1 Bayes teoria, Luku: maksimaalisen entropian priori

Olkoon X satunnaismuuttuja joka saa arvot $\{1, \dots, n\}$. Jos meillä ei ole muuta informaatiota, tapahtumat $\{X = x\}$ $x = 1, \dots, n$ ovat meidän kannalta samankaltaisia, ja siksi on loogista olettaa $P(\{X = x\}) = P(\{X = x'\})$. Additiivisuudesta seuraa $p_x = P(\{X = x\}) = 1/n$.

Tasainen jakauma on maksimaalisesti epäinformatiivinen.

Oletetaan olemme saanet tietoa, mittausdatasta tai teorian kautta että

$$E_P(f(X)) = \sum_{x=1}^{n} p_x f_x = \bar{f} \tag{1.1}$$

jossa $f_x, \bar{f} \in \mathbb{R}^d$.

Haluamme poimia tmaksimalisesti epäinformatiivinenodennäköisyys jakauma P^* joukosta

$$\{P : E_P(f(X)) = \bar{f}\}$$

.

Ensin meidän pitää vastata kysymykseen: miten mitataan kvantitatiivisti todennäköisyysjakauman epäinformatiivisuutta ?

2 Entropia

Määrätään askiomaatisesti todennäköisyysjakauman epäinformatiivisuusasteen $H(P) = H_n(p_1, \dots, p_n)$ ominaisuuksia.

$H(P)$ kutsutaan jakauman entropiaksi.

1. funktio $H_n(p_1, \dots, p_n)$ on jatkuva vektorin $P = (p_1, \dots p_n)$ suhteen.
2. $H(n) := H_n(\frac{1}{n}, \dots, 1/n)$ jossa argumenttina on tasajakauma on kasvava n:n suhteen.
3. Epäinformatiivisuusaste voidaan laskea marginaali ja ehdollista jakaumista seuraavaksi:

$$H_n(p_1, \dots, p_n) = H_2\big(p_1, (1 - p_1)\big) + (1 - p_1) H_{n-1}\left(\frac{p_2}{1 - p_1}, \dots, \frac{p_n}{1 - p_1}\right)$$

jos $Y = f(X)$ saa arvot $\{1, \dots, m\}$,

$$q_y := \sum_{x=1}^{n} p_x \mathbf{1}(f(x) = y),$$

$$H_n(p_1, \dots, p_n) = H_m(q_1, \dots, q_m) + \sum_{y=1}^{m} H_{|f^{-1}(y)|}\left(\frac{p_x}{q_y} : \, x f(x) = y\right)$$

Teoreema 2.1. *(Shannonin lause) Jos $H_n(P)$ toteuttaa edelliset ehdot*

$$H_n(p_1, \dots, p_n) = -C \sum_{x=1}^{n} p_x \log(p_x)$$

jossa $C > 0$ on vakio.

Tod. Osoitamme että väite on tosi kun $P = (p_1, \dots, p_n) \in Q^n$, jatkuvuuden oletuksesta seuraa kaikille P.

Olkoon $p_x = \frac{m_x}{M} \in \mathbb{Q}$, $m_x \in \mathbb{N}$, $M = \sum_{x=1}^n p_x$.

Oletuksesta 3) seuraa

$$H(n) = H_n(\frac{m_1}{M}, \dots \frac{m_n}{M}) = H_M(\frac{1}{M}, \dots, \frac{1}{M}) - \sum_{x=1}^{M} \frac{m_x}{M} H_{m_x}(\frac{1}{m_x}, \dots, \frac{1}{m_x})$$

$$= H(M) - \sum_{x=1}^{m} \frac{m_x}{M} H_{m_x}$$

Tasajakaumien entropiat määrävät kaikki muut.

Huomataan myös

$$H(nM) = H_{nM}(\frac{1}{nM}, \dots, \frac{1}{nM}) =$$
$$H_n(\frac{1}{n}, \dots, \frac{1}{n}) + \sum_{i=1}^{n} \frac{1}{n} H_M(\frac{1}{M}, \dots, \frac{1}{M}) = H(N) + H(M)$$

Tästä seuraa

$$H(n^k) = kH(n)$$

Funktio $H(N) = C\log(N)$ toteuttaa yhtälöä. Osoitetaan että ei ole muita vaihtoehtoja.

