$(3, -, -)$	3	1/8
$(1, 2, -)$	3	1/8
$(1, 3, -)$	4	1/8
$(2, 3, -)$	5	1/8
$(1, 2, 3)$	6	1/8

exact dist.

$$W = \begin{cases} 0, 1, 2, 4, 5, 6, & \text{w.p. } 1/8 \\ 3 & \text{w.p. } 2/8 \end{cases}$$

Thm: under $\theta = 0$ $\dfrac{W - EW}{\sqrt{Var(W)}} \xrightarrow[n\to\infty]{} N(0,1)$

$$EW = \sum_{i=1}^n E(\xi_i) = \sum_{i=1}^n \frac{i}{2} = \frac{n(n+1)}{4}$$

$$W = \sum_{i=1}^n R_i \mathbb{1}\{Z_i > 0\} = \sum_{i=1}^n \xi_i \quad \leftarrow \text{indep.}$$

$$\xi = \begin{cases} i, & \text{w.p. } 1/2 \\ 0, & \text{w.p. } 1/2 \end{cases}$$

$$Var\, W = \sum_{i=1}^n V(\xi_i) = \sum_{i=1}^n \Big[\underbrace{E(\xi_i^2)}_{i^2/2} - \underbrace{(E(\xi_i))^2}_{i^2/4} \Big] = \frac{1}{4} \sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{24}$$

$$\frac{W - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \xrightarrow{d} N(0,1)$$

e.g.) 5,4, 5,55 , 5,60

R 1 3 2

$$S = \frac{12n}{k(k+1)} \cdot \sum_{j=1}^{k} \left(\bar{R}_j - \frac{k+1}{2}\right)^2 \longrightarrow \chi^2_{k-1}$$

if $\beta_1 = \dots = \beta_k$ $\qquad \bar{R}_j \approx \frac{k+1}{2}$

$$R_{ij} = \begin{cases} 1 \\ \vdots \\ k \end{cases} \text{ w same p.} \qquad E\bar{R}_j = \frac{1}{n}\sum_{i=1}^{n} E(R_{ij})^2 = E R_{ij} = \frac{k+1}{2}$$

Mann - Whitney

(2) $(X_1, \dots, X_n, Y_1 \dots Y_m)$

$\text{rank}(X_i) = R_i \in \{1, \dots, (n+m)\}$

$\text{rank}(Y_i) = S_i \in \{1, \dots, (n+m)\}$

$W = \sum_{j=1}^{n} S_j$ – Wilcoxon statistics

$U = \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbb{1}\{X_i < Y_j\}$ – Mann - Whitney statistics

Lemma : $W = U + \frac{m(m+1)}{2}$

$\triangleright$ $W = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \mathbb{1}\{X_i < Y_j\} + \sum_{i=1}^{m} \mathbb{1}\{Y_i < Y_j\} + 1 \right) =$

$= U + \underbrace{\sum_{j=1}^{m} \sum_{i=1}^{n} \mathbb{1}\{Y_i < Y_j\}}_{} + m = U + \frac{m(m+1)}{2}$ $\square$

$= \sum_{i=1}^{n} \sum_{i'=1}^{m} \mathbb{1}\{Y_{(i)} < Y_{(j)}\} + 1 + 2 + \dots + (m-1)$

Exact distr.

$n = 3, m = 2$

(S_1, S_2)	W
(1,2)	3
(1,3)	4
(1,4)	5
(1,5)	6
(2,3)	5
(2,4)	6
(2,5)	7
(3,4)	7
(3,5)	8
(4,5)	9

$\Delta = 0$

$W =$	w.p.
3	1/10
4	1/10
5	1/5
6	1/5
7	1/5
8	1/10
9	1/10

Large sample approximation

Thm: under $\Delta = 0$ $\frac{W - EW}{\sqrt{\text{var}\, W}} \xrightarrow[n \to \infty]{} N(0,1)$

$EW = E\left[\sum_{j=1}^{n} S_j\right] = n\, E S_1 = \frac{n(n+m+1)}{2}$

$S = 1, \dots, n+m$ w. p. $\frac{1}{n+m}$

$ES_1 = \frac{1}{n+m} \sum_{i=1}^{n+m} i = \frac{n+m+1}{2}$

ex. $\text{cov}(S_i, S_j) = -\frac{N+1}{12}$

$\text{Var}\, W = \sum_{i=1}^{m} \text{Var}\, S_j + \sum_{i \neq j} \text{cov}(S_i, S_j)$

$N = n + m$

$$Var\, S_j = E S_j^2 - (E S_j)^2 = \underbrace{\frac{1}{N}\sum_{j=1}^{N} j}_{= \frac{N(N+1)(2N+1)}{6}} - \left(\frac{N+1}{2}\right)^2 =$$

$$= m \cdot \frac{N^2-1}{12} + 2 \underbrace{C_m^2}_{m(m-1)} \cdot \left(-\frac{N+1}{12}\right) = \frac{n(N+1)}{12}\left(N-1-(m-1)\right) - \frac{nm(N+1)}{12}$$

$$\Rightarrow \frac{W - \frac{N+1}{2}}{\sqrt{\frac{nm(N+1)}{12}}} \xrightarrow[n, m \to \infty]{} N(0,1)$$

(3) Kruskal - Wallis

many indep. samples >2

e.g. $X_{ij} = \theta + \Delta_j + \varepsilon_{ij}$

θ — common median

Δ_j — effect of treatment

ε_{ij} — same distributed $\forall i, j$ (med. $= 0$)

$H_0: \Delta_1 = \Delta_2 = \dots = \Delta_k$

$H_1: \exists\, \Delta_i \neq \Delta_j, (i \neq j)$

KW

$N = n_1 + \dots + n_k$

$R_{ij} = rank(X_{ij})$

$$KW = \frac{12}{N(N+1)} \sum_{j=1}^{k} n_j \left(\bar{R}_j - \frac{N+1}{2}\right)^2$$

$\bar{R}_j$ — average rank in the group

$$\bar{R}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} R_{ij}$$

(1) $\overline{R_{ij}} = \frac{1}{N}\sum_{i=1}^{N} R_{ij} = \frac{1}{N}\sum_{i=1}^{N} i = \frac{N+1}{2}$

if $\Delta_1 = \dots = \Delta_k = 0 \Rightarrow \bar{R}_j \approx \overline{R_{ij}}$ (no dif. between groups)

(2) ANOVA

F - test for ranks