دانشگاه علم و صنعت
دانشکده مهندسی کامپیوتر

استدلال احتمالاتی

«هوش مصنوعی: رهیافتی نوین»، فصل ۱۴

مدرس: آرش عبدی هجراندوست

نیم‌سال دوم ۱۴۰۱-۱۴۰۲

شبکه بیز

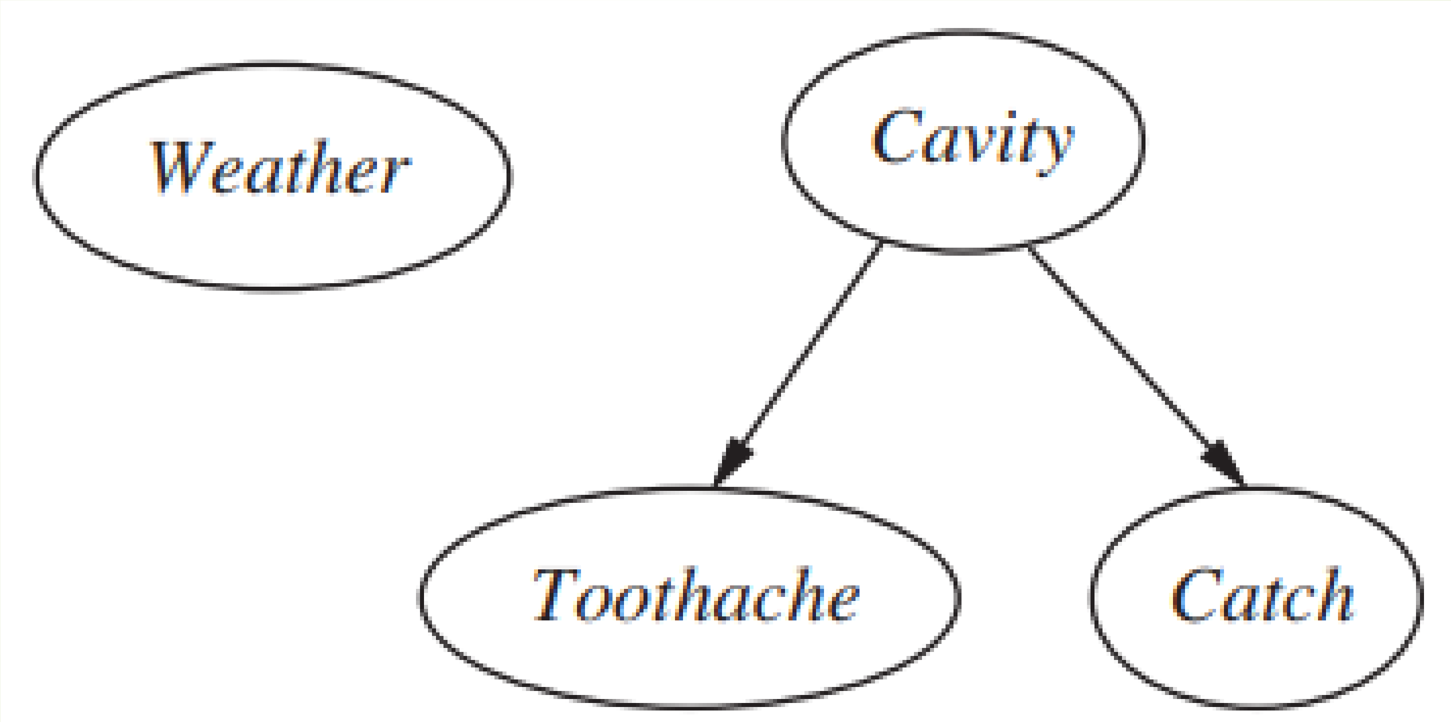
- ابزاری برای بازنمایی توزیع احتمالاتی توام
- نام های دیگر:
  - شبکه باور – Belief Network
  - شبکه احتمالاتی – Probability Network
  - شبکه عِلّی – Causal Network
  - نگاشت دانش – Knowledge Map

تعریف شبکه بیز

- شبکه بیز یک گراف جهت‌دار است که در آن:
  - هر گره متناظر با یک متغیر تصادفی است.
  - تعدادی یال جهت دار بین گره‌ها وجود دارد. اگر از X به Y یال وجود داشته باشد، X والد Y خوانده می‌شود.
  - گراف حلقه جهت دار ندارد
  - هر گره X_i دارای توزیع احتمالاتی شرطی به شکل $P(X_i|Parents(X_i))$ است که تاثیر والدها بر روی گره را مشخص می‌کند

