Zaključivanje u Bayesovim mrežama

Poglavlje 14.4–5

Prema slajdovima Stuarta Russella (hvala)!

Sadržaj

- Egzaktno zaključivanje enumeracijom (pobrojavanjem)
- Egzaktno zaključivanje eliminacijom varijabli
- Približno zaključivanje statističkom/stohastičkom simulacijom
- Približno zaključivanje metodom "Markov chain Monte Carlo"

Očita pitanja — problemi zaključivanja

Bayesova mreža je bazirana na uvjetnim vjerojatnostima i uvjetnim zavisnostima (nezavisnostima) pojedinih varijabli.

Prirodna pitanja — problemi zaključivanja:

– određivanje međuzavisnosti pojedinih čvorova,

– računanje a priori (početne) vjerojatnosti, a ne uvjetne,

– širenje (propagiranje) dokaza unaprijed kroz mrežu
$=$ duž puta u smjeru strelica $=$ kauzalni smjer,

– širenje (propagiranje) dokaza unatrag kroz mrežu
$=$ duž puta obrnuto od smjera strelica $=$ dijagnostički smjer,

– tzv. "skretanje" $=$ kombinacija natrag–naprijed,

– širenje (propagiranje) dokaza s višestrukim putevima,
kad imamo više puteva između dva čvora (ovo je teško).

Oznake i standardni upit — ponavljanje

Oznake — kao i ranije (kod zaključivanja enumeracijom), neka je

$\mathbf{X}$ = popis varijabli upita, tj. traženih varijabli,

$\mathbf{E}$ = popis dokaznih varijabli i

$\mathbf{e}$ = popis njihovih opaženih = dokaznih vrijednosti,

tj. $\mathbf{E} = \mathbf{e}$ je konkretni opaženi događaj,

$\mathbf{Y}$ = popis svih preostalih, tzv. skrivenih varijabli.

Dakle, unija $\mathbf{X} \cup \mathbf{E} \cup \mathbf{Y}$ daje skup svih varijabli u problemu (mreži).

U praksi je interesantan upit oblika $\mathbf{P}(\mathbf{X} \,|\, \mathbf{e})$, tj. traži se uvjetna združena distribucija varijabli $\mathbf{X}$, uz dane dokaze $\mathbf{E} = \mathbf{e}$.

Radi jednostavnosti, obično gledamo upite sa samo jednom varijablom, tj. $\mathbf{X} = \{X\}$ — odmah će biti jasno i zašto.

Ako je $\mathbf{E} = \emptyset$, tj. ako nema dokaza, onda, umjesto uvjetnih, dobivamo bezuvjetne ili a priori vjerojatnosti.

Zadaci zaključivanja — u Bayesovoj mreži

Jednostavni upiti:

izračunaj naknadnu (marginalnu) distribuciju $\mathbf{P}(X \mid \mathbf{E} = \mathbf{e})$,

na primjer, vjerojatnost da "nemam goriva u rezervoaru"

$$P(noGas \mid Gauge = empty, Lights = on, Starts = false)$$

(engl. gauge = mjerač za gorivo).

Konjunktivni upiti (upiti s više varijabli): svode se na jednostavne!

Koristimo produktno/lančano pravilo — prirodno za Bayesove mreže

$$\mathbf{P}(X_i, X_j \mid \mathbf{E} = \mathbf{e}) = \mathbf{P}(X_i \mid \mathbf{E} = \mathbf{e})\, \mathbf{P}(X_j \mid X_i, \mathbf{E} = \mathbf{e}).$$

Ovo vrijedi i bez uvjetne nezavisnosti X_i, X_j uz uvjete $\mathbf{E} = \mathbf{e}$.

Uvjetna nezavisnost $\Longrightarrow$ pojednostavljenje zadnjeg faktora (nema X_i).

Zadaci zaključivanja — općenito

Osnovni zadaci vjerojatnosnog zaključivanja u Bayesovoj mreži služe za donošenje složenijih zaključaka u problemu.

Optimalne odluke: mreže odlučivanja uključuju informacije o korisnosti; vjerojatnosno zaključivanje je potrebno za računanje/procjenu

$$P(outcome \mid action, evidence)$$

vjerojatnosti nekog rezultata (ishoda) za dane akcije i dokaze.

Vrijednost informacije: koji dokaz treba prvo pronaći ili ispitati?

Analiza osjetljivosti: koje vjerojatnosti su najkritičnije?

Objašnjenje (kad to nekog zanima): zašto trebam novi starter?

Najvjerojatnije objašnjenje $\arg\max_x P(X = x \mid \mathbf{E} = \mathbf{e})$,

Ovaj "zašto" nije samo pitanje osobne znatiželje, već pokretač istraživanja i razvoja u industriji, medicini, . . .

Zaključivanje enumeracijom — ponavljanje

To je najjednostavnija metoda za probabilističko zaključivanje.

Bazira se na potpunoj združenoj distribuciji svih varijabli i svodi se na
pobrojavanje svih mogućnosti iz upita i
zbrajanje pripadnih vjerojatnosti, uz normalizaciju zbroja 1.

Za jednostavni upit $\mathbf{P}(X \,|\, \mathbf{e})$, to znači
fiksiraj dokaze i zbroji preko svih ostalih $=$ skrivenih varijabli iz $\mathbf{Y}$.

Tražena distribucija $\mathbf{P}(X \,|\, \mathbf{e})$ računa se "marginalizacijom", kao zbroj

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

Suma ide po svim mogućim kombinacijama vrijednosti $\mathbf{Y} = \mathbf{y}$ skrivenih varijabli $\mathbf{Y}$, a α je normalizacijska konstanta da konačni zbroj bude 1.

Svrha marginalizacije je eliminacija skrivenih varijabli,
a algoritam za to $=$ sumiranje po njima.

Zaključivanje enumeracijom — Bayesove mreže

U Bayesovoj mreži, potpuna združena distribucija ima poseban prikaz, preko produkta tablica uvjetnih vjerojatnosti (CPT) u čvorovima mreže

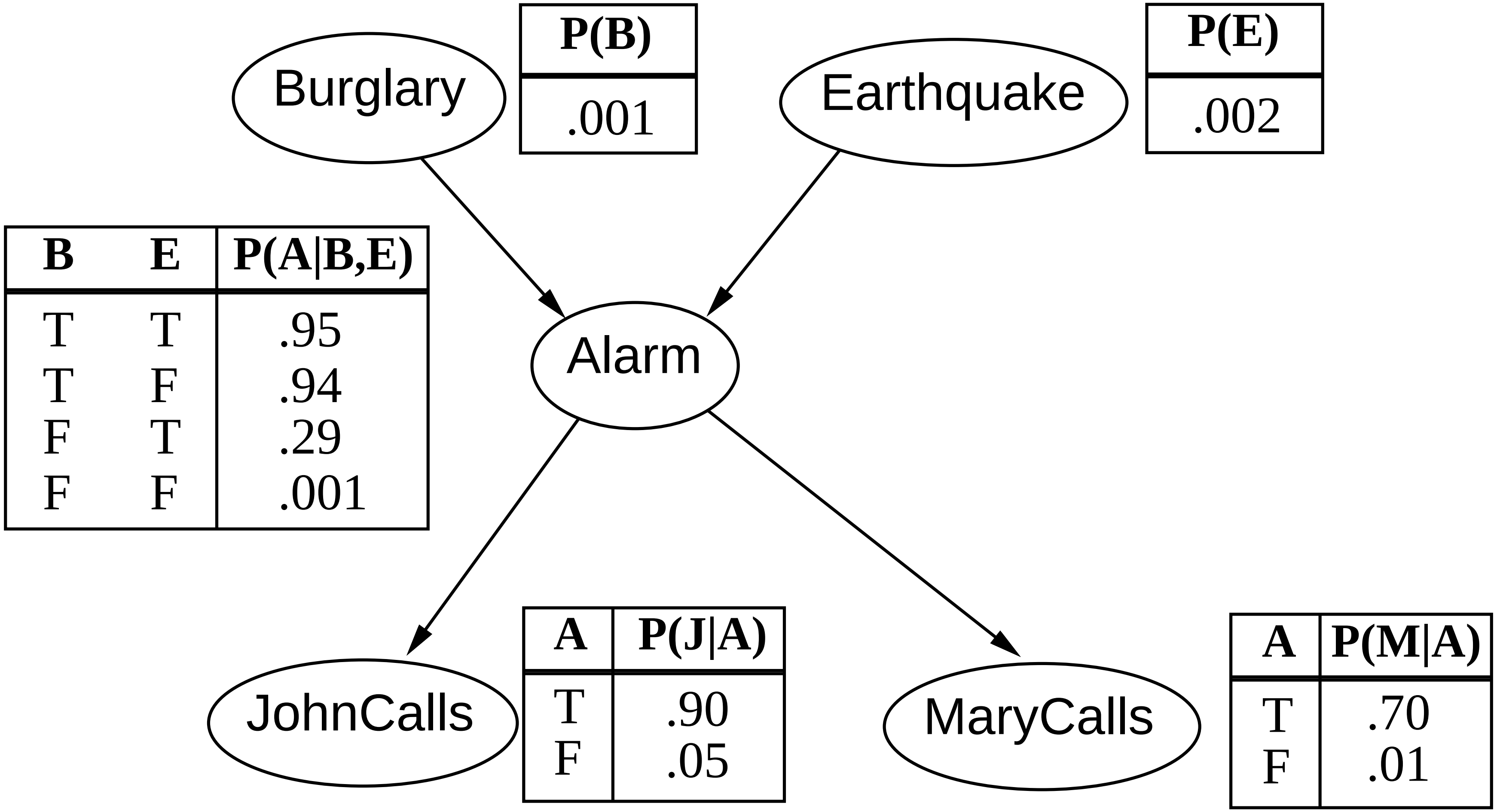
$$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i \,|\, Parents(X_i))$$

Ovo množenje je množenje vektora po točkama, svaki sa svakim.
Kroneckerov, direktni ili tenzorski produkt — standardna oznaka je $\times$.

Napomena: $X_1, \ldots, X_n$ su sve varijable u problemu (mreži), tj. ako je X varijabla upita, a $\mathbf{E}$ su dokazne varijable, onda su skrivene varijable
$\mathbf{Y} = \{X_1, \ldots, X_n\} \setminus (\{X\} \cup \mathbf{E})$.

Kod računanja upita $\mathbf{P}(X \,|\, \mathbf{e})$ u Bayesovoj mreži, gornja formula se može iskoristiti na dva načina:
– za "pametno" zbrajanje = pobrojavanje ili enumeracija,
– za izravnu eliminaciju skrivenih varijabli = eliminacija varijabli.

Primjer — Provalnička mreža



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Zaključivanje enumeracijom (pobrojavanjem)

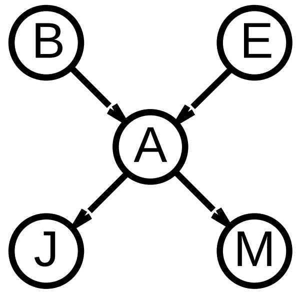
Algoritam enumeracije ili pobrojavanja: "ponešto pametan" način da "odsumiramo" skrivene varijable iz združene distribucije, bez da stvarno formiramo njezinu eksplicitnu reprezentaciju, već zbrajamo produkte uvjetnih vjerojatnosti iz mreže.

Primjer jednostavnog upita za provalničku mrežu

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

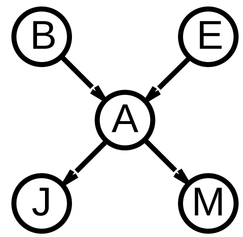
Skrivene varijable za ovaj upit su $Earthquake$ i $Alarm$.

