

1

אמינות פרמטרים

כמה נרצה לדעת מנתונים?
מה אילו פרמטרים? אילו בעיות נכסה? נתתי משהו
הפשוטה של קיומה פרמטרים עם תצפיות רגילות

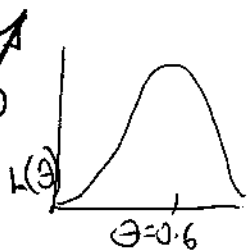
תצפיות: שנתק נסיון בינו: מחשבים $\theta = P(X="H")$
בעיות פונקציות $X[1], \dots, X[m]$

הנחות: - זכרי X יצוים קטן (D/H במקרה הזה)
- כי הצפונות בעיות טאית התפלגות
- הצפונות "בת" -
- המודל הישיר "מחלים" קמצטאות.

כמה טוב זרק מסומל של θ ? נעתי עבור H, T, T, H, H

$$L(\theta; D) = P(D|\theta) = \prod_{m=1}^m P(X[m]|\theta) = \theta^2(1-\theta)^3$$

פונקציות
פונקציות



מה עם האינו את הסברה H, T, T, H, H?

כדי קחם את הנאות מספר n_H, n_T

ולת הפלטיסטים המספיקים עבור ההתפלגות הבינומית

$$s(D) = s(D') \Rightarrow L(\theta; D) = L(\theta; D')$$

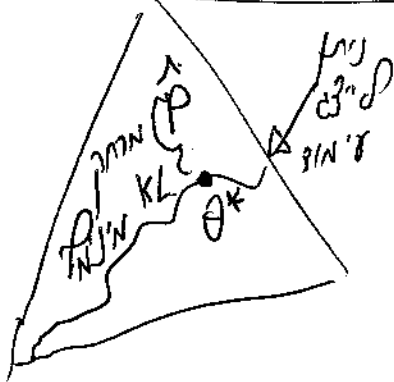
↑
sufficient
statistics

2)

$$\theta = \frac{N_H}{M}$$

MLE - מקסימום ליקור . L - פונקציית ליקור

$$\begin{aligned} \frac{\partial L}{\partial \theta} &= \frac{\partial \theta^{N_H} (1-\theta)^{N(T)} \cdot N(T)}{\partial \theta} = N(H) \cdot \theta^{N(H)-1} \cdot (1-\theta)^{N(T)} \\ &\quad - \theta^{N(H)} \cdot (1-\theta)^{N(T)-1} \cdot N(T) \\ &= N(H)(1-\theta) - \theta \cdot N(T) \\ &= N_H(1 - \frac{N_H}{M}) - N_H(1 - \frac{N_H}{M}) = 0 \end{aligned}$$



מה עולה מקורן ה-MLE באופן ?

$$D_{KL}(P||Q) = \sum_x p(x) \log \frac{p(x)}{q(x)} \quad \text{מרחק KL}$$

מדד אי-סמנטי ואי-פילי-מרחק בין התפלגויות.
 כאשר $\frac{n_x}{M} = \hat{p}(x)$ להיות מספר הופעות x באות x .

$$\begin{aligned} \frac{1}{M} \ell(\theta; D) &= \frac{1}{M} \sum_m \log P(X[m]|\theta) = \frac{1}{M} \sum_x n_x \log P(x|\theta) \\ &= \sum_x \hat{p}(x) \log p(x|\theta) = \sum_x \hat{p}(x) \log p(x|\theta) - \sum_x \hat{p}(x) \log \hat{p}(x) \\ &\quad + \sum_x \hat{p}(x) \log \hat{p}(x) \\ &= -D_{KL}[\hat{p}(x) || p(x|\theta)] - H(\hat{p}(x)) \\ \hat{\theta}_{MLE} &= \text{argmin}_{\theta} D_{KL}[\hat{p}(x) || p(x|\theta)] \quad \Leftarrow \end{aligned}$$

(3)

מטרי לרשת ביסטיות:

$$p(x; \theta) = \prod_i P(x_i | x_{pai}; \theta) = \prod_i \theta_{x_i | x_{pai}} \leftarrow \begin{matrix} \text{כמות} \\ \text{קובי"ס ומסתברות} \\ 1 - \delta \end{matrix}$$

$$\begin{aligned} L(\theta; D) &= \prod_m P(x_1[m], \dots, x_n[m]; \theta) \\ &= \prod_m \prod_i \theta_{x_i | x_{pai}} \\ &= \prod_i \prod_m \theta_{x_i | x_{pai}} \equiv \prod_i L_i(\theta_i; D) \end{aligned}$$