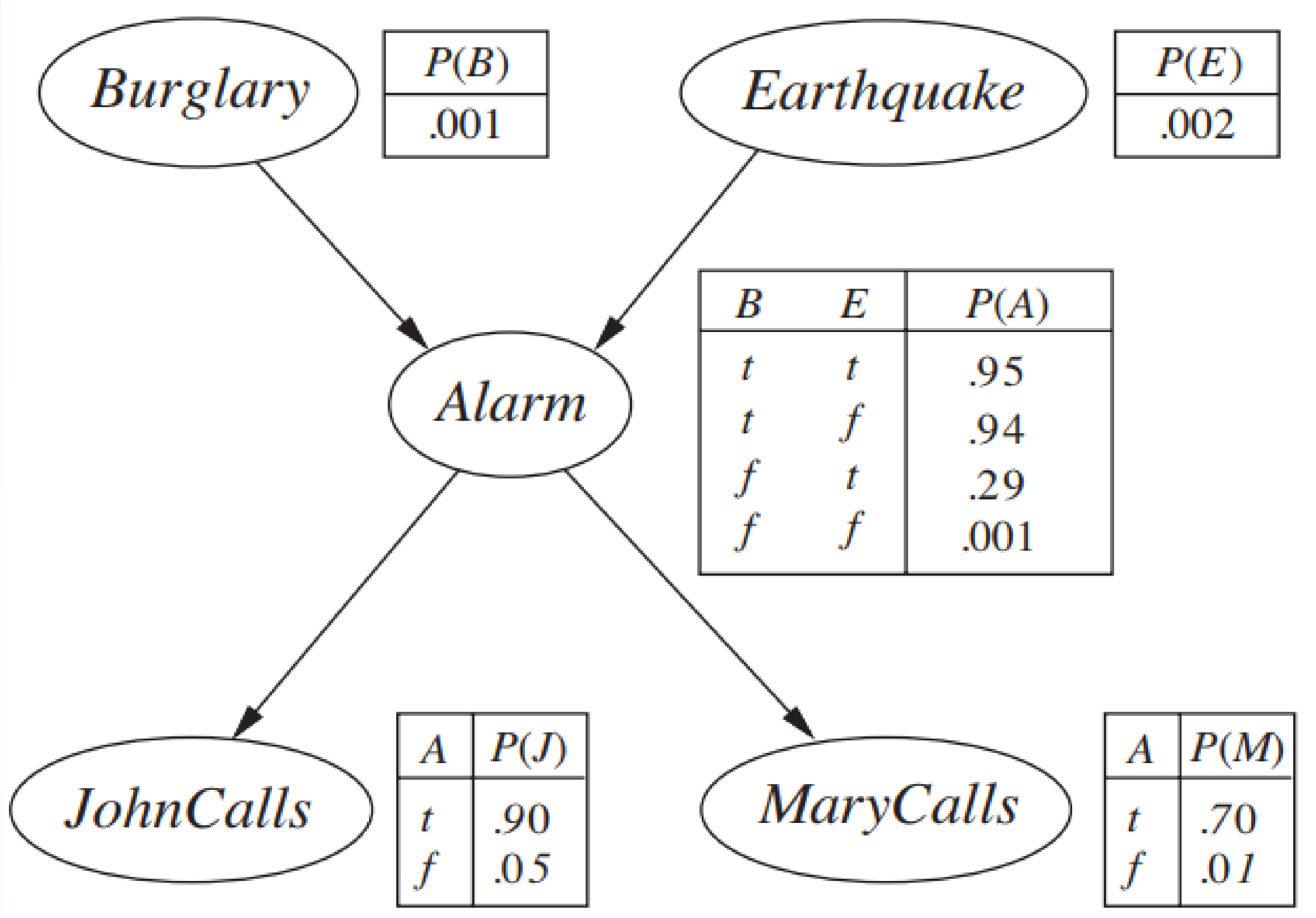
- معنی هر یال آن است که والد تاثیر **مستقیم** روی فرزند دارد.
- Cause والد Effect است.
- برای فرد خبره کار ساده‌ای است که روابط علی مستقیم را بین متغیر ها تعیین کند
  - تعیین توپولوژی، صرف نظر از تعیین احتمال هر یک
- اگر احتمالات شرطی مربوطه هم مشخص شود، با شبکه بیز می‌توان توزیع توام کامل تمام متغیرها را مشخص کرد

مثالی از شبکه بیز



مثالی دیگر

- زنگ دزدگیر جدید در خانه
- حساسیت زنگ، به دزد و نیز زمین لرزه
- دو همسایه با نام‌های John و Mary
  - قول دادهاند اگر صدای زنگ را شنیدند، تماس بگیرند و خبر دهند (وقتی خانه نیستی)
- John تقریبا هر وقت صدای زنگ را بشنود، تماس می‌گیرد، اما گاهی صدای تلفن را با زنگ اشتباه می‌گیرد.
- Mary صدای بلند موسیقی را دوست دارد و خیلی وقتها صدای زنگ را نمی‌شنود!



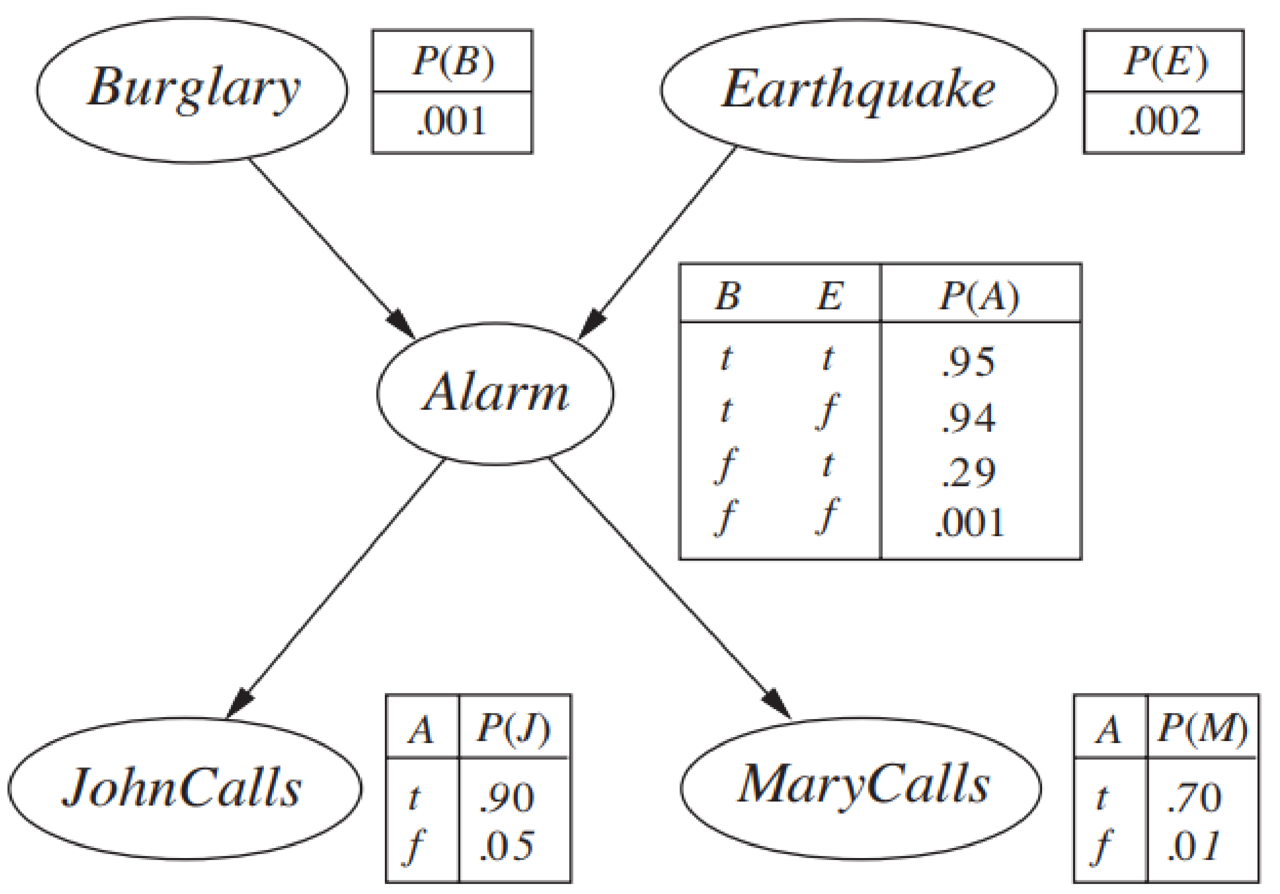
- سوال ۱: آیا متغیر زنگ و تماس جان، از هم مستقلند؟
  - چرا در جدول احتمالات زنگ، حرفی از تماس جان زده نشده؟
- سوال ۲: احتمال زنگ نخوردن چقدر است (در حالت های مختلف)؟