U skraćenim oznakama (jedno slovo), enumeracija daje



$$\begin{aligned}\mathbf{P}(B \mid j, m) &= \mathbf{P}(B, j, m)/P(j, m) \\ &= \alpha\, \mathbf{P}(B, j, m) \\ &= \alpha \sum_e \sum_a \mathbf{P}(B, e, a, j, m).\end{aligned}$$

Zaključivanje enumeracijom — zapis preko CPT



Vrijednosti $\mathbf{P}(B, e, a, j, m)$ iz potpune združene distribucije zapišemo kao produkte vrijednosti iz CPT u mreži — to je osnovna formula za provalničku Bayesovu mrežu:

$$\mathbf{P}(B, e, a, j, m) = \mathbf{P}(B)\, P(e)\, \mathbf{P}(a \,|\, B, e)\, P(j|a)\, P(m|a).$$

To uvrstimo u enumeracijsku sumu, pa dobijemo

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

U ovom obliku, za svaku vrijednost $B = b$, $B = \neg b$, treba izračunati sumu od 4 člana, po svim vrijednostima skrivenih varijabli $y = (e, a)$, svaki član je produkt od 5 faktora (ukupni broj čvorova).

Općenito, ako treba "odsumirati" većinu varijabli (skoro n), složenost u najgorem slučaju za mrežu s n Booleovih varijabli je $O(n2^n)$.

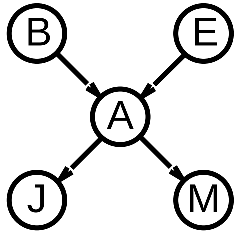
Zaključivanje enumeracijom — primjer

Pogledati: Eliminacija, str. 1.

Rezultat je $\mathbf{P}(B \mid j, m) = \langle\, 0.284172, 0.715818 \,\rangle$.

Zaključivanje enumeracijom — poboljšanje

Poboljšanje = izlučivanje faktora koji ne ovise o pojedinoj sumi.



Na primjer, $\mathbf{P}(B)$ ne ovisi o e, a, jer je na vrhu mreže.
Slično, $P(e)$ ne ovisi o a, jer je iznad njega u mreži.
Onda dobijemo

$$\mathbf{P}(B \mid j, m) = \alpha\, \mathbf{P}(B) \sum_e P(e) \sum_a \mathbf{P}(a \mid B, e)\, P(j|a)\, P(m|a).$$

Kad možemo izlučiti neki faktor ispred neke sume?

Svaki faktor = neka vrijednost iz CPT u danom čvoru,
a CPT u čvoru ovisi samo o čvoru i roditeljima čvora (ako ih ima)!

Dakle, faktor smije ispred svih suma koje nisu taj čvor ili njegovi roditelji.

Ostaje još pitanje dobrog "globalnog poretka" suma za izlučivanje.

Enumeracija — oblik izraza

Svaka suma = jedna skrivena varijabla ili čvor u mreži.

Sume poredamo (slijeva udesno) prema standardnom topološkom poretku čvorova u mreži — tako da roditelji dolaze prije djece.

Time dobivamo jedan korektan poredak suma za izlučivanje faktora.

Faktor iz danog čvora ide ispred svih onih suma koje idu po skrivenim varijablama (čvorovima) iza njega, po poretku čvorova u mreži.

$$\mathbf{P}(B \,|\, j, m) = \alpha\, \mathbf{P}(B) \sum_e P(e) \sum_a \mathbf{P}(a \,|\, B, e)\, P(j|a)\, P(m|a).$$

Čvor B je prvi, pa ide ispred svega (slučajno baš upit).

Prva skrivena varijable E je druga po redu. Pripadni faktor $P(e)$ možemo izlučiti iz svih naknadnih suma — ima još skrivenih varijabli!

Zadnja skrivena varijabla A je treća po redu, odmah iza prethodne — iz pripadne sume ne možemo ništa izlučiti.

Zaključivanje enumeracijom — skica algoritma

Računanje ovog izraza:

= petlja po svim čvorovima, prema poretku:

u svakom čvoru množimo odgovarajuću vrijednost iz pripadne CPT,

u skrivenim čvorovima još ide petlja za zbrajanje tih produkata,
po svim mogućim vrijednostima te varijable.

Normalizacijski faktor α se računa na kraju, zbrajanjem po dobivenom vektoru vrijednosti, a onda se distribucija normalizira.

Rekurzivno prebrojavanje u dubinu: $O(n)$ prostorno, $O(d^n)$ vremenski.

Enumeracijski algoritam

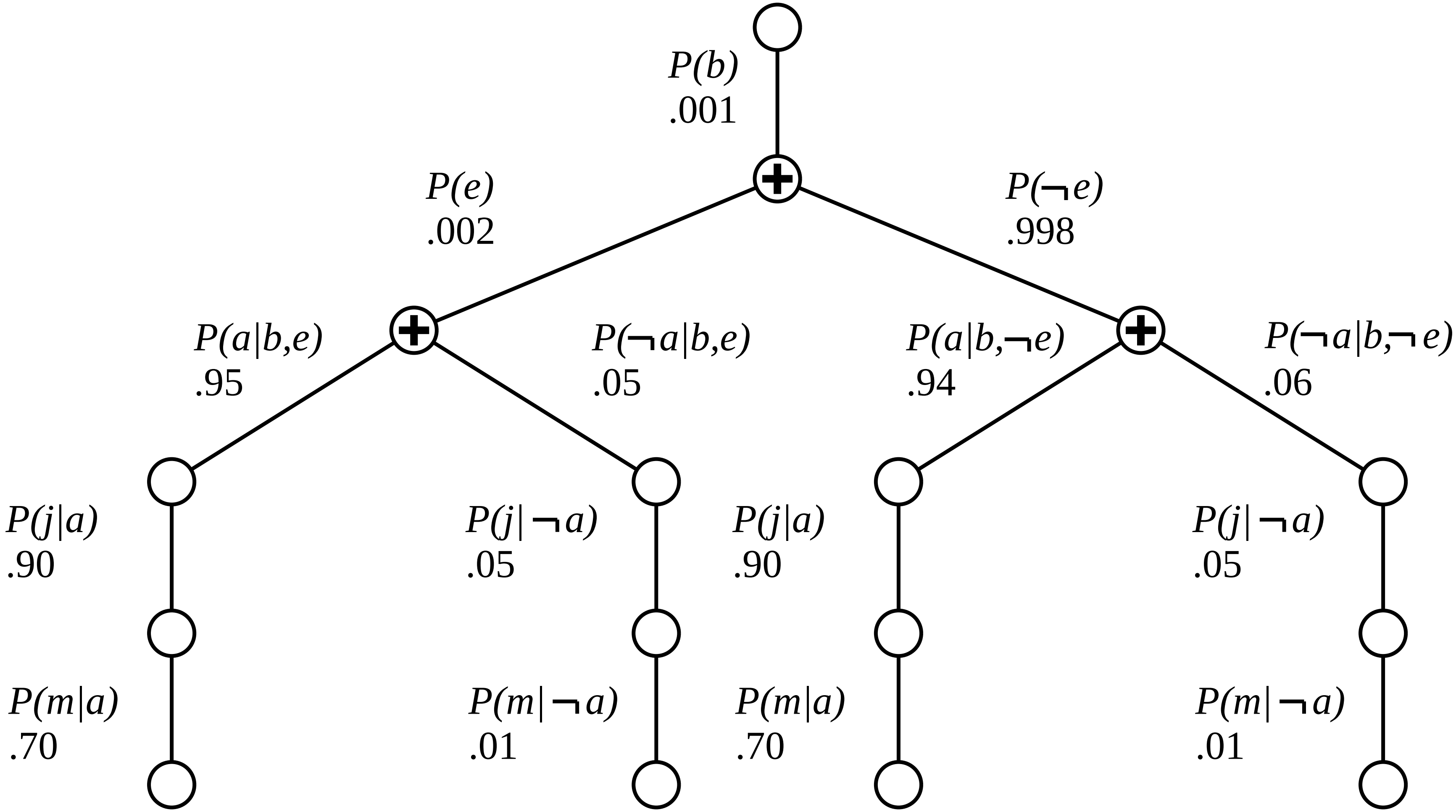
function ENUMERATION-ASK(X, $\mathbf{e}$, bn) **returns** a distribution over X
  inputs: X, the query variable
    $\mathbf{e}$, observed values for variables $\mathbf{E}$
    bn, a Bayesian network with variables $\{X\} \cup \mathbf{E} \cup \mathbf{Y}$

  $\mathbf{Q}(X) \leftarrow$ a distribution over X, initially empty
  for each value x_i of X **do**
    extend $\mathbf{e}$ with value x_i for X
    $\mathbf{Q}(x_i) \leftarrow$ ENUMERATE-ALL(VARS[bn], $\mathbf{e}$)
  return NORMALIZE($\mathbf{Q}(X)$)

---

function ENUMERATE-ALL($vars$, $\mathbf{e}$) **returns** a real number
  if EMPTY?($vars$) **then return** 1.0
  $Y \leftarrow$ FIRST($vars$)
  if Y has value y in $\mathbf{e}$
    then return $P(y \mid Pa(Y)) \times$ ENUMERATE-ALL(REST($vars$), $\mathbf{e}$)
    else return $\sum_y P(y \mid Pa(Y)) \times$ ENUMERATE-ALL(REST($vars$), $\mathbf{e}_y$)
      where $\mathbf{e}_y$ is $\mathbf{e}$ extended with $Y = y$

Evaluacijsko stablo



Enumeracija je neefikasna: ponovljeno računanje
na pr. računa $P(j|a)\,P(m|a)$ za svaku vrijednost e

Enumeracija — drugi oblik izraza

Ako prvo uzmemo čvor A, pa onda E, dobivamo izraz

$$\mathbf{P}(B \mid j, m) = \alpha \, \mathbf{P}(B) \sum_a P(j|a) \, P(m|a) \sum_e P(e) \, \mathbf{P}(a \mid B, e).$$

Optimalni poredak = NP-težak problem.

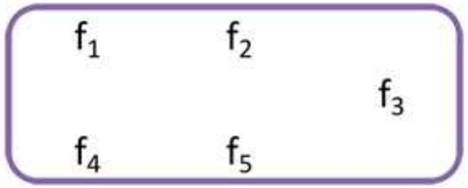
Zaključivanje eliminacijom varijabli

Eliminacija varijabli: treba napraviti sumacije zdesna nalijevo, spremiti međurezultate (faktore) za izbjegavanje ponovljenog računanja

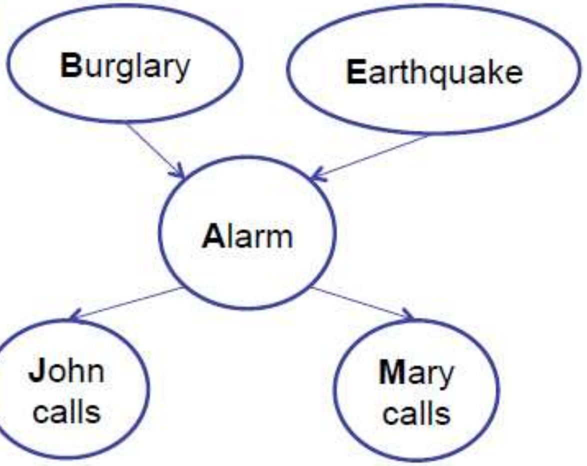
$$\mathbf{P}(B|j,m) = \alpha \underbrace{\mathbf{P}(B)}_{f_1(B)} \sum_e \underbrace{P(e)}_{f_2(E)} \sum_a \underbrace{\mathbf{P}(a|B,e)}_{f_3(A,B,E)} \underbrace{P(j|a)}_{f_4(A)} \underbrace{P(m|a)}_{f_5(A)} .$$

Primjer provalničke mreže i značenje faktora (na sljedećoj stranici).

