הסקה בתנאי אי ודאות

ייצוג ידע לא ודאי

- ייצוג ידע בלוגיקה מסדר ראשון הינו בעיתי: בד"כ אין לסוכן גישה לכל הידע על הסביבה שלו.
- חלק מהידע של הסוכן על העולם הינו לא ודאי כתוצאה מחיישנים לא מושלמים.
- חלק אחר אינו ודאי מכיוון שלא ידועים לסוכן כל הפרמטרים המשפיעים על נכונותו.
- על כן עלינו לצייד את הסוכן האינטליגנטי ביכולת לייצג ידע לא ודאי ולהסיק באמצעותו מסקנות.

דוגמא: דיאגנוסטיקה רפואית

- נניח שברצוננו לייצג מערכת לדיאגנוסטיקה רפואית באמצעות לוגיקה.
- מערכת זו תכיל חוקים כגון:

$$\forall p \; Symptom(p, Toothache) \Rightarrow Disaease(p, Cavity)$$

- חוק זה כמובן אינו נכון: יתכנו פציינטים עם כאב שיניים אבל ללא חור.
- נוכל להוסיף אפשרויות נוספות:

$$Symptom(p, Toothache) \Rightarrow Disaease(p, Cavity) \vee Disaease(p, GumDisease) \vee \cdots$$

- אולם כדי להפוך פסוק זה לנכון עלינו לפרט בחוק את כל הסיבות האפשריות לכאב שיניים.
- ניתן להפוך את החוק:

$$\forall p \; Disaease(p, Cavity) \Rightarrow Symptom(p, Toothache)$$

- אולם גם חוק זה אינו נכון: יתכן חור ללא כאבים. יתכן שניהם מתקיימים אולם אינם קשורים.

בעייתיות בשימוש בלוגיקה

- כתיבה ממצה של כל הסיבות האפשריות לכל הסימפטומים האפשריים מערבת עבודה רבה מידי.
- השימוש בחוקים סבוכים כל כך אינו יעיל.
- ברוב התחומים (כמו בתחום הדיאגנוסטיקה הרפואית) אין למומחה תיאוריה שלמה על התחום.
- קיימת אי-ודאות לגבי חלק מהעובדות הידועות לנו (לדוגמא: הפציינט אינו בטוח היכן בדיוק כואב לו, צילום השן אינו ברור וכו')

שימוש בתורת ההסתברות

- המצב שתיארנו הינו מצב אופייני בו איננו מסוגלים לתאר את הידע שלנו בתחום באמצעות היסקים לוגיים.
- תחת זאת נעדיף לתאר את מידת האמונה של הסוכן בפסוקים על העולם.
- הכלי הידוע ביותר למטרה זו הינה תורת ההסתברות המצמידה לפסוקים מספר המבטא את הוודאות של הסוכן לגבי נכונותם.
- תורת ההסתברות מאפשרת לנו לסכם בצורה חסכנית את אי הוודאות הנובעת מהסיבות שתיארנו למעלה.
- לדוגמא, יתכן שאיננו בטוחים ממה נובע כאב השיניים של פציינט מסוים, אולם הננו מאמינים שבהסתברות 0.8 יש לפציינט חור אם יש לו כאב שיניים.
- 20% הנותרים מסכמים את כל הסיבות האחרות לכאב שיניים אותן איננו יודעים או שאנו עצלנים מידי לציין במפורש.

הסתברות - תזכורת

- נעסוק כאן במשתנים אקראיים בדידים.
- לשם פשטות נצמצם את הדיון למשתנים בוליאניים בעלי שני הערכים True ו-False. הדיון ניתן להרחבה בקלות למשתנים אקראיים בעלי יותר ערכים.
- לשם קיצור נכתב את $P(Cavity=True)$ בצורה $P(Cavity)$.
- $P(A)$ הינה ההסתברות הבלתי מותנית (unconditional או prior) ש-A אמיתי.
- $P(A|B)$ הינה ההסתברות המותנית של A ב-B. (conditional או posterior).
- חוק המכפלה: $P(A|B)=P(A\wedge B)/P(B)$ לכל $P(B)>0$.
- ההתפלגות של משתנה (ההסתברויות של כל ערכיו האפשריים תסומן ב- $\mathbf{P}(X)=\langle 0.7,0.3\rangle$

הסתברות - תזכורת (המשך)

- ההתפלגות המשותפת של קבוצת משתנים (joint probability distribution) מצמידה הסתברות לכל איבר במכפלה הקרטזית של תחומי המשתנים.
- עבור n משתנים ההתפלגות המשותפת הינה טבלה בת 2^n כניסות.
- לדוגמא:

	Toothache	Not Toothache
Cavity	0.04	0.06
Not Cavity	0.01	0.89

- ההתפלגות המשותפת של קבוצת המשתנים המתארים תחום מסוים יכולה לשמש לענות על כל שאילתא הסתברותית בתחום.
- לתחומים מעשיים המתוארים ע"י עשרות או מאות משתנים לא ניתן לבנות את טבלת ההתפלגות המשותפת.