توزیع توام کامل از شبکه بیز

- با داشتن شبکه بیز، توزیع توام کامل بدین شکل قابل بیان است.

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

- حروف کوچک: مقدار خاص برای متغیر
- حروف بزرگ: متغیر تصادفی
- $parents(X_i)$: مقادیری از متغیرهای والد متغیر X_i که در $x_1, x_2, \ldots, x_n$ ظاهر شده‌اند.
- چرا رابطه فوق صحیح است؟



❖ مثال:

$$\begin{aligned} P(j, m, a, \neg b, \neg e) &= P(j \mid a) P(m \mid a) P(a \mid \neg b \wedge \neg e) P(\neg b) P(\neg e) \\ &= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.000628 \end{aligned}$$

قاعده زنجیره

❖ Product Rule

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

❖ Chain Rule:

$$\begin{aligned} P(x_1,\ldots,x_n) &= P(x_n \,|\, x_{n-1},\ldots,x_1)P(x_{n-1} \,|\, x_{n-2},\ldots,x_1) \cdots P(x_2 \,|\, x_1)P(x_1) \\ &= \prod_{i=1}^{n} P(x_i \,|\, x_{i-1},\ldots,x_1)\,. \end{aligned}$$

❖ مقایسه با $P(x_1,\ldots,x_n) = \prod_{i=1}^{n} P(x_i \,|\, parents(X_i))$ نتیجه میدهد:

$$\mathbf{P}(X_i \,|\, X_{i-1},\ldots,X_1) = \mathbf{P}(X_i \,|\, Parents(X_i))$$

❖ به فرض آنکه داشته باشیم: $Parents(X_i) \subseteq \{X_{i-1},\ldots,X_1\}$

$$\mathbf{P}(X_i \mid X_{i-1}, \ldots, X_1) = \mathbf{P}(X_i \mid Parents(X_i))$$

$$Parents(X_i) \subseteq \{X_{i-1}, \ldots, X_1\}$$

- برای تامین فرض مذکور باید شماره گذاری گره ها به ترتیبی باشد که در گراف بیز تعیین شده است.
- طبق رابطه فوق: شبکه بیز و تفسیرش از متغیرها صحیح است اگر:
- هر متغیر، به شرط داشتن وضعیت والدهایش، از سایر متغیرهای پیشین خود (طبق ترتیب شماره گذاری متغیرها) مستقل باشد.
- برای تامین این شرط، در ساختن شبکه بیز باید به آن دقت کرد.

ساختن شبکه بیز

- گام ۱: تعیین متغیرها و مرتب کردن آنها
  - هر نوع ترتیبی قابل قبول است
  - اگر ترتیب به گونه‌ای باشد که متغیرهای علت قبل از متغرهای معلول قرار گیرند، شبکه کوچکتر/فشرده تر خواهد شد.
- گام ۲: تعیین یالها. برای هر متغیر X_i (i=1,....,n) :
  - در مجموعه متغیرهای پیشین (از X_1 تا X_{i-1})، کوچکترین مجموعه والدهای X_i را پیدا کن، به گونه‌ای که $\mathbf{P}(X_i \mid X_{i-1}, \ldots, X_1) = \mathbf{P}(X_i \mid Parents(X_i))$
  - برای هر والد، یالی از والد به فرزند وصل کن
  - جدول احتمال شرطی متغیر را ایجاد کن: $\mathbf{P}(X_i|Parents(X_i))$
- به صورت شهودی: والدهای هر متغیر باید متغیرهایی باشند که تاثیر مستقیم روی این متغیر دارند