Zaključivanje eliminacijom varijabli



B	P(B)
+b	0.001
¬b	0.999



E	P(E)
+e	0.002
¬e	0.998

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	¬a	0.05
+b	¬e	+a	0.94
+b	¬e	¬a	0.06
¬b	+e	+a	0.29
¬b	+e	¬a	0.71
¬b	¬e	+a	0.001
¬b	¬e	¬a	0.999

A	J	P(J|A)
+a	+j	0.9
+a	¬j	0.1
¬a	+j	0.05
¬a	¬j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	¬m	0.3
¬a	+m	0.01
¬a	¬m	0.99

Kad uvrstimo dokaze, u terminima faktora, izraz se može napisati kao

$$\mathbf{P}(B|j,m) = \alpha f_1(B) \times \sum_e f_2(E) \times \sum_a f_3(A,B,E) \times f_4(A) \times f_5(A).$$

Koraci u eliminaciji varijabli

- Prvo, odsumiramo A iz produkta f_3, f_4 i f_5.
  To daje novi faktor $f_6(B, E)$, kojemu indeksi prolaze samo po B i E:

$$\begin{aligned} f_6(B, E) &= \sum_a f_3(A, B, E) \times f_4(A) \times f_5(A) \\ &= (f_3(a, B, E) \times f_4(a) \times f_5(a)) \\ &\qquad + (f_3(\neg a, B, E) \times f_4(\neg a) \times f_5(\neg a)). \end{aligned}$$

  Sada ostajemo s izrazom:

$$\mathbf{P}(B|j, m) = \alpha f_1(B) \times \sum_e f_2(E) \times f_6(B, E).$$

(za računanje pogledati: Eliminacija, str. 2–3.)

Koraci u eliminaciji varijabli (nastavak)

- Nadalje, odsumiramo E iz produkta f_2 i f_6:

$$f_7(B) = \sum_e f_2(E) \times f_6(B,E)$$
$$= f_2(e) \times f_6(B,e) + f_2(\neg e) \times f_6(B,\neg e)$$

  što daje

$$\mathbf{P}(B|j,m) = \alpha f_1(B) \times f_7(B).$$

  Ovaj izraz izvrednjava se množenjem po točkama
  i normalizacijom rezultata.

Prethodni primjer pokazuje potrebu za samo dvije operacije:
– produkt po točkama para faktora
– zbrajanje po varijabli iz produkta faktora.
Stvarno, treba i treća operacija — za uvrštavanje vrijednosti varijable
– fikisiranje neke vrijednosti u faktoru = projekcija!

Eliminacija varijabli: osnovne operacije

Zbrajanje po varijabli iz produkta faktora:

konstante se izbace ispred sumacije

zbroje se podmatrice u produkt po točkama iz preostalih faktora

$$
\begin{aligned}
\sum_x f_1 \times \cdots \times f_k &= f_1 \times \cdots \times f_i \sum_x f_{i+1} \times \cdots \times f_k \\
&= f_1 \times \cdots \times f_i \times f_{\bar{X}}
\end{aligned}
$$

uz pretpostavku da $f_1, \ldots, f_i$ ne ovisi o X

Produkt po točkama faktora f_1 i f_2:

$$
\begin{aligned}
f_1(x_1, \ldots, x_j, y_1, \ldots, y_k) \;\times\; & f_2(y_1, \ldots, y_k, z_1, \ldots, z_l) \\
= \;& f(x_1, \ldots, x_j, y_1, \ldots, y_k, z_1, \ldots, z_l)
\end{aligned}
$$

Na primjer $f_1(a, b) \times f_2(b, c) = f(a, b, c)$

Algoritam eliminacije varijabli

function ELIMINATION-ASK(X, $\mathbf{e}$, bn) **returns** a distribution over X
  inputs: X, the query variable
    $\mathbf{e}$, evidence specified as an event
    bn, a belief network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$

  $factors \leftarrow [\,]$; $vars \leftarrow$ REVERSE(VARS[bn])
  for each var **in** $vars$ **do**
    $factors \leftarrow$ [MAKE-FACTOR(var, $\mathbf{e}$)|$factors$]
    if var is a hidden variable **then** $factors \leftarrow$ SUM-OUT(var, $factors$)
  return NORMALIZE(POINTWISE-PRODUCT($factors$))

Za korektan algoritam, operacija MAKE-FACTOR mora generirati faktor s uvrštenim dokazima (za smanjenje veličine faktora).

Ovdje:

– pokazati slajdove Dana Kleina (CS 188, Lecture 15, str. 14–32)
– vrati se na računanje = Eliminacija, str. 2–3.

Eliminacija varijabli — drugi oblik izraza

Eliminaciju možemo napraviti i na drugi način — za izraz:

$$\mathbf{P}(B \mid j, m) = \alpha \, \mathbf{P}(B) \sum_a P(j|a) \, P(m|a) \sum_e P(e) \, \mathbf{P}(a \mid B, e).$$

Ovdje prvo eliminiramo varijablu E, pa onda A.

Pogledati: Eliminacija, str. 4.

Nebitne varijable

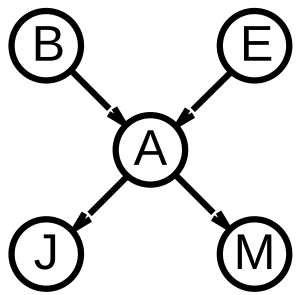
Razmotrimo upit $P(JohnCalls \mid Burglary = true)$

$$P(J|b) = \alpha P(b) \sum_e P(e) \sum_a P(a \mid b, e) P(J|a) \sum_m P(m|a)$$

Suma po m je identički jednaka 1; M je nebitan za upit!

Teorem 1. Y je nebitna osim ako $Y \in Ancestors(\{X\} \cup \mathbf{E})$.

Ovdje je $X = JohnCalls$, $\mathbf{E} = \{Burglary\}$ i
$Ancestors(\{X\} \cup \mathbf{E}) = \{Alarm, Earthquake\}$
pa je $MaryCalls$ nebitno.



(Slično s ulančavanjem unatrag iz upita u KB s Hornovim klauzulama).

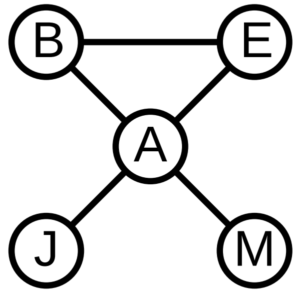
Nebitne varijable (nastavak)

Definicija. Moralni graf Bayesove mreže = poženi (spoji) sve roditelje i ispusti strelice.

Definicija. $\mathbf{A}$ je m-separiran od $\mathbf{B}$ s $\mathbf{C}$, akko je separariran s $\mathbf{C}$ u moralnom grafu.

Teorem 2. Y je nebitan, ako je m-separiran od X s dokazima $\mathbf{E}$.

Za upit $P(JohnCalls \mid Alarm = true)$, oboje, $Burglary$ i $Earthquake$ je nebitno.



Složenost egzaktnog zaključivanja

Jednostruko povezane mreže (ili poli-stabla):

– svaka dva čvora povezana su najviše jednim (neusmjerenim) putem

– vremenska i prostorna složenost eliminacija varijabli su $O(d^k n)$
($d =$ br. vrij. varijable, $k =$ br. roditelja, $n =$ broj čvorova).

Drugim riječima, linearno u ukupnom broju elemenata u CPT.

Složenost egzaktnog zaključivanja (nastavak)

Višestruko povezane mreže:

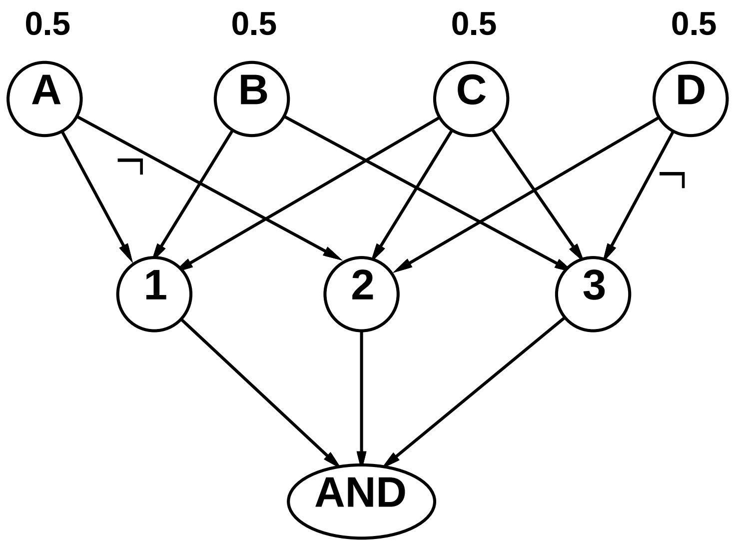
– mogu reducirati 3SAT u egzaktno zaključivanje $\Rightarrow$ NP-težak

– ekvivalentno **brojenju** 3SAT modela $\Rightarrow$ #P-potpun (striktno teži od NP-potpunih problema).

1. A v B v C

2. C v D v ¬A

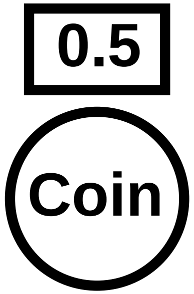
3. B v C v ¬D



Zaključivanje stohastičkom simulacijom

Osnovna ideja:

1) Uzmi N uzoraka iz uzoračke distribucije S
2) Izračunaj aproksimativnu a posteriori vjerojatnost $\hat{P}$
3) Pokaži da to konvergira prema pravoj vjerojatnosti P.



Sadržaj:

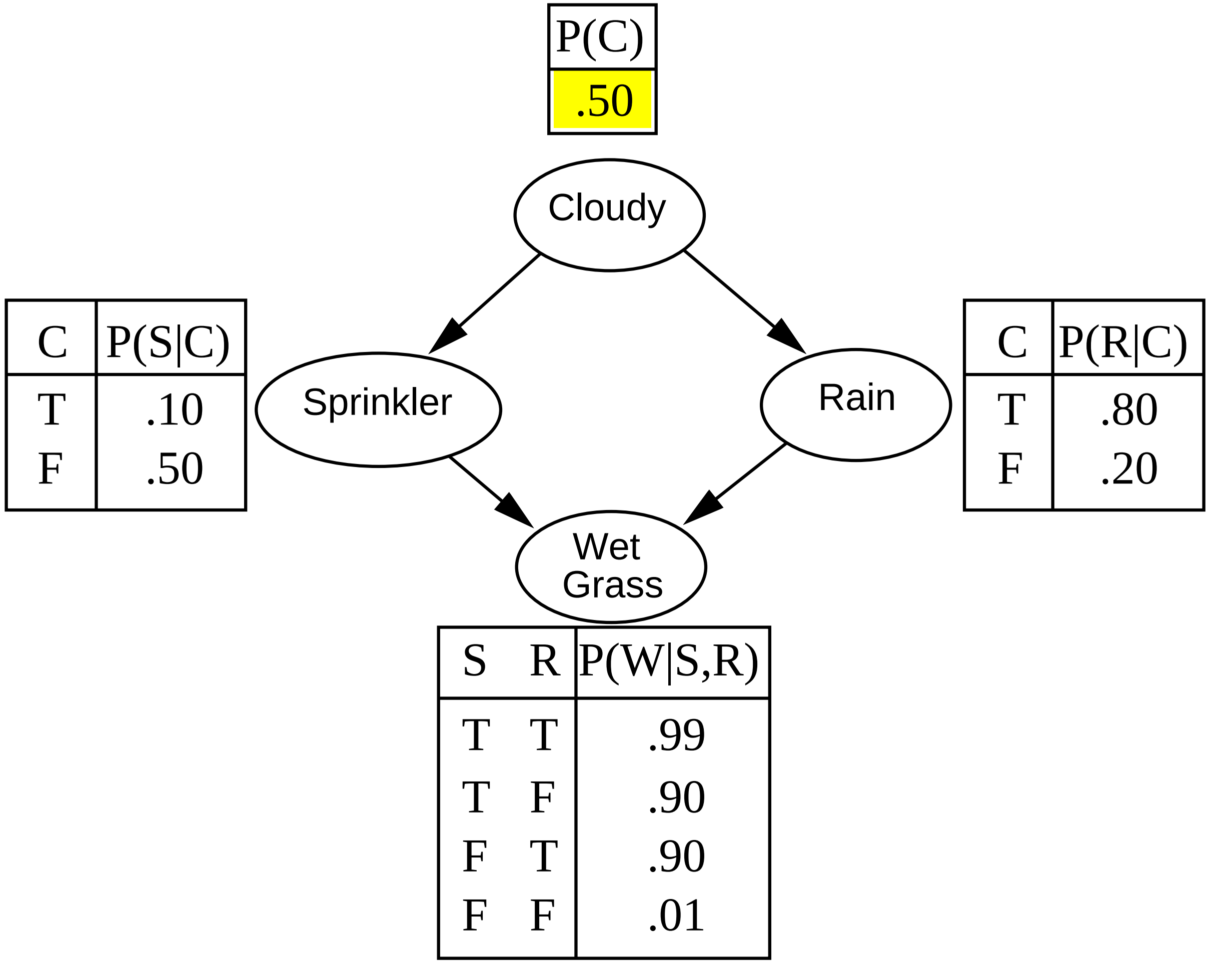
– Uzorkovanje iz prazne mreže
– Odbacivanje uzoraka: odbacuju se uzorci koji se ne slažu s dokazima
– Težinska izglednost/vjerodostojnost (engl. "likelihood weighting"): iskoriste se dokazi za stavljanje težina na uzorke
– Markov chain Monte Carlo (MCMC): uzme se uzorak iz stohastičkog procesa kojemu je stacionarna distribucija jednaka pravoj aposteriornoj distribuciji.

