Bayesova statistika - zapiski s predavanj prof. Smrekarja

Tomaž Poljanšek

študijsko leto 2023/24

Kazalo

Seznam uporabljenih kratic

kratica	pomen
s.v.	slučajni vektor
B	binomska porazdelitev
NEP	neodvisen in enako porazdeljen
s.s.	slučajna spremenljivka
p.v.	pričakovana vrednost

Povzetek

Poglavje 1

Uvod

Bayesova statistika je formalni okvir za „osveževanje“ vedenja/znanja o porazdelitvi nekega slučajnega vektorja.

Zgled. 1000, $\approx$ 400Č $\to$ 600B (apriorno znanje).
Izvedemo (statistični) poskus: izvlečemo 10, dobimo 6 črnih in 4 bele

1.1 Elementarna Bayesova statistika

Privzamemo popoln sistem dogodkov $E_1, E_2 \dots E_m : E_i \cap E_j = \emptyset$ za $i \neq j$ in $E_1 \cup E_2 \cup \dots \cup E_m = \Omega$.
Če imamo še neki dogodek A, velja t.i. zakon o popolni verjetnosti
$P(A) = \sum_{i=1}^{m} P(A \mid E_i) \cdot P(E_i)$ (interpretacija: 2-fazni poskus).
V Bayesovem okviru nas zanimajo $P(E_j \mid A)$ (verjetnost, da se je v „1. fazi“ zgodil E_j, če se je „2. fazi“ zgodil A). Ker je

$$P(E_j \mid A) = \frac{P(E_j \cap A)}{P(A)}$$

je

$$P(E_j \mid A) = \frac{P(A \mid E_j) \cdot P(E_j)}{P(A)} \quad \text{- elementarna pogojna verjetnost}$$

oziroma

$$P(E_j \mid A) = \frac{P(A \mid E_j) \cdot P(E_j)}{\sum_{i=1}^{m} P(A \mid E_i) \cdot P(E_i)} \quad \text{- elementarna Bayesova formula.}$$

Nadaljujemo zgled. V Bayesovi statistiki predhodno („apriorno“) vedenje formaliziramo kot realizacijo slučajnega eksperimenta. V našem primeru vpeljemo fukcijo, da smo število črnih frnikul θ (- realizacija) dobili kot rezultat slučajne spremenljivke $\Theta \in \{0, 1, 2 \ldots 1000\}$.
Informacijo $\theta \approx 400$ zakodiramo kot $E(\Theta) = 400$.
Privzamemo (kar!) $\Theta \sim B\left(1000, \frac{4}{10}\right)$
$\Longrightarrow \ P(\Theta = \theta) = \binom{1000}{\theta}\left(\frac{4}{10}\right)^{\theta}\left(1 - \frac{4}{10}\right)^{1000-\theta}$.
$P(k$ črnih od 10 izvlečenih$\mid \Theta = \theta) = \frac{\binom{\theta}{k}\binom{1000-\theta}{10-k}}{\binom{10}{k}}$ (*)
(*) pri omejitvah (k omejimo).
Osvežena porazdelitev - novo vedenje

$$P(\Theta = \theta \mid 6 \text{ črnih od 10 izvlečenih}) = \frac{P(6 \text{ črnih od 10 izvlečenih} \mid \Theta = \theta) \cdot P(B(1000, \frac{4}{10}) = \theta)}{\sum_{i=0}^{1000} P(6 \text{ črnih od 10 izvlečenih} \mid \Theta = i) \cdot P(B(1000, \frac{4}{10})) = i}.$$

Pravimo ji aposteriorna porazdelitev.

1.2 Proučevani slučajni vektor (vzročni) parametrični model

Naj bo $X = (X_1, X_2 \ldots X_n) \in \mathbb{R}^n$ preučevani slučajni vektor. Pogosto so neodvisni in enako porazdeljeni (NEP) realizacija danega slučajnega eksperimenta. S pomočjo statistike lahko „ocenjujemo“ porazdelitev slučajnega vektorja X. Zanjo privzamemo, da pripada nekemu modelu, t.j. neki množici dopustnih rešitev. Privzamemo, da je ta množica parametrizirana s parametričnim prostorom $\Theta \subset \mathbb{R}^r$. Tu si mislimo, da parameter $\theta \in \Theta$ dobimo kot realizacijo slučajnega vektorja (s.v.) Θ z vrednostmi v Θ (večinoma $r \geq 2$).

Porazelitvi s.v. X_i pogojno na $\Theta = \theta$ pravimo vzorčna porazdelitev. Privzeli bomo, da imamo gostote $f(x \mid \theta)$ ali verjetnostne funkcije

$$P(X = x \mid \theta) = f(x \mid \theta),$$

torej da velja

$$P(X \in B \mid \Theta = \theta) = \int_B f(x \mid \theta) d\nu(x)$$

(v Lebesgueovi meri) ali

$$P(X \in B \mid \Theta = \theta) = \sum_{x \in B} f(x \mid \theta).$$

Modelu pogojnih porazdelitev $(X \mid \Theta = \theta)$ pravino vrorčni model.