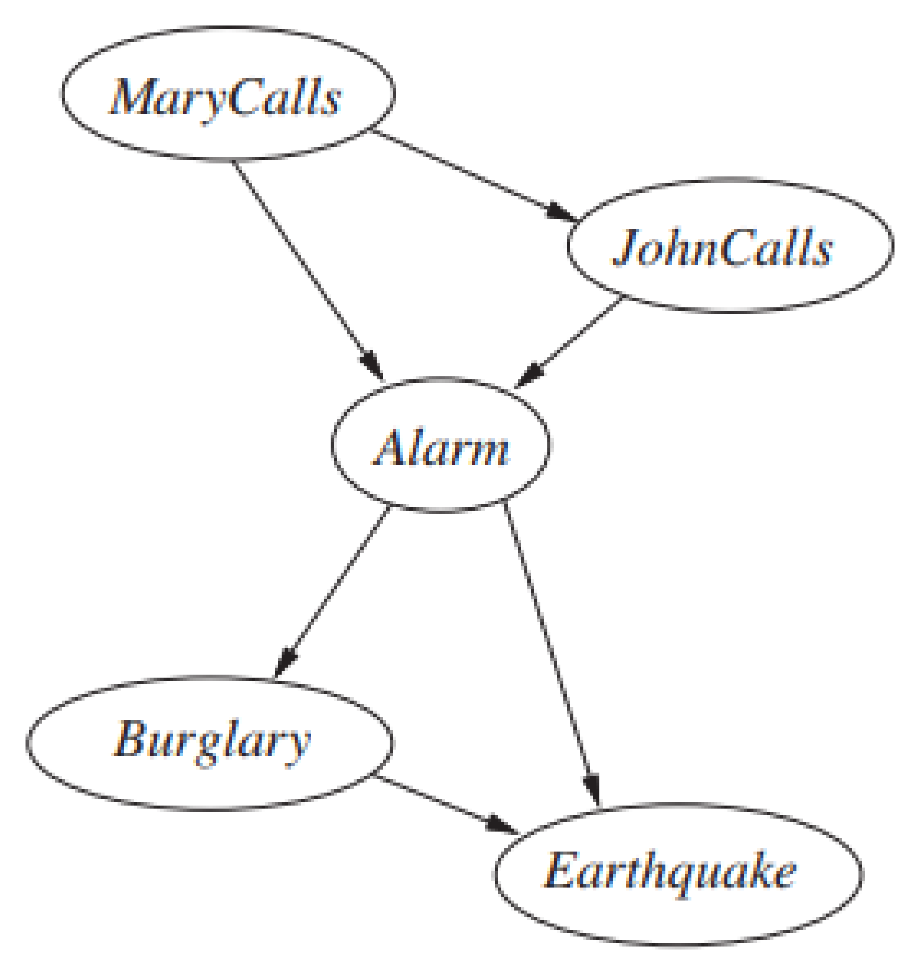
خصوصیات شبکه بیز

- فشرده بودن (نسبت به توزیع توام کامل)
  - Sparsity
  - ساختارمحلی
- می‌توان فرض کرد هر متغیر در عمل از k متغیر دیگر تاثیر مستقیم می‌پذیرد.
  - اندازه جدول احتمال هر متغیر 2^k
  - اندازه جدولهای کل شبکه: $n2^k$
  - اندازه جدول توزیع توام کامل (بدون بیز): 2^n
- این امکان وجود دارد که روابط ضعیف بین متغیرها حذف شود و مدل کوچکتر شود:
  - مثلا در مثال زنگ دزدگیر: ارتباط مستقیم بین زمین لرزه و تماس ماری و جان در نظر گرفته نشده.
  - اگر زمین لرزه را احساس کنند، حدس می‌زنند که صدای زنگ بابت دزد نیست، و ممکن است تماس نگیرند.

اهمیت ترتیب گره‌ها

- اگر ترتیب مشاهده گره‌ها چنین باشد:

