

שבוע 1 · Conjectures

התפלגות ברנולי, שבילי נתונים וסבירות

מהנתונים אל הפרמטר

פרופ' ניצן שחר

כיוון הזרימה של מדע אמפירי (תזכורת)

לפני שנצלול פנימה, נזכיר את סדר הדברים כפי שהם מתקיימים בטבע:

1. פרמטר (θ): הקבוע החבוי בטבע. "המספר בסוד".
2. המודל המייצר: הפונקציה שמתרגמת את החוקיות החבויה ואת המקריות לכדי מציאות נצפית.
3. משתנה מקרי (y): התצפית עצמה. התוצאה הממשית שנייצר או נמדוד. "המספר שאפשר לראות".

המודל הפשוט ביותר: התפלגות ברנולי

$$y \sim \text{Bernoulli}(\theta)$$

- y : המשתנה המקרי שלנו (0 או 1).
- θ : הפרמטר החבוי בטבע — ההסתברות ש- $y = 1$. הערך של θ נע תמיד בין 0 ל-1.



הגדרה כללית

מהי סבירות, ולמה היא פונקציה של הפרמטר ולא של הנתונים.

הבעיה האמפירית: מסתכלים הפוך

אם אנחנו יודעים מהו θ , אנחנו יכולים לחזות את התפלגות התצפיות (הסתברות).

אבל במדע, המצב הפוך: יש לנו רק את סדרת התצפיות, ואנחנו צריכים ללכת אחורה כדי לנחש מהו ה- θ .

כשאנו שואלים: “עד כמה הנתונים האלו צפויים תחת ערכים שונים של θ ?” אנו עוסקים ב**סבירות (Likelihood)**.

מהי סבירות (Likelihood)?

הגדרה

פונקציית הנראות (Likelihood) מתארת את ההסתברות לקבל את המדגם הספציפי שנאסף, תחת ערכים אפשריים שונים של פרמטר המודל.

- **נקודה קריטית:** ה-Likelihood אינה ההסתברות של הפרמטר להיות נכון!
- **ההבדל בין הסתברות לסבירות:**
 - פונקציית הסתברות (Probability) מסכמת את כל התצפיות האפשריות תחת פרמטר קבוע, ומתכנסת ל-1.
 - פונקציית סבירות (Likelihood) משווה ערכי פרמטרים שונים עבור **תצפית נתונה**. (השטח תחת עקומת ה-Likelihood לא מסתכם ל-1).



חישוב ידני באמצעות גן האפשרויות

סבירות כספירת מסלולים: שק עם ארבע גולות, ועץ שמסתעף עם כל שליפה.

הגן של שבילי הנתונים

איך מחשבים סבירות? נחזור לבסיס:

הגדרה

דברים שיש יותר דרכים שבהן הם יכולים לקרות, הם סבירים יותר.