MLE per family

$$\Rightarrow \ell_i(\theta_i; D) = \sum_m \sum_{x_i, x_{pai}} \log \theta_{x_i | x_{pai}} = \sum_i \sum_{x_i, x_{pai}} \underbrace{\hat{p}(x_i, x_{pai})}_{\text{מספר תצפיות}} \log \theta_{x_i | x_{pai}}$$

נכנסים למקסימום כוללת כולל פרמטרים:

$$L(\theta, \lambda) = \ell(\theta; D) - \sum_i \sum_{x_{pai}} \lambda_{i, x_{pai}} \left[\sum_{x_i} \theta_{x_i | x_{pai}} - 1 \right]$$

$$\frac{\hat{p}(x_i, x_{pai})}{\theta_{x_i | x_{pai}}} - \lambda_{i, x_{pai}} = 0$$

המשוואות נגזרות
(גזור $\theta_{x_i | x_{pai}}$ ו- λ)

$$\theta_{x_i | x_{pai}} = \frac{1}{\lambda_{i, x_{pai}}} \cdot \hat{p}(x_i, x_{pai})$$

$$\textcircled{4} \int \sum_{x_i} \theta_{x_i | x_{pai}} = \frac{1}{\lambda_{i, x_{pai}}} \sum_{x_i} \hat{p}(x_i, x_{pai})$$

PP: p_i vs p_{pai}

$$\Rightarrow \lambda_{i, x_{pai}} = \hat{p}(x_{pai})$$

$$\hat{\theta}_{x_i | x_{pai}} = \frac{\hat{p}(x_i, x_{pai})}{\hat{p}(x_{pai})} = \hat{p}(x_i | x_{pai}) \Leftarrow$$

"H" למה ומה 10 המלצות מלבד מטרה 7-1 מה לפי 11 ה
 מהו P? מהו λ ? מהו θ ? מהו λ ? $\lambda = 1,000,000$ $\lambda = 1,000,000$?

Frequentist

- המטרה היא למצוא ה
 - המטרה היא למצוא ה
 - המטרה היא למצוא ה

Bayesian

- המטרה היא למצוא ה
 - המטרה היא למצוא ה
 - המטרה היא למצוא ה

$$P(\theta | D) = \frac{P(\theta) P(D | \theta)}{P(D)} \leftarrow \text{likelihood}$$

posterior P(D)

ההסתברות למקרה
 היא כפי שה המטרה

$$P(X[M+1] | X[1] \dots X[M]) = \int P(X[M+1] | \theta, X[1] \dots X[M]) P(\theta | X[1], \dots, X[M])$$

likelihood posterior

5

$P(\theta)=1$ ז'ת'ו ו'ת'ו ו'ת'ו ו'ת'ו : ל'ת'ו

MLE: $\hat{\theta} = 4/5 \iff N_T = 1 \quad N_H = 4$ ו'ת'ו ו'ת'ו

$$P(X_{[M+1]} = H | D) = \alpha \int \theta \cdot \theta^5 (1-\theta) d\theta = \alpha \left[\int \theta^6 d\theta - \int \theta^7 d\theta \right] = \alpha \left(\frac{1}{6} - \frac{1}{7} \right) = \frac{1}{42}$$

$$P(X_{[M+1]} = T | D) = \alpha \int \theta^4 (1-\theta^2) d\theta = \alpha \int \theta^4 - 2\theta^5 + \theta^6 d\theta$$
$$= \frac{1}{5} - \frac{2}{6} + \frac{1}{7} = \frac{1}{105}$$

$\Rightarrow \alpha = 30 \Rightarrow P(H) = 5/7$
 $P(T) = 2/7$

? $P(T) > P(H)$ ו'ת'ו ו'ת'ו ו'ת'ו ו'ת'ו
? $M \rightarrow \infty$ ו'ת'ו ו'ת'ו ו'ת'ו ו'ת'ו

$$P(X_{[M+1]} = H | D) = \frac{N_H + 1}{N_H + 1 + N_T + 1} \quad \text{: (Laplace) ו'ת'ו ו'ת'ו}$$

מה דושה טות יס יות כטת? טלס כריר לטת?
 טלס β , טת כריר טת / טלס $P(D|\theta)$ כרר כטו $P(D|\theta)$
 קר טת טת טת (conjugate prior)