Uzorkovanje iz prazne mreže

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

  $\mathbf{x} \leftarrow$ an event with n elements
  for $i = 1$ **to** n **do**
    $x_i \leftarrow$ a random sample from $\mathbf{P}(X_i \mid parents(X_i))$
      given the values of $Parents(X_i)$ in $\mathbf{x}$
  return $\mathbf{x}$

Primjer



P(C)
.50

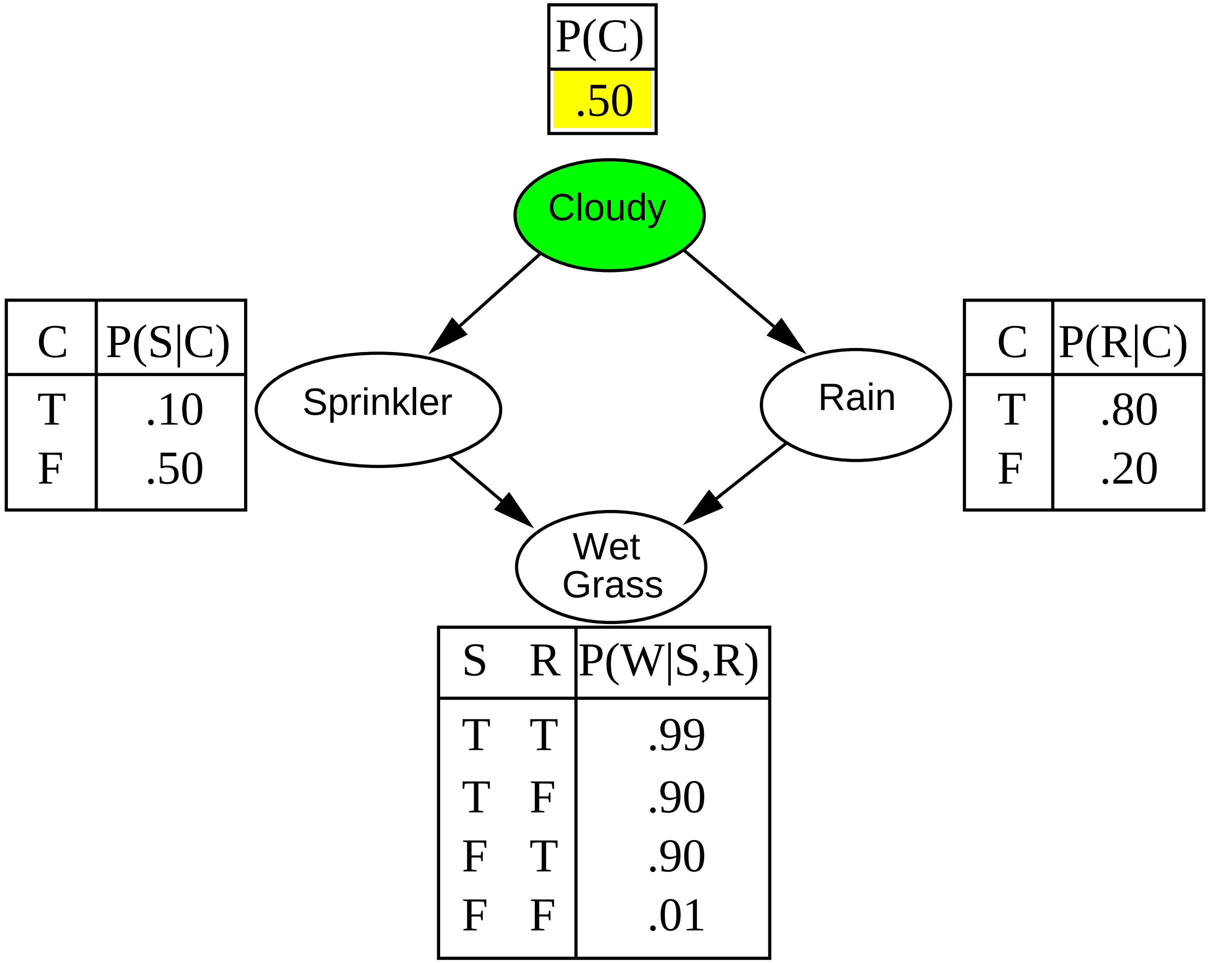
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Uzorak iz $\mathbf{P}(Cloudy) = \langle 0.5, 0.5 \rangle$

Primjer



P(C)
.50

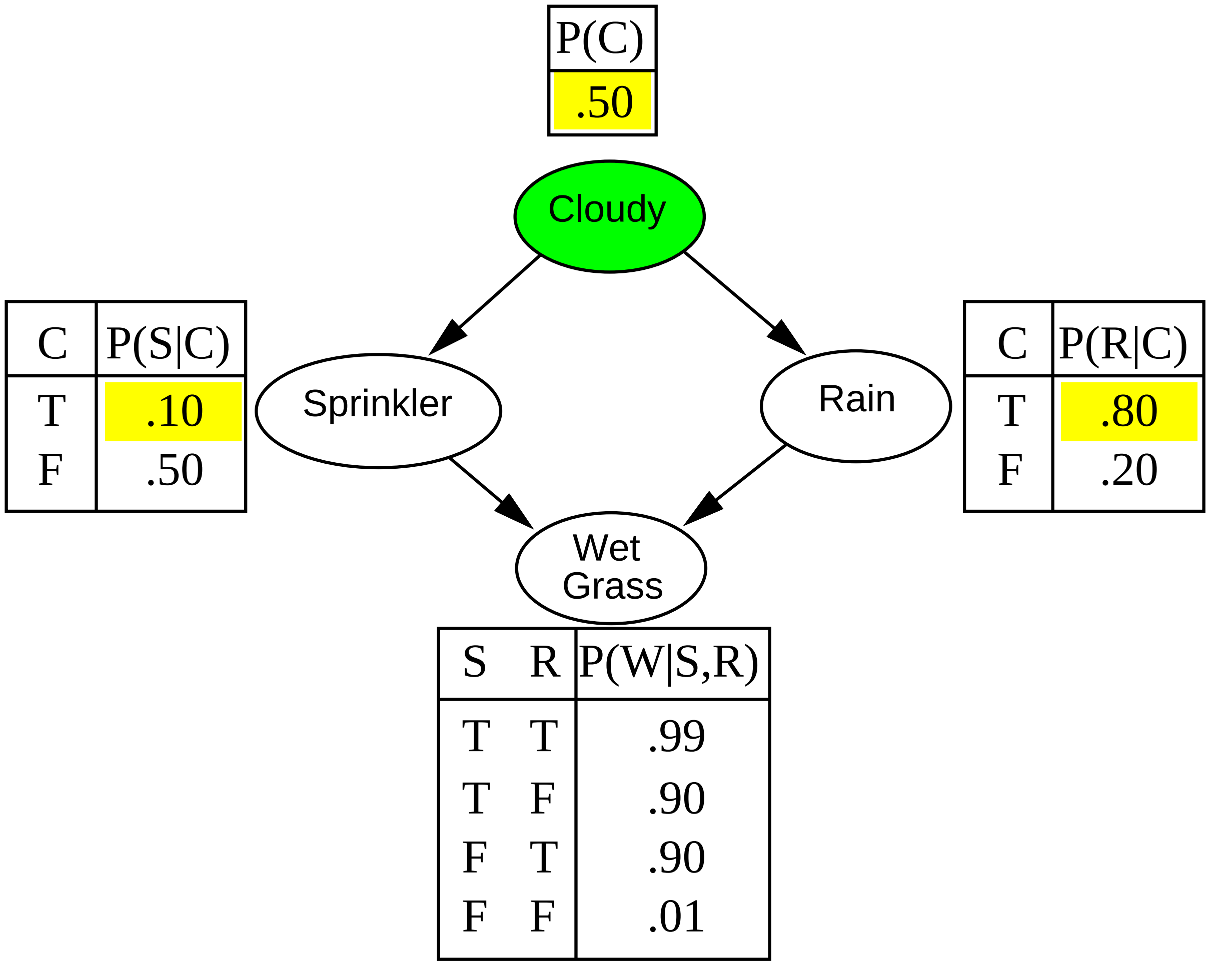
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Uzorak iz $\mathbf{P}(Cloudy) = \langle 0.5, 0.5 \rangle$, neka je $true$.