1.3 Apriorna in „robna“ porazdelitev

Porazdelitvi fiktivnega slučajnega vektorja Θ pravimo apriorna porazdelitev, brezpogojni (robni) porazdelitvni slučajnega vektorja X pa pravimo „robna“ porazdelitev
(*) v resnici sta obe porazdelitvi robni porazdelitvi družne porazdelitve vektorja (X, Θ) z vrednostmi v $\mathbb{R}^{n+r}$.

Zgled. Ocenjujemo Bernoullijevo porazdelitev. Predhodno vedenje je podano z apriorno prazdelitvijo na (0,1); mislimo si, da je p realizacija slučajne spremenljivke (s.s.) Π z vrednostmi v (0,1). Možnosti:

- nimamo apriornega mnenja o (dejanskem) p: tedaj bi (morda) vzeli zvezno porazdelitev z gostoto enakomerna porazdelitve,
- smo „zelo“ prepričani, da je (dejanski) $p \approx \frac{1}{2}$.

Recimo, da je $f(p)$ gostota apriorne porazdelitve. Tedaj so apriorne verjetnosti

$$P(\Pi \in (a,b)) = \int_a^b f(p) dp$$

in apriorna pričakovana vrednost

$$E(\Pi) = \int_0^1 p f(p) dp.$$

Pripomnimo, da pri $\Pi \sim U(0{,}1)$ dobimo $E(U(0{,}1)) = \frac{1}{2}$.
Privzemimo, da smo „vzorčili“ p, potem pa „neodvisno“ n-krat vržemo p-kovanec (P(cifra$=p$)), gre za slučajne spremenljivke $X_1, X_2 \dots X_n$, za katere je $(X_i \mid \Pi = p) \sim \mathit{Bernoulli}(p)$ in so $X_1 \dots X_n$ neodvisne pogojno na p. To ne pomeni, da do $X_1 \dots X_n$ brezpogojno neodvisne.
Za $i \neq j$ je

$$\begin{aligned} P(X_i = 1 \wedge X_j = 1) &= \int_0^1 P(X_i = 1 \wedge X_j = 1 \mid \Pi = p) f(p) dp = \\ &\overset{\text{pogojno neodvisne}}{=} \int_0^1 P(X_i = 1 \mid \Pi = p) P(X_j = 1 \mid \Pi = p) f(p) dp = \\ &= \int_0^1 p^2 f(p) dp = \\ &= E(\Pi^2). \end{aligned}$$

Ker je $P(X_i = 1) = \int_0^1 P(X_i = 1 \mid p) f(p) dp = \int_0^1 p f(p) dp = E(\Pi)$, je

$$Cov(X_i, X_j) = E(\Pi^2) - E(\Pi)^2 = D(\Pi)$$

za $i \neq j$, torej so X_i brezpogojno neodvisne $\iff$ $\Pi =$ konstantna (slučajna spremenljivka).
Tvorimo $X = X_1 + \cdots + X_n \in \{0, 1 \dots n\}$. To je „preučevana“ slučajna spremenljivka. Velja $(X \mid \Pi = p) \sim B(n{,}p)$. To je vzorčna porazdelitev; vzročni model je parametriziran s prostorom parametrov $(0{,}1) = \Theta$. Robna porazdelitev je podana z verjetnostmi

$$\begin{aligned} P(X = k) &= \int_0^1 P(X = k \mid p) f(p) dp = \\ &= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} f(p) dp. \end{aligned}$$

Recimo, da „opazimo“ $X = k$. Aposteriorna porazdelitev (osveženo vedenje o p) je sestavljeno iz verjetnosti

$$\begin{aligned} P(X \in (a{,}b) \mid X = k) &= \frac{P(X = k \wedge \Pi \in (a{,}b))}{P(X = k)} = \\ &= \frac{\int_0^1 P(X = k \wedge \Pi \in (a{,}b) \mid \Pi = p) f(p) dp}{P(X = k)} = \\ &= \int_a^b \frac{P(X = k \mid \Pi = p)}{P(X = k)} f(p) dp. \end{aligned}$$

Opazimo, da ima aposteriorna porazdelitev $(\Pi \mid X = k)$ gostoto

$$f_{(\pi|X)}(p \mid k) = \frac{P(X = k \mid p)f(p)}{P(X = k)}.$$

Zgornji formuli pravimo Bayesova formula.