חוק בייס (Bayes' Rule)

$$P(B|A)=\frac{P(A|B)P(B)}{P(A)}$$

דוגמא:

- M: דלקת קרום המוח.
- S : צוואר נוקשה
- $P(S|M)=0.5$
- $P(M)=1/50000$
- $P(S)=1/20$
- נפעיל את חוק בייס ונקבל את ההסתברות שפציינט בעל צוואר נוקשה חולה בדלקת קרום המוח:

$$P(M|S)=\frac{P(S|M)P(M)}{P(S)}=\frac{0.5\times 1/50000}{1/20}=0.0002$$

רשתות בייס (Bayesian Networks)

- רשת בייס הינה דרך חסכנית לציין את ההתפלגות המשותפת של משתני התחום.
- הרשת הינה גרף מכוון.
- הצמתים ברשת הינם המשתנים האקראיים בתחום.
- קיימת קשת מכוונת בין הצומת המייצג את X לבין הצומת המייצג את Y אם ל-X יש השפעה ישירה על Y.
- לכל צומת צמודה טבלה המתארת את ההסתברות המותנית של המשתנה בהוריו הישירים.

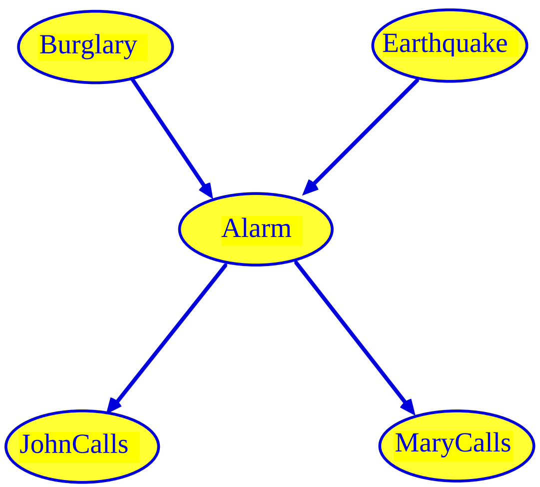
דוגמת האזעקה

- נניח שהתקנו אזעקה חדשה נגד גנבים.
- האזעקה מגיבה באמינות גבוהה לנוכחות גנבים.
- האזעקה מגיבה לעתים גם לרעידות אדמה קלות.
- ביקשנו משני שכנים, ג׳ון ומרי להתקשר אלינו לעבודה במקרה שהם שומעים אזעקה.
- ג׳ון תמיד מצלצל כשיש אזעקה אולם לעתים מצלצל בטעות כאשר הוא שומע את הטלפון מצלצל.
- מרי מקשיבה למוסיקה בעצמה רבה ולכן לעתים לא שומעת את האזעקה.

- בהינתן שג׳ון צלצל ומרי לא, מה ההסתברות שאירעה פריצה?

טופולוגית הרשת

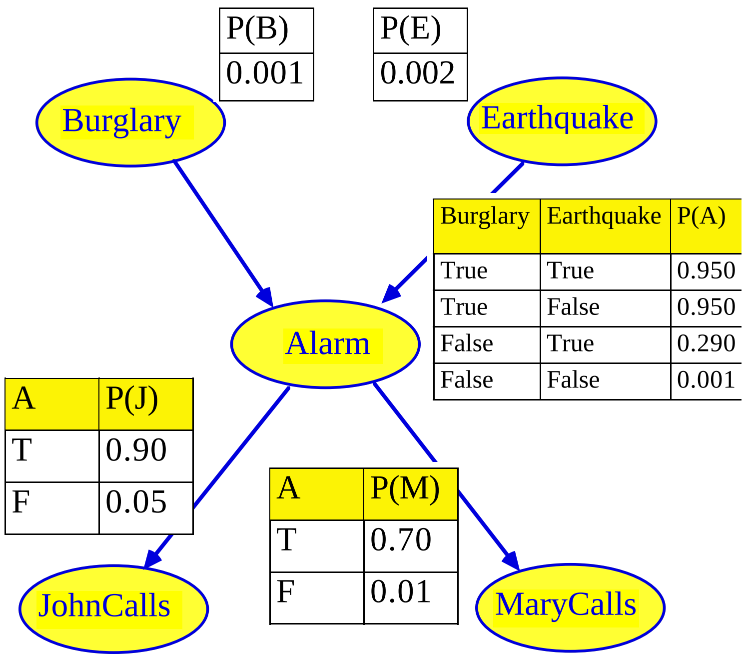
- טופולוגית הרשת משקפת את הידע שלנו על הסיבתיות בתחום.
- במקרה שלנו, לפריצה ורעידת אדמה השפעה סיבתית ישירה על הפעלת האזעקה.
- להפעלת האזעקה השפעה סיבתית ישירה על צלצול ע״י ג׳ון ומרי.
- אנו מניחים שג׳ון ומרי לא מרגישים ברעידות אדמה קלות ולא בפריצה ולכן אין קשר סיבתי ישיר בין המאורעות.