Primjer

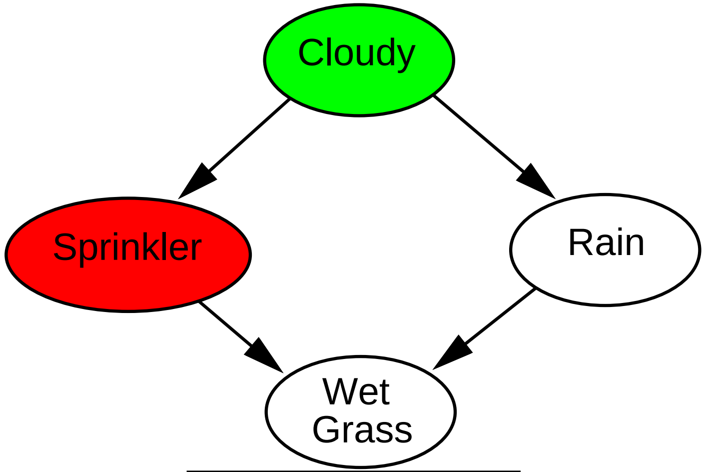


Uzorak iz $\mathbf{P}(Sprinkler \mid Cloudy = true) = \langle 0.1, 0.9 \rangle$

Uzorak iz $\mathbf{P}(Rain \mid Cloudy = true) = \langle 0.8, 0.2 \rangle$

Primjer

P(C)
.50



C	P(S|C)
T	.10
F	.50

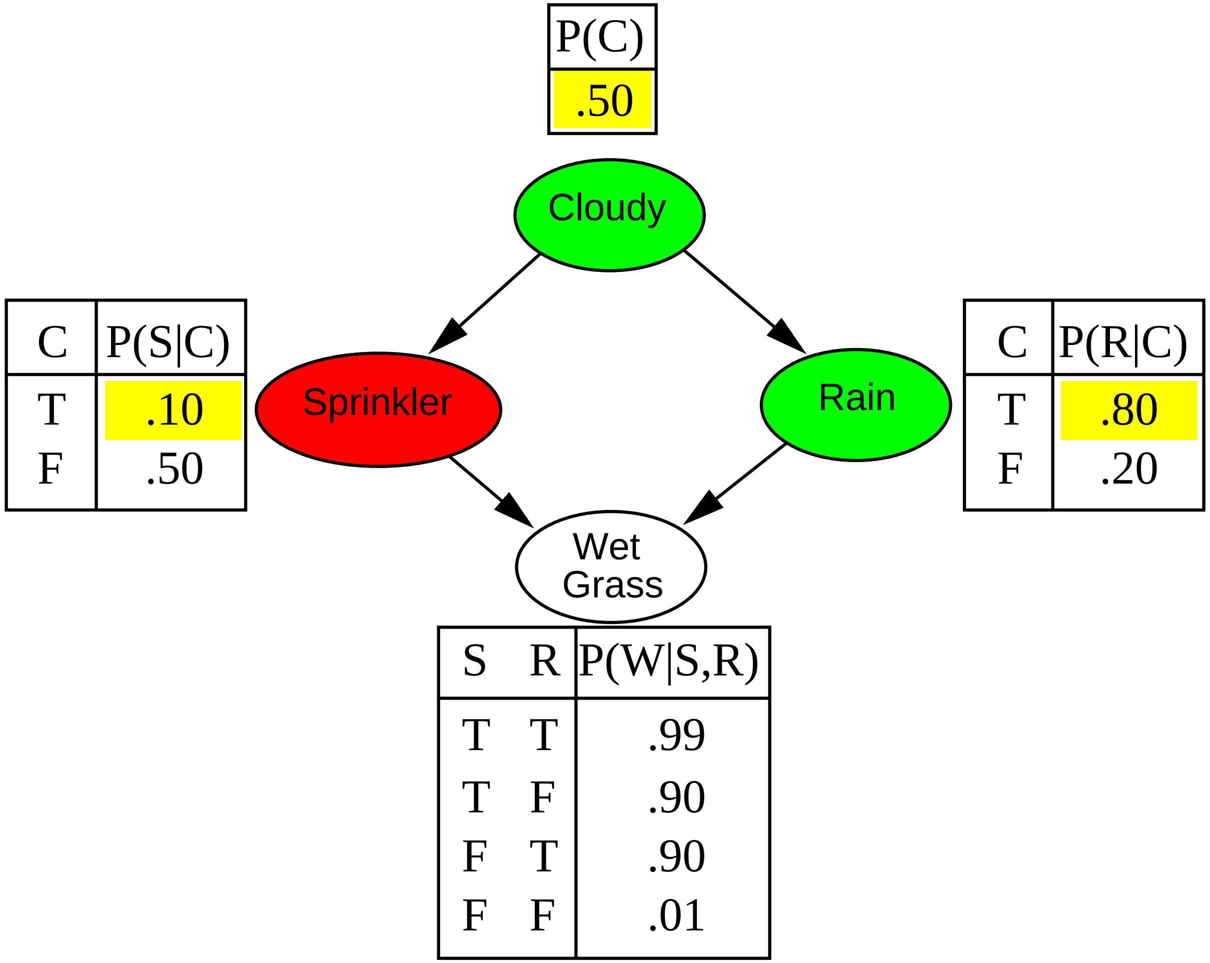
C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Uzorak iz $\mathbf{P}(Sprinkler \mid Cloudy = true) = \langle 0.1, 0.9 \rangle$ neka je *false*

Uzorak iz $\mathbf{P}(Rain \mid Cloudy = true) = \langle 0.8, 0.2 \rangle$

Primjer

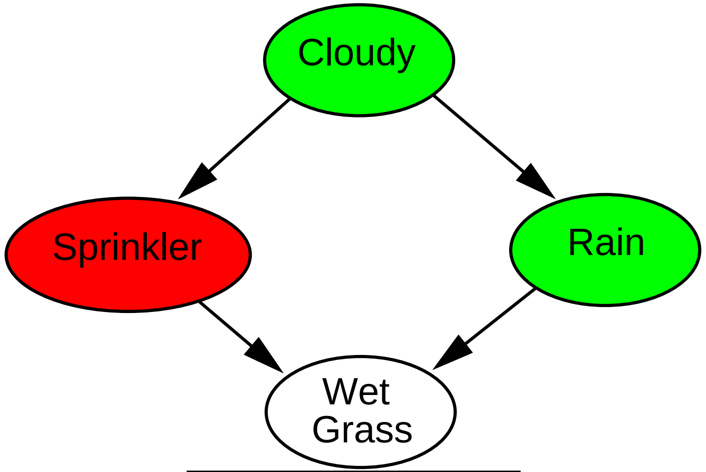


Uzorak iz $\mathbf{P}(Sprinkler \mid Cloudy = true) = \langle 0.1, 0.9 \rangle$ neka je *false*

Uzorak iz $\mathbf{P}(Rain \mid Cloudy = true) = \langle 0.8, 0.2 \rangle$ neka je *true*

Primjer

P(C)
.50



C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

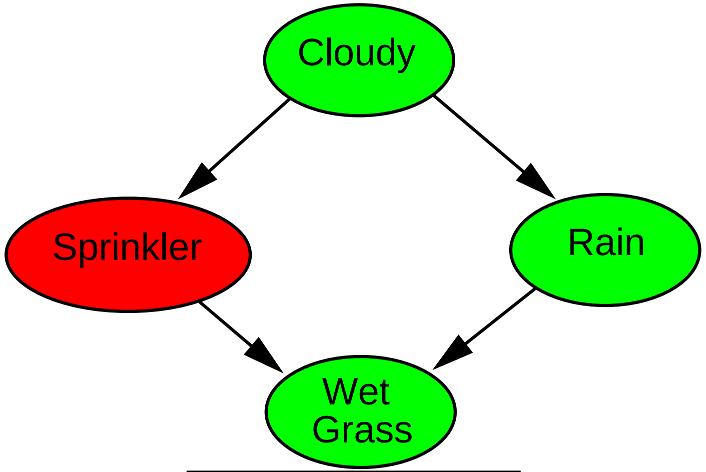
S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Uzorak iz $\mathbf{P}(WetGrass \mid Sprinkler = false, Rain = true) = \langle 0.9, 0.1 \rangle$

Primjer

P(C)
.50

C	P(S|C)
T	.10
F	.50



C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Uzorak iz $\mathbf{P}(WetGrass \mid Sprinkler = false, Rain = true) = \langle 0.9, 0.1 \rangle$
neka je *true*

Uzorkovanje iz prazne mreže (nastavak)

Vjerojatnost da PriorSample generira odgovarajući događaj

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

tj., stvarna apriori vjerojatnost!

Na pr. $S_{PS}(t, f, t, t) = 0.5 \times 0.9 \times 0.8 \times 0.9 = 0.324 = P(t, f, t, t)$.

Neka je $N_{PS}(x_1, \ldots, x_n)$ broj povoljnih uzoraka generiranih za događaj $x_1, \ldots, x_n$. Tada imamo

$$\begin{aligned} \lim_{N\to\infty} \hat{P}(x_1, \ldots, x_n) &= \lim_{N\to\infty} N_{PS}(x_1, \ldots, x_n)/N \\ &= S_{PS}(x_1, \ldots, x_n) \\ &= P(x_1, \ldots, x_n) \end{aligned}$$

To znači da su procjene izvedene iz PriorSample konzistentne.

Pokrata: $\hat{P}(x_1, \ldots, x_n) \approx P(x_1, \ldots, x_n)$.

Odbacivanje uzoraka

$\hat{\mathbf{P}}(X|\mathbf{e})$ procjene iz uzoraka slažu se s dokazima $\mathbf{e}$:

```
function REJECTION-SAMPLING(X, e, bn, N) returns an estimate of P(X|e)
   local variables: N, a vector of counts over X, initially zero

   for j = 1 to N do
       x ← PRIOR-SAMPLE(bn)
       if x is consistent with e then
           N[x] ← N[x]+1 where x is the value of X in x
   return NORMALIZE(N[X])
```

Na primjer, procjena $\mathbf{P}(Rain \mid Sprinkler = true)$ iz 100 uzoraka

27 uzoraka ima $Sprinkler = true$

od čega, 8 ima $Rain = true$ i 19 ima $Rain = false$.

$\hat{\mathbf{P}}(Rain \mid Sprinkler = true) = \textsc{Normalize}(\langle 8, 19 \rangle) = \langle 0.296, 0.704 \rangle$.

Slično osnovnim empirijskim procedurama procjene u realnom svijetu!

Analiza odbacivanja uzoraka

Provjera korektnosti (konzistentnosti) uzorkovanja:

$$\begin{aligned}\hat{\mathbf{P}}(X|\mathbf{e}) &= \alpha\, \mathbf{N}_{PS}(X,\mathbf{e}) && \text{(definicija algoritma)}\\ &= \mathbf{N}_{PS}(X,\mathbf{e})/N_{PS}(\mathbf{e}) && \text{(normalizirano s } N_{PS}(\mathbf{e})\text{)}\\ &\approx \mathbf{P}(X,\mathbf{e})/P(\mathbf{e}) && \text{(svojstvo } \textsc{PriorSample}\text{)}\\ &= \mathbf{P}(X|\mathbf{e}) && \text{(definicija uvjetne vjerojatnosti)}\end{aligned}$$

Odbacivanje uzoraka vraća konzistentnu a posteriori procjenu!

Problem: nevjerojatno skupo ako $P(\mathbf{e})$ je malen

$P(\mathbf{e})$ pada eksponencijalno s brojem dokaznih varijabli!

Težinska vjerodostojnost

Ideja: fiksiraj dokazne varijable, uzorkuj samo iz ne-dokaznih varijabli, stavi težinu na svaki uzorak prema vjerodostojnosti slaganja s dokazima

function LIKELIHOOD-WEIGHTING($X, \mathbf{e}, bn, N$) **returns** an estimate of $P(X|\mathbf{e})$
  local variables: $\mathbf{W}$, a vector of weighted counts over X, initially zero