- Marry→John→Allarm→Burglary→Earthquake

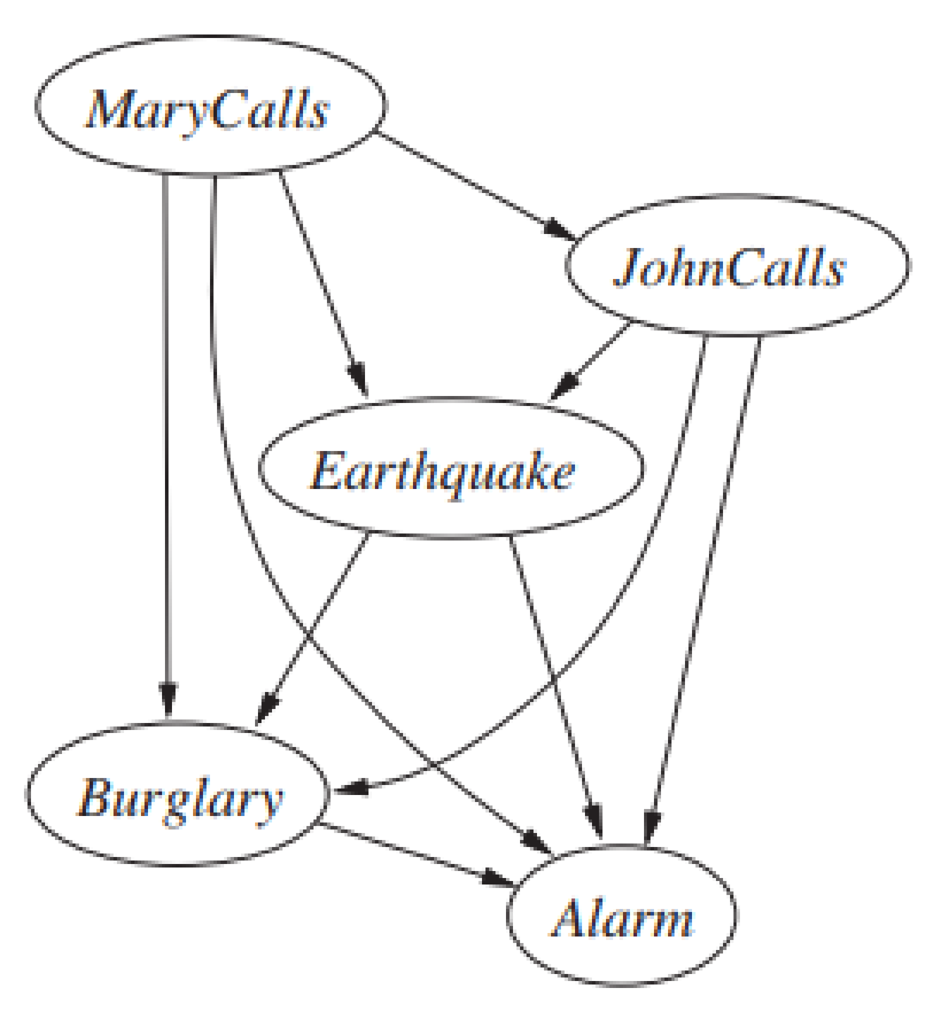


- با دانستن وضعیت زنگ، تماس ماری و جان اطلاعاتی درباره صدای موسیقی و صدای تلفن می‌دهد، نه درباره دزد
- وجود یا عدم وجود دزد، با دانستن وضعیت زنگ، احتمال زمین لرزه را بیشتر/کمتر می‌کند!
- نتیجه: دو یال بیشتر در شبکه

- ترتیب **تشخیصی** در گره‌ها ← شلوغی بیشتر
- ترتیب **عِلّی** در گره‌ها ← خلوتی بیشتر گراف

- اگر تریب گره‌ها بدتر باشد:

- Mary→John→Earthquake→Burglary→Alarm



- شلوغ ترین حالت ممکن
  - هم اندازه جدول توزیع توام کامل
  - ۳۱ احتمال مجزا

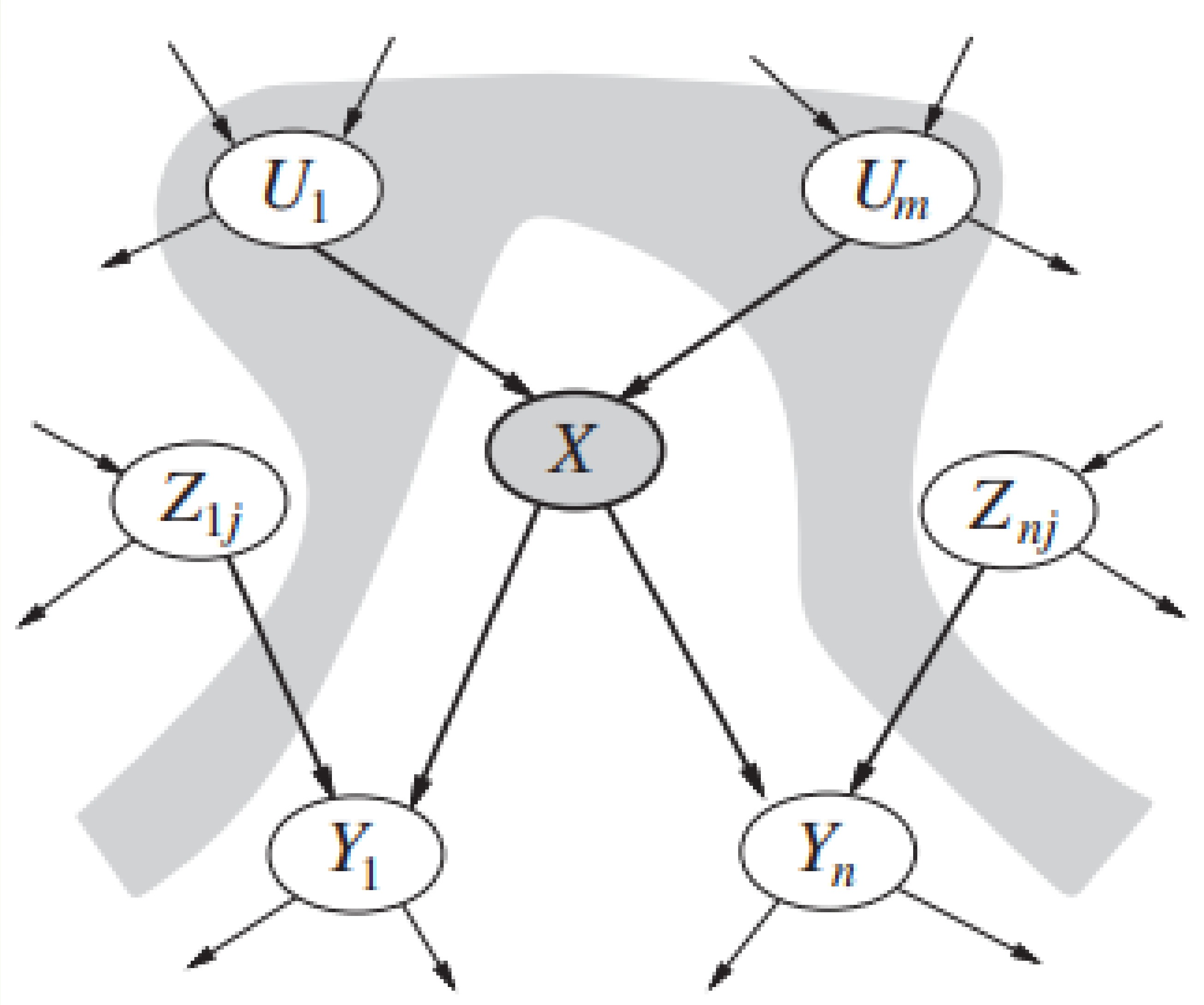
- هم این شبکه و هم شبکه قبلی، توزیع توام کامل یکسانی نسبت شبکه اصلی (درست) تولید می‌کنند
- دو شبکه اخیر، نتوانسته‌اند روابط استقلال شرطی را کامل احصا کنند.