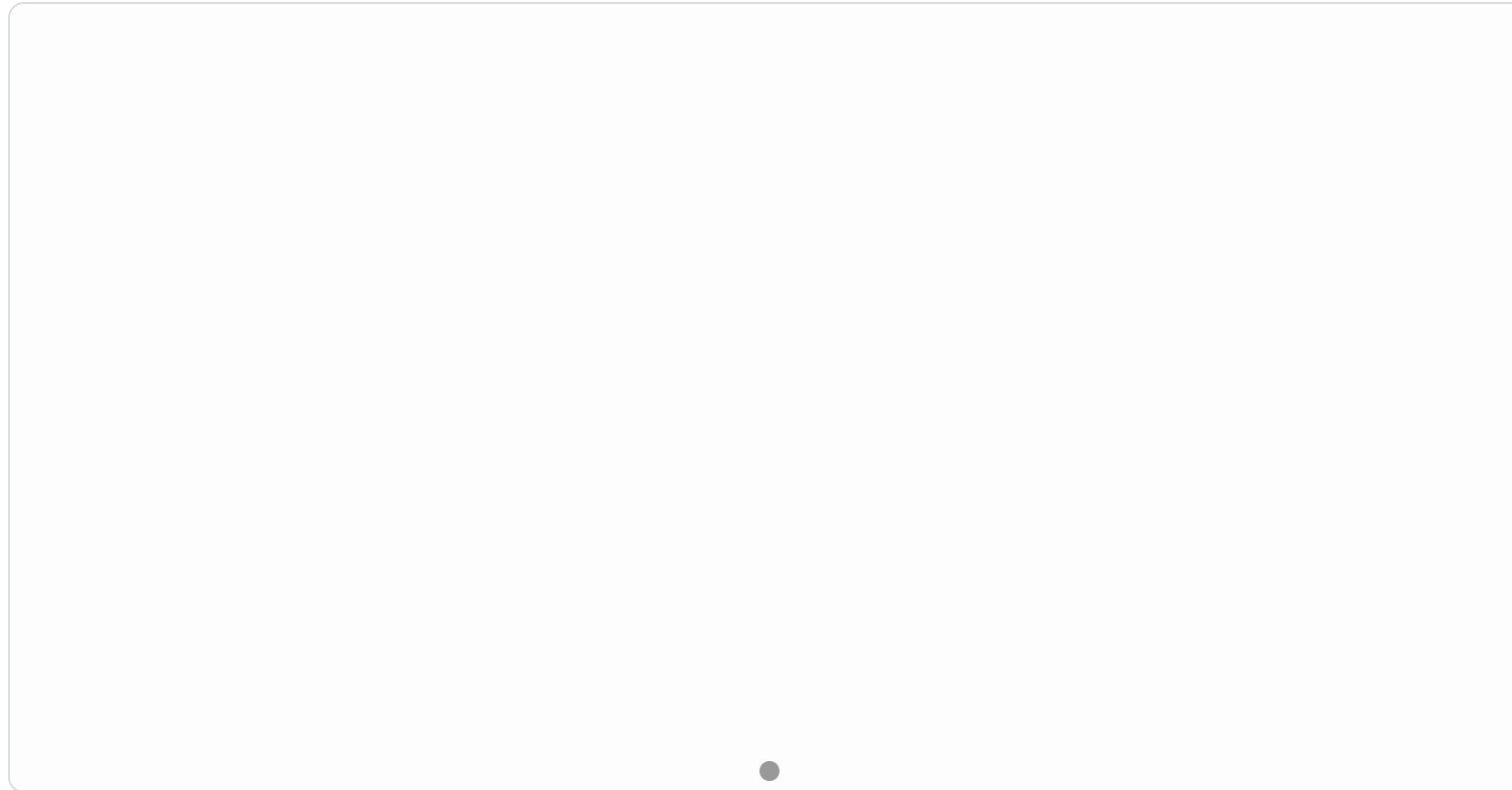
נניח שיש לנו שק ובו בדיוק 4 גולות (שחורות או לבנות). אנו דוגמים 3 פעמים (עם החזרה). התצפית שקיבלנו: שחור, לבן, שחור.

מה הסבירות שהשק מכיל גולה אחת שחורה ו-3 לבנות ($\theta = 0.25$)? בואו נשרטט את כל המסלולים האפשריים.

עץ האפשרויות: בחירת עולם ודגימה

עולם (הרכב השק): 1 שחורה, 3 לבנות ($\theta=0.25$)

הדגימה שנצפתה: שחור, לבן, שחור



●

גיזום מסלולים ☐

הצג שליפה 1

ספירת המסלולים ששרדו

→ כאשר גזמנו את העץ כך שיישארו רק המסלולים שתואמים בדיוק לתצפית שלנו (שחור → לבן → שחור), נשארו עם מעט שבילים.

איך נחשב זאת מתמטית עבור $\theta = 0.25$?

- שחור ראשון: 1 דרך (רק גולה אחת כזו)

- לבן שני: 3 דרכים (שלוש גולות כאלו)

- שחור שלישי: 1 דרך (שוב, רק גולה אחת)

סך הכל: $1 \times 3 \times 1 = 3$ מסלולים אפשריים.

השוואת השערות (סבירות הנתונים)

נבצע את הספירה עבור כל ההרכבים האפשריים של השק, עבור דגימה נבחרת:

הדגימה שנצפתה: שחור, לבן, שחור					
0	0.25	0.5	0.75	1	ההשערה (θ)
○○○○○	●○○○○	●●○○○	●●●○	●●●●●	השק
0	3	8	9	0	שבילים (Likelihood)
0%	15%	40%	45%	0%	הסיכוי היחסי שזה העולם

בהינתן התצפית שנבחרה, ההשערה $\theta = 0.75$ היא הסבירה ביותר (יש לה הכי הרבה שבילים).



חישוב רציף באמצעות חזקות

כשהפרמטר רציף, הספירה הופכת לחזקות, והסבירות הופכת לעקומה.