טבלאות ההסתברות המותנה

- אנו מוסיפים את ההסתברויות לרשת באמצעות טבלאות ההסתברויות המותנות המוצמדות לצמתים.
- טבלה כזו מגדירה את ההתפלגות של ערכי המשתנה המיוצג ע"י הצומת כתלות בכל צירוף של ערכי הוריו
- ההסתברויות בכל שורה מסתכמות ל-**1.0**
- עבור משתנה בוליאני עם **n** הורים מכילה הטבלה 2^n ערכים (2^n הערכים האחרים נגזרים מיידית)

Burglary	Earthquake	P(A|B,E)=True	P(A|B,E)=False
True	True	0.950	0.050
True	False	0.950	0.050
False	True	0.290	0.710
False	False	0.001	0.999



P(B)
0.001

P(E)
0.002

Burglary	Earthquake	P(A)
True	True	0.950
True	False	0.950
False	True	0.290
False	False	0.001

A	P(J)
T	0.90
F	0.05

A	P(M)
T	0.70
F	0.01

הערות

- שימו לב שהרשת אינה כוללת מאורעות שונים שנראים רלוונטיים לבעיה כגון המאורע שמרי מקשיבה למוסיקה חזקה.
- מאורעות אלו מבוטאים ע"י אי הודאות שתוצמד לקשר בין הפעלת האזעקה וצלצול ע"י מרי.
- ההסתברויות מסכמות למעשה קבוצה גדולה מאוד (אולי אינסופית) של מאורעות אפשריים.
- לדוגמא, יתכן שהאזעקה לא תפעל בשל תקלה אלקטרונית.

חישוב ההתפלגות המשותפת

- מכיוון שהסתברות של כל מאורע תלויה רק בהוריו הישירים ניתן לחשב את ההתפלגות המשותפת לפי הנוסחה (1):

$$P(x_1,\cdots,x_n)=\prod_{i=1}^{n}P\big(x_i\big|Parents(X_i)\big)$$

- נחשב למשל את ההסתברות של המאורע הבא: "נשמעה אזעקה, ג'ון ומרי צלצלו, אולם לא היתה פריצה ולא היתה רעידת אדמה."

$$P(J\wedge M\wedge A\wedge\neg B\wedge\neg E)$$

$$=P(J|A)P(M|A)P(A|\neg B\wedge\neg E)P(\neg B)P(\neg E)$$

$$=0.9\times0.7\times0.001\times0.999\times0.998=0.00062$$

- ההתפלגות המשותפת יכולה לשמש לענות על כל שאילתה בתחום.
- ניתן לכן לחשב אותה לפי הרשת ולענות על שאילתות. בהמשך נראה אלגוריתם יעיל יותר.

בניית רשתות

- ההתפלגות המשותפת ניתנת להכתב בצורה:

$$P(x_1,\cdots,x_n)=$$

$$P(x_n|x_{n-1},\cdots,x_1)P(x_{n-1},\cdots,x_1)$$

- נמשיך את הפיתוח:

$$P(x_1,\cdots,x_n)=P(x_n|x_{n-1},\cdots,x_1)P(x_{n-1}|x_{n-2},\cdots,x_1)$$

$$\cdots P(x_2|x_1)P(x_1)$$

- $$P(x_1,\cdots,x_n)=\prod_{i=1}^{n}P(x_i|x_{i-1},\cdots,x_1)$$

- אם מתקיים $Parent(X_i)\subseteq\{x_{i-1},\cdots,x_1\}$ אזי נוסחא 1 למעלה

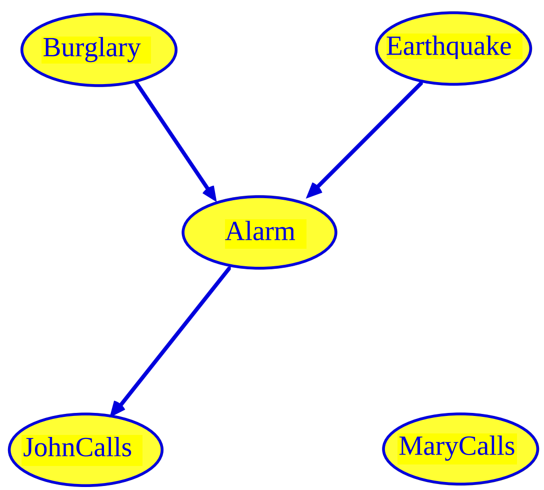
$$P(x_1,\cdots,x_n)=\prod_{i=1}^{n}P(x_i|Parents(X_i))$$