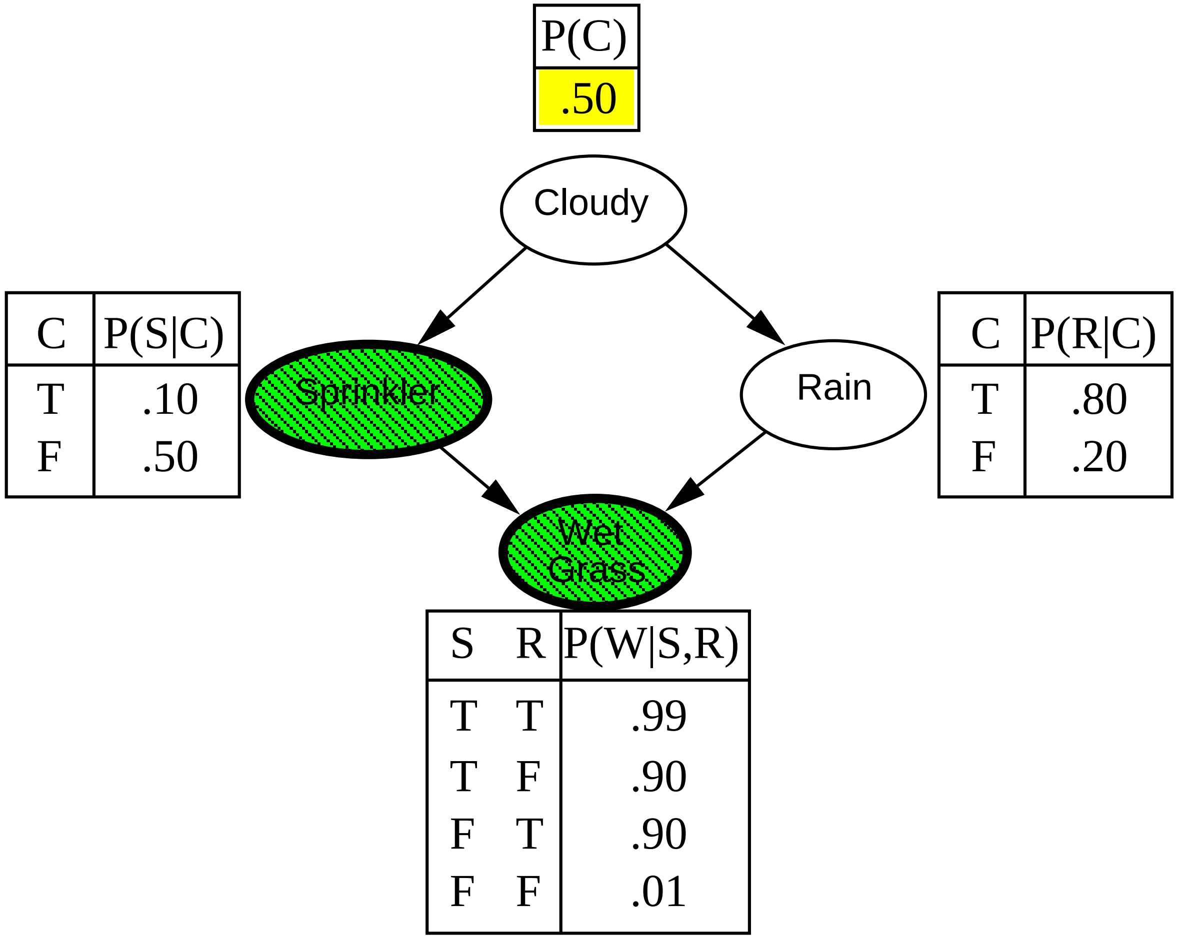
  for $j = 1$ to N **do**
    $\mathbf{x}$, $w \leftarrow$ WEIGHTED-SAMPLE(bn)
    $\mathbf{W}[x] \leftarrow \mathbf{W}[x] + w$ where x is the value of X in $\mathbf{x}$
  return NORMALIZE($\mathbf{W}[X]$)

---

function WEIGHTED-SAMPLE($bn, \mathbf{e}$) **returns** an event and a weight

  $\mathbf{x} \leftarrow$ an event with n elements; $w \leftarrow 1$
  for $i = 1$ **to** n **do**
    if X_i has a value x_i in $\mathbf{e}$
      then $w \leftarrow w \times P(X_i = x_i \mid parents(X_i))$
      else $x_i \leftarrow$ a random sample from $\mathbf{P}(X_i \mid parents(X_i))$
  return $\mathbf{x}$, w

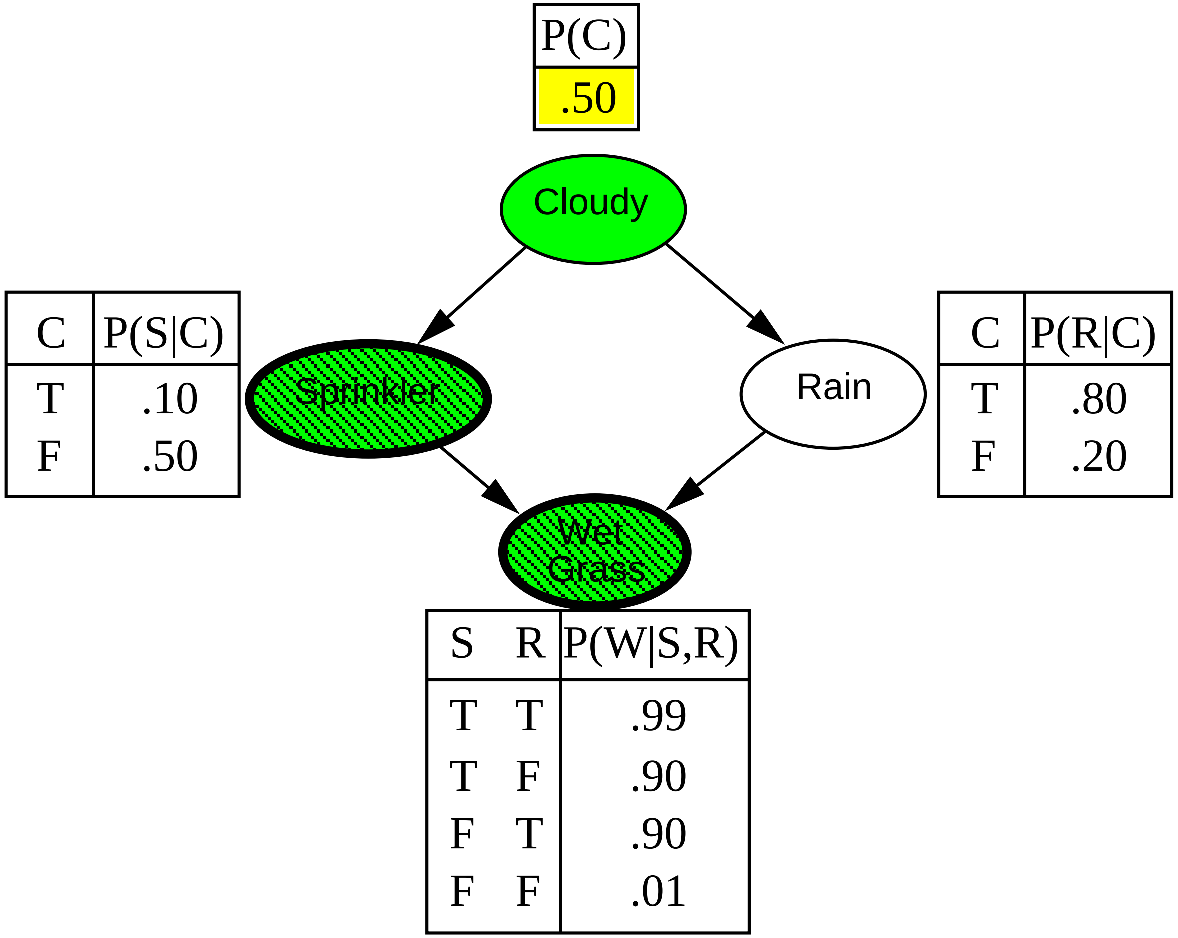
Primjer težinske vjerodostojnosti



$w = 1.0$ (početak), $\quad$ *Cloudy* nije dokazna varijabla

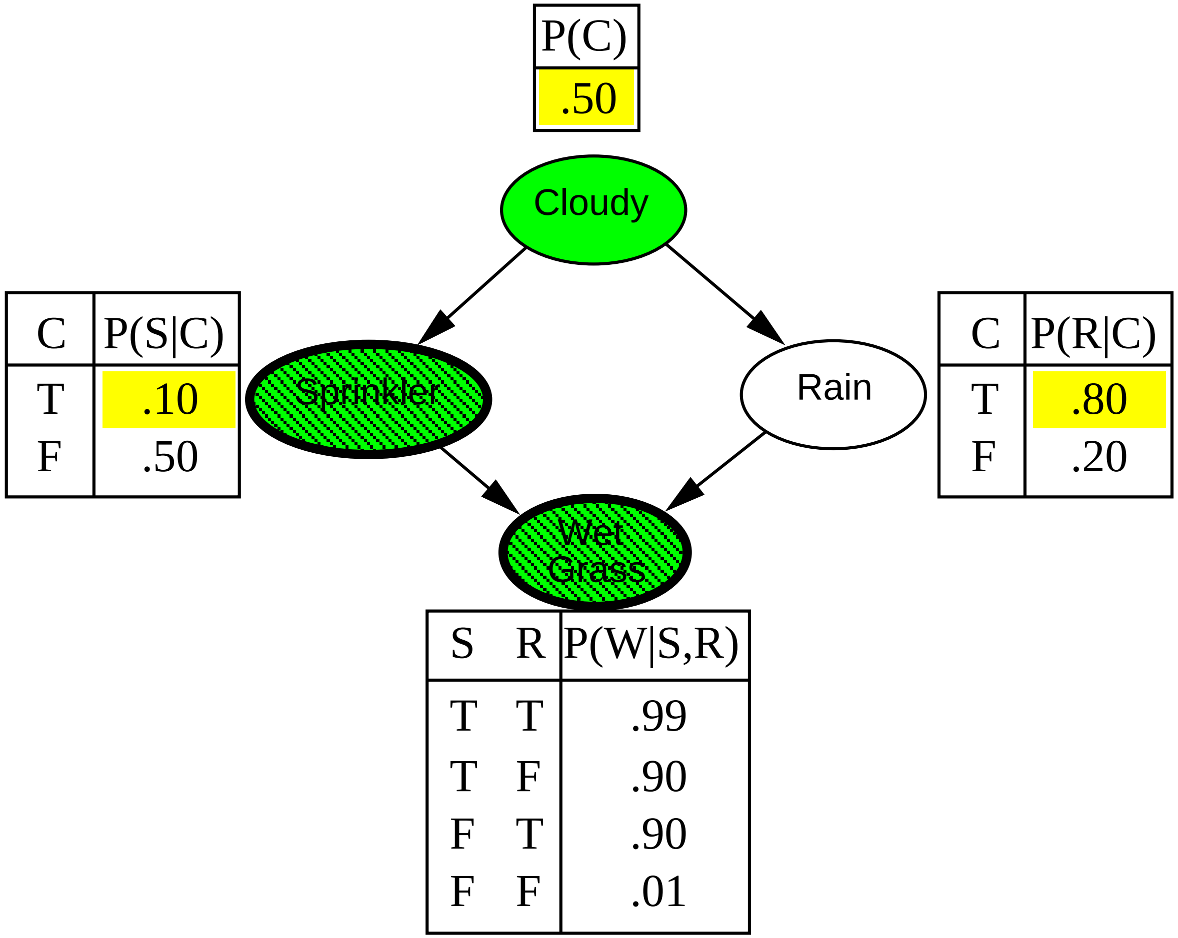
Uzorak iz $\mathbf{P}(Cloudy) = \langle 0.5, 0.5 \rangle$