מעבר לסבירות כפונקציה רציפה

במציאות θ רציף בין 0 ל-1. במקום "כמות פתקים", נחשב ישירות את הסבירות: הסבירות לתצפית 1 תחת θ היא פשוט θ . הסבירות לתצפית 0 תחת θ היא $(1 - \theta)$. מכיוון שהדגימות בלתי תלויות (הסדר לא משנה את המכפלה), אנו מכפילים את ההסתברויות כפי שהכפלנו שבילים:

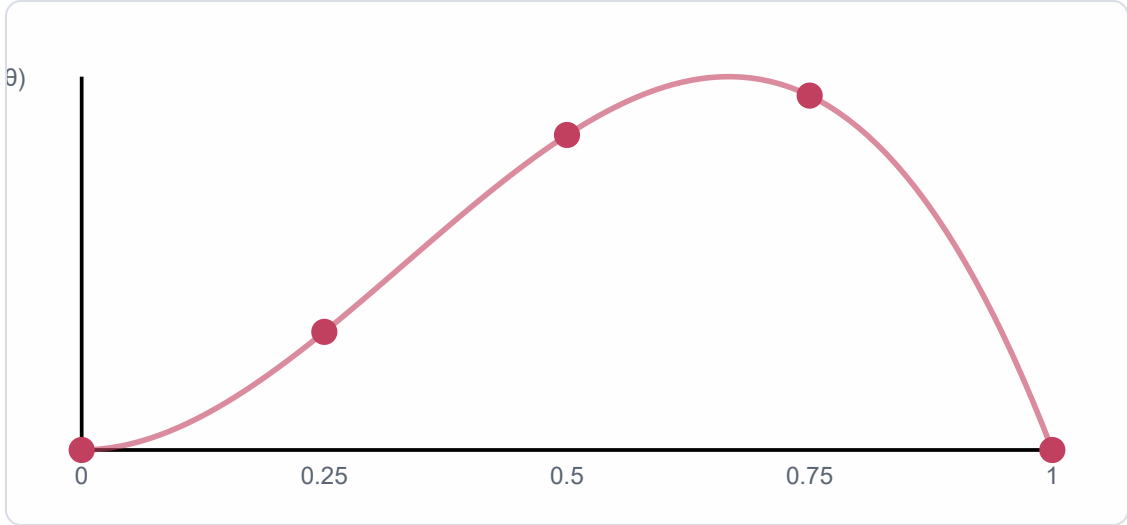
$$\mathcal{L}(\theta|\text{data}) = \theta \times (1 - \theta) \times \theta = \theta^2(1 - \theta)$$

פונקציה זו מתארת בדיוק את כמות "שבילי הנתונים" עבור כל θ אפשרי.

מנקודות בדידות אל העקומה הרציפה

לפני שנראה את העקומה המלאה, נחשב את הסבירות בכמה נקודות בדידות עבור דגימה נבחרת — ונגלה שאלו בדיוק אותם יחסים שכבר מצאנו בספירת השבילים.

הדגימה שנצפתה: שחור, לבן, שחור



0	0.25	0.5	0.75	1	θ
0.0000	0.0469	0.1250	0.1406	0.0000	$L(\theta) = \theta \times (1-\theta) \times \theta$
0.0%	15.0%	40.0%	45.0%	0.0%	הסיכוי היחסי

סיכום וחזרה

- פרמטר (θ): הקבוע החבוי בטבע, שאותו אנחנו רוצים להעריך.
- משתנה מקרי (y): התצפית הגלויה, שמשתנה בכל מדידה.
- התפלגות ברנולי: $y \sim \text{Bernoulli}(\theta)$. משתנה בינארי (0 או 1), ו- θ היא ההסתברות ש- $y = 1$.
- הסתברות מול סבירות: הסתברות מקבעת את θ ושואלת על הנתונים; סבירות מקבעת את הנתונים ושואלת על θ .
- סבירות (**likelihood**): עד כמה הנתונים שנצפו צפויים תחת כל ערך של θ .
- הגן של שבילי הנתונים: לספור את המסלולים שבהם הנתונים יכולים לקרות. יותר מסלולים = השערה סבירה יותר.
- מכפלה: כשהתצפיות בלתי תלויות, מכפילים את הסבירויות שלהן, למשל $\theta^2(1 - \theta)$.
- עקומת הסבירות: השטח תחתיה אינו 1. מה שחשוב הוא ההשוואה בין ערכי θ .