روابط استقلال شرطی در شبکه بیز

- در شبکه بیز، هر متغیر، به شرط داشتن وضعیت والدهایش، از سایر متغیرهای پیشین خود (طبق ترتیب شماره گذاری متغیرها) مستقل است.
- طبق توپولوژی شبکه بیز، می‌توان گفت:

هر متغیر، به شرط داشتن وضعیت والدهایش، از متغیرهای غیر فرزند خود مستقل است.

- به فرزندان وابسته است
- مثلا وضعیت تماس ماری، احتمال صدای زنگ را بیشتر می‌کند

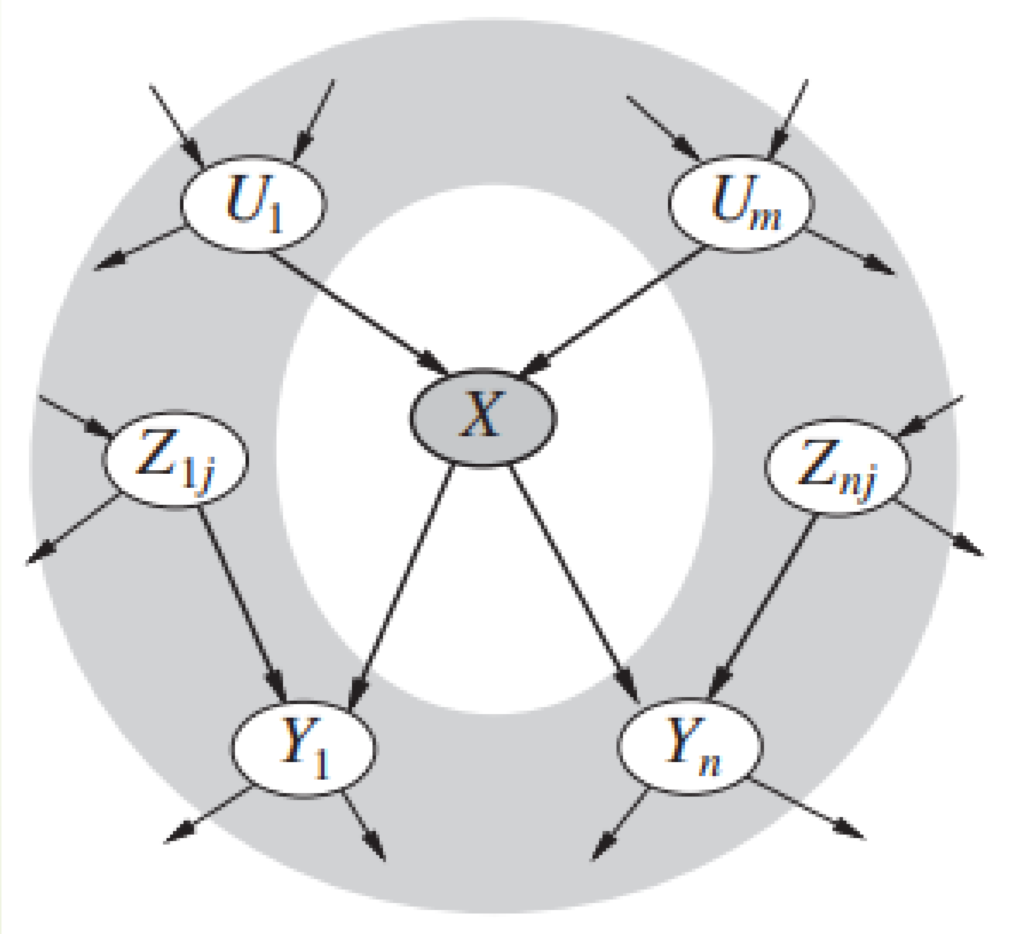


❖ X از متغیرهای Z مستقل است، اگر وضعیت والدها (متغیرهای U) را بدانیم.

❖ طبق توپولوژی شبکه بیز:

هر متغیر، به شرط داشتن وضعیت والدها، فرزندان و والدهای فرزندان، از همه متغیرهای دیگر مستقل است.

مجموعه فوق را لایه مارکف (Markov Blanket) می‌گویند.



استنتاج دقیق در شبکه بیزی

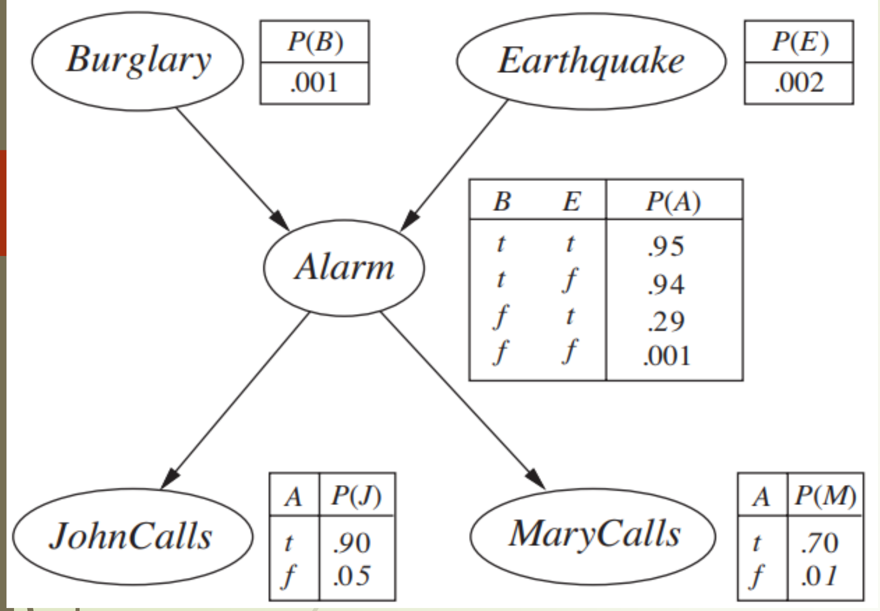
- وظیفه اصلی هر سیستم استنتاج احتمالاتی (مانند شبکه بیز) آن است که احتمال پسین برخی از متغیرها را به شرط دانستن برخی دیگر، تعیین کند.
  - برخی اول را متغیرهای پرسش (**query**) می‌نامیم
  - برخی دوم را وقایع (events) مشاهده شده یا متغیرهای مشاهده شده (**evidence**) می‌گوییم
- X: متغیر پرسش
- E متغیرهای مشاهده $(E_1,E_2,...E_m)$
  - e: یک مشاهده مشخص
- Y: سایر متغیرها (نه مشاهده شده، نه پرسش شده. $(Y_1,,...Y_l$
  - متغیرهای پنهان
- یک پرسش تیپیک: $P(X|e)$