Primjer težinske vjerodostojnosti



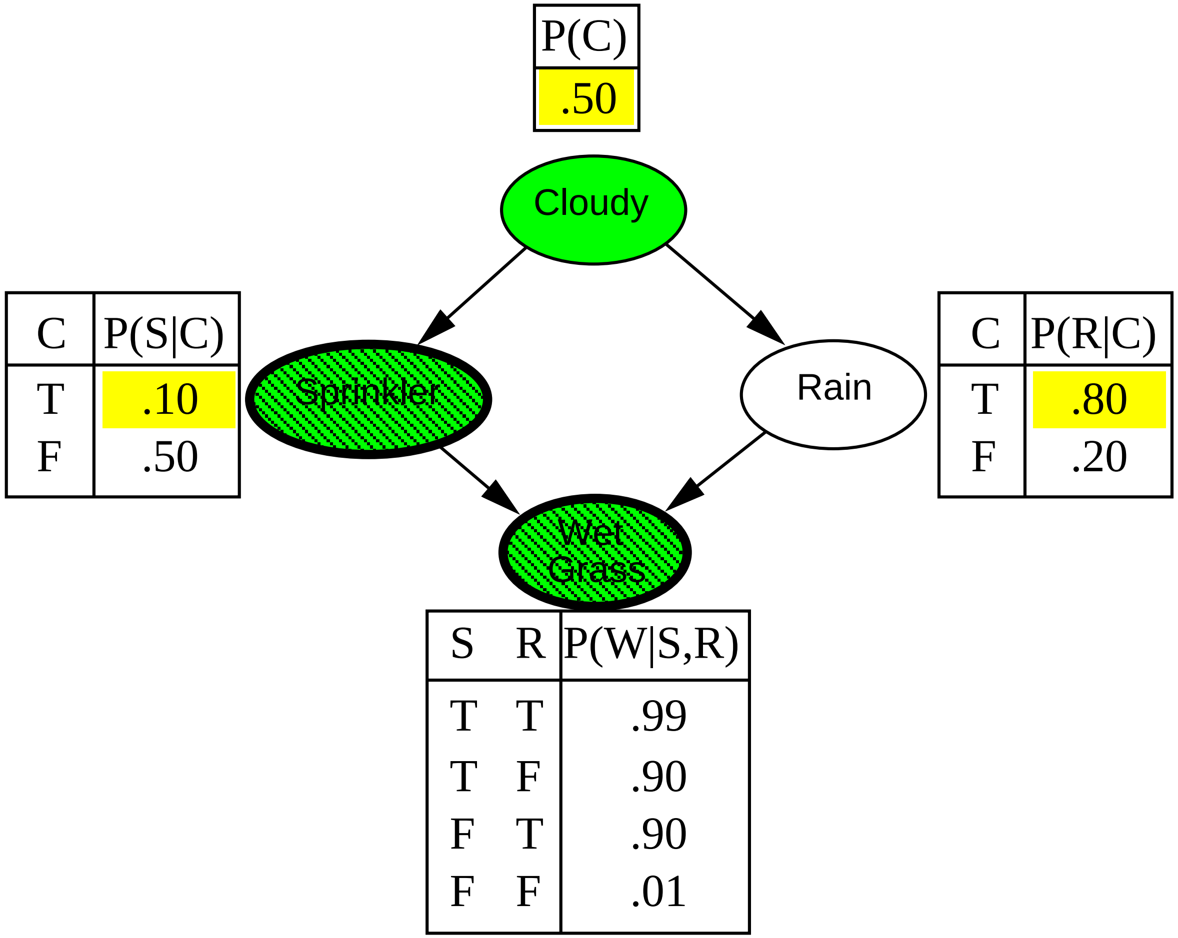
$w = 1.0$,
Uzorak iz $\mathbf{P}(Cloudy) = \langle 0.5, 0.5 \rangle$, neka je $true$.

Primjer težinske vjerodostojnosti



$w = 1.0$, $\quad$ $Sprinkler = true$ je dokaz
modificiraj težinu prema dokazu

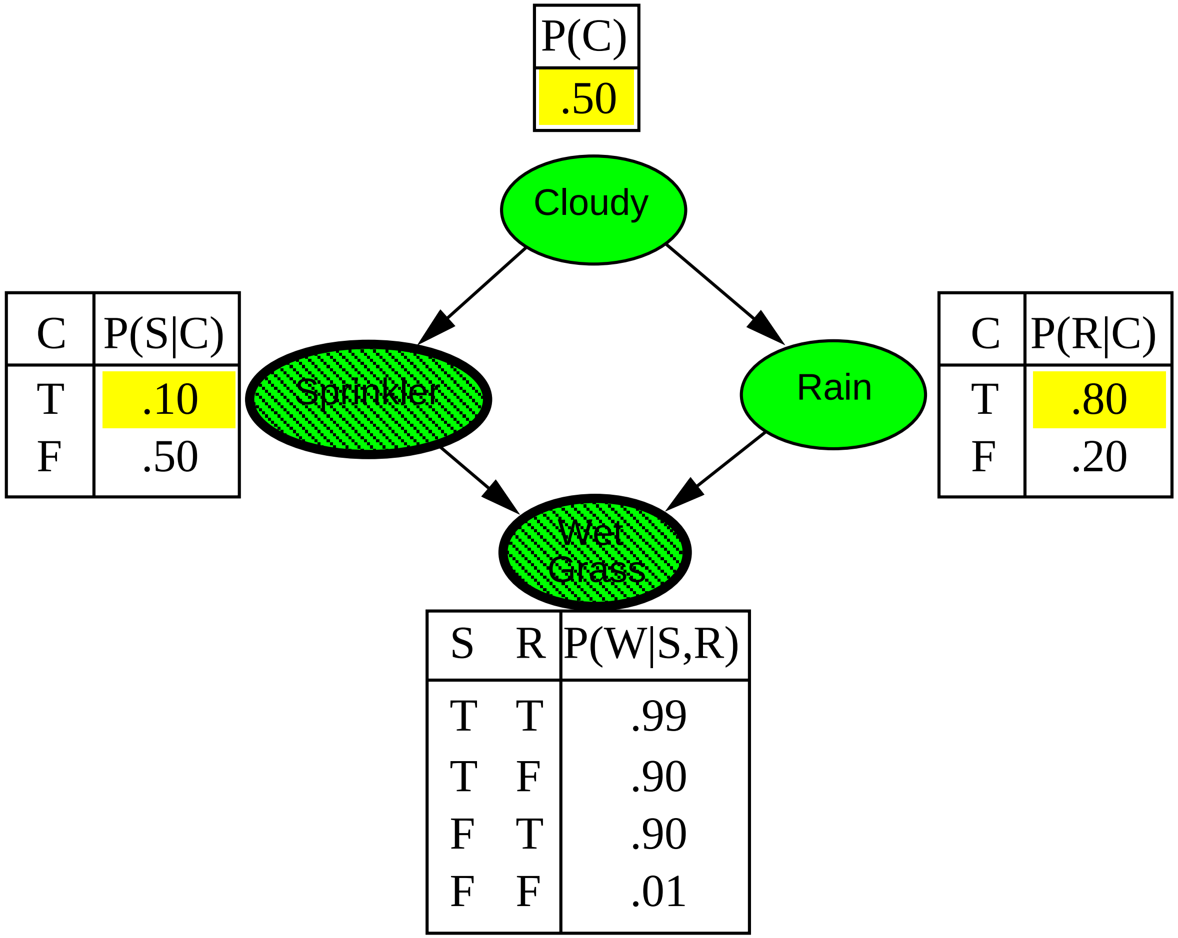
Primjer težinske vjerodostojnosti



$w = 1.0 \times 0.1$, $\quad$ *Rain* nije dokazna varijabla

Uzorak iz $\mathbf{P}(Rain \mid Cloudy = true) = \langle 0.8, 0.2 \rangle$

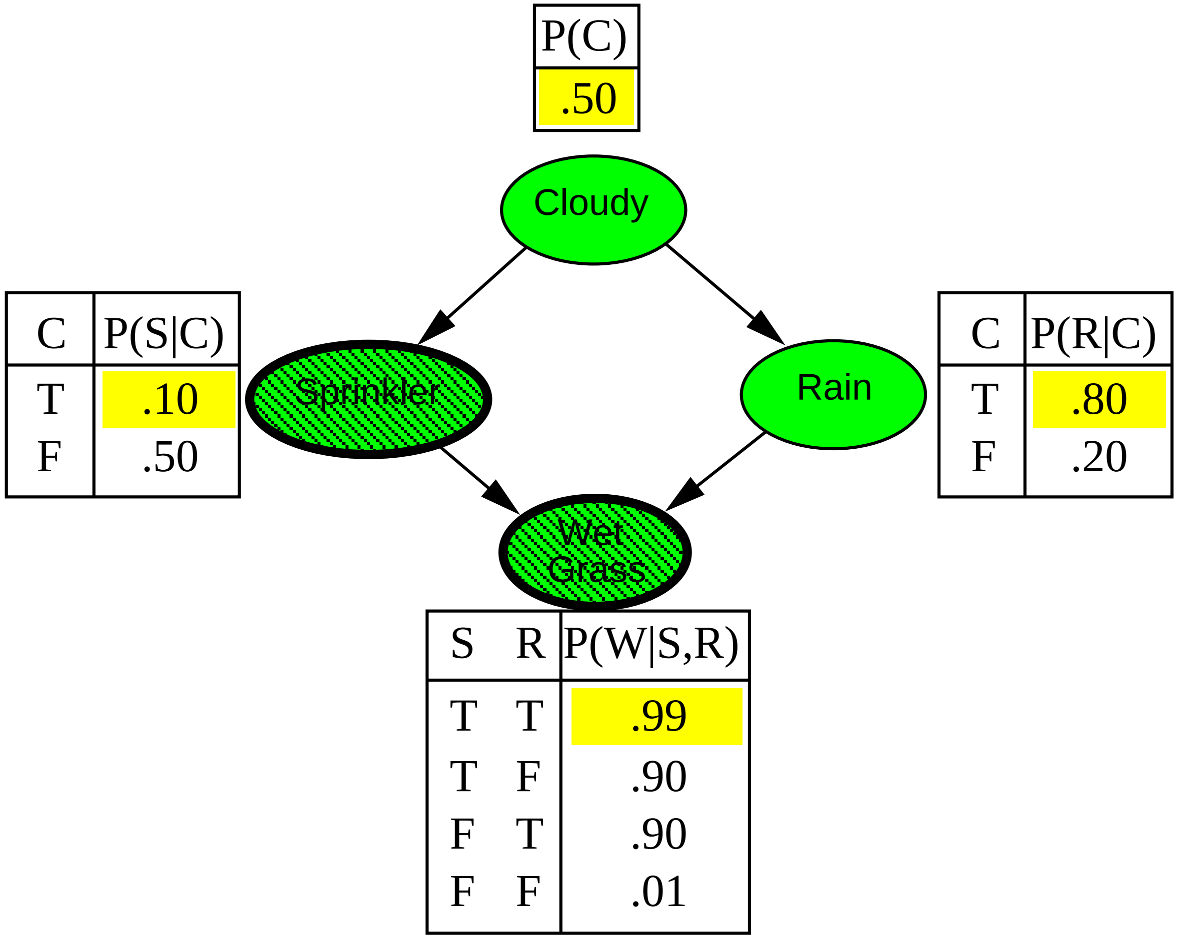
Primjer težinske vjerodostojnosti



$w = 1.0 \times 0.1$,

Uzorak iz $\mathbf{P}(Rain \mid Cloudy = true) = \langle 0.8, 0.2 \rangle$, neka je $true$

Primjer težinske vjerodostojnosti

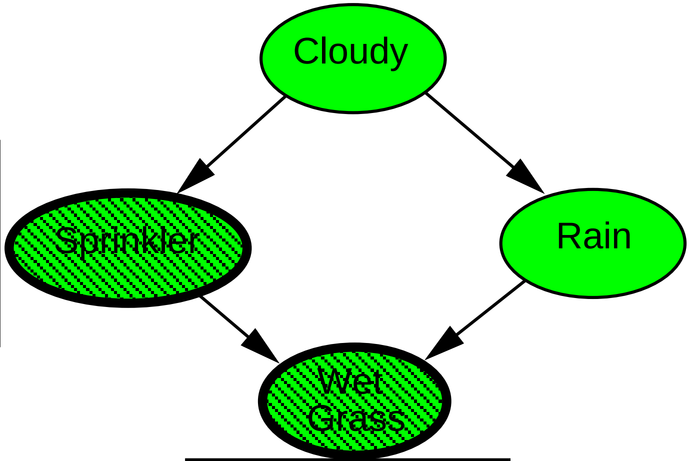


$w = 1.0 \times 0.1$ $\quad$ $WetGrass = true$ je dokaz

modificiraj težinu prema dokazu

Primjer težinske vjerodostojnosti

P(C)
.50



C	P(S|C)
T	.10
F	.50

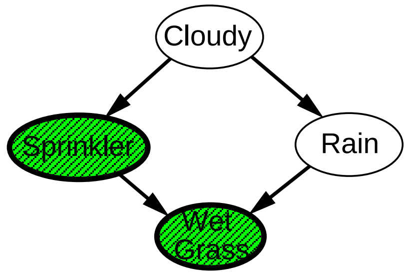
C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

$w = 1.0 \times 0.1 \times 0.99 = 0.099$

Analiza težinske vjerodostojnosti

Uzoračka vjerojatnost za WeightedSample je



$$S_{WS}(\mathbf{z}, \mathbf{e}) = \prod_{i=1}^{l} P(z_i \,|\, parents(Z_i))$$

gdje $\mathbf{Z}$ označava nedokazne varijable, uključivo i upit X.