Beta $P(\theta) \propto \theta^{\alpha_H-1} (1-\theta)^{\alpha_T-1}$

$\beta \Downarrow$

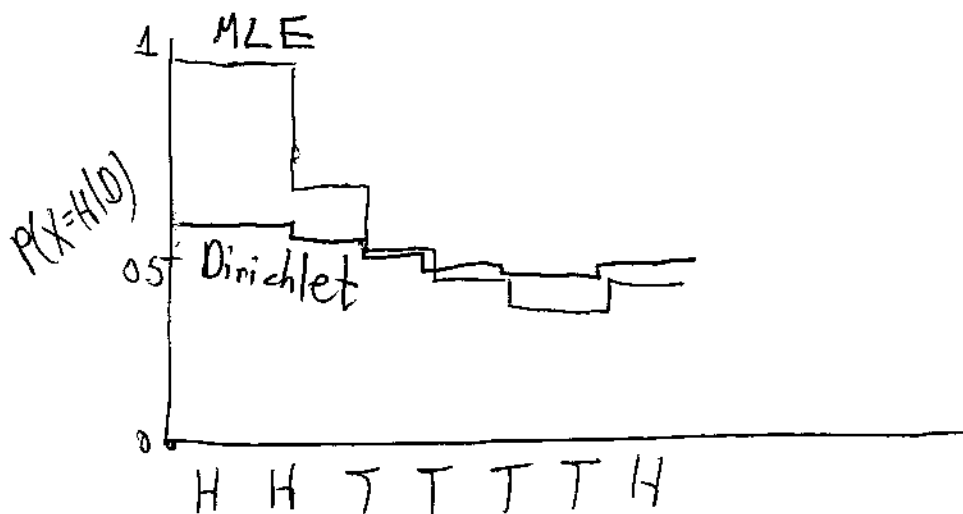
Dirichlet $P(\theta) \propto \prod_{k=1}^K \theta_k^{\alpha_k-1}$

$$P(\theta|D) \propto P(\theta) P(D|\theta) = \prod_{k=1}^K \theta_k^{\alpha_k-1} \prod_{k=1}^K \theta_k^{N_k}$$

301

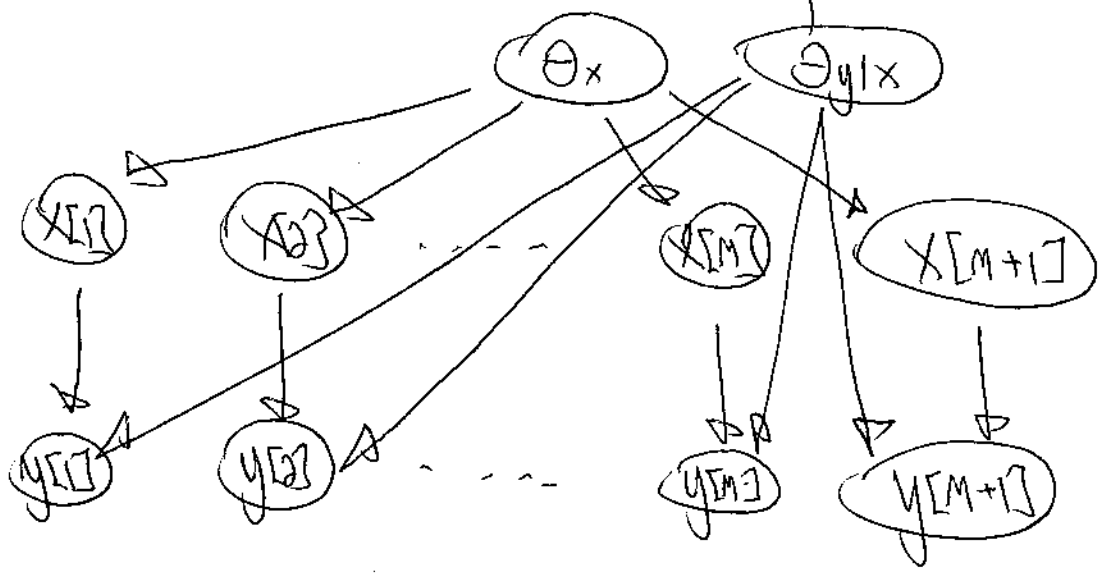
$$= \prod_{k=1}^K \theta_k^{N_k + \underbrace{\alpha_k-1}_{\text{imaginary counts}}}$$

טלס טת כריר? (1,1)? (5,5)? (1000,1000)?



7

$X \rightarrow Y$ — מודל של תלות בין X ל- Y .
 מודל של תלות בין X ל- Y — מודל של תלות בין X ל- Y .



$$\hat{\theta}_{x_i | x_{p(i)}} = \frac{\alpha(x_i, p(i)) + N(x_i, x_{p(i)})}{\alpha(p(i)) + N(p(i))}$$

איך מודל זה מתנהג?
 איך מודל זה מתנהג?