4./

1: V V V V

Învăţare automată

— Licenţă, anul III, 2016-2017, examenul parţial II —

Nume student: VIZITIU MONICA-MARIA Grupa: B1

1. (Distribuţia geometrică: estimarea parametrului, în sens MLE şi respectiv în sens MAP)

Considerăm $X_1, \ldots, X_n$ variabile aleatoare independente, toate urmând *distribuţia geometrică* (discretă) de parametru θ. Aceasta înseamnă că pentru oricare variabilă X_i şi pentru orice număr natural k avem $P(X_i = k) = (1-\theta)^k\theta$.[1]

a. Fie un set de date D, conţinând „observaţiile" $D = \{X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n\}$. Scrieţi expresia funcţiei de log-verosimilitate $\ell_D(\theta)$, ca funcţie de D şi θ. Este oare valoarea acestei funcţii afectată de ordinea în care sunt „observate" cele n variabile?

b. Pornind de la funcţia $\ell_D(\theta)$ dedusă la punctul precedent, calculaţi θ_{MLE}, estimarea de verosimilitate maximă (engl., Maximum Likelihood Estimation, MLE) pentru parametrul θ.

c. Fie următoarea secvenţă de 15 „observaţii":

$$X = (0, 21, 23, 8, 9, 2, 9, 0, 7, 8, 20, 9, 7, 4, 17).$$

Aplicând [eventual] formula dedusă la punctul precedent, calculaţi valoarea lui θ_{MLE}, mai întâi pentru mulţimea formată din primele cinci „observaţii", adică $\{(0, 21, 23, 8, 9)\}$, apoi pentru primele zece „observaţii" şi, în final, pentru toate cele cincisprezece „observaţii".

d. Pentru estimarea în sensul probabilităţii maxime a posteriori (engl., Maximum A posteriori Probability, MAP) a parametrului θ, vom folosi ca distribuţie *a priori* distribuţia (continuă) Beta.[2] Funcţia de densitate de probabilitate pentru distribuţia Beta este

$$p(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)},$$

unde $B(\alpha,\beta)$ este funcţia Beta de argumente $\alpha, \beta \in \mathbb{R}$.[3]
Deduceţi formula de calcul pentru θ_{MAP}, estimarea de probabilitate maximă a posteriori pentru parametrul θ.

e. Similar cu cerinţa de la punctul ~~b~~ c), calculaţi valorile celor trei estimări în sens MAP pentru parametrul θ, folosind [de fiecare dată] următoarele valori pentru parametrii distribuţiei Beta: $\alpha = 1$ şi $\beta = 2$.

[1] Simplu spus, distribuţia gemetrică poate fi gândită ca modelând următorul experiment aleatoriu: Fie o monedă a cărei probabilitate de apariţie a feţei-stemă este θ. Aruncăm moneda de una sau mai multe ori, până când apare stema. Notăm numărul de aruncări care au precedat apariţia stemei cu k. Acest număr (k) va fi [asociat cu] valoarea unei variabile aleatoare X (ca în enunţ), despre care spunem că urmează distribuţia geometrică. După cum s-a precizat deja mai sus, $P(X = k) = (1-\theta)^k\theta$.

[2] Distribuţia Beta este adeseori folosită ca „distribuţie conjugată" [în contextul estimării parametrilor în sens MAP] nu doar pentru distribuţia geometrică ci şi pentru distribuţia Bernoulli şi, mai general, pentru distribuţia categorială.