Za številsko oceno za p bi lahko vzeli pričakovano vrednost aposteriorne porazdelitve

$$\hat{p} = E(\Pi \mid X = k) = \int_0^1 p \cdot f(p \mid k)dp.$$

Pravimo ji aposteriorna pričakovana vrednost.

Posebej priročna družina apriornih porazdelitev (za binomske vzorčne porazdelitve) je t.i. $Beta = \{Beta(a,b) \mid a,b \in (0, \infty)\}$

$$f_{Beta(a,b)}(p) = \frac{1}{B(a,b)}p^{a-1}(1-p)^{b-1}1_{(0,1)}(p)$$

(tu je $B(a,b) = \int_0^1 p^{a-1}(1-p)^{b-1}dp$).

$$E(Beta(a,b)) = \frac{a}{a+b}$$
$$D(Beta(a,b)) = \frac{ab}{(a+b)^2(a+b+1)}.$$

$D(Beta(a,b))$ predstavlja „težo“ apriornega prepričanja; večji - manj sigurni smo.

$$E(Beta(a,b)) = 0.7.$$

Aposteriorna porazdelitev ima gostoto (če je $f(p) = f_{Beta(a,b)}(p)$)

$$f(p \mid k) = \frac{\binom{n}{k}p^k(1-p)^{n-k} \cdot \frac{1}{B(a,b)}p^{a-1}(1-p)^{b-1}}{P(X = k)} =$$
$$= \text{konst.} \cdot p^{a+k-1}(1-p)^{b+n-k-1}.$$

Vidimo, da je $(\Pi \mid X = k) \sim Beta(a + k, b + n - k)$.

Aposteriorna pričakovana vrednost (p.v.) je

$$\frac{a+k}{a+b+n} = \frac{(a+b)\frac{a}{a+b} + n\frac{k}{n}}{a+b+n} =$$
$$= \frac{a+b}{a+b+n} \cdot \frac{a}{a+b} + \frac{n}{a+b+n} \cdot \frac{k}{n}.$$

Tukaj je

- $\frac{a}{a+b}$ apriorna ocena,
- $\frac{k}{n}$ vzorčna ocena in
- $\frac{a+b}{a+b+n}$ in $\frac{n}{a+b+b}$ faktorja pri konveksni kombinaciji obeh ocen.

Vzorec velik $\to$ prevlada mnenje vzorca.

1.4 Disperzija aposteriornih porazdelitev

Gre pravzaprav za disperzijo pogojnih porazdelitev. Naj bosta $X : \Omega \to \mathbb{R}^m$ in $Y : \Omega \to \mathbb{R}^n$ in naj ima (X,Y) gostoto $f_{(X,Y)}$ glede na $\mu \times \nu$ Sledita gostoti
$f_X(x) = \int f_{(X,Y)}(x,y)d\nu(y)$ za X glede na μ in
$f_Y(y) = \int f_{(X,Y)}(x,y)d\mu(x)$ za Y glede na ν. Dalje definiramo pogojni porazdelitvi $(Y \mid X = x)$ in $(X \mid Y = y)$ preko gostot

$$f_{(Y|X)}(y \mid x) = \frac{f_{(X,Y)}(X,Y)}{f_X(x)}$$

glede na ν: gostota v $X \to \mu$ in simetrično za $f_{(X|Y)}(x \mid y)$.
$P(Y \in B) \mid X = x = \int_B f_{(Y|X)}(y \mid x)d\nu(y)$ - porazdelitev, opremljena z gostoto.

Definicija 1.4.1.

$$E(Y \mid X = x) = \int y f_{(Y|X)}(y \mid x)d\nu(y).$$

y lahko zamenjamo s $h(y)$.
Pišemo $E(Y \mid X = x) = u(X)$ - h je identiteta.

Definicija 1.4.2.

$$E(Y \mid X) = u(X) : \Omega \to \mathbb{R}^n.$$

Slučajni vektor $\to$ pogojna pričakovana vrednost,
oz.

$$E(Y \mid X)(\omega) = u(X(\omega)) = E(Y \mid X = X(\omega)).$$

$E(Y \mid X)(\omega)$: funkcija na X, kompozitum.
$X(\omega)$: vrednost.

Definicija 1.4.3. Pogojno varianco slučajnega vektorja Y, pogojno na $X = x$ definiramo kot varianco pogojne porazdelitve $(Y \mid X = x)$, t.j.

$$E((Y - u(X))(Y - u(X))^T \mid X = x) =: Var(Y \mid X = x).$$

Ker je E aditivna, velja

$$E((Y - u(X))(Y - u(X))^T \mid X = x) = E(YY^T \mid X = x) - u(X)u(X)^T =: v(X).$$

$v(X)$ je $n \times n$ matrika.

Definicija 1.4.4. Pogojna varianca slučajnega vektorja Y pogojno na slučajni vektor X je

$$Var(Y \mid X) = v(X).$$

Literatura

Dodatki