

פרדוקס סימפסון

יכול להיות ששני סטודנט אחד קיבל בכל סמסטר ציון גבוה יותר מסטודנט אחר, אבל בממוצע השנתי הסטודנט השני קיבל ציון יותר גבוה, כי בסמסטר שבו הציונים של שניהם היו יותר גבוהים (מאשר בסמסטר הקודם) הוא לקח יותר קורסים ולכן החלק של הסמסטר הזה בממוצע השנתי יותר גבוה.

ערבוב אקראיות עם הסתברות

הרבה פעמים דברים נראים לנו לא אקראיים, למרות שהם כן אקראיים. למשל אם מוודאים 100 השערות עם רמת מובהקות 5%, כנראה שנטעה ב-5 למרות שבכל אחד מהבדיקות בדקנו מדגם גדול. דוגמה אחרת - אם מסתכלים על כמות הפעמים שכל כדור יצא בלוטו בפרק זמן מסויים, כדורים מסויימים יצאו יותר פעמים מאחרים - זה לא בהכרח אומר שיש הטייה כלפי הכדורים האלה.

השפעות מנוגדות

- מתי יש יותר סיכוי להיפגע בתאונת דרכים - בקיץ או בחורף? בחורף תנאי הנהיגה יותר קשים, אבל יש פחות אנשים על הכביש.
- מתי כדי ללכת לדיסנילנד? בoffseason יש פחות אנשים, אבל יש יותר סיכוי שמתקנים יושבתו לצורך תחזוקה
- מתי נחכה פחות זמן בתור לקופה? בלילה יש פחות אנשים, אבל יש גם פחות קופאיות

הגדרות נוספות

- Prior probability - ההסתברות שאנו נותנים למאורע לפני שיש לנו מידע עליו
- Posterior probability - ההסתברות שאנו נותנים למאורע לאחר שמתברר לנו מידע חדש
- Probability revision - התהליך של התחשבות במידע חדש לצורך הערכה מחדש של ההסתברויות

ניתוח באמצעות עץ החלטה

- שלבים: בניית עץ החלטה
- חישוב Expected Value של כל אלטרנטיבה
- בחירת האלטרנטיבה עם expected value הגבוה ביותר
- שימוש בניתוח רגישות על מנת לבחון את תוצאות הניתוח

חשוב לשים לב - את הערך של מי אנחנו רוצים למקסם?

דוגמאות: רפואה: רופא רושם תרופה לחולה - את הרווח של מי הוא רוצה למקסם?

- של החולה?
- של הרופא? אולי הוא רושם תרופה פחות טובה כדי שהחולה יחזור שוב?
- של החברה? אולי הוא רושם תרופה יקרה כדי שהחברה תרוויח יותר?
- של הממשלה?
- של הביטוח הרפואי?
- הרווח של הקונה?

- הרווח של המוכר?
- הרווח של המתווך? למשל סוכן נסיעות - האם הוא ממקסם את הרווח ללקוח או לעצמו?
- של הממשלה?

בניית עץ החלטה

נרצה לזהות את התוצאות האפשריות של כל שילוב אפשרי של אלטרנטיבות ומצבי טבע, ולייצג אותם בתור עץ של צמתי אי וודאות (change nodes) וצמתי החלטה (decision node).

- נייצג החלטות בתור ריבועים מהם יוצאות אלטרנטיבות
 - נייצג אי וודאות בתור עיגולים מהם יוצאים קשתות, כאשר לכל קשת מצמידים הסתברות וסכום כל הסתברויות עבור צומת אי וודאות חייב להיות 1.
- עץ ההחלטה מתחיל מהחלטה או ממצב טבע - אבל לעולם לא מתמורה.

קיפול העץ

עוברים על העץ מהסוף להתחלה

- בצמתי אי וודאות, מכפילים את ההסתברויות בתוחלת הרווח
- בצמתי החלטה, בוחרים את ההחלטה שנותנת את התמורה הכי גבוהה

ניתוח רגישות

מה אם אנחנו טועים בהערכת ההסתברויות? יכול להיות שבגלל הטעות נבחר בתוצאה רצויה. כמה הבחירה שלנו רגישה לטעויות בהערכת הסתברות?
אם יש רק שני מצבי טבע, ניתן לתת לסיכוי שאחד יקרה את המשתנה p , ולבנות גרף של תוחלת התמורה שנקבל עבור החלופה הטובה ביותר עבור כל p . אם p שלנו נמצא קרוב לנקודת המפגש בין שתי אלטרנטיבות - אנחנו רגישים לטעויות, ועד כמה ניפגע אם הערכת ההסתברות שלנו לא נכונה.

סימולציה

אפשר לבצע סימולציה ככלי אלטרנטיבה לקיפול העץ. אפשר לרוץ על העץ מההתחלה לסוף עבור סט אלטרנטיבות מסויים. בכל פעם שמגיעים לצומת אי וודאות מגרילים לפי ההסתברות. אם חוזרים על זה הרבה פעמים אפשר לקבל את ממוצע התועלת עבור אותו סט של בחירות.

ערך מידע חדש

- עד כמה המידע החדש אמין?

– מידע מושלם (perfect information) - מידע שאנחנו בטוחים שהוא נכון
– מידע לא מושלם (imperfect information) - מידע שיכול להיות שגוי

- כמה כדאי לשלם על המידע?

ערך מידע מושלם הוא ההפרש שבין הרווח בודאות מוחלטת לבין הרווח במצב של אי-ודאות:

$$EVPI = \text{Expected value under certainty} - \text{Maximum EMV}$$

- כאשר: Expected value under certainty - הרווח הגבוה ביותר האפשרי לכל מצב טבע שהמידע המושלם יכול לחזות לנו, כפול ההסתברות לאותו מצב טבע
- Maximum EMV - תוחלת הרווחים באי וודאות

כלומר, חשוב להסתכל לא רק על כמה נרוויח אם יהיה לנו את המידע, אלא גם להשוות את זה לרווח הצפוי לנו ללא המידע. ההפרש ביניהם הוא החסם העליון לכמה שווה לנו לשלם על המידע הזה.