ORIS PROBLEMA

Dobimo določene podatke X. Zanima nas Y.

Primer: Problem trgovskega potnika:

X = graf z uteženimi povezavami
Y = optimalni obhod

Matematični model

$\downarrow$

Verjetnostni/stohastični model

$\downarrow$

Statistični model

Stohastični model vsebuje slučaj: Vsaj eno od količin X in Y modeliramo kot slučajno. Tako Y tipično ni natančno določen z X.

Statistični model običajno upošteva še ponavljanje.

Ena izmed glavnih nalog sklepne statistike je najti opazljivo količino $\hat{Y}$, $\hat{Y} = h(X)$, h merljiva izračunljiva funkcija, da bo $\hat{Y}$ čim bližje Y.

V Bayesovem modelu sta X in Y slučajni spremenljivki na $(\Omega, \mathcal{F}, P)$.

Če poznamo X in Y živi v $\mathbb{R}$, je smiselna napoved/prediktor za Y slučajna spremenljivka $E(Y|X) = h(X)$.

Diskreten primer: $\hat{Y} = h(X)$, $h(x) = E(Y|X=x)$

Poznamo $P(X=x|Y=y)$. Privzamemo apriorno porazdelitev napovedovalca/prediktanda Y. Po Bayesovi formuli izračunamo $P(Y=y|X=x)$. Iz slednjih dobimo napoved $E(Y|X=x)$.

Fisherjeva statistika na $(\Omega, \mathcal{F})$ pa gleda celo družino verjetnostnih mer P_θ, $\theta \in \Theta$.

Splošno: $X = F(\omega, \theta)$, $Y = G(\omega, \theta)$

Dostikrat je Y karakteristika modela, torej $Y = y = g(\theta)$.

Tipično je $X = X(\omega)$.

Primer: Opazimo $X_1, X_2, \dots, X_n$, ki so neodvisne in porazdeljene kot $N(\mu, \sigma^2)$.

$$\Theta = \mathbb{R} \times (0, \infty)$$

$$\theta = (\mu, \sigma)$$

Ocenjujemo na primer:

- $g(\mu, \sigma) = \mu$
- $g(\mu, \sigma) = \sigma^2$
- $g(\mu, \sigma) = \mu^2$
- $g(\mu, \sigma) = \mu + 2\sigma$ (VaR)

Terminološka pojasnila:

	deterministična	slučajna
poznamo	konstanta	opažanje
ne poznamo	karakteristika	predikant ali šum

Če je predmet našega zanimanja karakteristika, potem namesto o prediktorju govorimo o cenilki preden opazimo podatke, in o ceni potem ko jih opazimo.

Primer: Vrnimo se k prejšnjemu primeru ...

$$X_k = \mu + \varepsilon_k, \quad k = 1, 2, \ldots, n$$

$\varepsilon_k \sim N(0, \sigma^2)$ so šumi.

Smiselna cenilka za μ je $\hat{\mu} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \overline{X}$.

VREDNOTENJE CENILK IN PREDIKTORJEV

Naj bo $\hat{y} = h(X)$ cenilka za y.

slučajna spremenljivka

Definicija: Pričakovana ali srednja kvadratična napaka cenilke $\hat{y}$ karakteristike y je:

$$MSE_\theta(\hat{y} \mid y) := E_\theta[(\hat{y} - y)^2]$$

mean squared error

ne označuje pogojevanja

pričakovana vrednost pri verjetnostni meri P_θ

Definicija: Korenjena pričakovana kvadratična napaka cenilke $\hat{y}$ karakteristike y je:

$$RMSE_\theta(\hat{y} \mid y) := \sqrt{MSE_\theta(\hat{y} \mid y)}$$

Definicija: Pristranskost cenilke $\hat{y}$ za y je:

$$\mathrm{Bias}_\theta(\hat{y}|y) := E_\theta(\hat{y}) - y = E_\theta(\hat{y}-y)$$

Definicija: Cenilka je **nepristranska**, če za **vse** $\theta \in \Theta$ velja:

$$\mathrm{Bias}_\theta(\hat{y}|y) = 0$$

Opazka 4.1: Če je cenilka nepristranska, je $\mathrm{MSE}_\theta(\hat{y}|y) = \mathrm{var}_\theta(\hat{y}) = \mathrm{var}_\theta(\hat{y}-y)$.