[3] Funcţia Beta se defineşte astfel:

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

cu $\Gamma(x) = (x-1)!$ pentru orice $x \in \mathbb{N}^*$.

a) $\ell_D(\theta) = \ln(L(\theta))$

$L(\theta) = P(\text{Date}|\theta) = P(D|\theta) = P(x_1, x_2, \ldots, x_n|\theta) \overset{x_1,\ldots,x_n \text{ ind}}{=} \prod_{i=1}^{n} P(x_i|\theta) =$

$= P(x_1|\theta)\cdot P(x_2|\theta)\cdot\ldots\cdot P(x_n|\theta) = \prod_{i=1}^{n}(1-\theta)^{k_i}\cdot\theta = \theta^n \cdot \prod_{i=1}^{n}(1-\theta)^{k_i}$

$\ell_D(\theta) = \ln\left(\theta^n \cdot \prod_{i=1}^{n}(1-\theta)^{k_i}\right) = \ln\theta^n + \ln\left(\prod_{i=1}^{n}(1-\theta)^{k_i}\right)$

$= n\ln\theta + \sum_{i=1}^{n}\ln(1-\theta)^{k_i} = n\ln\theta + \sum_{i=1}^{n}k_i\cdot\ln(1-\theta)$

Valoarea funcţiei nu este afectată de ordinea în care sunt observate cele n variabile, deoarece ele sunt independente şi se face suma din produsul ~~valorilor~~ variabilei şi $\ln(1-\theta)$.

Deci, $\boxed{\ell_D(\theta) = n\ln\theta + \sum_{i=1}^{n}k_i\cdot\ln(1-\theta)} = n\ln\theta + \ln(1-\theta)\sum_{i=1}^{n}k_i$

b) $\theta_{MLE} = ?$ (Notăm $\hat\theta$)

$\ell'_D(\hat\theta) = n\cdot\frac{1}{\hat\theta} + \sum_{i=1}^{n}k_i\cdot\frac{1}{1-\hat\theta}\cdot(-1) = \frac{n}{\hat\theta} - \sum_{i=1}^{n}k_i\cdot\frac{1}{1-\hat\theta}$

$\ell''_D(\hat\theta) = -n\cdot\frac{1}{\hat\theta^2} - \sum_{i=1}^{n}k_i\cdot\frac{-1}{(1-\hat\theta)^2}\cdot(-1) =$

$= \underbrace{\frac{-n}{\hat\theta^2}}_{<0} - \underbrace{\sum_{i=1}^{n}k_i\cdot\frac{1}{(1-\hat\theta)^2}}_{(k_i\geq 0)\ >0} < 0 \Rightarrow \ell''_D$ concavă $\Rightarrow$

$\Rightarrow \ell'_D(\hat\theta) = 0$

$\Leftrightarrow \frac{n}{\hat\theta} - \sum_{i=1}^{n}k_i\cdot\frac{1}{1-\hat\theta} = 0 \Leftrightarrow \frac{n}{\hat\theta} = \sum_{i=1}^{n}k_i\cdot\frac{1}{1-\hat\theta} \Leftrightarrow$

$\Leftrightarrow \frac{n}{\hat\theta} = \frac{1}{1-\hat\theta}\sum_{i=1}^{n}k_i \Leftrightarrow n(1-\hat\theta) = \hat\theta\sum_{i=1}^{n}k_i \Leftrightarrow$

$\Leftrightarrow n - n\hat\theta = \hat\theta\sum_{i=1}^{n}k_i \Leftrightarrow n = n\hat\theta + \hat\theta\sum_{i=1}^{n}k_i \Leftrightarrow \hat\theta = \frac{n}{n+\sum_{i=1}^{n}k_i}$

Deci $\boxed{\theta_{MLE} = \frac{n}{n+\sum_{i=1}^{n}k_i}}$ ✓

Reminder: Maximum A posteriori Probability Estimation

Now we assume that we have some prior knowledge about the true parameter θ. We express it by treating θ itself as a random variable and definining a prior probability distribution over it. Precisely, we suppose that the data $X_1, \ldots, X_n$ are drawn as follows:

- θ is drawn from the prior probability distribution
- Then $X_1, \ldots, X_n$ are drawn independently from a Geometric distribution with θ as the parameter.