Uočiti: uzima u obzir samo dokaze u pretcima (bez djece)

$\Longrightarrow$ negdje "između" apriorne i aposteriorne distribucije!

Bez težina — procjene nisu konzistentne!

Moraju dobiti "težinu" za dokaze — prema danim roditeljima.

Analiza težinske vjerodostojnosti (nastavak)

Težina za dani uzorak $\mathbf{z}, \mathbf{e}$ je

$$w(\mathbf{z}, \mathbf{e}) = \prod_{i=1}^{m} P(e_i \,|\, parents(E_i)).$$

Težinska uzoračka vjerojatnost

$$
\begin{aligned}
S_{WS}(\mathbf{z}, \mathbf{e}) w(\mathbf{z}, \mathbf{e}) & \\
&= \prod_{i=1}^{l} P(z_i \,|\, parents(Z_i)) \ \prod_{i=1}^{m} P(e_i \,|\, parents(E_i)) \\
&= P(\mathbf{z}, \mathbf{e}) \qquad \text{(std. globalna semantika mreže).}
\end{aligned}
$$

Težinska vjerodostojnost vraća konzistentne procjene, jer
Z i e pokrivaju sve varijable u mreži.

Performansa LW ipak pada s mnogo dokaznih varijabli,
jer samo nekoliko uzoraka ima skoro cijelu ukupno težinu.

Aproksimativno zaključivanje korištenjem MCMC

"Stanje" mreže = trenutne dodijeljene vrijednosti za sve varijable.

Novo stanje generira se uzorkovanjem jedne varijable za dani Markovljev pokrivač. Na primjer, u petlji, uzorkuj redom sve, osim dokaza (fiksni)!

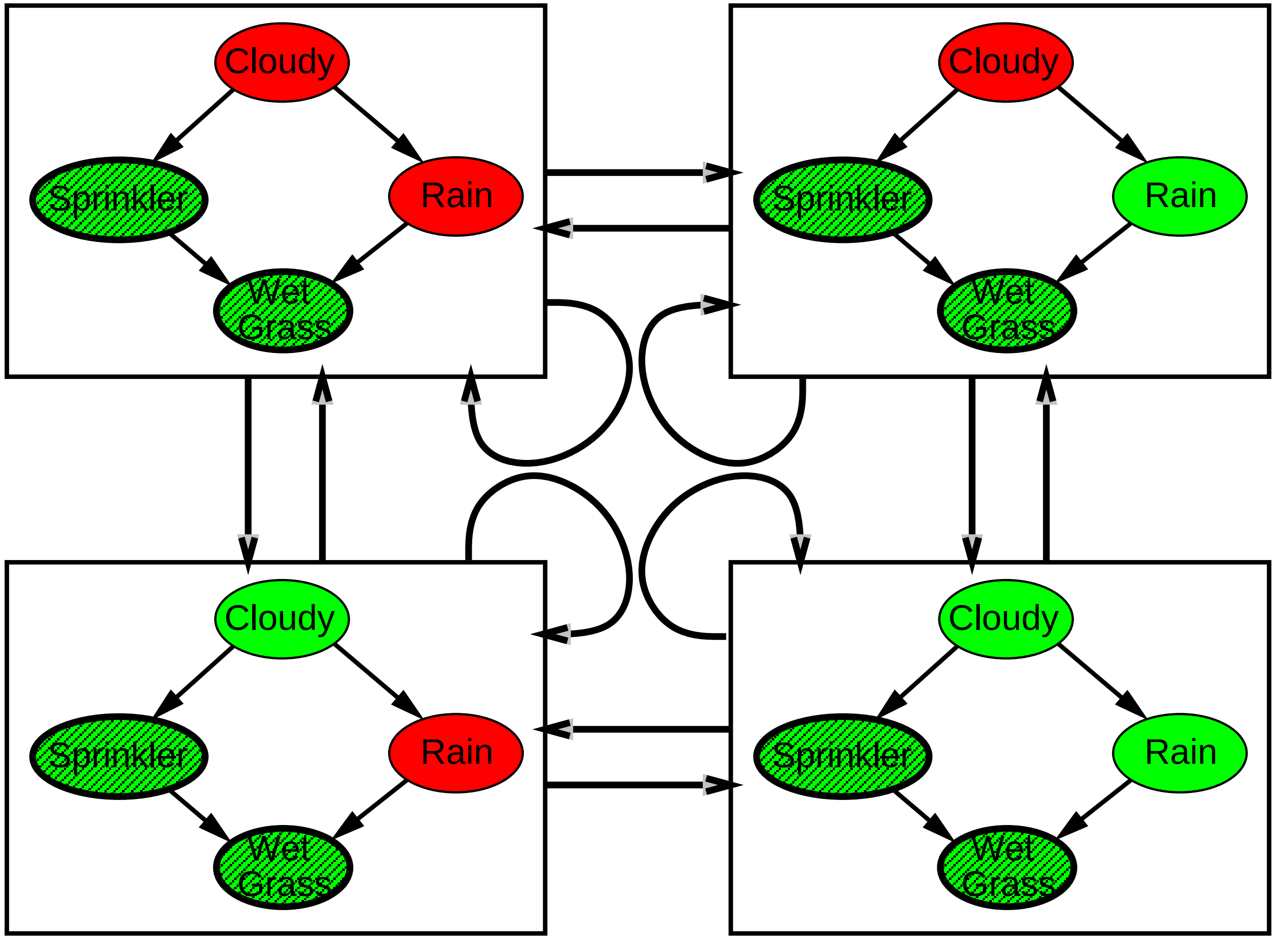
function MCMC-ASK($X, \mathbf{e}, bn, N$) **returns** an estimate of $P(X|\mathbf{e})$
  local variables: $\mathbf{N}[X]$, a vector of counts over X, initially zero
    $\mathbf{Z}$, the nonevidence variables in bn
    $\mathbf{x}$, the current state of the network, initially copied from $\mathbf{e}$

  initialize $\mathbf{x}$ with random values for the variables in $\mathbf{Y}$
  for $j = 1$ to N **do**
    for each Z_i in $\mathbf{Z}$ **do**
      sample the value of Z_i in $\mathbf{x}$ from $\mathbf{P}(Z_i|mb(Z_i))$
        given the values of $MB(Z_i)$ in $\mathbf{x}$
      $\mathbf{N}[x] \leftarrow \mathbf{N}[x] + 1$ where x is the value of X in $\mathbf{x}$
  return NORMALIZE($\mathbf{N}[X]$)

Napomena: Može se i slučajno izabrati varijabla koja se uzorkuje.

Markovljev lanac

$Sprinkler = true, WetGrass = true$, ima četiri stanja:



Putujte neko vrijeme, uzmite prosjek onog što vidite!

MCMC primjer (nastavak)

Procjena $\mathbf{P}(Rain|Sprinkler = true, WetGrass = true)$

Ponovljeno uzorkuj $Cloudy$ ili $Rain$ za njihov Markovljev pokrivač.
Izbroji koliko puta je $Rain$ istinito ili lažno u uzorcima.

Na primjer, u 100 stanja
31 ima $Rain = true$, 69 ima $Rain = false$

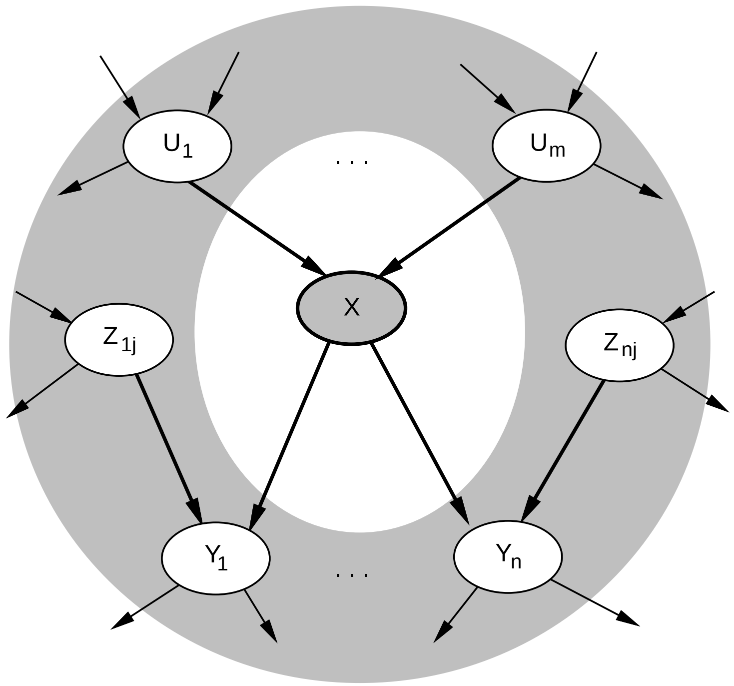
$\hat{\mathbf{P}}(Rain|Sprinkler = true, WetGrass = true)$
$= \textsc{Normalize}(\langle 31, 69\rangle) = \langle 0.31, 0.69\rangle$

Teorem: ulančani pristup stacionarnoj distribuciji:
uz dugotrajno ponavljanje, proporcija vremena u svakom stanju proporcionalna je a posteriori vjerojatnosti

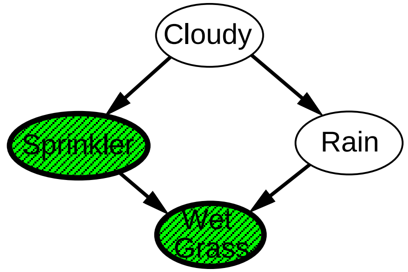
Markovljev pokrivač — ponavljanje

Markovljev pokrivač od X — oznaka $MB(X)$, (označen sivo na slici)
$=$ njegovi roditelji $+$ djeca $+$ roditelji djece.

Svaki čvor X je uvjetno nezavisan od svih ostalih čvorova u mreži, ako su dane (kao uvjet/dokaz) vrijednosti varijabli iz $MB(X)$.
Oznaka za dane vrijednosti tih varijabli je $mb(X)$.



Uzorkovanje Markovljevog pokrivača



Markovljev pokrivač za $Cloudy$ je
$Sprinkler$ i $Rain$
Markovljev pokrivač za $Rain$ je
$Cloudy$, $Sprinkler$ i $WetGrass$

Za Markovljeve pokrivače vjerojatnosti se računaju ovako:

$$P(x_i' \mid mb(X_i)) \\ = P(x_i'|parents(X_i)) \prod_{Z_j \in Children(X_i)} P(z_j|parents(Z_j))$$

Dakle, za promjenu vrijednosti neke varijable X_i, broj množenja u gornjoj formuli je jednak broju djece od X_i.

Glavni problemi u računanju:
1) Teško je reći da li je došlo do konvergencije
2) Može biti rasipno ako je Markovljev pokrivač velik:
$P(X_i|mb(X_i))$ se neće jako promijeniti (zakon velikih brojeva)

Sažetak

Egzaktno zaključivanje elimacijom varijabli:
- polinomno vrijeme na poli-stablima, NP-teško na općim grafovima
- prostor = vrijeme, vrlo osjetljivo na topologiju

Aproksimativno zaključivanje korištenjm LW, MCMC:
- LW je loše ako ima mnogo (engl. downstream) dokaza
- LW, MCMC općenito neosjetljivi na topologiju
- Konvergencija može biti vrlo spora s vjerojatnostima blizu 1 ili 0
- Rade za proizvoljnu kombinaciju diskretnih i neprekidnih varijabli