שקולה לטענה: $P(x_i|x_{i-1},\cdots,x_1)=P(x_i|Parents(X_i))$

- על כן בחירת ההורים נעשית כך שתתקיים תכונה זו.

דוגמא

- נניח שכבר בנינו את הרשת הבאה:

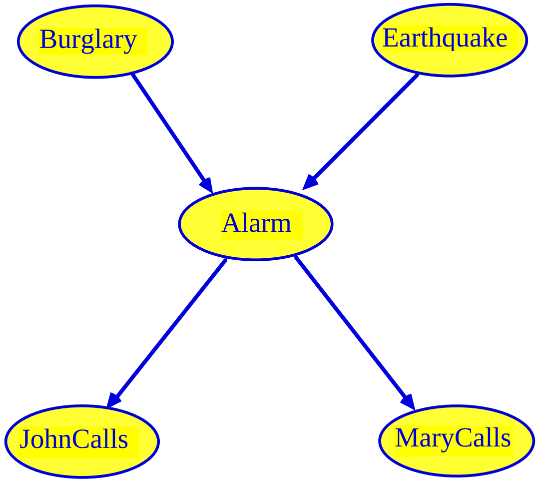


- עלינו להחליט מי יהיו ההורים של הצומת **MaryCalls**
- ברור שהמאורע **MaryCalls** מושפע ע"י המאורעות **Earthquake** ו-**Burglary**.
- אולם השפעה זו אינה ישירה. היא מגיעה דרך המאורע **Alarm**
- גם המאורע **JohnCalls** אינו משפיע על **MaryCalls** בהינתן **Alarm**
- מסקנות אלו שקולות לנוסחא:

$$P(MaryCalls|JohnCalls, Alram, Earthquake, Burglary))$$

$$=P(MaryCalls|Alarm)$$

- על כן אנו מחליטים ש- $Parents(maryCalls)=\{Alarm\}$



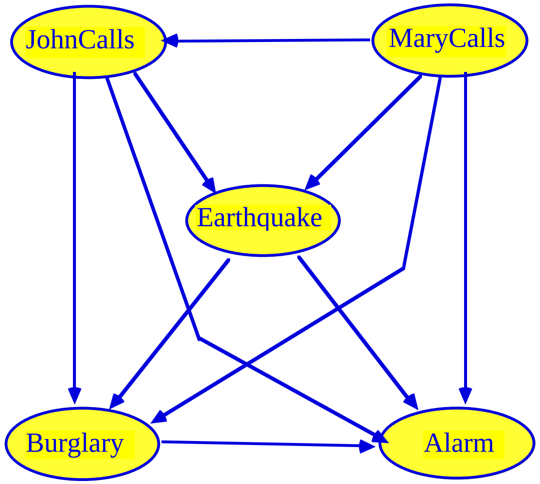
פרוצדורה כללית לבנית רשת

- בחר קבוצת משתנים אקראיים המתארים את התחום.
- בחר סדר עליהם.
- לכל משתנה X_i (לפי הסדר):
  - הוסף צומת עבור X_i ברשת
  - בחר קבוצת $Parent(X_i)$: קבוצה מינימלית מהמשתנים שכבר הוכנסו המקיימת את השוויון למעלה.
  - הגדר את טבלת ההסתברות המותנה עבור X_i

- מאחר ואנו מחברים צומת רק לצמתים קודמים לא יווצרו מעגלים.
- ניתן להוכיח שלא ניתן לבנות רשת שמפירה את אקסיומות תורת ההסתברות.

בחירת סדר טוב

- ניתן לייצג תחומים אמיתיים רבים באמצעות רשת קומפקטית (יחסית להתפלגות המשותפת המלאה)
- אם בתחום בין N משתנים מושפע כל משתנה <u>ישירות</u> מלכל היותר K משתנים ו $K << N$ אזי נזדקק למספר הסתברויות קטן בהרבה ($N2^K$ עבור משתנים בוליאניים במקום 2^N)
- חשוב לבנות את הרשת בסדר טוב: להתחיל בגורמים משפיעים לעבור לגורמים עליהם הם משפיעים וכו'. כלומר סדר התואם את הסיבתיות.
- בחירת סדר לא טוב תגרום לרשתות מסורבלות יותר ולטבלאות פחות אינטואיטיביות למילוי.
- לדוגמא, בחירת הסדר $\langle MaryCalls, JohnCalls, Earthquake, Burglary, Alram \rangle$ תביא לרשת הדורשת 31 הסתברויות - בדיוק כמו ההתפלגות המשותפת המלאה.