Now, both X_i and θ are random variables, and they have a joint probability distribution. We now estimate θ as follows

$$\theta_{MAP} = \arg\max_{\theta} P(\theta | X_1, \ldots, X_n).$$

This is called Maximum A posteriori Probability (MAP) estimation. Using Bayes rule, we can rewrite the posterior probability as follows:

$$P(\theta|X_1,\ldots,X_n) = \frac{P(X_1,\ldots,X_n|\theta)P(\theta)}{P(X_1,\ldots,X_n)}$$

Applying this to the MAP estimate, we get the following expression. Notice that we can ignore the denominator since it is not a function of θ.

$$\begin{aligned}\theta_{MAP} &= \arg\max_{\theta} P(X_1,\ldots,X_n|\theta=\theta)P(\theta)\\ &= \arg\max_{\theta} L(\theta)P(\theta)\\ &= \arg\max_{\theta} \left(\ell(\theta) + \log P(\theta=\theta)\right),\end{aligned}$$

where $L(\theta)$ is the likelihood function.

Thus, the MAP estimator maximizes the sum of the log-likelihood and the log-probability of the prior distribution on θ. When the prior is a continuous distribution with density function p, we have

$$\theta_{MAP} = \arg\max_{\theta} \left(\ell(\theta) + \log p(\theta)\right).$$

c) $X = \{0, 21, \cancel{23}\, 23, 8, 9, 2, 9, 0, 7, 8, 20, 9, 7, 4, 17\}$

I $\{0, 21, 23, 8, 9\}$

$$\Theta_{MLE} = \frac{n}{n+\sum_{i=1}^{n} k_i} = \frac{5}{5+(0+21+23+8+9)} = \frac{5}{5+61} = \frac{5}{66} = 0{,}075$$

$$\boxed{\Theta_{MLE} = 0{,}075}$$

II $\{0, 21, 23, 8, 9, 2, 9, 0, 7, 8\}$

$$\Theta_{MLE} = \frac{n}{n+\sum_{i=1}^{n} k_i} = \frac{10}{10+(0+21+23+8+9+2+9+0+7+8)} = \frac{10}{10+87} = \frac{10}{97} = 0{,}103$$

$$\boxed{\Theta_{MLE} = 0{,}103}$$

III $\{0, 21, 23, 8, 9, 2, 9, 0, 7, 8, 20, 9, 7, 4, 17\}$

$\Theta_{MLE} =$

3.

III $$\Theta_{MLE} = \frac{n}{n+\sum_{i=1}^{n} k_i} = \frac{15}{15+(0+21+23+8+9+2+9+0+7+8+20+9+7+4+17)}$$

$$= \frac{15}{15+\cancel{144}} = \frac{15}{159} = 0{,}094$$

$$\boxed{\Theta_{MLE} = 0{,}094}$$

d) $$\Theta_{MAP} = \arg\max P(\Theta \mid Date) = \arg\max \frac{P(Date \mid \Theta)\cdot P(\Theta)}{\underbrace{P(Date)}_{\text{constantă}}} \Rightarrow$$

$$\Rightarrow \Theta_{MAP} = \arg\max P(Date \mid \Theta)\cdot P(\Theta)$$

$$\cancel{\Theta_{MAP}}\; L(\Theta_{MAP}) = L(\Theta_{MLE})\cdot P(\Theta) = \Theta^n \cdot \left(\prod_{i=1}^{n} (1-\Theta)^{k_i}\right) \cdot \frac{\Theta^{\alpha-1}(1-\Theta)^{\beta-1}}{B(\alpha,\beta)}$$

$$\Leftrightarrow L(\Theta_{MAP}) = \Theta^n \cdot \frac{\Theta^{\alpha-1}(1-\Theta)^{\beta-1}}{B(\alpha,\beta)} \cdot \prod_{i=1}^{n} (1-\Theta)^{k_i}$$

$$\ell(\Theta_{MAP}) = \ln(L(\Theta_{MAP})) = n\ln\Theta + (\alpha-1)\ln\Theta + (\beta-1)\ln(1-\Theta) - \ln(B(\alpha,\beta)) + \sum_{i=1}^{n} \ln(1-\Theta)\cdot k_i =$$

$$= (n+\alpha-1)\ln\Theta + (\beta-1)\ln(1-\Theta) - \ln(\beta(\alpha,\beta)) + \ln(1-\Theta)\sum_{i=1}^{n} k_i$$