Olkoon $t, s \in \mathbb{N}$, $t, s \leq 2$ ja m, n jolla

$$\frac{m}{n} \leq \log(t)\log(s) < \frac{m+1}{n}$$

koska $H(n)$ on ei- vähenevä

$$H(s^m) \leq H(t^n) \leq H(s^{m+1})$$
$$\iff \quad mH(s) \leq nH(t) \leq (m+1)H(s)$$

eli

$$\frac{m}{n} \leq \frac{H(t)}{H(s)} \leq \frac{m+1}{n}$$

josta seuraa

$$\left|\frac{H(t)}{H(s)} - \frac{\log(t)}{\log(s)}\right| \leq \frac{1}{n}$$

jossa n on mielivaltainen.

Maksimaalisen entropian jakauma

Olkoon $E = \{1, \dots, n\}$, $f_x \in \mathbb{R}^d$, $x = 1, \dot{;} n$ ja $\bar{f} \in \mathbb{R}^d$

Etsitään todennäköisyysjakauma P joka toteuttaa $E_P f(X) = \sum_{x=1}^n f_x p_x = \bar{f}$. Tämä on olemassa jos ja vain jos $\bar{f}$ kuuluu konveksi joukon $\{f_1, \dots, f_n\}$ konveksi peittoon, mutta ei ole välttämättä yksikäsitteinen.

Etsitään minimoivan entropian ratkaisu. Maksimoidaan

$$-\sum_{x=1}^{n} p_x \log(p_x) \quad \text{rajoituksilla}$$

$$\sum_{x=1}^{n} p_x = 1, \quad E_P(f) = \sum_{x=1} f_x p_x = \bar{f} \in \mathbb{R}^d$$

Sovelletaan Lagrange kertoimien menetelmää. Jos P on $H(P)$:n lokaali maksimipiste, rajoituksilla $E_P(f(X)) = \bar{f}$,

$$\nabla_p H(p) - \lambda_0 \nabla_p \left(\sum_{x=1}^{n} p_x \right) - \sum_{i=1}^{d} \lambda_i \nabla E_P(f^{(i)}) = 0$$

Määritellään Langrangen funktio

$$\begin{aligned} &L(p_1, \dots, p_n; \lambda_1, \dots, \lambda_d) = \\ &-\sum_{x=1}^{n} p_x \log(p_x) - \lambda_0 \left(\sum_{x=1}^{n} p_x - 1 \right) - \sum_{i=1}^{d} \lambda_d \left(\sum_{x=1}^{d} f_x^{(i)} p_x - \bar{f} \right) \end{aligned}$$

ja ratkaistaan

$$\nabla_{p,\lambda} L(p_1, \dots, p_n; \lambda_1, \dots, \lambda_d) = 0$$

Saadaan

$$-\log p_x - 1 - \lambda_0 - \sum_{i=1}^{d} f_x^{(i)} \lambda_i = 0, \quad x = 1, \dots, n,$$

$$\sum_{x=1}^{n} f_x^{(i)} p_x = \bar{f}, \quad i = 1, \dots, d, \quad \sum_{x=1} p_x = 1$$

josta seuraa

$$p_x = \exp(-\sum_{i=1}^{d} f_x^{(i)} \lambda_i - \lambda_0 - 1) = \exp(-\sum_{i=1}^{d} f_x^{(i)} \lambda_i - \mu)$$

jossa $\mu = \lambda_0 + 1$, rajoitukset määrävät parametrien arvot.

Olkoon

$$Z(\lambda_1, \dots, \lambda_d) = \sum_{x=1}^{n} \exp(-\sum_{i=1}^{d} f_x^{(i)} \lambda_i)$$

Tämä normalisointi vakio kutsutaan partitio funktioksi.

Koska $\sum_{x=1}^{n} p_x = 1$, seuraa $\mu = \log Z(\lambda_1, \dots, \lambda_d)$ Seuraa myös

$$\begin{aligned} &-\frac{\partial}{\partial \lambda_i} \log Z(\lambda_1, \dots, \lambda_d) \\ &= \frac{1}{Z(\lambda_1, \dots, \lambda_d)} \sum_{x=1}^{n} \exp(-\sum_{j=1}^{d} f_x^{(j)} \lambda_i) f_x^{(i)} = \sum_{x=1}^{n} f_x^{(j)} p_x, \\ &-\nabla \log Z(\lambda_1, \dots, \lambda_d) = E_P\big(f(X)\big) \end{aligned}$$