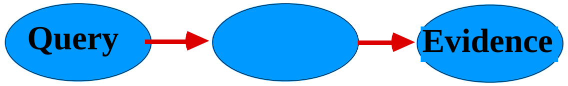


הסקה ברשתות בייס

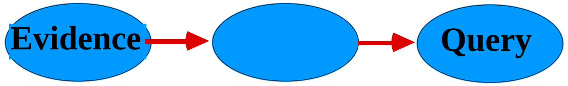
- בדרך כלל הסוכן מקבל ערכים עבור קבוצת משתני עדות (evidence).
- ערכים אלו מגיעים ממקור חיצוני (למשל חישנים).
- מטרת הסוכן להסיק את התפלגות הפוסטריור של משתני השאילתא (query variables) בהנתן הראייה: $\mathbf{P}(Query|Evidence)$
- בדוגמת האזעקה סביר שמשתנה השאילתא יהיה $Burglary$ ומשתני העדות יהיו $MaryCalls$ ו-$JohnCalls$
- אולם הרשת מאפשרת גמישות מוחלטת בקביעת משתני השאילתא ומשתני העדות.

סוגי הסקה ברשת בייס

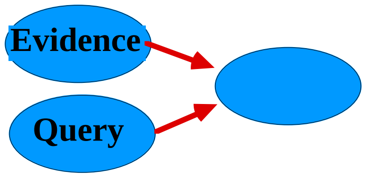
- הסקה דיאגנוסטית [מתוצאות לסיבות]. $P(Burglary|JohnCalled)$



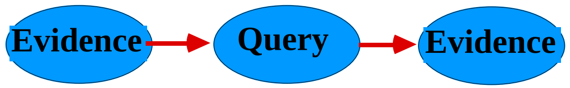
- הסקה סיבתית [מסיבות לתוצאות] $P(JohnCalled|Burglary)$



- הסקה בין סיבות משותפות [סיבות לאותה תוצאה] $P(Burglary|Alarm, Earthquake)$



- הסקה מעורבת [של שני סוגים] $P(Alarm|JohnCalled \wedge Burglary)$



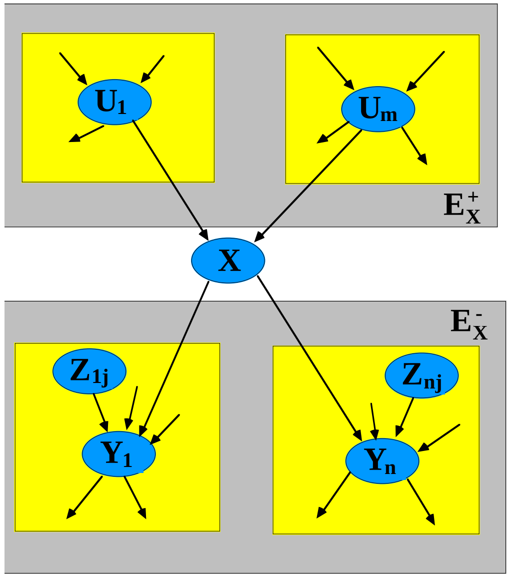
הסקה הסתברותית

- הסקה בייסינית מאפשרת לנו לחמוק מטעויות הנפוצות אצל בני אדם.
- לדוגמא, מה ההסתברות של $Burglary$ בהנתן $JohnCalls$?
- מכיון שהפעלת האזעקה גורמת ב-90% מהמקרים לטלפון מג׳ון ומכיון שב-95% מהמקרים בהם יש גניבה מופעלת האזעקה, אנו מתפתים להסיק ש $P(Burglary|JohnCalls) \approx 0.9$
- אולם חישוב באמצעות הרשת מביאה לתוצאה מפתיעה: $P(Burglary|JohnCalls) = 0.016$
- הסיבה: ההסתברות לפריצה היא נדירה - פעם ב-1000 ימים.
- ההסתברות שג׳ון מצלצל כשאין אזעקה הינה 0.05 (כלומר 50 פעם ב-1000 ימים).
- כלומר רוב צלצוליו של ג׳ון הינם אזעקות שווא.
- בנוסף יתכן שהאזעקה לא מופעלת בזמן פריצה.