V tem primeru označimo tudi $SE := \sqrt{\mathrm{var}_\theta(\hat{y})}$, čemur pravimo **standardna napaka** (standard error).

V splošnem pa velja:

$$\begin{aligned}
\mathrm{var}_\theta(\hat{y}) &= \mathrm{var}_\theta(\hat{y}-y) \\
&= E_\theta[(\hat{y}-y)^2] - (E_\theta(\hat{y}-y))^2 \\
&= \mathrm{MSE}_\theta(\hat{y}|y) - (\mathrm{Bias}_\theta(\hat{y}|y))^2
\end{aligned}$$

$$\Rightarrow \mathrm{MSE}_\theta(\hat{y}|y) = \mathrm{var}_\theta(\hat{y}) + (\mathrm{Bias}_\theta(\hat{y}|y))^2$$

Definicija: Zaporedje cenilk $\hat{y}_1, \hat{y}_2, \dots$ je **šibko dosledno** za y, če v porazdelitvi konvergira proti y:

$$\forall \theta \in \Theta: \quad \hat{y}_n \xrightarrow[n\to\infty]{d|\theta} y = y(\theta)$$

Trditev 4.2: Brž ko je limita pričakovanih kvadratičnih napak $\lim_{n\to\infty} \mathrm{MSE}_\theta(\hat{y}_n|y) = 0$ za vse $\theta \in \Theta$, je zaporedje cenilk $\hat{y}_1, \hat{y}_2, \dots$ šibko dosledno za y.

Dokaz: Rezultat sledi iz trditve 2.18:

$W_n \geq 0$
$p > 0$

$$\lim_{n\to\infty} E(W_n^p) = 0 \Rightarrow W_n \xrightarrow[n\to\infty]{d} 0$$

Po posledici 2.16 velja:

$$X_n \xrightarrow[n\to\infty]{d} c \iff d(X_n, c) \xrightarrow[n\to\infty]{d} 0$$

Vstavimo $X_n = \hat{\gamma}_n$, $c = \gamma$, $p = 2$, $W_n = |\hat{\gamma}_n - \gamma|$, $d(a,b) = |a-b|$.

Posledica 4.3: Če je zaporedje cenilk $\hat{\gamma}_n$ za γ asimptotično nepristransko ($\lim_{n\to\infty} E_\theta(\hat{\gamma}_n) = \gamma$ za vse $\theta \in \Theta$) in če je tudi $\lim_{n\to\infty} var_\theta(\hat{\gamma}_n) = 0$ za vse $\theta \in \Theta$, je to zaporedje šibko dosledno za γ.

Primer: Naj bodo $X_1, X_2, \ldots$ enako porazdeljene s pričakovano vrednostjo μ in varianco σ^2.

$$\mu = E_\theta(X_1) = g(\theta)$$

Tedaj je $\hat{\mu} = \overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$ nepristranska cenilka za μ.

Ni pa nujno tudi šibko dosledna. Lahko je denimo $X_1 = X_2 = \ldots$ in $\sigma > 0$.

Če privzamemo, da so $X_1, X_2, \ldots$ nekorelirane, pa je $\hat{\mu}$ tudi šibko dosledna:

$$MSE(\overline{X} \mid \mu) = var(\overline{X}) = \frac{\sigma^2}{n}$$

$$RMS(\overline{X} \mid \mu) = SE(\overline{X}) = \frac{\sigma}{\sqrt{n}}$$

Ko gre $n \to \infty$, gre MSE $\to 0$, torej je $\hat{\mu}$ šibko dosledna za μ.

Če so $X_1, X_2, \dots$ neodvisne, pa je po zakonu velikih števil (opazka 2.20 (1)) to velja tudi brez predpostavke o končnem drugem momentu.

Zaporedje cenilk $X_1, X_2, \dots$ pa ni šibko dosledno za μ.

V resnici je $\overline{X}$ najboljša nepristranska linearna cenilka (NNLC, best linear unbiased estimator, BLUE) za μ.

