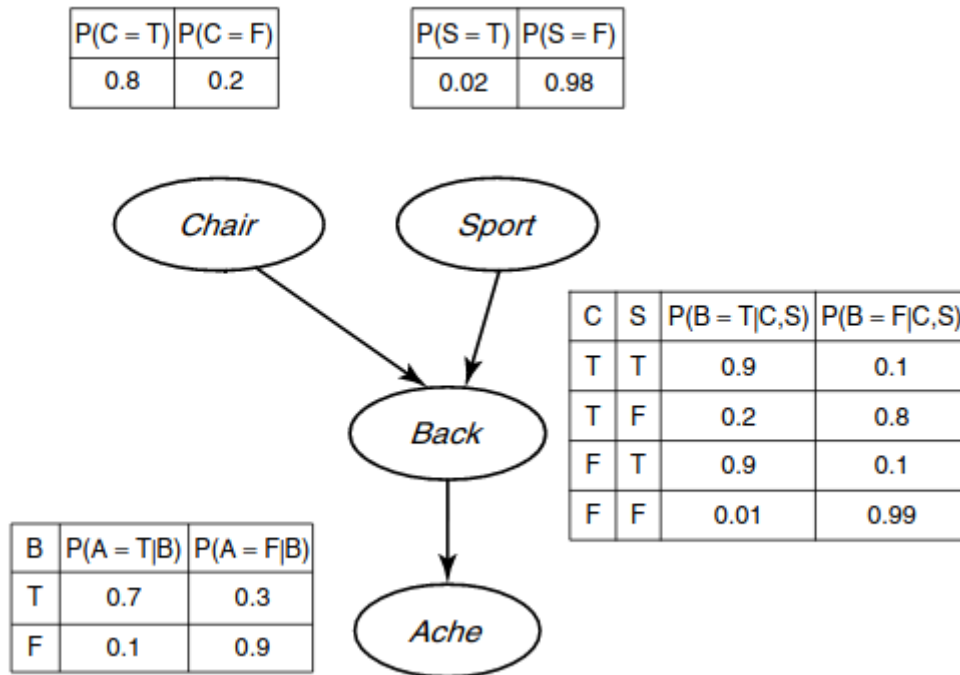


# WSI – Ćwiczenie 7

## Zadanie

Proszę zaimplementować losowy generator danych, który działa zgodnie z rozkładem reprezentowanym przez daną sieć bayesowską.



Sieć ta opisuje zależności między (zero-jedynkowymi) zmiennymi losowymi i dana jest w postaci opisu grafu połączeń oraz tabel prawdopodobieństw warunkowych. Wejście algorytmu: ile przykładów wygenerować, opis struktury prostej sieci (według własnego formatu) oraz tabele prawdopodobieństw należy wczytać z pliku tekstowego. Wyjście: plik tekstowy z przykładami. Strukturę sieci i tabele prawdopodobieństw widać na rysunku. Klasa to „Ache” (czy bolą plecy), pozostałe węzły to atrybuty („Back” to uszkodzenie kręgosłupa (drobne, czasem nie skutkujące bólem)). Wytworzony zbiór podzielić i użyć do treningu i testowania klasyfikatora utworzonego na wcześniejszych ćwiczeniach. Jakie uzyskujemy wyniki? Wnioski?

## Wyniki

Zaimplementowano losowy generator danych dla rozkładu reprezentowanego przez podaną w poleceniu sieć bayesowską, a następnie użyto tych danych do treningu i testowania klasyfikatora ID3 utworzonego na ćwiczeniach nr 4. Otrzymano następujące wyniki dla zbiorów składających się z 10 000 i 100 000 próbek:

[ Zbiór danych o wielkości 10 000 ] => Dokładność 80,35%

Przewidywany / Rzeczywisty	T	F
T	3214	786
F	0	0

[ Zbiór danych o wielkości 100 000 ] => Dokładność 79,89%

Przewidywany / Rzeczywisty	T	F
T	31955	8044
F	1	0

## Wnioski

W otrzymanych wynikach zauważalnie sporą część wskazań (ok. 20%) stanowią błędy 1-go rodzaju (false positive), czyli stwierdzamy, że plecy bolą, gdy tak naprawdę nie bolą. Popęlnienie błędu 2-go rodzaju (false negative), czyli że plecy nie bolą, choć w rzeczywistości bolą, jest bardzo rzadkie i dopiero przy próbie ok. 100 000 pojawił się jeden taki przypadek.