אלגוריתם הסקה ברשת בייס

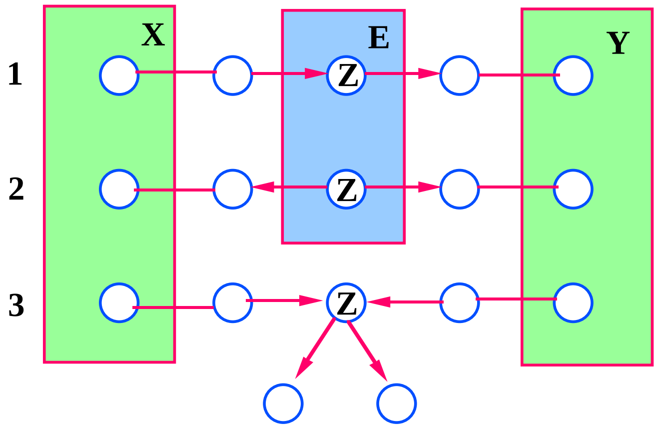
- האלגוריתם הבסיסי מתאים רק ל- polytree: רשת בה יש לכל היותר מסלול אחד (לא מכוון) בין שני צמתים.
- האלגוריתמים לרשתות כלליות מושתתות על האלגוריתם להסקה ב-polytree.
- האלגוריתם עובד ע"י הסקה "אחורנית" ממשתנה השאילתא למשתני העדות (evidence).
- האלגוריתמים לגרף כללי מושתתים על האלגוריתם ל- polytree.

Polytree



d-seperation

- קבוצת צמתים E מפרידה 2 קבוצות צמתים X ו- Y אם כל מסלול לא מכוון ביניהם חסום בהנתן E.
- מסלול חסום בהנתן קבוצת צמתים E אם קיים בו צומת Z עבורו מתקיים אחד משלושת התנאים:
- Z נמצא ב-E ויש במסלול חץ אחד שנכנס אליו וחץ אחד שיוצא ממנו.
- Z נמצא ב-E ושני החצים מצביעים ממנו.
- Z וצאצאיו אינם ב-E ושני החצים מובילים לתוכו.



- אם קבוצת צמתים X מופרדת מקבוצת צמתים Y ע"י קבוצת צמתים E אזי X ו-Y בלתי תלויים בהנתן E.

נורמליזציה

- חוק בייס בצורתו המוכללת מחזיר טבלת התפלגות:

$$\mathbf{P}(Y|X)=\frac{\mathbf{P}(X|Y)\mathbf{P}(Y)}{\mathbf{P}(X)}$$

- אם למשל $Domain(X)=\{x_1,x_2,x_3\}$ $Domain(Y)=\{y_1,y_2\}$ אזי $\mathbf{P}(Y|X)$ תראה לדוגמא כך:

	$Y=y_1$	$Y=y_2$
$X=x_1$	0.7	0.3
$X=x_2$	0.1	0.9
$X=x_3$	0.5	0.5

- שימו לב שבכל שורה X הינו קבוע.
- ניתן לכן לכתוב $\mathbf{P}(Y|X)=\alpha\mathbf{P}(X|Y)\mathbf{P}(Y)$ ולחשב את α כך שיגרום שהשורות יסתכמו ל-1. לתהליך כזה קוראים נורמליזציה
- פעמים רבות, כאשר רוצים להשוות את היחס בין שתי הסתברויות, אין צורך לחשב את α.

סימונים

- X הינו משתנה השאילתא.
- $\mathbf{Y}=\{Y_1,\cdots,Y_n\}$ הינם ההורים של X
- $\mathbf{U}=\{U_1,\cdots,U_n\}$ הינם הילדים של X
- E היא קבוצה של משתני עדות (נתון לנו ערכו של כל אחד מהם).
- E_X^+ הינם התמיכה הסיבתית: משתני העדות מעל X הקשורים להוריו.
- E_X^- הינם התמיכה התוצאתית: משתני העדות מתחת ל- X הקשורים לילדיו.
- $E^+_{Y_i\setminus X}$ כל משתני העדות (ב-E) הקשורים ל- Y_i שלא דרך X

מבנה האלגוריתם

- מטרת האלגוריתם למצוא את $\mathbf{P}(X|E)$ כאשר X הינו משתנה השאילתא ו-E הינם משתני העדות (evidence).
- כצעד ראשון מפרידים בין ההשפעה של E_X^+ ו-E_X^-.
- מחשבים את ההשפעה של E_X^+ על כל אחד מההורים של X (רקורסיבית) ומעבירים את שקלול ההשפעה ל-X
- מחשבים את ההשפעה של E_X^- על כל אחד מהילדים של X (רקורסיבית) ומעבירים את שקלול ההשפעה ל-X

