3: / 4: / 2:

1: [illegible]

Învăţare automată

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

Nume student: MĂRCULEŢ DAN **Grupa:** A4

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 $\beta = 2$.

e) pentru 15: $\hat\theta^{(15)}_{MAP} = \dfrac{15}{16+144} = 0.0937$ ✓

[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 $x \in \mathbb{N}^*$.

e) $\alpha = 1, \beta = 2 \Rightarrow \hat\theta_{MAP} = \dfrac{m + \cancel{1} - \cancel{1}}{m + \cancel{1} - \cancel{1} + 2 - 1 + \sum_{i=1}^m k_i} = \dfrac{m}{m + 1 + \sum_{i=1}^m k_i}$

pentru 5: $\hat\theta^{(5)}_{MAP} = \dfrac{5}{6+61} = 0.0746$ ✓

pentru 10: $\hat\theta^{(10)}_{MAP} = \dfrac{10}{11+87} = 0.10204$ ✓

1.

a) $L(\hat\theta) = P(D|\hat\theta) = P(X_1 = k_1, \ldots, X_m = k_m|\hat\theta) \overset{①}{=} \prod_{i=1}^{M} P(X_i = k_i|\hat\theta) = \prod_{i=1}^{M} (1-\hat\theta)^{k_i}\hat\theta$

① Observaţiile sunt independente condiţional în raport cu parametrul de estimat

$\ell_D(\hat\theta) = \ln L(\hat\theta) = \ln P(D|\hat\theta) = \ln \prod_{i=1}^{M} P(x_i = k_i|\hat\theta) = \sum_{i=1}^{M} \ln P(x_i = k_i|\hat\theta) =$

$= \sum_{i=1}^{M} \ln\left((1-\hat\theta)^{k_i}\hat\theta^1\right) = \sum_{i=1}^{M} k_i \ln(1-\hat\theta) + \ln\hat\theta = m\ln\hat\theta + \ln(1-\hat\theta)\cdot\sum_{i=1}^{m} k_i$

Valoarea funcţiei nu este afectată de ordinea observaţiilor, ~~[illegible]~~ datorită proprietăţilor de asociativitate şi comutativitate ale sumei.

b) $\ell'_D(\hat\theta) = \dfrac{m}{\hat\theta} + \left(\sum_{i=1}^{M} k_i\right)\cdot\dfrac{1}{1-\hat\theta}\cdot(-1) = \dfrac{m}{\hat\theta} - \left(\sum_{i=1}^{M} k_i\right)\dfrac{1}{1-\hat\theta}$

$\ell''_D(\hat\theta) = \dfrac{-M}{\hat\theta^2} + \left(\sum_{i=1}^{M} k_i\right)\dfrac{1}{(1-\hat\theta)^2}\cdot(-1) = \dfrac{-M}{\hat\theta^2} - \left(\sum_{i=1}^{M} k_i\right)\dfrac{1}{(1-\hat\theta)^2} \Rightarrow \ell''_D(\hat\theta) < 0 \Rightarrow$ funcţia $\ell_D(\hat\theta)$ are un singur punct de maxim şi anume: $\ell'_D(\hat\theta_{MLE}) = 0 \iff$ (vei folosi $\hat\theta$ ra mc)

$\iff \dfrac{m}{\hat\theta} - \left(\sum_{i=1}^{M} k_i\right)\dfrac{1}{1-\hat\theta} = 0 \,\Big|\cdot\hat\theta(1-\hat\theta) \iff (1-\hat\theta)m - \hat\theta\sum_{i=1}^{M} k_i = 0 \iff m - \hat\theta\left(m + \sum_{i=1}^{m} k_i\right) = 0$

$\iff \hat\theta_{MLE} = \dfrac{m}{m + \sum_{i=1}^{m} k_i} = \dfrac{1}{1+\mu_k} \in (0,1)$, $\mu_k = \dfrac{1}{m}\sum_{i=1}^{M} k_i$ ✓

c) pentru primele 5: $\mu_k^{(5)} = \dfrac{0+21+23+8+9}{5} = 12.2 \Rightarrow \hat\theta^{(5)}_{MLE} = \dfrac{1}{1+12.2} = 0.0(75)$ ✓

pentru primele 10: $\mu_k^{(10)} = \dfrac{61+2+9+0+7+8}{10} = \dfrac{87}{10} = 8.7 \Rightarrow \hat\theta_{MLE} = \dfrac{1}{1+8.7} = 0.10309$ ✓

pentru toate 15: $\mu_k^{(15)} = \dfrac{87+20+9+7+4+17}{15} = \dfrac{144}{15} = 9.6 \Rightarrow \hat\theta_{MLE} = \dfrac{1}{1+9.6} = 0.0943$ ✓

d) $\ell^{MAP}_D(\hat\theta) = \ln(L(\hat\theta)\cdot p(\hat\theta)) = \ln L(\hat\theta) + \ln p(\hat\theta) = \ell^{MLE}_D(\hat\theta) + \ln \dfrac{\hat\theta^{\alpha-1}(1-\hat\theta)^{\beta-1}}{B(\alpha,\beta)} =$

$= m\ln(\hat\theta) + \left(\sum_{i=1}^{M} k_i\right)\cdot\ln(1-\hat\theta) + (\alpha-1)\ln\hat\theta + (\beta-1)\ln(1-\hat\theta) - \ln B(\alpha,\beta) =$ şi $\Delta_k = \sum$

$= (m+\alpha-1)\ln\hat\theta + \left(\beta - 1 + \sum_{i=1}^{M} k_i\right)\ln(1-\hat\theta) - \ln B(\alpha,\beta)$ (notăm $\ell^{MAP}_D(\hat\theta) = \ell_D(\hat\theta)$ pentru cele ce urmează)

$\ell'_D(\hat\theta) = (m+\alpha-1)\hat\theta^{-1} + (\beta-1+\Delta_k)(1-\hat\theta)^{-1}\cdot(-1) = (m+\alpha-1)\hat\theta^{-1} - (\beta-1+\Delta_k)(1-\hat\theta)^{-1}$

$\ell''_D(\hat\theta) = -\underbrace{(m+\alpha-1)}_{>0}\underbrace{\hat\theta^{-2}}_{>0} - \underbrace{(\beta-1+\Delta_k)}_{>0}\cdot\underbrace{(1-\hat\theta)^{-2}}_{>0} \Rightarrow \ell''_D(\hat\theta) < 0 \Rightarrow \ell^{MAP}_D(\hat\theta)$ are un singur punct de maxim

(pt $\alpha, \beta \in \mathbb{N}^* \Rightarrow \alpha-1, \beta-1 > 0$)

$\ell'_D(\hat\theta) = 0 \iff (m+\alpha-1)\hat\theta^{-1} - (\beta-1+\Delta_k)(1-\hat\theta)^{-1} = 0 \,\Big|\cdot\hat\theta(1-\hat\theta) \Rightarrow (m+\alpha-1)(1-\hat\theta) - (\beta-1+\Delta_k)\hat\theta = 0$

$\iff m+\alpha-1 - \hat\theta(m+\alpha-1+\beta-1+\Delta_k) = 0 \Rightarrow \hat\theta_{MAP} = \dfrac{m+\alpha-1}{m+\alpha-1+\beta-1+\sum_{i=1}^{M} k_i}$ ✓

2.