Če je μ fiksen in so $c_1, c_2, \dots, c_n$ take konstante, da je $\hat{\mu} = c_1 X_1 + c_2 X_2 + \dots + c_n X_n$ nepristranska cenilka za μ, velja:

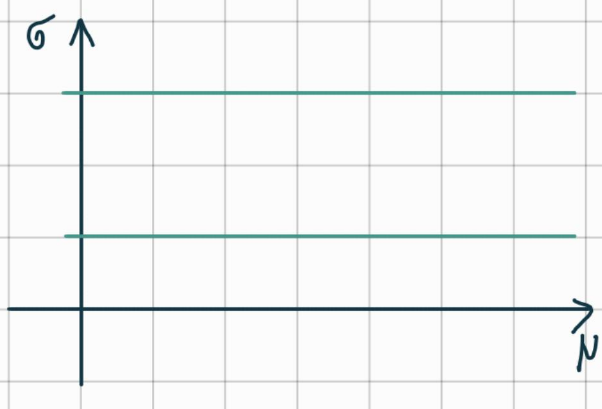
$$\underset{\substack{\| \\ \mathrm{var}(\hat{\mu})}}{\mathrm{MSE}(\hat{\mu} \mid \mu)} \geq \mathrm{MSE}(\overline{X} \mid \mu)$$

Za sledenje je dovolj privzeti nekoreliranost.

Če pa je model Gaussov ($X_1, X_2, \dots$ neodvisne in porazdeljene kot $N(\mu, \sigma^2)$) ter velja še eden izmed naslednjih pogojev:

i) σ je konstanten, torej znan

ii) μ in σ se spreminjata neodvisno drug od drugega



Potem je $\overline{X}$ tudi najboljša nepristranska cenilka za μ nasploh.

Za tem stoji Rao-Cramérjeva ocena.

(Calyampudi Radhakrishna Rao, 1920-2023)

(Carl Harald Cramér, 1893-1985)

Če dopustimo kar vse cenilke, pa tipično nobena ni enakomerno najboljša:

Padalska cenilka $\mu := 42$ je vsekakor najboljša, če je kar $\mu = 42$, sicer pa ne nujno.

Če pa ocenjujemo μ^2, je $\overline{X}^2$ tipično pristranska cenilka, saj v primeru nekoreliranosti velja:

$$E(X^2) = \mathrm{var}(\overline{X}) + (E(\overline{X}))^2 = \frac{\sigma^2}{n} + \mu^2$$

Je pa asimptotično nepristranska:

$$\mathrm{Bias}(\overline{X}^2 \mid \mu^2) = \frac{\sigma^2}{n} \xrightarrow[n \to \infty]{} 0 \quad \text{za vse } \theta \in \Theta$$

Je tudi šibko nedosledna, in sicer po trditvi 2.21:

Če gre $X_n \xrightarrow{d} X$ in je g zvezna, gre $g(X_n) \xrightarrow{d} g(X)$.

Splošni primer, ko je $\hat{Y} = h(X)$ prediktor za $Y = G(\omega, \theta)$:

Ostane pričakovana kvadratična napaka $MSE_\theta(\hat{Y} \mid Y) = E_\theta[(\hat{Y} - Y)^2]$.

Če za vsak θ velja $E_\theta(\hat{Y}) = E_\theta(Y)$, je nujno $MSE_\theta(\hat{Y}|Y) = var_\theta(\hat{Y}-Y)$, tipično pa to ni več enako $var_\theta(\hat{Y})$.

Primer: Naj bodo $X_1, X_2, \ldots$ nekorelirane in enako porazdeljene s pričakovano vrednostjo μ in varianco σ^2.

$$X_i = \mu + \varepsilon_i$$

$\varepsilon_1, \varepsilon_2, \ldots$ tvorijo beli šum, kar pomeni, da so nekorelirane z ničelno pričakovano vrednostjo in enakimi variancami.

Opazimo $X = (X_1, X_2, \ldots, X_n)$.

Zanima nas $Y = X_{n+1}$.

$$Y = \mu + \varepsilon_{n+1}$$

Prediktor za Y: $\hat{X}_{n+1} = \hat{Y} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \overline{X}$

Ta prediktor je nepristranski v prejšnjem smislu: $E(\hat{X}_{n+1}) = E(X_{n+1})$

$$MSE = var(\overline{X} - X_{n+1}) = var(\overline{X}) + var(X_{n+1})$$

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