הפרדת השפעות ההורים והילדים

- אנו מעוניינים ב- $\mathbf{P}(X|E)=\mathbf{P}(X|E_X^-,E_X^+)$
- קיימת הרחבה לחוק בייס המאפשרת משתני "רקע":

$$\mathbf{P}(Y|X,B)=\frac{\mathbf{P}(X|Y,B)\mathbf{P}(Y|B)}{\mathbf{P}(X|B)}$$

- נפעיל את החוק ונקבל: $\mathbf{P}(X|E_X^-,E_X^+)=\frac{\mathbf{P}(E_X^-|X,E_X^+)\mathbf{P}(X|E_X^+)}{\mathbf{P}(E_X^-|E_X^+)}$
- X מפריד בין E_X^+ ל-E_X^- לכן $\mathbf{P}(E_X^-|X,E_X^+)=\mathbf{P}(E_X^-|X)$
- נשתמש ב- $1/\mathbf{P}(E_X^-|E_X^+)$ כקבוע נרמול ונקבל:

$$\mathbf{P}(X|E_X^-,E_X^+)=\alpha\,\mathbf{P}(E_X^-|X)\mathbf{P}(X|E_X^+)$$

function Ploytree-Ask(X,E)

input: X, A random variable
E, A set of random variables with known values

output: $\mathbf{P}(X|E)$ A distribution over X given evidence E

return Support-Except(X,{},E)

function Support-Except(X,V,E)

input: X, a Random variable. V and E, sets of random variables.

output: $\mathbf{P}(X|E_{X\setminus V})$

[A distribution over X given evidence variables E except that reachable via V]

if $X \in E$ then **return** observed point distribution for X.

else calculate $\mathbf{P}(E^-_{X\setminus V}|X)$ = Evidence-Except(X,V,E)

U ← Parents[X]

if U is empty then **return Normalize** $\left(\mathbf{P}(E^-_{X\setminus V}|X)\mathbf{P}(X)\right)$

else for each U_i in U

calculate and store $\mathbf{P}(U_i|E_{U_i\setminus X})$ = Support-Except(U_i,{X},E)

return Normalize $\left(\mathbf{P}(E^-_{X\setminus V}|X)\sum_u \mathbf{P}(X|u)\prod_i \mathbf{P}(U_i|E_{U_i\setminus X})\right)$

function Evidence-Except(X,V,E)

input: X, a Random variable.
V and E, sets of random variables.

output: $\mathbf{P}(E^-_{X\setminus V}|X)$

[probability of downstream evidence (except evidence reachable via V) for each value of X]

U ← Children[X] - V

if Y is empty then return a vector of 1s
;No downstream evidence - true for all X

else

for each Y_i in Y do

if $Y_i \notin E$ then

calculate $\mathbf{P}(E^-_{Y_i}|Y_i)$=Evidence-Except($Y_i$,{},E)

$$Z_i \leftarrow Parents[Y_i] - X - E$$

$$Z_i' \leftarrow \left(Parents[Y_i] - X\right) \cap E$$

for each Z_{ij} in Z_i do calculate

$\mathbf{P}\left(Z_{i_j}|E_{Z_{i_j}\setminus Y_i}\right)$ = Support-Except(Z_{ij},{Y_i},E)

return $\prod_i$ **if** $Y_i \in E$ then

$$\sum_{Z_i} P(Y_i|X, Z_i, Z_i')\prod_j \mathbf{P}\left(Z_{i_j}|E_{Z_{i_j}\setminus Y_i}\right)$$

else

$$\sum_{Y_i}\mathbf{P}\left(E^-_{Y_i}|Y_i\right)\sum_{Z_i} P(Y_i|X, Z_i, Z_i')\prod_j \mathbf{P}\left(Z_{i_j}|E_{Z_{i_j}\setminus Y_i}\right)$$

הסברים

function Ploytree-Ask(X,E)

- אם נניח שמשתני העדות הינם $Burglary$ ו-$MaryCalls$ אזי הקלט יראה כך: $Burglary = True, MaryCalls = False$ או בקיצור $Burglary, \neg MaryCalls$
- הקלט יהיה ווקטור של ערכים שסכומו 1. אם למשל משתנה האילתא הינו $Alarm$ אזי הפלט יראה כך: $P(Alarm = True) = 0.8,$ $P(Alarm = False) = 0.2$ או בקיצור $\langle 0.8, 0.2\rangle$