استنتاج با سرشماری

- هر پرسش شرطی را میتوان با کمک جمع تعدادی توزیع توام کامل نوشت:

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

- چرا برحسب توزیع کامل؟
  - زیرا با شبکه بیز، توزیع کامل را به راحتی داریم، اما توزیع ناقص(!) مورد درخواست را نه!
- با شبکه بیز، توزیع کامل بر حسب ضرب توزیع‌های شرطی قابل بیان است.



همچنان استنتاج با سرشماری

- مثال:

$$\mathbf{P}(Burglary \mid JohnCalls = true, MaryCalls = true)$$

$$\mathbf{P}(B \mid j, m) = \alpha\, \mathbf{P}(B, j, m) = \alpha \sum_{e} \sum_{a} \mathbf{P}(B, j, m, e, a,)$$

- مثلا برای Burglary= true :

$$P(b \mid j, m) = \alpha \sum_{e} \sum_{a} P(b) P(e) P(a \mid b, e) P(j \mid a) P(m \mid a)$$

- و برای کاهش بار محاسباتی:

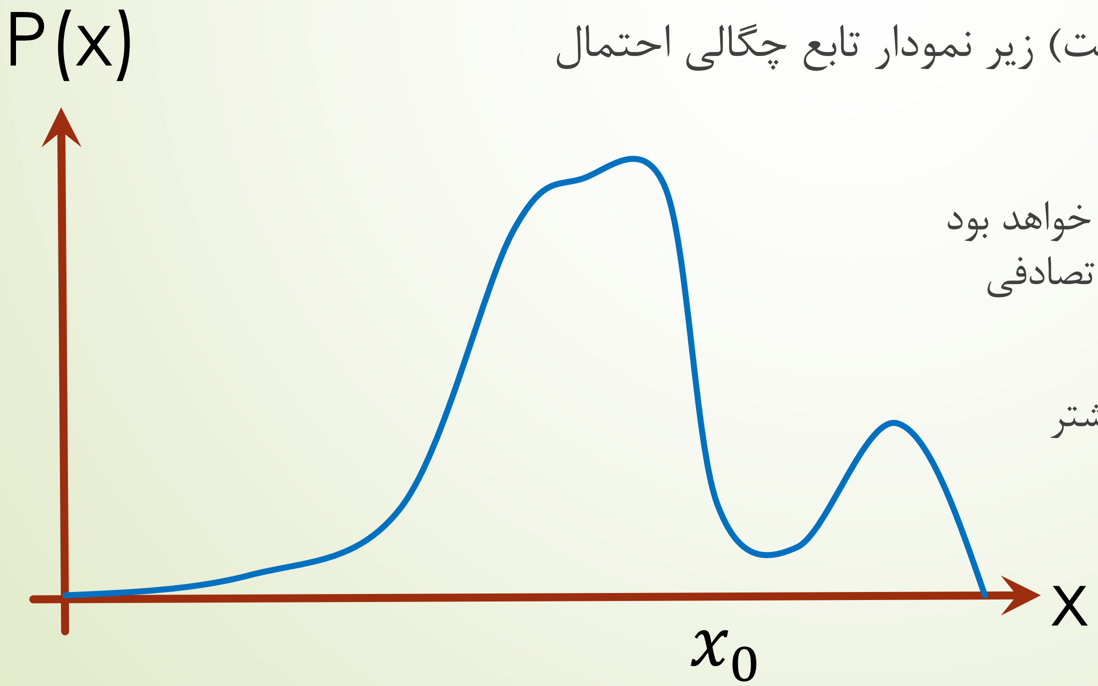
$$P(b \mid j, m) = \alpha\, P(b) \sum_{e} P(e) \sum_{a} P(a \mid b, e) P(j \mid a) P(m \mid a)$$

استنتاج تقریبی

- استنتاج دقیق، ممکن است خیلی هزینه بر باشد
  - برحسب شبکه و پرسش
  - تعداد متغیرهای پنهان بیشتر ← تعداد سرشماری های تودرتوی بیشتر ← هزینه بالاتر
  - گرچه با کمک شبکه بیز، محاسبه هر توزیع توام کامل کم هزینه شده است، ولی تعداد محاسبات ممکن است بر حسب پرسش، خیلی بالا باشد.
- نیاز به روش‌هایی کم هزینه‌تر
  - گرچه کم دقت‌تر
- روش‌های نمونه‌برداری (تولید نمونه تصادفی)
  - دقت وابسته به تعداد نمونه‌ها