$$\ell'(\Theta_{MAP}) = (n+\alpha-1)\cdot\frac{1}{\Theta} + (\beta-1)\cdot\frac{-1}{1-\Theta} - 0 + \frac{-1}{1-\Theta}\sum_{i=1}^{n} k_i$$

$$= \frac{n+\alpha-1}{\Theta} - \frac{\beta-1}{1-\Theta} - \frac{1}{1-\Theta}\sum_{i=1}^{n} k_i$$

$$\ell''(\Theta_{MAP}) = -\frac{n+\alpha-1}{\Theta^2} - \frac{-(\beta-1)}{(1-\Theta)^2}\cdot(-1) - \frac{-1}{(1-\Theta)^2}\cdot(-1)\sum_{i=1}^{n} k_i$$

$$= -\frac{n+\alpha-1}{\Theta^2} - \frac{\beta-1}{(1-\Theta)^2} - \frac{1}{(1-\Theta)^2}\sum_{i=1}^{n} k_i < 0 \text{ (evident)}$$

? dacă...

$\Rightarrow \ell''$ concavă $\Rightarrow \ell'(\Theta_{MAP}) = 0 \Leftrightarrow$

$$\Leftrightarrow \frac{n+\alpha-1}{\Theta_{MAP}} - \frac{\beta-1}{1-\Theta_{MAP}} - \frac{1}{1-\Theta_{MAP}}\sum_{i=1}^{n} k_i = 0 \Leftrightarrow$$

$$\Leftrightarrow \frac{n+\alpha-1}{\Theta_{MAP}} = \frac{1}{1-\Theta_{MAP}}\cdot\left(\beta-1+\sum_{i=1}^{n} k_i\right) \Leftrightarrow$$

$$\Leftrightarrow n+\alpha-1 - (n+\alpha-1)\cdot\Theta_{MAP} = \Theta_{MAP}\left(\beta-1+\sum_{i=1}^{n} k_i\right)$$

$$\Leftrightarrow \boxed{\Theta_{MAP} = \frac{n+\alpha-1}{\beta-1+\sum_{i=1}^{n} k_i + n+\alpha-1}}$$

4.

Exercițiul 1, punctul ②

$\alpha = 1,\ \beta = 2$

$$\Theta_{MAP} = \frac{n+\alpha-1}{n+\alpha+\beta-2+\sum_{i=1}^{n} k_i}$$

I $\{0, 21, 23, 8, 9\}$ ~~2, 9, 0, 78~~

$$\Theta_{MAP} = \frac{5+\cancel{1}-\cancel{1}}{5+1+\cancel{2}-\cancel{2}+(0+21+23+8+9)} = \frac{5}{6+61} = \frac{5}{67} = 0{,}074$$

$\boxed{\Theta_{MAP} = 0{,}074}$

II $\{0, 21, 23, 8, 9, 2, 9, 0, 78\}$ $\boxed{\Theta_{MAP} = 0{,}102}$

$$\Theta_{MAP} = \frac{10+\cancel{1}-\cancel{1}}{10+1+\cancel{2}-\cancel{2}+\sum_{i=1}^{10} k_i} = \frac{10}{11+87} = \frac{10}{98} = 0{,}102$$

III $\{0, 21, 23, 8, 9, 2, 9, 0, 78, 20, 9, 7, 4, 17\}$

$$\Theta_{MAP} = \frac{15+\cancel{1}-\cancel{1}}{15+1+\cancel{2}-\cancel{2}+\sum_{i=1}^{15} k_i} = \frac{15}{16+144} = \frac{15}{160} = 0{,}093$$

=> $\boxed{\Theta_{MAP} = 0{,}093}$

Se observă că diferența dintre Θ_{MLE} și Θ_{MAP} e foarte mică, ~~deci~~ cu 0,001 Θ_{MLE} este mai mare

$\Theta_{MLE} = \Theta_{MAP} + 0{,}001$

Foarte bine!