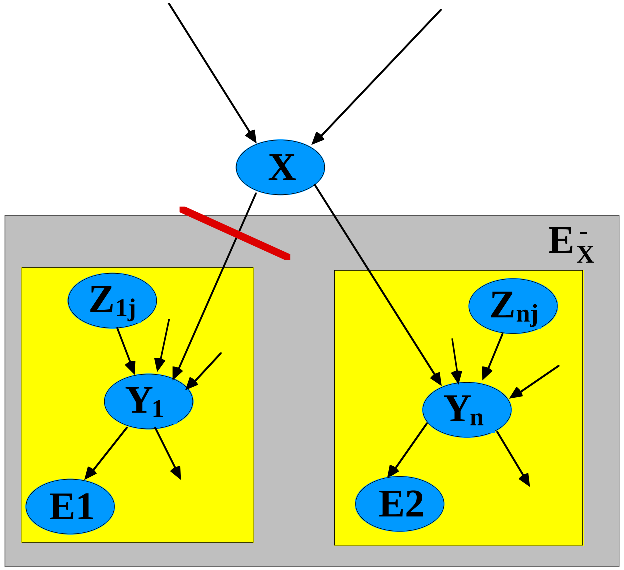
if $X \in E$ then **return** observed point distribution for X.

- זהו אחד מתנאי העצירה של הרקורסיה. אם הגענו למשתנה עדות אזי ידוע לנו ערכו. אין צורך להמשיך לטפס למעלה (או לרדת למטה) מכיון שהוא חוסם את הוריו (או ילדיו)

calculate $\mathbf{P}(E^-_{X\setminus V}|X)$ = Evidence-Except(X,V,E)

- זוהי הקריאה לחישוב ההשפעה של הילדים. אמנם בקריאה הראשונה V

יהיה ריק אולם בשאר הקריאות נרצה לחסום את ההשפעה של האחים:



- בחישוב הרקורסיבי של Y_1 נרצה לבלום את ההשפעה של E2 דרך X.

if U is empty

then return Normalize $\left(\mathbf{P}(E^-_{X\setminus V}|X)\mathbf{P}(X)\right)$

- זהו תנאי שני לעצירת הרקורסיה. הגענו לשורש (שאינו משתנה עדות).
- מחשבים לפי הנוסחא אולם מכיון שאין הורים אזי $\mathbf{P}(X|E^+_X)=\mathbf{P}(X)$.

for each U_i in U

calculate and store

$\mathbf{P}(U_i|E_{U_i\setminus X})$= Support-Except($U_i$,{X},E)

- זוהי הקריאה הרקורסיבית. מנתקים את הקשר ל-X ומפעילים את הפרוצדורה על כל אחד מההורים.

$$\left(\mathbf{P}(E^-_{X\setminus V}|X)\sum_u \mathbf{P}(X|u)\prod_i \mathbf{P}(U_i|E_{U_i\setminus X})\right)$$

- הביטוי $\mathbf{P}(E^-_{X\setminus V}|X)$ חושב ע"י הקריאה לפרוצדורה המחשבת את ההשפעה מלמטה.
- הביטוי $\mathbf{P}(X|u)$ נלקח מטבלת ההסתברות המותנה של המשתנה X.
- הביטוי $\mathbf{P}(U_i|E_{U_i\setminus X})$ חושב ע"י הקריאות הרקורסיביות.
- באופן כללי הביטוי $\sum_u \mathbf{P}(X|u)\prod_i \mathbf{P}(U_i|E_{U_i\setminus X})$ מסכם את ההסתברויות של X כתלות בכל צירוף אפשרי של הורי X.
- ההסתברות של הצירוף כתלות במשתני העדות חושב ע"י חישוב רקורסיבי לכל אחד מההורים והכפלת ההסתברויות שהתקבלו (עקב אי-תלות בין ההורים כפי שיוסבר בהמשך).

חישוב השפעת ההורים

- אנו מתחילים מ- $\mathbf{P}(X|E^+_X)$
- מכיוון ש X תלוי רק בהוריו (מהגדרת הרשת) אנו מסכמים את ההסתברויות בהתחשב בכל הצירופים האפשריים של ערכי הורי X משוקללים בהסתברויות שלהם. $\mathbf{P}(X|E^+_X)=\sum_u \mathbf{P}(X|u)P(u|E^+_X)$
- E^+_X מפריד בין ההורים (זכרו את התנאי השלישי) לכן ההסתברויות שלהם בלתי תלויות וניתן לכתוב אותן בצורה:

$$\mathbf{P}(X|E^+_X)=\sum_u \mathbf{P}(X|u)\prod_i P(u_i|E^+_X)$$

- לבסוף ניתן להפריד את משתני העדות של כל הורה

$$E^+_X = E^+_{U_1\setminus X}\cup\cdots\cup E^+_{U_m\setminus X}$$

- קבוצת משתנים כזו מפרידה את ההורה משאר המשתנים:

$$\mathbf{P}(X|E^+_X)=\sum_u \mathbf{P}(X|u)\prod_i P(u_i|E^+_X)$$