الگوریتم نمونه‌برداری مستقیم

- تولید نمونه از روی یک توزیع احتمال مشخص
- مثلا متغیر تصادفی سکه، دارای دو حالت با احتمال <0.5, 0.5> است (توزیع احتمال)
- تولید نمونه برای این توزیع احتمال، شبیه به پرتاب سکه به دفعات مکرر است.
- اگر منبعی برای تولید عدد تصادفی با توزیع یکنواخت بین بازه [0, 1] داشته باشیم، میتوان برای هر تک متغیر دارای هر نوع توزیعی، نمونه تصادفی تولید کرد
  - گسسته: تقسیم کردن بازه ۰ تا ۱ به بخشهایی با طول متناسب با احتمال مربوطه به هر خروجی برای متغیر
  - پیوسته: همین رویکرد با توجه به انتگرال (مساحت) زیر نمودار تابع چگالی احتمال
    - مساحت زیر نمودار = ۱
    - نمونه تصادفی تولید شده، نقطه‌ای روی محور X خواهد بود که انتگرال (مساحت) نمودار تا آنجا برابر با عدد تصادفی تولید شده باشد.
    - احتمال تولید نمونه در بازه‌های دارای چگالی بیشتر بیشتر است



تولید عدد تصادفی

- تمرین: تابعی بنویسید که یک عدد تصادفی تولید کند.
- عدد شبه تصادفی
- مثلا روش middle-square method
  - یک عدد مثلا ۵ رقمی اولیه به عنوان seed اولیه در نظر می‌گیرد
  - Seed را به توان ۲ می‌رساند، ۱۰ رقمی میشود (اگر نشد صفر اضافه میکند تا بشود)
  - سپس ۵ رقم وسط را به عنوان عدد تصادفی تولید شده ارائه می‌کند
  - همان ۵ رقم را به عنوان seed برای تولید عدد تصادفی بعدی در نظر می‌گیرد.
  - seed اولیه را هم مثلا از روی ساعت سیستم می‌توان پیدا کرد یا ...

ادامه الگوریتم نمونه‌برداری مستقیم

- یک الگوریتم اولیه نمونه‌برداری برای شبکه بیز (بدون در نظر گرفتن مشاهدات احتمالی)
  - تولید مقدار تصادفی برای هر متغیر، به ترتیب توپولوژیک شبکه (بالا به پایین)
- الگوریتم:
  - از بالای شبکه بیز شروع کن.
  - متغیر اول دارای احتمال پیشین است، با همان احتمال یک مقدار تصادفی برایش تولید کن.
  - برای متغیرهای پایین تر، توزیع احتمال هر متغیر را به شرط آنکه والدهایشان مقادیر قبلا تولید شده را داشته باشند، در نظر بگیر
  - طبق همان توزیع احتمال، برایشان عدد تصادفی تولید کن
  - برای تمام متغیرها این کار را انجام بده ← یک نمونه تصادفی کامل شامل کل متغیرها تولید شده است.
- با همین روش می‌توان مثلا ۱۰۰۰ نمونه تصادفی تولید کرد
- فراوانی هر نمونه متناسب با احتمال واقعی تولید همان نمونه خواهد بود، تقریبا.
- با مجموعه نمونه‌های تولید شده، می‌توان توزیع توام کامل را برای مقادیر مختلف متغیرها **تقریب** زد
  - چگونه؟

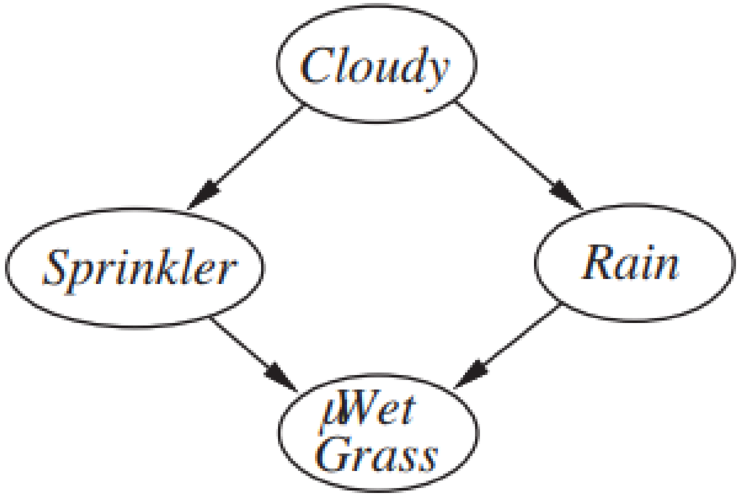
❖ الگوریتم تولید نمونه تصادفی از شبکه بیزی

function PRIOR-SAMPLE(bn) **returns** an event sampled from the prior specified by bn
  inputs: bn, a Bayesian network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$

  $\mathbf{x} \leftarrow$ an event with n elements
  foreach variable X_i **in** $X_1, \ldots, X_n$ **do**
    $\mathbf{x}[i] \leftarrow$ a random sample from $\mathbf{P}(X_i \mid parents(X_i))$
  return $\mathbf{x}$

❖ مثال:

$P(C)=.5$



C	$P(S)$
t	.10
f	.50

C	$P(R)$
t	.80
f	.20

S	R	$P(W)$
t	t	.99
t	f	.90
f	t	.90
f	f	.00

1. Sample from $\mathbf{P}(Cloudy) = \langle 0.5, 0.5 \rangle$, value is $true$.
2. Sample from $\mathbf{P}(Sprinkler \mid Cloudy = true) = \langle 0.1, 0.9 \rangle$, value is $false$.
3. Sample from $\mathbf{P}(Rain \mid Cloudy = true) = \langle 0.8, 0.2 \rangle$, value is $true$.
4. Sample from $\mathbf{P}(WetGrass \mid Sprinkler = false, Rain = true) = \langle 0.9, 0.1 \rangle$, value is $true$.

❖ در این مورد، الگوریتم نمونه [true, false, true, true] را تولید کرده است.