



האוניברסיטה
העברית
בירושלים
THE HEBREW
UNIVERSITY
OF JERUSALEM

אלגוריתמים בביו' חישובית

76558

הכרעה ומבחן יחס הניראות

תומי קפלן

7/1/2024

עימוד זוג רצפים ביולוגיים

- **אלגוריתם:** תכנון דינאמי [Global/Local/Overlap]
- **סקור:** תחת הנחת אי-תלות בין העמדות, מתפרק לסכום הציונים עמדה-עמדה
- **מטריצת הציונים:** $\sigma : \Sigma \times \Sigma \rightarrow \mathbb{R}$
מה משמעותה? כיצד נאמוד אותה?

קבלת החלטות אופטימלית

דוגמא 1

- ריצפנו חלבון חדש. ברצוננו לדעת מה הוא עושה.
- האם מוכרים חלבונים הדומים לו ברצף? מה תפקידם?
- מה זה "דומה"? מובהקות סטטיסטית? מודל רקע?



האוניברסיטה
העברית
בירושלים
THE HEBREW
UNIVERSITY
OF JERUSALEM

אלגוריתמים בביו' חישובית

76558

הכרעה ומבחן יחס הניראות

תומי קפלן

9/1/2024

כלל החלטה

- פונקציה שממפה תצפית לתיוג

$$\tau : \mathbb{R}^n \rightarrow \{-1, +1\}$$

- נניח שידועים לנו:

$$P(X|L = +1), P(X|L = -1)$$

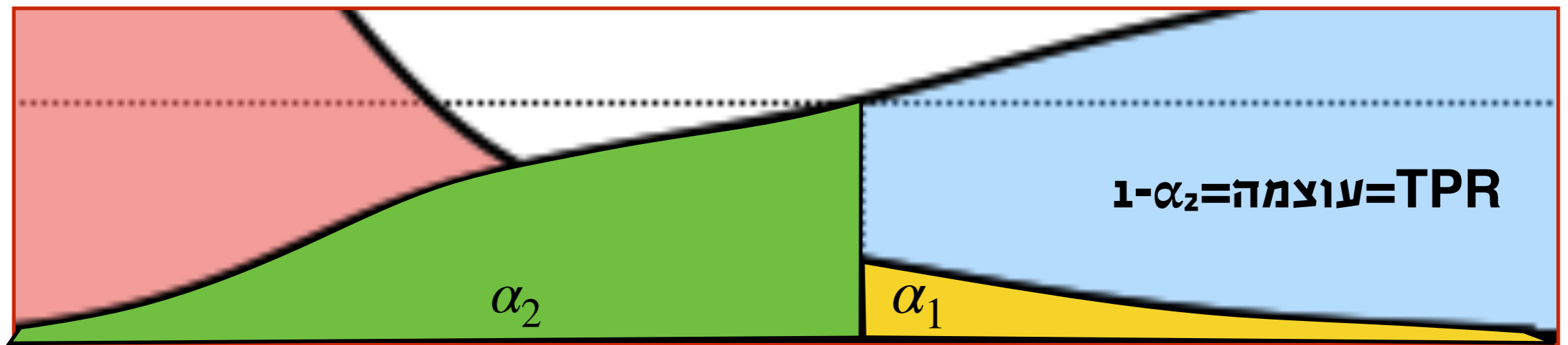
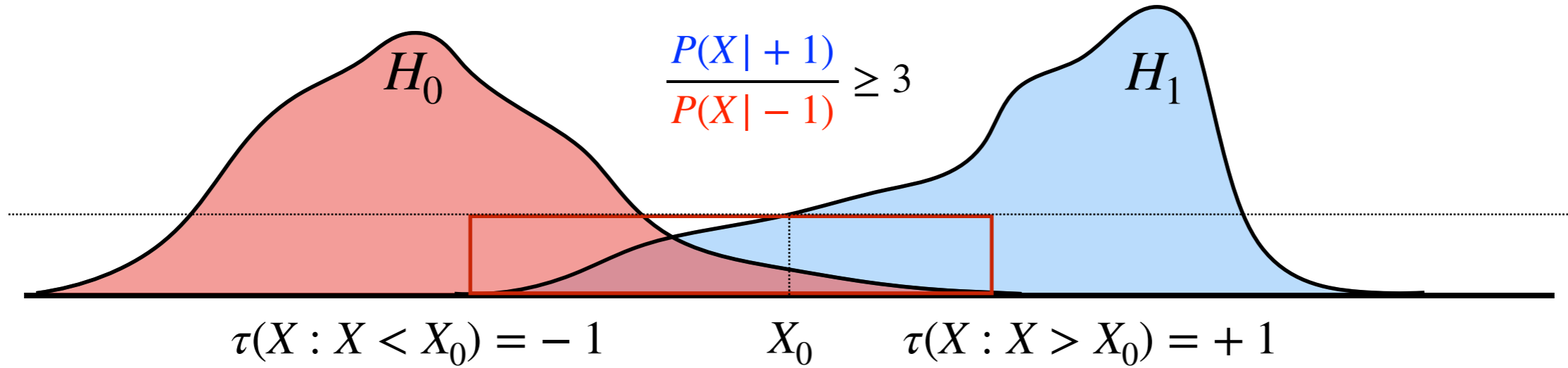
- אז אולי נבחר:

$$\tau(X) = \begin{cases} +1 & P(X|+1) \geq P(X|-1) * t \\ -1 & \text{otherwise} \end{cases}$$

$$\tau(X) = \begin{cases} +1 & \frac{P(X|+1)}{P(X|-1)} \geq t \\ -1 & \text{otherwise} \end{cases}$$

מסתבר שזה לא כזה רעיון גרוע

סוגי טעויות



טעות מסוג 2
 False Negative Rate
 קבלה מוטעית של H_0
 $\alpha_2 = P(\tau(X) = -1 | X \sim H_1)$

טעות מסוג 1
 False Positive Rate
 דחיה מוטעית של H_0
 $\alpha_1 = P(\tau(X) = +1 | X \sim H_0)$

רמת מובהקות

Sensitivity and specificity cheat sheet

		True condition			
Total population		Condition positive	Condition negative	Prevalence = $\frac{\Sigma \text{Condition positive}}{\Sigma \text{Total population}}$	Accuracy (ACC) = $\frac{\Sigma \text{True positive} + \Sigma \text{True negative}}{\Sigma \text{Total population}}$
Predicted condition	Predicted condition positive	True positive	False positive, Type I error	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{True positive}}{\Sigma \text{Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{False positive}}{\Sigma \text{Predicted condition positive}}$
	Predicted condition negative	False negative, Type II error	True negative	False omission rate (FOR) = $\frac{\Sigma \text{False negative}}{\Sigma \text{Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{True negative}}{\Sigma \text{Predicted condition negative}}$
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power = $\frac{\Sigma \text{True positive}}{\Sigma \text{Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm = $\frac{\Sigma \text{False positive}}{\Sigma \text{Condition negative}}$	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR-}}$
		False negative rate (FNR), Miss rate = $\frac{\Sigma \text{False negative}}{\Sigma \text{Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) = $\frac{\Sigma \text{True negative}}{\Sigma \text{Condition negative}}$	Negative likelihood ratio (LR-) = $\frac{\text{FNR}}{\text{TNR}}$	

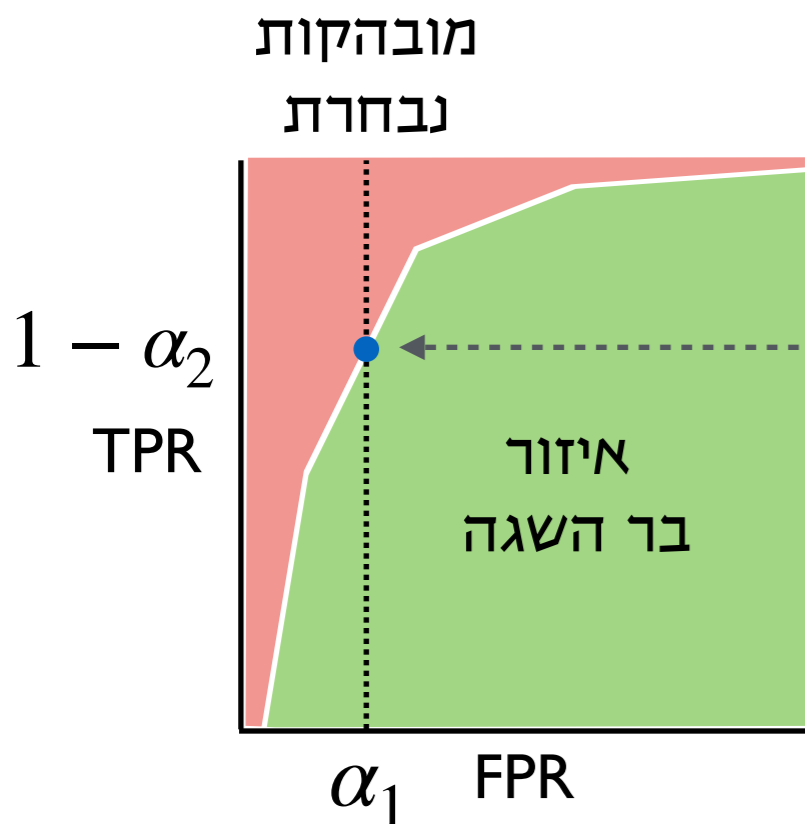
למת נימון-פירסון [33']

