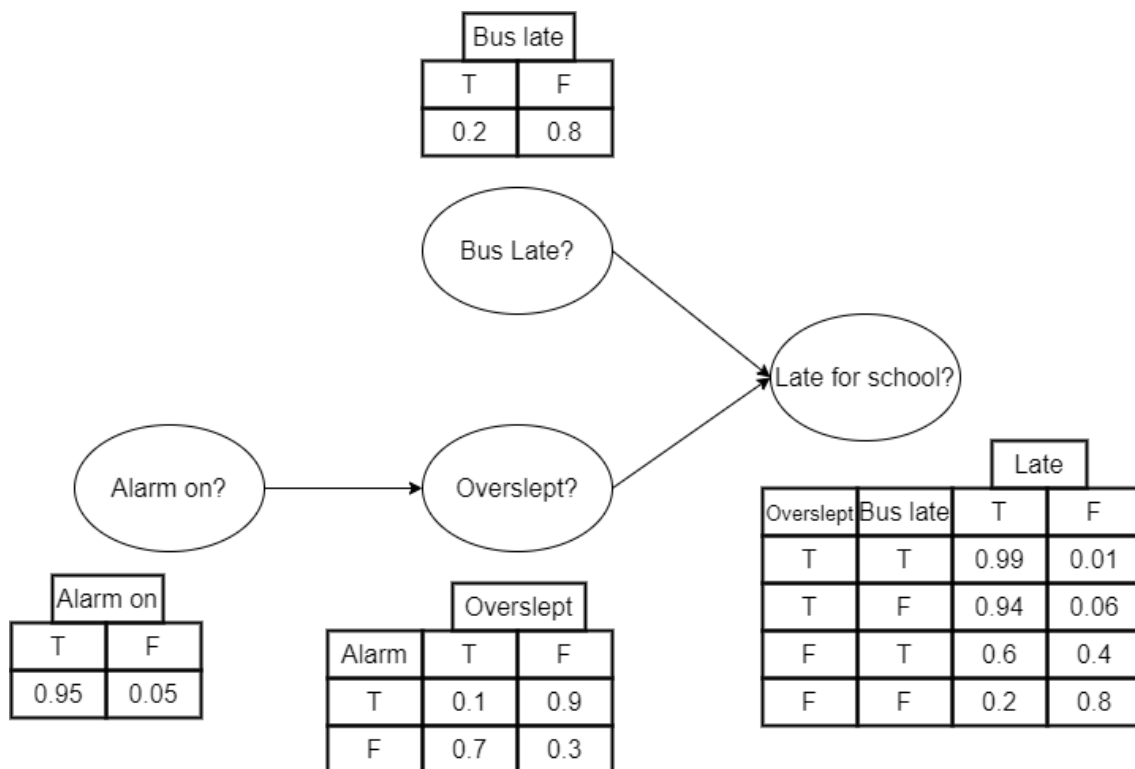


Napisać losowy generator danych działający zgodnie z rozkładem reprezentowanym przez daną sieć Bayes'a. Sieć powinna opisywać zależności między zmiennymi losowymi (binarne: 0/1) Na wejściu dany jest opis grafu połączeń/zależności między zmiennymi oraz tabele prawdopodobieństw warunkowych. Format danych wejściowych jest dowolny, może być i z pliku .py, ale nie jako `MyCustomFancyGraphClassThatIWillForgetHowToUseSoon(1, a, 2, 84)`... tylko raczej `{'a': 'b', ...}, {'a': [0, 1], 'b': [[0,1],[1,0]]...}` Wygenerować X próbek np. w liczbie [100, 1000, 10000] i sprawdzić czy są zgodne z rozkładem. Problem należy samemu zdefiniować, ale powinien zawierać kilka zmiennych i co najmniej jedną podwójną zależność.

Założenia wstępne

Zdefiniowano problem prawdopodobieństwa spóźnienia do szkoły, który reprezentuje poniższa sieć Bayesa.



Rys. 1 – sieć Bayesa reprezentująca problem spóźnienia do szkoły

Dane wejściowe

Dane wejściowe reprezentujące sieć Bayesa pobierane są z pliku .py. Należy przekazać listę węzłów. Przykładowy taki węzeł jest zmienną typu `dict` i został przedstawiony na rysunku 2. Węzeł ma następujące pola:

- „name” – nazwa węzła
- „dependencies” – lista węzłów, które są rodzicami danego węzła
- „prob_table” – tablica prawdopodobieństw dla tego węzła

```

"name": "Late for school", "dependencies": ["Overslept", "Bus late"], "prob_table":
[
  [True, True, True, 0.99],
  [True, True, False, 0.01],
  [True, False, True, 0.94],
  [True, False, False, 0.06],
  [False, True, True, 0.6],
  [False, True, False, 0.4],
  [False, False, True, 0.2],
  [False, False, False, 0.8]
]

```

Rysunek 2

Wyniki eksperymentów

Generator próbek działa na zasadzie propagacji zmiennych przez sieć. Tą metodą wygenerowano 100, 1000 i 10000 próbek losowych postaci:

{„Alarm on”: bool, „Overslept”: bool, „Bus late”: bool, „Late for school”: bool}.

Następnie korzystając z tablic prawdopodobieństw ze zdefiniowanej sieci obliczono prawdopodobieństwa wystąpienia każdej możliwej próbki (czyli 16 prawdopodobieństw, bo próbka ma 4 zmienne – 16 próbek). Prawdopodobieństwa wystąpienia próbek sumują się do 100%, więc zostały wyznaczone poprawnie. Wyniki eksperymentów i obliczone prawdopodobieństwa przedstawiono w tabeli 1.

Nr próbki	% wystąpień w 100 próbkach	% wystąpień w 1000 próbkach	% wystąpień w 10000 próbkach	Prawdopodobieństwa wystąpienia próbki [%]
1	1	2,1	1,74	1,881
2	0	0	0	0,019
3	10	10,3	10,21	10,26
4	11	7,5	6,57	6,84
5	2	0,4	0,73	0,693
6	0	0	0,02	0,007
7	0	0,1	0,2	0,18
8	1	0,1	0,13	0,12
9	6	7,6	7,27	7,144
10	0	0,5	0,46	0,456
11	12	13,7	14,08	13,68

12	55	53,9	54,7	54,72
13	1	2,7	2,64	2,632
14	0	0,3	0,19	0,168
15	0	0,2	0,21	0,24
16	1	0,6	0,85	0,96

Tabela 1

Z tabeli wynika, że generator próbek został zaimplementowany poprawnie. Procentowe wystąpienia danej próbki są zbliżone do prawdopodobieństwa wylosowania danej próbki. Wraz ze wzrostem wygenerowanych próbek te dwie wartości stają się coraz bliższe sobie.