Entropian maksimi arvo on

$$-\sum_{i=1}^{n} p_x \log p_x = -\sum_{i=1}^{n} p_x\big(-\mu - \sum_{i=1}^{d} \lambda_i f_x^{(i)}\big) = \mu + \lambda \cdot E_P(f(X))$$

Vaihtoehtoinen todistus Olkoon $p = (p_1, \dots, p_n)$ ja $q = (q_1, \dots, q_n)$ todennäköisyysmittoja. Koska $\log(p) = \log(1 + p - 1) \leq (p - 1)$, kun $0 \geq p$

$$\sum_{x=1}^{n} p_x \log(q_x/p_x) \leq \sum_{x=1}^{n} p_x \left(\frac{q_x}{p_x} - 1\right) = 0$$

jossa yhtäsuuruus on voimassa jos ja vain jos $p_x = q_x \ \forall x$.

Seuraa

$$H_n(p_1, \dots, p_n) \leq -\sum_{x=1}^{n} p_x \log(q_x)$$

Olkoon

$$q_x = \frac{1}{Z(\lambda_1, \dots, \lambda_d)} \exp\left(-\sum_{i=1}^{d} f_x^{(i)} \lambda_i\right)$$

Seuraa

$$\begin{aligned} H_n(p_1, \dots, p_n) &\leq \sum_{x=1}^{n} p_x \sum_{i=1}^{d} f_x^{(i)} \lambda_i + \log Z(\lambda_1, \dots, \lambda_d) \\ &= \log Z(\lambda_1, \dots, \lambda_d) + E_P(f(X)) \cdot \lambda = H_n(q_1, \dots, q_n) \end{aligned}$$

Kun P ja Q toteuttavat $E_P(f(X)) = \bar{f}$ ja $E_Q(f(X)) = \bar{f}$,

- $H_n(P) \leq H_n(Q)$
- ja $H_n(P) = H_n(Q)$ jos ja vain jos $P = Q$.

Esim. Olkoon $X \in \{1, \dots, 6\}$ nopan heitton tulos. Jos noppa on reilu, $E_P(X) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 7/2 = 3.5$.

Oletamme nyt että noppa ei ole reilu, jolla $E_P(X) = \mu = 3$. Mikä on maksimin entropian jakauma ?

R $p_x = \frac{1}{Z(\lambda)} \exp(\lambda x)$.

jossa $Z(\lambda) = \sum_{x=1}^{n} \exp(\lambda x) = \frac{\exp(n\lambda) - 1}{1 - \exp(-\lambda)}$, $n = 6$ ja

$$\begin{aligned} \mu = E_P(X) = \frac{d \log Z(\lambda)}{d\lambda} &\iff \frac{n \exp(n\lambda)}{\exp(n\lambda) - 1} - \frac{\exp(-\lambda)}{1 - \exp(-\lambda) -} = \mu \\ &\iff (n - \mu)\theta^{n+1} + (\mu - n - 1)\theta^n + \mu\theta + (1 - \mu) = 0 \end{aligned}$$

jossa $\theta = \exp(\lambda)$. Kun $n = 6, \mu = 3$ saadaan polynomiaalinen yhtälö

$$3\theta^7 - 4\theta^6 + 3\theta - 2 = 0$$

Matlabilla saadaan kompleksi ratkaisut

```
» roots( [ 3 -4 0 0 0 0 3 -2] )

-0.8099 + 0.4848i
-0.8099 - 0.4848i
0.0567 + 0.9422i
0.0567 - 0.9422i
1.0000 + 0.0000i
1.0000 - 0.0000i
0.8398
```

Koska $\mu = 3 < 3.5$ joka on tasajakauman odotusarvo, $\lambda^* < 0$ on odottavissa.
$\theta^* = 0.8398$ on positiivinen ratkaisu ja $\lambda^* = \log(\theta^*) = -0.1746$.
$Z(\lambda^*) = 3.4029$ ja $\mu^* = \log(Z(\lambda^*)) = 1.2246$.
Siis $p_x = \exp(-0.1746x - 1.2246)$,
$(p_1, \ldots, p_6) = (0.2468, 0.2072, 0.1740, 0.1461, 0.1227, 0.1031)$.
Matlabilla voidaan tarkistaa

```
» sum( exp( lambda*(1:6))'/z )
ans =
1.0000
» (1:6)*exp( lambda*(1:6))'/z
ans =
3.0000
```