- מבין כל המבחנים להכרעה בין שתי ההשערות H_0 ו- H_1 , מבחן יחס הנראות הוא בעל העוצמה המקסימלית בהנתן

$$1 - \alpha_2 = \text{TPR}$$

מובהקות נתונה

$$\alpha_1 = \text{FPR}$$



$$\tau(X) = \begin{cases} +1 & \frac{P(X|+1)}{P(X|-1)} \geq t \\ -1 & \text{otherwise} \end{cases}$$

מבט בייזיאני על יחס הנראות

$$LR = \text{Likelihood Ratio} = \frac{P(X|+)}{P(X|-)}$$

• נגדיר את יחס הנראות

$$P(+|X) \stackrel{?}{>} P(-|X)$$

• כדי לסווג, בעצם נתעניין בפוסטריוור

$$LR = \frac{P(X|+)}{P(X|-)} = \frac{P(+|X)}{P(-|X)} \cdot \frac{\cancel{P(X)}}{\cancel{P(X)}} \cdot \frac{P(+)}{P(-)}$$

Likelihood Ratio Posterior Ratio Prior Ratio

$$\frac{P(+|X)}{P(-|X)} \stackrel{?}{>} 1 \equiv \frac{P(X|+)}{P(X|-)} \stackrel{?}{>} \frac{P(+)}{P(-)}$$

מבט בייזיאני על יחס הנראות

$$LLR = \log \frac{P(X|+1)}{P(X|-1)}$$

- נגדיר את לוג יחס הנראות

$$0 \leq \frac{P(+1|X)}{P(+1|X) + P(-1|X)} \leq 1$$

- ונתעניין בפוסטריוור

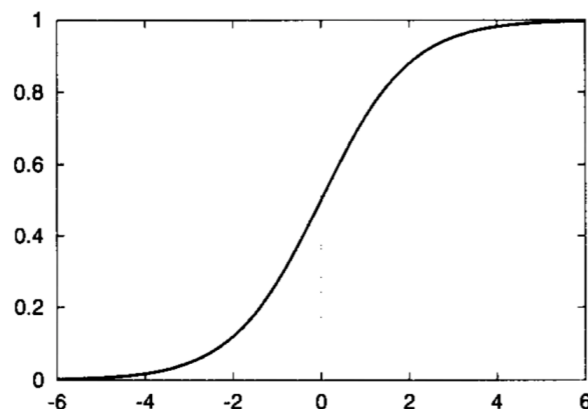
$$a = P(+|X), b = P(-|X)$$

- נסמן

$$\frac{P(+1|X)}{P(+1|X) + P(-1|X)} = \frac{a}{a+b} = \frac{1}{1 + \frac{b}{a}} = \frac{1}{1 + PR^{-1}} =$$

$$= \frac{1}{1 + \exp(-LPR)} = \frac{1}{1 + \exp(-LLR + t)}$$

פוסטריוור
 $P(+1|X)$



LLR לוג יחס הנראות

- ניזכר בפונקצית הסיגמואיד

$$s(x) = \frac{1}{1 + e^{-x}}$$

אז מה בין כללי החלטה לעימוד רצפים?

- נגדיר מחדש את שתי ההיפותיזות שלנו:

H_0 : אין קשר בין שני הרצפים

H_1 : לשני הרצפים יש אב קדמון משותף

$$LR = \frac{P(X|H_1)}{P(X|H_0)} = \frac{P_1(S, T)}{P_0(S, T)} = \frac{P_1(S, T)}{P_0(S) \cdot P_0(T)}$$

- ויחס הנראות:

$$P_1(S, T) = \prod_i P_1(S_i, T_i), \quad P_0(S) = \prod_i P_0(S_i)$$

- נניח אי-תלות בין העמדות

- ולוג יחס הנראות:

$$LLR = \log \frac{P_1(S, T)}{P_0(S) \cdot P_0(T)} = \log \frac{\prod P_1(S_i, T_i)}{\prod P_0(S_i) \cdot \prod P_0(T_i)} = \sum_i \log \frac{P_1(S_i, T_i)}{P_0(S_i) \cdot P_0(T_i)} \triangleq \sum_i \sigma(S_i, T_i)$$

כלומר ציון העימוד בין שני רצפים הוא
בדיוק לוג יחס הנראות בין שתי
ההיפותיזות (מקור משותף/לא)

והכרעה בעזרתו היא הדרך
האופטימלית לפי למת נימנ-פירסון